IBBJ

Autumn 2017, Vol 3, No 4

Original Article

In Silico Prediction and Docking of Tertiary Structure of Multifunctional Protein X of Hepatitis B Virus

Sharav Desai*, Pooja Tahilramani, Dhara Patel, Prachi Patel, Dhananjay Meshra

Department of Pharmacy, Pioneer Pharmacy Degree College, Sayajipura, Vadodara, India.

Submitted 27 Jun 2017; Accepted 19 Aug 2017; Published 23 Sep2017

Hepatitis B virus (HBV) infection is a universal health problem and may result into acute, fulminant, chronic hepatitis liver cirrhosis, or hepatocellular carcinoma. Sequence for protein X of HBV 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 evaluated by Chou & Fasman beta-turn prediction method. Five models were generated, and model 1 was having the greatest quality on the basis of the QMEAN6 score with 0367 ERRAT analysis, revealing the overall quality of 54054% whereas the initial model was having only 17730% quality. The mean force potential energy, as analyzed by ANOLEA, was better compared with the initial 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. Transactivation domain predicted by MOTIF program and 3D structure submitted to PMDB can be used to design new drugs against HBV protein X.

Keywords: Protein X, homology, docking, Uniprot, *in silico*, hepatitis B virus, AutoDock, Pyrx08

Hepatitis B virus (HBV) belongs to the genus of *Orthohepadnavirus* and the family of *Hepadnaviridae*. HBV also includes three other species into the *Orthohepadnavirus* genus that are called Ground squirrel, Woodchuck hepatitis virus, and Woolly monkey HBV (1). Two other genera involved in the *Hepadnaviridae* family are *Orthohepadnavirus* and *Avihepadnavirus*. There is a third genera which has still not been given a name. HBV 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 on the basis of this epitopes, four serotypes called air, adw, ayr, and ayw are available. On the basis of the nucleotide content of the virus genome, eight genotypes have been classified. From the genotype of the virus, it is possible 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 genotypes (3, 4). Hepatitis B is a serious and global infection of the liver, and is most of the time life threatening. Chronic conditions with liver cirrhosis and cancer are leading causes of the death due to such infections. 5% to 10 % of adults

*Correspondence: Department of Phamacy, Pioneer Pharmacy Degree College, Vadodara, Gujarat, India. E-mail: sharavdesai@gmail.com

are known to get an infection in the Saharan and East Asian regions. Amazon, Southern and Eastern part of the Central Europe are also known to have high rates of chronic infection of HBV. In India and Middle East, 2% to 5% of population 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 leading cause of death which arises after the chronic HBV infection. Protein X of the virus is found to play a very crucial role in the virally induced hepatocellular carcinoma. Protein X also known as HBx protein is a multifunctional protein with the molecular weight of 17 kDa. This protein is known to interact with host proteins to cause hepatocellular carcinoma. HBx protein can interfere with cellular processes like oxidative phosphorylation, DNA repair, signal transduction, transcription, and protein degradation. In most of the cases, the host genome is known to be integrated by the HBx open reading frame gene, and such integration will play a crucial role in the development of hepatocellular carcinoma (6, 7). Among all serotypes of the virus, and even in the *Hepadnaviridae* family, this protein is found to be conserved. The protein is known to be localized in the cytoplasm, but the nucleus of the hepatic cells may also get localized to some extent. Protein X also plays a crucial role in the virus replication and subsequent infection (8, 9). DNA is not directly targeted by this protein. Instead, several transcription factors such as are transcription factor I, transcription factor 2, RNA polymerase binding protein, a subunit of RNA polymerase, have been identified as targets of protein X. Further, it is possible that this protein can modify the cytoplasmic signal transduction pathways through 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 also suppresses the tumor regression activity of P53 (10, 11).

Materials 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 HBV. The sequence was searched from the proteomic section of the database. The sequence of protein X in the Uniprot is assigned by the accession number Q765E5, and the sequence 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 was 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 from 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 Pfam 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 generate 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, whole structure relaxation was done (23, 24). The refined models were assessed by several validation tools to select the best model and assess 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 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-score was calculated for overall model quality and deviation of total energy with respect to 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 on the basis of their relative propensity to possess the property described by the scale (35).

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 by X protein were obtained from the Protein Data Bank (PDB) (37). Most of the ligands selected were composed of very low molecular weight, and were also 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 HBV. PyRx 0.8 (Virtual Screening Tools) was used to screen the Ligand library of the PDB. File formats like SDF and PDB were used for the ligands. Single ligand file formats were then optimized using a program called Openbabel (40, 41).

Results

HBx is a HBV protein and its 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. HBx 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 was found to be 16.756 kDa, and the theoretical pI is 8.08, allowing purification by isoelectric focusing. Protein purification is measured mostly by spectrophotometric analysis. Extinction coefficient value represses the amount of light absorbed by the protein at a given wavelength. The extinction coefficient of the protein in water was found to be 8980 M^{-1} cm^{-1} at 280 nm. HMM scan (hidden Markov model) calculated the matching score between the query sequence and each domain found in Pfam library in bit score. Motifs found 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 contribute to the X protein transactivator domain (Figure 1). To predict the secondary structure of protein, full DSSP classification was adopted by SSPro8 server. The structure predicted comprises 20% α-helix, 22% extended β-strands, 8% turns, 3% bands, and the remaining 46% is random coils (17) (Figure 2). Figure 3 represents the antigenicity prediction of the HBx protein. On the graph, the Y-axis 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-axis represents the residue positions in the sequence. 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).

Phyre2 web server was used to build the 3-D model for the protein sequence. 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, the server performed loop modeling, and fold

MAARLYCQLDSSRDVLCLRPVGAESRGRPFARPLGTVSSPSPSAVPSDHGAHLSLRGLPV
CAFSSAGPCALRFTSARCMETTVNAHQILPKVLHKRTLGLPAMSTTDLEAYFKDRVFKDW
EELGEETRLMIFVLGGCRHKLVCAPSSCNFFTSA

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

CCCEEECCCCCCTTCEEECCCCCCCSSCCCCSCCCCCCCCCCCCCCCCTTCCCCCTTCCE
EEEETSSCEEEEEEEHHCCTCCCCHHCHHHHHHHHHHTTCCCCCHTCHHHHHHHHHHH
HHHHHCTCCEEEEEEEECCEEEEEECCCCCCEEECC

Figure 2. The secondary structure predicted by SsPro8. H: alpha-helix; G: 3-10-helix; E: ex-tended strand; B: beta-bridge; T: turn; S: bend; C: the rest.

recognition. The model produced and predicted by any *in silico* method must be refined by suitable algorithms to rebuild 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 were performed to find the best suitable refine model (Tables 1 and 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 initial model had many areas of high energy, which were greatly improved in the refined model, suggesting greater reliability.

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

Qualitative model energy analysis (QMEAN) is a composite scoring function describing the major geometrical aspects of protein structures. QMEAN 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 (Table 3).

In the Z-sore plot of Protein X, the particular value that contains the Z-scores of all experimentally determined protein chains in current PDB is displayed (Figure 8). 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.

Several of the ligands were selected from the PDB and drug bank. All the compounds were minimized for their energy and were lastly converted into PDBQT format. Amongst hundreds of the compounds, ten lead compounds were taken as lead molecules on the basis of their binding affinity to the protein (Figures 10 and 11).

Table 1. Refined models by Galaxy web servers

Sr.No.	Model	GDT-HA	RMSD	MolProbity	Clash score	Poor rotamers	Ramafavored
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

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

Sr.No.	Model	GDT-HA	RMSD	MolProbity	Clash scor	Poor rotamers	Ramafavored
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

Table 3. The QMEAN6 and component scores and their Z-scores of the initial and first models with respect to 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

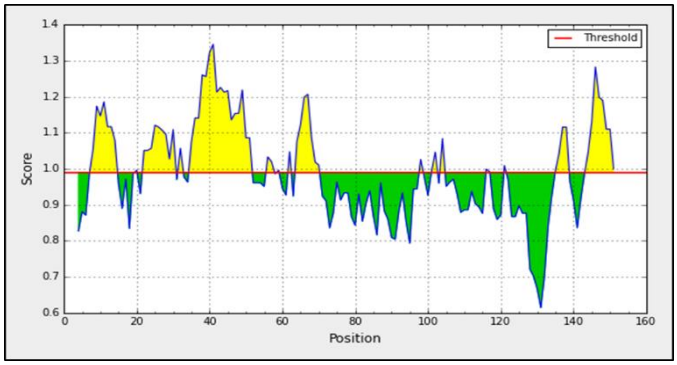


Figure 3. Chou & Fasman beta-turn prediction result.

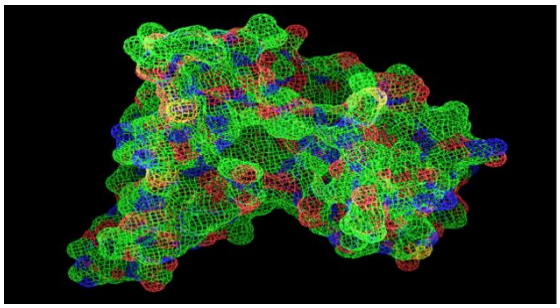


Figure 4. 3D model for protein X-generated from Phyre 2 unrefined.

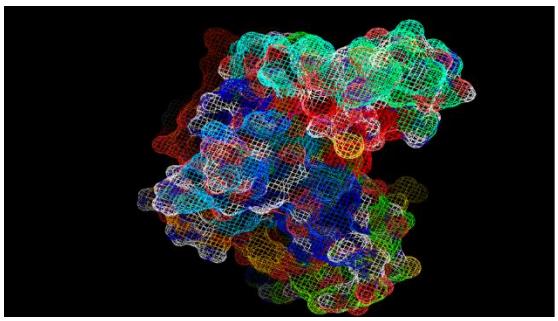


Figure 5. The three-dimensional structure of protein X, model 1, produced by Phyre2 and refined by Galaxy web servers.

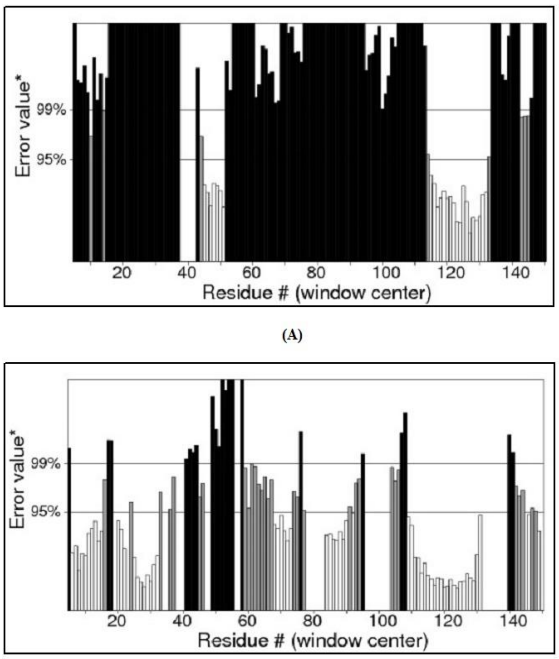


Figure 6. ERRAT plot shows error values for residues. The Y-axis represents the error value and the X-axis represents the amino acid residues of the protein model. An error value exceeding 99% confidence level indicates a poorly-modeled region. (A): the initial model with the quality of 17.730%; (B): the refined model 1 with the quality of 54.054%.

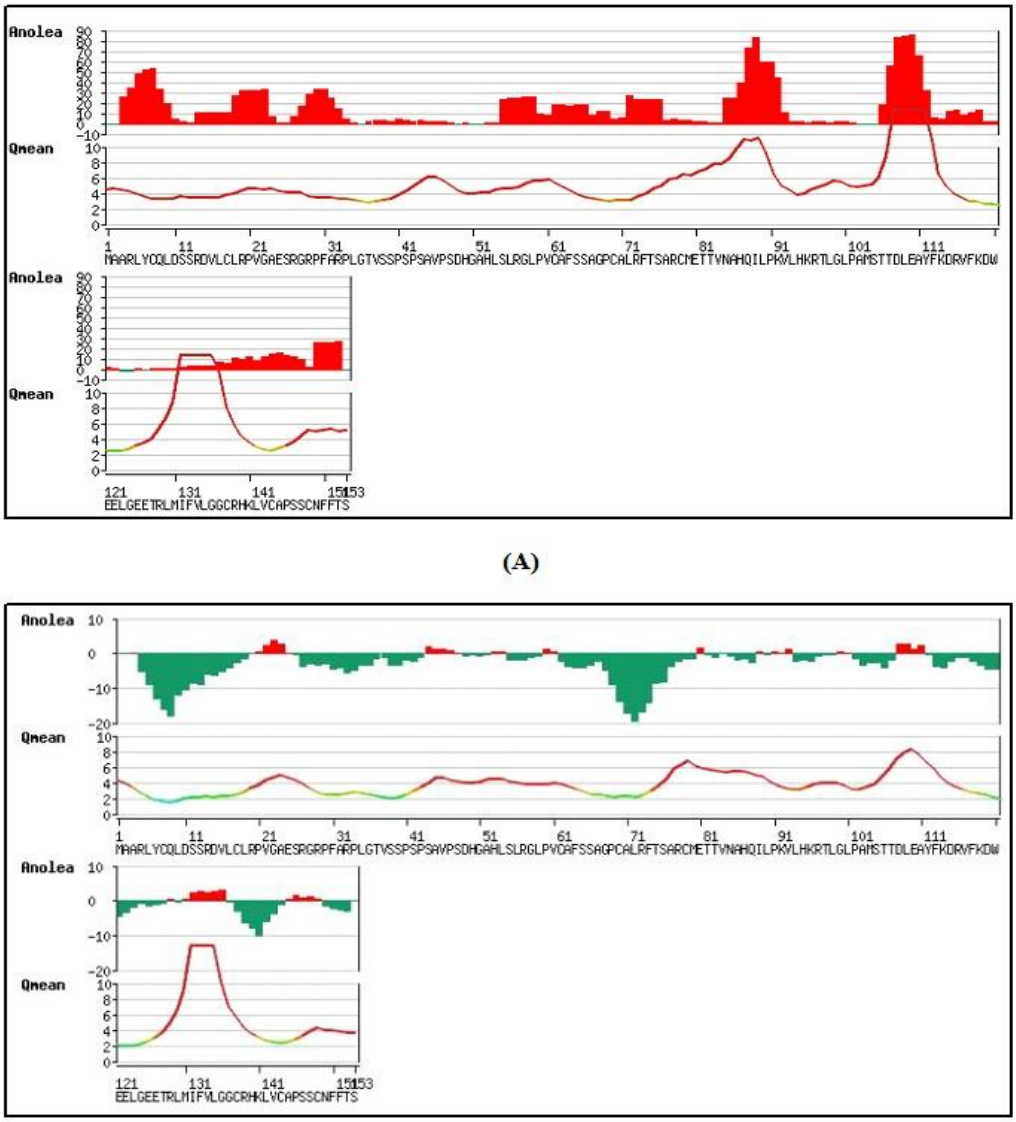


Figure 7. ANOLEA plots showing high energy zones as red. (A): the initial model; (B): the refined model 1 with the high energy zones greatly minimized and improved.

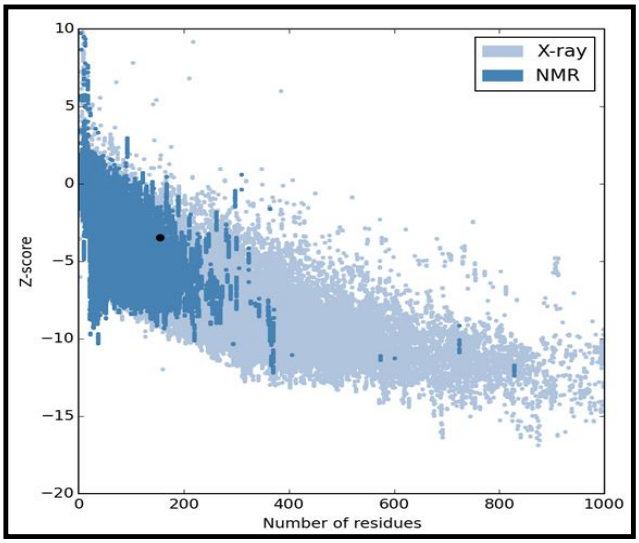


Figure 8. The Z-score plot of protein X (dot) determined by ProSA. The Z-score is -4.062, within the range of experimental native.

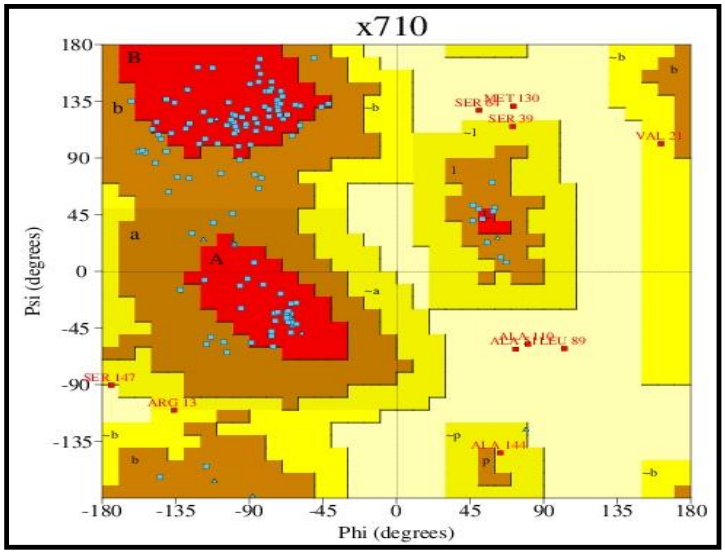
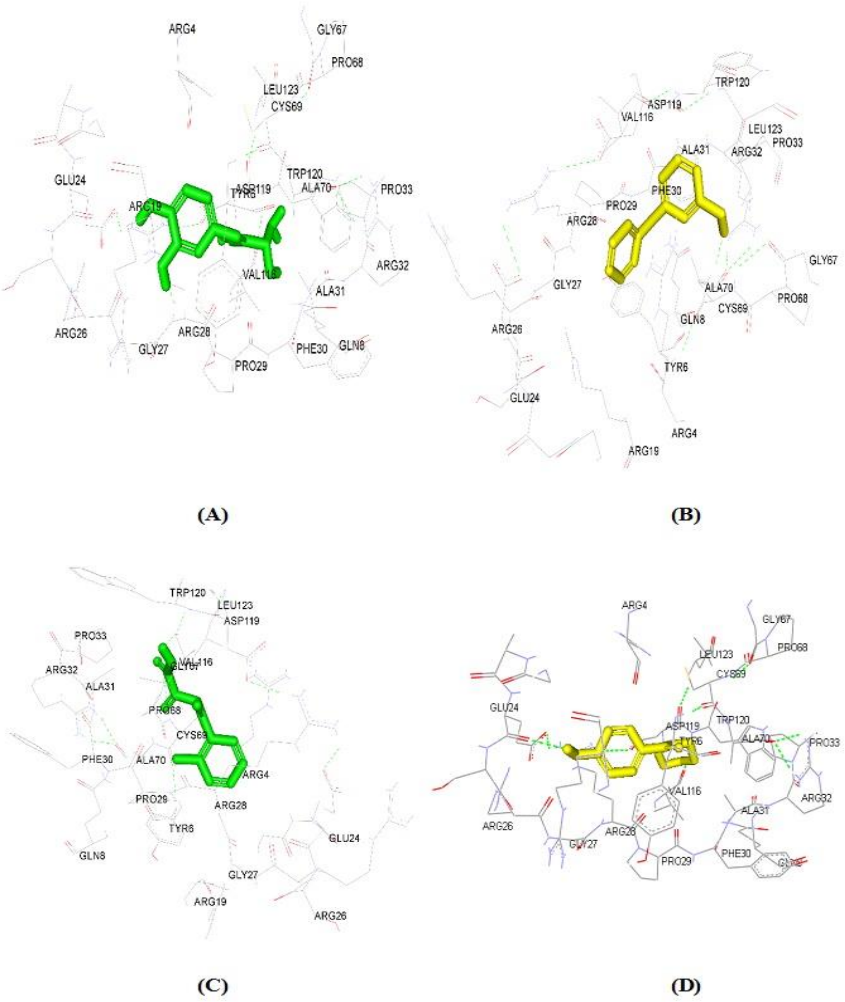


Figure 9. Ramachandran plot of the fourth model is determined by Procheck. The most favored regions are marked as A, B, and L. The additional allowed regions are marked as a, b, l, and p. All non-glycine and proline residues are shown as filled black squares, whereas glycines (non-end) are shown as filled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



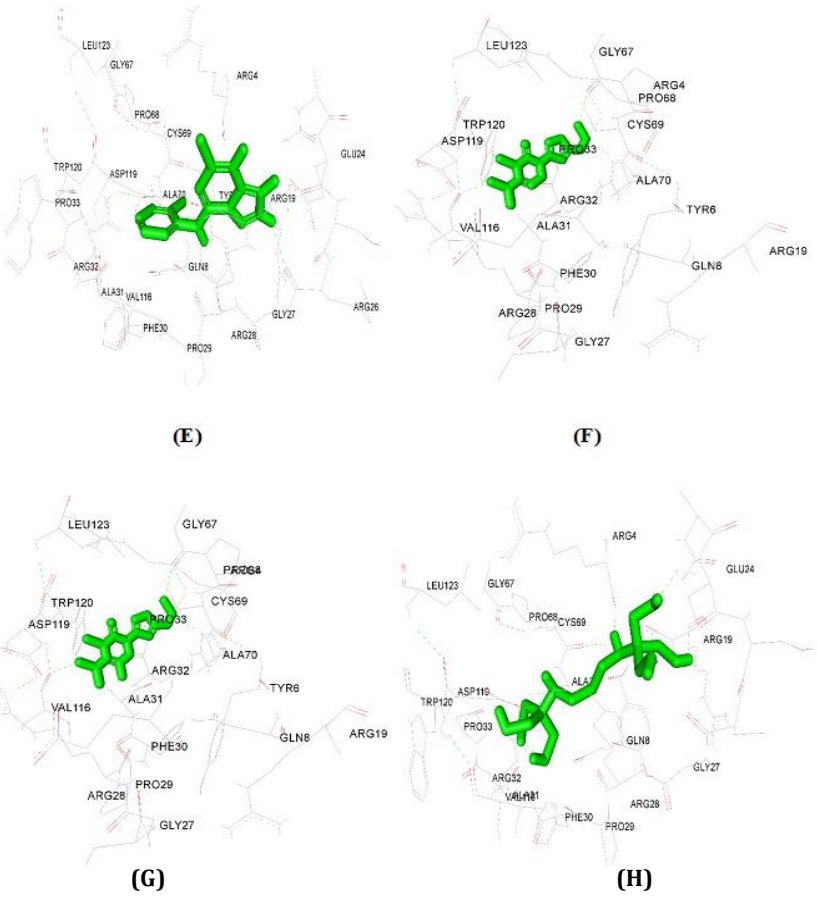


Figure 10. Docking interaction structure of lead molecules. (A): PDBOJW; (B): PDB0OO (C); DPB0OY; (D): DPB040; (E): DPB64Q; (F): DB00709; (G): DB00879; (H): DPBBP3.

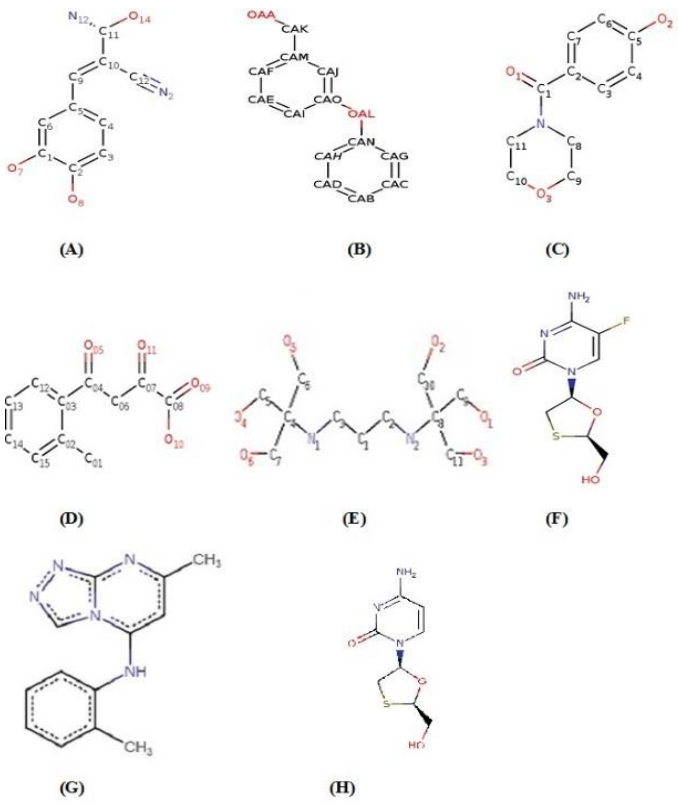


Figure 11. Chemical structures of lead molecules. (A): PDBOJW; (B): PDB0OO; (C): PDB0OY; (D): PDB040; (E): PDBB3P; (F): PDB00879; (G): PDB 64Q; (H): DB00879.

Discussion

HBx has an important role in HBV infection. The physico-chemical properties of the protein are crucial to characterize it. The biochemical parameters (16.756 kDa molecular weight, and pI of 8.08) are helpful to predict the location of the spot that may appear in the 2-D gel electrophoresis (42).

The extinction coefficient of the protein (8980 M^{-1} cm^{-1}) assumes that all cysteine pairs come from cysteine residues, and 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 above 40 indicates instability of the protein. The protein X was found to score of 60.28 and it is slightly unstable in the test tube (44). 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 a number of side chains of the aliphatic amino acids like alanine, valine, isoleucine, and leucine. The calculated aliphatic index of the protein was found to be 79.22 which represents slightly mesophilic nature (45).

The 3-D structure of a protein generally 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.

According to the QMEAN6 score, model 1 can be regarded as the best model amongst all model tested. This is further clear from the other contributors 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 been repeatedly found integrated into the genome of liver carcinoma cells (17, 48-49). 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).

According to 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.

From Ramachandran plot (Figure 9), it is very clear that for a refined model, model 1 carries a good percentage of residues that fall into favor regions.

From the above work, it is further recommended that the 3D structure of protein X could be utilized by 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 the protein X was found to be exceptionally fundamental for an infection.

Conflict of interest

The authors declared no conflict of interest.

References

1. Hassan M M, Li D, El-Deeb A S, et al. Association between hepatitis B virus and pancreatic cancer. J Clin Oncol. 2008;26:4557-62.

2. Dupinay T, Gheit T, Roques P, et al. Discovery of naturally occurring transmissible chronic hepatitis B virus infection among Macaca fascicularis from Mauritius Island. Hepatology. 2013;58:1610-20.

3. Kramvis A, Kew M, and Francois G. Hepatitis B virus genotypes. Vaccine. 2005;23:2409-23.

4. Magnius L O 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. Vos T, Barber R M, Bell B, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015;386:743-800.

6. Al-Khayyat M Z and Al-Dabbagh a G. In silico Prediction and Docking of Tertiary Structure of LuxI, an Inducer Synthase of Vibrio fischeri. Rep Biochem Mol Biol. 2016;4:66-75.

7. Miller R H and Robinson W S. Common evolutionary origin

of hepatitis B virus and retroviruses. Proc Natl Acad Sci U S A. 1986;83:2531-5.

8. Murakami S. Hepatitis B virus X protein: a multifunctional viral regulator. J Gastroenterol. 2001;36:651-60.

9. Gong D Y, Chen E Q, Huang F J, et al. Role and functional domain of hepatitis B virus X protein in regulating HBV transcription and replication in vitro and in vivo. Viruses. 2013;5:1261-71.

10. Diao J, Garces R, and Richardson C D. 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 Rev. 2001;12:189-205.

11. Feitelson M A and Lee J. Hepatitis B virus integration, fragile sites, and hepatocarcinogenesis. Cancer Lett. 2007;252:157-70.

12. Bastian F B, Chibucos M C, Gaudet P, et al. The Confidence Information Ontology: a step towards a standard for asserting confidence in annotations. Database (Oxford). 2015;2015:bav043.

13. Boutet E, Lieberherr D, Tognolli M, et al. UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View. Methods Mol Biol. 2016;1374:23-54.

14. UniProt: a hub for protein information. Nucleic Acids Res. 2015;43:D204-12.

15. Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy server, In: Walker, John M. (Ed.), The Proteomics Protocols Handbook. Humana Press;2005:571–607.

16. Magnan C N 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;30:2592-7.

17. Pollastri G, Przybylski D, Rost B, et al. Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. Proteins. 2002;47:228-35.

18. Kabsch W and Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983;22:2577-637.

19. Bateman A, Coin L, Durbin R, et al. The Pfam protein families database. Nucleic Acids Res. 2004;32:D138-41.

20. Kelley L A, Mezulis S, Yates C M, et al. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015;10:845-58.

21. Berman H M, Westbrook J, Feng Z, et al. The Protein Data Bank. Nucleic Acids Res. 2000;28:235-42.

22. Bennett-Lovsey R M, Herbert a D, Sternberg M J, et al. Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. Proteins. 2008;70:611-25.

23. Ko J, Park H, Heo L, et al. GalaxyWEB server for protein structure prediction and refinement. Nucleic Acids Res. 2012;40:W294-7.

24. Heo L, Park H, and Seok C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. Nucleic Acids Res. 2013;41:W384-8.

25. Colovos C and Yeates T O. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 1993;2:1511-9.

26. Sippl M J. Knowledge-based potentials for proteins. Curr Opin Struct Biol. 1995;5:229-35.

27. Sippl M J. Recognition of errors in three-dimensional structures of proteins. Proteins. 1993;17:355-62.

28. Wiederstein M and Sippl M J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 2007;35:W407-10.

29. Arnold K, Bordoli L, Kopp J, et al. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics. 2006;22:195-201.

30. Biasini M, Bienert S, Waterhouse A, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 2014;42:W252-8.

31. Guex N, Peitsch M C, and Schwede T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. Electrophoresis. 2009;30 Suppl 1:S162-73.

32. Kiefer F, Arnold K, Kunzli M, et al. The SWISS-MODEL Repository and associated resources. Nucleic Acids Res. 2009;37:D387-92.

33. Melo F and Feytmans E. Assessing protein structures with a non-local atomic interaction energy. J Mol Biol. 1998;277:1141-52.

34. Benkert P, Kunzli M, and Schwede T. QMEAN server for protein model quality estimation. Nucleic Acids Res. 2009;37:W510-4.

35. Benkert P, Tosatto S C, and Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. Proteins. 2008;71:261-77.

36. Chou P Y and Fasman G D. Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol Relat Areas Mol Biol. 1978;47:45-148.

37. Castrignano T, De Meo P D, Cozzetto D, et al. The PMDB Protein Model Database. Nucleic Acids Res. 2006;34:D306-9.

38. Shin J M and Cho D H. PDB-Ligand: a ligand database based on PDB for the automated and customized classification of ligand-binding structures. Nucleic Acids Res. 2005;33:D238-41.

39. Wishart D S, Knox C, Guo a C, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 2006;34:D668-72.

40. Trott O and Olson a J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31:455-61.

41. Morris G M, Huey R, Lindstrom W, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem. 2009;30:2785-91.

42. Sahoo M, Jena L, Daf S, et al. Virtual Screening for Potential Inhibitors of NS3 Protein of Zika Virus. Genomics Inform. 2016;14:104-11.

43. Bjellqvist B, Hughes G J, Pasquali C, et al. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis. 1993;14:1023-31.

44. Gill S C and Von Hippel P H. Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem. 1989;182:319-26.

45. Guruprasad K, Reddy B V, and Pandit M W. 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 Eng. 1990;4:155-61.

46. Ikai A. Thermostability and aliphatic index of globular proteins. J Biochem. 1980;88:1895-8.

47. Kidd-Ljunggren K, Oberg M, and Kidd a H. The hepatitis B virus X gene: analysis of functional domain variation and gene phylogeny using multiple sequences. J Gen Virol. 1995;76:2119-30.

48. Renner M, Haniel A, Burgelt E, et al. Transactivating function and expression of the x gene of hepatitis B virus. J Hepatol. 1995;23:53-65.

49. Sanchez R and Sali A. Large-scale protein structure modeling of the Saccharomyces cerevisiae genome. Proc Natl Acad Sci U S A. 1998;95:13597-602.