Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 43 (2008) 471–477
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

SAR and QSAR studies: Modelling of new DAPY derivatives

Abhilash Thakur [a,*], Mamta Thakur [b], Arpan Bharadwaj [c], Suprajnya Thakur [d]

[a] *Department of Chemistry, Rishiraj Institute of Technology, Gram Ravti, Sanwer Road, Indore, M.P., India*
[b] *Department of Chemistry, Kailash Chand Bansal Academy of Technology, Indore, India*
[c] *Department of Chemistry, Government Madhav Science College, Ujjain, India*
[d] *Department of Physics, Shri Vaishnav Institute of Technology & Science, Indore, India*

Received 11 November 2006; received in revised form 9 March 2007; accepted 2 April 2007
Available online 3 May 2007

Abstract

This work describes QSAR and SAR studies on the inhibition of reverse transcriptase by 31 novel DAPY (diarylpyrimidine) derivatives using both topological and physicochemical properties and molecular modelling parameters along with indicator parameters. The application of a multiple linear regression analysis indicated that a combination of topological and physicochemical descriptors and the indicator parameters yielded a statistically significant model for the prediction of the activity, log 1/*C* (50% of effective concentration of DAPY derivatives for RTs). The modelling of some new potential DAPY compounds and their maximum active comformers for the inhibition of reverse transcriptase are made by quantum molecular modelling.

Keywords: Reverse transcriptase; QSAR; Correlation; AIDS; DAPY derivatives; MM+ force field; Multiple linear regression

1. Introduction

The acquired immune-deficiency syndrome (AIDS) is one of the most hazardous diseases, which is caused by infection with the human immune-deficiency virus (HIV). Reverse transcriptase (RT) is the key for HIV replication and is not required for normal host cell replication. During the inspection of effective therapies facing HIV, reverse transcriptase (RT) has been identified as one of the most promising targets [1].

Substituted NNRTIs, (non-nucleotides reverse transcriptase inhibitors) derivatives find extensive applications in inhibition of reverse transcriptase [1]. At the same time NRTIs (nucleotides reverse transcriptase inhibitors) are also studied for their fast activity. It is, thus, important to understand the potential inhibition activity and its dependence on structural, electronic, and steric features. Thus, there is a compelling need to understand the mechanisms and correlation modes of potential inhibition activity with reverse transcriptase. Over the past one decade, several groups have studied the effect and significance of the NNRTIs [2–5]. In the recent past Ludovici [2] studied the biological activities for various binding sites and cells for inhibition of RTs by these novel classes of DAPY (diarylpyrimidine), NNRTIs. The study on biological activity of novel class of NNRTIs has become more important due to the emergence of multi drug resistance and toxic effects after the prolonged doses [3].

To understand the role of molecular, steric and hydrophobic properties in the biological activity of novel class of NNRTIs, derivatives and to model some new derivatives with high potential as RTIs, SAR and QSAR studies are performed on some new DAPY derivatives. Also efforts have been made to generate the new class of compounds having substituted alkyl chain as spacer to combine with the NRTIs' functional groups by without affecting the potential of the compounds as NNRTIs. To understand the structural and molecular effect on biological activity topological indices [6–10] and some physicochemical properties along with the indicator parameters are tested. To get insight into the binding with receptor

* Corresponding author. Tel.: +91 731 2792454; fax: +91 731 2763618.
E-mail address: abhilashthakur@yahoo.com (A. Thakur).

0223-5234/$ - see front matter © 2007 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2007.04.005

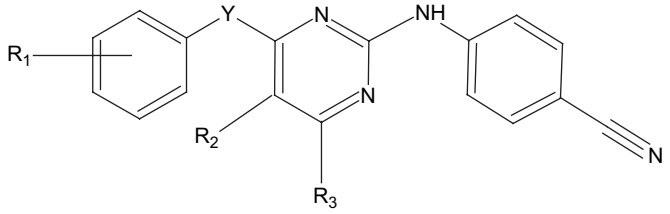

Fig. 1. Parent structure for the DAPY derivatives used in present study.

site and to find out suitable comformers for the binding with receptor we have gone through the energy minimization using molecular mechanics MM+ force field.

This work describes QSAR and SAR studies on the inhibition of reverse transcriptase by 31 novel DAPY derivatives using both topological and physicochemical properties and molecular modelling parameters along with indicator parameters. The application of a multiple linear regression analysis indicated that a combination of topological and physicochemical descriptors and the indicator parameters yielded a statistically significant model for the prediction of activity, log $1/C$ (50% of effective concentration of DAPY derivatives for RTs). The modelling of some new potential DAPY derivatives and their maximum active comformers for the inhibition of reverse transcriptase are made by quantum molecular modelling (Fig. 1).

2. Experimental

Biological activity: The reverse transcriptase inhibition activity mentioned (log $1/C$) is taken from the literature [2].

Topological indices: All the topological indices used are calculated from the hydrogen suppressed molecular graphs. Though their calculations are exclusively discussed in the literature [6–10] we give below the expressions used for their calculations.

Wiener index $W = W(G)$ and G is defined as the half-sum of the elements of the distance matrix:

$$W = W(G) = \frac{1}{2}\sum_{i=1,}\sum_{j=1} D_{ij} \tag{1}$$

where $(D)_{ij}$ is the ijth element of the distance matrix which denotes the shortest graph theoretical distance between sites i and j of G.

The *connectivity index* ${}^1\chi = {}^1\chi(G)$ and G is defined by Randic as:

$${}^1\chi = {}^1\chi(G) = \sum_{i,j}[d(i)d(j)]^{-0.5} \tag{2}$$

Balaban index $J = J(G)$ and G is defined as:

$$J = \frac{M}{\mu} + 1 \sum_{\text{bonds}} \left(d_i d_j\right)^{-0.5} \tag{3}$$

where M is the number of bonds in G, μ is the cyclomatic number of G, and d_i ($i = 1, 2, 3, \ldots N$; N is the number of vertices in G) is the distance sum. The cyclomatic number $\mu = \mu(G)$ of a cyclic graph G is equal to the minimum number of edges necessary to be erased from G in order to transform it into the related acyclic graph. In case of monocyclic graph $\mu = 1$ otherwise it is calculated by means of the following expression

$$M = M - N + 1 \tag{4}$$

Szeged index $\mathrm{Sz} = \mathrm{Sz}(G)$, is calculated according to the following expression:

$$\mathrm{Sz} = \mathrm{Sz}(G) = \sum_{\text{edges}} n_u n_v \tag{5}$$

where n_u is the number of vertices lying closer to one end of the edge $e = uv$; the meaning of n_v is analogous. Edges equidistant from both the ends of an edge, $e = uv$ are not taken into account.

Molecular descriptors: Physicochemical properties MR, MV, Pc, RI, ST and D used in proposed investigation, are calculated with help of ACD lab's [11] free wear and the modelling parameters DpM, TE, and RMSg, are calculated by the Hyperchem7 (Demo version) [12].

Indicator parameters: In the present study three different indicator parameters are used for the understanding of positional significance of substituents in their activities/properties.

Regression analysis: All the regressions were carried out using maximum R^2 method [13].

Software: The calculation of topological indices and regression analysis were performed using computer program developed for in house use only. In addition, we have used ACD lab's software [11] and Hyperchem [12] for calculations. We also used the molecular modelling pro (Demo version) [14] for the molecule to molecule docking.

3. Results and discussion

As mentioned in the Section 1, to analyse the relationship between biological activity and the structure of the molecule and various molecular properties of DAPY derivatives mentioned in Table 1, we test the topological indices along with the classical physicochemical properties presented in Table 2. Indicator parameters used in present study to understand the effect of substitution at various positions are also listed in the same table (Table 2). The univariate correlation shows that except hydrophobic parameter log P, none of the topological or physicochemical parameters have the statistically significant correlation with the biological activity. From the various bi, tri and tetra parametric combinations of log P with topological, physicochemical and indicator parameters very few give the statistically significant results but the best result is given by the combination of log P, and indicator parameters I_1 and I_2 with the r value of 0.6059 which demonstrates the domination of hydrophobicity in modelling of activity. It also shows that the increase in the hydrophobicity inhibits the activity of this specific series of compound. Magnitude of indicator parameter I_1 shows the importance

Table 1
DAPY derivatives and their biological activities used in present study

Compd. no.	R_1	R_2	R_3	Y
1	2,4,6-TriMe	–	–	N
2	2,6-DiMe-4-CN	–	–	O
3	2,6-DiMe-4-CN	–	–	N
4	2,6-DiMe-4-Br	–	–	O
5	2,6-DiMe-4-Br	–	–	S
6	2,6-DiMe-4-(HCC)	–	–	O
7	2,4,6-TriMe	–	–	S
8	2,4,6-TriMe	–	–	O
9	2,6-DiBr-4-F	–	–	N
10	2,4,6-TriCl	–	–	N
11	2,6-DiMe	–	–	N
12	2,4-DiCl-6-Me	–	–	N
13	2,6-DiMe-4-Cl	–	–	N
14	2,6-DiBr-4-Me	–	–	N
15	2,6-DiMe-4-Br	–	–	N
16	2,6-DiMe-4-CN	Br	–	N
17	2,6-DiMe-4-CN	Br	–	O
18	2,4,6-TriMe	Br	–	N
19	2,4,6-TriMe	HCC	–	N
20	2,6-DiMe-4-CN	Vinyl	–	N
21	2,6-DiMe-4-CN	CN	–	N
22	2,4,6-TriMe	CN	–	N
23	2,6-DiMe-4-CN	Cl	–	N
24	2,6-DiMe-4-CN	Cl	–	O
25	2,4,6-TriMe	Cl	–	N
26	2,6-DiMe-4-CN	Me	–	N
27	2,4,6-TriMe	Me	–	N
28	2,6-DiMe-4-CN	NO_2	–	O
29	2,6-DiMe-4-CN	NH_2	–	O
30	2,6-DiMe-4-CN	NHAc	–	O
31	2,6-DiMe-4-CN	Br	NH_2	O
32[a]	2-Pr-6-Me-4-CN	Br	–	O
33[a]	2-Pr-6-Me-4-CN	Br	–	N
34[a]	2-OMeEt-6-Me-4-CN	Br	–	N
35[a]	2-OMeEt-6-Me-4-CN	Br	–	O

[a] Newly modelled molecules (test compounds).

Table 2
Parameters used in present study for DAPY derivatives

Compd. no.	J	MV	log P	I_1	I_2
1	1.448	267.7	1.9	1	0
2	1.447	259.8	1.64	0	1
3	1.447	261.3	1.47	1	1
4	1.448	263.3	1.97	0	0
5	1.448	271	2.31	0	0
6	1.447	267.3	1.79	0	0
7	1.448	274	2.42	0	0
8	1.448	266.2	2.07	0	0
9	1.448	250.5	0.94	1	0
10	1.448	253.1	0.77	1	0
11	1.441	252	1.74	1	0
12	1.448	258	1.14	1	0
13	1.448	262.8	1.52	1	0
14	1.448	261.7	1.69	1	0
15	1.448	264.7	1.79	1	0
16	1.498	274	1.64	1	1
17	1.498	272.5	1.82	0	1
18	1.499	280.4	2.07	1	0
19	1.534	284.5	1.89	1	0
20	1.534	294.9	2.36	1	0
21	1.534	270.6	1.31	1	1
22	1.534	277	1.74	1	0
23	1.498	272.1	1.37	1	1
24	1.498	270.7	1.54	0	1
25	1.499	278.5	1.80	1	0
26	1.498	277	1.74	1	1
27	1.499	283.4	2.17	1	0
28	1.572	270.9	−0.98	0	1
29	1.498	262.9	0.05	0	1
30	1.582	296	−0.12	0	1
31	1.549	275.6	1.82	0	1

J = Balaban branching index; MV = molar volume; log P = octanol/water partition coefficient; I_1 = indicator parameter accounted for presence of –NH– group; I_2 = indicator parameter accounted for presence of –CN group on ring.

of –NH– substitution in the activity of these compounds, while negative coefficient of indicator parameter I_2 shows the inverse relationship between the presence of –CN group and the inhibition activity in quantitative manner. But this combination is not adequate to explain the relationship between the structure of the molecule and molecular properties with biological activity. This combination having two outliers and after the deletion of these two compounds (compd. nos. 28, 29) there is a significant increase in the value of R^2_A but at the same time the Durbin Watson parameter of the model decreases from 2.4 to 1.56, hence it is statistically insignificant.

For the QSAR study and modelling of the compounds further regression analysis has been done using the topological [6–10] and physicochemical properties.

Various bi and tri parametric combination gives the better statistics but the best result is given by the tetra parametric combination of Balaban branching index (J) [8], molar volume (MV), indicator parameters I_1 and I_2. The model obtained from this combination is presented below:

$$\log\frac{1}{C} = 0.0566(\pm 0.0224)\,J - 1.0020\times 10^{-4}\left(\pm 7.6380\times 10^{-5}\right)\mathrm{MV} - 0.0029(\pm 0.0011)\,I_1 - 0.0030(\pm 0.0014)\,I_2 - 0.0516$$
$$n = 31,\ \mathrm{Se} = 0.0023,\ R = 0.5749,$$
$$R^2 = 0.3305,\ R^2_A = 0.2275,\ F = 3.208 \quad (6)$$

Eq. (6) is marginally statistically significant and shows the domination of structural and volumetric properties in the modelling of inhibition activity of these DAPY derivatives. It also exhibits the role of the presence of –NH– group parent moiety in the inhibition activity and the role of presence of –CN group in the same. Magnitude of I_1 and MV shows the domination of presence of –NH– group and high volume on the activity. Negative coefficient of MV exhibits the increase in the activity with the increase in the volume of the molecule. Similar relation is present between the presence of –NH– group on moiety and the inhibition activity, i.e., presence of –NH– favours the inhibition activity of DAPY derivatives.

For the further improvement in the predictive potential and modelling efficiency outlier have been done, based on the residue from Eq. (6). In this process five compounds were outliered in three different steps. Model obtained after the deletion of these five compounds are given below:

$$\log\frac{1}{C} = -0.0152(\pm 0.0098)\,J + 9.5265\times10^{-5}\left(\pm 3.13123\times10^{-5}\right)\mathrm{MV} - 0.0014\left(\pm 4.21385\times10^{-4}\right)I_1 - 0.0016\left(\pm 5.04836\times10^{-4}\right)I_2 + 1.89686\times10^{-4}$$
$$n = 30,\ \mathrm{Se} = 0.0010,\ R = 0.77869,\ R^2 = 0.6192,\ R^2_{\mathrm{A}} = 0.5583,\ F = 10.164 \quad (7)$$

$$\log\frac{1}{C} = -0.0143(\pm 0.0083)\,J + 8.29798\times10^{-5}\left(\pm 2.67597\times10^{-5}\right)\mathrm{MV} - 0.0015\left(\pm 3.59056\times10^{-4}\right)I_1 - 0.0015\left(\pm 4.28981\times10^{-4}\right)I_2 + 0.0021$$
$$n = 29,\ \mathrm{Se} = 0.0009,\ R = 0.8241,\ R^2 = 0.6792,\ R^2_{\mathrm{A}} = 0.6257,\ F = 12.701 \quad (8)$$

$$\log\frac{1}{C} = -0.0116(\pm 0.0045)\,J + 6.44503\times10^{-5}\left(\pm 1.42581\times10^{-5}\right)\mathrm{MV} - 0.0011\left(\pm 2.03398\times10^{-4}\right)I_1 - 9.5138\times10^{-4}\left(\pm 2.33904\times10^{-4}\right)I_2 + 0.0024$$
$$n = 26,\ \mathrm{Se} = 0.0005,\ R = 0.8826,\ R^2 = 0.7790,\ R^2_{\mathrm{A}} = 0.7369,\ F = 18.504 \quad (9)$$

Comparison of the result with log P and Eqs. (6)–(9) indicate that the hydrophobicity plays no role in the prediction of inhibition activity of these compounds. The comparison of obtained models also indicates that, dependence of biological activity on structural and molecular properties is not justified by the five outlier compounds and these compounds behave exceptionally from their parent series. Model obtained for the set of 26 compounds gives the better statistics and the most suitable for the prediction of inhibition activity of the compounds against reverse transcriptase-1. It also demonstrates the domination of volumetric properties over the branching and other structural and physicochemical features. Eqs. also exhibits the significant presence of –NH– favouring the inhibition activity of this class of compounds. It also expresses the favourable presence of –CN group in modelling the activity. In Eq. (9) greater magnitude of indicator I_1 in comparison to I_2 explores the domination of presence of –NH– over the –CN on parent moiety to model the inhibition activity of the DAPY derivatives.

It is obvious that reduction in size of data set increases the regression value, but in present case significant lowering of Se (standard error of estimation) and a noted improvement in the F-statistics along with the improvement in the value of R^2_{A} from Eqs. (6) to (9) justify the improvement in statistics and deletion of the compounds.

At this stage, it is laudable to comment on R^2_{A} values. We pragmatic that as we pass from the model obtained for 31 compounds (Eq. (6)) to model obtained for 26 compounds (Eq. (9)) there is consistent increase in R^2_{A}, from 0.2275 to 0.7369, as we pass from Eqs. (6) to (9). Such an increase in R^2_{A} values indicates that the deleted compounds have the unfair share in the modelling of respective activity and also shows exceptional behavior from their parent series. The value of R^2_{A} will decrease if the deletion of the compounds does not reduce the unexplained variation in the model enough to off set the loss of degree of freedom [15–19]. The correlation between the estimated log 1/C and observed log 1/C is represented in Fig. 2.

In the second phase of our study based on the second category containing five compounds, compd. nos. 5, 6, 18, 19 and 28, all these compounds (outliers) taken together resulted into a model according to the following equations:

$$\log\frac{1}{C} = 0.0731(\pm 0.0449)\,J - 0.0992$$
$$n = 5,\ \mathrm{Se} = 0.0049,\ R = 0.6759,\ F = 2.523 \quad (10)$$

$$\log\frac{1}{C} = -0.0041\left(\pm 5.2458\times10^{-4}\right)\log P + 0.0136$$
$$n = 5,\ \mathrm{Se} = 0.0014,\ R = -0.9768,\ F = 62.414 \quad (11)$$

Model presented in the form of Eqs. (10) and (11) expresses the domination of branching and hydrophobic parameter log P in comparison with steric and volumetric properties for the set of five outliers.

Eqs. (10) and (11) also show the domination of hydrophobic properties over the branching and size specific properties for modelling the activity log 1/C for the set of five outliers. Equations also demonstrate the little favour of branching property to the biological activity log 1/C and confirm the previous statement about the justification to un-applicability of hydrophobicity in modelling of 31 DAPY derivatives.

In order to carry out quantum computations, we have first carried out the molecular geometry optimizations to find out the structural behavior of these compounds as a function of attached groups and their positions. The corresponding molecular modelling parameters are shown in Table 4.

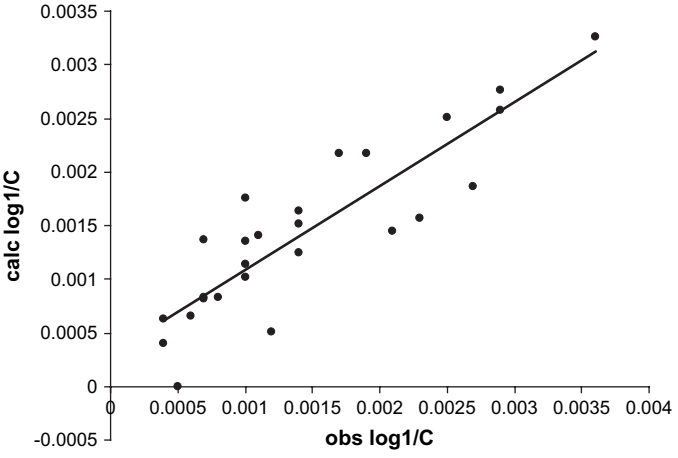

Fig. 2. Graph obtained between observed and calculated log 1/C from Eq. (9).

Table 3
Observed and calculated biological activities (Eq. (9)) for the DAPY derivatives

Compd. no.	log 1/C[a]	log 1/C[b]	Residue
1	0.0010	0.0018	−0.0008
2	0.0011	0.0014	−0.0003
3	0.0004	0.0004	0.0000
4	0.0029	0.0026	0.0003
5	0.0057[c]	0.0031	0.0026
6	0.0055[c]	0.0028	0.0027
7	0.0036	0.0033	0.0003
8	0.0029	0.0028	0.0001
9	0.0006	0.0006	0.0000
10	0.0007	0.0008	−0.0001
11	0.0007	0.0008	−0.0001
12	0.001	0.0011	−0.0001
13	0.0021	0.0014	0.0007
14	0.0007	0.0014	−0.0007
15	0.0023	0.0016	0.0007
16	0.0004	0.0006	−0.0002
17	0.0014	0.0016	−0.0002
18	0.0055[c]	0.0020	0.0035
19	0.0042[c]	0.0018	0.0024
20	0.0025	0.0025	0.0000
21	0.0005	−5.53E−06	0.0005
22	0.001	0.0014	−0.0003
23	0.0012	0.0005	0.0007
24	0.0014	0.0015	−0.0001
25	0.0027	0.0019	0.0008
26	0.0008	0.0008	0.0000
27	0.0017	0.0022	−0.0005
28	0.0180[c]	0.0007	0.0173
29	0.0010	0.0010	0.0000
30	0.0019	0.0022	−0.0003
31	0.0014	0.0012	0.0002

[a] Observed biological activity.
[b] Calculated biological activity.
[c] Data points not included in calculations.

Based on the above study and magnitude of residue from Eq. (9), demonstrated in Table 3 we have selected compounds viz., 3, 9, 20, 26, 29 to correlate their modelling parameters with the activities. This we have done to find out which DAPY derivative has the highest correlative and predictive potential for the same category.

In order to resolve our problem of selecting out the DAPY derivative with the best quality and correlation potential; we have carried out further regression analysis using the molecular modelling parameters from Table 4.

The models obtained from all the three modelling parameters are statistically insignificant with the high value of standard deviation than the correlation coefficient. From these five compounds various combinations of the compounds show that the removal of compd. no. 29 gives the informative model with the modelling parameters. The models obtained from modelling parameters are given below:

$$\log\frac{1}{C} = 6.3283\times10^{-4}\left(\pm5.82792\times10^{-4}\right)\mathrm{TE} + 0.0037$$
$$n = 4,\ \mathrm{Se} = 0.0009,\ R = -0.6090,\ F = 1.179 \qquad (12)$$

$$\log\frac{1}{C} = 2.42809\times10^{-4}\left(\pm6.48638\times10^{-5}\right)\mathrm{DpM} - 0.0066$$
$$n = 4,\ \mathrm{Se} = 0.0004,\ R = 0.9355,\ F = 14.013 \qquad (13)$$

$$\log\frac{1}{C} = 0.0044\left(\pm8.84375\times10^{-4}\right)\mathrm{RMSg} - 7.6778\times10^{-6}$$
$$n = 4,\ \mathrm{Se} = 0.0003,\ R = 0.9621,\ F = 24.870 \qquad (14)$$

All three Eqs. (12)–(14) exhibit the presence of these compounds in their minimum comformers energy state during the probable binding with receptor and their dependence on the dipole moment. Model also expresses that the increase in the dipole moment is unfavourable for the biological activity, i.e., binding with the receptor. Highest correlative potential of biological activity with RMSg shows that the flexibility favours the biological activity and helps to bind with the receptor for particular RT's inhibition activity.

All these equations also indicate about the rigidness of compd. no. 29 and its less probability towards the binding with receptors rather this compound shows the minimum residue with molecular parameters and also shows the positive participation in modelling the activity initially.

On the basis of modelling parameters and QSAR we model the four new compounds with the spacer to attach the NRTs and become bi-functional. To analyse the effect of spacer on binding with receptor and to calculate the activity of these compounds we use Eq. (9). The activity calculated for the new compounds is comparatively higher than that of most of the compounds from the parent series; it may be due to increase in volume and branching by the spacer. With the calculated activity we analyse these four compounds along with the four compounds tested with modelling parameters from Table 5. Models obtained with this test set of four new compounds are shown below:

$$\log\frac{1}{C} = 0.0033\left(\pm4.42355\times10^{-4}\right)\mathrm{DpM} + 0.0144$$
$$n = 4,\ \mathrm{Se} = 0.0003,\ R = -0.9828,\ F = 56.660 \qquad (15)$$

Table 4
Modelling parameters for the compounds having minimum residue

Compd. no.	DpM	TE	RMSg
3	5.059	27.379	0.08894
9	3.335	31.13	0.09029
20	3.274	36.416	0.5345
26	4.728	31.196	0.2682
29	6.683	16.174	1.017

DpM = dipole moment; TE = total energy; RMSg = root mean square gradient.

Table 5
Molecular and modelling parameters for the new test compounds[a]

Compd. no.	DpM	TE	RMSg	J	MV	I_1	I_2
1	3.427	17.7947	0.07731	1.541	304.8	0	1
2	3.599	83.307	0.07752	1.541	306.3	1	1
3	2.709	131.9336	0.7896	1.548	312.3	1	1
4	3.096	19.80788	0.08111	1.548	310.9	0	1

[a] Refer footnotes of Tables 2 and 4.

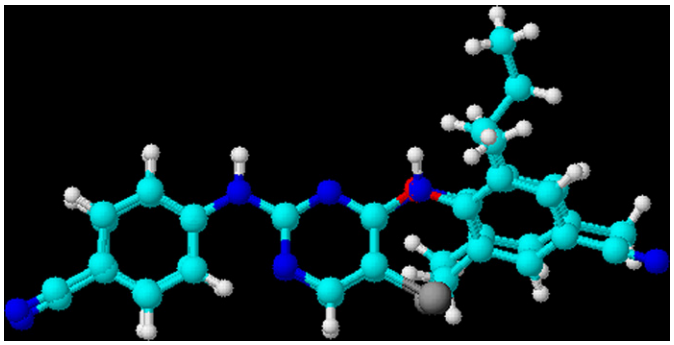

Fig. 3. Molecule to molecule docking of compd. no. 26 and test compd. no. 1.

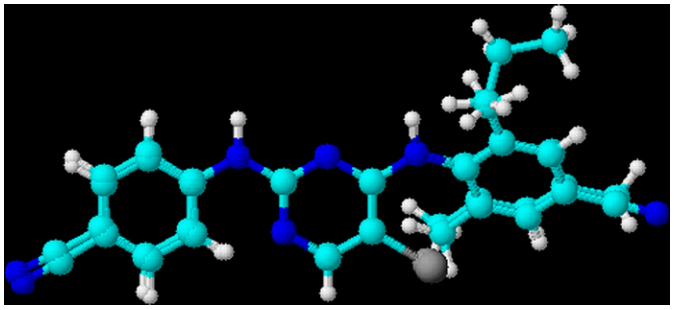

Fig. 4. Molecule to molecule docking of compd. no. 26 and test compd. no. 2.

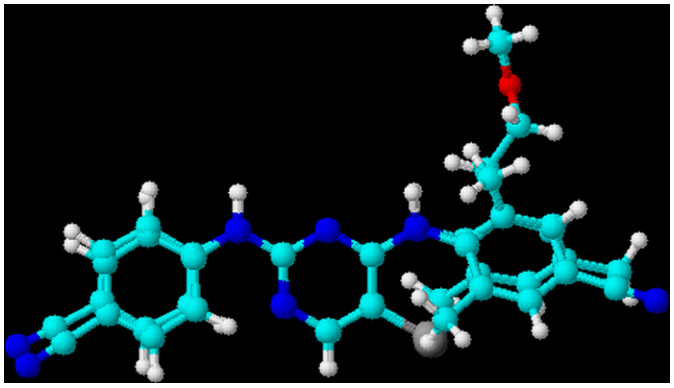

Fig. 5. Molecule to molecule docking of compd. no. 26 and test compd. no. 3.

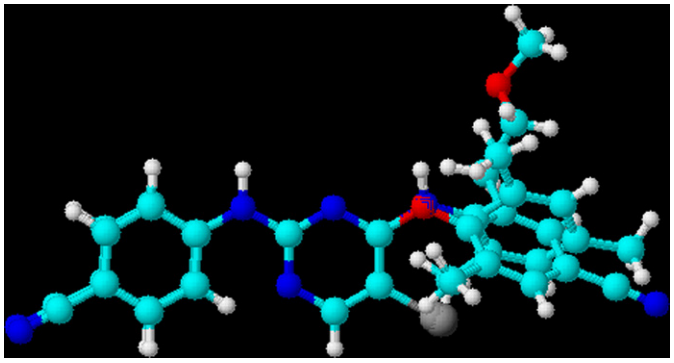

Fig. 6. Molecule to molecule docking of compd. no. 26 and test compd. no. 4.

Table 6
Distance gradient in molecule to molecule docking for DAPY derivatives

Docking case no.	Distance gradient
Case 1	0.77958
Case 2	0.12209
Case 3	0.10179
Case 4	0.12833

$$\log\frac{1}{C} = 7.92857 \times 10^{-6}\left(\pm 1.61572 \times 10^{-5}\right) \mathrm{TE} - 0.0032$$
$$n = 4,\ \mathrm{Se} = 0.0015,\ R = 0.3275,\ F = 0.241 \qquad (16)$$

$$\log\frac{1}{C} = 0.0029(\pm 0.0017)\ \mathrm{RMSg} + 0.0029$$
$$n = 4,\ \mathrm{Se} = 0.0010,\ R = 0.767,\ F = 2.857 \qquad (17)$$

These Eqs. (15)–(17) exhibit that the molecules are not in their minimum comformers energy state while binding with the receptor but in case of dipole moment the increase in the value of DpM reduces the inhibition activity for these molecules also. Similar effect is shown by the RMSg (parameter responsible for flexibility) that the high flexibility in molecule may have negative effect on the activity of this type of compounds.

For further information about the activity of these new compounds we dock these compounds over the most active compd. no. 26 and the docking figures are shown in Figs. 3–6. All these figures conform the similar activity of these new compounds. Figs. 3–6 also show that the test compd. no. 1 occupied the little less area over the active compd. no. 26 than that of the three other test compounds and it is also shown in terms of distance gradient in Table 6. But in all the cases of molecule to molecule docking, spacer is coming out from the plane of binding and this confirms the hypothesis for the binding with NRTs and exhibits the inactive nature of these spacers in case of binding with reverse transcriptase.

4. Conclusion

From the results and the discussion above, we conclude that the distance-based topological indices can be used successfully in combination with physicochemical properties for modelling the anti-HIV activities of DAPY compounds and that, for the present set of DAPY compounds, the branching is found to be the prominent one to affect the biological activity studied. The results also indicate that a combination of topological indices and molecular (3D) modelling can be used for understanding the structural behavior and selecting the compound with potential activity. We can also conclude that the test compd. no. 3 is found closer to the required properties for the particular biological activity and this compound may show good potential against HIV RTs.

References

[1] Rajni Garg, Satya P. Gupta, Hua Gao, Mekapati Suresh Babu, Asim Kumar Debnath, Corwin Hansch, Chem. Rev. 99 (1999) 3525–3601.
[2] D.W. Ludovici, Bart L. De Corte, Michael J. Kukla, Hong Ye, Chih Y. Ho, Bioorg. Med. Chem. 11 (2001) 2235–2239.

[3] Kalyan Das, A.D. Clark Jr., Paul J. Lewi, Jan Heeres, M.R. De Jonge, Eddy Arnold, J. Med. Chem. 47 (2004) 2550–2560.
[4] Abhilash Thakur, Mamta Thakur, Bioinformatics India 1 (1) (2003) 63–69.
[5] Abhilash Thakur, B.K. Tiwari, Mamta Thakur, N.D. Pandey, S.S. Narvi, Suprajnya Thakur, ARKIVOC 2 (2006) 213–225.
[6] H. Wiener, J. Am. Chem. Soc. 69 (1947) 17–21.
[7] M. Randic, J. Am. Chem. Soc. 97 (1975) 6609–6617.
[8] A.T. Balaban, Chem. Phys. Lett. 89 (1982) 399–412.
[9] P.V. Khadikar, N.V. Deshpande, P.P. Kale, A. Dobrynin, I. Gutman, G. Domotor, J. Chem. Inf. Comput. Sci. 35 (1995) 547–556.
[10] L.B. Kier, L.H. Hall, Molecular Connectivity in Chemistry and Drug Research, Academic Press, New York, 1976.
[11] ACD-Lab software for calculating the referred physicochemical parameters, Chemsketh S.O., <www.Acdlabs.com>.
[12] HyperChem 7 Demoversion from <www.Hyperchem.com>.
[13] S. Chaterjee, A.S. Hadi, B. Price, Regression Analysis by Examples, third ed. Wiley, New York, 2000.
[14] Molecular modelling pro (demo-version) from <http://www.chemsw.com/>.
[15] V.K. Agrawal, R. Sohgaura, P.V. Khadikar, A. Phadnis, Bioorg. Med. Chem. 9 (12) (2001) 3295–3299.
[16] P.V. Khadikar, K.C. Mathuhr, S. Singh, A. Phadnis, A. Shrivastava, M. Mandloi, Bioorg. Med. Chem. 10 (6) (2002) 1761–1766.
[17] P.V. Khadikar, S. Karmarker, S. Singh, A. Shrivastava, Bioorg. Med. Chem. 10 (10) (2002) 3163–3170.
[18] A. Thakur, A. Agrawal, M. Thakur, S. Thakur, Bioinformatics India 2 (3) (2004) 17–23.
[19] Abhilash Thakur, Mamta Thakur, P.V. Khadikar, Bioorg. Med. Chem. 11 (23) (2003) 5203–5207.