Journal of Computer-Aided Molecular Design **18:** 309–331, 2004.

Three-dimensional quantitative structure-activity and structure-selectivity relationships of dihydrofolate reductase inhibitors

Jeffrey J. Sutherland[a] & Donald F. Weaver[b,*]
[a]*Department of Chemistry, Queen's University, Kingston, Ontario, Canada K7L 3N6;* [b]*Departments of Medicine (Neurology) and Chemistry and School of Biomedical Engineering, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4J3*

Received 3 March 2004; accepted in revised form 21 July 2004

Key words: 3D-QSAR, CoMSIA, dihydrofolate reductase inhibitors, *Pneumocystis carinii*, rat liver, SIMCA

Summary

Three-dimensional quantitative structure-activity relationship (3D-QSAR) modelling using comparative molecular similarity indices analysis (CoMSIA) was applied to a series of 406 structurally diverse dihydrofolate reductase (DHFR) inhibitors from *Pneumocystis carinii* (pc) and rat liver (rl). X-ray crystal structures of three inhibitors bound to pcDHFR were used for defining the alignment rule. For pcDHFR, a QSAR model containing 6 components was selected using leave-10%-out cross-validation (n= 240, $q^2 = 0.65$), while a 4-component model was selected for rlDHFR (n= 237, $q^2 = 0.63$); both include steric, electrostatic and hydrophobic contributions. The models were validated using a large test set, designed to maximise its diversity and to verify the predictive accuracy of models for extrapolation. The pcDHFR model has $r^2 = 0.60$ and mean absolute error (MAE) = 0.57 for the test set after removing 4 outliers, and the rlDHFR model has $r^2 = 0.60$ and MAE = 0.69 after removing 4 test set outliers. In addition, classification models predicting selectivity for pcDHFR over rlDHFR were developed using soft independent modelling by class analogy (SIMCA), with a selectivity ratio of 2 ($IC_{50,rlDHFR}/IC_{50,pcDHFR}$) used for delimiting classes. A 5-component model including steric and electrostatic contributions has cross-validated and test set classification rates of 0.67 and 0.68 for selective inhibitors, and 0.85 and 0.72 for unselective inhibitors. The predictive accuracy of models, together with the identification of important contributions in QSAR and classification models, offer the possibility of designing potent selective inhibitors and estimating their activity prior to synthesis.

Introduction

The metabolism of folate plays an important role in the biosynthesis of nucleic acid precursors [1]. During the synthesis of purines and thymidylate, the cofactor tetrahydrofolate is oxidised to 7,8-dihydrofolate and subsequently converted back to tetrahydrofolate by the enzyme dihydrofolate reductase (DHFR). The inhibition of DHFR causes the depletion of tetrahydrofolate and disrupts DNA synthesis, leading to cell death. For this reason, DHFR inhibitors such as methotrexate (MTX, Figure 1) have been used as antitumor, antibacterial and antiprotozoan agents [2]. Because MTX and other classical antifolates require an active transport mechanism for their uptake, they are not effective for the treatment of infections caused by the opportunistic pathogens *Pneumocystis carinii* (pc) and *Toxoplasma gondii* (tg) that lack these mechanisms [3, 4]. *P. carinii* and *T. gondii* infections are the principal cause of death in patients with AIDS, and also affect patients with other immune disorders [5]. Trimetrexate (TMQ) and piritrexim (PTX) are potent lipophilic inhibitors of pcDHFR and tgDHFR taken up by passive diffusion, but inhibit mammalian DHFR to a greater extent [6,7]. This results in toxicity to mammalian tissue and requires that PTX or TMQ be co-administered with leucovorin (5-formyl-5,6,7,8-tetrahydrofolate), a

*To whom correspondence should be addressed. Fax: +1-902-494-1310; E-mail: weaver@chem3.chem.dal.ca

reduced folate which is taken up by active transport and protects the host tissue [6–8]. Treatment with leucovorin is costly and subject to serious side effects that may require interruption of treatment. As such, there is great interest in developing potent and selective inhibitors of *P. carinii* and *T. gondii* DHFR.

A large number of antifolates have been synthesised and tested by Queener and colleagues (see list of references in the footnotes of Table A2). Many of these are potent inhibitors of pcDHFR and tg-DHFR, but show moderate selectivity at best (ca. 20×) over the mammalian enzyme, represented by rat liver (rl) DHFR. Three-dimensional quantitative structure-activity relationships (3D-QSAR) attempt to relate the biological activity of compounds to their 3D properties. Using a large series of 406 diverse compounds, we have developed QSAR models explaining the potency of antifolates for pcDHFR and rlDHFR, and classification models explaining selectivity for pcDHFR.

Inhibitors of DHFR have been widely studied by QSAR; one particular series has been frequently used for assessing the predictive ability of novel methods [9–14]. Most of these studies concern only one class of inhibitors (e.g. pyrimidines, quinazolines). Fewer yet consider inhibition of pcDHFR [15,16]. The models presented in this work can be used for predicting the potency and selectivity towards pcDHFR of novel derivatives from multiple structural classes, while providing insights into the protein-inhibitor interactions that modulate these properties. This should facilitate the development of novel selective inhibitors.

Methods

(i) **Dataset**: A set of 723 inhibitors of DHFR has been assembled from the work of Queener et al. In vitro activities for pcDHFR and rlDHFR are reported as IC_{50} values for the inhibition of the enzymatic reduction that converts dihydrofolate to tetrahydrofolate. Using 2-D (structural) fingerprints with the Tanimoto coefficient [17] (T_c) to calculate pairwise similarity of compounds, a subset was selected using the coverage-based diversity algorithm [18] implemented in Cerius2 (Accelrys, Inc., San Diego, CA). This gave 406 compounds for which all pairs have $T_c < 0.975$. The excluded compounds were not used in any model development or evaluation. Some compounds have activities expressed in indeterminate form (e.g. $IC_{50} > 10\ \mu M$); these were assigned to an 'inactive' set used to verify if models can correctly identify inactives. The remaining compounds were divided between training and test sets. Approximately 33% were selected by 'cherry picking' with a maximum dissimilarity algorithm [19,20] and assigned to the test set, with the remaining compounds assigned to the training set. The sets were structured this way to maximise the diversity of the test set and to examine the predictive accuracy of methods when extrapolating outside the training set. The maximum dissimilarity algorithm (the MaxMin function in Cerius2) maximises the minimum squared distance from each compound to all other compounds in the selected subset, with pairwise distances determined using $1-T_c$. The optimisation uses a Monte Carlo procedure [20] that we have coupled to a simulated annealing protocol implemented in T_cl (up to 100000 trial sets per pseudo-temperature, which is lowered in 10% increments from 5000 K to 10 K). The MaxMin function was optimised under restraint, such that the selected compounds have a similar distribution of pcDHFR activities as the complete set. The penalty function was obtained by assigning compounds to 10 evenly separated bins covering the range of pIC_{50} values. The training and test set compounds belong to 11 structural classes with pIC_{50} values ranging from 3.5 to 9.5 for pcDHFR and 3.3 to 9.8 for rlDHFR (Figure 2). The composition of the sets, tabulated in the Appendix, is summarised in Table 1.

(ii) **Alignment**: The crystal structures for several ligand-pcDHFR complexes have been used for aligning compounds for 3D-QSAR analysis. The pcDHFR-

Table 1. Description of training and test sets.

	pcDHFR	rlDHFR
train[a]	240	237
test	119	126
inact	47	36
train sel/unsel[b]	33/204	–
test sel/unsel	19/97	–
train $< T_c >$[c]/min$\{T_c\}$[d]	0.47/0.91	0.47/0.91
test $< T_c >$/min$\{T_c\}$	0.43/0.75	0.43/0.77
test–train[e] $< T_c >$/min$\{T_c\}$	0.44/0.86	0.44/0.86

[a]The number of training, test, inactive set compounds; 3 training set compounds, 3 test set and 1 inactive compounds for pcDHFR have no data for rlDHFR activity; other differences in set size arise from inactivity at one enzyme but activity at the other; [b]the number of selective and unselective compounds in each set; [c]the average pairwise value of T_c for the set; [d]the average value of T_c calculated over pairs of most similar compounds; [e]calculated using test set-training set pairs.

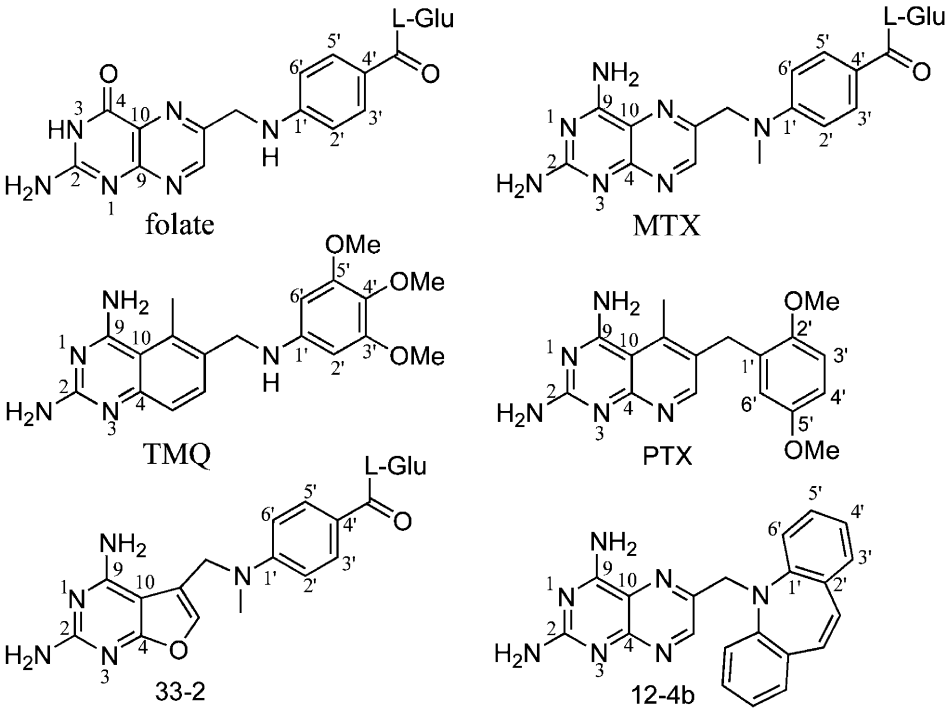


Figure 1. DHFR inhibitors. Numbers for MTX, TMQ, PTX, 33-2 and 12-4b show mapping of atoms to folate used for aligning compounds.

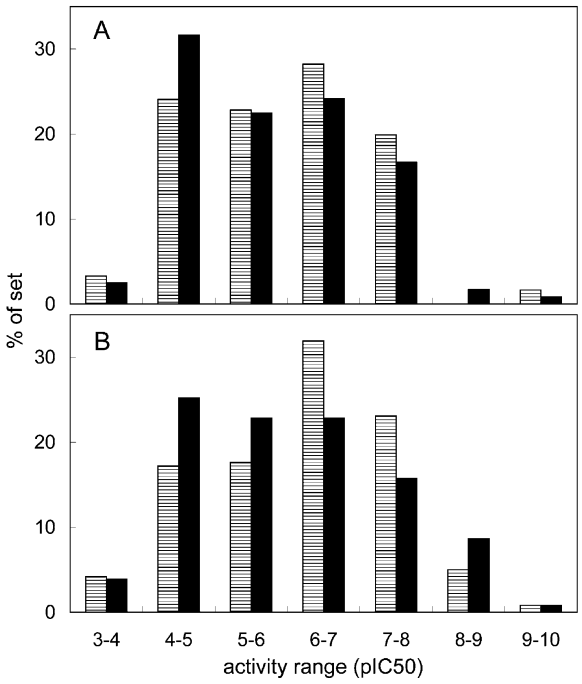


Figure 2. Distribution of activities of training set (hashed) and test sets (solid) for (a) pcDHFR and (b) rlDHFR.

bound structures of 33-2 (PDB 1daj) and 12-4b (PDB 1klk) were rigidly superposed onto folate (PDB 1cd2) using the atom mappings implied by folate labels shown in Figure 1; the root-mean-square deviations (RMSDs) with respect to folate are 0.3 Å for 33-2 and 0.4 Å for 12-4b when calculated with the atoms used for the superposition. All other compounds were flexibly fit onto the most similar of these three templates with the alignment module in Cerius2, using the pyridine (having un-primed numbers in Figure 1), phenyl ring (having primed numbers in Figure 1) and L-glutamate group (when present) for defining the superposition. Flexible fitting allows rotation of torsions to minimise the RMSD calculated from the atoms used for superposition, subject to bump-checking for ensuring that modified conformations are reasonable. With the positions of the pyridine, phenyl and L-glutamate fixed, structures were relaxed using the CFF97 force field [21] in Cerius2 with no cut-off for van der Waals or electrostatic interactions (other parameters default). Substituents on the phenyl ring were placed in a consistent fashion, with preference for the solvent-exposed edge of the phenyl ring: the 2′ position was occupied before 6′ and the 3′ position before 5′, with the ortho position taking precedence over the

meta position when both are substituted. This placement of substituents is consistent with that observed for the methoxy groups in 15-11 (PDB 1ly3) and the dibenzo[b,f]azepine ring in 12-4b. The same alignment was used for QSAR modelling of both pcDHFR and rlDHFR. The active site geometry of pc and human (for which rat liver is a surrogate) enzymes is very similar, and the heavy-atom RMSDs for several inhibitors complexed to pc and human DHFR following rigid superposition are small (15-11: 1ly3 vs. 1boz, 0.4 Å; folate: 1cd2 vs 1drf, 0.7 Å; 33-2: 1daj vs. 1hfp, 0.6 Å; the latter two are reduced by half if the solvent-exposed glutamate side chain is excluded). Some aligned representative inhibitors are shown in Figure 3.

(iii) **QSAR model development**: Comparative molecular similarity indices analysis [22] implemented in Sybyl 6.81 (Tripos Inc., St. Louis, MO) was used for developing 3D-QSAR models for pcDHFR and rlDHFR. The method is conceptually similar to CoMFA [23], in which steric and electrostatic fields are calculated at regularly spaced grid points of a lattice into which a series of aligned molecules are embedded. Partial least squares (PLS) is used to obtain a statistical model relating field values to observed activities. In CoMSIA, Coulomb and Lennard-Jones potentials are replaced with smooth gaussian potentials having the form

$$A^{q}_{F,k}(j) = -\sum_{i} w_{probe,k} w_{i,k} \exp(-\alpha r^{2}_{iq})$$

in which the 'field' k at grid point q for molecule j is obtained by adding the probe-atom potential for each atom i in the molecule; the coefficients w are the partial charge or atomic radius raised to the 3rd power for the electrostatic and steric fields, respectively, α is a smoothing parameter set to 0.3 and r_{iq} is the distance (Å) between atom i and grid point q. Similar definitions hold for hydrophobic [24] and hydrogen-bond donor/acceptor fields. The probe has a 1 Å radius, and has charge, hydrophobicity and hydrogen-bond properties of +1. After determining net formal charges by deprotonating carboxylic acids, scaled MNDO ESP-fit partial charges [25] were calculated with MOPAC 6.0 using atomic coordinates obtained by energy minimising the aligned molecules with the MMFF94S force field and MAXIMIN2 routine in Sybyl (200 steps, other parameters default). A lattice with 2 Å grid spacing and extending at least 4 Å in each direction beyond the aligned molecules was used for calculating fields. PLS analyses were performed after block-scaling the fields (CoMFA-standard scaling). The statistical significance of models was evaluated by 'leave-one-out' (LOO; using the SAMPLS routine) and 'leave-several-out' (LSO) cross-validation. For the latter, the training set was randomly divided into 10 groups that were used in turn as prediction sets for models derived with the other 9 groups. This was repeated for 5 cycles, giving estimates of q^2 and s_{PRESS} based on 50 models. The optimal number of components was determined by selecting the smallest s_{PRESS} value. Both the final PLS model and LSO cross-validation models were derived by setting the 'minimum-σ' standard deviation threshold to 1.0 kcal/mol.

(iv) **Classification model development**: Classification models for selective inhibition of pcDHFR were developed using Soft Independent Modelling by Class Analogy (SIMCA). The SIMCA method [26,27] applies principal component analysis (PCA) separately to each class of objects, and uses the principal components (PCs) to define (hyper)volumes in the descriptor space. Classification of test objects is achieved by comparing the orthogonal projection distance to each class model with the PCA residuals of objects within each class. The Sybyl implementation used in this work differs somewhat, as discussed in a recent evaluation of its application in drug design [28]. As for QSAR analyses, CoMSIA fields were block-scaled and 'minimum-σ' was set to 1.0 kcal/mol. Inhibitors selective for pcDHFR were assigned to class 1, while unselective inhibitors were assigned to class 0. The threshold $IC_{50,rlDHFR}/IC_{50,pcDHFR} = 2.0$ was used for determining class membership (no compounds with indeterminate pcDHFR activities were used). This gave 33 selective inhibitors for the training set, 20 for the test set (Table 1).

Results and discussion

(i) **QSAR models**: CoMSIA models were developed using three combinations of steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor fields. For pcDHFR, the addition of hydrophobic fields to the standard steric and electrostatic fields results in models with greater internal predictive accuracy, as measured by cross-validation (Table 2). The addition of donor and acceptor fields does not lead to improved statistics. Because of correlations among the fields (especially the donor/acceptor fields with the electrostatic field), additional fields are useful

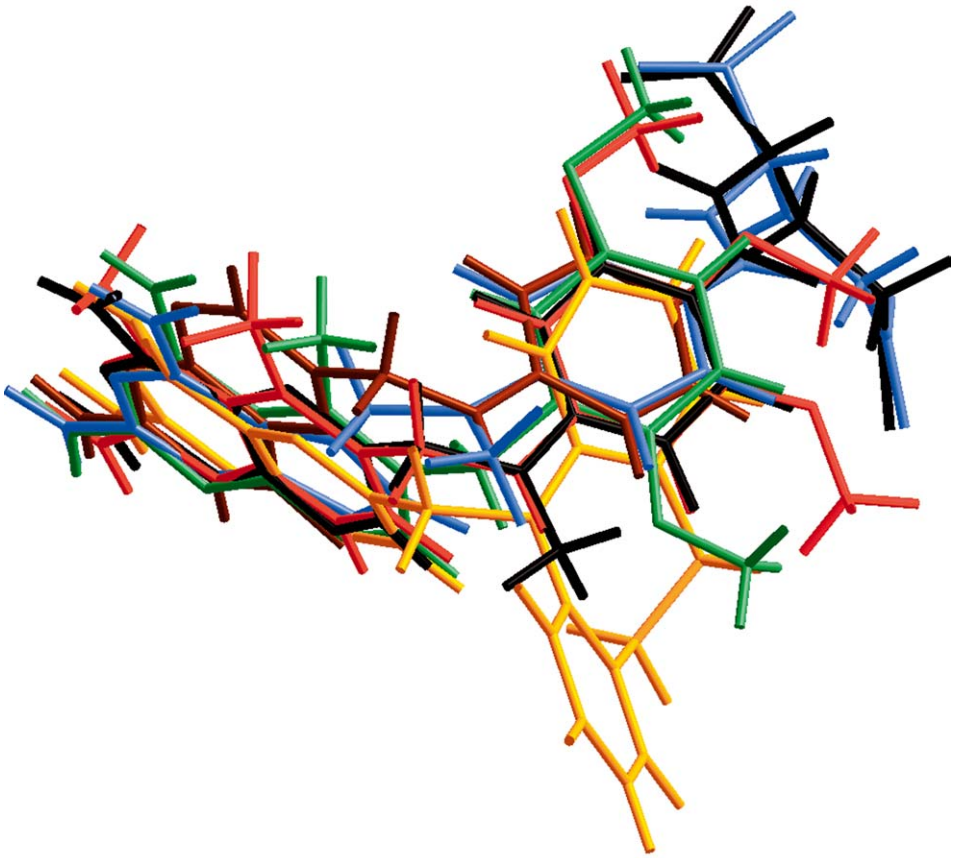


Figure 3. Alignment of representative compounds: folate (brown), MTX (black), TMQ (red), PTX (green), 33-2 (blue), and 12-4b (orange).

Table 2. QSAR cross-validation statistics.

		pcDHFR			rlDHFR		
		ste/ele	ste/ele/pho	all	ste/ele	ste/ele/pho	all
LOO	q^2	0.65	0.66	0.65	0.64	0.64	0.65
	s_{PRESS}	0.70	0.68	0.69	0.77	0.77	0.76
	comps	7	7	6	6	4	4
LSO	q^2	0.63	0.65	0.64	0.62	0.63	0.64
	s_{PRESS}	0.71	0.69	0.71	0.76	0.77	0.76
	comps	6	6	4	6	4	4

q^2 is the cross-validated correlation coefficient; s_{PRESS} is the standard error from cross-validation; LOO and LSO denote 'leave-one-out' and 'leave-several-out' cross-validation, respectively; fields are denoted by three-letter abbreviations: **ste**ric, **ele**ctrostatic, hydro**pho**bic or **all** three fields plus the donor and acceptor fields.

for enhancing the interpretation of the models, but generally do not increase substantially the predictive accuracy of models [22, 29]. For this study, the contour maps for donor/acceptor fields only duplicated features clearly conveyed by the electrostatic field contour map; as such, we use only the steric, electrostatic and hydrophobic fields for deriving the final models. The leave-one-out (LOO) and leave-several-out (LSO) procedures give similar estimates of statistical significance and optimal complexity (i.e. the number of PLS components). We use the number of components identified with LSO cross-validation for the final model.

Having identified the optimal number of PLS components, the steric, electrostatic and hydrophobic fields were used to develop CoMSIA models from the complete training set (Table 3). The models were used for making predictions on a test set designed to maximise its diversity. Accordingly, it allows assessment of the ability of models to generalise. The models have good predictive accuracy for both pcDHFR and rlDHFR (Table 3). Removal of a small number of outliers causes r^2_{test} to increase substantially (outliers have residuals greater than 2 times the standard deviation of all test set residuals). For pcDHFR, 6-3, 18-5, 22-20 and 22-21 are outliers; for rlDHFR, 7-7, 22-21, 29-2-PTX and 40-T-2-10 are outliers. Predicted vs. actual activities for the training and test sets are shown in Figure 4.

It is instructive to examine the predictions of CoMSIA models for compounds having IC_{50} values expressed in indeterminate form (e.g. $IC_{50} > 10\ \mu M$). This arises when a compound has low affinity for the receptor and cannot be dissolved at the concentration required for accurate determination of its IC_{50} value. Both models correctly predict inactives, with better performance obtained when the threshold for inactivity is decreased from 10 to 1 μM (Table 4).

Contour maps of the '$\sigma\times$ coeff' type that identify the dominant contributions to the model are shown in Figure 5 for the pcDHFR and rlDHFR models. Initial contour levels were set to default values (i.e. at 20% and 80% contribution), and adjusted iteratively to enhance their interpretability. In this graphical representation, grid points at which steric bulk causes an increase in activity are enclosed in green contours (i.e. they correspond to positive PLS coefficients) while grid points for which steric bulk causes a decrease in activity are enclosed in yellow contours (negative PLS coefficients). Blue contours usually identify regions for which ligand atoms bearing positive charge enhance activity, while red contours identify regions at which ligand atoms bearing negative charge enhance activity. For the hydrophobic field, activity is enhanced when hydrophobicity at brown-contoured regions increases, and reduced when hydrophobicity at pink-contoured regions increases. The use of the standard deviation of field values at a given grid point (σ in '$\sigma\times$ coeff') ensures that only regions for which there is significant variability of field values are identified as dominant contributions.

To better understand the dominant contributions in the QSAR models, we have examined in great detail the structural features that give rise to the observed contours for the pcDHFR model. We have tested assumptions regarding the effects represented by contours by developing additional models without the compounds that are thought to give rise to the ob-

Table 3. Summary statistics for CoMSIA models derived using steric, electrostatic and hydrophobic fields.

	pcDHFR	rlDHFR
r^2	0.80	0.76
s	0.52	0.63
F	157.5	186.9
q^2	0.65	0.63
s_{PRESS}	0.69	0.77
components	6	4
fraction		
steric	0.18	0.19
electrostatic	0.43	0.37
hydrophobic	0.39	0.44
r^2test	0.47 (0.60)	0.52 (0.60)
MAE_{test}	0.63 (0.57)	0.75 (0.69)
$n_{outliers,test}$	4	4

r^2 is the correlation coefficient, s the standard error of prediction; F is Fischer's test of statistical significance; fraction indicates the relative importance of the fields in the model; r^2_{test} is the correlation coefficient calculated from the test set, using the average activity from the training set; MAE_{test} is the mean absolute error on the test set; values in parentheses reflect the removal of n outliers.

Table 4. Percentage of inactives that are correctly predicted.

	pcDHFR	rlDHFR
$pIC_{50} < 5$	62	67
$pIC_{50} < 6$	100	94

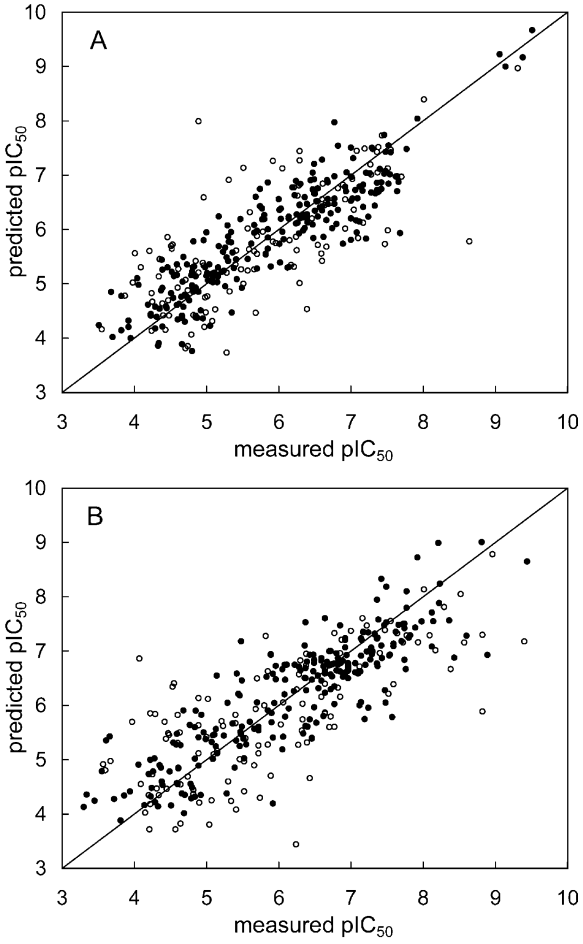


Figure 4. Scatter plots showing predicted vs. measured activities, with training set compounds shown using filled circles and test set compounds using empty circles. (A) pcDHFR, (B) rlDHFR.

served feature. For clarity, the following discussion refers to contour labels used in Figure 5A.

1. For pyrrole derivatives of folate (family N.3), the amine nitrogen has a less negative charge than the corresponding pyrimidine nitrogen in other inhibitors and results in lower activity (e.g. 43-10 has IC_{50} 22 μM, compared to TMQ with IC_{50} 0.042 μM; in all comparisons, the derivatives are identical except for the structural feature under consideration; references to positions in structures correspond to those for MTX, shown in the inset of Figure 5A).
2. There is a preference for hydrophobic atoms at position 8. For example, TMQ has IC_{50} 0.042 μM compared to 0.086 μM for 23-4a; the CH fragment of TMQ has hydrophobicity [24] of 0.34 while the pyridine N of 23-4a has hydrophobicity −0.11. Among the inhibitors of family N, the furan derivative 33-2 has IC_{50} 0.035 μM compared to 0.044 μM for the pyrrole derivative 35-10; the hydrophobicity of O in furan is 0.27, and that of NH in pyrrole is 0.09.
3. For 19 training set compounds lacking an aromatic ring separated by a flexible linker (i.e. within families F.2, G.2, H.2, J.3 and N.2; henceforth referred to as a linker-phenyl group), hydrophobic substituents at ring position 7 enhance activity: 6-22 (CF_3, hydrophobicity 1.06) has IC_{50} 31 μM compared to 65 μM for 6-21 (Me, hydrophobicity 0.65).
4. Compounds having an electronegative pyridine N at position 5 (e.g. 9-12q, IC_{50} 7.0 μM) have lower activity than compounds containing CH at the same position (e.g. 9-15c, IC_{50} 2.0 μM).
5. Related to 4, compounds with greater hydrophobicity at position 5 have higher activity: the hydrophobicity contributions of N, CH, C-Me, C-Cl at

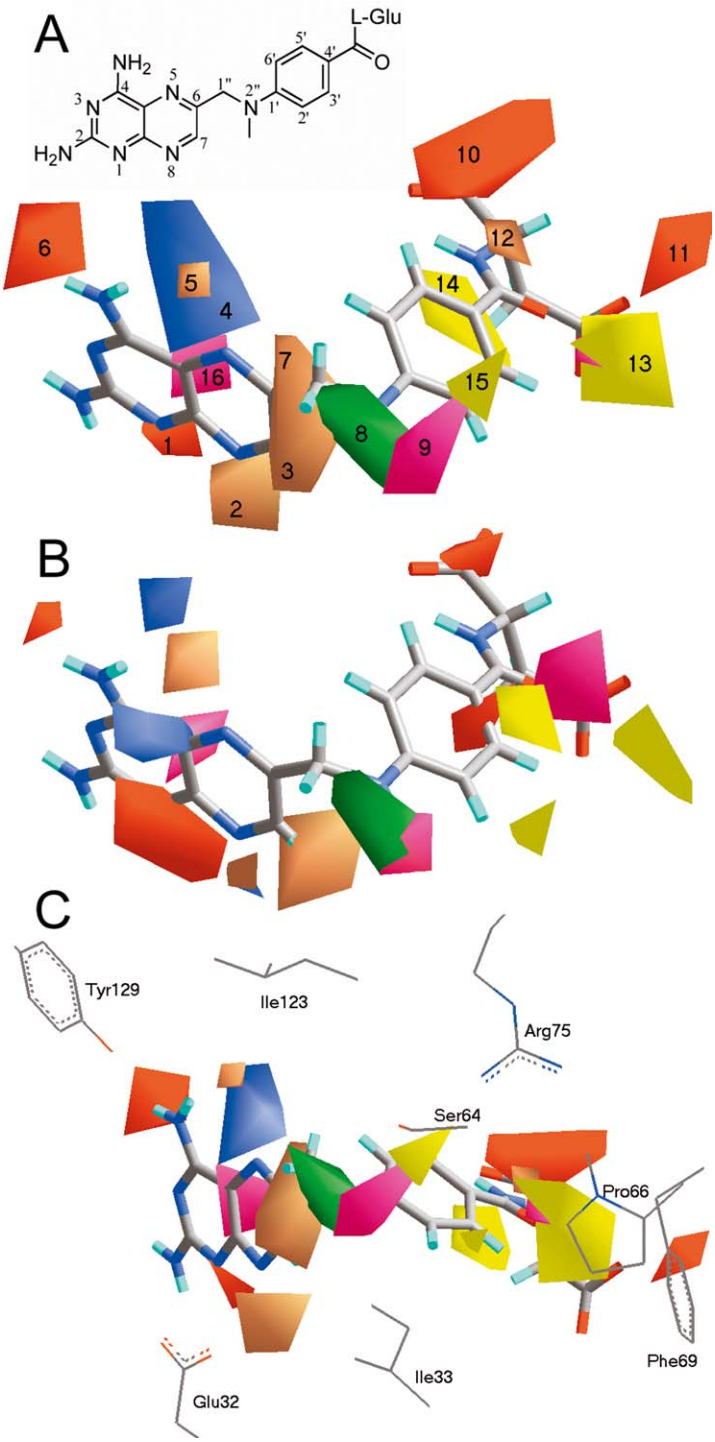


Figure 5. CoMSIA 'σ× coeff' contour maps showing important contributions in models. (A) pcDHFR with steric field contoured at 0.045 (green) and −0.004 (yellow), electrostatic field contoured at 0.045 (blue) and −0.044 (red) and hydrophobic field contoured at 0.058 (brown) and −0.033 (pink); (B) rlDHFR with steric field contoured at 0.045 (green) and −0.003 (yellow), electrostatic field contoured at 0.024 (blue) and −0.044 (red) and hydrophobic field contoured at 0.058 (brown) and −0.034 (pink); (C) pcDHFR map with active site residues from PDB 1daj. MTX is shown for reference.

position 5 are −0.11, 0.34, 0.81, 0.86. The derivatives 9-12q (N), 9-15c (CH) and 23-4a (C-Me) have IC_{50} values of 7.0 μM, 2.0 μM and 0.086 μM; the 5-Cl derivative 14-29 has IC_{50} 0.033 μM, compared to 0.042 μM for TMQ.

6. Compounds with short linkers (2 intervening bonds) between the bicyclic moiety (i.e. the pteridine in MTX) and the phenyl ring have their bicyclic moiety shifted with respect to that in compounds with longer linkers. As such, the amine at position 4 is also shifted. Compounds with short linkers generally have lower activity than compounds with longer linkers, although there are exceptions (e.g. PTX has IC_{50} 0.031 μM, while its 3-bond derivative, 24-2a, has IC_{50} 0.046 μM). The outliers PTX, 22-20 and 22-21 all have short linkers.
7. Compounds with hydrophobic groups at position 1″ of the linker have higher activity. For example, the derivative 25-11 with linker -$NHCH_2$- has IC_{50} 14.1 μM compared to 2.0 μM for 9-15c with linker -CH_2NH-. Also, for compounds lacking a linker-phenyl group, hydrophobic substituents at position 6 enhance activity: 6-15 (CF_3) has IC_{50} 2.7 μM compared to 20 μM for 6-14 (Me).
8. Compounds with substituents at linker position 2″ have higher activity. For example, MTX (Me) has IC_{50} 0.0013 μM, compared to 0.028 μM for 1-98535 (H) and 0.00035 μM for 1-169531 (Et).
9. Hydrophilic groups at 2″ are preferred; for example, 1-235777 having an NH group at that position (hydrophobicity 0.10) has IC_{50} 0.13 μM, compared to 4.2 μM for 9-12c having S (hydrophobicity 0.61) or 97 μM for 9-12t having CH_2 (hydrophobicity 0.40).
10. Compounds with glutamate at the 4′ position of MTX have higher activity; the contour arises from the α-carbon COOH.
11. Compounds with D-glutamate at the 4′ position of MTX have higher activity than L-glutamate derivatives (e.g. 1-117356, the D-glutamate analog of MTX has IC_{50} 0.00088 μM compared to 0.0013 μM for MTX). This is essentially a 2-level variable, with only 3 D-glutamate derivatives occupying the corresponding position in the alignment space.
12. Compounds with hydrophilic carbonyl substituents at the 4′ position have lower activity (e.g. 1-131463 having COOH).
13–15. Compounds with steric bulk within these contours have lower activity. For example, 41-7b (contour 13), 38-10 (contour 14) and 44-17 (contour 15) have phenyl rings at those positions in the alignment space. Contour 15 arises from compounds without a linker-phenyl group.

While our analysis of the pcDHFR contour map helps to understand the origin of important variables in the model, one must remember that the model weighs multiple contributions simultaneously and that analysing contributions independently may lead to erroneous conclusions. For example, contours 2 and 16 are related but have opposite meaning. Attempts to interpret contour 16 by itself would lead one to believe that hydrophilic groups in the bicylic ring increase activity. The contour disappears if thiophene derivatives are excluded. In contrast to other elements, there is no differentiation of aromatic and aliphatic sulfur in the method of Viswanadhan et al. [24] The hydrophobicity of S in thiophene may be overestimated. The same feature also demonstrates another important aspect of any field-based 3D-QSAR. The contoured region is not necessarily located at the position of substitution that gives rise to it (position 8 of MTX), because field values at a particular location are influenced by all atoms in the molecule.

A number of features in the contour map for pcDHFR are consistent with the position and identity of active site residues (Figure 5C). Contours 2 and 5, indicating that hydrophobicity increases activity, are near Ile33 and Ile123, respectively (residue numbers are those used in the PDB structure 1daj). Contours 1 and 6, which reflect electrostatic contributions to affinity, are located near Glu32 and Tyr129, respectively. Contour 10, indicating the favourable effect of negatively charged groups (i.e. COOH in glutamate), is near Arg75. Contours 13 and 15, indicating that steric bulk decreases activity, are located near Pro66 and Ser64, respectively.

(ii) **Selectivity classification models.** We have developed SIMCA classification models relating selectivity for pcDHFR to the CoMSIA fields of compounds. Because of the high correlation of pIC_{50} values for pcDHFR and rlDHFR ($r = 0.89$), it was not possible to develop regression models using differences (or quotients) of activities as the dependent variable. Indeed, if the rlDHFR model is used to predict pcDHFR activity, the value of r^2_{test} is 0.55 when 1-184692, 6-3, 18-5, 22-25 and 40-T-2-1 are excluded. Instead, the classification model simply makes pre-

Table 5. SIMCA cross-validation classification rates.

	ste/ele	ste/ele/ phob	all
C_1	0.67	0.67	0.67
C_0	0.85	0.82	0.80
comps	5	5	5

C_1 and C_0 are the classification rates for selective and unselective inhibitors, respectively.

Table 6. Classification rates for SIMCA selectivity models derived using steric and electrostatic fields.

train C_1	0.79
train C_0	0.88
test C_1	0.68
test C_0	0.72
components	5

dictions for 'selective' (category 1) or 'unselective' (category 0). As for the QSAR models, we have examined the use of three combinations of five field types and used 'leave-several-out' cross-validation to select the most predictive combination of fields and the optimal number of SIMCA components (Table 5). The classification rate, or fraction of compounds belonging to a particular class that are correctly classified, is used for assessing the performance of models. The number of components that maximise the classification rate for the less well predicted class (always the 'selective' class for this work) was chosen. The final model was developed using only the steric and electrostatic fields with 5 SIMCA components (Table 6).

For SIMCA models, the 'discriminating power' of each variable (i.e. the value at each grid point) is the ratio of sum of squared residuals when the variable is fit to the unselective class and that when fit to the selective class. Variables with larger values for the discriminating power are more effective at discriminating between classes. However, just as PLS coefficients are multiplied by the standard deviation of field values at a given grid point, it is necessary to take into account the magnitude of the difference between the field values of two classes. For the purpose of determining important contributions in the model, we contour using the difference of class means (selective − unselective) multiplied by the discriminating power for a given field at a grid point. The maps are analogous to the familiar 3D-QSAR '$\sigma\times$ coeff' maps, and use the same colour scheme for identifying favourable contributions to selectivity.

The contour map (Figure 6) indicates 3 important contributions to selectivity. The largest contribution (as determined by the levels used for contouring; see the caption of Figure 6) to selectivity results from the placement of steric bulk at the 3′ position (referring to MTX). Three selective training set inhibitors that have bulky aromatic groups within this contour are 12-4a (5H-dibenzo[b,f]azepine), 38-11 (biphenyl), and 41-4d (9H-fluorene). A less important steric contribution favours the absence of steric bulk between positions 4′ and 5′. A contribution favouring uncharged groups at position 8 explains the selectivity of compounds lacking a linker-phenyl system; all the compounds with bulky groups placed within the green contoured region actually have electronegative N and O atoms at position 8. It would be interesting to determine if inhibitors having both a bulky group within the green contoured region and less electronegative atoms at position 8 have greater selectivity.

A comparison of the active sites of pcDHFR (PDB 1daj) and human DHFR (PDB 1hfp) complexed with 33-2 reveals a high degree of similarity, underscoring the difficult task of identifying selective inhibitors (both were crystallised by the same group, hence the choice of inhibitor). Residues within 5 Å of 33-2 that differ between the enzymes are Phe (pc) vs. Asn (human) at residue 69, Ile (pc) vs. Gly (human) at residue 33, Lys (pc) vs. Gln (human) at residue 37 and Ile (pc) vs. Val (human) at position 123 (all residue numbers refer to 1daj). The features from the SIMCA contour map are in reasonable agreement with the first two residue substitutions (Figure 6). The pc enzyme allows interactions between Phe69 and bulky aromatic groups near the green-contoured region. A preference for uncharged groups at ring position 8 of MTX is consistent with the presence of a nearby Ile33.

The threshold for selectivity was chosen to be $IC_{50,rlDHFR}/IC_{50,pcDHFR} = 2.0$; using a higher threshold results in too few selective inhibitors. Others have used the threshold 1.0 [16]. While classification statistics are improved to 0.70 and 0.78 for C_1 and C_0 (perhaps because of greater representation in the 'selective' category), few would argue that an inhibitor having equal affinity at both targets is selective. Indeed, the contour maps obtained using the higher threshold bear little semblance to those presented here, and so the correspondence to amino acid substitutions in the organisms is lost.

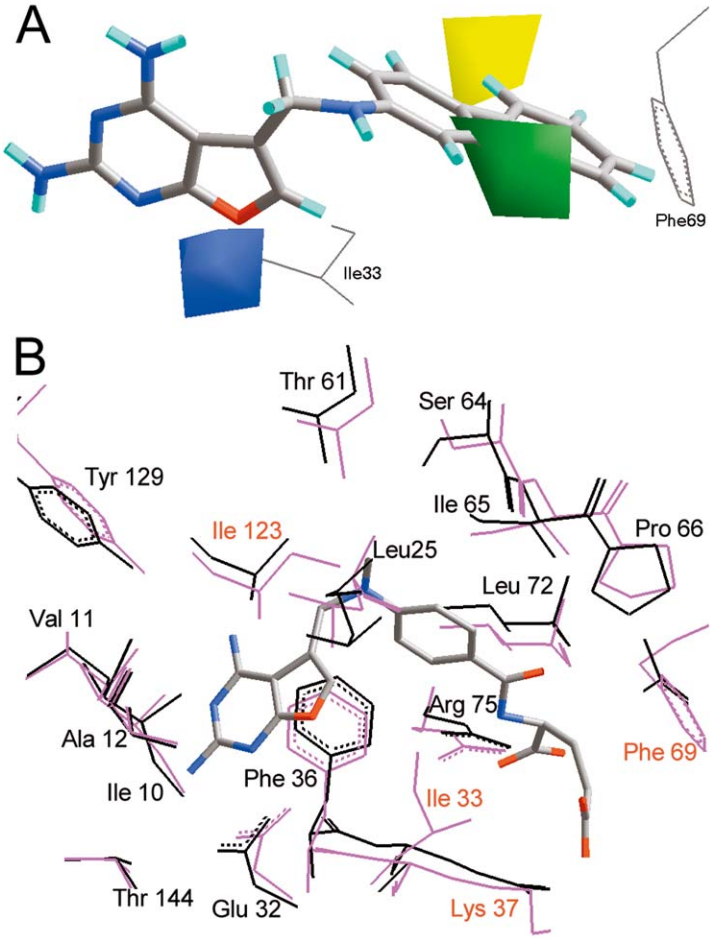


Figure 6. (A) SIMCA '(<sel>−<unsel>) × discrim. power' contour map for selectivity towards pcDHFR with steric field contoured at 22.0 (green) and −9.0 (yellow), electrostatic field contoured at 5.0 (blue) and −4.0 (red − no contour). The selective inhibitor 41-4d is shown for reference. (B) Residues within 5 Å of the inhibitor 33-2 in the pc (magenta; PDB 1daj) and human (black; PDB 1hfp) active sites after superposition using the inhibitor. Residues that differ between the enzymes are labelled in red.

As noted above, DHFR inhibitors have been widely studied by QSAR methods. However, only the study of Mattioni and Jurs [16] using neural networks and 2D descriptors addresses pcDHFR activity for a diverse set of inhibitors. The compounds and activities used in the present work were obtained from the publications of the same research group. A direct comparison of QSAR and classification statistics is not possible, as they have used a substantially smaller test set ($\leq$ 57 compounds) selected randomly from multiple activity bins. Their QSAR models have r^2_{test} and rms_{test} values of 0.64 and 0.66 for pcDHFR, and 0.72 and 0.64 for rlDHFR, while a classification model predicting selectivity for pcDHFR has classification rates for selective and unselective inhibitors of 0.81 and 0.78 (using a selectivity ratio of 1 as threshold). When using the training / test set from the present work, 2D-QSAR models obtained with a similar methodology have lower predictive accuracy than CoMSIA models (unpublished results). The principal advantage of the CoMSIA models presented in this work is the insight obtained from contour maps. Interpreting 2D QSAR models is a difficult task. In light of the high correlation between affinities for pcDHFR and rlDHFR, the small number of descriptors used in both 2D models underscores this difficulty.

Conclusions

In this paper, 406 structurally diverse compounds have been used to generate and validate 3D-QSAR mod-

els predicting affinity towards DHFR from *P. carinii* (pcDHFR) and rat liver (rlDHFR). Steric, electrostatic and hydrophobic CoMSIA fields were calculated using an alignment rule deduced from the crystal structures of three inhibitors complexed with pcDHFR. Models for both enzymes exhibit good predictive accuracy, as assessed by their performance on a large, designed test set. In addition, SIMCA classification models were developed using steric and electrostatic CoMSIA fields for predicting the selectivity of inhibitors for pcDHFR over rlDHFR. The principal features of QSAR and classification models were found to be consistent with active site residues. The predictive accuracy of QSAR and classification models, together with the structural insights gleaned from them, should aid in the development of novel inhibitors used for the treatment of *P. carinii* infections in immuno-compromised individuals.

Acknowledgements

D.F.W. acknowledges support from the Natural Sciences and Engineering Research Council (NSERC), the Canadian Institutes of Health Research (CIHR) and the Canada Research Chairs Program. J.J.S. acknowledges support from a CIHR doctoral research award.

References

1. Blakley, R.L. and Benkovic, S.J. Folates and Pterins, Wiley, New York, 1984.
2. Kisliuk, R.L., In Sirotnak, F.M., Burchall, J.J., Ensminger, W.D. and Montgomery, J.A. (Eds.), Folate Antagonists as Therapeutic Agents, Academic Press, New York, 1984, pp. 1–68.
3. Warren, E., George, S., You, J. and Kazanjian, P., Pharmacotherapy, 17 (1997) 900.
4. Behbahani, R., Moshfeghi, M. and Baxter, J.D., Ann. Pharmacother., 29 (1995) 760.
5. Kovacs, J.A., Hiemenz, J.W., Macher, A.M., Stover, D., Murray, H.W., Shelhamer, J., Lane, H.C., Urmacher, C., Honig, C., Longo, D.L., Parker, M.M., Natanson, C., Parrillo, J.E., Fauci, A.S., Pizzo, P.A. and Masur, H., Ann. Intern. Med., 100 (1984) 663.
6. Allegra, C.J., Kovacs, J.A., Drake, J.C., Swan, J.C., Chabner, B.A. and Masur, H., J. Clin. Invest., 79 (1987) 478.
7. Kovacs, J.A., Allegra, C.J., Swan, J.C., Drake, J.C., Parrillo, J.E., Chabner, B.A. and Masur, H., Antimicrob. Agents Chemother., 32 (1988) 430.
8. Kovacs, J.A., Allegra, C.J., Beaver, J., Boarman, D., Lewis, M., Parrillo, J.E., Chabner, B. and Masur, H., J. Infect. Dis., 160 (1989) 312.
9. Kroemer, R.T. and Hecht, P., J. Comput.-Aided Mol. Des., 9 (1995) 205.
10. Kroemer, R.T. and Hecht, P., J. Comput.-Aided Mol. Des., 9 (1995) 396.
11. Hopfinger, A.J., Wang, S., Tokarski, J.S., Jin, B.Q., Albuquerque, M., Madhav, P.J. and Duraiswami, C., J. Am. Chem. Soc., 119 (1997) 10509.
12. Dunn, W.J. and Hopfinger, A.J., Perspect. Drug Discov. Design, 12 (1998) 167.
13. Crippen, G.M., J. Comput. Chem., 20 (1999) 1577.
14. Agrafiotis, D.K., Cedeno, W. and Lobanov, V.S., J. Chem. Inf. Comput. Sci., 42 (2002) 903.
15. Marlowe, C.K., Selassie, C.D. and Santi, D.V., J. Med. Chem., 38 (1995) 967.
16. Mattioni, B.E. and Jurs, P.C., J. Mol. Graph. Model., 21 (2003) 391.
17. Willett, P. and Winterman, V., Quant. Struct.-Act. Relat., 5 (1986) 18.
18. Clark, R.D., J. Chem. Inf. Comput. Sci., 37 (1997) 1181.
19. Lajiness, M., Johnson, M.A. and Maggiora, G.M., In Fauchere, J.L. (Ed.) QSAR: Quantitative Structure-Activity Relationships in Drug Design, Alan R. Liss Inc., New York, 1989, pp. 173–176.
20. Hassan, M., Bielawski, J.P., Hempel, J.C. and Waldman, M., Mol. Div., 2 (1996) 64.
21. Maple, J.R., Dinur, U. and Hagler, A.T., Proc. Natl. Acad. Sci. USA, 85 (1988) 5350.
22. Klebe, G., Abraham, U. and Mietzner, T., J. Med. Chem., 37 (1994) 4130.
23. Cramer, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
24. Viswanadhan, V.N., Ghose, A.K., Revankar, G.R. and Robins, R.K., J. Chem. Inf. Comput. Sci., 29 (1989) 163.
25. Besler, B.H., Merz, K.M. and Kollman, P.A., J. Comput. Chem., 11 (1990) 431.
26. Wold, S., Pattern Recognit., 8 (1976) 127.
27. Dunn, W.J. and Wold, S., In van de Waterbeemd, H. (Ed.) Chemometric Methods in Molecular Design, VCH, New York, 1995, pp. 179–193.
28. Hunt, P.A., J. Comput.-Aided Mol. Des., 13 (1999) 453.
29. Bohm, M., Sturzebecher, J. and Klebe, G., J. Med. Chem., 42 (1999) 458.

Table A1. Structures (families) of DHFR inhibitors.

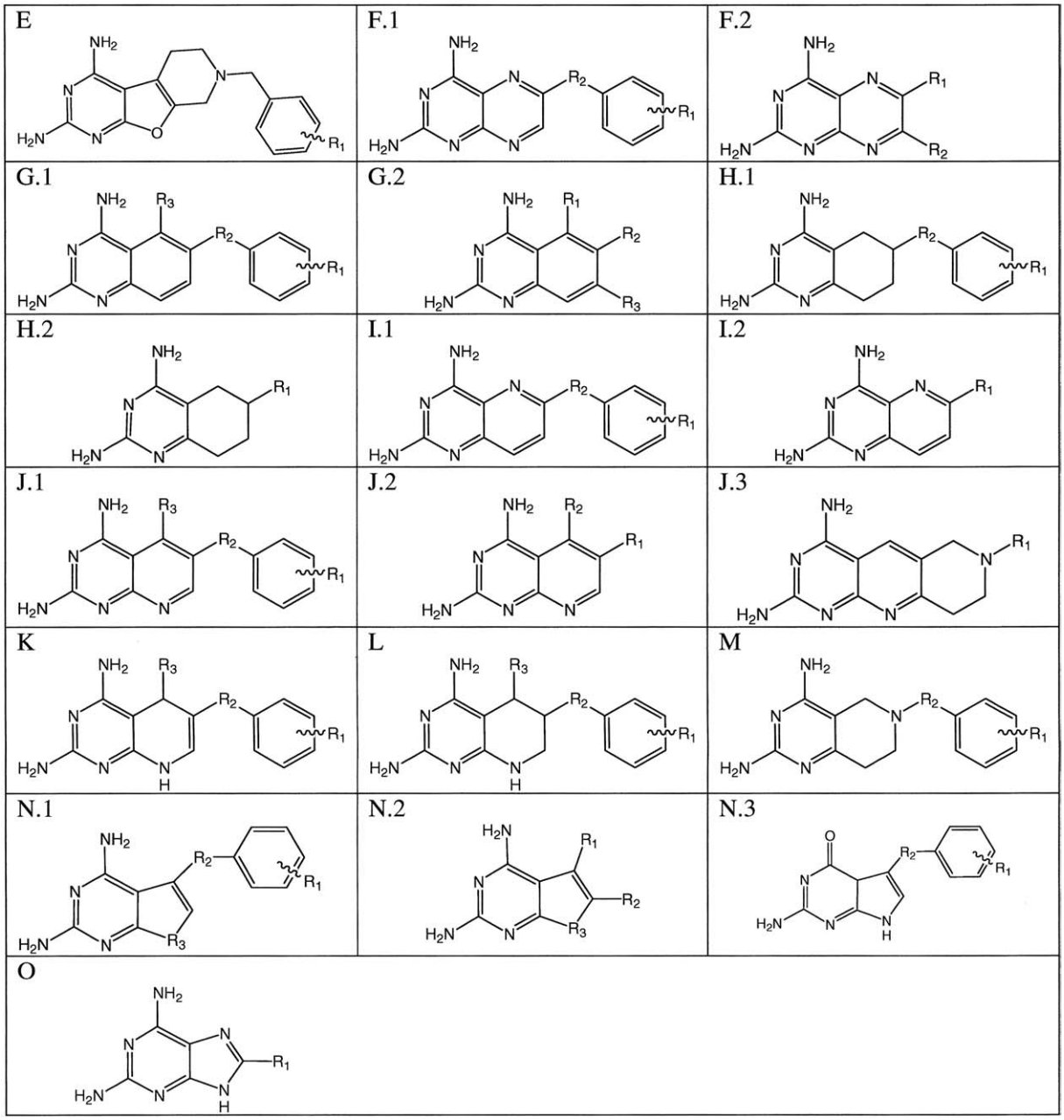


Table A2. DHFR inhibitors.

Name[a]	R_1	R_2	R_3	Family	IC_{50} (μM)		Set[b]
					PC	RL	
7-1	-H			E	>9	15	3
7-2	-3,4,5-OMe			E	>35	>35	3
7-3	-3,5-OMe			E	>22.8	22.8	3
7-4	-2,4-Cl			E	22.6	50.9	1
7-5	-3,4-Cl			E	24.1	20.6	1
7-6	-2,6-Cl			E	40.5	81.2	2
7-7	-4-CO-L-Glu			E	10.9	85.8	2
1-MTX	-4-CO-L-Glu	$-CH_2N(Me)-$		F.1	0.0013	0.0025	0
1-98535	-4-CO-L-Glu	$-CH_2NH-$		F.1	0.028	0.017	1
1-107146	-2-F,-4-D-Glu	$-CH_2N(Me)-$		F.1	0.00072	0.00036	1
1-117356	-4-D-Glu	$-CH_2N(Me)-$		F.1	0.00088	0.0062	1
1-127977	-4-CO-D-Lys	$-CH_2N(Me)-$		F.1	>8	>8	3
1-131463	$-4-CO_2H$	$-CH_2N(Me)-$		F.1	0.94	0.11	1
1-137545	$-2-Me,-4-CO_2H,-6-Me$	$-CH_2N(Me)-$		F.1	0.19	0.038	2
1-144698	$-2-CF_3$,-4-D-Glu	$-CH_2N(Me)-$		F.1	0.00031	0.00155	1
1-152737	-2-OMe,-4-D-Glu	$-CH_2N(Me)-$		F.1	0.00049	0.0011	2
1-169531	-4-CO-L-Glu	$-CH_2N(Et)-$		F.1	0.00035	0.0014	0
1-233903	$-4-CONH_2$	$-CH_2NH-$		F.1	>1000	220	1
1-233904	-H	$-CH_2N(Me)-$		F.1	10	0.83	1
1-233910	-4-NHCOMe	$-CH_2NH-$		F.1	7.1	0.17	2
1-235791	-4-COMe	$-CH_2NH-$		F.1	0.72	0.075	1
1-236642	$-4-CON(Me)_2$	$-CH_2NH-$		F.1	>5	>10	3
1-241522	-4-CONH-Pr	$-CH_2NH-$		F.1	3.1	0.48	2
9-12c	-2,3-(CH)4-	$-CH_2S-$		F.1	4.2	8.2	1
9-12d	-3-Me	$-CH_2S-$		F.1	21.2	8.48	1
9-12e	-4-Me	$-CH_2S-$		F.1	30	26	1
9-12g	-4-OMe	$-CH_2S-$		F.1	17.2	10.2	1
9-12h	-3,4-OMe	$-CH_2S-$		F.1	58.2	36.3	2
9-12i	-3-Cl	$-CH_2S-$		F.1	42.2	26.6	1
9-12j	-4-Cl	$-CH_2S-$		F.1	42.1	14.3	1
9-12k	-4-Cl	$-CH_2NH-$		F.1	1.5	0.3	1
9-12l	-2-Me,-5-OMe	$-CH_2NH-$		F.1	>3	>3	3
9-12m	-2-Me,-6-OMe	$-CH_2NH-$		F.1	15.8	5.7	1
9-12n	-2,5-OMe	$-CH_2NH-$		F.1	6.2	22.9	1
9-12o	-2,5-OMe	$-CH_2N(Me)-$		F.1	3.9	0.47	2
9-12p	-3,5-OMe	$-CH_2NH-$		F.1	0.96	0.88	1
9-12q	-3,4,5-OMe	$-CH_2NH-$		F.1	7	1.9	1
9-12r	$-2-OMe,-5-CF_3$	$-CH_2NH-$		F.1	21	21	1
9-12s	$-3-OMe,-5-CF_3$	$-CH_2NH-$		F.1	0.68	1.9	2
9-12t	$-2,3-(CH)_4-$	$-CH_2CH_2-$		F.1	97	60	2
9-12u	-2,5-OMe	$-CH_2CH_2-$		F.1	11.1	23.2	2
10-2d	-H	$-CH_2CO_2-$		F.1	9.8	7	2
9-12f	-3-OMe	$-CH_2S-$		F.1	9.9	7.7	1
1-232965	$-CH_2$-S-Ph	-H		F.2	9.5	246	1
1-235776	$-CH_2$-NH-Ph	-H		F.2	>1.9	>1.9	3
1-233912	$-CH_2$-O-Ph	-H		F.2	136	13	2
1-235777	$-CH_2$NH-1-naphthyl	-H		F.2	0.13	1.26	1
6-48	-1,2-naphthyl	$-R_1$		F.2	18.4	1.3	2

Table A2. Continued.

Name[a]	R_1	R_2	R_3	Family	IC_{50} (μM)		Set[b]
					PC	RL	
11-GR92754	-i-Butyl	$-R_1$		F.2	0.082	0.32	1
11-AH2503	-Et	$-R_1$		F.2	0.62	2.2	2
11-AH2504	$-C_6H_4$-2-OMe	-Et		F.2	2	0.41	2
12-4b	mdba[c]	-H		F.2	1.4	5.1	1
12-4d	mda-O[c]	-H		F.2	3.4	13	1
12-4g	$-CH_2N(Ph)_2$	-H		F.2	4.9	4	2
12-4c	mda-C[c]	-H		F.2	0.042	0.027	1
12-4e	mda-S[c]	-H		F.2	0.12	0.2	2
12-4a	mdba[c]	-H		F.2	0.21	4.4	1
10-18d	$-CH_2OH$	-H		F.2	>51	>51	3
1-122870	-4-CO-L-Asp	$-CH_2NH-$	-Me	G.1	0.00042	0.00016	1
1-TMQ	-3,4,5-OMe	$-CH_2NH-$	-Me	G.1	0.042	0.003	2
1-184692	-4-CO-L-Asp	$-CH_2NH-$	-Et	G.1	0.010	0.004	2
1-351521	-3,5-OMe	$-CH_2-$	-Me	G.1	0.031	0.002	2
13-10	-3,4,5-OMe	$-CH_2N(Me)-$	-Cl	G.1	0.012	0.012	1
13-11	-2,5-OMe	$-NHCH_2-$	-Cl	G.1	0.053	0.028	2
13-14	-3,4,5-OMe	$-NHCH_2-$	-Cl	G.1	0.033	0.006	1
13-17	-3,4,5-OMe	$-N(Me)CH_2-$	-Cl	G.1	0.17	0.038	1
14-28	-2,5-OMe	$-CH_2NH-$	-Cl	G.1	0.051	0.044	1
14-29	-3,4,5-OMe	$-CH_2NH-$	-Cl	G.1	0.033	0.006	1
9-16b	-4-Cl	$-CH_2NH-$	-H	G.1	0.6	0.073	1
9-16c	-2-Me,-4-Cl	$-CH_2NH-$	-H	G.1	0.33	0.023	2
15-3	-2,5-OMe	$-NHCH_2-$	-H	G.1	4.6	1.1	1
15-4	-3,5-OMe	$-NHCH_2-$	-H	G.1	2.2	0.84	1
15-5	-2,4-OMe	$-NHCH_2-$	-H	G.1	4.4	1.2	1
15-6	-3,4,5-OMe	$-NHCH_2-$	-H	G.1	6.8	0.9	1
15-7	-2,3,4-OMe	$-NHCH_2-$	-H	G.1	4.9	1.3	1
15-10	$-2,3-(CH)_4-$	$-NHCH_2-$	-H	G.1	0.72	0.19	1
10-1a	-H	$-CO_2CH_2-$	-H	G.1	2.5	0.42	2
10-3a	-3,4,5-OMe	$-CH_2CO_2-$	-H	G.1	0.56	0.026	2
10-4a	-2,5-OMe	$-CH_2CO_2-$	-H	G.1	2.4	0.42	1
10-5a	-2,3,4-OMe	$-CH_2CO_2-$	-H	G.1	1.4	0.5	1
10-6a	-3,5-OMe	$-CH_2CO_2-$	-H	G.1	0.46	0.35	1
16-9a	-2-Br,-3,4,5-OMe	$-CH_2NH-$	-Cl	G.1	0.028	0.010	2
16-11	-2-Br,-3,4,5-OMe	$-CH_2NH-$	-H	G.1	0.55	0.33	1
15-11	-2,5-OMe	$-N(Me)CH_2-$	-H	G.1	0.087	0.026	1
15-12	-3,5-OMe	$-N(Me)CH_2-$	-H	G.1	0.024	0.008	2
15-13	-2,4-OMe	$-N(Me)CH_2-$	-H	G.1	0.1	0.043	1
15-14	-2,3,4-OMe	$-N(Me)CH_2-$	-H	G.1	0.052	0.019	1
15-15	$-2,3-(CH)_4-$	$-N(Me)CH_2-$	-H	G.1	0.017	0.017	1
10-11a	-H	$-CH_2NH-$	-H	G.1	0.98	0.44	1
10-12a	-H	$-CH_2N(Me)-$	-H	G.1	0.096	0.069	2
6-1	-Me	-H	-H	G.2	1.5	7.4	1
6-2	$-CF_3$	-H	-H	G.2	82	215	2
6-3	-F	-H	-H	G.2	0.41	5	2
6-7	-OMe	-H	-H	G.2	4.5	1.2	1
6-13	-SMe	-H	-H	G.2	1.1	0.54	2
6-14	-H	-Me	-H	G.2	20	55	1

Table A2. Continued.

Name[a]	R_1	R_2	R_3	Family	IC_{50} (μM)		Set[b]
					PC	RL	
6-15	-H	$-CF_3$	-H	G.2	2.7	4.1	1
6-16	-H	-F	-H	G.2	119	452	1
6-20	-H	-OMe	-H	G.2	58	46	2
6-21	-H	-H	-Me	G.2	65	42	1
6-22	-H	-H	$-CF_3$	G.2	31	63	1
6-23	-H	-H	-F	G.2	41	189	1
6-26	-Cl	-Cl	-H	G.2	9.7	3.5	1
6-28	-H	-Me	-Me	G.2	5.3	7.9	2
18-5	-H	dca[c]	-H	G.2	13	1.5	2
18-6	-H	ddca[c]	-H	G.2	0.51	0.019	2
10-17a	-H	$-CO_2H$	-H	G.2	>32	>32	3
10-18a	-H	$-CH_2OH$	-H	G.2	51.7	40.3	1
19-1	-3,4,5-OMe	$-CH_2NH-$		H.1	4.6	0.29	1
19-2	-3,4,5-OMe	$-CH_2N(Me)-$		H.1	0.095	0.038	1
19-3	-3,4,5-OMe	$-CH_2N(CH_2C{\equiv}CH)-$		H.1	0.119	0.074	2
19-4	-2,5-OMe	$-CH_2NH-$		H.1	1.67	0.56	2
19-5	-2,5-OMe	$-CH_2N(Me)-$		H.1	0.3	0.26	1
19-6	-2,5-OMe	$-CH_2N(Et)-$		H.1	0.114	0.071	1
19-7	-2,5-OEt	$-CH_2NH-$		H.1	1.57	1.47	1
19-8	-2,5-OEt	$-CH_2N(Me)-$		H.1	0.319	0.116	1
19-9	-3,4-Cl	$-CH_2NH-$		H.1	6.8	0.15	1
19-10	-3,4-Cl	$-CH_2N(Me)-$		H.1	0.246	0.034	2
19-11	-2,5-Cl	$-CH_2NH-$		H.1	0.41	0.24	1
19-12	-2,6-Cl	$-CH_2NH-$		H.1	0.502	0.109	1
19-13	-4-Cl	$-CH_2NH-$		H.1	0.94	0.128	1
19-14	-4-Cl	$-CH_2N(Me)-$		H.1	0.171	0.067	1
19-15	-3-Br	$-CH_2NH-$		H.1	0.33	0.227	1
19-19	$-2,3-(CH)_4-$	$-CH_2NH-$		H.1	0.517	0.139	1
19-20	$-2,3-(CH)_4-$	$-CH_2N(Me)-$		H.1	0.1	0.047	1
20-5a	-H	$-CH_2-$		H.1	0.29	0.18	1
20-5b	-2-Me	$-CH_2-$		H.1	0.25	0.11	2
20-5c	-3-Me	$-CH_2-$		H.1	0.34	0.11	1
20-5d	-2-OMe	$-CH_2-$		H.1	0.45	0.12	1
20-5e	-3-OMe	$-CH_2-$		H.1	0.27	0.11	1
20-5f	-4-OMe	$-CH_2-$		H.1	0.44	0.077	1
20-5g	$-3-CF_3$	$-CH_2-$		H.1	0.4	0.19	1
20-5h	$-3-OCF_3$	$-CH_2-$		H.1	0.1	0.079	2
20-5i	$-4-OCF_3$	$-CH_2-$		H.1	0.58	0.17	1
20-5j	-2,5-OMe	$-CH_2-$		H.1	0.057	0.034	1
20-5k	-3,4-OMe	$-CH_2-$		H.1	0.1	0.063	1
20-5m	-3,4-Cl	$-CH_2-$		H.1	15	7.3	1
20-9a	-Et			H.2	6.9	3	1
20-9b	-t-Bu			H.2	0.18	0.065	1
20-9c	-Ph			H.2	2.2	1.9	2
14-5	-2,5-OMe	$-CH_2NH-$		I.1	1.4	0.43	1
14-8	-4-Cl	$-CH_2N(Me)-$		I.1	0.062	0.022	1

Table A2. Continued.

Name[a]	R_1	R_2	R_3	Family	IC_{50} (μM)		Set[b]
					PC	RL	
14-9	-3-Cl	-CH_2N(Me)-		I.1	2.1	0.067	1
14-10	-3,4-Cl	-CH_2N(Me)-		I.1	0.022	0.032	1
21-3	-H	-CH_2S-		I.1	2	0.52	2
21-4	-H	-CH_2NH-		I.1	1.7	0.26	1
21-5	-H	-CH_2N(Me)-		I.1	0.29	0.024	1
21-6	-2-OMe	-CH_2NH-		I.1	2.7	0.42	2
21-7	-2-OMe	-CH_2N(Me)-		I.1	0.51	0.12	1
21-8	-3-OMe	-CH_2NH-		I.1	1.7	0.2	1
21-9	-3-OMe	-CH_2N(Me)-		I.1	0.097	0.035	1
21-10	-4-OMe	-CH_2NH-		I.1	0.85	0.073	1
21-11	-4-OMe	-CH_2N(Me)-		I.1	0.25	0.018	1
21-12	-2-Cl	-CH_2NH-		I.1	0.53	0.14	1
21-13	-2-Cl	-CH_2N(Me)-		I.1	0.21	0.12	1
21-14	-3-Cl	-CH_2NH-		I.1	2	0.14	1
21-15	-2,4-OMe	-CH_2NH-		I.1	5.5	0.32	1
21-16	-2,4-OMe	-CH_2N(Me)-		I.1	0.16	0.016	1
21-18	-2,5-OMe	-CH_2N(Me)-		I.1	0.21	0.05	1
21-19	-3,4-OMe	-CH_2NH-		I.1	0.9	0.06	1
21-20	-3,4-OMe	-CH_2N(Me)-		I.1	0.091	0.003	2
21-21	-2,4-Cl	-CH_2NH-		I.1	0.73	0.088	1
21-22	-2,4-Cl	-CH_2N(Me)-		I.1	0.5	0.058	1
21-23	-2,5-Cl	-CH_2NH-		I.1	1.6	0.2	1
21-24	-2,5-Cl	-CH_2N(Me)-		I.1	0.15	0.047	2
21-27	-3,4-Cl	-CH_2NH-		I.1	0.41	0.054	1
21-34	-2,3-$(CH)_4$-	-CH_2S-		I.1	0.47	0.16	1
21-36	-2,3-$(CH)_4$-	-CH_2NH-		I.1	0.23	0.04	2
21-38	-2,3-$(CH)_4$-	-CH_2N(Me)-		I.1	0.04	0.007	1
21-40	-4-COMe	-CH_2NH-		I.1	0.41	0.003	1
21-41	-4-COMe	-CH_2N(Me)-		I.1	0.13	0.005	2
21-42	-4-COMe	-CH_2N(C≡CH)-		I.1	0.22	0.015	1
21-43	-4-$COCF_3$	-CH_2N(Me)-		I.1	0.25	0.032	1
21-44	-4-$COCF_3$	-CH_2N(C≡CH)-		I.1	0.12	0.008	1
22-4	-2-OMe	-S-		I.1	2.2	0.23	1
22-5	-4-OMe	-S-		I.1	0.7	0.075	2
22-6	-3,4-OMe	-S-		I.1	0.086	0.018	1
22-7	-2-OMe	-SO_2-		I.1	3.2	1.4	1
22-8	-4-OMe	-SO_2-		I.1	10.5	2	1
22-9	-3,4-OMe	-SO_2-		I.1	2.7	0.88	2
22-11	-2-OMe	-NH-		I.1	8.7	0.26	1
22-12	-4-OMe	-NH-		I.1	90.4	3.8	1
22-15	-3,4-OMe	-NH-		I.1	40.4	1.1	1
22-19	-4-OMe	-N(Me)-		I.1	0.22	0.007	2
22-21	-3,4-OMe	-N(Me)-		I.1	0.0023	0.0004	2
22-14	-2,5-OMe	-NH-		I.1	16.1	3.6	1
22-20	-2,5-OMe	-N(Me)-		I.1	0.034	0.004	2
22-16	-3,4,5-OMe	-NH-		I.1	25.9	3.2	1

Table A2. Continued.

Name[a]	R_1	R_2	R_3	Family	IC_{50} (μM)		Set[b]
					PC	RL	
22-22	-3,4,5-OMe	-N(Me)-		I.1	0.021	0.004	1
22-23	-2,3-$(CH)_4$-	-N(Me)-		I.1	5.1	3.3	1
22-17	-3,4-$(CH)_4$-	-NH-		I.1	15	2	2
22-10	-H	-NH-		I.1	8.3	0.43	1
22-18	-H	-N(Me)-		I.1	0.010	0.001	1
22-13	-4-Cl	-NH-		I.1	14.6	0.82	1
22-25	-3,4,5-OMe	-$NHCH_2$-		I.1	29.4	1.4	2
18-4	mdba[c]			I.2	0.037	0.053	2
8-3	-H	-CH_2S-	-H	J.1	1.3	1.9	1
8-6	-2,5-Cl	-CH_2S-	-H	J.1	5.9	2.5	1
8-7	-3,5-Cl	-CH_2S-	-H	J.1	11	38	2
8-8	-3,4-OMe	-CH_2S-	-H	J.1	2.2	4	2
23-6a	-3,4,5-OMe	-CH_2N (CHO)-	-Me	J.1	0.55	0.11	2
23-4a	-3,4,5-OMe	-CH_2NH-	-Me	J.1	0.086	0.0021	0
23-4b	-3,4-Cl	-CH_2NH-	-Me	J.1	0.32	0.053	1
23-6b	-3,4-Cl	-CH_2N (CHO)-	-Me	J.1	0.51	0.14	2
24-2a	-2,5-OMe	-CH_2NH-	-Me	J.1	0.046	0.128	1
24-2b	-3,5-OMe	-CH_2NH-	-Me	J.1	0.023	0.043	1
24-2c	-2,4-OMe	-CH_2NH-	-Me	J.1	0.316	0.214	1
24-2d	-3,4-OMe	-CH_2NH-	-Me	J.1	0.044	0.008	1
24-2e	-2,5-OEt	-CH_2NH-	-Me	J.1	0.077	0.017	1
24-3a	-2,5-OMe	-CH_2N(Me)-	-Me	J.1	0.216	0.407	1
24-3c	-2,4-OMe	-CH_2N(Me)-	-Me	J.1	0.32	0.044	2
24-3e	-2,5-OEt	-CH_2N(Me)-	-Me	J.1	3.1	3.0	2
24-4a	-3,4,5-OMe	-$CH_2N(CH_2C{\equiv}CH)$-	-Me	J.1	0.054	0.012	2
24-4c	-3,4,5-OMe	-CH_2N(Et)-	-Me	J.1	0.050	0.011	1
24-5a	-2,3-$(CH)_4$-	-CH_2NH-	-Me	J.1	0.573	0.030	1
24-5b	-2,3-$(CH)_4$-,-4-OMe	-CH_2NH-	-Me	J.1	0.041	0.054	1
14-22	-3,4-Cl	-CH_2N(Me)-	-Me	J.1	0.1	0.042	1
9-13b	-H	-CH_2S-	-Me	J.1	0.44	0.43	1
9-13d	-3-Me	-CH_2S-	-Me	J.1	0.17	0.33	1
9-13e	-4-Me	-CH_2S-	-Me	J.1	0.53	0.5	2
9-13f	-3-OMe	-CH_2S-	-Me	J.1	0.34	0.6	1
9-13g	-4-OMe	-CH_2S-	-Me	J.1	0.56	0.52	1
9-13h	-3,4-OMe	-CH_2S-	-Me	J.1	0.15	0.18	1
9-13i	-3-Cl	-CH_2S-	-Me	J.1	0.068	0.19	1
9-13j	-4-Cl	-CH_2S-	-Me	J.1	0.36	0.37	1
9-13l	-2-Me,5-OMe	-CH_2NH-	-Me	J.1	0.038	0.15	1
9-13m	-2-Me,-6-OMe	-CH_2NH-	-Me	J.1	1	0.32	1
9-13r	-2-OMe,-5-CF_3	-CH_2NH-	-Me	J.1	0.044	0.02	1
9-13s	-3-OMe,-5-CF_3	-CH_2NH-	-Me	J.1	0.02	0.017	2
9-13t	-2,3-$(CH)_4$-	-CH_2CH_2-	-Me	J.1	0.064	0.135	2
9-13u	-2,5-OMe	-CH_2CH_2-	-Me	J.1	0.34	0.77	1
9-14a	-2,5-Me	-CH_2NH-	-Me	J.1	0.03	0.12	1
9-14b	-2-Me,-4-OMe	-CH_2NH-	-Me	J.1	0.17	0.029	1
9-14c	-2-OMe,-5-Me	-CH_2NH-	-Me	J.1	0.068	0.16	1
9-14h	-3,4-OMe	-CH_2N(Me)-	-Me	J.1	0.32	0.004	1

Table A2. Continued.

Name[a]	R_1	R_2	R_3	Family	IC_{50} (μM)		Set[b]
					PC	RL	
9-14k	-2-OMe,-5-CF_3	-CH_2N(Me)-	-Me	J.1	0.093	0.23	1
9-14m	-2,3-$(CH)_4$-	-CH_2N(Me)-	-Me	J.1	0.15	0.14	1
9-15a	-2,3-$(CH)_4$-	-CH_2NH-	-H	J.1	0.26	0.23	2
9-15b	-4-Cl	-CH_2NH-	-H	J.1	0.97	0.72	1
9-15c	-3,4,5-OMe	-CH_2NH-	-H	J.1	2	0.81	1
9-16d	-2,5-OMe	-CH_2NH-	-H	J.1	0.75	0.46	1
25-10	-H	-N(Me)CH_2-	-H	J.1	0.068	0.14	2
25-11	-3,4,5-OMe	-$NHCH_2$-	-H	J.1	14.1	3.3	1
25-12	-3,4,5-OMe	-N(Me)CH_2-	-H	J.1	0.061	0.033	1
25-13	-3,4,5-OMe	-NHCH(Me)-	-H	J.1	9.2	1.27	1
25-14	-2,3,4-OMe	-$NHCH_2$-	-H	J.1	15.3	3.24	1
25-15	-2,3,4-OMe	-N(Me)CH_2-	-H	J.1	0.079	0.03	2
25-16	-2,4,6-OMe	-$NHCH_2$-	-H	J.1	20.7	1.2	1
25-19	-3,5-OMe	-$NHCH_2$-	-H	J.1	5.7	3.4	1
25-20	-3,5-OMe	-N(Me)CH_2-	-H	J.1	0.076	0.072	1
25-21	-2,5-OMe	-$NHCH_2$-	-H	J.1	3.8	0.35	1
25-22	-2,5-OMe	-N(Me)CH_2-	-H	J.1	0.084	0.057	1
25-23	-2,3-$(CH)_4$-	-$NHCH_2$-	-H	J.1	3.9	0.24	1
25-24	-2,3-$(CH)_4$-,-4-OMe	-$NHCH_2$-	-H	J.1	8.2	0.43	1
25-25	-2,3-$(CH)_4$-,-6-OMe	-$NHCH_2$-	-H	J.1	15.4	0.37	1
25-26	-4-O-Ph	-$NHCH_2$-	-H	J.1	24.3	2.9	2
26-7	-2,5-OMe,-4-pyrrolo	-CH_2NH-	-Me	J.1	0.35	0.23	2
26-8	-2-pyrrolo,-4,5-OMe	-CH_2NH-	-Me	J.1	1.8	3.5	1
26-9	-2,3,5,6-OMe,-4-pyrrolo	-CH_2NH-	-Me	J.1	0.62	0.17	2
26-10	-2-OMe,-5-Ph	-CH_2NH-	-Me	J.1	0.64	0.44	1
27-6	-2-OMe	-CH_2NH-	-Me	J.1	0.117	0.169	1
27-7	-3-OMe	-CH_2NH-	-Me	J.1	0.069	0.080	1
27-9	-2-Cl	-CH_2NH-	-Me	J.1	0.047	0.088	1
27-10	-3-Cl	-CH_2NH-	-Me	J.1	0.023	0.037	1
27-11	-4-Cl	-CH_2NH-	-Me	J.1	0.055	0.051	1
27-13	-3-OMe	-CH_2(Me)-	-Me	J.1	0.03	0.018	1
27-14	-4-OMe	-CH_2(Me)-	-Me	J.1	0.035	0.013	1
27-15	-2-Cl	-CH_2(Me)-	-Me	J.1	0.084	0.1	2
27-16	-4-Cl	-CH_2(Me)-	-Me	J.1	0.029	0.026	1
10-2b	-H	-CH_2CO_2-	-H	J.1	5.45	3.9	1
10-2c	-H	-CH_2CO_2-	-Me	J.1	0.25	0.23	2
28-3	-3,4,5-OMe	-CH_2N(Me)-	-H	J.1	0.24	0.28	1
28-4	-3,4,5-OMe	-CH_2N(Et)-	-H	J.1	0.19	0.12	1
28-5	-3,4,5-OMe	-CH_2N (CHO)-	-H	J.1	18.5	7.4	1
28-6	-3,4,5-OMe	-CH=CH-	-H	J.1	>5.0	12.9	3
28-7	-3,4-OMe	-CH=CH-	-H	J.1	2.6	2.1	2
28-8	-4-OMe	-CH=CH-	-H	J.1	5.3	11.8	1
28-9	-3,4,5-OMe	-CH_2CH_2-	-H	J.1	5	1.14	1
28-10	-3,4-OMe	-CH_2CH_2-	-H	J.1	1.4	0.61	1
28-11	-4-OMe	-CH_2CH_2-	-H	J.1	0.29	0.26	2
29-2-PTX	-2,5-OMe	-CH_2-	-Me	J.1	0.031	0.002	2
26-5	-CH_2-1-(indole-5-OMe)	-Me		J.2	0.57	0.47	2

Table A2. Continued.

Name[a]	R_1	R_2	R_3	Family	IC_{50} (μM)		Set[b]
					PC	RL	
30-3	-CH_2-1-(2,3-dihydro-indole-5-OMe)	-Me		J.2	0.25	0.17	1
30-2	-CH_2-1-(2,3-dihydro-indole-4-OMe)	-Me		J.2	0.29	0.15	2
30-4	-CH_2-1-(2,3-dihydro-indole-5,6-diOMe)	-Me		J.2	0.41	0.23	1
18-3	mdba[c]	-H		J.2	0.043	0.19	2
10-18b	-CH_2OH	-H		J.2	>104	>104	3
10-18c	-CH_2OH	-Me		J.2	10.5	8.5	2
31-4	-CH_2Ph			J.3	16		1
31-3	-$CH_2C_6H_4$-L-Glu			J.3	1.8		1
31-5	-$CH_2C_6H_4$-4-OMe			J.3	>15		3
31-6	-$CH_2C_6H_3$-3,5-OMe			J.3	14		2
31-7	-$CH_2C_6H_3$-2,4-Cl			J.3	4		2
31-8	-$CH_2C_6H_3$-3,4-Cl			J.3	5		1
23-7	-3,4,5-OMe	-CH_2CHO-	-Me	K	>2.6	>2.6	3
23-8b	-3,4,5-OMe	-CH_2CHO-	-Me	L	>1.2	>1.2	3
23-8	-3,4,5-OMe	-CH_2NH-	-Me	L	>3700	3700	3
28-12	-3,4,5-OMe	-CH_2CH_2-	-H	L	61.7	6.1	2
32-8	-2,5-OMe	-CH_2-		M	3.3	1.4	2
32-9	-3,4,5-OMe	-CH_2-		M	6.9	2.2	1
32-10	-2-Br,-3,4,5-OMe	-CH_2-		M	0.51	0.35	2
32-11	-3,4,5-OMe	-CH_2CH_2-		M	30	9.5	2
33-1	-4-L-Glu	-CH_2NH-	O	N.1	0.9	1.3	2
33-2	-4-L-Glu	-CH_2N(Me)-	O	N.1	0.035	0.43	1
33-3	-3,4,5-OMe	-CH_2NH-	O	N.1	>4	>37.0	3
33-5	-3,4-Cl	-CH_2NH-	O	N.1	>35	35.2	3
33-6	-2,5-OMe	-CH_2NH-	O	N.1	>21	>21	3
34-2	-3,4-OMe	-CH_2NH-	NH	N.1	119	116	1
34-3	-4-OMe	-CH_2NH-	NH	N.1	279	63	2
34-4	-2,5-OMe	-CH_2NH-	NH	N.1	45.7	156	1
34-5	-2,5-OEt	-CH_2NH-	NH	N.1	>21	70	3
34-6	-3,4-Cl	-CH_2NH-	NH	N.1	35.3	14.4	1
34-7	-2,3-$(CH)_4$-	-CH_2NH-	NH	N.1	307	59.3	1
34-9	-4-L-Glu	-CH_2NH-	NH	N.1	0.038	0.044	1
35-2	-2,5-OMe	-CH_2N(Me)-	NH	N.1	>12	>12	3
35-3	-3,4-Cl	-CH_2N(Me)-	NH	N.1	28.3	3	2
35-4	-2,3-$(CH)_4$-	-CH_2N(Me)-	NH	N.1	209	8.2	1
35-5	-3,4-OMe	-CH_2S-	NH	N.1	11.1	16.7	1
35-6	-3,4-Cl	-CH_2S-	NH	N.1	58.5	5.3	1
35-7	-2,3-$(CH)_4$-	-CH_2S-	NH	N.1	10.6	3	2
35-10	-4-L-Glu	-CH_2N(Me)-	NH	N.1	0.044	0.06	2
36-7a	-3,4,5-OMe	-CH_2NH-	S	N.1	>10	>10	3
36-7b	-2,5-OMe	-CH_2NH-	S	N.1	>10	>10	3
36-7c	-3,4,5-OMe	-CH_2N(Me)-	S	N.1	>10	>10	3
36-7d	-2,5-OMe	-CH_2N(Me)-	S	N.1	>10	>10	3
36-7e	-3,5-Cl,-4-(1-pyrrolo)	-CH_2NH-	S	N.1	>10	>10	3
37-1e	-H	-CH_2-	NH	N.1	>189	270	3
38-3	-H	-CH_2S-	O	N.1	>26	252	3

Table A2. Continued.

Name[a]	R_1	R_2	R_3	Family	IC_{50} (μM)		Set[b]
					PC	RL	
38-6	-2,3-$(CH)_4$-	-CH_2NH-	O	N.1	13.5	12	1
38-8	-3,4-$(CH)_4$-	-CH_2O-	O	N.1	14	60.3	2
38-9	-2-O-Ph	-CH_2NH-	O	N.1	>12	>12	3
38-10	-4-O-Ph	-CH_2NH-	O	N.1	8.1	16.2	1
38-11	-2-Ph	-CH_2NH-	O	N.1	7.7	137	1
38-12	-3,4-$(CH)_4$-	-CH_2N(Me)-	O	N.1	14.8	14.6	1
38-13	-2,5-Cl	-CH_2NH-	O	N.1	50.9	71.9	1
38-14	-3,4-Cl	-CH_2(Me)-	O	N.1	44.8	>27	2
38-16	-3-OMe	-CH_2NH-	O	N.1	>31.3	>31.3	3
38-17	-2,5-OMe	-CH_2(Me)-	O	N.1	>27	>27	3
39-3a	-4-L-Glu	-CH_2CH_2-	NH	N.1	1.2		2
40-3	-Me	-C_6H_3-2,5-OMe	S	N.2	4.8	0.37	2
40-4	-Me	-$CH_2C_6H_3$-2,5-OMe	S	N.2	14	0.4	1
40-5	-Me	-$(CH_2)_2$C6H3 -2,5-OMe	S	N.2	1.2	5.9	2
40-6	-Me	-C_6H_2-3,4,5-OMe	S	N.2	>8	1.8	3
40-7	-Me	-$CH_2C_6H_2$-3,4,5-OMe	S	N.2	>8	51	3
40-8	-Me	-$(CH_2)_2C_6H_2$-3,4,5-OMe	S	N.2	>8	25	3
40-9	-$CH_2CH_2C_6H_3$-2,5-OMe	-H	S	N.2	28	3.1	1
40-T2-1	-Me	-Ph	S	N.2	26	57	1
40-T2-2	-Me	-C_6H_3-3,4-Cl	S	N.2	16	4.6	2
40-T2-3	-Me	-CH_2Ph	S	N.2	35	14	1
40-T2-4	-C_6H_4-4-Cl	-Me	S	N.2	>100	>100	3
40-T2-5	-CH_2Ph	-Me	S	N.2	>100	>37	3
40-T2-6	-$(CH_2)_4$-	R1	S	N.2	2.1	3.9	2
40-T2-10	-$CH_2CH_2N(CH_2$-Ph)CH_2-	R1	S	N.2	5.2	0.57	2
16-8a	-Me	-$CH_2C_6H_1$-2-Br,-3,4,5OMe	S	N.2	>12	0.93	3
16-12	-$CH_2CH_2C_6H_1$-2-Br-3,4,5-OMe	-H	S	N.2	49	2.8	1
16-10	-$CH_2NHC_6H_1$-2-Br-3,4,5-OMe	-H	S	N.2	200	43	1
36-6a	-$CH_2NHC_6H_2$-3,4,5-OMe	-Br	S	N.2	13	17	1
36-6b	-$CH_2NHC_6H_3$-2,5-OMe	-Br	S	N.2	>100	33	3
36-6c	$CH_2N(Me)C_6H_2$-3,4,5-OMe	-Br	S	N.2	31	28	2
36-6d	$CH_2N(Me)C_6H_3$-2,5-OMe	-Br	S	N.2	>10	>10	3
36-6e	-$CH_2NC_6H_3$-3,5-Cl,-4-(1-pyrrolo)	-Br	S	N.2	7.5	10	2
37-1a	-Ph	-H	NH	N.2	>186	9.1	3
37-1b	-C_6H_4-4-Cl	-H	NH	N.2	>161	62	3
37-1c	-C_6H_3-3,4-Cl	-H	NH	N.2	33	23	2
37-1d	-C_6H_2-3,4,5-OMe	-H	NH	N.2	8.3	27	2
41-4c	-CH_2NH-1-fluorene	-H	O	N.2	>29	>29	3
41-4d	-CH_2NH-2-fluorene	-H	O	N.2	36.2	>10	1
41-4e	-CH_2NH-3-(2-methoxy-dibenzofuran)	-H	O	N.2	10.3	>32	2
41-4f	-CH_2NH-3-(N-ethyl-carbazole)	-H	O	N.2	16.2	12.6	2
41-4g	-CH_2NH-2-(9-hydroxy-fluorene)	-H	O	N.2	>13	>13	3
41-4h	-CH_2NH-2-(9-oxofluorene)	-H	O	N.2	>63	>63	3
41-4i	-CH_2NH-4-(9-oxofluorene)	-H	O	N.2	19.3	>38	2
41-5a	-CH_2N(Me)-3-(2-methoxy-dibenzofuran)	-H	O	N.2	>54	241	3
41-5b	-CH_2N(Me)-3-(N-ethyl-carbazole)	-H	O	N.2	18.7	24.7	1
41-7a	-CH_2S-2-biphenyl	-H	O	N.2	46	49	2

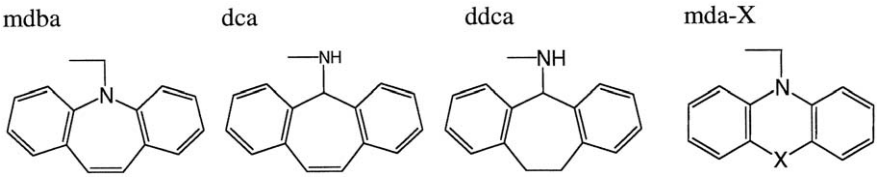
Table A2. Continued.

Name[a]	R_1	R_2	R_3	Family	IC_{50} (μM)		Set[b]
					PC	RL	
41-7b	-CH_2S-3-biphenyl	-H	O	N.2	22	31	1
41-7c	-CH_2S-4-biphenyl	-H	O	N.2	41	351	1
42-8	-4-Cl	-S-		N.3	40	24.6	2
42-9	-3,4-Cl	-S-		N.3	>20	>20	3
42-10	-4-NO_2	-S-		N.3	>15	7470	3
42-11	-3,4-OMe	-S-		N.3	>21	>21	3
43-7	-2,5-OMe	-CH_2NH-		N.3	47	47	1
43-8	-3,5-OMe	-CH_2NH-		N.3	20	20	1
43-9	-2,4-OMe	-CH_2NH-		N.3	16	16	1
43-10	-3,4,5-OMe	-CH_2NH-		N.3	22	22	1
43-11	-2,5-Cl	-CH_2NH-		N.3	25	25	1
43-13	-2,4-Cl	-CH_2NH-		N.3	15	15	2
43-14	-3-Cl	-CH_2NH-		N.3	26	26	1
43-15	-2,5-OMe	-CH_2N(Me)-		N.3	56	56	2
43-16	-3,5-OMe	-CH_2N(Me)-		N.3	87	87	1
43-17	-3,4,5-OMe	-CH_2N(Me)-		N.3	45	45	1
44-3	-C_6H_2-3,-4,5-OMe			O	59.4	13	1
44-2	-Ph			O	153	59.9	1
44-6	-C_6H_2-2,-4,5-OMe			O	22	7	1
44-7	-C_6H_3-2,-5-OMe			O	37.4	3.5	2
44-8	-C_6H_3-3,-5-OMe			O	9	32.4	1
44-9	-C_6H_3-3,-4-OMe			O	12	52.3	1
44-10	-C_6H_3-2,-4-OMe			O	14	0.9	1
44-11	-C_6H_3-3,-4-Cl			O	19.5	252	2
44-12	-C_6H_3-2,-6-Cl			O	11	11	1
44-14	-C_6H_2-2-NO_2-4,5-OMe			O	25	25	2
44-16	-4-pyridine			O	18	18	2
44-17	-2-naphthyl			O	113	280	1
44-18	-1-naphthyl-4-OMe			O	13	13	2
44-19	-9-fluorenyl			O	35	29.8	2
44-20	-CH_2-Ph			O	9.8	1.6	2
44-21	-CH_2NH-C_6H_2-3,4,5-OMe			O	34	34	2
44-22	-CH_2NH-C_6H_3-2,5-OMe			O	29	29	1
44-23	-CH_2S-2-naphthyl			O	105	107	2

[a]Names that begin with a number are formed by hyphenating the reference number and the label given to the compound in the reference.

[b]pc training set, test set and inactive compounds are indicated as 1, 2, and 3, respectively. Compounds identified with a 0 are included only for reference purpose; they were excluded during coverage-based set design. The rl training set is identical to that for pc; the rl test and inactive sets are identical to the pc sets, except that pc inactives with quantitative rl activities are rl test set compounds, and rl inactives with quantitative pc activities are pc test set compounds.

[c]Abbreviations for heterocyclic groups: mdba, dca, ddca, mda-X.



References

1. Broughton, M.C. and Queener, S.F., Antimicrob. Agents Chemother., 35 (1991) 1348.
6. Rosowsky, A., Hynes, J.B. and Queener, S.F., Antimicrob. Agents Chemother., 39 (1995) 79.
7. Gangjee, A., Elzein, E., Queener, S.F. and McGuire, J.J., J. Med. Chem., 41 (1998) 1409.
8. Gangjee, A., Adair, O. and Queener, S.F., Bioorg. Med. Chem., 9 (2001) 2929.
9. Piper, J.R., Johnson, C.A., Krauth, C.A., Carter, R.L., Hosmer, C.A., Queener, S.F., Borotz, S.E. and Pfefferkorn, E.R., J. Med. Chem., 39 (1996) 1271.
10. Graffner-Nordberg, M., Kolmodin, K., Aqvist, J., Queener, S.F. and Hallberg, A., J. Med. Chem., 44 (2001) 2391.
11. Jackson, H.C., Biggadike, K., McKilligin, E., Kinsman, O.S., Queener, S.F., Lane, A. and Smith, J.E., Antimicrob. Agents Chemother., 40 (1996) 1371.
12. Rosowsky, A., Cody, V., Galitsky, N., Fu, H.N., Papoulis, A.T. and Queener, S.F., J. Med. Chem., 42 (1999) 4853.
13. Rosowsky, A., Mota, C.E., Wright, J.E. and Queener, S.F., J. Med. Chem., 37 (1994) 4522.
14. Rosowsky, A., Forsch, R.A. and Queener, S.F., J. Med. Chem., 38 (1995) 2615.
15. Gangjee, A., Vidwans, A.P., Vasudevan, A., Queener, S.F., Kisliuk, R.L., Cody, V., Li, R.M., Galitsky, N., Luft, J.R. and Pangborn, S., J. Med. Chem., 41 (1998) 3426.
16. Rosowsky, A., Mota, C.E. and Queener, S.F., J. Heterocycl. Chem., 33 (1996) 1959.
18. Rosowsky, A., Fu, H.N. and Queener, S.F., J. Heterocycl. Chem., 37 (2000) 921.
19. Gangjee, A., Zaveri, N., Kothare, M. and Queener, S.F., J. Med. Chem., 38 (1995) 3660.
20. Rosowsky, A., Papoulis, A.T., Forsch, R.A. and Queener, S.F., J. Med. Chem., 42 (1999) 1007.
21. Gangjee, A., Zhu, Y.M., Queener, S.F., Francom, P. and Broom, A.D., J. Med. Chem., 39 (1996) 1836.
22. Gangjee, A., Zhu, Y.M. and Queener, S.F., J. Med. Chem., 41 (1998) 4533.
23. Gangjee, A., Shi, J.F., Queener, S.F., Barrows, L.R. and Kisliuk, R.L., J. Med. Chem., 36 (1993) 3437.
24. Gangjee, A., Vasudevan, A., Queener, S.F. and Kisliuk, R.L., J. Med. Chem., 38 (1995) 1778.
25. Gangjee, A., Vasudevan, A., Queener, S.F. and Kisliuk, R.L., J. Med. Chem., 39 (1996) 1438.
26. Gangjee, A., Vasudevan, A. and Queener, S.F., J. Med. Chem., 40 (1997) 479.
27. Gangjee, A., Adair, O. and Queener, S.F., J. Med. Chem., 42 (1999) 2447.
28. Gangjee, A., Devraj, R. and Queener, S.F., J. Med. Chem., 40 (1997) 470.
29. Rosowsky, A., Forsch, R.A. and Queener, S.F., J. Med. Chem., 45 (2002) 233.
30. Gangjee, A., Vasudevan, A. and Queener, S.F., In Ayling, J.E. et al. (Eds) Chemistry and Biology of Pteridines and Folates, Plenum Press, New York, 1993, pp. 449–453.
31. Gangjee, A., Zeng, Y.B., McGuire, J.J. and Kisliuk, R.L., J. Med. Chem., 45 (2002) 5173.
32. Rosowsky, A., Mota, C.E. and Queener, S.F., J. Heterocycl. Chem., 32 (1995) 335.
33. Gangjee, A., Devraj, R., McGuire, J.J., Kisliuk, R.L., Queener, S.F. and Barrows, L.R., J. Med. Chem., 37 (1994) 1169.
34. Gangjee, A., Mavandadi, F., Queener, S.F. and McGuire, J.J., J. Med. Chem., 38 (1995) 2158.
35. Gangjee, A., Mavandadi, F. and Queener, S.F., J. Med. Chem., 40 (1997) 1173.
36. Rosowsky, A., Papoulis, A.T. and Queener, S.F., J. Med. Chem., 40 (1997) 3694.
37. Rosowsky, A., Fu, H.N. and Queener, S.F., J. Heterocycl. Chem., 38 (2001) 1197.
38. Gangjee, A., Guo, X., Queener, S.F., Cody, V., Galitsky, N., Luft, J.R. and Pangborn, W., J. Med. Chem., 41 (1998) 1263.
39. Gangjee, A., Yu, J.M., McGuire, J.J., Cody, V., Galitsky, N., Kisliuk, R.L. and Queener, S.F., J. Med. Chem., 43 (2000) 3837.
40. Rosowsky, A., Mota, C.E., Wright, J.E., Freisheim, J.H., Heusner, J.J., McCormack, J.J. and Queener, S.F., J. Med. Chem., 36 (1993) 3103.
41. Gangjee, A., Dubash, N.P. and Queener, S.F., J. Heterocycl. Chem., 37 (2000) 935.
42. Gangjee, A., Mavandadi, F., Kisliuk, R.L., McGuire, J.J. and Queener, S.F., J. Med. Chem., 39 (1996) 4563.
43. Gangjee, A., Vidwans, A., Elzein, E., McGuire, J.J., Queener, S.F. and Kisliuk, R.L., J. Med. Chem., 44 (2001) 1993.
44. Gangjee, A., Vasudevan, A. and Queener, S.F., J. Med. Chem., 40 (1997) 3032.