Journal of Computer-Aided Molecular Design, 10 (1996) 145–152
ESCOM

J-CAMD 325

Structure–activity correlation study of HIV-1 inhibitors: Electronic and molecular parameters

Supa Hannongbua*, Luckhana Lawtrakul and Jumras Limtrakul

Department of Chemistry, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand

Received 14 August 1995
Accepted 6 December 1995

Keywords: Semiempirical AM1 method; Atomic net charges; Hydration energy; Molar refractivity; QSAR; HEPT

Summary

Quantitative structure–activity relationships (QSARs) for 40 HIV-1 inhibitors, 1-[(2-hydroxyethoxy)-methyl]-6-(phenylthio)thymine and its derivatives, were studied. Fully optimized geometries, based on the semiempirical AM1 method, were used to calculate electronic and molecular properties of all compounds. In order to examine the relation between biological activities and structural properties, multiple linear regression models were employed. A suitable QSAR model was obtained, showing not only statistical significance, but also predictive ability. The significant molecular descriptors used were atomic charges of two substituted carbon atoms in the thymine ring, hydration energies and molar refractivities of the molecules. These descriptors allowed a physical explanation of electronic and molecular properties contributing to HIV-1 inhibitory potency.

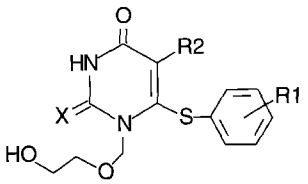
Introduction

During the past decade, compounds with a wide variety of structures have been synthesized and examined for the treatment of acquired immunodeficiency syndrome (AIDS), which is caused by the human immunodeficiency virus (HIV) [1–6]. One of the most potent and selective drugs displaying in vitro activity against HIV-1 reverse transcriptase (HIV-1 RT) is 1-[(2-hydroxyethoxy)-methyl]-6-(phenylthio)thymine (HEPT), discovered by Tanaka and co-workers [7,8]. Other animal retrovirus-associated RTs and even HIV-2 RT are totally unaffected by this compound. Due to its high specificity, this inhibitor is a promising candidate for the treatment of AIDS. Some successful 3D QSAR approaches in predicting affinity of designed HIV-1 inhibitors have been reported [9–12]. However, based on experimental investigations, the mechanism of action of HEPT is not yet properly understood. The electronic and steric character of molecules play an important role in drug–receptor interactions. As the electronic character relates directly to the electron distribution of interacting molecules at the active site, while steric effects deal with molecular geometry in the same manner, a correlation study based on electronic and molecular properties is expected to provide insight into the anti-HIV-1 mechanism of HEPT.

As is well known, electronic effects of substituents can influence the amount of interacting species available to the receptor as well as the strength of the drug–receptor interaction. Electronic variables such as π and σ have been used in the derivation of classical QSARs. However, it is difficult to guess model parameters for QSAR when the chemical mechanism of the drug–receptor molecule interaction is not known and hence, the type of model applicable to this interaction is not easily defined. In principle, quantum-chemical calculations can provide much help to overcome these difficulties. As low-priced computing time has become easily accessible nowadays, molecular orbital calculations at the semiempirical level have become very economical, even for series of larger drug molecules, when compared to the cost of synthesis and experimental drug testing. Thus, such calculations not only give the possibility to study the relationship between drug structure and activity at the electronic level, but also allow the design of new candidates for potent drugs of the same type with higher chances of success.

Atomic net charges obtained from quantum-chemical calculations have been reported as the suitable electronic

*To whom correspondence should be addressed.



Scheme 1. General structure of HEPT derivatives.

descriptor in QSAR studies and were found useful to indicate the active center of drugs [13–15]. However, our previous correlation study on HIV-1 inhibitors found that the use of only atomic net charges related to inhibitory activity was not sufficient to explain such correlations [16]. The quality of the investigated QSAR model was significantly improved by adding some molecular properties, such as molar refractivity and hydration energy, to the model. The contribution of molar refractivity in the interaction of biomacromolecules was emphasized by Pauling [17] and it was applied to biochemical structure–activity studies [18–21]. Another successful QSAR is the use of binding free energy between the HIV-1 inhibitor and the receptor, obtained from molecular dynamics simulations [22–24]. However, the computational effort needed to obtain each value of the free energy is excessive (ca. 100 h of CPU time on a CrayYMP for each novel inhibitor analog) [23]. Apart from the above approaches, relations between the experimental activity and lipophilicity in terms of logP (log of octanol/water partition coefficient) illustrate that logP is suitable to account for the hydrophobic interactions of a drug in the biological system [25–27].

The goal of this study is to develop QSARs for HIV-1 inhibitors in the class of HEPT and to determine if the obtained descriptors can help to understand the biological activity of the drugs in this category.

Method of calculation

HIV-1 inhibitors for QSAR study

Biological activities of 40 HEPT derivatives against the cytopathic effect of HIV-1 have been reported [7] as the effective concentration of the compound required to achieve 50% protection of MT-4 cells (EC_{50}, μg/ml). The potency has been defined as log(1/C) in the QSAR analysis, where C is the molar EC_{50} value of the compound, and is used as the dependent variable in the QSAR study (Table 1).

Calculation of electronic and molecular properties

Molecular geometries of all compounds were generated by the ALCHEMY III molecular modeling software package [28], running on a 486 PC. Molecular geometries were fully optimized and atomic net charges were calculated based on the semiempirical AM1 method, available in GAUSSIAN 92 [29] on an HP-715 workstation. The AM1 method has been demonstrated to represent well for the molecular geometries and hydrogen bond systems [30]. With the optimized geometry, molecular properties were determined by the ChemPlus 1.0 program [31]. Molar refractivities were calculated by a constrained least-squares technique using physicochemical parameters from 3D structure-directed QSAR [32]. Partition coefficients were calculated from the nonlinear regression model using parameters from molecular surface, volume, weight and charge densities on nitrogen and oxygen atoms of the molecules. This method was used to test the predictive power of the model by the accurate estimation of logP for complex molecules [33]. Semianalytical treatment, in which solvent is treated as a statistical continuum, was used to calculate hydration energies, providing both energies and derivatives analytically. This method yields molecular hydration energies with an accuracy comparable to those obtained from contemporary free energy perturbation methods [34].

Statistical method

In order to set up the QSAR equations, the following linear model was used:

$$\log(1/C) = \sum_{i=1}^{8}\sum_{j=1}^{8} P_i Q_j + \sum_{i=9}^{11}\sum_{j=1}^{3} P_i M_j + C \qquad (1)$$

where P_i are the fitting parameters, Q_j and M_j denote the electronic and molecular parameters, respectively, and C is the regression constant obtained from the fit.

Using 35 HEPT derivatives, multiple linear regression models were developed based on regression algorithms in the SPSS package [35] on a 486 PC. The quality of the model was considered as statistically satisfactory on the basis of multiple correlation coefficients (r), standard deviations (s) and F-statistics.

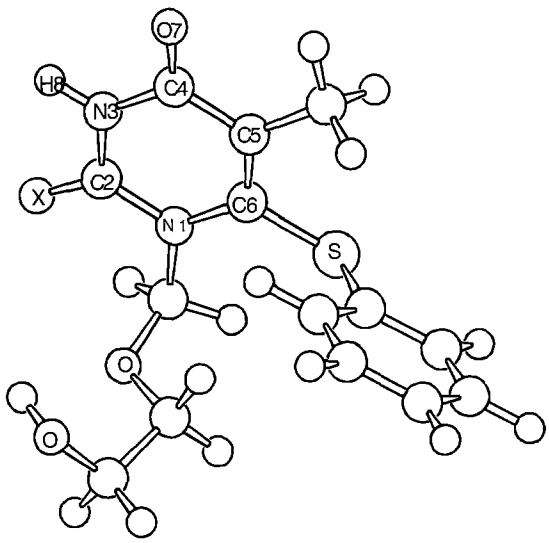


Fig. 1. Structure of 1-[(2-hydroxyethoxy)-methyl]-6-(phenylthio)thymine (HEPT). The atomic numbering as used in this study is indicated.

TABLE 1
STRUCTURE OF HEPT DERIVATIVES[a] AND EXPERIMENTAL AND CALCULATED log(1/C) AFFINITIES

No.	X	R1	R2	log(1/C)		Residual
				Experimental	Calculated[b]	
1	O	2-Me	Me	4.146	5.124	−0.978
2	O	2-Cl	Me	3.885	5.104	−1.219
3	O	2-NO_2	Me	3.853	3.857	−0.004
4	O	2-OMe	Me	4.720	5.186	−0.466
5	O	3-Et	Me	5.568	5.484	0.084
6	O	3-*t*Bu	Me	4.920	6.717	−1.797
7	O	3-CH_2OH	Me	3.534	4.023	−0.489
8	O	3-CF_3	Me	4.346	5.664	−1.318
9	O	3-F	Me	5.480	4.758	0.722
10	O	3-Cl	Me	4.885	4.978	−0.093
11	O	3-NO_2	Me	4.468	3.645	0.823
12	O	3-OMe	Me	4.657	4.842	−0.185
13	O	3,5-Me_2	Me	6.584	6.456	0.128
14	O	3,5-Cl_2	Me	5.885	5.459	0.426
15	S	3,5-Me_2	Me	6.656	6.442	0.214
16	O	3-COOMe	Me	5.102	4.769	0.333
17	O	3-COMe	Me	5.136	5.052	0.084
18	O	4-COMe	Me	3.958	4.788	−0.830
19	O	3-COOH	Me	3.453	3.566	−0.113
20	O	3-CN	Me	4.999	4.458	0.541
21	O	H	$CH_2CH{=}CH_2$	5.601	6.402	−0.801
22	O	H	COOMe	5.180	5.439	−0.259
23	O	H	CONHPh	4.744	4.360	0.384
24	S	H	Et	6.957	6.202	0.755
25	S	H	i-Pr	7.228	7.422	−0.194
26	S	3,5-Me_2	i-Pr	8.299	8.468	−0.169
27	S	3,5-Cl_2	Et	7.365	7.386	−0.021
28	O	H	Et	6.919	6.175	0.744
29	O	H	Pr	5.467	6.071	−0.604
30	O	H	i-Pr	7.199	7.007	0.192
31	O	3,5-Me_2	Et	7.885	7.360	0.525
32	O	3,5-Me_2	i-Pr	8.567	7.720	0.847
33	O	3,5-Cl_2	Et	7.852	6.603	1.249
34[c]	O	H	Me	5.154	4.527	0.627
35	S	H	Me	6.008	5.183	0.825
36[d]	S	3,5-Me_2	Et	8.108		
37[d]	O	3-Me	Me	5.585		
38[d]	O	3-OH	Me	4.086		
39[d]	O	4-NO_2	Me	3.721		
40[d]	O	4-OH	Me	3.558		

[a] See Scheme 1.
[b] Calculated by Eq. 5.
[c] HEPT.
[d] Not included in the derivation of the equations.

Reliability of the models

An analysis of predictive ability was carried out in terms of both predictive r^2 [36] and actual prediction. Predictive r^2 (Q^2) is defined as: $Q^2 = (SSY - PRESS)/SSY$, where SSY is the sum of squares of the deviation between the affinities of the fitted set and their mean affinity and PRESS is the prediction error sum of squares obtained from the leave-one-out method. The standard deviation of prediction (S_{PRESS}) is also considered and defined as: $S_{PRESS} = [PRESS/(n-k-1)]^{1/2}$, where k is the number of variables in the model. In addition, five compounds, selected from various ranges of inhibitory affinity, were kept to test for the actual prediction of the model.

Results

Atomic net charges of various atoms in the thymine ring, i.e. N1, C2, N3, C4, C5, C6, O7 and H8 (Fig. 1), were used as electronic parameters. Molar refractivity (MR), partition coefficient (logP) and hydration energy (HE, kcal/mol) were employed as molecular parameters. The values of all variables are given in Table 2.

All possible combinations of parameters were considered rather than employing the automatic stepwise method. This procedure yielded a large amount of satisfactory statistical models. The quality of the models, as indicated by r, s, F, Q^2 and S_{PRESS}, was tested statistically.

TABLE 2
ATOMIC NET CHARGES AND MOLECULAR PROPERTIES FOR COMPOUNDS IN TABLE 1 AND THE PARAMETERS USED IN THE CORRELATION STUDY

No.	N1	C2	N3	C4	C5	C6	O7	H8	HE[a]	MR	logP
1	–0.3924	0.4908	–0.5095	0.4102	–0.1894	–0.0727	–0.3498	0.3416	–12.737	50.303	6.440
2	–0.3878	0.4933	–0.5095	0.4120	–0.1739	–0.0845	–0.3497	0.3419	–14.083	50.190	6.754
3	–0.3893	0.4829	–0.5071	0.4108	–0.1795	–0.0878	–0.3425	0.3444	–19.159	51.692	6.162
4	–0.3926	0.4839	–0.5068	0.4089	–0.1859	–0.0680	–0.3517	0.3404	–15.008	52.117	5.885
5	–0.3898	0.4849	–0.5066	0.4120	–0.1955	–0.0657	–0.3497	0.3424	–13.039	54.904	6.837
6	–0.3913	0.4855	–0.5061	0.4118	–0.1962	–0.0592	–0.3519	0.3417	–12.166	63.928	7.600
7	–0.3896	0.4846	–0.5068	0.4120	–0.1962	–0.0676	–0.3503	0.3421	–18.567	52.078	5.438
8	–0.3981	0.4898	–0.5042	0.4106	–0.1867	–0.0611	–0.3448	0.3440	–13.329	51.236	6.856
9	–0.3946	0.4896	–0.5046	0.4120	–0.1918	–0.0636	–0.3466	0.3442	–13.584	45.601	6.376
10	–0.3915	0.4855	–0.5060	0.4114	–0.1929	–0.0647	–0.3494	0.3428	–14.173	50.190	6.754
11	–0.3976	0.4937	–0.5078	0.4124	–0.1942	–0.0802	–0.3476	0.3452	–18.303	51.692	6.162
12	–0.3912	0.4854	–0.5063	0.4117	–0.1948	–0.0611	–0.3516	0.3418	–16.027	52.117	5.885
13	–0.3972	0.4879	–0.5045	0.4133	–0.1911	–0.0540	–0.3529	0.3424	–11.423	54.912	6.754
14	–0.3920	0.4863	–0.5055	0.4113	–0.1885	–0.0689	–0.3468	0.3437	–13.887	54.684	7.383
15	–0.3502	0.1052	–0.4680	0.4016	–0.1808	–0.0796	–0.3449	0.3478	–12.432	62.400	5.470
16	–0.3914	0.4853	–0.5060	0.4113	–0.1932	–0.0654	–0.3490	0.3430	–16.849	54.062	7.466
17	–0.3922	0.4859	–0.5057	0.4114	–0.1927	–0.0641	–0.3485	0.3432	–15.041	52.479	7.084
18	–0.3894	0.4847	–0.5064	0.4113	–0.1922	–0.0710	–0.3467	0.3437	–15.149	52.479	7.804
19	–0.3900	0.4861	–0.5057	0.4132	–0.1877	–0.0724	–0.3470	0.3436	–20.461	49.024	7.533
20	–0.3928	0.4928	–0.5027	0.4112	–0.1868	–0.0702	–0.3427	0.3459	–17.742	51.457	5.938
21	–0.3968	0.4889	–0.5045	0.4157	–0.1860	–0.0455	–0.3515	0.3415	–11.023	47.476	6.877
22	–0.3908	0.4942	–0.5015	0.4367	–0.2436	0.0174	–0.3299	0.3492	–15.727	49.454	7.152
23	–0.3654	0.5126	–0.5055	0.4253	–0.2847	0.0424	–0.3156	0.3533	–16.748	49.908	8.685
24	–0.3535	0.1055	–0.4685	0.4005	–0.1800	–0.0691	–0.3453	0.3471	–13.699	57.785	5.238
25	–0.3551	0.1045	–0.4695	0.4050	–0.1614	–0.0732	–0.3459	0.3471	–13.183	62.334	5.569
26	–0.3538	0.1069	–0.4694	0.4051	–0.1716	–0.0678	–0.3484	0.3468	–10.936	71.550	6.197
27	–0.3521	0.1178	–0.4645	0.3998	–0.1701	–0.0798	–0.3378	0.3510	–12.165	66.774	6.495
28	–0.3908	0.4914	–0.5036	0.4135	–0.1984	–0.0466	–0.3531	0.3423	–10.417	50.296	6.523
29	–0.3933	0.4842	–0.5070	0.4114	–0.1869	–0.0590	–0.3497	0.3416	–13.367	54.897	6.919
30	–0.3997	0.4865	–0.5071	0.4170	–0.1804	–0.0489	–0.3555	0.3412	–12.077	54.845	6.853
31	–0.3986	0.4867	–0.5055	0.4126	–0.1925	–0.0435	–0.3545	0.3411	–10.746	59.513	7.151
32	–0.3987	0.4871	–0.5068	0.4168	–0.1781	–0.0496	–0.3533	0.3418	–11.465	60.347	7.501
33	–0.3936	0.4920	–0.5039	0.4102	–0.1801	–0.0697	–0.3429	0.3453	–12.238	59.285	7.779
34	–0.3958	0.4914	–0.5091	0.4104	–0.1793	–0.0846	–0.3483	0.3425	–13.609	45.695	6.126
35	–0.3537	0.1031	–0.4704	0.4009	–0.1688	–0.0994	–0.3425	0.3468	–13.376	53.184	4.842
36	–0.3532	0.1061	–0.4678	0.4008	–0.1756	–0.0701	–0.3451	0.3472	–11.739	67.001	5.866
37	–0.3960	0.4876	–0.5045	0.4106	–0.2013	–0.0507	–0.3516	0.3430	–12.804	50.303	6.440
38	–0.3969	0.4887	–0.5041	0.4129	–0.1888	–0.0568	–0.3508	0.3429	–18.017	47.079	5.952
39	–0.3894	0.4852	–0.5058	0.4108	–0.1860	–0.0794	–0.3411	0.3459	–20.399	51.692	6.162
40	–0.3955	0.4892	–0.5045	0.4134	–0.1937	–0.0532	–0.3516	0.3422	–20.768	47.079	5.952

[a] In kcal/mol.

The best of the one-term, two-term, three-term, four-term, five-term, six-term and seven-term equations, with regard to the relative importance of the various statistical criteria, were found to be Eqs. 2–8, respectively. The use of more than seven independent variables for this data set, however, is not satisfying because of possible chance correlations [37]. Consequently, they were omitted from our model.

$$\log(1/C) = 0.402(\pm 0.134)HE + 11.292(\pm 1.914)$$

$$(n = 35,\ r = 0.729,\ s = 0.980,\ F = 37.473,\ Q^2 = 0.484,\ S_{PRESS} = 1.029) \quad (2)$$

$$\log(1/C) = 0.297(\pm 0.123)HE + 0.105(\pm 0.054)MR + 4.085(\pm 4.015)$$

$$(n = 35,\ r = 0.829,\ s = 0.814,\ F = 35.098,\ Q^2 = 0.633,\ S_{PRESS} = 0.881) \quad (3)$$

$$\log(1/C) = 0.326(\pm 0.127)HE + 0.090(\pm 0.057)MR + 76.020(\pm 101.673)H8 - 20.833(\pm 33.561)$$

$$(n = 35,\ r = 0.842,\ s = 0.798,\ F = 25.143,\ Q^2 = 0.637,\ S_{PRESS} = 0.891) \quad (4)$$

$$\log(1/C) = 0.198(\pm 0.134)HE + 0.093(\pm 0.052)MR + 45.797(\pm 35.602)C5 + 38.482(\pm 27.062)C6 + 14.418(\pm 8.835)$$

$$(n = 35,\ r = 0.869,\ s = 0.743,\ F = 23.211,\ Q^2 = 0.665,\ S_{PRESS} = 0.870) \quad (5)$$

$$\log(1/C) = 0.224(\pm 0.130)HE + 411.489(\pm 223.484)H8 - 142.342(\pm 115.402)O7 + 46.670(\pm 36.525)C5 + 42.487(\pm 25.312)C6 - 170.690(\pm 118.342)$$

$$(n = 35,\ r = 0.888,\ s = 0.702,\ F = 21.659,\ Q^2 = 0.682,\ S_{PRESS} = 0.862) \quad (6)$$

$$\log(1/C) = 0.206(\pm 0.130)HE + 0.045(\pm 0.061)MR + 304.241(\pm 262.794)H8 - 104.928(\pm 123.926)O7 + 42.767(\pm 36.184)C5 + 39.491(\pm 25.135)C6 - 124.442(\pm 131.806)$$

$$(n = 35,\ r = 0.897,\ s = 0.687,\ F = 19.232,\ Q^2 = 0.679,\ S_{PRESS} = 0.881) \quad (7)$$

$$\log(1/C) = 0.229(\pm 0.131)HE + 0.050(\pm 0.060)MR + 510.817(\pm 384.848)H8 - 148.359(\pm 135.593)O7 + 49.730(\pm 36.776)C5 + 39.378(\pm 24.655)C6 - 31.728(\pm 43.891)N3 - 225.084(\pm 189.994)$$

$$(n = 35,\ r = 0.905,\ s = 0.673,\ F = 17.505,\ Q^2 = 0.668,\ S_{PRESS} = 0.913) \quad (8)$$

TABLE 3
THE 95% CONFIDENCE, F AND p STATISTICS FOR THE COEFFICIENTS OF VARIABLES IN EQS. 2–11

Eq.	Variable	Coefficient	95% confidence	F	p	Significance
2	HE	0.402	±0.134	37.473	0.000	
3	HE	0.297	±0.123	24.118	0.000	
	MR	0.105	±0.054	15.854	0.000	
4	HE	0.326	±0.127	27.440	0.000	
	MR	0.090	±0.057	10.441	0.003	
	H8	76.020	±101.673	2.325	0.137	not significant
5	HE	0.198	±0.134	9.026	0.005	
	MR	0.093	±0.052	13.454	0.001	
	C5	45.797	±35.602	6.902	0.013	
	C6	38.483	±27.062	8.434	0.007	
6	HE	0.224	±0.130	12.548	0.001	
	H8	411.489	±223.484	14.181	0.001	
	O7	−142.342	±115.402	6.364	0.017	
	C5	46.670	±36.526	6.829	0.014	
	C6	42.487	±25.312	11.785	0.002	
7	HE	0.206	±0.129	10.623	0.003	
	MR	0.045	±0.006	2.288	0.142	not significant
	H8	304.242	±262.794	5.624	0.025	
	O7	−104.928	±123.926	3.008	0.094	not significant
	C5	42.767	±36.184	5.862	0.003	
	C6	39.491	±25.135	10.357	0.003	
8	HE	0.229	±0.131	12.869	0.001	
	MR	0.050	±0.060	2.924	0.099	not significant
	H8	510.818	±384.848	7.417	0.011	
	O7	−148.359	±135.593	5.040	0.033	
	C5	49.730	±36.776	7.698	0.010	
	C6	39.378	±24.656	10.739	0.003	
	N3	−31.728	±43.891	2.200	0.150	not significant
9	HE	0.268	±0.135	16.437	0.000	
	H8	158.862	±96.948	11.199	0.002	
	C5	66.923	±35.373	14.930	0.001	
	C6	44.124	±27.404	10.813	0.003	
10	HE	0.277	±0.151	13.935	0.001	
	O7	52.390	±55.319	3.741	0.063	not significant
	C5	72.593	±40.377	13.482	0.001	
	C6	46.019	±30.237	9.661	0.004	
11	HE	0.228	±0.121	14.875	0.006	
	MR	0.118	±0.049	24.216	0.000	
	C5	31.071	±33.111	3.684	0.648	not significant
	C6	28.979	±24.851	5.689	0.238	

TABLE 4
ACTUAL PREDICTION OF EQS. 3, 5, 6, 9–11 FOR THE FIVE TESTED COMPOUNDS

Tested compound	log(1/C)						
	Experimental	Predicted					
		Eq. 3	Eq. 5	Eq. 6	Eq. 9	Eq. 10	Eq. 11
36	8.108	7.664	7.613	7.382	7.342	7.029	7.926
37	5.585	5.586	5.413	5.965	5.525	5.421	5.479
38	4.086	3.696	4.419	4.964	4.678	4.646	4.121
39	3.721	3.475	3.637	3.453	3.705	3.658	3.553
40	3.558	2.879	3.789	4.096	3.659	3.652	3.446
SSE[a]		0.871	0.446	1.803	0.951	1.516	0.086

[a] Sum of square error of prediction for five tested compounds.

In Eqs. 2–8, n is the number of compounds used in the fit, F is the overall F-statistics for the addition of each successive term, and values in parentheses are the 95% confidence limit of each coefficient. The statistics for the coefficients, i.e. 95% confidence level, F and p, are summarized in Table 3. The actual predictions for the five tested compounds applying the investigated equations are given in Table 4.

In order to verify the validity of the models, a cross-validation test was applied to the original data set. For a reasonable QSAR model, the ratio PRESS/SSY should be smaller than 0.4 or Q^2 should be greater than 0.6 [38]. Considering statistical criteria, Eq. 2 was the best single-variable model. However, the r value and the predictive ability were not acceptable. Equations 4, 7 and 8 contained at least one insignificant coefficient, as revealed by the confidence interval and F-statistics (Table 3). Only models 3, 5 and 6 were considered to be statistically significant, according to statistical criteria and predictive ability. In terms of actual prediction, these three equations predict well for the five tested compounds (Table 4). The predictions are even better than those of the fitted set (see Table 1 and Eq. 5 as an example). The values of s and S_{PRESS} in Eq. 6 are smaller than elsewhere; nevertheless, this equation contains O7 and H8, which are highly cross-correlated. In order to examine the reliability of both variables in the model, O7 or H8 were excluded from Eq. 6, producing Eqs. 9 and 10, respectively.

$$\log(1/C) = 0.268(\pm 0.135)HE + 158.862(\pm 96.948)H8 + 66.923(\pm 35.372)C5 + 44.124(\pm 27.404)C6 - 29.820(\pm 33.615)$$

$$(n = 35,\ r = 0.862,\ s = 0.763,\ F = 21.617,\ Q^2 = 0.667,\ S_{PRESS} = 0.868) \quad (9)$$

$$\log(1/C) = 0.277(\pm 0.151)HE + 52.390(\pm 55.319)O7 + 72.593(\pm 40.377)C5 + 46.019(\pm 30.237)C6 + 44.333(\pm 22.948)$$

$$(n = 35,\ r = 0.828,\ s = 0.843,\ F = 16.347,\ Q^2 = 0.591,\ S_{PRESS} = 0.961) \quad (10)$$

Without O7 in Eq. 9, the statistical significance was lower than that of Eq. 5. Exclusion of H8 in Eq. 10 was not acceptable, as this produced a coefficient that was not significant (Table 3) and an extreme change in the regression coefficient of O7 (from −142 to +52). In a comparison of the four-variable models obtained by Eqs. 5 and 9, the sum of square error of prediction for five tested compounds, shown in Table 4, revealed that Eq. 5 is more reliable. Consequently, Eq. 5 can be considered as the

TABLE 5
SQUARED CORRELATION MATRIX FOR PARAMETERS USED IN THE CORRELATION STUDY

	N1	C2	N3	C4	C5	C6	O7	H8	HE	MR	logP	log(1/C)
N1	1.000	−0.928	−0.234	−0.527	0.122	−0.099	0.417	0.719	0.123	0.550	−0.415	0.331
C2		1.000	−0.990	0.664	−0.415	0.312	−0.140	−0.538	−0.268	−0.636	0.573	−0.478
N3			1.000	−0.612	0.351	−0.234	0.211	0.610	0.274	0.639	−0.526	0.506
C4				1.000	−0.751	0.804	0.328	0.000	−0.200	−0.449	0.578	−0.226
C5					1.000	−0.886	−0.670	−0.372	0.279	0.347	−0.522	0.287
C6						1.000	0.0565	0.341	0.016	−0.153	0.553	0.053
O7							1.000	0.874	−0.329	−0.084	0.241	−0.139
H8								1.000	−0.149	0.259	−0.009	0.169
HE									1.000	0.434	0.017	0.729
MR										1.000	−0.066	0.671
logP											1.000	−0.043
log(1/C)												1.000

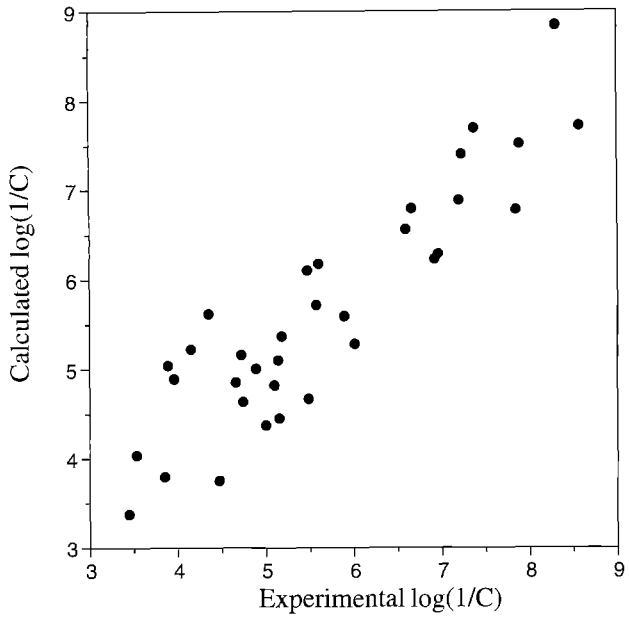


Fig. 2. Comparison of experimental log(1/C) with calculated log(1/C) obtained from Eq. 5.

most suitable model, with both high statistical significance and high predictive ability.

In essence, it was found from our study that Eq. 5 was the best QSAR equation. The multiple correlation coefficient of this equation is 0.869 and it can explain 75.58% of the variation in the biological data. The experimental versus the calculated affinities of Eq. 5 are reported in Table 1. Analysis of residuals from Eq. 5 indicates that compound 6 is the highest outlier. Removal of this compound from the regression caused a significant improvement of the r value from 0.869 (see Eq. 5) to 0.902, as shown in Eq. 11:

$$\log(1/C) = 0.228(\pm 0.121)HE + 0.118(\pm 0.049)MR + 31.072(\pm 33.111)C5 + 28.979(\pm 24.851)C6 + 10.195(\pm 8.339)$$

$$(n = 34,\ r = 0.902,\ s = 0.658,\ F = 31.583,\ Q^2 = 0.760,\ S_{PRESS} = 0.746) \quad (11)$$

The reason for the differences seen in connection with this outlier cannot be explained by the model. However, it was observed that R1 in compound 6 is 3-*t*Bu, and therefore it has the largest MR value (63.928) among the first 20 compounds (Table 1).

Discussion

Molecular parameters

Due to the high degree of correlation with log(1/C) of HE (0.729) and MR (0.671) (Table 5), it seems that both variables play strong roles in the proposed QSAR equations. Exclusion of HE causes not only low fitting quality, but also poor predictivity of the model. According to the HE definition used here, which stands for solvent–solvent cavity term, solute–solvent van der Waals term and solute–solvent electrostatic polarization term [32], the HE parameter is considered to describe the drug–solvent interaction. The positive values of the coefficient for this term point out that the HEPT derivatives displaying a high HE, binding more tightly to the solvent molecules than those with a low HE, diffuse slower through the biological environment to the receptor target. In addition, electrostatic polarizability induced by the HE can help to facilitate correct orientation of drug molecules for the first contact at the binding site. It is suggested from the model obtained in Eq. 5 that MR is a neccessary contributor to activity. This variable represents steric bulk and/or dispersion interactions due to substituents of the compounds. A positive sign of the coefficient for this term indicates binding of the substituents to a polar surface [39]. From analysis of logP in our correlation study, it was found that this parameter did not contribute towards HIV-1 inhibitory activity in this class of compounds. Attempts to use other regression methods to correlate activity with the parabolic lipophilicity term were also unsatisfactory. Some comment could be made here that the logP values for the HEPT derivatives, obtained from the described method [33], seem to be unrealistic, as the highest logP is 8.7 (compound 23). Therefore, using the relative values of logP, this molecular variable was not included in the derived QSAR models.

Electronic parameters

Among the atomic variables within the thymine ring, C5 and C6 were found to play an important role in the correlation equation. Substitution of C6 by the phenylthio

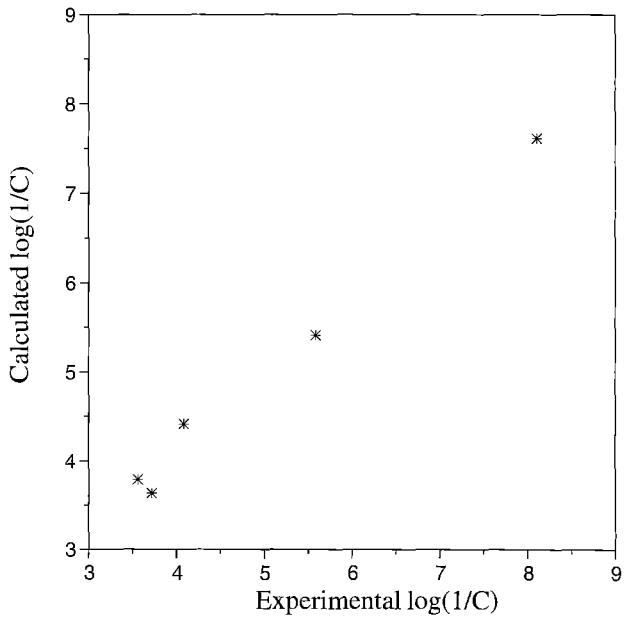


Fig. 3. Plot of actual prediction of Eq. 5 for five tested compounds.

R1 group results in negative charges on this atom. The same effect is observed on C5 upon substitution by alkyl groups. Addition of these two variables to Eq. 3, as shown in Eq. 5, produced the best QSAR model. Inclusion of O7 and H8 instead of MR, as shown in Eq. 6, gave better statistical results. However, a high intercorrelation between the O7 and H8 variables (0.874) was encountered, i.e., this model was found to be not reliable. Equation 5 was detected as being the most suitable correlation equation, based on statistical criteria and predictive ability for both internal and external sets of compounds.

Conclusions

We have attempted to evaluate electronic and molecular properties of HEPT analogs and to correlate their changes as a function of biological activity. The derived QSAR model in this study was reasonably satisfying, based on both statistical significance and predictive ability. The obtained model indicates a correlation between the HIV-1 inhibition of these HEPT compounds and atomic charges of two substituted carbon atoms in the thymine ring, hydration energies and molar refractivities of the molecules.

Acknowledgements

The authors would like to thank Suvaporn Luangkhamin and Kanlaya Chimkhlai for their supporting data, and Dr. Teerakiat Kerdcharoen for his helpful assistance. Financial support from KURDI, in the form of providing an HP-715 workstation, is also gratefully acknowledged.

References

1 Broder, S. (Ed.) AIDS, Modern Concepts and Therapeutic Challenges, Marcel Dekker, New York, NY, 1987.
2 Roey, P.V., Salerno, J.M., Duax, W.L., Chu, C.K., Ahn, M.K. and Schinazi, R.F., J. Am. Chem. Soc., 110 (1988) 2277.
3 Greengrass, C.W., Hoople, D.W.T., Street, S.D.A., Hamilton, F., Marriott, M.S., Bordner, J., Dalgleish, A.G., Mitsuya, H. and Broder, S., J. Med. Chem., 32 (1989) 618.
4 Tucker, T.J., Lumma Jr., W.C., Payne, L.S., Wai, J.M., de Solms, S.J., Giuliani, E.A., Darke, P.L., Hembach, J.C., Zugay, J.A., Schleif, W.A., Quintero, J.C., Emini, E.A., Huff, J.R. and Anderson, P.S., J. Med. Chem., 35 (1992) 2525.
5 Getman, D.P., DeCrescenzo, G.A., Heintz, R.M., Reed, K.L., Talley, J.J., Brayant, M.L., Clare, M., Houseman, K.A., Marr, J.J., Mueller, R.A., Vazquez, M.L., Shieh, H.-S., Stallings, W.C. and Stegeman, R.A., J. Med. Chem., 36 (1993) 288.
6 Clercq, E.D., J. Med. Chem., 38 (1995) 2491.
7 Tanaka, H., Takashima, H., Ubasawa, M., Sekiya, K., Nitta, I., Baba, M., Shigeta, S., Walker, R.T., Clercq, E. and Miyasaka, T., J. Med. Chem., 35 (1992) 337.
8 Tanaka, H., Takashima, H., Ubasawa, M., Sekiya, K., Nitta, I., Baba, M., Shigeta, S., Walker, R.T., Clercq, E. and Miyasaka, T., J. Med. Chem., 35 (1992) 4713.
9 Nicklaus, M.C., Milne, G.W.A. and Burke Jr., T.R., J. Comput.-Aided Mol. Design, 6 (1992) 487.
10 Klebe, G. and Abraham, U., J. Med. Chem., 36 (1993) 70.
11 Waller, C.L., Oprea, T.I., Giolitti, A. and Marshall, G.R., J. Med. Chem., 36 (1993) 4152.
12 Van de Waterbeemd, H. (Ed.) Advanced Computer-Assisted Techniques in Drug Discovery, VCH, Weinheim, 1994.
13 Kokpol, S.U., Hannongbua, S.V., Thongrit, N., Polman, S., Rode, B.M. and Swendinger, M.G., Anal. Sci., 4 (1988) 565.
14 Rode, B.M., Swendinger, M.G., Kokpol, S.U., Hannongbua, S.V. and Polman, S., Monatsschr. Chem., 120 (1989) 913.
15 Polman, S., Kokpol, S.U., Hannongbua, S.V. and Rode, B.M., Anal. Sci., 5 (1989) 641.
16 Hannongbua, S., In Proceedings of the 33rd Annual Conference of the Kasetsart University, Bangkok, January 30–February 1, 1995, pp. 273–279.
17 Pauling, L. and Pressman, D., J. Am. Chem. Soc., 67 (1945) 1003.
18 Agin, D., Hersh, L. and Holtzman, D., Proc. Natl. Acad. Sci. USA, 53 (1965) 952.
19 Silipo, C. and Hansch, C., J. Am. Chem. Soc., 97 (1975) 6849.
20 Selassie, C.D., Fang, S.-X., Li, R., Hansch, C., Debnath, G., Klien, T.E., Langridge, R. and Kaufman, B.T., J. Med. Chem., 32 (1989) 1895.
21 Andrea, T.A. and Kalayeh, H., J. Med. Chem., 34 (1991) 2824.
22 Ferguson, D.M., Radmer, R.J. and Kollman, P.A., J. Med. Chem., 34 (1991) 2654.
23 Reddy, M.R., Vishwanadhan, V.N. and Weinstein, J.N., Proc. Natl. Acad. Sci. USA, 88 (1991) 10287.
24 Lybrand, T.P. and McCammon, J.A., J. Comput.-Aided Mol. Design, 2 (1988) 259.
25 Hansch, C. and Fujita, T., J. Am. Chem. Soc., 86 (1964) 1616.
26 Hansch, C. and Leo, A., Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York, NY, 1979.
27 Verloop, A., In Ariëns, E.J. (Ed.) Drug Design, Academic Press, New York, NY, 1972.
28 ALCHEMY III, Tripos Associates Inc., St. Louis, MO, 1992.
29 GAUSSIAN 92, Gaussian Inc., Pittsburgh, PA, 1992.
30 Dewar, M.J., Zoebisch, E.G., Healy, E.F. and Stewart, J.P., J. Am. Chem. Soc., 107 (1985) 3902.
31 ChemPlus, Hypercube Inc., Waterloo, ON, 1993.
32 a. Ghose, A.K. and Crippen, G.M., J. Chem. Inf. Comput. Sci., 27 (1987) 21.
b. Viswanadhan, V.N., Ghose, A.K., Revankar, G.N. and Robins, R.K., J. Chem. Inf. Comput. Sci., 29 (1989) 163.
33 Bodor, N., Gabanyi, Z. and Wong, C., J. Am. Chem. Soc., 111 (1989) 3783.
34 Still, W.C., Tempczyk, A., Hawley, R.C. and Hendrickson, T., J. Am. Chem. Soc., 112 (1990) 6127.
35 Norusis, M.J., The SPSS Guide to Data Analysis for SPSS/PC+, 2nd ed., SPSS, Inc., Chicago, IL, 1991.
36 Cramer, R.D., Bunce, J.D. and Patterson, D.E., Quant. Struct.–Act. Relatsh., 7 (1988) 18.
37 Topliss, J.G. and Costello, R.J., J. Med. Chem., 15 (1973) 1066.
38 Wold, S., Quant. Struct.–Act. Relatsh., 10 (1991) 191.
39 Kubinyi, H., QSAR: Hansch Analysis and Related Approaches, VCH, Weinheim, 1993.