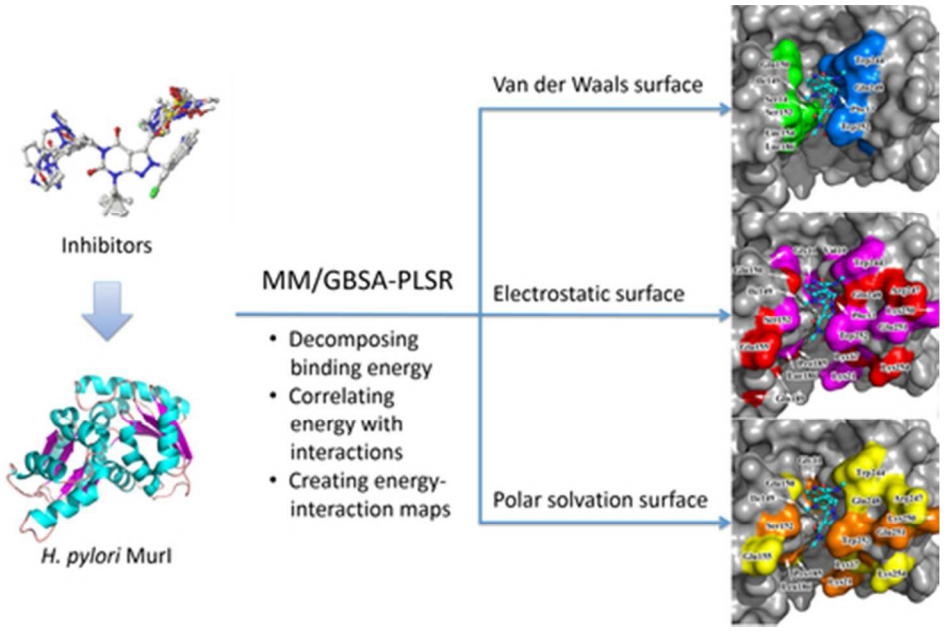
RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances



39x26mm (300 x 300 DPI)

Identifying MurI Uncompetitive Inhibitors by Correlating Decomposed Binding Energies with Bioactivity

Xiu Le, Qiong Gu* and Jun Xu*

Research Center for Drug Discovery, School of Pharmaceutical Sciences, Sun Yat-Sen University, Guangzhou 510006, China

Abstract

The glutamate racemase (MurI) is essential for *Helicobacter pylori* (*H. pylori*) cell wall biosynthesis. In this work, we report a new method that correlates decomposed binding free energies with MurI inhibition based upon the data from pyrazolopyrimidinedione series MurI uncompetitive inhibitors. With the molecular mechanics/generalized Born surface areas (MM/GBSA) approach, we were able to decompose the binding interaction into van der Waals, electrostatic, and polar solvation surfaces. The decomposed binding energies were correlated with MurI inhibitory activity with partial least squares regression (PLSR). Hence, the method is termed as MM/GBSA-PLSR. The non-cross-validation (R^2) and leave-one-out cross-validation (LOOCV) (Q^2) correlation coefficients of the 3D-QSAR model are 0.962 and 0.822, respectively. The external testing yields a predicted correlation coefficient (R^2_{pred}) of 0.817. This study demonstrated that the activity-contribution fractions from the three types of ligand-receptor interactions are 29.5% from van der Waal interactions, 38.2% from electrostatic interactions, and 32.3% from polar solvation interactions. Comparing with molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA), we find that the CoMFA/CoMSIA steric interaction fields can be interpreted as the MM/GBSA-PLSR

van der Waals interactions; CoMFA/CoMSIA electrostatic and H-bond acceptor/donor interaction fields can be interpreted as the MM/GBSA-PLSR electrostatic interactions. However, there is no explicit association between MM/GBSA-PLSR solvation interactions (polar or non-polar) and CoMFA/CoMSIA fields. It is worth to note that the solvation interaction is important for ligand design. Moreover, MM/GBSA-PLSR map the decomposed binding interactions on to pharmacophore surfaces (van der Waals, electrostatic, and polar solvation surfaces) to aid drug design.

Keywords: *Helicobacter pylori*; MurI; structure-based QSAR; MM/GBSA-PLSR; CoMFA; CoMSIA; pharmacophore surface.

1. Introduction

The human pathogen *Helicobacter pylori* (*H. pylori*) is a key cause of gastric inflammation and cancer. *H. pylori*-induced gastric inflammation does not cause symptoms in most infected people but is associated with an increased risk of developing duodenal ulcer disease, gastric ulcer disease, gastric adenocarcinoma, and gastric lymphoma.[1-5] Approximately 50% of the world's population suffers from *H. pylori* infection.[6] Currently, a cocktail therapy consisting of a proton pump inhibitor (e.g. omeprazole) and two broad-spectrum antibiotics (i.e. clarithromycin and amoxicillin) is most often used for the treatment of *H. pylori* infections; it is given over a one-week period.[7, 8] However, the treatment success is compromised by poor patient compliance due to diarrhea and other side effects resulting from the suppression of commensal bacteria. Additionally, *H. pylori* resistance to current therapies prompts the need for an alternative therapy with a new mode-of-action (MOA).[6, 9, 10]

Glutamate racemase (MurI) is a bacterial cytoplasmic enzyme that catalyzes the conversion of L-glutamate to D-glutamate, one of the essential amino acids in peptidoglycan synthesis.[11-13] Deletion of MurI prevents peptidoglycan construction and bacterial viability by disrupting the supply of D-glutamate.[14, 15] Therefore, MurI represents a promising target for the design of antibacterial drugs.[16] Glutamate analogs were reported to be competitive inhibitors that bound at the active site[17] of MurI and showed potent antibacterial activity.[18] However, it was not until AstraZeneca identified a series of uncompetitive inhibitors via high-throughput screening (HTS) that specifically bind to a cryptic allosteric site of MurI, the structural, kinetic and mutational studies of uncompetitive inhibitors emerged.[19]

The co-crystal structure of MurI and D-glutamate indicates that the MurI inhibitor can occupy an allosteric binding site that resides away from the substrate. The C-terminal helix movement induces Trp252 side-chain displacement and rotation to form a surface for π-stacking with the pyrazolopyrimidinedione core of the inhibitors among which compound **1** shows the best inhibitory activity at 6 nM (Figure 1). Sequence analyses of diverse, clinically relevant *H. pylori* isolates revealed that almost all the interacting residues in this binding pocket were conserved, demonstrating the suitability of the site for *H. pylori* MurI inhibition. Site-specific mutagenesis highlighted the importance of maintaining the interactions with these residues, which include Val10, Gly11, His183, Leu186, Glu150, Ser152, and Trp244.[19]

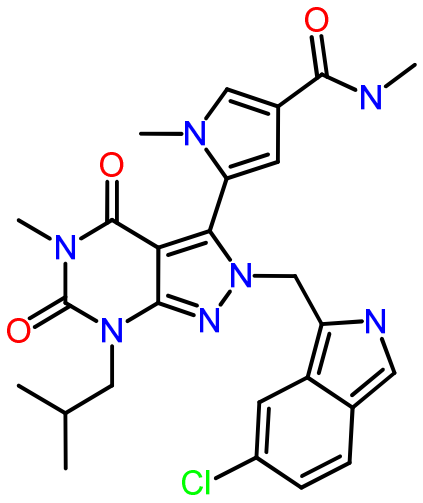


Figure 1. Chemical structure of the MurI uncompetitive inhibitor, Compound **1**.

On the basis of the MurI co-crystal structure, a number of studies were conducted on the competitive inhibitors, which utilized structure-based methods[20], HTS[19], docking virtual screening (VS)[21], and quantitative structure-activity relationship (QSAR) studies[18]. However, these results have proven problematic due to the flexibility of the enzyme and

species-specific hydrophobic pocket proximal to the active site. Therefore, the discovery of uncompetitive inhibitors is attractive. So far, there is only a type of SAR study towards molecular modeling of the MurI uncompetitive inhibitors that has been conducted on the aforementioned series of pyrazolopyrimidinedione derivatives.[22-26]

Empirical correlations between affinities and a set of physicochemical descriptors of a series of ligands have long been used in drug design, and extension to the three-dimensional properties of the ligands, namely 3D-QSAR, have proven greatly successful.[27-30] Comparative molecular field analysis (CoMFA)[31] and comparative molecular similarity index analysis (CoMSIA)[32] are two ligand-based 3D-QSAR methods among those which do not use structural data regarding the receptor. To compensate the underlying adverse impact irrespective of receptor, conformational alignment is required, whether based on the maximum common substructure (MCS), or fields (e.g. Surflex-Sim's morphological similarity[33], OpenEye's shape[34] and electrostatic[35] fields or Cresset's XED[36] force field), or using other methods (e.g. MOE's Flexible Alignment[37]). But this would still lead to excellent but unreliable statistical results. In contrast, structure-based (i.e. receptor-based) 3D-QSAR approaches modeling receptor-ligand interactions rely on receptor conformation data and receptor-ligand interaction calculations, which would effectively overcome such a problem. The molecular mechanics/generalized Born surface areas (MM/GBSA) free-energy calculation has been successfully used in structure-based studies.[38-41] Herein, we explore a new structure-based 3D-QSAR approach, which employs partial least squares regression (PLSR)[42] to correlate the decomposed binding free energies calculated from MM/GBSA with the MurI uncompetitive inhibitory activity. The 3D-QSAR approach is termed MM/GBSA-

PLSR; it takes structural information on receptor-ligand interactions and thus induced-fit effects as well as solvent effects into account.

To elucidate the model derived from the MM/GBSA-PLSR approach, we created ligand-based 3D-QSAR models with CoMFA and CoMSIA. By referencing the ligand-based 3D-QSAR models, we attempt to reveal the relationships between the MM/GBSA interactions and the CoMFA/CoMSIA fields to describe MurI uncompetitive inhibitor activity at the residue level. Our goal is to find a rational and efficient method for designing or optimizing MurI uncompetitive inhibitors.

2. Methods

2.1 Compounds and biological data

By an exhaustive literature search, a total of 69 pyrazolopyrimidinediones as potent MurI inhibitors were collected from literatures[25, 26] for modeling studies. This work focuses on describing the pharmacophore features of existing actives and it will be difficult to describe the pharmacophore features for the in-actives due to great structural diversity. Therefore, compounds used here are all active. The *in vitro* biological activities (i.e. IC_{50}) of these compounds were converted into the corresponding negative logarithmic values (pIC_{50}) and used as dependent variables for QSAR analyses. The structures and biological data, expressed in pIC_{50}, are listed in Table S1.

2.2 Ligand-receptor systems energy minimization

The co-crystal structure of MurI (PDB code: 2JFZ[19]; Resolution: 1.86 Å) containing compound **67** is used as the initial structure for construction of all ligand-receptor systems. The chain A and the substrate were preserved while water molecules were

removed, and missed residues were repaired by using the *Clean Protein* tool in Discovery Studio 3.5[43]. All compounds were then superimposed to the conformation of compound **67**. Each system's energy was minimized within the protein pocket using MOE 2013[37] with the built-in Amber12EHT force field (an all-atom force field combining 2D Extended Hueckel Theory[44] and Amber ff12[45]). The 69 aligned compounds are depicted in Figure 2.

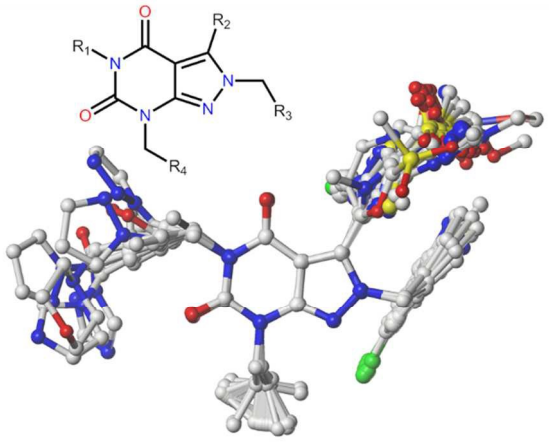


Figure 2. 69 compounds were aligned on the core structure (top left).

For construction of MM/GBSA models, an explicit solvent minimization protocol is required. It was applied within the AMBER 12 package[45] with the FF12SB force field. Hydrogen atoms were added to the system with *tLEaP*. The geometries of the small molecules were completely optimized at HF/6-31G* level of theory with Gaussian 09 suite[46]. The electrostatically derived atomic charges were computed via the RESP[47] method. All complexes were neutralized by adding sodium ions and solvated in a periodic truncated octahedron box of TIP3P[48] molecules with a margin of 10 Å.

The solvated system was initially minimized to remove bad van der Waals contacts via five stages, employing the steepest descent algorithm (800, 1000, 2000, 5000, and 0 cycles, respectively) and the conjugate gradient algorithm (1200, 2000, 3000, 5000, and 10000 cycles, respectively) with a non-bonded cutoff of 10.0 Å. The system incorporated gradually reduced positional restraints with force constants of 10.0, 5.0, 1.0, 1.0, and 0 kcal mol^{-1} $Å^{-2}$, respectively. In the first stage, all atoms were restrained except the solvent such that the added TIP3P water molecules could adjust their orientation. In the second and third stages, the protein backbone and the key residues Glu150, Leu186, Trp244, and Gln248 were restrained while the amino acid side-chains were allowed to move, which allowed the ligand to achieve a lower-energy position. In the fourth stage, a weak restraint potential was imposed only on Gln248 due to its relatively large flexibility identified in our previous molecular dynamic (MD) simulations (unpublished), and in the final stage, the whole system was fully minimized.

2.3 Binding free energy calculations

Averaging over snapshots during the MD trajectory is often required to improve binding free energy estimation, but this is not always the case.[49-52] The binding free energy was computed using the final snapshots of the energy minimization to reduce computational complexity. The binding free energy change is computed via Equation 1:

$$\Delta G_{bind} = G_{complex} - (G_{free\text{-}protein} + G_{free\text{-}ligand})$$

$$= \Delta G_{MM} + \Delta G_{sol} - T\Delta S \quad \textbf{(1)}$$

In a molecular mechanics system, the energy consists of electrostatic and van der Waals interaction terms:

$$\Delta G_{MM} = \Delta G_{ele} + \Delta G_{vdw} \quad \textbf{(2)}$$

The solvation free energy consists of the polar and nonpolar terms:

$$\Delta G_{sol} = \Delta G_{ele,sol} + \Delta G_{nonpol,sol} \quad \textbf{(3)}$$

where $\Delta G_{ele,sol}$ is obtained by solving the Poisson-Boltzmann (PB) equation or the generalized Born (GB) equation. $\Delta G_{nonpol,sol}$ is calculated via Equation 4:

$$\Delta G_{nonpol,sol} = \gamma SASA + b \quad \textbf{(4)}$$

where γ represents the surface tension, and b is a constant (0.0072 or 0 kcal mol^{-1} $Å^{-2}$). SASA is the solvent-accessible surface area ($Å^2$) determined via a linear combination of pairwise overlapping models. The conformational entropy contributions (translation, rotation, and vibration) are neglected. This method using the GB equation is termed MM/GBSA, and the binding free energies were then decomposed to residue-wise energy terms as a basis for the construction of the MM/GBSA-PLSR model.

2.4 Building the MM/GBSA-PLSR model

PLSR constructs linear combinations of the original variables and has been used to predict the biological behavior of peptides and their analogs.[53, 54] It assumes that the binding free energies or inhibition constants (i.e. pK_i or pIC_{50}) measured in experiments can be correlated by PLSR with weighted theoretical interaction energy terms. All compounds were divided into a training set consisting of 50 compounds and a test set of 19 compounds based on the distribution of biological data and structural diversity. The data source for the MM/GBSA-PLSR model is a matrix encompassing the MM/GBSA energy terms for the optimized receptor-ligand structures as well as their biological

activities. The statistical method underlying analysis has been previously described.[55-57] The residue-wise interaction energy terms were acquired via MM/GBSA approach. The terms are the electrostatic, van der Waals, polar, and nonpolar solvation free energies. The multivariate PLSR[58] technique, implemented in the R[59] statistical package, was used to extract the relevant trends between the binding free energies and pIC_{50} values. A threshold of 0.0027 for standard-deviation-weighted PLSR coefficients (StDev*Coeff) was used to filter energy terms. The regression coefficient (R^2) was calculated as follows:

$$R^2 = \left[\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \langle\hat{y}\rangle)\right]^2 \Big/ \sum_{i=1}^{N}(y_i - \bar{y})^2 \sum_{i=1}^{N}(\hat{y}_i - \langle\hat{y}\rangle)^2 \quad \textbf{(5)}$$

where $\langle\hat{y}\rangle = \sum_{i=1}^{N}\hat{y} \Big/ N$.

The third step is to perform LOOCV and predict the dependent variable for certain complexes that were excluded during model derivation. This method is often used to check whether the derived correlation is spurious and to assess the robustness of the resulting statistical model. The performance of the model was quantified with the cross-validated correlation coefficient Q^2 (Equation 6) and the root-mean-square error of prediction (RMSEP, Equation 7):

$$Q^2 = 1 - \sum_{i=1}^{N}(\hat{y}_i - \langle\hat{y}\rangle) \Big/ \sum_{i=1}^{N}(y_i - \bar{y})^2 \quad \textbf{(6)}$$

$$RMSEP = \sqrt{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 / N} \quad \textbf{(7)}$$

where $\bar{y}$ is the average value of activities.

In addition, to validate whether the performance of the MM/GBSA-PLSR model is a result of chance correlation, 100 trials of Y-randomization of the experimental activity values were executed. It consists of repeating the calculation procedure several times after shuffling the Y vector randomly.

The residuals of the experimental and predicted activities of compounds in the test set as well as R^2_{pred} were used to measure the predictive capacity of the model.

2.5 Building CoMFA and CoMSIA models

The data were divided into a training set (56 compounds) for model generation and a test set (13 compounds) for model validation based on the same rule of thumb as the MM/GBSA-PLSR model but with different proportion of training set compounds to obtain the optimal results. The 3D-QSAR models were built using the program SYBYL-X 1.1[60]. All molecules were placed within a lattice of 1.0 Å with a 2.0 Å margin for each dimension. To construct a CoMFA model, a probe atom having the van der Waals properties of sp3 carbon and a charge of +1.0 was used to calculate the steric (Lennard-Jones 6-12 potential) and electrostatic (Coulombic potential) field energies. To construct a CoMSIA model, five similarity indices were computed, including steric contributions, electrostatics, hydrophobic, hydrogen-bonding donor, and hydrogen-bonding acceptor using a probe atom with 1.0 Å radius, +1.0 charge, +1.0 hydrophobicity, and +1.0 H-bond donor and acceptor property. In PLSR analysis, the LOOCV was employed to determine the ONC, and the final 3D-QSAR models of CoMFA and CoMSIA were derived from each non-cross-validated analysis with the ONC.

3. Results and discussion

3.1 Quality of the MM/GBSA-PLSR, CoMFA, and CoMSIA models

The statistical parameters of the MM/GBSA-PLSR, CoMFA, and CoMSIA models are listed in Table 1. For simplicity, the best CoMSIA model constituted by electrostatic and hydrogen-bonding receptor components is displayed only. The cross-validated PLSR analysis of the training set resulted in correlation coefficients Q^2 of 0.822, 0.684, and 0.687 with the optimal number of components (ONC) of 8, 6, and 12 for the three models. The non-cross-validated PLSR yields R^2 coefficients of 0.962, 0.937, and 0.955, respectively. All models obtained by the Y-randomization test have much lower values for R^2 and Q^2 statistics, which verifies that the high internal validation performance of the MM/GBSA-PLSR model is not due to a chance correlation or structural dependency of the training set (see Supplementary Information Table S2). For the MM/GBSA-PLSR model, the proportions of receptor-ligand interactions for van der Waals, electrostatic, and polar solvation interactions are 0.295, 0.382, and 0.323, respectively. For the CoMFA model, the contributions of steric and electrostatic interaction fields were 0.446 and 0.554. And for the CoMSIA model, the electrostatic and H-bond acceptor interaction fields provide 0.633 and 0.367 contributions to the model, respectively.

Table 1. Summary of the ligand-based and structure-based 3D-QSAR Models [a]

Statistics	Ligand-based		Structure-based
	CoMFA	CoMSIA	MM/GBSA-PLSR
Q^2	0.684	0.687	0.822
ONC	6	12	8
R^2	0.937	0.955	0.962
SEE	0.138	0.125	0.103

F	121.758	75.675	129.743
P	< 0.01	< 0.01	< 0.01
R^2_{pred}	0.561[b]	0.748	0.817
Contributions			
Steric	0.446		
Electrostatic	0.554	0.633	0.382
H-bond acceptor		0.367	
Van der Waals			0.295
Polar solvation			0.323

[a] Abbreviations used: Q^2, leave-one-out cross-validation (LOOCV) correlation coefficient; ONC, optimum number of principal components; R^2, non-cross-validation correlation coefficient; SEE, standard error of the estimate.

[b] R^2_{pred} for the test set without the outlier compound **68**.

The external testing set consisted of 19 compounds, which were predicted by the MM/GBSA-PLSR model and yielded a correlation coefficient R^2_{pred} of 0.817. Another external testing set consisted of 13 compounds, which were predicted by both the CoMFA and CoMSIA models, and R^2_{pred} values were 0.561 (CoMFA) and 0.748 (CoMSIA). Figure 3 depicts the correlations between the observed activities and the predicted activities for the training set and testing set. Apparently, the predicted and observed activities agree significantly except for an outlier (detailed discussion on the outlier seen in the Supplementary Information) from the CoMFA model. The observed pIC_{50} values, the predicted pIC_{50} values, and the residuals between them are listed in the Supplementary Information Table S3.

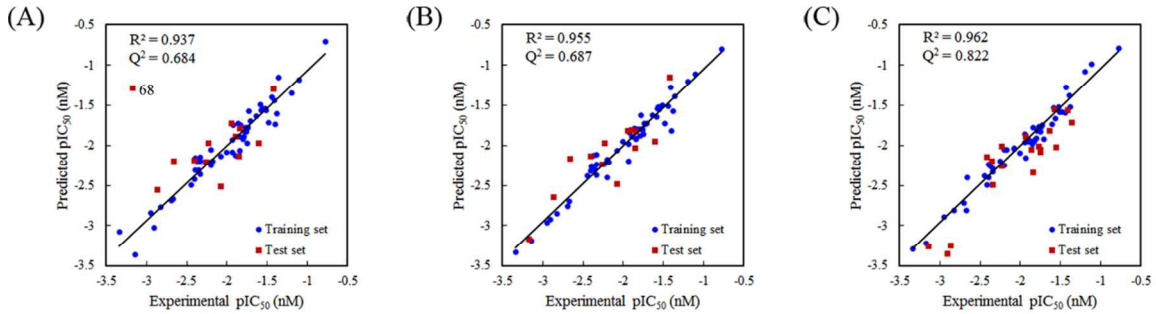


Figure 3. CoMFA (A), CoMSIA (B), and MM/GBSA-PLSR (C) predictions for the training sets (blue circle dots) and test sets (red square dots) regarding inhibitory activities against MurI. The solid line is the regression line for the training set predictions.

3.2 Mapping MM/GBSA-PLSR model and CoMFA model

StDev*Coeff is a quantitative index of relative contribution of the energy component to the inhibitory activity. A higher absolute value of the StDev*Coeff indicates a more crucial interaction in MurI inhibition. Notably, a negative coefficient corresponds to a favorable interaction and a positive coefficient corresponds to an unfavorable interaction. The key interaction energy components from the MM/GBSA-PLSR model include 14 van der Waals, 22 electrostatic, and 18 polar solvation interactions (Figure 4, Supplementary Information Table S4). By comparing the three models, we recognized that the van der Waals and electrostatic interactions generated from the MM/GBSA-PLSR model can be interpreted by the binding requirements demonstrated in the contour maps of the CoMFA and CoMSIA models (Figure 5). The MM/GBSA-PLSR model indicates that the van der Waals interactions at Trp244, Gln248, and Trp252 can improve binding affinities. For example, compound **32** is more active than compound **57** due to its additional nitrile group in the 1-methyl-1*H*-pyrrole moiety at the R_2 position providing more van der Waals interaction with the target. However, the van der Waals interactions

at Ile149, Glu150, Ser152, Leu154, and Leu186 are not favored to the binding affinities. For example, compound **52** is less potent than compound **17** because of the former has more van der Waals interactions with these residues by its substituent in the R_1 position. The green region close to Gln248 (side chain) and Trp252 (side chain) indicates a more bulky substituent is preferred (e.g. compounds **32** and **57**). The yellow regions close to Glu150 (backbone), Leu154 (side chain), Leu186, and Trp252 (side chain) indicate that a smaller substituent is preferred (e.g. compounds **52** and **17**). The observation, so far, suggests that the MM/GBSA van der Waals interactions correlate with the CoMFA steric interaction fields. A green polyhedron close to Glu150 (backbone) and Ser152 (side chain) suggests a steric contribution to the binding affinities, but it is unfavorable for van der Waals interactions according to the MM/GBSA-PLSR model. As a consequence, these results appear to be inconsistent. In reality, both van der Waals interaction of MM/GBSA-PLSR and steric field of CoFMA are Lennard-Jones potential. Hence, it is unsurprising that there is high correlation between both. However, the MM/GBSA-PLSR van der Waals interaction is calculated between atoms of ligand and residues while the CoMFA steric field is obtained between ligand atoms and a probe atom (usually sp3 carbon). This may be the substantial reason why in some region they do not agree. The different binding property in this region may also be due to the flexible conformation of the R_2 substituents. In modeling, we replace MD simulation with a simple energy minimization in order to reduce computational time, which may lead to a wrong conformational speculation of the R_2 group. To summarize, steric bulky group requirements (CoMFA) can be elucidated as van der Waals interactions (MM/GBSA-

PLSR). The smaller steric group requirements can be explained as a means of circumventing van der Waals repulsions (Figure 5A).

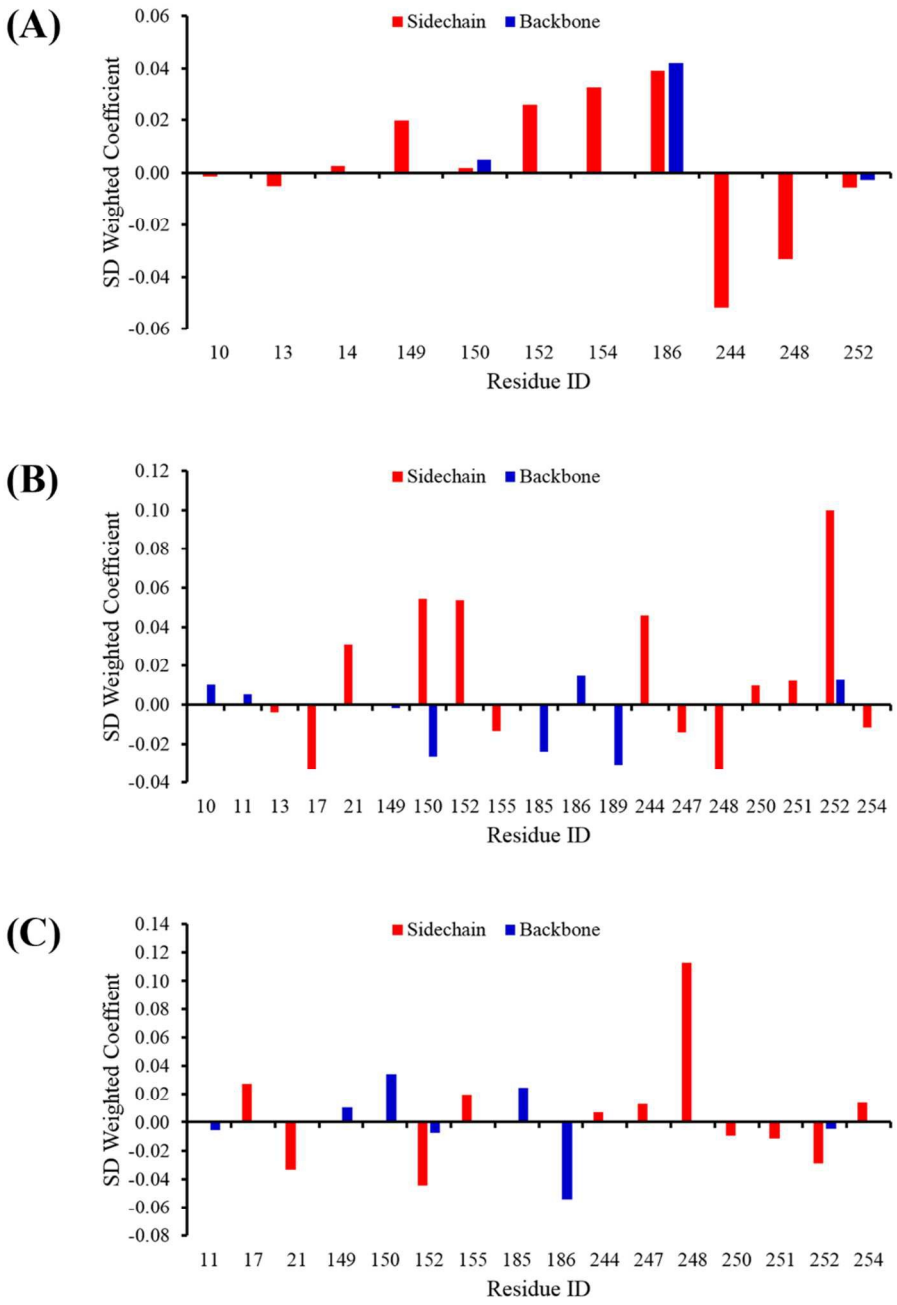


Figure 4. Partial least squares regression (PLSR) standard-deviation-weighted coefficients (StDev*Coeff) for interaction components in side-chain (red) and backbone

(blue) used in the MM/GBSA-PLSR model: A) van der Waals, B) electrostatic, and C) polar solvation interactions.

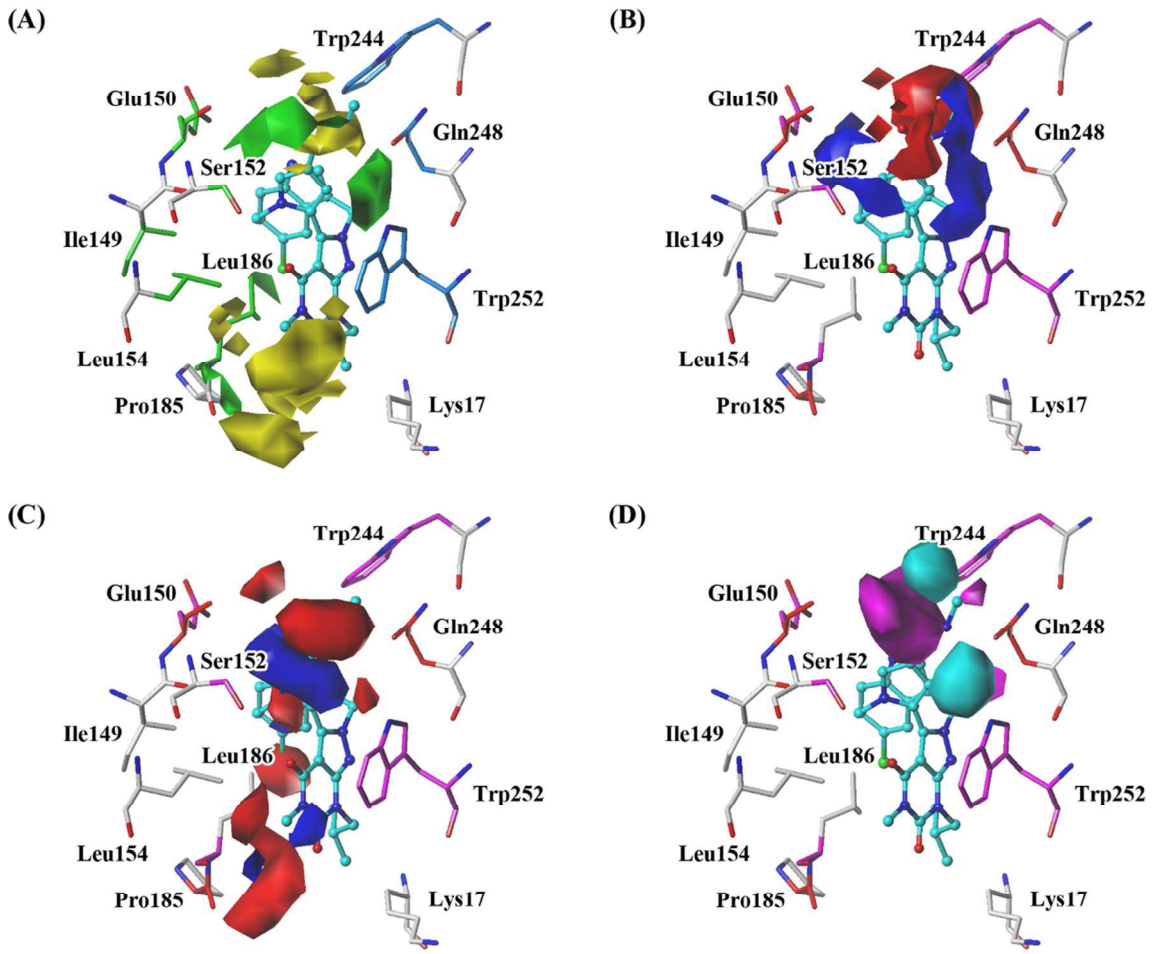


Figure 5. Correspondence between the key interactions identified by the MM/GBSA-PLSR model and the CoMFA and CoMSIA contour maps. Colored atoms indicate regions where: (A) van der Waals (marine or green) and (B, C, D) electrostatic (red or magenta) interactions are favorable or unfavorable. Polyhedra contour maps represent regions where: (A) more steric bulky (green) or less steric bulky (yellow) groups, (B, C) negative charge (red) or positive charge (blue) groups, and (D) groups having H-bond acceptor (magenta) or not (cyan) are preferred to enhance activity.

In a CoMFA electrostatic map, a red shape represents a negatively charged group increasing binding affinity and a blue shape represents positively charged group increasing binding affinity. However, the MM/GBSA-PLSR model does not recognize electrostatic types (positive or negative) since electrostatic interactions are calculated with different residue atoms rather than a unified +1.0 charged probe atom. Nevertheless, it can be recognized by residue types. Figure 5B demonstrates that negatively charged groups can increase binding affinities by interacting with Trp244 (side chain) and Gln248 (side chain), and positively charged groups can increase binding affinities by interacting with Ile149 (backbone), Glu150 (backbone), Ser152 (side chain), Gln248 (side chain), and Trp252 (side chain). The MM/GBSA-PLSR model supports favorable electrostatic interactions with Gln248 and Glu150 in agreement with the fact that Glu150 and Gln248 provide hydrogen bond acceptors. However, the electrostatic interactions do not always contribute to the binding affinities. Except for the controversial interaction with Trp252, the CoMFA electrostatic contour maps are apparently consistent with the fact that the indole ring of Trp244 is positively charged and Glu150, Ile149, Ser152, and Gln248 can provide electrons. However, the receptor-ligand interactions are not always independent of each other; increasing electrostatic interactions with Trp244 or Trp252 may reduce the electrostatic interactions with Gln248 because electrostatic type (positive charge) required for Gln248s is opposite to the electrostatic type of Trp244 or Trp252 (negative charge). In fact, several compounds (**13**, **59**, **29**, and **1**) are found to form hydrogen bonds with the amido oxygen atom of Gln248. But none are found to have hydrogen bonds with Trp244. And Trp255 is barely capable of forming hydrogen bonds, either. There is only one compound (compound **68**) that forms hydrogen bond with Trp255, but its activity is

low (IC_{50} = 1500 nM). Therefore, the CoMFA electrostatic fields derived simply based on ligands is not always in line with the actual circumstance while the MM/GBSA-PLSR model gives a more correct judgment. Nevertheless, they show a certain correspondence to each other both in theory and in practice.

3.3 Mapping MM/GBSA-PLSR model and CoMSIA model

A CoMSIA electrostatic map approximates the corresponding CoMFA map with slight differences. Negative electrostatic interactions are predominantly favorable in three regions, one of them close to Leu186, and the others close to the MM/GBSA-PLSR electrostatic-interaction favorable residues, Gln248 (Figure 5C). However, the amido oxygen atom of Leu186 does not have enough positive charge to attract an electron-donating group. The red polyhedron is close to Leu186 (side chain) and surrounded by Phe13, Ser14, Gly11, Thr182, and His183 (not shown in Figure 5). In this region, CoMSIA is unable to provide the correct structure activity relation. Observation and the MM/GBSA-PLSR model indicates that this region has a deep hydrophobic pocket and requires hydrophobic interactions instead of electrostatic interactions. On the other hand, the amido oxygen atoms of Glu150 (backbone) and Gln248 (side chain) are potential hydrogen-bond acceptors, but the CoMSIA model indicates that a hydrogen-bond acceptor is not allowed near Trp244 and Gln248 (Figure 5D). This is inconsistent with the prediction of the MM/GBSA-PLSR model. As shown in Figure 5D, the larger magenta polygon indicates that hydrogen-bond acceptor (HBA) between Glu150 and Trp244 may improve the activity. However, there is no evidence allowing a ligand to form a hydrogen bond at this point. This observation agrees with MM/GBSA-PLSR model, which proves that Ser152, Trp244, and Trp252 are not favorable for the formation

of hydrogen bonds in this region in order to enhance the activity. Therefore, the CoMSIA electrostatic and H-bond acceptor fields can be uniformly deciphered by the MM/GBSA-PLSR electrostatic interactions and MM/GBSA-PLSR model can identify CoMSIA models' defects.

3.4 Interpreting 3D-QSAR with the MM/GBSA-PLSR model

MM/GBSA-PLSR interaction components can be correlated to CoMFA and CoMSIA interaction fields. Consequently, the fields are mapped to the interacting sites of the receptor. The MM/GBSA-PLSR model results in three interacting factors contributing to the MurI inhibitory activity, i.e. van der Waals (29.5%), electrostatic (38.2%), and polar solvation (32.3%) interactions, which are elucidated as van der Waals and electrostatic interaction maps in Figure 6 and a polar solvation interaction map in Figure 7. A marine region formed by Val10, Phe13, Trp244, Gln248, and Trp252 suggests that increasing van der Waals interactions with those residues will improve binding affinity (Figure 6A and 5B). For example, compound **32** (IC_{50} = 70 nM) is more potent than compound **57** (IC_{50} = 260 nM). Compound **32** has a nitrile group surrounded by Trp244, Gln248, and Trp252, which enables compound **32** to have more van der Waals interactions with these residues than compound **57** has although the nitrile group is polar. For the same reason, compound **37** is more potent (86 nM) than compound **47** (170 nM). The favorable van der Waals interactions are also observed for compounds **6**, **16**, **38**, **26,** and **50** which have MurI inhibitory activities of 25, 39, 87, 60, and 220 nM, respectively. At the receptor side, the Phe13 side chain is a van der Waals favorable moiety. Compound **41** (IC_{50} = 103 nM) has stronger inhibitory activity than compound **69** (IC_{50} = 2200 nM) due to an additional chlorine atom attached to the indole ring for more van der Waals interactions with the

moiety. On the other hand, van der Waals interactions with Ser4, Ile149, Glu150, Ser152, Leu154, and Leu186 can reduce the binding affinity between the ligand and receptor. For example, compound **17** is more potent (IC_{50} = 41 nM) than compound **52** (IC_{50} = 220 nM) because compound **52** has more van der Waals interactions with these residues through its substituent in the R_1 position of 2*H*-pyrazolo[3,4-*d*]pyrimidine-4,6(*5H*,7*H*)-dione moiety.

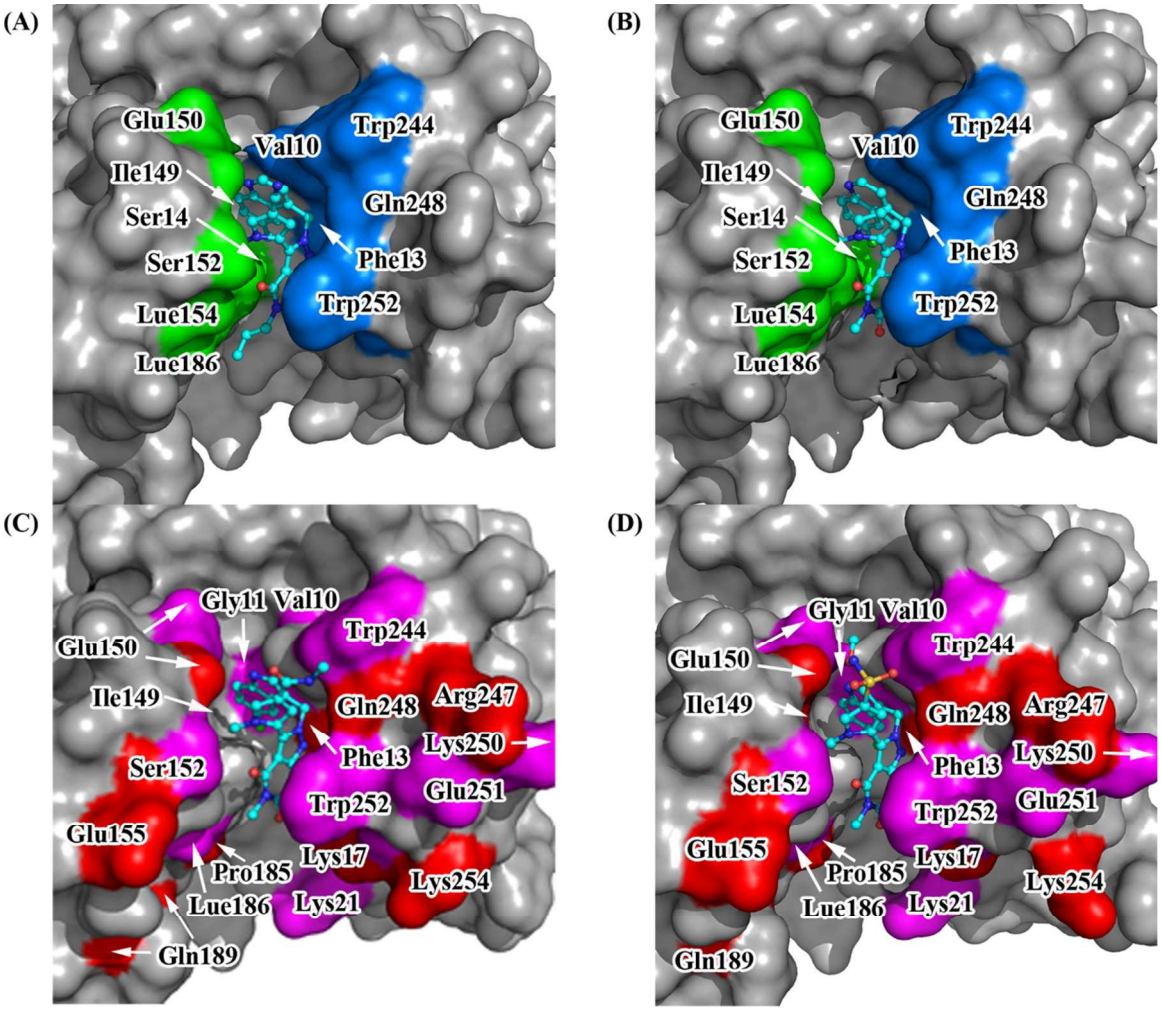


Figure 6. MM/GBSA-PLSR van der Waals and electrostatic interaction surfaces of compound **32** (A) and its counterpart **57** (B) as well as compound **1** (C) and its counterpart **13** (D). Marine regions indicate that van der Waals interactions are favorable

in activity enhancement, whereas green regions display that the interactions are unfavorable (A, B). Red areas suggest that electrostatic interactions increase activity, while the magenta area represents electrostatic interactions that can reduce inhibition (C, D).

Electrostatic interactions are more complicated, as demonstrated by the mixed red and magenta regions in Figure 6C and 5D. The model shows a multi-layered pattern of electrostatic interactions in the right flank of the pocket; electrostatic interactions with Trp244, Trp252, Glu251, and Lys21 are not preferred for the binding affinity, whereas those with Gln248, Arg247, Lys17, and Lys254 are preferred. Substituents at the R_2 position of 2*H*-pyrazolo[3,4-*d*]pyrimidine-4,6(5*H*,7*H*)-dione can improve binding affinity through electrostatic interactions (e.g. hydrogen bonding with Gln248 side chain). However, there are one red and two magenta regions nearby as well. A substituent for this area has to accommodate multiple types of binding interactions. For example, compound **1** (IC_{50} = 6 nM) is five-fold more potent than compound **13** (IC_{50} = 36 nM). The former has a $–CONHCH_3$ group at the 1-methyl-1H-pyrrole ring, while the latter has an $–SO_2NHOCH_3$ group (bulkier), generating more electrostatic interactions with Trp244 and Ser152 and reducing the binding affinity. A longer substituent group, such as compound **13**, may weaken hydrogen bonding with Gln248 but strengthens the van der Waals interaction with Glu150. Compound **13** still forms a hydrogen bond with Glu150 with a moderate activity although it experiences unfavorable interactions (Figure 6C, D). Another tricky spot is at the unfavorable electrostatic interactions with the Glu150 side chain, Ser152 side chain, and Leu186 backbone, and favorable interactions with the Glu150 backbone. For example, compounds **13**, **8**, and **7** have IC_{50} of 36, 27, and 26 nM,

respectively, due to $–SO_2NHOCH_3$, $–SO_2NHCH_3$, and $–SO_2CH_3$ groups. Compound **13** has the longest substituent which improves activity via the weakened hydrogen bond to the favored Glu150 backbone, but in the mean while its activity is reduced by unfavorable electrostatic and van der Waals interactions with the Glu150 side chain. The shorter substituent, $–SO_2NHCH_3$, reduces the conflict interaction to improve the activity for compound **8**. The shortest substituent, $–SO_2CH_3$, further deceases the conflict interactions though weakening the hydrogen bond interaction. Compounds having an $–SO_2R$ substituent with different R sizes demonstrate a consistent activity order: compound **3** ($–SO_2NH_2$, 16 nM) < compound **12** ($–SO_2CH_3$, 34 nM) < compound **18** ($–SO_2NHCH_3$, 44 nM) < compound **21** ($–SO_2NHOCH_3$, 55 nM). The last electrostatically favorable region is at Glu155, but it cannot improve the activity because it is far away from the native ligand.

The polar solvation interaction is only introduced by the MM/GBSA-PLSR approach. This type of interaction cannot be mapped onto a CoMFA field or CoMSIA field. Because polar solvation interactions are not for direct interactions between the ligand and receptor, they cannot be directly used for ligand design. However, the polar solvation interaction reflects an indispensable receptor-ligand interaction in solvent. By combining polar solvation interactions with steric and electrostatic interactions, the 3D-QSAR can be better articulated and mapped. For example, Trp252 colored in orange (Figure 6) representing favorable polar solvation interactions together with the van der Waals and electrostatic interactions indicates that a large hydrophobic group is required to interact with the Trp252 side chain, and fewer electrostatic interactions will improve the binding affinity. In Figure 7, Glu150 and Gln248 (in bright yellow) indicate there is a limit on

activity enhancement. That is, the activity cannot be unlimitedly improved by increasing electrostatic and van der Waals interaction because polar solvation interactions (another factor that improves activity) will be reduced simultaneously and *vice versa*.

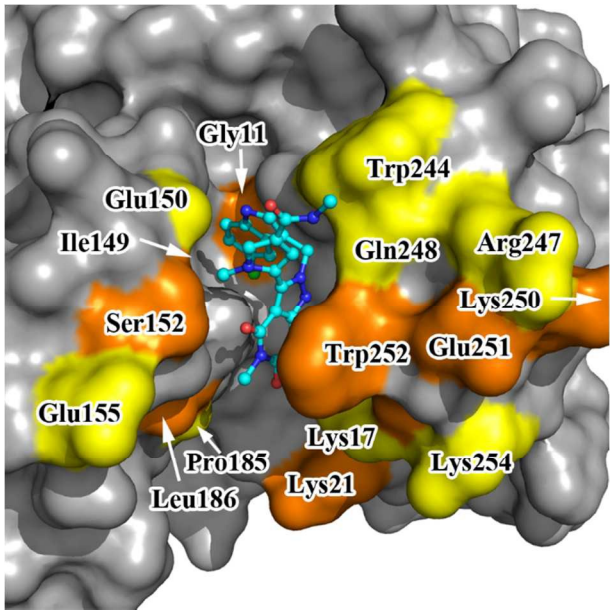


Figure 7. The protein surfaces showing the polar solvation interactions. Favorable residues are depicted in orange, while unfavorable residues are represented in yellow.

4. Conclusions

The binding features of MurI uncompetitive inhibitors have been articulated with a new structure-based 3D-QSAR approach, MM/GBSA-PLSR. To better understand this model, the ligand-based 3D-QSAR models of CoMFA and CoMSIA have been created and compared against the model. The structure-based 3D-QSAR results are interpreted with respect to the relations of the activity and the interaction descriptors (van der Waals, electrostatic, polar solvation, and nonpolar solvation). For the MurI inhibitors, the interacting factors contributing to the activity are van der Waals (29.5%), electrostatic (38.2%), and polar solvation (32.3%). By associating the different types of interactions of

the MM/GBSA-PLSR model with the fields of CoMFA and CoMSIA, the 3D-QSAR models are better elucidated. MM/GBSA-PLSR van der Waals interactions can be mapped to CoMFA/CoMSIA steric interaction fields; MM/GBSA-PLSR electrostatic interactions can be mapped to CoMFA/CoMSIA electrostatic and H-bond acceptor/donor interactions fields. There is no explicit mapping between MM/GBSA-PLSR solvation interactions (polar or non-polar) and CoMFA/CoMSIA fields. However, this type of interaction is still useful for ligand design. In general, MM/GBSA-PLSR takes advantages of receptor-ligand interactions, models induced-fit effects, considers solvent effect, substitutes rough exclusion volumes in modeling, and avoids putative conformational alignment, which enables itself to surmount the defects of the ligand-based models, and has distinguished itself from others. The information acquired in this study provides a tool for guiding further optimization of potent MurI uncompetitive inhibitors. And this approach may serve as a rational means for lead optimization and drug design by explicitly mapping the favorite/un-favorite pharmacophore regions onto the binding pocket.

Acknowledgements

This work was supported by a grant from the National High Technology Research and Development Program of China (863 Program) (No. 2012AA020307), the Guangdong Innovative Research Team Program (No. 2009010058), the National Natural Science Foundation of China (No. 81173470), the Special Funding Program for the National Supercomputer Center in Guangzhou (2012Y2-00048/2013Y2-00045, 201200000037), and Major Scientific and Technological Project of Guangdong Province (No. 2011A080300001).

Notes and references

XL E-mail: lexiu2012@163.com.

QG Email: guqiong@mail.sysu.edu.cn.

JX Email: junxu@biochemomes.com.

† The experiment design XL, JX. Implementation: XL. Manuscript revision and submission: QG and JX.

Electronic Supplementary Information (ESI):

Table S1. Chemical structures and pIC_{50} values of the 69 compounds used in development of the 3D-QSAR models.

Table S2. Y-randomization test for the MM/GBSA-PLSR model.

Table S3. Predicted pIC50 values and respective residuals generated from the ligand-based and structure-based 3D-QSAR models.

Table S4. Coefficients (Coeff) and standard-deviation-weighted coefficients (StDev*Coeff) of the MM/GBSA-PLSR model.

Figure S1. Discussion on the outlier compound **68** occurring in the CoMFA model.

1. L. E. Wroblewski, R. M. Peek, Jr. and K. T. Wilson, *Clinical microbiology reviews*, 2010, **23**, 713-739.
2. R. M. Peek, Jr. and M. J. Blaser, *Nature reviews. Cancer*, 2002, **2**, 28-37.
3. A. C. Ford, D. Forman, R. H. Hunt, Y. Yuan and P. Moayyedi, *Bmj*, 2014, **348**, g3174.
4. S. Z. Ding and P. Y. Zheng, *Gut pathogens*, 2012, **4**, 18.
5. C. P. Dooley, H. Cohen, P. L. Fitzgibbons, M. Bauer, M. D. Appleman, G. I. Perez-Perez and M. J. Blaser, *The New England journal of medicine*, 1989, **321**, 1562-1566.
6. N. R. Salama, M. L. Hartung and A. Muller, *Nature reviews. Microbiology*, 2013, **11**, 385-399.
7. P. Malfertheiner, F. Megraud, C. A. O'Morain, J. Atherton, A. T. Axon, F. Bazzoli, G. F. Gensini, J. P. Gisbert, D. Y. Graham, T. Rokkas, E. M. El-Omar, E. J. Kuipers and G. European Helicobacter Study, *Gut*, 2012, **61**, 646-664.
8. P. Malfertheiner, F. Megraud, C. O'Morain, F. Bazzoli, E. El-Omar, D. Graham, R. Hunt, T. Rokkas, N. Vakil and E. J. Kuipers, *Gut*, 2007, **56**, 772-781.

9. D. Y. Graham and L. Fischbach, *Gut*, 2010, **59**, 1143-1153.
10. A. G. O'Connor, *Nursing times*, 1984, **80**, suppl 5-6.
11. H. Barreteau, A. Kovac, A. Boniface, M. Sova, S. Gobec and D. Blanot, *FEMS microbiology reviews*, 2008, **32**, 168-207.
12. J. van Heijenoort, *Natural product reports*, 2001, **18**, 503-519.
13. M. E. Tanner, *Accounts of chemical research*, 2002, **35**, 237-246.
14. E. D. LoVullo, C. R. Molins-Schneekloth, H. P. Schweizer and M. S. Pavelka, Jr., *Microbiology*, 2009, **155**, 1152-1163.
15. K. Y. Shatalin and A. A. Neyfakh, *FEMS microbiology letters*, 2005, **245**, 315-319.
16. S. L. Fisher, *Microbial biotechnology*, 2008, **1**, 345-360.
17. S. Glavas and M. E. Tanner, *Biochemistry*, 2001, **40**, 6199-6204.
18. A. de Dios, L. Prieto, J. A. Martin, A. Rubio, J. Ezquerra, M. Tebbe, B. Lopez de Uralde, J. Martin, A. Sanchez, D. L. LeTourneau, J. E. McGee, C. Boylan, T. R. Parr, Jr. and M. C. Smith, *Journal of medicinal chemistry*, 2002, **45**, 4559-4570.
19. T. Lundqvist, S. L. Fisher, G. Kern, R. H. Folmer, Y. Xue, D. T. Newton, T. A. Keating, R. A. Alm and B. L. de Jonge, *Nature*, 2007, **447**, 817-822.
20. S. N. Ruzheinikov, M. A. Taal, S. E. Sedelnikova, P. J. Baker and D. W. Rice, *Structure*, 2005, **13**, 1707-1713.
21. K. L. Whalen, K. L. Pankow, S. R. Blanke and M. A. Spies, *ACS medicinal chemistry letters*, 2010, **1**, 9-13.
22. B. Geng, G. Breault, J. Comita-Prevoir, R. Petrichko, C. Eyermann, T. Lundqvist, P. Doig, E. Gorseth and B. Noonan, *Bioorganic & medicinal chemistry letters*, 2008, **18**, 4368-4372.
23. B. Geng, G. Basarab, J. Comita-Prevoir, M. Gowravaram, P. Hill, A. Kiely, J. Loch, L. MacPherson, M. Morningstar, G. Mullen, E. Osimboni, A. Satz, C. Eyermann and T. Lundqvist, *Bioorganic & medicinal chemistry letters*, 2009, **19**, 930-936.
24. G. A. Breault, J. Comita-Prevoir, C. J. Eyermann, B. Geng, R. Petrichko, P. Doig, E. Gorseth and B. Noonan, *Bioorganic & medicinal chemistry letters*, 2008, **18**, 6100-6103.
25. G. S. Basarab, P. J. Hill, A. Rastagar and P. J. Webborn, *Bioorganic & medicinal chemistry letters*, 2008, **18**, 4716-4722.
26. G. S. Basarab, P. Hill, C. J. Eyermann, M. Gowravaram, H. Kack and E. Osimoni, *Bioorganic & medicinal chemistry letters*, 2012, **22**, 5600-5607.
27. S. Avram, I. Svab, C. Bologa and M. L. Flonta, *Journal of cellular and molecular medicine*, 2003, **7**, 287-296.
28. C. C. Melo-Filho, R. C. Braga and C. H. Andrade, *Current computer-aided drug design*, 2014, **10**, 148-159.
29. A. R. Ortiz, M. T. Pisabarro, F. Gago and R. C. Wade, *Journal of medicinal chemistry*, 1995, **38**, 2681-2691.
30. J. Verma, V. M. Khedkar and E. C. Coutinho, *Current topics in medicinal chemistry*, 2010, **10**, 95-115.
31. R. D. Cramer, D. E. Patterson and J. D. Bunce, *Journal of the American Chemical Society*, 1988, **110**, 5959-5967.
32. G. Klebe and U. Abraham, *Journal of computer-aided molecular design*, 1999, **13**, 1-10.
33. E. R. Yera, A. E. Cleves and A. N. Jain, *Journal of medicinal chemistry*, 2011, **54**, 6771-6785.
34. *ROCS 3.2.0.4*, OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com.
35. *EON 2.2.0.5*, OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com.
36. J. G. Vinter, *Journal of computer-aided molecular design*, 1994, **8**, 653-668.

37. *Molecular Operating Environment (MOE), 2013.08*, Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2013.
38. R. A. Brown and D. A. Case, *Journal of computational chemistry*, 2006, **27**, 1662-1675.
39. Z. Zuo, N. S. Gandhi and R. L. Mancera, *Journal of chemical information and modeling*, 2010, **50**, 2201-2212.
40. L. Li, D. Li, H. Chen and J. G. Han, *Journal of biomolecular structure & dynamics*, 2013, **31**, 299-315.
41. T. Hou, J. Wang, Y. Li and W. Wang, *Journal of chemical information and modeling*, 2011, **51**, 69-82.
42. P. Geladi and B. R. Kowalski, *Analytica Chimica Acta*, 1986, **185**, 1-17.
43. *Discovery Studio, version 3.5*, Accelrys Software Inc., San Diego, CA, 2012.
44. P. R. Gerber and K. Muller, *Journal of computer-aided molecular design*, 1995, **9**, 251-268.
45. D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Götz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P. A. Kollman, AMBER 12, University of California, San Francisco, 2012.
46. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, Gaussian 09, Gaussian. Inc., Wallingford CT, 2009.
47. C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *The Journal of Physical Chemistry*, 1993, **97**, 10269-10280.
48. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *Journal of Chemical Physics*, 1983, **79**, 926-935.
49. F. M. Ytreberg and D. M. Zuckerman, *The Journal of chemical physics*, 2004, **120**, 10876-10879.
50. M. S. Lee and M. A. Olson, *Biophysical Journal*, 2006, **90**, 864-877.
51. A. Suenaga, N. Okimoto, Y. Hirano and K. Fukui, *PloS one*, 2012, **7**, e42846.
52. M. H. Joseph and A. Georgios, in *Molecular Dynamics - Studies of Synthetic and Biological Macromolecules*, ed. L. Wang, InTech, Rijeka, Kroatien, 2012.
53. P. Zhou, X. Chen, Y. Wu and Z. Shang, *Amino acids*, 2010, **38**, 199-212.
54. Y. Yang, J. Qin, H. Liu and X. Yao, *Journal of chemical information and modeling*, 2011, **51**, 680-692.
55. S. P. Brown and S. W. Muchmore, *Journal of chemical information and modeling*, 2006, **46**, 999-1005.
56. M. K. Gilson and H. X. Zhou, *Annual review of biophysics and biomolecular structure*, 2007, **36**, 21-42.

592 57. L. Wang, B. J. Berne and R. A. Friesner, *Proceedings of the National Academy of*
593 *Sciences of the United States of America*, 2012, **109**, 1937-1942.
594 58. B.-H. Mevik and R. Wehrens, *Journal of Statistical Software*, 2007, **18**.
595 59. R Core Team, R: A Language and Environment for Statistical Computing, R
596 Foundation for Statistical Computing, Vienna, Austria, 2014.
597 60. *SYBYL-X 1.1*, Tripos Inc., St. Louis, MO, USA, 2008.

598

599