



HAL
open science

In silico design and analysis of NS4B inhibitors against hepatitis C virus

Ismail Hdoufane, Imane Bjj, Mehdi Oubahmane, Mahmoud E S Soliman,
Didier Villemin, Driss Cherqaoui

► **To cite this version:**

Ismail Hdoufane, Imane Bjj, Mehdi Oubahmane, Mahmoud E S Soliman, Didier Villemin, et al.. In silico design and analysis of NS4B inhibitors against hepatitis C virus. *Journal of Biomolecular Structure and Dynamics*, 2022, 40 (5), pp.1915-1929. 10.1080/07391102.2020.1839561 . hal-03015951

HAL Id: hal-03015951






<https://normandie-univ.hal.science/hal-03015951>

Submitted on 20 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

In silico design and analysis of NS4B inhibitors against hepatitis C virus

Ismail Hdoufane^a , Imane Bjjj^{a,b} , Mehdi Oubahmane^a , Mahmoud E.S Soliman^b ,
Didier Villemin^c and Driss Chergaoui^a 

^aDepartment of Chemistry, Faculty of Science Semlalia, Laboratory of Molecular Chemistry, Marrakech, Morocco; ^bSchool of Health Sciences, University of KwaZulu-Natal, Westville, Durban, South Africa; ^cEcole Nationale Supérieure d'Ingénieurs (E.N.S.I.) I. S. M. R. A., LCMT, UMR CNRS n° 6507, Caen, France

Communicated by Ramaswamy H. Sarma

ABSTRACT

The hepatitis C virus is a communicable disease that gradually harms the liver leading to cirrhosis and hepatocellular carcinoma. Important therapeutic interventions have been reached since the discovery of the disease. However, its resurgence urges the need for new approaches against this malady. The NS4B receptor is one of the important proteins for Hepatitis C Virus RNA replication that acts by mediating different viral properties. In this work, we opt to explore the relationships between the molecular structures of biologically tested NS4B inhibitors and their corresponding inhibitory activities to assist the design of novel and potent NS4B inhibitors. For that, a set of 115 indol-2-ylpyridine-3-sulfonamides (IPSA) compounds with inhibitory activity against NS4B is used. A hybrid genetic algorithm combined with multiple linear regressions (GA-MLR) was implemented to construct a predictive model. This model was further used and applied to a set of compounds that were generated based on a pharmacophore modeling study combined with virtual screening to identify structurally similar lead compounds. Multiple filtrations were implemented for selecting potent hits. The selected hits exhibited advantageous molecular features, allowing for favorable inhibitory activity against HCV. The results showed that 7 out of 1285 screened compounds, were selected as potent candidate hits where Zinc14822482 exhibits the best predicted potency and pharmacophore features. The predictive pharmacokinetic analysis further justified the compounds as potential hit molecules, prompting their recommendation for a confirmatory biological evaluation. We believe that our strategy could help in the design and screening of potential inhibitors in drug discovery.

ARTICLE HISTORY

Received 18 June 2020
Accepted 2 October 2020

KEYWORDS

QSAR; pharmacophore modeling; virtual screening; GA-MLR; HCV; NS4B

1. Introduction

Over the past century, hepatitis C virus (HCV) is considered as a world-wide serious health-care issue with at least 400,000 deaths annually and almost 185 million afflicted people (~3% of the world's population) all over the world (Bartenschlager et al., 2018). HCV is an enveloped single-stranded positive sense RNA virus of 9.4 kb belonging to the Hepacivirus genus of the Flaviviridae family (Cox, 2015). Approximately 80% of HCV infections develop into chronic hepatitis that can ultimately lead to liver fibrosis, cirrhosis, and hepatocellular carcinoma (Baumert & Hoshida, 2019). Following the discovery of the virus, significant attempts have been made in order to reduce the burden of the global HCV epidemic. However, the development of antivirals against HCV has still been dominated by inhibitors of the HCV targets NS3 genotype 4A protease, RNA-dependent RNA polymerase, and NS5 leading to major adverse effects (De Clercq, 2014; El-Hasab et al., 2018; Lin et al., 2006; Pelosi et al., 2012; Wang et al., 2020). For instance, in recent years, two NS3-4A protease inhibitors (telaprevir and boceprevir) have been approved for the treatment of hepatitis C in combination with standard of care (pegylated interferon and

Ribavirin). Direct-acting antivirals (DAAs) therapy has significantly enhanced the sustained virologic response (De Clercq, 2014). Nevertheless, the use of these therapies has been associated with important side events, especially those related to tolerability and safety profile (Berman & Kwo, 2009; Zeuzem et al., 2011). Therefore, the quest for novel, safer, and highly-effective drugs along with new antiviral targets must be continually conducted (Kanwal et al., 2017). One such underdeveloped target is non-structural membrane bound NS4B protein.

NS4B is a 27 KDa protein with 261 amino acid residues that plays an essential role in HCV replication (Dvory-Sobol et al., 2010; Gouttenoire et al., 2010; Lundin et al., 2003; Yu et al., 2006; Zajac et al., 2019). It has two alpha helices each in its N and C terminal domains and 4 transmembrane domains in its central region. N-terminal domain resides in the ER-lumen while the C-terminal domain resides in the cytoplasm. The NS4B has recently emerged as a potential drug target for the treatment of chronic HCV infection. Initially, small molecules inhibiting the anti-apoptotic effect of NS4B have been identified by ViroPharma Inc. using the NS4B-expressing cell line (Chunduru et al., n.d.). A growing

number of publications have been disclosed focusing on novel chemotypes targeting HCV NS4B. (Manfroni & Cannalire, 2019; Tao et al., 2019) The last decade has highlighted that NS4B represents an appealing target in HCV drug discovery (Manfroni & Cannalire, 2019). Many hit compounds and effective inhibitors, belonging to different chemical families, have been reported (Bryson et al., 2010; Cho et al., 2010; Dufner-Beattie et al., 2014; Rai & Deval, 2011). Among them, to our knowledge, there are few modeling investigations of indole-based derivatives as HCV-NS4B enzyme inhibitors. Therefore, we have performed QSAR studies associated with principal pharmacophore features investigations on a set of 115 indol-2-ylpyridine-3-sulfonamides (IPSA) analogs in order to design new safer and potent inhibitors of this target protein.

Over last decades, computer-aided drug discovery/design (CADD) techniques have played a crucial role in the development of important therapeutic small molecules due to their marked pros of time-consuming, cost-reducing, and high efficiency in the in silico screening and prediction of drug candidates (Podlogar et al., 2001) QSAR models have been widely used in some aspects such as biological chemistry and related sciences (Ancuceanu et al., 2019; Hdoufane et al., 2019). Though, in the development of novel and effective indole-based inhibitors of the HCV NS4B protein, innovative protocols need to be considered and explored. In this study, QSAR models were established based on the structural features of compounds with potent inhibitory activity against the NS4B protein. A systemic external validation was employed to assess the predictive ability of the developed models. Thereafter, a pharmacophore model was also created in order to screen for new indole-based inhibitors with enhanced pharmacological profiles. Hence, this work could serve as a valuable platform in the development of novel and potent drugs for HCV inhibition.

2. Materials and methods

2.1. Data collection and preparation

We have collected a series of 6-(indol-2-yl)pyridine-3-sulfonamides (IPSA) derivatives. This family of molecules was synthesized and assessed for its NS4B inhibitory activity (Chen et al., 2015; Zhang et al., 2014; Zhang et al., 2013). All ISPA analogs of this research work along with their NS4B inhibitory activities in terms of pEC_{50} (i.e. $-\log_{10}(EC_{50})$) are summarized in Table 1. The chemical structures were drawn using the GaussView program and their geometries were optimized using the Gaussian09 program package (Frisch et al., 2009). The geometry of the compounds was optimized using the DFT/B3LYP/6-31G method. All structures were optimized in the gas phase. The energy minima for the optimized samples were verified by frequency analysis. Compounds were randomly split into a training set of 81 compounds (70%) and a test set of 34 compounds (30%). The training set was used to choose the best modeling approach and obtain the final model. The test set served only to validate and test the performance of the constructed model. The test set compounds are indicated in Table 1 by a superscript ^{tr}.

2.2. Molecular descriptors calculation

Molecular descriptors are at the core of QSAR modeling. They are commonly used to encode certain information related to chemicals in order to create cheminformatics models. In this current study, the descriptors have been computed using Dragon 7.1 software (Talet srl, n.d.), including a: constitutional descriptors; b: 1D-functional group counts, atom-centered fragments; c: 2D-topological descriptors, walk and path counts, connectivity indices, information indices, 2D-autocorrelations, edge adjacency indices, Burden eigenvalues, topological charge indices, eigenvalue-based indices, 2D-binary fingerprints, 2D-frequency fingerprints; d: 3D-Randic molecular profiles, geometrical descriptors, WHIM descriptors, GETAWAY descriptors; e: charge descriptors, and f: molecular properties.

To optimize the information content of the descriptors pool, the descriptors were initially screened by eliminating the constant and almost constant descriptors. Then, the descriptor pool was reduced by examining the pairwise correlations. If the pairwise correlation of two descriptors is very high, the one showing the highest pair correlation with all other descriptors is automatically excluded (K correlation coefficient greater than 0.95 was used in our study). Finally, among the 5225 descriptors that were computed, Table 2 lists only 10 descriptors that were found to be important and govern the potency of the studied compounds.

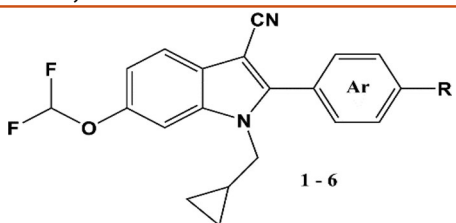
2.3. Models construction

In order to establish a possible relationship between the studied biological activity and the molecular descriptors, the selected parameters were utilized to build QSAR models using the Genetic Algorithm (GA) procedure combined with a multiple linear regression (MLR) method. This technique is implemented in the QSARINS software (Gramatica et al., 2013; 2014). The parameters used to build the models are as follows: all subset until 2, maximum generations: 10000 and mutation probability of 0.05. All other parameters were set as default values. The schematic flowchart of the methodology used in this work is depicted in Figure 1.

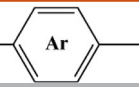
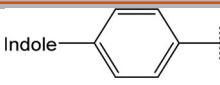
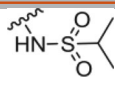
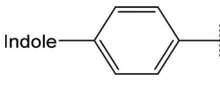
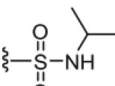
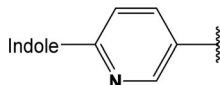
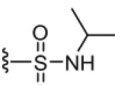
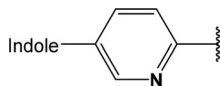
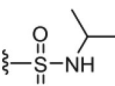
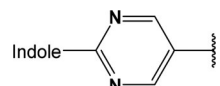
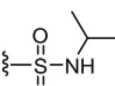
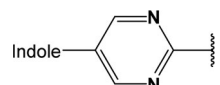
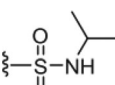
2.4. Model validation and applicability domain evaluation

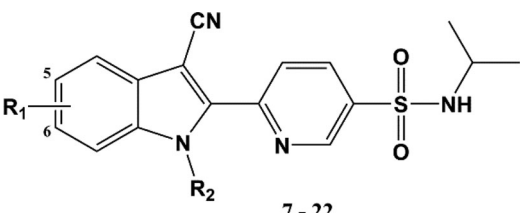
Model validation, including internal and external verifications, plays a crucial role in detecting the stability and reliability of the constructed model. In the internal validation, leave-one-out (LOO) cross-validation is performed to verify the reliability of the model. The correlation coefficient (R_{tr}^2), cross-validated Q_{LOO}^2 and the root mean squared error (RMSE) metrics were computed. Additionally, Y-randomization, which is a commonly used method to ensure model robustness and exclude randomness of the constructed model, was applied by calculating Y-scrambling (R^2) parameter as well. The Y-scramble procedure is repeated for 2000 times. In the external validation, the predictive ability was verified by computing the R_{test}^2 along with the concordance correlation

Table 1. Molecular structure and the studied biological activity data.



1 - 6

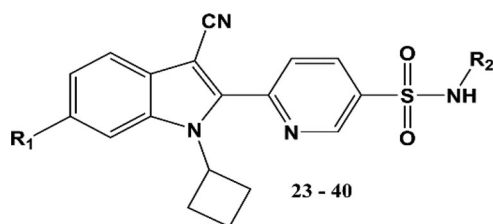
Compound	Indole- 	R	pEC ₅₀	Publication
1	Indole- 		6.41	(Zhang et al., 2013)
2	Indole- 		7.00	"
3 ^t	Indole- 		7.48	"
4	Indole- 		6.25	"
5	Indole- 		6.62	"
6 ^t	Indole- 		5.00	"



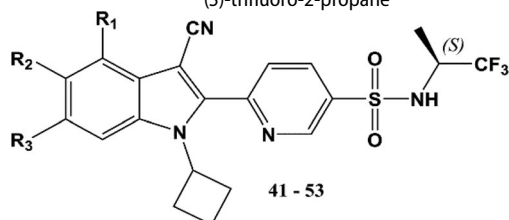
7 - 22

Compound	R ₁	R ₂	pEC ₅₀	Publication
7	6-OCHF ₂	<i>n</i> -Pr	7.14	(Zhang et al., 2013)
8 ^t	6-OCHF ₂	<i>i</i> -Pr	7.21	"
9	6-OCHF ₂	<i>c</i> -Bu	7.92	"
10 ^t	6-OCHF ₂	<i>c</i> -Pentyl	7.82	"
11 ^t	6-OCHF ₂	<i>c</i> -Hexyl	7.66	"
12	6-OCF ₃	<i>c</i> -Bu	7.17	"
13	6- <i>c</i> -Pr	<i>c</i> -Bu	7.68	"
14 ^t	6- <i>c</i> -Bu	<i>c</i> -Bu	6.26	"
15 ^t	6-OCF ₂	<i>c</i> -Bu	7.89	"
16	6-OCH ₂ CHF ₂	<i>c</i> -Bu	5.72	"
17	6- <i>i</i> -PrS	<i>c</i> -Bu	5.88	"
18	5-OCHF ₂	<i>c</i> -Bu	6.27	"
19	5-OCF ₃	<i>c</i> -Bu	5.96	"
20 ^t	5- <i>c</i> -Pr	<i>c</i> -Bu	6.24	"
21	5-F	<i>c</i> -Bu	7.59	"
22 ^t	5-Cl	<i>c</i> -Bu	7.55	"

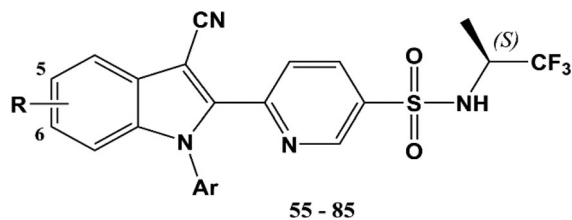
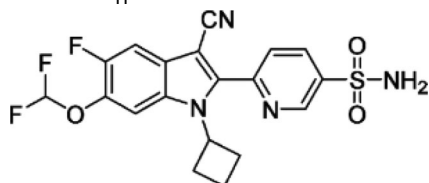
(continued)



Compound	R ₁	R ₂	pEC ₅₀	Publication
23	-OCHF ₂	2-methyl-2-propane	7.57	(Zhang et al., 2013)
24	-OCHF ₂	(S)-ethyl-cyclopropane	7.64	"
25	-OCHF ₂	1,3-difluoro-2-propane	8.40	"
26	-OCHF ₂	2-propane-1,3-diol	6.03	"
27	-OCHF ₂	1-fluoro-2-methyl-2-propane	6.28	"
28	-OCHF ₂	1,1-difluoro-3-cyclobutane	5.57	"
29	-OCHF ₂	(R)-trifluoro-2-propane	7.17	"
30	-OCHF ₂	(S)-trifluoro-2-propane	8.70	"
31	-OCHF ₂	trifluoromethyl-1-cyclopropane	6.74	"
32	-OCHF ₂	(trifluoromethyl)-1-cyclobutane	7.52	"
33	-OCHF ₂	1,1,1-trifluoro-2-methyl-2-propane	7.26	"
34	-OCHF ₂	1,1,1-trifluoro-2-ethane	6.38	"
35	-OCHF ₂	2-propanenitrile	6.66	"
36	-OCHF ₂	H	5.89	"
37	c-Pr	1,3-difluoro-2-propane	8.40	"
38	c-Pr	trifluoromethyl-1-cyclopropane	7.13	"
39	c-Pr	(R)-trifluoro-2-propane	6.90	"
40	c-Pr	(S)-trifluoro-2-propane	7.92	"



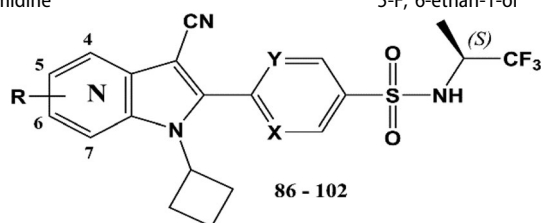
Compound	R ₁	R ₂	R ₃	pEC ₅₀	Publication
41	-H	-F	-CHF ₂ O	8.15	(Zhang et al., 2014)
42 ^t	-H	-OH	-CHF ₂ O	7.20	"
43	-H	-Me	-CHF ₂ O	6.97	"
44 ^t	-H	-Et	-CHF ₂ O	5.85	"
45	-H	-F	-Me	8.15	"
46 ^t	-H	-F	-Et	8.40	"
47	-H	-F	c-Pr	7.55	"
48	-H	-F	-OH	6.39	"
49 ^t	-H	-F	-Cl	6.34	"
50	-H	-F	-OMe	7.13	"
51	-H	-F	-SMe	7.46	"
52	-F	-H	-CHF ₂ O	7.74	"
53 ^t	-F	-H	-Me	6.82	"
54				6.48	"



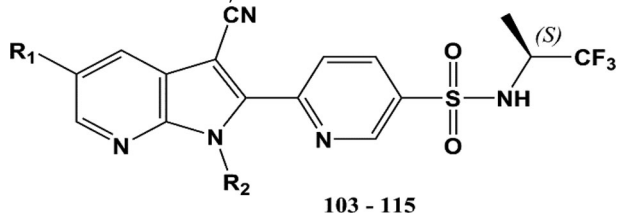
Compound	Ar	R	pEC ₅₀	Publication
55 ^t	2-pyrimidine	5-H, 6-OCHF ₂	8.15	(Zhang et al., 2014)
56	2-pyrimidine	5-H, 6-Me	8.52	"
57 ^t	2-pyrimidine	5-F, 6-OCHF ₂	7.80	"
58	2-pyrimidine	5-F, 6-Me	8.22	"
59 ^t	2-pyridine	5-F, 6-Me	7.74	"

(continued)

60	3-pyridine	5-F, 6-Me	7.96	"
61	4-pyridine	5-F, 6-Me	7.96	"
62	2-pyrazine	5-F, 6-Me	8.05	"
63	3-pyridazine	5-F, 6-Me	7.89	"
64 ^t	2-thiazole	5-F, 6-Me	8.52	"
65	4-thiazole	5-F, 6-Me	7.70	"
66	5-thiazole	5-F, 6-Me	8.15	"
67	2-thiophene	5-F, 6-Me	8.70	"
68 ^t	5-methyl-2-pyrimidine	5-F, 6-Me	7.49	"
69	5-fluoro-2-pyrimidine	5-F, 6-Me	8.40	"
70	3-fluoro-6-pyridine	5-F, 6-Me	8.10	"
71 ^t	3-methyl-6-pyridine	5-F, 6-Me	7.60	"
72	3-fluoro-2-pyridine	5-F, 6-Me	8.52	"
73	3-methyl-2-pyridine	5-F, 6-Me	6.40	"
74	2-pyrimidine	5-F, 6-Et	8.70	"
75 ^t	2-pyridine	5-F, 6-Et	8.30	"
76	4-pyridine	5-F, 6-Et	8.05	"
77	2-pyrazine	5-F, 6-Et	8.40	"
78	2-thiazole	5-F, 6-Et	8.70	"
79 ^t	5-fluoro-2-pyrimidine	5-F, 6-Et	8.52	"
80	5-methyl-2-pyrimidine	5-F, 6-Et	7.64	"
81 ^t	3-fluoro-6-pyridine	5-F, 6-Et	7.89	"
82	2-pyrimidine	c-Pr	8.22	"
83	2-pyridine	c-Pr	7.80	"
84 ^t	3-fluoro-6-pyridine	c-Pr	7.92	"
85	2-pyrimidine	5-F, 6-ethan-1-ol	7.17	"



Compound	R	X, Y	Azaindole	pEC ₅₀	Publication
86	5-OMe	N, CH	4-	6.64	(Chen et al., 2015)
87 ^t	6-Me	N, CH	4-	6.92	"
88	6-Cl	N, CH	4-	7.03	"
89	6-Me	N, CH	5-	6.64	"
90	6-Cl	N, CH	5-	6.59	"
91	5-Cl	N, CH	6-	6.28	"
92 ^t	5-Me	N, CH	6-	5.85	"
93	6-OMe	N, CH	5,7-di	6.62	"
94	6-OCHF ₂	N, CH	5,7-di	6.01	"
95 ^t	5-OMe	N, CH	7-	7.57	"
96	6-OMe	N, CH	7-	6.02	"
97 ^t	5-Cl	N, CH	7-	7.77	"
98 ^t	6-Cl	N, CH	7-	6.15	"
99	5-Me	N, CH	7-	7.89	"
100	6-Me	N, CH	7-	7.06	"
101	5-Me	CH, CH	7-	7.06	"
102	5-Me	N, N	7-	6.80	"



Compound	R ₁	R ₂	pEC ₅₀	Publication
103 ^t	-OMe	c-CH2Pr	6.52	(Chen et al., 2015)
104 ^t	-OMe	c-Pent	6.82	"
105	-OEt	c-Bu	6.70	"
106	<i>n</i> -OPr	c-Bu	5.67	"
107	<i>i</i> -OPr	c-Bu	5.12	"
108	-OCHF ₂	c-Bu	7.30	"
109	-Cl	c-CH2Pr	7.41	"
110	-Cl	c-Pent	7.57	"
111	-Et	c-Bu	7.82	"
112 ^t	c-Pr	c-Bu	7.40	"
113 ^t	-F	c-Bu	7.34	"
114 ^t	-CF ₃	c-Bu	8.70	"
115	-H	c-Bu	6.66	"

^t test molecules.

Table 2. List of the selected molecular descriptors and their physical-chemical meaning.

Abbreviation	Descriptor
MATS5p	Moran autocorrelation of lag 5 weighted by polarizability
VE1_B(m)	Coefficient sum of the last eigenvector (absolute values) from Burden matrix weighted by mass
CATS3D_07_AL	CATS3D Acceptor-Lipophilic BIN 07 (7.000 – 8.000 Å)
P_VSA_p_2	P_VSA-like on polarizability, bin 2
R6v+	R maximal autocorrelation of lag 6 / weighted by van der Waals volume ^a
Mor23v	Signal 23 / weighted by van der Waals volume ^b
SpPosA_X	Normalized spectral positive sum from chi matrix
RDF090s	Radial Distribution Function – 090 / weighted by I-state
H6s	H autocorrelation of lag 6 / weighted by I-state ^a
TIC1	Total Information Content index (neighborhood symmetry of 1-order)

^aGATEWAY class: Geometry, Topology and Atom-Weights Assembly.

^bMoRSE class: Molecule Representation of Structure based on Electron diffraction.

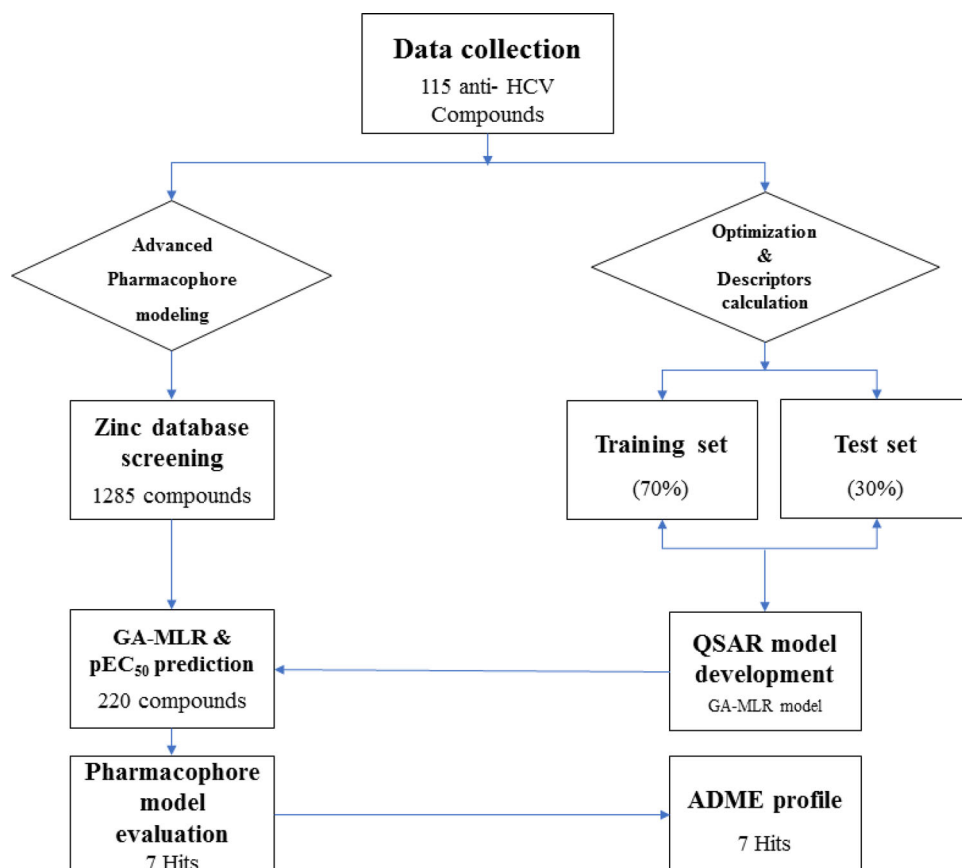


Figure 1. Schematic workflow illustrating the complete hierarchy involved in the development of QSAR models of IPSA derivatives.

coefficient CCC (Tropsha et al., 2003), Q_{F1}^2 (Chirico & Gramatica, 2012), Q_{F2}^2 (Shi et al., 2001), and Q_{F3}^2 (Schüürmann et al., 2008).

Each model has its own scope of application. Only samples that fall within the scope of the model are likely to be accurately predicted. In this study, leverage method (Eriksson et al., 2003) is adopted to analyze the model's applicability domain (AD), defined as follows:

$$h_i = x_i^T (X^T X)^{-1} x_i, \quad i = 1, \dots, n$$

where h_i and x_i are the leverage value and the descriptor vector of the considered compound, respectively. X is the descriptor matrix derived from the training set descriptor values. T stands for 'transposed'. The delimited domain of the model is defined by the leverage cut-off (h^*), set as $3(p+1)/n$, (p is the number of model descriptors and n is the

number of training set samples). A leverage greater than h^* indicates that the prediction of the model is unreliable.

All these research works regarding the GA-MLR model were executed using QSARINS software, which is free for academia and research groups. This program implements various tools for rigorous external validation of the established models, based on different validation criteria, as well as for the check of the model AD.

2.5. Pharmacophore modeling and virtual screening

Ligand-based pharmacophore modeling is one of the most significant procedures to classify and identify the key chemical features from a group of ligand compounds such as active molecules (Cruz et al., 2019; da Costa et al., 2019; Pinto et al., 2019; Ramos et al., 2020). However, virtual

screening is the technique used to filter the available chemo-libraries and databases of active molecules by mapping them on the generated pharmacophore model. Combining these techniques might contribute to the design of novel lead molecules that can enhance the expression of the target molecules. In the current work, chemical features from the general scaffold of the active compounds were used to create a preliminary pharmacophore model using the ZincPharmer online tool (Figure 2) (Ascher et al., 2014). This model was used to screen for new compounds from the Zinc database (a free database of commercially-available compounds for virtual screening). The retrieved lead compounds were subject to two filtration procedures. In the first step, we employed the established GA-MLR model as described above in order to select the most active from the pool of the retrieved ones. The second step aims to apply the pharmacophore fit model to evaluate the features fitting of each hit compound by means of the pharmacophore model features to rank the hit molecules screened using the GA-MLR QSAR model. The pharmacophore fit model was generated based on the initial 115 chemical structures using LigandScout software ("LigandScout, Version 4.3; Inte:Ligand GmbH, Clemesn-Maria-Hofbauer-G. 6, 2344, Maria Enzersdorf, Austria. HYPERLINK, <http://www.inteligand.com>," n.d.).

Among the chemical features described above, it was observed that four types including hydrogen-bond acceptor (HBA), hydrogen-bond donor (HBD), hydrophobic (H), and aromatic ring (AR) could effectively map the chemical features of all used compounds in the training set (Figure 3). These chemical features were generated in order to stay close to the common scaffold of the studied IPSA analogs and then were used to set up a series of lead molecules. For this reason, the Zinc database was screened out and as a result 1285 molecules were received as leads. The retrieved 1285 molecules were subjected to the constructed GA-MLR model. Subsequently, the top 220 compounds were selected for the eventual pharmacophore fit model. Finally, 7 hits with the higher pharmacophore fit score were subjected to the ADME analysis.

2.6. Pharmacological and metabolic properties

Complementary and interesting results, concerning the potential use of the selected hit molecules as a drug, are the prediction of its absorption, distribution, metabolic and excretion (ADME) properties in humans (C. B. R. Santos et al., 2020; K. L. B. dos Santos et al., 2020). These properties of the potent hit compounds were calculated using the SWISSADME online tool, which predicts about 50 ADMET endpoints (Daina et al., 2017). This analysis provides a criterion to eliminate potentially harmful molecules during the selection phases, for the synthesis of compounds and future experimental tests.

3. Results and discussion

3.1. GA-MLR modeling

QSAR strategies have been established based on a variety of mathematical methods to predict activities and properties of

untested molecules. The major differences between them are related to the molecular descriptors and the applied mathematical methodologies. In the present work, the hybrid GA-MLR technique was employed and several mathematical parameters were considered for QSAR validation. The ten molecular descriptors in the selection process were considered to develop QSAR models. Several models with low multicollinearity between descriptors and good correlation with the modeled response were developed using the QSARINS software. The best constructed GA-MLR model using the selected descriptors along with its statistical metrics is as follows:

$$\begin{aligned} \text{pEC}_{50} = & 16.72 + 10.02 * (\text{MATS5p}) - 1.85 * (\text{VE1B(m)}) - 0.16 * \\ & (\text{CATS3D}_07_{\text{AL}}) - 0.02 * (\text{P}_{\text{VSA}P_2}) - 203.06 * (\text{R6v+}) + 3.33 * \\ & (\text{Mor23v}) + 32.28 * (\text{SpPosA}_X) + 0.001 * (\text{RDF090s}) + 0.26 * \\ & (\text{H6s}) - 0.03 * (\text{TIC1}) \end{aligned} \quad (1)$$

$$\begin{aligned} N_{\text{tr}} = 81, R_{\text{tr}}^2 = 0.74, R_{\text{adj}}^2 = 0.71, Q_{\text{LOO}}^2 = 0.67, R_{\text{test}}^2 = 0.64, Q_{F1}^2 \\ = 0.64, Q_{F2}^2 = 0.64, Q_{F3}^2 = 0.65, \text{CCC} = 0.86, \text{RMSE}_{\text{tr}} \\ = 0.44, S = 0.48, F = 20.12. \end{aligned}$$

In this equation, N_{tr} is the number of training compounds, R_{tr}^2 is the correlation coefficient of the training set, R_{adj}^2 is the adjusted R_{tr}^2 , Q_{LOO}^2 is the square of the cross-validated correlation coefficient obtained from leave-one-out (LOO) procedure, R_{test}^2 is the square of correlation coefficient obtained from the test set, S is the standard error of estimate, and F is Fischer-ratio between the variances of calculated and observed activities.

From the above results, the model presented an R_{tr}^2 of 0.74, so it manifests a proper fit for modeling the anti-HCV inhibition. As can be seen, the fitness and robustness of the model are both acceptable according to the standard validation metrics (Golbraikh & Tropsha, 2002; Gramatica, 2007). These metrics (i.e. $Q_{\text{LOO}}^2 > 0.6$ and $R_{\text{test}}^2 > 0.5$) are based on the regression through origin. We have adopted these methods to validate the established model using the external test set of molecules that were not used for the model construction (Table 1). Moreover, Chirico and Gramatica had recommend the thresholds for other external validation parameters, including Q_{F1}^2 , Q_{F2}^2 and $Q_{F3}^2 > 0.6$ and $\text{CCC} > 0.85$ (Chirico & Gramatica, 2011). All these parameters of the obtained GA-MLR model are large enough to ensure its performance. Furthermore, the scatter plot of the experimental versus the predicted pEC_{50} by the established GA-MLR model is illustrated in Figure 4. From this figure we shall see that all the samples are slightly scattered near the regression line, which showed that the model denotes a significant correlation between the studied activity and the 10 selected descriptors.

3.2. Model validation and applicability domain assessment

After the Y-scrambling procedure, Figure 5 shows the distribution of R_{Yscr}^2 and Q_{Yscr}^2 versus the correlation between descriptors and pEC_{50} activity (Kxy) by applying 2000

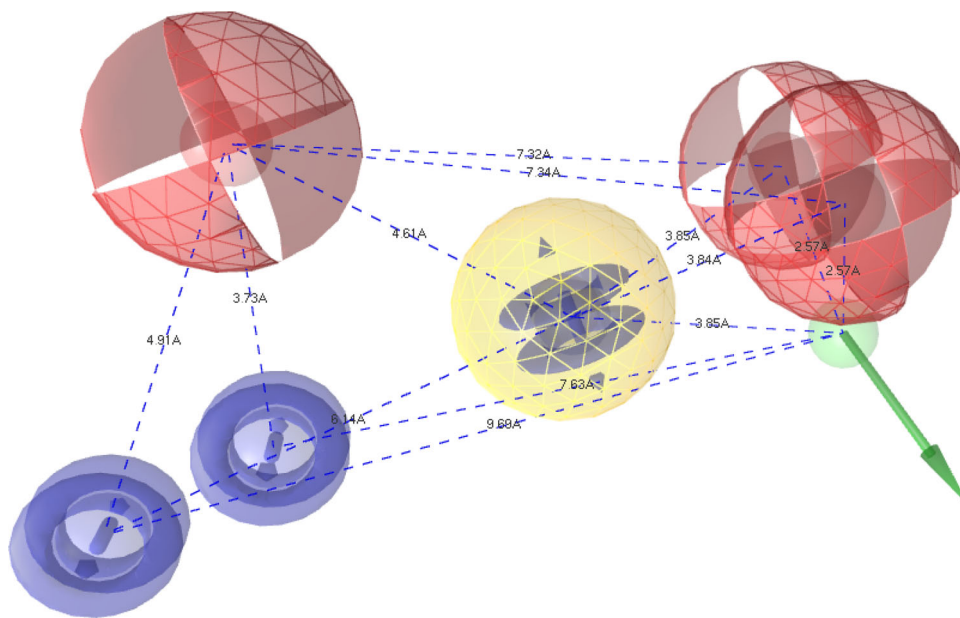


Figure 2. Pharmacophore model with distance constraints.

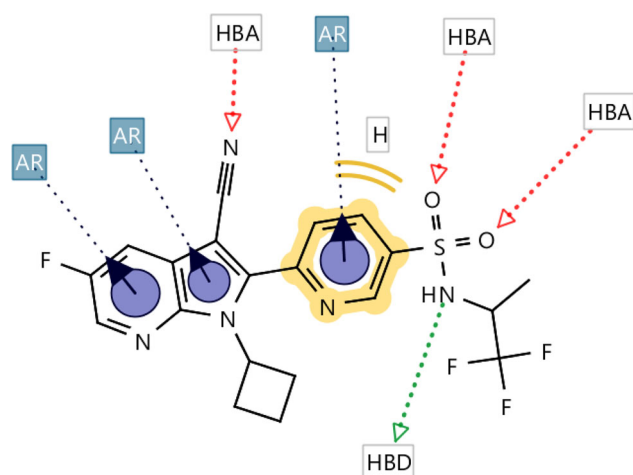


Figure 3. Structure-based pharmacophore model generated by LigandScout software. Three hydrogen-bonds acceptors (HBA), one hydrogen-bond donor (HBD), one Hydrophobic (H), and three aromatic rings (AR).

iterations. The correlation coefficients of the non-scrambled model ($R_{tr}^2 = 0.74$, $R_{adj}^2 = 0.71$, $Q_{LOO}^2 = 0.67$) tend to be seen higher than those from Y-scrambling, indicating that the relationship between the IPSA derivatives and their corresponding pEC_{50} do exist and the non-scrambled GA-MLR model was not accidentally acquired. The proposed GA-MLR model has a good reliability, robustness and stability.

The leverage method shown in Figure 6, illustrated as Williams plot, is employed to evaluate the practical applicability domain (AD) of the model. The dashed lines indicate the cutoff value of ± 3 standard deviation (s.d.). The warning line for X outlier (h^*) is 0.407. As can be seen from the Williams plot, all chemicals are within the scope of the AD except for compounds No. 26 and No. 65. In addition, all compounds are within standard residuals of ± 3 (s.d.). Even for compounds No. 26 and 65, whose h value are beyond h^* , the predicted pEC_{50} are slightly close to their experimental ones.

3.3. Mechanistic interpretation of the selected descriptors

Mechanistic interpretation of the QSAR model was performed in order to infer whether the important descriptors were selected reasonably. These descriptors belong to several categories (Table 2). Some of these categories are difficult to interpret, and therefore are delicate to use for the study of the synthesis of new derivatives. The ten molecular descriptors appearing in the GA-MLR model (Eq. 1) are divided into two classes. The first one comprises descriptors with favorable correlation to the inhibitory activity, whereas the other class is negatively correlated to it.

In the first class, the most important descriptor is SpPosA_X due to its highly positive coefficient. The lower the value of this descriptor (i.e. the greater the number of atoms in the molecule), the tendency of increased anti-HCV inhibitory activity tends to be higher. This may show that active inhibitors have a larger molecular size. This finding was addressed in a previous similar QSAR study (de Campos & de Melo, 2017).

The second important descriptor is the MAT5p that belongs to the 2D autocorrelations descriptor and describes the electronic properties of the molecules, particularly in terms of the atomic polarizability. It shows that the high polarizability stems from the larger size of the molecules (in this case, the higher the better) is significant to the activity.

The third descriptor is Mor23v, it belongs to the 3D-Morse family and weighted by atomic van der Waals volume (Devinyak et al., 2014). The higher Mor23v values lead to higher activity levels. Its positive coefficient might indicate the beneficial molecular volume in improving the pEC_{50} activity.

The last two descriptors H6s and RDF090s belong to the GETAWAY and Radial Distribution Function (RDF) classes, respectively. These two properties are both weighted by the intrinsic state and considered as relevant descriptors but

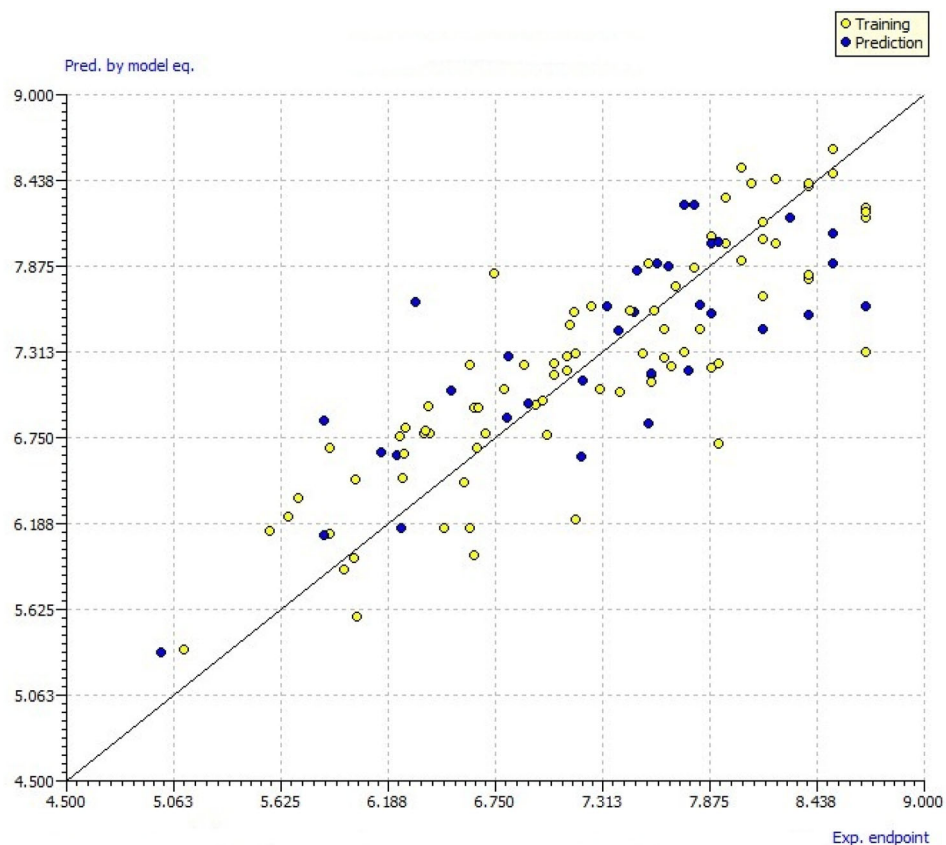


Figure 4. Scatter plot of experimental vs. predicted pEC_{50} values calculated by the GA-MLR model.

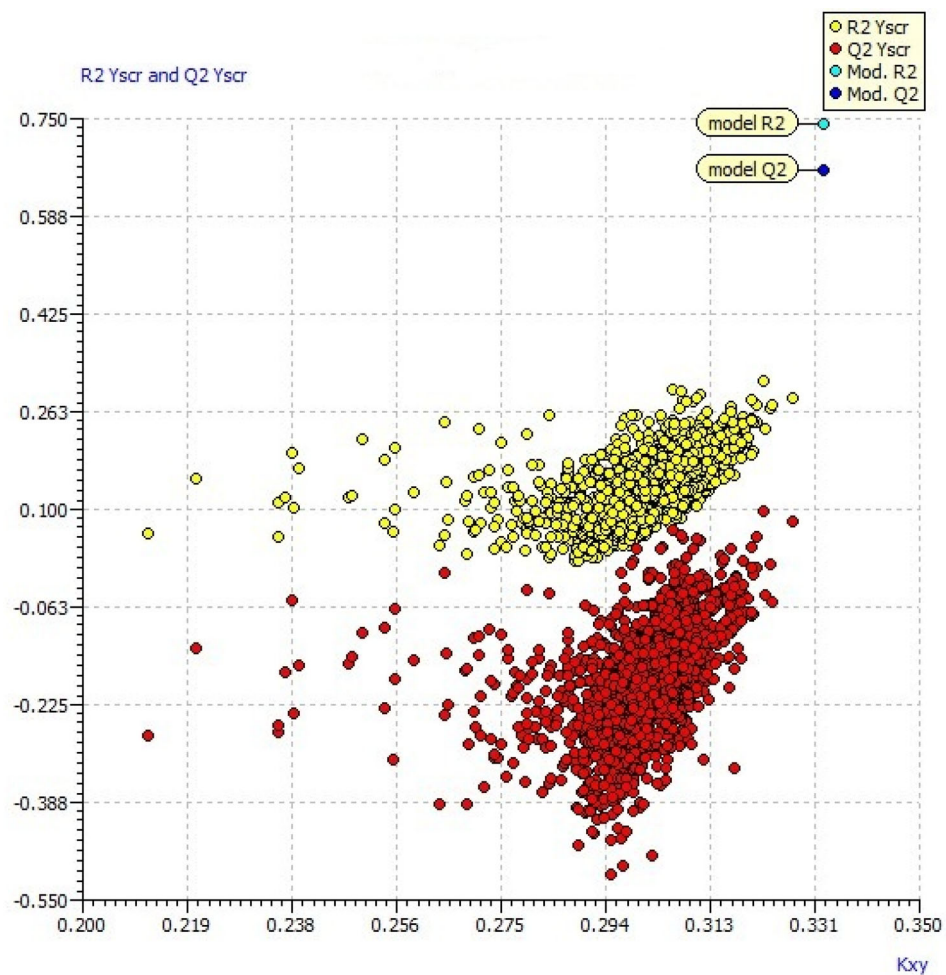


Figure 5. R^2_{Yscr} and Q^2_{Yscr} vs. K_{xy} from Y-scrambling procedure.

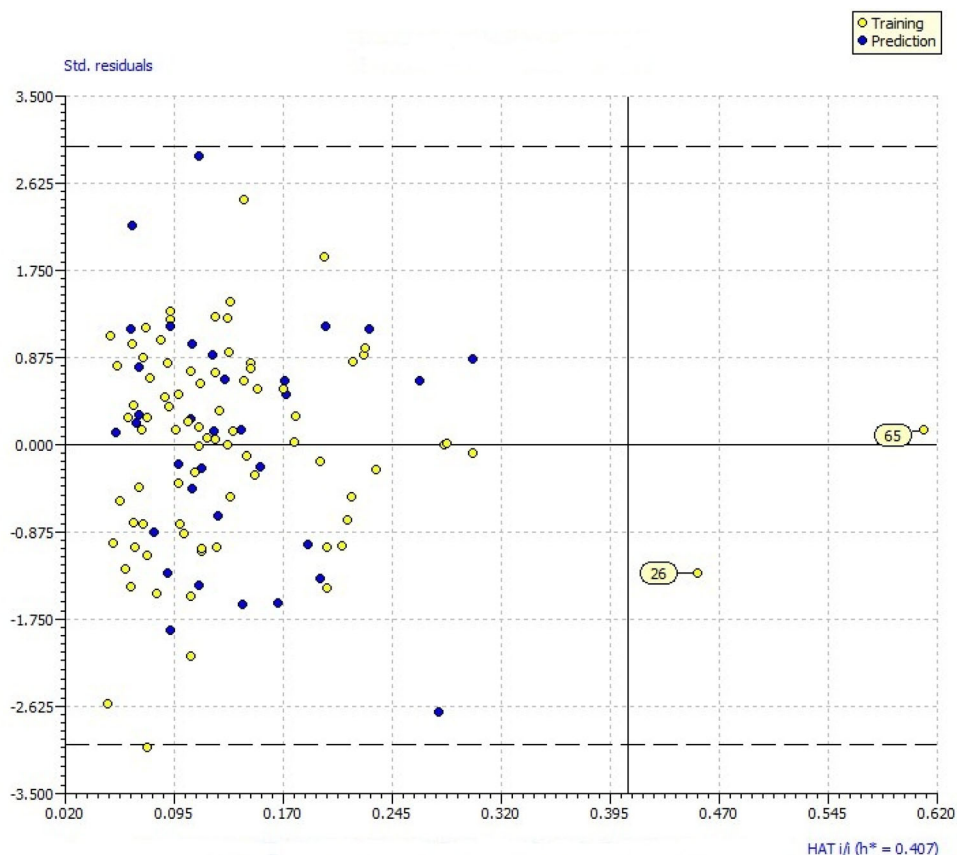


Figure 6. Williams plot. Training samples in yellow and test samples in blue. The dashes lines are the cutoff ± 3 s.d. and the warning leverage value h^* of 0.407.

difficult to interpret. Taking into consideration that they are weighted by intrinsic state, the values of these factors are higher for compounds with more electronegative atoms (H. Hall, 2012).

In the second class, the descriptor $R6v+$ that belongs to the GETAWAY class (Todeschini & Consonni, 2009) is found to be very important though its coefficient is highly negative. Indeed, it should be noted that the proposed hits should have lower $R6v+$ values. This parameter is directly related to the Van der Waals volume. Unlike the most important descriptor with a positive coefficient, it is important to note that $R6v+$ is not necessarily related to the total size of the compounds.

The second descriptor $VE1_B(m)$ is a 2D matrix-based descriptor and depends on a complex way on the molecular size, shape, presence of heavy heteroatoms and of multiple bonds (Consonni & Todeschini, 2012). $VE1_B(m)$ tends to increase when increasing the molecular size (number of non-H atoms), the branching (number of terminal atoms), and the number of cycles and of multiple bonds.

The third negatively correlated descriptor $CATS3D_07_AL$ encodes the conformation of a molecule in the form of a histogram that contains the normalized frequencies of all pairs of atoms in a molecule (Fechner et al., 2003). Furthermore, the TIC1 descriptor that belongs to the class of the total information index, represents the residual information contained in the molecular structures (Todeschini & Consonni, 2000).

Finally, $P_VSA_p_2$ is the least contributing one and is weighted by polarizability. Nevertheless, it is useful for

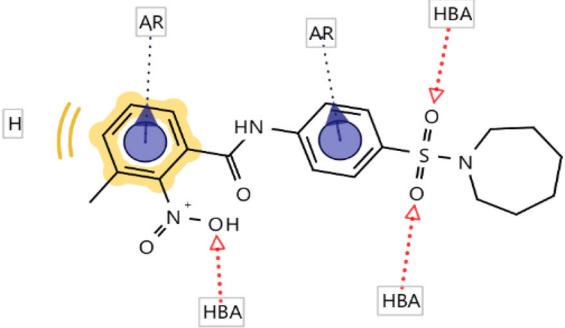
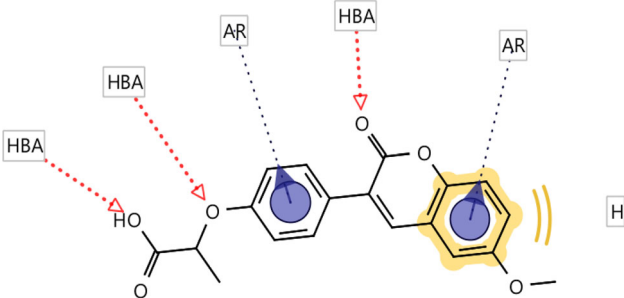
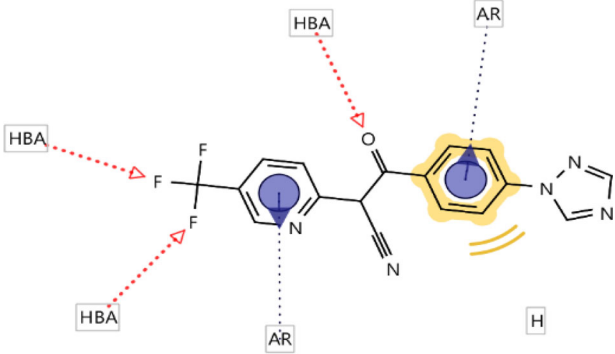
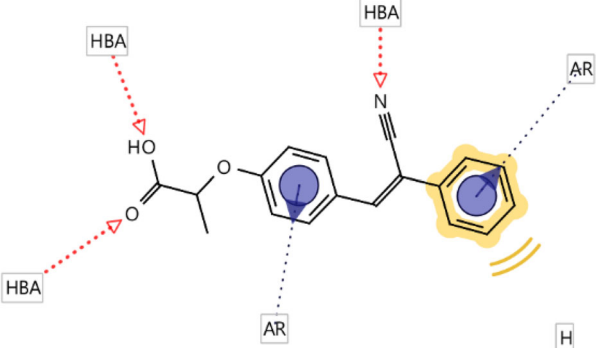
selecting new molecules in virtual screening since the weighting by polarizability is consistent with the pEC_{50} inhibitory activity.

3.4. Pharmacophore model generation and hits analysis

A virtual screening methodology, using a basic pharmacophore-based model, was firstly implemented to search in the Zinc database. A total number of 1285 of molecules from the Zinc database were retrieved to match the created pharmacophore model and were exposed to further analysis. After that, we have applied the established GA-MLR model on these molecules, which allows the prediction of their activities based on Equation 1. Only compounds with pEC_{50} in the range of 8.26 to 9.29 (i.e. EC_{50} in the range of 0.52 to 5.38 nM) were selected. Thereby, 220 compounds were chosen based on their pEC_{50} values and were subjected to the pharmacophore fit score calculations. The pharmacophore fit model was built based on the training and test sets of the 115 IPSA compounds. The best model was then used to select the best hits by means of their pharmacophore fitting scores. Finally, seven hits were selected for the ADME analysis (Table 3).

Several small molecule inhibitors were reported to have a good potency regarding the treatment of the HCV. Inhibitors bearing the sulfonamide moiety attached to the phenyl ring, as a preferred pharmacophore feature, lead to promising sub-micromolar potency in the replicon cell-based assay (Chen et al., 2013). Out of the selected hit molecules,

Table 3. Seven molecules that successfully passed the assigned filters.

Hit	Zinc ID	Chemical Structure	Predicted pEC50	Pharmacophore-Fit Score
<u>1</u>	Zinc14822482		9.04	65.18
<u>2</u>	Zinc11766170		8.68	64.93
<u>3</u>	Zinc41554609		8.55	64.91
<u>4</u>	Zinc39557374		8.50	64.63

(continued)

Zinc14822482 (Hit: 1), bearing the sulfonamide moiety attached to the phenyl ring, exhibited the best pharmacophore features with the highest fit score and also the best predicted inhibitory activity. This moiety is presented in compound hits 5 and 6 as well. The sulfonamide group could be identified in a core structure as a leading group during the optimization of novel compounds and in the finding of potent drugs with excellent pharmacokinetic properties (Manfroni & Cannalire, 2019). In addition, Zinc14822482 is

having a nitro-benzamide moiety in the core structure. This group increases the efficacy of the hit molecule and showed a very potent antiviral activity as reported in previous studies (Ganta et al., 2019; Walker, 1999).

Concerning the second ranked hit (2), which is the Zinc11766170, it is an oxochromen-based compound and belongs to the family of coumarin derivatives. The use of the coumarin-based chemicals had shown increased potency and selectivity as anti-HCV compounds (Manvar et al., 2016;

Table 3. Continued.

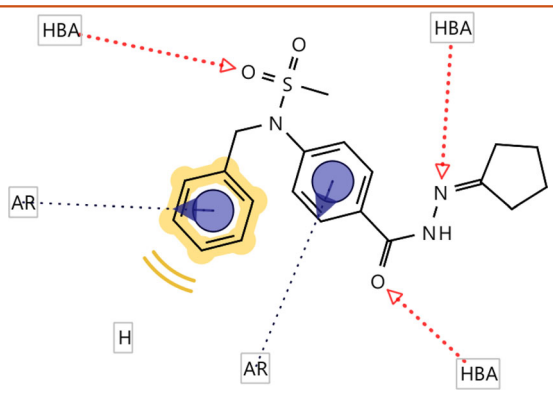
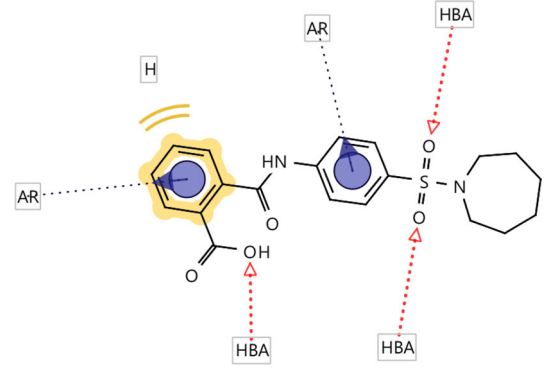
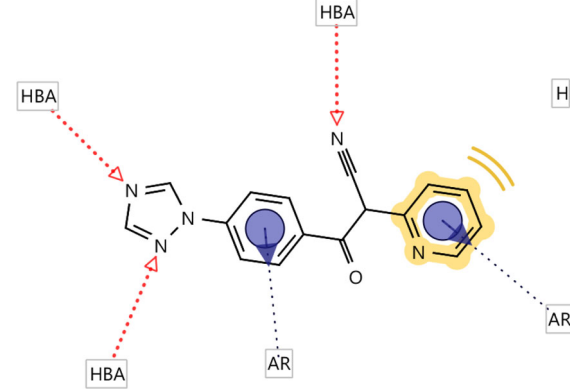
Hit	Zinc ID	Chemical Structure	Predicted pEC50	Pharmacophore-Fit Score
<u>5</u>	Zinc05205726		8.36	64.55
<u>6</u>	Zinc00670533		9.21	64.48
<u>7</u>	Zinc41493305		8.39	64.31

Table 4. ADME profiles of the seven top ranked hits.

Hit	Molecular Weight (g/mol)	Lipophilicity (logP)	HI Absorption	BBB Permeability	Druglikeness (Lipinski)	Synthetic Accessibility
<u>1</u>	417.48	2.48	High	No	Yes	3.05
<u>2</u>	339.32	2.65	High	No	Yes	3.54
<u>3</u>	357.29	1.77	High	No	Yes	2.96
<u>4</u>	292.31	2.50	High	Yes	Yes	2.80
<u>5</u>	385.48	2.93	High	No	Yes	3.11
<u>6</u>	401.46	2.19	High	No	Yes	2.67
<u>7</u>	289.29	1.99	High	No	Yes	2.70

Mishra et al., 2020). Despite the early described groups, the other hit molecules including Zinc41554609 and Zinc41493305 (**3** and **7**) are selected among the retrieved 1285 molecules due to the exhibition of the triazolyl substituent. Heterocyclic compounds bearing the triazolyl group showed a promising activity in the treatment of the HCV (Küçükgül & Çikla-Süzcü, 2015; Naud et al., 2008). All in all, the sulfonamide, the nitro-benzamide, the oxochromen,

and the triazolyl moieties were found to be the leading groups in the selected hits.

3.5. Pharmacokinetic ADME analysis

The well-known ADME analysis i.e. Absorption (A), Distribution (D), Metabolism (M) and Excretion (E) of any

organic compound is a crucial investigation in the drug development. It provides complimentary and interesting results regarding the potential use and the disposition of any pharmaceutical molecule inside an organism and consequently, affects its pharmacological activity (Tian et al., 2015). In the current study, we analyzed the ADME profile of the hits that successfully passed the allocated filters using the SWISSADME online tool (Daina et al., 2017). Hence, it has been shown from the obtained results that all the hits exhibit a high predicted pEC₅₀ inhibitory activity (as evident from GA-MLR results) and good ADME properties (Table 4).

The ADME properties shown in Table 4 are essential for any pharmacologically active molecule and play a vital role in the drug design process. It has been observed from these results that all the hits exhibit encouraging properties. Among these properties, the drug-likeness of Lipinski. All the selected hit compounds follow the Lipinski rule, according to which an orally active drug molecule should have its molecular weight < 500, logP < 5, hydrogen-bond donors < 5, and hydrogen-bond acceptors < 10. Moreover, the absorption of a molecule after oral administration is a significant characteristic and refers to Human Intestinal Absorption (HIA) and the blood brain barrier is also important and refers to Blood Brain Penetration (BBB). All the hit molecules show a positive absorption from the intestine into the bloodstream, while only compound 4 being the one exhibiting a favorable absorption permeability. Additionally, a closer look at the ADME analysis, reveals that the theoretical drug-like behavior of these hits is very promising and can be considered beneficial for further *in vitro* biological tests. Furthermore, in the tested *in silico* synthetic accessibility which provides a score from 1 (very easy to synthesize) to 10 (difficult and complex to synthesize), all the hit molecules are shown to be easier to prepare. The highest predicted score is less than 3.05 indicating that these hits are chemically synthesizable.

4. Conclusion

In this work, the main concern was to investigate the quantitative structure-activity relationships of a series of indol-2-ylpyridine-3-sulfonamides (IPSA) analogs and their corresponding inhibitory activities within the NS4B enzyme by using the hybrid GA-MLR method. Furthermore, the constructed model, which is found to be stable, robust and predictive with satisfactory performance, was used to predict the inhibitory activity of screened molecules from a chemical library. The IPSA compounds were used to generate a ligand-based pharmacophore model which was used to screen for potential, safe and potent inhibitor scaffolds of NS4B target. From the screened molecules, it has been shown that seven hits, which were checked by their pharmacokinetic ADME profiles, exhibit advantageous chemical features, allowing for favorable inhibitory activity against NS4B target. The combined strategies were utilized as a computational tool and demonstrate their abilities to contribute to the design of newly anti-HCV antivirals. The filtered hit compounds from this study can be further biologically evaluated for their efficiency and physiologic toxicity.

Acknowledgements

The authors would like to thank Professor Paola Gramatica for providing a copy of the QSARINS Software.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Ismail Hdoufane  <http://orcid.org/0000-0002-1435-5131>

Imane Bijj  <http://orcid.org/0000-0002-4393-6867>

Mehdi Oubahmane  <http://orcid.org/0000-0003-2465-0537>

Mahmoud E.S Soliman  <http://orcid.org/0000-0002-8711-7783>

Driss Cherqaoui  <http://orcid.org/0000-0003-1327-306X>

References

- Ancuceanu, R., Tamba, B., Stoicescu, C. S., & Dinu, M. (2019). Use of QSAR global models and molecular docking for developing new inhibitors of C-src tyrosine kinase. *International Journal of Molecular Sciences*, 21(1), 19. <https://doi.org/10.3390/ijms21010019>
- Ascher, D. B., Wielens, J., Nero, T. L., Doughty, L., Morton, C. J., & Parker, M. W. (2014). Potent hepatitis C inhibitors bind directly to NS5A and reduce its affinity for RNA. *Scientific reports*, 4(1), 4765. <https://doi.org/10.1038/srep04765>
- Bartenschlager, R., Baumert, T. F., Bukh, J., Houghton, M., Lemon, S. M., Lindenbach, B. D., Lohmann, V., Moradpour, D., Pietschmann, T., Rice, C. M., Thimme, R., & Wakita, T. (2018). Critical challenges and emerging opportunities in hepatitis C virus research in an era of potent antiviral therapy: Considerations for scientists and funding agencies. *Virus Research*, 248, 53–62. <https://doi.org/10.1016/j.virusres.2018.02.016>
- Baumert, T. F., & Hoshida, Y. (2019). Addressing the challenges of hepatitis C cure and persistent risk of hepatocellular carcinoma. *Viruses*, 11(5), 441. <https://doi.org/10.3390/v11050441>
- Berman, K., & Kwo, P. Y. (2009). Boceprevir, an NS3 protease inhibitor of HCV. *Clinics in Liver Disease*, 13(3), 429–439. <https://doi.org/10.1016/j.cld.2009.05.008>
- Bryson, P. D., Cho, N.-J., Einav, S., Lee, C., Tai, V., Bechtel, J., Sivaraja, M., Roberts, C., Schmitz, U., & Glenn, J. S. (2010). A small molecule inhibits HCV replication and alters NS4B's subcellular distribution. *Antiviral Research*, 87(1), 1–8. <https://doi.org/10.1016/j.antiviral.2010.03.013>
- Chen, G., Ren, H., Turpoff, A., Arefolov, A., Wilde, R., Takasugi, J., Khan, A., Almstead, N., Gu, Z., Komatsu, T., Freund, C., Breslin, J., Colacino, J., Hedrick, J., Weetall, M., & Karp, G. M. (2013). Discovery of N-(4'-(indol-2-yl)phenyl)sulfonamides as novel inhibitors of HCV replication. *Bioorganic & Medicinal Chemistry Letters*, 23(13), 3942–3946. <https://doi.org/10.1016/j.bmcl.2013.04.050>
- Chen, G., Ren, H., Zhang, N., Lennox, W., Turpoff, A., Paget, S., Li, C., Almstead, N., Njoroge, F. G., Gu, Z., Graci, J., Jung, S. P., Colacino, J., Lahser, F., Zhao, X., Weetall, M., Nomeir, A., & Karp, G. M. (2015). 6-(Azaindol-2-yl)pyridine-3-sulfonamides as potent and selective inhibitors targeting hepatitis C virus NS4B. *Bioorganic & Medicinal Chemistry Letters*, 25(4), 781–786. <https://doi.org/10.1016/j.bmcl.2014.12.093>
- Chirico, N., & Gramatica, P. (2011). Real external predictivity of QSAR models: How to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *Journal of Chemical Information and Modeling*, 51(9), 2320–2335. <https://doi.org/10.1021/ci200211n>
- Chirico, N., & Gramatica, P. (2012). Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. *Journal of Chemical Information and Modeling*, 52(8), 2044–2058. <https://doi.org/10.1021/ci300084j>
- Cho, N.-J., Dvory-Sobol, H., Lee, C., Cho, S.-J., Bryson, P., Masek, M., Elazar, M., Frank, C. W., & Glenn, J. S. (2010). Identification of a class

- of HCV inhibitors directed against the nonstructural protein NS4B. *Science Translational Medicine*, 2(15), 15ra6. <https://doi.org/10.1126/scitranslmed.3000331>
- Chunduru, S. K., Benatatos, C. A., Theodore, J., Nitz, Bailey, T. R. (n.d.). US 2007/0269420 A1 - Compounds, Compositions And Methods For Treatment And Prophylaxis Of Hepatitis C Viral Infections And Associated Diseases - The Lens - Free & Open Patent and Scholarly Search. Retrieved November 9, 2019, from <https://www.lens.org/lens/patent/192-743-375-133-003/fulltext>
- Consonni, V., & Todeschini, R. (2012). Multivariate analysis of molecular descriptors. In *Statistical Modelling of molecular descriptors in QSAR/QSPR* (pp. 111–147). Wiley-VCH Verlag GmbH & Co. KGaA. <https://doi.org/10.1002/9783527645121.ch4>
- Cox, A. L. (2015). MEDICINE. Global control of hepatitis C virus. *Science* (New York, N.Y.), 349(6250), 790–791. <https://doi.org/10.1126/science.aad1302>
- Cruz, J. N., Costa, J. F. S., Khayat, A. S., Kuca, K., Barros, C. A. L., & Neto, A. M. J. C. (2019). Molecular dynamics simulation and binding free energy studies of novel leads belonging to the benzofuran class inhibitors of Mycobacterium tuberculosis Polyketide Synthase 13. *Journal of Biomolecular Structure & Dynamics*, 37(6), 1616–1627. <https://doi.org/10.1080/07391102.2018.1462734>
- da Costa, K. S., Galúcio, J. M., da Costa, C. H. S., Santana, A. R., Dos Santos Carvalho, V., do Nascimento, L. D., Lima E Lima, A. H., Neves Cruz, J., Alves, C. N., & Lameira, J. (2019). Exploring the potentiality of natural products from essential oils as inhibitors of odorant-binding proteins: A structure- and ligand-based virtual screening approach to find novel mosquito repellents. *ACS Omega*, 4(27), 22475–22486. <https://doi.org/10.1021/acsomega.9b03157>
- Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*, 7(1), 42717. <https://doi.org/10.1038/srep42717>
- de Campos, L. J., & de Melo, E. B. (2017). A QSAR study of integrase strand transfer inhibitors based on a large set of pyrimidine, pyrimidone, and pyridopyrazine carboxamide derivatives. *Journal of Molecular Structure*, 1141, 252–260. <https://doi.org/10.1016/j.molstruc.2017.03.103>
- De Clercq, E. (2014). Current race in the development of DAAs (direct-acting antivirals) against HCV. *Biochemical Pharmacology* 89(4), 441–452. June 15). <https://doi.org/10.1016/j.bcp.2014.04.005>
- Devinyak, O., Havrylyuk, D., & Lesyk, R. (2014). 3D-MoRSE descriptors explained. *Journal of Molecular Graphics and Modelling*, 54, 194–203. <https://doi.org/10.1016/j.jmgs.2014.10.006>
- Dufner-Beattie, J., O'Guin, A., O'Guin, S., Briley, A., Wang, B., Balsarotti, J., Roth, R., Starkey, G., Slomczynska, U., Noueiry, A., Olivo, P. D., & Rice, C. M. (2014). Identification of AP80978, a novel small-molecule inhibitor of hepatitis C virus replication that targets NS4B. *Antimicrobial Agents and Chemotherapy*, 58(6), 3399–3410. <https://doi.org/10.1128/AAC.00113-14>
- Dvory-Sobol, H., Pang, P. S., & Glenn, J. S. (2010). The future of HCV therapy: NS4B as an antiviral target. *Viruses*, 2(11), 2481–2492. <https://doi.org/10.3390/v2112481>
- El-Hasab, M. A. E. M., El-Bastawissy, E. E., & El-Moselhy, T. F. (2018). Identification of potential inhibitors for HCV NS3 genotype 4a by combining protein-ligand interaction fingerprint, 3D pharmacophore, docking, and dynamic simulation. *Journal of Biomolecular Structure & Dynamics*, 36(7), 1713–1727. <https://doi.org/10.1080/07391102.2017.1332689>
- Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T. D., McDowell, R. M., & Gramatica, P. (2003). Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environmental Health Perspectives*, 111(10), 1361–1375. <https://doi.org/10.1289/ehp.5758>
- Fechner, U., Franke, L., Renner, S., Schneider, P., & Schneider, G. (2003). Comparison of correlation vector methods for ligand-based similarity searching. *Journal of Computer-Aided Molecular Design*, 17(10), 687–698. <https://doi.org/10.1023/B:JCAM.0000017375.61558.ad>
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., ... Fox, D. J. (2009). Gaussian 09, Wallingford CT. Retrieved from <http://gaussian.com/>
- Ganta, N. M., Gedda, G., Rathnakar, B., Satyanarayana, M., Yamajala, B., Ahsan, M. J., Jada, S. S., & Balaraju, T. (2019). A review on HCV inhibitors: Significance of non-structural polyproteins. *European Journal of Medicinal Chemistry*, 164, 576–601. <https://doi.org/10.1016/j.ejmech.2018.12.045>
- Golbraikh, A., & Tropsha, A. (2002). Beware of q2. *Journal of Molecular Graphics & Modelling*, 20(4), 269–276. [https://doi.org/10.1016/S1093-3263\(01\)00123-1](https://doi.org/10.1016/S1093-3263(01)00123-1)
- Gouttunoire, J., Penin, F., & Moradpour, D. (2010). Hepatitis C virus non-structural protein 4B: A journey into unexplored territory. *Reviews in Medical Virology*, 20(2), 117–129. <https://doi.org/10.1002/rmv.640>
- Gramatica, P. (2007). Principles of QSAR models validation: Internal and external. *QSAR & Combinatorial Science*, 26(5), 694–701. <https://doi.org/10.1002/qsar.200610151>
- Gramatica, P., Cassani, S., & Chirico, N. (2014). QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS. *Journal of Computational Chemistry*, 35(13), 1036–1044. <https://doi.org/10.1002/jcc.23576>
- Gramatica, P., Chirico, N., Papa, E., Cassani, S., & Kovarich, S. (2013). QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. *Journal of Computational Chemistry*, 34(24), 2121–2132. <https://doi.org/10.1002/jcc.23361>
- Hall, L. H. (2012). Development of structure information from molecular topology for modeling chemical and biological properties: A tribute to the creativity of Lemont Burwell Kier on his 80th birthday. *Current Computer-Aided Drug Design*, 8(2), 93–106. <https://doi.org/10.2174/157340912800492393>
- Hdoufane, I., Stoycheva, J., Tadjer, A., Villemain, D., Najdoska-Bogdanov, M., Bogdanov, J., & Cherqaoui, D. (2019). QSAR and molecular docking studies of indole-based analogs as HIV-1 attachment inhibitors. *Journal of Molecular Structure*, 1193, 429–443. <https://doi.org/10.1016/j.molstruc.2019.05.056>
- Kanwal, F., Kramer, J., Asch, S. M., Chayanupatkul, M., Cao, Y., & El-Serag, H. B. (2017). Risk of Hepatocellular cancer in HCV patients treated with direct-acting antiviral agents. *Gastroenterology*, 153(4), 996–1005.e1. <https://doi.org/10.1053/j.gastro.2017.06.012>
- Küçüküzgel, G., & Çikla-Süzgün, P. (2015). Recent advances bioactive 1,2,4-triazole-3-thiones. *European Journal of Medicinal Chemistry*, 97, 830–870. <https://doi.org/10.1016/j.ejmech.2014.11.033>
- LigandScout, V. 4.3; Inte:Ligand GmbH, Clemesn-Maria-Hofbauer-G. (2344). 6, Maria Enzersdorf, Austria. (n.d.). <http://www.inteligand.com>
- Lin, K., Perni, R. B., Kwong, A. D., & Lin, C. (2006). VX-950, a novel hepatitis C virus (HCV) NS3-4A protease inhibitor, exhibits potent antiviral activities in HCV replicon cells. *Antimicrobial Agents and Chemotherapy*, 50(5), 1813–1822. <https://doi.org/10.1128/AAC.50.5.1813-1822.2006>
- Lundin, M., Monne, M., Widell, A., von Heijne, G., & Persson, M. A. A. (2003). Topology of the membrane-associated hepatitis C virus protein NS4B. *Journal of Virology*, 77(9), 5428–5438. <https://doi.org/10.1128/JVI.77.9.5428-5438.2003>
- Manfroni, G., & Cannalire, R. (2019). Evolution of HCV NS4B inhibitors. In Sofia M. (Eds.), *HCV: The Journey from Discovery to a Cure. Topics in Medicinal Chemistry* (Vol. 32). Cham: Springer. https://doi.org/10.1007/7355_2018_46
- Manvar, P., Shaikh, F., Kakadiya, R., Mehariya, K., Khunt, R., Pandey, B., & Shah, A. (2016). Synthesis of novel imidazo[1,2-a]pyridine-4-hydroxy-2H-coumarins by Groebke-Blackburn-Bienaymé multicomponent reaction as potential NS5B inhibitors. *Tetrahedron*, 72(10), 1293–1300. <https://doi.org/10.1016/j.tet.2016.01.023>
- Mishra, S., Pandey, A., & Manvati, S. (2020). Coumarin: An emerging antiviral agent. *Heliyon*, 6(1), e03217. <https://doi.org/10.1016/j.heliyon.2020.e03217>
- Naud, J., Lemke, C., Goudreau, N., Beaulieu, E., White, P. D., Llinàs-Brunet, M., & Forgione, P. (2008). Potent triazolyl-proline-based inhibitors of HCV NS3 protease. *Bioorganic & Medicinal Chemistry Letters*, 18(11), 3400–3404. <https://doi.org/10.1016/j.bmcl.2008.04.012>

- Pelosi, L. A., Voss, S., Liu, M., Gao, M., & Lemm, J. A. (2012). Effect on Hepatitis C Virus Replication of Combinations of Direct-Acting Antivirals, Including NS5A Inhibitor Daclatasvir. *Antimicrobial Agents and Chemotherapy*, 56(10), 5230–5239. <https://doi.org/10.1128/AAC.01209-12>
- Pinto, V., Araújo, J., Silva, R., da Costa, G., Cruz, J., De A. Neto, M., Campos, J., Santos, C., Leite, F., & Junior, M. (2019). In silico study to identify new antituberculosis molecules from natural sources by hierarchical virtual screening and molecular dynamics simulations. *Pharmaceuticals*, 12(1), 36. <https://doi.org/10.3390/ph12010036>
- Podlogar, B. L., Muegge, I., & Brice, L. J. (2001a). Computational methods to estimate drug development parameters. *Current Opinion in Drug Discovery & Development*, 4(1), 102–109. <https://doi.org/10.1124/pr.112.007336>
- Rai, R., & Deval, J. (2011). New opportunities in anti-hepatitis C virus drug discovery: Targeting NS4B. *Antiviral Research*, 90(2), 93–101. <https://doi.org/10.1016/j.antiviral.2011.01.009>
- Ramos, R. S., Macêdo, W. J. C., Costa, J. S., da Silva, C. H. T. d P., Rosa, J. M. C., da Cruz, J. N., de Oliveira, M. S., de Aguiar Andrade, E. H., e Silva, R. B. L., Souto, R. N. P., & Santos, C. B. R. (2020). Potential inhibitors of the enzyme acetylcholinesterase and juvenile hormone with insecticidal activity: Study of the binding mode via docking and molecular dynamics simulations. *Journal of Biomolecular Structure & Dynamics*, 38(16), 4687–4623. <https://doi.org/10.1080/07391102.2019.1688192>
- Santos, K. L. B. d., Cruz, J. N., Silva, L. B., Ramos, R. S., Neto, M. F. A., Lobato, C. C., Ota, S. S. B., Leite, F. H. A., Borges, R. S., Silva, C. H. T. P. d., Campos, J. M., & Santos, C. B. R. (2020). Identification of novel chemical entities for adenosine receptor type 2A using molecular modeling approaches. *Molecules*, 25(5), 1245. <https://doi.org/10.3390/molecules25051245>
- Santos, C. B. R., Santos, K. L. B., Cruz, J. N., Leite, F. H. A., Borges, R. S., Taft, C. A., ... Silva, C. H. T. P. (2020). Molecular modeling approaches of selective adenosine receptor type 2A agonists as potential anti-inflammatory drugs. *Journal of Biomolecular Structure and Dynamics*, <https://doi.org/10.1080/07391102.2020.1761878>
- Schüürmann, G., Ebert, R.-U., Chen, J., Wang, B., & Kühne, R. (2008). External validation and prediction employing the predictive squared correlation coefficient test set activity mean vs. training set activity mean. *Journal of Chemical Information and Modeling*, 48(11), 2140–2145. <https://doi.org/10.1021/ci800253u>
- Shi, L. M., Fang, H., Tong, W., Wu, J., Perkins, R., Blair, R. M., Branham, W. S., Dial, S. L., Moland, C. L., & Sheehan, D. M. (2001). QSAR Models using a large diverse set of estrogens. *Journal of Chemical Information and Computer Sciences*, 41(1), 186–195. <https://doi.org/10.1021/ci000066d>
- Talete srl. (n.d.). Dragon (Software for Molecular Descriptor Calculation, version 7.0). Retrieved from https://chm.kode-solutions.net/products_dragon.php
- Tao, X., Wang, N., Wang, J., Fu, Z., Gu, Z., Zhang, Y., Chen, S., Wang, L., & Yu, L. (2019). Preclinical evaluation of Amphihevir, a first-in-class clinical Hepatitis C virus NS4B inhibitor. *Antimicrobial Agents and Chemotherapy*, 63(12), e01237-19. <https://doi.org/10.1128/AAC.01237-19>
- Tian, S., Wang, J., Li, Y., Li, D., Xu, L., & Hou, T. (2015). The application of in silico drug-likeness predictions in pharmaceutical research. *Advanced Drug Delivery Reviews*. Elsevier, 86, 2–10. <https://doi.org/10.1016/j.addr.2015.01.009>
- Todeschini, R., & Consonni, V. (2000). Handbook of *molecular descriptors* (Vol. 11). Wiley-VCH. <https://doi.org/10.1002/9783527613106>
- Todeschini, R., & Consonni, V. (2009). *Molecular descriptors for chemoinformatics*. Wiley-VCH.
- Tropsha, A., Gramatica, P., & Gombar, V. (2003). The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR & Combinatorial Science*, 22(1), 69–77. <https://doi.org/10.1002/qsar.200390007>
- Walker, M. A. (1999). Hepatitis C virus: An overview of current approaches and progress. *Drug Discovery Today*, 4(11), 518–529. [https://doi.org/10.1016/S1359-6446\(99\)01414-2](https://doi.org/10.1016/S1359-6446(99)01414-2)
- Wang, Z., Chen, Z., Li, J., Huang, J., Zheng, C., & Liu, J. P. (2020). Combined 3D-QSAR, molecular docking and molecular dynamics study on the benzimidazole inhibitors targeting HCV NS5B polymerase. *Journal of Biomolecular Structure & Dynamics*, 38(4), 1071–1082. <https://doi.org/10.1080/07391102.2019.1593244>
- Yu, G.-Y., Lee, K.-J., Gao, L., & Lai, M. M. C. (2006). Palmitoylation and polymerization of hepatitis C virus NS4B protein. *Journal of Virology*, 80(12), 6013–6023. <https://doi.org/10.1128/JVI.00053-06>
- Zajac, M., Muszalska, I., Sobczak, A., Dadej, A., Tomczak, S., & Jelińska, A. (2019). Hepatitis C – new drugs and treatment prospects. *European Journal of Medicinal Chemistry*, 165, 225–249. <https://doi.org/10.1016/j.ejmech.2019.01.025>
- Zeuzem, S., Andreone, P., Pol, S., Lawitz, E., Diago, M., Roberts, S., Focaccia, R., Younossi, Z., Foster, G. R., Horban, A., Ferenci, P., Nevens, F., Müllhaupt, B., Pockros, P., Terg, R., Shouval, D., van Hoek, B., Weiland, O., Van Heeswijk, R., ... Beumont, M. (2011). Telaprevir for retreatment of HCV infection. *New England Journal of Medicine*, 364(25), 2417–2428. <https://doi.org/10.1056/NEJMoa1013086>
- Zhang, X., Zhang, N., Chen, G., Turpoff, A., Ren, H., Takasugi, J., Morrill, C., Zhu, J., Li, C., Lennox, W., Paget, S., Liu, Y., Almstead, N., Njoroge, F. G., Gu, Z., Komatsu, T., Clausen, V., Espiritu, C., Graci, J., ... Karp, G. M. (2013). Discovery of novel HCV inhibitors: Synthesis and biological activity of 6-(indol-2-yl)pyridine-3-sulfonamides targeting hepatitis C virus NS4B. *Bioorganic & Medicinal Chemistry Letters*, 23(13), 3947–3953. <https://doi.org/10.1016/j.bmcl.2013.04.049>
- Zhang, N., Zhang, X., Zhu, J., Turpoff, A., Chen, G., Morrill, C., Huang, S., Lennox, W., Kakarla, R., Liu, R., Li, C., Ren, H., Almstead, N., Venkatraman, S., Njoroge, F. G., Gu, Z., Clausen, V., Graci, J., Jung, S. P., ... Karp, G. M. (2014). Structure-activity relationship (SAR) optimization of 6-(indol-2-yl)pyridine-3-sulfonamides: Identification of potent, selective, and orally bioavailable small molecules targeting hepatitis C (HCV) NS4B. *Journal of Medicinal Chemistry*, 57(5), 2121–2135. <https://doi.org/10.1021/jm401621g>