Insilico optimization of the first DNA-independent mechanism-based inhibitor of mammalian DNA methyltransferase Dnmt1

Patrik Nikolić[,2,4], Vedran Miletić[3,4], Ivica Odorcić[2], Željko M. Svedružić[1*]

[1] Department of Biotechnology and Faculty of Medicine, University of Rijeka, Croatia.
[2] Department of Biotechnology, University of Rijeka, Rijeka, Croatia.
[3] Department of Informatics and Faculty of Engineering, University of Rijeka, Rijeka, Croatia.
[4] Equal contributions
* corresponding author, E-mail: zeljko.svedruzic@biotech.uniri.hr   www.svedruziclab.com

ABSTRACT:

Background: Inhibitors of DNA methylation can be used to control functional organization of human genome in basic research and regenerative medicine.

Results: We describe a set of adenosyl-1-methyl-pyrimidin-2-one derivatives as novel mechanism-based suicide-inhibitors of mammalian DNA methyltransferase DNMT1. The inhibitors are designed to act as transition state analogs that can bind simultaneously to the cofactor binding site and the active site on DNMT1. The two binding sites and the mechanism-based suicide-inhibition are combined to provide highly potent and highly specific inhibition. The stability of presented DNMT1-inhibitor complex is described in detail using 58 modifications that affect flexibility and binding interactions at specific sites within the complex.

Conclusions: Presented results can guide synthesis and optimization of some highly specific mechanism-based inhibitors of mammalian DNA methylation with specific pharmacological properties.

Keywords: Epigenetics, Epigenome, cell differentiation, iPSC, molecular docking, Gromacs, molecular dynamics, CP2K, GAMESS, quantum mechanics, enzymatic mechanism, rational-drug-design, pharmacophoric groups.

1.1. Introduction

DNA methylation is a fundamental mechanism in functional organization of the human genome. DNA methylation is one of the first steps in epigenetic regulation and the most enduring epigenetic landmark . The inhibitors of DNA methylation can be used in studies of human epigenome , or in creation of induced pluripotent stem cells and cellular reprogramming . Inhibitors of DNA methylation can be also used in clinics for treatment of oncogenic transformation, viral infections, immunological disorders, or neurological and psychiatric impairments .

Dnmt1 is the principal DNA methyltransferase in mammalian cells . More than 20 different inhibitors of mammalian Dnmt1 have been described in the last 30 years . Unfortunately none of those studies gave consistent results and sustainable progress . The observed inconsistencies and the lack of progress could be in a large part attributed to assay design and regulation of Dnmt1 activity in cells. Dnmt1 in cells can interact with about 40 different proteins and with some RNA molecules . Thus, DNA methylation in cells can be affected by any change in DNA metabolism, DNA repair, chromatin organization, and cell cycle control . Cell-based studies of inhibition of DNA methylation cannot differentiate between compounds that target Dnmt1 directly from the compounds that can stall DNA methylation by causing DNA damage or other changes in DNA structure and metabolism. Thus, the screenings for inhibitors have to start with purified Dnmt1, and then well characterized compounds can be used to study inhibition of DNA methylation in cells. The screenings with purified Dnmt1 have to be designed to differentiate between compounds that bind to Dnmt1 from the compounds that interfere with DNA methylation by binding to the DNA substrate. None of the published studies of inhibition of DNA methylation has included all of the required precautions, and still none of the published studies found compounds with IC50 values significantly below 1 μM .

Mechanism-based inhibitors can give highest specificity and lowest toxicity . Different cytosine derivatives are the only true mechanism based inhibitors that have been developed in the last 40 years . The cytosine derivatives were introduced based on mechanistic similarities between DNA methyltransferase and tymidilate sythetase . The cytosine derivatives have been very useful in mechanistic studies of DNA methyltransferases , however their full applicability as inhibitors of DNA methylation was very limited due to high toxicity in cells . Very little improvements have been introduced following the initial studies, mostly due to unusually challenging mechanism of the target base attack in the process of DNA methylation . Consistent with earlier activity studies recent crystal structures showed that the target base attack depends on multiple flexible loops in the protein structure . Regulation of protein function by flexible loops in protein structure can be described very effectively by combining experimental results and computational methods .

In this study we present a novel strategy to design of a whole family of mechanism-based suicide-inhibitors. The inhibitors are designed to act as a transition-state analogue by binding simultaneously to the AdoMet site and the active site on Dnmt1 . The combination of two binding sites and the mechanism-based suicide-inhibition was chosen to provide highly potent and highly specific inhibition of DNA methyltransferases . The presented strategy is a combination of insights from extensive enzyme activity studies, recent crystal structures, and some of the most recent computational methods .

2.1. Lead structure for the mechanism-based transition-state analogue.

The structures 4.1 and 4.2 in figure 1 show the concept of mechanism-based suicide-inhibition, while the structure in figure 2 shows a realistic functional derivative of the conceptual structure 4.1. in figure 1 . In essence our mechanism-based inhibitor can mimic the last step in the catalytic cycle of DNA methyltransferases (Fig 1). This concept was first presented in our earlier paper on catalytic mechanism of Dnmt1 , and subsequently very similar concept was used to prepare novel inhibitor of histone methyltransferase DOT1L .

We will use the structure presented in Figure 2 as a lead compound in our insilico studies with a desire to design a whole family of novel improved mechanism based-inhibitors. The lead compound consists of four functional parts. The first two parts are adenosine and ribose rings that can anchor the inhibitor to the AdoMet/AdoHcys site on Dnmt1 . The third part of the inhibitor is a flexible linker that could mimic catalytic intermediates in the rate-limiting methyltransfer step . The linker is expected to bind in a 3.2 Å long cavity that represents reaction coordinate . The linker is expected to acts as a bridge between the AdoMet site and the active site . The fourth part of the lead compound is the target base ring (Fig 2). The target base ring can bind in the active site cavity, where it can form hydrogen bonds with Arg1310 and Arg1312 and a covalent adduct with active site Cys 1226 (Fig 1). The linker and the target base are expected to assure that the inhibitor will preferentially target DNA methyltransferases in cells rater than the other enzymes that can also bind adenine and ribose rings.

2.2 Evaluation of the proposed lead compound as a mechanism based inhibitor of Dnmt1.

We show that the presented lead-compound can bind to Dnmt1 and act as a transition state analogue and mechanism-based suicide-inhibitor using four different insilico approaches

(Figures 3 to 6). First, rigid-body-docking with Chimera-Vina programs showed the lead compound can dock in its cleft on Dnmt1 and form binding interactions (Figs. 3 and 4) . Second, Gromacs MM/MD studies were used to analyze how molecular flexibility and bulk solvent molecules can affect stability of Dnmt1-ligand complex (Table 1, lead compound entry). Third, specific modifications in the lead compound structure were chosen to evaluate binding interactions and flexibility at different positions within the Dnmt1-ligand complex (Tables 1-5). In all three steps the binding was described as successful if the ligand in the binding site can overlap with AdoHcys and target cytosine in the corresponding crystal structures (Fig 3). Such overlap can be expected if the inhibitor can act as a transition state analogue (Fig 3). In the final fourth step, QM/MM studies were used to test if the observed complex can support formation of a covalent adduct between the target base ring and the active site Cys1226 (Fig 6). Formation of the covalent adduct is the key step in process of suicide inhibition that can assure a high potency and a high specificity for DNA methyltransferase (Fig 1).

In the first step, we used rigid-body-docking studies with programs Vina-Chimera as a fast and computationally light approach that can test if the lead compound can dock in its binding site and form supportive interactions . The corresponding search box included the whole catalytic domain, the whole DNA binding domain and about 15 Å of the surrounding space. The protein binding site was treated as a rigid body, and the docking was restricted to a maximum of 9 conformers. The docking was successful if at least one of the nine binding conformers could overlap with AdoHcys and target cytosine in the corresponding crystal structures (Fig 3 and Table 1A, the lead compound entry).

In the second step, we subjected selected complex from the rigid-body-docking studies to MM/MD simulations using program Gromacs . The selected complex was first minimized using molecular mechanics protocols in presence of bulk solvent molecules. The molecular mechanics minimizations in essence represent evaluation of the initial rigid-body-docking complex as a flexible-docking problem. The minimizations resulted in only minor changes in Dnmt1 complex with the lead compound, indicating the initial rigid-body-docking was suitable for our application. Next, molecular dynamics simulations were used to analyze how bulk solvent molecules and molecular flexibility can affect interactions within the complex. First, we find that the adenine and the ribose rings are firmly anchored in their binding site within 20 nsec simulations (Figs 3 and 4). The solvent can break the hydrogen bonds between Dnmt1 and ribose or adenine ring (Fig 3), however such interruptions have very little effect on binding since the two rings are held in place by multiple interactions (Table 1A, RMSD the lead compound entry). Second, we find that unlike adenosine and ribose rings, the linker and the target base constantly swivel within confines of the binding site cavity (Fig 4). The swiveling results in two major conformations which show as two distinct steps in the plots of average RMSD values (Table 1A, the lead compound entry). MM/MD simulations plots also show that the distances between active site Cys 1226 and carbon 6 on the target base oscillate between 4 and 6 Å, with an average distance of 4.7 Å. Such oscillations can significantly affect rate of formation of the covalent adduct and the suicide inhibition (Fig 6). MM/MD simulations show that the oscillations are driven by two sets of motions. First, the mobility of the active site Cys 1226 is driven by conformational changes in the active site loop (amino acids 1220 to 1236). The loop forms the edges of the active site cavity (Fig 4), that is continually closing and opening with the oscillation amplitude of about 4 Å at the tip of the loop. The second set of motions within the active site is wobbling of the target base ring around its two hydrogen bonds with Arg 1310 and Arg 1312 (Fig 3). In some frames the ring can rotate by as much as 360 degrees, and such rotations can be

4

detected as the extremes on "Cys1226-carbon-6-distance" plots (Table 1A, the lead compound entry).

In the third step we prepared 58 modifications in the lead compound to analyze how specific changes in functional groups and ligand flexibility can affect the binding interactions (Tables 1-5). This ligand modification approach was a very rigorous test of the reliability of presented insilico studies. The modifications also gave valuable insights about interaction forces and showed a number of possible improvements in the lead compound structure. For example, we found a number of apparently subtle modifications in the lead structure that could prevent the initial rigid body docking (Tables 1-5). Some modifications did not significantly affect the initial rigid-body docking, but the subsequent MM/MD simulations showed differences in the stability of Dnmt1-ligand complex (Tables 1-5). The observed differences in binding interactions are illustrated by changes in the "average-ligand-RMSD" plots and by changes in "Cys1226-carbon-6-distance" plots relative to the lead compound (Tables 1-5). Ligand RMSD plots were used to track MM/MD frames that show conformations with significant changes in Dnmt1-ligand interaction. Distinct steps in RMSD plots show different conformations of the ligand within the complex, while uniform RMSD plots with low RMSD values indicates a stabile complex. Similarly "Cys1226-carbon-6-distance" plots were used to select MM/MD frames for QM/MM studies of covalent bond formation and mechanism-based suicide inhibition step (Fig 6). Covalent bond formation and the suicide inhibition step can be expected if the distance between the active site Cys 1226 and carbon 6 on the target base is within 3-4 Å, and if the two groups can meet within reactive angle (Fig 6). It is also necessary to add that many of the presented modifications gave higher more favorable LogD values for the ligand relative to the lead compound (Fig 2, and Table 1-4).

2.2.1. Evaluations of binding interactions at the adenine part of the lead compound (Table 1)

All MM/MD simulation frames showed that the adenine ring sits tightly in its hydrophobic pocket behind Met 1169 and fully overlaps with adenine ring of AdoHcys in the crystal structure (Fig 3 and). In addition to the tight binding pocket, the binding is stabilized by three highly dynamic hydrogen bonds between the ring nitrogen atoms and the protein backbone (Fig 3). MM/MD simulations showed that each of the hydrogen bonds is frequently broken or interrupted by bulk solvent molecules, but that does not affect position of the ring in its binding cavity (Fig. 4-5). We have analyzed two different modification sets. First modifications at the carbon 6 amino group and second modifications within the ring structure (Table 1 A-B). Different modifications in the adenine ring that can affect binding to Dnmt1 have been experimentally described .

The amino group on carbon 6 is positioned on a hydrophobic ridge out of the adenosine binding pocket where it makes maximum of three dynamic hydrogen bonds with protein backbone and bulk solvent molecules (Figs. 3-5). We have replaced amino group with hydrogen atom, nonpolar methyl group, or carbonyl group (Table 1A). The binding was not observed only when a hydrogen atom was placed on carbon 6, what is suggesting that a group of certain size is needed to guide the adenosine ring in its binding pocket (Fig 4-5). Both methyl group and carbonyl groups support binding, what indicates that the hydrogen bonds with the amino group on carbon 6 are not crucial for the Dnmt1-ligand interaction. Precisely, MM/MD simulations showed that methyl and carbonyl group can reposition the modified ring in the ligand binding cavity so that optimal interactions with other amino acids can be achieved. The repositioning can

also increase the ligand RMSD values, yet still "Cys1226-carbon-6" distance plots show that both modifications can support the suicide inhibition step (Table 1A).

We next introduced modifications within the ring structure (Table 1B). We made the ring more hydrophobic by replacing the polar nitrogen atoms with carbon atoms or by replacing the aromatic ring with saturated ring (Table 1B). The increase in hydrophobic surface can bury the modified ring in its binding pocket and decrease the ligand mobility in its binding site groove (Figs 4-5, and Table 1B). This results in decrease in RMSD values, and in a very favorable decrease in "Cys1226-carbon-6" distance plots (Table 1B). The only exception is fully saturated ring, which is too tick to slip in the binding site pocket that is suited for a flat aromatic ring. The ligand with the saturated ring can dock in the binding site groove, but the corresponding RMSD and "Cys1226-carbon-6" distance plots show that the ligand with saturated ring is only loosely bound in the binding site groove and cannot support the suicide inhibition step (Fig. 6).

In conclusion, we found several modifications in the adenine part of the ligand that can support binding of the ligand to Dnmt1 (Table 1 A-B). These modifications can be used to decrease the large negative contribution of the adenine ring to the LogD value (Fig 2B), and to possibly eliminate some of the nonspecific interaction with numerous other proteins that can bind adenine ring in cells . Preparation of different modifications in the adenine ring have been described in the literature , and many of the modified structures are commercially available.

2.2.2. Evaluations of binding interactions at the ribose part of the lead compound (Table 2)

In all Dnmt1 crystal structures the two OH groups on the ribose ring of AdoHcys and AdoMet make hydrogen bonds with the negatively charged Glu1168 . The same hydrogen bonds can be observed between Glu1168 and our lead compound (Fig 3). The Gromacs MM/MD simulations showed that ribose part of the inhibitor overlaps with the ribose part of crystalline AdoHcys in all simulation frames (Fig 3). The crystal structures and MM/MD simulations showed that both -OH groups and Glu1168 can make dynamic hydrogen bonds with the bulk water molecules . Despite those fluctuations the ribose is firmly anchored to its place. We have modified all six atoms in the ribose ring to test how each atom can contribute to binding of the lead compound (Table 2). We designed modifications in the ribose ring with two aims. The first aim is to increase specificity of the lead compound, since numerous molecules in cells can bind ribose. The second aim is to increase LogD value, and thus enhance ADME properties of the inhibitor. The two -OH group give the highest contribution to the LogD value of the presented lead compound (-0.52 per one OH group, Fig 2).

We find that our lead compound cannot bind to Dnmt1 if both OH groups are replaced with H atoms (Table 2). The ligand can bind if the OH group on carbon 3 is kept and OH group on carbon 2 is replaced with H atoms. However, there is no binding if the OH group on carbon 2 is kept and OH group on carbon 3 is replaced with H atoms. Gromacs MM/MD simulations show that the OH group on the position 3 has more influence on binding that on position 2, since it has more favorable interaction geometry with Glu 1168 and tighter interactions with the walls of the binding site grove. The binding will be also abolished when a double bond is introduced between carbon 2 and 3 (Table 2). In that case the two OH groups are lifted in the plane of the ribose ring, and thus cannot satisfy the optimal geometry for the hydrogen bond formation. We also prepared several positive controls to prove that interaction with Glu 1168 is crucial for binding (Table 2). In all cases the ligand can bind to its site when positive group is in the position

of the OH groups, while the binding is not observed when the OH groups are replaced with negative groups.

Apart from the interactions with Glu 1168, the binding of the lead-compound depends on a tight fitting of its ribose ring in its binding cavity. Precisely, all modifications at the acetal oxygen on position 6 can abolish binding by displacing the ring from it position. The acetal oxygen is in direct Van der Waals contact with Phe1145. Thus any group at the position 6 that is bigger than oxygen atom can push the ribose ring out of its binding grove. The only modification that can be tolerated at that position is replacement of oxygen with secondary nitrogen that has one hydrogen atom directly below the plane of the ring. We also found several surprises. First, the ligand binding is observed when both OH groups are removed and the carbon 3 is replaced with an oxygen atom (Table 2). MM/MD simulation showed that such protein-ligand complex is stabilized in its position by a water molecule that can wedge between the Glu 1168 and oxygen atom at the position 3. The second surprise was observed when the two OH groups were replaced with chlorine atoms. The resulting compound has optimal size for wedging into the binding site cleft in protein in rigid-body docking (Table 2). However such ligand lacks the binding interactions so it can quickly escape from the bind site cleft in MM/MD simulations.

In conclusion just like the adenosine ring, the ribose ring in the lead-compound can tightly overlap with the corresponding parts in crystal structures of AdoHcys (Fig 3). In difference to the adenosine ring with ribose ring we could not find modifications that can give desired stability to the Dnmt1-ligand complex.

2.2.3. Evaluation of binding contribution from the linker part of the lead compound (Table 3).

Both docking and MM/MD studies showed that a tertiary amine-linker can mimic the transition state intermediate of DNA methyltransferases (Fig 3). First, the ethyl branch on the linker is positioned in the same place as the methyl group on the target cytosine in crystal structure of Dnmt1 (Fig 3 and). Second, the tertiary nitrogen is positioned in the same place as sulfur in AdoHcys, while its methyl-tail is positioned in the same place as the Hcys tail of AdoHcys (Fig 3). MM/MD simulations showed that the linker can swivel out the binding site cavity together with the target base while the adenosyl part of the ligand is anchored in its place. Such swiveling can be explained with the binding interactions. The linker is positioned between three binding pockets on Dnmt1 surface where it can make only lose hydrogen bonds or hydrophobic interactions with Dnmt1. The linker is wedged in its place by firmly anchored adenosily part of the ligand and by the target base ring that makes two dynamic hydrogen bonds with Arg 1310 and 1312 in the active site cavity (Fig 3 and Fig 4). Thus we prepared modifications in the linker with three specific aims. The first aim was to test how changes in the linker flexibility can affect the binding of the ligand (Table 3B). Second aim was to replace the positively charged nitrogen atom with some group that has more desirable ADME properties (Table 3A). The third aim was to wedge-in the linker in the binding site grove by adding strategically placed methyl groups that can decrease flexibility and mobility of the linker in its binding site (Table 3C).

First we show that the lead-structure can bind to Dnmt1 with both protonated and unprotonated nitrogen atom on the linker (Table 3A). This is not surprising, since we have found in earlier experiments that a charge at that potion does not significantly affect binding of any ligand to Dnmt1. . Second, we tested different modifications in the tail section of the linker (figure 2, potion 4). The tail end of the linker is expected to fit in the same site as AdoHcys tail

(Fig 3), where it can anchor the linker it its position by acting as a tree-legged stool. Indeed a linker without its methyl-tail cannot support binding of the target base ring to the active site cavity (Fig 2 and Table 3A). Next we replaced the methyl-tail with an ethyl-tail, to see if a longer tail can be a better anchor. Surprisingly, docking studies showed that the longer tail supports ligand binding much less than the methyl-tail, and only when the ligand nitrogen is in its unprotonated position. The corresponding MM/MD studies showed that the long tail has more collisions with the protein surface than the short methyl-tail. Those collisions tend to push the linker out of its binding site and thus decrease the chances for the tail to fit in the Hcys binding cavity. In the last attempt to find the best anchor in the tail section, we analyzed binding of the lead compound with hydroxy-methyl and hydroxy-ethyl tail. The hydroxyl-methyl tail is attractive since it can lower the pKa of the tertiary amine so that the ligand will be mostly neutral at the physiological pH (Table 3A). The docking studies show that hydroxy-methyl will support binding while hydroxy-ethyl tail will not. The corresponding MM/MD studies show that the hydroxy-methyl tail does not make significant hydrogen bonds with the protein surface, rather it supports the binding by acting as a better steric anchor than hydroxy-ethyl tail.

Apart from its function as a tetrahedral steric anchor, the tertiary amine linker was selected as the most versatile structure for organic synthesis. Four atoms could be considered for the branch point of the linker, nitrogen, carbon, oxygen and sulfur. Linkers with a bivalent oxygen or sulfur atom in position 1 cannot have a tail that might be needed to anchor the linker in the required position (Fig 2). A linker with methyl-sulfonium at the position 1 does not have required chemical stability. Docking analysis showed that linkers with oxygen or sulfur at position 1 can support docking of the target base in the binding site cleft, but the linker with -CH$_2$- group cannot. MM/MD analysis showed that this surprising interaction can be attributed to a very tight Van der Waals interaction between oxygen and Pro 1224, which can accommodate oxygen atom but not slightly bigger -CH$_2$- group. The bulky sulfur can wedge-in the binding site cleft and force the target base towards the active site cavity. Finally, a linker with a tertiary carbon at the potion 1 can anchor the lead compound to its binding site just as tertiary nitrogen linker. The carbon linker has also more favorable LogD value (Fig 2), however the compounds with tertiary carbon at the branch point are more difficult to synthesize.

We have also attempted several strategies to stiffen the long arm of the linker and force the target base ring towards Cys 1226 in the active site. We found that introduction of E or Z double bond to any part of the linker cannot force the target base ring towards the active site Cys 1226. Similarly the target base cannot be forced towards Cys 1226 by replacing a flexible linker with some of the ring structures, since different rings structures were too big to fit in the required position. At the end, we found that the target base ring can be forced towards active site Cys 1226 when an *entgegen* double bond is introduced between the carbons 2 and 3 and with another *entgegen* double bond or an oxygen atom placed at the position 1 on the linker (Fig 2 and Table 3B). The binding of the ligand can be also improved by adding methyl groups to the flexible linker (Table 3C). Two methyl groups on the linker can decrease the ligand mobility in the active site grove by wedging the flexible linker against the sides of the binding site cleft (Table 3C).

In conclusion, we found several modifications that can significantly improve our lead compound. Modifications that can simplify preparation of the linker, increase the LogD value of the ligand, and increase stability of the Dnmt1-ligand complex. Some modifications can decrease the linker flexibility with specifically placed double bonds or methyl-groups that can wedge the ligand in its binding site. Some modifications can increase the linker flexibility and still give desired stability to the Dnmt1-ligand complex. The tail section of the ligand can stabilize the

Dnmt1-ligand complex to some degree, however these effects are very limited since the tail can also destabilize the complex by collisions with binding site surface. The presented modifications in the linker can be prepared together with the target base using Biginelli reaction with different commercially available aldehydes, and subsequent modification on the nascent chain using pyridinium chlorochromate oxidations .

2.2.4. QM/MM analysis of the target base ring in proximity of the active site Cys 1226 (Table 4)

We used QM/MM studies to show that the active site Cys 1226 can make a covalent bond with the carbon 6 on the target base ring when the lead compound is bound to Dnmt1 (Fig 6). Formation of the covalent bond is the key step in mechanism-based suicide-inhibition that can make our inhibitor highly potent and highly specific for DNA methyltransferase in cells (Figs 1 and 6). MM/MD plots of distances between cysteine 1226 and carbon 6 on the target base showed that those two atoms are separated by 3 to 4 Å when the lead compound or some of its derivatives are bond to Dnmt1 (Table 1-4). Earlier QM/MM studies of bacterial DNA methyltransferases showed that active site cysteine can form a covalent bond with the target base ring when the two groups are about 3-4 Å apart .

The two highest occupied orbitals on nucleophilic cysteine are perpendicular to each other and have almost equal energy (Fig. 6). The lowest unoccupied orbitals on the target base are positioned on carbon 6 and perpendicular to the plane of the ring. Our QM/MM simulations showed how the covalent bond can form when the active site Cys 1226 moves in the plane of the target base ring (Fig. 6). Cys 1226 can be driven in the reactive position by closing of the active site loop which makes the sides of bind site cleft (Fig 4 vs. Fig 7). The target base ring can turn towards the active site cysteine by wobbling around its two hydrogen bonds with Arg1310 and Arg 1312 (Fig. 3 and Fig. 6). The formation of the covalent bond also requires two water molecules within active site cavity (Fig 6). Interestingly, MM/MD studies showed that the methyl group on the position 1 or 4 of the target base ring can interfere with productive collisions between Cys 1226 and the carbon 6 on the target base ring. Thus, we could regulate the rate of the suicide inhibition step by placing the methyl group on the nitrogen 1 or carbon 4.

We analyzed lead compounds with both 1-methyl-pirimidin-2-one and cytosine as the target base ring. 1-Methyl-pirimidin-2-one has several advantages. First, 1-methyl-pirimidin-2-one has smaller polar surface and larger LogD value than cytosine. Second, 1-methyl-pirimidin-2-one lacks the electron rich amino group, so it is a better electrophile than cytosine (its LUMO is -1.4 eV, vs. LUMO for cytosine is -0.9 eV). Third, derivatives of 1-methyl-pirimidin-2-one are easier to prepare than cytosine derivatives using Biginelli reaction. Fourth, 1-methyl-pirimidin-2-one makes stabile covalent adduct that can be observed in crystal structures, while cytosine makes a reversible covalent adduct . Cytosine as a target base can also have some specific advantages. The lead compound with cytosine as the target base can bind tighter to Dnmt1 since the amino group on the cytosine carbon 4 can form hydrogen bonds with Glu1266 in the active site (Table 4). However, the amino group on carbon 4 is known to hydrolyze and turn into carbonyl group that has repulsive interaction with Glu1266 (Table 4). Taken together these results indicate that 1-methyl-pirimidin-2-one is a better choice when irreversible suicide inhibition is desired, while cytosine can be a better choice when short lasting reversible inhibition is desired. The reactivity of the target base can be also modulated by the methyl groups on carbons 1 and 4 (Table 4). The "Cys1226-to-carbon-6" distance plots show that the methyl-

groups can affect mobility of the target base ring and thus rate of formation of the covalent adduct (Fig 6).

Different modifications on the target base with different linkers at carbon 5 can be prepared in two or three synthetic steps using Biginelli reaction with different commercially available aldehydes and methyl-urea .

2.3 Optimization of binding interactions at multiple sites on the lead compound (Table 5)

The insights from different modifications at single site can be combined to create improved compounds with modifications at multiple sites. MM/MD and docking studies showed that the binding interactions at the ribose ring and the linker region are the two most challenging sites for modification (Tables 1-4). MM/MD studies showed that the loss of the bonding interaction between –OH group and Glu1168 will result in wobbling of the ribose ring in the binding site cleft and facilitate dissociation of the ligand. We added methyl-groups at the C2 and C3 opposite to the two -OH groups to wedge the ligand in its binding site and make the Dnmt1-ligand complex less dependent on interaction with Glu1168 (Table 5). Docking and MM/MD studies showed that such methyl-groups can stabilize the ribose ring and Dnmt1-ligand complex even with some modifications that otherwise do not support formation of the complex (Table 2). Ribose molecules with methyl groups on carbon 2 and carbon 3 are commercially available from several suppliers.

3.1 Binding of the mechanism based inhibitor to different conformations of Dnmt1 (Figure 7)

We next show that the presented lead compound can bind to different conformations of Dnmt1 (Fig 4 vs. Fig 7) . Different crystal structures show different conformations of Dnmt1 that represent different stages in process of DNA methylation in cells. We are specifically interested in function of the active site loop (amino acids 1220 to 1236). The active site loop is in its closed position when full-methylated DNA is bond to Dnmt1 (PDBcode: 4DA4). That structure represents the last step in the process of maintenance methylation when Dnmt1 is in complex with fully-methylated DNA substrate. The active site loop is in open position when mouse Dnmt1 is in complex with unmethylated DNA (PDB code: 3PT6), or when DNA is not bond to Dnmt1 (PDB code: 3AV6). We would like to exploit the function of the active site loop and its influence on the suicide inhibition to design inhibitors that could preferentially target different metabolic functions of Dnmt1 and thus provide specific pharmacological properties .

The simulations show that RMSD values for the lead compound bound to Dnmt1 are almost identical when the active site loop is open and closed (Table 6, average RMSD value 11.4 vs. 12.4 Å). This is not surprising since the active site loop has little effect on AdoMet binding sites which provide the majority of binding interactions for the lead compound (and Fig 3). However there are significant differences on plots of Cys 1226 and carbon 6 distances. The average values for 4DA4 structure is 4.7 Å, while the average for 3AV6 it is 9.1 Å (Table 6). Those differences can be attributed to the different positions in the active site loop (Fig 4 vs. Fig 7). The distance between the Cys 1226 and the carbon 6 on the target base can affect kinetics of the suicide inhibition step (Fig 6). Taken together these results indicate that the suicide inhibition can happen fast when Dnmt1 is bond to hemi-methylated substrate in the process of maintenance methylation. Much slower suicide inhibition can be expected when Dnmt1 is bond to

unmethylated DNA or when Dnmt1 is in its DNA free form. In those cases the slow suicide inhibition step will depend on the slow thermal swiveling of the active site loop between the open and closed position. The differences between fast and slow target base attack and the function of the active site loop have been described experimentally for Dnmt1 and HhaI DNA methyltransferases .

4.1. Conclusions and future directions

We have presented a combination of four different insilico approaches as an insightful strategy for design of different mechanism-based inhibitors of Dnmt1. The presented results show that it is possible to prepare a family of mechanism-based suicide inhibitors of Dnmt1 that can bind to the enzyme like transition state analogues. Different modifications in the original lead compound showed how to increase the stability of the complex and how to affect the rate of the suicide inhibition. Presented insights can guide different synthetic strategies to prepare a whole family of mechanism-based inhibitors of Dnmt1 with optimal ADME properties for desired treatments. Many of the presented compounds can be prepared for less than 2000 USD in material costs by combining different products of Biginelli reaction with different adenosine derivatives in total of 3 to 5 synthetic steps .

The presented insilico approaches can be also extended in the future to prepare compounds specific for Dnmt1, Dnmt2, Dnmt3, or some bacterial DNA methyltransferases . Selective compounds can be prepared by exploiting subtle differences between different DNA methyltransferase in their active site and interaction with the cofactor . The hope is to design inhibitors that will have preference for specific DNA methylation processes in cells

5.1 Figures

Step 1: activation of the target base by formation of the unstable covalent adduct intermediate

(1.1)   (1.2)
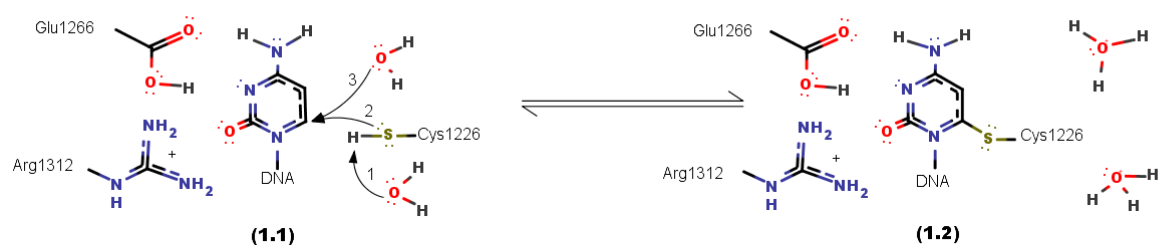
Step 2: the methyltransfer step by electrophilic addition

(2.1)   (2.2)
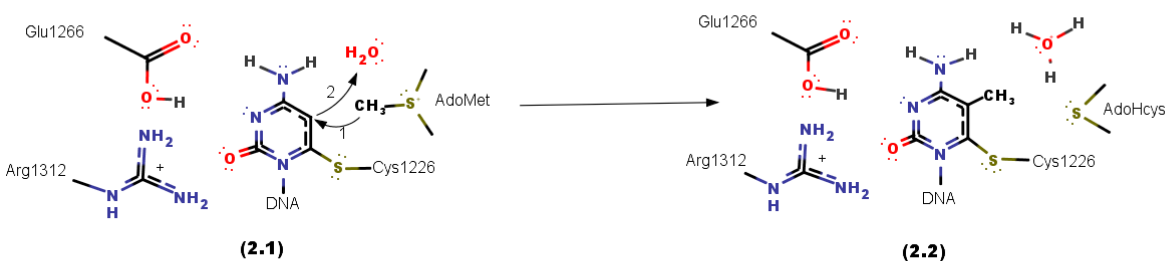
Step 3: elimination

(3.1)   (3.2)
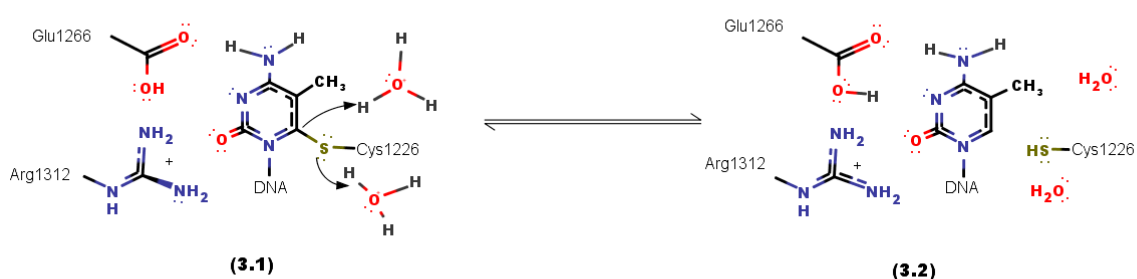
Conceptual mechanism-based transtion state analogue

(4.1)   (4.2)
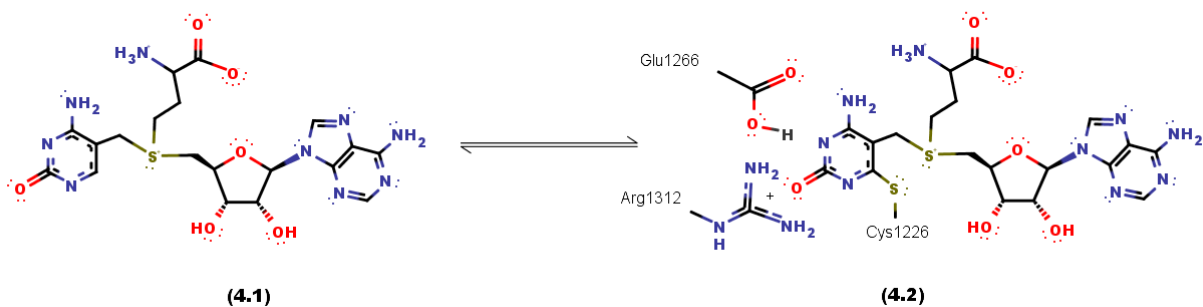
Figure 1. **Catalytic mechanism of DNA methyltransferase Dnmt1**. The catalytic mechanism of bacterial and mammalian cytosine-C5 DNA methyltransferase has been described in details in the last 40 years in different crystal structures, enzyme kinetics studies, enzyme substrate binding studies, and QM/MM studies . The catalysis consists of three main steps. The catalysis is initiated when the aromatic target base ring is posited in the active site cavity (step 1.1). In case of human and mouse enzyme the active site is composed of conserved Arg 1310, Arg 1312, Glu1266, and Cys 1126 residues (step 1.1).

12

The enzyme can form unstable covalent adduct intermediate when conjugated aromatic bonds in the target base ring are positioned in asymmetric active site between polar amino acids (steps 1.1 to 1.2). Kinetic and QM/MM studies showed that the formation of the covalent adduct intermediate is a reversible step unless the reaction is driven forward by the rate-limiting irreversible methyltransfer step . Following the methyltransfer step (steps 2.1 to 2.2), there is a reversible target base eliminations step (steps 3.1 to 3.2). We have decided to target this reversible step with a mechanism-based suicide inhibitor. The structure 4.1 is supposed to mimic the transition from structures 2.1 to 2.2. Thus, when structure 4.1 is positioned in the active site it should trigger formation of covalent adduct shown as the structure 4.2. This is in essence the same process as transition from structure 3.2 to structure 3.1. The figures were drawn in ChemAxon Marvin 15.0.1

A

B
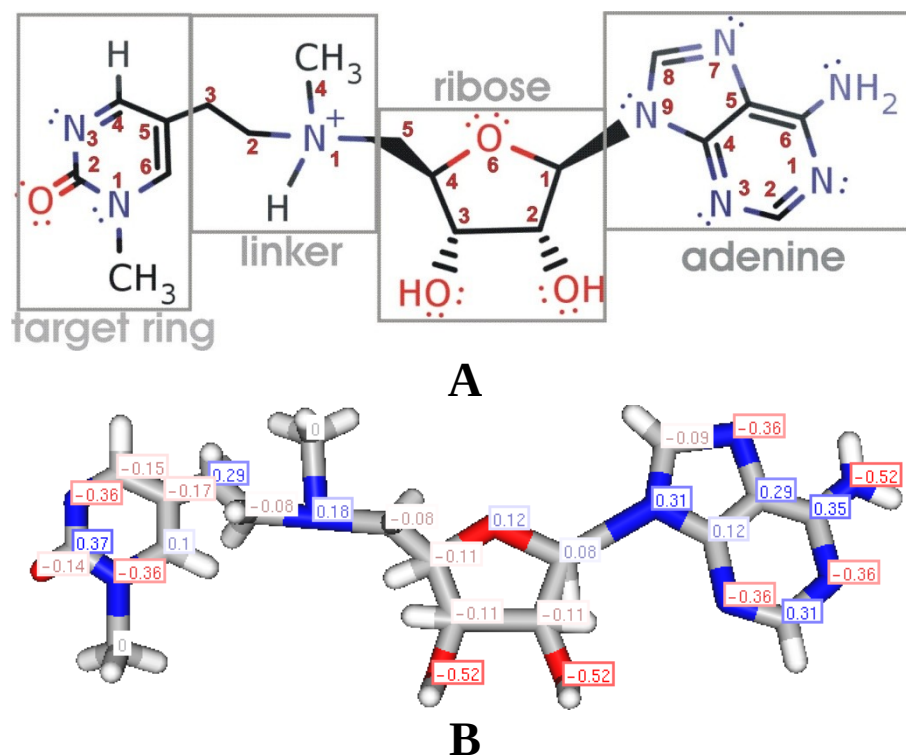
Figure 2 (A-B). **Lead structure for the mechanism-based suicide-inhibitor of Dnmt1**.
(**A**) The figure shows our first prototype for a mechanism-based inhibitor of Dnmt1. The structure is derived from the structure 4.1 in figure 1 as a reasonable challenge for organic synthesis with reasonable agreement with the Lipinski's rules . The structure is designed to act as a transition state analogue and consists of four functional parts: adenine, ribose, tertiary amine linker and 1-methyl-pyrimidin-2-one as the target base ring. The adenine and ribose parts are selected to assure that the lead compound can bind to AdoHcys/AdoMet sites. Tetrahedral tertiary amine linker is chosen as a flexible structure that can mimic the transition state structure (Fig 1). The 2-pyrmidine(2-one) was chosen as the target base ring instead of cytosine, to avoid problems with deamination at carbon 4 and to decrease polar surface on the target base ring. Different parts on the structure are labeled with numbers to help the reader in following different modifications throughout the text.

(**B**) The present structure is in agreement with the Lipinski's rules except for its low LogD value. The relative molecular mass is 417.45 Da. There are 6 rotatable bonds, and total of 58 bonds between 55 atoms. At pH 7.40 the polar surface area is 156.42 Å^2 while Van der Walls surface area is 562.17. There are five H bond donor sites and 13 acceptor sites. At pH 7.40, 94% of the ligand is protonated at the linker nitrogen and the compound is positively charged. The LogD value in protonated form is -2.63 and -1.88 in unprotonated form. To improve ADME properties the future modifications will be designed to increase the LogD value. This can be achieved by removing the positive charged nitrogen at the linker, and by removing hetero atoms at the ribose and adenine ring that give highest negative contribution to the LogD value. The numbers on the structure show contribution to the total LogD for each atom. All figures and calculations were prepared in ChemAxon Marvin suite version 15.3.16
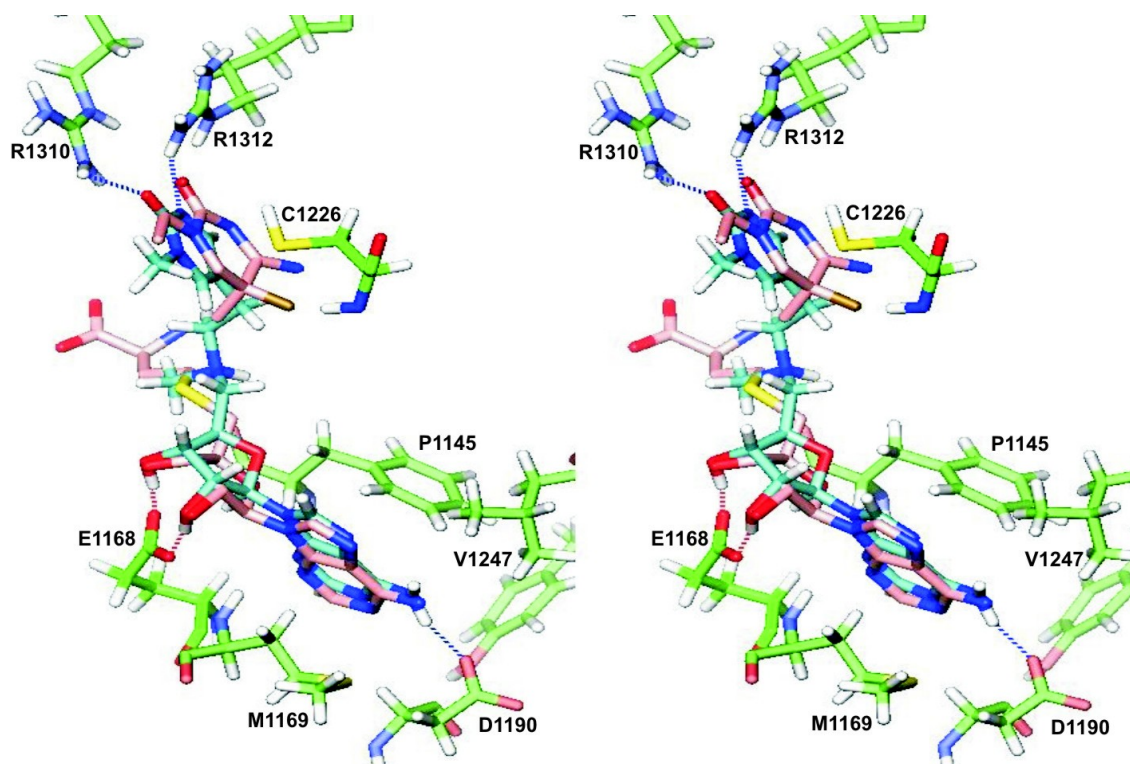
Figure 3. The lead compound can bind to crystal structure of mouse Dnmt1 (PDB: 4DA4). The stereo image was prepared in VMD and shows the lead compound (cyan), and the crystal structure of interacting amino acids (green) AdoHcys (pink) and the target cytosine (pink) (PDB: 4DA4). Hydrogen atoms are shown in white, oxygen in red, nitrogen in blue, and sulfur in yellow. The figure shows that our lead compound can dock to Dnmt1 in the same position as AdoHcy and the target cytosine. The figure was prepared by superimposing one of the MM/MD simulation frames with the lead-compound in complex with Dnmt1 to the crystal structure of Dnmt1 in complex with AdoHcys and the target cytosine (PDB: 4DA4).

First, we find that adenine and ribose part of the lead compound can overlap with AdoHcys in crystal structure in all simulation frames. The ribose rings are always anchored to their binding site by two hydrogen bonds with Glu 1168. The adenine rings are buried in a

hydrophobic pocket and anchored with several hydrogen bonds to Asp1190 and the corresponding section of peptide backbone. MM/MD simulations show that all of the hydrogen bonds are dynamic and sometimes interrupted by the bulk solvent molecules. However, adenine and ribose rings stay in place since the binding is always supported by multiple interactions.

Second, we find that linker on the lead compound can fully mimic the transition state structure. The ethyl part of the linker is parallel to the methyl group on carbon-5 of target cytosine in the place of expected reaction coordinate . The positively charged tertiary amine is in position of the sulfur atom on AdoHcys, while the methyl-tail of the linker overlaps with the homocysteine tail.

Third, target base ring on the lead compound is positioned in same orientation as the target cytosine (pink) so that the two structures can practically overlap. Both rings form dynamic hydrogen bounds between the carbonyl oxygen and Arg 1310 or Arg 1312. The interactions with the two Arg residues are crucial for positioning of carbon 6 atom on the target base ring within 4 Å from the active site Cys1226. This proximity is crucial for formation of covalent adduct between the enzyme and the target base (Fig 1, and).
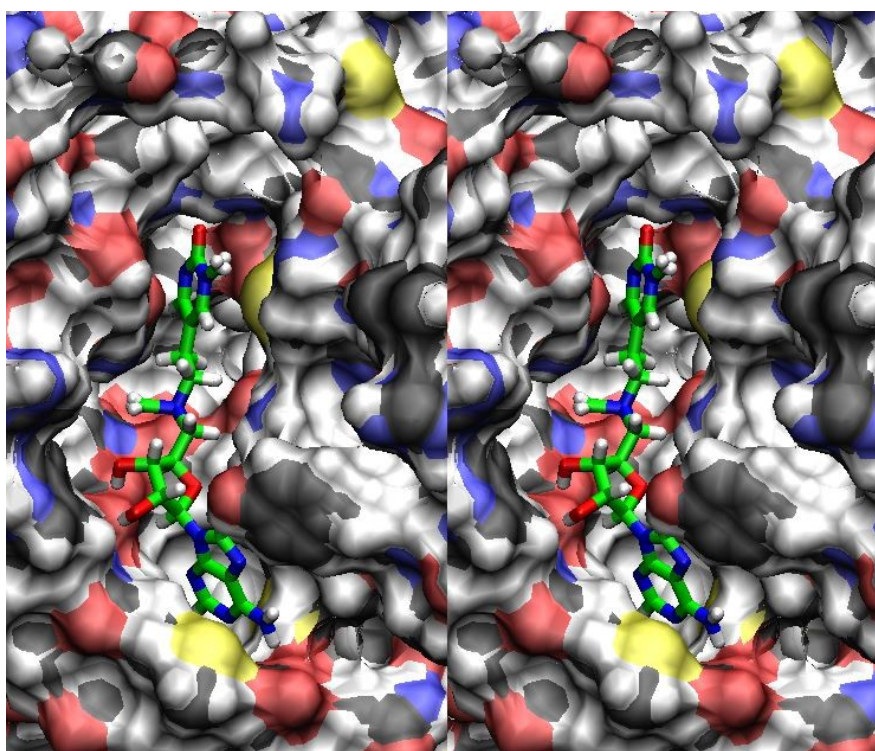
Figure 4 Stereo image of the lead compound in its binding cleft on mouse Dnmt1 (PDB: 4DA4). The image shows the lead compound bound to mouse Dnmt1 just as in figure 3, but this time in surface mode to highlight the depth and the structure of binding site cleft. The image was prepared in VMD , and shows the surface hydrogens in white, carbons in black, nitrogens in blue, oxygens in red, and sulfur in yellow. The lead compound is shown in stick mode and colored green. The image shows one of the most open Gromacs MM/MD simulation frames to show depth and surface of the active site cavity. All of the amino acids shown in figure 3 can be easily recognized. The surrounding solvent molecules were removed for clarity. Reader is looking at the binding site cleft from the position of the DNA substrate, therefore the methyl

group on the nitrogen 1 of the target base ring is positioned almost perpendicular to the plane of paper.

The adenine ring is positioned in a tight surface pocket enclosed by Met1169, while the amino group on carbon 6 is positioned towards Asp 1190. The ribose ring is anchored in a wide ridge with its two OH groups atop of Glu1168. The linker is positioned in the widest part of the binding site cleft just above AdoHcys/AdoMet binding cavity. The target base ring is positioned in the active site cavity where it is stabilized by two dynamic hydrogen bonds between its carbonyl group and Arg1310 and Arg1312. The carbon 6 of the target base ring is within 4 Å from the active Cys1226, which is close to the optimal distance for formation of the covalent adduct .
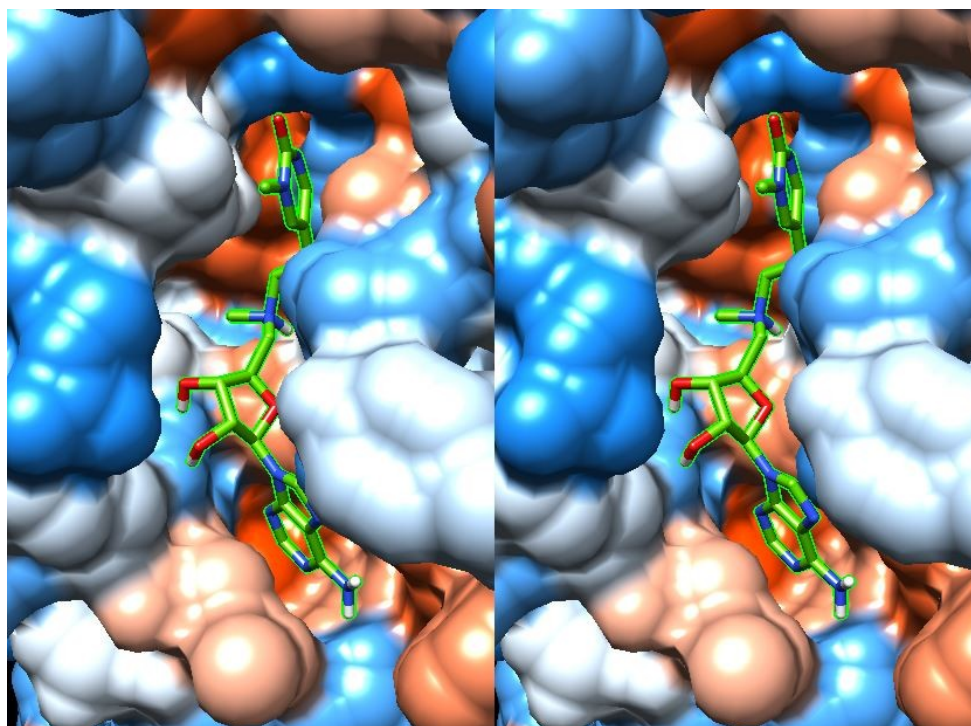
Figure 5. The lead structure binds to both hydrophobic and hydrophilic surface in its binding site cleft on mouse Dnmt1 (PDB: 4DA4). The stereo image shows the lead compound bound to mouse Dnmt1 just as in figures 3 and 4, but this time surface in the binding cavity is colored according to its hydrophobic (brown) and hydrophilic (blue) properties in program Chimera .

The figure shows that the adenine ring is positioned in a deep hydrophobic pocket, except for its amino group on carbon 6 which makes hydrogen bonds with hydrophilic Asp 1190. The ribose ring is positioned on a tight ridge so that its two OH groups sit atop of the hydrophilic surface of Glu 1168 while its acetal oxygen is buried in the slightly hydrophobic cavity. The nitrogen and the methyl tail part of the linker sit in the hydrophilic AdoMet/AdoCys cavity. The ethyl part of the linker that is positioned in the reaction coordinate grove sits on a hydrophobic surface. Finally, the target base 1-methyl-pyrimid-2-one ring is positioned in a hydrophobic active site cavity where it makes dynamic hydrogen bonds through its carbonyl oxygen and the hydrophilic surface of the active site Arg 1310 and 1312.
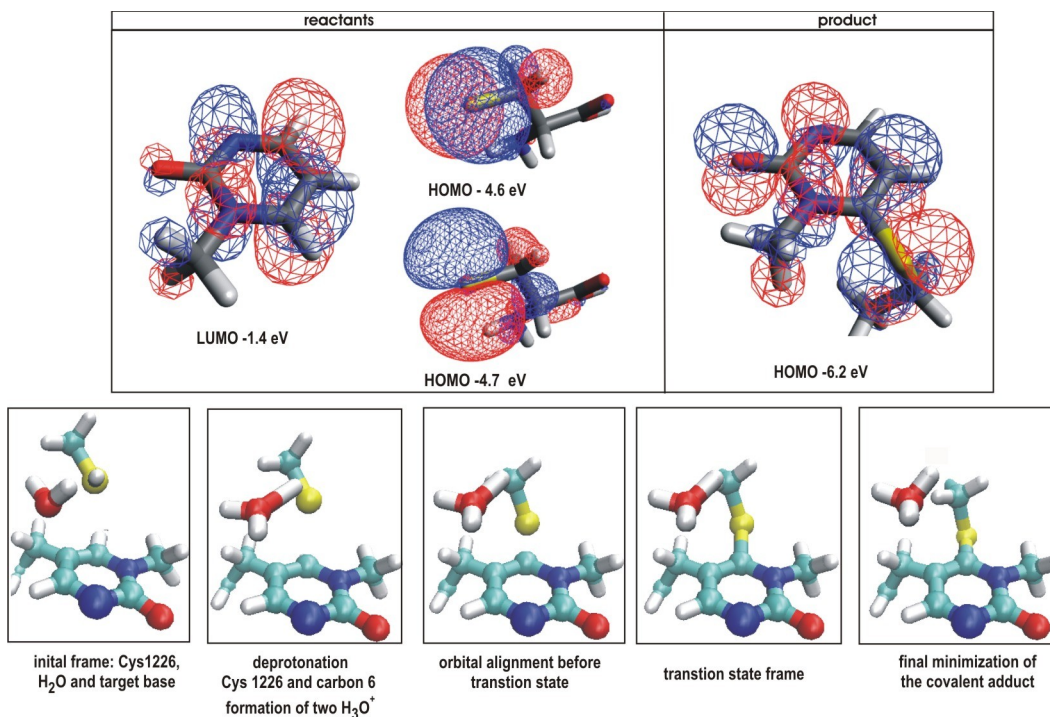
Figure 6. **Formation of a covalent bond between the lead compound and active site cysteine 1226 (all figures were prepared in VMD)**.

The target base ring has LUMO orbitals at carbon 6 that are perpendicular to the plane of the ring. The nucleophilic sulfur anion has two HOMO orbitals with almost identical energy that are perpendicular to each other. The reaction can start when one of the two HOMO orbitals on the sulfur atoms come in the same plane as the target base ring. The orbitals and their energies were calculated in program Gamess .
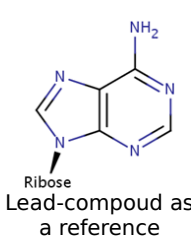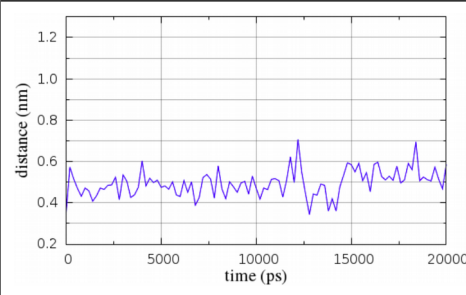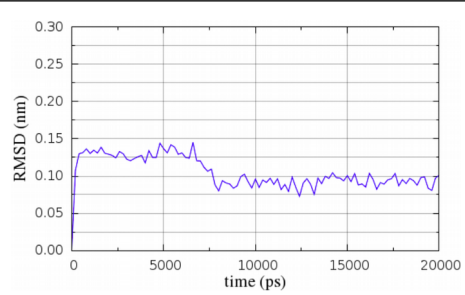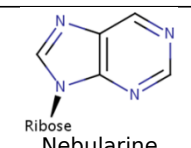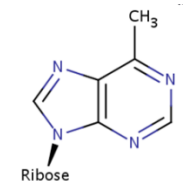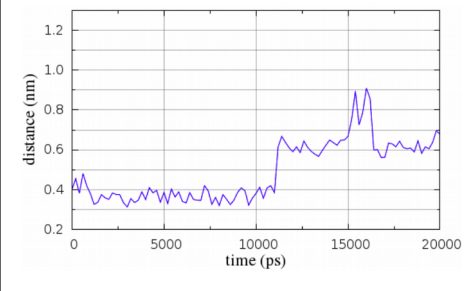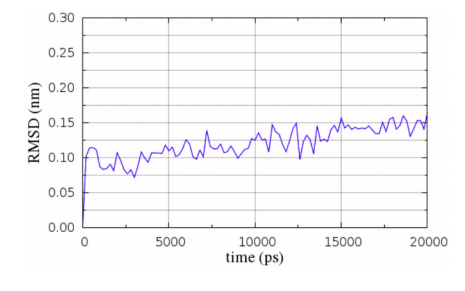
In the bottom panels, selected QM/MM simulation frames show how nucleophilic orbitals on sulfur anion can come in the plane with the electrophilic orbitals on carbon 6 of the target base. Hydrogen atoms are shown in white, carbon in cyan, oxygen in red, nitrogen in blue, and sulfur in yellow. The first panel shows one of the Gromacs MM/MD simulation frames that was used as a starting point for QM/MM simulations with program CP2K . The active site cysteine 1226 is directly above the plane of the target base ring 3.8 Å from the carbon 6. Both Cys 1226 and carbon 6 on the target base are protonated (1st frame). The reaction starts with deprotonation of the active site Cys by a water molecule (2nd frame). When Cys 1226 anion collides with carbon 6 and its hydrogen atom, the carbon 6 can get deprotonated (3rd frame, deprotonation of the carbon 6 takes place below the ring so it is not visible in the presented angle). Deprotonation is completed when the Cys 1226 anion gets in the same plane as the target base ring (3rd frame) so that the transition state complex and the covalent bond can be formed (4th frame). At the end the new bonds are minimized to achieve the optimal bond angles. Both MM/MD and QM/MM simulations show that the bond formation is driven by the wobbling of the target base ring around the hydrogen bond axis between carbonyl oxygen and Arg 1310 and 1312. Up-and-down swiveling motion of Cys 1226 is driven by wobbling in the active site loop . Looking at figure 4, the bond forming motions can be driven by the fluctuations in the width of the binding site cavity, and by the rotation of the target base ring around its hydrogen bonds axis with Arg 1310 and Arg 1312.
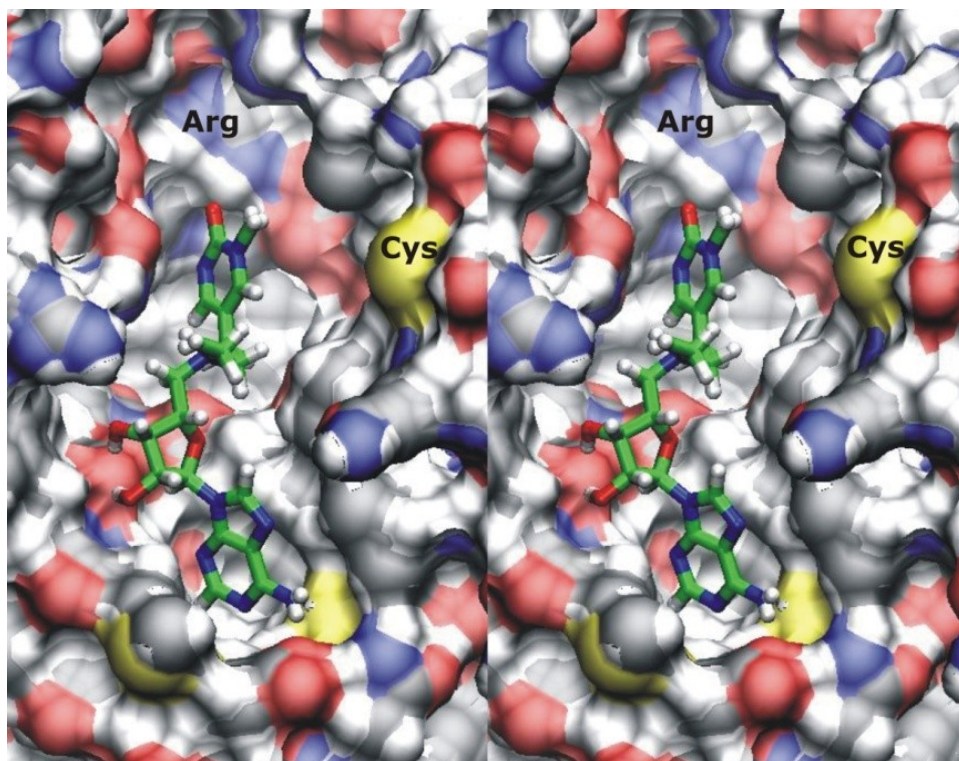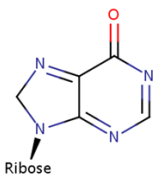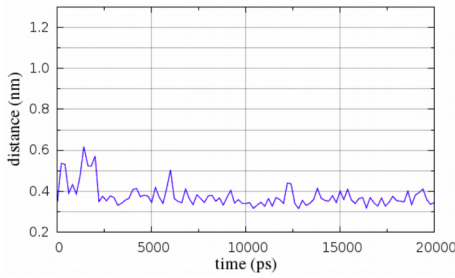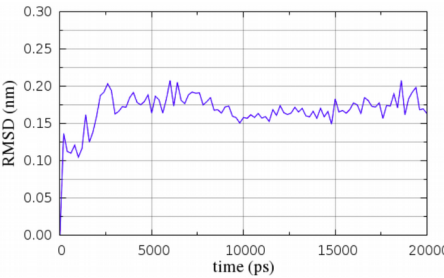
Figure 7. Stereo image of the lead structure bound to mouse Dnmt1 with the active site loop open (PDB code:3AV6). The image shows that the binding site for the lead compound is much wider when the active site loop is open (amino acids 1220 to 1236). Just as in previous figures, the inhibitor is held in its binding cavity by its three main interactions. The two -OH groups on the ribose ring, the adenine ring that is buried in the hydrophobic pocket, and the the hydrogen bonding with Arg 1310 and Arg 1312. Closing of the active site loop drives the active site cytosine Cys1226 to its catalytic position . Taken together these observations indicate that the lead compound can bind to Dnmt1 when its active site loop is in its open and in closed position. However the suicide inhibition is much more likely when the active site loop is closed. The closing of the loop can be driven by DNA binding or by slow thermodynamic swiveling in the loop structure from closed to open conformation. Image was prepared in VMD
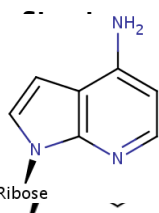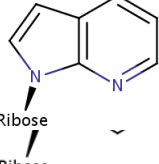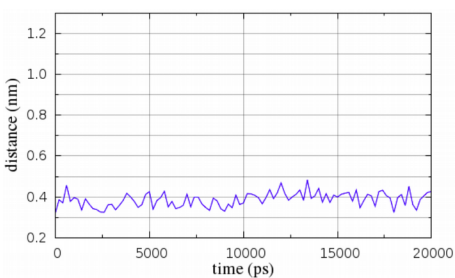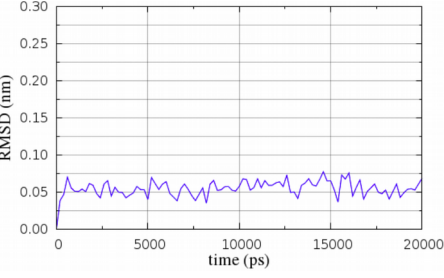
6.1 Tables
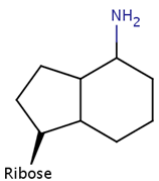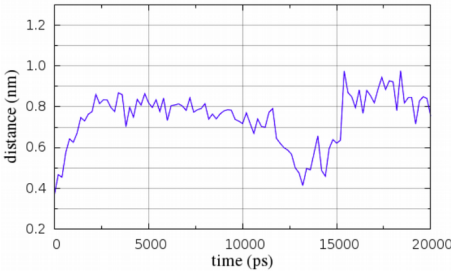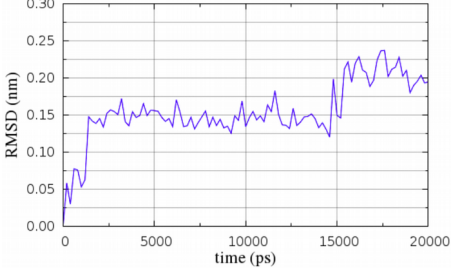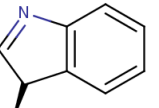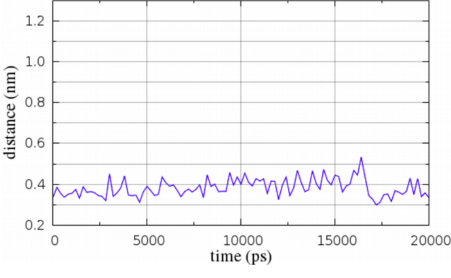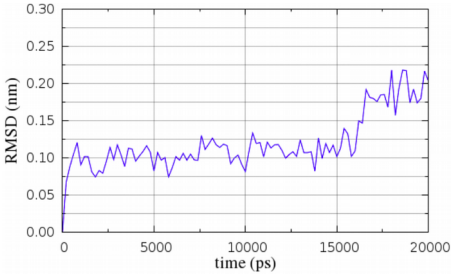
Table 1A. Optimization at the position 6 in adenine ring of the presented lead compound (the LogD values represent calculated values for the lead compound with presented modification using weighted option in ChemAxon Marvin . The LogD value for the lead compound is -3.07, thus any modification that has LogD value higher than -3.07 represents a favorable LogD modification)**:**

Structure modification	LogD @ pH=7.4	Chimera-Vina Docking interaction energy range H bonds with Dnmt1	Gromacs MM/MD frames Cys1226 ringC6 distance relative to the first frame	Gromacs MM/MD frames RMSD Ligand relative to the first frame
Lead-compoud as a reference	-3.07	-9.7 to -7.9 kcal/mol 7 hydrogen bonds: CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) MET 1169 – N3 (adenosine) GLU 1168 – O3 (ribose) GLU 1168 – O2 (ribose) PHE 1145 – N1 (linker) ARG 1312 – O2 (cytosine)		
Nebularine	-2.84	did not dock to the binding site cavity	n.a	n.a
(CH3 modification)	-2.60	-9.2 to -8.0 kcal/mol 3 hydrogen bonds: MET 1169 – N3 (adenosine) CYS 1191 – N1 (adenosine) HOH 1704 – O2 (linker)		
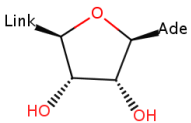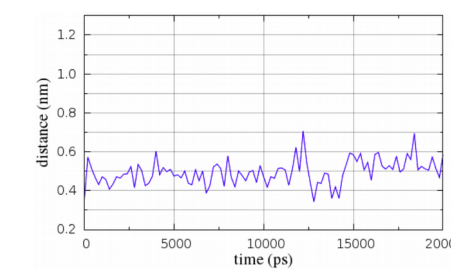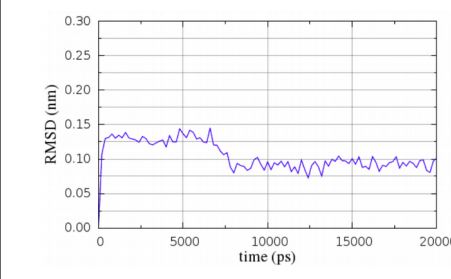
Ribose	-2.44	10.1 to -8.1 kcal/mol 7 hydrogen bonds MET 1169 – N3 (adenosine) CYS 1191 – O6 (adenosine) CYS 1191 – N2 (adenosine) GLU 1168 – O3 (ribose) GLY 1223 – N1 (linker) ARG 1312 – N3 (cytosine) ARG 1312 – O2 (cytosine)		
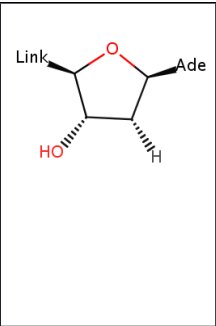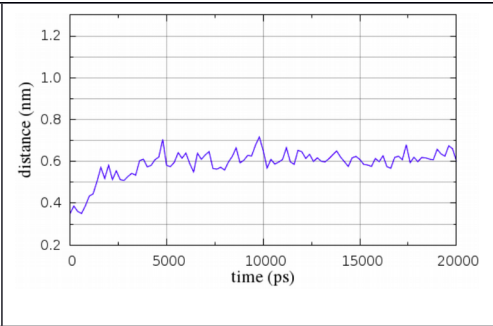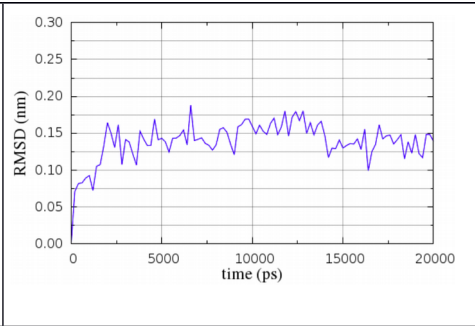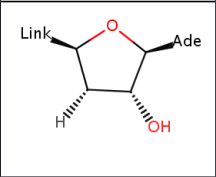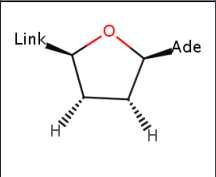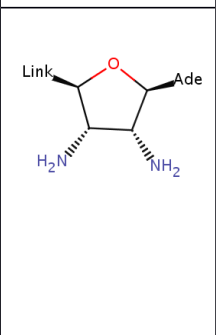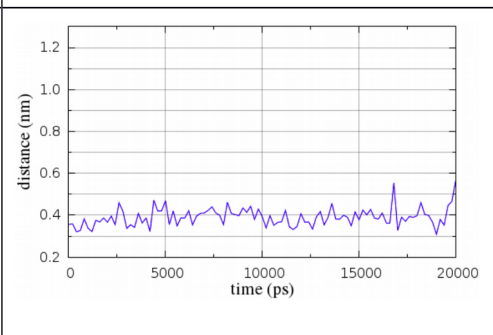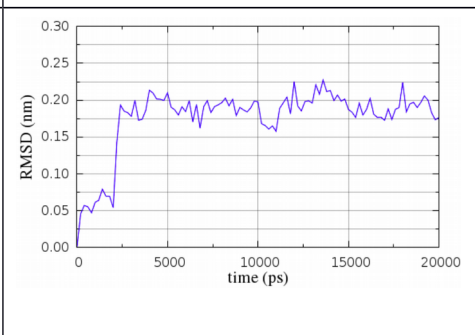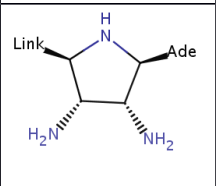
Table 1B. Optimization of heteroatoms in adenosine ring of the presented lead compound:
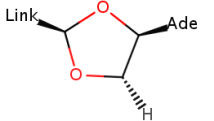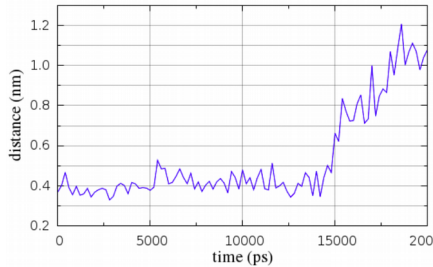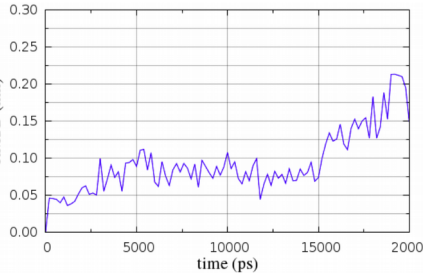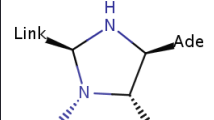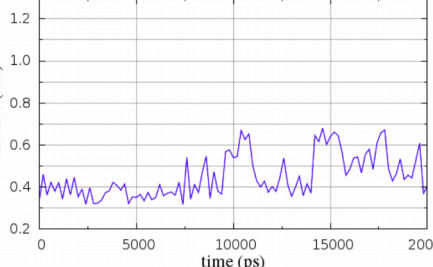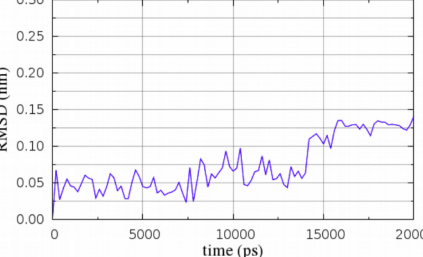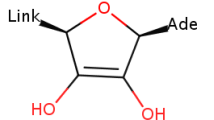
Structure	LogD @ pH=7.4	Chimera-Vina Docking interaction energy range H bonds with Dnmt1	Gromacs MM/MD frames Cys1226 ringC6 distance relative to the first frame	Gromacs MM/MD frames RMSD Ligand relative to the first frame
Ribose Ribose	-2.64	did not dock to the binding site cavity	n.a.	n.a.
	-2.84	-8.8 to -8.1 kcal/mol 4 hydrogen bonds CYS 1191 – N6 (adenosine) GLU 1168 – O2 (ribose) ARG 1310 – O2 (cytosine) ARG 1312 – O2 (cytosine)		

Structure	LogD	Chimera-Vina Docking interaction energy range H bonds with Dnmt1	Gromacs MM/MD frames Cys1226 ringC6 distance relative to the first frame	Gromacs MM/MD frames RMSD Ligand relative to the first frame
(structure with NH₂, Ribose)	-1.20	-8.5 to -7.4 kcal/mol 5 hydrogen bonds HOH 1773 – N6 (adenosine) HOH 1704 – O6 (ribose) GLN 1227 – O2 (ribose) ARG 1310 – O2 (cytosine) GLU 1312 – O2 (cytosine)	*(distance vs time plot, 0–20000 ps)*	*(RMSD vs time plot, 0–20000 ps)*
(indole-type structure with N, Ribose)	-1.31	-9.6 to -8.1 kcal/mol 5 hydrogen bonds HOH 1704 – O2 (ribose) HOH 1704 – O3 (ribose) GLU 1168 – O2 (ribose) ARG 1310 – O2 (cytosine) ARG 1312 – O2 (cytosine)	*(distance vs time plot, 0–20000 ps)*	*(RMSD vs time plot, 0–20000 ps)*

Table 2A. Optimization of the hydroxyl groups on ribose ring of the presented lead compound: (the LogD values represent calculated values for the lead compound with presented modification using weighted option in ChemAxon Marvin . The LogD value for the lead compound is -3.07, thus any modification that has LogD value higher than -3.07 represents a favorable modification)
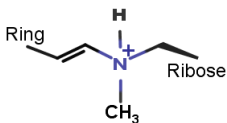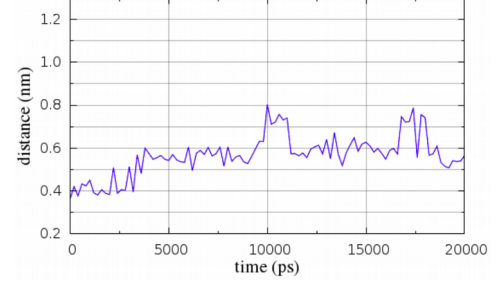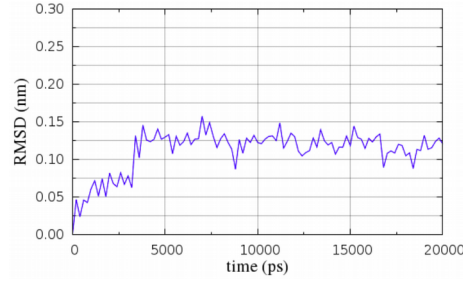
Structure modification	LogD @ pH=7.4	Chimera-Vina Docking interaction energy range H bonds with Dnmt1	Gromacs MM/MD frames Cys1226 ringC6 distance relative to the first frame	Gromacs MM/MD frames RMSD Ligand relative to the first frame
(ribose structure: Link, O, Ade, HO, OH) lead compound as reference	-3.07	-9.7 to -7.9 kcal/mol 7 hydrogen bonds: CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) MET 1169 – N3 (adenosine) GLU 1168 – O2 (ribose) GLU 1168 – O3 (ribose) PHE 1145 – N1 (linker) ARG 1312 – O2 (cytosine)	*(distance vs time plot, 0–20000 ps)*	*(RMSD vs time plot, 0–20000 ps)*

Structure	Score	Binding Info	Distance (nm) vs time (ps)	RMSD (nm) vs time (ps)
Link–O–Ade ring, HO, H	-2.46	-9.7 to -8.2 kcal/mol 5 hydrogen bonds: CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) MET 1169 – N3 (adenosine) ARG 1312 – O2 (cytosine) ARG 1312 – N3 (cytosine)		
Link–O–Ade ring, H, OH	-2.87	did not dock to the binding site cavity	n.a.	n.a.
Link–O–Ade ring, H, H	-2.25	did not dock to the binding site cavity	n.a.	n.a.
Link–O–Ade ring, H_2N, NH_2	-5.03	-9.0 to -7.9 kcal/mol 4 hydrogen bonds CYS 1191 – N6 (adenosine) GLU 1168 – N3 (ribose) ARG 1312 – O2 (cytosine) ARG 1312 – N3 (cytosine)		
Link–N(H)–Ade ring, H_2N, NH_2	-5.07	did not dock to the binding site cavity	n.a.	n.a.

Structure modification	LogD	Chimera-Vina Docking interaction energy range H bonds with Dnmt1	Gromacs MM/MD frames Cys1226 ringC6 distance relative to the first frame	Gromacs MM/MD frames RMSD Ligand relative to the first frame
(Link—O—O—Ade structure, with H)	-1.23	-9.9 to -8.3 kcal/mol 5 hydrogen bonds: CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) MET 1169 – N3 (adenosine) HOH 1704 – O3 (ribose) ARG 1312 – O2 (cytosine)		
(Link—NH—N—Ade structure, with H, H, H)	-2.03	-9.9 to -8.5 kcal/mol 4 hydrogen bonds: CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) MET 1169 – N3 (adenosine) ARG 1312 – O2 (cytosine)		
(Link—O—Ade structure, with HO, OH)	-4.16	did not dock to the binding site cavity	n.a.	n.a.

Table 2B. Optimization of acetal-oxygen at the position 6 in ribose ring of the presented lead compound (the LogD values represent calculated values for the lead compound with presented modification using weighted option in ChemAxon Marvin . The LogD value for the lead compound is -3.07, thus any modification that has LogD value higher than -3.07 represents a favorable modification)**:**

Structure modification	LogD @ pH=7.4	Chimera-Vina Docking interaction energy range H bonds with Dnmt1	Gromacs MM/MD frames Cys1226 ringC6 distance relative to the first frame	Gromacs MM/MD frames RMSD Ligand relative to the first frame

		-10.1 to -7.5 kcal/mol 2 hydrogen bonds: HOH 1773 – N1 (adenosine) ARG 1312 – O2 (cytosine)		
(structure: cyclopentane ring with Link, Ade, two Cl, two H)	-1.53		(distance vs time plot, 0–20000 ps)	(RMSD vs time plot, 0–20000 ps)
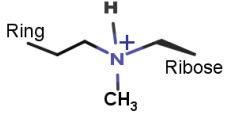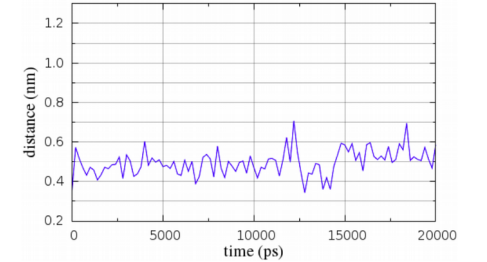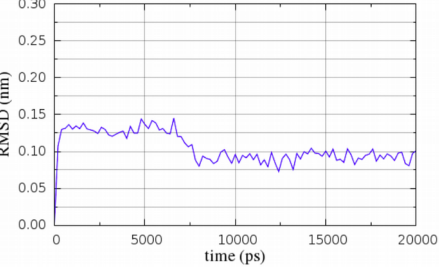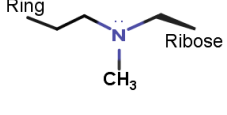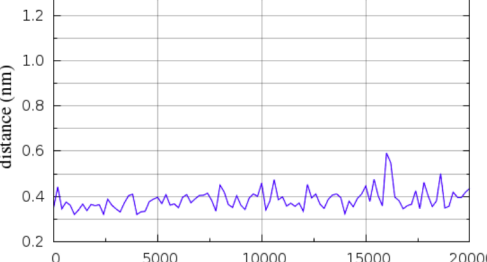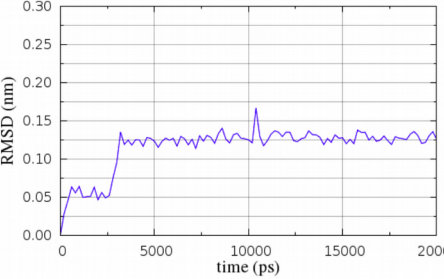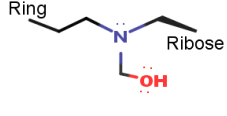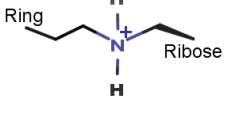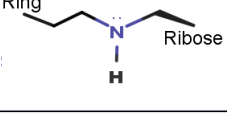
The following modifications at the acetal-oxygen could not dock into the binding cleft cavity

(structure: Link, Ade, HO, OH)	(structure: Link, Ade, H, OH)	(structure: Link, Ade, HO, H)
(structure: Link, Ade, H, H)	(structure: Link, Ade, HS, SH)	(structure: Link, Ade, F, F)
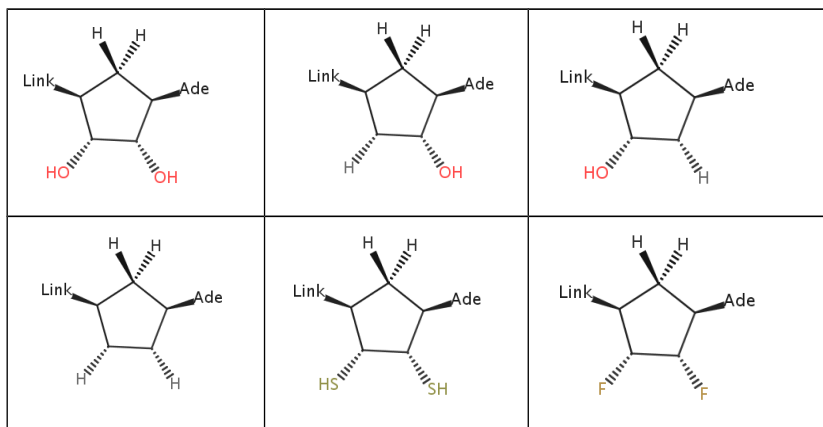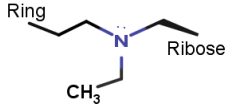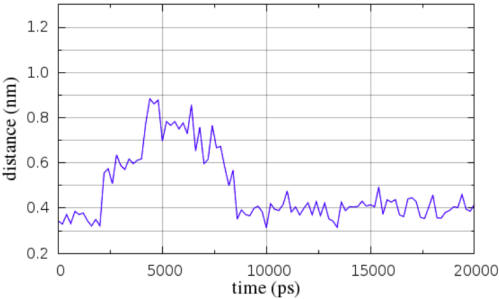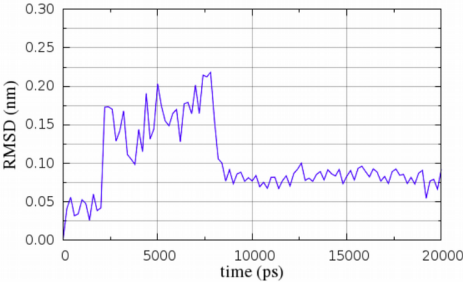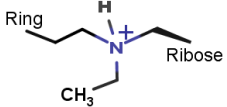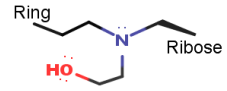
Table 3A. Optimization of the linker part of the presented lead compound:changes in the linker tail. (the LogD values represent calculated values for the lead compound with presented modification using weighted option in ChemAxon Marvin . The LogD value for the lead compound is -3.07, thus any modification that has LogD value higher than -3.07 represents a favorable modification)**:**

Structure modification	LogD @ pH=7.4	Chimera-Vina Docking interaction energy range H bonds with Dnmt1	Gromacs MM/MD frames Cys1226 ringC6 distance relative to the first frame	Gromacs MM/MD frames RMSD Ligand relative to the first frame

24

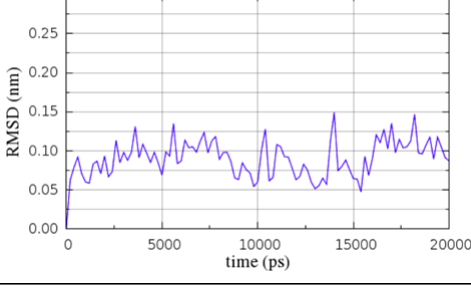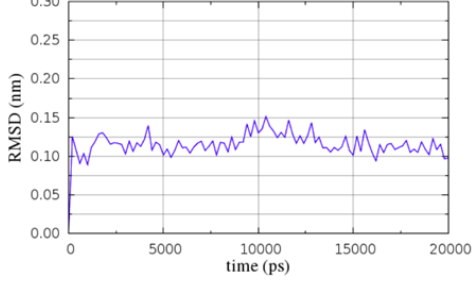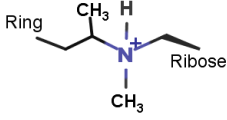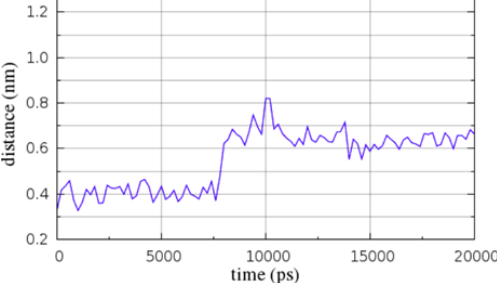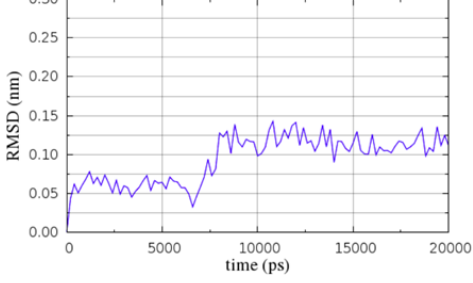| Structure	Value	Binding energy / H-bonds	Distance (nm) vs time	RMSD (nm) vs time
Ring — N⁺(H)(CH₃)—Ribose **Lead compound as reference**	-3.07	-9.7 to -7.9 kcal/mol 7 hydrogen bonds: CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) MET 1169 – N3 (adenosine) HOH 1704 – O3 (ribose) HOH 1704 – O2 (ribose) PHE 1145 – N1 (linker) ARG 1312 – O2 (cytosine)		
Ring — N(CH₃)—Ribose	-3.07	-9.8 to -8.7 kcal/mol 5 hydrogen bonds: CYS 1191 – N6 (adenosine) CYS 1191 – N1 (adenosine) MET 1169 – N3 (adenosine) ARG 1310 – O2 (cytosine) ARG 1312 – O2 (cytosine)		
Ring — N(—OH)—Ribose	-2.67	-10.4 to -8.5 kcal/mol 6 hydrogen bonds: CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) MET 1169 – N3 (adenosine) HOH 1704 – O2 (ribose) GLU 1168 – O3 (ribose) ARG 1312 – O2 (cytosine)	n.a.	n.a.
Ring — N⁺(H)(H)—Ribose	-2.17	did not dock to the binding site cavity	n.a.	n.a.
Ring — N(H)—Ribose	-2.17	did not dock to the binding site cavity	n.a.	n.a.

Structure modification	LogD @ pH=7.4	Chimera-Vina Docking interaction energy range H bonds with Dnmt1	Gromacs MM/MD frames Cys1226 ringC6 distance relative to the first frame	Gromacs MM/MD frames RMSD Ligand relative to the first frame
Ring—N(CH3)—Ribose	-1.34	-10.2 to -8.1 kcal/mol 6 hydrogen bonds: CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) MET 1169 – N3 (adenosine) GLU 1168 – O3 (ribose) GLU 1168 – O2 (ribose) ARG 1312 – O2 (cytosine)		
Ring—N+(H)(CH3)—Ribose	-1.34	did not dock to the binding site cavity	n.a	n.a
Ring—N(CH2CH2OH)—Ribose	-2.08	did not dock to the binding site cavity	n.a	n.a

Table 3B. Evaluation of the linker part of the lead compound; changes in the linker flexibility:
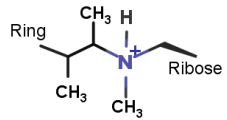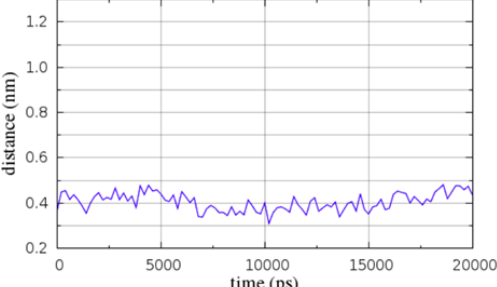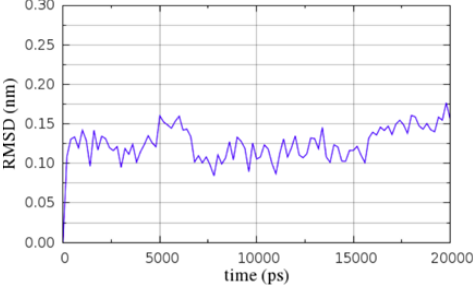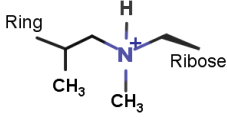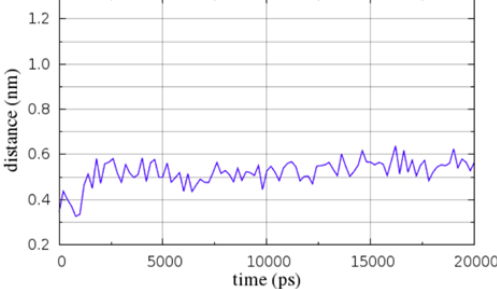
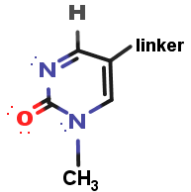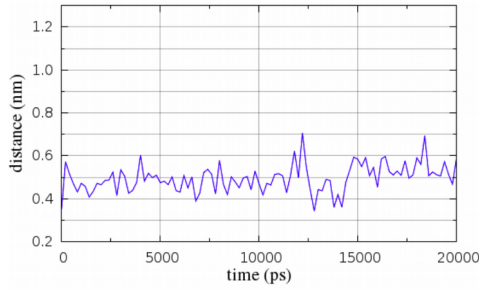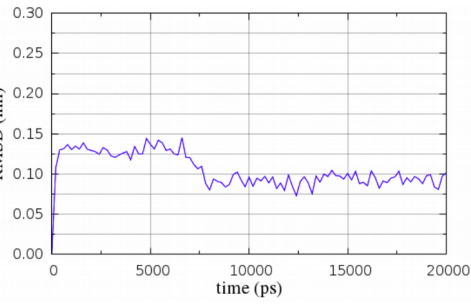| Structure modification	LogD @ pH=7.4	Chimera-Vina Docking interaction energy range H bonds with Dnmt1	Gromacs MM/MD frames Cys1226 ringC6 distance relative to the first frame	Gromacs MM/MD frames RMSD Ligand relative to the first frame
Ring—CH=CH—N+(H)(CH3)—Ribose	-2.16	-10.5 to -8.2 kcal/mol 7 hydrogen bonds: CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) MET 1169 – N3 (adenosine) GLU 1168 – O3 (ribose) GLU 1168 – O2 (ribose) PHE 1145 – N1 (linker) ARG 1312 – O2 (cytosine)		

Structure	Value	Binding details	Distance plot	RMSD plot
Ring =\N+(H)(H) CH2CH3 Ribose	-2.41	-11.1 to -9.1 kcal/mol 7 hydrogen bonds: CYS 1169 – N1 (adenosine) CYS 1169 – N6 (adenosine) MET 1169 – N3 (adenosine) PHE 1145 – N1 (linker) ARG 1310 – O2 (cytosine) ARG 1312 – O2 (cytosine) ARG 1312 – N3 (cytosine)		
Ring =N Ribose	-0.42	did not dock to the binding site cavity	n.a	n.a
Ring Ribose	1.44	did not dock to the binding site cavity	n.a	n.a
Ring Ribose	1.08	did not dock to the binding site cavity	n.a	n.a
Ring Ribose	1.08	did not dock to the binding site cavity	n.a	n.a
Ring Ribose	0.87	-10.2 to -8.3 kcal/mol 4 hydrogen bonds: CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) MET 1169 – N3 (adenosine) ARG 1312 – O2 (cytosine)		

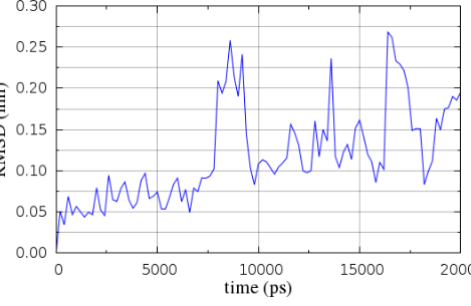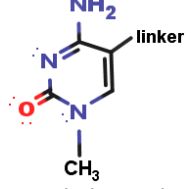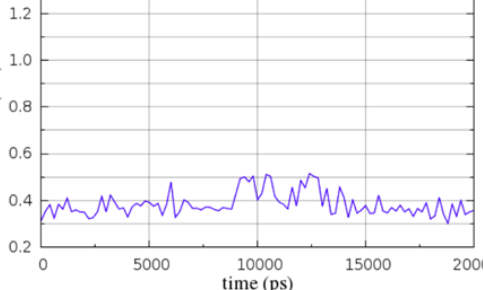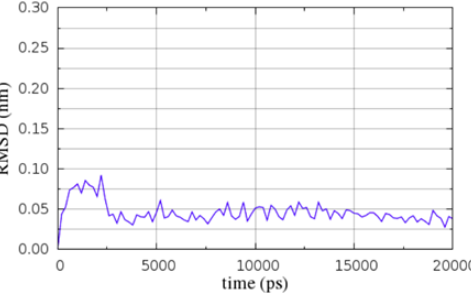| Structure modification	LogD @ pH=7.4	Chimera-Vina Docking interaction energy range H bonds with Dnmt1	Gromacs MM/MD frames Cys1226 ringC6 distance relative to the first frame	Gromacs MM/MD frames RMSD Ligand relative to the first frame
Ring —\\—O—/ Ribose	0.06	-10.2 to -8.9 kcal/mol 8 hydrogen bonds: CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) MET 1169 – N3 (adenosine) GLU 1168 – O3 (ribose) GLU 1168 – O2 (ribose) HOH 1704 – O3 (ribose) HOH 1704 – O2 (ribose) ARG 1312 – O2 (cytosine)		
Ring —\\—S—/ Ribose	-1.21	-10.4 to -8.2 kcal/mol 5 hydrogen bonds CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) MET 1169 – N3 (adenosine) ARG 1310 – O2 (cytosine) ARG 1312 – O2 (cytosine)		
Ring —\\—O—/ Ribose	0.07	-10.9 to -8.3 kcal/mol 7 hydrogen bonds CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) MET 1169 – N3 (adenosine) GLU 1168 – O3 (ribose) GLU 1168 – O2 (ribose) ARG 1310 – O2 (cytosine) ARG 1312 – O2 (cytosine)		
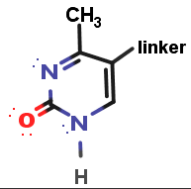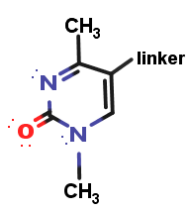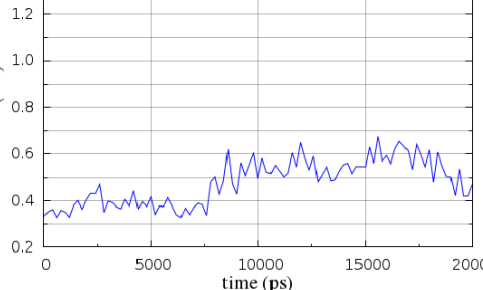
Table 3C. Evaluation of the linker part of the lead compound; changes in the linker mobility:
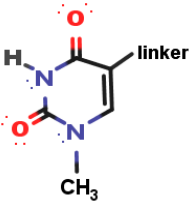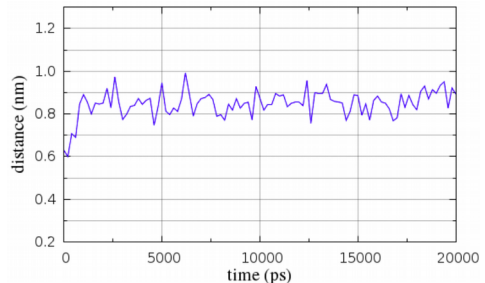
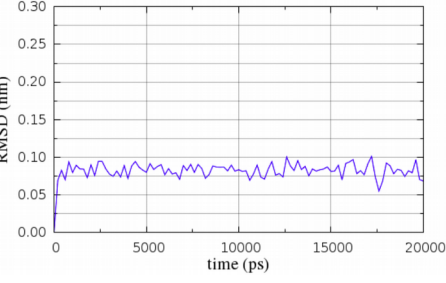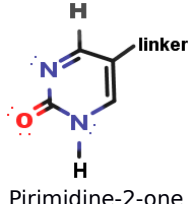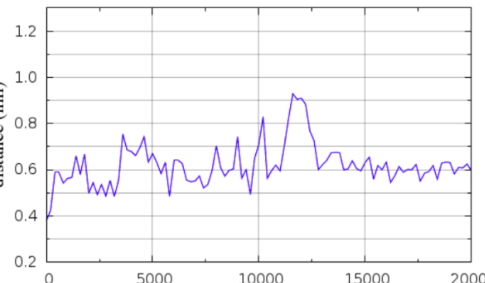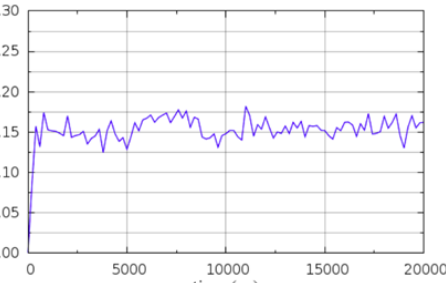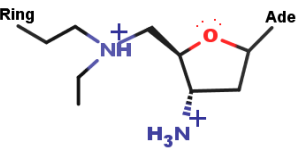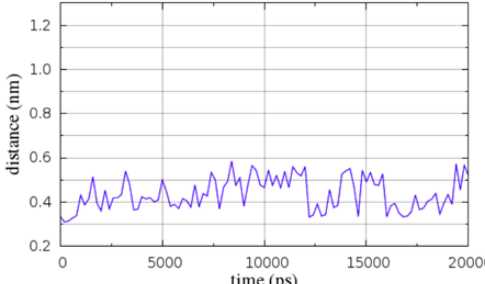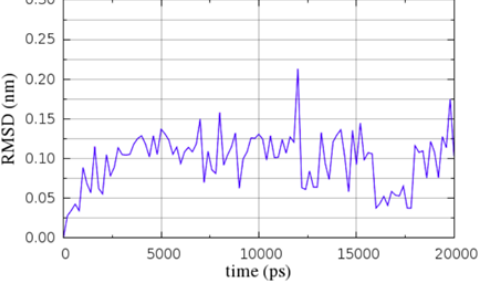| Structure	LUMO eV	Chimera-Vina Docking interaction energy range H bonds with Dnmt1	Gromacs MM/MD frames Cys1226 ringC6 distance relative to the first frame	Gromacs MM/MD frames RMSD Ligand relative to the first frame
	-2.84	-9.4 to -7.8 kcal/mol 5 hydrogen bonds CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) MET 1169 – N3 (adenosine) GLU 1168 – O3 (ribose) GLU 1168 – O2 (ribose)		
	-2.99	-9.8 to – 8.0 kcal/mol 6 hydrogen bonds CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) MET 1169 – N3 (adenosine) GLU 1168 – O3 (ribose) GLU 1168 – O2 (ribose) ARG 1312 – O2 (cytosine)		
	-2.92	-9.3 to -8.2 kcal/mol 8 hydrogen bonds CYS 463 – N1 (adenosine) CYS 463 – N6 (adenosine) MET 441 – N3 (adenosine) GLU 1168 – O3 (ribose) GLU 1168 – O2 (ribose) ARG 1310 – O2 cytosine) ARG 1310 – O2 (cytosine) ARG 1312 – N3 (cytosine)		

Table 4. Evaluation of binding interactions at the target base ring (LUMO energies were calculated using program Gamess **)**

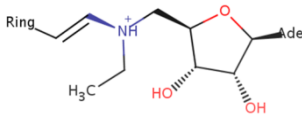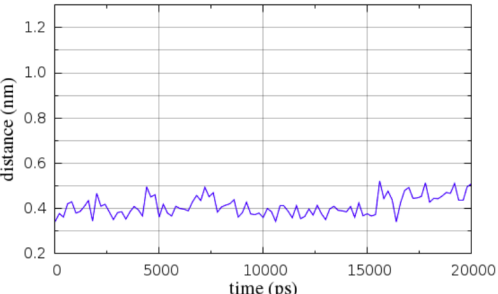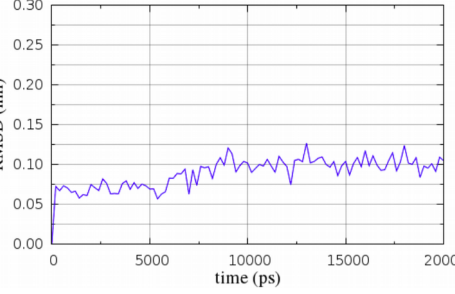| Structure			
1-methyl-pirimidin-2-one (lead compound)	-1.4	-9.7 to -7.9 kcal/mol 7 hydrogen bonds: CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) MET 1169 – N3 (adenosine) GLU 1168 – O3 (ribose) GLU 1168 – O2 (ribose) PHE 1145 – N1 (linker) ARG 1312 – O2 (cytosine)	
	-1.4	did not dock to the binding site cavity	n.a n.a
	-1.5	-9.2 to -6.5 kcal/mol 6 hydrogen bonds: CYS 1191 – N1 (adenosine) MET 1169 – N3 (adenosine) GLU 1168 – O3 (ribose) GLU 1168 – O2 (ribose) PHE 1145 – N1 (linker) ARG 1312 – O2 (cytosine)	
1-methyl-cytosine	-0.9	-10.1 to -8.6 kcal/mol 12 hydrogen bonds: CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) ASP 1190 – N6 (adenosine) MET 1169 – N3 (adenosine) HOH 1704 – O2 (ribose) HOH 1704 – O3 (ribose) GLU 1168 – O2 (ribose) PRO 1224 – N4 (cytosine) GLU 1266 – N4 (cytosine) ARG 1312 – N3 (cytosine) ARG 1312 – O2 (cytosine) ARG 1310 – O2 (cytosine)	

30
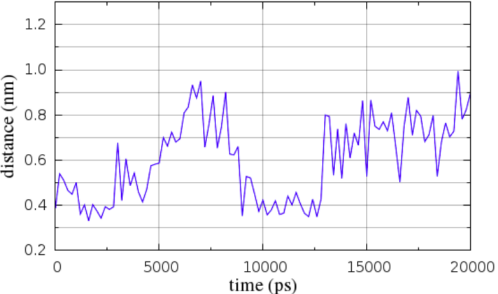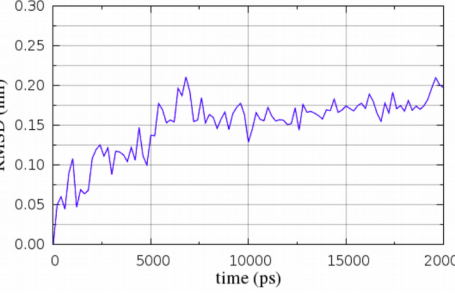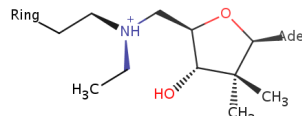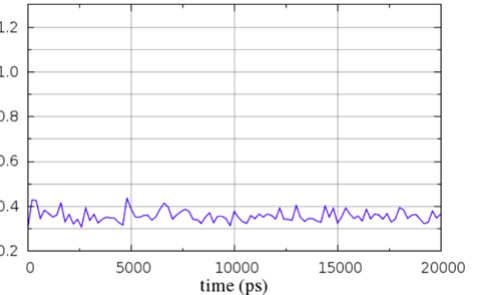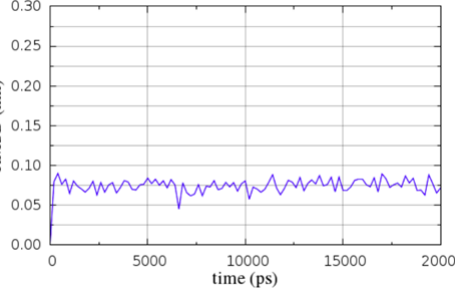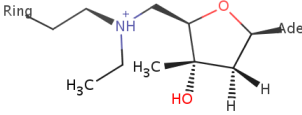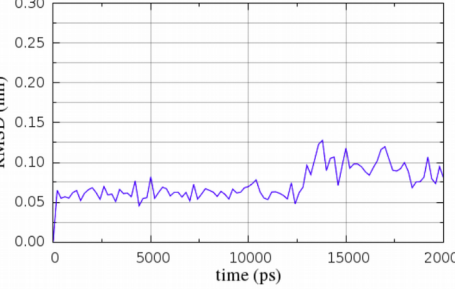
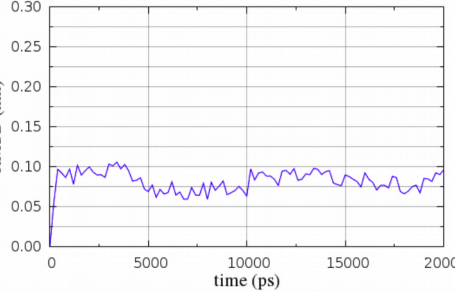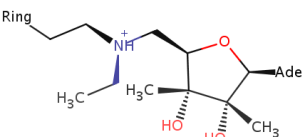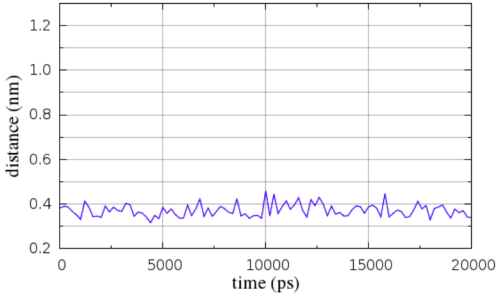| Structure modification	LogD @ pH=7.4	Chimera-Vina Docking interaction energy range H bonds with Dnmt1	Gromacs MM/MD frames Cys1226 ringC6 distance relative to the first frame	Gromacs MM/MD frames RMSD Ligand relative to the first frame
1-methyl-uracil	-0.6	-10.0 to -8.3 kcal/mol 7 hydrogen bonds: CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) MET 1169 – N3 (adenosine) GLU 1168 – O2 (ribose) CYS 1226 – O4 (cytosine) GLU 1266 – N3 (cytosine) VAL 1579 – O2 (cytosine)		
Pirimidine-2-one	-1.4	-10.6 to -8.4 kcal/mol 10 hydrogen bonds: CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) MET 1169 – N3 (adenosine) GLU 1168 – O2 (ribose) HOH 1704 – O2 (ribose) GLU 1168 – O3 (ribose) HOH 1704 – O3 (ribose) PHE 1145 – N1 (linker) ARG 1312 – N3 (cytosine) ARG 1312 – O2 (cytosine)		
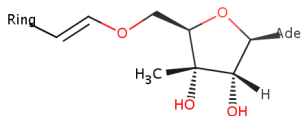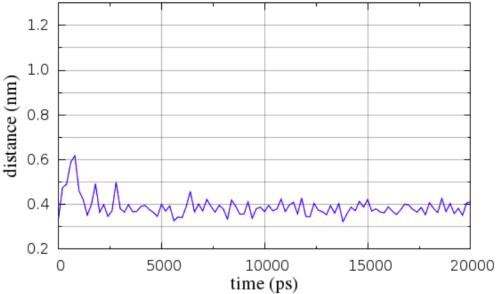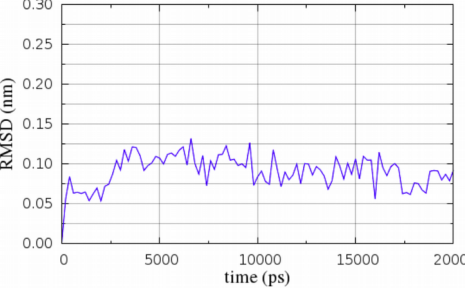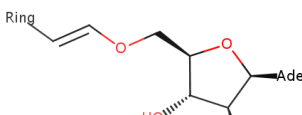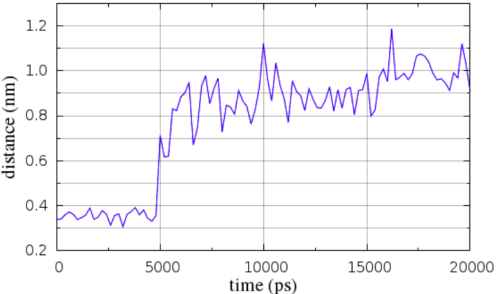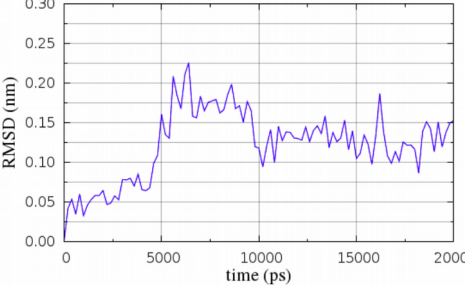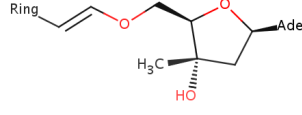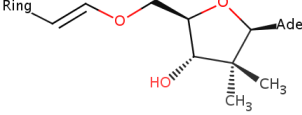
Table 5. Evaluation of the rigidity of the Dnmt1-lignad complex (the LogD values represent calculated values for the lead compound with presented modification that were calculated using weighted option in ChemAxon Marvin . The LogD value for the lead compound is -3.07, thus any modification that has LogD value higher than -3.07 represents a favorable modification)**:**

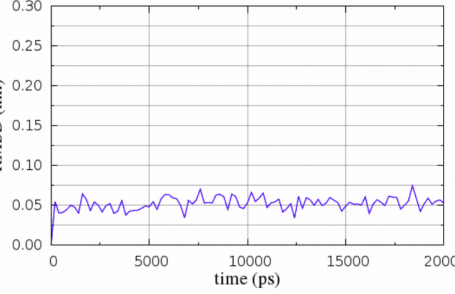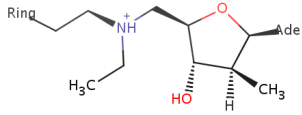| Structure modification	LogD @ pH=7.4	Chimera-Vina Docking interaction energy range H bonds with Dnmt1	Gromacs MM/MD frames Cys1226 ringC6 distance relative to the first frame	Gromacs MM/MD frames RMSD Ligand relative to the first frame
Ring ... Ade	-4.24	-9.3 to -8.3 kcal/mol 6 hydrogen bonds: CYS 1191 – N6 (adenosine) CYS 1191 – N1 (adenosine) GLU 1168 – N3 (ribose) GLY 1223 – N (linker) ARG 1312 – N3 (citozine) ARG 1312 – O (citozine)		

	-1.96	-10.1 to -8.5 kcal/mol 3 hydrogen bonds: CYS 1191 – N6 (adenosine) CYS 1191 – N1 (adenosine) HOH 1704 – O3 (ribose)		
	-1.43	-9.0 to -7.4 kcal/mol 6 hydrogen bonds: CYS 1191 – N6 (adenosine) CYS 1191 – N1 (adenosine) MET 1169 – N3 (adenosine) HOH 1704 – O2 (ribose) HOH 1704 – O3 (ribose) ARG 1312 – O2 (cytosine)		
	-2.46	-9.3 to -7.9 kcal/mol 5 hydrogen bonds: CYS 1191 – N6 (adenosine) CYS 1191 – N1 (adenosine) MET 1169 – N3 (adenosine) ARG 1310 – O2 (citozine) ARG 1312 – O2 (citozine)		
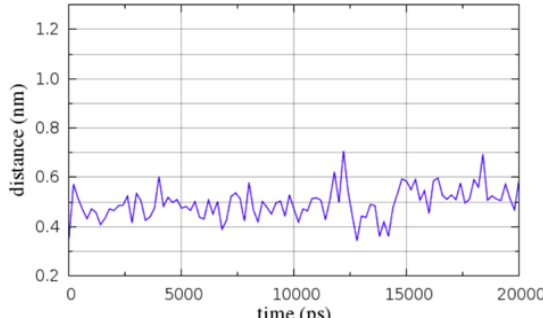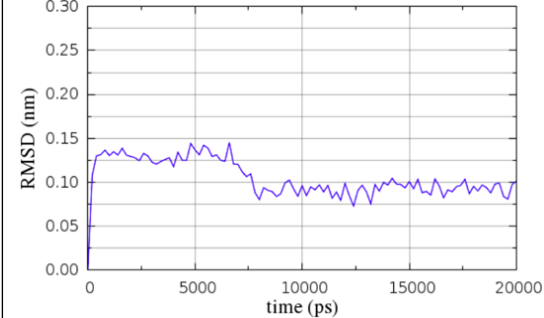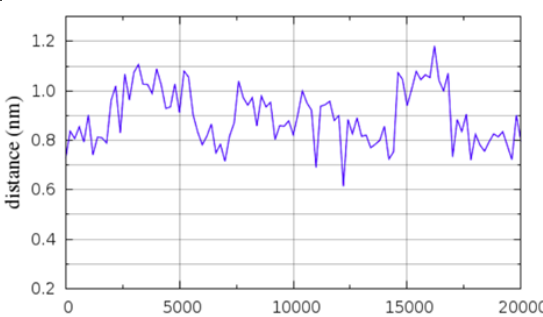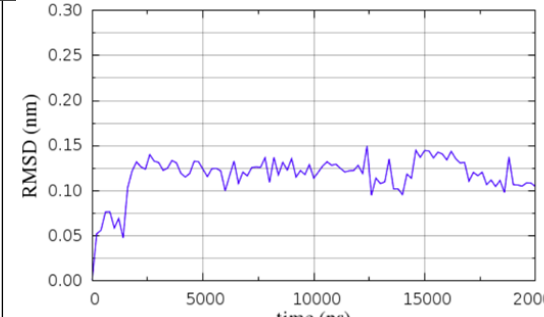
Structure	Value	Binding details	Distance (nm) vs time	RMSD (nm) vs time
	-2.08	-9.9 to -7.8 kcal/mol 5 hydrogen bonds: CYS 1191 – N6 (adenosine) CYS 1191 – N1 (adenosine) MET 1169 – N3 (adenosine) GLU 1168 – O3 (ribose) ARG 1312 – O2 (citozine)		
	-1.70	-8.7 to -7.3 kcal/mol 5 hydrogen bonds: CYS 1191 – N6 (adenosine) CYS 1191 – N1 (adenosine) MET 1169 – N3 (adenosine) GLU 1168 – O3 (ribose) ARG 1312 – O2 (citozine)		
	-2.13	did not dock to the binding site cavity	n.a	n.a
	-1.38	-10.1 to -7.9 kcal/mol 6 hydrogen bonds: CYS 1191 – N6 (adenosine) CYS 1191 – N1 (adenosine) MET 1169 – N3 (adenosine) HOH 1704 – O3 (ribose) ARG 1312 – O2 (cytosine) ARG 1312 – N3 (cytosine)		

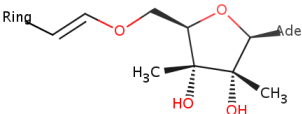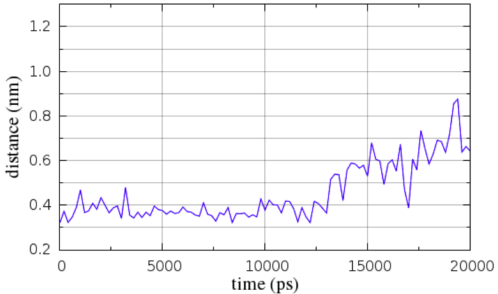| PDB structure	Chimera-Vina Docking interaction energy range H bonds with Dnmt1	Gromacs MM/MD frames Cys1226 ringC6 distance relative to the first frame	Gromacs MM/MD frames RMSD Ligand relative to the first frame
(structure)	-1.66	-10.0 to -7.5 kcal/mol 7 hydrogen bonds: CYS 1191 – N6 (adenosine) CYS 1191 – N1 (adenosine) MET 1169 – N3 (adenosine) HOH 1704 – O2 (ribose) GLU 1168 – O3 (ribose) ARG 1310 – O2 (citozine) ARG 1312 – O2 (citozine)		
(structure)	-0.66	-10.6 to -8.2 kcal/mol 5 hydrogen bonds: CYS 1191 – N6 (adenosine) CYS 1191 – N1 (adenosine) MET 1169 – N3 (adenosine) GLU 1168 – O3 (ribose) ARG 1312 – O2 (citozine)		
(structure)	-0.76	did not dock to the binding site cavity	n.a.	n.a.
(structure)	-0.28	did not dock to the binding site cavity	n.a.	n.a.

Table 6. Binding of the lead compound to Dnmt1 with the active site loop in closed and in open position

4DA4 with the lead compound	-9.7 to -7.9 kcal/mol 7 hydrogen bonds: CYS 1191 – N1 (adenosine) CYS 1191 – N6 (adenosine) MET 1169 – N3 (adenosine) GLU 1168 – O3 (ribose) GLU 1168 – O2 (ribose) PHE 1145 – N1 (linker) ARG 1312 – O2 (cytosine)		
3AV6 with the lead compound	-9.5 to -8.1 kcal/mol 8 hydrogen bonds: CYS 838 – N1 (adenosine) CYS 838 – N6 (adenosine) MET 816 – N3 (adenosine) GLU 815 – O2 (ribose) ASN 1224 – O3 (ribose) GLY 870 – N1 (linker) ARG 957 – O2 (cytosine) ARG 957 – N3 (cytosine)		

7.1. METHODOLOGY

DOCKING ANALYSIS

The initial "rigid-body-docking' calculations used programs Autodock Vina as a plug-in application in UCSF-Chimera . The protein coordinates were acquired from protein data bank, while the ligand structures were drawn in ChemAxon Marvin . Prior to docking the ligand structures were tested with several programs to make sure that conversion between file formats did not distort chiral centers or aromatic bonds. In first step, both the protein and the ligand were protonated at pH=7.2 using Amber99SB force-field tool in Chimera program. Than the search volume box 25.0 X 23.5 X 19.0 Å was prepared, which included AdoHcy binding site, the active site where the target cytosine binds, and the interface between the catalytic domains . The crystal water molecules have been deleted, except for the molecules that had hydrogen bonds with the residues at the AdoMet sites. In the later case, we did docking with and without the water molecules. Docking was judged as successful if at least one of the nine conformers showed overlap with adenine rings and the target cytosine in the corresponding crystal structure (Fig. 3). The hydrogen bonds, and the interaction energies reported by Vina were used only as reference points and have to be taken with caution.

MD SIMULATION AND ANALYSIS

Classical molecular mechanics and molecular dynamics (MM/MMD) simulations were calculated using program package GROMACS . Position-restrained equilibrated MD simulations for large biomolecular systems (>200,000 atoms) were running on heterogeneous CPU/GPU workstations with CUDA-enabled NVIDIA GeForce and Tesla GPUs. Our systems were composed from protein (standard amino acid residues), ligand (non-standard residue) and bulk solvent molecules. Each of these elements was prepared separately. The protein was processed with pdb2gmx, while the ligand was processed with ACPYPE10 . ACPYPE is an interface for Antechamber (part of AmberTools11) which is used to generate various topology types for various forcefields (CHARMM, GROMOS, AMBER, OPLS) . Amber99SB forcefield was used for MD simulations that have non-standard residues such as ligands. We merged all coordinate files into single coordinate file, and all topology files into single topology file after parameterization of the protein and the ligand.

The next step was preparation of solvent box and solvatation of the system. We used rhombic dodecahedron solvent box type optimized to minimize the simulation wall-clock time. TIP3P model was used for water since the solvent can significantly affect protein-ligand interactions. The prepared system was next minimized and equilibration. The minimizations used a combination of steepest descent and conjugate gradient algorithms. When the most stabile state was achieved the temperature was introduced and the system was equilibrated to 310 K (NVT equilibration, V-rescale). Next the pressure was equilibrated to 1 atm (NPT equilibration, Parrinello-Rahman). No restraints were used for the protein or the ligand when the system was minimized, but in NPT and NVT equilibration protein and ligand were restrained. The outputs of minimization and equilibration MD simulations provide insight into the potential energy of the system based on minimization of the temperature in NVT simulation, and minimization of the pressure in NPT simulation.

We calculated position-restrained MD simulations for mammalian Dnmt1 in total duration of 20 ns, 10 million steps, 2 fs step-time, with total of 100 recorded frames. Position-restrained MD simulations were chosen since a large fraction of amino acids in Dnmt1 does not take part when ligands bind to the AdoMet site and the active site. Position restrained MD runs did not restrain the amino acids in catalytic domain (residues 1135 to 1295), and amino acids in the adjacent DNA binding domain (residues 1455 to 1597) . Such strategy was justified by the observed b-factor values and by our earlier experiences with smaller bacterial DNA methyltransferase HhaI (data not published). Restraining could save us about 60% of the wall clock time in each simulation run. Following MM/MD simulations RMSD plots and "Cys 1226 and carbon 6 distance" plots were calculated using built-in GROMACS and VMD utilities .

QM and QM/MM simulations

The frontier molecular orbitals were calculated using program Gamess . The molecules were drawn in Avogadro , the calculations were set-up using "SCFTYP=RHF RUNTYP=ENERGY DFTTYP=B3LYP" protocols with 6-31G basis set, water medium, and initial MO guess set as Huckel. The resulting molecular orbitals and the calculated energies have been visualized using VMD .

CP2K program was used for quantum mechanics (QM) and hybrid quantum mechanics and molecular mechanics (QM/MM) simulation. QM simulations used density functional theory (DFT) that was based on hybrid Gaussian and plane waves density functional . TZV2P-GTH basis sets and BLYP exchange-correlation energy functional were used for 19 QM atoms in the active site. CHARMM force field parameters were used for MM atoms. Forcefield parameters were generated by Acpype for ligand , and taken from VMD for amino acids . PSF topology used in CP2K is generated by VMD using automatic PSF builder plugin .

Nudged elastic band method was used for transition state search . PyMOL is used for selection of QM atoms in QM/MM simulation, and PyMOL plugin was used for generation of QMMM section in CP2K input file. Replicas used in the nudged elastic band were drawn in Avogadro .

8.1 References

Aravind, L., 2011. Natural History of Eukaryotic DNA Methylation Systems Lakshminarayan M. Iyer, Saraswathi Abhiman, and. Modifications of Nuclear DNA and Its Regulatory Proteins, 25.

Babenko, O., Kovalchuk, I., Metz, G.A., 2012. Epigenetic programming of neurodegenerative diseases by an adverse environment. Brain research 1444, 96-111.

ChemAxon, Marvin was used for drawing, displaying and characterizing chemical structures, substructures and reactions, Marvin 15.0.1 (version number), 2015, ChemAxon

Covic, M., Karaca, E., Lie, D., 2010. Epigenetic regulation of neurogenesis in the adult hippocampus. Heredity 105, 122-134.

Daigle, S.R., Olhava, E.J., Therkelsen, C.A., Basavapathruni, A., Jin, L., Boriack-Sjodin, P.A., Allain, C.J., Klaus, C.R., Raimondi, A., Scott, M.P., Waters, N.J., Chesworth, R., Moyer, M.P., Copeland, R.A., Richon, V.M., Pollock, R.M., 2013. Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. Blood 122, 1017-1025.

Di Ruscio, A., Ebralidze, A.K., Benoukraf, T., Amabile, G., Goff, L.A., Terragni, J., Figueroa, M.E., Pontes, L.L.D.F., Alberich-Jorda, M., Zhang, P., 2013. DNMT1-interacting RNAs block gene-specific DNA methylation. Nature 503, 371-376.

Estabrook, R.A., Nguyen, T.T., Fera, N., Reich, N.O., 2009. Coupling sequence-specific recognition to DNA modification. J Biol Chem 284, 22690-22696.

Fersht, A., 1998. Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding (Hardcover), 1st ed. W. H. Freeman; 1st edition

Gordon, M.S., Schmidt, M.W., 2005. Advances in electronic structure theory: GAMESS a decade later. Theory and Applications of Computational Chemistry: the first forty years, 1167-1189.

Gros, C., Fahy, J., Halby, L., Dufau, I., Erdmann, A., Gregoire, J.-M., Ausseil, F., Vispé, S., Arimondo, P.B., 2012. DNA methylation inhibitors in cancer: recent and future approaches. Biochimie 94, 2280-2296.

Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., Tian, C., Wang, J., Cong, Y., Xu, Y., 2015. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature 517, 640-644.

Hanwell, M.D., Curtis, D.E., Lonie, D.C., Vandermeersch, T., Zurek, E., Hutchison, G.R., 2012. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J Cheminform 4, 17.

Heerboth, S., Lapinska, K., Snyder, N., Leary, M., Rollinson, S., Sarkar, S., 2014. Use of epigenetic drugs in disease: an overview. Genetics & epigenetics 6, 9.

Henkelman, G., Jónsson, H., 2000. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. The Journal of chemical physics 113, 9978–9985

Henkelman, G., Uberuaga, B.P., Jónsson, H., 2000. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. The Journal of chemical physics 113, 9901–9904.

Hess, B., Kutzner, C., van der Spoel, D., Lindahl, E., 2008. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. . J. Chem. Theory Comput. 4, 435-447.

Humphrey, W., Dalke, A., Schulten, K., 1996. VMD – Visual Molecular Dynamics. J. Mol. Graph 14, 33-38.

Isakovic, L., Saavedra, O.M., Llewellyn, D.B., Claridge, S., Zhan, L., Bernstein, N., Vaisburg, A., Elowe, N., Petschner, A.J., Rahil, J., 2009. Constrained (l-)-S-adenosyl-l-homocysteine (SAH) analogues as DNA methyltransferase inhibitors. Bioorganic & medicinal chemistry letters 19, 2742-2746.

Khavari, D.A., Sen, G.L., Rinn, J.L., 2010. DNA methylation and epigenetic control of cellular differentiation. Cell Cycle 9, 3880-3883.

Li, S., Du, J., Yang, H., Yin, J., Ding, J., Zhong, J., 2013. Functional and structural characterization of DNMT2 from Spodoptera frugiperda. J Mol Cell Biol 5, 64-66.

Medina-Franco, J.L., Caulfield, T., 2011. Advances in the computational development of DNA methyltransferase inhibitors. Drug discovery today 16, 418-425.

Merino, P., 2013. Chemical Synthesis of Nucleoside Analogues. John Wiley & Sons.

Pennarossa, G., Maffei, S., Campagnol, M., Tarantini, L., Gandolfi, F., Brevini, T.A., 2013. Brief demethylation step allows the conversion of adult human skin fibroblasts into insulin-secreting cells. Proceedings of the National Academy of Sciences 110, 8948-8953.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera--a visualization system for exploratory research and analysis vesrion J Comput Chem 25, 1605-1612.

Qin, W., Leonhardt, H., Pichler, G., 2011. Regulation of DNA methyltransferase 1 by interactions and modifications. Nucleus 2, 392-402.

Romanoski, C.E., Glass, C.K., Stunnenberg, H.G., Wilson, L., Almouzni, G., 2015. Epigenomics: Roadmap for regulation. Nature 518, 314-316.

Santi, D.V., Hardy, L.W., 1987. Catalytic mechanism and inhibition of tRNA (uracil-5-) methyltransferase: evidence for covalent catalysis. Biochemistry 26, 8599-8606.

Sen, G.L., Reuter, J.A., Webster, D.E., Zhu, L., Khavari, P.A., 2010. DNMT1 maintains progenitor function in self-renewing somatic tissue. Nature 463, 563-567.

Song, J., Rechkoblit, O., Bestor, T.H., Patel, D.J., 2011. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science 331, 1036-1040.

Song, J., Teplova, M., Ishibe-Murakami, S., Patel, D.J., 2012. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science 335, 709-712.

Sousa da Silva, A.W., Vranken, W.F., 2012. ACPYPE - AnteChamber PYthon Parser interfacE. BMC Res Notes 5, 367.

Svedruzic, Z.M., 2008. Mammalian cytosine DNA methyltransferase Dnmt1: enzymatic mechanism, novel mechanism-based inhibitors, and RNA-directed DNA methylation. Curr Med Chem 15, 92-106.

Svedruzic, Z.M., 2011. Dnmt1 structure and function. Prog Mol Biol Transl Sci 101, 221-254.

Svedruzic, Z.M., Reich, N.O., 2004. The mechanism of target base attack in DNA cytosine carbon 5 methylation. Biochem. 43, 11460-11473.

Svedruzic, Z.M., Reich, N.O., 2005a. DNA cytosine C5 methyltransferase Dnmt1: catalysis-dependent release of allosteric inhibition. Biochem. 44, 9472-9485.

Svedruzic, Z.M., Reich, N.O., 2005b. Mechanism of allosteric regulation of Dnmt1's processivity. Biochem. 44, 14977-14988.

Takeshita, K., Suetake, I., Yamashita, E., Suga, M., Narita, H., Nakagawa, A., Tajima, S., 2011. Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). Proc Natl Acad Sci U S A 108, 9055-9059.

Trott, O., Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31, 455-461.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., Berendsen, H.J., 2005. GROMACS: fast, flexible, and free. J Comput Chem 26, 1701-1718.

VandeVondele, J., Krack, M., Mohamed, F., Parrinello, M., Chassaing, T., Hutter, J., 2005. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. Computer Physics Communications 167, 103-128

Vorbrüggen, H., Ruh-Pohlenz, C., 2001. Handbook of Nucleoside Synthesis. Wiley.

Wu, J.C., Santi, D.V., 1987. Kinetic and catalytic mechanism of HhaI methyltransferase. J Biol Chem 262, 4778-4786.

Xu, F., Mao, C., Ding, Y., Rui, C., Wu, L., Shi, A., Zhang, H., Zhang, L., Xu, Z., 2010. Molecular and enzymatic profiles of mammalian DNA methyltransferases: structures and targets for drugs. Current medicinal chemistry 17, 4052.

Yang, J., Lior-Hoffmann, L., Wang, S., Zhang, Y., Broyde, S., 2013. DNA cytosine methylation: structural and thermodynamic characterization of the epigenetic marking mechanism. Biochemistry 52, 2828-2838.

Yoo, J., Kim, J.H., Robertson, K.D., Medina-Franco, J.L., 2012. Molecular modeling of inhibitors of human DNA methyltransferase with a crystal structure: discovery of a novel DNMT1 inhibitor. Advances in protein chemistry and structural biology 87.

Yoo, J., L Medina-Franco, J., 2012. Inhibitors of DNA methyltransferases: insights from computational studies. Current medicinal chemistry 19, 3475-3487.

Zhou, L., Cheng, X., Connolly, B.A., Dickman, M.J., Hurd, P.J., Hornby, D.P., 2002. Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. J Mol Biol 321, 591-599.