



Cite this: *Phys. Chem. Chem. Phys.*,
2019, 21, 12931

Insight into the selective binding mechanism of DNMT1 and DNMT3A inhibitors: a molecular simulation study†

Tianli Xie,^{‡a} Jie Yu,^{‡a} Weitao Fu,^a Zhe Wang,^a Lei Xu,^b Shan Chang,^{ib}
Ercheng Wang,^a Feng Zhu,^{ib} Su Zeng,^a Yu Kang,^{ib}*^a and Tingjun Hou,^{ib}*^{ac}

DNA methyltransferases (DNMTs), responsible for the regulation of DNA methylation, have been regarded as promising drug targets for cancer therapy. However, high structural conservation of the catalytic domains of DNMTs poses a big challenge to design selective inhibitors for a specific DNMT isoform. In this study, molecular dynamics (MD) simulations, end-point free energy calculations and umbrella sampling (US) simulations were performed to reveal the molecular basis of the binding selectivity of three representative DNMT inhibitors towards DNMT1 and DNMT3A, including SFG (DNMT1 and DNMT3A dual inhibitors), DC-05 (DNMT1 selective inhibitor) and GSKex1 (DNMT3A selective inhibitor). The binding selectivity of the studied inhibitors reported in previous experiments is reproduced by the MD simulation and binding free energy prediction. The simulation results also suggest that the driving force to determine the binding selectivity of the studied inhibitors stems from the difference in the protein–inhibitor van der Waals interactions. Meanwhile, the per-residue free energy decomposition reveals that the contributions from several non-conserved residues in the binding pocket of DNMT1/DNMT3A, especially Val1580/Trp893, Asn1578/Arg891 and Met1169/Val665, are the key factors responsible for the binding selectivity of DNMT inhibitors. In addition, the binding preference of the studied inhibitors was further validated by the potentials of mean force predicted by the US simulations. This study will provide valuable information for the rational design of novel selective inhibitors targeting DNMT1 and DNMT3A.

Received 10th April 2019,
Accepted 16th May 2019

DOI: 10.1039/c9cp02024a

rsc.li/pccp

Introduction

DNA methylation is a major epigenetic modification that regulates gene expression in the genome of higher eukaryotes. In addition to genetic alterations, DNA methylation can serve as a secondary mechanism for the inactivation of tumor suppressor genes (TSGs).¹ In many types of cancer cells, hypermethylation of promoter CpG islands can turn off the expression of TSGs.^{2,3} Unlike genetic alterations, epigenetic events, such as DNA

methylation, are reversible. Therefore, removing aberrant methylation can restore the TSG expression⁴ and then revert cancer cells to their normal states, which provides possibility to develop new strategies for cancer therapy.⁵

DNA methyltransferases (DNMTs) are a family of enzymes that catalyze the addition of a methyl group to DNA. DNMT, DNA and its cofactor *S*-adenosyl-*L*-methionine (SAM) are the three key components in DNA methylation. First, DNMT forms a complex with DNA and the deoxycytidine flips into the catalytic pocket

^a College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, Zhejiang, China. E-mail: yukang@zju.edu.cn, tingjunhou@zju.edu.cn

^b Institute of Bioinformatics and Medical Engineering, School of Electrical and Information Engineering, Jiangsu University of Technology, Changzhou 213001, Jiangsu, China

^c State Key Lab of CAD&CG, Zhejiang University, Hangzhou 310058, Zhejiang, China

† Electronic supplementary information (ESI) available: Table S1. MM/GBSA free energy calculation for the 26 complexes based on the 5 ns MD simulations, Fig. S1. The best binding poses of GSKex1 predicted by the Glide SP/XP docking compared with that of DNMT1/SFG (PDB entry: 3SWR). SFG is colored blue, and the best binding poses of GSKex1 predicted by the Glide XP and SP are colored red and green, respectively. The binding pose of GSKex1 on DNMT3A predicted by the Glide XP is colored magenta, Fig. S2. The RMSDs of the C_α atoms of the protein in the six studied complexes as a function of simulation time, Fig. S3. The RMSDs of the heavy atoms of the ligands in the six studied complexes as a function of simulation time, Fig. S4. Residue-based RMSFs relative to the initial structure during the 150 ns MD simulations, Fig. S5. Comparison of the representative structures of DNMT1/DC-05 and DNMT3A/GSKex1 with the lowest RMSDs relative to the averaged structures extracted from the 150 ns MD trajectories. DNMT1 is colored green, and DNMT3A is colored magenta. The carbon backbone of DC-05 is colored yellow, Fig. S6. Comparison of the representative structures of DNMT1/DC-05 and DNMT3A/GSKex1 with the lowest RMSDs relative to the averaged structures extracted from the 150 ns MD trajectories. DNMT1 is colored green, and DNMT3A is colored magenta. The carbon backbone of GSKex1 is colored yellow, Fig. S7. Convergence of the PMFs given by the US simulations for (A) DNMT1/SFG, (B) DNMT1/DC-05, (C) DNMT1/GSKex1, (D) DNMT3A/SFG, (E) DNMT3A/DC-05, and (F) DNMT3A/GSKex1. See DOI: 10.1039/c9cp02024a

‡ Equivalent authors.

of DNMT. Then, DNMT catalyzes the transfer of a methyl group from SAM to the C5-cytosine.⁶ In mammals, four DNMT isoforms, including DNMT1, DNMT3A, DNMT3B and DNMT3L, have been identified.⁷ DNMT1 is the most abundant isoform, and it is responsible for the maintenance of the existing methylation patterns after DNA replication.² DNMT3A and DNMT3B are two *de novo* methyltransferases that establish methylation patterns during mammalian embryonic development, with the help of the stimulatory factor DNMT3L.⁸ The accumulated experimental data show that abnormal DNA methylation is associated with many cancers, such as acute myelogenous leukemia (AML),⁹ breast cancer,¹⁰ hepatocellular carcinoma,¹¹ and so on. Specific inhibition of gene promoter hypermethylation by DNMT inhibitors (DNMTi) has been validated as an attractive and novel approach for cancer therapy.^{12,13} However, it is still obscure which DNMT isoform would be the best therapeutic target because the reported outcomes are heavily dependent on the pathologies.^{14,15} For example, it has been reported that in glioblastoma DNMT1 is found to be overexpressed;¹⁶ in certain leukemias¹⁷ and melanomas,¹⁸ DNMT3A is overexpressed or mutated, while in pancreatic cancer, both DNMT1 and DNMT3A are highly overexpressed.¹⁹

Over the past two decades, two types of DNMTi have been discovered, namely nucleoside derivatives and non-nucleoside compounds. Two nucleoside derivatives, *i.e.*, 5-azacytidine (azacitidine) and 5-aza-2'-deoxycytidine (decitabine), have been approved by the US Food and Drug Administration (FDA) for the treatment of myelodysplastic syndromes (MDS), AML and chronic myelomonocytic leukemia (CMML).²⁰ However, such nucleoside drugs need to be incorporated into DNA through the covalent modification of azanucleosides, leading to poor chemical stability, low specificity and high toxicity.¹⁵ To overcome these limitations, extensive efforts have been dedicated to the discovery and development of non-nucleoside inhibitors such as SGI-1027,²¹ RG108,³ DC-05,²² *etc.* Although a number of non-nucleoside inhibitors have been reported, most of them are not quite promising because of weak inhibition activity, low specificity or lack of capability to induce DNA methylation at the cellular level. Until now, none of the non-nucleoside DNMTi has been approved by the FDA.^{23,24} Therefore, discovery of

novel non-nucleoside DNMTi with specific selectivity and high potency is quite urgent. Structure-based investigations on the mechanism of the protein–ligand recognition at the molecular level are thus promising to provide valuable information for the rational design of selective DNMTi.

DNMT1 and DNMT3A share a similar domain architecture with two functional components, an N-terminal regulatory domain and a C-terminal catalytic domain. The N-terminal domains of DNMTs play central roles in nucleosome recognition where that of DNMT1 contains several subdomains including an independently folded domain (NTD), a replication foci-targeting sequence (RFTS) domain, a zinc finger with the Cys–X–X–Cys (CXXC) motif, and two bromo adjacent homology (BAH1 and BAH2) domains.^{25–27} The N-terminal domain of DNMT3A is comparatively smaller and composed of only two defined subdomains namely the Pro–Trp–Trp–Pro (PWWP) domain and the ATRX–DNMT3–DNMT3L (ADD) domain.²⁸ The C-terminal domains of DNMTs involved in cofactor binding and substrate catalysis are highly conserved,¹⁵ which brings a big challenge to design selective inhibitors towards a specific DNMT isoform.

The crystal structures of DNMT1^{25–27} and DNMT3A²⁸ offer the possibility of exploring the binding selectivity of inhibitors towards different DNMTs by molecular simulation techniques, which may improve the success rate of designing selective DNMTi. Therefore, in this study, the binding selectivity of three representative non-nucleoside DNMTi, including SFG, DC-05 and GSKex1 (Fig. 1), towards DNMT1 and DNMT3A was studied by molecular dynamics (MD) simulations, Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) free energy calculations and umbrella sampling (US) simulations. SFG (sinefungin) is a DNMT1/DNMT3A dual inhibitor with $IC_{50} = 7.4 \mu\text{M}$ towards DNMT1 and $IC_{50} = 1.8 \mu\text{M}$ towards DNMT3A.^{29,30} As a SAM-competitive inhibitor, SFG is structurally similar to SAM, which has been widely used as a reference inhibitor of both DNMT1³⁰ and DNMT3A.³¹ DC-05 is a DNMT1 selective inhibitor with $IC_{50} = 10.3 \mu\text{M}$ towards DNMT1 and $IC_{50} > 200 \mu\text{M}$ towards DNMT3A. DC-05 was discovered through docking-based virtual screening by Chen *et al.*,²² and it has a much higher activity towards DNMT1 than DNMT3A and several other

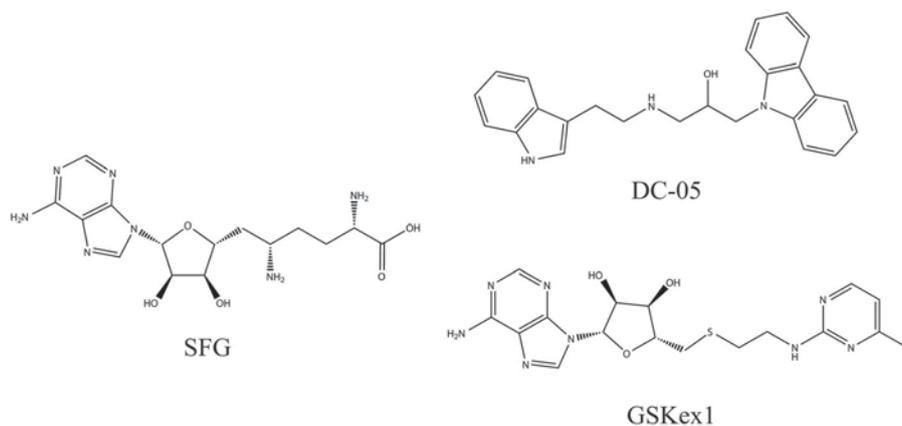


Fig. 1 2D structures of the three studied inhibitors (SFG, DC-05 and GSKex1).

methyltransferases. GSKex1 is a DNMT3A selective inhibitor patented by GlaxoSmithKline Intellectual Property Development Limited with $IC_{50} = 300 \mu\text{M}$ towards DNMT1 and $IC_{50} = 3 \mu\text{M}$ towards DNMT3A.³²

Materials and methods

Initial structures and system preparations

The initial crystal structures of DNMT1 in complex with SFG (PDB entry: 3SWR, unpublished data) and DNMT3A (PDB entry: 4U7T²⁸) were retrieved from the RCSB Protein Data Bank (PDB). The missing side chains and loop structures of the proteins were added by using the Loops/Refine Structure module in the UCSF Chimera package.³³ A total of six protein/inhibitor complexes, including DNMT1/SFG, DNMT1/DC-05, DNMT1/GSKex1, DNMT3A/SFG, DNMT3A/DC-05 and DNMT3A/GSKex1, were prepared for the MD simulations. The structure of the DNMT1/SFG complex was directly obtained from the PDB, and those of the other five complexes were predicted by molecular docking. It should be noted that DC-05 contains a chiral center, but the chirality effect on the binding affinity is not obvious and (*R*)-DC-05 ($IC_{50} = 7.3 \mu\text{M}$) is only slightly more active than (*S*)-DC-05 ($IC_{50} = 15.6 \mu\text{M}$).²² Moreover, the predicted binding poses of the (*R*)- and (*S*)-enantiomers within DNMT1 are quite similar.²² In this study, only (*R*)-DC-05 was used in the simulations.

The *Glide* module in Schrödinger (version 11.1)³⁴ was used to predict the binding pose of each inhibitor in the SAM site of DNMT1 or DNMT3A. The two proteins were prepared by using the *Protein Preparation Wizard* in Schrödinger,³⁵ including removing all non-bonded hetero-atoms and water molecules, adding missing hydrogen atoms, and optimizing the structure to relieve steric clashes using the OPLS3 force field.³⁶ The scoring grid for docking was generated by enclosing the residues in a box with a size of $15 \text{ \AA} \times 15 \text{ \AA} \times 15 \text{ \AA}$ centered on the centroid of SAM using the *Receptor Grid Generation* module with default settings in Schrödinger. The prepared inhibitors were then docked by *Glide* with the extra precision (XP) scoring mode. A previous study shows that combination of short MD simulations and MM/GBSA is effective to determine the near-native binding conformations.³⁷ Therefore, in this study, for each protein–ligand complex, the top five scored binding poses were saved, and then rescored by MM/GBSA based on the MD simulations. First, 5 ns MD simulations were run for each protein–ligand complex, and then the MM/GBSA calculations based on the 500 snapshots extracted from the 3–5 ns MD trajectory were performed to estimate the binding affinity for each protein–ligand complex (the details about MM/GBSA can be found in the following sections). Finally, the binding pose with the lowest binding free energy was determined as the best pose for each inhibitor in DNMT1 or DNMT3A.

Conventional molecular dynamics (MD) simulations

Prior to MD simulations, the pK_a values of the ionizable residues were calculated using PROPKA,^{38,39} and the protonated states of the ionizable residues in each system were assigned using the

PDB2PQR software (version 2.1.0).⁴⁰ Then, the partial charges of the three inhibitors (SFG, DC-05, and GSKex1) were generated by the restrained electrostatic potential (RESP) fitting technique based on the electrostatic potentials computed by Gaussian09 at the Hartree–Fock (HF) SCF/6-31G* level of theory.^{38,41,42} Afterward, the missing force field parameters for the studied inhibitors were generated using the *antechamber* module in AMBER16.^{43,44} Finally, the topology and parameter files of the studied protein–inhibitor complexes were generated using the *tleap* module in AMBER16. The general AMBER force field (GAFF),⁴⁵ the ff14SB force field⁴⁶ and the Zinc AMBER force field (ZAFF) were used for the inhibitors, proteins and Zn^{2+} , respectively.⁴⁷ Each protein–inhibitor complex was solvated in a periodic box filled with the TIP3P water molecules with a hydration shell of 15 \AA .^{48,49} In order to achieve the electroneutrality of each protein–ligand system, 16 chloride ions for the DNMT1 system and 6 sodium ions for the DNMT3A system were added.

The molecular mechanics (MM) minimizations and MD simulations were conducted using the *sander* and *pmemd* modules in AMBER16, respectively. The Particle Mesh Ewald (PME) algorithm was used to compute the long-range electrostatic interactions under periodic boundary conditions, and a cutoff of 10 \AA was used for the real-space interactions.⁵⁰ The minimizations of the initial structures were divided into three stages. First, the non-hydrogen atoms of the protein were restrained with a force constant of $5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, and the structure was optimized by 2500 cycles of the steepest descent and 2500 cycles of conjugate gradient minimizations. Then, the protein C_α and the ligand atoms were restrained with a force constant of $4 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, and each system was optimized by 2500 cycles of the steepest descent and 2500 cycles of conjugate gradient minimizations. Finally, all the atoms were optimized by 5000 cycles of minimizations (2500 cycles of the steepest descent and 2500 cycles of the conjugate gradient) without any restraint.

After minimizations, each system was heated to 300 K at a constant volume over a period of 100 ps. Subsequently, the whole system was equilibrated over 100 ps in the *NPT* ($T = 300 \text{ K}$ and $P = 1 \text{ bar}$) ensemble using a Langevin thermostat. Finally, 150 ns MD simulations were carried out in the *NPT* ($T = 300 \text{ K}$ and $P = 1 \text{ bar}$) ensemble. The SHAKE algorithm was used to constrain the covalent bonds involving hydrogen atoms and the time step was set to 2 fs.^{51,52} The coordinates were saved every 2 ps for the subsequent analysis.

MM/GBSA free energy calculations and decomposition

The binding free energy (ΔG_{bind}) between the receptor and the ligand in each system was calculated by the MM/GBSA methodology according to the following equations:^{53–58}

$$\begin{aligned} \Delta G_{\text{bind}} &= G_{\text{complex}} - (G_{\text{receptor}} + G_{\text{ligand}}) \\ &= \Delta E_{\text{MM}} + \Delta G_{\text{solvation}} - T\Delta S \quad (1) \\ &= \Delta E_{\text{MM}} + \Delta G_{\text{GB}} + \Delta G_{\text{SA}} - T\Delta S \end{aligned}$$

where G_{complex} , G_{receptor} , and G_{ligand} represent the free energies of the complex, the receptor and the ligand, respectively;

ΔE_{MM} is the gas-phase interaction energy between the protein and the ligand, including the electrostatic (ΔE_{elec}) and van der Waals interactions (ΔE_{vdW}); ΔG_{GB} and ΔG_{SA} are the polar and nonpolar parts of the desolvation free energy, respectively. ΔG_{GB} was calculated using the modified generalized Born (GB) model developed by Onufriev *et al.* ($\text{igb} = 2$);⁵⁹ ΔG_{SA} was computed based on the solvent-accessible surface area (SASA) estimated by a fast linear combination of the pairwise overlap (LCPO) algorithm using a probe radius of 1.4 Å: $\Delta G_{\text{SA}} = 0.0072 \times \Delta \text{SASA}$;⁶⁰ and $-T\Delta S$ is the change of the conformational entropy upon ligand binding, which was not considered here due to expensive computational cost and low prediction accuracy.⁶¹ In the ΔG_{GB} calculations, the solvent and solute dielectric constants were set to 80 and 1, respectively. 1000 snapshots evenly extracted from the equilibrated 50 ns trajectories of each system were used for the binding free energy calculation with the *MMPBSA.py* module in AMBER16.⁶²

Residue free energy decomposition was performed to identify the key residues responsible for ligand binding by splitting the total free energy into the energy contributions from individual residue–ligand pairs ($\Delta G_{\text{residue-ligand}}$).^{63,64} The residue–inhibitor interaction has four terms, including ΔE_{elec} , ΔE_{vdW} , ΔG_{GB} and ΔG_{SA} . Except for the ΔG_{SA} term, which was calculated based on the SASA using the ICOSA algorithm,⁶³ the other terms were calculated based on the same parameters used in the binding free energy calculations.

Analysis of MD trajectories

The MD trajectories were post-processed and analyzed by the *cpptraj* module in Amber16. The root mean square deviation (RMSD) of the C_{α} atoms with respect to the initial structure as a function of time was computed to monitor the stability of the MD simulations. The root mean square fluctuation (RMSF) of the atomic positions after fitting to a reference structure as a function of residue number was computed to monitor local conformational flexibility. The following criteria were used to identify hydrogen bonds (H-bonds): the distance between the hydrogen donor and acceptor heavy atoms < 3.5 Å and the donor–hydrogen–acceptor angle $> 135^{\circ}$.

Principal components analysis (PCA) transforms a series of potentially coordinated observations into a set of orthogonal vectors called principal components (PCs), and it is an effective tool to explore the essential motion governing the conformation transition during the simulations.^{65,66} The input to PCA will be a coordinate covariance matrix. To obtain an appropriate trajectory matrix, the overall translation or rotation motion was removed by fitting the coordinate data to the average structure. Then, the fitted trajectory data were utilized to generate a positional covariance matrix between the C_{α} atoms of any two residues, which is defined in eqn (2):⁶⁷

$$\sigma_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \quad (i, j = 1, 2, 3, \dots, 3N) \quad (2)$$

where $x_i(x_j)$ is the Cartesian coordinate of the i th(j th) C_{α} atom, $\langle x_i \rangle$ or $\langle x_j \rangle$ denotes the time average over all sampled conformations, and N represents the number of the C_{α} atoms considered. The symmetrical covariance matrix σ is diagonalized to produce

eigenvectors γ_n , namely the principal component PC_n , and the corresponding eigenvalues λ_n . The eigenvectors γ_n stand for the directions of atomic motions that are independent of each other in the multidimensional space, and the eigenvalues λ_n describe the corresponding magnitude. The γ_n and λ_n are arranged in a descending order so that λ_1 represents the largest eigenvalue. Significantly, the first several eigenvectors of σ are sufficient to qualitatively describe the large amplitude conformational changes for most cases.

The dynamic correlative motions between any pair of residues were explored by the dynamic cross-correlation map (DCCM) analysis for the C_{α} atoms. The cross-correlation coefficient C_{ij} between the C_{α} atoms of the i th and j th residues defined by eqn (3) was used to measure the correlation of their atomic fluctuations relative to their average positions:⁶⁸

$$C_{ij} = \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{\sqrt{\langle \Delta r_i \cdot \Delta r_i \rangle \langle \Delta r_j \cdot \Delta r_j \rangle}} \quad (3)$$

where Δr_i and Δr_j represent the displacement vectors of the i th and j th residues, respectively. The angle bracket $\langle \dots \rangle$ denotes the time average over the trajectory. C_{ij} ranges from -1 to 1 . A positive value of C_{ij} suggests that the two correlated residues move in the same direction, and a negative value suggests that the two anti-correlated residues move in opposite directions.

Umbrella sampling simulations

Among the enhanced sampling methodologies, US may be the most classic and widely used technique to explore the unbinding pathway of a ligand from the binding site of its target.^{69–71} The last equilibrated snapshot extracted from the trajectory of the conventional MD simulations was used as the initial structure for the US simulations. To construct the reaction coordinates (RCs) for the US simulations, the dissociation channel, which is defined as the direction along the largest pocket, was determined using the Caver3.0 plugin in PYMOL.⁷² For SFG, the distance between the C5' in SFG and the C_{α} atom of Gly1123 in DNMT1 (corresponding to Gly707 in DNMT3A) was defined as the RC; for DC-05, the distance between the C17 in DC-05 and the C_{α} atom of Glu1266 in DNMT1 (Val687 in DNMT3A) was defined as the RC; for GSKex1, the distance between the C11 in GSKex1 and the C_{α} atom of Glu1266 in DNMT1 (Glu756 in DNMT3A) was defined as the RC. For all the systems, the US simulations were executed across 21 windows in the range of 0–10 Å away from the initial position with each window of 0.5 Å. For each window, 8 ns MD simulations were performed to ensure the convergence of the simulations, and the elastic constant of the restraint potential was set to 6 kcal mol⁻¹ Å⁻². In order to prevent the drifting of the receptor–ligand complex, a restraint force of 5 kcal mol⁻¹ Å⁻² was added to the atoms of the terminal residues in each system. Then, the weighted histogram analysis method (WHAM) was employed to reconstruct the free energy profile along the RC from the biased probability distribution of each window.^{73,74} Herein, the RC was separated into 2000 bins for the WHAM calculation after each cycle of the US simulations. The tolerance

for iteration was set to 0.0001 to generate the convergent potential of mean force (PMF) curves.

Results and discussion

Molecular docking and MM/GBSA rescoring

In this study, the complex structures for DNMT1/DC-05, DNMT1/GSKex1, DNMT3A/SFG, DNMT3A/DC-05 and DNMT3A/GSKex1 were predicted by the Glide docking method. In order to examine the prediction accuracy of Glide with the XP scoring (Glide XP) for the DNMT systems, the ligands in the crystal structures of 3SWR (DNMT1) and 4U7T (DNMT3A) were extracted and redocked by Glide XP. The RMSDs between the predicted pose and the experimentally observed conformation are 0.36 Å and 0.44 Å for 3SWR and 4U7T, respectively, suggesting that the experimentally-determined binding conformations of the ligands could be accurately reproduced by the Glide docking. According to previous studies, the correct binding conformations of ligands in some complexes could not be identified as the best scored conformation by molecular docking calculations, and the combination of MM/GBSA and MD simulations is more accurate to identify the correct binding conformations.³⁷ Therefore, for each protein–ligand complex, the top five binding poses scored by the Glide XP scoring were submitted to 5 ns MD simulations and rescored by MM/GBSA. The binding free energy for each docking pose was calculated by averaging 500 snapshots evenly extracted from 3 to 5 ns MD simulations. It should be noted that, for each complex, the top five poses predicted by Glide XP were saved and rescored by MM/GBSA, but it seems that the top five poses for the DNMT1/GSKex1 complex might not be correct according to our experience. The structures of GSKex1, SFG and SAM are quite similar, and therefore their binding conformations should be similar. However, as shown in Fig. S1 (ESI[†]), the binding pose of GSKex1 predicted by Glide XP shows a reverse orientation compared with that of SFG, which is also different from the binding pose predicted by Glide with the standard precision (SP) scoring mode and that of GSKex1 in DNMT3A. Therefore, for the DNMT1/GSKex1 complex, the top 3 binding conformations predicted by Glide XP, and the binding conformation that is similar to that of SAM and the best binding conformation predicted by Glide SP were used as the initial structures for the MM/GBSA rescoring.

As shown in Table S1 in the ESI[†], the best binding pose predicted by Glide XP does not always correspond to the strongest MM/GBSA binding free energy. According to the binding free energies, the five structures (no. 4, 10, 15, 19, and 22) were identified as the correct binding conformations for the DNMT1/DC-05, DNMT1/GSKex1, DNMT3A/SFG, DNMT3A/DC-05, and DNMT3A/GSKex1 complexes. The predicted binding structures of the five complexes were then compared by structural alignment. The best scored conformation for DNMT1/DC-05 (no. 4) is quite similar to that for DNMT3A/DC-05 (no. 19), which is also consistent with the predicted binding conformation for DNMT1/DC-05 reported by Chen *et al.*²² The predicted binding pose of SFG in DNMT3A (no. 15) is highly similar to that in the crystal structure of DNMT1/SFG (PDB entry: 3SWR), highlighting the

consistency of the binding poses of SFG in DNMT1 and DNMT3A. As we expected, the binding poses of GSKex1 in DNMT1 (no. 10) and DNMT3A (no. 22) are also similar to those of SAM in DNMT1 (PDB entries: 3PTA and 4DA4) and DNMT3A (PDB entry: 4U7T).^{25,26,28} Generally, due to the highly conserved catalytic pockets of DNMTs, the predicted binding poses of the studied inhibitors in DNMT1 and DNMT3A are quite similar (Section S1 in the ESI[†]).

Structural stability and flexibility

The RMSDs of the C_α atoms of the protein in each snapshot relative to the initial structure were calculated to assess the structural stability during the 150 ns MD simulations. As shown in Fig. S2 (ESI[†]), the three DNMT3A systems reached stability earlier than the three DNMT1 systems (~20 ns for DNMT3A/SFG, DNMT3A/DC-05 and DNMT3A/GSKex1, ~40 ns for DNMT1/SFG, and ~60 ns for DNMT1/DC-05 and DNMT1/GSKex1). However, as shown in Fig. 2, the binding sites of the six studied systems are quite stable during the 150 ns simulations by evaluating the RMSDs of the C_α atoms of the residue within 5 Å of the ligands. The RMSDs of the non-hydrogen atoms of the ligands were also calculated. As shown in Fig. S3 (ESI[†]), the selective DNMT1 inhibitor DC-05 showed a large RMSD fluctuation in DNMT3A through the MD simulations (Fig. S3E, ESI[†]), suggesting that the binding of DC-05 to DNMT3A is obviously more unstable than that to DNMT1. Similarly, as shown in Fig. S3C (ESI[†]), the selective DNMT3A inhibitor GSKex1 showed a large RMSD fluctuation in DNMT31, suggesting that the inhibitors with low inhibitory activity cannot form stable interactions with the corresponding targets. The RMSFs of the protein C_α atoms were calculated to assess the flexibility of each residue. As shown in Fig. S4 (ESI[†]), in the DNMT1/inhibitor system, the CXXC domain (residues 646–692), the autoinhibitory linker (residues 693–754), the BAH2 domain (residues 972–1100) and the target recognition domain (TRD) (residues 1336–1551) are more flexible than the BAH1 domain (residues 755–880) and the catalytic domain (residues 1139–1335 and 1552–1599) excluding the TRD. The flexible part contains more loop regions that need to interact with DNA or some

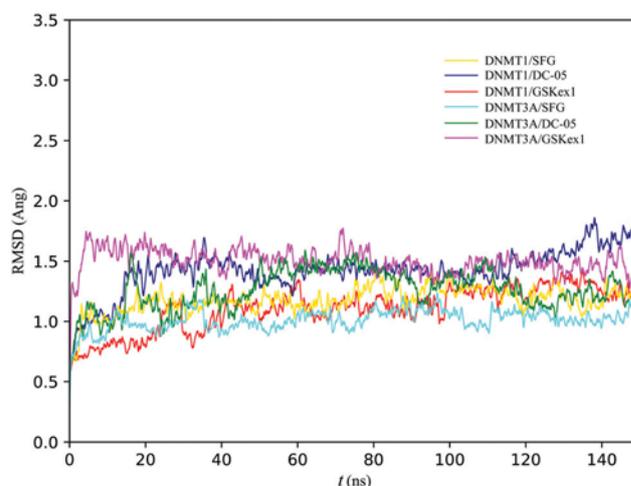


Fig. 2 RMSDs of the main residue C_α atoms within 5 Å of the ligands as a function of simulation time.

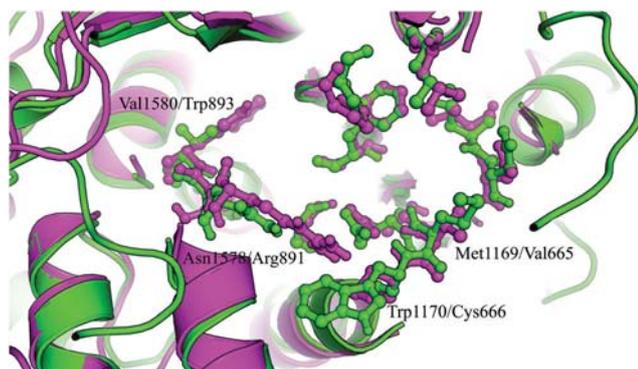


Fig. 3 Structural alignment of the SAM binding pocket of DNMT1 (PDB entry: 3SWR) colored green and DNMT3A (PDB entry: 4U7T) colored magenta.

functional proteins. Similarly, the catalytic domain (residues 627–912) is generally more stable than the ADD domain (residues 476–609) except for the regions (loop L2, residues 806–849) that may interact with DNA in the DNMT3A/inhibitor system.²⁸ In summary, the functional domains of DNMT1 and DNMT3A are more flexible than the conserved catalytic domains. By comparing the three DNMT1 systems or the three DNMT3A systems, we can observe the binding pockets of DNMT1 and DNMT3A with high structural stability, and therefore the flexible region out of the catalytic domain may not have a large impact on the binding of inhibitors.

Comparison of DNMT1 and DNMT3A structures from MD simulations

It is well known that the catalytic pockets of DNMTs are highly conserved in their sequences and structures, and therefore

design of a selective inhibitor towards a specific DNMT isoform is quite challenging. However, there still exist different residues which may generate different interactions with ligands and surface-based local binding sites on the protein, contributing to the selectivity of the inhibitors. In order to compare the binding sites of DNMT1 and DNMT3A, the structures of DNMT1 (3SWR) and DNMT3A (4U7T/A) were aligned by using the *Binding Site Alignment module* in Maestro (Schrodinger, Inc.). As shown in Fig. 3, we can observe that most residues around the binding pockets are hydrophobic and only two residues are hydrophilic, including Glu1168/Glu664 (the former residue is in DNMT1 and the latter is in DNMT3A) and Asp1190/Asp686. In addition, three pairs of the residues in DNMT1/DNMT3A are quite different, including Asn1578/Arg891, Val1580/Trp893 and Trp1170/Cys666. Asn1578 has a polar uncharged side chain while Arg891 has a larger positively charged side chain. The side chain of Trp1170 in DNMT1 is much more hydrophobic and larger than that of Cys666 in DNMT3A. Both Val1580 in DNMT1 and Trp893 in DNMT3A are hydrophobic, but the side chain of Val1580 is much smaller than that of Trp893, indicating distinct hydrophobicity and steric effects. According to the calculation results of Sitemap (Schrodinger, Inc.), the volumes of the SAM binding sites for DNMT1 and DNMT3A are 681.198 Å³ and 566.979 Å³, respectively, suggesting that the presence of Arg891 and Trp893 makes the SAM binding pocket of DNMT3A smaller than that of DNMT1. The structural difference discussed above may be utilized to design novel DNMT inhibitors with good selectivity.

Fig. 4 and 5 show the snapshots of DNMT1/GSKex1 and DNMT3A/DC-05 at four different time points (0 ns, 50 ns, 100 ns, and 150 ns) extracted from the MD trajectories. In the DNMT1/GSKex1 system, the chloropyrimidin group of GSKex1 does not undergo large conformational rearrangement during the first

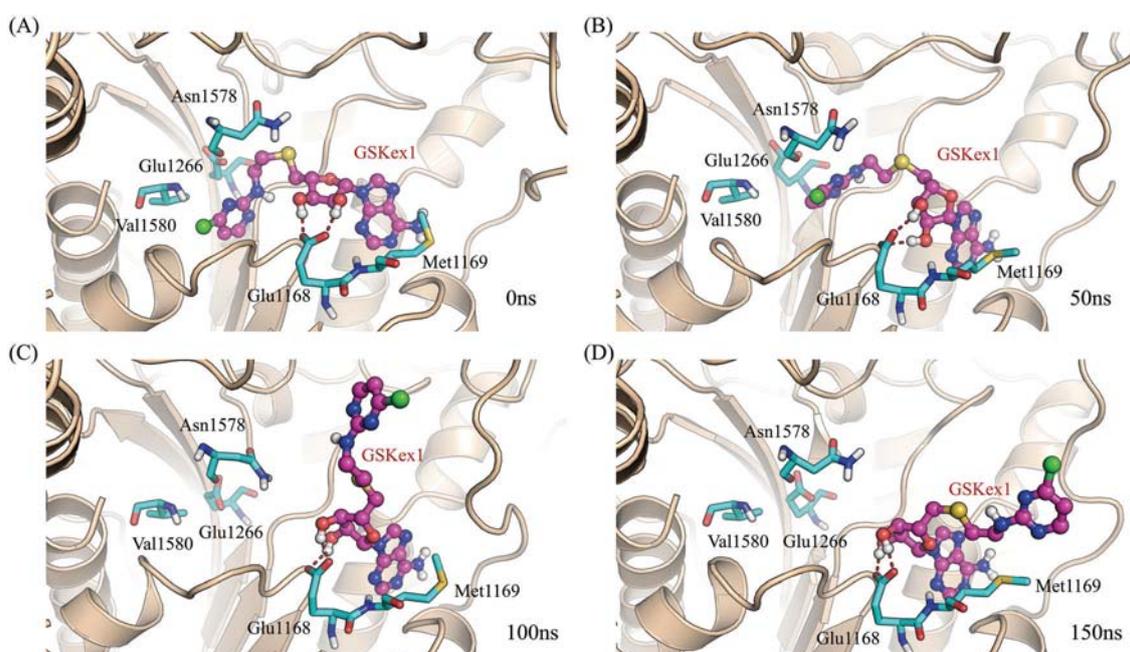


Fig. 4 The snapshots of DNMT1/GSKex1 at (A) 0 ns, (B) 50 ns, (C) 100 ns, and (D) 150 ns extracted from the MD trajectory. The key residues are colored cyan. The H-bonds between GSKex1 (magenta) and Glu1168 are shown in dashed red lines.

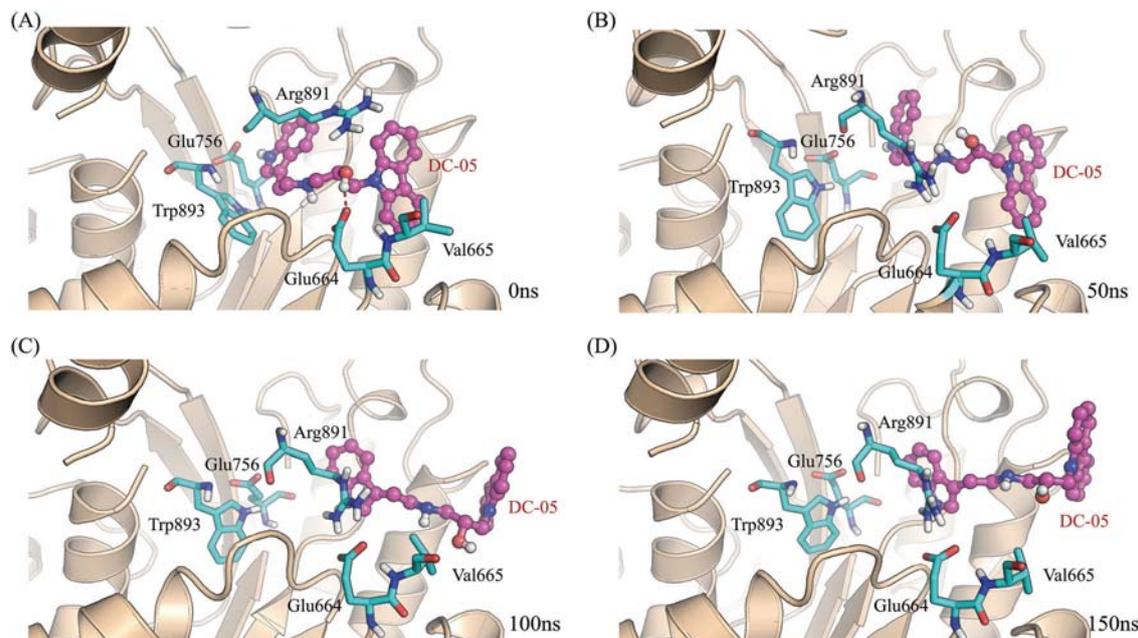


Fig. 5 The snapshots of DNMT3A/DC-05 at (A) 0 ns, (B) 50 ns, (C) 100 ns, and (D) 150 ns extracted from the MD trajectory. The key residues are colored cyan. The H-bonds between DC-05 (magenta) and Glu664 are shown in a dashed red line. After ~ 50 ns MD simulations, there is no any H-bond between DC-05 and DNMT3A.

50 ns simulation (Fig. 4A and B). However, it begins to dissociate from the bottom of the SAM binding pocket at ~ 60 ns, which corresponds to the significant change of the ligand RMSDs shown in Fig. S3C (ESI[†]). As shown in Fig. 4A, C and D, the locations of the chloropyrimidin group of GSKex1 undergo a large conformational change during the MD simulations, but the tetrahydrofuran group is quite stable due to the formation of the two H-bonds between GSKex1 and Glu1168 in DNMT1. Similarly, DC-05 is unstable in the SAM binding pocket of DNMT3A during the MD simulations. However, different from the stepped conformational change of GSKex1 in DNMT1, DC-05 deviates from its initial position at DNMT3A very quickly (Fig. 5 and Fig. S3E, ESI[†]). This phenomenon may be explained by the fact that the tetrahydrofuran group of GSKex1 can form two stable H-bonds with DNMT1 while DC-05 cannot form any stable H-bond with DNMT3A. Additionally, consistent with the time evolutions of RMSDs for the inhibitors (Fig. S3, ESI[†]), the inhibitors in the other four systems are very stable (snapshots not shown). It appears that our MD results reproduce the selectivity of the studied inhibitors well with the experimental data since the potent inhibitors towards DNMT1 or DNMT3A tend to be more stable.^{22,29,30,32}

To further explore the influence of non-inhibitors and inhibitors on the structural movements of DNMT1 or DNMT3A, the PCA and DCCM analyses were performed based on the 1000 snapshots extracted from the last 50 ns equilibrium MD trajectory using the Amber16 *cpptraj* module and Bio3D.⁷⁵ As shown in Fig. 6A–C, the motion patterns of the catalytic domains of DNMT1/SFG and DNMT1/DC-05 are quite similar, but the moving orientation of the TRD region in DNMT1/GSKex1 is obviously different from those of the TRD region in

DNMT1/SFG and DNMT1/DC-05. The residues 877–888 in DNMT3A/DC-05 exhibit a much larger fluctuation than those in DNMT3A/SFG and DNMT3A/GSKex1 (the black cycles in Fig. 6D–F). In addition, it can be observed that the key residues of the six systems are quite conservative and stable. According to the results of the DCCM analysis, it was not difficult to observe that the overall correlated motions of DNMT1/SFG and DNMT1/DC-05 are more dramatic than those of DNMT1/GSKex1 (Fig. 7), and the overall correlated motions of DNMT3A/SFG and DNMT3A/GSKex1 are substantially more dramatic than those of DNMT3A/DC-05 (Fig. 7). The PCA and DCCM results indicate that the inhibitor binding shows certain selectivity towards DNMT1 and DNMT3A, and the conformational changes of the conserved catalytic domains induced by the binding of an inhibitor and a noninhibitor to DNMT may be correspondingly different.

Binding selectivity of inhibitors predicted by MM/GBSA

For most enzyme systems, the bioactivities of inhibitors are largely related to binding affinities, and therefore the selectivity of the studied inhibitors should be correlated with the binding affinities.^{70,76–78} Therefore, in this study, the binding free energies for the six systems were predicted by using the MM/GBSA approach based on the 1000 snapshots evenly extracted from the last 50 ns MD trajectories. As shown in Table 1, the predicted binding free energies (ΔG_{bind}) for DNMT1/SFG, DNMT3A/SFG, DNMT1/DC-05, DNMT3A/DC-05, DNMT1/GSKex1 and DNMT3A/GSKex1 are -53.65 , -57.02 , -37.39 , -22.65 , -40.03 and -59.14 kcal mol⁻¹, respectively. That is to say, DC-05 forms a much stronger interaction with DNMT1 than with DNMT3A, GSKex1 forms a much stronger interaction with DNMT3A than

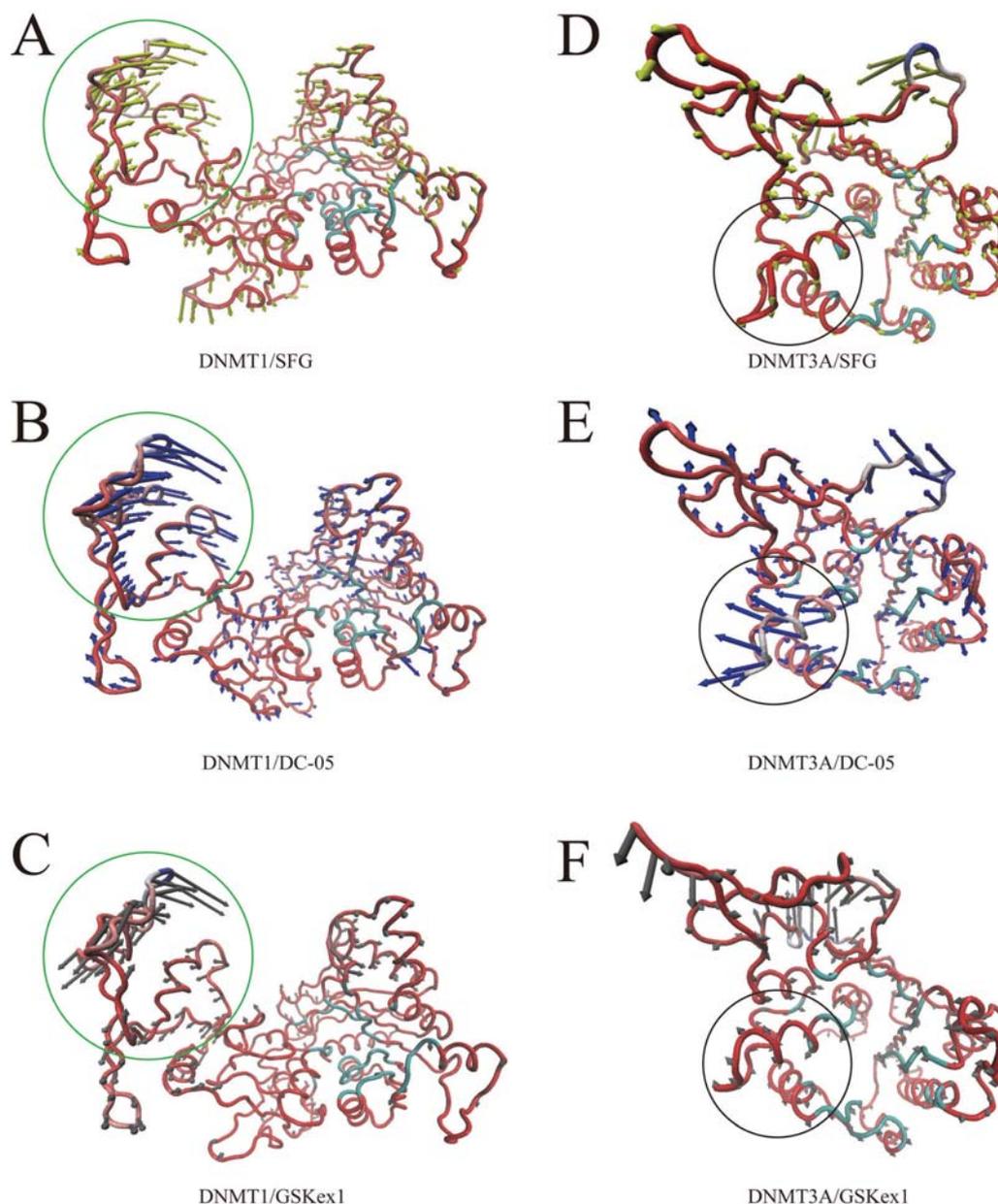


Fig. 6 The first slowest motion modes (PC1) of catalytic domains in (A) DNMT1/SFG, (B) DNMT1/DC-05, (C) DNMT1/GSKex1, (D) DNMT3A/SFG, (E) DNMT3A/DC-05 and (F) DNMT3A/GSKex1 complexes. The C_{α} motion magnitude ($>1 \text{ \AA}$) is reflected by the length of the cone, and the orientation of the cone indicates the motion direction. The green circles in (A)–(C) highlight the different motion magnitudes between DNMT1/GSKex1 and the other two complexes. The black circles in (D)–(F) highlight the different motion magnitudes between DNMT3A/DC-05 and the other two complexes. The cyan regions are the main residue C_{α} atoms within 5 \AA of the ligands.

with DNMT1, and SFG forms relatively similar interactions with DNMT1 and DNMT3A. Therefore, according to the prediction results, DC-05 is a selective DNMT1 inhibitor, GSKex1 is a selective DNMT3A inhibitor and SFG is a DNMT1/DNMT3A dual inhibitor, which is in good agreement with both the above structural analysis and the experimental conclusions. It is worth noting that, for the selective inhibitors DC-05 and GSKex1, the difference in the binding affinities towards DNMT1 and DNMT3A is primarily contributed by the difference in the van der Waals interactions (ΔE_{vdw}) as shown in Table 1.

Key residues for binding specificity predicted by free energy decomposition and hydrogen bonds analysis

In order to highlight the key residues to determine the binding specificity of the studied inhibitors, the residue-specific binding free energies between each inhibitor and protein were calculated by the MM/GBSA free energy decomposition analysis.⁶³ First, for each system, the total binding free energy was decomposed and the top 10 residues with the largest contributions to ligand binding were recorded. As shown in Fig. 8, it can be found that Glu1168, Phe1145, Asn1578,

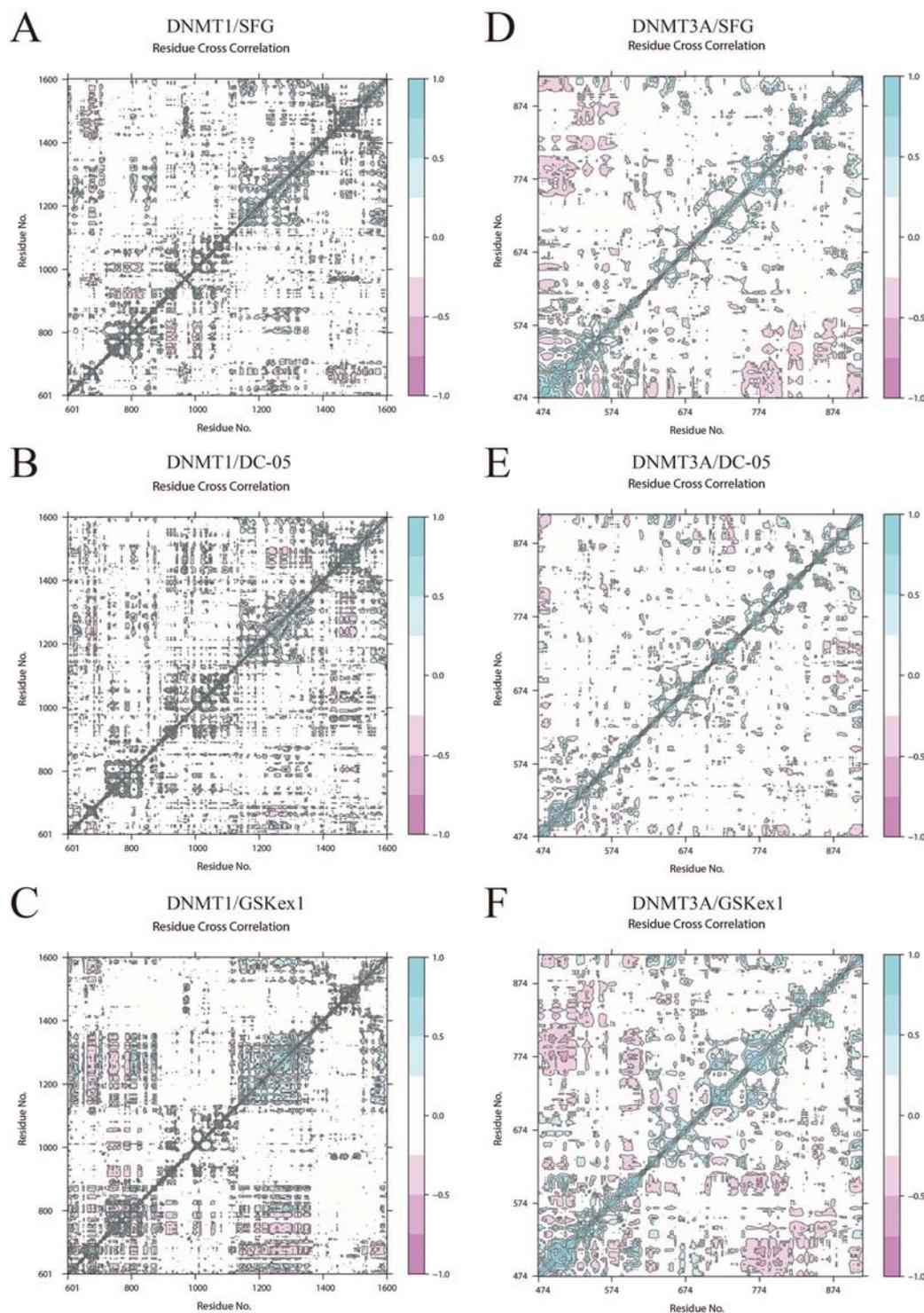


Fig. 7 The DCCMs of the fluctuations of the C_{α} atoms in (A) DNMT1/SFG, (B) DNMT1/DC-05, (C) DNMT1/GSKex1, (D) DNMT3A/SFG, (E) DNMT3A/DC-05 and (F) DNMT3A/GSKex1 around their corresponding mean positions. The extent of the correlated motions and anticorrelated motions is color-coded in the panel. Dark blue indicates highly positive correlation whereas dark red refers to strong anticorrelation.

Met1169 and several other residues are the most important residues in the DNMT1 systems. As shown in Fig. 8D and F, it can be determined that Glu664, Trp893, Phe640 and Asp686 play an important role in stabilizing the DNMT3A/inhibitor interactions. However, for the DNMT3A systems, the top 10

residues are not quite similar. For example, the top 10 residues for DNMT3A/DC-05 in Fig. 8E are quite different from those for DNMT3A/SFG and DNMT3A/GSKex1. This may be explained by the fact that DC-05 deviates from the initial conformation at the beginning of the MD simulations mentioned above. Then, the

Table 1 The binding free energies and energy components predicted by MM/GBSA (kcal mol⁻¹)

System	ΔE_{vdw}	ΔE_{elec}	ΔG_{GB}	ΔG_{SA}	ΔG_{bind}	$\Delta\Delta G_{\text{bind}}^a$	$\Delta\Delta E_{\text{vdw}}^b$	$\Delta E_{\text{elec}} + \Delta G_{\text{GB}}$	IC ₅₀ (μM) ^c
DNMT1/SFG	-44.36 ± 0.25	-71.66 ± 2.47	68.53 ± 2.52	-6.16 ± 0.02	-53.65 ± 0.22	3.38	-3.71	-3.13	7.4
DNMT3A/SFG	-40.65 ± 0.40	-154.56 ± 1.15	144.00 ± 0.50	-5.82 ± 0.00	-57.02 ± 0.26			-10.55	1.8
DNMT1/DC-05	-42.85 ± 2.24	60.28 ± 0.93	-48.87 ± 0.78	-5.95 ± 0.20	-37.39 ± 0.74	-14.74	-10.97	11.41	10.3
DNMT3A/DC-05	-31.88 ± 2.86	-14.27 ± 2.35	27.43 ± 2.44	-3.94 ± 0.25	-22.65 ± 3.02			13.16	>200
DNMT1/GSKex1	-37.89 ± 1.45	-54.80 ± 1.21	57.39 ± 1.51	-4.73 ± 0.15	-40.03 ± 1.29	19.11	18.10	2.59	>300
DNMT3A/GSKex1	-55.99 ± 0.84	-57.80 ± 0.66	60.98 ± 0.22	-6.32 ± 0.11	-59.14 ± 0.51			3.17	3

^a $\Delta\Delta G_{\text{bind}} = \Delta G_{\text{bind DNMT1/inhibitor}} - \Delta G_{\text{bind DNMT3A/inhibitor}}$; ^b $\Delta\Delta E_{\text{vdw}} = \Delta E_{\text{vdw DNMT1/inhibitor}} - \Delta E_{\text{vdw DNMT3A/inhibitor}}$; ^c IC₅₀^{22,29,30,32} half maximal inhibitory concentration.

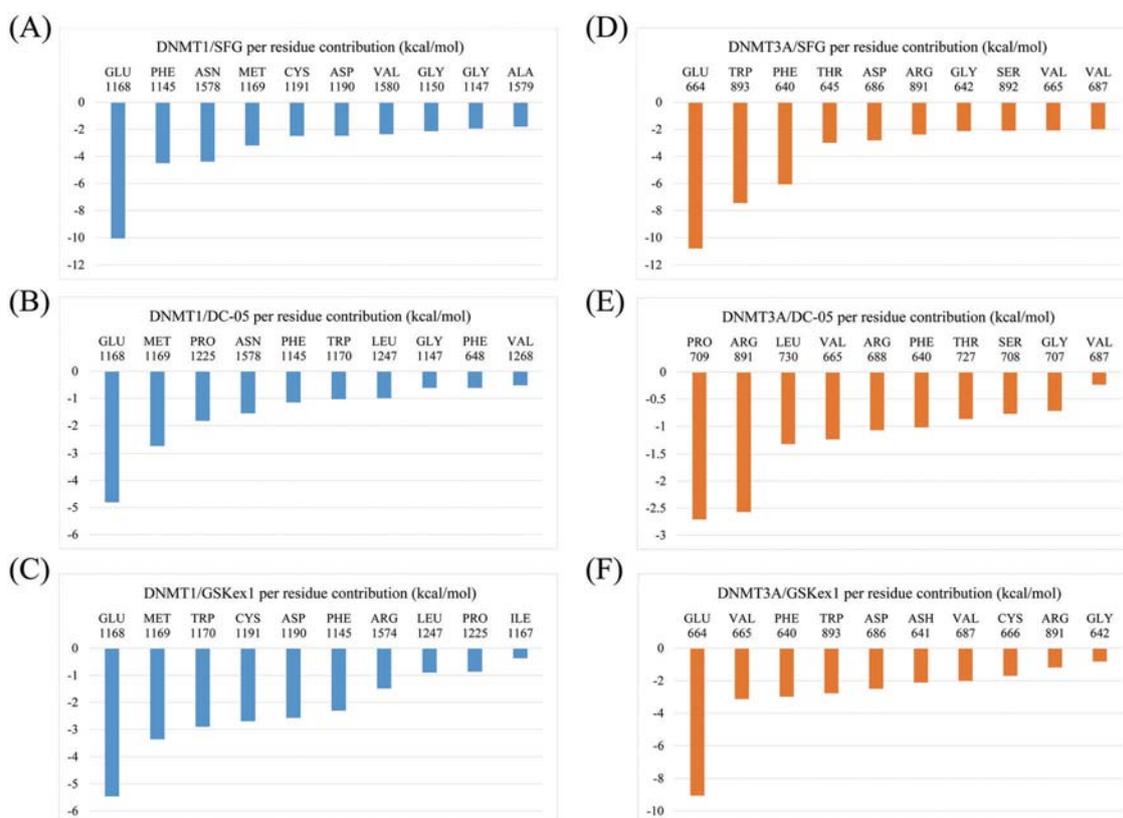


Fig. 8 Contributions of the top ten residues to the binding of (A) SFG to DNMT1, (B) DC-05 to DNMT1, (C) GSKex1 to DNMT1, (D) SFG to DNMT3A, (E) DC-05 to DNMT3A, and (F) GSKex1 to DNMT3A predicted by the MM/GBSA binding free energy decomposition.

total contribution of each DNMT1 or DNMT3A residue to the binding of the three inhibitors was calculated, and 13 important residue pairs with noticeable contributions to ligand binding were determined, including Cys1191/Val687, Phe1145/Phe640, Gly1147/Gly642, Gly1150/Thr645, Glu1168/Glu664, Met1169/Val665, Trp1170/Cys666, Asp1190/Asp686, Pro1225/Pro709, Leu1247/Leu730, Asn1578/Arg891, Ala1579/Ser892, and Val1580/Trp893. Afterwards, we calculated the energy difference between each DNMT1/inhibitor and DNMT3A/inhibitor ($\Delta\Delta G_{\text{res_inh}} = \Delta G_{\text{res_inh/DNMT1}} - \Delta G_{\text{res_inh/DNMT3A}}$) for the 13 important residue pairs.

Firstly, we calculated the sum of $\Delta\Delta G_{\text{bind_res}}$ of the 13 residue pairs, and these values are 5.70, -6.31 and 6.96 kcal mol⁻¹ for SFG, DC-05 and GSKex1, respectively, suggesting that the interactions of SFG and GSKex1 with the key residues of DNMT3A are

stronger than those with the key residues of DNMT1, and the interaction of DC-05 with the key residues of DNMT3A is weaker than that with the key residues of DNMT1. The results are basically consistent with the binding selectivity of the three inhibitors, suggesting that the selected key residues play an important role in determining the binding selectivity of the inhibitors. Then, we systematically compared the contributions of the 13 important residues in DNMT1 and DNMT3A to the binding of SFG and GSKex1 (Fig. 9A and C). We can observe that most of the conservative residues in DNMT1 and DNMT3A have similar contributions to the binding of SFG or GSKex1. However, the inhibitors can form a stronger interaction with Trp893 in DNMT3A than those with Val1580 in DNMT1. The side chain of Trp893 is much larger and more hydrophobic than that of

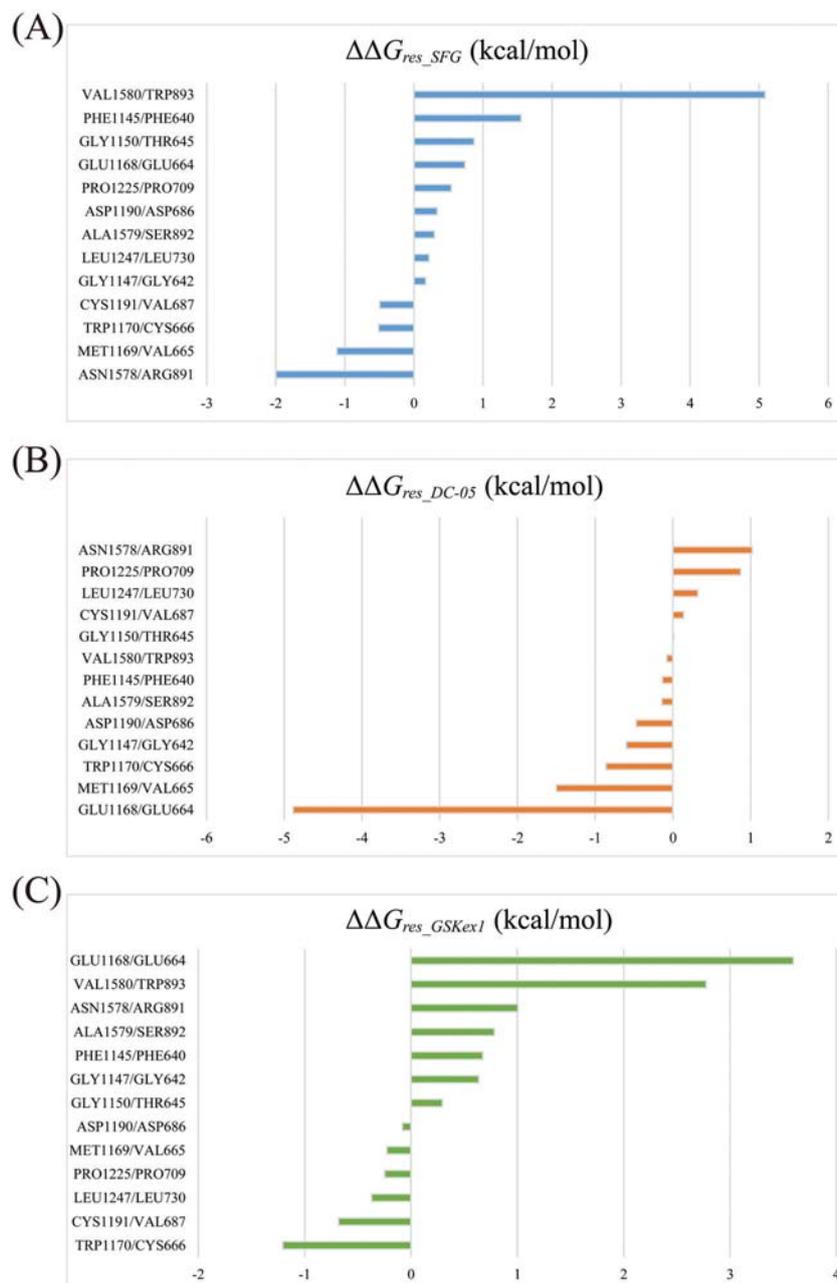


Fig. 9 Energy difference ($\Delta\Delta G_{res_inh} = \Delta G_{res_inh/DNMT1} - \Delta G_{res_inh/DNMT3A}$) of the selected 13 important residue pairs between system (A) DNMT1/SFG and DNMT3A/SFG, (B) DNMT1/DC-05 and DNMT3A/DC-05, or (C) DNMT1/GSKex1 and DNMT3A/GSKex1.

Val1580, and therefore it can form more favorable van der Waals interactions with the inhibitors. Furthermore, Met1169/Val665 is also an important residue pair to determine the binding specificity of inhibitors. Met1169 in DNMT1 provides more favorable energetic contributions to the binding of the three inhibitors than Val665 in DNMT3A.

We then analyzed the distributions of the H-bonds formed between the proteins and inhibitors. The average numbers of H-bonds are 9.79, 8.92, 2.79, 0.80, 5.25 and 6.88 for DNMT1/SFG, DNMT3A/SFG, DNMT1/DC-05, DNMT3A/DC-05, DNMT1/GSKex1 and DNMT3A/GSKex1, respectively. As shown in Fig. 10, the

stable H-bonds formed between SFG and 5 residues (Glu1168/Glu664, Asp1190/Asp686, Cys1191/Val987, Asn1578/Arg891, and Val1580/Trp893) stabilize the structure of the ligand in both proteins. However, the H-bonding patterns of DC-05 with DNMT1 and DNMT3A are completely different. In DNMT1/DC-05, DC-05 forms an H-bond with Glu1168, which is consistent with the observations reported by Chen *et al.*²² In DNMT3A/DC-05, DC-05 is located in the very shallow position of the binding pocket in DNMT3A. Because of the huge side chain of Trp893, DC-05 that bears larger substituents at the both ends than SFG cannot enter the bottom of the pocket and form any stable H-bond with the

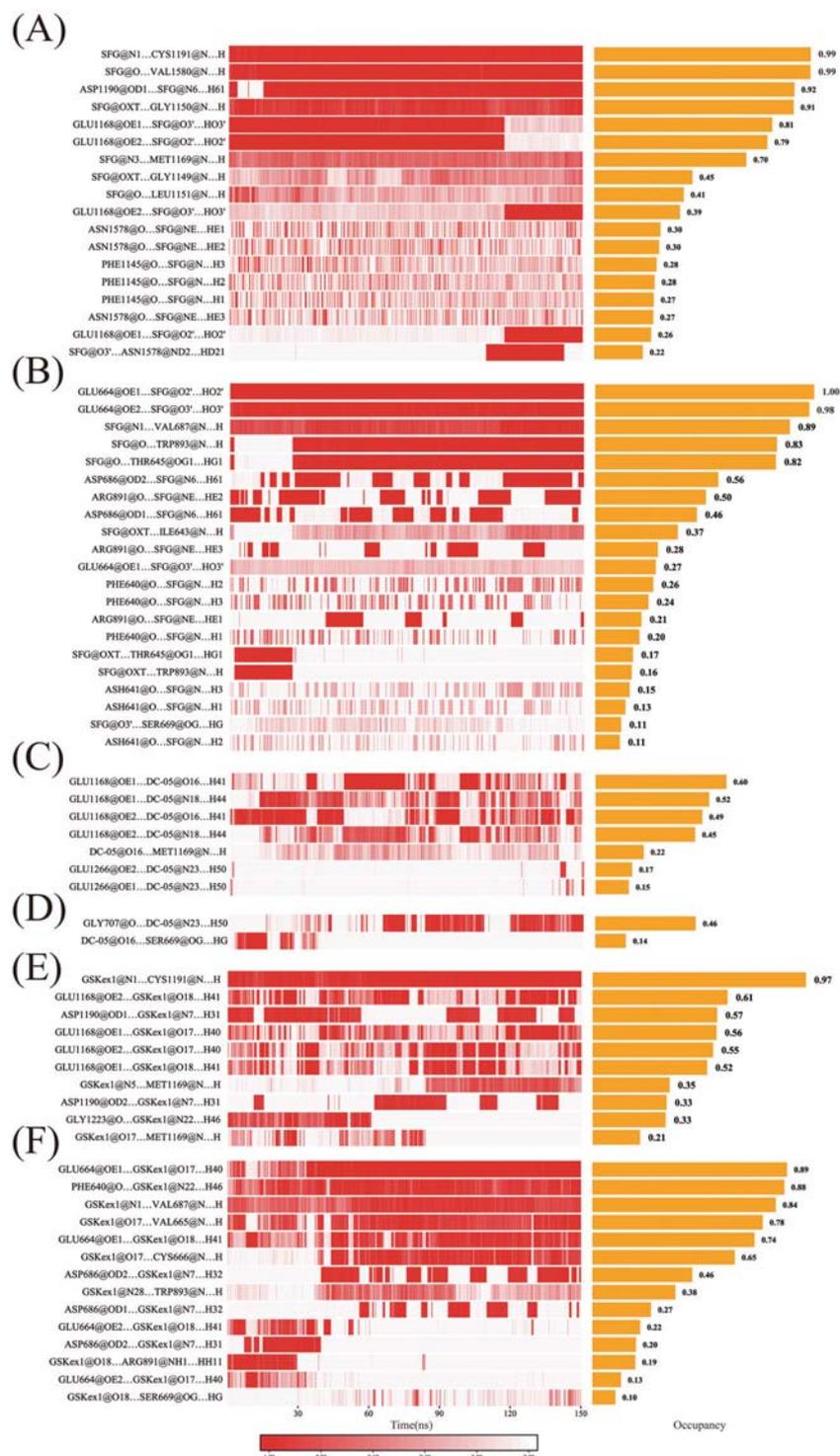


Fig. 10 H-bond analysis for (A) DNMT1/SFG, (B) DNMT3A/SFG, (C) DNMT1/DC-05, (D) DNMT3A/DC-05, (E) DNMT1/GSKex1 and (F) DNMT3A/GSKex1 complexes. Occupancy represents the occupancy of the H-bonds in all 75 000 frames. Each MD trajectory was split into 100 equal time windows. The differentiated shades and deepness of the red color are used to distinguish the length of the H-bond lifetime over time windows. And the H-bond is ignored if the occupancy is less than 10%.

corresponding Glu residues (Fig. S5, ESI[†]). In addition, although Arg891 plays an active role in the binding of DC-05 to DNMT3A, its huge side chain also has a substantial impact on the conformation of DC-05 in the pocket. Therefore, we can conclude

that the spatial constraint is a main reason why DC-05 cannot form favorable interactions with DNMT3A. According to the H-bond analysis, three stable H-bonds between GSKex1 and DNMT1/DNMT3A can be observed. It is also worth noting that the

π - π interaction between Trp893 and the chloropyrimidin group of GSKex1 is favorable to stabilize this structure (Fig. S6, ESI[†]). Miletic *et al.* have also observed that Trp893 can form π - π stacking interaction with mechanism-based inhibition.⁷⁹ In DNMT1, since Trp893 is replaced by Val1580, the π - π interaction does not exist and the chloropyrimidin group of GSKex1 in the deep pocket cannot be stabilized. Besides, compared with the carboxyl group of SFG, the chloropyrimidin group of GSKex1 is not so easy to form H-bonds with the proteins. According to the above analysis, we found that the residue pair Val1580/Trp893 is critical for the binding selectivity of inhibitors. The huge and flexible side chain of Trp893 is a double-edged sword for the binding of inhibitors.

In general, Trp893 can stabilize the conformation of an inhibitor, but if the substituent of an inhibitor towards the inside of the binding site is large, Trp893 may hinder the binding of the inhibitor to DNMT3A.

Unbinding pathways of inhibitors dissociating from DNMT1 and DNMT3A

Understanding the dissociation process of an inhibitor from its target can elucidate the dynamic mechanisms of interactions and is valuable for the design of novel selective inhibitors.^{69–71} Although recent advances in computer resources allow for using long-time-scale conventional MD simulations, sampling the dissociation pathway of a ligand is still a challenging task.^{80–82} Herein, US simulation,^{83–86} one of the most classical enhanced sampling methodologies, was employed to characterize the unbinding processes of SFG, DC-05 and GSKex1 from the binding pockets in DNMT1 and DNMT3A to further clarify the dynamic mechanisms of binding selectivity. The equilibrated snapshots extracted from the MD trajectories were used as the initial structures for the US simulations. To guarantee the sampling convergence of the US simulations, 8 ns US simulations for each window were performed for all the systems, and the convergence of each PMF curve was checked after each nanosecond. As shown in Fig. S7 (ESI[†]), the PMFs achieved stability after ~ 5 ns US simulations for each window, and therefore

Table 2 The PMFs for the dissociation of the studied inhibitors from the binding sites of DNMT1 and DNMT3A predicted by the US simulations

System	ΔW_{PMF}^a (kcal mol ⁻¹)	$\Delta\Delta G_{\text{US}}^b$ (kcal mol ⁻¹)	IC ₅₀ (μM)
DNMT1/SFG	-15.50 ± 0.55	0.33	7.4
DNMT3A/SFG	-15.83 ± 0.62		1.8
DNMT1/DC-05	-11.84 ± 0.56	-5.2	10.3
DNMT3A/DC-05	-6.64 ± 0.99		>200
DNMT1/GSKex1	-5.11 ± 0.38	1.55	>300
DNMT3A/GSKex1	-6.66 ± 1.13		3

^a The ΔW_{PMF} and the standard deviations were estimated by averaging the bins across 8–10 Å of the RC based on the last 3 ns US simulations. ^b The energy difference was calculated by $\Delta\Delta G_{\text{US}} = \Delta W_{\text{PMF, DNMT3A}} - \Delta W_{\text{PMF, DNMT1}}$.

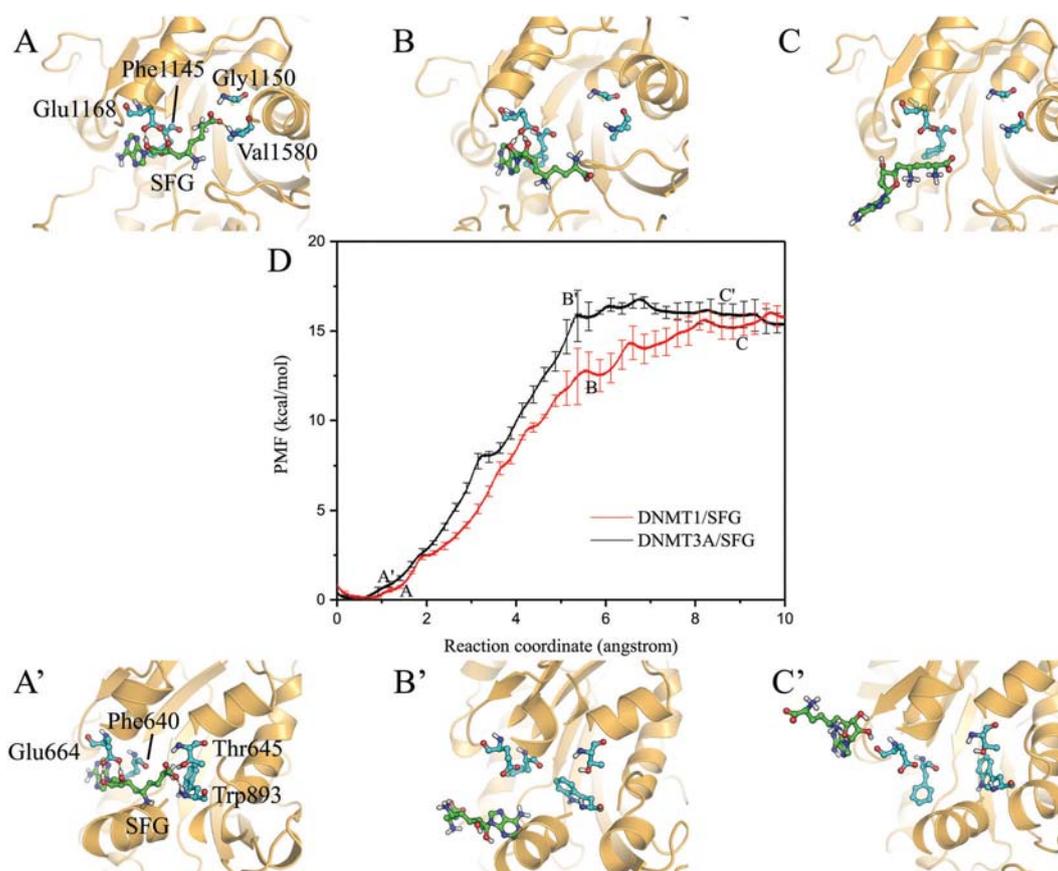


Fig. 11 Dissociation process of SFG from the binding site of (A–C) DNMT1 or (A'–C') DNMT3A along the RCs and (D) the corresponding PMF curves.

the PMF curves were generated based on the last 3 ns US samples (6–8 ns).

As listed in Table 2, the predicted binding affinities (PMF depth, ΔW_{PMF}) of DNMT1/SFG, DNMT3A/SFG, DNMT1/DC-05, DNMT3A/DC-05, DNMT1/GSKex1 and DNMT3A/GSKex1 are -15.50 , -15.83 , -11.84 , -6.64 , -5.11 and -6.66 kcal mol $^{-1}$, respectively. This also proves that SFG is a DNMT1/DNMT3A dual inhibitor, and DC-05 is a DNMT1 selective inhibitor while GSKex1 is a DNMT3A selective inhibitor, which in satisfactory agreement with previous results.

We should notice that the PMF curves for DNMT1/SFG are not quite different from those for DNMT3A/SFG if the calculation error was considered (Fig. 11D). It could be conferred that during the unbinding processes, the binding poses of SFG in DNMT1 and DNMT3A are quite similar (Fig. 11A–C and A'–C'). As shown in Fig. 11B and B', the inhibitor moves to the entrance of the binding pocket with the attenuation of the original interactions (Val1580, Phe1145 and Gly1150 in DNMT1, and Trp893, Phe640 and Thr645 in DNMT3A). Subsequently, the inhibitor continues to move horizontally with the gradual disappearance of the strongest H-bond between SFG and Glu1168 in DNMT1 (Glu664 in DNMT3A), and eventually moves out of the binding pocket to the solvent (Fig. 11C and C').

For the dissociation of DC-05 from DNMT1 (Fig. 12A), starting with the rotation of the indolyl group of DC-05, the H-bond between indolyl and Glu1266 vanishes quickly, leading to an increase of the free energy of ~ 3 kcal mol $^{-1}$ (Fig. 12B). The free

energy continuously increases to ~ 11 kcal mol $^{-1}$ when DC-05 moves vertically to the entrance of the binding cavity, accompanied by the break of the H-bond between DC-05 and Glu1168 (Fig. 12C). Then, the system reaches an energy balance with the carbazole group almost moving out of the binding pocket. Finally, DC-05 rotates and totally dissociates from the binding pocket (Fig. 12D). As shown in Fig. 12E, the PMF profile of DC-05 from DNMT1 is distinctly higher than that from DNMT3A, with a difference of 5.2 kcal mol $^{-1}$, suggesting a deeper energy potential depth and thus a longer residence time. The dissociation of DC-05 from DNMT3A is much easier than that from DNMT1, which can be explained by fewer obstacles of the pocket (Fig. 12A'–D'). These prediction results provide solid theoretical evidence to understand the DNMT1 selectivity of DC-05.

For the DNMT3A/GSKex1 system, with the increase of the biasing potential, the inhibitor wiggles out to the edge of the binding cavity, and then overcomes a free energy barrier of ~ 5 kcal mol $^{-1}$, where the π - π stacking interaction between GSKex1 and Trp893 vanishes (Fig. 13B'), and finally moves away from the binding pocket (Fig. 13C'). In comparison, the PMF profile of GSKex1 from DNMT1 is lower than that from DNMT3A, with a difference of 1.55 kcal mol $^{-1}$ (Fig. 13D), further confirming the binding selectivity of GSKex1 to DNMT3A. In the DNMT1/GSKex1 system, a rotation of GSKex1 was observed at ~ 4.5 Å of the RC because of the unstable H-bond between Gly1223 and GSKex1 (Fig. 13B), which is consistent with the results provided by the MM/GBSA binding free energy decomposition.

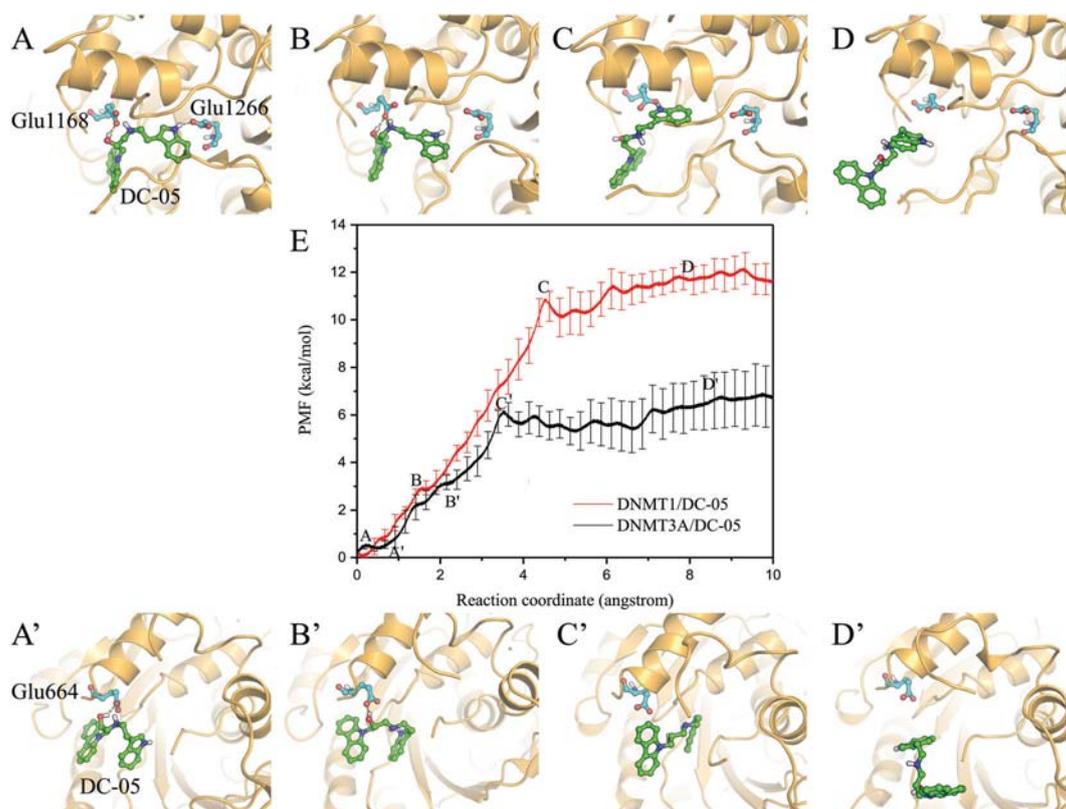


Fig. 12 Dissociation process of DC-05 from the binding site of (A–D) DNMT1 or (A'–D') DNMT3A along the RCs and (E) the corresponding PMF curves.

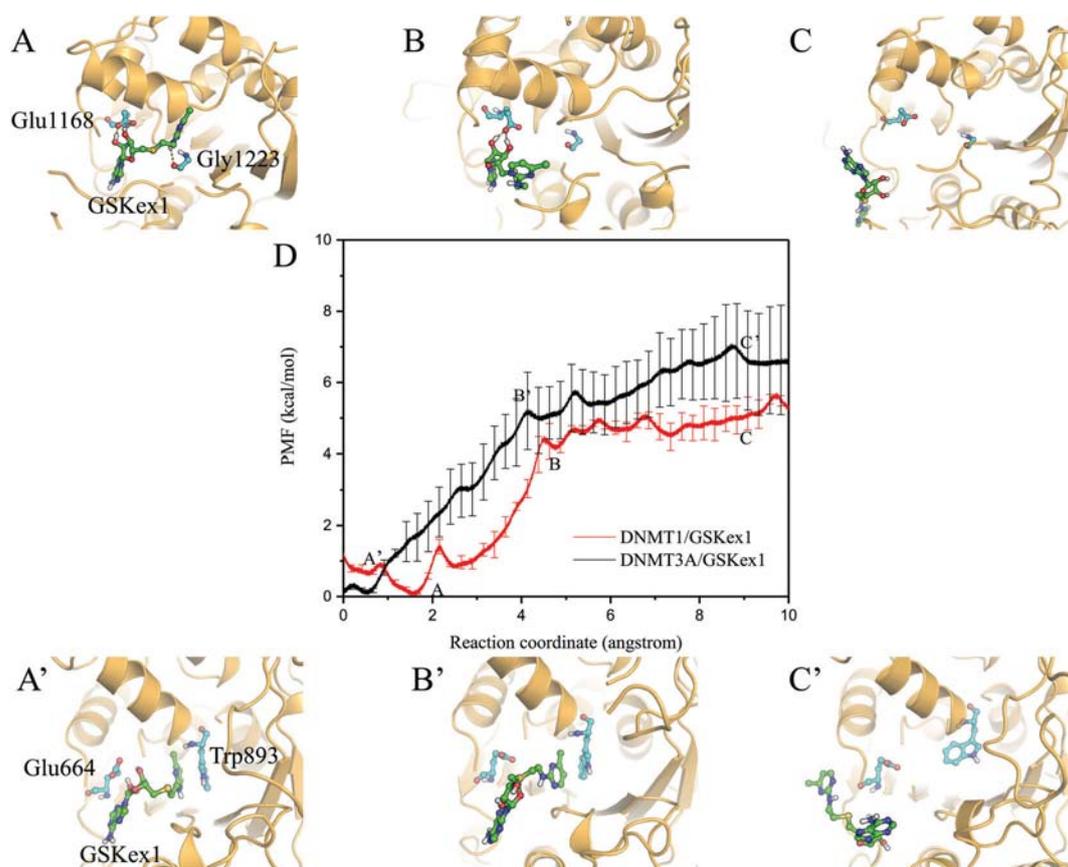


Fig. 13 Dissociation process of GSKex1 from the binding site of (A–C) DNMT1 or (A'–C') DNMT3A along the RCs and (D) the corresponding PMF curves.

Conclusion

In our study, conventional MD simulations, MM/GBSA free energy calculations and US simulations were employed to clarify the molecular principles of the binding selectivity of three DNMT inhibitors (SFG, DC-05, and GSKex1) towards DNMT1 and DNMT3A. The binding specificity of the studied inhibitors can be roughly predicted by the binding free energies through MM/GBSA calculations. The energy decomposition analysis suggests that the variance of the van der Waals interactions was contributed by the non-conserved residues in the SAM binding pocket. In particular Val1580/Trp893, Asn1578/Arg891 and Met1169/Val665 in DNMT1/DNMT3A have a significant impact on the selectivity of DNMT inhibitors reflected by the π - π stacking and H-bond interactions. Furthermore, the US simulations confirmed the predictions on the selectivity given by the structural analysis and MM/GBSA free energy calculations. We expect that our study can help to gain a deeper understanding of the selectivity mechanism of DNMT inhibitors and provide useful information for the design of novel selective DNMT inhibitors.

Authors' contributions

TH and YK conceived the idea and supervised the study. TX, JY, and WF conducted the simulations. TX, JY, WF, ZW, LX, SC and EW analyzed the data. TX and JY drafted the manuscript.

TH, YK, SZ and FZ edited the manuscript. All authors read and approved the final manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was financially supported by the National Science and Technology Major Project of China "Key New Drug Creation and Manufacturing Program" (2018ZX09711002-007), the National Science Foundation of China (21575128, 81773632, 21603189), and the Zhejiang Provincial Natural Science Foundation of China (LZ19H300001).

References

- 1 P. A. Jones and P. W. Laird, *Nat. Genet.*, 1999, **21**, 163–167.
- 2 K. D. Robertson and A. P. Wolffe, *Nat. Rev. Genet.*, 2000, **1**, 11–19.
- 3 P. Siedlecki, R. G. Boy, T. Musch, B. Brueckner, S. Suhai, F. Lyko and P. Zielenkiewicz, *J. Med. Chem.*, 2006, **49**, 678–683.

- 4 J. G. Herman, A. Umar, K. Polyak, J. R. Graff, N. Ahuja, J. P. J. Issa, S. Markowitz, J. K. V. Willson, S. R. Hamilton, K. W. Kinzler, M. F. Kane, R. D. Kolodner, B. Vogelstein, T. A. Kunkel and S. B. Baylin, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 6870–6875.
- 5 G. Egger, G. N. Liang, A. Aparicio and P. A. Jones, *Nature*, 2004, **429**, 457–463.
- 6 C. B. Yoo and P. A. Jones, *Nat. Rev. Drug Discovery*, 2006, **5**, 37–50.
- 7 M. Okano, S. P. Xie and E. Li, *Nat. Genet.*, 1998, **19**, 219–220.
- 8 D. Jia, R. Z. Jurkowska, X. Zhang, A. Jeltsch and X. D. Cheng, *Nature*, 2007, **449**, 248–251.
- 9 S. Mizuno, T. Chijiwa, T. Okamura, K. Akashi, Y. Fukumaki, Y. Niho and H. Sasaki, *Blood*, 2001, **97**, 1172–1179.
- 10 A. T. Agoston, P. Argani, S. Yegnasubramanian, A. M. De Marzo, M. A. Ansari-Lari, J. L. Hicks, N. E. Davidson and W. G. Nelson, *J. Biol. Chem.*, 2005, **280**, 18302–18310.
- 11 Z. J. Zhao, Q. X. Wu, J. A. Cheng, X. M. Qiu, J. Q. Zhang and H. Fan, *J. Biomed. Biotechnol.*, 2010, **2010**, 737535.
- 12 K. E. Bachman, B. H. Park, I. Rhee, H. Rajagopalan, J. G. Herman, S. B. Baylin, K. W. Kinzler and B. Vogelstein, *Cancer Cell*, 2003, **3**, 89–95.
- 13 F. Chik and M. Szyf, *Carcinogenesis*, 2011, **32**, 224–232.
- 14 D. Subramaniam, R. Thombre, A. Dhar and S. Anant, *Front. Oncol.*, 2014, **4**, 80.
- 15 A. Erdmann, L. Halby, J. Fahy and P. B. Arimondo, *J. Med. Chem.*, 2015, **58**, 2569–2583.
- 16 G. Rajendran, K. Shanmuganandam, A. Bendre, D. Mujumdar, A. Goel and A. Shiras, *J. Neuro-Oncol.*, 2011, **104**, 483–494.
- 17 L. I. Shlush, S. Zandi, A. Mitchell, W. C. Chen, J. M. Brandwein, V. Gupta, J. A. Kennedy, A. D. Schimmer, A. C. Schuh, K. W. Yee, J. L. McLeod, M. Doedens, J. J. F. Medeiros, R. Marke, H. J. Kim, K. Lee, J. D. McPherson, T. J. Hudson, A. M. K. Brown, Q. M. Trinh, L. D. Stein, M. D. Minden, J. C. Y. Wang, J. E. Dick and H. P.-L. G. Panel, *Nature*, 2014, **506**, 328–333.
- 18 T. Deng, Y. Kuang, L. Wang, J. Li, Z. Wang and J. Fei, *Biochem. Biophys. Res. Commun.*, 2009, **387**, 611–616.
- 19 S. He, F. Wang, L. Yang, C. Guo, R. Wan, A. Ke, L. Xu, G. Hu, X. Xu, J. Shen and X. Wang, *PLoS One*, 2011, **6**, e27684.
- 20 M. Esteller, *N. Engl. J. Med.*, 2008, **358**, 1148–1159.
- 21 B. Brueckner, R. G. Boy, P. Siedlecki, T. Musch, H. C. Kliem, P. Zielenkiewicz, S. Suhai, M. Wiessler and F. Lyko, *Cancer Res.*, 2005, **65**, 6305–6311.
- 22 S. Chen, Y. Wang, W. Zhou, S. Li, J. Peng, Z. Shi, J. Hu, Y.-C. Liu, H. Ding, Y. Lin, L. Li, S. Cheng, J. Liu, T. Lu, H. Jiang, B. Liu, M. Zheng and C. Luo, *J. Med. Chem.*, 2014, **57**, 9028–9041.
- 23 D. Pechalrieu, C. Etievant and P. B. Arimondo, *Biochem. Pharmacol.*, 2017, **129**, 1–13.
- 24 Y. Pan, G. Liu, F. Zhou, B. Su and Y. Li, *Clin. Exp. Med.*, 2018, **18**, 1–14.
- 25 J. Song, O. Rechkoblit, T. H. Bestor and D. J. Patel, *Science*, 2011, **331**, 1036–1040.
- 26 J. Song, M. Teplova, S. Ishibe-Murakami and D. J. Patel, *Science*, 2012, **335**, 709–712.
- 27 Z.-M. Zhang, S. Liu, K. Lin, Y. Luo, J. J. Perry, Y. Wang and J. Song, *J. Mol. Biol.*, 2015, **427**, 2520–2531.
- 28 X. Guo, L. Wang, J. Li, Z. Ding, J. Xiao, X. Yin, S. He, P. Shi, L. Dong, G. Li, C. Tian, J. Wang, Y. Cong and Y. Xu, *Nature*, 2015, **517**, 640–644.
- 29 C. Gros, L. Chauvigne, A. Poulet, Y. Menon, F. Ausseil, I. Dufau and P. B. Arimondo, *Nucleic Acids Res.*, 2013, **41**, e185.
- 30 S. Valente, Y. W. Liu, M. Schnekenburger, C. Zwergel, S. Cosconati, C. Gros, M. Tardugno, D. Labella, C. Florean, S. Minden, H. Hashimoto, Y. Q. Chan, X. Zhang, G. Kirsch, E. Novellino, P. B. Arimondo, E. Miele, E. Ferretti, A. Gulino, M. Diederich, X. D. Cheng and A. Mai, *J. Med. Chem.*, 2014, **57**, 701–713.
- 31 Z. Y. Shao, P. Xu, W. Xu, L. J. Li, S. Liu, R. K. Zhang, Y. C. Liu, C. H. Zhang, S. J. Chen and C. Luo, *Bioorg. Med. Chem. Lett.*, 2017, **27**, 342–346.
- 32 Glaxosmithkline Intellectual Property Development Limited, New compounds, WO2013062943, 2013.
- 33 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.
- 34 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, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *J. Med. Chem.*, 2004, **47**, 1739–1749.
- 35 G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju and W. Sherman, *J. Comput.-Aided Mol. Des.*, 2013, **27**, 221–234.
- 36 E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyran, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel and R. A. Friesner, *J. Chem. Theory Comput.*, 2016, **12**, 281–296.
- 37 T. Hou, J. Wang, Y. Li and W. Wang, *J. Comput. Chem.*, 2011, **32**, 866–877.
- 38 C. R. Sondergaard, M. H. M. Olsson, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comput.*, 2011, **7**, 2284–2295.
- 39 M. H. Olsson, C. R. Søndergaard, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comput.*, 2011, **7**, 525–537.
- 40 T. J. Dolinsky, J. E. Nielsen, J. A. McCammon and N. A. Baker, *Nucleic Acids Res.*, 2004, **32**, W665–W667.
- 41 C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.
- 42 M. Frisch, G. Trucks, H. B. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, B. Mennucci and G. Petersson, *Gaussian 09, revision A. 02*, Gaussian, Inc., Wallingford, CT, 2009, vol. 19, pp. 227–238.
- 43 D. A. Case, D. S. Cerutti, T. E. Cheatham, III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, *AMBER 2017*, University of California, San Francisco, 2017.
- 44 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graphics Modell.*, 2006, **25**, 247–260.

- 45 J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
- 46 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
- 47 M. B. Peters, Y. Yang, B. Wang, L. Fuesti-Molnar, M. N. Weaver and K. M. Merz Jr., *J. Chem. Theory Comput.*, 2010, **6**, 2935–2947.
- 48 D. Paschek, R. Day and A. E. Garcia, *Phys. Chem. Chem. Phys.*, 2011, **13**, 19840–19847.
- 49 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
- 50 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
- 51 S. G. Lambrakos, J. P. Boris, E. S. Oran, I. Chandrasekhar and M. Nagumo, *J. Comput. Phys.*, 1989, **85**, 473–486.
- 52 J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
- 53 F. Chen, H. Liu, H. Sun, P. Pan, Y. Li, D. Li and T. Hou, *Phys. Chem. Chem. Phys.*, 2016, **18**, 22129–22139.
- 54 T. Hou, J. Wang, Y. Li and W. Wang, *J. Chem. Inf. Model.*, 2011, **51**, 69–82.
- 55 H. Sun, Y. Li, S. Tian, L. Xu and T. Hou, *Phys. Chem. Chem. Phys.*, 2014, **16**, 16719–16729.
- 56 L. Xu, H. Sun, Y. Li, J. Wang and T. Hou, *J. Phys. Chem. B*, 2013, **117**, 8408–8421.
- 57 P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. H. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case and T. E. Cheatham, *Acc. Chem. Res.*, 2000, **33**, 889–897.
- 58 F. Chen, H. Sun, J. Wang, F. Zhu, H. Liu, Z. Wang, T. Lei, Y. Li and T. Hou, *RNA*, 2018, **24**, 1183–1194.
- 59 A. Onufriev, D. Bashford and D. A. Case, *J. Phys. Chem. B*, 2000, **104**, 3712–3720.
- 60 J. Weiser, P. S. Shenkin and W. C. Still, *J. Comput. Chem.*, 1999, **20**, 217–230.
- 61 T. Hou and R. Yu, *J. Med. Chem.*, 2007, **50**, 1177–1188.
- 62 B. R. Miller III, T. D. McGee Jr, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, *J. Chem. Theory Comput.*, 2012, **8**, 3314–3321.
- 63 H. Gohlke, C. Kiel and D. A. Case, *J. Mol. Biol.*, 2003, **330**, 891–913.
- 64 T. J. Hou, N. Li, Y. Y. Li and W. Wang, *J. Proteome Res.*, 2012, **11**, 2982–2995.
- 65 K. Pearson, *Philos. Mag.*, 1901, **2**, 559–572.
- 66 C. Bergonzo, N. M. Henriksen, D. R. Roe, J. M. Swails, A. E. Roitberg and T. E. Cheatham III, *J. Chem. Theory Comput.*, 2014, **10**, 492–499.
- 67 D. M. F. vanAalten, J. B. C. Findlay, A. Amadei and H. J. C. Berendsen, *Protein Eng.*, 1995, **8**, 1129–1135.
- 68 T. Ichiye and M. Karplus, *Proteins: Struct., Funct., Genet.*, 1991, **11**, 205–217.
- 69 H. Sun, S. Tian, S. Zhou, Y. Li, D. Li, L. Xu, M. Shen, P. Pan and T. Hou, *Sci. Rep.*, 2015, **5**, 8457.
- 70 X. Kong, H. Sun, P. Pan, S. Tian, D. Li, Y. Li and T. Hou, *Phys. Chem. Chem. Phys.*, 2016, **18**, 2034–2046.
- 71 W. Fu, L. Chen, Z. Wang, Y. Kang, C. Wu, Q. Xia, Z. Liu, J. Zhou, G. Liang and Y. Cai, *Phys. Chem. Chem. Phys.*, 2017, **19**, 3649–3659.
- 72 E. Chovancova, A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor and J. Damborsky, *PLoS Comput. Biol.*, 2012, **8**, e1002708.
- 73 S. Kumar, *J. Comput. Chem.*, 1992, **13**, 1011–1021.
- 74 M. Souaille and B. T. Roux, *Comput. Phys. Commun.*, 2001, **135**, 40–57.
- 75 B. J. Grant, A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon and L. S. D. Caves, *Bioinformatics*, 2006, **22**, 2695–2696.
- 76 D. Shi, Q. Bai, S. Zhou, X. Liu, H. Liu and X. Yao, *Proteins: Struct., Funct., Bioinf.*, 2018, **86**, 43–56.
- 77 Q. Wang, Y. Li, J. Xu, Y. Wang, D. Shi, L. Liu, E. L.-H. Leung and X. Yao, *Proteins*, 2019, **87**, 3–11.
- 78 J.-R. Fan, Q.-C. Zheng, Y.-L. Cui, W.-K. Li and H.-X. Zhang, *J. Biomol. Struct. Dyn.*, 2015, **33**, 2360–2367.
- 79 V. Miletić, I. Odorčić, P. Nikolić and Ž. M. Svedružić, *PLoS One*, 2017, **12**, e0174410.
- 80 H. Sun, Y. Li, M. Shen, D. Li, Y. Kang and T. Hou, *J. Chem. Inf. Model.*, 2017, **57**, 1895–1906.
- 81 R. C. Bernardi, M. C. R. Melo and K. Schulten, *Biochim. Biophys. Acta, Gen. Subj.*, 2015, **1850**, 872–877.
- 82 D. R. Roe, C. Bergonzo and T. E. Cheatham, *J. Phys. Chem. B*, 2014, **118**, 3543–3552.
- 83 S. A. Adcock and J. A. McCammon, *Chem. Rev.*, 2006, **106**, 1589–1615.
- 84 G. M. Torrie and J. P. Valleau, *J. Chem. Phys.*, 1977, **66**, 1402–1408.
- 85 G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187–199.
- 86 J. Kästner, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 932–942.