

QSAR Modeling of a Ligand-based Pharmacophore Derived from Hepatitis B Virus Surface Antigen Inhibitors

Alireza Mohebbi (✉ mohebbi-a@goums.ac.ir)

Golestan University of Medical Sciences and Health Services Medical School <https://orcid.org/0000-0003-2489-585X>

Fatemeh Sana Askari

Golestan University of Medical Sciences and Health Services

Reyhane Shaddel

Golestan University of Medical Sciences and Health Services

Azam Mirarab

Golestan University of Medical Sciences and Health Services

Morteza Oladnabi

Golestan University of Medical Sciences and Health Services

Research Article

Keywords: Ligand-based pharmacophore modeling, Hepatitis B virus, Quantitative structure-activity relationship, Hepatitis B virus surface antigen, HBsAg secretion inhibitor

Posted Date: June 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-611525/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Functional cure for Hepatitis B virus (HBV) by inhibiting HBV surface antigen (HBsAg) is crucial. We aimed to develop a predictive quantitative structure-activity relationship (QSAR) model on a ligand-based pharmacophore (LBP) derived from already known HBsAg secretion inhibitors in the present study.

Methods: A LBP model was developed using active HBsAg secretion inhibitors as both trainings- and test-sets using LigandScout v3.12 software. The best model with the highest score was used for high throughput screening (HTS) screening of a virtual library comprising 720,000 compounds. A QSAR model was developed by a stepwise multiple linear regression (MLR) on ~2700 descriptors with a confidence interval (CI) of 95%. The test set validated the QSAR model. The goodness of fit statistics evaluated the fitness of the model. A comparable R^2 and adjusted R^2 were considered as the lack of overfitting. Further RMSE and Q^2 statistics were measured for testing the model on the validation set. Principal component analysis (PCA) was also evaluated to estimate the predictor variables' associations and impact on the model.

Results: 34 active anti-HBsAg compounds were used to develop an LBP model. 9/34 of compounds with higher clustering pharmacophore-fit scores were tagged as the training set, and the rest of the inhibitors were used as the test set. The best model had a 0.8832 fit score. HTS resulted in 10 potential hit compounds with a fit score of 101.44 ± 0.65 . A QSAR model was developed with two response variables, including Yindex and GATS8m, with substantial variance information ($p < 0.05$). The model was well fitted ($R^2 = 0.9563$, $MSE = 0.0023$). The model was not predictive on the test set ($Q^2 = 0.00$, $RMSE = 0.8153$). The PCA results of two factors demonstrated a substantial variance data of both predictor variables.

Conclusion: The present study showed a reliable pharmacophore modeling based on known active inhibitors of HBsAg and a well-fitted predictive QSAR model on the LBP. The model can be applied to the chemical libraries fitted to the LBP model, and the QSAR equation would estimate the biological activities of the hit compounds with 95.63% accuracy with only two Yindex and GATS8m descriptors.

Background

One of the most treatment-challenging viral diseases is caused by the hepatitis B virus (HBV) infection. Chronic HBV (CHB) has influenced millions of people worldwide. HBV remains at the top10 of viral-caused human deaths, despite the availability of preventive vaccines and some treatment regimens. It is also associated with various clinical liver diseases, including cirrhosis, liver failure, and hepatocellular carcinoma (HCC). Endemic areas suffer from higher rates of liver diseases associated with HBV [1]. Moreover, co-infection of HBV with other infectious agents such as the Hepatitis Delta virus (HDV) also increases mortality rates of infected patients [2]. The key to potential treatment is to uncover the life cycle of HBV.

Seven proteins are encoded by ~3.2 kbp genome of HBV, of which surface antigen (HBsAg) has specific determinative roles, including an indication of HBV seroepidemiology, viral attachment/penetration, viral envelopment [3], trans-activator [4], virus-host response/irresponsiveness, and an indicator of viral replication status. It is also understood that sustained clearance of HBsAg from chronically infected patients indicates a functional cure [5]. There are three types of HBsAg, including L (large), M (middle), and S (small). Large HBsAg is responsible for attachment to the human sodium taurocholate co-transporting polypeptide receptor (hNTCP) on the hepatocytes [6–8]. HBsAg transcribes from residual cccDNA and integrated form of HBV genome [9,10]. This makes it challenging to achieve a sustained HBsAg clearance.

In the CHB patients, HBsAg-based recombinant vaccine does not induce well-tolerated immunity due to some reasons, like what has been observed in those with escape mutant quasi-species, carriers with occult infection untreated CHB patients. However, FDA-approved nucleos(t)ide analogs (NAs) are frequently used for these patients [11]. In this regard, a systematic analysis of 36 randomized controlled trials has shown that NAs are safe to use and successful in HBeAg seroconversion, HBeAg loss, and the achievement of undetectable HBV DNA [12]. However, in CHB patients undergoing long-term therapy with NAs, HBsAg seroclearance is rare [13]. A functional cure can therefore not be obtained alone by using current NA therapy. In addition, in patients under NA medication, cccDNA in residual HBV-infected hepatocytes will remain unaffected [14]. This is due to the evolutionary process of cccDNA formation via the repair mechanism of host DNA damage, and NAs only target viral polymerase. Therefore, the use of combination therapy of novel HBsAg suppressors and the elimination of cccDNA would lead to an HBV cure.

Hundreds of studies have been performed identifying the anti-HBsAg activities of novel natural products [15–17], synthetic compounds [18–20], and other macromolecules [21]. Meanwhile, no well-established model defines the role of physio-chemical descriptors of compounds in their anti-HBsAg activities. In this respect, the Quantitative Structure-Property Relationship (QSAR) modeling is a powerful analytical tool for developing a robust linear equation model between the biological activities of chemical compounds and their physio-chemical properties [22]. Soni et al. used this model and showed that the electron-withdrawing group's aromatic ring and polarizability parameters are correlated with the anti-HBV activity of 4-Aryl-6-Chloro-Quinolin-2-Ones 5-Aryl-7-Chloro-1,4-Benzodiazepine compounds [23]. QSAR is a routine analysis along with ligand-based drug discovery (LBDD) approaches. Since there is no crystallographic data on HBsAg, there might be a bias toward the discovery of ligands with similar physio-chemical properties.

Therefore, we used clustering and validation of a pharmacophore model by already known compounds with anti-HBsAg activities. The pharmacophore was used for HTS of an extensive virtual library of chemical compounds. A highly accurate QSAR model was established based on the validated pharmacophore model. The model was evaluated to explain the relation of the descriptors of compounds with their activities. Our findings showed an applicable linear model that could be used for drug discovery of potential anti-HBsAg compounds.

Materials And Methods

Data Sets and clustering

Thirty-four active inhibitors of HBsAg secretion have been derived from the literature (Figure1). ACD/ChemSketch (Advanced Chemistry Development, Inc.) has drawn up 3D schemes of these compounds and checked for the best tautomeric shape. After drawing, compounds were maintained in the format of mol files. The mol files have been converted to the structure data file (SDF) using OpenBabel software [24]. The compounds were structurally clustered using the fast clustering algorithm of LigandScout 3.12 [25]. For clustering, the measure of similarity was retained as its norm, Max.num. Conformances were updated to 7, and the calculation method of clustering distance was set to maximum.

3D pharmacophore modeling

Ligand-based modeling of pharmacophores was carried out by the software LigandScout 3.12. The compounds in the same cluster were flagged as training sets based on the clustering, and other structurally distant related compounds were used as the test set for pharmacophore model(s) validation. Settings have been retained by default for modeling, but the number of omitted features has decreased to zero. This decreases the number of hits that do not have adequate feature patterns that match the models, leading to an increased pharmacophore-fit score.

Virtual screening

A virtual library containing more than 720,000 compounds was established using LifeChemicals, ASINEX®, and Drug Bank databases. The library was developed from these compounds with the LigandScout Omega Best performance to keep the compounds in their best conformational states. Virtual screening is performed by using the library, as mentioned earlier, with default software settings. Hits with the highest pharmacophore fitting score were used for further analysis.

Descriptor prediction and quantitative structure-property relationship

Version 5.5 of the Dragon software was used to classify compounds' descriptors. These data are selected as independent (X) variables. A further dependent (Y) variable of compound biological activity was obtained from Liu et al. Descriptors were checked for non-zero variance, and those with zero value were removed. Multiple linear regression (MLR) was used to generate a QSAR model with MS-Excel v2019 and XLSTAT V2020. The model was established with fixed zero intercept and selected with a stepwise process of variable entry with the threshold probability of 0.05 (95% CI). The model was validated by the test set of compounds used for pharmacophore modeling. The model with good fitness was applied on the hit compounds resulted in the HTS. The goodness of fitness statistics was reported for evaluation of the model predictivity. Close values of R² and adjusted R², as well as threshold of the variables, were considered as the lack of overfitted model

Model validation and PCA

The QSAR model was validated by its accuracy in estimation of the activity of training-set ligands. Scatter plots for the biological and predicted activities of ligands were plotted to elicit a significant linear correlation. The efficacy of the model was also examined with external validation. The test used in the modeling of the pharmacophore was used for the external test set. This external validation was done to check further reliability of the model. Principal Component Analysis (PCA) was also performed to investigate further the role of descriptors in the biological activities of anti-HBV compounds. Data was autoscaled when the proper variance of descriptors was less than 1e-24. Two PCs were highlighted to investigate scores and loadings for the independent descriptor variables. The PC data were graphed by MS-Excel v2019.

Applicability domain

In order to apply the QSAR model for new compounds, the validated model was used to predict the activity of the screened hit compounds. The compounds were chosen based on their pharmacophore-fit scores of the screened virtual library. We assumed the model would work fine since the hit compounds were obtained by screening through the pharmacophore model. However, The Tanimoto similarity was calculated on binary linear fingerprints of the training set on which the MLR model was established.

$$\text{Tanimoto coefficient} = \frac{c}{a + b - c}$$

The variable c is the number of features (or on-bits in the binary fingerprint) common in both compounds, while a and b are the number of unique features in one or the other compound, respectively. The Tanimoto ranges from 0 to 1, with higher values indicating more remarkable similarity than lower ones.

Results

Pharmacophore modeling

Thirty-four chemical compounds with known anti-HBsAg secretion activities were used for clustering. Based on conformational structures of the compounds, one model with the highest pharmacophore-fit score (90.99±8.49) was selected for virtual screening. The model has resulted in a structural relationship of nine compounds, which was validated by remaining non-clustered compounds as the test set (n = 25). The compounds were including Osthol (score = 72.69), Compound 39 (score = 85.76), Compound 40 (score = 93.81), Compound 41 (score = 97.93), Compound 42 (score = 96.63), Compound 43 (score = 84.38), Compound 47 (score = 88.54), Compound 48 (score = 99.67) and Compound 49 (score = 99.59). The pharmacophore model is provided in the supplemented

materials (see Additional file 1). The pharmacophore model (Figure 2) consisted of nine features: four hydrophobic regions, three aromatic rings, one hydrogen bond acceptor, and one hydrogen bond donor site.

Virtual screening of chemical compounds

An HTS was performed with the validated pharmacophore model on a virtual library consisting of chemicals of three LifeChemicals, ASINEX®, and DrugBank databases. The HTS results showed ten compounds with the best matching features with the highest pharmacophore-fit score. The hit compounds were aligned with the pharmacophore model (Figure 3). The average pharmacophore-fit score of the hits was 101.44 ± 0.65 . Hit number 1, with a score of 102.934, had the highest fitness within the pharmacophore.

Establishment of QSAR model based on validated pharmacophore

A QSAR model was generated with 2489 descriptors including, constitutional descriptors ($n = 48$), topological descriptors ($n = 119$), walk and path counts ($n = 47$), information indices ($n=47$), 2D autocorrelations ($n = 96$), connectivity indices ($n = 33$), edge adjacency indices ($n = 107$), burden eigenvalues ($n = 64$), topological charge indices ($n = 21$), eigenvalue-based indices ($n = 44$), functional group counts ($n = 154$), atom-centered fragments ($n = 120$), molecular properties ($n = 29$), 2D binary fingerprints ($n = 780$), and 2D frequency fingerprints ($n = 780$). The model was established on two predictors, including GATS8m and Yindex and further validated by the test set. 95.63% of the biological activity of the compounds as explained by the model ($R^2 = 0.9563$, $MSE = 0.0023$, $PRESS = 0.0371$, $Q^2 = 0.8978$). The RMSE of the test set was 0.8153 ($Q^2 = 0.0$), indicating no predictivity of the model for the test set. No overfitting was observed, as the predictor variables had significant information in the model (p -value < 0.05) and the R^2 (0.9563) was very close to the Adjusted R^2 (0.8893). The analysis of variance was investigate to evaluate the effect of each predictor of the presented model. As a result, both variables had significant amount of variance in the model. The linear equation is shown below:

$$\text{Activity (mM)} = -0.2032 \times \text{Yindex} + 0.1386 \times \text{GATS8m}$$

The result of model parameters showed a substantial role of each predictor on the biological activity of the compounds (Table 1). The residuals showed normal distribution (Shapiro-Wilk $W = 0.9394$, p -value = 0.574) of the linear model (Figure 4A). Standardized residuals were also plotted against the test set values (Figure 4B and C). No specific trends were observed, suggesting the model is correct.

The model's applicability domain was checked by the Tanimoto similarity matrix of linear fingerprints of training-set compounds. The results (Table 2) showed a relationship between most of the compounds. However, Osthol, with a 0.072 ± 0.010 score, had the lowest relationship with other compounds. This might be because of the small size of Osthol lacking some other topological fingerprints identical to other compounds. However, as the Osthol had an acceptable pharmacophore-fit score, it was used for QSAR modeling. The model was applied to the hit compounds to predict their biological activity (Figure 5). Hit-1 with the predicted activity of 0.0009 ± 0.016 mM showed the lowest biological activity. The predicted biological activity of the hits number 2 and 9 was -0.007 ± 0.016 mM and -0.022 ± 0.017 mM, indicating a range of positive to negative activity values. This could result from a minor standard error of the model and the values can still be acceptable.

Principal component analysis of the predictors

PCA was used to reduce the dimension of variance of the two predictors into principal factors (F) or eigenvectors. As shown in the Scree plot (Figure 6A), 100% of the variance of data are allocated into F1 (73.04%) and F2 (26.96%). As the model was established and fitted on two variables, it was rational that all variance was divided into Yindex and GATS8m. The monoplots (Figure 6B) of the two factors illustrates the relationship between variables. The exact size of the predictors indicates the equal role of each variable in the F1. The angle between Yindex and GATS8m demonstrated the almost independent role of each descriptor in the linearity of the model.

Discussion

The unavailability of a functional cure for CHB has led to WHO's implementation of the first global strategy to eliminate HBV infection as a public health threat by 2030 [26]. The viral surface antigen determines both functional and complete cure by its secretion levels in the serum. It also determines the host antiviral response, and its level in blood is positively correlated with intrahepatic HBV replication [27]. The sustained suppression of HBsAg secretion as a functional cure can be pursued or combined with conventional anti-HBV medicines or novel inhibitors of cccDNA formation to complete the treatment. In this study, we developed and validated a pharmacophore and QSAR model to discover potent compounds with anti-HBsAg activities. The model can be applied to compounds that are screened through the pharmacophore model.

In the present study, a pharmacophore model was developed and validated by 34 active inhibitors of HBsAg secretion. The complete information about the inhibitors is reviewed by Liu et al. [28]. Nine compounds having 6-chloro-4-(2-chlorophenyl) quinolone residues were clustered into the training set based on their structural relationship and validated by the rest of the compounds as the test set. Screening results have also shown fitted compounds to have the same matching pharmacological features. In this regard, chlorophenyl and quinolone groups provided a hydrophobic axis. In addition, the quinolone group also provides amide residue as an HBD. This implicates similar prerequisite features resulted as LBP modeling of anti-HBsAg compounds.

As the test set validated the pharmacophore, a predictive QSAR model was developed by a stepwise MLR and internally validated by the test set. Two independent predictors, including Yindex and GATS8m, were identified as good predictors of the response variable. Correlation matrix and PCA factors showed no significant associations between descriptors. This implies a substantial impact of each variable separately on the predictivity of the QSAR model. The model's applicability on the compounds screened with the pharmacophore model was also accompanied by the Tanimoto similarity search. We observed less similarity between Osthol and other compounds of training-set. This was partly due to the lack of some chemical residues in the Osthol. However, due to the high pharmacophore-fit score and match pharmacophore features of Osthol to the model, we assumed that its removal from the model would not change the outcome of the QSAR model. The linear regression model was used to predict the biological activities of the identified hits. The results showed a range of 0.0009 mM to 0.1 mM anti-HBsAg activity of the compounds. The activity values of the LBP-derived compounds were comparable to the training-set (0.141±0.143 mM), suggesting the chance of discovery of even more potential anti-HBsAg inhibitors. The coefficient and the PCA factors suggest a direct association of biological activity with GATS8m and an indirect relationship with Yindex values.

Yindex (Balaban Y index) is a topological index [29] and had a significant association with compounds' biological properties, as shown by PCA. GATS8m (Geary autocorrelation of lag 8 weighted by mass) is a 2D autocorrelation descriptor defined by the Geary coefficient [30]. Eigenvalues and score analysis showed a direct association of the descriptors with factor 1, suggesting both predictors have an identical variance for the QSAR model. In addition, there was no association between Yindex and GATS8m due to ~90° angle between the two descriptors.

The study was designed to build a ligand-based pharmacophore modeling and HTS of possible anti-HBsAg small molecules and establish a well-fitted predictive QSAR equation based on the LBP model. The QSAR model was established hardly by using different sets of descriptors (data are not shown here), and we also suffered from model overfitting due to higher numbers of predictors, which their existence in the model was not significantly necessary. After several modeling, we achieved a model well fitted using two independent descriptors with significant information in the model. The pharmacophore model can be used for future discovery in the setting of different servers (as we have previously shown [31]) of huge virtual libraries of compounds with predictable anti-HBsAg activity (with a range of nM to mM) by implementing the presented QSAR model.

Conclusions

The present study demonstrated a reliable QSAR equation based on a ligand-based pharmacophore modeling procedure to discover potent active HBsAg inhibitors. The highlighted hits with predicted anti-HBsAg activities could be used for future in vitro or in vivo studies.

List of abbreviations: Hepatitis B virus, HBV; HBV surface antigen, HBsAg; Quantitative Structure-Property Relationship, QSAR; high throughput screening, HTS; Chronic HBV, CHB; hepatocellular carcinoma, HCC; Hepatitis Delta virus, HDV; human sodium taurocholate co-transporting polypeptide receptor, hNTCP; ligand-based drug discovery, LBDD; ligand-based drug discovery, LBDD; Geary autocorrelation of lag 8 weighted by mass, GATS8m; Balaban Y index, Yindex.

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: All the authors gave their consent for submitting the manuscript to the journal.

Availability of data and materials: The dataset(s) supporting the conclusions of this article is(are) included within the article (and its additional file(s)).

Competing interests: Not applicable.

Funding: #110671

Authors' contributions: All the authors have contributed to supervision, data gathering, analysis, and academic writing.

Ethics approval: Not applicable.

Acknowledgements: Not applicable.

References

1. Rybicka M, Bielawski KP. Recent advances in understanding, diagnosing, and treating hepatitis B virus infection. *Microorganisms*. 2020;8:1–29.
2. Miao Z, Zhang S, Ou X, Li S, Ma Z, Wang W, et al. Estimating the global prevalence, disease progression, and clinical outcome of hepatitis delta virus infection. *J Infect Dis*. 2020;221:1677–87.
3. Ueda K, Tsurimoto T, Matsubara K. Three envelope proteins of hepatitis B virus: large S, middle S, and major S proteins needed for the formation of Dane particles. *J Virol*. 1991;65:3521–9.
4. Hildt E, Saher G, Bruss V, Hofschneider PH. The hepatitis B virus large surface protein (LHBs) is a transcriptional activator. *Virology*. 1996;225:235–9.
5. Mohebbi A, Lorestani N, Tahamtan A, Kargar NL, Tabarraei A. An overview of hepatitis B virus surface antigen secretion inhibitors. *Front Microbiol*. 2018;9:1–9.

6. Erratum to: Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *eLife* 1:e00049. doi: 10.7554/eLife.00049. *Elife*. 2014;3.
7. Eller C, Heydmann L, Colpitts CC, Verrier ER, Schuster C, Baumert TF. The functional role of sodium taurocholate cotransporting polypeptide NTCP in the life cycle of hepatitis B, C and D viruses. *Cell Mol Life Sci*. 2018;75:3895–905.
8. Lempp FA, Wiedtke E, Qu B, Roques P, Chemin I, Vondran FWR, et al. Sodium taurocholate cotransporting polypeptide is the limiting host factor of hepatitis B virus infection in macaque and pig hepatocytes. *Hepatology*. 2017;66:703–16.
9. Tout I, Loureiro D, Mansouri A, Soumelis V, Boyer N, Asselah T. Hepatitis B surface antigen seroclearance: Immune mechanisms, clinical impact, importance for drug development. *J Hepatol*. 2020;73:409–22.
10. Tu T, Budzinska MA, Shackel NA, Urban S. HBV DNA integration: Molecular mechanisms and clinical implications. *Viruses*. 2017;9.
11. Rezanezhadi M, Mohebbi A, Askari FS, Hosseini SD, Tabarraei A. Hepatitis B virus reverse transcriptase polymorphisms between treated and treatment-naïve chronically infected patients. *VirusDisease*. 2019;30:219–26.
12. Geng JS, Bao HN, Chen YL, Shi LL, Geng J, Wang Q, et al. Nucleos(t)ide analogues for the treatment of chronic hepatitis B: a systematic review with network meta-analysis. *Expert Rev Anti Infect Ther*. 2020;18:823–34.
13. Wong GLH, Chan HLY, Yuen BWY, Tse YK, Luk HWS, Yip TCF, et al. The safety of stopping nucleos(t)ide analogue treatment in patients with HBeAg-negative chronic hepatitis B. *Liver Int*. 2020;40:549–57.
14. Martinez MG, Villeret F, Testoni B, Zoulim F. Can we cure hepatitis B virus with novel direct-acting antivirals? *Liver Int*. 2020;40:27–34.
15. Hu Z, Hu J, Ren F, Xu H, Tan M, Wang Q, et al. Nobiletin, a novel inhibitor, inhibits HBsAg production and hepatitis B virus replication. *Biochem Biophys Res Commun*. 2020;523:802–8.
16. Khani H, Ghorbani M, Nojoomi F, Mohebbi A. Honey Bee Dry Venom Reduces Hepatitis B Virus Surface Antigen Secretion in PLC/PRF/5 Cell Line. *Int J Med Lab*. 2019;
17. Mohebbi A, Azadi F, Hashemi MM, Askari FS, Razzaghi N. Havachoohe (Onosma dichroanthum Boiss) Root Extract Decreases the Hepatitis B Virus Surface Antigen Secretion in the PLC/PRF/5 Cell Line. *Intervirolgy [Internet]*. 2021;64:22–6. Available from: <https://www.karger.com/DOI/10.1159/000512140>
18. Vaillant A. Nucleic acid polymers: Broad spectrum antiviral activity, antiviral mechanisms and optimization for the treatment of hepatitis B and hepatitis D infection. *Antiviral Res*. 2016;133:32–40.
19. Roehl I, Seiffert S, Brikh C, Quinet J, Jamard C, Dorfler N, et al. Nucleic Acid Polymers with Accelerated Plasma and Tissue Clearance for Chronic Hepatitis B Therapy. *Mol Ther - Nucleic Acids*. 2017;8:1–12.
20. Quinet J, Jamard C, Burtin M, Lemasson M, Guerret S, Sureau C, et al. Nucleic acid polymer REP 2139 and nucleos(T)ide analogues act synergistically against chronic hepadnaviral infection in vivo in Pekin ducks. *Hepatology*. 2018;67:2127–40.
21. Qian L, Fan H, Ju Y, Chen L, Li X, Ye X, et al. A peptide-based inhibitor of gp96 suppresses HBsAg expression and HBV replication by upregulation of p53. *J Gen Virol*. 2019;100:1241–51.
22. Muhammad U, Uzairu A, Ebuka Arthur D. Review on: quantitative structure activity relationship (QSAR) modeling. *J Anal Pharm Res*. 2018;7.
23. Soni P Das, Tripathi IP, Dwivedi MK. QSAR Study of Some Anti-Hepatitis B Virus Agents Comprising 4-Aryl-6-Chloro-Quinolin-2-Ones and 5-Aryl-7-Chloro-1,4-Benzodiazepines. 2017;
24. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An Open chemical toolbox. *J Cheminform*. 2011;3.
25. Wolber G, Langer T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J Chem Inf Model*. 2005;45:160–9.
26. Catton M, Gray G, Griffin D, Hasegawa H, Kent SJ, Mackenzie J, et al. 2017 international meeting of the Global Virus Network. *Antiviral Res*. 2018;153:60–9.
27. Ren F, Yang X, Hu ZW, Wong VKW, Xu HY, Ren JH, et al. Niacin analogue, 6-Aminonicotinamide, a novel inhibitor of hepatitis B virus replication and HBsAg production. *EBioMedicine*. 2019;49:232–46.
28. Liu N, Zhao F, Jia H, Rai D, Zhan P, Jiang X, et al. Non-nucleoside anti-HBV agents: Advances in structural optimization and mechanism of action investigations. *Medchemcomm*. 2015;6:521–35.
29. Balaban AT. Highly discriminating distance-based topological index. *Chem Phys Lett*. North-Holland; 1982;89:399–404.

30. Geary RC. The Contiguity Ratio and Statistical Mapping. Inc Stat. JSTOR; 1954;5:115.

31. Mohebbi A, Askari FS, Sammak AS, Ebrahimi M, Najafimemar Z. Druggability of cavity pockets within SARS-CoV-2 spike glycoprotein and pharmacophore-based drug discovery. Future Virol [Internet]. Future Medicine Ltd London, UK ; 2021 [cited 2021 Jun 4];fvl-2020-0394. Available from: <https://doi.org/10.1089/fvl.2020.0394>.

Tables

Table 1. The parameters of the QSAR model.

Source	Value	Standard error	t	Pr > t	Lower bound (95%)	Upper bound (95%)
Intercept	0.0000					
Yindex	-0.2032	0.0483	-4.2065	0.0040	-0.3174	-0.0890
GATS8m	0.1386	0.0166	8.3580	<0.0001	0.0994	0.1778

Table 2. Tanimoto results of similarity between linear fingerprints of the training-set.

Compounds	Osthol	Compound39	Compound40	Compound41	Compound42	Compound43	Compound47	Compound48	Compound49
Osthol	1.000	0.081	0.067	0.073	0.064	0.095	0.056	0.073	0.071
Compound39	0.081	1.000	0.756	0.366	0.355	0.478	0.189	0.353	0.339
Compound40	0.067	0.756	1.000	0.515	0.501	0.360	0.303	0.499	0.463
Compound41	0.073	0.366	0.515	1.000	0.780	0.589	0.466	0.920	0.859
Compound42	0.064	0.355	0.501	0.780	1.000	0.581	0.354	0.757	0.710
Compound43	0.095	0.478	0.360	0.589	0.581	1.000	0.240	0.570	0.549
Compound47	0.056	0.189	0.303	0.466	0.354	0.240	1.000	0.426	0.403
Compound48	0.073	0.353	0.499	0.920	0.757	0.570	0.426	1.000	0.836
Compound49	0.071	0.339	0.463	0.859	0.710	0.549	0.403	0.836	1.000
Mean±Std.dev	0.072±0.010	0.480±0.235	0.550±0.199	0.687±0.206	0.630±0.198	0.546±0.195	0.422±0.222	0.670±0.212	0.670±0.212

Figures

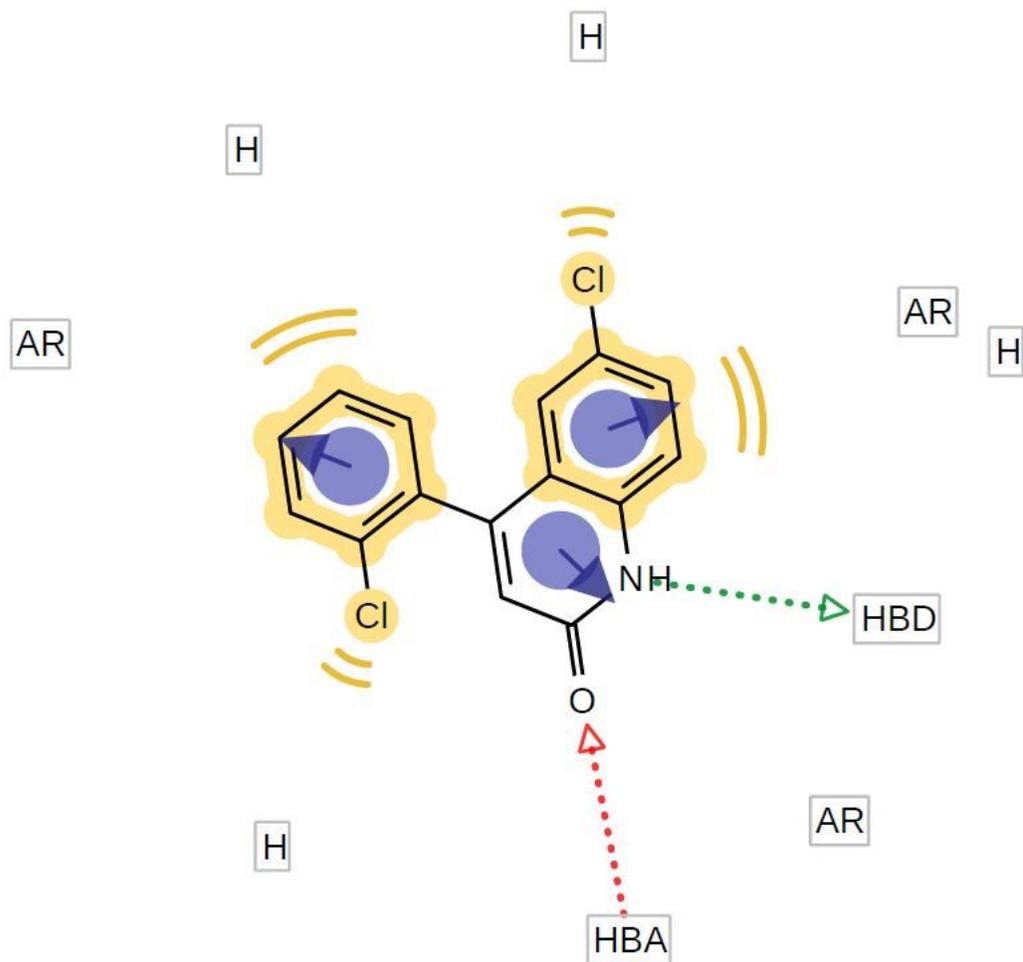


Figure 2

Schematic representation of the Pharmacophore model. As it is shown, Compound 40 is well fitted in the pharmacophore model. Of 10 predicted models, this model with a score of 0.8832 was selected for virtual LBDD. H: Hydrophobic region; AR: Aromatic ring; HBA: Hydrogen bond acceptor; HBD: Hydrogen bond donor.

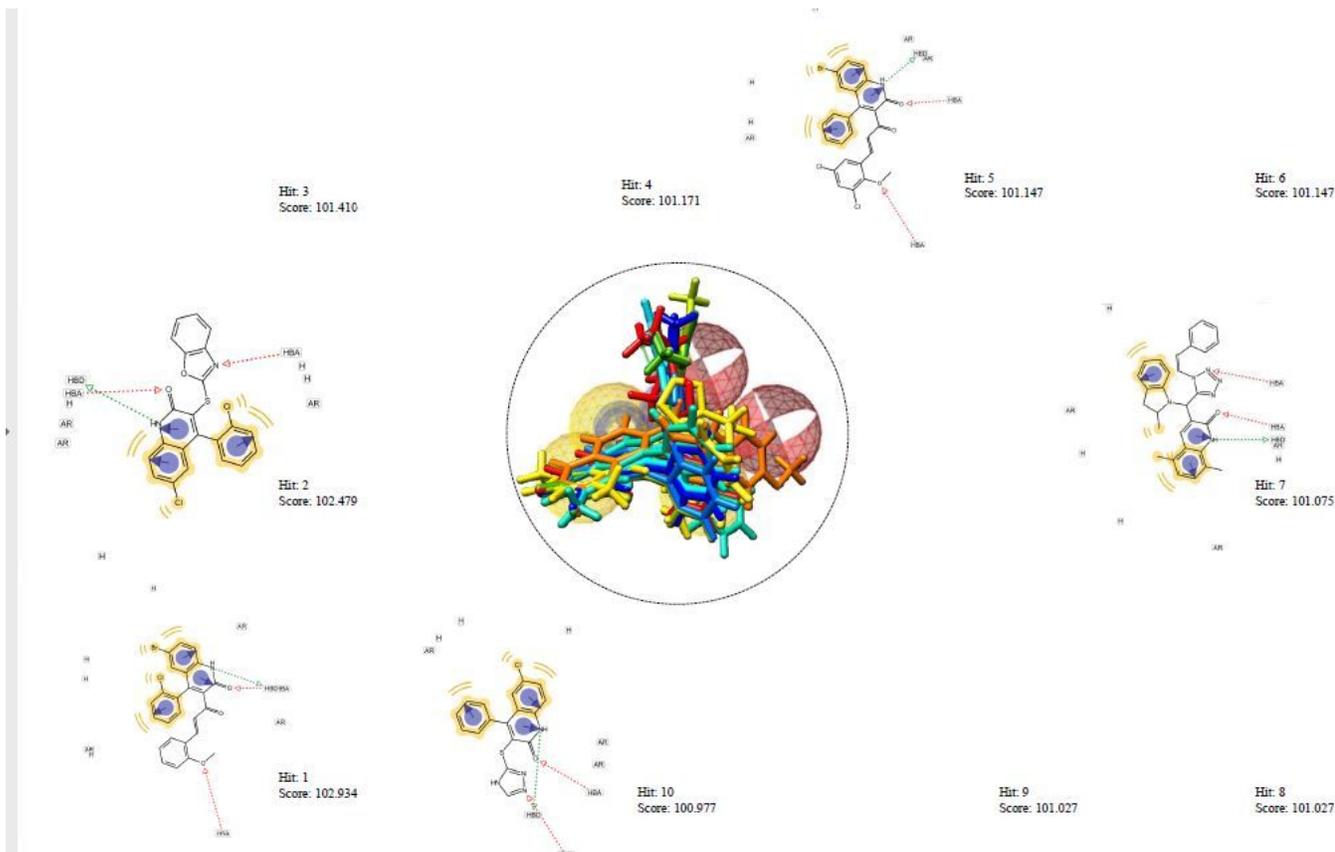


Figure 3

Schematic representation of top-10 hits with the highest pharmacophore-fit score. All hits had matched features with the pharmacophore. The pharmacophore-hit score of hit no. Three was the highest.

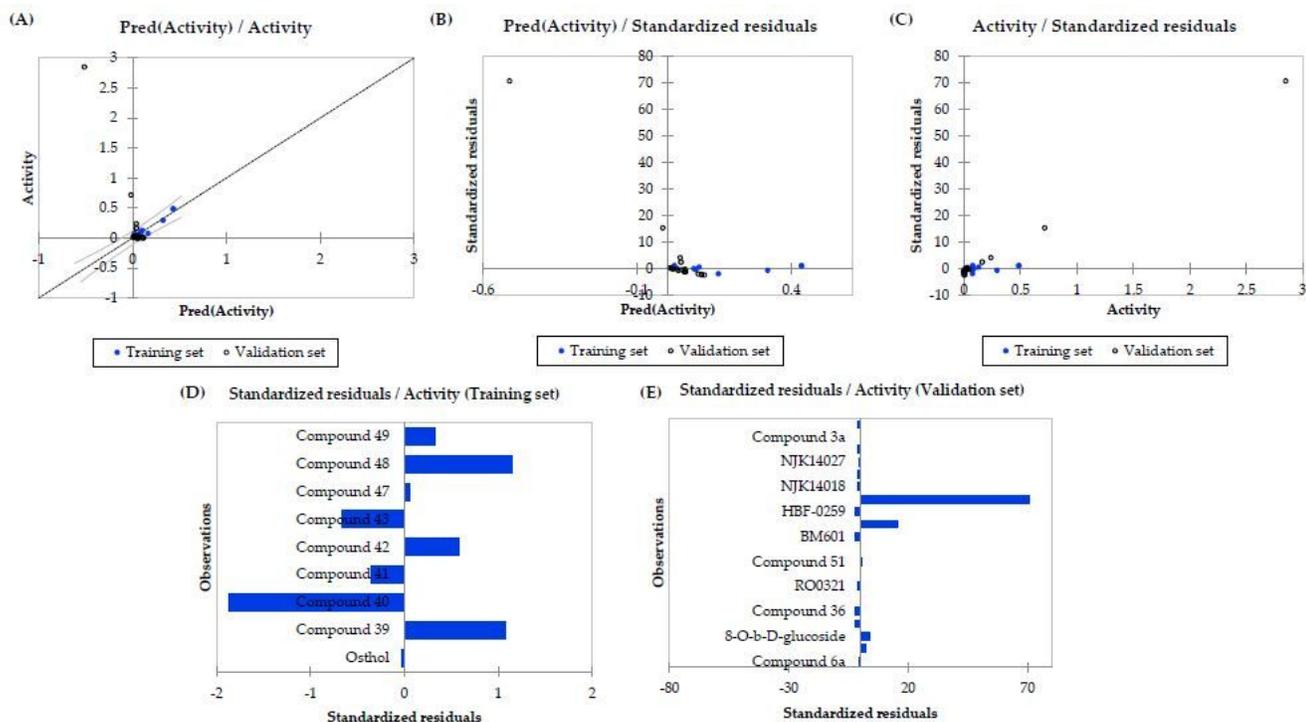


Figure 4

Series of plots generated to identify the predictive power of the QSAR model. A) shows the residuals of the training-set activity versus the predicted ones. The values reside within 95% CI. B) and C) are standardized residuals plots for training- and test-sets. Both plots represent the excellent quality of the model with

no significant trends. D) and E) histograms illustrate the standardized residuals of training- and test-sets. The standardized residuals of the training set are very narrow related to that in the test set.

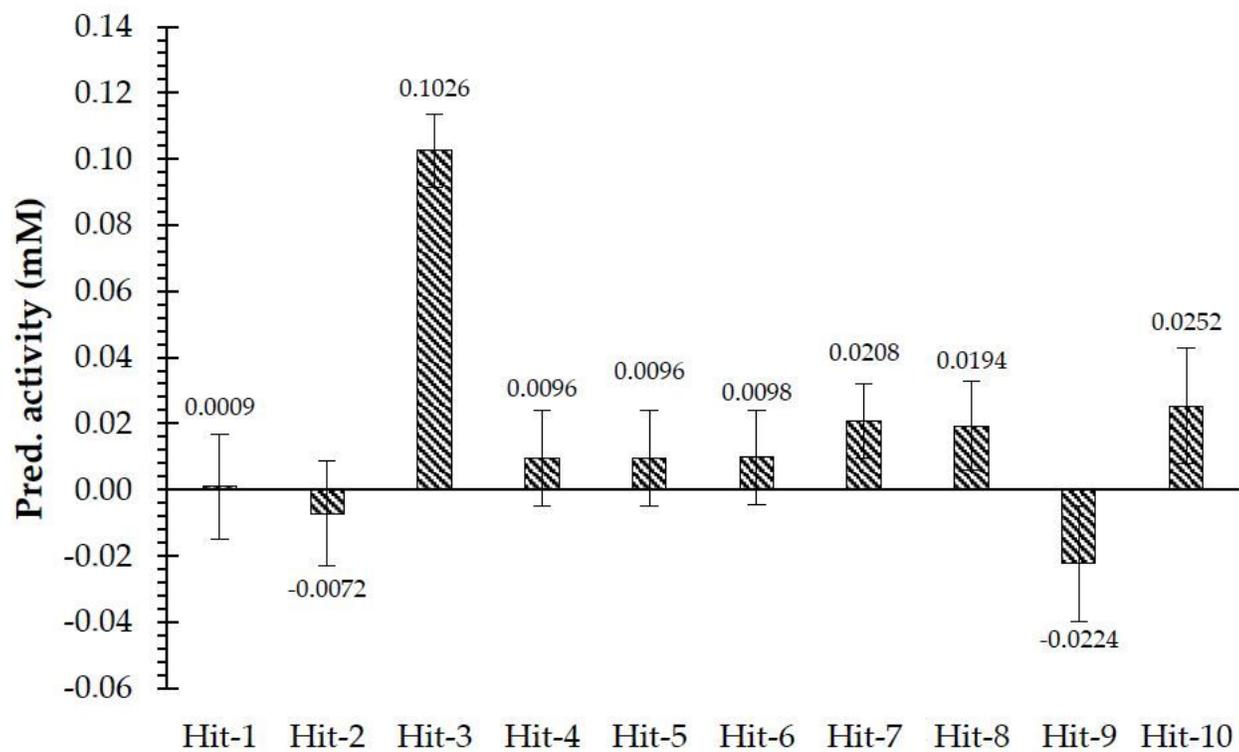


Figure 5

The predicted biological activities of the hit compounds. Hits number 1, 2, and nine had the lowest predicted anti-HBsAg activities. All the hit compounds' predicted activities were comparable and lower than that observed in the training set (0.141 ± 0.143 mM).

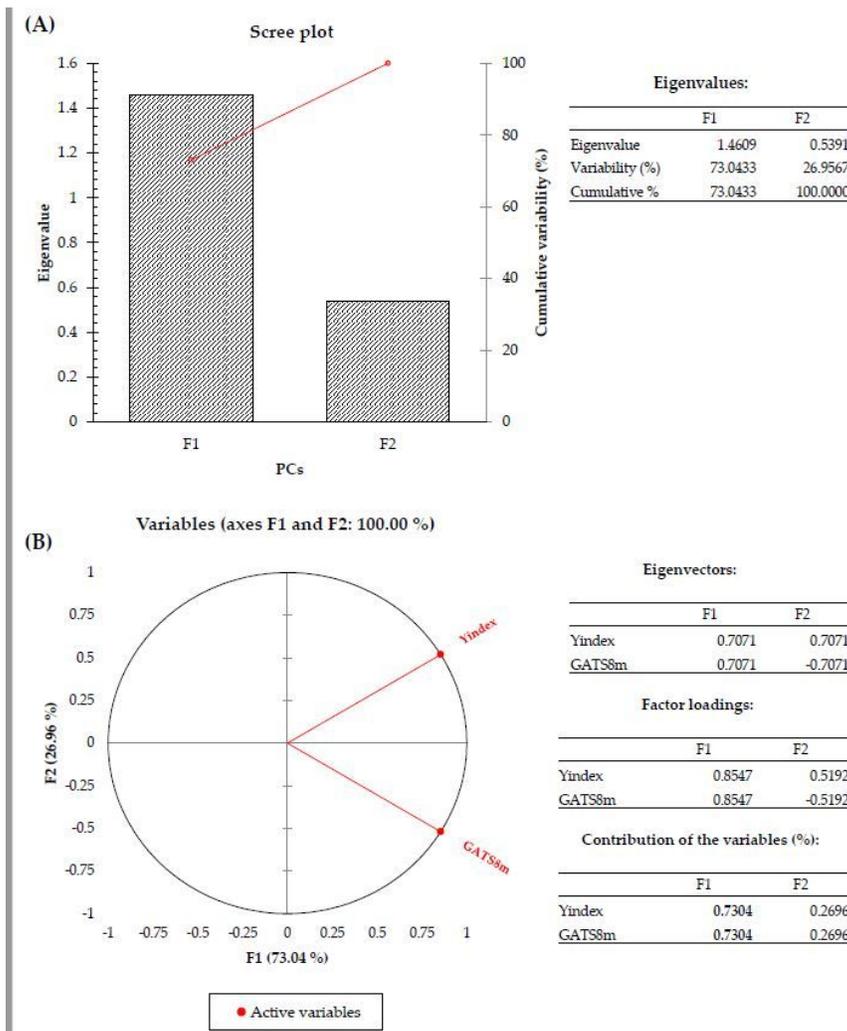


Figure 6

The PCA results of the predictors of the MLR model. A) Scree plot of two predictors that are reduced into two factors. Moreover, B) illustrates PCs of loading values. Values in bold correspond for each variable to the factor for which the squared cosine is the largest.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Model.pml](#)