DE GRUYTER OPEN

ACTA FACULTATIS MEDICAE NAISSENSIS

UDC:547.587.51:616.98:578.828

DOI: 10.2478/afmnai-2014-0011

Scientific Journal of the Faculty of Medicine in Niš 2014;31(2):95-103

Original article ■

Monte Carlo Method Based QSAR Modeling of Coumarin Derivates as Potent HIV-1 Integrase Inhibitors and Molecular Docking Studies of Selected 4-phenyl Hydroxycoumarins

Jovana Veselinović[1], Aleksandar Veselinović[2], Andrey Toropov[3], Alla Toropova[3], Ivana Damnjanović[1], Goran Nikolić[2]

[1]*University of Niš, Faculty of Medicine, Department of Pharmacy, Serbia*

[2]*Univerity of Niš, Faculty of Medicine, Department of Chemistry, Serbia*

[3]*IRCCS-Instituto di Ricerche Farmacologiche Mario Negri, Milano, Italy*

SUMMARY

In search for new and promising coumarin compounds as HIV-1 integrase inhibitors, chemoinformatic methods like quantitative structure-activity relationships (QSAR) modeling and molecular docking have an important role since they can predict desired activity and propose molecule binding to enzyme.

The aim of this study was building of QSAR models for coumarin derivatives as HIV-1 integrase inhibitors with the application of Monte Carlo method. SMILES notation was used to represent the molecular structure and for defining optimal SMILES-based descriptors. Molecular docking into rigid enzyme active site with flexible molecule was performed.

Computational results indicated that this approach can satisfactorily predict the desired activity with very good statistical significance. For best built model statistical parameters were: a) 3' Processing activity: R^2=0.9980 and Q^2=0.9977 for training set and R^2=0.9788 for test set and b) Integration activity: R^2=0.9999 and Q^2=0.9998 for training set and R^2= 0.9213 for test set. Built QSAR models were applied to selected 4-phenyl hydroxycoumarins for calculating desired activity and for HIV-1 integrase inhibition estimation. Additionally, molecular docking study was performed to a newly identified pocket in the HIV-1 integrase enzyme structure for determination of selected 4-phenyl hydroxycoumarins binding mode.

Monte Carlo method proved to be an efficient approach to build up a robust model for estimating HIV-1 integrase inhibition of coumarin compounds. Based on QSAR and molecular docking studies, 4-phenyl hydroxycoumarins can be considered as promising model compounds for developing new HIV-1 integrase inhibitors.

Key words: **coumarins, HIV-1 integrase inhibition, QSAR, molecular docking**

Corresponding author:

Jovana Veselinović •

phone: +381 18 4570029•

e-mail: milosavljevic.jovana@hotmail.com •

INTRODUCTION

Acquired immunodeficiency syndrome (AIDS), reported in 1981 (1), is a fatal disorder resulting from a chronic persistent infection by the human retrovirus, human immunodeficiency virus (HIV) (2). Today, AIDS is considered as one of the most devastating diseases faced by mankind, with an estimation of 34 million people living with HIV worldwide at the end of 2010 according to Joint United Nations Programme on HIV/AIDS (3). Up to now, successful chemotherapy has not been developed. Currently, Reverse Transcriptase (RT) and Protease (PT) inhibitors are the main targets for the majority of available drugs for HIV treatment. However, toxicity and rapid development of resistance to these inhibitors are the main issues related to the current therapy (4). Therefore, the development of new anti-HIV agents with varied structure and mechanisms of action is of great importance. HIV-1 integraze (HIV-1IN) is a very attractive and unexplored target for developing of new anti-HIV drugs as it plays a vital role in replication cycle and it has no cellular counterpart (5-7).

Various compounds exhibit HIV-1IN inhibitory activity, including lignanolides (8), curcumins (9), aurintricarboxylic acids (10), dicaffeoyl quinic acids and analogues (11, 12), diaryl sulfones (13). Unfortunately, all of stated inhibitors have the 1,2-dihydroxy (catechol) moiety, separated by an appropriate linker, so all of them have significant cytotoxicity because of catechol moiety autoxidation to reactive quinone species (14, 15). To overcame this problem, a series of coumarin derivatives which do not contain catechol functionality but possess good HIV-1IN inhibition activity was synthesized (16).

The importance of quantitative structure-reactivity relationship (QSAR) studies in modern drug design is well established since QSAR can make the early prediction of activity-related characteristics of drug candidates and can eliminate molecules with undesired properties (17). The main goal of QSAR approach is to correlate the biological activity of a series of compounds with the calculated molecular properties in terms of descriptors (18). Thousands of molecular descriptors are used in QSAR studies for the purpose of encoding molecules chemical and structural features (19, 20) with great importance of topological descriptors calculated on the basis of molecular graphs (21). The simplified molecular input line entry system (SMILES) is an alternative to molecular graphs and it can be used for representation of molecular structures (22). Recent papers have reported the applicability of SMILES based descriptors in QSAR analysis with models built on the basis of Monte Carlo method (23-27). Several QSAR studies dealing with coumarin compounds as HIV-1IN inhibitors are reported (28-31).

The aim of this research is to build QSAR modes for coumarin derivates as HIV-1IN inhibitors with SMILES based optimal descriptors and application of Monte Carlo method by using CORAL software. Built QSAR models were applied to selected 4-phenyl hydroxycoumarins with good antioxidant properties (32) but with no literature data about their HIV-1IN inhibition activity. Further, docking study is performed to a newly identified pocket right behind catalytic core domain (CCD) helix 4 (33) in the HIV-1IN enzyme for determinating the possible binding mode of selected 4-phenyl hydroxycoumarins.

METHOD

Data. A dataset of 26 coumarin derivatives with determined HIV-1 integrase inhibition activity was selected for QSAR study (16). Figure 1 presents general structures of used coumarin compounds for QSAR modeling. As an endpoint for QSAR model building pIC_{50} for enzyme 3' processing and integration was used.

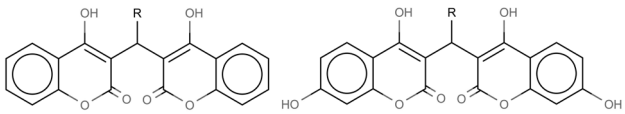


Figure 1. *General molecular structures of used molecules*

Canonical SMILES for all compounds were generated with the ACD/ChemSketch program (ACD/Chem Sketch v.11.0) in order to preserve consistency because different software may generate different SMILES notations. One random split into the training and test set was examined (20% of molecules are taken as test compounds). The role of the training set is in developing of the model. The role of test set is selection of preferable values for the number of epoch of the Monte Carlo optimization and the threshold value.

Optimal descriptors. SMILES is a representation of the molecular structure by sequence of symbols. Some symbols represent molecular fragments, such as atoms or bonds (e.g. 'C', 'N','=', '#', etc.). Some of these fragments are represented by two symbols (e.g. 'Br', 'Cl', '@@', etc.) which cannot be separated. Optimal SMILES-based descriptors, determined by descriptor correlation weight ($DCW(T,N_{epoch})$), were calculated with CORAL software (http://www.insilico.eu/coral) as:

$$DCW(T,N_{epoch}) = \alpha\Sigma CW(Sk) + \beta\Sigma CW(SSk) + \gamma\Sigma CW(SSSk) \quad (1)$$

where Sk, SSk, and SSSk are one-, two-, and three-component SMILES attributes, respectively; the component of SMILES attribute is SMILES symbol previously defined (27).

Two parameters in Eq. 1 should be defined for the Monte Carlo optimization: threshold (T) and the number of epochs (N_{epoch}). The classification of components of the representation of the molecular structure into two classes is done with the following criteria: rare and active which is defined with the T. The correlation weight of a rare component is fixed as zero, because this component brings noise to the model, so rare component is discarded from building up of the model and T is zero. The N_{epoch} is the number of epochs of the Monte Carlo

optimization (one epoch is the cycle of modifications of all correlation weights involved in the model). The predictive potentials of the model are mathematical functions T and N_{epoch} in the Monte Carlo optimization. The searches for the most predictive combination of T and N_{epoch} were concluded from values 0-7 for T and 0-70 for N_{epoch} for all models, according to previously published methodology (23-27).

Having numerical data on these correlation weights (CW) one can calculate DCW (T,N_{epoch}) for compounds of training and test set. Least squares method was used to calculate endpoint from theses data.

$$Endpoint=C_0+C_1\times DCW(T,N_{epoch}) \quad (2)$$

Molecular docking. 3D structures of the compounds for docking simulation were constructed using MarvinSketch 6.1.0, 2013, ChemAxon (http://www. chemaxon. com). Geometry optimization was carried out by employing MMFF94 molecular force field (34). To date, no full strength structure of HIV-1IN is available to elucidate the spatial arrangement of its three domains: N-terminal (NTD), catalytic core (CCD) and C-terminal (CTD). In the field of the development of allosterically targeted HIV-1IN inhibitors a new advantageous approach for the discovery of compounds effective against HIV-1IN strand -transfer drug-resistant viral strains has been proposed recently (35). A new site in integrase, a valid region for the structure-based design of allosteric integrase inhibitors, has been identified using a structure-based design process (protein data bank code: 3NF7) (33). The compounds were docked into enzyme binding sites using the MolegroVirtual Docker (MVD) (36). The Molegro Virtual Docker (MVD v. 2013.6.0.1.) software was employed for docking ligands to the rigid enzyme model for identification of hydrogen bonds and hydrophobic interactions between residues at the active site. The binding site was computed with a grid resolution of 0.3 Å. The MolDock SE as a search algorithm was used and the number of runs was set to 100. The parameters of docking procedure were: population size 50, maximum number of iterations 1500, energy threshold 100.00 and maximum number of steps 300. The number of generated poses was 10. The estimation of ligand–receptor interactions was described by the MVD-related scoring functions: MolDock Score, Rerank Score, Hbond Score, Similarity Score, and Docking Score. The ligand was docked into computed cavity instead ligand from 3NF7 using the MolDock Optimizer algorithm and its interactions were monitored using detailed energy estimates. A maximum population of 100 and maximum iterations of 10.000 were used for each run and the 5 best poses were retained.

RESULTS

The chemical structures represented with SMILES notation, the experimental activity for 3' Processing and Integration (expr) data, the calculated data (calc) with CORAL and difference (diff) between expr and calc are presented in Table 1. Statistical criteria of the predictability of the models are represented in Table 2 (37).

Using Eq.2 for predicting pIC_{50} following equations were calculated from best Monte Carlo runs:

3' Processing:

$$pIC_{50}=1.8584\ (\pm0.0036)+0.0187\ (\pm0.0000215)\times DCW(0,3) \quad (3)$$

Integration:

$$pIC_{50}=2.5636\ (\pm0.0004)+0.0180\ (\pm0.0000034)\times DCW(0,3) \quad (4)$$

Built QSAR models were applied for predicting pIC_{50} values of selected 4-phenyl hydroxycoumarins (7-hydroxy-4-phenyl coumarin (7C), 5,7-dihydroxy-4-phenyl coumarin (5,7C) and 7,8-dihydroxy-4-phenyl coumarin (7, 8C)) with good antioxidant properties (32) but with no literature data about their HIV-1IN inhibition activity. Eq. 3 and Eq. 4 were applied for calculation of pIC_{50} for enzyme 3' Processing and Integration for selected 4-phenyl hydroxycoumarins. Calculated values and molecular structures of used coumarin derivates are presented in Table 3.

Monte Carlo method can be used for classification of molecular features (SAk) calculated with SMILES notation based descriptors. The list of the SAk together with correlation weights for the three probes of the Monte Carlo optimization for all enzyme activities is given in the Table 4.

In order to gain insight into the plausible mechanism for 3' Processing and Integration actions docking simulations were performed for 7C; 5,7C and 7,8C. Figure 2 presents the best docking poses for all investigated coumarins inside enzyme binging pocket.

Two dimensional representation of the best docking poses for all investigated coumarins inside enzyme binding pocket are shown in Figure 3 (38).

Table 1. *Structures of 26 examined coumarin derivates as a HIV-1 integrase inhibitors represented with SMILES notations, calculated values for DCW, the experimental activity data (pIC_{50}) - expr (16), calculated values for pIC_{50} with application of CORAL - calc and difference (diff) between expr and calc*

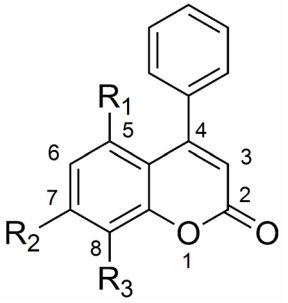
			3' PROCESSING				INTEGRATION			
	SMILES NOTATION	Set	DCW	Expr	Calc	Diff	DCW	Expr	Calc	Diff
1	OC=1c5ccccc5OC(=O)C=1C(C2=C(O)c3ccccc3OC2=O)c4ccccc4	Train	133.901	4.367	4.362	0.005	130.045	4.411	4.423	-0.012
2	Oc1ccc(cc1)C(C2=C(O)c3ccccc3OC2=O)C4=C(O)c5ccccc5OC4=O	Test	112.152	3.893	3.956	-0.063	117.992	4.131	4.191	-0.06
3	Oc1ccc(cc1OC)C(C2=C(O)c3ccccc3OC2=O)C4=C(O)c5ccccc5OC4=O	Train	101.154	3.752	3.75	0.002	114.24	4.119	4.119	0
4	CN(C)c1ccc(cc1)C(C2=C(O)c3ccccc3OC2=O)C4=C(O)c5ccccc5OC4=O	Train	116.902	4.055	4.044	0.011	123.728	4.301	4.301	0
5	[O-][N+](=O)c1ccc(cc1)C(C2=C(O)c3ccccc3OC2=O)C4=C(O)c5ccccc5OC4=O	Train	130.123	4.301	4.292	0.009	156.225	4.921	4.925	-0.004
6	O=C(O)c1ccc(cc1)C(C2=C(O)c3ccccc3OC2=O)C4=C(O)c5ccccc5OC4=O	Train	131.395	4.319	4.315	0.004	124.463	4.31	4.315	-0.005
7	OC=4c5ccccc5OC(=O)C=4C(C=1C(=O)Oc2ccccc2C=1O)c3cccs3	Train	114.362	4	3.997	0.003	99.275	3.83	3.832	-0.002
8	OC=1c5ccccc5OC(=O)C=1C(C2=C(O)c3ccccc3OC2=O)c4ccco4	Test	129.901	4.468	4.288	0.18	126.551	3.951	4.355	-0.404
9	OC=1c6ccccc6OC(=O)C=1C(C2=C(O)c3ccccc3OC2=O)c4cc5ccccc5nc4	Train	142.612	4.538	4.525	0.013	152.307	4.854	4.85	0.004
10	OC=1c6ccccc6OC(=O)C=1C(C2=C(O)c3ccccc3OC2=O)c4cc5ccccc5cc4	Train	155.618	4.721	4.768	-0.047	163.071	5.046	5.057	-0.011
11	OC=1c6ccccc6OC(=O)C=1C(C2=C(O)c3ccccc3OC2=O)c4ccc(cc4)c5ccccc5	Train	160.847	4.854	4.866	-0.012	154.801	4.886	4.898	-0.012
12	OC=1c7ccccc7OC(=O)C=1C(C2=C(O)c3ccccc3OC2=O)c4ccc5c6ccccc6Cc5c4	Train	167.13	5	4.984	0.016	165.807	5.108	5.109	-0.001
13	OC=1c6ccccc6OC(=O)C=1C(C2=C(O)c3ccccc3OC2=O)c5ccc(/C=C/c4ccccc4)cc5	Train	180.856	5.26	5.24	0.02	182.81	5.432	5.436	-0.004
14	OC=1c6ccccc6OC(=O)C=1C(C2=C(O)c3ccccc3OC2=O)c5ccc(OCc4ccccc4)cc5	Test	154.346	5.071	4.745	0.326	157.301	4.854	4.946	-0.092
15	OC=1c7ccccc7OC(=O)C=1C(C2=C(O)c3ccccc3OC2=O)c6cc(OCc4ccccc4)cc(OCc5ccccc5)c6	Train	172.077	5.097	5.076	0.021	183.03	5.432	5.44	-0.008
16	OC=1c9ccccc9OC(=O)C=1C(C2=C(O)c3ccccc3OC2=O)c4ccc(cc4)C(C5=C(O)c6ccccc6OC5=O)C7=C(O)c8ccccc8OC7=O	Train	204.771	5.699	5.688	0.011	217.421	6.097	6.1	-0.003
17	Oc3ccc4C(O)=C(CC1=C(O)c2ccc(O)cc2OC1=O)C(=O)Oc4c3	Train	131.927	4.334	4.325	0.009	126.441	4.348	4.353	-0.005
18	Oc1ccc2C(O)=C(C(=O)Oc2c1)C(C3=C(O)c4ccc(O)cc4OC3=O)c5ccccc5	Train	150.123	4.764	4.666	0.098	141.188	4.654	4.637	0.017
19	Oc1ccc2C(O)=C(C(=O)Oc2c1)C(C3=C(O)c4ccc(O)cc4OC3=O)c5ccc(cc5)C(C6=C(O)c7ccc(O)cc7OC6=O)C8=C(O)c9ccc(O)cc9OC8=O	Train	244.311	6.432	6.427	0.005	237.543	6.481	6.487	-0.006
20	Oc1ccc2C(O)=C(C(=O)Oc2c1)C(C3=C(O)c4ccc(O)cc4OC3=O)c5ccccn5	Train	122.112	4.076	4.142	-0.066	124.932	4.31	4.324	-0.014
21	Oc1ccc2C(O)=C(C(=O)Oc2c1)C(C3=C(O)c4ccc(O)cc4OC3=O)c5cccnc5	Train	123.609	4.208	4.17	0.038	139.441	4.602	4.603	-0.001
22	Oc1ccc2C(O)=C(C(=O)Oc2c1)C(C3=C(O)c4ccc(O)cc4OC3=O)c5cc6ccccc6cc5	Test	161.091	5.377	4.871	0.506	171.71	5.45	5.223	0.227
23	Oc1ccc2C(O)=C(C(=O)Oc2c1)C(C3=C(O)c4ccc(O)cc4OC3=O)c6ccc(/C=C/c5ccccc5)cc6	Train	177.588	5.155	5.179	-0.024	199.447	5.745	5.755	-0.01
24	Oc5ccc6C(O)=C(C2c4ccccc4OC=1c3ccc(O)cc3OC(=O)C=12)C(=O)Oc6c5	Train	109.849	3.917	3.913	0.004	103.785	3.914	3.918	-0.004
25	Oc6ccc7C(O)=C(C2C=5C(=O)Oc1ccccc1C=5OC4=C2C(=O)Oc3cc(O)ccc34)C(=O)Oc7c6	Test	124.82	4.447	4.193	0.254	140.916	4.648	4.631	0.017
26	Oc1ccc2c(c1)OC(=O)C=C2O	Train	88.717	3.523	3.517	0.006	81.441	3.488	3.489	-0.001

Table 2. *Statistical quality of built QSAR models*

	3' PROCESSING							INTEGRATION						
	Training			Test				Training			Test			
	R^2	Q^2	s	R^2	$r_{m(av)}^2$	Δr_m^2	s	R^2	Q^2	s	R^2	$r_{m(av)}^2$	Δr_m^2	s
1	0.9993	0.9992	0.020	0.9671	0.8083	0.0577	0.305	0.9999	0.9998	0.010	0.9185	0.5230	0.1825	0.248
2	0.9977	0.9974	0.033	0.9368	0.6138	0.1677	0.265	0.9999	0.9998	0.005	0.9213	0.5678	0.2041	0.268
3	0.9980	0.9977	0.032	0.9788	0.6213	0.1467	0.341	0.9999	0.9998	0.008	0.9186	0.5798	0.1965	0.239
Av	0.9983	0.9981	0.028	0.9609	0.6811	0.1240	0.304	0.9999	0.9998	0.008	0.9195	0.5569	0.1945	0.252

Av is average value from three independant Monte Carlo runs (1, 2 and 3)
R^2 is correlation coefficient
Q^2 is cross-validated correlation coefficient
s is standard error of estimation
$r_{m(av)}^2$ should be > 0.5 (37)
Δr_m^2 should be < 0.2 (37)

Table 3. *Molecular structures of used coumarin derivates with calculated pIC_{50} values for enzyme 3' Processing and Integration activities using Eq. 3 and 4.*



Molecule	R1	R2	R3	pIC_{50} (3'Processing)	pIC_{50} (Integration)
7C	H	OH	H	3.329	3.640
5,7C	OH	OH	H	3.138	3.451
7,8C	H	OH	OH	3.259	3.402

Table 4. *The list of the SAk with correlation weights for three independent Monte Carlo optimization runs for best QSAR model*

	3' PROCESSING				INTEGRATION			
	SAk	**Run 1**	**Run 2**	**Run 3**	**SAk**	**Run 1**	**Run 2**	**Run 3**
Decrease	n.....	-3.997	-3.248	-5.004	n...c...c...	-3.999	-3.002	-2.497
	s...c.....	-1.002	-1.502	-4.005	n.....	-2.998	-2.248	-2.247
	N...+.....	-0.745	-1.252	-0.004	O...C...=...	-1.253	-0.996	-1.245
	C...(...C...	-0.502	-1.249	-0.999	O...=...(...	-1	-0.998	-0.998
	N...(...C...	-0.496	-1.246	-2.251	O...=.....	-0.496	-0.997	-1.005
Increase	O...(...C...	0.252	1.003	1.25	c...O.....	0.245	0.753	1.001
	c.....	0.504	0.746	0.998	O...C.....	0.502	0.996	0.999
	o.....	0.997	0.996	0.997	C.../.....	0.997	0.504	0.004
	C...=.....	4.254	4.5	2.748	c...C.....	1.001	1.501	1
	C...C.....	4.5	3.496	3.497	C...(.....	1.505	0.747	0.998

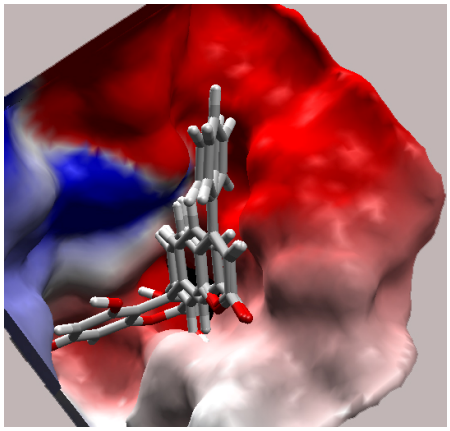


Figure 2. *Surface diagram showing docked selected 4-phenyl hydroxycoumarins*

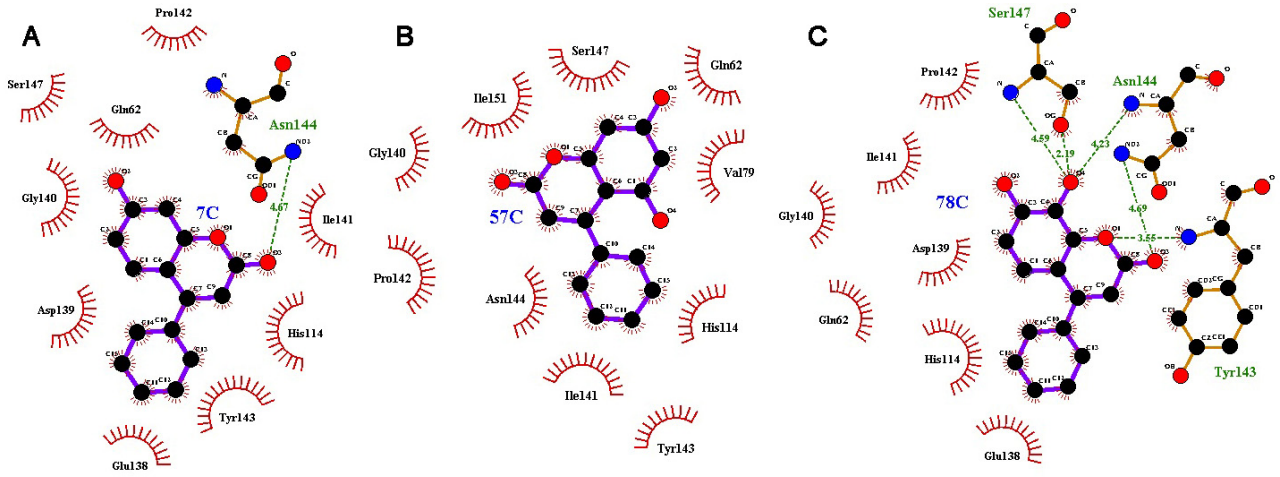


Figure 3. *Two dimensional representations of the best docking pose for a) 7-hydroxy-4-phenyl coumarin, b) 5,7-dihydroxy-4-phenyl coumarin and c) 7,8-dihidroxy-4-phenyl coumarin inside binding pocket*

DISCUSSION

QSAR study

Results from Table 2 show that the predictability for all models is good. Also, the results are satisfactory from the point of view of new criteria (37).

The correlation weights for molecular features calculated with SMILES can be used for classification of the aforementioned features according to their values from three probes for defined Monte Carlo model. They could be divided into three categories: features with stable positive values of correlation weights (promoters of increase of an endpoint); features with stable negative values of correlation weights (promoters of decrease of an endpoint); and unstable features which have positive values of correlation weights together with negative correlation weights values for several models (26, 27). For example, if the correlation weight of Sk CW(Sk) is >0 in all three runs of the optimization, then the Sk is promoter of Ac increase. However, if CW(Sk) is <0 in all three runs of the optimization, then the Sk is promoter of Ac decrease. In the end, if there are both CW(Sk) >0 and CW(Sk) <0, or Sk is blocked in three runs of optimization then Sk has an undefined role. Same rule is applied for all SAk. It must be noted that SAk have mechanistic interpretation and according to presented results from Table 4 SAk can be classified as following. For example, for 3' Processing 'n.....', 'N...+.....' and 'C...(...C...' are promoters of decrease while 'C...C.....', 'C...=.....' and 'o.....' are promoters of increase. 'n.....' can be interpreted as aromatic nitrogen atom, 'N...+.....' as sp^3 nitrogen atom with positive charge and 'C...(...C...' as sp^3 carbon atoms with branching. 'C...C.....' can be interpreted as two sp^3 carbon atoms without branching, 'C...=.....' as sp^2 carbon atom since '=' is a symbol for double bond, 'o.....' as aromatic oxygen atom.

Molecular docking

The results of molecular docking studies are presented in Figure 2 and 3. On the surface diagram (Figure 2) it can be seen that hydrophobic parts of the molecules are oriented toward the hydrophobic parts of the enzyme binding pocket (red colored surface). Two dimensional representations of the best docking pose for selected coumarins (Figure 3) give a more detailed insight into the interactions with particular amino acids in enzyme binding pocket. Based on the presented results, it can be concluded that hydrophobic interactions between investigated coumarins and binding pocket play an important role.

However, number, bond length and bond energy of hydrogen bonds formed between ligand and enzyme has an important role in ligand effect on investigated activity. It was observed from *in silico* studies of compounds binding to 3NF7 that 7C oxygen from carbonyl group forms hydrogen bond with Asn-144 (bond length 4.67 Å). 5,7C does not form any hydrogen bonds with enzyme. Compound 7,8C hydroxyl group in position 8 forms three hydrogen bonds, two with Ser-147 (2.19 Å and 4.59 Å) and one with Asn-144 (4.23 Å). Oxygen from carbonyl group forms one with Asn-144 (4.69 Å) and sp^3 oxygen one with Tyr-143 (3.55 Å).

CONCLUSION

QSAR models for coumarin compounds as potent HIV-1 integrase inhibitors were built. Monte Carlo method proved to be an efficient tool to build up a robust model for estimating HIV-1 integrase inhibition. For suggested modeling process optimal descriptors were based on SMILES notation. The predictive potential of the applied approach was tested with one split into the training and test set. The robustness of model was confirmed with different methods. The SMILES attributes which are promoters of increase/decrease of HIV-1 integrase inhibition were identified. Built QSAR models were applied to selected 4-phenyl coumarins for inhibition prediction. Further, the correlation between calculated inhibitory activity and the *in silico* molecular docking scores of these compounds was obtained through hydrogen bonding interactions. Our results suggest that 4-phenyl hydroxycoumarins may be considered as good molecular templates for potential HIV-1 integrase inhibitors.

Acknowledgment

This work has been financially supported by Ministry of Education and Science, Republic of Serbia, under Project Numbers OI 172044 and TR 31060.

References

1. Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, Wolf RA, Saxon A. Pneumocystis carinii Pneumonia and Mucosal Candidiasis in Previously Healthy Homosexual Men - Evidence of a New Acquired Cellular Immunodeficiency. N Engl J Med 1981; 305(24): 1425–31. http://dx.doi.org/10.1056/NEJM198112103052401
2. Koup RA, Merluzzi VJ, Hargrave KD, Adems J, Grozinger KJ. Inhibition of human immunodeficiency virus type 1 replication by the dipyridocliazepinone. J Infect Dis 1991; 163(5):966–70. http://dx.doi.org/10.1093/infdis/163.5.966
3. UNAIDS, 2011. World AIDS Day Report 2011

4. Erickson JW, Burt SK. Structural mechanisms of HIV drug resistance. Annu Rev Pharmacol Toxicol 1996; 36:545–71. http://dx.doi.org/10.1146/annurev.pa.36.040196.002553
5. Goldgur Y, Craigie R, Cohen GH, et al. Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: A platform for antiviral drug design. Proc Natl Acad Sci 1999; 96(23):13040–3. http://dx.doi.org/10.1073/pnas.96.23.13040
6. Bhatt H, Patel P, Pannecouque C. Discovery of HIV-1 integrase inhibitors: pharmacophore mapping, virtual screening, molecular docking, synthesis, and biological evaluation. Chem Biol Drug Des 2014; 83(2):154–66. http://dx.doi.org/10.1111/cbdd.12207
7. de Carvalho LL, Maltarollo VG, de Lima EF, Weber KC, Honorio KM, da Silva AB. Molecular features related to HIV integrase inhibition obtained from structure- and ligand-based approaches. PLoS One. 2014; 9(1): e81 301. http://dx.doi.org/10.1371/journal.pone.0081301
8. Eich E, Pertz H, Kaloga M, et al. (−)-Arctigenin as a Lead Structure for Inhibitors of Human Immunodeficiency Virus Type-1 Integrase. J Med Chem 1996; 39 (1):86-95. http://dx.doi.org/10.1021/jm950387u
9. Mazumder A, Raghavan K, Weinstein J, Kohn KW, Pommier Y. Inhibition of human immunodeficiency virus type-1 integrase by curcumin. Biochem Pharmacol 1995; 49(8):1165-70. http://dx.doi.org/10.1016/0006-2952(95)98514-A
10. Cushman M, Sherman P. Inhibition of HIV-1 integration protein by aurintricarboxylic acid monomers, monomer analogs, and polymer fractions. Biochem Biophys Res Commun 1992; 185(1):85–90. http://dx.doi.org/10.1016/S0006-291X(05)80958-1
11. Robinson WE, Cordeiro M, Abdel-Malek S, et al. Dicaffeoylquinic acid inhibitors of human immunodeficiency virus integrase: inhibition of the core catalytic domain of human immunodeficiency virus integrase. Mol Pharmacol 1996; 50(4):846-855.
12. Robinson WE, Reinecke MG, Abdel-Malek S, Jia Q, Chow SA. Inhibitors of HIV-1 replication [corrected; erratum to be published] that inhibit HIV integrase. Proc Natl Acad Sci 1996; 93(13):6326-31. http://dx.doi.org/10.1073/pnas.93.13.6326
13. Mazumder A, Neamati N, Sommadossi JP, et al. Effects of nucleotide analogues on human immunodeficiency virus type 1 integrase. Mol Pharmacol 1996; 49 (4):621-8.
14. Galati G, Sabzevari O, Wilson JX, O'Brien PJ. Prooxidant activity and cellular effects of the phenoxyl radicals of dietary flavonoids and other polyphenolics. Toxicology 2002; 177(1):91-104. http://dx.doi.org/10.1016/S0300-483X(02)00198-1
15. Cavalieri EL, Li KM, Balu N, et al. Catechol orthoquinones: The electrophilic compounds that form depurinating DNA adducts and could initiate cancer and other diseases. Carcinogenesis 2002; 23(6):1071–7. http://dx.doi.org/10.1093/carcin/23.6.1071
16. Zhao H, Neamati N, Hong H, et al. Coumarin-based inhibitors of HIV integrase. J Med Chem 1997; 40(2): 242-9. http://dx.doi.org/10.1021/jm960450v
17. Hansch C, Hoekman D, Gao H. Comparative QSAR: toward a deeper understanding of chemicobiological interactions. Chem Rev 1996; 96(3):1045-76. http://dx.doi.org/10.1021/cr9400976
18. Tropsha A, Gramatica P, Gombar VK. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Comb Sci 2003; 22(1):69-77. http://dx.doi.org/10.1002/qsar.200390007
19. Karelson M. Molecular Descriptors in QSAR/QSPR. Wiley-Interscience, New York, 2000.
20. Todeschini R, Consonni V. Handbook of Molecular Descriptors, Wiley-VCH, Weinheim, Germany, 2000. http://dx.doi.org/10.1002/9783527613106
21. Duchowicz PR, Talevi A, Bruno-Blanch LE, Castro EA. New QSPR study for the prediction of aqueous solubility of drug-like compounds. Bioorg Med Chem 2008; 16(17):7944-55. http://dx.doi.org/10.1016/j.bmc.2008.07.067
22. Daylight Chemical Information Systems, Inc., 2008. http://www.daylight.com (accessed 11.02.14.).
23. Toropov AA, Toropova AP, Raska I Jr. QSPR modeling of octanol/water partition coefficient for vitamins by optimal descriptors calculated with SMILES. Eur J Med Chem 2008; 43(4):714-40. http://dx.doi.org/10.1016/j.ejmech.2007.05.007
24. Toropov AA, Benfenati E. Additive SMILES-based optimal descriptors in QSAR modelling bee toxicity: using rare SMILES attributes to define the applicability domain. Bioorg Med Chem 2008; 16(9) 4801-9. http://dx.doi.org/10.1016/j.bmc.2008.03.048
25. Toropov AA, Toropova AP, Lombardo A, Roncaglioni A, Benfenati E, Gini G. CORAL: building up the model for bioconcentration factor and defining it's applicability domain. Eur J Med Chem 2011; 46(4) 1400-3. http://dx.doi.org/10.1016/j.ejmech.2011.01.018
26. Veselinović AM, Milosavljević JB, Toropov AA, Nikolić GM. SMILES-based QSAR models for arylpiperazines as high-affinity 5-HT1A receptor ligands using CORAL. Eur J Pharm Sci 2013; 48(3) 532-41. http://dx.doi.org/10.1016/j.ejps.2012.12.021
27. Veselinović AM, Milosavljević JB, Toropov AA, Nikolić GM. SMILES based QSAR models for the calcium channel antagonistic effect of 1,4-dihydropyridines. Arch Pharm 2013; 346(2):134-39. http://dx.doi.org/10.1002/ardp.201200373
28. Yuan H, Parrill AL. QSAR studies of HIV-1 integrase inhibition. Bioorg Med Chem 2002; 10(12):4169-83. http://dx.doi.org/10.1016/S0968-0896(02)00332-2
29. Cheng Z, Zhang Y, Fu W. QSAR study of carboxylic acid derivatives as HIV-1 Integrase inhibitors. Eur J Med Chem 2010; 45(9):3970–80. http://dx.doi.org/10.1016/j.ejmech.2010.05.052
30. Kaushik S, Gupta SP, Sharma PK, Anwar Z. A QSAR study on some series of HIV-1 integrase inhibitors. Med Chem 2011; 7(6):553-60. http://dx.doi.org/10.2174/157340611797928361

31. Srivastav VK, Tiwari M. QSAR and docking studies of coumarin derivatives as potent HIV-1 integrase inhibitors. Arabian J Chem 2013; article in press; http://dx.doi.org/10.1016/j.arabjc.2013.01.015 http://dx.doi.org/10.1016/j.arabjc.2013.01.015
32. Veselinović JB, Veselinović AM, Vitnik ZJ, Vitnik VD, Nikolić GM. Antioxidant properties of selected 4-phenyl hydroxycoumarins: Integrated in vitro and computational studies. Chem Biol Interact 2014; 214:49-56. http://dx.doi.org/10.1016/j.cbi.2014.02.010
33. Rhodes DI, Peat TS, Vandegraaff N, et al. Structural basis for a new mechanism of inhibition of HIV-1 integrase identified by fragment screening and structure-based design. Antivir Chem Chemother 2011; 21(4): 155-68. http://dx.doi.org/10.3851/IMP1716
34. Halgren TA. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MM-FF94. J Comput Chem 1996; 17(5-6):490-519.
35. Al-Mawsawi LQ, Neamati N. Allosteric inhibitor development targeting HIV-1 integrase. Chem Med Chem 2011;6(2):228-41. http://dx.doi.org/10.1002/cmdc.201000443
36. Thomsen R, Christensen MH. MolDock: a new technique for high-accuracy molecular docking. J Med Chem 2006; 49(11):3315-21. http://dx.doi.org/10.1021/jm051197e
37. Roy PP, Roy K. QSAR studies of CYP2D6 inhibitor aryloxypropanolamines using 2D and 3D descriptors. Chem Biol Drug Des 2009; 73(4):442-55. http://dx.doi.org/10.1111/j.1747-0285.2009.00791.x
38. Laskowski RA, Swindells MB. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J Chem Inf Model 2011; 51(10):2778-86. http://dx.doi.org/10.1021/ci200227u

QSAR MODELOVANJE DERIVATA KUMARINA KAO POTENTNIH INHIBITORA HIV-1 INTEGRAZE BAZIRANO NA MONTE KARLO METODI I STUDIJE MOLEKULARNOG DOKINGA ODABRANIH 4-FENIL HIDROKSIKUMARINA

Jovana Veselinović[1], Aleksandar Veselinović[2], Andrey Toropov[3], Alla Toropova[3], Ivana Damnjanović[1], Goran Nikolić[2]

[1]Univerzitet u Nišu, Medicinski fakultet, Katedra za farmaciju, Srbija
[2]Univerzitet u Nišu, Medicinski fakultet, Katedra za hemiju, Srbija
[3]IRCCS- Instituto di Ricerche Farmacologiche Mario Negri, Milano, Italija

Sažetak

U potrazi za novim i obećavajućim kumarinskim jedinjenjima koja su inhibitori HIV-1 integraze veoma značajnu ulogu imaju hemoinformatičke metode, kao što je modelovanje kvantitativnog odnosa strukture i aktivnosti (QSAR) i molekularni doking, jer mogu da predvide željenu aktivnost i objasne način vezivanja molekula za enzim.

Cilj ovog rada bio je uspostavljanje QSAR modela za derivate kumarina kao inhibitore HIV-1 integraze korišćenjem Monte Karlo metode. SMILES notacija je korišćenja kao reprezentacija molekulske strukture i za definisanje optimalnih deskriptora baziranih na SMILES notaciji. Primenjen je i molekularni doking u kruto aktivno mesto enzima sa fleksibilnim molekulom.

Rezultati kompjuterske studije ukazuju da ovaj pristup može na zadovoljavajući način da predvidi željenu aktivnost sa veoma dobrom statističkom značajnošću. Za najbolji model statistički parametri su bili: a) "3' Processing" aktivnost: R^2=0.9980 i Q^2=0.9977 za skup molekula koji su korišćeni za konstruisanje modela (trening set) i R^2=0.9788 za skup molekula koji su korišćeni za proveru modela (test set) i b) "Integration" aktivnost: R^2=0.9999 i Q^2=0.9998 za trening set i R^2=0.9213 za test set. Uspostavljeni QSAR modeli su primenjeni na odabrane 4-fenil hidroksikumarine radi izračunavanja željene aktivnosti u cilju procene inhibicije HIV-1 integraze. Dodatno je urađena doking studija u novom identifikovanom džepu unutar strukture enzima HIV-1 integraze radi određivanja načina vezivanja ispitivanih 4-fenil hidroksikumarina.

Monte Karlo metoda se pokazala kao efikasan pristup u uspostavljanju robusnog modela za procenu inhibicije HIV-1 integraze od strane kumarinskih jedinjenja. Na osnovu rezultata QSAR i doking studija, 4-fenil kumarini se mogu okarakterisati kao dobra model jedinjenja za razvijanje novih inhibitora HIV-1 integraze.

Ključne reči: **kumarini, Inhibitori HIV-1 integraze, QSAR, molekularni doking**