

RESEARCH

Open Access



Computer-aided identification of a series of novel ligands showing high potency as hepatitis C virus NS3/4A protease inhibitors

Stephen Ejeh^{*} , Adamu Uzairu, Gideon Adamu Shallangwa and Stephen E. Abechi

Abstract

Background: Hepatitis C virus (HCV) is a global medical condition that causes several life-threatening chronic diseases in the liver. The conventional interferon-free treatment regimens are currently in use by a blend of direct-acting antiviral agents (DAAs) aiming at the viral NS3 protease. However, major concerns may be the issue of DAA-resistant HCV strains and the limited availability to the DAAs due to their high price. Due to this crisis, the developments of a new molecule with high potency as an NS3/4A protease inhibitor of the hepatitis-C virus remain a high priority for medical research. This study aimed to use in-silico methods to identify high potent molecule as an NS3/4A protease inhibitor and investigating the binding energy of the identified molecule in comparison with approved direct-acting antiviral agents (Telaprevir, Simeprevir, and Voxilaprevir) through molecular docking.

Results: The model obtained by in-silico method have the following statistical records, coefficient of determination (r^2) of 0.7704, cross-validation ($q^2_{\text{LOO}} = 0.6914$); external test set ($r^2_{\text{(pred)}} = 0.7049$) and Y-randomization assessment ($^cR^2_p = 0.7025$). The results from the model were used to identify 12 new potential human HCV NS3/4A protease inhibitors, and it was observed that the identified molecule is well-fixed when docked with the receptor and was found to have the lowest binding energy of -10.7 , compared to approved direct-acting antiviral agents (Telaprevir, Simeprevir, and Voxilaprevir) with -9.5 , -10.0 , -10.5 binding energy, respectively.

Conclusion: The binding affinity (-10.7) of the newly identified molecule docked with 3D structures of HCV NS3/4a protease/helicase (PDB ID: 4A92) was found to be better than that of Telaprevir, Simeprevir, and Voxilaprevir (approved direct-acting antiviral agents) which are -9.5 , -10.0 , and -10.5 , respectively. Hence, a novel molecule was identified showing high potency as HCV NS3/4a protease inhibitors.

Keywords: Computer-aided drug design, QSAR, HCV NS3/4A protease inhibitors, Molecular docking, Molecular descriptors

Background

Hepatitis C virus (HCV) is a global medical condition that causes several life-threatening chronic diseases in the liver and the hepatitis C virus (HCV) is primarily liable for chronic liver diseases worldwide (Jia et al. 2020). In many countries, HCV is indeed a primary risk factor for liver failure and liver transplantation which is

a growing public health problem (El-Kassem et al. 2019; González-Grande et al. 2016). The World Health Organization (WHO) estimated seventy-one million people were infected with HCV in 2015, representing one percent of the world's population. The infection is widely dispersed in diverse parts of the world, with an incidence of 0.5 to 6.5 percent in the wide-ranging population (Kucherenko et al. 2016; World Health Organization 2018).

HCV is an associate of the viral envelop (Flaviviridae) consisting of a positive single-strand ribonucleic acid

*Correspondence: ejehstephen@gmail.com
Department of Chemistry, Faculty of Physical Sciences, Ahmadu Bello University, Zaria, Nigeria

(RNA) encoding a polyenzyme. Both host and viral proteolytic enzymes cleave this polyenzyme to the structural and non-structural (NS) enzymes. The NS3/4A area is trypsin similar to the protein complex which performs a very important function in viruses' replication and facilitates procedures for attenuating and evading the natural immune defense of the host cell. Extensive studies indicate that blocking the NS3/4A enzyme function would successfully prevent HCV replication (Shi et al. 2015). HCV serine protease NS3/4A is an enzyme that arises from the mixture of two distinct enzymes, the macromolecule NS3, and the macromolecule NS4A. NS3 macromolecule is a bi-functional macromolecule containing the N-terminal serine protease and a C-terminal endonuclease region. The HCV poly-macromolecules at the NS3/4A, NS4A/4B, NS4B/5A, and NS5A/5B intersections are catalyzed by the N-terminal serine protease. However, NS3 only is not sufficient for cleavage at these intersections. NS4A macromolecule is essential for efficient cleavages. Mainly, as a cofactor, the central portion of NS4A is essential to offer NS3 with a fitted complex and stimulate the catalytic process. At these intersections, abnormalities in NS4A disrupt the NS3 protease to cleave. The exact function of NS4A is to balance the NS3 from cellular protease by protecting it from breakdown and stimulating the proteolytic response (Subedi 2019). There are seven recognized HCV genotypes (GTs), and sixty-seven verified subgroups (Chahine et al. 2017; Petruzzello et al. 2016). The global dispersion of the HCV genotypes differs across various geographic locations. HCV GT1 is the most common in the world and has a broad geographic range, accounting for a larger percentage (46%) of HCV infections worldwide. HCV GT3 is the next most common GT in South Asia, Australia, as well as some European nations with thirty percent of infections worldwide. HCV GT2 and GT4 account for nine to thirteen percent of the infections with a narrow geographic dispersion. HCV GT2 is greater in Asia and West Africa, whereas HCV GT4 infections are high in Central and Eastern Sub-Saharan Africa, North Africa, and the Middle East. HCV GTs 5, 6, and 7 are perhaps the most geographically limited, with GT5 accessible in South Africa and GT6 common in eastern and south-eastern Asia, whereas GT7 also was stated in a small percentage of people in DR Congo (Coppola et al. 2019; Rabaan et al. 2019; Smith et al. 2014).

Though some hepatitis viruses have vaccines, there are none available for HCV (El-Kassem et al. 2019). In the last few years, conventional interferon (IFN)-free therapeutic regimens in blend with ribavirin have been widely recognized as a model of excellence in antiviral therapy (Liu et al. 2018). However, this may have several side effects such as thyroid deficiency, neurological problems,

digestive problems, as well as other negative reactions. Lately, the therapy is primarily dependent on direct-acting antiviral agents in which HCV NS protease is seen as the main target of antiviral inhibitor's development (El-Kassem et al. 2019). According to Liu et al., "in 2011, telaprevir and boceprevir were successively approved as the first direct-acting antiviral agents (DAAs) used as the HCV NS3/4A protease inhibitors, which initiated a breakthrough in the treatment of HCV" (Liu et al. 2018; Poordad et al. 2011). DAAs have culminated in a significantly improved tolerability and effectiveness compared with the traditional regimen of severe HCV infection (Bidell et al. 2016). The development of resistance-mechanism, like antibiotics, also encourages the viable discovery of new compounds or the modification of existing ones (El-Kassem et al. 2019).

The strategy to a correlation existed between structure and activity (QSAR) is indeed very helpful for the estimation of biological responses, particularly in drug development. This strategy is built on the hypothesis that differences in the properties of the molecules (biological activities) may be strongly linked with variations in their physicochemical features (molecular descriptors) (Arthur et al. 2020; Bhadoriya et al. 2015; Veerasamy et al. 2011). Virtual screening (VS) utilizes computer-driven tools and techniques to explore hidden organic molecules that are similar in structure. VS has surfaced in drug development as a computationally intensive strategy to evaluate different databases of chemical compounds for unique hits with improved characteristics, which could then be tested empirically. Just like other computational techniques, VS would not aim to substitute *in vitro* and also *in vivo* assays, but instead to facilitate the development process, lessen the number of candidates to be tried empirically, then justify their selection. Such techniques are typically applied to get hits that seem to be much more likely to offer good clinical candidates (Arthur et al. 2020; Neves et al. 2018; Vyas et al. 2008).

The earlier process of the drug development process is preceded by guesswork, and it is costly in terms of capital, time, and resources. Nevertheless, with the introduction of computational strategies of drug design, the drug development and design process can be successfully carried out saving huge capital resources (Arthur et al. 2016). Over the random screening of existing chemical libraries, the ligand-based strategy has proven successfully (Roy et al. 2012). It provides a theoretical tool that can be used to suggest the actions of recognized and proposed drug molecules. Ligand-based and 3D-QSAR approaches for the discovery of unique and effective NS5B inhibitors were also explored by Therese et al. (2014). In the present research, computational methods were applied to derive a reliable QSAR model and to use

the data provided by the model to proposal novel molecules with high potency as an NS3/4A protease inhibitor and to investigate the binding energy of designed molecules in comparison with an approved direct-acting antiviral agent (Telaprevir, Simeprevir, and Voxilaprevir) through molecular docking.

Methods

Dataset

The molecules utilized in this study were 63 *N*-methyl-6-(*N*-methylmethylsulfonamido)-5-(4-oxo-3,4-dihydroquinazolin-6-yl)benzofuran-3-carboxamide derivatives retrieved from datasets database (<https://pubchem.ncbi.nlm.nih.gov/>) as HCV inhibitors with pubchem AID: 1344392 deposited on 8th September 2018 by ChEMBL (External ID: ChEMBL3888610) obtained as IC₅₀ (μM) and was transformed to pIC₅₀ (pEC₅₀ = -logIC₅₀) (Tropsha 2010).

Computed descriptors

The descriptors were computed by first optimized the dataset molecules with density functional theory (DFT) using B3LYP functional and 631G** basis set in Spartan 14 software (Shao et al. 2006). The optimized structures are first transferred to another software (PaDEL-Descriptors), which computed the structural properties (molecular descriptors) for each molecule (Yap 2011).

Dataset division

In the current analysis, the dataset was mainly split into two parts containing 70% dataset for constructing the model and 30% dataset which is unused during model construction but was used in the determination of the model's predictive ability (Tropsha 2010).

Model generation

The correlation analysis was achieved by Material Studio software and Genetic Function Approximation (GFA) was incorporated in the process to define the ideal QSAR models. In regression analysis, *X* (descriptors) relies on the conditional value of predictor variables *Y* (pIC₅₀) (Veerasingam et al. 2011). GFA is the technique used to generate statistical data models using the evolution process. Substituting regression study further with the GFA algorithm enables model-building to be comparable with, or better to conventional approaches, and provides additional information accessible that is not given by other methods. As with most methods for extrapolation, GFA offers various models for the user (Rogers 1997).

Assessment of the generated model

The established model was assessed by the following numerical measures: cross-validated correlation

coefficient (q_{CV}^2), external explained variance (r_{pred}^2), random R^2 (cR_p^2), variance inflation factor (VIF), and mean effect (MF), which are defined as follows:

$$q_{CV}^2 = 1 - \frac{\sum (y_{exp} - y_{est})^2}{\sum_{i=1}^n (y_{exp} - \bar{y})^2}$$

y_{exp} , y_{est} and \bar{y} represents the experimental, estimated, and average data point of experimental biological response, respectively.

The external explained variance (r_{pred}^2) was computed using the equation:

$$r_{pred}^2 = 1 - \frac{\sum (y_{exp(Test)} - y_{est(Test)})^2}{\sum (y_{exp(Test)} - \bar{y}_{Training})^2}$$

$y_{exp(Test)}$ and $y_{est(Test)}$ represent experimental and estimated activity data for the test set molecules, and $\bar{y}_{Training}$ represents the average experimental biological response of the training set.

The random R^2 values (cR_p^2) of the model were estimated from the equation:

$$cR_p^2 = R \times \sqrt{R^2 - \bar{R}_r^2}$$

where R , R^2 and \bar{R}_r^2 represents correlation coefficient, coefficient of determination, and mean of randomized coefficient of determination, respectively.

The variance inflation factor (VIF) of each descriptor in the model was estimated by the equation:

$$VIF = \frac{1}{1 - R^2}$$

where R^2 is the various correlation coefficient of one descriptor's influence regressed in the model over the other molecular descriptors (Beheshti et al. 2016).

Ery descriptor's mean effect (MF) value had been used to determine the descriptor's comparative impact on the model. The MF was determined by the Formula:

$$MF = \frac{\beta_j \sum_{i=1}^{i=n} d_{ij}}{\sum_j \beta_j \sum_i d_{ij}}$$

where β_j , d_{ij} , m , and n represents the descriptor coefficient j in that model, the descriptor's value in the sample space for each compound in the training dataset, the number of descriptors in the model, and the number of compounds in the training dataset, respectively (Arthur et al. 2020; Oluwaseye et al. 2020).

Applicability domain (AD)

Williams's plot was used to measure the established QSAR model's AD. The actual reference space (h_i) of a molecule, as well as the threshold value (h^*), are evaluated using the equations below:

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

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

For which X is the row-matrix descriptor of the query item, which is distinctive array of the training dataset, n represents the total of query item and q represents the total descriptors in the model (Arthur et al. 2020; Eriksson et al. 2003; Li et al. 2011). The standardized residual (SDR) of the model AD is estimated by the equation:

$$\text{SDR} = \frac{\bar{Y} - Y}{\sqrt{\sum_{i=1}^n \frac{(\bar{Y} - Y)^2}{n}}}$$

In which Y is the observed response value for which ever the set (training or validation sets), \bar{Y} is the model's predicted activity value, and the total of compound present in the dataset is represented by n . The conventional dimension prediction for a given molecule is usually demarcated by $0 < h_i < h^*$ and $-3 < \text{SDR} < 3$ boundaries. Consequently, whichever molecule through SDR less than -3 or greater than $+3$ are labeled an outlier in the variable response area, as well as any molecule with control higher value than h^* , is labeled a distinguished molecule foreign to the most compounds used during model construction. $\text{pIC}_{50} = -50.5082(\pm 22.1927) - 0.0021(\pm 0.0004)ATSC5i$

$$+ 47.1967(\pm 8.8627)SpMin3_Bhs + 23.7391(\pm 7.2362)SpMax3_Bhs$$

$$+ 13.2688(\pm 1.8530)MDEN33 - 23.4365(\pm 3.7989)piPC3$$

Docking studies

Ligand structure preparation

ChemBio Ultra 19.2.0 was used to draw the ligand structures (Evans 2014; Li et al. 2004). The density functional theory (DFT) technique in Spartan 14 was used to minimize the energy of each ligand in the dataset and input into PyRx in PDB file format (Huey et al. 2012).

Protein structure preparation

The structure of HCV NS3/4a protease was extracted via Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) with the HCV NS3/4a protease structural PDB ID being 4A92. The co-crystallized ligand macrocyclic protease inhibitor in complex with the HCV NS3/4a protease crystal structure was discarded, hydrogen atoms were introduced, slight use residue structures were discarded, partials side chains were

substituted with the use of discovery studio (Danishuddin et al. 2010). The structures are again saved in PDB form for use in PyRx tools (Huey et al. 2012).

Docking procedure and evaluation

A rectangular grid measuring $65.5217 \times 72.7141 \times 80.3011$ Å, centered on 5.2017, 15.6939, 30.8304 was built across the binding site of ligand on HCV NS3/4a protease by autodock tools. The grid center was fixed at ligand, and grid energy measurements were performed. The Autodock docking computation used default settings, and 10 docked alignments were produced for each molecule. The bonded ligand was deleted from the complexes test the validity and reliability of the docking computations and forwarded for one-ligand run computation. This replicated core-scoring sequences of 4 falling from bonded X-ray verification for HCV NS3/4a protease roots mean square deviation (rmsd) standards of 0.71–0.74 Å, proposing that this process is sufficiently valid to be enough for docking studies of other molecules. The results were transferred for thorough observation of the binding relationships and correlations between the molecules and amino acid sequences at the active spots using discovery-studio software (Trott and Olson 2010).

Results

A QSAR method for investigating the structure–activity relationship of 63 HCV NS3/4a protease inhibitors was implemented in the present research, and the QSAR model is presented as:

$$n_{\text{train}} = 44, r_{\text{train}}^2 = 0.7704, K = 5, F = 25.4979, q_{\text{LOO}}^2 = 0.6914, \text{RMSE}_{\text{train}} = 0.3880,$$

$n_{\text{test}} = 19, r_{\text{test}}^2 = 0.7047, \text{RMSE}_{\text{test}} = 0.3392, \text{Outliers} > 3.0 = 3, \text{Influential molecules} > h^* = 4.$

where n_{train} and n_{test} are amounts of data present in the training and validation dataset, respectively, r_{train}^2 and r_{test}^2 are the coefficients of correlation for internal and external validation, respectively, Q_{LOO}^2 is the squared cross-validation coefficients for leave one out, F is the Fisher F statistic, and $\text{RMSE}_{\text{train}}$ and $\text{RMSE}_{\text{test}}$ are the roots mean square error for training and test set, respectively, K is the predictor parameters (descriptors) present in the model.

Additional file 1: Table S1 displays the chemical structure, Pubchem SID and CID, experimental IC_{50} , and Estimated pIC_{50} of all the compounds used for this analysis. Table 1 describes the model figures. The model's correlation matrix and VIF are stated in Table 2. Table 3 presents 10 iterations of γ -randomization test.

Table 1 QSAR models validation parameters and scores

Statistical parameter	Values for the developed model	Threshold value	Comments	Source
Internal validation (validation with the training set data)				
LOF	0.1574			
R^2	0.7704	> 0.6	Passed	a
R^2_{adj}	0.7402	> 0.6	Passed	b
F	25.4979			
Q^2_{LOO}	0.6914	> 0.5	Passed	b
SEE	0.4175			
RMSE	0.3880			
$R^2_{m(LOO)}$	0.6507	> 0.5	Passed	c
$R^2'_{m(LOO)}$	0.5106	> 0.5	Passed	c
$\Delta R^2_{m(LOO)}$	0.1401			
$^cR^2_p$	0.7025	> 0.5	Passed	d
External validation (validation with test set data)				
$R^2_{(pred)}$	0.7047	> 0.6	Passed	a
r^2	0.6359	> 0.5	Passed	e
r^2_0	0.6235	> 0.5	Passed	e
r^2_0	0.5651	> 0.5	Passed	e
$ r^2_0 - r^2_0 $	0.0584	< 0.3	Passed	e
$r^2 - r^2_0/r^2$	0.0195	< 0.1	Passed	f
k	1.0004	$0.85 < k < 1.15$	Passed	f
$r^2_0 - r^2_0/r^2_0$	0.0937	< 0.1	Passed	f
k'	0.9979	$0.85 < k < 1.15$	Passed	f

a = (Veerasamy et al. 2011), b = (Tropsha 2010), c = (Damme and Bultinck 2007), d = (Roy et al. 2012), e = (Golbraikh and Tropsha 2002), f = (Tropsha et al. 2003)

Table 2 Pearson's correlation, Variance Inflation Factor (VIF) of descriptors used in the model

Descriptors	ATSC5i	SpMin3_Bhs	SpMax3_Bhs	MDEN33	piPC3	VIF
ATSC5i	1					1.659
SpMin3_Bhs	-0.324	1				2.539
SpMax3_Bhs	-0.712	0.622	1			3.280
MDEN33	-0.528	0.608	0.623	1		2.029
piPC3	-0.675	0.841	0.865	0.754	1	1.259

The descriptions of the descriptors used in the model and the computed mean effect (MF) of each descriptor present in the model are stated in Table 4. Table 5 represents the Template molecule, designed molecules, and approved direct-acting antiviral agents (Telaprevir, Simeprevir, and Voxilaprevir) with their estimated pIC₅₀ and leverages while Table 6 presents the docking results of the template molecule, designed molecule with the highest activity, and approved direct-acting antiviral agents (Telaprevir, Simeprevir, and Voxilaprevir). In Fig. 1 the model applicability domain (AD) was represented using Williams's plot. Figure 2 presents a plot of the model estimated against

experimental anti-hepatitis C activity values for both the training and test sets. Figure 3 represents the chart of SDR vs estimated pIC₅₀ values for the whole data point. The structure of the precursor molecule which is compound 33 in Additional file 1: Table S1 is presented in Fig. 4. The three-dimensional and two-dimensional interactions of the template molecule (Molecule 1, see Table 6), the newly designed molecule (molecule 7, see Table 6), Telaprevir (molecule 14, see Table 6), Simeprevir (molecule 15, see Table 6) and Voxilaprevir (molecule 16, see Table 6) with the binding pocket of 3D structures of HCV NS3/4a protease/helicase are presented in Figs. 5, 6, 7, 8 and 9, respectively.

Table 3 Y-randomization test scores

Iteration	R	R ²	Q ²
Random 1	0.251435	0.063219	-0.37456
Random 2	0.55169	0.304362	0.124734
Random 3	0.599885	0.359861	0.063428
Random 4	0.177018	0.031335	-0.74622
Random 5	0.257091	0.066096	-0.19249
Random 6	0.537332	0.288726	0.116066
Random 7	0.357103	0.127522	-0.38471
Random 8	0.344553	0.118717	-1.45487
Random 9	0.284942	0.081192	-0.21375
Random 10	0.241769	0.058452	-0.48915
Random models parameters			
Average r	0.360282		
Average r ²	0.149948		
Average Q ²	-0.35515		
cR _p ²	0.702485		

Discussion

The developed model explains seventy-seven percent (77%) and predicts seventy percent (70%) of the variances of the considered molecules with anti-hepatitis-C virus activity against HCV NS3/4a protease. The model statistics described in Table 1 meet the criteria for validating a QSAR model developed by OECD (Roy et al. 2012; Veerasamy et al. 2011). The findings show that the R² and Q² for the model's internal evaluation have been stated as 0.7704 and 0.6914, respectively. This implies that the model correctly interpreted the data when evaluated and that the model can estimate the fitted training set, as the model predicted approximately 70% of the data and thus met the minimum condition of 50% (Veerasamy et al. 2011). The error statistics, such as SEE and RMSE, have also been documented in Table 2 and have been found to backing model robustness.

Qin et al. reported a QSAR study of the bioactivity of hepatitis C virus (HCV) NS3/4A protease inhibitors by multiple linear regression (MLR) and support vector machine (SVM) in the literature, and results show R² values for internally and externally evaluation were, respectively, 0.75 and 0.72 (Qin et al. 2017), which seem to be

close in values compared to R² values for internally and externally evaluation of 0.77 and 0.70, respectively, as noted in this article.

From Table 2, it was detected that all the descriptors have VIF scores below 5, which means the model obtained has significant results and that the descriptors were considered to be fairly orthogonal (Eriksson et al. 2003). The mean effect (MF) value offers significant details on the impact of the model's molecular descriptors, the size and the signs of these descriptors MF show their intensity and direction in manipulating the activities of the study compounds and are observed to be in decreasing order of piPC3 > SpMax3_Bhs > SpMin3_Bhs > ATSC3i > MDEN33 (see Table 4). SpMax3_Bhs, SpMin3_Bhs, and MDEN33 contribute positively, while piPC3 and ATSC3i contribute negatively to the activity of the anti-hepatitis-C virus compounds. The y-randomization test computed shows that the value for the model's random R² (cR²_p = 0.7025) is substantially higher compared to the target value of 0.50, meaning the model is never the product of simple possibility alone (Arthur et al. 2016).

The square area in Fig. 1 represents the model AD. Where h^* (0.41) boundary is the model warning leverage and SDR is the standardized residual of the models. The outcome shows that 89% of the molecules considered were inside the AD of the model while 5% formed the Outliers which are compound 16, 18, and 32 in Additional file 1: Table S1 as identified and indicated in Fig. 1 (> ± 3.0) and 6% of the studied molecules are Influential molecules which are compound 2, 6, 21, and 39 in Additional file 1: Table S1 as identified and indicated in Fig. 1 (> h^*). In summary, the suggested model had high potential and efficiency. Thus, it can be used as an instrument for optimizing the activity of any of the compounds considered.

Figure 2 indicates a meaningful correlation among the models' experimental and estimated activity values, and there was also propagation of the models residual along with the axis SDR equivalent to zero in Fig. 3. Such results showed that the models had the high predictive potential both internally and externally and were free of systemic bias. Consequently, they could be used to

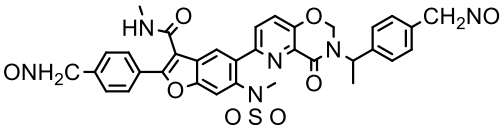
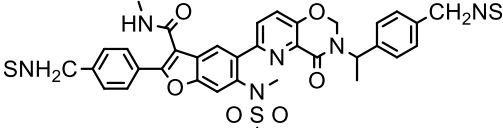
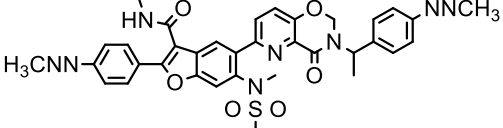
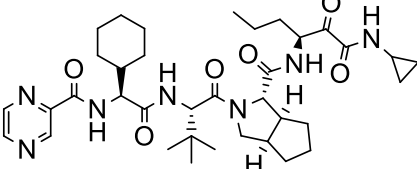
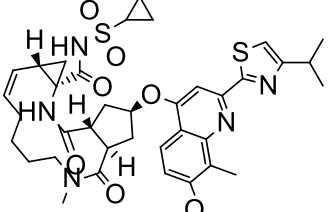
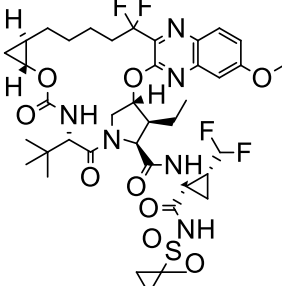
Table 4 A description of the descriptor used in the model and the MF

S/N	Descriptor	Description	Class	bj	MF
1	ATSC5i	Centered Broto-Moreau autocorrelation—lag 5/weighted by first ionization potential	2D	-0.002	0.032
2	SpMin3_Bhs	Smallest absolute eigenvalue of Burden-modified matrix—n 3/weighted by the relative I-state	2D	47.197	1.592
3	SpMax3_Bhs	Largest absolute eigenvalue of Burden-modified matrix—n 3/weighted by relative I-state	2D	23.739	1.594
4	MDEN33	Molecular distance edge between all tertiary nitrogens	2D	13.269	0.0273
5	piPC3	Conventional bond order ID number of order 3 (ln(1 + x))	2D	-23.437	-2.246

Table 5 Template molecule, designed molecules, Telaprevir, Simeprevir, and Voxilaprevir with their estimated pIC_{50} and leverages

Molecule	Structure	Estimated pIC_{50}	Leverages
1 ^x		8.3572	0.3155
2 ^y		9.0010	0.2157
3		8.2936	0.4872
4		8.0445	0.2300
5 ^y		9.1787	0.2976
6 ^y		9.1765	0.2858
7 ^{yy}		17.3373	0.2339
8 ^y		11.6416	0.4126
9 ^y		8.4136	0.3166
10		8.3165	0.3628

Table 5 (continued)

Molecule	Structure	Estimated pIC ₅₀	Leverages
11 ^y		10.2050	0.2694
12 ^y		9.5078	0.3581
13 ^y		8.5560	0.4126
14 ^f		13.5142	0.3446
15 ^s		15.1774	0.3811
16 ^t		16.9516	0.2767

^x Template molecule, ^ydesigned molecules with improved activity, ^{yy}designed molecule with better activity compared to Telaprevir, Simeprevir, and Voxilaprevir, ^fTelaprevir, ^sSimeprevir, ^tVoxilaprevir (f, s, t are first, second and third generation DAAs, respectively)

predict known molecules lacking activity, as long as the molecule is inside the AD of the model.

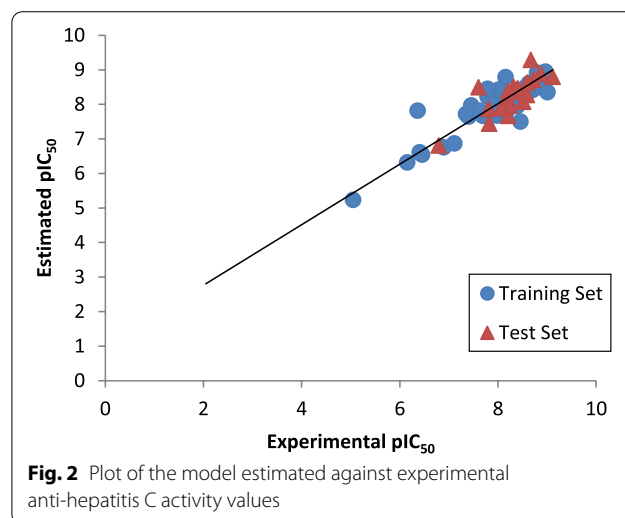
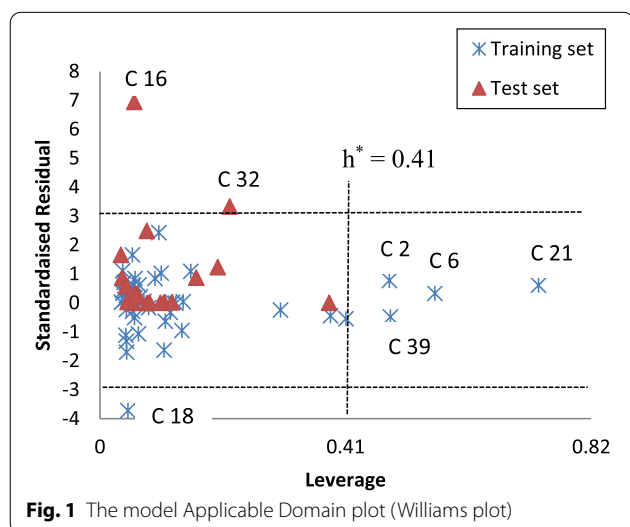
Explanation of descriptors utilized in the established QSAR model

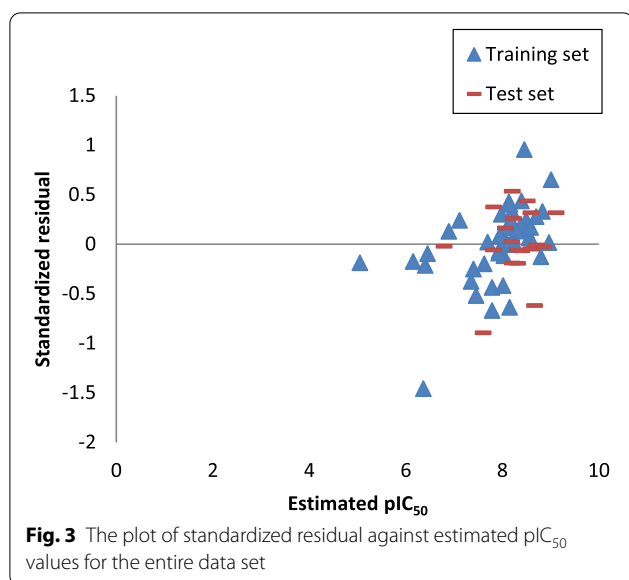
The first predictor parameter in the model is a two-dimensional Autocorrelation descriptor (ATSC5i) and is characterized as based Broto-Moreau autocorrelation of lag 5 measured by first ionization potential which

explains how the first ionization potential is spread along with the topological structures of the molecules (Gramatica et al. 2000), and its presence in the model linked the first ionization potential of pairs atoms that are disconnected by five bonds (lag 5) of the researched molecules with anti-hepatitis-C virus activity on HCV NS3/4a protease. It was found from the developed model that ATSC5i negatively influences the activity of the compounds when increased.

Table 6 Docking results of Template molecule, designed molecule with the highest activity, first, second, and third generation approved direct-acting antiviral agents

The compound from Table 5	Binding energy kCal/mol	Residues interacting with ligand	Types of interaction	Bond length (Å)
1	-7.5	THR298, SER229 SER297 GLU291 SER294, TRP501	Conventional Hydrogen Bond Carbon Hydrogen Bond Pi-Anion Pi-Donor Hydrogen Bond	2.17, 3.60 3.69 3.51 3.78, 3.54
7	-10.7	ALA497, PRO230, HIS293 THR416, THR416, SER294, THR295, GLY484 TYR391, HIS293 ASP454 SER457, THR295, SER483 ARG393, VAL456	Pi-Alkyl Conventional Hydrogen Bond Carbon Hydrogen Bond Pi-Anion Pi-Donor Hydrogen Bond Pi-Alkyl	4.48, 4.79, 5.48 2.91, 2.79, 3.02, 2.85, 3.10 3.70, 3.10 3.43 3.75, 4.05, 3.56 4.69, 4.64
14	-9.5	ARG481, MET485, GLY484, THR295, THR295, HIS369 VAL490, VAL490, PRO523, VAL456 VAL432	Conventional Hydrogen Bond Alkyl Pi-Alkyl	3.18, 1.95, 3.11, 3.31, 3.10, 2.32 4.99, 4.71, 4.34, 4.61 5.01
15	-10.0	ALA413, GLN434 ASN556, SER489, GLU493 ASP454 THR295, THR433 VAL456, VAL490 VAL456	Conventional Hydrogen Bond Carbon Hydrogen Bond Pi-Anion Pi-Donor Hydrogen Bond Alkyl Pi-Alkyl	2.80, 3.21 3.58, 3.14, 3.63 4.09 3.71, 3.88 3.96, 3.57 4.93
16	-10.5	GLU493, TRP501, TYR502, GLY253 THR269 ASP412, ASP412 GLY271 ALA413, PRO558 ALA497, TRP501	Conventional Hydrogen Bond Carbon Hydrogen Bond Halogen (Fluorine) Amide-Pi Stacked Alkyl Pi-Alkyl	2.39, 3.16, 2.50, 3.17 3.47 3.31, 3.40 4.59 4.19, 4.50 4.86, 5.06





SpMin3_Bhs and SpMax3_Bhs have positive influences on the anti-hepatitis-C virus activity of the analogs on HCV NS3/4a protease. They are both Burden-Modified Eigenvalues Descriptor. SpMin3_Bhs reveals the smallest absolute eigenvalue of the Burden modified matrix— $n \ 3 /$ measured by the relative I-state. Fluorine/fluorine-containing substituents, e.g., C_6H_5F , drop the estimated value and so are unfavorable to the activity. The SpMax3_Bhs is defined as the largest absolute eigenvalue of Burden-modified matrix— $n \ 3 /$ measured by relative I-state. A transformed correlation matrix (Burden matrix) is considered, the diagonal element of which is being replaced by the comparative interpretative condition of the atoms in the molecule, and the two bonded atoms are represented off-diagonally. The descriptor encrypts facts about the underlying molecular structure feature and is usually

employed for similarity/differences searching (Todeschini and Consonni 2008).

MDEN33 is another descriptor found in the model and is described as a Molecular distance edge among all tertiary nitrogen. The descriptors are symbolized as MDEXst in which X represents the element, s stand for the first atom kind, and t stand for the second atom kind. The kind or level of the atom (primary, secondary, or tertiary) is derived from the number of non-hydrogen atoms connected to a specified atom (Todeschini and Consonni 2008). It is positively related to the activity of researched molecules. This is an indication that the N-containing moiety introduction rises the activity values of the researched molecules. A molecule with an extra nitrogen atom had a high value of MDEN33 in its structure, hence better activity e.g., molecule, 7, 8, 11, and 12 in Table 6.

In the model, the last descriptor is piPC3 which is a two-dimensional descriptor and is known as a Standard bond order 3 ($\ln(1+x)$, and was found to have a negative impact on the activity of the researched molecules when improved (Todeschini and Consonni 2008).

New molecule proposal and estimation of activity

Based on the built QSAR model and evaluated results, compound 33 in Additional file 1: Table S1 shown in Fig. 4 was used as a pattern to improve the molecular structure. Compound 33 was used as a template for designing novel molecules because it was carefully chosen from Fig. 1, detecting the compound with high activity, low standardized residual, and was discovered inside the established model's AD. The previously established QSAR model was used to estimate the activity of the template molecule, newly designed molecules, and approved direct-acting antiviral agents (Telaprevir, Simeprevir, and Voxilaprevir). The result shows that all designed derivatives and DAAs have enhanced

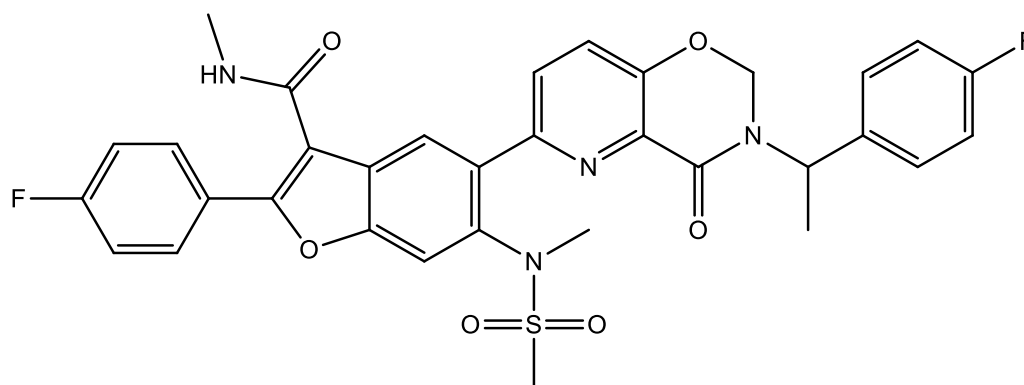
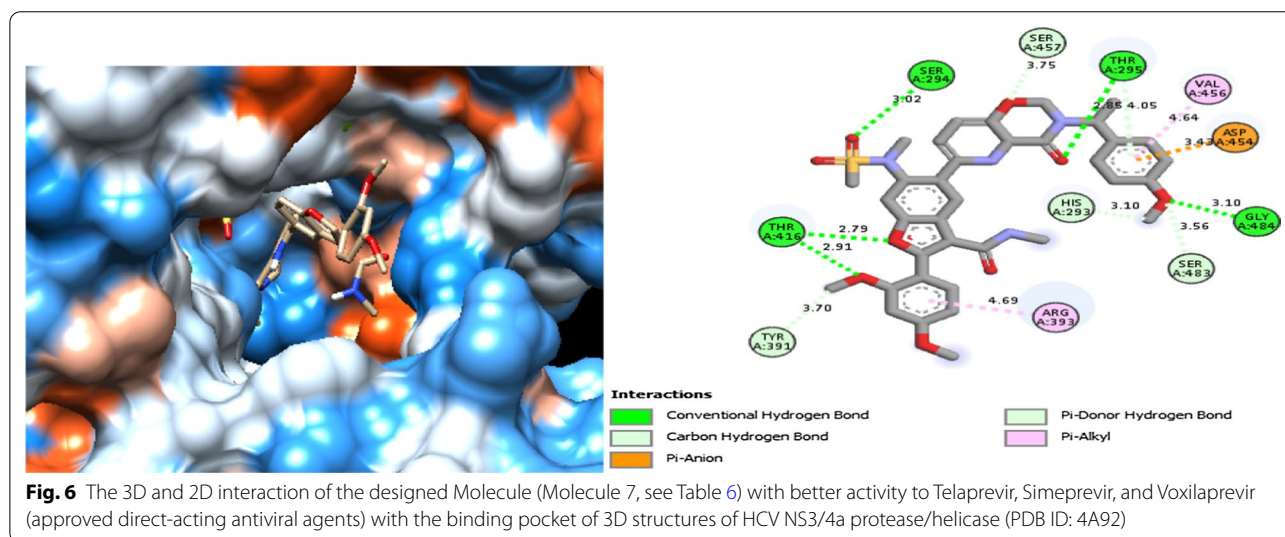
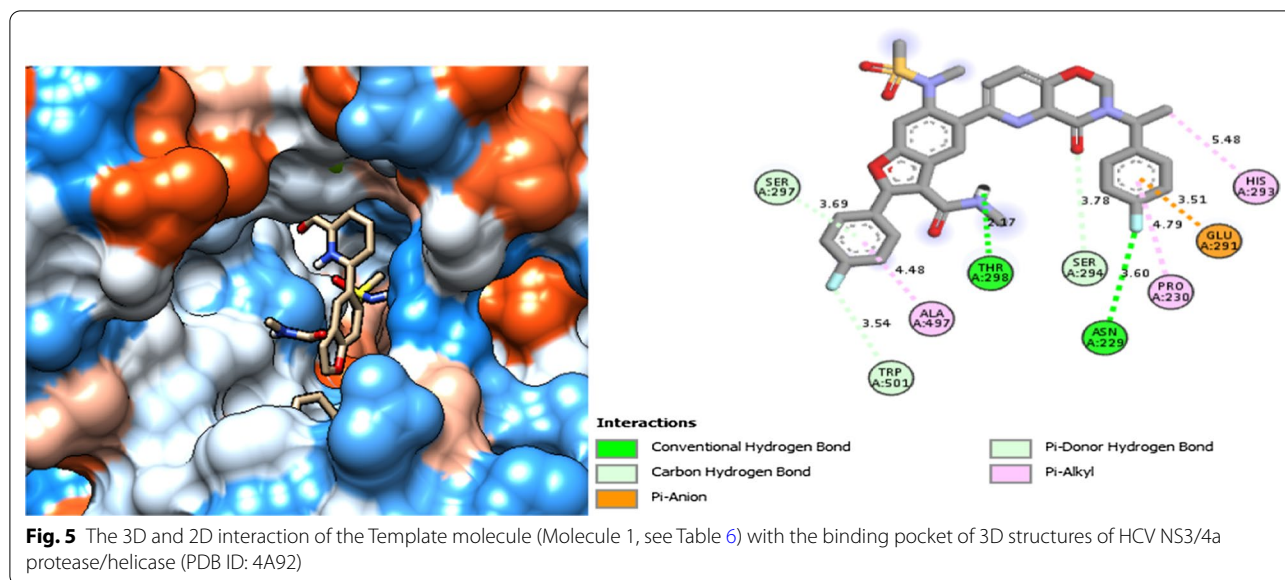


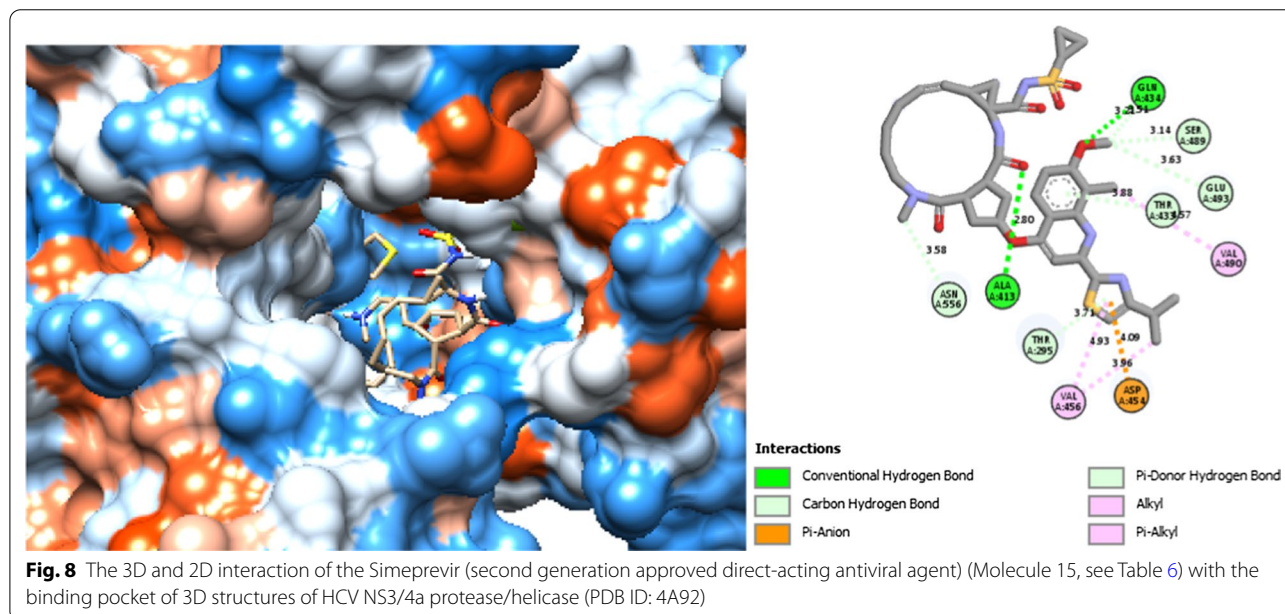
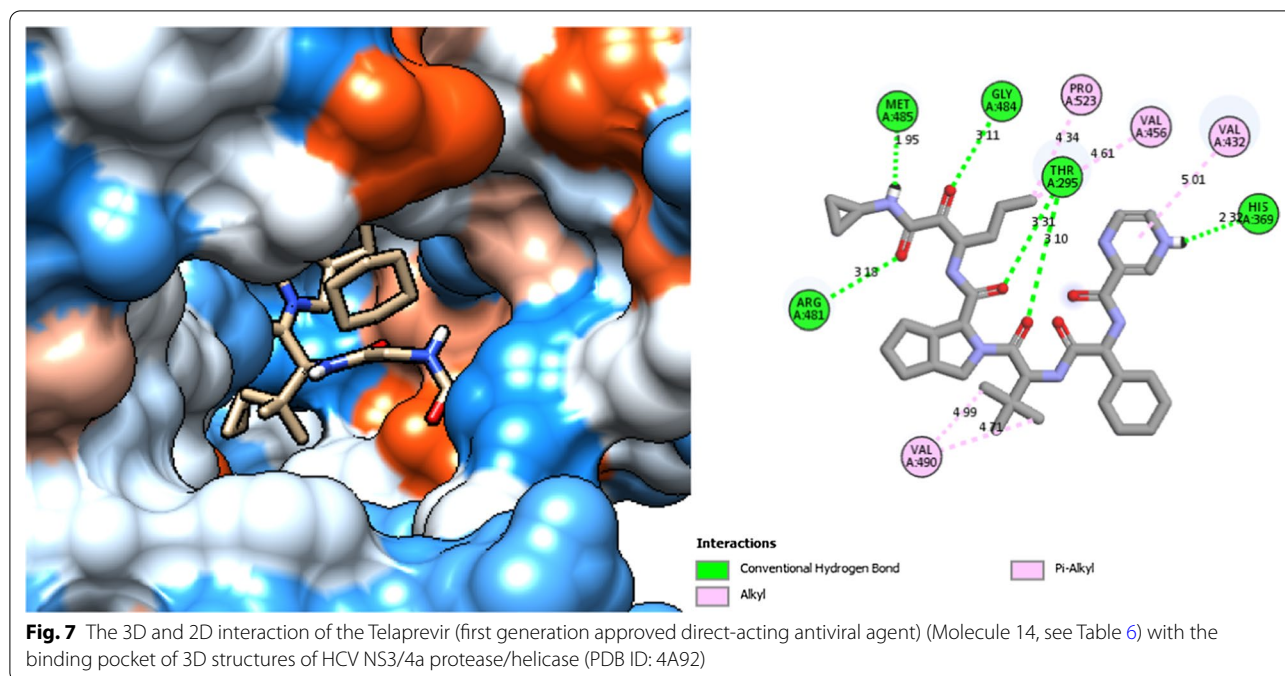
Fig. 4 The structure of the template molecule see Additional file 1: Table S1, C 33 (2-(4-fluorophenyl)-5-(3-(1-(4-fluorophenyl)ethyl)-4-oxo-3,4-dihydro-2H-pyrido[2,3-e][1,3]oxazin-6-yl)-N-methyl-6-(N-methylmethylsulfonamido)benzofuran-3-carboxamide)



pIC_{50} value than the template except molecule 3, 4, and 10 (see Table 5). Also, molecule 7 in Table 5 has the highest activity among them even better compared to DAAs. The structure of the template, newly designed molecules, and approved direct-acting antiviral agents (Telaprevir, Simeprevir, and Voxilaprevir) together with their estimated activity and Leverages are presented in Table 5. The outcomes of the leverage were good and found to be lower than the leverage threshold ($h^* = 0.41$), this implied that all the designed molecules, as well as approved direct-acting antiviral agents (Telaprevir, Simeprevir, and Voxilaprevir), were within the model's sphere of applicability.

Molecular docking results and analysis

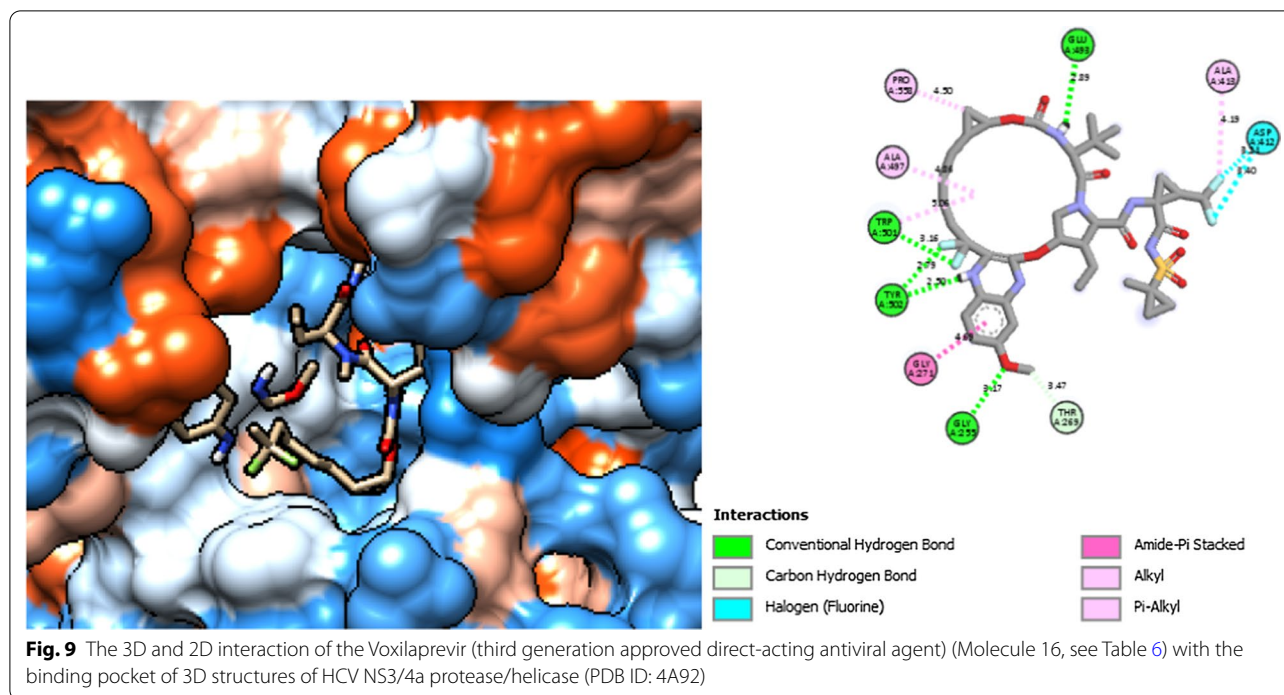
Among all the molecules in Table 5 including approved direct-acting antiviral agents, it was observed that molecule 7 has the highest predicted pIC_{50} value (17.3373) and therefore was exposed to a molecular docking study. Also, Telaprevir, Simeprevir, and Voxilaprevir (approved direct-acting antiviral agents) and the template molecule with predicted pIC_{50} of 13.5142, 15.1774, 16.9516, and 8.3572, respectively, are exposed to a similar docking analysis for comparison. The results of the docking analysis such as Binding Energy (kCal/mol), Interactions with amino acid, Types of Interaction, Bond length (Å) of the template molecule (1 in Table 5), designed molecule (7 in Table 5), and



the approved direct-acting antiviral agents (Telaprevir, Simeprevir, and Voxilaprevir) is reported in Table 6, while Figs. 5, 6, 7, 8, 9 show the three-dimensional and two-dimensional interaction of the template molecule (molecule 1 in Table 5), designed molecule (molecule 7 in Table 5), and the approved direct-acting antiviral agents (molecule 14, 15, and 16, respectively, in Table 5) with the binding pocket of 3D structures of HCV NS3/4a protease/helicase (PDB ID: 4A92), respectively.

As shown in Fig. 6, we observe that the designed molecule is excellently placed in the active pocket of the receptor. Because it has the highest activity and the lowest binding energy (17.3373 and -10.7) compared to molecule 1 (8.3572 and -7.5), molecule 14 (13.5142 and -9.5), molecule 15 (15.1774 and -10.0), and molecule 16 (16.9516 and -10.5).

Figure 5 shows that THR298, SER229, SER297, GLU291, SER294, TRP50, ALA497, PRO230, and



HIS293 amino acid of the target receptor are involved in the interaction with template molecule (molecule 1 in Table 6). Figure 6 shows that THR416, THR416, SER294, THR295, GLY484, TYR391, HIS293, ASP454, SER457, THR295, SER483, ARG393, and VAL456 are the amino acid of the target receptor involved in the interaction with a designed molecule (molecule 7 in Table 5), while Fig. 7 shows that ARG481, MET485, GLY484, THR295, THR295, HIS369, VAL490, VAL490, PRO523, VAL456, and VAL432 are the amino acid of the target receptor involved in the interaction with the approved direct-acting antiviral agents (molecule 14 in Table 5). Also Fig. 8 shows that ALA413, GLN434, ASN556, SER489, GLU493, ASP454, THR295, THR433, VAL456, and VAL490 are the amino acid of the target receptor involved in the interaction with molecule 15 in Table 5 and Fig. 9 shows that GLU493, TRP501, TYR502, GLY253, THR269, ASP412, ASP412, ALA413, PRO558, ALA497 are the amino acid of the target receptor involved in the interaction with molecule 16 in Table 5. It was observed from the docking results presented in Table 6 that the target amino acid THR is involved in the interaction with all the docked molecules. This implies the importance of this amino acid in the inhibition of HCV NS3/4a protease/helicase. The bound complex of telaprevir with HCV NS3/4a protease/helicase is accompanied by the establishment of a covalent bond between the serine nucleophile of the HCV protease catalytic triad and the ketoamide connectivity of telaprevir which forms

a stable, covalent and reversible complex with the serine protease, but the designed molecule accounts for both covalent and non-covalent interactions of the inhibitor with HCV NS3/4a protease/helicase. It was also observed that molecule 7 shows more interaction with the target receptor when compared to the template molecule and approved direct-acting antiviral agents which imply, the more the interaction the better the inhibition.

Limitations of the present study

Non-availability of reliable experimental datasets on hepatitis C virus.

Conclusion

The theoretically verified QSAR model found offered rationales to describe the anti-hepatitis-C virus activities of researched molecules. The model is theoretically reliable with sound statistical record ($r^2 = 0.7704$ and $r^2_{\text{pred}} = 0.7047$). And meet the conditions of a satisfactory QSAR model suggested by various groups. Different molecules having improved anti-hepatitis-C virus activity compared to the better active molecule in the data collection (compound 33), have been proposed for deep investigation. The binding affinity (-10.7) of this newly identified molecule docked into the binding pocket of 3D structures of HCV NS3/4a protease/helicase (PDB ID: 4A92) were found to be better than that of compound 33 (-7.5) in the datasets as well as approved direct-acting antiviral agents (Telaprevir, Simeprevir, and Voxilaprevir)

which are -9.5 , -10.0 , and -10.5 , respectively. Hence, a novel molecule was identified showing high potency as HCV NS3/4a protease inhibitors.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s42269-020-00467-w>.

Additional file 1: Table S1. Molecular structure, Pubchem CID, Pubchem SID, IC₅₀ [μ M], pIC₅₀ and predicted pIC₅₀ of the datasets used in the present research.

Acknowledgements

Authors are grateful to the Physical chemistry research team head by Prof. Uzairu of department of the chemistry Ahmadu Bello University Zaria-Nigeria for their meaningful contributions.

Authors' contributions

SE and AU were liable for selecting the study instrument. SE, AU, GAS, and SEA conceived the research, prepared the manuscript, and participated in the selection and collection of data and statistical analysis. All authors read and approved the final manuscript.

Funding

No fund received.

Availability of data and materials

Not applicable.

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: 2 November 2020 Accepted: 7 December 2020

Published online: 06 January 2021

References

- Arthur DE, Uzairu A, Mamza P, Abechi S (2016) Quantitative structure–activity relationship study on potent anticancer compounds against MOLT-4 and P388 leukemia cell lines. *J Adv Res* 7(5):823–837
- 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* 40:1–20
- Beheshti A, Pourbasheer E, Nekoei M, Vahdani S (2016) QSAR modeling of antimalarial activity of urea derivatives using genetic algorithm–multiple linear regressions. *J Saudi Chem Soc* 20(3):282–290
- Bhadoriya KS, Sharma MC, Jain SV (2015) 2, 4-Dihydropyrano [2, 3-c] pyrazole: discovery of new lead as through pharmacophore modelling, atom-based 3D-QSAR, virtual screening and docking strategies for improved anti-HIV-1 chemotherapy. *J Taibah Univ Sci* 9(4):521–530
- Bidell MR, McLaughlin M, Faragon J, Morse C, Patel N (2016) Desirable characteristics of hepatitis C treatment regimens: a review of what we have and what we need. *Infect Dis Therapy* 5(3):299–312
- Chahine EB, Sucher AJ, Hemstreet BA (2017) Sofosbuvir/velpatasvir: the first pangenotypic direct-acting antiviral combination for hepatitis C. *Ann Pharmacother* 51(1):44–53
- Coppola N, Alessio L, Onorato L, Sagnelli C, Macera M, Sagnelli E, Pisaturo M (2019) Epidemiology and management of hepatitis C virus infections in immigrant populations. *Infect Dis Poverty* 8(1):17
- Damme SV, Bultinck P (2007) A new computer program for QSAR-analysis: ARTE-QSAR. *J Comput Chem* 28(11):1924–948
- 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
- El-Kassem LA, Hawas UW, El-Souda S, Ahmed EF, El-Khateeb W, Fayad W (2019) Anti-HCV protease potential of endophytic fungi and cytotoxic activity. *Biocatal Agric Biotechnol* 19:101170
- Eriksson L, Jaworska J, Worth AP, Cronin MT, McDowell RM, Gramatica P (2003) Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs. *Environ Health Perspect* 111(10):1361–1375
- Evans DA (2014) History of the Harvard ChemDraw project. *Angew Chem Int Ed* 53(42):11140–11145
- Golbraikh A, Tropsha A (2002) Beware of q²! *J Mol Graph Model* 20(4):269–276
- González-Grande R, Jiménez-Pérez M, Arjona CG, Torres JM (2016) New approaches in the treatment of hepatitis C. *World J Gastroenterol* 22(4):1421
- Gramatica P, Corradi M, Consonni V (2000) Modelling and prediction of soil sorption coefficients of non-ionic organic pesticides by molecular descriptors. *Chemosphere* 41(5):763–777
- Huey R, Morris GM, Forli S (2012) Using AutoDock 4 and AutoDock vina with AutoDockTools: a tutorial. The Scripps Research Institute Molecular Graphics Laboratory, La Jolla
- Jia S, Zhou W, Wu J, Liu X, Guo S, Zhang J et al (2020) A biomolecular network-based strategy deciphers the underlying molecular mechanisms of Bupleuri Radix/Curcumae Radix medicine pair in the treatment of hepatitis C. *Eur J Integr Med* 33:101043
- Kucherenko A, Pampukha V, Romanchuk KY, Chernushyn SY, Bobrova I, Moroz L, Livshits L (2016) IFNL4 polymorphism as a predictor of chronic hepatitis C treatment efficiency in Ukrainian patients. *Cytol Genet* 50(5):330–333
- Li Z, Wan H, Shi Y, Ouyang P (2004) Personal experience with four kinds of chemical structure drawing software: review on ChemDraw, ChemWindow, ISIS/Draw, and ChemSketch. *J Chem Inf Comput Sci* 44(5):1886–1890
- Li F, Li X, Liu X, Zhang L, You L, Zhao J, Wu H (2011) Docking and 3D-QSAR studies on the Ah receptor binding affinities of polychlorinated biphenyls (PCBs), dibenzo-p-dioxins (PCDDs) and dibenzofurans (PCDFs). *Environ Toxicol Pharmacol* 32(3):478–485
- Liu M, Xu Q, Guo S, Zuo R, Hong Y, Luo Y et al (2018) Design, synthesis, and structure-activity relationships of novel imidazo [4, 5-c] pyridine derivatives as potent non-nucleoside inhibitors of hepatitis C virus NS5B. *Bioorg Med Chem* 26(9):2621–2631
- Neves BJ, Braga RC, Melo-Filho CC, Moreira Filho JT, Muratov EN, Andrade CH (2018) QSAR-based virtual screening: advances and applications in drug discovery. *Front Pharmacol* 9:1275
- Oluwaseye A, Uzairu A, Shallangwa GA, Abechi SE (2020) A novel QSAR model for designing, evaluating, and predicting the anti-MES activity of new 1H-pyrazole-5-carboxylic acid derivatives. *J Turk Chem Soc Sect A Chem* 4(3):739–774
- Petruzzello A, Marigliano S, Loquerio G, Cozzolino A, Ciacciapuoti C (2016) Global epidemiology of hepatitis C virus infection: An up-date of the distribution and circulation of hepatitis C virus genotypes. *World J Gastroenterol* 22(34):7824
- Poordad F, McCone J Jr, Bacon BR, Bruno S, Manns MP, Sulkowski MS et al (2011) Boceprevir for untreated chronic HCV genotype 1 infection. *N Engl J Med* 364(13):1195–1206
- Qin Z, Wang M, Yan A (2017) QSAR studies of the bioactivity of hepatitis C virus (HCV) NS3/4A protease inhibitors by multiple linear regression (MLR) and support vector machine (SVM). *Bioorg Med Chem Lett* 27(13):2931–2938
- Rabaan AA, Al-Ahmed SH, Bazzi AM, Alfouzan WA, Alsuliman SA, Aldrazi FA, Haque S (2019) Overview of hepatitis C infection, molecular biology, and new treatment. *J Infect Public Health* 13:773–783
- Rogers D (1997) Evolutionary statistics: using a genetic algorithm and model reduction to isolate alternate statistical hypotheses of experimental data. Paper presented at the ICGA

- Roy K, Mitra I, Kar S, Ojha PK, Das RN, Kabir H (2012) Comparative studies on some metrics for external validation of QSPR models. *J Chem Inf Model* 52(2):396–408
- Shao Y, Molnar LF, Jung Y, Kussmann J, Ochsenfeld C, Brown ST et al (2006) Advances in methods and algorithms in a modern quantum chemistry program package. *Phys Chem Chem Phys* 8(27):3172–3191
- Shi F, Zhang Y, Xu W (2015) Discovery of a series of novel compounds with moderate anti-hepatitis C virus NS3 protease activity in vitro. *Bioorg Med Chem* 23(17):5539–5545
- Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, Stapleton JT, Simmonds P (2014) Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. *Hepatology* 59(1):318–327
- Subedi R (2019) NS3/4A serine protease inhibitors for hepatitis C virus therapy (Telaprevir and Boceprevir). *EC Microbiol* 15:732–740
- Therese PJ, Manvar D, Kondepudi S, Battu MB, Sriram D, Basu A, 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
- Todeschini R, Consonni V (2008) *Handbook of molecular descriptors*, vol 11. Wiley, New York
- Tropsha A (2010) Best practices for QSAR model development, validation, and exploitation. *Mol Inf* 29(6–7):476–488
- Tropsha A, Gramatica P, Gombar VK (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb Sci* 22(1):69–77
- Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multi-threading. *J Comput Chem* 31(2):455–461
- Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK (2011) Validation of QSAR models-strategies and importance. *Int J Drug Des Discov* 3:511–519
- Vyas V, Jain A, Jain A, Gupta A (2008) Virtual screening: a fast tool for drug design. *Sci Pharm* 76(3):333–360
- World Health Organization (2018) Guidelines for the care and treatment of persons diagnosed with chronic hepatitis C virus infection
- Yap CW (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem* 32(7):1466–1474

Publisher's Note

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

Submit your manuscript to a SpringerOpen[®] 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)
