

# A 3D QSAR Study of a Series of HEPT Analogues: The Influence of Conformational Mobility on HIV-1 Reverse Transcriptase Inhibition

Dmitri B. Kireev, Jacques R. Chrétien,\* David S. Grierson,† and Claude Monneret‡

Laboratoire de Chimiométrie, Université d'Orléans, B.P. 6759, Orléans Cedex 2, 45067, France, Institut de Chimie des Substances Naturelles, UPR CNRS 2301, 91198, Gif-sur-Yvette, France, and Institut Curie, Section Recherche, UMR CNRS 176, 26 r. d'Ulm, 75231, Paris, France

Received February 24, 1997<sup>⊗</sup>

Quantitative structure–activity relationships (QSAR) have been established for 87 analogues of 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine (HEPT), a potent inhibitor of the HIV-1 reverse transcriptase (RT). Of these 87 nonnucleoside RT inhibitors, 9 novel HEPT analogues were used in the study and the others were taken from the literature. The predictive ability of these relationships has been evaluated using a large set of 54 compounds which were not used to derive the activity model. Descriptors related to the conformational changes were found to be an important factor which underlies RT inhibitory activity in the HEPT series. Indeed, the QSAR model provides evidence concerning the conformational transformations the molecules may undergo during the inhibition process. The established relationships are supplementary to the experimental study on the binding of HEPT type inhibitors to RT by Hopkins *et al.* (*J. Med. Chem.* **1996**, *39*, 1589–1600). The present study suggests a quantitative interpretation of the structure–activity relationships which otherwise cannot be explained within the framework of the crystal inhibitor–protein model. This information is pertinent to the further design of new HEPT type RT inhibitors.

## Introduction

Reverse transcriptase (RT) is a key enzyme of the human immunodeficiency virus (HIV), catalyzing the RNA-dependent and DNA-dependent synthesis of double-strand viral DNA. From random screening programs, a number of potent and structurally different compounds which are nonnucleoside type inhibitors (NNI's) of RT have been identified. These include TIBO, HEPT, nevirapine, pyridinone, BHAP, and  $\alpha$ -APA.<sup>1</sup> In general, these NNIs display fewer side effects compared to the nucleoside-based inhibitors (NIs) 3'-azido-3'-deoxythymidine (AZT), 2',3'-dideoxycytidine (ddC), 2',3'-dideoxyinosine (ddI) and (-)-2',3'-dideoxythiacytidine (3TC).<sup>2</sup> However, the efficiency of both nucleoside and nonnucleoside RT inhibitors is limited by the high rate of the virus mutation which rapidly leads to the emergence of drug-resistant viral strains.<sup>3</sup>

The results of recent studies on the "knock out approach"<sup>4</sup> suggest that, by using NNI's at elevated, but nontoxic concentrations in cell culture, HIV-1 replication can be completely suppressed. Even more pertinent, in the combination therapy approach, in which two or more nucleoside inhibitors and/or NNI's together with a protease inhibitor are administered simultaneously to AIDS patients,<sup>5</sup> the presence of HIV *in sera* is found to drop below detectable limits. To follow up on these strategies it will be necessary to have on hand an arsenal of new compounds with improved activities and/or less vulnerability to viral drug resistance.

The rational design of new NNI's interacting with the allosteric pocket will require a more detailed knowledge of the mechanism of RT inhibition by this class of compounds. The availability of three-dimensional X-ray crystal structures of RT, complexed with a variety of

NNI's, has enabled considerable progress in this direction.<sup>6</sup> In a complementary manner, clinical and biochemical studies have provided valuable knowledge concerning the mutagenic substitutions at the allosteric binding site in RT and other aspects of the inhibition mechanism.<sup>7,8</sup>

Computer simulation techniques potentially offer further means to probe inhibition mechanisms. The quantitative structure–activity relationships (QSAR)<sup>9</sup> represent one of the most effective computational approaches in drug design. While QSAR is largely used to predict activities and to define pharmacophore models, it can also be involved in understanding the behavior of an inhibitor in biological systems. Particularly, QSAR models can provide information on the types of intermolecular and intramolecular interactions the active molecules are exposed to during the course of their incorporation into the enzyme.

In the present study on the inhibition of HIV-1 reverse transcriptase by NNI's, a QSAR analysis was performed on a series of 87 HEPT type RT inhibitors I–VI. HEPT (1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine) is the parent compound in a series of NNI's which display activity against HIV-1 RT down to subnanomolar levels in MT-4 cell culture.<sup>10</sup>

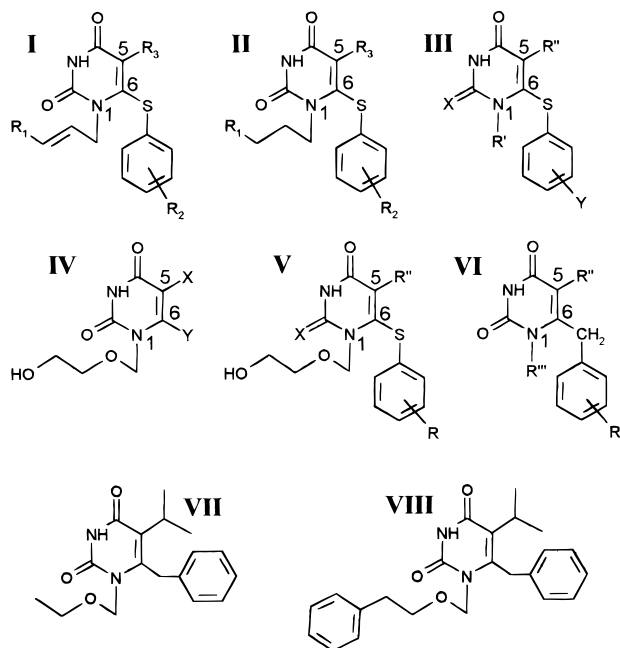
While this paper was in preparation, the crystal structures of HIV-1 RT complexed with HEPT and two more potent HEPT analogues, MKC-442 (VII) and TNK-651 (VIII), became available.<sup>11</sup> This crystal study clearly shows the conformation in which HEPT analogues bind to the p66 enzyme unit. Another finding of this experimental study is a switching role of the C-5 substituent which forces the Tyr181 protein residue to change its conformation, thereby, remarkably increasing the protein–inhibitor affinity. The following results illustrate the complementarity of the present QSAR analysis with respect to the experimental study.<sup>11</sup>

The objectives of our study were both to provide

† UPR CNRS 2301.

‡ UMR CNRS 176.

⊗ Abstract published in *Advance ACS Abstracts*, November 15, 1997.

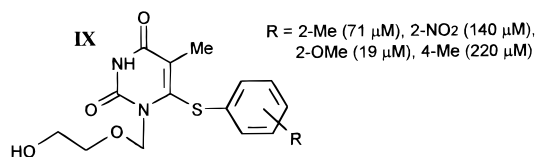


supplementary information concerning the behavior of these compounds and to further define the criteria necessary for the rational design of new generation HEPT type anti-HIV agents. Specifically, this study indicates which conformational transformations the inhibitors may undergo and why the fragments which ought to be good for attaining high affinity for the allosteric binding site may, in fact, decrease the inhibitory activity.

### Computational Methodology

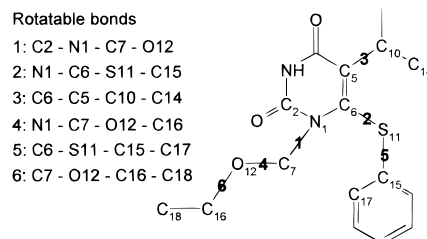
**Selection of the Dataset.** In this study, a series of nine 1-arylallyl and 1-arylpropyl HEPT analogues (I, II) prepared in our laboratories<sup>12</sup> was combined with four different families of HEPT derivatives described by Tanaka *et al.* (III,<sup>13</sup> IV,<sup>14</sup> V,<sup>15</sup> VI<sup>10</sup>) for which the *in vitro* inhibitory activity was measured in MT-4 cell culture.

Of the 82 compounds described by Tanaka, 78 were included in the training set. The four remaining compounds containing an ortho-substituted thiophenyl moiety and/or a para-substituted thiophenyl moiety (see structure IX) were discarded, since they do not suffice to quantify the influence of this type of substitution. Moreover, these four compounds are much less active than HEPT itself, and *a priori* substitution at these positions is ultimately unfavorable.



The structural formulas and *in vitro* RT inhibitory activities for the HEPT analogues studied are shown in Table 1.

One more issue concerns the compatibility between the biological data of Tanaka *et al.*<sup>10,13-15</sup> and our data.<sup>12</sup> Two reference compounds, HEPT and BPT [1-[(2-benzyloxy)methyl]-6-(phenylthio)thymine],<sup>15</sup> were used in this work in order to estimate the activity difference



**Figure 1.** The numbering of the rotatable bonds in compound 30.

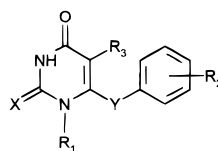
between the two testing laboratories. The RT inhibitory activity of HEPT in the MT-4 cell culture is 11 μM in our study<sup>12</sup> while the activity of HEPT cited by Tanaka *et al.*<sup>10,13-15</sup> is 7 μM; hence the corrective ratio (11/7) is 1.6. Similarly, the activity for BPT is 0.74 μM in our study<sup>12</sup> while the activity measured by Tanaka *et al.*<sup>15</sup> is 0.088; hence the ratio (0.74/0.088) is 8.4. An average between 1.6 and 8.4, *i.e.* 5, may be used as a correction factor between the two series. The variation of the correction factor for the two sample compounds is rather large. However, our biological tests repeated on different samples of HEPT and BPT always resulted in the same values of EC<sub>50</sub>. A search for these or other reference compounds in the literature gave no result. Finally, we kept the correction factor 5, and the IC<sub>50</sub> values from our laboratory<sup>12</sup> were divided by 5 before proceeding with the QSAR analyses.

**Molecular Modeling.** The structures of the HEPT analogues were modeled using Sybyl 6.0 (Tripos Associates, St. Louis, MO). The initial low-energy conformations were optimized using the molecular mechanics (MM) algorithm with the Tripos Force Field.<sup>16</sup> All other low energy conformations were then obtained using the SYBYL's *Systematic Search* procedure by rotation around all rotatable bonds with an angle step of 30°. *Systematic Search* creates a database which contains the conformers whose energies are within a specified limit (we fixed this limit at 10 kcal/mol). But *Systematic Search* does not optimize the geometry of the found conformers.

Compound 30, one of the most constrained of active compounds, was taken to explore the conformational diversity of the HEPT analogues. This compound counts six rotatable bonds as shown in Figure 1 (rotations of the methyl groups were neglected). Hence, the theoretical number of conformations is 12<sup>6</sup>. The systematic conformational search resulted in 3514 sterically accessible conformations within a 10 kcal/mol interval.

Among these 3514, there is a number of isoenergetic conformers obtainable by changing the sign of some torsion angles. For example, a conformer whose torsion angles 1 and 2 are both equal to 300° may be obtained by the reflection of a corresponding conformer whose torsion angles 1 and 2 are both equal to 60°. The energies and other physical properties of these two conformers are identical. As we cannot obtain different properties, *i.e.* different values of molecular descriptors, from these two isoenergetic conformers, we may eliminate one of them. As each conformer from the 3514 has its homologue with inverted values of torsion angles 1 and 2, the total number of eliminated conformers is 1757.

Geometries of the remaining 1757 conformers were optimized using the Tripos force field. This optimization

**Table 1.** Structures and RT Inhibitory Activity of HEPT Analogues

no.	no. <sup>ref</sup>	X	R <sub>1</sub>	Y	R <sub>2</sub>	R <sub>3</sub>	EC <sub>50</sub> , μM	no.	no. <sup>ref</sup>	X	R <sub>1</sub>	Y	R <sub>2</sub>	R <sub>3</sub>	EC <sub>50</sub> , μM
1	2 <sup>a</sup>	O	CH <sub>2</sub> O(CH <sub>2</sub> ) <sub>2</sub> OMe	S	H	Me	8.7	45	20 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3-OMe	Me	22
2	26 <sup>a</sup>	O	CH <sub>2</sub> OMe	S	H	Me	2.1	46	28 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3,5-Me	Me	0.26
3	27 <sup>a</sup>	O	CH <sub>2</sub> OEt	S	H	Me	0.33	47	29 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3,5-Cl	Me	1.3
4	28 <sup>a</sup>	O	CH <sub>2</sub> OPr	S	H	Me	3.6	48	30 <sup>b</sup>	S	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3,5-Me	Me	0.22
5	29 <sup>a</sup>	O	CH <sub>2</sub> OAu	S	H	Me	4.7	49	33 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3-COOMe	Me	7.9
6	31 <sup>a</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> Ph	S	H	Me	0.088	50	34 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3-COMe	Me	7.3
7	32 <sup>a</sup>	S	CH <sub>2</sub> OEt	S	H	Et	0.026	51	40 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3-CN	Me	10
8	33 <sup>a</sup>	S	CH <sub>2</sub> OEt	S	3,5-Me	Et	0.0044	52	48 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	H	Et	0.11
9	34 <sup>a</sup>	S	CH <sub>2</sub> OEt	S	3,5-Cl	Et	0.013	53	50 <sup>b</sup>	S	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	H	<i>i</i> -Pr	0.059
10	35 <sup>a</sup>	S	CH <sub>2</sub> - <i>i</i> -Pr	S	H	Et	0.22	54	51 <sup>b</sup>	S	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3,5-Me	Et	0.0078
11	36 <sup>a</sup>	S	CH <sub>2</sub> O-c-Hex	S	H	Et	1.6	55	52 <sup>b</sup>	S	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3,5-Me	<i>i</i> -Pr	0.005
12	37 <sup>a</sup>	S	CH <sub>2</sub> OCH <sub>2</sub> -c-Hex	S	H	Et	0.35	56	53 <sup>b</sup>	S	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3,5-Cl	Et	0.043
13	38 <sup>a</sup>	S	CH <sub>2</sub> OCH <sub>2</sub> Ph	S	H	Et	0.0078	57	54 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	H	Et	0.12
14	39 <sup>a</sup>	S	CH <sub>2</sub> OCH <sub>2</sub> Ph	S	3,5-Me	Et	0.0069	58	56 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	H	<i>i</i> -Pr	0.063
15	40 <sup>a</sup>	S	CH <sub>2</sub> OCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-Me)	S	H	Et	0.078	59	57 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3,5-Me	Et	0.013
16	41 <sup>a</sup>	S	CH <sub>2</sub> OCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-Cl)	S	H	Et	0.012	60	58 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3,5-Me	<i>i</i> -Pr	0.0027
17	42 <sup>a</sup>	S	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> Ph	S	H	Et	0.091	61	59 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3,5-Cl	Et	0.014
18	43 <sup>a</sup>	S	CH <sub>2</sub> OEt	S	H	<i>i</i> -Pr	0.014	62	8 <sup>c</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	c-Hex	Me	3.5
19	44 <sup>a</sup>	S	CH <sub>2</sub> OCH <sub>2</sub> Ph	S	H	<i>i</i> -Pr	0.0068	63	13 <sup>c</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	H	Me	8.9
20	45 <sup>a</sup>	S	CH <sub>2</sub> OEt	S	H	c-Pr	0.095	64	27 <sup>c</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	CH <sub>2</sub>	H	Me	5.7
21	46 <sup>a</sup>	O	CH <sub>2</sub> OEt	S	H	Et	0.019	65	44 <sup>c</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	H	I	4.3
22	47 <sup>a</sup>	O	CH <sub>2</sub> OEt	S	3,5-Me	Et	0.0054	66	55 <sup>c</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	H	CH=CH <sub>2</sub>	4.7
23	48 <sup>a</sup>	O	CH <sub>2</sub> OEt	S	3,5-Cl	Et	0.0074	67	27 <sup>d</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	CH <sub>2</sub>	H	Et	0.35
24	49 <sup>a</sup>	O	CH <sub>2</sub> - <i>i</i> -Pr	S	H	Et	0.34	68	28 <sup>d</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	CH <sub>2</sub>	3,5-Me	Et	0.013
25	50 <sup>a</sup>	O	CH <sub>2</sub> O-c-Hex	S	H	Et	4.0	69	29 <sup>d</sup>	O	CH <sub>2</sub> OEt	CH <sub>2</sub>	H	Et	0.041
26	51 <sup>a</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> -c-Hex	S	H	Et	0.45	70	30 <sup>d</sup>	O	CH <sub>2</sub> OEt	CH <sub>2</sub>	3,5-Me	Et	0.0016
27	52 <sup>a</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> Ph	S	H	Et	0.0059	71	31 <sup>d</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	CH <sub>2</sub>	H	<i>i</i> -Pr	0.063
28	53 <sup>a</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> Ph	S	3,5-Me	Et	0.0032	72	32 <sup>d</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	CH <sub>2</sub>	3,5-Me	<i>i</i> -Pr	0.0027
29	54 <sup>a</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> Ph	S	H	Et	0.096	73	33 <sup>d</sup>	O	CH <sub>2</sub> OEt	CH <sub>2</sub>	H	<i>i</i> -Pr	0.0042
30	55 <sup>a</sup>	O	CH <sub>2</sub> OEt	S	H	<i>i</i> -Pr	0.012	74	34 <sup>d</sup>	O	CH <sub>2</sub> OEt	CH <sub>2</sub>	3,5-Me	<i>i</i> -Pr	0.0006
31	56 <sup>a</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> Ph	S	H	<i>i</i> -Pr	0.0027	75	37 <sup>d</sup>	O	Au	CH <sub>2</sub>	H	Et	0.21
32	57 <sup>a</sup>	O	CH <sub>2</sub> OEt	S	H	c-Pr	0.1	76	38 <sup>d</sup>	O	Au	CH <sub>2</sub>	H	<i>i</i> -Pr	0.042
33	60 <sup>a</sup>	O	Et	S	H	Me	2.2	77	39 <sup>d</sup>	O	CH <sub>2</sub> CH <sub>2</sub> OMe	CH <sub>2</sub>	H	Et	0.25
34	61 <sup>a</sup>	O	Bu	S	H	Me	1.2	78	40 <sup>d</sup>	O	CH <sub>2</sub> CH <sub>2</sub> OMe	CH <sub>2</sub>	H	<i>i</i> -Pr	0.052
35	9 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3-Me	Me	2.6	79	39 <sup>a</sup>	O	CH <sub>2</sub> CH=CHPh	S	H	Me	2.5
36	10 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3-Et	Me	2.7	80	40 <sup>a</sup>	O	CH <sub>2</sub> CH=CHPh	S	H	Et	0.19
37	11 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	<i>t</i> -Bu	Me	12	81	41 <sup>a</sup>	O	CH <sub>2</sub> CH=CHPh	S	3,5-Me	Et	0.039
38	13 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3-CF <sub>3</sub>	Me	45	82	40 <sup>d</sup>	O	CH <sub>2</sub> CH=CH-thienyl	S	H	Et	0.08
39	14 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3-F	Me	3.3	83	41 <sup>d</sup>	O	CH <sub>2</sub> CH=CH-thienyl	S	3,5-Me	Et	0.04
40	15 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3-Cl	Me	13	84	40 <sup>c</sup>	O	CH <sub>2</sub> CH=CH-furyl	S	H	Et	0.17
41	16 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3-Br	Me	5.7	85	41 <sup>c</sup>	O	CH <sub>2</sub> CH=CH-furyl	S	3,5-Me	Et	0.04
42	17 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3-I	Me	10	86	39 <sup>b</sup>	O	CH <sub>2</sub> CH=CH-3-pyridyl	S	H	Me	0.1
43	18 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3-NO <sub>2</sub>	Me	34	87	—	O	(CH <sub>2</sub> ) <sub>3</sub> Ph	S	H	Et	0.29
44	19 <sup>b</sup>	O	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	3-OH	Me	82								

<sup>a</sup> Compounds from ref 12. <sup>b</sup> Compounds from ref 14. <sup>c</sup> Compounds from ref 13. <sup>d</sup> Compounds from ref 9. <sup>e</sup> Compounds from ref 11.

served to eliminate many energetically unstable conformers, and in the end, a total of 144 different low-energy conformers were obtained. In the database of optimized conformers, torsion angle 1 takes only two values, 60° (±5°) and 300° (±5°), torsion angle 2 takes a single value of 65° (±5°), and torsion angle 3 takes two values, 0° and 180°.

Switching torsion angle 1 from 60° to 300° significantly changes the shape of the molecule (see Figure 2). In the conformer **a**, whose torsion angle 1 is equal to 60°, the 6-phenylthio group and the terminal atom of the N-1 substituent are on the same side of the uracil ring plane. This conformer will be further referred to as a 1,6-*cis* conformer. The conformer **b**, whose torsion angle 1 is equal to 300°, has the 6-phenylthio group and the terminal atom of the N-1 substituent on different sides of the uracil ring plane and will be further referred to as a 1,6-*trans* conformer. The **a** and **b** structures

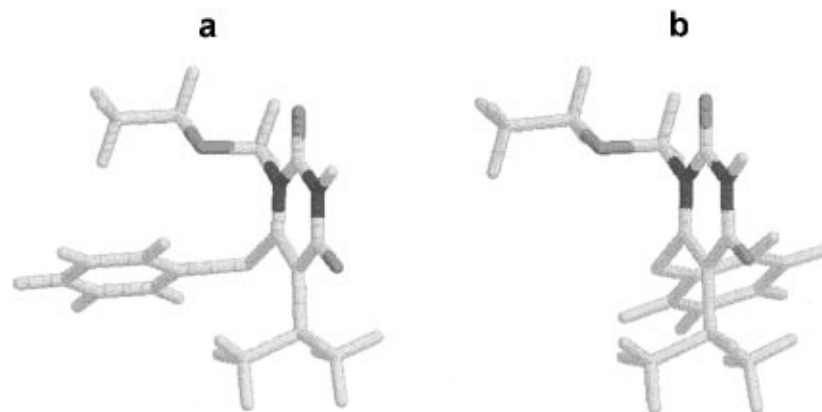
belong to two different families (1,6-*cis* and 1,6-*trans*) of geometrically similar conformers. In each family the conformers differ in the arrangement of the N-1 chain and/or in the position of the C-5 substituent (torsion angle 3 may be equal to 0 or 180°).

Then *Systematic Search* with subsequent geometry optimization was performed on five more compounds (**14**, **24**, **70**, **75**, **81**) having high activity values. All conformers of these compounds also belong to two (*i.e.* 1,6-*cis* and 1,6-*trans*) families.

The low-energy conformations so obtained served as initial geometries for semiempirical quantum chemical calculations by the AM1 method<sup>17</sup> in order to get partial atomic charges.

## QSAR

**Molecular Descriptors.** A set of molecular descriptors related to physicochemical, electronic and geometric



**Figure 2.** Two most stable conformational families found for the HEPT analogues.

properties of the molecules was used for this study. The majority of the descriptors were also calculated for the separate substituents, *i.e.* the N-1 side chain, C-5 hydrophobic moiety, and C-6 SAR.

These include the octanol/water partition coefficient ( $\log P$ ), used as a descriptor of the hydrophobic molecular properties (calculated by the fragmental Hansch and Leo method<sup>18</sup>), and descriptors based upon partial atomic charges, *i.e.* the sum and mean value over negative or positive atomic charges ( $\Sigma q^+$ ,  $\Sigma q^-$ ,  $\bar{q}^+$ ,  $\bar{q}^-$ ), except  $sp^3$  carbons with adjacent hydrogens. The partial atomic charges, as well as the dipole moment ( $\mu$ ), were calculated from the AM1<sup>17</sup> wave functions by the MO-PAC quantum chemical package in SYBYL 6.0.

The size and shape of the substituents were quantified by their van der Waals volume ( $V^{dW}$ ) and molecular dimensions: length ( $L$ ), width ( $W$ ), and height ( $H$ ), *i.e.* the projections on the inertial axes of the molecules and substituents. The ratios  $L/W$ ,  $L/H$ ,  $W/H$  which express the oblong degree of the molecules were also calculated.

As all of the HEPT analogues are extremely flexible, it was anticipated that molecular flexibility would play an important role in the structure–activity relationships. For this reason, in addition to the “classical” electronic, physicochemical, and geometric descriptors, properties related to the conformational changes, *i.e.* conformational barriers, were also included in the analysis. For this series of HEPT analogues, there are four rotatable bonds (bonds 1, 2, 3, and 5 in Figure 1) which can be found in every molecule of the set and, therefore, serve for producing molecular descriptors. The rotational barriers for these bonds were determined in the following way. First, we defined a rotational barrier in a manner which would allow us to unambiguously calculate it for any rotatable bond. In this definition, a rotational barrier is the energetic height of the barrier on the least energetically expensive way from one energetic minimum to another. According to this definition, a rotatable bond with one or two energetic minima produce only one value of rotational barrier. Another problem is that rotational barriers are very sensitive to the torsion angles of neighboring substituents. If the torsion angle of a neighboring substituent has two or more energetic minima, then two or more different values may be obtained for a given rotational barrier. In our case, each of bonds 1, 2, 3 has two minima and bond 5 has one minimum. This means that a particular rotational barrier value for a given bond

corresponds to each possible rearrangement of the other bonds. The number of possible rearrangements for all neighboring rotatable bonds is equal to the product of numbers of energetic minima for each neighboring rotatable bond. For bonds 1, 2, 3, the numbers of rotational barriers are equal to  $2 \times 2 \times 1 = 4$  and, for bond 2, the number is  $2 \times 2 \times 2 = 8$ . However, some of these rearrangements are isoenergetic (see the section entitled Molecular Modeling for details). Taking that into account leads to only two barriers for bond 3. The barrier values were determined through the rotation around bonds 1, 2, 3, 5 with the angle step of  $5^\circ$ . The energy was calculated at each step with no geometry optimization. When calculating these barriers, we found that switching bond 3 from one minimum to another has no effect on the barriers for bonds 1, 5. Thus, finally, we obtained two barrier values for bonds 1, 3, 5 and four barrier values for bond 4. These 10 values were intended for use as molecular descriptors in our QSAR analyses. Because of a gap between maximal and mean values of these descriptors, they were smoothed by taking log values.

To address the problem of chance correlation which may occur if a large number of descriptors is screened,<sup>19</sup> an *ad hoc* preselection of pertinent descriptors was made for each partial model. Thus, for the first partial model derived on a series of 17 compounds with variable C-6 substituents, 12 descriptors were preselected. The dipole moment ( $\mu$ ) was taken as a whole molecule descriptor, and 11 other descriptors were related to the C-6 substituent.  $V^{dW}$ ,  $L$ ,  $W$ , and  $L/W$ , were selected as shape descriptors for this series. The descriptors related to the height of the C-6 substituent were not taken because the 17 substituents within the series are mostly plane. Four descriptors, *i.e.*  $\Sigma q^+$ ,  $\Sigma q^-$ ,  $\bar{q}^+$ ,  $\bar{q}^-$ , were taken as electronic charge descriptors. Rotation barriers for the rotatable bond 5 in 1,6-*cis* and 1,6-*trans* conformers, referred to as RB5<sub>cis</sub> and RB5<sub>trans</sub>, were taken as flexibility descriptors. Finally,  $\log \pi$  was taken as a lipophilicity descriptor. No significant intercorrelations exceeding  $r = 0.8$  were detected among these 12 descriptors (except  $\Sigma q^-$  and  $\bar{q}^-$  whose intercorrelation is 0.9).

For the second partial model derived on a series of four compounds differing in C-5 substituent, four descriptors were preselected. These include two shape descriptors, *i.e.*  $V^{dW}$  and  $L$ . The width and height descriptors had a poor variation on this series. The other two descriptors include a flexibility descriptor, *i.e.*

a rotation barrier for the rotatable bond **3** referred to as RB3, and  $\log \pi$ . This rotation barrier for these compounds does not depend on the type of conformer (1,6-*cis* or 1,6-*trans*). It was found, however, that all four descriptors are strongly intercorrelated. Moreover, none of them correlates with activity. A possible solution to the problem is discussed in the Results and Discussion section.

The third subseries consists of 12 compounds with the N-1 and C-6 substituent varied. Again, electronic descriptors were not selected because of the nonpolar character of the varying fragments. Fourteen whole molecule shape descriptors were calculated for both 1,6-*cis* and 1,6-*trans* conformers. These are referred to as  $(V^{dW})_{cis}$ ,  $L_{cis}$ ,  $W_{cis}$ ,  $H_{cis}$ ,  $L/W_{cis}$ ,  $L/H_{cis}$ ,  $W/H_{cis}$ ,  $(V^{dW})_{trans}$ ,  $L_{trans}$ ,  $W_{trans}$ ,  $H_{trans}$ ,  $L/W_{trans}$ ,  $L/H_{trans}$ ,  $W/H_{trans}$ . Seven shape descriptors are related to the N-1 substituent. These are referred to as  $(V^{dW})_1$ ,  $L_1$ ,  $W_1$ ,  $H_1$ ,  $L/W_1$ ,  $L/H_1$ ,  $W/H_1$ . Three shape descriptors, *i.e.*  $L_5$ ,  $W_5$ ,  $L/W_5$ , were calculated for the plane C-5 substituent. The rotational barriers related to the rotatable bond 5 for both 1,6-*cis* and 1,6-*trans* conformers (RB5<sub>cis</sub> and RB5<sub>trans</sub>) were selected to serve as flexibility descriptors. The above 26 descriptors were then submitted to the correlation analysis. A significant correlation ( $r = 1.0$ ) was found between the  $L_5$  and  $W_5$  descriptors. Hence,  $W_5$  and  $W/L_5$  were removed from the descriptor set. Finally, 24 descriptors were preselected to search a model for the third subseries.

**Data Analysis.** Multiple linear regression (MLR) has been used to relate the RT inhibitory activity to the molecular descriptors. The stepwise procedure combining the forward and backward algorithms,<sup>20</sup> and thus addressing the multicollinearity problem, was used to select descriptors for the multiple regression equations. A cross-validation (CV) procedure<sup>21</sup> was employed to test the validity and predictive ability of the models. Five validation groups were randomly selected. When combining variable selection with cross-validation, we followed a suggestion of Wold who pointed out that "cross-validation does not work well... when cross-validation is applied after variable selection methods in stepwise regression (or other variable selection methods)".<sup>22</sup> In this work, after leaving out each cross-validation group, the stepwise selection of the variables was repeated, and we verified whether the stepwise algorithm always selects the same variables (this being an effective way to avoid chance correlations).

The results are expressed in terms of the regression coefficient estimates, the conventional and CV correlation coefficients ( $r$  and  $cv-r$ ), the standard deviations for estimation and prediction ( $s$  and  $cv-s$ ), and the Fisher statistics ( $F$ ).

## Results and Discussion

As a substituted uracil ring is a common element in the structure of the 87 HEPT analogues considered in this study, the observed RT inhibition activity for these molecules will be a function of the modification of the substituents on this central fragment. As a consequence, there are two principal effects which should be revealed by this QSAR study: (i) the influence of the variation at each substituent site and (ii) the interaction of the influences for two or more distinct substituents.

In this study, the following approach was used to account for the two aforementioned effects. First, three chemically sound partial models were derived from the compound subseries with a single substituent varied. A limited number of descriptors selected for the partial models was then used to build a general model on a subseries of *ca.* 40% of all available compounds. The remaining 60% of the HEPT analogues, including a small number of compounds with two and three simultaneously varied substituents, were employed as an external test set to validate the general model. Finally, after the validation of the descriptors used in the general model, the regression coefficients were readjusted on the whole series of 87 compounds.

At each step of the model building, the chance correlation problem initially raised by Topliss and Edwards<sup>19</sup> was addressed. The chance correlation is being produced when a large number of descriptors is screened, and therefore, descriptors which have nothing to do with the investigated activity occasionally give the best fit equation. Predictions made by such an equation on an external series of compounds should be very poor. A good prediction will also be a matter of chance. First, an *ad hoc* descriptor preselection was made for each of the three subseries. The descriptors which have no *a priori* physical relation to the inhibitory activity within a given series of compounds were removed from the screened descriptor set. Second, the cross-validation technique was combined in an appropriate manner with the variable selection method.<sup>22</sup> Each time when a next group of compounds was left out of the training set, the stepwise variable selection was repeated anew. Third, the predictive capacity of the model was tested on a test set which contained even more compounds than the training set.

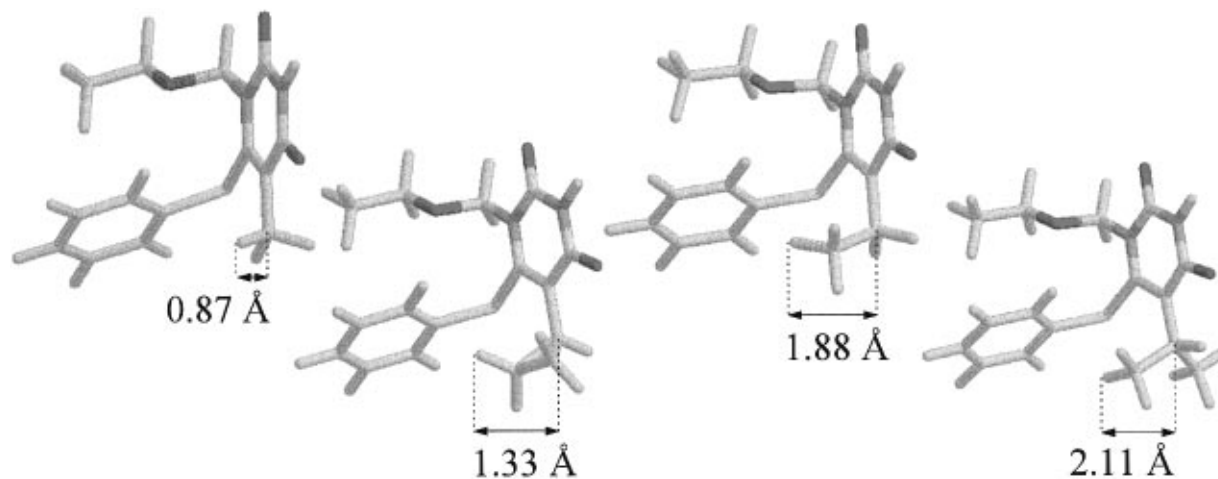
It should be noted that a technique referred to as experimental design<sup>23</sup> is usually suggested as a tool for selection of an optimal training set which would allow to study the two aforementioned effects. According to the principles of experimental design one should use in a training set the compounds which are simultaneously varied at several positions. However, this requirement is often difficult to meet due to synthetic difficulties. Thus, as compared to HEPT, only single modifications were made in the overwhelming majority of the analogues examined. Hence, in the present work, this technique of retrospective experimental design could not be employed.

**First Partial Model.** The first partial model was derived from a series of 17 HEPT analogues (**35–47**, **49–51**, HEPT) covering the range of activity from 82 to 0.26  $\mu\text{M}$ . All these compounds are 1-[(2-hydroxyethoxy)methyl]thymines with a meta-substituted C-6 thiophenyl substituent. Both geometric and electronic descriptors were found to be important. The best MLR for this series is the following

$$\log(1/EC_{50}) = 0.446(\pm 0.0973)(W_6) - 1.38(\pm 0.185)(\sum q^-)_6 - 1.97(\pm 0.545) \quad (1)$$

$$n = 17, \quad r = 0.92, \quad s = 0.26, \quad F = 36, \quad cv-r = 0.88, \\ cv-s = 0.31$$

where  $W_6$  is the width of the C-6 substituent, *i.e.* the projection on its second inertial axis, and  $(\sum q^-)_6$  is the



**Figure 3.** The geometric property of the C-5 substituent which correlates with the RT inhibitory activity.

total negative charge over the atoms of the C-6 substituent.

**Second Partial Model.** Another partial relationship has been built upon a small group of 5-alkyl-substituted 1-(ethoxymethyl)-6-(phenylthio)uracils (**3**, **21**, **30**, **32**). This series, like the other series with a single modification at the fifth uracil position (*e.g.* **7**, **18**, **19**), unambiguously shows that inhibitory activity increases smoothly with modification of the C-5 substituent in the following order, Me  $\rightarrow$  *c*-Pr  $\rightarrow$  Et  $\rightarrow$  *i*-Pr, where the influence of isopropyl is only slightly greater than that of ethyl. No descriptors from the standard preselected set follow this ordering. The only imaginable property which fits correctly with this relationship is the length of half of the C-5 substituent projected on the plane perpendicular to the uracil cycle (Figure 3). Though the definition of the descriptor may seem to be ponderous and artificial, it is easily understandable in the context of interactions with residues in the polymerase which lie to the left and toward the bottom from the uracil moiety as shown in Figure 3. The linear relationship of the activity to this descriptor is as follows:

$$\log(1/EC_{50}) = 1.19(\pm 0.0692)(W^5) - 0.557(\pm 0.0692) \quad (2)$$

$$n = 4, \quad r = 1.00, \quad sd = 0.0412, \quad F = 774$$

This smooth relationship between the activity and the structure of the C-5 substituent of the uracil ring suggests that the role of this substituent is probably more complex than conformational switching of the Tyr181 RT residue as suggested by the experimental study.<sup>11</sup> Indeed, from the crystal structures of RT complexed with HEPT and two more potent HEPT analogues, MKC-442 (**VII**) and TNK-651 (**VIII**), Hopkins *et al.*<sup>11</sup> suggested that the C-5 substituent forces this protein residue to change its conformation thereby increasing remarkably the protein-inhibitor affinity. From this mechanistic point of view the C-5 substituent should only be large enough to push the residue away from its normal position. Then, the 5-ethyl-, 5-cyclopropyl-, and 5-isopropyl-substituted compounds should have approximately the same activities. However, the activity of the 5-cyclopropyl-substituted compound **32** (0.1  $\mu$ M) is substantially lower than the activity of the compounds **21** (0.019  $\mu$ M), **30** (0.012  $\mu$ M). Such a

difference suggests that the C-5 substituent may have some additional functions in the protein–ligand docking process.

**Third Partial Model.** The third partial model relates the RT inhibitory activity to modification of the N-1 substituent, and to the interaction of the influences of the substituents at positions 1 and 6 in the uracil moiety. The necessity to treat simultaneously the compounds with the N-1 and C-6 substituents varied may be seen from the observed structure–activity relationships. For example, the 1-ethoxymethyl-substituted 5-ethyl-6-(phenylthio)thiouracil **7** has moderate activity of 0.026  $\mu$ M, while its 1-benzylmethoxymethyl analogue **13** (0.0078  $\mu$ M) is a 3.3 times more potent inhibitor. If there is no interaction between the influences of the N-1 and C-6 substituents on activity, then the same ratio is expected for compounds **8** and **14**. Yet, the ratio **8/14** = 0.64 is inverted, *i.e.* the 1-ethoxymethylated compound **8** (0.0044  $\mu$ M) is more active than the 1-benzylmethoxymethylated compound **14** (0.0069  $\mu$ M).

To summarize, the ratio between the activity values of 1-ethoxymethylated and 1-benzylmethoxymethylated 5-ethyl-6-(arylthio)uracils strongly depends on the structure of the substituent at the sixth position of the uracil. This fact leaves no chance to derive a good QSAR model using a method such as comparative molecular field analysis (CoMFA),<sup>24</sup> which considers the activity as a sum of structural increments.

A series of 12 5-ethyl-6-(arylthio)uracils (**21–29**, **57**, **59**, **61**) was employed as the training set to study the combined effect of the substituents at the positions 1 and 6 of the uracil ring. These compounds constitute the most representative subseries in which the relevant substituents are varied. For these systems the rotational barriers were taken to be important, because they express numerically the interactions between different fragments of the molecules.

The correlations have shown that two descriptors are relevant to the activity changes in this subseries. The first ( $W_{cis}$ ) is the width of the molecule, and the second descriptor ( $RB4_{cis}$ ) is the conformational barrier related to the rotation of the aryl group belonging to the 6-thioaryl moiety when the molecule is in the 1,6-*cis* conformation.

$$\log(1/EC_{50}) = 0.546(\pm 0.104)(W_{cis}) - 0.749(\pm 0.0992)(RB5_{cis}) + 4.97(\pm 0.981) \quad (3)$$

$$n = 12, \quad r = 0.95, \quad s = 0.343, \quad F = 38.4, \\ cv-r = 0.92, \quad cv-s = 0.356$$

The high values of the statistical and cross-validation criteria for eq 3 indicate that it probably addresses properties closely related to the mechanism of action of the HEPT analogues. The presence of the  $RB5_{cis}$  descriptor shows that the 1,6-*cis* conformation should be important for the inhibitory action, even though it is not the conformation in which these inhibitors are bound to the protein. A positive contribution of this descriptor suggests that the more stable a molecule is in the 1,6-*cis* conformation the more active it is. This conformation corresponds to the minimal volume and minimal linear dimensions of the molecules; hence it may be implicated in the entering of the inhibitor into the enzyme's hydrophobic pocket or in some earlier phase of the inhibitor-enzyme docking.

Again, as well as eq 2, this partial model points out an effect, which cannot be quantitatively interpreted from the crystal data,<sup>11</sup> *i.e.* that there is a significant variation of activity among the molecules with almost identical physicochemical, electronic and steric properties (see for example the molecules **67–78**). Indeed, it is the 1,6-*cis* conformation of the HEPT inhibitor which is present in the QSAR model and not the 1,6-*trans* conformation which was detected in the crystal structure.

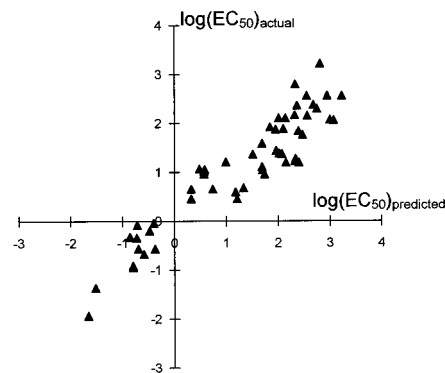
**General Model.** While deriving the partial QSAR models, five descriptors were selected as the most pertinent to the inhibitory activity. Finally, the five descriptors were submitted to the variable selection procedure in order to compose the best equation on the series of 33 (17 + 4 + 12) compounds which were previously used to derive the partial models. The following four-term equation was obtained:

$$\log(1/EC_{50}) = -1.55(\pm 0.302)(\sum q^-)_6 + 2.10(\pm 0.190)(^{1/2}W_5) - 0.728(\pm 0.122)(W_{cis}) + 0.362(\pm 0.0889)(RB5_{cis}) + 2.86(\pm 0.995) \quad (4)$$

$$n = 33, \quad r = 0.95, \quad sd = 0.43, \quad F = 70$$

At the next stage, this model was tested for its predictive power using a large series of 54 compounds which were not involved in the model development phase. The RT inhibitory activity values for these compounds were calculated using the general model (eq 4). The predicted inhibitory activities are plotted against the actual activity values in Figure 4. An appropriate measure of the model's predictive ability is the PRESS/SSY ratio,<sup>22</sup> where PRESS is predictive residual sum of squares and SSY is the sum of the squares of the experimental activity values. The PRESS/SSY ratio for this test set of 54 compounds is 0.11, indicating a good predictive quality of the model.

After chemically sound descriptors were selected and the predictive power was tested, the readjustment of the regression coefficients on a wider data set was done in order to generalize the model. This readjustment, increasing the interpolative abilities of the model, resulted in eq 5:



**Figure 4.** The actual activities plotted against the predicted activities for the test set of 54 RT inhibitors.

$$\log(1/EC_{50}) = -1.92(\pm 0.277)(\sum q^-)_6 + 1.43(\pm 0.113)(^{1/2}W_5) - 0.469(\pm 0.0719)(W_{cis}) + 0.410(\pm 0.0501)(RB5_{cis}) + 1.64(\pm 0.651) \quad (5)$$

$$n = 87, \quad r = 0.94, \quad sd = 0.46, \quad F = 147$$

**Comments on the Model Application.** The validation by a test set and the high values of the statistical estimates show that the model may be used for the prediction of RT inhibitory activity for new compounds. Another important problem which may be addressed by using eq 5 is the structural modification of the HEPT analogues in order to improve the RT inhibitory activity. However, the search for new HEPT type inhibitors is essentially constrained by previous SAR studies. In fact, the optimal C-5 and C-6 substituents are already found. It was demonstrated by the studies<sup>10,13–15</sup> that changing the size or attaching polar groups to these substituents decreases the activity. In the present QSAR model this qualitative observation is expressed by presence of the  $(\sum q^-)_6$  and  $W_{cis}$  descriptors.

The only site where new important modifications may be tried is the N-1 position of the uracil. Indeed, the substituent at N-1 may have larger volume and length than C-5 and C-6 substituents. The N-1 substituent may also contain H-bond donor and acceptor groups with no loss in activity, *e.g.*  $\beta$ -oxygen and terminal hydroxy group in many compounds. The synthesized compounds already provide information on the rational modifications at N-1 site. This information is expressed by the presence of the  $RB5_{cis}$  descriptor in the general model (eq 6). In fact, the variation of the N-1 substituent does not substantially influence on the physicochemical and geometric properties of the molecules. However, the inhibitory activity is widely varied even for closely related compounds. For example, the activity range for compounds **67–78** is from 0.35 to 0.0006  $\mu$ M, and the descriptors  $W_{cis}$  and  $RB5_{cis}$  accurately describe and predict this variation. It is apparent from the results obtained that the stability of the 1,6-*cis* conformation is one of the most important properties to check when planning a synthesis of new candidate molecules. Another important point which was not yet quantified is the role of the  $\beta$  ether oxygen in the 1-substituent. According to the crystallographic model of the enzyme-inhibitor complex, this oxygen interacts with the Tyr318 residue of the RT nonnucleoside hydrophobic pocket. More potent H-bond acceptor groups may be introduced in place of the  $\beta$  ester oxygen to explore in detail the role of this structural feature. Moreover, a new H-bond

acceptor group in the  $\beta$  position of the N-1 side chain, e.g. a carbonyl group, will also increase the stability of the 1,6-*cis* conformation and, hence, increases the probability of improving the activity.

### Conclusion

It has been shown in this study that the QSAR approach may be used as a complement to the experimental research techniques for better understanding of the HEPT type inhibitor's behavior in the biological system. A general model has been formulated through combination of three partial models, for each of which chemically sound descriptors have been thoroughly selected. The model's predictive ability has then been evaluated using a large test set of compounds. The technique of the combination of the partial models has permitted us to avoid using the retrospective experimental design whose multiple substituent variation requirement was not met by this series.

The derived quantitative model of the RT inhibitory activity is easily interpretable. A descriptor related to the conformational flexibility of the molecules has been found to be responsible for the activity variation among the most potent inhibitors.

Work is underway to generalize the model by involving some other classes of nonnucleoside HIV-1 RT inhibitors.

**Acknowledgment.** We would like to acknowledge the Fondation de Recherche Médicale for a postdoc fellowship grant for D.B.K. on the period of the present work. We also thank the Agence Nationale de Recherche contre le SIDA for research grants to Laboratoire de Chimométrie (Orléans), UPR CNRS 2301 (Gif-sur-Yvette), UMR CNRS 176 (Paris). We highly appreciate the constructive discussions with Dr. E. Bisagni (Institut Curie, Orsay) and Professor H. Buc (Institut Pasteur, Paris).

### References

- (1) De Clercq, E. HIV-1 specific RT inhibitors: Highly selective inhibitors of human immunodeficiency virus type 1 that are specifically targeted at the viral reverse transcriptase. *Med. Res. Rev.* **1993**, *13*, 229–258.
- (2) Schinazi, R. F. Competitive inhibitors of human immunodeficiency virus reverse transcriptase. *Perspect. Drug Disc. Des.* **1993**, *1*, 151–180.
- (3) Young, S. D. Nonnucleoside inhibitors of HIV-1 reverse transcriptase. *Perspect. Drug Disc. Des.* **1993**, *1*, 181–192.
- (4) De Clercq, E. HIV resistance to reverse transcriptase inhibitors. *Biochem. Pharmacol.* **1994**, *47*, 155–169.
- (5) Jonson, V. A. Combination Therapy: More Effective Control of HIV Type 1? *AIDS Res. Human Retrovir.* **1994**, *10*, 907–912.
- (6) Tantillo, Ch.; Ding, J.; Jacobo-Molina, A.; Nanni, R. G.; Boyer, P. L.; Hughes, S. H.; Pauwels, R.; Andries, K.; Janssen, P. A.; Arnold, E. Locations of Anti-AIDS Drug Binding Sites and Resistance Mutations in the Three-dimensional Structure of HIV-1 Reverse Transcriptase. *J. Mol. Biol.* **1994**, *243*, 369–387.
- (7) Balzarini, J.; Karlsson, A.; De Clercq, E. Human Immunodeficiency Virus Type 1 Drug-Resistance Patterns with Different 1-[(2-Hydroxyethoxy)methyl]-6-(phenylthio)thymine Derivatives *Mol. Pharmacol.* **1993**, *44*, 694–701.
- (8) Spence, R. A.; Kati, W. M.; Anderson, K. S.; Johnson, K. A. Mechanism of Inhibition of HIV-1 Reverse Transcriptase by Nonnucleoside Inhibitors. *Science* **1995**, *267*, 988–993.
- (9) van de Waterbeemd, H. Chapter 1. Introduction. In *Chemometric Methods in Drug Design*, van de Waterbeemd, H., Ed; VCH: Weinheim, 1995; pp 1–14.
- (10) Tanaka, H.; Takashima, H.; Ubasawa, M.; Sekiya, K.; Inouye, N.; Baba, M.; Shigeta, Sh.; Walker, R. T.; De Clercq, E.; Miyasaka, T. Synthesis and antiviral activity of 6-Benzyl Analogs of 1-[(2-Hydroxyethoxy)methyl]-6-(phenylthio)thymine (HEPT) as Potent and Selective Anti-HIV-1 Agents. *J. Med. Chem.* **1995**, *38*, 2860–2865.
- (11) Hopkins, A. L.; Ren, J.; Esnouf, R. M.; Willcox, B. E.; Jones, E. Y.; Ross, C.; Miyasaka, T.; Walker, R. T.; Tanaka, H.; Stammers, D. K.; Stuart, D. I. Complexes of HIV-1 Reverse Transcriptase with Inhibitors of the HEPT Series Reveal Conformational Changes Relevant to the Design of Potent Nonnucleoside Inhibitors. *J. Med. Chem.* **1996**, *39*, 1589–1600.
- (12) Pontikis, H.; Benhida, R.; Aubertin, A.-M.; Grierson, D. S.; Monneret, C. Synthesis and Anti-HIV Activity of Novel N-1 Side Chain Modified Analogs of 1-[(2-Hydroxyethoxy)methyl]-6-phenylthio)thymine (HEPT). *J. Med. Chem.* **1997**, *40*, 1845–1854.
- (13) Tanaka, H.; Takashima, H.; Ubasawa, M.; Sekiya, K.; Nitta, I.; Baba, M.; Shigeta, Sh.; Walker, R. T.; De Clercq, E.; Miyasaka, T. Synthesis and antiviral activity of Deoxy Analogs of 1-[(2-Hydroxyethoxy)methyl]-6-(phenylthio)thymine (HEPT) as Potent and Selective Anti-HIV-1 Agents. *J. Med. Chem.* **1992**, *35*, 4713–4719.
- (14) Tanaka, H.; Baba, M.; Hayakawa, H.; Sakamaki, T.; Miyasaka, T.; Ubasawa, M.; Takashima, H.; Sekiya, K.; Nitta, I.; Shigeta, Sh.; Walker, R. T.; Balzarini, J.; De Clercq, E. A New Class of HIV-1 Specific 6-Substituted Acylouridine Derivatives: Synthesis and Anti-HIV-1 Activity of 5- or 6-Substituted Analogues of 1-[(2-Hydroxyethoxy)methyl]-6-(phenylthio)thymine (HEPT). *J. Med. Chem.* **1991**, *34*, 349–357.
- (15) Tanaka, H.; Takashima, H.; Ubasawa, M.; Sekiya, K.; Nitta, I.; Baba, M.; Shigeta, Sh.; Walker, R. T.; De Clercq, E.; Miyasaka, T. Structure–Activity Relationships of 1-[(2-Hydroxyethoxy)methyl]-6-(phenylthio)thymine Analogues: Effect of Substitutions at the C-6 Phenyl Ring and at the C-5 Position on Anti-HIV-1 Activity. *J. Med. Chem.* **1992**, *35*, 337–345.
- (16) Clarc, M.; Cramer, R. D. III; Opendbosch, N. V. Validation of the general purpose Tripos 5.2 force field. *J. Comput. Chem.* **1989**, *10*, 982–1012.
- (17) Dewar, M. J. S.; Zoebich, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- (18) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley Interscience: New York, 1979.
- (19) Topliss, J. G.; Edwards, R. P. Chance Factors in Studies of Quantitative Structure–Activity Relationships. *J. Med. Chem.* **1979**, *22*, 1238–1244.
- (20) *STATlab User's Guide*, SLP Statistiques, 51, 59 rue Ledru Rollin, 94200, Ivry, France.
- (21) Cramer, R. D.; Bunce, J. D.; Patterson, D. E.; Frank, I. E. Crossvalidation, Bootstrapping, and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies *Quantum Struct.-Act. Relat.* **1988**, *7*, 18–25.
- (22) Clementi, S.; Wold, S. How to Choose the Proper Statistical Method. Section 5.2.6 Conclusions. In *Chemometric Methods in Drug Design*, van de Waterbeemd, H., Ed; VCH: Weinheim, 1995; pp 319–338.
- (23) Austel, V. Chapter 3. Experimental Design. In *Chemometric Methods in Drug Design*, van de Waterbeemd, H., Ed; VCH: Weinheim, 1995; pp 49–62.
- (24) Cramer, R. D. III; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.

JM970110P