Medicinal Chemistry Research (2019) 28:2270–2278
https://doi.org/10.1007/s00044-019-02455-w

MEDICINAL CHEMISTRY RESEARCH

ORIGINAL RESEARCH

Identification of prodigious and under-privileged structural features for RG7834 analogs as Hepatitis B virus expression inhibitor

Vijay H. Masand[1] · Nahed N. E. El-Sayed[2,3] · Vesna Rastija[4] · Mithilesh M. Rathore[1] · Maja Karnaš[4]

Received: 21 May 2019 / Accepted: 5 October 2019 / Published online: 19 October 2019

Abstract

In the present work, QSAR (quantitative structure−activity relationship) analysis has been executed for RG7834 analogs. RG7834 is a first-in-class selective and orally available dihydroquinolizinone (DHQ)-based small molecule Hepatitis B virus expression inhibitor. OECD's guidelines have been followed for developing multiple QSAR models for Hepatitis B virus expression inhibitory activity of 73 RG7834 analogs. The present multiple QSAR models are not only easily interpretable but possess high external predictive ability, as well. These are effective in the recognition of many privileged and underprivileged molecular descriptors, which could be very valuable for the use of these models by the experts and nonexperts of QSAR in future optimizations. The models satisfy threshold values for many fitting, internal and external validation parameters, such as $R^2 = 0.83$, $Q^2 = 0.80$, $CCC_{ext} = 0.88$, etc., thereby demonstrating good external predictive ability of the models. The multiple QSAR and pharmacophoric models successfully identified a good number of important positively and negatively related structural features of RG7834 analogs that govern their Hepatitis B virus expression inhibitory activity. The results could be very beneficial to synthetic/medicinal chemists for future alterations of RG7834 analogs as better drug candidates.

Keywords RG7834 analogs · Hepatitis B virus · Dihydroquinolizinone · QSAR · Pharmacophore modeling

Abbreviations

HBV	Hepatitis B virus
MLR	Multiple linear regression
QSAR	Quantitative structure−activity analysis
WHO	World Health Organization
ADMET	Absorption, Distribution, Metabolism, Excretion and Toxicity
OLS	Ordinary least square
QSARINS	QSAR Insubria
OECD	Organisation for Economic Cooperation and Development
OFS	Objective feature selection
SFS	Subjective feature selection

Introduction

Hepatitis B virus (HBV) infection, which affects the liver, often leads to both acute and chronic hepatitis; thereby, it remains a global serious health problem. In the absence of a strong antibody or cellular immune response, chronic HBV infection generally results in viral persistence. HBV is also related with a great risk of developing cirrhosis and hepatocellular carcinoma. The number of chronic HBV-infected people is estimated to be more than 250 million worldwide with 8 million deaths in 2015, with approximately 4 million deaths from the resulting cirrhosis and hepatocellular carcinoma (Han et al. 2018; Rajbhandari and Chung 2016; Tang et al. 2018; Yuen et al. 2018). Therefore, the number of people as carriers of hepatitis B surface antigen (HBsAg) is a cause of concern (Han et al. 2018; Mueller et al. 2018).

Supplementary information The online version of this article (https://doi.org/10.1007/s00044-019-02455-w) containssupplementary material, which is available to authorized users.

✉ Vijay H. Masand
vijaymasand@gmail.com

[1] Department of Chemistry, Vidya Bharati Mahavidyalaya, Amravati, Maharashtra 444 602, India

[2] Department of Chemistry, College of Science, "Girls Section", King Saud University, PO Box 2457, Riyadh 11451, Saudi Arabia

[3] National Organization for Drug Control and Research, Giza 35521, Egypt

[4] Department of Chemistry, Faculty of Agrobiotechnical Sciences, Josip Juraj Strossmayer University of Osijek, Osijek, Croatia

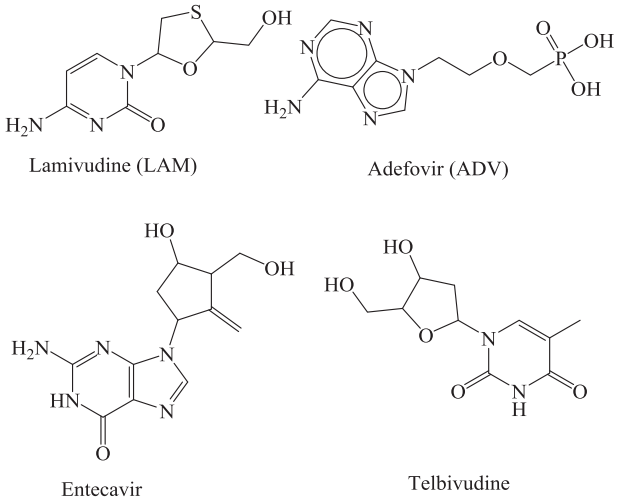

Fig. 1 Marketed drugs (nucleoside analogs) for the clinical treatment of hepatitis B

The drugs employed for the clinical treatment of hepatitis B could be classified as (i) interferons or immune system modulators and (ii) nucleotides analogs (Han et al. 2018; Rajbhandari and Chung 2016; Tang et al. 2018; Yuen et al. 2018). Even though, α-interferons (interferon alpha-2a and PEGylated interferon alpha-2a) are useful in treating chronic HBV, the side effects such as influenza-like symptoms, depression, and insomnia are very serious. The nucleoside analogs such as Lamivudine (LAM), Adefovir (ADV) and Entecavir can be orally administered and have lesser side effects (Li et al. 2015; Mueller et al. 2018) (Fig. 1). Unfortunately, the nucleos(t)ides therapies require a lifelong treatment in most patients. In addition, both types of drugs cannot reduce HBsAg levels effectively (Han et al. 2018; Mueller et al. 2018). The search for new anti-HBV reagents with a novel mechanism of action, better ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) profile with high selectivity is in progress.

Recently, Han et al. reported RG7384 (Han et al. 2018; Mueller et al. 2018), a dihydroquinolizinone (DHQ)-based small molecule, as a new oral HBV viral gene expression inhibitor, which obstructs viral antigen and virion production. It was found to be highly selective for HBV with a novel mechanism of action that can be clearly differentiated from nucleos(t)ide analogs. Consequently, it has been selected for further clinical investigations. Generally, during clinical investigations, the structure of lead compounds is optimized further to retain activity and have desired ADMET profile. In such a situation knowing the important structural features of RG7834 analogs that govern their Hepatitis B virus expression inhibitory activity will be highly advantageous. As many analogs of RG7384 contain a multitude of substituents as well as positional and chain isomers, therefore it is essential that some hidden or underprivileged correlations of structural features, which cannot be identified by visual inspection during SAR analysis, should be identified using advanced techniques like computer-aided drug designing.

CADD has emerged as a popular and thriving alternative to reduce repeated cycles of "synthesis-testing" and for conventional "trial and error" approach. QSAR (quantitative structure−activity relationship), molecular docking, pharmacophore modeling, and other branches of CADD are employed for identifying key structural features and pharmacophoric patterns for optimizing lead and drug candidates. In QSAR, mathematical models that are based on correlations between the desired activity and structural features are built using statistical methodologies. To achieve these objectives, in general, the structural features are represented in numerical form (known as molecular descriptors), statistically validated models are built and subjected to mechanistic interpretation (Aswathy et al. 2018; Chtita et al. 2019; Jisha et al. 2017; (Masand et al. 2018, 2019; Rastija et al. 2018). The interpretation of QSAR model in terms of structural features is useful for drug optimization (Fujita and Winkler 2016; Polishchuk 2017).

The QSAR analysis is generally expected to satisfy two main objectives (Fujita and Winkler 2016): (1) Qualitative QSAR: to recognize the structural features, which have strong correlation with the activity/toxicity profile of a congeneric series of molecules and (2) Quantitative QSAR: to predict the activity/toxicity of a molecule prior to its synthesis and/or biological testing. In the present work, importance is on developing QSAR models that satisfy the criteria for both objectives i.e. qualitative as well as quantitative.

Though Han et al. (Han et al. 2018; Mueller et al. 2018) have reported extensive structure−activity relationship (SAR) for RG7384 analogs, in the present work, extensive QSAR analysis has been performed to identify additional pharmacophoric features. The main objective of this work is to derive statistically robust and easily interpretative, in terms of structural fragments or specific atom, QSAR models with high external predictive ability and to develop consensus pharmacophoric model.

Experimental methodology

Dataset

The dataset (Han et al. 2018; Mueller et al. 2018) selected for the present work comprises 73 structurally diverse DHQ-based small molecules (RG7384 analogs) with different type of substituents like $–NH_2$, –OH, –OMe, etc. at various positions. The activity values expressed as IC_{50}

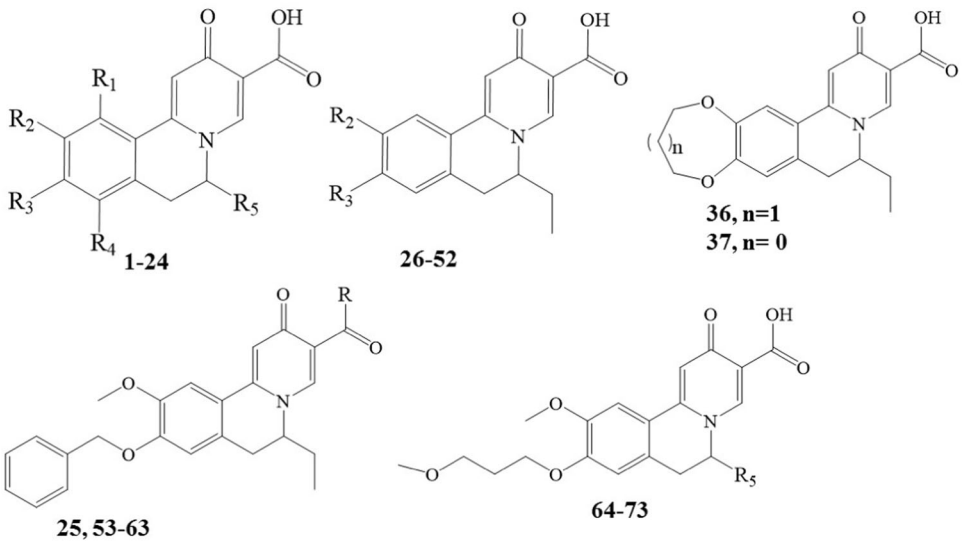

Fig. 2 General structures of molecules in the dataset

have been reported by Yang et al. (Han et al. 2018; Mueller et al. 2018) and follow the same experimental protocol. The activity values IC_{50} represent the concentration required to decrease enzymatic activity by 50%. The reported IC_{50} (μM) activity values were converted to pIC_{50} (M) ($pIC_{50} = -\log_{10} IC_{50}$) for QSAR analysis. The experimental activity values IC_{50}, pIC_{50} and structural substituents have been listed in supplementary material in Table ST0. The general structures of molecules in the dataset have been presented in Fig. 2.

QSAR model building and validation

Structure drawing and optimization

ChemSketch 12 freeware (www.acdlabs.com) was used to draw all the structures and four different sets were prepared, each set comprising 73 molecules. Then, the structures in each set were optimized using different force filed or semi-empirical method (MMFF94, AM1, PM3 and PM6).

Molecular descriptor calculation, pruning and model building

The complete procedure of descriptor calculation, pruning, selection as well as model building and validation was performed on all the four sets. It was observed that the best QSAR models were obtained with MMFF94 optimized set.

MMFF94 force field available in Avogadro software (v. 1.2) (https://avogadro.cc/) using following settings: ForceField: MMFF94, Algorithm: Steepest Descent, number of steps used for optimization: 1000. The molecular descriptor calculations were accomplished using PyDescriptor (Masand and Rastija 2017) and PaDEL (Yap 2011). This led to a myriad cluster of more than 28,000 molecular descriptors. Then, redundant molecular descriptors, that is, highly correlated ($R > 0.99$), constant and nearly constant variables were eliminated using QSARINS 2.2.2 (Gramatica et al. 2013, 2014). This resulted in a diminished set of descriptors encompassing 723 molecular descriptors only, still covering broad structural and chemical space viz. constitutional (0D−), mono-dimensional (1D−), bi-dimensional (2D−) and three-dimensional (3D−).

For thriving analysis, multiple QSAR models were derived using undivided and randomly divided (training-80% and prediction-20% sets prior to descriptor selection) dataset to ensure the capturing of all the relevant structural and activity information. The training set was used for model development, and the prediction set for the evaluation of predictive ability on new chemicals.

Subjective descriptor selection (SDS) involves selection of optimum number and set of molecular descriptors using appropriate feature selection technique like stepwise regression, genetic algorithm, ant colony, etc. In this work, genetic algorithm (GA) available in QSARINS 2.2.2 was used. QSARINS 2.2.2 was used to select optimal number and set of five descriptors with default settings employing Q^2 as the fitness function, which helps to avoid naive Q^2, as well (Hawkins et al. 2003). For every procedure used for model building, descriptor selection was completed separately every time, that is when using undivided dataset, divided dataset (training: 80%, prediction: 20%). The analysis indicated that up to five descriptors there was growth in the value of Q^2, but then, it had visible and significant reduction. Hence, to eliminate overfitting and generate easy and informative Genetic Algorithm-Multilinear Regression (GA-MLR)-based QSAR models, the molecular descriptor selection was restricted to a set of five descriptors. The selected molecular descriptors used in models 1–3 are presented in the supplementary information. OECD principles for QSAR model

validation were followed while developing the statistically acceptable QSAR models; only models with high internal and external predictive abilities have been reported.

The newly developed QSAR models were accepted only when they have values above the recommended threshold values for different statistically criteria for internal validation and external predictive ability along with Y-randomization. The following recommended threshold values for statistical criteria were considered while accepting any QSAR model (Chtita et al. 2019; Mantoani et al. 2019; Masand et al. 2018, 2019; Masand and Rastija 2017): $R^2_{tr} \geq 0.6$, $Q^2_{loo} \geq 0.5$, $Q^2_{LMO} \geq 0.6$, $R^2 > Q^2$, $R^2_{ex} \geq 0.6$, $RMSE_{tr} < RMSE_{cv}$, $\Delta K \geq 0.05$, $CCC \geq 0.80$, $Q^2\text{-}F^n \geq 0.60$, $r^2_m \geq 0.5$, $(1 - r^2/r_o^2) < 0.1$, $0.9 \leq k \leq 1.1$ or $(1 - r^2/r'^2_o) < 0.1$, $0.9 \leq k' \leq 1.1$, $|r_o^2 - r'^2_o| < 0.3$ with RMSE and MAE as low as possible. The formulae for calculating various statistical parameters is available in supplementary material. This confirmed purging of overfitting and spurious models. The mean value of Q^2_{LMO} has been reported after repeating the LMO for 2000 times with 30% of the objects left out randomly from the training set each time. The external validation parameters viz. $RMSE_{ex}$, MAE_{ex}, R^2_{ex}, Q^2_{F1}, Q^2_{F2}, Q^2_{F3}, and CCC_{ex} were used to assess the external predictive ability of models (Cherkasov et al. 2014; Consonni et al. 2009, 2019; Golbraikh et al. 2003); (Gramatica 2007, 2012−2014; Roy et al. 2011; Tropsha 2012; Tropsha et al. 2003). Any QSAR model having poor internal validation or external predictivity was rejected. Model validation was carried out also by checking the model applicability domain (AD), i.e. Williams plots.

Consensus pharmacophore modeling

The method involves selection of top five most active molecules (see Table 1). The structures were optimized using MMFF94 force field, followed by their alignment on the basis of the common scaffold using Open3DAlign (http://open3dalign.sourceforge.net/). The aligned structures were used to generate consensus pharmacophore model using PyMOL 2.2 (www.pymol.org) and its plugin LIQUID 1.0 with default settings.

Results and discussion

QSAR models

Generally, researchers employ step-wise regression, GA, etc. algorithms for subjective feature selection (SFS), which leads to generation of several MLR models, often having highly comparable statistical performance but comprise dissimilar descriptors. In such a situation, a QSAR modeler selects only one MLR model on the basis of model's statistical robustness ("first among equals" approach (Chtita et al. 2019; Mantoani et al. 2019; Masand et al. 2014a, b, 2018, 2019; Masand and Rastija 2017). But this approach has many drawbacks (Chtita et al. 2019; Mantoani et al. 2019; (Masand et al. 2014a, b, 2018, 2019; Masand and Rastija 2017): (1) if the QSAR model contains complex/esoteric descriptors, then a straight and easier correlation of descriptors with appropriate structural features is quite difficult,

(2) the chosen QSAR model may be biased due to:

i. splitting pattern and method,
ii. composition of training and prediction sets,
iii. algorithm employed for feature selection, and
iv. the presence of high influencing molecules in the training/prediction set.

To circumvent these limitations of "first among equals" approach, developing multiple QSAR models based on multiple splitting is a possible solution. One more advantage of this solution lies in its ability to recognize underprivileged but significant pharmacophoric features associated with the HBV activity of RG-7384 analogs. Therefore, in the present study, multiple QSAR models have been constructed. The five-parameter-based QSAR models are given as follows.

Model 1 (full set model)

$$pIC_{50} = 12.59 \quad (\pm 1.55) + 9.12 \quad (\pm 1.25) \times \text{logGGI6} - 0.2 \quad (\pm 0.1) \times \text{com\_aroC\_6A} - 1.81 \quad (\pm 0.28) \times \text{sqrt (don\_N\_3B)} - 0.11 \quad (\pm 0.04) \times \text{ringC\_MSA3} - 0.71 \quad (\pm 0.43) \times \text{flipoacc5B}.$$

Table 1 The experimental activity values IC_{50} and structures of DHQ used for consensus pharmacophore modeling

S.N.	SMILES	IC_{50} (μM)	pIC_{50} (M)
61	N32C(Cc1c(cc(c(c1)OCCCOC)OC)C2=CC(=O)C(=C3)C(=O)O)C(C)(C)C	0.001	9
62	N32C(Cc1c(cc(c(c1)OCCCOC)OC)C2=CC(=O)C(=C3)C(=O)O)C4(CC4)C	0.002	8.699
73	s1c(ccc1)C4N2C(=CC(=O)C(=C2)C(=O)O)c3c(cc(c(c3)OC)OCCCOC)C4	0.003	8.523
65	N32C(Cc1c(cc(c(c1)OCCCOC)OC)C2=CC(=O)C(=C3)C(=O)O)C(COC)(C)C	0.005	8.301
67	N32C(Cc1c(cc(c(c1)OCCCOC)OC)C2=CC(=O)C(=C3)C(=O)O)C(CO)(C)C	0.006	8.222

Divided set models (80% training, 20% prediction)

Model 2

$pIC_{50} = 10.76 \quad (\pm 1.49) + 9.33 \quad (\pm 1.33) \times logGGI6 - 0.26 \quad (\pm 0.11) \times com\_aroC\_6A + 1.84 \quad (\pm 0.56) \times acc\_N\_4Ac + 1.36 \quad (\pm 0.54) \times AD2D102 - 2.05 \quad (\pm 0.74) \times H\_acc\_3Bc.$

Model 3

$pIC_{50} = 2.28 \quad (\pm 1.87) - 0.41 \quad (\pm 0.16) \times com\_aroC\_6A + 0.47 \quad (\pm 0.08) \times com\_C\_7A + 0.6 \quad (\pm 0.23) \times com\_ringCminus\_4A - 1.28 \quad (\pm 0.32) \times sqrt \ (don\_N\_3B) - 0.27 \quad (\pm 0.27) \times com\_ringChyd\_5A.$

The statistical performance of all the developed models 1–3 have been tabulated in Table 2 and ST2 (see supplementary material).

The high values of R^2, Q^2, R^2_{ex}, Q^2_{Fn}, CCC_{ex} and other statistical parameters (see Table 2 and ST2 in supplementary material) indicate that the models 1–3 are statistically acceptable and possess good external predictive ability (Cherkasov et al. 2014; Consonni et al. 2009, 2019; Golbraikh et al. 2003; Gramatica 2007, 2014; Gramatica et al. 2012, 2013; Roy et al. 2011; Tropsha 2012; Tropsha et al. 2003). These parameters, along with low correlation among the molecular descriptors, point out that these models are not developed by chance (see supplementary information). In addition, for a better validation of derived models, the model applicability domain (AD) was assured by plotting Williams plots for models 1–3 and included in supplementary material. The experimental, predicted pIC_{50}, by models 1–3, values and status for various molecules have been included in Table ST1 (see supplementary material) (Fig. 3).

Table 2 Statistical parameters of the developed QSAR models

S. no.	Statistical parameter	Model 1	Model 2	Model 3
1.	N_{tr}	73	59	59
2.	N_{ex}	00	14	14
3.	R^2_{tr}	0.83	0.83	0.83
4.	$RMSE_{tr}$	0.47	0.48	0.49
5.	CCC_{tr}	0.91	0.91	0.91
6.	R^2_{cv} (Q^2loo)	0.8	0.80	0.80
7.	$RMSE_{cv}$	0.51	0.53	0.53
8.	CCC_{cv}	0.89	0.89	0.89
9.	Q^2_{LMO}	0.79	0.78	0.78
10.	R^2_{Yscr}	0.07	0.09	0.09
11.	$RMSE_{ex}$	–	0.52	0.52
12.	R^2_{ex}	–	0.77	0.79
13.	Q^2-F^1	–	0.77	0.75
14.	Q^2-F^2	–	0.76	0.75
15.	Q^2-F^3	–	0.81	0.81
16.	CCC_{ex}	–	0.88	0.88

As stated earlier, a QSAR model is built for two main reasons: (1) to predict the activity of as-yet untested molecules and (2) to get the relevant information useful for lead optimization i.e. interpretation of the model for mechanistic details. It is clear that the models 1–3 are based on different types of molecular descriptors. Multiple QSAR models development, based on different splitting, led to capturing of different important descriptors of high and low privilege.

The molecular descriptor logGGI6 is a topological charge descriptor that stands for topological charge index of order 6. The logarithm of this 2D- molecular descriptor has positive coefficients in models 1 and 2. This indicates that its value must be kept higher to have good activity. For e.g. the two active molecules **65** ($IC_{50} = 0.005\ \mu M$) and **67** ($IC_{50} = 0.006\ \mu M$) have high value of this descriptor 0.19 and 0.18, respectively. In the case of the least active molecule of the present congeneric series, the molecule **4** ($IC_{50} = 52\ \mu M$) has the lowest value for this descriptor ($logGGI6 = -0.41$).

The molecular descriptor com_aroC_6A, which represents number of carbon atoms of aromatic ring that are present within 6 Å from center of mass (com) of a molecule (see Fig. 4a), has negative coefficients in all the three models. Consequently, its value must be lowered to have a better activity profile. This could be achieved by avoiding aromatic rings (like benzene, naphthalene, etc.) within 6 Å from center of mass (com) of a molecule. A similar molecular descriptor com_ringChyd_5A (number of hydrophobic carbon atoms of the ring that are present within 5 Å from center of mass (com) of a molecule) has a negative coefficient in model 3. Therefore, such carbon atoms must be avoided to have a better activity. The negative effect of the presence of the ring carbon atoms is further indicated by the negative coefficient for the molecular descriptor ringC_MSA3 in model 1. The molecular descriptor ringC_MSA3 stands for the molecular surface area of carbon atoms possessing partial charge in the range 0 to 0.099. A similar molecular descriptor but having positive coefficient in model 3 is com_ringCminus_4A, which stands for number of negatively charged carbon atoms of the ring that are present within 4 Å from center of mass (com) of a molecule. Therefore, a good strategy to increase activity is to increase such carbon atoms. To add further, the molecular descriptor com_C_7A (number of carbon atoms that are present within 7 Å from center of mass (com) of a molecule) has a positive coefficient in model 3. Thereby, increasing the number of carbon atoms within 7 Å from com could be beneficial for enhancing the activity. This descriptor has been depicted in Fig. 4b.

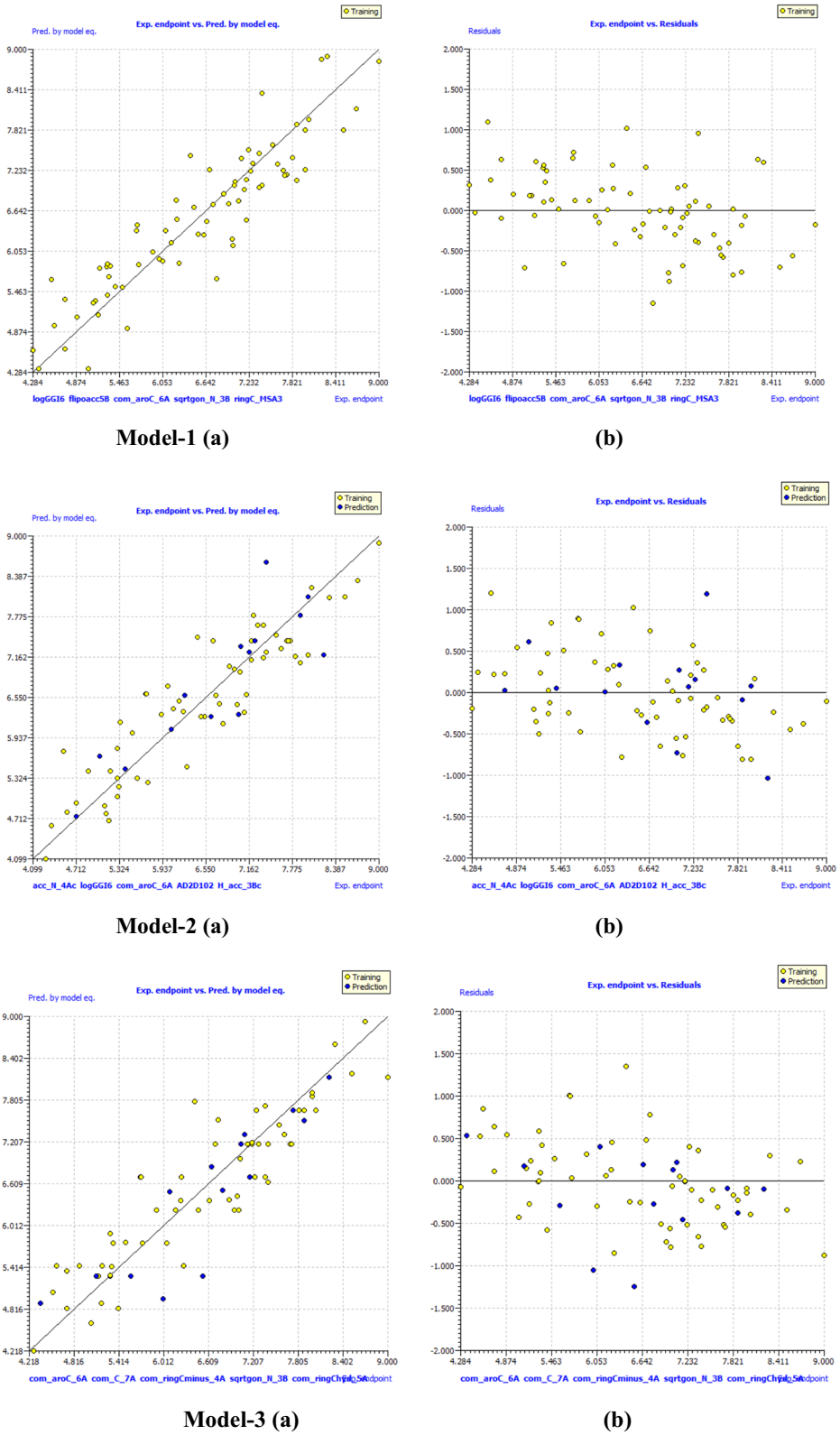

Fig. 3 **a** Graph of experimental vs. predicted values pIC_{50} values; **b** graph of experimental vs. residual values for models 1–3

The molecular descriptors com_aroC_6A, com_ringChyd_5A, ringC_MSA3, com_C_7A and com_ringCminus_4A are connected with the presence of carbon atoms within 4–7 Å from center of mass (com) of a molecule, except ringC_MSA3. Interestingly, all of these descriptors provide different level and type of consensus and complementary information. A harmonious consideration of these molecular descriptors indicates that these molecular

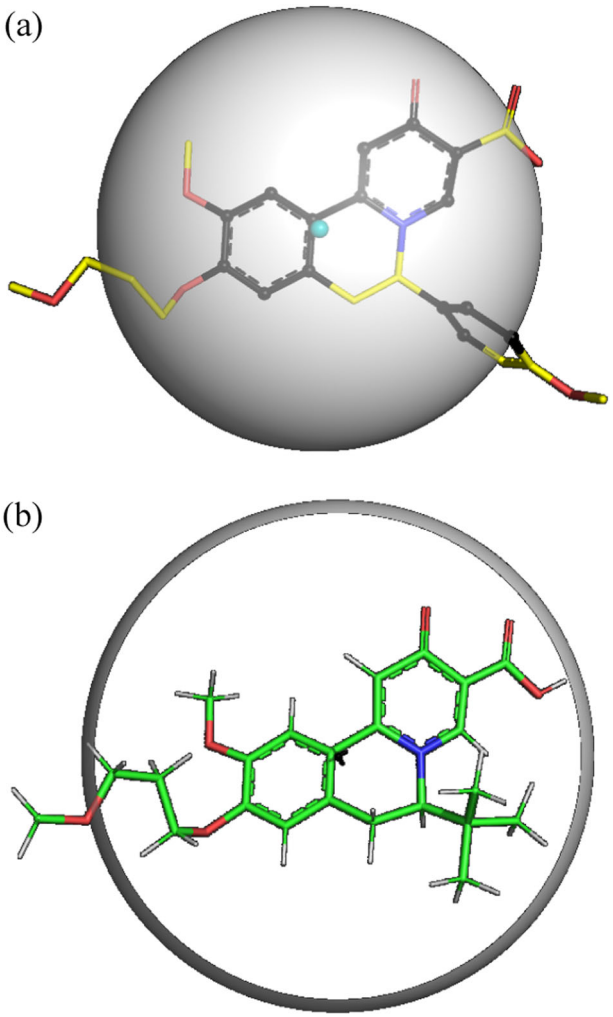

Fig. 4 Pictorial representation of some of the selected molecular descriptors used to build the QSAR models. (**a**) Com_aroC_6A (using 3r-74 as a representative only) (aromatic Carbons are black while com as cyan colored); (**b**) com_C_7A (using 3r-74 as a representative only) (com as black colored) (color figure online)

descriptors point out for the local environment in the vicinity of center of mass of a molecule which could be used for favorable activity profile. To summarize the combined effect of these molecular descriptors, in future optimizations, high priority be given to negatively charged carbon atoms in the vicinity of com, while hydrophobic (neutral) and aromatic/ring carbon atoms must be discarded. These observations are supported by the fact that the less active molecules **4** ($IC_{50} = 52\ \mu M$) and **9** ($IC_{50} = 29\ \mu M$) have the lowest number of negatively charged carbon atoms of the ring that are present within 4 Å from center of mass (com) of a molecule (com_ringCminus_4A = 4 for both the molecules). These results could be useful to extend the search of a suitable lead candidate for future drug research.

The molecular descriptor acc_N_4Ac, which stands for sum of partial charges on N atoms present within 4 Å from H-bond acceptor atoms, has a positive coefficient in model 2. Hence, its value must be as high as possible. To achieve this, a good solution is to increase substitutions on such N atoms using equitably electronegative atoms.

The molecular descriptor flipoacc5B represents frequency of occurrence of H-bond acceptor atoms within five bonds from lipophilic atoms. It has a negative coefficient in model 1; therefore, its value must be minimized. To achieve this goal, such a combination of H-bond acceptors and lipophilic atoms must be avoided.

The molecular descriptor sqrt (don_N_3B), which stands for square root of number of nitrogen atoms present within three bonds from H-bond donor atoms, is a descriptor with negative coefficients in models 1 and 3. Therefore, there is a requirement to minimize the value of this descriptor. A simple comparison of molecule **54** ($IC_{50} = 43\ \mu M$) with **55** ($IC_{50} = 13\ \mu M$), and **56** ($IC_{50} = 6.6\ \mu M$) with **57** ($IC_{50} = 0.52\ \mu M$) vindicates this observation.

Another molecular descriptor with a negative coefficient is H_acc_3Bc in model 2. This descriptor represents sum of partial charges on hydrogen atoms which are separated from H-bond acceptor atom by three bonds. Therefore, decreasing its value could lead to a better activity. For e.g. the low activity molecule **39** ($IC_{50} = 26.1\ \mu M$) has high value for H_acc_3Bc (=1.69).

One more molecular descriptor with a positive coefficient in model 2, thus having a promising influence on activity, is AD2D102. It represents the presence of two oxygen atoms at a topological distance of 2. This observation is supported by the difference in the activity of **25** ($IC_{50} = 0.043\ \mu M$) and **58** ($IC_{50} = 19\ \mu M$) as well as by the consensus pharmacophore model (depicted in Fig. 5). The consensus pharmacophoric model indicates that the presence of H-bond acceptor region (big red contour) and H-bond donor region (blue contour) in the vicinity of the ring A are beneficial for activity augmentation. These regions are created due to the presence of two oxygen atoms of carboxylic acid group attached to ring A. Therefore, this -COOH group must be retained in future optimizations.

Consensus pharmacophore model

The consensus pharmacophore model has been depicted in Fig. 5. It indicates that the activity has profound relation with one lipophilic region (green contour), three H-bond acceptor regions (red contours) and one H-bond donor region (blue contour). The green contour in the vicinity of the ring carbon atoms highlights the importance of the presence of lipophilic groups in this region. The red contours in the proximity of ring A and C are important for a

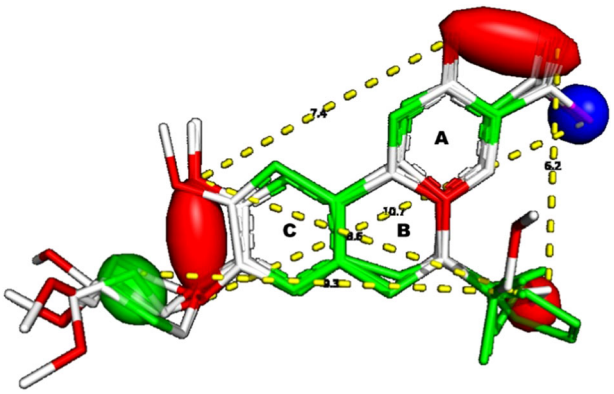

Fig. 5 Consensus pharmacophore model for (red: H-Bond acceptor, blue: H-Bond donor and green: Lipophilic region) (color figure online)

better activity profile. The presence of H-bond acceptor and donor in the vicinity of the ring A is also vindicated by the QSAR analysis (see molecular descriptor AD2D102).

It is important that QSAR and pharmacophore analyses provided equivalent as well as complementary results. The results of these analyses pointed out that some structural features like retaining the -COOH group on ring A, higher importance to negatively charged carbon atoms in the vicinity of com (center of mass) as well as limiting the number of hydrophobic (neutral) and aromatic/ring carbon atoms. A coordinated use of various results of this work could significantly extend the drug research for dihydroquinolizinone (DHQ)-based small molecule Hepatitis B virus expression inhibitor.

Conclusion

The QSAR and pharmacophore models derived in the present work pointed out that specific combination of H-bond acceptor and donor with each other and with certain atoms at specific distances (AD2D102, H_acc_3Bc, don_N_3B, flipoacc5B, etc.) have significant influence on the anti-hepatitis activity of RG7834 analogs.

The present work was not only successful in identifying many additional pharmacophoric features which have neither been reported earlier using SAR nor any other technique. These models could be used for prediction of activity prior to synthesis-testing, as the developed QSAR models have high external predictive ability.

Acknowledgements The authors would like to extend their sincere appreciation to the Deanship of Scientific Research at King Saud University for funding this research (Group No. RG-1435-083). Authors are grateful to Dr. Paola Gramatica and QSARINS-Chem developing team for providing QSARINS-Chem 2.2.2.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

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

References

Aswathy L, Jisha RS, Masand VH, Gajbhiye JM, Shibi IG (2018) Design of novel amyloid β aggregation inhibitors using QSAR, pharmacophore modeling, molecular docking and ADME prediction. Silico Pharmacol 6(1):12

Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, Dearden J, Gramatica P, Martin YC, Todeschini R, Consonni V, Kuz'min VE, Cramer R, Benigni R, Yang C, Rathman J, Terfloth L, Gasteiger J, Richard A, Tropsha A (2014) QSAR modeling: where have you been? Where are you going to? J Med Chem 57(12):4977–5010

Chtita S, Ghamali M, Ousaa A, Aouidate A, Belhassan A, Taourati AI, Masand VH, Bouachrine M, Lakhlifi T (2019) QSAR study of anti-Human African Trypanosomiasis activity for 2-phenylimidazopyridines derivatives using DFT and Lipinski's descriptors. Heliyon 5(3):e01304

Consonni V, Ballabio D, Todeschini R (2009) Comments on the definition of the Q2 parameter for QSAR validation. J Chem Inf Model 49(7):1669–1678

Consonni V, Todeschini R, Ballabio D, Grisoni F (2019) On the misleading use of QF32 for QSAR model comparison. Mol Inform 38(1−2):e1800029

Fujita T, Winkler DA (2016) Understanding the roles of the "two QSARs". J Chem Inf Modeling 56(2):269–274

Golbraikh A, Shen M, Xiao Z, Xiao YD, Lee KH, Tropsha A (2003) Rational selection of training and test sets for the development of validated QSAR models. J Comput Aided Mol Des 17(2−4):241–253

Gramatica P (2007) Principles of QSAR models validation internal and external. QSAR Combinatorial Sci 26(5):694–701

Gramatica P, Cassani S, Roy PP, Kovarich S, Yap CW, Papa E (2012) QSAR modeling is not push a button and find a correlation: a case study of toxicity of (benzo-)triazoles on algae. Mol Inform [Online] 31:817–835

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. J Computational Chem 34 (24):2121–2132

Gramatica P (2014) External evaluation of QSAR models, in addition to cross-validation verification of predictive capability on totally new chemicals. Mol Inform 33(4):311–314

Gramatica P, Cassani S, Chirico N (2014) QSARINS-chem: insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS. J Computational Chem 35(13):1036–1044

Han X, Zhou C, Jiang M, Wang Y, Wang J, Cheng Z, Wang M, Liu Y, Liang C, Wang J, Wang Z, Weikert R, Lv W, Xie J, Yu X, Zhou X, Luangsay S, Shen HC, Mayweg AV, Javanbakht H, Yang S (2018) Discovery of RG7834: the first-in-class selective and orally available small molecule hepatitis B virus expression inhibitor with novel mechanism of action. J Medicinal Chem 61 (23):10619–10634

Hawkins DM, Basak SC, Mills D (2003) Assessing model fit by cross-validation. J Chem Inf Comput Sci 43(2):579–586

Jisha RS, Aswathy L, Masand VH, Gajbhiye JM, Shibi IG (2017) Exploration of 3,6-dihydroimidazo(4,5-d)pyrrolo(2,3-b)pyridin-2

(1H)-one derivatives as JAK inhibitors using various in silico techniques. Silico Pharmacol 5(1):9

Li X, Jie Y, You X, Shi H, Zhang M, Wu Y, Lin G, Li X, Gao Z, Chong Y (2015) Optimized combination therapies with adefovir dipivoxil (ADV) and lamivudine, telbivudine, or entecavir may be effective for chronic hepatitis B patients with a suboptimal response to ADV monotherapy. Int J Clin Exp Med 8(11):21062–21070

Mantoani SP, de Andrade P, Chierrito TPC, Figueredo AS, Carvalho I (2019) Potential triazole-based molecules for the treatment of neglected diseases. Curr Med Chem 26(23):4403–4434

Masand VH, Mahajan DT, Ben Hadda T, Jawarkar RD, Alafeefy AM, Rastija V, Ali MA (2014a) Does tautomerism influence the outcome of QSAR modeling? Medicinal Chem Res 23:1742–1757

Masand VH, Mahajan DT, Gramatica P, Barlow J (2014b) Tautomerism and multiple modelling enhance the efficacy of QSAR: antimalarial activity of phosphoramidate and phosphorothioamidate analogues of amiprophos methyl. Medicinal Chem Res 23(11):4825–4835

Masand VH, Rastija V (2017) PyDescriptor: A new PyMOL plugin for calculating thousands of easily understandable molecular descriptors. Chemometrics Intell Lab Syst 169:12–18

Masand VH, El-Sayed NNE, Bambole MU, Quazi SA (2018) Multiple QSAR models, pharmacophore pattern and molecular docking analysis for anticancer activity of α, β-unsaturated carbonyl-based compounds, oxime and oxime ether analogues. J Mol Struct 1157:89–96

Masand VH, El-Sayed NNE, Bambole MU, Patil VR, Thakur SD (2019) Multiple quantitative structure−activity relationships (QSARs) analysis for orally active trypanocidal N-myristoyltransferase inhibitors. J Mol Struct 1175:481–487

Mueller H, Wildum S, Luangsay S, Walther J, Lopez A, Tropberger P, Ottaviani G, Lu W, Parrott NJ, Zhang JD, Schmucki R, Racek T, Hoflack J-C, Kueng E, Point F, Zhou X, Steiner G, Lütgehetmann M, Rapp G, Volz T, Dandri M, Yang S, Young JAT, Javanbakht H (2018) A novel orally available small molecule that inhibits hepatitis B virus expression. J Hepatol 68(3):412–420

Polishchuk P (2017) Interpretation of quantitative structure–activity relationship models: past, present, and future. J Chem Inf Modeling 57(11):2618–2639

Rajbhandari R, Chung RT (2016) Treatment of Hepatitis B: a concise review. Clin Transl Gastroenterol 7(9):e190

Rastija V, Molnar M, Siladi T, Masand VH (2018) QSAR analysis for antioxidant activity of dipicolinic acid derivatives. Combinatorial Chem High Throughput Screen 21(3):204–214

Roy PP, Kovarich S, Gramatica P (2011) QSAR model reproducibility and applicability: a case study of rate constants of hydroxyl radical reaction models applied to polybrominated diphenyl ethers and (benzo-)triazoles, Journal of Computational Chemistry Volume 32, Issue 11. J Computational Chem 32 (11):2386–2396

Tang LSY, Covert E, Wilson E, Kottilil S (2018) Chronic hepatitis B infection. JAMA 319(17):1802–1813

Tropsha A, Gramatica P, Gombar VK (2003) The importance of being earnest validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Combinatorial Sci 22(1):69–77

Tropsha A (2012) Recent trends in statistical QSAR modeling of environmental chemical toxicity. EXS 101:381–411

Yap CW (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Computational Chem 32(7):1466–1474

Yuen M-F, Chen D-S, Dusheiko GM, Janssen HLA, Lau DTY, Locarnini SA, Peters MG, Lai C-L (2018) Hepatitis B virus infection. Nat Rev Disease Primers 4(1):18035