A Simple and Efficient Docking Method to the Cyclin-Dependent Kinase 2

Kwang-Su Park, Jinyoung Kim, Youhoon Chong,[*] and Hyunah Choo[†,*]

Division of Bioscience and Biotechnology, Institute of Biomedical Science and Technology, Konkuk University,
Seoul 143-701, Korea. [*]E-mail: chongy@konkuk.ac.kr
[†]Life Sciences Division, Korea Institute of Science and Technology, P.O. Box 131, Cheongryang, Seoul 130-650, Korea
[*]E-mail: hchoo@kist.re.kr

The subtle but significant differences and thereby the lack of consensus in active site structures among the crystal structures of cyclin-dependent kinase 2 (CDK2) has hampered structure-based drug design. In this study, we devised a simple but effective 'mutation, pharmacophore-guided docking, followed by mutation' strategy to generate an "average" CDK2 structure, which was used for ligand docking study to successfully reproduce 30 out of 32 X-ray ligand positions within 2.0 Å of heavy atom RMSD. This novel docking method was applied for structure-based 3D QSAR with CoMSIA study of a series of structurally related ligands, which showed a good discrimination between CDK2 binders and nonbinders.

Key Words : CDK2, Docking, Mutation, CoMSIA

Introduction

Cyclin-dependent kinase 2 (CDK2) is an enzyme involved in the regulation of the cell cycle, which undergoes a series of conformational changes upon binding to cyclin and phosphorylation to yield a fully active complex.[1] ATP and inhibitor binding at the ATP- binding site also induce further conformational changes.[2] Thus, more than 60 CDK2 crystal structures are publicly available at present, which have subtle but significant structural differences.[2] In a crystal structure, however, a protein conformation is optimally adapted for interaction with one specific ligand, which makes it difficult to choose a basal protein structure for use in high-throughput docking of thousands of ligands of diverse structure.

There have been several approaches to tackle this problem, which include i) docking a set of ligands into several different structures of the same protein and take the hits from each structure,[3] and ii) statistical model-fitting approach of fitting multiple linear regression models to important geometrical features of the active site to identify the conformational change of the enzyme caused by binding of the ligand.[4] However, in spite of the advances provided by these studies in identifying the best protein structure for ligand docking, there is an ongoing need for simple, intuitive, straightforward and easily reproducible method which can be generally used for docking study of conformationally flexible enzymes.

Recently, Thomas *et al.* showed that there are some CDK2 structures that are clearly better than others for docking study, and the main determinants of this are the volume of the binding site into which the ligands are docked.[3] Moreover, by comparison of 20 CDK2 structures, they concluded that the more the side chains of Lys33, Phe80, Lys89, and Asp145 protruded into the ATP binding site, the smaller the number of correctly docked ligands tended to be. Taken

together, even though the relationship was not clearly notified by the authors, it was obvious that the volume of the ligand binding site would be determined by conformations of the four amino acid residues (Lys33, Phe80, Lys89, and Asp145). Superposition of 32 CDK2 structures (Fig. 1) clearly shows that they share almost the same backbone conformation but differ in the side chain conformation of the four amino acid residues (Lys33, Phe80, Lys89, and Asp145).

In this study, we focused to minimize the effect of the side chain conformation on docking of various ligands by widening the ligand binding site of the enzyme. Thus, the four
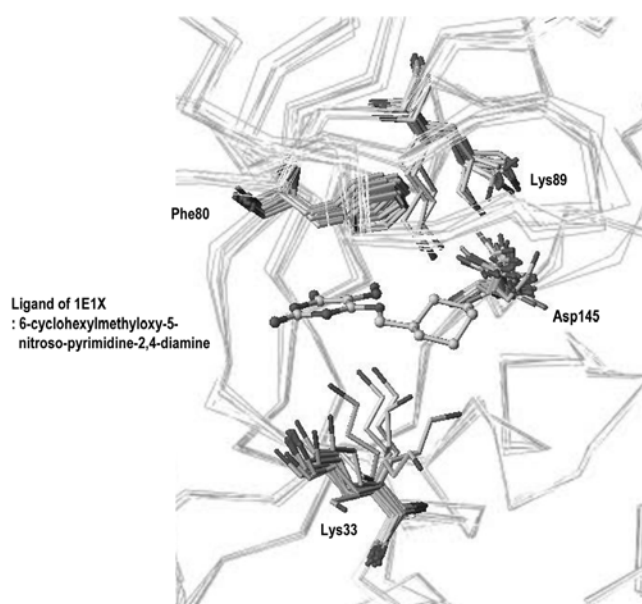

Figure 1. Superposition of 32 CDK2 structures used in this study. Fine gray lines indicate superimposed backbone Cα's, and the four residues (Lys33, Phe80, Lys89 and Asp145) are shown as capped sticks. A ball and stick model at the center is a ligand of the PDB ID 1E1X.

critical amino acid residues (Lys33, Phe80, Lys89, and Asp145) were mutated into glycines, and then ligands extracted from the crystal structures of 32 CDK2 complexes were allowed to dock into the wide open active site. After docking, the four residues were mutated back into their initial forms and the side chain conformations were optimized accordingly to generate enzyme-ligand complexes. The docking result was examined in terms of the ability to reproduce X-ray ligand positions within 2.0 Å heavy atom root-mean-square deviation (RMSD). Also, structure-based 3D QSAR study of a set of CDK2 inhibitors was compared with ligand-based 3D QSAR study to examine how well the docking protocol discriminates the CDK2-binders from the nonbinders.

Materials and Methods

Protein Structures. The following CDK2 protein structures, taken from the Protein Data Bank,[5] were used in this work: 1B38,[6] 1CKP,[7] 1DI8,[8] 1DM2,[9] 1E1V,[10] 1E1X,[10] 1E9H,[11] 1FVT,[12] 1FVV,[12] 1G5S,[13] 1GZ8,[14] 1H00,[15] 1H08,[15] 1H0V,[14] 1H0W,[14] 1H1Q,[16] 1H1R,[16] 1H1S,[16] 1JSV,[17] 1JVP,[18] 1KE5,[19] 1KE6,[19] 1KE7,[19] 1KE8,[19] 1OGU,[20] 1OIQ,[21] 1OIR,[21] 1OIT,[21] 1P2A,[2] 1PF8,[22] 1PXK,[23] and 1PXL.[23] These 32 structures were chosen because of their diversity in terms of resolution, size, the presence or absence of other proteins, phosphorylation status, and ATP binding site conformation. Details of the 32 ligand structures identified by their CDK2 crystal structure PDB identifiers are presented in Figure 2.

The side chains of lysine and arginine residues are protonated and the carboxylate groups of aspartic and glutamic acid are ionized. Water molecules contained in the PDB file have been removed. In order to define the active sites of the proteins, all 32 enzyme structures were superimposed together with their ligand structure using the "Align Structures by Homology" module of Sybyl 7.2 (Tripos Inc., St. Louis, MO). All atoms are selected that are located less than 10 Å apart from an atom of any ligand of the crystal structures. In addition, the complete amino acid is selected if at least one of its atoms is picked.

Mutation. A crystal structure of CDK2 was randomly chosen (1E1X[10]). The four amino acid residues (Lys33, Phe80, Lys89, and Asp145) were mutated into glycines using "mutate monomers" module in Sybyl 7.2 to provide the protein structure which was used for docking ligands.

Ligand Structures. First, the ligand coordinates of the non-hydrogen atoms were extracted from the original PDB. They are used as reference for the calculation of the RMSD values later on. The ligand input files were obtained by defining correct atom types (including hybridization states) and correct bond types, adding hydrogen atoms, assigning formal charges to each atom, and finally energy-minimizing the reference structure. The energy minimization guarantees a low-energy conformation with suitable bond length and angles. This new geometry and the fact that the minimized structures is not translated according to its original crystal structure guarantees that there is no implicit docking

information about the protein-ligand complex of the PDB structure in the ligand input file. In general, all carboxylic acid groups are ionized while all amino groups but no amide groups are protonated.

Docking. Ligands were docked into the mutated enzyme structures using the pharmacophore-guided docking program FlexX-Pharm [running under Sybyl version 7.2][24] in order to guide the construction of the ligand in the active site, ensuring the formation of hydrogen bonds to Leu83 CO and Leu83 NH because most of the CDK inhibitors form a hydrogen bond with the backbone of Glu81/Leu83. After docking, the four residues (Gly33, Gly80, Gly89, and Gly145) were mutated back to their original forms (Lys33, Phe80, Lys89, and Asp145) with concurrent minimization of the side chains of the mutated residues by using the 'Set Side-chain Conformation' module in Sybyl 7.2. The final adjustment of the conformation of the bound ligands was performed by energy minimization of the enzyme-ligand complex.

Energy Minimization. Gästeiger-Hückel charge[25,26] was given to the ligand. Then, Kollman-All-Atom charges[27,28] were loaded to the enzyme site from the biopolymer option in Sybyl 7.2. The enzyme-inhibitor complexes were minimized by using Kollman-All-Atom force field until the RMS (root mean square) tolerance value of 0.05 kcal/mol was achieved.

3D QSAR Study. CDK2 inhibitory activity of twenty five 2-amino-4-(thiazol-5-yl)pyrimidine compounds was obtained from the literature.[29] Both ligand-based and structure-based three dimensional structure-activity relationship (3D QSAR) studies were performed. All compounds were constructed by the 'Sketch' module in SYBYL base, protonated and assigned with MMFF94s charges. For more flexible compounds, systematic searches were performed with an interval of 10° on every rotatory bond to ensure their lowest energy conformations. Finally, repeated minimization was performed using the steepest descent and conjugated gradient method until the RMS tolerance value of 0.001 kcal/mol was achieved. The CoMSIA method is based on molecular similarity indices. Using a common probe atom, similarity indices were calculated for a data set of prealigned molecules at regularly spaced grid points and evaluated in a PLS analysis following the usual CoMFA protocol. The regression analysis of CoMSIA field energies was performed using the partial least squares (PLS) algorithm with the leave-one-out (LOO) method adopted for cross validation.[30] The CoMSIA descriptors served as independent variables and pK_i values as a dependent variable in PLS regression analysis in deducing the 3D-QSAR models. Normally, cross-validation is used to check the predictivity of the derived model. Results of the analyses correspond to a regression equation with thousands of coefficients. The performance of models was calculated using the LOO cross-validation method. The optimum number of components (Nc) used to derive the non-cross-validated model was defined as the number of components leading to the highest r^2 cross-validated and lowest standard error of prediction (SEP). To obtain the statistical confidence limit in analyses, PLS ana-
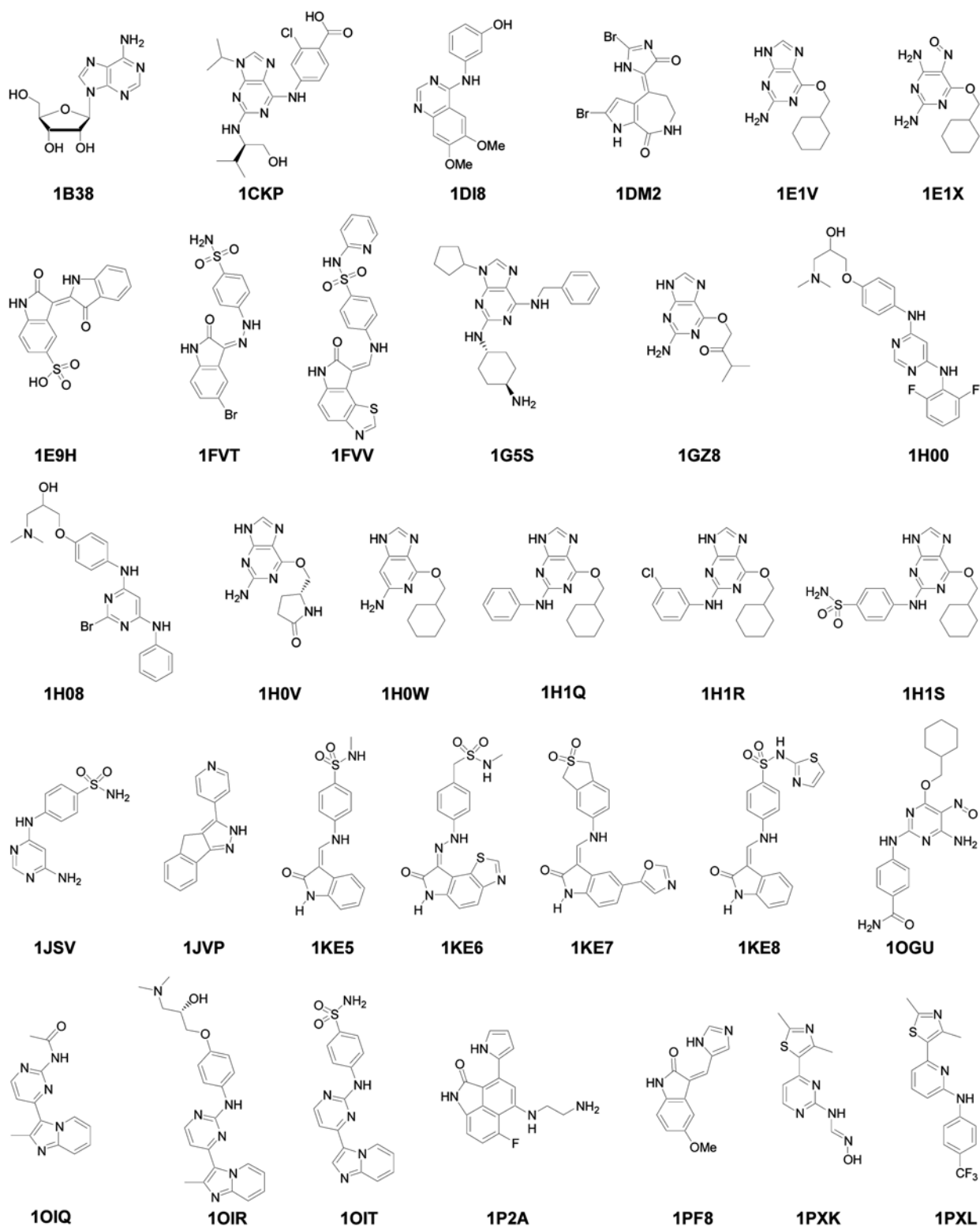
Figure 2. Structures of the 32 ligands identified by their CDK2 crystal structure PDB identifiers.

lysis using 100 bootstrap groups within the optimum number of components was performed.

Results and Discussion

Mutant Enzyme Structure. Generally, in a crystal struc-

ture, a protein conformation is optimally adapted for inter-action with one specific ligand. As a result, there have been numerous efforts to abolish the conformational bias between the ligand and enzyme in a specific crystal structure and devise an "average" CDK2 structure which can be generally used for docking study.[3,4,31-34] Our goal in this study was to
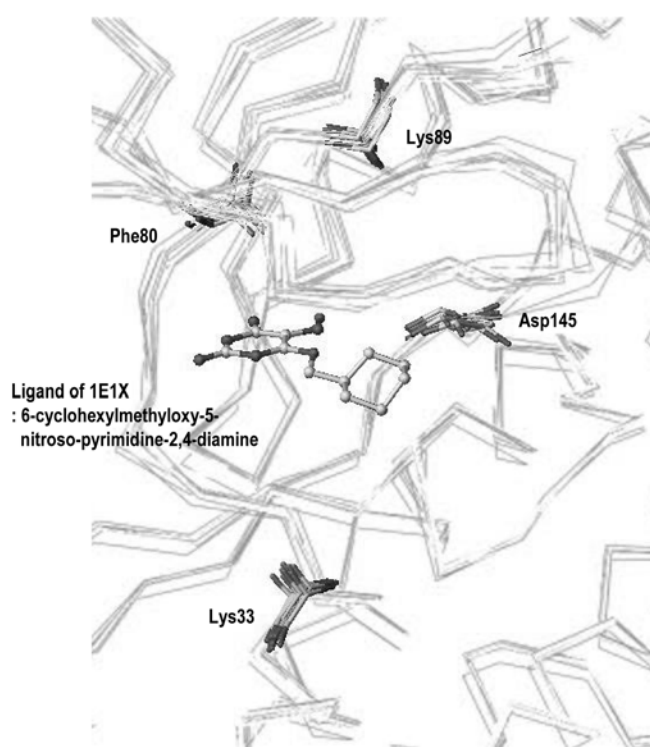
Figure 3. Superposition of 32 CDK2 enzymes mutated at the four residues (Lys33, Phe80, Lys89, and Asp145). Fine gray lines indicate backbone Cα superimposed but the four mutated residues (Gly33, Gly80, Gly89 and Gly145) are shown as capped sticks. A ball and stick model at the center is a ligand of the PDB ID 1E1X.

provide a docking method which can be easily exploited and generally applied for docking studies of conformationally flexible enzymes. For this purpose, an enzyme structure which is not biased to a specific ligand had to be identified first, and thus, we generated a mutant enzyme structure which shares common structural motifs among the series of CDK2 crystal structures. Superposition of the 32 CDK2 crystal structures used in this study (Fig. 1) revealed that CDK2 structures share almost the same backbone conformations but differ in the side chain conformations at the four residues (Lys33, Phe80, Lys89, and Asp145), which are responsible for the different volume and thereby different three dimensional structure of the ATP-binding site of CDK2. As a result, mutation of the four key residues in each CDK2 crystal structures into glycines generated almost identical mutant enzyme structures which could be superimposed atom by atom (Fig. 3).

In other words, our study shows that a random choice of any CDK2 structure would result in the same structure after mutation of the four amino acid residues (Lys33, Phe80, Lys89, and Asp145). In this study, a CDK2 structure obtained by mutation of 1E1X (PDB ID) was used for docking study.

Docking. The use of mutant CDK2 structure in docking study is based on an assumption that the side chains of the mutated four glycine residues do not significantly affect the ligand docking at the ATP-binding site of CDK2. In this regard, it should be noted that CDK2 crystal structures

identified to date shows that ligands binding at the ATP-binding-site should form a hydrogen bond with the backbone of Leu83.[3] Even though additional hydrogen bonds with Lys33, Lys89, or Gln131 take part in fine-tuning the binding affinities of some ligands, the key interaction is hydrogen bonding between ligand and Leu83 CO and Leu83 NH. Thus, dockings were performed with the restraint that docked ligands should form hydrogen bonds to Leu83 CO and Leu83 NH by using the pharmacophore-guided docking (FlexX-Pharm) module in Sybyl 7.2. In FlexX-Pharm, which is an extended version of FlexX, a previously defined set of pharmacophore features in the active site constraints the docking calculation so that only solutions are produced that match the specified set of features. For docking with FlexX-Pharm, interacting groups and the corresponding interaction type (hydrogen bond donor, acceptor, hydrophobic pocket) in the active site must be specified. FlexX-Pharm then ensures that an interaction is formed between the specified interaction group in the active site and the ligand in a valid docking solution. A compound was regarded as having been docked correctly if, by visual inspection, it formed the expected hydrogen bonds (as defined by formation of the requisite hydrogen bonds with Leu83 and correct orientation of the ligand in the ATP-binding site). Interestingly, all 32 ligands were successfully docked with the correct hydrogen bonds formed with Leu83, which is in sharp contrast with the docking ratio of the same set of ligands docked with the wild type 1E1X (10 out of 32, 31.3%). Due to the formation of rigid multiple hydrogen bonds, the correct docking pose was always reproduced. However, they failed to adopt the exact conformations observed crystallographically due to the lack of the side chains of the mutated four residues (Lys33, Phe80, Lys89, and Asp145) with which they can form additional hydrogen bonding networks. Fine-tuning of the docking pose was achieved by mutating the four residues back to the corresponding original amino acids with concurrent adjustment of the side chain conformations. This process was automatically performed by the 'Set Sidechain Conformation' module in Sybyl 7.2. In order to avoid steric hindrance but to accommodate the formation of additional hydrogen bonds between

**Table 1**. RMSD of the docked ligands from the X-ray ligand positions

PDB	RMSD (Å)	PDB	RMSD (Å)	PDB	RMSD (Å)
1B38	2.34	1H00	1.67	1KE7	0.54
1CKP	0.24	1H08	1.42	1KE8	1.79
1DI8	1.84	1H0V	0.67	1OGU	1.23
1DM2	0.64	1H0W	0.69	1OIQ	0.27
1E1V	0.65	1H1Q	0.25	1OIR	1.39
1E1X	1.21	1H1R	2.27	1OIT	0.82
1E9H	1.24	1H1S	1.11	1P2A	0.63
1FVT	1.21	1JSV	1.76	1PF8	0.61
1FVV	1.93	1JVP	0.54	1PXK	1.00
1G5S	1.35	1KE5	1.16	1PXL	0.95
1GZ8	1.18	1KE6	1.86		

the bound ligands and enzyme, the obtained enzyme-ligand complexes were fully relaxed by energy minimization to give the final models of CDK2-ligand complex.

The results of our novel docking protocol were examined in terms of the ability to reproduce X-ray ligand positions within 2.0 Å heavy atom RMSD. Adjustment of the mutated side chains followed by energy-minimization of the enzyme-ligand complexes placed 30 out of 32 ligands with an RMSD to the crystal structure below 2.0 Å (Table 1). It is noteworthy that the docking poses of 32 ligands were well
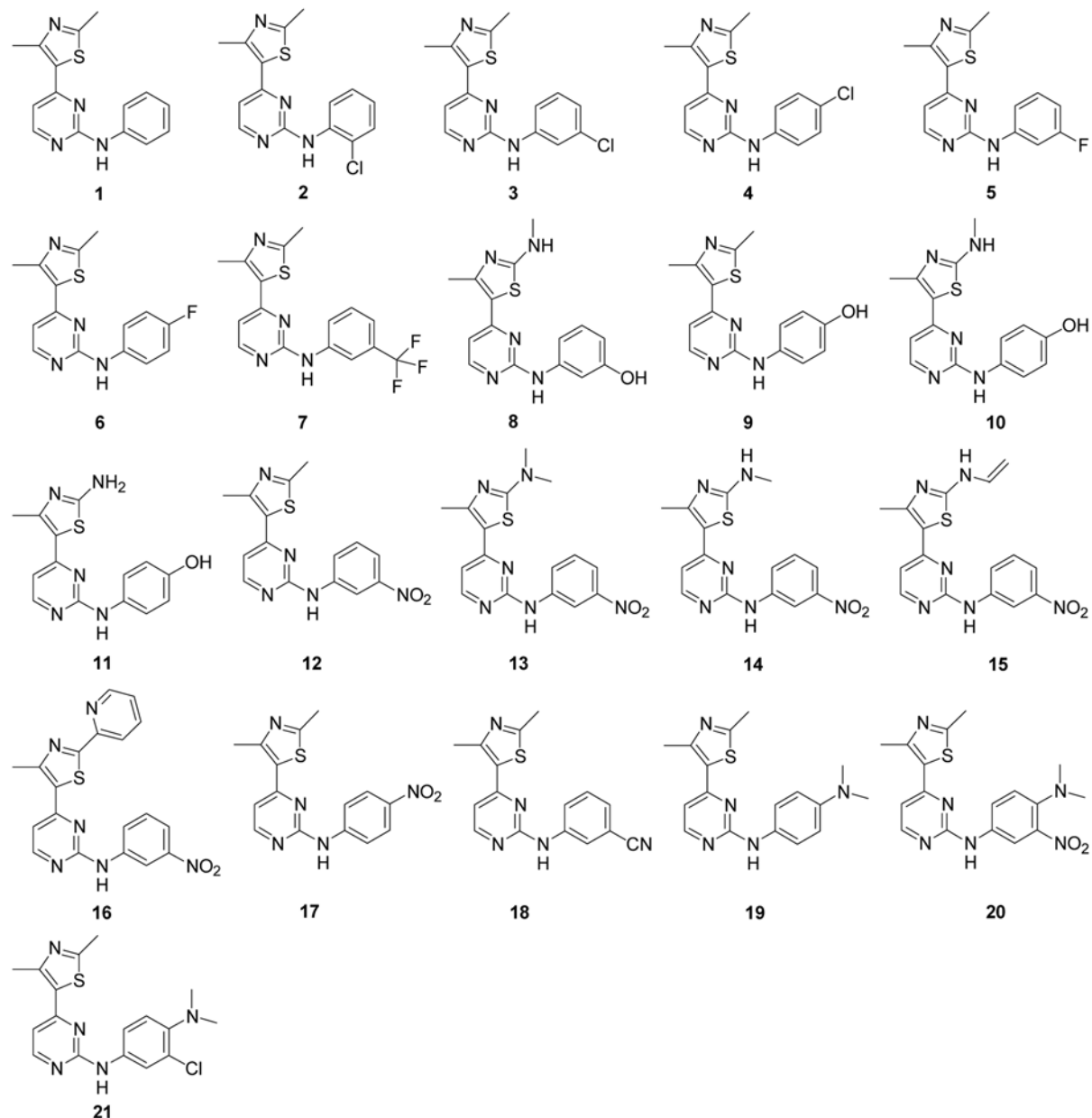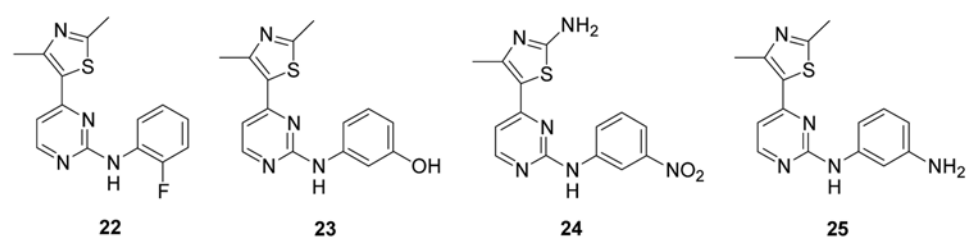

Figure 4. 2-Amino-4-(thiazol-5-yl)pyrimidine compounds as CDK2 inhibitors.

reproduced in the RMSD range of 0.2-2.3 Å and, in many cases, the ligand-RMSDs were even better than that of the self-docking of ligand 1E1X (Fig. 2) into its original crystal structure (1.21 Å, Table 1). This result shows that the randomly chosen mutated 'average' enzyme structure is not structurally biased to its original crystal structure which is specific for the bound ligand.

3D-Quantitative Structure-Activity Relationship (QSAR) Study. As our docking protocol is based on mutation of amino acid residues and thereby widening the ligand binding site by mutation, there could be a tendency for underestimation of the side chains of the mutated amino acid residues as well as overestimation of volume of the ligand binding site and thereby generation of false positives during docking. Thus, as a measure to examine how well the docking protocol discriminates the binders from the nonbinders, comparative 3D QSAR studies were performed on the dataset of Wang *et al.* (Fig. 4).[29] We first applied a ligand-based strategy which uses atom-based fit to align training set as well as test set molecules. Second, the availability of newly developed docking method allowed us to apply a structure-based technique by docking database molecules into the ligand binding site of an 'average' CDK2 structure obtained by mutation. This way we derived two comparative 3D QSAR models, either neglecting or including X-ray information. The validity check of the ligand-based versus the structure-based model comprised statistical aspects (fitting, prediction, PLS coefficients).[35] In the ligand-based 3D QSAR study, compounds were aligned by atom based fit around the thiazolpyrimidine moiety (Fig. 5a).

The basic principle of CoMSIA[36,37] is the same as that of CoMFA,[36,38] but CoMSIA includes some additional descriptors such as hydrophobicity, hydrogen bond donor and hydrogen bond acceptor. In this study, CoMSIA was chosen as the method of choice in preference to CoMFA because hydrogen bonding interaction field is included in the CoMSIA model, which might play a key role in fine-tuning the binding affinity of ligand molecules to the ATP-binding site of CDK2.[3] A ligand-based CoMSIA model was established of 21 molecules (Fig. 4a) by using four different fields (steric, electrostatic, hydrogen bond donor, and hydrogen bond acceptor) [$q^2 = 0.466$, $r^2 = 0.603$, F value = 13.672 with 2 components, and SEE = 0.382] (Table 2), but the predictive power of the model was not significant (low q^2
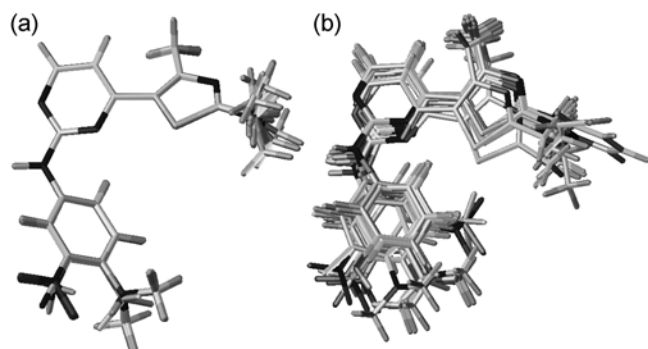

Figure 5. Ligand alignment by (a) atom based fit; (b) docking.

and r^2 values). In addition to LOO method to validate the CoMSIA model, a test set made up of 4 molecules (Fig. 4b) was used for model validation. Similar to cross-validated q^2 values of LOO method, the predictive performance of models on the test set was estimated by predictive r^2 value (0.445, Table 2), which is a measure of the predictivity of the model to predict biological activity of new derivatives accurately. The predictive r^2 value for the test set was lower than 0.5, which indicates the predictive power of the model is insignificant.

The structure-based CoMSIA model was also established of the same set of 21 molecules by using the docking-based alignment (Fig. 5b). Thus, ligand molecular superimposition on the template structure was performed by docking the ligands into the ATP-binding site of CDK2 by using our docking protocol (mutation-docking-mutation-minimization) described above (Fig. 5b). The statistically significant CoMSIA model was established of 21 molecules by using four different fields (steric, electrostatic, hydrogen bond donor, and hydrogen bond acceptor) [$q^2 = 0.501$, r^2 (non-cross-validated square of correlation coefficient) = 0.957, F value = 67.536 with five components, and SEE (standard error of

Table 2. CoMSIA analysis on 2-amino-4-(thiazol-5-yl)pyrimidine compounds by using four fields (steric, electrostatic, H-bond donor, and H-bond acceptor)

	q^2	N	r^2	SEE	F	Pred.r^2
Structure-Based	0.501	5	0.957	0.167	67.536	0.572
Ligand-Based	0.466	2	0.603	0.382	13.672	0.445

q^2-leave one out (LOO) cross-validated correlation coefficient, N-optimum number of components, r^2-non-cross-validated correlation coefficient, SEE-standard error of estimate, F-F-test value, Pred.r^2-CoMSIA predictive q^2 values on the test set.

Table 3. Conventional fit values on training set and prediction values on test set made by the structure-based CoMSIA model

Training Set							
	pK_i				pK_i		
Compd.	Actual	Pred.	Residual	Compd.	Actual	Pred.	Residual
1	8.31	8.12	0.19	**12**	9.00	8.84	0.16
2	7.80	7.81	−0.01	**13**	8.55	8.60	−0.05
3	8.77	8.44	0.33	**14**	8.47	8.54	−0.07
4	7.90	7.98	−0.08	**15**	9.30	9.34	−0.03
5	8.70	8.63	0.07	**16**	7.92	7.84	0.09
6	8.17	8.14	0.03	**17**	6.88	6.89	−0.01
7	7.96	8.06	−0.10	**18**	8.82	8.91	−0.09
8	8.80	8.86	−0.06	**19**	9.22	9.16	0.06
9	8.70	8.71	−0.01	**20**	9.22	9.36	−0.14
10	9.00	8.97	0.03	**21**	9.70	9.69	0.01
11	9.70	9.66	0.04				

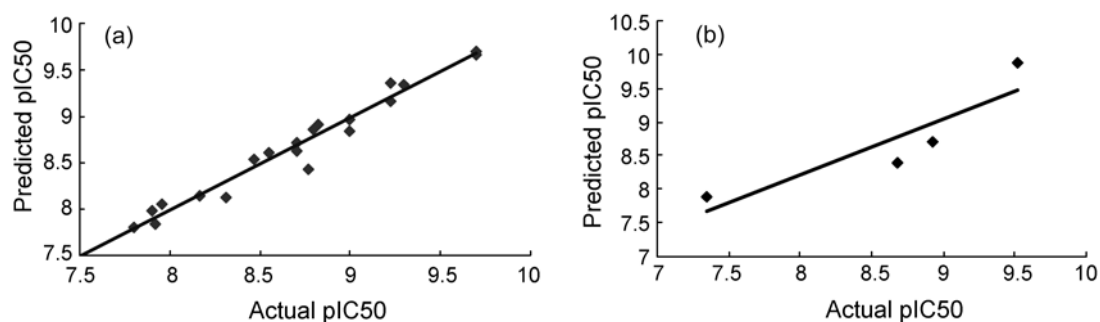| Test Set							
	pK_i				pK_i		
Compd.	Actual	Pred.	Residual	Compd.	Actual	Pred.	Residual
22	8.35	7.88	−0.53	**24**	9.52	9.87	−0.35
23	8.92	8.71	0.21	**25**	8.68	8.40	0.28

Figure 6. Relationship curve between observed values versus conventional fit values (prediction values) on the training set (a) and test set (b)
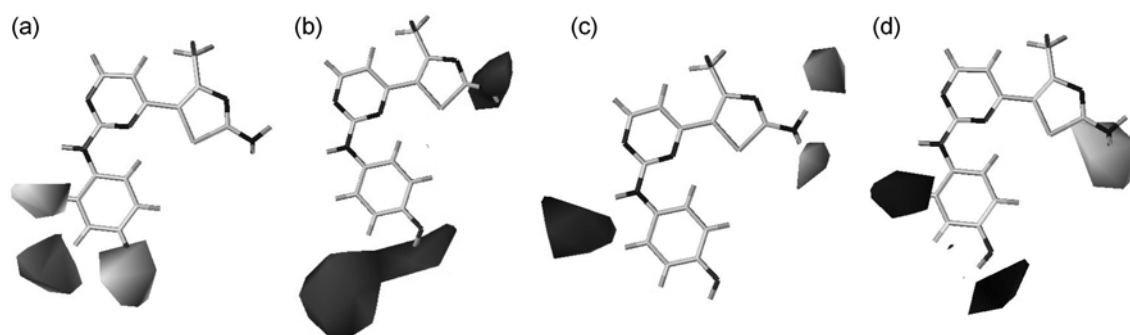

Figure 7. (a) Hydrophobic contour plot. Light gray contours and a dark gray contour indicate hydrophobicity favored and disfavored regions, respectively; (b) H-bond acceptor contour plot. The indicated contour represents hydrogen bond acceptor favored region; (c) H-bond donor contour plot. Light gray contours and a dark gray contour indicate hydrogen bond donor favored and disfavored regions, respectively; (d) Superposition of the steric and electrostatic contour plots. A light gray contour and dark gray contours indicate sterically disfavored and electrostatically favored regions, respectively.

estimate) = 0.167] (Table 2). The predictive performance of the structure-based CoMSIA model on the test set composed of 4 molecules was also estimated by predictive r^2 value (0.572, Table 2), which indicated that the built 3D-QSAR model was relatively reliable and able to predict biological activity of new derivatives (Table 2). Also, the predictive r^2 value for the test set was greater than 0.5, which indicates significant predictive power of the model.

The conventional fit values on training set and prediction values on the test set made by the structure-based CoMSIA model is shown in Table 3.

The relationship curve between observed values versus conventional fit values (prediction values) on the training set and the test set are also displayed in Figure 6. Comparing the statistical quality of the models shows that the structure-based approach yields better results in terms of fitting (r^2 and SEE value) and prediction (predictive q^2 value).

Thus, the comparison of the results of 3D QSAR with CoMSIA studies obtained from the same data set but using different alignment methods clearly shows that our docking study is able to reproduce the actual binding modes of ligand molecules at the ATP-binding site of CDK2. Also, it is noteworthy that the least active ligand molecules from the literature data (compounds with K_i values >20 μM)[29] failed to dock into the ATP-binding site and thereby could not be included in the ligand data set, which indicates that, in spite of a widened binding site due to mutation of four amino acid

residues, our docking method discriminates the nonbinders from the binders.

Graphical representations of CoMSIA maps of the structure-based model obtained by the field type "stDev*coeff" are displayed in Figure 7. The contour maps were superimposed on the most active compound **11** shown as a capped stick. Light gray contour region where hydrophobicity is associated with enhanced affinity (Fig. 7a) is located near the ortho and para positions of the aromatic ring, whereas the substitution at the meta position of the aromatic ring (dark gray region, Fig. 7a) with bulky substituents suggests decreased binding affinity of 2-amino-4-(thiazol-5-yl)pyrimidine analogues. The regions where hydrogen bond acceptor is associated with enhanced affinity are found near the para-hydroxy group of aromatic ring and amino group of thiazole ring (Fig. 7b). The thiazole amino group is also favored as a hydrogen bond donor, but, on the other hand, amino group bridging pyrimidine and aromatic rings disfavors hydrogen bond donor (Fig. 7c). Superposition of the CoMSIA steric and electrostatic contour plots (Fig. 7d) shows that the ortho and para positions of the aromatic ring favor substitution with electronegative substituents but the thiazole amino group needs to be unsubstituted for higher biological activity.

Conclusions

In a crystal structure, a protein conformation is optimally

adapted for interaction with one specific ligand, which makes it difficult to choose a basal protein structure for use in docking studies of conformationally flexible enzymes. There have been several efforts to tackle this problem but, due to the complexities of the specially designed approaches (docking multiple ligands to multiple enzyme structures),[3] there is an ongoing need for simple, intuitive, straightforward and easily reproducible method which can be generally used for docking study of conformationally flexible enzymes. In this study, by assuming the major conformational difference among the numerous CDK2 crystal structures lies at the side chains of four amino acid residues (Lys33, Phe80, Lys89, and Asp145) of the ATP-binding site, we used 'mutation, pharmacophore-guided docking, followed by mutation' strategy to generate an "average" CDK2 structure, which was used for ligand docking study to successfully reproduce 30 out of 32 X-ray ligand positions within 2.0 Å of heavy atom RMSD. Also, structure-based 3D QSAR study of a set of CDK2 inhibitors was compared with ligand-based 3D QSAR study to prove that our docking study is able to reproduce the actual binding modes of ligand molecules at the ATP-binding site of CDK2. Even though the docking study was performed on a limited number of CDK2-binders, our simple, intuitive and straightforward docking method which was found to reproduce the actual binding modes of various ligands warrants further application to the docking study of conformationally flexible enzymes as well as structure-based 3D-QSAR.

References

1. (a) Huse, M.; Kuriyan, J. *Cell* **2002**, *109*, 275. (b) Kim, J.; Lee, Y.; Kim, H.; Kang, S.-Y.; Park, K.-S.; Cho, J.-H.; Lee, Y.-Y.; Kim, B.-S.; Lim, Y.; Chong, Y. *Bull. Korean Chem. Soc.* **2005**, *26*, 2065.
2. Wu, S. Y.; McNae, I.; Kontopidis, G.; McClue, S. J.; McInnes, C.; Stewart, K. J.; Wang, S.; Zheleva, D. I.; Marriage, H.; Lane, D. P.; Taylor, P.; Fischer, P. M.; Walkinshaw, M. D. *Structure* **2003**, *11*, 399.
3. Thomas, M. P.; McInnes, C.; Fischer, P. M. *J. Med. Chem.* **2006**, *49*, 92.
4. Subramanian, J.; Sharma, S.; B-Rao, C. *J. Med. Chem.* **2006**, *49*, 5434.
5. 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.
6. Brown, N. R.; Noble, M. E.; Lawrie, A. M.; Morris, M. C.; Tunnah, P.; Divita, G.; Johnson, L. N.; Endicott, J. A. *J. Biol. Chem.* **1999**, *274*, 8746.
7. Gray, N. S.; Wodicka, L.; Thunnissen, A. M.; Norman, T. C.; Kwon, S.; Espinoza, F. H.; Morgan, D. O.; Barnes, G.; LeClerc, S.; Meijer, L.; Kim, S. H.; Lockhart, D. J.; Schultz, P. G. *Science* **1998**, *281*, 533.
8. Shewchuk, L.; Hassell, A.; Wisely, B.; Rocque, W.; Holmes, W.; Veal, J.; Kuyper, L. F. *J. Med. Chem.* **2000**, *43*, 133.
9. Meijer, L.; Thunnissen, A. M.; White, A. W.; Garnier, M.; Nikolic, M.; Tsai, L. H.; Walter, J.; Cleverley, K. E.; Salinas, P. C.; Wu, Y. Z.; Biernat, J.; Mandelkow, E. M.; Kim, S. H.; Pettit, G. R. *Chem. Biol.* **2000**, *7*, 51.
10. Arris, C. E.; Boyle, F. T.; Calvert, A. H.; Curtin, N. J.; Endicott, J. A.; Garman, E. F.; Gibson, A. E.; Golding, B. T.; Grant, S.; Griffin, R. J.; Jewsbury, P.; Johnson, L. N.; Lawrie, A. M.; Newell, D. R.; Noble, M. E.; Sausville, E. A.; Schultz, R.; Yu, W. *J. Med. Chem.* **2000**, *43*, 2797.
11. Davies, T. G.; Tunnah, P.; Meijer, L.; Marko, D.; Eisenbrand, G.; Endicott, J. A.; Noble, M. E. *Structure* **2001**, *9*, 389.
12. Davis, S. T.; Benson, B. G.; Bramson, H. N.; Chapman, D. E.; Dickerson, S. H.; Dold, K. M.; Eberwein, D. J.; Edelstein, M.; Frye, S. V.; Gampe Jr., R. T.; Griffin, R. J.; Harris, P. A.; Hassell, A. M.; Holmes, W. D.; Hunter, R. N.; Knick, V. B.; Lackey, K.; Lovejoy, B.; Luzzio, M. J.; Murray, D.; Parker, P.; Rocque, W. J.; Shewchuk, L.; Veal, J. M.; Walker, D. H.; Kuyper, L. F. *Science* **2001**, *291*, 134.
13. Dreyer, M. K.; Borcherding, D. R.; Dumont, J. A.; Peet, N. P.; Tsay, J. T.; Wright, P. S.; Bitonti, A. J.; Shen, J.; Kim, S. H. *J. Med. Chem.* **2001**, *44*, 524.
14. Gibson, A. E.; Arris, C. E.; Bentley, J.; Boyle, F. T.; Davies, N. J.; Curtin, T. G.; Endicott, J. A.; Golding, B. T.; Grant, S.; Griffin, R. J.; Jewsbury, P.; Johnson, L. N.; Mesguiche, V.; Newell, D. R.; Noble, M. E.; Tucker, J. A.; Whitfield, H. J. *J. Med. Chem.* **2002**, *45*, 3381.
15. Beattie, J. F.; Breault, G. A.; Ellston, R. P. A.; Green, S.; Jewsbury, P. J.; Midgley, C. J.; Naven, R. T.; Minshull, C. A.; Pauptit, R. A.; Tucker, J. A.; Pease, J. E. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2955.
16. Davies, T. G.; Bentley, J.; Arris, C. E.; Boyle, F. T.; Curtin, N. J.; Endicott, J. A.; Gibson, A. E.; Golding, B. T.; Griffin, R. J.; Hardcastle, I. R.; Jewsbury, P.; Johnson, L. N.; Mesguiche, V.; Newell, D. R.; Noble, M. E. M.; Tucker, J. A.; Wang, L.; Whitfield, H. J. *Nat. Struct. Biol.* **2002**, *9*, 745.
17. Clare, P. M.; Poorman, R. A.; Kelley, L. C.; Watenpaugh, K. D.; Bannow, C. A.; Leach, K. L. *J. Biol. Chem.* **2001**, *276*, 48292.
18. Furet, P.; Meyer, T.; Strauss, A.; Raccuglia, S.; Rondeau, J. M. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 221.
19. Bramson, H. N.; Corona, J.; Davis, S. T.; Dickerson, S. H.; Edelstein, M.; Frye, S. V.; Gampe, Jr., R. T.; Harris, P. A.; Hassell, A.; Holmes, W. D.; Hunter, R. N.; Lackey, K. E.; Lovejoy, B.; Luzzio, M. J.; Montana, V.; Rocque, W. J.; Rusnak, D.; Shewchuk, L.; Veal, J. M.; Walker, D. H.; Kuyper, L. F. *J. Med. Chem.* **2001**, *44*, 4339.
20. Sayle, K. L.; Bentley, J.; Boyle, F. T.; Calvert, A. H.; Cheng, Y.; Curtin, N. J.; Endicott, J. A.; Golding, B. T.; Hardcastle, I. R.; Jewsbury, P.; Mesguiche, V.; Newell, D. R.; Noble, M. E. M.; Parsons, R. J.; Pratt, D. J.; Wang, L. Z.; Griffin, R. J. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3079.
21. Anderson, M.; Beattie, J.; Breault, G.; Breed, J.; Byth, K.; Culshaw, J.; Ellston, R.; Green, S.; Minshull, C.; Norman, R.; Pauptit, R.; Stanway, J.; Thomas, A.; Jewsbury, P. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3021.
22. Liu, J.-J.; Dermatakis, A.; Lukacs, C. M.; Konzelmann, F.; Chen, Y.; Kammlott, U.; Depinto, W.; Yang, H.; Yin, X.; Chen, Y.; Schutt, A.; Simcox, M. E.; Luk, K.-C. *Bioorg. Med. Chem.* **2003**, *13*, 2465.
23. Moshinsky, D. J.; Bellamacina, C. R.; Boisvert, D. C.; Huang, P.; Hui, T.; Jancarik, J.; Kim, S. H.; Rice, A. G. *Biochem. Biophys. Res. Comm.* **2003**, *310*, 1026.
24. Hindle, S. A.; Rarey, M.; Buning, C.; Lengaue, T. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 129.
25. Gästeiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219.
26. Purcell, W. P.; Singer, J. A. *J. Chem. Eng. Data* **1967**, *12*, 235.
27. Blaney, J. M.; Weiner, P. K.; Dearing, A.; Kollman, P. A.; Jorgensen, E. C.; Oatley, S. J.; Burridge, J. M.; Blake, C. C. F. *J.*

Am. Chem. Soc. **1982**, *104*, 6424.

28. Wipff, G.; Dearing, A.; Weiner, P. K.; Blaney, J. M.; Kollman, P. A. *J. Am. Chem. Soc.* **1983**, *105*, 997.

29. Wang, S.; Meades, C.; Wood, G.; Osnowski, A.; Anderson, S.; Yuill, R.; Thomas, M.; Mezna, M.; Jackson, W.; Midgley, C.; Griffiths, G.; Fleming, I.; Green, S.; McNae, I.; Wu, S.-Y.; McInnes, C.; Zheleva, D.; Walkinshaw, M. D.; Fischer, P. M. *J. Med. Chem.* **2004**, *47*, 1662.

30. Wold, S.; Albano, C.; Dunn, W.; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjosstrom, M. *Mathematics and Statistics in Chemistry* in *Chemometrics*; Kowalski, B., Ed.; Reidel: Dordrecht, 1987.

31. Sims, P. A.; Wong, C. F.; McCammon, J. A. *J. Med. Chem.* **2003**, *46*, 3314.

32. Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. *J. Med. Chem.* **2004**, *47*, 45.

33. Zavodszky, M. I.; Kuhn, L. A. *Protein Sci.* **2005**, *14*, 1104.

34. Barril, X.; Morley, S. D. *J. Med. Chem.* **2005**, *48*, 4432.

35. Sciabola, S.; Carosati, E.; Baroni, M.; Mannhold, R. *J. Med. Chem.* **2005**, *48*, 3756.

36. Bohm, M.; Sturzebecher, J.; Klebe, G. *J. Med. Chem.* **1999**, *42*, 458.

37. Klebe, G. *Comparative Molecular Similarity Indices: CoMSIA* in *3D QSAR in Drug Design;* Kubinyi, H.; Folkers, G.; Martin, Y. C., Eds.; Kluwer Academic Publishers: Great Britain, 1998; Vol 3, p 87.

38. (a) Cramer III, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959. (b) Kim, J.; Lee, M.; Kang, S.-Y.; Park, J.; Lim, Y.; Koh, D.; Park, K. H.; Chong, Y. *Bull. Korean Chem. Soc.* **2006**, *27*, 1025.