A combined approach of docking and 3D QSAR study of β-ketoacyl-acyl carrier protein synthase III (FabH) inhibitors

Ali Ashek[a,b] and Seung Joo Cho[a,*]

[a]*Biochemicals Research Center, Korea Institute of Science and Technology, PO Box 131, Cheongryang, Seoul 130-650, Republic of Korea*
[b]*Department of Biomolecular Science, University of Science and Technology, Daejon, Republic of Korea*

Abstract—The enzyme FabH catalyzes the initial step of fatty acid biosynthesis via a type II fatty acid synthase. The pivotal role of this essential enzyme combined with its unique structural features and ubiquitous occurrence in bacteria has made it an attractive new target for the development of antibacterial and antiparasitic compounds. Three-dimensional quantitative structure–activity relationship (3D QSAR) studies such as comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) and docking simulations were conducted on a series of potent benzoylaminobenzoic acids. Docking studies were employed to position the inhibitors into the FabH active site to determine the probable binding conformation. A reasonable correlation between the predicated binding free energy and the inhibitory activity was found. CoMFA and CoMSIA were performed based on the docking conformations, giving q^2 of 0.637 and 0.697 for CoMFA and CoMSIA models, respectively. The predictive ability of the models was validated using a set of compounds that were not included in the training set and progressive scrambling test. Mapping the 3D QSAR models to the active site of FabH related that some important amino acid residues are responsible for protein–inhibitor interaction. These results should be applicable to the prediction of the activities of new FabH inhibitors, as well as providing structural understanding.
© 2005 Elsevier Ltd. All rights reserved.

1. Introduction

Fatty acid biosynthesis in bacteria, plants, and animals is carried out by the ubiquitous fatty acid synthase (FAS) 1 system. In the type I system of animals, including humans, FAS is a homodimer of two large polypeptides, each composed of several distinct enzyme domains and an integral acyl carrier protein (ACP).[1,2] In the type II systems of bacteria,[3] plants,[4] and protozoa,[5] the FAS components, including the ACP, exist as discrete proteins. The corresponding enzymes of the type I and II FASs are related in structure and function but generally lack overall sequence homology. The essentiality of type II FAS for bacterial viability together with major differences between it and type I FAS suggests that broad-spectrum antibacterial drugs may be obtained by screening for inhibitors of the bacterial components.[6–8]

A key enzyme responsible for initiation of bacterial fatty acid biosynthesis has so far escaped serious attention by the drug discovery industry. FabH, a α-keto-acyl-ACP synthase, is the bacterial condensing enzyme in Gram-positive and -negative bacteria that initiates the FAB cycle by catalyzing the first condensation step between acetyl-CoA and malonyl-ACP. The other bacterial condensing enzymes FabB and FabF, functioning later in the cycle, differ significantly from FabH in that they use acyl-ACP rather than acetyl-CoA as the primer for subsequent condensations and are hence nonredundant. As a result, no other known enzyme in the pathway appears to be able to accomplish this essential reaction, and thus, FabH appears to play a key role in the bacterial FAB cycle. FabH proteins from both Gram-positive and -negative bacteria are highly conserved at the sequence and structural level, while there are no significantly homologous proteins in humans. Importantly, the residues that comprise the active site are essentially invariant among Gram-positive and -negative organisms. These attributes suggest that small molecule inhibitors of FabH enzymatic activity could be potential development candidates leading to selective, nontoxic, and broad-spectrum antibacterial.[9]

Three-dimensional quantitative structure–activity relationship (3D QSAR) methods, such as comparative

Keywords: Docking; CoMFA; CoMSIA; FabH; Antibiotic.
* Corresponding author. Tel.: +82 2 958 5134; fax: +82 2 958 5189; e-mail: chosj@kist.re.kr

molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA), were applied to potent FabH inhibitors to gain insights into how steric, electrostatic, hydrophobic, and hydrogen-bonding interactions influence their activity. Derived 3D QSAR model can be used for designing and forecasting the activity of FabH inhibitors of this class. In addition to the 3D QSAR analyses, docking simulations were performed using the X-ray crystallographic structure of the FabH of *Enterccocus faecalis* in complex with an inhibitor to explore the binding modes of these compounds at the active site. Such information is essential to understand the SAR of benzoylaminobenzoic acid derivatives and subsequently for the design of novel and selective inhibitors of FabH.

2. Result and discussion

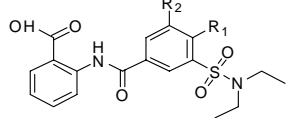### 2.1. Correlation between docking energy and inhibitory activity

Flexible docking of all data sets used for the QSAR study was carried out on the active site of a FabH. The predicted binding energies of these inhibitors into the active site are listed in Tables 1–4. Least-squares fit analyses were performed to explore whether the binding energies could be correlated with activities. The equation was obtained for the inhibitory activities represented as pIC_{50} values, using the FlexX score (ΔE) as the sole descriptor variable. A model with the least-squares fit $r^2 = 0.55$ was obtained for all 43 compounds (Fig. 1)

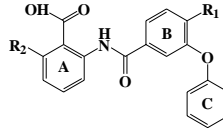$$pIC_{50} = -0.613 - 0.217 * \Delta E.$$

2.2. CoMFA model

CoMFA analysis was performed based on the conformational alignment obtained from the docking of 37 compounds (six compounds were selected as test set for model validation). To derive a predictive relationship, model analysis was conducted by correlating

Table 1. Diethyl sulfonamide derivatives as inhibitors of FabH

Compound	Structure		pIC_{50}	Flex × Score
	R_1	R_2		
1	F	H	5.08	−26.81
2*	Br	H	5.80	−24.64
3	Ph	H	5.80	−29.31
4*	Br	Me	3.80	−25.04
5	OMe	H	4.94	−26.71
6	(N-piperidinyl)	H	5.66	−28.1
7	(O-phenyl)	H	5.21	−30.78
8	(O-cyclopentylmethyl)	H	5.68	−26.78

Table 2. [3-Phenoxybenzoylamino] benzoic acid derivatives as inhibitors of FabH

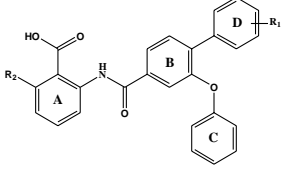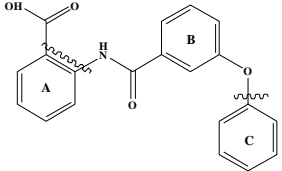| Compound	Structure		pIC_{50}	Flex × Score
	R_1	R_2		
9	H	H	5.57	−28.08
10	F	H	5.42	−28.03
11*	Br	H	4.96	−23.62
12	(4-methylpiperazin-1-yl)	H	4.60	−28.34
13	(methylpiperazine NH)	H	4.42	−27.56
14	(morpholine O)	H	5.49	−31.5
15	(thiomorpholine S)	H	5.92	−26.52
16	(piperidinyl)	H	6.54	−26.75
17	(dimethylpiperidinyl)	H	6.57	−33.07
18	(methylpyrazolyl NH)	H	4.66	−26.16
19	(methylpyridinyl)	H	6.96	−32.32
20*	(methylphenyl)	H	7.25	−31.88
21	Br	OH	7.21	−31.02

Table 3. Aromatic substitutions on the *para* position of [3-phenoxybenzoylamino] benzoic acid derivatives

Compound	Structure		pIC_{50}	Flex × Score
	R_1	R_2		
22	CF_3	H	7.02	−33.64
23	Me	H	6.80	−34.26
24	CO_2H	H	5.68	−28.21
25	OH	H	6.39	−32.82
26	OEt	H	6.66	−33.96
27	SO_2Me	H	7.55	−35.98
28	OCF_3	H	6.33	−33.51
29*	iPr	H	6.10	−34.36
30	3-Me-4-F	H	6.62	−33.28
31	2,4-di-F	H	6.80	−32.27
32	3,4-di-F	H	6.48	−33.08
33	3-Me-4-Cl	H	6.60	−32.59
34*	3-Cl-4-F	H	6.24	−34.05
35	H	OH	8.40	−33.25

variations in the biological activities of the different inhibitors with variations in their CoMFA fields using the partial least-squares (PLS) method. PLS analysis

Table 4. [3-Phenoxybenzoylamino] benzoic acid derivatives as inhibitors of FabH: modification of ring A and C

Compound	Structure		pIC$_{50}$	Flex × Score
	A	C		
36	(3,4-dimethylthiophene)	Ph	5.00	−25.49
37	(thiophene)	Ph	4.37	−24.09
38	(F-phenyl)	Ph	5.22	−29.19
39	(OH-phenyl)	Ph	6.39	−29.32
40	Ph	4-Pyr	5.00	−29.01
41	Ph	3-CO$_2$H-Ph	5.43	−31.59
42	Ph	4-CO$_2$H-Ph	5.36	−30.94
43	Ph	4-F-Ph	5.30	29.02
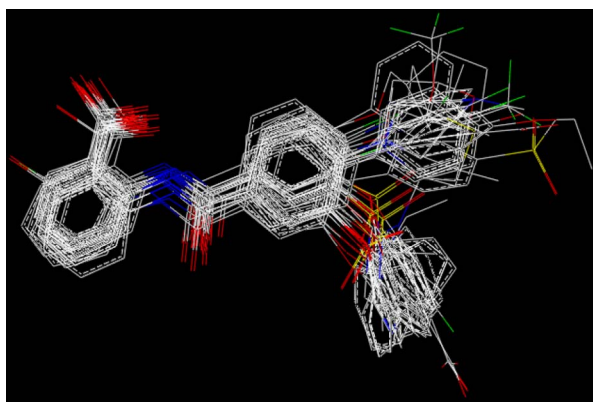
Figure 1. The alignment of all 43 molecules.

Figure 2. Correlation between experimental pIC$_{50}$ and CoMFA predicted pIC$_{50}$ for training and test sets.

yielded a crossvalidated correlation coefficient q^2 of 0.637 with a standard error of prediction (SEP) of 0.618. The results of CoMFA analysis are summarized in Table 6. The noncross-validated PLS analysis gave a conventional r^2 of 0.982 with a SEE of 0.142. These values indicate a good statistical correlation and reasonable predictability of the CoMFA model. Figure 2 shows correlation between the actual values and the predicted values from the final CoMFA model.

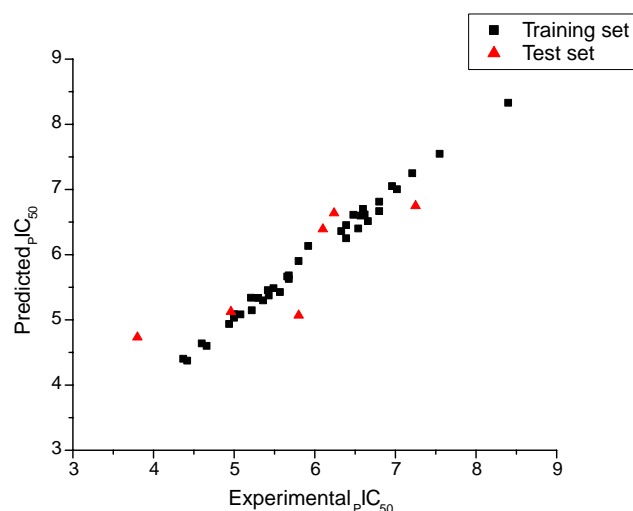The CoMFA contour plots of steric and electrostatic interactions are shown in Figure 3. In Figure 3a, the green contour around the benzene ring D indicates that steric bulk is favored there. As a consequence when the *para* position of the ring B is substituted by phenyl, it shows great-

er activity. This can be seen with compound 20 in which the incorporation of phenyl group in the *para* position of ring B results in an increase of activity as compared to compound 9. Another green contour near the *ortho* to the COOH on the benzene ring A suggests that any bulky substitution there may enhance the activity. As a consequence when the *ortho* hydrogen of ring A of compounds 9, 11, and 20 is replaced by an OH group, it shows greater activity (Compounds 39, 21, and 35). Actually the activity increases mostly due to the formation of an additional hydrogen bond, which is described later.

The yellow contour near the ring C indicates that steric bulk is disfavored; as a consequence when diethyl sulfonamide group is replaced by a less bulky phenoxy group, it is favored. This is in agreement with the fact that the inhibitory activities of compounds 10 and 20 with less bulky phenoxy group are higher than those of their diethyl sulfonamide derivatives 1 and 3.

In Figure 3b, the red contour near the benzene ring D indicates that any negative charge is favored there. As a consequence when a positively charged group substitutes the *para* position of ring B, it causes significant loss of potency. This can be seen with compounds 12, 13, and 18 where positively charged groups are oriented in a red contour, which may lose important interactions at the active site and will be detrimental to activity. But as the same position is also sterically favored, when the *para* position is substituted by charge neutral benzene ring it shows higher activity such as compound 20. We can see another big red contour near the COOH group of ring A which indicates the favorable region for a negatively charged COOH group.

The blue contour close to the ortho of the COOH on the benzene ring A indicates a favorable region for positive charge, which can be explained by the fact that compounds 36 and 37 are less active than compound 9 as they contain a relatively electronegative thiophene ring instead of benzene.
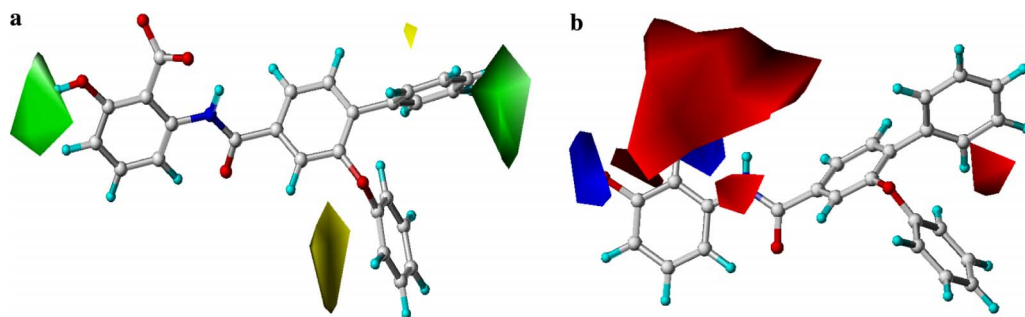
Figure 3. CoMFA contour map (a) Regions where increasing the molecular volume increases bioactivity are in green and regions where increasing the volume decreases the activity are in yellow. (b) Regions where increasing the positive charges increases activity are in blue and regions where increasing the negative charges increases the activity are in red. Compound **35**, the most active ligand, is superimposed on both maps.
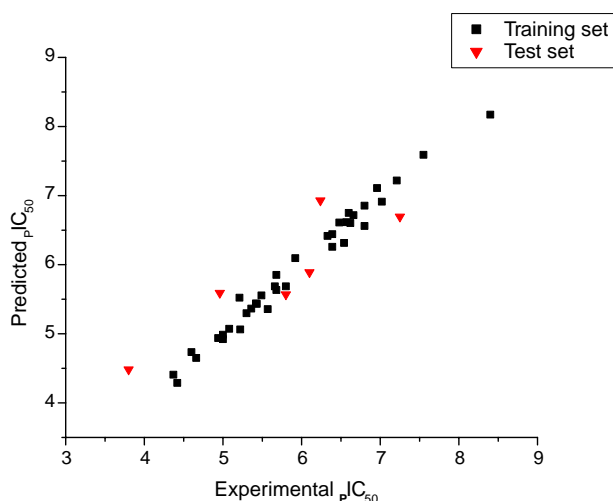
Figure 4. Correlation between experimental pIC_{50} and CoMSIA predicted pIC_{50} for training and test sets.

2.3. CoMSIA model

CoMSIA analyses were performed using the following descriptor fields: steric electrostatic, lipophilicity, and hydrogen bond donor and acceptor in different combinations. The same procedure used for the CoMFA study was also used for the CoMSIA analyses. The CoMSIA results are summarized in Table 6. A cross-validated value, q^2 of 0.697, and a conventional r^2 of 0.972 were obtained when all the five descriptors were considered. The F value and standard error of estimation are 275.6 and 0.137, respectively. These data also indicate that a reliable CoMSIA model was successfully constructed. Figure 4 shows the actual versus predicted values from the final CoMSIA model considering all the five descriptors.

Figure 5a shows the steric contour map of CoMSIA. The green contour on the ring D indicates that any bulky substitution is favored there, whereas a yellow contour around ring C indicates a sterically unfavored region. These contours are also seen in the CoMFA map and are described there. In addition to these, one sterically unfavored yellow contour on ring B is seen. As a result, when the *meta*-positioned proton of compound **2** is replaced by a methyl group in compound **4**, it becomes less active.

Figure 5b shows the electrostatic contour map of CoMSIA. Besides CoMFA contours, one additional red isopleth is present at the junction of ring B and ring C. It indicates that oxygen is more favored than sulfonamide as a linker. It can be explained by the fact that compounds **10**, **16**, and **20** showed more activity than their respective diethyl sulfonamide-substituted analogues 1, 6, and 3. Another additional blue contour is also found near the rind D, which can be explained by the fact that when ring D *para* hydrogen of compound **20** is replaced by any electronegative group it causes significant loss of potency. This can be seen with compounds **22**, **30**, **31**, and **34**.

Figure 6a shows the hydrophobic contour map of CoMSIA. The large yellow isopleths indicate that any hydrophobic substituent in the B ring para position is favored. This can be explained by the fact that when the proton
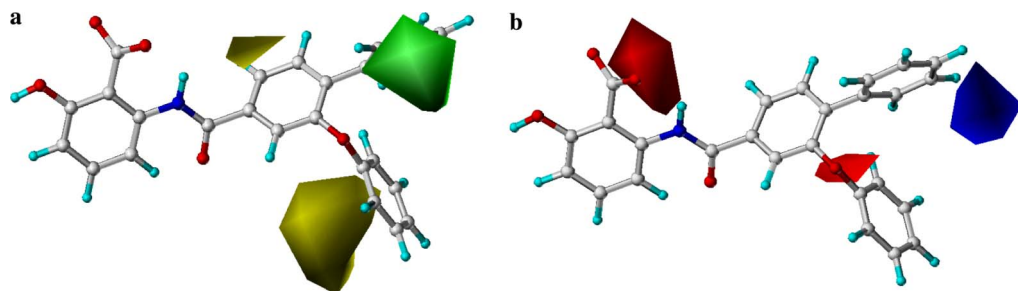
Figure 5. (a) CoMSIA steric contour map. Green and yellow represent sterically favored and disfavored regions, respectively (b) CoMSIA electrostatic contour map. Blue and red represent electrically favored and disfavored regions, respectively. **35**, the most active compound, is overlaid in each plot.
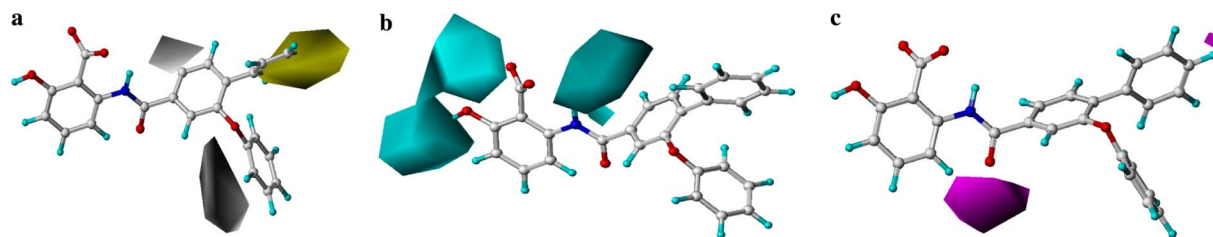
Figure 6. (a) CoMSIA hydrophobic field contour map. Yellow regions indicate where hydrophobic groups increased activity and white regions indicate areas where hydrophilic groups increase activity. (b) CoMSIA H-bond donor contour map. Cyan contour indicates regions where hydrogen bond acceptor groups on the receptor increase activity. (c) CoMSIA H-bond acceptor contour map. Magenta contour indicates regions where hydrogen bond donor groups on the receptor increase activity. **35**, the most active compound, is overlaid in each plot.

of compound **9** is substituted by the phenyl ring (**20**), it shows a 50-fold improvement in potency. Any hydrophilic groups at the *para* position were not tolerated and resulted in significant decreases in potency (**12**, **13**, **18**).

The hydrophilic favored white contour on the ring B implies that any hydrophilic substitution is favored there. This can be explained from the fact that when a methyl group is placed on the *meta* position of ring B (compound **4**), it resulted in a 100-fold loss in inhibition activity.

Figure 6b and c show the CoMSIA H-bond donor and H-bond acceptor contour map, respectively. The cyan isopleths besides the ring A denote the regions where proton acceptors are expected in the receptor and thus a hydrogen-bond-donating substituent should be favored in these regions. As a consequence, when the ortho hydrogen of ring A of compounds **9**, **11**, and **20** is replaced by an OH group, its shows greater activity (compounds **39**, **21**, and **35**). Another cyan isopleth is also present near the amide nitrogen. In the contour plot for the hydrogen-bond-acceptor properties, the area highlighted in magenta represents the region where the hydrogen-bond donors are expected on the receptor site. Therefore, placement of acceptor functions of the ligands directing to these magenta isopleths is correlated with enhanced receptor binding.

2.4. Superimposition of contour map

Figure 7a depicts an overlay of the steric CoMSIA field with the active site. (For better visualization some unimportant amino acids are not shown.) It reveals that the green isopleths are found unoccupied on the receptor site, a consequence of which, any bulky substitution is favored. Moreover when a phenyl ring is present there, it creates a π–π interaction with Trp 38. The presence of small yellow isopleths can be explained by the fact that Asn 253 is only 1.6 Å away from the meta position of B ring. So when a *meta* hydrogen is substituted by any bulky molecule, it is unfavored. The yellow isopleths near C ring are occupied by Phe 224 and Arg 221, so any kind of substituent bulkier than benzene is disfavored there.

Figure 7b depicts an overlay of the electrostatic CoMSIA field with the active site. The contribution of electrostatic energy in the active site is very small (13%) and it is mostly in the solvent exposed region where no direct interaction between ligand and receptor takes place. The red contour near the carboxylic group can be satisfied by the presence of a positively charged His250 amino acid.

Figure 8a depicts an overlay of the hydrophobic CoMSIA field with the active site. The yellow contour on the D ring, which indicates preference for hydrophobic environment, is partly surrounded by a hydrophobic pocket formed by Trp 38, Ser 41, Thr 43, and Val 160,
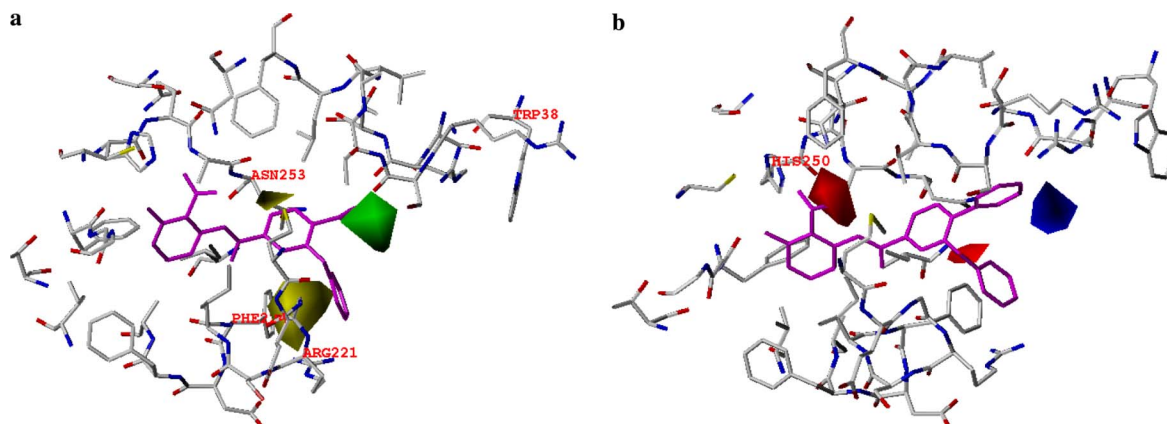
Figure 7. (a) CoMSIA steric, (b) CoMSIA electrostatic fields are projected in the active site of FabH. **35**, the most active compound, is overlaid in each plot.
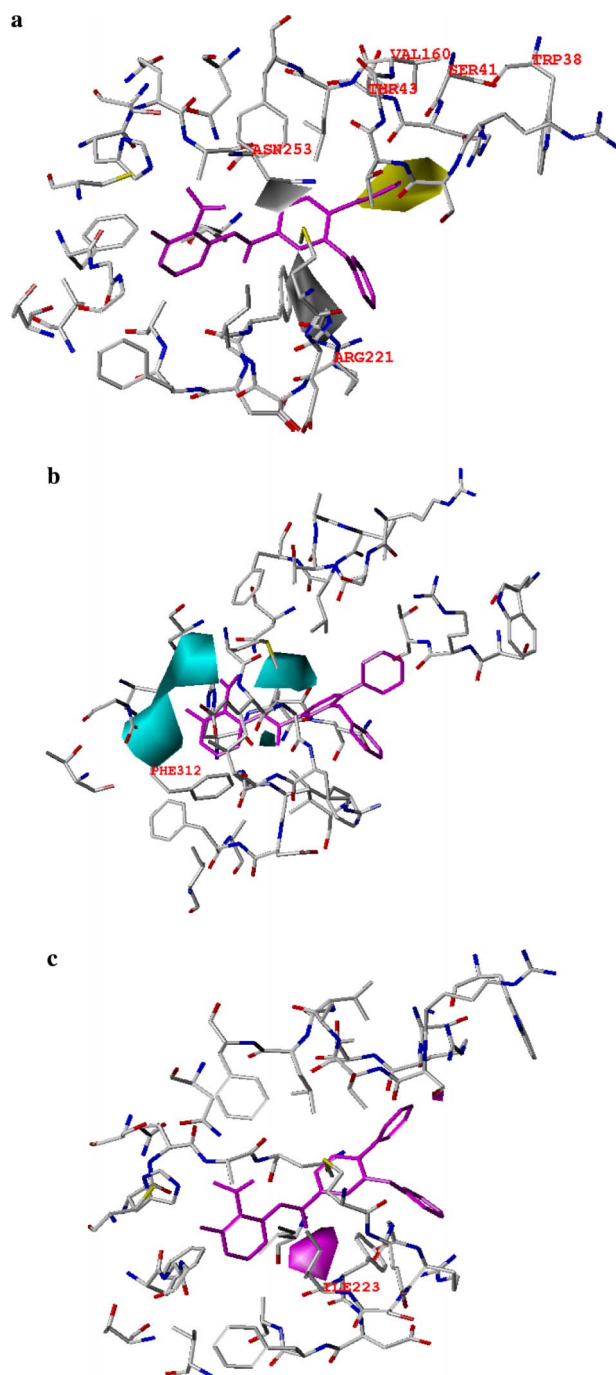
Figure 8. (a) CoMSIA hydrophobic, (b) H-bond donor, and (c) H-bond acceptor fields are projected in the active site of FabH. **35**, the most active compound, is overlaid in each plot.

and is partly solvent exposed. The presence of hydrophilic Asn 253 satisfies the hydrophilic favored white contour near the *meta* position of B ring, whereas the big white hydrophilic contour, besides the D ring, is satisfied by the presence of a positively charged Arg 221.

Figure 8b and c depicts an overlay of H-bond donor and acceptor field with the active site. In Figure 8b, the cyan isocontour denotes regions where proton acceptors are expected in the receptor, which is satisfied by the presence of Phe 312. No acceptor was found for the other

cyan isopleths near the amide nitrogen. In Figure 8c, contour plot for the hydrogen-bond-acceptor properties highlighted in magenta represents regions where hydrogen-bond donors are expected on the receptor site. This is satisfied by the presence of Ile 223 near the magenta isopleths.

2.5. Validation of the 3D QSAR models

2.5.1. Predictive r^2. The six manually selected compounds (compounds **2**, **4**, **11**, **20**, **29**, and **34**) were used as testing set to verify the constructed CoMFA and CoMSIA models. The calculated results are also listed in Table 5 (labeled with an asterisk) and are displayed in Figures 2 and 4 (in triangles). The predicted pIC_{50} with the QSAR models are in good agreement with the experimental data within a statistically tolerable er-

Table 5. Experimental activities (pIC_{50}) and predicted activities (PA) with residuals (Δ) by CoMFA and CoMSIA

Compound	pIC_{50}	CoMFA		CoMSIA	
		PA	Δ	PA	Δ
1	5.08	5.08	0.00	5.07	0.01
2[*]	5.80	5.07	0.73	5.57	0.23
3	5.80	5.90	−0.10	5.69	0.11
4[*]	3.80	4.73	−0.93	4.48	−0.68
5	4.94	4.94	0.00	4.94	0.00
6	5.66	5.66	0.00	5.69	−0.03
7	5.21	5.34	−0.13	5.52	−0.31
8	5.68	5.63	0.05	5.63	0.05
9	5.57	5.43	0.14	5.36	0.21
10	5.42	5.45	−0.03	5.44	−0.02
11[*]	4.96	5.13	−0.17	5.59	−0.63
12	4.60	4.64	−0.04	4.73	−0.13
13	4.42	4.37	0.05	4.29	0.13
14	5.49	5.48	0.01	5.56	−0.07
15	5.92	6.13	−0.21	6.09	−0.17
16	6.54	6.40	0.14	6.32	0.22
17	6.57	6.60	−0.03	6.62	−0.05
18	4.66	4.60	0.06	4.65	0.01
19	6.96	7.05	−0.09	7.11	−0.15
20[*]	7.25	6.75	0.50	6.69	0.56
21	7.21	7.25	−0.04	7.22	−0.01
22	7.02	7.00	0.02	6.91	0.11
23	6.80	6.81	−0.01	6.85	−0.05
24	5.68	5.68	0.00	5.85	−0.17
25	6.39	6.25	0.14	6.26	0.13
26	6.66	6.51	0.15	6.72	−0.06
27	7.55	7.55	0.00	7.59	−0.04
28	6.33	6.36	−0.03	6.42	−0.09
29[*]	6.10	6.39	−0.29	5.89	0.21
30	6.62	6.62	0.00	6.60	0.02
31	6.80	6.67	0.13	6.56	0.24
32	6.48	6.61	−0.13	6.61	−0.13
33	6.60	6.70	−0.10	6.75	−0.15
34[*]	6.24	6.64	−0.40	6.93	−0.69
35	8.40	8.33	0.07	8.17	0.23
36	5.00	5.03	−0.03	4.99	0.01
37	4.37	4.40	−0.03	4.41	−0.04
38	5.22	5.15	0.07	5.06	0.16
39	6.39	6.45	−0.06	6.44	−0.05
40	5.00	5.09	−0.09	4.92	0.08
41	5.43	5.37	0.06	5.43	0.00
42	5.36	5.30	0.06	5.37	−0.01
43	5.30	5.33	−0.03	5.30	0.00

Table 6. Summary of CoMFA and CoMSIA results for various models

Component	CoMFA	CoMSIA			
		SE	SEH	SEDA	ALL
r^2	0.982	0.958	0.923	0.961	0.972
SEE	0.142	0.226	0.257	0.176	0.137
F-value	318.9	131.4	116.6	165.8	275.6
q^2	0.637 (6)	0.583 (7)	0.640 (4)	0.681 (4)	0.697 (6)
SEP	0.618	0.565	0.549	0.547	0.569
r^2_{pred}	0.74	0.43	0.54	0.62	0.76
Field contribution (%)					
S	68	44	21	22	16
E	32	56	31	20	13
H	—	—	48	—	25
D	—	—	—	22	19
A	—	—	—	36	27
r^2_{bs}	0.99	0.98	0.96	0.98	0.99
SD_{bs}	0.003	0.006	0.014	0.006	0.004

Table 7. Results of progressive scrambling of CoMFA and CoMSIA analyses

Component	CoMFA			CoMSIA		
	q^2	cSDE P	$dq^{2'}/dr^2_{yy'}$	q^2	cSDEP	$dq^{2'}/dr^2_{yy'}$
2	0.33	0.78	0.22	0.51	0.67	0.53
3	0.42	0.74	0.58	0.54	0.65	0.68
4	0.48	0.71	0.89	0.56	0.66	0.73
5	0.51	0.69	1.039	0.57	0.66	0.85
6	0.50	0.71	1.24	0.58	0.66	0.93
7	0.48	0.73	1.37	0.57	0.68	1.00
8	—	—	—	0.55	0.71	1.12

ror range, with a correlation coefficient of $r^2 = 0.74$ and 0.76 for CoMFA and CoMSIA models, respectively (Table 6). The testing results indicate that the CoMFA and CoMSIA models could be reliably used in a new inhibitor design for developing drug leads against FabH.

2.5.2. Progressive scrambling. Due to the limited number of compounds demonstrating a statistically significant activity, progressive scrambling was used to validate these models, which gauges the dependence of the model on chance correlations. SAMPLS can potentially overstate the q^2 value when sets with redundant data are examined. Progressive scrambling of the CoMFA model gave the best correlation at five components, with $q^2 = 0.51$ and $dq^{2'}/dr^2_{yy'} = 1.039$ (Table 7). In the case of CoMSIA, the best analysis was performed with six components, producing a q^2 of 0.58 with a $dq^{2'}/dr^2_{yy'}$ value of 0.930. Progressive scrambling of both CoMFA and CoMSIA PLS demonstrates that the models were not dependent on chance correlation.

3. Conclusion

Using the alignment scheme generated from the docking study, highly predictive 3D QSAR models were developed on benzoylaminobenzoic acid FabH inhibitors. The best prediction was obtained by the CoMSIA model with q^2 of 0.697 and r^2 of 0.972. The models were validated with high reliability in the prediction of inhibitory activities by using external test set and progressive

scrambling test. The consistency between the CoMFA and CoMSIA field distributions shows the robustness of the 3D QSAR models. Moreover, there was a satisfactory correlation between the inhibitory activities and the FlexX scores. When we docked the 3D QSAR models into the active site of FabH, we could identify the important amino acids of FabH with high reliability. For example, Trp 38 is important for π–π interaction, Val 160 and Thr 43 are responsible for hydrophobicity, and Phe 312 and Ile 223 are important for hydrogen bonding. Thus, the information obtained from this study can be utilized to further develop novel inhibitors.

4. Materials and methods

4.1. Data sets

Forty-three molecules selected for the present study were taken from the published work by Nie et al.[9] The structures of the compounds and their biological data are given in Tables 1–4. The 3D QSAR models were generated using a training set of 37 molecules. Predictive power of the resulting models was evaluated using a test set of six molecules (Tables 1–4 marked with *). The test compounds were selected manually such that the structural diversity and wide range of activity in the data set were included. The biological activity used in the present study was expressed as

$$pIC_{50} = -\log IC_{50},$$

where IC_{50} is the concentration (μM) of the inhibitor producing 50% inhibition of *E. faecalis* FabH. pIC_{50} values were used as dependent variables in the CoMFA and CoMSIA analyses.

4.2. Molecular modeling and alignment

Three-dimensional structure building and all modeling were performed using the SYBYL program package,[10] version 6.9 on a Silicon Graphics Octane (R1200) workstation with the IRIX 6.5 operating system. Energy minimizations were performed using the Tripos force field[11] and the Gasteiger–Huckel charge with a distance-depen-

dent dielectric and conjugate gradient method. The convergence criterion was 0.01 kcal/mol.

CoMFA and CoMSIA studies require that the 3D structures of the molecules to be analyzed be aligned according to a suitable conformational template, which is assumed to be a 'bioactive' conformation.[12] Twenty conformations were obtained through FlexX for each ligand. The conformation corresponding to the lowest free energies was selected as the most probable binding conformation. These conformations were aligned together inside active site and used directly for CoMFA and CoMSIA analyses. Figure 1 depicts the aligned structures.

4.3. Docking

The FlexX program[13] version 1.9 interfaced with SYBYL 6.9 was used to dock all the compounds onto the active site of FabH in order to determine the probable binding conformations of these inhibitors. FlexX is a fast automated docking program that considers ligand conformational flexibility by an incremental fragment placing technique.[13,14] We used the program to dock the training set as well as the test set molecules into the active site of FabH. The crystal structure of FabH complexed with compound **17** collected from Nie et al.[9] (not deposited in PDB). The crystal structure was stripped of inhibitor and water molecules, and hydrogen atoms were generated in standard geometry. Protonation states were assumed to be those most common at pH 7, that is, lysines, arginines, aspartates, and glutamates are ionized. The active site for docking was defined as all atoms within 6.5 Å radius of the co-crystallized ligand. The default SYBYL FlexX parameters were used. The docking reliability was validated using the known X-ray structure of FabH in complex with a small molecular ligand. The ligand 17 was re-docked to the binding site of protein and the docked conformation corresponding to the lowest free energies was selected as the most probable binding conformation. The root-mean-square deviation (RMSD) of the docked conformation to the experimental conformation was 0.8 Å, suggesting a high reliability of FlexX in reproducing the experimentally observed binding mode for FabH inhibitors.

4.4. CoMFA and CoMSIA 3D QSAR models

The steric and electrostatic potential fields for CoMFA were calculated at each lattice intersection of a regularly spaced grid of 2.0 Å. The lattice was defined automatically and is extended 4 Å units past van der Waals volume of all molecules in X, Y, and Z directions. The van der Waals potential (Lennard-Jones 6–12) and coulombic term, which represent steric and electrostatic fields, respectively, were calculated using Tripos force field.[11] A distance-dependent dielectric expression $\varepsilon = \varepsilon_0 \cdot R_{ij}$ with $\varepsilon_0 = 1.0$ was used. An sp^3 carbon atom with a van der Waals radius of 1.52 Å and +1.0 charge served as the probe atom to calculate steric and electrostatic fields. The steric and electrostatic contributions were truncated to ±30 kcal/mol and electrostatic contributions were ignored at lattice intersections with maximum steric inter-

actions. The CoMFA steric and electrostatic fields generated were scaled by CoMFA standard option given in SYBYL. CoMSIA similarity indices descriptors were derived according to Klebe et al.[15] with the same lattice box as was used for the CoMFA calculations, with a grid spacing of 2 Å employing a C^{+1} probe atom with a radius of 1.0 Å as implemented in SYBYL. CoMSIA similarity indices (A_F) for a molecule j with atoms i at a grid point q are calculated by Eq. 1 as follows:

$$A_{F,K}^q(j) = -\sum \omega_{probe,k} \omega_{ik} e^{-\alpha r_{iq}^2}. \qquad (1)$$

Five physicochemical properties k (steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor) were evaluated using the probe atom. A Gaussian type distance dependence was used between the grid point q and each atom i of the molecule. The default value of 0.3 was used as the attenuation factor (α). The steric indices are related to the third power of the atomic radii, the electrostatic descriptors are derived from atomic partial charges, the hydrophobic fields are derived from atom-based parameters developed by Viswanadhan et al.,[16] and the hydrogen bond donor and acceptor indices are obtained by a rule-based method derived from experimental values.[17]

4.5. PLS calculations and validations

Partial least-square (PLS)[18,19] methodology was used for all 3D QSAR analyses. Column filtering was set to 1.0 kcal/mol to speed up the analysis and reduce the noise. The CoMFA and CoMSIA descriptors were used as independent variables, and pIC$_{50}$ values were used as dependent variables in partial least-squares regression analyses to derive 3D QSAR models using the standard implementation in the SYBYL package. The predictive value of the models was evaluated first by leave-one-out (LOO)[20,21] crossvalidation. The crossvalidated coefficient, q^2, was calculated using Eq. 2

$$q^2 = 1 - \frac{\sum (Y_{predicted} - Y_{observed})^2}{\sum (Y_{observed} - Y_{mean})^2}, \qquad (2)$$

where $Y_{predicted}$, $Y_{observed}$, and Y_{mean} are predicted, actual, and mean values of the target property (pIC$_{50}$), respectively. $\sum (Y_{predicted} - Y_{observed})^2$ is the predictive sum of squares (PRESS). To maintain the optimum number of PLS components and minimize the tendency to overfit the data, the number of components corresponding to the lowest PRESS value was used to derive the final PLS regression models. In addition to the q^2 and number of components, the conventional correlation coefficient r^2 and its standard errors (SEE) were also computed. To further assess the robustness and statistical confidence of the derived models, bootstrapping analysis[19] (100 runs) was performed and the mean r^2 is given as q^2 bootstrap.

4.6. Predictive r squared (r² pred)

To validate the derived CoMFA and CoMSIA models, biological activities of external test set were predicted using models derived from the training set. The predictive ability of the models is expressed by predictive r^2

value, which is analogous to crossvalidated $r^2(q^2)$ and is calculated by using Eq. 3

$$r^2_{\text{pred}} = \frac{\text{SD-PRESS}}{\text{SD}},\qquad(3)$$

where SD is the sum of squared deviation between the biological activities of the test set molecule and the mean activity of the training set molecules and PRESS is the sum of squared deviations between the observed and the predicted activities of the test molecules.

4.7. Progressive scrambling

Progressive scrambling allows for evaluation of the sensitivity of a QSAR model to chance correlations.[22] Thirty scramblings were carried out with a maximum of 10 bins and a minimum of two bins. The critical point was 0.85. Progressive scrambling of the biological data produces three statistics. q^2 and cSDEP are the predictivity of the model and the calculated crossvalidated standard error, respectively. The susceptibility of the model to chance correlation can be gauged by the slope of q^2 (as originally determined using SAMPLS) with respect to the correlation of the original biological activity versus the scrambled biological activity. This is denoted as $\mathrm{d}q^{2\prime}/\mathrm{d}r^2_{yy\prime}$. In a model where the correlation of activity is not due to chance, the slope should be in-between 0.8 and 1.2. It should be equal to one for an ideal model.

References and notes

1. Jayakumar, A.; Tai, M. H.; Huang, W. Y.; Al Feel, W.; Hsu, M.; Abu-Elheiga, L.; Chirala, S. S.; Wakil, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 8695.
2. Chirala, S. S.; Huang, W. Y.; Jayakumar, A.; Sakai, K.; Wakil, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 5588.
3. Cronan, J. E., Jr.; Rock, C. O. Biosynthesis of Membrane Lipids. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*; Neidhardt, F. C., et al., Eds., Second Ed.; American Society for Microbiology: Washington, DC, 1996; pp 612–636.
4. Clough, R. C.; Matthis, A. L.; Barnum, S. R.; Jaworski, J. G. *J. Biol. Chem.* **1992**, *267*, 20992.
5. Waller, R. F.; Keeling, P. J.; Donald, R. G.; Striepen, B.; Handman, E.; Lang-Unnasch, N.; Cowman, A. F.; Besra, G. S.; Roos, D. S.; McFadden, G. I. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12352.
6. Qiu, X.; Janson, C. A.; Court, R. I.; Smyth, M. G.; Payne, D. J.; Abdel-Meguid, S. S. *Protein Sci.* **1999**, *8*, 2529.
7. Banerjee, A.; Dubnau, E.; Quemard, A.; Balasubramanian, V.; Um, K. S.; Wilson, T.; Collins, D.; de Lisle, G.; Jacobs, W. R. *Science* **1994**, *263*, 227.
8. Mdluli, K.; Slayden, R. A.; Zhu, Y.; Ramaswamy, S.; Pan, X.; Mead, D.; Crane, D. D.; Musser, J. M.; Barry, C. E., III *Science* **1998**, *280*, 1607.
9. Nie, Z.; Perretta, C.; Lu, J.; Su, Y.; Margosiak, S.; Gajiwala, K. S.; Cortez, J.; Nikulin, V.; Yager, K. M.; Appelt, K.; Chu, S. *J. Med. Chem.* **2005**, *48*, 1596.
10. SYBYL 6.9. Tripos Inc., St. Louis, MO 63144, USA.
11. Clark, M.; Cramer, R. D., III; Van Opdenbosch, N. *J. Comput. Chem.* **1989**, *10*, 982.
12. Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
13. Kramer, B.; Rarey, M.; Lengauer, T. *Proteins* **1999**, *37*, 228.
14. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470.
15. Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130.
16. Viswanadhan, V. N.; Ghose, A. K.; Revenkar, G. R.; Robins, R. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163.
17. Klebe, G. *J. Mol. Biol.* **1994**, *237*, 212.
18. Wold, S.; Albano, C.; Dunn, W. J.; Edlund, U.; Esbenson, K.; Geladi, P.; Hellberg, S.; Lindburg, W.; Sjostrom, M. In *Chemometrics*; Kowalski, B., Ed.; Reidel: Dordrecht, The Netherlands, 1984; Vol. 17.
19. Geladi, P. *J. Chemon.* **1998**, *2*, 231.
20. Cramer, R. D., III; Bunce, J. D.; Patterson, D. E. *Quant. Struct. Act. Relat.* **1988**, *7*, 18.
21. Wold, S. *Technometrics* **1978**, *4*, 397.
22. Clark, R. D.; Sprous, D. G.; Leonard, J. M. *Prous Science*; SA: Barcelona, 2001; pp 475.