J Comput Aided Mol Des (2011) 25:31–49
DOI 10.1007/s10822-010-9399-4

Computer-assisted combinatorial design of bicyclic thymidine analogs as inhibitors of *Mycobacterium tuberculosis* thymidine monophosphate kinase

Vladimir Frecer · Pierfausto Seneci · Stanislav Miertus

Received: 17 September 2010 / Accepted: 28 October 2010 / Published online: 17 November 2010

Abstract Thymidine monophosphate kinase ($TMPK_{mt}$) is an essential enzyme for nucleotide metabolism in *Mycobacterium tuberculosis*, and thus an attractive target for novel antituberculosis agents. In this work, we have explored the chemical space around the 2′,3′-bicyclic thymidine nucleus by designing and *in silico* screening of a virtual focused library selected via structure based methods to identify more potent analogs endowed with favorable ADME-related properties. In all the library members we have exchanged the ribose ring of the template with a cyclopentane moiety that is less prone to enzymatic degradation. In addition, we have replaced the six-membered 2′,3′-ring by a number of five-membered and six-membered heterocyclic rings containing alternative proton donor and acceptor groups, to exploit the interaction with the carboxylate groups of Asp9 and Asp163 as well as with several cationic residues present in the vicinity of the $TMPK_{mt}$ binding site. The three-dimensional structure of the $TMPK_{mt}$ complexed with 5-hydroxymethyl-dUMP, an analog of dTMP, was employed to develop a QSAR model, to parameterize a scoring function specific for the $TMPK_{mt}$ target and to select analogues which display the highest predicted binding to the target. As a result, we identified a small highly focused combinatorial subset of bicyclic thymidine analogues as virtual hits that are predicted to inhibit the mycobacterial TMPK in the submicromolar concentration range and to display favorable ADME-related properties.

Keywords *Mycobacterium tuberculosis* · TMPK inhibitors · Combinatorial library design · QSAR model · *In silico* screening · Gibbs free energy of binding · ADME properties

Abbreviations

ADME	Adsorption, distribution, metabolism and excretion
AZTMP	3′-azido-2′-deoxythymidine monophosphate
CFF91	Consistent class II force field
dTTP	Deoxythymidine triphosphate
dUMP	Deoxyuridine monophosphate
HB	Hydrogen bond
TEM	5′-CH_2OH dUMP inhibitor
$TMPK_h$	Human thymidine monophosphate kinase
$TMPK_{mt}$	*Mycobacterium tuberculosis* thymidine monophosphate kinase

V. Frecer · P. Seneci · S. Miertus (✉)
International Centre for Science and High Technology, UNIDO, AREA Science Park, Padriciano 99, 34012 Trieste, Italy
e-mail: stanislav.miertus@ics.trieste.it

V. Frecer
Cancer Research Institute, Slovak Academy of Sciences, 83391 Bratislava, Slovakia

V. Frecer
Department of Physical Chemistry of Drugs, Faculty of Pharmacy, Comenius University, 83232 Bratislava, Slovakia

P. Seneci
Department of Organic and Industrial Chemistry and CISI, University of Milan, 20133 Milan, Italy

Introduction

More than 2 billion people, equal to one-third of the world's population, are currently infected with *Mycobacterium tuberculosis*, the causative agent of tuberculosis

(TB). TB is an airborne infectious disease leading annually to more than 9 million of new cases and claiming almost 2 million of lives each year [1, 2]. Mycobacterial infection can remain dormant for decades and become active later, typically after the immune system of an infected individual becomes compromised by factors such as aging, viral infections, alcohol and drug abuse. The World Health Organization (WHO) reports that in Africa over 80% of TB patients are HIV infected [1]. HIV infection is a potent risk factor for converting latent TB into active transmissible TB while mycobacteria accelerate the progress of AIDS infection, making thus the combination of both diseases more destructive than either alone.

Currently active TB is treated with first-line drugs introduced to clinical practice more than 40 years ago (isoniazid, rifampicin, pyrazinamide, and ethambutol) and requires treatment with a duration of at least six months. As a consequence of the prolonged regimen, low patient's compliance, elevated costs, and highly adaptive nature of *M. tuberculosis* to its surroundings, multidrug-resistant (MDR) and extensively drug-resistant (XDR) strains of the causative agent have emerged. According to WHO, 50–80% of treated TB cases develop drug-resistance [3]. Extended treatment for up to two years is needed for M/XDR TB cases, using expensive injectable second-line drugs which unfortunately cause much more severe side effects [4]. Consequently, TB has re-emerged as a serious public health threat worldwide and currently represents one of the most common causes of death in less-developed high-burden countries. The global emergency posed by the M/XDR TB has stimulated the search for new targets and discovery of innovative more effective antibiotic drugs acting on novel targets.

Determination of the complete genome sequence of *M. tuberculosis* strain H37Rv allowed identification of new potential antimycobacterial drug targets [5]. Regulatory proteins, enzymes involved in the biosynthesis of essential amino acids, enzymes for cell wall biosynthesis, and DNA metabolism are considered as crucial drug targets for anti-TB chemotherapy [6, 7]. *M. tuberculosis* thymidine monophosphate kinase (EC2.7.4.9, $TMPK_{mt}$) has emerged as an attractive target for the design of a novel class of anti-TB agents [8]. $TMPK_{mt}$ catalyses the phosphorylation of deoxythymidine monophosphate (dTMP, Fig. 1) to deoxythymidine diphosphate (dTDP) utilizing ATP as its preferred phosphoryl donor. This step lies at the junction of the de novo and salvage pathways of deoxythymidine triphosphate (dTTP) metabolism and is the last specific enzyme for its synthesis [9]. $TMPK_{mt}$ was shown to be essential in providing the bacteria with dTTP and for DNA synthesis [10]. Low sequence identity (22%) of $TMPK_{mt}$ with its human isozyme ($TMPK_h$) and identification of 3′-azido-2′-deoxythymidine monophosphate (AZTMP, $K_i = 10$ μM, Fig. 1) as a selective competitive inhibitor made $TMPK_{mt}$ an attractive target for blocking mycobacterial DNA synthesis [8, 11].

The three-dimensional structure of $TMPK_{mt}$ bound to dTMP was solved, thus allowing to perform structure-based drug design studies [11, 12]. $TMPK_{mt}$ is a homodimer with 214 amino acids per monomer and a global folding similar to that of other TMPKs [11]. The active site of $TMPK_{mt}$ complexed with dTMP is in a fully closed conformation, with the ATP binding site being already preformed and the LID region (a highly flexible stretch of residues covering the ATP binding site) well ordered into a α-helical conformation [11, 12]. The active site contains an Mg^{2+} ion and the Tyr39 residue, which interact with two opposite non-bridging oxygens of the phosphate moiety of dTMP [11, 12, 15]. Closer examination of the catalytic site revealed that the main interactions between dTMP and the enzyme include a stacking interaction between the pyrimidine ring of thymine and Phe70, as well as a number of hydrogen bonds (HBs). A HB is formed between the 4-carbonyl of dTMP and the side chain of Arg74, another between Asn100 and the 3-NH of the pyrimidine ring, and another between the 3′-OH of deoxyribose and the terminal carboxyl of Asp9, which also interacts with the Mg^{2+} ion that is responsible for positioning the phosphate oxygens of dTMP. Other HBs and ionic interactions exist between the 5′-O-phosphoryl group and Tyr39, Phe36, Arg95 and Mg^{2+}, respectively [11, 12, 15].

Inspection of the enzyme-substrate interactions suggested that substrate-based competitive inhibitors could be

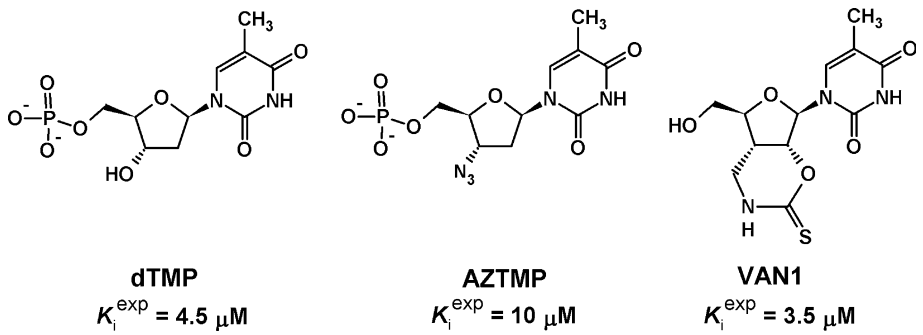

Fig. 1 Chemical structures of the substrate and of some competitive inhibitors of $TMPK_{mt}$ [8, 13, 14]

designed by replacement of the 3′-OH and 2′-H of deoxyribose, substitutions on the 2- and 5-position of the pyrimidine ring, as well as by taking advantage of several cavities occupied by crystallographic water [15]. Single or multiple chemical modifications have been introduced into the pyrimidine moiety and the sugar part of dTMP, leading to micromolar inhibitors (K_i) endowed with a high selectivity index for $TMPK_{mt}$ [13–23]. A $TMPK_{mt}$ inhibitor derived from AZTMP and bearing a bicyclic thymidine analogue with a six-membered ring fused to the C2′-C3′ ribose bond (VAN1, $K_i = 3.5$ μM, Fig. 1), showed selective activity against *M. tuberculosis* [14]. Despite of its potent inhibition of $TMPK_{mt}$, the compound failed to suppress the growth of the *M. tuberculosis* H37Rv strain [20]. As the cell wall of mycobacterium is covered by a waxy coating composed primarily of mycolic acids, this makes the penetration of polar anti-TB drugs such as VAN1 through this barrier rather difficult. Van Daele et al. [20] have synthesized a series of bicyclic thymidine derivatives of VAN1 and identified a more potent 5′-deoxy analog with a K_i value of 2.3 μM. Such a less polar compound displayed also in vitro inhibition of bacterial growth against *Mycobacterium bovis* BCG, with an inhibitory effect causing 99% growth reduction (IC_{99}) at 100 μg mL^{-1}, confirming thus the potential of bicyclic dTMP analogs for further design of $TMPK_{mt}$ inhibitors [20].

In this work, we have virtually explored in an extensive manner the chemical space around VAN1 by designing and *in silico* screening a virtual library of 2′,3′-bicyclic thymidine analogues using computer-assisted combinatorial techniques [24–28]. Our goal was to select more potent $TMPK_{mt}$ inhibitors endowed with favorable ADME-related properties, and to prioritize them for future synthesis. As to library design, we exchanged the ribose ring of VAN1 with a cyclopentane moiety that is less prone to enzymatic degradation. In addition, we replaced the 2′,3′-fused six-membered ring of VAN1 by a number of five-membered and six-membered heterocyclic rings containing proton donor and acceptor groups able to interact with the carboxylate group of the key Asp9 and Asp163 binding site residues of $TMPK_{mt}$ [11]. Moreover, 5 and 5′ substitutions on the dTMP-like scaffold were made by methyl and phosphate group isosteres that are likely to enhance the interaction with the active site of $TMPK_{mt}$. The three-dimensional structure of $TMPK_{mt}$ complexed with 5-hydroxymethyl-dUMP [15], an analog of dTMP, was employed to develop a QSAR model, to parameterize a target-specific scoring function for $TMPK_{mt}$ and to select VAN1 analogues which display the highest predicted binding to the target. We identified a small, highly focused combinatorial subset of bicyclic thymidine analogues, which contains virtual hits predicted to inhibit the mycobacterial $TMPK_{mt}$ with sub-micromolar potency and to display comparable or better ADME-related properties than the 5′-deoxy analogue of VAN1.

Results and discussion

Library design

Vanheusden reported that when VAN1 (Fig. 1) was modeled in the active site of $TMPK_{mt}$ [11, 12] its 2′,3′-fused six-membered ring directed the thiourethane N atom to form a HB with the Asp9 active site residue, while its S atom established hydrophobic interactions with Tyr103 and Tyr165 residues [14]. Our modeling results observed an HB between the thiourethane N and Asp9, but showed that the S atom is surrounded by both hydrophobic (Tyr103, Tyr165 and Leu171) and polar residues (Gln172 and Glu166). The latter residues can form weaker HB interactions with the thiouretane S atom (Fig. 2). Replacement of the thiourethane moiety in the six-membered ring of VAN1 by other polar function groups could thus enhance inhibitor-enzyme interactions which involve Gln172, Glu166 and potentially even the adjacent Arg153 residue.

Enzyme inhibition data from Munier-Lehmann et al. [8] and Haouz et al. [15] revealed the tolerance of the $TMPK_{mt}$ binding pockets for halogen atoms and the hydroxyl group on the 5-methyl group of the thymine ring. Van Daele et al. [21] designed potent 5′-arylthiourea substituted thymidine analogs and showed that $TMPK_{mt}$ well tolerates larger polar groups in the 5′ position. In addition, Vanheusden

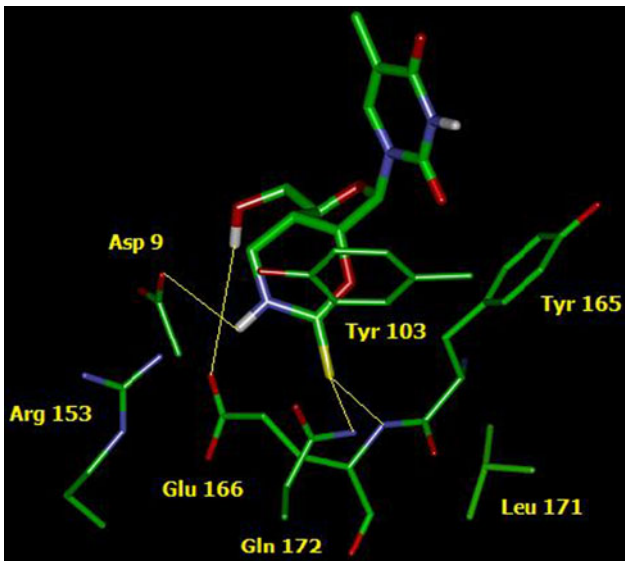

Fig. 2 VAN1 [14] (Fig. 1) in the active site of $TMPK_{mt}$. Residues surrounding, and interacting with the 2′,3′-fused six-membered ring of the inhibitor are shown. Hydrogen bonds are indicated as *sallow lines*

et al. [14] showed that even much larger molecules than dTMP, such as thymidine dinucleosides, can be potent competitive inhibitors of $TMPK_{mt}$. Based on these observations, we designed a virtual library of bicyclic thymidine analogues targeted around the VAN1 template inhibitor (Fig. 1). We considered a large selection of unsaturated heterocyclic five-membered and six-membered 2′,3′-fused rings, a variety of charged and neutral phosphate group isosteres and some methyl group replacements on the thymine 5-position. It must be mentioned that a small number of similar five-membered and six-membered bicyclic sugar nucleosides was previously described by Vanheusden et al., [14] Van Daele et al. [20] and Kifli et al. [30].

We first decided to replace the ribose ring of VAN1 with a cyclopentane moiety that is less prone to enzymatic degradation via glycosidic bond cleavage by human thymidine phosphorylase [29]. A set of reagents (fragments) shown on Fig. 3 was then selected to make the initial diversity library made by: 6 (R_1) × 29 ($R_{2,3}$) × 12 (R_4) = 2088 analogs. The size, composition and property ranges of selected $R_{2,3}$-groups and R_4-groups with respect to the structure of the template inhibitor VAN1 were somewhat enlarged so as to introduce more diverse building blocks, explore a wider subset of the chemical space and propose novel bicyclic thymidine analogs. This is in accordance with the elevated tolerance of $TMPK_{mt}$ for larger residues in these positions, and with the relatively strict specificity of the pocket containing the R_1-groups [11, 12].

The initial virtual library was generated by attaching the R-groups onto the dTMP-like scaffold (Fig. 3) using the CombiChem module of the Cerius^2 program [31]. Retaining full diversity of the initial library, all of the 2088 generated analogs were submitted to a structure-based focusing step.

$TMPK_{mt}$ target and structure-based library focusing

The 3D model of the closed form of $TMPK_{mt}$ was prepared by modification of the crystal structure of the complex comprising 5′-CH2OH dUMP at the binding site of $TMPK_{mt}$, [15] by modeling the thiourethane ring of the VAN1 template fused to the C2′-C3′ bond of the ribose (TEM, Fig. 4) [14, 20]. The $TMPK_{mt}$ binding site contains a magnesium ion which coordinates to and neutralizes the 5′-O-phosphate of the substrate, and four crystallographic water molecules located in the vicinity of the Mg^{2+}, which were kept in our $TMPK_{mt}$ model. The molecular structure of the complex was carefully refined by molecular mechanics energy-minimization (see Computational Methods section for details). In the complex, the TEM inhibitor is almost completely enclosed in the dTMP binding site by the α-helical LID region. The refined 3D structure of the complex was used in the structure-based evaluation of the library.

To select a small, highly focused combinatorial subset of the initial virtual library of VAN1 analogs that display high predicted binding affinities to $TMPK_{mt}$ binding site, structure-based evaluation and *in silico* screening procedures have been applied. Each analog was docked into the binding pocket of the $TMPK_{mt}$ model represented by a binding site grid (Fig. 5) by employing a Monte Carlo procedure to generate ligand conformers, and a semi-flexible ligand fitting algorithm [32] of the LigFit module of Cerius^2 for the docking [31]. Ten best binding conformers per analog yielded by the docking procedure were clustered into five conformational families depending on their r.m.s. deviations [33]. In each cluster, the conformer with the highest docking score was selected for subsequent virtual screening.

QSAR model and target-specific scoring function

Target-specific scoring functions exhibit better predictivities on average than their generic counterparts [34]. To train a scoring function specific for the $TMPK_{mt}$ target, we have correlated the scores implemented in the Cerius^2 program [31] with the experimental activities (K_i^{exp}) of a training set of 18 thymidine $TMPK_{mt}$ inhibitors T1–T18 (Table 1) prepared by Vanheusden et al. [13, 14, 17, 18] and Van Daele et al. [20]. The compounds of the training set were docked into the binding site of the $TMPK_{mt}$ model using the LigFit algorithm [32] of Cerius^2 [31]. Comparison of the docked analogs with the bound conformation of the template TEM at the active site of the target demonstrated that the docked poses of the training set inhibitors were similar to the binding mode of TEM (Fig. 6).

We have tested several QSAR models which relate the observed K_i^{exp} to the computed scores LigScore, LUDI, PMF, PLP1 and PLP2 [32, 35–41]. From the set of available empirical scoring functions, the PLP1 score led to the best fit of the experimental activities to the computed enzyme-inhibitor binding energies. The following QSAR equation was obtained by linear regression (Fig. 7):

$$pK_i = -\log_{10} K_i[TMPK_{mt}] = -6.98745 + 0.05110 \cdot PLP1 \quad (1)$$

(number of samples $n = 18$, correlation coefficient $R^2 = 0.88$, leave-one-out cross validated correlation coefficient $R^2_{xv} = 0.84$, standard error $\sigma = 0.21$, Fisher F-test = 116.8, statistical significance of the correlation $\alpha > 95\%$). The leave-one-out cross validated correlation coefficient R^2_{xv} of 0.84 indicates that the major portion of the variance of the training set data was well described by this QSAR model.

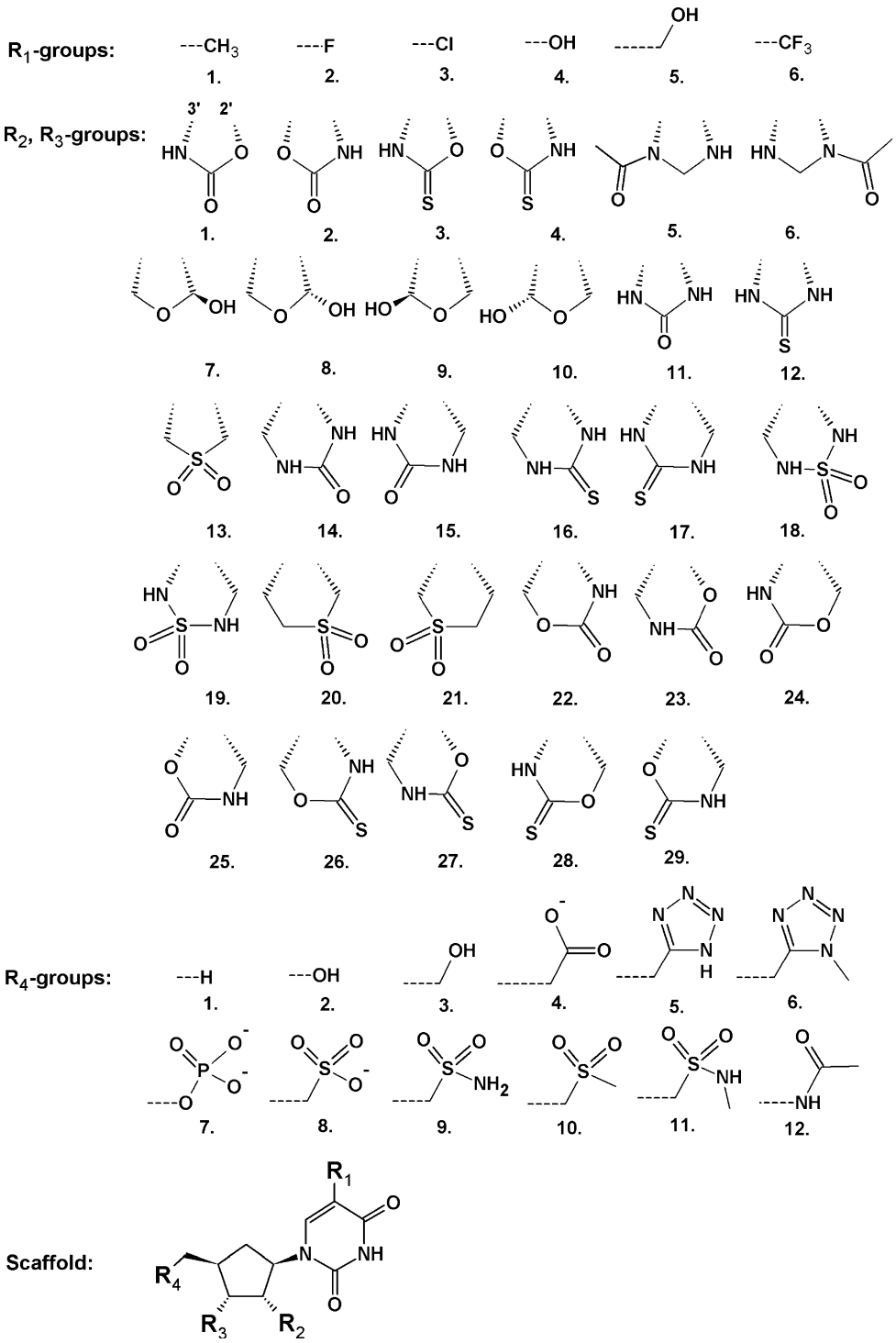

Fig. 3 R_1–R_4 groups used in the design of the initial diversity library of bicyclic thymidine analogues based on the shown scaffold

Empirical scoring functions, which estimate the free energy of binding of a ligand to a protein receptor with known 3D structure, tend to fall in general into two major classes emphasizing either hydrogen bonding and ionic interactions or van der Waals and hydrophobic interactions. As PLP1 belongs to the first class of functions, it was able to reproduce the binding of the training set of charged and polar thymidine analogs to the $TMPK_{mt}$ fairly successfully. The PLP1 score was previously shown to perform comparably or better than other more rigorous computationally intensive methods in identifying and ranking of tight-binding ligands within drug discovery programs [42–45].

The quality of the QSAR model was also confirmed by predicting the $TMPK_{mt}$ inhibition constants for a validation

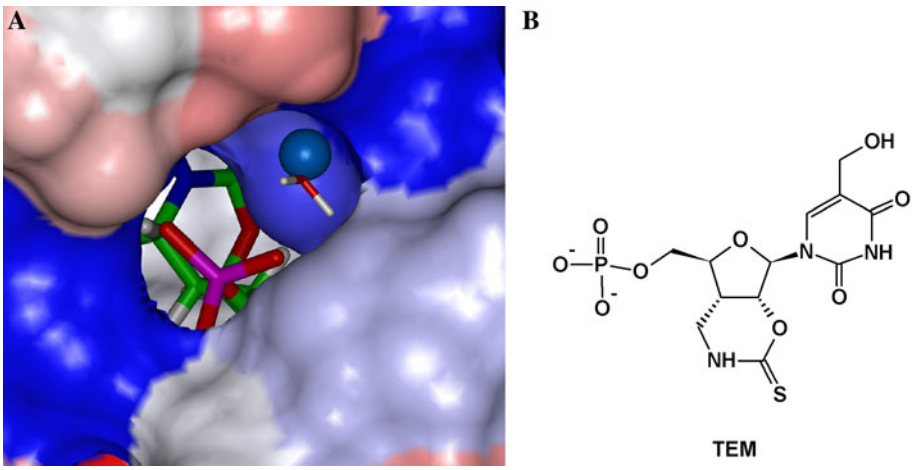

Fig. 4 **a** Molecular surface of $TMPK_{mt}$ showing the "mouth" of the dTMP binding site containing the 5′-CH_2OH dUMP inhibitor (TEM) [15]. Color coding of the surface: *blue*—cationic, *red*—acidic, and *white*—non-polar residues. *Blue* sphere—Mg^{2+}. **b** Chemical structure of TEM

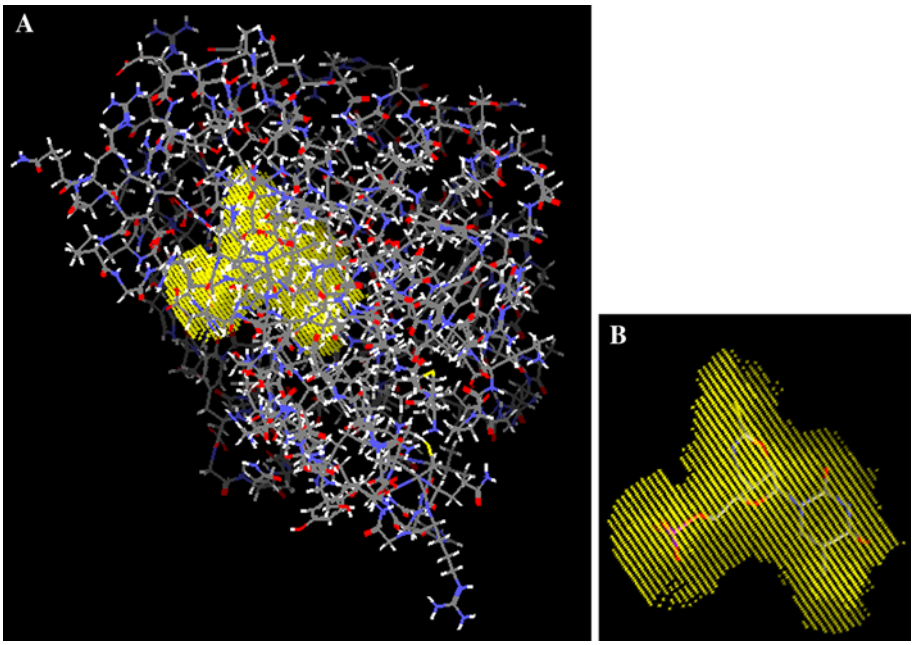

Fig. 5 **a** Binding site grid of $TMPK_{mt}$ with bound bicyclic inhibitor TEM (Fig. 4), which was used for the grid size and shape definition. **b** Grid close-up

set of five analogs V1–V5 not included into the training set (Table 1) [13, 14, 17, 18, 20]. The ratio between the predicted activities derived from Eq. (1) and observed activities pK_i^{pre}/pK_i^{exp}, close to 1, confirmed the predictive power of the QSAR model. The training and validation sets displayed somewhat limited variation in the R-groups space due to restricted availability of experimental data. Prediction of inhibitory potencies by the trained target-specific scoring function, which slightly exceed the activity ranges of the training set is, however, still possible. QSAR models using ligand-receptor binding affinity of docked analogs estimated via empirical scoring functions reflect closely the mode of the action for a given ligand. They are, thus, less sensitive to the size of the training set than QSAR models relying on simpler descriptors which are only distantly related to the receptor binding [24–28].

The parameterized scoring function (1), which is specific for $TMPK_{mt}$, was subsequently used for *in silico* screening and ranking of the best conformers docked to the active site of the drug target in the structure-based evaluation step.

In silico screening

Generated poses of the analogs were screened *in silico* by using the PLP1 score [40, 41]. Predicted $TMPK_{mt}$ inhibitory activities K_i^{pre} of the analogs were calculated from the target-specific scoring function (1). The analogs were then rank-ordered according to their K_i^{pre} (Table 2). The chemical structures of the 16 analogs displaying the highest predicted potencies in the submicromolar range (virtual hits) are shown on Fig. 8.

Table 1 Training and validation sets of thymidine inhibitors of $TMPK_{mt}$ [13, 14, 17, 18, 20] used in the QSAR model of the inhibitory effect and parameterization of target-specific scoring function

Training set	R_1[a]	R_2	R_3	R_4	Q_M[b] [è]	K_i^{exp}[c] [μM]
T1 AZTMP	$-CH_3$	–H	$-N_3$	$-OPO_3^{2-}$	−2	10
T2	$-CH_3$	–H	$-CH_2N_3$	$-OPO_3^{2-}$	–2	12
T3	$-CH_3$	–H	$-CH_2NH_2$	$-OPO_3^{2-}$	−2	10.5
T4	$-CH_3$	–H	$-CH_2F$	$-OPO_3^{2-}$	−2	15
T5	$-CH_3$	–H	$-CH_2OH$	$-OPO_3^{2-}$	−2	29
T6	$-CH_3$	–F	–OH	$-OPO_3^{2-}$	−2	43
T6	$-CH_3$	–H	$-N_3$	–OH	0	28
T7 AZT	$-CH_3$	–H	$-CH_2N_3$	–OH	0	40
T8	$-CH_3$	–H	$-NH_2$	–OH	0	230
T9	$-CH_3$	–F	$-CH_2N_3$	–OH	0	165
T10	$-CH_3$	–H	–OH	$-NHCOCH_3$	0	90
T11	$-CH_3$	–F	$-CH_2N_3$	–F	0	80
T12	–F	–OH	–OH	–OH	0	521
T13	–F	–OH	–OH	–H	0	560
T14	$-CH_3$	–H	$-N_3$	$-OPO_3^{2-}$	−2	10
T15 VAN1	$-CH_3$	2′-OC(S)NHCH$_2$-3′		–OH	0	3.5
T16	$-CH_3$	2′-OC(O)NHCH$_2$-3′		–OH	0	13.5
T17	$-CH_3$	2′-NHC(O)NHCH$_2$-3′		–OH	0	59
T18	$-CH_3$	2′-NHC(NH)NHCH$_2$-3′		–OH	0	46
Validation set	**R_1**	**R_2**	**R_3**	**R_4**	**Q_M [è]**	**pK_i^{pre}/pK_i^{exp}[d]**
V1	$-CH_3$	–OH	$-NH_2$	$-OPO_3^{2-}$	−2	1.14
V2	$-CH_3$	–H	$-CH_2F$	–OH	0	1.09
V3	$-CH_3$	–F	–OH	–OH	0	0.89
V4	$-CH_3$	2′-NHC(S)NHCH$_2$-3′		–OH	0	1.12
V5	$-CH_3$	2′-OC(S)NH-3′		–OH	0	1.22

[a] For the position of R_1–R_4-groups see Fig. 3

[b] Q_M—total molecular charge at neutral pH

[c] K_i^{exp}—experimental inhibition constants of $TMPK_{mt}$ [13, 14, 17, 18, 20]

[d] ratio of predicted pK_i^{pre} and experimental pK_i^{exp} inhibition constants ($pK_i^{pre} = -\log_{10}K_i^{pre}$) of the validation set. pK_i^{pre} were computed from the QSAR Eq. (1)

The superposition of bound conformations of the virtual hits and TEM template is shown in Fig. 9. The relatively tight overlap of the poses of the high ranking analogs with the experimental structure of the template inhibitor TEM indicates that the LigFit docking algorithm of Cerius^2 [31, 32] generated inhibitor geometries close to the observed binding mode of dTMP, and the PLP1 score correctly discriminated between realistic and artificial docking solutions.

Inhibitor: $TMPK_{mt}$ interactions

The designed analog with the highest predicted inhibitory potency D1 obtained from the semi-flexible docking and virtual screening contains a sulfondiamide group substituting the thiourethane moiety present in the added six-membered ring of the template inhibitors VAN1 and TEM, and a 5′-N-methyl-sulfonamide group replacing the phosphate group of the natural substrate (Fig. 8). Similar to VAN1, D1 is a neutral molecule that contains a number of proton acceptor groups, which match well with the high concentration of cationic side chains both from the LID region and from the phosphate-binding P-loop with arginines 74, 95, 149, 153, and 160 in the binding site of the $TMPK_{mt}$ [11, 15]. The bound conformation of D1 is stabilized by HB interactions of the fused ring with Arg95, Asp163, Gln172, the thymine moiety with Arg74 and the 5′-N-methyl-sulfonamide with Arg160, His53 and by ionic interaction with the Mg^{2+}. The pose of the D1 overlapped at the active site of $TMPK_{mt}$ closely with the crystal

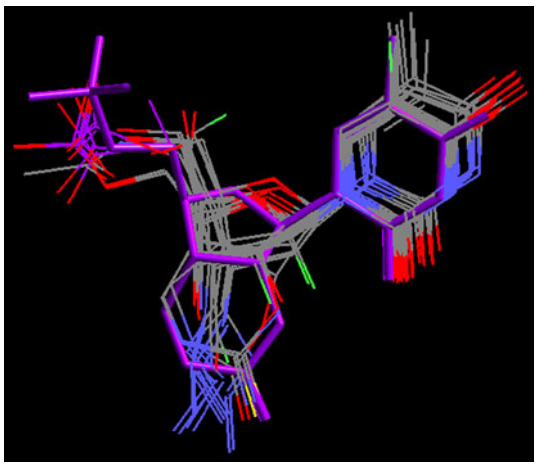

Fig. 6 Docked poses of training set inhibitors superimposed with TEM (cylinder rendering, purple color). TEM was used for the definition of the binding site grid (Fig. 4)

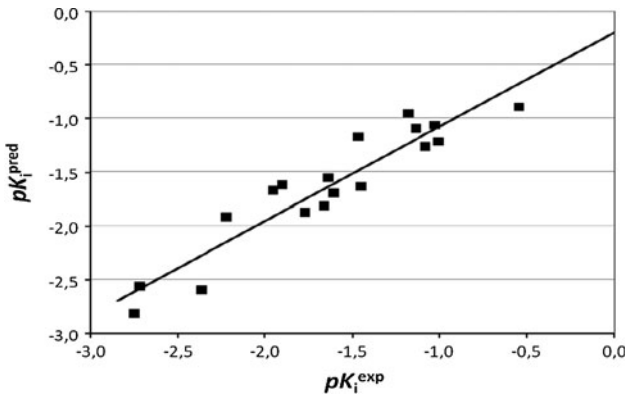

Fig. 7 Plot of the regression Eq. (1) showing the dependence of the predicted (pK_i^{pred}) and observed inhibitory activity (pK_i^{exp}) for the training set of 18 inhibitors

structure pose of the 5-CH_2OH dUMP inhibitor [15] (Fig. 10). Due to the larger size of the fused six-membered ring of the D1 analogue residues Asp 9 and Leu 52 were shifted up to 1 Å from their original crystal structure position and Lys 13, Arg 95, Tyr 103, Asp 163 and Gln 172 displayed slightly rotated side chains in the $TMPK_{mt}$-D1 complex.

We have analyzed the 16 best virtual hits D1–D16 (Table 2) in terms of the individual building block contributions (PHO, SUG, THY, Fig. 11) to the inhibitor-enzyme binding after a careful molecular mechanics refinement of the complexes resulting from the docking (see Computational Methods for the refinement details). The refinement carried out by molecular mechanics minimization using Insight-II program [46] partially compensated for the lost receptor flexibility that was neglected during the semi-flexible docking. After relaxation, most of

Table 2 Virtual thymidine hits predicted to inhibit $TMPK_{mt}$ derived by structure-based evaluation and virtual screening

Virtual hit	R-groups[a] Analog-R_1-$R_{2,3}$-R_4	Q_M^b [è]	PLP1[c]	K_i^{pre} [d] [μM]
D1	Analog-5-19-11	0	150.1	0.29
D2	Analog-5-19-5	0	148.7	0.34
D3	Analog-5-19-6	0	148.5	0.35
D4	Analog-5-19-10	0	147.3	0.40
D5	Analog-5-23-9	0	146.9	0.42
D6	Analog-5-27-6	0	144.6	0.53
D7	Analog-5-27-5	0	144.5	0.54
D8	Analog-5-24-5	0	144.0	0.57
D9	Analog-2-15-11	0	143.8	0.58
D10	Analog-5-19-9	0	143.7	0.59
D11	Analog-1-19-11	0	143.7	0.59
D12	Analog-1-19-5	0	143.1	0.64
D13	Analog-5-15-11	0	143.0	0.64
D14	Analog-5-23-5	0	142.9	0.65
D15	Analog-5-23-6	0	142.0	0.72
D16	Analog-3-19-5	0	141.8	0.73

[a] See Fig. 2 for the structure of individual R_1–R_4-groups and Fig. 8 and for the structure of the analogs D1–D16

[b] Q_M—total molecular charge at neutral pH

[c] empirical scoring function estimates the free energy of ligand—receptor binding [40, 41]

[d] inhibition constants predicted from the scoring function (Eq. (1)) specific to the $TMPK_{mt}$ target: $K_i^{pre} = 10^{-(a + b\ PLP1)}$

the virtual hits displayed improved total inhibitor-enzyme non-bonding interaction energies (E_{int}) with $TMPK_{mt}$ with respect to the template inhibitor VAN1 ($K_i^{exp} = 3.5$ μM) [14] (Table 3). The building block contribution analysis revealed that the hydroxymethyl group (R_1 fragment 5, Fig. 3) represents the most suitable substituent for the position 5 of thymine, the six-membered ring-containing sulfondiamide moiety ($R_{2,3}$ fragment 19, Fig. 3) fused with the cyclopentane ring is predicted to yield the best replacement of the bicyclic ribose moiety of VAN1, and the sulfonamides (R_4 fragments 9 and 11) correspond to the best replacement of the phosphate group of TEM. The analogs displaying the highest predicted binding affinity to the relaxed $TMPK_{mt}$ receptor are D10, D4, D1 and D13.

The largest contribution to E_{int} arises from the polar building blocks, which contain multiple proton acceptor groups. For the virtual hits, which are neutral molecules ($Q_M = 0$ è), the major stabilization at the $TMPK_{mt}$ binding site originates from the thymine building block, while the best SUG and PHO blocks contribute to the enzyme binding about equally.

We have computed the contributions of individual amino acid residues lining the binding site of $TMPK_{mt}$ to the E_{int} for the analog D10 (Table 4) to identify the

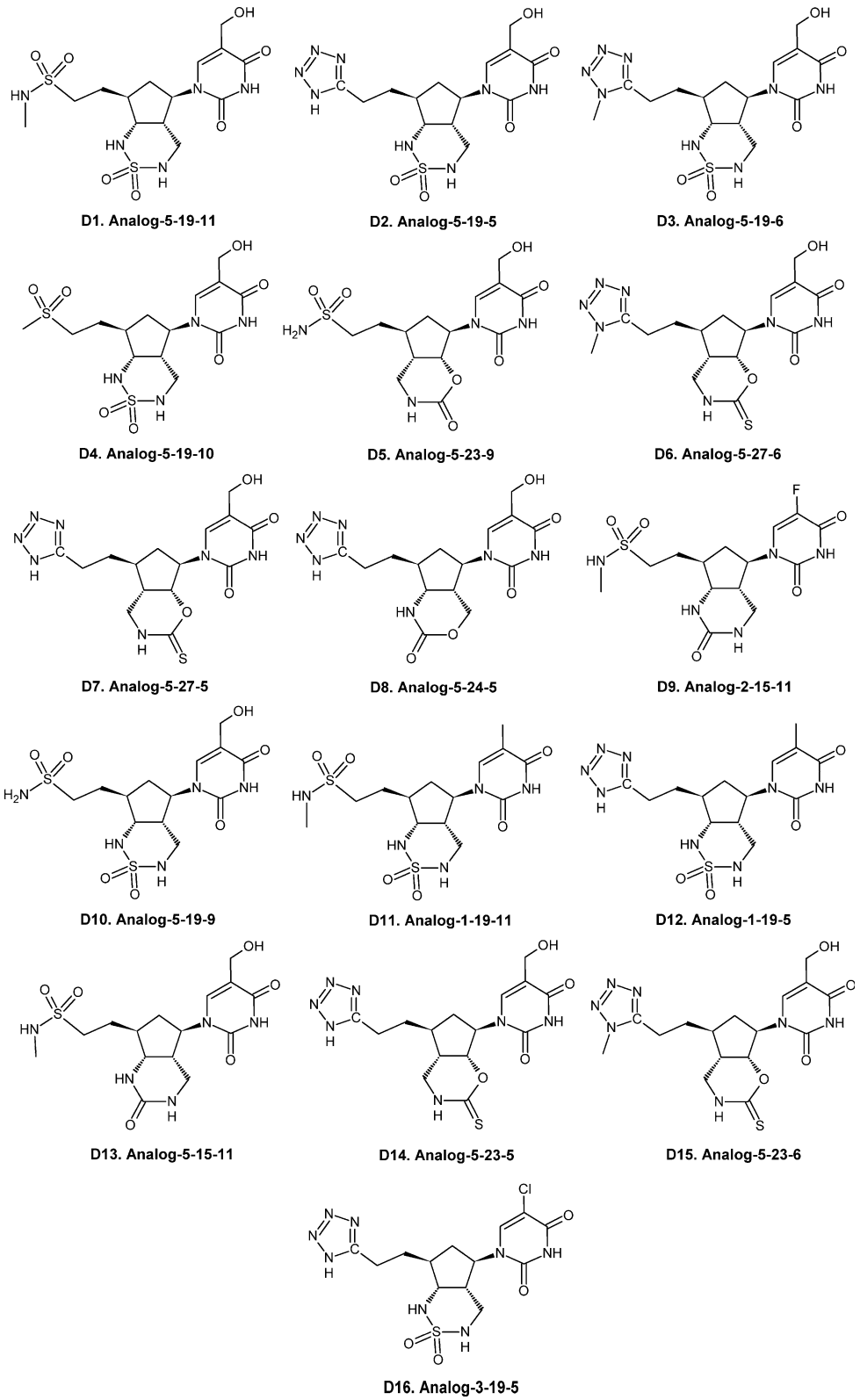

Fig. 8 Chemical structures of 16 best scoring thymidine analogs

strongest interactions. This analysis showed that the SUG fragment (Fig. 11) of D10 interacts strongly with Asp163, Tyr103 and Tyr165, while the PHO fragment binds strongly to Arg95, Arg153 and especially to the magnesium ion. Not surprisingly, out of 18 binding site residues located within 3.5 Å from the bound inhibitor, with exception of Asp163 that forms a HB with one of the sulfonediamide nitrogens, the stabilizing interactions

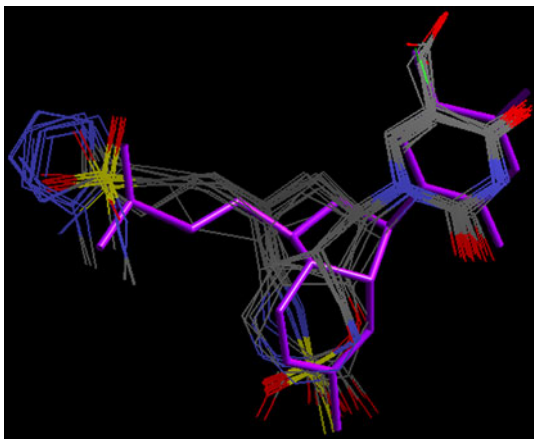

Fig. 9 Superposition of the docked poses of 16 best scoring analogs with the inhibitor TEM used for the binding grid definition (Fig. 4)

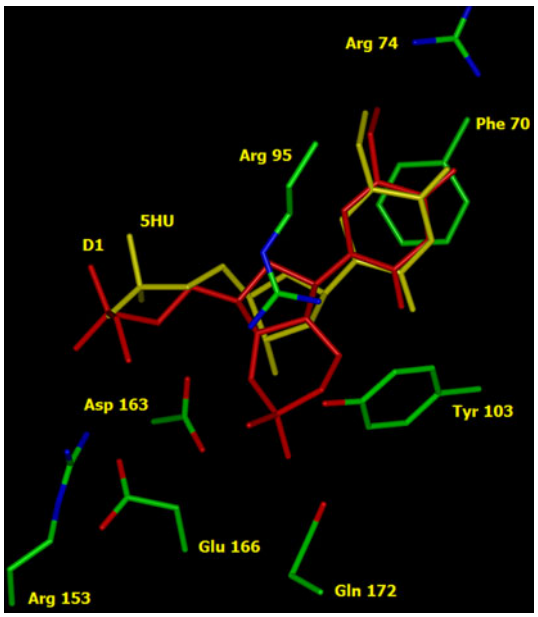

Fig. 10 Overlay of docked D1 (*red*) with the refined crystal structure of 5-CH_2OH dUMP inhibitor (5HU, *yellow*) at the active site of $TMPK_{mt}$ (PDB entry code 1MRS, [15]) obtained by superposition of backbone atoms of the receptors (r.m.s.d. = 0.59 Å). Side chains of key residues interacting with the inhibitor are shown

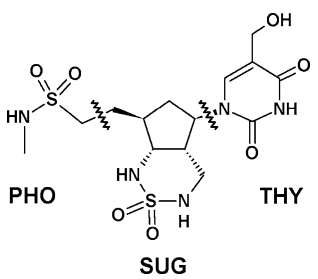

Fig. 11 Building blocks of inhibitor considered in the inhibitor-enzyme interaction energy partitioning

Table 3 Partitioning of total inhibitor—$TMPK_{mt}$ interaction energy into individual building block contributions for the 16 best virtual hits

Virtual hit[b]	E_{int}^{a} [kcal mol^{-1}]			
	THY[c] (R_1-group)	SUG ($R_{2,3}$-group)	PHO (R_4-group)	Whole inhibitor
D1	−38.5	−28.0	−19.2	−85.7
D2	−37.7	−21.4	−17.7	−76.8
D3	−39.5	−25.3	−12.3	−77.1
D4	−39.8	−29.8	−18.4	−88.0
D5	−43.6	−16.1	−20.1	−79.8
D6	−40.7	−15.6	−10.6	−66.9
D7	−42.1	−12.6	−18.6	−73.3
D8	−43.0	−19.5	−13.2	−75.7
D9	−37.4	−22.2	−22.0	−81.6
D10	−38.0	−27.5	−23.2	−88.7
D11	−35.5	−27.9	−17.6	−81.0
D12	−35.7	−21.5	17.6	−74.8
D13	−40.7	−23.9	−20.9	−85.5
D14	−40.5	−17.4	−18.6	−76.5
D15	−44.6	−17.9	−12.5	−75.0
D16	−37.0	−21.8	−17.6	−76.4
VAN1[d]	−38.9	−11.3	−15.8	−66.0

[a] E_{int} inhibitor-enzyme non-bonding interaction energy (1-6-9 potential) as defined in CFF91 force field [48], a dielectric constant ε of 2 was used

[b] See Fig. 8 for the structures of the virtual hits

[c] See Fig. 11 for the definition of the building blocks

[d] reference inhibitor VAN1 is listed for comparison

derive from proton donor residues, mainly arginines. For the D10, interaction with the Asp9, which is considered to be one of the crucial residues in the binding site of $TMPK_{mt}$ [11], is repulsive. Its role in the HB interaction with the ligands observed for VAN1 is in the case of D10 taken over by the Asp163 (Fig. 12).

Inhibitor: $TMPK_{mt}$ Gibbs free energy of binding

We have also calculated the Gibbs free energy of binding (ΔG_{bind}) of the hits to the $TMPK_{mt}$, in order to estimate more precisely the activities of the virtual hits predicted from the empirical PLP1 score. The approach computes ΔG_{bind} as the sum of the ligand-receptor interaction term derived by molecular mechanics, of the solvent effect contribution estimated from implicit model of solvation and of the entropic term derived from vibrational analysis [47] (see Computational Methods section for details). The relative binding affinities predicted by the complexation model and their components for the 16 best virtual hits computed with respect to the reference inhibitor VAN1 (K_i^{exp} = 3.5 μM) [14] are given in Table 5. Six of the

Table 4 Analysis of D10—$TMPK_{mt}$ interaction energy in terms of highest residue contributions to the inhibitor binding

Residue[b]	E_{int}^{a} [kcal mol^{-1}]			
	THY[c] (R_1-group)	SUG ($R_{2,3}$-group)	PHO (R_4-group)	D10 Total
Phe70	−6.0	−1.0	0.0	−7.0
Arg95	−8.0	2.7	−3.6	−8.9
Ser99	−4.3	−1.0	0.0	−5.3
Tyr103	−1.6	−3.6	−0.1	−5.3
Arg153	−3.3	−0.1	−3.8	−7.2
Asp163	4.9	−8.4	−1.3	−4.8
Tyr165	−1.5	−3.0	−0.2	−4.7
Asp9	3.7	−3.2	2.3	2.7
Mg^{2+}	−7.8	9.3	−12.0	−10.5
Binding site[d]	−37.8	−15.0	−18.7	−71.5
H_2O[e]	0.1	−1.7	0.4	−1.2

[a] E_{int} inhibitor-enzyme non-bonding interaction energy (1-6-9 potential) as defined in CFF91 force field [48], a dielectric constant ε of 2 was used

[b] selected residues lining the binding site of $TMPK_{mt}$

[c] for definition of the building blocks see Fig. 11

[d] interaction with binding site residues distant up to 3.5 Å form the inhibitor

[e] three crystallographic water molecules coordinating the Mg^{2+} ion and binding site residues

virtual hits display higher predicted binding affinities to $TMPK_{mt}$ than the reference inhibitor, with analogs D10 and D11 showing most promising $\Delta\Delta G_{bind}$ values. It is interesting to notice that receptor binding for D10 and D11 is predicted to be driven by a mixture of inhibitor-enzyme interactions ($\Delta\Delta H_{MM} < 0$), solvation ($\Delta\Delta G_{solv} < 0$) and entropic effects ($-\Delta\Delta TS_{vib} < 0$).

To demonstrate that the approach is able to identify potential $TMPK_{mt}$ inhibitors, we have performed a literature search for thymidine analogs known to inhibit $TMPK_{mt}$ [21–23] (Fig. 13), which are dissimilar to the designed bicyclic compounds (Fig. 8) and calculated their Gibbs free energies of binding to the $TMPK_{mt}$ relative to the reference inhibitor VAN1. By comparing the computed $\Delta\Delta G_{bind}$ values with the experimental inhibition constants K_i^{exp} (Table 6) we can see a reasonable correlation between the predicted and observed quantities. This correspondence demonstrates that the complexation model is able to identify known $TMPK_{mt}$ inhibitors and discriminate between active and inactive analogs.

Combinatorial subset selection

Over 100 analogs displayed activities predicted from the PLP1 score in submicromolar range. The relative frequency of occurrence of the individual R-groups was monitored in this set of potent analogs (Fig. 14). A clear preference for the hydroxymethyl group (fragment 5, Fig. 3) in R_1, the six-membered fused ring containing sulfondiamide moiety (fragment 19) in $R_{2,3}$ and the tetrazole (fragment 5) in R_4, is evident from Fig. 14.

The fragments that displayed the highest frequency of occurrence were selected to constitute a highly focused combinatorial subset of the initial diversity library (Fig. 15). This subset displays greatly increased probability to contain potent VAN1 analogs with $TMPK_{mt}$ inhibition activities in the submicromolar range. The size of the resulting combinatorial subset was narrowed down to only: 1 (R_1) × 7 ($R_{2,3}$) × 5 (R_4) = 35 analogs, which will allow the rapid synthesis and testing for their inhibitory activity.

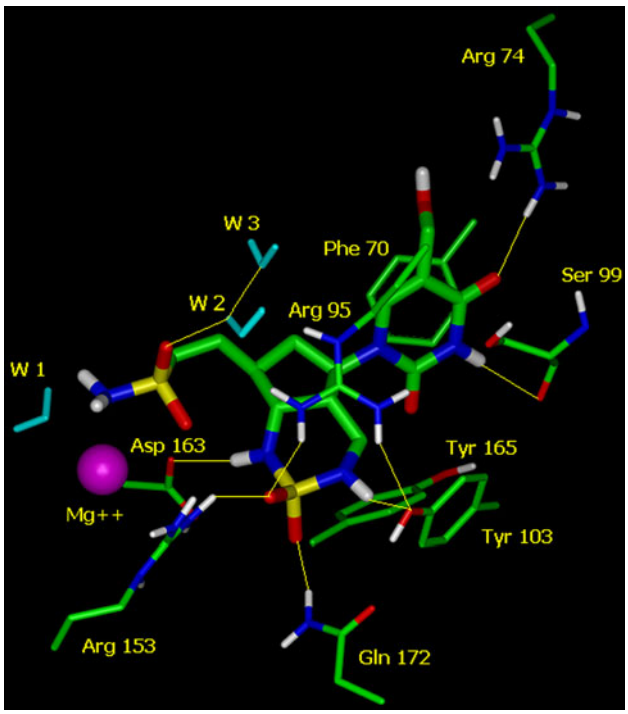

Fig. 12 Hydrogen bonding interactions of D10 at the active site of $TMPK_{mt}$ are displayed as yellow lines. D10 is shown in thick stick rendering. Only selected residue side chains with polar hydrogens are shown for better clarity. Measured bond lengths of hydrogen bonds: D10-Arg74: 2.2 Å, D10-Arg99: 2.7 Å, D10-Ser99: 2.8 Å, D10-Tyr103: 2.8 Å, D10-Arg153: 2.7 Å, D10-Arg163: 2.0 Å, D10-Gln172: 2.6 Å, D10-W2: 2.5 Å

ADME-related properties and hit prioritization

Frequently, drug candidates and lead compounds fail at late stages of pharmaceutical development due to unsatisfactory pharmacokinetic properties. It is therefore useful to incorporate early ADME properties prediction into the lead compound selection process.

We have computed a number of physicochemical molecular properties (descriptors) relevant to the

Table 5 Relative Gibbs free energy of binding to $TMPK_{mt}$ and its components for 16 best virtual hits

Virtual hit	$\Delta\Delta H^{a}_{MM}$ [kcal mol^{-1}]	$\Delta\Delta G^{b}_{solv}$ [kcal mol^{-1}]	$-\Delta\Delta TS^{c}_{vib}$ [kcal mol^{-1}]	$\Delta\Delta G^{d}_{bind}$ [kcal mol^{-1}]
D1	−0.8	−0.2	0.5	−0.5
D2	4.1	2.4	1.0	7.5
D3	6.7	2.2	−0.5	8.4
D4	−1.8	−0.7	0.4	−2.1
D5	3.7	3.5	2.5	9.6
D6	1.8	1.5	−1.4	2.0
D7	1.9	−0.9	−0.5	0.5
D8	4.0	−0.1	−2.5	1.5
D9	−2.4	1.5	−1.1	−2.0
D10	−4.2	−2.4	−1.9	−8.5
D11	−2.1	−1.9	−1.7	−5.7
D12	7.8	0.4	−0.5	7.6
D13	−1.3	−1.3	0.1	−2.4
D14	3.7	2.6	−3.0	3.3
D15	1.0	3.1	−0.2	3.9
D16	7.3	0.1	0.8	8.2
VAN1[e]	0.0	0.0	0.0	0.0

[a] $\Delta\Delta H_{MM}$ is the relative enthalpic contribution to the Gibbs free energy change of the enzyme-inhibitor complex formation derived by molecular mechanics: $\Delta\Delta H_{MM} \cong [E_{MM}\{TMPK_{mt}{:}D_i\} - E_{MM}\{D_i\}] - [E_{MM}\{TMPK_{mt}{:}VAN1\} - E_{MM}\{VAN1\}]$, VAN1 is the reference inhibitor

[b] $\Delta\Delta G_{solv}$ is the relative solvation Gibbs free energy contribution to the binding affinity: $G_{solv} = [G_{solv}\{TMPK_{mt}{:}D_i\} - G_{solv}\{D_i\}] - [G_{solv}\{TMPK_{mt}{:}VAN1\} - G_{solv}\{VAN1\}]$

[c] $-\Delta\Delta TS_{vib}$ is the relative entropic contribution to the binding affinity: $\Delta\Delta TS_{vib} = [TS_{vib}\{D_i\}_{TMPK} - TS_{vib}\{D_i\}] - [TS_{vib}\{VAN1\}_{TMPK} - TS_{vib}\{VAN1\}]$, $\{\}_{TMPK}$ means inhibitor at the binding site of 'frozen' receptor

[d] $\Delta\Delta G_{bind}$ is the total relative binding affinity: $\Delta\Delta G_{bind} \cong \Delta\Delta H_{MM} + \Delta\Delta G_{solv} - \Delta\Delta TS_{vib}$

[e] inhibitor VAN1 with K_i^{exp} of 3.5 μM [14] was used as the reference compound to derive the relative quantities

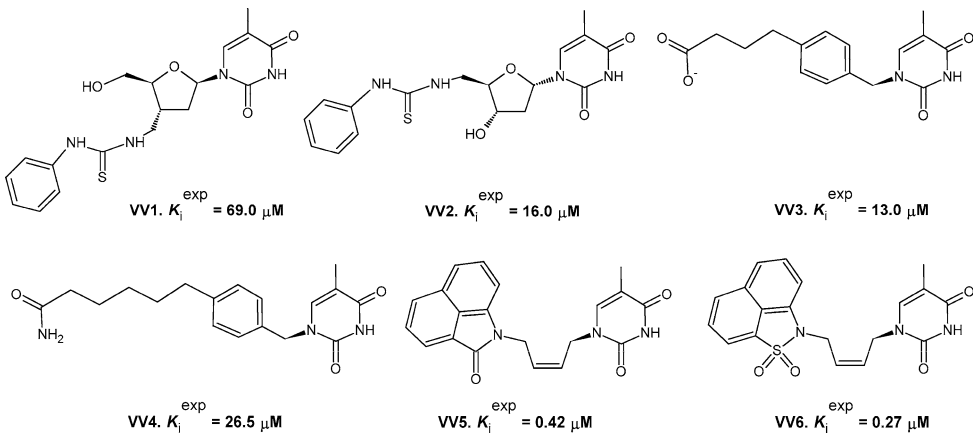

Fig. 13 Known thymidine analog inhibitors of $TMPK_{mt}$ and their experimental inhibiton constants [21–23]

pharmacokinetic profile of compounds that can be used for prioritization of virtual hits for further development. A set of 13 ADME-related properties out of total of 35 descriptors calculated by the QikProp program [49] based on the method of Jorgensen [50–52], is given in Table 7. We have used an overall score-drug likeness parameter (#stars, Table 7), to assess the ADME profiles of the hits. The #stars parameter indicates the number of descriptors computed by QikProp that fall outside of the optimum range of values fulfilled by 95% of known drugs and can thus serve as an overall indicator of favorable ADME-related properties and drug-like character.

All 16 best virtual hits with the highest predicted $TMPK_{mt}$ inhibitory potencies display overall ADME-related properties that comply with the expected pharmacokinetic profile of drug-like compounds. Out of them, six

Table 6 Inhibitory activities and computed relative Gibbs free energies of binding of an assessment set of known $TMPK_{mt}$ inhibitors (Fig. 13)

Inhibitor	$\Delta\Delta H^{a}_{MM}$ [kcal mol^{-1}]	$\Delta\Delta G_{solv}$ [kcal mol^{-1}]	$-\Delta\Delta TS_{vib}$ [kcal mol^{-1}]	$\Delta\Delta G_{bind}$ [kcal mol^{-1}]	$K_i^{exp\ b}$ [μM]
VV1	3.5	2.6	2.4	8.5	69.0[c]
VV2	3.6	1.4	0.4	5.4	16.0[c]
VV3	−20.6	23.3	2.9	5.6	13.0[d]
VV4	1.3	4.9	1.4	7.6	26.5[d]
VV5	−1.7	0.3	1.1	−0.3	0.42[e]
VV6	−4.9	2.8	0.2	−1.9	0.27[e]
VAN1	0.0	0.0	0.0	0.0	3.5[f]

[a] See the footnote of Table 5

[b] experimental inhibition constants of $TMPK_{mt}$

[c] Ref. [21]

[d] Ref. [22]

[e] Ref. [23]

[f] Ref. [14]

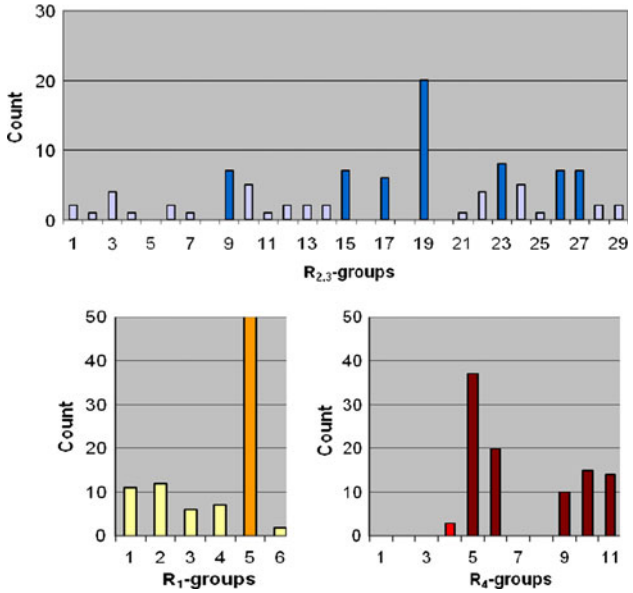

Fig. 14 Frequency of occurrence of R-groups in 100 best scoring VAN1 analogs

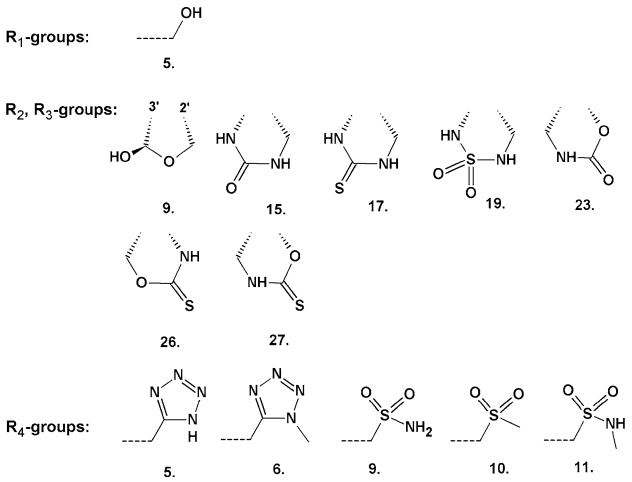

Fig. 15 Selected R-groups forming highly focused combinatorial subset

analogs display a profile approximately equal or better than that of the template inhibitor VAN1 (Table 7). Therefore, we can expect that some of the virtual hits could be further developed into lead compounds.

Conclusions

The presented study yielded a small highly focused virtual combinatorial subset of new bicyclic thymidine derivatives of the template compound VAN1, a potent selective inhibitor of TMPK of *Mycobacterium tuberculosis*. The potencies of the analogs from this combinatorial subset predicted from the QSAR model fall into the submicromolar range, and the D1 analog is predicted to be 10 times more potent than the template inhibitor VAN1, the most potent compound from the training set [14]. The most promising analogs D1, D4, D9, D10, D11 and D13 out of the 16 virtual hits are predicted to bind to the $TMPK_{mt}$ with an improved binding affinity compared to the template compound. The virtual hits are endowed with ADME-related properties comparable to those of VAN1, and appear to be drug-like compounds. In particular, D7 displays a lower polarity ($logP_{o/w}$) than VAN1, while being predicted to possess activity comparable to VAN1, and may thus also display in vitro inhibition of bacterial growth against the mycobacteria. The proposed set of R_1 to R_4 fragments optimized by structure-based focusing and virtual screening methods may represent optimum choice of building blocks for bicyclic nucleoside derivatives.

Table 7 Predicted ADME-related properties of the 16 best virtual hits computed by QikProp program [49] used to prioritize the hit selection

Virtual hits[a]	#stars[b]	M_w^c [g mol^{-1}]	S_{mol}^d [Å^2]	$S_{mol,hfob}^e$ [Å^2]	V_{mol}^f [Å^3]	RotB[g]	HB_{don}^h	HB_{acc}^i	$logP_{o/w}^j$	$logS_{wat}^k$ [g dm^{-3}]	$logK_{HSA}^l$	logB/B[m]	BIP_{caco}^n [nm s^{-1}]	#metab[o]	K_i^{pre} [p] [μM]
D1	3	437	645	248	1185	6	4	12	−0.9	−3.4	−0.5	−3.4	2.5	1	0.29
D2	4	412	634	203	1130	5	4	10	−0.9	−3.7	−0.4	−3.7	1.3	2	0.34
D3	3	426	679	270	1212	5	3	10	−0.2	−4.3	−0.2	−3.7	2.0	2	0.35
D4	3	422	632	246	1140	5	3	11	−0.8	−3.4	−0.5	−3.3	3.1	1	0.40
D5	3	388	596	206	1065	6	4	12	−1.6	−2.6	−0.8	−3.3	2.9	1	0.42
D6	0	407	672	273	1201	5	2	10	0.8	−4.5	−0.3	−2.5	15.6	2	0.53
D7	1	393	625	209	1122	5	3	10	0.1	−3.9	−0.4	−2.6	9.9	2	0.54
D8	3	377	616	214	1090	5	3	11	−1.0	−3.3	−0.5	−3.4	2.3	2	0.57
D9	1	389	590	226	1073	4	4	10	−0.1	−3.5	−0.4	−2.2	16.3	0	0.58
D10	3	423	618	203	1121	6	5	12	−1.5	−3.1	−0.6	−3.5	1.6	1	0.59
D11	1	421	646	290	1168	4	4	11	−0.4	−3.9	−0.4	−2.9	6.4	1	0.59
D12	3	396	622	237	1114	3	4	10	−0.4	−4.0	−0.2	−3.1	3.4	2	0.64
D13	3	401	636	259	1153	6	4	11	−0.6	−3.5	−0.4	−3.3	3.6	1	0.64
D14	4	377	611	209	1092	5	3	11	−0.9	−3.2	−0.5	−3.3	2.5	2	0.65
D15	2	391	653	273	1167	5	2	11	−0.3	−3.7	−0.4	−3.3	4.0	2	0.72
D16	2	417	615	152	1100	3	4	10	−0.2	−4.2	−0.3	−3.0	3.1	1	0.73
VAN1[q]	1	313	515	221	893	2	3	10	−0.1	−2.9	−0.6	−1.1	139.0	3	3.5[r]

[a] See Fig. 8 for the chemical structures of the virtual hits

[b] drug likeness, number of property descriptors (from the full list of 35 descriptors of QikProp, ver. 3.2) that fall outside the range of values for 95% of known drugs

[c] molecular weight, in Da (range for 95% of drugs: 130–725 Da) [49]

[d] total solvent-accessible molecular surface, in Å^2 (probe radius 1.4 Å) (range for 95% of drugs: 300–1000 Å^2);

[e] hydrophobic portion of the solvent-accessible molecular surface, in Å^2 (probe radius 1.4 Å) (range for 95% of drugs: 0–750 Å^2);

[f] total volume of molecule enclosed by solvent-accessible molecular surface, in Å^3 (probe radius 1.4 Å) (range for 95% of drugs: 500–2000 Å^3);

[g] number of rotatable bonds (range for 95% of drugs: 0–15)

[h] number of hydrogen bonds donated by the molecule (range for 95% of drugs: 0–6)

[i] number of hydrogen bonds accepted by the molecule (range for 95% of drugs: 2–20)

[j] logarithm of partitioning coefficient between n-octanol and water phases (range for 95% of drugs: −2 to 6.5)

[k] logarithm of aqueous solubility (range for 95% of drugs: −6.0 to 0.5)

[l] logarithm of predicted binding constant to human serum albumin (range for 95% of drugs: −1.5 to 1.5)

[m] logarithm of predicted blood/brain barrier partition coefficient (range for 95% of drugs: −3.0 to 1.2)

[n] predicted apparent Caco-2 cell membrane permeability in Boehringer-Ingelheim scale, in [nm/s] (range for 95% of drugs: <25 low, >500 high)

[o] number of likely metabolic reactions (range for 95% of drugs: 1–8)

[p] predicted K_i^{pre} were estimated from the QSAR Eq. (1) and the computed PLP1 score

[q] for comparison we list also the ADME-related properties of the reference inhibitor VAN1

[r] experimental activity of VAN1 [14]

Naturally, the predicted activities derived by this computational approach need to be verified by synthesis and inhibitory activity testing. Nevertheless, this study can help to direct the interest of medicinal chemists working on the preparation of antimycobacterial chemotherapeutics to this particular subset of the chemical space, which is predicted to contain bicyclic thymidine analogs with elevated inhibitory potencies towards the $TMPK_{mt}$.

Computational methods

Virtual library generation

Generation of the virtual library of thymidine analogs was accomplished in the Cerius^2 program [31] using the class II consistent force field CFF91 [48] and Rappé and Goddard equilibrated charges [53]. The library of analogs was enumerated by attaching the R-groups (fragments) onto the

dTMP-like scaffold (Fig. 3), which was kept in its refined bound conformation [15], using the CombiChem module of Cerius^2. Reagents (fragments) considered in this study were taken from the directories of available chemicals [54]. Each analog was built as a neutral molecule or molecular ion containing a phosphate, sulphonate or carboxylate R_4-group. Molecular geometry of the attached R-groups was refined by molecular mechanics (MM) energy minimization by employing smart minimizer algorithm of the Cerius^2 with high convergence criteria (energy difference of 10^{-4} kcal mol^{-1}, r.m.s. displacement of 10^{-5} Å) and a dielectric constant of 2.

$TMPK_{mt}$ target

To prepare a suitable 3D model of *M. tuberculosis* $TMPK_{mt}$ drug target we have first modified the complex of $TMPK_{mt}$ with 5-CH_2OH dUMP inhibitor (PDB entry code 1MRS) [15] by building in the thiourethane moiety of the second 2′,3′-fused ring of the bicyclic inhibitor VAN1 (Fig. 1) [14, 20]. The TEM template inhibitor (Fig. 4) contains a bicyclic sugar moiety of VAN1, 5′-O-phosphoryl group and 5-hydroxymethyl group representing thus compound that is large enough to define a binding site grid that can accommodate analogs derivatized with R_1- to R_4-groups. The complex included 4 crystallographic water molecules located near the Mg^{2+} binding site [15]. Then the enzyme-inhibitor complex was carefully refined by MM energy-minimization using Insight-II program [46], all-atom representation and CFF91 force field [48]. A non-bonding interaction distance cut-off was set to 15 Å and dielectric constant of 2 was used to take into account dielectric shielding effects in proteins upon the long-term electrostatic interactions. Minimization of the complex was carried out by relaxing the structure gradually, starting with hydrogen atoms, then proceeding with all inhibitor atoms, followed by $TMPK_{mt}$ residue side chains and concluding the relaxation with freeing of all atoms including the enzyme backbone. In the geometry optimization, a sufficient number of conjugate gradient iterative cycles, with a convergence criterion set to the average gradient of 0.01 kcal mol^{-1} Å^{-1}, have been used.

Structure-based library evaluation

3D structure of the $TMPK_{mt}$-TEM complex was used as the target receptor for docking of virtual library of VAN1 analogs. The binding site of $TMPK_{mt}$ receptor was mapped onto a 3D energy grid of the size (48 × 62 × 65) points with a resolution of 0.33 Å per unit after augmenting it by 3 Å in all directions. The grid was enumerated by the LigFit algorithm of the Cerius^2 [31] using a non-bonding interaction cut-off distance of 12 Å and a dielectric constant of 2.

The shape and size of the ligand binding site was defined from the bound TEM (see previous paragraph). The binding site model was enlarged in areas around the R_1- to R_4-groups by two grid layers. Then conformers of each analog were generated by randomizing the dihedral angles (10^4 Monte Carlo steps) and subsequently docked to the $TMPK_{mt}$ binding site model. The generated conformers were fitted to the site model via flexible fit algorithm by comparing principal moments of inertia of the site and the analog after 50 rigid body minimization steps of 4 ligand orientations and 200 steps of final minimization [32]. Subsequently, the analogs were energy minimized at the binding site of $TMPK_{mt}$ with a maximum number of iterations set to 800 steps using the smart minimizer of Cerius^2 [31]. The docking score (ligand-receptor interaction energy) was computed as the non-bonding MM energy term using the CFF91 force field [48] for each docked conformer. An energy penalty of 200 kcal mol^{-1} was applied to each ligand atom placed outside of the grid. During the docking score calculation a grid representation of rigid $TMPK_{mt}$ receptor, a cut-off distance of 12 Å applied to the non-bonded interactions and a dielectric constant of 2, were used. Ten best-fitting conformers were saved and clustered into five conformational families according to their mutual r.m.s. deviations by means of Jarvis–Patrick complete linkage clustering method [33]. The best representative of each cluster was considered in the virtual screening of the analogs.

QSAR model

A training set of 18 and a validation set of 5 dTMP analogs prepared by Vanheusden et al. [13, 14, 17, 18] and Van Daele et al. [20] with inhibitory potencies (K_i^{exp}) determined in a spectrophotometric binding assay [55] on a recombinant $TMPK_{mt}$ enzyme, were used. Molecular models of the inhibitors were docked to the binding site of $TMPK_{mt}$ receptor using the LigFit docking procedure [32] of Cerius^2 [31]. Various scores implemented in the Cerius^2 (such as LigScore, LUDI, PMF, and PLP) [32, 35–41] were computed for the best clustered poses and several QSAR models, which relate the experimental activity data K_i^{exp} to the computed scores, were prepared by linear regression analysis encoded in the QSAR module of the Cerius^2. The regression equation $pK_i = f(PLP1)$, which correlates the computed piecewise linear potential type 1 score (PLP1) [40, 41] with the experimental K_i^{exp} of the $TMPK_{mt}$ [13–18, 20], displayed the highest statistical significance. The PLP1 score models the attraction and repulsion (hydrogen bonding and steric interactions) between protein and ligand atoms (both heavy atoms and hydrogens were included)

and emphasizes the H-bonding interactions [40, 41]. The statistical accuracy of this QSAR model was verified by leave-one-out cross validation method. The predictive power of the regression equation, which was then used as the target-specific scoring function for *in silico* screening, was also verified by applying it to external validation set of 5 $TMPK_{mt}$ inhibitors with known K_i^{exp} values, which were not included into the training set.

In silico screening

The best member in each of the 5 clusters of conformational families obtained from analog docking to the $TMPK_{mt}$ binding site was selected for virtual screening carried out with help of PLP1 score [40, 41] and the parameterized scoring function $pK_i = f(PLP1)$. The PLP1 score was used for predicting the $TMPK_{mt}$ inhibitory potencies (K_i^{pre}) of the structure-focused virtual library by employing this parameter in the target-specific scoring function. This scoring function, specific for the TMPK of *Mycobacterium tuberculosis* in the form $pK_i^{pre} = a + b{\cdot}PLP1$, was parameterized in the QSAR model described above.

Receptor flexibility and interaction energy partitioning

The effect of the receptor flexibility was introduced partially by full relaxation of the $TMPK_{mt}$-inhibitor complexes of 16 best scoring virtual hits obtained by semi-flexible docking from the LigFit algorithm of Cerius^2 [31]. The inhibitor-enzyme interaction energy (E_{int}) was computed as the non-bonding interaction energy (1-6-9 non-bonding pairwise interatomic interaction potential that includes coulombic, dispersion and repulsion terms as defined in CFF91 force field [48]). E_{int} was partitioned into contributions of the main building blocks (thymine, bicyclic component and phosphate isostere) in the minimized complexes using Insight-II modeling package [46]. Original net atomic charges, van der Waals and repulsion atomic parameters of CFF91 were used together with a dielectric constant of 2 to account partially for the dielectric shielding effect within the target protein.

Calculation of ligand-receptor Gibbs free energy of binding

Calculation of binding affinity

The standard Gibbs free energy change of the enzyme-inhibitor complex (E:I) formation, $\Delta G_{bind} = -RT\ln K_i$, can be derived by molecular simulations of the E:I and free enzyme (E) and inhibitor (I) as $\Delta G_{bind} = G\{E:I\} - G\{E\} - G\{I\}$, assuming the equilibrium in solution: $\{E\}_{aq} + \{I\}_{aq} \leftrightarrow \{E:I\}_{aq}$. We approximate the exact values of standard Gibbs free energies in the complexation model [47] by the expression: $G\{E:I\} \cong [E_{MM}\{E:I\} + RT - TS_{trv}\{E:I\}] + G_{solv}\{E:I\}$, where $E_{MM}\{E:I\}$ stands for molecular mechanics total energy of E:I, $G_{solv}\{E:I\}$ is the solvation Gibbs free energy and $TS_{trv}\{E:I\}$ the entropic term: $TS_{trv}\{E:I\} = TS_{trans}\{E:I\} + TS_{rot}\{E:I\} + TS_{vib}\{E:I\}$, a sum of translational, rotational and vibrational contributions. When assuming that translational and rotational terms for E and E:I are approximately equal and postulate ideal gas behaviour for the rotational and translational motions of the inhibitor, we obtain for the relative changes in the binding affinity ($\Delta\Delta G_{bind}$) between an analog and a reference inhibitor (I_{ref}): $\Delta\Delta G_{bind} = \Delta G_{bind}(I) - \Delta G_{bind}(I_{ref}) \cong [E_{MM}\{E:I\} - E_{MM}\{I\}] - [E_{MM}\{E:I_{ref}\} - E_{MM}\{I_{ref}\}] + [G_{solv}\{E:I\} - G_{solv}\{I\}] - [G_{solv}\{E:I_{ref}\} - G_{solv}\{I_{ref}\}] - [TS_{vib}\{E:I\} - TS_{vib}\{I\}] - [TS_{vib}\{E:I_{ref}\} - TS_{vib}\{I_{ref}\}] = \Delta\Delta H_{MM} + \Delta\Delta G_{solv} - \Delta\Delta TS_{vib}$. The evaluation of relative changes is prefered as it is expected to lead to partial cancellation of errors caused by the approximate nature of the MM method as well as solvent and entropic effects description.

This approach to fast estimation of Gibbs free energy of binding bears similarity to the MM-PBSA methods of P. Kollman [56]. The MM-PBSA has been previously successfully applied to binding free energy calculations for several biological systems [57–59].

Conformational search

Free I conformations were derived from their bound conformations in the E:I complexes by gradual relaxation to the nearest local energy minimum. Then a Monte Carlo search (with an upper limit of 50000 iterations) for low energy conformations over all rotatable bonds except those in the rings, was carried out using Cerius^2 [31]. Five hundred unique conformations were generated per each I by randomly varying torsion angles of the last accepted conformer by ±10 deg at 5000 K followed by subsequent energy minimization. During the minimization a dielectric constant $\varepsilon = 80$ was used to account approximately for the dielectric screening effect of hydration upon the generated conformers. The conformer with the lowest total energy was re-minimized at $\varepsilon = 2$.

Solvation Gibbs free energies

The electrostatic component of solvation Gibbs free energy that incorporates also the effects of ionic strength through the solution of nonlinear Poisson-Boltzmann equation [60, 61] was computed by the DelPhi software module of Insight-II [46]. This program treats the solvent as a continuous medium of high dielectric constant ($\varepsilon_o = 80$) and the solute as a cavity with low dielectric ($\varepsilon_i = 2$) with

boundaries linked to the solute's molecular surface, which encloses all atomic charges of the solute. DelPhi uses a finite difference method to solve for the molecular electrostatic potential and reaction field around the solute. Calculations were carried out on a (175 × 175 × 175) cubic lattice grid for the E:I and free E and (75 × 75 × 75) grid for the free I with full Coulombic boundary conditions. Two subsequent focusing steps led in both cases to a similar final resolution of about 0.3 Å per grid unit at the final 70% filling of the grid by the solute. Physiological ionic strength of 0.145 mol·dm^{-3}, atomic partial charges and radii defined in the CFF91 parameter set [48] and a probe sphere radius of 1.4 Å were used. The electrostatic component of the solvation Gibbs free energy was calculated as the corrected reaction field energy [60, 61].

Entropic term

The vibrational entropy change during binding of I to E was calculated by normal mode analysis of the I using a simplified method of Fischer et al. [62, 63]. In this approach vibrational analyses of I bound at the active site of a 'frozen' E replaces the analysis of the E:I complex: $TS_{\mathrm{vib}}\{\mathrm{E{:}I}\} \sim TS_{\mathrm{vib}}\{\mathrm{I}\}_{\mathrm{E}}$. The normal mode analysis of the bound I and of the lowest-energy conformer of free I were computed for fully minimized structures using the Discover package of Insight-II [46]. It has been shown previously that for small and relatively stiff ligands this method gives a good approximation of the vibrational entropy change of the fully flexible system, i.e. including the degrees of freedom of the protein receptor [62, 63]. Vibrational motions represent an indicator of conformational flexibility of the molecule. Namely, low frequency vibrations, which correspond to collective motions of a number of atoms with larger amplitudes, i.e. conformational changes, contribute most to the TS_{vib} term. Relative values of $\Delta\Delta TS_{\mathrm{vib}}$ with respect to the $\mathrm{I}_{\mathrm{ref}}$ were used to compensate partially for the restricted flexibility of E.

ADME-related properties

Prediction of descriptors related to adsorption, distribution, metabolism and excretion (ADME) of the analogs was carried out by QikProp program [49] based on the method of Jorgensen [50–52]. The QikProp computes besides the Lipinski parameters [64] also properties such as hydrophobic molecular surface, octanol/water partitioning coefficient, aqueous solubility, brain/blood partition coefficient, Caco-2 cell permeability, serum protein binding, number of likely metabolic reactions, predicted blockage of potassium channels, and others. Drug likeness (#stars)—the number of molecular property descriptors from the list of 24 principal descriptors computed by the program, that fall outside the optimum range of values determined for 95% of known drugs, was used as an additional compound selection filter.

Acknowledgments This work has been done within the ICS-UNIDO global program on Rational Drug Design and Discovery. Overall support of this work by ICS-UNIDO is gratefully acknowledged. We thank to Dr. Eugene Megnassan for technical assistance.

References

1. Global Tuberculosis Control—Epidemiology, Strategy, Financing (2009) WHO report 2009.411, World Health Organization, Geneva. Retrieved from http://www.who.int/tb/publications/global_report/2009/en/index.html. 1/9/2010
2. Ginsberg AM, Spigelman M (2007) Challenges in tuberculosis drug research and development. Nat Med 13:290–294
3. WHO Global Task Force: Outlines Measures to Combat XDR-TB Worldwide (2006) WHO press release, World Health Organization, Geneva. Retrieved from http://www.who.int/mediacentre/news/notes/2006/np29/en/index.html. 1/9/2010
4. Mitnick CD, McGee B, Peloquin CA (2009) Tuberculosis pharmacotherapy: strategies to optimize patient care. Expert Opin Pharmacother 10:381–401
5. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry C III, Tekaia CF, Badcock K, Basham D, Brown D, Chillingworth DT, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Barrell BG (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393:537–544
6. Long MC, Parker WB (2006) Structure-activity relationship for nucleoside analogs as inhibitors or substrates of adenosine kinase from *Mycobacterium tuberculosis*. I. Modifications to the adenine moiety. Biochem Pharmacol 71:1671–1682
7. Williams KJ, Duncan K (2007) Current strategies for identifying and validating targets for new treatment-shortening drugs for TB. Curr Mol Med 7:297–307
8. Munier-Lehmann H, Chafotte 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
9. Anderson E (1973) In: Boyer PD (ed) The enzymes, vol 8. Academic Press, New York, pp 49–96
10. Sclafani RA, Fangman WL (1984) Yeast gene CDC8 encodes thymidylate kinase and is complemented by herpes thymidine kinase gene TK. Proc Natl Acad Sci USA 81:5821–5825
11. Li de la Sierra I, Munier-Lehmann H, Gilles AM, Bârzu O, Delarue M (2001) X-ray structure of TMP kinase from *Mycobacterium tuberculosis* complexed with TMP at 1.95 A resolution. J Mol Biol 311:87–100
12. Li de la Sierra I, Munier-Lehmann H, Gilles AM, Bârzu O, Delarue M (2000) Crystallization and preliminary X-ray analysis of the thymidylate kinase from *Mycobacterium tuberculosis*. Acta Crystallogr Sect D Biol Crystallogr 56:226–228
13. Vanheusden V, Munier-Lehmann H, Pochet S, Herdewijna 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
14. 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
15. Haouz A, Vanheusden V, Munier-Lehmann H, Froeyen M, Herdewijn P, Van Calenbergh S, Delarue M (2003) Enzymatic and structural analysis of inhibitors designed against *Mycobacterium tuberculosis* thymidylate kinase. New insights into the phosphoryl transfer mechanism. J Biol Chem 278:4963–4971
16. Pochet S, Dugué L, Douguet D, Labesse G, Munier-Lehmann H (2002) Nucleoside analogues as inhibitors of thymidylate kinases: possible therapeutic applications. ChemBioChem 3:108–110
17. Vanheusden V, Van Rompaey P, Munier-Lehmann H, Pochet S, Herdewijn P, Van Calenbergh S (2003) Thymidine and thymidine-5′-O-monophosphate analogues as inhibitors of *Mycobacterium tuberculosis* thymidylate kinase. Bioorg Med Chem Lett 13:3045–3048
18. 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
19. Pochet S, Dugué L, Labesse G, Delepierre M, Munier-Lehmann H (2003) Comparative study of purine and pyrimidine nucleoside analogues acting on the thymidylate kinases of *Mycobacterium tuberculosis* and of humans. ChemBioChem 4:742–747
20. Van Daele I, Munier-Lehmann H, Hendrickx PM, Marchal G, Chavarot P, Froeyen M, Qing L, Martins CJ, Van Calenbergh S (2006) Synthesis and biological evaluation of bicyclic nucleosides as inhibitors of *M. tuberculosis* thymidylate kinase. ChemMedChem 1:1081–1090
21. Van Daele I, Munier-Lehmann H, Froeyen M, Balzarini J, Van Calenbergh S (2007) Rational design of 5′-thiourea-substituted alpha-thymidine analogues as thymidine monophosphate kinase inhibitors capable of inhibiting mycobacterial growth. J Med Chem 50:5281–5292
22. Gasse C, Douguet D, Huteau V, Marchal G, Munier-Lehmann H, Pochet S (2008) Substituted benzyl-pyrimidines targeting thymidine monophosphate kinase of *Mycobacterium tuberculosis*: synthesis and in vitro anti-mycobacterial activity. Bioorg Med Chem 16:6075–6085
23. 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. ChemMedChem 3:1083–1093
24. Frecer V, Burello E, Miertus S (2005) Combinatorial design of nonsymmetrical cyclic urea inhibitors of aspartic protease of HIV-1. Bioorg Med Chem 13:5492–5501
25. Frecer V, Megnassan E, Miertus S (2009) Design and in silico screening of combinatorial library of antimalarial analogs of triclosan inhibiting Plasmodium falciparum enoyl-acyl carrier protein reductase. Eur J Med Chem 44:3009–3019
26. Rungrotmongkol T, Frecer V, De-Eknamkul W, Hannongbua S, Miertus S (2009) Design of oseltamivir analogs inhibiting neuraminidase of avian influenza virus H5N1. Antivir Res 82:51–58
27. Rungrotmongkol T, Udommaneethanakit T, Frecer V, Miertus S (2010) Combinatorial design of avian influenza neuraminidase inhibitors containing pyrrolidine core with a reduced susceptibility to viral drug resistance. Comb Chem High Throughput Screen 13:268–277
28. Frecer V, Miertus S (2010) Design, structure-based focusing and in silico screening of combinatorial library of peptidomimetic inhibitors of Dengue virus NS2B-NS3 protease. J Comput Aided Mol Des 24:195–212
29. Pérez-Pérez M-J, Priego EM, Hernández AI, Camarasa MJ, Balzarini J, Liekens S (2005) Thymidine phosphorylase inhibitors: recent developments and potential therapeutic applications. Mini-Rev Med Chem 5:1113–1123
30. Kifli N, De Clercq E, Balzarini J, Simons C (2004) Novel bicyclic sugar modified nucleosides: synthesis, conformational analysis and antiviral evaluation. Bioorg Med Chem 12:3247–4252
31. Cerius^2 Life Sciences software, version 4.6 (2002) Accelrys, San Diego, CA
32. Peters KP, Fauck J, Frommel C (1996) The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria. J Mol Biol 256:201–213
33. Jarvis RA, Patrick EA (1973) Clustering using a similarity measure based on shared near neighbors. IEEE Trans Comput C22:1025–1034
34. Seifert MH (2009) Targeted scoring functions for virtual screening. Drug Discov Today 14:562–569
35. Böhm HJ (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. J Comput Aided Mol Des 8:243–256
36. Böhm HJ (1994) On the use of LUDI to search the Fine Chemicals Directory for ligands of proteins of known three-dimensional structure. J Comput Aided Mol Des 8:623–632
37. Muegge I, Martin YC (1999) A general and fast scoring function for protein-ligand interactions: a simplified potential approach. J Med Chem 42:791–804
38. Muegge I, Martin YC, Hajduk PJ, Fesik SW (1999) Evaluation of PMF scoring in docking weak ligands to the FK506 binding protein. J Med Chem 42:2498–2503
39. Muegge I (2001) Effect of Ligand Volume Correction on PMF Scoring. J Comput Chem 22:418–425
40. Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DB, Fogel LJ, Freer ST (1995) Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. Chem Biol 2:317–324
41. Verkhivker GM, Bouzida D, Gehlhaar DK, Reijto PA, Arthurs S, Colson AB, Freer ST, Larson V, Luty BA, Marrone T, Rose PW (2000) Deciphering common failures in molecular docking of ligand-protein complexes. J Comput Aided Mol Des 14:731–751
42. Holloway MK, McGaughey GB, Coburn CA, Stachel SJ, Jones KG, Stanton EL, Gregro AR, Lai M-T, Crouthamel M-C, Pietrak BL, Munshi SK (2007) Evaluating scoring functions for docking and designing beta-secretase inhibitors. Bioorg Med Chem Lett 17:823–827
43. Kontoyianni M, Sokol GS, McClellan LM (2005) Evaluation of library ranking efficacy in virtual screening. J Comput Chem 26:11–22
44. Krovat EM, Langer T (2004) Impact of scoring functions on enrichment in docking-based virtual screening: an application study on renin inhibitors. J Chem Inf Comput Sci 44:1123–1129
45. Wang R, Lu Y, Wang S (2003) Comparative evaluation of 11 scoring functions for molecular docking. J Med Chem 47:2287–2303
46. Insight-II Life Sciences software, version (2005) Accelrys, San Diego, CA
47. Frecer V, Berti F, Benedetti F, Miertus S (2008) Design of peptidomimetic inhibitors of aspartic protease of HIV-1 including–PheΨPro–core and favorable ADME properties. J Mol Graphics Model 27:376–387
48. Maple JR, Hwang MJ, Stockfish TP, Dinur U, Waldman M, Ewing CS, Hagler AT (1994) Derivation of class II force fields. I. Methodology and quantum force field for the alkyl functional group and alkane molecules. J Comput Chem 15:162–182
49. QikProp ADME Prediction software, version 3.2, Schrödinger, New York, NY

50. Duffy EM, Jorgensen WL (2000) Prediction of properties from simulations: free energies of solvation in hexadecane, octanol, and water. J Am Chem Soc 122:2878–2888
51. Jorgensen WL, Duffy EM (2000) Prediction of drug solubility from Monte Carlo simulations. Bioorg Med Chem Lett 10:1155–1158
52. Jorgensen WL, Duffy EM (2002) Prediction of drug solubility from structure. Adv Drug Deliv Rev 54:355–366
53. Rappé AK, Goddard WA III (1991) Charge equilibration for molecular dynamics simulations. J Phys Chem 95:3358–3363
54. Available Chemicals Directory, version 3.0, Symyx Technologies, Santa Clara, CA
55. 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 220:219–222
56. Srinivasan J, Miller J, Kollman PA, Case DA (1998) Continuum solvent studies of the stability of RNA hairpin loops and helices. J Biol Struct Dyn 16:671–682
57. Chong LT, Duan Y, Wang L, Massova I, Kollman PA (1999) Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. Proc Natl Acad Sci USA 96:14330–14335
58. Lee MR, Duan Y, Kollman PA (2000) Use of MM-PBSA in estimating the free energies of proteins: application to native, intermediates, nand unfolded villin headpiece. Proteins Struct Funct Genet 39:309–316
59. Wang J, Morin P, Wang W, Kollman PA (2001) Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. J Am Chem Soc 123:5221–5230
60. Gilson MK, Honig B (1991) The inclusion of electrostatic hydration energies in molecular mechanics calculations. J Comput Aided Mol Des 5:5–20
61. Rocchia W, Sridharan S, Nicholls A, Alexov E, Chiabrer A, Honig B (2002) Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. J Comput Chem 23:128–137
62. Fischer S, Smith JC, Verma C (2001) Dissecting the vibrational entropy change on protein/ligand binding: burial of a water molecule in bovine pancreatic trypsin inhibitor. J Phys Chem B 105:8050–8055
63. Schwarzl SM, Tschopp TB, Smith JC, Fischer S (2002) Can the calculation of ligand binding free energies be improved with continuum solvent electrostatics and an ideal-gas entropy correction? J Comput Chem 23:1143–1149
64. 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 Delivery Rev 46:3–26