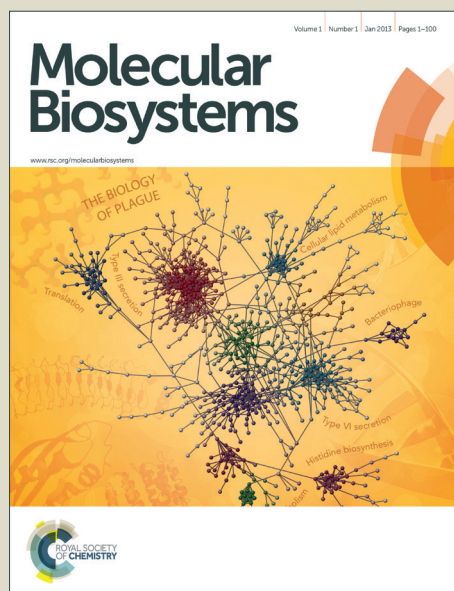


Molecular BioSystems

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



www.rsc.org/molecularbiosystems

ARTICLE

Studies of N⁹-arenthenyl purines as novel DFG-in and DFG-out dual Src/Abl inhibitors using 3D-QSAR, docking and molecular dynamics simulations

Cite this: DOI: 10.1039/x

Received 00th January 2012,
Accepted 00th January 2012

DOI: MB-ART-06-2014-000350

www.rsc.org/

Shaojie Ma,^a Guohua Zeng,^a Danqing Fang,^b Juping Wang,^a Wenjuan Wu,^{*a}
Wenguo Xie,^a Shepei Tan^a and Kangcheng Zheng^c

Recently, the development of Src/Abl (c-Src/Bcr-Abl tyrosine kinases) dual inhibitors attracts extensive attention of research for treatment of malignancies. In order to explore the difference of structural features impacting the Src and Abl activities of N⁹-arenthenyl purines and to investigate the molecular mechanisms of ligand-receptor interactions, a molecular modeling study combining the three-dimensional quantitative structure-activity relationship (3D-QSAR), molecular docking and molecular dynamics (MD) simulations was performed. The obtained CoMFA (comparative molecular field analysis) models exhibited satisfactory internal and external predictability. The plots of the CoMFA fields could explain the structural differences between DFG-in (targeting the active enzyme conformation) and DFG-out (targeting the inactive enzyme conformation) inhibitors. The key amino acid residues were identified by docking studies, and the detailed binding modes of the compounds with different activities were determined by MD simulations. The binding free energies gave a good correlation with the experimental activity. An energetic analysis, the MM-PBSA (molecular mechanics Poisson-Boltzmann Surface) energy decomposition revealed that the van der Waals interaction was the major driving force for the binding of compounds to DFG-in and DFG-out Src and Abl, especially the hydrophobic interactions between ligands and residues Ala403/380, Asp404/381, and Phe405/382 in DFG-out Src and Abl complexes help to stabilize the DFG-out conformations. These results can offer useful references for designing novel potential DFG-in and DFG-out dual Src/Abl inhibitors.

Introduction

As a non-receptor tyrosine kinase, c-Src proto-oncogene plays an important role in multiple signaling pathways relating with cell development, growth, progression, metastasis, and apoptosis. There is experimental evidence that the overexpression or hyperactivation of c-Src kinase is associated with many human malignancies, including colon, breast, pancreatic, lung, and brain carcinomas.^{1,2} As well as having a role in solid tumors, c-Src is also involved in the progression of chronic myeloid and acute lymphoid (CMLs and ALLs) that are positive for the Philadelphia chromosome (Ph+).³ Bcr-Abl is a constitutively activated cytoplasmic tyrosine kinase encoded by the Philadelphia chromosome, which derives from a reciprocal translocation between chromosomes 9 (Abelson oncogene, Abl) and 22 (breakpoint cluster region gene Bcr).⁴ The finding that Bcr-Abl is the causative event of chronic myeloid Leukemia (CML) and that the Abl activity is fundamental Bcr-Abl-mediated transformation makes this kinase an important target for the development of specific therapies.⁵⁻⁷ However, due to

the development of resistance, such as resistance to imatinib, a Bcr-Abl kinase inhibitor and the first choice drug for CML therapy, the search for new drugs capable of circumventing clinically relevant resistant mutations has captured much attention.^{8,9} Moreover, recent data have demonstrated that c-Src kinases have a function in imatinib-resistant CML and ALL^{10,11} and that compounds inhibiting both Src and Abl might be useful in the treatment of patients who have relapsed on imatinib.

In recent years, even if the fact that the targeted therapy era started as a hunt for selective kinase inhibitors, but the research aim has now changed to the identification of compounds acting on dual or multiple targets in order to overcome the drug resistance.¹² Dual or multi-targeted compounds, usually inhibiting different cell pathways or compensatory mechanisms, could be more effective than selective inhibitors, especially in tumors. Moreover, since Src shares significant sequence homology with Abl, several Src inhibitors showed also potent Bcr-Abl inhibitory activity and were successfully used as

antileukemia drugs, and thus the search for dual or multi-kinase inhibitors is a very attractive field. So far, the most recent and important dual Src/Abl inhibitors combating imatinib resistance include dasatinib, bosutinib, AP23464, PD166326, AZD0530, and CGP70630.¹³⁻¹⁵ Recently, Huang and coworkers have synthesized a novel series of N⁹-arenethenyl purine derivatives and assessed their activities. They discovered that these compounds have potent inhibitory activities on c-Src and Abl enzymes at low nanomolar concentrations and potentially inhibit several human cancer cell lines, demonstrating the great potential of developing N⁹-arenethenyl purine derivatives as a novel class of dual Src/Abl inhibitors for cancer therapy.^{16, 17} Dependent on the different binding interactions with the kinase domains of two receptors, these dual Src/Abl inhibitors can be divided into two chemical series, DFG-in and DFG-out inhibitors.¹⁸ The former binds with kinase domain in an active conformation, which is adjacent to the active ATP pocket whereas the latter binds to the inactive conformations of Src and Abl, associated with the conformational transforms in the activation loop Asp-Phe-Gly (DFG) residues promoted by the inhibitor.¹⁹ Although some progresses have been made in experimental researches, so far the theoretical studies on the mechanisms of these compounds toward DFG-in and DFG-out Src and Abl kinases and the structural features influencing their anticancer activities remain largely unknown.

To predigest the drug discovery process, it is a very significant work to carry out detailed theoretical studies on the mechanisms of ligand-receptor interactions.²⁰ 3D-QSAR model, especially, the popular comparative molecular field analysis (CoMFA) has been successfully used in understanding the pharmacological properties of the studies compounds and modern drug design, because not only CoMFA model is visualized but also the obtained steric and electrostatic maps may help to understand the detail 3D-structure of active site of receptor.^{21,22} Meanwhile, molecular docking is an approach to predict the possible orientations of ligand in the active pocket of receptor to study the interaction mechanisms.²³ In addition, molecular dynamics simulation is a useful methodology providing vivid pictures to depict the fluctuations and conformational changes of molecules, and further investigate the interaction mechanism of protein complex with ligands at the atomic level.^{24,25} Therefore, a combined 3D-QSAR, molecular docking and MD simulation study can offer the deep insight into understanding the structural features of ligand-receptor interactions.^{26,27}

In this work, we focused on a novel series of N⁹-arenethenyl purine derivatives acting as DFG-in and DFG-out dual Src/Abl inhibitors to perform a molecular modeling study by using molecular docking, MD simulation and 3D-QSAR CoMFA analysis. The optimum 3D-QSAR CoMFA models were developed and the key structural features contributing to the inhibitory activities were also identified. The stability and predictive ability of the constructed CoMFA models were estimated with internal and external validation. Moreover, the rational conformations and detailed interactions for these

compounds interacting with Src and Abl kinases in both DFG-in and DFG-out binding modes were analyzed based on the results from molecular docking and MD simulations. We expect the obtained results can help to understand the binding process and provide useful information for the development of novel potent dual Src/Abl inhibitors, especially DFG-out inhibitors.

Materials and methods

Data set

A set of 49 N⁹-arenethenyl purine derivatives with well-expressed inhibitory activities against dual Src and Abl receptors were taken from the literature^{16,17} to perform this study. The general structural formulae of the studied compounds are displayed in Fig. 1. These compounds consisted of two classes: DFG-in (compounds 1-24 in Table S1, ESI†) and DFG-out inhibitors (compounds 25-49 in Table S1, ESI†). Each class was divided into a training set for model generation and a test set for model validation, containing 17 and 7 compounds for DFG-in series, and 18 and 7 compounds for DFG-out class, respectively. The test compounds were selected manually considering the structural diversity and wide range of activities in the data set.²⁸ All original IC₅₀ values were converted to pIC₅₀ values and used as dependent variables in the CoMFA study.

The 3D-structures of purine derivatives were constructed by the sketch molecule module in Sybyl 6.9 software. Structural energy minimization was performed using Powell gradient algorithm and the tripos force field with a convergence criterion of 0.001 kcal/(mol·Å) and a maximum of 1000 iterations. MMFF94 charges were assigned to each compound. The minimized structure was used as the initial conformation for molecular docking.

Molecular docking

To locate the probable binding conformations and orientations of these purine derivatives interacting with both Src and Abl kinases, docking studies were performed with the surflex module in Sybyl 6.9 software package.^{29,30} The four X-ray crystal structures of Src and Abl (PDB entry: 2BDJ and 3KF4 for DFG-in Src and Abl, 3G6H and 3KFA for DFG-out Src and Abl) were obtained from the Protein Data Bank and used to dock. Prior to docking, all the water molecules and ligands were extracted, the polar hydrogen-atoms and Kollman all atom charges were added to the proteins. In the current study, the protomol were generated by the ligand-based mode, with two important parameters, i.e., protomol threshold and protomol bloat at their default values of 0.5 and 0, respectively. With the other defaulted parameters used, 20 conformations per ligand were produced by Surflex-dock method, and the binding conformation with the highest docking score and the orientation of the conformation being similar with that of ligand was selected for further 3D-QSAR studies.³¹ In the molecular docking, the ligand compounds were considered to be flexible and the proteins were regarded as being rigid.

Molecular modeling and alignment

ARTICLE

CoMFA studies were performed by using SYBYL 6.9 molecular modeling software package running on an SGI R2400 workstation. All parameters used in CoMFA were default except for explained.

Active conformation selection and structural alignment of these compounds are key steps for 3D-QSAR analysis.^{32,33} To derive the optimal 3D-QSAR models, two different alignment methods were employed. The first one is ligand-based alignment, i.e., all compounds were aligned to the most active compound **20** (or compound **32** for DFG-out series) by the Align Database command in SYBYL 6.9 software. The other one is receptor-based alignment. In this process, the optimal conformations of all compounds derived from docking studies were assigned MMFF94 charges and imported into molecular alignment for 3D-QSAR analysis. The common skeleton (the atoms numbered from 1 to 9) shown in Fig. 1.

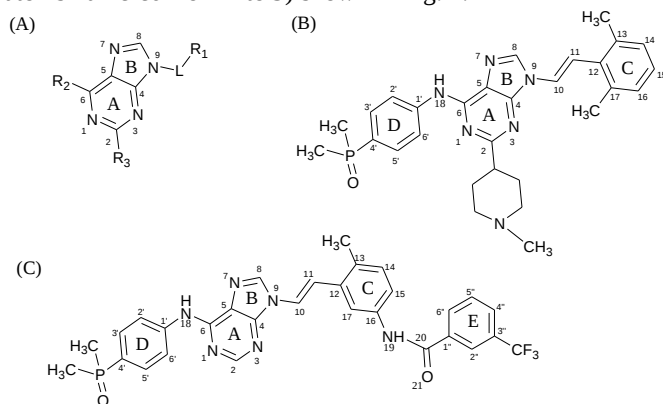


Fig. 1 General structural formula and numbering of purine derivatives (A) and template molecules (B, compound **20** and C, compound **32**).

Generation of CoMFA model

In CoMFA analysis, models of steric and electrostatic fields were based on both Lennard-Jones and Coulombic potentials.^{34,35} The steric and electrostatic energies were calculated using Tripos force field with a dielectric constant of 80, an sp³ carbon atom with Vander Waals radius of 1.52 Å, +1 charge, and 2 Å grid spacing. The truncation for both the steric and the electrostatic energies was set to 30 kcal/mol.^{36,37}

Partial least squares (PLS) analysis and validation of QSAR models

To generate statistically significant 3D-QSAR CoMFA models, the partial least-squares (PLS) statistical method was used to correlate the CoMFA interactions fields to the pIC₅₀ values. In the process of PLS algorithm, the leave-one-out (LOO) cross-validation method

was exploited to reduce the cross-validation correlation coefficient (q^2) and the optimum number of components N . The non-cross-validation analyses were evaluated by the correlation coefficient (R^2), standard error of estimates (SEE), and F value.³⁸ To further assess the robustness and statistical validity of the derived models, bootstrapping analysis for 100 runs was also applied.³⁹

To assess the predictive abilities of 3D-QSAR models generated from the training set, the biological activities of compounds in two external test sets were predicted. The predictive ability of the model is expressed by the predictive correlation coefficient (R^2_{pred}) calculated by the formula: $R^2_{pred} = (SD - PRESS) / SD$, where SD is the sum of the squared deviations between the actual activities of the test set compounds and the mean activity of the training set compounds, and $PRESS$ is the sum of the squared deviations between the actual and predicted activities of the test set compounds.^{40,41}

Molecular dynamics simulations

To confirm the docking results, the MD simulations were performed with AMBER 9.0 software package.^{42,43} The docked complexes of 2BDJ and 3KF4 with **20** and **4** as well as 3G6H and 3KFA with **32** and **40**, respectively, were used as the initial structures for MD simulations. The electrostatic potentials (ESP) of the ligands were calculated at B3LYP/6-31G(d) level in the Gaussian 09 program^{44,45} and the partial atomic charges for ligand atoms were assigned by using the RESP protocol implemented in the Antechamber module. The FF03 AMBER force field and the general AMBER force field (GAFF) were respectively used to describe the protein and the ligands.⁴⁶⁻⁴⁸ Each complex was neutralized by adding sodium ions and solvated in a box of TIP3P water molecules with a margin distance of 12 Å.^{49,50}

Then, two-stage energy minimizations were carried out to avoid possible steric stress. Firstly, each complex was fixed with restraint constant of 2.0 kcal mol⁻¹ Å⁻² and the waters and sodium ions were minimized with steepest descent (SD) method for 2000 steps followed by conjugated gradient (CG) method for 3000 steps. Secondly, the whole relaxed complex was optimized by 5000 steps steepest descent minimization and 5000 steps conjugated gradient minimization. Then, the systems were gradually heated from 0 to 300 K in 200 ps with a weak constraint of 1.0 kcal mol⁻¹ Å⁻² at constant volume and equilibrated for 500 ps at 300 K and 1 atm. Finally, 10 ns production MD simulation was performed in a NPT (constant composition, $T = 300$ K and $P = 1.0$ atm) ensemble. During the simulation, the particle mesh Ewald (PME) method was used to treat the long-range electrostatic interactions with non-bonded cutoff of 8.0 Å,⁵¹ and SHAKE algorithm was applied to constrain all covalent bonds involving hydrogen atoms with 2

fs time step.⁵² The coordinated trajectories were saved every 1 ps for further binding free energy calculation and decomposition analysis.

Binding Free Energy Calculations

The binding free energies of the ligand–protein were calculated by the MM-PBSA procedure encoded in AMBER 9.0 software.^{53–56} For each system, a total of 200 snapshots of the simulated structure extracted from the last 2 ns stable MD trajectory were used for the calculations. The binding free energy (ΔG_{bind}) is calculated as follows:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}}) = \Delta G_{\text{MM}} + \Delta G_{\text{sol}} - T\Delta S \quad (1)$$
 where ΔG_{MM} is the molecular mechanics free energy, ΔG_{sol} is the solvation free energy, and $T\Delta S$ is the entropy contribution. The ΔG_{MM} contains the van der Waals energy (ΔG_{vdw}) and electrostatic (ΔG_{ele}) energy:

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

The ΔG_{sol} is the sum of electrostatic solvation, including the polar solvation free energy ($\Delta G_{\text{ele,sol}}$) and the nonpolar solvation free energy ($\Delta G_{\text{nonpol,sol}}$):

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

The $\Delta G_{\text{ele,sol}}$ was determined by Poisson Boltzmann (PB) equation with the dielectric constant for solute and solvent set to 1.0 and 80.0, respectively. The $\Delta G_{\text{nonpol,sol}}$ was determined by using:

$$\Delta G_{\text{nonpol,sol}} = \gamma \times \text{SASA} + \beta \quad (4)$$

where γ , standing for the surface tension, and β , being the offset value constant, were set to 0.0072 kcal/(mol Å²) and 0, respectively. SASA is the solvent accessible surface area that was calculated using the Molsurf module in AMBER 9.0. As the calculation of entropy term was time-consuming and its value seldom converge, the entropy contribution has been omitted in this study.^{57–59}

For discerning the difference of the binding modes of two classes inhibitors, the binding free energies were decomposed to each residue using MM/GBSA (molecular mechanics/generalized Born Surface area) method. Each inhibitor-residue pair includes four energy terms: van der Waals contribution (ΔG_{vdw}), electrostatic contribution (ΔG_{ele}), polar solvation contribution ($\Delta G_{\text{ele,sol}}$) and nonpolar solvation contribution ($\Delta G_{\text{nonpol,sol}}$), which can be summarized as the following equation:

$$\Delta G_{\text{inhibitor-residue}} = \Delta G_{\text{vdw}} + \Delta G_{\text{ele}} + \Delta G_{\text{ele,sol}} + \Delta G_{\text{nonpol,sol}} \quad (5)$$

where ΔG_{vdw} and ΔG_{ele} were calculated with Sander program. The polar contribution was determined by the generalized Born model (GB^{OC}, igb = 2)⁶⁰ and the nonpolar part was computed using the solvent accessible surface area (SASA).⁶¹

Results and Discussion

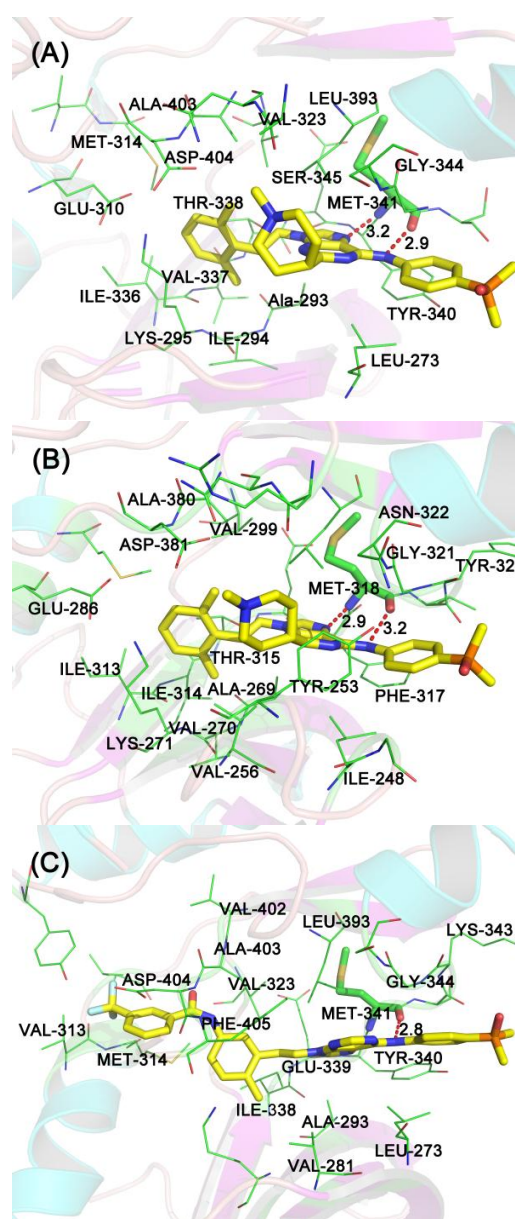
Molecular Docking

Before docking, it is indispensable to verify the reliability of the docking project. The four native ligands were extracted from the X-ray structures of Src (PDB entry: 2BDJ, 3G6H) and Abl (PDB entry: 3KF4, 3KFA) and redocked into their corresponding protein pockets.

As a result, the RMSD (root means square derivation) values between the optimal docked poses and the crystallized ones were 0.698 Å, 0.599 Å, 0.512 Å and 0.404 Å, respectively, suggesting that the docking procedure was reliable enough for this system.

All the studied compounds were successfully docked into the corresponding binding pockets of Src and Abl. In order to elucidate the interaction mechanisms, compounds **20** and **32**, the most potential inhibitors of DFG-in and DFG-out classes were selected for the further analyses.

For DFG-in binding model. The binding models of compound **20** docked into c-Src (2BDJ) and Abl (3KF4) were displayed in Fig. 2A and B. It can be found that compound **20** is suitably situated at the ATP binding site adopted similar poses for two enzymes. In the complex structures, the dimethyl phosphinoxide (DMPO)-phenyl ring is orientated almost coplanar to the purine core whereas the plane of the dimethylphenyl is nearly perpendicular to the core.



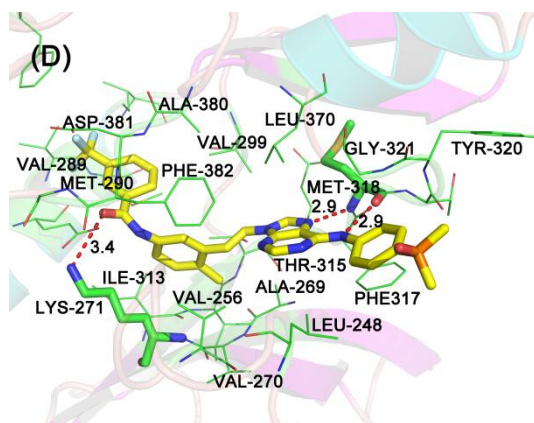


Fig. 2 Docking modes of compounds **20** and **32** in the binding sites of Src-in (2BDJ, A), Abl-in (3KF4, B), Src-out (3G6H, C) and Abl-out (3KFA, D), in which the red and blue regions represent oxygen and nitrogen atoms, respectively, whereas white regions indicate carbon or hydrogen atoms. Hydrogen bonds are depicted as red dotted line.

The substituent R_1 penetrates deeply into a large hydrophobic pocket, in Van der Waals contact with Ala293/269, Ile294/314, Glu310/286, Ile336/313, Val337/270, Thr338/315 and Ala403/380 for Src-in or Abl-in (in Fig. 2). The side chain of substituent R_2 at the entrance of the active pocket created by Tyr340/Phe317, and Gly344/321 is placed in the solvent accessible region, especially the DMPO moiety, which is largely solvent-exposed and makes almost no contacts with two proteins. At the same time, the substituent R_3 which is blocked by the side chain of Ser345/Asn322, limiting its prolongation. In addition, the N_7 atom of purinerine and the NH group of aniline ring at C_6 -position can form hydrogen bonds with the NH backbone and the carbonyl oxygen of Met341/318 for Src-in and Abl-in, respectively.

For DFG-out binding model. As illustrated in Fig. 2C and D, compound **32** binds to Src-out (3G6H) and Abl-out (3KF4) in an extended conformation. While the purine template and the methylphenyl group are largely co-planar with the DMPO-phenyl ring and the trifluoromethylphenyl group, respectively, the two chemical groups themselves lie almost perpendicular to each other. As with compound **20**, the purine template of compound **32** occupied the adenine binding pocket and the substituent R_1 is bound in the selectivity pocket. The trifluoromethylphenyl moiety of substituent R_1 is buried completely in a small hydrophobic pocket vacated by Phe405/382 of the DFG-motif and in Van der Waals contact with the side chains of Val313/289, Met314/290 and Asp404/381 for Src-out or Abl-out. On the other hand, the aniline ring of the substituent R_2 binds in a pocket near the hinge and is in close contact with the Ser342/Thr319, Lys343/Tyr320 and Gly344/321, the DMPO moiety is also largely exposed to solvent. Finally, compound **32** can form only one hydrogen bond with Src-out, from the NH group of substituent R_2 to the carbonyl oxygen of Met341, whereas its purine ring moved toward to Met318 and formed three hydrogen bonds with Abl-out, from the N_7 and the NH of substituent R_2 to the NH backbone and the carbonyl oxygen of

Met341, and the third one is made between the carboxyl oxygen of substituent R_1 and the NH of Lys271. Compared Src-out with Abl-out, the hydrogen bond interaction and conformation have slightly changed. The detailed interactions will be further investigated and discussed in the following section.

Comparison between DFG-in and DFG-out conformations. Compounds **20** and **32** share a common purine template, a vinyl linkage and a DMPO-phenylamine group at C_6 , but they are bound to Src and Abl kinases in different binding modes, namely DFG-in and DFG-out. In order to ascertain the division justification of these two binding models and direct the design of novel potential dual Src/Abl inhibitors, the structure alignments for two different conformations of both compounds are performed (Fig. 3). As can be seen in Fig. 3, the binding positions of these common groups in both inhibitors are found to be nearly identical.

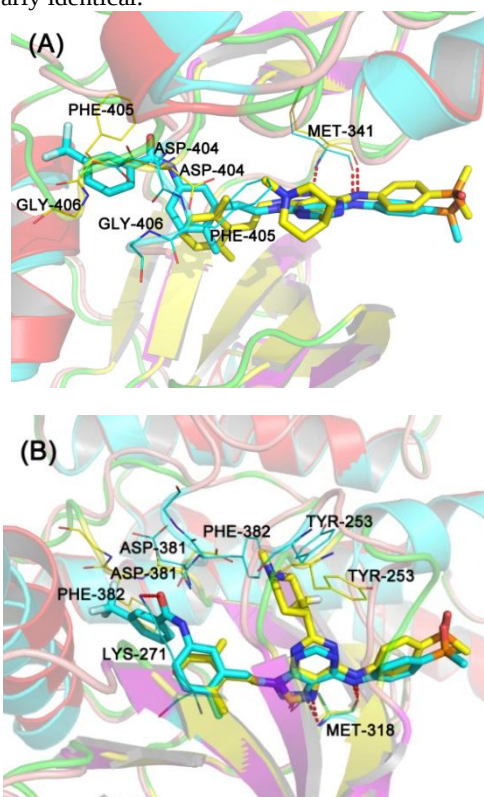


Fig.3 Structural comparison for DFG-in and DFG-out series: (A) overlay of Src-in (2BDJ):**20** (yellow) with Src-out (3G6H):**32** (cyan), (B) overlay of Abl-in (3KF4):**20** (yellow) and Abl-out (3KFA):**32** (cyan). Hydrogen bonds are depicted as red dotted line.

Moreover, the dimethylphenyl group in **20** binds in almost the same position as methylphenyl group in **32**, whereas the trifluoromethylphenyl moiety occupies a hydrophobic pocket vacated by Phe405/382 of the DFG-motif as kinase conformation switches from DFG-in to DFG-out conformation. The activation loop in DFG-in conformation makes sufficient space to accommodate substituent R_3 of compound **20**, which may make collision with Phe405/382 in DFG-out conformation.

Similarly, the trifluoromethylphenyl of **32** may also be steric clashed by the DFG-in loop.

As the different conformations of DFG-in and DFG-out, compound **32** moves slightly toward to Met341/318, and makes more hydrogen bonds to the protein than **20**. Meanwhile, Phe405/382 binds near the adenine binding site and makes multiple vdw interactions with **32** but forms no such interactions with **20**, due to the far distance. On the contrary, Tyr253 in 3KF4:**20** is positioned near the bound inhibitor and makes extensive vdw contacts with **20**, but it can only stabilize the DFG-out conformation in 3KFA:**32**, because it is far away from **32**. Through the structural comparisons, we can find that the differing molecular recognition exists in DFG-in and DFG-out inhibitors. These differences are evident in the activation loop conformations of the different structures, and particularly in the differing interactions of Phe405/382 and Tyr253 with inhibitors.

CoMFA statistical results

Statistically significant CoMFA models were generated based on the training set followed by validation with the test set, and their statistical parameters are summarized in Table 1.

For all four targets, the correlation coefficients based on training (R^2 in Table1) and test set compounds (q^2) were 0.999 and 0.699, 0.995 and 0.614, 0.995 and 0.562, and 0.996 and 0.629 for Src-in, Abl-in, Src-out and Abl-out kinases, respectively. The bootstrapping results have R^2_{bs} of 1.000, 0.999, 0.999 and 0.998, and SD_{bs} of 0.000, 0.001, 0.001 and 0.002 for these four models, respectively, suggesting that good internal consistencies exist within the two training sets. Furthermore, the predicted R^2_{pred} for the two test sets were 0.765, 0.668, 0.614 and 0.662, respectively, indicating that these CoMFA models were reliable. In the built CoMFA models, we can also find that the contributions of the steric and electrostatic fields are almost the same for Src-in and Src-out kinases, indicating that both the steric and electrostatic fields have important influences on the interaction of the ligand and Src receptor, but in Abl-in and Abl-out studies, the steric feature is found making larger contribution than the electrostatic one to Abl inhibitory activity.

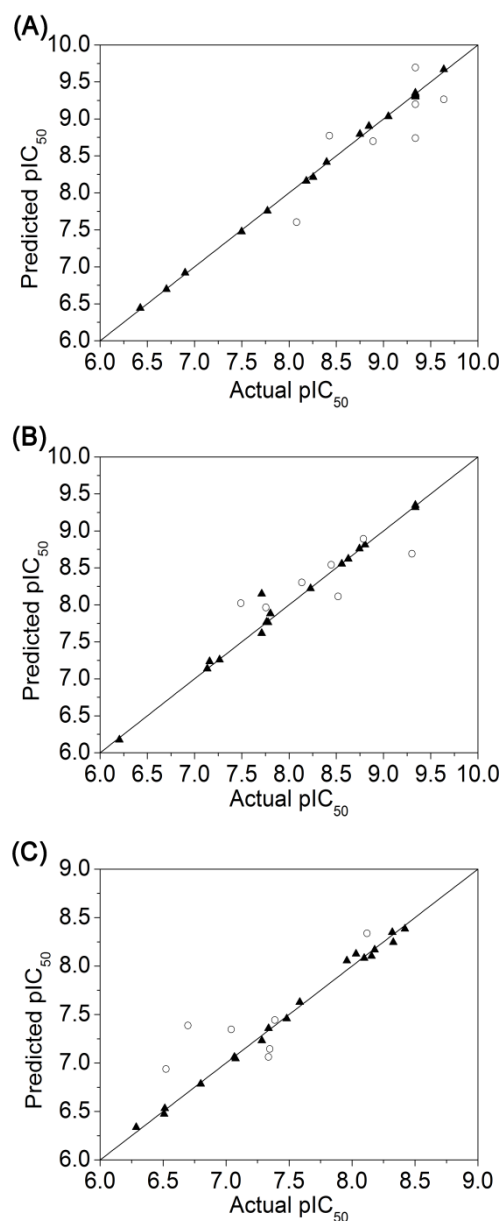
Table 1 Statistical results of CoMFA models

	DFG-in		DFG-out	
	Src	Abl	Src	Abl
PLS statistic	0.999	0.995	0.995	0.996
R^2				
N	6	5	1	4
q^2	0.699	0.614	0.562	0.629
SEE	0.036	0.078	0.061	0.036
F	2139.768	346.378	365.353	428.007
R^2_{bs}	1.000	0.999	0.999	0.998
SD_{bs}	0.000	0.001	0.001	0.002
R^2_{pred}	0.765	0.668	0.614	0.662
Field distribution %				
S	54.5	68.6	45.4	65.4
E	45.5	31.4	54.6	34.6

Note: N is the optimal number of components, q^2 is the square of

leave-one-out (LOO) cross-validation coefficient, R^2 is the square of non-cross-validation coefficient, SEE is the standard error of estimation, F is the F -test value, R^2_{bs} is the mean R^2 of bootstrapping analysis (100 runs), SD_{bs} is the mean standard deviation by bootstrapping analysis. Abbreviations; S(steric), E(electrostatic).

The predicted pIC_{50} values and the residual values of compounds for four CoMFA models are listed in Table S2 (ESI†). The plots of the predicted pIC_{50} values versus the experimental ones are displayed in Fig. 4, in which most points are evenly distributed along the line $Y = X$, suggesting that the CoMFA models have good quality.



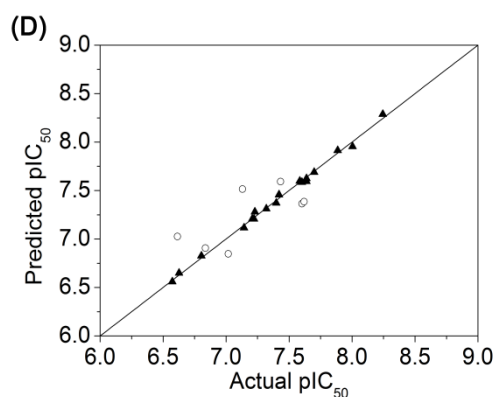


Fig. 4 Plots of the predicted versus actual values using the training set (triangle) and test set (dot) based on the CoMFA models. (A) Src-in, (B) Abl-in, (C) Src-out, (D) Abl-out.

CoMFA contour maps analyses

DFG-in series. For each enzyme, the contour maps for two CoMFA fields were generated and analyzed.

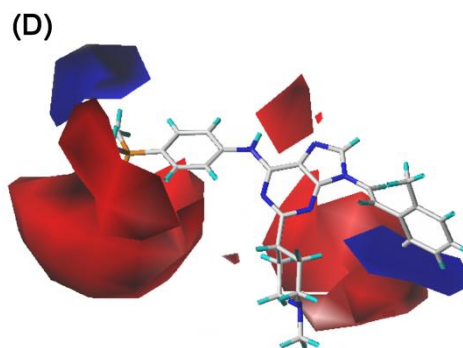
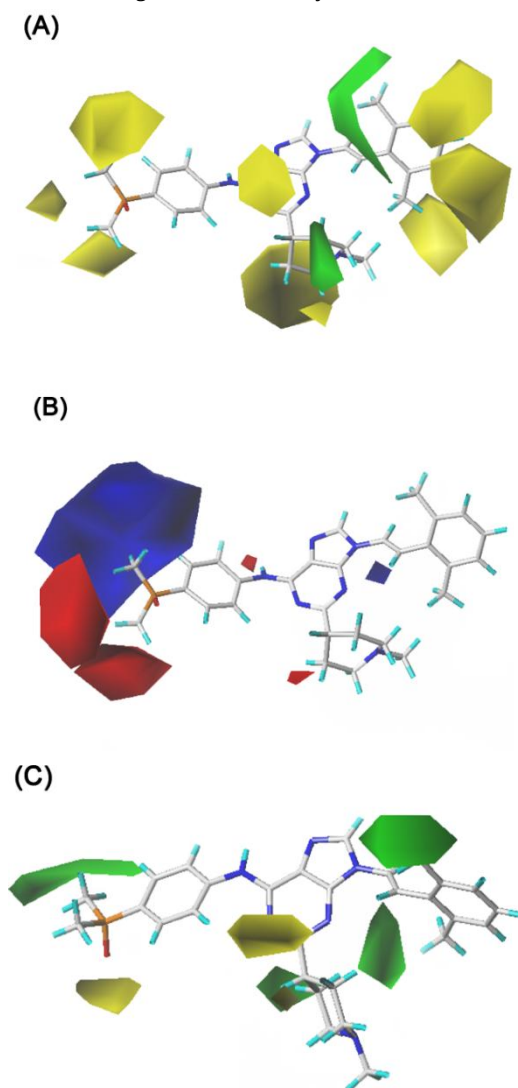


Fig. 5 CoMFA contour maps of the highly active compound **20** in the DFG-in series. (A) Steric contour map for Src, (B) Electrostatic contour map for Src, (C) Steric contour map for Abl, (D) Electrostatic contour map for Abl.

Fig. 5A shows the sterically favorable (green) and unfavorable (yellow) regions for Src-in model with molecule **20** as a reference structure. There is a big green contour near the C₁₃- and C₁₇-positions of ring-C, suggesting that bulky groups in these sites are favorable. This is consistent with the experimental fact that compounds **1** and **2** with -CH₃ on these positions have higher activities than **3** and **4** with H in these sites. The yellow contours near the two methyl groups of substituent R₂, embedding the C₁₄-, C₁₅- and C₁₆-positions and the NH-bridge at the C₆, indicate that bulky groups in these sites would decrease the activity. The fact that compounds **8** and **9** (with indole and indazole as substituent R₁, respectively) have less activities than compound **5** with no substituents at the C₁₄-, C₁₅- and C₁₆- positions of ring-C. A green and two yellow contours are mapped enclosing the piperidine ring of substituent R₃, demonstrating that medium-size groups are favored. This may explain why compounds **14-21** all possessing a relative medium-size substituent R₃ exhibit higher activity of almost 10 times than C-2 unsubstituted compound **5**.

On the CoMFA electrostatic contour map (Fig. 5B), the blue regions indicate that the electropositive groups are favorable and the red regions indicate the electronegative groups are favorable. Red contours at the terminal of substituent R₂ and near the bridge amino nitrogen atom of substituent R₂ suggest that the electronegative substituents at these regions are favorable. Positively charged groups are favored towards the lower regions of ring-D and near H atom of the vinyl linker. This may result in electrostatic interactions between the electropositive part of ring-D and the electron-rich O atoms of OH of residues Tyr340 and Ser342 in docking, and can form a C-H-O nonclassical hydrogen bond interaction between the olefinic protons in the vinyl linker with the side chain of Thr338.

The CoMFA steric and electrostatic contours for Abl-in model (Fig. 5C and D) were found to be nearly identical to the corresponding Src-in model except for several slight different. The steric field shows that a yellow and a narrow green regions near the DMPO moiety of substituent R₂ in Abl-in, while three yellow contours are found near this site in Src-in. This is not

surprised, docking study (Fig. 2B) shows the DMPO moiety of substituent R_1 locates at the entrance of the solvent region, due to the orientation of substituent R_1 , one methyl group may be blocked by the near residues Tyr319 and Gly321, but another CH_3 group has not steric hindrance, so that certain bulky group reaching this green area would improve the Abl activity. With respect to the electrostatic field, an additional favorable positive charge is found embedding the methyl of R_1 , which may make contacts with the electron-rich oxygen atom of $-\text{OH}$ of residue Asp381. The fact that compound **9** has larger potency than **8** can be interpreted by this blue contour. Another additional red contour near substituent R_3 reveals the electronegative group may enhance the activity. Therefore, compounds **24**, **22**, and **23** have an order for activity of **24**>**22**>**23**, with the corresponding substituent R_3 $-\text{CH}_2\text{CH}_2\text{CN}$, $-\text{iPr}$, and $-\text{cyclopentyl}$, respectively.

This similarity is corresponding to the high degree of conservation of the amino acid sequence within the ATP binding sites of Src-in (2BDJ) and Abl-in (3KF4). The crystal structural superposition, created using <http://cl.sdcc.edu>, revealed that only three active site residues vary between the sequence of Src-in model and that of Abl-in model (Tyr340-Phe317, Lys343-Tyr320 and Ser345-Asn322). Three residues present their backbone to the binding site cavity, while their side chains point away from the cavity. Hence, the architecture of the ATP binding site of Src-in and Abl-in can be considered as rather similar.

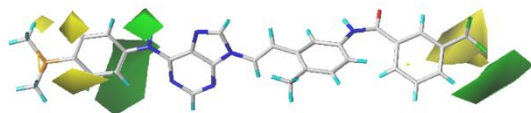
From the above, we can conclude:

(1) Bulky groups at the C_{13} -, and C_{17} -positions of ring-C and moderate-sized groups as substituents R_3 could improve the activity.

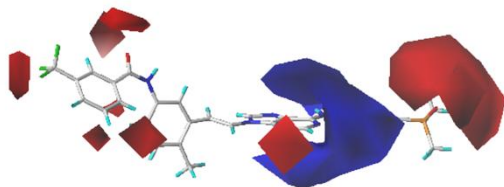
(2) Electron-withdrawing groups linking to ring-D of substituent R_2 may be crucial for the activity.

DFG-out series. Like the DFG-in Src/Abl studies, the CoMFA models for DFG-out Src/Abl kinases are visualized in Fig. 6, with compound **32** as a reference structure.

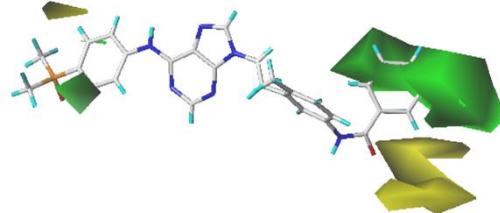
(A)



(B)



(C)



(D)

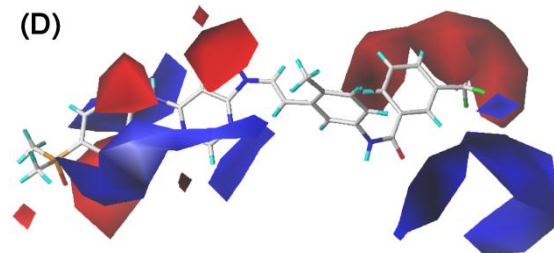


Fig. 6 CoMFA contour maps based on the highly active compound **32** in the DFG-out series: (A) Steric contour map for Src, (B) Electrostatic contour map for Src, (C) Steric contour map for Abl, (D) Electrostatic contour map for Abl.

Fig. 6A shows the contour maps of sterically favored (green) and disfavored (yellow) regions for DFG-out Src kinase. A green and a yellow regions near $-\text{CF}_3$ group of substituent R_1 suggest the medium-size substituent is favorable, because too large group may make collision with residues Tyr382 and Leu317 in docking. Moreover, two green and three yellow contours are mapped near the NH and ring-D of substituent R_2 , demonstrating that a medium-size substituent R_2 benefits the activity. This can be used to explain that compounds **44** and **43** with cyclopropyl amine as substituent R_2 exhibit higher activities of almost 10 and 45 times than **48** and **49** with H atom as R_2 , which may also lead to the loss of hydrogen bond donor to the hinge region as the removal of the C_6 -substituent.

On the electrostatic contour map (Fig. 6B), several red contours near the trifluoromethyl, the carbonyl oxygen atom, the terminal of substituent R_2 imply the preference for negative charges at these regions are favorable. Compound **34** with $-\text{CF}_3$ linking to ring-E of substituent R_1 shows higher activity than **39** existing $-\text{iPr}$ at the same location is a good illustration. The fact that compound **32** of the highest activity possesses strong electronegative oxygen atom at the terminal of substituent R_2 proves this finding. It is in good agreement with the CoMFA study of DFG-in Src. In contrast, the blue region around the ring-D shows key positive-favorable property, which is consistent with the docking study.

The CoMFA steric and electrostatic contours for DFG-out Abl (Fig. 6C and D) are similar to the ones obtained from the CoMFA model of DFG-out Src, and only slight differences are found. The electrostatic field shows an additional large blue contour (positive charge favored) in the vicinity of the $\text{C}_{2'}$ -position of ring-E (Fig. 6D). The fact that compound **36** with $-\text{F}$ on $\text{C}_{2'}$ -position is almost the inactive compound is a good example. This similarity is also because of the high structural homology between Src and Abl, the structural superposition of

3G6H (Src-out) and 3KFA (Abl-out) also demonstrates that only three active site residues vary between these two sequences (Ile338-Thr315, Tyr340-Phe317 and Lys343-Tyr320), which intimates that the docking domain for DFG-out Src and Abl may have almost the same active sites.

From the above, we can find that moderate-sized and electronegative groups at C_{3'}-position of ring-E and C_{4'}-position of ring-D may be good for the activity.

MD simulations

The MD simulations of eight above-mentioned docking complexes (compounds **20**-2BDJ, **20**-3KF4, **4**-2BDJ, **4**-3KF4, **32**-3G6H, **32**-3KFA, **40**-3G6H and **40**-3KFA) were performed to further investigate ligand-receptor interactions in the binding process. To ensure the dynamic stability and rationality of these complexes, the RMSD values of the protein backbone atoms were calculated and displayed in Fig. 7.

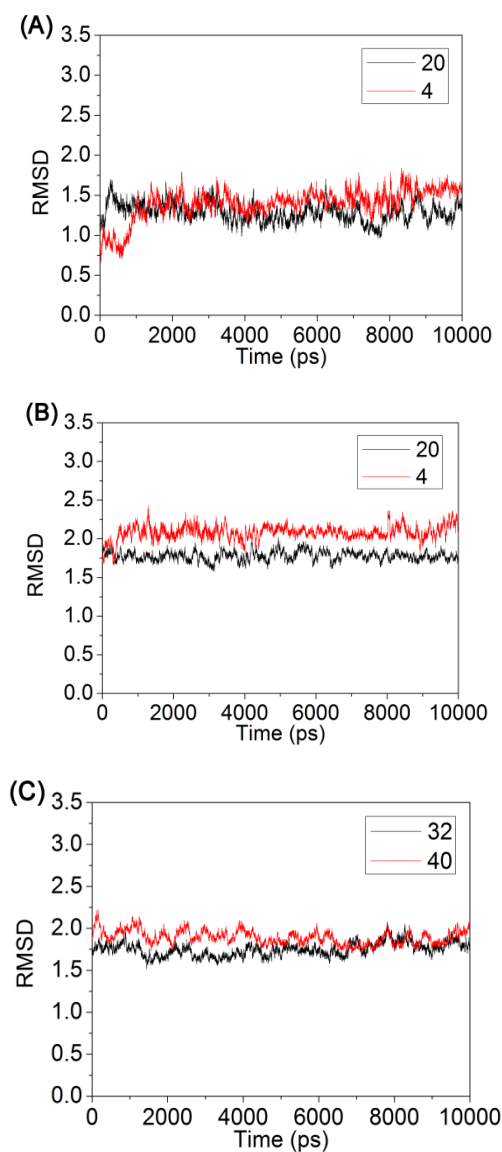


Fig.7 The rmsd of backbone atoms of eight complexes during MD simulations, in which the black lines represent compounds **20** (A and B) and **32** (C and D) systems, respectively, while red lines indicate compounds **4** (A and B) and **40** (C and D) systems, respectively. (A) for **20**-2BDJ and **4**-2BDJ complexes, (B) for **20**-3KF4 and **4**-3KF4 complexes, (C) for **32**-3G6H and **40**-3G6H complexes, (D) for **32**-3KFA and **40**-3KFA complexes.

As showed in Fig. 7, both the complexes (**20**-2BDJ and **4**-2BDJ systems) reach equilibrium after 1 ns, whereas the other six systems achieve equilibrium around 0.5 ns. The mean RMSD values of these systems were 1.2 Å, 1.8 Å, 1.4 Å, 2.2 Å, 1.7 Å, 1.6 Å, 2.0 Å and 2.1 Å, respectively, and the relative RMSD fluctuations were very small, suggesting that the systems were stable during the MD simulations. Moreover, the superimpositions of the average structure of the last 2 ns trajectory and the initial docked structures for eight systems are displayed in Fig.S1 (ESI†), where the yellow line represents the initial structure of the docked complex, and the magenta line represents the average MD simulated structure. The initial and the final average structures of the ligand are shown in yellow and magenta sticks, respectively. It can be obviously noted that the docked complexes and the average structures were well overlapped at the same binding site with only slight positional derivations, which further verified the reasonability and stability of the docking results.

To further examine the residue contribution of the receptor in the binding process, analyses of root-mean-square fluctuation (RMSF) versus the residue number for these systems were illustrated in Fig. 8. From Fig. 8, we can see that the protein structures of two complexes in each diagram display the similar RMSF distributions and similar trends of dynamic features. The fluctuations of the residues are higher in the four lowly active compounds than those in their corresponding highly active compounds. For instance, in Fig. 8A, the active site, including Leu273, Val281, Ile294 and Met341, has large conformational drift for **4**-2BDJ system than that for **20**-2BDJ, suggesting that the inhibitor **20** should have more stable interaction with the receptor than **4**. Overall, these analyses for binding stabilization consist with the experimental activities.

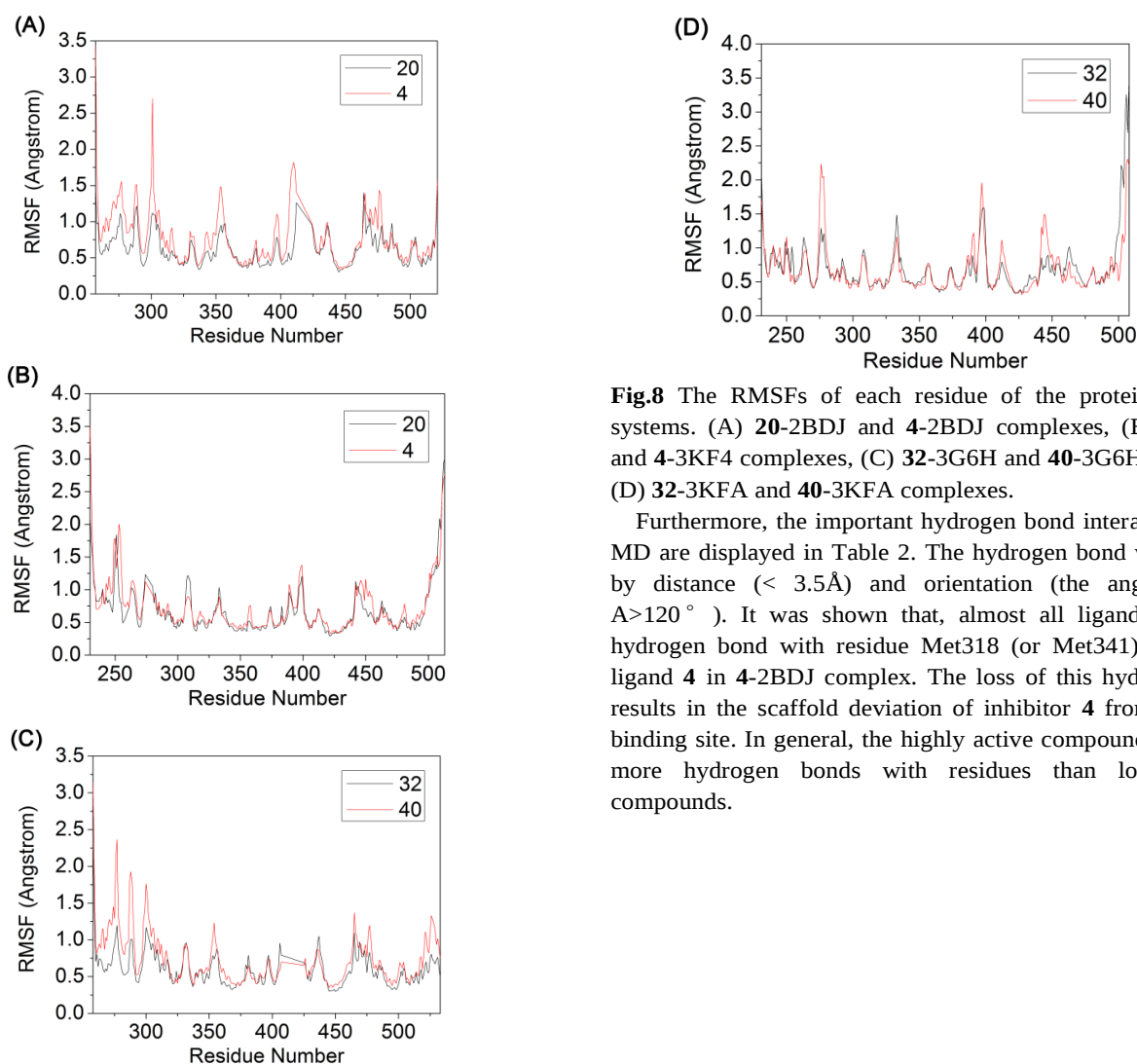


Fig.8 The RMSFs of each residue of the protein for eight systems. (A) **20**-2BDJ and **4**-2BDJ complexes, (B) **20**-3KF4 and **4**-3KF4 complexes, (C) **32**-3G6H and **40**-3G6H complexes, (D) **32**-3KFA and **40**-3KFA complexes.

Furthermore, the important hydrogen bond interactions from MD are displayed in Table 2. The hydrogen bond was defined by distance ($< 3.5\text{\AA}$) and orientation (the angle $\text{D-H} \cdots \text{A} > 120^\circ$). It was shown that, almost all ligands form the hydrogen bond with residue Met318 (or Met341) except for ligand **4** in **4**-2BDJ complex. The loss of this hydrogen bond results in the scaffold deviation of inhibitor **4** from the exact binding site. In general, the highly active compounds can form more hydrogen bonds with residues than lowly active compounds.

Table 2 Hydrogen bonds analysis from MD

system	donor	acceptor	occupancy (%)	Distance($\text{\AA}$)	Angle ($^\circ$)
20 -2BDJ	Met341@O	ligand@N-H	95.00	3.011	33.92
	ligand@N7	Met341@N-H	70.98	3.234	30.83
20 -3KF4	ligand@N18	Asn322@ND2-HD22	97.12	3.043	28.86
	Met318@O	ligand@N-H	83.47	3.009	42.16
	ligand@N7	Met318@N-H	81.08	3.244	24.50
4 -3KF4	ligand@N7	Met318@N-H	96.41	3.085	29.93
	Met318@O	ligand@N-H	69.66	3.220	29.40
32 -3G6H	Val323@O	ligand@N18-H	48.64	3.196	20.76
	ligand@O21	Val323@N-H	42.76	3.166	43.02
	Met341@O	ligand@N-H	28.18	3.184	40.15
	ligand@N7	Met341@N-H	21.76	3.292	22.69
40 -3G6H	ligand@N7	Met341@N-H	91.30	3.187	22.19
32 -3KFA	ligand@N7	Met318@N-H	91.25	3.159	27.94
	ligand@O21	Lys271@NZ-HZ	66.02	2.896	44.30
	Met318@O	ligand@N-H	63.79	3.198	40.82
40 -3KFA	Glu286@OE1	ligand@N18-H	98.19	2.942	34.42
	ligand@N7	Met318@N-H	97.29	3.113	22.26
	ligand@O21	Asp381@N-H	96.13	3.060	23.49
	Met318@O	ligand@N-H	78.32	3.021	40.20

ARTICLE

Binding free energy analysis

To identify the stabilities of the eight complexes, the binding free energy calculations using MM-PBSA method were performed and the corresponding results were listed in Table 3. It can be noted that the calculated ΔG_{bind} values for the highly active compound systems (e.g., 20-2BDJ, 20-3KF4, 32-3G6H, 32-3KFA) were higher than the ones of their corresponding lowly active compound systems (e.g., 4-2BDJ, 4-3KF4, 40-3G6H, 40-3KFA), indicating that the former exhibit a stronger potency of binding to the receptor than the latter, which was in line with the experimental activities. According to the energy individual component of the binding free energies (Table 3), we can also see that, in eight complexes, the van der Waals interactions (ΔG_{vdw}) make significant contributions to the binding, and there are great differences of van der Waals energies between the highly active compound systems and their corresponding lowly active compound systems. Meanwhile, the electrostatic energy (ΔG_{ele}) and the nonpolar solvation free energy ($\Delta G_{\text{nonpol,sol}}$) are also favorable for the binding, while the polar solvation free energy ($\Delta G_{\text{ele,sol}}$) is largely unfavorable for the binding in all complexes, which could be contributed to the weaker electrostatic interactions between ligands and receptors compared with stronger force between ligands and solvents. Usually, the van der Waals energy is closely correlative with the hydrophobic interaction, so we can further validate that the hydrophobic interactions play a key role for stabilizing the ligand in the receptor.

To further investigate the detailed protein-ligand interactions, the binding free energy in eight systems was decomposed into individual residue located within 6 Å of the ligand. In the DFG-in models (Fig. 9A and B), it can be seen that the interactions between **20** (or **4**) and 2BDJ (Src-in) are mainly determined by residues (larger than 0.5 kcal/mol) Leu273, Val281, Ala293, Ile294, Lys295, Ile336, Thr338, Met341, Lys343, Gly344, and Leu393. And residues such as

Leu248, Tyr253, Val256, Ala269, Lys271, Val299, Thr315, Phe317, Met318, and Gly321 of 3G6H (Abl-in) undergo strong interactions with these two compounds. As far as most of the previous residues are nonpolar, it is apparent that ligands can form strong van der Waals interactions with these residues and the van der Waals part makes major contributions to the binding free energy. The residues Met341 (Src-in) and Met318 (Abl-in) take strong contacts with two compounds by not only vdw interaction but also forming H-bonds with both 2BDJ and 3G6H. Meanwhile, it is observed that almost all residues energetically contribute more for the binding of compound **20** than that of compound **4**, suggesting that the interactions of compound **20** with both 2BDJ and 3G6H are stronger than those of **4**. This further demonstrates that these dual DFG-in Src/Abl inhibitors undergo similar interactions with the two proteins in the binding pockets, which had been confirmed by molecular docking and 3D-QSAR results. For the DFG-out models (Fig. 9C and D) the residues with the most favorable contributions to the binding free energy in **32**-3G6H and **40**-3G6H systems (Src-out) are Leu273, Val281, Ala293, Met314, Leu322, Val323, Ile338, Tyr340, Met341, Lys343, Gly344, Leu393, Ala403, Asp404 and Phe405, and those in **32**- and **40**-3KFA systems (Abl-out) are Leu248, Val256, Ala269, Lys271, Met290, Val299, Phe317, Met318, eu370, Ala380, Asp381 and Phe382. It is shown that most residues are hydrophobic, which can form strong van der Waals interactions with the inhibitors. Moreover, almost all those important residues for two receptors make more contributions to the binding of highly active compound **32** than those of lowly active compound **40**. Especially, the residues Ala403, Asp404, and Phe405 in 3G6H (Src-out), and Ala380, Asp381, and Phe382 in 3KFA (Abl-out), involved in the lipophilic pocket of DFG-out loop, contribute more favorably to compound **32** than to **40**. Form these results, we can conclude that hydrophobic interactions play an important role in the binding affinity with these dual Src/Abl inhibitors.

Table 3 The binding free energy of eight systems*

system	polar contributions		nonpolar contributions		ΔG_{bind}	ΔG_{exp}	$\text{IC}_{50}(\text{nM})$
	ΔG_{ele}	$\Delta G_{\text{ele,sol}}$	ΔG_{vdw}	$\Delta G_{\text{nonpol,sol}}$			
20 -2BDJ	-15.55	35.94	-60.45	-7.36	-47.42	-12.83	0.46
4 -2BDJ	-5.74	29.79	-52.68	-7.42	-36.04	-8.83	376
20 -3KF4	-11.20	37.55	-68.12	-8.13	-49.90	-12.83	0.46
4 -3KF4	-13.97	32.38	-50.05	-7.64	-39.27	-8.52	626
32 -3G6H	-14.06	38.42	-66.08	-8.39	-50.12	-11.57	3.8
40 -3G6H	-6.31	29.15	-61.86	-8.30	-47.32	-8.94	310
32 -3KFA	-16.97	45.46	-76.27	-8.57	-56.35	-11.33	5.7
40 -3KFA	-14.37	42.27	-73.97	-8.25	-54.30	-9.03	267

*All energies are in kcal/mol. ΔG_{exp} is the experimental binding free energy estimated from IC_{50} values by $\Delta G \sim -RT \ln \text{IC}_{50}$.

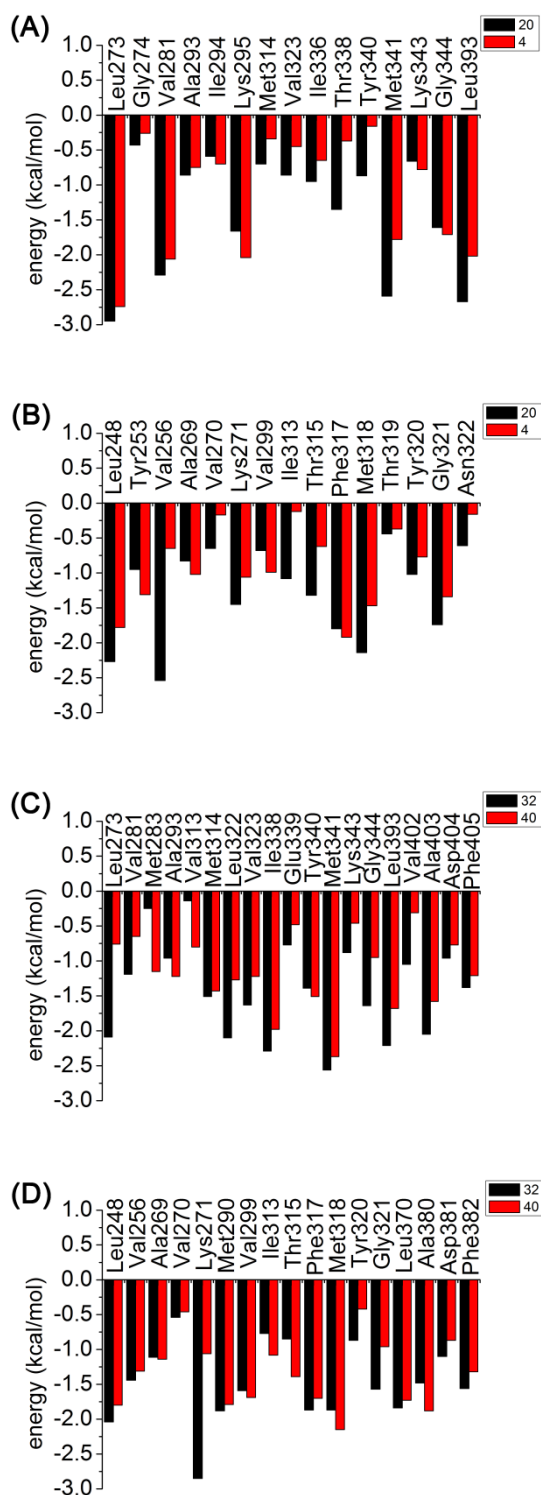


Fig.9 Free energy decomposition plots for eight systems. (A) 20-2BDJ and 4-2BDJ complexes, (B) 20-3KF4 and 4-3KF4 complexes, (C) 32-3G6H and 40-3G6H complexes, (D) 32-3KFA and 40-3KFA complexes.

Conclusions

In this paper, a molecular modeling study was applied to gain insight into the structural basis and inhibitory mechanism for novel DFG-in or DFG-out dual Src/Abl inhibitors. The constructed 3D-QSAR models not only exhibited good predictive powers in both the internal and external validations, but also identified the critical structure features influencing the inhibitory activity. The docking studies revealed that the inhibitors bound to Src in a similar mode to that observed in Abl, whereas the bound pattern of DFG-in inhibitors are quite different to that of DFG-out ones, due to the different structure of the activation loop. Based on the docking results, the MD simulations were performed and confirmed the reasonable binding modes of these complexes and the key interaction features. The calculated binding free energies were good consistent with the experimental activities. The decomposition of binding free energy to each residue revealed that the van der Waals interactions play an important role in stabilizing the binding of inhibitors, and the key residues with the most favorable contributions to the binding free energy in eight systems were also identified. The pivotal interactions with Ala403/380, Asp404/381, and Phe405/382 could help stabilize the DFG-out conformations. The obtained results could provide useful insights into the binding mechanism between targets and inhibitors and direct the further design of novel potent DFG-in and DFG-out dual Src/Abl inhibitors.

Acknowledgements

We gratefully acknowledge supports of this research by the National Natural Science Foundation of China (No.20903026), the Natural Science Foundation of Guangdong (No.S2011010002483, S2011010002964), the Science and Technology planning Project of Guangzhou (No.2013J4100071). We also heartily thank the computation environment support by the Department of Biochemistry, College of Life Sciences, Sun Yat-Sen University for SYBYL 6.9.

Notes and references

^aDepartment of Physical Chemistry, College of Pharmacy, Guangdong Pharmaceutical University, Guangzhou 510006, China.

E-mail: wuwenjuan83@126.com; Tel: +86 20 39352119

^bDepartment of Cardiothoracic Surgery, Affiliated Second Hospital of Guangzhou Medical University, Guangzhou 510260, China

^cSchool of Chemistry and Chemical Engineering, SunYat-Sen University, Guangzhou 510275, China

† Electronic supplementary information (ESI) available: including Table S1-S2 and Fig.S1.

- 1 Q. P. Ly and T. J. Yeatman, in *Targeted Interference with Signal Transduction Events*, Springer, 2007, pp. 169-188.

- 2 L. C. Kim, L. Song and E. B. Haura, *Nature Reviews Clinical Oncology*, 2009, **6**, 587-595.
- 3 Y. Hu, Y. Liu, S. Pelletier, E. Buchdunger, M. Warmuth, D. Fabbro, M. Hallek, R. A. VanEtten and S. Li, *Nat. Genet.*, 2004, **36**, 453-461.
- 4 S. Dhut, T. Chaplin and B. Young, *Leukemia*, 1990, **4**, 745-750.
- 5 A. M. Pendergast, A. J. Muller, M. H. Havlik, Y. Maru and O. N. Witte, *Cell*, 1991, **66**, 161-171.
- 6 C. L. Sawyers, *N. Engl. J. Med.*, 1999, **340**, 1330-1340.
- 7 A. Quint &-Cardama and J. Cortes, *Blood*, 2009, **113**, 1619-1630.
- 8 T. Tauchi and K. Ohyashiki, *Int. J. Hematol.*, 2006, **83**, 294-300.
- 9 T. O'Hare, A. S. Corbin and B. J. Druker, *Curr. Opin. Genet. Dev.*, 2006, **16**, 92-99.
- 10 E. Weisberg, P. W. Manley, S. W. Cowan-Jacob, A. Hochhaus and J. D. Griffin, *Nat. Rev. Cancer*, 2007, **7**, 345-356.
- 11 J. Baselga, A. Cervantes, E. Martinelli, I. Chirivella, K. Hoekman, H. I. Hurwitz, D. I. Jodrell, P. Hamberg, E. Casado and P. Elvin, *Clin. Cancer Res.*, 2010, **16**, 4876-4883.
- 12 J. Bain, L. Plater, M. Elliott, N. Shpiro, C. Hastie, H. Mclauchlan, I. Klevernic, J. Arthur, D. Alessi and P. Cohen, *Biochem. J.*, 2007, **408**, 297-315.
- 13 M. M. Schittenhelm, S. Shiraga, A. Schroeder, A. S. Corbin, D. Griffith, F. Y. Lee, C. Bokemeyer, M. W. Deininger, B. J. Druker and M. C. Heinrich, *Cancer Res.*, 2006, **66**, 473-481.
- 14 U. Rix, O. Hantschel, G. Dürnberger, L. L. R. Rix, M. Planavsky, N. V. Fernbach, I. Kaupe, K. L. Bennett, P. Valent and J. Colinge, *Blood*, 2007, **110**, 4055-4063.
- 15 H. Kantarjian, N. P. Shah, A. Hochhaus, J. Cortes, S. Shah, M. Ayala, B. Moiraghi, Z. Shen, J. Mayer and R. Pasquini, *N. Engl. J. Med.*, 2010, **362**, 2260-2270.
- 16 Y. Wang, W. C. Shakespeare, W. S. Huang, R. Sundaramoorthi, S. Lentini, S. Das, S. Liu, G. Banda, D. Wen and X. Zhu, *Bioorg. Med. Chem. Lett.*, 2008, **18**, 4907-4912.
- 17 W. S. Huang, X. Zhu, Y. Wang, M. Azam, D. Wen, R. Sundaramoorthi, R. M. Thomas, S. Liu, G. Banda and S. P. Lentini, *J. Med. Chem.*, 2009, **52**, 4743-4756.
- 18 T. Zhou, L. Commodore, W. S. Huang, Y. Wang, T. K. Sawyer, W. C. Shakespeare, T. Clackson, X. Zhu and D. C. Dalgarno, *Chem. Biol. Drug Des.*, 2010, **75**, 18-28.
- 19 Y. Liu and N. S. Gray, *Nat. Chem. Biol.*, 2006, **2**, 358-364.
- 20 F. Ntie-Kang, S. Kannan, K. Wichapong, L. C. O. Owono, W. Sippl and E. Megnassan, *Mol. Biosyst.*, 2014, **10**, 223-239.
- 21 Z. Zhang, M. Zheng, L. Du, J. Shen, X. Luo, W. Zhu and H. Jiang, *J. Comput. Aided Mol. Des.*, 2006, **20**, 281-293.
- 22 J. Caballero, M. Quiliano, J. H. Alzate-Morales, M. Zimic and E. Deharo, *J. Comput. Aided Mol. Des.*, 2011, **25**, 349-369.
- 23 K. Kumar, A. Anbarasu and R. Sudha, *Mol. Biosyst.*, 2014.
- 24 L. Sacconay, D. Smirlis, E. F. Queiroz, J. L. Wolfender, M. B. P. Soares, P.-A. Carrupt and A. Nurisso, *Mol. Biosyst.*, 2013, **9**, 2223-2230.
- 25 P. Kirubakaran, G. Kothandan, S. J. Cho and K. Muthusamy, *Mol. Biosyst.*, 2014, **10**, 281-293.
- 26 M. R. Reddy, M. D. Varney, V. Kalish, V. N. Viswanadhan and K. Appelt, *J. Med. Chem.*, 1994, **37**, 1145-1152.
- 27 M. R. Reddy and M. D. Erion, *J. Am. Chem. Soc.*, 2001, **123**, 6246-6252.
- 28 X. Li, L. Ye, X. Wang, W. Shi, H. Liu, X. Qian, Y. Zhu and H. Yu, *Chemosphere*, 2013, **92**, 795-802.
- 29 A. N. Jain, *J. Comput. Aided Mol. Des.*, 2007, **21**, 281-306.
- 30 Y. Nataraja Sekhar, M. Ravi Shashi Nayana, M. Ravikumar and S. Mahmood, *Chem. Biol. Drug. Des.*, 2007, **70**, 511-519.
- 31 X. Li, L. Ye, X. Wang, X. Wang, H. Liu, X. Qian, Y. Zhu and H. Yu, *Sci. Total. Environ.*, 2012, **441**, 230-238.
- 32 S. J. Cho and A. Tropsha, *J. Med. Chem.*, 1995, **38**, 1060-1066.
- 33 M. D. M. AbdulHameed, A. Hamza, J. Liu and C.-G. Zhan, *J. Chem. Inf. Model*, 2008, **48**, 1760-1772.
- 34 M. Böhm, J. Stürzebecher and G. Klebe, *J. Med. Chem.*, 1999, **42**, 458-477.
- 35 X. H. Liu, Y. X. Shi, Y. Ma, C. Y. Zhang, W. L. Dong, L. Pan, B. L. Wang, B. J. Li and Z. M. Li, *Eur. J. Med. Chem.*, 2009, **44**, 2782-2786.
- 36 R. D. Cramer, D. E. Patterson and J. D. Bunce, *J. Am. Chem. Soc.*, 1988, **110**, 5959-5967.
- 37 M. Hao, Y. Li, Y. Wang, Y. Yan and S. Zhang, *J. Chem. Inf. Model.*, 2011, **51**, 2560-2572.
- 38 L. Ståhle and S. Wold, *J. Chemom.*, 1987, **1**, 185-196.
- 39 Y. Yang, J. Qin, H. Liu and X. Yao, *J. Chem. Inf. Model*, 2011, **51**, 680-692.
- 40 S. A. DePriest, D. Mayer, C. B. Naylor and G. R. Marshall, *J. Am. Chem. Soc.*, 1993, **115**, 5372-5384.
- 41 H. Kubinyi, *Drug Discov. Today*, 1997, **2**, 457-467.
- 42 D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668-1688.
- 43 D. Case, T. Darden, T. Cheatham Iii, C. Simmerling, J. Wang, R. Duke, R. Luo, M. Crowley, R. C. Walker and W. Zhang, *Focus*, 2008, **32**.
- 44 M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, B. Mennucci and G. Petersson, *Inc., Wallingford, CT*, 2009.
- 45 P. Dobeš, J. Fanfrlik, J. Řezáč, M. Otyepka and P. Hobza, *J. Comput. Aided Mol. Des.*, 2011, **25**, 223-235.
- 46 Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo and T. Lee, *J. Comput. Chem.*, 2003, **24**, 1999-2012.
- 47 M. C. Lee and Y. Duan, *Proteins*, 2004, **55**, 620-634.
- 48 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157-1174.
- 49 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
- 50 Y. Yang, Y. Shen, S. Li, N. Jin, H. Liu and X. Yao, *Mol. Biosyst.*, 2012, **8**, 3049-3060.
- 51 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089-10092.

- 52 J. P. Ryckaert, G. Ciccotti and H. J. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327-341.
- 53 P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan and W. Wang, *Acc. Chem. Res.*, 2000, **33**, 889-897.
- 54 B. Kuhn, P. Gerber, T. Schulz-Gasch and M. Stahl, *J. Med. Chem.*, 2005, **48**, 4040-4048.
- 55 T. Hou, J. Wang, Y. Li and W. Wang, *J. Chem. Inf. Model*, 2010, **51**, 69-82.
- 56 N. Homeyer and H. Gohlke, *Mol. Inform.*, 2012, **31**, 114-122.
- 57 U. Bren, V. Martinek and J. Florián, *J. Phys. Chem. B*, 2006, **110**, 12782-12788.
- 58 J. Zeng, W. Li, Y. Zhao, G. Liu, Y. Tang and H. Jiang, *J. Phys. Chem. B*, 2008, **112**, 2719-2726.
- 59 X. L. Shen, M. Takimoto-Kamimura, J. Wei and Q. Z. Gao, *J. Mol. Model*, 2012, **18**, 203-212.
- 60 A. Onufriev, D. Bashford and D. A. Case, *Proteins*, 2004, **55**, 383-394.
- 61 H. Gohlke, C. Kiel and D. A. Case, *J. Mol. Biol.*, 2003, **330**, 891-913.