INDO AMERICAN JOURNAL OF PHARMACEUTICAL RESEARCH

IN SILICO PREDICTION AND DOCKING OF TERTIARY STRUCTURE OF PROTEIN X, MULTIFUNCTIONAL PROTEIN OF HEPATITIS B VIRUS (HBV)

Sharav Desai[*], Pooja Tahilramani[1], Dhara Patel[2], Dhananjay Meshram[2], Prachi Patel[1]

[1]*Department of Pharmaceutical Microbiology and Biotechnology, Pioneer Pharmacy Degree College, Sayajipura, Vadodara-390019, Gujarat, India.*

[2]*Department of Quality Assurance, Pioneer Pharmacy Degree College, Sayajipura, Vadodara-390019, Gujarat, India.*

ARTICLE INFO

Article history
Received 03/09/2017
Available online
30/09/2017

Keywords
Protein X,
Homology,
Docking,
Uniprot,
In Silico,
Hepatitis B,
Autodock,
Pyrx0.8.

ABSTRACT

Hepatitis B virus (HBV) infection is a universal health problem and may result in acute, fulminant, chronic hepatitis liver cirrhosis, or hepatocellular carcinoma (HCC). The sequence for Protein X of hepatitis B virus was retrieved from UniProt database ProtParam from ExPAsy server was used to investigate the physicochemical properties of the protein. Homology modeling was carried out using Phyre2 server, and Refinement studies were done with Galaxy web browser.Five models were generated and evaluated by ERRAT, ANOLEA, QMEAN6, and PROCHECK. Antigenicity of the protein was also assessed by Chou & Fasman Beta-Turn Prediction method. Five models were generated, and model 1 was having the greatest quality by the QMEAN6 score with 0367 ERRAT analysis reveals the overall quality of 54054% whereas the initial model was having only 17730% quality The mean force potential energy, as analyzed by ANOLEA, were better compared to the original model. Stereochemical quality estimation by Procheck showed that the refined Model 1 had a reliable structure, and was therefore submitted to the protein model database PyRx with Autodock vina was used to screen the compounds from Drug bank and Protein Data Bank to find the molecules that can bind to the active site between 1 to 142 amino acids.Ten compounds with highest negative energy were selected as lead molecules. Protein X structurally evaluated computational methods can be further be experimentally and clinically investigated to develop as potential drug target in Hepatitis B infection. The steps utilized in the article can also be used to investigate the drug target for other infection.

Corresponding author
Dr. Sharav Desai
Department of Pharmaceutical Microbiology and Biotechnology,
Pioneer Pharmacy Degree College,
Ajwa-Nimeta Road,
Sayajipura. Vadodara, Gujarat, India.
sharavdesai@gmail.com

*Please cite this article in press as **Dr. Sharav Desai** et al. In Silico Prediction and Docking Of Tertiary Structure of Protein X, Multifunctional Protein of Hepatitis B Virus (HBV). Indo American Journal of Pharmaceutical Research.2017:7(09).*

INTRODUCTION

HBV is the acronym for the hepatitis B virus. This virus belongs to the genus of Orthohepadnavirus and the family of Hepadnaviridae. Infection with this virus will cause a condition called hepatitis B. Hepatitis B virus also includes three other species of the Orthohepadnavirus genus, and these are called Ground squirrel. Woodchuck hepatitis virus, and Woolly monkey hepatitis B virus[1]. Two other families are involved in the Hepadnaviridae family one is Orthohepadnavirus, the other is Avihepadnavirus, and third is still not given a name. Virus and its homologs are found to have an ancient origin by appearing in old world monkeys and new world primate[2]. The envelope of the virus contains antigenic epitopes, and by this epitopes, four serotypes are available these are air, adw, ayr, ayw. By the nucleotide content of the virus genome, eight genotypes have been classified. From the genotype of the virus, it is possible for us to trace the location of the origin of infection as they all are having diversity in their geographical distribution. Vaccination, treatment, and severity of the disease are also known to variate between different genotype[3,4]. Hepatitis B is a serious and global infection of the liver most of the time it is life threating also. Chronic condition with liver cirrhosis and cancer are leading cause of the death due to such infections. 5% to 10 % adults are known to get an infection in the Saharan and East Asia. Amazon and southern and eastern part of the central Europe are also known to have high rates of chronic infection of Hepatitis B virus. In India and the Middle East, 2% to 5% are known to get an infection. Northern America and Western Europe are having 1% of chronic infections[5] . Today amongst many of the conditions hepatocellular carcinoma is also a one of the leading cause of the death which arise after the chronic Hepatitis B infection. Protein X of the virus is found to play a very crucial role in the virally induced hepatocellular carcinoma. Protein X also was known as HBx protein and is a multifunctional protein with the molecular weight of 17 kDa. This protein is known to involve and interacts with the list of the host protein to cause the hepatocellular carcinomas. It is observed that HBx protein can interfere with cellular processed like oxidative phosphorylation, DNA repair and signal transduction, transcription and protein degradation. In most of the cases, the host genome is known to integrated by the open reading frame of the HBx virus gene, and such integration will play a crucial role in the development of hepatocellular carcinoma[6,7]. Among all the serotypes of the virus and even in the Hepadnaviridae family this protein is found to be conserved. Localization of the protein known to occur in the cytoplasm but to some extent nucleus of the hepatic cells may also get localized. Protein X also plays a crucial role in the virus replication and subsequent infection[8,9]. DNA is not direct target by this protein instead several transcription factors are identified, and some are transcription factor I, transcription factor 2, RNA polymerase binding protein, a subunit of RNA polymerase. Further, it is possible that this protein can modify the cytoplasmic signal transduction pathways, Ras-Raf nitrogen activated protein kinase, Janus Kinase, focal adhesion kinase and proline-rich tyrosine kinase. Lastly, this protein is also known to induce the hepatic cell proliferation through its transactivation domain, and it is also observed that this protein also suppresses the tumor regression activity of P53[10,11].

As it is discussed above the said protein X is found to be part of virus essential life cycle and involved in the establishment of infection in the host. Further, traditional drug discovery process can take a huge amount of time and effort. Where as in this article computational approach is used to develop the Protein X as a drug target for HBV infection. Currently the structure for HBx protein is not available to study the active compounds against it which is the first step in drug target development. Build structure refinement and validation also increases the failure of experimental results. In this article protein X form hepatitis B virus is structurally evaluated to conduct docking studies. Furthermore Protein X structure is docked using ligand from the protein data bank and lead compounds resulted from the studies should be evaluated by experimental analysis.

MATERIAL AND METHODS

Sequence retrieval, Physiochemical properties, and secondary structure.

UniProt is a universal protein database that contains information about the structure and functions of the protein. The UniProt server is freely accessible, and the same was used to retrieve the amino acid sequence of the Protein X of Hepatitis B virus. The sequence was searched from the proteomic section of the database. The sequence in the UniProt is assigned by the Accession number, and the Accession number Q765E5 was obtained as protein X series and was used in the FASTA format for further studies[12-14]. ExPAsy server (Expert Protein Analysis System, Biozentrum, and University of Basel, Switzerland) is a resource portal for the bioinformatics tools and is governed by Swiss Institute of Bioinformatics (SIB). ExPAsy server contains many scientific portals with databases, which are freely available. One of such portal ProtParam was used to study the physicochemical properties of the protein X. one-letter code amino acid sequence obtained from the UniProt database was submitted to the tool in simple file format and analyzed for the properties like molecular weight, amino acid composition, instability index, half-life, aliphatic index, theoretical pI and extinction coefficient[15]. For predicting the secondary structure of the protein X scratch server was used Donald Bren School of Informatics and Computer Sciences, California, USA). The server suite includes the tools)for predicting secondary structure, relative solvent accessibility, disorder profile and much more. Amino acid sequence in the plain format without header and spaces was submitted to SSpro8 tool. SSpro8 is an advanced version of the SSpro, and instead of using Helix, strand and rest classification full eight class DSPP (Dictionary of protein secondary structure prediction) output was used for prediction[16-18].

Functional domain prediction

Sequence MOTIF is the composition of amino acid sequence which could have biological significance. MOTIF is a program from GenomeNet, Japan which uses a Pfam and prosite database. Here in this study Pfam database was used for the MOTIF prediction[19]. The sequence was submitted to MOTIF database into plain single letter format. Cut off score of 1.0 *E for P-fam database was kept.

Homology modeling, refinement, and evaluation of the 3D structure.

Protein structure prediction and construction were done using a suite of tools available on the network called Phyre 2 from the Imperial College of London[20]. FASTA formatted sequence was submitted to the server, and the results were obtained through the Email address provided. Intensive modeling mode was used to generate the structure. Structural classification of protein provide (SCOP) collection of available protein structures and also augment them with the newer database like Protein Data Bank (PDB)[21]. The non-redundant database was searched for the similarity with the user submitted query, profile constructed and secondary structure was built. Finally, secondary structure was screened against the fold library using profile- profile alignment algorithm detailed in Bennett-Lovsey et al.[22]. Alignment score was generated for ranked structures, and top ten structures were taken to produce the full three-dimensional structure of the submitted query. Loop library was used to repair the missing and delete inserts. Side chains were placed on the model using a fast graph-based algorithm and side chain rotamer library. The quality of a structure produced by prediction program depends on the similarity between target and available templates. Therefore it is always advisable to improve the quality of predicted structures. Here Galaxy Refine web server (Computational Biology Lab in the Department of Biochemistry, Seoul National University) was used which works on the method for refinement that is approved by CASP 10 (Critical Assessment of techniques for protein Structure Prediction). First side chain building and repacking were done. Through molecular dynamic simulation, full structure relaxation was done[23,24]. The refined models were assessed by several validation tools to select the best model and evaluate the quality of that model. National Health Institute, University of California, USA provide and maintain an algorithm called ERRAT, and it was used to study the improvements in the model building and refinement studies[25]. PROSA is a web tool available from the Center for Applied Molecular Engineering, Division, of Bioinformatics and University of Salzburg, Austria. Model coordinates were supplied by submitting a Structure in the PDB file format. All the calculations were done using C^{α}Potentials. The Z- the score was calculated for overall model quality and deviation of total energy concerning an energy distribution derived from random conformation[26-27]. The SWISS-MODEL workspace server was utilized for calculation of ANOLEA score and QMEAN6 (Biozentrum, University of Basel, Switzerland[28-31]. Packing quality of the model was estimated by atomic empirical mean force (ANOLEA) score. Energy calculations were performed on protein chain by evaluating Non- local environment of each heavy atom in the molecule [32]. Qualitative Model Energy Analysis score for both global and local quality was calculated using the QMEAN6 tool[33,34]. All prediction calculations were based on propensity scales for each of 20 amino acids. Each scale consists of 20 values assigned to each of amino acid residues by their relative propensity to possess the property described by the scale [35]. The refined and validated structure was submitted to PMDB (Protein Model Data Base) web server. PMDB is a joint project between CASPUR and the Biocomputing group of the Department of Biochemical Sciences of the University of Rome "La Sapienza [36]. The compounds used to screen Transactivation Protein X were obtained from the Protein Data Bank [37]. The most of the ligands selected were composed of very small molecular weight and also were newly introduced in PDB ligand library. Chemical structure of three ligands Lamivudine, Rimantadine and zalcitanib were obtained from Drugbank [38]. PDBQT files of the ligands and Receptors were prepared. Autodock vina was used to perform the docking studies [39]. PDBQT file of three drugs was taken as ligand and PDBQT file of Protein X was taken as protein. These drug compounds were docked with Protein X of Hepatitis B virus. PyRx 0.8 (Virtual Screening Tools) was used to screen the Ligand library of the protein data bank. File formats like SDF and PDB were utilized for the ligands respectively. Single ligand file formats were then optimized using a program called Openbabel [40-41].

RESULTS AND DISCUSSION

Hbx is hepatitis B virus protein, and the amino acid sequence was retrieved from the UniProt database (Accession Number Q765E5. The sequence obtained from the database was then submitted to ProtParam web tool (ExPAsy Server), and physicochemical properties were computed. Protein is composed of 154 amino acids with the highest percentage of Leucine (11.4%) followed by Serine (10.4%). The molecular weight of the protein is found to be of 16.756 kDa, and theoretical pI found to be of 8.08 allowing purification by isoelectric focusing. These parameters are helpful to predict the location of band those may appear in the 2-D gel electrophoresis [42]. Protein purification is measured mostly by spectrophotometric analysis. Extinction coefficient value represses the amount of light absorbed by the protein at given wavelength. The extinction coefficient of the protein in water was found to be 8980 M^{-1} cm^{-1} at 280nm, assuming that all cysteine pairs come from cysteine residue and 8480 M^{-1} cm^{-1} assuming that all cysteine residues are reduced [43]. The instability index of the protein represents the stability of the protein in the test tube. A score below 40 represents the stability of the protein and score above 40 indicates instability of the protein. The protein X found to score of 60.28, and it is slightly unstable in the test tube [44].

Table 1. Refined Models by Galaxy Web servers.

Sr.No.	Model	GDT-HA	RMSD	Mol Probity	Clash score	Poor rotamer	Rama favored
1.	Initial	1.0000	0.000	4.438	172.5	12.4	60.5
2.	MODEL 1	0.9091	0.529	2.288	12.1	0.8	82.9
3.	MODEL 2	0.9140	0.524	2.270	13.3	0.8	86.2
4.	MODEL 3	0.9058	0.541	2.270	12.9	0.0	85.5
5.	MODEL 4	0.9123	0.536	2.174	10.4	0.0	86.2
6.	MODEL 5	0.9026	0.524	2.439	13.3	1.6	84.9

It has been found that aliphatic index of the protein for thermophilic bacteria is relatively higher compared to normal bacteria. The aliphatic index is represented by some side chains of the aliphatic amino acids like alanine, valine, isoleucine, and leucine. The aliphatic calculated index of the protein found to be 79.22 represents slightly Mesophilic nature [45]. HMM scan (hidden Markov model) calculated for the matching score between the query sequence and each domain found in Pfam library in bit score. Found motifs with smaller E-value compared to the threshold value were listed. From the results, it is quite clear that the amino acids from 1 to 142 contributing for the transactivator protein X domain for Protein X of Hepatitis B virus (Figure 1).

Figure 1. The amino acid sequence of Protein X in Hepatitis B Virus showing the transactivation protein X domain in red.

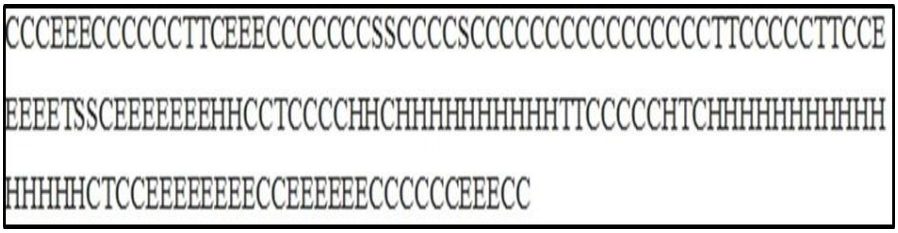


Figure 2. The secondary structure predicted by SsPro8, where H: alpha-helix G: 3-10-helix, E: extended strand, B: beta-bridge, T: turn, S: Bend, C: the rest.

To, Predict SSPro8 server adopted the secondary structure of protein full DSSP classification. The structure predicted comprises of 20% α-Helix, 22% of extended β-strands, 8% of turns, 3% of bands and remaining 46% is of random coils(Pollasriet et al.; 2002) (Figure 2). Figure 3 represents the antigenicity prediction of the HBx protein. On the graphs, the Y-axes represents residual correspondent score (averaged in the specified window), be it a BepiPred score or a residue score on the Karplus and Schulz flexibility scale; while the X-axes represents the residue positions in the sequence.

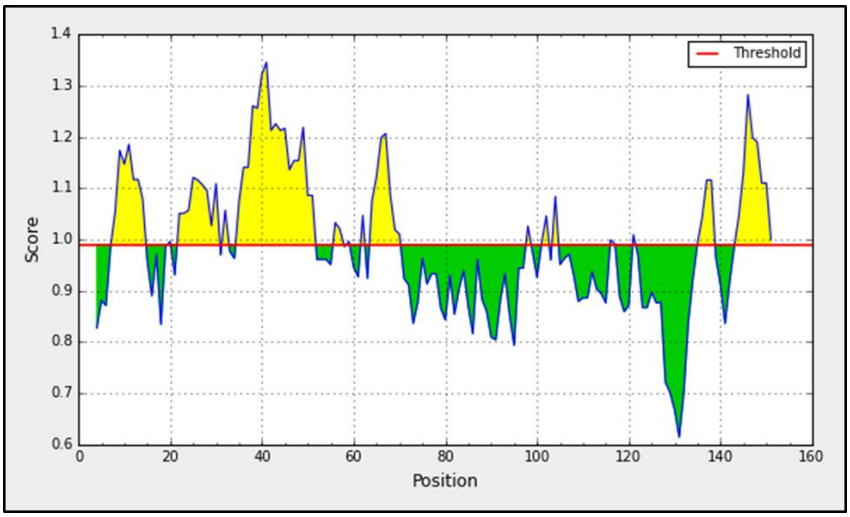


Figure 3. Chou & Fasman Beta-turn Prediction Result.

Table 2. The refined model produced by Galaxy web browser and it's scored.

Model	ProSA Z-Score	ERRAT Quality	QMEAN6 score	Procheck Ramachandran plot			
				Co[1] (%)	AA[2] (%)	GA[3] (%)	DA[4] (%)
Initial	-2.88	17.730	0.227	51.1	32.1	9.2	7.6
MODEL 1	-3.45	54.054	0.367	65.6	26.7	3.1	4.6
MODEL 2	-3.12	35.185	0.325	67.2	25.2	4.6	3.1
MODEL 3	-3.16	41.667	0.317	68.7	23.7	3.1	4.6
MODEL 4	-3.15	45.794	0.276	65.6	26.7	3.8	3.8
MODEL 5	-3.27	47.573	0.304	67.2	25.2	3.8	3.8

1Residues in the most-favored regions, 2residues in the additionally allowed regions, 3residues in the generously allowed regions, 4 residues in the disallowed regions.

The larger score for the residues is interpreted as that the residue might have a higher probability of being part of the epitope (those residues are colored in yellow on the graphs). The 3-D structure of a protein provides more information about its function than its sequence because interactions of a protein with other molecules are determined by amino acid residues that are close in space but are frequently distant in sequence. Phyre2 web server was used to build the 3-D model for the protein sequence.

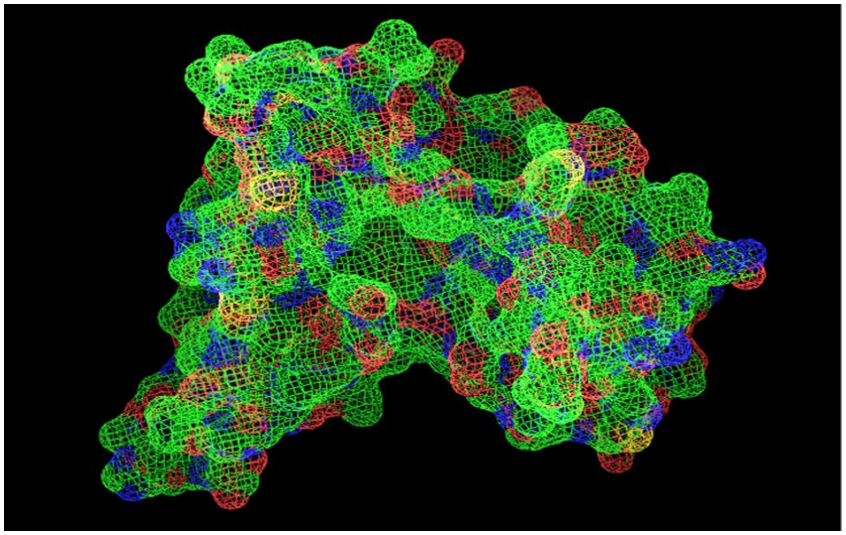


Figure 4. 3D Model for Protein X-Generated from Phyre 2 Unrefined.

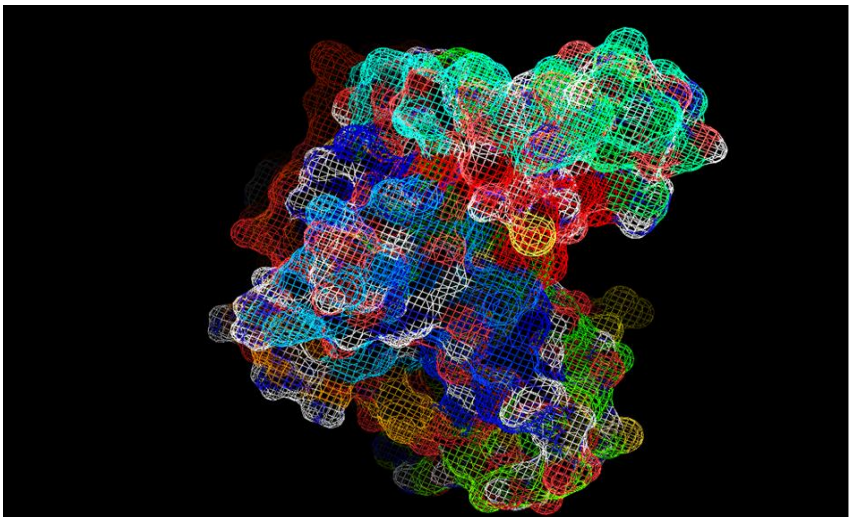


Figure 5. The three-dimensional structure of Protein X, Model 1, produced by Phyre2 and refined by Galaxy WEB servers.

The sequence was submitted to the server in single letter format. The server generates a model by looking for a homologous sequence in the available database. Next server performed loop modeling and fold recognition. The model produces and predicted by any *in silico* method must be refined by suitable algorithms to rebuilt the several of the side chains. The initial model was submitted to the Galaxy Refine server for the refinement. Five models were generated amongst these generated model several validation steps performed to find the best suitable refine model (Table 1 & Table 2). ERRAT is a novel method that can detect incorrect regions of protein structures according to errors leading to random distributions of atoms, which can be distinguished from correct distributions. Figure 6 shows the refined Model 1 with the quality of 54.054% which is greater than the initial model containing many erroneous regions and a quality of 17.730%. In ANOLEA profile (Figure 7), the original model had many areas of high energy, which were significantly improved in the refined model, suggesting greater reliability.

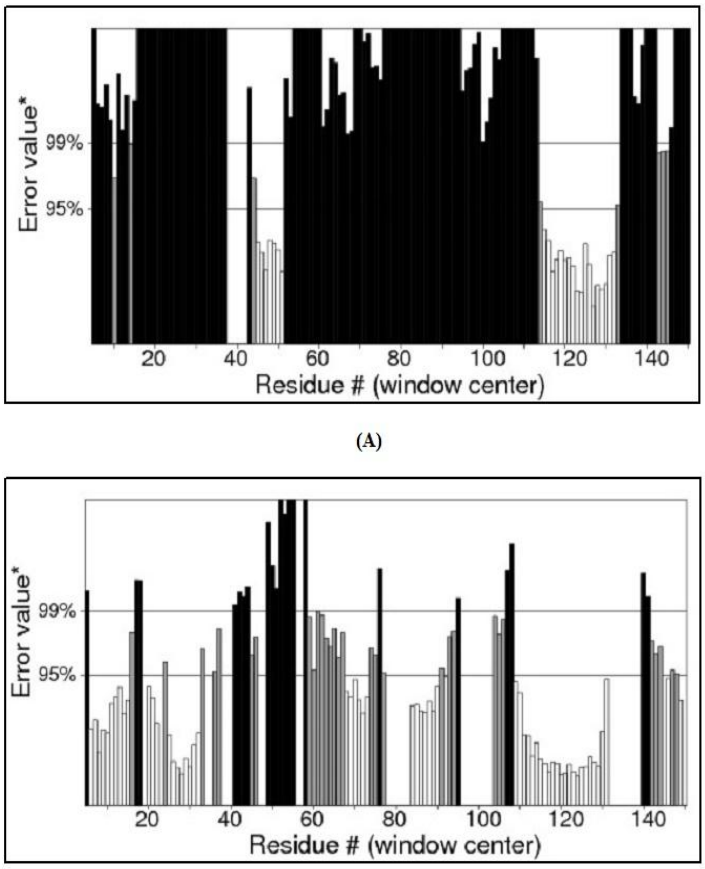
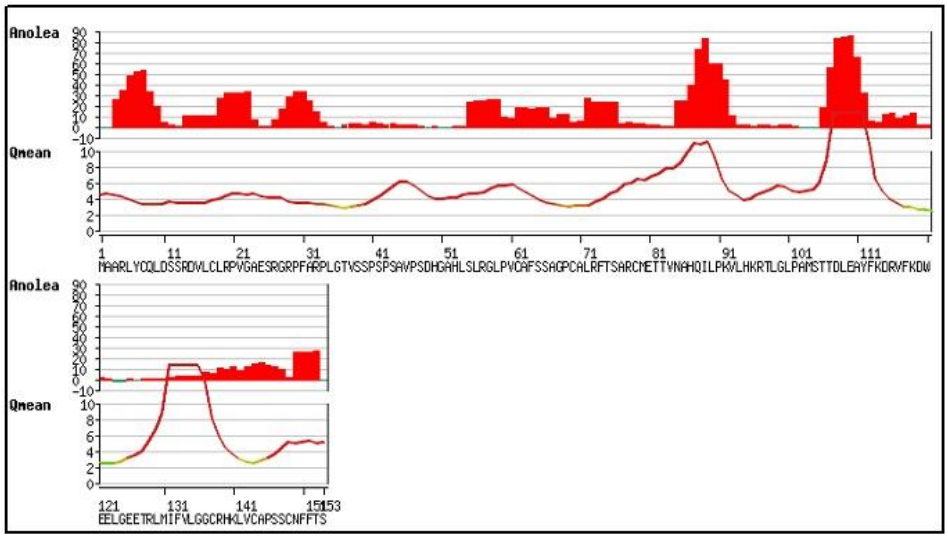


Figure 6. ERRAT plot shows error values for residues. The Y-axis represents the error value, and the X-axis represents the amino acid residues of the protein model. An error value exceeding 99% confidence level indicates a poorly-modeled region (A) The initial model with the quality of 17.730% (B) the refined Model 4 with the quality of 54.054%.



(A)

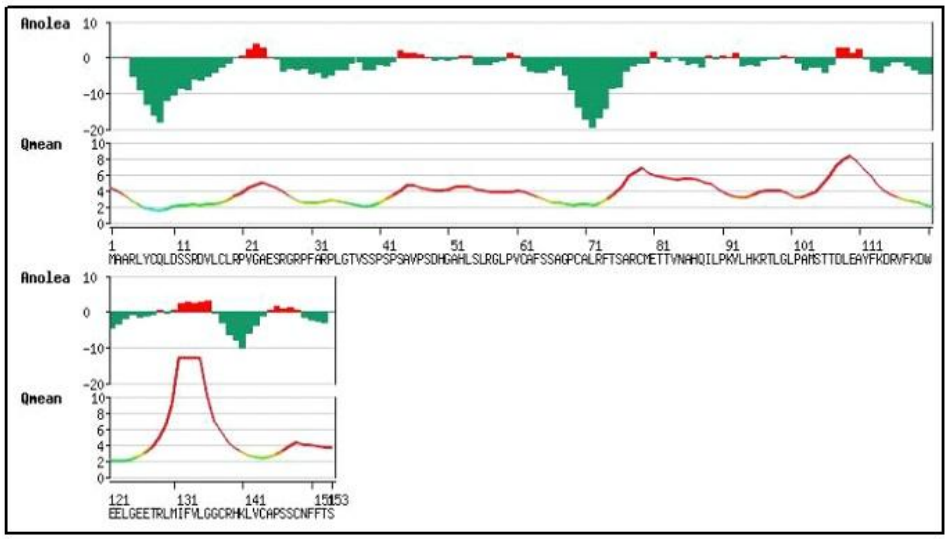


Figure 7. ANOLEA plots are showing high energy zones as red (A) The initial model (B) The refined Model 1 with the high energy zones substantially minimized and improved.

The Z-scores of all the models is similar to the typical values commonly found in native structures determined by NMR spectroscopy and X-ray crystallography (Figure 8).

Table 3. The QMEAN6and component scores and their Z-scores of the initial and first models concerning the experimental structure of similar sizes.

Scoring Function Term	Initial Model		Refined Model 1	
	Raw Score	Z-score	Raw Score	Z-Score
C_beta interaction energy	-8.77	-2.74	-9.67	-2.70
All-atom pairwise energy	-306.20	-3.45	-1077.74	2.73
Solvation energy	16.78	-4.67	14.03	-4.26
Torsion angle energy	29.04	-5.56	14.25	-4.37
Secondary structure agreement	76.0%	-1.07	76.6%	-0.97
Solvent accessibility agreement	57.8%	-3.56	63.6%	-2.51
QMEAN6 score	0.227	-5.55	0.367	-4.06

QMEAN, which stands for Qualitative Model Energy Analysis, is a composite scoring function describing the major geometrical aspects of protein structures was used to validate the quality of the structures produced through homology modeling. The QMEAN6 score for the model 1 was found to be highest, and model 1 can be regarded as the best model amongst all model tested. This is further clear from the other contributor to the overall score (Table 3)[46]. The transactivating function is probably associated with a tumorigenic potential of HBx, since x gene sequences, encoding functional HBx, have repeatedly been found integrated into the genome of liver carcinoma cells[47-50]. In addition to the raw scores, Z-scores of the QMEAN composite score as well as all terms were investigated relating the quality estimates to scores obtained for high-resolution reference structures solved experimentally by X-ray crystallography (Table 3).

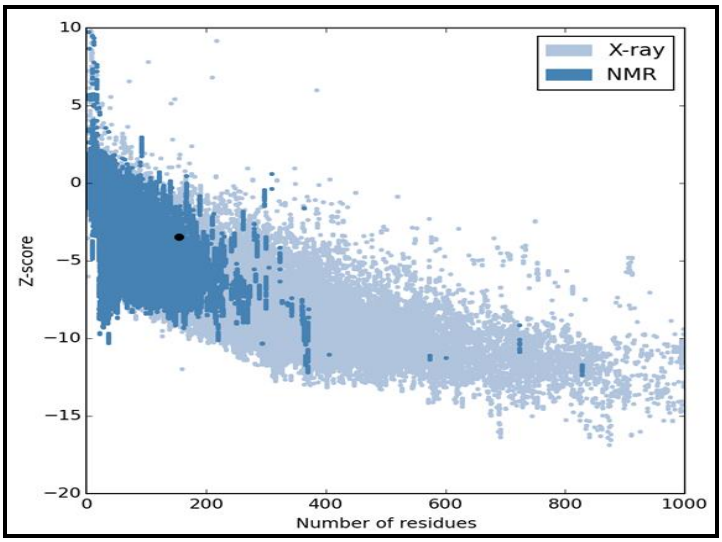


Figure 8. The Z-score plot of Protein X (dot) determined by ProSA. The Z-score is -4.062, within the range of experimental native.

In the Z-sore plot of Protein X, the Particular value is displayed that contains the z-scores of all experimentally determined protein chains in current PDB (Figure 8). The Z- score of the protein and from the plot (Figure 8) it is clear that refined model 1 is within the range of experimental native structures of similar sizes. Ramachandran plot (Figure 9) shows a plot of phi/psi dihedral angels amongst N-Cα and Cα-C peptide bond in the protein's backbone, and it is very clear from the plot that for a model refined model 1 carries a good percentage of residues fall in to favor regions that unrefined model.

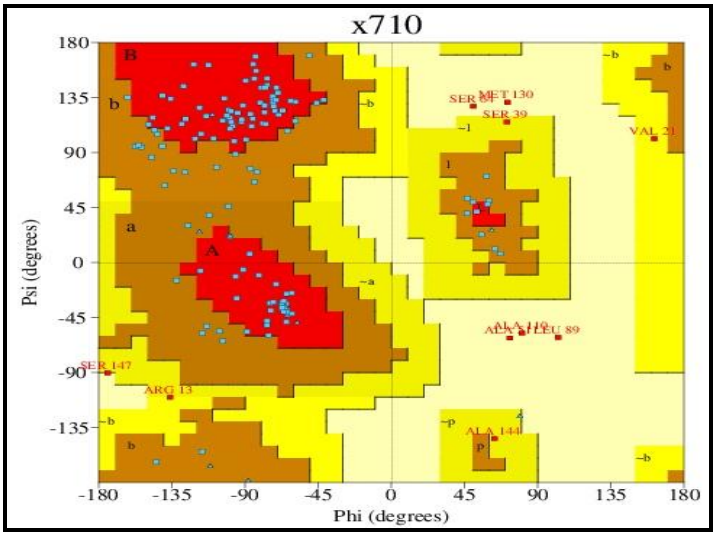


Figure 9. Procheck determines Ramachandran plot of the fourth model. The most preferred regions are marked as A, B, and L. The additional allowed regions are indicated as a, b, l, and p. All non-glycine and proline residues are shown as filled black squares, whereas glycines (non-end) are shown as filled black triangles. Disallowed residues are colored red.

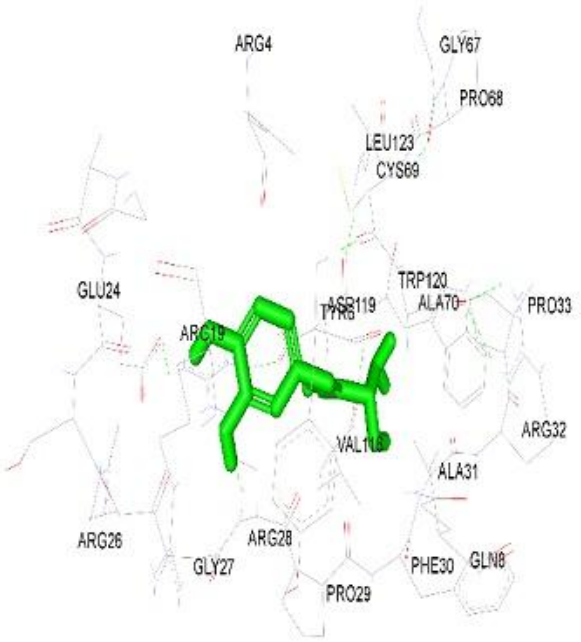
Several of the ligands were selected from the Protein data bank and drug bank. All the compounds were minimized for their energy and lastly converted to PDBQT format.

Table 4. Docking Results of Protein X against Experimental Ligands from PDB.

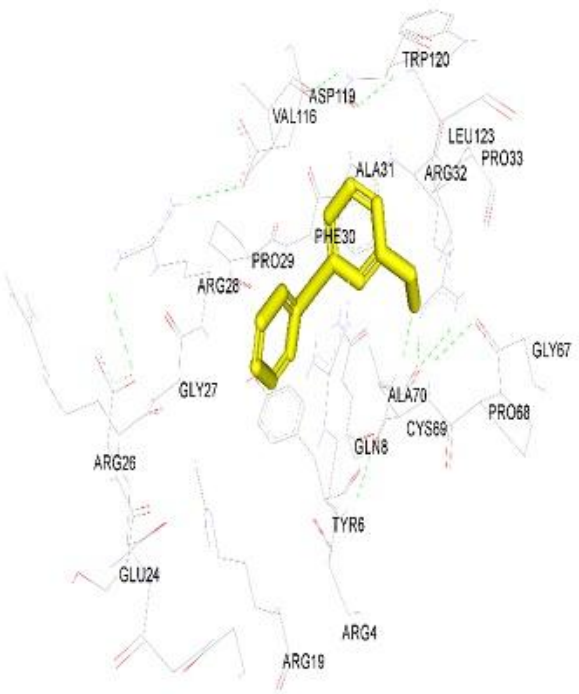
DB and PDB ID	Molecular Weight	Ligand Formula	Binding Energy Kcal/Mol
DB00709	229.25	$C_8H_{11}N_3O_3S$	-8.0
DB00478	179.30	$C_{12}H_{21}N$	-7.3
DB00879	247.24	$C_8H_{10}FN_3O_3S$	-8.4
64Q	239.28	$C_{13}H_{13}N_5$	-8.1
7U3	263.27	$C_{13}H_{14}FN_3O_2$	-8.6
B3P	282.33	$C_{11}H_{26}N_2O_6$	-5.9
LQH	264.3	$C_{17}H_{13}FN_2$	-9.0
TPO	199.1	$C_4H_{10}NO_6P$	-5.3
040	206.195	C11 H10 O4	-6.6
0OO	200.233	C13 H12 O2	-7.1
OJW	206.198	C10 H10 N2 O3	-6.8
0OY	207.226	C11 H13 N O3	-7.2

DB-DrugBank, PDB-Protein Data Bank

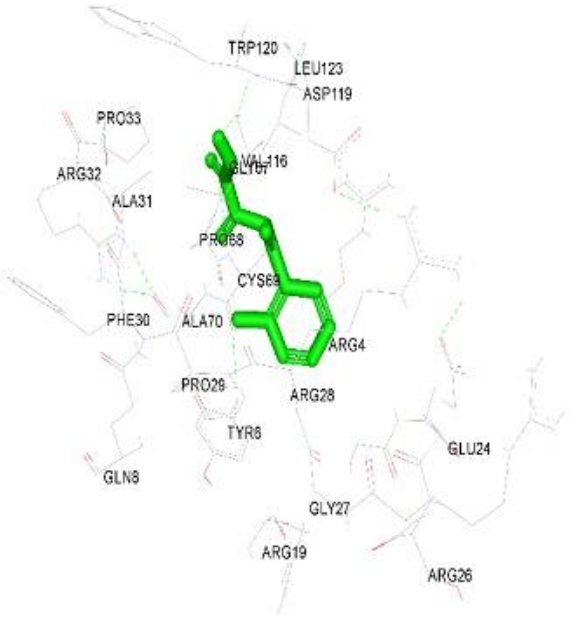
Amongst hundreds of the compounds, ten lead compounds were taken as lead molecules by their binding affinity to the protein (Figure10 &Figure11& Table 4).


ARG4
GLY67
PRO68
LEU123
CYS69
GLU24
TRP120
ASP119
ALA70
PRO33
TYR6
ARG19
VAL116
ARG32
ALA31
ARG26
ARG28
GLY27
PRO29
PHE30
GLN8

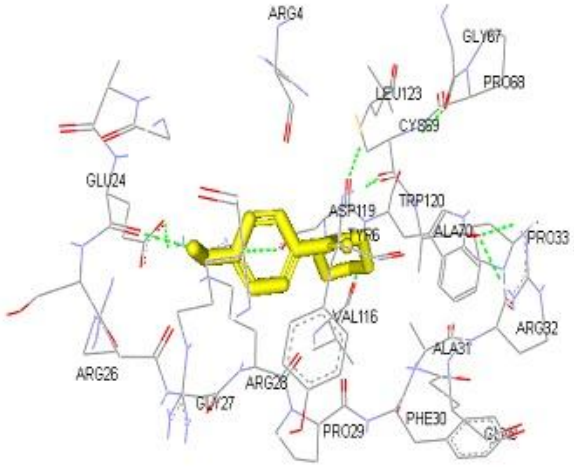
(A)


TRP120
ASP119
VAL116
LEU123
PRO33
ALA31
ARG32
PHE30
PRO29
ARG28
GLY67
GLY27
ALA70
PRO68
CYS69
GLN8
ARG26
TYR6
GLU24
ARG4
ARG19

(B)


TRP120
LEU123
ASP119
PRO33
ARG32
VAL116
GLY67
ALA31
PRO68
CYS69
PHE30
ALA70
ARG4
PRO29
ARG28
TYR6
GLU24
GLN8
GLY27
ARG19
ARG26

(C)


ARG4
GLY67
LEU123
PRO68
CYS69
GLU24
ASP119
TRP120
TYR6
ALA70
PRO33
VAL116
ARG32
ALA31
ARG26
ARG28
GLY27
PRO29
PHE30

(D)

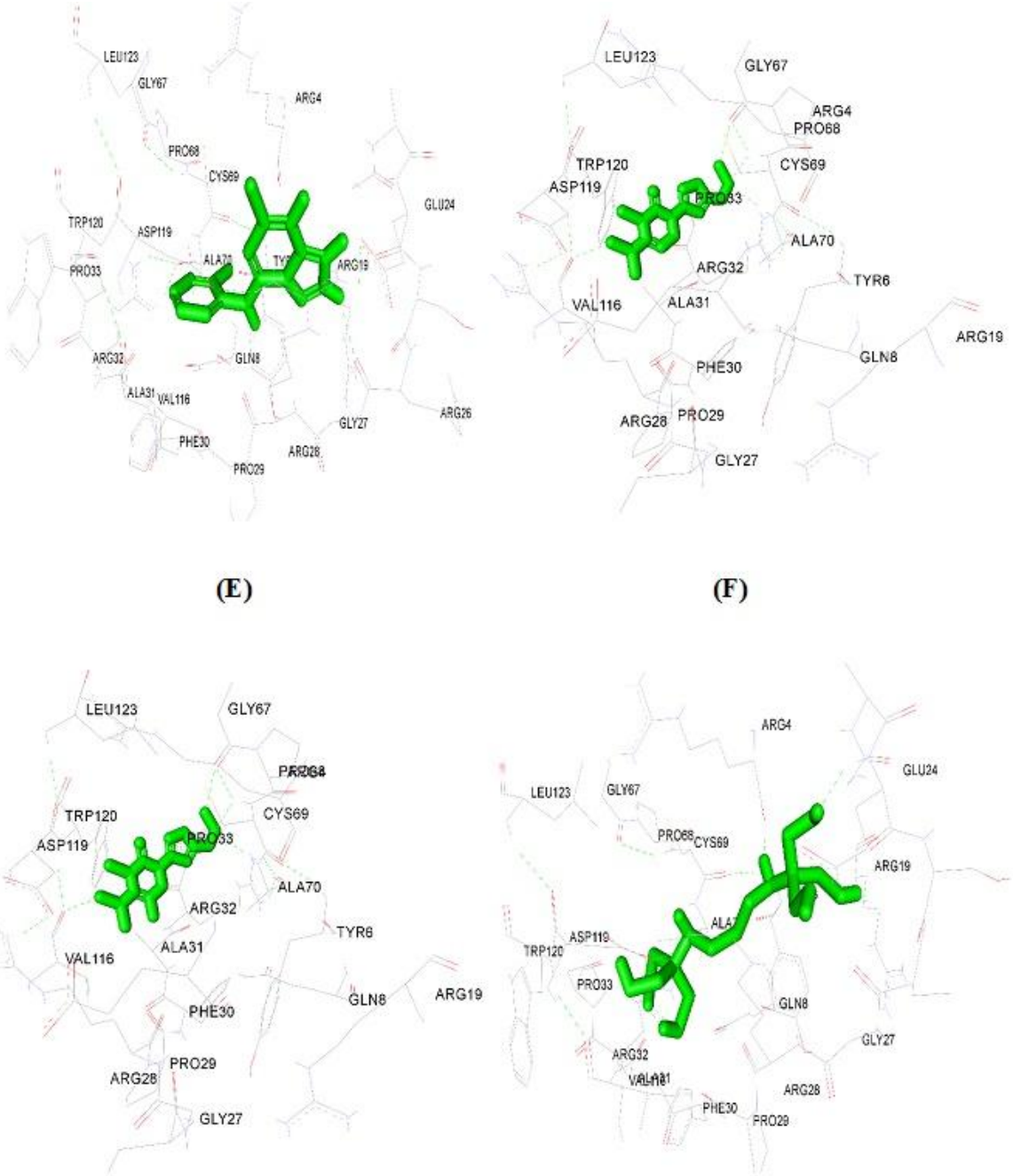


Figure 10. Docking interaction structure of lead molecules (A), PDBOJW (B), PDB0OO (C), DPB0OY (D), DPB040 (E), DPB64Q (F), DB00709(G),DB00879(H), DPBBP3.

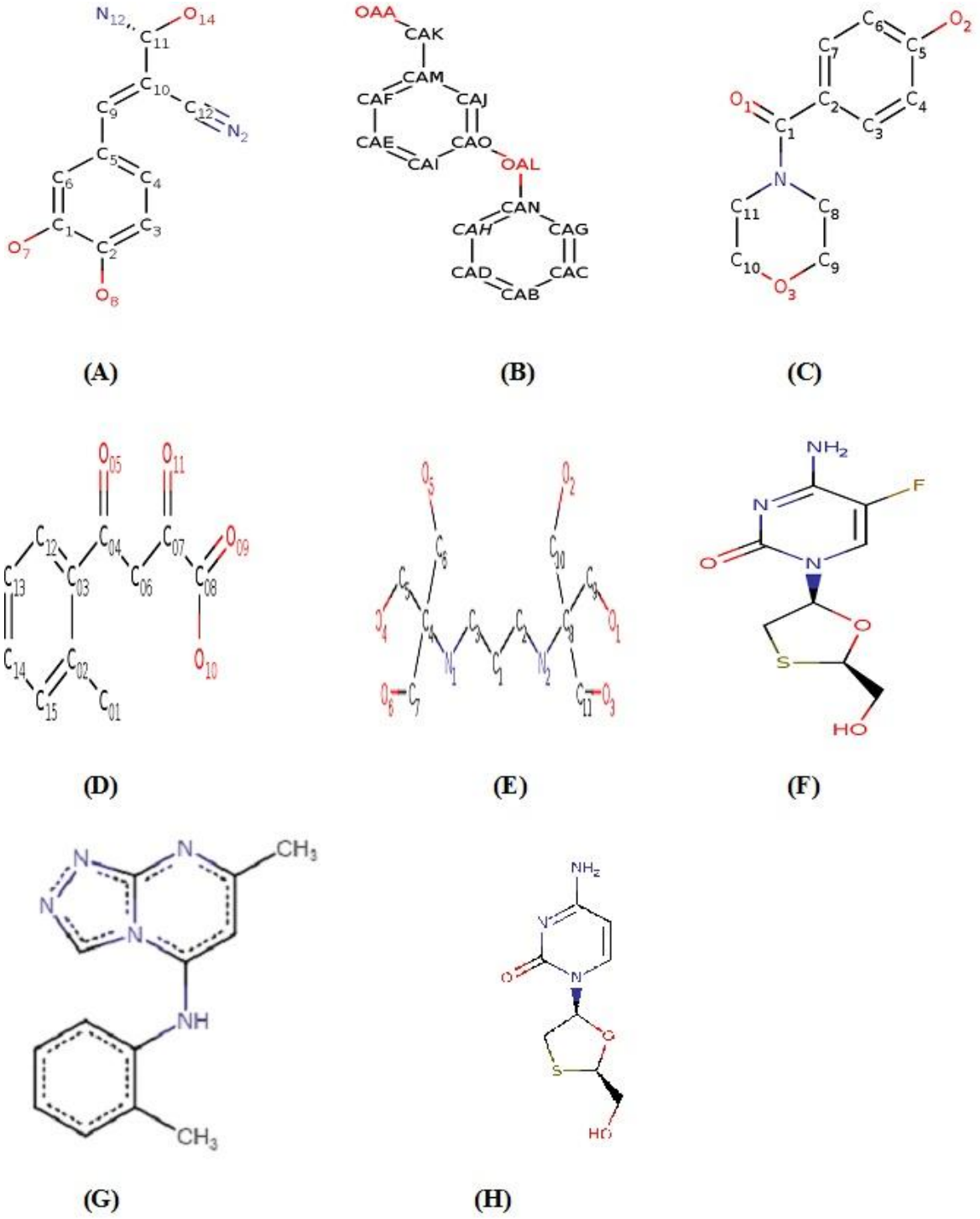


Figure 11. Chemical Structures of lead Molecules. (A), PDBOJW (B), PDB0OO (C), PDB0OY (D), PDB040 (E), PDBB3P (F), PDB00879 (G), PDB 64Q (H), DB00879.

CONCLUSION

From presented work, it further recommended that 3D structure of Protein X could be utilized to model Inhibitors with homologous matching in different microorganisms. This work also likewise demonstrated that the Transactivation Domain site of activity might be abused in ligand configuration to hinder the viral development, as it is watched that the protein X found to exceptionally fundamental for an infection for a few of the capacities. Further binding of protein X can also disturb the host-pathogen interactions, thus authors of studies strongly suggest that lead molecules resulted from the docking studies should further be evaluated experimentally and clinically to develop as potential target.

Conflict of interests:

Author declares no conflict of interest.

Abbreviations:

HBV - Hepatitis B Virus
DB - DrugBank
PDB - Protein Data Bank
PROSA - Protein Structure Analysis
CASP - Critical Assessment of techniques for protein Structure Prediction
DSPP - Dictionary of protein secondary structure prediction) output was used for prediction
SIB - Swiss Institute of Bioinformatics
ExPASy - Expert Protein Analysis System

REFERENCES

1. Hassan MM, Li D, El-Deeb AS, Wolff RA, Bondy ML, Davila M, and Abbruzzese JL. Association Between Hepatitis B Virus and Pancreatic Cancer. Journal of Clinical Oncology. 2008; 26:4557-4562
2. Dupinay T, Gheit T, Roques P, Cova L, Chevallier-Queyron P, Tasahsu SI, Le Grand R, Simon F, Cordier G, Wakrim L. Discovery of naturally occurring transmissible chronic hepatitis B virus infection among Macaca fascicularis from Mauritius Island. Hepatology (Baltimore, Md). 2013; 58: 1610-1620
3. Kramvis A, Kew M, and Francois G. Hepatitis B virus genotypes. Vaccine. 2005 ;23: 2409-2423
4. Magnius LO, and Norder H. Subtypes, genotypes and molecular epidemiology of the hepatitis B virus as reflected by sequence variability of the S-gene. Intervirology, 1995; 38:24-34
5. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015; 385(9963):117-71 doi: 101016/S0140-6736(14)61682-2
6. Al-Khayyat MZS, and Al-Dabbagh AGA. In silico Prediction and Docking of Tertiary Structure of LuxI, an Inducer Synthase of Vibrio fischeri Reports of .Biochemistry and Molecular Biology. 2016;4:66-75
7. Miller RH, and Robinson. Common evolutionary origin of hepatitis B virus and retroviruses. Proceedings of the National Academy of Sciences of the United States of America. 1996; 83:2531-2535
8. Murakami S. Hepatitis B virus X protein: a multifunctional viral regulator. Journal of Gastroenterology. 2001; 36:651-660
9. Gong DY, Chen EQ, Huang FJ, Leng XH, Cheng X and Tang H. Role and functional domain of hepatitis B virus X protein in regulating HBV transcription and replication in vitro and in vivo. Viruses, 2013;5:1261-1271 doi: 103390/v5051261
10. Diao J, Garces R, and Richardson CD. (2001) X protein of hepatitis B virus modulates cytokine and growth factor related signal transduction pathways during the course of viral infections and hepatocarcinogenesis .Cytokine & growth factor reviews.2001; 12:189-205
11. Feitelson MA, and Lee J.Hepatitis B virus integration, fragile sites, and hepatocarcinogenesis. Cancer letters,2007; 252: 157-170
12. Bastian FB, Chibucos MC, Gaudet P, Giglio M, Holliday GL, Huang H, Lewis SE, Niknejad A, Orchard S, Poux S. The Confidence Information Ontology: a step towards a standard for asserting confidence in annotations Database. The Journal of Biological Databases and Curation.2015. 101093/database/bav043
13. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, Poux S, Bougueleret L, and Xenarios I. UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View Methods in molecular biology (Clifton, NJ), 2016; 1374: 23-54
14. UniProt C. UniProt: a hub for protein information.Nucleic Acids Res.2015; 43: D204-212
15. Gasteiger E HC, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. (2005) The Proteomics Protocols Handbook, Humana Press, (Humana Press)
16. Magnan CN, and Baldi P.SSpro/ACCpro 5: almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning, and structural similarity. Bioinformatics, 2014; 30: 2592-2597
17. Pollastri G, Przybylski D, Rost B, and Baldi P. Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles . Proteins: Structure, Function, and Bioinformatics, 47, 228-235
18. Kabsch, W, and Sander, C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features Biopolymers, 22: 2577-2637
19. Bateman, A, Coin, L, Durbin, R, Finn, RD, Hollich, V, Griffiths-Jones, S, Khanna, A, Marshall, M, Moxon, S, Sonnhammer, ELL, et al. (2004) The Pfam protein families database Nucleic Acids Research, 32, D138-D141
20. Kelley, LA, Mezulis, S, Yates, CM, Wass, MN, and Sternberg, MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis Nat Protocols, 10: 845-858
21. Berman, HM, Westbrook, J, Feng, Z, Gilliland, G, Bhat, TN, Weissig, H, Shindyalov, IN, and Bourne, PE (2000) The Protein Data Bank Nucleic Acids Research, 28: 235-242
22. Bennett-Lovsey, RM, Herbert, AD, Sternberg, MJ, and Kelley, LA (2008) Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre Proteins, 70: 611-625
23. Ko, J, Park, H, Heo, L, and Seok, C (2012) GalaxyWEB server for protein structure prediction and refinement Nucleic Acids Research, 40: W294-W297
24. Heo, L, Park, H, and Seok, C (2013) GalaxyRefine: Protein structure refinement driven by side-chain repacking Nucleic Acids Res, 41: W384-388
25. Colovos, C, and Yeates, TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions Protein Science : A Publication of the Protein Society, 2, 1511-1519
26. Sippl, MJ (1995) Knowledge-based potentials for proteins Current Opinion in Structural Biology, 5, 229-235
27. Sippl, MJ (1993) Recognition of errors in three-dimensional structures of proteins Proteins, 17: 355-362
28. Wiederstein, M, and Sippl, MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins Nucleic Acids Research, 35, W407-W410
29. Arnold, K, Bordoli, L, Kopp, J, and Schwede, T (2005) The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling Bioinformatics, 22: 195-201
30. Biasini, M, Bienert, S, Waterhouse, A, Arnold, K, Studer, G, Schmidt, T, Kiefer, F, Cassarino, TG, Bertoni, M, Bordoli, L, et al (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information Nucleic Acids Research, 42: W252-W258

31. Guex, N, Peitsch, MC, and Schwede, T (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective Electrophoresis, 30: S162-S173
32. Kiefer, F, Arnold, K, Künzli, M, Bordoli, L, and Schwede, T (2008) The SWISS-MODEL Repository and associated resources Nucleic Acids Research, 37: D387-D392
33. Melo, F, and Feytmans, E (1998) Assessing protein structures with a non-local atomic interaction energy Journal of molecular biology, 277, 1141-1152
34. Benkert, P, Künzli, M, and Schwede, T (2009) QMEAN server for protein model quality estimation Nucleic Acids Research, 37, W510-W514
35. Benkert, P, Tosatto, SC, and Schomburg, D (2008) QMEAN: A comprehensive scoring function for model quality assessment Proteins, 71, 261-277
36. Chou, PY, and Fasman, GD (1978) Prediction of the secondary structure of proteins from their amino acid sequence Advances in enzymology and related areas of molecular biology, 47, 45-148
37. Castrignanò, T, De Meo, PDO, Cozzetto, D, Talamo, IG, and Tramontano, A (2006) The PMDB Protein Model Database Nucleic Acids Research, 34, D306-D309
38. Shin, J-M, and Cho, D-H (2005) PDB-Ligand: a ligand database based on PDB for the automated and customized classification of ligand-binding structures Nucleic Acids Research, 33, D238-D241
39. Wishart, DS, Knox, C, Guo, AC, Shrivastava, S, Hassanali, M, Stothard, P, Chang, Z, and Woolsey, J (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration Nucleic Acids Research, 34, D668-D672
40. Trott, O, and Olson, AJ (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading Journal of Computational Chemistry, 31, 455-461
41. Morris, GM, Huey, R, Lindstrom, W, Sanner, MF, Belew, RK, and Goodsell, DS (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility J Comput Chem, 30
42. Sahoo, M, Jena, L, Daf, S, and Kumar, S (2016) Virtual Screening for Potential Inhibitors of NS3 Protein of Zika Virus Genomics & Informatics, 14, 104-111
43. Bjellqvist, B, Hughes, GJ, Pasquali, C, Paquet, N, Ravier, F, Sanchez, JC, Frutiger, S, and Hochstrasser, D (1993) The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences Electrophoresis, 14, 1023-1031
44. Gill, SC, and von Hippel, PH (1989) Calculation of protein extinction coefficients from amino acid sequence data Analytical biochemistry, 182, 319-326
45. Guruprasad, K, Reddy, BV, and Pandit, MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence Protein engineering, 4, 155-161
46. Ikai, A (1980) Thermostability and Aliphatic Index of Globular Proteins The Journal of Biochemistry, 88, 1895-1898
47. Pollastri, G, Przybylski, D, Rost, B, and Baldi, P (2002) Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles Proteins, 47, 228-235
48. Kidd-Ljunggren, K, Öberg, M, and Kidd, AH (1995) The hepatitis B virus X gene: analysis of functional domain variation and gene phylogeny using multiple sequences Journal of General Virology, 76, 2119-2130
49. Renner, M, Haniel, A, Bürgelt, E, Hofschneider, PH, and Koch, W (1995) Transactivating function and expression of the x gene of hepatitis B virus J Hepatol, 23: 53-65
50. Sánchez, R, and Sali, A (1998) Large-scale protein structure modeling of the Saccharomyces cerevisiae genome Proceedings of the National Academy of Sciences of the United States of America, 95, 13597-13602.

5487847845**170903**