Journal of Computer-Aided Molecular Design, **16:** 287–295, 2002.
KLUWER/ESCOM

HEPT derivatives as non-nucleoside inhibitors of HIV-1 reverse transcriptase: QSAR studies agree with the crystal structures

Anderson Coser Gaudio[a,*] and Carlos Alberto Montanari[b]
[a]*Departamento de Física, Centro de Ciências Exatas, Universidade Federal do Espírito Santo, Campus de Goiabeiras, 29.060-900 Vitória - ES, Brazil;* [b]*Departamento de Química, Universidade Federal de Minas Gerais, Campus da Pampulha, 31.270-901 Belo Horizonte - MG, Brazil*

Received 22 October 2001; Accepted 18 June 2002

Key words: human immunodeficiency virus type 1, reverse transcriptase, quantitative structure-activity relationships, molecular modeling, acyclouridine derivatives

Summary

The interest in the non-nucleoside inhibitors (NNIs) to the reverse transcriptase (RT) as anti-AIDS agents has grown in the last ten years. The compound 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine (HEPT) is the precursor of the most studied class of NNIs, from which hundreds of derivatives have been synthesized and tested. There are at least twelve QSAR studies about the HEPT derivatives as RT inhibitors. Most of the predictions derived by these studies are related to the nature of the active site near the substituents at positions N-1 and C-5, and at the C-6 phenyl ring. The validity of these models has been checked against the 3-D structure of HIV 1 RT-HEPT complexes available. Most of these predictions were confirmed at the molecular level.

Abbreviations: AIDS, acquired immunodeficiency syndrome; CoMFA, comparative molecular field analysis; DNA, deoxyribonucleic acid; HEPT, 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine; HIV, human immunodeficiency virus; PDB, protein databank; PLS, partial least squares; QSAR, quantitative structure-activity relationships; RCSB, Research Collaboratory for Structural Bioinformatics; RT, reverse transcriptase.

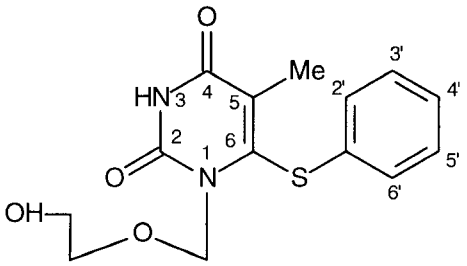
Introduction

The acquired immunodeficiency syndrome (AIDS) is probably the most investigated disease since it was discovered in the earlier eighties. Huge amounts of money and human effort have been spent in the search of a possible cure, but in spite of these, the AIDS virus is still resisting and no absolute successful chemotherapy is available to date. The AIDS virus is a retrovirus that belongs to the lentivirus subfamily. Two of its enzymes became the preferred targets for developing anti-AIDS drugs: reverse transcriptase (RT) [1–3] and protease [4–8]. The RT is responsible for making a copy of the viral RNA into a complementary DNA as soon as the virus enters the host cell, a process named transcription. Protease is mainly involved in the virion maturation, a later stage in the virus life cycle.

Although many drugs are currently available in the market, the search for new HIV enzyme inhibitors is still going on, especially because of the induced resistance most of these drugs cause. Tanaka and co-workers studied a series of RT inhibitors derived from 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine (HEPT; Structure **1**) [9–11]. They also performed structure-activity studies about the influence of the chemical groups of structure **1** on the HIV-1 RT inhibitory activity [12–16]. The quantitative biological data presented by Tanaka and co-workers in these papers is the basis of all the quantitative structure-activity relationship (QSAR) studies published to date.

The HEPT derivatives are probably the most studied group of non-nucleoside inhibitors to HIV-1 RT.

*Corresponding author: Phone: 0055-27-3335.2483; Fax: 0055-27-3335.2823; e-mail: anderson@npd.ufes.br



The crystal structure of RT in complex with many non-nucleoside inhibitors, including HEPT, was elucidated in 1995 [17]. Therefore, the structure of the non-nucleoside binding site at RT and the binding mode of HEPT derivatives are known since then. There are at least twelve QSAR studies about these compounds described in the literature [18–29]. Considering that most of the these studies have been published after 1995, it is surprising that just a minor part of them [22, 29] has made some considerations about the QSAR results and the 3-D structure of HIV-1 RT-HEPT complex.

The purpose of the present paper is to review the main conclusions of the published QSAR studies on HEPT derivatives and to check their validity against the 3-D structure of HIV-1 RT-HEPT complex.

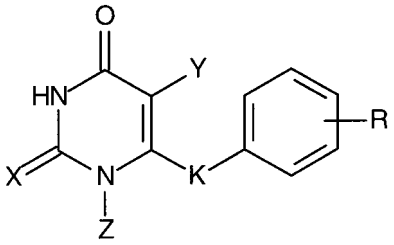
Methods

Most of the calculation in the QSAR models presented in this paper has been checked with the program BuildQSAR, developed by De Oliveira and Gaudio [30]. The crystal structures of the HEPT-RT complexes were taken from the Protein Data Bank of the Research Collaboratory for Structural Bioinformatics (RCSB) [31]. The receptor sites of the substituents at positions N-1, N-3, C-5, and C-2′-C-6′ were mapped by selecting all the amino acid residues in the range of 6 Å from the non-hydrogen atoms of the substituent at the position considered. The computer program BioMedCAChe [32] has been used to import, refine and relax the molecular structures in the PDB files. The program WebLab Viewer [33] generated the images of the HEPT derivatives in the active site of the HIV-1 RT.

QSAR studies on HEPT derivatives

The first attempt to establish structure-activity relationships for the HEPT derivatives was made by Tanaka and co-workers in 1992 [15]. At that time, the investigations indicated that the anti-HIV activity of HEPT (Structure **1**) could be increased by the attachment of a hydrophobic group to the C-3′ and C-5′ positions of the C-6 thioaryl group. It was noticed that substitutions at the C-2′ and C-4′ positions appear to be ineffective, while the replacement of the C-5 methyl by a more lipophilic group, specially an isopropyl, promoted a marked increase in the anti-HIV activity. It was also reported that the hydrogen atom at N-3 position is essential for the anti-HIV activity.

Hansch and Zhang built the first QSAR model (Eq. 1) for the HEPT derivatives (Structure **2**, where X = O (27 compounds), S (6); Y = Me (20), Et (6), Pr (2), i-Pr (4), $CH_2CH{=}CH_2$ (1); Z = $CH_2OCH_2CH_2OH$; K = S; R = Miscellaneous) [18].



$$\log(1/EC_{50}) = 0.88(\pm 0.39)\Sigma\pi_{Y,R} + \\ +12.1(\pm 3.8)\, L_Y - 1.59(\pm 0.49)\, L_Y{}^2 + \\ +1.17(\pm 0.86)\, B_{1,R3'} + 1.53(\pm 0.82)\, E_{s,R2'} - \\ -15.3(\pm 7.4) \qquad (1)$$

$$(n = 33;\ r = 0.941;\ s = 0.500;\ L_{Yopt} = 3.8)$$

In this equation, EC_{50} is the molar concentration of drug affording 50% protection to MT-4 cells [15], $\Sigma\pi_{Y,R}$ is the sum of π_Y and π_R, where π is the lipophilic parameter [34], L_Y is the sterimol parameter [35] for the length of the substituent attached to position C-5, $B_{1,R3'}$ is the sterimol parameter for the width of R in C-3′ position, and $E_{s,R2'}$ is the Taft steric constant [36] for the ortho substituents (R at position C-2′). The numbers in parenthesis are the 95% confidence limits to the coefficients, n is the number of compounds considered in the model, r is the correlation coefficient, s is the standard deviation, and L_{Yopt} is the (optimum) value of the L_Y parameter that results in maximum activity. Equation 1 clearly suggests that Y and R most influence the inhibitory activity. Substituent Y should be more lipophilic, although limited in size, in order to be more effective. In the case of R, there are two consequences. Considering that all

non-zero values of $E_{s,R2'}$ that contributed to the development of Eq. 1 are negative, the positive coefficient of this parameter tell us that the ortho position must have a small group, preferably C-2′ = H, for optimum activity. The positive coefficient of $B_{1,R3'}$ indicates that a group other than hydrogen should be attached to C-3′ or/and C-5′ positions.

Garg and co-workers [21] made an extensive QSAR study from which two models are presented. The first one included 36 compounds derived from structure **2**, where X = O (28), S (8); Y = Me (23), Et (6), Pr (2), i-Pr (4), $CH_2CH{=}CH_2$ (1); Z = $CH_2OCH_2CH_2OH$; K = S; R = Miscellaneous (Eq. 2).

$$\log(1/EC_{50}) = 12.215(\pm4.297)\pi_Y - 5.716(\pm2.197)(\pi_Y)^2 + 1.876(\pm0.808)I_{iso,Y} + 1.305(\pm0.411)I_{R3',R5'} - 0.955(\pm0.521)\, I_{R2',R4'} - 0.018 \quad (2)$$

$(n = 36; r = 0.949; s = 0.46; F = 54.80; r_{cv}^2 = 0.839)$

In this equation, $I_{iso,Y}$ is an indicator variable which accounts for a specific effect of an isopropyl group at position C-5, $I_{R3',R5'}$ and $I_{R2',R4'}$ are indicator variables which account for an R-substituent at positions C-3′ and C-5′, and C-2′ and C-4′, respectively, F is the variance ratio and r_{cv} is the cross-validation correlation coefficient. Equation 2 is a quantitative evidence of the qualitative observations made by Tanaka [15]. The coefficient of parameter $I_{iso,Y}$ indicates that an isopropyl group at C-5 would increase log $1/EC_{50}$ by approximately 1.9 units when compared to other groups. The coefficients of $I_{R3',R5'}$, and $I_{R2',R4'}$ indicate that positions C-3′ and C-5′ should be substituted (R ≠ H), while C-2′ and C-4′ should remain unsubstituted (R = H) in order to increase the activity. It is important to notice that Eq. 2 detected a non-linear dependence of the biological response on the lipophilicity. A possible interpretation is that substituent Y may be involved in a hydrophobic interaction with some hydrophobic pocket of the active site, which seems to have limited bulk tolerance [21]. This result is in accordance with Eq. 1, which showed a non linear dependence of the inhibitory activity in terms of the length of Y.

The second model (Eq. 3) included 35 compounds: X = O (21), S (14), Y = Me (9), Et (20), i-Pr (4), c-Pr (2); Z = Miscellaneous hydrocarbons; K = S; R = Miscellaneous.

$$\log(1/EC_{50}) = 2.325(\pm0.618)\pi_Y - 1.457(\pm0.883)\, V_Z + 1.562(\pm0.553)\, I_{phe,Z} + 1.061(\pm0.447)\, I_{R3',R5'} + 5.444 \quad (3)$$

$(n = 35; r = 0.917; s = 0.47; F = 39.68; r_{cv}^2 = 0.790)$

In this equation, V_Z is the van der Waals volume of Z substituent and $I_{phe,Z}$ is a indicator variable which has a value of unity for Z substituents that contain an aromatic phenyl moiety. Garg and co-workers published slightly modified versions of Eqs. 1 and 3 in a recent review study [25]. In Eq. 3, the negative coefficient of V_Z denotes the existence of a region of limited size in the receptor site around position N-1, which is in accordance with the previous observations [15, 18, 19]. The meaning of the parameter $I_{phe,Z}$ is somewhat contradictory. While V_Z points to a small chemical group at N-1 for higher activity, the coefficient of $I_{phe,Z}$ indicates that if there was a phenyl moiety in the Z substituent, the activity would be increased by approximately 1.6 when compared to a Z with no phenyl. The contradiction comes up when we check the dataset that generated Eq. 3 [21] and verify that the Z groups with a phenyl moiety are the ones with greater volume. A possible explanation is that when Z has a phenyl, it may be involved in a stacking interaction with some residue in the receptor site.

Kireev and co-workers [22] performed a 3-D QSAR study involving 87 HEPT derivatives, which generated Eq. 4: X = O (68), S (19); Y = Me (30), Et (39), i-Pr (14), c-Pr (2), I (1), $CH{=}CH_2$ (1); Z = Miscellaneous; K = S (73), CH_2 (13), O (1); R = Miscellaneous.

$$\log(1/EC_{50}) = -1.92(\pm0.277)(\Sigma\, q^-)_{C6} + 1.43(\pm0.113)^{1/2}W_Y - 0.469(\pm0.0719)\, W_{cis} + 0.410(\pm0.0501)\, RB_{cis,Y} + 1.64(\pm0.651) \quad (4)$$

$(n = 87; r = 0.94; s = 0.46; F = 147)$

In this equation, $(\Sigma q^-)_{C6}$ is the total negative charge over the atoms of the C-6 substituent, W_Y is the width of the C-5 substituent, W_{cis} is the width of the molecule and $RB_{cis,Y}$ is the conformational barrier related to the rotation of the aryl group belonging to the C-6 thioaryl moiety when the molecule is in the 1,6-cis conformation. Equation 4 is the result of a variable selection process that used a huge dataset with dozens of calculated parameters (the dataset has not been published). That is why some unusual parameters, such as W_{cis} and $RB_{cis,Y}$, appeared in the model. Among the

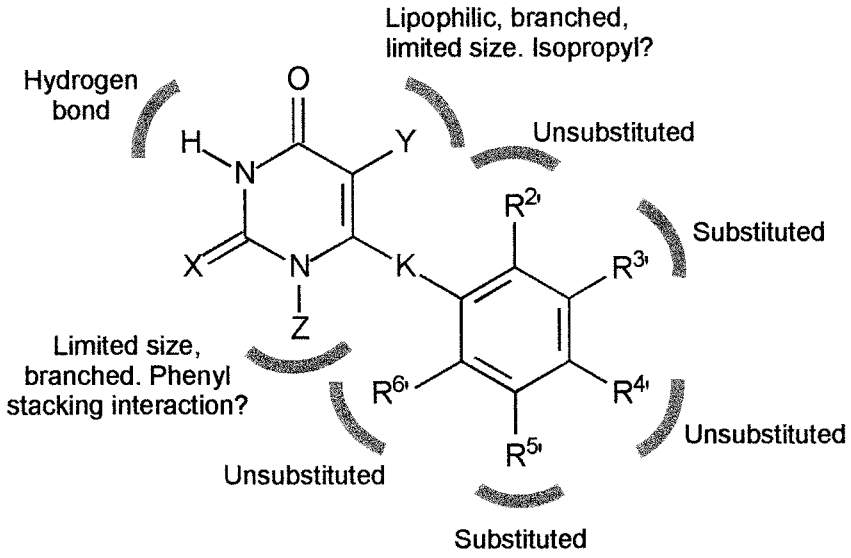


Figure 1. Main conclusions of the QSAR studies involving HEPT derivatives.

variables in Eq. 4, only W_Y seems to have a direct relationship with the previous results. The presence of two variables related to the 1,6-cis conformation is a point that must be seen with caution. The crystal structure of the enzyme inhibitor complex revealed that the thymine ring of the HEPT derivatives bound to the receptor in almost planar conformation [16]. The authors suggested that W_{cis} and $RB_{cis,Y}$ may be implicated in the entering of the inhibitor into the hydrophobic pocket of the enzyme in some earlier phase of the enzyme-inhibitor interaction. The meaning of the parameter $(\Sigma q^-)_{C6}$ may be related with a possible existence of a positively charged region near position C-6.

Luco and Ferretti [23] built a QSAR model (Eq. 5) including 107 HEPT derivatives: X = O (85), S (22); Y = Me (52), Et (26), I-Pr (8), Miscellaneous (21); Z = $CH_2OCH_2CH_2OH$ (64), CH_2OCH_2Me (11), CH_2OCH_2Ph (7), Miscellaneous (25); K = S; R = Miscellaneous.

$$\begin{aligned}\log(1/EC_{50}) = {} & 0.65(\pm0.18)\,\Sigma\pi_{YR} + \\ & +130(\pm34)\,{}^1\chi_Y^N - 198(\pm46)({}^1\chi_Y^N)^2 - \\ & -1.14(\pm0.38)\,B_{1,R3'} + 1.60(\pm0.56)\,E_{s,R2'} + \\ & +63(\pm14)\,{}^4\chi_Y^N + 12.0(\pm4.6)V/100 - \\ & -2.13(\pm0.86)(V/100)^2 - 2.27(\pm0.60)\,I_{ch,Z} - \\ & -1.25(\pm0.40)\,I_{R4'} - 0.72(\pm0.42)\,{}^0\Delta X_Z - \\ & -47.7(\pm5.92) \\ & (n = 107;\ r = 0.951;\ s = 0.49;\ F = 82.08; \\ & r_{cv}{}^2 = 0.67)\end{aligned} \tag{5}$$

In this model, χ's are connectivity molecular indexes, ${}^0\Delta X_Z$ is a zeroth order differential molecular connectivity index related to substituent Z, V is the molecular volume, $I_{ch,Z}$ is an indicator variable which takes the value of 1 or 0 for the presence or absence of a six-membered saturated ring in Z. Many of the previous results are confirmed by Eq. 5. The biological response is favored by the lipophilicity of Y and R. The substituent at C-2′ should be small, preferably R2′ = H, and the overall molecular volume should have an optimum value. The parameter $I_{ch,Z}$ tell us that if Z had a six-membered saturated ring, the activity should decrease. In Eq. 3, the parameter $I_{phe,Z}$ indicated that a phenyl moiety in Z is advantageous. The meaning of parameters $I_{ch,Z}$ and $I_{phe,Z}$ seems to be consistent if a possible stacking interaction with the receptor is supposed. A planar ring such as phenyl would have a more favored interaction with a phenylalanine

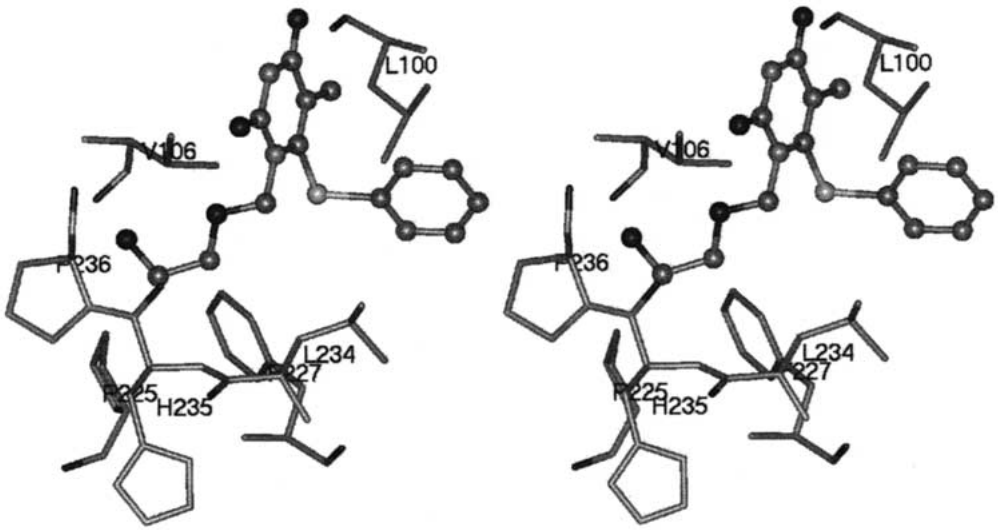


Figure 2. Stereoview of the region surrounding the Z group of HEPT (Structure **1**) in its binding site of HIV-1 RT (PDB file 1RTI [17]). The hydrogen atoms have been omitted for clarity.

or a tyrosine residue in the receptor than a boat-like or a chair-like ring such as cyclohexyl. The authors interpreted the molecular connectivity indexes in Eq. 5 as an indication that the activity is favored by increasing the number of R substituents ($^4\chi_p^N$), by increasing the branching of Y ($^4\chi_p^N$ and $^1\chi_Y^N$), and by decreasing the polarity of Z ($^0\Delta X_Z$) [23]. Although highly significant statistically, the low r_{cv} value of Eq. 5 reveals its low predictability when compared to the other models presented above. According to the authors, this fact may be due to inaccuracies in the biological testing of some of the compounds incorporated in the analysis. In fact, a model very similar to Eq. 5 presented in the same paper, but including only n = 79 compounds, showed the following statistics: ($r = 0.949$; $s = 0.439$; $F = 69.45$; $r_{cv}^2 = 0.745$). Probably what contributed most to the low predictability of Eq. 5 was the high degree of the collinearity among the descriptors, especially those that make the principal contribution to the activity (the intercorrelation allowed was 0.7). A PLS model [23], which is insensible to the collinearity among the descriptors, based on all compounds and variables as used in Eq. 5 produced the following statistics: ($r = 0.944$; $s = 0.503$; $F = 283.57$; $r_{cv}^2 = 0.867$).

Finally, Hannongbua and co-workers [29] performed a 3D-QSAR study, using structures of 101 HEPT derivatives calculated by ab initio methods. The molecular geometry used in the calculation was taken from the crystal structure of HIV-1 RT-HEPT complexes [16]. This study attempted to find explanations for the steric and the electrostatic interactions of the molecules with the amino acids of the inhibition pocket of the enzyme. The most relevant PLS model was derived from the data obtained with the $sp^3C(+1)$ probe atom. This included four principal components and accounted for electrostatic and steric field contributions, being the later responsible for 63.9% of the total effect. The statistics of this model is: ($r = 0.964$; $s = 0.371$; $F = 255.4$; $r_{cv}^2 = 0.837$). The results indicated that the steric contours they form are consistent with possible interactions between the Y side chain and the aromatic ring of Tyr181. Once again, the findings pointed that Y should be branched, but not too large, which suggests that Y should be an isopropyl group. Substituents R and Z, on the other hand, should not be branched in order to preserve the interactions between these groups and the active site. In terms of electrostatic interactions, the results showed that a positively charged environment near positions C-5 and C-6 are important for the mechanism of the drug-receptor interaction. A previous CoMFA analysis due to the same authors [20] also suggested the existence of a favorable steric space along the plane of the C-6 thioaryl group, closed to positions C-2′ and C-3′, and near position C-5 in the thymine ring. This study also

showed that the space near position N-1 is sterically unfavorable.

As it was shown, most of the QSAR studies involving HEPT derivatives described so far attempted to identify specific contributions of the different parts of their molecular structure to the inhibitory activity. The knowledge collected indicates that the activity of these compounds is mainly sensitive to the molecular modification of three regions of their structure: (a) the substituent Z attached to N-1, which should be preferably of limited size and not branched, but if it had a phenyl moiety, the volume limitation would be overcome; (b) the Y at C-5 position, which should be small, branched and lipophilic, preferably an isopropyl group; (c) the meta positions of the C-6 thiophenyl substituent, which should have R $\neq$ H. It was also revealed that modifications at position C-2 (X) and the replacement of the sulfur atom by oxygen (K) do not influence the level of the inhibitory potency. The replacement of the C-2′ and C-4′ hydrogen atoms results in deleterious effect to the activity. Finally, the hydrogen atom at N-3 position cannot be substituted. Its replacement will abolish the anti-HIV-1 activity. These results can be summarized in Figure 1.

Results and discussion

The N-1 position

The site of the substituent $CH_2OCH_2CH_2OH$ of HEPT in HIV-1 RT is formed by seven residues: Leu100, Val106, Pro225, Phe227, Leu234, His235 and Pro236 (Figure 2). These amino acids form a lipophilic environment around the Z site. The only polar residue in this group (His235) is projected to the opposite direction of the site.

Equation 3 is the only model that made explicit reference to the substituent Z, which indicated that Z should have small volume, but if it had a phenyl group this should increase the biological response. Garg and co-workers [21] suggested that the presence of a phenyl ring in Z may result in a good stacking with the receptor, which could overcome the low volume restriction. A quick examination on Figure 2 shows that in fact there are a phenylalanine and two proline residues in the Z site in good conditions to have a van der Waals interaction with a phenyl of Z. The crystal structure of compound 6-benzyl-1-benzyloxymethyl-5-isopropyluracyl (X = O, Y = i-Pr; Z = CH_2OCH_2Ph, K = CH_2, R = H) in complex with HIV-1 RT has been determined (PDB file 1RT2 [16]) and can be used to shed some light to this point. Figure 3 shows that the phenyl moiety of Z substituent is being sandwiched by Pro236, the strongest interaction, and Val106. The phenylalanine residue plays a minor part in the stabilization of Z.

Among the compounds included in the composition of Eq. 3, mainly two types of Z substituents containing phenyl were used: CH_2OCH_2Ph and $CH_2OCH_2CH_2Ph$. Most of the time the phenyl group is unsubstituted. Only in few cases there was a substituent attached to its para position. Considering only the unsubstituted phenyl moieties, the compounds with Z = CH_2OCH_2Ph presented higher activity than with Z = $CH_2OCH_2CH_2Ph$. For example, the inhibitory activity of the compound defined by X = O, Y = Et, Z = CH_2OCH_2Ph, K = S, R = H, measured as log $1/EC_{50}$, is 8.23 [15]. If Z = $CH_2OCH_2CH_2Ph$, the activity drops to 7.02. The same effect is observed in the case of structure X = S, Y = Et, K = S, R = H. If Z = CH_2OCH_2Ph, the inhibitory activity is 8.11. Replacing Z by $CH_2OCH_2CH_2Ph$, the activity drops to 7.04. It can be concluded that a shorter connection between N-1 and phenyl favor the ring stacking. A longer connection chain pushes the phenyl ring away from the stacking zone, lowering the possibility of a reasonable ring parallelism.

The N-3 position

In the space defined by 6 Å from the atom N-3 in the active site of HIV-1 RT there are four amino acids: Leu100, Lys101, Lys103, and Lys103. The N3–H moiety is in a hydrogen bond interaction with the carbonyl group of Lys101. This interaction is so important for the anti-HIV-1 activity that many studies reported that a substitution at this position makes the activity to be abolished [10-12, 14, 15, 37].

The C-5 position

In the space surrounding the C-5 position of the HEPT derivatives in the active site of HIV-1 RT (PDB structure 1RT1 [16]) there are six residues, which form a pocket that accommodates the Y substituent (Figure 4): Val106, Val179, Tyr181, Tyr188, Val189, and Gly190. Both OH groups of the tyrosine residues are projected far away from the pocket region, letting only their hydrocarbon moieties making part of the Y site. As tyrosine is the only polar amino acid among those six in the site, these make it a lipophilic place.

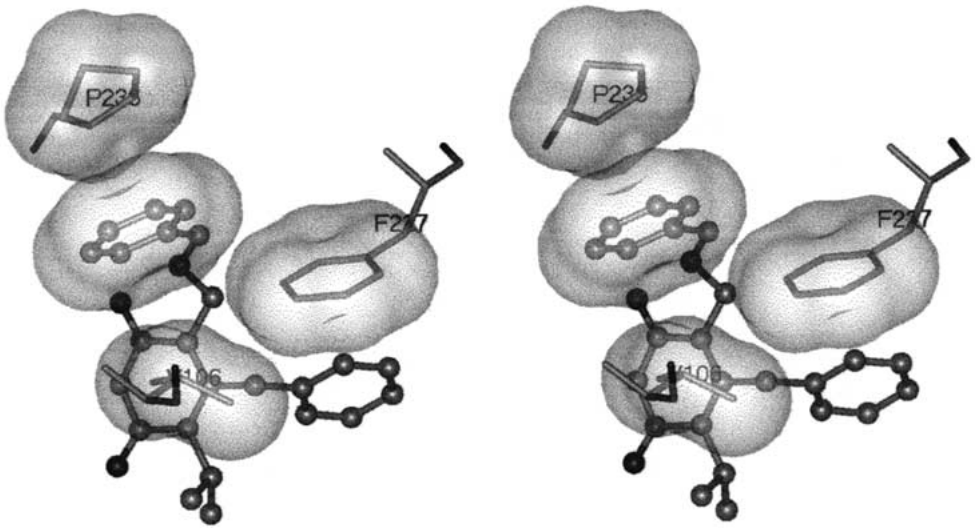


Figure 3. Stereoview of the region surrounding the Z group of a HEPT derivative (X = O, Y = i-Pr; Z = CH_2OCH_2Ph, K = CH_2, R = H) in its binding site of HIV-1 RT (PDB file 1RT2 [16]) showing that the phenyl moiety of Z substituent is being sandwiched by Pro236, the strongest interaction, and Val106. The Phe227 plays a minor part in the stabilization of Z.

The space available for the Y group in its pocket is of very limited size. Substituent Y is placed in between the valine residues, which are located near the thymine ring, and in between the Cβ atoms of the tyrosine residues, which are spatially oriented almost in parallel to the phenyl ring of HEPT (Figure 4). The problems related to the space in the lipophilic pocket available to Y have been foreseen in Eqs. 1, 2, 4 and 5. In these models, Y should be lipophilic, but of limited length and branched, which suggest an isopropyl group as the ideal ligand. In Figure 4, it is clear that the isopropyl group in the C-5 position has a perfect match to its lipophilic pocket. Smaller substituents, such as methyl or ethyl should leave some unfilled space when interacting with the Y site. Larger substituents, such as COOMe, should have problems to fit the available space and, as consequence, this may force the thymine ring to be displaced from its ideal position in the binding site. The lack of adjustment of the ligand of course results in loss of affinity. This can also be easily checked. The inhibitory activity of the HEPT derivative defined by X = O, Y = i-Pr, Z = $CH_2OCH_2CH_2OH$, K = S, R = H is 7.23 [15]. If the isopropyl is replaced by a methyl, the activity drops to 6.01. Replacing methyl by ethyl, the activity becomes 6.96. Making Y = Pr results in a strong decrease in the activity (log $1/EC_{50}$ = 5.00). Although isopropyl and propyl are isomers, the later has a greater length than the former, which seems to be the ideal substituent. Some extreme cases are: Y = $COCHMe_2$ (log $1/EC_{50}$ = 4.92), COPh (4.89) and SPh (4.68) [14].

The C-6 position

The site of the SPh group of HEPT in active site of HIV-1 RT (PDB file 1RTI [17]) is composed by seven residues (Fig. 5): Pro95, Leu100, Val106, Tyr181, Tyr188, Phe227 and Trp229. As four of these amino acids are hydrophobic and three are polar, these give the site a double character.

It was previously reported that the HEPT derivatives substituted at C-2′ or C-6′ position are less active than the equivalent C-3′ or C-5′ substituted ones. For example, the compound X = O, Y = Me, Z = $CH_2OCH_2CH_2OH$, K = S and R = H show log $1/IC_{50}$ = 5.15 [15]. Making R = 2-Me, the activity becomes 4.15, while if R = 3-Me the activity goes to 5.59. Considering only the configuration of the active site environment of SPh, it is difficult to find a reason why ortho and meta-substituted derivatives of HEPT present so remarkably different activities. At least visually, there is space in the binding site to well accommodate both derivatives. Of course one of the ortho positions, the one nearest to the thymine ring, is less able to accommodate a substituent due to its proximity to the rest of the molecule. So, how can one explain the drop in the activity when a single methyl group is added to C-2′? Kireev and co-workers [22] raised one possible explanation in their model (Eq. 4).

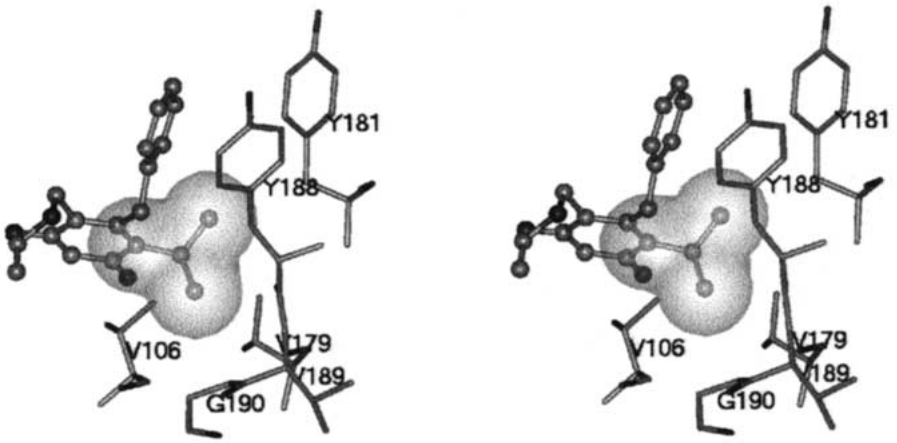


Figure 4. Stereoview of the lipophilic region surrounding the Y group of an HEPT derivative (6-benzyl-1-ethoxymethyl-5-isopropyl-uracil (X = O, Y = i-Pr, Z = CH_2OEt, K = CH_2, R = H) in its binding site of HIV-1 RT (PDB structure 1 RT1 [16]). A van der Waals surface was applied to the isopropyl group bound to the C-5 position of the inhibitor in order to check the occupancy of the Y site.

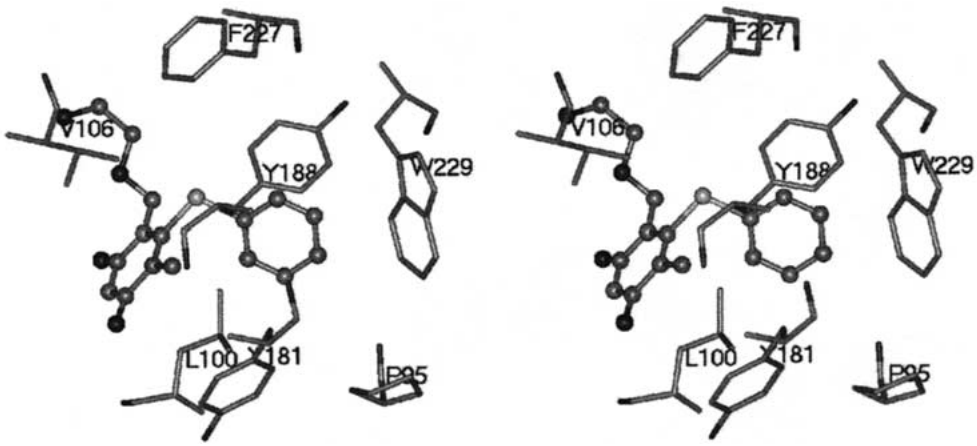


Figure 5. Stereoview of the region surrounding the C-6 substituent of HEPT (Structure **1**) in its binding site of HIV-1 RT (PDB structure 1RTI [17]).

They suggested that the entering of the drug into the receptor, which could be forced to adopt different conformations to get there, could influence the activity level. If this is a really important factor, one can suppose that the ortho derivatives show lower activity due to their difficulty to change the conformation of the SPh group, especially when this involves a phenyl rotation.

In the previous example, if the methyl group is placed at C-4′ (para position), the activity becomes 3.66, being the lowest of the four values. The poor activity of the para substituted derivatives cannot be explained by the above hypothesis. One possible explanation could be the proximity of the C-4′ substituent to the Pro95 and to the huge Trp229 residues in the active site (Figure 5). A molecular graphics substitution of the 4′-H by Me showed that the distance between the hydrogen atoms of the methyl group and the closest atoms of Pro95 and Trp229 are in the range of 2.1 and 2.4 Å. It seems that both Pro95 and Trp229 can be pointed as the cause of the negative influence of the para substituents in the SPh group of the HEPT derivatives. The C-3′ and the C-5′ substituted derivatives are subject to neither of these influences and, following this line, they can reach the active site and have favored interactions with the residues around the SPh group.

Conclusions

In the last ten years, nearly a dozen of QSAR studies about HEPT derivatives as HIV-1 RT inhibitors have been published. The quantitative models in these papers used a diverse set of properties, such as classical substituent constants, atomic and molecular quantum-chemical parameters, connectivity indexes, CoMFA derived parameters, indicator variables, etc. Although looking at a glance most of those models seem to be different, a deeper analysis shows that the biochemical implications coming from their predictions are not too many. The main predictions were related to the substituents at positions N-1, C-5, and to the substituents at the C-6 phenyl ring. As it was shown in the present paper, most of the predictions have been confirmed at the molecular level. This is an evidence of the power of the QSAR methodology.

Acknowledgements

The authors thank the CAChe Group of Fujitsu America Inc. for providing a temporary license for program BioMedCAChe. We also thank Dr. Márcia Miguel Castro Ferreira from the Instituto de Química of Universidade Estadual de Campinas for the use of program WebLab Viewer. This work received financial support from Programa de Pós-Graduação em Física of Universidade Federal do Espírito Santo - PPGFIS-UFES, Fundação de Amparo à Pesquisa de Minas Gerais - FAPEMIG and Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq.

References

1. Pedersen, O.S. and Pedersen, E.B., Antivir. Chem. Chemother., 10 (1999) 285.
2. De Clercq, E., Farmaco, 54 (1999) 26.
3. De Clercq, E., Antiviral Res., 38 (1998) 153.
4. Molla, A. and Japour, A., Curr. Opin. Infect. Dis., 10 (1997) 491.
5. Flexner, C.N., Engl. J. Med., 338 (1998) 1281.
6. Patick, A.K. and Potts, K.E., Clin. Microbiol. Rev., 11 (1998) 614.
7. Kakuda, T.N., Struble, K.A. and Piscitelli, S.C., Am. J. Health-Syst. Pharm., 55 (1998) 233.
8. Rana, K.Z. and Dudley, M.N., Pharmacotherapy, 19 (1999) 35.
9. Miyasaka, T., Tanaka, H., Baba, M., Hayakawa, H., Walker, R.T., Balzarini, J. and De Clercq, E., J. Med. Chem., 32 (1989) 2507.
10. Tanaka, H., Baba, M., Hayakawa, H., Sakamaki, T., Miyasaka, T., Ubasawa, M., Takashima, H., Sekiya, K., Nitta, I., Shigeta, S., Walker, R.T., Balzarini, J. and De Clercq, E., J. Med. Chem., 34 (1991) 349.
11. Tanaka, H., Baba, M., Ubasawa, M., Takashima, H., Sekiya, K., Nitta, I., Shigeta, S., Walker, R.T., De Clercq, E. and Miyasaka, T., J. Med. Chem., 34 (1991) 1394.
12. Tanaka, H., Takashima, H., Ubasawa, M., Sekiya, K., Inouye, N., Baba, M., Shigeta, S., Walker, R.T., De Clercq, E. and Miyasaka, T., J. Med. Chem., 38 (1995) 2860.
13. Tanaka, H., Baba, M., Yamamoto, T., Mori, S., Walker, R.T., De Clercq, E. and Miyasaka, T., Bioorg. Med. Chem. Lett., 3 (1993) 1681.
14. Tanaka, H., Takashima, H., Ubasawa, M., Sekiya, K., Nitta, I., Baba, M., Shigeta, S., Walker, R.T., De Clercq, E. and Miyasaka, T., J. Med. Chem., 35 (1992) 4713.
15. Tanaka, H., Takashima, H., Ubasawa, M., Sekiya, K., Nitta, I., Baba, M., Shigeta, S., Walker, R.T., De Clercq, E. and Miyasaka, T., J. Med. Chem., 35 (1992) 337.
16. 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. and Stuart, D.I., J. Med. Chem., 39 (1996) 1589.
17. Ren, J.S., Esnouf, R., Garman, E., Somers, D., Ross, C., Kirby, I., Keeling, J., Darby, G., Jones, Y., Stuart, D. and Stammers, D., Nature Struct. Biol., 2 (1995) 293.
18. Hansch, C. and Zhang, L.T., Bioorg. Med. Chem. Lett., 2 (1992) 1165.
19. Hannongbua, S., Lawtrakul, L., Sotriffer, C.A. and Rode, B.M., Quant. Struct.-Act. Relat., 15 (1996) 389.
20. Hannongbua, S., Lawtrakul, L. and Limtrakul, J., J. Comput.-Aided Mol. Des., 10 (1996) 145.
21. Garg, R., Kurup, A. and Gupta, S.P., Quant. Struct.-Act. Relat., 16 (1997) 20.
22. Kireev, D.B., Chrétien, J.R., Grierson, D.S. and Monneret, C., J. Med. Chem., 40 (1997) 4257.
23. Luco, J.M. and Ferretti, F.H., J. Chem. Inf. Comp. Sci., 37 (1997) 392.
24. Tronchet, J.M.J., Grigorov, M., Dolatshahi, N., Moriaud, F. and Weber, J., Eur. J. Med. Chem., 32 (1997) 279.
25. Garg, R., Gupta, S.P., Gao, H., Babu, M.S., Debnath, A.K. and Hansch, C., Chem. Rev., 99 (1999) 3525.
26. Lawtrakul, L. and Hannongbua, S., Sci. Pharm., 67 (1999) 43.
27. Jalali-Heravi, M. and Parastar, F., J. Chem. Inf. Comp. Sci., 40 (2000) 147.
28. Alves, C.N., Pinheiro, J.C., Camargo, A.J., Ferreira, M.M.C. and Da Silva, A.B.F., J. Mol. Struct. (Theochem), 530 (2000) 39.
29. Hannongbua, S., Nivesanond, K., Lawtrakul, L., Pungpo, P. and Wolschann, P., J. Chem. Inf. Comp. Sci., 41 (2001) 848.
30. De Oliveira, D.B. and Gaudio, A.C., Quant. Struct.-Act. Relat., 19 (2000) 599.
31. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E., Nucl. Acids Res., 28 (2000) 235.
32. BioMedCAChe 5.0, Fujitsu Ltd., Tokio, 2000.
33. WebLab Viewer Pro 4.0, Molecular Simulations, Inc., San Diego-CA, 2000.
34. Fujita, T., Hansch, C. and Iwasa, J., J. Am. Chem. Soc., 86 (1964) 5175.
35. R. Mannhold, P. Krogsgaard-Larsen and H. Timmerman (Eds.), QSAR: Hansch Analysis and Related Approaches, Vol. 1, VCH, Weinheim, 1993.
36. Taft, R.W., Jr., J. Am. Chem. Soc., 74 (1952) 3120.
37. Tanaka, H., Baba, M., Takahashi, E., Matsumoto, K., Kittaka, A., Walker, R.T., De Clercq, E. and Miyasaka, T., Nucleosides Nucleotides, 13 (1994) 155.