Theor Chem Acc (2010) 125:407–418
DOI 10.1007/s00214-009-0681-2

REGULAR ARTICLE

Implications of fast-time scale dynamics of human DNA/RNA cytosine methyltransferases (DNMTs) for protein function

David A. Evans · Agnieszka Katarzyna Bronowska

Received: 2 February 2009 / Accepted: 31 October 2009 / Published online: 18 November 2009

Abstract The role of protein dynamics in the control of substrate recognition, catalysis, and protein–protein interactions is often underestimated. Recently, a number of studies have examined the contribution of protein dynamics to the thermodynamics of ligand binding in detail, mostly using NMR relaxation measurements and molecular dynamics (MD) simulations. The results unequivocally demonstrate that conformational dynamics play a pivotal role in the properties and functions of proteins, and ignoring this contribution is likely to lead to substantial errors when explaining the biological function of proteins and in predictions of the binding affinities of their cognate ligands. However, the details of the interplay between structure and dynamics and the way it affects the biological function of the target protein remain poorly understood. In this study, the changes in fast (picosecond-to-nanosecond time scale) dynamics of catalytic domains of four human cytosine DNA methyltransferases (DNMTs) were studied using molecular dynamics (MD) simulations. The results provide insight into the protein dynamics changes that occur upon binding of the cofactor, *S*-adenosylmethionine (SAM). Contrary to expectations, increased amplitude of motions of backbone amide (N–H) and terminal heavy atom (C–C) bond vectors was observed in all studied DNMTs upon binding of SAM. These results imply that the cofactor binding causes a global increase in the extent of protein dynamics in the short time scale. This global dynamic change constitutes a favourable entropic contribution to the free energy of SAM binding. These results suggest that cytosine DNA methyltransferases may exploit changes in their fast scale dynamics to reduce the entropic cost of the substrate binding.

Keywords MD simulations · Protein dynamics · Conformational entropy · DNA/RNA methyltransferase · Drug design · Ligand–protein interactions · Dynamic allostery · Enthalpy–entropy compensation

1 Introduction

The methylation of mammalian DNA, primarily at CpG dinucleotides, plays a pivotal role in controlling gene expression, embryogenesis, and pathophysiology of cancer [1, 2]. Aberrant methylation, in particular hypermethylation of promoter regions, is observed in nearly all steps of tumour progression [3]. Improper methylation patterns are also the cause of several major pathologies, including developmental disorders involving chromosome instability [4] and mental retardation [5]. Despite their biological relevance and the substantial body of structural and biochemical data available for proteins catalysing the methylation reaction, the picture of molecular mechanisms

Dedicated to Professor Sandor Suhai on the occasion of his 65th birthday and published as part of the Suhai Festschrift Issue.

Electronic supplementary material The online version of this article (doi:10.1007/s00214-009-0681-2) contains supplementary material, which is available to authorized users.

D. A. Evans
School of Electronic and Electrical Engineering,
University of Leeds, Leeds LS2 9JT, UK

A. K. Bronowska (✉)
Institute of Organic Chemistry and Biochemistry,
Czech Academy of Sciences, Flemingovo Namesti 2,
166-10 Prague, Czech Republic
e-mail: bronka@tiger.chem.uw.edu.pl

governing the enzymatic machinery involved in methylation and binding specificity of inhibitors targeting DNA-specific methyltransferases (DNMTs) is far from being complete.

In general, it is accepted that protein function is controlled by both structure and dynamics. The 'structural' aspects of structure/activity relationships of DNMTs are fairly well described and understood, while the 'dynamic' aspects are often neglected. Since the crystal structures of DNMT2, and more recently DNMT3A, have been solved, and sequence similarity of the catalytic domains of DNMTs is quite high, making accurate predictions of 3-D structures of DNMT complexes of interest is possible using methods such as homology modelling. Moreover, the likelihood of such models accurately representing the structures of target DNMTs is high. This should allow rational design of selective and highly specific inhibitors of the DNMT of choice. However, such a structure-centred approach, neglecting the influence of the protein dynamics, may lead to misinterpretation of the data and is likely to cause discrepancies between in silico predictions and in vitro experiments.

In this study, a closer look into the influence of the cofactor binding on the global dynamics of several human DNMTs was taken. A common view of ligand–protein interactions regards the shape complementarity and favourable enthalpy, which arises directly from the ligand–protein interactions, as their major driving forces. Solvation effects and ligand conformational entropy loss, which result from the restriction in conformational degrees of freedom upon binding, and loss of free diffusion of one component with respect to the other are often quite rightly taken into account. In such a view, however, ligand–protein interactions would be expected to be characterised by an unfavourable entropy change [6], and this is not the case [7–12]. Some interactions are entropy-driven, and others, despite being enthalpy-driven, have entropic contributions more favourable than may be expected [13]. In these circumstances, the often-neglected contribution from the protein is likely to compensate for this entropy loss. Changes in protein dynamics associated with ligand/interactor binding are often assumed to be negative, i.e. it is believed that protein becomes more rigid upon ligand binding. Obviously, these two views cannot co-exist, and enthalpy/entropy interplay upon ligand binding requires closer scrutiny.

This study aims to examine the protein dynamics contribution to the thermodynamic landscape of cofactor binding and its potential influence on biological function of DNMTs. In 2005, Gowher et al. [14] suggested a molecular mechanism of recognition and binding a substrate by DNMT3A and DNMT3B. This was based on vast conformational changes of DNMT3A/3B upon protein–protein interactions with DNMT3L. Such changes, resulting in the DNMT catalytic domain more 'open', would allow efficient binding of substrate and cofactor molecules. In order to explore structure/dynamics changes, accompanying cofactor binding to DNMTs, all-atom molecular dynamics (MD) simulations of the catalytic domains of human DNMT1, DNMT2, DNMT3A, and DNMT3B were performed. However, no large conformational changes were observed in the course of long (multi-nanosecond) MD simulation on any investigated human DNMTs. Thus, other factors are likely to play a prominent role in regulating of the biological activity of cytosine DNMTs. Herein, we focused on one of those factors, namely, changes in fast-scale protein dynamics.

MD simulations can provide detailed information about the protein motions by representing the state of a protein as a conformational ensemble, which is controlled by the laws of statistical thermodynamics. Importantly, it is possible to compare directly the results of MD simulations with experimental data, mostly obtained by NMR spectroscopy. Such comparisons, performed by numerous groups, indicate an excellent agreement between NMR measurements: heteronuclear spin relaxation and residual dipolar coupling data and MD simulations [13, 15–18]. This enables great confidence in MD results, provided appropriate force field and sufficient trajectory length.

The results obtained shed light on the mechanism of cofactor-dependent thermodynamic compensation employed by DNMTs, and may be useful for rationalising the binding affinities of potent and selective DNMT inhibitors.

2 Methods

2.1 Structure preparation

Initial coordinates of DNMT2 and DNMT3A were based on crystal structures (PDB codes 1G55 and 2QRV, respectively). Hydrogen atoms and some missing heavy atoms were added using the XleaP module of Amber [19]. The structure of SAM was based on the initial geometry of its derivative, *S*-adenosylhomocysteine (SAH), present in crystal structures of DNMT2 and DNMT3A. The missing methyl group was added in MOLDEN [20], and the molecule was optimised with the ab initio Hartree–Fock (HF) method and the 6-31G* (HF/6-31G*) basis set using Gaussian 98 [21]. Restrained electrostatic potential (RESP) charges were subsequently generated and fitted [22]. The ligand molecule was modelled using parm99SB force field, originally developed by Cornell et al. [23] and modified by Hornak et al. [24], with additional parameters for SAM specified and described by Markham et al. [25], Pavelites et al. [26], and Walker et al. [27].

2.2 Construction of homology models

Initial models of the catalytic domains of human DNMT1 (residues 1,138–1,593) and DNMT3B (residues 566–853) were constructed by homology to the catalytic domain of bacterial HhaI (PDB code 2HMY) and human DNMT3A (PDB code 2QRV), respectively. Since the degree of homology between DNMT3A and DNMT3B is very high (about 80% sequence identity and a further 15% similarity), the amino acid residues that required substitution were changed manually. Missing hydrogen atoms were added using XleaP and the apo and holo structures were processed in the same way as structures derived directly from the crystal data.

In the case of DNMT1, due to significantly lower degree of sequence homology between target and template, an initial structure was generated by the MODELLER programme [28, 29]. Hydrogen atoms were added to the model, and the subsequent refinement procedure consisted of: 1,500 cycles of all-atom molecular mechanics (MM) energy minimisation; 500 ps of MD simulation, during which helices and beta-sheets of the protein were kept fixed, while unstructured loops were allowed to move; 500 ps of MD simulation, during which protein side chains were allowed to move, while the backbone conformation remained fixed; and 9 ns of MD simulation with no restraints. The temperature was kept at 300 K. The coordinates obtained during last 6 ns of MD were used for generating starting structures of 32 replicas, which were then subjected to 25-ns replica-exchange MD simulation (REMD) with implicit solvent (Generalised Born, 15 Å cut-off for non-bonded interactions). The coordinates of the lowest-temperature replica were averaged and energy-minimised.

2.3 MD simulation protocol

All simulations were carried out using AMBER 8, with the parm99SB field [24]. All apo-DNMTs and SAM-DNMT complexes were initially subjected to 5,000 cycles of MM energy minimisation. Because the purpose of MD simulations was to investigate the dynamics of the complexes, long simulation trajectories were required, each greater than 30 ns. All MD simulations were carried out at constant temperature and pressure (300 K, 1 atm), with periodic boundary conditions, 12 Å cutoff for non-bonded interactions and 2 fs time-step. A TIP3P explicit water model was used in order to include solvation effects [30]. In all systems, the solute was immersed in a cubic water box with boundary >10 Å. Potassium counter-ions were appropriately placed in order to maintain neutrality of the simulated system. Particle Mesh Ewald was used and SHAKE constraints [31] were applied to all hydrogen atoms during MD simulations to eliminate the fastest X–H vibrations and allow a longer simulation time-step. Translational and rotational centre-of-mass motions were removed every 10 ps. Equilibration started with the gradual heating of the complexes to the target temperature, while the atomic positions of the protein backbone atoms were harmonically restrained. As the temperature increased, the restraints were gradually released, from 25 kcal/(mol Å^2) to zero. The equilibration period took 10 ns, while the production phase took the remaining 35 ns. The coordinates were saved every 1 ps during the production period, averaged, and, finally, energy-minimised.

2.4 Generalised order parameters (S^2) and conformational entropy calculations

In order to understand the thermodynamic implications of the measured changes in dynamics, generalised order parameters and entropies were calculated from the MD trajectories. The method is described in detail by MacRaild et al. [13]. Briefly, it exploits the relationship between the Lipari–Szabo generalised order parameter (S^2) and conformational entropy derived by Yang and Kay [32]. Generalised order parameters are calculated from the MD trajectory of individual bond vectors as:

$$S^2_{\mathrm{LZ}} = \frac{3}{2}\left[\langle x^2\rangle^2 + \langle y^2\rangle^2 + \langle z^2\rangle^2\right] + 2\left[\langle xy\rangle^2 + \langle xz\rangle^2 + \langle yz\rangle^2\right] - \frac{1}{2}$$

where x, y, and z are components of a unit vector along the amide bond, and angular brackets denote the time-average over the trajectory.

Changes in conformational entropy associated with the observed changes in S^2_{LZ} parameters were determined using the relationship described by Yang and Kay [32]. This relationship is essentially independent of motional model for $S^2_{\mathrm{LZ}} < 0.95$, so only residues consistent with this criterion were included in considerations of entropy change.

Trajectories were post-processed and analysed using the ptraj module of AMBER. Solvent molecules were removed and protein backbone atoms were superimposed on the corresponding reference structure. RMS deviations, atomic fluctuations, and structure averaging were calculated directly. The post-processed trajectories were then used to calculate the Lipari-Szabo generalised order parameters [33] and per-residue conformational entropies. Convergence of the dynamics was tested using the approach described by Best and Vendruscolo [34]. Estimation of the statistical error was described by Stöckmann et al. [17].

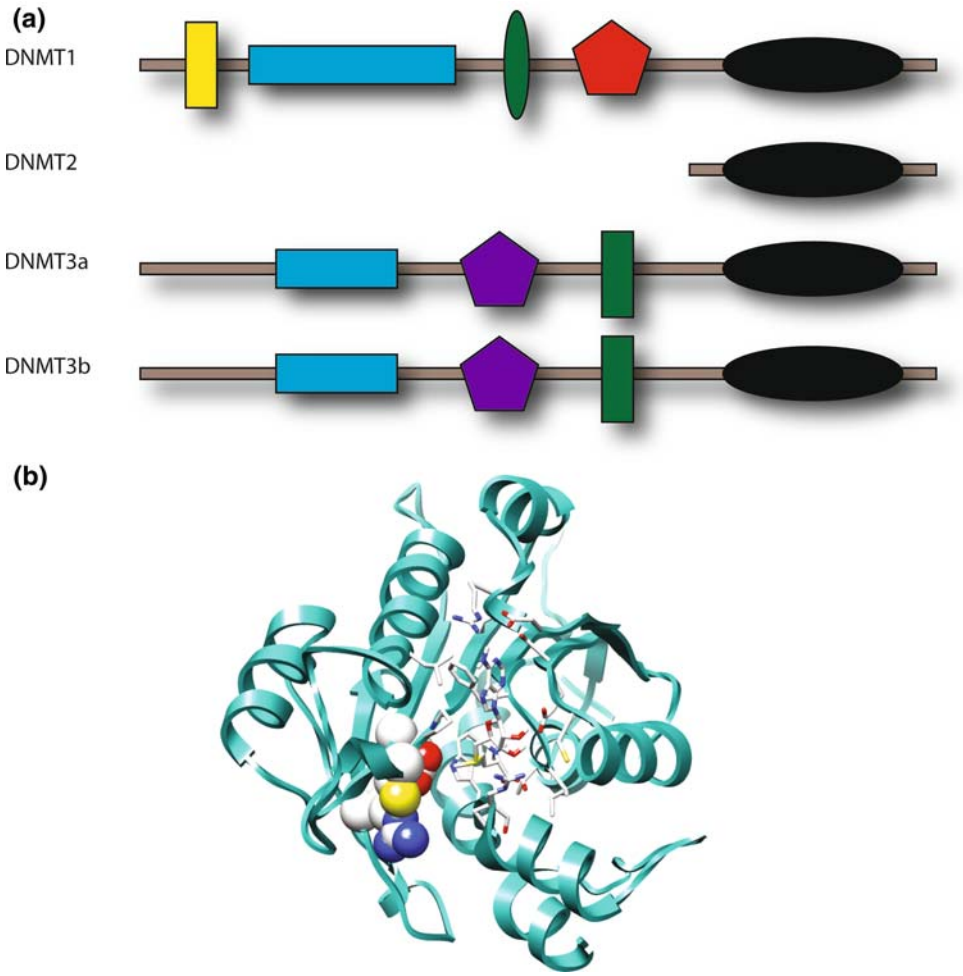

Fig. 1 **a** Graphical representation of the protein domains of four mammalian DNMTs. The domains are represented and coloured as follows: *yellow rectangle* DMAP1 binding domain, *blue rectangle* region with several low-complexity domains, *green ellipse* CXXC-type zinc finger domain, *red pentagon* bromo-adjacent homology domain (BAH), *black ellipse* catalytic DNMT domain, *purple pentagon* PWWP domain, *green rectangle* PHD-type zinc finger domain. **b** Overall 3-D structure of the C-terminal catalytic domain of mammalian DNMTs. The catalytic residues (involving highly conserved PC motif) are coloured by atom and displayed as spheres. The SAM molecule is displayed and coloured by atom

2.5 Trajectory analysis and principal component analysis (PCA)

In order to filter global, collective motions of the protein from local, fast motions, we used principal component analysis (PCA). PCA is a standard mathematical tool used to detect correlations in large data sets, such as MD simulation trajectories. This analysis is based on the diagonalisation of the covariance matrix built from the atomic fluctuations after the removal of the translational and rotational movements. In all simulations, PCA was carried out using the ptraj module of AMBER and PCAZIP software developed by Laughton et al., which uses the methodology described by Meyer et al. [35]. The procedure was performed on the post-processed trajectories, taking into account the alpha carbons, backbone carbonyl carbons and oxygens, and backbone amide hydrogens and nitrogens.

3 Results

Mammalian SAM-dependent DNA methyltransferases (DNMTs) have similar structural organisations. Typically, they consist of an N-terminal region of variable length with multiple regulatory functions, joined to a C-terminal catalytic domain (Fig. 1a). Very long N-terminal domains of DNMT1, DNMT3A, and DNMT3B play important roles in protein–protein interactions [36–39], transcriptional co-repressing [40], and thus maintaining the biological function of DNMTs. In contrast, the catalytic domain of DNMT2 begins at the start of the amino acid sequence. The central part of the protein (residues 124–340) is likely to serve additional functions, such as maintaining interactions with relevant cellular targets.

The overall structures of the catalytic domains of all DNMTs are very similar, resembling a classic Rossmann-fold (Fig. 1b). They can be topologically divided into two

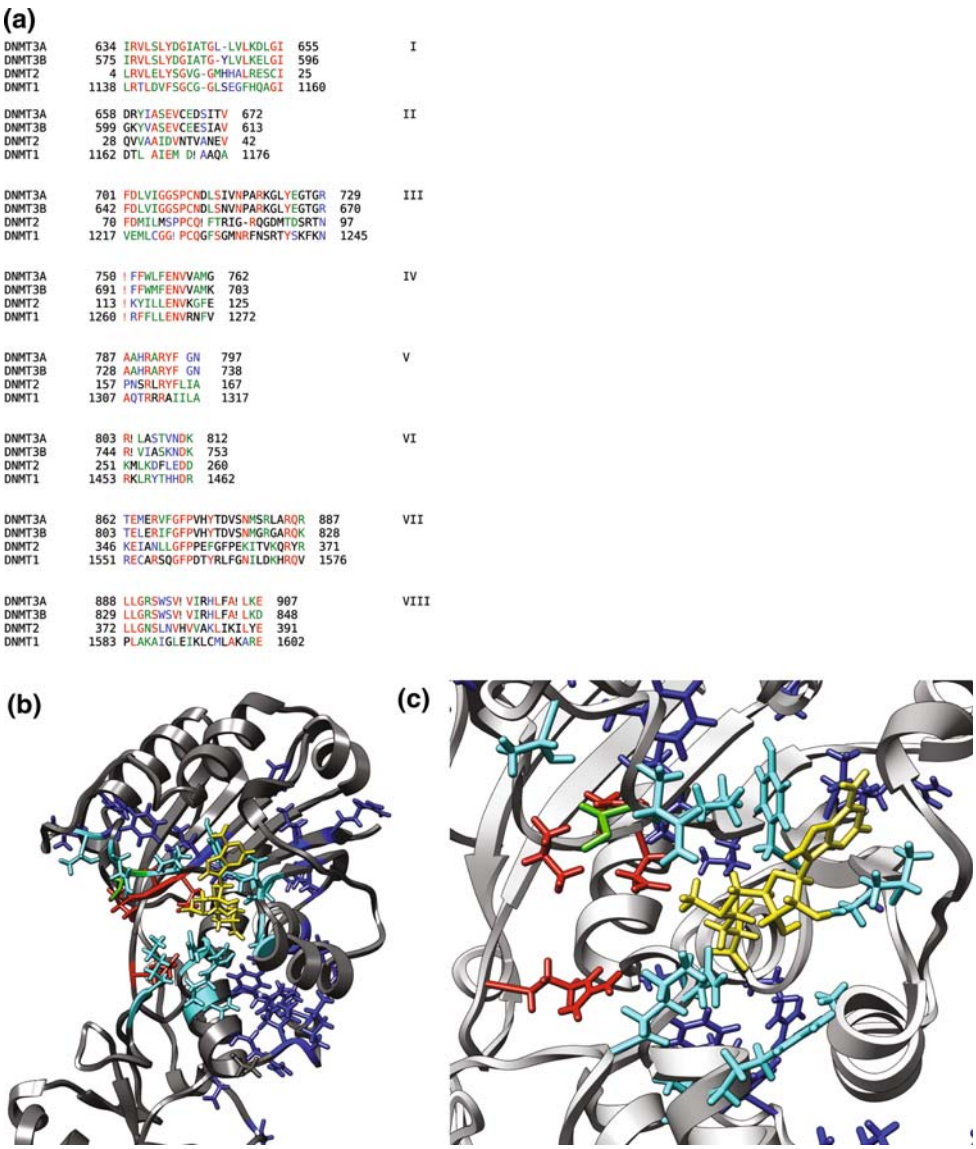

Fig. 2 **a** Eight highly conserved sequence motifs present in catalytic domains of investigated DNMTs. **b** The spatial distribution of highly conserved residues throughout the catalytic domain of DNMT2. The protein backbone is coloured *grey*, the catalytic cysteine (C79) is coloured *green*, the highly conserved ENV motif (residues 119–121), and highly conserved arginine (R160), which are crucial for catalysis, are coloured *red*. The residues, which are involved in direct interactions with either substrate or cofactor, are coloured *cyan*. The remaining conserved residues are coloured *blue*. The cofactor (SAM) molecule is displayed and coloured *yellow*. **c** The orientation of the CFTxxYxxY sequence motif—characteristic for DNMT2 family—towards the binding site of the catalytic domain. The motif is coloured *magenta*, the catalytic cysteine is coloured *green*, and the SAM molecule is coloured *yellow*

distinct regions, linked together by a flexible 'hinge'. The substrate and the cofactor molecules are bound near the interface between these two regions.

The binding pockets of all investigated DNMTs appeared to be well solvated. In both the apo state and in the SAM-DNMT complex, there were a number of ordered water molecules in the binding pocket. On average, we found 11 water molecules in the binding pocket of the SAM-DNMT complex, and 15 water molecules in the binding pocket of the apo-DNMT. These water molecules were hydrogen-bonded to several conserved residues of the DNMTs and, in case of the SAM-DNMT complexes, also to the cofactor molecule. However, all water molecules in the binding pocket were found to be quite mobile and no 'confined' water molecules, which would very likely play the prominent role in the thermodynamics of ligand–protein binding [9], was observed in the course of our simulations.

Although the overall sequence homology of the catalytic domains of all four investigated DNMTs is not very high, sequence analysis showed eight sequence motifs that are highly conserved between them (Fig. 2a). The vast majority of these residues are scattered throughout the substrate binding region and the cofactor binding site (Figs. 2b, 3). Many of these residues are critical for the biological activity of DNMTs and their mutation leads to either partial impairment, or a complete blockage of the protein enzymatic activity [41, 42].

Differences in the binding site architecture of different DNMTs, such as amino acid substitutions (Fig. 3), are evident from the sequence analysis and structural data available. Although these differences can be linked to the observed differences in biological function, it is very likely that factors other than structure play an important role in the control of substrate recognition and binding specificity.

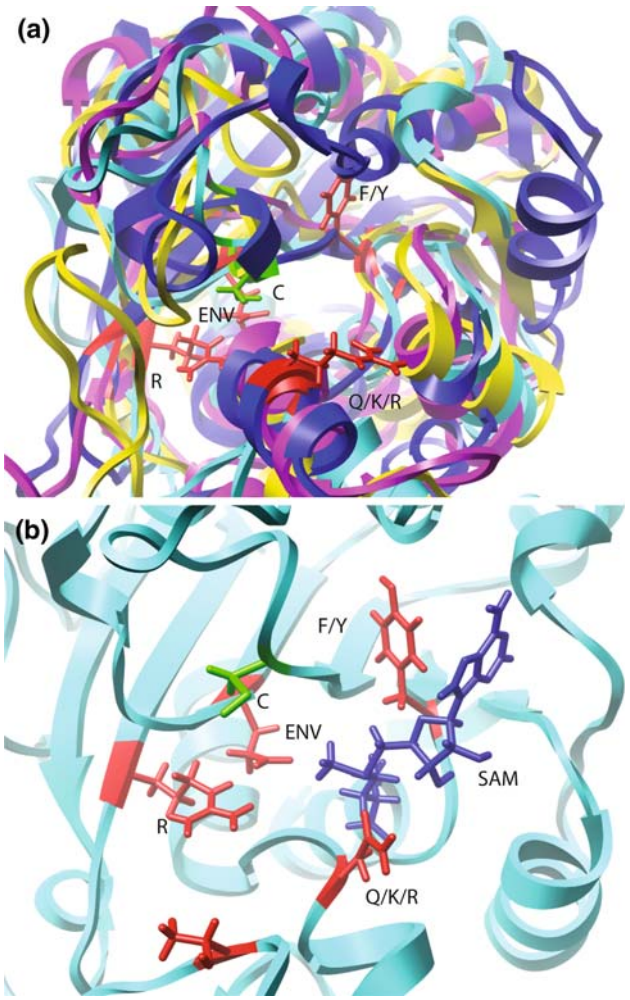

Fig. 3 **a** Superimposed binding pockets of DNMT1 (*blue*), DNMT2 (*cyan*), DNMT3A (*yellow*), and DNMT3B (*magenta*). Several conserved residues are highlighted and displayed. Catalytic cysteine (C) is coloured *green*, glutamic acid from the ENV pocket, Tyr/Phe (Y/F) residue from the motif I, conserved arginine, and Arg/Gln/Lys (R/Q/K) from motif VIII are coloured *red*. **b** More detailed view of the binding pocket of DNMT2. The protein backbone is coloured *cyan*, critical and highly conserved residues are displayed and coloured *red* and *green* (catalytic cysteine), Threonine from the conserved CTFxxYxxY motif is displayed and coloured *red*. The cofactor (SAM) molecule is coloured *blue*

3.1 Protein conformational changes

For each DNMT, three systems were studied: apo protein, SAM-bound, and SAH-bound complexes. Surprisingly, no dramatic conformational changes occurred upon the ligand (either SAM or SAH) binding in any of the investigated DNMTs on the nanosecond time scale. We detected, however, differences in dynamical behaviour of investigated DNMTs on multi-nanosecond time scale, using principal component analysis (PCA). These differences were particularly strong between DNMT2 and remaining DNMTs. In Fig. 4, the projections of the Ca trajectory snapshots of all SAM-bound DNMTs on the first two eigenvectors are shown. The first principal component was a 'sideways' twisting rather than an 'open-to-close' transition of the two parts of the catalytic domain (Fig. 5). Moreover, the loop containing the CFTxxYxxY sequence motif, characteristic for the DNMT2 family, showed a rigid body-like 'wagging' movement, enhanced by SAM binding.

From all studied DNMTs, DNMT2 showed the largest conformational changes (Fig. 4). In terms of PCA, its dynamic behaviour was characterised by a lower amplitude of the opening and closing-like motions of the two subdomains. Also, the averaged structure was the most widely opened, which suggests that the contribution of the 'open' conformation to the ensemble is higher than for the other DNMTs.

3.2 Fast dynamics of the protein backbone and per-residue entropy

To assess the thermodynamic contribution of fast (picosecond-to-nanosecond) motions of the protein backbone to the binding of SAM, a series of MD simulations of apo- and holo- (i.e. SAM-bound) DNMTs were performed. The Lipari-Szabo generalised order parameters (S^2_{LS}) and per-residue entropies for backbone amide bond vectors of the structured core of the apo- and holo- protein were calculated from the equilibrated MD trajectories.

The obtained results indicate that despite similar conformations, the SAM-DNMT complexes and apo-DNMT proteins display very different dynamic behaviour. Overall increases in backbone dynamics over 50 ns of MD simulation were observed in the SAM-containing complexes. The increase was least pronounced in the DNMT2-SAM complex measured with respect to apo-DNMT2 [$\Delta(T\Delta S) = +2.67$ kJ/mol for SAM-DNMT2 vs. apo-DNMT2 at $T = 300$ K], but the trend was, nevertheless, preserved.

To examine whether this trend was specific to SAM, 50 ns of MD simulations of SAH-DNMT2 and SAH-DNMT3A complexes were performed. The conditions of these MD simulations were identical to those used for SAM-DNMTs and apo-DNMTs. The initial conformation of SAH closely resembled that of SAM, with the exception of the missing methyl group, and was based on the crystal structure coordinates (1G55 and 2QRV, respectively).

The results obtained suggested that the observed trend of increasing backbone dynamics upon cofactor binding was SAM-specific. In contrast to the SAM-bound complexes, the SAH-bound complexes showed a decrease in backbone dynamics when compared to apo-DNMTs [$\Delta(T\Delta S) = -8.39$ kJ/mol for DNMT2, -4.77 kJ/mol for DNMT3A, respectively, $T = 300$ K]. SAM and SAH, despite having similar overall structures, have very different molecular

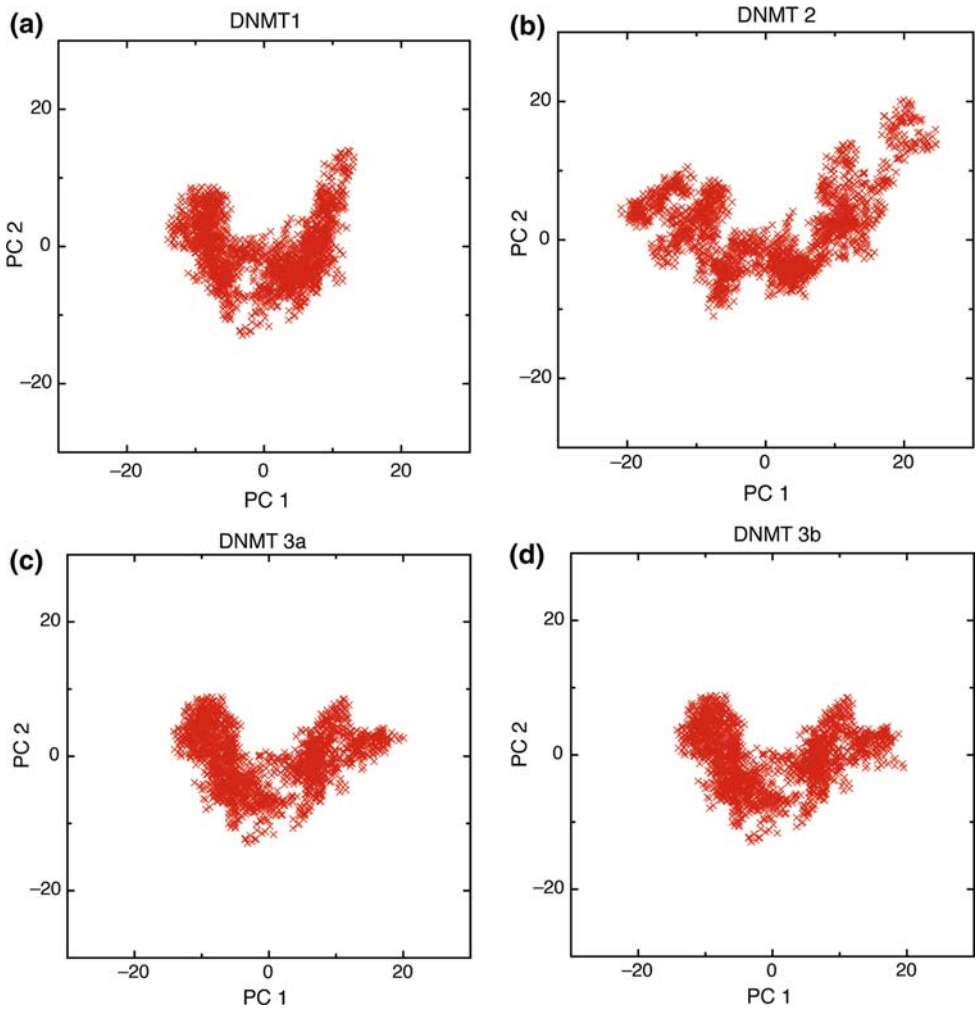

Fig. 4 Projections of trajectory snapshots on first (PC1) and second (PC2) eigenvectors for four investigated DNMTs (SAM-bound)

properties, such as net charge and its distribution. As a result, their interactions with DNMTs would be different, and it is likely that they would affect the protein's structure and dynamics in different ways. The number of charged residues in the DNMT binding pockets means that this influence is likely to be strong. Indeed, the difference in interaction energy measured for SAM- and SAH-bound complexes suggested that SAH is a better binder than SAM, if only the enthalpic contribution is addressed (data not shown). This result, although qualitative rather than quantitative, is supported by isothermal calorimetric (ITC) data for the cytosine DNA methyltransferase, HhaI [43], which is a bacterial orthologue of the investigated DNMTs. Observed changes in protein dynamics, which were of contrasting signs, were, nevertheless, unexpected. On the other hand, such enthalpy–entropy compensation has been described for several related systems, and its presence and role in the thermodynamics of ligand–protein binding is increasingly acknowledged [17].

Differences in atomic fluctuations of the apo-DNMTs and the respective SAM-DNMT complexes, along with differences between the amide backbone S^2_{LS} in the apo-DNMTs and the S^2_{LS} in SAM-DNMTs are shown in Fig. 6. These differences represent changes in 'fast' (picosecond to nanosecond time scale) dynamics of the protein backbone, induced by SAM. It is evident that observed changes in atomic fluctuations and generalised order parameters are consistent with each other and clearly show the gain in 'fast' dynamics upon SAM binding. However, the observed increase in global dynamics upon SAM binding was not uniform. Several residues (often highly conserved ones) in the binding pocket region became more dynamic (e.g. conserved R/K/N gating residue from motif VIII). In contrast, the highly conserved glutamic acid (from the ENV motif) and arginine (motif V) became more rigid in all SAM-DNMT complexes investigated. All DNMTs investigated displayed decreased dynamics of the regions reported to be involved in the protein–protein interactions (such as dimerisation) upon binding of SAM. This is quite surprising since SAM does not interact directly with any of the residues involved in dimerisation. The remaining parts of the catalytic domains of the investigated DNMTs did not

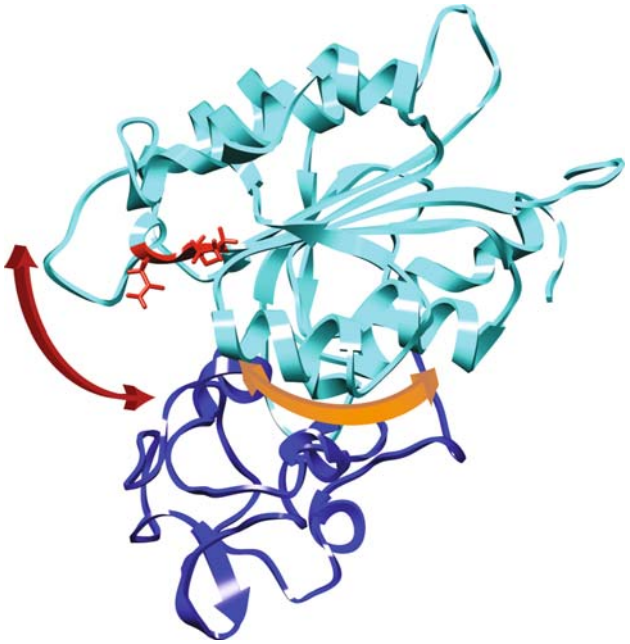

Fig. 5 Schematic representation of two main contributions to protein motions, depicted by principal component analysis: 'opening-and-closing' (*red arrow*) and 'sideways twisting' (*yellow arrow*), showed on the backbone of DNMT3B as an example. For clarity of the presentation, two subdomains of the catalytic domain connected by a flexible 'hinge' region are coloured *cyan* and *blue*. The catalytic residues (PCX) are coloured *red*

show any detectable trend, but overall the SAM binding seemed to increase the global backbone dynamics of the catalytic domains.

Some differences in the fast dynamics could be observed among the different DNMTs. Whether these differences could reflect observed differences in biological function remains speculative, however, particularly for the dynamical behaviour of DNMT2, it seems likely.

The global changes in backbone conformational entropy are shown in Table 1. In all four cases, the global difference between the SAM-DNMT complex and apo-DNMT was always positive and yielded a favourable entropic contribution to the cofactor binding. This difference was largest (i.e. conformational entropy most favourable) in the case of DNMT1. Interestingly, entropic contributions were virtually the same (within error) for DNMT3A and DNMT3B despite very different absolute values of backbone conformational entropy.

3.3 Dynamic contributions of side-chains

In order to investigate the influence of the cofactor on the dynamics of DNMTs, we focused on dynamics of side chains all over the protein, and then specifically those in the binding pocket. Generalised Lipari-Szabo order parameters and entropies were calculated for each terminal heavy-atom bond vector for each side chain for SAM complexes and apo-DNMTs. The results, shown in Table 2, are consistent with the results obtained for backbone amide bond vectors.

Increase in dynamics upon SAM binding was observed for several residues in the binding pocket. This increase was particularly prominent for the highly conserved Y/F of motif I, the N/Q residue, adjacent to the catalytic cysteine (motif III), the nucleophile cysteine, R166, the V residue of the ENV motif, and the K/Q/R and L/I/W residues in motif VIII. The dynamics of several side chains, such as the highly conserved D and N residues in ENV motif, the catalytic arginine from RXR patch (motif V), and the CFTxxYxxY loop, systematically decreased. Nonetheless, the overall dynamics of the binding pocket increased upon SAM binding (Fig. 5). This increase was largest for DNMT3A and smallest for DNMT2.

In combination with different 'slow' dynamical behaviour such as the larger amplitude of 'sideway opening/closing', and the rigid-body movement of CFTxxYxxY loop depicted by PCA, this behaviour may reflect the dramatically different substrate specificity of DNMT2 (e.g. tRNA methylation) as compared to DNA-specific DNMTs.

4 Discussion

In this study, we focused on structural and dynamical features of four mammalian DNMTs, using a combination of homology modelling and molecular dynamics MD simulations.

For structural analysis, the available crystal data of DNMT2 and DNMT3A were used for comparison of 'static' structures, and as a starting point for molecular dynamics simulations and homology modelling of DNMT3B and DNMT1. Our DNMT1 and DNMT3B models were created using a combined approach, which included homology modelling and molecular dynamics (MD) simulations. It is useful to notice that the application potential of homology modelling is limited by the sequence similarities between the template and the target. As a rule of the thumb, homology models based on 30% and greater sequence similarity to experimentally resolved structures may be used to address basic structural questions and to plan experiments such as site-directed mutagenesis [44]. Homology models, which are based on 50% or greater sequence similarity are usually accurate and can be successfully used in drug discovery protocols, including high-throughput docking of small molecule libraries. In this study, the homology model is based on very high conservation of catalytic motifs between individual DNMTs (some of which are of known structures) and high (>50%)

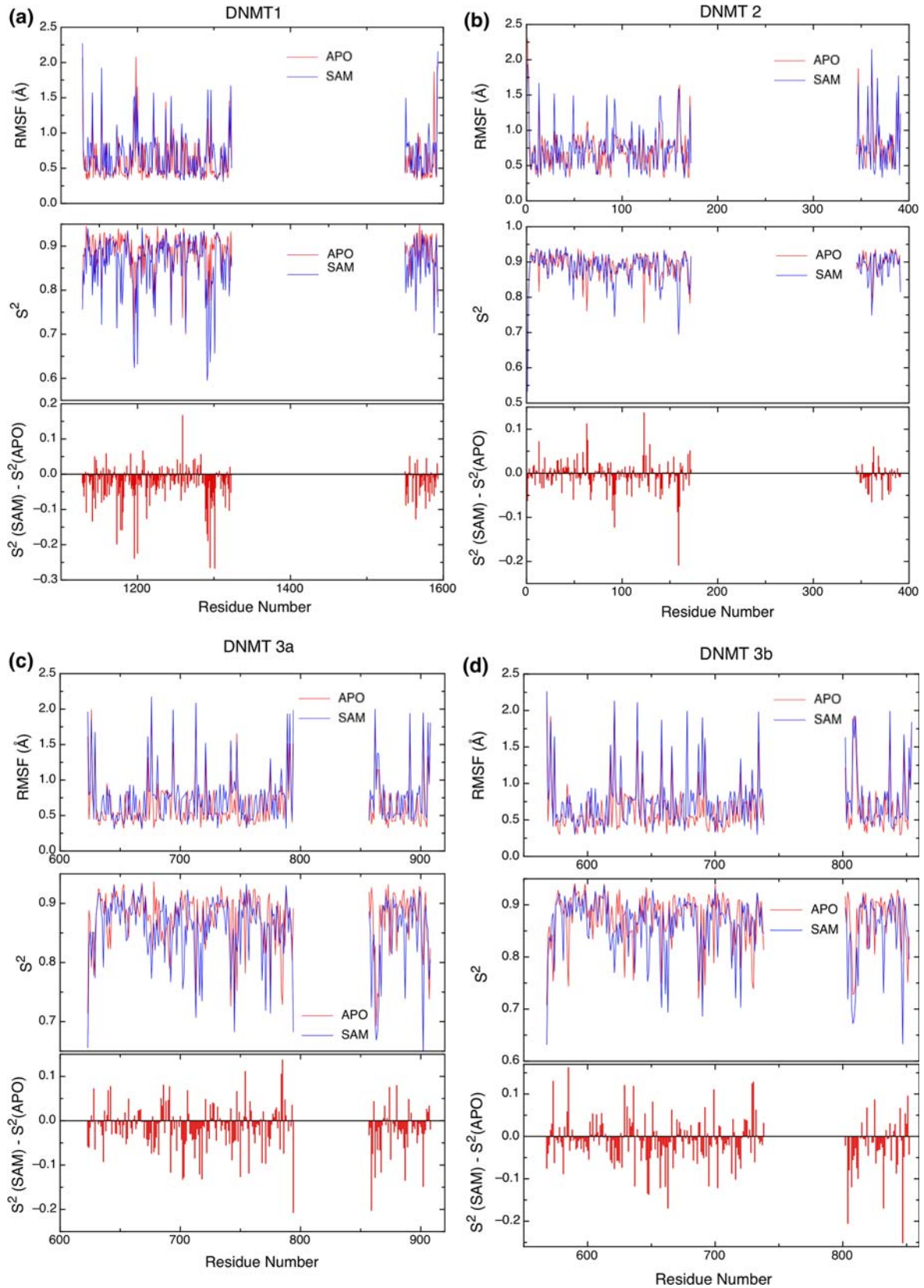

Fig. 6 Per-residue plots of atomic fluctuations (RMSF), generalised order parameters (S^2), and differences in S^2 parameters for apo-DNMTs (*red*) and SAM-bound DNMTs (*blue*)

Table 1 Calculated backbone (N–H) conformational entropy differences (given as $T\Delta S$, where $T = 300$ K) between holo (i.e. SAM-bound) and apo cytosine DNA methyltransferases

	DNMT1	DNMT2	DNMT3A	DNMT3B
$\Delta(T\Delta S)$ (N–H) (HOLO-APO) (kJ/mol)	+35.0	+2.7	+25.8	+25.6

The differences are stated in kJ/mol

Table 2 Calculated side chain (terminal C–C) and binding site region only (BS) conformational entropy differences (given as $T\Delta S$, where $T = 300$ K) between holo (i.e. SAM-bound) and apo cytosine DNA methyltransferases

	DNMT1	DNMT2	DNMT3A	DNMT3B
$\Delta(T\Delta S)$ (C–C) (HOLO-APO) (kJ/mol)	+20.7	+4.9	+20.8	+18.3
$\Delta(T\Delta S)$ (BS) (HOLO-APO) (kJ/mol)	+11.5	−1.1	+8.5	+8.3

The differences are stated in kJ/mol

sequence conservation of the amino acids constituting the ligand binding site. The structure of this model was further improved by a series of MD simulations. Long (50 ns) unrestrained MD simulation was carried out in order to obtain information about protein dynamics in the picosecond-to-nanosecond time scale.

The fundamental question that needs to be addressed relates to the molecular mechanism which is exploited by DNA cytosine DNMTs in order to recognise and bind the substrate. The model proposed by Gowher et al. [14] for DNMT3A and DNMT3B was based on substantial conformational changes of DNMT3A/3B upon protein–protein interactions with DNMT3L. Such changes would allow for the required conformation of the binding site of the catalytic domain of DNMT3A/3B, allowing efficient binding of substrate and cofactor molecules. However, in the course of 50 ns of MD simulation of DNMT3A and three other DNMTs (apo- and SAM-bound), no substantial conformational changes were observed. Valuable information on slower scale motions (approximate range of tens of ns) can be derived from the principal component analysis (PCA), in particular about large domain reorientations. In our study, the displacement of backbone atoms along the first eigenvector having the largest eigenvalue indicated the opening-and-closing of two domains of apo-DNMT3A/3B monomers in a manner resembling the rigid body motion of two subdomains. Such motions have been observed for several structurally related proteins, such as glycolytic enzymes (e.g. lactate and phosphoglycerate dehydrogenases) [45], transcription co-repressors (CtBP1) [46], and carbohydrate binding proteins (ABP) [13]. However, in the case of the DNMTs studied, the difference between 'open' and 'closed' conformation was very minor (superimposed, the 'open' and 'closed' conformation RMSD on backbone heavy atoms was 1.8 Å). PCA indicated, however, that the presence of the cofactor restricted such slow motions, and a similar trend was observed for all DNMTs. Moreover, time-averaged energy minimised SAM-DNMT complexes were very similar to the crystal structures of the DNM3A-SAH and DNMT2-SAH complexes; the starting structures were used as reference points. These observations indicate that the model based on large conformational changes, although ostensively attractive, has certain limitations. This view is also supported by calorimetric data for bacterial cytosine DNA methyltransferase, HhaI, which has a very similar structure and binding properties. Very small changes in heat capacity, measured upon SAM binding to HhaI, rule out very large conformational changes in the interacting species; therefore, some other factors not related to substantial conformational changes (such as domain rearranging) are likely to play a prominent role in regulating of the biological activity of cytosine DNMTs.

Our results also indicate that the binding specificity of investigated cytosine DNA methyltransferases is likely to be regulated by things other than structural differences in their catalytic domains. Although the importance of structure-related factors, such as single amino acid substitutions in strategic positions—depicted in the course of this study by sequence analysis combined with the direct comparison of the structures in binding specificity control—is undisputable; it is very likely that protein dynamics will contribute to the observed differences in target binding specificity. A pivotal role of picosecond-to-nanosecond time scale dynamics in control of ligand–protein associations is becoming increasingly acknowledged in recent years. To date, such fast-scale protein motions were demonstrated to significantly reduce the thermodynamic cost of ligand binding in several protein systems [13, 17].

Counter-intuitively, protein conformational entropy increased upon binding of the cofactor, i.e. protein entropic contribution to binding was favourable. This trend was observed in all investigated SAM-DNMT complexes. This conformational entropy increase was induced by SAM specifically—when SAM was replaced by SAH, the protein became more rigid, i.e. protein entropic contribution to SAH binding was unfavourable. This is consistent with the observed enthalpy–entropy compensation in SAM/SAH binding to HhaI [43]. Large, favourable enthalpic contributions to the interaction energy in SAH-DNMT2 and SAH-DNMT3A complexes, as assessed by molecular mechanical energy minimisation of time-averaged structures, was partially offset by unfavourable protein entropic contribution. In SAM-DNMT2 and SAM-DNMT3A, favourable entropic contribution from protein dynamics was found to compensate for less favourable enthalpic

contribution from ligand–protein interactions when compared to SAH-containing complexes.

ITC measurements demonstrated that binding of SAH was tighter than binding of SAM. Free energies of SAH and SAM binding to HhaI were −31.94 kJ/mol and −28.34 kJ/mol, respectively (at 37°). This is consistent with our observations for DNMT3A and DNMT2. In both cases SAH was found to be a tighter binder. Swaminathan et al. [43] attributed this difference in binding affinity to differences in solvent reorganisation around SAM and SAH. In the course of MD simulations, we observed some differences in amount and distribution of water molecules around SAM and SAH. However, in all cases the binding pockets were well solvated. Though there is no doubt that solvation effects play a crucial role in the control of SAM/SAH binding and are key in governing catalysis, the present results suggest that role of protein dynamics in enthalpy–entropy compensation should also be emphasised.

In the present study, fast protein dynamics were assessed using the Lipari-Szabo model-free formalism [33]. A significant assumption entailed by this analysis is that the rotational diffusion of the protein can be fully characterised by a single diffusive process which is fully separated from the fast internal motions of the protein. This assumption generally holds well for globular, single-domain proteins, but might be called into question in the case of the cytosine DNA methyltransferases, where flexibility of the domain hinge, which is proximal to the binding pocket, may result in a complex interaction between local protein conformation and the rotational diffusion. The study of arabinose-binding protein (ABP), whose overall fold and dynamical behaviour are similar to those of DNMTs, indicates that inter-domain flexibility has only a minor influence on the protein relaxation rates [13]. Moreover, results obtained from the MD simulations indicate that inter-domain motions of DNMTs have little impact on measured S^2 parameters and per-residue conformational entropy.

In the course of the present work, we found that the SAM-induced changes in fast protein dynamics extended beyond the binding pocket. This was unexpected, because such motions are almost exclusively local in their character, with few (if any) correlations over distances longer than a few angstroms. Another surprising finding was the direction of these changes in dynamics. Counter-intuitively, cofactor binding increased global protein dynamics, which was measured for both backbone amides and side chains. Although in the case of DNMT2 the binding of SAM decreased the local dynamics in the binding pocket, the global change in dynamics upon SAM binding was still positive, i.e. protein conformational entropy increased upon the binding of the cofactor molecule.

The results obtained in this study clearly show a consistent, global trend in increase of ‘fast’ dynamics, initiated by the local event of binding of the small molecule (here, *S*-adenosylmethionine). The physical basis for such a change is not clear; however, it is not without precedent in the literature. Several studies have identified favourable changes in pico-to-nanosecond backbone dynamics on ligand binding to be similarly dispersed throughout the entire protein [17].

Considering the association between mammalian cytosine DNMTs and SAM, the average conformational entropy ($T\Delta S$) increase upon this process, measured for backbone as well as side chains, varied from 0.02 kJ/mol per residue (DNMT2) to 0.12 kJ/mol per residue (DNMT1). Clearly, these values are very likely to be overestimated, as they result from adding all relevant bond vectors, and the assumption of non-correlated motion is unlikely to hold for all residues. It is evident, however, that the change in pico-to-nanosecond time scale dynamics associated with the SAM binding provides a favourable entropic contribution to the binding free energy. This may indirectly contribute to the biological function (cytosine methylation) by reducing the thermodynamic cost of the cofactor binding, and hence, making the catalytic event more likely to occur. It is also possible that local changes in a global structure/dynamics landscape of the protein, which result from the cofactor binding, facilitate the substrate binding. Thus, SAM seems to play a dual role—acting not merely as a donor of the methyl group for the target cytosine methylation but also facilitating the molecular association between DNMT and substrate through the ‘dynamic allostery’.

In summary, our results suggest that the fast-scale protein dynamics are likely to contribute to the differences in substrate recognition and binding specificity, which are observed throughout the entire DNMT family.

5 Conclusions

It has been acknowledged for some time that protein dynamics may control certain aspects of protein function, including ligand–protein and protein–protein interactions and catalysis. Evidence for such a functional role, principally from NMR relaxation measurements and MD simulations, has been reported [17]. The results obtained in the course of this study, although qualitative rather than quantitative in nature, indicate that cytosine DNA/RNA methyltransferases exploit internal dynamic events to exert their biological function (here, facilitate the cofactor binding for the catalysis). Our data suggest the existence of dispersed dynamics networks in cytosine DNMTs, which can regulate the thermodynamics of their interactions with cofactor and perhaps also methylation targets. The enthalpy–entropy compensation, which we observed, is supported by ITC data for HhaI.

Considerable research efforts in rational drug discovery have aimed, so far, at establishing structure/affinity relationships in ligand–protein complex. The success of these attempts has been limited, partly because of the approximations used, specifically, the neglect of the role of protein conformational dynamics in determining the thermodynamics of ligand–protein interactions.

The present results are likely to indicate some general phenomenon, which may be useful to understand the mechanisms controlling the cofactor and target binding and in DNA methyltransferases and, possibly, other related ligand–protein systems. This information may ultimately be used for rational design of selective inhibitors of DNA/RNA cytosine methyltransferases.

Acknowledgments This work was supported by BBSRC (grant no. B19388 to AKB) and German Cancer Research Centre (DKFZ) funds. All calculations were performed on HPC clusters, University of Leeds. We are also grateful to Prof. Pavel Hobza and Prof. Steve Homans for helpful suggestions and comments. Special thanks to Prof. Alexander Suhai for his support and a lot of inspiring discussions.

References

1. Song F, Mahmood S, Ghosh S, Liang P, Smiraglia DJ, Nagase H, Held WA (2009) Genomics 93(2):130–139
2. Wilson IM, Davies JJ, Weber M, Brown CJ, Alvarez CE, MacAulay C, Schübeler D, Lam WL (2006) Cell Cycle 5(2):155–158
3. Smith LT, Otterson GA, Plass C (2007) Trends Genet 23(9):449–456
4. Kanai Y (2008) Pathol Int 58(9):544–558
5. Ehrlich M (2003) Clin Immunol 109(1):17–28
6. Klebe G, Boehm HJ (1997) J Recept Sign Transduct Res 17(1–3):459–473
7. Diehl C, Genheden S, Modig K, Ryde U, Akke M (2009) J Biomol NMR 45(1–2):157–169
8. Dastidar SG, Lane DP, Verma CS (2008) J Am Chem Soc 130(41):13514–13515
9. Fischer S, Verma CS (1999) Proc Natl Acad Sci USA 96(17):9613–9615
10. Young L, Post CB (1996) Biochemistry 35(48):15129–15133
11. Reddi AR, Guzman TR, Breece RM, Tierney DL, Gibney BR (2007) J Am Chem Soc 129(42):12815–12827
12. Yin F, Cao R, Goddard A, Zhang Y, Oldfield E (2006) J Am Chem Soc 128(11):3524–3525
13. MacRaild CA, Daranas AH, Bronowska A, Homans SW (2007) J Mol Biol 368(3):822–832
14. Gowher H, Liebert K, Hermann A, Xu G, Jeltsch A (2005) J Biol Chem 280(14):13341–13348
15. Lindorff-Larsen K, Best RB, Depristo MA, Dobson CM, Vendruscolo M (2005) Nature 433(7022):128–132
16. Showalter SA, Brüschweiler R (2007) J Am Chem Soc 129(14):4158–4159
17. Stöckmann H, Bronowska A, Syme NR, Thompson GS, Kalverda AP, Warriner SL, Homans SW (2008) J Am Chem Soc 130(37):12420–12426
18. Wong KB, Daggett V (1998) Biochemistry 37(32):11182–11192
19. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Crowley M, Walker RC, Zhang W, Merz KM et al (2006) AMBER 8. University of California, San Francisco
20. Schaftenaar G, Noordik JH (2000) J Comput Aided Mol Design 14:123–134
21. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA Jr et al (1998) Gaussian 98. Gaussian, Inc, Pittsburgh
22. Cornell WD, Cieplak P, Bayly CI, Kollman PA (1993) J Am Chem Soc 115:9620–9631
23. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) J Am Chem Soc 117:5179–5197
24. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Proteins: structure. Funct Genet 3:712–725
25. Markham GD, Norrby PO, Bock CW (2002) Biochemistry 41(24):7636–7646
26. Pavelites JJ, Gao JL, Bash PA, Mackerell AD (1997) J Comput Chem 18:221–239
27. Walker RC, de Souza MM, Mercer IP, Gould IR, Klug DR (2002) J Phys Chem B 106(44):11658–11665
28. Eswar N, Marti-Renom MA, Webb B, Madhusudhan MS, Eramian D, Shen M, Pieper U, Sali A (2000) Current protocols in bioinformatics. Wiley, New York, pp 5.6.1–5.6.30
29. Fiser A, Sali A (2003) Methods Enzymol 374:461–491
30. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926
31. Amber 8 user manual (2004) http://ambermd.org/doc8/amber8.pdf
32. Yang D, Kay LE (1996) J Mol Biol 263(2):369–382
33. Lipari G, Szabo A (1982) J Am Chem Soc 104(17):4559–4570
34. Best RB, Vendruscolo M (2004) J Am Chem Soc 126(26):8090–8091
35. Meyer T, Ferrer-Costa C, Perez A, Rueda M, Bidon-Chanal A, Luque FJ, Laughton CA, Orozco M (2006) J Chem Theor Comp 2:251–258
36. Callebaut I, Courvalin JC, Mornon JP (1999) FEBS Lett 446(1):189–193
37. Robertson KD, Ait-Si-Ali S, Yokochi T, Wade PA, Jones PL, Wolffe AP (2000) Nat Genet 25(3):338–342
38. Sarraf SA, Stancheva I (2004) Mol Cell 15(4):595–605
39. Shikauchi Y, Saiura A, Kubo T, Niwa Y, Yamamoto J, Murase Y, Yoshikawa H (2009) Mol Cell Biol 29:1944–1958
40. Rountree MR, Bachman KE, Baylin SB (2000) Nat Genet 25(3):269–277
41. Sankpal UT, Rao DN (2002) Crit Rev Biochem Mol Biol 37(3):167–197
42. Wyszynski MW, Gabbara S, Kubareva EA, Romanova EA, Oretskaya TS, Gromova ES, Shabarova ZA, Bhagwat AS (1993) Nucleic Acids Res 21(2):295–301
43. Swaminathan CP, Sankpal UT, Rao DN, Surolia A (2002) J Biol Chem 277(6):4042–4049
44. Siedlecki P, Boy RG, Musch T, Brueckner B, Suhai S, Lyko F, Zielenkiewicz P (2006) J Med Chem 49(2):678–683
45. Lorentzen E, Hensel R, Knura T, Ahmed H, Pohl E (2004) J Mol Biol 341(3):815–828
46. Nardini M, Spanò S, Cericola C, Pesce A, Damonte G, Luini A, Corda D, Bolognesi M (2002) Acta Crystallogr D Biol Crystallogr 58(Pt 6 Pt 2):1068–1070