PCCP

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.

1

Importance of Protein Flexibility in Ranking Inhibitor Affinities:

Modeling the Binding Mechanisms of Piperidine Carboxamides

as Type I$^1/_2$ ALK Inhibitors

5

Xiaotian Kong[a], Peichen Pan[b], Dan Li[b], Sheng Tian[a], Youyong Li[a*], Tingjun Hou[a,b*]

7

[a]Institute of Functional Nano and Soft Materials (FUNSOM), Soochow University,
Suzhou, Jiangsu 215123, China.
[b]College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang
310058, China

12

13

Corresponding authors:
Tingjun Hou
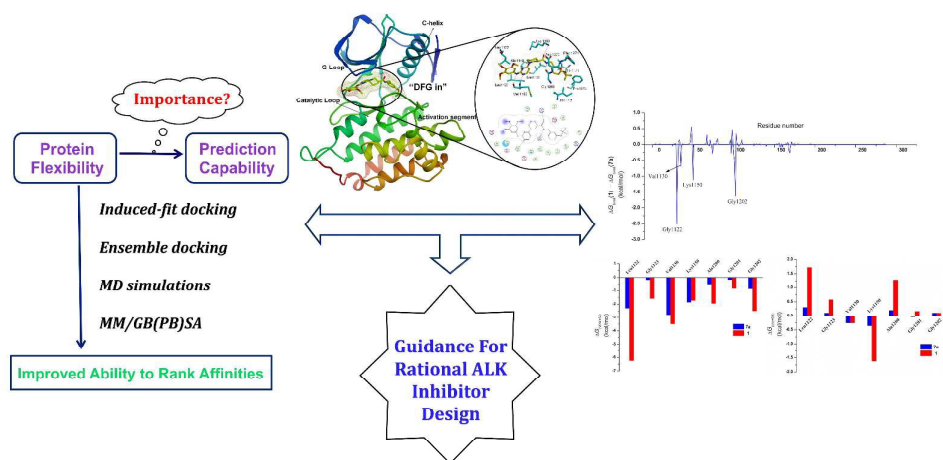**E-mail**: tingjunhou@zju.edu.cn   or   tingjunhou@hotmail.com

17

Youyong Li
E-mail: yyli@suda.edu.cn

20

23

24

25

26

27

28

29

30 **For Table of Contents Use Only**

31

32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

Abstract

Anaplastic lymphoma kinase (ALK) has gained increased attention as an attractive therapeutic target for the treatment of various cancers, especially non-small-cell lung cancer (NSCLC). Recently, piperidine carboxamides were reported as Type $I^1/_2$ inhibitors of ALK, which occupy both the ATP binding site and the back ATP hydrophobic cavity in DFG-in conformation. Due to the dynamic behavior of ALK in the binding of Type $I^1/_2$ inhibitors, the accurate predictions of the binding structures and relative binding potencies of these inhibitors are quite challenging. In this study, different modeling techniques, including molecular docking, ensemble docking based on multiple receptor conformations, molecular dynamics simulations and free energy calculations, were utilized to explore the binding mechanisms of piperidine carboxamides. Our predictions show that the conventional docking protocols are not sufficient to predict the relative binding potencies of the studied inhibitors with high accuracy, but incorporating protein flexibility before or after docking is quite effective to improve the prediction accuracy. Notably, the binding free energies predicted by MM/GBSA or MM/PBSA based on the MD simulations for the docked poses give the highest correlation with the experimental data, highlighting the importance of the inclusion of receptor flexibility on the accurate predictions of the binding potencies for Type $I^1/_2$ inhibitors of ALK. Furthermore, the comprehensive analysis of several pairs of representative inhibitors demonstrates the importance of hydrophobic interactions in improving the binding affinities of the inhibitors with the hot-spot residues surrounding the binding pocket. This work is expected to provide valuable clues for further rational design of novel and potent Type $I^1/_2$ ALK inhibitors.

Introduction

Anaplastic lymphoma kinase (ALK) is a trans-membrane protein, which belongs to the insulin receptor (IR) superfamily of receptor tyrosine kinases (RTKs). ALK is normally expressed in central nervous system and plays an important role in physiological development.[1-3] However, it has been reported that abnormal chimeric ALK proteins generated from chromosomal rearrangements are involved in the oncogenesis of various human cancers,[4] including anaplastic large cell lymphoma (ALCL),[5] inflammatory myofibroblastic tumor (IMT),[6] diffuse large B-cell lymphoma (DLBCL),[7] non-small cell lung cancer (NSCLC),[8] etc. Besides, amplification of ALK gene or constitutively activated point mutations in the full-length ALK gene are also the key oncogenic drivers in a subset of tumors.[9-12] Crizotinib (Compound 1), the first-in-class inhibitor of ALK approved by US Food and Drug Administration (FDA), has exhibited impressive therapeutic effects on most ALK fusion protein positive cancers.[13-15] Recently, great efforts have been made to discover and develop new generations of ALK inhibitors to meet the clinical requirements.[16-21]

The majority of reported small-molecule inhibitors of kinases target the ATP binding site (Type I inhibitors), and a small set of inhibitors occupy not only the ATP binding site but also an adjacent allosteric pocket (Type II inhibitors) formed by the conformational change of the Asp-Phe-Gly (DFG) motif from an active DFG-in conformation to an inactive DFG-out conformation. In 2012, Bryan and coworkers discovered piperidine carboxamide 1 as a potent and selective ALK inhibitor from a high-throughput screening of a proprietary sample collection.[22] Different from both Type I and II inhibitors, piperidine carboxamide 1 is a Type $I^1/_2$ inhibitor of ALK and occupies both the ATP binding site and the back ATP hydrophobic cavity in DFG-in conformation.[23] A series of analogues of piperidine carboxamide 1 were then synthesized and tested for their *in vitro* inhibitory activities against ALK. Epstein and coworkers subsequently obtained the crystal structure of human ALK in complex with piperidine carboxamide 2 (11j) (PDB entry: 4FNZ).[24] Nevertheless, elucidation of the binding mechanisms of ALK Type $I^1/_2$ inhibitors is of great importance to the future

4

105 design of ALK inhibitors, which cannot be afforded merely by experimental

106 approaches. Thus, in this study, in order to gain in-depth understanding of the

107 interactions between ALK and Type $I^1/_2$ inhibitors and identify the key structures

108 important to the potencies of inhibitors, a combined computational modeling strategy,

109 based on molecular docking, molecular dynamics (MD) simulations, free energy

110 calculations and free energy decomposition analysis, was employed.

111 Considering the dynamic behavior of ALK in the binding of the 36 derivatives of

112 piperidine carboxamides, the use of rigid receptor structures may hamper the correct

113 prediction of the ligand binding poses and relative binding potencies.[25] Herein,

114 various molecular modeling techniques were employed to deal with the flexibility of

115 ALK, such as induced-fit docking (IFD), ensemble docking and MD simulations. IFD

116 models the induced-fit phenomenon by refining the side chain conformations of

117 important active site residues. Ensemble docking incorporates protein flexibility by

118 using multiple receptor conformations (MRC) in molecular docking.[26,27] In this study,

119 an ensemble of MRC for ALK were generated by MD simulations and structural

120 clustering, and utilized for rigid receptor docking. Moreover, MD simulations after

121 docking were employed to relax the docked conformations. Based on the stable

122 snapshots derived from the MD simulations, both Molecular Mechanics/Generalized

123 Born Surface Area (MM/GBSA) and Molecular Mechanics/Poisson-Boltzmann

124 Surface Area (MM/PBSA) technologies were used to predict the binding affinities of

125 these 36 inhibitors. We expect that the detailed analysis of the binding modes, both

126 structurally and energetically, between ALK and Type $I^1/_2$ inhibitors can provide a

127 valuable strategy in rational design of novel, potent and selective kinase inhibitors

128 with a controlled activity profile.

129

Materials and methods

130

131 **1. Preparation of Proteins and Ligands.** The co-crystal structures of two piperidine

132 carboxamide inhibitors in complex with ALK (PDB entries: 4FNZ and 4DCE)

133 retrieved from the RCSB Brookhaven Protein Data Bank (PDB) were used as the

134   templates for molecular docking,[22, 24] and the missing loop segments and residues near

135   the active site were constructed using Discovery Studio 2.5.[26] Furthermore, each

136   complex was prepared by the *Protein Preparation Wizard* in Schrodinger 9.0,[27]

137   including adding hydrogen atoms, deleting crystallographic water molecules,

138   assigning protonation states and partial charges, and optimizing the structure using the

139   OPLS-2005 force field.[28]

140   The 3D structures of all the 36 piperidine carboxamides were sketched using

141   *Maestro* and minimized with *Macromodel* in Schrodinger[27] using the OPLS-2005

142   force field.[28] The 2D structures of the 36 inhibitors and their biological activities

143   against ALK (pIC_{50}) are summarized in Table 1. All these compounds were processed

144   by using the *ligprep* module in Schrodinger with the protomers and tautomers

145   enumerated using *Epik* at pH=7.0.[29] Default settings were used for the other

146   parameters.

147

148   **2. Generation of Representative Multiple Protein Conformations.** In our scheme,

149   based on the two crystal structures (PDB entries: 4DCE and 4FNZ)[22, 24], MD

150   simulations were employed to generate the representative protein conformations for

151   ensemble docking.

152   The two inhibitors in 4DCE and 4FNZ were optimized by semi-empirical AM1

153   method and the electrostatic potentials were computed at Hartree-Fock (HF)

154   SCF/6-31* level in Gaussian 09,[30] and then the atomic partial charges were obtained

155   by fitting the electrostatic potentials using the RESP fitting technique.[31] The partial

156   charges and the force field parameters were generated with the *antechamber* suite in

157   AMBER11.[43] The AMBER ff99SB force field[45] and the general AMBER force field

158   (gaff)[32] were used for the inhibitors and proteins, respectively. All missing atoms in

159   the protein were added using the *tleap* program, and the counter ions of Cl− were

160   placed into each system to neutralize the charge.[33] Each protein-ligand complex was

161   immersed into a periodic TIP3P water box extended 8 Å from any solute atom. The

162   long-range electrostatics was handled by the particle Mesh Ewald (PME) algorithm,[34]

163   and the non-bonded cutoff of 8 Å was used for real-space interactions.

164   Prior to MD simulations, three-stage minimizations were performed to relax each

165 system using the *sander* program.[35] Firstly, 500 cycles of steepest descent and 500

166 cycles of conjugate gradient minimizations were performed with the protein backbone

167 restrained (50 kcal/mol/Å^2). Secondly, another 1000 cycles of minimizations with

168 relatively weaker restrain (10 kcal/mol/Å^2) were carried out. Finally, the whole

169 system was minimized without any restraints (1000 cycles of steepest descent and

170 4000 cycles of conjugate gradient minimizations).

171   Each system was gradually heated from 0 to 300 K over 50 ps with 2.0

172 kcal/mol/Å^2 restrain on the protein, followed by 50 ps MD simulations at 300 K with

173 the same restrain. Afterwards, 50 ns NPT MD simulations under constant temperature

174 and pressure (T = 300K and P = 1 atm) were employed to produce trajectories.

175 SHAKE algorithm was employed to restrain all covalent bonds involving hydrogen

176 atoms, and the time interval was set to 2.0 fs.[36] During the sampling process, the

177 coordinates of each complex were saved every 20 ps.

178   For each system, 200 conformations were evenly extracted from the last stable 40

179 ns MD trajectory. The previous studies show that the use of too many receptor

180 conformations may not necessarily improve the docking performance because a large

181 number of more false positives may reduce the enrichment rate in VS.[37, 38] Thus, the

182 k-means clustering algorithm[39] based on the pairwise root-mean-square displacements

183 (RMSDs) between the extracted MD conformations was utilized to reduce the initial

184 size of MRC. Through repeated iterative minimization of the sum of distances from

185 each object to its cluster centroid over all clusters, 10 representative structures were

186 eventually generated for each system.[40, 41]

187

188 **3. Docking protocols.** All the inhibitors were docked into the active site of 4DCE or

189 4FNZ,[22, 24] and the docking calculations were performed using the programs

190 implemented in Schrodinger 9.0.[27] For each system, the receptor grid box for docking

191 was generated and centered on the ligand in the active site of ALK with the box size

192 of 20 Å × 20 Å × 20 Å using the *Receptor Grid Generation* protocol of Schrodinger

193 *9.0*. The scaling factor for van der Waals radii was set to 1.0.[29]

194   First, 36 inhibitors were docked into the active site of ALK by using the rigid

195   receptor docking (RRD) protocol with the extra precision (XP) scoring mode of

196   *Glide*.[42] During the docking process, the protein was fixed while the inhibitors were

197   flexible. Then, considering the importance of receptor flexibility to ligand binding, the

198   induced fit docking (IFD) protocol, which incorporates the flexibility of the side

199   chains of the receptor, was then employed. In the IFD calculations, the side chains of

200   the residues within 5 Å of each inhibitor were relaxed by *Prime* in Schrodinger 9.0.[29]

201   The best receptor-ligand complex was evaluated by the XP scoring mode.

202   In order to overcome the difficulty in accurately estimating the electrostatic

203   interaction, the QM-Polarized Ligand Docking (QPLD) protocol in Schrodinger,

204   which combine *Glide* and the QM/MM method Q-site, was employed.[29] The best

205   binding poses were firstly generated by the regular *Glide* docking with the XP scoring

206   mode, and then the QM-ESP charges at B3LYP/6-31G* level of theory within the

207   protein environment was computed. Finally, the resulting poses with the QM-ESP

208   atomic charges were re-docked by *Glide* and rescored by the XP scoring mode in

209   *Glide*.

210   In ensemble docking, the 36 inhibitors were successively docked into the 10

211   representative structures generated from the MD simulations based on the crystal

212   structure of 4DCE or 4FNZ by using the RRD protocol with the XP scoring mode of

213   *Glide* in Schrodinger, and one pose with the best XP score was saved for each ligand.

214

215   **4. Molecular Dynamics (MD) Simulations.** The binding structures of the 36

216   inhibitors bound with ALK predicted by RRD based on the crystal structure of 4FNZ

217   were submitted to 5 ns NPT MD simulations (T = 300K and P = 1 atm). The details of

218   the MM minimizations and MD simulations for each system are described in the

219   previous section "Generation of Representative Multiple Protein Conformations".

220   During the sampling process, the coordinates of each complex were saved every 10

221   ps.

222

223   **5. MM/GBSA and MM/PBSA Binding Free Energy Calculations.** Based on the

224   150 snapshots evenly extracted from the stable 2~5 ns MD trajectory, the binding free

225   energy for each system was calculated by using the MM/PBSA and MM/GBSA

226   approaches through Equations 1-4[27, 43-54]:

227
$$\Delta G_{bind} = G_{com} - (G_{rec} + G_{lig})$$
    (1)

228
$$\Delta G_{bind} = \Delta H - T\Delta S \approx \Delta E_{MM} + \Delta G_{sol} - T\Delta S$$
    (2)

229
$$\Delta E_{MM} = \Delta E_{internal} + \Delta E_{electrostatic} + \Delta E_{vdw}$$
    (3)

230
$$\Delta G_{sol} = \Delta G_{PB/GB} + \Delta G_{SA}$$
    (4)

231   where ΔE_{MM} is the gas-phase interaction energy between receptor and ligand, which

232   consists of the internal energy ($\Delta E_{internal}$), electrostatic (ΔE_{ele}) and van der Waals

233   (ΔE_{vdw}) terms but $\Delta E_{internal}$ is canceled by using the single-trajectory protocol;

234    $\Delta G_{GB/PB}$ and ΔG_{SA} represent the polar and non-polar contributions of the solvation

235   free energy (ΔG_{sol}), respectively; the entropy term ($-T\Delta S$), which can be estimated

236   using normal-mode analysis, was neglected for congeneric series due to the high

237   computational cost and low prediction accuracy.[27] The van der Waals (ΔE_{vdw}) and

238   electrostatic (ΔE_{ele}) terms were calculated using the *sander* module in AMBER11.[35]

239   The modified GB model (GB^{OBC1}) with the parameters endowed by Onufriev and

240   coworker (*igb=2*), or Poisson-Boltzmann (PB) equation solved by Rocchia et al, was

241   utilized to calculate the polar solvation free energy.[55, 56] The non-polar contribution of

242   solvation energy (ΔG_{SA}) was predicted by the solvent accessible surface area (SASA)

243   computed from the LCPO method[57]. The solute dielectric constant (ε_{in}) was set to 1, 2

244   or 4, and the solvent dielectric constant was set to 80.

245

246   **6. MM/GBSA Free Energy Decomposition.** In order to explore the contribution of

247   each residue to the interaction between ALK and each inhibitor, the total binding free

248   energy of each inhibitor was decomposed into inhibitor-residue pairs using the

249   MM/GBSA decomposition analysis supported by AMBER 11.[58-60] Each

250   inhibitor-residue interaction consists of four parts: van der Waals contribution (ΔG_{vdw}),

251   electrostatic contribution (ΔG_{ele}), polar solvation contribution (ΔG_{GB}) and non-polar

252   solvation contribution (ΔG_{SA}). The polar part of the solvation free energy (ΔG_{GB}) was

253   calculated by the GB model with the parameter developed by Onufriev and coworkers

254   (igb = 2),[55] and the non-polar part (ΔG_{SA}) was computed by the SASA based on the

255   ICOSA technique.[58]

256

Results and Discussions

258   **1. Static Structure Analysis of the Binding Mode of Type I$^1/_2$ Inhibitor.** As shown

259   in the X-ray crystal structure of piperidine caboxmide 2 (11j) in complex with ALK,

260   this Type I$^1/_2$ inhibitor binds in two regions: the ATP binding pocket and the back

261   ATP hydrophobic activation loop region in DFG-in conformation. As illustrated in

262   Figure 1, the ATP binding pocket is mainly surrounded by the residues Leu1122,

263   Lys1150, Leu1196, Leu1198, Met1199, Gly1202 and Leu1256, and the activation

264   loop binding region is surrounded by the residues Ile1171, Phe1174, Ile1179, His1247,

265   Gly1269, Asp1270 and Phe1271. The anilinic NH and the N1 of the aminopyrimidine

266   ring of 11j form two H-bonds with the backbone NH and carbonyl O of Met1199 in

267   the hinge region, respectively. The side chain NH of Lys1150 forms a H-bond with

268   the amide carbonyl located between the piperidine and aromatic ring, and the amide

269   NH between which establishes a H-bond with the backbone carbonyl of Gly1269.

270

271   **2. Molecular Docking based on Single Protein Conformation.** At first, for each

272   crystal structure (4DCE or 4FNZ), the inhibitor was extracted from the complex and

273   re-docked into the binding pocket. The RMSD between the predicted binding pose

274   and the respective experimental structure was calculated. For 4DCE and 4FNZ, the

275   RMSDs predicted by *Glide* based on XP scoring mode are 0.64 and 0.50 Å,

276   respectively. Obviously, the *Glide* docking can reproduce the experimental binding

277   poses with high accuracy.

278   Then, all the 36 compounds listed in Table 1 were docked into the active site of

279   ALK (4FNZ) using three different docking protocols in Schrodinger: RRD, IFD and

280   QPLD, and the corresponding docking scores are summarized (Table 1). The

281   performance of each docking protocol was evaluated by the linear correlation between

282   the docking scores and the experimental pIC_{50} values. As shown in Figure 2, the

283   correlation coefficients (r^2) for RRD, IFD and QPLD are 0.10, 0.17 and 0.02,

284   respectively. The performance of IFD is better than those of RRD and QPLD,

285   implying that incorporating protein flexibility can enhance the docking accuracy.

286   However, it seems that the protein flexibility cannot be well characterized by these

287   three docking protocols. RRD only handles the flexibilities of the ligands, and IFD

288   merely introduces limited flexibility into the side chains of the receptor while not the

289   flexibility of the whole protein. Based on the above analysis, any of the traditional

290   docking methodologies used above is not necessarily sufficient to correctly evaluate

291   the binding potencies with relatively high confidence.

292

293   **3. Incorporating Flexibility and Dynamics into Receptor**

294   **3.1 Docking into an Ensemble of ALK Conformations.** There have been many

295   attempts to incorporate protein flexibility into the prediction of protein-ligand

296   interactions, such as molecular docking based on MRC,[40, 61-63] MD simulations,[64]

297   normal modes analysis (NMA),[65] side chain rotamers,[66] and others[67, 68]. Some

298   successes have been achieved, but none of them is acknowledged as a universal

299   strategy to handle protein flexibility.[63, 69]

300   For each complex, 200 conformations evenly extracted from the last stable 40 ns

301   MD trajectory (Figure 3) were clustered by the *k*-means clustering algorithm, and 10

302   representative structures were eventually obtained for each system. All inhibitors were

303   then docked into the 10 representative conformations of each system by using RRD.

304   As shown in Figure S1 of Supporting Information, most conformations generated by

305   the MD simulations even exhibit better capability to rank the binding potencies than

306   the crystal structures. The protein backbones of the conformations extracted from the

307   MD simulations have obvious conformational rearrangement, highlighting the

308   importance of protein flexibility on the prediction of protein-ligand interactions.

309   However, not all MD conformations perform better than the crystal structure

310   ($r^2_{4FNZ\_1550} = 0.05$, $r^2_{4DCE\_960} = 0.02$, $r^2_{4DCE\_1650} = 0.00$, $r^2_{4DCE\_2060} = 0.00$). When an

311   inappropriate protein structure is used in RRD, the docking method may fail to give

312   the correct binding pose of a ligand because the ligand binding space may be partially

313   occupied by the side chains of some residues. This phenomenon makes us wonder

314   how to choose the "good" docking conformations while avoid the "bad" ones. In fact,

315   it is difficult to determine the "good" or "bad" structures beforehand. Therefore, we

316   used both the highest and average docking scores of all the 10 protein conformations

317   for each inhibitor to rank the binding affinities of the inhibitors (Figure 4). The results

318   show that for both templates (4FNZ and 4DCE) the highest docking scores and the

319   average docking scores have better correlations ($r^2_{4DCE\_average} = 0.26$, $r^2_{4FNZ\_average} =$

320   0.40, $r^2_{4DCE\_highest} = 0.30$, and $r^2_{4FNZ\_highest} = 0.41$) with the experimental data than the

321   docking scores based on any single protein structure ($r^2_{4FNZ} = 0.10$ and $r^2_{4DCE} = 0.00$).

322   Apparently, the results shown above illustrate that the use of an ensemble of protein

323   structures in molecular docking can indeed improve the prediction accuracy.

324

325   **3.2 MD Simulations based on Docking Results and Binding Free Energy**

326   **Calculations.** MD simulation is recognized as an effective way to deal with the

327   flexibility and dynamics for both protein and ligand. Since the *Glide* docking based on

328   4FNZ yields similar binding geometries for all the 36 inhibitors, the binding

329   complexes of the docked inhibitors with the protein in 4FNZ were used as the initial

330   structures for the 5 ns MD simulations. To explore the overall stabilities during the

331   MD simulations, the RMSDs of the representative inhibitor-ALK complexes (7a, 1,

332   11a, 11d, 11k, 11q, and 11s) were analyzed. As shown in Figure 5, all the studied

333   systems achieve equilibria after ~1 ns. Furthermore, the root-mean-square fluctuation

334   (RMSF) *versus* each individual residue of the selected systems are illustrated in

335   Figure 6. We can observe that the RMSF distributions and trends of the fluctuations

336   of the selected systems are roughly identical. The residues with higher dynamics are

337   mostly located in the flexible loop regions (Gly-rich loop, C-loop and activation loop),

338   and the residues having strong interactions with inhibitors, both in the ATP binding

339   pocket and activation loop binding region, show relatively higher rigidity.

340   The binding free energies and the energy components for each inhibitor-ALK

341   system were predicted by the MM/PBSA and MM/GBSA approaches based on the

342   150 snapshots extracted evenly from the stable 2~5 ns MD trajectory. The results

343   shown in Figure 7 demonstrate that the performance of MM/PBSA ($r^2 = 0.46$ at $\varepsilon_{in}=1$,

344    $r^2 = 0.42$ at $\varepsilon_{in} = 2$, and $r^2 = 0.40$ at $\varepsilon_{in} = 4$) is comparative to that of MM/GBSA ($r^2 =$

345   0.44 at $\varepsilon_{in} = 1$, $r^2 = 0.43$ at $\varepsilon_{in} = 2$, and $r^2 = 0.41$ at $\varepsilon_{in} = 4$). Both methods exhibit much

346   stronger capability in ranking the binding affinities of the inhibitors than RRD, IFD,

347   QPLD and even ensemble docking. When the solute dielectric constant (ε_{in}) was set to

348   1, the linear correlation between the predicted binding affinities and the experimental

349   pIC$_{50}$ values is the highest for both MM/PBSA and MM/GBSA.

350   In addition, the correlations between the experimental activities and individual

351   energy terms were compared. It can be observed that the non-polar ($\Delta E_{vdw} + \Delta G_{SA}$;

352   MM/GBSA: $r^2 = 0.39$) contributions shows much higher linear correlations with the

353   experimental data than the polar contributions ($\Delta E_{ele} + \Delta G_{GB}$; MM/GBSA: $r^2 = 0.21$ at

354    $\varepsilon_{in} = 1$, $r^2 = 0.21$ at $\varepsilon_{in} = 2$ and $r^2 = 0.21$ at $\varepsilon_{in} = 4$) (Figure S2). Therefore, it is quite

355   possible that the non-polar contributions are more important than the polar

356   contributions to determine the discrepancy of the binding affinities of the studied

357   inhibitors. The energy components of the binding free energy for each complex are

358   also listed in Table S1 of Supporting Information.

359   Furthermore, by comparing the total binding free energies (Table S1 of

360   Supporting Information), we can observe that the predicted binding free energy of

361   compound 1 is higher (-50.17 kcal/mol) than those of the analogues with hinge region

362   (7a-7h) and alkyl linker (11a-11g, 14) modifications, which is consistent with the

363   experimental results. In addition, most compounds with phenyl substituent located in

364   the activation loop binding region (11h-11v) display improved binding affinities

365   compared with 11b. Moreover, the position of one specific substituent can obviously

366   affect the binding of inhibitors, such as 11o ($\Delta G_{bind} = -53.63$ kcal/mol), 11p ($\Delta G_{bind} =$

367   -50.51 kcal/mol), 11q ($\Delta G_{bind} = -57.26$ kcal/mol) and 11r ($\Delta G_{bind} = -52.97$ kcal/mol)

368   with strong electron withdrawing group located in different positions of phenyl. The

369   above results obtained from MM/GBSA ($\varepsilon_{in} = 1$) are in relatively good accordance

370   with the experimental data.

371

372   **4. Identification of Key Residues Responsible for Inhibitor Binding.** To reveal the

373   key residues involved in the binding process and understand the possible molecular

374   mechanism of the important substituents that can improve the binding affinity with

375   ALK, the binding free energies of the representative inhibitors (7a, 1, 11k, 11q and

376   11s) predicted by MM/GBSA (ε_{in} = 1) were decomposed into the contributions from

377   inhibitor-residue pairs. The pIC$_{50}$ values of the compounds 11k, 11q and 11s with

378   different substituents that interact with the activation loop binding region is about 10

379   times higher than that of compound 1, while compound 7a with the R$_1$ modification in

380   the hinge region shows the lowest activity. According to Figures 8, 9 and 10, the

381   favorable energy contributions primarily originate from the residues around two

382   regions: ATP binding pocket (Leu1122, Val1130, Lys1150, Leu1196, Leu1198,

383   Met1199, Gly1202 and Leu1256) and activation loop binding region (Ile1171,

384   Phe1174, Ile1179, Phe1271, Gly1269 and Asp1270), where the hydrogen bonding and

385   hydrophobic interactions between the inhibitors and ALK play the major roles. Except

386   for the residues Gly1269 and Asp1270, the contribution of the polar residues (e.g.

387   Gly1123, His1123, Lys1150, Gly1101, Gly1202, Asp1203 and His1247) are

388   relatively not significant and yield slight energetic difference among different

389   inhibitors. The contributions of the non-polar residues account for a large proportion

390   of the binding free energies, especially the residues Leu1122, Val130, Leu1196,

391   Leu1198, Met1199, Leu1256, Ile1171, Phe1174, Ile1179 and Phe1271. The binding

392   features of the selected compounds were further compared to understand the structural

393   requirements of inhibitors for improved binding affinities, which will guide rational

394   design of more potent and selective Type I^1/$_2$ inhibitors of ALK. Because the

395   compounds modified in the activation loop binding region show improved activities,

396   the analyses of these inhibitors will be emphasized below.

397

398   **4.1 Comparison of Binding Modes of 7a and 1.** The replacement of

399   trimethoxyphenyl in compound 1 by methyl (compound 7a) at R$_1$ significantly

400   decreases the binding affinity. As we can see from Table 2, the predicted binding free

401   energy of 1 (-50.17 kcal/mol) is much stronger than that of 7a (-42.48 kcal/mol),

402   which is consistent with the observed activities.[22] The difference of the non-polar

403   contributions between compounds 1 (-66.83 kcal/mol) and 7a (-54.97 kcal/mol) can

404   reach 11.86 kcal/mol. Although the unfavorable polar contribution of 7a (12.50

405   kcal/mol) is lower than that of 1 (16.66 kcal/mol), it cannot compensate the decrease

406   of the binding free energy caused by the van der Waals interactions. In Figure 8(a),

407   we can find that the substituent of R_1 locates in a narrow groove sandwiched by the

408   G-loop region and hinge region, and the trimethoxyphenyl group in compound 1

409   forms more effective interactions with the residues surrounding the G-loop region and

410   hinge region, especially the residues Leu1122, Gly1123, Ala1200, Gly1201 and

411   Gly1202. Meanwhile, the replacement of R_1 may result in the pose change of the

412   inhibitors and have more effective interactions with residues far away from R_1 group,

413   such as residue Lys1150. As shown in the residue-inhibitor interaction spectra (Figure

414   9), the difference of the energy contributions of Leu1122 between 1 (-4.52 kcal/mol)

415   and 7a (-2.00 kcal/mol) is the largest. This is because the trimethoxyphenyl group in

416   compound 1 can form stronger non-polar interactions with the hydrophobic side chain

417   of Leu1122 than the methyl group in 7a. In addition, the residues Lys1150 and

418   Gly1202 also play critical roles in rendering the difference of the binding free

419   energies, which are primarily determined by the van der Waals interactions as well.

420   Besides, the contributions of the hinge residues (Met1123, Ala1200 and Gly1201) and

421   Val1130 to the binding of compound 1 are higher than those to the binding of

422   compound 7a, but the differences are not obvious.

423

424   **4.2 Comparison of Binding Modes of 1, 11k, 11q and 11s.** Compounds 1, 11k, 11q,

425   and 11s have different phenyl substitutions, which occupy the extended binding

426   pocket formed by the "DFG-in" activation loop. The predicted binding free energies

427   of 11k, 11q, 11s are relatively stronger than those of the other modified inhibitors,

428   which is roughly consistent with the experimental data. As is shown in Figure 10,

429   almost all the residues surrounding the activation loop binding pocket, including

430   Ile1171, Ile1179, Leu1240, Phe1245, His1247, Ile1268, Gly1269 and Phe1271, are

431 hydrophobic and have stronger interactions with compounds 11k, 11s and 11q than

432 with compound 1. Detailed analysis indicates that the contribution of Ile1179 flanking

433 the extended hydrophobic pocket to the binding of compound 1 is only -1.24 kcal/mol,

434 which is obviously weaker than those to the binding of the compounds 11k, 11q and

435 11s (-2.24, -2.24 and -2.22 kcal/mol). The contributions of the residues Ile1171 and

436 Phe1271 to the binding of compounds 11k, 11q and 11s are more favorable than those

437 to the binding of compound 1, which can be explained by the fact that the side chains

438 of Ile1171 and Phe1271 tend to form stronger interactions with the trifuoromethoxy

439 group, methy carboxylate group and phenyl ring than with the methyl group of

440 compound 1 (Figure 8). Besides, it can be seen in Figure 8(c) that the strong electron

441 withdrawing methyl carboxylate of compound 11q forms a stable H-bond with the

442 lmidazolidine ring of the positively-charged residue His1247, which is supported by

443 the per-residue energy decomposition analysis that His1247 shows more favorable

444 contribution to the binding of compound 11q (-2.46 kcal/mol) than to the binding of

445 the compounds 1, 11k and 11s (-0.32, -0.58 and -0.80 kcal/mol).

446

447 **5. Suggestions for the Design of Improved Type I$^1/_2$ ALK Inhibitors.** Although the

448 results given by molecular modeling may be not satisfactory, we are ready to provide

449 several guidelines for the design of Type I$^1/_2$ ALK inhibitors.

450 (1) Incorporation of the flexibility and dynamics of ALK is quite essential to give

451 correct predictions of the binding structures of Type I1/2 inhibitors. Based on the

452 structures afforded by the MD simulations, ensemble docking can obviously improve

453 the ranking ability of the binding potencies of the studied inhibitors, and the predicted

454 binding affinities based on the MD simulations for the docked structures yield the best

455 correlation to the experimental data, highlighting the importance of incorporating

456 protein flexibility in predicting protein-ligand interactions.

457 (2) According to the per-residue energy decomposition analysis, the favorable

458 electrostatic contribution (ΔE_{ele}) is counteracted by the unfavorable polar desolvation

459 energy, and the balanced contribution of the polar interactions shows negative effect

460 on ligand binding. Thus, increasing the hydrophobic contributions may be helpful to

16

461   improve activity. Furthermore, both the predicted and experimental results suggest

462   that the increase of the hydrophobic interactions with the back ATP activation loop is

463   highly preferred. Potential ALK inhibitors can be designed by forming more effective

464   hydrophobic interactions with the surrounding key residues, such as Ile1171, Phe1174,

465   Ile1179, and Phe1271.

466   (3) Hydrogen bonds play critical roles in stabilizing the inhibitors in both the ATP

467   binding pocket and the back ATP activation loop region. The piperidine carboxamide

468   scaffold forms four H-bonds with the backbone NH and carbonyl O of Met1199, the

469   side chain NH of Lys1150, and the backbone carbonyl O of Gly1269. Except for the

470   shared H-bonds, the formation of the H-bonds between the modified moieties and the

471   protein would be beneficial for inhibitor binding; for example, the improved activity

472   of compound 11q can be explained by the formation of a stable H-bond between 11q

473   and the imidazolidine ring of the positively charged residue His1247.

474

475   **Conclusion**

476   In this study, a computational strategy, which combines molecular docking, ensemble

477   docking, MD simulations and free energy calculations, was employed to explore the

478   binding mechanisms of the Type $I^1/_2$ inhibitors of ALK. Based on our predictions, the

479   conventional docking methodologies, such as glide docking, QM-polarized docking

480   and induced-fit docking, are not sufficient to predict the relative binding potencies of

481   the studied inhibitors with high accuracy. Incorporating protein flexibility before or

482   after docking is relatively more effective to improve the prediction accuracy. Among

483   them, the predicted binding free energies based on the MD simulations for the docked

484   poses give the highest correlation with the experimental data, highlighting the

485   importance of incorporating protein flexibility in predicting protein-ligand

486   interactions. The MM/GBSA binding free energy calculations illustrate that the

487   non-polar interactions dominated by the van der Waals energies play major roles in

488   the binding of the studied inhibitors to ALK and determine the difference of the

489   binding affinities of the inhibitors. The decomposition analysis of the binding free

energy on a per-residue basis has identified possible several key residues for ligand

binding. In addition, the comprehensive analysis of several pairs of representative

inhibitors also demonstrates the importance of hydrophobic interactions in improving

the binding affinities of the inhibitors with the hot-spot residues surrounding the ATP

binding pocket and the back ATP activation loop region. These findings provide a

better structural understanding of Type $I^{1}/_{2}$ inhibitors targeting the DFG-in form of

ALK and valuable clues for further rational design of new potent inhibitors of ALK.

Acknowledgment

Supporting information

Table S1. The individual energy components of the binding free energies predicted by MM/GBSA based on three different solute dielectric constants (kcal/mol); **Figure S1**. Correlations between the experimental bioactivities and the docking scores based on different conformations extracted from the MD trajectories for (a) 4FNZ and (b) 4DCE; **Figure S2**. Correlations between the experimental bioactivities and (a) nonpolar energy components or (b) polar energy components based on three different solute dielectric constants.

References

1.   Dirks, W. G.; Fähnrich, S.; Lis, Y.; Becker, E.; MacLeod, R. A.; Drexler, H. G., Expression and functional analysis of the anaplastic lymphoma kinase (ALK) gene in tumor cell lines.

519   *International Journal of cancer* **2002**, 100, 49-56.

520   2.   Palmer, R.; Vernersson, E.; Grabbe, C.; Hallberg, B., Anaplastic lymphoma kinase: signalling in
521   development and disease. *Biochem. J* **2009**, 420, 345-361.

522   3.   Cheng, M.; R Ott, G., Anaplastic lymphoma kinase as a therapeutic target in anaplastic large cell
523   lymphoma, non-small cell lung cancer and neuroblastoma. *Anti-Cancer Agents in Medicinal*
524   *Chemistry-Anti-Cancer Agents)* **2010**, 10, 236-249.

525   4.   Elenitoba-Johnson, K. S.; Crockett, D. K.; Schumacher, J. A.; Jenson, S. D.; Coffin, C. M.;
526   Rockwood, A. L.; Lim, M. S., Proteomic identification of oncogenic chromosomal translocation
527   partners encoding chimeric anaplastic lymphoma kinase fusion proteins. *Proceedings of the*
528   *National Academy of Sciences* **2006**, 103, 7402-7407.

529   5.   Morris, S. W.; Kirstein, M. N.; Valentine, M. B.; Dittmer, K. G.; Shapiro, D. N.; Saltman, D. L.;
530   Look, A. T., Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's
531   lymphoma. *Science* **1994**, 263, 1281-1284.

532   6.   Griffin, C. A.; Hawkins, A. L.; Dvorak, C.; Henkle, C.; Ellingham, T.; Perlman, E. J., Recurrent
533   involvement of 2p23 in inflammatory myofibroblastic tumors. *Cancer Research* **1999**, 59,
534   2776-2780.

535   7.   Reichard, K. K.; McKenna, R. W.; Kroft, S. H., ALK-positive diffuse large B-cell lymphoma:
536   report of four cases and review of the literature. *Modern pathology* **2007**, 20, 310-319.

537   8.   Soda, M.; Choi, Y. L.; Enomoto, M.; Takada, S.; Yamashita, Y.; Ishikawa, S.; Fujiwara, S.-i.;
538   Watanabe, H.; Kurashina, K.; Hatanaka, H., Identification of the transforming EML4–ALK fusion
539   gene in non-small-cell lung cancer. *Nature* **2007**, 448, 561-566.

540   9.   Tuma, R. S., ALK gene amplified in most inflammatory breast cancers. *Journal of the National*
541   *Cancer Institute* **2012**, 104, 87-88.

542   10. Salido, M.; Pijuan, L.; Martínez-Avilés, L.; Galván, A. B.; Canadas, I.; Rovira, A.; Zanui, M.;
543   Martínez, A.; Longarón, R.; Sole, F., Increased ALK gene copy number and amplification are
544   frequent in non-small cell lung cancer. *Journal of Thoracic Oncology* **2011**, 6, 21.

545   11. George, R. E.; Sanda, T.; Hanna, M.; Fröhling, S.; Luther II, W.; Zhang, J.; Ahn, Y.; Zhou, W.;
546   London, W. B.; McGrady, P., Activating mutations in ALK provide a therapeutic target in
547   neuroblastoma. *Nature* **2008**, 455, 975-978.

548   12. Janoueix-Lerosey, I.; Lequin, D.; Brugières, L.; Ribeiro, A.; de Pontual, L.; Combaret, V.;
549   Raynal, V.; Puisieux, A.; Schleiermacher, G.; Pierron, G., Somatic and germline activating
550   mutations of the ALK kinase receptor in neuroblastoma. *Nature* **2008**, 455, 967-970.

551   13. Christensen, J. G.; Zou, H. Y.; Arango, M. E.; Li, Q.; Lee, J. H.; McDonnell, S. R.; Yamazaki, S.;
552   Alton, G. R.; Mroczkowski, B.; Los, G., Cytoreductive antitumor activity of PF-2341066, a novel
553   inhibitor of anaplastic lymphoma kinase and c-Met, in experimental models of anaplastic
554   large-cell lymphoma. *Molecular Cancer Therapeutics* **2007**, 6, 3314-3322.

555   14. Cui, J. J.; Tran-Dubé, M.; Shen, H.; Nambu, M.; Kung, P.-P.; Pairish, M.; Jia, L.; Meng, J.; Funk,
556   L.; Botrous, I., Structure based drug design of crizotinib (PF-02341066), a potent and selective
557   dual inhibitor of mesenchymal–epithelial transition factor (c-MET) kinase and anaplastic
558   lymphoma kinase (ALK). *Journal of medicinal chemistry* **2011**, 54, 6342-6363.

559   15. Zou, H. Y.; Li, Q.; Lee, J. H.; Arango, M. E.; McDonnell, S. R.; Yamazaki, S.; Koudriakova, T.
560   B.; Alton, G.; Cui, J. J.; Kung, P.-P., An orally available small-molecule inhibitor of c-Met,
561   PF-2341066, exhibits cytoreductive antitumor efficacy through antiproliferative and
562   antiangiogenic mechanisms. *Cancer Research* **2007**, 67, 4408-4417.

563   16.   Choi, Y. L.; Soda, M.; Yamashita, Y.; Ueno, T.; Takashima, J.; Nakajima, T.; Yatabe, Y.;
564   Takeuchi, K.; Hamada, T.; Haruta, H., EML4-ALK mutations in lung cancer that confer
565   resistance to ALK inhibitors. *New England Journal of Medicine* **2010**, 363, 1734-1739.

566   17.   Galkin, A. V.; Melnick, J. S.; Kim, S.; Hood, T. L.; Li, N.; Li, L.; Xia, G.; Steensma, R.;
567   Chopiuk, G.; Jiang, J., Identification of NVP-TAE684, a potent, selective, and efficacious
568   inhibitor of NPM-ALK. *Proceedings of the National Academy of Sciences* **2007**, 104, 270-275.

569   18.   Katayama, R.; Khan, T. M.; Benes, C.; Lifshits, E.; Ebi, H.; Rivera, V. M.; Shakespeare, W. C.;
570   Iafrate, A. J.; Engelman, J. A.; Shaw, A. T., Therapeutic strategies to overcome crizotinib
571   resistance in non-small cell lung cancers harboring the fusion oncogene EML4-ALK. *Proceedings*
572   *of the National Academy of Sciences* **2011**, 108, 7535-7540.

573   19.   Lovly, C. M.; Heuckmann, J. M.; de Stanchina, E.; Chen, H.; Thomas, R. K.; Liang, C.; Pao, W.,
574   Insights into ALK-driven cancers revealed through development of novel ALK tyrosine kinase
575   inhibitors. *Cancer Research* **2011**, 71, 4920-4931.

576   20.   Mesaros, E. F.; Burke, J. P.; Parrish, J. D.; Dugan, B. J.; Anzalone, A. V.; Angeles, T. S.; Albom,
577   M. S.; Aimone, L. D.; Quail, M. R.; Wan, W., Novel 2, 3, 4, 5-tetrahydro-benzo [< i> d</i>]
578   azepine derivatives of 2, 4-diaminopyrimidine, selective and orally bioavailable ALK inhibitors
579   with antitumor efficacy in ALCL mouse models. *Bioorganic & medicinal chemistry letters* **2011**,
580   21, 463-466.

581   21.   Sakamoto, H.; Tsukaguchi, T.; Hiroshima, S.; Kodama, T.; Kobayashi, T.; Fukami, T. A.;
582   Oikawa, N.; Tsukuda, T.; Ishii, N.; Aoki, Y., CH5424802, a selective ALK inhibitor capable of
583   blocking the resistant gatekeeper mutant. *Cancer cell* **2011**, 19, 679-690.

584   22.   Bryan, M. C.; Whittington, D. A.; Doherty, E. M.; Falsey, J. R.; Cheng, A. C.; Emkey, R.; Brake,
585   R. L.; Lewis, R. T., Rapid development of piperidine carboxamides as potent and selective
586   anaplastic lymphoma kinase inhibitors. *Journal of medicinal chemistry* **2012**, 55, 1698-1705.

587   23.   Zuccotto, F.; Ardini, E.; Casale, E.; Angiolini, M., Through the "gatekeeper door": exploiting the
588   active kinase conformation. *Journal of medicinal chemistry* **2009**, 53, 2681-2694.

589   24.   Epstein, L. F.; Chen, H.; Emkey, R.; Whittington, D. A., The R1275Q neuroblastoma mutant and
590   certain ATP-competitive inhibitors stabilize alternative activation loop conformations of
591   anaplastic lymphoma kinase. *Journal of Biological Chemistry* **2012**, 287, 37447-37457.

592   25.   Frączek, T.; Siwek, A.; Paneth, P., Assessing molecular docking tools for relative biological
593   activity prediction: a case study of triazole HIV-1 NNRTIs. *Journal of chemical information and*
594   *modeling* **2013**, 53, 3326-3342.

595   26.   Users'Manual, D. S., Version 2.5. 5. *Accelrys Inc., San Diego, CA* **2010**.

596   27.   Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the performance of the MM/PBSA and
597   MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular
598   dynamics simulations. *Journal of chemical information and modeling* **2011**, 51, 69-82.

599   28.   Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L., Evaluation and
600   reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum
601   chemical calculations on peptides. *The Journal of Physical Chemistry B* **2001**, 105, 6474-6487.

602   29.   Schrödinger, L., Maestro, Version 9.0. *New York, NY, USA* **2009**.

603   30.   Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.;
604   Barone, V.; Mennucci, B.; Petersson, G., Gaussian 09, Gaussian. *Inc., Wallingford, CT* **2009**.

605   31.   Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., A well-behaved electrostatic potential
606   based method using charge restraints for deriving atomic charges: the RESP model. *The Journal*

607   *of Physical Chemistry* **1993**, 97, 10269-10280.

608   32.   Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of
609   a general amber force field. *Journal of computational chemistry* **2004**, 25, 1157-1174.

610   33.   Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An W log (N) method for Ewald sums
611   in large systems. *The Journal of chemical physics* **1993**, 98, 10089-10092.

612   34.   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of
613   simple potential functions for simulating liquid water. *The Journal of chemical physics* **1983**, 79,
614   926-935.

615   35.   Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.;
616   Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *Journal*
617   *of computational chemistry* **2005**, 26, 1668-1688.

618   36.   Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of
619   multiple Amber force fields and development of improved protein backbone parameters.
620   *Proteins: Structure, Function, and Bioinformatics* **2006**, 65, 712-725.

621   37.   Lill, M. A., Efficient incorporation of protein flexibility and dynamics into molecular docking
622   simulations. *Biochemistry* **2011**, 50, 6157-6169.

623   38.   Mangoni, M.; Roccatano, D.; Di Nola, A., Docking of flexible ligands to flexible receptors in
624   solution by molecular dynamics simulation. *Proteins: Structure, Function, and Bioinformatics*
625   **1999**, 35, 153-162.

626   39.   Hartigan, J. A.; Wong, M. A., Algorithm AS 136: A k-means clustering algorithm. *Applied*
627   *statistics* **1979**, 100-108.

628   40.   Tian, S.; Sun, H. Y.; Li, Y. Y.; Pan, P. C.; Li, D.; Hou, T. J., Development and Evaluation of an
629   Integrated Virtual Screening Strategy by Combining Molecular Docking and Pharmacophore
630   Searching Based on Multiple Protein Structures. *Journal Of Chemical Information And Modeling*
631   **2013**, 53, 2743-2756.

632   41.   Zhou, S.; Li, Y.; Hou, T., Feasibility of using molecular docking-based virtual screening for
633   searching dual target kinase inhibitors. *Journal of chemical information and modeling* **2013**, 53,
634   982-996.

635   42.   Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J.
636   L., Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in
637   database screening. *Journal of Medicinal Chemistry* **2004**, 47, 1750-1759.

638   43.   Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the Performance of the Molecular
639   Mechanics/Poisson Boltzmann Surface Area and Molecular Mechanics/Generalized Born Surface
640   Area Methods. II. The Accuracy of Ranking Poses Generated From Docking. *J Comput Chem.*
641   **2011**, 32, 866-877.

642   44.   Sun, H.; Li, Y.; Tian, S.; Xu, L.; Hou, T., Assessing the Performance of MM/PBSA and
643   MM/GBSA Methods. 4. Accuracies of MM/PBSA and MM/GBSA Methodologies Evaluated by
644   Various Simulation Protocols using PDBbind Data Set. *Phys. Chem. Chem. Phys.* **2014**, 16,
645   16719-16729.

646   45.   Wang, J.; Hou, T.; Xu, X., Recent Advances in Free Energy Calculations with a Combination of
647   Molecular Mechanics and Continuum Models. *Curr. Comput.-Aided Drug Des.* **2006**, 2, 287-306.

648   46.   Xu, L.; Sun, H.; Li, Y.; Wang, J.; Hou, T., Assessing the Performance of MM/PBSA and
649   MM/GBSA Methods. 3. The Impact of Force Fields and Ligand Charge Models. *J. Phys. Chem. B*
650   **2013**, 117, 8408-8421.

651   47. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan,
652   Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., Calculating
653   structures and free energies of complex molecules: Combining molecular mechanics and
654   continuum models. *Acc. Chem. Res.* **2000**, 33, 889-897.

655   48. Sun, H. Y.; Li, Y. Y.; Tian, S.; Wang, J. M.; Hou, T. J., P-loop Conformation Governed
656   Crizotinib Resistance in G2032R-Mutated ROS1 Tyrosine Kinase: Clues from Free Energy
657   Landscape. *PLoS Comput. Biol.* **2014**, 10, 13.

658   49. Sun, H.; Li, Y.; Shen, M.; Tian, S.; Xu, L.; Pan, P.; Guan, Y.; Hou, T., Assessing the performance
659   of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute
660   dielectric constant MM/GBSA and MM/PBSA rescoring. *Physical Chemistry Chemical Physics*
661   **2014**, 16, 22035-22045.

662   50. Li, L.; Li, Y. Y.; Zhang, L. L.; Hou, T. J., Theoretical Studies on the Susceptibility of Oseltamivir
663   against Variants of 2009 A/H1N1 Influenza Neuraminidase. *Journal Of Chemical Information*
664   *And Modeling* **2012**, 52, 2715-2729.

665   51. Xue, W. W.; Pan, D. B.; Yang, Y.; Liu, H. X.; Yao, X. J., Molecular modeling study on the
666   resistance mechanism of HCV NS3/4A serine protease mutants R155K, A156V and D168A to
667   TMC435. *Antiviral Res.* **2012**, 93, 126-137.

668   52. Yang, Y.; Qin, J.; Liu, H. X.; Yao, X. J., Molecular Dynamics Simulation, Free Energy
669   Calcuiation and Structure-Based 3D-QSAR Studies of B-RAF Kinase Inhibitors. *J. Chem Inf.*
670   *Model.* **2011**, 51, 680-692.

671   53. Homeyer, N.; Gohlke, H., Free Energy Calculations by the Molecular Mechanics
672   Poisson-Boltzmann Surface Area Method. *Mol. Inf.* **2012**, 31, 114-122.

673   54. Liu, H.; Yao, X., Molecular Basis of the Interaction for an Essential Subunit PA-PB1 in Influenza
674   Virus RNA Polymerase: Insights from Molecular Dynamics Simulation and Free Energy
675   Calculation. *Mol. Pharmaceut.* **2009**, 7, 75-85.

676   55. Onufriev, A.; Bashford, D.; Case, D. A., Exploring protein native states and large‐scale
677   conformational changes with a modified generalized born model. *Proteins: Structure, Function,*
678   *and Bioinformatics* **2004**, 55, 383-394.

679   56. Rocchia, W.; Alexov, E.; Honig, B., Extending the applicability of the nonlinear
680   Poisson-Boltzmann equation: Multiple dielectric constants and multivalent ions. *The Journal of*
681   *Physical Chemistry B* **2001**, 105, 6507-6514.

682   57. Weiser, J.; Shenkin, P. S.; Still, W. C., Approximate atomic surfaces from linear combinations of
683   pairwise overlaps (LCPO). *Journal of computational chemistry* **1999**, 20, 217-230.

684   58. Gohlke, H.; Kiel, C.; Case, D. A., Insights into protein–protein binding by binding free energy
685   calculation and free energy decomposition for the Ras–Raf and Ras–RalGDS complexes. *Journal*
686   *of molecular biology* **2003**, 330, 891-913.

687   59. Hou, T.; Li, N.; Li, Y.; Wang, W., Characterization of Domain-peptide Interaction Interface:
688   Prediction of SH3 Domain-Mediated Protein-protein Interaction Network in Yeast by Generic
689   Structure-Based Models. *J. Proteome Res.* **2012**, 11, 2982.

690   60. Hou, T.; Zhang, W.; Case, D. A.; Wang, W., Characterization of domain-peptide interaction
691   interface: A case study on the amphiphysin-1 SH3 domain. *J. Mol. Biol.* **2008**, 376, 1201-1214.

692   61. Craig, I. R.; Essex, J. W.; Spiegel, K., Ensemble docking into multiple crystallographically
693   derived protein structures: an evaluation based on the statistical analysis of enrichments. *Journal*
694   *of chemical information and modeling* **2010**, 50, 511-524.

695   62. Huang, S. Y.; Zou, X., Ensemble docking of multiple protein structures: considering protein
696   structural variations in molecular docking. *Proteins: Structure, Function, and Bioinformatics*
697   **2007**, 66, 399-421.
698   63. Slynko, I.; Scharfe, M.; Rumpf, T.; Eib, J.; Metzger, E.; Schule, R.; Jung, M.; Sippl, W., Virtual
699   screening of PRK1 inhibitors: ensemble docking, rescoring using binding free energy calculation
700   and QSAR model development. *Journal of chemical information and modeling* **2013**.
701   64. Whalen, K. L.; Chang, K. M.; Spies, M. A., Hybrid Steered Molecular Dynamics‐Docking: An
702   Efficient Solution to the Problem of Ranking Inhibitor Affinities Against a Flexible Drug Target.
703   *Molecular informatics* **2011**, 30, 459-471.
704   65. Floquet, N.; Marechal, J.-D.; Badet-Denisot, M.-A.; Robert, C. H.; Dauchez, M.; Perahia, D.,
705   Normal mode analysis as a prerequisite for drug design: application to matrix metalloproteinases
706   inhibitors. *FEBS letters* **2006**, 580, 5130-5136.
707   66. Ding, F.; Yin, S.; Dokholyan, N. V., Rapid flexible docking using a stochastic rotamer library of
708   ligands. *Journal of chemical information and modeling* **2010**, 50, 1623-1632.
709   67. Novoa, E. M.; Pouplana, L. R. d.; Barril, X.; Orozco, M., Ensemble docking from homology
710   models. *Journal of Chemical Theory and Computation* **2010**, 6, 2547-2557.
711   68. Totrov, M.; Abagyan, R., Flexible ligand docking to multiple receptor conformations: a practical
712   alternative. *Current opinion in structural biology* **2008**, 18, 178-184.
713   69. Carlson, H. A., Protein flexibility and drug design: how to hit a moving target. *Current opinion in*
714   *chemical biology* **2002**, 6, 447-452.

715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738

739　**Legend of the Figures**

740　**Figure 1**. Ribbon diagram of human ALK (right-hand of the illustration). The key

741　residues in the binding pocket of ALK and the interactions between piperidine

742　carboxamides 2 (11j) with ALK are highlighted.

743　**Figure 2**. Correlation between the experimental bioactivities (pIC50) and the docking

744　scored predicted by RRD, IFD or QPLD.

745　**Figure 3**. RMSD Map for the 200 conformations extracted from the MD trajectory

746　based on the crystal structure of 4FNZ by using the *k*-means clustering algorithm.

747　**Figure 4**. Correlation between the experimental pIC_{50} and (a) the average docking

748　scores or (b) the highest docking scores predicted by ensemble docking.

749　**Figure 5**. RMSDs of the backbone C_{α} atoms of the selected inhibitor/ALK complexes

750　(1, 7a, 11a, 11d, 11k, 11q, and 11s) as a function of simulation time.

751　**Figure 6**. RMSF of each residue of the selected complexes (1, 7a, 11k, 11q, and 11s)

752　obtained from the last 3ns MD simulations.

753　**Figure 7**. Correlation between the experimental pIC50 and the binding free energies

754　calculated by (a) MM/GBSA or (b) MM/PBSA.

755　**Figure 8**. Comparison of the averaged structures for (a) compounds 1 and 7a, (b)

756　compounds 1 and 11k, (c) compounds 1 and 11q, and (d) compounds 1 and 11s.

757　**Figure 9**. (a) Comparison of the inhibitor-residue interaction spectra for compounds 1

758　and 7a; Comparison of the (b) nonpolar (van der Waals and nonpolar solvation

759　energy) and (c) polar (electrostatic and polar solvation energy) interactions between

760　compounds 7a and 1 for the key residues in the active site.

761　**Figure 10**. Comparison of the inhibitor-residue interaction spectra for compounds 1

762　and the right-hand modified compounds (a) 11k, (b) 11q, and (c) 11s; Comparison of

763　the (d) nonpolar (van der Waals and nonpolar solvation energy) and (e) polar

764　(electrostatic and polar solvation energy) interactions between the right-hand

765　modified compounds (11k, 11q, and 11s) and 1 for the key residues in the active site.

766
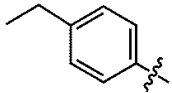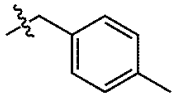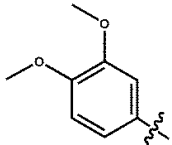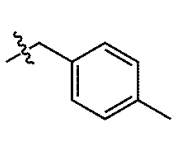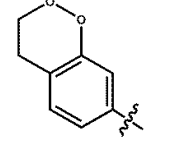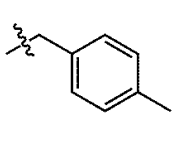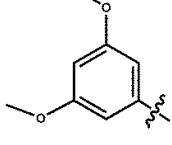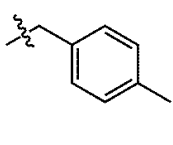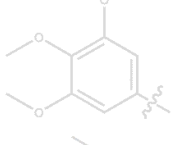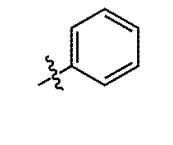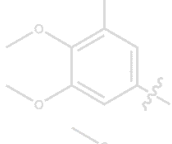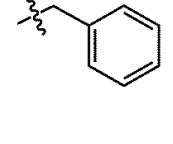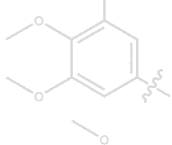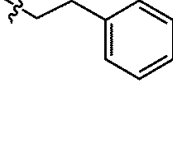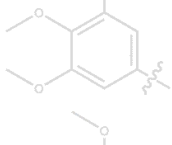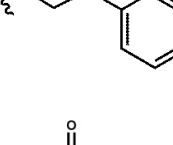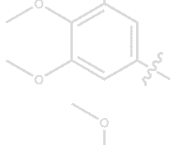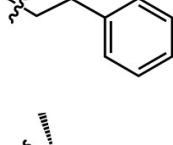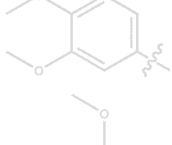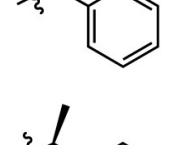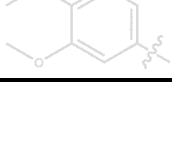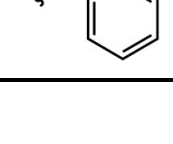
767
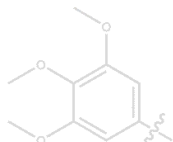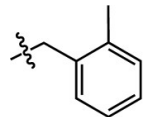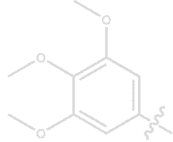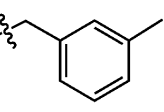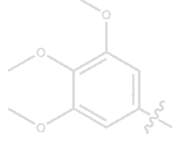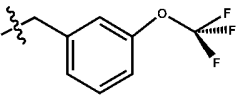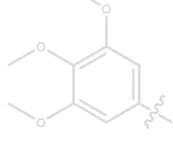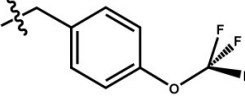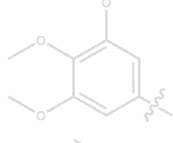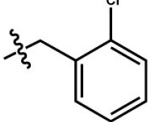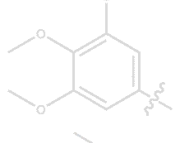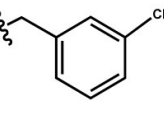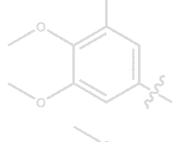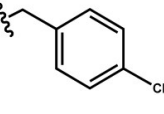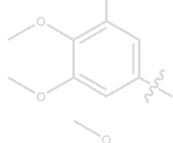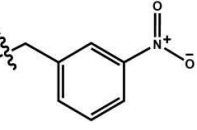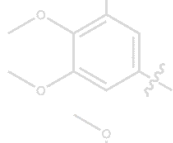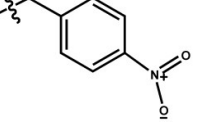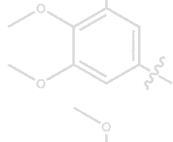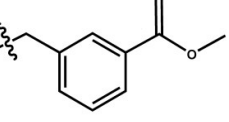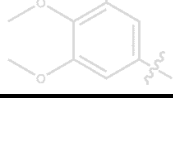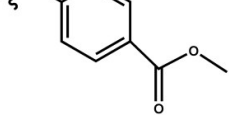
768   **Table 1**. Structures, biological activities and predicted scores of the studied ALK
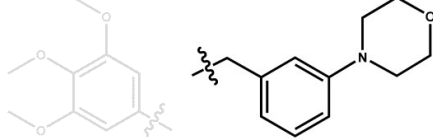
769   inhibitors.
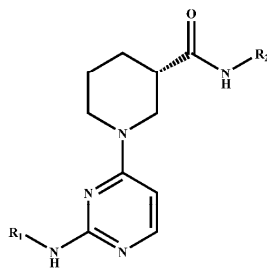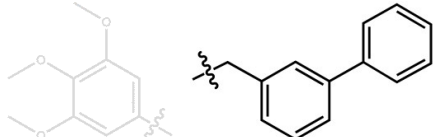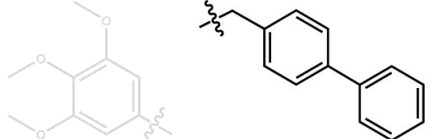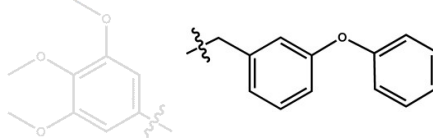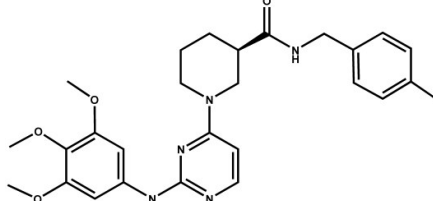
770

771

No.	R$_1$	R$_2$	IC$_{50}$ (μM)	pIC$_{50}$	RRD	IFD	QPLD
1			0.17	6.76	-11.10	-10.93	-12.75
7a			25	4.60	-11.96	-11.18	-12.74
7b			2.32	5.63	-8.63	-9.67	-7.81
7c			25	4.60	-11.57	-11.81	-12.81
7d			2.01	5.70	-11.98	-13.75	-13.17
7e			6.30	5.20	-12.50	-12.34	-13.53
7f			1.02	5.99	-11.93	-13.76	-12.93
7g			3.10	5.51	-11.51	-10.25	-12.48
7h			0.91	6.04	-11.63	-12.45	-12.77

25

7i			25	4.60	-11.79	-13.27	-13.26
7j			1.73	5.76	-12.32	-12.37	-13.25
7k			1.48	5.83	-11.53	-14.32	-14.11
7l			0.33	6.49	-12.07	-11.71	-13.09
11a			0.83	6.08	-11.38	-13.11	-13.14
11b			0.36	6.44	-12.76	-11.44	-13.28
11c			0.36	6.45	-12.14	-10.83	-12.87
11d			2.94	5.17	-10.66	-9.88	-12.49
11e			2.60	5.53	-10.97	-8.61	-11.90
11f			2.01	5.59	-11.41	-13.11	-13.72
11g			2.01	5.70	-11.09	-10.03	-12.43

11h		0.34	6.47	-13.57	-14.27	-15.02
11i		0.08	7.08	-10.68	-13.67	-10.27
11j		0.02	7.80	-9.62	-12.76	-5.66
11k		0.01	8.00	-14.11	-13.40	-13.66
11l		0.59	6.23	-11.06	-10.32	-12.17
11m		0.08	7.10	-11.18	-11.21	-12.91
11n		0.10	7.02	-11.02	-12.95	-12.56
11o		0.07	7.15	-11.68	-13.24	-12.61
11p		2.93	5.53	-10.85	-10.31	-12.31
11q		0.06	7.22	-11.55	-11.33	-12.65
11r		1.70	5.77	-11.28	-11.26	-12.80

11s		0.02	7.72	-11.70	-11.33	-12.80
11t		0.14	6.86	-11.14	-9.52	-12.23
11u		0.19	6.72	-11.89	-11.40	-12.88
11v		0.03	7.51	-11.71	-11.81	-12.68
14		4.12	5.39	-12.21	-11.87	-13.12

772

773

774

775

776

777

778

779

780

781

782

783

784 **Table 2**. Calculated binding free energies and individual energy components predicted

785 by MM/GBSA (kcal/mol)

System	ΔE_{ele}	ΔG_{GB}	ΔE_{vdW}	ΔG_{SA}	$\Delta G_{polar}{}^{a}$	$\Delta G_{nonpolar}{}^{b}$	ΔG_{bind}	pIC_{50}
1/ALK	-28.92±1.85	45.58±1.20	-61.60±0.04	-5.23±0.01	16.66±0.65	-66.83±0.03	-50.17±0.62	6.76
7a/ALK	-28.42±1.48	40.92±0.73	-50.47±0.59	-4.50±0.01	12.50±0.74	-54.97±0.58	-42.48±0.16	4.60
11k/ALK	-31.69±1.46	48.18±1.18	-66.40±0.05	-5.83±0.05	16.49±0.27	-72.23±0.00	-55.74±0.27	8.00
11q/ALK	-35.23±1.21	52.71±1.44	-68.93±0.62	-5.82±0.04	17.48±0.24	-74.75±0.66	-57.26±0.43	7.22
11s/ALK	-26.65±1.76	47.07±1.02	-68.00±0.73	-5.84±0.03	20.42±0.74	-73.84±0.76	-53.44±0.02	7.72

786 $^{a}\Delta G_{polar}=\Delta E_{ele}+\Delta G_{GB}$; $^{b}\Delta G_{nonpolar}=\Delta E_{vdW}+\Delta G_{SA}$

787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817

29

818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834

835

836   **Figure 1**

837
838
839
840
841
842
843
844
845
846
847
848
849
850

851

852

853 **Figure 2**

854
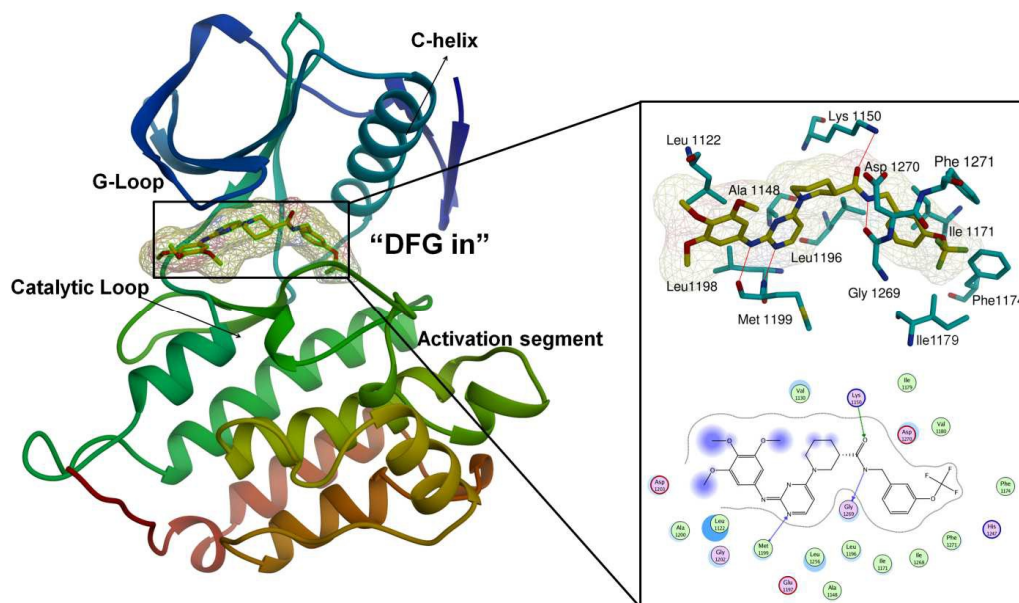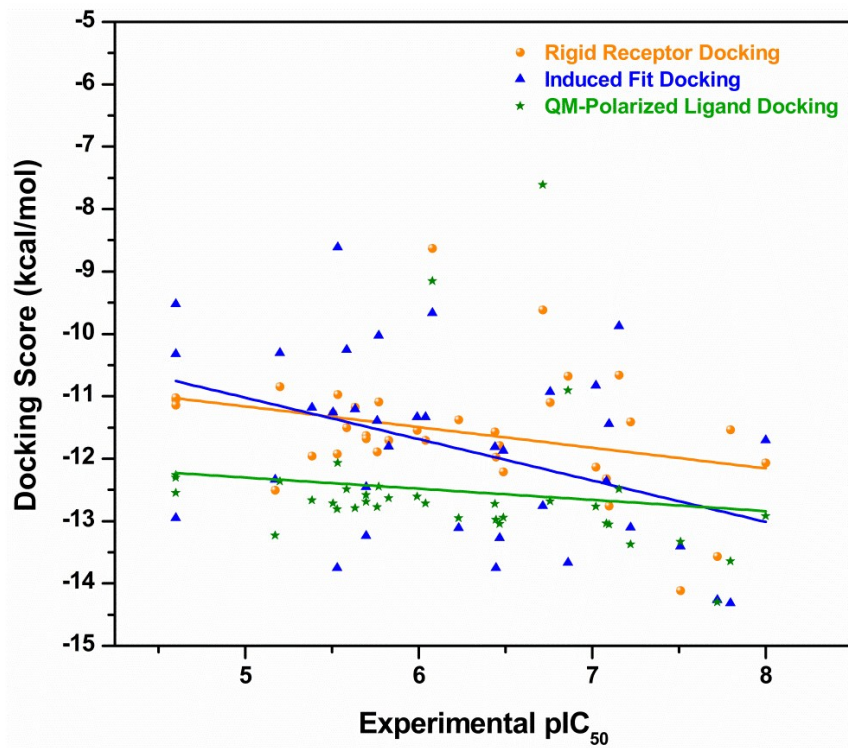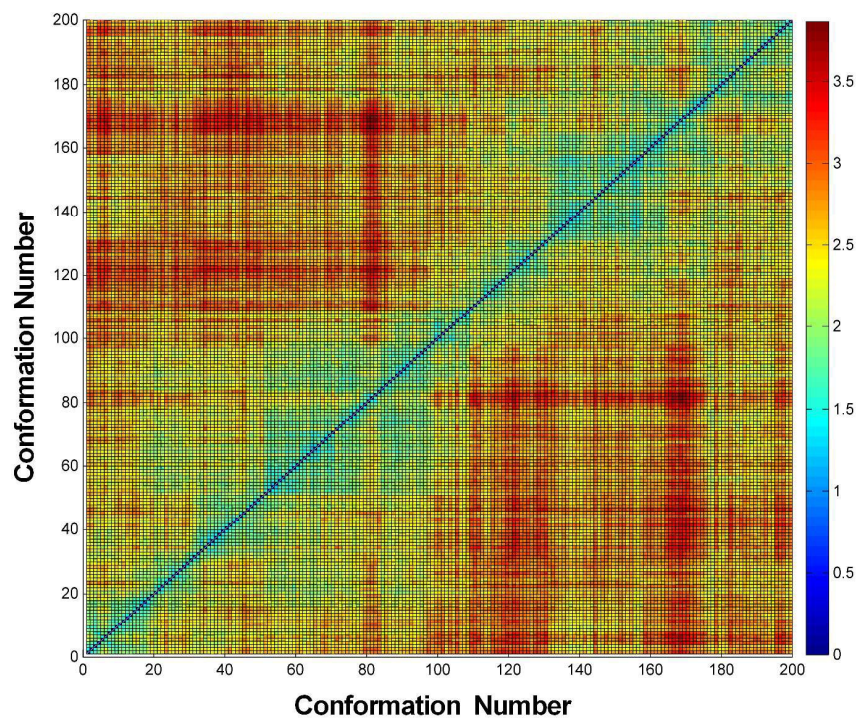
855

856

857 **Figure 3**
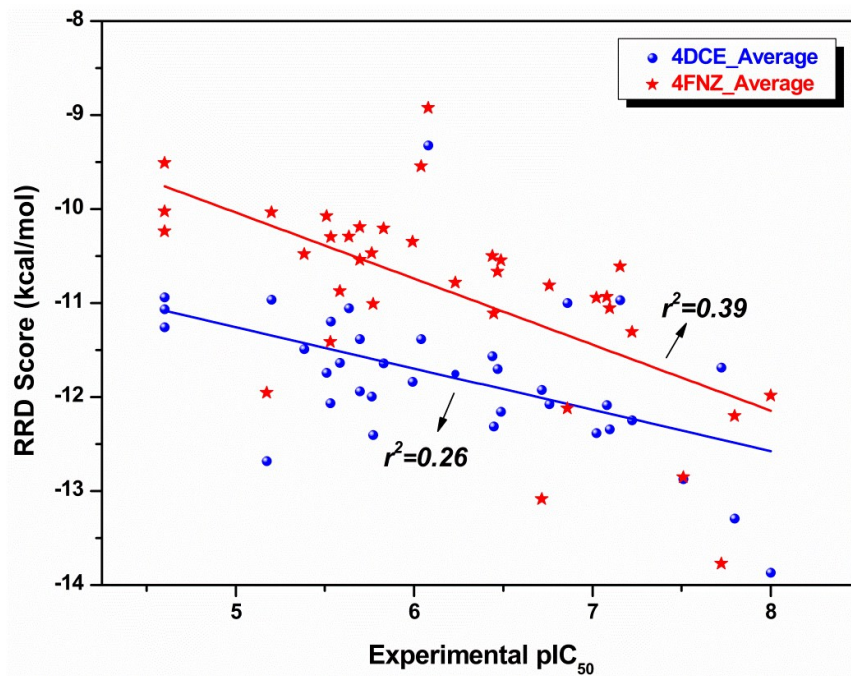
(a)
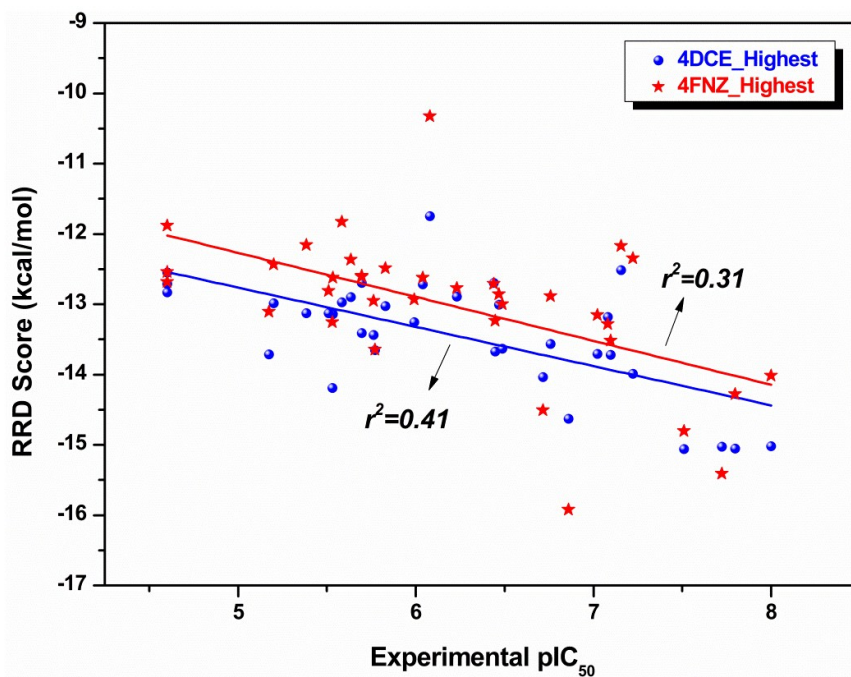
(b)

Figure 4
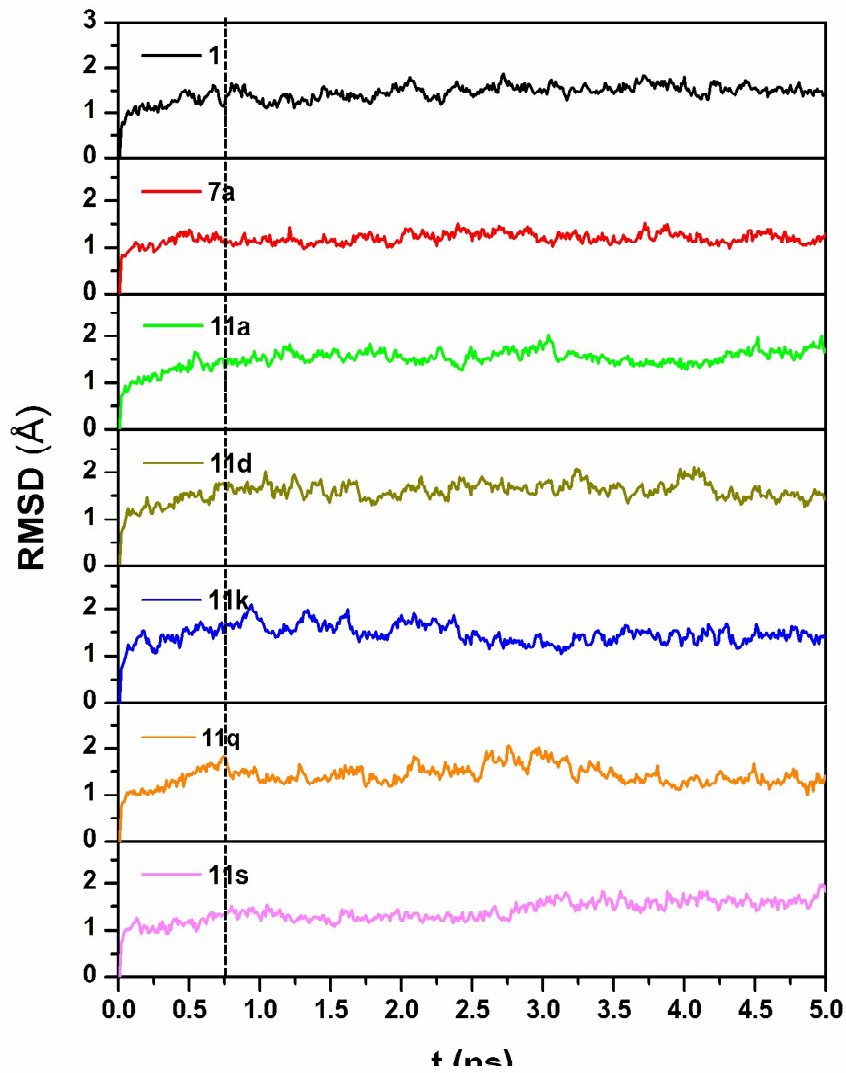
Figure 5

938

939   **Figure 6**
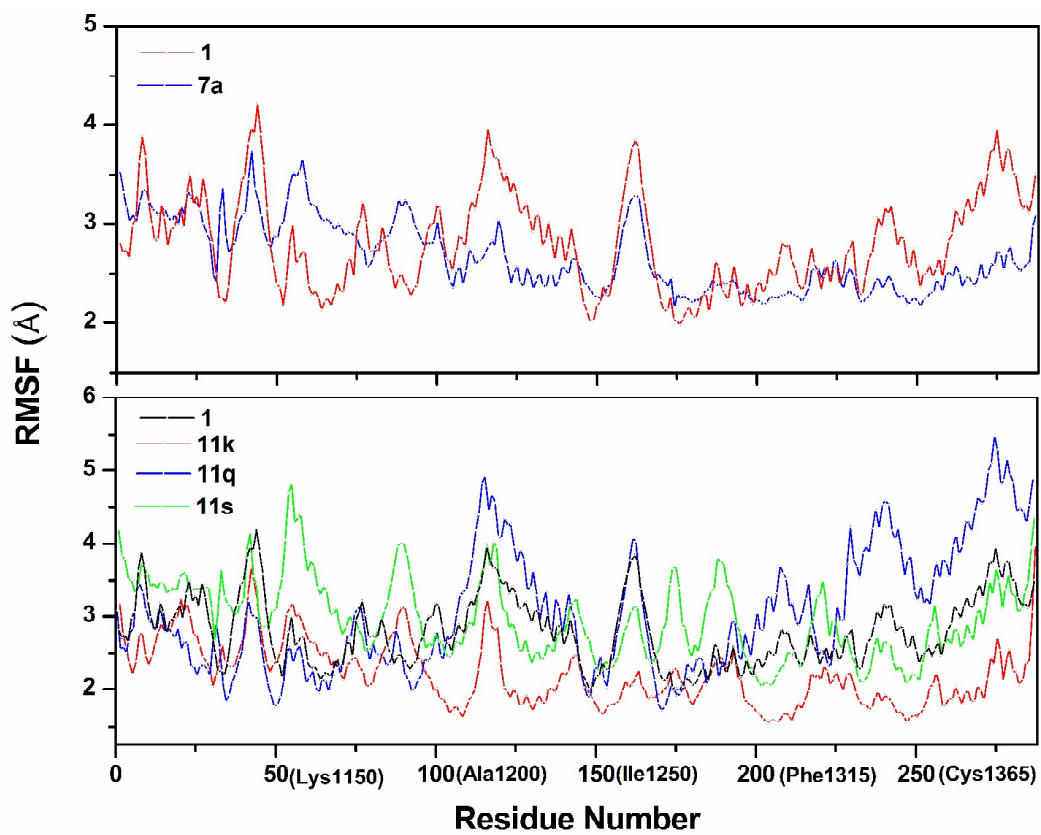
940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

(a)

(b)

Figure 7
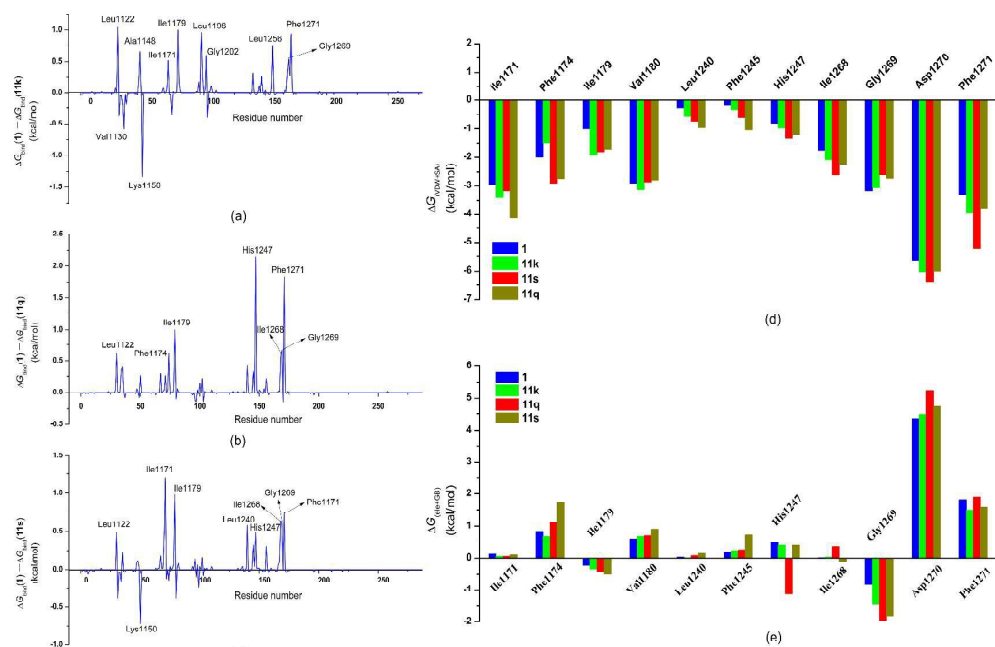
(a)

(b)

(c)

(d)

Figure 8

(a)

(b)   (c)

Figure 9

1062

1063

1064

1065

Figure 10