J Comput Aided Mol Des (2009) 23:241–252
DOI 10.1007/s10822-008-9254-z

Interactions between cycloguanil derivatives and wild type and resistance-associated mutant *Plasmodium falciparum* dihydrofolate reductases

Phornphimon Maitarad · Sumalee Kamchonwongpaisan · Jarunee Vanichtanankul · Tirayut Vilaivan · Yongyuth Yuthavong · Supa Hannongbua

Received: 20 June 2008 / Accepted: 20 November 2008 / Published online: 21 January 2009

Abstract Comparative molecular field analysis (CoMFA) and quantum chemical calculations were performed on cycloguanil (Cyc) derivatives of the wild type and the quadruple mutant (Asn51Ile, Cys59Arg, Ser108Asn, Ile164-Leu) of *Plasmodium falciparum* dihydrofolate reductase (*Pf*DHFR). The represented CoMFA models of wild type ($r^2_{cv} = 0.727$ and $r^2 = 0.985$) and mutant type ($r^2_{cv} = 0.786$ and $r^2 = 0.979$) can describe the differences of the Cyc structural requirements for the two types of *Pf*DHFR enzymes and can be useful to guide the design of new inhibitors. Moreover, the obtained particular interaction energies between the Cyc and the surrounding residues in the binding pocket indicated that Asn108 of mutant enzyme was the cause of Cyc resistance by producing steric clash with *p*-Cl of Cyc. Consequently, comparing the energy contributions with the potent flexible WR99210 inhibitor, it was found that the key mutant residue, Asn108, demonstrates attractive interaction with this inhibitor and some residues, Leu46, Ile112, Pro113, Phe116, and Leu119, seem to perform as second binding site with WR99210. Therefore, quantum chemical calculations can be useful for investigating residue interactions to clarify the cause of drug resistance.

Keywords Dihydrofolate reductase · Cycloguanil · WR99210 · CoMFA · Quantum chemical calculations

P. Maitarad · S. Hannongbua (✉)
Department of Chemistry, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand
e-mail: fscisph@ku.ac.th

P. Maitarad · S. Hannongbua
Center of Nanotechnology, Kasetsart University, Bangkok 10900, Thailand

P. Maitarad · S. Kamchonwongpaisan · J. Vanichtanankul · Y. Yuthavong
National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Pathumthani 12120, Thailand

T. Vilaivan
Department of Chemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand

Introduction

Dihydrofolate reductase (DHFR) of *Plasmodium falciparum* (*Pf*) is one of the targets in malarial chemotherapy. DHFR is the enzyme that catalyzes the nicotinamide adenine dinucleotide phosphate (NADPH)-dependent reduction of 7,8-dihydrofolate (DHF) to 5,6,7,8-tetrahydrofolate (THF) [1–4]. X-ray crystal structures of *Pf*DHFR-TS (thymidylate synthase) are available for the bifunctional and dimeric enzyme, consisting of *Pf*DHFR in the form of wild type (PDB code 1J3I), double mutant type (PDB code 1J3J), and quadruple mutant type (PDB code 1J3K) [5]. Antifolate drugs are used for inhibiting the reduction of DHF to THF, the former of which is a product from TS. Cycloguanil (Cyc) is one of the potent antifolate drugs and is the metabolic product of proguanil, as shown in Fig. 1. This drug has some structural similarities to the substrate DHF and acts as a competitive inhibitor [6–8].

Unfortunately, the resistance of *Pf*DHFR to Cyc has rapidly reduced its clinical utility [9–12]. It has been reported that antifolate drug resistance is consistent with point mutations in *Pf*DHFR. The reports of many previous studies suggested and also provided further evidence that

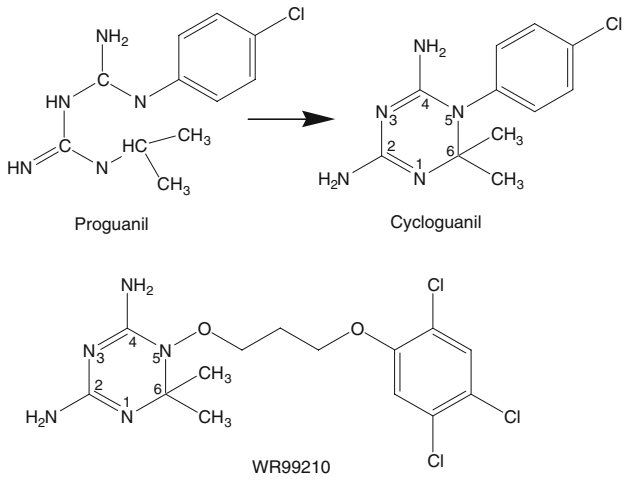

Fig. 1 Structures of dihydrotriazine antimalarial antifolates

point mutation from Ser108 to Asn108 in *Pf*DHFR is probably causing resistance to antifolates including pyrimethamine (Pyr) and Cyc in malaria. Additional point mutations in Asn51Ile, Cys59Arg, and Ile164Leu were associated with higher levels of resistance [13–17]. Another set of mutations, Ala16Val and Ser108Thr, was also known to cause resistance to Cyc specifically. In 2004, Kamchonwongpaisan and co-workers [18] reported the Pyr and the Cyc antifolate derivatives with their inhibition constants (K_i) in both wild type and multiple mutant *Pf*DHFRs. This research has led to further work to establish a three-dimensional quantitative structure–activity relationship (3D-QSAR) and to its use to develop new and effective antifolate antimalarials against *Pf*DHFRs. Very recently, we successfully investigated the 3D-QSAR of the Pyr derivatives using comparative molecular field analysis (CoMFA) technique to guide the design of new inhibitors of quadruple mutant (Asn51Ile, Cys59Arg, Ser108Asn, and Ile164Leu) *Pf*DHFR [19]. Consequently, the CoMFA technique has been chosen to find the relationship between the Cyc derivatives structures and their biological activity. The advantages of CoMFA are the ability to predict the biological activities of the molecules and to represent the relationships between steric and electrostatic properties, calculated according to Lennard-Jones and Coulomb potentials, respectively, and also the biological activity in the form of contour maps to provide the key features of both the ligand–receptor interaction and the topology of the receptor [20–24].

In addition, a comparison of the role of the amino acids surrounding the binding pocket of Cyc in the wild type and the quadruple mutant type *Pf*DHFRs was investigated by using AMBER molecular mechanic minimizations and quantum chemical calculations. The strong repulsive energy of some significant amino acids can lead to the cause of Cyc resistance. On the other hand, the strong attractive interaction should be used in the design of new compounds. Therefore, an understanding of the particular interaction energy of individual amino acids with the Cyc inhibitor at molecular level is required to support the identification of these new structural modifications. Ligand- and structure-based drug design will need to be merged to identify the best structural features for potent Cyc derivatives active against the wild type and the quadruple mutant type *Pf*DHFRs.

Computational methods

Data set preparation

A set of 25 Cyc derivatives was selected for CoMFA study. Their structures and their inhibition constants (K_i) against wild type and quadruple mutant type *Pf*DHFR [18] were as listed in Table 1. Among these, 20 compounds were used as a training set and five with sampling from different ranges of biological activity were kept as a test set for model validation. A dependent variable in CoMFA was defined as pK_i ($-\log K_i$), where K_i values were measured in vitro under the same experimental conditions. Three-dimensional structures of these compounds were built using the Sybyl 7.0 program package [25] on a Silicon Graphics Octane2 workstation. Geometry optimization of each molecule was carried out using a Tripos force field, and the Gasteiger-Hückle charges were calculated.

CoMFA analysis

An atom-base superimposition alignment of 25 compounds was performed using a common template with 14 aligned positions as marked by asterisks in the structure of the figure insert in Table 1. Three-dimensional (3D) cubic lattices with 2-Å grid spacing were generated around the aligned compounds based on the molecular volume of the structure. The lattices were defined automatically, and were extended by 4.0 Å in all directions. Three types of probe atoms were placed at each lattice point, namely sp^3 carbon atom with +1 charge, sp^3 oxygen atom with −1 charge, and H atom with +1 charge. The interactions of the steric and electrostatic fields with each atom in the molecule were calculated using CoMFA standard scaling. The default value of 30 kcal/mol was used as the maximum electrostatic and steric energy cutoff. The CoMFA fields were scaled by CoMFA-STD method in Sybyl and a partial least-squares technique (PLS) was employed to derive a CoMFA model expressing the correlation between the steric and electrostatic properties and the biological

Table 1 Structure and data set of Cyc and its derivatives used for CoMFA analysis; probe atoms are indicated by an asterisk

Comp	X	Y	R_1	R_2	K_i (wt) (nM)	pK_i (wt)	K_i (mut)[a] (nM)	pK_i (mut)
1 (C21)	H	Cl	Me	Me	1.5	8.82	420	6.37
2 (C97)	Cl	H	Me	Me	3.0	8.52	25	7.60
3 (C66)[b,c]	H	Cl	Me	*n*Pr	3.5	8.45	4,432	5.35
4 (C433)	Cl	H	Me	*i*Pr	25.5	7.59	472	6.32
5 (C22)	H	Cl	Me	*i*Pr	36.5	7.44	10,000	5.00
6 (C434)	Cl	H	Me	*n*Pr	4.6	8.34	39.4	7.40
7 (C71)	H	Cl	Me	*n*Hex	0.6	9.22	69.2	7.16
8 (C435)[c]	Cl	H	Me	*n*Hex	2.4	8.62	15.9	7.79
9 (C17)	H	Cl	H	Me	4.1	8.38	2,427	5.61
10 (C248)[b]	Cl	H	H	Me	10.2	7.99	14.3	7.84
11 (C53)	H	Cl	H	C_6H_5	4.5	8.34	2,436	5.61
12 (C96)	Cl	H	H	C_6H_5	11.7	7.93	984	6.01
13 (C121)	H	Cl	H	4-$C_6H_5OC_6H_4$	0.4	9.39	222	6.65
14 (C138)	Cl	H	H	4-$C_6H_5OC_6H_4$	0.7	9.15	1.3	8.88
15 (C133)[b,c]	H	Cl	H	3-$C_6H_5OC_6H_4$	0.5	9.30	239	6.62
16 (C110)	Cl	H	H	3-$C_6H_5OC_6H_4$	1.1	8.96	5.8	8.23
17 (C111)	H	Cl	H	3-$C_6H_5CH_2OC_6H_4$	0.7	9.15	638	6.19
18 (C185)[c]	Cl	H	H	3-$C_6H_5CH_2OC_6H_4$	2.3	8.64	7.7	8.11
19 (C188)[b]	H	Cl	H	3-(4-ClC_6H_4O)C_6H_4	1.4	8.85	121	6.92
20 (C143)	Cl	H	H	3-(4-ClC_6H_4O)C_6H_4	1.3	8.88	8.0	8.09
21 (C372)	Cl	H	H	*n*C_7H_{15}	2.7	8.56	0.8	9.09
22 (C229)[b]	Cl	H	H	4-$PrOC_6H_4$	3.3	8.48	2.7	8.57
23 (C186)[c]	Cl	H	H	3-(3,5-$Cl_2C_6H_3O$)C_6H_4	1.8	8.74	2.7	8.57
24 (C313)	Cl	H	H	3-[2,4,5-$Cl_3C_6H_2O(CH_2)_3O$]C_6H_4	4.0	8.40	4.5	8.34
25 (C299)	Cl	H	H	3-(3-$CF_3C_6H_4O$)C_6H_4	2.7	8.57	4.8	8.32

[a] Quadruple mutation (Asn51Ile, Cys59Arg, Ser108Asn, Ile164Leu) of *Pf*DHFR

[b] Test set compounds for wild type *Pf*DHFR CoMFA model

[c] Test set compounds for quadruple mutant type *Pf*DHFR CoMFA model

activities. The orthogonal latent variables were extracted by the nonlinear iterative partial least squares (NIPALS) algorithm and subjected to full cross-validation with the leave-one-out (LOO) method. In order to speed up the analysis and reduce the amount of noise, the minimum sigma value was set to 2.0 kcal/mol as a default. The analyses were first carried out with a maximum of six components, then using the number of component (noc) at which the difference in the r^2_{cv} value compared with the next one was less than 0.05.

The predictive ability of the model derived from the training set is expressed as the cross-validation predictive (r^2_{cv}) value. The r^2_{cv} value is defined as

$$r^2_{cv} = 1.0 - \frac{\text{PRESS}}{\text{SSY}}, \quad (1)$$

where SSY is the variance of the biological activities around the mean value, and PRESS is the prediction error sum of squares derived from the LOO.

$$\text{PRESS} = \sum\nolimits_y \left(y_{\text{pred}} - y_{\text{actual}}\right)^2 \quad (2)$$

$$\text{SSY} = \sum\nolimits_y \left(y_{\text{actual}} - y_{\text{mean}}\right)^2. \quad (3)$$

The uncertainty of the prediction is defined as

$$S_{\text{PRESS}} = \sqrt{\frac{\text{PRESS}}{n-k-1}}, \quad (4)$$

where k is the number of variables in the model and n is the number of compounds used in the study [26–28].

AMBER molecular mechanic minimizations

The models of wild type and the quadruple mutant *Pf*DHFR complex with the Cyc inhibitor were constructed from the X-ray complex structures of WR99210/wild type *Pf*DHFR (1J3I.pdb) and WR99210/quadruple mutant type *Pf*DHFR

(1J3K.pdb), respectively. The missing residues (residues 86–95) of the structures were added using Sybyl program, then the partial minimizations were performed by using Tripos force field, implemented in the Sybyl program, with unrestrained residues 83–98, including the inserted residues. The Cyc structure was fully optimized at the B3LYP/6-31G(*) level of calculations, and then the restrained electrostatic potential (RESP) charges [29] of all atoms in the Cyc inhibitor were also calculated using the Gaussian 03 package [30]. The resultant structure and electrostatic charges were used to prepare the molecular mechanical AMBER force field parameters using the Antech AMBER module. The AMBER parameters of NADPH cofactor were taken from the AMBER parameter database [31, 32]. TIP3PBOX water molecules were used to solvate both complexes. Four chloride ions were added to neutralize the system. The minimization process of the complex structures was performed to eradicate bad contacts and to relax the models. A cutoff distance of 12 Å was set for the nonbonded pair interactions. All molecular mechanical minimizations were carried out using the AMBER 9.0 simulation package [33].

Quantum chemical calculations

The obtained molecular mechanical structures of the wild type and the quadruple mutant *Pf*DHFRs complexed with the Cyc ligand were used as the starting geometry of the particular interaction energy calculations at the molecular level for the Cyc inhibitor with respect to both target enzymes. All residues located with at least one atom interacting with any atoms of the inhibitor within the interatomic distance of 4 Å were selected to calculate the interaction energy with the inhibitor based on quantum chemical calculations. The 20 selected residues were Ile14, Cys15, Ala16, Val45, Leu46, Trp48, Cys50, Asn51Ile, Asp54, Met55, Tyr57, Phe58, Met104, Ser108Asn, Ser111, Ile112, Leu119, Ile164Leu, Gly165, and Thr185. Cys59Arg, Pro113, and Phe116 were also included in the systems to compare their particular interaction energies contributed to Cyc and WR99210. The three-dimensional (3D) scheme of 23 residues is shown in Fig. 2. In addition, the backbone amino acids of 14 inserted residues, Pro47, Lys49, Ser52, Leu53, Lys56, Gly105, Arg106, Thr107, Trp109, Glu110, Lys114, Lys115, Lys117, and Pro118, were added to the system to complete the connection between the cutting regions for peptide bonds of amino acids. The N- and C-terminals were capped with a methyl amino group (–$NHCH_3$) and an acetyl group (CH_3CO–), respectively, which were retained from the backbone geometries of the nearby residues. Geometry optimization of the hydrogen atoms was carried out based on the HF/6-31G(d,p) level of calculations, implemented in the Gaussian 03

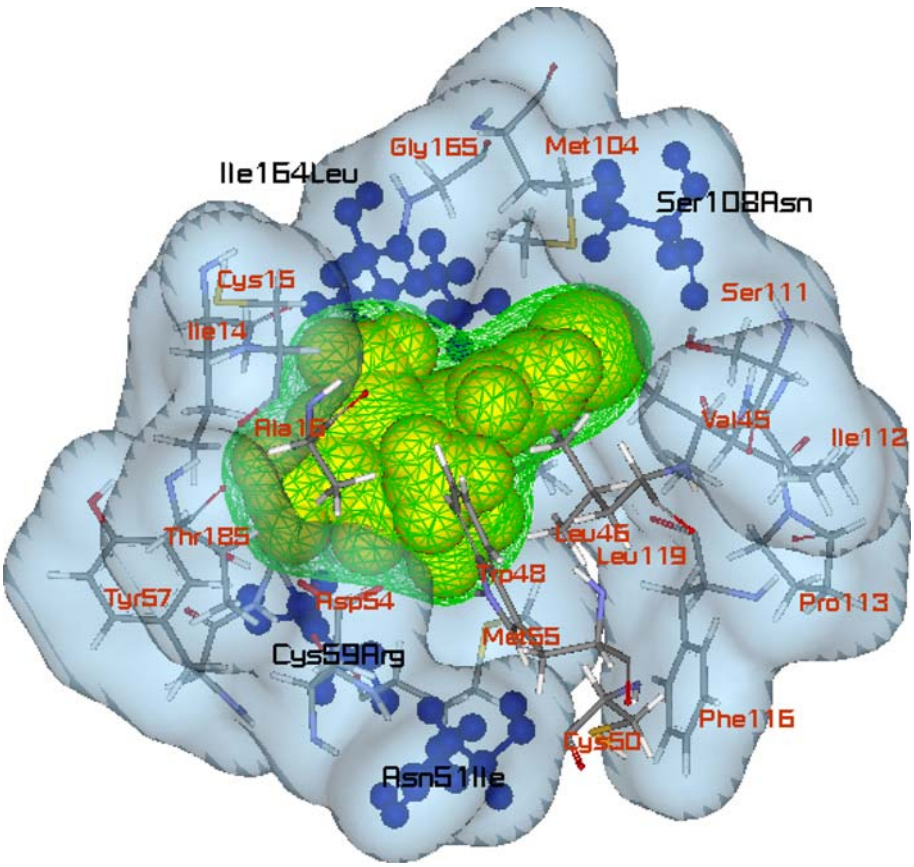

Fig. 2 Three-dimensional diagram of the adopted binding model of Cyc in *Pf*DHFR active site

program running on a Linux super cluster (four 3.2-GHz CPUs, 3.0 GB RAM). The resulting geometries were used to provide different model systems of the Cyc and the residues for MP2/6-31G(d,p) calculations, which then provided information on the particular interaction energy between the inhibitor and each residue surrounding the binding site as shown in the interaction energy formula:

Table 2 PLS statistical results of CoMFA models of Cyc derivatives against wild type and quadruple mutant type *Pf*DHFRs

Parameters	Wild type *Pf*DHFR			Quadruple mutant *Pf*DHFR			
	Model I C_{sp^3} (+1)	Model II O_{sp^3} (−1)	Model III H (+1)	Model IV C_{sp^3} (+1)	Model V O_{sp^3} (−1)	Model VI H (+1)	Model VII Combined probe atoms
r^2_{cv}	0.650	0.645	0.727	0.786	0.729	0.676	0.786
S_{PRESS}	0.365	0.368	0.322	0.682	0.713	0.837	0.650
noc	6	6	6	6	4	6	5
r^2_{nv}	0.977	0.979	0.985	0.983	0.962	0.986	0.979
s	0.094	0.090	0.075	0.192	0.267	0.171	0.202
F value	91.085	99.690	144.780	125.390	95.141	157.924	132.386
Steric (%)	0.604	0.596	0.565	0.674	0.682	0.644	0.678
Electrostatic (%)	0.396	0.404	0.435	0.326	0.318	0.356	0.322

Table 3 Actual (Act) and predicted (Pred) pK_i values and the residuals (Δ) of the training set and test set molecules for the wild type model

Comp	Act p*K*i	Model I		Model II		Model III	
		Pred pK_i	Δ	Pred pK_i	Δ	Pred pK_i	Δ
Wild type							
1	8.82(4)	8.78(6)	0.038	8.77(2)	0.052	8.78(9)	0.035
2	8.52(3)	8.60(8)	−0.085	8.57(6)	−0.053	8.55(5)	−0.032
4	7.59(3)	7.39(6)	0.197	7.39(2)	0.201	7.43(4)	0.159
5	7.43(8)	7.61(1)	−0.173	7.61(6)	−0.178	7.67(2)	−0.234
6	8.33(7)	8.21(7)	0.120	8.24(8)	0.089	8.27(3)	0.064
7	9.22(2)	9.13(2)	0.090	9.10(4)	0.118	9.11(7)	0.105
8	8.62(0)	8.68(3)	−0.063	8.75(4)	−0.134	8.82(1)	−0.201
9	8.38(7)	8.36(3)	0.024	8.38(5)	0.002	8.35(5)	0.032
11	8.34(7)	8.34(3)	0.004	8.33(3)	0.014	8.24(2)	0.105
12	7.93(2)	8.06(4)	−0.132	8.07(4)	−0.142	8.03(7)	−0.105
13	9.39(8)	9.41(9)	−0.021	9.42(4)	−0.026	9.40(5)	−0.007
14	9.15(5)	9.12(9)	0.026	9.14(0)	0.015	9.14(9)	0.006
16	8.95(9)	8.99(4)	−0.035	8.98(9)	−0.030	8.99(6)	−0.037
17	9.15(5)	9.13(0)	0.025	9.15(7)	−0.002	9.17(8)	−0.023
18	8.63(8)	8.63(2)	0.006	8.61(1)	0.027	8.61(6)	0.022
20	8.88(6)	8.87(3)	0.013	8.87(7)	0.009	8.90(0)	−0.014
21	8.56(9)	8.61(8)	−0.049	8.56(3)	0.006	8.48(1)	0.088
23	8.74(5)	8.71(9)	0.026	8.71(6)	0.029	8.73(6)	0.009
24	8.39(8)	8.42(1)	−0.023	8.41(8)	−0.020	8.40(0)	−0.002
25	8.56(9)	8.55(7)	0.012	8.54(7)	0.022	8.54(1)	0.028
3[a]	8.45(6)	8.49(9)	−0.043	8.33(9)	0.117	8.40(2)	0.054
10[a]	7.99(1)	8.31(4)	−0.323	8.38(2)	−0.391	8.37(3)	−0.382
15[a]	9.30(1)	9.23(6)	0.065	9.24(5)	0.056	9.22(5)	0.076
19[a]	8.85(4)	9.17(5)	−0.321	9.15(1)	−0.297	9.14(9)	−0.295
22[a]	8.48(1)	8.82(4)	−0.343	8.80(6)	−0.325	8.74(1)	−0.260

[a] Test set compounds

$$E^{\mathrm{INT}}_{(\mathrm{ligand-aminoacid})} = E^{AB}_{(\mathrm{ligand-aminoacid})} - E^{A}_{(\mathrm{ligand})} - E^{B}_{(\mathrm{aminoacid})}, \quad (5)$$

where A and B are the number of basis sets of ligands and amino acids, respectively, $E^{AB}_{(\mathrm{ligand-aminoacid})}$ is the energy of the ligand–amino acid complex with the basis set of A plus B. $E^{A}_{(\mathrm{ligand})}$ and $E^{B}_{(\mathrm{aminoacid})}$ are the energies of the ligand and the amino acid with its number of basis sets.

Furthermore, the basis set superposition error (BSSE) based on the counterpoise scheme of Boys–Bernardi [34] was also computed to define the interaction energy with BSSE as shown in Eq. 6:

$$E^{\mathrm{INT}}_{(\mathrm{ligand-aminoacid})} = E^{AB}_{(\mathrm{ligand-aminoacid})} - E^{AB}_{(\mathrm{ligand})} - E^{AB}_{(\mathrm{aminoacid})} \quad (6)$$

where $E^{AB}_{(\mathrm{ligand})}$ and $E^{AB}_{(\mathrm{aminoacid})}$ are the energies of the ligand and the amino acid, respectively, with the number of basis sets of A plus B [35–37].

Results and discussion

CoMFA statistical results

Statistically significant CoMFA models were derived from the Cyc derivatives based on their inhibitory activity against the wild type and the quadruple mutant *Pf*DHFRs. Each CoMFA models was varied by changing the type of probe atoms at the grid spacing as representative of mainly atom types of amino acids for in silico receptor. The C_{sp^3} (+1) is generally standed for steric interaction, O_{sp^3} (−1) for electrostatic interaction, and H (−1) for steric and electrostatic equality. The obtained statistical CoMFA models and predicted pK_i values are summarized, respectively, in Table 2 and Tables 3 and 4. The representative CoMFA models of Cyc for the two types of *Pf*DHFRs were determined using the terms of r^2_{cv} and nonvalidated r^2 (r^2_{nv}). By considering the statistical results of Cyc against the

Table 4 Actual (Act) and predicted (Pred) pK_i values and the residuals (Δ) of the training set and test set molecules for the quadruple mutant type model

Comp	Act	Model IV		Model V		Model VI		Model VII	
	pK_i	Pred pK_i	Δ	Pred pK_i	Δ	Pred pK_i	Δ	Pred pK_i	Δ
Quadruple mutant type									
1	6.37(7)	6.73(3)	−0.356	6.74(1)	−0.364	6.51(8)	−0.141	6.19(9)	0.171
2	7.60(2)	7.22(3)	0.379	7.27(4)	0.328	7.43(1)	0.171	7.48(8)	0.112
4	6.32(6)	6.48(8)	−0.162	6.38(3)	−0.057	6.32(9)	−0.003	6.38(6)	−0.066
5	5.00(0)	5.20(8)	−0.208	5.19(0)	−0.190	5.05(4)	−0.054	4.93(5)	0.065
6	7.40(5)	7.21(9)	0.186	7.19(6)	0.209	7.22(1)	0.184	7.22(5)	0.175
7	7.16(0)	6.95(7)	0.203	6.97(4)	0.186	6.98(1)	0.179	7.15(6)	0.004
9	5.61(5)	5.74(6)	−0.131	5.70(8)	−0.093	5.60(9)	0.006	5.74(8)	−0.138
10	7.84(5)	7.73(0)	0.115	7.76(6)	0.079	7.77(4)	0.071	7.69(7)	0.143
11	5.61(3)	5.03(8)	0.575	5.18(8)	0.425	5.44(0)	0.173	5.30(9)	0.301
12	6.00(7)	6.54(5)	−0.538	6.38(7)	−0.380	6.31(4)	−0.307	6.21(3)	−0.213
13	6.65(4)	6.84(1)	−0.187	6.98(4)	−0.330	6.92(2)	−0.268	6.76(3)	−0.113
14	8.88(6)	8.75(0)	0.136	8.69(4)	0.192	8.73(9)	0.147	8.86(3)	0.017
16	8.23(7)	8.19(8)	0.039	8.14(9)	0.088	8.06(9)	0.168	8.17(3)	0.057
17	6.19(5)	6.10(3)	0.092	6.03(8)	0.157	6.22(2)	−0.027	6.15(6)	0.034
19	6.91(7)	6.90(8)	0.009	6.89(8)	0.019	6.84(9)	0.068	6.99(8)	−0.088
20	8.09(7)	8.21(0)	−0.113	8.23(1)	−0.134	8.15(6)	−0.059	8.13(8)	−0.038
21	9.09(7)	9.13(3)	−0.036	9.21(5)	−0.118	9.32(8)	−0.231	9.19(8)	−0.098
22	8.56(9)	8.68(6)	−0.117	8.72(2)	−0.153	8.63(6)	−0.067	8.26(8)	0.052
24	8.34(7)	8.30(2)	0.045	8.24(6)	0.101	8.42(6)	−0.079	8.28(0)	0.060
25	8.31(9)	8.24(9)	0.070	8.28(6)	0.033	8.25(0)	0.069	8.60(7)	−0.037
3[a]	5.35(3)	5.84(5)	−0.492	5.86(9)	−0.110	5.75(9)	0.086	5.75(3)	−0.400
8[a]	7.79(9)	8.52(6)	−0.727	8.56(0)	−0.129	8.43(1)	0.095	8.44(6)	−0.647
15[a]	6.62(2)	6.89(8)	−0.276	6.74(6)	0.082	6.82(8)	0.070	6.95(2)	−0.330
18[a]	8.11(4)	7.52(4)	0.590	7.68(5)	−0.125	7.56(0)	−0.036	7.43(5)	0.679
23[a]	8.56(9)	8.34(5)	0.224	8.18(1)	0.123	8.30(4)	0.041	8.11(9)	0.450

[a] Test set compounds

wild type in Table 2, CoMFA model III shows the highest r^2_{cv} (0.727) and r^2_{nv} (0.985) values. Consequently, CoMFA model III was used for further contour map discussion. In the case of the quadruple mutant enzyme, models IV-VI, with r^2_{cv} values higher than 0.6 (0.786, 0.729, and 0.676, respectively) were acceptable and the r^2_{nv} values were 0.983, 0.962, and 0.986, respectively. These results suggest that the three tested probes (C_{sp^3}, O_{sp^3}, and H) produce qualitatively very similar models. The results also suggest that all three types of probes are equally important in the enzyme–ligand interactions. The combination of the three probe atoms was then used, resulting in model VII with $r^2_{cv} = 0.786$ and $r^2_{nv} = 0.979$. Thus, based on the CoMFA of the Cyc derivatives against the wild type *Pf*DHFR, the best model is model III using a probe atom H(+1) with $r^2_{cv} = 0.727$, $S_{PRESS} = 0.322$, noc = 6, $r^2_{nv} = 0.985$, $s = 0.075$, and F value = 144.780. The combination of the three probe atoms in model VII gave satisfactory results ($r^2_{cv} = 0.786$, $S_{PRESS} = 0.650$, noc = 5, $r^2_{nv} = 0.979$, $s = 0.202$, F value = 132.386) and therefore was the best CoMFA model of the Cyc derivatives active against the mutant enzyme. The graphical plots between the actual and the predicted pK_i of the training set and the test set of both the CoMFA models are shown in Fig. 3. By comparing CoMFA model III of wild type enzyme and model VII of the mutant enzyme, the electrostatic contribution of both models were about the same (43.5% in model III versus 32.2% in model VII), whereas the steric contributions of model VII was higher (56.5% in model III versus 67.8% in model VII). This implies that the steric property is important for the binding of Cyc compounds in the mutant enzyme compared with the wild type enzyme.

CoMFA contour maps

CoMFA with hundreds or thousands of terms is usually represented as the scalar product of the associated coefficients and the standard deviation of all values in the corresponding column of the data table of the (STDEV×COEFF) contour plots. For greater clarity, the resultant CoMFA contours were merged into the binding pocket of the two enzyme targets, namely the wild type and quadruple mutant type *Pf*DHFRs, using compound **14** as the template for the display of the merged CoMFA contour maps. The CoMFA steric and electrostatic fields based on PLS analysis are shown as three-dimensional contour plots for the wild type (Fig. 4a, b) and the quadruple mutant type (Fig. 4c, d). The steric contour map indicates areas in which the molecular steric bulk might have a favorable (green) or unfavorable (yellow) effect on the activity of an analogue. The electrostatic contour map displays blue contours to indicate positive-charge-favoring areas and red contours to indicate negative-charge-favoring areas.

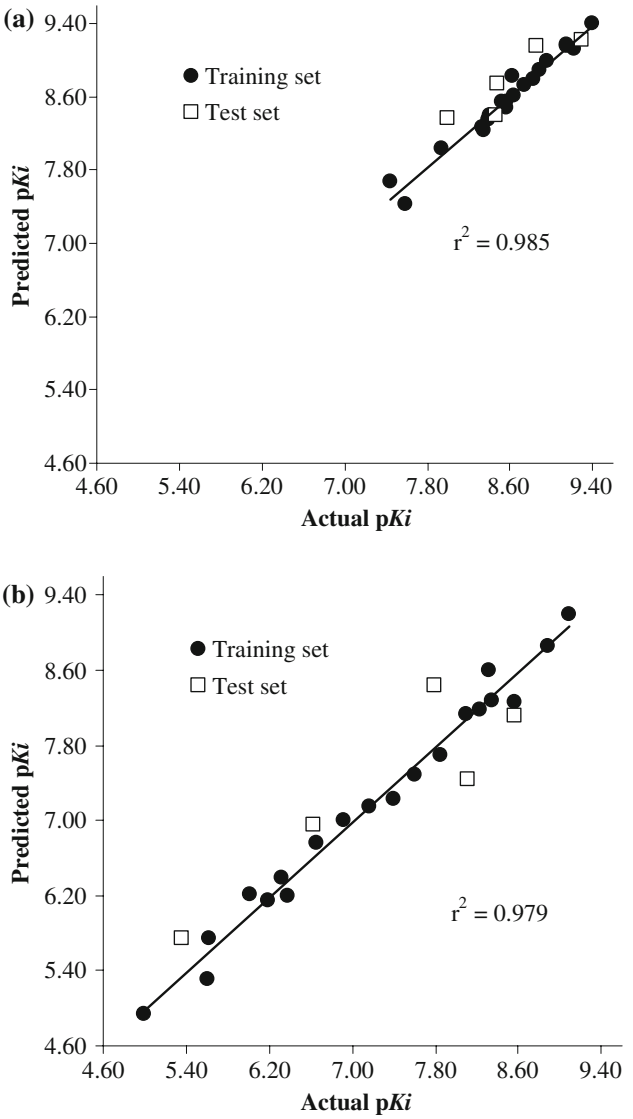

Fig. 3 Plots of the predicted versus actual biological activities (pK_i) of the training set (•) and the test set (□) molecules. The predicted activities were calculated from wild type model III (**a**) and mutant model VII (**b**)

For the CoMFA contour plots of the wild type *Pf*DHFR, the steric contour (Fig. 4a) shows two green regions close to the **Y** and $\mathbf{R_2}$ substitutions, which indicates that bulky groups are preferred in these positions to give higher activity. There is a yellow region near the $\mathbf{R_1}$ substitution, which implies that bulky substitution here is likely to decrease the activity effect. Figure 4b shows red regions near the 2- and 4-amino groups of the triazine ring and along the side-chain of the $\mathbf{R_2}$ substitution. The blue areas are close to the **X** and **Y** substitutions and the 2-amino of triazine ring positions. All steric/electrostatic contour plots are favorably consistent with the experimental data. The Cl substitute at the **Y** position gives a higher activity than the H atom, the phenoxyphenyl side-chain of $\mathbf{R_2}$ shows good

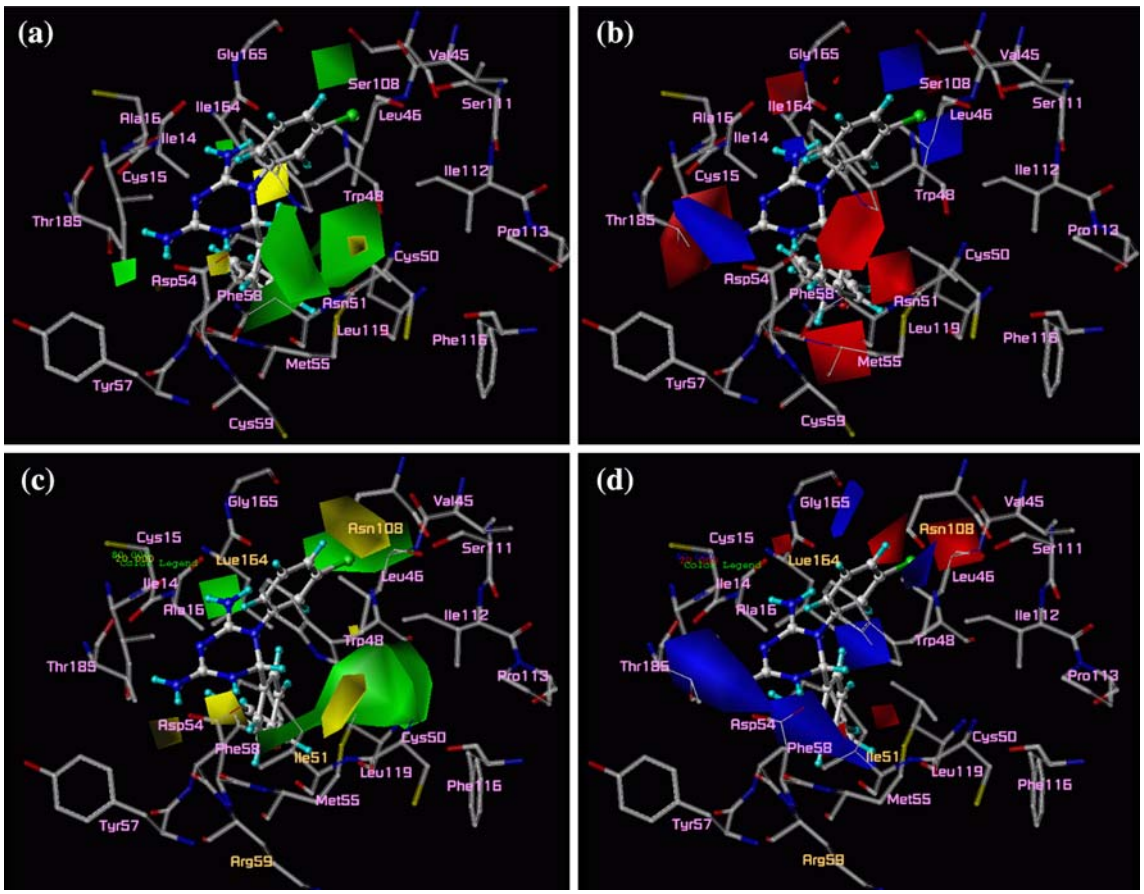

Fig. 4 Steric and electrostatic contour maps of CoMFA models (SD × coeff.) of wild type (**a**, **b**) and mutant *Pf*DHFRs (**c**, **d**). Favorable (contribution levels 80%) and unfavorable (contribution levels 20%) steric areas (**a**, **c**) are indicated by *green* and *yellow regions*, respectively. Favorable electrostatic areas (**b**, **d**) with positive charges (contribution levels 80%) and negative charges (contribution levels 20%) are indicated by *blue* and *red regions*, respectively. Compound **14** and the active site residues are depicted by *balls* and *sticks*

activity, the 2- and 4-amino groups of the triazine ring can interact as a hydrogen-bonding interaction with Asp54, Ile14, and Ile164, and bulkier groups larger than CH_3 are not favorable at the $\mathbf{R_1}$ substitution.

In the case of the CoMFA quadruple mutant steric contour map (Fig. 4c), the major sterically favored green regions are found near the **X** or meta-phenyl ring and the $\mathbf{R_2}$ substitution. On the other hand, there are yellow unfavored steric regions close to the $\mathbf{R_1}$ substitution and the **Y** or *para*-phenyl ring. Considering the merging amino acids, Asn108 displays closer to the **Y** position (Fig. 4c) than Ser108 of wild type *Pf*DHFR (Fig. 4a), therefore this region needs unfavorable bulky group to avoid the steric clash with the amino side-chain of Asn108. For the electrostatic contours (Fig. 4d), the **X** position favored negative-charge substitutes, as indicated by the red colour. Regarding the positive charge electrostatic, represented by the blue contour, there are regions around $\mathbf{R_2}$ and the 2-amino group of triazine ring, suggesting that substituents with low electron density would give a higher level of activity. The distributions of the steric and electrostatic contours are supported by experimental data, with the Cl substitute at the **Y** position showing lower activity than the H atom, which favored nonsteric and positive charge. On the other hand, the Cl substituent gives higher activity than that of H for **X** substitution, favoring steric and negative charge. The $\mathbf{R_2}$ position favors bulky groups with low electron densities, whereas these groups are not favored in the wild type model. Consequently, compound **21** shows a higher level of activity than the compounds that have bulky groups with high electron densities substituents, and $\mathbf{R_1}$ shows unfavorable steric distribution, the same as in the wild type CoMFA steric contour.

Comparison of particular interaction energies of the Cyc compound with the wild type and mutant enzymes

The interactions between individual amino acids and Cyc of the wild type and the quadruple mutant *Pf*DHFR binding

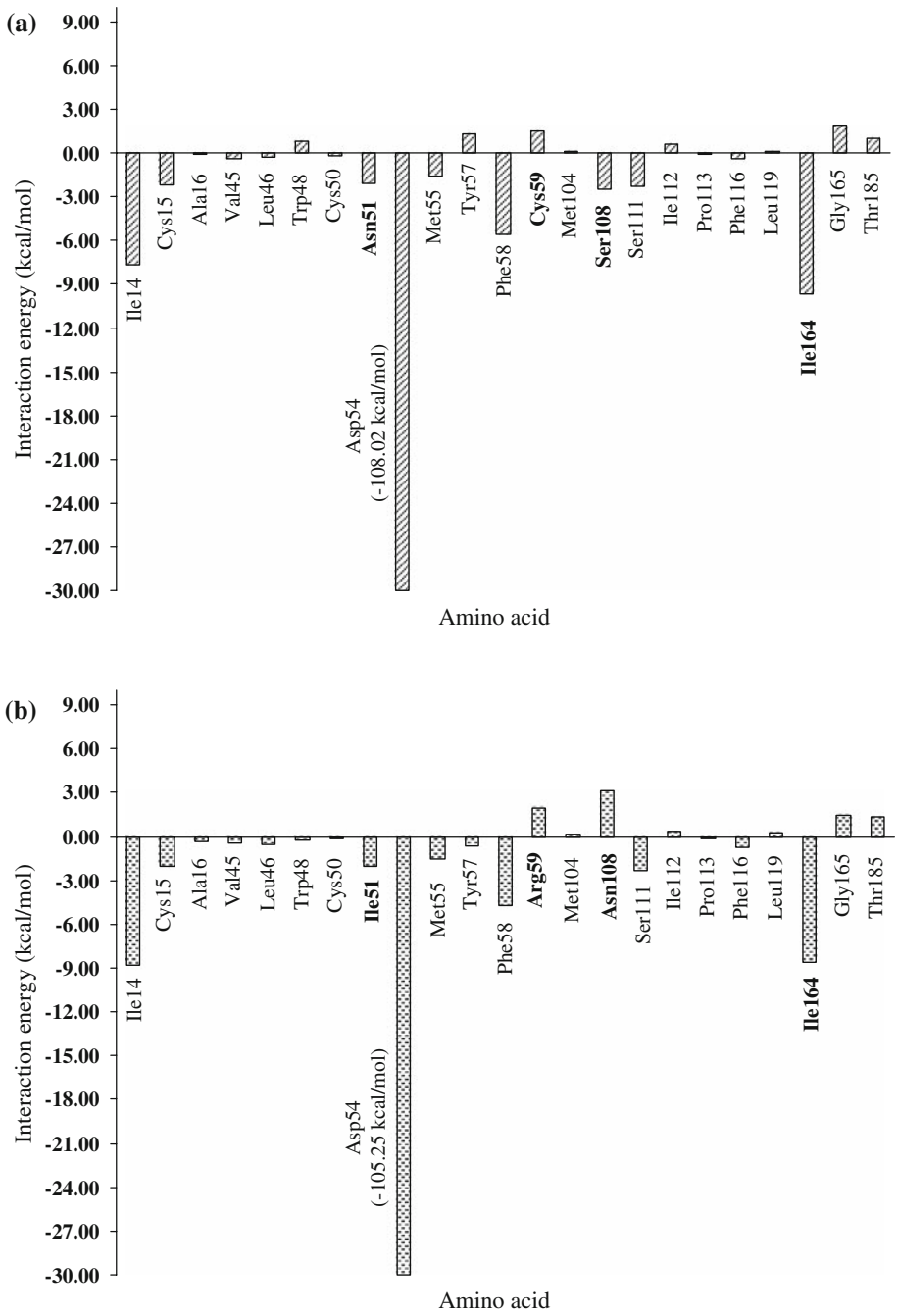

Fig. 5 Obtained MP2/6–31G(d,p) with BSSE interaction energies of Cyc and individual amino acids surrounding the binding pocket of wild type (**a**) and quadruple mutant type (**b**) *Pf*DHFRs

pockets were investigated using MP2/6–31G(d,p)-level calculations, followed by BSSE energy corrections. The obtained corrected interaction energies of Cyc which surrounding the binding pocket of both the wild type and quadruple mutant enzymes are plotted in Fig. 5a and b, respectively. Asp54 shows the strongest interaction energy with the wild type and mutant enzymes. In molecular-level investigation, Asp54 formed a strong hydrogen-bonding interaction with the 2-amino group. Ile14 also formed hydrogen-bonding interaction with the 4-amino group. In the case of Phe58, a pi–pi interaction could be present between the phenyl ring of Phe58 and the 1,3,5-dihydrotriazine ring of the inhibitor. Figure 6 shows the main attractive amino acids, Asp54, Ile14, Ile14-Leu, and Phe58, and the Cyc which are in the binding sites of the wild type and quadruple mutant type.

The two mutations at Asn51Ile and Cys59Arg did not have any significantly different interaction with Cyc in either the wild type and mutant enzyme. Although Ile164 mutated to Leu164, its backbone amino acid still produced hydrogen-bonding interaction with the 4-amino group of the 1,3,5-dihydrotriazine ring. Finally, changing Ser to Asn at the 108 position showed the largest difference in

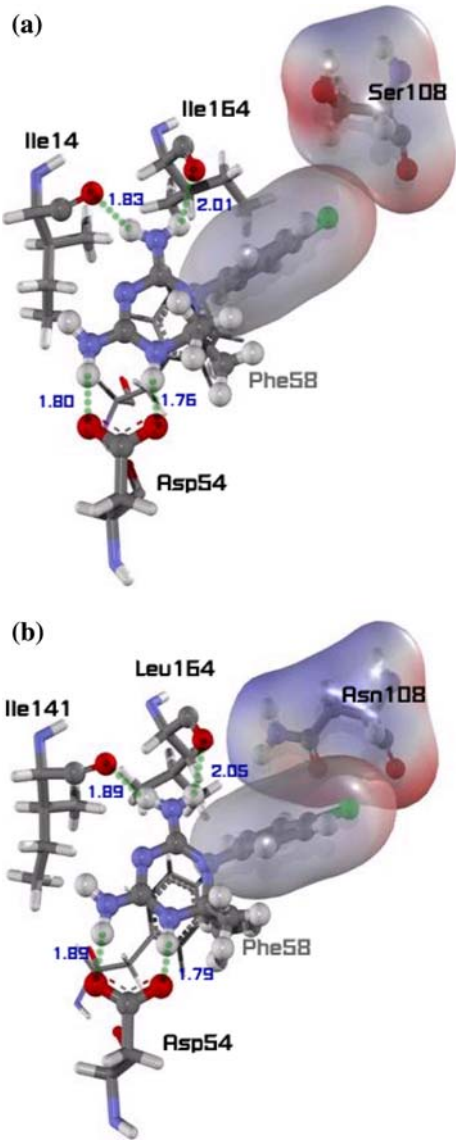

Fig. 6 Hydrogen bonding distances (Å) between Cyc inhibitor and 14, 54, and 164 residues in the binding pocket; **a** wild type and **b** quadruple mutant type *Pf*DHFRs, merged with van der Waals electrostatic interaction of Cyc and Ser108 (**a**) and Asn108 (**b**). *Red* and *blue color* indicated negative and positive electrostatic values

repulsive interaction energy, approximately 4 kcal/mol. This is due to a steric clash between the *p*-Cl phenyl substitute of Cyc and the larger side-chain of Asn108, as shown by the van der Waals surface in Fig. 6. This evidence corresponds to the different favorable wild type and mutant type CoMFA contour map of the **Y** position.

Different interaction energies between the Cyc and WR99210 inhibitors

WR99210 is a good inhibitor against both wild type and quadruple mutant type *Pf*DHFRs because of the greater flexibility of side-chains, as shown by its structure in Fig. 1. The individual interaction energy of the WR99210 inhibitor bound to the two target enzymes of *Pf*DHFRs was calculated by following the Cyc steps of the calculations detailed above in the "AMBER molecular mechanical minimization" and "Quantum chemical calculations" sections. The energy comparisons between the Cyc and WR99210 inhibitors in both the wild type and the quadruple mutant type *Pf*DHFR are depicted in Fig. 7a and b, respectively. Most of the amino acids of the wild type and the mutant *Pf*DHFRs surrounding the inhibitors show the same trend of interaction energy, except for Asn108, which gives the highest repulsive interaction energy for the Cyc inhibitor in the mutant model and shows the largest different interaction energy compared with WR99210 by approximately 5 kcal/mol. However, the flexible WR99210 inhibitor has higher attractive energy in summation than the rigid Cyc inhibitor. Of particular note is the fact that the basic structural information proves that the cause of the Cyc resistances in the quadruple mutant type *Pf*DHFR came from the steric clash of the Asn108 mutation with the repulsive interaction energies, as shown in Fig. 6b, similar to that found with Pyr [19]. On the other hand, the potent WR99210 inhibitor can avoid the steric clash with Asn108, so it does not generate any positive energy value. Moreover, the particular interactions of WR99210 are observed that produced the second binding site in the area of Leu46, Ile112, Pro113, Phe116, and Leu119. In conclusion, quantum chemical calculations can be applied to investigate the relative levels of specific interaction energy of different DHFR inhibitors at a molecular level in biomolecular systems.

Conclusion

This paper presents an application of the CoMFA technique to understanding binding and interaction of Cyc derivatives against wild type and quadruple mutant type *Pf*DHFRs. Satisfactory CoMFA models of both the wild type and the mutant were obtained, with LOO cross-validation r^2_{cv} values of 0.727 and 0.786, respectively. The steric and electrostatics contour maps of Cyc derivatives for wild type enzymes indicate that the Cl substitute at **Y** position is better than H, the long chain with O-linkage between the aromatic rings gives a good activity for $\mathbf{R_2}$, and the bulky group is not suitable for $\mathbf{R_1}$. On the other hand, the CoMFA contour maps of Cyc derivatives active against the quadruple mutant type *Pf*DHFR described the different Cyc structural requirements compared with the wild type. For example, the use of a Cl substitute on *para*-phenyl or **Y** position decreased binding affinity against the quadruple mutant and the long aliphatic side-chain is

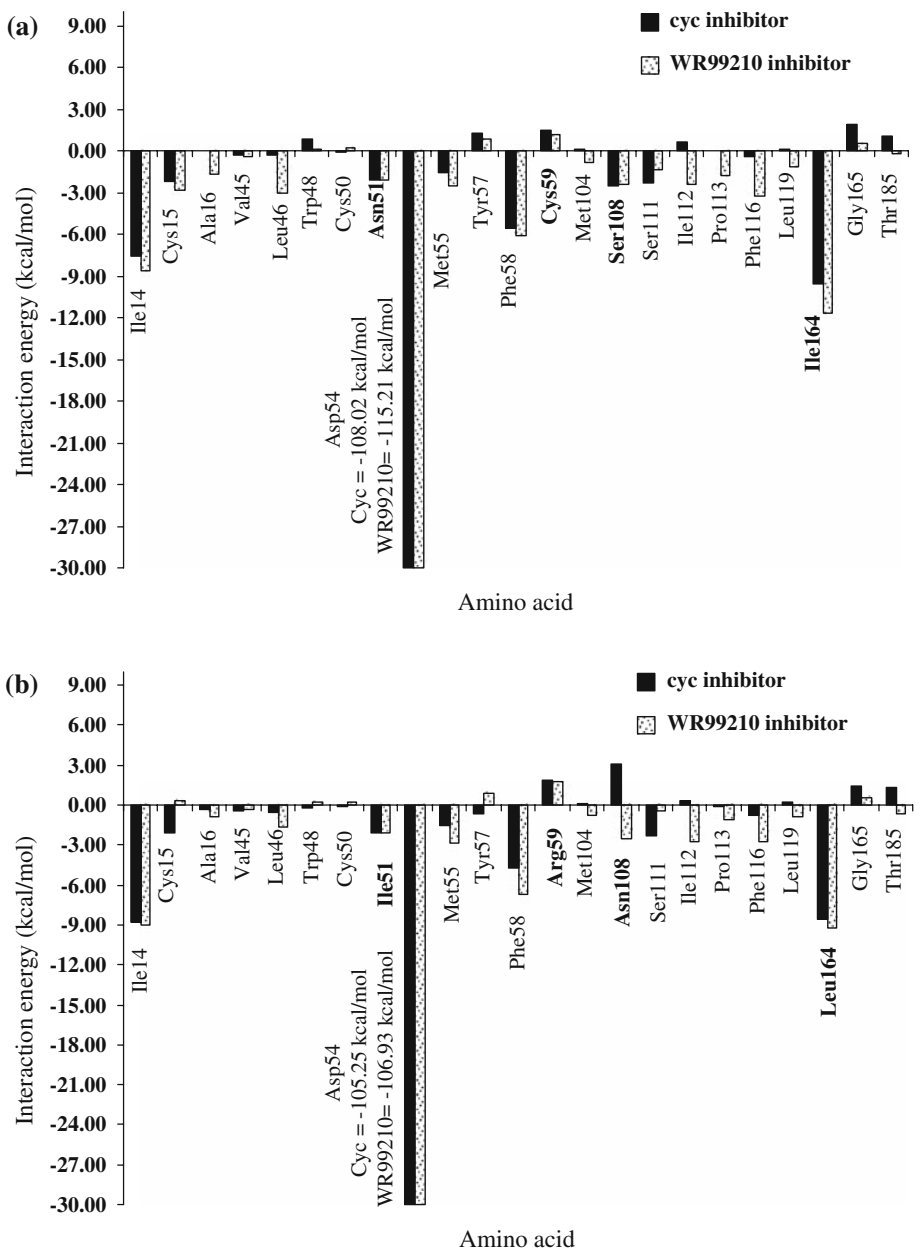

Fig. 7 MP2/6–31G(d,p) with BSSE energetic comparison between Cyc and WR99210 in case of wild type (**a**) and quadruple mutant type (**b**) *Pf*DHFR binding sites

suitable for $\mathbf{R_2}$, which differed from the result for the wild type activity. Therefore, CoMFA appears to be a valuable tool to help understanding of different structure-activity relationships. In addition, the deep molecular energy investigations using quantum chemical calculations of the Cyc bound into binding sites of wild type and quadruple mutant *Pf*DHFRs are shown to be the cause of resistance, arising from the key mutant residue at Ser108Asn. This is caused by the steric clash between the *p*-Cl phenyl of Cyc and the side-chain of the residue Asn108. On the other hand, WR99210, a potent flexible inhibitor does not result in a repulsive energy when the residue changes from Ser108 to Asn108. The CoMFA and quantum chemical approaches are powerful methods for determining the ligand requirements of the structural steric and electrostatic properties in various types of enzymes and the details of the key molecular interactions which cannot be solved by experimentally. The benefit of the two approaches which combined the particular of residue-inhibitor interaction and the structural requirement of inhibitor against different types of *Pf*DHFR, has provided valuable informations to guide for high potent inhibitors.

Acknowledgements This work was supported by the Thai Graduate Institute of Science and Technology (TGIST) Funds (TG-22-11-845D) under the National Science and Technology Development Agency, Thailand. S.H. is grateful to the Thailand Research Fund (RTA5080005). The Postgraduate Education and Research Program in Petroleum Petrochemical Technology and Advanced Materials, the Commission on Higher Education (CHE), the Kasetsart University Research and Development Institute (KURDI), and the Graduate School Kasetsart University Scholarship are gratefully acknowledged for financial support. The authors would like to thank the Large-Scale Research Laboratory of the National Electronics and Computer Technology (NECTEC), the National Center for Genetic Engineering and Biotechnology (BIOTEC), LCAC, and the computing center of KU for computing research facilities. This work has been partially supported by the National Nanotechnology Center (NANOTEC), NSTDA, Ministry of Science and Technology, Thailand, through its Center of Excellence Network program. Medicines for Malaria Venture (MMV), Wellcome Trust, and UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR) for Y.Y. and S.K. S.K. is an international research scholar of Howard Hughes Medical Institute (HHMI), USA. Finally, P.M. would like to thank Dr. Chak Sangma and Dr. Witcha Treesuwan for their AMBER supports.

References

1. Prapunwattana P, Sirawaraporn W, Yuthavong Y, Santi DV (1996) Mol Biochem Parasitol 83:93
2. Dasgupta T, Anderson KS (2008) Biochemistry 47(5):1336
3. Yuthavong Y, Kamchonwongpaisan S, Leartsakulpanich U, Chitnumsub P (2006) Future Microb 1(1):113
4. Yuthavong Y, Yuvaniyama J, Chitnumsub P, Vanichtanankul J, Chusacultanachai S, Tarnchompoo B, Vilaivan T, Kamchonwongpaisan S (2005) Parasitology 130:249
5. Yuvaniyama J, Chitnumsub P, Kamchonwongpaisan S, Vanichtanankul J, Sirawaraporn W, Taylor P, Walkinshaw MD, Yuthavong Y (2003) Nat Struct Biol 10:357
6. Yuthavong Y (2002) Microbes Infect 4:175
7. Nzila A (2006) Drug Discov Today 11(19–20):936
8. Cody V, Schwalbe CH (2006) Crystallog Rev 12(4):301
9. Martin S (2007) ChemMedChem 2(7):944
10. Warhurst DC (2002) Sci Prog 85:89
11. Hastings IM, Donnelly MJ (2005) Drug Resist Updates 8:43
12. Nzila A (2006) J Antimicrob Chemother 57:1043
13. Gregson A, Plowe CV (2005) Pharmacol Rev 57:117
14. Uhlemann AC, Yuthavong Y, Fidock DA (2005) Mol Approaches Malar. ASM Press, Washington DC, pp 429–461
15. Rastelli G, Sirawaraporn W, Sompornpisut P, Vilaivan T, Kamchonwongpaisan S, Quarrell R, Lowe G, Thebtaranonth Y, Yuthavong Y (2000) Bioorg Med Chem 8:1117
16. Peterson DS, Milhous WK, Wellems TE (1990) Proc Natl Acad Sci USA 87:3018
17. Sirawaraporn W, Yongkiettrakul S, Sirawaraporn R, Yuthavong Y, Santi DV (1997) Exp Parasitol 87:245
18. Kamchonwongpaisan S, Quarrell R, Charoensetakul N, Ponsinet R, Vilaivan T, Vanichtanankul J, Tarnchompoo B, Sirawaraporn W, Lowe G, Yuthavong Y (2004) J Med Chem 47:673
19. Maitarad P, Saparpakorn P, Hannongbua S, Kamchonwongpaisan S, Tarnchompoo B, Yuthavong Y (2008) Particular interaction between pyrimethamine derivatives and quadruple mutant type dihydrofolate reductase of *Plasmodium falciparum*: CoMFA and quantum chemical calculations studies. J Enzym Inhib Med Chem. doi:JEIMC07/1339
20. Kubinyi H (1993) QSAR: hansch analysis and related approaches. VCH, Weinheim
21. Hansch C, Leo A (1995) Exploring QSAR. Fundamentals and applications in chemistry and biology. American Chemical Society, Washington, DC
22. Cramer RDIII, Patterson DE, Bunce JD (1988) J Am Chem Soc 110:5959
23. Kubinyi H, Flokers G, Martin YC (1998) Perspect Drug Discov Des 12–14:v–vii
24. Cramer RDIII, Patterson DE, Bunce JD (1989) Prog Clin Biol Res 291:161
25. SYBYL 7.0, Tripos Associates Inc., 1699 South Hanley Road, Suite 303, St Louis, Missouri 63144, USA
26. Golbraikh A, Tropsha A (2002) J Mol Graph Mod 20:269
27. Nilsson J, De Jong S, Smilde AK (1997) J Chemom 11:511
28. Hannongbua S, Nivesanond K, Lawtrakul L, Pungpo P, Wolschann P (2001) J Chem Inf Comput Sci 41:848
29. Cornell WD, Cieplak P, Bayly CI, Kollman PA (1993) J Am Chem Soc 115:9620
30. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2003) Gaussian 03, revision B.05. Gaussian, Inc., Pittsburgh
31. AMBER parameter database. http://www.pharmacy.manchester.ac.uk/bryce/amber/
32. Cummins PL, Ramnarayan K, Singh UC, Gready JE (1991) J Am Chem Soc 113(22):8247
33. Case DA, Darden TA, Cheatham III TE, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA (2006) AMBER9.0
34. Boys SF, Bernardi F (1970) Mol Phys 19:553
35. Saen-oon S, Kuno M, Hannongbua S (2005) Proteins 61:859
36. Kuno M, Hongkrengkai R, Hannongbua S (2006) Chem Phys Lett 424:172
37. Saen-oon S, Aruksakunwong O, Wittayanarakul K, Sompornpisut P, Hannongbua S (2007) J Mol Graph Model 26:720