Informatics in Medicine Unlocked 9 (2017) 68–75

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu

Multi-template homology based structure prediction and molecular docking studies of protein 'L' of Zaire ebolavirus (EBOV)

Jayasree Ganugapati *, Sai Akash

Department of Biotechnology, Sreenidhi Institute of Science and Technology, Ghatkesar, Hyderabad, Telangana 501301, India

ARTICLE INFO

Keywords:
EBOV
Large Structural Protein – L
EBOLA
Auto Dock

ABSTRACT

Ebola is one of the most dangerous pathogenic RNA virus that causes severe hemorrhagic fever in humans and is considered to be a threat to humanity. The RNA genome of EBOV encodes seven proteins viz., glycoprotein (GP), nucleoprotein (NP), RNA-dependent RNA polymerase (protein 'L'), VP35, VP30, VP40 andVP24. The objective of the present study is to find a suitable inhibitor for protein 'L'. The large structural protein 'L', is made up of 2212 amino acid residues. This protein works as an RNA-dependent RNA polymerase (RdRp) and a methyl transferase. It is carried by the virus during the infection as the host mechanisms cannot be used to transcribe the –ss RNA genome of the virus. As the protein is crucial for the replication of the viral genome and no other host enzyme can perform the same function, this viral protein 'L' was considered as a potential drug target to design inhibitors. The 3D structure of protein 'L' is not available to date. This is a limitation in understanding the protein's function. Hence, the present work is aimed at predicting the first homology-based model of protein 'L' and elucidating the function by providing insight into the molecular details of the protein. As there is no drug available for the treatment of EBOV infection our findings play a crucial a role to identify an inhibitor of the protein 'L' of EBOV. HTS against ZINC database resulted in identification of few possible inhibitors. Molecular docking studies resulted in finding a suitable inhibitor for protein 'L'.

1. Introduction

Ebola is one of the most dangerous and unknown of all the pathogenic viruses and is considered to be a threat to humanity. This virus has the fatality rate of over 90% and mortality rate of over 79% [1]. This viral infection has no vaccine and neither any effective cure as it pathogenicity has not been understood properly [2]. The natural reservoirs of this virus are considered to be fruit bats, which spreads the infection to humans, non-human primates and other mammals through a zoonotic cycle. This virus can spread to humans on the consumption of improperly cooked bush meat [3].

The genus Ebolavirus has five species Bundibugyo ebolavirus, Reston ebolavirus, Sudan ebolavirus, Tai forest ebolavirus and Zaire ebolavirus. All these species of Ebola virus cause hemorrhagic fever with varying mortality and fatality rates. The first recorded infection was in the year 1976 in Sudan by Sudan ebolavirus and the most recent in 2014 in Liberia and spread to Sierra Leone, Nigeria, Guinea even to the USA and Spain [4]. EBOV has a negative-stranded ssRNA genome as per the genetic studies conducted on this virus. The genome is around 18,959 nt long [5]. The genome consists of seven genes that code for seven proteins, each of these proteins has a specific function and some proteins perform multiple functions to complete its life cycle. As the virus has a –ssRNA, it must carry its own polymerase to replicate its genome and transcribe its genome into an mRNA. This process can be explained by Baltimore scheme for viral genome replication [6].

The genome of EBOV, is 18,959 nucleotides in length and contains seven transcriptional units that produce primary translation products: the nucleoprotein (NP), virion protein (VP) 35, VP40, glycoprotein (GP), soluble glycoprotein (sGP), small soluble glycoprotein (ssGP), VP30, VP24, and the large protein 'L' [7]. The matrix protein of this virus is VP40 which helps the virus in budding and assembly into a new virion [8]. The viral protein VP24 is a peripheral viral membrane protein that helps in viral binding and also plays an important role in the suppression of host interferon activity [9]. There exist two more proteins in EBOV that are responsible for the suppression of the host immune system and tolerating host immune responses, they are VP30 and VP 35. The protein

* Corresponding author.
E-mail address: gpljayasree@sreenidhi.edu.in (J. Ganugapati).

http://dx.doi.org/10.1016/j.imu.2017.06.002
Received 31 January 2017; Received in revised form 24 June 2017; Accepted 24 June 2017
Available online 27 June 2017

VP35 is also known to bind to dsRNA [10]. NP is a ribonucleoprotein that binds with the viral genome in the virion and plays a major role in releasing the viral genome when the endosome is opened [11]. EBOV, viral RNA synthesis requires the viral NP, VP35, VP30, and L proteins. 'L' is the catalytic subunit of the viral polymerase complex in the host, the viral RNA polymerase (protein 'L') begins to copy the negative strand (–ve) RNA to make the positive strand (+ve) transcripts which are translated by host ribosomes.

The objective of the present study is to find a suitable inhibitor for protein 'L'. The large structural protein 'L', is made up of 2212 amino acid residues. This protein works as an RNA-dependent RNA polymerase (RdRp) and a methyl transferase [12]. It is carried by the virus during the infection as the host mechanisms cannot be used to transcribe the –ss RNA genome of the virus. As the protein is crucial for the replication of the viral genome and no other host enzyme can perform the same function, this viral protein 'L' was considered as a potential drug target to design inhibitors. The 3D structure of protein 'L' is not available to date. This is a limitation in understanding the protein's function. Hence, the present work is aimed at predicting the first homology-based model of protein 'L' and elucidating the function by providing insight into the molecular details of the protein. As there is no drug available for the treatment of EBOV infection, our findings play a crucial a role to identify an inhibitor of the protein 'L' of EBOV.

Computational methods were used to predict the structure of the protein based on protein threading and homology modelling.

2. Material and methods

2.1. Sequence retrieval

The amino acid sequence of the Large Structural Protein (protein 'L') of Zaire Ebola virus was retrieved from NCBI (https://www.ncbi.nlm.nih.gov/genbank). The protein sequence with the Accession number 'Q6V1Q2.1' was selected for the present study.

2.2. Protein threading and multi-template homology modelling

Homology modelling of the protein 'L' was carried out through Modeller 9.14 [13]. As the complete protein did not have enough similarity with other proteins and length of the protein was too long (2212 residues), structure prediction of the whole protein was performed by protein threading and the protein was truncated into 8 fragments, keeping the functional domains intact. Each fragment was between 150 and 400 amino acid residues long. Blast search for each of the fragment was done against the PDB database to ensure that these protein fragments had enough similarity with existing proteins so that their structure can be predicted. As these protein fragments had a good similarity with existing proteins whose structures were known, they were submitted to the I-TASSER [14,15], server for the prediction of their 3D structures. I-TASSER produced 5 models for every fragment and the best model for each fragment was selected based on the confidence score. These were then submitted to Mod Refiner [16] for refinement. The 3D structures of these 8 fragments were then used as the templates for modelling the complete protein using Modeller 9.14. The structure files of the fragments and the alignment files were used as templates for Homology modelling to generate an initial model.

2.3. Model refinement and evaluation

The 3D structure of the protein initial model was then submitted to PDB-sum (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) maintained by EBI. Ramchandran plot was generated by PROCHECK [17], which highlight the residues in the outlier region. These residues in the outlier region were then loop modelled through 'scan loop database' of SPDBV [18]. 3D structure of the protein was then submitted to PDB-Sum after loop modelling to assess the quality of the final model.

2.4. Energy minimization, molecular dynamics simulations and active-site prediction

Protein models produced through homology modelling are not very accurate. The structure of the protein model can be refined using Molecular Dynamics (MD) simulations. The final model of the protein 'L' was then subjected to Energy Minimization (EM) and MD Simulations using GROMACS 5.0.2 [19] on a Linux OS based computer. The simulation was performed in a cubic box containing the protein solvated in water molecules using OPLS/AA force field. In order to describe long-range interaction in the simulations, eight Cl− ions were added to neutralise the extra positive charges. The temperature of the simulation was kept constant at 303 K and pressure at 1 atm. EM was performed using the steepest descent algorithm with the energy step size at 0.01 s and a maximum of 50,000 steps to be performed. The resultant potential energy graph was visualized through Xmgrace. For MDS the maximum number of steps was set to 500,000, the step size was 0.002 s and the coordinates of the protein were noted every 10 ps throughout the simulation time of 1 ns. The resultant backbone RMSD graph was visualized using Xmgrace.

2.5. Active site analysis

Active site of the protein was predicted through RESIDUE DEPTH (http://cospi.iiserpune.ac.in/depth/htdocs/index.html) a server maintained by IISER Pune. The protein was submitted to the DEPTH server to predict the binding site, where the solvent neighbourhood radius was set to a default of 4.2 Å.

2.6. Virtual screening and toxicity studies

HTS was performed using DOCKblaster [20] which is a structure based ligand discovery tool maintained by University of California, San Francisco. The protein was submitted to the server and the active site predicted by DEPTH was picked from the possible ligand binding sites shown by DOCKblaster. That active site of the protein was docked against the clean-fragment subset of the ZINC database.

The toxicity of the ligand molecules was studied using OSIRIS property calculator (http://openmolecules.org/propertyexplorer/index.html). It is an online server that reveals various properties of ligand molecules like Mutagenic, Tumorigenic, Irritant, Reproductive effective, Clogp value, Solubility, Molecular weight, Drug-likeness and finally the Drug score.

2.7. Molecular docking

MGL tools 1.5.7 [21–22] and Auto Dock Vina [23] were used for Molecular Docking. The pdb structures of the protein and the ligand opened on MGL tools 1.5.7 and were converted into .pdbqt files, which were then used for molecular docking. Auto dock Tools 4.2 was used to add partial charges by Gasteiger method and polar hydrogen atoms were added to the protein that was saved in pdbqt format. The ligand was prepared in pdbqt by setting flexible.

Torsion angles at all rotatable bonds, while the protein is kept as a rigid structure. The grid with X, Y and Z centre of 21.45, 32.359, and 15.701 was created around the active site of the protein. The size of the grid was X*Y*Z = 40*40*40. Lamarckian Genetic Algorithm (LGA) was used to obtain ten conformations. Best hits were selected based affinity energy (kcal/mol) and the Root-Mean-Square Deviation (RMSD) < 2.0 Å, score. The docked complexes were visualized on Accelrys Discovery Studio 4.1.

3. Results

3.1. Protein sequence retrieval

The sequence of the protein was retrieved from NCBI. The search for