Quantitative Structure–Activity Relationships (QSARs) of Pyrimidine Nucleosides as HIV-1 Antiviral Agents

Massoud Mahmoudian[1]

Received August 18, 1989; accepted July 3, 1990

The structural requirements for the antiviral activity of pyrimidine nucleosides against HIV-1 virus was evaluated with the Hansch SAR analysis. Antiviral activity is best related to the hydrophobicity and steric (*L* and *B*3) properties of the substituent at the C5 of pyrimidine ring. Further, the antiviral activity is related to *B*4 of the substituent at position 3′ of the sugar ring with a positive slope. The activity of both uracil and cytosine derivatives can be related to their structure by the same equations, which indicates that the SARs are similar in these two groups of congeners. These results suggest that compounds with a small substituent at the 5 position of the pyrimidine ring and a flat substituent at the 3′ position of the sugar ring will be the most active compounds against HIV-1 virus.

KEY WORDS: antiviral; pyrimidine nucleosides; quantitative structure–activity relationships (QSAR); HIV-1.

INTRODUCTION

The finding that 3′-azido-3′-deoxy thymidine (zidovudine) is an effective anti HIV-1 agent (1) has resulted in its approval for the treatment of patients with AIDS (2). However, its high incidence of hematological side effects (3) calls for the design of better drugs. Recently Chu *et al.* (4,5) have synthesized several pyrimidine nucleosides and evaluated their antiviral activity against HIV-1 virus. The present study was carried out to investigate the structural requirements for antiviral activity of these compounds using Hansch analysis (6) to provide some guidelines for the design of more effective compounds.

METHODS

Quantitative structure–activity relationships (QSAR) were examined by computerized multiple regression analysis (6). The numerical values of the substituent parameters of various pyrimidine nucleosides (Table I) substituted at the C5 position of the pyrimidine ring or the C3′ position of the sugar ring were correlated with their potency against HIV-1 virus (4,5). It should be mentioned that both groups (4,5) have applied a similar method for estimating antiviral activity of the pyrimidine nucleosides using a culture of human peripheral blood mononuclear cells. Therefore, it is justifiable to pool their data. The following physiochemical parameters were used in this study:

(a) Verloop's steric parameters, various widths (*B*1–*B*4), and length (*L*) of individual substituents (7); and
(b) aromatic substituent constants of individual substituents at the C5 position of the pyrimidine ring (8). The values of substituent constants for each compound are shown in Table II and their intervariable correlation matrix is reported in Table III.

The validity of regression was judged by (i) standard deviation of estimate, SD; (ii) *F*-test value; (iii) level of significance, *P*; (iv) and correlation coefficient, *r*. The estimated standard error of each coefficient is shown in parentheses following its value.

RESULTS AND DISCUSSION

The application of Hansch analysis to relate the structure of pyrimidine nucleosides to their activity as anti-HIV-1 agents showed that various equations could be found which relate the biological activity to the physiochemical parameters of substituents at different places. The most significant equations are presented here.

Effect of Substitution on the Uracil Ring on the Antiviral Activity

The potency for anti-HIV-1 activity of pyrimidine nucleosides [expressed as $-\log(EC_{50})$] when a N3 group is present at the 3′ position of the sugar ring (compounds 1–6, 9, 10, 13–16) is correlated best with the P_i and length of the substituent at the C5 position of the uracil ring as described by Eq. (1).

$$-\log(EC_{50}) = 7.5433(+0.633) + 1.1701(+0.421)P_i - 0.1723(+0.048)\,L^2 \quad (1)$$

$n = 12$, $r = 0.772$, SD = 0.82, $F(2,9) = 6.63$, $P < 0.05$

Fifty-nine and six-tenths percent of the variance in the biological activity [$-\log(EC_{50})$] is explained by Eq. (1).

Equation (1) suggests that the correlation between the antiviral activity and the length (*L*) of the substituent at R5 is curvilinear (Fig. 1).

From this equation, it can be deduced that the C-5 substituent must be a small and hydrophobic group. Therefore, compounds such as 11 and 12 (Table I), which have a highly hydrophilic group [$NHCH_3$ and $N(CH_3)_2$, respectively], at this position would have a rather low activity.

Effect of Substitution at the 3′ Position of the Sugar Ring on the Antiviral Activity

The antiviral activity of pyrimidine nucleosides when R5 = H (compounds 1, 19, 24, and 27) correlates best with the *B*4 of the substituent at the C3′ position of the sugar ring as described by Eq. (2) and shown in Fig. 2.

$$-\log(EC_{50}) = 3.0329(+0.297) + 0.8199(+0.115)\,B4 \quad (2)$$

$n = 4$, $r = 0.98$, SD = 0.269, $F(1,2) = 50.6$, $P < 0.025$

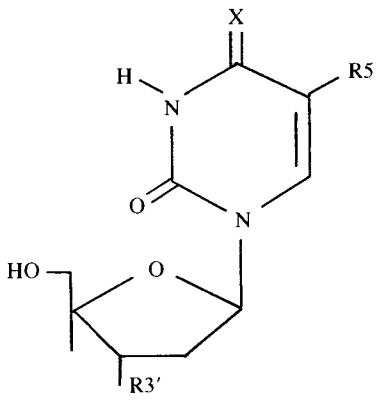
Ninety-six and two-tenths percent of the variance in the biological activity [$-\log(EC_{50}$] is explained by Eq. (2).

However, no significant correlation was found between the biological activity and the *L*, *B*1, MR, HA, and HD pa-

[1] Department of Pharmacology, University of Medical Sciences of Iran Firoozgar Institute, Ali Valady Street, Tehran 15934, Iran.

0724-8741/91/0100-0043$06.50/0

Table I. The Structure and Biological Activity of Pyrimidine Nucleosides Used in this Analysis[a]



No.	Compound	R-5	R-3'	X	−log (EC$_{50}$)	Ref. No.
1	CS-87	H	N3	O	6.50	4
2	Zidovudine	CH_3	N3	O	8.26	4
3	CS-85	C_2H_5	N3	O	6.28	4
4		C_3H_7	N3	O	4.20	4
5		Br	N3	O	5.98	4
6		I	N3	O	5.94	4
7		CH=CHBr	N3	O	<4	4
8		CF_3	N3	O	<4	5
9		F	N3	O	5.32	5
10		NH_2	N3	O	5.21	5
11		$NHCH_3$	N3	O	<4	5
12		$N(CH_3)_2$	N3	O	<4	5
13		OH	N3	O	5	5
14		OCH_3	N3	O	4.15	5
15		OC_2H_5	N3	O	4.27	5
16		SCN	N3	O	5.29	5
17		SCH_3	N3	O	<4	5
18		F	N3	NH	6.0	5
19		H	N3	NH	6.03	4
20		CH_3	N3	NH	6.82	4
21		H	NH2	O	4.22	4
22		CH_3	NH2	O	<4	4
23		C_2H_5	NH2	O	4.26	4
24		H	I	O	4.92	4
25		CH_3	I	O	4.33	4
26		C_2H_5	I	O	4.07	4
27		H	H	O	4.01	4
28		CH_3	H	O	6.77	4
29		C_2H_5	H	O	5.31	4

[a] The −log(EC$_{50}$), median antiviral concentration against HIV-1, is taken from Refs. 4 and 5.

rameters of the substituent at the 3′ position of the sugar ring. Therefore, R-3′ should be a flat group, and the wider the substituent, the higher the activity. Hence, among the tested compounds (4,5), the derivatives with an azido group at 3′ are the most active ones.

The Whole Set of Data

The biological activity of uracil derivatives in Table I correlates best with substitution parameters as described by Eqs. (3) and (4).

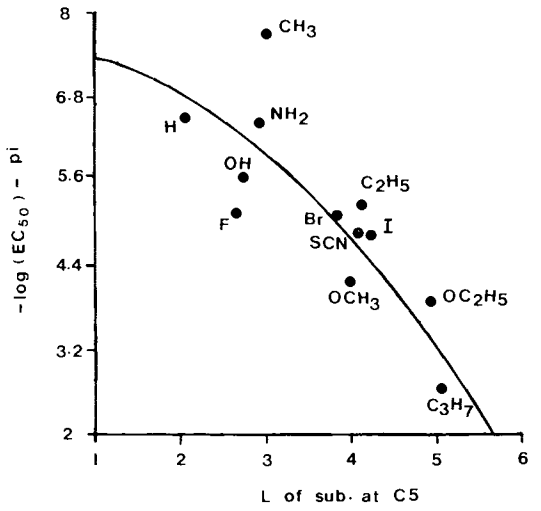


Fig. 1. The correlation between the length (L) of the substituent at C-5 of the uracil ring and the potency of the pyrimidine nucleosides as anti-HIV-1 agents. R-3′ is N_3 in all compounds. The line is drawn according to Eq. (1).

$$-\log(EC_{50}) = 1.9192(+1.046) + 2.4005(+0.986)\,\mathcal{R}(C5) - 0.1731(+0.047)L^2\,(C5) + 2.6132(+0.786)\,B3(C5) + 0.4854(+0.169)\,B4(C3') \quad (3)$$

$n = 20$, $r = 0.746$, SD = 0.85, $F(4,15) = 4.70$, $P < 0.05$

Fifty-five and six-tenths percent of the variance in the biological activity [−log(EC$_{50}$)] is explained by Eq. (3).

$$-\log(EC_{50}) = 2.7371(+1.015) + 1.0271(+0.431)\,P_i(C5) - 0.2123(+0.054)\,L^2\,(C5) + 1.8907(+0.756)B3(C5) + 0.4889\,(+0.172)\,B4(C3') \quad (4)$$

$n = 20$, $r = 0.742$, SD = 0.85, $F(4,15) = 4.60$, $P < 0.05$

Fifty-five and one-tenth percent of the variance in the biological activity [−log(EC$_{50}$)] is explained by Eq. (4).

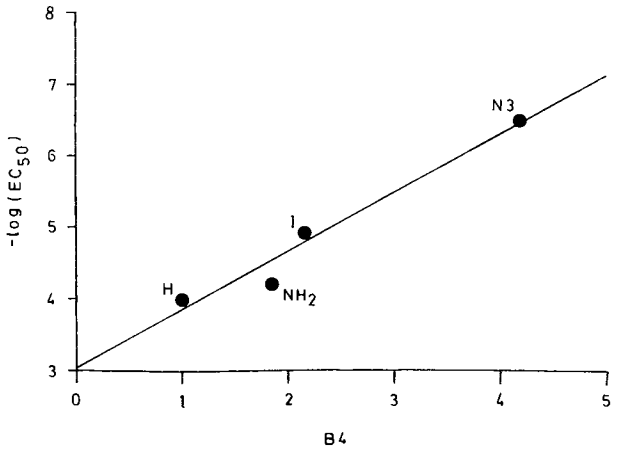


Fig. 2. The correlation between the $B4$ of the substituent at C-3′ of the sugar ring and the potency of pyrimidine nucleosides as anti-HIV-1 agents. R-5 is H in all compounds. The line is drawn according to Eq. (2).

Table II. Substituent Constants for C5- and C3′-Substituted Pyrimidine Nucleosides[a]

Compound No.	P_i(C5)	$\mathcal{F}$(C5)	$\mathcal{R}$(C5)	MR(C5)	L(C5)	$B1$(C5)	$B2$(C5)	$B3$(C5)	$B4$(C5)	L^2(C5)	$B4$(C3′)
1	0.00	0.00	0.00	1.03	2.06	1.00	1.00	1.00	1.00	4.24	4.18
2	0.56	−0.04	−0.13	5.65	3.00	1.52	1.90	1.90	2.04	9.00	4.18
3	1.02	−0.05	−0.10	10.30	4.11	1.52	1.90	1.90	2.97	16.89	4.18
4	1.55	−0.06	−0.08	14.96	5.05	1.52	1.90	1.90	3.49	25.50	4.18
5	0.86	0.44	−0.17	8.88	3.83	1.95	1.95	1.95	1.95	14.67	4.18
6	1.12	0.40	−0.19	13.94	4.23	2.15	2.15	2.15	2.15	17.89	4.18
9	0.14	0.43	−0.34	0.92	2.65	1.35	1.35	1.35	1.35	7.02	4.18
10	−1.23	0.02	−0.68	5.42	2.93	1.50	1.50	1.84	1.84	8.58	4.18
13	−0.67	0.29	−0.64	2.85	2.74	1.35	1.35	1.35	1.93	7.51	4.18
14	−0.02	0.26	−0.51	7.87	3.98	1.35	1.90	1.90	2.87	15.84	4.18
15	0.38	0.22	−0.44	12.47	4.92	1.35	1.90	1.90	3.36	24.21	4.18
16	0.41	0.36	0.19	13.40	4.08	1.70	1.70	1.70	4.45	16.65	4.18
18	0.14	0.43	−0.34	0.92	2.65	1.35	1.35	1.35	1.35	7.02	4.18
19	0.00	0.00	0.00	1.03	2.06	1.00	1.00	1.00	1.00	4.24	4.18
20	0.56	−0.04	−0.13	5.65	3.00	1.52	1.90	1.90	2.04	9.00	4.18
21	0.00	0.00	0.00	1.03	2.06	1.00	1.00	1.00	1.00	4.24	1.84
23	1.02	−0.05	−0.10	10.30	4.11	1.52	1.90	1.90	2.97	16.86	1.84
24	0.00	0.00	0.00	1.03	2.06	1.00	1.00	1.00	1.00	4.24	2.15
25	0.56	−0.04	−0.13	5.65	3.00	1.52	1.90	1.90	2.04	9.00	2.15
26	1.02	−0.05	−0.10	10.30	4.11	1.52	1.90	1.90	2.97	16.86	2.15
27	0.00	0.00	0.00	1.03	2.06	1.00	1.00	1.00	1.00	4.24	1.00
28	0.56	−0.04	−0.13	5.65	3.00	1.52	1.90	1.90	2.04	9.00	1.00
29	1.02	−0.05	−0.10	10.30	4.11	1.52	1.90	1.90	2.97	16.89	1.00

[a] The values of constants are taken from Refs. 7 and 8. For the structure and biological activity see Table I.

The addition of the cytosine derivatives (compounds 18–20) to the above set of data results in a better correlation as shown by Eqs. (5) and (6) and Fig. 3.

$$-\log(EC_{50}) = 2.0954(+0.901) + 2.3094(+0.864)\,\mathcal{R}\,(C5) - 0.1711(+0.040)\,L^2\,(C5) + 2.4723(+0.663)B3(C5) + 0.4968\,(+0.144)\,B4(C3') \quad (5)$$

$n = 23$, $r = 0.774$, SD = 0.78, $F(4,18) = 6.72$, $P < 0.01$

Fifty-nine and nine-tenths percent of the variance in the biological activity $[-\log(EC_{50})]$ is explained by Eq. (5).

$$-\log(EC_{50}) = 2.8660(+0.869) + 1.0247(+0.377)\,P_i(C5) - 0.2090(+0.045)\,L^2\,(C5) + 1.7851(+0.636)B3(C5) + 0.4913\,(+0.142)\,B4(C3') \quad (6)$$

$n = 23$, $r = 0.777$, SD = 0.78, $F(4,18) = 6.84$, $P < 0.01$

Sixty and three-tenths percent of the variance in the biological activity $[-\log(EC_{50})]$ is explained by Eq. (6).

Considering the fact that there is a small intervariable correlation between p_i and $\mathcal{R}$ (Table III), it is surprising that $\mathcal{R}$ could be replaced by p_i in Eqs. (4) and (6). It is possible that the biological activity might be related to both parameters. However, due to the limited number of analogues available for the analysis, it is unjustifiable to include more than four parameters in one equation. One needs more than 25 analogues to include both P_i and $\mathcal{R}$ in the same equation.

Compounds 7, 8, 11, 12, and 17 were not included in the regression analysis because of uncertainty regarding the value of their biological activity. The correlation between the observed values of biological activity and those calculated from Eq. (6) is shown in Fig. 3.

Table III. Intervariable Correlation Matrix for the Substituent Constants Shown in Table II

	−logEC	P_i(C5)	$\mathcal{F}$(C5)	$\mathcal{R}$(C5)	MR(C5)	L(C5)	$B1$(C5)	$B2$(C5)	$B3$(C5)	$B4$(C5)	L^2(C5)	$B4$(C3′)
−logEc	1.000											
P_i(C5)	0.040	1.000										
$\mathcal{F}$(C5)	0.007	−0.145	1.000									
$\mathcal{R}$(C5)	0.112	0.510	−0.351	1.000								
MR(C5)	−0.155	0.667	0.052	0.064	1.000							
L(C5)	−0.223	0.646	0.120	−0.100	0.958	1.000						
$B1$(C5)	0.210	0.508	0.373	−0.113	0.730	0.668	1.000					
$B2$(C5)	0.111	0.641	0.054	−0.134	0.821	0.829	0.844	1.000				
$B3$(C5)	0.105	0.537	0.037	−0.223	0.809	0.810	0.850	0.984	1.000			
$B4$(C5)	−0.207	0.502	0.043	0.033	0.891	0.890	0.544	0.719	0.703	1.000		
L^2(C5)	−0.268	0.649	0.095	−0.067	0.949	0.992	0.599	0.769	0.747	0.871	1.000	
$B4$(C3′)	0.339	−0.152	0.506	−0.363	0.134	0.177	0.238	0.100	0.126	0.126	0.175	1.000

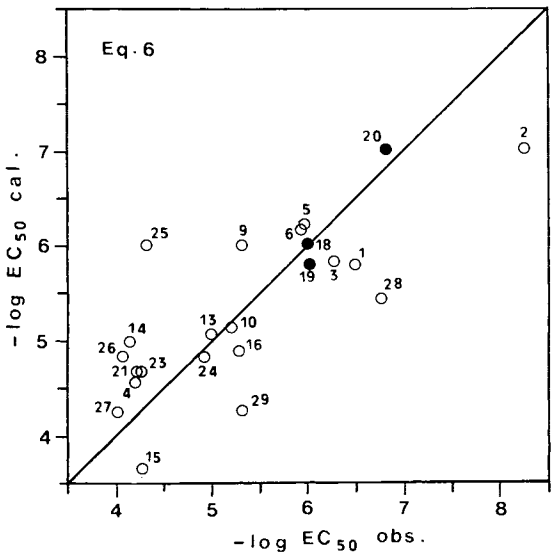


Fig. 3. The correlation between the observed potency of various pyrimidine nucleosides as anti-HIV-1 agents and the calculated values which have been obtained from Eq. (6). The overall regression coefficient (r) is equal to 0.777. The numbers refer to the analogues as shown in Table I. (○) Uracil and (●) cytosine derivatives.

From Eq. (6), the potency of the bromovinyl analogue of zidovudine (compound 7, Table I) is calculated to be equal to 3.22, which is in agreement with its low activity (4). However, the potencies of compounds 8, 17, and 22 according to this equation are 7.9, 5.07, and 5.78, which are rather high compared with their reported activities [Table I (4,5)]. This discrepancy may be due to either a different pattern of metabolism of these compounds (i.e., a higher rate of metabolic inactivation or a lower rate of activation via phosphorylation) or inaccuracy in the determination of their EC_{50}'s.

The good correlation of biological activity with the hydrophobicity and steric properties of the C5 substituents indicates that the interaction of this part of molecule with its receptor site is mostly governed by steric factors. The receptor site for the binding of pyrimidine moiety must be very selective since only a small group at C5 is suitable for maximum activity. Also, the C nucleosides (where the sugar is attached to C5 instead of N1) are not active against HIV-1 (4). However, the replacement of the O of the uracil ring with the NH of cytosine does not affect the structure activity relationships [compare Eqs. (3) and (4) with Eqs. (5) and (6)]. This may result from the possibility that both groups have the capability to form hydrogen binding with the receptor site.

The presence of a hydroxy group at C-5′ is also necessary for the anti HIV-1 activity (4), because such compounds cannot be phosphorylated to active metabolites (4). Further various processes may contribute to "biological activity," e.g., the activation of antiviral compounds to phosphorylated species, binding of the phosphorylated active metabolites to the target enzyme, and finally, inactivation of these compounds by catabolic metabolism. Therefore the structure activity relationships may be affected by any or all of these processes.

In summary, it can be concluded that for maximum anti-HIV-1 activity, a molecule must possess a small group at C-5, and a rather flat group at C-3′, and a OH at C-5′ must be present in the molecule of 2′,3′-dideoxy pyrimidine nucleosides.

REFERENCES

1. H. Mitsuya, K. J. Weinhold, P. A. Furman, M. H. St. Clair, S. N. Lehrman, R. L. Gallo, D. Bolognesi, D. W. Barry, and S. Broder. 3′-Azido, 3′-deoxythymidine (BW A509U): An antiviral agent that inhibits the infectivity and cytopathic effects of human T cell lymphotropic virus III/lymphadenophath-associated virus in-vitro. *Proc. Natl. Acad. Sci. USA* 82:7096–7100 (1985).
2. P. A. Furman and D. W. Barry. Spectrum of antiviral activity and mechanism of action of zidovudine. *Am. J. Med.* 85:176–181 (1988).
3. D. D. Richman. Results of continued monitoring of participants in the placebo-controlled trial of zidovudine for serious human immunodeficiency virus infection. *Am. J. Med.* 85:208–213 (1988).
4. K. C. Chu, R. F. Schinazi, M. K. Ahn, G. V. Ullas, and Z. P. Gu. Structure-activity relationships of pyrimidine nucleosides as antiviral agents for human immunodeficiency virus type I in peripheral blood mononuclear cells. *J. Med. Chem.* 32:612–617 (1989).
5. T.-S. Lin, J.-Y. Guo, R. F. Schinazi, C. K. Chu, J.-N. Xiang, and W. H. Prusoff. Synthesis and antiviral activity of various 3′-azido analogues of pyrimidine deoxyribonucleosides against human immunodeficiency virus (HIV-1, HTLV-III/LAV). *J. Med. Chem.* 31:336–340 (1988).
6. C. Hansch. In C. J. Cavallito (ed.). *Structure Activity Relationships*, Pergamon Press, Oxford, 1973, pp. 75–165.
7. A. Verloop, W. Hoogenstraaten, and J. Tipker. Development and application of new steric substituent parameters in drug design. In E. J. Ariens (ed.), *Drug Design, Vol. 7*, Academic Press, New York, 1976, pp. 165–207.
8. S. H. Unger. Consequences of the Hansch paradigma for the pharmaceutical industry. In E. J. Ariens (eds.), *Drug Design, Vol. 9*, Academic Press, New York, 1980, pp. 48–119.