Journal of Research in Pharmacy

Research Article

www.jrespharm.com

In silico design of potential HCV NS5B inhibitors: A comprehensive approach combining combinatorial library generation, ensemble docking, MM-GBSA calculations, QSAR model development, and molecular dynamics

Berin KARAMAN MAYACK [1, 2, *], Muhammed Moyasar ALAYOUBI [3], Mikail Hakan GEZGINCI [2]

1 Department of Pharmacology, School of Medicine, University of California, Davis, Davis, CA, United States.
2 Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Istanbul University, Istanbul 34116, Türkiye.
3 Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul 34956, Türkiye.

* Corresponding Author. E-mail: bkaramanmayack@ucdavis.edu, karaman.berin@gmail.com (B.K.M.); Tel. +01-916-305 99 41.

Received: 06 February 2025 / Revised: 18 March 2025 / Accepted: 18 March 2025

ABSTRACT: HCV is a blood-borne RNA virus that causes acute and chronic hepatitis, cirrhosis, liver failure, and hepatocellular carcinoma. In the present work, a large in silico combinatorial library was generated using the privileged substructures of existing inhibitors of the HCV NS5B protein. Next, we performed a multistep virtual screening process to identify novel HCV NS5B inhibitors. Additionally, we assessed the hit compounds' pharmacokinetic characteristics to evaluate their potential as drugs. Hit molecules with drug-like properties were classified with fingerprint-based chemical similarity clustering. Molecular dynamics simulations confirmed the stability of complexes and provided a comprehensive understanding of the molecular interactions between the novel molecule classes and HCV NS5B polymerase. The results of this study set the stage for developing new scaffolds as allosteric inhibitors of HCV NS5B protein for drug designing objectives and highlight the promising prospects of using privileged substructures for screening library construction in pharmaceutical research.

KEYWORDS: NS5B; Docking; MM-GBSA; QSAR; ADME; Molecular Dynamics

1. INTRODUCTION

Hepatitis C Virus (HCV) is a blood-borne RNA virus belonging to the Hepacivirus genus within the Flaviviridae family [1]. There are currently at least six major identified genotype variants of HCV, each with a unique geographic distribution and numerous subtypes [2, 3]. HCV can lead to acute and chronic hepatitis, cirrhosis, liver failure, and hepatocellular carcinoma. According to the World Health Organization (WHO), approximately 58 million people are chronically infected with HCV, with 1.5 million new cases reported annually [4]. Given its status as a significant global health threat, the pharmaceutical industry and academia are keen on developing new molecules based on previous medicinal chemistry efforts and structural information [5].

NS5B is a non-structural protein component of HCV that has gained global attention due to the absence of a comparable mammalian counterpart [6]. NS5B plays a crucial role in HCV's transcription and genome replication. It consists of a catalytic core with distinct subdomains such as 'fingers', 'palm', and 'thumb'. [7] Scientists have discovered several small molecules targeting the active site and the allosteric binding sites of NS5B polymerase, including nucleoside inhibitors (NIs) and non-nucleoside inhibitors (NNIs) [8-10]. The complexity of nucleoside pharmacology for NIs makes it challenging to predict their effectiveness and safety in humans. Although there has been progress in making them more potent, there is

How to cite this article: Karaman Mayack B, Alayoubi MM, Gezginci MH. In silico design of potential HCV NS5B inhibitors: a comprehensive approach combining combinatorial library generation, ensemble docking, MM-GBSA calculations, QSAR model development, and molecular dynamics. J Res Pharm. 2025; 29(2): 871-891.

ISSN: 2630-6344

https://doi.org/10.12991/jrespharm.1634330

still a need for new structural elements that can improve their pharmacokinetic properties, as well as their off-target effects. On the other hand, non-nucleoside inhibitors (NNIs) target the allosteric binding pocket within NS5B [11]. They allow for the development of selective anti-HCV drugs with fewer non-targeted side effects compared to NIs. Dasabuvir, an NNI, inhibits the action of NS5B palm polymerase [12, 13, 14]. It was approved by the FDA in 2014. Dasabuvir is used in combination with ombitasvir (an NS5A inhibitor), paritaprevir (an NS3-4A serine protease inhibitor), and ritonavir (a cytochrome P450-3A4 inhibitor) for the treatment of hepatitis C virus genotypes 1a and 1b [15]. Another NNI, Beclabuvir, which binds to the thumb site 1 pocket and inhibits NS5B, was approved in Japan in 2016 for treating genotype 1 chronic hepatitis C [16]. It is used in combination with the NS5A inhibitor Daclatasvir and NS3 protease inhibitor Asunaprevir [16]. Despite advances in treatment, many patients with HCV still need to undergo retreatment due to limited effectiveness against different HCV genotypes, variations in ethnicity, ineffective long-term treatments, resistance to HCV, side effects, and limited affordability of drugs [17, 18]. This emphasizes the ongoing need for further advancements in therapy. Furthermore, no medications currently target the thumb site 2 pocket, which prompted our research initiative.

Modulating the biological functions of HCV NS5B polymerase has been the focus of several drug discovery projects. Therefore, several crystal structures of HCV NS5B protein in complex with various ligands have been resolved in the Protein Data Bank (PDB) [19]. Molecular docking studies often involve a static protein structure and flexible ligands [20, 21]. However, the binding region's conformation can undergo significant changes based on the structure of the co-crystallized ligand and the internal flexibility of the protein's binding site. Molecular dynamics (MD) simulations are commonly used to generate the protein's conformational substates [22]. Nonetheless, it is difficult to sample the conformational space of proteins using MD simulations and to identify the most biologically relevant structural conformations. On the other hand, X-ray crystals or cryo-EM structures only provide a single native conformation of protein structures, leading to limited conformational heterogeneity. However, the structural data for biomolecules with various co-crystallized ligands is increasing rapidly. This provides a collection of valuable experimental protein conformations. Despite this, screening large compound libraries on a range of protein conformations can significantly increase the computational time required for docking calculations, making it impractical. A subset of representative protein conformations can be used to minimize the computing time [23-26]. To analyze the conformational changes in the thumb site 2 pocket caused by the ligand during our docking experiments, we initially grouped the existing crystal structures of HCV NS5B in the PDB. Subsequently, we conducted ensemble docking studies to pinpoint the most probable poses with the highest achievable binding affinity score.

Molecular docking experiments are valuable for identifying the likely way in which a ligand binds to a protein surface and the interactions between them. However, it remains challenging to estimate the changes in free energy upon ligand desolvation and intramolecular and conformational entropy changes upon binding, which are essential for evaluating the binding affinity of ligands to the receptor [21, 27-29]. Instead, the endpoint binding free energy calculations such as Molecular Mechanics Poisson–Boltzmann Surface Area (MM-PBSA) and MM-GBSA methods are popular tools for understanding the binding affinities of small ligands to their target macromolecules and postprocessing the docking solutions [24, 30-33]. These methods offer a balance between accuracy and computational efficiency. They provide improved accuracy compared to many molecular docking scoring functions and require fewer computational resources than alchemical-free energy methods. As a result, they represent a valuable intermediate option between empirical scoring and strict alchemical perturbation methods. Therefore, we carried out molecular mechanics generalized Born surface area (MM/GBSA) calculations to rescore the docking poses and compute the binding energies with a more physically reasonable description of the contributions to the binding.

QSAR modeling is an in-silico approach that helps identify differences in structural properties that lead to variations in biological activities [34-37]. This method uses regression and classification techniques to build a mathematical model that links the structural modifications to continuous (IC50, EC50, Ki, etc.) or categorical/binary (active, inactive, etc.) biological properties. As a result, this model can prioritize a large number of chemicals based on their desired biological activities and reduce the number of chemicals to be tested in the laboratory. In our current project, we developed both categorical and continuous QSAR models to forecast the activities of new chemical compounds created during combinatorial library enumeration. The compounds anticipated to be active were then analyzed for their drug-likeness and ADME properties. MD

simulations were carried out on the reference molecule of each of the four clusters, which were grouped based on their chemical similarity. These compounds exhibited stable interactions with the key residues Ser476 and Tyr477, as well as other commonly associated residues, Arg422, Leu474, Leu497, and Arg501, within the thumb site 2 pocket of NS5B, suggesting that they could serve as innovative allosteric inhibitors of HCV NS5B, offering a new starting point for developing novel therapeutics [38-43]. The overall workflow implemented in this work is presented in Figure 1.

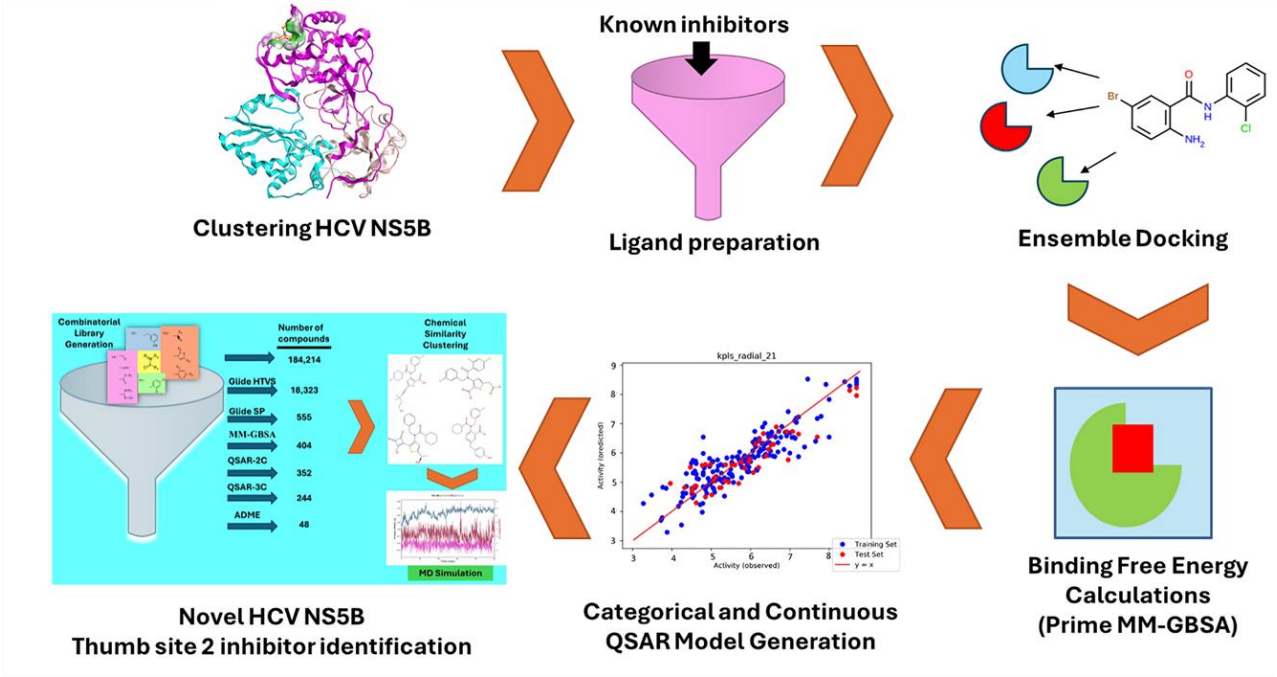

Figure 1. Overall workflow applied in the present work.

2. RESULTS and DISCUSSION

2.1. Clustering the thumb site 2 pockets of the known HCV NS5B crystals

To decrease the expense of the computing time of ensemble docking of our focused virtual combinatorial library, we used a subset of HCV NS5B protein conformations. We first mapped the topologies of the binding sites of all the protein structures using the SiteMap module of Schrödinger software. SiteMap analyses the surface of a protein, detects all the possible binding sites, and then ranks them based on SiteScore to eliminate the pharmaceutically irrelevant ones. By providing the names of the present ligands in the co-crystalized structures and a search distance of 5 Å, we mapped only the thumb site 2 regions of each of the 81 protein-ligand complexes. Next, we clustered the thumb site 2 pocket surfaces using Schrödinger's volume clustering script, which uses a hierarchical clustering with average linkage for clustering the structures. In total, 8 clusters were generated, and the protein structure that represents the cluster's center was chosen for the subsequent in-silico experiments.

Conducting ensemble docking studies of the newly generated in-house compound library of 184,214 molecules on this number of protein structures would require high computational power, rendering the study infeasible. Fortunately, clustering of binding pockets and selecting a representative grid of each cluster has shown to be successful in virtual screening studies. Thus, as a first step, we created a KNIME workflow to cluster the thumb site 2 pockets of the known HCV NS5B crystal structures in complex with thumb site 2 inhibitors. In total, eight protein structures with PDB IDs 1NHV [42], 1YVZ [79], 2I1R [65], 2WHO [80], 2O5D [66], 2HWH [64], 4J04 [63], and 4TLR [57] were predicted as representatives of different conformations adopted by thumb site 2 pockets upon inhibitor binding.

2.2. Ensemble docking and rescoring of docking poses with MM-GBSA method: Known HCV NS5B inhibitors

In the next step, 314 known thumb site 2 inhibitors of HCV NS5B were docked onto the eight representative structures using the Glide docking tool implemented in the Schrödinger program. Only the top-scored pose out of the eight poses retrieved from the ensemble docking campaign was used to generate the box plots showing the distribution of the Glide SP docking score, and the derived ligand efficiency scores, Glide ligand efficiency score, Glide ligand efficiency_ln score, and Glide ligand efficiency_sa score (Figure S1). In the case of the Glide SP docking score, the threshold value for obtaining the first quartile of the dataset was -7.87, -7.5, and -7.36 for the active, weak, and inactive compounds, respectively. According to these values, with a Glide SP score equal to or higher than -7.87, we can eliminate more than 75% of the weak binders and inactives from a prospective dataset. On the other hand, the threshold Glide ligand efficiency score to obtain the first quartile of the dataset was -0.23, -0.24, and -0.25 for the active, weak, and inactive compounds, respectively. Similarly, Glide ligand efficiency_ln and ligand efficiency_sa scores could not discriminate actives from weak and inactive compounds. For the Glide ligand efficiency_ln-based scoring, a threshold value of -1.43, 1.44, and -1.43, and for the ligand efficiency_sa-based scoring, -0.67, -0.73, and -0.69 was predicted to obtain the first quartile of the dataset for the active, weak, and inactive compounds, respectively. This shows us that the derived ligand efficiency scores performed poorly compared to the Glide SP docking score. To assess the discriminative power of each score produced by Glide SP docking, we further generated a ROC plot (Figure 2a) and calculated the AUC values. The Glide SP docking score had the highest performance, with an AUC value of 0.66, compared to the other ligand efficiency metrics obtained from Glide SP docking. As a result, it was used as a parameter to filter the compounds in the combinatorial library.

Consequently, docking solutions obtained from the Glide SP docking experiment of 314 known thumb site 2 inhibitors of HCV NS5B were analyzed using BFE calculations. The correlation coefficient (R2) of predicted and experimental binding free energies for different groups of ligands has been enhanced by rescoring an ensemble of docking poses rather than a single pose. Since the accuracy of the calculated relative binding free energies relies on the quality of modeled poses, all docking poses retrieved against eight representatives of inhibited HCV NS5B thumb site 2 conformations were used for rescoring with the Prime MM-GBSA method. Then, docking poses were re-ranked based on their MM-GBSA dG binding score, and only the top-ranked pose obtained for each ligand was used to generate box plots (Figure 4) and the ROC plot (Figure 2b). The distribution of the MM-GBSA dG binding score and ligand efficiency metrics (MM-GBSA ligand efficiency score, MM-GBSA ligand efficiency_ln score, and MM-GBSA ligand efficiency_sa score) is shown in Figure S2. Based on the box plots, we observed that the first quartile of the dataset corresponds to the following threshold MM-GBSA dG binding score values: -65.57 for active compounds, -62.9 for weak compounds, and -60.39 for inactive compounds. Using a threshold value of -65.57 or higher for the MM-GBSA dG binding score allows us to filter out more than 75% of weak and inactive molecules. Similarly, the threshold values for the MM-GBSA ligand efficiency_ln score were -14.78 for active compounds, -14.36 for weak compounds, and -13.8 for inactive compounds. However, the MM-GBSA ligand efficiency (-2.29 for actives, -2.21 for weak binders, and -2.37 for inactives) and MM-GBSA ligand efficiency_sa scores (-3.44 for actives, -3.31 for weak binders, and -3.55 for inactive molecules) were not as effective in discriminating actives from weak and inactive compounds as the first two metrics. It has already been reported in the literature that choosing a docking solution based on binding free energy may improve the accuracy of the docking pose [24, 25, 33, 79-81]. In addition, we also extracted information on the Glide SP score and related ligand efficiency metrics predicted for the top-ranked pose based on its MM-GBSA dG score. We compared the predictive power of different scores obtained from processing the docking poses using Prime MM-GBSA calculations. The MM-GBSA dG binding score outperformed the others, with an AUC of 0.71 distinguishing between active and inactive compounds. Therefore, we selected the MM-GBSA dG binding score as a secondary parameter to screen the hits from the molecular docking experiment of the combinatorial library.

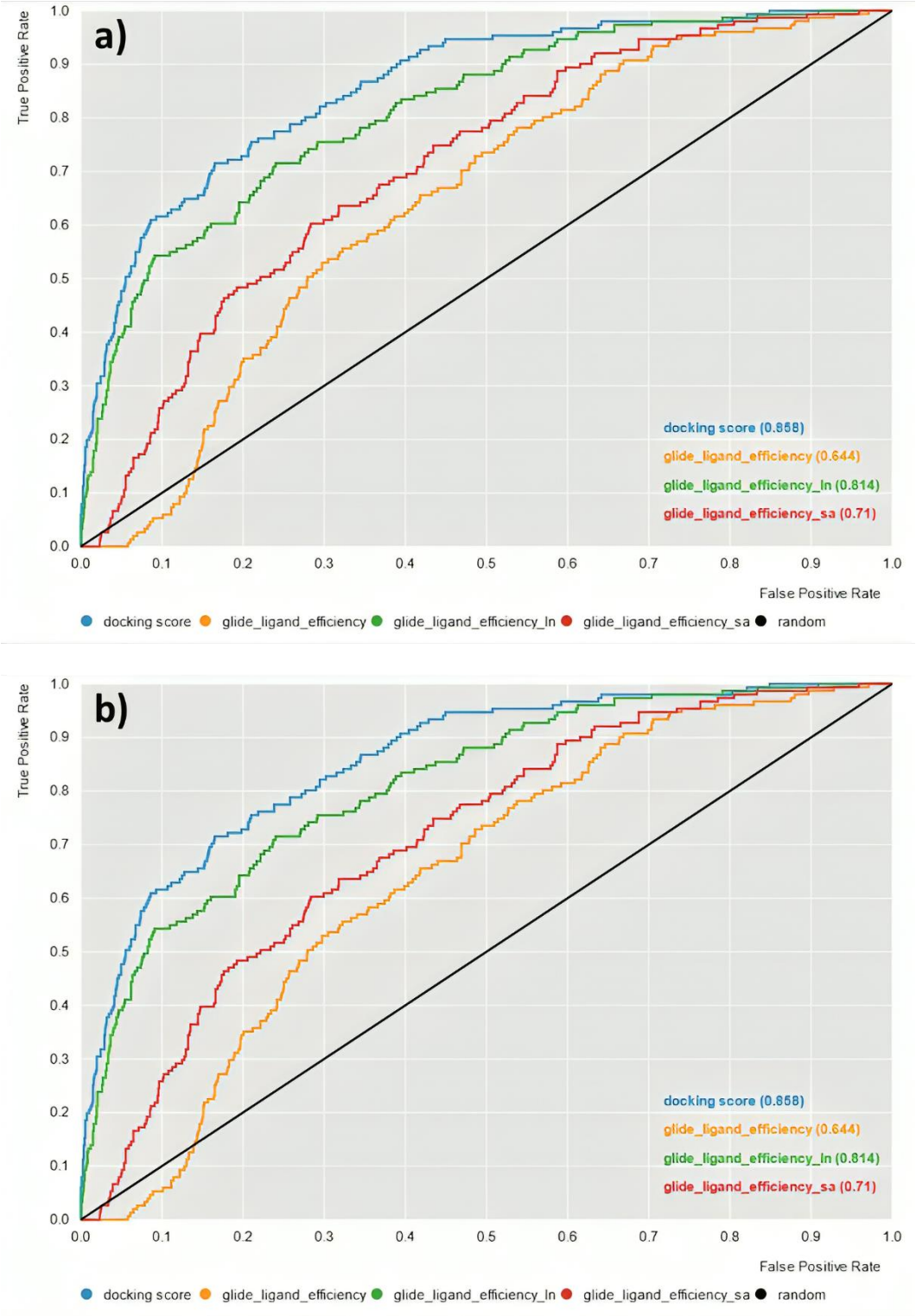

Figure 2. The ROC curve generated using the top-pose selected based on **(a)** the Glide SP docking score and **(b)** the MM-GBSA dG binding score of the known 314 inhibitors of HCV NS5B.

2.3. Ensemble Docking and Rescoring of Docking Results with MM-GBSA Method: Combinatorial Library

A virtual combinatorial library of was generated using the privileged substructures of known HCV NS5B thumb site 2 inhibitors from different series of compounds, including phenylalanine derivatives,

thiophene-2-carboxylic acid derivatives, and anthranilic acid derivatives in the literature. We have developed a step-by-step screening process to improve the effectiveness of our virtual screening efforts. This method involves gradually increasing computational resources and accuracy. Initially, we filter the combinatorial library using the Glide HTVS precision option to dock all the molecules. We then chose only the top 10% of the ligands based on the HTVS docking score, resulting in 18,323 selected molecules. These selected molecules undergo a second round of screening using a more thorough docking algorithm, the Glide SP scoring function. Only compounds with a Glide SP score of -7.87 or higher proceed to the next step, resulting in 555 hit molecules. Subsequently, we conduct binding free energy calculations using the Prime MM-GBSA method. Compounds with an MM-GBSA dG binding score of -65.57 or higher move on to the next stage, resulting in 404 compounds. Finally, these compounds undergo screening using QSAR models.

2.4. Generation and Application of QSAR Models

In our study, we developed three models for predicting the activity of inhibitors of HCV NS5B polymerase. The models are QSAR-2C, QSAR-3C, and QSAR-N. The statistical details for the top ten QSAR-2C models are provided in **Table 1**. All models performed well, scoring 0.82 or higher out of 1.00. The confusion matrices for the external validation set using the rp_32 model and the consensus QSAR-2C model are shown in **Table 2**.

Table 1. Accuracy rate per category for the top-scored ten models for QSAR-2C

Code	Score	Train (inactive)	Train (active)	Test (inactive)	Test (active)
rp_32	0.9510	0.9792	0.9286	1.0000	0.9048
rp_23	0.8885	0.8542	0.9167	1.0000	0.8571
rp_14	0.8841	0.8542	0.8929	0.9167	0.8095
rp_35	0.8496	0.8542	0.9286	0.8333	0.9048
rp_33	0.8449	0.7708	0.9167	0.7500	0.9524
bayes_linear_27	0.8304	0.9375	0.9881	0.8333	0.9524
bayes_desc_49	0.8275	0.7708	0.8690	0.9167	0.8571
bayes_desc_8	0.8239	0.7708	0.8690	0.8333	0.8571
bayes_desc_3	0.8236	0.7500	0.8810	0.9167	0.8571
bayes_desc_35	0.8206	0.8125	0.8214	0.9167	0.7619

Table 2. Confusion matrix of the external validation set using the top-ranked QSAR-2C model (rp_32) and consensus QSAR-2C model

	Top-ranked QSAR-2C Model		**Consensus QSAR-2C Model**	
	Active predicted	**Inactive predicted**	**Active predicted**	**Inactive predicted**
Active (46)	40	6	38	8
Inactive (9)	3	6	2	7
	The prediction ability			
Accuracy	(40+6)/ 55 = 83.64%		(38+7)/ 55 = 81.82%	
Sensitivity	40 / (40+6) = 86.96%		38 / (38+8) = 82.61%	
Specificity	6/ (3+6) = 66.67%		7/ (7+2) = 77.78%	

The consensus QSAR-2C model demonstrated superior performance in identifying true inactive molecules, with an accuracy of 81.82%, sensitivity of 82.61%, and specificity of 77.78%, compared to the top-ranked QSAR-2C model, rp_32, which achieved an accuracy of 83.64%, sensitivity of 86.96%, and specificity of 66.67%. To enrich the hit dataset with probable actives and eliminate as many inactive molecules as possible, we employed both QSAR-2C models to predict inactive molecules. Of the 404 hit compounds, 362 and 357 molecules were predicted as actives based on the top-ranked rp_32 and consensus QSAR-2C models, respectively. Compounds predicted as actives by both models were selected for further filtering. In total, 352 compounds were subjected to filtering using QSAR-3C models.The statistical details for the top ten QSAR-3C models are provided in **Table 3**.

Table 3. Accuracy rate per category for the top-scored ten models for QSAR-3C

Code	Score	Train (inactive)	Train (weak)	Train (active)	Test (inactive)	Test (weak)	Test (active)
bayes_desc_13	0.6892	0.6279	0.5357	0.8696	0.6000	0.6923	0.8636
rp_36	0.6878	0.7209	0.7321	0.9348	0.4000	0.8462	0.9545
bayes_desc_33	0.6850	0.5581	0.6071	0.8696	0.7000	0.5385	0.8636
bayes_desc_42	0.6845	0.6279	0.5714	0.8370	0.6000	0.5385	0.9545
bayes_desc_45	0.6761	0.5349	0.6250	0.8370	0.7000	0.5385	0.8636
bayes_desc_5	0.6652	0.6047	0.5179	0.8587	0.7000	0.4615	0.8636
bayes_desc_21	0.6500	0.6977	0.4107	0.8478	0.5000	0.6923	0.8182
bayes_desc_2	0.6510	0.5349	0.5000	0.8261	0.7000	0.6923	0.8182
bayes_desc_7	0.6508	0.6512	0.4643	0.8478	0.6000	0.5385	0.8182
rp_40	0.6489	0.7907	0.7500	0.9565	0.5000	0.7692	0.9091

Table 4. Confusion matrix of the external validation set using the top-ranked (bayes_desc_13) and the consensus QSAR-3C models

	Top-ranked QSAR-3C Model			Consensus QSAR-3C Model		
	Active predicted	Weak predicted	Inactive predicted	Active predicted	Weak predicted	Inactive predicted
Active (37)	30	6	1	30	6	1
Weak (25)	4	10	11	5	11	9
Inactive (16)	5	1	10	5	2	9
Accuracy	(30+10+10)/78 = 64.10%			(30+11+9)/ 78 = 64.10%		
Sensitivity*	30 / (30+6+1) = 81.08%			30 / (30+6+1) = 81.08%		
Specificity**	10/ (5+1+10) = 62.50%			9/ (5+2+9) =56.25 %		

* Only active molecules are considered.
** Only inactive molecules were considered.

Overall, the prediction accuracy rate of QSAR-3C models was lower than QSAR-2C models. The highest-scoring QSAR-3C model achieved a score of approximately 0.69 out of 1.00. This result wasn't surprising, as previous literature has shown that distinguishing weak binders from active and inactive molecules is generally more challenging than distinguishing actives from inactive molecules [82, 83]. Additionally, the QSAR code of the AutoQSAR module was not optimized for more than two classes. When we looked at the confusion matrix of the external validation set using the top-ranked QSAR-3C model (bayes_desc_13) and the consensus QSAR-3C model, we found a similar predictivity rate for both models (Table 4). As a result, we only selected the compounds predicted as actives in both QSAR-3C models as hit molecules for further analysis. For categorical response endpoints, it is essential to have at least five compounds in each class, and the distribution of compounds across the classes should be relatively equal. If one class contains significantly more compounds than the others, developing a predictive model that performs well becomes challenging. Since the QSAR models in this work have been generated using the known Thumb Site 2 inhibitors collected from literature and different laboratories —often seen in high-throughput screening (HTS) datasets—compounds bearing a similar scaffold for each class were not distributed equally. It is important to avoid data heterogeneity, meaning we should not mix data from different species or experimental protocols. This is particularly important if chiral compounds or racemates are used in the dataset. One should exclude all chiral compounds or the activity data from racemic mixtures. However, the available activity data tested against HCV NS5B thumb Site2 was limited to a few hundred.

When excluding all chiral compounds is not feasible, one can consider including only one of the forms. However, in order to retain all available data points and explore the potential for creating more global QSAR models for prefiltering purposes—rather than for lead optimization—we decided to include all available compounds in our analysis. We selected the form that acquired the highest binding affinity during docking and binding free energy calculations. However, the imbalanced number of compounds between different classes of compounds, in terms of chemistry and activity, might have caused lower specificity values, as seen in the QSAR-2C model (66.67%) and QSAR-3C model (62.50%), and a lower accuracy rate for QSAR-3C model (64.10%).

In addition, when building a categorical model based on activity cutoffs, it is recommended to discard compounds that fall in some gray area. In accordance, we discarded the ligands having an activity of 2 µM < weak binders ≤ 20 µM in QSAR-2C model generation, and these ligands were accepted as weak binders. Since the specificity of the QSAR-2C model was low, and we aimed to enrich the datasets with active molecules as much as possible, we also wanted to assess the power of a QSAR model where weak binders were included. Since the QSAR-3C model's sensitivity was 81.08%, we also included this model to our pre-filtering workflow. Thus, compounds predicted as actives by both QSAR-2C and QSAR-3C models could pass for further evaluation. In total, 244 compounds were predicted as active and further evaluated using QSAR-N models.

Statistical details for the top 10 ranked models obtained during the generation of the QSAR_N model are presented in Table 5. All models had a coefficient of determination (R2) above 0.70 for the training set and above 0.72 for the test set (Q [2]). Generally, for a QSAR model to be predictable, it is required that (R2 > 0.60 and Q [2] > 0.50). Furthermore, all the top-ranked models were based on the KPLS model. This is not surprising as the KPLS method performs well when there is a significant difference between the number of independent variables and the number of samples, which is typical in most drug discovery studies. The method has been demonstrated to have better correlation and prediction power in the literature, making it a valuable QSAR tool in various drug discovery projects [84]. The top-ranked model, kpls_radial_21, was built using KPLS fitting with linear fingerprints using the 21st split of the learning set into a training set and a test set. The scatter plot depicting the performance of the kpls_radial_21 model in predicting experimental binding affinity for the learning set and the external validation set is shown in Figure 3.

This plot shows that the top-ranked QSAR model predictions reproduce experimental binding affinities with an R2 of 0.55. Compounds with a domain alert are colored in green. These compounds fall outside the applicability domain of training set compounds where predictions aren't expected to be accurate. One compound out of 74 was flagged as an outlier. Visual inspection showed that it was part of the same congeneric series as the training set and was not an outlier. We also assessed the predictive ability of the consensus QSAR-N model (y = 0.57x + 2.48, R2 = 0.58) using external validation set compounds. The predicted pIC50 values of the 244 hits using the top-ranked QSAR model ranged from 4.497 to 6.643, while the predicted values using the consensus QSAR model ranged from 4.996 to 6.495.

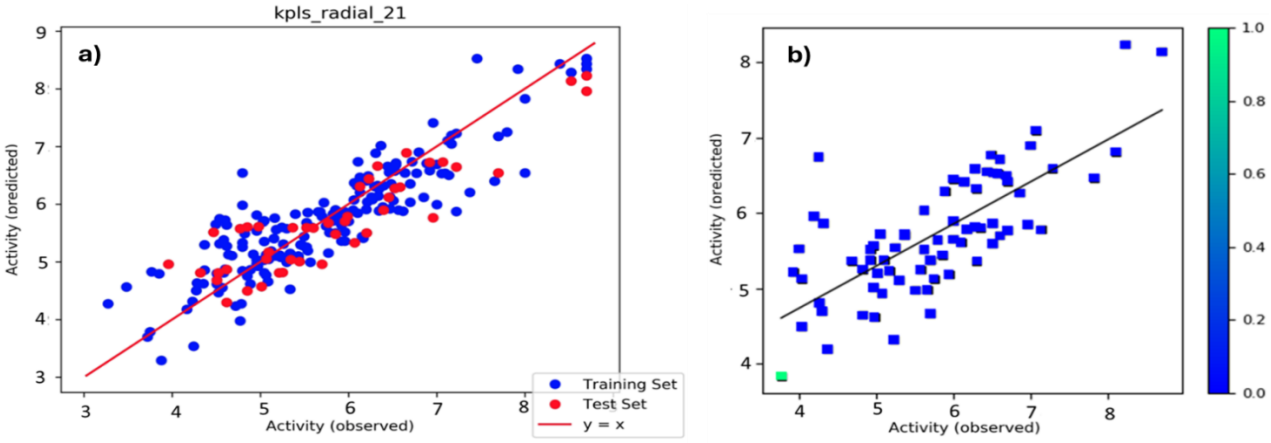

Figure 3. Comparison of experimental and predicted HCV NS5B polymerase inhibitory activity (pIC_{50}) of the learning dataset (a) and external validation set (b; y = 0.56x + 2.50 (R^2 = 0.55)) using kpls_radial_21 QSAR-N model.

Table 5. The predictive power of the first ten top-scored QSAR-N models

Model Code	Score	S.D.	R^2	RMSE	Q [2]	Q [2] MW
kpls_radial_21	0.7682	0.5125	0.7831	0.5221	0.7907	0.0213
kpls_molprint2D_21	0.7646	0.5340	0.7659	0.5331	0.7819	0.0213
kpls_dendritic_21	0.7532	0.4432	0.8378	0.5034	0.8055	0.0213
kpls_linear_21	0.7438	0.4303	0.8471	0.5048	0.8044	0.0213
kpls_molprint2D_15	0.7388	0.5531	0.7533	0.5555	0.7429	0.0674
kpls_molprint2D_20	0.7189	0.5390	0.7635	0.5632	0.7444	0.0227
kpls_linear_32	0.7156	0.5566	0.7431	0.5745	0.7452	-0.0478
kpls_radial_6	0.7030	0.5045	0.7983	0.5569	0.7214	-0.0574
kpls_linear_20	0.6981	0.5679	0.7360	0.5884	0.7211	0.0227
kpls_dendritic_48	0.6961	0.4546	0.8355	0.5432	0.7407	0.0759

2.5. Evaluation of Drug-likeness and ADME Properties

We predicted physiochemically significant descriptors and pharmaceutically relevant (ADME) properties of hit molecules by using the QikProp program (Schrödinger Release 2022-1: QikProp, Schrödinger, LLC, New York, NY, 2022). Then, the drug-likeness of the 244 hit compounds was assessed by considering Lipinski's rule of 5 (Pfizer filter) [85], Ghose [86], Veber (GSK) [87], Egan (Pharmacia) [88], and Muegge (Bayer) [89] filters. When assessing compounds that adhere to Lipinski's Rule of 5 for oral drug activity, the molecular weight (MW) should be ≤ 500 g/mol, the lipophilicity coefficient (octanol/water partition coefficient) Moriguchi Log P (MLogP) value should be MLogP ≤ 4.15, the number of hydrogen bond acceptors (HBA) should be ≤ 10, and the number of hydrogen bond donors (HBD) should be ≤ 5. According to Veber, the total number of rotatable bonds (RB) should be ≤ 10, and the total polar surface area (TPSA) should be ≤ 140 Å2. According to Ghose's rules, a drug candidate compound should have lipophilicity in the range of -0.4 ≤ WLOGP ≤ -5.6, a molar refraction value in the range of 40 ≤ MR ≤ 130, a MW in the range of 160-480 daltons, and a total number of atoms in the range of 20-70. According to Egan's filter, the compound's lipophilicity should be lower than WLOGP ≤ 5.88, and the TPSA should be ≤ 131.6 Å2. According to Muegge's rules, a drug-like compound should have a MW of 200-600 Da, lipophilicity in the range of -2 ≤ XLOGP ≤ -5, a TPSA ≤ 150 Å2, the number of rings should be ≤ 7, the number of carbon atoms should be > 4, the number of heteroatoms should be > 1, the RB should be ≤ 15, the HBA should be ≤ 10, and the HBD should be ≤ 5. We screened compounds based on the following physicochemical parameters: MW ≤ 600 g/mol, HBA ≤ 10, HBD ≤ 5, RB ≤ 15, TPSA ≤ 150 Å2, total number of atoms ≤ 70, MR ≤ 130, and logP o/w ≤ 6.5. Out of 244 compounds, 48 met these criteria and were considered as hit compounds.

2.6. Clustering of Hit Molecules and Molecular Dynamics (MD) Simulations

The Canvas Similarity and Clustering tool (Schrödinger Release 2022-1: Canvas, Schrödinger, LLC, New York, NY, 2022) was used to cluster the selected 48 molecules into four representative groups. The following settings were used for fingerprints: 64-bit precision, MolPrint2D fingerprint type, atom typing scheme, Daylight invariant atom types, and bonds distinguished by bond order. Tanimoto was used as the similarity metric, and the average linkage method was chosen for clustering. Next, MD simulations were performed for the ligand representing the structure nearest to the centroid in each cluster group. Molecular dynamics offers a comprehensive understanding of the interactions between proteins and ligands, as suggested by molecular docking programs. It also helps assess the stability of the proposed binding mode of the ligands with their target biomolecules. Additionally, MD simulations uncover the conformational changes the protein-ligand complex may undergo over time. In order to assess the average change in displacement of a selection of atoms for a particular frame with respect to a reference frame, the Root Mean Square Deviation (RMSD) is calculated. RMSD analysis involved aligning the generated frames with reference frames, using backbone atoms for the protein and heavy atoms for the ligand RMSD is calculated for all frames in the trajectory (Figure 4).

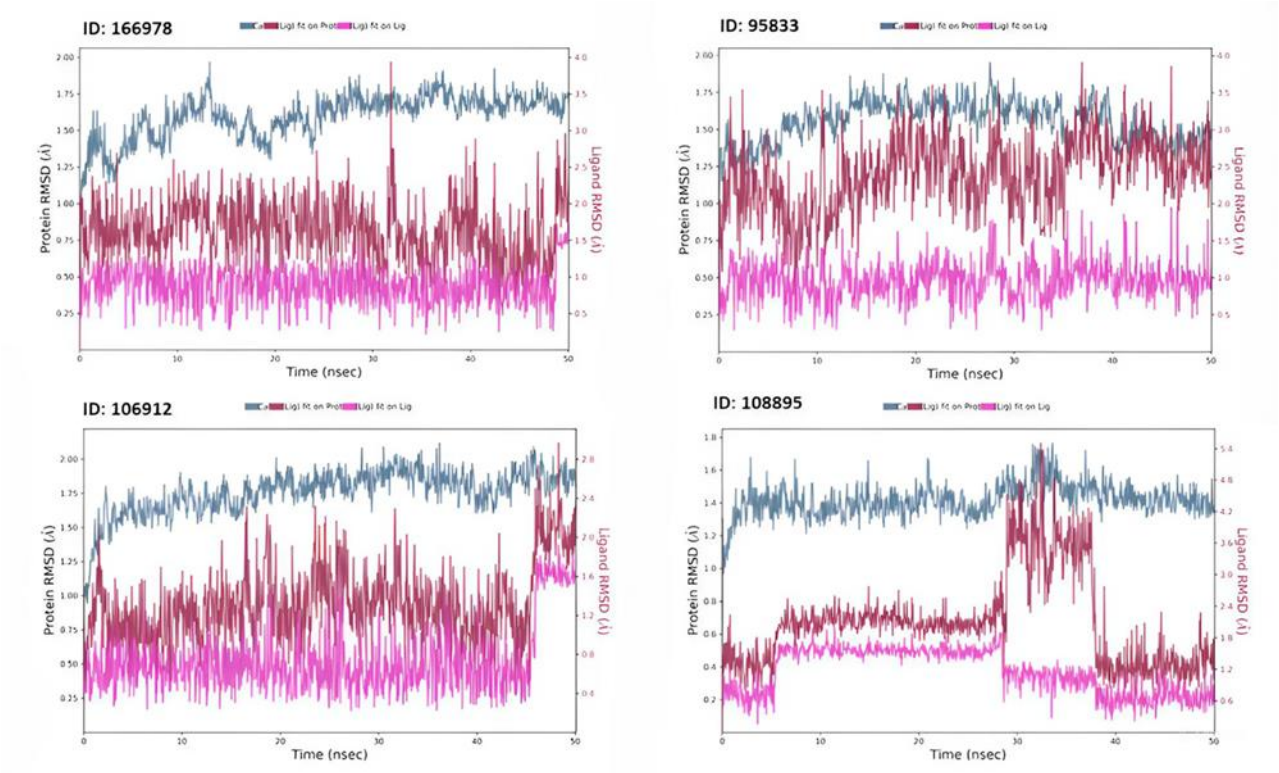

Figure 4. MD simulation analysis of HCV NS5B protein and ligand complexes. Protein Cα RMSD values are depicted on the left y-axis using pale blue lines. 'Lig fit on Prot' RMSD values are depicted on the right Y-axis using red lines. 'Lig fit on Lig' RMSD values are depicted on the right Y-axis using pink lines. Compound ID numbers are shown in the upper left corner of each RMSD plot.

RMSD analysis of the protein provides insights into the fluctuations in its structural conformation during the simulation. Moreover, we calculated the Root Mean Square Fluctuation (RMSF) to characterize local changes along the protein chain (Figure S3, Figure S8, Figure S13, Figure S18). On these plots, peaks indicate areas of the protein that fluctuate the most during the simulation. As expected, the proteins' tails (N- and C-terminal) and the loop regions fluctuated the most. During MD simulations, the secondary structure elements like alpha helices and beta strands were usually more rigid than the unstructured part of the proteins during MD simulations (Figure S4-5, Figure S9-10, Figure S14-15, Figure S19-20). The RMSD values for the protein Cα atoms (pale blue line, left X-axis) fluctuated between 1-2 Å during MD simulations, showing that the simulation has equilibrated well and that the protein adopted some thermal average structure. Moreover, we performed ligand-protein contact analysis. For each interacted residue, we assessed four types of interactions: hydrogen bonds, hydrophobic, ionic, and water bridges, and plotted the fraction and a timeline representation of each of these interactions (Figure S6-7, Figure S11-12, Figure S16-17, Figure S21-22). Protein-ligand interactions that occurred over 30% of the simulation time are shown in Figure 5.

Cluster-1 representative molecule 166978. The RMSD (Root Mean Square Deviation) values for the 'Lig fit Prot' (red line, right Y-axis) fluctuated between approximately 1-3 Å during the MD simulation. This indicates that the ligand remained stable in relation to the protein and its binding pocket. The RMSD of the ligand was calculated by aligning the protein-ligand complex with the protein backbone of the reference, and then determining the RMSD of the ligand's heavy atoms. Moreover, the 'Lig fit Lig' RMSD values (pink line, right Y-axis) of the ligand were also measured to determine the internal fluctuations of the ligand atoms. For this purpose, the ligand was aligned on its reference conformation, and RMSD was measured. 'Lig fit Lig' RMSD values of 166978 fluctuated approximately between 0.5-1 Å during the MD simulation except for the last few ns where fluctuations reached 1.5 Å. However, visual analysis showed that the ligand was still bound to the binding pocket. Cluster-1 ligand 166978 showed hydrophobic contacts with residues Leu419, Tyr477, Leu497, and Trp528. Stable H-bond interactions were observed between the ligand carbonyl

group and the backbone N-H of the residue Tyr477 (60 % simulation time of the MD trajectory) and ligand OH group and the backbone N-H of the residue Leu497 (63 % simulation time of the MD trajectory).

Cluster-2 representative molecule 95833. The 'Lig fit Prot' RMSD values and the 'Lig fit Lig' RMSD values fluctuated approximately 1.5-3.5 Å and 0.5-1.5 Å, respectively, during the MD simulation, indicating that the ligand remained stable in the binding pocket. Ligand 95833 showed hydrophobic contacts with residues Leu419, Tyr477, and Trp528. Stable H-bond interactions were observed between the carboxylic acid carbonyl group and the oxygen with the backbone N-H of the residue Tyr477 for 50% and 39%, respectively, during the simulation time of the MD trajectory. Interestingly, bridging water molecule H-bond interactions occurred between Tyr477 and the ligand carbonyl group during MD simulation for 33% of the trajectory time. Similarly, the backbone N-H of the residue Ser476 made H-bond interactions with the carboxylic acid carbonyl group and the oxygen for 39% and 60%, respectively, during the simulation time of the MD trajectory.

Cluster-3 representative molecule, 106912. The 'Lig fit Prot' RMSD values and the 'Lig fit Lig' RMSD values fluctuated between approximately 0.8-1.6 Å and 0.4-1.2 Å, respectively, during the first 45 ns of the MD simulation, indicating that the ligand was highly stable within the binding pocket. However, in the last ~5 ns, there was an approximately 0.8 Å RMSD jump for both 'Lig fit Prot' RMSD values and the 'Lig fit Lig' RMSD values. Visual analysis revealed that the Arg501 sidechain adopted a different conformation, and the 3,4-dioxocyclobut-1-en-1-olate moiety of the ligand was pulled over towards the sidechain to maintain the H-bond and ionic interactions with this residue and then returned to the original orientation, which is further explained later. Additionally, 106912 formed multiple hydrogen bond interactions during the simulation. One notable interaction involved the ligand's hydroxyl group, which formed hydrogen bonds with the sidechain -NH2+ of the Arginine residue and the backbone carbonyl group of Leucine 474 for 93% and 65% of the MD trajectory, respectively. Furthermore, water-bridging interactions were observed between the ligand's carbonyl group and the backbone -NH of Ser476 and Tyr477 for 50% and 38% of the simulation time, respectively. Persistent interactions were also noted between the sidechain -NH2+ groups of Arginine 501 residue and the O- of the 3,4-dioxocyclobut-1-en-1-olate moiety of the ligand, spanning 105% of the MD simulation. It is worth noting that interactions lasting longer than 100% of the simulation time are possible due to the Arginine side chain having four H-bond donors that can all hydrogen bond to a single H-bond acceptor. Additionally, the Lysine -NH3+ group formed hydrogen bond interactions with the O- of the 3,4-dioxocyclobut-1-en-1-olate moiety of the ligand for 31% of the trajectory time during the MD simulation.

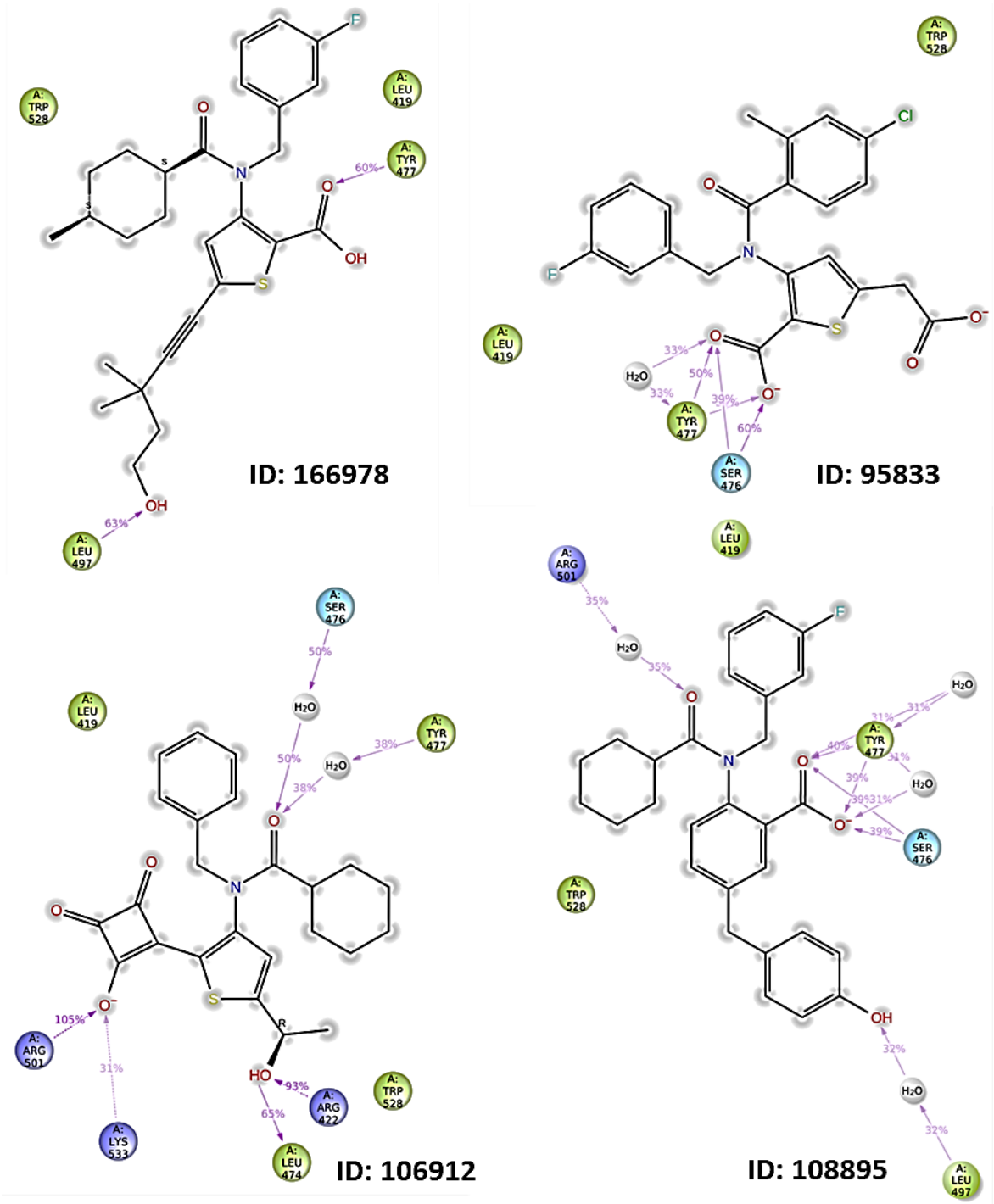

Figure 5. HCV NS5B protein and ligand interactions exist more than 30% of the time during MD simulation. Compound ID numbers for each ligand are shown in the lower right corner. Hydrophobic (green), polar (water blue), and charged (purple) residues are shown as spheres. Water molecules are shown as grey spheres. Solvent exposure is illustrated as grey spheres with transparent centers.

Cluster-4 representative molecule, 108895. During the MD simulation, the 'Lig fit Prot' RMSD values and the 'Lig fit Lig' RMSD values fluctuated approximately between 1.2-5.4 Å and 0.6-1.8 Å, respectively. Visual analysis revealed that the ligand remained in the binding pocket during the MD simulation. Significant H-bond interactions were observed between the ligand, the carboxylic acid carbonyl group, and the oxygen with the backbone -NH of Ser476 and Tyr477 residues for 39% and ~40%, respectively, during the simulation time of the MD trajectory. Interestingly, bridging water molecule H-bond interactions occurred between

Tyr477 and the ligand carboxylic acid group during MD simulation for 31% of the trajectory time. There were 2 other important water-bridging interactions between the ligand hydroxyl group and the backbone -NH of Leu497 residue (32% simulation time of the MD trajectory), as well as the ligand amide carbonyl group and the sidechain -NH2+ of Arg501 residue (35% simulation time of the MD trajectory). The slight RMSD jump observed for the ligand between ~5.5-28 ns resulted from the fluorophenyl and hydroxyphenyl moieties obtaining different conformations (**Figure S12**). Then, in the next 10 ns, the ligand lost the stable H-bond interactions between its carboxylic acid functional group and Ser476 and Tyr477 residues, resulting in the 'Lig fit Prot' RMSD values jumping from ~2Å to 5.4Å. Nevertheless, the last 12ns ligand adopts its initial binding mode and regains these significant interactions again. This indicates that the simulation has equilibrated, and that the protein-ligand complex adopted some thermal average structure.
In summary, each representative ligand from the four clusters showed stable interactions with important residues, Arg422, Leu474, Ser476, Tyr477, Leu497, and Arg501 in the thumb site 2 pocket of NS5B. These results suggest that they could potentially function as novel allosteric inhibitors of HCV NS5B.

The NS5B is an essential protein for the replication of the RNA genome of HCV. The absence of a comparable mammalian counterpart of NS5B polymerase and the enzyme's confirmed druggable allosteric binding sites makes it an attractive therapeutic target for HCV infections. Sofosbuvir [90] and Dasabuvir [13, 14, 12] are two NS5B inhibitors approved by the FDA. They have varying degrees of sustained virological response for different genotypes and subtypes. Sofosbuvir (GS-7977) is a nucleotide analog and prodrug. Once inside liver cells, it is converted into its active uridine triphosphate form and works by stopping viral genome replication. On the other hand, Dasabuvir is a non-nucleoside polymerase inhibitor that binds to the palm site. Due to the risk of resistance development, it is combined with other agents. The incidence of new acute HCV infections is increasing annually, and the mortality rate for chronic HCV infections is high in several countries [4, 5, 91]. The high mutation rate in the HCV viral strains and the immune response to HCV are believed to be the main barriers to effective anti-HCV therapy, and there is an urgent need for affordable, broad-spectrum, and direct-acting antiviral drugs [92-94]. Consequently, there is a pressing need for the development of new treatments.

In this study, we used molecular modeling techniques to identify novel HCV NS5B inhibitors that target the thumb site 2 pocket. First, we generated a combinatorial library of 184,214 molecules using the privileged substructures (Figure S1) of known NS5B inhibitors from different series of compounds. Next, a structure-based virtual screening search was applied to filter over compounds with increasing scoring precision, resulting in 555 compounds. Application of the Prime MM-GBSA rescoring enabled us to filter these molecules further, resulting in 404 compounds. In addition, classification-based QSAR models were generated to enrich the dataset with active molecules. The top-ranked QSAR-2C model (rp_32) and the consensus QSAR-2C model demonstrated excellent performance in accurately identifying active molecules, with an accuracy rate of over 80%. Therefore, only compounds predicted as active molecules by both models were chosen for additional filtering. Out of 404 molecules, 352 of them were predicted to be active. In the next step, the top-ranked QSAR-3C model (bayes_desc_13) and the consensus QSAR-3C models, which had a predictive accuracy of over 64% for distinguishing active molecules from weak and inactive ones, were used for filtering. In total, 244 compounds were predicted to be active. We calculated physiochemically significant descriptors and pharmaceutically relevant properties of these hits. Consequently, 48 molecules were considered druglike molecules. Fingerprint similarity clustering of these molecules revealed four novel scaffolds as potential HCV NS5B thumb site 2 inhibitors. MD simulations confirmed the binding stability of these compounds within the thumb site 2 of HCV NS5B protein. They revealed the presence of key interactions previously identified that play a critical role in potency and protein function [38-43]. Compounds that produce a stable protein-ligand complex with a high pIC50 value predicted using the QSAR-N model could be considered for future synthetic work as a follow-up project.

3. CONCLUSION

Based on this study, three main objectives have been attained: 1) developing in silico combinatorial libraries by integrating known bioactive moieties and then enumerating them to novel molecules can be an easy alternative in the search for new intellectual properties for HCV NS5B inhibitors, 2) balancing the number of molecules for different chemical and activity classes with further literature search and inclusion of physiochemically similar decoys can be considered to improve the predictivity and specificity rates of

categorical QSAR models, and 3) introducing another layer of post-filtering method for instance, field-based 3D QSAR models based on fields, such as electrostatic, hydrophobic, or steric fields, may overcome or mitigate the limitations of the categorical QSAR models. Our ongoing efforts of designing more effective and selective antiviral drugs against HCV targeting NS5B thumb site 2 inhibitors using known bioactive molecules and replacing or enumerating different functionalities resulted in the recent discovery of novel phenylalanine derivatives with single-digit micromolar EC50 values against HCV gt1b replicon . The suitability of the identified hits as starting points for lead discovery against HCV NS5B will depend on testing these molecules through traditional wet lab experiments, which require further investigation. We plan to further explore the effects of these compounds on HCV and optimize the hits or generate new series with potential anti-HCV activities.

4. MATERIALS AND METHODS

4.1. Dataset Preparation

A validated dataset of known inhibitors of HCV NS5B polymerase inhibitors with activities on genotype 1b (gt1b) was collected from the literature. The structure of the molecules was drawn using the 2D Sketcher module implemented in the Maestro (Schrödinger Release 2022-1: Maestro, Schrödinger, LLC, New York, NY, 2022) interface of the Schrödinger platform. In total, 314 compounds were collected [39, 44-67] and prepared using LigPrep within the Schrödinger software for in silico experiments. The options were set to generate all possible ionization states at pH 7.0. We kept the specified chirality from the input file. Low-energy 3D structures of the compounds were obtained using the Optimized Potentials for Liquid Simulations (OPLS) force field OPLS3e [68]. All other parameters for ligand preparation were kept as default. If the compounds' exact HCV NS5B inhibition activities (IC50, µM) were known, we converted them into the corresponding pIC50 and -log10(IC50) values.

4.2. Combinatorial Library Generation

We generated a virtual combinatorial library where privileged substructures (**Table S1**) of known NS5B inhibitors from different series of compounds, including phenylalanine derivatives, thiophene-2-carboxylic acid derivatives, and anthranilic acid derivatives. We generated the library using R Group Creator and Custom R Group Enumeration tools implemented in Schrödinger software within Maestro. We first sketched each R group using the 2D Sketcher panel implemented in Maestro and used R-group libraries to enumerate the core-containing molecule. Compounds were then subjected to the LigPrep module (Schrödinger Release 2022-1: LigPrep, Schrödinger, LLC, New York, NY, 2022) for ligand preparation. The settings were configured to create probable ionization states at pH 7.0, maintain the specified chirality from the input file, and generate all possible isomers of other chiral centers. In total, 184,214 molecules were obtained after the ligand generation and preparation.

4.3. Preparation of protein-inhibitor complexes

The preparation of protein-inhibitor complexes has already been reported before [69]. In short, crystal structures of HCV NS5B in complex with thumb site 2 inhibitors were collected from PDB by April 2020 and prepared using the Protein Preparation Wizard module (Schrödinger Release 2022-1: Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2022; Impact, Schrödinger, LLC, New York, NY; Prime, Schrödinger, LLC, New York, NY, 2022). A total of 47 X-ray crystal structures with various chains were processed. After splitting the chains during protein preparation, we obtained 81 protein-ligand complexes. A KNIME [70] workflow (**Figure S2-4**) was generated to cluster the thumb site 2 pockets of the HCV NS5B-ligand complexes using the SiteMap module (Schrödinger Release 2022-1: SiteMap, Schrödinger, LLC, New York, NY, 2022). In total, 8 clusters were generated, and the protein structure that represents the cluster's center was chosen for the subsequent in silico experiments.

4.4. Ensemble Docking and Rescoring of Known HCV NS5B Thumb Site 2 Inhibitors

An automated KNIME workflow was generated to perform the Glide SP (Standard Precision) [71, 72] Docking calculations and post-processing of docking poses with the Prime MM-GBSA (Schrödinger Release 2022-1: Prime, Schrödinger, LLC, New York, NY, 2022) method (**Figures S5-8**). Ensemble docking was performed using the eight representative protein structures retrieved from clustering of all known complexes of thumb site 2 inhibitors with HCV NS5B polymerase deposited in PDB. The top-ranked docking

pose produced for each ligand and each protein structure was used for binding free energy (BFE) calculations. Thus, in total, eight poses were subjected to Prime MM-GBSA calculations.

4.5. Ensemble Docking and Rescoring of Combinatorial Library

We also generated a KNIME workflow for ensemble docking combinatorial library compounds into the eight representative protein structures of HCV NS5B with the Glide HTVS (High throughput Virtual Screening) precision option. We selected only the top 10% of the ligands ranked based on their HTVS docking score (**Figure S9**). Thus, only 18,323 molecules were subjected to Glide docking using the SP scoring function, which has a much more exhaustive conformational sampling than HTVS docking (**Figure S10**). During this docking run, the top-ranked docking pose was saved again for each ligand and each protein structure. Thus, eight poses for each ligand were collected for further BFE calculations using the Prime MM-GBSA method for the known HCV NS5B inhibitors (**Figure S11**).

4.6. Generation of Numeric- and Classification-based QSAR Models

We used 314 known thumb site 2 inhibitors of HCV NS5B to generate numeric- and classification-based QSAR models. The ligand dataset was divided into three activity classes. Ligands having an inhibitory activity of ≤ 2 μM were annotated as actives (151 molecules). Ligands having an activity of 2 μM < weak binders ≤ 20 μM were accepted as weak binders (94 molecules). Finally, ligands with an inhibitory activity of > 20 μM were grouped as inactives (69 molecules).

First, we generated a two-categorical QSAR model (QSAR Model-2C) based on two activity classes: actives and inactives. The ligand dataset (220 compounds) was randomly split into two sets: a modeling set consisting of %75 (165 compounds) of the total dataset and an external validation set of 55 compounds (%25 of the whole dataset). During model building, the modeling set was randomly split into a training set (%80 of the entire modeling set, 132 molecules) and a test set (%20 of the whole modeling set, 33 molecules). This process was repeated 50 times to develop 50 different models per supervised learning techniques, Naïve Bayes classification, and ensemble recursive partitioning (RP, rp). Only the top-scored ten models were saved for further evaluation.

Furthermore, we generated a three-categorical QSAR model (QSAR-3C) based on three activity classes: actives, weak binders, and inactives. The ligand dataset (314 compounds) was randomly split into two sets: a modeling set consisting of %75 (236 compounds) of the total dataset and an external validation set of 78 compounds (%25 of the whole dataset). During model building, the modeling set was randomly split into again a training set (%80 of the entire modeling set, 188 molecules) and a test set (%20 of the whole modeling set, 48 molecules). This step was repeated 50 times to generate 50 different models for both Bayes and RP techniques used during model generation. For further assessment, only the top-scored ten models were kept for each method used in model generation.

We aimed to use these models to filter and rank the hit compounds selected from the hierarchical filtering of the combinatorial library based on Glide docking and Prime MM-GBSA calculations. To generate QSAR models, we used the AutoQSAR module implemented in Schrödinger software. In the AutoQSAR model generation step, several 2D descriptors were automatically generated, including molecular and topological descriptors and feature counts. As several of these descriptors are highly correlated, only the most informative subset for all the correlation coefficients below a specified threshold was selected. The maximum allowed correlation between any pair of descriptors was set to 0.80 by default. Descriptors were removed before the model-building stage if more than 90% of the ligands in the dataset had the same value for that particular property. Moreover, 2D fingerprints for Naïve Bayes classification (Bayes, bayes) and kernel-based partial least-squares regression (KPLS) models were generated. The 10,000 most informative bits for each fingerprint type (linear, radial, dendritic, and molprint2D) were retained.

In addition, we built a numeric QSAR model (QSAR Model-N) using the ligands with exact pIC50 values. In AutoQSAR, the generation of numeric QSAR models was done using four different techniques: 1) multiple linear regression (MLR) [73], 2) partial least-squares regression (PLS) [74], 3) KPLS [74, 75], and 4) principal components regression (PCR) [76]. The ligand dataset was randomly split into two sets: a modeling set consisting of %75 (221 compounds) of the total dataset and an external validation set of 74 compounds (%25 of the whole dataset). During model building, the modeling set was randomly split into a training set (%80 of the entire modeling set, 176 molecules) and a test set (%20 of the entire modeling set, 45 molecules). Like the categorical QSAR model generation steps, this process was repeated 50 times to generate 50 models

per supervised learning techniques used in the model generation step. Only the top-scored ten models were kept for further evaluation.

4.7. Molecular Dynamics (MD) Simulation

The preparation of protein-inhibitor complexes for MD simulation has already been reported before [77]. The MD simulations were conducted using the Desmond module (Schrödinger Release 2022-1: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2022. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2022). The ligand-protein complexes were preprocessed before MD to assign the correct bond orders, adjust the formal charges, and cap the termini. OPLS4 [78] force field parameters were employed in all simulations. The system was solvated using the simple point-charge (SPC) water model and an orthorhombic water boundary box with a margin of 10 Å. The system was neutralized by counterions, followed by adding 0.15 M NaCl salt to the system, and the iso-osmotic state was maintained during the simulation. Before the production MD run, the system was relaxed using the default relaxation procedure for the NPT ensemble. First, the system was relaxed in the NVT ensemble with Brownian dynamics at a temperature of 10 K. Then, it was simulated in the NPT ensemble using a Langevin thermostat and a Langevin barostat. During relaxation, heavy atoms of the solute were restrained, and the temperature and the pressure were kept constant at 10 K and 1 bar, respectively. Next, the system's temperature was increased to 300 K, and two consecutive simulations were performed, first keeping the restraints on the heavy atoms of the solute and then switching them off. The production MD simulation was run for 50 ns with the NPT (isothermal-isobaric) ensemble. The pressure was kept constant at 1.01325 bar using the Martyna-Tobias-Klein barostat, and the constant temperature was maintained at 300 K using the Nosé- Hoover thermostat. The trajectories were saved at 50 ps intervals, producing 1000 frames, which were analyzed using the Simulation Interaction Diagram tool implemented in the Schrödinger program.

Acknowledgements: This work was supported by The Scientific and Technological Research Council of Türkiye (Project Number: 1109B321801603). Authors would like to thank E. D. Dincel and O. Soylu Eter (Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Istanbul University, 34134, Istanbul, Turkey) for their data collection support.
Author contributions: Concept – B.K.M.; Design – B.K.M.; Data Collection and/or Processing – B.K.M., M.M.A.; Analysis and/or Interpretation – B.K.M., M.M.A.; Literature Search – B.K.M., M.M.A., M.H.G.; Writing – B.K.M.; Critical Reviews – B.K.M., M.M.A.
Conflict of interest statement: The authors declare no conflict of interest.

REFERENCES

[1] Li HC, Lo SY. Hepatitis C virus: Virology, diagnosis and treatment. World J Hepatol. 2015;7(10):1377-1389. https://doi.org/10.4254/wjh.v7.i10.1377.
[2] Zein NN. Clinical significance of hepatitis C virus genotypes. Clin Microbiol Rev. 2000;13(2):223-235. https://doi.org/10.1128/cmr.13.2.223.
[3] Morozov VA, Lagaye S. Hepatitis C virus: Morphogenesis, infection and therapy. World J Hepatol. 2018;10(2):186-212. https://doi.org/10.4254/wjh.v10.i2.186.
[4] Pimenov N, Kostyushev D, Komarova S, Fomicheva A, Urtikov A, Belaia O, Umbetova K, Darvina O, Tsapkova N, Chulanov V. Epidemiology and Genotype Distribution of Hepatitis C Virus in Russia. Pathogens. 2022;11(12):1482. https://doi.org/10.3390/pathogens11121482.
[5] Yang J, Qi JL, Wang XX, Li XH, Jin R, Liu BY, Liu HX, Rao HY. The burden of hepatitis C virus in the world, China, India, and the United States from 1990 to 2019. Front Public Health. 2023;11:1041201. https://doi.org/10.3389/fpubh.2023.1041201.
[6] Gu M, Rice CM. Structures of hepatitis C virus nonstructural proteins required for replicase assembly and function. Curr Opin Virol. 2013;3(2):129-136. https://doi.org/10.1016/j.coviro.2013.03.013.
[7] Moradpour D, Penin F. Hepatitis C virus proteins: from structure to function. Curr Top Microbiol Immunol. 2013;369:113-142. https://doi.org/10.1007/978-3-642-27340-7_5.
[8] Kumar A, Narang RK, Bhatia R. Recent advancements in NS5B inhibitors (2011-2021): Structural insights, SAR studies and clinical status. J Mol Struct. 2023;1293:136272. https://doi.org/10.1016/j.molstruc.2023.136272.
[9] Ganta NM, Gedda G, Rathnakar B, Satyanarayana M, Yamajala B, Ahsan MJ, Jadav SS, Balaraju T. A review on HCV inhibitors: Significance of non-structural polyproteins. Eur J Med Chem. 2019;164:576-601. https://doi.org/10.1016/j.ejmech.2018.12.045.
[10] Membreno FE, Lawitz EJ. The HCV NS5B nucleoside and non-nucleoside inhibitors. Clin Liver Dis. 2011;15(3):611-626. https://doi.org/10.1016/j.cld.2011.05.003.

[11] Abdurakhmanov E, Øie Solbak S, Danielson UH. Biophysical Mode-of-Action and Selectivity Analysis of Allosteric Inhibitors of Hepatitis C Virus (HCV) Polymerase. Viruses. 2017;9(6):151. https://doi.org/10.3390/v9060151.
[12] Trivella JP, Gutierrez J, Martin P. Dasabuvir : a new direct antiviral agent for the treatment of hepatitis C. Expert Opin Pharmacother. 2015;16(4):617-624. https://doi.org/10.1517/14656566.2015.1012493.
[13] Poordad F, Sedghi S, Pockros PJ, Ravendhran N, Reindollar R, Lucey MR, Epstein M, Bank L, Bernstein D, Trinh R, Krishnan P, Polepally AR, Unnebrink K, Martinez M, Nelson DR. Efficacy and safety of ombitasvir/paritaprevir/ritonavir and dasabuvir with low-dose ribavirin in patients with chronic hepatitis C virus genotype 1a infection without cirrhosis. J Viral Hepat. 2019;26(8):1027-1030. https://doi.org/10.1111/jvh.13109.
[14] El Kassas M, Elbaz T, Hafez E, Wifi MN, Esmat G. Discovery and preclinical development of dasabuvir for the treatment of hepatitis C infection. Expert Opin Drug Discov. 2017;12(6):635-642. https://doi.org/10.1080/17460441.2017.1322955.
[15] Mantry PS, Pathak L. Dasabuvir (ABT333) for the treatment of chronic HCV genotype I: a new face of cure, an expert review. Expert Rev Anti Infect Ther. 2016;14(2):157-165. https://doi.org/10.1586/14787210.2016.1120668.
[16] Gentles RG. Discovery of Beclabuvir: A Potent Allosteric Inhibitor of the Hepatitis C Virus Polymerase. HCV: The Journey from Discovery to a Cure. 2019;31:193-228. https://doi.org/10.1007/7355_2018_38.
[17] Bernal LA, Soti V. Hepatitis C Virus: Insights Into Its History, Treatment, Challenges, and Future Directions. Cureus. 2023;15(8):e43924. https://doi.org/10.7759/cureus.43924.
[18] Parlati L, Hollande C, Pol S. Treatment of hepatitis C virus infection. Clin Res Hepatol Gastroenterol. 2021;45(4):101578. https://doi.org/10.1016/j.clinre.2020.11.008.
[19] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res. 2000;28(1):235-242. https://doi.org/10.1093/nar/28.1.235.
[20] Chen YC. Beware of docking! Trends Pharmacol Sci. 2015;36(2):78-95. https://doi.org/10.1016/j.tips.2014.12.001.
[21] Yuriev E, Agostino M, Ramsland PA. Challenges and advances in computational docking: 2009 in review. J Mol Recognit. 2011;24(2):149-164. https://doi.org/10.1002/jmr.1077.
[22] Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. BMC Biol. 2011;9:71. https://doi.org/10.1186/1741-7007-9-71.
[23] Amaro RE, Baudry J, Chodera J, Demir Ö, McCammon JA, Miao Y, Smith JC. Ensemble Docking in Drug Discovery. Biophys J. 2018;114(10):2271-2278. https://doi.org/10.1016/j.bpj.2018.02.038.
[24] Greenidge PA, Kramer C, Mozziconacci JC, Sherman W. Improving docking results via reranking of ensembles of ligand poses in multiple X-ray protein conformations with MM-GBSA. J Chem Inf Model. 2014;54(10):2697-2717. https://doi.org/10.1021/ci5003735.
[25] Slynko I, Scharfe M, Rumpf T, Eib J, Metzger E, Schüle R, Jung M, Sippl W. Virtual Screening of PRK1 Inhibitors: Ensemble Docking, Rescoring Using Binding Free Energy Calculation and QSAR Model Development. J Chem Inf Model. 2014;54(1):138-150. https://doi.org/10.1021/ci400628q.
[26] Amaro RE, Li WW. Emerging methods for ensemble-based virtual screening. Curr Top Med Chem. 2010;10(1):3-13. https://doi.org/10.2174/156802610790232279.
[27] Bolcato G, Cuzzolin A, Bissaro M, Moro S, Sturlese M. Can We Still Trust Docking Results? An Extension of the Applicability of DockBench on PDBbind Database. Int J Mol Sci. 2019;20(14):3558. https://doi.org/10.3390/ijms20143558.
[28] Plewczynski D, Łaźniewski M, Augustyniak R, Ginalski K. Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. J Comput Chem. 2011;32(4):742-755. https://doi.org/10.1002/jcc.21643.
[29] Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS. A critical assessment of docking programs and scoring functions. J Med Chem. 2006;49(20):5912-5931. https://doi.org/10.1021/jm050362n.
[30] Wang E, Sun H, Wang J, Wang Z, Liu H, Zhang JZH, Hou T. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. Chem Rev. 2019;119(16):9478-9508. https://doi.org/10.1021/acs.chemrev.9b00055.
[31] Hayes J, Archontis G, editors. MM-GB(PB)SA Calculations of Protein-Ligand Binding Free Energies. 2012. https://doi.org/10.5772/37107.
[32] Wichapong K, Lawson M, Pianwanit S, Kokpol S, Sippl W. Postprocessing of protein-ligand docking poses using linear response MM-PB/SA: application to Wee1 kinase inhibitors. J Chem Inf Model. 2010;50(9):1574-1588. https://doi.org/10.1021/ci1002153.
[33] Wichapong K, Rohe A, Platzer C, Slynko I, Erdmann F, Schmidt M, Sippl W. Application of docking and QM/MM-GBSA rescoring to screen for novel Myt1 kinase inhibitors. J Chem Inf Model. 2014;54(3):881-893. https://doi.org/10.1021/ci4007326.
[34] Achary PGR. Applications of Quantitative Structure-Activity Relationships (QSAR) based Virtual Screening in Drug Design: A Review. Mini Rev Med Chem. 2020;20(14):1375-1388. https://doi.org/10.2174/1389557520666200429102334.

[35] Neves BJ, Braga RC, Melo-Filho CC, Moreira-Filho JT, Muratov EN, Andrade CH. QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. Front Pharmacol. 2018;9:1275. https://doi.org/10.3389/fphar.2018.01275.

[36] Tropsha A. QSAR in drug discovery. In: Merz JKM, Ringe D, Reynolds CH, editors. Drug Design: Structure- and Ligand-Based Approaches. Cambridge: Cambridge University Press; 2010. p. 151-64.

[37] Bastikar V, Bastikar A, Gupta P. Quantitative structure–activity relationship-based computational approaches. Computational Approaches for Novel Therapeutic and Diagnostic Designing to Mitigate SARS-CoV-2 Infection. 2022:191-205. https://doi.org/10.1016/B978-0-323-91172-6.00001-7.

[38] Wei Y, Li J, Qing J, Huang M, Wu M, Gao F, Li D, Hong Z, Kong L, Huang W, Lin J. Discovery of Novel Hepatitis C Virus NS5B Polymerase Inhibitors by Combining Random Forest, Multiple e-Pharmacophore Modeling and Docking. PLoS One. 2016;11(2):e0148181. https://doi.org/10.1371/journal.pone.0148181.

[39] Lazerwith SE, Lew W, Zhang J, Morganelli P, Liu Q, Canales E, Clarke MO, Doerffler E, Byun D, Mertzman M, Ye H, Chong L, Xu L, Appleby T, Chen X, Fenaux M, Hashash A, Leavitt SA, Mabery E, Matles M, Mwangi JW, Tian Y, Lee Y-J, Zhang J, Zhu C, Murray BP, Watkins WJ. Discovery of GS-9669, a Thumb Site II Non-Nucleoside Inhibitor of NS5B for the Treatment of Genotype 1 Chronic Hepatitis C Infection. J Med Chem. 2014;57(5):1893-1901. https://doi.org/10.1021/jm401420j.

[40] Eltahla AA, Luciani F, White PA, Lloyd AR, Bull RA. Inhibitors of the Hepatitis C Virus Polymerase; Mode of Action and Resistance. Viruses. 2015;7(10):5206-5224. https://doi.org/10.3390/v7102868.

[41] Wang M, Ng KK, Cherney MM, Chan L, Yannopoulos CG, Bedard J, Morin N, Nguyen-Ba N, Alaoui-Ismaili MH, Bethell RC, James MN. Non-nucleoside analogue inhibitors bind to an allosteric site on HCV NS5B polymerase. Crystal structures and mechanism of inhibition. J Biol Chem. 2003;278(11):9489-9495. https://doi.org/10.1074/jbc.m209397200.

[42] Nasr T, Aboshanab AM, Mpekoulis G, Drakopoulos A, Vassilaki N, Zoidis G, Abouzid KAM, Zaghary W. Novel 6-Aminoquinazolinone Derivatives as Potential Cross GT1-4 HCV NS5B Inhibitors. Viruses. 2022;14(12):2767. https://doi.org/10.3390/v14122767.

[43] Elhefnawi M, ElGamacy M, Fares M. Multiple virtual screening approaches for finding new hepatitis C virus RNA-dependent RNA polymerase inhibitors: structure-based screens and molecular dynamics for the pursue of new poly pharmacological inhibitors. BMC Bioinformatics. 2012;13 Suppl 17(Suppl 17):S5. https://doi.org/10.1186/1471-2105-13-s17-s5.

[44] Beaulieu PL, Coulombe R, Duan J, Fazal G, Godbout C, Hucke O, Jakalian A, Joly MA, Lepage O, Llinàs-Brunet M, Naud J, Poirier M, Rioux N, Thavonekham B, Kukolj G, Stammers TA. Structure-based design of novel HCV NS5B thumb pocket 2 allosteric inhibitors with submicromolar gt1 replicon potency: discovery of a quinazolinone chemotype. Bioorg Med Chem Lett. 2013;23(14):4132-4140. https://doi.org/10.1016/j.bmcl.2013.05.037.

[45] LaPlante SR, Forgione P, Boucher C, Coulombe R, Gillard J, Hucke O, Jakalian A, Joly MA, Kukolj G, Lemke C, McCollum R, Titolo S, Beaulieu PL, Stammers T. Enantiomeric atropisomers inhibit HCV polymerase and/or HIV matrix: characterizing hindered bond rotations and target selectivity. J Med Chem. 2014;57(5):1944-1951. https://doi.org/10.1021/jm401202a.

[46] Stammers TA, Coulombe R, Duplessis M, Fazal G, Gagnon A, Garneau M, Goulet S, Jakalian A, LaPlante S, Rancourt J, Thavonekham B, Wernic D, Kukolj G, Beaulieu PL. Anthranilic acid-based Thumb Pocket 2 HCV NS5B polymerase inhibitors with sub-micromolar potency in the cell-based replicon assay. Bioorg Med Chem Lett. 2013;23(24):6879-6885. https://doi.org/10.1016/j.bmcl.2013.09.102.

[47] Chan L, Reddy TJ, Proulx M, Das SK, Pereira O, Wang W, Siddiqui A, Yannopoulos CG, Poisson C, Turcotte N, Drouin A, Alaoui-Ismaili MH, Bethell R, Hamel M, L'Heureux L, Bilimoria D, Nguyen-Ba N. Identification of N,N-disubstituted phenylalanines as a novel class of inhibitors of hepatitis C NS5B polymerase. J Med Chem. 2003;46(8):1283-1285. https://doi.org/10.1021/jm0340400.

[48] Reddy TJ, Chan L, Turcotte N, Proulx M, Pereira OZ, Das SK, Siddiqui A, Wang W, Poisson C, Yannopoulos CG, Bilimoria D, L'Heureux L, Alaoui HM, Nguyen-Ba N. Further SAR studies on novel small molecule inhibitors of the hepatitis C (HCV) NS5B polymerase. Bioorg Med Chem Lett. 2003;13(19):3341-3344. https://doi.org/10.1016/s0960-894x(03)00670-x.

[49] Chan L, Pereira O, Reddy TJ, Das SK, Poisson C, Courchesne M, Proulx M, Siddiqui A, Yannopoulos CG, Nguyen-Ba N, Roy C, Nasturica D, Moinet C, Bethell R, Hamel M, L'Heureux L, David M, Nicolas O, Courtemanche-Asselin P, Brunette S, Bilimoria D, Bédard J. Discovery of thiophene-2-carboxylic acids as potent inhibitors of HCV NS5B polymerase and HCV subgenomic RNA replication. Part 2: tertiary amides. Bioorg Med Chem Lett. 2004;14(3):797-800. https://doi.org/10.1016/j.bmcl.2003.10.068.

[50] Court JJ, Poisson C, Ardzinski A, Bilimoria D, Chan L, Chandupatla K, Chauret N, Collier PN, Das SK, Denis F, Dorsch W, Iyer G, Lauffer D, L'Heureux L, Li P, Luisi BS, Mani N, Nanthakumar S, Nicolas O, Rao BG, Ronkin S, Selliah S, Shawgo RS, Tang Q, Waal ND, Yannopoulos CG, Green J. Discovery of Novel Thiophene-Based, Thumb Pocket 2 Allosteric Inhibitors of the Hepatitis C NS5B Polymerase with Improved Potency and Physicochemical Profiles. J Med Chem. 2016;59(13):6293-6302. https://doi.org/10.1021/acs.jmedchem.6b00541.

[51] Eltahla AA, Tay E, Douglas MW, White PA. Cross-genotypic examination of hepatitis C virus polymerase inhibitors reveals a novel mechanism of action for thumb binders. Antimicrob Agents Chemother. 2014;58(12):7215-7224. https://doi.org/10.1128/aac.03699-14.

[52] Gentile I, Buonomo AR, Zappulo E, Coppola N, Borgia G. GS-9669: a novel non-nucleoside inhibitor of viral polymerase for the treatment of hepatitis C virus infection. Expert Rev Anti Infect Ther. 2014;12(10):1179-1186. https://doi.org/10.1586/14787210.2014.945432.

[53] Jiang M, Zhang EZ, Ardzinski A, Tigges A, Davis A, Sullivan JC, Nelson M, Spanks J, Dorrian J, Nicolas O, Bartels DJ, Rao BG, Rijnbrand R, Kieffer TL. Genotypic and phenotypic analyses of hepatitis C virus variants observed in clinical studies of VX-222, a nonnucleoside NS5B polymerase inhibitor. Antimicrob Agents Chemother. 2014;58(9):5456-5465. https://doi.org/10.1128/aac.03052-14.

[54] Xue W, Jiao P, Liu H, Yao X. Molecular modeling and residue interaction network studies on the mechanism of binding and resistance of the HCV NS5B polymerase mutants to VX-222 and ANA598. Antiviral Res. 2014;104:40-51. https://doi.org/10.1016/j.antiviral.2014.01.006.

[55] Yi G, Deval J, Fan B, Cai H, Soulard C, Ranjith-Kumar CT, Smith DB, Blatt L, Beigelman L, Kao CC. Biochemical study of the comparative inhibition of hepatitis C virus RNA polymerase by VX-222 and filibuvir. Antimicrob Agents Chemother. 2012;56(2):830-837. https://doi.org/10.1128/aac.05438-11.

[56] Antonysamy SS, Aubol B, Blaney J, Browner MF, Giannetti AM, Harris SF, Hébert N, Hendle J, Hopkins S, Jefferson E, Kissinger C, Leveque V, Marciano D, McGee E, Nájera I, Nolan B, Tomimoto M, Torres E, Wright T. Fragment-based discovery of hepatitis C virus NS5b RNA polymerase inhibitors. Bioorg Med Chem Lett. 2008;18(9):2990-2995. https://doi.org/10.1016/j.bmcl.2008.03.056.

[57] Barnes-Seeman D, Boiselle C, Capacci-Daniel C, Chopra R, Hoffmaster K, Jones CT, Kato M, Lin K, Ma S, Pan G, Shu L, Wang J, Whiteman L, Xu M, Zheng R, Fu J. Design and synthesis of lactam-thiophene carboxylic acids as potent hepatitis C virus polymerase inhibitors. Bioorg Med Chem Lett. 2014;24(16):3979-3985. https://doi.org/10.1016/j.bmcl.2014.06.031.

[59] Biswal BK, Wang M, Cherney MM, Chan L, Yannopoulos CG, Bilimoria D, Bedard J, James MN. Non-nucleoside inhibitors binding to hepatitis C virus NS5B polymerase reveal a novel mechanism of inhibition. J Mol Biol. 2006;361(1):33-45. https://doi.org/10.1016/j.jmb.2006.05.074.

[59] Canales E, Carlson JS, Appleby T, Fenaux M, Lee J, Tian Y, Tirunagari N, Wong M, Watkins WJ. Tri-substituted acylhydrazines as tertiary amide bioisosteres: HCV NS5B polymerase inhibitors. Bioorg Med Chem Lett. 2012;22(13):4288-4292. https://doi.org/10.1016/j.bmcl.2012.05.025

[60] Hucke O, Coulombe R, Bonneau P, Bertrand-Laperle M, Brochu C, Gillard J, Joly MA, Landry S, Lepage O, Llinàs-Brunet M, Pesant M, Poirier M, Poirier M, McKercher G, Marquis M, Kukolj G, Beaulieu PL, Stammers TA. Molecular dynamics simulations and structure-based rational design lead to allosteric HCV NS5B polymerase thumb pocket 2 inhibitor with picomolar cellular replicon potency. J Med Chem. 2014;57(5):1932-1943. https://doi.org/10.1021/jm4004522.

[61] Kumar DV, Rai R, Brameld KA, Somoza JR, Rajagopalan R, Janc JW, Xia YM, Ton TL, Shaghafi MB, Hu H, Lehoux I, To N, Young WB, Green MJ. Quinolones as HCV NS5B polymerase inhibitors. Bioorg Med Chem Lett. 2011;21(1):82-87. https://doi.org/10.1016/j.bmcl.2010.11.068.

[62] Love RA, Parge HE, Yu X, Hickey MJ, Diehl W, Gao J, Wriggers H, Ekker A, Wang L, Thomson JA, Dragovich PS, Fuhrman SA. Crystallographic identification of a noncompetitive inhibitor binding site on the hepatitis C virus NS5B RNA polymerase enzyme. J Virol. 2003;77(13):7575-7581. https://doi.org/10.1128/jvi.77.13.7575-7581.2003

[63] Stammers TA, Coulombe R, Rancourt J, Thavonekham B, Fazal G, Goulet S, Jakalian A, Wernic D, Tsantrizos Y, Poupart MA, Bös M, McKercher G, Thauvette L, Kukolj G, Beaulieu PL. Discovery of a novel series of non-nucleoside thumb pocket 2 HCV NS5B polymerase inhibitors. Bioorg Med Chem Lett. 2013;23(9):2585-2589. https://doi.org/10.1016/j.bmcl.2013.02.110.

[64] Yan S, Appleby T, Larson G, Wu JZ, Hamatake R, Hong Z, Yao N. Structure-based design of a novel thiazolone scaffold as HCV NS5B polymerase allosteric inhibitors. Bioorg Med Chem Lett. 2006;16(22):5888-5891. https://doi.org/10.1016/j.bmcl.2006.08.056.

[65] Yan S, Larson G, Wu JZ, Appleby T, Ding Y, Hamatake R, Hong Z, Yao N. Novel thiazolones as HCV NS5B polymerase allosteric inhibitors: Further designs, SAR, and X-ray complex structure. Bioorg Med Chem Lett. 2007;17(1):63-67. https://doi.org/10.1016/j.bmcl.2006.09.095.

[66] Yan S, Appleby T, Larson G, Wu JZ, Hamatake RK, Hong Z, Yao N. Thiazolone-acylsulfonamides as novel HCV NS5B polymerase allosteric inhibitors: convergence of structure-based drug design and X-ray crystallographic study. Bioorg Med Chem Lett. 2007;17(7):1991-1995. https://doi.org/10.1016/j.bmcl.2007.01.024.

[67] Yang H, Hendricks RT, Arora N, Nitzan D, Yee C, Lucas MC, Yang Y, Fung A, Rajyaguru S, Harris SF, Leveque VJ, Hang JQ, Pogam SL, Reuter D, Tavares GA. Cyclic amide bioisosterism: strategic application to the design and synthesis of HCV NS5B polymerase inhibitors. Bioorg Med Chem Lett. 2010;20(15):4614-4619. https://doi.org/10.1016/j.bmcl.2010.06.008.

[68] Roos K, Wu C, Damm W, Reboul M, Stevenson JM, Lu C, Dahlgren MK, Mondal S, Chen W, Wang L, Abel R, Friesner RA, Harder ED. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. J Chem Theory Comput. 2019;15(3):1863-1874. https://doi.org/10.1021/acs.jctc.8b01026.

[69] Mayack BK, Alayoubi MM, Gezginci MH. Fingerprint-based QSAR Model Generation to Identify Structural Determinants of HCV NS5B Inhibition. J Res Pharm. 2023; 27(4): 1421-1430. http://dx.doi.org/10.29228/jrp.429.

[70] Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, Ohl P, Sieb C, Thiel K, Wiswedel B, editors. KNIME: The Konstanz Information Miner. Data Analysis, Machine Learning and Applications; 2008 2008//; Berlin, Heidelberg: Springer Berlin Heidelberg.

[71] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J Med Chem. 2004;47(7):1739-1749. https://doi.org/10.1021/jm0306430.

[72] Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem. 2004;47(7):1750-1759. https://doi.org/10.1021/jm030644s.

[73] Peter SC, Dhanjal JK, Malik V, Radhakrishnan N, Jayakanthan M, Sundar D. Quantitative Structure-Activity Relationship (QSAR): Modeling Approaches to Biological Applications. In: Ranganathan S, Gribskov M, Nakai K, Schönbach C, editors. Encyclopedia of Bioinformatics and Computational Biology. Oxford: Academic Press; 2019. p. 661-676.

[74] Chun H, Keleş S. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. J R Stat Soc Series B Stat Methodol. 2010;72(1):3-25. https://doi.org/10.1111/j.1467-9868.2009.00723.x.

[75] Allegrini F, Olivieri AC. Two sides of the same coin: Kernel partial least-squares (KPLS) for linear and non-linear multivariate calibration. A tutorial. Talanta Open. 2023;7:100235. https://doi.org/10.1016/j.talo.2023.100235.

[76] Gu F, Cheung MW. A model-based approach to multivariate principal component regression: Selecting principal components and estimating standard errors for unstandardized regression coefficients. Br J Math Stat Psychol. 2023;76(3):605-622. https://doi.org/10.1111/bmsp.12301.

[77] Mayack BK. Modeling disruption of Apis mellifera (honey bee) odorant-binding protein function with high-affinity binders. J Mol Recognit. 2023;36(5):e3008. https://doi.org/10.1002/jmr.3008.

[78] Lu C, Wu C, Ghoreishi D, Chen W, Wang L, Damm W, Ross GA, Dahlgren MK, Russell E, Von Bargen CD, Abel R, Friesner RA, Harder ED. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. J Chem Theory Comput. 2021;17(7):4291-4300. https://doi.org/10.1021/acs.jctc.1c00302.

[79] Biswal BK, Cherney MM, Wang M, Chan L, Yannopoulos CG, Bilimoria D, Nicolas O, Bedard J, James MN. Crystal structures of the RNA-dependent RNA polymerase genotype 2a of hepatitis C virus reveal two conformations and suggest mechanisms of inhibition by non-nucleoside inhibitors. J Biol Chem. 2005;280(18):18202-18210. https://doi.org/10.1074/jbc.m413410200.

[80] Ontoria JM, Rydberg EH, Di Marco S, Tomei L, Attenni B, Malancona S, Martin Hernando JI, Gennari N, Koch U, Narjes F, Rowley M, Summa V, Carroll SS, Olsen DB, De Francesco R, Altamura S, Migliaccio G, Carfi A. Identification and biological evaluation of a series of 1H-benzo[de]isoquinoline-1,3(2H)-diones as hepatitis C virus NS5B polymerase inhibitors. J Med Chem. 2009;52(16):5217-5227. https://doi.org/10.1021/jm900517t

[81] Karaman B, Sippl W. Docking and binding free energy calculations of sirtuin inhibitors. Eur J Med Chem. 2015;93:584-598. https://doi.org/10.1016/j.ejmech.2015.02.045.

[82] Tropsha A. Best Practices for QSAR Model Development, Validation, and Exploitation. Mol Inform. 2010;29(6-7):476-488. https://doi.org/10.1002/minf.201000061.

[83] Sedykh A, Zhu H, Tang H, Zhang L, Richard A, Rusyn I, Tropsha A. Use of in vitro HTS-derived concentration-response data as biological descriptors improves the accuracy of QSAR models of in vivo toxicity. Environ Health Perspect. 2011;119(3):364-370. https://doi.org/10.1289/ehp.1002476.

[84] Falchi F, Bertozzi SM, Ottonello G, Ruda GF, Colombano G, Fiorelli C, Martucci C, Bertorelli R, Scarpelli R, Cavalli A, Bandiera T, Armirotti A. Kernel-Based, Partial Least Squares Quantitative Structure-Retention Relationship Model for UPLC Retention Time Prediction: A Useful Tool for Metabolite Identification. Anal Chem. 2016;88(19):9510-9517. https://doi.org/10.1021/acs.analchem.6b02075.

[85] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 2001;46(1-3):3-26. https://doi.org/10.1016/s0169-409x(00)00129-0.

[86] Ghose AK, Viswanadhan VN, Wendoloski JJ. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. J Comb Chem. 1999;1(1):55-68. https://doi.org/10.1021/cc9800071.

[87] Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem. 2002;45(12):2615-2623. https://doi.org/10.1021/jm020017n.

[88] Egan WJ, Merz KM, Jr., Baldwin JJ. Prediction of drug absorption using multivariate statistics. J Med Chem. 2000;43(21):3867-3877. https://doi.org/10.1021/jm000292e.

[89] Muegge I, Heald SL, Brittelli D. Simple Selection Criteria for Drug-like Chemical Matter. J Med Chem. 2001;44(12):1841-1846. https://doi.org/10.1021/jm015507e.
[90] Stedman C. Sofosbuvir, a NS5B polymerase inhibitor in the treatment of hepatitis C: A review of its clinical potential. Therap Adv Gastroenterol. 2014;7(3):131-140. https://doi.org/10.1177/1756283x13515825.
[91] Dennis BB, Naji L, Jajarmi Y, Ahmed A, Kim D. New hope for hepatitis C virus: Summary of global epidemiologic changes and novel innovations over 20 years. World J Gastroenterol. 2021;27(29):4818-4830. https://doi.org/10.3748/wjg.v27.i29.4818.
[92] Bailey JR, Barnes E, Cox AL. Approaches, Progress, and Challenges to Hepatitis C Vaccine Development. Gastroenterology. 2019;156(2):418-430. https://doi.org/10.1053/j.gastro.2018.08.060.
[93] Dustin LB. Innate and Adaptive Immune Responses in Chronic HCV Infection. Curr Drug Targets. 2017;18(7):826-843. https://doi.org/10.2174/1389450116666150825110532.
[94] Martinez MA, Franco S. Therapy Implications of Hepatitis C Virus Genetic Diversity. Viruses. 2020;13(1):41. https://doi.org/10.3390/v13010041.
[95] Camci M, Şenol H, Kose A, Karaman Mayack B, Alayoubi MM, Karali N, Gezginci MH. Bioisosteric replacement of the carboxylic acid group in Hepatitis-C virus NS5B thumb site II inhibitors: phenylalanine derivatives. Eur J Med Chem. 2024;279:116832. https://doi.org/10.1016/j.ejmech.2024.116832.