Transcriptional regulation by histone modifications: towards a theory of chromatin re-organization during stem cell differentiation

2013 Phys. Biol. 10 026006

(http://iopscience.iop.org/1478-3975/10/2/026006)

Phys. Biol. **10** (2013) 026006 (16pp) doi:10.1088/1478-3975/10/2/026006

Transcriptional regulation by histone modifications: towards a theory of chromatin re-organization during stem cell differentiation

Hans Binder[1,2], **Lydia Steiner**[1,3], **Jens Przybilla**[1], **Thimo Rohlf**[1,4], **Sonja Prohaska**[1,3] **and Jörg Galle**[1,5]

[1] Interdisciplinary Centre for Bioinformatics, University of Leipzig, D-04107 Leipzig, Härtelstr. 16-18, Germany
[2] Leipzig Interdisciplinary Research Cluster of Genetic Factors, Clinical Phenotypes and Environment, University of Leipzig, D-04103 Leipzig, Philipp-Rosenthalstr. 27, Germany
[3] Computational EvoDevo Group, Institute of Computer Science, University of Leipzig, D-04107 Leipzig, Härtelstr. 16-18, Germany
[4] Max-Planck-Institute for Mathematics in the Sciences, Inselstr. 22, D-04103 Leipzig, Germany

E-mail: lydia@bioinf.uni-leipzig.de, przybilla@izbi.uni-leipzig.de, galle@izbi.uni-leipzig.de, binder@izbi.uni-leipzig.de, sonja@bioinf.uni-leipzig.de and rohlf@izbi.uni-leipzig.de

Received 6 August 2012
Accepted for publication 29 January 2013
Published 12 March 2013
Online at stacks.iop.org/PhysBio/10/026006

Abstract
Chromatin-related mechanisms, as e.g. histone modifications, are known to be involved in regulatory switches within the transcriptome. Only recently, mathematical models of these mechanisms have been established. So far they have not been applied to genome-wide data. We here introduce a mathematical model of transcriptional regulation by histone modifications and apply it to data of trimethylation of histone 3 at lysine 4 (H3K4me3) and 27 (H3K27me3) in mouse pluripotent and lineage-committed cells. The model describes binding of protein complexes to chromatin which are capable of reading and writing histone marks. Molecular interactions of the complexes with DNA and modified histones create a regulatory switch of transcriptional activity. The regulatory states of the switch depend on the activity of histone (de-) methylases, the strength of complex-DNA-binding and the number of nucleosomes capable of cooperatively contributing to complex-binding. Our model explains experimentally measured length distributions of modified chromatin regions. It suggests (i) that high CpG-density facilitates recruitment of the modifying complexes in embryonic stem cells and (ii) that re-organization of extended chromatin regions during lineage specification into neuronal progenitor cells requires targeted de-modification. Our approach represents a basic step towards multi-scale models of transcriptional control during development and lineage specification.

[5] Author to whom any correspondence should be addressed.

Introduction

Regulation of transcript abundance derived from a genomic locus can be governed by diverse mechanisms to ensure that functional regulatory states can be maintained and adapted to variable environments. Related molecular systems often involve binding of transcription factors forming cis-regulatory networks and are modulated by chromatin reorganization. Unravelling the interplay between different regulatory mechanisms, their hierarchy and, particularly, their changes in the course of the life cycle represent central challenges to Molecular Systems Biology (Mohammad and Baylin 2010). For instance, during gametogenesis and early embryonic development modules of transcription factors networks regulate differentiation (Reik 2007, Kota and Feil 2010) whereas in subsequent stages new chromatin marks accumulate, and take control of the cell fate (Dahl *et al* 2010, Prohaska *et al* 2010).

Mathematical models of transcription factor networks controlling development and differentiation are a matter of intense debate (reviewed in Karlebach and Shamir (2008)), models of chromatin reorganization during these processes, however, are rather rare (reviewed in Rohlf *et al* (2012)). A number of different hypotheses have been put forward on how chromatin modification states are established and maintained, and how combinatorial patterns of these modification states may contribute to transcriptional regulation and cellular memory. In contrast, very few approaches have been developed so far that rigorously formalize basic dynamical properties of chromatin modification. The approaches proposed by Dodd *et al* (2007) and Sedighi and Sengupta (2007) for the first time explained how stable cell fate decisions can emerge from cooperative action of chromatin modifiers. These approaches focused on the dynamics of histone modifications in yeast. Here, we introduce a model of transcriptional regulation by histone modification that is motivated by experimental findings in higher eukaryotes.

Particularly, our study is motivated by the observed functions of polycomb group (PcG) and trithorax group (trxG) proteins, both forming large protein complexes. These complexes act as epigenetic readers and writers, i.e. their binding to chromatin is governed by specific histone modifications and enables modification of histones by attaching chemical groups (Ringrose and Paro 2004). Binding of these complexes contributes to transcriptional control during development and lineage specification (Schuettengruber *et al* 2007, Orlando 2003).

In the absence of transcriptional induction, trxG and PcG complexes bind and remain attached to chromatin target loci and modulate the chromatin structure. Specifically, the complexes can induce trimethylation of histone H3 at lysin 4 (H3K4me3) and 27 (H3K27me3) (Schuettengruber *et al* 2007), respectively. These methylation reactions are catalyzed by SET domains of the trxG and PcG proteins. The so established modification marks both can be reversed by specific de-methlyases (Pedersen and Helin 2010, Klose and Zhang 2007). The trimethylation marks at K4 and K27 were classified as activating and repressive marks, respectively, according to their overall effect on transcriptional regulation. Occurrence of both marks at a single histone is thought to constitute an intermediate (bivalent or poised) state that subsequently can lead to either activation or repression. It is considered as a hallmark for embryonic stem cells (ESCs) (Bernstein *et al* 2007).

Experimental studies reveal a differential distribution of H3K4me3 chromatin marks at CpG-rich versus CpG-poor promoters (Meissner *et al* 2008). In ESCs nearly all genes with CpG-rich promoters are bound by trxG complexes (Mikkelsen *et al* 2007). The bound complexes protect the CpG-motifs from becoming methylated and thus the associated genes from becoming silenced (Ooi *et al* 2007). Consequently, most of these genes are actively transcribed (Mohn and Schubeler 2009). Silencing of genes under the influence of trxG complexes can be achieved by association with PcG complexes which is the case for about 22% of the CpG-rich promoters of mouse ESCs (Mikkelsen *et al* 2007). Contiguous genomic regions associated with H3K4me3 and H3K27me3 modified histones have been found to reach lengths that exceed 10 000 base pairs (bp). Thus, they can extend over a length associated with more than 50 nucleosomes (Pan *et al* 2007). This large range of concerted modification may be explained by cooperative recruitment of the modifying protein complexes.

The mechanism of recruitment of trxG and PcG complexes is largely unknown. Binding them to DNA sequence motifs has been favoured for a long time. However, only little evidence has been acquired so far (Mendenhall *et al* 2010). Binding profiles of PcG complexes in Drosophila suggest that they bind DNA sequence motifs clustered in the so-called polycomb response elements (PREs (Schwartz *et al* 2006)). Nevertheless, only for the polycomb group protein Pho such a binding motif has been clearly identified (Schuettengruber *et al* 2009, Klymenko *et al* 2006). The apparent absence of such binding sites in mammalian genomes suggests that different developmental strategies entailed different repressive mechanisms in higher eukaryotes (Kohler and Villar 2008). Several studies report a relationship between the CpG-density of the DNA and the recruitment of trxG- and PcG-complexes in mammals. Actually, trxG-complexes contain the CXXC1 protein including a CXXC domain capable of binding CpGs (Allen *et al* 2006). In addition, specific recruitment of PcG complexes has been suggested to proceed via interactions with modified histones, non-coding RNA and transiently associated transcription factors (Kohler and Villar 2008, Simon and Kingston 2009).

In the following, we introduce a mathematical model of an epigenetic switch of transcriptional activity. This model describes the binding of protein complexes to chromatin which are able to read and to write histone marks. Molecular interactions of the complexes with DNA and with modified histones can create a regulatory switch changing gene activity between transcribed and silenced states. We analyse and discuss the switching regimes predicted by the model, the ability of the model to describe memory effects and the role of noise in the system. We apply the model to genome wide ChIP-seq data on H3K4me3 and H3K27me3 modifications

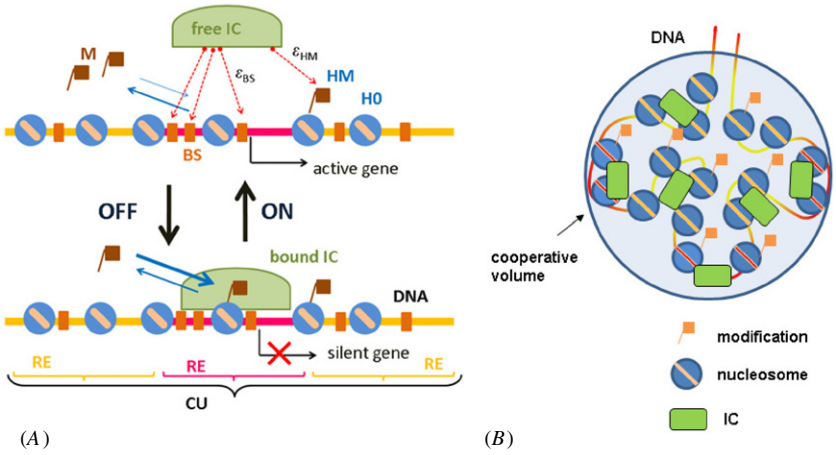

Figure 1. Transcriptional regulation by histone modification. (*A*) Schematic plot of the epigenetic switch model: the transcription of a selected gene is active if the interaction complex (IC) is in a free, i.e. unbound state. Reversible IC-binding to a response element (RE) is facilitated by interactions with DNA-binding sites (BS) located within the RE and with modified histones (HM, unmodified nucleosomes are marked as H0) within the CU containing the RE. ε_{BS} and ε_{HM} are the free energy increments per BS and modified nucleosome upon IC-binding, respectively. Bound ICs catalyse histone modifications, giving rise to a positive feedback loop between IC-binding and histone modification. Gene activity is repressed after IC-binding and increased after IC-release. (*B*) Spatial organization of the epigenetic switch. A CU clusters in a cooperative volume. Binding of individual ICs to DNA is improved in this volume by the presence of modified nucleosomes.

in pluripotent and lineage-committed mouse cells taken from Mikkelsen *et al* (2007). Crucial parameters of the model are fitted to these data and, by comparison of the parameter sets for the different cell types, hypotheses on the dynamics of histone modifications in these cells are derived.

1. Results

1.1. Mathematical model of an epigenetic switch controlling gene activity

1.1.1. General assumptions. In this section, we introduce a model of transcriptional regulation based on reversible binding of protein complexes to genomic loci which, in turn, is governed by histone modifications via a feedback mechanism. In our model we make the following basic assumptions (see figure 1(*A*) for illustration).

(i) Interaction complexes (ICs) represent proteins or complexes of proteins that are capable to bind to DNA in a sequence-specific manner. Particularly, we assume that each IC binds to one DNA-response element (RE) which contains a variable number n_{BS} of binding sites (BS). Each binding site additively contributes to the free energy of IC-binding to DNA.

(ii) Adjacent REs form cooperative units (CUs) of length L^{CU} given in units of the number of base pairs (bp) involved. Each CU is associated with $N_H = L^{CU}/200$ bp nucleosomes, where n_{HM} of them are in a modified (HM) and the remaining $N_H - n_{HM}$ are in an un-modified (H0) histone state. The nucleosomes of the CU containing modified histones (in short: 'modified nucleosomes'), facilitate the binding of free ICs to all REs within the CU. Hence, free ICs are capable of 'reading' histone modifications.

(iii) ICs bound to a RE are capable to 'write' histone modifications, i.e. they catalyze the histone modification reaction by facilitating the association of the modifier-enzyme with the histones. The inverse process, the histone de-modification reaction, is assumed to occur permanently, i.e. independent of the presence of bound ICs. Besides active (catalyzed) reactions, overall de-modification dynamics also includes passive processes as, e.g. dilution of modified histones during transcription.

(iv) Each CU is associated with one (or several) genes. Binding of ICs to REs within the CU silences these gene(s) (OFF-state) the transcription of which is otherwise active (ON-state).

ICs thus act here as repressive complexes which repress transcription according to assumption (iv). Alternatively one can apply the same assumptions (i)–(iii) but reverse the action in (iv), namely that a bound IC activates gene expression. In the following, we will focus on repressive ICs related to PcG complexes as potential example. Activating ICs such as trxG complexes are thought to function analogously but in reverse direction.

Formation of CUs that enable modifications of extended genomic regions may occur via chromatin looping as e.g. proposed by Tiwari *et al* (2008). A sketch of such a hypothetical chromatin region after cooperative binding of ICs is shown in figure 1(*B*). The region is folded into multiple loops of a length of a few kbp, each of them containing a subset of the REs. All REs of the CU cluster into a relatively small cooperative volume (coloured area).

Our model thus includes a series of mutually interacting structural elements which are organized in a hierarchical fashion: a number of BSs form an RE representing the elementary binding domain for an IC. Upon DNA binding the IC cooperatively catalyzes the modification of all histones within a CU formed by several adjacent REs. This IC-binding is, in turn, affected by the mean modification state of all nucleosomes involved. Below we introduce the rate equations

describing the model and solve them (i) by applying steady-state equilibrium thermodynamics to discover stationary solutions and (ii) by applying computer simulations to study the dynamics of histone modifications.

1.1.2. Binding equilibrium and modification reactions. We assume a binding equilibrium between free ICs, free REs and mutual RE/IC complexes according to the mass action law. The equilibrium constant of the binding reaction is governed by the concentrations [IC], [RE] and [RE/IC] of the respective species:

$$K_{\mathrm{RE/IC}} \equiv \frac{[\mathrm{RE/IC}]}{[\mathrm{RE}][\mathrm{IC}]} = \frac{1}{v}\exp(-\Delta G/kT), \tag{1}$$

where ΔG denotes the standard free enthalpy change upon binding and v is the reaction volume. If REs and ICs approach each other inside v they are assumed to react. The fraction of occupied REs, the so-called RE occupancy, is given by

$$\Theta = \frac{[\mathrm{RE/IC}]}{[\mathrm{RE}] + [\mathrm{RE/IC}]} = \frac{1}{\exp(\Delta g) + 1}. \tag{2}$$

The right-hand side is obtained after inserting equation (1) and rearrangement. The dimensionless free enthalpy of association, $\Delta g \equiv \Delta G/kT + \ln(v/[\mathrm{IC}])$, can be decomposed according to

$$\Delta g = \varepsilon_0 + n_{\mathrm{BS}}\varepsilon_{\mathrm{BS}} + n_{\mathrm{HM}}\varepsilon_{\mathrm{HM}}, \tag{3}$$

where $\varepsilon_0 \propto \ln(v/[\mathrm{IC}])$ denotes a basal contribution per IC to Δg and $\varepsilon_{\mathrm{BS}}$ and $\varepsilon_{\mathrm{HM}}$ are the free enthalpy changes per RE-binding site and per modified nucleosome, respectively. All these dimensionless quantities are given in units of the thermal energy kT. We assume $\varepsilon_0 > 0$ meaning that the basal contribution to Δg hampers IC-binding. In contrast, $\varepsilon_{\mathrm{BS}}$ and $\varepsilon_{\mathrm{HM}}$ are typically set to negative values. Consequently, IC binding is progressively facilitated with an increasing number of involved binding sites and modified nucleosomes n_{BS} and n_{HM}, respectively.

The number n_{BS} of binding sites in each RE is constant. In contrast, the number of modified nucleosomes in a CU can vary due to permanent (de-)modification activities. We consider a simple rate equation:

$$\frac{\mathrm{d}n_{\mathrm{HM}}}{\mathrm{d}t} = -k_m^- n_{\mathrm{HM}} + k_m^+ \cdot \Theta \cdot (N_H - n_{\mathrm{HM}}), \tag{4}$$

where k_m^- is the rate constant of de-modification whereas $k_m^+ \cdot \Theta$ defines the rate of modification. The latter is assumed to scale with the RE-occupancy ($0 \leqslant \Theta \leqslant 1$). Hence, histones are modified in the presence of bound ICs only. The constants k_m^+ and k_m^- will in general depend on the abundance of the respective modifiers.

Note that equations (3) and (4) apply a mean field approximation. It assumes that IC/RE binding is affected by the number of modified histones of all REs forming the CU and, in turn, that histone modification is affected by the mean IC-occupancy averaged over all REs of the CU. Consequently, n_{BS} in equation (3) defines the mean number of binding sites per RE in the CU.

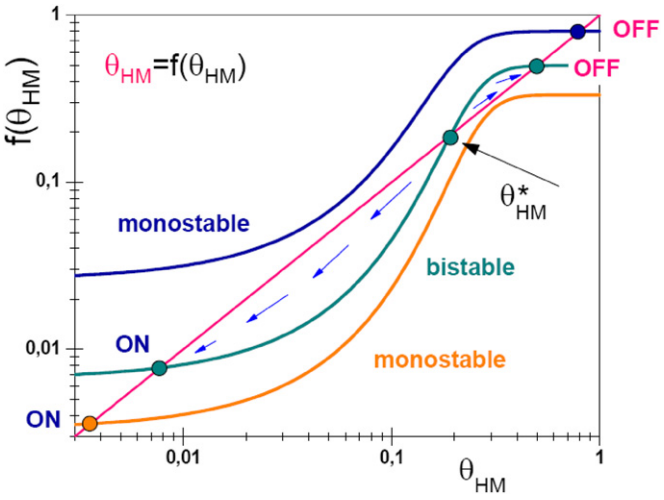

Figure 2. Stationary modification states. Double-logarithmic plot of the self-consistent equation (7) as a function of θ_{HM}. Plotted are the left hand (red) and the right hand side (orange, green, blue) of the equation for different parameter sets. Fixpoints of equation (7) are given by the intersections of the graphs (dots). One solution refers to monostable (ON or OFF), while three solutions refer to bistable behaviour (ON and OFF). In the bistable case an unstable solution $\theta_{\mathrm{HM}}{}^*$ exists between the respective stable ON and OFF solutions. It defines the range of the attractors of the stable solutions (see the text).

1.1.3. Stationary solutions. The stationary solution of equation (4) meets the condition $\mathrm{d}n_{\mathrm{HM}}/\mathrm{d}t = 0$. It provides the average fraction of modified nucleosomes per CU:

$$\theta_{\mathrm{HM}} \equiv \frac{\bar{n}_{\mathrm{HM}}}{N_H} = \frac{k_m^+ \cdot \Theta}{k_m^+ \cdot \Theta + k_m^-} = \frac{1}{1 + K_m/\Theta}, \tag{5}$$

where $\bar{n}_{\mathrm{HM}}$ is the mean number of modified nucleosomes per CU referring to stationary conditions. $K_m = k_m^-/k_m^+$ defines the nominal equilibrium constant of histone de-modification at maximum RE-occupancy $\Theta = 1$. Note that the equilibrium is governed by an effective constant K_m/Θ which inversely scales with Θ. The equilibrium value of the RE-occupancy used in equation (5) is a function of θ_{HM} given by equations (2) and (3):

$$\Theta = (\exp(\varepsilon_0 + n_{\mathrm{BS}}\varepsilon_{\mathrm{BS}} + N_H \cdot \theta_{\mathrm{HM}}\varepsilon_{\mathrm{HM}}) + 1)^{-1}. \tag{6}$$

Insertion into equation (5) finally provides the conditional equation linking the degree of histone modification with the parameters describing the IC/RE-binding equilibrium:

$$\theta_{\mathrm{HM}} = (1 + K_m \cdot (\exp(\varepsilon_0 + n_{\mathrm{BS}}\varepsilon_{\mathrm{BS}} + N_H \cdot \theta_{\mathrm{HM}}\varepsilon_{\mathrm{HM}}) + 1))^{-1}. \tag{7}$$

As demonstrated in figure 2, equation (7) has either one or three solutions depending on the parameter set chosen. In the case of one solution, the system is monostable. Thereby, the solutions may either refer to small or large values of θ_{HM}. The respective θ_{HM}-values transform into values of small and large RE-occupancy Θ after applying equation (6) which refer to predominantly active and silent genes, respectively. We will assign the respective solutions as 'ON' and 'OFF' with the definition $0 \leqslant \Theta_{\mathrm{ON}} \ll \Theta_{\mathrm{OFF}} \leqslant 1$.

In the case that equation (7) has three solutions the system is bi-stable. The smallest and the largest solution

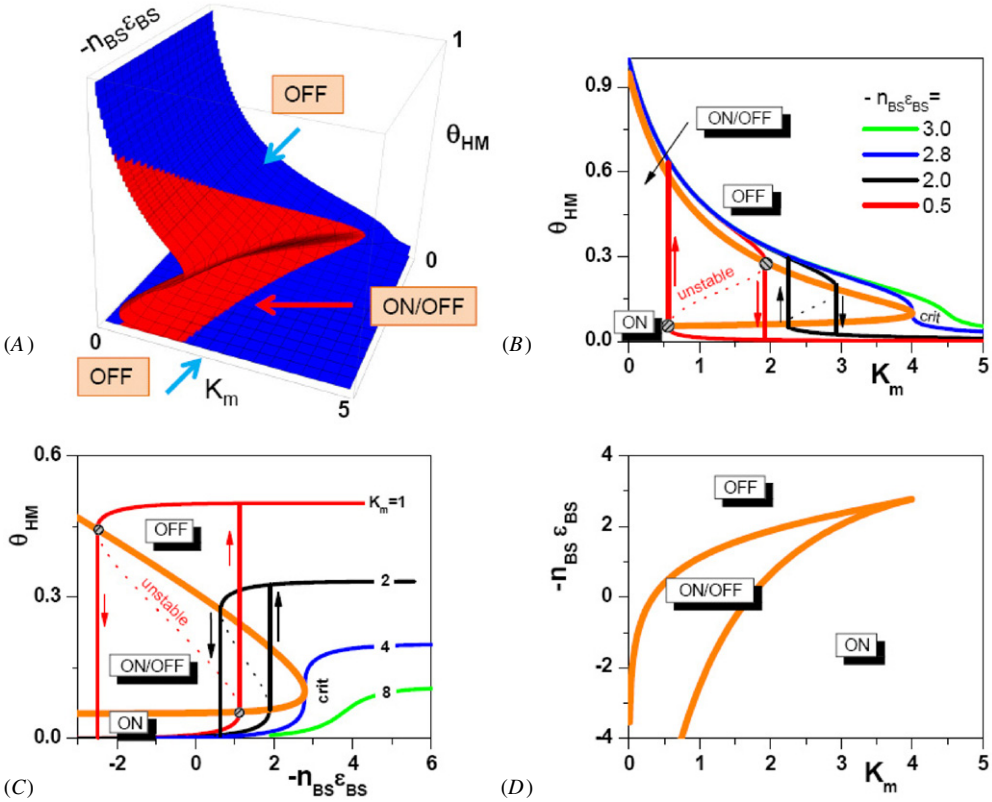

Figure 3. Regimes of gene activity. (*A*) Fraction of modified nucleosomes θ_{HM} in dependence of the equilibrium constant of histone de-modification (K_m) and the total IC/BS interaction free energy per RE ($n_{BS} \cdot \varepsilon_{BS}$) as given by equation (7) ($\varepsilon_0 = 5$, $-N_{HM}\varepsilon_{HM} = 20$). Monostable (ON or OFF) and bistable (ON/OFF) ranges of the solution are shown in blue and red colour, respectively. (*B*), (*C*) Projections of the solution shown in (*A*) onto the vertical coordinate planes in terms of iso-K_m and iso-BS trajectories, respectively. The bistable ranges (orange lines) are defined by the turning points of the iso-trajectories. For increasing and decreasing θ_{HM} the turning points differ, indicating hysteretic behaviour that is associated with memory. (*D*) Projection of the range of ON/OFF-bistability into the $n_{BS}\varepsilon_{BS}$-versus-K_m-coordinate plane, providing a phase diagram of the epigenetic switch.

for θ_{HM}, referred to as $\theta_{HM}{}^{ON}$ and $\theta_{HM}{}^{OFF}$, respectively, define the stable solutions while the intermediate one, $\theta_{HM}{}^{*}$, is unstable. $\theta_{HM}{}^{*}$ defines the range of the attractors of the stable solutions, i.e. following perturbations the system always converges to the OFF state for $\theta_{HM} > \theta_{HM}{}^{*}$, whereas for $\theta_{HM} < \theta_{HM}{}^{*}$ it converges to the ON state. The relations between the degree of histone modification, θ_{HM}, and its molecular determinants are disentangled in figure 3 using K_m and $n_{BS} \cdot \varepsilon_{BS}$ as examples. Figure 3(*A*) shows the surface defined by solutions of equation (7) depending on K_m and $n_{BS}\varepsilon_{BS}$. Projections of the solutions onto the θ_{HM}-versus-K_m (figure 3(*B*)) and θ_{HM}-versus-$n_{BS} \cdot \varepsilon_{BS}$ (figure 3(*C*)) coordinate planes provide iso-BS ($n_{BS} \cdot \varepsilon_{BS} = \text{const}$) and iso-$K_m$ ($K_m = \text{const}$) curves.

Essentially two types of iso-curves can be distinguished: (i) monotonically increasing or decaying curves with a constant sign of their slope over the whole range of the respective control parameter (blue, green curves) and curves that inverse the sign of their slope in a defined range of the control parameter (red, black curves). The region enclosed between the two turning points of the second type of curves (dotted part of the curves) defines unstable solutions of equation (7). Thus, moving along an iso-curve of the second type, the system switches from the OFF to the ON state or *vice versa* along a vertical line starting at the respective turning point. The two turning points of a particular iso-curve refer to different values of the control parameter. Thus, the state of the system will depend on its history, i.e. on whether the state was reached after increasing or decreasing the control parameter. Accordingly, the system provides a regulatory memory. The continuum of turning points in figures 3(*B*) and (*C*) (orange lines) defines the region of bi-stability with respect to K_m and $n_{BS} \cdot \varepsilon_{BS}$, respectively. They are obtained as described in the appendix A.1. Figure 3(*D*) shows the bistable region in K_m-versus-$n_{BS} \cdot \varepsilon_{BS}$ coordinates.

The degree of histone modification is used as the state function describing the status of the switch in figure 3. It can be transformed into the occupancy of the REs by IC complexes Θ via equation (6) which more directly defines the regimes of gene activity of the epigenetic switch. Related iso-curves in the Θ-versus-θ_{HM} coordinate system are provided and discussed in appendix A.2.

The above results demonstrate that our model is capable of describing an epigenetic switch of gene expression which is governed by the free energy of RE-specific IC binding $n_{BS} \cdot \varepsilon_{BS}$ and by the equilibrium constant of histone de-modification K_m. Additionally, the systems behaviour depends on the maximum interaction free energy of IC-binding to modified nucleosomes, $N_H\varepsilon_{HM}$, i.e. it is a function of the length of the CU given as the number of nucleosomes included.

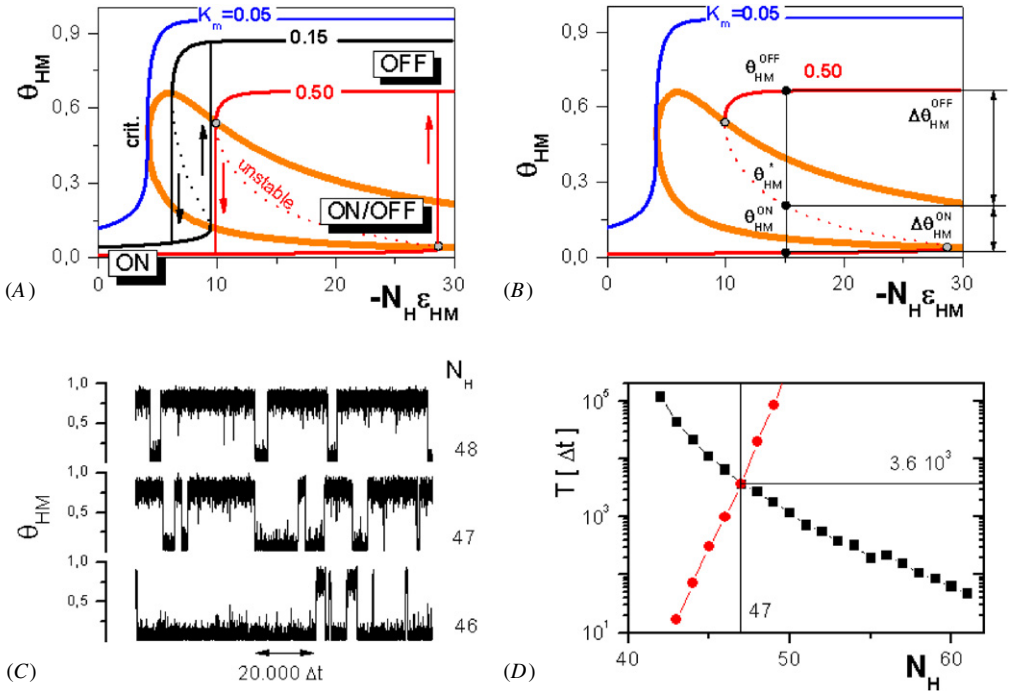

Figure 4. Length of cooperative units affects gene activity. (*A*) Solutions of equation (7) for variable $N_H\varepsilon_{\rm HM}$. Shown are iso-K_m trajectories assuming a constant value $n_{\rm BS}\varepsilon_{\rm BS} = 0$. A minimal value $K_m = 0.05$ (blue curve) is required for bistability. Details as in figures 3(*B*) and (*C*). (*B*) The minimum increment $\Delta\theta_{\rm HM}$ of $\theta_{\rm HM}$ that must be 'overleaped' by fluctuations to switch between the ON and the OFF state is given by $\Delta\theta_{\rm HM}{}^{\rm ON} = \theta_{\rm HM}{}^{*} - \theta_{\rm HM}{}^{\rm ON}$ for the ON state and $\Delta\theta_{\rm HM}{}^{\rm ON} = \theta_{\rm HM}{}^{\rm OFF} - \theta_{\rm HM}{}^{*}$ for the OFF state. (*C*) Under stochastic (de-) modification dynamics the bistable states have a finite life time. Shown are solutions for three CUs of different lengths $N_H = 46, 47, 48$, referring to $L^{\rm CU}$ of about 10 kb. $\varepsilon_{\rm HM}$ was set to -0.15 and K_m to 0.15. Time is measured in units of the sampling time Δt. (*D*). Mean life time in the ON state (black) and the OFF state (red). The state of balanced occupancy is defined by the intersection of both curves and refers to $N_H = 47$. Parameters are chosen as in (*C*).

Figure 4(*A*) shows K_m-isocurves in $\theta_{\rm HM}$- $N_H\varepsilon_{\rm HM}$-coordinates. The critical iso-curve demonstrates that an epigenetic switch requires a minimal value $K_m \approx 0.05$ in our example. Above this value, i.e. upon shifting the equilibrium towards de-modification of the histones, the system becomes bistable in general.

1.1.4. The impact of fluctuations. The maintenance of a stable, stationary chromatin structure, as described by the IC/RE binding equilibrium (see equation (7)), requires that fluctuations of $\theta_{\rm HM}$ due to histone (de-)modification are relatively small, because otherwise the system might switch between the ON and OFF states in the range of bistability.

Let us define the feedback-to-noise ratio, $F = \Delta\theta_{\rm HM}/\delta\theta_{\rm HM}$, as the ratio of the minimum increment $\Delta\theta_{\rm HM}$ of $\theta_{\rm HM}$ that has to be surpassed by noise to switch between the ON and the OFF state, and the (average) amplitude $\delta\theta_{\rm HM}$ of fluctuations present in our system. $\Delta\theta_{\rm HM}$ is given by the distance between the stable and the unstable fixpoint, i.e. $\Delta\theta_{\rm HM}{}^{\rm ON\to OFF} = \theta_{\rm HM}{}^{*} - \theta_{\rm HM}{}^{\rm ON}$ for ON→OFF transitions and $\Delta\theta_{\rm HM}{}^{\rm OFF\to ON} = \theta_{\rm HM}{}^{\rm OFF} - \theta_{\rm HM}{}^{*}$ for the OFF→ON transitions (see figure 4(*B*)). So far we neglected fluctuations assuming $\delta\theta_{\rm HM} = 0$, or equivalently, infinite values of F. Accordingly, all states were stable for an infinite life time. However, for finite values F spontaneous transitions between ON and OFF states become possible.

We studied this scenario for different lengths of the CU. In particular, the stochastic process underlying equation (4) was explicitly solved assuming modification and de-modification of the histones with fixed rates given by k^+ and k^-, respectively. The time step size Δt applied in the stochastic simulations thus refers to the characteristic time of the catalytic reactions governing histone modification dynamics.

Figure 4(*C*) shows time-dependent solutions of the fraction of modified nucleosomes for three CUs of different length N_H, for a parameter set that leads to bistable solutions of equation (4). The system clearly switches between the ON and the OFF state in all three examples. With increasing N_H the CUs progressively prefer the OFF state.

We estimated the mean life time of the ON- and OFF-states calculating the time T needed to observe a fluctuation $\delta\theta_{\rm HM} > \Delta\theta_{\rm HM}{}^{\rm OFF}$ (and $\Delta\theta_{\rm HM}{}^{\rm ON}$). Figure 4(*D*) shows the obtained life times as a function of the number of cooperatively acting nucleosomes. The intersection point defines the state of balanced occupancy of both states of the switch. For the chosen parameter set, it is realized by a CU containing 47 nucleosomes. Shorter and longer CUs prefer the ON and OFF states, respectively.

The life time of the more stable state (ON for $N_H < 47$ and OFF for $N_H > 47$, respectively) exceeds 3.6×10^3 time steps of the simulation. Taking this value and assuming characteristic catalytic times of histone modifications in the range of minutes (Zhang and Bruice 2008, Dirk *et al* 2007) provides a life time of the most stable state in the range of a few days ($\geqslant 3.6\ 10^3$ min). This value roughly agrees with measured turnover times of histone methylation marks (Zee *et al* 2010).

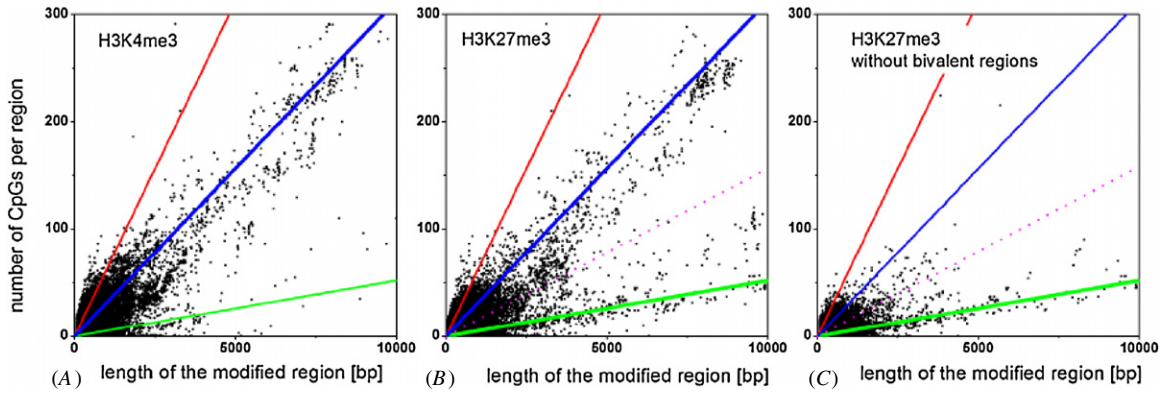

Figure 5. Chromatin modifications in ESCs. (*A*) Number of CpGs in H3K4me3 modified chromatin regions of defined length. The modification is associated with CpG-rich regions with an about six-fold higher CpG-density ρ_H (blue line) than the background (green line). Typically the CpG-density at CpG-rich promoters (red line) is about twice as large as that found for the enriched regions (blue line). Note that the promoter and background CpG-densities remain almost unoccupied for lengths greater 1000 bp and 4000 bp, respectively. (*B*) H3K27me3 modified chromatin regions are associated with both, CpG-rich regions and regions showing a CpG-content characteristic for the background. (*C*) H3K27me3 regions without H3K4me3 modification show no enrichment for CpGs. Thick solid lines refer to CpG densities that are actually observed in the particular system. Dotted lines refer to a CpG-density of $\rho_H/2$ which was considered as a threshold density for CpG-rich regions (see the text).

In contrast, the formation of multiple histone methylations has been observed to take up to some hours (Patel *et al* 2009). In this case the life time of the most stable state extends to several months or years ($>3.6 \times 10^3$ h) and thus implies long term stability of the switch.

In summary, catalytic reaction times ranging from minutes to hours, in dependence of the particular modification reaction, transform into life times of the epigenetic switch ranging from days to years referring to intermediate and long term stability of the associated states. However, additional large scale fluctuations, as e.g. observed after histone dilution during replication, can shorten the life time of modification states considerably (Micheelsen *et al* 2010, David-Rus *et al* 2009). Details of this mechanism are discussed in appendix A.3.

2. Experimental indications

Maintenance, lineage commitment and differentiation of various stem cell systems are under epigenetic control (Lee *et al* 2006, Oguro *et al* 2010). A genome-wide combined ChIP-seq and gene expression microarray data set has been published by Mikkelsen *et al* (2007) addressing the state of histone modification and of gene activity in murine embryonic stem cells (ESCs), murine embryonic fibroblasts (MEFs) and neuronal progenitor cells (NPCs). In these experiments, both ESCs and MEFs were obtained from C57BL/6 mice. NPCs were derived from the ESCs by differentiation through embryoid body formation and selection in insulin-transferrin-selenium fibronectin media. We re-analysed these data in order to characterize the length distribution of H3K4me3 and H3K27me3 modified chromatin regions in these cells and its relation to the CpG-content of the associated DNA sequences.

2.1. CpG content of modified regions

It is known that H3K4me3 and H3K27me3 are catalyzed by trxG- and the PcG-complexes, respectively. The recruitment of these ICs to a genomic locus has been suggested to depend on the local CpG-density (Thomson *et al* 2010). In particular, H3K4me3 preferentially associate with CpG-rich promoter regions (Mikkelsen *et al* 2007). Here, we investigated whether such association can be observed for any CpG-rich region throughout the genome and not only for promoter regions. After mapping of the ChIP-seq read data using the mapper 'segemehl' (Hoffmann *et al* 2009), we inferred H3K4me3 and H3K27me3 modified chromatin regions. Subsequently, we computed the absolute amount of CpG within these regions by counting each CG dinucleotide in the plus strand (mouse genome version mm9) as one CpG (see 'Materials and Methods' for details).

Figures 5(*A*) and (*B*) relate these CpG counts in H3K4me3 and H3K27me3 modified regions in ESCs to the length of the associated genomic regions. A constant CpG-density transforms into a straight line of positive slope in this plot. The 'background' CpG-density of 0.0052 CpGs/bp is provided by the green lines. The steep red lines refer to the CpG-density in 1.3 kbp-wide regions around the transcription start site of CpG-rich promoters (0.062 CpGs/bp (Mohn and Schubeler, 2009)). Our analysis reveals a third, intermediate CpG-density of $\rho_H =$ 0.031 CpGs/bp in DNA-regions associated with H3K4me3 modification in ESCs (blue lines in figure 5), which is slightly below the lower limit for CpG islands (0.038 (Bernstein *et al* 2007)). It represents a typical upper limit of the CpG-density found in longer DNA-fragments (>1.5 kbp), which is about half as large as that near CpG-rich promoters.

For H3K4me3 the background branch remains virtually empty, showing that most of the modified histones are associated with CpG-rich DNA (figure 5(*A*)). In contrast, H3K27me3-modified histones were found to associate with CpG-enriched regions, as well as with regions of CpG-background density. This bimodal CpG-density distribution suggests two types of recruitment for H3K27me3 either requiring or not requiring high CpG-density.

Figure 5(*B*) shows all H3K27me3 modified regions including monovalent and bivalent ones as well. After removal

of regions with bivalent-modifications the CpG-rich branch nearly de-populates (figure 5(*C*)), showing that CpG-driven PcG-binding strongly correlates with H3K4me3. However, to our best knowledge there is no evidence for a direct interaction between PcG and trxG complexes in the literature so far. Moreover, there is recent experimental evidence that if both marks are occurring on one nucleosome, they are generally found on different histone copies (Voigt *et al* 2012), supporting the idea that they are set by distinct (or, at least, only weakly coupled) processes. Consequently, we treated them independently. Remarkably, many of the detected CpG-rich DNA-regions associated with bivalent modifications exceed 3 kbp but only a very few of them exceed 10 kbp. This suggests a typical upper limit of about 50 nucleosomes per modified region which translates into an estimate $N_H \leqslant 50$ in our model.

In summary, the data provide evidence that H3K4me3 modification is genome-wide positively correlated with high CpG-density. Thus, in case H3K27me3 is found in CpG-rich regions it will almost always coexist with H3K4me3, giving rise to bivalent domains. In case H3K27me3 is found in domains devoid of H3K4me3, it does not depend on particularly high CpG-density.

2.2. *Length distribution of modified regions*

Differentiation and lineage commitment have been shown to modulate the regulatory environment of histone modifications as e.g. the activity of (de-)methylases (Reik 2007, Pedersen and Helin 2010). Changed modification probabilities of the histones potentially lead to an altered length distribution of the modified regions. To prove this expectation, we calculated the length distribution of H3K4me3 and H3K27me3 modified chromatin regions in ESCs and compared it with the length distribution of the sub-regions that remain modified in MEFs and NPCs (figures 6(*A*) and (*B*), respectively). However, in the latter case we considered CpG-rich regions only to exclude CpG-independent modification mechanisms discussed above. We set a threshold halfway between the CpG levels of the background and the CpG enriched regions (see dotted line in figures 5(*B*), (*C*)). Panels A and B of figure 6 show that the obtained frequency distributions of modified chromatin regions strongly decay with their length L_{mod} in all systems studied (see symbols).

To gain further insight into the (cooperative) maintenance of modified regions during differentiation, we normalized the data with respect to the length-dependent modification frequency in ESCs (black symbols in figures 6(*A*) and (*B*)) which provides the fraction of regions modified in ESCs that remain modified in MEFs and NPCs, respectively (figure 6(*C*) and (*D*)). The results show that the modification status of long regions (>3 kbp) remains virtually unchanged in MEFs compared with ESCs whereas in NPCs longer regions nearly completely lose their histone marks compared with ESCs.

We fitted these data applying our model of histone modification to chromatin regions of different length. For this purpose, we assumed that CUs are separated from each other by insulator motifs (Lunyak 2008) randomly distributed along the DNA, leading to a geometric distribution for the lengths of CUs (see Material and Methods). The overall decay of the length distribution can be explained by a simple approximation. In our model the modification probability θ_{HM} depends on the length of the CUs. It switches from a low to a high value increasing the length of the CU (see figure 4(*A*)). Thus, in the frequent short CUs the probability to find a modified chromatin region of length L_{mod}, i.e. a chain of $L_{\mathrm{mod}}/200$ bp modified nucleosomes, will be proportional to $\left(\theta_{\mathrm{HM}}^{\mathrm{ON}}\right)^{L_{\mathrm{mod}}/200\mathrm{bp}}$ while in the rare long CUs it will be proportional to $\left(\theta_{\mathrm{HM}}^{\mathrm{OFF}}\right)^{L_{\mathrm{mod}}/200\mathrm{bp}}$.

Accordingly, one can expect a roughly double exponential decay of the analysed frequency distribution of L_{mod}:

$$w(L_{\mathrm{mod}}) \approx A_{\mathrm{ON}} \exp(-L_{\mathrm{mod}}/\lambda_{\mathrm{ON}}) + A_{\mathrm{OFF}} \exp(-L_{\mathrm{mod}}/\lambda_{\mathrm{OFF}}), \quad (8)$$

with decay lengths $\lambda_{\mathrm{ON}} \approx -200\mathrm{bp}/\ln\left(\theta_{\mathrm{HM}}^{\mathrm{ON}}\right) \ll \lambda_{\mathrm{OFF}} \approx -200\mathrm{bp}/\ln\left(\theta_{\mathrm{HM}}^{\mathrm{OFF}}\right)$ which estimate the typical CU-length scales associated with the two alternative states of the switch and with amplitudes $A_{\mathrm{ON}} \gg A_{\mathrm{OFF}}$.

Actually, double exponential fits of the H3K4me3 and H3K27me3 data in ESCs (blue curves in figures 6(*A*) and (*B*)) provided in both cases $\lambda_{\mathrm{ON}} \approx 150$ bp and $\lambda_{\mathrm{OFF}} \approx 1500$ bp referring to $\theta_{\mathrm{ON}} \approx 0.26$ and $\theta_{\mathrm{OFF}} \approx 0.87$, respectively. These latter values are in good agreement with respective values obtained for N_H-dependence of the K_m-isolines of the cooperative switch (see figure 4(*A*)). The characteristic lengths refer to single histones (λ_{ON}) and to about 7–8 cooperatively acting histones (λ_{OFF}) ensuring IC-free and IC-bound chromatin regions, respectively. This result clearly shows that the observed length distribution can be explained by the existence of essentially two modification states of the nucleosomes which are governed by the CU-length dependence of the epigenetic switch model introduced above.

In order to gain quantitative insights into the changes occurring during lineage specification, we fitted the data applying our model of histone modification in detail. The model parameter sets ensuring the best least-squares fits of the experimental data are listed in table 1. Simulation details are given in the 'Material and Methods' section. As seen in figures 6(*A*) and (*B*) our model well describes the overall decay of the length distribution of modified chromatin regions in all systems studied. De-modification in MEFs compared to ESCs is well described assuming a CU-length-independent increase of the de-modification constant K_m (cyan curves in figures 6(*C*) and (*D*)). Surprisingly, the same approach fails in describing NPC data (dotted and dashed magenta curves in figures 6(*C*) and (*D*)). However, NPC data can nicely be fitted by assuming larger K_m values for long CUs (>3 kbp) than for shorter ones (solid magenta curves in figures 6(*C*) and (*D*)). This suggests that a targeted de-modification of long CUs (or spatial reorganization of long chromatin loops) is taking place for H3K4me3 and H3K27me3 during specification of ESCs into NPCs.

From the fitted model, we subsequently calculated the average fraction of modified nucleosomes throughout the genome. Applying the parameter set given in table 1, our model predicts that in ESCs histones of about 25% of all nucleosomes

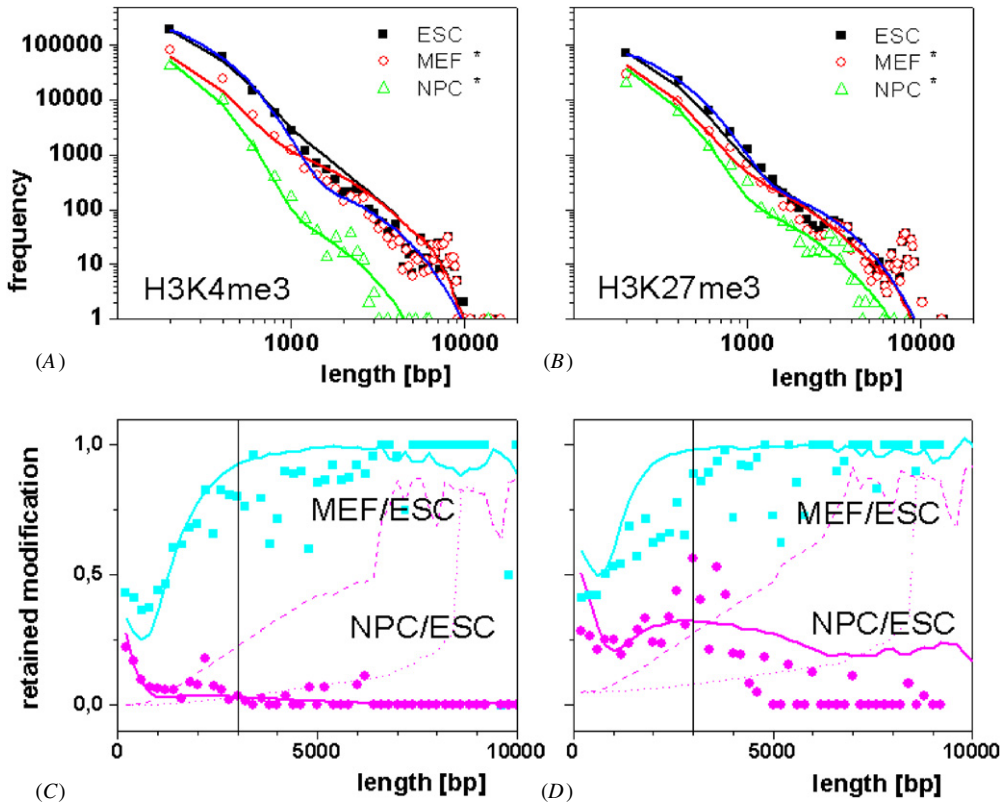

Figure 6. Changes of histone modifications during lineage commitment. (*A*) and (*B*) Length distribution of chromatin regions in ESCs, MEFs and NPCs modified by H3K4me3 and H3K27me3, respectively. We considered only CpG-rich regions which are modified in ESCs. Lineage commitment results in a globally decreased number of modified regions. Data points refer to the experimental distributions. Blue curves are double exponential fits of the ESC data. The other curves are theoretical distributions calculated applying our model. The differences observed between ESCs and MEFs can be described by globally increased K_m. The effect in NPCs requires assuming a local and thus targeted de-modification of long regions. The parameters used are summarized in table 1. (*C*) and (*D*) Fraction of H3K4me3 and H3K27me3 modified regions in ESCs, that remain modified in MEFs (cyan) and NPCs (pink) as a function of their length. Symbols are experimental data and lines are simulated ones. Modelling results for $K_m = 0.05$ (dashed curve) and $K_m = 0.07$ (dotted curve) show that for both marks a length-independent, global increase of K_m does not reproduce NPC data (see the text).

Table 1. Fitted model parameters. Parameters found by minimizing the sum of squared residuals between the experimental and theoretical length distributions.

Modification	Cell Type	ε_0	ε_{BS}	ε_{HM}	K_M
H3K4me3	ESC	5	−0.075	−0.15	0.0370
	MEF				0.0392
	NPC				0.0395 (L^{CU} < 15 nucleosomes), 0.16 ($L^{CU} \geqslant$ 15 nucleosomes)
H3K27me3	ESC	5	−0.025	−0.15	0.0370
	MEF				0.0385
	NPC				0.0388 (L^{CU} < 15 nucleosomes), 0.07 ($L^{CU} \geqslant$ 15 nucleosomes)

become H3K4me3-modified. Interestingly, the experimentally obtained averaged coverage with this modification is about 3% only, i.e. one magnitude smaller than expected by theory. This difference can be explained by assuming that only part of the whole genome is accessible to trxG complexes. In fact, under the assumption that about 12% of the nucleosomes are actually accessible and 25% of them are modified (as suggested by the model) one recovers a total of 3% modified nucleosomes. This estimation of the accessible part of the genome nicely agrees with the estimated fraction of nucleosomes that is associated with genomic regions of CpG-densities $\rho > \rho_H/2$ (∼11%, see 'Material and Methods'). Accordingly, we suggest that only regions of sufficiently high CpG-density are accessible to trxG complexes. Our model formally describes inaccessible genomic regions by large positive values of the basal interaction term $\varepsilon_0 \gg 5$ which practically excludes IC-binding. Thus, our model suggests that high CpG-density reduces the energy barrier that has to be overcome for trxG recruitment in ESCs from values $\varepsilon_0 \gg 5$ to about 5, a value that enables sufficient IC-binding.

Note that all proper fits of the length distributions of L_{mod} yielded absolute values of $|n_{BS}\varepsilon_{BS}| \ll 1$ suggesting that CpG-independent, sequence-specific recruitment of ICs plays only a minor role. Moreover, our fits provide values of $\varepsilon_{HM} \approx -0.15$ for both H3K4me3 and H3K27me3. This suggests that only CU-lengths larger than 6 kbp ($N_H > 30$)

can acquire bi-stable modification states capable to create a memory (recall that the critical iso-curve for bi-stable modification refers to -$N_H\varepsilon_{HM}$ ~4, see also 'Materials and Methods' and figure 4(*A*)).

In summary, our model predicts that binding of trxG (PcG) complexes to chromatin regions in ESCs is mainly governed by the local CpG-density and the capabilities of the complexes to read histone marks. Moreover, the data suggest that long modified regions in ESCs are subject to a targeted de-modification or spatial reorganization of associated chromatin loops during lineage specification into NPCs.

3. Discussion

3.1. The epigenetic switch model predicts different modes of gene activation

Chromatin modifications maintain regulatory states and also provide effective modes for changing them as required to adapt complex cell fates to variable environments. In our model this ambivalence is related to a positive feedback topology inducing bistable switch-like systems behaviour that establishes a molecular memory. There is growing evidence that this kind of epigenetic regulation underlies many fate decisions in development and stem cell differentiation (Pietersen and van Lohuizen 2008).

The states of such regulatory switches and the respective pattern of gene expression are governed by complex molecular interactions. Nonetheless, it has been demonstrated that already a relatively small toolkit of DNA binding, nucleosome binding and enzymatic activities can provide a flexible and reliable epigenetic regulation (Dodd and Sneppen 2011). In our model these components are, for the first time, explicitly considered in terms of binding isotherms of the enzymatic complexes, allowing us to distinguish contributions from DNA binding, histone modification states and spatial organization.

Accordingly, the regimes of gene activity are determined by the interaction free energy of the ICs with DNA ($\varepsilon_0 + n_{BS} \cdot \varepsilon_{BS}$) and modified nucleosomes ($N_H \cdot \varepsilon_{HM}$) as well as by the equilibrium constant of histone de-modification K_m. The latter constant depends on the local abundance of the respective (de)-modifying enzymes which are mostly under transcriptional and post-transcriptional control. Thus, variation of enzymatic activity can occur on the time scale of development and differentiation. In general, the activity of the modifying enzymes is determined by their abundance in the medium surrounding the DNA. One particular value of K_m will consequently apply to several CUs located for example in the same functional compartment of the cell nucleus such as transcriptional factories (Chakalova *et al* 2005).

The number n_{BS} and the strength ε_{BS} of DNA binding sites are linked to DNA sequence which typically adjusts by mutational mechanisms on an evolutionary time scale. The same time scale usually applies to changes of the structure and composition of the ICs. However, these complexes are composed of several molecular building blocks where their alternative compositions can be functional as well (Kouzarides 2007). Thus, the variation of $n_{BS} \cdot \varepsilon_{BS}$ or ε_{HM} due to changes in the composition of the IC may also provide an option for transcriptional regulation on the shorter time scales of development and differentiation.

The length of the CUs and thus the number N_H of cooperative nucleosomes has been suggested to self-organize along boundaries between domains of competitive histone marks (Lunyak 2008). Accordingly, N_H could vary depending on the regulatory environment. Here, we assumed a fixed distribution of N_H. This can be motivated by the idea of a sequence-specific binding of a barrier insulator ((Lunyak 2008) see also 'Materials and Methods'). In general, variation of the distribution of N_H may contribute to regulation of the binding equilibrium on both the time scales of differentiation and evolution.

Importantly, the approximation of the length dependence of modified regions by a double exponential fit function allows the simple and straightforward interpretation of data in terms of the characteristic CU-lengths and modification probabilities of the two alternative states of the switch. This function virtually links the properties of the single switch with the effective behaviour of an ensemble of such switches associated with a large number of cooperative chromatin regions of different lengths.

3.2. Spatial chromatin organization affects cooperativity of the switch

Early models for nucleosome modification, including histone modifications, invoke a linear stepwise process, where a modified nucleosome stimulates the modification of its nearest neighbours (Grewal and Elgin 2002). Nucleosome modifications were thus envisioned to spread in a continuous fashion along the DNA. Alternatively, Dodd *et al* (2007) assumed that during recruitment of modifying complexes any nucleosome can act on any other nucleosome in the region, and thus modification states can 'jump' across many nucleosomes that are in arbitrary modification states. This kind of non-local transmission is suggested to be facilitated by higher-order chromatin structure, e.g. by DNA-looping, or by more complex processes such as the passing of an RNA polymerase.

Our model assumes that all histones in the region of a CU are cooperatively affected by modification reactions without explicitly assuming a propagation mechanism for modification. In other words, our model assumes non-local interactions between ICs and the modified nucleosomes, e.g. mediated through a particular 'biochemical milieu' created by the modified histones (Wachsmuth *et al* 2008). The interpretation raises questions about mechanisms of chromatin compartmentalization that define the CUs (Lunyak 2008) and about the general relevance of chromatin looping for gene expression (Chakalova *et al* 2005, Deng and Blobel 2010). Recent experimental findings on frequent long range interactions among H3K27me3 modified domains support the idea of higher order folding of chromatin (Tolhuis *et al* 2011).

3.3. The inheritance of histone modification states is limited

We have shown that fluctuations due to ongoing modification and de-modification reactions of the nucleosomes can

modulate the dynamic regime of the epigenetic switch as a function of the length of the CU. Another process that potentially changes the nucleosome modification state is cell division. An accepted assumption states that parental histones distribute randomly onto daughter strands upon DNA replication. Afterwards the full population of histones is restored by *de novo* synthesized, unmodified histones (Probst *et al* 2009, Margueron and Reinberg 2010). In consequence the modification degree of the nucleosomes on average halves upon cell division. Thus, cell division represents a large scale fluctuation of nucleosome modification degree, which can rapidly change histone modification states (David-Rus *et al* 2009, Micheelsen *et al* 2010). Accordingly the parental state of nucleosome modification may either recover and become inherited or not. In the latter case the modification degree is altered in at least one of the daughter cells. Our model predicts that, also during cell division, the stability of the switch is modulated by the length of the CUs. Thereby, long term stability is observed in the case of long CUs and high de-modification constants (see appendix A.3). This limited cellular capability of histone modification states has been suggested to catalyze cellular ageing processes and can explain experimentally observed hypo- and hyper-methylation of genes (Przybilla *et al* 2012) known to be involved in tumor formation (Calvanese *et al* 2009).

4. Conclusions

Our epigenetic switch model provides new insights into an essential epigenetic layer of transcriptional regulation. Experimentally observed histone modification patterns, e.g. the length distribution of modified regions, can directly be related to a small set of model parameters, allowing to formulate hypotheses on the interplay of key players driving epigenetic reorganization during differentiation and development. The model connects local molecular dynamics and overall chromatin structure, e.g. assumptions on the length distribution of cooperative units (chromatin loops), in a straightforward manner. Hence, it provides a powerful toolkit for future extensions into multi-scale and multi-level models of epigenetic regulation.

Integrated models of this kind, however, require explicit consideration of cross-talk with other regulatory layers, which is conveniently provided e.g. by integration into the framework of an artifical genome model (Binder *et al* 2010, Rohlf and Winkler 2009). Such extensions will enable new approaches to study cross-talk between cis-regulatory networks of transcription factors and transcriptional regulation by histone modifications and DNA methylation (Rohlf *et al* 2012). Our model may help to identify crucial interactions in the epigenetic machinery shaping the different regulatory phenotypes of living cells, and thus provide a basic step to unravel the syntax of the 'chromatin language' underlying transcriptional regulation.

5. Materials and methods

5.1. Data analysis

We re-analysed the raw data set provided by Mikkelsen *et al* (GSE12241: H3K27me3, H3K4me3, H3K9me3 and whole cell extract of ESCs, MEFs and NPCs as well as H3 of ESCs). The read data were mapped against the UCSC mouse genome version mm9 using the in-house tool 'segemehl' (Hoffmann *et al* 2009). This tool has been designed for fast mapping of short reads under consideration of mismatches, insertions and deletions. We considered only seeds with at most two sequence differences. The top-scored hits were selected under the condition that the obtained accuracy was at minimum 90%. On average, 90.21% of all reads could be mapped by this procedure. Afterwards, we counted the number of mapped reads for each position in the genome.

ChIP-seq data from experiments using an antibody against the H3 histone provide reliable information about the positioning of relevant nucleosomes. In the case of ESC data sets, we used them in order to verify the H3 modification data. Only reads overlapping with reads from the H3 data, i.e. with defined nucleosome positions, were passed on. Since less than 1% of the reads were not passed on, we did not assume a bias in comparison with MEF or NPC data. All data were normalized by the read count of the corresponding whole cell extracts. We required a 3-fold enrichment of reads in the modification data compared to the whole cell extracts to consider a site as modified. Modified sites were merged together if the distance between them was less than 100 nt (about one-half of the average nucleosome spacing) into modified regions.

Analysing the stability of modified regions during differentiation of ESCs into MEFs or NPCs, we focused on the set of regions modified in ESCs. We started with this set and calculate for each region the number of nucleotides covered with regions from the same modification in another cell type. This coverage value was normalized with the length of the region in ESCs. The set of regions remaining modified in MEFs and NPCs was defined as the set of those regions in ESCs having a normalized coverage of at least 0.95 by MEF and NPC data, respectively.

The dependence of the modification probabilities on the length of the modified region was analysed based on the enrichment values for the nucleotides. For each region, the average sum of enrichment values for each nucleosome is calculated. Hereby, the regions' length is divided by 200 (approximately the number of nucleotides occupied by one nucleosome including linker DNA) and rounded. Regions with the same number of nucleosomes are grouped together. For each group, a standardized length of 200 times the number of nucleosome is used and the average nucleosome-wise enrichment of all regions in the group is calculated.

The number of CpGs in each modified region was simply counted as the number of CG di-nucleotides in the respective genomic sequence from the plus strand defined in the mm9 genome. We counted each CG di-nucleotide as one CpG. The CpG content of randomly selected DNA regions (size distribution as that of the modified regions) was considered

as 'background' CpG-density. The calculation of the CpG-content of randomly selected DNA regions was further used to estimate the overall fraction ξ of nucleosomes associated with local genomic regions with a CpG-density above $\rho_H/2$, in the following called $\rho_H/2$-regions. The amount of DNA typically associated with a single nucleosome provides a representative length scale $L \approx 200$ bp for the respective measurement. Randomly selecting $n_0 = 2 \times 10^6$ regions of length $L = 200$ bp from the whole mouse genome, we counted the number n_1 of $\rho_H/2$-regions among them and used the fraction n_1/n_0 as an estimate of ξ. Applying this procedure, we obtained $\xi = 0.11$ for the mouse genome.

5.2. Estimating the length distribution of modified regions

In the following, we outline the model which has been applied to generate the theoretical probability $w(L_{\mathrm{mod}})$ to find modified chromatin regions of length L_{mod} throughout the genome. We first applied our model to CUs of different lengths and then calculated the probability to find patches of uniquely modified nucleosomes within these CUs which actually provides $w(L_{\mathrm{mod}})$.

The length of the DNA sequence associated with one CU is given by $L^{\mathrm{CU}} = 200$ bp N_H. We assume that adjacent CUs are separated by so-called barrier insulators (Lunyak 2008) representing short sequence motifs of length l_0. To estimate the probability distribution to find a CU of length L^{CU} we make use of an artificial genome which is generated as random integer sequences with an alphabet of size 4, similar to the ATGC alphabet in a random genome model developed previously (Binder *et al* 2010, Rohlf and Winkler 2009). Accordingly, a specific insulator sequence of length l_0 occurs with probability 4^{-l_0} and the probability that two neighbouring insulators are separated by a sequence of length L^{CU} is given by the geometric distribution:

$$p(L^{\mathrm{CU}}) = 4^{-l_0}(1 - 4^{-l_0})^{L^{\mathrm{CU}}}. \tag{9}$$

We have chosen $l_0 = 5$ bp referring to an average L^{CU} equal to 1024 bp. Larger values of l_0 lead to slightly different parameter sets in the fitting procedure described below however without consequences for the interpretation of the results.

Then, we generated random 'nucleosome strings' of $N_H(i) = L^{\mathrm{CU}}(i)/200$ bp nucleosomes in a row (i is the running index of the CU) with probability $p(L^{\mathrm{CU}})$ that can potentially be modified after binding of interaction complexes. Each nucleosome $j = 1 \ldots N_H(i)$ in CU number i carries a binary variable:

$$h_{ij} = \begin{cases} 1 & \text{with probability} \quad \theta^i_{\mathrm{HM}} \\ 0 & \text{with probability} \quad 1 - \theta^i_{\mathrm{HM}} \end{cases}, \tag{10}$$

where θ^i_{HM} is the mean fraction of modified nucleosomes provided by the solution of equation (7) for CU number i:

$$\theta^i_{\mathrm{HM}} = (1 + K_m(\exp(\varepsilon_0 + n_{\mathrm{BS}}(i)\varepsilon_{\mathrm{BS}} + N_H(i)\theta_{\mathrm{HM}}\varepsilon_{\mathrm{HM}}) + 1))^{-1}. \tag{11}$$

h_{ij} consequently characterizes the histone modification state of all nucleosomes considered. In the case of bistable CUs, we selected either low $\left(\theta^{\mathrm{ON}}_{\mathrm{HM}}\right)$ or high $\left(\theta^{\mathrm{OFF}}_{\mathrm{HM}}\right)$ modification probability at random proportional to the size of their attractors $\theta_{\mathrm{HM}}{}^*$ and 1-$\theta_{\mathrm{HM}}{}^*$, respectively (see figure 4(*B*)). Note that the discretization introduced by equation (10) mimics the corresponding discretization of histone states by means of an enrichment threshold applied to the experimental data. It assumes that enrichment increases with modification probability, implicating that both quantities tend to increase with the length of modified regions (compare figure A3).

The number of binding sites per RE in CU number i was assumed to scale with the length of the RE with 1/200 bp. We set an upper cut-off limit of $n_{\mathrm{BS,max}} = 5$. This implicates a maximum RE-length of 1 kb, which is a lower bound of observed REs (Cuddapah *et al* 2012). $w(L_{\mathrm{mod}})$ was obtained by counting the number of adjacent nucleosomes with $h_{ij} = 1$, irrespectively of their membership to the same or to different REs and subsequent normalization with respect to the total number of modified nucleosomes. The length of each modified chromatin region L_{mod} is simply given by applying the scaling factor 200 bp as explained above. Finally, $w(L_{\mathrm{mod}})$ was calculated as average over one hundred independent genome realizations.

We varied $\varepsilon_{\mathrm{BS}}$, $\varepsilon_{\mathrm{HM}}$ and K_m while fixing $\varepsilon_0 = 5$ and $p(L^{\mathrm{CU}})$ to find theoretical length distributions which best fit the experimental distributions in ESCs by minimizing the sum of squared residuals (SSR):

$$\mathrm{SSR} = \frac{1}{L^{\max}} \sum_{L_{\mathrm{mod}}=1}^{L^{\max}} (\log[w_{\exp}(L_{\mathrm{mod}}) + \delta] - \log[w_{\mathrm{theo}}(L_{\mathrm{mod}}) + \delta])^2, \tag{12}$$

where $L^{\max}$ is the maximal observed length of modified regions and δ is a small offset (typically set to 10^{-6}) to avoid divergence of the logarithms for cases where $w = 0$. Only the parameter K_m was varied (globally or dependent on L^{CU}, see the text), while all other parameters were fixed to their ESC values to obtain the length distributions in MEFs and NPCs. The parameter sets ensuring the best fits are listed in table 1.

For estimating the length distributions in MEFs and NPCs, we considered coverage values c of the corresponding modifications relative to their distribution in ESCs. We define

$$c^i(*) = \frac{\theta^i_{\mathrm{HM}}(*)}{\theta^i_{\mathrm{HM}}(\mathrm{ESC})}, \tag{13}$$

where $\theta^i_{\mathrm{HM}}(*)$ is the modification rate in MEFs or NPCs. Only sites modified in ESC are taken into account. In order to classify a site as 'still modified', we demand $c > 0.95$, as it was required for the corresponding experimental data.

Note that the model, due to its mean field character, cannot explain why certain sites within CUs are more stable than close neighbours by principal reasons. Hence, $c(i)$ has to be interpreted as an average tendency to resist de-modification during lineage specification.

Acknowledgments

This work was supported by the BMBF grant MAGE (grant number 50500541). LS was supported by the DFG-Projects STA 850/6 and STA 850/2, the EU-Project Quantomics (grant

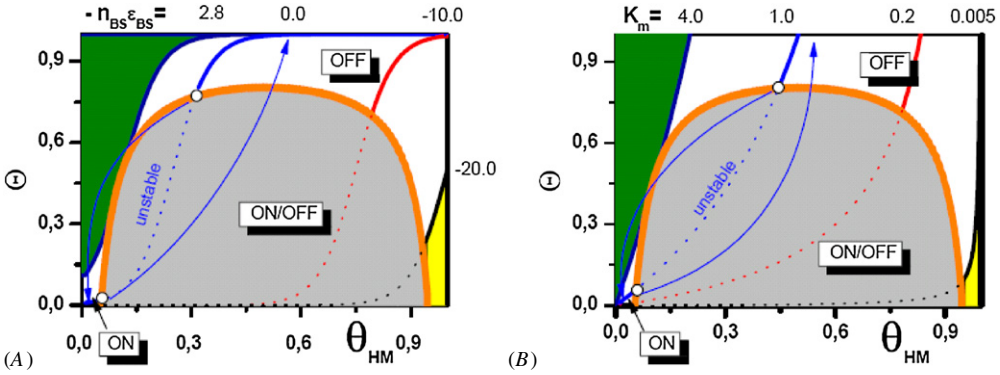

Figure A1. Phenotypic maps of the switch. The maps link the occupancy of the REs Θ with the degree of histone modification θ_{HM}. Panels (*A*) and (*B*) show iso-curves for $-n_{\mathrm{BS}}\varepsilon_{\mathrm{BS}}$ and K_m, respectively ($\varepsilon_0 = 5$, $-N_{\mathrm{HM}}\varepsilon_{\mathrm{HM}} = 20$). The ranges of ON/OFF bi-stability are shown by grey areas. The continuous regime of the switch is determined by the respective critical iso-curve (dark blue) forming the tangent to the bi-stability range. The blue iso-curves illustrate hysteretic behaviour as shown in figures 3(*B*) and (*C*). The small circles and the blue arrowheads indicate the onset and completion points of the respective transition. For parameters associated with the yellow region the system does not function as regulatory switch of gene activity (see the text).

number 222664) and the VW-Project 'Komplexe Netzwerke'. The project LIFE is financially supported by the European Fonds for Regional Development (EFRE) and the State of Saxony (Ministry for Science and the Arts).

Appendix

A.1. Calculating the range of bistability

The two solutions of $\theta_{\mathrm{HM}}{}^{(1,2)}$ referring to the upper and lower branches of the ON/OFF-range in figures 3(*B*) and (*C*) are obtained in two steps. Firstly, we rearranged equation (7) according to

$$n_{\mathrm{BS}} \cdot \varepsilon_{\mathrm{BS}} = \ln\left[\frac{1-(1+K_m)\cdot\theta_{\mathrm{HM}}}{K_m\cdot\theta_{\mathrm{HM}}}\right] - (\varepsilon_0 + N_H\cdot\theta_{\mathrm{HM}}\cdot\varepsilon_{\mathrm{HM}}) \tag{A.1a}$$

and

$$K_m = \frac{(1-\theta_{\mathrm{HM}})}{\theta_{\mathrm{HM}}\cdot(1+\exp(\varepsilon_0 + n_{\mathrm{BS}}\varepsilon_{\mathrm{BS}} + N_H\varepsilon_{\mathrm{HM}}\cdot\theta_{\mathrm{HM}}))}, \tag{A.1b}$$

respectively. Secondly, we took the first derivative of equation (A.1*a*) and (A.1*b*) with respect to θ_{HM} and calculated the roots:

$$\theta_{\mathrm{HM}}^{(1,2)}(K_m) = \frac{1}{2(1+K_m)}\left[1 \mp \sqrt{\frac{K_m^{\mathrm{crit}} - K_m}{K_m^{\mathrm{crit}}+1}}\right] \quad \text{and}$$

$$K_m^{\mathrm{crit}} \equiv \frac{N_H\varepsilon_{\mathrm{HM}}}{4} - 1 \tag{A.2a}$$

and

$$\left(\theta_{\mathrm{HM}}^{(1,2)}(\varepsilon)\right)^2 - \theta_{\mathrm{HM}}^{(1,2)}(\varepsilon) + \frac{1}{N_H\varepsilon_{\mathrm{HM}}}\left(1+\exp\left(-\varepsilon_0 - n_{\mathrm{BS}}\varepsilon_{\mathrm{BS}} - N_H\varepsilon_{\mathrm{HM}}\cdot\theta_{\mathrm{HM}}^{(1,2)}(\varepsilon)\right)\right) = 0 \tag{A.2b}$$

either analytically equation (A.2*a*) or solving equation (A.2*b*) numerically. The onset-points of ON/OFF bi-stability shown in figure 4(*A*) were calculated analogously.

A.2. Stationary solutions: IC binding probabilities determine gene activity

The regimes of gene activity of the epigenetic switch indicated in figures 3(*A*)–(*D*) are defined by the occupancy Θ of the REs by ICs via equation (6). Figure A1 shows selected iso-curves in the Θ-versus-θ_{HM} coordinate system focusing on process control under iso-K_m (figure A1A) and iso- $n_{\mathrm{BS}}\cdot\varepsilon_{\mathrm{BS}}$ (figure A1B) conditions. The critical iso-curves border the continuous regime of the switch in the left upper part of the figure. The other iso-curves refer to the hysteretic behaviour shown in figure 3. The dotted parts define their unstable regions. Upon approaching the bi-stability range along the stable part of the iso-curves, the system switches from the OFF into the ON state and vice versa along the respective arrows where the grey circles indicate the onset points of the switch and the arrowheads their completion points.

The ON regime is enclosed within a narrow region at low RE occupancies and low degrees of histone modification (both in the order of magnitude of a few per cent). Contrarily, the OFF regime occupies a larger area. It refers to large values of the RE-occupancy at intermediate degrees of histone modifications. Interestingly, at large θ_{HM}-degrees the 'OFF regime' progressively expands towards small Θ-values referring to unblocked and thus active genes contradicting its intended OFF status. The black iso-curves shown in figures A1A and B reveal that this subregion refers to large positive values of $n_{\mathrm{BS}}\varepsilon_{\mathrm{BS}}$ and small values of the de-modification constant K_m, respectively.

For such parameter values (see yellow areas in figures A1A and B) the system is dominated by the repulsion between the IC and the DNA. Although up to 100% of the nucleosomes are modified, due to vanishing de-modification activity, the resulting attraction of the IC by the modified histones is not sufficient to overcome the repulsive interaction with the DNA binding sites. Within these parts of the phase diagram the system does not function as epigenetic switch, since transitions between low and high degrees of histone modifications do not change the state of gene activity, which is assumed to depend on Θ.

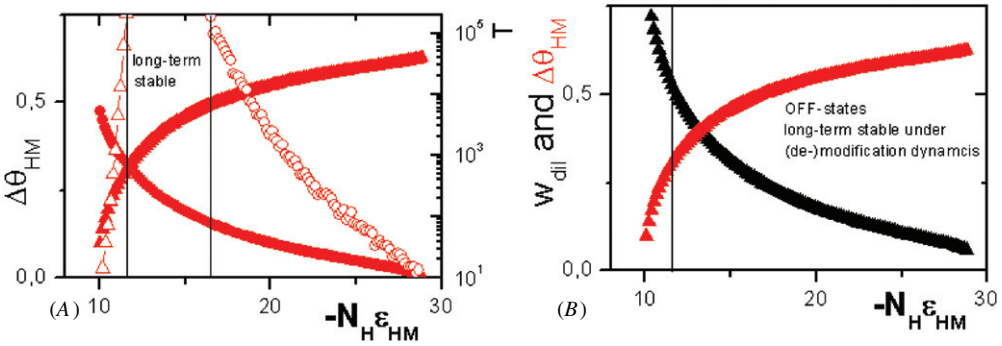

Figure A2. Impact of fluctuations due to permanent (de)-modification (*A*) and spontaneous de-modification during replication (*B*). (*A*) For $K_m = 0.5$ part of the bistable states is long-term stable under demodiftion and modification fluctuations. Shown are the increments $\Delta\theta_{HM}^{OFF–ON}$ (solid triangles) and $\Delta\theta_{HM}^{ON–OFF}$ (solid circles) in dependence of $N_H \cdot \varepsilon_{HM}$ as well as the respective mean life times T to observe these fluctuations (open symbols). (*B*) Shown is the probability w_{dil} (black symbols) that the amount of histone dilution in at least one daughter cell is larger than $\Delta\theta_{HM}^{OFF–ON}$ (red symbols). It can be seen that also modification states that are long-term stable under (de-) modification dynamics (range to the right of the vertical line, see also panel (*A*)) have a high probability up to 0.5, to get lost in at least one daughter cell via switching into the ON-state after dilution.

A.3. Fluctuations due to high de-modification activity and cell replication

As shown in figure 4(*A*), our model predicts that increasing the de-modification constant K_m at constant interaction energy ($n_{BS}\varepsilon_{BS}$) the system remains bi-stable. Thereby, the turning points of the isotherms move to larger values of $-N_H \cdot \varepsilon_{HM}$ paralleled by a strongly increased horizontal range of the hysteresis and slightly decreased vertical switching amplitude $\Delta\theta^{ON} + \Delta\theta^{OFF}$ (compare K_m-isocurves for $K_m =$ 0.15 and 0.5 in figure 4(*A*)). The simulation of the fluctuations for $K_m = 0.15$ reveal a monotonically increasing mean life time T of the OFF-state and a decreasing T of the ON-state of the switch in the bistable region with increasing length of the CU (see figures 4(*C*) and (*D*), $\varepsilon_{HM} = -0.15$). Figure A2A shows analogous results for the de-modification constant $K_m = 0.5$. Importantly, in this case of high de-modification activity both the mean life times of the ON- and of the OFF- states exceed $T > 10^5$ for values of $-N_H\varepsilon_{HM}$ between 12 and 16. In other words, the ON- and OFF-states become virtually fully stable in this range meaning that the system provides a real long term memory. However, for $\varepsilon_{HM} = -0.15$, as obtained fitting the ESC data, this would apply for very long CUs comprising more than 80 nucleosomes, i.e. lengths $L^{CU} > 16\,000$ bp, only.

Our simulation considers fluctuations of θ_{HM} caused by permanent modifications and de-modifications of the nucleosomes. Another kind of fluctuations of θ_{HM} can occur during cell replication. A simple model of this process assumes that the modified parental histones are distributed randomly onto daughter strands and are complemented by *de novo* synthesized unmodified histones (Probst *et al* 2009, Margueron and Reinberg 2010). On average this process leads to dilution of the modified histones per strand down to half their equilibrium value in the parent cell (i.e. $\theta_{HM} \rightarrow \theta_{HM}/2$ for parent→daughter). Inheritance requires that the original state can recover from these diluted states. This is always the case in monostable states. For bi-stable states our model predicts that this recovery into, for example, the OFF-state requires that associated fluctuations $\delta\theta_{HM}$ in both daughter cells are smaller than $\Delta\theta_{HM}^{OFF}$ because otherwise the system overleaps the instable fixpoint θ_{HM}^* and switches into the 'opposite' ON-state. As many histones are involved in the process at the same time, such perturbations can switch also states that are long term stable under the (de-)modification dynamics described above.

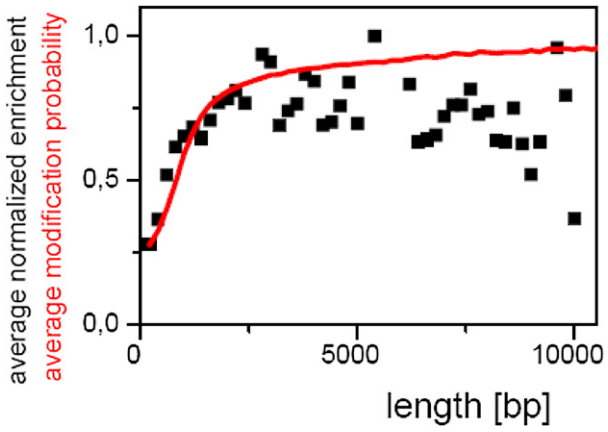

Figure A3. Length dependent modification probability. Average normalized enrichment of reads per nucleosome as a function of the length of modified regions (points), compared to the average modification probability in the model (red line), for H3K4me3 in ESCs. Three data points exceeding the background trend by a factor of 3 were disregarded as outliers.

Accordingly, regulatory OFF-states are not necessarily inherited in the range of bistable solutions, i.e. their life time is limited by the replication cycle. This can be seen in figure A2B providing the probability that changes in θ_{HM} due to histone dilution during replication are larger than $\Delta\theta_{HM}^{OFF}$ in at least one daughter cell. A comparable behaviour has also been described by David-Rus *et al* (2009) and by Micheelsen *et al* (2010).

A.4. On the length-dependent modification probability

Our model assumes a length-dependent modification probability of nucleosomes. To verify this assumption, we calculated average enrichment of reads per nucleosome for

H3K4me3 modified regions in ESCs in a length-dependent manner relative to the whole cell extract serving as reference (see 'Materials and Methods'). This normalization accounts for artefacts arising from the experimental procedure and subsequent analysis. As shown in figure A3, the so normalized average enrichment of reads per nucleosome, as a function of the length of the modified regions they are found in, correlates with the corresponding average modification probability, calculated by means of our switch model. We applied the spatial extension of our model using the respective parameter set for H3K4me3 in ESCs (see table 1). Long modified regions show a trend to higher enrichment per nucleosome than the shorter regions, thus supporting our model predictions. Comparable results are obtained for H3K27me3 (not shown).

References

Allen M D, Grummitt C G, Hilcenko C, Min S Y, Tonkin L M, Johnson C M, Freund S M, Bycroft M and Warren A J 2006 Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase *EMBO J.* **25** 4503–12

Bernstein B E, Meissner A and Lander E S 2007 The mammalian epigenome *Cell* **128** 669–81

Binder H, Wirth H and Galle J 2010 Gene expression density profiles characterize modes of genomic regulation: theory and experiment *J. Biotechnol.* **149** 98–114

Calvanese V, Lara E, Kahn A and Fraga M F 2009 The role of epigenetics in aging and age-related diseases *Ageing Res. Rev.* **8** 268–76

Chakalova L, Debrand E, Mitchell J A, Osborne C S and Fraser P 2005 Replication and transcription: shaping the landscape of the genome *Nature Rev. Genet.* **6** 669–77

Cuddapah S, Roh T Y, Cui K, Jose C C, Fuller M T, Zhao K and Chen X 2012 A novel human polycomb binding site acts as a functional polycomb response element in Drosophila *PLoS One* **7** e36365

Dahl J A, Reiner A H, Klungland A, Wakayama T and Collas P 2010 Histone H3 lysine 27 methylation asymmetry on developmentally-regulated promoters distinguish the first two lineages in mouse preimplantation embryos *PLoS One* **5** e9150

David-Rus D, Mukhopadhyay S, Lebowitz J L and Sengupta A M 2009 Inheritance of epigenetic chromatin silencing *J. Theor. Biol.* **258** 112–20

Deng W and Blobel G A 2010 Do chromatin loops provide epigenetic gene expression states? *Curr. Opin. Genet. Dev.* **20** 548–54

Dirk L M, Flynn E M, Dietzel K, Couture J F, Trievel R C and Houtz R L 2007 Kinetic manifestation of processivity during multiple methylations catalyzed by SET domain protein methyltransferases *Biochemistry* **46** 3905–15

Dodd I B, Micheelsen M A, Sneppen K and Thon G 2007 Theoretical analysis of epigenetic cell memory by nucleosome modification *Cell* **129** 813–22

Dodd I B and Sneppen K 2011 Barriers and silencers: a theoretical toolkit for control and containment of nucleosome-based epigenetic states *J. Mol. Biol.* **414** 624–37

Grewal S I and Elgin S C 2002 Heterochromatin: new possibilities for the inheritance of structure *Curr. Opin. Genet. Dev.* **12** 178–87

Hoffmann S, Otto C, Kurtz S, Sharma C M, Khaitovich P, Vogel J, Stadler P F and Hackermuller J 2009 Fast mapping of short sequences with mismatches, insertions and deletions using index structures *PLoS Comput. Biol.* **5** e1000502

Karlebach G and Shamir R 2008 Modelling and analysis of gene regulatory networks *Nature Rev. Mol. Cell. Biol.* **9** 770–80

Klose R J and Zhang Y 2007 Regulation of histone methylation by demethylimination and demethylation *Nature Rev. Mol. Cell. Biol.* **8** 307–18

Klymenko T, Papp B, Fischle W, Kocher T, Schelder M, Fritsch C, Wild B, Wilm M and Muller J 2006 A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities *Genes Dev.* **20** 1110–22

Kohler C and Villar C B 2008 Programming of gene expression by Polycomb group proteins *Trends Cell. Biol.* **18** 236–43

Kota S K and Feil R 2010 Epigenetic transitions in germ cell development and meiosis *Dev. Cell* **19** 675–86

Kouzarides T 2007 Chromatin modifications and their function *Cell* **128** 693–705

Lee T I *et al* 2006 Control of developmental regulators by Polycomb in human embryonic stem cells *Cell* **125** 301–13

Lunyak V V 2008 Boundaries. Boundaries . . . Boundaries??? *Curr. Opin. Cell. Biol.* **20** 281–7

Margueron R and Reinberg D 2010 Chromatin structure and the inheritance of epigenetic information *Nature Rev. Genet.* **11** 285–96

Meissner A *et al* 2008 Genome-scale DNA methylation maps of pluripotent and differentiated cells *Nature* **454** 766–70

Mendenhall E M, Koche R P, Truong T, Zhou V W, Issac B, Chi A S, Ku M and Bernstein B E 2010 GC-rich sequence elements recruit PRC2 in mammalian ES cells *PLoS Genet.* **6** e1001244

Micheelsen M A, Mitarai N, Sneppen K and Dodd I B 2010 Theory for the stability and regulation of epigenetic landscapes *Phys. Biol.* **7** 026010

Mikkelsen T S *et al* 2007 Genome-wide maps of chromatin state in pluripotent and lineage-committed cells *Nature* **448** 553–60

Mohammad H P and Baylin S B 2010 Linking cell signaling and the epigenetic machinery *Nature Biotechnol.* **28** 1033–8

Mohn F and Schubeler D 2009 Genetics and epigenetics: stability and plasticity during cellular differentiation *Trends Genet.* **25** 129–36

Oguro H, Yuan J, Ichikawa H, Ikawa T, Yamazaki S, Kawamoto H, Nakauchi H and Iwama A 2010 Poised lineage specification in multipotential hematopoietic stem and progenitor cells by the polycomb protein Bmi1 *Cell Stem Cell* **6** 279–86

Ooi S K *et al* 2007 DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA *Nature* **448** 714–7

Orlando V 2003 Polycomb, epigenomes, and control of cell identity *Cell* **112** 599–606

Pan G, Tian S, Nie J, Yang C, Ruotti V, Wei H, Jonsdottir G A, Stewart R and Thomson J A 2007 Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells *Cell Stem Cell* **1** 299–312

Patel A, Dharmarajan V, Vought V E and Cosgrove M S 2009 On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex *J. Biol. Chem.* **284** 24242–56

Pedersen M T and Helin K 2010 Histone demethylases in development and disease *Trends Cell. Biol.* **20** 662–71

Pietersen A M and Van Lohuizen M 2008 Stem cell regulation by polycomb repressors: postponing commitment *Curr. Opin. Cell. Biol.* **20** 201–7

Probst A V, Dunleavy E and Almouzni G 2009 Epigenetic inheritance during the cell cycle *Nature Rev. Mol. Cell. Biol.* **10** 192–206

Prohaska S J, Stadler P F and Krakauer D C 2010 Innovation in gene regulation: the case of chromatin computation *J. Theor. Biol.* **265** 27–44

Przybilla J, Galle J and Rohlf T 2012 Is adult stem cell ageing driven by conflicting modes of chromatin remodeling? *Bioessays* **34** 841–8

Reik W 2007 Stability and flexibility of epigenetic gene regulation in mammalian development *Nature* **447** 425–32

Ringrose L and Paro R 2004 Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins *Annu. Rev. Genet.* **38** 413–43

Rohlf T, Steiner L, Przybilla J, Prohaska S, Binder H and Galle J 2012 Modeling the dynamic epigenome: from histone modifications towards self-organizing chromatin *Epigenomics* **4** 205–19

Rohlf T and Winkler C 2009 Emergent network structure, evolvable robustness, and nonlinear effects of point mutations in an artificial genome model *Adv. Complex Syst.* **12** 293–310

Schuettengruber B, Chourrout D, Vervoort M, Leblanc B and Cavalli G 2007 Genome regulation by polycomb and trithorax proteins *Cell* **128** 735–45

Schuettengruber B, Ganapathi M, Leblanc B, Portoso M, Jaschek R, Tolhuis B, Van Lohuizen M, Tanay A and Cavalli G 2009 Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos *PLoS Biol.* **7** e13

Schwartz Y B, Kahn T G, Nix D A, Li X Y, Bourgon R, Biggin M and Pirrotta V 2006 Genome-wide analysis of Polycomb targets in Drosophila melanogaster *Nature Genet* **38** 700–5

Sedighi M and Sengupta A M 2007 Epigenetic chromatin silencing: bistability and front propagation *Phys. Biol.* **4** 246–55

Simon J A and Kingston R E 2009 Mechanisms of polycomb gene silencing: knowns and unknowns *Nature Rev. Mol. Cell. Biol.* **10** 697–708

Thomson J P *et al* 2010 CpG islands influence chromatin structure via the CpG-binding protein Cfp1 *Nature* **464** 1082–6

Tiwari V K, Mcgarvey K M, Licchesi J D, Ohm J E, Herman J G, Schubeler D and Baylin S B 2008 PcG proteins, DNA methylation, and gene repression by chromatin looping *PLoS Biol.* **6** 2911–27

Tolhuis B, Blom M, Kerkhoven R M, Pagie L, Teunissen H, Nieuwland M, Simonis M, De Laat W, Van Lohuizen M and Van Steensel B 2011 Interactions among Polycomb domains are guided by chromosome architecture *PLoS Genet.* **7** e1001343

Voigt P, Leroy G, Drury W J III, Zee B M, Son J, Beck D B, Young N L, Garcia B A and Reinberg D 2012 Asymmetrically modified nucleosomes *Cell* **151** 181–93

Wachsmuth M, Caudron-Herger M and Rippe K 2008 Genome organization: balancing stability and plasticity *Biochim. Biophys. Acta* **1783** 2061–79

Zee B M, Levin R S, Xu B, Leroy G, Wingreen N S and Garcia B A 2010 *In vivo* residue-specific histone methylation dynamics *J. Biol. Chem.* **285** 3341–50

Zhang X and Bruice T C 2008 Product specificity and mechanism of protein lysine methyltransferases: insights from the histone lysine methyltransferase SET8 *Biochemistry* **47** 6671–7