ELSEVIER

QSAR and molecular graphics analysis of N^2-phenylguanines as inhibitors of herpes simplex virus thymidine kinases

Anderson Coser Gaudio*, W. Graham Richards,† and Yuji Takahata‡

**Departamento de Física, Centro de Ciências Exatas, Universidade Federal do Espírito Santo, Vitória, Brazil*
†New Chemistry Laboratory, University of Oxford, Oxford, United Kingdom
‡ Instituto de Química, Universidade Estadual de Campinas, Campinas, Brazil

A quantitative structure-activity relationship study of N^2-(substituted)-phenylguanines (PHG) as inhibitors of herpes simplex virus thymidine kinase (HSV TK) was performed. The activity of a set of PHG derivatives were analyzed against the thymidine kinase of herpes simplex virus types 1 (HSV1 TK) and 2 (HSV2 TK). Classic and calculated physicochemical parameters were included in the analysis. The results showed that there is an important difference in the activity of the meta substituted PHG derivatives against HSV1 TK and HSV2 TK. The activity of the meta derivatives against HSV2 TK is influenced by a steric effect, which is not observed against HSV1 TK. The superposition of the three-dimensional structures of the active sites of HSV1 TK (crystal structure) and HSV2 TK (homology model) revealed that the amino acid Ile97 is located near the meta position in the HSV1 TK active site, whereas the amino acid Leu97 is located near the meta position in the HSV2 TK active site. This single difference in the active sites of both enzymes can explain the source of the steric effect and serves as an indication that our previously proposed binding mode for the PHG derivatives is plausible. However, another observed mutation in the active site region, Ala168 by Ser168, suggests that an alternative binding mode, similar to that of ganciclovir, could be possible. © 2000 by Elsevier Science Inc.

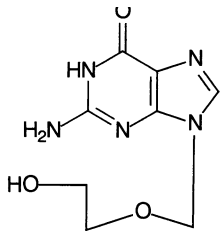
Keywords: herpes simplex virus, thymidine kinase, phenylguanine binding mode, quantitative structure-activity relationships

Corresponding author: A.C. Gaudio, Departamento de Física, Centro de Ciências Exatas, Universidade Federal do Espírito Santo, Campus de Goiabeiras, Vitória, ES, 29060-900, Brazil. Tel.: 0055-27-3352834; fax: 0055-27-3352823.

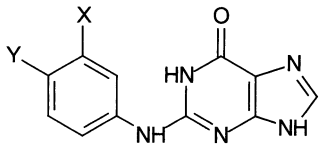
E-mail address: anderson@cce.ufes.br (A.C. Gaudio)

INTRODUCTION

The herpes simplex virus (HSV) is widely distributed in humans and is the cause of a number of important occurrences in medical practice. HSV type 1 (HSV1) usually affects the upper parts of the body, such as the mouth, eyes, and brain. HSV2 generally is associated with genital infections.[1] HSV infections can be efficiently prevented through the combined inhibition of the viral thymidine kinase (TK) and DNA polymerase. At present, acyclovir (**1**) is the drug of choice against HSV infections. Acyclovir is a competitive inhibitor to HSV TK, being monophosphorylated during the interaction with the enzyme. The triphosphorylated form of acyclovir acts as substrate to DNA polymerase. After being added to replicating DNA, acyclovir-3P promotes chain termination due to the lack of an equivalent 3′-OH group.[2]



1



2

The N^2-(m,p-substituted)-phenylguanine (PHG) derivatives (**2**) have been proposed as competitive inhibitors to HSV1 and HSV2 TKs.[3] Qualitative[4] and quantitative[5] structure-activity relationship studies revealed the nature of some interactions that occur between the PHG derivatives and the TK active site. Equations 1 and 2 (recalculated) were

Journal of Molecular Graphics and Modelling 18, 33–41, 2000

655 Avenue of the Americas, New York, NY 10010

1093-3263/00/$–see front matter
PII S1093-3263(00)00032-2

derived by Gambino et al.[5] from a set of 36 meta and para substituted-PHG derivatives.

$$log(1/C_{HSV1\ TK}) = 0.594(\pm 0.229)\pi_m - 0.414(\pm 0.200) \times \pi_p + 1.048(\pm 0.746)\sigma_m + + 2.381(\pm 0.871)\Re_p + 5.077(\pm 0.244)$$
$$(n = 34;\ r = 0.897;\ s = 0.399;\ F = 29.767;\ Q^2 = 0.741;\ s_{PRESS} = 0.459) \quad (1)$$

$$\log(1/C_{HSV2\ TK}) = 0.424(\pm 0.239)\pi_m - 0.496(\pm 0.221) \times \pi_p + 2.055(\pm 0.771)\sigma_m + + 1.468(\pm 0.898)\Re_p + 5.238(\pm 0.253)$$
$$(n = 33;\ r = 0.896;\ s = 0.410;\ F = 28.431;\ Q^2 = 0.736;\ s_{PRESS} = 0.474) \quad (2)$$

In these equations, $C_{HSV1\ TK}$ and $C_{HSV2\ TK}$ are the PHG molar concentrations that provide 50% of HSV1 TK and HSV2 TK inhibition, respectively; π is the Hansch lipophilic substituent constant; σ is the Hammett electronic substituent constant; $\Re$ is the Swain-Lupton resonance parameter; subscripts m and p stand for meta and para positions of the phenyl ring of the PHG derivative molecule, respectively; n is the number of compounds included in the models; r is the correlation coefficient; s is the overall standard deviation; F is the Fischer variance ratio; Q is the cross validation correlation coefficient; and s_{PRESS} is the cross validation standard deviation. The numbers in parentheses are the 95% confidence interval of the respective regression coefficients.

The similarity of Equations 1 and 2 is an indication that the mechanism of interaction of PHG derivatives with HSV1 and HSV2 TKs is similar. This also can be an indication that the binding sites of HSV1 and HSV2 TKs do not differ very much.

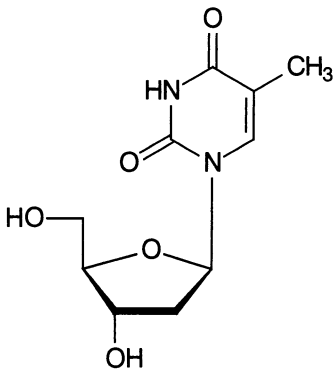
Hadjipavlou-Litina[6] reanalyzed the Gambino data set and presented the bilinear model given by Equation 3, which is a recalculated version of the original model presented in reference 6 (the Hadjipavlou-Litina data set had few wrong parameter values). In the derivation of Equation 3, parameters were taken from reference 7.

$$\log(1/C_{HSV1\ TK}) = 1.28(\pm 0.77)\pi_{3,4} - 2.22(\pm 1.17) \times \log(\beta \cdot 10^{\pi 3,4} + 1) + 1.79(\pm 0.59) \times B_{1-3} - 1.42(\pm 0.82)\Im + 3.20(\pm 0.86)$$
$$(n = 30;\ r = 0.831;\ s = 0.530;\ F = 10.74;\ \pi_o = 0.53;\ \log\beta = -0{,}401) \quad (3)$$

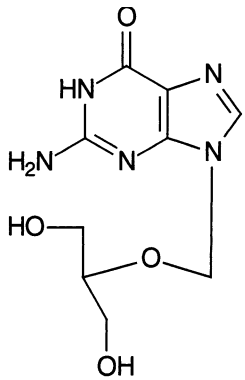
In Equation 3, $\pi_{3,4}$ is the sum of the π constant of the substituents attached to meta and para positions of the phenyl ring of PHG derivatives, β is the bilinear constant, B_{1-3} is the Sterimol parameter B_1, which is a measure of the width of the substituent, on the meta position, $\Im$ is the Swain-Lupton field substituent parameter, and π_o is the optimum $\pi_{3,4}$ value.

The fitting quality of Equation 3 is much worse than in the original published form (r = 0.907; s = 0.401; F = 22.30; π_o = 0.141; log β = 0.198). Therefore, much of the possible biochemical interpretation one could give to the original equation may not be true. A curious aspect of Equation 3 is the dependence of the inhibitory activity on the size of the meta substituent, in which the activity increases as the width of the substituent becomes larger. In general, such a dependence is seldom observed because of the limited size of binding pockets.[8] It is interesting to note that five of the six compounds eliminated from the original Gambino data set had a substituent at the meta position of the phenyl ring.

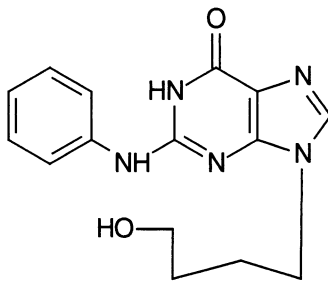
The binding mode of N^2-PHGs in the active site of HSV TK is still a source of some controversy. Gaudio et al.[9] and Sun[10] proposed different models for the binding mode of N^2-PHG (**2**, X = H, Y = H). Our model is based on the binding mode of the natural HSV TK substrate 2′-deoxythimidine (**3**), whereas the Sun model is based on the binding mode of the competitive inhibitor ganciclovir (**4**). The binding modes of 2′-deoxythimidine and ganciclovir became known when the crystal structure of HSV1 TK was first elucidated.[11,12]



3



4



5

Bennett et al.[13] determined the crystal structure of HSV1 TK in complex with the PHG derivative analogue 9-(4-hydroxybutyl)-N^2-phenylguanine (HBPG) (**5**) and concluded that its binding mode agreed well with the Sun model. However, the presence of the 4-hydroxybutyl group on position 9 of the PHG molecule makes the HBPG molecule very similar to

ganciclovir (**4**), especially with respect to the ability of the hydroxy group to form a hydrogen bond with Arg163 or Tyr101 in the HSV1 TK active site.[9,11] This may force the HBPG molecule to adopt a ganciclovir-like binding mode.

The purpose of the present work is to perform a more focused QSAR analysis restricted to the meta PHG derivatives synthesized by Gambino et al.,[5] in order to investigate the nature of the steric effect proposed by Hadjipavlou-Latina[6] and detected by us in a previous calculation.[14] The investigation is performed on HSV1 and HSV2 TKs, and the results are analyzed with respect to the three-dimensional structure of the proposed binding mode of the PHG derivatives in the HSV TK active site.[9]

MATERIALS AND METHODS

QSAR Analysis: The following classic substituent parameters were tested in the QSAR analysis: π_m and its square π_m^2, σ_m, MR_m, the molar refractivity, $\Im_m$, and L, B_1, and B_5, the Verloop sterimol parameters. These parameters were taken from reference 7. Parameters for the substituent CH_2N_3 were estimated. See reference 8 for a detailed discussion about substituent parameters used in the QSAR analysis.

The calculated parameters tested in the QSAR study were the following: the energy of the highest occupied molecular orbital (ϵ_{HOMO}); the energy of the lowest unoccupied molecular orbital (ϵ_{LUMO}); the molecular hardness (η); the molecular dipole moment (μ); the molecular polarizability (α); the van der Waals molecular volume (V_{vdW}); the liquid atomic charge of the n^{th} atom of the PHG molecule (Q_n); the atomic frontier electron density ($F_n^{(e)}$), the frontier orbital density ($F_n^{(o)}$), and the frontier radical density ($F_n^{(r)}$). For a discussion about the applications of calculated parameters in molecular pharmacology, the reader is referred to references [15] and [16] The parameters relevant to the present QSAR study are shown in Table 1.

The quantum mechanical properties were calculated with the AM1 method,[17] implemented in the MOPAC 6.0 package.[18] The atomic charges were derived from the fitting of the molecular electrostatic potential generated by the AM1 wave function. The van der Waals molecular volume was calculated using a previously described algorithm.[19] The molecular geometry adopted in the calculation procedure was that obtained in the determination of the binding mode of the PHG derivatives.[9] The values adopted for the torsion angles θ_1 (N_1-C_2-N_{11}-$C_{1'}$) and θ_2 (C_2-N_{11}-$C_{1'}$-$C_{2'}$) (Figure 1) were approximately 210° and 270°, respectively.

The regression equations presented in this article were built using an unpublished computer program, called MRA, written by one of the authors (A.C.G.). The physicochemical parameters that appear in the regression equations were selected by systematic search, in which all possible parameter combinations, considering models with no more than five parameters, were analyzed.

HSV1 and HSV2 TKs amino acid alignment: The amino acid sequences of HSV1 TK (376 residues) and HSV2 TK (375 residues) were downloaded from the Internet site of the Swiss-Prot Protein Sequence Databank*. These sequences were aligned with the Multal Data Matrix algorithm,[20] implemented in the program Cameleon.[21] The total identity percentage was computed manually by counting the exact amino acid matches along the two sequences and comparing the number of matches with the total number of amino acids in HSV1 TK (i.e., 276 matches in 376 amino acids [73%]).

Force field and molecular dynamics: Molecular mechanics and dynamics calculations were performed using CHARMm 22.3 force field.[22] The atomic charges was gener-

Table 1. Matrix of physicochemical parameters and the inhibitory activities of PHG derivatives against HSV1 TK and HSV2 TK

		Log 1/C[a]										
No	X	HSV1	HSV2	π_m[b]	σ_m[b]	MR_m[b]	B_{1m}[b]	V_{vdW}[c]	$F_1^{(e)}$[d]	$F_2^{(e)}$[d]	$F_3^{(e)}$[d]	$F_1^{(e)}$[d]
1	CF_3	6.82	7.00	0.88	0.43	0.50	1.99	210.5	0.484	1.489	2.905	−32.33
2	CH_2N_3	6.10	5.82	[e]1.12	[e]0.00	[e]1.47	[e]1.50	221.3	0.431	1.446	3.039	−23.12
3	CH_2Br	6.00	5.92	0.79	0.12	1.34	1.52	223.1	0.420	1.429	3.015	−24.87
4	I	5.96	5.96	1.12	0.35	1.39	2.15	214.9	0.424	1.432	3.024	−27.41
5	CH_2Cl	5.85	5.87	0.17	0.11	1.05	1.52	216.8	0.415	1.429	3.039	−25.22
6	Cl	5.82	6.46	0.71	0.37	0.60	1.80	200.0	0.419	1.424	3.019	−27.36
7	Br	5.82	6.22	0.86	0.39	0.89	1.95	206.6	0.423	1.427	3.009	−27.48
8	Et	5.60	5.40	1.02	−0.07	1.03	1.52	217.4	0.355	1.352	3.147	−21.88
9	n-Pr	5.48	5.52	1.55	−0.06	1.50	1.52	232.9	0.357	1.355	3.141	−21.81
10	CN	5.30	6.40	−0.57	0.56	0.63	1.60	200.1	0.468	1.471	2.930	−29.20
11	H	5.10	5.80	0.00	0.00	0.10	1.00	185.0	0.369	1.372	3.138	−22.68
12	CH_2OH	4.60	4.70	−1.03	0.00	0.72	1.52	225.5	0.384	1.383	3.087	−21.84
13	CH_2NH_2	4.52	4.70	−1.04	−0.03	0.91	1.52	212.5	0.355	1.353	3.140	−20.74

[a] C is the molar concentration of a drug that produces 50% of inhibition in the enzyme activity.
[b] Taken from reference 7.
[c] In $Å^3$, calculated by a self-made program.[19]
[d] Calculated through the AM1 wave function. Values multiplied by 100.
[e] Estimated.

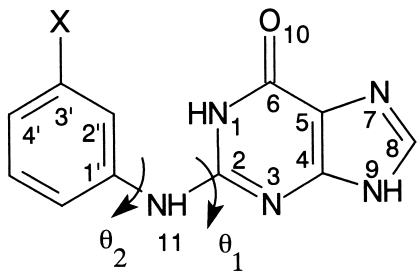


Figure 1. Structure and atom numbering of the meta substituted PHG molecule, of which the substituents X are shown in Table 1. The torsion angles θ_1 and θ_2, referred to in the text, are the dihedrals N_1-C_2-N_{11}-$C_{1'}$ and C_2-N_{11}-$C_{1'}$-$C_{2'}$, respectively.

ated by the Gasteiger-Huckel method, which produced very stable equilibration and simulation steps. A nonbonded cutoff of 10 Å was used. The hydrogen bonds were treated as electrostatic nonbonded interactions with proper directional properties.[22] A dielectric constant of unity was used. Molecular dynamics simulation was performed under NVE conditions. The Verlet algorithm for integrating the equations of motion and a SHAKE constraint with a 1 fs time step also were used.

HSV1 TK structure preparation: The prepared HSV1 TK structure used in the present work is the same one used in our previous article,[9] to which the reader is referred for details about the process.

HSV2 TK structure building: The three-dimensional structure of HSV2 TK has not been determined yet. The Protein Design module of program Quanta-CHARMm[22,23] was used to build a three-dimensional model of the HSV2 TK from the HSV1 TK crystal structure,[11] used as a template. Starting from the original crystal structure of HSV1 TK, the water molecules and the sulfate ions were eliminated from the system. The next step was to convert the amino acid sequence of the HSV1 TK to the amino acid sequence of HSV2 TK. This was done by carrying out individual amino acid mutations, in which every amino acid of the HSV1 TK crystal structure that was not in the HSV2 TK sequence was substituted by an appropriate one. After each mutation, a local relaxation was performed, comprising 100 steps of steepest descent energy minimization. The spatial arrangement of the side chain of each new substituted amino acid also was observed visually.

As in the case of HSV1 TK, the missing carboxy and amino terminal sequences of the available crystal structure have been left out. Some loops on the surface of the original HSV1 TK structure were absent. These loops, defined by the amino acid sequences 73-76, 150-151, and 263-277, had to be completed using the corresponding amino acid sequences of HSV2 TK. A loop search was performed through the Protein Design module of program Quanta. Loop Gly73–Asp76 was selected from the enzyme 1LYS, starting at Ala10, loop Pro150–Gln151 was selected from 3CYT, starting on Thr64, and the long loop Thr263–Gly277 was selected from 5CHA, starting at Gly13. After being connected to the free ends, the chosen loops and some residues from the anchors were submitted to 500 steps of steepest descent minimization in order to reduce the tension in the enzyme structure. The missing hydrogen atoms also were added and the charged residues were adjusted to the isoelectric point. The structure of the N^2-PHG were built using ChemNote module of Quanta and manually adjusted in the HSV1 TK active site to match the proposed binding mode.[9] CHARMm 22.3 force field parameters were assigned to the inhibitor molecules using Quanta. The HSV2 TK-PHG structure near the active site was relaxed with molecular dynamics. The calculation included only the inhibitor molecule and those complete amino acid residues in which at least one of their atoms was ≤8 Å distant from the nearest inhibitor atom. The carbon and nitrogen atoms (atom types C, N, and CA) from the enzyme backbone were submitted to a harmonic constraint potential in which the force constant was equal to 100 $kcal.mol^{-1}.Å^{-2}$. As the HSV1 TK crystallographic structure is a dimer, two active sites were included in the simulation.

The simulation was performed as follows: (a) Relaxation: 1000 steps of steepest descent minimization in which the inhibitor molecule was kept frozen, and 1000 steps of conjugate gradient minimization on the whole system; (b) Heating: from 0 to 310 K over 4 ps, with temperature increase every 0.05 ps; (c) Equilibration: at 310 K over 4 ps with atomic velocity assignment every 0.05 ps; (d) Simulation: at 310 K over 8 ps; (e) Quenching: from 310 K to 0 K over 4 ps with atomic velocities assignment every 0.05 ps; (f) Final relaxation: 1000 steps of steepest descent minimization on the whole system, and 1000 steps of conjugate gradient minimization on the whole system.

QSAR ANALYSIS

In order to perform the present QSAR analysis, the activity of the meta PHG derivatives of the Gambino data set were correlated with classic and calculated parameters. Equation 4 is the best two-variable model found for the inhibitory activity of the meta PHG derivatives against HSV1 TK. No significant three-variable model could be found.

$$\log(1/C_{HSV1\ TK}) = 0.486(\pm 0.160)\pi_m + 8.201 \times (\pm 3.000)F_2^{(e)} - 6.180(\pm 4.230)$$
$$(n = 13;\ r = 0.953;\ s = 0.209;\ F_{(2,10)} = 49.057;\ Q^2 = 0.805;\ s_{PRESS} = 0.303) \quad (4)$$

According to Equation 4, the activity of the meta PHG derivatives against HSV1 TK is a function of the lipophilic character of the substituent and an electronic property, namely the frontier electron density on atom C2. The replacement of $F_2^{(e)}$ by $F_1^{(e)}$ or $F_3^{(e)}$ in Equation 4 also resulted in equally highly significant models, which is a consequence of the high correlation between $F_1^{(e)}$, $F_2^{(e)}$, and $F_3^{(e)}$ ($r > 0.9$) (Table 2). The best combination of classic substituent constants that generate a model equivalent to Equation 4 is π_m and σ_m. However, the fitness quality of the model $\log 1/IC_{HSV1\ TK} = f(\pi_m, \sigma_m)$ is not as good as in Equatioon 4 ($r = 0.857$; $s = 0.354$; $F = 13.815$). The steric effect on the meta position of the PHG derivatives against HSV1 TK, proposed by Hadjipavlou-Litina,[6] was not observed here.

The atomic frontier electron density $F_n^{(e)}$ is a property that denotes the electron density in the HOMO of a given atom in a molecule. Its value can be thought of as the atom's ability to act as electron donor when interacting with an approaching electrophilic species. Conversely, if that atom is not in direct

Table 2. Correlation matrix for the parameters used in the QSAR analyses, in terms of r^2

	π_m	σ_m	RM_m	B_{1m}	V_{vdW}	$F_1^{(e)}$	$F_2^{(e)}$	$F_3^{(e)}$	$F_{1'}^{(e)}$
π_m	1	0.002	0.259	0.122	0.071	0.016	0.024	0.008	0.018
σ_m		1	0.079	0.443	0.184	0.698	0.638	0.770	0.670
RM_m			1	0.063	0.652	0.015	0.006	0.019	0.018
B_{1m}				1	0.027	0.301	0.271	0.355	0.326
V_{vdW}					1	0.040	0.034	0.034	0.059
$F_1^{(e)}$						1	0.989	0.974	0.943
$F_2^{(e)}$							1	0.943	0.936
$F_3^{(e)}$								1	0.880
$F_{1'}^{(e)}$									1

intermolecular interaction with any other species, the $F_n^{(e)}$ value simply tells us about its HOMO electron density. The $F_n^{(e)}$ value is always positive. In Equation 4, the positive coefficient of $F_2^{(e)}$ indicates that the inhibitory activity against HSV1 TK increases when the HOMO electron density on atom C2 gets higher. The same thing is observed when $F_2^{(e)}$ is replaced by $F_1^{(e)}$. Conversely, if $F_2^{(e)}$ is replaced by $F_3^{(e)}$, the coefficient of $F_3^{(e)}$ is negative. This is possibly related to the ability of atom N3 to act as hydrogen bond acceptor.

Considering the variables tested, the best three-variable model one can build for the activity of the meta PHG derivatives against HSV2 TK is Equation 5, which is composed of a lipophilic, an electronic, and a steric parameter.

$$\log(1/C_{HSV2\,TK}) = 0.292(\pm 0.096)\pi_m - 0.205 \times (\pm 0.030)F_{1'}{}^{(n)} - 0.922(\pm 0.370)B_{1m} + 2.067(\pm 0.570)$$
$$(n = 13;\ 4 = 0.988;\ s = 0.117;\ F_{(3,9)} = 120.946;\ Q^2 = 0.952;\ s_{PRESS} = 0.166) \quad (5)$$

```
                1        10        20        30        40        50        60
                ....:....+....:....+....:....+....:....+....:....+....:....+....:
HSV1 TK         MASYPCHQHASAFDQAARSRGHSNRRTALRPRRQQEATEVRLEQKMPTLLRVYIDGPHGMGKTTT
HSV2 TK         MASHAGQQHAPAFGQAARASGPTDGRAASRPSHRQGASEARGDPELPTLLRVYIDGPHGVGKTTT

Consensus:      MAS    QHA AF QAAR  G %  R A RP $ Q A%E!R!~   PTLLRVYIDGPHG GKTTT
Identity:       ***    *** ** **** *    * * **   * * * *     ************* *****

                66        75        85        95        105       115       125
                ....+ ....:....+....:....+....:....+....:....+....:....+....:....
HSV1 TK         TQLLV-ALGSRDDIVYVPEPMTYWRVLGASETIANIYTTQHRLDQGEISAGDAAVVMTSAQITMG
HSV2 TK         SAQLMEALGPRDNIVYVPEPMTYWQVLGASETLTNIYNTQHRLDRGEISAGEAAVVMTSAQITMS

Consensus:      %  L  ALG RD IVYVPEPMTYW VLGASET! NIY TQHRLD GEISAG~AAVVMTSAQITM
Identity:          *  *** ** ********** *******  *** ****** ****** ************

                130       140       150       160       170       180       190
                +....:....+....:....+....:....+....:....+....:....+....:....+....
HSV1 TK         MPYAVTDAVLAPHIGGEAGSSHAPPPALTLIFDRHPIAALLCYPAARYLMGSMTPQAVLAFVALI
HSV2 TK         TPYAATDAVLAPHIGGEAVGPQAPPPALTLVFDRHPIASLLCYPAARYLMGSMTPQAVLAFVALM

Consensus:       PYA!TDAVLAPHIGGEA!   APPPALTL!FDRHPIA LLCYPAARYLMGSMTPQAVLAFVAL
Identity:        *** *************    ******** ******* *************************

                195       205       215       225       235       245       255
                :....+....:....+....:....+....:....+....:....+....:....+....:....
HSV1 TK         PPTLPGTNIVLGALPEDRHIDRLAKRQRPGERLDLAMLAAIRRVYGLLANTVRYLQCGGSWREDW
HSV2 TK         PPTAPGTNLVLGVLPEAEHADRLARRQRPGERLDLAMLSAIRRVYDLLANTVRYLQRGGRWREDW

Consensus:      PPT!PGTN!VLG!LPE  H!DRLA$RQRPGERLDLAML AIRRVY LLANTVRYLQ GG WREDW
Identity:       *** **** *** ***  * **** ************* ****** ********** ** *****
```

Figure 2. HSV1 and HSV2 TKs amino acid sequence alignment made by program Cameleon,[21] using the algorithm Multal Data Matrix.[20] The highlighted amino acid residues form the active site, but are not necessarily in close interaction with the ligand. The differences in the two sequences are the amino acids Ile97, Ala98, and Ala168 in HSV1 TK, which are substituted by Leu97, Thr98, and Ser168 in HSV2 TK. In the Identity *line, an asterisk indicates that the same amino acid in present in equivalent positions of the two sequences.*

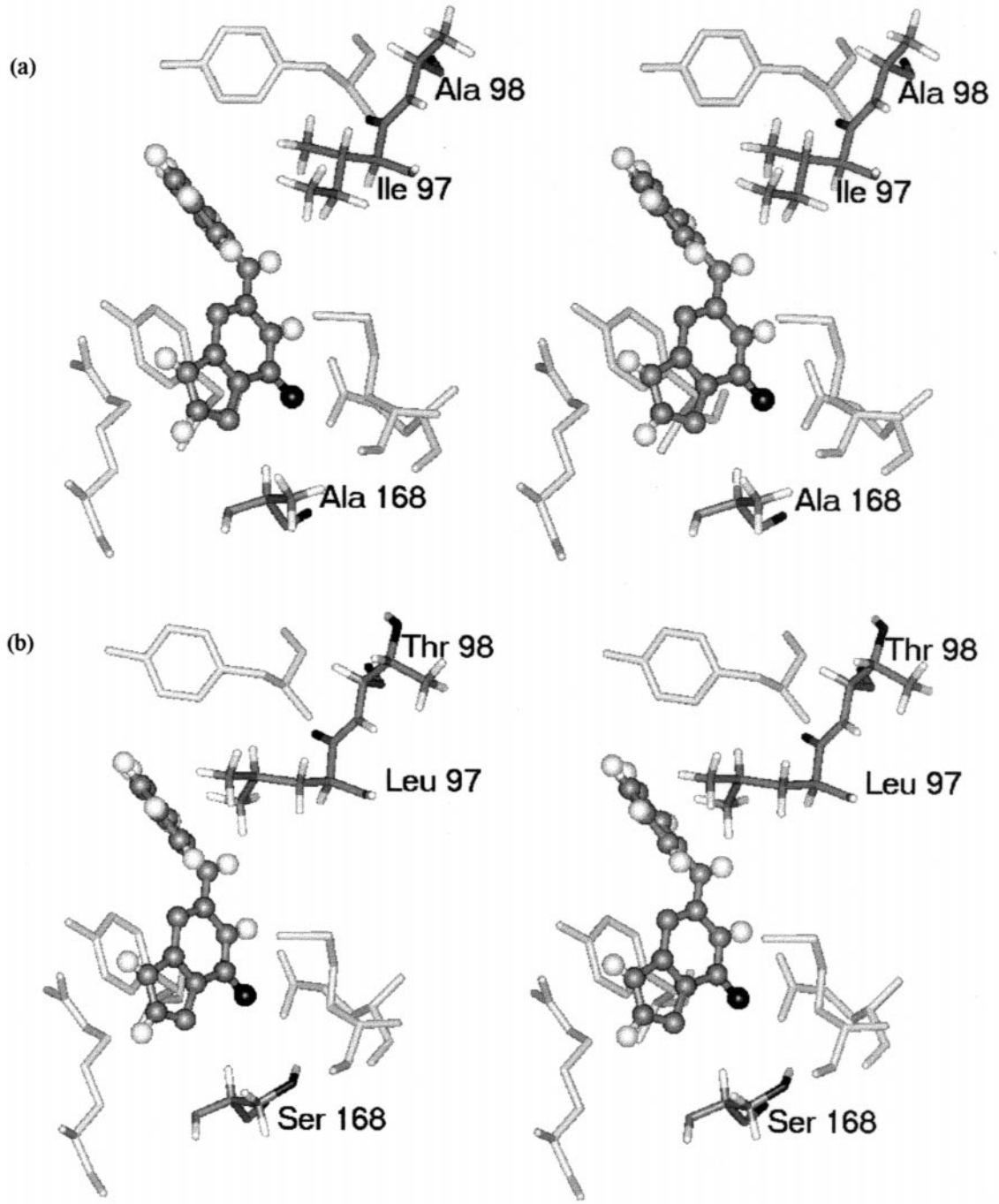


Figure 3. Stereoview of the unsubstituted PHG derivative in the active sites of (a) HSV1 and (b) HSV2 TKs. Although there are three different amino acid residues in these active sites (Ile97, Ala98, and Ala168 in HSV1 TK; Leu97, Thr98, and Ser168 in HSV2 TK), only position 97 appears to be responsible for the source of the steric effect observed in HSV2 TK. The majority of the hydrogen atoms from the amino acid residues were hidden for clarity.

A good model also results for the activity against HSV2 TK if a steric parameter is added to a model similar to Equation 4: log $1/IC_{HSV2\ TK} = f(\pi_m, F_3^{(e)}, V_{vdW})$, (n = 13, r = 0.980; s = 0.150; F = 72.213; $Q^2 = 0.923$; $s_{PRESS} = 0.209$). Considering just classic substituent constants, the only meaningful model that is still equivalent to Equation 5 is: log $1/IC_{HSV2\ TK} = f(\pi_m, \sigma_m, B_{1m})$, (n = 13, r = 0.948; s = 0.239; F = 26.750; $Q^2 = 0.765$; $s_{PRESS} = 0.365$).

In Equation 5, the steric parameter has negative coefficient, which is consistent with the classic steric effect. As previously noted, the inhibitory activity of the meta substituted PHG derivatives against HSV1 TK does *not* depend on any *steric* parameter. It only depends on lipophilic and electronic parameters (Equation 4). It remains to be explained why steric parameter appears in Equation 5, while no steric parameter appears in Equation 4.

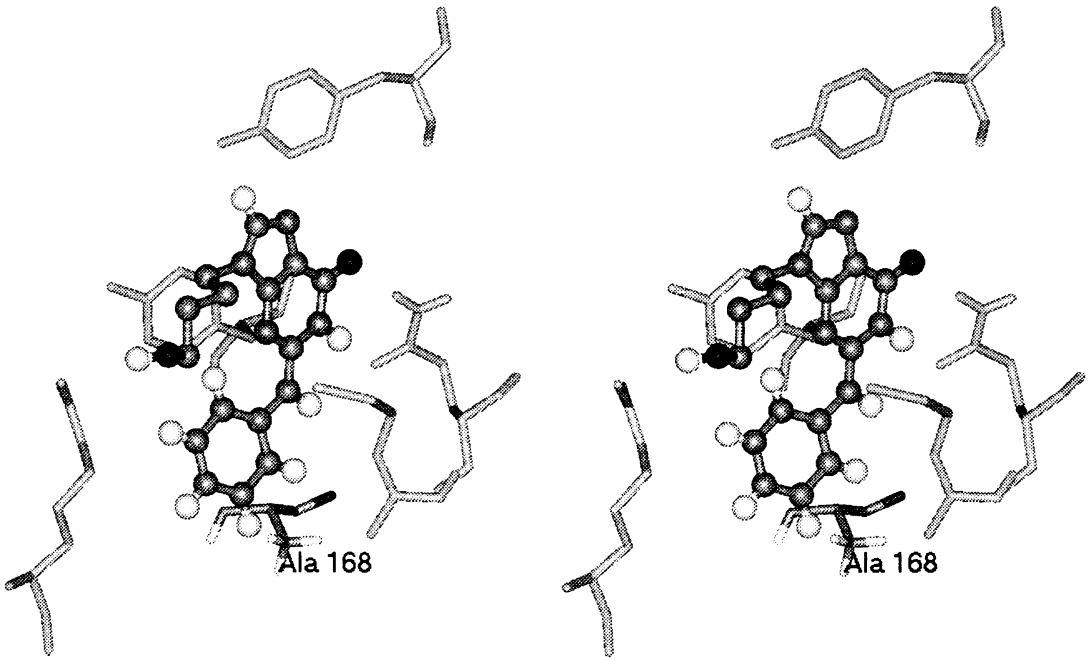


*Figure 4. Stereoview of the crystal data of PHG derivative 9-(4-hydroxybutyl)-N^2-phenylguanine (HBPG) (**5**) in the active site of HSV1 TK.[13] The binding mode of HBPG is similar to that of ganciclovir.[11] In this binding mode, the methyl group of Ala168 is located very close to the phenyl group of the inhibitor, being approximately 2 Å the distance between the closest hydrogen atoms of these groups (the hydrogen from phenyl is that one bound to position meta). Some hydrogen atoms from the inhibitor molecule and the majority of the hydrogen atoms from the amino acid residues were hidden for clarity.*

A possible hypothesis to explain the presence of a steric parameter in Equation 5 and its absence in Equation 4 is that the active receptor sites of HSV1 TK and HSV2 TK are somewhat similar, but not identical. The amino acid composition near the region where the meta substituent of the phenyl ring of PHG binds is most likely to be different.

The results of the QSAR analysis obtained so far can be used to check if the recently proposed binding mode of the PHG derivatives[9] in the TK active site is consistent. In order to do this, we need to examine the active site region of both enzymes and look for differences in the amino acid arrangement that could explain the QSAR results.

MOLECULAR GRAPHICS ANALYSIS

The crystal structure of the HSV1 TK was determined simultaneously in two different laboratories in 1995.[11,12] According to these crystallographic studies, HSV1 TK has a dimeric alpha-beta structure, with each monomer being 376 amino acids long. The HSV2 TK crystal structure has not yet been determined, but the amino acid sequences of both enzymes are well known.[24–26] Figure 2 shows part of the HSV1 and HSV2 TK amino acid sequence alignment. These sequences share approximately 73% (276/376) of identity, considering exact matches. Such a high degree of homology in the primary structures suggest that the secondary and tertiary structures of HSV1 and HSV2 TKs also are very similar. One could expect that the alpha-beta arrangement observed in the crystal structure of HSV1 TK also will be observed in the HSV2 TK structure. Based on this hypothesis, the crystal structure of HSV1 TK was used as a template in order to build a model of the HSV2 TK. Figure 2 also shows, in bold type face, the amino acid residues that form the active site, together with some other close residues. It is possible to see that there are only three amino acid mutations between the two enzymes within the highlighted region: these mutations are Ile97, Ala98, and Ala168 of HSV1 TK being replaced by Leu97, Thr98, and Ser168 in HSV2 TK, respectively. However, it is impossible to be sure whether one particular amino acid interacts directly with the ligand in the active site by just looking at the amino acid sequences. In order to know the possible interaction between a particular amino acid and the ligand, we have to take a look at the three-dimensional arrangement of the amino acid residues located near the active site.

Figure 3 shows the N^2-PHG molecule bound to the active site of HSV1 and HSV2 TKs, according to our previously proposed binding mode.[9] Although there are three amino acid mutations in the active site region of the two enzymes, only one of the three interacts directly with the ligand. This is the residue Ile97 of HSV1 TK, which is substituted by the residue Leu97 in the corresponding spatial position of HSV2 TK structure. The residue Ile97 is located near one of the meta positions of the phenyl ring of the PHG molecule in the receptor-ligand complex. The presence of Leu97 in place of Ile97 in HSV2 TK can explain the source of the steric effect observed in the QSAR analysis (Equation 5). Because the side chain of the isoleucine residue has the tertiary carbon atom close to the protein backbone, just one methyl group is left in close contact with the PHG phenyl ring. On the other hand, the intrinsic configuration of the leucine residue locates the tertiary carbon atom at the end of the side chain. In this case, there are two

methyl groups in close contact with the phenyl ring of the PHG molecule, just near the meta position. The presence of the two methyl groups around the meta position of the PHG phenyl ring probably causes the binding problems when the PHG derivative has large substituents attached to the meta position. There would not be enough space to accommodate the large meta substituent in that region. In this case, the whole ligand molecule could be displaced from the ideal binding mode, decreasing inhibitory potency.

Amino acid 98 of both HSV1 and HSV2 TK, which also undergoes mutation, has its side chain projected to the opposite side where the active site is located (Figure 3). This means that whatever binding mode PHG derivatives may come to adopt, residue 98 cannot be directly responsible for any steric effect on the PHG molecule.

The mutation observed in position 168 (Ala replaced by Ser) is an interesting one. The side chain of Ala is a methyl group whereas that of Ser is a hydroxymethylene. As they are pointing toward the inside of the active site, it is necessary to make further considerations. The size of the Ala and Ser side chains differ only by an oxygen atom, so that the mutation in position 168 may not be capable of explaining possible steric hindrances that could be observed in the active site. However, the highly polar hydroxyl group of Ser could represent an important source of dipole interaction with the meta substituent of PHG molecule if the meta substituent was close enough to it. In the binding mode presented in Figure 3, which mimics the 2-deoxythymidine binding mode,[9] there is no possibility of observing such interaction. Therefore, it is necessary to consider alternative binding modes, specially those that could place the phenyl group at the neighborhood of position 168.

Our previous study,[9] which included energy criteria in the selection of PHG binding modes, indicated that the only other possible binding mode for the PHG derivatives is the ganciclovir binding mode.[11] In fact, there is an alternative proposition for the binding mode of the PHG derivatives that is similar to the ganciclovir binding mode.[10] In addition, the PHG derivative HBPG (**5**) was shown to bind to the HSV1 TK active site in the same way as ganciclovir.[13] Figure 4 shows the crystal data of the ganciclovir-like binding mode of the HBPG molecule. It is possible that the phenyl group of HBPG is really very close to the methyl side chain of Ala168, being approximately 2 Å the distance between the closest hydrogen atoms of these groups. It is difficult to be sure if the replacement of Ala by Ser can explain the steric effect predicted in the present QSAR analysis, at least in terms of the substituent size. From the polar point of view, however, a possible explanation of the steric effect remains to be fully explored. Although we have already presented some qualitative and quantitative arguments to support the binding mode shown in Figure 3,[9] we should not discard the ganciclovir-like binding mode as a possible binding mode to the PHG derivatives.

CONCLUSIONS

The present QSAR study indicates that the activity of meta substituted PHG derivatives as competitive inhibitors of HSV2 TK is influenced by a steric effect, which is absent when those derivatives act against HSV1 TK. Our previously proposed binding mode for the PHG derivatives is consistent with this observation. The molecular graphics analysis shows that there is an important difference in the active sites of the two enzymes around the meta region of the PHG phenyl ring, which could explain the source of the steric effect. The amino acid Ile97 of HSV1 TK is replaced by Leu97 in HSV2 TK. This result is an indication that the adopted PHG binding mode seems to be reasonable. However, another mutation observed in the HSV TK active site must be considered as a possible source for the differences found in the QSAR study for both enzymes. This is the amino acid Ala168 of HSV1 TK which is substituted by Ser168 in HSV2 TK. If a binding mode similar to that of ganciclovir is considered for the PHG derivatives, the phenyl group is set very close to the residue 168. Therefore, a ganciclovir-like binding mode for the PHG derivatives should not be discarded.

It is important to consider the possibility that different PHG derivatives could bind to the HSV TK active site adopting different binding modes. Such occurrence in principle could be detected by the construction of a significant bilinear model just like the one presented in Equation 3. However, many attempts to build different kinds of bilinear models including more than 20 compounds, using π_m, π_p, and $\pi_{m,p}$ as the bilinear variable, have failed. Unfortunately, it does not seem possible to test bilinear variables in models such as Equations 4 and 5 due to the low number of degrees of freedom available.

ACKNOWLEDGMENTS

The crystallographic structures used in this work were provided by Dr. Mark R. Sanderson and Dr. John Champness, to whom the authors are very grateful. This work was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp), Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (Capes), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Programa Institucional de Capacitação Docente de Universidade Federal do Espírito Santo (PICDT-UFES).

REFERENCES

1 Collier, L., and Oxford, J. *Human Virology: A Text for Students of Medicine, Dentistry, and Microbiology*. Oxford University Press, Oxford, 1993

2 Richman, D.D. *Antiviral Drug Resistance*. John Wiley & Sons, Guilford, 1996

3 Focher, F., Hildebrand, C., Freese, S., Ciarrocchi, G., Noonan, T., Sangalli, S., Brown, N., Spadari, S., and Wright, G. N^2-phenyldeoxyguanosine: A novel selective inhibitor of herpes simplex thymidine kinase. *J. Med. Chem.* 1988, **31**, 1496–1500

4 Hildebrand, C., Sandoli, D., Focher, F., Gambino, J., Ciarrocchi, G., Spadari, S., and Wright, G. Structure-activity relationships of N^2-Substituted guanines as inhibitors of HSV1 and HSV2 thymidine kinases. *J. Med. Chem.* 1990, **33**, 203–206

5 Gambino, J., Focher, F., Hildebrand, C., Maga, G., Noonan, T., Spadari, S., and Wright, G. Quantitative structure-activity relationships of N^2-phenylguanines as inhibtors of herpes simplex virus thymidine kinases. *J. Med. Chem.* 1992, **35**, 2979–2983

6 Hadjipavlou-Litina, D. QSAR of N^2-phenylguanidines as inhibitors of herpes simplex virus-1 thymidine kinase. *Pharmazie* 1995, **50**, 796–798

7 Hansch, C., Leo, A., and Hoekman, D. *Exploring QSAR:*

Hydrophobic, Electronic, and Steric Constants. American Chemical Society, Washington, DC, 1995
8 Kubinyi, H. QSAR: Hansch analysis and related approaches. In: *Methods and Principles in Medicinal Chemistry*, Volume 1 (Mannhold, R., Krogsgaard-Larsen, P., and Timmerman, H., Eds.). VCH, Weinheim, 1993, pp. 21–55
9 Gaudio, A.C., Takahata, Y., and Richards, W.G. Prediction of the binding mode of N^2-phenylguanines derivative inhibitors to herpes simplex virus type 1 thymidine kinase. *J. Comput. Aided Mol. Des.* 1998, **12**, 15–25
10 Sun, H. Ph.D thesis. Clark University, Worcester, MA, 1997
11 Brown, D.G., Visse, R., Sandhu, G., Davies, A., Rizkallah, P.J., Melitz, C., Summers, W.C., and Sanderson, M.R. Crystal structure of the thymidine kinase from herpes simplex virus type-I in complex with deoxythymidine and ganciclovir. *Nat. Struct. Biol.* 1995, **2**, 876–881
12 Wild, K., Bohner, T., Aubry, A., Folkers, G., and Schulz, G.E. The three-dimensional structure of thymidine kinase from herpes simplex virus type 1. *FEBS Lett.* 1995, **368**, 289–292
13 Bennett, M.S., Wien, F., Champness, J.N., Batuwangala, T., Rutherford, T., Summers, W.C., Sun, H., Wright, G., and Sanderson, M.R. Structure to 1.9 Å resolution of a complex with herpes simplex virus type-1 thymidine kinase of a novel, non substrate inhibitor: X-ray crystallographic comparison with binding of acyclovir. *FEBS Lett.* 1999, **443**, 121–125
14 Gaudio, A.C. PhD thesis. Universidade Estadual de Campinas, Campinas, 1998
15 Richards, W.G. *Quantum Pharmacology*, 2nd Edition. Butterworths, London, 1983
16 Karelson, M., Lobanov, V.S., and Katritzky, A.R. Quantum-chemical descriptors in QSAR/QSPR studies. *Chem. Rev.* 1996, **96**, 1027–1043
17 Dewar, M.J.S., Zoebisch, E.G., Healy, E.F., and Stewart, J.J.P. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* 1985, **107**, 3902–3909
18 Stewart, J.J.P. MOPAC: A semiempirical molecular orbital program. *J. Comput. Aided Mol. Des.* 1990, **4**, 1–105
19 Gaudio, A.C., and Takahata, Y. Calculation of molecular-surface area with numerical factors. *Comp. Chem.* 1992, **16**, 277–284
20 Dayhoff, M.O. (Ed.). *Atlas of Protein Sequences and Structure*, Volume 5, Supplement 3. NBRF, Washington, DC, 1979
21 *Cameleon*, Release 3.1. Oxford Molecular Ltd., Oxford, 1994
22 *CHARMm*, Release 22.3. Molecular Simulations, San Diego, 1994
23 *Quanta*, Release 4.0. Molecular Simulations, San Diego, 1994
24 Swain, M.A., and Galloway, D.A. Nucleotide sequence of the herpes simplex virus type 2 thymidine kinase gene. *J. Virol.* 1983, **46**, 1045–1050
25 Kit, S., Kit, M., Qavi, H., Trkula, D., and Otsuka, H. Nucleotide sequence of the herpes simplex virus type 2 (HSV-2) thymidine kinase gene and predicted amino acid sequence of thymidine kinase polypeptide and its comparison with the HSV-1 thymidine kinase gene. *Biochim. Biophys. Acta* 1983, **741**, 158–170
26 McGeoch, D.J., Dalrymple, M.A., Davison, A.J., Dolan, A., Frame, M.C., McNab, D., Perry, L.J., Scott, J.E., and Taylor, P. The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. *J. Gen. Virol.* 1988, **69**, 1531–1574

*The Swiss-Prot Protein Sequence Databank can be accessed at *http://www.expasy.ch/sprot/sprot-top.html*.