

# Prediction of dihydrofolate reductase inhibition and selectivity using computational neural networks and linear discriminant analysis

Brian E. Mattioni, Peter C. Jurs\*

Department of Chemistry, The Pennsylvania State University, 152 Davey Laboratory, University Park, PA 16802, USA

Received 3 August 2002; accepted 22 October 2002

## Abstract

A data set of 345 dihydrofolate reductase inhibitors was used to build QSAR models that correlate chemical structure and inhibition potency for three types of dihydrofolate reductase (DHFR): rat liver (rl), *Pneumocystis carinii* (pc), and *Toxoplasma gondii* (tg). Quantitative models were built using subsets of molecular structure descriptors being analyzed by computational neural networks. Neural network models were able to accurately predict log IC<sub>50</sub> values for the three types of DHFR to within  $\pm 0.65$  log units (data sets ranged  $\sim 5.5$  log units) of the experimentally determined values. Classification models were also constructed using linear discriminant analysis to identify compounds as selective or nonselective inhibitors of bacterial DHFR (pcDHFR and tgDHFR) relative to mammalian DHFR (rlDHFR). A leave-*N*-out training procedure was used to add robustness to the models and to prove that consistent results could be obtained using different training and prediction set splits. The best linear discriminant analysis (LDA) models were able to correctly predict DHFR selectivity for  $\sim 70\%$  of the external prediction set compounds. A set of new nitrogen and oxygen-specific descriptors were developed especially for this data set to better encode structural features, which are believed to directly influence DHFR inhibition and selectivity.

© 2002 Elsevier Science Inc. All rights reserved.

**Keywords:** Prediction; Inhibition; Selectivity; Structural descriptors; Genetic algorithm

## 1. Introduction

The dihydrofolate reductase (DHFR) enzyme appears in many living organisms such as bacteria, plants, and mammals and is responsible for catalyzing the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF), which is crucial to facilitate cell growth and reproduction [1]. Effective inhibition of DHFR activity lowers THF levels causing a disruption in RNA, DNA, and protein synthesis ultimately leading to cell death [2]. Therefore, DHFR inhibition is the basis of treatment for immuno-deficient patients who lack a healthy immune system to combat diseases caused by the opportunistic pathogens, *Pneumocystis carinii* (pc) and *Toxoplasma gondii* (tg). Diseases originating from these pathogens (*Pneumocystis carinii* Pneumonia and Toxoplasmosis) are among the leading causes of death for people suffering from AIDS [3,4].

The goal of treatment is to employ drugs that decrease bacterial cell growth via DHFR inhibition without affect-

ing the essential functions of human DHFR. Unfortunately, existing treatments are beset with high costs and moderately low selectivity resulting in discontinued therapy because of side effects. Also, current drugs often must be co-administered with other drugs for synergistic effects, which is a relatively expensive approach to therapy. Therefore, one of the continuous goals by several research laboratories is to develop stand-alone drugs which not only show the desired potency but which are also selective towards the bacterial forms of DHFR [5–8].

The development of new DHFR inhibitors can be a long and expensive task for individual researchers and the pharmaceutical industry, making it impractical to experimentally synthesize and test every possible lead structure. The ability of quantitative structure–activity relationship (QSAR) models to screen large groups of drug candidates would be advantageous to predict drug selectivity and potency. Instead of time and money being wasted on the testing of undesirable DHFR inhibitors, computational techniques could cheaply predict and label promising new drugs that are not only strong DHFR inhibitors, but selective towards pcDHFR and tgDHFR relative to a mammalian standard (rlDHFR). Despite the limited insight that some QSAR models bring to the understanding of the overall mechanism of activity, accurate correlations have been achieved between chemical

*Abbreviations:* CNN, computational neural networks; DHFR, dihydrofolate reductase; LDA, linear discriminant analysis; QSAR, quantitative structure–activity relationships

\* Corresponding author. Tel.: +1-814-865-3739; fax: +1-814-865-3314.  
E-mail address: pcj@psu.edu (P.C. Jurs).

structure and drug activity for various biologically relevant data sets [9–12]. Furthermore, classification techniques allow for models to be built that can discriminate between compounds belonging to different activity classes (e.g. strong/weak inhibitor, high/low selectivity) [9,13,14].

In the current literature, a variety of models for predicting DHFR inhibition are reported [15–46]. Most use homologous data sets pertaining to specific classes (e.g. quinazolines, pyrimidines) of DHFR inhibitors to build the models. Also, very few studies have been performed with inhibition data from the *P. carinii* and *T. gondii* forms of DHFR. The goal of this work is to build robust QSAR models to predict the inhibition values for a larger, more diverse DHFR data set, which is comprised of different classes of inhibitors with some relatively newly developed DHFR inhibitors. In addition, the challenging and rarely explored problem of predicting the selectivity of the inhibitors will be addressed through the development of classification models to label compounds as selective or nonselective inhibitors of bacterial DHFR. Furthermore, it has been proposed by several researchers that specific structural features of the compounds play a key role during the inhibition process. Therefore, new descriptor routines were developed especially for this study to better encode these types of interactions. It was believed that the information provided by these new study-specific descriptors would assist in producing more accurate QSAR models. The models presented in this work could be used as a possible screening tool to help uncover new lead compounds with potent DHFR inhibition with the added benefit of also labeling their selectivity.

## 2. Methodology

### 2.1. Data

All data used in this study were collected from literature articles by Rosowsky et al. [7,47–51], Gangjee et al. [5,52–65], Piper et al. [6], and Stevens et al. [8]. For each compound, they obtained  $IC_{50}$  values with inhibition assays done on DHFR derived from three different organisms: rat liver, *Pneumocystis carinii*, and *Toxoplasma gondii*. Where possible, a compound selectivity index ( $IC_{50(r)} / IC_{50(pc)}$ ,  $IC_{50(r)} / IC_{50(tg)}$ ) was also calculated for each compound. This ratio was also used by the above authors to assess the selectivity of drug molecules to inhibit *pcDHFR* and *tgDHFR* relative to *rlDHFR*. The reported  $IC_{50}$  and selectivity values for all 345 compounds can be found in Appendix A.

When using data obtained from different sources, it is important to ensure that experimental values are the result of consistent assay conditions.  $IC_{50}$  values for a particular compound can vary greatly when exposed to dissimilar experimental environments, thus jeopardizing the integrity of a QSAR model. All data used in this study were collected using the same experimental conditions for the DHFR in-

hibition assay and the same procedure for determining the  $IC_{50}$  value described in [66,67]. The protocols described in these two papers do not include methods to assess or report experimental errors associated with the observed  $IC_{50}$  values. The 23 papers cited above which provided the  $IC_{50}$  values likewise do not report experimental errors. The sole exception is that for four standard drug compounds (TMP, TMQ, PTX, PM) used for quality control experiments, where the authors report standard deviations with the  $IC_{50}$  values. These standard deviations range between 6 and 29% of the  $IC_{50}$  values [5–7,49,50].

For computational ease, the logarithm of the  $IC_{50}$  ( $\mu M$ ) value was used as the dependent variable to even the distribution of inhibition values that spanned many orders of magnitude. Log  $IC_{50}$  values ranged from  $-3.40$  to  $2.45$  (mean =  $-0.21$ ),  $-2.64$  to  $2.97$  (mean =  $0.15$ ), and  $-3.07$  to  $2.10$  (mean =  $-0.70$ ), respectively, for *rlDHFR*, *pcDHFR*, and *tgDHFR*. For the classification studies, selectivity indices ranged from  $0.0061$  to  $29$  (mean =  $1.3$ ) and  $0.0629$  to  $319$  (mean =  $6.42$ ), respectively, for the *pcDHFR* and *tgDHFR* selectivity data. The molecular weight of the compounds ranged from  $192$  to  $484$  amu with an average of  $334$  amu.

### 2.2. Structure entry and optimization

Structures were drawn on a PC using the HyperChem software package (HyperCube Inc., Waterloo, ON). They were then transferred to a Unix platform and optimized using the semi-empirical molecular orbital package MOPAC [68]. The PM3 Hamiltonian [69] was employed to obtain low-energy, 3-D geometry-optimized structures. To extract appropriate charge information needed by various descriptor routines, the PM3-optimized structures were run through a single-pass AM1 [70] calculation. Two Hamiltonians are used due to the suitability of each in extracting pertinent information from molecular structure [71].

### 2.3. Set generation

The distribution of training, cross-validation, and prediction set compounds for the QSAR models are shown in Table 1. Cross-validation and prediction set compounds were chosen pseudo-randomly by separating compounds into bins along the range of log  $IC_{50}$  values. Compounds were then chosen randomly from each bin and placed into cross-validation and prediction sets. The remaining

Table 1  
Distribution of compounds into training, cross-validation, and prediction sets for the three QSAR studies presented in this article

QSAR	Sets		
	Training	Cross-validation	Prediction
rlDHFR	224	58	57
pcDHFR	231	51	51
tgDHFR	244	38	38

compounds were used as the training set. This was done to ensure that the ranges of the dependent variables for all sets were approximately the same.

For classification, nine models were developed for each study using a leave-*N*-out procedure. The data set was split into three groups of equal size with one group being withheld to serve as an external prediction set while the remaining compounds were used for model development. The groups were systematically shuffled until each group served as a prediction set once. The entire procedure was repeated three times forming a total of nine models where each compound appeared in three prediction sets. This was done to examine the robustness of the models and to ensure that results were not a consequence of one particular training and prediction set split.

#### 2.4. Descriptor generation

Various types of descriptors were calculated using the Automated Data Analysis and Pattern Recognition Toolkit (ADAPT) [72,73] to describe topological, geometric, and electronic aspects of molecular structure. After descriptor generation, subsets of descriptors are examined to form predictive models using two computational methodologies (computational neural networks [74] and linear discriminant analysis [75]). Topological descriptors extract information ranging from simple atom/bond counts to more detailed data derived from graph-invariant techniques [76–82]. Geometric descriptors can capture structural information about overall 3-D molecular size, surface area and volume [83–85]. Utilization of the AM1 Hamiltonian allowed for charge information (e.g. partial charges, HOMO–LUMO energies, dipole moment) to be extracted from MOPAC output and used as descriptors for each compound. Several descriptor routines combined information about surface area and charge to form hybrid descriptors that encode the potential for polar intermolecular or hydrogen bond interactions that can occur with each molecule. Examples include the charged partial surface area (CPSA) [86] and hydrogen bond [87,88] descriptors along with atom-specific CPSA routines that combine surface area and charge information for individual atom types (e.g. N, O, S, halogens).

In addition to the ADAPT routines, constitutional [89], functional group, empirical [89], and molar refractivity [90,91] descriptors were calculated using the DRAGON [92] software package. Log *P* values were obtained from the website, <http://www.logp.com>, using the AlogPS model developed by Tetko et al. [93].

It is believed that both of the amino group nitrogen atoms on the base 2,4-diaminopyrimidine ring (Fig. 1) play a key role in the inhibition process [94,95]. Based on this, specific descriptor routines were developed and used to calculate the solvent-accessible surface areas and partial charges for the two amino nitrogen atoms. In addition, hydrogen bonding has also been proposed as a key factor when considering binding aspects between the substrate and the inhibitors

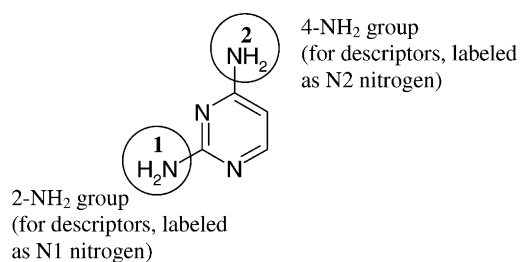


Fig. 1. Structure of the base ring appearing in 97% of the compounds. Around 3% of the compounds replace the 4-NH<sub>2</sub> group with a keto group. For those cases, partial charge and solvent-accessible surface area information was calculated using the keto groups.

[96]. A second set of descriptors was calculated which determined the maximum and minimum surface areas over all nitrogen and oxygen atoms separately. Partial charge information over nitrogen and oxygen atoms was also included to expand on this theory. The existing ADAPT routine which calculates molecular distance edge (MDE) [81] descriptors was modified to calculate MDE indices between nitrogen and oxygen atoms.

#### 2.5. Objective feature selection

Initially, over 200 descriptors were calculated for each compound in the data set. Reduction of the descriptor pool was performed to remove redundant or highly correlated descriptors. Identical tests were carried out to remove any descriptor that had the same value for over 90% of the training set compounds. Furthermore, if two descriptors were pairwise correlated above a value of 0.80, one was randomly removed. Since different training and prediction set splits were used to develop the classification models, the entire data set was used in the identical test and pairwise correlation procedures to filter out unwanted information. Descriptor reduction was done to ensure that the ratio between the number of descriptors to training set compounds was kept below a value of 0.6, thereby decreasing the risk of chance correlations during model development [97]. The reduced descriptor pools used to develop the models reported in this work contained 82–84 descriptors.

#### 2.6. Model development and subjective feature selection

Descriptor generation, CNN, and LDA analyses were performed on a DEC 3000 AXP Model 500 workstation using the ADAPT software. Predictive models were developed using quantitative and classification algorithms combined with evolutionary optimization routines to search the descriptor space for subsets that would accurately predict or classify activity. Simulated annealing (SA) [98] and genetic algorithm (GA) [99] optimization algorithms, written in-house, were used to find information-rich subsets of descriptors

that possessed superior predictive ability. For CNN models, the optimal model was chosen as that which minimized a cost function. Classification models were deemed optimal for those that minimized the number of misclassifications for the training set. These models were then validated for predictive ability using the external prediction set compounds.

## 2.7. QSAR

### 2.7.1. Feed-forward neural networks

Detailed descriptions of the optimization procedures and methods used for developing fully connected, feed-forward, computational neural network models have been published extensively in the literature [9,74,98,100]. Nonlinear transformations performed in the CNN analysis are done employing the sigmoid function shown in (1).

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (1)$$

Nonlinear feature selection was carried out by surveying the descriptor space with a 10-3-1 CNN accompanied by a GA optimizer to find optimal subsets of descriptors that minimized the cost function shown in (2).

$$\text{CNN cost} = \text{TSET}_{\text{rms}} + 0.4 |\text{TSET}_{\text{rms}} - \text{CVSET}_{\text{rms}}| \quad (2)$$

While obtaining low training set errors is the ultimate goal in QSAR modeling, the cost function (2) preserves the ability of a model to generalize by penalizing models that have large cross-validation set errors. After feature selection, a full CNN optimization (e.g. adjusting the number of hidden layer neurons, CNN committees) is performed on the best models to form final predictive models.

## 2.8. Classification

### 2.8.1. Linear discriminant analysis

Classification models were developed by using LDA combined with a GA optimization routine. LDA represents compounds in descriptor space using a weighted summation of their descriptor values in a discriminant function. An example of a discriminant function is shown in (3)

$$D = \mu_0 + \mu_1 x_1 + \mu_2 x_2 + \dots + \mu_p x_p \quad (3)$$

where  $\mu_1, \dots, \mu_p$  are discriminant coefficients (or weights) for each descriptor and  $x_1, \dots, x_p$  are the descriptor values for each compound. The offset from the origin is represented as  $\mu_0$ . The goal is to find a function which best separates the two classes but which also maintains a low within-class variance. In other words, a discriminant is desired which maximizes the distance between class clusters while trying to keep the clusters as small as possible. An optimal cut-off score is determined which ultimately decides class membership. Compounds exceeding the cut-off will be labeled as class 1, while compounds falling below the cut-off will be labeled as class 2.

Subsets of 2–10 descriptors were examined to find those that minimized the cost shown in (4).

$$\text{LDA cost} = (100.0 - \text{TSET}\%) + 0.5 |\text{TSET}\% - \text{CVSET}\%| \quad (4)$$

High training set classification rates are desired but the cost function penalizes models with poor generalization through examination of the percent correct for the cross-validation set. A leave-10%-out cross-validated training procedure was used for model development. For each model, the cost value over all 10 runs are averaged and used to represent the fitness of the model. The descriptors comprising the top models are then used to form the final models by using the entire training set to calculate final discriminant values with the external prediction set being used for final model validation.

## 3. Results and discussion

### 3.1. QSAR

#### 3.1.1. Rat liver DHFR

Network trainings were performed on the best 10 models found with the GA search. For each model, CNN architectures ranging from 10-2-1 to 10-8-1 were examined. An averaged committee of 10 10-7-1 CNN trainings produced rms errors of 0.51 log IC<sub>50</sub> ( $r^2 = 0.829$ ), 0.61 log IC<sub>50</sub> ( $r^2 = 0.780$ ), and 0.64 log IC<sub>50</sub> ( $r^2 = 0.716$ ) for the training, cross-validation, and prediction sets, respectively. A calculated versus experimental plot of the best rIDHFR model is shown in Fig. 2. Compound **211** appears as a prediction set outlier and explains 16.2% of the prediction set variance. This compound (shown in Fig. 3) is structurally similar to the 17 compounds reported in [8]. However, from a holistic point of view, these compounds only represent ~5% of the data set and are generally dissimilar when compared to the other compounds in terms of their molecular configuration due to the N=N–N (R, R') moiety attached to the side benzyl group. Perhaps the under-representation for these types of structures in the training set could explain why the compound did not predict accurately when compared to other compounds that were more structurally similar to a majority of the training set compounds. The individual descriptor values for compound **211** were embedded within the range of training set descriptor values. However, removal of this outlier compound reduced the prediction set rms error to 0.59 log IC<sub>50</sub> ( $r^2 = 0.755$ ).

The descriptors comprising this model are shown in Table 2. Out of 10, 5 descriptors in this model are simple counts of structural features: double bonds (NDB-13), sp<sup>2</sup>-hybridized carbons attached to three other carbons (3SP2-1), sp<sup>3</sup>-hybridized carbons attached to two other carbons (2SP3-1), number of hydrogen bond acceptor atoms (CTAA-0) [87,88], and number of 10-membered rings

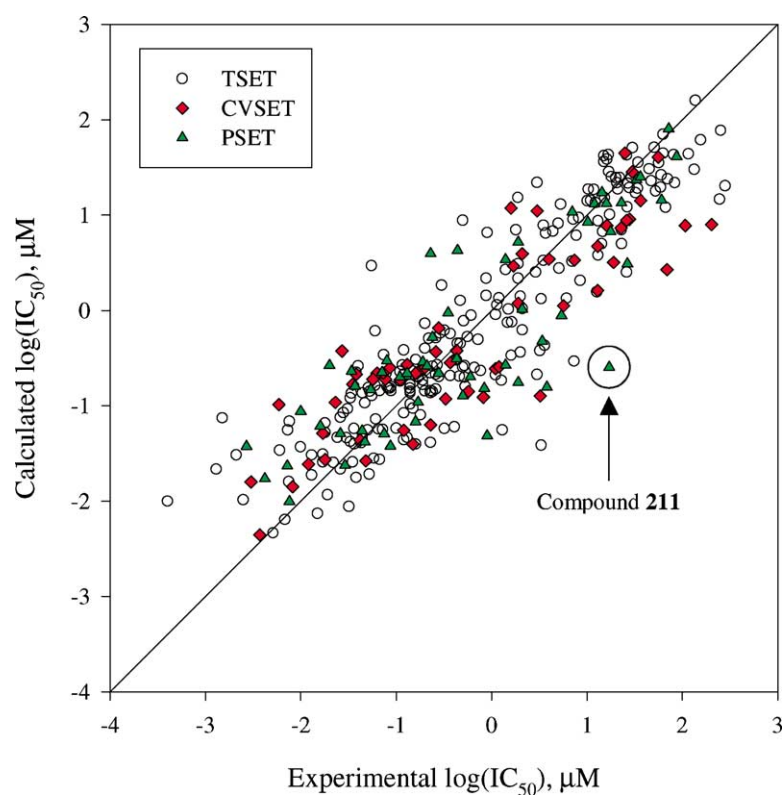
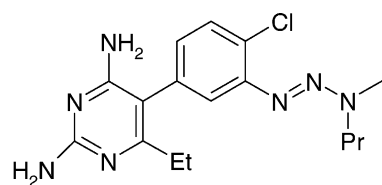
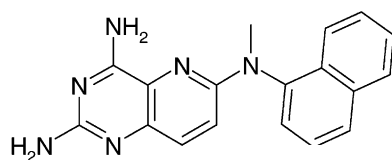


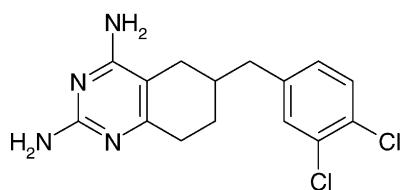
Fig. 2. Plot of calculated vs. experimental log  $IC_{50}$  values for the best rIDHFR model using a committee of 10 10-7-1 CNN trainings.



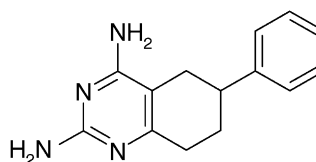
Compound **211**



Compound **46**



Compound **134**



Compound **137**

Fig. 3. Prediction set outliers for the three QSAR studies presented in this manuscript (Et, Pr = ethyl and propyl groups, respectively).



Table 2  
Descriptors used in CNN modeling for prediction of rIDHFR inhibition

Descriptor <sup>a</sup>	Type	Range
NDB-13	Topo.	0–7
3SP2-1	Topo.	0–7
2SP3-1	Topo.	0–6
MDEC-14	Topo.	0–28.1
CTAA-0	Topo. (H-bond)	4–12
nR10-42	Topo.	0–3
HARD-4	Elec.	3.4–4.6
DPSA-1	Hybrid (CPSA)	–62.9 to 439.2
RPCS-1	Hybrid (CPSA)	0–8.1
OSLO-4	Geom. (O-specific)	0–42.1

<sup>a</sup> NDB-13, number of non-aromatic double bonds; 3SP2-1, number of sp<sup>2</sup>-hybridized carbons attached to three other carbons; 2SP3-1, number of sp<sup>3</sup>-hybridized carbons attached to two other carbons; MDEC-14, molecular distance edge index between primary and quaternary carbons [81]; CTAA-0, number of hydrogen bond acceptor atoms (N, O, F, S) [87,88]; nR10-42, number of 10-membered rings [90,93]; HARD-4, [0.5(*E*<sub>HOMO</sub> – *E*<sub>LUMO</sub>)]; DPSA-1, difference in charged partial surface areas [86]; RPCS-1, relative positively charged surface area [86]; OSLO-4, surface area of the least solvent-accessible oxygen atom.

(nR10-42) [89]. Due to their relative ease of computation, descriptor counts are desirable for quickly screening large groups of drug candidates. Inclusion of NDB-13 (excluding aromatic bonds) suggests that interactions involving  $\pi$  electrons and the degree of molecular unsaturation are important parameters for predicting rIDHFR inhibition. The two ‘SP’ descriptors assess the relative degree of branching that originate from sp<sup>2</sup>-hybridized and sp<sup>3</sup>-hybridized carbons and can provide information about the overall topological size and branching of the molecules. The number of hydrogen bond acceptor atoms (CTAA-0) conveys a crude estimate for the potential of hydrogen bonding to occur in each molecule relative to other compounds in the data set. The number of 10-membered rings (nR10-42) can assess the degree of flexibility or rigidity for each molecule. Generally, molecules that contain more 10-membered rings will be less prone to bond rotations. Molecular distance edge information between primary and quaternary carbons is supplied by MDEC-14. MDE descriptors have been shown to provide good discrimination between structural isomers in a highly homologous data set [81]. The only purely electronic descriptor, HARD-4, describes the absolute hardness of each molecule. Calculated as one-half of the HOMO–LUMO energy gap, this descriptor encodes the relative chemical stability of the molecules. Many descriptors appear that encode information about polar interactions of an inhibitor [86]. DPSA-1 is the difference between the positively and negatively charged partial surface areas of each structure. RPCS-1 [86] represents the relative positively charged surface area of the molecules. This is calculated by multiplying the surface area of the most positively charged atom by the relative positive charge on the molecule (where *rpcg* is the charge on the most positive atom divided by the total charge summed over all positive atoms). These two descriptors suggest that intermolecular interactions that are

polar in nature are important when predicting rIDHFR inhibition. Finally, OSLO-4 is the surface area of the least accessible oxygen atom. This is placing importance on the solvent-accessible surface area of certain oxygen atoms in determining rIDHFR inhibition.

### 3.1.2. *Pneumocystis carinii* DHFR

The best 10 descriptor models found with the GA-search were evaluated using the CNN committee approach by sequentially increasing the number of hidden layer neurons until a minimum cost was achieved. The rms errors of 0.45 log IC<sub>50</sub> ( $r^2 = 0.839$ ), 0.49 IC<sub>50</sub> ( $r^2 = 0.785$ ), and 0.66 IC<sub>50</sub> ( $r^2 = 0.643$ ) for the training, cross-validation, and prediction sets, respectively, were obtained by averaging network predictions from ten CNN trainings. A calculated versus experimental plot of the best *pc*DHFR model is shown in Fig. 4. Note that one major prediction set outlier (compound 46) accounts for 22% of the prediction set variance. The chemical structure of compound 46 appears in Fig. 3. One notable difference between this compound and the rest of the data set is the presence of the naphthyl group. Even so, a fair representation (29 of 333 compounds, ~9%) of compounds containing naphthyl groups appear in the data set for *pc*DHFR so no conclusion can be drawn as to why this molecule did not predict well. The descriptor values for this molecule did not exceed the ranges of descriptor values for the training set. Some of the other naphthyl-containing structures predicted very well with low residual errors. By excluding this outlier, the rms error of the prediction set improves to 0.59 log IC<sub>50</sub> ( $r^2 = 0.713$ ).

The descriptors used to develop this model are shown in Table 3. NDB-13 and CTAA-0 reappear as descriptors

Table 3  
Descriptors used in CNN modeling for prediction of *pc*DHFR inhibition

Descriptor <sup>a</sup>	Type	Range
NDB-13	Topo.	0–4
CTAA-0	Topo. (H-bond)	4–12
V5C-10	Topo.	0.03–0.48
MDEC-13	Topo.	0–13.9
MDEC-23	Topo.	0–23.9
MDEC-24	Topo.	0–33.1
SCDH-2	Hybrid (H-bond)	3.3–6.8
CHAA-2	Hybrid (H-bond)	–0.39 to –0.14
CHAA-3	Hybrid (H-bond)	–0.006 to –0.0015
OSLO-4	Geom. (O-specific)	0–42.1

<sup>a</sup> NDB-13, number of non-aromatic double bonds; CTAA-0, number of hydrogen bond acceptor atoms (N, O, F, S) [87,88]; V5C-10, valence-corrected fifth order cluster  $\chi$  index [77,78]; MDEC-{13, 23, 24}, molecular distance edge index between primary/tertiary, secondary/tertiary, secondary/quaternary carbons, respectively [81]; SCDH-2, {sum of (solvent-accessible surface area  $\times$  partial charge) for all donatable H-bond hydrogen atoms/total number of donatable H-bond hydrogen atoms} [87,88]; CHAA-2, {sum of partial charges on H-bond acceptor atoms/total number of H-bond acceptor atoms} [87,88]; CHAA-3, {sum of partial charges on H-bond acceptor atoms/total molecular solvent-accessible surface area} [87,88]; OSLO-4, surface area of the least solvent-accessible oxygen atom.

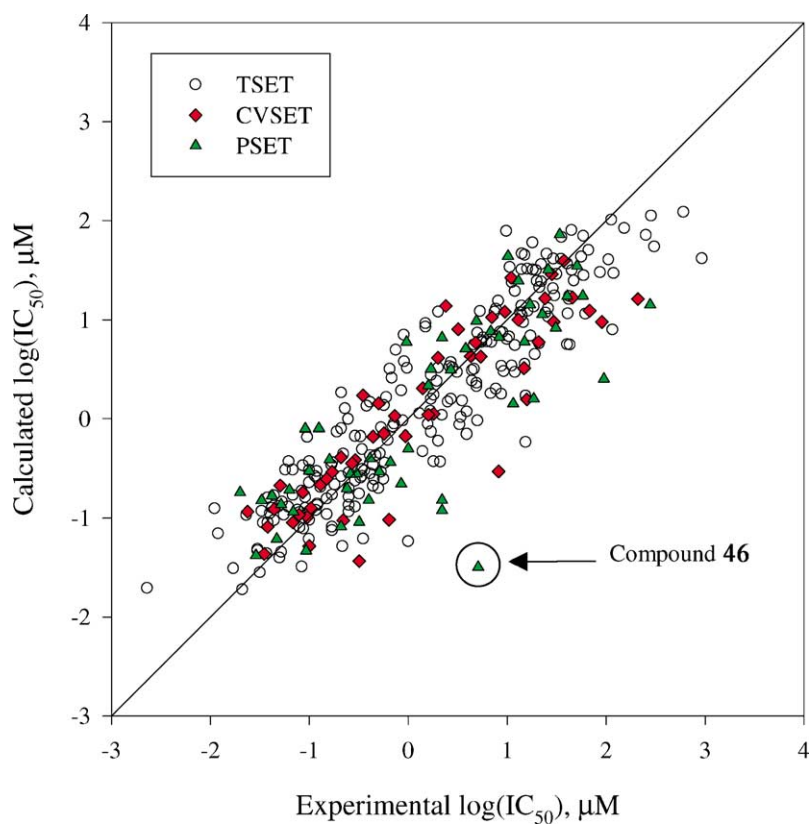


Fig. 4. Plot of calculated vs. experimental  $\log \text{IC}_{50}$  values for the best *pcDHFR* model using a committee of 10 10-6-1 CNN trainings.

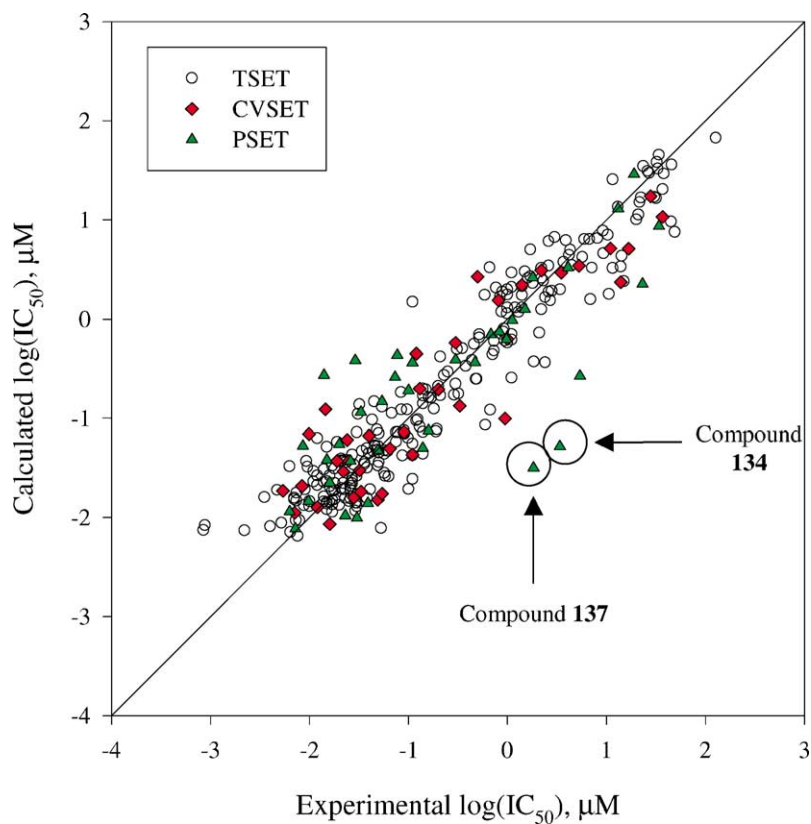


Fig. 5. Plot of calculated vs. experimental  $\log \text{IC}_{50}$  values for the best *tgDHFR* model using a committee of 10 10-7-1 CNN trainings.

in the best *pcDHFR* model. This implies that the number of non-aromatic double bonds and the number of hydrogen bond acceptor atoms are two influential parameters when predicting both *rlDHFR* and *pcDHFR* inhibition. V5C-10 [77,78] is a heteroatom and aromatic valence-corrected  $\chi$  index and encodes branching information between fifth order clusters embedded within each molecule. Essentially, this descriptor describes the relative degree of branching that can occur between clusters of six connectable heavy atoms in a molecule. Molecular distance edge information reappears in the descriptors MDEC-13, MDEC-23, and MDEC-24. This is not surprising due to the many structural isomers that appear in the data set. The inclusion of the MDE descriptors suggests that the ability to discriminate between structural isomers is important for *pcDHFR* inhibition prediction. Three hybrid descriptors are also present in this model that encode the potential for a molecule to undergo hydrogen bond interactions (SCDH-2, CHAA-2, and CHAA-3) [87,88]. SCDH-2 is the sum of the (surface area  $\times$  charge) information on donatable hydrogen atoms divided by the total number of donatable hydrogen atoms. CHAA-2 is the sum of partial charges on hydrogen bond acceptor atoms divided by the total number of hydrogen bond acceptor atoms while CHAA-3 is the sum of acceptor atom partial charges divided by the total molecular solvent-accessible surface area for each molecule. Finally, OSLO-4 resurfaces in the best model for *pcDHFR* and proposes that the surface area of the least accessible oxygen atom is significant in regards to both *rlDHFR* and *pcDHFR* inhibition.

### 3.1.3. *Toxoplasma gondii* DHFR

The best models found for *tgDHFR* were evaluated as described above. The best model was developed using a committee of 10 10-7-1 CNN trainings that produced a training set rms error of 0.33 log IC<sub>50</sub> ( $r^2 = 0.917$ ), cross-validation set rms error of 0.43 log IC<sub>50</sub> ( $r^2 = 0.852$ ), and prediction set rms error of 0.66 log IC<sub>50</sub> ( $r^2 = 0.626$ ). A calculated versus experimental plot of the best *tgDHFR* model is shown in Fig. 5. Two major outliers account for nearly 40% of the prediction set variance, compounds **134** and **137**, shown in Fig. 3. Descriptor values for the two compounds fell in the range of training set descriptor values, and both compounds are structurally similar to a majority of the training set members. Yet, when they were removed from the prediction set, the rms error dropped to 0.52 log IC<sub>50</sub> ( $r^2 = 0.772$ ).

The descriptors used to generate this model appear in Table 4. V5C-10 and MDEC-23 reappear as descriptors in this model, conveying they are important parameters when predicting both *pcDHFR* and *tgDHFR*. Another  $\chi$  index appears in N6CH-16 that counts the number of sixth order molecular chains. Specific to this data set, molecular configurations that would produce one count for N6CH-16 are sole six-membered rings and five-membered rings with one substituent—with each substituent contributing one

Table 4

Descriptors used in CNN modeling for prediction of *tgDHFR* inhibition

Descriptors <sup>a</sup>	Type	Range
N6CH-16	Topo.	2–10
2SP3-1	Topo.	0–6
V5C-10	Topo.	0.03–0.48
MDEC-23	Topo.	0–23.9
EDIF-1	Topo.	5.5–17.6
LOGP-1	Topo.	−0.12 to 5.0
HOMO-1	Elec.	−9.2 to −7.5
NCHI-1	Elec. (N-specific)	−0.29 to −0.04
NSHI-3	Geom. (N-specific)	15.0–25.5
OSHI-3	Geom. (O-specific)	0–46.1

<sup>a</sup> N6CH-16, number of sixth order chains  $\chi$  index [77,78]; 2SP3-1, number of sp<sup>3</sup>-hybridized carbons attached to two other carbons; V5C-10, valence-corrected fifth order cluster  $\chi$  index [77,78]; MDEC-23, molecular distance edge index between secondary and tertiary carbons [81]; EDIF-1, (max. E-state−min. E-state) [79]; LOGP-1, log of *n*-octanol/water partition coefficient [94]; HOMO-1, energy of the highest occupied molecular orbital; NCHI-1, charge on least negatively charged nitrogen atom; NSHI-3, surface area of the most solvent-accessible nitrogen atom; OSHI-3, surface area of the most solvent-accessible oxygen atom.

count. 2SP3-1 is also present in both models for *rlDHFR* and *tgDHFR* prediction. This suggests that branching information originating from sp<sup>3</sup>-hybridized carbons plays a key role for predicting both types of DHFR. EDIF-1 is the difference between the maximum electrotopological state (E-state) value and the minimum E-state value for each molecule. E-state [79] indices combine topological and electronic information to assess the degree of interaction an atom will have with a neighboring atom or molecule. Lower E-state values correspond to atoms that are buried within the molecular structure and that contain low electronegativity values. HOMO-1 is the energy of the highest occupied molecular orbital and is a measure of the reactivity for each compound. Molecules with high HOMO energies can more readily donate electrons when involved in chemical reactions. The logarithm of the *n*-octanol/water partition coefficient is represented by LOGP-1 [93]. The degree of hydrophobicity has been shown to be a useful parameter in many QSAR studies because of the key role it plays in estimating drug solubility.

In addition, descriptors that include specific nitrogen and oxygen information appear in this model: NCHI-1, NSHI-3, and OSHI-3. The charge on the least negatively charged nitrogen atom is supplied by NCHI-1. NSHI-3 and OSHI-3 both account for the solvent-accessible surface areas of the most accessible nitrogen and oxygen atom, respectively. The inclusion of NCHI-1 and NSHI-3 suggest that the partial charges and solvent-accessible surface areas of certain nitrogen atoms play in key role in prediction of *tgDHFR* inhibition. Furthermore, specific information about the surface areas of the most accessible oxygen atoms is labeled as an important parameter for *tgDHFR* inhibition prediction.



Table 5  
Distribution of compounds into training, cross-validation, and prediction sets for the *pcDHFR* and *tgDHFR* two-class problems

	Training	Prediction
<i>pcDHFR</i>		
Selective, $IC_{50(r)} / IC_{50(pc)} > 1$	42	21
Nonselective, $IC_{50(r)} / IC_{50(pc)} = 1$	178	90
<i>tgDHFR</i>		
Selective, $IC_{50(r)} / IC_{50(tg)} > 1$	167	84
Nonselective, $IC_{50(r)} / IC_{50(tg)} = 1$	44	22

Table 6  
Results for the nine models built using the *pcDHFR* selectivity data

Model	Training set (%)	Prediction set (%)
1	72.3	72.1
2	77.4	74.6
3	76.0	72.7
4	77.3	78.4
5	78.3	75.5
6	73.8	78.2
7	80.0	73.0
8	78.3	74.6
9	72.0	74.6
Average	76.1	74.8
Standard deviation	2.9	2.2

### 3.2. Classification

For both classification studies, compounds were considered nonselective for selectivity indices less than or equal to 1.0 and selective if their selectivity indices exceeded 1.0. The distribution of class membership in the training and prediction sets for the nine models is shown in Table 5.

#### 3.2.1. *pcDHFR* relative to *rlDHFR*

The training and prediction set classification rates for all nine models for *pcDHFR* are shown in Table 6. Note the low standard deviations for both training and prediction sets

shows that consistent results could be obtained independent of the different training and prediction set splits. The sizes of the nine models ranged from 4 to 10 descriptors with each compound appearing in a total of three prediction sets. A “majority rules” averaging procedure was performed to produce an average prediction for each compound. This revealed that 76.7% (254 of 331 compounds) of the data set was correctly classified when using the prediction averaging scheme. Several different classes of descriptors appeared frequently in the nine models. In particular, the MDE, molecular fragment, and nitrogen/oxygen-specific classes proved to be very useful in discriminating between nonselective and selective inhibitors of *pcDHFR*.

The best model contained the six descriptors shown in Table 7. This model was chosen by examining the training set classification rate but by also looking at the number of false negative misclassifications for each model. One of the important goals in developing DHFR inhibitors is to find drugs that are *selective* inhibitors of *pcDHFR* and not *rlDHFR*. Keeping this in mind, models with marginally better training rates were found, but the importance of correctly labeling selective compounds took precedence. The molecular distance edge class of descriptors is represented by MDEC-13. The number of 10-membered rings (nR10-42) appears as a useful molecular fragment descriptor. Furthermore, one of the nitrogen/oxygen-specific descriptors is present as the solvent-accessible surface area of the 4-NH<sub>2</sub> nitrogen atom (N2) on the base ring. The model produced a training set classification rate of 77.3% and a prediction set classification rate of 78.4%. A confusion matrix for this model is shown in Table 8.

Further examination of the magnitudes of the discriminant coefficients (weights) revealed that nR10-42 and NSUR-2 were the most influential descriptors to appear in the model. The general trend was that selective compounds tended to have lower N2 nitrogen surface areas and lower counts for the number of 10-membered rings. This suggests that steric considerations with the N2 nitrogen are important when determining the selectivity. Lowering the solvent-accessibility of the nitrogen atom acts in a favorably way with regards

Table 7  
Descriptors used in the best model to solve the two-class problem for *pcDHFR* selectivity

Descriptor <sup>a</sup>	Type	Discrim. coeff.	Training set range		Training set average	
			Selective	Nonselective	Selective	Nonselective
3SP2-1	Topo.	0.0054	0–5	0–7	2.6	1.8
V5P-6	Topo.	0.00075	1.3–3.4	1.0–2.9	1.9	1.9
MDEC-13	Topo.	−0.0035	0–11	0–14	2.7	3.2
nR10-42	Topo.	−0.0071	0–3	0–3	0.64	0.92
SCAA-3	Hybrid (H-bond)	0.0019	−0.077–0.0092	−0.15 to −0.0021	−0.038	−0.042
NSUR-2	Geom. (N-specific)	−0.0083	5.4–16.4	2.3–37.5	13.4	15.2

<sup>a</sup> 3SP2-1, number of sp<sup>2</sup>-hybridized carbons attached to three other carbons; V5P-6, valence-corrected fifth order path  $\chi$  index [77,78]; MDEC-13, molecular distance edge index between primary and tertiary carbons [81]; nR10-42, number of 10-membered rings [90]; SCAA-3, (sum of (surface area  $\times$  charge) on H-bond acceptor atoms)/(total molecular surface area) [87,88]; NSUR-2, solvent-accessible surface area for 4-NH<sub>2</sub> nitrogen atom on base ring.

Table 8  
Confusion matrix for the optimal *pcDHFR* selectivity model

Actual class	Selective	Nonselective	Class (%)
Training set ( <i>calculated class</i> )			
Selective	33	9	78.6
Nonselective	41	137	77.0
Overall = 77.3%			
Prediction set ( <i>predicted class</i> )			
Selective	17	4	81.0
Nonselective	20	70	77.8
Overall = 78.4%			

Table 9  
Results for the nine models built using the *pcDHFR* selectivity data

Model	Training set (%)	Prediction set (%)
1	72.0	75.5
2	72.0	68.9
3	73.1	71.4
4	76.3	67.0
5	73.5	72.6
6	76.4	74.3
7	66.4	69.8
8	73.5	62.3
9	66.5	70.5
Average	72.2	70.3
Standard deviation	3.6	4.0

to the selectivity of the molecules. Furthermore, molecules possessing high counts of 10-membered rings usually contained bulkier side groups as the side ring structure for the molecules. This suggests that having bulky side groups with high counts of 10-membered rings produces a detrimental effect on the selectivity of inhibitors for *pcDHFR*. These two trends are supported by the fact that misclassified compounds in the prediction set contained descriptor values, which deviated from the trend described above.

Table 10  
Descriptors used in the best model to solve the two-class problem for *tgDHFR* selectivity

Descriptor <sup>a</sup>	Type	Discrim. coeff.	Training set range		Training set average	
			Selective	Nonselective	Selective	Nonselective
N5C-14	Topo.	0.0043	2–11	2–12	4.2	5.0
MOLC-9	Topo.	−0.003	1.5–2.3	1.4–2.2	1.8	1.8
NN-4	Topo.	0.0033	4–7	5–7	5.6	5.8
nR06-39	Topo.	−0.0048	2–5	2–4	2.9	3.0
MDEC-13	Topo.	−0.0032	0–13.9	0–7.95	3.3	3.1
LUMO-2	Elec.	−0.013	−1.1 to 0.41	−1.6 to −0.17	−0.46	−0.66
NCHI-1	Elec. (N-specific)	−0.0047	−0.29 to −0.04	−0.29 to −0.04	−0.162	−0.156
NSUR-1	Geom. (N-specific)	−0.0033	13.5–18.1	13.4–17.9	16.4	16.2
NSHI-3	Geom. (N-specific)	−0.0034	15.4–23.6	15.4–25.5	19.7	19.2

<sup>a</sup> N5C-14, number of fifth order clusters c index [77,78]; MOLC-9, average distance sum connectivity (topological index J) [76]; NN-4, number of nitrogen atoms; nR06-39, number of 6-membered rings [90]; MDEC-13, molecular distance edge index between primary and tertiary carbon atoms [81]; LUMO-2, energy of the lowest unoccupied molecular orbital; NCHI-1, charge on the least negatively charged nitrogen atom; NSUR-1, surface area of the 2-NH<sub>2</sub> nitrogen atom on base ring; NSHI-3, surface area of the most accessible nitrogen atom.

Table 11  
Confusion matrix for the best *tgDHFR* selectivity model

Actual class	Selective	Nonselective	Class (%)
Training set ( <i>calculated class</i> )			
Selective	131	37	78.0
Nonselective	13	31	70.5
Overall = 76.4%			
Prediction set ( <i>predicted class</i> )			
Selective	64	19	77.1
Nonselective	8	14	63.6
Overall = 74.3%			

### 3.2.2. *tgDHFR* relative to *rlDHFR*

The results for the nine *tgDHFR* classification models are shown in Table 9. Unfortunately, the consistency of the models was lower than the models built with *pcDHFR* selectivity data, but the models still produced average prediction set rates with ~70% accuracy. The sizes of the nine models ranged from 3 to 10 descriptors where each compound appeared as a prediction set member three times. A “majority rules” averaging procedure revealed that 234 of the 317 compounds (73.8%) were predicted correctly. Again, descriptor classes that appeared frequently in the models were the MDE, molecular fragment, and nitrogen/oxygen-specific descriptors.

The best model contained the nine descriptors shown in Table 10. A confusion matrix for this model is shown in Table 11. Through examination of the discriminant magnitudes, the most influential descriptor was LUMO-2 (energy of the lowest unoccupied molecular orbital). Unfortunately, this descriptor is less structurally intuitive than the two most influential descriptors for the best *pcDHFR* model. However, the general trend was that selective inhibitors had higher LUMO energies than their nonselective counterparts. Again, it was seen that prediction set misclassifications were a result of compounds with descriptor values that differed from the general trend seen with the LUMO-2 descriptor.

Table 12

Root-mean-square errors and classification rates for models built with the random dependent variables and random class labels

	Prediction set error
Quantitative	
rlDHFR	1.46 ( $r^2 = 0.001$ )
pcDHFR	1.23 ( $r^2 = 0.108$ )
tgDHFR	1.47 ( $r^2 = 0.041$ )
Classification (%)	
pcDHFR <sup>a</sup>	55.9 ± 5.5
tgDHFR <sup>a</sup>	57.4 ± 4.5

<sup>a</sup> These results were obtained by averaging five class label scramblings for each of the nine models.

### 3.3. Random testing

To further ensure that the regression and classification models were built with a low probability of chance effects, randomization experiments were performed. For regression models, the dependent variable ( $\log IC_{50}$ ) was randomly scrambled and used in the subjective feature selection routines to find descriptor subsets that minimized the CNN cost function. For all random models, experimental parameters were set equal to those of the real models for comparative purposes. For classification models, class labels were scrambled within each set five times and used in the ensuing classification analyses. The results from the random classification models were averaged and compared to the results obtained with the real class labels. If the QSAR or classification models built with the real dependent variable convey true structure–activity relationships, then the models built randomly should be very poor in terms of the prediction set error (and correlation coefficients for quantitative models). The rms errors and classification rates for the random models are shown in Table 12. For QSAR, the poor rms errors and  $r^2$ -values for the prediction set indicate that the real models were developed with a low probability of chance correlations. Since consistently better models were obtained when using the real dependent variable versus the random class labels, it was concluded that the real models contained a low probability of chance effects during model development.

## 4. Conclusions

QSAR and classification models were developed to predict DHFR inhibition and selectivity with a high degree of accuracy using CNN and LDA methodologies. Summaries of the results obtained for the three QSAR and two classification models are shown in Table 13. CNN models were

Table 13

Summary of results obtained for the three QSAR and two classification models presented in this study

	Training	Cross-validation	Prediction
RMSE			
rlDHFR	0.51 ( $r^2 = 0.829$ )	0.61 ( $r^2 = 0.780$ )	0.64 ( $r^2 = 0.716$ )
pcDHFR	0.45 ( $r^2 = 0.839$ )	0.49 ( $r^2 = 0.785$ )	0.66 ( $r^2 = 0.643$ )
tgDHFR	0.33 ( $r^2 = 0.917$ )	0.43 ( $r^2 = 0.852$ )	0.66 ( $r^2 = 0.626$ )
Rate (%)			
pcDHFR	77.3	–	78.4
tgDHFR	76.4	–	74.3

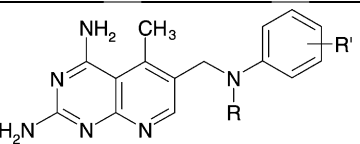
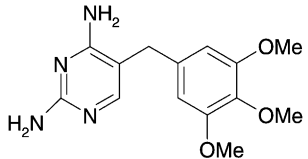
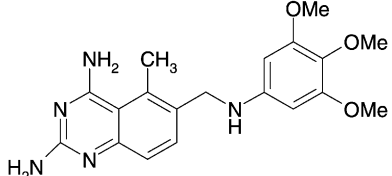
built to predict inhibition values for the three types of DHFR to within  $\sim 0.65 \log IC_{50}$  units from the experimentally determined values. Furthermore, classification models that label compounds as selective or nonselective inhibitors of pcDHFR and tgDHFR relative to rlDHFR were constructed with  $\sim 70\%$  accuracy. A leave- $N$ -out approach was applied to build models with different training and prediction set splits to add robustness to the models and to ensure that consistent results could be obtained with different training sets. In addition, it was seen that the new descriptors developed especially for this study appeared in the best models for quantitation and classification proving that the information provided by these descriptors were important to model inhibitor activity. For both regression and classification, randomization experiments were performed that provided evidence that the original models were likely not due to chance effects.

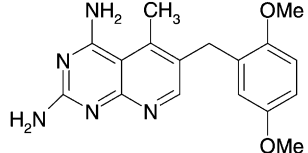
One precaution needs to be noted in that only structurally similar compounds can be predicted with the models presented above, due to the limited structural diversity of the data set used to train the models. The authors would also like to note that although the models provide limited insight into the design of new DHFR inhibitors, they could be applied to screen large groups of drug candidates to identify possible lead inhibitors of DHFR. Keeping in mind the overall drug development scheme for new DHFR inhibitors, these QSAR models could aid in revealing potentially potent inhibitors with the added advantage of labeling their selectivity relative to rlDHFR.

## Appendix A

Compounds used in this study accompanied by their corresponding  $IC_{50}$  inhibition values towards *Pneumocystis carinii* DHFR, *Toxoplasma gondii* DHFR, and rat liver DHFR, and selectivity indexes for pcDHFR and tgDHFR relative to rlDHFR are shown in Table A.1.

Table A.1  
Supplementary material showing the selectivity values for the 345 compounds

								
Dan	R	R'	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
1	H	3,4,5-triOCH <sub>3</sub>	0.086	0.0021	0.024	0.0074	0.28	[5]
2	CH <sub>3</sub>	3,4,5-triOCH <sub>3</sub>	0.013	0.0076	0.58	0.00085	8.9	[5]
3	H	3,4,5-triCl	0.063	0.033	0.52	0.012	2.8	[5]
4	CH <sub>3</sub>	3,4,5-triCl	0.1045	0.0363	0.35	0.0381	0.95	[5]
5	H	3,4-diOCH <sub>3</sub>	0.044	0.0076	0.17	0.0088	0.86	[5]
6	CH <sub>3</sub>	3,4-diOCH <sub>3</sub>	0.32	0.044	0.14	0.029	1.5	[5]
7	CH <sub>3</sub>	2,5-diOCH <sub>3</sub>	0.216	0.407	1.9	0.0301	14	[5]
8	H	3,4-diCl	0.32	0.053	0.17	0.028	1.9	[5]
9	CH <sub>3</sub>	3,4-diCl	0.1	0.042	0.42	0.027	1.6	[5]
10	H	H	0.08	0.17	2.1	0.017	10	[5]
11	H	2-OCH <sub>3</sub>	0.117	0.169	1.4	0.023	7.3	[5]
12	H	3-OCH <sub>3</sub>	0.0689	0.0801	1.2	0.0074	11	[5]
13	H	4-OCH <sub>3</sub>	0.0954	0.0556	0.59	0.012	4.7	[5]
14	H	2-Cl	0.047	0.088	1.9	0.0071	12	[5]
15	H	3-Cl	0.023	0.037	1.6	0.011	3.4	[5]
16	H	4-Cl	0.0554	0.051	0.93	0.019	2.7	[5]
17	H	4-Br	0.0808	0.0349	0.43	0.0095	3.7	[5]
18	CH <sub>3</sub>	3-OCH <sub>3</sub>	0.03	0.018	0.6	0.0063	2.9	[5]
19	CH <sub>3</sub>	4-OCH <sub>3</sub>	0.035	0.013	0.37	0.0073	1.8	[5]
20	CH <sub>3</sub>	2-Cl	0.084	0.1	1.2	0.018	5.6	[5]
21	CH <sub>3</sub>	4-Cl	0.029	0.026	0.9	0.0054	4.8	[5]
22	CH <sub>3</sub>	4-Br	0.037	0.036	0.97	0.03	1.2	[5]
								
23 (TMP)			12	133	11	2.7	49	[5]
								
24 (TMQ)			0.042	0.003	0.071	0.01	0.3	[5]

25 (PTX)		0.038	0.0015	0.039	0.011	0.14	[5]
----------	---	-------	--------	-------	-------	------	-----

Dan	R	R'	pcDHFR	rlDHFR	rlpc	tgDHFR	rltg	Reference
26	S	2-OCH <sub>3</sub>	2.2	0.23	0.1	0.058	4	[61]
27	S	4-OCH <sub>3</sub>	0.7	0.075	0.11	0.045	1.7	[61]
28	S	3,4-diOCH <sub>3</sub>	0.086	0.018	0.21	0.019	0.95	[61]
29	SO <sub>2</sub>	2-OCH <sub>3</sub>	3.2	1.4	0.44	0.21	6.7	[61]
30	SO <sub>2</sub>	4-OCH <sub>3</sub>	10.5	2	0.19	1	2	[61]
31	SO <sub>2</sub>	3,4-diOCH <sub>3</sub>	2.7	0.88	0.33	0.94	0.94	[61]
32	NH	2-OCH <sub>3</sub>	8.7	0.26	0.03	0.46	0.57	[61]
33	NH	4-OCH <sub>3</sub>	90.4	3.8	0.04	2.8	1.4	[61]
34	NH	3,4-diOCH <sub>3</sub>	40.4	1.1	0.028	0.68	1.6	[61]
35	NCH <sub>3</sub>	4-OCH <sub>3</sub>	0.22	0.0068	0.031	0.0086	0.79	[61]
36	NCH <sub>3</sub>	3,4-diOCH <sub>3</sub>	0.0023	0.0004	0.17	0.00088	0.45	[61]
37	NH	2,5-diOCH <sub>3</sub>	16.1	3.6	0.22	0.73	4.9	[61]
38	NCH <sub>3</sub>	2,5-diOCH <sub>3</sub>	0.034	0.0042	0.12	0.041	0.1	[61]
39	NH	3,4,5-triOCH <sub>3</sub>	25.9	3.2	0.12	2.4	1.3	[61]
40	NCH <sub>3</sub>	3,4,5-triOCH <sub>3</sub>	0.021	0.0037	0.18	0.0076	0.49	[61]
41	NH	3,4-C <sub>4</sub> H <sub>4</sub>	15	2	0.13	1.1	1.8	[61]
42	NCH <sub>3</sub>	3,4-C <sub>4</sub> H <sub>4</sub>	0.0317	0.0099	0.31	0.0145	0.68	[61]
43	NH	H	8.3	0.43	0.052	0.3	1.4	[61]
44	NCH <sub>3</sub>	H	0.0995	0.0013	0.013	0.0022	0.59	[61]
45	NH	4-Cl	14.6	0.82	0.055	0.83	0.99	[61]
46	NCH <sub>3</sub>	2,3-C <sub>4</sub> H <sub>4</sub>	5.1	3.3	0.65	2.1	1.6	[61]
47	NHCH <sub>2</sub>	3,4,5-triOCH <sub>3</sub>	29.4	1.4	0.048	0.49	2.9	[61]

Dan	n	Ar	pcDHFR	rlDHFR	rlpc	tgDHFR	rltg	Reference
48	0	C <sub>6</sub> H <sub>5</sub>	153	59.9	0.39	28	2.1	[56]
49	0	3,4,5-OCH <sub>3</sub> C <sub>6</sub> H <sub>2</sub>	59.4	13	0.22	2.2	5.9	[56]
50	0	2,3,4-OCH <sub>3</sub> C <sub>6</sub> H <sub>2</sub>	17	3.7	0.22	1.8	2.1	[56]
51	0	2,4,6-OCH <sub>3</sub> C <sub>6</sub> H <sub>2</sub>	31.5	1.88	0.06	0.84	2.2	[56]
52	0	2,4,5-OCH <sub>3</sub> C <sub>6</sub> H <sub>2</sub>	22	7	0.32	1.5	4.7	[56]
53	0	2,5-OCH <sub>3</sub> C <sub>6</sub> H <sub>3</sub>	37.4	3.5	0.094	1.7	2.1	[56]

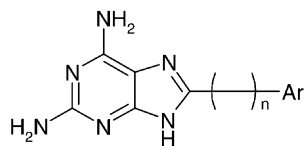
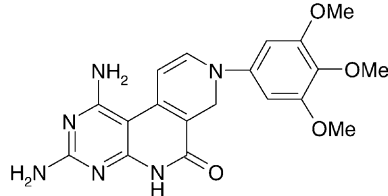
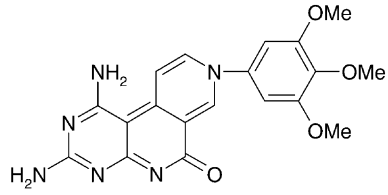
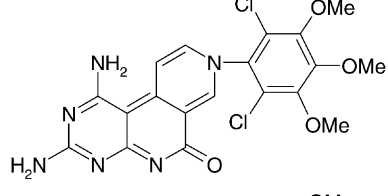
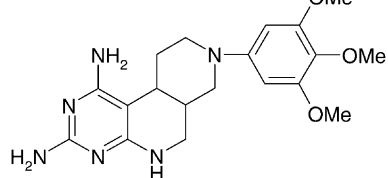
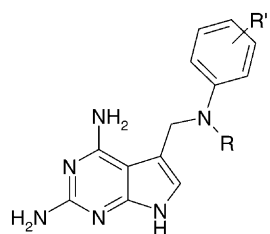


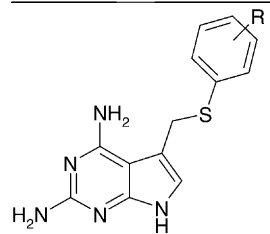


Table A.1 (Continued)

Dan	<i>n</i>	Ar	<i>pc</i> DHFR	rIDHFR	rl <i>pc</i>	<i>tg</i> DHFR	rl <i>tg</i>	Reference
54	0	3,5-OCH <sub>3</sub> C <sub>6</sub> H <sub>3</sub>	9	32.4	3.6	3	11	<a href="#">[56]</a>
55	0	3,4-OCH <sub>3</sub> C <sub>6</sub> H <sub>3</sub>	12	52.3	4.4	14.2	3.7	<a href="#">[56]</a>
56	0	2,4-OCH <sub>3</sub> C <sub>6</sub> H <sub>3</sub>	14	0.9	0.064	2.4	0.38	<a href="#">[56]</a>
57	0	3,4-Cl <sub>2</sub> C <sub>6</sub> H <sub>3</sub>	19.5	252	13	6.7	38	<a href="#">[56]</a>
58	0	2,6-Cl <sub>2</sub> C <sub>6</sub> H <sub>3</sub>	11	11	1	11	1	<a href="#">[56]</a>
59	0	3,5-OCH <sub>3</sub> -4-OHC <sub>6</sub> H <sub>2</sub>	68	15.3	0.23	1	15	<a href="#">[56]</a>
60	0	4-C <sub>6</sub> H <sub>5</sub> -C <sub>6</sub> H <sub>4</sub>	605	30.1	0.05	32.9	0.91	<a href="#">[56]</a>
61	0	4-C <sub>5</sub> H <sub>4</sub> N	18	18	1	20	0.9	<a href="#">[56]</a>
62	0	2-naphthyl	113	280	2.5	27	10	<a href="#">[56]</a>
63	0	4-OCH <sub>3</sub> -1-naphthyl	13	13	1	13	1	<a href="#">[56]</a>
64	0	9-fluorenyl	35	29.8	0.85	30.5	0.98	<a href="#">[56]</a>
65	1	C <sub>6</sub> H <sub>5</sub>	9.8	1.6	0.16	0.5	3.2	<a href="#">[56]</a>
66	1	NH-3,4,5-OCH <sub>3</sub> C <sub>6</sub> H <sub>2</sub>	34	34	1	34	1	<a href="#">[56]</a>
67	1	NH-2,5-OCH <sub>3</sub> C <sub>6</sub> H <sub>3</sub>	29	29	1	49	0.59	<a href="#">[56]</a>
68	1	S-2-naphthyl	105	107	1	45	2.4	<a href="#">[56]</a>
Dan			<i>pc</i> DHFR	rIDHFR	rl <i>pc</i>	<i>tg</i> DHFR	rl <i>tg</i>	Reference
69		4.28	0.24	0.06	1.12	0.21	<a href="#">[57]</a>	
70		2.4	0.086	0.036	0.52	0.17	<a href="#">[57]</a>	
71		–	0.5	–	2.3	0.22	<a href="#">[57]</a>	
72		2.2	0.16	0.073	0.29	0.55	<a href="#">[57]</a>	

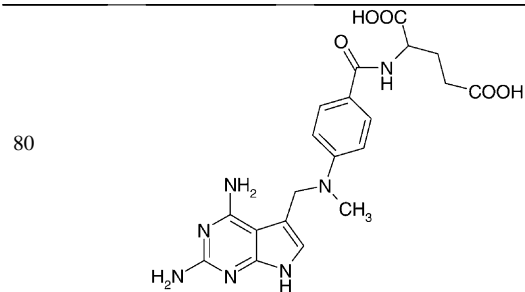


Dan	R	R'	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
73	CH <sub>3</sub>	2,5-diOCH <sub>3</sub>	—	—	—	3.4	—	[58]
74	CH <sub>3</sub>	3,4-diCl	28.3	3	0.11	1	3	[58]
75	CH <sub>3</sub>	2,3-(CH) <sub>4</sub>	209	8.2	0.039	0.87	9.4	[58]



Dan	R	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
76	3,4-diOCH <sub>3</sub>	11.1	16.7	1.5	2.6	6.4	[58]
77	3,4-diCl	58.5	5.3	0.091	11.6	0.46	[58]
78	2,3-(CH) <sub>4</sub>	10.6	3	0.28	0.81	3.7	[58]
79	3,4-(CH) <sub>4</sub>	929	82.9	0.089	9.2	9.0	[58]

Dan	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
-----	----------------	----------------	-------------	----------------	-------------	-----------



80	0.044	0.06	1.4	0.15	0.4	[58]
----	-------	------	-----	------	-----	------

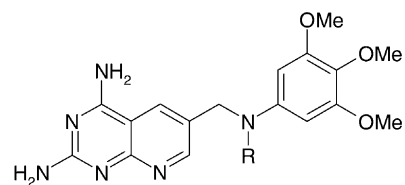
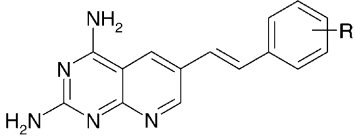


Table A.1 (Continued)

Dan	R	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
81	H	1.5	1.9	1.3	0.3	6.3	[59]
82	CH <sub>3</sub>	0.24	0.28	1.2	0.009	31	[59]
83	CH <sub>2</sub> CH <sub>3</sub>	0.19	0.12	0.63	0.049	3.2	[59]
84	CHO	18.5	7.4	0.4	1.1	6.7	[59]

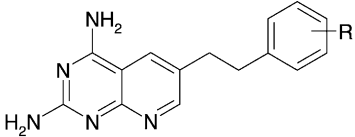
  



Dan	R	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
85	3,4,5-triOCH <sub>3</sub>	–	12.9	–	1.4	9.2	[59]
86	3,4-diOCH <sub>3</sub>	2.6	2.1	0.81	1.4	1.5	[59]
87	4-OCH <sub>3</sub>	5.3	11.8	2.2	1.5	7.9	[59]

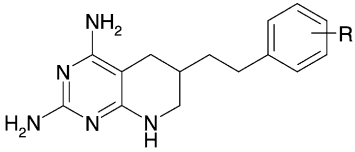
  



Dan	R	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
88	3,4,5-triOCH <sub>3</sub>	5	1.14	0.23	0.2	5.7	[59]
89	3,4-diOCH <sub>3</sub>	1.4	0.61	0.44	0.2	3	[59]
90	4-OCH <sub>3</sub>	0.29	0.26	0.9	0.25	1	[59]

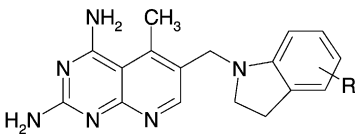
  



Dan	R	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
91	3,4,5-triOCH <sub>3</sub>	61.7	6.1	0.099	0.47	13	[59]
92	3,4-diOCH <sub>3</sub>	7.7	2.1	0.27	1.1	1.9	[59]

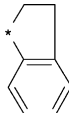
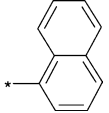
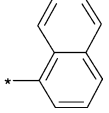
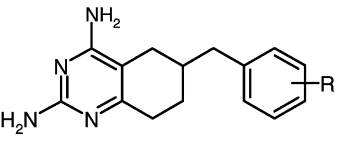
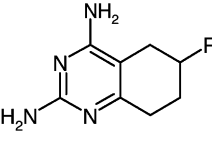


Dan	R	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
93	4-OCH <sub>3</sub>	0.29	0.15	0.52	0.048	3.1	[60]
94	5-OCH <sub>3</sub>	0.25	0.17	0.68	0.057	3	[60]
95	5,6-diOCH <sub>3</sub>	0.41	0.23	0.56	0.049	4.7	[60]

Dan		<i>pc</i> DHFR	rlDHFR	rl <i>pc</i>	<i>tg</i> DHFR	rl <i>tg</i>	Reference	
96		0.57	0.47	0.83	0.077	6.1	[60]	
97		0.85	0.13	0.15	0.11	1.2	[60]	
Dan	R		<i>pc</i> DHFR	rlDHFR	rl <i>pc</i>	<i>tg</i> DHFR	rl <i>tg</i>	Reference
98	2,5-diOCH <sub>3</sub> -4-pyrrolo		0.35	0.23	0.7	0.033	7	[60]
99	4,5-diOCH <sub>3</sub> -2-pyrrolo		1.8	3.5	1.9	0.6	5.8	[60]
100	2,3,5,6-tetraOCH <sub>3</sub> -4-pyrrolo		0.62	0.17	0.27	0.075	2.3	[60]
101	2-OCH <sub>3</sub> -5-phenyl		0.64	0.44	0.69	0.068	6.5	[60]
Dan	R	Ar	<i>pc</i> DHFR	rlDHFR	rl <i>pc</i>	<i>tg</i> DHFR	rl <i>tg</i>	Reference
102	H	3,4,5-(triCH <sub>3</sub> ) C <sub>6</sub> H <sub>2</sub>	4.6	0.29	0.063	0.054	5.4	[52]
103	CH <sub>3</sub>	3,4,5-(triCH <sub>3</sub> ) C <sub>6</sub> H <sub>2</sub>	0.095	0.038	0.4	0.007	5.4	[52]
104	CH <sub>2</sub> C≡CH	3,4,5-(triCH <sub>3</sub> ) C <sub>6</sub> H <sub>2</sub>	0.119	0.074	0.62	0.012	6.2	[52]
105	H	2,5-(diOCH <sub>3</sub> ) C <sub>6</sub> H <sub>3</sub>	1.67	0.56	0.34	0.181	3.1	[52]
106	CH <sub>3</sub>	2,5-(diOCH <sub>3</sub> ) C <sub>6</sub> H <sub>3</sub>	0.3	0.26	0.870	0.015	17	[52]
107	C <sub>2</sub> H <sub>5</sub>	2,5-(diOCH <sub>3</sub> ) C <sub>6</sub> H <sub>3</sub>	0.114	0.071	0.62	0.017	4.2	[52]
108	H	2,5-(diOCH <sub>2</sub> CH <sub>3</sub> ) C <sub>6</sub> H <sub>3</sub>	1.57	1.47	0.94	0.14	10.5	[52]
109	CH <sub>3</sub>	2,5-(diOCH <sub>2</sub> CH <sub>3</sub> ) C <sub>6</sub> H <sub>3</sub>	0.319	0.116	0.36	0.017	6.8	[52]
110	H	3,4-(diCl) C <sub>6</sub> H <sub>3</sub>	6.8	0.15	0.022	0.11	1.4	[52]
111	CH <sub>3</sub>	3,4-(diCl) C <sub>6</sub> H <sub>3</sub>	0.246	0.034	0.14	0.021	1.6	[52]
112	H	2,5-(diCl) C <sub>6</sub> H <sub>3</sub>	0.41	0.24	0.59	0.097	2.5	[52]
113	H	2,6-(diCl) C <sub>6</sub> H <sub>3</sub>	0.502	0.109	0.22	0.0099	11	[52]
114	H	4-Cl C <sub>6</sub> H <sub>4</sub>	0.94	0.128	0.14	0.078	1.6	[52]
115	CH <sub>3</sub>	4-Cl C <sub>6</sub> H <sub>4</sub>	0.171	0.067	0.39	0.022	3	[52]
116	H	3-Br C <sub>6</sub> H <sub>4</sub>	0.33	0.227	0.69	0.03	7.6	[52]
117	H	-CH <sub>2</sub> -3,4,5-(triOCH <sub>3</sub> ) C <sub>6</sub> H <sub>2</sub>	18.5	12.7	0.69	–	–	[52]
118	CH <sub>3</sub>	-CH <sub>2</sub> -3,4,5-(triOCH <sub>3</sub> ) C <sub>6</sub> H <sub>2</sub>	3.1	1.63	0.53	0.33	4.9	[52]

Table A.1 (Continued)

Dan	R	Ar	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
119		R, Ar = 	0.21	0.16	0.76	0.027	5.9	<a href="#">[52]</a>
120	H		0.517	0.139	0.27	0.036	3.9	<a href="#">[52]</a>
121	CH <sub>3</sub>		0.1	0.047	0.47	0.023	2	<a href="#">[52]</a>
								
Dan	R	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference	
122	H	0.29	0.18	0.62	0.032	5.6	<a href="#">[51]</a>	
123	2-CH	0.25	0.11	0.44	0.023	4.8	<a href="#">[51]</a>	
124	3-CH <sub>3</sub>	0.34	0.11	0.32	0.036	3.1	<a href="#">[51]</a>	
125	2-OCH <sub>3</sub>	0.45	0.12	0.27	0.014	8.6	<a href="#">[51]</a>	
126	3-OCH <sub>3</sub>	0.27	0.11	0.41	0.021	5.2	<a href="#">[51]</a>	
127	4-OCH <sub>3</sub>	0.44	0.077	0.18	0.05	1.5	<a href="#">[51]</a>	
128	3-CF <sub>3</sub>	0.40	0.19	0.48	0.14	1.4	<a href="#">[51]</a>	
129	3-OCF <sub>3</sub>	0.1	0.079	0.79	0.014	5.6	<a href="#">[51]</a>	
130	4-OCF <sub>3</sub>	0.58	0.17	0.29	0.073	2.3	<a href="#">[51]</a>	
131	2,5-diOCH <sub>3</sub>	0.057	0.034	0.6	0.021	1.6	<a href="#">[51]</a>	
132	3,4-diOCH <sub>3</sub>	0.1	0.063	0.63	0.023	2.7	<a href="#">[51]</a>	
133	3,4,5-triOCH <sub>3</sub>	0.091	0.038	0.42	0.024	1.6	<a href="#">[51]</a>	
134	3,4-diCl	15	7.3	0.49	2.6	2.8	<a href="#">[51]</a>	
								
Dan	R	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference	
135	Et	6.9	3	0.43	1.1	2.7	<a href="#">[51]</a>	
136	t-Bu	0.18	0.065	0.36	0.018	3.6	<a href="#">[51]</a>	
137	Ph	2.2	1.9	0.86	1.3	1.5	<a href="#">[51]</a>	



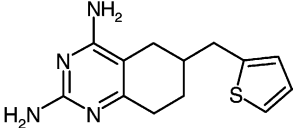
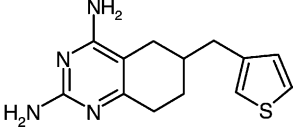
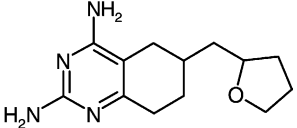
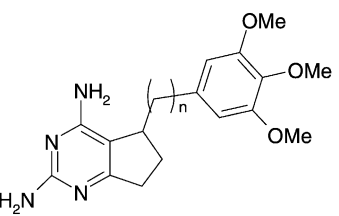
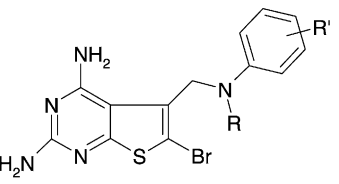
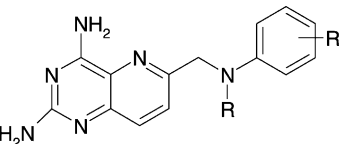
138		0.21	0.08	0.38	0.062	1.3	[51]
139		1.7	0.13	0.076	0.049	2.7	[51]
140		0.88	0.27	0.31	0.04	6.8	[51]
							
Dan	<i>n</i>	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
141	1	–	69	–	21	3.3	[50]
142	2	1.8	0.51	0.28	0.14	3.6	[50]
143	3	1.3	0.22	0.17	0.14	1.6	[50]
							
Dan	R	R'	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	Reference
144	H	3,4,5-triOCH <sub>3</sub>	13	17	1.3	34	0.5 [49]
145	H	2,5-diOCH <sub>3</sub>	–	33	–	–	[49]
146	CH <sub>3</sub>	3,4,5-triOCH <sub>3</sub>	31	28	0.9	127	0.22 [49]
147	H	3,5-diCl-4-[1-pyrrolo]	7.5	10	1.3	26	0.38 [49]
							

Table A.1 (Continued)

Dan	R	R'	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
148	CH <sub>3</sub>	3,4,5-triOCH <sub>3</sub>	0.13	0.026	0.2	0.0047	5.5	[48]
149	CH <sub>3</sub>	4-Cl	0.062	0.022	0.35	0.015	1.5	[48]
150	CH <sub>3</sub>	3-Cl	2.1	0.067	0.03	0.02	3.4	[48]

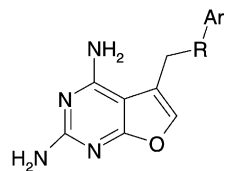
Dan	R	R'	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
151	H	2,5-diOCH <sub>3</sub>	0.051	0.044	0.86	0.03	1.5	[47]
152	H	3,4,5-triOCH <sub>3</sub>	0.033	0.0059	0.18	0.0052	1.1	[47]
153	CH <sub>3</sub>	3,4,5-triOCH <sub>3</sub>	0.012	0.012	1	0.0064	1.9	[47]

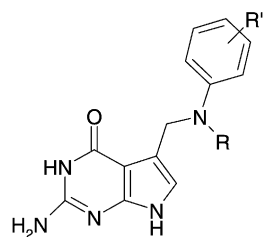
Dan	R	<i>n</i>	R'	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
154	H	1	2,5-diOCH <sub>3</sub>	0.053	0.028	0.53	0.017	1.6	[47]
155	H	2	2,5-diOCH <sub>3</sub>	12	2.1	0.18	0.99	2.1	[47]
156	H	3	2,5-diOCH <sub>3</sub>	26	1.9	0.073	1	1.9	[47]
157	H	1	3,4,5-triOCH <sub>3</sub>	0.033	0.006	0.18	0.007	0.86	[47]
158	H	2	3,4,5-triOCH <sub>3</sub>	43	5.4	0.13	2.6	2.1	[47]
159	H	3	3,4,5-triOCH <sub>3</sub>	95	19	0.2	14	1.4	[47]
160	CH <sub>3</sub>	1	3,4,5-triOCH <sub>3</sub>	0.17	0.038	0.22	0.016	2.4	[47]

Dan	R	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
161	CH <sub>2</sub> CH <sub>2</sub>	1.4	5.1	3.6	0.91	5.6	[7]
162	O	3.4	13	3.8	2.2	5.9	[7]
163	None	4.9	4	0.82	1.3	3.1	[7]
164	CH <sub>2</sub>	0.042	0.027	0.64	0.029	0.93	[7]
165	S	0.12	0.02	0.17	0.11	0.18	[7]
166	Direct	0.1	0.055	0.55	0.012	4.6	[7]
167	CH=CH	0.21	4.4	21	0.043	102	[7]



Dan	R	Ar	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
168	S	phenyl	–	252	–	–	–	[62]
169	S	1-naphthyl	19	23	1.2	19	1.2	[62]
170	S	2-naphthyl	0.65	12.3	19	11.6	1.1	[62]
171	NH	1-naphthyl	13.5	12	0.89	37	0.32	[62]
172	NH	2-naphthyl	41	36.5	0.89	38	0.96	[62]
173	O	2-naphthyl	14	60.3	4.3	–	–	[62]
174	NH	4-phenoxyPh	8.1	16.2	2	32.4	0.5	[62]
175	NH	2-phenylPh	7.7	137	18	45.4	3	[62]
176	N(CH <sub>3</sub> )	2-naphthyl	14.8	14.6	0.99	23.6	0.62	[62]
177	NH	2,5-diClPh	50.9	71.9	1.4	–	–	[62]
178	N(CH <sub>3</sub> )	3,4-diClPh	44.8	–	–	–	–	[62]
179	N(CH <sub>3</sub> )	3,4,5-triClPh	284	34.3	0.12	21.5	1.6	[62]



Dan	R	R'	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
180	H	2,5-diOCH <sub>3</sub>	47	47	1	2.2	21	[65]
181	H	3,5-diOCH <sub>3</sub>	20	20	1	2.6	7.7	[65]
182	H	2,4-diOCH <sub>3</sub>	16	16	1	7.1	2.3	[65]
183	H	3,4,5-triOCH <sub>3</sub>	22	22	1	9.3	2.4	[65]
184	H	2,5-diCl	25	25	1	0.66	38	[65]
185	H	3,5-diCl	66.2	63	0.95	4.1	15	[65]
186	H	2,4-diCl	15	15	1	2.2	6.8	[65]
187	H	3-Cl	26	26	1	3.5	7.4	[65]
188	CH <sub>3</sub>	2,5-diOCH <sub>3</sub>	56	56	1	10.3	5.4	[65]
189	CH <sub>3</sub>	3,5-diOCH <sub>3</sub>	87	87	1	4.2	21	[65]
190	CH <sub>3</sub>	3,4,5-triOCH <sub>3</sub>	45	45	1	37	1.2	[65]

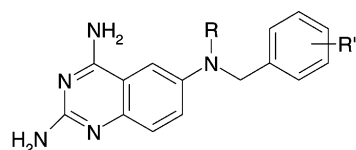
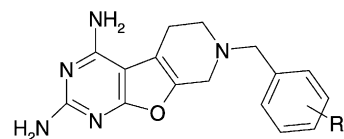
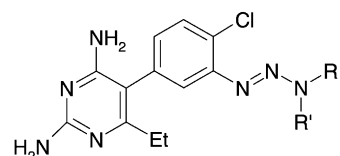


Table A.1 (Continued)

Dan	R	R'	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
191	H	H	8.7	0.66	0.076	0.32	2.1	[63]
192	H	2,5-diOCH <sub>3</sub>	4.6	1.1	0.24	0.16	6.9	[63]
193	H	3,5-diOCH <sub>3</sub>	2.2	0.84	0.38	0.12	7	[63]
194	H	2,4-diOCH <sub>3</sub>	4.4	1.2	0.27	0.17	7.1	[63]
195	H	3,4,5-triOCH <sub>3</sub>	6.8	0.9	0.13	0.084	11	[63]
196	H	2,3,4-triOCH <sub>3</sub>	4.9	1.3	0.27	0.19	6.8	[63]
197	H	2,4,5-triOCH <sub>3</sub>	5.4	1.6	0.3	0.124	13	[63]
198	H	2,4,6-triOCH <sub>3</sub>	116	22.7	0.2	0.95	24	[63]
199	H	2,3-C <sub>4</sub> H <sub>4</sub>	0.72	0.19	0.26	0.099	1.9	[63]
200	CH <sub>3</sub>	2,5-diOCH <sub>3</sub>	0.087	0.026	0.3	0.03	0.87	[63]
201	CH <sub>3</sub>	3,5-diOCH <sub>3</sub>	0.0238	0.0082	0.34	0.0093	0.88	[63]
202	CH <sub>3</sub>	2,4-diOCH <sub>3</sub>	0.1	0.043	0.43	0.039	1.1	[63]
203	CH <sub>3</sub>	2,3,4-triOCH <sub>3</sub>	0.052	0.019	0.37	0.017	1.1	[63]
204	CH <sub>3</sub>	2,3-C <sub>4</sub> H <sub>4</sub>	0.017	0.017	1	0.021	0.81	[63]

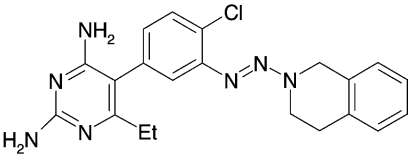
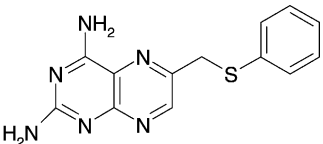
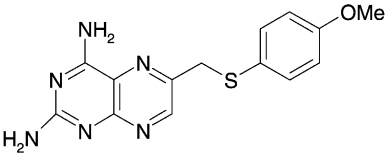
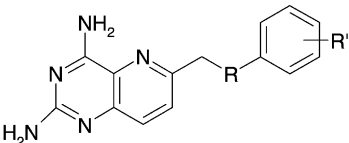


Dan	R	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
205	H	–	15	–	1.4	11	[64]
206	3,5-diOCH <sub>3</sub>	–	22.8	–	–	–	[64]
207	2,4-diCl	22.6	50.9	2.3	13.1	3.9	[64]
208	3,4-diCl	24.1	20.6	0.85	22.3	0.92	[64]
209	2,6-diCl	40.5	81.2	2	31.7	2.6	[64]



Dan	R	R'	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
210	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>3</sub>	14.1	202	14	–	–	[8]
211	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	10.2	–	–	–	–	[8]
212	CH <sub>3</sub>	CH <sub>2</sub> Ph	2.8	1.8	0.64	–	–	[8]
213	CH <sub>3</sub>	(CH <sub>2</sub> ) <sub>2</sub> N(CH <sub>2</sub> CH <sub>3</sub> ) <sub>2</sub>	18.2	17.3	0.95	–	–	[8]
214	CH <sub>2</sub> CH <sub>3</sub>	CH <sub>2</sub> CH <sub>3</sub>	29.3	17.7	0.6	–	–	[8]
215		–(CH <sub>2</sub> ) <sub>5</sub> –	40.4	66.8	1.7	–	–	[8]
216		–(CH <sub>2</sub> ) <sub>2</sub> NOCH <sub>3</sub> (CH <sub>2</sub> ) <sub>2</sub> –	8.5	16.2	1.9	–	–	[8]
217		–(CH <sub>2</sub> ) <sub>2</sub> CHOH(CH <sub>2</sub> ) <sub>2</sub> –	3.3	7.2	2.2	–	–	[8]
218		–(CH <sub>2</sub> ) <sub>2</sub> O(CH <sub>2</sub> ) <sub>2</sub> –	1.7	26.3	15	–	–	[8]
219	CH <sub>3</sub>	(CH <sub>2</sub> ) <sub>2</sub> OH	3.5	27.9	8	–	–	[8]
220	CH <sub>2</sub> CH <sub>3</sub>	(CH <sub>2</sub> ) <sub>2</sub> OH	11.5	27	2.3	–	–	[8]

221	CH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	(CH <sub>2</sub> ) <sub>2</sub> OH	2.5	0.44	0.18	–	–	[8]
222	<i>tert</i> -Butyl	(CH <sub>2</sub> ) <sub>2</sub> OH	4.9	10.3	2.1	–	–	[8]
223	CH <sub>2</sub> Ph	(CH <sub>2</sub> ) <sub>2</sub> OH	0.26	7	27	–	–	[8]
224	(CH <sub>2</sub> ) <sub>2</sub> OH	(CH <sub>2</sub> ) <sub>2</sub> OH	0.44	5	11	–	–	[8]
225	(CH <sub>2</sub> ) <sub>2</sub> OCH <sub>3</sub>	(CH <sub>2</sub> ) <sub>2</sub> OCH <sub>3</sub>	0.91	26.1	29	–	–	[8]
226	CH <sub>3</sub>		0.57	0.5	0.88	–	–	[8]

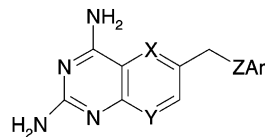
Dan			<i>pc</i> DHFR	rlDHFR	rl <i>pc</i>	<i>tg</i> DHFR	rl <i>tg</i>	Reference
227			5	1.7	0.34	–	–	[8]
228			9.5	246	26	0.77	320	[54]
229			0.56	0.52	0.93	0.063	8.3	[54]
								

Dan	R	R'	<i>pc</i> DHFR	rlDHFR	rl <i>pc</i>	<i>tg</i> DHFR	rl <i>tg</i>	Reference
230	S	H	2	0.52	0.26	0.13	4	[54]
231	NH	H	1.7	0.26	0.15	0.085	3.1	[54]
232	NCH <sub>3</sub>	H	0.29	0.024	0.08	0.0084	2.9	[54]
233	NH	2-OCH <sub>3</sub>	2.7	0.42	0.16	0.12	3.5	[54]
234	NCH <sub>3</sub>	2-OCH <sub>3</sub>	0.51	0.12	0.24	0.026	4.6	[54]
235	NH	3-OCH <sub>3</sub>	1.7	0.2	0.12	0.1	2	[54]
236	NCH <sub>3</sub>	3-OCH <sub>3</sub>	0.097	0.035	0.36	0.015	2.3	[54]
237	NH	4-OCH <sub>3</sub>	0.85	0.073	0.086	0.054	1.4	[54]
238	NCH <sub>3</sub>	4-OCH <sub>3</sub>	0.25	0.018	0.072	0.016	1.1	[54]
239	NH	2-Cl	0.53	0.14	0.26	0.11	1.3	[54]
240	NCH <sub>3</sub>	2-Cl	0.21	0.12	0.57	0.015	8	[54]
241	NH	3-Cl	2	0.14	0.07	0.13	1.1	[54]
242	NCH <sub>3</sub>	2,4-diOCH <sub>3</sub>	5.5	0.32	0.058	0.14	2.3	[54]



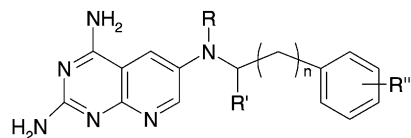
Table A.1 (Continued)

Dan	R	R'	pcDHFR	rlDHFR	rlpc	tgDHFR	rltg	Reference
243	NH	2,4-diOCH <sub>3</sub>	0.16	0.016	0.1	0.014	1.1	[54]
244	NCH <sub>3</sub>	2,5-diOCH <sub>3</sub>	4.4	0.28	0.064	0.12	2.3	[54]
245	NH	2,5-diOCH <sub>3</sub>	0.21	0.05	0.24	0.025	2	[54]
246	NCH <sub>3</sub>	3,4-diOCH <sub>3</sub>	0.9	0.06	0.067	0.09	0.67	[54]
247	NH	3,4-diOCH <sub>3</sub>	0.091	0.0027	0.03	0.0098	0.28	[54]
248	NCH <sub>3</sub>	2,4-diCl	0.73	0.088	0.12	0.05	1.8	[54]
249	NH	2,4-diCl	0.5	0.058	0.12	0.05	1.2	[54]
250	NCH <sub>3</sub>	2,5-diCl	1.6	0.2	0.13	0.091	2.2	[54]
251	NH	2,5-diCl	0.15	0.047	0.31	0.025	1.9	[54]
252	NCH <sub>3</sub>	2,6-diCl	1	0.082	0.082	0.028	2.9	[54]
253	NH	2,6-diCl	0.17	0.048	0.28	0.03	1.6	[54]
254	NCH <sub>3</sub>	3,4-diCl	0.41	0.054	0.13	0.057	0.95	[54]
255	NH	3,4-diCl	0.038	0.017	0.45	0.027	0.63	[54]
256	NCH <sub>3</sub>	3,4,5-triOCH <sub>3</sub>	2	0.2	0.1	0.04	5	[54]
257	NH	3,4,5-triCl	0.66	0.044	0.067	0.087	0.51	[54]
258	NCH <sub>3</sub>	3,4,5-triCl	0.25	0.087	0.35	0.038	2.3	[54]
259	NH	2,4,6-triCl	2	0.57	0.29	0.046	12	[54]
260	NCH <sub>3</sub>	2,4,6-triCl	0.12	0.052	0.43	0.044	1.2	[54]
261	S	2,3-C <sub>4</sub> H <sub>4</sub>	0.47	0.16	0.34	0.049	3.3	[54]
262	S	3,4-C <sub>4</sub> H <sub>4</sub>	0.38	0.086	0.23	0.048	1.8	[54]
263	NH	2,3-C <sub>4</sub> H <sub>4</sub>	0.23	0.04	0.17	0.026	1.5	[54]
264	NH	3,4-C <sub>4</sub> H <sub>4</sub>	1.6	0.21	0.13	0.16	1.3	[54]
265	NCH <sub>3</sub>	2,3-C <sub>4</sub> H <sub>4</sub>	0.04	0.0073	0.18	0.018	0.41	[54]
266	NCH <sub>3</sub>	3,4-C <sub>4</sub> H <sub>4</sub>	0.052	0.0072	0.14	0.016	0.45	[54]
267	NH	4-COCH <sub>3</sub>	0.41	0.0025	0.0061	0.027	0.093	[54]
268	NCH <sub>3</sub>	4-COCH <sub>3</sub>	0.13	0.0051	0.039	0.015	0.34	[54]
269	N-propargyl	4-COCH <sub>3</sub>	0.22	0.015	0.068	0.02	0.75	[54]
270	NCH <sub>3</sub>	4-COCF <sub>3</sub>	0.25	0.032	0.13	0.046	0.7	[54]
271	N-propargyl	4-COCF <sub>3</sub>	0.12	0.0075	0.063	0.054	0.14	[54]



Dan	X	Y	Z-Ar	pcDHFR	rlDHFR	rlpc	tgDHFR	rltg	Reference
272	CH	CH	NHC <sub>6</sub> H <sub>3</sub> (OCH <sub>3</sub> ) <sub>2</sub> -2,5	0.75	0.46	0.61	0.14	3.3	[6]
273	CCH <sub>3</sub>	N	SC <sub>6</sub> H <sub>5</sub>	0.44	0.43	0.98	0.034	13	[6]
274	CH	CH	NHC <sub>6</sub> H <sub>3</sub> -4-Cl-2-CH <sub>3</sub>	0.33	0.023	0.07	0.033	0.7	[6]
275	N	N	SC <sub>6</sub> H <sub>4</sub> CH <sub>3</sub> -3	21.2	8.48	0.4	1.8	4.7	[6]
276	CCH <sub>3</sub>	N	SC <sub>6</sub> H <sub>4</sub> CH <sub>3</sub> -3	0.17	0.33	1.9	0.065	5.1	[6]
277	N	N	SC <sub>6</sub> H <sub>4</sub> CH <sub>3</sub> -4	30	26	0.87	15	1.7	[6]
278	CCH <sub>3</sub>	N	SC <sub>6</sub> H <sub>4</sub> CH <sub>3</sub> -4	0.53	0.5	0.93	0.057	8.7	[6]
279	N	N	SC <sub>6</sub> H <sub>4</sub> OCH <sub>3</sub> -3	9.9	7.7	0.78	1.6	4.8	[6]
280	CCH <sub>3</sub>	N	SC <sub>6</sub> H <sub>4</sub> OCH <sub>3</sub> -3	0.34	0.6	1.8	0.11	5.3	[6]
281	N	N	SC <sub>6</sub> H <sub>4</sub> OCH <sub>3</sub> -4	17.2	10.2	0.59	14	0.73	[6]

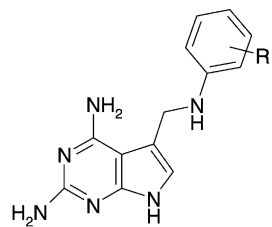
282	N	N	SC <sub>6</sub> H <sub>3</sub> (OCH <sub>3</sub> ) <sub>2</sub> -3,4	58.2	36.3	0.62	23.2	1.6	[6]
283	CCH <sub>3</sub>	N	SC <sub>6</sub> H <sub>3</sub> (OCH <sub>3</sub> ) <sub>2</sub> -3,4	0.15	0.18	1.2	0.03	5.9	[6]
284	N	N	SC <sub>6</sub> H <sub>4</sub> Cl-3	42.2	26.6	0.63	5.7	4.7	[6]
285	CCH <sub>3</sub>	N	SC <sub>6</sub> H <sub>4</sub> Cl-3	0.068	0.19	2.8	0.09	2.1	[6]
286	N	N	SC <sub>6</sub> H <sub>4</sub> Cl-4	42.1	14.3	0.34	16.8	0.85	[6]
287	CCH <sub>3</sub>	N	SC <sub>6</sub> H <sub>4</sub> Cl-4	0.36	0.37	1	0.09	4.1	[6]
288	N	N	NHC <sub>6</sub> H <sub>4</sub> Cl-4	1.5	0.3	2.1	0.59	0.52	[6]
289	N	N	NHC <sub>6</sub> H <sub>3</sub> -2-CH <sub>3</sub> -5-OCH <sub>3</sub>	–	–	–	0.44	–	[6]
290	CCH <sub>3</sub>	N	NHC <sub>6</sub> H <sub>3</sub> -2-CH <sub>3</sub> -5-OCH <sub>3</sub>	0.038	0.15	3.9	0.023	6.3	[6]
291	N	N	NHC <sub>6</sub> H <sub>3</sub> -2-CH <sub>3</sub> -6-OCH <sub>3</sub>	15.8	5.7	0.36	1.85	3.1	[6]
292	CCH <sub>3</sub>	N	NHC <sub>6</sub> H <sub>3</sub> -2-CH <sub>3</sub> -6-OCH <sub>3</sub>	1	0.32	0.32	0.1	3.2	[6]
293	N	N	NHC <sub>6</sub> H <sub>3</sub> (OCH <sub>3</sub> ) <sub>2</sub> -2,5	6.2	22.9	3.7	6.9	3.3	[6]
294	CCH <sub>3</sub>	N	NHC <sub>6</sub> H <sub>3</sub> (OCH <sub>3</sub> ) <sub>2</sub> -2,5	0.011	0.01	0.94	0.014	0.73	[6]
295	N	N	N(CH <sub>3</sub> )C <sub>6</sub> H <sub>3</sub> (OCH <sub>3</sub> ) <sub>2</sub> -2,5	3.9	0.47	0.12	0.21	2.2	[6]
296	N	N	NHC <sub>6</sub> H <sub>3</sub> (OCH <sub>3</sub> ) <sub>2</sub> -3,5	0.96	0.88	0.92	0.11	8	[6]
297	CCH <sub>3</sub>	N	NHC <sub>6</sub> H <sub>3</sub> (OCH <sub>3</sub> ) <sub>2</sub> -3,5	0.049	0.031	0.67	0.0035	8.9	[6]
298	N	N	NHC <sub>6</sub> H <sub>2</sub> (OCH <sub>3</sub> ) <sub>3</sub> -3,4,5	7	1.9	0.27	1	1.9	[6]
299	N	N	NHC <sub>6</sub> H <sub>3</sub> -2-OCH <sub>3</sub> -5-CF <sub>3</sub>	21	21	1	10.6	2.2	[6]
300	CCH <sub>3</sub>	N	NHC <sub>6</sub> H <sub>3</sub> -2-OCH <sub>3</sub> -5-CF <sub>3</sub>	0.044	0.02	0.45	0.022	0.91	[6]
301	N	N	NHC <sub>6</sub> H <sub>3</sub> -3-OCH <sub>3</sub> -5-CF <sub>3</sub>	0.68	1.9	2.8	0.89	2.1	[6]
302	CCH <sub>3</sub>	N	NHC <sub>6</sub> H <sub>3</sub> -3-OCH <sub>3</sub> -5-CF <sub>3</sub>	0.02	0.017	0.85	0.018	0.94	[6]
303	CH	CH	NHC <sub>6</sub> H <sub>4</sub> Cl-4	0.6	0.073	0.12	0.075	0.97	[6]
304	CH	N	NHC <sub>6</sub> H <sub>4</sub> Cl-4	0.97	0.72	0.74	0.3	2.4	[6]
305	N	N	CH <sub>2</sub> C <sub>6</sub> H <sub>3</sub> (OCH <sub>3</sub> ) <sub>2</sub> -2,5	11.1	23.2	2.1	5.4	4.3	[6]
306	CCH <sub>3</sub>	N	CH <sub>2</sub> C <sub>6</sub> H <sub>3</sub> (OCH <sub>3</sub> ) <sub>2</sub> -2,5	0.34	0.77	2.3	0.0079	98	[6]
	CCH <sub>3</sub>	N	NHC <sub>6</sub> H <sub>3</sub> (CH <sub>3</sub> ) <sub>2</sub> -2,5	0.03	0.12	4	0.016	7.5	[6]
308	CCH <sub>3</sub>	N	NHC <sub>6</sub> H <sub>3</sub> -2-CH <sub>3</sub> -4-OCH <sub>3</sub>	0.17	0.029	0.17	0.0072	4	[6]
309	CCH <sub>3</sub>	N	NHC <sub>6</sub> H <sub>3</sub> -2-OCH <sub>3</sub> -5-CH <sub>3</sub>	0.068	0.16	2.4	0.015	11	[6]
310	CCH <sub>3</sub>	N	NHC <sub>6</sub> H <sub>3</sub> (OCH <sub>3</sub> ) <sub>2</sub> -2,4	0.091	0.061	0.67	0.01	6.1	[6]
311	CCH <sub>3</sub>	N	N(CH <sub>3</sub> )C <sub>6</sub> H <sub>3</sub> (OCH <sub>3</sub> ) <sub>2</sub> -2,4	0.35	0.23	0.66	0.053	4.3	[6]
312	CCH <sub>3</sub>	N	N(CH <sub>3</sub> )C <sub>6</sub> H <sub>3</sub> (OCH <sub>3</sub> ) <sub>2</sub> -3,5	0.05	0.013	0.26	0.004	3.2	[6]
313	CCH <sub>3</sub>	N	NHC <sub>6</sub> H <sub>3</sub> (OC <sub>2</sub> H <sub>5</sub> ) <sub>2</sub> -2,5	0.125	0.37	3	0.022	16	[6]
314	CCH <sub>3</sub>	N	N(CH <sub>3</sub> )C <sub>6</sub> H <sub>3</sub> (OC <sub>2</sub> H <sub>5</sub> ) <sub>2</sub> -2,5	0.167	0.53	3.2	0.029	18	[6]
315	CCH <sub>3</sub>	N	N(CH <sub>3</sub> )C <sub>6</sub> H <sub>3</sub> -2-OCH <sub>3</sub> -5-CF <sub>3</sub>	0.093	0.23	2.5	0.038	6.2	[6]
316	CCH <sub>3</sub>	N	NHC <sub>6</sub> H <sub>3</sub> (CH <sub>2</sub> ) <sub>4</sub> -2,3	0.1	0.13	1.3	0.026	5	[6]
317	CCH <sub>3</sub>	N	N(CH <sub>3</sub> )C <sub>6</sub> H <sub>3</sub> (CH <sub>2</sub> ) <sub>4</sub> -2,3	0.15	0.14	0.93	0.016	8.8	[6]



Dan	R	R'	n	R''	pcDHFR	riDHFR	rlpc	tgDHFR	rltg	Reference
318	H	H	0	H	2.7	2.1	0.78	0.52	4	[55]
319	CH <sub>3</sub>	H	0	H	0.068	0.14	2.1	0.032	4.4	[55]
320	H	H	0	3,4,5-triOCH <sub>3</sub>	14.1	3.3	0.23	0.35	9.4	[55]
321	CH <sub>3</sub>	H	0	3,4,5-triOCH <sub>3</sub>	0.061	0.033	0.54	0.014	2.4	[55]
322	H	CH <sub>3</sub>	0	3,4,5-triOCH <sub>3</sub>	9.2	1.27	0.14	0.194	6.5	[55]
323	H	H	0	2,3,4-triOCH <sub>3</sub>	15.3	3.24	0.21	0.67	4.8	[55]

Table A.1 (Continued)

Dan	R	R'	<i>n</i>	R''	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
324	CH <sub>3</sub>	H	0	2,3,4-triOCH <sub>3</sub>	0.079	0.03	0.38	0.026	1.2	[55]
325	H	H	0	2,4,6-triOCH <sub>3</sub>	20.7	1.2	0.058	0.23	5.2	[55]
326	H	H	0	2,4,5-triOCH <sub>3</sub>	5.5	1.1	0.2	0.48	2.3	[55]
327	H	H	0	3,4-diOCH <sub>3</sub>	4.8	1.5	0.31	0.73	2.1	[55]
328	H	H	0	3,5-diOCH <sub>3</sub>	5.7	3.4	0.6	1.2	2.8	[55]
329	CH <sub>3</sub>	H	0	3,5-diOCH <sub>3</sub>	0.076	0.072	0.95	0.031	2.3	[55]
330	H	H	0	2,5-diOCH <sub>3</sub>	3.8	0.35	0.092	0.31	1.1	[55]
331	CH <sub>3</sub>	H	0	2,5-diOCH <sub>3</sub>	0.084	0.057	0.68	0.0063	9	[55]
332	H	H	0	2,3-C <sub>4</sub> H <sub>4</sub>	3.9	0.24	0.062	0.98	0.24	[55]
333	H	H	0	4-OCH <sub>3</sub> -2,3-C <sub>4</sub> H <sub>4</sub>	8.2	0.43	0.052	0.38	1.1	[55]
334	H	H	0	6-OCH <sub>3</sub> -2,3-C <sub>4</sub> H <sub>4</sub>	15.4	0.37	0.024	0.71	0.52	[55]
335	H	H	0	4-O-C <sub>6</sub> H <sub>5</sub>	24.3	2.9	0.12	3.7	0.78	[55]
336	H	H	1	H	1.94	0.28	0.14	4.45	0.063	[55]
337	CH <sub>3</sub>	H	1	H	1.8	1.4	0.78	0.92	1.5	[55]



Dan	R	<i>pc</i> DHFR	<i>rl</i> DHFR	<i>rlpc</i>	<i>tg</i> DHFR	<i>rltg</i>	Reference
338	3,4,5-triOCH <sub>3</sub>	–	56.3	–	8.1	7	[53]
339	3,4-diOCH <sub>3</sub>	119	116	0.97	4.3	27	[53]
340	4-OCH <sub>3</sub>	279	63	0.23	6	11	[53]
341	2,5-diOCH <sub>3</sub>	45.7	156	3.4	1.7	92	[53]
342	2,5-diOCH <sub>3</sub>	–	70	–	5.3	13	[53]
343	3,4-diCl	35.3	14.4	0.41	1.4	10	[53]
344	2,3-(CH) <sub>4</sub>	307	59.3	0.19	1.1	54	[53]
345	H	252	–	–	3.9	–	[53]

## References

- [1] Folates and pterins, R.L. Blakley, S.J. Benkovic (Eds.), Wiley, New York, 1984.
- [2] J.E. Gready, Dihydrofolate reductase: binding of substrates and inhibitors and catalytic mechanism, *Adv. Pharmacol. Chemother.* 17 (1980) 37–102.
- [3] R. Behbahani, M. Moshfeghi, J.D. Baxter, Therapeutic approaches for AIDS-related toxoplasmosis, *Ann. Pharmacother.* 29 (1995) 760–768.
- [4] E. Warren, S. George, J. You, P. Kazanjian, Advances in the treatment and prophylaxis of pneumocystis carinii pneumonia, *Pharmacotherapy* 17 (1997) 900–916.
- [5] A. Gangjee, O. Adair, S.F. Queener, Pneumocystis carinii and toxoplasma gondii dihydrofolate reductase inhibitors and antitumor agents: synthesis and biological activities of 2,4-diamino-5-methyl-6-[(monosubstituted anilino)methyl]-pyrido[2,3-d]pyrimidines, *J. Med. Chem.* 42 (1999) 2447–2455.
- [6] J.R. Piper, C.A. Johnson, C.A. Krauth, R.L. Carter, C.A. Hosmer, S.F. Queener, S.E. Borotz, E.R. Pfefferkorn, Lipophilic antifolates as agents against opportunistic infections. I. Agents superior to trimetrexate and piritrexim against toxoplasma gondii and pneumocystis carinii in in vitro evaluations, *J. Med. Chem.* 39 (1996) 1271–1280.
- [7] A. Rosowsky, V. Cody, N. Galitsky, H. Fu, A.T. Papoulis, S.F. Queener, Structure-based design of selective inhibitors of dihydrofolate reductase: synthesis and antiparasitic activity of 2,4-diaminopteridine analogues with a bridged diarylamine side chain, *J. Med. Chem.* 42 (1999) 4853–4860.
- [8] M.F.G. Stevens, K.S. Phillip, D.L. Rathbone, D.M. O'Shea, S.F. Queener, C.H. Schwalbe, P.A. Lambert, Structural studies on bioactive compounds. 28. Selective activity of triazenyl-substituted pyrimethamine derivatives against pneumocystis carinii dihydrofolate reductase, *J. Med. Chem.* 40 (1997) 1886–1893.
- [9] G.W. Kauffman, P.C. Jurs, QSAR and *k*-nearest neighbor classification analysis of selective cyclooxygenase-2 inhibitors using topologically-based numerical descriptors, *J. Chem. Inf. Comput. Sci.* 41 (2001) 1553–1560.
- [10] P.D. Mosier, A.E. Counterman, P.C. Jurs, D.E. Clemmer, Prediction of peptide ion collision cross sections from topological molecular structure and amino acid parameters, *Anal. Chem.* 74 (2002) 1360–1370.
- [11] S.J. Patankar, P.C. Jurs, Prediction of IC<sub>50</sub> values for ACAT inhibitors from molecular structure, *J. Chem. Inf. Comput. Sci.* 40 (2000) 706–723.
- [12] M.D. Wessel, P.C. Jurs, J.W. Tolan, S.M. Muskal, Prediction of human intestinal absorption of drug compounds from molecular structure, *J. Chem. Inf. Comput. Sci.* 38 (1998) 726–735.
- [13] G.A. Bakken, P.C. Jurs, Classification of multidrug-resistance reversal agents using structure-based descriptors and linear discriminant analysis, *J. Med. Chem.* 43 (2000) 4534–4541.
- [14] L. Xue, J. Bajorath, Accurate partitioning of compounds belonging to diverse activity classes, *J. Chem. Inf. Comput. Sci.* 42 (2002) 757–764.
- [15] O. Ivanciuc, T. Ivanciuc, D. Cabrol-Bass, QSAR for dihydrofolate reductase inhibitors with molecular graph structural descriptors, *Theochem. J. Mol. Struct.* 582 (2002) 39–51.
- [16] R. Burbidge, M. Trotter, B. Buxton, S. Holden, Drug design by machine learning: support vector machines for pharmaceutical data analysis, *Comput. Chem.* 26 (2001) 5–14.
- [17] S. Garg, L.E.K. Achenie, Mathematical programming assisted drug design for nonclassical antifolates, *Biotechnol. Prog.* 17 (2001) 412–418.
- [18] W. Zheng, A. Tropsha, Novel variable selection quantitative structure–property relationship approach based on the *k*-nearest-neighbor principle, *J. Chem. Inf. Comput. Sci.* 40 (2000) 185–194.
- [19] C.D. Selassie, W.-X. Gan, L.S. Kallander, T.E. Klein, Quantitative structure–activity relationships of 2,4-diamino-5-(2-X-benzyl) pyrimidines versus bacterial and avian dihydrofolate reductase, *J. Med. Chem.* 41 (1998) 4261–4272.
- [20] F.R. Burden, B.S. Rosewarne, D.A. Winkler, Predicting maximum bioactivity by effective inversion of neural networks using genetic algorithms, *Chemometr. Intell. Lab.* 38 (1997) 127–137.
- [21] J. Kyngas, J. Valjakka, Evolutionary neural networks in quantitative structure–activity relationships of dihydrofolate reductase inhibitors, *Quant. Struct. Act. Relat.* 15 (1996) 296–301.
- [22] O. Ivanciuc, Artificial neural networks applications. Part 2. Using theoretical descriptors of molecular structure in quantitative structure–activity relationships: analysis of the inhibition of dihydrofolate reductase, *Revue Roumaine de Chimie* 41 (1996) 645–652.
- [23] C.K. Marlowe, C.D. Selassie, D.V. Santi, Quantitative structure–activity relationships of the inhibition of *Pneumocystis carinii* dihydrofolate reductase by 4,6-diamino-1,2-dihydro-2,2-dimethyl-1-(X-phenyl)-s-triazines, *J. Med. Chem.* 38 (1995) 967–972.
- [24] J.D. Hirst, R.D. King, M.J.E. Sternberg, Quantitative structure–activity relationships by neural networks and inductive logic programming. II. The inhibition of dihydrofolate reductase by triazines, *J. Comput. Aid. Mol. Des.* 8 (1994) 421–432.
- [25] J.D. Hirst, R.D. King, M.J.E. Sternberg, Quantitative structure–activity relationships by neural networks and inductive logic programming. I. The inhibition of dihydrofolate reductase by pyrimidines, *J. Comput. Aid. Mol. Des.* 8 (1994) 405–420.
- [26] D.T. Stanton, W.J. Murray, P.C. Jurs, Comparison of QSAR and molecular similarity approaches for a structure–activity relationship study of DHFR inhibitors, *Quant. Struct. Act. Relat.* 12 (1993) 239–245.
- [27] S.-S. So, W.G. Richards, Application of neural networks: quantitative structure–activity relationships of the derivatives of 2,4-diamino-5-(substituted-benzyl)pyrimidines as DHFR inhibitors, *J. Med. Chem.* 35 (1992) 3201–3207.
- [28] C.D. Selassie, R.-L. Li, M. Poe, C. Hansch, On the optimization of hydrophobic and hydrophilic substituent interactions of 2,4-diamino-5-(substituted-benzyl)pyrimidines with dihydrofolate reductase, *J. Med. Chem.* 34 (1991) 46–54.
- [29] A.K. Debnath, R.L. Lopez de Compadre, G. Debnath, A.J. Shusterman, C. Hansch, Structure–activity relationship of mutagenic aromatic and heteroaromatic nitro compounds: correlation with molecular orbital energies and hydrophobicity, *J. Med. Chem.* 34 (1991) 786–797.
- [30] T.A. Andrea, H. Kalayeh, Applications of neural networks in quantitative structure–activity relationships of dihydrofolate reductase inhibitors, *J. Med. Chem.* 34 (1991) 2824–2836.
- [31] C.D. Selassie, Z.-X. Fang, R.-L. Li, C. Hansch, G. Debnath, T.E. Klein, R. Langridge, B.T. Kaufman, On the structure selectivity problem in drug design: a comparative study of benzylpyrimidine inhibition of vertebrate and bacterial dihydrofolate reductase via molecular graphics and quantitative structure–activity relationships, *J. Med. Chem.* 32 (1989) 1895–1905.
- [32] R.-L. Li, M. Poe, Quantitative structure–activity relationships for the inhibition of *Escherichia coli* dihydrofolate reductase by 5-(substituted benzyl)-2,4-diaminopyrimidines, *J. Med. Chem.* 31 (1988) 366–370.
- [33] R.G. Booth, C.D. Selassie, C. Hansch, D.V. Santi, Quantitative structure–activity relationship of triazine-antifolate inhibition of *Leishmania* dihydrofolate reductase and cell growth, *J. Med. Chem.* 30 (1987) 1218–1224.
- [34] C.D. Selassie, Z.-X. Fang, R.-L. Li, C. Hansch, T.E. Klein, R. Langridge, B.T. Kaufman, Inhibition of chicken liver dihydrofolate reductase by 5-(substituted benzyl)-2,4-diaminopyrimidines: a quantitative structure–activity relationship and graphics analysis, *J. Med. Chem.* 29 (1986) 621–626.

- [35] A.K. Ghose, G.M. Crippen, Use of physicochemical parameters in distance geometry and related three-dimensional quantitative structure–activity relationships: a demonstration using *Escherichia coli* dihydrofolate reductase inhibitors, *J. Med. Chem.* 28 (1985) 333–346.
- [36] C. Hansch, B.A. Hathaway, Z.-R. Guo, C.D. Selassie, S.W. Dietrich, J.M. Blaney, R. Langridge, K.W. Volz, B.T. Kaufman, Crystallography, quantitative structure–activity relationships, and molecular graphics in a comparative analysis of the inhibition of dihydrofolate reductase from chicken liver and *Lactobacillus casei* by 4,6-diamino-1,2-dihydro-2,2-dimethyl-1-(substituted-phenyl)-s-triazines, *J. Med. Chem.* 27 (1984) 129–143.
- [37] B.A. Hathaway, Z.-R. Guo, C. Hansch, T.J. Delcamp, S.S. Susten, J.H. Freisheim, Inhibition of human dihydrofolate reductase by 4,6-diamino-1,2-dihydro-2,2-dimethyl-1-(substituted-phenyl)-s-triazines: a quantitative structure–activity relationship analysis, *J. Med. Chem.* 27 (1984) 144–149.
- [38] A.J. Hopfinger, Theory and application of molecular potential energy fields in molecular shape analysis: a quantitative structure–activity relationship study of 2,4-diamino-5-benzylpyrimidines as dihydrofolate reductase inhibitors, *J. Med. Chem.* 26 (1983) 990–996.
- [39] T.A. Khwaja, S. Pentecost, C.D. Selassie, Z.-R. Guo, C. Hansch, Comparison of quantitative structure–activity relationships of the inhibition of leukemia cells in culture with the inhibition of dihydrofolate reductase from leukemia cells and other cell types, *J. Med. Chem.* 25 (1982) 153–156.
- [40] C. Hansch, R.-L. Li, J.M. Blaney, R. Langridge, Comparison of the inhibition of *Escherichia coli* and *Lactobacillus casei* dihydrofolate reductase by 2,4-diamino-5-(substituted-benzyl)pyrimidines: quantitative structure–activity relationships, X-ray crystallography, and computer graphics in structure–activity analysis, *J. Med. Chem.* 25 (1982) 777–784.
- [41] A.K. Ghose, G.M. Crippen, Quantitative structure–activity relationship by distance geometry: quinazolines as dihydrofolate reductase inhibitors, *J. Med. Chem.* 25 (1982) 892–899.
- [42] R.-L. Li, S.W. Dietrich, C. Hansch, Quantitative structure–selectivity relationships. comparison of the inhibition of *Escherichia coli* and bovine liver dihydrofolate reductase by 5-(substituted-benzyl)-2,4-diaminopyrimidines, *J. Med. Chem.* 24 (1981) 538–544.
- [43] E.A. Coats, C.S. Genther, S.W. Dietrich, Z.-R. Guo, C. Hansch, Comparison of the inhibition of methotrexate-sensitive and -resistant *Lactobacillus casei* cell cultures with purified *Lactobacillus casei* dihydrofolate reductase by 4,6-diamino-1,2-dihydro-2,2-dimethyl-1-(3-substituted-phenyl)-s-triazines: use of quantitative structure–activity relationships in making inferences about the mechanism of resistance and the structure of the enzyme in situ compared with the enzyme in vitro, *J. Med. Chem.* 24 (1981) 1422–1429.
- [44] G.M. Crippen, Quantitative structure–activity relationships by distance geometry: systematic analysis of dihydrofolate reductase inhibitors, *J. Med. Chem.* 23 (1980) 599–606.
- [45] S.W. Dietrich, J.M. Blaney, M.A. Reynolds, P.Y.C. Jow, C. Hansch, Quantitative structure–selectivity relationships. comparison of the inhibition of *Escherichia coli* and bovine liver dihydrofolate reductase by 5-(substituted-benzyl)-2,4-diaminopyrimidines, *J. Med. Chem.* 23 (1980) 1205–1212.
- [46] J.M. Blaney, C. Hansch, C. Silipo, A. Vittoria, Structure–activity relationships of dihydrofolate reductase inhibitors, *Chem. Rev.* 84 (1984) 333–407.
- [47] A. Rosowsky, C.E. Mota, J.E. Wright, S.F. Queener, 2,4-Diamino-5-chloroquinazoline analogues of trimetrexate and piritrexim: synthesis and antifolate Activity, *J. Med. Chem.* 37 (1994) 4522–4528.
- [48] A. Rosowsky, R.A. Forsch, S.F. Queener, 2,4-Diaminopyrido[3,2-*d*]pyrimidine inhibitors of dihydrofolate reductase from *Pneumocystis carinii* and *Toxoplasma gondii*, *J. Med. Chem.* 38 (1995) 2615–2620.
- [49] A. Rosowsky, A.T. Papoulis, S.F. Queener, 2,4-Diaminothieno[2,3-*d*]pyrimidine lipophilic antifolates as inhibitors of *Pneumocystis carinii* and *Toxoplasma gondii* dihydrofolate reductase, *J. Med. Chem.* 40 (1997) 3694–3699.
- [50] A. Rosowsky, A.T. Papoulis, S.F. Queener, 2,4-Diamino-6,7-dihydro-5*H*-cyclopenta[*d*]pyrimidine analogues of trimethoprim as inhibitors of *Pneumocystis carinii* and *Toxoplasma gondii* dihydrofolate reductase, *J. Med. Chem.* 41 (1998) 913–918.
- [51] A. Rosowsky, A.T. Papoulis, R.A. Forsch, S.F. Queener, Synthesis and antiparasitic and antitumor activity of 2,4-diamino-6-(arylmethyl)-5,6,7,8-tetrahydroquinazoline analogues of piritrexim, *J. Med. Chem.* 42 (1999) 1007–1017.
- [52] A. Gangjee, N. Zaveri, M. Kothare, S.F. Queener, Nonclassical 2,4-diamino-6-(aminomethyl)-5,6,7,8-tetrahydroquinazoline antifolates: synthesis and biological activities, *J. Med. Chem.* 38 (1995) 3660–3668.
- [53] A. Gangjee, F. Mavandadi, S.F. Queener, J.J. McGuire, Novel 2,4-diamino-5-substituted-pyrrolo[2,3-*d*]pyrimidines as classical and nonclassical antifolate inhibitors of dihydrofolate reductases, *J. Med. Chem.* 38 (1995) 2158–2165.
- [54] A. Gangjee, Y. Zhu, S.F. Queener, P. Francom, A.D. Broom, Nonclassical 2,4-diamino-8-deazaolates analogues as inhibitors of dihydrofolate reductases from rat liver, *Pneumocystis carinii*, and *Toxoplasma gondii*, *J. Med. Chem.* 39 (1996) 1836–1845.
- [55] A. Gangjee, A. Vasudevan, S.F. Queener, R.L. Kisliuk, 2,4-Diamino-5-deaza-6-Substituted pyrido[2,3-*d*]pyrimidine antifolates as potent and selective nonclassical inhibitors of dihydrofolate reductases, *J. Med. Chem.* 39 (1996) 1438–1446.
- [56] A. Gangjee, A. Vasudevan, S.F. Queener, Conformationally restricted analogues of trimethoprim: 2,6-diamino-8-substituted purines as potential dihydrofolate reductase inhibitors from *Pneumocystis carinii* and *Toxoplasma gondii*, *J. Med. Chem.* 40 (1997) 3032–3039.
- [57] A. Gangjee, J. Shi, S.F. Queener, Synthesis and biological activities of conformationally restricted, tricyclic nonclassical antifolates as inhibitors of dihydrofolate reductases, *J. Med. Chem.* 40 (1997) 1930–1936.
- [58] A. Gangjee, F. Mavandadi, S.F. Queener, Effect of N9-methylation and bridge atom variation on the activity of 5-substituted 2,4-diaminopyrrolo[2,3-*d*]pyrimidines against dihydrofolate reductases from *Pneumocystis carinii* and *Toxoplasma gondii*, *J. Med. Chem.* 40 (1997) 1173–1177.
- [59] A. Gangjee, R. Devraj, S.F. Queener, Synthesis and dihydrofolate reductase inhibitory activities of 2,4-diamino-5-deaza and 2,4-diamino-5,10-dideaza lipophilic antifolates, *J. Med. Chem.* 40 (1997) 470–478.
- [60] A. Gangjee, A. Vasudevan, S.F. Queener, Synthesis and biological evaluation of nonclassical 2,4-diamino-5-methylpyrido[2,3-*d*]pyrimidines with novel side chain substitutes as potential inhibitors of dihydrofolate reductases, *J. Med. Chem.* 40 (1997) 479–485.
- [61] A. Gangjee, Y. Zhu, S.F. Queener, 6-Substituted 2,4-diaminopyrido[3,2-*d*]pyrimidine analogues of piritrexim as inhibitors of dihydrofolate reductase from rat liver, *Pneumocystis carinii*, and *Toxoplasma gondii* and as antitumor agents, *J. Med. Chem.* 41 (1998) 4533–4541.
- [62] A. Gangjee, X. Guo, S.F. Queener, V. Cody, N. Galitsky, J.R. Luft, W. Pangborn, Selective *Pneumocystis carinii* dihydrofolate reductase inhibitors: design, synthesis, and biological evaluation of new 2,4-diamino-5-substituted-furo[2,3-*d*]pyrimidines, *J. Med. Chem.* 41 (1998) 1263–1271.
- [63] A. Gangjee, A.P. Vidwans, A. Vasudevan, S.F. Queener, R.L. Kisliuk, V. Cody, R. Li, N. Galitsky, J.R. Luft, W. Pangborn, Structure-based design and synthesis of lipophilic 2,4-diamino-6-substituted quinazolines and their evaluation as inhibitors of dihydrofolate reductases and potential antitumor agents, *J. Med. Chem.* 41 (1998) 3426–3434.



- [64] A. Gangjee, E. Elzein, S.F. Queener, J.J. McGuire, Synthesis and biological activities of tricyclic conformationally restricted tetrahydropyrido annulated furo[2,3-d]pyrimidines as inhibitors of dihydrofolate reductases, *J. Med. Chem.* 41 (1998) 1409–1416.
- [65] A. Gangjee, A. Vidwans, E. Elzein, J.J. McGuire, S.F. Queener, R.L. Kisliuk, Synthesis, antifolate, and antitumor activities of classical and nonclassical 2-amino-4-oxo-5-substituted-pyrrolo[2,3-d]pyrimidines, *J. Med. Chem.* 44 (2001) 1993–2003.
- [66] M.C. Broughton, S.F. Queener, *Pneumocystis carinii* dihydrofolate reductase used to screen potential antipneumocystis drugs, *Antimicrob. Agents Chemother.* 35 (1991) 1348–1355.
- [67] L.-C. Chio, S.F. Queener, Identification of highly potent and selective inhibitors of *Toxoplasma gondii* dihydrofolate reductase, *Antimicrob. Agents Chemother.* 37 (1993) 1914–1923.
- [68] J.P.P. Stewart, MOPAC, v. 6.0, Quantum Chemistry Program Exchange, Program 455, Indiana University, Bloomington, IN.
- [69] J.P.P. Stewart, MOPAC: a semiempirical molecular orbital program, *J. Comput.-Aided Mol. Des.* 4 (1990) 1–105.
- [70] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.P.P. Stewart, AM1: a new general purpose quantum mechanical molecular model, *J. Am. Chem. Soc.* 107 (1985) 3902–3909.
- [71] C. Aleman, F.J. Luque, M. Orozco, Suitability of the PM3-derived molecular electrostatic potentials, *J. Comput. Chem.* 14 (1993) 799–808.
- [72] P.C. Jurs, J.T. Chou, M. Yuan, Studies of chemical structure-biological activity relations using pattern recognition, in: E.C. Olsen, R.E. Christoffersen (Eds.), *Computer-Assisted Drug Design*, American Chemical Society, Washington, DC, 1979.
- [73] A.J. Stuper, W.E. Brugger, P.C. Jurs, *Computer-Assisted Studies of Chemical Structure and Biological Function*, Wiley, New York, 1979.
- [74] X. Lu, J.W. Ball, S.L. Dixon, P.C. Jurs, Quantitative structure–activity relationships for toxicity of phenols using regression analysis and computational neural networks, *Environ. Toxicol. Chem.* 13 (1994) 841–851.
- [75] S.K. Kachigan, *Statistical Analysis*, Radius Press, New York, 1986.
- [76] A.T. Balaban, Highly discriminating distance-based topological index, *Chem. Phys. Lett.* 89 (1982) 399–404.
- [77] L.B. Kier, Distinguishing atom differences in a molecular graph shape index, *Quant. Struct. Act. Relat.* 5 (1986) 7–12.
- [78] L.B. Kier, L.H. Hall, *Molecular Connectivity in Structure–Activity Analysis*, Wiley, 1986.
- [79] L.B. Kier, L.H. Hall, An electrotopological-state index for atoms in molecules, *Pharma. Res.* 7 (1990) 801–807.
- [80] L.B. Kier, L.H. Hall, Intermolecular accessibility: the meaning of molecular connectivity, *J. Chem. Inf. Comput. Sci.* 40 (2000) 792–795.
- [81] S. Liu, C. Cao, Z. Li, Approach to estimation and prediction for normal boiling point (NBP) of alkanes based on a novel molecular distance-edge (MDE) vector,  $\lambda$ , *J. Chem. Inf. Comput. Sci.* 38 (1998) 387–394.
- [82] H. Wiener, Structural determination of paraffin boiling points, *J. Am. Chem. Soc.* 69 (1947) 17.
- [83] A.R. Katritzky, L. Mu, V.S. Lobanov, M. Karelson, Correlation of boiling points with molecular structure. 1. A training set of 298 diverse organics and a test set of 9 simple inorganics, *J. Phys. Chem.* 100 (1996) 10400–10407.
- [84] R.S. Pearlman, Molecular surface areas and volumes and their use in structure/activity relationships, in: S.H. Yalkowsky, A.A. Sinkula, S.C. Valvani (Eds.), *Physical Chemical Properties of Drugs*, Marcel Dekker, New York, 1980.
- [85] T.R. Stouch, P.C. Jurs, A simple method for the representation, quantification, and comparison of the volumes and shapes of chemical compounds, *J. Chem. Inf. Comput. Sci.* 26 (1986) 4–12.
- [86] D.T. Stanton, P.C. Jurs, Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure–property relationship studies, *Anal. Chem.* 62 (1990) 2323–2329.
- [87] S.N. Vinogradov, R.H. Linnell, *Hydrogen Bonding*, Nostrand Reinhold (Van), New York, 1971.
- [88] A.I. Vogel, *Textbook of Organic Chemistry*, Chaucer, 1977.
- [89] R. Todeschini, V. Consonni, *Methods and Principles in Medicinal Chemistry*, R. Mannhold, H. Kubinyi, H. Timmerman (Eds.), *Handbook of Molecular Descriptors*, Wiley, Germany, 2000.
- [90] A.K. Ghose, G.M. Crippen, Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure–activity relationships. 2. Modeling dispersive and hydrophobic interactions, *J. Chem. Inf. Comput. Sci.* 27 (1987) 21–35.
- [91] V.N. Viswanadhan, A.K. Ghose, G.R. Revankar, R.K. Robins, Atomic physicochemical parameters for three dimensional structure directed quantitative structure–activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics, *J. Chem. Inf. Comput. Sci.* 29 (1989) 163–172.
- [92] R. Todeschini, V. Consonni, M. Pavan, DRAGON, v. 2.1, Milano Chemometrics and QSAR Research Group, University of Milano, Bicocca, Milano, Italy.
- [93] I.V. Tetko, V.Y. Tanchuk, A.E.P. Villa, Prediction of *n*-octanol/water partition coefficients from PHYSPROP database using artificial neural networks and E-state indices, *J. Chem. Inf. Comput. Sci.* 41 (2001) 1407–1421.
- [94] B.R. Baker, *Design of Active-Site-Directed Irreversible Enzyme Inhibitors*, Wiley, New York, 1967.
- [95] A.-M. Perault, B. Pullman, Structure Electronique et Mode D’Action des Antimetabolites de L’Acide Folique, *Biochim. Biophys. Acta* 52 (1961) 266–280.
- [96] S.F. Zakrzewski, The mechanism of binding of folate analogues by folate reductase, *J. Biol. Chem.* 238 (1963) 1485–1490.
- [97] J.G. Topliss, R.P. Edwards, Chance factors in studies of quantitative-structure property relationships, *J. Med. Chem.* 22 (1979) 1238–1244.
- [98] J.M. Sutter, S.L. Dixon, P.C. Jurs, Automated descriptor selection for quantitative structure–activity relationships using generalized simulating annealing, *J. Chem. Inf. Comput. Sci.* 35 (1995) 77–84.
- [99] B.T. Luke, Evolutionary programming applied to the development of quantitative structure–activity relationships and quantitative structure-property relationships, *J. Chem. Inf. Comput. Sci.* 34 (1994) 1279–1287.
- [100] M.D. Wessel, P.C. Jurs, Prediction of reduced ion mobility constants from structural information using multiple linear regression analysis and computational neural networks, *Anal. Chem.* 66 (1994) 2480–2487.