Chemometrics and Intelligent Laboratory Systems 154 (2016) 52–61

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Exploring the structure–activity relationship of oxazolidinones as HIV-1 protease inhibitors—QSAR and pharmacophore modelling studies

Veerasamy Ravichandran [a,*], Krishnamoorthy Venkateskumar [a], Sivadasan Shalini [a], Rajak Harish [b]

[a] Faculty of Pharmacy, AIMST University, Semeling 08100, Kedah, Malaysia
[b] SLT Institute of Pharmaceutical Sciences, Guru Ghasidas University, Bilaspur, India

ARTICLE INFO

Article history:
Received 24 August 2014
Received in revised form 13 March 2016
Accepted 14 March 2016
Available online 24 March 2016

Keywords:
HIV-1protease
Inhibition
QSAR
Oxazolidinones
Pharmacophore mapping

ABSTRACT

In the present study, 2D QSAR and 3D QSAR models and pharmacophore hypothesis were evaluated for a series of N-aryl-oxazolidinone-5-carboxamides to predict their HIV-1 protease inhibitory activity. The developed QSAR models were validated by external validation method, leave-one-out and leave-many-outcross validation, Y-randomization method and applicability domain analysis. The primary findings of this study were that the number of carbon atoms separated from any specific carbon atom by 2- and 7-bond distances, and the number of fluorine atoms separated from any specific fluorine atom by a 5-bond distance in a molecule, altered the compounds' inhibitory action on HIV-1 protease. Further, 3D QSAR study results indicated that the presence of electrostatic and steric field descriptors in N-aryl-oxazolidinone-5-carboxamides significantly inhibited HIV-1 protease. The generated pharmacophore hypothesis of the compounds indicated the significance of the two aromatic and three hydrogen bond acceptor features on HIV-1 protease's inhibitory activity. The proposed model also provided a better understanding of HIV-1 protease's inhibitory activity on oxazolidinones and could be used as guidance for the proposition of new anti-HIV agents.

1. Introduction

HIV-1 (human immunodeficiency virus type-1) is the pathogenic retrovirus and causative agent of AIDS or AIDS-related complex (ARC) [1,2]. Acquired immunodeficiency syndrome (AIDS) is a dreadful virulent disease still causing havoc worldwide. The shattering potential of this viral disease has not been fully realized. The causative moiety of this disease is human immunodeficiency virus (HIV), which is a retrovirus of the lentivirus family [3]. Enzymes such as reverse transcriptase, protease and integrase encoded by the gag and gag-pol genes of HIV play important roles in viral replication cycle [4]. Highly active antiretroviral therapy (HAART) made dramatic impacts on mortality and morbidity associated with HIV infection. However, current drug regimens include major limitations such as long-term toxicity, drug resistance and mutations that lead to development of resistance to HIV [5]. Therefore, development of a novel, specifically targeted antiviral therapy for HIV is imminent and meaningful in the anti-HIV research area.

Quantitative structure–activity relationship (QSAR) models have found fertile grounds in the fields of chemical/biological chemistry and related sciences, especially within the areas of computer aided drug design. Various QSAR models have been developed using characteristic parameters of molecular structures and experimental property/activity of compounds to improve the optimal performance of new chemicals and reduce the cost of exploring drugs.

The 2D QSAR studies [6–20], comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) have been carried out among different groups of compounds in rational drug design and related applications [21–33] of anti-HIV activity of molecules. As part of on-going efforts to design novel molecules with potent HIV-1 protease inhibitory activity, 2D and 3D QSAR and pharmacophore mapping analysis were performed on N-aryl-oxazolidinone-5-carboxamides [34] (Table 1) to correlate their HIV-1 protease inhibitory activity with physicochemical, sterical and electrostatic parameters, as well as pharmacophoric features. The models were validated by dividing the dataset into training and test sets, and by using several validation criteria to evaluate the predictive ability of the established models. The established models could provide some valuable information about structural modifications for designing new possible lead compounds with higher activity.

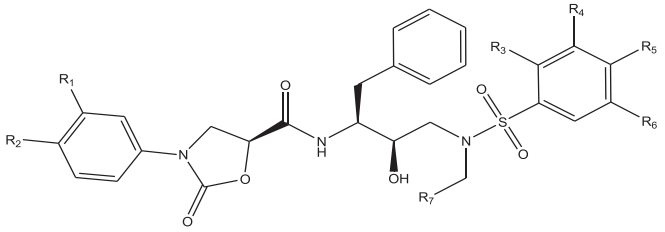
2. Materials and methods

Molecular modelling studies were performed using the software VLife MDS 4.3 (a product of VLife Sciences Technologies Private Limited, India: www.vlifesciences.com). Some of the statistical parameters were calculated using QSARINS (www.qsar.it) [35].

* Corresponding author at: Pharmaceutical Chemistry Unit, Faculty of Pharmacy, AIMST University, Semeling 08100, Kedah, Malaysia. Tel.: +6 4 4298000x1029; fax: +6 44298007.
E-mail address: phravi75@rediffmail.com (V. Ravichandran).

http://dx.doi.org/10.1016/j.chemolab.2016.03.017
0169-7439/

Table 1
Structures and HIV-1 protease inhibitory activity of the oxazolidinone derivatives.



Compd. no.	R_1	R_2	R_3	R_4	R_5	R_6	R_7	pKi (nM)
1	H	H	H	H	OCH_3	H	iPr	10.000
2	F	H	H	H	OCH_3	H	iPr	10.080
3	F	F	H	H	OCH_3	H	iPr	10.180
4	CF_3	H	H	H	OCH_3	H	iPr	11.222
5	Ac	H	H	H	OCH_3	H	iPr	12.097
6	H	Ac	H	H	OCH_3	H	iPr	11.398
7	OCH_3	H	H	H	OCH_3	H	iPr	10.347
8	H	H	H	H	NH_2	H	iPr	9.276
9	F	H	H	H	NH_2	H	iPr	9.769
10	F	F	H	H	NH_2	H	iPr	9.638
11	CF_3	H	H	H	NH_2	H	iPr	10.377
12	Ac	H	H	H	NH_2	H	iPr	10.495
13	H	Ac	H	H	NH_2	H	iPr	9.735
14	F	H	H	–O–CH_2–O–		H	iPr	9.971
15	F	F	H	–O–CH_2–O–		H	iPr	10.071
16	CF_3	H	H	–O–CH_2–O–		H	iPr	10.796
17	Ac	H	H	–O–CH_2–O–		H	iPr	11.222
18	H	Ac	H	–O–CH_2–O–		H	iPr	10.796
19	F	H	H	F	OCH_3	H	iPr	10.155
20	F	F	H	F	OCH_3	H	iPr	9.465
21	CF_3	H	H	F	OCH_3	H	iPr	10.143
22	Ac	H	H	F	OCH_3	H	iPr	9.876
23	H	Ac	H	F	OCH_3	H	iPr	10.097
24	CF_3	H	H	H	OCH_3	H	iPr	8.000
25	Ac	H	H	H	OCH_3	H	iPr	8.699
26	H	H	H	OCH_3	H	H	iPr	8.420
27	H	Ac	H	OCH_3	H	H	iPr	9.076
28	F	H	H	H	H	H	cPr	9.590
29	F	F	H	H	OCH_3	H	cPr	9.237
30	H	Ac	H	H	OCH_3	H	cPr	9.097
31	H	H	H	OCH_3	H	H	2-TP	6.622
32	F	H	H	OCH_3	H	H	2-TP	6.724
33	H	Ac	H	OCH_3	H	H	2-TP	7.530
34	H	H	F	H	F	F	2-TP	6.769
35	F	H	F	H	F	F	2-TP	6.795
36	H	Ac	F	H	F	F	2-TP	6.775
37	H	H	H	OCH_3	H	H	2-THF	7.377
38	F	H	H	OCH_3	H	H	2-THF	6.824

2-TP = 2-thiophene, 2THF = 2-tetrahydrofuran.

2.1. Biological data

Collection of high-quality and diverse data is commonly recommended for establishing a QSAR model. Biological and chemical data of 38 N-aryl-oxazolidinone-5-carboxamides derivatives from the work of Yeung *et al.* (2013) [34] (Table 1) were selected. All the 38 compounds in this study were synthesized by same method with the same scaffold and their HIV-1 protease inhibitory activity was determined by same method. This series of compounds was found to possess high structural diversity and a sufficient range of biological activity. This formed the rationale for selecting the 38 compounds for our present study. The HIV-1 protease inhibitory activities used in the present study were expressed as $pKi = -\log_{10}Ki$, where Ki is the micro-molar concentration of the compounds producing 50% reduction in HIV-1 protease activity stated as the mean of at least two experiments.

2.2. Sketching of molecules

The 2D structures of the compounds in the respective series were drawn in modelling software CS Chem Office 2004 using its drawing tools. The structures were then checked for errors, cleaned up and saved as .mol files, which were further transferred to VLife MDS 4.3 software, where the structures were converted from two-dimensional form to three-dimensional form [19].

2.3. Energy minimization

The geometry of the 3D structure was optimized to local minima by Merck Molecular Force Field (MMFF) by considering a distance-dependent dielectric constant of 1.0, convergence criterion or a root-mean-square (RMS) gradient at 0.001 kcal/mol Å and an iteration limit of 10,000. Most stable structure for each compound was generated and saved as .mol2 files for computing various physico-chemical and alignment-independent (AI) descriptors [19].

2.4. 2D QSAR analyses

2.4.1. Calculation of descriptor (independent variable)

Various physico-chemical and AI descriptors of energy-minimized molecules were determined using VLife MDS 4.3 software. Energy-minimized geometry was used for calculating various 2D descriptors (i.e., individual, chi, chiv, path count, chi chain, chiv chain, chain path count, cluster, path cluster, Kappa, element count, estate number, estate contribution, semi-empirical, hydrophilic–hydrophobic and polar surface area).

Various AI descriptors were also calculated. Independent descriptors were assigned the three most important attributes for the calculation of AI. The first attribute was 'T', characterizing the topology of molecule. The second attribute was the atom type and the third attribute was assigned to atoms involved in the formation of double or triple bonds. The software then developed a model with a total of 200 physicochemical descriptors and more than 700 AI descriptors.

The independent variables (i.e., 2D descriptors) were pre-processed by removing the invariable (constant column), which resulted in a total of 120 descriptors to be used for QSAR analysis. Variable exclusion was done for a constant variable or near constant variable at paired correlation [19]. The total number of descriptors involved in the study was found to be high for the given series of compounds; hence significant descriptors have been elaborated in results and discussion.

2.4.2. Training and test set selection

The training and test datasets were selected using the sphere exclusion (SE) [36] and random selection methods. The dissimilarity values (i.e., SE radius) used in the SE method were 2 and 2.5, and the most active compound in the dataset was selected as the starting point for building a sphere [37]. In the random selection method, ten trials (70, 75, 80, 85 and 90%) were run. The trials were based on structural diversity and a wide range of activity—the range of biological activity of the test-set molecules was similar to that of the training set. Thus, the test set chosen was a true representative of the training set [38]. The selection of training and test set was further justified by uni-column statistics calculated for each case of the study.

2.4.3. Feature selection and model development

Among several search algorithms, feature selection procedures based on stepwise (SW) forward–backward variable selection method [39], genetic algorithms (GA) [40] and simulated annealing (SA) [41] were found to be most popular for building QSAR models to explain the features more effectively.

To build QSAR equations, the cross-correlation limit was set at 0.7, the number of variables at 5 and the term selection criteria at q^2. *F* value was specified to evaluate the significance of the variable. The variance cut-off was set at 0, with auto-scaling where the number of random iteration was set at 100.

In the SW forward–backward variable selection algorithm, the model was repeatedly altered from the previous version by adding or