Transcriptional memory emerges from cooperative histone modifications

Hans Binder[1,2], Lydia Steiner[1,3], Thimo Rohlf[1], Sonja Prohaska[1,3], Jörg Galle[1§]

[1] Interdisciplinary Centre for Bioinformatics of Leipzig University, D-04107 Leipzig, Härtelstr. 16-18, Germany
[2] Leipzig Interdisciplinary Research Cluster of Genetic Factors, Clinical Phenotypes and Environment (LIFE); Universität 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

[§]Corresponding author

Email addresses:
HB: binder@izbi.uni-leipzig.de
LS: lydia@bioinf.uni-leipzig.de
TR: rohlf@izbi.uni-leipzig.de
SP: sonja@bioinf.uni-leipzig.de
JG: galle@izbi.uni-leipzig.de

Abstract

Background

Transcriptional regulation in cells makes use of diverse mechanisms to ensure that functional states can be maintained and adapted to variable environments; among them are chromatin-related mechanisms. While mathematical models of transcription factor networks controlling development are well established, models of transcriptional regulation by chromatin states are rather rare despite they appear to be a powerful regulatory mechanism.

Results

We here introduce a mathematical model of transcriptional regulation governed by histone modifications. This model describes binding of protein complexes to chromatin which are capable of reading and writing histone marks. Molecular interactions between these complexes and DNA or histones create a regulatory switch of transcriptional activity possessing a regulatory memory. The regulatory states of the switch depend on the activity of histone (de-) methylases, the structure of the DNA-binding regions of the complexes, and the number of histones contributing to binding.

We apply our model to transcriptional regulation by trithorax- and polycomb-complex binding. By analyzing data on pluripotent and lineage-committed cells we verify basic model assumptions and provide evidence for a positive effect of the length of the modified regions on the stability of the induced regulatory states and thus on the transcriptional memory.

Conclusions

Our results provide new insights into epigenetic modes of transcriptional regulation. Moreover, they implicate well-founded hypotheses on cooperative histone modifications, proliferation induced epigenetic changes and higher order folding of chromatin which await experimental validation. Our approach represents a basic step towards multi-scale models of transcriptional control during development and lineage specification.

Background

Regulation of the amount of transcript derived from a genomic locus makes use of diverse mechanisms to ensure that functional regulatory states can be maintained and adapted to variable environments. Such mechanisms can be based on binding of transcription factors forming cis-regulatory networks or can be related to chromatin reorganisation. Unravelling the interplay of different regulatory mechanisms represents a challenge to Molecular Systems Biology [1], in particular, since the functional role of the mechanisms changes in the course of the life cycle. For instance, during gametogenesis and early embryonic development modules of transcription factors networks regulate differentiation while the bulk of the epigenome is reset to default [2,3]. In subsequent stages new chromatin marks accumulate, e.g. via enzymatic activation of histone modifiers, and soon take control of the cell fate [4,5].

While mathematical models of transcription factor networks controlling development are a matter of intense debate [6,7], models of transcriptional regulation by chromatin states are rather rare [8-10]. Here, we introduce a model of chromatin-based transcriptional regulation for Metazoa. The model is motivated by the function of Polycomb group (PcG) and trithorax group (trxG) genes in heterochromatin and euchromatin formation. Both, PcG and trxG proteins form protein complexes, which are epigenetic readers and writers. Their binding to chromatin on one hand is governed by specific histone modifications and on the other hand modifies histones by attaching chemical groups [11]. These interactions contribute to transcriptional control during development and lineage specification [12,13].

In the absence of transcriptional induction, trxG and PcG complexes bind and remain attached to chromatin target loci and modulate the chromatin structure. Respectively, the complexes can induce, among others, trimethylation of lysin 4 (H3K4me3) and 27 (H3K27me3) of histone H3 [12]. These methylation reactions are catalyzed by SET domains of the trxG and PcG proteins which are responsible for methyltransferase activity in many proteins. The trimethylation marks at K4 and K27 were classified as activating and repressive marks, respectively, according to the overall effect of their occurrence on transcriptional regulation. Occurrence of both marks at a single histone is described as hallmark for embryonal stem cells (ESCs) [14]. Furthermore, histone methylation has been found to be reversible. Specific de-methlyases have been identified for both H3K4me3 and H3K27me3 [15,16].

Recent studies show a differential distribution of H3K4me3 chromatin mark, at CpG-rich versus CpG-poor promoters [17]. In ESCs nearly all genes with CpG-rich promotors are bound by trxG complexes but only a small subset of CpG-poor promoters [18]. There is increasing evidence that these complexes protect the CpG-motifs from becoming methylated and silenced. Consequently, most of these genes are actively transcribed [19]. Stable 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 promotors of mouse ESCs as reported by Mikkelsen et al. [18].

Despite the detailed knowledge about correlations of trxG/PcG complex binding with histone modifications and genomic loci the basal mechanism of recruitment to specific genomic loci is currently unknown. While direct binding of trxG and PcG proteins to DNA sequence motifs via DNA-binding domains have been favoured for a long time, only little evidence has been acquired. In Drosophila, Pho is the only sequence-specific DNA-binding polycomb group protein [20,21]. The absence of common binding sites and polycomb response elements in mammalian genomes suggest that different developmental strategies entailed different repressive mechanisms [22]. Several studies report a causal relationship between CpG-rich regions and recruitment of trxG- and the PcG-complexes in mammals. While trxG-complexes contain the CXXC1 protein including a CXXC domain capable of binding CpGs [23], the responsible factors for recruitment of polcomb repressive complex 2 [24] have not been determined yet. Specific recruitment of PcG complexes to genomic loci by interactions of PcG proteins with histone modifications, ncRNAs, and transiently associated transcription factors has been proposed [22,25].

In the following, we first introduce a mathematical model of epigenetic transcriptional regulation. This model describes the binding of protein complexes to chromatin. These complexes can read and write histone marks. Molecular interactions between the protein complex and the modified histones can create a memory of the regulatory state. This memory is in general heritable. Subsequently, we validate basic model assumptions analysing ChIP-seq data on H3K4me3 and H3K27me3 modifications in pluripotent and lineage-committed mouse cells [18]. Fitting our model to these data we provide evidence that the modification process throughout the genome is actually cooperative.

Results

Mathematical model of an epigenetic switch controlling gene activity

General assumptions

In this section we introduce a model of transcriptional regulation based on reversible protein complex binding 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 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 ICs associates with a response element (RE) which contains a variable number n_{BS} of binding sites (BS). Each binding site contributes to IC –BS association.

(ii) We assume a fixed RE length L^{RE}. Each RE is associated with N_H histones where n_{HM} of them are in a modified (HM), and the remaining N_H-n_{HM} are in a un-modified (H0) state. Modified histones facilitate the binding of free ICs to RE. Hence, free ICs are capable of 'reading' histone modifications.

(iii) On the other hand, ICs bound to a RE are capable to 'write' histone modifications by catalyzing the association reaction of the modifier with the histones. The inverse process, the histone de-modification reaction is independent of the presence of bound IC.

(iv) Each RE is associated with one (or several) genes. Binding of ICs to the RE 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 however in reverse direction.

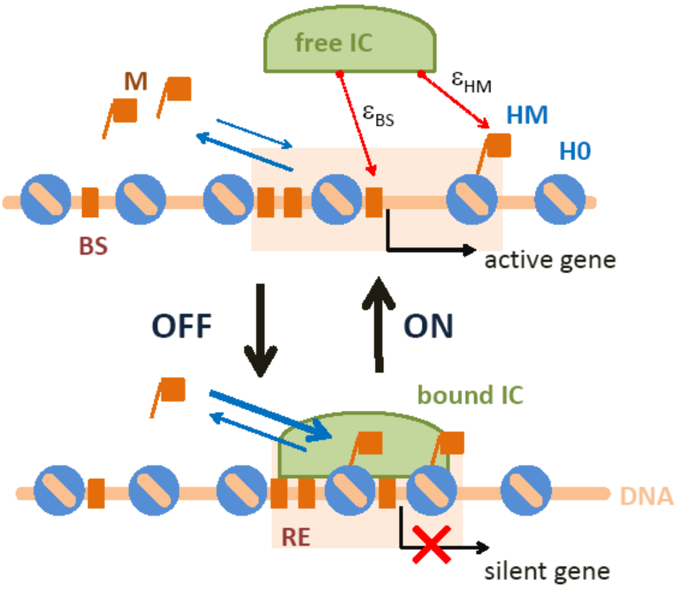

Figure 1: **Transcriptional regulation by histone modification.** Schematic plot of the epigenetic switch model: The transcription of a selected gene is active if the interaction complex (IC) is in a free, i.e. unbound state. Its reversible binding to a response element (RE) is facilitated by interactions with binding sites (BS) located within the RE and with modified histones (HM, unmodified histones are marked as H0). ε_{BS} and ε_{HM} are the free energy increments per BS and modified histone upon IC-binding, respectively. Bound ICs catalyze 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.

Binding equilibrium and modification reactions

We assume a binding equilibrium between free IC, free RE 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,

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

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 RE, i.e. RE occupancy is

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

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

$$\Delta g = \varepsilon_0 + n_{BS}\varepsilon_{BS} + n_{HM}\varepsilon_{HM}, \qquad (3)$$

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

The number of binding sites in each RE is a constant, n_{BS}, given by, e.g., the presence of specific sequence motifs. In contrast, the number of modified histones can vary due to the changing binding and de-binding activities of the modifiers. We consider a simple rate equation:

$$\frac{dn_{HM}}{dt} = -k_m^- n_{HM} + k_m^+ \cdot \Theta \cdot \left(N_H - n_{HM}\right) \qquad (4)$$

where k_m^- is the rate constant of de-modification $k_m^+ \cdot \Theta$ whereas defines the rate of modification. The latter is assumed to scale with the RE-occupancy ($0 \le \Theta \le 1$). 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. We assume stationary conditions for the modification dynamics ($dn_{HM}/dt=0$). In this case Eq. (4) provides the average fraction of modified histones per RE:

$$\theta_{HM} \equiv \frac{\bar{n}_{HM}}{N_H} = \frac{k_m^+ \cdot \Theta}{k_m^+ \cdot \Theta + k_m^-} = \frac{1}{1 + K_m / \Theta} \qquad (5)$$

where $K_m = k_m^- / k_m^+$ is the nominal equilibrium constant of histone de-modification at maximum RE-occupancy $\Theta=1$. Note that the equilibrium is governed by an effective constant, $K_m^{eff}=K_m/\Theta$ which inversely scales with Θ. The equilibrium value of the RE-occupancy used in Eq. (5) is a function of θ_{HM} given by Eqs. (2) and (3) with the substitution $\bar{n}_{HM} \equiv N_H \cdot \theta_{HM}$,

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

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

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

Eq. (7) has either one or three solutions depending on the parameter values chosen (see Figure 2). In case of one solution the system is monostable. Solutions may refer to small or large values of θ_{HM}. The respective θ_{HM}-values, in turn, transform into small and large RE-occupancy values, Θ (see Eq. (6)) and thus to predominantly active and silent genes, respectively. For short we will assign the respective solutions as 'ON' and 'OFF' with the definition $0 \le \Theta_{ON} < \Theta_{OFF} \le 1$.

In case Eq. (7) has three solutions the system is bi-stable. Thereby, the smallest and the largest solution for θ_{HM} (θ_{HM}^{OFF}, θ_{HM}^{ON}) define the stable solutions while the 'middle' one, θ_{HM}^{*}, is unstable. For $\theta_{HM}>\theta_{HM}^{*}$, the system always converges to the OFF state, whereas for $\theta_{HM}<\theta_{HM}^{*}$, it is attracted into the ON state (see Figure 2**)**. The feedback-to-noise ratio, $F=\Delta\theta_{HM}/\delta\theta_{HM}$ defines the relation between the minimum increment $\Delta\theta_{HM}$ of θ_{HM} that must be 'overleaped' by fluctuations of amplitude $\delta\theta_{HM}$ to switch between the ON and the OFF state. In the following we will assume the limiting case F→ ∞ referring to vanishing fluctuations.

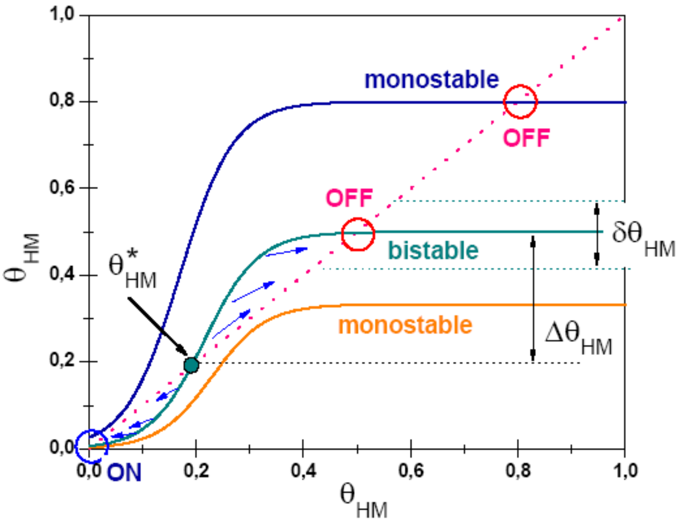

Figure 2: Graphical representation of solutions of Eq. (7). The solutions are given by the intersections of the curves (right hand side of the equation) with the diagonal line (left hand side). One gets either a bistable (ON and OFF) or monostable (ON or OFF) solution for θ_{HM} (circles). In the bistable case a third, unstable solution θ_{HM}^{*} exists at an intermediate position between the respective stable ON and OFF solutions. It defines the range of the attractors of the stable solutions (see text).

Switching regimes

The relations between the degree of histone modification, θ_{HM}, and its molecular determinants K_m and $n_{BS}\cdot\varepsilon_{BS}$ are further disentangled in Figure 3. Figure 3A shows the surface defined by solutions of Eq. 7 in dependence of K_m and $n_{BS}\varepsilon_{BS}$. Projections of the solutions onto the θ_{HM}-vs- K_m and θ_{HM}-vs- $n_{BS}\cdot\varepsilon_{BS}$ coordinate planes provide iso-BS ($n_{BS}\cdot\varepsilon_{BS}$=const) and iso-$K_m$ (K_m=const) curves. Examples are shown in Figure 3B,C.

Essentially two types of such iso-curves can be distinguished. The first type monotonically changes over the whole range of the respective control parameter. The second type inverses the slope in the intermediate part of the curves. The region enclosed between the two turning points defines unstable solutions of Eq. (7). 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 continuum of turning points defines the region of bi-stability (orange lines, see also section "Methods"). Following the iso-curves either in direction of increasing or decreasing argument gives rise to different positions of the turning points and thus to a hysteresis. In case of bi-stable behavior the state of the system consequently depends on its history and, particularly, whether the state variables K_m or $n_{BS}\cdot\varepsilon_{BS}$ are increased or decreased, i.e. the system provides a regulatory memory. Figure 3D shows phase diagram of the epigenetic switch. It was obtained via a projection of $g(\theta_{HM})$ into the $n_{BS}\varepsilon_{BS}$-vs-K_m- coordinate plane.

The regimes of gene activity of the epigenetic switch indicated in Figure 3A-D are defined by the occupancy of the REs by IC complexes Θ via Eq. (6). Figure 4 shows selected iso-curves in the Θ-vs-θ_{HM} coordinate system focusing on process control under iso-K_m (Figure 3A) and iso- $n_{BS}\cdot\varepsilon_{BS}$ (Figure 3B) 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.

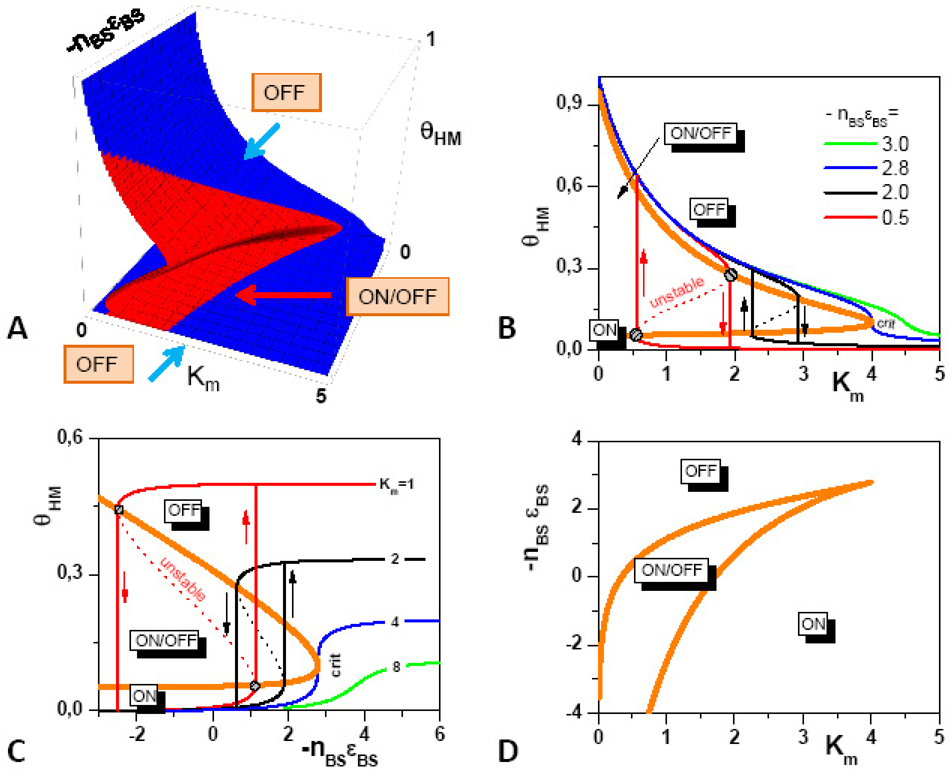

Figure 3: **Regimes of gene activity.** A): Fraction of modified histones θ_{HM} in dependence of the equilibrium constant of histone de-modification (K_m) and the total IC/BS interaction free energy per RE ($n_{BS}\cdot\varepsilon_{BS}$) as given by Eq. (7) ($\varepsilon_0$=5, $N_{HM}\varepsilon_{HM}$=20). 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 behavior (memory). D): Projection of the range of ON/OFF-bistability into the $n_{BS}\varepsilon_{BS}$-vs-K_m- coordinate plane, providing a phase diagram of the epigenetic switch.

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 percent). 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 Figure 4A and B reveal that this subregion refers to large values of $–n_{BS}\varepsilon_{BS}$ and small values of the de-modification constant K_m, respectively.

For such parameter values (see yellow areas in Fig. 4A and B) the system is dominated by the repulsion between the IC and the DNA. Although up to 100% of the histones 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.

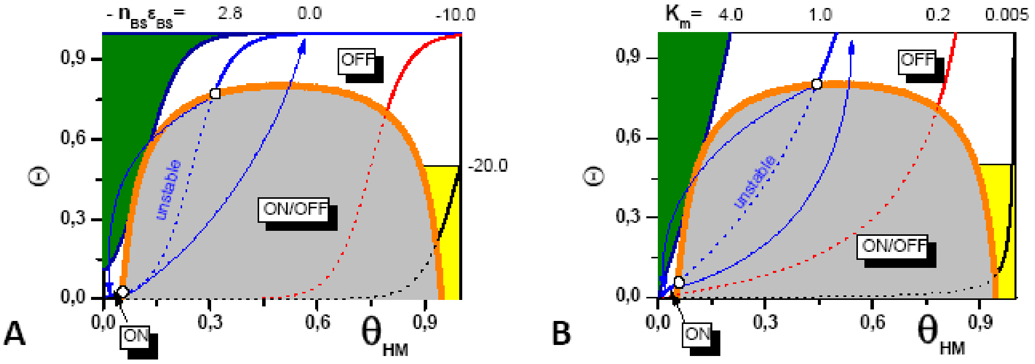

Figure 4: Phenotypic maps of the switch. The maps link the occupancy of the REs Θ with the degree of histone modification θ_{HM}. Panels (A) and (B) show iso-curves for $-n_{BS}\varepsilon_{BS}$ and K_m, respectively (ε_0=5, $N_{HM}\varepsilon_{HM}$=20). The ranges of ON/OFF bi-stability are shown by 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 behavior as shown in Figure 3B 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 text).

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 the equilibrium constant of histone de-modification K_m. Additionally, the switch behaviour depends on the maximum interaction free energy of the IC with the histones of a RE, $N_H\varepsilon_{HM}$. Figure 5A reveals that with increasing -$N_H\varepsilon_{HM}$ the range of bistability markedly expands to larger values of K_m. (see also section “Methods”, Eq. 10a). Changing -$N_H\varepsilon_{HM}$ while keeping all other systems parameter constant a hysteretic behavior can be observed as well (Figure 5B). Thus, the system can provide a regulatory memory also in the case that the length L^{RE} of the RE (here in terms of N_H) and/or the free energy of interaction between the IC and the histones ε_{HM} is varied.

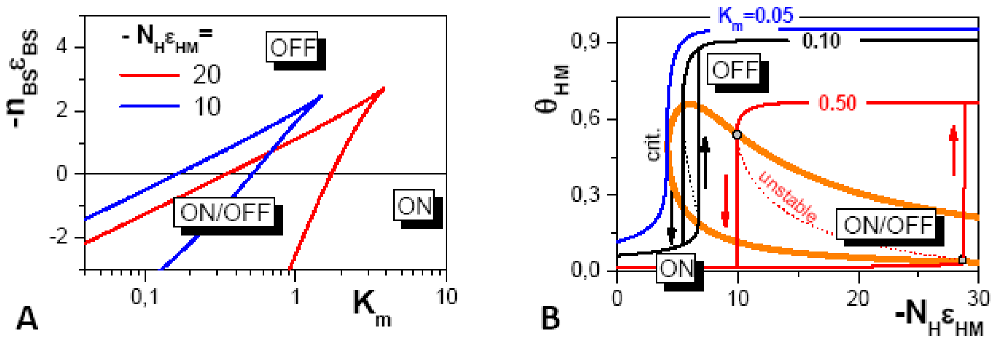

Figure 5: Length of cooperative units affects gene activity. A) Regions of ON/OFF bistability for -$N_{H}.\varepsilon_{HM}$=10 (blue) and 20 (red). Note that K_m is scaled logarithmically in contrast to Figure 3B,C. It scales in free energy units and thus it adequately matches the $n_{BS}\varepsilon_{BS}$ axis. B) Solutions of Equ. 7 for variable $N_H\varepsilon_{H,}$. Shown are iso-K_m trajectories assuming a constant value $n_{BS}\varepsilon_{BS}$=0. A minimal value K_m = 0.05 (blue curve) is required to surpass the critical value -$N_{H}.\varepsilon_{HM}$=4.2 below which no bistable dynamics occurs. Details as in Fig. 3 B,C.

Experimental indications

Maintenance, lineage commitment and differentiation of various stem cell systems are under epigenetic control [26,27]. A genome-wide combined ChIP-seq and gene expression microarray data set has been published by Mikkelsen et al. [18] addressing the state of histone modification and of gene activity in murine embryonic stem cells (ESCs), lineage-committed murine embryonic fibroblasts (MEFs) and neuronal progenitor cells (NPCs). We re-analyzed this data set to study whether properties of the observed histone modifications i) provide information about IC binding sites, lengths of REs and their genome-wide distribution and ii) would support our model results on a cooperative binding of ICs. While our model provides information about equilibrium states, biological data represent snapshots of dynamic systems. As a consequence, modified histones detected in the experiments may not be bound by ICs. This has to be considered for the analysis of binding properties [28], but will not affect the identification of REs. We here assume that lineage committed cells have converged to a stationary regulatory state characterized by a well defined binding equilibrium.

Identifying response elements

Our model predicts that an increasing number of binding sites per RE strengthens IC binding and facilitates the modification of the histones involved. The recruitment of ICs such as trxG and PcG complexes to a genomic locus has been suggested to depend on the local CpG-density. Actually it has been shown that histones tri-methylated at H3K4 preferentially associate with CpG-rich promoter regions [18].

In a first step we asked whether such association can be observed not only for promoters but for any CpG-rich region throughout the genome. After mapping the ChIP-seq data using the mapper 'segemehl' [29], we could infere H3K4me3 and H3K27me3 modified chromatin regions. Moreover, 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 section “Methods” for more details). Figure 6 relates CpG counts to the length of the respective regions. A constant CpG-density transforms into a line of positive slope in this plot. The ‘background’ CpG-density of 0.0052 CpGs/bp is provided by the green line in each of the figures. 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; [19]). Our analysis reveals a third, intermediate characteristic CpG-density of $\rho_H = 0.031$ CpGs/bp in DNA-regions associated with H3K4me3 modification in ESCs (blue lines in Figure 6), which is slightly below the lower limit for CpG islands (0.0375, [14]). It represents an upper bound of the CpG-density found in longer DNA-fragments (>1.5 kbp), which is about half as large as that near the CpG-rich promoters. Interestingly, the background branch remains virtually empty for H3K4me3 showing that all H3K4me3 are associated with CpG-rich DNA.

In contrast, H3K27me3 modified histones were found to associate not only with CpG-enriched regions, but also with regions of CpG-background density. This bimodal CpG-density distribution suggests two types of REs for H3K27me3 either requiring or not requiring an enriched CpG-density.

It is known that H3K4me3 and H3K27me3 are catalyzed by different ICs, namely the trxG- and the PcG-complexes, respectively. TrxG-complexes contain the CXXC1 protein including a CXXC domain capable of binding CpGs [23]. This explains the association of H3K4me3 with CpG-rich chromatin regions. In contrast PcG-complexes do not contain such binding domains. A possible explanation of the association of H3K27em3 with CpG-rich domains is provided by experimental findings suggesting that trxG-complexes recruit polycomb repressor proteins [30]. As a consequence PcG-complexes may associate with trxG complexes and associate only indirectly with CpG-rich DNA regions.

In the line of these arguments, one expects that the CpG-rich branch observed for H3K27me3 is reminiscent for bivalent, i.e. H3K4me3 and H3K27me3 modified regions. Indeed, after removal of such bivalent-modifications the respective CpG-rich branch nearly completely de-populates (Figure 6C). Interestingly many of the detected CpG-rich DNA-regions associated with bivalent modifications exceed 3kbp but only a few regions larger than 10kbp are identified. This suggests an upper limit of the number of modified histones of about 50 (assuming a DNA-length including linker DNA, of 200bp) per modified region, i.e. $N_H \leq 50$.

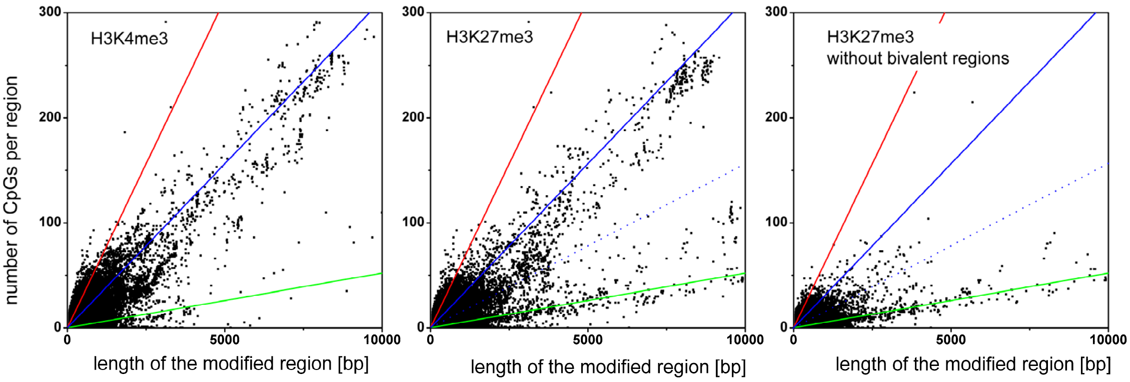

Figure 6: **Chromatin modifications in ESCs.** A) Number of CpG's in H3K4me3 modified chromatin regions of defined length. The modification is associated with CpG-rich regions with an about six-fold higher CpG-density ρ_H (blue line) than the background (green line). Typically the CpG-density at CpG-rich promotors (red line) is about twice that found for the enriched regions. B) H3K27me3 is associated both with CpG-rich regions and regions showing a CpG-content characteristic for the background. C) Non-bivalent H3K27me3 regions, i.e. regions with H3K27me3, but without H3K4me3 modifications, show no enrichment for CpGs. The dotted lines refer to a CpG-density of $\rho_H/2$ which was considered as a threshold density for CpG-rich regions (section 3.2).

In summary, our data analysis provides evidence that local CpG density, and hence the number of potential DNA binding sites in REs, is positively correlated with histone modification rate, as predicted by our model. A more detailed analysis of the length distribution of modified regions is carried out in the following section.

Cooperative binding of interaction complexes

Our model establishes a novel mechanism for cooperative control of histone states, mediated through a feedback loop established by ICs acting both as readers and modifiers of histone methylation states. The strength of IC binding increases with the number of modified histones, which in turn increases the modification rate and eventually the stability of modified states. As a consequence, larger modified chromatin regions are more stable against changes of the regulatory environment than shorter ones. Hence differentiation and lineage commitment are expected to change the length distribution of modified regions.

To prove this expectation we calculated the length distribution of modified chromatin regions for H3K4me3 and H3K27me3 in ESC, MEF and NPC (Figure 7). In case of

H3K27me3 we considered only CpG-rich regions with a CpG-density larger than $\rho_H/2$ (see Figure 6). Regarding the NPC and MEF data only those regions are taken into account, which are enriched in ESCs as well. In other words, we focused on the length distribution directly referring to changes associated with lineage-commitment of ESCs into either NPCs or MEFs. It turned out that the number of modified chromatin regions strongly decays with their length L_{mod} in all systems studied. The linear decays for $L_{mod}>500$ in the double logarithmic plot indicate that the tail of the frequency distributions can be approximated by a power law.

The theoretical curves in Figure 7 are calculated using a spatially extended version of our model assuming a geometrically decaying distribution of the number of cooperatively acting histones N_H and independent modification of adjacent histones. This model well describes the overall decay of the length distribution of modified chromatin regions (see Fig. 7A, B). Thereby, an increase of the de-modification constant K_m systematically reproduces the trends for MEF and NPC compared to ESCs. As the de-methylation constant K_m is governed by the activity of the respective de-methylases and/or methylases we suggest their changes to associate with changes of the expression degree of genes decoding these enzymes.

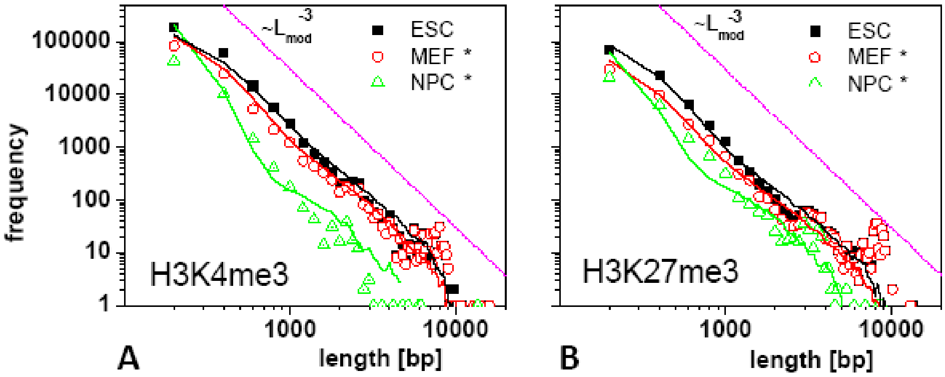

Figure 7: Changes of histone modifications during lineage commitment. A), B) Length distribution of chromatin regions in pluripotent ESC and lineage-committed MEF and NPC cells modified by H3K4me3 and H3K27me3. Considered are only CpG-rich regions which are modified in ESCs (MEF*, NPC*). Lineage commitment results in a decreased number of modified regions. While in MEF short regions are preferentially affected, in NPC also long regions become de-modified. Data points refer to the experimental distributions, while lined curves through the data points are theoretical distributions. The differences observed between ESC, MEF and NPC are described by changing K_m only: K_m = 0.037 (ESC), 0.048 (MEF), 0.179 (NPC) for H3K4me3; and K_m = 0.037 (ESC), 0.046 (MEF), 0.136 (NPC) for H3K27me3. The full set of parameters and details of the fitting procedure can be found in section 5.3. The decaying lines are power laws.

Applying the spatial extension of our model as described above, optimal fits of the experimental length distributions of modified regions (setting: $\varepsilon_0 = 5$) yielded values of $n_{BS}\varepsilon_{BS}$ below 1. This suggests that sequence specific recruitment of the ICs is driven by relatively weak interactions with the DNA. Moreover, we obtained values of about 0.1 for K_m. Accordingly, the H3K4me3 and H3K27me3 modification states of individual REs are approximately described by the black iso-K_m trajectories (K_m=0.1, $n_{BS}\varepsilon_{BS}$=0) in Figure 5B. For ε_{HM} our fits provided values of about 0.15 for both the H3K4me3 and the H3K27me3 modification. Thus, the critical value of $N_H\varepsilon_{HM}$ (~4) below which no bistable behavior can be observed (see section "Methods", Eq. 10a) refers to lengths of the RE of about 6kbp. This result suggests that only the longest REs are capable to create a memory effect.

In summary, our results demonstrate a positive effect of the length of the modified regions on the stability of the modifications. Increasing the de-modification activity consequently de-stabilizes first short regions and only subsequently larger regions. In the framework of our model this behavior can be explained by cooperative modes of histone modifications for both H3K4me3 and H3K27me3. Moreover, the spatially extended version of our model on randomly generated genomes predicts that IC binding is mainly governed by the capabilities of the ICs to read and write histone marks, while interaction with DNA play a minor role. Thus, the length of the RE and the de-modification constant K_m largely determine the regulatory state of the associated genes.

Discussion

Chromatin modifications can not only maintain regulatory states but 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 behavior that establishes a molecular memory. There is growing evidence that this kind of regulatory switches underlies many fate decisions in development and stem cell differentiation [31].

The states of the regulatory switch and the respective pattern of gene expression are governed by complex molecular interactions. In our simple model the regimes of gene activity are determined by the interaction free energy of the ICs with DNA binding sites $n_{BS}\cdot\varepsilon_{BS}$ and histones $N_H\cdot\varepsilon_{HM}$ as well as by the equilibrium constant of histone de-modification K_m. Accordingly, for fixed properties of the interaction with histones ($N_H\varepsilon_{HM}$), the state of the switch can be modulated by two complementary modes, either by varying the number and strength of binding sites per RE or by altering the activity of histone modifying enzymes in terms of the constant of histone de-modification. In the following, we will briefly discuss selected aspects of our model related to the parameter settings that ensure functionality as an epigenetic switch, the time scales of regulation, the impact of fluctuations and non-local cooperative effects.

Model parameters associated to functional switch behaviour

In our model the free energy term ε_0 depends on the concentration of the ICs and comprises all the interactions related to non-specific IC-RE association. It increases with decreasing IC-concentration and with increasing steric repulsion. This parameter was set to 5kT to ensure that spontaneous IC binding without the presence of specific binding sites is virtually impossible. In fact, recruitment of IC is governed effectively by the difference (ε_0 - n_{BS} ε_{BS}). The energy term n_{BS} ε_{BS} was chosen in the same range as ε_0. Accordingly, IC binding to RE alone allows only a weak association and the regulatory capacity is modulated actually by association with chromatin modifications. This is in contrast to relatively strong binding of typical transcription factors to DNA with effective binding energies to DNA in the range of 14 kT [32].

The term $N_H\varepsilon_{HM}$, describing the maximal free energy contribution of IC-histone interactions, was set to 10 to 20 kT which refers to about ε_{HM} = 0.2 – 0.4 kT per histone considering N_H=50. Actually, the fits of the length distributions suggest even smaller values for ε_{HM} between 0.1 to 0.2 kT. Accordingly binding to individual histones seems to be relatively weak which implies fast modulations of IC binding in agreement with experimental findings [33]. The total free energy at maximum occupancy of a RE has to be strongly negative to guarantee function of the regulatory switch. Our model suggests that in case of trxG- and PcG-omplex binding this requires long REs with a large number of associated histones (N_H>30). As we discuss in section below, it is unrealistic to assume that a single IC can interact with such an extended region. Hence our results may be indicative of cooperative behavior of ensembles of consecutive ICs.

Time scales of regulation

The number n_{BS} and strength ε_{BS} of DNA binding sites are linked to DNA sequence which typically adjusts by mutational mechanisms on an evolutionary time scale. Consequently, they can be considered invariant during processes like development and differentiation. The same time scale usually applies to changes of the structure and composition of the interaction complexes. However, the variation of $n_{BS}\cdot\varepsilon_{BS}$ due to changes in the composition of the IC may provide also an option for transcriptional regulation on short times scales. These complexes are composed of several molecular building blocks and thus alternative compositions of the IC can be functional as well [34]. Accordingly, monitoring changes of the composition of the interaction complexes will be of particular interest in understanding the systems behavior.

The equilibrium constant of histone de-modification K_m depends on the local abundance of the respective (de)-modifying enzymes which are mostly under transcriptional and post-transcriptional control. This enables variation on the time scale of development and differentiation. The activity of the modifying enzymes is mainly determined by their abundance in the medium surrounding the DNA. One particular value of K_m will consequently apply to several REs located for example in the same functional compartment of the cell nucleus. Thus, beside the composition of the interacting complexes, the spatial distribution of the modifying enzymes will be of interest for future extensions of our model. The length of REs and thus the number N_H of associated histones has been suggested to self-organize along boundaries between domains of competitive histone marks [35]. Accordingly, N_H could vary depending on the regulatory environment. We here assumed a fixed length and thus a fixed N_H. This can be motivated by the idea of a sequence specific binding of a barrier insulator [35] (see also section “Methods”). In general, variation of N_H may contribute to regulation of the binding equilibrium on both the time scales of differentiation and evolution.

The impact of fluctuations

A stationary chromatin structure, as described by the RE/IC binding equilibrium given by Eq. (7), requires that fluctuations of θ_{HM} occurring in processes of reading and writing histone modifications are small. This can be expected if the number of cooperatively acting histones per RE and thus its length is large. Hence, we predict that the length of the regulated chromatin region has a clear impact on the stability of the regulatory states of the switch.

In our model we neglect fluctuations of the binding energy and consequently of the degree of histone modification assuming infinite values of the feedback-to-noise ratio F. Under these conditions the hysteretic behaviour of the switch is most pronounced. With decreasing F, i.e. with increasing fluctuations, spontaneous transitions between ON and OFF states in the bistable region become relevant. At small F one expects a transition at a defined 'melting point' which resembles a first-order phase transition such as melting in a solid-liquid system. The fluctuations of the molecular interactions and local concentrations that lead to fluctuations of the degree of histone modification take here the role of thermal noise leading to density fluctuations at solid-liquid phase coexistence.

A process which unavoidably induces large fluctuations of the degree of histone modification is proliferation. These fluctuations originate in mechanisms of the inheritance of histone marks during cell division. In order to conserve the regulatory state of the parent cell its histone modification state has to be hand down to the daughter cells. A simple model of this process assumes that the modified parental histones are distributed randomly onto daughter strands and are complemented by *de novo* synthesised unmodified histones [36,37]. On average this process leads to the 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 be recovered from these diluted states (i.e. $\theta_{HM}/2 \rightarrow \theta_{HM}$). Our model predicts that this recovery will be successful under defined conditions only. It requires that the diluted states of both daughter cells are part of the attractor basin of the modification state of the parent cell. Monostable parent states always meet this condition, whereas bistable parent states do not. In the latter case, the diluted states of a high modification parent state can be part of the attractor basin of the corresponding low modification state and consequently will relax into this state (Figure 8A). Accordingly, in the range of bistable solutions regulatory OFF- states are not necessarily inherited. A comparable behaviour has been described also by Micheelsen et al. [10].

Non-local effects in cooperative dynamics

Early models for nucleosome modification, including histone modifications invoke a linear stepwise process, where a modified nucleosome stimulates the modification of its nearest neighbors [38]. Nucleosome modifications were thus envisioned to spread in a continuous fashion along the DNA. Alternatively, Dodd et al. [8] assume that in the recruitment reaction 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 RE are potentially affected by modification reactions without explicitly assuming a propagation mechanism for modification.

We found rather extended genomic regions associated with modified H3K4me3 and H3K27me3 of a length up to about 10kbp. According to our model, this result suggests that one IC interacts with to up to 50 histones which appears to be very unlikely for individual trxG and PcG complexes. These extended regions contain up to about 300 CpGs which are rather equally distributed throughout the regions. We therefore assume that adjacent REs form cooperative units that enable modifications of extended genomic regions; for example via chromatin looping as proposed by Tiwari et al. [28].

Our model allows to consider this scenario assuming that IC binding to several adjacent REs is affected by all the histones associated with them. In this case Eq. 4 describes the modification of N_H histones interacting with an ensemble of ICs. A sketch of such a hypothetical chromatin region after cooperative binding of ICs is shown in Figure 8B. The region is folded into multiple loops of a length of a few kbp, each of them containing a subset of the REs. Thereby, the adjacent REs cluster into a relatively small volume (dashed area).

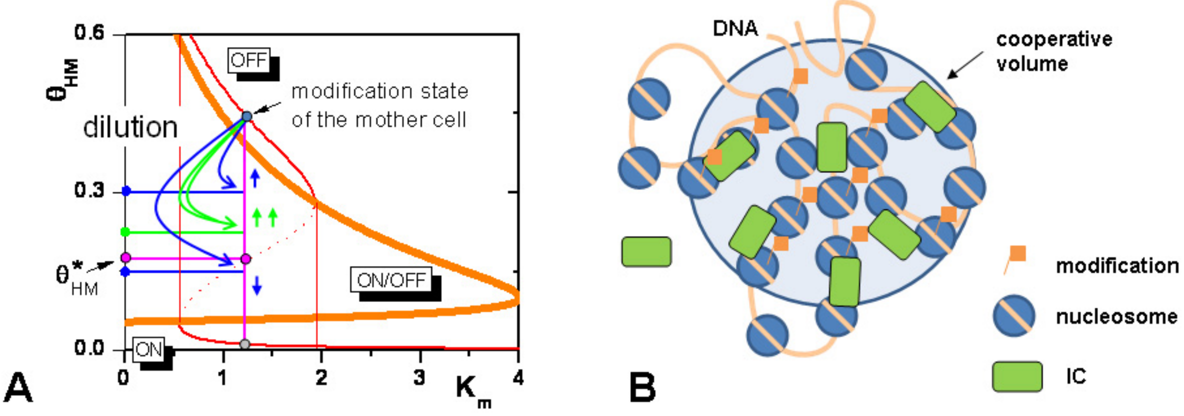

Figure 8: Inheritance of regulatory states and spatial organization of the epigenetic switch. A) Model of limited inheritance of histone modification states. Upon cell division the split of the parental histones onto the daughter strands decreases the fraction of modified histones per RE (dilution). Accordingly OFF-states can switch into ON-states if the fraction of modified histones reaches values below the instable fix point θ^*_{HM} . Shown are examples of symmetric (1:1, green) and asymmetric (2:1, blue) splitting. In the symmetric case the diluted states θ_{HM} are both part of the attractor basin of the mother state (θ_{HM}>θ_{HM}*), whereas in the asymmetric case one of the diluted states is not. This leads to a histone de-modification on the respective strand. B) Hypothetical chromatin structure of an epigenetic switch. Adjacent REs cluster in a cooperative volume. Binding of individual IC to DNA is improved in this volume by the presence of modified histones.

This generalization of the model does not require a change of the mathematical description of the switch. However, it implies non-local interactions between the IC and the modified histones, e.g. mediated through a particular ‘biochemical milieu’ created by the histones [39]. The interpretation raises questions about mechanisms of chromatin compartmentalization that define REs [35] and about the general relevance of chromatin looping for gene expression [40]. Recent experimental findings on frequent long range interactions among H3K27me3 modified domains support the idea of higher order folding of chromatin [41].

While our epigenetic switch model provides new insights into an essential epigenetic layer of transcriptional regulation it does not address any cross-talk with other regulatory layers. Thus, in a next step we plan integrating the model into the framework of an Artifical Genome Model [32,42]. This will provide a straight forward opportunity to study cross-talk between cis-regulatory networks of transcription factors and transcriptional regulation by histone modifications as explored here. Thereby, we will consider the combinatorics of different activating and repressing histone modifications [5]. Currently, the evolution of a versatile 'chromatin code' is still an intriguing hypothesis. Unraveling the syntax of the 'chromatin language' represents an important step towards realistic multi-scale models of transcriptional regulation.

Methods

Data analysis

ChIP-seq data: We re-analysed the data set provided by Mikkelsen et al. (GSE12241: H3K27me3, H3K4me3, H3K9me3 and whole cell extract of ESC, MEF and NPC cells as well as H3 of ESC cells) in order to enable some general comparisons with our model. The read data were mapped against the UCSC mouse genome version mm9 using the in-house tool 'segemehl' [29]. 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 100nt (about one half of the average nucleosome spacing) into modified regions.

In order to count the number of CpGs in each modified region, we simply counted 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 of the same size distribution as that of the modified regions was calculated as ‘background’ CpG-density. This approach accounts for size dependent variations of the CpG-density.

Calculating the range of bistability

The onset-points of ON/OFF bi-stability are given by curves meeting the maximum and minimum condition of the respective iso-curves in Figure 3 B and C. The two solutions of $\theta_{HM}^{(1,2)}$ referring to the upper and lower branches of the ON/OFF-range are obtained in two steps. Firstly, we rearranged Eq. (7) according to:

$$n_{BS} \cdot \varepsilon_{BS} = \ln\left[\frac{1-(1+K_m)\cdot\theta_{HM}}{K_m \cdot \theta_{HM}}\right] - (\varepsilon_0 + N_H \cdot \theta_{HM} \cdot \varepsilon_{HM}) \qquad (9a)$$

and

$$K_m = \frac{(1-\theta_{HM})}{\theta_{HM} \cdot (1+\exp(\varepsilon_0 + n_{BS}\varepsilon_{BS} + N_H \varepsilon_{HM} \cdot \theta_{HM}))}, \qquad (9b)$$

respectively. Secondly, we took the first derivative of Eq. (9a,b) with respect to θ_{HM} and calculated the roots:

$$\theta_{HM}^{(1,2)}(K_m) = \frac{1}{2(1+K_m)}\left[1 \mp \sqrt{\frac{K_m^{crit} - K_m}{K_m^{crit} + 1}}\right] \quad \text{and} \quad K_m^{crit} \equiv \frac{N_H \varepsilon_{HM}}{4} - 1 \qquad (10a)$$

and

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

(10b)

either analytically (Eq. 10a) or solving Eq. 10b numerically.

The onset-points of ON/OFF bi-stability shown in Figure 5B were calculated analogously.

Model of the length distribution of modified regions

In the following, we will outline the model which has been applied to generate the theoretical probability $w(L_{mod})$ to find modified chromatin regions of length L_{mod} throughout the genome (compare Fig. 7).

1) The number of histones N_H defines the length of the DNA sequence L^{RE} of the respective RE under control of a given IC (L^{RE} = 200 N_H bp). We further assume that adjacent REs are separated by so-called barrier insulators [35] representing short sequence motifs of length l_0. To estimate the probability distribution of L^{RE} 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 random genome model [32,42]. Accordingly, specific insulator sequences of length l_0 occur with probability 4^{-l_0} and the probability that two neighbouring insulators are separated by a sequence of length L^{RE} is given by the geometric distribution:

$$p(L^{RE}) = 4^{-l_0}(1 - 4^{-l_0})^{L^{RE}} \qquad (11)$$

We have chosen $l_0 = 5$ referring to an average L^{RE} equal to 1024.

Based on the distribution $p(L^{RE})$, we generated a random "histone strings" of $N_H(i) = L^{RE}(i)/200$ histones in a row (i is the running index of the RE) that can potentially be modified after binding of an interaction complex. Each histone j=1...$N_H(i)$ in RE no. i carries a binary variable:

$$h_{ij} = \begin{cases} 1 & \textit{with probability} \quad \theta_{HM}^{i} \\ 0 & \textit{with probability} \quad 1 - \theta_{HM}^{i} \end{cases} \qquad (12)$$

where $\theta_{HM}^{i} = \theta_{HM}(\varepsilon_0, \varepsilon_{BS}, \varepsilon_{HM}, n_{BS}(i), N_H(i))$ is the mean fraction of modified histones given by the solution of Eq. (7) for RE no. i. h_{ij} consequently characterizes the modification state of all histones considered. The number of binding sites in region i scales with their length, as suggested by the conserved CpG densities in Fig 6. We set an upper cut-off limit of $n_{BS,max}$ = 5 because ICs are likely to be limited in the number of binding sites they can bind simultaneously.

3) $w(L_{mod})$ was obtained by counting the number of adjacent histones with $h_{ij}=1$, irrespectively of their membership to the same or to different RE and subsequent normalization with respect to the total number of modified histones. The length of each modified chromatin region L_{mod} is simply given by the scaling factor 200 as explained above. Finally, $w(L_{mod})$ was calculated as mean averaged over hundred independent genome realizations.

4) We vary ε_{BS}, ε_{HM} and K_m while fixing $\varepsilon_0 = 5$ and $p(L^{RE})$ to find solutions which best fit the experimental distributions by minimizing the sum of squared residuals (SSR) between the experimental and theoretical distributions:

$$SSR = \frac{1}{L^{max}} \sum_{L_{mod}=1}^{L^{max}} \left(\log\left[w_{exp}\left(L_{mod}\right)+\delta\right]-\log\left[w_{theo}\left(L_{mod}\right)+\delta\right]\right)^2 \qquad (13)$$

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 $w = 0$. The optimum fit parameters are listed in Table 1.

Authors' contributions

Wrote the paper: HB SP JG. Essential ideas regarding the model: HB JG. Spatial extension of the model: TR. Analysis of ChIP-seq data: LS. Supervising biological issues: SP.

Acknowledgements

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 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).

References

1. Mohammad HP, Baylin SB: **Linking cell signaling and the epigenetic machinery.** *Nat Biotechnol* 2010, **28**: 1033-1038.
2. Reik W: **Stability and flexibility of epigenetic gene regulation in mammalian development.** *Nature* 2007, **447**: 425-432.
3. Kota SK, Feil R: **Epigenetic transitions in germ cell development and meiosis.** *Dev Cell* 2010, **19**: 675-686.
4. Dahl JA, Reiner AH, Klungland A, Wakayama T, Collas P: **Histone H3 lysine 27 methylation asymmetry on developmentally-regulated promoters distinguish the first two lineages in mouse preimplantation embryos.** *PLoS One* 2010, **5**: 9150.
5. Prohaska SJ, Stadler PF, Krakauer DC: **Innovation in gene regulation: the case of chromatin computation.** *J Theor Biol* 2010, **265**: 27-44.
6. Karlebach G, Shamir R: **Modelling and analysis of gene regulatory networks.** *Nat Rev Mol Cell Biol* 2008, **9**: 770-780.
7. Ben-Tabou de-Leon S, Davidson EH: **Modeling the dynamics of transcriptional gene regulatory networks for animal development.** *Dev Biol* 2009, **325**: 317-328.
8. Dodd IB, Micheelsen MA, Sneppen K, Thon G: **Theoretical Analysis of Epigenetic Cell Memory by Nucleosome Modification.** *Cell* 2007, **129**: 813-822.
9. Sneppen K, Micheelsen MA, Dodd IB: **Ultrasensitive gene regulation by positive feedback loops in nucleosome modification.** *Mol Syst Biol* 2008, **4**: e182.
10. Micheelsen MA, Mitarai N, Sneppen K, Dodd IB: **Theory for the stability and regulation of epigenetic landscapes.** *Phys Biol* 2010, **7**: 026010.
11. Ringrose L, Paro R: **Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins.** *Annu Rev Genet* 2004, **38**: 413-443.
12. Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G: **Genome Regulation by Polycomb and Trithorax Proteins.** *Cell* 2007, **128**: 735-745.
13. Orlando V: **Polycomb, epigenomes, and control of cell identity.** *Cell* 2003, **112**: 599-606.
14. Bernstein BE, Meissner A, Lander ES: **The mammalian epigenome.** Cell 2007, **128**:669-81.
15. Pedersen MT, Helin K: **Histone demethylases in development and disease.** *Trends Cell Biol* 2010, **20**: 662-671.
16. Klose JR, Zhan Y: **Regulation of histone methylation by demethylimination and demethylation.** Nat. Rev. Mol. Cell Biol. 2007, **8**: 307-318.
17. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES: **Genome-scale DNA methylation maps of pluripotent and differentiated cells.** *Nature* 2008, **454**: 766-770.
18. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim T-K, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE: **Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.** *Nature* 2007, **448**: 553-560

19. Mohn F, Schübeler D: **Genetics and epigenetics: stability and plasticity during cellular differentiation.** *Trends Genet* 2009, **25**: 129-136.
20. Schuettengruber B, Ganapathi M, Leblanc B, Portoso M, Jaschek R, Tolhuis B, van Lohuizen M, Tanay A, Cavalli G: **Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos.** *PLoS Biol.* 2009, **7**: e13.
21. Klymenko T, Papp B, Fischle W, Köcher T, Schelder M, Fritsch C, Wild B, Wilm M, Müller J: **A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities.** *Genes Dev* 2006, **20**: 1110-1122.
22. Köhler C, Villar CB: **Programming of gene expression by Polycomb group proteins.** *Trends Cell Biol* 2008, **18**: 236-243.
23. Allen MD, Grummitt CG, Hilcenko C, Min SY, Tonkin LM, Johnson CM, Freund SM, Bycroft M, Warren AJ: **Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase.** *EMBO J* 2006, **25**: 4503-4512.
24. Mendenhall EM, Koche RP, Truong T, Zhou VW, Issac B, Chi AS, Ku M, Bernstein BE: **GC-rich sequence elements recruit PRC2 in mammalian ES cells.** *PLoS Genet* 2010, **6**: e1001244.
25. Simon JA, Kingston RE: **Mechanisms of polycomb gene silencing: knowns and unknowns.** *Nat Rev Mol Cell Biol* 2010, **10**: 697-708.
26. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, Koseki H, Fuchikami T, Abe K, Murray HL, Zucker JP, Yuan B, Bell GW, Herbolsheimer E, Hannet NM, Sun K, Odom DT, Otte AP, Volkert TL, Bartel DP, Melton DA, Gifford DK, Jaenisch R, Young RA: **Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells.** *Cell* 2006, **125**: 301-313.
27. Oguro H, Yuan J, Ichikawa H, Ikawa T, Yamazaki S, Kawamoto H, Nakauchi H, Iwama A: **Poised lineage specification in multipotential hematopoietic stem and progenitor cells by the polycomb protein Bmi1.** *Cell Stem Cell.* 2010, **6**: 279-286.
28. Tiwari VK, McGarvey KM, Licchesi JDF, Ohm JE, Herman JG, Schübeler D, Baylin SB: **PcG Proteins, DNA Methylation, and Gene Repression by Chromatin Looping.** *PLoS Biol* 2008, **6**: 306.
29. Hoffmann S, Otto C, Kurtz S, Sharma CM, Khaitovich P, Vogel J, Stadler PF, Hackermüller J: **Fast mapping of short sequences with mismatches, insertions and deletions using index structures.** *PLoS Comput Biol* 2009, **5**: e1000502.
30. Xia ZB, Anderson M, Diaz MO, Zeleznik-Le NJ: **MLL repression domain interacts with histone deacetylases, the polycomb group proteins HPC2 and BMI-1, and the corepressor C-terminal-binding protein.** *PNAS* 2003, **100**: 8342-8347.
31. Pietersen AM, van Lohuizen M: **Stem cell regulation by polyomb repressors: postponing commitment.** *Curr. Op. Cell Biol.* 2008, **20**: 201-207.
32. Binder H, Wirth H, Galle J: **Gene expression density profiles characterize modes of genomic regulation: theory and experiment.** *J Biotechnol* 2010, **149**: 98-114.
33. Ficz G, Heintzmann R, Arndt-Jovin DJ: **Polycomb group protein complexes exchange rapidly in living Drosophila.** *Development* 2005, **132**: 3963-3976.

34. Kouzarides T: **Chromatin modifications and their function.** *Cell* 2007, **128**: 693-705.
35. Lunyak VV: **Boundaries, Boundaries…Boundaries???** *Curr. Op. Cell Biol.* 2008, **20**: 281-287.
36. Probst AV, Dunleavy E, Almouzni G: **Epigenetic inheritance during the cell cycle.** *Nat. Rev. Mol. Cell Biol.* 2009, **10**: 192-206.
37. Margueron R, Reinberg D: **Chromatin structure and the inheritance of epigenetic information.** *Nat Rev Genet* 2010, **11**: 285-296.
38. Grewal SI, Elgin SC: **Heterochromatin: new possibilities for the inheritance of structure.** *Curr Opin Genet Dev.* 2002, **12**: 178-187.
39. Wachsmuth M, Caudron-Herger M, Rippe K: **Genome organization: balancing stability and plasticity.** *Biochim Biophys Acta* 2008, **1783**: 2061-2079.
40. Deng W, Blobel GA: **Do chromatin loops provide epigenetic gene expression states?** *Curr. Op. Gen. Dev.* 2010, **20**: 548-554.
41. Tolhuis B, Blom M, Kerkhoven RM, Pagie L, Teunissen H, Nieuwland M, Simonis M, de Laat W, van Lohuizen M, van Steensel B: **Interactions among Polycomb domains are guided by chromosome architecture.** *PLoS Genet.* 2011, **7**: e1001343.
42. Rohlf T, Winkler C: **Emergent Network Structure, evolvable Robustness, and nonlinear Effects of Point Mutations in an Artificial Genome Model.** *Advances in Complex Systems* 2009, **12**: 293–310.

Tables

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}	$\varepsilon_{\mathrm{HM}}$	K_{M}
H3K4me3	ESC	5	-0.075	-0.15	0.037
	MEF				0.048
	NPC				0.179
H3K27me3	ESC	5	-0.025	-0.15	0.037
	MEF				0.046
	NPC				0.136