Showcasing research from the Group of Assoc. Prof. Weiwei Xue and Prof. Feng Zhu at Chongqing University and Zhejiang University, China

Prediction of the binding mode and resistance profile for a dual-target pyrrolyl diketo acid scaffold against HIV-1 integrase and reverse-transcriptase-associated ribonuclease H

The Innovative Drug Research and Bioinformatics (IDRB) group is working on the binding mechanism identification and structure-based design of multi-target drugs for complex diseases, such as mood disorders and HIV infection. Multi-target drugs have the benefits of rapid onset and/or higher efficacy for complex disease treatment. This computational work predicts the binding mode and potential resistance profile of the dual-target pyrrolyl diketo acid scaffold against HIV-1 integrase (IN) and the reverse-transcriptase-associated ribonuclease H (RNase H) active sites.

As featured in:

Prediction of the binding mode and resistance profile for a dual-target pyrrolyl diketo acid scaffold against HIV-1 integrase and reverse-transcriptase-associated ribonuclease H†

Fengyuan Yang,ab Guoxun Zheng,ab Tingting Fu,ab Xiaofeng Li,ab Gao Tu,ab Ying Hong Li,ab Xiaojun Yao,c Weiwei Xueab,* and Feng Zhuab,*

The rapid emergence of drug-resistant variants is one of the most common causes of highly active antiretroviral therapeutic (HAART) failure in patients infected with HIV-1. Compared with the existing HAART, the recently developed pyrrolyl diketo acid scaffold targeting both HIV-1 integrase (IN) and reverse transcriptase-associated ribonuclease H (RNase H) is an efficient approach to counteract the failure of anti-HIV treatment due to drug resistance. However, the binding mode and potential resistance profile of these inhibitors with important mechanistic principles remain poorly understood. To address this issue, an integrated computational method was employed to investigate the binding mode of inhibitor JMC6F with HIV-1 IN and RNase H. By using per-residue binding free energy decomposition analysis, the following residues: Asp64, Thr66, Leu68, Asp116, Tyr143, Gln148 and Glu152 in IN, Asp443, Glu478, Trp536, Lys541 and Asp549 in RNase H were identified as key residues for JMC6F binding. And then computational alanine scanning was carried to further verify the key residues. Moreover, the resistance profile of the currently known major mutations in HIV-1 IN and 2 mutations in RNase H against JMC6F was predicted by in silico mutagenesis studies. The results demonstrated that only three mutations in HIV-1 IN (Y143C, Q148R and N155H) and two mutations in HIV-1 RNase H (Y501R and Y501W) resulted in a reduction of JMC6F potency, thus indicating their potential role in providing resistance to JMC6F. These data provided important insights into the binding mode and resistance profile of the inhibitors with a pyrrolyl diketo acid scaffold in HIV-1 IN and RNase H, which would be helpful for the development of more effective dual HIV-1 IN and RNase H inhibitors.

Introduction

Human immunodeficiency virus 1 (HIV-1) remains one of the most devastating health problems all over the world, which leads to 70 million infections and 36 million deaths.1–3 Current therapeutic regimens for HIV-1 infection are primarily based on highly active antiretroviral therapy (HAART).4–6 This therapy usually combines multiple drugs aiming at various targets such as HIV-1 protease, reverse transcriptase (RT) and integrase (IN) as well as the host chemokine receptor (CCR5).7–9 In particular, both RT and IN play essential roles in viral replication.10–12 and three integrase strand transfer inhibitors (INSTI; raltegravir (RAL), elvitegravir (EVG) and dolutegravir (DTG), as shown in Fig. 1) and a dozen RT inhibitors (RTIs) have been approved by the U.S. Food and Drug Administration (FDA).13–15 In spite of the clinical success of HAART, several factors (extensive drug resistance,16–18 severe adverse effects,19 limited number of efficacious combinations,20 drug–drug interactions21 and the requirement for lifelong patient compliance22) have seriously affected its efficacy. Therefore, the discovery of anti-HIV agents of novel mechanisms other than drug combination is urgently needed.23–25

HIV-1 IN consists of a highly conserved DDE motif (Fig. 2A) that chelates with two Mg2+ ions in its catalytic core domain (CCD).26 Since the catalytic site of the ribonuclease H domain (RNase H) in HIV-1 RT shared a very similar topology with that of the CCD in IN(Fig. 2B),27 these 2 proteins are reported as a potential pair for discovering multi-target drugs.28–30

Cite this: Phys. Chem. Chem. Phys., 2018, 20, 23873

Received 22nd March 2018, Accepted 12th June 2018
DOI: 10.1039/c8cp01843j
rsc.li/pccp
The design of dual-action drugs with an anti-IN/RNase H pharmacological profile has therefore emerged as a promising strategy for treating HIV-1 infection. Different from the traditional approved RTIs, this new strategy blocks the catalytic activity of RT by inhibiting the RNase H domain (RHI), which is beneficial for suppressing viral resistance, reducing toxic side effects and ameliorating drug-drug interactions. 4 or more molecular scaffolds based on this strategy have been discovered, including hydroxytropolone, 2-hydroxyisoquinoline-1,3(2H,4H)-dione, madurahydroxylactone and diketo derivatives. Pyrrolyl diketo acid derivatives showed good effectiveness against the RAL resistant HIV-1 strain and the most recently discovered dual-action inhibitor (JMC6F in Fig. 2), a pyrrolyl diketo acid derivative, showed good correlation between IN and RNase H inhibition. Because JMC6F is reported to be one of the best dual inhibitors of this class, it can be selected as a representative of drugs with an anti-IN/RNase H pharmacological profile.

However, the molecular basis of (1) why the pyrrolyl diketo acid derivatives can simultaneously bind to IN and RNase H and (2) how these currently available drug-resistant mutations for FDA approved INSTIs and RTIs affect the affinity of this dual-action scaffold remains elusive. Molecular docking has provided some information on the interaction between the diketo derivatives and their corresponding target, but more detailed information such as dynamic and quantitative identification of the interactions between key residues and drugs is still not available. This information is especially valuable when considering the fact that the mutation of the key residues is the main reason leading to the resistance of anti-HIV drugs.
With the success of computational methods, such as molecular dynamics (MD) simulation and the binding free energy calculation tool, in structure-based drug design, it is of great interest to evaluate the application of these methods in order to address the elusive molecular basis discussed above. The MD simulations can provide plentiful dynamic structural information about protein–ligand interactions. Based on MD trajectories, several approaches can be applied to predict binding free energies, including free energy perturbation (FEP), thermodynamic integration (TI), and molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) methods. Compared to FEP or TI, due to the good balance between the speed and accuracy of the MM/PBSA method it has been widely applied in various protein–ligand systems, including the prediction of the binding mode and mutation-induced drug resistance profile of antiviral drugs to their targets (like influenza neuraminidase, HCV RNA polymerase and HIV protease). Herein, an integrated computational strategy was therefore used to identify the binding mode of JMC6F in HIV-1 IN and RNase H. In summary, this study revealed the resistance profile of antiviral drugs to their targets (like influenza neuraminidase, HCV RNA polymerase and HIV protease).

Molecular Docking of JMC6F into HIV-1 IN and RNase H

Molecular docking and MD simulations were employed to investigate the binding of JMC6F in HIV-1 IN and RNase H. The 3D structure of JMC6F was constructed with the Site-directed mutagenesis and binding free energy calculations were applied to assess the resistance profile of JMC6F based on the known resistance mutations in HIV IN or RNase H. In summary, this study revealed the resistance profile of the most promising dual inhibitor of HIV IN and RNase H, which could facilitate the discovery of new anti-HIV drugs of this class with an improved resistance profile.

Materials and methods

Homology modeling of HIV-1 IN and construction of the HIV-1 intasome complex

Homology modeling of the HIV-1 IN full-length structure was performed using Prime in Maestro. The sequence of HIV-1 IN was extracted from the GenBank (accession number: AAC83551.1). The crystal structure of PFV IN (PDB ID: 3OYA) was used as a model template. Before the construction of the homology model, sequence alignment of HIV/PFV IN and viral US DNA was done and adjusted manually to obtain reasonable matching. Then, the stereochemical quality of the models was evaluated using the Ramachandran plot of PROCHECK. Compared with the free HIV-1 IN, INSTIs preferentially bind to the intasome complex (IN, vDNA and Mg2+). The construction of the HIV-1 intasome complex was done using PyMOL. First, the 19-base-pair mimic of the pre-processed terminal vDNA from the structure of PFV IN (PDB ID: 3OYA) and the Mg2+ ions from the inhibited structure of PFV IN (PDB ID: 3S3M) were introduced into the HIV-1 IN model. Then, vDNA from the structure of the PFV intasome was mutated based on the sequence alignment of vDNA between HIV and PFV. In addition, three FDA approved INSTIs (RAL, EVG and DTG) from the inhibited structures of PFV IN (PDB ID: 3OYA, 3L2W and 3S3M) were fitted into the HIV-1 intasome.

Prior to docking JMC6F into HIV-1 IN and RNase H, cross-docking was performed on the co-crystal structures of PFV IN in complex with RAL and DTG (PDB ID: 3OYA and 3S3M) and HIV-1 RNase H in complex with MK1 and MK2 (PDB ID: 3LP0 and 3LP1) by Glide with standard precision (SP). In order to imitate the real biological environment of inhibition, Mg2+ ions were substituted for Mn2+ in the reported co-crystal structures. The structures of RAL, DTG, MK1 and MK2 were preprocessed by LigPrep using OPLS-2005 force field, and the ionized state was assigned by Epik at a pH value of 7.0 ± 2.0. OPLS-2005 is a popular force field used to perform molecular docking, and was successfully adopted previously for revealing the binding mechanisms between HIV protein targets (IN or RT) and their corresponding inhibitors. Moreover, the OPLS force field has emerged as the most recent force field developed by Schrödinger with significantly enhanced performance. In this study, to be consistent with previous studies, OPLS-2005 force field is applied. Co-crystal structures were prepared by adding hydrogen atoms, assigning partial charges and protonation states, and minimizing the structure using the Protein Preparation Wizard module in Maestro. And the docking grid box was defined by centering on the corresponding ligand in each complex using the Receptor Grid Generation module in Maestro. In order to ensure that there are no alternative binding sites with higher (similar) affinity, additional calculations were performed using the SiteMap code included by Schrödinger and the results are shown in Table S1. As shown in Table S1, seven potential binding sites were found and JMC6F more favorably binds at S1. Therefore, the S1 site was found with the highest docking performance which corresponds to the site we chose in this study. In molecular docking, 300 poses were generated during the initial phase of the docking procedure and all the poses were chosen for energy minimization by 100 steps of conjugate gradient minimization. Fig. S1 (ESI†) illustrated the superposition of the cross-docking poses of RAL, DTG, MK1 and MK2 onto their corresponding co-crystallized poses. The RMSD values between the docking pose and crystal pose of DTG, RAL, MK1 and MK2 were calculated (2.71, 3.40, 1.63 and 1.14, respectively). As shown, the RMSD values of RAL and DTG were higher than those of MK1 and MK2. This was mainly due to the flexible linker of RAL and DTG. But the docking pose at the catalytic site was quite similar. These results indicated that the Glide SP docking procedure is able to produce poses that are in close agreement with the crystal structures.

After validation of the docking procedure, the same parameter settings in cross-docking were applied to dock the prepared JMC6F into the INSTI binding site of IN and the RHI binding site of RT. The 3D structure of JMC6F was constructed with the...
building tool in Maestro. Docking grid boxes in IN and RNase H were defined by centering on DTG and MK1, respectively. The docking poses with the highest docking score were chosen for further MD simulation.

MD simulation

**System setup.** Six wild type (WT) and sixty-six mutant type systems were constructed for MD simulation (Table 1). Before the simulation, each system was prepared using the LEaP module of AMBER14 package. The AMBER force field ff14SB was applied for describing proteins, vDNA as well as Mg\textsuperscript{2+} ions. The force field parameters for optimizing ligands were created using the Antechamber program, using the General AMBER Force Field and restrained electrostatic partial charge. The geometric optimization and the electrostatic potential calculation of ligands were performed at the HF/6-31G* level of the Gaussian09 suite. Finally, each system was immersed into a rectangular periodic box of pre-equilibrated TIP3P water with at least 10 Å distance around the complexes. Finally, appropriate numbers of sodium counter ions were added to maintain the electroneutrality of the simulation system.

**Equilibration and production runs.** All MD simulations were performed using GPU-accelerated PMEMD integrated in AMBER14 package. For each simulation, a sophisticated protocol (minimization, heating, and equilibration) was followed. Initially, the prepared system was subjected to energy minimization (minimization, heating, and equilibration) was followed. Initially, a 50 ps NVT ensemble heating process was applied to gradually increase the temperature from 0 to 300 K and the density of the system to 1 g cm\textsuperscript{−3} in this step. Then, to equilibrate the system, three MD equilibrations of 500 ps at 300 K were performed with decreased restraint weights (5.0, 1.0 and 0.1 kcal mol\textsuperscript{−1} Å\textsuperscript{−2}). Finally, a 50 ns production run in the NPT ensemble was carried out without any restraint on the simulation system. During MD simulations, periodic boundary conditions were employed and the direct space interaction was calculated by using the particle-mesh Ewald (PME) method with a long range electrostatic interaction (cutoff = 12.0 Å). All bonds involving hydrogen atoms were constrained with the SHAKE algorithm allowing an integration time step of 2 fs.

**Thermodynamic analysis**

**Binding free energy calculation.** The mm_pbsa.pl script was used to calculate the binding free energy ($\Delta G_{\text{calc}}$) by molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) methods. The free energy perturbation (FEP) method could also be applied to perform qualitative predictions in this analysis. To be consistent with previous publications, MM/PBSA methods were adopted for analysis in this study. For each simulation system, a total of 500 snapshots were taken from the last equilibrium trajectory with equal intervals. And for each snapshot, $\Delta G_{\text{calc}}$ was calculated by:

$$\Delta G_{\text{calc}} = \Delta E_{\text{vdw}} + \Delta E_{\text{ele}} + \Delta G_{\text{PB}} + \Delta G_{\text{SA}} \quad (1)$$

where $\Delta E_{\text{vdw}}$ and $\Delta E_{\text{ele}}$ represent van der Waals and electrostatic energy changes in the gas phase. $\Delta G_{\text{PB}}$ and $\Delta G_{\text{SA}}$ represent polar and non-polar solvent interaction energy changes. $\Delta G_{\text{SA}}$ is calculated as the description of the PB equation, with dielectric constants for the solute and the solvent set to 1 and 80, respectively. While $\Delta G_{\text{SA}}$ is estimated by the solvent-accessible surface area (SASA) determined by a water probe radius of 1.4 Å and the surface tension constant $\gamma$ was set to 0.0072 kcal (mol Å\textsuperscript{−1})$^{-1}$ Å$^{-2}$.

**Per-residue binding free energy decomposition.** To quantitatively evaluate the contribution of each residue for drug’s binding, $\Delta G_{\text{calc}}$ was decomposed at a per-residue basis, including the van der Waals term ($\Delta G_{\text{per-residue}}^{\text{vdw}}$), the electrostatic term ($\Delta G_{\text{per-residue}}^{\text{ele}}$), the polar term ($\Delta G_{\text{per-residue}}^{\text{PB}}$) and the nonpolar term ($\Delta G_{\text{per-residue}}^{\text{SA}}$) for ligands and each residue, as shown in eqn (2):

$$\Delta G_{\text{calc}} = \Delta E_{\text{vdw}} + \Delta E_{\text{ele}} + \Delta G_{\text{PB}} + \Delta G_{\text{SA}}$$

$$\Delta G_{\text{per-residue}} = \Delta G_{\text{per-residue}}^{\text{vdw}} + \Delta G_{\text{per-residue}}^{\text{ele}} + \Delta G_{\text{per-residue}}^{\text{PB}} + \Delta G_{\text{per-residue}}^{\text{SA}} \quad (2)$$

**Computational alanine scanning (CAS).** CAS was performed to verify the key residues of IN and RNase H contributing significantly to JMC6F ($\geq 0.5$ kcal mol\textsuperscript{−1}). The whole process includes the generation of mutated snapshots and the calculation of the binding free energy difference between the WT and mutant (MUT) complex. Firstly, 500 snapshots were collected from the last 20 ns WT trajectory. The alanine

<table>
<thead>
<tr>
<th>Systems</th>
<th>Wide type</th>
<th>Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>IN-JMC6F</td>
<td>JMC6F docked into the HIV-1 IN model</td>
<td>T66A/I/K, E92Q, E138A/K, G140C/S, Y143C/H/R,</td>
</tr>
<tr>
<td>IN-DTG</td>
<td>DTG in the crystal structure (PDB ID: 3S3M)</td>
<td>S147G, Q148H/K/R, N153H</td>
</tr>
<tr>
<td>IN-RAL</td>
<td>RAL in the crystal structure (PDB ID: 3OYA)</td>
<td>T66A/I/K, E92Q, E138A/K, G140C/S, Y143C/H/R,</td>
</tr>
<tr>
<td>IN-EVG</td>
<td>EVG in the crystal structure (PDB ID: 3L2W)</td>
<td>S147G, Q148H/K/R, N153H</td>
</tr>
<tr>
<td>RNase H-JMC6F</td>
<td>JMC6F docked into the crystal structure (PDB ID: 3L0P)</td>
<td>T66A/I/K, E92Q, E138A/K, G140C/S, Y143C/H/R,</td>
</tr>
<tr>
<td>RNase H-MK1</td>
<td>Crystal structure (PDB ID: 3L0P)</td>
<td>Y501R, Y501W</td>
</tr>
</tbody>
</table>

* ND, not determined in this study.
mutation was generated by truncating the selected mutation residue at Cγ and by replacing Cγ with a hydrogen atom at a distance of 1.09 Å from Cβ along the direction of the Cγ-Cβ bond. And the topology files with alanine mutations were regenerated by the LEaP module in AMBER14 and the MM/PBSA method was used to calculate the relative binding free energy (ΔAGCAS) defined by the difference between WT and MUT complexes shown as below:

$$\Delta G_{\text{CAS}} = \Delta G_{\text{MUT}} - \Delta G_{\text{WT}}$$

where ΔGWt and ΔGMUT refer to the binding free energy of the WT and MUT complexes, respectively.

In silico site-directed mutagenesis

Collecting the reported drug resistant mutations in HIV-1 IN and RNase H. A total of 16 IN and 2 RNase H mutations were collected from the Stanford HIV Drug Resistance Database and the experiment previously reported by Arion et al. to validate the MD simulation models was used in this study. The resistance profile of JMC6F was further analyzed based on these previously reported resistance mutations.

In silico site-directed mutagenesis was performed to validate the accuracy of the constructed model and to predict the drug resistance profile of JMC6F. Based on the representative structure extracted from the equilibrated WT trajectories, a total of 48 mutant complexes (Table 1) were generated using PyMOL and prepared by LEAP in AMBER14. For each system, the MD simulation (10 ns) and binding free energy calculations were performed following the same way as described in previous sections.

Results and discussion

Modeled structure of the full-length HIV-1 IN

The resolving of HIV-1 intasome complexed with inhibitors has been hampered by many technical hurdles, such as the low solubility of IN. Herein, the homology model of HIV-1 IN was built based on the crystal structure of PFV IN. The sequence alignment of PFV/HIV-1 IN and viral U5 DNA is shown in Fig. S2 (ESI†). From the sequence alignment, the overall identity between HIV-1 IN and PFV IN was up to 50%, showing the high degree of homology and highly conserved DDE motif (Asp64, Asp116 and Glu152). The homology model of HIV-1 IN represents a strand transfer conformation for inhibitor binding. The overall architecture of the HIV-1 IN model comprised three domains including NTD (residues 1–46), CCD (residues 56–202) and CTD (residues 220–272) connected with each other by linkers (Fig. 2A).

The stereochemical quality of the model was evaluated using Ramachandran plot analysis in PROCHECK. As shown in Fig. S3 (ESI†), 98.2% residues were found in the allowed regions, and 83.5% were in the favored regions indicating that the homology mode was reliable.

Initial poses of JMC6F in the binding sites of HIV-1 IN and RNase H

The inhibitor JMC6F was found active to inhibit both HIV-1 IN and RT associated RNase H. Therefore, JMC6F was docked into the INSTI binding site of HIV-1 IN and RT associated RNase H. The docking poses of JMC6F in HIV-1 IN and RNase H were similar to the co-crystal structures of DTG in PFV IN and MK1 in HIV-1 RNase H, respectively. Since JMC6F, DTG and MK1 shared a similar structure portion (Fig. 1) chelation occurs with two Mg2+ ions at the IN and RNase H active site. The superposition of poses of JMC6F with DTG in PFV IN (PDB ID: 3S3M) and MK1 in HIV-1RT (PDB ID: 3LP0) is displayed in Fig. 3. As shown in Fig. 3, JMC6F occupied the active sites of HIV-1 IN and RNase H with the DKA moiety chelating with Mg2+. The initial poses of JMC6F presented in docking could render a prediction of protein–ligand interactions. The initial pose obtained from the docking procedure was subjected to MD simulation to explore the detailed structural and energetic properties for protein–ligand recognition.

MD simulation and stability of the systems. For each system, 50 ns explicit solvent MD simulation was carried, and the fluctuation of simulation trajectory was monitored by root-mean-square deviation (RMSD) relative to the initial structure. The RMSD plots of backbone heavy atoms of proteins and vDNA, heavy atoms of ligands and backbone atoms of the binding site (the residues within 5 Å around the ligand) for each system are shown in Fig. S4 (ESI†).
the RMSD values of the binding site and ligand were within ~2 Å and ~1.5 Å respectively, while the RMSD values of the overall protein and vDNA were within ~8 Å and ~6 Å. In order to further depict the changes of conformation, RMSFs of the HIV-1 integrase in the simulation trajectory of the JM6C6F-IN system were calculated as shown in Fig. S5 (ESI†). And this indicates that the fluctuation was rather higher in the flexible loop regions (residues 15–48, residues 60–72, residues 135–148, residues 200–250), especially the loop region (residues 140–149) at the binding site, which is required for IN function as described previously in a reported experiment.64 By combining RMSD and RMSF analyses, it was observed that the large fluctuation of the overall protein was caused by the flexible linkers65 in HIV-1 IN as shown in Fig. S5 (ESI†). For bound HIV-1 RNase H complexes (Fig. S4E and F, ESI†), the overall proteins, binding sites and ligands showed low fluctuations within a RMSD value of ~3 Å. In summary, the results indicated that 50 ns MD simulation was sufficient to obtain an equilibrated simulation trajectory. Thus, the last 20 ns equilibrated trajectory was used for the following structural and energetic analysis.

Validation of the models constructed by MD simulation

Correlation of the binding affinity between the simulation and experimental results. Based on the 500 snapshots extracted from the last 20ns MD simulation, the MM/PBSA method26 was used to analyze the binding free energy (ΔGcalc) and the results are listed in Table 2. The ΔGcalc for the six WT systems were −21.89 kcal mol⁻¹ (IN-JM6C6F), −17.30 kcal mol⁻¹ (RNase H-JM6C6F), −17.50 kcal mol⁻¹ (RNase H-MK1), −32.83 kcal mol⁻¹ (IN-DTG), −31.34 kcal mol⁻¹ (IN-RAL) and −28.22 kcal mol⁻¹ (IN-EVG). And the detailed energy components of the binding free energy are shown in Table S2 (ESI†). Both van der Waals and electrostatic contributions play key roles in JM6C6F binding with IN and RNase H. And it is interesting to note that the electronic contribution is higher when JM6C6F binds with IN, while the van der Waals contribution is lower when JM6C6F binds with RNase H, which may contribute to the balanced binding affinity of JM6C6F against HIV-1 IN and RNase H (ratio ~ 1).28 In addition, ΔGexp = RTlnIC50 was used to convert the experimental IC50 values25,26,29 into the binding free energies (ΔGexp). As shown in Fig. S6A (ESI†), the relative binding free energy of the ΔGcalc correlated very well (R² = 0.95) with the experimental value (ΔGexp). The good correlation coefficient between ΔGcalc and ΔGexp can be good evidence for verifying the resulting model from MD simulation.47

Resistence profile identified by in silico site-directed mutagenesis analysis. To further validate the model constructed in the study, the resistance profile associated with 16 single point mutations of the approved INSTIs (RAL, EVG and DTG) (Table 1) was investigated by in silico site mutagenesis analysis. It is noted that the resistance profile of the 16 mutations from experimental results45 was reproduced well (R² = 0.85) by the calculated ΔGcalc (Fig. S6B and Table S3, ESI†). More detailed information of the binding affinity is listed in Tables S4–S6 (ESI†). As the in silico site-directed mutagenesis analysis was based on the WT model from MD simulation in this work, the good correlation further indicated that the resulting models from MD simulation were capable of predicting the mutation-induced drug resistance profile.47

Comparison of MD refined and cryo-EM solved HIV-1 IN structures. It is important to indicate that the cryo-EM structure of the core tetrameric HIV-1 strand transfer complex (STC) binding without inhibitors was solved40 at the end of this work. Solving the HIV-1 IN cryo-EM structure provided an opportunity for validating the model from our MD simulation. Thus, a structure comparison between the refined HIV-1 IN and the cryo-EM HIV-1 IN was conducted. As shown in Fig. S7 (ESI†), the overall distribution of NTD, CCD, CTD and the linkers matched pretty well. The RMSDs of the backbone atoms for NTD, CTD, CCD and the residues at the binding site between the MD refined and cryo-EM solved HIV-1 IN structures were 3.21 Å, 2.26 Å, 3.08 Å and 1.40 Å, respectively. These suggested that the HIV-1 IN model was reliable.

**Table 2.** The calculated and experimentally estimated binding free energies of JMC6F, DTG, EVG, RAL and MK1 binding to the wide type HIV-1 IN and RNase H, respectively. (ΔG is in kcal mol⁻¹ and the IC50 value is in μM).

<table>
<thead>
<tr>
<th>Drugs</th>
<th>Targets</th>
<th>ΔGcalc</th>
<th>ΔGexp</th>
<th>ΔGcalc</th>
<th>ΔGexp</th>
<th>ΔIC50</th>
<th>IC50</th>
</tr>
</thead>
<tbody>
<tr>
<td>JMC6F</td>
<td>IN</td>
<td>−21.89</td>
<td>−8.08</td>
<td>3.62</td>
<td>1.2000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RNase H</td>
<td>−17.30</td>
<td>−7.84</td>
<td>3.86</td>
<td>1.8000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MK1</td>
<td>RNase H</td>
<td>−17.50</td>
<td>−7.58</td>
<td>4.12</td>
<td>2.8000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DTG</td>
<td>IN</td>
<td>−32.83</td>
<td>−11.70</td>
<td>0.00</td>
<td>0.0007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EVG</td>
<td>IN</td>
<td>−28.22</td>
<td>−11.20</td>
<td>0.50</td>
<td>0.0060</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAL</td>
<td>IN</td>
<td>−31.34</td>
<td>−11.60</td>
<td>0.10</td>
<td>0.0033</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Calculated binding free energy by the MM/PBSA method in this work. b Binding free energy variation ΔGcalc values were calculated using the IN-DTG complex as a reference. c Estimated binding free energy ΔGexp based on IC50 values using ΔGexp = RTlnIC50. d Experimental IC50 value from the reported work in ref. 32, 38, 94 and 95.

**Paper**

Published on 12 June 2018. Downloaded by Zhejiang University on 4/24/2020 2:56:45 PM.

**PCCP**

This journal is © the Owner Societies 2018

Identification and verification of the key residues contributing to JMC6F inhibition. Based on equilibrated MD simulation trajectories, the energy contribution of each residue for the inhibition of JMC6F was calculated. As shown in Fig. 4A2 and B2, residues Asp64, Thr66, Leu68, Asp116, Tyr143, Gln148 and Glu152 at the INSTI binding site of HIV-1 IN and residues Asp443, Gly444, Trp535, Lys540 and Asp549 at the RHI binding site of HIV-1 RNase H were identified as key residues ($Z_{0.5}$ kcal mol$^{-1}$) for the inhibition of JMC6F. The computational alanine scanning (CAS) method has been widely applied to verify the reliability of key residues identified in protein–ligand complexes by MD simulations. Therefore, the identified seven HIV-1 IN residues (Asp64, Thr66, Asp116, Thr143, Gln148, Glu152 and Lys159) and five HIV-1 RNase H residues (Asp443, Glu478, Trp535, Lys540 and Asp549) were further analyzed by CAS calculations (Fig. 5 and Table S7, ESI†). By comparing Fig. 4 and 5, it was obvious that the per-residue binding free energy and CAS results were consistent with each other. Besides, as shown in Table S7 (ESI†) and Fig. 5, in all the simulated complexes, residues Asp64, Leu68, Asp116, Thr143, Lys159, Asp443, Glu478 and Asp549 were found to be hotspots for JMC6F.

Shared binding modes of JMC6F in HIV-1 IN and RNase H
The extracted structures as well as the identified key residues inferred the similarity of the binding mode of JMC6F with
HIV-1 IN and RNase H. The similar binding mode shared by the two complexes was then generalized and is schematically represented in Fig. 6. As illustrated, the shared binding mode could be characterized with two pharmacophores: the diketo acid group chelated with Mg$^{2+}$ ions at the enzyme catalytic sites which was coordinated by the DDE motif (site I) of IN and RNase H, while the pyrrole substitutions were exposed to a hydrophobic cavity (site II and site III) out of the catalytic sites. Compared with the binding site of IN, the RNase H binding site is largely hydrophobic, which apparently affects the interactions of JMC6F with the two targets.

**Resistance profile of JMC6F identified based on the available mutations**

Though three INSTIs (RAL, EVG and DTG) of HIV-1 IN were approved by the U.S. FDA, the clinical drug resistance as well as the cross resistance between these drugs due to the emergence of mutations could not be ignored. Moreover, combination of G140S/Q148H/N155H substitutions was associated with the reduced susceptibility to the second-generation inhibitor DTG. The major goal of developing dual inhibitors was employed to predict sensitivity of JMC6F based on known mutations. In silico site-directed mutagenesis analysis of MD simulations based on binding free energy ($\Delta G_{\text{cal}}$) calculations was used to predict sensitivity of JMC6F based on known major single mutations (Table 1). The resistance profile of certain residues towards inhibitors can be reflected by the relative binding free energy ($\Delta \Delta G_{\text{cal}}$) before and after *in silico* mutation on the corresponding residues. However, the sensitivity of JMC6F based on the available mutations remains not well understood.

In this study, *in silico* site-directed mutagenesis analysis of MD simulations based on binding free energy ($\Delta G_{\text{cal}}$) calculations was used to predict sensitivity of JMC6F based on known major single mutations (Table 1). The resistance profile of certain residues towards inhibitors can be reflected by the relative binding free energy ($\Delta \Delta G_{\text{cal}}$) before and after *in silico* mutation on the corresponding residues. However, the sensitivity of JMC6F based on the available mutations remains not well understood.

In this study, *in silico* site-directed mutagenesis analysis of MD simulations based on binding free energy ($\Delta G_{\text{cal}}$) calculations was used to predict sensitivity of JMC6F based on known major single mutations (Table 1). The resistance profile of certain residues towards inhibitors can be reflected by the relative binding free energy ($\Delta \Delta G_{\text{cal}}$) before and after *in silico* mutation on the corresponding residues. However, the sensitivity of JMC6F based on the available mutations remains not well understood.

In this study, *in silico* site-directed mutagenesis analysis of MD simulations based on binding free energy ($\Delta G_{\text{cal}}$) calculations was used to predict sensitivity of JMC6F based on known major single mutations (Table 1). The resistance profile of certain residues towards inhibitors can be reflected by the relative binding free energy ($\Delta \Delta G_{\text{cal}}$) before and after *in silico* mutation on the corresponding residues. However, the sensitivity of JMC6F based on the available mutations remains not well understood.

In this study, *in silico* site-directed mutagenesis analysis of MD simulations based on binding free energy ($\Delta G_{\text{cal}}$) calculations was used to predict sensitivity of JMC6F based on known major single mutations (Table 1). The resistance profile of certain residues towards inhibitors can be reflected by the relative binding free energy ($\Delta \Delta G_{\text{cal}}$) before and after *in silico* mutation on the corresponding residues. However, the sensitivity of JMC6F based on the available mutations remains not well understood.

In this study, *in silico* site-directed mutagenesis analysis of MD simulations based on binding free energy ($\Delta G_{\text{cal}}$) calculations was used to predict sensitivity of JMC6F based on known major single mutations (Table 1). The resistance profile of certain residues towards inhibitors can be reflected by the relative binding free energy ($\Delta \Delta G_{\text{cal}}$) before and after *in silico* mutation on the corresponding residues. However, the sensitivity of JMC6F based on the available mutations remains not well understood.

In this study, *in silico* site-directed mutagenesis analysis of MD simulations based on binding free energy ($\Delta G_{\text{cal}}$) calculations was used to predict sensitivity of JMC6F based on known major single mutations (Table 1). The resistance profile of certain residues towards inhibitors can be reflected by the relative binding free energy ($\Delta \Delta G_{\text{cal}}$) before and after *in silico* mutation on the corresponding residues. However, the sensitivity of JMC6F based on the available mutations remains not well understood.

In this study, *in silico* site-directed mutagenesis analysis of MD simulations based on binding free energy ($\Delta G_{\text{cal}}$) calculations was used to predict sensitivity of JMC6F based on known major single mutations (Table 1). The resistance profile of certain residues towards inhibitors can be reflected by the relative binding free energy ($\Delta \Delta G_{\text{cal}}$) before and after *in silico* mutation on the corresponding residues. However, the sensitivity of JMC6F based on the available mutations remains not well understood.

In this study, *in silico* site-directed mutagenesis analysis of MD simulations based on binding free energy ($\Delta G_{\text{cal}}$) calculations was used to predict sensitivity of JMC6F based on known major single mutations (Table 1). The resistance profile of certain residues towards inhibitors can be reflected by the relative binding free energy ($\Delta \Delta G_{\text{cal}}$) before and after *in silico* mutation on the corresponding residues. However, the sensitivity of JMC6F based on the available mutations remains not well understood.
respectively. The relatively large decrease of binding free energy induced by Y501W and Y501R mutations implies their potential resistance to JMC6F.

**Conclusion**

In the present study, the binding mode and resistance profile of a novel dual inhibitor JMC6F targeting HIV-1 IN and RNase H were analyzed by computational methods. The models from MD simulations were verified by the good correlation ($R^2 = 0.95$) of the relative binding free energy between calculated and experimental values, the good reproducibility of the in silico site-directed mutagenesis study ($R^2 = 0.85$) and the structural comparison of the model with the very recently reported HIV-1 IN cryo-EM. By using integrated computational methods, seven residues [Asp64, Thr66, Asp116, Thr143, Gln148, Glu152 and Lys159] of HIV-1 IN and five residues (Asp443, Glu478, Trp535, Lys540 and Asp549) of HIV-1 RNase H were identified as key residues which contributed to the inhibition of JMC6F binding. Thus, a shared binding mode of JMC6F with IN and RNase H was summarized in terms of the diketo acid group chelating with Mg$^{2+}$ ions at the enzyme catalytic sites which was coordinated by the DDE motif (site I) of IN and RNase H, while the pyrrole substitutions were exposed to a hydrophobic cavity (site II and site III) out of the catalytic sites. The drug resistance profile of JMC6F based on the currently known major single mutations in HIV-1 IN and RNase H suggested only three mutations in HIV-1 IN (Y143C, Q148R and N155H) and the key to overcoming drug resistance is to design optimized ligands with substituted groups in the pyrrolyl ring which will strengthen the interaction with residues at the binding site. Finally, two mutations in HIV-1RNase H (Y501R and Y501W) lead to a reduction in JMC6F potency, implying their potential role in providing resistance to JMC6F. The observations obtained in this study will be useful for the structure-based design of more efficient HIV-1 IN and RNase H dual targeting inhibitors.
Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (21505009), the Precision Medicine Project of the National Key Research and Development Plan of China (2016YFC0902200), the Innovation Project on Industrial Generic Key Technologies of Chongqing (cstc2015zdcy-ztzx120003) and the Fundamental Research Funds for Central Universities (10611CDJXZ238826, CDJZR14468801, and CDJKXB14011).

References
