



The index of ideality of correlation and the variety of molecular rings as a base to improve model of HIV-1 protease inhibitors activity

Andrey A. Toropov¹ · Alla P. Toropova¹ · Edoardo Carnesecchi^{1,2} · Emilio Benfenati¹ · Jean Lou Dorne³

Received: 11 February 2020 / Accepted: 19 March 2020 / Published online: 4 April 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Computational prediction of HIV-1 protease inhibitor via quantitative structure–activity relationships (QSARs) is a popular study in the field of computational chemistry. The aim of the present study was building up of QSAR models for anti-HIV activity by means of the CORAL software (<http://www.insilico.eu/coral>). Applying of correlation weights for five and six-member molecular rings as components of the target function in the Monte Carlo optimization that is aimed to build up correlation between activity of HIV-1 protease inhibitors expressed as $pIC_{50} = \lg[1/(IC_{50} \times 109)]$ and optimal descriptor improves the predictive potential. Simplified molecular input-line entry system (SMILES) is used as the representation of the molecular structure of HIV-1 protease inhibitors for the QSAR analysis. New criterion of predictive potential so-called index of ideality of correlation (*IIC*) has been applied to improve the model for HIV-1 protease inhibitors activity. High correlation coefficients were observed between the experimental and predicted anti-HIV activity. Applying of special ring code as component of the Monte Carlo calculation significantly improves the statistical quality of the model. Furthermore, applying of the *IIC* as component of the Monte Carlo optimization improves the predictive potential of the CORAL models. The presence of the rings and the quality of these are very important molecular features which are able to improve statistical quality of model for anti-HIV activity. In addition, the ability of *IIC* to improve predictive potential of a model has been confirmed.

Keywords QSAR · Anti-HIV-1 · Monte Carlo method · CORAL software

Introduction

Nowadays, acquired immunodeficiency syndrome (AIDS) is regarded as one of the most devastating diseases [1–6]. Human immunodeficiency virus (HIV) is the causative agent for AIDS, and millions of people are suffering from AIDS all over the world. The AIDS-related mortality is increasing, but AIDS has

gradually become a controllable, chronic disease [7]. In this regard, HIV protease inhibitor is one of the most important components in the combination therapy. Therefore, the HIV protease is a prime target for drug design of anti-HIV treatment. Generally, drug design methodologies involve quantitative structure–activity relationships (QSARs) in order to understand the relationship between chemical structures and their biological activities as well as to predict the activities of any unknown compounds. In the case of drug discovery aimed to detect effective anti-HIV1 agents, QSARs models allow identifying the most promising molecules for further synthesis and clinical testing as shown in recent studies [1, 8]. Important components of the QSAR analysis are (i) definition of the domain of applicability, and (ii) mechanistic interpretation of QSAR models [9].

There are groups of conceptually different approach to build up QSAR related to anti-HIV activity: (i) multiple linear regression [8, 10]; (ii) nonlinear iterative partial least squares method [11]; (iii) artificial neural networks [8, 12–14]; (iv) 3D-QSAR [15]; (v) the support vector machine [8, 10, 16–18]; and (vi) Monte Carlo technique [19–22]. The above researches have quite different logistic, e.g., interpretation of anti-HIV activity

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11224-020-01525-9>) contains supplementary material, which is available to authorized users.

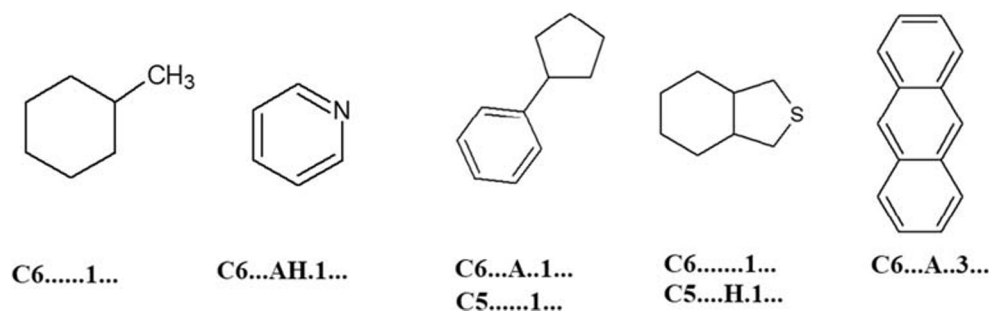
✉ Alla P. Toropova
alla.toropova@marionegri.it

¹ Laboratory of Environmental Chemistry and Toxicology, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via Mario Negri 2, 20156 Milan, Italy

² Institute for Risk Assessment Sciences (IRAS), Utrecht University, 3584 Utrecht, Netherlands

³ Scientific Committee and Emerging Risks Unit, European Food Safety Authority, Via Carlo Magno 1A, 43126 Parma, Italy

Fig. 1 Generalized representation of codes for various molecular five- and six-member rings



via numerical data on physicochemical data (molar volume, dipole moment) [8]; support vector machine and artificial neuron networks based on the molecular docking [10]; chemogeometric analysis for extraction of the most reliable chemical elements in corresponding 3D positions [11]; search for effective pyridinone derivatives [12]; group center overlapping [15], drug-drug interaction and multiple QSAR [16]. In other words, each approach based on defined hypothesis: 3D geometry; and representability (sufficiency) of physicochemical and/or biochemical data (thermodynamics, drug-drug interactions, etc.).

Here the Monte Carlo technique [23–25] applied with two new component of building up models for anti-HIV-1 activity. These are (i) special coefficients (correlation weights) which are used to the assessment of influence of molecular rings to anti-HIV-1 activity; factually, it is a way to improve the transparency of mechanistic interpretation of corresponding models [23–27]; and (ii) the target function of the Monte Carlo optimization involves a new component, so-called index of ideality of correlation (*IIC*) [28–31].

Thus, the aim of the present study is to check up two hypotheses which are the following:

- Molecular five- and six-member rings have influence to anti-HIV-1 activity;
- The *IIC* improves the predictive potential of models for anti-HIV-1 activity.

Method

Data

Numerical data on the activity of HIV-1 protease inhibitors expressed as $\text{pIC}_{50} = \lg[1/(\text{IC}_{50} \times 10^9)]$ were taken from the literature [1]. The total number of compounds examined here is 450. It should be noted that the CORAL software [5, 6, 23] gives possibility to build up a model with structured training set that involves three subsets which are the following: the active training, invisible passive training, and calibration sets. The fourth set is the external validation set. Each subset has a special task. The task for the training set is to calculate correlation weights which give as large as possible correlation between experimental and

predicted endpoint for the training set. The task for the invisible training set is inspection: whether these data give reasonable correlation coefficient for the similar compounds. The task of the calibration set is to detect the overtraining. The task for the validation set is the final estimation of the predictive potential of the model. Thus, the above-mentioned compounds ($n = 450$) were randomly distributed into the training ($\approx 25\%$), invisible training ($\approx 25\%$), calibration ($\approx 25\%$), and validation ($\approx 25\%$) sets. Table S1 in *Supplementary materials* confirms that these three splits are far from be identical ones.

Optimal descriptor

Two kinds of optimal descriptors used here to build up QSAR models for anti-HIV-1 activity

$$DCW(T^*, N^*) = \sum_{k=1}^{NA} CW(S_k) + \sum_{k=1}^{NA-1} CW(SS_k) + \sum_{k=1}^{NA-2} CW(SSS_k) \quad (1)$$

$$DCW(T^*, N^*) = CW(C5) + CW(C6) + \sum_{k=1}^{NA} CW(S_k) + \sum_{k=1}^{NA-1} CW(SS_k) + \sum_{k=1}^{NA-2} CW(SSS_k) \quad (2)$$

The C5 and C6 are SMILES attributes related to different kinds of five and six-member molecular rings. The S_k is the “SMILES-atom”, i.e., one symbol (e.g., “C,” “N,” “O,” etc.) or two symbols which cannot be examined separately (e.g., “Cl,” “Br,” “Si,” etc.); the SS_k is a combination of two SMILES atoms; the SSS_k is a combination of three SMILES atoms; the $CW(C5)$, $CW(C6)$, $CW(S_k)$, $CW(SS_k)$, and $CW(SSS_k)$ are so-called correlation weights of the above-mentioned attributes of SMILES.

Figure 1 demonstrated generalized conception of the codes for five- and six-member rings.

The numerical data on the $CW(S_k)$, $CW(SS_k)$, and $CW(SSS_k)$ are calculated with the Monte Carlo method, i.e., the optimization procedure which gives maximal value of a target function (*TF*). The *NA* is the number of attributes in SMILES. Having the numerical data on the corresponding correlation

Table 1 The statistical characteristics of the models for anti-HIV-1 activity for training (TRN), invisible training (iTRN), calibration (CLB), and validation (VLD) sets

Eq.	Split	Run	Set*	n	R ²	s	MAE
1	1	1	TRN	111	0.7661	0.810	0.646
			iTRN	114	0.7949	0.920	0.744
			CLB	114	0.7264	0.728	0.558
			VLD**	111	0.8154	0.687	0.547
		2	TRN	111	0.7439	0.847	0.673
			iTRN	114	0.7822	0.958	0.783
			CLB	114	0.7273	0.724	0.580
			VLD	111	0.8281	0.674	0.527
		3	TRN	111	0.7705	0.802	0.633
			iTRN	114	0.8120	0.888	0.702
			CLB	114	0.7303	0.733	0.580
			VLD	111	0.8319	0.669	0.530
1	2	1	TRN	112	0.7697	0.863	0.693
			iTRN	112	0.7693	0.853	0.725
			CLB	113	0.7874	0.760	0.635
			VLD	113	0.7987	0.710	0.572
		2	TRN	112	0.7937	0.817	0.645
			iTRN	112	0.8032	0.779	0.637
			CLB	113	0.7543	0.818	0.631
			VLD	113	0.7855	0.714	0.561
		3	TRN	112	0.7915	0.821	0.626
			iTRN	112	0.7948	0.804	0.626
			CLB	113	0.7696	0.801	0.628
			VLD	113	0.7876	0.706	0.545
1	3	1	TRN	114	0.7179	0.957	0.763
			iTRN	110	0.7107	0.978	0.781
			CLB	113	0.7756	0.702	0.522
			VLD	113	0.8031	0.722	0.567
		2	TRN	114	0.8083	0.789	0.614
			iTRN	110	0.7755	0.856	0.649
			CLB	113	0.7364	0.789	0.636
			VLD	113	0.7263	0.855	0.659
		3	TRN	114	0.7732	0.858	0.683
			iTRN	110	0.7608	0.886	0.706
			CLB	113	0.7657	0.729	0.597
			VLD	113	0.8204	0.703	0.569
$\overline{R^2_{VLD}} = 0.800 \pm 0.0303; \overline{s_{VLD}} = 0.716 \pm 0.0521; \overline{MAE_{VLD}} = 0.564 \pm 0.0370$							
2	1	1	TRN	111	0.6080	1.050	0.919
			iTRN	114	0.6900	1.130	0.984
			CLB	114	0.8434	0.528	0.451
			VLD	111	0.9016	0.511	0.405
		2	TRN	111	0.6528	0.986	0.817
			iTRN	114	0.7308	1.050	0.876
			CLB	114	0.7904	0.610	0.470
			VLD	111	0.9008	0.502	0.381
		3	TRN	111	0.6731	0.957	0.794
			iTRN	114	0.7791	0.971	0.782
			CLB	114	0.8268	0.572	0.456
			VLD	111	0.8853	0.574	0.426
2	2	1	TRN	112	0.6321	1.090	0.882
			iTRN	112	0.7079	0.971	0.828
			CLB	113	0.8048	0.710	0.532
			VLD	113	0.8242	0.639	0.476
		2	TRN	112	0.6719	1.030	0.839
			iTRN	112	0.6847	0.987	0.808
			CLB	113	0.8604	0.660	0.502
			VLD	113	0.8420	0.612	0.486
		3	TRN	112	0.6337	1.090	0.894

Table 1 (continued)

Eq.	Split	Run	Set*	n	R ²	s	MAE
2	3	1	iTRN	112	0.6725	1.020	0.869
			CLB	113	0.8540	0.631	0.502
			VLD	113	0.8423	0.605	0.452
			TRN	114	0.6679	1.040	0.869
			iTRN	110	0.6625	1.060	0.913
			CLB	113	0.7986	0.659	0.514
		2	VLD	113	0.8637	0.583	0.463
			TRN	114	0.6399	1.080	0.913
			iTRN	110	0.6062	1.140	0.957
			CLB	113	0.8415	0.605	0.465
			VLD	113	0.8900	0.531	0.406
			TRN	114	0.7471	0.906	0.725
4	1	1	iTRN	110	0.6750	1.030	0.851
			CLB	113	0.7788	0.737	0.593
			VLD	113	0.8617	0.608	0.470
			TRN	111	0.6468	0.995	0.868
			iTRN	114	0.6613	1.130	0.994
			CLB	114	0.9169	0.398	0.309
		2	VLD	111	0.9344**	0.396	0.318
			TRN	111	0.6302	1.020	0.896
			iTRN	114	0.6921	1.090	0.959
			CLB	114	0.8897	0.462	0.369
			VLD	111	0.9195	0.454	0.356
			TRN	111	0.6226	1.030	0.908
4	2	3	iTRN	114	0.6619	1.140	1.030
			CLB	114	0.9143	0.394	0.313
			VLD	111	0.9322	0.425	0.341
		1	TRN	112	0.6355	1.090	0.935
			iTRN	112	0.6623	1.020	0.866
			CLB	113	0.9174	0.463	0.365
4	3	2 s	VLD	113	0.9211	0.429	0.331
			TRN	112	0.6072	1.130	0.974
			iTRN	112	0.6318	1.080	0.958
			CLB	113	0.9234	0.451	0.368
			VLD	113	0.9325	0.403	0.325
			TRN	112	0.5247	1.240	1.080
4	3	3	iTRN	112	0.5658	1.180	1.060
			CLB	113	0.9027	0.503	0.400
			VLD	113	0.9113	0.458	0.368
		1	TRN	114	0.5351	1.230	1.070
			iTRN	110	0.5538	1.210	1.040
			CLB	113	0.9044	0.428	0.330
4	3	2	VLD	113	0.9325	0.389	0.301
			TRN	114	0.5831	1.160	0.993
			iTRN	110	0.5820	1.180	1.000
			CLB	113	0.8617	0.515	0.406
			VLD	113	0.9178	0.450	0.333
			TRN	114	0.5384	1.220	1.080
4	3	3	iTRN	110	0.5514	1.220	1.060
			CLB	113	0.8596	0.527	0.411
			VLD	113	0.9038	0.476	0.380
			TRN	114	0.5384	1.220	1.080
			iTRN	110	0.5514	1.220	1.060
			CLB	113	0.8596	0.527	0.411

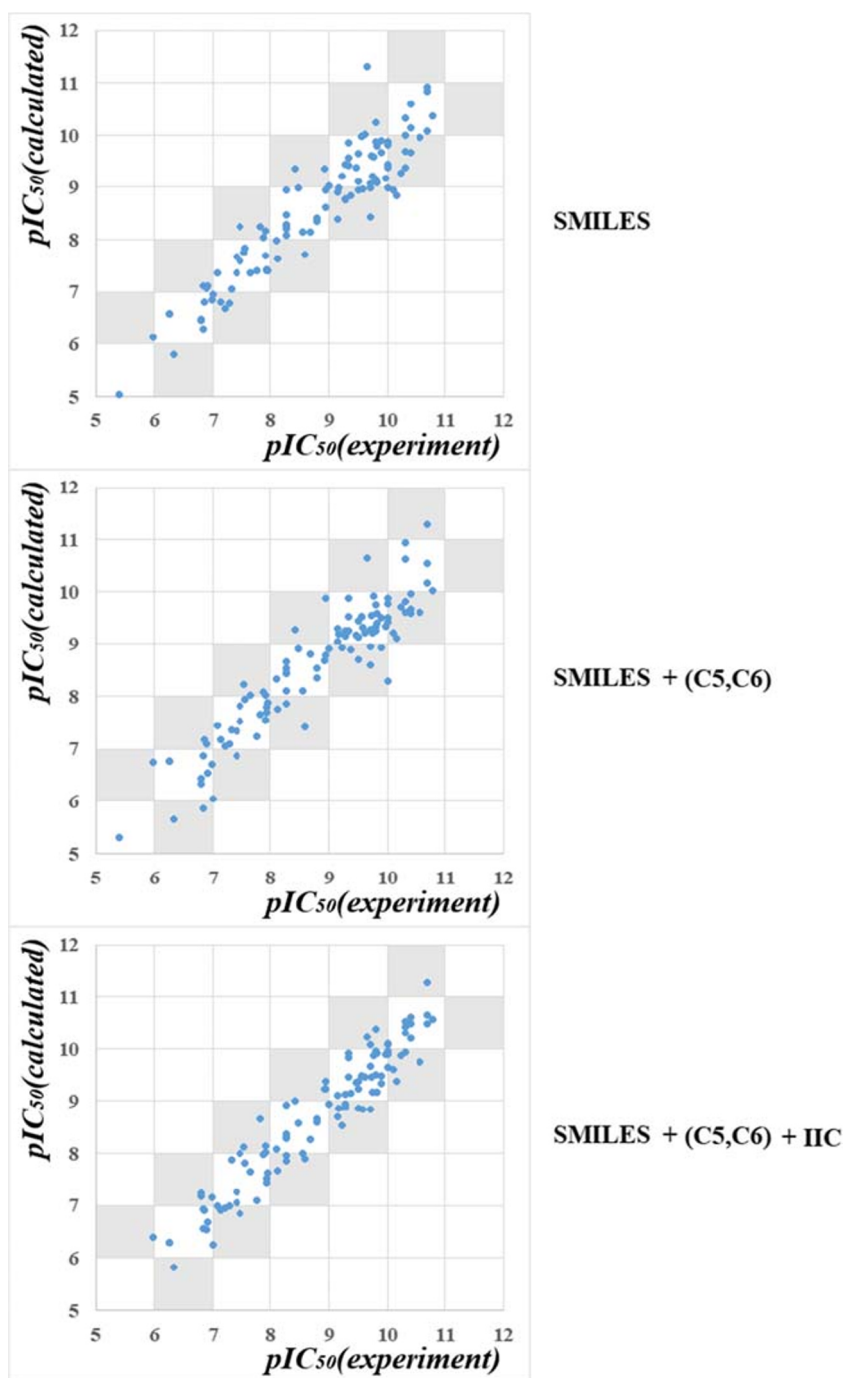
$$\overline{R^2_{VLD}} = 0.868 \pm 0.0265; \overline{s_{VLD}} = 0.574 \pm 0.0458; \overline{MAE_{VLD}} = 0.441 \pm 0.0350$$

$$\overline{R^2_{VLD}} = 0.923 \pm 0.0102; \overline{s_{VLD}} = 0.431 \pm 0.0288; \overline{MAE_{VLD}} = 0.339 \pm 0.0236$$

*) TRN training set; iTRN invisible training set; CLB calibration set; VLD validation set

**) Most attractive models are indicated in bold

Fig. 2 Graphical representations of models of pIC_{50} for external validation sets obtained with optimal descriptors calculated with Eqs. 1, 2, and 4 in the case of split 1



weights, one can build up the predictive model using compounds from the training set:

$$pIC_{50} = C_0 + C_1 * DCW(T^*, N^*) \quad (3)$$

The predictive potential of the model calculated with Eq. 3 should be checked up with external validation set.

Recently, so-called index of ideality of correlation (IIC) [28–31] has been tested as a tool to improve the predictive potential of a model. Factually, the improvement is the Monte Carlo optimization of correlation weights for the molecular

features above extracted from SMILES for Eq. 2 with the application of the IIC as additional component of the target function:

$$DCW(T^*, N^*, IIC) = CW(C5) + CW(C6) + \sum_{k=1}^{NA} CW(S_k) + \sum_{k=1}^{NA-1} CW(SS_k) + \sum_{k=1}^{NA-2} CW(SSS_k) \quad (4)$$

Results and discussion

The statistical quality of examined models estimated via the statistical characteristics for the validation set.

The optimal descriptor calculated with Eq. 1 (solely SMILES) gives the following models for three random splits:

$$pIC50(split\ 1) = 2.4214 (\pm 0.0278) + 0.1993 (\pm 0.0009) * DCW(1, 2) \quad (5)$$

$$pIC50(split\ 2) = 2.2202 (\pm 0.0307) + 0.2102 (\pm 0.0009) * DCW(1, 2) \quad (6)$$

$$pIC50(split\ 3) = 2.1413 (\pm 0.0305) + 0.1581 (\pm 0.0007) * DCW(1, 2) \quad (7)$$

The optimal descriptor calculated with Eq. 2 (SMILES together with rings C5 and C6) gives the following best models for three random splits:

$$pIC50(split\ 1) = 2.1077 (\pm 0.0521) + 0.0923 (\pm 0.0007) * DCW(1, 3) \quad (8)$$

$$pIC50(split\ 2) = 1.3947 (\pm 0.0446) + 0.1914 (\pm 0.0012) * DCW(1, 3) \quad (9)$$

$$pIC50(split\ 3) = 1.7914 (\pm 0.0436) + 0.1288 (\pm 0.0008) * DCW(1, 3) \quad (10)$$

The optimal descriptor calculated with Eq. 4 (SMILES together with C5 and C6, plus the target function of the Monte Carlo optimization contains the *IIC*) gives the following best models for three random splits:

$$pIC50(split\ 1) = 0.6941 (\pm 0.0543) + 0.1437 (\pm 0.0010) * DCW(1, 10) \quad (11)$$

$$pIC50(split\ 2) = 1.5934 (\pm 0.0459) + 0.1617 (\pm 0.0011) * DCW(1, 10) \quad (12)$$

$$pIC50(split\ 3) = 1.2136 (\pm 0.0544) + 0.1069 (\pm 0.0008) * DCW(1, 10) \quad (13)$$

Table 1 contains the statistical characteristics of these models. Table 1 indicates that the dispersion of these models in several runs of the corresponding Monte Carlo optimization actually remains similar for the cases of descriptor calculated with Eqs. 1, 2, and 4. However, according to average of the determination coefficient for validation set, Eq. 2 gives better model: $\overline{R^2_{VLD}} = 0.800 (Eq. 1)$ vs. $\overline{R^2_{VLD}} = 0.868 (Eq. 2)$. Consequently, taking into account of the presence of molecular rings [24–27] improves the predictive potential of the CORAL models. These models calculated that taking into account the *IIC* gives addition improvement of these models $\overline{R^2_{VLD}} = 0.923 (Eq. 4)$.

Figure 2 contains graphical representations of models based on optimal descriptors calculated with Eqs. 1, 2, and 4

Table 2 Molecular fragments related to different rings, which are promoters for increase anti-HIV-1 activity

SMILES attribute*	Run 1	Run 2	Run 3	N1	N2	N3	Defect
Split 1							
C5...A...1...	0.26267	0.72379	0.46564	42	45	32	0.0013
C6...AH.4...	0.16106	0.40993	0.06486	30	26	44	0.0016
C5...AH.2...	0.24249	0.26925	0.18428	15	13	25	0.0021
C6...AH.3...	0.16196	0.49875	0.81998	9	14	19	0.0031
Split 2							
C5...A...1...	0.62778	1.34055	0.47117	44	43	37	0.0008
C6...A...3...	0.33554	0.60756	0.68471	36	39	22	0.0022
C6...AH.4...	0.27742	0.41411	0.18363	26	30	38	0.0016
C5...AH.2...	0.15125	0.36622	0.39355	14	13	20	0.0015
Split 3							
C6...A...3...	0.38266	0.04536	0.48122	37	37	23	0.0020
C6...AH.4...	0.51998	0.26144	0.67700	29	28	34	0.0007
C5...H.1...	0.34284	0.51002	0.43070	17	19	25	0.0017
C6...AH.5...	0.34670	0.12587	0.39795	16	12	15	0.0002

*C5 and C6 are indicators of five-member and six-member rings, respectively; A is indicator of presence of aromaticity; H is indicator of presence of heteroatoms in a ring; the digit (1–4) indicates the number of rings in a molecule; N1, N2, and N3 are frequencies of SMILES attribute in the training, invisible training, and calibration sets, respectively; the defect is the statistical defect of SMILES attribute [24–26]

for external validation sets (i.e., for substances invisible during building up a model).

Table 2 contains example of promoters for increase of the pIC₅₀ values. Indeed, the presence of five and six-member molecular rings as well as aromaticity and heteroatoms are factors with significant influence upon the anti-HIV-1 activity [1].

The statistical characteristics of the best CORAL model for the validation set suggested in the literature [1] are the following: $n = 104$, $R^2 = 0.762$, $MAE = 0.731$. The comparison with the average statistics for the validation set (Table 1) indicates that the model herein developed (where kinds of molecular rings and the *IIC* taken into account) seems more suitable than the model suggested in the literature [1].

It is to be noted that predictive models should be checked up with several splits (into the training set and the validation set), because as it is shown above, the dispersions of statistical characteristics exist. Consequently, examination of solely one split gives far of completed information about an approach.

Supplementary materials contain technical details of the CORAL model.

Conclusions

The CORAL software (<http://www.insilico.eu/coral>) gives good predictive models for HIV protease inhibitor activity (pIC₅₀). In other words, the Monte Carlo technique that applied to build up QSAR models gives satisfactory models for the anti-HIV-1 activity. Results showed that taking into account of the correlation weights for different kinds of molecular rings improved the predictive potential of these models. Applying of the index of ideality of correlation gives further improvement of the described models. There are dispersions of the statistical characteristics of the examined models in several runs of the Monte Carlo optimization as well as for different splits into the training and validation sets. However, in spite of the above-mentioned dispersions, the approach gives reproducible models.

Author's contributions Authors have done equivalent contributions to this work.

Funding information The authors are grateful for the contribution of the project LIFE-CONCERT contract (LIFE17 GIE/IT/000461) for the financial support.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest

References

- Halder AK (2018). SAR QSAR Environ Res 29(11):911–933. <https://doi.org/10.1080/1062936X.2018.1529702>
- Toropov AA, Toropova AP, Benfenati E, Leszczynska D, Leszczynski J (2010). J Comput Chem 31(2):381–392. <https://doi.org/10.1002/jcc.21333>
- Toropova AP, Toropov AA, Benfenati E, Leszczynska D, Leszczynski J (2010). J Math Chem 48(4):959–987. <https://doi.org/10.1007/s10910-010-9719-x>
- Toropova AP, Toropov AA, Veselinović JB, Miljković FN, Veselinović AM (2014). Eur J Med Chem 77:298–305. <https://doi.org/10.1016/j.ejmech.2014.03.013>
- Toropov AA, Toropova AP, Benfenati E, Dorne JL (2018). Chem Biol Interact 290:1–5. <https://doi.org/10.1016/j.cbi.2018.04.030>
- Toropova AP, Toropov AA, Marzo M, Escher SE, Dorne JL, Georgiadis N, Benfenati E (2018). Food Chem Toxicol 112:544–550. <https://doi.org/10.1016/j.fct.2017.03.060>
- Lv Z, Chu Y, Wang Y (2015). HIV AIDS (Auckl) 7:95–104. <https://doi.org/10.2147/HIV.S79956>
- Darnag R, Minaoui B, Fakir M (2017). Arab J Chem 10:S600–S608. <https://doi.org/10.1016/j.arabjc.2012.10.021>
- OECD (2006) Report on the regulatory uses and applications in OECD member countries of (quantitative) structure-activity relationship [(Q)SAR] models in the assessment of new and existing chemicals. OECD Pap 6
- Hdoufane I, Stoycheva J, Tadjer A, Villemin D, Najdoska-Bogdanov M, Bogdanov J, Cherqaoui D (2019). J Mol Struct 1193:429–443. <https://doi.org/10.1016/j.molstruc.2019.05.056>
- Muthukumaran P, Rajiniraja M (2019). J Theor Biol 469:18–24. <https://doi.org/10.1016/j.jtbi.2019.02.019>
- Barzegar A, Zamani-Gharehchamani E, Kakhodaie-Ilkhchi A (2017). Future Med Chem 9(11):1175–1191. <https://doi.org/10.4155/fmc-2017-0040>
- Sapre NS, Jain N, Gupta S, Sapre N (2013). RSC Adv 3(26): 10442–10451. <https://doi.org/10.1039/C3RA40685G>
- Fernández M, Caballero J (2006). Bioorg Med Chem 14(1):280–294. <https://doi.org/10.1016/j.bmc.2005.08.022>
- Sapre NS, Gupta S, Pancholi N, Sapre N (2009). J Comput Chem 30(6):922–933. <https://doi.org/10.1002/jcc.21114>
- Wei Y, Li W, Du T, Hong Z, Lin J (2019). Int J Mol Sci 20(14): 3572. <https://doi.org/10.3390/ijms20143572>
- Hdoufane I, Bjij I, Soliman M, Tadjer A, Villemin D, Bogdanov J, Cherqaoui D (2018). Pharmaceuticals 11(3):69. <https://doi.org/10.3390/ph11030069>
- Shiri F, Pirhadi S, Rahmani A (2018). J Recept Sig Transd 38(1): 37–47. <https://doi.org/10.1080/10799893.2017.1414844>
- Bhargava S, Adhikari N, Amin SA, Das K, Gayen S, Jha T (2017). SAR QSAR Environ Res 28(12):973–990. <https://doi.org/10.1080/1062936X.2017.1388281>
- Putz MV, Dudas NA (2013). Molecules 18(8):9061–9116. <https://doi.org/10.3390/molecules18089061>
- Zhou Z, Lin X, Madura JD (2006). Infect Disord-Drug Targets 6(4): 391–413. <https://doi.org/10.2174/187152606779025833>
- Zhou Z, Madura JD (2004). J Chem Inf Comput Sci 44(6):2167–2178. <https://doi.org/10.1021/ci049893v>
- Toropov AA, Toropova AP, Marzo M, Dorne JL, Georgiadis N, Benfenati E (2017). Environ Toxicol Pharmacol 53:158–163. <https://doi.org/10.1016/j.etap.2017.05.011>
- Toropov AA, Toropova AP (2018) In: Bidoia E, Montagnoli R (eds) Toxicity and biodegradation testing. Methods in Pharmacology and Toxicology. Humana Press, New York, NY, pp 147–183. https://doi.org/10.1007/978-1-4939-7425-2_8

25. Toropov AA, Toropova AP, Benfenati E, Gini G, Leszczynska D, Leszczynski J (2012). *Anti Cancer Agents Med Chem* 12(7):807–817. <https://doi.org/10.2174/187152012802650255>
26. Velázquez-Libera JL, Caballero J, Toropova AP, Toropov AA (2019). *Chemometr Intell Lab Syst* 184:14–21. <https://doi.org/10.1016/j.chemolab.2018.11.008>
27. Toropov AA, Toropova AP, Benfenati E, Salmona M (2018). *Toxicol Mech Methods* 28(5):321–327. <https://doi.org/10.1080/15376516.2017.1422579>
28. Toropov AA, Toropova AP (2017). *Mutat Res Genet Toxicol Environ Mutagen* 819:31–37. <https://doi.org/10.1016/j.mrgentox.2017.05.008>
29. Toropova AP, Toropov AA (2017). *Sci Total Environ* 586:466–472. <https://doi.org/10.1016/j.scitotenv.2017.01.198>
30. Kumar P, Kumar A, Sindhu J (2019). *SAR QSAR Environ Res* 30(2):63–80. <https://doi.org/10.1080/1062936X.2018.1564067>
31. Ahmadi S (2020). *Chemosphere* 242:125192. <https://doi.org/10.1016/j.chemosphere.2019.125192>

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