# Rational Design and 3D-Pharmacophore Mapping of 5′-Thiourea-Substituted α-Thymidine Analogues as Mycobacterial TMPK Inhibitors

Carolina H. Andrade,*[,†,‡] Kerly F. M. Pasqualoto,[†] Elizabeth I. Ferreira,[†] and Anton J. Hopfinger[‡]

Faculty of Pharmaceutical Sciences, Av. Prof. Lineu Prestes, 580, University of Sao Paulo, Sao Paulo, 05508-900, Brazil, and College of Pharmacy, MSC09 5360, 1 University of New Mexico, Albuquerque, New Mexico 87131-0001

Received December 19, 2008

Thymidine monophosphate kinase (TMPK) has emerged as an attractive target for developing inhibitors of *Mycobacterium tuberculosis* growth. In this study the receptor-independent (*RI*) 4D-QSAR formalism has been used to develop QSAR models and corresponding 3D-pharmacophores for a set of 5′-thiourea-substituted α-thymidine inhibitors. Models were developed for the entire training set and for a subset of the training set consisting of the most potent inhibitors. The optimized (*RI*) 4D-QSAR models are statistically significant ($r^2 = 0.90$, $q^2 = 0.83$ entire set, $r^2 = 0.86$, $q^2 = 0.80$ high potency subset) and also possess good predictivity based on test set predictions. The most and least potent inhibitors, in their respective postulated active conformations derived from the models, were docked in the active site of the TMPK crystallographic structure. There is a solid consistency between the 3D-pharmacophore sites defined by the QSAR models and interactions with binding site residues. This model identifies new regions of the inhibitors that contain pharmacophore sites, such as the sugar-pyrimidine ring structure and the region of the 5′-arylthiourea moiety. These new regions of the ligands can be further explored and possibly exploited to identify new, novel, and, perhaps, better antituberculosis inhibitors of TMPKmt. Furthermore, the 3D-pharmacophores defined by these models can be used as a starting point for future receptor-dependent antituberculosis drug design as well as to elucidate candidate sites for substituent addition to optimize ADMET properties of analog inhibitors.

## INTRODUCTION

Tuberculosis (TB) is a growing global health problem causing nearly two million deaths, and having eight million new cases reported, each year.[1] Although TB can be cured with the compliant use of proper drugs for at least six months, the emergence of multi- and extensively drug-resistant forms has created very significant new challenges for the treatment of the disease.[2] The combination of drug-resistance and TB/HIV coinfection is a major epidemiologic factor responsible for the global burden of TB.[1−3] Therefore, there is an urgent need for new drugs that can overcome resistance and are safe and effective for use in humans.

The development of new drugs that shorten current long therapy treatment as well as the discovery of new mycobacterial targets are essential to eradicate TB strains currently resistant to different types of drugs. A recently discovered target is thymidine monophosphate kinase (TMPKmt), which plays an essential and unique role in the DNA synthesis of the bacillus.[4] The elucidation of the TMPK X-ray structures of both human[5,6] and mycobacteria,[7] and their low (22%) sequence homology, enhances the consideration of TMPKmt as an attractive target for the development of selective inhibitors.

Thymidine is a moderate inhibitor of TMPKmt ($K_i \sim 27$ $\mu$M). Both the sugar[8−10] and the base[11,12] moiety of thymidine have been the subject of different modifications to enhance affinity and selectivity for the bacterial enzyme. Recently, Van Daele and co-workers[13] discovered a series of 5′-arylthiorea α-thymidine analogues with significant inhibitory activity against *M. tuberculosis* TMPK ($K_i = 0.6$ $\mu$M) and low human cytotoxicity.

4D-QSAR analysis[14] has been used to develop 3D pharmacophore models for ligand−receptor data sets. This methodology is able to extensively explore both conformational and alignment degrees of freedom in the search for an active conformation and binding mode to incorporate into a QSAR model. In this study, we report the application of the *receptor-independent* (*RI*) 4D-QSAR formalism to a series of 5′-thiourea-substituted α-thymidine derivatives which have been evaluated as inhibitors of thymidine monophosphate kinase from *M. tuberculsosis.* Although a crystal structure of the biomacromolecule target is available, the (*RI*) 4D-QSAR approach was applied in this study because of uncertainty in the binding mode of the ligands. The hypothesized active conformations resulting from (*RI*) 4D-QSAR analysis can be used as structural design templates, which includes their deployment as the molecular geometries of each of the ligands in structure-based ligand−receptor binding research.

## METHODS

**1. Biological Data.** A data set of 34 compounds reported as thymidine monophosphate kinase inhibitors was selected

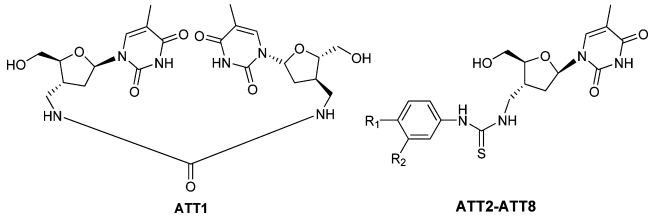
* Corresponding author phone: + 55 11 3091-3793; fax: +55 11 3815-4818; e-mail: carolhorta@usp.br. Corresponding author address: Department of Pharmacy, Faculty of Pharmaceutical Sciences, Av. Prof. Lineu Prestes 580, Bloco 13, Sao Paulo, SP 05508-900, Brazil.

† University of Sao Paulo.

‡ University of New Mexico.

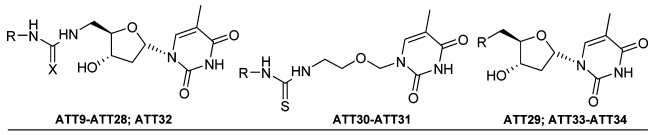
10.1021/ci8004622 CCC: $40.75 
Published on Web 03/19/2009

**Table 1.** Structural Features and p$K_i$ Values of Compounds ATT1−ATT8




| cpd | $R_1$ | $R_2$ | p$K_i$ |
|---|---|---|---|
| **ATT1** | − | − | 4.432 |
| **ATT2** | H | H | 4.161 |
| **ATT3** | Cl | H | 4.678 |
| **ATT4*** | $OCH_3$ | H | 4.337 |
| **ATT5** | $CH_3$ | H | 4.444 |
| **ATT6** | Cl | Cl | 5.143 |
| **ATT7** | $OCH_2Ph$ | H | 4.921 |
| **ATT8** | Cl | $CF_3$ | 5.301 |

**Table 2.** Structural Features and p$K_i$ Values of Compounds ATT9−ATT34

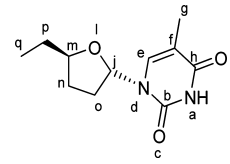



| cpd | R | X | p$K_i$ |
|---|---|---|---|
| **ATT9** | Ph | S | 4.796 |
| **ATT10*** | 4-Cl-Ph | S | 5.495 |
| **ATT11** | 4-MeO-Ph | S | 5.000 |
| **ATT12** | 4-Me-Ph | S | 5.108 |
| **ATT13** | 3,4-di-Cl-Ph | S | 6.000 |
| **ATT14** | 3-$CF_3$-4-Cl-Ph | S | 6.222 |
| **ATT15** | 4-morpholinophenyl | S | 4.717 |
| **ATT16*** | 1-adamantyl | S | 4.824 |
| **ATT17** | 3-pyridil | S | 4.745 |
| **ATT18** | fluoroesceinyl | S | 5.268 |
| **ATT19** | phenylmethyl | S | 5.268 |
| **ATT20** | benzhydryl | S | 5.310 |
| **ATT21** | benzoyl | S | 5.143 |
| **ATT22** | 3-$CF_3$-4-Cl-phenylmethyl | S | 5.585 |
| **ATT23** | phenylethyl | S | 5.420 |
| **ATT24** | 3,4-di-Cl-phenylethyl | S | 5.658 |
| **ATT25** | 3,4-di-Cl-Ph | O | 5.959 |
| **ATT26**[c] | 3-$CF_3$-4-Cl-Ph | O | 5.721 |
| **ATT27**[a] | 3-$CF_3$-4-Cl-Ph | S | 5.638 |
| **ATT28**[b] | 3-$CF_3$-4-Cl-Ph | S | 5.420 |
| **ATT29** | benzamido | − | 4.456 |
| **ATT30** | phenyl | S | 3.585 |
| **ATT31** | 3-$CF_3$-4-Cl-Ph | S | 4.432 |
| **ATT32** | 5′-deoxy-$\beta$-d-adenosin-5′-yl | S | 4.585 |
| **ATT33** | $N_3$ | − | 4.577 |
| **ATT34** | $NH_2$ | − | 4.796 |

[a] 3′-Deoxyribonucleoside. [b] 3′-Deoxy-2′−3′-didehydronucleoside. [c] The test set comprises the compounds ATT4, ATT10, ATT16, and ATT26. ATT = arylthiorea thymidine derivatives.

from ref[13] (Tables 1 and 2). The experimental inhibitory activities were converted into their corresponding p$K_i$ (−log-$K_i$) measures, where a $K_i$ value represents the inhibitor concentration that produces complete inhibition of dTMP phosphorylation by TMPKmt. All the $K_i$ values were obtained by the same assay method,[15] and the p$K_i$ values spanned a sufficiently wide range from 3 to 6. The structures and corresponding p$K_i$ values for the set of TMPKmt inhibitors are included in Tables 1 and 2. The compounds of both training and test sets were randomly selected subject to the constraint to ensure complete and representative coverage across the entire range of p$K_i$ values. The best models were externally validated using a four compound test set (**ATT4**, **ATT10**, **ATT16**, and **ATT26**) randomly selected from the original 34 compounds and not included in the 4D-QSAR model development process.

**Table 3.** Set of Trial Alignments Used in Constructing the 4D-QSAR Models



| align no. | first atom | second atom | third atom |
|---|---|---|---|
| 1 | a | b | d |
| 2 | i | d | m |
| 3 | a | c | l |
| 4 | g | j | p |
| 5 | l | m | p |
| 6 | a | g | q |
| 7 | a | i | q |
| 8 | a | d | q |
| 9 | a | c | q |
| 10 | a | l | q |

**2. (RI) 4D-QSAR Analysis.** *Structure Building and Conformational Sampling.* The three-dimensional (3D) structures of each of the 34 inhibitors in their neutral forms were constructed using the HyperChem 7.05 software.[16] The crystallized structure of deoxythymidine monophosphate (dTMP) cocrystallized with TMPKmt (1G3U, resolution 1.95 Å)[7] was used as a reference structure in constructing the set of inhibitors. The 3D model for each structure was subsequently energy-minimized, and the partial atomic charges were computed using the AM1[17] semiempirical method implemented in the HyperChem program.[16] The energy-minimized structures were used as the initial structures in each molecular dynamics simulation (MDS) employed to generate the conformational ensemble profile (CEP) of each inhibitor. The Molsim 3.2 software[18] was used to perform the MDS and to generate the trajectories for, in turn, deriving the CEP. The MDS protocol employed 100 000 steps for each inhibitor, the step size was 0.001 ps (1 fs), and the simulation temperature was 303 K, the same used in the biological assay.[15] An output trajectory file was generated by saving every twenty simulation steps leading to a CEP containing 5000 conformations.

*Alignment Definitions.* The (*RI*) 4D-QSAR methodology[19] uses three-ordered atom alignments to compare the molecules of a training set. Alignments are typically chosen to systematically span the common scaffold of the compounds. Atoms from all parts of the common scaffold of the compounds are represented in the chosen alignments to ensure a thorough alignment analysis. Spanning the scaffold provides information regarding the sensitivity for a preferred alignment and perhaps a corresponding binding mode, with respect to both the structure of the scaffold as well as the substituents of the training set analogs. In this study ten alignments that systematically explore the entire scaffold and substituents were selected and are defined in Table 3.

*Interaction Pharmacophore Elements (IPEs).* The 4D-QSAR methodology[14,19] currently defines seven types of interaction pharmacophore elements (IPEs), corresponding

to the atom types that may occupy the set of reference grid cells [see below] under a selected alignment. These IPEs correspond to the types of interactions that may occur upon ligand−receptor binding and are used to the ultimately define the pharmacophore groups. In this study all seven types of IPEs were used, and they are defined as follows: any type (Any), nonpolar (NP), polar-positive charge (P+), polar-negative charge (P-), hydrogen bond acceptor (HA), hydrogen bond donor (HB), and atom in aromatic systems (Ar).

*Grid Cell Occupancy Descriptors (GCODs) Generation.* Each conformation from the CEP, which consists of 5000 conformations generated by the MDS sampling for each inhibitor, was placed in a reference grid cell space according to the trial alignment under consideration. In this study, the selected size of the cubic grid cell was 1 Å, and the size of the overall grid cell lattice was chosen to enclose all inhibitors of the training set. The grid cell occupancy profiles for each of the seven IPEs atom types are then computed and used as the trial set of descriptors, which are referred to as grid cell occupancy descriptors (GCODs). Moreover, the absolute occupancy of each grid cell (i, j, k), where i, j, and k define the xyz-coordinate location in the Cartesian space of the grid cell, was computed at time t in the MDS ensemble sampling for the IPE atoms of each compound.[14] The normalized absolute occupancy of each grid cell by each IPE atom type over the CEP for a given alignment forms a set of GCODs unique to the training set. The GCODs were computed and used as the trial set of descriptors in this (*RI*) 4D-QSAR analysis.[14]

*Partial Least Squares (PLS) Data Reduction.* A 4D-QSAR analysis generates an enormous number of trial GCOD descriptors because of the large number of grid cells and the seven IPEs. Partial least-squares (PLS) regression analysis[20] is used to perform a data reduction fit between the observed dependent variable (the experimental biological activities, p$K_i$; Tables 1 and 2) and the corresponding GCOD values for each of the trial (ten) alignments. In this study, a variance-filtering constraint was applied to the entire set of GCODs prior to the PLS analysis. GCODs having a variance (self-variance) over the set of analogues less than 2.0 (prechosen fraction) were eliminated. The data reduction by PLS analysis provides the selection of those descriptors having the highest individual weightings to the observed biological activity measures.

*(RI) 4D-QSAR Model Optimization.* The GCODs with the highest individual weightings were used to generate optimized (*RI*) 4D-QSAR models using the GA-MLR approach.[21,22] The genetic function approximation (GFA) algorithm[21] in the GA-MLR approach combines Holland's genetic algorithm (GA) and Friedman's multivariate adaptive regression splines (MARS) algorithm.[22] GFA is implemented using multidimensional linear regression (MLR) analysis for performing data fitting.[22] Concisely, in the GA-MLR approach,[21,22] initial models (equations) are generated by randomly selecting features (descriptors) from the training data set, building basis functions (terms) from these descriptors using user-defined basis function types (e.g., linear most often), and then constructing a population of QSAR-genetic models (equations) from random combinations of these basis functions. Improved models are generated through repetitive crossover operations to recombine the terms of the better-scored models, according to the Friedman's lack-of-fit (LOF) measure.[21−23] The LOF score is a penalized least-squares error (LSE) measure, i.e., when two models have the same value of LSE, that one which has the lowest number of terms will have the lowest (best score) value of the LOF.[21,22] In addition to the crossover operations, a rate-of-mutation operation is allowed in order to ensure a diversity of models is integrated into the model population. At the end of the optimization process (evolution), the LOF value remains constant with respect to crossover operations, and QSAR model optimization has been performed.[21,22]

GFA optimizations were initiated using 200 randomly generated (*RI*) 4D-QSAR models. The mutation probability over the crossover optimization process was set as 10%. The smoothing factor controls the number of independent variables, GCODs, in the models and along with the LSE controls model overfitting. Smoothing factor values over the range of 1.0−2.5 were tested in order to determine the optimal number of descriptors in the (*RI*) 4D-QSAR models.[22,24,25] Other non-GCOD descriptors, like Clog P (the calculated water/octanol partition coefficient), molar refractivity (MR), etc., can be added to the trial descriptor set at the start of the GFA-MLR model optimization.[19]

Lipophilicity is a major determinant property for pharmacokinetic and pharmacodynamic profiles of drug molecules.[26] Thus, ClogP values were calculated for all inhibitors of the training set and included to the set of GCODs in the initial descriptors pool. The Ghose, Pritchett, and Crippen method (1998)[27] was used to compute the ClogP values using HyperChem 7.05.[16]

**3. Internal and External (*RI*) 4D-QSAR Model Validation.** The ten best scored models found by GA-MLR analysis,[21,22] according to their values of Friedman's LOF[23] measure, were tested for internal validation by the leave-one-out (LOO) cross-validation method available in the 4D-QSAR program.[19] Overall, the main statistical measures evaluated were as follows: $r^2$ (linear correlation coefficient), $q^2$ (LOO cross-validated correlation coefficient), SE (standard error), LOF (Friedman's lack-of-fit score), residuals (difference between the observed and calculated p$K_i$ values), and SD (standard deviation of the residuals). Compounds were considered as outliers when the difference between the observed and calculated biological data (residuals) exceeded twice the SD value.

Each alignment was evaluated using the composite procedure described above. For the best alignment, a cross-correlation matrix of the residuals in error between pairs of the top 10 (*RI*) 4D-QSAR models, based on maximizing $q^2$ as a function of number of descriptor terms, was constructed.[14,19] This has been done to determine if the top 10 (*RI*) 4D-QSAR models are providing common, or distinct, structure−activity information. In other words, it is possible to identify the set of best and unique (*RI*) 4D-QSAR models. Pairs of models with highly correlated residuals of fit ($R^2 \sim 1$) are judged to be nearly the same model, while pairs of models with poorly correlated residuals ($R^2 < 0.5$) are considered to be distinct from one another. Also, the linear cross-correlation matrix of the GCODs for the best (*RI*) 4D-QSAR model for the best alignment is used to determine if these significant GCODs are correlated to one another.

The robustness of each of the overall best models were explored and validated using an external four compound test set (**ATT4**, **ATT10**, **ATT16**, and **ATT26**). These four compounds were used to evaluate the best (*RI*) 4D-QSAR models regarding their ability to predict the p$K_i$ values of compounds not included in the training data set.

**Table 4.** Statistical Measures, Number of GCODs, and Number of Outliers for the Top Ten (*RI*) 4D-QSAR Models for Each Trial Alignments Using a Smoothing Factor of 2.0

| align | no. GCODs | $r^2$ | $q^2$ | no. outliers | LSE |
|---|---|---|---|---|---|
| 1 | 6−8 | 0.79−0.83 | 0.64−0.72 | 0−1 | 0.05−0.06 |
| 2 | 6−8 | 0.82−0.87 | 0.75−0.80 | 0−2 | 0.03−0.05 |
| 3 | 5−7 | 0.75−0.82 | 0.65−0.72 | 0−2 | 0.05−0.07 |
| 4 | 6−8 | 0.81−0.85 | 0.71−0.73 | 1−2 | 0.04−0.05 |
| 5 | 4−6 | 0.71−0.78 | 0.56−0.61 | 0−2 | 0.06−0.08 |
| 6 | 5−6 | 0.79−0.81 | 0.63−0.68 | 0−2 | 0.05−0.06 |
| **7** | **6−7** | **0.89−0.90** | **0.83−0.84** | **0−2** | **0.02−0.03** |
| 8 | 6−8 | 0.76−0.86 | 0.65−0.72 | 0−2 | 0.04−0.07 |
| 9 | 6−7 | 0.81−0.85 | 0.70−0.78 | 0−1 | 0.04−0.05 |
| 10 | 6−7 | 0.79−0.81 | 0.72−0.78 | 0−2 | 0.03−0.05 |

**Table 5.** Statistical Measures, Number of GCODs, and Number of Outliers for the Top Ten (*RI*) 4D-QSAR Models from Alignment 7

| model | no. GCODS | $r^2$ | $q^2$ | LSE | LOF | no. outliers |
|---|---|---|---|---|---|---|
| 1 | 7 | 0.90 | 0.84 | 0.02 | 0.10 | 1 |
| 2 | 6 | 0.90 | 0.84 | 0.02 | 0.08 | 1 |
| 3 | 7 | 0.90 | 0.84 | 0.02 | 0.10 | 1 |
| **4** | **6** | **0.90** | **0.83** | **0.02** | **0.08** | **0** |
| 5 | 7 | 0.90 | 0.83 | 0.02 | 0.10 | 1 |
| 6 | 7 | 0.90 | 0.83 | 0.02 | 0.10 | 1 |
| 7 | 7 | 0.90 | 0.83 | 0.02 | 0.10 | 1 |
| 8 | 6 | 0.89 | 0.83 | 0.03 | 0.09 | 1 |
| 9 | 7 | 0.89 | 0.83 | 0.03 | 0.12 | 2 |
| 10 | 6 | 0.90 | 0.83 | 0.03 | 0.08 | 0 |

**4. 4D-QSAR Model and Bioactive Conformation Selections.** A final best (*RI*) 4D-QSAR model was selected based upon the alignment and statistical fitting measures. The lowest-energy conformer states (allowing up to 2.0 kcal/mol from the global minimum energy conformation of the CEP), which predicted the maximum p$K_i$ value using the optimum 4D-QSAR model, was defined as the active conformation.[14] Postulated bioactive conformations can be used as structure design templates in other CAMD approaches including structure-based design, as reality checks of the 4D-QSAR models and for virtual screening.[19]

The putative bioactive conformations of the inhibitors were manually docked to the crystal structure active site of TMPKmt (PDB code 1G3U), using the HyperChem 7.05 software,[16] in order to elucidate the possible types of interactions and pharmacophore sites considering the surrounding amino acid residues. The MOLSIM 3.2 program[18] was used to perform energy minimization and MDS relaxation of each of the docked complexes.

## RESULTS

(*RI*) 4D-QSAR models were constructed and optimized for each of the 10 trial alignments listed in Table 3. The number of GCODs, statistical measures, and the number of outliers of each of the top 10 models are presented in Table 4 for each selected alignment when using a smoothing factor of 2.0 in the GA optimization.

Alignment 7 was judged to provide the best (*RI*) 4D-QSAR models because it yields models which collectively have the highest $q^2$ values. Table 5 shows the top 10 (*RI*) 4D-QSAR models built from alignment 7. High values of both $q^2$ and $r^2$ for all models can be observed. However, models 1−3 and 5−9 have at least one outlier and consequently were not considered further in the analysis.

**Table 6.** Linear Cross-Correlation Matrix of the Residuals of Fit for the Top Ten (*RI*) 4D-QSAR Models from Alignment 7

| model | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.00 | | | | | | | | | |
| 2 | 0.92 | 1.00 | | | | | | | | |
| 3 | 0.90 | 0.90 | 1.00 | | | | | | | |
| 4 | 0.85 | 0.94 | 0.96 | 1.00 | | | | | | |
| 5 | 0.83 | 0.94 | 0.94 | 0.98 | 1.00 | | | | | |
| 6 | 0.85 | 0.75 | 0.91 | 0.82 | 0.77 | 1.00 | | | | |
| 7 | 0.93 | 0.95 | 0.83 | 0.86 | 0.84 | 0.72 | 1.00 | | | |
| 8 | 0.85 | 0.95 | 0.96 | 1.00 | 0.98 | 0.80 | 0.86 | 1.00 | | |
| 9 | 0.87 | 0.74 | 0.92 | 0.81 | 0.78 | 0.95 | 0.73 | 0.80 | 1.00 | |
| 10 | 0.85 | 0.94 | 0.96 | 1.00 | 0.98 | 0.82 | 0.86 | 1.00 | 0.81 | 1.00 |

The cross-correlation matrix of the residuals of fit between pairs of models from alignment 7 was computed and is given in Table 6. The data in Table 6 indicate that all of the top 10 4D-QSAR models have residuals of fit that are highly correlated to one another. Thus, there is a single unique 4D-QSAR model among these 10 models, and that model with the highest $q^2$ value (model 4) was selected as most exemplary of this set and is given by

$$\begin{aligned} p\mathit{K}_i = {} & 5.40\,\mathrm{GC1(Any)} - 4.77\,\mathrm{GC2(P+)} + 2.18\,\mathrm{GC3(Any)} + \\ & 14.11\,\mathrm{GC4(Any)} + 1.14\,\mathrm{GC5(Any)} + 0.17\,(\mathrm{ClogP})^2 + 2.71 \\ & N = 30; r^2 = 0.90; q^2 = 0.83, \mathrm{LSE} = 0.02 \quad (1) \end{aligned}$$

This model is composed of only two types of IPEs, namely polar-positive charge (P+) and any type of atom (Any). Moreover, the best 4D-QSAR model includes the calculated quadratic descriptor of lipophilicity $(\mathrm{ClogP})^2$ with a positive contribution to inhibition potency. Increasing occupancy of only one GCOD (GC2) is seen to decrease p$K_i$ [the negative regression coefficient]. GC2 corresponds to occupancy by a (P+) IPE type. The observed and predicted p$K_i$ values of the training set, using eq 1, are plotted in Figure 1.

In Table 7 are presented the linear cross-correlation matrix of the descriptors found in the best (*RI*) 4D-QSAR model (eq 1). Only one pair of GCODs is moderately correlated to one another ($r > 0.7$), and their $r$ value is given in bold print (see Table 7). Some cross-correlations are negative. One interpretation of such a negative relationship is that occupancy of one GCOD by the appropriate inhibitor atom types forces a conformational change in the inhibitor preventing occupancy of the other GCOD. But it is important to note that the correlation coefficient values are all rather small, save for the one in bold. Thus, the small cross-correlations between pairs of GCODs, be they positive or negative, may not have much overall significance.

The "bioactive" conformation of each inhibitor in the training set was hypothesized using model 4 (represented by eq 1) from alignment 7 by first identifying all conformer states sampled for each inhibitor within $\Delta E$ equal to 2 kcal/mol of the global minimum energy conformation of its CEP. The GCODs of each resulting set of low-energy conformations were employed to predict the activities for each inhibitor using eq 1, and the conformer with the highest predicted p$K_i$ value was selected as the "bioactive" conformation of each inhibitor. Figure 2 shows the most active compound, **ATT14**, and the most inactive compound, **ATT30**, each in its respective predicted bioactive conformation added by a representation of the 3D-pharmacophore embedded in the (*RI*) 4D-QSAR model given by eq 1.

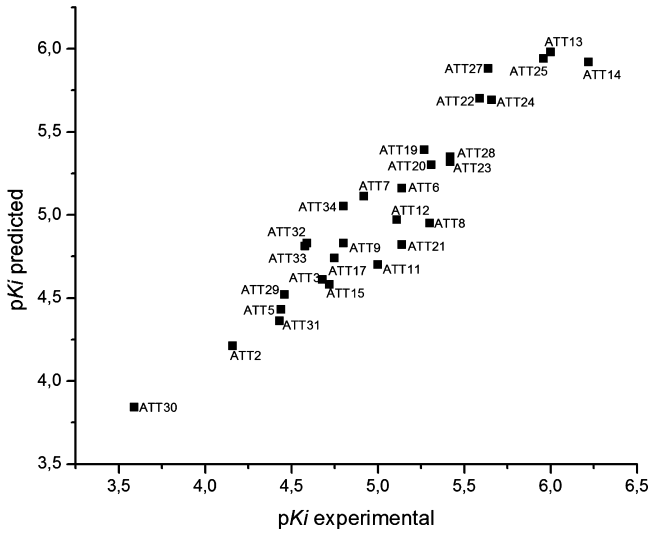



**Figure 1.** Predicted and observed activities ($pK_i$) of each compound of the training set for alignment 7 using eq 1.

**Table 7.** Linear Cross-Correlation Matrix of the GCODs for the Optimal (*RI*) 4D-QSAR Model (Eq 1)

| | GC1 | GC2 | GC3 | GC4 | GC5 | $(ClogP)^2$ |
|---|---|---|---|---|---|---|
| GC1 | 1.00 | | | | | |
| GC2 | −0.41 | 1.00 | | | | |
| GC3 | −0.39 | −0.41 | 1.00 | | | |
| GC4 | −0.39 | 0.41 | −0.35 | 1.00 | | |
| GC5 | **0.84** | −0.44 | −0.29 | −0.28 | 1.00 | |
| $(ClogP)^2$ | −0.39 | 0.20 | −0.01 | −0.15 | −0.36 | 1.00 |

GCODs that increase $pK_i$ are shown as yellow spheres (GC1, GC3, GC4, and GC5), and the GCOD that decreases the inhibition potency is shown as a red sphere (GC2). The yellow color intensity of the GCODs (spheres) is proportional to the magnitude of the regression coefficients in eq 1. The larger the absolute value of the regression coefficient, the more intense is the yellow color of its corresponding GCOD sphere (see Figure 2).

To ascertain the predictive power of model 4, the four compound test set (see Tables 1 and 2) was employed as an external validation set. The $pK_i$ value of each test set inhibitor was calculated using eq 1. Table 8 lists the test set GCOD indices found for model 4 (eq 1). The test set predictions are given in Table 9.

Three of the four inhibitors of the test set have residuals whose absolute values are less than or equal to 0.90, which is the SD value. This finding indicates that model 4 has a predictability of 75%, which is a reasonably good external predictive power. The only test set compound presenting poorly predicted activity was compound **ATT26**.

It is noteworthy that for most models from all alignments the most potent compound (**ATT14**) was frequently an outlier. Thus, in order to see if a separate (*RI*) 4D-QSAR model was needed to better represent the more potent inhibitors, some of the least potent inhibitors were removed (**ATT1, ATT2, ATT5, ATT30**, and **ATT31**) from the training set. A (*RI*) 4D-QSAR analysis was then carried out using only alignment 7 (the best alignment) for this new "most potent" training set. Table 10 contains the statistical measures, the number of GCODs, and the number of outliers of the top ten models from alignment 7 of the "most potent" training set.

The best (*RI*) 4D-QSAR model for the "most potent" training set is defined by

$$pK_i = -2.33\ GC1(P+) + 3.67\ GC2(Any) - 8.25\ GC3(Any) + 2.79\ GC4(HA) + 2.88\ GC5(Any) + 0.40\ ClogP + 4.28$$
$$N = 25; r^2 = 0.86; q^2 = 0.80, LSE = 0.03 \quad (2)$$

This model contains three classes of IPEs: Any, HA, and P+. In addition the "most potent" model also has the lipophilicity property, ClogP, as a linear term descriptor which presents a positive contribution to the inhibition potency. The $q^2$ and $r^2$ values of eq 2 decreased in comparison to eq 1 owing to a more narrow range in $pK_i$. Figure 3 shows the graphical representation of the bioactive conformations of the most active and least active inhibitors, **ATT14** and **ATT29**, of the "most potent" training set based upon the model represented by eq 2.

In order to gain a better understanding of the behavior of the fitted data to the models, the descriptors of the best models (eqs 1 and 2) were inspected and compared. GC2 from eq 1 corresponds to GC1 of the "most potent" (*RI*) 4D-QSAR model (eq 2). This descriptor is occupied by polar positive atom type (P+) with negative regression coefficients in both models, suggesting that the occupation of this region with a polar positive atom is detrimental to inhibition potency. Also, GC3 in eq 1 and GC2 in eq 2 are equivalent descriptors terms. These descriptors have a common IPE type (Any) and a positive regression coefficient in each model, indicating $pK_i$ increases with increasing occupancy by any IPE type sampled in the training set. Finally, GC5 (eq 1) and GC4 (eq 2), which are close in space, are occupied by any IPE type in eq 1 and a hydrogen-bond acceptor IPE type in eq 2. Each descriptor term has a positive regression coefficient. As already mentioned, both models (eqs 1 and 2) contain ClogP as a descriptor term indicating it is important to inhibition potency. The descriptors GC3 (Any) and GC5 (Any) of the "most potent" model do not have any direct corresponding GCODs in the model given by eq 1.

In terms of specifying a 3D-pharmacophore, the pharmacophore sites, as defined by the GCODs of eq 1, are largely restricted to the inhibitor region between the pyrimidine and sugar rings. However, the model 9 (eq 2) distributes the GCODs more fully across the entire molecule. Moreover, GC3 and GC5 from eq 2, which are not present in eq 1, are related to the 5′-arylthiorea moiety of the inhibitors (the 'tail') and may define an important binding feature only observed by these more active inhibitors.

Finally, the postulated bioactive conformation, as determined from eqs 1 and 2 [the best (*RI*) 4D-QSAR models], for compound **ATT14,** the most potent member of the training set, was also taken as the trial bioactive conformation for the series of inhibitors represented by the training set. The training set inhibitors in this postulated initial bioactive conformation were each docked into the active site of the crystal structure of TMPKmt (1G3U).[7] The binding pose was correspondingly based upon alignment 7. MOLSIM 3.2[18] was then used to perform a coupled energy minimization and MDS relaxation on each of the docked complexes. The goal of these docking-relaxation studies was to elucidate the possible

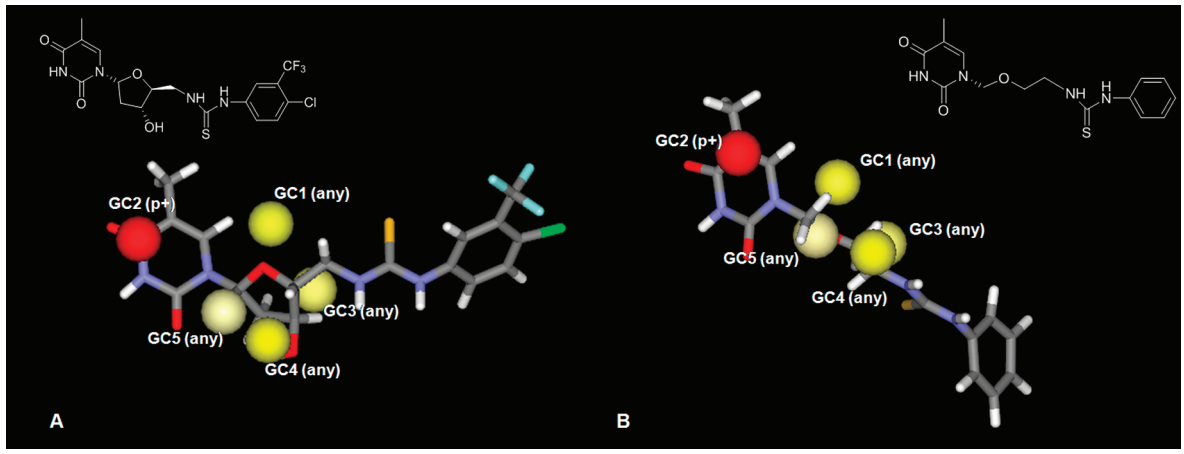



**Figure 2.** Graphical representation of the predicted bioactive conformations for the most active compound ATT**14** (A) and the most inactive compound ATT**30** (B), using eq 1 (Accelrys DS Visualizer).[28] Inhibition-enhancing and inhibition diminishing grid cells are shown, respectively, as yellow and red spheres. The bioactive conformations are presented as stick models. Carbon atoms are in gray, oxygen in red, nitrogen in blue, sulfur in yellow, chlorine in green, fluorine in cyan, and hydrogen atoms are in white.

**Table 8.** GCOD Coordinates and Normalized Occupancy Values and the Calculated Value of ClogP, for Each Test Set Inhibitor

| test set | GC1 (−1, 5, −1) | GC2 (2, 1, 1) | GC3 (−1, 5, −4) | GC4 (2, 3, −5) | GC5 (1, 2, −3) | $(ClogP)^2$ |
|---|---|---|---|---|---|---|
| **ATT10** | 0.137 | 0.000 | 0.136 | 0.027 | 0.496 | 3.312 |
| **ATT16** | 0.303 | 0.000 | 0.030 | 0.000 | 0.489 | 1.440 |
| **ATT26** | 0.089 | 0.036 | 0.125 | 0.040 | 0.327 | 1.563 |
| **ATT4** | 0.068 | 0.004 | 0.117 | 0.089 | 0.448 | 0.608 |

**Table 9.** Test Set Predictions

| test set | $pK_i$ obs | $pK_i$ pred | $\Delta pK_i$ residuals |
|---|---|---|---|
| **ATT10** | 5.49 | 5.26 | 0.23 |
| **ATT16** | 4.82 | 5.22 | −0.39 |
| **ATT26** | 5.72 | 4.48 | **1.23** |
| **ATT4** | 4.33 | 5.18 | −0.84 |

**Table 10.** Statistical Measures, Number of GCODs, and Number of Outliers for the Top Ten (*RI*) 4D-QSAR Models from Alignment 7 Using Only the "Most Potent" Training Set ($N = 25$)

| model | no. GGODS | $r^2$ | $q^2$ | LSE | no. outliers |
|---|---|---|---|---|---|
| 1 | 7 | 0.92 | 0.80 | 0.03 | 1 |
| 2 | 8 | 0.93 | 0.79 | 0.03 | 0 |
| 3 | 8 | 0.92 | 0.78 | 0.04 | 1 |
| 4 | 7 | 0.91 | 0.77 | 0.05 | 1 |
| 5 | 7 | 0.92 | 0.76 | 0.03 | 2 |
| 6 | 6 | 0.89 | 0.75 | 0.03 | 1 |
| 7 | 7 | 0.89 | 0.75 | 0.03 | 1 |
| 8 | 6 | 0.90 | 0.75 | 0.03 | 1 |
| **9** | **6** | **0.86** | **0.75** | **0.03** | **0** |
| 10 | 7 | 0.90 | 0.73 | 0.03 | 1 |

types of interactions between the inhibitors regarding the surrounding amino acid residues lining the binding site. This docking procedure can be visualized in Figure 4 for the most potent inhibitor of the training set, **ATT14**, using the predicted bioactive conformation from eq 1 (A) and eq 2 (B).

## DISCUSSION

The model 4 (eq 1) indicates that $pK_i$ will increase with the increasing inhibitor atom occupancy of grid cells of GC1, GC3, GC4, and GC5 by the appropriate IPE types. GC2 is the only descriptor in eq 1 that decreases $pK_i$ when an increasing occupancy occurs by polar positive charge (P+) IPE type. GC4 can be responsible for the largest increase in $pK_i$ among the descriptors of eq 1 (based upon its regression coefficient = 14.11). This descriptor is located near the 3′-hydroxyl group of ribose ring (1.9 Å). The IPE atom type of GC4 is Any, which means occupancy by any type of atom sampled across the training set increases $pK_i$. In the crystal complex with the substrate dTMP (1G3U) the 3′-hydroxyl group of dTMP plays a unique role in catalysis, adopting a C2′-endo conformation in the active site and participating in $Mg^{2+}$ stabilization through interactions involving a water molecule (W1009), Asp163, and Asp9.[7,30] The bioactive conformation of compound **ATT14** based on eq 1 (Figure 4A) leads to a binding mode where the inhibitor may be participating in interactions involving residues Met66, Tyr165, and Leu52 of the active site. These interactions are important because they promote the destabilization of the interaction with $Mg^{2+}$ and suspend the catalytic process of the enzyme. The nonspecificity in IPE type for GC4 might be explained through the location of GC4 in the active site. GC4 can establish a hydrogen bond with a water molecule (W1002) held by the side chain of Met66 or have hydrophobic interactions considering other amino acid residues of the active site. Furthermore, GC4 could be a hydrogen bond donor or acceptor site. Thus, these multiple types of interactions all leading to the suspension of the catalytic process may be the source of the selection of an IPE type of Any to cover all these possibilities. Overall, molecular modification in this region is a good target strategy to develop better antituberculosis agents, that is, more potent inhibitors of TMPKmt. Examples of such modifications could be the replacement of the 3′-hydroxyl group by 3′-amino or

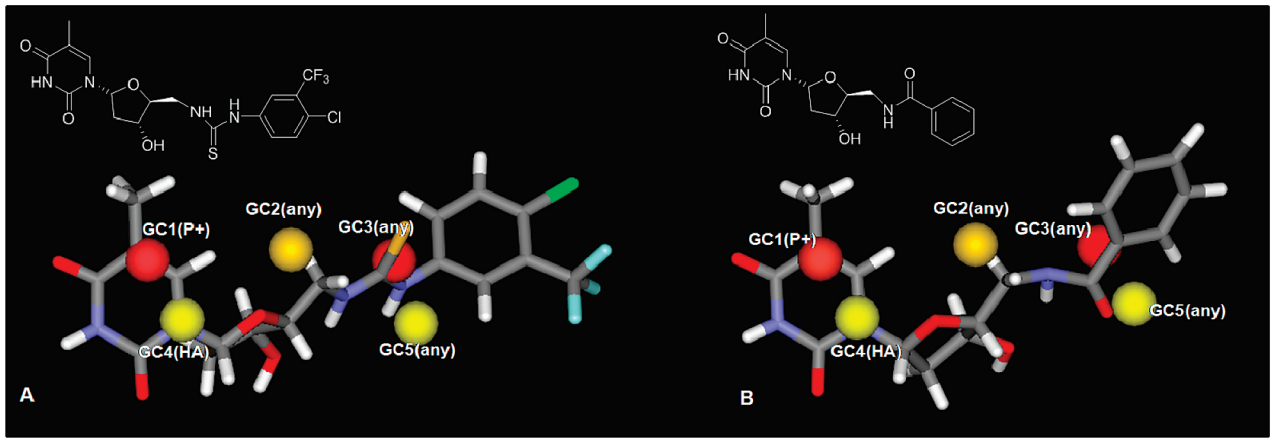



**Figure 3.** Graphical representation of the predicted active conformations for the most potent compound **ATT14** (A) and least potent compound **ATT29** (B) using model 9 (eq 2) (Accelrys DS Visualizer).[28] Inhibition-enhancing and inhibition diminishing grid cells are shown, respectively, as yellow and red spheres. The bioactive conformations are presented as stick models. Carbon atoms are in gray, oxygen in red, nitrogen in blue, sulfur in yellow, chlorine in green, fluorine in cyan, and hydrogen atoms are in white.

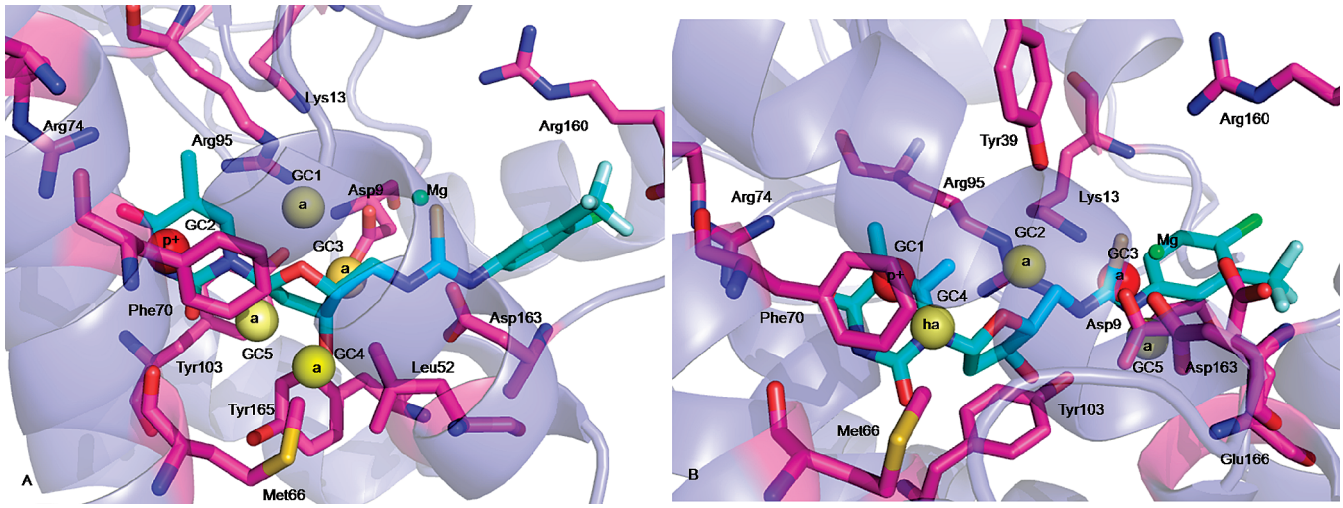



**Figure 4.** The postulated (*RI*) 4D-QSAR bioactive conformations of the most potent inhibitor (**ATT14**) from eq 1 (A) and eq 2 (B), respectively, and its respective GCODs from each 4D-QSAR model, docked in the active site of the crystal structure of TMPKmt (Pymol Viewer).[29] Only the main interacting residues in the pocket of the binding site are shown as a stick model. The carbon atoms are in magenta, nitrogen in blue, oxygen in red, and sulfur in yellow. The inhibitor **ATT14** is presented in a stick model, but the carbon atoms are in cyan. The GCODs are represented as spheres of 1 Å radius. (A) The GC2 descriptor has IPE type = P+. The GC1, GC3, GC4, and GC5 descriptors correspond to IPE type = Any. (B) The GC1 descriptor has IPE type = P+. The GC4 descriptor corresponds to the IPE type = HA. The GC2, GC3, and GC5 descriptors have IPE = Any.

3′-methyl groups in order to verify the importance of the hydroxyl group, and an isobutyl, or NH-alkyl group in order to explore the extent of steric tolerance at this position.

The only descriptor in eq 1 whose increasing occupancy by P+ can decrease p$K_i$ (GC2) is approximately 2.6 Å from the O4 of the pyrimidine ring. The main O4 interaction in the binding site is one of two possible hydrogen bonds, or a combination of them, with the Arg74 residue. Arginine is a polar positively charged amino acid. Thus, occupancy of GC2 by polar positively charged inhibitor atoms is detrimental to binding potency owing to electrostatic polar positive repulsions between the GC2 P+ site and Arg74.

Since the GCODs are normalized descriptors all having values between 0 and 1, the descriptor terms of eq 1 are predicted to increase the p$K_i$ in the order GC1 > GC3 > GC5 based upon their regression coefficients. The GC1 and GC5 descriptors from eq 1 are slightly cross-correlated (see Table 7). Although they are relatively distantly separated in space (4.1 Å), they possibly define a single pharmacophore binding site (see Figure 4). Both descriptors are located between the pyrimidine and sugar rings. These two GCODs reveal the pharmacophore sites embedded in the sugar-pyrimidine ring structure. The amino acid residues that could be participating in the ligand−receptor interaction identified by GC1 are Arg95 and Tyr103. Thus, GC1 can reflect both hydrogen bond and aromatic ring stacking interactions. These possible multiple types of interactions involving GC1 could be the reason for the nature of its IPE (Any). The amino acid residues that could interact with GC5 are Met66 and Phe70. The IPE atom type of GC5 is also Any, which might be explained by the location of GC5 in the active site. GC5 can establish a hydrogen bond involving a water molecule held by Met66 or have hydrophobic interactions with other amino acid residues of the active site. Furthermore, the side chain of Phe70 is located adjacent to the pyrimidine ring of compound **ATT14**, allowing its participation in an aromatic ring stacking interaction. The distance between the two aromatic rings is the same for the ligands of the crystallized complex (1G3U) and the predicted bioactive conformation of the most active compound **ATT14** as determined from

eq 1 and then docked into the active site based upon alignment 7.

The GC3 descriptor in eq 1 is positioned between the C5′of the sugar ring (1.5 Å) and the nitrogen of thiourea moiety (2.2 Å). The amino acids residues that possibly interact with inhibitor atoms that are captured by this GCOD descriptor include Asp163, Asp9, and Arg95. The IPE type of GC3 is also Any. This nonspecific IPE type for GC3 may be explained considering the participation of atoms across the training set acting as both hydrogen bond acceptors and donors with the amino acid residues Asp163, Asp9, and Arg95 of the active site.

As already mentioned, some descriptor terms from eq 2 are equivalent to descriptor terms from eq 1. The only descriptors of the "most active" model 9 (eq 2) that did not have any direct corresponding GCODs to the model 4 (eq 1) were GC3 (Any) and GC5 (Any). A comparison of the amino acid residues, which define the binding pocket (see Figure 4) to the GCODs obtained for the "most potent" set of inhibitors (eq 2), provides the identification of interactions between substituents of the inhibitors occupying GC5 (Any) and polar negatively charged residues as Asp9 and Glu166 or polar positively charged residues as Arg95 (see Figure 4B). Thus, GC5 can be both a hydrogen bond donor or acceptor pharmacophore site depending on the inhibitor. This might explain the use of an Any IPE type for this descriptor. In addition, repulsive steric and/or electrostatic interactions from inhibitor substituents in the region corresponding to the location of GC3 (Any) (eq 2) involving polar positively charged residues as the side chains of Arg95 and Arg160, or polar negative charged residues as Glu166 (see Figure 4B), may be a source for the lost in inhibitory potency.

It is noteworthy that in both, the best full training set and the best "most potent" (*RI*) 4D-QSAR models, p$K_i$ has a favorable contribution when the lipophilicity increases. This finding is consistent with previously reported results,[13] showing that α-thymidine analogues bind in the active site 'upside down' as compared to the natural substrate. The tails [5′-arylthiorea moieties] of these molecules are oriented to the outside of the enzyme through a channel, which is surrounded by nonpolar and aromatic residues including Ala35, Phe36, Pro37, and Arg160. The presence of lipophilic substituents on the 5′-aryl moiety is an important feature of the α-thymidine derivatives possibly representing an additional pharmacophore site responsible for the higher inhibition potency of these derivatives.

The GCODs of the models developed in this study suggest novel 3D-pharmacophore sites across pseudoanalogs to the training set inhibitors that can be explored in the search for better antituberculosis agents having the TMPKmt as target. Such investigation can be done to optimize binding potency by focusing upon substituents that will occupy GCODs that increase potency. One such region is that of the sugar-pyrimidine ring structure, and another is the region around 5′-arylthiourea moiety. Alternatively, placing substituents in regions of the inhibitors where no GCODs are found in the current (*RI*) 4D-QSAR models can focus upon identifying new favorable inhibitor interaction sites regarding the receptor binding site. Regions of the inhibitors with no GCODs also are candidate sites for adding substituents to optimize requisite ADMET properties, presumably because those regions are not interacting with the receptor binding site. Thus, although this data set is relatively small and not too chemically diverse, it seems to provide a robust set of structural options useful for the design of new TMPKmt inhibitors. We are now in the process of expanding this investigation by doing a receptor-independent 4D-QSAR analysis of a larger and more chemically diverse set of TMPKmt inhibitors.

## ACKNOWLEDGMENT

C.H.A. is grateful to the CAPES foundation for scholarship support. This work was also funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant 1 R21 GM075775. Information on Novel Preclinical Tools for Predictive ADME-Toxicology can be found at http://grants.nih.gov/grants/guide/rfa-files/RFA-RM-04-023.html. Links to nine initiatives are found at http://nihroadmap.nih.gov/initiatives.asp. Resources of the Laboratory of Molecular Modeling and Design at UNM and The Chem21 Group, Inc. were used in performing this work.

## REFERENCES AND NOTES

(1) Dye, C.; Floyd, K.; Uplekar, M. Global tuberculosis control: surveillance, planning, financing. WHO Report 2008; World Health Organization Document, 2008. WHO/HTM/TB/2008.393.
(2) Vilchèze, C.; Weissbrod, T. R.; Chen, B.; Kremer, L.; Hazbón, M. H.; Wang, F.; Alland, D.; Sacchettini, J. C.; Jacobs, W. R., Jr. Altered NADH/NAD+ ratio mediates coresistance to isoniazid and ethionamide in mycobacteria. *Antimicrob. Agents Chemother.* **2005**, *49*, 708–720.
(3) Dye, C. Global epidemiology of tuberculosis. *Lancet* **2006**, *367*, 938–940.
(4) Munier-Lehmann, H.; Chafotte, A.; Pochet, S.; Labesse, G. Thymidylate kinase of *Mycobacterium tuberculosis*: a chimera sharing properties common to eukaryotic and bacterial enzymes. *Protein Sci.* **2001**, *10*, 1195–1205.
(5) Ostermann, N.; Schlichting, I.; Brundiers, R.; Konrad, M.; Reinstein, J.; Veit, T.; Goody, R. S.; Lavie, A. Insights into the phosphoryltransfer mechanism of human thymidylate kinase gained from crystal structures of enzyme complexes along the reaction coordinate. *Structure* **2000**, *8*, 629–642.
(6) Lavie, A.; Schlichting, I.; Vetter, I. R.; Konrad, M.; Reinstein, J.; Goodey, R. S. The bottleneck in AZT activation. *Nat. Med.* **1997**, *3*, 922–924.
(7) Li de la Sierra, I.; Munier-Lehmann, H.; Gilles, A. M.; Bârzu, O.; Delarue, M. X-ray structure of TMP kinase from *Mycobacterium tuberculosis* complexed with TMP at 1.95 Å resolution. *J. Mol. Biol.* **2001**, *311*, 87–100.
(8) Vanheusden, V.; Munier-Lehmann, H.; Froeyen, M.; Busson, R.; Rozenski, J.; Herdewijn, P.; Van Calenbergh, S. Discovery of bicyclic thymidine analogues as selective and high affinity inhibitors of *Mycobacterium tuberculosis* thymidine monophosphate kinase. *J. Med. Chem.* **2004**, *47*, 6187–6194.
(9) Vanheusden, V.; Munier-Lehmann, H.; Pochet, S.; Herdewijn, P.; Van Calenbergh, S. Synthesis and evaluation of thymidine-5′-O-monophosphate analogues as inhibitors of *Mycobacterium tuberculosis* thymidylate kinase. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2695–2698.
(10) Vanheusden, V.; Van Rompaey, P.; Munier-Lehmann, H.; Pochet, S.; Herdewijn, P.; Van Calenbergh, S. Thymidine and Thymidine-5′-O-monophosphate analogues as Inhibitors of *Mycobacterium tuberculosis* Thymidylate Kinase. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3045–3048.
(11) Haouz, A.; Vanheusden, V.; Munier-Lehmann, H.; Froeyen, M.; Herdewijn, P.; Van Calenbergh, S.; Delarue, M. J. Enzymatic and structural analysis of inhibitors designed against *M. tuberculosis* thymidylate kinase: new insights into the phosphoryl transfer mechanism. *Biol. Chem.* **2003**, *278*, 4963–4971.
(12) Pochet, S.; Dugue, L.; Labesse, G.; Delepierre, M.; Munier-Lehmann, H. Comparative study of purine and pyrimidine nucleoside analogues acting on the thymidylate kinases of *Mycobacterium tuberculosis* and of humans. *ChemBioChem* **2003**, *4*, 742–747.
(13) Van Daele, I.; Munier-Lehmann, H.; Froeyen, M.; Balzarini, J.; Van Calenbergh, S. Rational design of 5′-thiourea-substituted α-thymidine analogues as thymidine monophosphate kinase inhibitors capable of inhibiting mycobacterial growth. *J. Med. Chem.* **2007**, *50*, 5281–5292.

(14) Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J. Am. Chem. Soc.* **1997**, *119*, 10509–10524.
(15) Blondin, C.; Serina, L.; Wiesmü ller, L.; Gilles, A. M.; Bârzu, O. Improved spectrophotometric assay of nucleoside monophosphate kinase activity using pyruvate kinase/lactate dehydrogenase coupling system. *Anal. Biochem.* **1994**, *220*, 219–222.
(16) *HyperChem Program Release 7.05 for Windows*; Hybercube, Inc.: Gainesville, FL, 2005.
(17) Dewar, M. J. S. E.; Zoebisch, G.; Healy, E. F.; Stewart, J. J. P. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
(18) Doherty, D. C. *MOLSIM Package version 3.2*; The Chem21 Group, Inc.: Lake Forest, IL, 2001.
(19) *4D-QSAR Package version 2.0*; The Chem21 Group Inc.: Lake Forest, IL, 1997.
(20) Glen, W. G.; Dunn, W. J.; Scott, D. R. Principal components analysis and partial least-squares regression. *Tetrahedron Comput. Methodol.* **1989**, *2*, 349–354.
(21) Rogers, D. G.; Hopfinger, A. J. Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854–866.
(22) Dunn, W. J., III; Rogers D. Genetic Partial Least Squares in QSAR. In *Genetic Algorithms in Molecular Modeling*; Devillers, J., Ed.; Academic: London, 1996; pp 109−130.
(23) Friedman, J. H. *Multivariate adaptative regression splines*; Technical Report No. 102; Laboratory for Computational Statistics, Departament of Statistics, Stanford University: Stanford, CA, 1988.
(24) Rogers, D. WOLF Genetic Function Approximation. *Reference Manual, version 5.5*; Molecular Simulation Inc.: Burlington, MA, 1994.
(25) Rogers, D. G/SPLINES: A hybrid of Friedman's multivariate adaptative regression splines (MARS) algorithm with Holland's genetic algorithm. In *The Proceedings of the Fourth International Conference on Genetic Algorithms*; Belew, R. K., Booker, L. B., Eds.; Morgan Kaufmann Publishers: San Francisco, 1991; pp 38−46.
(26) Mannhold, R.; van de Waterbeemd, H. Substructure and whole molecule approaches for calculating log P. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 337–354.
(27) Ghose, A. K.; Pritchett, A.; Crippen, G. M. Atomic physicochemical parameters for three-dimensional structure directed quantitative structure-activity relationships III: Modeling hydrophobic interactions. *J. Comput. Chem.* **1988**, *9*, 80–90.
(28) *Discovery Studio Visualizer version 2.0*; Accelrys Software Inc.: San Diego, CA, 2007.
(29) DeLano, W. L. *The Pymol Molecular Graphics System version 1.0*; Delano Scientific LLC: Palo Alto, CA, 2004.
(30) Van Calenbergh, S. Structure-aided design of inhibitors of *Mycobacterium tuberculosis* thymidylate kinase. *Verh K Acad. Geneeskd Belg.* **2006**, *68*, 223–48.

CI8004622