

RESEARCH

Open Access



# Computational insight to design new potential hepatitis C virus NS5B polymerase inhibitors with drug-likeness and pharmacokinetic ADMET parameters predictions

Stephen Ejeh<sup>\*</sup> , Adamu Uzairu, Gideon A. Shallangwa and Stephen E. Abechi

## Abstract

**Background:** Hepatitis C virus (HCV) is considered a worldwide health problem since it affects over 3% of the population and causes 300,000 fatalities per year. Chronic infection causes high morbidity and mortality in patients, leading to liver cirrhosis, hepatocellular carcinoma, fibrosis, liver cancer, and other HCV-related illnesses. Finding novel and better HCV treatments is a top international health goal right now. As a result, the pressing need for new HCV antiviral drugs has fueled research into the structural requirements of NS5B polymerase inhibitors at a molecular basis.

**Results:** In this study, an in silico technique was applied to study 79 compounds with HCV inhibitory bioactivity, with the best statistical results ( $R^2 = 0.7051$ ,  $Q^2 = 0.6455$ ,  $R^2_{\text{pred}} = 0.6992$ ,  $^cR^2_r = 0.6570$ ,  $\text{SEE} = 0.2694$ ).

**Conclusions:** This QSAR investigation allowed the research team to evaluate the influence of straightforward descriptors, shedding insight into the critical elements that guide the design of innovative effective molecules. Most of the innovative effective molecules exhibited better binding affinity ( $-195.6$  kcal/mol) than dasabuvir the reference drug ( $-171.0$  kcal/mol) with the target receptor (hepatitis C virus NS5B RNA polymerase). ADMET prediction disclosed enhanced pharmacokinetic properties with lower MRTD (maximum tolerated dose) of some new derivatives.

**Keywords:** Ligand-based design, HCV NS5B polymerase, Pharmacokinetic profiles

## Background

Hepatitis C Virus (HCV) is considered a worldwide health problem since it affects over 3% of the population and causes 300,000 fatalities per year [1]. Chronic infection causes high morbidity and mortality in patients, leading to liver cirrhosis, hepatocellular carcinoma, fibrosis, liver cancer, and other HCV-related illnesses [2]. HCV is an encapsulated single-strand linear positively oriented RNA virus belonging to the species Hepacivirus,

which belongs to the Flaviviridae family. It has six genotypes and various subtypes, each with a different geographic distribution [1].

The host ribosomes translate HCV-RNA into a poly-protein, which is then processed by host and viral proteases into ten proteins: three structural proteins (core, E1, and E2), an ion channel (p7), and six non-structural proteins (NS2, NS3, NS4A, NS4B, NS5A, and NS5B) [2]. The virus nucleocapsid is made up of structural proteins, while non-structural proteins are involved in virus development, maturation, and reproduction. The HCV NS5B protein is a multifunctional RNA binding protein that is critical for virus duplication. NS5B protein inhibition

\*Correspondence: ejehstephen@gmail.com  
Department of Chemistry, Ahmadu Bello University, P.M.B. 1044, Zaria, Kaduna, Nigeria

has been shown in multiple studies to successfully suppress viral replication [3]. As a result, the NS5B protein is a well-established therapeutic target for which several medicines have been licensed and released.

Anti-HCV vaccination is currently unavailable; nonetheless, in the 1990s, interferon-alpha was used to treat HCV, and in the next decade, pegylated interferon-alpha in a mixture with ribavirin was used to treat HCV. Interferon-alpha inhibitors, DNA and RNA polymerase inhibitors, NS3/4A RNA protease inhibitors, NS5 RNA serine protease inhibitors, and NS5B RNA polymerase inhibitors have all been licensed for use in clinical trials since 2011 [1, 4].

Current HCV medication has a low continued viral success rate, quick onset of drug resistance, especially in genotype 1 HCV patients, and considerable adverse effects, all of which lead to therapy cessation [5]. As a result, better anti-HCV medications with fewer problems are urgently needed. Finding the most essential molecular descriptors can aid in the discovery of new biologically active chemicals that can be used as anti-HCV medications and provide clinically successful treatment [6].

With advances in computational algorithms and simulation software, computer drug design has become widely used for drug discovery and development because of the advantages of being less time-consuming, cost-effective, and high effectiveness in silico screening and prediction of candidate medications [7]. Throughout the previous few decades, QSAR procedures have been developed and widely employed in a diversity of sectors, including chemical/biological, chemistry, and associated fields [4]. In this study, a QSAR model was proposed based on the structure of drugs having significant NS5B polymerase inhibitor bioactivity, and the model was externally validated to evaluate its genuine predictive capabilities. Based on the good efficacy of QSAR and molecular docking, the proposed model was efficient in understanding the relationship between structure and bioactivity and making suitable structure modifications for more promising suppression. Drug-likeness and ADMET pharmacokinetic characteristics were also tested on the novel compounds.

## Methods

### Dataset

79 compounds of HCV NS5B polymerase inhibitors were employed, which were sourced from online resources (<https://pubchem.ncbi.nlm.nih.gov/>). The EC<sub>50</sub> (μM) values of the identified compounds' bioactivity were provided. The bioactivity in EC<sub>50</sub> was transformed to Log (1/EC<sub>50</sub>), which gives pEC<sub>50</sub> [8]. Additional file 1: Table S1 in the Additional file 1 contains the datasets.

### Optimization and descriptors calculation

The optimization was carried out using the quantum mechanical approach in Spartan 14 and the density function theory B3LYP/631G\*\* [9]. PaDEL-Descriptors were used to create roughly 2000 molecular descriptors for every species using the preprocessing molecules.

### Dataset treatment, division, and Selection of optimal descriptors

The Drug Theoretics and Cheminformatics Laboratory software (DTC-QSAR v1.0.5) freely available to the public at <http://dtclab.webs.com/software-tools> was used in performing the dataset pre-treatment, division, and selection of optimal descriptors for Model building [10]. The total dataset was split into two parts: 70% model building data and 30% model test data [8]. The DTC-QSAR tool v1.0.5's Genetic Function Algorithm (GFA) was used to choose a set of descriptors that best explained the diversity in bioactivity values of the investigated compounds. GFA is a heuristic search approach for locating exact or approximate answers to optimization and search issues. It has the advantage of constructing many models, use the lack of fit function to avoid overfitting, and allowing the user to regulate the length of the equation [11, 12].

### Model applicability domain (AD)

The length of projection methodology was used to set up the AD for QSAR models in the research, which is established on leverage principles of the research compounds and standardized residual (SR) of the model. The horizontal component of the hat state vector was used to create compound leverages ( $h_i$ ) as given in Eq. (1):

$$h_i = X(X^T X)^{-1} X^T \quad (1)$$

$X^T$  is the transpose of  $X$ , and  $X$  is the descriptor array. SR is calculated using the model's error function with Eq. (2)

$$SR = \frac{\bar{p} - p}{\sqrt{\sum_{i=1}^n \frac{(\bar{p} - p)^2}{n}}} \quad (2)$$

$p$  represents the actual bioactivity value for the trained or test data,  $\bar{p}$  represents the model's estimated bioactivity value and  $n$  stands for the number of molecules in the trained or test data.

The borders  $> -3$ ,  $< 3$  and  $< h^*$  were used to design the AD models, where  $h^*$  stands for caution leverage, which is calculated using Eq. (3)

$$h^* = \frac{3(t + 1)}{n} \quad (3)$$

$t$  is the proportion of structural properties in the model and  $n$  has the same meaning as in Eq. (2) [13].

### Design of new chemical entities

New chemical entities are designed by a careful selection of template molecule (*T*) with high bioactivity, low standardized residual and within the model AD from the dataset and making modifications based on the description of the descriptor selected for optimal model (see Additional file 1: Table S3).

### Accessing the design entities

To determine the reliability of these new design entities, they were evaluated using the online tools SwissADME ([www.swissadme.ch/](http://www.swissadme.ch/)) and pkCSM (<http://structure.bioc.cam.ac.uk/pkcsml>) to estimate the drug-likeness and pharmacokinetics ADMET profile.

### Docking process

The target receptor was accessed from a protein data bank deposited in 2013 by Schoenfeld and co-worker [14] with the following information: PDB ID of 4MKA, the resolution of 2.05 Å, and the *R* value free of 0.217. The co-crystallized molecule was removed after which

the target receptor was imported into the Molegro Virtual Docker tools (MVD) and the Preparation tab was selected on the dialog and Assign All was set to Always. This ensures that all preparation will be done by MVD. The detailed docking process was applied as described in our previous work [13]. The target receptor is presented in Fig. 1

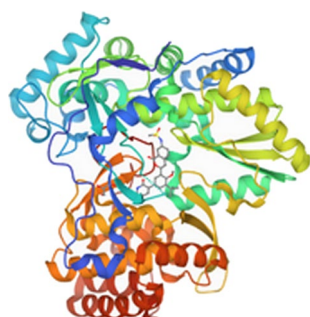
### Results

Three mathematical models relating the bioactivity of hepatitis C virus NS5B RNA polymerase inhibitors with its structure were built in the present research, and the QSAR models are presented in Table 1 with their coefficient of determination ( $R^2$ ), standard error of estimate (SEE), and mean average error (MAE) values.

The  $R^2$  and the error measurement (SEE and MAE) were used to assess the performance of QSAR models, resulting in model 2 being chosen for further investigation. The better the model, the higher the  $R^2$  scores or the lower the error values. In addition, the Tropsha parameters [8] were utilized to assess the top models' reliability. All the statistical tools and necessary predictive capability as a benchmark for a good QSAR model are listed in Table 2.

The descriptors included in the selected model, as well as their descriptions, are listed in the Additional file 1: Table S2, and the numerical values of the descriptor used in the selected model are presented in Additional file 1: Table S3, while Table 3 presents observed  $EC_{50}$ ,  $pEC_{50}$ , predicted  $pEC_{50}$ , and residual of the entire datasets.

William's plots were generated and are presented in Fig. 2 for the selected QSAR model, based on the plot of standardized residual (SR) versus leverage values ( $h_i$ ), to get an immediate and straightforward graphical detection of both the *Y* outliers and the *X* outliers of the model, and Fig. 3 depicts a display of estimated vs actual



**PDB ID: 4MKA**

**Fig. 1** Structure of the target receptor (hepatitis C virus NS5B RNA polymerase) in complex with dasabuvir

**Table 1** Developed models and the coefficient of determination ( $R^2$ ), standard error of estimate (SEE), and mean average error (MAE) values

Model	Equation	$R^2$	SEE	MAE
1	$pEC_{50} = 96.89265 - 0.72022 * ATSC5e + 0.0737 * bpol$ $+ 0.00432 * ATSC3i + 0.00949 * MDEC - 22 - 15.60553$ $* SpMax5_{Bhm} - 12.73215 * SpMin6_{Bhv} - 9.24232 * SpMin5_{Bhm}$	0.7044	0.2697	0.1553
2	$pEC_{50} = 2.42246 + 0.06217 * VR2_{Dzi} - 0.587 * GATS6p$ $- 0.24467 * nX + 5.9672 * GGI10 - 0.6542 * ATSC5e$ $+ 3.05767 * GATS2p - 0.14821 * MDEC - 24$	0.7051	0.2694	0.1307
3	$pEC_{50} = -24.0199 - 4.8512 * SpMax5_{Bhv} + 15.7646 * SpMax6_{Bhp}$ $- 0.6268 * ATSC5e + 2.9609 * SpMin8_{Bhs} - 0.0319$ $* AATS7m - 3.9737 * SpMin5_{Bhs} + 0.2919 * AATS8s$	0.7038	0.2700	0.1448

$R^2$ , coefficient of determination; SEE, standard error of estimate; MAE, mean average error

**Table 2** Model parameters, formula, threshold values, and models scores

Parameter	Formula	Threshold	Model score		
			Model 1	Model 2	Model 3
$R^2$	$\frac{[\sum \{(Y_{obs} - \bar{Y}_{obs}) \times (Y_{pred} - \bar{Y}_{pred})\}]^2}{\sum (Y_{obs} - \bar{Y}_{obs})^2 \times \sum (Y_{pred} - \bar{Y}_{pred})^2}$	$R^2 > 0.6$	0.7044	0.7051	0.7038
$R^2_{adj}$	$\frac{(N-1) \times R^2 - P}{N-1-P}$	$R^2_{adj} > 0.6$	0.6604	0.6611	0.6597
$R^2_{pr}$	$1 - \frac{\sum (Y_{exp(Tests)} - Y_{est(Tests)})^2}{\sum (Y_{exp(Tests)} - \bar{Y}_{Training})^2}$	$R^2_{pr} > 0.6$	0.6109	0.6992	0.6552
$Q^2$	$1 - \frac{\sum_{i=1}^n (Y_{exp} - Y_{pred})^2}{\sum_{i=1}^n (Y_{exp} - \bar{Y})^2}$	$Q^2 > 0.6$	0.6217	0.6455	0.6211
$F$	$\frac{\sum (Y_{pred} - \bar{Y}_{obs})^2}{\sum (Y_{obs} - \bar{Y}_{pred})^2} \times \frac{N-P-1}{P}$	$F > 0.33$	.00	16.05	15.95
SEE	Standard Error of Estimate	Smaller the better	0.2697	0.2694	0.2700
MAE	Mean Average Error	Smaller the better	0.1553	0.1307	0.1448
$R_r$	An average of the correlation coefficient for randomized data	$R_r < 0.5$	0.3224	0.3047	0.3712
$R^2_r$	An average of determination coefficient for randomized data	$R^2_r < 0.5$	0.096	0.1036	0.1496
$^cR^2_r$	$R \times \sqrt{R^2 - R^2_r}$	$^cR^2_r > 0.5$	0.6504	0.6570	0.6311

bioactivity over the whole data, revealing a strong link between actual and estimated bioactivity levels in the models.

Based on the proven and verified QSAR model, molecule 16 in dataset Table (Additional file 1: Table S1), pictured in Fig. 4, was utilized as a standard to characterize the molecule. Molecule 16 was utilized as a foundation for developing new molecules because this was chosen cautiously from Fig. 2, recognizing the data point as more active, low SR, and detected inside the existing model's AD.

The bioactivity of the precursor drug, newly created molecules, and reference molecule (dasabuvir) was predictable using the previously proved QSAR model (model 2). All modified variants, as well as dasabuvir, had a greater pEC<sub>50</sub> value than the prototype, according to the data except N5 (see Additional file 1: Table S4).

The interaction of molecules N15, R, and T with the target receptor (hepatitis C virus NS5B RNA polymerase, PDB ID: 4MKA) was visualized using the Discovery Studio software and the Molecular Operating Environment (MOE) software, as shown in Figs. 5, 6, and 7, respectively.

A molecular docking assessment utilizing Molegro Virtual Docker version 6.0 verified that molecule N15 had the most energetically favorable ligand conformation with the MolDock scoring function value of −195.6 kcal/mol which is better than that of the reference drug of −171 kcal/mol. The results of the ligand–receptor interactions of the novel molecule (N15) in Fig. 5, the

reference molecule (R) in Fig. 6, and the template molecule (T) in Fig. 7 are summarized in Table 4.

To establish the viability of these molecules as medications, the research team looked at the drug-like and pharmacokinetic properties (ADMET) of these novel compounds, using dasabuvir as a reference. The ADMET properties given in Table 6 were estimated using the pkCSM software, while the drug-likeness attributes presented in Table 5 were originally assessed using the web tool SwissADME. Drug-likeness characteristics are an important criterion for drug candidates in the early stages of development [15]. The practice of correlating a molecule's physicochemical characteristics with its biopharmaceutical characteristics in the human body, notably its impact on oral bioavailability, is also known as the drug-likeness approach [16].

## Discussion

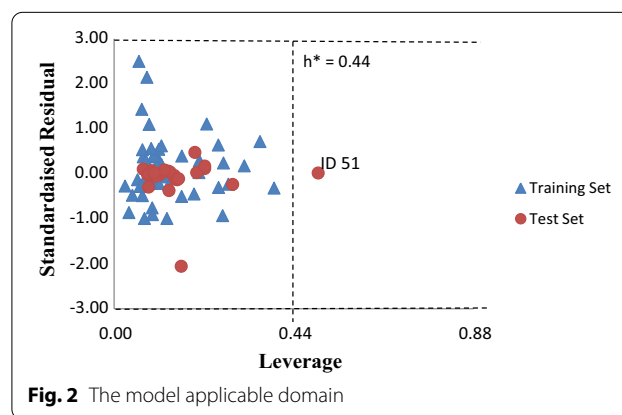
The selected QSAR model is statistically sound because it met the conditions listed in Table 2 as a threshold, and hence, has an appropriate predictive capability. Model 2 met Tropsha's and the Organization for Economic Cooperation and Development's (OECD) requirements [12, 17] as it explains 71% and predicts 70% of the variances of the HCV NS5B polymerase inhibitors with their bioactivity as presented in Table 2. It indicates that the models precisely regressed the data and that it can predict the fitting training set for the model, as it predicted around 70% of the data and so met the minimum criteria of 60% [11]. The result of the y-randomization test demonstrates that

**Table 3** The PubChem compound ID (CID),  $EC_{50}$ ,  $pEC_{50}$ , predicted  $pEC_{50}$ , and residual of the entire datasets

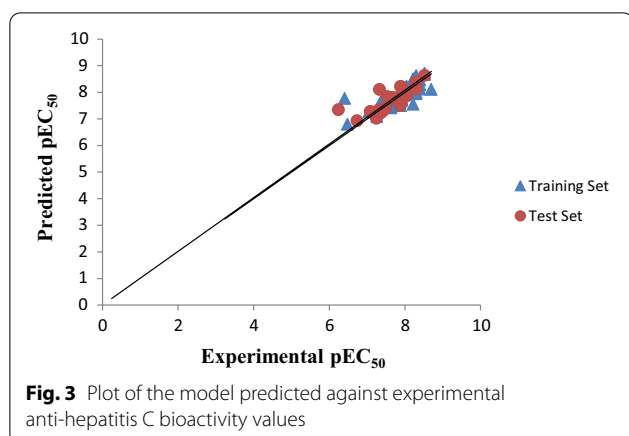
ID	PubChem CID	$EC_{50}$ [ $\mu$ M]	$pEC_{50}$	Predicted $pEC_{50}$	Residual
<i>Training set compounds</i>					
2	118728148	0.020	7.699	7.720	-0.021
3	118730993	0.011	7.959	8.091	-0.132
4	118730994	0.010	8.000	8.036	-0.036
5	118730995	0.013	7.886	7.876	0.010
6	118730996	0.008	8.097	8.174	0.077
7	118730997	0.015	7.824	7.790	0.034
8	118730999	0.009	8.046	8.232	-0.186
9	118731000	0.013	7.886	7.500	0.386
10	118731179	0.020	7.699	7.615	0.084
11	118731180	0.053	7.276	7.348	-0.072
12	118731181	0.022	7.658	7.627	0.031
13	118731182	0.089	7.051	7.263	-0.212
14	118731184	0.012	7.921	7.706	0.215
15	118731185	0.006	8.222	8.511	-0.289
16	118731186	0.002	8.699	8.113	0.586
17	118731187	0.005	8.301	8.291	0.010
19	118731189	0.007	8.155	8.215	-0.06
21	118731191	0.005	8.301	8.327	-0.026
22	118731192	0.003	8.523	8.665	-0.142
23	118731193	0.012	7.921	7.975	-0.054
24	118731194	0.044	7.357	7.628	-0.271
26	118731196	0.006	8.222	7.555	0.667
27	118731197	0.036	7.444	7.697	-0.253
31	118731202	0.004	8.398	8.119	0.279
32	118731203	0.004	8.398	8.460	-0.062
33	118731204	0.004	8.398	8.280	0.118
35	118731206	0.005	8.301	7.944	0.357
36	118731207	0.005	8.301	8.239	0.062
37	118731208	0.005	8.301	8.630	-0.329
38	118731209	0.006	8.222	8.122	0.100
39	118731210	0.016	7.796	7.739	0.057
41	118731212	0.013	7.886	7.743	0.143
42	118731213	0.005	8.301	8.000	0.301
43	118731214	0.018	7.745	7.628	0.117
44	118731216	0.006	8.222	8.215	0.007
46	118731218	0.015	7.824	7.720	0.104
47	118731219	0.012	7.921	7.769	0.152
49	118731221	0.013	7.886	7.807	0.079
50	118731222	0.050	7.301	7.455	-0.154
58	118731232	0.017	7.770	7.689	0.081
61	118731235	0.012	7.921	7.920	0.001
63	118731237	0.016	7.796	7.861	-0.065
64	118731238	0.023	7.638	7.420	0.218
65	118731240	0.003	8.523	8.722	-0.199
66	118731241	0.017	7.770	7.659	0.111
67	118731242	0.050	7.301	7.316	-0.015
68	118731243	0.026	7.585	7.432	0.153
70	118731245	0.055	7.260	7.085	0.175

**Table 3** (continued)

ID	PubChem CID	$EC_{50}$ [ $\mu$ M]	$pEC_{50}$	Predicted $pEC_{50}$	Residual
71	118731246	0.330	6.481	6.802	-0.321
72	118731248	0.027	7.569	7.493	0.076
74	118731250	0.031	7.509	7.636	-0.127
75	118731251	0.030	7.523	7.749	-0.226
76	118731252	0.040	7.398	7.459	-0.061
77	118731376	0.390	6.409	7.774	-1.365
79	118731379	0.014	7.854	7.924	-0.07
<i>Test set compounds</i>					
1	118728143	0.013	7.886	8.223	-0.337
18	118731188	0.005	8.301	8.132	0.169
20	118731190	0.012	7.921	7.829	0.092
25	118731195	0.018	7.745	7.760	-0.015
28	118731198	0.082	7.086	7.281	-0.195
29	118731200	0.012	7.921	8.190	-0.269
30	118731201	0.003	8.523	8.637	-0.114
34	118731205	0.005	8.301	8.386	-0.085
40	118731211	0.011	7.959	8.045	-0.086
45	118731217	0.009	8.046	7.862	0.184
48	118731220	0.013	7.886	7.511	0.375
51	118731224	0.010	8.000	7.946	0.054
52	118731225	0.032	7.495	7.415	0.080
53	118731226	0.043	7.367	7.233	0.134
54	118731227	0.037	7.432	7.308	0.124
55	118731228	0.056	7.252	7.029	0.223
56	118731229	0.049	7.310	7.331	-0.021
57	118731230	0.185	6.733	6.935	-0.337
59	118731233	0.047	7.328	8.113	0.169
60	118731234	0.021	7.678	7.811	0.092
62	118731236	0.012	7.921	7.766	-0.015
69	118731244	0.044	7.357	7.287	-0.195
73	118731249	0.029	7.538	7.838	-0.269
78	118731377	0.575	6.240	7.351	-0.114







the model's random  $R^2$  ( $cR^2p=0.6570$ ) is significantly greater than the recommended value of 0.50, indicating that the model is not the result of pure chance [12].

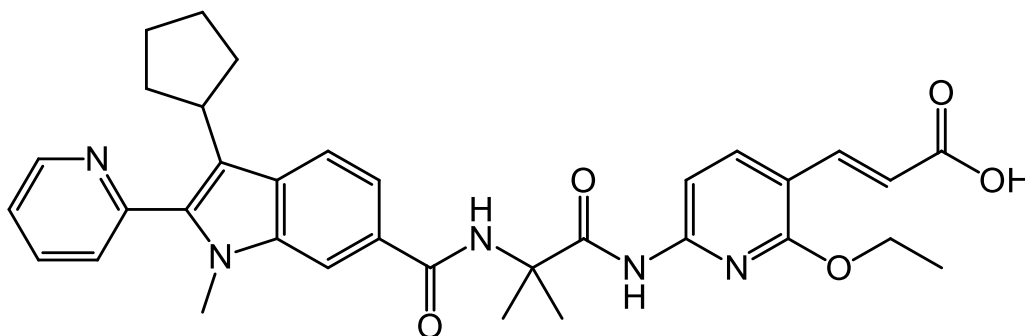
In this study, no Y outlier was detected and one X outlier was detected (molecule 51) accounting for 1% of the entire dataset (see Fig. 2). The majority of the compounds fall within the application domain, indicating that the chemical molecules adhere to a well-defined applicability region. In summary, the optimization techniques have a great deal of potential and are extremely successful. As a consequence, it can be utilized to enhance the bioactivity of any of the chemicals being studied.

Figure 3 shows the estimated vs. actual bioactivity levels across the entire dataset, indicating a close relationship between actual and predicted bioactivity levels in the models which are further confirmed by low residual values presented in Table 3. Because the models have the internal and external statistical capacity and are structurally error-free, they can be used to infer whether or not a known molecule is active if it falls inside the models' AD. As indicated in Additional file 1: Table S4, N15 has the highest predicted bioactivity of all the molecules, much

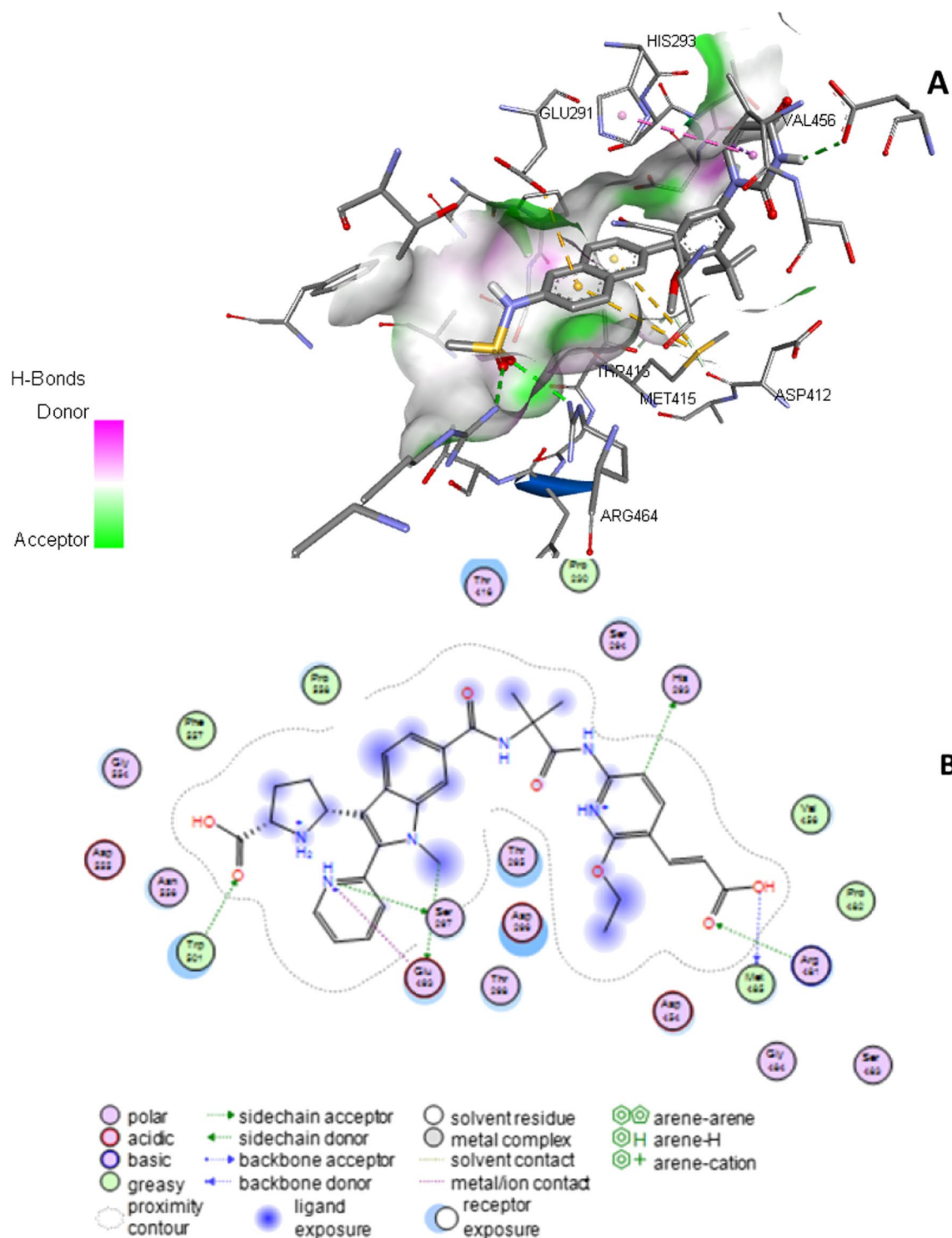
exceeding dasabuvir. It shows the template structure, newly improved molecules, and licensed molecule (dasabuvir), as well as their expected responses, leverages, and MolDock score (binding affinity). Except for molecule 51, the majority of the molecules' leveraging findings were logical and below the leveraging cutoff point ( $h^*=0.43$ ), showing that the improved molecules, as well as dasabuvir, met the requirements and were within the model's applicability restrictions.

GGI10 is a topological descriptor that measures atom-to-atom charge transfer. It has the largest positive coefficient in the calculated linear equation in the chosen model. This means that increasing the charge transfer between the pair of atoms will increase the  $pEC_{50}$  values, and the more active the molecule will be. Conversely, decreasing the charge transfer between the pair of atoms will result in a decrease in the  $pEC_{50}$  values, and the less active the molecule will be [18]. GATS2p and VR2<sub>Dzi</sub> are the other positive correlation descriptors in the chosen model, in order of positive contribution as measured by their coefficients. On the other hand, the ATSC5e is the most negative descriptor contributor in the chosen model, and it is defined as a based Broto–Moreau autocorrelation of lag 5 measured by electronegativities that explains how electronegativities effects are spread along with the topological structures of the molecule [19], and its presence in the model linked the electronegativities of pairs atoms that are bonded together. The built model revealed that increasing ATSC5e reduces the activity of the compounds and that other descriptors such as GATS6p, nX, and MDEC-24 have a significant negative impact on reported anti-HCV bioactivity in decreasing order.

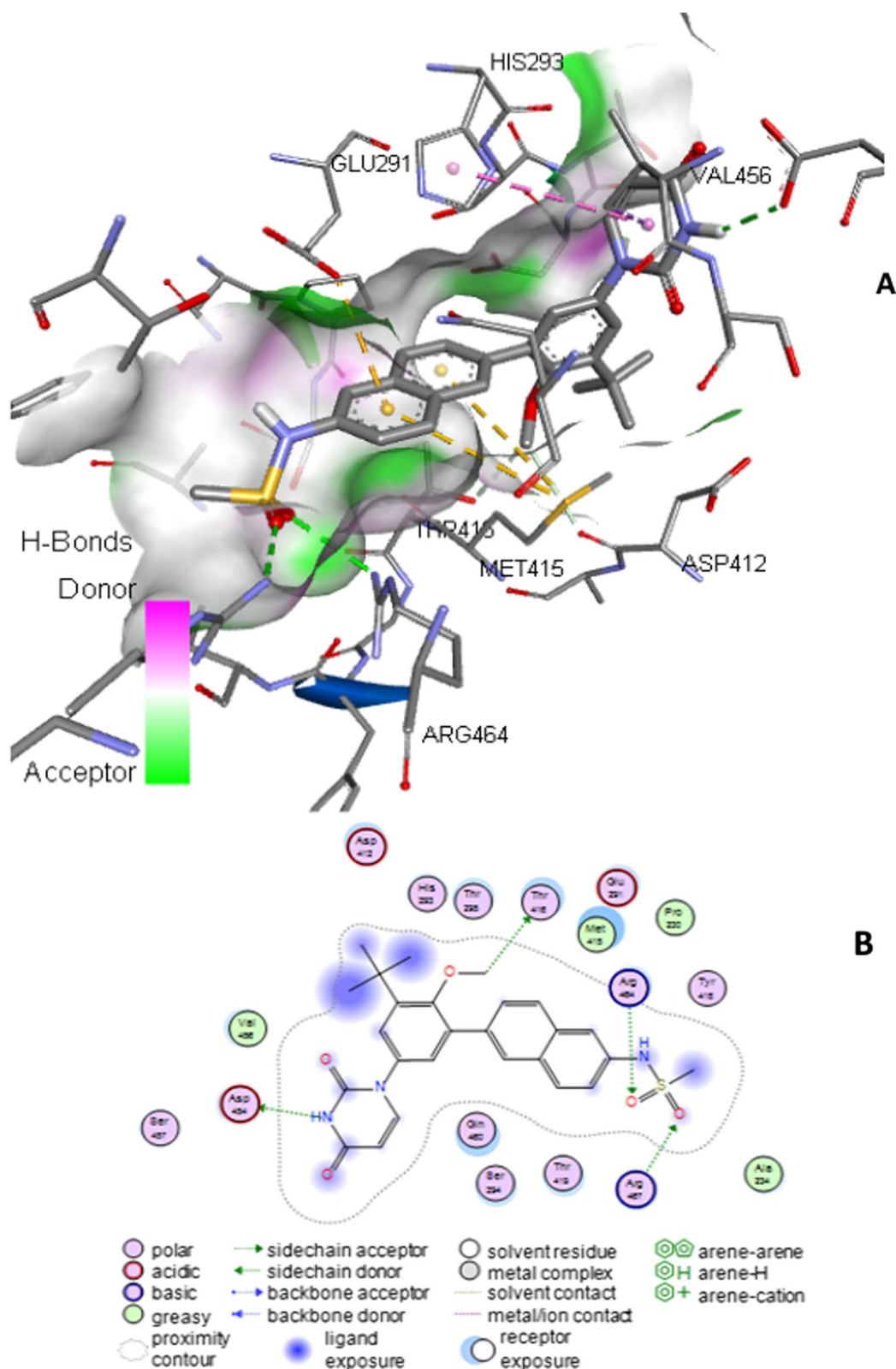
The molecular docking method determines the most thermodynamically favorable ligand conformation using an energy-based scoring function. Lower energy scores, in comparison with higher energy values, are thought



**Fig. 4** (E)-3-[6-[[2-[(3-cyclopentyl-1-methyl-2-pyridin-2-yl)indole-6-carbonyl]amino]-2-methylpropanoyl]amino]-2-ethoxypyridin-3-yl]prop-2-enoic acid (Template Molecule)

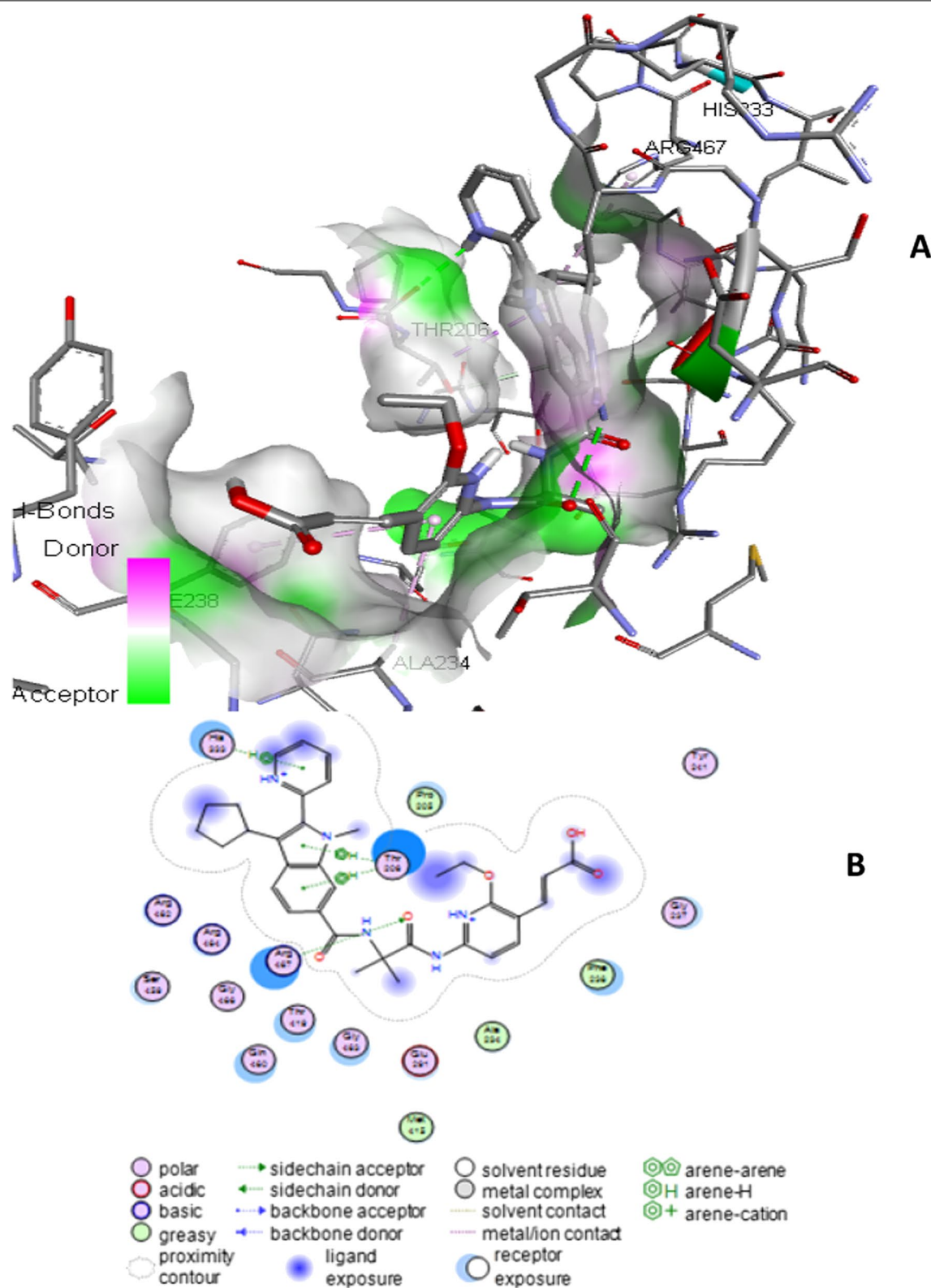


**Fig. 5** **A** 3D diagram for the interaction of molecule 15 (N15) in complex with the target receptor (hepatitis C virus NS5B RNA polymerase, PDB ID: 4MKA), **B** 2D diagram for the interaction of molecule 15 (N15) in complex with the target receptor (hepatitis C virus NS5B RNA polymerase),



**Fig. 6** **A** 3D diagram for the interaction of reference molecule R (dasabuvir) in complex with the target receptor (hepatitis C virus NS5B RNA polymerase, PDB ID: 4MKA), **B** 2D diagram for the interaction of reference molecule R (dasabuvir) in complex with the target receptor (hepatitis C virus NS5B RNA polymerase)





**Fig. 7** **A** 3D diagram for the interaction of template molecule T in complex with the target receptor (hepatitis C virus NS5B RNA polymerase, PDB ID: 4MKA), **B** 2D diagram for the interaction of template molecule T in complex with the target receptor (hepatitis C virus NS5B RNA polymerase)

**Table 4** In-depth report of the ligand–receptor interaction of new molecule 15 (N15), reference molecule (R) and template molecule (T)

Compound	Ligand		Receptor			Interaction	Distance	Energy (kcal/mol)
N15	C	10	OE2	GLU	493	H—donor	3.37	−1.2
	C	23	ND1	HIS	293	H—donor	3.61	−0.9
	O	33	O	MET	485	H—donor	2.90	−1.7
	N	53	OG	SER	297	H—donor	2.92	−1.4
	O	32	NH1	ARG	467	H—acceptor	3.03	−5.7
	O	45	NE1	TRP	501	H—acceptor	3.05	−1.4
	N	53	OE2	GLU	493	Ionic	3.92	−0.7
R	C	24	OG1	THR	416	H—donor	3.31	−0.7
	N	31	OD1	ASP	454	H—donor	3.31	−3.0
	O	21	NH2	ARG	467	H—acceptor	3.16	−2.5
	O	22	NH2	ARG	467	H—acceptor	3.02	−1.8
T	O	11	NH2	ARG	467	H—acceptor	3.30	−1.1
	5—ring		CB	THR	206	Pi—H	3.88	−0.8
	6—ring		OG1	THR	206	Pi—H	3.78	−1.3
	6—ring		NE2	HIS	333	Pi—H	3.68	−1.2

to imply better protein–ligand bindings (binding affinity). As a result, molecular docking can be thought of as an optimization problem in which the goal is to discover the lowest-energy ligand-binding configuration. The results presented in Additional file 1: Table S4 show that N15 has the highest predicted  $pEC_{50}$  value of 12.05 and Molegro Virtual Docker version 6.0 was used to further confirm the result by performing molecular docking analysis which revealed that molecule N15 had the lowest-energy ligand-binding configuration (see Additional file 1: Table S4). The predicted chemical N15, according to the data, outperforms both the reference and template molecules. It had the most favorable MolDock score (binding affinity) of  $-195.6$  kcal/mol, and total interaction energy of  $-13.0$  kcal/mol resulting from six H-bonds and one ionic bond as reported in Table 4, wherewith NH of the piperidine, the OH of the carboxyl group of the acetic acid moiety attached to benzene ring formed H—bond donor with GLU493, HIS293, MET485 and SER297 of the target residue, respectively, while the oxygen atom (O) of the carboxyl group of the acetic acid moiety attached to pyrrolidine and piperidine formed H—bond acceptor with ARG481 and TRP501, respectively, and finally the NH of piperidine moiety formed ionic interaction with GLU493 of the target receptor residue. All these interactions stabilize the N15 in the binding pocket of the receptor, making N15 the most potent novel inhibitor of the target receptor as observed in Table 4 and Fig. 5, and these could be attributed to the presence of pyrrolidine and piperidine groups. The dasabuvir (reference molecule) had a lower MolDock score

(binding affinity) of  $-171.0$  kcal/mol, and total interaction energy of  $-8.0$  kcal/mol resulting from four H-bonds. The methylsulfonyl group attached to dimethylamine connected to the naphthalene ring formed two H—bond acceptors with ARG467 residue, while  $-NH$  attached to pyridine ring and  $-OCH_3$  attached to benzene ring formed H—bond donor each with ASP454 and THR416 target residue, respectively, as reported in Table 4.

Discovery Studio and Molecular Operating Environment (MOE) software were used for virtual evaluation which shed more light on the interaction of molecules N15, R, and T with the target receptor (hepatitis C virus NS5B RNA polymerase, PDB ID: 4MKA). When compared to the reference molecule, molecule N15 has a higher degree of interaction with the target receptor.

The Lipinski idea of five, which argues that if a molecule violates more than two of the qualities given in Table 5, it would be poorly absorbed [15], is one of the most popular and useful directions in the early clinical phase of drug development. The selected compounds, except for N6, N7, and N11, are regarded to meet Lipinski's criterion because the majority of them did not breach more than two, and so maybe categorized as drug-like molecules. Furthermore, an evaluation was carried out using ABS standards, with the template and all generated molecules, including the standard molecule, having a value in the range of 0.11 to 0.56. These criteria were established based on the probability value of a molecule with an optimal permeability and bioavailability outline, where this range represents total compliance of the Lipinski concept of five, which is a

**Table 5** Computed drug-likeness based on Lipinski's rule, bioavailability (BA), and synthetic accessibility (SA)

Molecule Lipinski's rule →	MWt ≤ 500	HBA ≤ 10	HBD ≤ 5	Log <i>p</i> ≤ 5	TPSA (Å <sup>2</sup> ) ≤ 40	Violation	BA	SA
N1	593.72	6	2	4.24	115.21	1	0.55	4.79
N2	609.71	7	2	4.67	124.44	1	0.55	4.79
N3	637.72	8	3	3.67	152.51	2	0.11	4.93
N4	594.70	6	3	3.67	141.23	2	0.55	4.71
N5	686.20	6	2	5.18	115.21	2	0.55	5.04
N6	664.79	8	2	5.08	144.64	3	0.17	5.22
N7	663.81	8	3	5.44	151.42	3	0.17	5.26
N8	622.71	8	2	4.10	144.64	2	0.17	4.87
N9	638.78	7	2	0.00	159.66	2	0.55	4.89
N10	621.73	8	2	5.09	139.93	2	0.17	4.92
N11	637.73	9	2	5.15	149.16	3	0.17	5.01
N12	619.75	6	2	5.00	115.21	2	0.55	4.97
N13	683.82	8	2	4.17	157.73	2	0.17	5.17
N14	567.63	7	3	3.36	135.44	1	0.56	4.39
N15	640.69	10	5	3.43	184.77	2	0.11	4.45
N16	585.61	9	4	3.06	172.74	2	0.11	4.36
N17	542.59	7	4	2.70	161.46	2	0.11	4.14
N18	541.60	7	3	3.04	135.44	1	0.56	4.19
N19	570.60	9	3	3.69	164.87	2	0.11	4.29
N20	586.66	8	3	0.00	179.89	2	0.11	4.29
N21	569.61	9	3	3.20	160.16	2	0.11	4.48
N22	585.61	10	3	3.95	169.39	2	0.11	4.56
N23	631.70	9	3	3.22	177.96	2	0.11	4.74
N24	570.60	8	4	2.23	178.53	2	0.11	4.26
N25	571.58	9	4	2.27	172.74	2	0.11	4.19
T	595.69	7	3	3.78	135.44	1	0.56	4.65
R	493.57	5	2	2.89	118.64	0	0.55	3.46

MWt, molecular weight; HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; TPSA, topological polar surface area; BA, bioavailability; SA, synthetic accessibility; T, Template molecule; N, new designed molecules, R, reference molecule (dasabuvir)

characteristic value of more than 10% [20]. On a scale of 1 to 10, the synthetic accessibility of the chemicals in question was also graded (extremely easy to synthesize to complicated to synthesize) [20]. Except for N5, N6, N7, N11, and N13, which have synthetic accessibility of 5.01–5.26, making them easy to synthesize, the majority of the compounds studied have synthetic accessibility of ≤ 5 (see Table 5).

The ADMET profiles of identified compounds and the reference molecule were assessed and are displayed in Table 6 using web-based pkCSM tools. BBB penetrations values help determine whether or not a chemical will pass across the blood–brain barrier (BBB). A Log BB > 0.3 indicates that a molecule can easily pass the blood–brain barrier, whereas a Log BB < −1 indicates that the molecule is poorly distributed [21]. Except for N7 all of the compounds studied easily crossed the blood–brain barrier in this investigation (see Table 6).

A molecule with a  $P_{app} > 8 \times 10^{-6}$  cm/s is considered to have a high Caco-2 permeability as a credible in vitro model for estimating oral drug permeability [21]. The findings of this study revealed that all of the compounds studied have only modest cell uptake in Caco-2 cells.

The HIA forecasts the percentage of a substance that will be absorbed through the human gut. Poorly absorbed molecules are those that have absorption of less than 30% [15]. The study findings revealed that all of the compounds studied have excellent HIA scores.

The maximum tolerated dose (MRDT) is a measurement of a chemical's hazardous dosage threshold in humans. An MRTD of less than or equal to 0.477 log(mg/kg/day) is considered low for a specific chemical, while significantly greater than 0.477 log(mg/kg/day) is considered high [15, 20]. Except for N14, all of the compounds in this investigation had a low toxicity.

**Table 6** Pharmacokinetic ADMET profiles of newly designed molecules, template molecule, and reference molecule

Molecule	BBB	CaCo2	HIA	MRTD
N1	−0.329	1.460	100.00	0.264
N2	−0.515	1.388	97.962	0.268
N3	1.657	−0.224	67.746	0.506
N4	1.330	0.612	92.884	0.253
N5	−0.235	1.527	95.364	0.292
N6	−0.645	0.618	99.377	0.380
N7	−1.690	0.568	96.158	0.395
N8	−0.682	0.606	96.767	0.298
N9	−0.661	1.206	95.543	0.317
N10	−0.625	0.600	98.053	0.274
N11	−0.849	0.541	97.795	0.309
N12	−0.244	1.337	100.00	0.316
N13	−0.832	0.658	98.614	0.258
N14	1.520	1.138	73.264	0.858
N15	1.098	−0.323	89.430	0.263
N16	1.864	−0.261	38.246	0.297
N17	1.637	−0.156	59.978	0.310
N18	1.565	1.306	67.467	0.293
N19	1.763	−0.120	65.339	0.304
N20	1.739	−0.062	64.114	0.213
N21	1.820	−0.097	67.286	0.267
N22	2.044	−0.115	64.097	0.393
N23	1.952	−0.188	63.637	0.238
N24	1.726	−0.210	54.103	0.198
N25	1.918	−0.261	37.023	1.153
T	1.522	1.223	74.912	0.445
R	1.162	−0.598	83.660	0.216

BBB, blood–brain barrier (Log BB); CaCo2, cell absorption (Log Papp in  $10^{-6}$  cm/s); HIA, human intestinal absorption (% absorbed); MRTD, maximum tolerated dose (Log mg/kg/day); T, template molecule; N, new designed molecules; R, reference molecule (dasabuvir)

## Conclusions

The statistically tested QSAR models produced justified describing the antiviral properties of the compounds investigated. The models are statistically valid, with  $R^2$  values of 0.7051 and 0.6992 for internal and external validation, respectively, and match the criteria for a good QSAR model, as advocated by many organizations. According to the molecular docking simulation, most of the modified compounds had a higher binding affinity for the target receptor (hepatitis C virus NS5B RNA polymerase) than dasabuvir, the reference medication. The enhanced therapeutic features of some novel derivatives were revealed by ADMET prediction, with lowered MRTD (maximum tolerable dosage). This inquiry can be useful to comprehend the physicochemical and biological bioactivities of hepatitis C virus

NS5B RNA polymerase inhibitors and their modified derivatives using a QSAR model, molecular docking, drug-likeness, and ADMET analysis to find new and improved HCV antiviral medicines.

## Abbreviations

HCV: Hepatitis C virus; NS: Non-structural proteins; QSAR: Quantitative structure–activity relationship; PDB: Protein data bank; AD: Applicability domain; SR: Standardized residual;  $IC_{50}$ : Half maximal inhibitory concentration; OECD: Organization for economic co-operation and development; MLR: Multiple linear regression.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s43094-021-00373-6>.

**Additional file 1. Table S1.** Chemical structures of the study compounds. **Table S2.** The descriptor used in the model and their descriptions. **Table S3.** Numerical values of the descriptor used in the selected model. **Table S4.** Chemical Structure, predicted pEC<sub>50</sub>, leverage values, and MolDock scores of template molecule (T), designed molecules (N), and reference Molecule (R).

## Acknowledgements

Not applicable.

## Authors' contributions

SE and AU were in charge of selecting the research tool. SE, AU, GAS, and SEA were involved in the study design, paper writing, and subject selection and data collecting. The statistical analysis was carried out by SE, AU, GAS, and SEA, who also revised the final text. The completed paper was read and endorsed by all writers. All authors read and approved the final manuscript.

## Funding

No fund received.

## Availability of data and materials

All data and materials are available upon request.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

Received: 27 August 2021 Accepted: 20 October 2021

Published online: 29 October 2021

## References

1. Zając M, Muszalska I, Sobczak A, Dadej A, Tomczak S, Jelińska A (2019) Hepatitis C: new drugs and treatment prospects. *Eur J Med Chem* 165:225–249. <https://doi.org/10.1016/j.ejmech.2019.01.025>
2. Abuelizz HA, Marzouk M, Bakheit AH, Al-Salahi R (2020) Investigation of some benzoquinazoline and quinazoline derivatives as novel inhibitors of HCV-NS3/4A protease: biological, molecular docking and QSAR studies. *RSC Adv* 10(59):35820–35830. <https://doi.org/10.1039/d0ra05604a>
3. Qin Z, Yan A (2020) QSAR studies on hepatitis C virus NS5A protein tetracyclic inhibitors in wild type and mutants by CoMFA and CoMSIA,

- SAR QSAR. *Environ Res* 31(4):281–311. <https://doi.org/10.1080/1062936X.2020.1740889>
4. Li W, Si H, Li Y, Ge C, Song F, Ma X, Duan Y, Zhai H (2016) 3D-QSAR and molecular docking studies on designing inhibitors of the hepatitis C virus NS5B polymerase. *J Mol Struct* 1117:227–239. <https://doi.org/10.1016/j.molstruc.2016.03.073>
  5. Therese PJ, Manvar D, Kondepudi S, Battu MB, Sriram D, Basu A, Yogeewari P, Kaushik-Basu N (2014) Multiple e-pharmacophore modeling, 3D-QSAR, and high-throughput virtual screening of hepatitis C virus NS5B polymerase inhibitors. *J Chem Inf Model* 54(2):539–552. <https://doi.org/10.1021/ci400644r>
  6. Algamal ZY, Lee MH, Al-Fakih AM, Aziz M (2017) High-dimensional QSAR classification model for anti-hepatitis C virus activity of thiourea derivatives based on the sparse logistic regression model with a bridge penalty. *J Chemom* 31(6):1–8. <https://doi.org/10.1002/cem.2889>
  7. Zheng M, Liu X, Xu Y, Li H, Luo C, Jiang H (2013) Computational methods for drug design and discovery: focus on China. *Trends Pharmacol Sci* 34(10):549–559. <https://doi.org/10.1016/j.tips.2013.08.004>
  8. Tropsha A (2010) Best practices for QSAR model development, validation, and exploitation. *Mol Inform* 29(6–7):476–488. <https://doi.org/10.1002/minf.201000061>
  9. Shao Y, Molnar LF, Jung Y, Kussmann J, Ochsenfeld C, Brown ST, Gilbert ATB, Slipchenko LV, Levchenko SV, O'Neill DP, DiStasio RA, Lochan RC, Wang T, Beran GJO, Besley NA, Herbert JM, Yeh Lin C, Van Voorhis T, Hung Chien S et al (2006) Advances in methods and algorithms in a modern quantum chemistry program package. *Phys Chem Chem Phys* 8(27):3172–3191. <https://doi.org/10.1039/b517914a>
  10. Ambure P, Gajewicz-Skretna A, Cordeiro MNDS, Roy K (2019) New workflow for QSAR model development from small data sets: small dataset curator and small dataset modeller. Integration of data curation, exhaustive double cross-validation, and a set of optimal model selection techniques. *J Chem Inf Model*. <https://doi.org/10.1021/acs.jcim.9b00476>
  11. Arthur DE, Ejeh S, Uzairu A (2020) Quantitative structure-activity relationship (QSAR) and design of novel ligands that demonstrate high potency and target selectivity as protein tyrosine phosphatase 1B (PTP 1B) inhibitors as an effective strategy used to model anti-diabetic agents. *J Recept Signal Transduct* 40(6):501–520. <https://doi.org/10.1080/10799893.2020.1759092>
  12. Ejeh S, Uzairu A, Shallangwa GA, Abechi SE (2021) Computer-aided identification of a series of novel ligands showing high potency as hepatitis C virus NS3/4A protease inhibitors. *Bull Natl Res Cent*. <https://doi.org/10.1186/s42269-020-00467-w>
  13. Ejeh S, Uzairu A, Shallangwa GA, Abechi SE (2021) In silico design, drug-likeness and ADMET properties estimation of some substituted thienopyrimidines as HCV NS3/4A protease inhibitors. *Chem Afr*. <https://doi.org/10.1007/s42250-021-00250-y>
  14. Schoenfeld RC, Bourdet DL, Brameld KA, Chin E, de Vicente J, Fung A, Harris SF, Lee EK, Le Pogam S, Leveque V, Li J, Lui A-T, Najera I, Rajyaguru S, Sangi M, Steiner S, Talamas FX, Taygerly JP, Zhao J (2013) Discovery of a novel series of potent non-nucleoside inhibitors of hepatitis C virus NS5B. *J Med Chem* 56(20):8163–8182. <https://doi.org/10.1021/jm401266k>
  15. Pires DEV, Blundell TL, Ascher DB (2015) pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J Med Chem* 58(9):4066–4072. <https://doi.org/10.1021/acs.jmedchem.5b00104>
  16. Veerasamy R, Rajak H, Jain A, Sivasadan S, Varghese CP, Agrawal RK (2011) Validation of QSAR models: strategies and importance. *Int J Drug Des Disocov* 2(3):511–519
  17. Yan F, Liu T, Jia Q, Wang Q (2019) Multiple toxicity endpoint-structure relationships for substituted phenols and anilines. *Sci Total Environ* 663:560–567. <https://doi.org/10.1016/j.scitotenv.2019.01.362>
  18. Adawara SN, Shallangwa GA, Mamza PA, Ibrahim A (2020) QSAR model for prediction of some non-nucleoside inhibitors of dengue virus serotype 4 NS5 using GFA-MLR approach. *J Chem Lett* 1:69–76
  19. Todeschini R, Consonni V (2008) Handbook of molecular descriptors, vol 11. Wiley, New York
  20. Umar AB, Uzairu A, Shallangwa GA, Uba S (2020) Docking-based strategy to design novel flavone-based arylamides as potent V600E-BRAF inhibitors with prediction of their drug-likeness and ADMET properties. *Bull Natl Res Cent*. <https://doi.org/10.1186/s42269-020-00432-7>
  21. Danishuddin M, Khan SN, Khan AU (2010) Molecular interactions between mitochondrial membrane proteins and the C-terminal domain of PB1-F2: an in silico approach. *J Mol Model* 16(3):535–541. <https://doi.org/10.1007/s00894-009-0555-5>

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)