



# Targeting non-structural proteins of Hepatitis C virus for predicting repurposed drugs using QSAR and machine learning approaches



Sakshi Kamboj<sup>a,b</sup>, Akanksha Rajput<sup>a</sup>, Amber Rastogi<sup>a,b</sup>, Anamika Thakur<sup>a,b</sup>, Manoj Kumar<sup>a,b,\*</sup>

<sup>a</sup>Virology Unit and Bioinformatics Centre, Institute of Microbial Technology, Council of Scientific and Industrial Research (CSIR), Sector 39A, Chandigarh 160036, India  
<sup>b</sup>Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India

## ARTICLE INFO

### Article history:

Received 31 March 2022  
Received in revised form 27 June 2022  
Accepted 27 June 2022  
Available online 30 June 2022

### Keywords:

Hepatitis C Virus  
Antiviral  
Prediction  
Drug repurposing  
Machine learning  
QSAR  
Non-structural protein  
NS3-NS3/NS4A-NS5A-NS5B

## ABSTRACT

Hepatitis C virus (HCV) infection causes viral hepatitis leading to hepatocellular carcinoma. Despite the clinical use of direct-acting antivirals (DAAs) still there is treatment failure in 5–10% cases. Therefore, it is crucial to develop new antivirals against HCV. In this endeavor, we developed the “Anti-HCV” platform using machine learning and quantitative structure–activity relationship (QSAR) approaches to predict repurposed drugs targeting HCV non-structural (NS) proteins. We retrieved experimentally validated small molecules from the ChEMBL database with bioactivity (IC<sub>50</sub>/EC<sub>50</sub>) against HCV NS3 (454), NS3/4A (495), NS5A (494) and NS5B (1671) proteins. These unique compounds were divided into training/testing and independent validation datasets. Relevant molecular descriptors and fingerprints were selected using a recursive feature elimination algorithm. Different machine learning techniques viz. support vector machine, k-nearest neighbour, artificial neural network, and random forest were used to develop the predictive models. We achieved Pearson’s correlation coefficients from 0.80 to 0.92 during 10-fold cross validation and similar performance on independent datasets using the best developed models. The robustness and reliability of developed predictive models were also supported by applicability domain, chemical diversity and decoy datasets analyses. The “Anti-HCV” predictive models were used to identify potential repurposing drugs. Representative candidates were further validated by molecular docking which displayed high binding affinities. Hence, this study identified promising repurposed drugs viz. naltifine, butalbital (NS3), vinorelbine, epicriptine (NS3/4A), pipecuronium, trimethaphan (NS5A), olodaterol and vemurafenib (NS5B) etc. targeting HCV NS proteins. These potential repurposed drugs may prove useful in antiviral drug development against HCV.

© 2022 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Hepatitis C Virus (HCV) is a pathogenic virus of global health concern. It is known to cause viral hepatitis which leads to liver cirrhosis, hepatorenal syndrome, liver failure, hepatocellular carcinoma and eventually death [1]. It is estimated to infect about 170 million people across the globe, with around 58 million people developing chronic HCV infection [2]. HCV chronic infection leads to about 300,000 to 400,000 deaths worldwide per year <https://www.who.int/news-room/fact-sheets/detail/hepatitis-c>. As HCV has a mortality rate of 5–7% of infected persons per annum, it could also be the cause of the next pandemic [3].

HCV is a positive-sense single stranded enveloped RNA virus belonging to the genus *hepacivirus* of the family *flaviviridae* [4]. HCV genome is of approximately 9.6 kilobases comprising 5' UTR, four structural genes - Core, E1, E2, and p7, six non-structural genes - NS2, NS3, NS4A, NS4B, NS5A, and NS5B and 3'UTR [5]. The genome encodes into a single polyprotein of around 3100 amino acids. This precursor polyprotein undergoes proteolytic cleavage by viral and host proteases to form viral structural and non-structural proteins [1].

Various direct acting antivirals (DAAs) targeting specific viral proteins have been developed to treat HCV infection [6]. Most of these DAAs have been developed for viral non-structural (NS) proteins namely NS3, NS3/NS4A, NS5A and NS5B. For example, NS3/4A inhibitor voxilaprevir and pibrentasvir, NS5A inhibitor glecaprevir and velpatasvir and NS5B RNA-dependent RNA polymerase (RdRp) inhibitor sofosbuvir and dasabuvir [7]. Use of these DAAs eliminated the side-effects of previous interferon therapy and improved

\* Corresponding author at: Virology Unit and Bioinformatics Centre, Institute of Microbial Technology, Council of Scientific and Industrial Research (CSIR), Sector 39A, Chandigarh 160036, India.

E-mail address: [manojk@imtech.res.in](mailto:manojk@imtech.res.in) (M. Kumar).

the quality of life of HCV infected patients. Treatment using DAAs achieves sustained virologic response (SVR) in about 92–95% cases [8]. However, DAA therapy is quite costly and fails in about 5–10% cases due to pre-existing or new generation of resistance-associated substitutions (RAS) [9–10].

RAS renders the treatment regimens involving combinations of DAAs ineffective in patients with resistant variants of the virus [6]. Thus, it is required to develop new drugs for HCV. In this endeavor, repurposing of FDA-approved drugs for HCV is quite promising. Various experimental studies have been conducted to look for repurposed drugs activity against HCV. For instance He et al., 2015 experimentally tested chlorcyclizine for repurposing against HCV [11]. Similarly, Perin et al., 2016 checked for flunarizine activity for HCV [12]. However, taking a computational approach for predicting the promising repurposed drugs could be helpful to deal with the time and resources constraint of the drug discovery process. In this context, quantitative structure–activity relationship approach (QSAR) and machine learning based predictive methods have already been used for different viruses viz., AVCpred [13], AVPPred [14] HIVproti [15], anti-flavi [16] anti-corona [17], HCVpred [18].

Some computational studies for identifying drugs for HCV have also been carried out. Like, da Cunha et al. 2013 used the QSAR approach to look for NS3 protease inhibitors [19]. Venkatesan et al. 2018 developed pharmacophore features based predictive models to identify HCV NS3/4A inhibitors [20]. A web server named HCVpred was developed for predicting the bioactivity of HCV NS5B inhibitors using the classification structure–activity relationship (CSAR) method [18]. Similarly, StackHCV provides classification based models developed by employing QSAR based machine learning techniques to identify NS5B inhibitors [21]. However, there is a need for an integrated platform to predict the antiviral activity of molecules against major HCV NS proteins namely NS3, NS3/4A, NS5A and NS5B using machine learning techniques (MLTs).

In this study, we developed the “Anti-HCV” framework having recursive regression based predictive models to predict antivirals targeting HCV NS3, NS3/4A, NS5A and NS5B proteins. We used different machine learning techniques namely support vector machine (SVM), artificial neural network (ANN), *k*-nearest neighbour (*k*NN), and random forest (RF) for the algorithm development. In addition, we have also predicted the promising repurposed candidates for all four NS proteins using the best developed models by scanning the “DrugBank” database. Few predicted molecules were further validated by molecular docking. This study will be helpful in finding the new drugs targeting major HCV NS proteins.

## 2. Methodology

Overall architecture of “Anti-HCV” is depicted in Fig. 1.

### 2.1. Data collection

The experimentally validated bioactive compounds for different “targets” of HCV were retrieved from the ChEMBL database. ChEMBL is a comprehensive repository for bioactive compounds with their inhibitory properties [22]. The data retrieving steps include:

1. We collected data of bioactive molecules against HCV NS3, NS3/4A, NS5A and NS5B proteins from ChEMBL. We obtained 1121, 1106, 2522 and 3011 entries respectively for the above proteins.
2. The data was filtered for getting inhibitors with  $IC_{50}/EC_{50}$  values and SMILES followed by removing the redundant entries.

3. Finally, we obtained 454, 495, 494 and 1671 unique small molecules targeting HCV NS3, NS3/4A, NS5A and NS5B proteins respectively.
4. These entries were respectively used to develop predictive models for each target utilizing different machine learning techniques.
5. The half-maximal inhibitory concentration  $IC_{50}/EC_{50}$  of these unique entries was changed to  $pIC_{50}$  using equation  $pIC_{50} = -\log_{10}(IC_{50})$ , here  $IC_{50}$  or  $EC_{50}$  are of molar concentration.

The inhibitor datasets used for the model development are provided in Supplementary Table S1, Table S2, Table S3 and Table S4 for NS3, NS3/4A, NS5A and NS5B proteins respectively.

### 2.2. Format conversion

The chemical structures of dataset compounds were converted from SMILES format into the structure-data file (3D-SDF) format using openbabel version 3.1.1 command line [23]. These converted files were later on used as input for extracting the chemical descriptors and fingerprints.

### 2.3. Molecular descriptors calculation

In order to develop QSAR based predictive models for different HCV targets, an open source PaDEL-descriptor software [24] was used to calculate molecular descriptors. We calculated 17,968 chemical descriptors for each molecule present in each dataset. These molecular descriptors and fingerprints depict the information about molecular structure such as molecular weight, number of bonds, solvent accessible area etc. The descriptors are classified into 1D, 2D and 3D features based on their dimensionality and are necessary to understand quantitative structure–activity relationship of compounds [17].

### 2.4. Feature selection

From the 17,968 chemical descriptors for the chemical datasets, the relevant top 50 features to provide input variables for machine learning methods were selected. Feature selection is necessary to avoid overfitting and the curse of dimensionality. We used support vector regression, decision tree regression and perceptron methods in the recursive feature elimination (RFE) method of SciKit library coded in Python for feature selection [25,26,17].

### 2.5. Machine learning methods

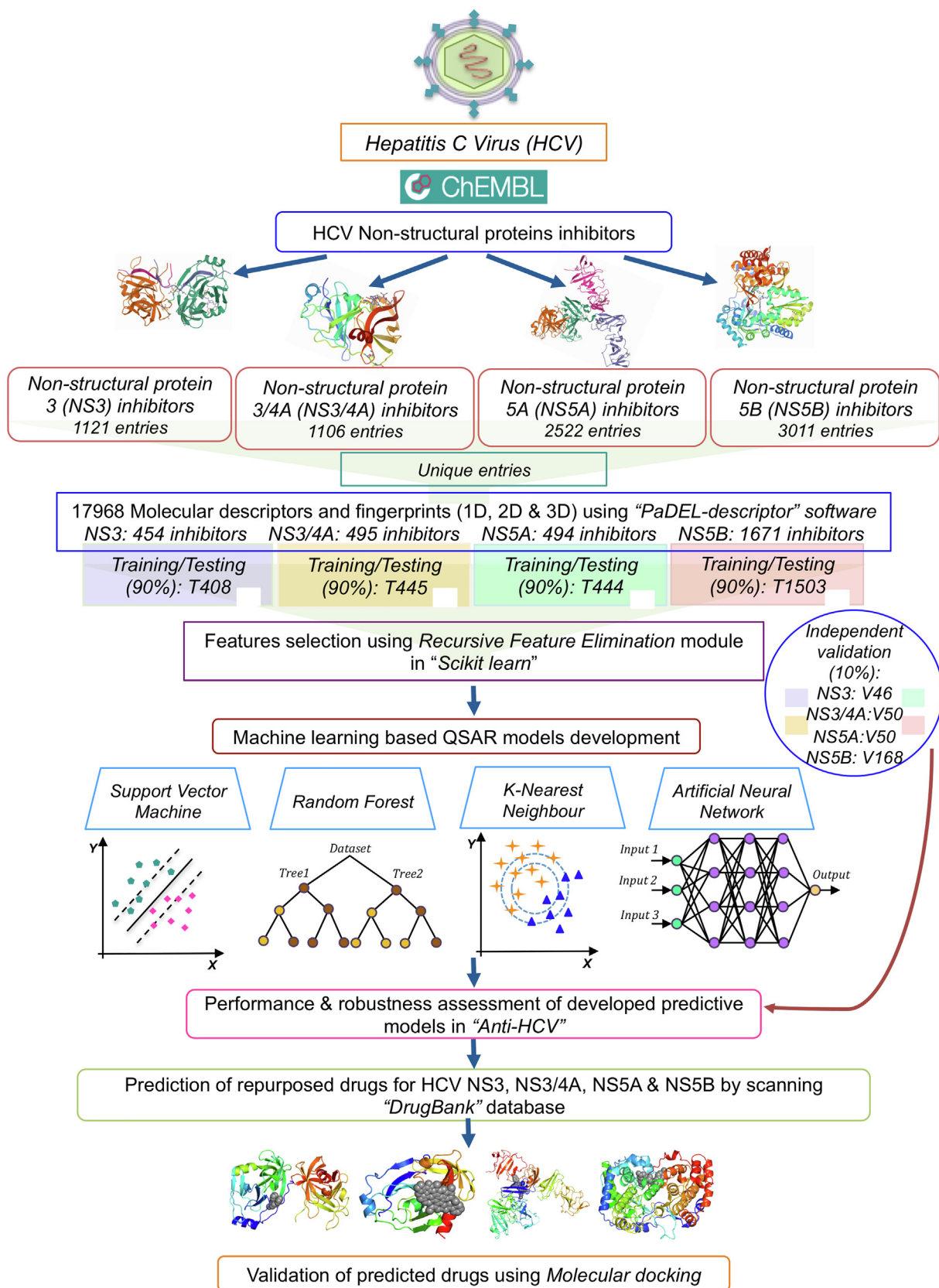
We developed predictive algorithms for targets using four machine learning techniques viz., SVM, ANN, *k*NN, and RF.

SVM is one of the robust supervised machine learning methods used for both classification and regression based problems. SVM works by defining the kernel function and identifying the data points by looking for hyperplanes in very high dimensional space.

*k*NN is a non-parametric supervised machine learning method that looks for matches in training dataset to assign values to the new data points. This method can use different distance matrices for calculation of euclidean distances.

ANN is a supervised machine learning technique consisting of nodes or connecting units in the same as neurons in the animal brain. It learns by processing different inputs having a certain probability-weighted link between the input and output. The weight adjusts itself during the process.

RF is also a supervised machine learning algorithm which works by creating decision trees for data points and developing models and then selecting the best model out of all. This algorithm could be used for both classification as well as regression analysis [27].



**Fig. 1.** Overall methodology used in "Anti-HCV" to develop predictive algorithms to identify inhibitors targeting HCV non-structural proteins – NS3, NS3/4A, NS5A and NS5B. HCV non structural proteins inhibitors were taken from ChEMBL. Molecular descriptors were calculated using PaDEL-descriptor software followed by feature selection using support vector regression (SVR), decision tree regression (DTR) and perceptron method. Selected features were used to develop predictive models using support vector machine (SVM), random forest (RF), k-nearest neighbour (kNN), and artificial neural network (ANN) machine-learning techniques during ten-fold cross validation on training/testing and independent validation datasets. Predictive models performance were assessed along with prediction of repurposed drugs for these NS proteins as well as their structural validation using molecular docking.

## 2.6. Randomized datasets

We randomly selected ~ 10 % molecules from the overall data for each NS protein to be used as an independent validation dataset. The remaining ~ 90% molecules were used for training/testing of the model. This process was repeated five times to generate five such training/testing and independent validation datasets. The final datasets for HCV NS3, NS3/4A, NS5A and NS5B comes out to be 454 (T408 + V46), 495 (T445 + V50), 494 (T444 + V50) and 1671 (T1503 + V168) respectively [17].

## 2.7. Ten-fold cross validation

For ten-fold cross validation, we divided the training/testing datasets into 10 sets randomly. Nine out of ten datasets were used for training the model while the remaining one dataset was used for testing. This was iterated for 10 times such that every dataset was used as a testing dataset. Then, performance of ten iterations averaged for the developed model [17].

## 2.8. Model performance evaluation

The performance of the developed models was assessed by calculating the Pearson's correlation coefficient (R or PCC), mean absolute error (MAE), coefficient of determination ( $R^2$ ) and root mean absolute error (RMSE) values using the formulas given below:

$$PCC = \frac{n \sum_{i=1}^n E_i^{act} E_i^{pred} - \sum_{i=1}^n E_i^{act} \sum_{i=1}^n E_i^{pred}}{\sqrt{n \sum_{i=1}^n (E_i^{act})^2 - (\sum_{i=1}^n E_i^{act})^2} \sqrt{n \sum_{i=1}^n (E_i^{pred})^2 - (\sum_{i=1}^n E_i^{pred})^2}}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |E_i^{pred} - E_i^{act}|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (E_i^{pred} - E_i^{act})^2}$$

where,  $n$ ,  $E^{act}$ , and  $E^{pred}$  are dataset size, actual and predicted efficiencies HCV NS inhibition respectively.

## 2.9. Applicability domain

In addition to model performance assessment, reliability on developed models for new predictions is also important. Applicability domain provides the boundary or chemical space of the developed model for its reliable performance [28]. We used a distance-based leverage approach to assess the applicability domain of the developed models for different HCV NS proteins. The applicability domain space is depicted by the squared area within the  $\pm 3$  band of leverage threshold ( $h^*$ ) and standardized residuals. The leverage threshold is calculated as:

$$h^* = 3(p + 1)/n$$

where,  $n$  is the dataset size and  $p$  is the number of features.

We plotted William's plot using leverage values against standardized residuals to get the applicability domain space for each HCV NS protein datasets. We also plotted actual inhibitory concentration (pIC50) against predicted pIC50 values to check the robustness of the models.

## 2.10. Decoy dataset

Decoys were generated for four HCV NS proteins – HCV NS3, NS3/4A, NS5A, and NS5B protein inhibitors using DecoyFinder 2.0

tool [29]. We used the molecular weight-based approach given in DecoyFinder 2.0 to generate decoys for each HCV NS protein. A subset of about 4.78 million drug-like molecules from the ZINC20 database was used as a source to generate the decoys [30]. Random decoys for each active molecule were selected to develop decoy datasets for each HCV NS protein. The decoy datasets contain 454, 495, 494 and 1671 randomly selected decoys for HCV NS3, NS3/4A, NS5A, and NS5B proteins respectively. Molecular descriptors for each decoy dataset were calculated to predict the inhibitory activity (pIC50). Finally correlation is determined between predicted pIC50 of decoys and actual pIC50 of their corresponding active molecules for each decoy dataset.

## 2.11. Chemical analysis

Chemical diversity of the drugs/compounds used to develop models for HCV NS proteins was checked by performing chemical clustering. We used the multidimensional scaling algorithm with a similarity score of 0.6 of ChemMine tools for chemical clustering [31]. Binning clustering using the Tanimoto coefficient with the same similarity score was also performed.

## 2.12. Drug repurposing

We used the best developed machine learning models to predict the promising repurposed drugs from approved drugs taken from the "DrugBank" database [32]. For this, we collected the "approved" drugs from the DrugBank repository. Format conversion and chemical descriptors calculation was performed for 2468 approved drugs [17]. These descriptors were used to predict the repurposing drug candidates for HCV NS proteins.

## 2.13. Molecular docking

After the prediction of highly efficient drugs for HCV NS proteins, top 5 drugs for each category not yet tested for activity against HCV were selected for docking. The AutoDock tool (ADT) was used to customize the inhibitor molecule and protein, also molecular structure saved in pdbqt format [33]. To dock HCV NS3, NS3/4A, NS5A and NS5B protein structures and a potent inhibitor molecule, AutoDock Vina (v1.1.2) was used at the default parameters [34]. Default settings were used to generate the grid boxes for each protein. Next, the nine best docking postures were created for proteins and ligand molecules. With the exhaustiveness parameter set to 10, we calculated the minimum binding affinity between protein and ligand [17]. The interactions between protein structure and ligands (inhibitors) were interpreted using Discovery Studio Visualizer (DSV) and Pymol version 2.5.2.

## 3. Results

### 3.1. Performance of developed machine learning based QSAR models

The robust prediction models to predict inhibitors for four non-structural proteins of HCV were developed. These models were developed using four machine learning techniques – SVM, ANN, kNN, and RF.

The predictive models for HCV NS proteins inhibitors developed utilized the top 50 features selected by using recursive feature elimination method of Scikit python module. For this, three regression based algorithms namely Support Vector Regression (SVR), perceptron and Decision tree regression (DTR) were used. The selected top 50 features for each NS protein for each method is provided in [Supplementary Table S5](#). Using 10-fold cross validation,



we select top performing models for all 4 NS proteins from randomly generated 5 models for each dataset.

The performance of developed models was evaluated by calculating different statistical measures using 10-fold cross-validation viz, Pearson's Correlation Coefficient (R or PCC), Mean Absolute Error (MAE), coefficient of determination ( $R^2$ ), and Root Mean Absolute Error (RMSE). In this, PCC value depicts the correlation of predicted with actual pIC50 values of the inhibitor. Generally, PCC values range from  $-1$  to  $+1$ , where  $-1$  shows negative correlation while  $+1$  depicts positive correlation. Similarly,  $R^2$  provides the likelihood of estimation of real data from the regression line. Its value also ranges from 0 to 1. More the  $R^2$  value inclines towards 1, higher is the efficiency of estimation. MAE and RMSE estimates give the measures of magnitude of error in prediction of values. More negative the values of MAE and RMSE, the better is the prediction model.

For HCV NS3 protease training/testing data during ten-fold cross-validation, PCC values of 0.84 to 0.86 for SVM, 0.80 to 0.84 for RF, 0.79 to 0.80 for kNN whereas 0.76 to 0.85 for ANN algorithms were achieved. Similarly, the independent validation dataset achieved the PCC of 0.83 to 0.92 for SVM, 0.81 to 0.89 for RF, 0.87 to 0.88 for kNN and 0.81 to 0.90 for ANN after ten-fold cross-validation. The performance measures of best models for each selected feature set, developed using SVM, RF, kNN and ANN for NS3 protein are given in Table 1. The remaining models developed for NS3 are provided in the Supplementary Table S6.

Likewise, for NS3/4A heterodimer protease complex inhibitor developed predictive model showed PCC values ranging from 0.83 to 0.92 for SVM, 0.84 to 0.89 for RF, 0.83 to 0.88 for kNN and 0.77 to 0.89 for ANN for the training/testing dataset. In the case of independent validation dataset, PCC was found to be 0.88 to 0.96 for SVM, 0.90 to 0.91 for RF, 0.86 to 0.91 for kNN while 0.89 to 0.93 for ANN algorithm after ten-fold validation. The performance measures of best models for each selected feature set, developed using SVM, RF, kNN and ANN for NS3/4A heterodimer protein are given in Table 2. The detailed information about mod-

els developed for the NS3/4A heterodimer protein complex are provided in the Supplementary Table S7.

The prediction models developed for NS5A displayed PCC values for SVM to be 0.80 to 0.88, 0.81 to 0.87 for RF, 0.80 to 0.86 for kNN and 0.73 to 0.87 in ANN models for training/testing data with 10-fold cross validation. For independent validation dataset, PCC were ranging from 0.78 to 0.87 for SVM models, 0.81 to 0.88 for RF, 0.83 to 0.86 for kNN and 0.72 to 0.84 for ANN models. The performance measures of best models for each selected feature set, developed using SVM, RF, kNN and ANN for NS5A protein are given in Table 3. The detailed information about models developed for the NS5A protein are provided in the Supplementary Table S8.

For NS5B RdRp, the predictive models developed showed PCC of 0.62 to 0.84 for SVM, 0.67 to 0.85 for RF, 0.64 to 0.83 for kNN, and 0.59 to 0.81 for ANN for training/testing data with ten-fold cross validation. For independent validation dataset, PCC ranges from 0.60 to 0.84 for SVM, 0.64 to 0.86 for RF, 0.61 to 0.85 for kNN and 0.62 to 0.84 for ANN. The performance measures of best models for each selected feature set, developed using SVM, RF, kNN and ANN for NS5B protein are given in Table 4. The detailed information about models developed for the NS5B protein are provided in the Supplementary Table S9.

### 3.2. Analysis of applicability domain

The applicability domain analysis showed the robustness of the developed models. The threshold leverage ( $h^*$ ) values of 1.375, 1.28, 1.344 and 1.27 for NS3, NS3/4A, NS5A and NS5B respectively depicted that the developed models are highly reliable Fig. 2. The plots between actual and predicted pIC50 for training/testing and validation datasets also revealed that most points fall near the trend line showing the robustness of the developed models Fig. 3. The additional data for applicability domain analysis and actual vs predicted pIC50 plots is provided in Supplementary Table S10 and S11.

**Table 1**

The statistical measures of performance of the best predictive models developed for NS3 protein using different machine-learning techniques and selected features during ten-fold cross validation on training/testing and independent validation datasets.

Machine learning model	Feature selection algorithm	Machine learning model information	Dataset	MAE	RMSE	$R^2$	PCC
Support Vector Machine	SVR	gamma:0.0005 C:100	T408	0.40	0.64	0.72	0.86
			V46	0.22	0.47	0.83	0.92
	DTR	gamma:0.01 C:10	T408	0.45	0.71	0.69	0.84
			V46	0.42	0.65	0.69	0.83
	PCT	gamma:0.001 C:200	T408	0.43	0.65	0.71	0.85
			V46	0.25	0.50	0.81	0.90
Random Forest	SVR	n:500 depth:None split:2 leaf:4	T408	0.55	0.76	0.62	0.80
			V46	0.47	0.68	0.65	0.81
	DTR	n:100 depth:12 split:5 leaf:2	T408	0.44	0.69	0.70	0.84
			V46	0.29	0.54	0.78	0.89
	PCT	n:500 depth:8 split:2 leaf:1	T408	0.56	0.74	0.62	0.80
			V46	0.42	0.65	0.68	0.83
<i>k</i> -Nearest Neighbour	SVR	k:5	T408	0.56	0.76	0.62	0.80
			V46	0.31	0.56	0.77	0.88
	DTR	k:9	T408	0.55	0.77	0.62	0.79
			V46	0.32	0.57	0.76	0.87
	PCT	k:5	T408	0.56	0.76	0.62	0.80
			V46	0.33	0.57	0.76	0.87
Artificial Neural Network	SVR	solver:sgd activation:tanh learning:constant	T408	0.40	0.64	0.71	0.85
			V46	0.25	0.50	0.81	0.90
	DTR	solver:sgd activation:tanh learning:constant	T408	0.65	0.81	0.54	0.76
			V46	0.52	0.72	0.61	0.81
	PCT	solver:sgd activation:tanh learning:constant	T408	0.40	0.63	0.71	0.85
			V46	0.28	0.53	0.79	0.89

\* SVR = Support Vector Regression, DTR = Decision tree regression, PCT = Perceptron method, MAE = Mean absolute Error; RMSE = Root Mean Absolute Error, PCC = Pearson's correlation coefficient,  $R^2$  = Coefficient of Determination, T = Training or Testing dataset, V = Validation dataset (independent).

**Table 2**

The statistical measures of performance of the best predictive models developed for NS3/4A heterodimer protein complex using different machine-learning techniques and selected features during ten-fold cross validation on training/testing and independent validation datasets.

Machine learning model	Feature selection algorithm	Machine learning model information	Dataset	MAE	RMSE	R <sup>2</sup>	PCC
Support Vector Machine	SVR	gamma:0.001 C:100	T445	0.38	0.62	0.82	0.92
			V50	0.20	0.44	0.92	0.96
	DTR	gamma:0.01 C:10	T445	0.53	0.72	0.75	0.88
			V50	0.41	0.64	0.83	0.91
	PCT	gamma:0.05 C:1	T445	0.74	0.85	0.68	0.83
			V50	0.52	0.72	0.78	0.88
Random Forest	SVR	n:300 depth:10 split:5 leaf:1	T445	0.54	0.69	0.76	0.88
			V50	0.47	0.68	0.80	0.90
	DTR	n:100 depth:8 split:2 leaf:1	T445	0.48	0.67	0.78	0.89
			V50	0.39	0.62	0.84	0.91
	PCT	n:200 depth:8 split:10 leaf:1	T445	0.72	0.79	0.69	0.84
			V50	0.44	0.67	0.81	0.90
k-Nearest Neighbour	SVR	k:3	T445	0.53	0.74	0.76	0.88
			V50	0.46	0.68	0.80	0.90
	DTR	k:7	T445	0.65	0.79	0.70	0.85
			V50	0.62	0.79	0.74	0.86
	PCT	k:5	T445	0.77	0.86	0.67	0.83
			V50	0.43	0.65	0.82	0.91
Artificial Neural Network	SVR	solver:sgd activation:tanh learning:adaptive	T445	0.50	0.68	0.76	0.89
			V50	0.32	0.57	0.86	0.93
	DTR	solver:sgd activation:tanh learning:adaptive	T445	0.81	0.81	0.60	0.81
			V50	0.49	0.70	0.79	0.89
	PCT	solver:sgd activation:tanh learning:adaptive	T445	1.08	0.92	0.47	0.77
			V50	0.52	0.72	0.78	0.89

\* SVR = Support Vector Regression, DTR = Decision tree regression, PCT = Perceptron method, MAE = Mean absolute Error; RMSE = Root Mean Absolute Error, PCC = Pearson's correlation coefficient, R2 = Coefficient of Determination, T = Training or Testing dataset, V = Validation dataset (independent).

**Table 3**

The statistical measures of performance of the best predictive models developed for NS5A protein using different machine-learning techniques and selected features during ten-fold cross validation on training/testing and independent validation datasets.

Machine learning model	Feature selection algorithm	Machine learning model parameters	Dataset	MAE	RMSE	R <sup>2</sup>	PCC
Support Vector Machine	SVR	gamma:0.001 C:300	T444	0.77	0.82	0.78	0.88
			V50	1.01	1.01	0.73	0.86
	DTR	gamma:0.01 C:50	T444	0.89	0.91	0.74	0.87
			V50	0.96	0.98	0.74	0.87
	PCT	gamma:0.05 C:10	T444	1.30	1.12	0.62	0.80
			V50	1.53	1.24	0.59	0.78
Random Forest	SVR	n:500 depth:None split:2 leaf:1	T444	1.10	1.01	0.68	0.83
			V50	1.37	1.17	0.64	0.81
	DTR	n:500 depth:12 split:2 leaf:2	T444	0.86	0.90	0.75	0.87
			V50	0.83	0.91	0.78	0.88
	PCT	n:100 depth:8 split:10 leaf:2	T444	1.20	1.04	0.64	0.81
			V50	1.12	1.06	0.70	0.85
k-Nearest Neighbour	SVR	k:3	T444	0.99	0.97	0.71	0.85
			V50	1.17	1.08	0.69	0.83
	DTR	k:5	T444	0.92	0.96	0.73	0.86
			V50	0.99	0.99	0.74	0.86
	PCT	k:7	T444	1.27	1.10	0.63	0.80
			V50	1.16	1.08	0.69	0.83
Artificial Neural Network	SVR	solver:sgd activation:tanh learning:constant	T444	0.87	0.89	0.75	0.87
			V50	1.15	1.07	0.69	0.84
	DTR	solver:sgd activation:tanh learning:constant	T444	1.05	1.05	0.69	0.84
			V50	1.13	1.06	0.70	0.84
	PCT	solver:sgd activation:tanh learning:constant	T444	1.87	1.44	0.47	0.73
			V50	2.53	1.59	0.32	0.72

\* SVR = Support Vector Regression, DTR = Decision tree regression, PCT = Perceptron method, MAE = Mean absolute Error; RMSE = Root Mean Absolute Error, PCC = Pearson's correlation coefficient, R2 = Coefficient of Determination, T = Training or Testing dataset, V = Validation dataset (independent).

### 3.3. Validation using decoy set

Decoys are generally regarded as inactive molecules that are unable to bind targets, unlike active molecules. For validating the developed models' robustness, the inhibitory activity (pIC50) of each decoy was predicted and compared with the inhibitory activity (pIC50) of its corresponding active molecule. PCC values for each HCV NS protein decoy dataset were calculated. The decoy datasets showed PCC values of -0.074, -0.037, 0.052 and 0.117

for HCV NS3, NS3/4A, NS5A, and NS5B proteins respectively and displayed through scatter plot in Fig. 4.

### 3.4. Chemical diversity analysis

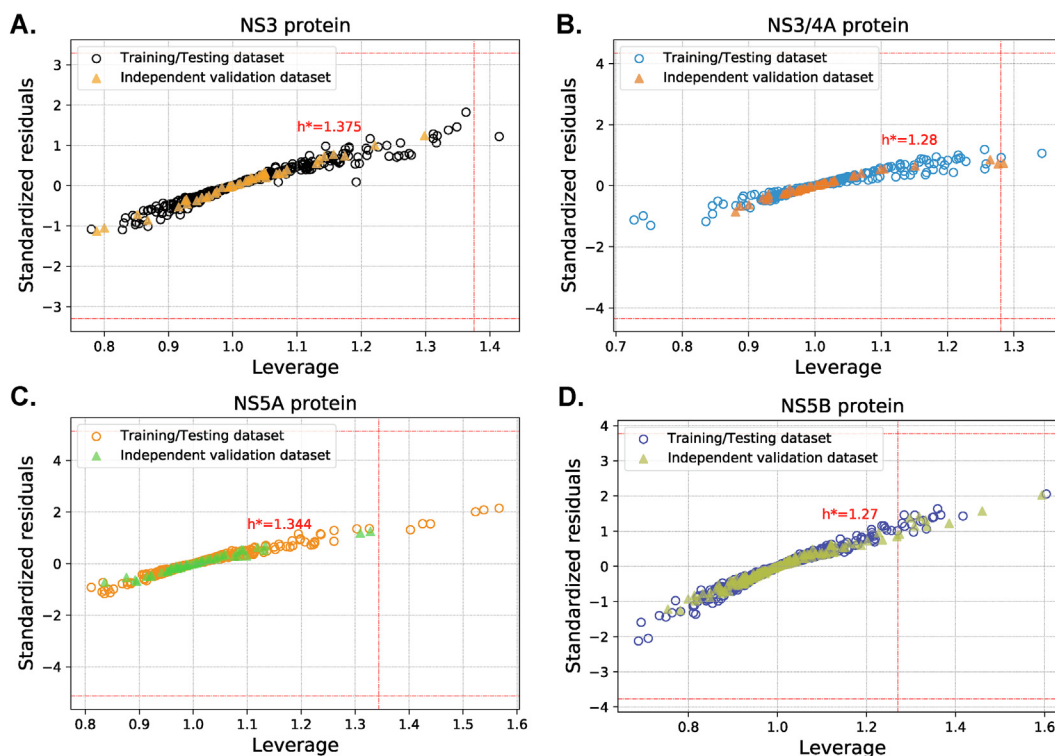
Chemical clustering was performed to assess the diversity in chemical structures of compounds for NS3, NS3/4A, NS5A and NS5B proteins. Binning clustering showed compounds clustered into 159 bins/clusters for NS3 protein, 25 bins for NS3/4A pro-

**Table 4**

The statistical measures of performance of the best predictive models developed for NS5B protein using different machine-learning techniques and selected features during ten-fold cross validation on training/testing and independent validation datasets.

Machine learning model	Feature selection algorithm	Machine learning model parameters	Dataset	MAE	RMSE	R <sup>2</sup>	PCC
Support Vector Machine	SVR	gamma:0.05 C:1	T1503	0.54	0.74	0.70	0.84
			V168	0.57	0.76	0.70	0.84
	DTR	gamma:0.01 C:10	T1503	0.58	0.78	0.67	0.82
			V168	0.65	0.81	0.66	0.81
	PCT	gamma:1 C:10	T1503	1.10	1.04	0.38	0.62
			V168	1.24	1.11	0.35	0.60
Random Forest	SVR	n:200 depth:None split:2 leaf:1	T1503	0.56	0.75	0.69	0.83
			V168	0.66	0.81	0.65	0.81
	DTR	n:400 depth: None split:2 leaf:1	T1503	0.51	0.71	0.71	0.85
			V168	0.52	0.72	0.73	0.86
	PCT	n:100 depth: 12 split:5 leaf:1	T1503	0.99	0.99	0.44	0.67
			V168	1.12	1.06	0.41	0.64
<i>k</i> -Nearest Neighbour	SVR	k:5	T1503	0.59	0.78	0.67	0.82
			V168	0.60	0.78	0.68	0.83
	DTR	k:7	T1503	0.57	0.76	0.68	0.83
			V168	0.55	0.74	0.71	0.85
	PCT	k:9	T1503	1.06	1.03	0.40	0.64
			V168	1.20	1.09	0.37	0.61
Artificial Neural Network	SVR	solver:adam activation:tanh learning:constant	T1503	0.60	0.76	0.66	0.81
			V168	0.58	0.76	0.70	0.84
	DTR	solver:adam activation:tanh learning:constant	T1503	0.62	0.81	0.65	0.81
			V168	0.70	0.84	0.63	0.80
	PCT	solver:sgd activation:tanh learning:constant	T1503	1.19	1.08	0.33	0.59
			V168	1.17	1.08	0.39	0.62

\* SVR = Support Vector Regression, DTR = Decision tree regression, PCT = Perceptron method, MAE = Mean absolute Error; RMSE = Root Mean Absolute Error, PCC = Pearson's correlation coefficient, R2 = Coefficient of Determination, T = Training or Testing dataset, V = Validation dataset (independent).

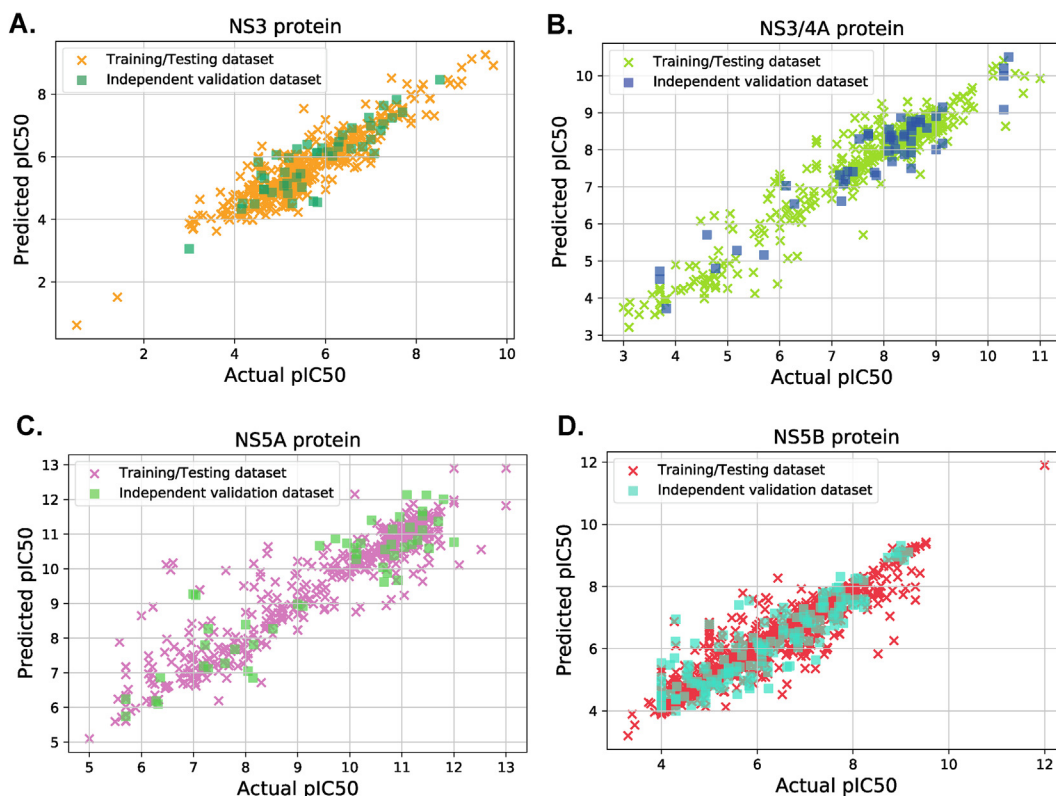


**Fig. 2.** William plots for applicability domain analysis of the support vector machine based predictive models developed for each HCV NS protein – (A) NS3, (B) NS3/4A, (C) NS5A and (D) NS5B.

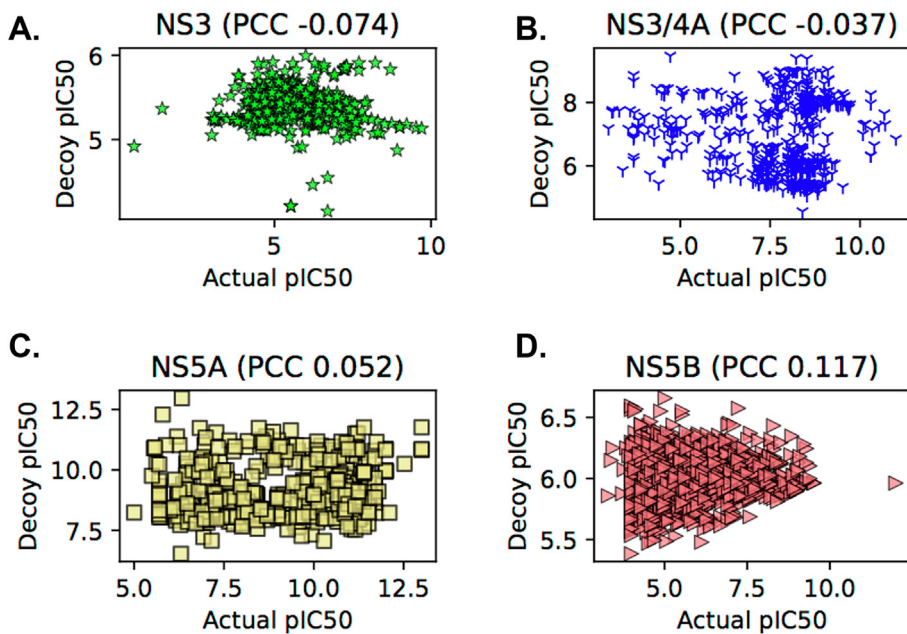
tein, 25 bins for NS5A protein and 142 bins for NS5B protein with similarity threshold of 0.6 [Supplementary Table S12](#). The 3D multidimensional scaling plots showed the chemical diversity of compounds for NS3, NS3/4A, NS5A and NS5B proteins [Fig. 5](#).

### 3.5. Prediction of repurposed drugs targeting HCV non-structural protein NS3

The best performing SVM predictive models developed for HCV NS3 were used to predict the promising repurposing drugs target-

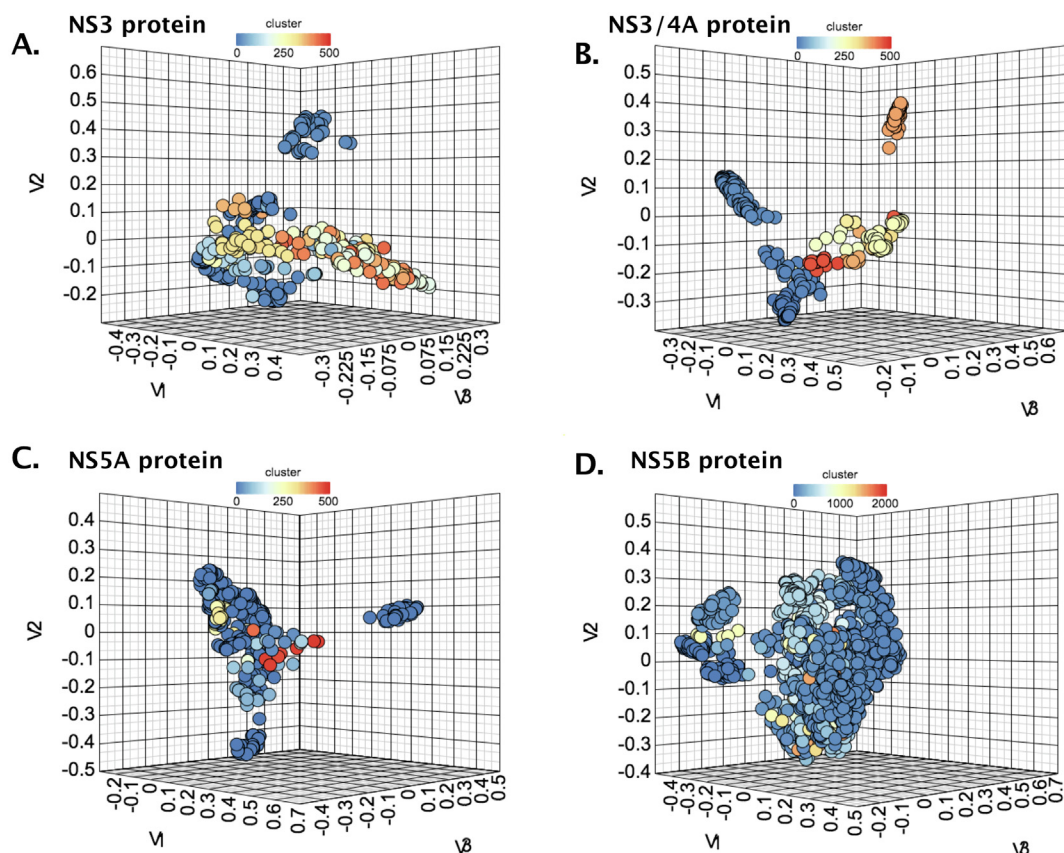


**Fig. 3.** Support vector machine based developed predictive models robustness shown by the plots between actual and predicted pIC50 of molecules for each HCV NS protein - (A) NS3, (B) NS3/4A, (C) NS5A and (D) NS5B.



**Fig. 4.** Scatter plots to display correlation between actual and predicted pIC50 of decoys and active molecules for HCV NS proteins - (A) NS3, (B) NS3/4A, (C) NS5A and (D) NS5B.





**Fig. 5.** The chemical analysis of inhibitors shown by 3-dimensional multiscale plots among the compounds for each HCV NS protein - (A) NS3, (B) NS3/4A, (C) NS5A and (D) NS5B.

**Table 5**

Table showing information for top 10 predicted repurposed drugs for HCV NS3 protein namely drug, DrugBank ID, primary use, predicted pIC50 and clinical status for HCV.

DrugBank Id	Name of Drug	Primary use	pIC50 (predicted against NS3)	Clinical status
DB00970	Dactinomycin	Anticancer	8.95	Not yet tested
DB00735	Naftifine	Antifungal drug	8.80	Not yet tested
DB01410	Ciclesonide	Obstructive airway diseases	8.70	Not yet tested
DB13253	Proxibarbal	Migraines treatment	8.61	Not yet tested
DB00241	Butalbital	Treatment of tension-type headache	8.51	Not yet tested
DB13170	Plecanatide	Chronic idiopathic constipation and IBS	8.48	Not yet tested
DB00474	Methohexital	Anesthetic for deep sedation	8.48	Not yet tested
DB15465	Benzhydrocodone	Pain reliever	8.42	Not yet tested
DB06711	Naphazoline	Vasoconstrictor to relieve eyes itching and redness	8.24	Not yet tested
DB01091	Butenafine	Antifungal	8.12	Not yet tested

**Table 6**

Table showing information for top 10 predicted repurposed drugs for HCV NS3/4A protein namely drug, DrugBank ID, primary use, predicted pIC50 and clinical status for HCV.

DrugBank Id	Name of Drug	Primary use	pIC50 (predicted against NS3/4A)	Clinical status
DB01395	Drospirenone	Oral contraceptive pills	13.48	Not yet tested
DB06402	Telavancin	Antibacterial agent	13.14	Not yet tested
DB00361	Vinorelbine	Metastatic non-small cell lung carcinoma (NSLC)	12.57	Not yet tested
DB11275	Epicriptine	Idiopathic decline in mental capacity	12.46	Not yet tested
DB00320	Dihydroergotamine	Migraine and cluster headache	12.44	Not yet tested
DB11273	Dihydroergocornine	Idiopathic decline in mental capacity	12.42	Not yet tested
DB00696	Ergotamine	Treatment of migraine disorders	11.77	Not yet tested
DB04911	Oritavancin	Antibacterial	11.52	Not yet tested
DB06663	Pasireotide	Cushing's disease treatment	11.24	Not yet tested
DB00256	Lymecycline	Acne vulgaris and other infections	11.10	Not yet tested

ing NS3. The repurposed drugs were predicted from the approved drugs available in the “DrugBank ” database. Top 10 predicted repurposed drug candidates for NS3 protein are given in [Table 5](#).

### 3.6. Prediction of repurposed drugs targeting HCV non-structural protein complex NS3/4A

We selected the best performing prediction model for NS3/4A and used this model to predict the repurposing drugs from the “DrugBank” database. Top 10 predicted repurposed drug candidates for NS3/4A protein are given in [Table 6](#).

### 3.7. Prediction of repurposed drugs targeting HCV non-structural protein NS5A

The potential repurposing drugs were predicted from the approved drugs available in the DrugBank database. The best prediction model was used to predict the repurposing molecules for HCV NS5A. Top 10 predicted repurposed drug candidates for NS5A protein are given in [Table 7](#).

### 3.8. Prediction of repurposed drugs targeting HCV non-structural protein NS5B

The best predictive model developed for NS5B inhibitors based on SVM was used to predict the potential repositioning drug candidates for NS5B from the DrugBank approved drugs. Top ten predicted repurposed candidates for NS5B are given in [Table 8](#).

The top 200 repurposing drug candidates for each NS protein are provided in [Supplementary Table S13](#).

### 3.9. Molecular docking of predicted inhibitors with NS proteins

Docking is an advantageous method to determine the affinity and interactions of proteins with different ligands and vice versa.

**Table 7**

Table showing information for top 10 predicted repurposed drugs for HCV NS5A protein namely drug, DrugBank ID, primary use, predicted pIC50 and clinical status for HCV.

DrugBank Id	Name of Drug	Primary use	pIC50 (predicted against NS5A)	Clinical status
DB11585	Drometizole trisiloxane	UV ray absorbing agent	13.66	Not yet tested
DB00728	Rocuronium	Facilitate tracheal intubation and relax skeletal muscles during surgery	13.33	Not yet tested
DB01338	Pipecuronium	Neuromuscular blocking agent, used as anesthetic	13.17	Not yet tested
DB00210	Adapalene	Acne vulgaris	13.04	Not yet tested
DB01116	Trimethaphan	Ganglionic blocker in hypertension	12.82	Not yet tested
DB14879	Cefiderocol	Cephalosporin antibiotic for urinary tract infections	12.57	Not yet tested
DB11951	Lemborexant	Insomnia treatment	12.52	Not yet tested
DB01190	Clindamycin	Bacterial infections	12.30	Not yet tested
DB13284	Meticrane	Diuretic	12.17	Not yet tested
DB01180	Rescinnamine	Antihypertensive drug	12.16	Not yet tested

**Table 8**

Table showing information for top 10 predicted repurposed drugs for HCV NS5B protein namely drug, DrugBank ID, primary use, predicted pIC50 and clinical status for HCV.

DrugBank Id	Name of Drug	Primary use	pIC50 (predicted against NS5B)	Clinical status
DB13125	Lusutrombopag	Thrombocytopenia treatment	8.06	Not yet tested
DB05294	Vandetanib	Symptomatic or progressive medullary thyroid cancer treatment	7.97	Not yet tested
DB00365	Grepafloxacin	Antibiotic to treat gram positive and gram negative bacterial infections	7.92	Not yet tested
DB09080	Olodaterol	Treatment of chronic obstructive pulmonary disease (COPD)	7.68	Not yet tested
DB12035	Sarecycline	Inflammatory lesions or acne vulgaris treatment	7.58	Not yet tested
DB08881	Vemurafenib	For the treatment of metastatic melanoma	7.67	Not yet tested
DB00334	Olanzapine	Antipsychotic drug	7.37	Not yet tested
DB14033	Acetyl sulfisoxazole	Antibacterial agent	7.35	Not yet tested
DB01044	Gatifloxacin	Treatment of different infections	7.27	Not yet tested
DB12792	Boscalid	Glaucoma and Schirmers treatment	7.17	Not yet tested

We selected the top three molecules for each HCV NS protein depending on the predicted pIC50 value for docking. Further, these compounds were sequentially docked for their respective protein. In the case of NS3 repurposed drug, the lowest binding energy was observed in naftifine and NS3 protein (PDB Id 2XCF), i.e.,  $-7.8$  Kcal/mol and indicated four different types of interactions as mentioned in [Table 9](#). with NS3 protein. In contrast, the remaining two molecules, namely, butalbital and proxibarbal, have binding energy  $\sim -6.3$  and  $-6.2$  Kcal/mol, respectively. Additionally, three molecules, namely vinorelbine, epicutipine, and drospirenone, were docked on HCV NS3/4A protein (PDB Id 4WF8) showed the lowest binding energy ranged from  $-8.9$  to  $-8.1$  Kcal/mol as well as showed different types of interactions as mentioned in [Table 9](#). Apart from this, three molecules were docked on NS5A protein (PDB Id 4CL1) of HCV. Interestingly, the pipecuronium molecule showed the lowest minimum binding energy  $\sim -9.8$  Kcal/mol. In addition, the remaining molecules, trimethaphan and cefiderocol showed minimum binding energy  $-9.4$  and  $-9.2$  Kcal/mol, and their interacting residues were mentioned in [Table 9](#). Additionally, three molecules i.e., olodaterol, vemurafenib, and grepafloxacin were docked on NS5B protein (PDB Id 3VQS). In the case of NS5B all three molecules showed  $-8.0$  Kcal/mol given in [Table 9](#). Ribbon structure of proteins NS3, NS3/4A, NS5A, NS5B binding with respective ligand molecules are displayed in [Fig. 6](#). Whereas, their molecular interactions in two dimensions form are shown in [Fig. 7](#).

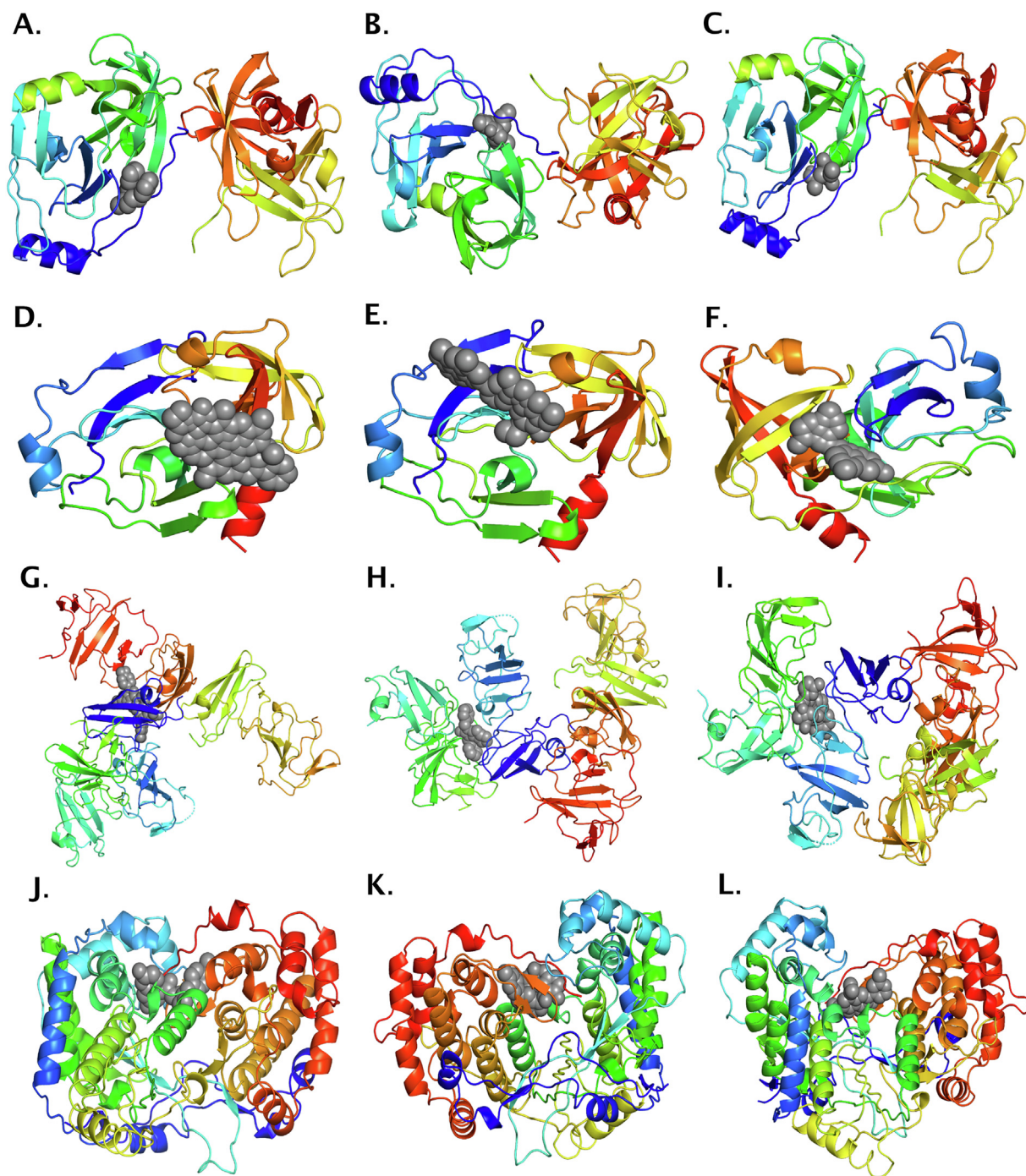
## 4. Discussion

Hepatitis C Virus (HCV) causes viral hepatitis characterized by acute liver inflammation to severe conditions like liver cirrhosis, hepatic encephalopathy, liver failure, hepatorenal syndrome and hepatocellular carcinoma [35 36]. About 55–85% of infected HCV patients develop chronic infection which is associated with the incidence of liver cancer [37 38]. HCV is a positive-sense single

**Table 9**

Table represents the ligand, protein name, protein Id (PDB id), binding affinity, interacting residues, distance between interacting residues (Å), types of molecular interactions.

Protein (PDB id)	Inhibitory ligand (Drugbank id)	Affinity (Kcal/mol)	Interacting residues	Distance (Å)	Molecular interactions
NS3 (2XCF)	Naftifine (DB00735)	−7.8	ALA-A:5	5.08	Van der waals
			TYR-A:6	3.76	Pi-anion
			ALA-A:7	4.76	Pi-Alkyl
			GLU-A:32	6.45	Alkyl
	Butalbital (DB00241)	−6.3	VAL-A:33	4.26, 4.95	
			ALA-A:5	4.57	Van der waals
			TYR-A:6	3.49	Conventional hydrogen bond
			GLU-A:32	3.62	Pi-Alkyl
	Proxibarbal (DB13253)	−6.2	VAL-A:33	3.64, 3.70, 6.32	Alkyl
			VAL-A:107	4.55	
			TYR-A:6	5.35	Van der waals
			GLN-A:8	3.48, 4.74	Conventional hydrogen bond
NS3/4A (4WF8)	Vinorelbine (DB00361)	−8.9	GLU-A:32	4.26	Pi-Alkyl
			VAL-A:33	3.64, 3.70, 6.32	Alkyl
			VAL-A:107	4.55	
	Epicriptine (DB11275)	−8.4	ASP-A:1081	5.21	Unfavorable Donar-Donar
			ASP-A:1168	5.26, 5.71, 5.73	Van der waals
			ARG-A:1155	8.21	Attractive charge
					Pi-cation
	Drospirenone (DB01395)	−8.1	ALA-A:1013	4.2, 5.2, 6	Van der waals
			GLN-A:1041	4.33	Carbon hydrogen bond
			THR-A:1042	5.64	Conventional hydrogen bond
			PHE-A:1043	5.20	Unfavorable Donar-Donar
	NS5A (4CL1)	Pipecuronium (DB01338)	−9.8	HIS-A:1057	5.73
GLY-A:1137					
ALA-A:1005				4.26	Van der waals
Trimethaphan (DB01116)		−9.4	VAL-A:1113	5.45	Alkyl
			ARG-A:12, 15	4.3, 5.8,	Carbon hydrogen bond
			ARG-C:49	5.59,	Conventional hydrogen bond
			ARG-D:15	5.08, 5.30	Van der waals
Cefiderocol (DB14879)		−9.2	GLY-A:13	5.74	Attractive charge
			ALA-D:74	4.91	Alkyl
			LEU-A:124	4.9	
			ASP-A:125	4.65, 5.76, 5.78	
NS5B (3VQS)		Oladaterol (DB09080)	−8.8	ARG-A:131	4.31
	PHE-B:132			5.62	Van der waals
	ARG-B:83			4.63, 4.77	Pi-Donor hydrogen bond
	Vemurafenib (DB08881)	−8.7	TRP-B:82	5.74	Pi-sigma
					Pi-Pi stacked
					Pi- Alkyl
			THR-A:65	4.64	Van der waals
	Grepafloxacin (DB00365)	−8.4	ARG-A:83	4.93	Attractive charge
			GLU-A:119	6.61	Conventional hydrogen bond
			HIS-A:130	4.09	Pi-Anion
			ARG-A:131	5.60	Pi-Donar hydrogen atom
	NS5B (3VQS)	Oladaterol (DB09080)	−8.8	PHE-A:132	5.26, 6.24
TRP-B:82				6.88	Pi-Pi T-shaped
GLU-B:119				5.94	Pi-Alkyl
Vemurafenib (DB08881)		−8.7	PHE-B:120	4.20	
			HIS-B:130	3.95	
			SER-A:180	5.47	Van der waals
			PHE-A:193	4.51, 4.67	Conventional hydrogen bond
Grepafloxacin (DB00365)		−8.4	GLN-A:194	6.99	Pi-Donar hydrogen atom
			SER-A:288	4.19	Pi-Sulfur
			MET-A:414	6.76, 6.45	Pi-Pi stacked
			ILE-A:447	5.26	Pi-Pi T-shaped
NS5B (3VQS)		Oladaterol (DB09080)	−8.8	TYR-A:452	5.26, 6.22, 6.63, 6.63
	ILE-A:454			5.76	Pi-Alkyl
	LEU-A:547			4.0, 5.81	
	Vemurafenib (DB08881)	−8.7	SER-A:196	4.44	Van der waals
			MET-A:414	6.19	Conventional hydrogen bond
			ILE-A:447	5.28	Carbon-hydrogen bond
			ILE-A:454	5.41	Pi-sigma
	Grepafloxacin (DB00365)	−8.4	ILE-A:462	4.78	Pi-Pi stacked
			LEU-A:466	6.08	Alkyl
			LEU-A:547	4.48, 4.59	Pi-Alkyl
			TYR-A:191	6.60	Van der waals
	Grepafloxacin (DB00365)	−8.4	PHE-A:193	4.10	Conventional hydrogen bond
SER-A:288			4.76	Pi-Donar hydrogen atom	
ALA-A:450			6.70	Alkyl	
LEU-A:540			4.04	Pi-Alkyl	

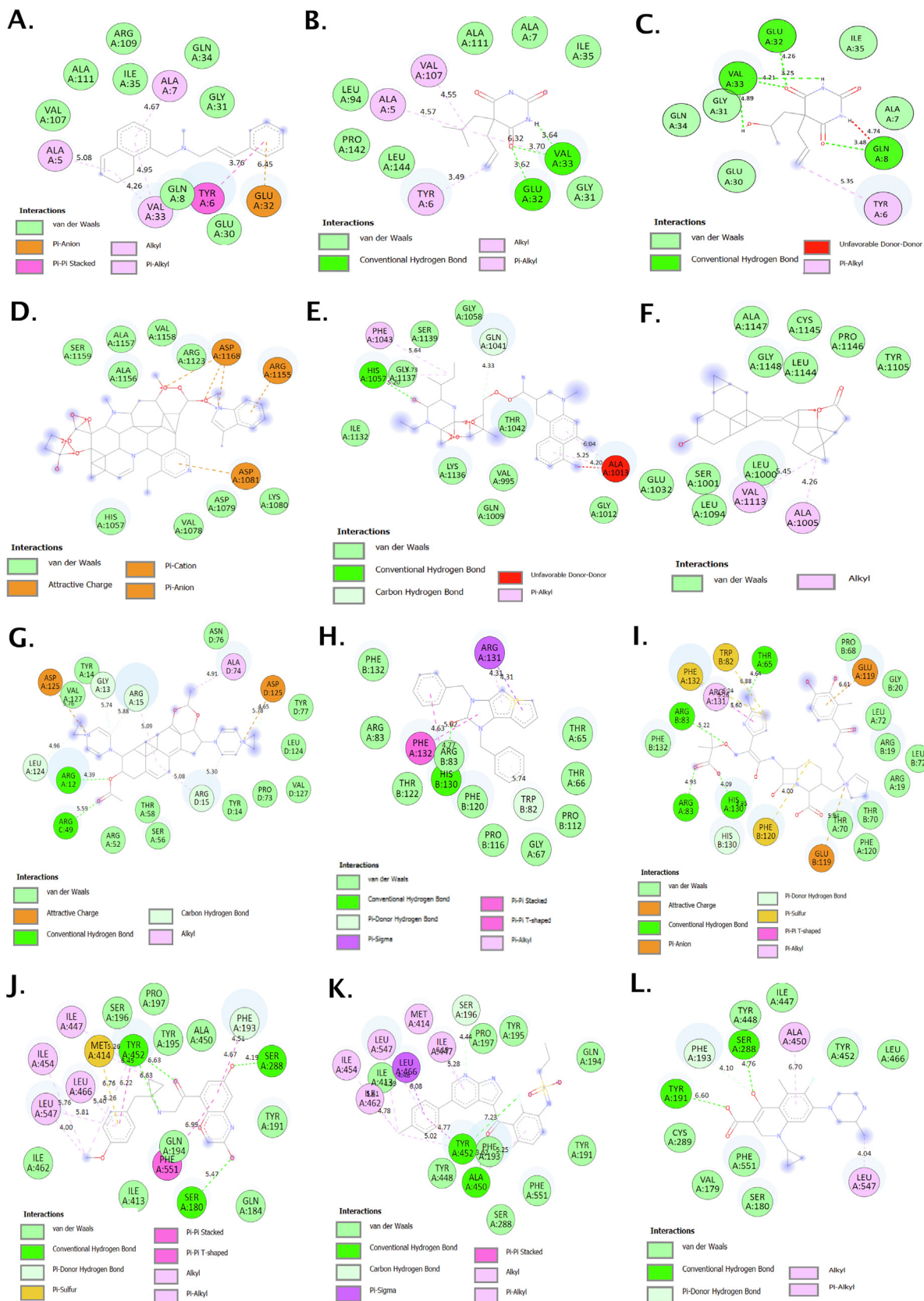


**Fig. 6.** Ribbon structure of proteins NS3, NS3/4A, NS5A, NS5B binding with respective ligand molecules (A) Represents the structure of NS3 protein and naftifine (B) structure of NS3 and butalbital (C) structure of NS3 protein and proxibarbalb (D) structure of NS3/4a protein and vinorelbine (E) structure of NS3/4a protein and epicriptine (F) structure of NS3/4a protein and drospirenone (G) structure of NS5A and pipecuronium (H) structure of NS5A trimethaphan (I) structure of NS5A and cefiderocol (J) structure of NS5B and olodaterol (K) structure of NS5B and vemurafenib (L) structure of NS5B and grepafloxacin (Protein in Rainbow color and ligand molecule is gray color sphere).

stranded RNA virus exhibiting error-prone replication due to RNA-dependent RNA polymerase [39]. Due to high mutation rate, there is generation of various viral variants called quasispecies [40–41]. The traditional regimen of pegylated interferon and ribavirin therapy have been used for years but are reported to show SVR rates of only 55% [42]. In recent years, various antivirals targeting specific viral proteins of HCV known as direct acting antivirals (DAAs) have been developed, achieving SVR in about 90% cases [43]. However, treatment cost and failure in about 5–10% cases makes it necessary to look for new drugs against HCV [9].

Since traditional drug discovery is complex and time consuming, computational interventions assist in speeding up this process. For this, different in-silico techniques like CSAR, QSAR and machine learning based methods have been used in the development of desirable predictive models. Further, drug repurposing is being used as an alternative to look for new drugs [44]. Many predictive algorithms have been developed to help in antiviral drug discovery. For instance, AVCpred [13] and Antiflavi [16] used a QSAR based approach to predict inhibitors for many viruses including HCV.





**Fig. 7.** An illustration of molecular interactions of proteins NS3, NS3/4A, NS5A, NS5B binding with respective ligand molecules in two dimensions form depicting NS3 with ligands (A) Naftifine (B) Butalbital (C) Proxibarbal; NS3/4A with ligands (D) Vinorelbine (E) Epicriptine (F) Drospirenone; NS5A with ligands (G) Pipecuronium (H) Trimethaphan (I) Cefiderocol and NS5B with ligands (J) Olodaterol (K) Vemurafenib (L) Grepafloxacin.

In the “*Anti-HCV*”, we used the experimentally validated compound with their activity against different NS proteins namely NS3, NS3/4A, NS5A and NS5B to develop predictive models. We employed the four MLTs viz, Support Vector Machine (SVM), Random Forest (RF), *k*-Nearest Neighbour (*k*NN) and Artificial Neural Network (ANN) for predictive algorithms development. We investigated 17,968 molecular descriptors (1D, 2D and 3D) and fingerprints of the inhibitors and used a robust recursive feature elimination method for feature selection. Robust predictive models were developed and assessed by different methods like applicability domain, independent validation, decoy datasets validation and chemical diversity analysis. The best performing models were found to be developed from SVM-SVR methods with PCC values of 0.86, 0.92, 0.88 and 0.84 on training/testing data and 0.92, 0.96, 0.86, and 0.84 on independent validation datasets for NS3, NS3/4A, NS5A and NS5B respectively. Likewise, for RF-SVR we achieved PCC values of 0.80, 0.88, 0.83 and 0.83 on training/testing data and 0.81, 0.90, 0.81, and 0.81 on independent validation datasets for above proteins. Similarly, for *k*NN-SVR we obtained PCC values of 0.80, 0.88, 0.85 and 0.82 on training/testing data and 0.88, 0.90, 0.83, and 0.83 on independent validation datasets for respective proteins. Moreover, for ANN-SVR we obtained PCC values of 0.85, 0.89, 0.87 and 0.81 on training/testing data and 0.90, 0.93, 0.84, and 0.84 on independent validation datasets for NS3, NS3/4A, NS5A and NS5B respectively.

A few regression based studies have been reported to identify inhibitors against HCV NS3 drug target [45,19]. Lafridi et al. 2022 used the QSAR based Multiple Linear Regression (MLR) approach to study the interaction between macrocyclic inhibitors and NS3 protease. They showed multiple correlation coefficient ( $R^2$ ) of 0.84 [45]. da Cunha et al., 2013 developed QSAR based models using 93 boceprevir analogs achieving the  $R^2$  of 0.66 for NS3 protease. They also performed molecular docking of promising analogs for binding affinity to NS3 protease [19]. In addition, we found a classification based method using 413 NS3 protease inhibitors having Matthew's correlation coefficient (MCC) of 0.79 [46]. The predictive “*Anti-HCV*” NS3 models were developed utilizing four MLTs in comparison to limited MLTs. Further, our method showed the PCC values of 0.86 on training/testing with 0.92 for independent validation dataset. Thus, “*Anti-HCV*” NS3 regression based algorithm using multi MLTs is performing equal or better than the pre-existing methods.

Similarly, a few regression based methods reported to target HCV NS3/4A protein for drug designing. Qin et al., 2017 used MLR and SVM methods to develop QSAR based predictive models for NS3/4A inhibitors. They achieved  $R^2$  values of 0.75 to 0.87 for training data and 0.72 to 0.85 on testing data [47]. Alqahtani et al., 2021 developed QSAR models using CORAL software for HCV NS3/4A inhibitors prediction employing the ideality of correlation method, achieving 0.86 to 0.88 of coefficient of determination [48]. In addition, QSAR and pharmacophore based in-silico studies have also been conducted. We found a study which used pharmacophore mapping based approach developing seven featured pharmacophores to identify HCV NS3/4A inhibitors [49]. Venkatesan et al., 2018 have developed pharmacophore features based predictive models using the PHASE module of Schrodinger suite and screened 197 HCV inhibitors for their activity against NS3/4A protein [20]. Whereas, “*Anti-HCV*” NS3/4A regression based algorithms showed better performance with PCC of 0.92 on training/testing with 0.96 on independent validation dataset.

“*Anti-HCV*” NS5A regression based algorithms achieved PCC of 0.88 from training/testing with 0.80 for independent validation dataset. Since we could not find any MLTs based method to design inhibitors targeting HCV NS5A protein. Therefore, this is the first *in-silico* study which incorporated HCV NS5A as a target to identify antivirals against HCV.

NS5B protein, being RNA-dependent RNA polymerase, has also been targeted for drug identification against HCV. A few regression based methods have been developed to predict inhibitors against NS5B protein. For instance, Wang et al., 2014 developed QSAR based predictive models using SVM and MLR approaches using 333 NS5B inhibitors, obtaining correlation coefficient of 0.91 [50]. Similarly, Z. Wang et al., 2020 used comparative molecular field and similarity indices analysis for NS5B inhibitors identification and achieved 0.74 to 0.91 correlation coefficient [51]. In addition, a few classification based methods are also available. HCVpred developed classification structure–activity relationship (CSAR) based models using the set of 578 HCV NS5B inhibitors. They achieved Matthew's correlation coefficient (MCC) of 0.7 to 0.8 [18]. StackHCV is a web server which employs MLTs using 124 active and 124 inactive compounds to develop a predictor for NS5B inhibitors [21]. “*Anti-HCV*” NS5B regression based algorithms are developed using the largest dataset (1671) employing four MLTs, having better performance with PCC values of 0.85 on training/testing and 0.84 for independent validation dataset.

Simultaneously, the robustness of the developed “*Anti-HCV*” models were assessed by applicability domain analysis by calculating the leverage threshold and plotting William's plots. We also plotted the actual  $pIC_{50}$  with predicted  $pIC_{50}$  of the model datasets for checking robustness of developed models and found that developed models are highly robust and reliable. The reliability and robustness of the developed “*Anti-HCV*” models was also validated by comparing the inhibitory activity of ‘inactive’ decoy molecule with the inhibitory activity of the corresponding active molecules for each HCV NS protein. We observed PCC values of  $-0.074$ ,  $-0.037$ ,  $0.052$ , and  $0.117$  for HCV NS3, NS3/4A, NS5A, and NS5B protein decoy datasets respectively suggesting the robustness of our developed predictive models. Moreover, chemical diversity of the compounds used for model development was also assessed. Chemical analysis using binning clustering based on Tanimoto coefficient ( $T_c$ ) with similarity index of 0.6 for each NS protein inhibitor was carried out. The compounds were found to be highly diverse clustering into many clusters, indicating the chemical space of the model developed to be quite large. The multidimensional scaling (MDS) based chemical clustering which uses classical MDS ‘cmdscale’ function implemented in R showed the dispersion of chemical compounds in 2D and 3D chemical space. The MDS plots showed that compounds used in model development are dispersed across the chemical space indicating the diverse nature of compounds.

Potential repurposed drug candidates for each target NS protein - NS3, NS3/4A, NS5A and NS5B are also predicted along with their  $pIC_{50}$  using the developed respective predictive models. A number of predicted repurposed drugs for different NS proteins are also found to be already experimentally validated for HCV in different studies. This showed the reliability and efficacy of our developed models in “*Anti-HCV*”. From the NS3 predicted inhibitors, terbinafine, an antifungal drug was used against HCV in a case report (predicted  $pIC_{50}$  = 8.50) [52]. candesartan cilexetil, an angiotensin receptor blocker used for treatment of hypertension and diabetic neuropathy is used for treatment of HCV patients with lichen planus condition (predicted  $pIC_{50}$  = 9.77) [53].

Among the NS3/4A predicted inhibitors from the study, vinorelbine (predicted  $pIC_{50}$  = 12.01) is used in phase II study for lymphoma regimen including HCV infected patients [54]. temsirolimus (predicted  $pIC_{50}$  = 11.06), an antineoplastic agent is reported to be in phase II trial for the advanced hepatocellular carcinoma treatment [55]. Similarly, from NS5A predicted drugs, rimantadine (predicted  $pIC_{50}$  = 11.06) a RNA synthesis inhibitor, is experimentally tested as HCV viroporin inhibitor [56]. Vildagliptin (predicted  $pIC_{50}$  = 11.06) a dipeptidyl peptidase 4 inhibitor, is experimentally validated for its activity against hepatocellular carcinoma progres-

sion [57]. For NS5B predicted inhibitors, cabozantinib (predicted pIC<sub>50</sub> = 7.15) and regorafenib (predicted pIC<sub>50</sub> = 6.97) have cleared the phase III clinical trial for the treatment of hepatocellular carcinoma [58].

In addition to the activity of predicted repurposing drug candidates against HCV, several predicted drugs were also found to show their antiviral activity against other viruses. From NS3 predicted inhibitors, Hou, H.Y., et al., 2016 looked for the antiviral activity of idarubicin (DB01177) against Enterovirus replication. It showed EC<sub>50</sub> of 0.493 μM of idarubicin against Enterovirus 71 strain [59]. Abrams, R.P.M., et al., 2020 identified methacycline (DB00931) as Zika protease inhibitor in addition to other four inhibitors [60]. Among NS3/4A inhibitors, demeclocycline (DB00618) showed inhibition of West Nile virus (WNV) replication and WNV-induced apoptosis [61]. Daptomycin (DB00080) was validated for its activity against Zika virus [62]. Kato, F., et al., 2019 showed anti-Dengue virus activity of bromocriptine (DB01200) using luciferase assay [63].

From NS5A predicted inhibitors, Zhou, N., et al., 2016 revealed that telavancin (DB06402) and Oritavancin (DB04911) blocks the entry of Ebola virus, Middle East Respiratory Syndrome Coronavirus (MERS-CoV), and Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) [64]. Paromomycin (DB01421) showed capsid protease inhibitory activity against Chikungunya virus (CHIKV). It showed EC<sub>50</sub> of 22.91 μM against CHIKV [65]. Similarly, among the NS5B predicted inhibitors, micafungin (DB01141) was revealed to inhibit Dengue virus infection by disruption of virus binding, entry and stability [66]. Another antifungal drug, anidulafungin (DB00362) also shows inhibitory activity against Zika virus [67] and SARS-CoV-2 [68]. Mefloquine (DB00358) showed antiviral activity against Human polyomavirus 2, also called JC virus [69]. It also inhibits SARS-CoV-2 infection with EC<sub>50</sub> of 1.2 μM [70].

Selected predicted repurposed drugs were also validated for their interactions with the respective NS proteins through molecular docking approach. Upon docking potential predicted inhibitors of NS3, plecanatide (DB13170), naftifine (DB00735), and butalbital (DB00241) showed comparable binding affinity of −6.3, −7.8, and −6.2 Kcal/mol respectively for NS3 protein as compared to approved NS3 inhibitor, asunaprevir (−7.4 Kcal/mol) [71–72]. Similarly, For NS3/4A, three docked ligands, vinorelbine (DB00361), epicutripte (DB11275), and drospirenone (DB01395) have lower binding energy of −8.9, −8.4, and −8.1 Kcal/mol respectively. These binding energies also correspond to the binding energy (−7.3 Kcal/mol) of telaprevir, an approved HCV NS3/4a inhibitor [73]. Likewise, for predicted inhibitors for NS5A, the binding affinity of −9.8, −9.4, and −9.2 Kcal/mol for pipecuronium (DB01338), trimethaphan (DB01116) and cefiderocol (DB14879) respectively was observed. These affinities are also close to the approved NS5A inhibitor, daclatasvir [74]. For the NS5B predicted repurposed drugs, oladaterol (DB09080), vemurafenib (DB08881), and grepafloxacin (DB04876) have quite lower binding energies of −8.8, −8.7, and −8.4 Kcal/mol respectively. Even though not much comparable with binding energy of −14.3 Kcal/mol for approved NS5B inhibitory drug, sofosbuvir [75]. This shows that the docked molecules are effectively interacting with the target proteins which confirms the efficiency of the developed machine learning models in our study. Thus, these models can be very helpful in predicting the activity of inhibitors against important targets for HCV i.e. NS3, NS3/4A, NS5A and NS5B proteins.

## 5. Conclusion

In this study, we have developed “Anti-HCV” QSAR regression based algorithms utilizing MLTs Support Vector Machine (SVM), Random Forest (RF), *k*-Nearest Neighbour (*k*NN) and Artificial Neu-

ral Network (ANN). In this method, predictive models were developed to identify inhibitors against important drug targets in HCV namely NS3, NS3/4A, NS5A and NS5B. These predictive models performed well with PCC of 0.85 to 0.92 on training/testing and 0.84 to 0.92 on independent validation datasets. Applicability domain, decoy datasets and chemical diversity analysis suggest these methods to be robust and reliable. We also scanned the “DrugBank” database to identify potential repurposing drug candidates. Selective repurposed molecules were also shown to be effective by molecular docking technique. Thus, it will help in easy and fast development of new antivirals against HCV targeting non-structural proteins.

## 6. Code availability

All the codes used for developing the predictive models for anti-HCV are available at GitHub (<https://github.com/manojk-imtech/anti-HCV>).

## Authors contribution

Study conceptualization, design, analysis and supervision was done by MK; Data curation, descriptor calculations, features selection carried out by SK, AT; Machine learning, analysis, drug repurposing were performed by ARaj, SK; Molecular docking and analysis was done by ARas; and manuscript is written by SK, ARas, MK.

## CRediT authorship contribution statement

**Sakshi Kamboj:** Data curation, Methodology, Software, Formal analysis, Validation, Visualization, Writing – original draft, Writing – review & editing. **Akanksha Rajput:** Methodology, Software, Formal analysis, Validation, Visualization, Writing – original draft. **Amber Rastogi:** Methodology, Formal analysis, Validation, Visualization, Writing – original draft. **Anamika Thakur:** Data curation, Methodology, Formal analysis, Validation, Writing – original draft. **Manoj Kumar:** Conceptualization, Supervision, Formal analysis, Software, Investigation, Funding acquisition, Project administration, Writing – original draft, Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We acknowledge the funding support from CSIR-Institute of Microbial Technology (STS-038).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2022.06.060>.

## References

- [1] Li HC, Lo SY. Hepatitis C virus: Virology, diagnosis and treatment. *World J Hepatol* 2015;7(10):1377–89.
- [2] Webster DP, Klenerman P, Dusheiko GM. Hepatitis C. *Lancet* 2015;385(9973):1124–35.
- [3] Meshram RJ, Kathwate GH, Gacche RN. Progress, evolving therapeutic/diagnostic approaches, and challenges in the management of hepatitis C virus infections. *Arch Virol* 2022;167(3):717–36.



- [4] Alzahrani N et al. Delayed by Design: Role of Suboptimal Signal Peptidase Processing of Viral Structural Protein Precursors in Flaviviridae Virus Assembly. *Viruses* 2020;12:10.
- [5] Ahmed A, Felmlee DJ. Mechanisms of Hepatitis C Viral Resistance to Direct Acting Antivirals. *Viruses* 2015;7(12):6716–29.
- [6] Halfon P, Locarnini S. Hepatitis C virus resistance to protease inhibitors. *J Hepatol* 2011;55(1):192–206.
- [7] Jensen SB et al. Evolutionary Pathways to Persistence of Highly Fit and Resistant Hepatitis C Virus Protease Inhibitor Escape Variants. *Arch Virol* 2019;70(3):771–87.
- [8] Vermehren J et al. Challenges and perspectives of direct antivirals for the treatment of hepatitis C virus infection. *Hepatology* 2018;69(5):1178–87.
- [9] Pawlotsky JM. Treatment failure and resistance with direct-acting antiviral drugs against hepatitis C virus. *Hepatology* 2011;53(5):1742–51.
- [10] Shokri A et al. Flexible Acyclic Polyol-Chloride Anion Complexes and Their Characterization by Photoelectron Spectroscopy and Variable Temperature Binding Constant Determinations. *J Phys Chem A* 2016;120(10):1661–8.
- [11] He S et al. Repurposing of the antihistamine chlorcyclizine and related compounds for treatment of hepatitis C virus infection. *Sci Transl Med* 2015;7(282):p. 282ra49.
- [12] Perin PM et al. Flunarizine prevents hepatitis C virus membrane fusion in a genotype-dependent manner by targeting the potential fusion peptide within E1. *Hepatology* 2016;63(1):49–62.
- [13] Qureshi A, Kaur G, Kumar M. AVCpred: an integrated web server for prediction and design of antiviral compounds. *Chem Biol Drug Des* 2017;89(1):74–83.
- [14] Thakur N, Qureshi A, Kumar M. AVppred: collection and prediction of highly effective antiviral peptides. *Nucleic Acids Res* 2012;40(W1):W199–204.
- [15] Qureshi A et al. HIVprot: an integrated web based platform for prediction and design of HIV proteins inhibitors. *J Cheminform* 2018;10(1):12.
- [16] Rajput A, Kumar M. Anti-flavi: A Web Platform to Predict Inhibitors of Flaviviruses Using QSAR and Peptidomimetic Approaches. *J Comput Chem* 2018;9:3121.
- [17] Rajput A et al. Prediction of repurposed drugs for Coronaviruses using artificial intelligence and machine learning. *J Comput Aided Mol Des* 2021;19:3133–48.
- [18] Malik, A.A., et al., HCVpred: A web server for predicting the bioactivity of hepatitis C virus NS5B inhibitors. *2020*. **41**(20): p. 1820–1834.
- [19] da Cunha EF, Matos KS, Ramalho TC. QSAR and docking studies of HCV NS3 serine protease inhibitors. *Med Chem* 2013;9(6):774–805.
- [20] Venkatesan, A., et al., Pharmacophore feature prediction and molecular docking approach to identify novel anti-HCV protease inhibitors. *2018*. **119**(1): p. 960–966.
- [21] Malik, A.A., et al., StackHCV: a web-based integrative machine-learning framework for large-scale identification of hepatitis C virus NS5B inhibitors. *2021*. **35**(10): p. 1037–1053.
- [22] Gaulton A et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res* 2012;40(Database issue):D1100–7.
- [23] O'Boyle NM et al. Open Babel: An open chemical toolbox. *J Cheminform* 2011;3:33.
- [24] Yap CW. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem* 2011;32(7):1466–74.
- [25] Lin X et al. A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables and mutual information. *J Chromatogr B Analyt Technol Biomed Life Sci* 2012;910:149–55.
- [26] Gholami B et al. Recursive feature elimination for brain tumor classification using desorption electrospray ionization mass spectrometry imaging. *Annu Int Conf IEEE Eng Med Biol Soc* 2012;2012:5258–61.
- [27] Subgupta J, Alzbutas R. Prediction and Risk Assessment Models for Subarachnoid Hemorrhage: A Systematic Review on Case Studies. *Biomed Res Int* 2022;2022:5416726.
- [28] Kar S, Roy K, Leszczynski J. Applicability Domain: A Step Toward Confident Predictions and Decidability for QSAR Modeling. *Methods Mol Biol* 2018;1800:141–69.
- [29] Cereto-Massagué A et al. DecoyFinder: an easy-to-use python GUI application for building target-specific decoy sets. *Bioinformatics* 2012;28(12):1661–2.
- [30] Irwin JJ et al. ZINC20-A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *J Chem Inf Model* 2020;60(12):6065–73.
- [31] Backman, T.W., Y. Cao, and T. Girke, *ChemMine tools: an online service for analyzing and clustering small molecules*. *Nucleic Acids Res*, 2011. **39**(Web Server issue): p. W486–91.
- [32] Wishart DS et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;46(D1):D1074–82.
- [33] Morris GM et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* 2009;30(16):2785–91.
- [34] Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 2010;31(2):455–61.
- [35] El-Serag HB. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology* 2012;142(6):1264–1273.e1.
- [36] Wang W et al. Progress in evaluating the status of hepatitis C infection based on the functional changes of hepatic stellate cells (Review). *Mol Med Res* 2020;22(5):4116–24.
- [37] Alberti A, Chemello L, Benvegñù L. Natural history of hepatitis C. *J Hepatol* 1999;31(Suppl 1):17–24.
- [38] Hoonagale JH. Course and outcome of hepatitis C. *Hepatology* 2002;36(5 Suppl 1):S21–9.
- [39] Takamizawa A et al. Structure and organization of the hepatitis C virus genome isolated from human carriers. *J Virol* 1991;65(3):1105–13.
- [40] Domingo E, Gomez J. Quasispecies and its impact on viral hepatitis. *Virus Res* 2007;127(2):131–50.
- [41] Franco S et al. Genetic and catalytic efficiency structure of an HCV protease quasispecies. *Hepatology* 2007;45(4):899–910.
- [42] Fried MW et al. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* 2002;347(13):975–82.
- [43] Pawlotsky JM. New hepatitis C therapies: the toolbox, strategies, and challenges. *Gastroenterology* 2014;146(5):1176–92.
- [44] Rajput A et al. Computational identification of repurposed drugs against viruses causing epidemics and pandemics via drug-target network analysis. *Comput Biol Med* 2021;136:104677.
- [45] Lafridi H, Almalki FA. In silico evaluation of molecular interactions between macrocyclic inhibitors with the HCV NS3 protease. In: Docking and identification of antiviral pharmacophore site. p. 1–14.
- [46] Wang M et al. Classification models of HCV NS3 protease inhibitors based on support vector machine (SVM). *Comb Chem High Throughput Screen* 2015;18(1):24–32.
- [47] Qin Z, Wang M, Yan A. QSAR studies of the bioactivity of hepatitis C virus (HCV) NS3/4A protease inhibitors by multiple linear regression (MLR) and support vector machine (SVM). *J Cell Biochem* 2017;27(13):2931–8.
- [48] Alqahtani AM et al. The index of ideality of correlation: QSAR studies of hepatitis C virus NS3/4A protease inhibitors using SMILES descriptors. *J Biomol Struct Dyn* 2021;32(6):495–520.
- [49] Wadood A et al. In silico identification and evaluation of leads for the simultaneous inhibition of protease and helicase activities of HCV NS3/4A protease using complex based pharmacophore mapping and virtual screening. *PLoS One* 2014;9(2):e89109.
- [50] Wang M et al. Quantitative structure and bioactivity relationship study on HCV NS5B polymerase inhibitors. *SAR QSAR Environ Res* 2014;25(1):1–15.
- [51] Wang Z et al. Combined 3D-QSAR, molecular docking and molecular dynamics study on the benzimidazole inhibitors targeting HCV NS5B polymerase. *J Comput Chem* 2020;38(4):1071–82.
- [52] Carducci M et al. Erythema multiforme during cytomegalovirus infection and oral therapy with terbinafine: a virus-drug interaction. *J Eur Acad Dermatol Venereol* 2004;18(2):201–3.
- [53] Katsuta M, Asahina A, Shiohara T. Multiple Fixed Drug Eruption Mimicking Parapsoriasis en Plaqué in a Patient with Hepatitis C Virus Infection. *Molecules* 2020;12(1):25–32.
- [54] Mizoroki F et al. A phase II study of VEPA/FEPP chemotherapy for aggressive lymphoma in elderly patients: Japan Clinical Oncology Group Study JCOG9203. *Int J Hematol* 2006;83(1):55–62.
- [55] Kelley RK et al. Phase II Trial of the Combination of Temsirolimus and Sorafenib in Advanced Hepatocellular Carcinoma with Tumor Mutation Profiling. *Liver Cancer* 2021;10(6):561–71.
- [56] Wei S et al. Inhibitor Development against p7 Channel in Hepatitis C. *Virus* 2021;26(5).
- [57] Nishina S et al. Dipeptidyl Peptidase 4 Inhibitors Reduce Hepatocellular Carcinoma by Activating Lymphocyte Chemotaxis in Mice. *Cell Mol Gastroenterol Hepatol* 2019;7(1):115–34.
- [58] Llovet JM, Kelley RK, Villanueva A. Hepatocellular carcinoma 2021;7(1):6.
- [59] Hou HY et al. Idarubicin is a broad-spectrum enterovirus replication inhibitor that selectively targets the virus internal ribosomal entry site. *J Gen Virol* 2016;97(5):1122–33.
- [60] Abrams RPM et al. Therapeutic candidates for the Zika virus identified by a high-throughput screen for Zika protease inhibitors. *Proc Natl Acad Sci U S A* 2020;117(49):31365–75.
- [61] Michaelis M et al. Minocycline inhibits West Nile virus replication and apoptosis in human neuronal cells. *J Antimicrob Chemother* 2007;60(5):981–6.
- [62] Barrows NJ et al. A Screen of FDA-Approved Drugs for Inhibitors of Zika Virus Infection. *Cell Host Microbe* 2016;20(2):259–70.
- [63] Kato F et al. Identification of inhibitors of dengue viral replication using replicon cells expressing secretory luciferase. *Antiviral Res* 2019;172:104643.
- [64] Zhou N et al. Glycopeptide Antibiotics Potently Inhibit Cathepsin L in the Late Endosome/Lysosome and Block the Entry of Ebola Virus, Middle East Respiratory Syndrome Coronavirus (MERS-CoV), and Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV). *J Biol Chem* 2016;291(17):9218–32.
- [65] Fatma B et al. Alphavirus capsid protease inhibitors as potential antiviral agents for Chikungunya infection. *Antiviral Res* 2020;179:104808.
- [66] Chen YC et al. Micafungin Inhibits Dengue Virus Infection through the Disruption of Virus Binding, Entry, and Stability. *Pharmaceuticals (Basel)* 2021;14(4).
- [67] Lu JW et al. Synergistic in-vitro antiviral effects of combination treatment using anidulafungin and T-1105 against Zika virus infection. *Antiviral Res* 2021;195:105188.
- [68] Ahamad S et al. Anti-Fungal Drug Anidulafungin Inhibits SARS-CoV-2 Spike-Induced Syncytia Formation by Targeting ACE2-Spike Protein Interaction. *Front Genet* 2022;13:866474.
- [69] Brickelmaier M et al. Identification and characterization of mefloquine efficacy against JC virus in vitro. *Antimicrob Agents Chemother* 2009;53(5):1840–9.
- [70] Sacramento CQ et al. Unlike Chloroquine, Mefloquine Inhibits SARS-CoV-2 Infection in Physiologically Relevant Cells. *Viruses* 2022;14(2).



- [71] Chandramohan V et al. Evaluating Andrographolide as a Potent Inhibitor of NS3-4A Protease and Its Drug-Resistant Mutants Using In Silico Approaches. *Adv Virol* 2015;2015:972067.
- [72] Patil VS, Harish DR. Hepatitis C Virus NS3/4A Inhibition and Host Immunomodulation by Tannins from *Terminalia chebula*: A Structural Perspective 2022;27(3).
- [73] Hegde HV, Sun X. Conformational study on telaprevir by HPLC-DAD-MS and theoretical calculation. *Molecules* 2019;33(11):e4644.
- [74] Barakat KH et al. A Refined Model of the HCV NS5A protein bound to daclatasvir explains drug-resistant mutations and activity against divergent genotypes. *J Chem Inf Model* 2015;55(2):362–73.
- [75] Khalid H, Ashfaq UA. Molecular Docking and Pharmacoinformatics Studies Reveal Potential Phytochemicals Against HCV NS5B Polymerase. *Comb Chem High Throughput Screen* 2022;25(2):335–46.