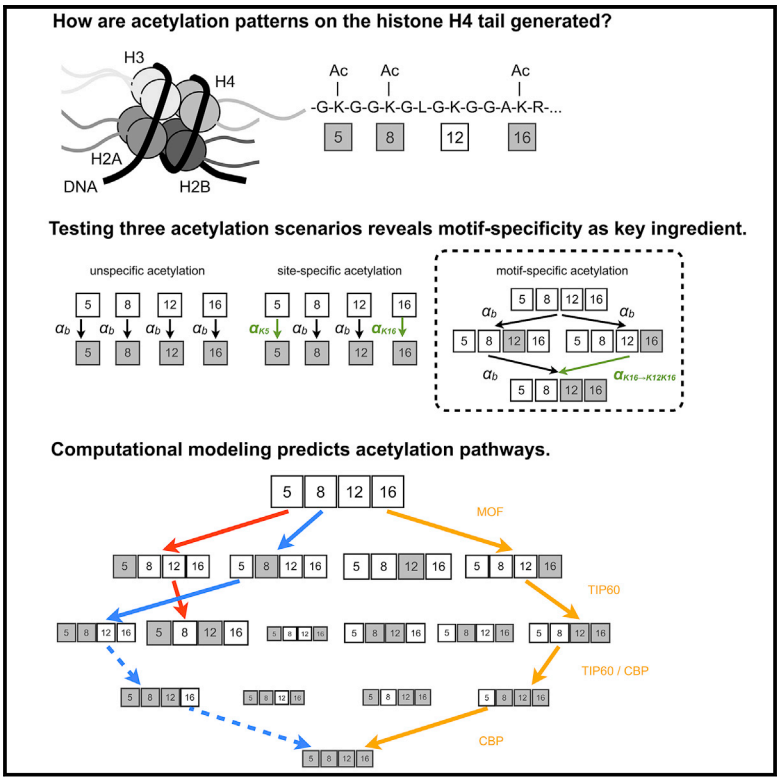
Cell Systems

Report

Combinatorial Histone Acetylation Patterns Are Generated by Motif-Specific Reactions

Graphical Abstract



Highlights

- Theoretical framework describes histone acetylation pattern abundances
- More than 10^9 computational models were evaluated and compared
- Motif-specific reaction rates are essential for the histone H4 acetylation network
- Enzymatic pathways contribute to combinatorial acetylation patterns

Authors

Thomas Blasi, Christian Feller, Justin Feigelman, ..., Fabian J. Theis, Peter B. Becker, Carsten Marr

Correspondence

feller@imsb.biol.ethz.ch (C.F.), carsten.marr@helmholtz-muenchen.de (C.M.)

In Brief

Developing and applying a computational framework for protein modification datasets, Blasi et al. find that combinatorial histone acetylation patterns emerge through motif-specific enzymatic pathways.

Blasi et al., 2016, Cell Systems *2*, 49–58
January 27, 2016
http://dx.doi.org/10.1016/j.cels.2016.01.002

CellPress

Cell Systems
Report

Combinatorial Histone Acetylation Patterns Are Generated by Motif-Specific Reactions

Thomas Blasi,[1,2] Christian Feller,[3,4,*] Justin Feigelman,[1,2,4] Jan Hasenauer,[1,2] Axel Imhof,[3] Fabian J. Theis,[1,2] Peter B. Becker,[3] and Carsten Marr[1,*]
[1]Institute of Computational Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany
[2]Chair of Mathematical Modeling of Biological Systems, Technische Universität München, Center for Mathematics, 85748 Garching, Germany
[3]Biomedical Center and Center for Integrated Protein Science Munich, Ludwig-Maximilians-Universität, 82152 Planegg, Germany
[4]Present address: Department of Biology, Institute of Molecular Systems Biology, ETH Zürich, 8093 Zürich, Switzerland
*Correspondence: feller@imsb.biol.ethz.ch (C.F.), carsten.marr@helmholtz-muenchen.de (C.M.)
http://dx.doi.org/10.1016/j.cels.2016.01.002

SUMMARY

Post-translational modifications (PTMs) are pivotal to cellular information processing, but how combinatorial PTM patterns (“motifs”) are set remains elusive. We develop a computational framework, which we provide as open source code, to investigate the design principles generating the combinatorial acetylation patterns on histone H4 in *Drosophila melanogaster*. We find that models assuming purely unspecific or lysine site-specific acetylation rates were insufficient to explain the experimentally determined motif abundances. Rather, these abundances were best described by an ensemble of models with acetylation rates that were specific to motifs. The model ensemble converged upon four acetylation pathways; we validated three of these using independent data from a systematic enzyme depletion study. Our findings suggest that histone acetylation patterns originate through specific pathways involving motif-specific acetylation activity.

INTRODUCTION

Post-translational protein modifications (PTMs) are central to the regulation of most cellular systems (Beltrao et al., 2013). Depending on their specific localization, PTMs can fine-tune protein function, modulate protein-protein interactions, or target proteins for destruction. Recent advances in mass-spectrometry-based proteomics have greatly improved the mapping of PTMs (Olsen and Mann 2013). Individual modifications mark specific residues, or “sites,” but if PTMs reside close to each other on a protein domain they may constitute a modification pattern, or “motif” (Figure 1A). The observation of clustered modifications suggests functions for motifs beyond that of individual site modifications. However, little is known about the mechanisms that generate modification motifs. Specifically, how do pre-existing chemical marks affect further modifications in their neighborhood? Do they arise from the cumulative effects of uncoordinated enzymatic activities or via dedicated, ordered pathways? Answering these questions requires the quantitative measurement of modification motif abundances and a theoretical framework that enables generating testable hypotheses about the nature of the modification network (see Supplemental Experimental Procedures, section 1).

Histones, the main constituents of chromatin and mediators of genome function and stability, provide a well-studied example of regulation by PTM motifs. Histones are multiply modified and the resulting motifs have been suggested to constitute a “code” (Strahl and Allis 2000; Turner 2000; Jenuwein and Allis 2001) or, alternatively, to integrate cellular and environmental signals and mediate appropriate responses (Smith and Shilatifard 2010). These modifications generate a rich landscape that is read out by interacting proteins to regulate gene expression programs and establish persistent “epigenetic” states (Kouzarides 2007; Suganuma and Workman 2011).

Of all the histone modifications, acetylation at specific lysine residues is particularly critical for the establishment of transcriptional programs and governs cell differentiation and organismal development (Voss and Thomas 2009; Haberland et al., 2009). Moreover, deregulation of acetylation contributes to aging and diseases (Dawson and Kouzarides 2012; Gräff and Tsai 2013; Cosentino and Mostoslavsky 2013; Morgan and Shilatifard 2015).

Acetylation of specific motifs is likely to convey specific functions. Indeed, the role of one combinatorial histone acetylation motif is already known: the doubly acetylated histone H4 at lysines 5 and 12 (H4.K5acK12ac) is generated by the acetyltransferase HAT1 (Parthun, 2012) and exhibits functions distinct from the singly acetylated K5 or K12 isoforms (see Supplemental Experimental Procedures, section 1).

In order to comprehensively characterize the histone acetylation system, we recently developed and employed a liquid chromatography-mass spectrometry (LC-MS)-based method (Feller et al., 2015). This technique enabled the precise and accurate quantification of the 16 possible motifs formed on the H4 N-terminal “tail” domain by combinatorial acetylation of lysines K5, K8, K12, and K16 (Figure 1B). As a proof of concept, we focused on *Drosophila melanogaster*, specifically KC cells, which serve as a well-established model for histone acetylation biology, with reduced complexity compared to human cells

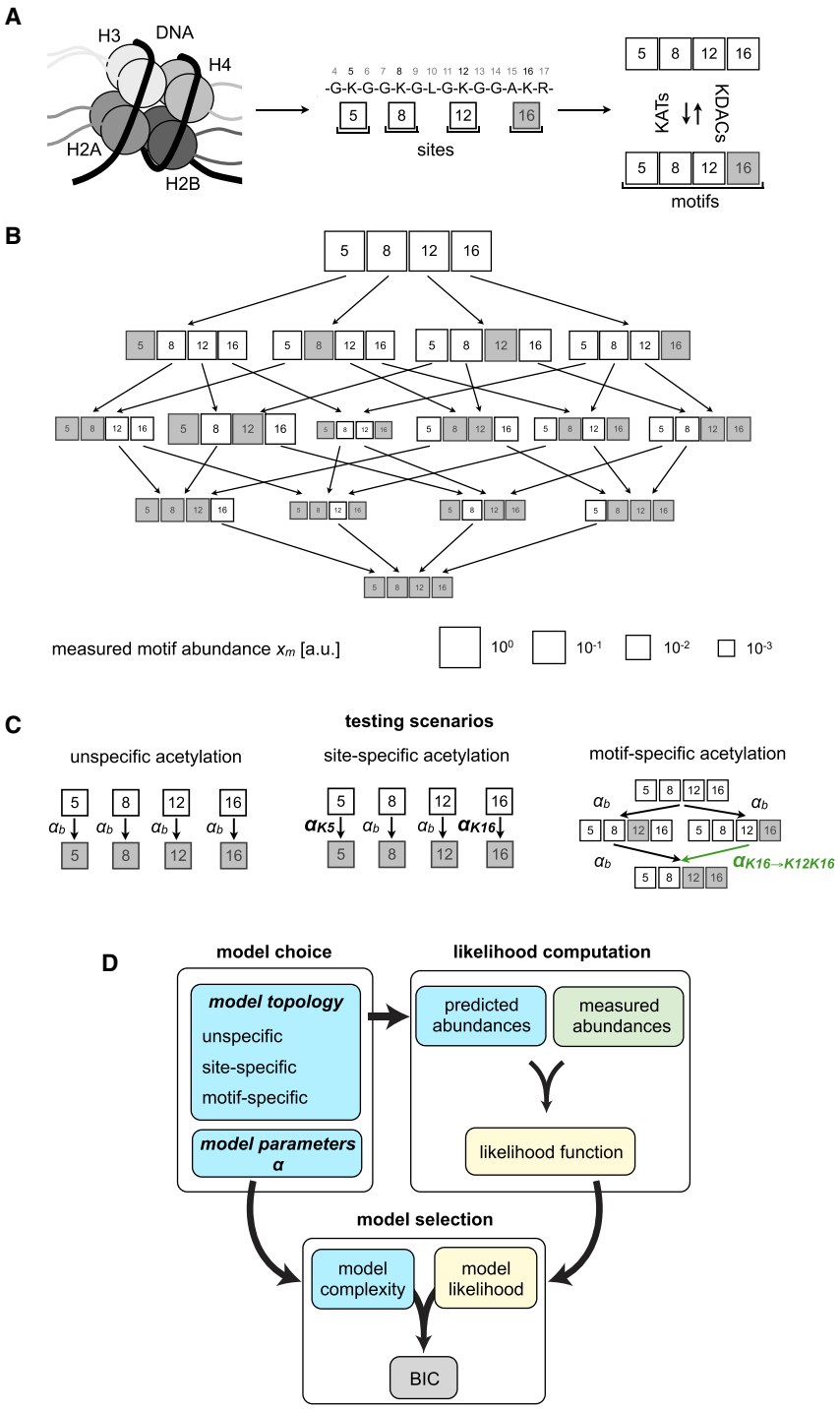


Figure 1. Overview on Modeling Framework to Test Histone H4 Acetylation Scenarios

(A) Nucleosome with protruding N-terminal histone "tail" domains (left), with four potential lysine (K) acetylation residues ("sites") on the histone H4 peptide G4-R17 (center), which together form the histone H4 N-terminal acetylation "motifs" (right). The motif K16 (right bottom) arises by acetylation of lysine 16 from the unmodified 0ac motif (right top). Acetylation and deacetylation reactions are mediated by lysine acetyltransferases (KATs) and lysine deacetylases (KDACs), respectively.

(B) Network representation of the abundances for the 16 histone H4 acetylation motifs as determined by liquid chromatography-mass spectrometry (LC-MS) (data from Feller et al., 2015). Box sizes depict log10 abundances. Note that the box size for K5K16 was set to scale with the lowest quantifiable H4 motif (K5K8K16), because the K5K16 motif is below the quantification limit (see Supplemental Experimental Procedures, section 1).

(C) Testing hypothetical acetylation scenarios: unspecific acetylation scenario with the same unspecific, basal reaction rate α_b for all lysines (left); site-specific acetylation scenario with site-specific reaction rate(s) for individual site(s) (center); and motif-specific acetylation scenario with rates specific to the modification context of lysines 5, 8, 12, and 16. In the depicted case, acetylation of lysine 12 requires pre-acetylated lysine 16 (green).

(D) Outline of computational modeling framework. All models are based on mass action kinetics with rate constants α_r for the r reactions. The models are fitted to the LC-MS data from (B) via maximum likelihood estimation (MLE). To find the model with an optimal trade-off between complexity and goodness of fit, we perform model selection based on the Bayesian Information Criterion (BIC), a score that penalizes more complex models. Applying this approach allows to test the different acetylation scenarios shown in (C) with a rigorous quantitative assessment.

(Straub and Becker 2011; Lucchesi and Kuroda 2015). In this previous study, we found that histone H4 acetylation in wild-type KC cells is unequally distributed across lysines and certain combinations of acetylations occur much more frequently than others (Figure 1B). Using a second, independent approach, we investigated the pathways by which those acetylation motifs are generated by systematically depleting all known and suspected lysine acetyltransferases (KATs) and lysine deacetylases (KDACs) expressed in KC cells (Feller et al., 2015). However, due to the complex interplay between the network components, the principles underlying the observed acetylation motif abundances could not be deduced using standard analysis.

Here, to interrogate data like these, we introduce a computational framework and use lysine acetylation in *Drosophila* KC cells as a test case. Our framework allows us to investigate the genesis of combinatorial histone acetylation motifs. Specifically, we assess the relative importance of dedicated synthesis pathways versus uncoordinated enzymatic activity in producing the complement of lysine acetylation marks observed in bulk measurements. We trained our model on published data from

unperturbed cells and validated predicted pathways using the independent KAT depletion dataset. Our modeling strategy provides insight into the design principles of PTM motifs, which will become increasingly relevant as further LC-MS datasets become available.

RESULTS

A Computational Framework for Modeling Endogenous Acetylation Motif Abundances

To help understand the complement of H4 acetylation isoforms observed in the KC cell-derived LC-MS datasets, it is useful to conceptualize the data in terms of "layers." The number of layers equals the number of lysines in the H4 tail, and each layer is composed of the set of H4 isoforms containing the same total number of acetylations (Figure 1B). The abundance of each of the possible 16 histone H4 acetylation isoforms (here called "motifs") is unevenly distributed both within layers (representing the diversity of motifs with the same total number of acetylations) and across layers (that is, the total number of acetylated lysines, see Figure 1B). However, the general design principles underlying this complex acetylation patterns remain elusive. We hypothesized that *motif-specific* reactions contribute to the skewed abundance distribution, where enzymes are sensitive to adjacent modifications (or their absence), and hence catalyze (de)acetylations only for specific acetylation motifs (Feller et al., 2015).

To test this hypothesis, we assess three hypothetical acetylation scenarios (Figure 1C):

(1) Lysine acetylation could be *unspecific*, i.e., not dependent on the site or motif, and therefore be governed by a single, basal acetylation rate constant $\alpha_r = \alpha_b$ for all reactions r (Figure 1C, left).
(2) Acetylation could be *site-specific* with some or all of the four lysine sites K5, K8, K12, K16 being acetylated by site-specific enzymes, resulting in a common acetylation rate for each reaction that targets that site, independent of neighboring modifications (e.g., $\alpha_{K5} \neq \alpha_{K16} \neq \alpha_b$; Figure 1C, middle).
(3) Acetylation could be *motif-specific* with enzymes being sensitive to the modification state of nearby lysines. In the depicted example, the acetylation rate at site K12 is different from the basal rate when K16 (but not K5 or K8) is already acetylated ($\alpha_{K16\rightarrow K12K16} \neq \alpha_b$; Figure 1C, right).

We developed a series of mathematical models to encompass these three acetylation scenarios. For each model, we predicted the abundance of all acetylation motifs as a function of the acetylation rates and infer the most likely acetylation rates by fitting the model to the measured abundances (Figure 1D).

In the following, we use the symbol m to denote a motif (e.g., m = 0ac for the state with no acetylations, m = K5 for acetylated lysine 5 and no modifications at lysines 8, 12, 16, etc.) and x_m for the relative abundance of m, for each of the 16 acetylation motifs. In our model, we assume:

- Acetylation and deacetylation occurs stepwise, i.e., only one acetylation or deacetylation event can occur at a time. Thus, a motif can be generated via a single acetylation of a less acetylated state or a single deacetylation of a more acetylated state (Figure 1C).
- Deacetylation is unspecific, such that the rate constant is the same for all reactions r connecting two motifs, consistent with current views on deacetylation and measured response upon KDAC depletion in *Drosophila* cells (Seto and Yoshida 2014; Feller et al., 2015; Supplemental Experimental Procedures, sections 2 and 3).
- Histone H4 acetylation follows mass action kinetics: the rate of each acetylation or deacetylation reaction is proportional to the abundance of the substrate of that reaction.
- The measured motif abundances are assumed to be in steady state. This is consistent with the fact that acetylation and deacetylation progresses much faster than the cell cycle and histone protein turnover (Katan-Khaykovich and Struhl 2002; Toyama et al., 2013).

For a more detailed explanation of this approach, see Supplemental Experimental Procedures, sections 2 and 3.

For the site-specific and motif-specific scenarios, it is unknown which of the sites or motifs are acetylated with a non-basal rate; therefore, we considered models containing any combination of site-specific or motif-specific reactions. In the case of site-specific acetylation, there are 11 different combinations of site-specific and basal acetylation rates: four possibilities with one site-specific rate, six possibilities with two site-specific rates, and one possibility where every site has a site-specific acetylation rate. Note that the case of three site-specific rates and a basal rate is identical to four site-specific rates. For the motif-specific acetylation scenario, there are a total of $2^{32} = 4.295 \times 10^9$ possible combinations of motif-specific reactions among the 32 acetylation reactions.

In each case, we evaluate the likelihood of the model using a log-error model for the measured abundances from unperturbed *Drosophila* cells. We then optimized model parameters to obtain the best fit, given by the maximum likelihood estimator (MLE). This approach also allows us to calculate confidence intervals for the estimated parameters and abundances, which is not possible using a likelihood-free approach, such as least-squares fitting. Using these likelihoods, we determined for each model the Bayesian Information Criterion (BIC), which penalizes additional model parameters. With the BIC, we can compare different models with each other: the smaller the BIC the more suitable a model is to describe the data. This prevents overfitting the data with too complex models and permits the identification of the simplest models capable of describing the data (Figure 1D and Supplemental Experimental Procedures, section 3).

The basic assumptions explained above, along with the inference procedure and model selection strategy, enable us to identify quantitative candidate models of the acetylation network (Figure 1B), which is necessary for evaluating the hypothetical acetylation scenarios and deriving testable predictions.

Testing Three Hypothetical Acetylation Scenarios: Models Containing Motif-Specific Reactions Are Favored over Models with Only Basal and Site-Specific Reactions

We start by analyzing the simplest scenario, which contains only one basal reaction rate governing all acetylation reactions. In this

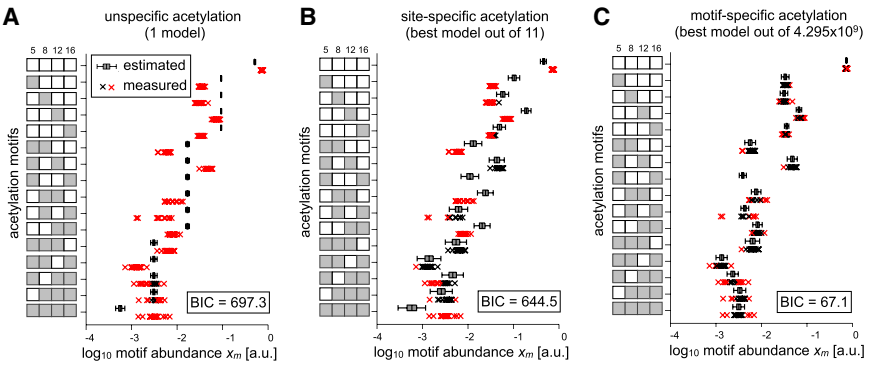


Figure 2. Motif-Specific Models Are Preferred over Unspecific and Site-Specific Models for Explaining the Measured Motif Abundances

(A–C) Comparison between modeling-derived predicted motif abundances (gray whisker box, percentiles 0.5/0.95) and LC-MS measured motif abundances (red and black "x", data with 18 biological replicates from Feller et al., 2015) for different acetylation scenarios. Red x: measurement falls outside of confidence interval (CI) of predicted motif abundance. Black x: measurement is within CI of predicted abundance value. Model selection based on Bayesian Information Criterion (BIC) shows strong preference for the best motif-specific acetylation model over the best site-specific model (ΔBIC = 577.4) and unspecific model (ΔBIC = 630.2).

(A) The unspecific model is inadequate to explain the measured abundances, because the mean abundance values from 13 out of 15 measured motifs do not fall within the 95% CIs of the model.

(B) The site-specific models are better than the unspecific model but still insufficient to explain most measurements (shown are data for the best out of 11 possible models: 8 out of 15 fall within 95% CI).

(C) The motif-specific acetylation models explain the measured abundances best. Shown are the data for the best motif-specific model out of 4.295×10^9 possible models (15/15 within 95% CI). Only the best 2.366×10^8 models were tested, allowing for up to 11 motif-specific reaction rates (BIC stopping criterion see Figure S1, Results, and Supplemental Experimental Procedures, section 2).

unspecific model, all motifs with the same total number of acetylations (0, 1, 2, 3, and 4), known as positional isoforms, will have the same abundance due to symmetry of the model. Using MLE, we determined that a basal rate constant α_b = 0.182 (0.170, 0.195) (95% confidence interval estimated using the profile likelihood) best fits the data. The basal reaction rate is hence approximately 18% of the deacetylation rate. The net effect of this bias toward deacetylation is the prevalence of motifs in layers with less acetylation. This simple scenario already suffices to capture the overall trend of decreasing abundance with increased degree of acetylation apparent in the data (Figure 2A). However, comparison of the predicted and the experimentally determined values shows that the abundance for motifs with a single or two acetylated lysines (except K5K12) is overestimated by this very simple model. Moreover, it cannot capture the large abundance variability present for the multiply acetylated motifs. In total, we find that 86.6% (13 out of 15 measured abundances) are not explained by this simple, unspecific scenario, because they fall outside the estimated 95% confidence intervals of the model. We therefore conclude that the unspecific scenario is insufficient to explain the majority of observed lysine acetylation abundances.

Next, we analyze the site-specific acetylation scenario, in which the acetylation rates of each site K5, K8, K12, and K16 are allowed to differ. We constructed all 11 possible models and in each case obtained the MLE of the parameters. We compared the models using the BIC, and found that the most complex model with four different site-specific reaction rates best explains the measured abundances. This model yields the site-specific acetylation rate constants α_{K5} = 0.225 (0.181, 0.280), α_{K8} = 0.126 (0.104, 0.154), α_{K12} = 0.417 (0.331, 0.528), and α_{K16} = 0.107 (0.088, 0.131). We note that, as for the unspecific scenario, the acetylation rates are all less than 1, indicating that deacetylation is faster than acetylation in this dynamic equilibrium. Since in this model, each site possesses a different acetylation rate, it is possible for motifs within a single layer to achieve different abundances. Indeed, this extra flexibility allows the model to achieve better agreement with the data compared to the unspecific model (Figure 2B) and is favored by the BIC (644.5 for the site-specific model as compared to 697.3 for the unspecific model). We find that 46.6% of all replicates of the measured motif abundances (0ac, K5, K12, K5K8, K8K12, K12K16, and 4ac) fall outside the 95% confidence intervals of the predicted abundances, representing a roughly 2-fold improvement compared to the purely unspecific model. However, the substantial fraction of measurements not explained by the model suggests that a more complex model, which accounts for differences among individual motifs, is required.

Finally, we analyzed the motif-specific acetylation scenario, which takes into account the context of nearby modifications. We computed the BIC score of each model and find that the BIC reaches its minimum for seven motif-specific rates and then increases again monotonically (Figure S1). Thus, we conservatively limit the analysis to maximally 11 motif-specific rates, thereby reducing the number of candidate models from 4.295×10^9 to 2.366×10^8 models. Of the motif-specific models, the model that best fits the data has seven motif-specific reaction rates, while all other reactions are governed by the same basal reaction rate. Six motif-specific rates are faster than the basal rate of $\alpha_b = 0.067 \pm 0.004$ and three reactions (K5 → K5K12, K12K16 → K8K12K16 and K8K12K16 → 4ac) have an acetylation rate constant $\alpha_r > 1$, indicating faster acetylation than deacetylation (Table S1). The obtained BIC of 67.1 indicates a strong preference of the motif-specific scenario over the unspecific (ΔBIC = 631.2) and the best site-specific scenario (ΔBIC = 577.4), which is corroborated by the observation that the best motif-specific model captures all measured mean motif abundances (Figure 2C).

Motif Abundances Are Best Described by an Ensemble of Similar Models with Distinct Motif-Specific Rates

Our analysis of the three hypothetical scenarios provides strong evidence that motif specificity is a necessary component of a

model that accurately recapitulates experimentally measured motif abundances. Until now, we considered only unspecific or *exclusively* site- or motif-specific models. However, by allowing both site- and motif specificity simultaneously, it might be possible to obtain an even better candidate model. Thus, we re-fit all models containing up to 11 motif or site-specific acetylation rates (9.3×10^8 total models) and compared the models using the BIC score. Interestingly, we found that permitting both site- and motif-specific rates improves the corresponding model BICs only minimally (Figure S2A). The model comparison reveals that the best model in this mixed scenario is exactly the same candidate found in the motif-specific scenario, that is, no site specific reactions included (Figure 2C). Taken together, this implies that site specificity is not essential to describe the observed abundances and confirms our previous finding (Figures 2A–2C) that motif specificity strongly contributes to the acetylation network.

Although the best model candidate has been independently identified twice, it is possible that other candidates are nearly as good at explaining the data. Thus, we compared the best model with all other models by examining the distribution of BIC scores over all scenarios. This analysis reveals that (1) the BIC scores of motif-specific models are vastly reduced compared to site-specific or unspecific models, confirming the essentiality of motif specificity (Figure S2A), and (2) there are approximately 100 motif-specific and motif-and-site-specific models with very similar BIC scores to the best model ($\Delta BIC < 6$) (Figure S2B). However, a ΔBIC of less than 6 is insufficient to reject one model in favor of another (see Supplemental Experimental Procedures, section 2); therefore, we retained an ensemble of candidate models for further consideration.

Ensemble Analysis: The Best Models Share Common Features

Next, we examined the ensemble of the best 100 models to ascertain similarities and differences between them. All models in the ensemble contain between six and nine motif-specific reaction rates (Figures S2C and S2E; Table S2) and 89 of the best 100 models have only one or no site-specific rate (Figures S2D and S2E). Not all specific rates appear equally within the model ensemble: particular motif-specific and site-specific rates are overrepresented (Figure 3A). We find that the motif-specific reaction K5 → K5K12 is common to almost all models in the ensemble (99%). With an average acetylation rate constant of $\alpha_{K5 \rightarrow K5K12} = 2.255 \pm 0.505$, it is on average 35 times faster than the basal acetylation rate constant $\alpha_b = 0.064 \pm 0.012$. Notably, this observation resonates with independent genetic and biochemical observations: K5K12 is the only acetylation motif that has been demonstrated to have a dedicated genesis pathway via the KAT HAT1 (see Introduction and Supplemental Experimental Procedures, section 1).

Average rate constants within the ensemble of 100 best models also have literature support. We find that the ensemble average effective basal acetylation rate constant, i.e., the ratio of basal acetylation to deacetylation, is of the same order of magnitude as the one previously reported using chromatin immunoprecipitation (ChIP) and pulse-chase LC-MS metabolic-labeling experiments (0.1–0.2 in Katan-Khaykovich and Struhl, 2002; Zheng et al., 2012; Evertts et al., 2013). The relatively high abundance of multiply acetylated states suggests the need for acetylation rates that are higher than the basal acetylation rates, which in the model is implemented by site-specific and motif-specific acetylation rate enhancement. In total, we conclude that our modeling approach robustly identifies motif specificity as a system property of histone acetylation, capable of capturing the only yet known motif-specific reaction rate (K5 → K5K12) and recovering a basal effective acetylation rate with values similar to previous estimates.

Computational Prediction of Pathways that Lead to Combinatorial Acetylation Motifs

In order to explore the commonalities among the best models, we grouped them using an unsupervised clustering algorithm on the reaction rates of the models. By doing so, we identified three large model families (at least five models per family), distinguished by similar patterns among their reaction rates (Figure 3C).

We examined the frequency distribution of motif-specific reactions within the ensemble (Figure 3A) and discovered a subset of overrepresented motif-specific reactions, which occur in more than 50% of the candidate models (Figure 3B). Each family also contained a different collection of frequently occurring motif-specific rates (Figures S3A–S3C). This procedure allows us to analyze properties of the model families in terms of their common features. In each model family, we conservatively threshold the motif-specific reaction frequencies at 60% to yield a set of essential motif-specific reactions (Figures S3D–S3F) and connect adjacent motif-specific reactions from this set to construct hypothetical acetylation pathways (Figures 4A–4D).

We identify four pathways, which are distinct between the three model families (Figures 4A–4D). Pathway 1 (supported by 45.8% of the models in family 3, Figures 4A, S3C, and S3F) suggests that K5K12 is generated not only via K5 → K5K12 (100% of the models in family 3), but also via K12 → K5K12 (45.8% of the models in family 3). Pathway 1 further suggests that K5K12 is subsequently acetylated at K8 to generate K5K8K12 (78.0% of the models in family 3). Furthermore, most models (84.8%) in family 3 contain pathway 4 (Figures 4D, S3C, and S3F), which yields the fully acetylated isoform 4ac via the serial acetylation K16 → K12K16 → K8K12K16 → 4ac. We also discover alternative pathways to the fully acetylated 4ac state in model families 1 and 2. Pathway 2 (contained in 81.8% of the models in family 1, Figures 4B, S3B, and S3E) suggests K8 → K5K8 → K5K8K12 → 4ac, while pathway 3 (contained in 42.3% of the models in family 2, Figures 4C, S3A, and S3D) suggests K8K12 → K5K8K12 → 4ac .

Validation of Computationally Predicted Pathways against a Systematic KAT Depletion Dataset

After having trained our model ensemble on the dataset derived from unperturbed *Drosophila* cells, we validated the pathways it predicted using an independent dataset. To generate these data, individual KATs were depleted by RNA interference (Figure S4; Feller et al., 2015). For a detailed discussion on enzyme annotation and methods used to generate these data, refer to Supplemental Experimental Procedures, section 4.

Pathway 1 (Figure 4A) clearly resembles the known HAT1-mediated acetylation of K5K12 via K5. Upon HAT1 depletion

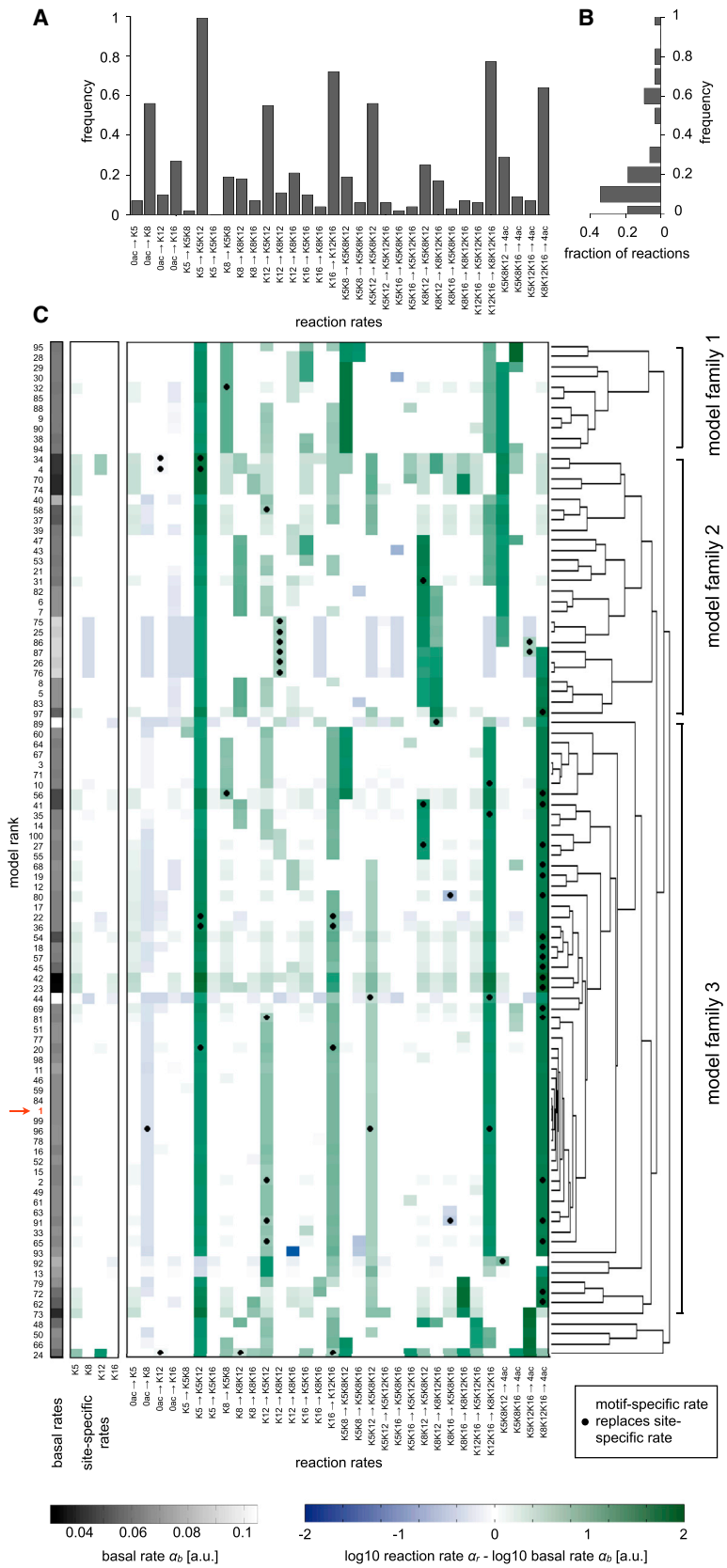


Figure 3. Hierarchical Clustering Reveals Distinct Model Families Characterized by Motif-Specific Reaction Rates

(A) The histogram shows the abundance distribution for all motif-specific reaction rates among the best 100 models. K5 → K5K12 is supported by 99/100 best models. Seven reaction rates occur with a frequency of more than 50%.

(B) The motif-specific reactions occur with variable frequency within the best 100 models.

(C) We performed hierarchical clustering to group models according to similarity patterns in their estimated acetylation rates. The resulting clusters can be categorized into three model families and a small group of five uncategorized models. The black dots in the heatmap indicate cases where the motif-specific reaction rate supersedes the site-specific rate (left columns). The color key shows the magnitude of the log fold change of reaction rates (colored, right) relative to the basal rate of each model (gray, left). Best model (ranked 1) is highlighted in red (line 75).

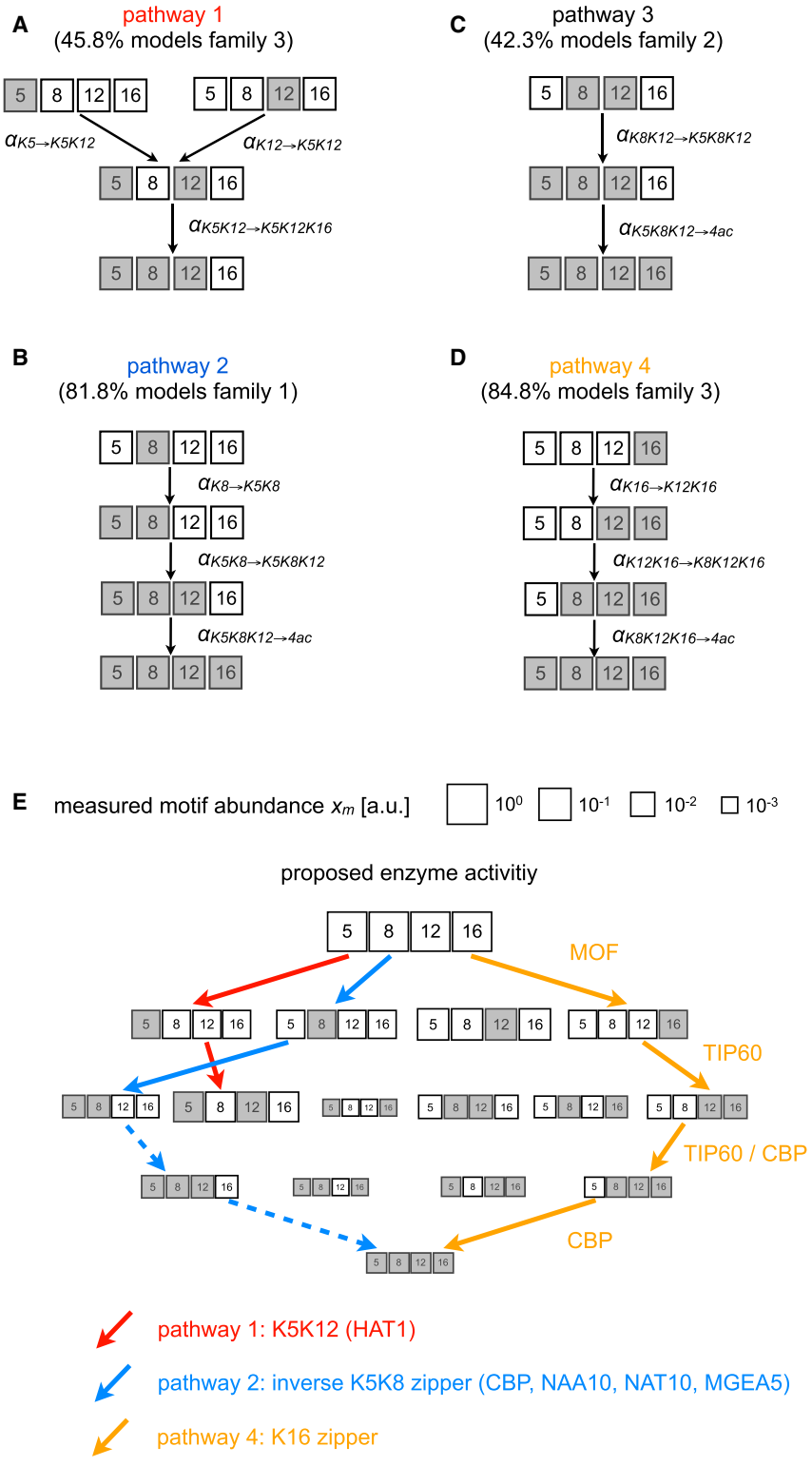


Figure 4. Hypothetical Acetylation Pathways Derived from the Model Ensemble and Schematic Representation of Proposed Enzyme Activity

(A–D) Predicted acetylation pathways that are composed of connected motif-specific acetylation rates with more than 60% support within a family (see Figure S3). Each model family is characterized by distinct acetylation reaction pathways.

(E) We validate the predicted pathways by an independent KAT depletion dataset (Figure S4) and propose candidate enzymes for the acetylation pathways. We find evidence for pathway 1, 2, and 4 (Figures 4A, 4B, and 4D), whereas we exclude pathway 3 (Figure 4C). In pathway 1 (red), HAT1 catalyzes the two main steps of the K5K12 pathway and a potential third "promiscuous" step to yield K5K8K12 (data not shown). CBP, NAA10, NAT10, or MGEA5 are putative candidates for the first two reactions of the inverse K5K8 zipper (pathway 2, blue, solid). The remaining links of pathway 2 (blue, dashed) are inferred by the model, but the dataset does not allow the assignment of associated enzymes. In pathway 4 (orange), the K16 zipper is likely generated by the subsequent actions of MOF (0 → K16), TIP60 (K16 → K12K16 and K12K16 → K8K12K16), and CBP (K12K16 → K8K12K16 and K8K12K16 → 4ac).

(Figure S4), we see a strong decrease in the abundances of K5 and K5K12, while the unused substrate K12 increases. Pathway 1 additionally reveals a promiscuous acetylation activity of HAT1 for K5K12 → K5K8K12, an observation consistent with a previous biochemical report (Makowski et al., 2001) and the observation that the newly synthesized histones engaged in chaperone assemblies contain low levels of K5K8K12 (Verreault et al., 1996).

Pathway 2 (Figure 4B) suggests a route to the 4ac motif via K8 → K5K8 → K5K8K12. By analyzing our depletion dataset, we propose CBP, NAA10, NAT10, and MGEA5 to contribute to pathway 2, whose ablation leads to a reduction of the corresponding motifs. We find no clear evidence for pathway 3 (Figure 4C) after having visited our depletion dataset.

The acetylation pathway 4 (Figure 4D) suggests successive acetylation of the H4 tail starting from K16 and propagating toward the N terminus until lysine 5. By analyzing our depletion dataset, we propose this pathway to be generated by subsequent acetylation activity of MOF (0ac → K16), TIP60 (K16 → K12K16), and CBP (K8K12K16 → 4ac), since the reaction products K16, K12K16, and 4ac are strongly reduced upon depletion of MOF, TIP60, and CBP, respectively. Importantly, as predicted by the model, the unused substrate for the first TIP60 reaction (i.e., K16) is increased after TIP60 depletion. Additionally, the depletion dataset provides support for the K12K16 → K8K12K16 reaction by both TIP60 and CBP. Consistent with the predicted pathway, the abundance of K12K16 is increased if the enzyme catalyzing K12K16 → K8K12K16 is depleted, which is consistent with the

outcome of the CBP depletion experiment, unless the enzyme also contributes to the generation of K12K16, as depicted above for TIP60.

In summary, our analysis identifies three acetylation pathways that can be linked to known and recently reported enzyme activity (Figure 4E). The combinatorial acetylation motif K5K12 is generated by HAT1 via K5 (red pathway in Figure 4E). The fully acetylated 4ac state appears to originate from two main routes: the "inverse K5K8 zipper pathway" (blue pathway in Figure 4E) relies on the enzymes CBP, NAA10, NAT10, MGEA5, and the "K16 zipper pathway" (orange pathway in Figure 4E) is brought about by the coordinated activity of MOF, TIP60, and CBP. The "K16 zipper pathway" was previously proposed based on the skewed abundance distribution in many human cell types as well as in other selected species (Zhang et al., 2002; Garcia et al., 2007). In those species, the abundance of K16 exceeds by far that of other mono-acetylated motifs, a trend that continues also for di- and tri-acetylated motifs that contain acetylated K16. While established for other species, such a pathway has not been previously described for *Drosophila*, nor is it evident from the measured abundances alone (Figure 1B). The identification of the K16 zipper pathway is therefore an unexpected result from our modeling framework and has direct implications for the evolution of the histone H4 acetylation system and the specialized chromosome-wide transcriptional control system present in male *Drosophila* cells (dosage compensation system; Prestel et al., 2010; Bachtrog et al., 2014; Lucchesi and Kuroda 2015).

DISCUSSION

In this study, we developed a modeling framework to investigate how combinatorial motifs of PTM modifications are generated and maintained using acetylation of the histone H4 N-terminal tail domain as a model system. Our findings strongly suggest that motif-specific acetylation is an essential feature of the histone H4 acetylation system in *Drosophila*. In particular, our results argue that (1) site-specific reactions are not the main driving force shaping the skewed distribution of histone abundances, as commonly assumed, but (2) motif-specific reaction rates exist and are essential components of the histone H4 acetylation network, and (3) motif-specific reactions create pathways that generate the combinatorial acetylation patterns. These conclusions arise directly from aspects of the modeling framework we have introduced here. The framework is exhaustive, in that we compare all model topologies; this allows us to discriminate between possible models. It utilizes an established information-theoretic technique to rigorously evaluate candidate models, and it is directly interpretable, revealing acetylation rates, and site and motif specificity.

Our framework can also be extended. Using the presented approach, one can identify pathways that are directly testable, aiding in hypothesis generation and experimental design. One may also "reverse-engineer" complex and diverse PTM networks. Besides including additional PTM types, it is also possible to change the parameters of the model. For example, while our assumption of a universal, constant deacetylation rate is justified by our current knowledge on KDAC activities in *Drosophila* cells (Feller et al., 2015), the higher complexity of the human KDAC network and the differential response upon KDAC inhibitor treatment in some human cells may motivate an extension, which relaxes this assumption (Yang and Seto 2007; Joshi et al., 2013; also see Supplemental Experimental Procedures, section 3).

Implications for Histone Modification Biology and Beyond

A common notion in histone acetylation biology is that modifications at neighboring sites may serve largely redundant functions, either through unspecific "reader" proteins that bind acetylated lysines with low affinity and specificity for individual acetylation sites or through charge neutralization (Dion et al., 2005; Ruthenburg et al., 2007; Allahverdi et al., 2011). Furthermore, KATs are generally thought to target either a single site with high specificity (e.g., MOF acetylates unmodified histone H4 at lysine 16) or multiple residues with relaxed specificity (e.g., CBP for lysine residues K5, K8, and several sites on histone H3). However, the comprehensive characterization of the motif targets of even a single KAT is greatly limited by the inadequacy of current biochemical methods (see Supplemental Experimental Procedures, section 1, for details).

Our computational results on the acetylation genesis side (that is, the "writer" proteins that perform lysine acetylation) are congruent with emerging evidence from the "reader" side, which recently demonstrated that the affinity and specificity of some bromodomains to interact with acetylated lysines is modulated by additional nearby modifications (Morinière et al., 2009; Filippakopoulos et al., 2012). Importantly, our findings also are consistent with the only acetylation motif with a suspected functional role and dedicated genesis, K5K12 (see Supplemental Experimental Procedures, section 1). Together, these independent observations on the histone acetylation system are congruent with alike observations from histone methylation studies (Strahl and Allis 2000; Jenuwein and Allis 2001; Smith and Shilatifard 2010; Suganuma and Workman 2011).

In this work, we comprehensively investigated lysine acetylation at the N terminus of histone H4. However, this represents only a small part of the vast histone modification network, which includes over 20 known PTM types and over 150 modification sites. Thus, a comprehensive characterization would require expanding the current model to include additional PTM types and motifs and their respective interactions. Such extensions will be possible to include in our framework once these datasets become available, as heralded by recent and foreseeable progress in MS-based technologies and complementary (proteo-) genomic, systems biochemistry and genetics methodologies (Günesdogan et al., 2010; Tran et al., 2011; Cox and Mann 2011; Gillet et al., 2012; Bensimon et al., 2012; Nguyen et al., 2014; Stasevich et al., 2014). Supplied with such data, quantitative, dynamic models of the chromatin modification system in healthy and diseased cells will yield a detailed characterization of PTM network and may provide the means to rationally design specific modulators of epigenetic regulation as a basis for effective therapy.

EXPERIMENTAL PROCEDURES

Model Definition

To model the reaction network of histone H4 tail acetylation and deacetylation, we use a stepwise reaction network (Figure 1B) wherein only a single acetyl

group can be added or removed at a time. We use mass action kinetics (Van Kampen 2007; Zheng et al., 2012) to describe the acetylation and deacetylation reactions r (r = 0ac → K5, r = K5 → K5K12, etc.) with acetylation rate constants a_r and deacetylation rate constants d_r. The change of the abundances of a particular motif x_m is given by the difference between influx from and the outflux to neighboring motif abundances

$$d/dt\, x_m = \sum_{m'} R_{m,m'}(a_r, d_r) \cdot x_{m'}. \qquad \text{(Equation 1)}$$

The reaction matrix $R_{m,m}$ incorporates all possible reactions between the motifs within the stepwise reaction network. For example, the change of the abundance of motif K8,

$$d/dt\, x_{K8} = a_{K8}x_{0ac} + d_{K5}x_{K5K8} + d_{K12}x_{K8K12} + d_{K16}x_{K8K16} - (a_{K5} + a_{K12} + a_{K16} + d_{K8})x_{K8},$$

is given as the difference between the influx from (terms with positive sign) and the outflux to (terms with negative sign) neighboring motifs (i.e., motifs differing by exactly one acetyl group).

We and others previously found that deacetylation occurs broadly with only little evidence for motif specificity in the analyzed cell system (Feller et al., 2015). Thus, we additionally assume that deacetylation is unspecific, and therefore proceeds with the same rate constant d for all motifs. We then simplify by dividing Equation 1 by the deacetylation rate constant d, thus converting all acetylation rate constants a_r to rescaled, effective rate constants $\alpha_r = a_r / d_r$.

The timescale of acetylation and deacetylation is much faster than the cell cycle and measured histone protein half-life. Therefore, we assume the reactions to be at steady state

$$d/dt\, x_m = 0. \qquad \text{(Equation 2)}$$

To take into account that the abundances of the motifs are not measured in total numbers but are given as fractions that are normalized to one, we solve for the steady states subjected to the constraint of unity total abundance; i.e., we solve

$$\sum_{m'} R_{m,m'}(\alpha) x_{m'} = 0 \quad s.t. \quad \sum_m x_m = 1. \qquad \text{(Equation 3)}$$

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.cels.2016.01.002.

AUTHOR CONTRIBUTIONS

T.B. and C.F. wrote the draft manuscript, and J.F., T.B., C.F., P.B.B., and C.M. created the final version of the manuscript. T.B., J.F., J.H., and C.M. conceived the model with input from C.F. T.B. and C.F. analyzed the data. T.B. conducted the computational analysis. C.F. conceived the project and C.M., C.F., J.F., and T.B. designed the study. F.J.T., A.I., and P.B.B. critically commented throughout the project. C.M. supervised the study. Correspondence regarding experimental data can be addressed to C.F.; correspondence regarding computational modeling and other inquiries can be addressed to C.M.

ACKNOWLEDGMENTS

T.B., J.F., J.H., F.J.T., and C.M. were supported by the Helmholtz Alliance on Systems Biology (project CoReNe), the European Research Council (starting grant LatentCauses), and the Deutsche Forschungsgemeinschaft (SPP 1356 Pluripotency and Cellular Reprogramming). T.B. was supported by the Studienstiftung des deutschen Volkes. J.H. was supported by the German Federal Ministry of Education and Research (BMBF) within the SYS-Stomach project (grant no. 01ZX1310B), and the Postdoctoral Fellowship Program (PFP) of the Helmholtz Zentrum München. A.I. was supported by a grant from the European Union (Epigenesys 257082). P.B.B. acknowledges support by the European Research Council (FP7/2007-2013) grant agreement number 293948. C.F. acknowledges support by the Deutsche Forschungsgemeinschaft (DFG postdoctoral fellowship, grant number FE 1544/1-1). We are grateful to Andreas Raue and Matthias Heinig for advice on the modeling strategy, Dennis Rickert for support with computational resources, and three anonymous reviewers for constructive criticism. C.F. acknowledges the "Zurich cellular memory club" for insightful discussions and helpful comments to improve the manuscript, in particular, Haochen Yu, Daniel Sévin, Karl Kochanowski, Alexander Leitner, Marieke Buffing, and Laura De Vargas Roditi.

Received: September 20, 2015
Revised: November 18, 2015
Accepted: January 5, 2016
Published: January 27, 2016

REFERENCES

Allahverdi, A., Yang, R., Korolev, N., Fan, Y., Davey, C.A., Liu, C.-F., and Nordenskiöld, L. (2011). The effects of histone H4 tail acetylations on cation-induced chromatin folding and self-association. Nucleic Acids Res. *39*, 1680–1691.

Bachtrog, D., Mank, J.E., Peichel, C.L., Kirkpatrick, M., Otto, S.P., Ashman, T.-L., Hahn, M.W., Kitano, J., Mayrose, I., Ming, R., et al.; Tree of Sex Consortium (2014). Sex determination: why so many ways of doing it? PLoS Biol. *12*, e1001899.

Beltrao, P., Bork, P., Krogan, N.J., and van Noort, V. (2013). Evolution and functional cross-talk of protein post-translational modifications. Mol. Syst. Biol. *9*, 714.

Bensimon, A., Heck, A.J.R., and Aebersold, R. (2012). Mass spectrometry-based proteomics and network biology. Annu. Rev. Biochem. *81*, 379–405.

Cosentino, C., and Mostoslavsky, R. (2013). Metabolism, longevity and epigenetics. Cell. Mol. Life Sci. *70*, 1525–1541.

Cox, J., and Mann, M. (2011). Quantitative, high-resolution proteomics for data-driven systems biology. Ann. Rev. Biochem. *80*, 273–299.

Dawson, M.A., and Kouzarides, T. (2012). Cancer epigenetics: from mechanism to therapy. Cell *150*, 12–27.

Dion, M.F., Altschuler, S.J., Wu, L.F., and Rando, O.J. (2005). Genomic characterization reveals a simple histone H4 acetylation code. Proc. Natl. Acad. Sci. USA *102*, 5501–5506.

Evertts, A.G., Zee, B.M., Dimaggio, P.A., Gonzales-Cope, M., Coller, H.A., and Garcia, B.A. (2013). Quantitative dynamics of the link between cellular metabolism and histone acetylation. J. Biol. Chem. *288*, 12142–12151.

Feller, C., Forné, I., Imhof, A., and Becker, P.B. (2015). Global and specific responses of the histone acetylome to systematic perturbation. Mol. Cell *57*, 559–571.

Filippakopoulos, P., Picaud, S., Mangos, M., Keates, T., Lambert, J.-P., Barsyte-Lovejoy, D., Felletar, I., Volkmer, R., Müller, S., Pawson, T., et al. (2012). Histone recognition and large-scale structural analysis of the human bromodomain family. Cell *149*, 214–231.

Garcia, B.A., Hake, S.B., Diaz, R.L., Kauer, M., Morris, S.A., Recht, J., Shabanowitz, J., Mishra, N., Strahl, B.D., Allis, C.D., and Hunt, D.F. (2007). Organismal differences in post-translational modifications in histones H3 and H4. J. Biol. Chem. *282*, 7641–7655.

Gillet, L.C., Navarro, P., Tate, T., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol. Cell Proteomics *11*, O111.016717.

Gräff, J., and Tsai, L.-H. (2013). Histone acetylation: molecular mnemonics on the chromatin. Nat. Rev. Neurosci. *14*, 97–111.

Günesdogan, U., Jäckle, H., and Herzig, A. (2010). A genetic system to assess in vivo the functions of histones and histone modifications in higher eukaryotes. EMBO Rep. *11*, 772–776.

Haberland, M., Montgomery, R.L., and Olson, E.N. (2009). The many roles of histone deacetylases in development and physiology: implications for disease and therapy. Nat. Rev. Genet. *10*, 32–42.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074–1080.

Joshi, P., Greco, T.M., Guise, A.J., Luo, Y., Yu, F., Nesvizhskii, A.I., and Cristea, I.M. (2013). The functional interactome landscape of the human histone deacetylase family. Mol. Syst. Biol. *9*, 672.

Katan-Khaykovich, Y., and Struhl, K. (2002). Dynamics of global histone acetylation and deacetylation in vivo: rapid restoration of normal histone acetylation status upon removal of activators and repressors. Genes Dev. *16*, 743–752.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705.

Lucchesi, J.C., and Kuroda, M.I. (2015). Dosage compensation in Drosophila. Cold Spring Harb. Perspect. Biol. *7*, a019398.

Makowski, A.M., Dutnall, R.N., and Annunziato, A.T. (2001). Effects of acetylation of histone H4 at lysines 8 and 16 on activity of the Hat1 histone acetyltransferase. J. Biol. Chem. *276*, 43499–43502.

Morgan, M.A., and Shilatifard, A. (2015). Chromatin signatures of cancer. Genes Dev. *29*, 238–249.

Morinière, J., Rousseaux, S., Steuerwald, U., Soler-López, M., Curtet, S., Vitte, A.-L., Govin, J., Gaucher, J., Sadoul, K., Hart, D.J., et al. (2009). Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. Nature *461*, 664–668.

Nguyen, U.T., Bittova, L., Müller, M.M., Fierz, B., David, Y., Houck-Loomis, B., Feng, V., Dann, G.P., and Muir, T.W. (2014). Accelerated chromatin biochemistry using DNA-barcoded nucleosome libraries. Nat. Methods *11*, 834–840.

Olsen, J.V., and Mann, M. (2013). Status of large-scale analysis of post-translational modifications by mass spectrometry. Mol. Cell. Proteomics *12*, 3444–3452.

Parthun, M.R. (2012). Histone acetyltransferase 1: more than just an enzyme? Biochim. Biophys. Acta (3-4), 256–263.

Prestel, M., Feller, C., and Becker, P.B. (2010). Dosage compensation and the global re-balancing of aneuploid genomes. Genome Biol. *11*, 216.

Ruthenburg, A.J., Li, H., Patel, D.J., and Allis, C.D. (2007). Multivalent engagement of chromatin modifications by linked binding modules. Nat. Rev. Mol. Cell Biol. *8*, 983–994.

Seto, E., and Yoshida, M. (2014). Erasers of histone acetylation: the histone deacetylase enzymes. Cold Spring Harb. Perspect. Biol. *6*, a018713.

Smith, E., and Shilatifard, A. (2010). The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes. Mol. Cell *40*, 689–701.

Stasevich, T.J., Hayashi-Takanaka, Y., Sato, Y., Maehara, K., Ohkawa, Y., Sakata-Sogawa, K., Tokunaga, M., Nagase, T., Nozaki, N., McNally, J.G., and Kimura, H. (2014). Regulation of RNA polymerase II activation by histone acetylation in single living cells. Nature *516*, 272–275.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41–45.

Straub, T., and Becker, P.B. (2011). Transcription modulation chromosome-wide: universal features and principles of dosage compensation in worms and flies. Curr. Opin. Genet. Dev. *21*, 147–153.

Suganuma, T., and Workman, J.L. (2011). Signals and combinatorial functions of histone modifications. Annu. Rev. Biochem. *80*, 473–499.

Toyama, B.H., Savas, J.N., Park, S.K., Harris, M.S., Ingolia, N.T., Yates, J.R., 3rd, and Hetzer, M.W. (2013). Identification of long-lived proteins reveals exceptional stability of essential cellular structures. Cell *154*, 971–982.

Tran, J.C., Zamdborg, L., Ahlf, D.R., Lee, J.E., Catherman, A.D., Durbin, K.R., Tipton, J.D., Vellaichamy, A., Kellie, J.F., Li, M., et al. (2011). Mapping intact protein isoforms in discovery mode using top-down proteomics. Nature *480*, 254–258.

Turner, B.M. (2000). Histone acetylation and an epigenetic code. BioEssays *22*, 836–845.

Van Kampen, N.G. (2007). Stochastic Processes in Physics and Chemistry Elsevier (North-Holland).

Verreault, A., Kaufman, P.D., Kobayashi, R., and Stillman, B. (1996). Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. Cell *87*, 95–104.

Voss, A.K., and Thomas, T. (2009). MYST family histone acetyltransferases take center stage in stem cells and development. BioEssays *31*, 1050–1061.

Yang, X.-J., and Seto, E. (2007). HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. Oncogene *26*, 5310–5318.

Zhang, K., Williams, K.E., Huang, L., Yau, P., Siino, J.S., Bradbury, E.M., Jones, P.R., Minch, M.J., and Burlingame, A.L. (2002). Histone acetylation and deacetylation: identification of acetylation and methylation sites of HeLa histone H4 by mass spectrometry. Mol. Cell. Proteomics *1*, 500–508.

Zheng, Y., Sweet, S.M.M., Popovic, R., Martinez-Garcia, E., Tipton, J.D., Thomas, P.M., Licht, J.D., and Kelleher, N.L. (2012). Total kinetic analysis reveals how combinatorial methylation patterns are established on lysines 27 and 36 of histone H3. Proc. Natl. Acad. Sci. USA *109*, 13549–13554.