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.

Molecular BioSystems

PAPER

Assessing ligand selectivity of sphingosine kinases with molecular dynamics and MM-PBSA binding free energy calculations

Liang Fang[a,b], Xiaojian Wang*[a,b], Meiyang Xi[a], Tianqi Liu[b] and Dali Yin*[a,b]

The dynamic balance of sphingolipids plays a crucial role in diverse biological processes such as mitogenesis, cell migration and angiogenesis. Sphingosine kinases (SKs) including SK1 and SK2 phosphorylate sphingosine to sphingosine 1-phosphate (S1P), and control the critical balance. SK1 overexpression was reported to increase cell survival and proliferation. Although several SK1 selective inhibitors have been reported, detail analysis toward the selectivity to understand the molecular mechanism has not been investigated to our knowledge. Herein, the crystal structure of SK1 and a homology model of SK2 were used to dock to five inhibitors (**1**, **2**, **3**, **4** and **5**). Protein-ligand complexes were then submitted to molecular dynamics study and MM-PBSA binding free energy calculations. By analyzing the binding model of these inhibitors, we found that residues ILE 170, PHE 188 and THR 192 in SK1 significantly contribute a favorable binding energy to the selectivity.

Introduction

The dynamic balance of sphingolipids including ceramide, sphingosine, and sphingosine 1-phosphate (S1P) plays a crucial role in diverse biological processes such as cell proliferation, apoptosis motility, and cell differentiation.[1-5] Both ceramide and sphingosine induce apoptosis in cancer cells, while S1P has the opposite effect[1]. Sphingosine kinases (SKs) that phosphorylate sphingosine to S1P control the critical balance.[6,7] There are two isoforms of SKs existing in mammalian cells, namely SK1 and SK2, which are encoded by two distinct genes located on two separate chromosomes. The expression pattern of the two SK subtypes is different: SK1 distributes mostly in the cytoplasm and migrates to the plasma membrane upon phosphorylation,[8,9] while SK2 can localize in the nucleus to inhibit DNA synthesis and regulate HDAC1/2 activity.[10,11] The two SKs have been reported to possess different biological effects. SK1 overexpression was reported to increase cell survival and proliferation,[12,13] while the role of SK2 is still in debate.[14,15].

A variety of sphingosine kinase inhibitors have been developed,[7,16-18] including SK1-selective inhibitors and SK2-selective inhibitors. However, design of highly SK1 selective ligands is quite challenging. SK1 and SK2 are highly homologous and share a high similarity and identity in their sequence. Moreover, the ligand binding domains are evolutional conserved.[19] Although a few previous works have attempted to interpret the possible mechanism of inhibitor selectivity in SK isoforms, exact mechanism is still not clear.[20-22] The first crystal structure of SK1 was reported in 2013(PDB ID: 3VZD).[23] From then on several crystal structures of SK1 are available in the protein data bank. Based on the crystal structure of SK1, our group has successfully conducted a virtual screening and found several promising inhibitors against SK1.[24]

In this study, homology modeling followed by molecular dynamics (MD) simulation was performed to achieve a 3D structure of SK2. Complexes of five inhibitors (Figure 2) obtained from docking study were subjected to the second round MD simulations. Molecular mechanics Poisson–Boltzmann solvent-accessible surface area (MM-PBSA) method was used to estimate free binding energies.[25-30] The detailed analyses of three inhibitors (**1**, **2** and **3**) and an additional study of **4** and **5** provided insight into the protein-inhibitor binding mechanism and helped to elucidate the structural and energetic basis for achieving selectivity. We found that residues ILE170, PHE 188 and THR192 in SK1 contributed significantly to the selectivity. We expect that this study will be helpful for the rational design of potential and selective inhibitors of SK1.

Materials and methods

Homology modeling

SK2 has three isoforms. The shortest one is SK2a which is generally referred as "SK2" in literature.[15,31] SK2b is the predominant form of SK2 in several human cell lines and tissues.[15] SK2c is the one with an N-terminal extension and an

[a.] State Key Laboratory of Bioactive Substances and Functions of Natural Medicines, Institute of Materia Medica, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing 100050, P. R. China. E-mail: wangxiaojian@imm.ac.cn; yindali@imm.ac.cn; Fax: +86 10 63165248; Tel: +86 10 63165248
[b.] Department of Medicinal Chemistry, Beijing Key Laboratory of Active Substances Discovery and Drugability Evaluation, Institute of Materia Medica, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing 100050, P. R. China
Electronic Supplementary Information (ESI) available. See DOI: 10.1039/x0xx00000x

additional C-terminal sequence.[32] The size of SK2 is almost twice that of SK1 (618 amino acids for SK2a, while 384 for SK1a),[31] but they share as high as 80% similarity and 45% overall sequence identity.[33] The active regions which is included in the five highly conserved regions are shared by all knew eukaryotic SKs. These indicate that homology modeling of SK2 based on SK1 is plausible.

Additional residues of SK2, compared to SK1, result from its N-terminal region (residue 1-176) and central proline-rich region (residue 381-513).[10,34] Both the N-terminal region and the center proline-rich region are not found in SK1 or any other proteins. The sequence of SK2 was then analysed by VSL2B[35] and PONDR-FIT[36] methods, as shown in Figure 1. Both the N-terminal region and the center proline-rich region are located in intrinsically disordered protein regions, which may partly explain the reason why the crystal structure of SK2 is still unavailable. The N-terminal region possesses a nuclear localization signal outside the active region, while the central proline-rich region was reported to coincide with the sphingosine binding region of these enzymes.[37] Herein we truncated the N-terminal region and kept the central proline-rich region while building the homology model. The sequence of SK2 was obtained from Uniprot database (Uniprot ID: Q9NRA0). A BLAST (Basic Local Alignment Search Tool) search was carried out with protein databank database for identification of a template structure.[38] Only the structure of SK1 showed an acceptable similarity, as expected. Crystal structure of SK1 bound with the sphingosine (PDB ID: 3VZB) was chosen as a template for homology modeling.[23] Modeler compiled in Discovery Studio 4.1 was used to construct the homology models.[39] A model with the lowest DOPE Score was picked from the constructed 100 models. Loops were refined by Loop Refinement tool compiled in Discovery Studio 4.1 before the model was subjected to MD simulation.

Molecular Dynamics Simulation to Stabilize Homology Model

MD simulation was performed using the AMBER v14 software.[40] The structure from homology modeling was prepared with *LEap* module employing ff14SB force field

parameters.[41] An octahedral box of TIP3P water molecules extending 9 Å from protein was created to solvate the system.[42] Sodium ions were added to neutralize the system. SHAKE algorithm was used to constrain heavy atom bond lengths with hydrogens.[43] The Particle Mesh Ewald method with 9.0 Å cut-off for nonbond interactions was used for long-range electrostatic interactions.[44] 5000 steps of the steepest descent algorithm and 5000 steps of conjugate gradient algorithm were employed to minimize the system, then the temperature was increased gradually from 0 K to 325 K within 100 ps using harmonic positional restraints of 10 kcal/mol/Å. The restraint was gradually removed through three phases of 200 ps relaxation at constant isotropic pressure of 1 atm and temperature of 325K. The production simulation was carried out without restrains for 500 ns in the NPT ensemble (T=325K, thermostat relaxation time = 2.0 ps; P=1 atm). The coordinates were extracted every 5 ps, resulting in a total of 100000 snapshots. Equilibration was monitored by assessing root-mean-square deviations (RMSD) using the *cpptraj* module.[45]

Preparation of receptor-ligand complexes

1, **2** and **3** with experimental determined IC$_{50}$ values were inspected in detail in this study. The activity and selectivity of the ligands are reported in Table 1.[46] Another two SK1 selective inhibitors, PF-54318 (**4**) and VPC96091[47] (**5**) with respectively larger than 100-folds and 10-folds selectivity, were also evaluated. Structure of **2** was extracted from the crystal structure of SK1 (PDB ID: 4L02), while **1**, **3**, **4** and **5** were constructed manually. All these ligands were prepared using the Prepare Ligands tool compiled in Discovery Studio 4.1.

Both the structures of SK1 and the homology model of SK2 were prepared by the Prepare Protein tool compiled in Discovery Studio 4.1. The active site of SK1 was defined from PDB site record, while the active site of the homology model of SK2 was identified from receptor cavities and comparison with the structure of SK1. Molecular docking was performed by using LibDock tool.[48] The poses with the highest LibDockScore were used to create receptor-ligand complexes. All ten complexes were then submitted to molecular dynamic simulation for further study.

Molecular dynamics simulations and MM-PBSA calculations

MD simulations of receptor-ligand complexes were performed
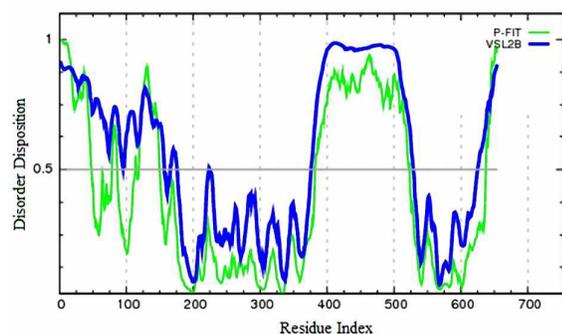
Figure 1. Disordered protein regions profile of SK2. Intrinsic disorder was predicted by the VSL2B (shown in green line) and PONDR-FIT (shown in blue line) methods. Segments with scores above 0.5 indicate disordered regions, while those below 0.5 correspond to ordered regions. Most of the N-terminal region and the central proline-rich region are located in intrinsically disordered protein regions.
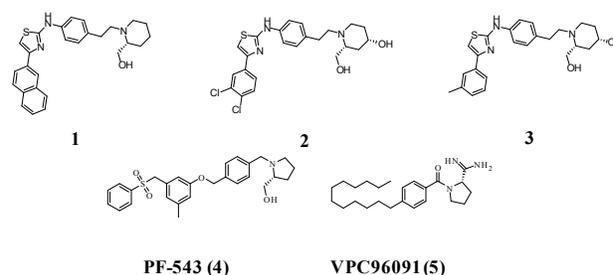
Figure 2. Molecular structures of inhibitors
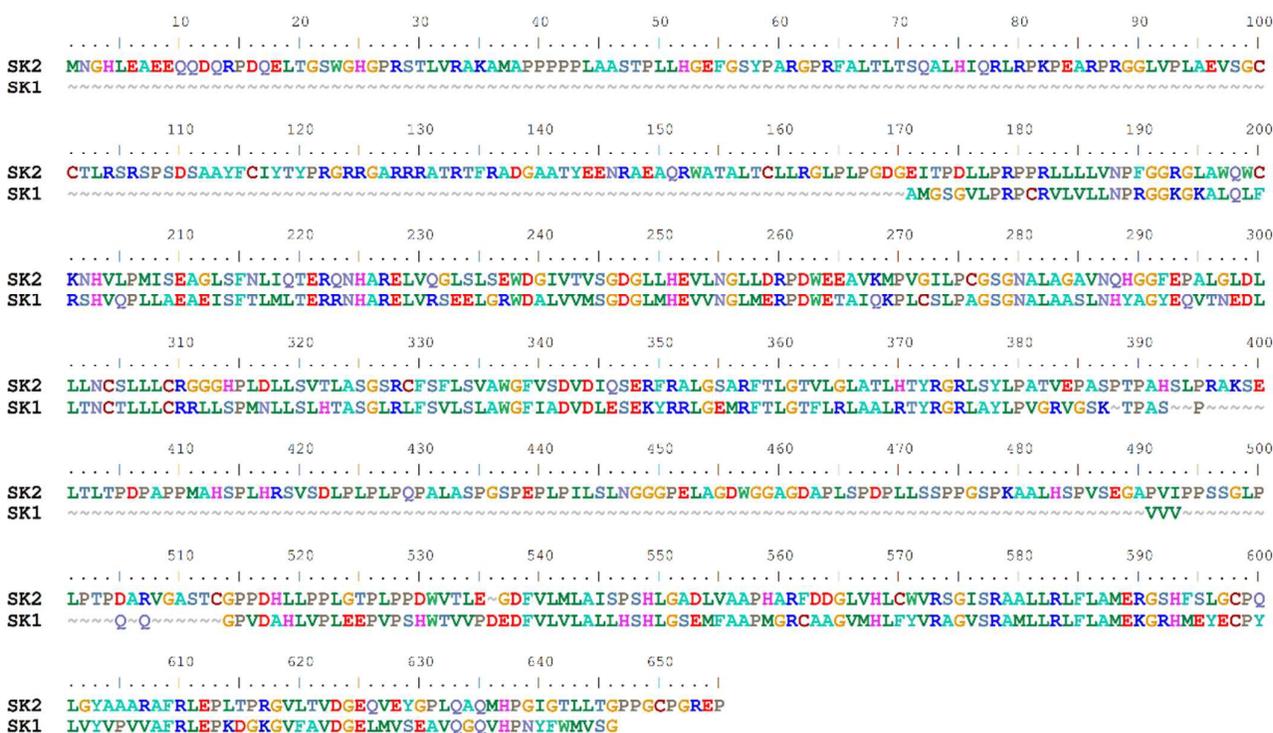
Figure 3 Sequence alignment result between SK2 and the template sequence of SK1 (PDB ID: : 3VZD)

using the GPU-PMEMD program of AMBER 14 software. The receptors were assigned with ff14SB force field parameters,
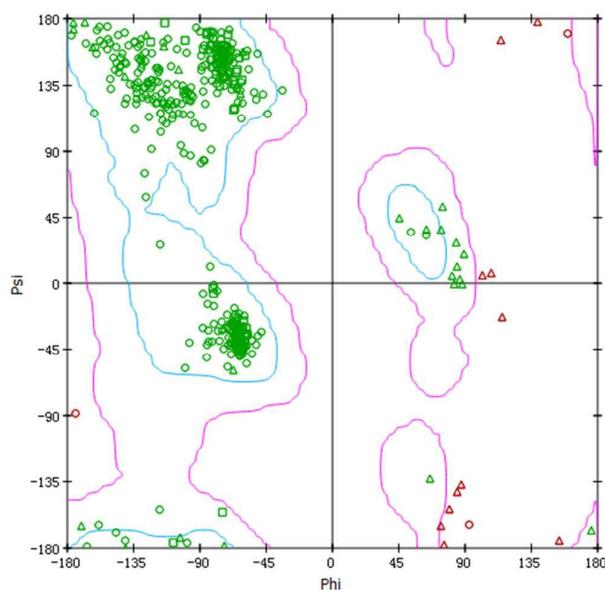
Figure 4 Ramachandran plot of MD stabilized homology model

while the ligands employed General Amber Force Field (GAFF)[49] with AM1-BCC charges.[50, 51] Receptor-ligand complexes were prepared with *LEap* program. An octahedral box of TIP3P water molecules extending 9Å from the solute was created to solvate the complex. Sodium ions or chloride ions were added to neutralize the system. SHAKE algorithm was used to constrain heavy atom bond lengths with hydrogens. The Particle Mesh Ewald method with 9.0 Å cut-off for nonbond interactions was used for long-range electrostatic interactions. 5000 steps of the steepest descent algorithm and 5000 steps of conjugate gradient algorithm were employed to minimize the system, then the temperature was increased gradually from 0 K to 300K within 100 ps using harmonic positional restraints of 10 kcal/mol/Å. The restraint was gradually removed through three phases of 200 ps relaxation at constant isotropic pressure of 1 atm and temperature of 300K. The production simulation for each complex was carried out without restrains for 40 ns in the NPT ensemble (T=300K, thermostat relaxation time = 2.0 ps; P=1 atm). Equilibration was monitored by assessing RMSD using the *cpptra*j module.

Binding free energy calculations were performed using the trajectories described earlier to investigate the differences of binding affinity between inhibitors and the binding pockets of SKs. For different receptor-ligand systems, the process of stabilized status was used for the binding free energy calculation. The calculations were performed using MM-PBSA

method implemented in AMBER 14 according to the following equation.

$$\Delta G_{bind} = \Delta G_{complex} - \Delta G_{protein} - \Delta G_{ligand} \quad\quad (1)$$

$$\Delta G_{bind} = \Delta H - T\Delta S \quad\quad (2)$$

The ΔH refers to the enthalpy contribution which is consisted of the molecular mechanics energy in the gas phase (ΔG_{MM}) and the solvation free energy (ΔG_{solv}), while $-T\Delta S$ stands for the conformational entropy effect due to binding.

$$\Delta G_{bind} = \Delta G_{MM} + \Delta G_{solv} - T\Delta S \quad\quad (3)$$

Although Normal-mode analysis (NMA) can be used to estimate the entropy contribution, this method is considered to be problematic and time-consuming. Moreover the NMA approach also does not involve the solvent entropy. Previous studies have illustrated that normal-mode analysis would not greatly improve the correlation between the calculated binding free energy and the experimental binding energy.[52] In that case, we omitted the solute entropy term in this study.

ΔG_{MM} can be further divided into a van der Waals contribution (ΔE_{vdw}) and electrostatic interaction energies (ΔE_{ele}). And the solvation free energy ΔG_{solv} can be computed as the summation of a polar component ($\Delta G_{pol,solv}$) and a nonpolar component ($\Delta G_{np,solv}$).

$$\Delta G_{MM} = \Delta E_{vdw} + \Delta E_{ele} \quad\quad (4)$$

$$\Delta G_{solv} = \Delta G_{pol,solv} + \Delta G_{np,solv} \quad\quad (5)$$

$\Delta G_{pol,solv}$ is calculated by solving the Poisson–Boltzmann equation using the PBSA module of Amber. $\Delta G_{np,solv}$ is calculated from the solvent-accessible surface area (SASA) using the equation:

$$\Delta G_{np,solv} = \gamma SASA + \beta \quad\quad (6)$$

where γ is set to 0.0072 kcal∕mol∕Å^2 and β to 0 kcal∕mol.

Free energy decomposition analysis can be performed by MM-PBSA decomposition or MM-GBSA decomposition.[53, 54] In this paper, MM-PBSA decomposition was performed to get a detailed insight into the interactions between the ligand and each residue in the binding site. The binding interaction of each ligand–residue pair includes three terms: the van der Waals contribution (ΔE_{vdw}), the electrostatic contribution (ΔE_{ele}), and the solvation contribution (ΔG_{solv}).

Results and discussion

Model building and validation

The sequence of SK2 was retrieved from the Uniprot database(Uniprot ID: Q9NRA0).[55] The crystal structure of SK1 (PDB ID: 3VZB) was selected as the template for building the homology model of SK2 as the result of BLAST search. Figure 3 shows the final alignment. The N-terminal region not found in SK1 or any other proteins was truncated while building the homology model. Modeler compiled in Discovery Studio 4.1 was used to construct the homology models. A model with the highest score was picked from the constructed 100 models.

Loops were refined by Loop Refinement compiled in Discovery Studio 4.1. The validation of the model was performed by Verify 3D[56] which shows 82.11% of the residues have an averaged 3D-1D score ≥ 0.2(ESI,† Figure S1). This suggests that the model is good enough to be a starting point for the next phases of studies.

To adapt its physiological environment, the homology model was submitted to molecular dynamic simulation. During the MD simulation, the RMSD change of backbone atoms of the model rose gradually during the initial 300ns, and reached a stable state after 300ns, which indicated the system attained equilibrium (Figure 7a). Compared with the initial structure, the major change came from the central proline-rich region (Figure 5). The structure with the lowest RMSD to the average structure of the last 10ns was used to create the receptor-ligand complexes. The Ramachandran plot(Figure 4) was obtained to inspect the quality of the stabilized structure.[57] The chosen structure shows a good model quality with 98.1% residues in the allowed region, 1.4% in the marginal region and only 0.6% residues in the disallowed region.
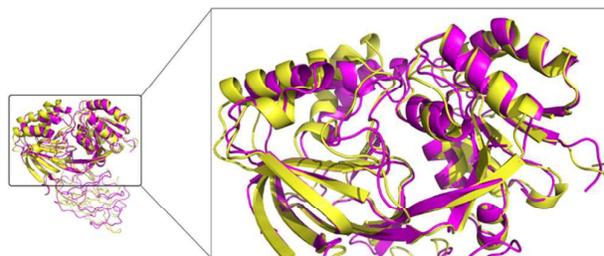
Figure 5 MD simulation stabilized structure of SK2 (yellow) overlap the initial homology model (magenta)

Table 1. Binding free energies calculations for SK1 and SK2 complexes[a].

Term	SK1-1	SK1-2	SK1-3	SK2-1	SK2-2	SK2-3
ΔE_{vdw}	−68. 66	-57.60	-55.25	-57.82	-52.49	-48.71
	(1. 94)	(1.41)	(1.56)	(2.03)	(1.89)	(1.61)
ΔE_{ele}	−7. 35	−19. 55	-24.73	-8.56	-17.13	-23.10
	(2. 04)	(1. 11)	(3.09)	(1.42)	(1.18)	(2.28)
ΔG_{MM}	-76.00	-77.14	-79.98	-66.38	-69.61	-71.80
	(2. 63)	(3.65)	(3.18)	(2.92)	(3.24)	(5.30)
$\Delta G_{pol,solv}$	47. 37	47.41	49.36	41.46	41.84	45.40
	(3. 89)	(4.77)	(2.79)	(3.66)	(5.36)	(3.55)
$\Delta G_{np,solv}$	−13. 51	-13.39	−13. 50	-10.36	-9.52	-11.17
	(0. 19)	(2.19)	(1. 51)	(0.57)	(3.25)	(0.37)
ΔG_{solv}	33.85	34.01	35.86	31.10	32.31	34.22
	(3. 88)	(3.61)	(4.41)	(3.34)	(5.40)	(4.26)
ΔG_{bind}	-42.15	-43.13	-44.12	-35.28	-37.30	-39.58
	(3. 99)	(4.52)	(4.64)	(3.39)	(4.42)	(3.39)
ΔG_{exp}[b]	-40.48	-43.92	-45.64	-33.07	-33.07	-35.11

a All values are given in kcal/mol with standard deviations in parentheses

b The experimental binding free energies calculated according to the IC_{50} by $\Delta Gexp$= - RTln IC_{50}

(a)

(b)

(c)

Figure 6   Comparisons between $\Delta G_{\text{bind-residue}}$ of SK1 and SK2, (a) **1**, (b) **2** and (c) **3**

MM-PBSA calculations and binding energy decomposition analysis

The binding poses of the five ligands to SKs were searched and ranked by docking analysis based on their docking scores. The poses with the highest LibDockScore, which suggested the most possible binding model, were selected to generate the

receptor-ligand complexes. The binding poses of these five inhibitors to
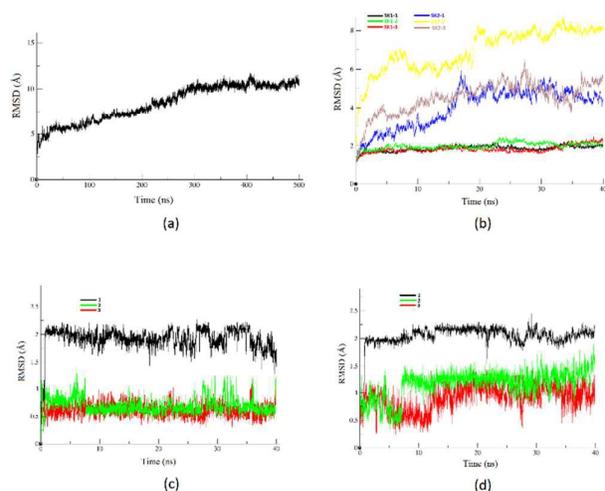
Figure 7(a) RMSD of backbone atoms of the homology model; (b) RMSDs of backbone atoms of protein in SKs complexes; (c) RMSDs of ligands in SK1 complexes; (d) RMSDs of ligands in SK2 complexes

SK1 and SK2 were similar. The hydrophilic part of the ligands extends into the binding pocket, while the more hydrophobic part exposes to the opening of the pocket.

MD simulation was applied to refine the receptor-ligand complexes by allowing both the ligands and receptors to adjust their conformations obtained from molecular docking. RMSD changes relative to their starting structures were plotted to estimate the dynamic stabilities of the 6 detailed inspected complexes. As shown in Figure 7b, the RMSD values of each system tend to converge after 20ns, indicating the systems reached a stable and equilibrated condition. As shown in Figure 7c and 7d, the RMSD values of each ligand tend to converge quickly, indicating that conformations these ligands adapted were stable during the simulations.

To compare the binding affinities of the three ligands with SK1 and SK2, MM-PBSA method was used to calculate the binding free energy. Table 1 shows experimental binding affinity and binding free energy estimated from over 100 snapshots extracted evenly from the last 10 ns MD trajectories. Although the calculated binding free energies are slightly different from the experimental value, the ranking of calculated binding free energies is in consonant with the experimental values. This suggests that MD simulation coupled with MM-PBSA simulations is able to reproduce the trend of affinities of the inhibitor among the set of SKs.

Selectivity of 1

Table 2 shows the decomposition values on a per-residue basis of the surrounding residues. Comparisons of ΔG_{bind} between SK1 and SK2 for three primary inhibitors are shown in Figure 6 and Figure 8. The RMSD deviation of **1** is larger than that of **2** and **3**, which can be explained by the larger substituting naphthalene group. The selectivity of 1 is mainly from five

Table 2. Energy contributions of SK1 and SK2 residues to the binding of ligands[a]

SKs	Residue	1				2				3			
		ΔE_{vdw}	ΔE_{ele}	ΔG_{solv}	ΔG_{bind}	ΔE_{vdw}	ΔE_{ele}	ΔG_{solv}	ΔG_{bind}	ΔE_{vdw}	ΔE_{ele}	ΔG_{solv}	ΔG_{bind}
SK1	ALA 111	-0.30	0.08	-0.40	-0.62	-0.16	-0.06	-0.07	-0.29	-0.49	-0.04	0.51	-0.03
SK2	ALA 281	-0.43	-0.02	0.51	0.07	-0.19	-0.24	0.94	0.50	-0.47	-0.07	0.09	-0.45
SK1	LEU 163	-0.58	-0.03	0.16	-0.45	-0.38	-0.10	-0.04	-0.53	-0.70	0.13	-0.21	-0.77
SK2	LEU 333	-0.68	-0.06	-0.04	-0.78	-0.75	-0.12	0.09	-0.78	-0.20	0.03	-0.05	-0.22
SK1	PHE 169	-1.35	-0.29	0.90	-0.74	-1.04	-0.27	0.93	-0.38	-1.18	-0.35	1.27	-0.26
SK2	PHE 338	-0.24	0.09	0.14	-0.01	-0.83	-0.28	0.58	-0.53	-0.91	-0.26	0.58	-0.60
SK1	ILE 170	-3.29	-1.07	2.17	-2.18	-2.86	-1.07	1.84	-2.08	-3.15	-0.72	1.79	-2.07
SK2	VAL 339	-0.83	-0.01	0.19	-0.65	-3.53	-0.89	3.57	-0.86	-2.16	-0.41	1.07	-1.50
SK1	VAL 173	-0.82	0.17	-0.22	-0.86	-0.74	0.32	-0.22	-0.64	-0.79	0.20	-0.18	-0.77
SK2	VAL 343	-0.03	-0.03	0.03	-0.03	-0.66	0.09	-0.08	-0.66	-0.54	0.11	-0.13	-0.56
SK1	PHE 188	-2.49	-0.50	1.36	-1.63	-2.38	-0.75	0.69	-2.43	-2.57	-0.50	1.43	-1.64
SK2	PHE 358	-2.33	-0.18	0.41	-2.10	-2.13	-0.55	1.20	-1.47	-2.71	-1.14	1.39	-2.46
SK1	THR 192	-0.97	-2.12	1.80	-1.29	-0.93	-2.20	1.06	-2.07	-0.95	-1.97	1.16	-1.75
SK2	THR 362	-0.72	-0.24	0.45	-0.51	-0.92	-2.25	2.13	-1.03	-1.02	-1.68	2.02	-0.68
SK1	LEU 255	-1.34	-0.26	0.22	-1.38	-1.53	-0.09	0.26	-1.36	-1.47	-0.22	0.25	-1.44
SK2	LEU 540	-1.56	-0.16	0.15	-1.57	-1.75	-0.31	0.21	-1.86	-1.64	-0.18	0.16	-1.67
SK1	LEU 257	-0.99	-0.14	0.10	-1.02	-1.52	-0.41	0.18	-1.76	-0.76	-0.04	0.16	-0.63
SK2	LEU 542	-1.00	-0.01	0.17	-0.84	-0.75	-0.03	0.15	-0.64	-0.56	0.03	0.00	-0.53
SK1	LEU 264	-2.45	-0.24	2.17	-0.51	-1.99	-0.03	1.12	-0.90	-2.50	-0.49	2.27	-0.72
SK2	LEU 549	-1.99	-0.23	1.35	-0.87	-2.22	-0.48	1.64	-1.06	-1.66	-0.12	1.66	-0.12
SK1	MET 268	-0.99	0.00	0.29	-0.70	-0.71	0.08	0.23	-0.40	-0.63	0.01	0.19	-0.43
SK2	LEU 553	-0.15	-0.06	0.20	-0.01	-0.20	-0.10	0.18	-0.12	-0.32	-0.10	0.13	-0.29
SK1	LEU 295	-1.14	-0.48	0.98	-0.65	-1.26	-0.90	1.35	-0.81	-0.97	-0.40	0.74	-0.62
SK2	LEU 580	-0.38	-0.15	0.13	-0.40	-0.73	-0.38	0.22	-0.89	-0.74	-0.22	0.04	-0.91
SK1	PHE 299	-2.30	-0.25	0.69	-1.86	-2.56	-0.10	0.80	-1.86	-2.51	-0.26	0.84	-1.92
SK2	PHE 584	-1.51	-0.02	0.35	-1.18	-2.43	-0.14	0.46	-2.11	-2.15	-0.32	0.50	-1.97
SK1	MET 302	-1.70	-0.27	0.91	-1.06	-1.45	0.10	0.75	-0.60	-1.46	-0.29	0.90	-0.85
SK2	MET 587	-1.70	0.01	0.58	-1.11	-1.54	0.15	0.54	-0.85	-1.72	-0.02	0.68	-1.06
SK1	GLY 338	-0.76	-0.43	0.66	-0.52	-0.62	-0.16	0.32	-0.46	-0.72	-0.46	0.66	-0.52
SK2	GLY 623	-0.64	-0.50	0.86	-0.28	-0.71	-0.42	0.67	-0.47	-0.80	-0.31	1.05	-0.06

[a] All values are given in kcal/mol. ΔE_{vdw}: van der Waals energy contributions, ΔE_{ele}: electrostatic energy contributions, ΔG_{solv}: solvation energy contributions, ΔG_{bind}: total free energy contributions

residues ALA111 (ALA281), ILE170 (VAL339), THR192 (THR362), MET268 (LEU553) and PHE299 (PHE584), with the absolute differences of ΔG_{bind} between SK1 and SK2 no less than 0.5 kcal/mol.

MM-PBSA decomposition studies were carried out to inspect key residues. Table 2 summarizes 15 residues contributed significant binding energies. As shown in Table 2, the replacement of SK2 VAL 339 by SK1 ILE 170 causes the free energy contribution decrease from -0.65 kcal/mol to -2.18 kcal/mol, and the replacement of SK2 leu553 by SK1 MET268 causes the free energy contribution decrease from -0.01 kcal/mol to -0.70 kcal/mol. The favorable free energies for both pair residues are mainly from the van der Waals energy of the side chain. As shown in Figure 8a, the methyl of the SK1 MET268 mainly interacts with the middle benzene ring of the inhibitor and forms alkyl-π interaction, while SK2 leu553 is too far away from the inhibitor to form efficient contents. As illustrated in Table 2, Figure 8a and 8d, residue ILE 170 forms a strong van de Waals contact and electronic interaction with the middle benzene ring of the inhibitor. Thus it is possible that a more potent interaction between the aromatic ring of the inhibitor and SK1 ILE 170 could lead to an improvement of SK1 selectivity.

The selectivity of PHE299 (PHE584) toward SK1 is majorly driven by van de Waals contact, as shown in Table 2. The distances between PHE299 (PHE584) and 1 are 3.6Å and 3.9Å, and PHE299 adopts a more favourable formation to interact with the thiazole ring of the inhibitor. Thus PHE299 has a lower
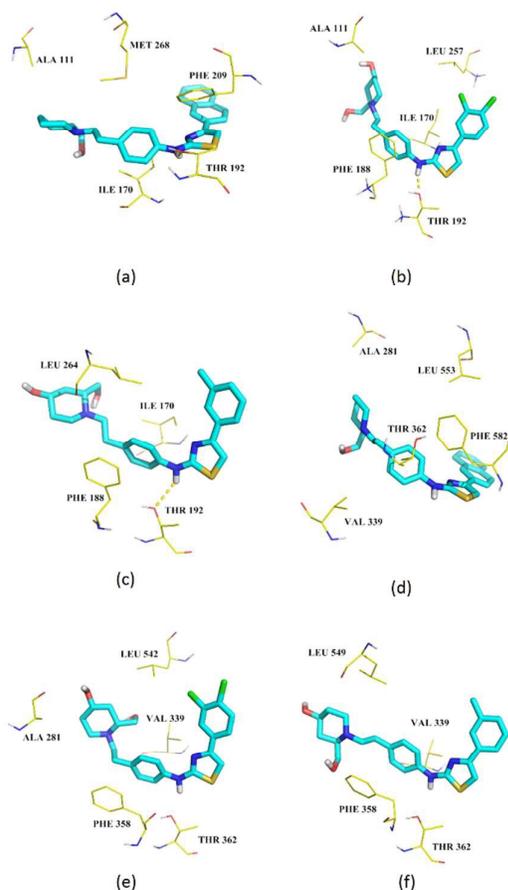
Figure 8 Comparisons the position of residues between SK1 and SK2

(a)Binding mode of **1** and key residues in SK1; (b)Binding mode of **2** and key residues in SK1; (c)Binding mode of **3** and key residues in SK1；(d)Binding mode of **1** and key residues in SK2; (e)Binding mode of **2** and key residues in SK2; (f)Binding mode of **3** and key residues in SK2；

Table 3a

Key residues responsible for selectivity of **4**[a]

SKs	Residues	ΔE_{vdw}	ΔE_{ele}	ΔG_{solv}	ΔG_{bind}
SK1	ILE 170	-3.07	-3.45	3.43	-3.09
SK2	VAL 339	-1.23	-0.42	1.42	-0.23
SK1	PHE 188	-2.03	-0.26	0.97	-1.32
SK2	PHE 358	-0.64	-0.01	0.23	-0.42
SK1	MET 302	-1.76	-0.59	0.85	-1.49
SK2	MET 587	-1.22	-0.04	0.61	-0.65
SK1	PHE 299	-2.28	-0.31	0.73	-1.86
SK2	PHE 584	-2.24	-0.35	1.21	-1.39

Table 3b

Key residues responsible for selectivity of **5**[a]

SKs	Residues	ΔE_{vdw}	ΔE_{ele}	ΔG_{solv}	ΔG_{bind}
SK1	ILE 170	-2.67	-0.07	0.50	-2.55
SK2	VAL 339	-1.86	-0.10	0.64	-1.51
SK1	PHE 188	-1.57	-0.05	0.65	-1.20
SK2	PHE 358	-0.15	-0.02	0.20	0.02
SK1	LEU 264	-2.05	0.07	0.47	-1.79
SK2	LEU 549	-1.52	-0.25	0.77	-1.21
SK1	LEU 298	-1.10	-0.61	0.26	-1.53
SK2	LEU 583	-1.09	0.17	0.00	-1.01
SK1	PHE 299	-1.96	-0.34	0.88	-1.62
SK2	PHE 584	-0.93	0.02	0.26	-0.71

[a] All values are given in kcal/mol. ΔE_{vdw}: van der Waals energy contributions, ΔE_{ele}: electrostatic energy contributions, ΔG_{solv}: solvation energy contributions, ΔG_{bind}: total free energy contributions

van der Waals energy and electrostatic energy contribution toward SK1. The distances between **1** and ALA111 (ALA281) are 4.2 Å and 3.7 Å respectively, however, ALA111 has a low solvation energy as shown in Table 2. This explains the favourable interaction between ALA111 and **1**.

A strong hydrogen bond is formed between the O atom of the side chain of SK1 THR192 and **1**, while hydrogen bond is relatively weak between SK2 THR362 and inhibitor. The distances between O atom of the side chain of THR192 (THR362) and the N atom of **1** are 3.3 Å and 6.1 Å respectively, and the electrostatic energy contributions are -2.12 and -0.24 kcal/mol, illustrating that the contribution of electrostatic energy is the main driven force. This indicates that the formation of a hydrogen bond to THR192 could lead to an improvement of SK1 selectivity.

Selectivity of 2

As shown in Figure 6b, selectivity of **2** is mainly from five residues ALA111 (ALA281), ILE170 (VAL339), THR192 (THR362),

As shown in Figure 6b, selectivity of **2** is mainly from five residues ALA111 (ALA281), ILE170 (VAL339), THR192 (THR362), PHE 188 (PHE 358) and LEU257 (LEU542) with the absolute differences of ΔG_{bind} between SK1 and SK2 no less than 0.5 kcal/mol. The mechanism of favourable selectivity contributes by ALA111 (ALA281), ILE170 (VAL339), and THR192 (THR362) is similar to that of **1**.

The distances between chlorine in the *meta* position of the end benzene ring of **2** and LEU257 (LEU542) are 3.6 Å and 4.0 Å respectively. This explains the more favourable van der Waal energy contribution of SK1 LEU257 is much lower, which indicates that halogen substitution in the *meta* position of the end benzene ring may have a more positive effect on improving the selective ability of SK1 inhibitors.

Comparing the binding position of SK1 PHE 188 and SK2 PHE358 in the complexes with **2**, the dipole moment of the aromatic rings of SK1 PHE188 aligns with that of middle benzene ring of the inhibitor in a less parallel fashion which results in a more efficient stacking.[58] This implies that an effective stacking with PHE 188 should be helpful to the selectivity of SK1 inhibitors.

Molecular BioSystems Accepted Manuscript

While PHE299 (PHE584) has a favorable contribution to selectivity of **1** binding to SK1, it is unfavorable for selectivity of **2** binding to SK1. As shown in Figure 8b and 8e, although PHE299 adopts a similar conformation as PHE584 in the complexes of 2, PHE584 has a lower solvation energy.

Selectivity of 3

As shown in Figure 6c, selectivity of **1** is mainly from four residues, ILE170 (VAL339), THR192 (THR362), PHE 188 (PHE 358) and LEU264 (LEU549) with the absolute differences of ΔG_{bind} between SK1 and SK2 no less than 0.5 kcal/mol.

The mechanism of favorable selectivity contributes by PHE 188 (PHE 358), ILE170 (VAL339) and THR192 (THR362) is similar to that of **1**.

Unlike ALA111 (ALA281) to **1** and **2**, solvation energy of ALA111 to **3** is higher than ALA281 to **3**. And it causes ALA111 (ALA281) to have an unfavorable contribution to SK1 selectivity.

The averaged distances between the inhibitor and side chain of LEU264 (LEU549) are 4.1 Å and 4.9 Å. This is in accordance with the van der Waals interaction.

Additional study of 4 and 5

Two more SK1 selective inhibitors with distinct structural types, 4 and 5, were chosen to conduct MD simulation and each system reached a stable and equilibrated condition (ESI, † Figure S2). MM-PBSA decomposition studies were performed based on each stabilized trajectory. Although the IC_{50} of these two inhibitors have not been reported, the Ki value of them indicates their strong inhibitory activity and selectivity. Table 3a shows that residues ILE170, PHE 188, MET 302 and PHE 299 in SK1 contribute significantly to selectivity of **4**, while Table 3b shows that ILE170, PHE 188, LEU 264, LEU 298 and PHE 299 contribute the most of selectivity of **5**. In all, both inhibitors emphasize the importance of contribution from residue ILE170 and PHE 188, which is consistent with key residues confirmed by **1**, **2** and **3**.

Conclusions

The 3D homology structure of SK2 constructed from the structure of SK1 were stabilized by MD simulation. Three SK1 selective inhibitors (**1**, **2**, and **3**) with experimental determined IC_{50} and two additional inhibitors (**4**, **5**) were docked into the binding pocket of SKs. The protein-ligand complexes were then submitted to MD simulations, and analysed by MM-PBSA method. The calculated binding free energies of inhibitor **1**, **2** and **3** are consistent with the experimental binding free energies. The decomposition of free energy contributions into individual residue shows that there are mainly five residues primarily contributing to selectivity for **1**, five residues for **2**, and four residues for **3**. Converging with the results of **4** and **5**, ILE170, THR192 and PHE 188 in SK1 were found to hold the most important position in selectivity of SK1 inhibitor. The chlorine in the *meta* position of the end benzene ring may also lead to an increasing selectivity resulting from a more favorable van der Waals energy. The information obtained from this study provides better structural understanding of

inhibitors binding to SKs, and can be conveniently used for future structure-based design of selective SK1 inhibitors.

Notes and references

1.   N. J. Pyne and S. Pyne, *Nat. Rev. Cancer*, 2010, **10**, 489-503.
2.   N. J. Pyne, F. Tonelli, K. G. Lim, J. S. Long, J. Edwards and S. Pyne, *Biochem. Soc. Trans.*, 2012, **40**, 94-100.
3.   G. T. Kunkel, M. Maceyka, S. Milstien and S. Spiegel, *Nat. Rev. Drug Discov.*, 2013, **12**, 688-702.
4.   Y. Zhang, V. Berka, A. Song, K. Sun, W. Wang, W. Zhang, C. Ning, C. Li, Q. Zhang, M. Bogdanov, D. C. Alexander, M. V. Milburn, M. H. Ahmed, H. Lin, M. Idowu, J. Zhang, G. J. Kato, O. Y. Abdulmalik, W. Zhang, W. Dowhan, R. E. Kellems, P. Zhang, J. Jin, M. Safo, A. L. Tsai, H. S. Juneja and Y. Xia, *J. Clin. Invest.*, 2014, **124**, 2750-2761.
5.   N. Takuwa, W. Du, E. Kaneko, Y. Okamoto, K. Yoshioka and Y. Takuwa, *Am. J. Cancer Res.*, 2011, **1**, 460.
6.   J. W. Antoon and B. S. Beckman, *Cancer Biol. Ther.*, 2011, **11**, 647-650.
7.   D. Plano, S. Amin and A. K. Sharma, *J. Med. Chem.*, 2014, **57**, 5509-5524.
8.   S. M. Pitson, P. A. Moretti, J. R. Zebol, H. E. Lynn, P. Xia, M. A. Vadas and B. W. Wattenberg, *The EMBO journal*, 2003, **22**, 5491-5500.
9.   K. E. Jarman, P. A. Moretti, J. R. Zebol and S. M. Pitson, *J. Biol. Chem.*, 2010, **285**, 483-492.
10.   N. Igarashi, T. Okada, S. Hayashi, T. Fujita, S. Jahangeer and S.-i. Nakamura, *J. Biol. Chem.*, 2003, **278**, 46832-46839.
11.   N. C. Hait, J. Allegood, M. Maceyka, G. M. Strub, K. B. Harikumar, S. K. Singh, C. Luo, R. Marmorstein, T. Kordula, S. Milstien and S. Spiegel, *Science*, 2009, **325**, 1254-1257.
12.   A. Olivera, T. Kohama, L. Edsall, V. Nava, O. Cuvillier, S. Poulton and S. Spiegel, *J. Cell Biol.*, 1999, **147**, 545-558.
13.   D. Kapitonov, J. C. Allegood, C. Mitchell, N. C. Hait, J. A. Almenara, J. K. Adams, R. E. Zipkin, P. Dent, T. Kordula, S. Milstien and S. Spiegel, *Cancer Res.*, 2009, **69**, 6915-6923.
14.   H. Liu, R. E. Toman, S. K. Goparaju, M. Maceyka, V. E. Nava, H. Sankala, S. G. Payne, M. Bektas, I. Ishii and J. Chun, *J. Biol. Chem.*, 2003, **278**, 40330-40336.
15.   T. Okada, G. Ding, H. Sonoda, T. Kajimoto, Y. Haga, A. Khosrowbeygi, S. Gao, N. Miwa, S. Jahangeer and S. Nakamura, *J. Biol. Chem.*, 2005, **280**, 36318-36325.
16.   J. A. Hengst, X. Wang, U. H. Sk, A. K. Sharma, S. Amin and J. K. Yun, *Bioorg. Med. Chem. Lett.*, 2010, **20**, 7498-7502.
17.   T. P. Mathews, A. J. Kennedy, Y. Kharel, P. C. Kennedy, O. Nicoara, M. Sunkara, A. J. Morris, B. R. Wamhoff, K. R. Lynch and T. L. Macdonald, *J. Med. Chem.*, 2010, **53**, 2766-2778.
18.   E. S. Mark, D. M. Matthew, K. Tom, Y. Matthew, J. Gina, W. R. John, H. Troii, C. Jill, K. Michelle and N. C. Ciaran, *Biochem. J*, 2012, **444**, 79-88.
19.   T. M. Leclercq and S. M. Pitson, *IUBMB life*, 2006, **58**, 467-472.

20.   M. R. Raje, K. Knott, Y. Kharel, P. Bissel, K. R. Lynch and W. L. Santos, *Bioorg. Med. Chem.*, 2012, **20**, 183-194.

21.   K. Knott, Y. Kharel, M. R. Raje, K. R. Lynch and W. L. Santos, *Bioorg. Med. Chem. Lett.*, 2012, **22**, 6817-6820.

22.   N. N. Patwardhan, E. A. Morris, Y. Kharel, M. R. Raje, M. Gao, J. L. Tomsig, K. R. Lynch and W. L. Santos, *J. Med. Chem.*, 2015, **58**, 1879-1899.

23.   Z. Wang, X. Min, S. H. Xiao, S. Johnstone, W. Romanow, D. Meininger, H. Xu, J. Liu, J. Dai, S. An, S. Thibault and N. Walker, *Structure*, 2013, **21**, 798-809.

24.   X. Wang, C. Sun, L. Fang and D. Yin, *Med. Chem. Commun.*, 2015, **6**, 413-417.

25.   B. R. Miller III, T. D. McGee Jr, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, *J Chem Theory Comput*, 2012, **8**, 3314-3321.

26.   T. Hou, J. Wang, Y. Li and W. Wang, *J Chem Inf Model*, 2011, **51**, 69-82.

27.   T. Hou, J. Wang, Y. Li and W. Wang, *J. Comput. Chem.*, 2011, **32**, 866-877.

28.   H. Sun, Y. Li, S. Tian, L. Xu and T. Hou, *Phys. Chem. Chem. Phys.*, 2014, **16**, 16719-16729.

29.   H. Sun, Y. Li, M. Shen, S. Tian, L. Xu, P. Pan, Y. Guan and T. Hou, *Phys. Chem. Chem. Phys.*, 2014, **16**, 22035-22045.

30.   L. Xu, H. Sun, Y. Li, J. Wang and T. Hou, *J. Phys. Chem. B*, 2013, **117**, 8408-8421.

31.   H. Liu, M. Sugiura, V. E. Nava, L. C. Edsall, K. Kono, S. Poulton, S. Milstien, T. Kohama and S. Spiegel, *J. Biol. Chem.*, 2000, **275**, 19513-19520.

32.   R. Alemany, C. J. van Koppen, K. Danneberg, M. Ter Braak and D. Meyer Zu Heringdorf, *Naunyn Schmiedebergs Arch. Pharmacol.*, 2007, **374**, 413-428.

33.   H. A. Neubauer and S. M. Pitson, *FEBS J.*, 2013, **280**, 5317-5336.

34.   A. S. Don and H. Rosen, *Biochem. Biophys. Res. Commun.*, 2009, **380**, 87-92.

35.   Z. Obradovic, K. Peng, S. Vucetic, P. Radivojac and A. K. Dunker, *Proteins*, 2005, **61 Suppl 7**, 176-182.

36.   B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker and V. N. Uversky, *Biochim. Biophys. Acta*, 2010, **1804**, 996-1010.

37.   S. M. Pitson, *Trends Biochem. Sci*, 2011, **36**, 97-107.

38.   S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, *Nucleic Acids Res.*, 1997, **25**, 3389-3402.

39.   A. Šali and T. L. Blundell, *J. Mol. Biol.*, 1993, **234**, 779-815.

40.   V. B. D.A. Case, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E., H. G. Duke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko,, S. L. T.S. Lee, T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui,, G. S. R. Salomon-Ferrer, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. and A. Wu and P.A. Kollman (2014), University of California, San Francisco.

41.   J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J Chem Theory Comput*, 2015, **11**, 3696-3713.

42.   W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J Chem Phys*, 1983, **79**, 926-935.

43.   J.-P. Ryckaert, G. Ciccotti and H. J. Berendsen, *J Comput Phys* 1977, **23**, 327-341.

44.   T. Darden, D. York and L. Pedersen, *J Chem Phys*, 1993, **98**, 10089-10092.

45.   D. R. Roe and T. E. Cheatham, 3rd, *J Chem Theory Comput*, 2013, **9**, 3084-3095.

46.   D. J. Gustin, Y. Li, M. L. Brown, X. Min, M. J. Schmitt, M. Wanska, X. Wang, R. Connors, S. Johnstone, M. Cardozo, A. C. Cheng, S. Jeffries, B. Franks, S. Li, S. Shen, M. Wong, H. Wesche, G. Xu, T. J. Carlson, M. Plant, K. Morgenstern, K. Rex, J. Schmitt, A. Coxon, N. Walker, F. Kayser and Z. Wang, *Bioorg. Med. Chem. Lett.*, 2013, **23**, 4608-4616.

47.   K. Yugesh, P. M. Thomas, M. G. Amanda, L. T. Jose, C. K. Perry, L. M. Morgan, L. M. Timothy and R. L. Kevin, *Biochem. J*, 2011, **440**, 345-353.

48.   S. N. Rao, M. S. Head, A. Kulkarni and J. M. LaLonde, *J Chem Inf Model*, 2007, **47**, 2159-2171.

49.   J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graph. Model.*, 2006, **25**, 247-260.

50.   A. Jakalian, B. L. Bush, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2000, **21**, 132-146.

51.   A. Jakalian, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2002, **23**, 1623-1641.

52.   G. Fu, H. Liu and R. J. Doerksen, *J. Phys. Chem. B*, 2012, **116**, 9580-9594.

53.   H. Gohlke, C. Kiel and D. A. Case, *J. Mol. Biol.*, 2003, **330**, 891-913.

54.   T. Hou, N. Li, Y. Li and W. Wang, *J. Proteome Res.*, 2012, **11**, 2982-2995.

55.   U. Consortium, *Nucleic Acids Res.*, 2011, gkr981.

56.   R. Liithy, J. Bowie and D. Eisenberg, *Nature*, 1992, **356**, 83-85.

57.   R. A. Laskowski, M. W. MacArthur, D. S. Moss and J. M. Thornton, *J. Appl. Crystallogr.*, 1993, **26**, 283-291.

58.   M. Harder, B. Kuhn and F. Diederich, *ChemMedChem*, 2013, **8**, 397-404.

Molecular BioSystems Accepted Manuscript