

Structural and chemical basis for enhanced affinity to a series of mycobacterial thymidine monophosphate kinase inhibitors: fragment-based QSAR and QM/MM docking studies

Renata V. Bueno · Ney R. Toledo · Bruno J. Neves · Rodolpho C. Braga · Carolina H. Andrade

Received: 27 March 2012 / Accepted: 3 July 2012 / Published online: 31 July 2012
© Springer-Verlag 2012

Abstract Tuberculosis (TB) still remains one of the most deadly infectious diseases. *Mycobacterium tuberculosis* thymidine monophosphate kinase (TMPKmt) has emerged as an attractive molecular target for the design of a novel class of anti-TB agents since blocking it will affect the pathways involved in DNA replication. Aiming at shedding some light on structural and chemical features that are important for the affinity of thymidine derivatives to TMPKmt, we have employed a special fragment-based method to develop robust quantitative structure-activity relationship models for a large and chemically diverse series of thymidine-based analogues. Significant statistical parameters ($r^2=0.94$, $q^2=0.76$, $r^2_{pred}=0.89$) were obtained, indicating the reliability of the hologram QSAR model in predicting the biological activity of untested compounds. The 2D model was then used to predict the potency of an external test set, and the predicted values obtained from the HQSAR model were in good agreement with the experimental results. We have accordingly designed novel TMPKmt inhibitors by utilizing the fragments proposed by HQSAR analysis and predicted with good activity in the developed models. The new designed compounds also presented *drug-like* characteristics based on Lipinski's rule of 5. The generated molecular recognition patterns gathered from the HQSAR analysis combined with quantum mechanics/molecular mechanics (QM/MM) docking studies, provided

important insights into the chemical and structural basis involved in the molecular recognition process of this series of thymidine analogues and should be useful for the design of new potent anti-TB agents.

Keywords Fragment-based · HQSAR · QM/MM docking · Thymidine monophosphate kinase · Tuberculosis

Introduction

Tuberculosis (TB) is an infectious disease spread worldwide. Two billion people (one-third of the world population) are infected with tuberculosis. *Mycobacterium tuberculosis* is responsible for 8.8 million new infections and 1.6 million deaths each year [1]. The rise of multi- and extensively drug-resistant strains contributes to the spread and worsens the situation; threatening both developing and industrialized countries. Novel drugs with activity against drug-resistant strains are therefore urgently needed to restrain the disease that was once thought to be under control [2–5].

M. tuberculosis thymidine monophosphate kinase (EC 2.7.4.9, TMPKmt), also called thymidylate kinase, recently emerged as an attractive target for the design of a novel class of anti-TB agents. TMPKmt catalyses the phosphorylation of deoxythymidine monophosphate (dTMP) to deoxythymidine diphosphate (dTDP) using ATP as a phosphoryl donor. This step lies at the junction of the *de novo* and salvage pathways of thymidine triphosphate (TTP) biosynthesis, which is essential for DNA replication [6]. Thus, TMPKmt is crucial for cell proliferation as well as survival of the organism. The elucidation of the TMPK X-ray structures of both human [7] and mycobacteria [8], and their low (22 %) sequence homology, enhances the consideration of TMPKmt as an attractive target for the development of selective inhibitors. Different families

Electronic supplementary material The online version of this article (doi:10.1007/s00894-012-1527-8) contains supplementary material, which is available to authorized users.

R. V. Bueno · N. R. Toledo · B. J. Neves · R. C. Braga · C. H. Andrade (✉)
Laboratório de Planejamento de Fármacos e Estudos de Metabolismo por Modelagem Molecular (LabMol), Faculdade de Farmácia, Universidade Federal de Goiás, 74605-220 Goiania, GO, Brazil
e-mail: carolina@farmacia.ufg.br

of nucleosidic and non nucleosidic compounds were reported with K_i values in the micromolar and low nanomolar range [9–16]. Various scaffolds were able to inhibit mycobacterial growth validating TMPKmt as a new target for the search of novel antituberculosis agents [17].

Previous QSAR analyses have been applied to smaller subsets of thymidine-based inhibitors. Based on a restricted set of 47 inhibitors of TMPKmt, Gopalakrishnan et al. [18] have developed 3D-QSAR molecular field analysis (MFA) models using different alignment methods. Receptor-based alignment proved to be more predictive and was claimed to be informative about the nature of the substituents required for activity. More recently, we have applied the 4D-QSAR formalism to develop QSAR models and corresponding 3D-pharmacophores for a set of 34 5'-thiourea-substituted α -thymidine derivatives [19], and to an extended set of 81 thymidine analogues with two corresponding subsets [20]. These studies have highlighted the importance of lipophilic substituents on the 5'-aryl moiety, which suggests the possibility of an additional pharmacophore site responsible for the higher inhibition potency of these derivatives. The drawbacks to doing this 4D-QSAR analysis were largely a consequence of the relatively limited size and structural diversity of the data set used in the study, which are important requirements for QSAR studies [20].

This scenario prompted us to investigate a larger and more chemically diverse series of 97 thymidine-based analogues reported as inhibitors of TMPKmt, using a specialized conformation-independent fragment-based method to develop predictive 2D-QSAR models by employing the hologram QSAR (HQSAR) method. It is noteworthy that structure- and ligand-based approaches have become vital components of many modern drug design projects [19–22]. Thus, the generated molecular recognition patterns from HQSAR were combined with three-dimensional molecular modeling studies, by using quantum mechanics/molecular mechanics (QM/MM) docking, as a fundamental step on the path to understanding the molecular basis of ligand – receptor interactions within this new series of potent anti-TB agents. The results revealed important molecular requirements for the design of new TMPKmt inhibitors with improved affinity.

Computational methods

Data set

The 97 TMPKmt inhibitors ranging from low micromolar to low nanomolar inhibition were collected from a series of papers by Van Calenbergh et al. and Familiar et al. [9–16]. The chemical diversity of the data set is very significant, since it encompasses acyclic nucleoside analogues (ATA), bicyclic thymidine analogues (ATB), 2' and 3' modified

thymidine analogues and thymidine 5'-O-monophosphate analogues (ddTMP) and 5'-arythiourea thymidine derivatives (ATT). The chemical structures and biological activity values for all compounds studied are listed in Table 1. The biological activity expressed in K_i were converted into their corresponding pK_i ($-\log K_i$) measures and used as dependent variables in the HQSAR analyses. It is important to mention that the values of K_i were selected from literature and measured under the same experimental conditions [23], which is considered essential for successful QSAR studies [24, 25]. The chemical structures of all TMPKmt inhibitors used in the modeling studies were constructed in the SYBYL-X 1.2 package (Tripos Inc., St. Louis, USA) and energetically optimized using the single point AM1-BCC semi-empiric method [26] as implemented in QUACPAC v1.5.0 (Open Eye Scientific Software, Santa Fe, NM, USA). Hierarchical cluster analyses of the data set were carried out within SYBYL-X 1.2.

HQSAR analysis

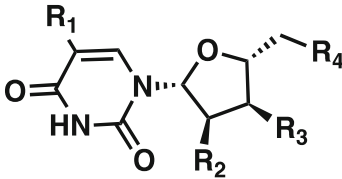
In this work, we have explored the molecular features related to the biological activity presented by a series of TMPKmt inhibitors using hologram QSAR (HQSAR) methodology [27], as this technique is considered a powerful ligand-based strategy in drug design. The HQSAR modeling analyses, calculations and visualizations were performed using the SYBYL-X 1.2 package (Tripos Inc., St. Louis, USA) running on Red Hat Enterprise Linux workstations. Parameters that affect hologram generation such as hologram length, fragment size and fragment distinction are crucial parameters that were evaluated during HQSAR model generation. The molecular holograms generated were used as independent variables during the partial least squares (PLS) regression analyses to derive the HQSAR models. Leave-one-out (LOO) cross-validation was applied to determine the number of components that yield optimally predictive models. The patterns of fragment counts related to increasing inhibition were used in the three-dimensional molecular modeling studies.

QM/MM docking

To explore the 3D molecular basis of ligand – receptor interactions for this series of thymidine analogs within the TMPKmt binding pocket, we used the quantum mechanics/molecular mechanics (QM/MM) docking by using the Schrödinger QM-polarized ligand docking (QPLD) [28] workflow.

The 3D ligand structures were prepared with LigPrep v.2.5 (Schrödinger, LCC, New York, 2011). All possible ionization states were generated at pH 7.0 \pm 2.0 using Epik [29]. For chiral ligands, specified chiralities were retained. The number

Table 1 Chemical structure and biological activity (pK_i) of training and test sets compounds

Cpd	Structure	R ₁	R ₂	R ₃	R ₄	X	pK_i
ddTMP1 ^a		CH ₃	H	OH	NHCOCH ₃	-	4.05
ddTMP2		CH ₃	H	OH	N ₃	-	5.15
ddTMP3		CH ₃	H	OH	NH ₂	-	4.92
ddTMP4		CH ₃	OH	OH	OH	-	3.15
ddTMP5		CF ₃	H	OH	OH	-	4.01
ddTMP6		C ₂ H ₅	H	OH	OH	-	2.94
ddTMP7 ^a		F	OH	OH	OH	-	3.28
ddTMP8		F	OH	OH	H	-	3.25
ddTMP9		CH ₃	H	NH ₂	OH	-	3.64
ddTMP10		CH ₃	H	F	OH	-	4.55
ddTMP11		CH ₃	F	OH	OH	-	3.91
ddTMP12 ^a		CH ₃	OH	NH ₂	H	-	2.72
ddTMP13		CH ₃	OH	NHC(NH)NH ₂	OH	-	3.85
ddTMP14		CH ₃	H	CH ₂ N ₃	OPO ₃ ²⁻	-	4.92
ddTMP15 ^a		CH ₃	H	CH ₂ NH ₂	OPO ₃ ²⁻	-	4.98
ddTMP16		CH ₃	H	CH ₂ F	OPO ₃ ²⁻	-	4.82
ddTMP17 ^a		CH ₃	H	CH ₂ OH	OPO ₃ ²⁻	-	4.54
ddTMP18		CH ₃	OH	CH ₂ N ₃	OPO ₃ ²⁻	-	3.93
ddTMP19		CH ₃	OH	CH ₂ NH ₂	OPO ₃ ²⁻	-	3.50
ddTMP20		CH ₃	H	CH ₂ N ₃	OH	-	4.40
ddTMP21		CH ₃	H	CH ₂ NH ₂	OH	-	4.24
ddTMP22		CH ₃	H	CH ₂ F	OH	-	4.35
ddTMP23		CH ₃	H	CH ₂ OH	OH	-	4.39
ddTMP24		CH ₃	OH	CH ₂ N ₃	OH	-	3.11
ddTMP25		CH ₃	OH	CH ₂ NH ₂	OH	-	3.41
ddTMP26 ^a		CH ₃	H	CH ₂ CH ₂ OH	OH	-	3.81
ddTMP27		CH ₃	NH ₂	OH	OPO ₃ ²⁻	-	4.26
ddTMP28		CH ₃	Cl	OH	OPO ₃ ²⁻	-	4.72
ddTMP29 ^a		CH ₃	F	OH	OPO ₃ ²⁻	-	4.37
ddTMP30		C ₆ H ₅ CH ₂	H	OH	OPO ₃ ²⁻	-	4.55
ddTMP31 ^a		CH ₃	H	N ₃	OPO ₃ ²⁻	-	5.00
ddTMP32 ^a		CH ₃	H	N ₃	OH	-	4.55
ddTMP33		Br	H	N ₃	OH	-	4.98
ddTMP34		CH=CHBr	H	OH	OH	-	3.20
ddTMP35		CH ₂ OH	H	OH	OH	-	3.09
ddTMP36		Cl	H	N ₃	OH	-	4.80
ddTMP37 ^a		CH ₃	H	OH	OH	-	4.57
ddTMP38		H	H	OH	OH	-	2.99
ddTMP39		F	H	OH	OH	-	3.67
ddTMP40		OH	H	OH	OH	-	3.57

of stereoisomers per molecule was changed from its default value of 32 to 128. For each structure, four low-energy ring conformations were generated. The ligand conformations were then generated using the MacroModel package by a mixed torsional/low-mode sampling method with the OPLS-

2005 force field [30] in implicit water without any constraints. The lowest potential energy conformers were retained as input for docking studies.

The X-ray crystallographic structure of TMPKmt in complex with the substrate deoxythymidine monophosphate

Table 1 (continued)

Cpd	Structure	R ₁	R ₂	R ₃	R ₄	X	p <i>K_i</i>	
ATT41		H	H	-	-	-	4.16	
ATT42		Cl	H	-	-	-	4.68	
ATT43		OCH ₃	H	-	-	-	4.34	
ATT44 ^a		CH ₃	H	-	-	-	4.44	
ATT45		Cl	Cl	-	-	-	5.14	
ATT46 ^a		Cl	CF ₃	-	-	-	5.30	
ATT47 ^a		-	-	-	-	S	4.80	
ATT48		-	-	-	-	S	5.50	
ATT49 ^a		-	-	-	-	S	5.00	
ATT50		-	-	-	-	S	5.11	
ATT51		-	-	-	-	S	6.00	
ATT52		-	-	-	-	S	6.22	
ATT53		-	-	-	-	S	4.72	
ATT54			-	-	-	-	S	4.82
ATT55 ^a			-	-	-	-	S	4.75
ATT56			-	-	-	-	S	5.27
ATT57		-	-	-	-	S	5.27	

(dTMP) was retrieved from the Protein Data Bank (PDB ID 1G3U, resolution 1.95 Å) [8]. The crystal waters were removed with the exception of the active site water molecules (1002, 1009, 1014, 1018, 1026, 1041 and 1050), which mediate the interaction between the protein and ligand. Then, the structure was prepared with the Protein Preparation Wizard workflow as follows: hydrogen atoms

were added according to Epik [29] calculation for p*K_a* values and protonation states were assigned, assigning partial charges using the OPLS-2005 force field [30]. Following this step, the structure underwent restrained minimization in vacuum. The minimization was carried out using the OPLS-2005 force field and terminated when the root-mean-square deviation (RMSD) reached a maximum cutoff of 0.30 Å. The grid with

Table 1 (continued)

Cpd	Structure	R ₁	R ₂	R ₃	R ₄	X	pK _i
ATT58			-	-	-	S	5.31
ATT59			-	-	-	S	5.14
ATT60			-	-	-	S	5.59
ATT61			-	-	-	S	5.42
ATT62 ^a			-	-	-	S	5.66
ATT63			-	-	-	O	5.96
ATT64			-	-	-	O	5.72
ATT65 ^{a*}			-	-	-	S	5.64
ATT66 ^{**}			-	-	-	S	5.42
ATT67			-	-	-	S	4.58
ATT68			-	-	-	-	3.59
ATT69			-	-	-	-	4.43
ATT70			-	-	-	-	4.46
ATT71			-	-	-	-	4.80
ATA72 ^a		-	-	-	-	CO	6.38
ATA73		-	-	-	-	SO ₂	6.57

dimensions of 22 Å × 22 Å × 22 Å centered on the thymidine ring of dTMP was generated.

In QPLD protocol we used Glide Extra Precision (XP) docking. For the Glide-XP docking runs, standard

parameters were applied including van der Waals scaling for non-polar atoms (by 0.8) to include modest ‘induced fit’ effects, with up to 5 poses per ligand input conformation saved. For QPLD docking runs, the output docking poses

from Glide-XP docking with OPLS-AA (2001) partial charges were used to obtain electrostatic potential (ESP) fit ligand charges using the program QSite and the QM/MM method in the ‘field’ of the receptor, and for comparison, ESP fit ligand charges in the gas phase (free ligand; no receptor) using QM. Single point energy calculation using the B3LYP density functional theory (DFT) method and 6–31 G*/LACVP* basis sets on the ligand (QM part) and the OPLS-AA force field for the receptor (MM part) were used for this purpose. The ligands with QM/MM modified charges were then re-docked and re-scored into the catalytic site using Glide-XP, and ten poses of each ligand were saved. Finally, the poses were selected by Emodel Energy value.

Results and discussion

The dataset of 97 thymidine analogs used for the HQSAR analyses is shown in Table 1. The values of pK_i span approximately five orders of magnitude and are acceptably distributed across the pK_i range values. The generation of consistent statistical models depends on the quality of both training and test sets in terms of structural diversity and property value distributions. Training and test sets were carefully selected in such a way that structurally-diverse molecules of a wide range of biological activities were included in both sets (Fig. 1). From the original data set, 74 compounds were selected as members of the training set for model generation, whereas the other 23 compounds were selected as members of the test set for external model validation. A statistical cluster analysis confirmed that the composition of both training and test sets is representative of the whole data set, as can be seen in Fig. 1. Thus, the data set is appropriate for the purpose of QSAR model development.

HQSAR models were resultant of a series of TMPKmt inhibitors with antituberculosis activity (Table 1). HQSAR relates biological activity to structural fragments. As HQSAR

models can be affected by a number of parameters concerning hologram generation, several combinations of fragment distinction were considered during the QSAR modeling runs. Holograms were generated using different combinations of atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor and acceptor (DA) as fragment distinctions. We performed the HQSAR analyses by screening the 12 default series of hologram length values ranging from 53 to 401 bins, initially using the fragment size default (4–7). The patterns of fragment counts from the training set inhibitors were related to the experimental biological activity using the partial least square PLS analysis. The statistical results obtained from PLS analyses using several fragment distinction combinations and the default fragment size (4–7) are presented in Table 2.

According to Table 2, the best statistical results among all models using the training set compounds were obtained for model 14 ($r^2=0.94$, $q^2=0.76$), which was resultant using the following combination of fragment distinctions: B, C and DA, with six being the optimum number of PLS components. This indicates that bonds, connections and donor and acceptor atoms are essential features of the molecular structures for biological activity.

The influence of fragment size is of fundamental importance in the generation of HQSAR models, as this parameter controls the minimum and maximum lengths of fragments to be encoded in the hologram fingerprint. Hence, distinct fragment size combinations (2–5, 3–6, 4–7, 5–8, 6–9 and 7–10) were investigated for the best model (model 14, Table 2) generated with the fragment size default (4–7). The HQSAR results obtained for several fragment sizes are displayed in Table S1 (Supplementary material), but no improvements were achieved in the statistical parameters.

Although a measure of internal consistency, available in the forms of q^2 and r^2 , is certainly important and significant, the most valuable test of a QSAR model is its ability to predict the activity of compounds not included in the training set. The structure encoded within a 2D fingerprint is directly related to biological activity of molecules within the training set, the high quality HQSAR models generated in this study can predict the activity of new structurally-related thymidine analogs from its fingerprint. In this way, the predictive power of the best HQSAR model derived from the training set molecules (fragment distinction B/C/DA; fragment size 4–7, Table 1) was assessed by predicting the pK_i values for the test set compounds. The external validation results are listed in Table 3, and the graphic results for the experimental versus predicted activities of both compound sets (training and test sets) are displayed in Fig. 2.

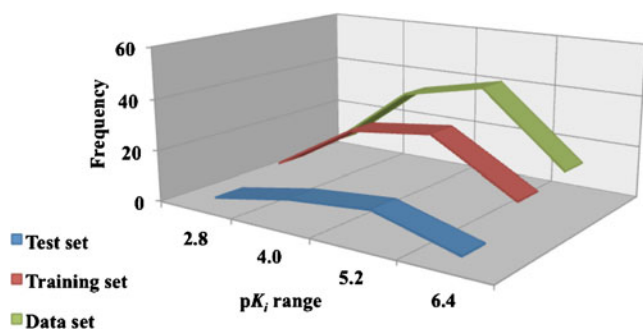


Fig. 1 Distribution of the biological activity values for the training set (red), test set (blue) and complete data set (green)

Table 2 HQSAR analyses for various fragment distinctions on the key statistical parameters using fragment size default (4–7)

Model	Distinction	q^2	r^2	SEE	HL	N
1	A/B	0.56	0.86	0.37	257	6
2	A/B/C	0.54	0.87	0.35	353	6
3	A/B/C/H	0.61	0.88	0.34	307	6
4	A/B/C/H/Ch	0.67	0.90	0.32	199	6
5	A/B/C/H/Ch/DA	0.73	0.93	0.27	353	6
6	A/B/C/H/DA	0.73	0.92	0.27	353	6
7	A/B/C/Ch/DA	0.69	0.89	0.32	401	5
8	A/B/H/Ch/DA	0.74	0.91	0.29	401	6
9	A/H/Ch/DA	0.73	0.92	0.28	307	6
10	A/B/H/CH	0.54	0.85	0.38	199	6
11	A/B/Ch/DA	0.64	0.88	0.33	307	6
12	A/B/H/DA	0.69	0.91	0.30	401	6
13	A/B/DA	0.67	0.88	0.33	199	6
14	B/C/DA	0.76	0.94	0.24	401	6

q^2 cross-validated correlation coefficient, r^2 noncross-validated correlation coefficient, *SEE* standard error of estimate, *HL* hologram length, N optimal number of components. Fragment distinction: *A* atoms, *B* bonds, *C* connections, *H* hydrogen atoms, *Ch* chirality, *DA* donor and acceptor

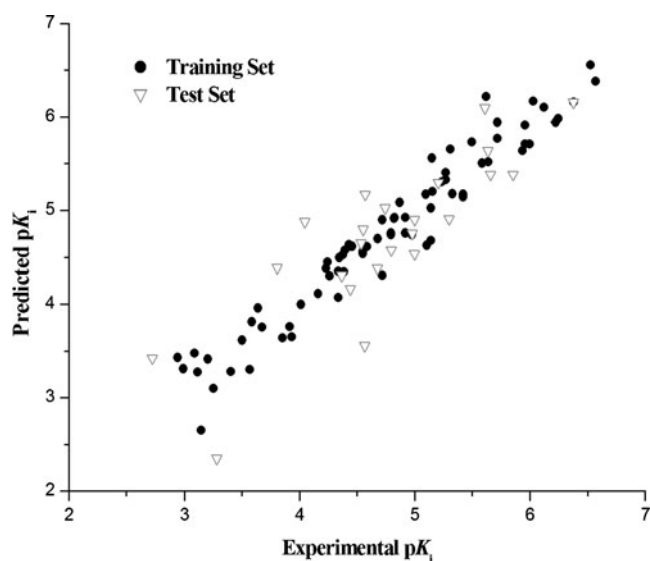
The external validation is a reliable process, as the test set compounds are completely excluded during the training of the

Table 3 Experimental and predicted biological property (pK_i), along with residual values, for the test set

Compound	pK_i		
	Experimental	Predicted	Residual
ddTMP1	4.05	4.88	-0.83
ddTMP7	3.28	2.35	0.94
ddTMP12	2.72	3.42	-0.70
ddTMP15	4.98	4.75	0.23
ddTMP17	4.54	4.65	-0.11
ddTMP26	3.81	4.39	-0.58
ddTMP29	4.37	4.31	0.06
ddTMP31	5.00	4.90	0.10
ddTMP32	4.55	4.80	-0.25
ddTMP37	4.57	3.56	1.01
ATT44	4.44	4.16	0.28
ATT46	5.30	4.91	0.39
ATT47	4.80	4.58	0.22
ATT49	5.00	4.53	0.47
ATT55	4.75	5.02	-0.28
ATT62	5.66	5.38	0.28
ATT65	5.64	5.64	0.00
ATA72	6.38	6.16	0.94
ATA78	5.61	6.10	0.22
ATA85	5.21	5.30	-0.49
ATA86	5.85	5.38	0.47
ATB91	4.68	4.38	0.30
ATB96	4.57	5.17	-0.60

model. The very good agreement between experimental and predicted pK_i values for the test set compounds indicates the robustness of the HQSAR model ($r^2_{pred}=0.89$). As can be seen in Table 3, the predicted values fall close to the experimental pK_i values, deviating by less than 0.50 log units for 16 compounds, and by less than 0.95 log units for six compounds. For only one compound (**ddTMP37**) the predicted value is more substantially in error (1.01 log unit). The HQSAR model obtained is reliable and can be used to predict the biological activity of novel compounds within this structural class.

Besides predicting the biological activity of untested molecules, HQSAR models should also provide important

**Fig. 2** Predicted vs. experimental values of pK_i for training and test sets compounds

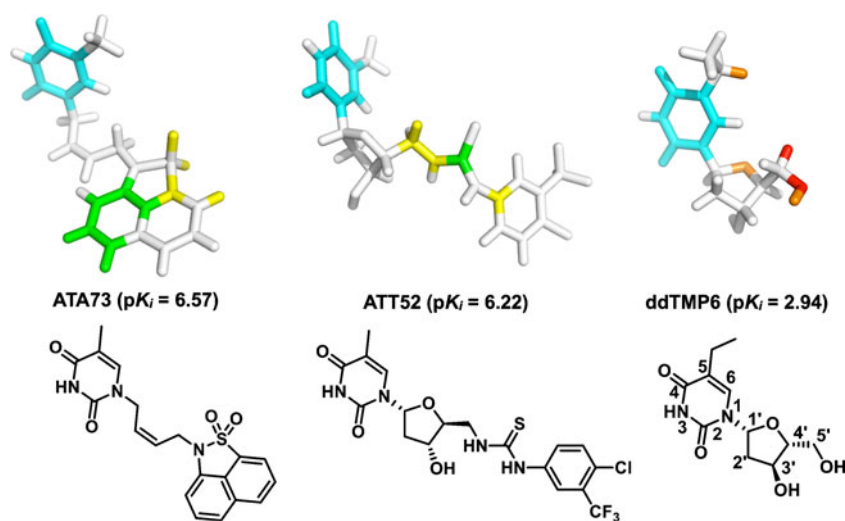
hints as to what molecular fragments are directly related to biological activity. This can be achieved through a careful interpretation of the structural fragments incorporated to the hologram-based QSAR models. HQSAR models can be graphically represented in the form of contribution maps where the color of each molecular fragment reflects the contribution of an atom or a small number of atoms to the activity of the molecule under study. The contribution map obtained from the HQSAR module implemented in SYBYL-X 1.2 uses a color scheme to discriminate individual atomic contributions to activity. The colors at the red end of the spectrum (red, red-orange and orange) reflect poor contributions, whereas colors at the green end (yellow, green-blue and green) reflect favorable contributions. Atoms with intermediate contributions are colored white. Atoms corresponding to the maximal common structure (MCS) are colored cyan, since it is common to all compounds and contributes in the same manner for all molecules in the training set. The most important fragments of compounds **ATT52** and **ATA73** (two of the most potent inhibitors of the data set) along with compound **ddTMP6** (the least potent inhibitor of the training set) are shown in Fig. 3. According to the contribution maps, the molecular fragments corresponding to the naphthosultam moiety (**ATA73**) and the thiourea moiety (**ATT52**) are strongly related to biological affinity (colored in green and yellow). The thymine ring is a common backbone to all compounds into the training set, and therefore, is colored cyan.

We also observed important structural features such as regions with poor contributions (colored in orange and red) that can be identified as potential targets for molecular modification and further SAR studies (Fig. 3). The main regions that negatively contribute to biological activity include the ethyl group at the 5-position of the thymine ring, the heteroatom oxygen of the sugar ring and the hydroxyl

group of the 5' position of the sugar ring (**ddTMP6**). These groups could be replaced by other substituents with different structural and physicochemical features with the aim to increase the affinity and potency of the compounds studied in this work. Some examples of such modifications could be the replacement of the ethyl group at the 5-position of the thymine ring by hydroxyl group in order to verify the importance of electronic interactions at this position. It is noteworthy that two of the most potent compounds in this data set (**ATA73** and **ATT52**) are an acyclic nucleoside analogue and a 5'-arylthiourea derivative, respectively, without the hydroxyl group at the 5'-position. This hydroxyl group has been found as an unfavorable contribution, according to our HQSAR model. Furthermore, the main fragments highlighted by the HQSAR model are directly related to important interactions that determine the preferred binding mode of the compounds studied and TMPKmt.

A valuable drug design strategy is the integration of structural and chemical analyses (ligand-based) with molecular modeling (receptor-based) studies to better understand the main protein–ligand interactions essential to biological activity. It is now well recognized that the accuracy of electric charges plays an important role in protein–ligand docking. The QM/MM docking aims to achieve the docking accuracy through improving the description of partial charges on the ligand atoms. The polarization of the charges on the ligand by the receptor is taken into account by replacing them with charges derived from quantum mechanical calculations in the field of the receptor, and re-docking of the ligands with QM/MM modified charges can result in improved docking accuracy. For this purpose, each dataset compound was docked in the active site of TMPKmt, along with its structural fragments highly related to biological activity (positive contributions), to be used to guide the search for intermolecular interactions within the thymidine

Fig. 3 HQSAR contribution maps and 2D chemical structures of the TMPKmt inhibitors **ATA73** ($pK_i=6.57$), **ATT52** ($pK_i=6.22$) and **ddTMP6** ($pK_i=2.94$)



analogues and TMPKmt, using the QPLD docking protocol. To validate the docking approach, re-docking was performed with XP/QPLD using the X-ray structure of TMPKmt in complex with the substrate dTMP [8]. The RMSD of the docked pose when compared to the crystallographically observed position of the dTMP was 0.95 Å, and thus, produced a docking mode that closely resembled the X-ray crystal structure. The binding mode of the whole data set was then explored by the QPLD protocol.

Figure 4 shows the best conformations for the most potent compound **ATA73** (a–b) and the least potent inhibitor **ddTMP6** (c–d) of the training set in the active site of TMPKmt obtained from QM/MM docking studies.

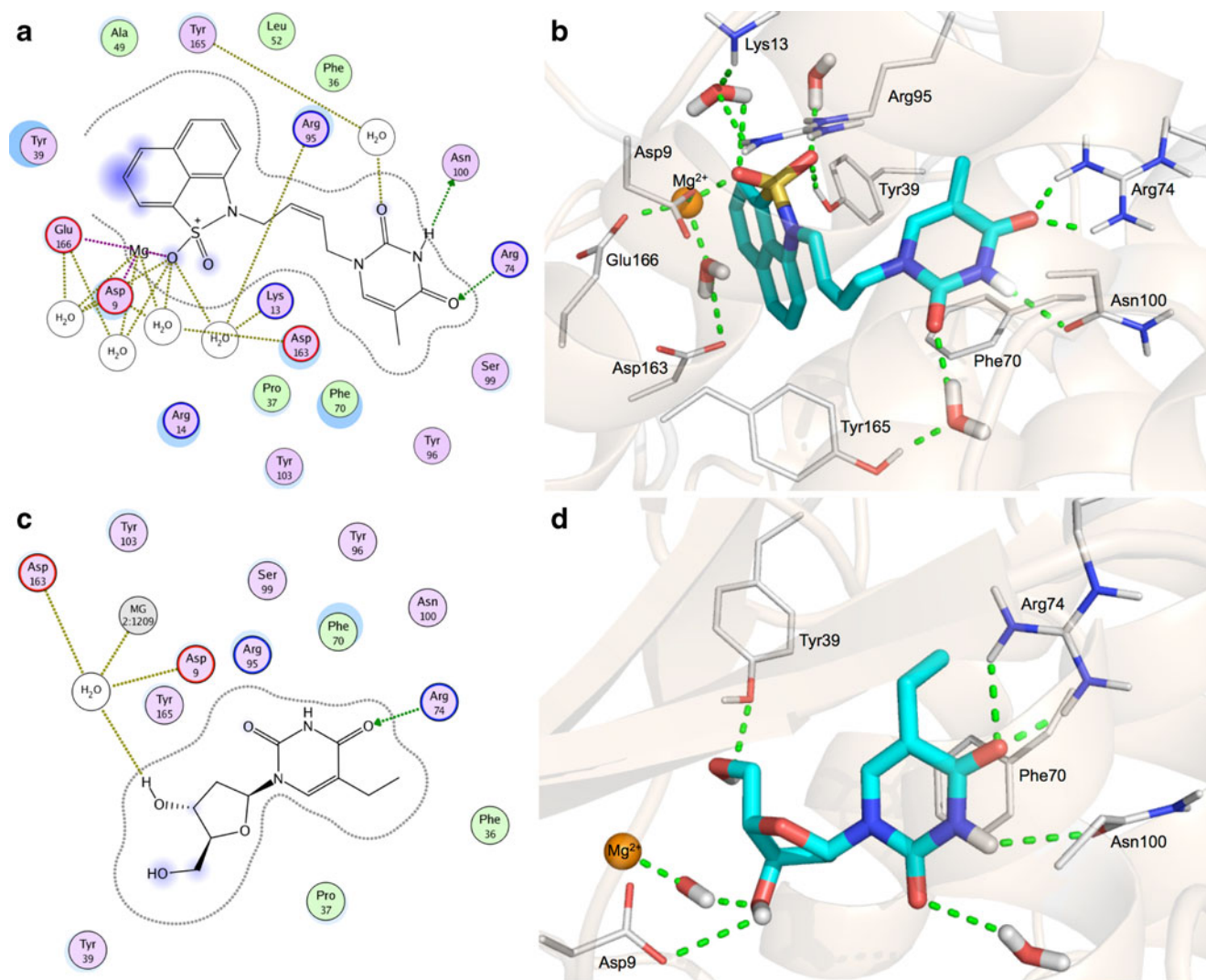


Fig. 4 The binding pose of the most potent compound **ATA73** and the least potent inhibitor **ddTMP6** of the training set in the active site of TMPKmt obtained from QM/MM docking studies. (a) and (c) Schematic representation between the compounds **ATA73** and **ddTMP6**, respectively, and the TMPKmt. (b) and (d) 3D representation of the inhibitors **ATA73** and **ddTMP6**, respectively, bound at the TMPKmt active site. Only the main interacting residues in the pocket of the

The substituents strongly related to the inhibitory activity, which were emphasized in the fragment-based HQSAR model, establish important hydrogen bonds and hydrophobic interactions within the TMPKmt binding cavity as shown in Fig. 4 (a–b). The thymine moiety of both inhibitors (**ATA73** and **ddTMP6**) binds inside the dTMP-binding cavity, with a stacking interaction between the thymine ring and Phe70, a hydrogen bond between the N3 of the thymine ring and the side-chain oxygen group of Asn100 and a hydrogen bond between O4 of thymine and the terminal guanidinium nitrogen atoms of Arg74. Importantly, these interactions of the thymine base are crucial for finding selectivity for TMPKmt versus TMPKh [16].

binding site are shown in line model representations (carbon atoms are in gray). The active site water molecules are presented in sticks (oxygen atoms in red). The inhibitors are also presented in stick models, but the carbon atoms are in cyan, nitrogen in blue, oxygen in red, and sulfur in yellow. The green dashes represent the proposed hydrogen bonds involved in binding of the inhibitor. The magnesium ion is represented as an orange sphere

According to the best docking conformation of the most potent inhibitor (**ATA73**), Figs. 4a and b, the SO₂ group at the distal substituent established key interactions with the hydroxyl group of Tyr39, with Asp163, Arg95 and Lys13, mediated by two water molecules and with the magnesium ion. Moreover, the Mg²⁺ appeared to be involved in a net of hydrogen bond interactions involving the active site water molecules and other amino acid residues, like Asp9 and Glu166. These interactions are important because they promote the destabilization of the interaction of the substrate dTMP with Mg²⁺ and suspend the catalytic process of the enzyme [8]. Additionally, the crucial interaction of Tyr39 residue with the most potent inhibitor was emphasized in the fragment-based HQSAR model and well observed by QM/MM docking study, that is responsible for keeping the active conformation of **ATA73** with the double bond of the butenyl unit fixed in the (*Z*) isomer. Notably, these observations are in agreement with previous studies that showed that the (*Z*)-butenyl derivatives have improved inhibitory potency against TMPKmt compared with the (*E*) counterparts [10]. The structural fragments of the naphthosultam moiety were highly related to biological activity, according to the HQSAR analysis, and therefore, could be explained by its interactions with Tyr39, Asp163, active-site water molecules and Mg²⁺.

Figure 4 (c–d) reveals that the least potent inhibitor of the data set, **ddTMP6**, binds to the phosphate acceptor binding site in a very similar fashion as the substrate dTMP [31]. The 3'-hydroxyl group of **ddTMP6** makes two polar contacts: one with a water molecule involved in the Mg²⁺ coordination and one with Asp9. The 3'-OH group of the substrate dTMP, which adopts a C2'-endo conformation in the active site, participates in Mg²⁺ stabilization through interactions involving a water molecule and Asp9, the conserved acidic residue of the P-loop, positioning the phosphate oxygen of dTMP. Modifications of the 3'- position aimed at destabilizing the interaction with the metal could form a firm basis to design compounds that specifically inhibit the parasite enzyme [17, 31]. Tyr39 makes a polar contact with the oxygen atom at the 5'-position of **ddTMP6**, which is the same interaction observed with the substrate, dTMP. According to our HQSAR model, the hydroxyl group of the 5' position of the sugar ring reflects poor contributions to biological activity, which could be explained by its interaction with Tyr39 in the same position of the substrate.

The obtained binding energies from QM/MM docking studies for two representatives of each subset, the most and the least potent inhibitor (**ddTMP**, **ATT**, **ATA** and **ATB**) are listed in Table 4.

The QPLD protocol gives a more accurate treatment of the electrostatic interactions, which results in an improvement of the docking accuracy. The choice of

Table 4 QM/MM docking results of selected TMPKmt inhibitors

Cpd ^a	p <i>K</i> _i ^b	GlideScore	Emodel ^c
ddTMP2	5.15	-7.51	-114.79
ddTMP6	2.94	-5.12	-78.88
ATT52	6.22	-8.90	-127.34
ATT68	3.59	-6.44	-154.49
ATA73	6.57	-9.42	-73.04
ATA81	4.38	-7.40	-69.81
ATB94	5.64	-8.10	-58.15
ATB92	4.23	-7.01	-116.70

^a For each subset, the first compound presented in the table is the most potent and the second one the least potent of this subset

^b Experimental values of biological activity

^c Emodel is a specific combination of GlideScore, the non-bonded interaction energy between the ligand and the receptor and the internal torsional energy of the ligand conformer, expressed in kcal/mol

best-docked pose for each individual inhibitor was made using a model energy score (Emodel) that combines the energy grid score, the binding affinity predicted by GlideScore, and the internal strain energy for the model potential used to direct the conformational-search algorithm. The ranking of ligands was based on the GlideScore. GlideScore, used for computing binding affinity, is an extension of the empirical score function ChemScore [32], but includes a steric-clash term, adds other rewards and penalties such as buried polar terms, amide twist penalties, hydrophobic enclosure terms, and excluded volume penalties, and has modifications to other terms:

$$\text{GlideScore} = 0.05 \cdot \text{vdW} + 0.15 \cdot \text{Coul} + \text{Lipo} + \text{Hbond} \\ + \text{Metal} + \text{Rewards} + \text{RotB} + \text{Site}$$

where, vdW => van der Waals energy; Coul => Coulomb energy; Lipo => lipophilic contact term; HBond => hydrogen-bonding term; Metal => metal-binding term; Rewards => penalties for various features, such as buried polar groups, hydrophobic enclosure, correlated hydrogen bonds, amide twists; RotB => penalty for freezing rotatable bonds; Site => polar interactions at the active site [33].

From examination of Table 4, one can conclude that the binding affinities (GlideScore) correlate quite well with the experimental p*K*_i values.

Finally, we have proposed modifications and designed new compounds based on the information gathered from the HQSAR contribution maps. The most important individual contributions (structural fragments) to the observed biological activity (p*K*_i) were inspected in order to provide a more complete interpretation of

the five best predictive HQSAR models (models 14, 8, 9, 5 and 6) in terms of their chemical and biological significance (Fig. S1, Supplementary material). The structural fragments with intermediate or poor contributions highlighted by the HQSAR models were replaced by bioisosteric groups using as scaffolds the compounds **ATT52** and **ATA73** in order to design new thymidine analogues having improved activity. The designed molecules were processed similarly as data set molecules and had their biological activities predicted by a consensus of the five best HQSAR models (Table S2, Supplementary material). The hits proposed were also docked into the active site of TMPKmt, using the same QM/MM docking protocol, and showed similar binding interaction pattern with a high docking score of -10.03 and -9.42 in comparison to acyclic nucleoside analogues (ATA) and score of -8.68 in comparison to arylthiourea thymidine derivatives (ATT). Table 5 shows the structure, HQSAR consensus predicted activity, Glide docking score, and some calculated physicochemical properties of the newly designed molecules. The physicochemical properties of the hits were calculated using QikProp v.3.4 (Schrödinger, LCC, New York, 2011).

The best predicted activities were found for **Hit1** and **Hit2**, which were designed based on the compound **ATA73**. It is noteworthy that the non-nucleosidic compound **ATA73** is the most potent inhibitor of the data set, with affinity in the sub-micromolar range ($pK_i=6.57$; $K_i=0.27 \mu\text{M}$) and was proved

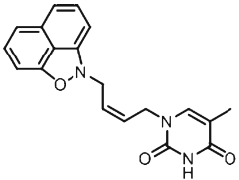
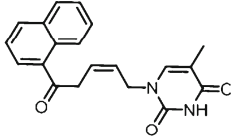
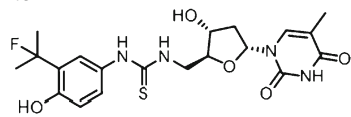
to be highly selective for TMPKmt versus TMPKh, without cytotoxicity on Vero cells ($100 \mu\text{g/mL}$). However, this compound showed no effect on the growth inhibition of mycobacteria, claimed to be due to its poor solubility [9]. Aiming at solving this problem, the newly designed compounds have lower molecular weight and are less lipophilic than **ATA73** ($\text{MW}=383.42$; $\log P=1.78$). Moreover, the three hits satisfy all the criteria of Lipinski's rule of five [34], indicating that these compounds are likely to be drug candidates.

In summary, the structural fragments highly related to the inhibitory activity were emphasized in the fragment-based HQSAR model. The presence of larger and more lipophilic groups at the 5'-position, encompassing the ATT, ATA and ATB derivatives, possibly represent an additional interaction site responsible for the higher inhibition potency of these derivatives. This finding is consistent with previously reported results [11], showing that α -thymidine analogues bind in the active site 'upside down' as compared to the natural substrate. The tails [5'-substituent] of these molecules are oriented to the outside of the enzyme through an exit channel, which is surrounded by nonpolar and aromatic residues including Ala35, Phe36, Pro37, and Arg160.

Conclusions

Our aim was to ascertain the state-of-the-art in the application of fragment-based and structure-based molecular

Table 5 2D structures, predicted biological activity, GlideScore, and physicochemical properties calculated by QikProp for three proposed hits

Compound	pK_i^a	GlideScore	MW ^b	$\log P^c$	HBD ^d	HBA ^e
Hit 1 	6.05	-10.03	335.36	2.48	1	5
Hit 2 	5.91	-9.42	334.37	3.03	1	5
Hit 3 	5.79	-8.68	452.50	1.79	5	10

^a Predicted biological activity using a consensus of the five best HQSAR models, ^b Molecular weight, ^c $\log P$ (octanol/water), ^d H-bond donors, ^e H-bond acceptors

modeling in order to better understand the intermolecular interactions between the thymidine analogues and TMPKmt binding site. The HQSAR model obtained shows both good internal and external consistency ($r^2=0.94$, $q^2=0.76$), with substantial predictive power ($r^2_{pred}=0.89$), indicating the reliability of the HQSAR model in predicting the biological activity of untested compounds. The generated molecular recognition patterns obtained from HQSAR combined with QM/MM docking provided important information into the chemical and structural basis involved in the molecular recognition process of this series of thymidine analogues, elucidating important interactions between the compounds and amino acids residues that promote the destabilization of the substrate interaction and suspend the catalytic process of the enzyme. The information gathered from the *in silico* approaches was used to design new active analogues of thymidine derivatives against TMPKmt with improved physicochemical properties. The proposed modifications aimed to keep the structural fragments that were more important to biological activity and replace the regions with intermediate contribution. Therefore, the insights provided by the combination of ligand- and structure-based techniques should be useful to design new anti-TB agents with improved affinity.

Acknowledgments The authors would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG) for financial support.

References

- WHO (2010) Global tuberculosis control. WHO/HTM/TB/2010.7.
- Andrade CH, Salum LDB, Mesquita Pasqualoto KF, Ferreira EI, Andricopulo AD (2008) Three-dimensional quantitative structure-activity relationships for a large series of potent antitubercular agents. *Lett Drug Des Discov* 5:377–387. doi:10.2174/157018008785777289
- Andrade CH, Pasqualoto KFM, Zaim MH, Ferreira EI (2008) Abordagem racional no planejamento de novos tuberculostáticos: inibidores da InhA, enoil-ACP redutase do *M. tuberculosis*. *Rev Bras Cienc Farm* 44. doi:10.1590/S1516-93322008000200002
- Burki T (2010) Tuberculosis-resistance, funding, and drugs. *Lancet Infect Dis* 10:297–298
- Dorman SE, Chaisson RE (2007) From magic bullets back to the magic mountain: the rise of extensively drug-resistant tuberculosis. *Nat Med* 13:295–298. doi:10.1038/nm0307-295
- Munier-Lehmann H, Chaffotte A, Pochet S, Labesse G (2001) Thymidylate kinase of *Mycobacterium tuberculosis*: a chimera sharing properties common to eukaryotic and bacterial enzymes. *Protein Sci* 10:1195–1205. doi:10.1110/ps.45701
- Ostermann N, Schlichting I, Brundiers R, Konrad M, Reinstein J, Veit T, Goody RS, Lavie A (2000) Insights into the phosphoryl-transfer mechanism of human thymidylate kinase gained from crystal structures of enzyme complexes along the reaction coordinate. *Structure* 8:629–642. doi:10.1016/S0969-2126(00)00149-0
- de la Sierra L, Munier-Lehmann H, Gilles AM, Bâzu O, Delarue M (2001) X-ray structure of TMP kinase from *Mycobacterium tuberculosis* complexed with TMP at 1.95 Å resolution. *J Mol Biol* 311:87–100. doi:10.1006/jmbi.2001.4843
- Familiar O, Munier-Lehmann H, Aínsa JA, Camarasa M-J, Pérez-Pérez M-J (2010) Design, synthesis and inhibitory activity against *Mycobacterium tuberculosis* thymidine monophosphate kinase of acyclic nucleoside analogues with a distal imidazoquinolinone. *Eur J Med Chem* 45:5910–5918. doi:10.1016/j.ejmech.2010.09.056
- Familiar O, Munier-Lehmann H, Negri A, Gago F, Douguet D, Rigouts L, Hernández A-I, Camarasa M-J, Pérez-Pérez M-J (2008) Exploring acyclic nucleoside analogues as inhibitors of *Mycobacterium tuberculosis* thymidylate kinase. *Chem Med Chem* 3:1083–1093. doi:10.1002/cmdc.200800060
- Van Daele I, Munier-Lehmann H, Froeyen M, Balzarini J, Van Calenbergh S (2007) Rational design of 5'-thioureia-substituted alpha-thymidine analogues as thymidine monophosphate kinase inhibitors capable of inhibiting mycobacterial growth. *J Med Chem* 50:5281–5292. doi:10.1021/jm0706158
- Van Daele I, Munier-Lehmann H, Hendrickx PMS, Marchal G, Chavarot P, Froeyen M, Qing L, Martins JC, Van Calenbergh S (2006) Synthesis and biological evaluation of bicyclic nucleosides as inhibitors of *M. tuberculosis* thymidylate kinase. *Chem Med Chem* 1:1081–1090. doi:10.1002/cmdc.200600028
- Vanheusden V, Munier-Lehmann H, Froeyen M, Dugué L, Heyerick A, De Keukeleire D, Pochet S, Busson R, Herdewijn P, van Calenbergh S (2003) 3'-C-branched-chain-substituted nucleosides and nucleotides as potent inhibitors of *Mycobacterium tuberculosis* thymidine monophosphate kinase. *J Med Chem* 46:3811–3821. doi:10.1021/jm021108n
- Vanheusden V, Munier-Lehmann H, Pochet S, Herdewijn P, van Calenbergh S (2002) Synthesis and evaluation of thymidine-5'-O-monophosphate analogues as inhibitors of *Mycobacterium tuberculosis* thymidylate kinase. *Bioorg Med Chem Lett* 12:2695–2698
- Vanheusden V, Munier-Lehmann H, Froeyen M, Busson R, Rozenski J, Herdewijn P, Van Calenbergh S (2004) Discovery of bicyclic thymidine analogues as selective and high-affinity inhibitors of *Mycobacterium tuberculosis* thymidine monophosphate kinase. *J Med Chem* 47:6187–6194. doi:10.1021/jm040847w
- Vanheusden V, Rompaeya PV, Munier-Lehmann H, Pochet S, Herdewijn P, Calenbergh SV (2003) Thymidine and thymidine-5'-O-monophosphate analogues as inhibitors of *Mycobacterium tuberculosis* thymidylate kinase. *Bioorg Med Chem Lett* 13:3045–3048. doi:10.1016/S0960-894X(03)00643-7
- Van Calenbergh S, Pochet S, Munier-Lehmann H (2012) Drug design and identification of potent leads against mycobacterium tuberculosis thymidine monophosphate kinase. *Curr Top Med Chem* 12:694–705
- Aparna V, Jeevan J, Ravi M, Desiraju GR, Gopalakrishnan B (2006) 3D-QSAR studies on antitubercular thymidine monophosphate kinase inhibitors based on different alignment methods. *Bioorg Med Chem Lett* 16:1014–1020. doi:10.1016/j.bmcl.2005.10.086
- Andrade CH, Pasqualoto KFM, Ferreira EI, Hopfinger AJ (2009) Rational design and 3D-pharmacophore mapping of 5'-thioureia-substituted alpha-thymidine analogues as mycobacterial TMPK inhibitors. *J Chem Inf Model* 49:1070–1078. doi:10.1021/ci8004622
- Andrade CH, Pasqualoto KFM, Ferreira EI, Hopfinger AJ (2010) 3D-Pharmacophore mapping of thymidine-based inhibitors of TMPK as potential antitubercular agents. *J Comput Aided Mol Des* 24:157–172. doi:10.1007/s10822-010-9323-y
- Salum LB, Dias LC, Andricopulo AD (2009) Structural and chemical basis for anticancer activity of a series of beta-Tubulin ligands: molecular modeling and 3D QSAR studies. *J Braz Chem Soc* 20:693–703

22. Salum LB, Valadares NF (2010) Fragment-guided approach to incorporating structural information into a CoMFA study: BACE-1 as an example. *J Comput Aided Mol Des* 24:803–817. doi:10.1007/s10822-010-9375-z
23. Blondin C, Serina L, Wiesmüller L, Gilles AM, Bârzu O (1994) Improved spectrophotometric assay of nucleoside monophosphate kinase activity using the pyruvate kinase/lactate dehydrogenase coupling system. *Anal Biochem* 220:219–221. doi:10.1006/abio.1994.1326
24. Young DC (2009) *Computational drug design: a guide for computational and medicinal chemists*. Wiley, Hoboken
25. Andrade CH, Pasqualoto KFM, Ferreira EI, Hopfinger AJ (2010) 4D-QSAR: perspectives in drug design. *Molecules* 15:3281–3294. doi:10.3390/molecules15053281
26. Jakalian A, Jack DB, Bayly CI (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* 23:1623–1641. doi:10.1002/jcc.10128
27. Salum LB, Andricopulo AD (2009) Fragment-based QSAR: perspectives in drug design. *Mol Divers* 13:277–285. doi:10.1007/s11030-009-9112-5
28. Cho AE, Guallar V, Berne BJ, Friesner R (2005) Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach. *J Comput Chem* 26:915–931. doi:10.1002/jcc.20222
29. Shelley JC, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M (2007) Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules. *J Comput Aided Mol Des* 21:681–691. doi:10.1007/s10822-007-9133-z
30. Banks JL, Beard HS, Cao Y, Cho AE, Damm W, Farid R, Felts AK, Halgren TA, Mainz DT, Maple JR, Murphy R, Philipp DM, Repasky MP, Zhang LY, Berne BJ, Friesner RA, Gallicchio E, Levy RM (2005) Integrated modeling program, applied chemical theory (IMPACT). *J Comput Chem* 26:1752–1780. doi:10.1002/jcc.20292
31. Van Calenbergh S (2006) Structure-aided design of inhibitors of Mycobacterium tuberculosis thymidylate kinase. *Verh K Acad Geneesk Belg* 68:223–248
32. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP (1997) Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des* 11:425–445
33. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 47:1739–1749. doi:10.1021/jm0306430
34. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 46:3–26