PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2018, **20**, 4851

Importance of protein flexibility in molecular recognition: a case study on Type-I1/2 inhibitors of ALK†

Xiaotian Kong,[ab] Huiyong Sun,[a] Peichen Pan,[a] Feng Zhu,[a] Shan Chang,[c] Lei Xu,[c] Youyong Li[b] and Tingjun Hou*[ab]

Anaplastic lymphoma kinase (ALK) has been regarded as a promising target for the therapy of various cancers. A large number of ALK inhibitors with diverse scaffolds have been discovered, and most of them belong to Type-I inhibitors that only occupy the ATP-binding pocket. Recently, we reported a series of novel and potent Type-I1/2 inhibitors of ALK with the 1-purine-3-piperidinecarboxamide scaffold, which can bind to both the ATP-binding site of ALK and the adjacent hydrophobic allosteric pocket. In this study, the binding mechanisms of these Type-I1/2 ALK inhibitors were elucidated by multiple molecular modeling techniques. The calculation results demonstrate that the ensemble docking based on multiple protein structures and the MM/PB(GB)SA calculations based on molecular dynamics (MD) simulations yield better predictions than conventional rigid receptor docking (Glide, Surflex-Dock, and Autodock Vina), highlighting the importance of incorporating receptor flexibility in the predictions of binding poses and binding affinities of Type-I1/2 ALK inhibitors. Furthermore, the umbrella sampling (US) simulations and MM/GBSA binding free energy decomposition analyses indicate that Leu1122, Leu1198, Gly1202 and Glu1210 in the hinge region and Glu1197, Ile1171, Phe1174, Ile1179, His1247, Ile1268, Asp1270 and Phe1271 in the allosteric pocket of ALK are the key residues for determining the relative binding strength of the studied inhibitors. Besides, we found that the most potent inhibitor (**001-017**) tends to form stronger transient interactions with residues along the dissociation channel due to the high electronegativity of its bulky 4-(trifluoromethoxy) phenylamine tail. As a whole, both the stronger binding affinity and the higher energetic barrier (which may prolong the drug-target residence time) of **001-017** contribute to its excellent anti-proliferation activity against ALK-positive cancer cells.

Received 8th December 2017,
Accepted 16th January 2018

DOI: 10.1039/c7cp08241j

rsc.li/pccp

Introduction

Anaplastic lymphoma kinase (ALK), a receptor tyrosine kinase (RTK) of the insulin receptor (IR) superfamily,[1] is recognized as an important therapeutic target for the treatment of non-small cell lung cancer (NSCLC) with EML4-ALK fusion proteins occurring in nearly 3–7% of NSCLC patients.[2] As the first-generation ALK inhibitor, Crizotinib was approved by the Food and Drug Administration (FDA) in 2011 for the treatment of advanced NSCLC harboring ALK rearrangements.[3,4] Unfortunately, the success of Crizotinib in ALK$^+$ NSCLC was rapidly overshadowed by the emergence of drug resistance caused by the point mutations in the kinase domain, such as L1196M, C1156Y and G1269A/S in both full length ALK and EML4-ALK, as well as the later identified L1152R, I1171T/N/S, F1174V/L/C, G1202R, D1203N, S1206Y, V1180L and 1151T-ins.[5–8] Then, two second-generation inhibitors, including Ceritinib and Alectinib, were approved by the FDA in 2014 and 2015, respectively, for the treatment of ALK$^+$ metastatic NSCLC patients who experience

[a] *College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, P. R. China. E-mail: tingjunhou@zju.edu.cn; Tel: +86-517-88208412*
[b] *Institute of Functional Nano and Soft Materials (FUNSOM), Soochow University, Suzhou, Jiangsu 215123, P. R. China*
[c] *Institute of Bioinformatics and Medical Engineering, School of Electrical and Information Engineering, Jiangsu University of Technology, Changzhou 213001, P. R. China*

† Electronic supplementary information (ESI) available: Table S1: Compound structures and their corresponding biological activities; Table S2: Binding free energies and their average values for the five representative ALK inhibitors determined by umbrella sampling and MM/GBSA (kcal mol^{-1}); Table S3: Energetic contributions of the key residues to the binding of the representative ALK inhibitors (kcal mol^{-1}); Fig. S1: Anti-proliferation activity against NCI-H2228, NCI-H3122, and Karpas-299 cell lines of the studied Type-I1/2 inhibitors of ALK; Fig. S2: RMSDs (root mean square deviations) of the representative systems; Fig. S3: The RCs through the ATP channels of the (A) ALK/001-007, (B) ALK/001-013, (C) ALK/**001–017**, (D) ALK/002-012, and (E) ALK/002-016 systems; Fig. S4: RMSFs (root mean square fluctuations) of the representative systems; Fig. S5: The correlation between the experimental binding affinities and the electrostatic (ΔE_{ele}), van der Waals (ΔE_{vdW}), solvation (ΔG_{GB}), and SASA (ΔG_{SA}) components; Fig. S6: Convergence of the PMFs of the representative systems along the ATP channel in US simulations. See DOI: 10.1039/c7cp08241j

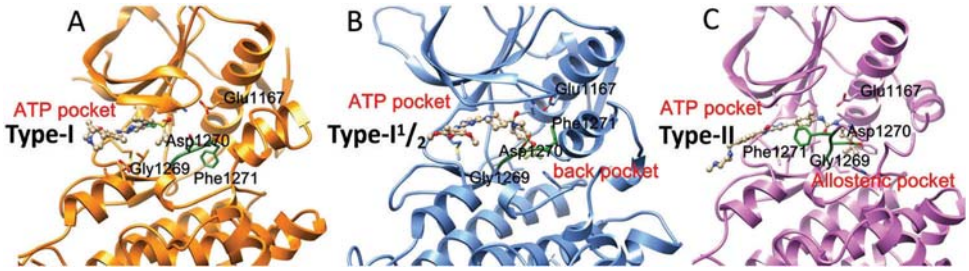

Fig. 1 The binding modes of (A) Type-I (PDB entry: 4MCK), (B) Type-I1/2 (PDB entry: 4FNZ), and (C) Type-II (PDB entry: 5IUG) inhibitors of ALK.

disease progression or who are resistant to Crizotinib.[9] On April 28, 2017, the FDA granted the accelerated approval to Brigatinib (Alunbrig), which is also classified as a second-generation inhibitor. In addition, a number of ALK inhibitors have been pushed into clinical trials, including Entrectinib (phase II), Lorlatinib (phase II), X-396 (phase I/II), ASP3026 (phase I), and CEP-37440 (phase I).[9]

All of the ALK drugs and drug candidates in ongoing clinical trials are Type-I inhibitors that bind with the ATP-binding pocket in the active state (referred to as the DFG-in state), and they are quite sensitive to the resistant mutations surrounding the ATP pocket (*e.g.* L1196M, G1202R, G1269A, and S1206Y). Recently, other kinds of ALK inhibitors, called Type-I1/2 inhibitors, have been reported, and they can not only bind at the ATP-binding site as Type-I inhibitors (Fig. 1A) but can also extend into the back hydrophobic pocket adjacent to the ATP-binding site, which can induce a unique DFG-shifted loop conformation (Fig. 1B).[10] It has been reported that Type-I1/2 and Type-II inhibitors are more selective and exhibit better capability to combat drug resistance than Type-I inhibitors.[11–13] Moreover, they generally exhibit lower dissociation rate constants (k_{off}) and prolonged residence time than Type-I inhibitors, which might enhance their efficacy both *in vitro* and *in vivo*.[14,15] To date, only several Type-II or Type-I1/2 inhibitors of ALK have been reported, and their inhibitory activity is generally weak.[16,17]

In our previous study, a series of Type-I1/2 inhibitors of ALK with the 1-purine-3-piperidinecarboxamide scaffold have been reported, and several of them, such as compound **001-017**, even exhibit unparallel inhibitory activity against ALK (**001-017**: IC_{50} = 0.27 nM) and better anti-proliferation activity against NCI-H2228, NCI-H3122, and Karpas-299 cell lines than two clinically used drugs, Crizotinib and Ceritinib (Table S1 and Fig. S1 in the ESI†).[10] It is noteworthy that compound **001-017** can also effectively inhibit several resistant ALK mutants (*e.g.* L1196M, C1156Y, R1275Q and F1174L), which are intolerant to Crizotinib and Ceritinib. Compound **001-017**, to our knowledge, is the most potent Type-I1/2 inhibitor of ALK reported so far. In this study, in order to gain reliable structural information to guide the optimization of these Type-I1/2 inhibitors, multiple molecular modeling techniques were used to elucidate their binding mechanisms. First, conventional docking calculations were carried out to predict the protein–ligand binding structures and binding affinities. Unfortunately, conventional docking did not afford satisfactory predictions due to the ignorance of protein flexibility and essential water molecules. Then, in order to incorporate the conformational changes of the target into the ligand-binding process, the ensemble docking and MM/GBSA calculations based on molecular dynamics (MD) simulations were subsequently utilized to mimic the ''conformational selection'' and ''induced-fit'' phenomena in protein–ligand recognition. Moreover, the enhanced sampling technique, umbrella sampling (US), was used to reveal the possible obstacles and transient conformations along the unbinding pathways of the ligands, and the essential residues surrounding the binding pocket were quantitatively highlighted by per-residue free energy decomposition. It is expected that our results can guide further design of novel Type-I1/2 ALK inhibitors with improved binding affinity and pharmacokinetics.

Materials and methods

Generation of a representative conformational ensemble

The crystal structure of piperidine carboxamide inhibitor 2 in complex with ALK (PDB entry: 4FNZ,[18] Fig. 2A) was used as the initial structure for MD simulations. The missing regions, including the G-loop and A-loop, were reconstructed using the loop module in SYBYL-X2.1 package (Tripos Associates, St Louis, MO, USA).

The protein and the ligand were parameterized using the Amber ff14SB force field[19] and the general Amber force field (gaff),[20] respectively. The electrostatic potential of the ligand was calculated at the Hartree–Fock (HF) SCF/6-31* level of theory using the Gaussian 09 package[21] and the atomic charges were subsequently fitted by using the restrained electrostatic potential (RESP) technology[22] in the antechamber module[23] of Amber14.[24] Then, the complex was immersed into a cubic TIP3P water box[25] extended 10 Å from any solute atom, and 4 Na^+ were added to neutralize the unbalanced charge of the system. For both molecular mechanics (MM) minimization and MD simulations, the long-range electrostatics was handled using the particle mesh Ewald (PME) algorithm,[26] and the non-bonded cutoff for the real-space interactions was set to 10 Å.[27] Prior to MD simulations, the system was optimized by the following four steps: (1) 5000 steps of minimization (1000 cycles of steepest descent and 4000 cycles of conjugate gradient) with all the heavy atoms restrained at 5 kcal mol^{-1} $Å^{-2}$, (2) 5000 steps of minimization (1000 cycles of steepest descent and

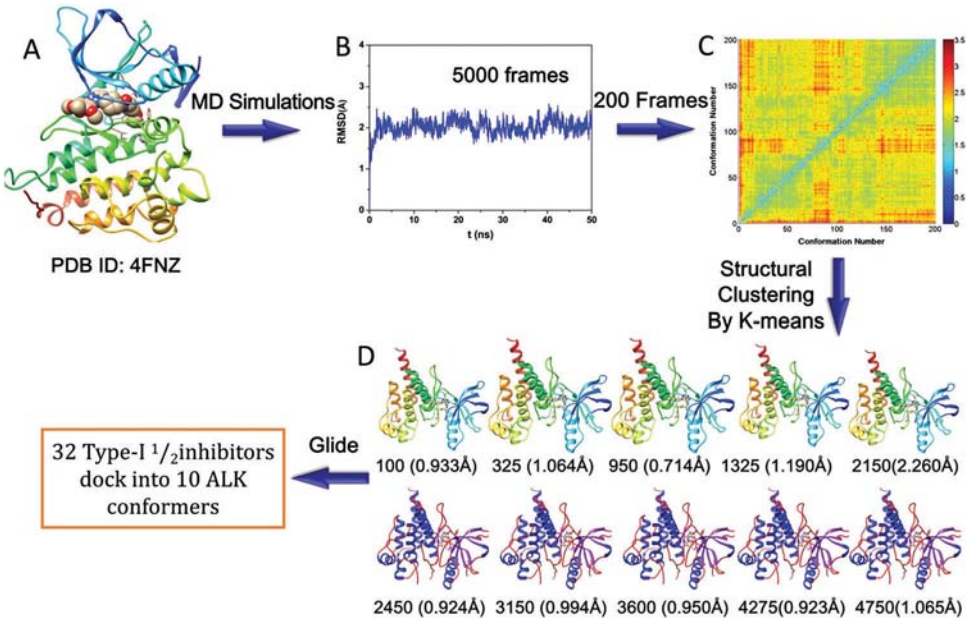

Fig. 2 Generation of representative conformational ensembles.

4000 cycles of conjugate gradient) with the protein and the ligand restrained at 5 kcal mol^{-1} $Å^{-2}$, (3) 5000 steps of minimization (1000 cycles of steepest descent and 4000 cycles of conjugate gradient) with the backbone heavy atoms of the protein restrained at 5 kcal mol^{-1} $Å^{-2}$, (4) 10 000 steps of minimization (5000 cycles of steepest descent and 5000 cycles of conjugate gradient) without any restrain.

Subsequently, MD simulations were performed for the system. The SHAKE algorithm[28] was used to restrain the covalent bonds between heavy atoms and hydrogen atoms. The time step was set to 2.0 fs. The system was heated from 0 to 310 K within 50 ps under the restraint of 2 kcal mol^{-1} $Å^{-2}$. Then, the system was equilibrated for another 50 ps in an *NPT* ensemble (T = 310 K and P = 1 atm). Finally, a 50 ns run was carried out in the *NPT* ensemble using the *pmemd.cuda* module in Amber14. The coordinates were saved with an interval of 10 ps.

Eventually, 200 conformations were evenly extracted from the 50 ns MD trajectory (Fig. 2B), and 10 representative structures for ensemble docking were chosen using the *k*-means clustering algorithm[29] based on the pairwise root-mean-square displacements (RMSDs) between any two of the extracted structures[30,31] (Fig. 2C and D).

Molecular docking by different protocols

The 32 inhibitors shown in Table S1 (ESI†) reported by our group[10] were designed based on piperidine carboxamides, which were reported as the Type-I1/2 inhibitors of ALK.[16] Then, based on the crystal structure of ALK complexed with piperidine carboxamide 2 (PDB ID: 4FNZ[18]), various molecular docking methods were used to predict the binding structures of the studied inhibitors at the active site of ALK, including Glide docking, Glide docking with Prime-MM/GBSA rescoring (Prime-MM/GBSA-rigid), Autodock Vina docking, Surflex-Dock docking, induced-fit docking (IFD), and induced-fit docking with Prime-MM/GBSA rescoring (Prime-MM/GBSA-flexible).

Glide docking and Prime-MM/GBSA rescoring. The protein and inhibitors were prepared using the *Protein Preparation Wizard* and *ligprep* modules[32] in Schrodinger 2016, respectively. The inhibitors were docked into the active site of ALK by rigid-receptor docking (RRD) using the extra precision (XP) scoring mode implemented in the *Glide* module[33] of Schrodinger (Schrödinger, LLC, New York, NY, USA). The docking box was set to 20 Å × 20 Å × 20 Å with the center located at the center of mass of the co-crystallized ligand of 4FNZ. Then, MM/GBSA in the *Prime* module in Schrodinger was used to rescore the binding poses generated by *Glide*. Because the RRD algorithm implemented in *Glide* was employed, the Prime-MM/GBSA rescoring procedure was termed as "*rigid*" to distinguish it from the following flexible rescoring for the induced-fit docking.

AutoDock Vina docking. The *AutoGrid* and *AutoDock* procedures were used to conduct the grid point energy calculations of the receptor and the binding pose scoring of the ligands, respectively.[34] The binding conformations of the ligands were optimized using the Lamarckian Genetic Algorithm (LGA),[35] where the initial population size for each ligand was set to 2 500 000 and the searching box was set to 24 Å × 22 Å × 20 Å.

Surflex-Dock docking. Surflex-Dock supported using the *Protomol*[36] module of SYBYL-X2.1 was utilized to predict the binding poses of the inhibitors, where all the inhibitors were minimized using the *Powell* module with the standard *Tripos* force field.[37]

Induced-fit docking and Prime-MM/GBSA rescoring. The induced-fit docking (IFD)[38] was also conducted, where the residues within 6 Å away from each inhibitor were relaxed using the *Prime* module in Schrodinger. Then, MM/GBSA in the *Prime* module in Schrodinger was used to rescore the binding poses generated by IFD.

Ensemble docking. In the ensemble docking, by using the 6 docking and rescoring methods, including Glide docking, Glide docking with Prime-MM/GBSA rescoring, Autodock Vina docking, Surflex-Dock docking, IFD, and IFD with Prime-MM/GBSA rescoring, the 32 inhibitors were successively docked into the 10 representative structures generated from the 50 ns MD trajectory, and the best scored pose for each ligand was saved for the further analyses (Fig. 2D).

End-point binding free energy calculation based on the classical MM/GB(PB)SA and water-MM/GBSA

The binding poses predicted by Glide were used as the starting structures for the following 5 ns *NPT* MD simulations. The parameters of the MM minimizations and MD simulations for each system can be found in the section of "*Generation of Representative Conformational Ensemble*". During the MD simulations, the coordinates of each system were saved with an interval of 10 ps.

The classical Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) and Molecular Mechanics/Poisson Boltzmann Surface Area (MM/PBSA) approaches in Amber14 were used to estimate the ligand-target binding free energies (ΔG_{bind}, eqn (1)–(4)).[39–46]

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

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

$$\Delta E_{\text{MM}} = \Delta E_{\text{int}} + \Delta E_{\text{ele}} + \Delta E_{\text{vdW}} \quad (3)$$

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

where ΔE_{int} (intra-molecular interactions, including bond, angle, and dihedral energies) in eqn (3) can be completely canceled because the single trajectory strategy was used for the MM/GBSA and MM/PBSA calculations.[47] Thus, the MM energy (ΔE_{MM}) only contains the van der Waals (ΔE_{vdW}) and electrostatic (ΔE_{ele}) interactions. The non-polar part of the solvation free energy (ΔG_{SA}) was calculated by the solvent-accessible surface area (SASA) through the LCPO algorithm: $\Delta G_{\text{SA}} = \gamma \times \text{SASA} + \beta$,[48,49] where the surface tension constants γ and β were set to 0.0072 and 0, respectively. The polar part of the solvation energy (ΔG_{GB} or ΔG_{PB}) in the MM/GBSA or MM/PBSA calculations was estimated using the GB model proposed by Onufriev *et al.*[48] (GB^{OBC1}, igb = 2) or the PB model optimized by Tan *et al.*[50] (PB^{pbsa}), respectively. The interior and exterior dielectric constants were set to 1 and 80, respectively. The ΔE_{vdW}, ΔE_{ele}, $\Delta G_{\text{GB/PB}}$ and ΔG_{SA} terms were computed based on the last 400 snapshots extracted from the 1–5 ns MD trajectories. However, due to the high computational cost of normal mode analysis (NMA), 32 snapshots evenly extracted from the 1–5 ns MD trajectories (Fig. S2, ESI†) were used to calculate the conformational entropies using the nmode module in Amber14.[24] The occupancy of H-bonds was calculated using the HBonds Plugin in Visual Molecular Dynamics (VMD)[51] software and the cutoffs for the donor–acceptor distance and angle cutoff were set to 3.0 Å and 20 degrees, respectively.

The water-MM/GBSA protocol[52] was employed to explicitly deal with the effects of the essential water molecules that mediate protein–ligand interactions on the MM/GBSA calculations. As shown in eqn (5), the essential water molecules are treated as the component of the protein.

$$\Delta G_{\text{bind}} = G_{\text{com+wat}} - (G_{\text{rec+wat}} + G_{\text{lig}}) \quad (5)$$

Here, the water molecules whose O atoms locate within 3.5 Å of both the protein and the ligand are considered in the water-MM/GBSA calculations. All the other parameters are the same as those mentioned in the classical MM/GBSA calculations.

Umbrella sampling simulations

As widely recognized, many biological phenomena, such as protein–ligand recognition, probably require microseconds or an even longer timescale, which are difficult to be simulated using conventional all-atom MD simulations because of the limited simulation timescale.[53,54] Enhanced sampling methods, such as umbrella sampling (US), have been developed to overcome the limitations of computational resources for many long timescale problems.[55–57] Here, to construct the reaction coordinate (RC) for the US simulations, the direction along the ATP binding pocket determined using Caver 2.0[58] was used as the dissociation channel. In the ALK/001-007 system, the reaction coordinate (RC) along the ATP channel is the distance between the carbon atom (CB) of Ile79 in the receptor and the nitrogen atom (N4) of **001-007** (ALK/**001-013**: carbon atom (CB) of Ile78, carbon atom (C12) of **001-013**; ALK/**001-017**: carbon atom (CB) of Ile78, carbon atom (C9) of **001-017**; ALK/**002-012**: carbon atom (CB) of Ile78, carbon atom (C12) of **002-012**; ALK/**002-016**: carbon atom (CG1) of Ile78, nitrogen atom (N7) of **002-016**; Fig. S3, ESI†). The harmonic potential added in window i was calculated by $u_i = 1/2k_i(r_{\text{current position}} - r_{\text{initial position}})^2$, where k_i represents the elastic constant of the restraint potential (5 kcal mol Å^{-2}). The US simulations were carried out for 41 continuous windows with each 0.5 Å in length, leading to 20 Å away from the original position. Besides, in order to achieve convergence, 5 ns US simulations were executed for each window. Finally, the biased distribution of the samples in each window was reconstructed and combined to a normal one using the weighted histogram analysis method (WHAM),[59] which was used for the calculation of the potential of mean force (PMF).

Results and discussion

Assessing molecular docking tools for binding affinity prediction

As shown in previous studies, there are no universal computational methods that can achieve accurate predictions for all protein–inhibitor systems; that is to say, system dependence is prevailing in computational science.[60–62] Therefore, in our case, various docking techniques were evaluated. According to the Pearson's correlation coefficients (R) between the docking scores and the experimental data (pIC_{50}) shown in Fig. 3, Glide and Glide with Prime-MM/GBSA rescoring exhibit relatively better prediction capability than Autodock Vina and Surflex-Dock (R: 0.52 and 0.51 *versus* 0.41 and 0.17). Besides, IFD performs

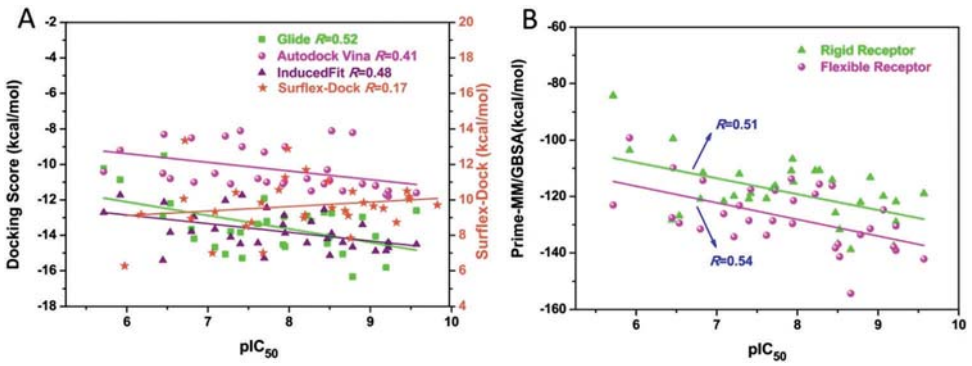

Fig. 3 Correlations between the docking scores and the experimental pIC_{50}.

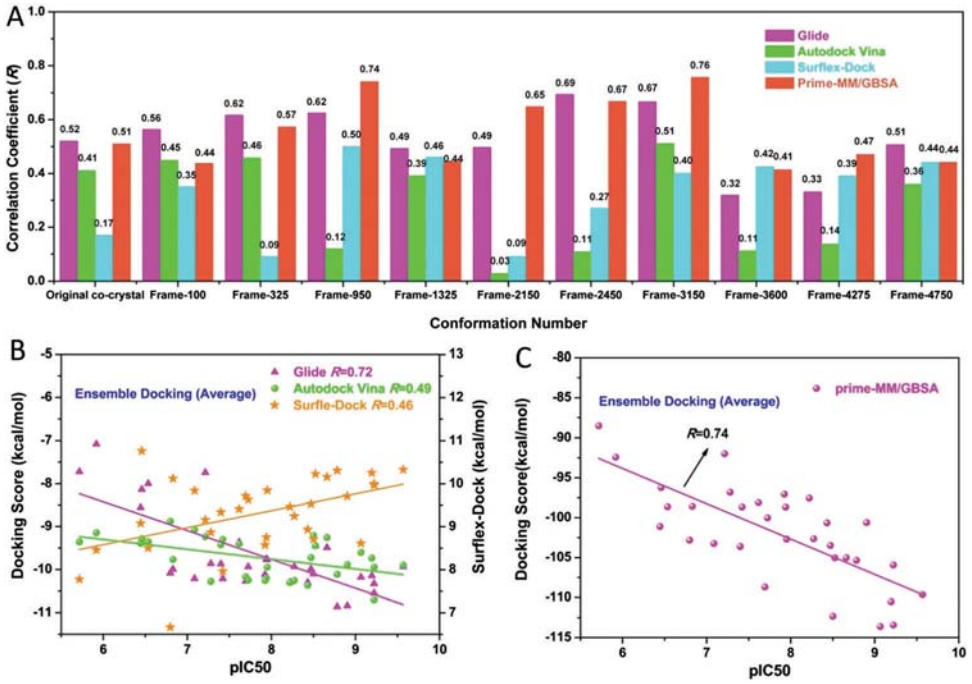

Fig. 4 Pearson correlation coefficients between the experimental pIC_{50} and (A) the docking results for each ALK conformer, or (B and C) the average docking scores of ensemble docking.

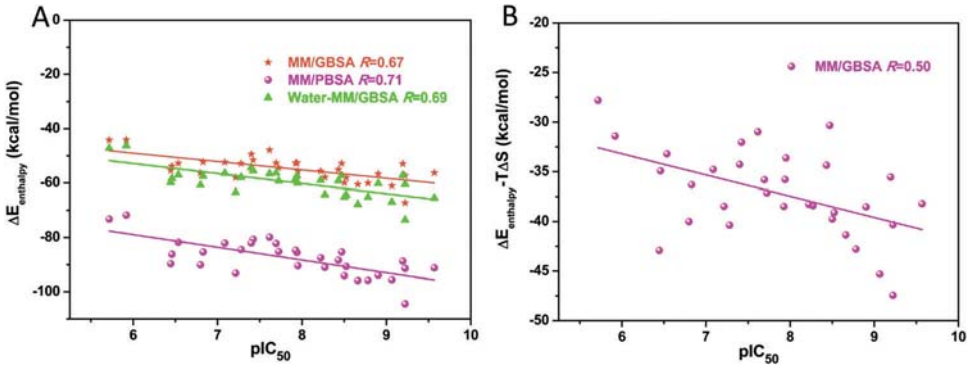

Fig. 5 Pearson correlation coefficients between the experimental pIC_{50} and the binding free energies calculated by (A) classical MM/GBSA (dark red), MM/PBSA (red) and water-MM/GBSA (green), or (B) MM/GBSA combined with entropic contribution ($T\Delta S$).

worse than the rigid-receptor docking (R: 0.48 and 0.52), which may be inconsistent with the commonly accepted knowledge that incorporating protein flexibility can improve the prediction accuracy of molecular docking. Here we only incorporated the conformational change of the side chains of a limited set of residues into flexible docking (residues within 6 Å radius of the ligand, that is, a too simplified method), and it may inevitably introduce inappropriate and/or insufficient dynamics or fluctuations that will impair the accuracy of molecular docking. Therefore, our results imply that traditional docking methodologies have limited capability to correctly incorporate protein flexibility in the binding of Type-I1/2 ALK inhibitors.

Behaviors of "conformational selection" and "induced-fit" models in molecular recognition

Docking into the MD ensemble. Protein flexibility (conformational change) involved in the selective binding of a ligand can be explained by conformational selection theory

Table 1 Binding free energies and individual energy components of several representative ALK inhibitors predicted by MM/GBSA and US (kcal mol^{-1})

Name	001-007	001-013	001-017	002-012	002-016
ΔE_{ele} [a]	−32.1 ± 0.84	−118.6 ± 6.29	−110.15 ± 5.99	−119.01 ± 4.81	−93.19 ± 3.78
ΔE_{vdW} [b]	−63.0 ± 0.26	−60.1 ± 0.50	−66.46 ± 0.48	−62.53 ± 0.28	−53.31 ± 0.43
ΔG_{GB} [c]	48.3 ± 0.48	132.5 ± 5.63	126.01 ± 5.27	133.90 ± 5.24	106.59 ± 4.79
ΔG_{SA} [d]	−5.5 ± 0.05	−5.3 ± 0.03	−5.7 ± 0.007	−5.1 ± 0.05	−4.3 ± 0.10
$\Delta E_{non\text{-}polar}$ [e]	−68.5 ± 0.32	−65.4 ± 0.53	−72.1 ± 0.47	−67.7 ± 0.33	−57.6 ± 0.53
ΔE_{polar} [f]	16.2 ± 0.35	13.8 ± 0.66	15.9 ± 0.72	14.9 ± 0.42	13.4 ± 1.01
ΔG_{bind} [g]	−52.3 ± 0.03	−51.6 ± 1.19	−56.3 ± 0.25	−52.8 ± 0.09	−44.2 ± 0.47
ΔG_{off} [h]	9.5 ± 0.13	12.5 ± 0.08	15.6 ± 0.29	9.0 ± 0.18	1.9 ± 0.04
ΔW_{PMF} [i]	11.4 ± 0.29	13.3 ± 0.12	16.5 ± 0.22	12.5 ± 0.08	6.4 ± 0.26
pIC_{50}	−6.83	−7.42	−9.57	−6.54	−5.72
$R(\Delta E_{non\text{-}polar})$	0.79		$R(\Delta E_{polar})$	0.53	
$R(\Delta G_{bind})$	0.82		$R(\Delta W_{PMF})$	0.90	

[a] Electrostatic interaction. [b] van der Waals interaction. [c] Polar contribution of the solvation effect. [d] Non-polar contribution of the solvation effect. [e] Non-polar interaction. [f] Polar interaction. [g] Binding free energy. The standard deviations were estimated based on five blocks. [h] Activation free energy of dissociation. The standard deviations were estimated from the PMF values of the last 3–5 ns US simulations. [i] PMF depth determined by the last 3–5 ns US simulations.

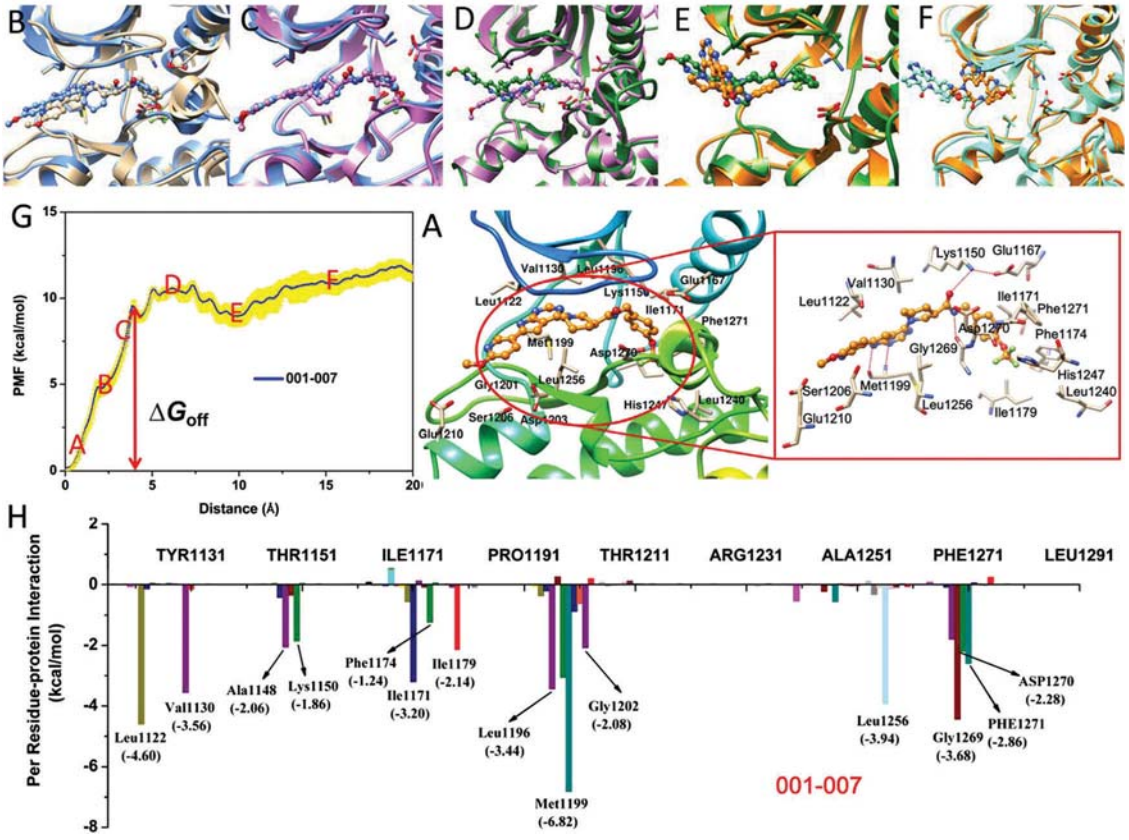

Fig. 6 Dissociation process of **001-007** from the binding site of ALK along the ATP channel (panel A–F) and the corresponding PMF curve (panel G). The initial crystal structure and inhibitor–residue interaction spectra of the ALK/**001-007** complex are illustrated in panels A and H, respectively.

(population-shift mechanism),[63–67] and a feasible strategy to account for protein flexibility in molecular docking may be the use of multiple structures or conformational ensembles derived from MD simulations.[68,69] Here, 10 representative structures extracted from the 50 ns MD trajectory (Fig. 2) were used for the docking of the 32 Type-I1/2 ALK inhibitors listed in Table S1 (ESI†). As shown in Fig. 4A, several ensemble members yield better performance than the reference crystal structure (such as Glide: R_{100} = 0.56, R_{325} = 0.62, R_{950} = 0.62, R_{2450} = 0.69, R_{3150} = 0.67 *versus* $R_{reference}$ = 0.52; Prime-MM/GBSA-rigid: R_{325} = 0.57, R_{950} = 0.74, R_{2150} = 0.65, R_{2450} = 0.67, R_{3150} = 0.76 *versus* $R_{reference}$ = 0.51; Autodock Vina: R_{100} = 0.45, R_{325} = 0.46, R_{3150} = 0.51 *versus* $R_{reference}$ = 0.41; Surflex-Dock: R_{100} = 0.35, R_{950} = 0.50, R_{1325} = 0.46, R_{2450} = 0.27, R_{3150} = 0.40, R_{3600} = 0.42, R_{4275} = 0.39, R_{4750} = 0.44 *versus* $R_{reference}$ = 0.17), which, to a large extent, might benefit from their preferable geometries to form better energetic complementarity with the bound partners. An alternative explanation would be that the well-behaved conformers might favor the formation of the initial contacts with the ligands by some specific residues. Meanwhile, when unfavorable protein structures are employed, the docking methods may fail to produce the correct binding orientation of the ligands because the crucial ligand binding regions might have been occupied by the disturbed residues of the protein. In addition, the C_α-RMSDs between the active pocket of each conformer and the reference were calculated (Fig. 2D), where we found that the MD simulations have introduced small-scale structural disturbance (backbone and side-chain fluctuations) to the structure ensemble (0.7–2.5 Å), and the success rate of docking for a conformer has no apparent correlation with its similarity to the reference bound state. Considering the one-sidedness of using a single conformer to describe the realistic binding feature, all the docking scores generated by different conformers were averaged to get a mean ensemble docking score. As shown in Fig. 4B and C, the ensemble docking scores have much higher correlations (Glide: R = 0.72; Prime-MM/GBSA-rigid: R = 0.74) with the experimental data than those based on the crystal structure. Overall, the above analyses demonstrate that the molecular docking based on a validated conformational ensemble could be utilized to improve the prediction accuracy of the docking based on a single crystal structure.

Incorporating flexibility and solvent effect in molecular recognition. Apart from the ''conformational selection'', an alternative approach to mimic the binding mechanism of a ligand is the ''induced-fit'' process,[70] where the flexibility can be introduced explicitly by MD simulations to imitate sidechain and/or backbone fluctuations.[62,71,72] Herein, MD simulations were launched to incorporate the induced-fit effect into protein–ligand recognition. The initial structures were generated from the above Glide docking because of its relatively better performance among the investigated docking methods. As shown in Fig. S2 (ESI†), the RMSDs of the representative inhibitor–ALK complexes

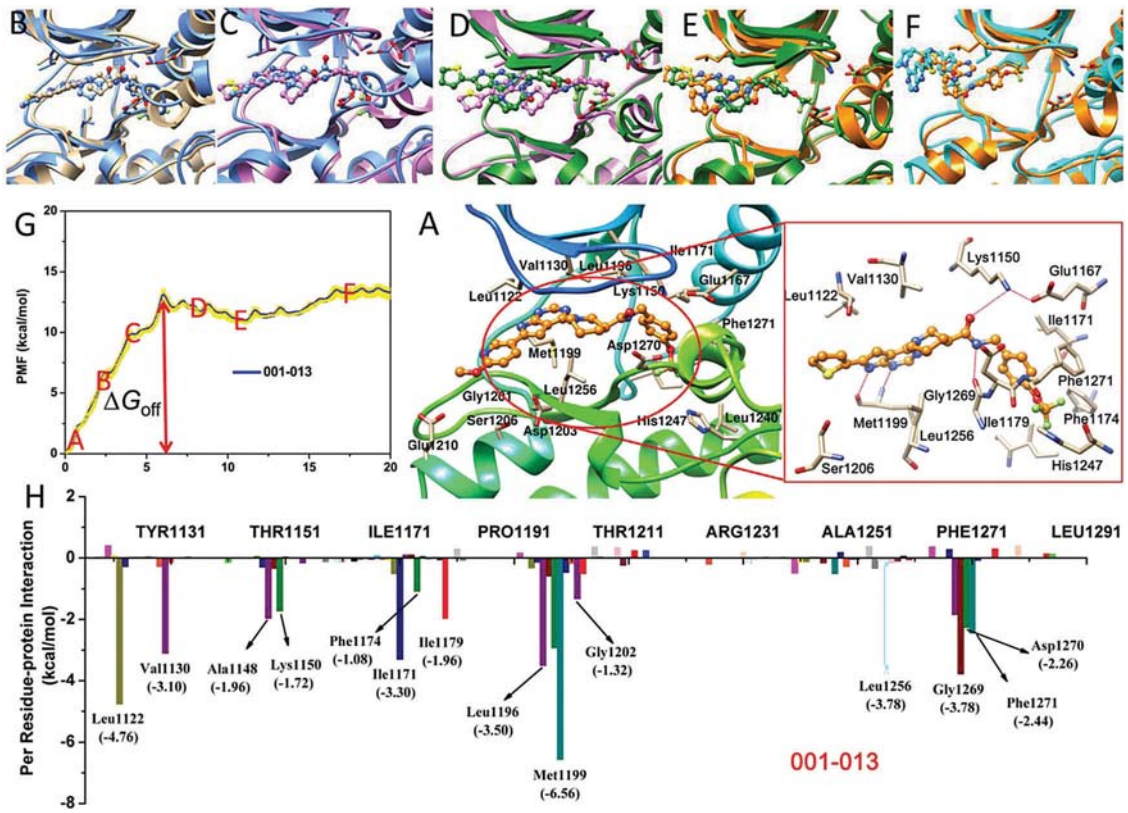

Fig. 7 Dissociation process of **001-013** from the binding site of ALK kinase along the ATP channel (panel A–F) and the corresponding PMF curve (panel G). The initial crystal structure and inhibitor–residue interaction spectra of the ALK/**001-013** complex are illustrated in panels A and H, respectively.

(**001-007**, **001-013**, **001-017**, **002-012** and **002-016**) (Fig. S2A, ESI†), as well as the binding site and the bound inhibitors (Fig. S2B, ESI†), achieve stability after ~1 ns MD simulation. Moreover, the root-mean-square fluctuations (RMSFs) *versus* per residue of the selected systems are illustrated in Fig. S3 (ESI†), where we can observe that the RMSF distributions and the trends of the fluctuations are roughly consistent with each other, implying that the results are suitable for further analyses.

Then the binding free energies of all the inhibitor–ALK systems were predicted using the MM/GBSA and MM/PBSA methods based on the 400 snapshots derived from the last 4 ns MD trajectories. It can be observed from Fig. 5A that MM/PBSA performs relatively better than MM/GBSA (R = 0.71 *versus* R = 0.67). Obviously, the MD-based MM/PBSA and MM/GBSA calculations exhibit advantages in predicting the binding affinities of the Type-I1/2 inhibitors of ALK over rigid receptor docking. Besides, although previous studies indicated that the changes in vibrational, translational, and rotational entropies are closely associated with protein–ligand interactions, but unfortunately, the inclusion of entropic contributions in our case cannot improve the accuracy of the binding free energy calculation (R = 0.50, Fig. 5B). Moreover, the limited improvement of incorporating explicit water into MM/GBSA compared with classical MM/GBSA (R = 0.69 *versus* R = 0.67) indicates that dealing with the essential water molecules is still a big issue in MM/GBSA and MM/PBSA calculations (Fig. 5A). The failure of the above attempts may attribute to the fact that the inclusion of more contributions to energy calculations may bring additional noise into binding pose prediction and binding affinity estimation. To sum up, MD simulations can indeed improve the prediction accuracy of molecular recognition by providing structural flexibility, and we should also put emphasis on avoiding the undesirable fluctuations and unfavorable contributions to the end-point binding free energy calculations.

Binding mechanism described by a combined molecular modeling strategy

Although none of the above mentioned methods yielded satisfactory prediction accuracy (such as $R > 0.8$), the MD-based MM/PBSA or MM/GBSA calculation and ensemble docking are sufficient to distinguish the tight-binding inhibitors from the loose-binding ones for the ALK system (Fig. 4 and 5). Moreover, we can observe that the non-polar parts exhibit higher linear correlations with the experimental data (ΔE_{vdW}: R = 0.71, ΔG_{SA}: R = 0.72) than the polar parts (ΔE_{ele}: R = 0.06, ΔG_{GB}: R = 0.06) (Fig. S5, ESI†), suggesting that the non-polar contributions are the determinant factors for the binding of these Type-I1/2 inhibitors.

Considering the relatively short timescale and limited sampling capacity of MD simulations, the US simulations (41 × 5 ns = 205 ns) and the MD-based (4 ns) free energy decompositions were combined to investigate the binding mechanisms of the studied

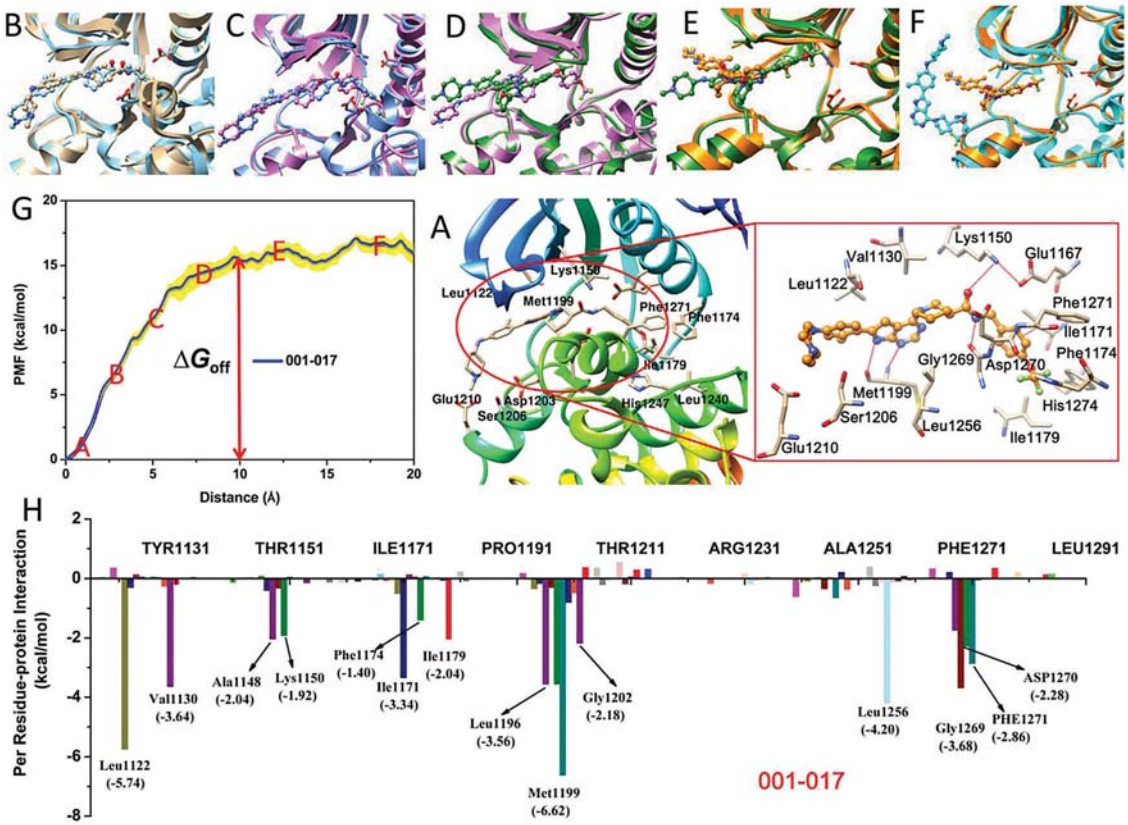

Fig. 8 Dissociation process of **001-017** from the binding site of ALK kinase along the ATP channel (panel A–F) and the corresponding PMF curve (panel G). The initial crystal structure and inhibitor–residue interaction spectra of the ALK/**001-017** complex are illustrated in panels A and H, respectively.

ALK inhibitors. Because of the high computational cost, only several representative inhibitors (**001-007**, **001-013**, **001-017**, **002-012**, and **002-016**) were used for the investigation. As shown in Fig. S4 (ESI†), the systems converged after ~2 ns US simulation (2–3, 3–4, and 4–5 ns), and thus the PMF curves were depicted based on the last 3 ns US samples (3–5 ns). As illustrated in Table 1, the predicted binding affinities (PMF depth, ΔW_{PMF} and ΔG_{bind}) exhibit good consistency with the experimental data ($R(\Delta W_{PMF}) = 0.90$, $R(\Delta G_{bind}) = 0.82$). Furthermore, previous studies suggest that a larger activation energy of dissociation (ΔG_{off}) corresponds to a prolonged residence time: $\Delta G_{off} \propto \ln(1/k_{off})$, where $1/k_{off}$ represents the residence time.[56,73,74] It can be found in Fig. 6G–10G that a large barrier located at ~9.8 Å of the RC in the **001-017** system, which may significantly lengthen the residence time with the ΔG_{off} of ~15.6 kcal mol^{-1}, much larger than those in the systems of **001-007**, **001-013**, **002-012**, and **002-016** (9.5, 12.5, 9.0, and 1.9 kcal mol^{-1}, respectively, Table 1), and further confirms **001-017** as a more potential ALK inhibitor.

Comparison of reaction coordinates with inhibitors sharing R1 modifications. As we can see from Fig. 6G–8G, the increasing trend of the PMF curves of **001-007**, **001-013** and **001-017** is similar while the energy profile of **001-017** is relatively higher than those of the other two systems (ΔW_{PMF}), suggesting that **001-017** needs to overcome more obstacles before escaping from the binding pocket (namely, exhibiting a longer residence time). With the increase of the biasing potentials on **001-007**, **001-013** and **001-017**,

Table 2 Hydrogen bond analysis extracted from the MD trajectories

Donor	Acceptor	**001-007** (%)[a]	**001-013** (%)	**001-017** (%)[a]	**002-012** (%)[a]	**002-016** (%)[a]
Met1199-NH	Ligand-N1	(65.0 ± 8.87)	(68.6 ± 2.59)	(70.0 ± 7.56)	(70.8 ± 3.20)	(73.0 ± 3.67)
Ligand-NH	Met1199-CO	(64.4 ± 14.22)	(54.4 ± 11.62)	(65.6 ± 4.21)	(61.0 ± 9.60)	(54.6 ± 6.16)
Lys1150-NH	Ligand-CO	(30.2 ± 7.75)	(24.0 ± 13.46)	(34.2 ± 10.70)	(26.2 ± 2.48)	(58.0 ± 7.36)
Ligand-NH	Gly1269-CO	(63.4 ± 1.78)	(69.6 ± 2.06)	(68.4 ± 11.96)	(69.0 ± 6.40)	(7.4 ± 7.12)
Ligand-NH	Glu1197-CO	(6.8 ± 2.34)	(19.8 ± 3.76)	(19.0 ± 4.54)	(10.6 ± 2.48)	(13.0 ± 2.34)

[a] The occupancy of the Hbonds were calculated based on the 101–500 frames from MD trajectories by VMD software, and the uncertainties were defined by $s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$ (101–200, 201–300, 301–400, and 401–500).

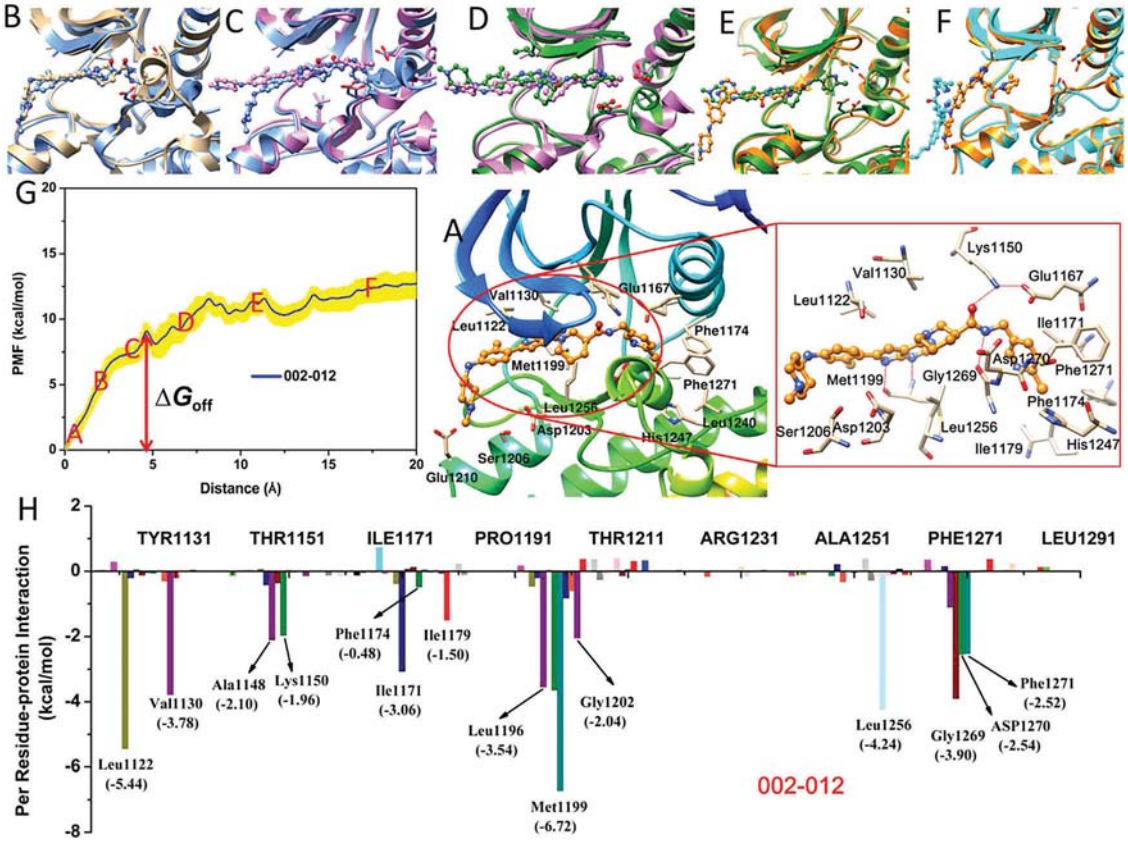

Fig. 9 Dissociation process of **002-012** from the binding site of ALK kinase along the ATP channel (panel A–F) and the corresponding PMF curve (panel G). The initial crystal structure and inhibitor–residue interaction spectra of the ALK/**002-012** complex are illustrated in panels A and H, respectively.

the hydrogen bonds between the carboxamide carbonyl (–CO) group and the sidechain ($–NH_2$) of Lys1150 vanish quickly in the three systems (Fig. 6B–8B). Subsequently, the ligands continue to move horizontally along the pocket channel, accompanied by the attenuation of the original interactions (residues both in the allosteric and ATP pockets) and the formation of new interactions, such as the p–π interaction between the phenyl ring of the ligands and Asp1270. Under a biasing potential (PMF curve) of $\sim$5 kcal mol^{-1}, the amide NH of **001-007** loses the hydrogen interaction with the backbone carbonyl of Gly1269 at $\sim$2.5 Å of the RC (Fig. 6B), and both **001-017** and **001-013** get rid of this constraint at $\sim$4 Å and $\sim$3.5 Å of the RCs, respectively, with a PMF value of $\sim$9 kcal mol^{-1} (Fig. 8C and 7C). Then, when the biasing potential reaches $\sim$11 kcal mol^{-1} in the PMF curve, other two H-bonding interactions in the **001-017** and **001-013** systems formed by N1 of the aminopyrimidine ring in Met1199 and NH of the imidazole ring of the ligands are disrupted (Fig. 8D and 7D). As for the **001-007** system, these two H-bonding interactions are lost under a biasing potential of $\sim$7 kcal mol^{-1} (Fig. 6C). Without the constraint of hydrogen bonds, the drugs collapse soon and constantly adjust the posture to accommodate themselves in the binding cavity. The flat regions in Fig. 6D and 7D suggest that the strain energies released in both the **001-007** and **001-013** systems just compensate for the biasing potential added to the drugs to shake off the other non-Hbond interactions. Whereas, in the **001-017** system, enhanced biasing potential is required to handle these remaining constraints that cannot be offset by simply conformational adjustment (upgrading region at 5–9 Å of the RCs in Fig. 8G).

As shown in Fig. 6A–8A, although the fragments of *N*-phenylpiperazine, 2-methoxypyridine and sulfur heterocyclic (R_1 moieties) in the initial structures of **001-007**, **001-013**, and **001-017** are all sandwiched by Leu1122 and the hinge region of ALK, different positions with a slightly different posture are shown for these moieties. Nevertheless, these inhibitors are tightly stabilized in the narrow gorge of the binding pocket with a "DFG-shifted" conformation (Type-I1/2 inhibitor conformation) by intricate interaction networks, such as the H-bonding interactions with residues Lys1150, Gly1269, and Met1199. The quantitative analyses of the H-bonds, as listed in Table 2, indicate that the H-bond contributions are almost equal to each other in all the three systems, implying that the non-Hbond interactions may be the key factor in determining the different dissociation behaviors of the above studied systems. The per-residue energy analyses further confirm the conjecture, where **001-017** forms more favorable interactions with most residues located in the ATP pocket than **001-007** and **001-013**, such as the H–π interaction with the sidechain of Leu1122 (−5.74, −4.6 and −4.76 kcal mol^{-1}), van der Waals interactions with Leu1198 (−3.56, −3.06 and −2.92 kcal mol^{-1}), and Gly1202 (−2.18, −2.08 and −1.32 kcal mol^{-1}), *etc.*

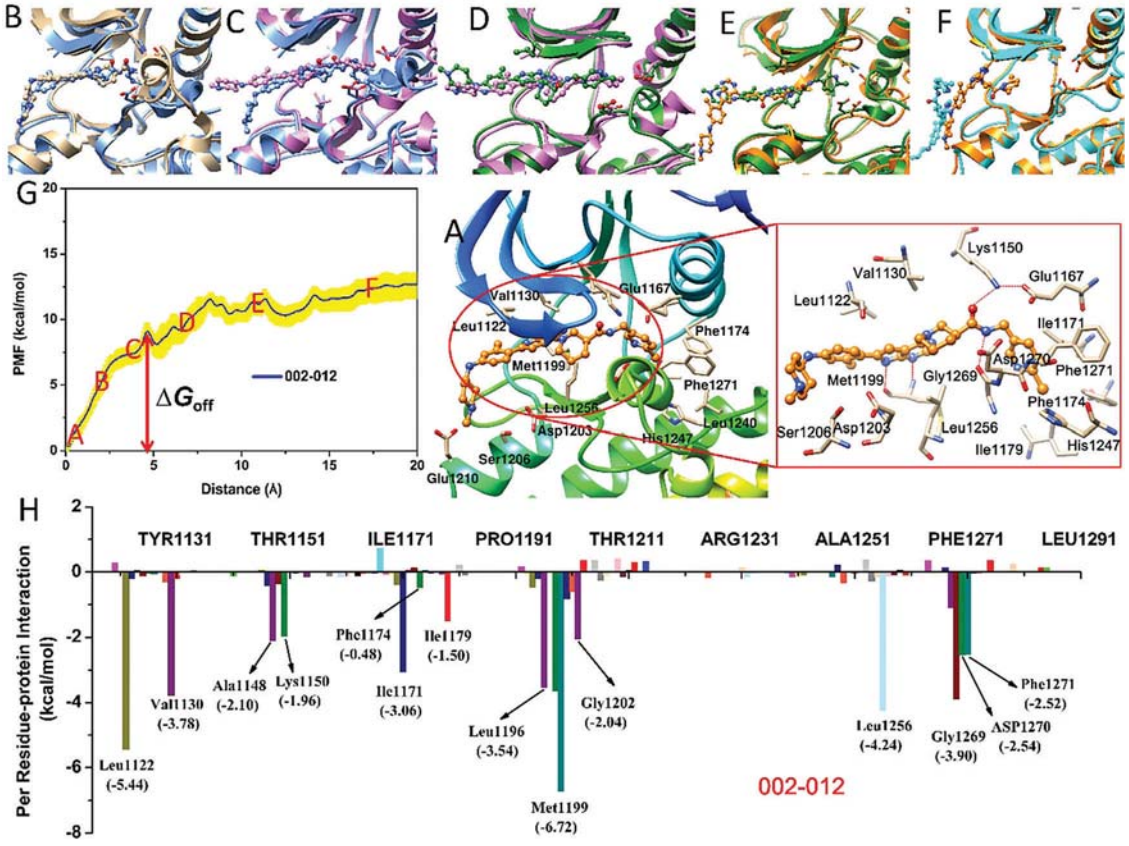

Fig. 10 Dissociation process of **002-016** from the binding site of ALK kinase along the ATP channel (panel A–F) and the corresponding PMF curve (panel G). The initial crystal structure and inhibitor–residue interaction spectra of the ALK/**002-016** complex are illustrated in panels A and H, respectively.

Comparison of reaction coordinates with inhibitors sharing modifications located in the allosteric pocket. The moieties that occupy the extended allosteric pocket of ALK in the **001-017**, **002-012**, and **002-016** systems are 4-(trifluoromethoxy) phenylamine, 2-chloropyridine (R_2), and *N*-piperidine (R_3) fragments, respectively, and the PMF curves of their dissociation processes are illustrated in Fig. 8–10, respectively. As shown in the regions A–C of the RCs in Fig. 8G and 9G, the biasing potentials added to **001-017** and **002-012** gradually increase at the beginning of the unbinding processes. Then the drugs move horizontally toward the entrance of the binding cavity (Fig. 8B, C and 9B, C) accompanied by the breakage of the H-bond interaction networks in the allosteric pocket. The per-residue energy analyses indicate that the residues in the allosteric pocket are more favorable in forming the interactions with the 4-(trifluoromethoxy) phenylamine fragment of **001-017** than the 2-chloropyridine (R_2) fragment of **002-012** (Phe1174: −1.40 *versus* −0.48 kcal mol^{-1}, Ile1179: −2.04 *versus* −1.50 kcal mol^{-1}, His1247: −0.64 *versus* −0.14 kcal mol^{-1}, Ile1268: −1.74 *versus* −1.10 kcal mol^{-1}, *etc.*, Fig. 8H *versus* Fig. 9H and Table S2, ESI†), which can explain the steeper and higher PMF curve of the **001-017** system. As for the **002-016** system, after a short rising stage of the PMF (0–1 Å, Fig. 10B), the biasing potential remains unchanged across 1–5 Å of the RC as shown in Fig. 10G. Further analysis shows that the drugs in the **001-017** and **002-012** systems can form effective H-bonding interactions with the backbone of Gly1269 (occupancy ratio: 68.4% and 69.0%, and Gly1269: −3.68 and −3.90 kcal mol^{-1}, respectively), whereas, in the case of **002-016**, this interaction becomes very weak (occupancy ratio: 7.4%, Gly1269: 0.4 kcal mol^{-1}). According to the initial structure of **002-016**/ALK, we speculate that it may be difficult for the tiny *N*-piperidine group to form effective interactions with residues surrounding the relatively spacious allosteric pocket. And this has been quantitatively confirmed by the energy decomposition analyses in Fig. 10H (such as Ile1171: −3.34 *versus* −1.50 kcal mol^{-1}, Phe1174: −1.40 *versus* −0.02 kcal mol^{-1}, Ile 1179: −2.04 *versus* −0.22 kcal mol^{-1}, His1247: −0.64 *versus* 0 kcal mol^{-1}, Ile1268: −1.74 *versus* −0.26 kcal mol^{-1}, Asp1270: −2.28 *versus* −1.60 kcal mol^{-1}, Phe1271: −2.86 *versus* −1.28 kcal mol^{-1} for **001-017** and **002-016**, respectively). Then, these three inhibitors struggle to shake off the nearly equal attractions (Fig. 8H–10H and Table S2, ESI†) from residues surrounding the ATP pocket (Leu1122, Gly1202, Val1130, Glu1210, Leu1256, Met1199, *etc.*), and finally move out of the pocket. The higher electronegativity of the relatively bulky 4-(trifluoromethoxy)phenylamine fragment of compound **001-017** tends to form stronger transient interactions with residues along the dissociation channel than the 2-chloropyridine (R_2) fragment and the *N*-piperidine fragment of **002-012** and **002-016**, which is reflected by the larger ΔG_{off} value in Fig. 8H–10H.

Conclusion

Herein, by employing a combined computational strategy, the molecular recognition mechanisms between the Type-I1/2 inhibitors and ALK kinase were exhaustively explored. The binding structures of these inhibitors were preliminarily obtained by molecular docking, but unfortunately the ranking capabilities of the docking protocols are far from satisfactory due to the complexity of ligand–protein interactions. The improved prediction accuracy of ensemble docking and MD-based MM/PBSA or MM/GBSA binding free energy calculations indicate that it is very important to incorporate receptor flexibility as well as solvent effect in the prediction of binding affinity. The US simulations show that the inhibitors with a tail moiety bearing stronger electronegativity and relatively larger bulkiness will be more favorable to form effective interactions with residues surrounding the allosteric pocket, and this is quantitatively consistent with the results provided by MM/GBSA binding free energy decomposition. Moreover, the strongest binding affinity, together with the highest energetic barrier (which means the longest residence time), further confirms **001-017** as a more potential lead compound among all the designed Type-I1/2 ALK inhibitors.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This study was supported by the National Key R&D Program of China (2016YFA0501701, 2016YFB0201700), the National Science Foundation of China (81603031, 21575128, 81773632), and the Ministry of Science and Technology of China (2016YFA0501700).

References

1 B. Hallberg and R. H. Palmer, *Nat. Rev. Cancer*, 2013, **13**, 685–700.
2 T. Sasaki and P. A. Jänne, *Clin. Cancer Res.*, 2011, **17**, 7213–7218.
3 A. T. Shaw, D.-W. Kim, K. Nakagawa, T. Seto, L. Crinó, M.-J. Ahn, T. De Pas, B. Besse, B. J. Solomon and F. Blackhall, *N. Engl. J. Med.*, 2013, **368**, 2385–2394.
4 B. J. Solomon, T. Mok, D.-W. Kim, Y.-L. Wu, K. Nakagawa, T. Mekhail, E. Felip, F. Cappuzzo, J. Paolini and T. Usari, *N. Engl. J. Med.*, 2014, **371**, 2167–2177.
5 Y. L. Choi, M. Soda, Y. Yamashita, T. Ueno, J. Takashima, T. Nakajima, Y. Yatabe, K. Takeuchi, T. Hamada and H. Haruta, *N. Engl. J. Med.*, 2010, **363**, 1734–1739.
6 T. Sasaki, J. Koivunen, A. Ogino, M. Yanagita, S. Nikiforow, W. Zheng, C. Lathan, J. P. Marcoux, J. Du and K. Okuda, *Cancer Res.*, 2011, **71**, 6051–6060.
7 R. Katayama, A. T. Shaw, T. M. Khan, M. Mino-Kenudson, B. J. Solomon, B. Halmos, N. A. Jessop, J. C. Wain, A. T. Yeo and C. Benes, *Sci. Transl. Med.*, 2012, **4**, 120ra117.
8 C. M. Lovly and W. Pao, *Sci. Transl. Med.*, 2012, **4**, 120ps122.
9 R. Roskoski, *Pharmacol. Res.*, 2017, **117**, 343–356.

10 P. Pan, H. Yu, Q. Liu, X. Kong, H. Chen, J. Chen, Q. Liu, D. Li, Y. Kang, H. Sun, W. Zhou, S. Tian, S. Cui, F. Zhu, Y. Li, Y. Huang and T. Hou, *ACS Cent. Sci.*, 2017, **3**, 1208–1220.
11 Z. Zhao, H. Wu, L. Wang, Y. Liu, S. Knapp, Q. Liu and N. S. Gray, *ACS Chem. Biol.*, 2014, **9**, 1230–1241.
12 S. C. Meyer, M. D. Keller, S. Chiu, P. Koppikar, O. A. Guryanova, F. Rapaport, K. Xu, K. Manova, D. Pankov and R. J. O'Reilly, *Cancer Cell*, 2015, **28**, 15–28.
13 C. C. Smith, K. Lin, A. Stecula, A. Sali and N. P. Shah, *Leukemia*, 2015, **29**, 2390–2392.
14 A. Gupta, C.-L. Yong, J. B. Madwed, H. Staehle and C. C. Wood, *J. Allergy Clin. Immunol.*, 2002, **109**, S67.
15 C. Pargellis, T. Liang, L. Churchill, P. F. Cirillo, T. Gilmore, A. G. Graham, P. M. Grob, E. R. Hickey, N. Moss and S. Pav, *Nat. Struct. Mol. Biol.*, 2002, **9**, 268.
16 M. C. Bryan, D. A. Whittington, E. M. Doherty, J. R. Falsey, A. C. Cheng, R. Emkey, R. L. Brake and R. T. Lewis, *J. Med. Chem.*, 2012, **55**, 1698–1705.
17 C.-H. Tu, W.-H. Lin, Y.-H. Peng, T. Hsu, J.-S. Wu, C.-Y. Chang, C.-T. Lu, P.-C. Lyu, C. Shih and W.-T. Jiaang, *J. Med. Chem.*, 2016, **59**, 3906–3919.
18 L. F. Epstein, H. Chen, R. Emkey and D. A. Whittington, *J. Biol. Chem.*, 2012, **287**, 37447–37457.
19 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
20 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
21 F. Ogliaro, M. Bearpark, J. Heyd, E. Brothers, K. Kudin, V. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari and A. Rendell, *Gaussian 09*, Gaussian Inc., Wallingford, CT, 2009.
22 K. Verschueren, F. SeljÃ, H. J. Rozeboom, K. H. Kalk and B. W. Dijkstra, *Nature*, 1993, **363**, 693.
23 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Am. Chem. Soc.*, 2001, **222**, U403.
24 D. A. Case, V. Babin, J. Berryman, R. Betz, Q. Cai, D. Cerutti, T. Cheatham Iii, T. Darden, R. Duke and H. Gohlke, Amber 14, University of California, San Francisco, 2014.
25 D. Paschek, R. Day and A. E. García, *Phys. Chem. Chem. Phys.*, 2011, **13**, 19840–19847.
26 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
27 M. Levitt, *J. Mol. Biol.*, 1974, **82**, 393–420.
28 S. Lambrakos, J. Boris, E. Oran, I. Chandrasekhar and M. Nagumo, *J. Comput. Phys.*, 1989, **85**, 473–486.
29 A. Likas, N. Vlassis and J. J. Verbeek, *Pattern Recogn.*, 2003, **36**, 451–461.
30 L. Song, D. Li, X. Zeng, Y. Wu, L. Guo and Q. Zou, *BMC Bioinf.*, 2014, **15**, 298.
31 Q. Zou, S. Wan, Y. Ju, J. Tang and X. Zeng, *BMC Syst. Biol.*, 2016, **10**, 114.
32 G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju and W. Sherman, *J. Comput.-Aided Mol. Des.*, 2013, **27**, 221–234.
33 R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley and J. K. Perry, *J. Med. Chem.*, 2004, **47**, 1739–1749.
34 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455–461.
35 J. Fuhrmann, A. Rurainski, H. P. Lenhof and D. Neumann, *J. Comput. Chem.*, 2010, **31**, 1911–1918.
36 A. N. Jain, *J. Med. Chem.*, 2003, **46**, 499–511.
37 M. Clark, R. D. Cramer and N. Van Opdenbosch, *J. Comput. Chem.*, 1989, **10**, 982–1012.
38 W. Sherman, T. Day, M. P. Jacobson, R. A. Friesner and R. Farid, *J. Med. Chem.*, 2006, **49**, 534–553.
39 H. Sun, Y. Li, S. Tian, L. Xu and T. Hou, *Phys. Chem. Chem. Phys.*, 2014, **16**, 16719–16729.
40 B. R. Miller III, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, *J. Chem. Theory Comput.*, 2012, **8**, 3314–3321.
41 F. Chen, H. Liu, H. Sun, P. Pan, Y. Li, D. Li and T. Hou, *Phys. Chem. Chem. Phys.*, 2016, **18**, 22129–22139.
42 T. Hou, J. Wang, Y. Li and W. Wang, *J. Chem. Inf. Model.*, 2011, **51**, 69–82.
43 T. Hou, J. Wang, Y. Li and W. Wang, *J. Comput. Chem.*, 2011, **32**, 866–877.
44 T. Hou and R. Yu, *J. Med. Chem.*, 2007, **50**, 1177–1188.
45 L. Xu, H. Sun, Y. Li, J. Wang and T. Hou, *J. Phys. Chem. B*, 2013, **117**, 8408–8421.
46 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.
47 J. Wang, T. Hou and X. Xu, *Curr. Comput.-Aided Drug Des.*, 2006, **2**, 287–306.
48 J. Weiser, P. S. Shenkin and W. C. Still, *J. Comput. Chem.*, 1999, **20**, 217–230.
49 D. E. Tanner, J. C. Phillips and K. Schulten, *J. Chem. Theory Comput.*, 2012, **8**, 2521–2530.
50 C. Tan, L. Yang and R. Luo, *J. Phys. Chem. B*, 2006, **110**, 18680–18687.
51 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.
52 Y.-L. Zhu, P. Beroza and D. R. Artis, *J. Chem. Inf. Model.*, 2014, **54**, 462–469.
53 D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.
54 R. O. Dror, D. H. Arlow, P. Maragakis, T. J. Mildorf, A. C. Pan, H. Xu, D. W. Borhani and D. E. Shaw, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 18684–18689.
55 H. Sun, P. Chen, D. Li, Y. Li and T. Hou, *J. Chem. Theory Comput.*, 2016, **12**, 851–860.
56 H. Sun, Y. Li, S. Tian, J. Wang and T. Hou, *PLoS Comput. Biol.*, 2014, **10**, e1003729.
57 H. Sun, S. Tian, S. Zhou, Y. Li, D. Li, L. Xu, M. Shen, P. Pan and T. Hou, *Sci. Rep.*, 2015, **5**, 8457.
58 M. Petřek, M. Otyepka, P. Banáš, P. Košinová, J. Koča and J. Damborský, *BMC Bioinf.*, 2006, **7**, 316–325.
59 J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok and K. A. Dill, *J. Chem. Theory Comput.*, 2007, **3**, 26–41.
60 M. A. Lill, *Biochemistry*, 2011, **50**, 6157–6169.
61 T. Frączek, A. Siwek and P. Paneth, *J. Chem. Inf. Model.*, 2013, **53**, 3326–3342.

62 C. Pons, R. B. Fenwick, S. Esteban-Martín, X. Salvatella and J. Fernandez-Recio, *J. Chem. Theory Comput.*, 2013, **9**, 1830–1837.
63 B. Ma, S. Kumar, C.-J. Tsai and R. Nussinov, *Protein Eng.*, 1999, **12**, 713–720.
64 C. J. Tsai, S. Kumar, B. Ma and R. Nussinov, *Protein Sci.*, 1999, **8**, 1181–1190.
65 S. Kumar, B. Ma, C. J. Tsai, N. Sinha and R. Nussinov, *Protein Sci.*, 2000, **9**, 10–19.
66 D. D. Boehr, R. Nussinov and P. E. Wright, *Nat. Chem. Biol.*, 2009, **5**, 789–796.
67 O. Korb, T. S. Olsson, S. J. Bowden, R. J. Hall, M. L. Verdonk, J. W. Liebeschuetz and J. C. Cole, *J. Chem. Inf. Model.*, 2012, **52**, 1262–1274.
68 S. Tian, H. Sun, Y. Li, P. Pan, D. Li and T. Hou, *J. Chem. Inf. Model.*, 2013, **53**, 2743–2756.
69 S. Tian, H. Sun, P. Pan, D. Li, X. Zhen, Y. Li and T. Hou, *J. Chem. Inf. Model.*, 2014, **54**, 2664–2679.
70 M. Perutz, *Nature*, 1970, **228**, 726–734.
71 C. Dominguez, R. Boelens and A. M. Bonvin, *J. Am. Chem. Soc.*, 2003, **125**, 1731–1737.
72 J. Fernández-Recio, M. Totrov and R. Abagyan, *Proteins*, 2003, **52**, 113–117.
73 R. A. Copeland, D. L. Pompliano and T. D. Meek, *Nat. Rev. Drug Discovery*, 2006, **5**, 730–739.
74 P. J. Tummino and R. A. Copeland, *Biochemistry*, 2008, **47**, 5481–5492.