Journal of Theoretical Biology 310 (2012) 14–20

Contents lists available at SciVerse ScienceDirect

Journal of Theoretical Biology

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

A dynamic multi-compartmental model of DNA methylation with demonstrable predictive value in hematological malignancies

Andrew P. McGovern [a], Benjamin E. Powell [b], Timothy J.T. Chevassut [a,*]

[a] Medical Research Building, Brighton and Sussex Medical School, University of Sussex, Falmer, Brighton BN1 9PS, UK
[b] Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, MA 02142, USA

HIGHLIGHTS

- DNA methylation is perturbed in myelodysplasia and leukemia.
- We model DNA methylation using a dynamic multicompartmental approach.
- Levels of methylation depend on enzymatic activity of Dnmt and Tet family proteins.
- Predicted methylation levels are consistent with published experimental data.
- Increased Tet activity maintains hypomethylation at CpG islands.

ARTICLE INFO

Article history:
Received 14 October 2011
Received in revised form
12 June 2012
Accepted 13 June 2012
Available online 20 June 2012

Keywords:
CpG island
Methylcytosine
Hydroxymethylcytosine
DNA methyltransferase (Dnmt)
Tet protein

ABSTRACT

Recent advances have highlighted the central role of DNA methylation in leukemogenesis and have led to clinical trials of epigenetic therapy, notably hypomethylating agents, in myelodysplasia and acute myeloid leukemia. However, despite these advances, our understanding of the dynamic regulation of the methylome remains poor. We have attempted to address this shortcoming by producing a dynamic, six-compartmental model of DNA methylation levels based on the activity of the Dnmt methyltransferase proteins. In addition, the model incorporates the recently discovered Tet family proteins which enzymatically convert methylcytosine to hydroxymethylcytosine. A set of first order, partial differential equations comprise the model and were solved via numerical integration. The model is able to predict the relative abundances of unmethylated, hemimethylated, fully methylated, and hydroxymethylated CpG dyads in the DNA of cells with fully functional Dnmt and Tet proteins. In addition, the model accurately predicts the experimentally measured changes in these abundances with disruption of Dnmt function. Furthermore, the model reveals the mechanism whereby CpG islands are maintained in a hypomethylated state via local modulation of Dnmt and Tet activities without any requirement for active demethylation. We conclude that this model provides an accurate depiction of the major epigenetic processes involving modification of DNA.

1. Introduction

Methylation of DNA bases is recognized to be one of the major epigenetic mechanisms regulating gene expression during development (Latham et al., 2008) and is also strongly implicated in tumourigenesis (Ehrlich, 2009; Sharma et al., 2010). The key proteins that regulate this process are the *de novo* DNA methyltransferases Dnmt3a and Dnmt3b and the maintenance DNA methyltransferase Dnmt1 (Latham et al., 2008). The fourth member of this protein family, Dnmt3L, has been shown to modulate the activity of Dnmt3a but lacks intrinsic methyltransferase activity of its own. A small proportion of methylated cytosine bases subsequently undergo hydroxylation by the α-ketoglutarate dependent, ten-eleven translocation (Tet) family of enzymes (Szwagierczak et al., 2010) which consists of Tet1, Tet2 and Tet3 (Iyer et al., 2009; Tahiliani et al., 2009).

Recent discoveries have highlighted the importance of DNA methylation to leukemogenesis. Inactivating mutations of Tet2 occur frequently in patients diagnosed with myelodysplastic syndromes (MDS) (Ko et al., 2011) leading to transcriptional deregulation and leukemogenesis (Wu et al., 2011). Moreover, gain of function mutations in the isocitrate dehydrogenase genes, Idh1 and Idh2, result in aberrant production of 2-hydroxyglutarate that inhibits the Tet2 reaction causing hypermethylation of

* Corresponding author.
E-mail address: t.chevassut@bsms.ac.uk (T.J.T. Chevassut).

0022-5193/$ - see front matter © 2012 Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.jtbi.2012.06.018

leukemia cells in patients with acute myeloid leukemia (AML) (Figueroa et al., 2010). Also in AML, inactivating mutations in Dnmt3a occur frequently and are associated with poor survival outcomes (Ley et al., 2010). Therefore, DNA methylation is clearly fundamental to the process of leukemogenesis and this is the basis of epigenetic therapies in patients with MDS and AML.

The density of CpGs throughout the genome is highly variable. Multiple clusters of high CpG density are observed upstream of many genes and have been termed CpG islands (Bird et al., 1985; Gardiner-Garden and Frommer, 1987). These islands appear to have a crucial role in regulating gene transcription and are principally unmethylated (Deaton and Bird, 2011). By contrast, the remainder of the DNA is generally methylated and has a relative paucity of CpG dinucleotides (Lander et al., 2001). This is presumably due to spontaneous deamination of methylcytosine to thymine for which there is no DNA repair mechanism (Fryxell and Zuckerkandl, 2000; Jabbari and Bernardi, 2004). For the purpose of this model we refer to these regions of methylated DNA as ^{m}CpG oceans to distinguish them from the unmethylated CpG islands.

The association between epigenetic dysregulation and malignancy has culminated in recent clinical trials of hypomethylating agents (5-azacytidine and 5-aza-2′-deoxycytidine) in MDS that show improved hematological responses, delayed leukemogenesis and improved overall survival (Fenaux et al., 2009). Despite these advances, however, an overall understanding of the dynamic regulation of the methylome remains elusive. The model we have developed attempts to address this issue and to provide insights into the intrinsic properties of the system comprised by the methylating enzymes. In particular, we address whether this system produces two intrinsically stable states which could account for the differential levels of methylation between the CpG islands and ^{m}CpG oceans.

A previous model of this system using a Markov chain approach, without the incorporation of hydroxymethylation, demonstrated the two metastable states of hypermethylation and hypomethylation (Sontag et al., 2006). However, recent analysis of the impact of the Tet proteins on methylation levels (Williams et al., 2011) and the discovery that hypomethylated regions correspond to the CpG islands suggests that rather than two intrinsically stable states, the relative enzyme activity is likely to be a function of the CpG density. This theory is further supported by Bestor et al. (1992) who suggest that CpG islands are inherently refractory to methylation by the *de novo* methyltransferases. It therefore appears that preferential activity of epigenetic enzymes for certain areas of genomic DNA, specifically, CpG islands in the case of Tet and ^{m}CpG oceans in the case of Dnmt3a and 3b, provides the mechanism by which variable methylation levels are maintained throughout the genome. A successful model of this system should, therefore, demonstrate a single stable state which can be induced into either a hypermethylated or a hypomethylated state through varying the activity of these enzymes. Such a model may be of benefit in understanding the dynamics of the methylome and in particular its importance to differentiation, tumorigenesis, and epigenetic therapy.

2. Model and methods

Our model is an extension of previous DNA methylation models (Genereux et al., 2005; Pfeifer et al., 1990; Sontag et al., 2006) with the additional incorporation of hydroxymethylation by the Tet protein family. Six principal dyad groups are considered; unmethylated CpG dyads (CpG/GpC), hemimethylated dyads (^{m}CpG/GpC), fully methylated dyads (^{m}CpG/GpC^{m}), dyads

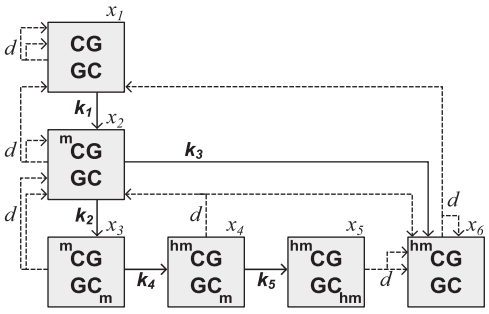

Fig. 1. A schematic of the six compartment model. Solid arrows indicate the flow between model compartments generated by the actions of the Dnmt or Tet enzyme families. Flow due to DNA replication is indicated by dashed lines, with each compartment contributing to two daughter compartments.

with a single methyl group and an hydroxymethyl group on the opposing strand (^{hm}CpG/GpC^{m}), fully hydroxymethylated dyads (^{hm}CpG/GpC^{hm}) and hemihydroxymethylated dyads (^{hm}CpG/GpC). The transitions between these groups are demonstrated in Fig. 1. In addition to transitions due to the methylating and hydroxymethylating actions of the Dnmt and Tet enzymes, transitions due to DNA replication at cell division are also represented. During DNA replication only unmethylated cytosines are incorporated into the newly formed daughter strand opposite to CpGs on the parent DNA strand. Therefore, during cell division each fully methylated CpG dyad produces two hemimethylated dyads, each hemimethylated dyad produces one unmethylated dyad and one hemimethylated dyad, and so forth.

The total number of dyads in each state are given by x_1, x_2, x_3, x_4, x_5 and x_6 for CpG/GpC, ^{m}CpG/GpC, ^{m}CpG/GpC^{m}, ^{hm}CpG/GpC^{m}, ^{hm}CpG/GpC^{hm}, and ^{hm}CpG/GpC respectively. The rate of *de novo* methylation in a population of non-differentiating cells is assumed to be time-invariant and proportional to the total quantity of unmethylated CpGs (x_1). The constant of proportionality is termed k_1 giving the rate of *de novo* methylation as k_1x_1. The value of k_1 is principally determined by the cumulative rate of action of the *de novo* methyltransferases (Dnmt3a, Dnmt3b and Dnmt3L). Similarly the rate of maintenance methylation is given by k_2x_2, where k_2 is the proportion of hemimethylated CpGs which are fully methylated per unit time. k_2 is determined by the copying fidelity of the methylcytosines to the daughter DNA strand, prior to replication, and this is principally mediated by the maintenance methyltransferase, Dnmt1. Similarly the rate of hydroxymethylation is given by k_3x_2 in the case of hemimethylated dyads, k_4x_3 for fully methylated dyads and k_5x_4 for the ^{hm}CpG/GpC^{m} dyads. Having defined these proportions, the rate of change of abundance of the six CpG dyads can be described by a set of first order, inhomogeneous, partial differential equations

$$\frac{\partial x_1}{\partial t} = (d-l-k_1)x_1 + dx_2 + dx_6 \tag{1}$$

$$\frac{\partial x_2}{\partial t} = k_1x_1 - (l+k_2+k_3)x_2 + 2dx_3 + dx_4 \tag{2}$$

$$\frac{\partial x_3}{\partial t} = k_2x_2 - (d+l+k_4)x_3 \tag{3}$$

$$\frac{\partial x_4}{\partial t} = k_4x_3 - (d+l+k_5)x_4 \tag{4}$$

$$\frac{\partial x_5}{\partial t} = k_5x_4 - (d+l)x_5 \tag{5}$$

$$\frac{\partial x_6}{\partial t} = k_3x_2 + dx_4 + 2dx_5 - lx_6 \tag{6}$$

where d is the cell division rate and l is the cell loss rate (due to cell death). During exponential growth, a value $l=0$ can be assumed. These partial derivatives were solved using fourth order Runge–Kutta numerical integration.

2.1. Application to the global methylome

As these model compartments are likely to be genome invariant the model can be applied equally to the global methylome as well as to localized patterns of DNA methylation. The global fidelity of maintenance methylation by Dnmt1 has been variously estimated to be between 97.7% and 99.9% (Jackson et al., 2004; Pfeifer et al., 1990). Selecting an approximate mean value within this estimated range of 99%, k_2 can be calculated in terms of the cell division rate d from the relative probabilities of progression from the hemimethylated state, i.e. 99% become fully methylated and 1% remain hemimethylated until subsequent cell division. This requires the maintenance methylation constant to be greater than the cell division rate constant by a factor of 99 ($k_2=99d$). The *de novo* methylation fraction occurring primarily through the action of Dnmt3a and Dnmt3b has been previously estimated at around 5–9% (Genereux et al., 2005; Pfeifer et al., 1990). However a hairpin-bisulfite analysis by Laird et al. (2004) demonstrated a highly variable *de novo* methylation efficiency of 17% at hypermethylated sites and 1% at hypomethylated sites. In order to model the dominant effect of hypermethylated ^{m}CpG oceans on the global methylome we have elected to use this higher value of 17%. This requires modeling a ratio between the cell division rate constant and the *de novo* methylation rate constant of 83:17. Therefore a value of $k_1=17d/83=0.205d$ was selected.

Values for k_3, k_4, and k_5 are more difficult to reliably determine. There is currently no evidence to suggest that Tet has preferential activity for any particular methylated dyad. Therefore we assume $k_3=k_4/2=k_5$. No previous calculation for the rate of action of Tet has been published, however it is possible to fit the value of k_3 to measured levels of ^{hm}C. Tahiliani et al. (2009) measured the genomic content of ^{hm}C and ^{m}C in embryonic stem (ES) cells, as 4–6% and 55–60% of MspI cleavage site (C\widehatCGG) respectively. Assuming that these values are representative of CpG sites as a whole this corresponds to a frequency of ^{hm}C of 0.04–0.06% relative to all bases. However due to the structure of its cleavage site Msp1 is more likely to cleave at CpG rich regions and hence produce an over estimate for the global levels of ^{m}C and ^{hm}C density. We measured a value of 0.03% for ^{hm}C relative to all bases for ES cells, using mass spectrometry analysis, which is not subject to this bias (Ramsahoye, personal communication). Assuming all of this measured ^{hm}C exists as part of CpG dyads this corresponds to a relative abundance of ^{hm}CpG of 3% compared to all CpG dyads. This may still represent an overestimate if there is hydroxymethylation of cytosines at non-CpG sites. Although other investigators have found a higher frequency of 0.3% relative to all cytosines (0.063% relative to all bases) (Szwagierczak et al., 2010), subsequent studies have also produced results consistent with our lower value (Song et al., 2011; Stroud et al., 2011). Based on this data, and in the current absence of evidence for non-CpG hydroxymethylation, we have chosen to model a value for the relative global abundance of ^{hm}CpG dinucleotides of 3% relative to all CpGs. A value of $k_3=0.04d$ was selected to fit this abundance of ^{hm}C, in the steady state solution, when used in combination with the values already selected for k_1 and k_2. This corresponds to a global hydroxymethylation efficiency of approximately 3.8% before cell division.

2.2. Application to local epigenetic regions

CpG islands demonstrate highly contrasting epigenetic properties to the ^{m}CpG oceans in that they are relatively hypomethylated (Lister et al., 2009; Meissner et al., 2008) and also contain an increased abundance of hydroxymethylated cytosines (Williams et al., 2011). The enzyme dynamics that maintain this hypomethylated state have not been fully elucidated, however it has been demonstrated that these regions are resistant to *de novo* methylation (Bestor et al., 1992). In addition the Tet protein family have an increased affinity for sites with a high CpG content (Williams et al., 2011). This preferential activity is likely to be mediated by their CpG binding CXXC domains (Iyer et al., 2009; Pastor et al., 2011; Tahiliani et al., 2009). Utilizing these discoveries, we have applied our model to simulate the local effects dynamic changes of methylation within CpG islands. This requires us to estimate new values for both the *de novo* methylation constant k_1 and the hydroxylation constant k_3 to reflect the decreased Dnmt3a and 3b activity and increased Tet activity within these regions.

Synergistic activity between Dnmt3a and 3b and Dnmt1 has been demonstrated and increases the rate of *de novo* methylation up to 15-fold (Chen et al., 2003; Kim et al., 2002; Liang et al., 2002). Subsequently the resistance of CpG islands to the *de novo* methyltransferases is likely to result in disruption of this synergism and reduction of *de novo* methylation rates by a similar magnitude. Therefore we have selected a value of 1.1% for the *de novo* methylation efficiency (a 15-fold decrease) to reflect this reduced activity. This equates to a revised value for the *de novo* methylation constant of $k_1=0.012d$. This is also consistent with the analysis by Laird et al. (2004) who measured a 17-fold decrease in *de novo* methylation efficiency in hypomethylated regions to 1%.

Williams et al. (2011) show that there is an approximate 2.5-fold increase in the density of a Tet1 binding tag at transcriptional start sites and this corresponds to measurements by Pastor et al. (2011) who show approximately 2.5-fold increase in the average ^{hm}C content also at transcriptional start sites. Whilst these measurements demonstrate strong positive evidence for an increased affinity of Tet for CpG islands, quantification of this increased affinity can only be a speculative extrapolation at present. However we have chosen a value for $k_3=0.11d$ (a 2.5-fold increase to 9.6% hydroxymethylation efficiency) to reflect this increased affinity.

3. Results and discussion

A phase space analysis was conducted for the entire phase space range to identify stable attractors of the system. This region is comprised of a 6-simplex, with a single orthogonal vertex, where the relative abundance of each dyad varied from 0% to 100% along each edge running from the orthogonal vertex. For both the global and CpG island models this analysis demonstrated a single, fixed point attractor with a basin of attraction encompassing all possible initial conditions. Projections of this analysis are demonstrated in Fig. 2 in the plane displaying varying abundance of unmethylated and fully methylated dyads. This analysis leads us to the conclusion that there is only one stable solution for this system which is rapidly converged upon, irrespective of the initial conditions. The relative abundance of each dyad group at this unique solution is dependent on the value assigned to the methylation and hydroxymethylation rate constants; k_1, k_2, and k_3.

The constant values selected for the global methylome model produced a stable state in which the overall methylcytosine level within CpG dinucleotides was 73.1% and hydroxymethylcytosine level was 2.9%. 98% of the methylated CpG dinucleotides were found to be part of a fully methylated dyad, which is in close agreement with the genome wide value of 99% measured by Lister et al. (2009). We found that if heterozygous loss of Dnmt3a

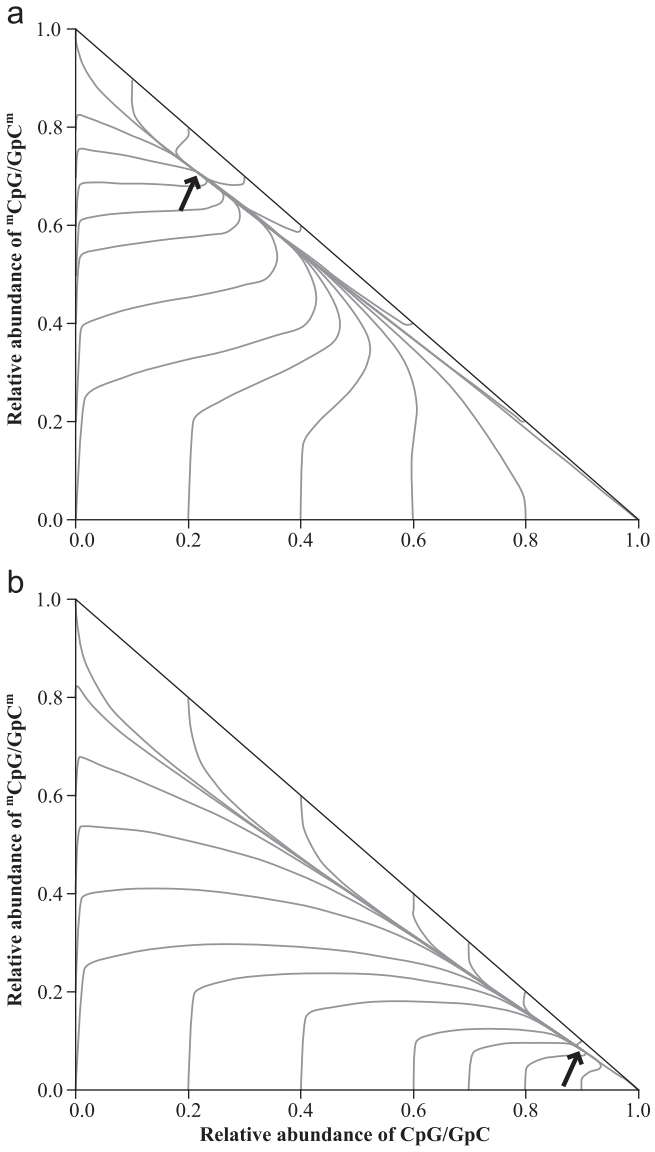

Fig. 2. Phase portrait projections of the six-simplex phase space illustrating methylation changes in a cell population with fully functional epigenetic enzymes. Multiple trajectories for varying initial values of relative abundance of CpG/GpC and ^{m}CpG/GpC^{m} are demonstrated to fall into a single fixed point attractor (arrows) for both (a) the global model and (b) the CpG island model application.

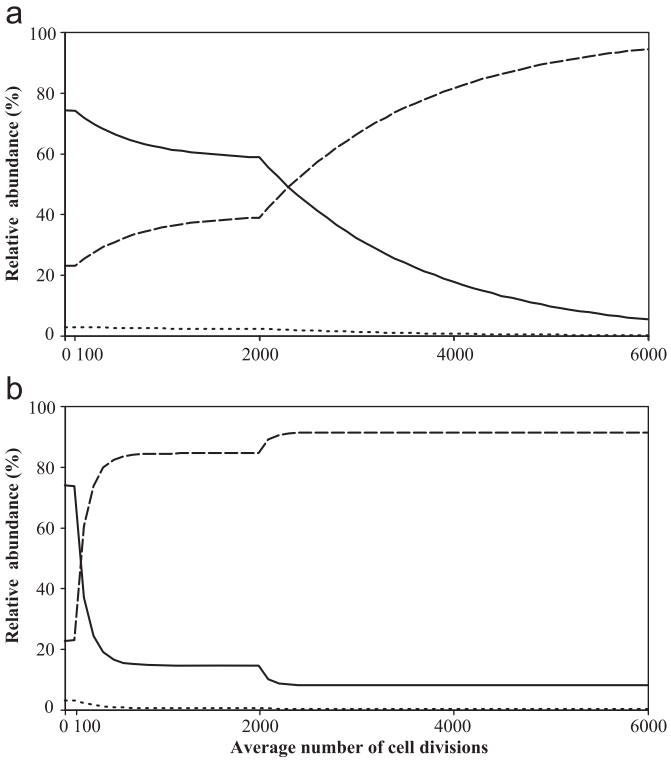

Fig. 3. The effects of homozygous and heterozygous loss of the Dnmt enzymes on the genome wide abundance of unmethylated CpG (dashed), methylated CpG (solid) and hydroxymethylated CpG (dotted). (a) Initially wild type dyad abundances are perturbed by heterozygous loss of both Dnmt3a and Dnmt3b after 100 cell divisions and homozygous loss of the same after 2000 cell divisions. (b) Initially wild type dyad abundances are perturbed by heterozygous loss of Dnmt1 after 100 cell divisions and homozygous loss after 2000 cell divisions.

or Dnmt3b is modeled (by a 25% reduction in the *de novo* methylation efficiency) the effect is a minimal decrease of global methylation levels to 67.2%. The combined effect of heterozygous loss of *both* Dnmt3a and Dnmt3b is expected to reduce the *de novo* methylation efficiency by around 50%. This causes a more significant reduction in global methylation levels to 58.0% as illustrated in Fig. 3a. Ley et al. (2010) performed an analysis of methylation levels on genomic DNA from AML patients with heterozygous inactivating mutations of Dnmt3a. In these cells they demonstrated hypomethylation at a small number of discrete locations but were unable to demonstrate a statistically significant reduction in global methylation levels when compared with wild type samples. This is fully consistent with our analysis showing a minimal change in global methylation levels due to Dnmt3a heterozygousity. Contrastingly, a double knockout of Dnmt3a and Dnmt3b is likely to produce a value of $k_1 \approx 0$ for *de novo* methylation, with a large accompanying decline in global methylation (Fig. 3a). ES cells containing targeted homozygous knockouts of both the Dnmt3a and Dnmt3b genes have been produced (Okano et al., 1999) and these have been shown to progressively lose methylation over serial passaging until they are profoundly hypomethylated cells containing less than 1% methylated CpGs (Jackson et al., 2004). Chen et al. (2003) also demonstrate progressive loss of methylation in double knockout of Dnmt3a and Dnmt3b ES cells with almost no residual methylation after five months of serial passaging. This protracted loss of methylation is also consistent with our analysis. Furthermore Chen et al. (2003) demonstrate, after restoration of Dnmt3a/3b function in these ES cells, the return to normal methylation levels after a further four weeks of passaging. Modeling the restoration of Dnmt3a/3b function in double knockouts produced a return to normal methylation levels after 500–1000 cell divisions, consistent with this analysis (data not shown).

With regard to the maintenance methyltransferase Dnmt1, our model shows that a 50% loss in enzyme activity ($k_2 = 1d$) produces a rapid reduction in global methylation levels to 14.7%. Complete loss of enzyme activity as caused by homozygous knockout of Dnmt1 genes (reduction of the maintenance methylation constant to $k_2 \approx 0$) results in reduction of the global methylation level to 8.2% (Fig. 3b). This low level methylation is comprised entirely of hemimethylated dyads due to the *de novo* methyltransferase activity of Dnmt3a and Dnmt3b. Some residual maintenance methylation is to be expected however as Dnmt3a and Dnmt3b have the same propensity for methylating hemimethylated dyads as unmethylated dyads (Latham et al., 2008). Subsequently, even with a complete knockout of Dnmt1, a residual non-zero value of k_2 for maintenance methylation is to be expected. Assuming Dnmt3a/3b has half the propensity for methylating hemimethylated dyad as unmethylated dyads a residual efficiency for maintenance methylation of 8.5% should

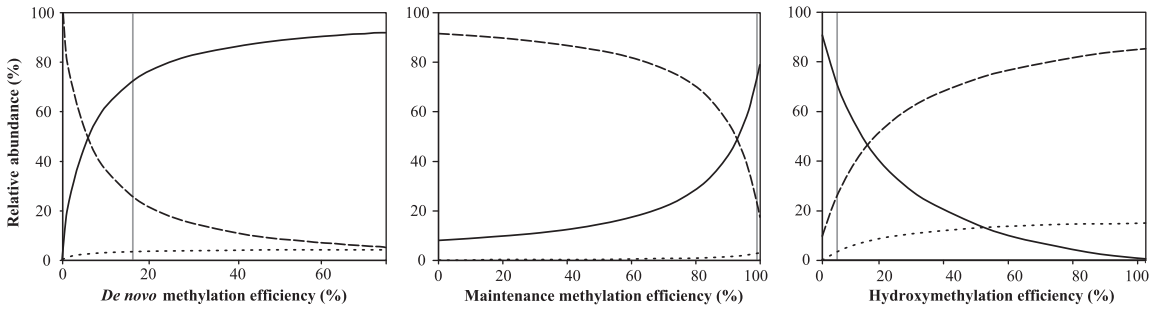

Fig. 4. Projected effects of perturbation Dnmt3, Dnmt1 and Tet activity on the genome wide abundance of unmethylated CpG (dashed), methylated CpG (solid) and hydroxymethylated CpG (dotted). The estimated values for wild type ES cells are demonstrated by the vertical gray line.

be used to represent this residual activity. This value is selected to be half the value of the *de novo* methylation efficiency as there are half the number of unmethylated CpGs in hemimethylated dyads. This results in a steady state methylation level of 8.8% as expected for a complete Dnmt1 knockout. Such experiments have been performed in ES cells by targeted knockout of the gene (Jackson et al., 2004; Ramsahoye et al., 2000) and these have revealed a residual global methylation level of approximately 20%. It is possible that this higher than expected level is due to upregulation of *de novo* methyltransferase activity in Dnmt1 knockout cells. Indeed, a two-fold increase in *de novo* methylation was found to produce this residual level of methylation. With regard to heterozygous loss of Dnmt1 we are unable to know whether this leads to a 50% reduction in maintenance methylation efficiency. It seems likely that there is excess Dnmt1 in order to facilitate high maintenance methylation efficiency and indeed halving the maintenance methylation constant ($k_2 = 49d$) only reduces the methylation efficiency to 98%. This was found to cause only a 5% reduction in global methylation levels. Howell et al. (2001) found that heterozygous mutations of Dnmt1 in murine ES cells displayed near-normal methylation patterns, but that their progeny were hypomethylated at a few discrete loci, consistent with this analysis. Mice generated to express Dnmt1 at around 10% of normal levels, exhibit marked hypomethylation (Gaudet et al., 2003), also consistent with our model. The effects of varying enzyme activity are illustrated in Fig. 4.

The small, but non-zero, value of k_3, reflecting the aggregate activity of the Tet protein family, has a significant impact on the relative abundance of methylation. Reduction of the value of the hydroxymethylation constant k_3 below $0.04d$ results in an approximately linear increase in overall methylation level, primarily through an increase in the proportion of fully methylated dyads. Mutations of Idh1/2 are likely to have a significant impact on the value of k_3 via the generation of hydroxygluterate which antagonizes Tet activity. Our model predicts that this will generate a significant increase in global methylation levels, an effect demonstrated by Figueroa et al. (2010) and Turcan et al. (2012). Interestingly, Ko et al. (2011) have shown that Tet2 mutated AML samples frequently contain aberrant hypomethylation rather than hypermethylation at differentially methylated CpG sites. We believe that this might be the consequence of clonal expansion of hypomethylated leukemic cells with selective proliferative growth advantage. Alternatively, loss of Tet2 function in myeloid cells may reactivate a stem cell-like state characterized by generalized hypomethylation and consequent genomic instability (Ehrlich, 2009; Lengauer, 2003). Of relevance in this regard is the fact that hypomorphic Dnmt1 mutations, that result in hypomethylation, lead to skewed hematopoietic differentiation towards myeloerythroid lineages (Broske et al., 2009).

The CpG island application of the model predicts a methylcytosine level within CpG dinucleotides of 16.2% and a hydroxymethylcytosine level of 2.4%. This is a 63.5% reduction in the abundance of methylcytosine when compared to the ^{m}CpG oceans, broadly predicted by the global model. This demonstrates that the combination of reduced *de novo* methylation activity and increased Tet activity at these sites effectively induces hypomethylation. As discussed, the selected 2.5-fold increase in the selected value of the hydroxmethylation constant k_3 is an approximate extrapolation from the existing data. Fig. 4 demonstrates that increasing Tet activity leads to a significant increase in the proportion of unmethylated dinucleotides. In other words, hydroxylation of methylated cytosines by Tet family proteins provides a highly efficient way of achieving genomic demethylation. This implies that no active method of demethylation is required. However, the recent discovery that Tet enzymes can further oxidize 5hmC into the higher oxidation substituents 5-formylcytosine and 5-carboxylcytosine provides a possible mechanism whereby demethylation of DNA can occur, possibly involving base excision repair by thymine–DNA glycosylase (He et al., 2011; Ito et al., 2011). Clearly this could be incorporated into a future revision of our model, however we suggest that this is unlikely to have any significant impact on the methylation dynamics as the relative abundance of these higher oxidation substituents is small (He et al., 2011; Ito et al., 2011).

Inevitably the model presented here has several limitations. A number of factors not incorporated into the model indirectly influence DNA methylation levels, including histone modification and chromatin accessibility. These factors limit the ability of the Dnmt methyltransferase enzymes to act on the DNA bases. The model here utilizes globally measured levels of maintenance methylation fidelity and *de novo* methylation activity levels for ES cells, which therefore indirectly accounts for these factors. Any significant change in these factors extrinsic to the model, however, requires revised values for these methylation activity levels to be input into the model. Furthermore, the activity of Dnmt3a/3b/3L, Dnmt1, and Tet1/2/3 vary between cell types. The model here uses enzyme activity levels measured for ES cells but could be extended to other cell types, where these levels are also known or can be estimated, as the model compartments are invariant between cell types.

4. Conclusions

Our model demonstrates that levels of methylation and hydroxymethylation within genomic DNA are determined by the relative activity of the Dnmt and Tet family proteins. This results in a unique stable solution rather than multiple solutions despite regional variation of methylation and hydroxymethylation throughout the genome. These variable levels are maintained by preferential affinity of these proteins for different genomic locations, in particular reduced activity of Dnmt3a/3b/3L and increased activity of the Tet proteins at CpG rich regions. This finding is in contrast to previous modeling attempts

where a metastable hypomethylated state was found to coexist with a hypermethylated state (Sontag et al., 2006). The model is also able to accurately predict the effects of inactivating heterozygous and homozygous mutations in the Dnmt3a, Dnmt3b, and Dnmt1 genes in ES cells. Furthermore, our model demonstrates that the combination of decreased Dnmt3a/3b/3L activity coupled to increased Tet activity at CpG rich regions provides a highly efficient means of causing local demethylation. This implies that no active means of demethylation is required by the cell to induce hypomethylation. We regard this as a major insight to emerge from our model.

We believe that our model has important relevance for understanding the biology of DNA methylation during processes of differentiation and oncogenesis. Murine ES cells lose approximately 50% of hydroxymethylcytosine upon withdrawal of leukemia inhibitory factor which leads to down-regulation of Tet family proteins, notably Tet1 (Tahiliani et al., 2009), although the importance of this with regard to differentiation is currently unknown. In the case of the pre-leukemic state found in MDS, inactivating mutations of Tet2 are the most commonly found recurrent mutation occurring in over 20% of cases (Bejar et al., 2011). In AML, inactivating mutations of Dnmt3a occur in 22% of *de novo* cases with enrichment amongst patients with intermediate-risk cytogenetic profiles where they are associated with an adverse outcome (Ley et al., 2010). Thus, Tet2 and Dnmt3a are involved in disease progression, via loss of function mutations, highlighting the importance of DNA methylation to leukemogenesis. What is intriguing is the fact that in the majority of cases these inactivating mutations of Tet2 and Dnmt3a cause functional haploinsufficiency suggesting that gene dosage and therefore enzyme activity is critically important. Our model provides a tool for understanding how differential activities of these enzymes affect DNA methylation. Not only is this important for understanding leukemogenesis but also for predicting the consequence of epigenetic therapies such as the hypomethylating agents 5-azacytidine and 5-aza-2′-deoxycytidine that are the first agents to demonstrate improved survival outcomes in MDS/AML.

Acknowledgments

We would like to thank Drs. Helen Stewart, John Brewin and Gillian Horne (University of Sussex) and Dr. Bernard Ramsahoye (University of Edinburgh) for critical review of the manuscript and the Elimination of Leukaemia Fund who supported this work.

References

Bejar, R., Stevenson, K., Abdel-Wahab, O., Galili, N., Nilsson, B., Garcia-Manero, G., Kantarjian, H., Raza, A., et al., 2011. Clinical effect of point mutations in myelodysplastic syndromes. N. Engl. J. Med. 364, 2496–2506.

Bestor, T.H., Gundersen, G., Kolsto, A., Prydz, H., 1992. CpG islands in mammalian gene promoters are inherently resistant to de novo methylation. Genet. Anal. Tech. Appl. 9, 48–53.

Bird, A., Taggart, M., Frommer, M., Miller, O.J., Macleod, D., 1985. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell 40, 91–99.

Broske, A.M., Vockentanz, L., Kharazi, S., Huska, M.R., Mancini, E., Scheller, M., Kuhl, C., Enns, A., et al., 2009. DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. Nat. Genet. 41, 1207–1215.

Chen, T., Ueda, Y., Dodge, J.E., Wang, Z., Li, E., 2003. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol. Cell. Biol. 23, 5594–5605.

Deaton, A.M., Bird, A., 2011. CpG islands and the regulation of transcription. Genes Dev. 25, 1010–1022.

Ehrlich, M., 2009. DNA hypomethylation in cancer cells. Epigenomics 1, 239–259.

Fenaux, P., Mufti, G.J., Hellstrom-Lindberg, E., Santini, V., Finelli, C., Giagounidis, A., Schoch, R., Gattermann, N., et al., 2009. Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. Lancet Oncol. 10, 223–232.

Figueroa, M.E., Abdel-Wahab, O., Lu, C., Ward, P.S., Patel, J., Shih, A., Li, Y., Bhagwat, N., et al., 2010. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer Cell. 18, 553–567.

Fryxell, K.J., Zuckerkandl, E., 2000. Cytosine deamination plays a primary role in the evolution of mammalian isochores. Mol. Biol. Evol. 17, 1371–1383.

Gardiner-Garden, M., Frommer, M., 1987. CpG islands in vertebrate genomes. J. Mol. Biol. 196, 261–282.

Gaudet, F., Hodgson, J.G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J.W., Leonhardt, H., Jaenisch, R., 2003. Induction of tumors in mice by genomic hypomethylation. Science 300, 489–492.

Genereux, D.P., Miner, B.E., Bergstrom, C.T., Laird, C.D., 2005. A population-epigenetic model to infer site-specific methylation rates from double-stranded DNA methylation patterns. Proc. Natl. Acad. Sci. USA 102, 5802–5807.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., et al., 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science 6047, 1303–1307.

Howell, C.Y., Bestor, T.H., Ding, F., Latham, K.E., Mertineit, C., Trasler, J.M., Chaillet, J.R., 2001. Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell 104, 829–838.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., Zhang, Y., 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science 6047, 1300–1303.

Iyer, L.M., Tahiliani, M., Rao, A., Aravind, L., 2009. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. Cell. Cycle 8, 1698–1710.

Jabbari, K., Bernardi, G., 2004. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. Gene 333, 143–149.

Jackson, M., Krassowska, A., Gilbert, N., Chevassut, T., Forrester, L., Ansell, J., Ramsahoye, B., 2004. Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. Mol. Cell. Biol. 24, 8862–8871.

Kim, G.D., Ni, J., Kelesoglu, N., Roberts, R.J., Pradhan, S., 2002. Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. EMBO J. 21, 4183–4195.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., et al., 2011. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature 468, 839–843.

Laird, C.D., Pleasant, N.D., Clark, A.D., Sneeden, J.L., Hassan, K.M., Manley, N.C., Vary Jr., J.C., Morgan, T., et al., 2004. Hairpin-bisulfite PCR: assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. Proc. Natl. Acad. Sci. USA 101, 204–209.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., et al., 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Latham, T., Gilbert, N., Ramsahoye, B., 2008. DNA methylation in mouse embryonic stem cells and development. Cell. Tissue Res. 331, 31–55.

Lengauer, C., 2003. Cancer. An unstable liason. Science 300, 442.

Ley, T.J., Ding, L., Walter, M.J., McLellan, M.D., Lamprecht, T., Larson, D.E., Kandoth, C., Payton, J.E., et al., 2010. DNMT3A mutations in acute myeloid leukemia. N. Engl. J. Med. 363, 2424–2433.

Liang, G., Chan, M.F., Tomigahara, Y., Tsai, Y.C., Gonzales, F.A., Li, E., Laird, P.W., Jones, P.A., 2002. Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. Mol. Cell. Biol. 22, 480–491.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., et al., 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462, 315–322.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., et al., 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454, 766–770.

Okano, M., Bell, D.W., Haber, D.A., Li, E., 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99, 247–257.

Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., et al., 2011. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature 473, 394–397.

Pfeifer, G.P., Steigerwald, S.D., Hansen, R.S., Gartler, S.M., Riggs, A.D., 1990. Polymerase chain reaction-aided genomic sequencing of an X chromosome-linked CpG island: methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability. Proc. Natl. Acad. Sci. USA 87, 8252–8256.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P., Jaenisch, R., 2000. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc. Natl. Acad. Sci. USA 97, 5237–5242.

Sharma, S., Kelly, T.K., Jones, P.A., 2010. Epigenetics in cancer. Carcinogenesis 31, 27–36.

Song, C.X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.H., et al., 2011. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat. Biotechnol. 29, 68–72.

Sontag, L.B., Lorincz, M.C., Georg Luebeck, E., 2006. Dynamics, stability and inheritance of somatic DNA methylation imprints. J. Theor. Biol. 242, 890–899.

Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S., Jacobsen, S.E., 2011. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome Biol. 12, R54.

Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F., Leonhardt, H., 2010. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res. 38, e181.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., et al., 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324, 930–935.

Turcan, S., Rohle, D., Goenka, A., Walsh, L.A., Fang, F., Yilmaz, E., Campos, C., Fabius, A.W., et al., 2012. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. Nature 483, 479–483.

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A., Rappsilber, J., Helin, K., 2011. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature 473, 343–348.

Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., 2011. Eve Sun, 2011. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature 473, 389–393.