Research Article

Chemical Review and Letters
journal homepage: www.chemrevlett.com
ISSN (online): 2645-4947 (print) 2676-7279

Computational techniques in designing a series of 1,3,4-trisubstituted pyrazoles as unique hepatitis C virus entry inhibitors

Stephen Ejeh*, Adamu Uzairu, Gideon A. Shallangwa, Stephen E. Abechi

Department of Chemistry, Ahmadu Bello University P.M.B. 1044, Zaria-Nigeria

ARTICLE INFO

Article history:
Received 17 September 2020
Received in revised form 23 December 2020
Accepted 31 January 2021
Available online 7 May 2021

Keywords:
Silico design;
molecular docking; PaDEL-descriptors;
HCV entry inhibitors

ABSTRACT

In this study, we developed a QSAR model for studying the antiviral activity of 1,3,4-trisubstituted pyrazoles derivatives on hepatitis C virus infected in human HuH-7 cell lines. We employed random analysis to split the data sets. Statistically robust model was generated with R^2, Q^2 and R^2_{pred} values of 0.777, 0.731 and 0.774 respectively. The reliability of this model was confirmed by acceptable validation parameters, and this model also fulfilled the Golbraikh and Tropsha standard model conditions. Through the evaluation of selected molecular descriptors we observed that, topological charge index of order 4 (GGI4), mean topological charge index of order 1 (JGI1), octanol water partition coefficient (XlogP), 3D topological distance based autocorrelation lag5/weighted by polarizabilities (TDB5p) and total molecular surface area (FPSA-2) are the molecular properties determining biological activities of the study compounds, which shed light on the vital features that aid in the design of unique potent hepatitis C virus entry inhibitors using computer-aided drug design tools.

1. Introduction

Hepatitis C virus (HCV) is a universal health predicament that causes several life-threatening chronic diseases in the liver and the hepatitis C virus (HCV) is responsible for the diseases [1]. In many countries, HCV is indeed one of the primary sources of liver failure and liver transplant surgery which is a growing public health problem [2, 3]. The World Health Organization (WHO) estimated seventy-one million people were infected with HCV infection in 2015, representing one percent of the world's populace. 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 populace [4, 5].

Though some hepatitis viruses have vaccines, there are none available for HCV [2]. In the last few years, conventional interferon (IFN)-free treatment regimens in blend with ribavirin have been widely recognized as a model of excellence in antiviral therapy [6]. However, this might trigger patients to suffer 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 [2]. 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 revolution in the field of HCV treatment' [6, 7]. DAAs have culminated in a significantly improved tolerability and effectiveness compared with the traditional regimen of severe HCV infection [8]. The development of resistance-mechanism, like antibiotics, also encourages the viable discovery of new compounds or the modification of existing ones [2].

The strategy to a correlation existed between structure and activity is indeed very helpful for the estimation of biological activities, 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) [9 -11]. 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 tested empirically, and justify their selection. Such techniques are typically used to get hits (or leads) that seem to be more likely to offer good clinical candidates [9, 12, 13].

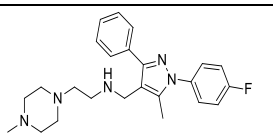
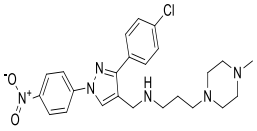
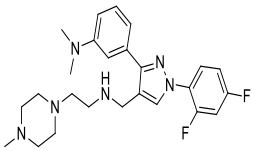
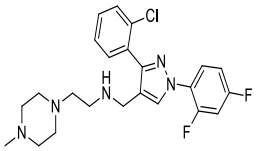
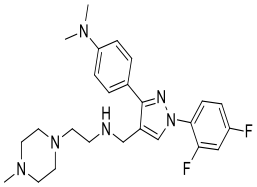
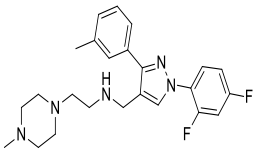
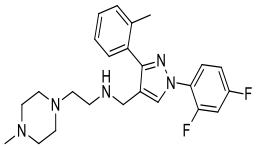
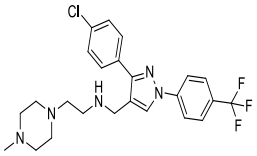
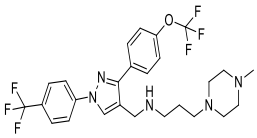
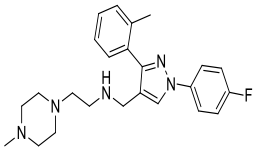
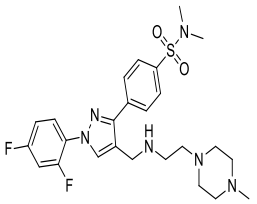
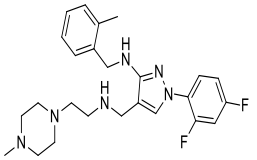
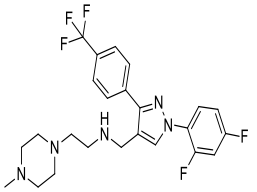
The earlier process of the drug development process is preceded by guesswork, and it is costly in terms of capital, time, and resources. However, with the introduction of computer-aided drug design strategies, the drug development and design process can be effectively carried out saving huge capital resources [14]. Over the random screening of existing chemical libraries, the ligand-based strategy has proven successful [15]. 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 development of novel and potent NS5B inhibitors were also explored by Therese et al., [16]. In the present research, computational methods were applied to generate a robust QSAR model and to use the data provided by the model to proposal new pyrazoles derivatives with high potency as hepatitis C virus entry inhibitors and to investigate the binding energy of designed molecules in comparison with an approved direct-acting antiviral agent (Boceprevir) through molecular docking.

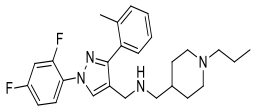
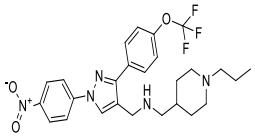
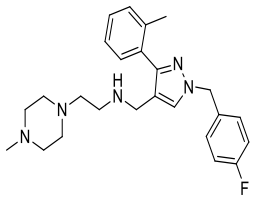
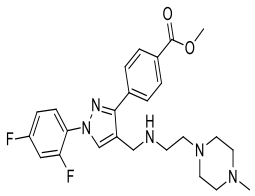
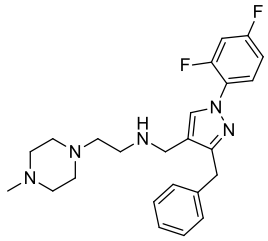
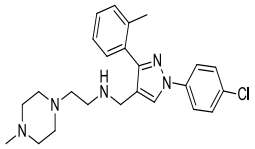
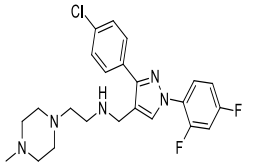
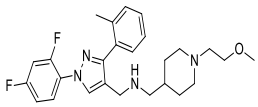
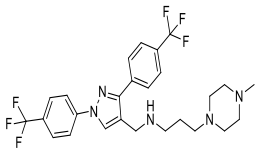
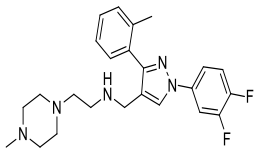
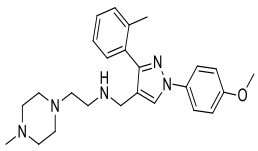
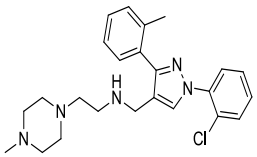
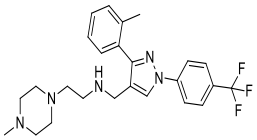
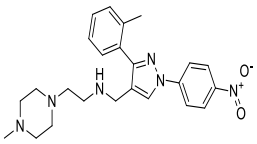
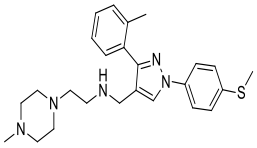
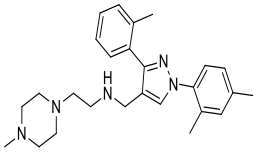
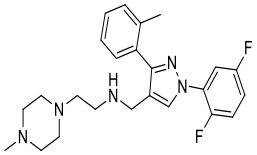
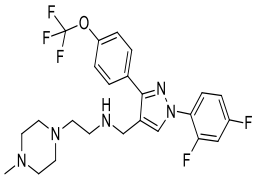
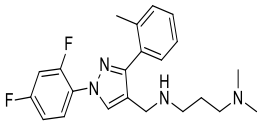
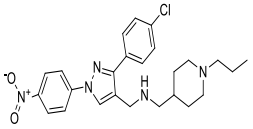
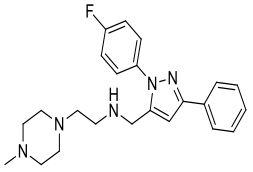
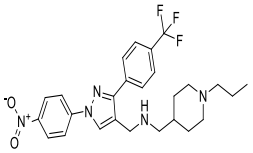
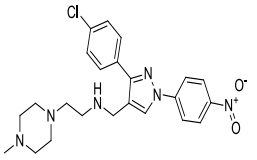
2. Materials and methods

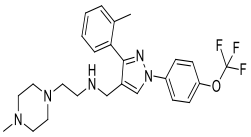
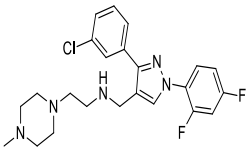
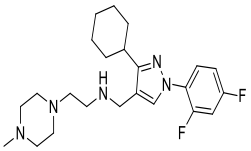
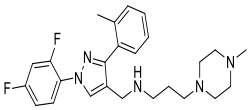
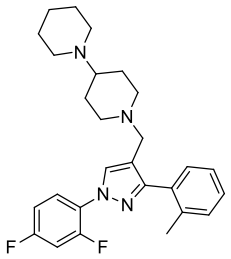
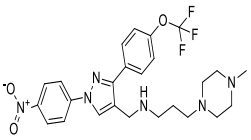
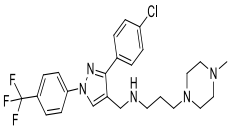
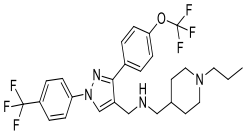
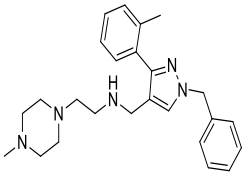
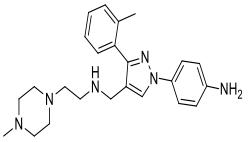
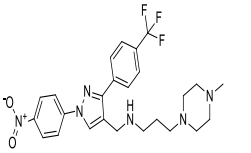
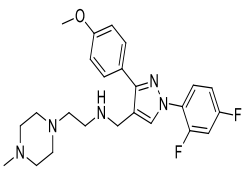
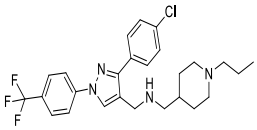
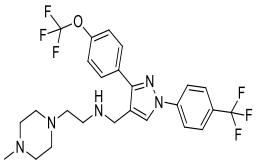
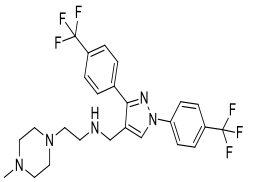
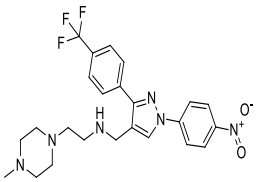
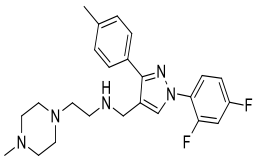
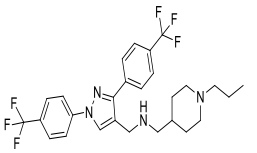
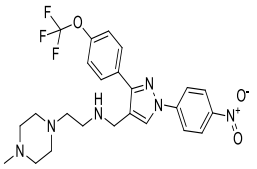
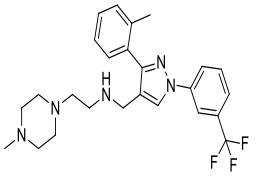
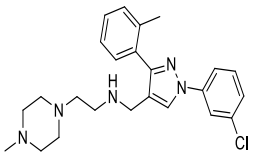
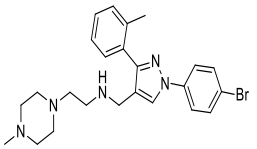
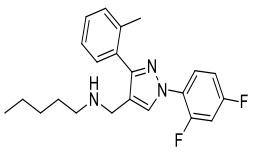
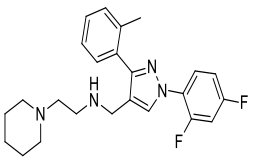
2.1. Dataset

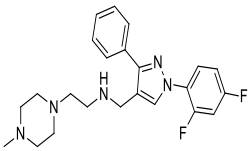
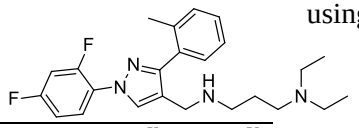
The dataset used for this study was 62 compounds of 1,3,4-trisubstituted pyrazoles derivatives retrieved from U.S. National Library of Medicine and National Center for Biotechnology Information which is accessible to the general public on the pubchem Web site (https://pubchem.ncbi.nlm.nih.gov/) as potent hepatitis C virus (HCV) entry inhibitor with pubchem AID: 781190 deposited on 3rd May 2014 and modified 28th September 2018 by CHEMBL (External ID: 990348) obtained as EC_{50} (μM) and was converted to pEC_{50} (pEC_{50} = -logIC_{50}), this is to reduce the skewness in the data [17]. The Pubchem CID and SID, chemical structure, experimental $EC_{50,}$ pEC_{50} and Predicted pEC_{50} of the datasets used in this research are reported in Table 1.

Table 1. Pubchem CID, pEC_{50}, predicted pEC_{50} and Chemical structure of the dataset

Entry	Pubchem CID	X	Y	Chemical Structure
C 1[*a]	72545083	5.1	5.5	
C 2	73346556	6.9	6.7	
C 3	72546770	5.5	5.4	
C 4	72546771	5.5	5.6	
C 5	72546531	5.9	5.6	
C 6[*]	72546533	5.8	5.6	
C 7	72545819	5.6	5.5	
C 8	73348042	6.4	6.6	
C 9	73348043	6.1	6.4	
C 10	72547539	5.4	5.6	
C 11[b]	72546534	5.7	5.7	
C 12[*]	72546773	5.4	5.3	
C 13	72546287	6.3	6.1	

C 14	72545573	5.6	5.5	
C 15	73349579	6.2	6.4	
C 16	72547543	5.0	5.9	
C 17	72546772	5.4	5.3	
C 18	72547034	4.8	5.6	
C 19	72547037	6.2	5.8	
C 20	72546285	6.4	6.1	
C 21	72545820	5.5	5.4	
C 22	73351139	6.4	6.6	
C 23	72547297	5.6	5.7	
C 24	72547538	5.6	5.4	
C 25	72547540	5.4	5.2	
C 26	72547035	6.3	6.2	
C 27*	72547036	6.3	6.2	
C 28	72547038	6.2	5.8	
C 29	72547295	5.6	5.7	
C 30*	72547296	5.6	5.5	
C 31	72546286	6.4	6.1	
C 32	16187217	5.2	5.3	
C 33	73352610	6.5	6.7	
C 34	72545084	4.9	5.6	
C 35*	73352611	6.8	6.6	
C 36	73352612	6.5	6.7	

C 37	72547039	6.0	6.2	
C 38	72546288	6.2	6.0	
C 39	72546530	5.9	5.5	
C 40	72545572	5.7	5.5	
C 41	72545822	5.1	5.2	
C 42	73354107	6.5	6.5	
C 43	73354108	6.4	6.6	
C 44	73354109	6.4	6.3	
C 45	72547541	5.1	5.3	
C 46	72547542	5.0	5.5	
C 47	73354110	6.9	6.7	
C 48*	72546769	5.7	5.4	
C 49	73355627	6.5	6.6	
C 50*	73355628	6.4	6.5	
C 51*	73355629	6.3	6.6	
C 52	73355631	7.0	6.6	
C 53*	72546529	6.1	5.7	
C 54	73357174	6.5	6.6	
C 55	73357175	6.4	6.5	
C 56*	72547292	6.0	5.9	
C 57	72547293	6.0	5.5	
C 58	72547294	5.9	5.8	
C 59[b]	72545823	4.7	4.8	
C 60	72546284	4.7	4.8	

C 61*	72546532	5.8	5.7	
C 62	72545821	5.4	5.3	

* Test Set, [a] Outliers, [b] Influential molecules, [X] pEC_{50}, [Y] predicted pEC_{50}

2.2. 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 [18]. The optimized structures are then transferred to another software (PaDEL-Descriptors), which computed the structural properties (molecular descriptors) for each molecule [19].

2.3. Dataset division

In the current analysis, the dataset was mainly split into two parts containing 80% (Training set) for constructing the model and 20% (Test set) which is unused during model construction but was used in the determination of the model's predictive ability [17].

2.4. Model generation

The correlation analysis was achieved by Material Studio software and Genetic Function Approximation (GFA) was applied in the process to define the ideal QSAR models. In regression analysis, X (descriptors) relies on the conditional value of predictor variables Y (pIC_{50}) [11]. GFA is the technique used to generate statistical data models using the evolution process. Substituting regression study further with 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 [20].

2.5. 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:

$$Q_{CV}^2 = 1 - \frac{PRESS}{\sum_{i=1}^{n}(y_{obs} - \bar{y})^2}$$

From the above equation, PRESS is the predicted error sum of square computed by

$$PRESS = \sum \left(y_{obs} - y_{pred}\right)^2$$

y_{obs}, y_{pred} and $\bar{y}$ represents the observed, predicted, and mean values of observed activity respectively.

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

$$R_{pred}^2 = 1 - \frac{\sum\left(y_{obs(Test)} - y_{pred(Test)}\right)^2}{\sum\left(y_{obs(Test)} - \bar{y}_{Training}\right)^2}$$

$y_{obs(Test)}$ and $y_{pred(Test)}$ represent observed and predicted activity data for the test set molecules, and $\bar{y}_{Training}$ represents the mean observed activity of the training set.

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

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

were 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 multiple correlation coefficient of one descriptor's effect regressed over the remaining molecular descriptors in the model [21].

Every 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:

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

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 [9].

2.6. Model Applicability Domain (AD)

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

$$h_i = X_i(X^T X_i)^{-1} X_i^T \quad (i = 1, 2, \ldots, n)$$

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

For which X_i is the row-vector descriptor of the request ith sample, it is the distinctive vector of the training set, n is the number of request samples and p is the number of selected variables (descriptors) in the model [9, 22, 23].

The standardized residual (SDR) of the model AD is estimated by the equation:

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

In which Y is the observed activity value for whichever 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 < SDR < 3 boundaries. Consequently, whichever molecule by means of SDR less than -3 or greater than +3 are labeled an outlier in the variable response area, as well as any molecule with leverage value larger than h^*, is labeled an influential molecule foreign to the most molecules used during model construction.

2.7. Docking studies

2.7.1. Ligand structure preparation

ChemBio Ultra 12.0 was used to draw 2D Ligand structures [24, 25]. 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 [26].

2.7.2. Protein Structure Preparation

The crystal structure of HCV NS5B polymerase was taken from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) with the HCV NS5B polymerase structural PDB ID being 2XWY. The co-crystallized ligand MK-3281, a Potent Non-Nucleoside Finger-Loop inhibitor in complex with the crystal structure of HCV NS5B Polymerase was discarded, hydrogen atoms were introduced, lower occupancy residue structures were discarded, partials side chains were substituted with the use of discovery studio [27]. The structures were then saved in PDB format for input into AutoDock version 1.5.6. in PyRx software [26].

2.7.3. Docking Procedure and evaluation

A rectangular grid measuring 65.5217 × 72.7141 × 80.3011 Å, centered on 5.2017, 15.6939, 30.8304 was built across the binding site of ligand on HCV NS5B polymerase using autodock tools. The grid center was set 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 conformations of 4 falling from bonded X-ray verification for HCV NS5B polymerase root-mean-square deviation (rmsd) standards of 0.71 to 0.74 Å, proposing that this process is sufficiently valid to be enough for docking studies of other molecules. The results were exported for thorough observation of the binding relationships and correlations between the molecules and amino acid residues at the active spots using discovery-studio software [28].

2. Results and Discussion

In the present study, we developed QSAR model for the antiviral activity of 1,3,4-trisubstituted pyrazoles derivatives on hepatitis C virus infected in human HuH-7 cell lines using 62 datasets and the QSAR model is presented as:

$$\boldsymbol{pEC_{50}} = 7.55735 + 0.29138 * \mathbf{GGI4} + 10.90034 * \boldsymbol{JGI1} - 0.17042 * \boldsymbol{XLogP} - 0.64795 * \boldsymbol{TDB5p} - 2.9115 * \boldsymbol{FPSA-2}$$

$N_{train} = 50$, $R^2_{train} = 0.777$, K = 5, $Q^2_{LOO(train)} = 0.731$, $N_{test} = 12$, $R^2_{test} = 0.774$, Outliers > $\pm$3.0 = 1, Influential molecules > h* = 2

Where N_{train} and N_{test} are sums of data present in the training and test set respectively, R^2_{train} and R^2_{test} are the coefficients of correlation for internal and external validation respectively, Q^2_{LOO} is the squared cross-validation coefficients for leave one out, and K is the predictor parameters (descriptors) present in the model. The developed model explains seventy-eight percent (78%) and predicts seventy-seven percent (77%) of the variances of the 1,3,4-trisubstituted pyrazoles derivatives activity on hepatitis C virus infected in human HuH-7 cell lines. The descriptions of the descriptors used in the model, the computed mean effect (MF) and the computed Variance Inflation Factor (VIF) of each descriptor present in the model were stated in Table 2.

Table 2: A description of the descriptor used in the model, the MF and VIF

E	Description	Descriptor	C	MF	VIF
1	Topological charge index of order 4	GGI4	2D	2.9	1.8
2	Mean topological charge index of order 1	JGI1	2D	1.2	1.8
3	XLogP	XLogP	2D	0.3	1.4
4	3D topological distance based autocorrelation - lag 5 / weighted by polarizabilities	TDB5p	3D	2.3	1.7
5	PPSA-2 / total molecular surface area	FPSA-2	3D	2.6	1.9

[E] Entry, [C] Class of descriptor, [MF] Mean Effect, [VIF] Variance Inflation Factor

The mean effect (MF) value offers significant details on the impact of the model's molecular descriptors, the size of these descriptors MF show their intensity in manipulating the activities of the study compounds and was graphically represented using histogram as shown in Figure 1 and the contribution of the descriptors to the activity of the study compounds was observed to be in increasing order of XLogP < JGI1< TDB5p < FPSA-2 < GGI4 (see Figure 1). From Table 2, it was detected that all the descriptors have VIF values of less than 5, which means that the model obtained has statistical significance and that the descriptors were considered to be fairly orthogonal [22].

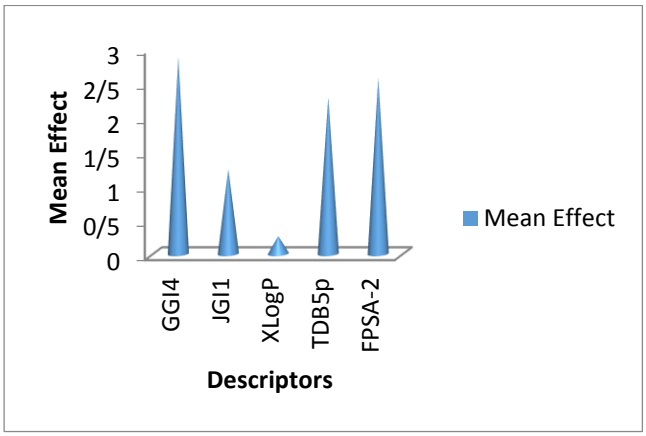


Figure 1: Plot of molecular descriptors against Mean Effect (descriptor's contribution)

Table 3 represent the minimum recommended values of validated parameters for generally acceptable QSAR model reported by Golbraikh and Tropsha [29] and the validated parameters of the model reported in this research.

Table 3: Minimum recommended values of validated parameters for generally acceptable QSAR model and the validated parameters of the model reported in this research.

Symbol	Name	Threshold value	Model value	Comments	Ident
R^2	Coefficient of determination	≥ 0.6	0.777	Passed	a
Q^2	Cross-validation coefficient	>0.5	0.731	Passed	a
R^2_{ext}	Coefficient of determination for external test set	≥ 0.6	0.774	Passed	a
cR^2_p	y-randomization test value	>0.5	0.721	Passed	b
$N_{test\ set}$	Minimum number of external test set	≥ 5	12	Passed	a
$P_{(95\%)}$	Confidence interval at 95% confidence level	< 0.05	0.0005	Passed	a

[a] [29], [b] [14]

The model statistics described in Table 3 meet the criteria for validating a QSAR model developed by OECD [11, 15]. The findings show that the R^2 and Q^2 for the model's internal evaluation have been stated as 0.777 and 0.731 respectively. It indications that the model correctly interpreted the data when regressed and that the model can estimate the fitted training set, as the model predicted approximately 78% of the data and thus met the minimum condition of 60% [11]. The y-randomization test computed shows that the value for the model's random R^2 (cR^2_p = 0.721, see Table 3) is substantially higher than the recommended value of 0.50, meaning the model is not the product of simple chance alone [14].

Qin, Wang, and Yan 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 R^2 values for internally and externally evaluation were respectively 0.75 and 0.72, [30] which seem to be close in values compared to R^2 values for internally and externally evaluation of 0.777 and 0.774 respectively as reported in this paper.

The square area 0 < hi < h* and -3 < SDR < 3 represented the model AD using Williams's plot (see Figure 2). Where h* (0.36) boundary is the model cautioning leverage and SDR is the standardized residual of the models. The outcome shows that 96.8% of the molecules considered were inside the AD of the model while 1.6% formed the Outliers which is compound 1 in Table 1 as identified and indicated in Figure 2 with SDR of 3.4 and 1.6% of the studied molecules formed the influential molecule which is compound 11 in Table 1 as identified and indicated in Figure 2 with leverage value of 0.43 greater than cautioning leverage (i.e 0.43> h*=0.69). 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 3 presented a plot of the model predicted against experimental anti-hepatitis C activity values for both the training and test sets and it shows that there are a strong correlation between the models observed and estimated activity values for both training and test sets data. Such results showed that the models had the high predictive potential both internally and externally and were free of systemic bias [31]. Consequently, they could be used to predict known molecules lacking activity, as

long as the molecule is inside the AD of the model. The structure of the template molecule which is compound 47 in Table 1 is presented in Figure 4.

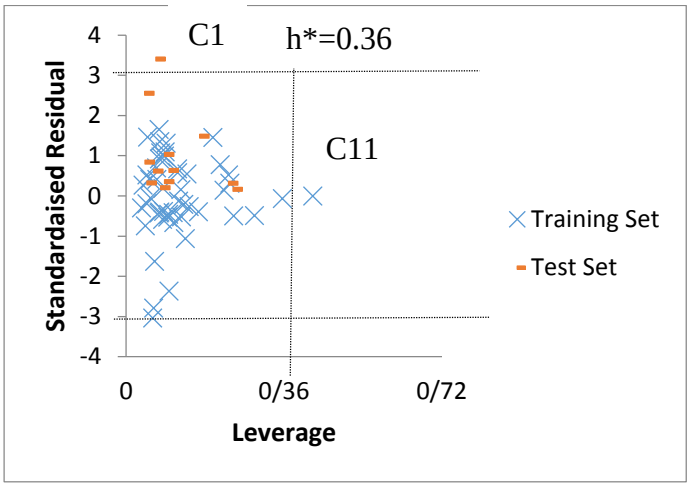


Figure 2: The model Applicable Domain plot (Williams plot)

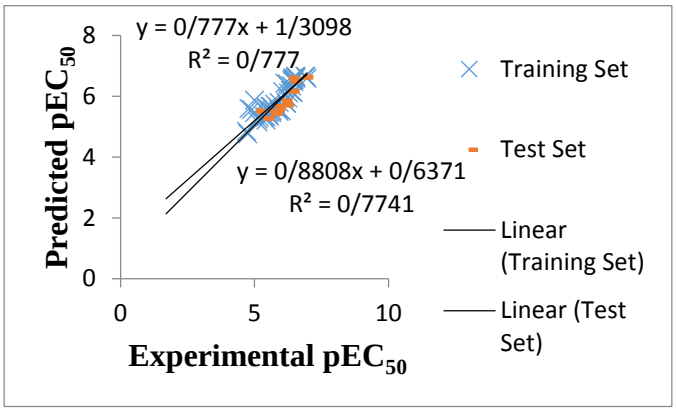


Figure 3: Plot of the model predicted against experimental anti-hepatitis C virus activities values for both Training and Test sets.

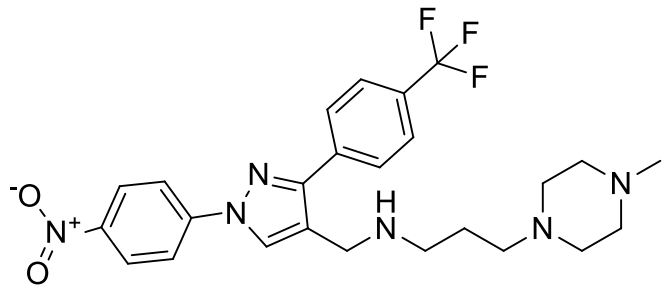


Figure 4: The structure of the hit molecule, see Table 1, M 47 (3-(4-methylpiperazin-1-yl)-N-[[1-(4-nitrophenyl)-3-[4-(trifluoromethyl)phenyl]pyrazol-4-yl]methyl]propan-1-amine)

3.1. Interpretation of descriptors used in the developed QSAR model

GGI4 is a Topological Charge Descriptor and is the first selected descriptor in the established model which defined the topological charge index of order 4. GGI4 measured the charge transfer among atoms. The positive sign of this descriptor in derived linear equation indicates that increasing the charge transfer among the pair of atoms would result in increase of the pEC_{50} values and the more active the molecule, also decreasing the charge transfer among the pair of atoms would result in decrease of the pEC_{50} values respectively [32].

The second selected descriptor in the established model is the mean topological charge index of order 1 (JGI1) [32], and it shows a positive sign in derived linear equation which indicates positive impact on the human HuH-7 cell line inhibition. The JGI1 can be increased by adding high-electronegativity molecules to the ring of the lead molecule (3-(4-methylpiperazin-1-yl)-N-[[1-(4-nitrophenyl)-3-[4-(trifluoromethyl)phenyl]pyrazol-4-yl]methyl]propan-1-amine) example molecule 2, 5 and 6 in Table 4.

The third selected descriptor in the established model is the Wang octanol water partition coefficient (XlogP). It describes lipophilicity of compounds, which expressed the ability to penetrate lipid-rich zones from aqueous solutions [33]. The negative sign in derived linear equation for this descriptor indicates that an increase in the activity is observed with a decrease in its lipophilicity.

TDB5p is the fourth selected descriptor in the established model and is an autocorrelation 3D descriptor which is defined as 3D topological distance based autocorrelation lag 5/weighted by polarizabilities.TDB5p measures the strength of the connection between atomic charges 5 bonds apart [32]. The negative sign in the derived linear equation for this descriptor indicates that an increase in the activity is observed with a decrease in the strength of the connection between atomic charges 5 bonds apart.

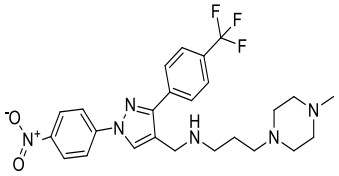
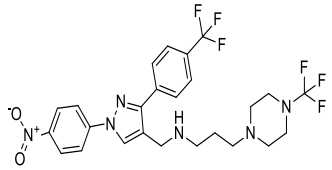
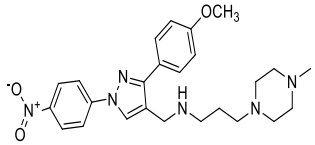
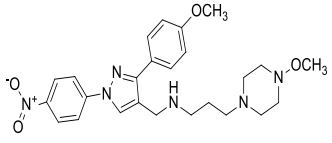
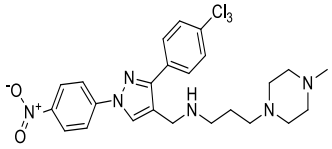
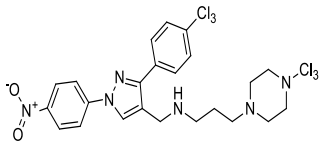
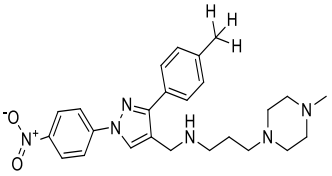
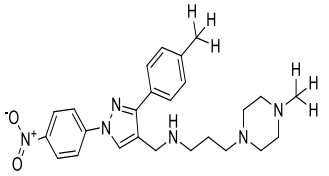
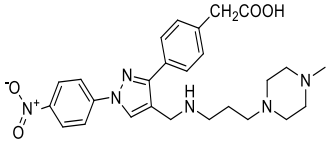
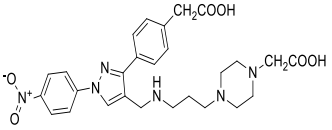
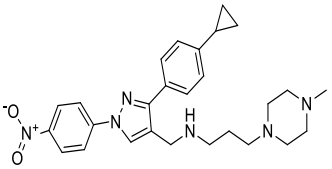
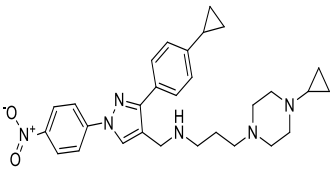
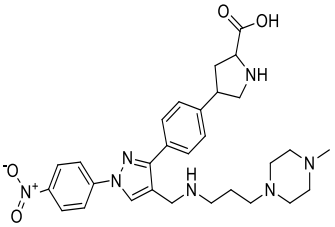
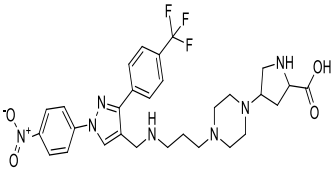
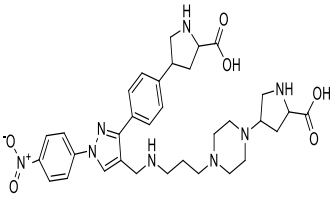
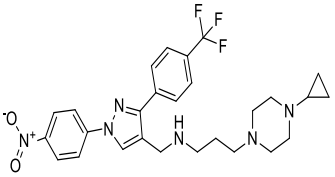
The fifth and final selected descriptor in the established model is the total molecular surface area (FPSA-2). FPSA-2 is extensively considered in the improvement of cellular permeation of a drug and it exemplifies the area due to nitrogen and oxygen and any attached hydrogen [34-36]. The negative sign in the derived linear equation for this descriptor indicates that an increase in the activity is observed with a decrease in the total molecular surface area of the observed molecules. This shows that nitrogen and oxygen atoms have a notable negative impact on observed anti-HCV activity.

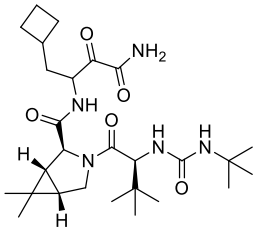
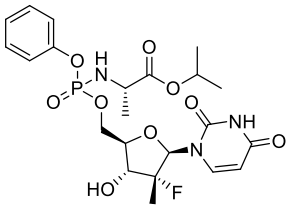
3.2. New compound design and activity prediction

Based on the built QSAR model and related results analysis and descriptors Interpretation, the compound 47 in Table 1 shown in Figure 4 (3-(4-methylpiperazin-1-yl)-N-[[1-(4-nitrophenyl)-3-[4-(trifluoromethyl)phenyl]pyrazol-4-yl]methyl]propan-1-amine) was used as a template for designing new molecules by molecular structure modification. Molecule 47 was used as a template for designing unique molecules because it was carefully chosen from the Williams plot

by identifying the molecule that is more active (having higher pEC_{50}), low standardized residual, and low leverage i.e. was found within the applicability domain of the built model. The previously established QSAR model was used to predict the activity of the template molecule, newly designed molecules, and approved direct-acting antiviral agents (Sofosbuvir and Boceprevir). The result shows that some of the designed molecules and Boceprevir (approved direct-acting antiviral agents) have improved activity values compared to template molecule except molecule 3, 4, 7, 8, 9, 10, 11, 12, 13, and 15 (see Table 4). Also, molecule 2 in Table 4 is the most active compound (pEC50 = 7.83) because the better the value of the pEC50, the more active the compound. The structure of the template, newly designed molecules, and approved direct-acting antiviral agents (Sofosbuvir and Boceprevir) together with their predicted activity and Leverages are presented in Table 4. The majority of the molecules shows good leverage (lower leverage than the threshold ($h^* = 0.36$). These implied that the majority of the designed molecules, as well as approved direct-acting antiviral agents (Boceprevir), are within the applicability domain of the model.

Table 4: Template molecule, Designed molecules and approved direct-acting antiviral agents (Sofosbuvir and Boceprevir) with their predicted pEC_{50} and leverages

Molecule	Structure	Predicted pEC_{50}	Leverages
1[x]		6.66	0.08
2[yy]		7.83	0.15
3		6.21	0.14
4		5.92	0.36
5[y]		6.97	0.39
6[y]		6.83	1.36
7		6.40	0.13
8		6.39	0.12
9		5.85	0.20
10		6.02	0.68
11		5.94	0.15
12		5.99	0.37
13		5.66	0.76
14[y]		6.85	0.86
15		5.41	3.15
16		6.49	0.16

17[zz]		6.74	0.31
18[z]		4.63	2.91

[x] Template molecule, [y] designed molecules with improved activity, [yy] designed molecule with better activity compared to approved direct-acting antiviral agents and are within the model AD, [z] Sofosbuvir (approved direct-acting antiviral agents), [zz] approved direct-acting antiviral agents that is within the model AD (Boceprevir)

3.3. Molecular docking results and analysis

Among all the molecules in Table 4 including Sofosbuvir and Boceprevir (approved direct-acting antiviral agents) it was observed that molecule 2 has the highest predicted pEC_{50} value (7.83) and hence was subjected to a molecular docking study. Also, Boceprevir (approved direct-acting antiviral agent that is within the AD of the model) with predicted pEC_{50} of 6.74 is subjected to a similar molecular docking study for comparison. The results of the molecular docking study such as Binding Energy (kCal/mol), Interactions with amino acid, Types of Interaction, Bond length (Å) of the designed molecule (molecule 2 in Table 4) and the approved direct-acting antiviral agent (Boceprevir, molecule 17 in Table 4) was reported in Table 5, while figure 5 and 6 show the 3D and 2D interaction of the designed molecule (molecule 2 in Table 4), and the approved direct-acting antiviral agents (molecule 17 in Table 4) with the binding pocket of the 3D crystal structure of Hepatitis C Virus NS5B polymerase (PDB ID: 2XWY) respectively. As shown in Figure 5, we observed that the designed molecule is well-superimposed in the active pocket of the receptor this is because it has the highest activity and the lowest binding energy (7.83 and -7.5) compared to molecule 17 (6.74 and -6.5).

Figure 5 shows that ASN291, VAL284, LYS172, LEU285, SER96, GLY449, GLY283, ARG168, GLU171, THR287, ARG168 and PRO93 are the residues of the target receptor involved in the interaction with a designed molecule (molecule 2 in Table 4), while Figure 6 shows that ASP220, ASP318, ASN291, PRO558 and ARG158 are the residues of the target receptor involved in the interaction with the approved direct-acting antiviral agents (molecule 17 in Table 4). It was observed from the docking results presented in Table 5 that the target residue ASN, PRO and ARG are involved in the interaction with all the docked molecules. This implies the importance of this residue in the inhibition of HCV NS5B polymerase. It was also observed that molecule 2 interacted better with the target receptor when compared to the approved direct-acting antiviral agent.

Table 5: Docking results of designed molecule with the highest activity and Boceprevir (approved direct-acting antiviral agent)

Molecules from Table 4	Binding Energy kCal/mol	Residues interacting with Ligand	Types of Interaction	Bond length (Å)
2	-7.5	ASN291, VAL284, LYS172, LEU285	Conventional Hydrogen Bond	3.08, 3.05, 3.35, 3.59
		SER96, GLY449	Carbon Hydrogen Bond	3.65, 3.54
		GLY283, ARG168, GLU171	Halogen (Fluorine)	3.62, 3.28, 3.17
		THR287	Pi-Sigma	3.58
		ARG168, PRO93	Alkyl	4.34, 5.45
17	-6.5	ASP220, ASP318, ASN291	Conventional Hydrogen Bond	2.41, 2.48, 2.42
		PRO558	Unfavorable Donor-Donor	2.01
		ARG158	Alkyl	5.32

The 3D and 2D interactions of the designed molecule (molecule 2 in Table 4), and the approved direct-acting antiviral agent (Boceprevir which is molecule 17 in Table 4) with the binding pocket of 3D crystal structure of Hepatitis C Virus NS5B polymerase (PDB ID: 2XWY) are presented in Figures 5 and 6 respectively.

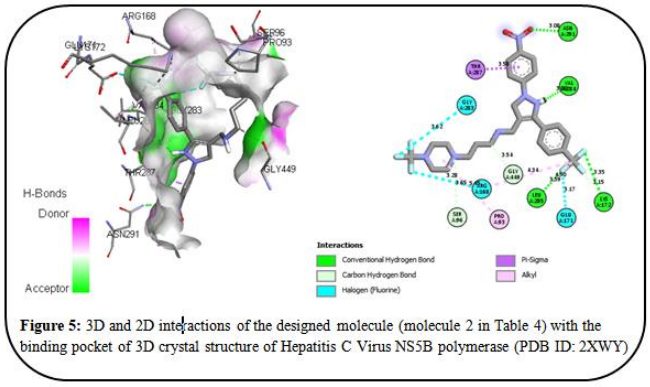


Figure 5: 3D and 2D interactions of the designed molecule (molecule 2 in Table 4) with the binding pocket of 3D crystal structure of Hepatitis C Virus NS5B polymerase (PDB ID: 2XWY)

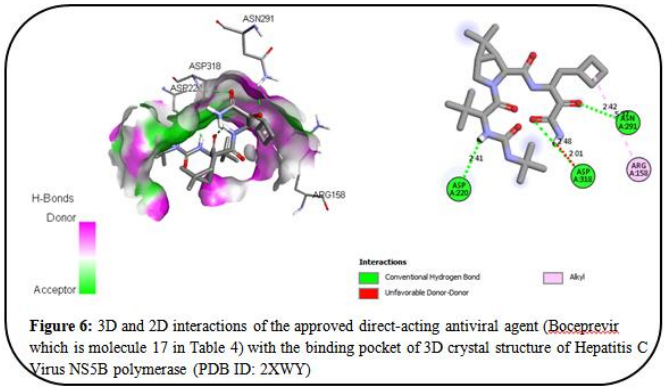


Figure 6: 3D and 2D interactions of the approved direct-acting antiviral agent (Boceprevir which is molecule 17 in Table 4) with the binding pocket of 3D crystal structure of Hepatitis C Virus NS5B polymerase (PDB ID: 2XWY)

4. Conclusion

The statistically validated QSAR model obtained provided rationales to explain the anti-hepatitis-C virus activities of 1,3,4-trisubstituted pyrazoles derivatives. The model is statistically robust with both internal and external validation R^2 values of 0.777 and 0.774 respectively and met the conditions for acceptable QSAR model proposed by different groups. Through the evaluation of selected molecular descriptors we observed that, topological charge index of order 4 (GGI4), mean topological charge index of order 1 (JGI1), octanol water partition coefficient (XlogP), 3D topological distance based autocorrelation lag5/weighted by polarizabilities (TDB5p) and total molecular surface area (FPSA-2) are the molecular properties determining biological activities of the study compounds, which shed light on the vital features that aid in the design of unique potent hepatitis C virus entry inhibitors using computer-aided drug design tools. Few new compounds having better anti-hepatitis-C virus activity than the highest active compound in the data set (compound 47 in Table 1), have been suggested for further exploration. The binding affinity (-7.5) of this newly identified molecule docked into the binding pocket of the crystal structure of Hepatitis C Virus NS5B polymerase (PDB ID: 2XWY) were found to be better than that of Boceprevir (approved direct-acting antiviral agent) which is -6.5. Hence, a novel molecule was identified showing high potency as hepatitis C virus entry inhibitors.

Acknowledgments

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.

.

References

[1] S. Jia, W. Zhou, J. Wu, X. Liu, S. Guo, J. Zhang, and X. Zhang, A biomolecular network-based strategy deciphers the underlying molecular mechanisms of Bupleuri Radix/Curcumae Radix medicine pair in the treatment of hepatitis C. *European Journal of Integrative Medicine*, *33* (2020) 101043.

[2] L. A. El-Kassem, U. W. Hawas, S. El-Souda, E. F. Ahmed, W. El-Khateeb, and W. Fayad, Anti-HCV protease potential of endophytic fungi and cytotoxic activity. *Biocatalysis and Agricultural Biotechnology*, *19* (2019) 101170.

[3] R. González-Grande, M. Jiménez-Pérez, C. G. Arjona, and J. M. Torres, New approaches in the treatment of hepatitis C. *World journal of gastroenterology*, *22*(4) (2016) 1421.

[4] A. Kucherenko, V. Pampukha, K. Y. Romanchuk, S. Y. Chernushyn, I. Bobrova, L. Moroz, and L. Livshits, IFNL4 polymorphism as a predictor of chronic hepatitis C treatment efficiency in Ukrainian patients. *Cytology and Genetics*, *50*(5) (2016), 330-333

[5] World Health Organization; Guidelines for the care and treatment of persons diagnosed with chronic hepatitis C virus infection. Geneva. Licence: CC BY-NC-SA 3.0 IGO (2018)

[6] M. Liu, Q. Xu, S. Guo, R. Zuo, Y. Hong, Y. Luo, and Y. Liu, Design, synthesis, and structure-activity relationships of novel imidazo [4, 5-c] pyridine derivatives as potent non-nucleoside inhibitors of hepatitis C virus NS5B. *Bioorganic & medicinal chemistry*, *26*(9) (2018) 2621-2631.

[7] F. Poordad, Jr. J. McCone, B. R. Bacon, S. Bruno, M. P. Manns, M. S. Sulkowski, and N. Boparai, Boceprevir for untreated chronic HCV genotype 1 infection. *New England Journal of Medicine*, *364*(13) (2011) 1195-1206.

[8] M. R. Bidell, M. McLaughlin, J. Faragon, C. Morse, and N. Patel, Desirable characteristics of hepatitis C treatment regimens: a review of what we have and what we need. *Infectious diseases and therapy*, *5*(3) (2016) 299-312.

[9] D. E. Arthur, S. Ejeh, and A. Uzairu, 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. *Journal of Receptors and Signal Transduction*, (2020) 1-20.

[10] K. S. Bhadoriya, M. C. Sharma, and S. V. Jain, 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.

Journal of Taibah University for Science, *9*(4) (2015) 521-530.

[11] R. Veerasamy, H. Rajak, A. Jain, S. Sivadasan, C. P. Varghese, and R. K. Agrawal, Validation of QSAR models-strategies and importance. *International Journal of Drug Design & Discovery*, *3*, (2011) 511-519.

[12] B. J. Neves, R. C. Braga, C. C. Melo-Filho, J. T. Moreira Filho, E. N. Muratov, and C. H. Andrade, QSAR-based virtual screening: advances and applications in drug discovery. *Frontiers in pharmacology*, *9*, (2018) 1275.

[13] V. Vyas, A. Jain, and A. Gupta, Virtual screening: a fast tool for drug design. *Scientia Pharmaceutica*, *76*(3), (2008) 333-360.

[14] D. E. Arthur, A. Uzairu, P. Mamza, and S. Abechi, Quantitative structure–activity relationship study on potent anticancer compounds against MOLT-4 and P388 leukemia cell lines. *Journal of Advanced Research*, *7*(5), (2016) 823-837.

[15] K. Roy, I. Mitra, S. Kar, P. K. Ojha, R. N. Das, and H. Kabir, Comparative studies on some metrics for external validation of QSPR models. *Journal of chemical information and modeling*, *52*(2), (2012) 396-408.

[16] P. J.Therese, D. Manvar, S. Kondepudi, M. B. Battu, D. Sriram, A. Basu, and N. Kaushik-Basu, Multiple e-pharmacophore modeling, 3D-QSAR, and high-throughput virtual screening of hepatitis C virus NS5B polymerase inhibitors. *Journal of chemical information and modeling*, *54*(2) (2014) 539-552.

[17] A. Tropsha, Best practices for QSAR model development, validation, and exploitation. *Molecular informatics*, *29*(6- 7), (2010) 476-488.

[18] Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, and D. P. O'Neill, Advances in methods and algorithms in a modern quantum chemistry program package. *Physical Chemistry Chemical Physics*, *8*(27) (2006) 3172-3191.

[19] C. W. Yap, PaDEL- descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of computational chemistry*, *32*(7) (2011) 1466-1474.

[20] D. Rogers, *Evolutionary Statistics: Using a Genetic Algorithm and Model Reduction to Isolate Alternate Statistical Hypotheses of Experimental Data*. Paper presented at the ICGA (1997).

[21] A. Beheshti, E. Pourbasheer, M. Nekoei, and S. Vahdani, QSAR modeling of antimalarial activity of urea derivatives using genetic algorithm–multiple linear regressions. *Journal of Saudi Chemical Society*, *20*(3) (2016) 282-290.

[22] L. Eriksson, J. Jaworska, A. P. Worth, M. T. Cronin, R. M. McDowell, and P. Gramatica, Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs. *Environmental health perspectives*, *111*(10) (2003) 1361-1375.

[23] F. Li, X. Li, X. Liu, L. Zhang, L. You, J. Zhao, and H. Wu, Docking and 3D-QSAR studies on the Ah receptor binding affinities of polychlorinated biphenyls (PCBs), dibenzo-p-dioxins (PCDDs) and dibenzofurans (PCDFs). *environmental toxicology and pharmacology*, *32*(3) (2011) 478-485.

[24] D. A. Evans, History of the Harvard ChemDraw project. *Angewandte Chemie International Edition*, *53*(42) (2014) 11140-11145.

[25] Z. Li, H. Wan, Y. Shi, and P. Ouyang, Personal experience with four kinds of chemical structure drawing software: review on ChemDraw, ChemWindow, ISIS/Draw, and ChemSketch. *Journal of Chemical Information and Computer Sciences*, *44*(5) (2004) 1886-1890.

[26] R. Huey, G. M. Morris, and S. Forli, Using AutoDock 4 and AutoDock Vina with AutoDockTools: A Tutorial. *The Scripps Research Institute Molecular Graphics Laboratory*, (2012)

[27] M. Danishuddin, S. N. Khan, and A. U. Khan, Molecular interactions between mitochondrial membrane proteins and the C-terminal domain of PB1-F2: an in silico approach. *Journal of molecular modeling*, 16(3) (2010) 535-541.

[28] O. Trott, and A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2) (2010) 455-461.

[29] A. Golbraikh, and A. Tropsha, Beware of q2! *Journal of molecular graphics and modelling*, *20*(4) (2002) 269-276.

[30] Z. Qin, M. Wang, and A. Yan, QSAR studies of the bioactivity of hepatitis C virus (HCV) NS3/4A protease inhibitors by multiple linear regression (MLR) and support vector machine (SVM). *Bioorganic & Medicinal Chemistry Letters*, *27*(13) (2017) 2931-2938.

[31] S. N. Adawara, G. A. Shallangwa, P. A. Mamza, A. Ibrahim, QSAR Model for Prediction of some Non-Nucleoside inhibitors of Dengue virus serotype 4 NS5 Using GFA-MLR Approach. *J. Chem. Lett.* 1 (2020) 69-76

[32] R. Todeschini, and V. Consonni, *Handbook of molecular descriptors* (Vol. 11): John Wiley & Sons (2008).

[33] R. Wang, Y. Gao, and L. Lai, Calculating partition coefficient by atom additive method, *Perspectives in Drug Discovery and Design*, 19 (2000) 47–66.

[34] P. Stenberg, K. Luthman, H. Ellens, C. P. Lee, P. L. Smith, A. Lago, J. D. Elliot, and P. Artursson, Prediction of the intestinal absorption of endothelin receptor antagonists using three theoretical methods of increasing complexity. *Pharm. Res*, 16 (1999) 1520–1526. doi:10.1023/A:1015092201811

[35] D. E. Clark, Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 1. Prediction of intestinal absorption. *J. Pharm. Sci*, 88 (1999) 807–816.

[36] M. R. J. Sarvestani, S. Majedi, A DFT study on the interaction of alprazolam with fullerene (C_{20}). *J. Chem. Lett.* 1 (2020) 32-38.