Journal of Molecular Graphics and Modelling 51 (2014) 193–202

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Investigating the binding preferences of small molecule inhibitors of human protein arginine methyltransferase 1 using molecular modelling

Wei Hong[a,b], Jingyang Li[c], Charles A. Laughton[d], Lee Fah Yap[b], Ian C. Paterson[b], Hao Wang[c,*]

[a] School of Chemistry and Chemical Engineering, Beifang University of Nationalities, Yinchuan, PR China
[b] Department of Oral Biology and Biomedical Sciences, and Oral Cancer Research and Coordinating Centre, Faculty of Dentistry, University of Malaya, Kuala Lumpur, Malaysia
[c] School of Pharmacy, Ningxia Medical University, Yinchuan, PR China
[d] School of Pharmacy and Centre for Biomolecular Sciences, University of Nottingham, Nottingham, UK

ARTICLE INFO

Article history:
Accepted 28 May 2014
Available online 4 June 2014

Keywords:
PRMT1 inhibitors
Homology modelling
Active Site Pressurisation
Molecular dynamic simulation
Binding mode

ABSTRACT

Protein arginine methyltransferases (PRMTs) catalyse the methylation of arginine residues of target proteins. PRMTs utilise S-adenosyl methionine (SAM) as the methyl group donor, leading to S-adenosyl homocysteine (SAH) and monomethylarginine (mMA). A combination of homology modelling, molecular docking, Active Site Pressurisation, molecular dynamic simulations and MM-PBSA free energy calculations is used to investigate the binding poses of three PRMT1 inhibitors (ligands 1–3), which target both SAM and substrate arginine binding sites by containing a guanidine group joined by short linkers with the SAM derivative. It was assumed initially that the adenine moieties of the inhibitors would bind in sub-site 1 (PHE44, GLU137, VAL136 and GLU108), the guanidine side chain would occupy sub-site 2 (GLU 161, TYR160, TYR156 and TRP302), with the amino acid side chain occupying sub-site 3 (GLU152, ARG62, GLY86 and ASP84; pose 1). However, the SAH homocysteine moiety does not fully occupy sub-site 3, suggesting another binding pose may exist (pose 2), whereby the adenine moiety binds in sub-site 1, the guanidine side chain occupies sub-site 3, and the amino acid side chain occupies sub-site 2. Our results indicate that ligand 1 (pose 1 or 2), ligand 2 (pose 2) and ligand 3 (pose 1) are the predominant binding poses and we demonstrate for the first time that sub-site 3 contains a large space that could be exploited in the future to develop novel inhibitors with higher binding affinities.

1. Introduction

Protein arginine methyltransferases (PRMTs) can catalyse the methylation of arginine residues of a variety of proteins involved in cell signalling, RNA splicing and gene regulation, such as histones H2A, H3 and H4 [1]. All PRMTs utilise S-adenosyl methionine (SAM) as the methyl group donor, leading to S-adenosyl homocysteine (SAH, Fig. 1) and monomethylarginine (mMA) [2]. Further methylation often follows, with Type I PRMTs delivering asymmetrical dimethylarginine (aDMA) and Type II PRMTs symmetrical dimethylarginine (sDMA), respectively [2,3]. PRMT1 is the predominant Type I PRMT in mammals and its activity is deregulated in some human pathologies, such as cardiovascular disease and cancer [4,5]. Therefore, specific PRMT1 inhibitors could prove useful as therapeutic agents.

The crystal structure of human PRMT1 has not been resolved, but the crystal structures of rat PRMT1 (rPRMT1, PDB 1OR8) [6] and rat PRMT3 (rPRMT3, PDB 1F3L) [7] homologues are available. Both these crystal structures show that the active site of PRMT can be divided into two parts, namely the SAH and substrate arginine binding sites. These two sites are close to each other, thus enabling the methyl group to be transferred from SAM to the substrate arginine (Fig. 1A). Recently, three bisubstrate inhibitors of hPRMT1 (ligands 1–3, Fig. 1B; IC_{50} ~5 μM for each inhibitor) have been described [8]. These ligands target both SAM (blue and yellow in Fig. 1C and D) and substrate arginine (red in Fig. 1C and D) binding sites by containing a guanidine group joined by short linkers with the SAM derivative. When these three ligands were designed, it was assumed that the adenine moieties of the ligands would bind in sub-site 1 (composed of PHE44, GLU137, VAL136 and GLU108, blue), the guanidine side chain would occupy sub-site 2 (composed of GLU 161, TYR160, TYR156 and TRP302, red), with the amino acid side chain occupying sub-site 3 (composed of GLU152, ARG62, GLY86 and ASP84, yellow), as shown in Fig. 1C; in

* Corresponding author. Tel.: +86 9516880582; fax: +86 9516980193.
E-mail address: paxhw@yahoo.co.uk (H. Wang).

http://dx.doi.org/10.1016/j.jmgm.2014.05.010
1093-3263/

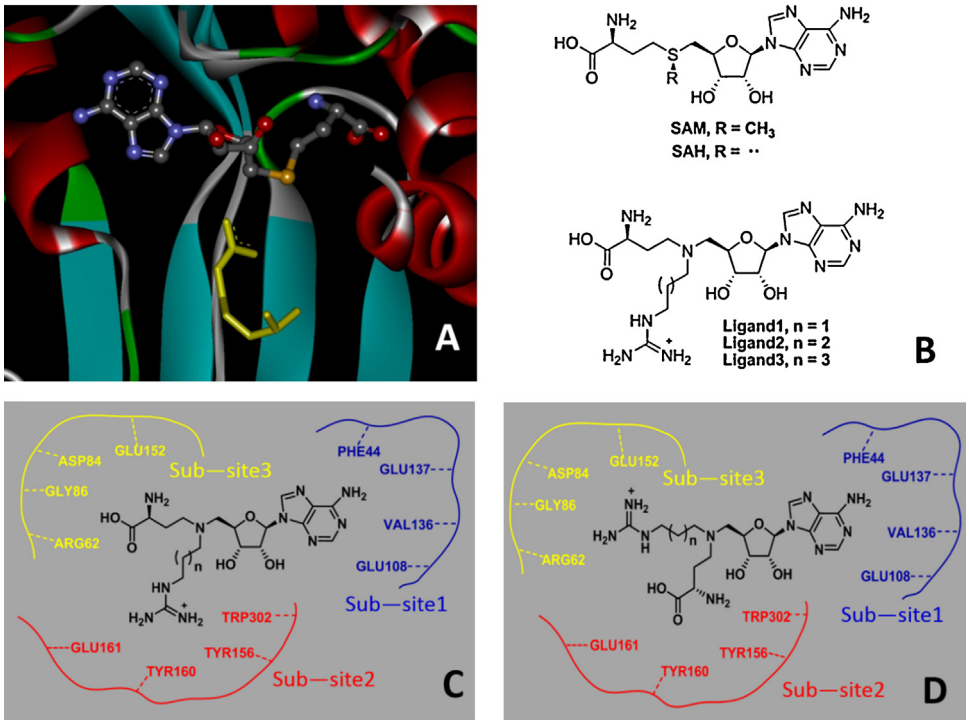

Fig. 1. (A) Active site of PRMT1 in complex with SAH (ball and stick) and substrate arginine (yellow tube); (B) the chemical structures of SAM and related compounds; (C) binding pose 1; (D) binding pose 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

the present study, this is referred to as binding mode pose 1. However, the cavity of sub-site 3 in Fig. 2 is in fact quite large and the SAH homocysteine moiety only occupies part of the space. Therefore, another binding pose is possible (pose 2), whereby the adenine moiety binds in sub-site 1, the guanidine side chain occupies subsite 3, and the amino acid side chain occupies sub-site 2. A clear understanding of how these inhibitors occupy the PRMT1 binding site is required to fully understand the mechanism of inhibition and to inform medicinal chemistry efforts to synthesise even more effective inhibitors.

Molecular modelling and computational aided drug design have been used extensively to study the reaction mechanism, discovery of potential inhibitors, and interactions between ligands with arginine methyltransferases [9–13]. Consideration of receptor flexibility is essential in computational aided drug design, and it is particularly important when the crystal structure of the receptor is not available [14–16]. Various docking software can take receptor flexibility into account during the docking process. However, they cannot fully explore the binding site conformations and are not suitable for large chemical database screening. An alternative approach is to consider the flexibility of receptor prior to the docking progress using molecular dynamic simulations [17,18]. However, with limited computational time, the accessible trajectories (typically in the nanosecond to microsecond time scale) suffer from not being able to overcome high conformational barriers in the potential energy surface of the receptors and complexes. This limits the technique to exploring only conformations of the ensemble that are around the starting structures. An additional problem is that when there is no ligand binding to maintain the conformation of the active site, the conformation of the receptor will tend to convert to that of the apo form during the simulation, which was an issue in the current study. Active Site Pressurisation (ASP) [19] was created recently to examine the flexibility of the active sites of receptors. The basic idea of ASP is to 'pump' Lennard–Jones (LJ) particles into a protein cavity and see how this expands by exploiting natural directions of the weakest resistance in the protein structure, which usually cannot be achieved by unperturbed MD with limited simulation time.

In the present study, we used a combination of homology modelling, molecular docking, ASP, molecular dynamic simulations and MM-PBSA (Molecular Mechanics – Poisson–Boltzmann Surface Area) free energy calculations to investigate the binding poses of

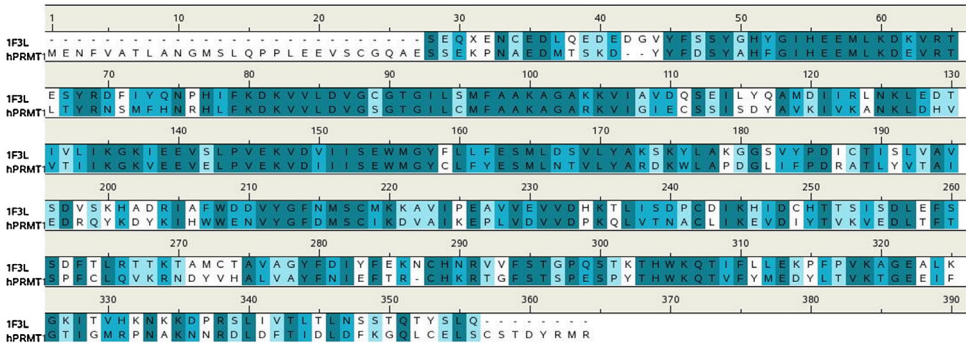

Fig. 2. Sequence alignment of hPRMT1 and rPRMT3 (PDB 1F3L).

the three PRMT1 inhibitors (ligands 1–3). Our results indicate that ligand 1 (pose 1 or 2), ligand 2 (pose 2) and ligand 3 (pose 1) are the predominant binding poses and this information should guide medicinal chemistry efforts to generate inhibitors with higher binding affinities.

2. Methods

2.1. Protein modelling

The crystal structure of rat PRMT subtype 3 (rPRMT3 PDB 1F3L) was used as the template to build the homology model of hPRMT1 using Modeller 9.11 [20]. The sequences of hPRMT1 and rPRMT3 were aligned using the align2d programme within Modeller. The 3D structure of the hPRMT1 model was generated and loops were refined by Modeller functions and the lowest value of the MODELLER objective function conformation was selected and used as the starting point for further optimisation.

2.2. Molecular dynamic simulations of the homology model of hPRMT1

The homology model of hPRMT1 was explicitly solvated in a truncated octahedral box of TIP3P model water (at least 12 Å from the complex to avoid periodic artefacts from occurring) by using Amber 12 with the amber ff12SB force field [21], and 13 Na^+ ions were added to neutralise the charges of the system by using the "addions" command line within the tleap module (AmberTools 12, which adds counterions around the complex using a Coulombic potential on a grid). The final system contained 361 protein residues, 13 Na^+ ions and 17,168 water molecules, giving a total of 57,296 atoms. The whole system was first optimised by energy minimisation, followed by 28 ns molecular dynamic simulation.

The standard equilibration protocol used is as follows:

(1) The solvent was energy minimised by 50 steps of the steepest descent method and then followed by 10,000 steps of the conjugate gradient method with the default nonbonded cutoff of 8 Å, while the solute was held fixed.
(2) The entire system was energy minimised with the same settings as the previous step.
(3) The solvent was subjected to a short (20 ps) MD simulation at a temperature of 100 K, under constant pressure conditions. For this and all following simulations, MD simulations were performed with explicit solvent models and in the NPT ensemble (T = 300 K; P = 1 atm). Periodic boundary conditions (PBC) and particle-mesh-Ewald method (PME) [22] were used to model long-range electrostatic effects, with the temperature coupled to an external bath using a weak coupling algorithm [23]. The cutoff non-bonded interaction was set as 8 Å. The bond interactions involving H-atoms were constrained by using the SHAKE algorithm.
(4) Over 20 ps, the solvent temperature was raised to 300 K. During both this phase and the last, position restraints on every solute atom (force constant 100 kcal/mol/$Å^2$) maintained them in their energy-minimised conformations.
(5) Over a series of 20 ps constant-pressure simulations at 300 K, the restraints on the solute were gradually relaxed (50, 25, 10, 5, 2, and then 1 kcal/mol/$Å^2$).
(6) The final 200 ps MD simulations were performed as the equilibration runs without any restraints.

After the energy minimisation and equilibrations, production MD was run for 28 ns in an NPT ensemble at 1 atm and 300 K. The time step necessary to solve the Newton's equations was chosen to be equal to 2 fs and the trajectory files were collected every 10 ps for the subsequent analysis. All trajectory analyses were performed with the Ptraj module in the AmberTools 12 and examined visually using VMD software [24].

2.3. Active Site Pressurisation methodology

The current version of ASP is implemented in the AMBER 12 Nucleic acid builder (NAB) language. Initially, ASP generates a grid box around the active site. With this version, the LJ particles grid can be of an arbitrary shape, defined for example as a volume around a template ligand in the active site. This grid usually should be larger than the active site to ensure the flexibility of the receptor can be fully explored. In the first step, ASP fills the active site (empty space) by "switching on" LJ particles at every grid point where they have no repulsive interaction with the protein. At the grid points that define the boundary layer, "virtual" LJ particles are placed which evaluate and record interactions with the protein but do not actually exert a force on it. A series of MD simulations (ASP cycles) are then performed. During each ASP cycle the protein interacts with all the "switched on" LJ particles while the "virtual" particles evaluate and accumulate the interaction energies they would have experienced had they been "switched on". After each ASP cycle, all "virtual" LJ particle sites that have accumulated energies below a set threshold (or the single lowest energy particle, if the threshold is exceeded) are "switched on", and the boundary surface recomputed. Full details of the theory of ASP have been described previously [19].

In the present study, the SAH and substrate arginine (overlaid by the reference of the crystal structure of rPRMT3) were used to generate the particles grid. In the ASP simulation, the non-bonded cutoff between protein and grid particles was set at 12 Å, and between protein atoms set at 25 Å. The MD simulations use an MM-GBSA implicit solvent model, and all C-alpha atoms were restrained to their initial coordinates with a force constant of 10 kcal/mol/$Å^2$. Before any LJ particles are activated, 100 steps MD simulations were performed on the system. In total, 1480 LJ particles were added in 585 ASP cycles, and between each ASP cycle, 100 steps MD simulations were performed to relax the system. After all LJ particles were added, another 895 'empty' ASP cycles (without adding any more LJ particles) were continually performed to relax the system. The ASP input file is available in the supplementary information.

2.4. Molecular docking

GOLD [25] (v 5.1; Genetic Optimisation for Ligand Docking) is a well-known protein–ligand docking programme released by the Cambridge Crystallographic Data Centre (CCDC). In this study, SAH and substrate arginine were aligned to hPRMT1 (homology model or ASP expanded) by overlaying the crystal structure of rat-PRMT3 to indicate the binding site (all protein atoms within 5.0 Å). GOLD was used to dock each ligand (ligands 1–3) 10 times, starting each time from a different random population of ligand orientations and using the default automatic genetic algorithm parameter settings. All torsion angles in each compound were allowed to rotate freely.

2.5. Molecular dynamic simulations of ligand-hPRMT1 complexes

The final complexes of ligand and ASP expended hPRMT1 obtained following molecular docking (ligands 1–3, with binding poses 1 and 2) were stabilised and refined by MD simulations for 20 ns each, following the procedure described above. To obtain the

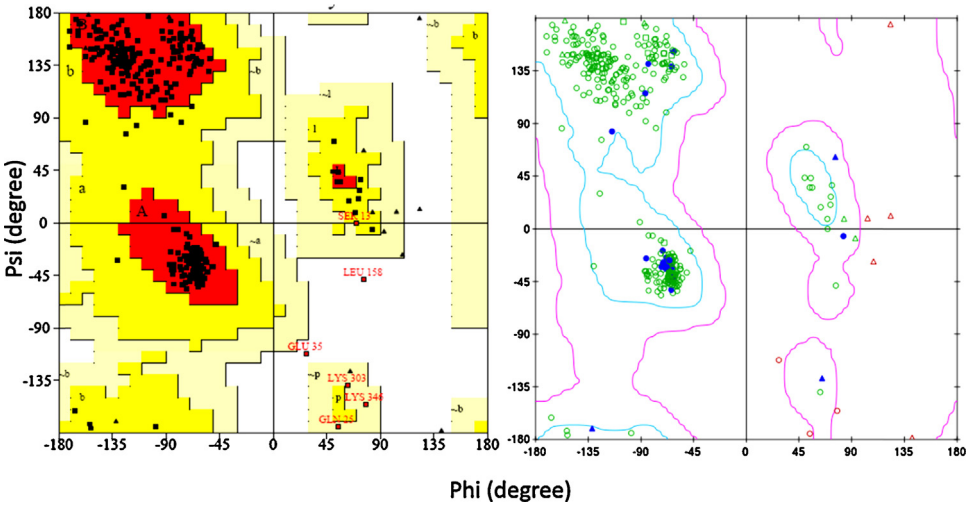

Fig. 3. Ramachandran plot of hPRMT1 generated by PROCHECK (left), and the residues in the active site are highlighted in blue (right, generated by Discovery Studio, Accelrys [35]). (For interpretation of the references to color in the text, the reader is referred to the web version of the article.)

parameters for the ligands, QM calculations were performed by using the B3LYP 6-31G* basis set within Gaussian03 [26] to optimise molecular geometries, and the atom-centred point charges were calculated to fit the electrostatic potential using RESP [27]. Other parameters were adopted from the Generalised Amber Force field (GAFF) [28]. We used the Principal Component Analysis (PCA) approach, to check for equilibration and sampling. At least the last 10 ns stable, equilibrated trajectory of each simulation was taken for energy analysis.

2.6. Binding free energy calculations

The binding free energy can be calculated using Eq. (1) from a well-equilibrated molecular dynamics simulation:

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

where G is the average free energy calculated from a set of structures taken from the equilibrated simulation (snapshots). Two popular choices to calculate the binding free energy of the snapshots are the Poisson–Boltzmann (PBSA) and generalised Born (GBSA) models [29–33], where SA corresponds to an estimation of the non-polar solvation free energy based on a simple surface area term.

In the present study, 1000 snapshots collected from the last 10 ns stable simulations at 10 ps intervals were used, and MM-PBSA was chosen to calculate the binding free energies with the Python script, MMPBSA.py, included in AMBERTOOLS 13. The nonpolar solvation free energy (ΔG_{np}) was determined by the solvent accessible surface area (SASA) according to Eq. (2).

$$\Delta G_{\text{np}} = \gamma \text{SASA} + \beta \quad (2)$$

where the surface tension γ and the offset β were set to the standard values of 0.00542 kcal/mol/Å^2 and 0.92 kcal/mol, respectively. Other options were set to default settings. Contributions to the free energy calculation from the configurational entropy, which mainly concerns rotations about single bonds, and the vibrational entropy are not included in the PBSA calculations. These two terms were ignored due to the difficulty in accurately predicting these effects. In the present study, as we are comparing a set of similar ligands in similar environments, these two terms would largely cancel out so that the accuracy would unlikely be affected significantly.

3. Results and discussion

3.1. The homology model and structure optimisation

The sequence of hPRMT1 was aligned to rPRMT3 (PDB IF3L) using the sequence alignment protocol of modeller9.11. It showed that the sequence identity is around 45.5%, and a strong similarity (64%) was observed. It should be noted that the first 30 residues are miss matched, but they are not located in the active site (Fig. 2).

The model was verified with PROCHECK [34], and the Ramachandran plot (Fig. 3) shows most residues (300 in 361 residues, 90.9%) are in the allowed regions (red), 24 residues (7.3%) are in the additional allowed regions (yellow), and 4 resides (1.2%) are in the generously allowed regions (dark yellow). Only 2 residues (0.6%) are in the disallowed regions (white), which are LEU158 and GLU35 (not present in the active site). The residues in the active site are highlighted in blue generated by Discovery Studio, Accelrys [35]).

The structure was also verified by Profile-3D (built in Discovery Studio, Accelrys) [35] and the self-compatibility score of this model is 127.52 (the highest score is 164.513 and the lowest is 74.0308). There are several sections of residues close to or lower than zero, which indicates they may not in the 'right' position (Fig. 4). We were concerned about residues TYR43 and PHE44, because they are located within the active site and have VDW interaction with adenine, suggesting further optimisation of the structure may be necessary.

3.2. Verification of the homology model by using molecular docking

SAH was docked back to the active site. As shown in Fig. 5, SAH was docked successfully into the homology model of hPRMT1. The SAH binding site is composed of two parts, sub-sites 1 and 3, which are indicated by blue and yellow respectively (Fig. 1B). Sub-site 1 accommodates the adenine moiety, and sub-site 3 accommodates the SAH homocysteine moiety. The SAH homocysteine moiety only occupies a part of sub-site 3, which can be expanded to ASP84. More detailed H-bond analysis indicated that the carboxylate group of the SAH homocysteine moiety forms two H-bonds with Arg62, and one H-bond is formed between the amino group of the SAH homocysteine moiety and GLU152, and both of them have static electric interactions. The adenine moiety forms two H-bonds with GLU137 and one with VAL136, respectively. The adenosine ribose

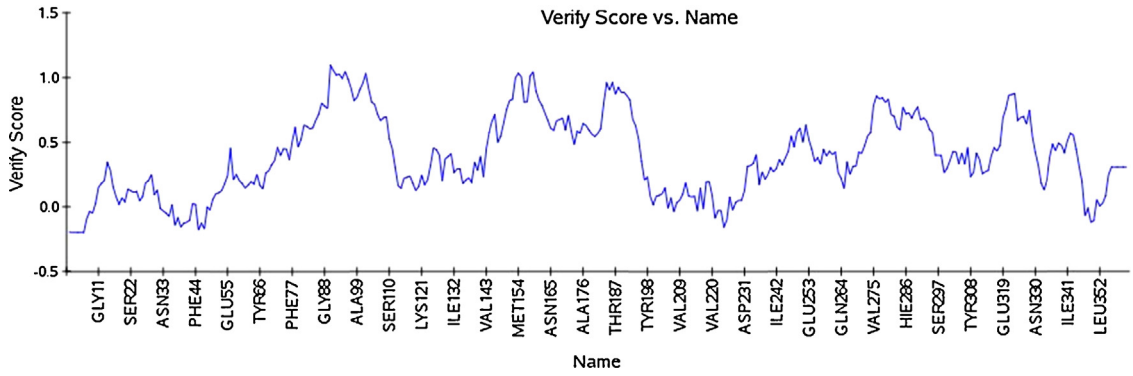

Fig. 4. 3D profiles verified the results of hPRMT1 model. Residues with positive compatibility score are reasonably folded.

moiety forms one H-bond with GLU108. The H-bond analysis is consistent with a previous report [36], which demonstrates that GOLD can generate reliable docking data.

We next docked ligands 1–3 into the active site using Gold, and the top 10 binding poses were recorded. Among these binding poses, ligands 1 and 2 showed both poses 1 and 2 exist, however, ligand 3 could only be docked in binding pose 2 using the current settings. There are two possible explanations for this result. First, the substrate arginine binding site is not large enough for the long guanidine side chain of ligand 3 to bind and; second, the flexibility of the receptor was not considered in the docking. In order to consider the flexibility of the receptor, ASP was performed, but prior to this, MD simulation was used to optimise the homology model.

3.3. Molecular dynamic simulations studies on hPRMT1

Following 28 ns free MD simulation, the potential energy along the trajectory was plotted and the results showed that the energy decreased in the first 6 ns and then became stable (Fig. 6). Comparing the active site of the homology model and the samples from the trajectory, we observed that the active site closed during the simulation, with MET163 and TYR47 moving closer to each other from around 6.48 Å to 3.83 Å (the closest distance between heavy atoms, Fig. 7). This is important because SAH binds between the two residues and, when they are in close proximity, SAH cannot be docked into the active site. This is because the bound ligands (SAH and substrate arginine) were not included in the simulation, so the conformation of hPRMT1 changed to the apo form. Normally, when SAH binds to the active site, hPRMT1 will undergo a conformational change to fit the ligand. This result highlights the need to consider the flexibility of the receptor in computer simulations, particularly when performing molecular docking studies.

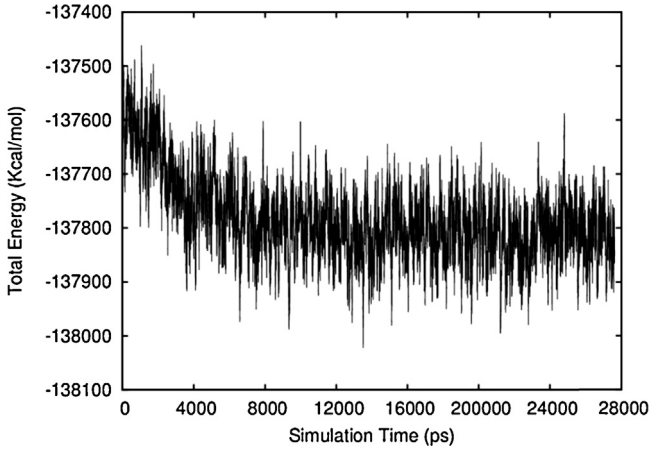

Fig. 6. The potential energy profile along the free molecular dynamic simulation of the homology model of hPRMT1.

3.4. ASP expansion

To fully explore the flexibility of the receptor, we used ASP to expand the active site before performing molecular docking. In total 1480 LJ particles were 'pumped' to expand the active site, and the energy along the trajectory was analysed. After 1373 (480 cycles) particles were pumped in, the energy increased dramatically and then became stable in a higher state.

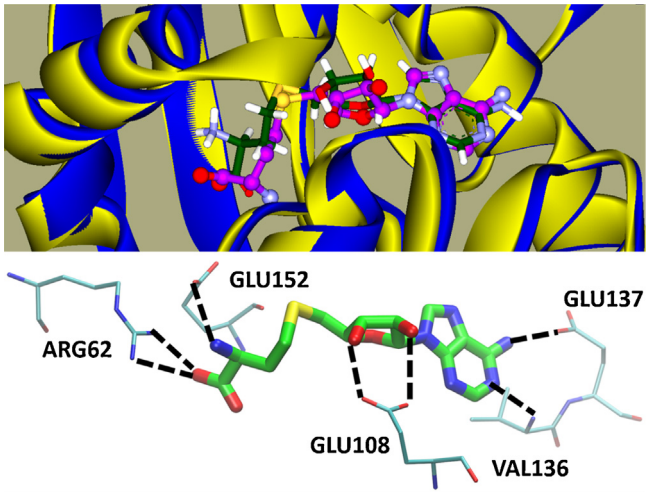

Fig. 5. Overlay of ratPRMT3-SAH (blue ribbons for protein and dark green for SAH) and hPRMT1-SAH (yellow ribbons for protein and purple for SAH) (top). The docking of SAH to the homology model of hPRMT1 (bottom). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

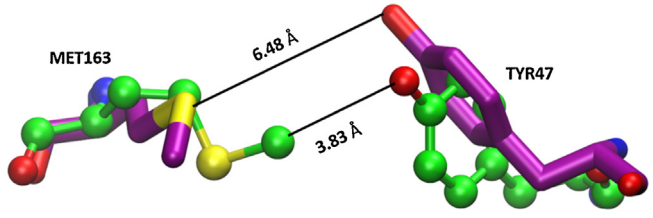

Fig. 7. The structures of homology model (purple, sticks) and the final snapshot of the MD simulation (green, ball and sticks) were overlaid, and the active site is closed when MET163 and TYR47 move closer to each other. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

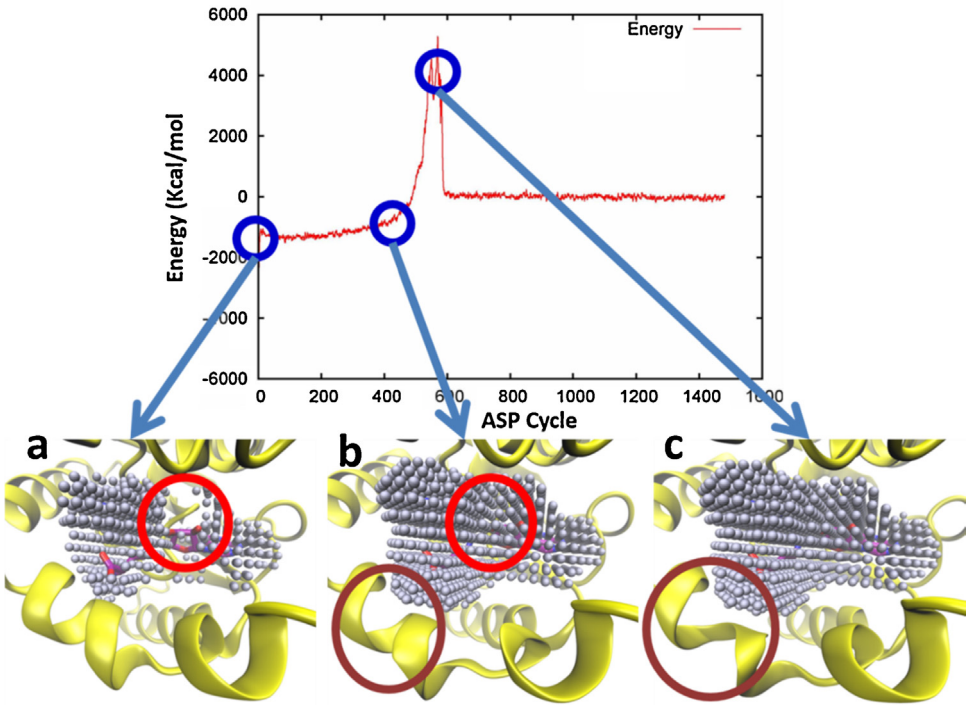

Fig. 8. The energy profile along the ASP simulation. (a) LJ particles were activated in the binding site, except the site corresponding to the middle of the adenosine ribose moiety (in sub-site 1, red circle). (b) After 480 ASP cycles (1373 particles were pumped in), the energy starts increasing dramatically. (c) At 600 ASP cycles (in total 1480 LJ particles were activated), the energy is increased dramatically and the α helix near sub-site 3 (brown circle) is distorted. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Through the analysis of the trajectory, correlations between the energy curve and the conformational changes were observed. Ligand 2 (with pose 1) was overlaid into the active site for a reference and this showed that originally the particles were 'pumped' into the binding site, except the site corresponding to the middle of the adenosine ribose moiety (in sub-site 1, Fig. 8A). This is because MET163 and TYR47 moved towards each other to close the site as shown in Fig. 7. As more particles were pumped into the space, most of them were filled in the position of the adenosine ribose moiety, which indicates this part of receptor is the most flexible area. The other parts (sub-sites 2 and 3) were also expanded, but not as significantly as sub-site 1. During this process (Fig. 8B), the energy is stable, which indicates that expansion at this stage is with a small resistance. Subsequently, more particles were pumped into the remaining space, and the site of the amino acid side chain (sub-site 3) was expanded. However, this expansion became more 'difficult', as it was noticed that the α helix near the amino acid side chain binding site was distorted and this corresponded to the energy peak in Fig. 8C. After all particles were pumped in, the protein was relaxed and the energy was stable in a higher energy state. The result of ASP expansion suggests that conformation for drug discovery should be obtained before the 480th cycle and, if a series of snapshots are collected for receptors, flexible docking can be achieved. The active site volume was measured by using mdpocket [37]. It can be seen that the active site was expanded from around 800 Å^3 to around 1800 Å^3 within 480 ASP cycles, and then the increment became slow and stabilised at around 2100 Å^3 (Fig. 9). On comparing the snapshot of the 480th ASP cycle, the structure of the homology model and the crystal structure of rPRMT3 (PDB IF3L) (supplementary Fig. 1a and b), it was observed that the backbone of the protein in the binding site did not change significantly, and most of the obvious differences come from the side chain movements. The RMSD of the homology model with respect to the crystal structure of rPRMT3 (PDB IF3L) is only 1.85 Å, and the one for the structure after MD equilibration and pressurisation is 3.03 Å. In the active site, the secondary structure of the protein remained stable, and only some loops showed a small degree of flexibility, particularly within the region from Val82 to ILE91. Taken together, these observations suggest that ASP expands or perturbs the binding site 'carefully' to avoid major conformational distortions.

3.5. *Molecular dynamic simulations studies on complexes and free energy calculations*

Ligands 1–3 were docked with Gold into the ASP expanded hPRMT1 (the snapshot of the 480th cycle) and both poses 1 and 2 were observed for each case. Each complex (ligands 1–3 with binding poses 1 and 2) was used as the starting conformation for 20 ns molecular dynamic simulation with explicit solvent.

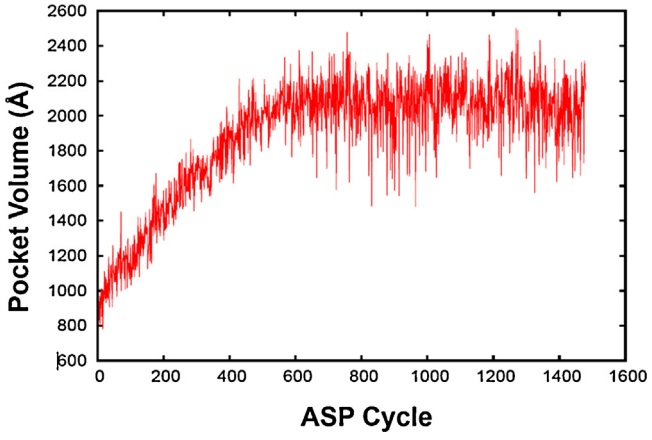

Fig. 9. The active site volume changes along the ASP expansion.

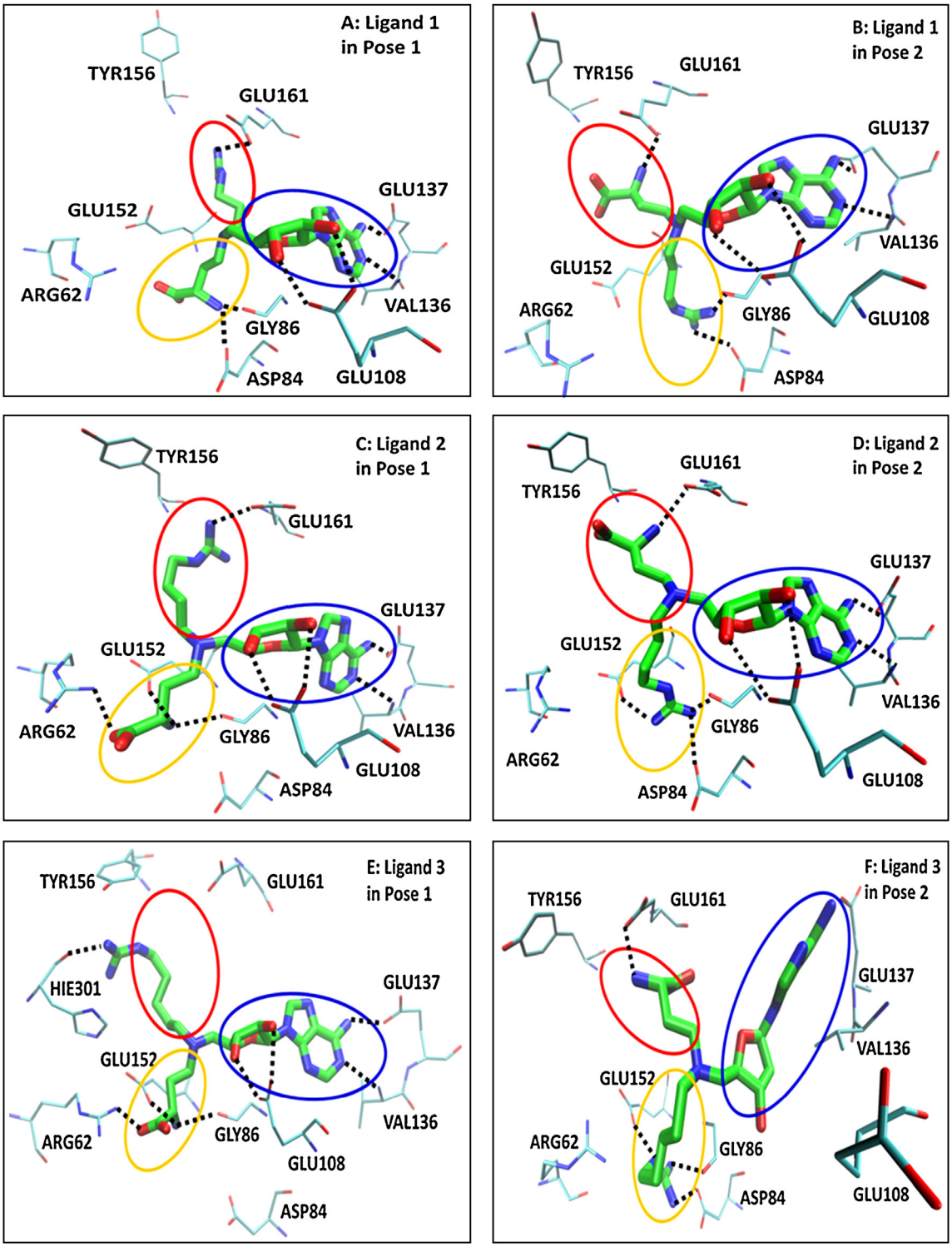

Fig. 10. The binding poses of ligands 1–3. The hydrogen bonds are represented by dotted lines. The sub-sites 1, 2 and 3 are indicated by the blue, red and yellow circles, respectively.

After the simulations were performed, the active site volume for each binding pose in the equilibrium trajectories was analysed. The average volumes for all the ligands were between 1100 and 1300 $Å^3$, which are larger than active site volume before ASP expansion and smaller than after 480 ASP cycles (Fig. 9). Therefore, ASP can expand the active site to maximise the potential volume to accommodate large ligands, and the subsequent MD can optimise the conformation of the active site to fit the bound ligand (induced fit theory).

For binding pose 1 (Fig. 10A, C and E), all three ligands form very similar interactions with hPRMT1. The guanidine groups (bound in sub-site 2, red circle) have electrostatic interactions and form H-bonds with GLU161, however, this interaction is not stable for ligand 3 because longer carbon chain provides more flexibility and increases the distance between guanidine group and GLU161. Instead, ligand 3 forms one hydrogen bond with the backbone of HIE301. The amino acid side chains (bound in sub-site 3, yellow circle) of ligands 1–3 have electrostatic interactions and form H-bonds

with ARG62, GLY86, and GLU152 (or ASP84 for ligand 1). The binding of adenosine ribose moieties (bound in sub-site 1, blue circle) of these three ligands are also very similar, and they all form H-bonds with GLU137, VAL136 and GLU108. The average distances between ligands and hPRMT1 are shown in supplementary Table S1.

For pose 2 (Fig. 10B, D and F), the binding of ligands 1 and 2 share some similarities. The amino groups on amino acid side chains (bound in sub-site 2, red circle) of these two ligands form H-bonds and electrostatic interactions with GLU161. The guanidine groups (bound in sub-site 3, yellow circle) of ligand 1 forms H-bonds with GLY86 and ASP84, and have electrostatic interactions with ASP84 and GLU152. Ligand 2 forms one more H-bond with Glu152 compared to ligand 1, possibly because the carbon chain of guanidine group of ligand 2 is longer than that of ligand 1. The adenosine ribose moieties (bound in sub-site 1, blue circle) of ligands 1 and 2 form similar interactions with GLU137, VAL136 and GLU108. Ligand 3 in pose 2 is significantly different from the other two ligands. The amino acid side chain shifts towards the adenosine ribose moieties instead occupying the arginine binding site (sub-site 2, red circle). It forms one H-bond with GLU161. The adenosine ribose moieties (sub-site 1, blue circle) also shifts towards the amino acid side chain, while the guanidine group (bound in sub-site 3, yellow circle) has similar interactions with those of ligands 1 and 2. This significant difference is perhaps because of the long carbon chain of the guanidine group. In order to gain enough space for the guanidine group, the whole molecule shifts towards the adenosine ribose moieties binding site (sub-site 1, blue circle).

MM-PBSA was used to calculate the binding free energies for these binding poses (Table 1), and the contribution of binding free energy of each residue was calculated and only those lower than −1.5 kcal/mol are presented (Fig. 11). The free energy calculations show that the total binding free energies of both binding poses of ligand 1 are very similar, which indicates that these two binding poses may exist at the same time. In pose 1 (Fig. 11A), there are the free energy contributions from ASP84, ARG62 and GLY86 (interact with amino acid side chain), and from the GLU161 (interact with guanidine group). In pose 2 (Fig. 11B), the contribution from ASP84 (interact with guanidine group) is large, and the contribution from GLY86 (interact with guanidine group) is similar to pose 1, whilst the contribution from GLU161 (interact with amino acid side chain) becomes weak.

For ligand 2, the free energy of pose 2 is lower than that of pose 1, which indicates that the binding pose 2 of ligand 2 is its predominant pose. In pose 1 (Fig. 11C), ARG62 and GLY86 have strong interactions with the amino acid side chain, and GLU161 has a strong interaction with the guanidine group. In pose 2 (Fig. 11 D), the guanidine group establishes a strong interaction with ASP84, maintains the interaction with GLY86, and establishes a strong interaction with GLU152. Similar to ligand 1 in pose 2, the free energy contribution from GLU161 (interacts with amino acid side chain) becomes weak.

It was noticed that in both poses 1 and 2 of ligand 1 (Fig. 10A and B), the orientation of GLU152 is always pointed to the guanidine group of the ligand, and the free energy contribution of this residue is small (Fig. 11A and B). While in pose 1 of ligand 2 (Fig. 10C), because the longer carbon chain increases the distance between GLU152 and the guanidine group, the orientation of GLU152 does not follow the guanidine group because of the existence of ARG62, so that the guanidine group only has strong electrostatic interactions with one negative charged residue (GLU161). While in pose 2 of ligand 2 (Fig. 10D), the guanidine group can have electrostatic interactions with two negative charged residues (GLU152 and ASP84). It is interesting to note that the free energy contribution of GLU152 is stronger in pose 2 than in pose 1 of ligand 2 (Fig. 11C and D), and this is perhaps one of the reasons that can explain why

Table 1
The free energies (kcal/mol, mean ± SEM) calculated by MM-PBSA and experimental IC_{50}s (mean ± SEM) for ligands 1–3 in both binding poses 1 and 2.

Ligand	Pose	ΔE_{vdw}	ΔE_{ele}	ΔG_{pb}	ΔG_{np}	ΔG_{gas}	ΔG_{solv}	ΔG	IC_{50} (μM) [8]
1	Pose 1	−63.53 ± 0.14	−390.15 ± 0.73	392.34 ± 0.58	−6.69 ± 0.003	−453.68 ± 0.75	385.65 ± 0.58	−68.02 ± 0.27	6.2 ± 3.8
	Pose 2	−66.78 ± 0.13	−357.08 ± 0.58	358.85 ± 0.35	−6.73 ± 0.004	−423.86 ± 0.54	352.11 ± 0.35	−71.75 ± 0.29	
2	Pose 1	−67.04 ± 0.23	−367.10 ± 0.63	377.54 ± 0.58	−6.99 ± 0.006	−434.14 ± 0.63	370.55 ± 0.58	−63.59 ± 0.36	2.9 ± 0.8
	Pose 2	−63.59 ± 0.19	−361.40 ± 0.54	358.18 ± 0.41	−7.08 ± 0.006	−424.99 ± 0.51	351.10 ± 0.41	−73.88 ± 0.30	
3	Pose 1	−68.93 ± 0.14	−372.14 ± 0.60	379.23 ± 0.60	−7.25 ± 0.004	−441.08 ± 0.59	371.98 ± 0.60	−69.09 ± 0.22	5.6 ± 4.5
	Pose 2	−60.88 ± 0.18	−302.13 ± 0.67	317.17 ± 0.75	−7.44 ± 0.010	−363.02 ± 0.68	309.73 ± 0.74	−53.29 ± 0.44	

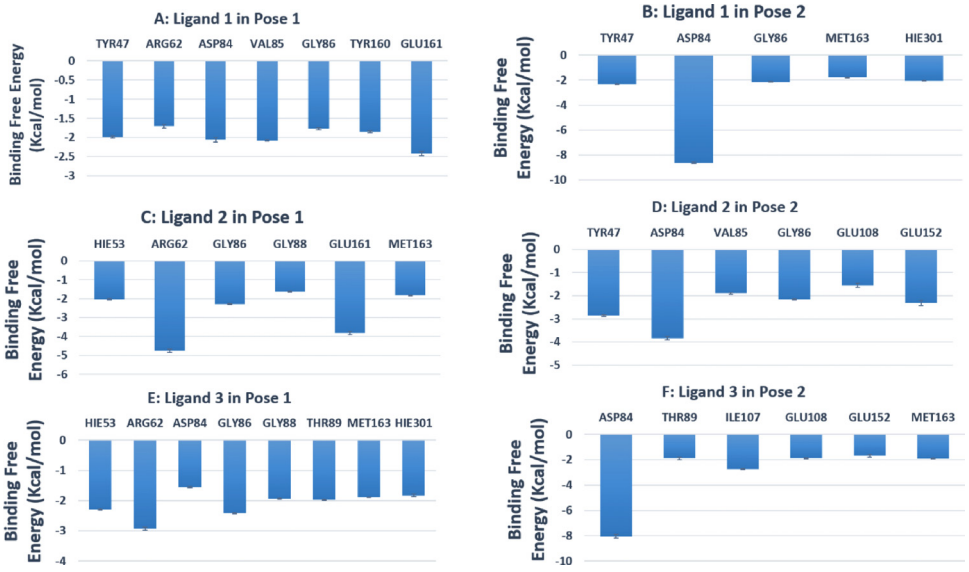

Fig. 11. The contribution of binding free energy of each residue and only those lower than −1.5 kcal/mol are presented.

both poses 1 and 2 exist in the case of ligand 1, while pose 2 is more favourable in the case of ligand 2.

For ligand 3, the free energy of pose 1 is significantly lower than that of pose 2, which indicates that binding pose 1 is its predominant pose. In pose 1 (Fig. 11E), there are large free energy contributions from ASP84, ARG62 and GLY86 (interact with amino acid side chain), and from HIE301 (interact with guanidine group). In pose 2 (Fig. 11F), there are free energy contributions from ASP84 and GLY152 (interact with guanidine group), and a weak contribution from GLU161 (interacts with amino acid side chain, −1.2 kcal/mol). Because sub-site 3 is not large enough to fully accommodate the guanidine group of ligand 3 in pose 2, in order to gain enough space, the whole molecule shifts towards the adenosine ribose moieties binding site (sub-site 1, blue circle), which leads ligand 3 into an unfavourable binding pose. Importantly, the free energies of the predominant poses of these three ligands (ligand 1 (pose 1 or 2), ligand 2 (pose 2) and ligand 3 (pose 1)) are similar, which matches the experimental data (Table 1). However, the absolute values of free energy calculation cannot be converted to binding constants, as several terms were negated (e.g. entropy contributions), so relative energies of binding are more likely to be meaningful.

A recent report [38] highlighted the important flexible regions within the dengue virus methyltransferase. However, the AdoHcy binding site of dengue virus methyltransferase is more flexible compared to human PRMT1, and is located between the loops LYS101-MET116 and ILE147-ASN153. The LYS101-MET116 loop is flexible in MD simulation in the absence of AdoHcy, but is quite stable when bound to AdoHcy. The SAH binding site of human PRMT1 is covered by two α-helices, that make the binding site more rigid than that of dengue virus methyltransferase. The most flexible part during the ASP expansion is VAL82 to ILE91 (corresponds to the volume of subsites 1 and 3). Both these two regions (LYS101-MET116 within dengue virus methyltransferase, and VAL82 to ILE91 with hPRMT1) are close to the SAH binding site, but they are different because the loop LYS101-MET116 in dengue virus methyltransferase corresponds to a more rigid α-helix on human PRMT1 (supplementary Fig. 2a and b). The loop LYS101-MET116 in dengue virus methyltransferase is flexible, and it locks the binding site after AdoHcy binding. The role of VAL82 to ILE91 region in human PRMT1 can only influence the size of the active site.

4. Conclusions

In the present study, a reliable homology model of hPRMT1 was successfully created. Ligands 1–3 were docked into the active site of hPRMT1 and both binding poses 1 and 2 were observed for ligands 1 and 2, but only binding pose 2 was observed for ligand 3. In order to identify the binding mode of these three ligands, free MD simulations were performed to optimise the homology model. Along free molecular dynamics simulations, the conformation of the model was optimised, and the binding site was closed because of the lack of bound ligand. The active site was then expanded by ASP, and the snapshot was collected before the energy increased dramatically. With this snapshot as the receptor, ligands 1–3 were docked and both poses 1 and 2 were obtained for each ligand. 20 ns molecular dynamic simulations were performed on each pose and MM-PBSA was used to calculate the binding free energies. Based on the trajectories and energy analysis, we conclude that the predominant binding poses are pose 1 or 2 for ligand 1, pose 2 for ligand 2 and pose 1 for ligand 3, but crystallography studies will be required for definitive proof.

The analysis of the binding poses of ligands 1–3 in the present study demonstrates for the first time that sub-site 3 contains a larger space than the amino acid chain if the ligand binds in pose 1, and a larger group like the guanidine group of ligand 2 (4 carbons in the chain) can be accommodated. We believe that these observations indicate a new path with which to develop new inhibitors by using medicinal chemistry to introduce larger groups to occupy this sub-binding site. Further, our data also demonstrate the usefulness of ASP to explore the active site and consider the flexibility of receptors in the drug discovery process. Therefore, similar approaches will significantly increase opportunities to identify potential ligands or different binding poses of known compounds.

Acknowledgements

This work was supported by National Natural Science Foundation of China [81260481]; Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry; Scientific Research Project for Ningxia Colleges and Universities, Department of Education of Ningxia [2011JY004]; Early Stage Project for NSFC, Beifang University of Nationalities [2012QZP08]

and University of Malaya-MOHE High Impact Research grant (UM.C/625/1/HIR/MOHE/DENT/22).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2014.05.010.

References

[1] M.T. Bedford, S.G. Clarke, Mol. Cell 33 (2009) 1–13.
[2] A. Di Lorenzo, M.T. Bedford, FEBS Lett. 585 (2011) 2024–2031.
[3] M.T. Bedford, J. Cell Sci. 120 (2007) 4243–4246.
[4] R.A. Copeland, M.P. Moyer, V.M. Richon, Oncogene 32 (2013) 939–946.
[5] O. Obianyo, T.C. Osborne, P.R. Thompson, Biochemistry 47 (2008) 10420–10427.
[6] X. Zhang, X.D. Cheng, Structure 11 (2003) 509–520.
[7] X. Zhang, L. Zhou, X.D. Cheng, EMBO J. 19 (2000) 3509–3519.
[8] J. Dowden, W. Hong, R.V. Parry, R.A. Pike, S.G. Ward, Bioorg. Med. Chem. Lett. 20 (2010) 2103–2105.
[9] R. Zhang, X. Li, Z. Liang, K. Zhu, J. Lu, X. Kong, S. Ouyang, L. Li, Y.G. Zheng, C. Luo, PLoS ONE 8 (2013) e72424.
[10] Y. Chu, G. Li, H. Guo, Can. J. Chem. 91 (2013) 605–612.
[11] J. Yao, Y. Chu, R. An, H. Guo, J. Chem. Inf. Model. 52 (2012) 449–456.
[12] J. Wang, L. Chen, S.H. Sinha, Z. Liang, H. Chai, S. Muniyan, W. Yu, C. Chou, L.Y. Yang, Y. Feng, K.K. Li, M.F. Lin, H. Jiang, Y.G. Zheng, C. Luo, J. Med. Chem. 55 (2012) 7978–7987.
[13] J. Dowden, R.A. Pike, R.V. Parry, W. Hong, U.A. Muhsen, S.G. Ward, Org. Biomol. Chem. 9 (2011) 7814–7821.
[14] M.I. Zavodszky, L.A. Kuhn, Protein Sci. 14 (2005) 1104–1114.
[15] K. Gunasekaran, R. Nussinov, J. Mol. Biol. 365 (2007) 257–273.
[16] A. Gutteridge, J. Thornton, J. Mol. Biol. 346 (2005) 21–28.
[17] K.L. Meagher, H.A. Carlson, J. Am. Chem. Soc. 126 (2004) 13276–13281.
[18] J.H. Lin, A.L. Perryman, J.R. Schames, J.A. McCammon, J. Am. Chem. Soc. 124 (2002) 5632–5633.
[19] I.M. Withers, M.P. Mazanetz, H. Wang, P.M. Fischer, C.a. Laughton, J. Chem. Inf. Model. 48 (2008) 1448–1454.
[20] A. Sali, T.L. Blundell, J. Mol. Biol. 234 (1993) 779–815.
[21] D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Götz, I. Kolossváry, K.F.Wong, F. Paesani, J. Vanicek, R.M.Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman (2012), AMBER 12, University of California, San Francisco.
[22] T. Darden, D. York, L. Pedersen, J. Chem. Phys. 98 (1993) 10089–10092.
[23] H.J.C. Berendsen, J.P.M. Postma, W.F. Vangunsteren, A. Dinola, J.R. Haak, J. Chem. Phys. 81 (1984) 3684–3690.
[24] W. Humphrey, A. Dalke, K. Schulten, J. Mol. Graphics 14 (1996) 33–38.
[25] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, J. Mol. Biol. 267 (1997) 727–748.
[26] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery, T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A. D Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. AlLaham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople; Gaussian, Inc., Wallingford CT, 2004.
[27] C.I. Bayly, P. Cieplak, W.D. Cornell, P.A. Kollman, J. Phys. Chem. 97 (1993) 10269–10280.
[28] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, J. Comput. Chem. 25 (2004) 1157–1174.
[29] P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, T.E. Cheatham, Acc. Chem. Res. 33 (2000) 889–897.
[30] W.C. Still, A. Tempczyk, R.C. Hawley, T. Hendrickson, J. Am. Chem. Soc. 112 (1990) 6127–6129.
[31] M. Orozco, F.J. Luque, Chem. Rev. 100 (2000) 4187–4226.
[32] H. Wang, C.A. Laughton, Methods 42 (2007) 196–203.
[33] H. Wang, C.A. Laughton, Phys. Chem. Chem. Phys. 11 (2009) 10722–10728.
[34] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, J. Appl. Crystallogr. 26 (1993) 283–291.
[35] R. Luthy, J.U. Bowie, D. Eisenberg, Nature 356 (1992) 83–85.
[36] A. Spannhoff, R. Heinke, I. Bauer, P. Trojer, E. Metzger, R. Gust, R. Schuele, G. Brosch, W. Sippl, M. Jung, J. Med. Chem. 50 (2007) 2319–2325.
[37] P. Schmidtke, A. Bidon-Chanal, J. Luque, X. Barril, Bioinformatics 10 (2011) 1–10.
[38] V. Luzhkov, E. Decroly, B. Canard, B. Selisko, J. Åqvist, Mol. Inform. 32 (2013) 155–164.