



CoMFA analysis of tgDHFR and rLDHFR based on antifolates with 6–5 fused ring system using the all-orientation search (AOS) routine and a modified cross-validated r^2 -guided region selection (q^2 -GRS) routine and its initial application [☆]

Aleem Gangjee ^{a,*}, Xin Lin ^a, Lisa R. Biondo ^a, Sherry F. Queener ^b

^a Division of Medicinal Chemistry, Graduate School of Pharmaceutical Sciences, Duquesne University, Pittsburgh, PA 15282, United States

^b Department of Pharmacology and Toxicology, School of Medicine, Indiana University, Indianapolis, IN 46202, United States

ARTICLE INFO

Article history:

Received 27 August 2009

Revised 18 December 2009

Accepted 28 December 2009

Available online 6 January 2010

Keywords:

Antifolates

Dihydrofolate reductase

CoMFA analysis

Selective inhibitors

ABSTRACT

We report the development of CoMFA analysis models that correlate the 3D chemical structures of 80 compounds with 6–5 fused ring system synthesized in our laboratory and their inhibitory potencies against tgDHFR and rLDHFR. In addition to conventional CoMFA analysis, we used two routines available in the literature aimed at the optimization of CoMFA: all-orientation search (AOS) and cross-validated r^2 -guided region selection (q^2 -GRS) to further optimize the models. During this process, we identified a problem associated with q^2 -GRS routine and modified using two strategies. Thus, for the inhibitory activity against each enzyme (tgDHFR and rLDHFR), five CoMFA models were developed using the conventional CoMFA, AOS optimized CoMFA, the original q^2 -GRS optimized CoMFA and the modified q^2 -GRS optimized CoMFA using the first and the second strategy. In this study, we demonstrate that the modified q^2 -GRS routines are superior to the original routine. On the basis of the steric contour maps of the models, we designed four new compounds in the 2,4-diamino-5-methyl-6-phenylsulfanyl-substituted pyrrolo[2,3-*d*]pyrimidine series. As predicted, the new compounds were potent and selective inhibitors of tgDHFR. One of them, 2,4-diamino-5-methyl-6-(2',6'-dimethylphenylthio)pyrrolo[2,3-*d*]pyrimidine, is the first 6–5 fused ring system compound with nanomolar tgDHFR inhibitory activity. The HCl salt of this compound was also prepared to increase solubility. Both forms of the drug were tested in vivo in a *Toxoplasma gondii* infection mouse model. The results indicate that both forms were active with the HCl salt significantly more potent than the free base.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Infections caused by opportunistic pathogens *Pneumocystis carinii* (pc) and *Toxoplasma gondii* (tg) are the leading cause of morbidity and mortality in immunocompromised patients such as those with AIDS.² Dihydrofolate reductase (DHFR) inhibitors are the current drugs of choice for the treatment of these infections. Ideally, these drugs should efficiently inhibit the growth of pathogenic cells via DHFR inhibition without affecting the mammalian DHFR. Unfortunately, due to their lack of potency and/or selectivity, combinations of current DHFR inhibitors with other agents such as sulfa drugs are often required for synergistic effects or to decrease host toxicity, which leads to high costs. Discontinuation of therapy is necessary in many cases as a result of severe side effects. Therefore, efforts continue to be directed toward the development of sin-

gle agents which not only display high potency but are also selective against DHFR from *P. carinii* and/or *T. gondii* over mammalian DHFR, such as rat liver (rl) DHFR. In recent reports, DHFR inhibitors with a 6–5 fused ring system including furo[2,3-*d*]pyrimidine, pyrro[2,3-*d*]pyrimidine and purine derivatives have shown good activity and selectivity against these pathogenic organisms, especially *T. gondii*.^{3–16} Computational techniques such as QSAR models can be used to assist the rational design of more potent and selective DHFR inhibitors with 6–5 fused ring system.

Most models for predicting DHFR inhibition in the literature^{17–47} published to-date use homologous data sets of DHFR inhibitors with a specific heterocyclic core (e.g., quinazolines, pyrimidines). Mattioli and Jurs⁴⁸ recently developed QSAR models that correlated chemical structure and inhibition potency for three types of DHFR: rLDHFR, pcDHFR, and tgDHFR. The results, however, did not give structural information about the binding sites.

More recently, Sutherland and Weaver⁴⁹ generated comparative molecular similarity indices analysis (CoMSIA) three-dimensional quantitative structure–activity relationship (3D-QSAR)

[☆] See Ref. 1.

* Corresponding author. Tel.: +1 412 396 6070; fax: +1 412 396 5593.

E-mail address: gangjee@duq.edu (A. Gangjee).

models for the inhibitory activities against pcDHFR and rIDHFR using a data set of 406 structurally diverse DHFR inhibitors. Gangjee and Lin⁵⁰ also reported comparative molecular field analysis (CoMFA) and CoMSIA analysis of pcDHFR, tgDHFR and rIDHFR based on the biological data of 179 compounds synthesized in our laboratory.⁵⁰ These are general models for pcDHFR, tgDHFR and rIDHFR inhibitors. Even though some compounds with 6–5 fused ring system were included in the training set and/or test set, the majority of the drugs used in the development of these models were compounds with 6–6 fused ring system, such as quinazolines and pyridopyrimidines. In addition, certain monocyclic, tricyclic and tetracyclic compounds were also included in the data sets.

As a continuation of our previous work and the importance of both potency and selectivity associated with the 6–5 fused ring systems, we report the development of CoMFA analysis models that correlate the 3D chemical structures of 80 compounds^{3,4,6–9,12,14,15} with 6–5 fused ring system synthesized in our laboratory and their inhibitory potencies for tgDHFR and rIDHFR. In addition to conventional CoMFA analysis, we used two routines available in the literature aimed at the optimization of CoMFA: all-orientation search (AOS)⁵¹ and cross-validated r^2 -guided region selection (q^2 -GRS)⁵² to further optimize the models. During this process, we identified a problem associated with q^2 -GRS routine and modified it with two different strategies. Thus, for the inhibitory activity against each enzyme (tgDHFR and rIDHFR), five CoMFA models were developed using the conventional CoMFA, AOS optimized CoMFA, the original q^2 -GRS optimized CoMFA and the modified q^2 -GRS optimized CoMFA with strategy one and two.

2. Computational details

2.1. Data set and biology activity

The structures of the compounds used to develop the CoMFA model along with their DHFR inhibitory activities are listed in Table 1.

2.2. Structure alignment

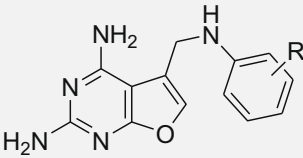
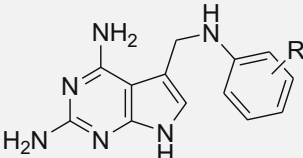
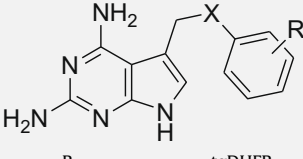
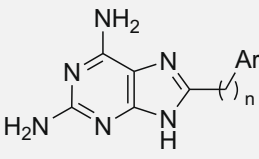
The pcDHFR bound structures of **81** (PDB 1daj)⁵³ (Fig. 1) was used as the template. For each class of compounds only the structure with the plain phenyl side chain was flexibly fit to the template in MMFF94 force field using the Flexible Align module implemented in MOE 2004.03⁵⁴ with the default options followed by the addition of the substituents on the phenyl in a consistent fashion, with preference for the solvent-exposed edge of the phenyl ring: the 2'-position was occupied before the 6'-position and the 3'-position before the 5'-position, with the ortho position taking precedence over the meta position when both are substituted. The newly added substituents were then relaxed in MOE with the rest of the molecule being fixed using MMFF94 force field with default option. The resulting structures were exported as MOL2 files and were subsequently imported into SYBYL 7.0,⁵⁵ followed by the calculation of Gasteiger–Hückel charges. The hydrogen atoms of each molecule were further optimized, while the heavy atoms were held still, with a gradient convergence of 0.05 kcal/Å and up to 1000 iterations using the Powell method in the TRIPOS force field. The final compound aggregate is shown in Figure 2.

2.3. Training set and test set

Among the 80 compounds listed in Table 1, and 76 compounds were active (had detectable IC_{50} values) against tgDHFR and 71 compounds were active against rIDHFR. A dataset containing only the


Table 1

Structures and IC_{50} values (μ M) of the compounds used in developing the models

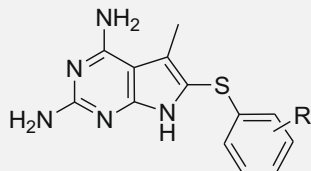
|  | | | | | |
|---------------------------------------------------------------------------------------|----------------------------------------------------|---------------------------------------------------------------------------------------|--------|----------|----------|
| Compd | R | tgDHFR | rIDHFR | Ref. | |
| 1 | 3,4,5-Cl ₃ | >3.9 | 25.6 | 3 | |
| 2 | 3,4-Cl ₂ | 89.3 | 35.2 | 3 | |
|  | | | | | |
| Compd | R | tgDHFR | rIDHFR | Ref. | |
| 3 | 3,4,5-(OCH ₃) ₃ | 8.1 | 56.3 | 4 | |
| 4 | 3,4-(OCH ₃) ₂ | 4.3 | 116.0 | 4 | |
| 5 | 4-OCH ₃ | 6.0 | 63.0 | 4 | |
| 6 | 2,5-(OCH ₃) ₂ | 1.7 | 156.0 | 4 | |
| 7 | 2,5-(OC ₂ H ₅) ₂ | 5.3 | 70.0 | 4 | |
| 8 | 3,4-Cl ₂ | 1.4 | 14.4 | 4 | |
| 9 | 2,3-(CH ₃) ₄ | 1.1 | 59.3 | 4 | |
| 10 | H | 3.9 | >252 | 4 | |
|  | | | | | |
| Compd | X | R | tgDHFR | rIDHFR | Ref. |
| 11 | N(CH ₃) | 2,5-(OCH ₃) ₂ | 3.40 | >12.0 | 6 |
| 12 | N(CH ₃) | 3,4-Cl ₂ | 1.00 | 3.00 | 6 |
| 13 | N(CH ₃) | 2,3-(CH ₃) ₄ | 0.87 | 8.20 | 6 |
| 14 | S | 3,4-(OCH ₃) ₂ | 2.60 | 16.7 | 6 |
| 15 | S | 3,4-Cl ₂ | 11.6 | 5.3 | 6 |
| 16 | S | 2,3-(CH ₃) ₄ | 0.81 | 3.0 | 6 |
| 17 | S | 3,4-(CH ₃) ₄ | 9.20 | 82.9 | 6 |
|  | | | | | |
| Compd | n | Ar | tgDHFR | rIDHFR | Ref. |
| 18 | 0 | C ₆ H ₅ | 28 | 59.9 | 7 |
| 19 | 0 | 3,4,5-(OCH ₃) ₃ C ₆ H ₂ | 2.2 | 13 | 7 |
| 20 | 0 | 2,3,4-(OCH ₃) ₃ C ₆ H ₂ | 1.8 | 3.7 | 7 |
| 21 | 0 | 2,4,6-(OCH ₃) ₃ C ₆ H ₂ | 0.84 | 1.88 | 7 |
| 22 | 0 | 2,4,5-(OCH ₃) ₃ C ₆ H ₂ | 1.5 | 7 | 7 |
| 23 | 0 | 2,5-(OCH ₃) ₂ C ₆ H ₃ | 1.7 | 3.5 | 7 |
| 24 | 0 | 3,5-(OCH ₃) ₂ C ₆ H ₃ | 3.0 | 32.4 | 7 |
| 25 | 0 | 3,4-(OCH ₃) ₂ C ₆ H ₃ | 14.2 | 52.3 | 7 |
| 26 | 0 | 2,4-(OCH ₃) ₂ C ₆ H ₃ | 2.4 | 0.9 | 7 |
| 27 | 0 | 3,4-Cl ₂ C ₆ H ₃ | 6.7 | 252 | 7 |
| 28 | 0 | 3,5-(OCH ₃) ₂ -4-OHC ₆ H ₂ | 1 | 15.3 | 7 |
| 29 | 0 | 2-NO ₂ -3,4-(OCH ₃) ₂ C ₆ H ₂ | 16.8 | >25 | 7 |
| 30 | 0 | 4-C ₆ H ₅ C ₆ H ₄ | 32.9 | 30.1 | 7 |
| 31 | 0 | 2-Naphthyl | 27 | 280 | 7 |
| 32 | 0 | 9-Fluorenyl | 30.5 | 29.8 | 7 |
| 33 | 1 | C ₆ H ₅ | 0.5 | 1.6 | 7 |
| 34 | 1 | NH-2,5-OCH ₃ C ₆ H ₃ | 49 | >29 | 7 |
| 35 | 1 | S-2-Naphtyl | 45 | 107 | 7 |

(continued on next page)

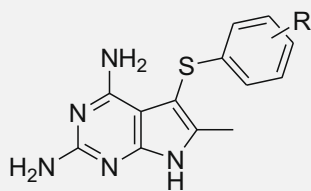
Table 1 (continued)



| Compd | X | Ar | tgDHFR | rIDHFR | Ref. |
|-------|---------------------|----------------------------------------------------------------|--------|--------|------|
| 36 | S | C ₆ H ₅ | >26 | 252 | 8 |
| 37 | S | 1-Naphthyl | 19 | 23 | 8 |
| 38 | S | 2-Naphthyl | 11.6 | 12.3 | 8 |
| 39 | NH | 1-Naphthyl | 37 | 12 | 8 |
| 40 | NH | 2-Naphthyl | 38 | 36.5 | 8 |
| 41 | O | 2-Naphthyl | >42 | 60.3 | 8 |
| 42 | NH | 4-C ₆ H ₅ OC ₆ H ₄ | 32.4 | 16.2 | 8 |
| 43 | NH | 2-C ₆ H ₅ C ₆ H ₄ | 45.4 | 137 | 8 |
| 44 | N(CH ₃) | 2-Naphthyl | 23.6 | 14.6 | 8 |
| 45 | NH | 2,5-Cl ₂ C ₆ H ₃ | >47 | 71.9 | 8 |
| 46 | N(CH ₃) | 3,4,5-Cl ₃ C ₆ H ₂ | 21.5 | 34.3 | 8 |
| 47 | NH | 1-Anthracene | 15.4 | >51 | 9 |
| 48 | NH | 2-Fluorene | 27.7 | >10 | 9 |
| 49 | NH | 3-(N-Ethylcarbazole) | 4.5 | 12.6 | 9 |
| 50 | NH | 2-(9-Hydroxyfluorene) | 63 | >13 | 9 |
| 51 | N(CH ₃) | 3-(2-Methoxydibenzofuran) | 392 | 241 | 9 |
| 52 | N(CH ₃) | 3-(N-Ethylcarbazole) | 22.6 | 24.7 | 9 |
| 53 | S | 2-Biphenyl | 105 | 49 | 9 |
| 54 | S | 3-Biphenyl | 23 | 31 | 9 |
| 55 | S | 4-Biphenyl | 79 | 351 | 9 |
| 56 | S | 2-C ₆ H ₅ OC ₆ H ₅ | 38 | 18 | 9 |
| 57 | S | 3-C ₆ H ₅ OC ₆ H ₅ | 19 | 39 | 9 |
| 58 | S | 4-C ₆ H ₅ OC ₆ H ₅ | 259 | 65 | 9 |

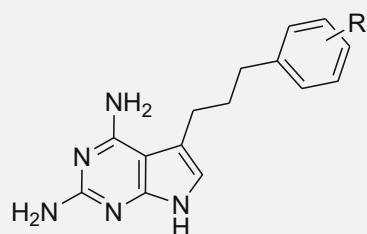


| Compd | R | tgDHFR | rIDHFR | Ref. |
|-------|--------------------------------------|--------|--------|------|
| 59 | 2,3-(CH) ₄ | 0.16 | 4.57 | 12 |
| 60 | 3-Cl | 6.94 | 47.7 | 12 |
| 61 | 4-Cl | 8.93 | 19 | 12 |
| 62 | 2-OCH ₃ | 1.02 | 2.32 | 12 |
| 63 | 3-OCH ₃ | 4.76 | 2.06 | 12 |
| 64 | 4-OCH ₃ | 2.69 | 31.3 | 12 |
| 65 | 2,4-Cl ₂ | 20 | 110 | 12 |
| 66 | 2,5-(OCH ₃) ₂ | 0.17 | 7.8 | 12 |

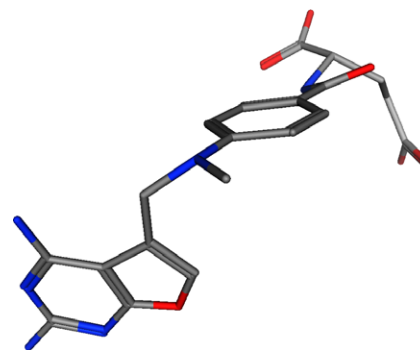
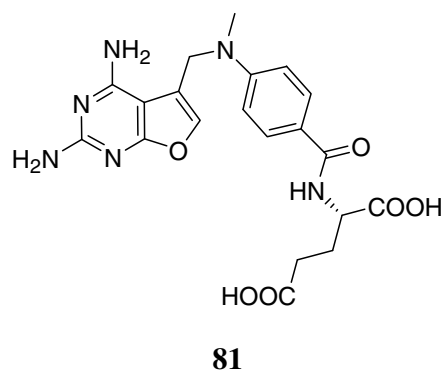


| Compd | R | tgDHFR | rIDHFR | Ref. |
|-------|--------------------------------------|--------|--------|------|
| 67 | 2,3-(CH) ₄ | 18.8 | 40 | 14 |
| 68 | 4-Cl | 31.8 | 37.4 | 14 |
| 69 | 3,4-(CH) ₄ | 31 | 31 | 14 |
| 70 | 3-Cl | 76.8 | 64 | 14 |
| 71 | 2,4-Cl ₂ | 28.2 | 34.6 | 14 |
| 72 | 2,5-(OCH ₃) ₂ | 87 | >46 | 14 |
| 73 | H | 81 | >33 | 14 |

Table 1 (continued)



| Compd | R | tgDHFR | rIDHFR | Ref. |
|-------|----------------------------------------|--------|--------|------|
| 74 | 2,5-Cl ₂ | 1.24 | 4.8 | 15 |
| 75 | 2,5-(OCH ₃) ₂ | 4.05 | 11.8 | 15 |
| 76 | 3,4,5-(OCH ₃) ₃ | 1.03 | 4.8 | 15 |
| 77 | 4-Benzoyloxy | 1.6 | 0.66 | 15 |
| 78 | 2,3-(CH) ₄ | 1.3 | 1.5 | 15 |
| 79 | 3,4-(CH) ₄ | 1.01 | 1.09 | 15 |
| 80 | H | 2.7 | 8.6 | 15 |

Figure 1. The structure and pcDHFR bound conformation of **81**.

active compounds against a particular enzyme were used to develop the CoMFA model. A diversity subset⁵⁶ which contained approximately 20% of the compounds (14 for rIDHFR and 15 for tgDHFR) with the farthest distance from the least active compound based on the 2-D fingerprints (MACCS Structural Keys bit packed version) using Tanimoto coefficient⁵⁶ as the similarity metric were extracted using MOE and was used as the test set. The remaining compounds were used as the training set. This was carried out to make sure that the test set is diverse enough to justify the models. Instead of the least active compound, any compound within the dataset could have been chosen to calculate the diversity subset, however, we used the least active compound here in all cases just to be consistent.

2.4. Conventional CoMFA

CoMFA was performed using the QSAR module in SYBYL 7.0. For each training set compound, the CoMFA descriptors, steric (Lenard-Jones 6–12 potential) and electrostatic (Coulombic potential)

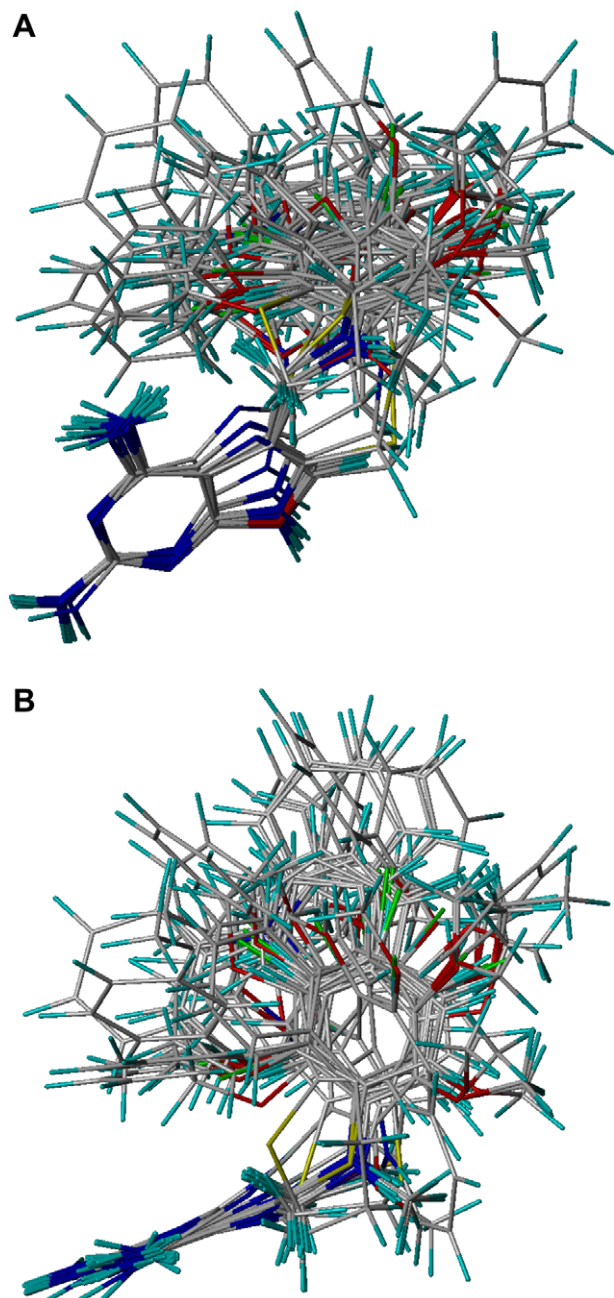


Figure 2. The compound aggregate. (A) Front view; (B) top view.

field energies were calculated using the *SYBYL* default parameters. The CoMFA region was defined to extend beyond the van der Waals envelopes of all molecules by 4.0 Å along the principal axes of the Cartesian coordinate system. A distance dependent dielectric constant was used. An sp^3 carbon atom with +1.0 charge was used as the probe atom to calculate steric and electrostatic fields. The steric and electrostatic contributions were truncated at 30 kcal/mol, and electrostatic contributions were dropped at lattice intersections with maximum steric interactions. The CoMFA steric and electrostatic fields generated were scaled by the CoMFA standard option in *SYBYL*. In order to improve the cross-validated q^2 to an acceptable level, a small number of outliers were dropped from the training set.

2.4.1. All-orientation search (AOS) CoMFA

As first reported by Cho and Tropsha⁵² the cross-validated r^2 (q^2) value of CoMFA analysis, which serves as a quantitative mea-

sure of the predictivity, fluctuates with the orientation of the aligned molecular aggregate on the computer screen by up to 0.5 q^2 unit. The reason for this fluctuation in q^2 values lies in the fact that conventional CoMFA samples the continuous molecular field at discrete lattice points and calculates the steric and electrostatic field energies at each lattice point with distance-sensitive functions, such as the Lennard–Jones 6–12 potential. When the molecular aggregate rotates, so does the molecular field surrounding the aggregate. The lattice box in CoMFA, however, is always axis-aligned and does not rotate along with the field. Thus, different points in the same molecular field are mapped onto the lattice points resulting in different field energy values. These values, when processed subsequently by partial least squares (PLS) to produce the final model causes a variation in the q^2 value and, consequently, the predictivity of the model.

The AOS routine⁵¹ optimizes the field sampling by rotating the molecular aggregate systematically and picking the orientation that produces the highest q^2 value. The details of the AOS routine were described previously elsewhere.⁵¹ Briefly, the compounds aggregate was rotated about the *x*, *y*, and *z* axes systematically with an increment of 30° using the *STATIC ROTATE* command in *SYBYL*. For each orientation, a conventional CoMFA was performed as described above and the predictive value of the model was evaluated using leave-one-out (LOO) cross-validation with sample-distance partial least squares (SAMPLS). The orientation that gave the highest q^2 value was selected to produce the final model. A *SYBYL* Programming Language (SPL) script was written to perform the AOS routine as described⁵¹ automatically. For the inhibitory activity against each enzyme, a new model was obtained by optimizing the model generated with the conventional CoMFA using the AOS routine.

2.4.2. The original q^2 -GRS CoMFA

The q^2 -GRS routine was developed by Cho and Tropsha⁵² to overcome the deficiency of CoMFA analysis mentioned above. The q^2 -GRS routine first subdivides the rectangular box obtained initially with conventional CoMFA, which, as mentioned above, extends beyond the van der Waals envelopes of all molecules by 4.0 Å along the principal axes of the Cartesian coordinate system, into 125 small boxes and perform 125 independent analyses using

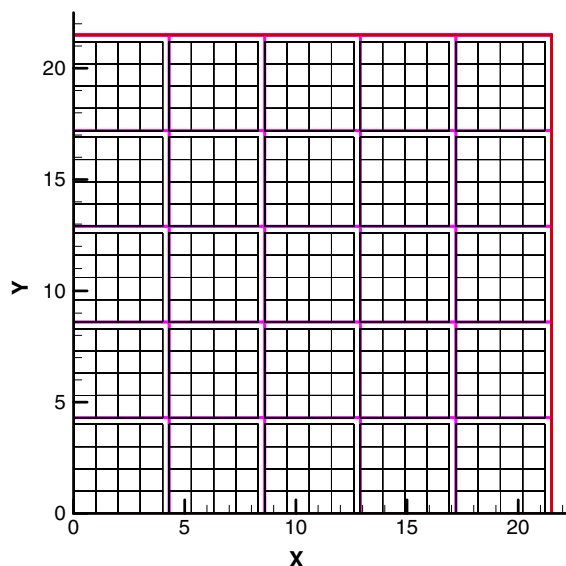


Figure 3. Front view of the q^2 -GRS grids (black), apparently, these grids do not coincide with that of the conventional CoMFA, and the distances between the grid points at the opposite side of the border (purple) of two adjacent subregions can be much less than 1.0 Å.

probe atoms placed within each box with the step size of 1.0 Å. Followed by the selection of only those small boxes for which a q^2 is higher than a specified optimal cutoff value (default 0.1). Finally, CoMFA analysis is repeated with the combined region of small boxes selected at the previous step to generate the final model.

Cho and Tropsha⁵² argued that it is contradictory for the conventional CoMFA routine to emphasize the limited areas of three-dimensional space as important for biological activity in the final result even though it assumes equal importance of all lattice points during the PLS analysis, which happens to be the calculation that leads to the final result. Thus q^2 -GRS was designed to eliminate those areas of three-dimensional space where changes in steric and electronic fields do not correlate with changes in biological activity from the analyses. For the inhibitory activity against each enzyme, a new model was obtained by optimizing the model generated with the conventional CoMFA using the original q^2 -GRS routine.

2.4.3. The modified q^2 -GRS CoMFA

A close inspection of the final combined region file generated by the original q^2 -GRS routine revealed a potential problem. For the training set of tgDHFR, the region box generated by conventional CoMFA has a dimension of $19.95 \text{ Å} \times 20.49 \text{ Å} \times 18.66 \text{ Å}$. If this region were to be defined using 1 Å as step size, there would be $20 \times 21 \times 19 = 7980$ lattice points. In the final combined region generated by q^2 -GRS using the default parameters, 74 boxes or

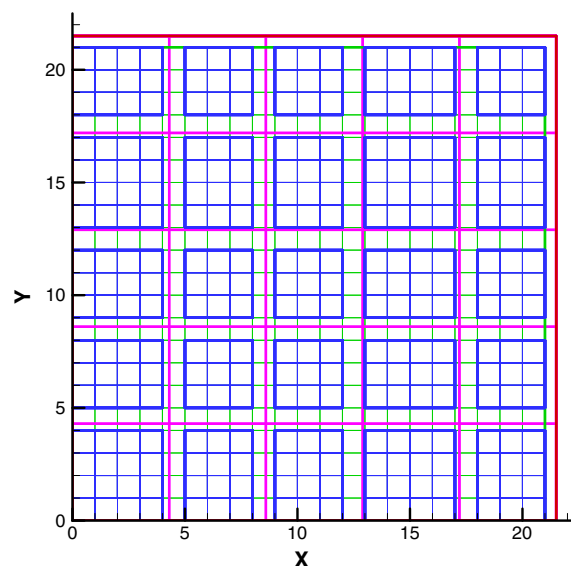


Figure 5. Front view of the modified q^2 -GRS grids using the first strategy (blue), which coincide with that of the conventional CoMFA (green).

subregions (59.2%) of the total 125 were selected and the rest 51 boxes (40.8%) were eliminated. Each of the subregions had the

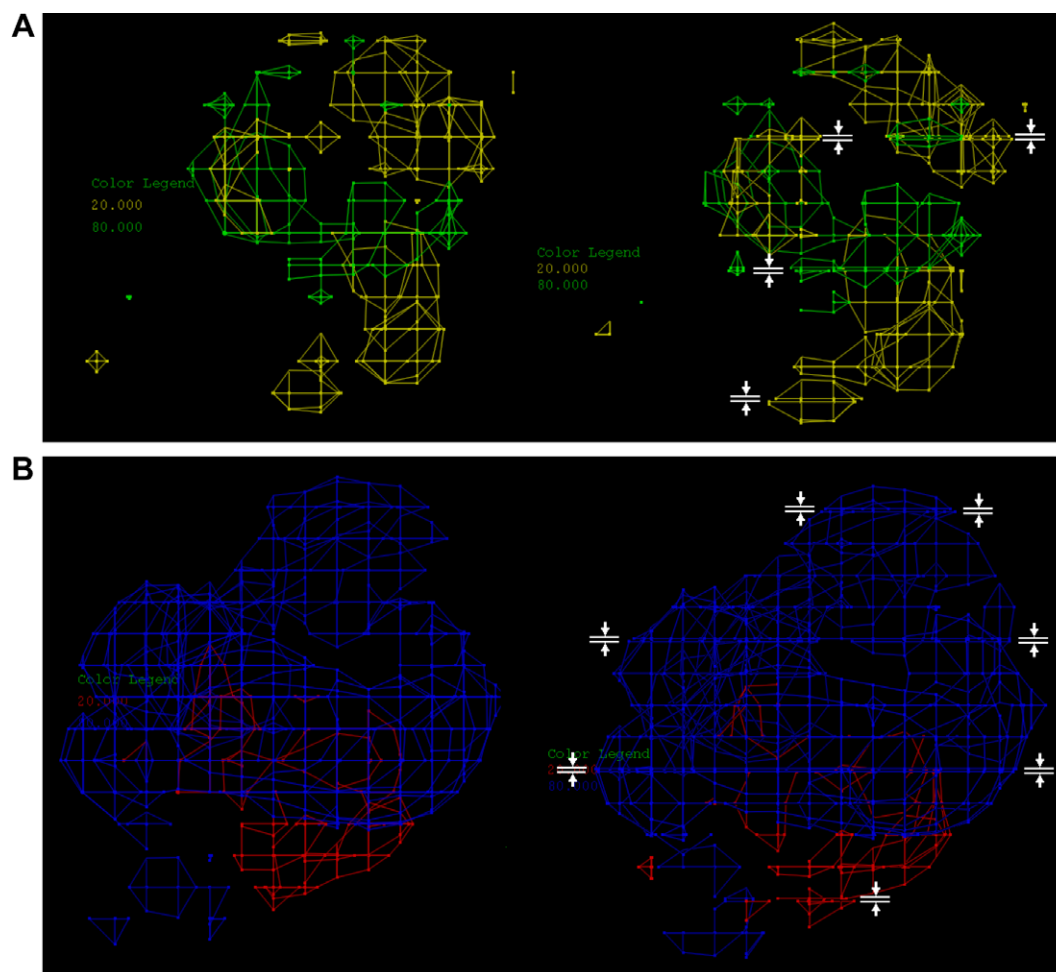


Figure 4. A comparison of the contour maps of the tgDHFR training set generated with the conventional CoMFA (left) and with the original q^2 -GRS with the cutoff set at -1000.0 (so that all subregions are selected) (right). The areas of the q^2 -GRS contour maps where the grid points are very close to each other (these areas are not present in the conventional CoMFA maps) are highlighted with the white symbols. (A) The steric contour maps, (B) the electrostatic contour maps.

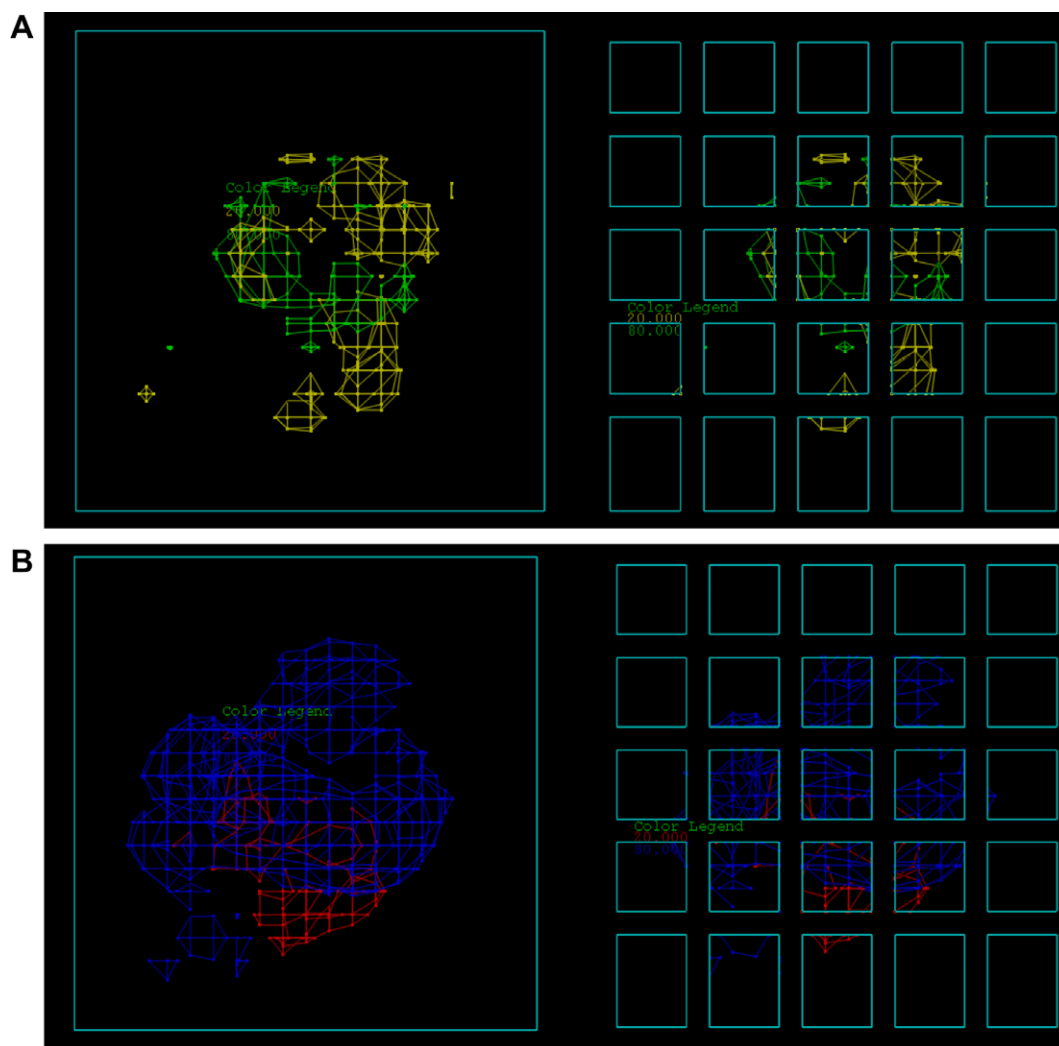


Figure 6. A comparison of the contour maps of the tgDHF training set generated with the conventional CoMFA (left) and with the first strategy modified original q^2 -GRS with the cutoff set at -1000.0 (so that all subregions are selected) (right). Superimposed on their respective region box(es) (cyan). (A) The steric contour maps, (B) the electrostatic contour maps.

dimension of one fifth of the region box generated by conventional CoMFA. In this case, it is $3.99 \text{ \AA} \times 4.10 \text{ \AA} \times 3.73 \text{ \AA}$. Thus, each box contained $5 \times 5 \times 4 = 100$ grid points. Thus the total number of grid points in the combined region is $74 \times 100 = 7400$ grid points, which is 96.5% of the total 7980 lattice points found in the conventional CoMFA region box. Apparently, the original q^2 -GRS routine forced 92.7% of the lattice points into 59.2% of the space. In the conventional CoMFA region, the distance between any given pair of immediately adjacent grid points is 1 \AA . In the combined region generated by the original routine of q^2 -GRS, in order to fit more grid points into the space, the distance between some of the lattice points must be less than 1 \AA . Since within each subregion, the distance of any given pair of immediately adjacent to each other is also 1 \AA , the shorter distances could only occur between grid points from adjacent boxes.

To understand this behavior of the original q^2 -GRS routine, let us consider a putative cubic region box, whose lowest corner is at $(0,0,0)$ in the Cartesian space, with the dimension of $21.5 \text{ \AA} \times 21.5 \text{ \AA} \times 21.5 \text{ \AA}$. From our observation, CoMFA normally rounds the dimensions down to the nearest integers. If the decimal part of a dimension is equal to or greater than 0.99, however, CoMFA treats this dimension as the rounded up integer, for example,

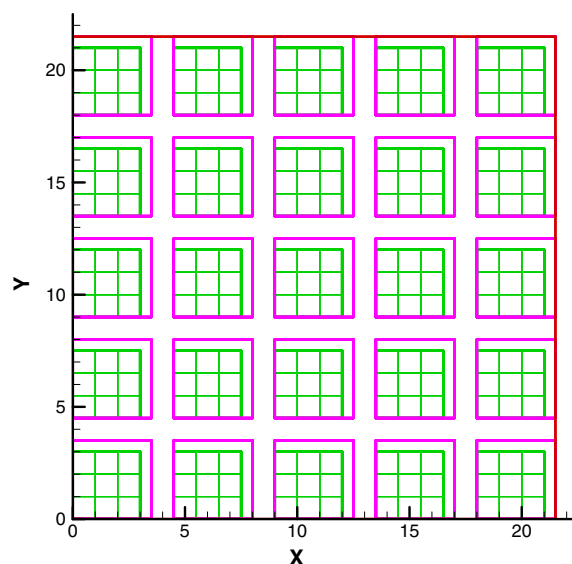


Figure 7. Front view of the modified q^2 -GRS subregions (purple) and grids (green) using the second strategy.

21.99 will be rounded up to 22 whereas 21.989 will still be rounded down to 21. Thus, the dimension of the conventional CoMFA grid in our example would be $21 \text{ \AA} \times 21 \text{ \AA} \times 21 \text{ \AA}$ and the number of grid points would be $22^3 = 10648$.

The q^2 -GRS routine divides the region box into 125 subregions. As in our example, the size of each subregion box is $4.3 \text{ \AA} \times 4.3 \text{ \AA} \times 4.3 \text{ \AA}$. CoMFA treats each individual subregion the same way as the conventional region box. In our example, the size of the grid within each subregion is rounded down to $4 \text{ \AA} \times 4 \text{ \AA} \times 4 \text{ \AA}$. The number of grid points within each subregion is $5^3 = 125$. Assuming that all the subregions are selected in the final combined region, there would be a total of $125^2 = 15625$ grid points. That is 47% more than the 10648 grid points in the conventional CoMFA grid. As shown in Figure 3, the closest distance occurs between the grid points at the opposite side of the border of two adjacent subregions. This distance is the decimal part of the dimension of the subregion, as in our example, it is 0.3 \AA . Apparently, in the worst case scenario, this distance could be 0 \AA , and in this extreme scenario some of the grid points would be counted twice, thrice or more (up to eight times).

Thus, in the final combined region, this uneven distribution of grid points will always be present, unless in the unlikely scenario where none of the selected subregions are adjacent to each other. The original q^2 -GRS introduces unnecessary weight on the three-

dimensional space near the borders of the adjacent subregions by putting more grid points at these sites.

To further illustrate the shortcoming of the original q^2 -GRS routine, we first conducted a conventional CoMFA analysis of the tgDHFR training set. For this analysis, the step size of the region file was set to 1.0 \AA instead of the default 2.0 \AA . The q^2 of this analysis was 0.627. Next, we did the analysis using the original q^2 -GRS routine, with the cutoff set to -1000.0 instead of the default 0.1 , so that all of the 125 regions were selected in the final region. At only 6.06, the q^2 of this analysis was significantly lower than that of the conventional CoMFA analysis. The steric and electrostatic stdev * coefficient contour maps derived from these two analyses were compared in Figure 4. There were areas in the q^2 -GRS contour maps where the grid points were in close vicinity to each other.

To overcome this problem, we modified the original q^2 -GRS routine with two different strategies.

In the first strategy, the lowest corner of each modified subregion is the lowest grid point of the conventional CoMFA grid enclosed by the original q^2 -GRS subregion, and the highest corner of each modified subregion is the highest grid point of the conventional CoMFA grid enclosed within the subregion of the original q^2 -GRS routine (Fig. 5). The advantage of this strategy is that only the grid points of the conventional CoMFA grid would be included in the final combined region.

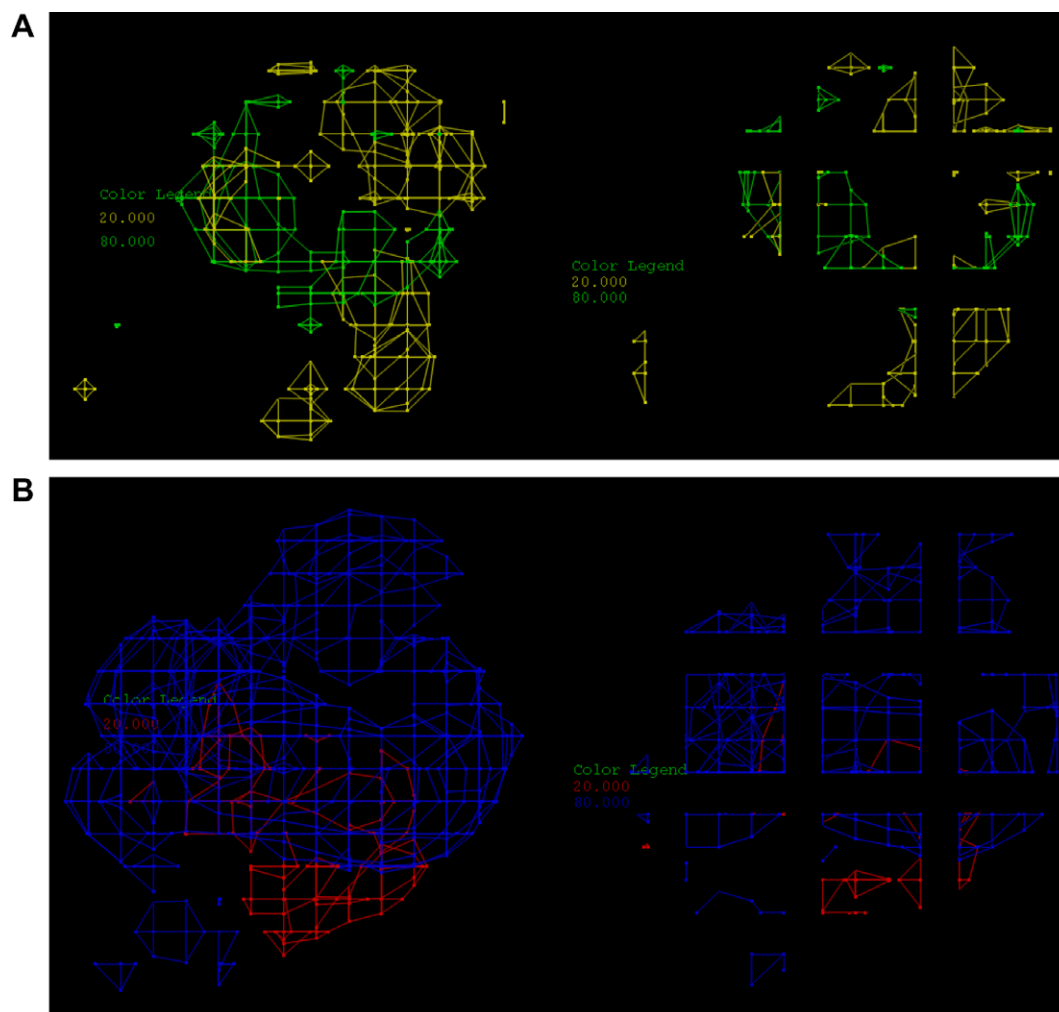


Figure 8. Comparison of the contour maps of the tgDHFR training set generated with the conventional CoMFA (left) and with the second strategy modified original q^2 -GRS with the cutoff set at -1000.0 (so that all subregions are selected) (right). (A) the steric contour maps, (B) the electrostatic contour maps.

We tested this modified routine on the tgDHFR training set by lowering the q^2 cutoff to -1000.0 so that all the 125 subregions were included in the final combined region, which gave exactly the same q^2 (0.627) as that of the conventional CoMFA (with 1.0 \AA as the step size instead of the default 2.0 \AA). As shown in Figure 6, except for those gaps between subregions, the stdev * coefficient contour maps derived from this analysis were exactly the same as that of the original CoMFA analysis.

The drawback of this strategy is that the size of the subregions varies. In our putative example illustrated in Figure 5, the number of grid points in each subregion varies from 64 to 125, making the subregions less comparable to each other. Therefore, a modification using a second strategy was devised. In this second strategy, the region box is divided into 125 equal-sized subregions and the distance between the adjacent subregions is 1 \AA (Fig. 7), whereas in the original routine, the adjacent subregions touch each other. In this case, even though the grid points do not match that of the conventional region, the size of the subregions and the number of grid points within these subregions are the same, much like the original routine.

This second modified routine was also tested on the tgDHFR training set by lowering the q^2 cutoff to -1000.0 , which gave a q^2 of 0.631; this was slightly higher than that of the conventional CoMFA using step size 1.0 \AA (0.627). The stdev * coefficient contour maps from this analysis was compared with the conventional CoMFA in Figure 8.

For the inhibitory activity against each enzyme, models were obtained by using both modified q^2 -GRS routine.

2.5. PLS analysis

The CoMFA, descriptors derived above were used as explanatory variables, and pIC_{50} ($-\log \text{IC}_{50}$) values were used as the target variable in PLS regression analyses to derive 3D QSAR models using the implementation in the *SYBYL* package. The predictive value of the models was evaluated by leave-one-out (LOO) cross-validation with SAMPLS. The cross-validated coefficient, q^2 , was calculated using Eq. 1,

$$q^2 = \frac{\sum (Y_{\text{pred}} - Y_{\text{actual}})^2}{\sum (Y_{\text{actual}} - Y_{\text{mean}})^2} \quad (1)$$

Where Y_{pred} , Y_{actual} , and Y_{mean} are predicted, actual, and mean values of the target property (pIC_{50}), respectively. $\sum (Y_{\text{pred}} - Y_{\text{actual}})^2$ is the predictive sum of squares (PRESS). The number of components giving the lowest PRESS value or the optimal number of components (ONC) was used to generate the final PLS regression models. To prevent overfitting, no more than 7 components were allowed. The conventional correlation coefficient r^2 and its standard error, s , were subsequently computed for the final PLS models using the following formulae, respectively.

$$r^2 = \frac{\sum (Y'_{\text{pred}} - Y_{\text{actual}})^2}{\sum (Y_{\text{actual}} - Y_{\text{mean}})^2} \quad (2)$$

$$S = \sqrt{\frac{\sum (Y'_{\text{pred}} - Y_{\text{actual}})^2}{N}} \quad (3)$$

Where Y'_{pred} is the predicted value of the target property (pIC_{50}) using the final model. The statistical F -test was also performed and the results indicated that all models were statistically significant.

CoMFA coefficient maps were generated by interpolation of the pairwise products between the PLS coefficients and the standard deviations of the corresponding CoMFA descriptor values.

3. Results and validation

The tgDHFR training set has a total of 61 compounds, which include **2–12**, **14–27**, **30**, **31**, **33**, **37–39**, **42–44**, **46–50**, **53–55**, **57–64**, **67–73**, **76**, **78–80** (Table 1). The tgDHFR test set contains 15 compounds, which are **13**, **28**, **29**, **32**, **34**, **35**, **40**, **51**, **52**, **56**, **65**, **66**, **74**, **75** and **77** (Table 1). A total of five models were generated using the conventional CoMFA, AOS optimized CoMFA, the original q^2 -GRS optimized CoMFA and the modified q^2 -GRS optimized CoMFA using the first and the second strategy. The key statistical parameters associated with these models are shown in Table 2. The graph the actual pIC_{50} versus the predicted pIC_{50} values for the training set and the test set by the five CoMFA models are shown in Figure 9A–E, respectively. The predicted pIC_{50} values of the training set compounds using the five models as well as the residual values are given in Appendix Tables 1, 3, 5, 7 and 9, respectively. The predicted pIC_{50} values of the test set compounds and the residual values using the five models are given in Appendix Tables 2, 4, 6, 8 and 10, respectively.

Though, AOS CoMFA gave the best internal cross-validated q^2 value 0.699 (ONC = 7) among the five models, this q^2 value was only obtained with the maximum allowed seven components. Furthermore, the AOS CoMFA model only gave the lowest external predictive r^2 (0.365), which was worse than that of the conventional CoMFA (0.422). In contrast, the original q^2 -GRS CoMFA model showed the lowest internal cross-validated q^2 value 0.622 (ONC = 3), lower than that of the conventional CoMFA model (0.639, ONC = 5), it however had the second best external predictive r^2 (0.465), whereas the modified q^2 -GRS model using the first strategy had a q^2 value of 0.639 (ONC = 3), which was the same as that of the conventional CoMFA, and its external predictive r^2 value (0.447) was only slightly better than that of the conventional model.

Finally, the q^2 -GRS model modified with the second strategy afforded the best external predictive r^2 (0.499) and also gave a satisfactory internal cross-validated q^2 at 0.647 (ONC = 3), which was the second best among the five models only after the AOS CoMFA. The stdev * coefficient contour maps generated using this model is further discussed in the next section.

The rIDHFR training set has a total of 57 compounds, which include **1–6**, **8**, **13–27**, **30**, **31**, **33**, **36–41**, **43**, **44**, **52–54**, **56–63**, **66–71**, **74–76**, **78–80** (Table 1). The rIDHFR test set contains 14 compounds, which are **7**, **9**, **12**, **28**, **32**, **35**, **42**, **45**, **49**, **51**, **55**, **64**, **65** and **77** (Table 1). A total of five models were generated using the conventional CoMFA, AOS optimized CoMFA, the original q^2 -GRS optimized CoMFA and the modified q^2 -GRS optimized CoMFA using the first and the second strategy. The key statistical parameters

Table 2
Statistical data for tgDHFR CoMFA results

| | Conventional | AOS | q^2 -GRS (original) | q^2 -GRS (modified 1) | q^2 -GRS (modified 2) |
|------------------------|--------------|---------|-----------------------|-------------------------|-------------------------|
| CV- r^2 (q^2) | 0.639 | 0.699 | 0.622 | 0.639 | 0.647 |
| Opt. no. of comp. | 5 | 7 | 3 | 3 | 3 |
| Std. error of estimate | 0.236 | 0.141 | 0.322 | 0.291 | 0.311 |
| Non-CV- r^2 | 0.897 | 0.964 | 0.800 | 0.837 | 0.814 |
| F value | 95.613 | 204.900 | 76.108 | 97.300 | 83.198 |
| Predictive r^2 | 0.422 | 0.365 | 0.465 | 0.447 | 0.499 |

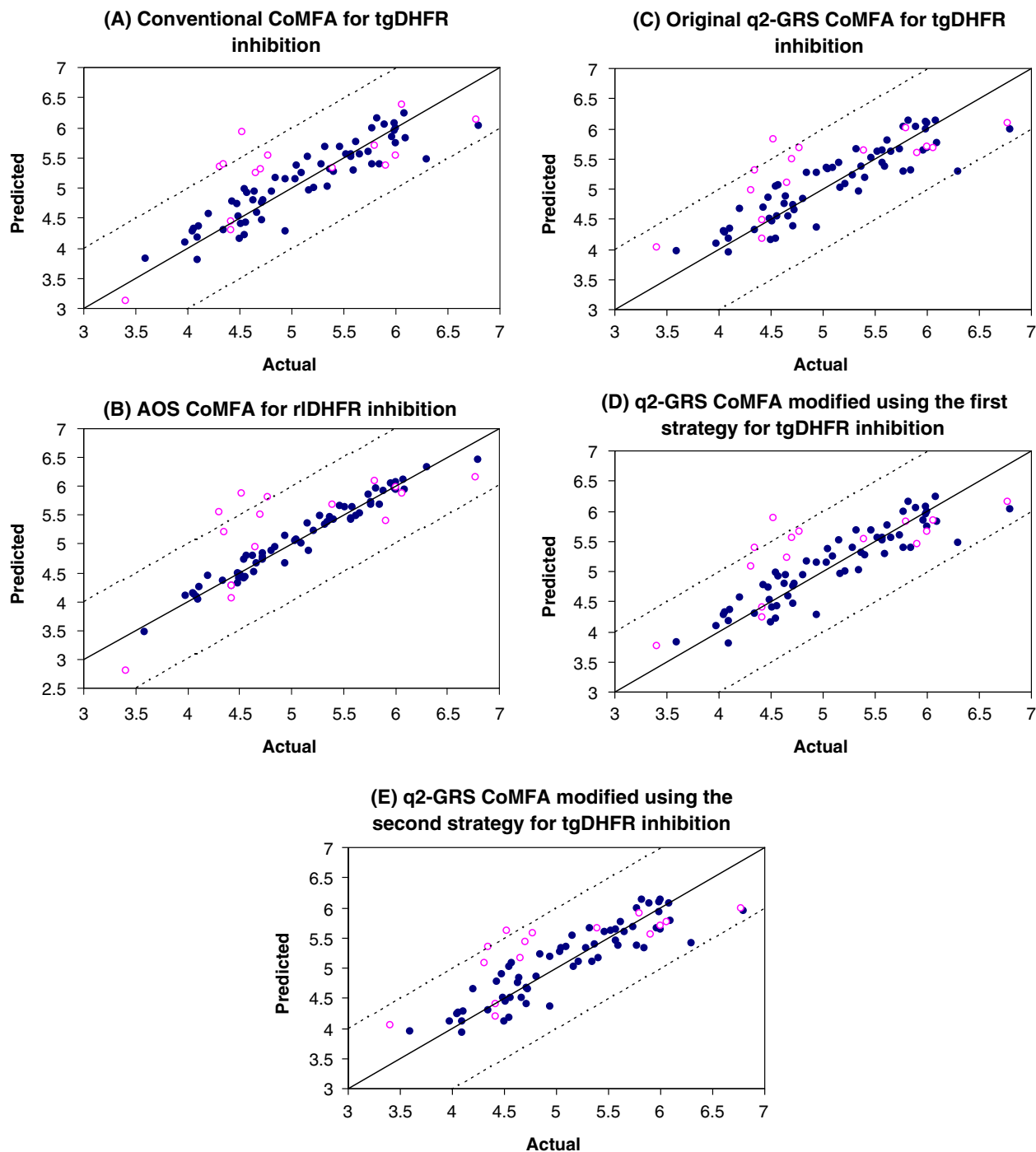


Figure 9. Predictions for the training (●) and test (○) sets for tgDHFR inhibitory activities. The solid line is the regression line for the training set predictions whereas the dotted lines indicate the ± 1.0 log margins.

Table 3

Statistical data for rIDHFR CoMFA (value from original training set/value from the training set after dropping three outliers)

| | Conventional | AOS | q ² -GRS (original) | q ² -GRS (modified 1) | q ² -GRS (modified 2) |
|-----------------------------------------------------|---------------|---------------|--------------------------------|----------------------------------|----------------------------------|
| CV- <i>r</i> ² (<i>q</i> ²) | 0.367/0.465 | 0.478/0.536 | 0.375/0.452 | 0.415/0.468 | 0.419/0.431 |
| Opt. no. of comp. | 3/3 | 3/3 | 3/3 | 3/3 | 3/3 |
| Std. error of estimate | 0.348/0.317 | 0.338/0.294 | 0.380/0.334 | 0.366/0.329 | 0.367/0.339 |
| Non-CV- <i>r</i> ² | 0.703/0.715 | 0.720/0.755 | 0.646/0.682 | 0.672/0.692 | 0.671/0.673 |
| <i>F</i> value | 41.822/41.812 | 45.837/51.373 | 32.246/35.810 | 36.150/37.454 | 36.003/34.252 |
| Predictive <i>r</i> ² | −0.025/0.267 | 0.163/0.281 | 0.171/0.296 | 0.233/0.326 | 0.188/0.320 |

associated with these models are shown in Table 3. The graph of the actual pIC_{50} versus the predicted pIC_{50} versus the predicted pIC_{50} for the training set and test set by the five CoMFA models are shown in Figure 10A–E, respectively. The predicted pIC_{50} values of the training set compounds using the five models as well as the residual values are given in Tables 11, 13, 15, 17 and 19 in Appendix, respectively. The predicted pIC_{50} values of the test set compounds and the residual values using the five models are given in Tables 12, 14, 16, 18 and 20 in Appendix, respectively.

A $q^2 \geq 0.5$ is generally considered as an indication that the model is internally predictive. In this case, however, none of the models afforded a q^2 greater than 0.5. However, it was evident that

both modified q^2 -GRS routines gave better q^2 values than that of the original routine. AOS CoMFA continued to give the highest q^2 values among the five models and the first strategy modified q^2 -GRS model afforded the best predictive r^2 .

In order to obtain a reasonable CoMFA model for rIDHFR, three outliers **21**, **26** and **33**, which had an absolute residual value greater than 1.0 during the cross-validation of the conventional CoMFA model, were dropped from the previous rIDHFR training set, while the test set compounds remained the same. A total of five models were again generated using the conventional CoMFA, AOS optimized CoMFA, the original q^2 -GRS optimized CoMFA and the modified q^2 -GRS optimized CoMFA using the first and the second

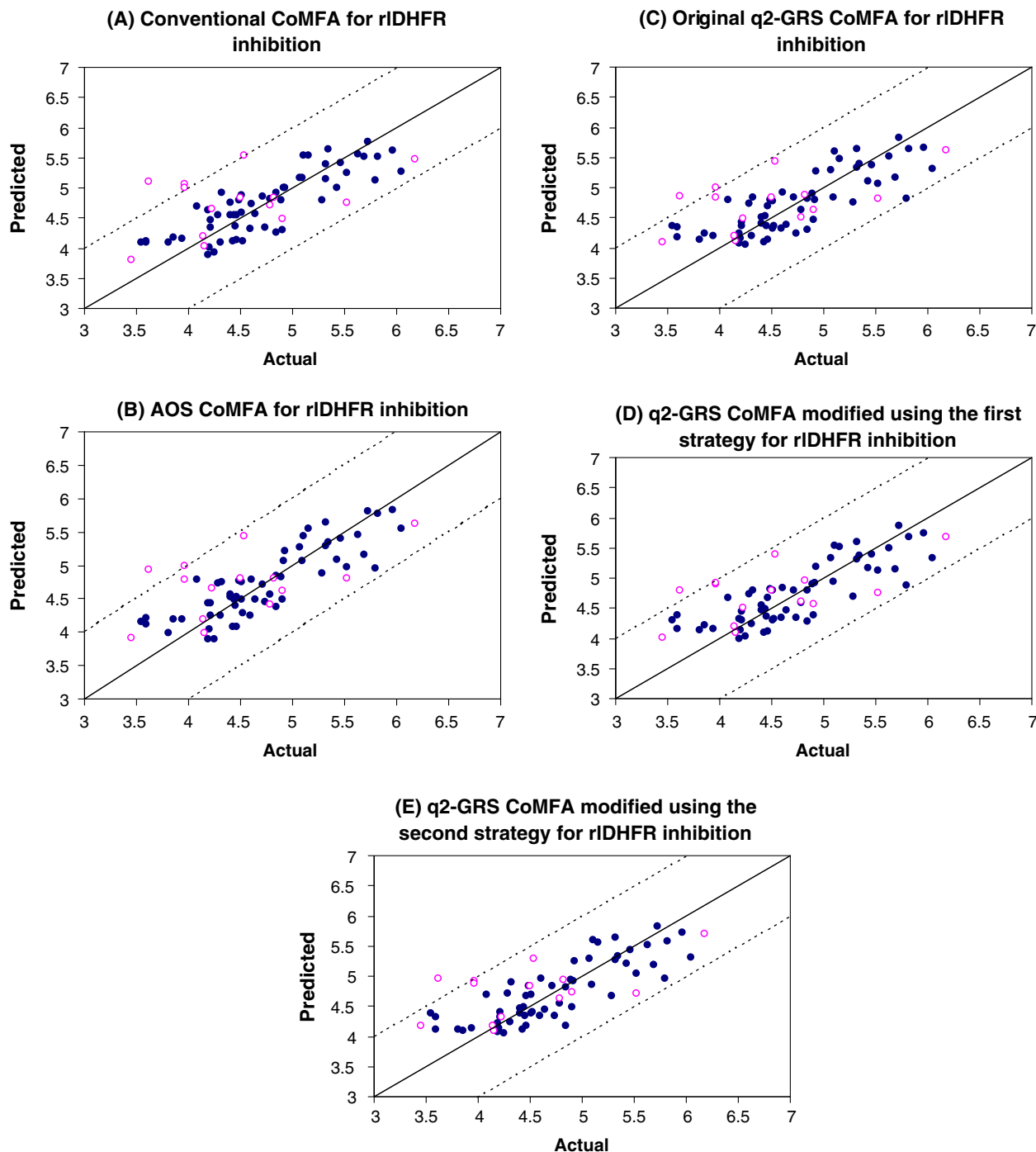


Figure 10. Predictions for the training (●) and test (○) sets for rIDHFR inhibitory activities. The solid line is the regression line for the training set predictions whereas the dotted lines indicate the ± 1.0 log margins.

strategy. The key statistical parameters associated with these models are shown in Table 3. The graph of the actual pIC_{50} versus the predicted pIC_{50} values for the training set and test set by the five CoMFA models are shown in Figure 11A–E, respectively. The predicted pIC_{50} values of the training set compounds using the five models as well as the residual values are given in Tables 21, 23, 25, 27 and 29 in Appendix, respectively. The predicted pIC_{50} values of the test set compounds and the residual values using the five models are given in Tables 22, 24, 26, 28 and 30 in Appendix, respectively.

In this case, only AOS CoMFA model gave a q^2 value greater than 0.5 (0.536, $\text{ONC} = 3$). The modified q^2 -GRS model using

the first strategy gave the second best q^2 value (0.468, $\text{ONC} = 3$) and the best predictive r^2 (0.326). The stdev * coefficient contour maps of the AOS CoMFA is further discussed in the following section.

4. The stdev * coefficient contour maps

4.1. tgDHFR

The second strategy modified q^2 -GRS CoMFA model was used to construct the stdev * coefficient contour maps for the tgDHFR (Fig. 12).

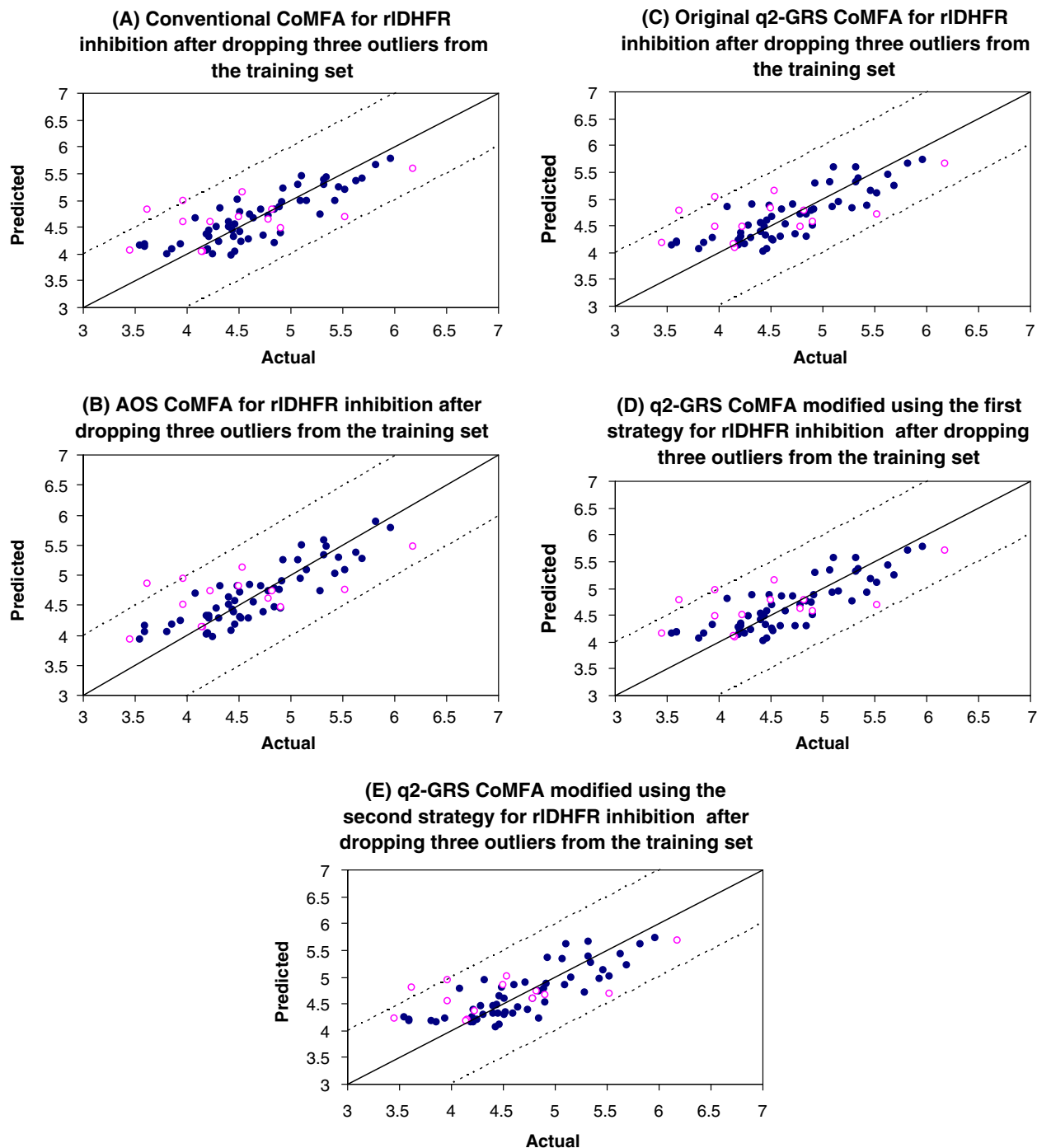


Figure 11. Predictions for the training (●) and test (○) sets for rIDHFR inhibitory activities (three outliers were dropped from the training set). The solid line is the regression line for the training set predictions whereas the dotted lines indicate the ± 1.0 log margins.

In the CoMFA steric field (Fig. 12A), the green (sterically favorable) and yellow (sterically unfavorable) contours represent 80% and 20% level contributions, respectively.

The 2'-substituent (on the phenyl ring) of 5-substituted pyrrolo[2,3-*d*]pyrimidines and furo[2,3-*d*]pyrimidines such as compounds **6**, **11**, **45**, **72** and **74** (Table 1) lies between a yellow region and a green region. A cluster of several small green regions which includes the aforementioned green region is found near the 3'-substituent of 5-substituted pyrrolo[2,3-*d*]pyrimidines and furo[2,3-*d*]pyrimidines such as compounds **2**, **4**, **14**, **46**, **70** and **76** (Table 1). Yellow regions are found near 4'-substituent and 5'-substituent of the same classes of compounds. The 1-naphthyl moiety, attached to the 5-position of a pyrrolo[2,3-*d*]pyrimidine or furo[2,3-*d*]pyrimidine including compounds **9**, **13**, **16**, **37**, **39**, **67** and **78** has more contact with green regions than yellow regions. However, the reversed can be said for the 2-naphthyl moiety (compounds **17**, **31**, **38**, **40**, **44** and **79**). This is evident because compounds **16**, **37**, **39**, **78** that contain the 1-naphthyl moiety all have greater predicted activities than their 2-naphthyl analogs **17**, **38**, **40** and **79**, respectively. A small green region is also found near the bridge atom of all 5-substituted compounds that is directly attached to the 5-position of pyrrolo[2,3-*d*]pyrimidines or furo[2,3-*d*]pyrimidines. A green region is also located near the 9-methyl group of compounds **11**, **12**, **13**, **44**, **46**, **51** and **52** (Table 1). The 6-methyl group of compounds **67–73** (Table 1) lies within a yellow region.

For the 5-methyl-6-phenylsulfanyl-substituted pyrro[2,3-*d*]pyrimidines including compounds **59–66** (Table 1), both 2'-

and 3'-substituents are surrounded by a number of small green regions. The 6-S bridge is also close to a yellow region. There are two green regions also near the 5' and 6'-substituent position. Some small green regions are found near the 5-methyl group.

For the purine analogs (compounds **18–35**, Table 1), green regions are found near the 2'-, 3'- and 5'-substituents and yellow regions near 4'- and 6'-substituents.

Some yellow regions are found near the 7-position of the ring systems indicating that the furo[2,3-*d*]pyrimidine is more sterically conducive to the activity than pyrrolo[2,3-*d*]pyrimidines, whose 7-NH is slightly bulkier than the oxygen. The ring of the purine compounds is slightly tilted as compared to the other analogs, positioning it further away from these yellow regions, indicating that the purine ring is probably better than the other two ring systems in terms of steric effect for the inhibitory activity against tgDHFR.

The CoMFA electrostatic contour map for tgDHFR inhibitory activity is depicted in Figure 12B. The red (negative charge favorable) and blue (negative charge unfavorable) contours in the CoMFA electrostatic field represent 80% and 20% level contributions, respectively. The molecules are enveloped by blue regions. Some patches of red regions are also present.

Large blue regions are found on top of the bridge atoms. Blue regions are also found near the amino protons. Red regions are present in the vicinity of the 7-NH and 7-O of the pyrrolo[2,3-*d*]pyrimidine and furo[2,3-*d*]pyrimidines, respectively. The 6-methyl group of compounds **67–73** (Table 1) lies near a red region.

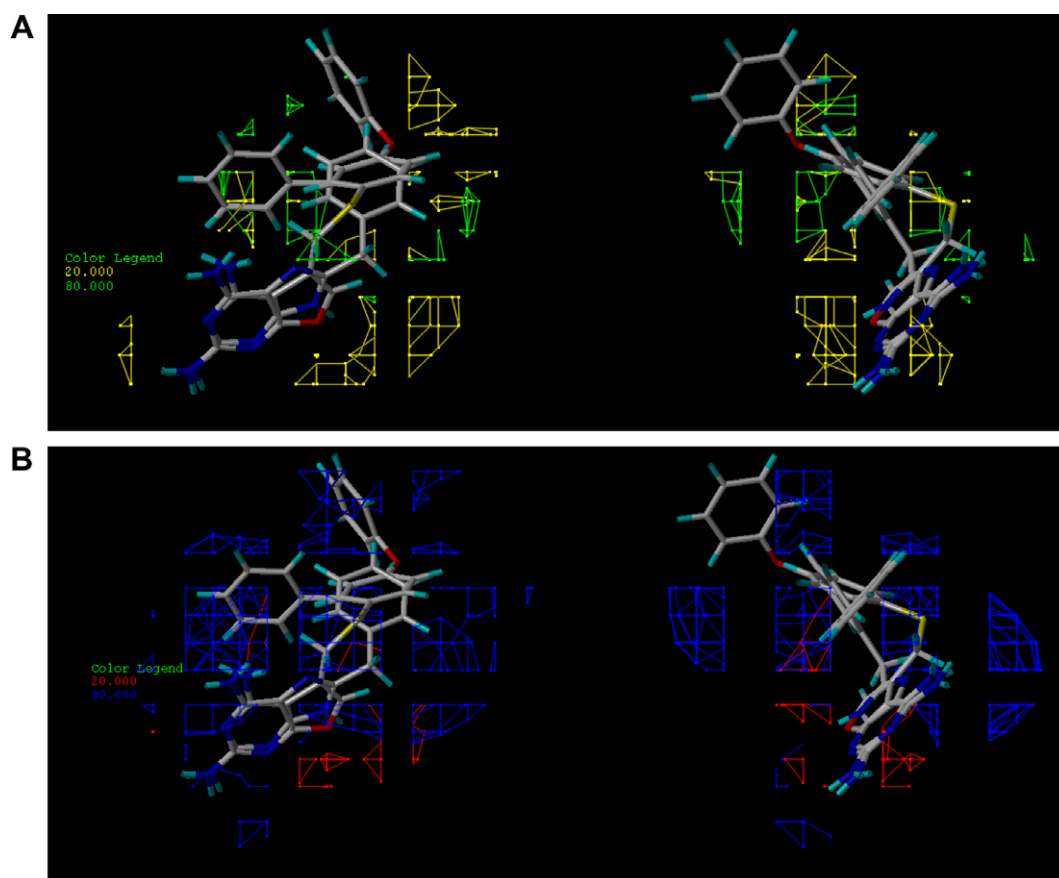


Figure 12. (A) Orthogonal view of the steric fields generated with the second strategy modified q^2 -GRS CoMFA model of tgDHFR: yellow indicates regions where bulky groups decrease activity, whereas green indicates regions where bulky groups increase activity. (B) Orthogonal view of the electrostatic fields generated with the second strategy modified q^2 -GRS CoMFA model of tgDHFR: red indicates regions where more negatively charged groups increase activity, whereas blue indicates regions where more positively charged groups increase activity.

For the purine analogs (compounds **18–35**, Table 1), two red regions and one blue region are found near the 2'-methoxy oxygen (e.g., compound **21**, Table 1) and one red region near the 2'-methoxy methyl. A few blue regions are close to the 3'-methoxy methyl group. Some blue regions are found near the 4'-methoxy methyl of the purine analogs. One blue region is found near 5'-methoxy methyl of the purine analogs.

For the 5-methyl-6-phenylsulfanyl-substituted pyrro[2,3-*d*]-pyrimidines including compounds **59–66** (Table 1), a red region is near the 2'-methoxy oxygen. The 3'-, 4'-, 5'- and 6'-substituents are surrounded by blue regions. A blue region is located near the 5-methyl group. Some red regions are found near the 6-sulfur bridge.

4.2. rIDHFR

The AOS CoMFA model was used to construct the stdev * coefficient contour maps for the rIDHFR (Fig. 13). In the CoMFA steric field (Fig. 13A), the green (sterically favorable) and yellow (sterically unfavorable) contours represent 80% and 20% level contributions, respectively.

The 1-naphthyl moiety of 5-substituted pyrrolo[2,3-*d*]pyrimidines and furo[2,3-*d*]pyrimidines with a $-\text{CH}_2\text{-NH}-$ bridge is in contact only with the yellow region whereas the 2-naphthyl contacts both the green and the yellow regions. In contrast, the 2-naphthyl moiety of 5-substituted pyrrolo[2,3-*d*]pyrimidines and furo[2,3-*d*]pyrimidines with a $-\text{CH}_2\text{-S}-$ bridge is in contact only with the yellow region whereas the 1-naphthyl contacts both the green and the yellow regions. For the 3-carbon bridge at the 5-position of pyrrolo[2,3-*d*]pyrimidines and furo[2,3-*d*]pyrimidines, both the 1- and 2-naphthyl contact both the green and yellow re-

gions, but 2-naphthyl seems to have more contact with the green region. A green region is found near the 9-methyl group of compounds **11**, **12**, **13**, **44**, **46**, **51** and **52** (Table 1).

The 2'-substituents of the purines lie in the green region, part of 3'-substituent lies inside the green region, but it is also in contact with the yellow region. The 4'-OCH₃ is found inside the yellow region and the 5'-OCH₃ hardly touches that yellow region. The 2-biphenyl contacts some small yellow regions, however, 3-biphenyl is in the previously mentioned green area. The 2-naphthyl of the purine analogs lies in a yellow region.

For the 5-methyl-6-substituted pyrrolo[2,3-*d*]pyrimidines, the 1-naphthyl is in contact with the green region. 2'-OCH₃ and 3'-OCH₃ lie within the green region.

The CoMFA electrostatic contour map for rIDHFR inhibitory activity is depicted in Figure 13B. The red (negative charge favorable) and blue (negative charge unfavorable) contours in the CoMFA electrostatic field represent 80% and 20% level contributions, respectively.

For the 5-substituted pyrrolo[2,3-*d*]pyrimidines and furo[2,3-*d*]pyrimidines, one relatively large blue region is found near the 2'- and 3'-substituents along with two small red regions located near the 2'-substituent alone. The 4'-substituent overlaps with a red region. A moderately sized blue region is found between the 5'- and 6'-substituents. A small blue region is near the 6'-substituent alone. This small blue region is also close to the sulfur atom of the $-\text{CH}_2\text{-S}-$ bridge in compounds **14–17**, **36–38** and **53–58** (Table 1) and the nitrogen atom of the $-\text{CH}_2\text{-NH}-$ or $-\text{CH}_2\text{-N}(\text{CH}_3)-$ bridge of compounds **1–13**, **39**, **40** and **42–52** (Table 1). One blue region is found near the bridge carbon atom which is directly attached to the 5-position of the furo[2,3-*d*]pyrimidines or

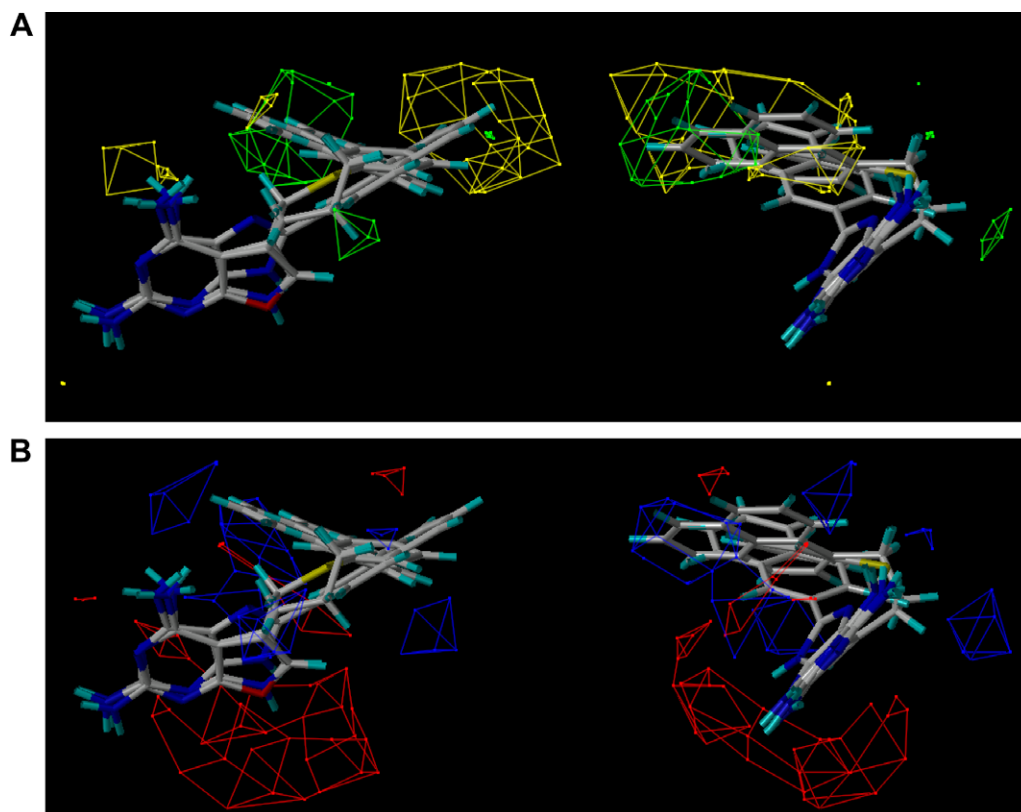


Figure 13. (A) Orthogonal view of the steric fields generated with AOS CoMFA model of rIDHFR (after dropping three outliers from the training set): yellow indicates regions where bulky groups decrease activity, whereas green indicates regions where bulky groups increase activity. (B) Orthogonal view of the electrostatic fields generated with the AOS CoMFA model of rIDHFR (after dropping three outliers from the training set): red indicates regions where more negatively charged groups increase activity, whereas blue indicates regions where more positively charged groups increase activity.

pyrrolo[2,3-*d*]pyrimidines. Three large red regions are located near the 7-position –NH– or –O– for the pyrrolo[2,3-*d*]pyrimidine and furo[2,3-*d*]pyrimidine, respectively.

For the purine analogs (compounds **18–35**, Table 1), one relatively large blue region is found near the 2'- and 3'-substituents with two small red regions overlapping with the 2-substituents and a red region near the 3'-substituent. One small blue region is found near the oxygen of the 5'-methoxy group.

In the case of 5-methyl-6-substituted pyrrolo[2,3-*d*]pyrimidines including compounds **59–66** (Table 1), the 2'- and 3'-substituent each overlaps with a red region and a small blue region is located near the 5'-region. The sulfur bridge is sandwiched between a large red region and a relatively small blue region.

4.2.1. Application of the model

As an initial application we elected to use the steric contour maps of the tgDHFR and rIDHFR CoMFA (Fig. 14A and B, respectively), which indicated that selective tgDHFR inhibitory activity

over rIDHFR activity may be possible for the 2,4-diamino-5-methyl-6-phenylsulfanyl-substituted pyrrolo[2,3-*d*]pyrimidines, if both sides of the phenyl ring are substituted (i.e., any combination of 2', 3' and 5', 6'). Indeed, among the eight previously reported 5-methyl-6-phenylsulfanyl-substituted pyrrolo[2,3-*d*]pyrimidines **59–66**,¹² the only compound that fits this criteria was **66** (2',5'-diOMePhenyl), which was the most potent tgDHFR inhibitor among the series (**59–66**) with an IC₅₀ value of 0.17 μM similar to **59** (1-naphthyl, IC₅₀ = 0.16 μM) and was also the most selective tgDHFR inhibitor as compared to rIDHFR with a selectivity ratio of 45.9 in contrast to 28.6 for **59**.

Thus it was of interest to synthesize additional analogs with disubstitution on either side of the phenyl ring as suggested by the steric contour maps. We synthesized and evaluated compounds **90–94** (Scheme 1). Four of these compounds (**91–94**) had substituents on both sides of the phenyl ring. The 2-naphthyl compound **90** was designed on the basis of the excellent activity and selectivity of the closely related 1-naphthyl compound **59** described above.

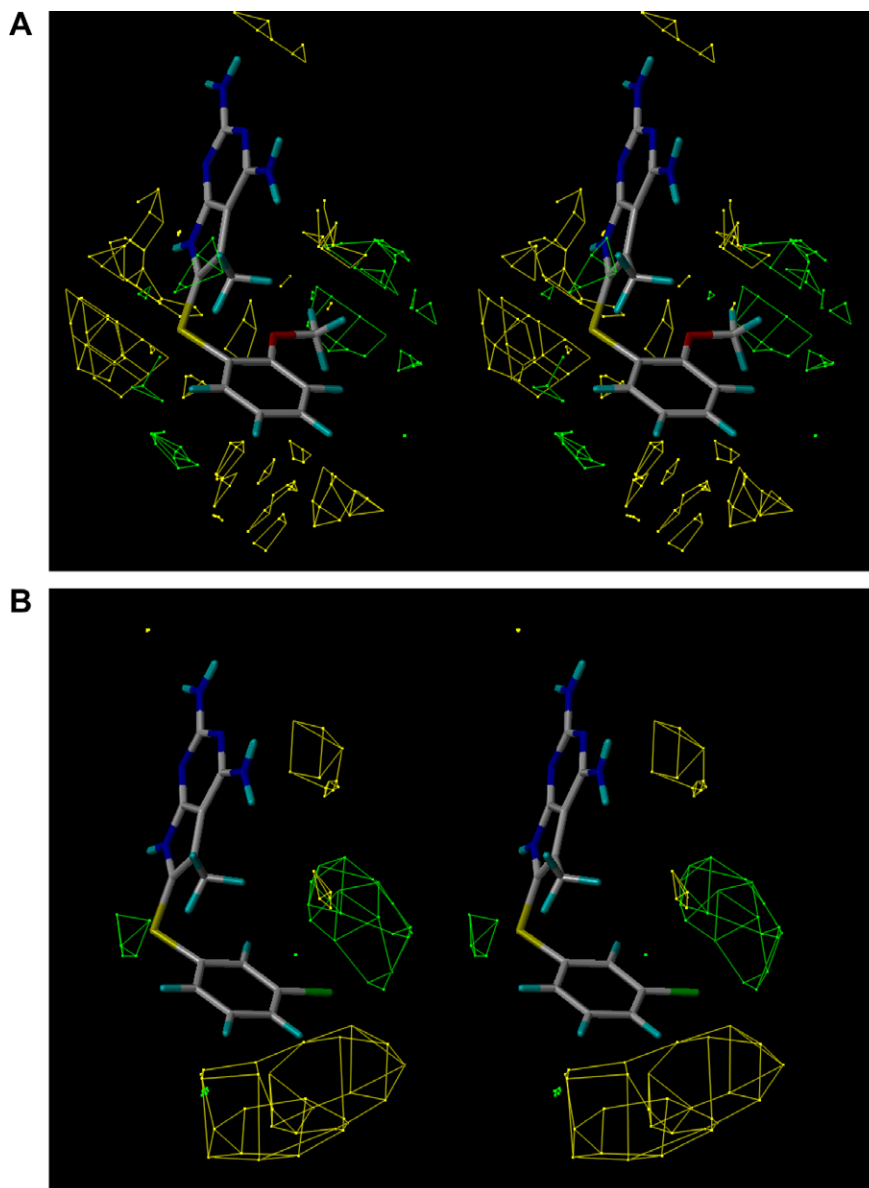


Figure 14. (A) Stereo view of the steric fields of the tgDHFR CoMFA model with the structure of compound **62**, there are green (steric favorable) regions near both sides of the phenyl moiety. (B) Stereo view of the steric fields of the rIDHFR CoMFA model with the structure of compound **60**, only one side of the phenyl moiety is close to a green (steric favorable) region.

These compounds were obtained via a previously reported an oxidative thiolation by Gangjee et al.¹² of the known compound **84**⁵⁷ (Scheme 1) with appropriately substituted thiols **85–89** in the presence of iodine as shown in Scheme 1.¹²

The thiol **89** was the only one that was not commercially available and was synthesized from the corresponding aniline.⁵⁸ Compounds **90–94** were evaluated as inhibitors of tgDHFR and rDHFR, and the results (IC₅₀) are reported in Table 4.

All four compounds designed on the basis of the CoMFA models (**91–94**) were sub-micromolar inhibitors of tgDHFR and had selectivity ratios of more than 10-fold against tgDHFR as compared to the mammalian standard rDHFR. Compound **91** was remarkable in terms of both its IC₅₀ value and selectivity ratio (5.11 nM and 75.9-fold, respectively) and is the first nanomolar inhibitor of tgDHFR with a 6–5 fused ring system. In contrast, the analog **90** was a micromolar inhibitor (IC₅₀ = 1.6 μM) of tgDHFR with a selectivity ratio of only fivefold.

The promising result for compound **91** against isolated tgDHFR prompted us to further investigate this compound in cell culture and in vivo. However, **91** had poor solubility properties. To improve the solubility characteristics, **91** was converted to its more soluble hydrochloride salt **95** (Scheme 1) by carefully adding 1 N hydrochloric acid to a suspension of **91** in methanol under sonication until all the compound went into solution. Filtration and evaporation of the filtrate to dryness afforded the HCl salt as a white powder.

Compounds **91** and **95** were initially evaluated in vitro as inhibitors of the growth of *T. gondii* cells in culture along with the standard compound pyrimethamine and the IC₅₀ values are tabulated in Table 5. These results suggest that the free base **91** and the HCl salt **95** do not differ in the inhibition of *T. gondii* cells in culture, and both are potent inhibitors of *T. gondii* cell growth in vitro.

Cytotoxicity of compound **91** was then evaluated against human embryonic lung fibroblast. Compound **91** inhibited these fibroblast cells only 9% at 260 μM indicating that compound **91** is relatively nontoxic against human fibroblast cells compared to *T. gondii* cells in culture.

Table 4

Inhibition concentrations (IC₅₀, in μM) dihydrofolate reductases from *T. gondii* and rat liver and selectivity ratios^a

| Compd | tgDHFR | rDHFR | rl/tg |
|-----------|---------|-------------------------|-------|
| 90 | 1.6 | 8.1 | 5.1 |
| 91 | 0.00511 | 0.388 | 75.9 |
| 92 | 0.572 | 7.6 | 13.3 |
| 93 | 0.212 | 3 | 14.2 |
| 94 | 0.686 | 24.6 (12%) ^b | ND |

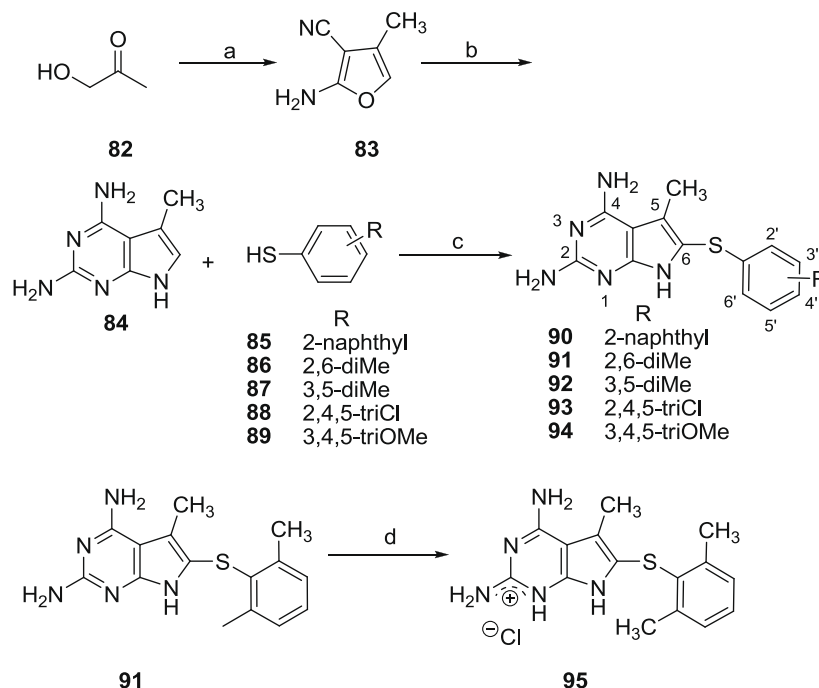
^a These assays were carried out at 37 °C under conditions of substrate (90 μM dihydrofolic acid) and cofactor (119 μM NADPH) in the presence of 150 mM KCl.

^b Numbers in parentheses are percent inhibition at the given concentration.

On the basis of the results in isolated tgDHFR and the cell culture inhibitory data compound **91** was evaluated in a *T. gondii* infection mouse model. The HCl salt **95** was also evaluated in a side-by-side evaluation with **91** and the data are presented in Table 6.

Compound **91** has been studied in two acute infection mouse models of RH strain of *T. gondii*. The intraperitoneal route was chosen, based on past experience showing activity by this route with other antifolates and on the knowledge that **91** lacked adequate solubility to administer appropriate doses intravenously. The dose of 50 mg/kg/day reflects the amount of compound injected but upon harvest of the mice in the first study, we observed drug deposits in the peritoneal cavity for many of the animals; therefore, the actual available dose was below 50 mg/kg/day for many mice. In the first study, folic acid was given to one treated group to offset any potential toxicity due to antifolate activity in mammalian cells; in fact, we saw no gross evidence of toxicity and folic acid had little effect on the activity of the compound. Hence, folic acid was not used in the second study.

The rationale for the second study was to determine if the HCl salt **95** was more effective than the parent compound. Control levels of infection and the results with **91** alone were comparable to the first study, in that **91** reduced the median counts per 1000×



Scheme 1. Reagents and conditions: (a) CH₂(CN)₂, CH₃OH, NEt₃, room temperature; (b) EtOH, NH₂CH=NH·HCl, NaOH, NaOCH₃, reflux; (c) EtOH–H₂O 2:1, I₂, reflux; (d) hydrochloric acid, MeOH, sonication.

Table 5IC₅₀ of cell growth of **91**, **95** and pyrimethamine in culture from *T. gondii*^a

| Compd | IC ₅₀ (μM) |
|---------------|-----------------------|
| 91 | 1.92 |
| 95 | 2.15 |
| Pyrimethamine | 0.7 |

^a *T. gondii* cell culture inhibition was assessed by measuring the incorporation of [³H]uracil by *T. gondii* cells.

field by more than twofold; the drop was statistically significant ($P < 0.035$). The group treated with **95** showed even more effective lowering of the counts of surviving organisms; the difference in counts for this group was statistically different from both the control and **91** alone, with P values of < 0.0001 . Furthermore, there was 100% survival of the mice for **91** and **95** compared to 25% for control on the 5th day. The mice were harvested on the 6th day. Clindamycin was used as the positive control and also provided a 100% survival of mice.

In summary, CoMFA analysis models that correlate the 3D chemical structures of 80 compounds with 6–5 fused ring system synthesized in our laboratory and their inhibitory potencies for tgDHFR and rIDHFR were developed. In addition to conventional CoMFA analysis, two routines available in the literature aimed at the optimization of CoMFA: all-orientation search (AOS) and cross-validated r^2 -guided region selection (q^2 -GRS) were used to further optimize the models. During this process, a potential problem associated with q^2 -GRS routine was identified and corrected by modifications using two strategies. Thus for the inhibitory activity against each enzyme (tgDHFR and rIDHFR), five CoMFA models were developed using the conventional CoMFA, AOS optimized CoMFA, the original q^2 -GRS optimized CoMFA and the modified q^2 -GRS optimized CoMFA using the first and the second strategy. This work demonstrated that the modified q^2 -GRS routines are superior to the original routine.

On the basis of the steric contour maps, four new compounds (**91–94**) were designed, synthesized and biologically evaluated. All of the compounds had IC₅₀ values that were at least sub-micromolar against tgDHFR inhibitors with 10-fold or better selectivity ratios as compared to rIDHFR. In particular, compound **91** was identified as the first nanomolar 6–5 fused ring system inhibitor of tgDHFR.

Compound **91** and its HCl salt **95** were potent inhibitors of *T. gondii* cell growth in culture with similar activities. On the basis of the cell culture inhibitory data, both compounds were evaluated in a *T. gondii* infection mouse model. The in vivo results confirmed that compound **91** has demonstrable activity against *T. gondii* in an acute in vivo model, and the HCl salt **95** is significantly more active than the parent compound **91**, a result likely due to increased solubility and better uptake from the peritoneal cavity.

Table 6Statistical analysis of two studies in *T. gondii* mouse model

| Group | Route | Dose (mg/kg/day) | Median | Mean | SEM | P value ^a |
|----------------|-------|------------------|--------|------|-----|------------------------|
| Control | N/A | N/A | 14.0 | 17.9 | 3.3 | N/A |
| 91 | ip | 50 | 7.0 | 12.5 | 2.5 | 0.055 |
| 91 + FA | ip | 50 | 6.0 | 9.8 | 2.0 | 0.019 |
| Control | N/A | N/A | 13.5 | 12.6 | 2.0 | N/A |
| 91 | ip | 50 | 5.0 | 8.9 | 1.5 | 0.035 |
| 95 | ip | 50 | 1.0 | 2.4 | 0.4 | < 0.0001 |

FA = folic acid, which was administered in drinking water at 10 mg/mL.

The data analyzed are counts of *T. gondii* per 1000x field in total harvested peritoneal, the 10 microliters sampled onto a 1 cm² area of a microscope slide for analysis.

^a Relative to the control within each experiment. The non-parametric Mann-Whitney test was used to compare two groups and the non-parametric Kruskal-Wallis ANOVA was used to compare several groups within one experiment.

5. Experimental section

All evaporations were carried out in vacuo with a rotary evaporator. Analytical samples were dried in vacuo (0.2 mmHg) in an Abderhalden drying apparatus over P₂O₅ at 70 °C. Thin-layer chromatography (TLC) was performed on silica gel plates with fluorescent indicator. Spots were visualized by UV light (254 and 365 nm). All analytical samples were homogeneous on TLC in at least two different solvent systems. Purification by column and flash chromatography was carried out using Merck Silica Gel 60 (200–400 mesh). The amount (weight) of silica gel for column chromatography was in the range of 50–100 times the amount (weight) of the crude compounds being separated. Columns were dry-packed unless specified otherwise. Solvent systems are reported as volume percent of mixture. Melting points were determined on a Mel-Temp II melting point apparatus and are uncorrected. Proton nuclear magnetic resonance (¹H NMR) spectra were recorded on a Bruker WH-300 (300 MHz) spectrometer. The chemical shift (δ) values are reported as parts per million (ppm) relative to tetramethylsilane as internal standard; s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br s = broad singlet. Elemental analyses were performed by Atlantic Microlab, Inc., Norcross, GA. Elemental compositions were within $\pm 0.4\%$ of the calculated values. Fractional moles of water or organic solvents frequently found in some analytical samples of antifolates could not be removed despite 24 h of drying in vacuo and were confirmed, where possible, by their presence in the ¹H NMR spectrum. All solvents and chemicals were purchased from Aldrich Chemical Co. and Fisher Scientific and were used as received.

5.1. 2,4-Diamino-5-methyl-6-(2-naphthylthio)pyrrolo[2,3-d]pyrimidine (**90**)

To a solution of **84** (0.10 g, 0.61 mmol) in a mixture of ethanol/water (2:1, 50 mL) was added 2-naphthylthiol **85** (0.20 g, 1.22 mmol) and the reaction mixture was heated to reflux for 4 h. The mixture was cooled to room temperature and concentrated under reduced pressure. After evaporation of the solvent under reduced pressure, the residue was washed with ethyl acetate followed by the addition of 200 mL of methanol and the pH adjusted to 10 with concentrated NH₄OH. The suspension was left at room temperature for 30 min and filtered. The residue was washed well with methanol and air dried to give compound **90** (0.077 g, 39%); mp 276.2–282.2 °C (decomp.); TLC R_f 0.50 (CHCl₃/MeOH, 5:1, silica gel); ¹H NMR (Me₂SO-*d*₆) δ 2.35 (s, 3H, 5-CH₃), 5.68 (s, 2H, 2-NH₂), 6.33 (s, 2H, 4-NH₂), 7.15–7.84 (m, 7H, C₁₀H₇), 11.06 (s, 1H, 7-H). HRMS (EI) m/e calcd for C₁₇H₁₅N₅S 321.104817, found (M^+) 321.105293.

5.2. 2,4-Diamino-5-methyl-6-(2',6'-dimethylphenylthio)pyrrolo[2,3-d]pyrimidine (**91**)

Compound **91** (0.94 g, 51%) was obtained from **84** (1.0 g, 6.1 mmol), 2,6-dimethyl-benzenethiol **86** (1.7 g, 12.3 mmol), and iodine (3.1 g, 12.3 mmol): mp 286.8–288.4 °C (decomp.); TLC R_f 0.70 (CHCl₃/MeOH, 5:1, silica gel); ¹H NMR (Me₂SO-*d*₆) δ 2.20 (s, 3H, 5-CH₃), 2.33 (s, 6H, 2',6'-diMe), 5.44 (s, 2H, 2-NH₂), 6.07 (s, 2H, 4-NH₂) 7.07 (s, 3H, aromatic), 10.62 (s, 1H, 7-H). Anal. (C₁₅H₁₇N₅S·0.20H₂O) C, H, N, S.

5.3. 2,4-Diamino-5-methyl-6-(3',5'-dimethylphenylthio)pyrrolo[2,3-d]pyrimidine (**92**)

Compound **92** (0.065 g, 7.0%) was obtained from **84** (0.50 g, 3.06 mmol), 3',5'-dimethyl-benzenethiol **87** (0.85 g, 6.13 mmol)

and iodine (1.56 g, 6.13 mmol): mp 285.5–288.9 °C (decomp.); TLC R_f 0.67 ($\text{CHCl}_3/\text{MeOH}$, 5:1, silica gel); ^1H NMR ($\text{Me}_2\text{SO}-d_6$) δ 2.17 (s, 6H, 3,5-(CH_3)₂), 2.32 (s, 3H, 5- CH_3), 5.63 (s, 2H, 2- NH_2), 6.29 (s, 2H, 4- NH_2), 6.62 (s, 2H, 2',6'- H_2), 6.75 (s, 2H, 4'-H), 10.93 (s, 1H, 7-H). Anal. ($\text{C}_{15}\text{H}_{17}\text{N}_5\text{S}$) C, H, N, S.

5.4. 2,4-Diamino-5-methyl-6-(2',4',5'-trichlorophenylthio)pyrrolo[2,3-d]pyrimidine (93)

Compound **93** (0.036 g, 8.4%) was obtained from **84** (0.20 g, 1.23 mmol), 2',4',5'-trichloro-benzenethiol **88** (0.52 g, 2.45 mmol) and iodine (0.62 g, 2.45 mmol): mp >300 °C (decomp.); TLC R_f 0.65 ($\text{CHCl}_3/\text{MeOH}$, 5:1, silica gel); ^1H NMR ($\text{Me}_2\text{SO}-d_6$) δ 2.30 (s, 3H, 5- CH_3), 5.72 (s, 2H, 2- NH_2), 6.36 (s, 2H, 4- NH_2), 6.53 (d, 1H, aromatic), 7.90 (s, 1H, aromatic), 11.10 (s, 1H, 7-H). Anal. ($\text{C}_{13}\text{H}_{10}\text{Cl}_3\text{N}_5\text{S}$) C, H, N, Cl, S.

5.5. 2,4-Diamino-5-methyl-6-(3',4',5'-trimethoxyphenylthio)pyrrolo[2,3-d]pyrimidine (94)

Compound **94** (0.023 g, 5.4%) was obtained from **84** (0.20 g, 1.23 mmol), 3',4',5'-trimethoxy-benzenethiol **89** (0.49 g, 2.45 mmol) and iodine (0.62 g, 2.45 mmol). 230.5–233.1 °C (decomp.); TLC R_f 0.55 ($\text{CHCl}_3/\text{MeOH}$, 5:1, silica gel); ^1H NMR ($\text{Me}_2\text{SO}-d_6$) δ 2.34 (s, 3H, 5- CH_3), 3.59 (s, 3H, 4'- OCH_3), 3.66 (s, 6H, 3',5'-(OCH_3)₂), 5.60 (s, 2H, 2- NH_2), 6.25 (s, 2H, 4- NH_2), 6.38 (s, 2H, 2',6'- H_2), 7.04 (d, 2H, 2', 6'-2H), 10.94 (s, 1H, 7-H). HRMS (EI) m/e calcd for $\text{C}_{16}\text{H}_{19}\text{N}_5\text{O}_3\text{S}$ 361.120862, found (M^+) 361.120657.

5.6. 2,4-Diamino-5-methyl-6-(2',6'-dimethylphenylthio)pyrrolo[2,3-d]pyrimidine hydrochloride salt (95)

Compound **91** (0.8 g) suspended in 100 mL of methanol was submerged in water inside of a sonicator. 1 N Hydrochloric acid was added drop-wise to the suspension until a clear solution was obtained, which was filtered and the filtrate evaporated to dryness under reduced pressure to afford the salt as a white powder: ^1H NMR ($\text{Me}_2\text{SO}-d_6$) δ 2.26 (s, 3H, 5- CH_3), 2.34 (s, 6H, 2',6'-diMe), 7.12 (s, 3H, aromatic), 7.28 (s, 2H, 2- NH_2), 8.02 (s, 2H, 4- NH_2), 11.76 (s, 1H, 7-H), 12.01 (s, 1H, 1-H). Anal. ($\text{C}_{15}\text{H}_{17}\text{N}_5\text{S} \cdot 1.0\text{HCl} \cdot 0.50\text{H}_2\text{O}$) C, H, N, S, Cl.

5.7. Dihydrofolate reductase (DHFR) assay

The spectrophotometric assay for DHFR was modified to optimize for temperature, substrate concentration, and cofactor concentration for each enzyme form assayed. The standard assay contained sodium phosphate buffer pH 7.4 (40.7 mM), 2-mercaptoethanol (8.9 mM), NADPH (0.117 mM), 1–3.7 IU of enzyme activity (1 IU = 0.005 OD units/min), and dihydrofolic acid (0.092 mM). KCl (150 mM) was included in the assay for *T. gondii* and rat liver DHFR, because it stimulated the enzymes 1.4- and 2.63-fold, respectively. The first three reagents were combined in a disposable cuvette and brought to 37 °C. Drug dilutions were added at this stage. The enzyme was added 30 s before the reaction was initiated with dihydrofolic acid. The reaction was followed for 5 min with continuous recording. Activity under these conditions of assay was linear with enzyme concentration over at least a fourfold range. Background activity measured with no added dihydrofolic acid was zero with the enzyme obtained from cultured *T. gondii* and near zero for other forms of DHFR. All DHFR inhibitors were tested against rat liver DHFR as well as against pathogen DHFR to allow assessment of selectivity.

5.8. Determination of IC_{50} values

DHFR was assayed without inhibitor and with a series of concentrations of inhibitors to produce 10–90% inhibition. At least three concentrations were required for calculation. Semilogarithmic plots of the data yielded normal sigmoidal curves for most inhibitors. The 50% inhibitory concentration (IC_{50}) was calculated from these curves using Prism 3.0.

5.9. Source of *T. gondii*

A clinical isolate of *T. gondii* was obtained from the Department of Pathology, Indiana University School of Medicine, after a single passage in a female BALB/c mouse (Harlan Industries, Indianapolis). The organisms were passaged in mice twice more, increasing the number of mice at each passage. After the final passage, the peritoneal exudate was pooled and centrifuged, and the organisms were resuspended in RPMI medium containing 10% fetal calf serum. Frozen stocks were prepared by adding 5% DMSO to the medium and freezing slowly over 8–15 h. Stocks were stored in liquid nitrogen.

5.10. Culture of *T. gondii* for enzyme production

By using a Chinese hamster ovary cell line that lacks DHFR (American Type Culture Collection, 3952 CL, CHO/dhfr-), *T. gondii* cells were grown and maintained in Iscove's modified Eagle's medium with 10% fetal calf serum, 1% penicillin/streptomycin, 100 mM hypoxanthine, and 10 mM thymidine. To each 75 cm² tissue culture flask containing the monolayer of cells was added an inoculum of approximately 10^7 organisms. Organisms (4×10^8) were harvested from each flask, within 6–8 days.

5.11. Preparation of enzymes from *T. gondii*

T. gondii organisms are minimally contaminated with mammalian host cells when harvested from tissue culture, without detectable mammalian DHFR activity.⁵⁸ When prepared as noted above, DHFR from cultured *T. gondii* has been shown to yield IC_{50} values similar to those reported in the literature.⁵⁸ The kinetics for cofactor and substrate are also similar to reported values in the literature.⁵⁸

5.12. Uracil incorporation by cultured *T. gondii*

For *T. gondii* grown in culture, uracil is incorporated into nucleic acid, but mammalian cells do not. Thus incorporation of uracil is used as an index of growth of *T. gondii* on host cells.⁵⁸ *T. gondii* is grown on HEL (human embryonic lung) cells with Minimum Essential Medium (MEM) supplemented with glutamine (2 mM), penicillin/streptomycin (100 units/mL and 100 $\mu\text{g}/\text{mL}$, respectively), and fetal bovine serum (10%). The experiment was carried out as described previously.⁵⁸

5.13. In vivo testing of drugs against *T. gondii*

Female BALB/c mice (18–20 g) were injected intraperitoneally with 5×10^3 trophozoites of *T. gondii* from culture. Drug treatment started immediately if the drugs were given in drinking water of food; if the drugs were injected, treatment started 4 h after inoculation. Survival or counts of *T. gondii* present in peritoneal exudates or liver were monitored as an index of drug efficacy.

Acknowledgments

This work was supported in part by the National Institutes of Health, National Institute of Allergy and Infectious Diseases AI047759 (A.G.) and NCI CA98850 (A.G.).

Supplementary data

Supplementary data (Elemental analysis and Appendix Tables 1–30.) associated with this article can be found, in the online version, at [doi:10.1016/j.bmc.2009.12.066](https://doi.org/10.1016/j.bmc.2009.12.066).

References and notes

- Presented in part at the 231st ACS National Meeting, Atlanta, GA, March 26–30, 2006; MEDI-349.
- Klepser, M. E.; Klepser, T. B. *Drugs* **1997**, *53*, 40.
- Gangjee, A.; Devraj, R.; McGuire, J. J.; Kisliuk, R. L.; Queener, S. F.; Barrows, L. R. *J. Med. Chem.* **1994**, *34*, 1169.
- Gangjee, A.; Mavandadi, F.; Queener, S. F.; McGuire, J. J. *J. Med. Chem.* **1995**, *38*, 2158.
- Gangjee, A.; Mavandadi, F.; Kisliuk, R. L.; McGuire, J. J.; Queener, S. F. *J. Med. Chem.* **1996**, *39*, 4563.
- Gangjee, A.; Mavandadi, F.; Queener, S. F. *J. Med. Chem.* **1997**, *40*, 1173.
- Gangjee, A.; Vasudevan, A.; Queener, S. F. *J. Med. Chem.* **1997**, *40*, 3032.
- Gangjee, A.; Guo, X.; Queener, S. F.; Cody, V.; Galitsky, N.; Luft, J. R.; Pangborn, W. *J. Med. Chem.* **1998**, *41*, 1263.
- Gangjee, A.; Dubash, N. P.; Queener, S. F. *J. Heterocycl. Chem.* **2000**, *37*, 935.
- Gangjee, A.; Vidwans, A.; Elzein, E.; McGuire, J. J.; Queener, S. F.; Kisliuk, R. L. *J. Med. Chem.* **2001**, *44*, 1993.
- Gangjee, A.; Yu, J.; Kisliuk, R. L. *J. Heterocycl. Chem.* **2002**, *39*, 833.
- Gangjee, A.; Lin, X.; Queener, S. F. *J. Med. Chem.* **2004**, *47*, 3689.
- Gangjee, A.; Zeng, Y.; McGuire, J. J.; Mehraein, F.; Kisliuk, R. L. *J. Med. Chem.* **2004**, *47*, 6893.
- Gangjee, A.; Jain, H. D.; Queener, S. F. *J. Heterocycl. Chem.* **2005**, *42*, 589.
- Gangjee, A.; Ye, Z.; Queener, S. F. *J. Heterocycl. Chem.* **2005**, *42*, 1127.
- Gangjee, A.; Zeng, Y.; Ihnat, M.; Warnke, L. A.; Green, D. W.; Kisliuk, R. L.; Lin, F.-T. *Bioorg. Med. Chem.* **2005**, *13*, 5475.
- Ivanciuc, O.; Ivanciuc, T.; Cabrol-Bass, D. *Theochem. J. Mol. Struct.* **2002**, *582*, 39.
- Burbidge, R.; Trotter, M.; Buxton, B.; Holden, S. *Comput. Chem.* **2001**, *26*, 5.
- Garg, S.; Achenie, L. E. K. *Biotechnol. Prog.* **2001**, *17*, 412.
- Zheng, W.; Tropsha, A. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 185.
- Selassie, C. D.; Gan, W.-X.; Kallander, L. S.; Klein, T. E. *J. Med. Chem.* **1998**, *41*, 4261.
- Burden, F. R.; Rosewarne, B. S.; Winkler, D. A. *Chemom. Intell. Lab. Syst.* **1997**, *38*, 127.
- Ivanciuc, O. *Rev. Roum. Chim.* **1996**, *41*, 645.
- Marlowe, C. K.; Selassie, C. K.; Santi, D. V. *J. Med. Chem.* **1995**, *38*, 967.
- Hirst, J. D.; King, R. D.; Sternberg, M. J. E. *J. Comput. Aided Mol. Des.* **1994**, *8*, 421.
- Hirst, J. D.; King, R. D.; Sternberg, M. J. E. *J. Comput. Aided Mol. Des.* **1994**, *8*, 405.
- Stanton, D. T.; Murray, W. J.; Jurs, P. C. *Quant. Struct. Act. Relat.* **1993**, *12*, 239.
- So, S.-S.; Richards, W. G. *J. Med. Chem.* **1992**, *35*, 3201.
- Selassie, C. D.; Li, R.-L.; Poe, M.; Hansch, C. *J. Med. Chem.* **1991**, *34*, 46.
- Debnath, A. K.; Lopez de Compadre, R. L.; Debnath, G.; Shusterman, A. J.; Hansch, C. *J. Med. Chem.* **1991**, *34*, 786.
- Andrea, T. A.; Kalayeh, H. *J. Med. Chem.* **1991**, *34*, 2824.
- Selassie, C. D.; Fang, Z.-X.; Li, R.-L.; Hansch, C.; Debnath, G.; Klein, T. E.; Langridge, R.; Kaufman, B. T. *J. Med. Chem.* **1989**, *32*, 1895.
- Li, R.-L.; Poe, M. *J. Med. Chem.* **1988**, *31*, 366.
- Booth, R. G.; Selassie, C. D.; Hansch, C.; Santi, D. V. *J. Med. Chem.* **1987**, *30*, 1218.
- Selassie, C. D.; Fang, Z.-X.; Li, R.-L.; Hansch, C.; Klein, T. E.; Langridge, R.; Kaufman, B. T. *J. Med. Chem.* **1986**, *29*, 621.
- Ghose, A. K.; Crippen, G. M. *J. Med. Chem.* **1985**, *28*, 333.
- Hansch, C.; Hathaway, B. A.; Guo, Z.-R.; Selassie, C. D.; Dietrich, S. W.; Blaney, J. M.; Langridge, R.; Volz, K. W.; Kaufman, B. T. *J. Med. Chem.* **1984**, *27*, 129.
- Hathaway, B. A.; Guo, Z.-R.; Hansch, C.; Delcamp, T. J.; Susten, S. S.; Freisheim, J. H. *J. Med. Chem.* **1984**, *27*, 144.
- Hopfinger, A. J. *J. Med. Chem.* **1983**, *26*, 990.
- Khwaja, T. A.; Pentecost, S.; Selassie, C. D.; Guo, Z.-R.; Hansch, C. *J. Med. Chem.* **1982**, *25*, 153.
- Hansch, C.; Li, R.-L.; Blaney, J. M.; Langridge, R. *J. Med. Chem.* **1982**, *25*, 777.
- Ghose, A. K.; Crippen, G. M. *J. Med. Chem.* **1982**, *25*, 892.
- Li, R.-L.; Dietrich, S. W.; Hansch, C. *J. Med. Chem.* **1981**, *24*, 538.
- Coats, E. A.; Genther, C. S.; Dietrich, S. W.; Guo, Z.-R.; Hansch, C. *J. Med. Chem.* **1981**, *24*, 1422.
- Crippen, G. M. *J. Med. Chem.* **1980**, *23*, 599.
- Dietrich, S. W.; Blaney, J. M.; Reynolds, M. A.; Jow, P. Y. C.; Hansch, C. *J. Med. Chem.* **1980**, *23*, 1205.
- Blaney, J. M.; Hansch, C.; Silipo, C.; Vittoria, A. *Chem. Rev.* **1984**, *84*, 333.
- Mattioni, B. E.; Jurs, P. C. *J. Mol. Graph. Modell.* **2003**, *21*, 391.
- Sutherland, J. J.; Weaver, D. F. *J. Comput. Aided Mol. Des.* **2004**, *18*, 309.
- Gangjee, A.; Lin, X. *J. Med. Chem.* **2005**, *48*, 1448.
- Wang, R.; Gao, Y.; Liu, L.; Lai, L. *J. Mol. Model.* **1998**, *4*, 276.
- Cho, S. J.; Tropsha, A. *J. Med. Chem.* **1995**, *38*, 1060.
- Cody, V.; Galitsky, N.; Luft, J. R.; Pangborn, W.; Gangjee, A.; Devraj, R.; Queener, S. F.; Blakley, R. L. *Acta Crystallogr., Sect. D* **1997**, *53*, 638.
- Molecular Operating Environment (MOE 2004.03), C. C. G., Inc, 1255 University St., Suite 1600, Montreal, Quebec, Canada, H3B 3X3.
- SYBYL Version 7.0; Tripos Associates: St. Louis, M.
- Godden, J. W.; Xue, L.; Bajorath, J. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 163.
- Taylor, E. C.; Patel, H. H.; Jun, J.-G. *J. Org. Chem.* **1995**, *60*, 6684.
- Offer, J.; Boddy, C. N. C.; Dawson, P. E. *J. Am. Chem. Soc.* **2002**, *124*, 4642.