

# Molecular docking and 3D-QSAR studies of *Yersinia* protein tyrosine phosphatase YopH inhibitors

Xin Hu and C. Erec Stebbins\*

Laboratory of Structural Microbiology, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA

Received 7 September 2004; revised 9 November 2004; accepted 15 November 2004

Available online 8 December 2004

**Abstract**—Three-dimensional quantitative structure–activity relationship (QSAR) studies were conducted on two classes of recently explored compounds with known YopH inhibitory activities. Docking studies were employed to position the inhibitors into the YopH active site to determine the probable binding conformation. Good correlations between the predicated binding free energies and the inhibitory activities were found for two subsets of phosphate mimetics:  $\alpha$ -ketocarboxylic acid and squaric acid ( $R^2 = 0.70$  and  $0.68$ , respectively). The docking results also provided a reliable conformational alignment scheme for 3D-QSAR modeling. Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) were performed based on the docking conformations, giving  $q^2$  of  $0.734$  and  $0.754$  for CoMFA and CoMSIA models, respectively. The 3D-QSAR models were significantly improved after removal of an outlier ( $q^2 = 0.829$  for CoMFA and  $q^2 = 0.837$  for CoMSIA). The predictive ability of the models was validated using a set of compounds that were not included in the training set. Mapping the 3D-QSAR models to the active site of YopH provides new insight into the protein–inhibitor interactions for this enzyme. These results should be applicable to the prediction of the activities of new YopH inhibitors, as well as providing structural implications for designing potent and selective YopH inhibitors as antiplague agents.

© 2004 Elsevier Ltd. All rights reserved.

## 1. Introduction

*Yersinia* spp. are causative agents of human bacterial diseases ranging from gastrointestinal syndromes to the plague.<sup>1</sup> The *Yersinia* genus utilizes a type III secretion machinery to translocate virulence proteins, referred to as Yops (*Yersinia* Outer Proteins), into the cytosol of the host cell.<sup>2,3</sup> One of these virulence factors, YopH, is an essential determinant for *Yersinia* pathogenicity.<sup>4,5</sup> YopH belongs to the family of protein tyrosine phosphatases (PTPs). Two of the major cellular targets of YopH have been identified as the focal adhesion kinase (FAK) and the focal adhesion protein p130<sup>Cas</sup>.<sup>6</sup> Upon translocation into host cell, YopH acts on the pTyr-containing protein targets to disrupt the normal cytoskeletal structure and regulate numerous cellular functions.<sup>7,8</sup>

PTPs represent an important class of drug targets for a wide variety of diseases including cancer, diabetes, immune-related diseases, and infection.<sup>9,10</sup> YopH has

recently emerged as a therapeutic target of interest due to its status as an essential virulence determinant in a weaponized organism.<sup>11–13</sup> A number of compounds have been discovered to inhibit the activity of YopH.<sup>14–16</sup> Progress has recently made on the development of highly potent and specific YopH inhibitors, offering promise in finding effective drug candidates that may serve as a novel class of antibacterial agents that target virulence.<sup>17–20</sup>

We previously reported a comparative docking study of *Yersinia* PTP YopH and *Salmonella* PTP SptP with respect to the eukaryotic PTP1B, aimed at identifying structural features in bacterial PTPs that can be used for structure-based drug design.<sup>21</sup> In this paper, we report a 3D-QSAR analysis on YopH inhibitors in order to seek new insights into the relationship between the structural information and the inhibitory potency. The probable binding conformations of YopH inhibitors were determined using a flexible docking approach. Highly reliable comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) models were derived based on the binding conformations in the active site of YopH. The comparisons of 3D-QSAR models with the 3D topology of YopH structure provide insight into the molecular

**Keywords:** YopH inhibitors; Docking; 3D-QSAR; CoMFA; CoMSIA.

\*Corresponding author. Tel.: +1 212 327 7190; fax: +1 212 327 7191; e-mail: [stebbins@rockefeller.edu](mailto:stebbins@rockefeller.edu)

basis of protein–inhibitor interactions, and are thereby useful in the development of novel and selective inhibitors for YopH.

## 2. Results and discussion

### 2.1. Molecular docking

**2.1.1. Binding conformations of YopH inhibitors.** Thirty-nine YopH inhibitors were selected from Seto's group (Table 1).<sup>14,15,19,20</sup> The compounds belong to two different chemical classes of phosphate mimetics:  $\alpha$ -ketocarboxylic acid and squaric acid.<sup>10</sup> In order to determine the probable binding conformations of these YopH inhibitors, the program AutoDock 3.0<sup>22</sup> was used to dock all compounds into the active site of the target protein. The docking reliability was validated using the known X-ray structure of YopH in complex with a small molecular ligand (PDB entry 1PA9, Fig. 1A).<sup>18</sup> The ligand *p*-nitrocatechol (pCNS) was re-docked to the binding site of protein employing the LGA algorithm,<sup>22</sup> 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 0.45 Å, suggesting a high docking reliability of AutoDock in reproducing the experimentally observed binding mode for YopH inhibitors.

All selected compounds were docked into the active site of YopH using the same protocol. Figure 1B shows the 39 compounds docked into the binding site, providing a well-defined binding mode of YopH inhibitors, as we observed in a previous study.<sup>21</sup> Firstly, all inhibitors are characterized with a phosphate mimetic ( $\alpha$ -ketocarboxylic acid or squaric acid) that is inserted into the pTyr catalytic binding pocket. Figure 1C and D depict the two different phosphate mimetics interacting with the residues of P-loop and WPD loop in the catalytic binding pocket. These interactions are well conserved and play a crucial role in ligand binding. Secondly, all inhibitors share a common aromatic structure that participates in a  $\pi$ – $\pi$  stacking interaction with residue Phe229. Residues Arg230 and Asp231 form a charged region that contributes to inhibitory potency as well as selectivity. Thirdly, as observed in PTP1B, YopH possesses a second phosphate binding site that allows simultaneous interactions of an inhibitor at both the active site and an adjacent peripheral site to achieve high potency and selectivity of inhibition.<sup>21</sup>

**2.1.2. The prediction of binding free energies.** An important application of AutoDock in structure-based drug design is to predict the binding free energies while determining the binding conformation of an inhibitor in the active site. The scoring reliability of AutoDock was evaluated by correlating the predicted binding free energies with the experimental inhibitory activities for the 39 YopH inhibitors. The squared correlation coefficient ( $R^2$ ) is 0.41 with a standard error (SE) of 1.34 (Eq. 1). This value is generally lower than that observed in other similar studies.<sup>21,23</sup> Since the phosphate mimetics of

YopH inhibitors dominate the binding at the catalytic binding pocket, it could be argued that the predicted binding free energies for the compounds bearing the same phosphate mimetic would be more correlated with the experimental data than that of combined data set. We therefore evaluated the correlations of the two classes of YopH inhibitors separately. As expected, the correlations within the subset were improved remarkably. The  $R^2$  was increased to 0.70 for the  $\alpha$ -ketocarboxylic acid and 0.68 for the squaric acid (Eqs. 2 and 3).

$$-\log \text{IC}_{50} = -0.6635 - 0.4855 \times \Delta G$$

$$(n = 39, \quad R^2 = 0.409, \quad \text{SE} = 1.335) \quad (1)$$

$\alpha$ -ketocarboxylic acid:

$$-\log \text{IC}_{50} = -5.8573 - 0.9624 \times \Delta G$$

$$(n = 25, \quad R^2 = 0.703, \quad \text{SE} = 0.759) \quad (2)$$

squaric acid:

$$-\log \text{IC}_{50} = -6.7618 - 1.38 \times \Delta G$$

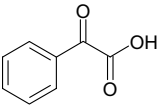
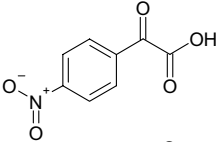
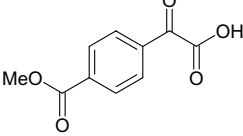
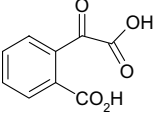
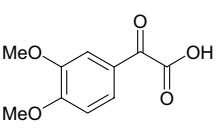
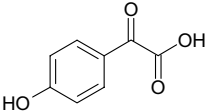
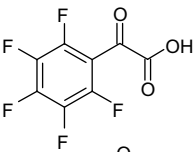
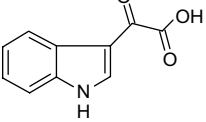
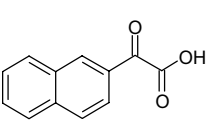
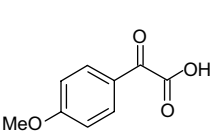
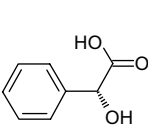
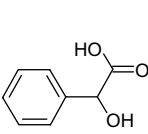
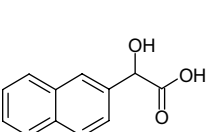
$$(n = 14, \quad R^2 = 0.683, \quad \text{SE} = 0.202) \quad (3)$$

In addition, as shown in Figure 2, compound C17 obviously deviates from other members of the  $\alpha$ -ketocarboxylic acid compounds. In fact, compared to other structurally similar compounds in the  $\alpha$ -ketocarboxylic acid group, C17 appears to have a higher activity than expected. Omission of this outlier increased the  $R^2$  to 0.82 for the series of  $\alpha$ -ketocarboxylates. The better correlations within the two classes of YopH inhibitors suggest that the binding of  $\alpha$ -ketocarboxylic acid and squaric acid in the conserved catalytic binding pocket of YopH is likely different. The predicted binding free energies of the  $\alpha$ -ketocarboxylic acid are  $-1.0$  kcal/mol lower than that of the squaric acid in general (Fig. 2). This is most probably because the electrostatic interactions of the negatively charged  $\alpha$ -ketocarboxylic acid in the catalytic pocket, especially with the positive charged residue Arg409, appear much stronger than that of the neutral phosphate mimetic squaric acid.

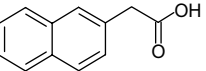
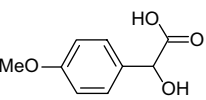
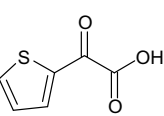
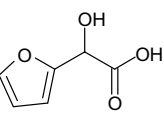
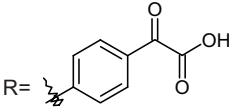
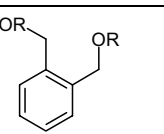
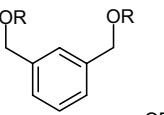
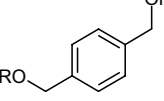
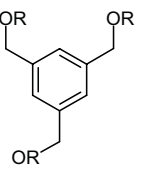
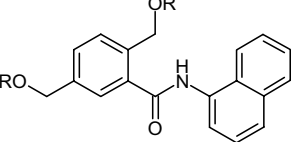
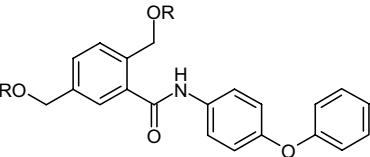
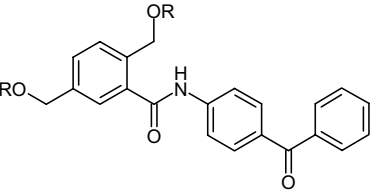
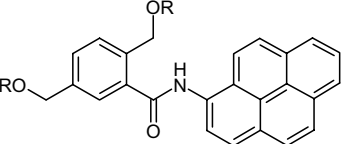
### 2.2. 3D-QSAR studies

**2.2.1. CoMFA analysis.** CoMFA analysis was performed based on the conformational alignment obtained from the docking of 35 compounds (four compounds were selected as test set for model validation). To derive a predictive relationship model, analysis was conducted by correlating variations in the biological activities of the different inhibitors with variations in their CoMFA fields using the partial least squares (PLS) method. The results of CoMFA analysis are summarized in Table 2. PLS analysis yielded a high cross-validated correlation coefficient  $q^2$  of 0.734 with standard error of prediction SEP of 0.693. The non-cross-validated PLS analysis gave a conventional  $r^2$  of 0.914 with SE of 0.414. These values indicate a good statistical correlation and reasonable predictability of the CoMFA model. The steric field descriptors explain 60% of the variance, while the electrostatic descriptors explain 40%. Analysis on the two classes of  $\alpha$ -ketocarboxylic acid and squaric acid separately did not show any

**Table 1.** Structures and inhibitory activities of 39 YopH inhibitors

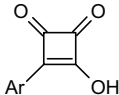
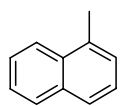
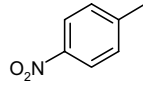
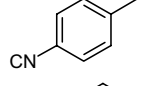
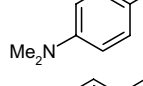
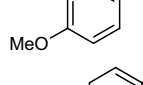
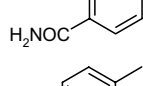
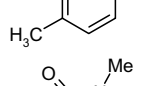
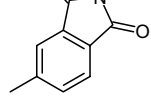
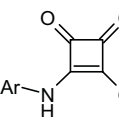
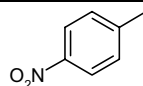
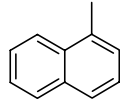
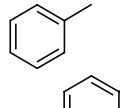
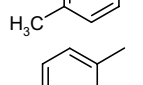
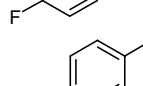
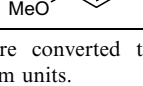
ID	Structure	pIC <sub>50</sub> <sup>a</sup>
C1		2.57
C2		2.57
C3		2.64
C4		2.82
C5		2.77
C6		3.70
C7 <sup>b</sup>		3.10
C8		3.41
C9		3.72
C10		3.82
C11		2.26
C12		2.31
C13		2.48

**Table 1 (continued)**

ID	Structure	pIC <sub>50</sub> <sup>a</sup>
C14		2.46
C15		2.44
C16		3.02
C17		4.57
R = 		
C18		5.07
C19		5.39
C20		6.15
C21		5.77
C22		6.298
C23		6.628
C24 <sup>b</sup>		6.178
C25		6.57

(continued on next page)

Table 1 (continued)

ID	Structure	pIC <sub>50</sub> <sup>a</sup>
		
C26		4.33
C27 <sup>b</sup>		4.25
C28		4.17
C29		4.00
C30		3.72
C31		3.64
C32		3.62
C33		3.59
		
C34		3.92
C35		3.46
C36		3.11
C37		2.92
C38 <sup>b</sup>		2.85
C39		2.38

<sup>a</sup> The IC<sub>50</sub> values were converted to “molar” quantities before expressing in logarithm units.

<sup>b</sup> Test set.

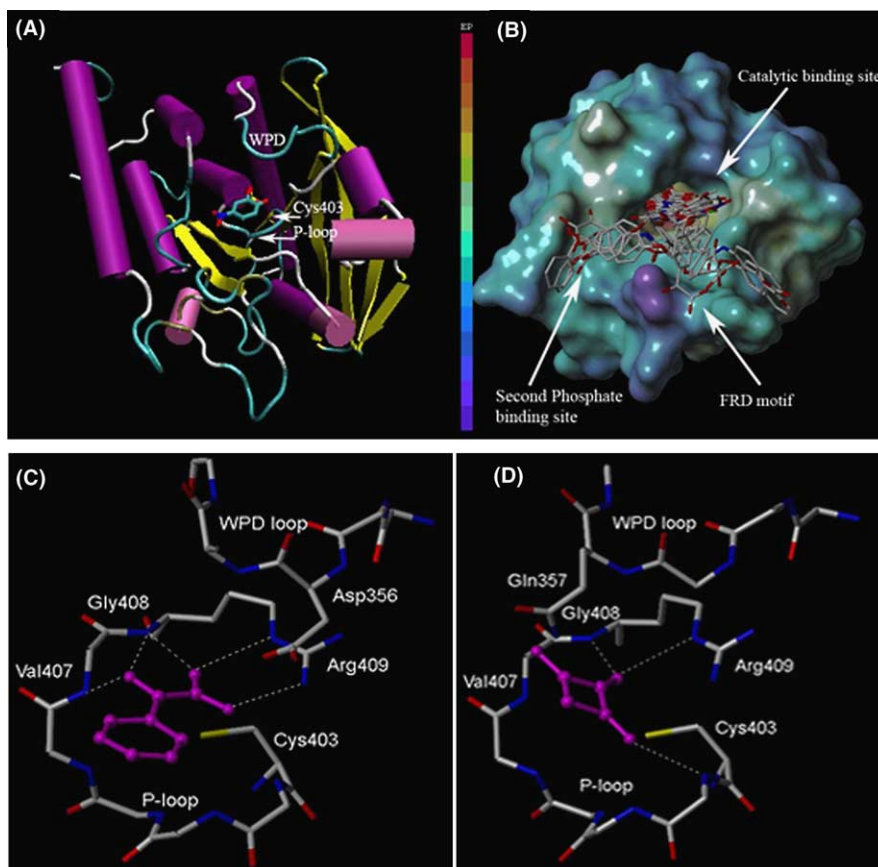
improvement with the CoMFA models, most probably because the samples in each data set are insufficient for QSAR analysis. For compound C17, the CoMFA

model cannot explain the sharp increase in activity observed when simply replacing the phenyl ring with furan. Omission of this outlier increased  $q^2$  of the CoMFA model to 0.829 (SEP = 0.592) and  $r^2$  to 0.968 (SE = 0.257).

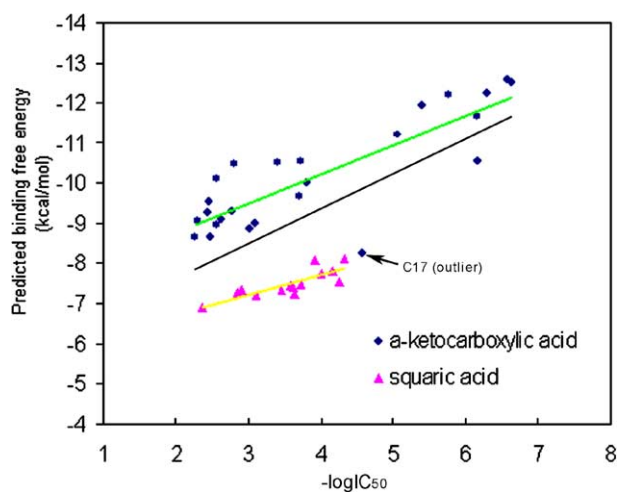
The 3D-CoMFA steric contour plot is shown in Figure 3A. Compound C1 is displayed in the map to assist in illustration. The CoMFA steric contours indicate that areas in which steric bulk substituents might have a favorable (green) or disfavorable (yellow) effect on the activity of an inhibitor. Sterically favorable green contours appear around the phenyl ring of compound C1, suggesting that bulky groups in this region are favored for higher activity. This also can be seen with compounds C6 and C9, in which introducing the bulky group of 4-methoxy, or replacing the phenyl ring with naphthyl ring, result in a 10-fold increase of activity as compared to C1. In contrast, there is not much of an increase in activity for compound C5, which bears an additional methoxy group at the 3-position and overlaps with the sterically disfavorable yellow region. In fact, most of sterically disfavorable regions are found surrounding the phosphate mimetics of  $\alpha$ -ketocarboxylic acid and squaric acid groups, which define a rather restricted pocket of the active site of YopH. It is noteworthy that there is an additional sterically favorable green region in the cavity of active site, indicating a possibility of further structural modification of the functional phosphate mimetics to improve the inhibitory activity.

Figure 3B shows the electrostatic contour map of the CoMFA model in combination with compound C21. The blue contours represent the regions where positively charged groups enhance the activity and red contours indicate regions where the negatively charged groups increase the activity. An important feature of the CoMFA model is that the electrostatic contour map is dominated by regions favorable to negative charges. Such a region is mostly observed surrounding the  $\alpha$ -ketocarboxylic acid and squaric acid functional mimetics in the catalytic binding cavity, indicating that compounds with high electron density are preferred in the active binding pocket. This is consistent with the fact that most of mimetics used in PTP inhibitors are negatively charged. In addition, there is another red contour region contributed to the proximal  $\alpha$ -ketocarboxylate group of compound C21. The only blue contour region favorable to positive charges is found in the vicinity of the phosphate mimetics  $\alpha$ -ketocarboxylate and squaric group, located at the bottom of the active binding pocket.

**2.2.2. CoMSIA analysis.** CoMSIA analysis using steric, electrostatic, and lipophilic fields as descriptors gave a model with  $q^2$  of 0.721 and  $r^2$  of 0.922. Omission of outlier C17 resulted in a more robust model with  $q^2$  of 0.837 and  $r^2$  of 0.973 (Table 3). The CoMSIA steric, electrostatic and lipophilic fields explain variance of 29%, 37%, 34%, respectively. This indicates that the hydrophobic interaction is necessary to fully describe the field properties of YopH inhibitors. Further attempts to combine the hydrogen bond fields with the standard



**Figure 1.** (A) the X-ray structure of YopH in complex with ligand pCNS (PDB entry 1PA9). (B) binding conformations of docked compounds at the active site of YopH. The molecular surface was mapped by lipophilicity. (C) interactions of  $\alpha$ -ketocarboxylic acid (magenta) at the active site of YopH. (D) interactions of squaric acid (magenta) at the active site of YopH.



**Figure 2.** Correlations between the experimental inhibitory activities ( $-\log IC_{50}$ ) and the predicted binding free energies (kcal/mol) by AutoDock.

steric, electrostatic, and lipophilic fields did not lead to any significant improvement ( $q^2 = 0.836$ ,  $r^2 = 0.978$ ). Indeed, since hydrophobic and electrostatic interactions dominate the inhibitor binding in YopH, the number of correlations between the activity and the hydrogen bond field is expected to be low.

Figure 4 illustrates the CoMSIA contour maps of steric, electrostatic, and hydrophobic fields. The steric field distribution of the CoMSIA model is in agreement with the field distribution of the CoMFA model. The hydrophobic contours show a large cyan region around the  $\alpha$ -ketocarboxy phenyl ring of compound C22, indicating that hydrophobic substituents in this area are likely to lead to compounds with enhanced activity (Fig. 4A). The hydrophobic favorable region overlaps the steric region. Indeed, many bulky groups in YopH inhibitors are hydrophobic. Hydrophobic disfavored white regions are embedded in the sterically disfavored yellow regions. The electrostatic field distribution of the CoMSIA model is also consistent with the CoMFA model. Negative charge favored regions are dominant at the site of the catalytic binding pocket and the proximal site of FDR motif that contribute to the  $\alpha$ -ketocarboxylate groups of compound C21. Interestingly, a red contour region (favorable to negative charges) is present in the CoMSIA model, mapping to the second phosphate binding site of YopH (Fig. 4B).

The CoMSIA model was therefore superimposed with the 3D topology of YopH structure to gain further insight into the properties of the different fields. As shown in Figure 5, the field contour maps correlate well with the environmental characteristics of the active site of YopH. The steric and hydrophobic favored regions lie



**Table 2.** Summary of CoMFA analysis results

	Model A (35 training compds)	Model B (34 training compds, omission of C17)
$q^{2a}$	0.734	0.829
$N^b$	3	4
$S_{cv}^c$	0.693	0.592
$r^{2d}$	0.914	0.968
$S^e$	0.414	0.257
$F^f$	61.625	167.016
Steric	60.3	61.0
Electrostatic	39.7	39.0
$r_{bs}^{2g}$	0.952	0.979
$S_{bs}^h$	0.027	0.006

<sup>a</sup> Cross-validation correlation coefficient.<sup>b</sup> Number of components.<sup>c</sup> Standard error of prediction.<sup>d</sup> Correlation coefficient.<sup>e</sup> Standard error of estimate.<sup>f</sup> *F*-ratio.<sup>g</sup> Bootstrapped correlation coefficient.<sup>h</sup> Bootstrapped standard deviation.

in the entrance of the catalytic binding pocket, which harbors the hydrophobic residues Phe229, Ile232, Val407, Arg404, and Ile443. The electrostatic contours show the importance and orientation of the phosphate mimetic functional group in the catalytic binding pocket. The negative charge favored regions that are mapped onto the other two proximal binding sites are consistent with the docking results, suggesting that simultaneous binding to a secondary site might substantially improve the inhibitory activity. Therefore, the 3D-QSAR model establishes the molecular basis for the inhibition of YopH, providing a guide to the design and improvement of drug candidates.

**2.2.3. Validation of the 3D-QSAR models.** Four compounds that were not included in the training set were selected as test data set to validate the QSAR

models. The predicted inhibitory activities using the CoMFA and CoMSIA models are shown in Table 4. All of the test compounds are well predicted. The mean and standard deviation of prediction errors are 0.58 and 0.10 for CoMFA model (model B using 34 training data), and only 0.33 and 0.03 for CoMSIA model (model D using three fields and 34 training data). The predictive  $r_{pred}^2$ , which is analogous to cross-validated correlation coefficient  $q^2$ , is 0.983 for the CoMFA and 0.995 for the CoMSIA, suggesting a high reliability for these models.

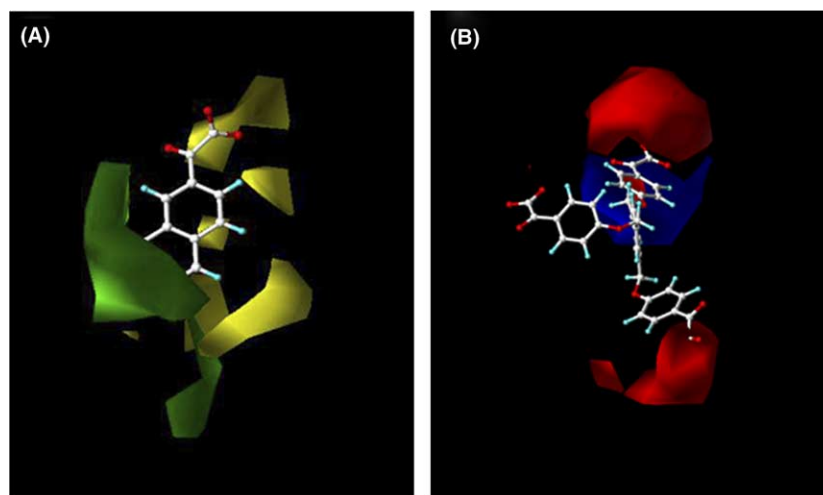
### 3. Conclusions

Using the alignment scheme generated from the docking study, highly predictive 3D-QSAR models were developed on *Yersinia* protein tyrosine phosphatase YopH inhibitors. The best prediction was obtained by CoMSIA model with  $q^2$  of 0.837 and  $r^2$  of 0.978. The models were validated with high reliability in the prediction of inhibitory activities. The consistency between the CoMFA/CoMSIA field distributions and the 3D topology of the protein structure shows the robustness of the 3D-QSAR models. These results, together with the good correlations between the inhibitory activities and the binding free energies predicated by AutoDock, demonstrate the power of combined docking/QSAR approach to explore the probable binding conformations of compounds at the active site of the protein target, and further develop reliable quantitative models for rational drug design.

### 4. Experimental

#### 4.1. Data set

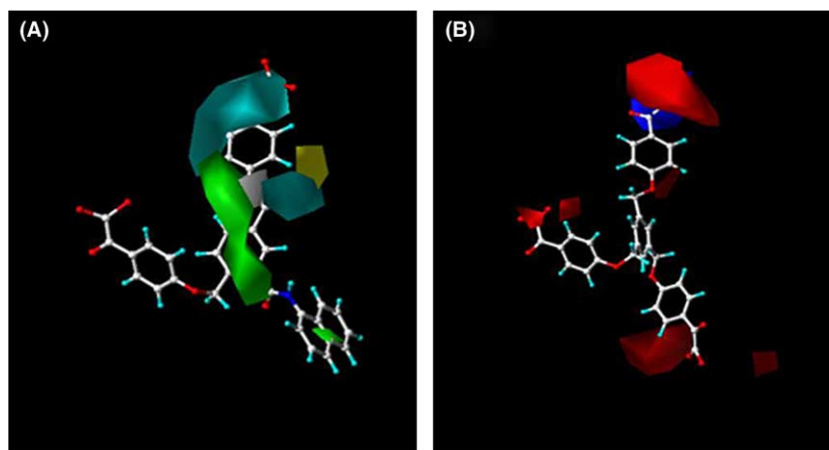
A total of 39  $\alpha$ -ketocarboxylic acid and squaric acid YopH inhibitors were used in this study. The structures



**Figure 3.** CoMFA StDev \* Coeff contour plots. (A) steric contour map displayed with compound C1. Green indicates regions where bulky groups increase activity; yellow indicates regions where bulky groups decrease activity. (B) electrostatic contour map displayed with compound C21. Red indicates regions where negative charges increase activity; blue indicates regions where positive charges increase activity.

**Table 3.** Summary of CoMSIA analysis results

	Three fields		Five fields	
	Model C (35 compds)	Model D (34 compds)	Model E (35 compds)	Model F (34 compds)
$q^{2a}$	0.721	0.837	0.754	0.836
$N^b$	3	4	3	4
$S_{cv}^c$	0.720	0.572	0.677	0.575
$r^{2d}$	0.922	0.973	0.950	0.978
$S^e$	0.394	0.236	0.317	0.213
$F^f$	68.656	199.841	109.428	246.995
Steric	29.0	30.2	17.5	20.6
Electrostatic	36.8	35.2	20.7	22.1
Hydrophobic	34.2	34.6	19.4	21.2
Donor	/	/	25.3	21.5
Acceptor	/	/	17.1	14.7
$r_{bs}^{2g}$	0.955	0.983	0.979	0.986
$S_{bs}^h$	0.026	0.008	0.013	0.009

<sup>a</sup> Cross-validation correlation coefficient.<sup>b</sup> Number of components.<sup>c</sup> Standard error of prediction.<sup>d</sup> Correlation coefficient.<sup>e</sup> Standard error of estimate.<sup>f</sup> *F*-ratio.<sup>g</sup> Bootstrapped correlation coefficient.<sup>h</sup> Bootstrapped standard deviation.

**Figure 4.** CoMSIA StDev \* Coeff contour plots. (A) steric (green/yellow) and hydrophobic (cyan/white) contour map combined with compound C22. Green indicates regions where bulky groups increase activity; yellow indicates regions where bulky groups decrease activity. Cyan indicates regions where hydrophobic substituents enhance activity; white indicates regions where hydrophobic substituents decrease activity. (B) electrostatic contour map (red/blue) in combination with compound C21. Red indicates regions where negative charges increase activity; blue indicates regions where positive charges increase activity.

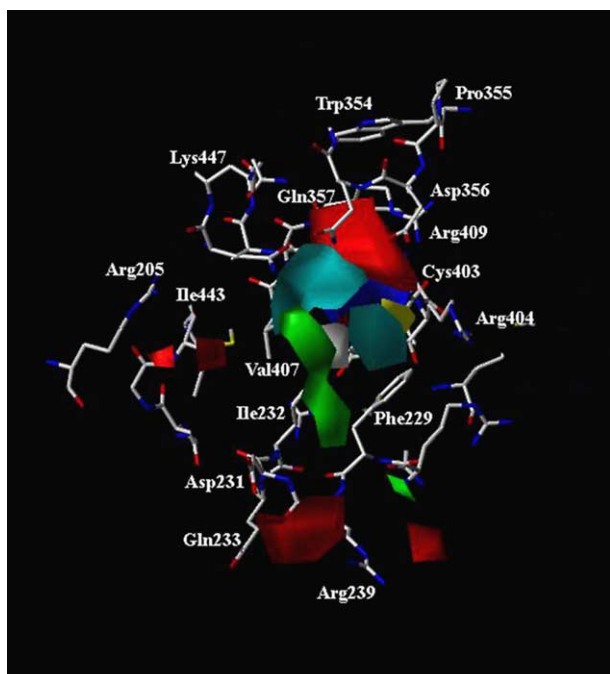
and their inhibitory activities are shown in Table 1. The inhibitory activities for YopH were measured under similar experimental conditions.<sup>14</sup> The  $IC_{50}$  values were converted to the corresponding  $pIC_{50}$  ( $-\log IC_{50}$ ) and used as dependent variables in the CoMFA and CoMSIA analysis. The  $pIC_{50}$  values span a range of 4-log units, providing a broad and homogeneous data set for 3D-QSAR study. Thirty-five compounds were used as training set and four compounds were used as test set. The test compounds were selected manually such that the structural diversity and wide range of activity in the data set were included.

#### 4.2. Molecular docking

The initial structures of the 39 YopH inhibitors were constructed using the SYBYL 6.9 software.<sup>24</sup> The geome-

tries of these compounds were subsequently optimized using the Tripos force field. The Powell method was used for energy minimization with an energy convergence gradient value of 0.001 kcal/(mol Å). The structure of YopH protein (PDB code 1PA9)<sup>18</sup> was obtained from the Protein Data Bank.<sup>26</sup> The hetero atoms (cofactors, water molecules, and ligands) were removed and polar hydrogens were added. The Kollman unit-atom charges were assigned to protein atoms using SYBYL 6.9.

The automated molecular docking calculations were carried out using AutoDock 3.0.<sup>22</sup> AM1-BCC charges were calculated on the inhibitors employing the antechamber module of Amber 8.0.<sup>25</sup> The AUTOTORS module of AutoDock defined the active torsions for each docked compound. The active site of the protein was defined using AutoGrid. The grid size was set to



**Figure 5.** CoMSIA steric (green/yellow), electrostatic (red/blue), and hydrophobic (cyan/white) fields projected in the active site of YopH.

**Table 4.** Observed versus predicted activities of test compounds using CoMFA and CoMSIA models

Compounds	Observed	CoMFA (model B)	CoMSIA (model D)
C7	3.10	3.32	3.05
C24	6.17	6.26	6.29
C27	4.25	4.01	4.31
C38	2.92	2.95	2.82
Mean <sup>a</sup>	/	0.58	0.33
SD <sup>a</sup>	/	0.10	0.03
$r^2_{\text{pred}}$ <sup>b</sup>	/	0.983	0.995

<sup>a</sup> Mean and standard deviation (SD) of prediction standard error ( $\text{pIC}_{50}^{\text{pred}} - \text{pIC}_{50}^{\text{obsd}}$ ).

<sup>b</sup>  $r^2_{\text{pred}} = 1 - \sum (\text{pIC}_{50}^{\text{obsd}} - \text{pIC}_{50}^{\text{pred}})^2 / \sum (\text{pIC}_{50}^{\text{obsd}} - \text{pIC}_{50}^{\text{mean}})^2$ .

$70 \times 70 \times 70$  points with a grid spacing of  $0.375 \text{ \AA}$  centered at the center of mass of the ligand. The Lamarckian Genetic Algorithm (LGA) was used as search method. Each LGA job consisted of 50 runs, and the number of generation in each run was 27,000 with an initial population of 50 individuals. The maximum number of energy evaluations was set to 1,000,000. Operator weights for cross-over, mutation, and elitism were 0.80, 0.02, and 1, respectively. The docking results were clustered by less than  $1.5 \text{ \AA}$  in positional root-mean-square deviation (RMSD), and the most favorable binding conformation with the lowest free energies was selected within the top-ranked cluster.

After docking, the complexes were energetically minimized permitting only the inhibitor and the side-chain atoms of the protein to relax. Energy minimizations were carried out employing SYBYL 6.9. In each step, MMFF94 force field was applied with  $0.05 \text{ kcal/\AA}$  con-

vergence and 5000 steps using the Powell method. The refined binding conformations of compounds were used for the 3D-QSAR studies.

### 4.3. CoMFA analysis

The CoMFA steric and electrostatic fields were calculated at grid lattice points with grid spacing ( $2.0 \text{ \AA}$ ) using the Lennard–Jones and the Coulomb potential functions of the Tripos force field. The default  $\text{sp}^3$  carbon atom was used as the steric probe atom and a  $+1.00$  charge for the electrostatic probe. The energy cutoff of  $30 \text{ kcal/mol}$  was adopted and the steric and electrostatic fields were scaled by the CoMFA standard method in SYBYL. The regression analysis was carried out using the full cross-validated partial least squares (PLS) method (leave-one-out). The minimum column filtering value was set to  $3.0 \text{ kcal/mol}$  to improve the signal-to-noise ratio by omitting those lattice points whose energy variation was below this threshold. The optimal number of components was selected to yield the highest cross-validated coefficient ( $q^2$ ) values and was set to 5. The cross-validated coefficient  $q^2$  was calculated using formula:

$$q^2 = 1 - \sum (\text{pIC}_{50}^{\text{obsd}} - \text{pIC}_{50}^{\text{pred}})^2 / \sum (\text{pIC}_{50}^{\text{obsd}} - \text{pIC}_{50}^{\text{mean}})^2$$

where obsd, pred, and mean refer to the observed, predicted, and mean values of the  $\text{pIC}_{50}$ , respectively.

### 4.4. CoMSIA analysis

As defined in SYBYL, the five CoMSIA similarity index fields (steric, electrostatic, lipophilic, hydrogen bond donor, and hydrogen bond acceptor) were evaluated at grid lattice point using a common probe atom of  $1 \text{ \AA}$  radius, as well as the charge, hydrophobicity, and hydrogen bond properties of  $+1$ . The partial least squares (PLS) method was used to derive a linear relationship, and cross-validation was performed using the leave-one-out method with a  $3.0 \text{ kcal/mol}$  column filter. CoMSIA similarity indices ( $A_{F,K}$ ) for a molecule  $j$  with atoms  $i$  at a grid point  $q$  were computed as follow,

$$A_{F,K}^q(j) = - \sum_{i=1}^n W_{\text{probe},k} W_{ik} e^{-\alpha r_{iq}^2}$$

where  $W_{ik}$  is the actual value of the physicochemical property  $k$  of atom  $i$ , and  $W_{\text{probe},k}$  is the value of the probe atom. A Gaussian type distance dependence was used between the grid point  $q$  and each atom  $i$  of the molecule, where  $r$  represents the distance. The default value of  $0.3$  was used as the attenuation factor ( $\alpha$ ).

### Acknowledgements

We thank R. Bennett, C. Pepper, A. Gazes, and G. Latter of the Rockefeller University Information Technology Resource Center for computational facilities. This work was funded in part by program grant 1U19AI056510-01 from the National Institute of Allergy and Infectious Disease.



## References and notes

1. Brubaker, R. R. *Clin. Microbiol. Rev.* **1991**, *4*, 309.
2. Cornelis, G. R. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 742.
3. Ramamurthi, K. S.; Schneewind, O. *Annu. Rev. Cell Dev. Biol.* **2002**, *18*, 107.
4. Guan, K. L.; Dixon, J. E. *Science* **1990**, *249*, 553.
5. Bliska, J. B.; Guan, K. L.; Dixon, J. E.; Falkow, S. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 1187.
6. Black, D. S.; Bliska, J. B. *Embo J.* **1997**, *16*, 2730.
7. Juris, S. J.; Shao, F.; Dixon, J. E. *Cell Microbiol.* **2002**, *4*, 201.
8. Bliska, J. B. *Trends Microbiol.* **2000**, *8*, 205.
9. Harley, E. A.; Levens, N. *Curr. Opin. Invest. Drugs* **2003**, *4*, 1179.
10. Zhang, Z. Y. *Curr. Opin. Chem. Biol.* **2001**, *5*, 416.
11. Inglesby, T. V.; Dennis, D. T.; Henderson, D. A.; Bartlett, J. G.; Ascher, M. S. *Jama* **2000**, *283*, 2281.
12. Cohen, M. L. *Nature* **2000**, *406*, 762.
13. Hawley, R. J.; Eitzen, E. M., Jr. *Annu. Rev. Microbiol.* **2001**, *55*, 235.
14. Chen, Y. T.; Onaran, M. B.; Doss, C. J.; Seto, C. T. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1935.
15. Chen, Y. T.; Seto, C. T. *J. Med. Chem.* **2002**, *45*, 3946.
16. Lee, K.; Gao, Y.; Yao, Z. J.; Phan, J.; Wu, L., et al. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2577.
17. Liang, F.; Huang, Z.; Lee, S. Y.; Liang, J.; Ivanov, M. I., et al. *J. Biol. Chem.* **2003**, *278*, 41734.
18. Sun, J. P.; Wu, L.; Fedorov, A. A.; Almo, S. C.; Zhang, Z. Y. *J. Biol. Chem.* **2003**, *278*, 33392.
19. Chen, Y. T.; Seto, C. T. *Bioorg. Med. Chem.* **2004**, *12*, 3289.
20. Xie, J.; Comeau, A. B.; Seto, C. T. *Org. Lett.* **2004**, *6*, 83.
21. Hu, X.; Vujanac, M.; Stebbins, C. E. *J. Mol. Graph. Model* **2004**, *23*, 175.
22. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639.
23. Hu, X.; Balaz, S.; Shelver, W. H. *J. Mol. Graph. Model* **2004**, *22*, 293.
24. SYBYL Molecular Modeling Software, v6.9; Tripos Associates, St. Louis, MO.
25. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.
26. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.