# nucleosomes – gatekeepers of the genome?

## Computer simulations and high throughput sequencing reveal how genome access is controlled

by Karsten Rippe and Gero Wedemann

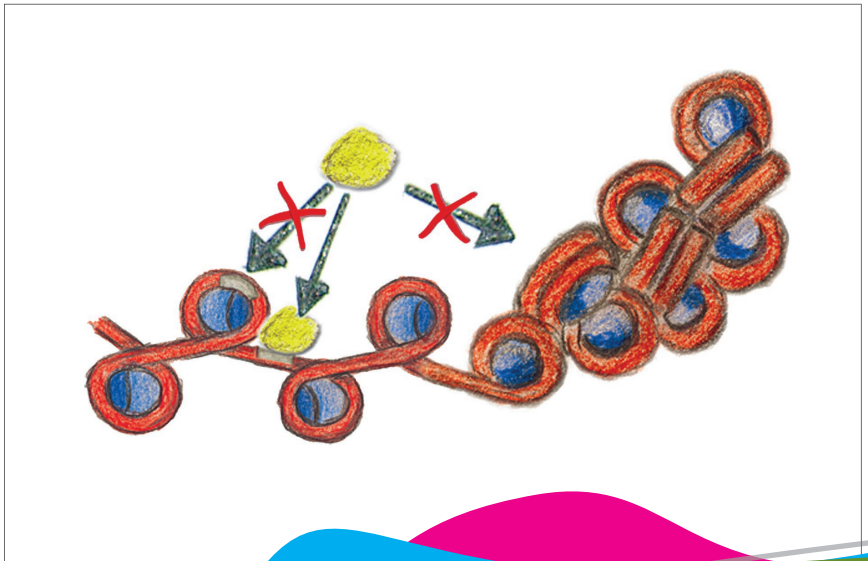
The human genome is packaged into a chain of nucleosomes — complexes of a protein core with DNA wrapped around. Many transcription factors can bind much better to the linker DNA between nucleosomes, and thus their positioning can directly determine access to the genome. Nucleosomes control genome access also indirectly: a regular spacing promotes folding of the chain into higher order fibers in which also the linker DNA can be occluded. Thus, predicting gene expression from the binding of transcription factors becomes dependent not only on DNA sequence and protein concentration but also on the spatial organization of the genome by nucleosomes.

### What chapters to read in the book of DNA?

All the cells of the human body contain essentially the same DNA sequence and with it the identical genetic information. The DNA sequence encodes the construction and maintenance plan of the organism. However, each cell 'reads' only selected chapters of the bodies 'DNA book', and uses only some of its plans and recipes to exert specific functions. But how does a undifferentiated stem cell establish the appropriate DNA program during development that makes it for example a muscle cell, a liver cells or a skin cell? One of the sophisticated mechanisms involved in this process is the packaging of the DNA genome in the cell nucleus into a structure called *chromatin* (the term originates from the Greek word *chroma*, i.e. color, because it can be easily stained for imaging by light microscopy): The chromatin in a human cell nucleus contains DNA chains that are almost 2 meters in length when put together. About ¾ of this DNA is part of the ~30 million nucleosomes that consist of DNA wrapped around a histone protein core in almost two turns.

Nucleosomes are connected by segments of protein-free linker DNA between them. As depicted in Fig. 1, DNA within the nucleosome is less accessible for other protein factors than linker DNA. Hence, the positions of the nucleosomes determine whether certain DNA sequences are in the more easily acces-

Figure 1: Nucleosome chain with a protein binding to the linker DNA between two nucleosomes



Nucleosome chain with a transcription factor (in yellow) binding to the linker DNA (red) between the nucleosomes (protein core in blue). The open 'beads-on-a-string' conformation of a nucleosome chain can associate into a chromatin fiber. Thus, both the position of a nucleosome along the DNA, as well as the higher order folding of the nucleosome chain, control access to the DNA (Source: Gernot Längst).

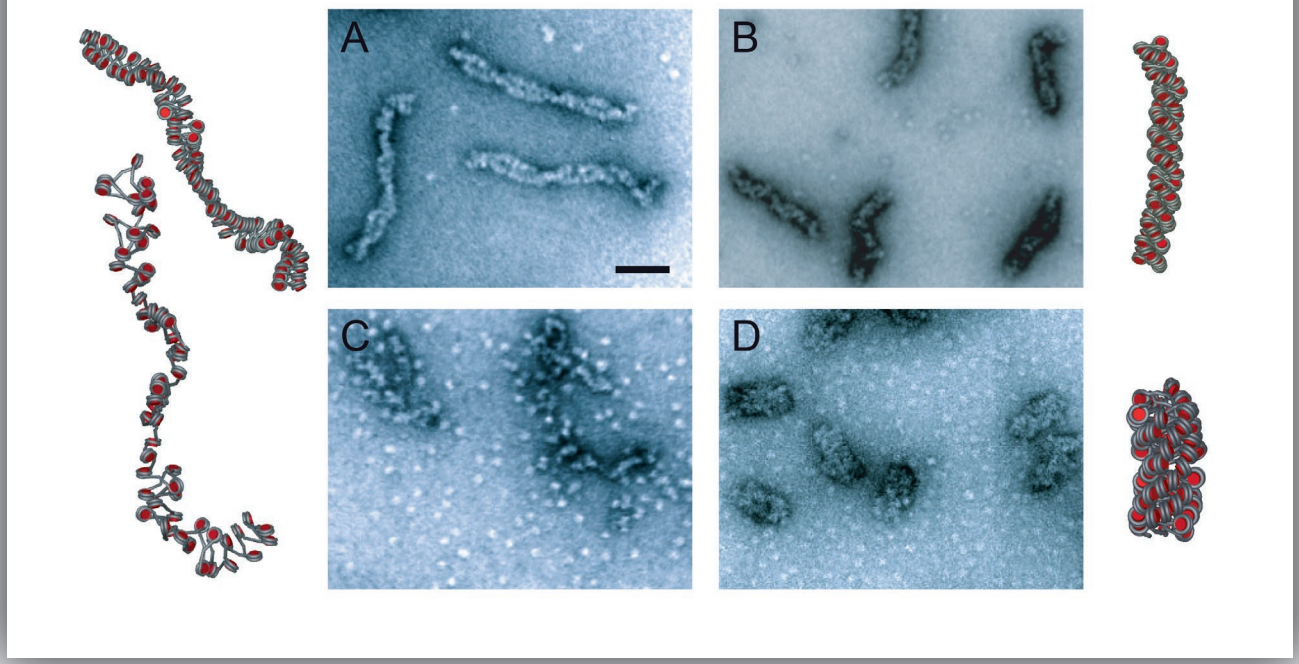



**Figure 2: Chromatin fibers observed on electron microscopy images and model structures from Monte Carlo computer simulations**
Samples were different with respect to the spacing of nucleosomes, i.e. the nucleosome repeat length (NRL) and the presence or absence of the linker histone H1. Scale bar: 50 nm.
**A)** Unfolded 167-bp NRL fiber in the absence of linker histone H1.
**B)** Addition of H1 induces the folding of the 167-bp NRL fiber.
**C)** Unfolded 197-bp NRL nucleosome chain with more irregular structures in the absence of linker histone.
**D)** Fully folded 197-bp NRL fibers form in the presence of saturating linker histone concentrations.
[Source: Daniela Rhodes (electron microscopy images), Nick Kepper (chromatin fiber simulation images)].

sible DNA linker region between nucleosomes or are masked by histone-DNA interactions. Thus, differences in nucleosome position, in conjunction with other regulatory mechanisms, decide whether genes are active or inactive and different cellular functions can be selected.

## From models to experiments and back

Our theoretical studies at the level of a single nucleosome revealed a complex linkage between the dynamic structure of the nucleosome, DNA interactions of the unstructured tails of histone proteins and the binding of transcription factors (Ettig *et al.*, 2011; Teif *et al.*, 2010). Likewise, we developed and applied coarse-grained models to predict accessibility of the linker DNA as well as partial unwrapping of nucleosomal DNA for larger chains of up to 1000 nucleosomes (Rippe *et al.*, 2012). These models were designed and parameterized with experimental data obtained from the analysis of short nucleosome chains with regular nucleosome spacing studied *in vitro*. As illustrated in Fig. 2, the comparison of fiber structures observed experimentally by electron microscopy for different nucleosome repeat lengths and their predicted conformation from the Monte Carlo computer simulations is very good.

However, until recently it was not possible to investigate the linkage between nucleosome positions and transcription factor binding in a systematic manner for mammalian genomes. Rather, most studies focused on the analysis of simple model organisms like yeast and the fruit fly *Drosophila* or were restricted to studies of nucleosome positions at a few selected genes. With recent advancements in high-throughput DNA sequencing methods it became possible to map all individual nucleosome positions in mammalian genome at single base-pair resolution and to link models and experiments (Fig. 3). Taking advantage of this exciting new possibility the *EpiGenSys* consortium supported by the BMBF set out to map for the first time all nucleosome positions in mouse embryonic stem cells in comparison to their differentiated counterparts (Teif *et al.*, 2012). The results revealed a wealth of features of nucleosome positioning at DNA sites that are important for cell differentiation. The start and the end of active genes displayed nucleosome-depleted regions as compared to silent genes where these sites were covered with nucleosomes.

Interestingly, the nucleosome positioning profiles were found to change according to specific modifications of the histone proteins. These modifications include posttranslational modifications with methyl and acetyl groups that can be attached and removed from the histones and represent so-called 'epigenetic' signals that serve to transmit a certain chromatin state through cell division.

As predicted, many proteins that play a central role for cell development were indeed found to bind to the free linker DNA between nucleosomes in embryonic stem cells, supporting the gatekeeper model for the role of nucleosomes in regulating DNA access. However, some proteins like three of the master regulators of the stem cell state, the proteins Nanog, Oct4 and Sox2 showed a different binding pattern. They were found also at regions with high nucleosome occupancy, suggesting that they act as 'pioneering' factors that can bind to the nucleosomal DNA. Thus, the simplified view that nucleosomes always impede transcription factor binding needs to be revisited.

An additional challenge for modeling work that uses the experimental nucleosome positions as input for the Monte Carlo simulations for 3D genome folding arises from the heterogeneity of the cell populations studied. In many instances the experimentally determined peaks that demarcate nucleosome positions were incompatible with a single nucleosome chain conformation. To obtain specific nucleosome configurations and optimized solutions for the complex positioning patterns from experimental data, we developed a scheme that combines binary-variable analysis and a Monte Carlo approach with a simulated annealing method (Schöpflin *et al.*, 2013). With this framework we are now conducting a new iteration of computer simulations based on the experimental data from the population average for specific genomic loci of up to 1000 nucleosomes to investigate their spatial organization and DNA accessibility (Fig. 3).

## Where to go from here?

Nucleosome positioning is a dynamic process that is regulated in the cell by energy consuming molecular machines that are referred to as chromatin remodeling complexes (CRCs). They can move or evict nucleosomes along the DNA chain to switch between 'on' and 'off' states of the DNA. Interestingly, genome-wide sequencing studies, as conducted for example in the International Cancer Genome Consortium (http://dcc.icgc.org), reveal a surprisingly high number of mutations in different CRC types in cancer. Potential 'driver' mutations were identified for members of the CHD and SMARCA type CRC families in breast, ovarian, kidney, lung, colon, uterus and liver cancer. Furthermore, it is known that aberrant epigenetic patterns can misguide CRC activity to establish pathological chromatin states in tumors.

Thus, important new questions emerge from the insights already gained within the *EpiGenSys* project: How is active nucleosome translocation regulated in the cell? Our current model proposes that signals targeting CRCs and modifying their activity are encoded by the nucleosomal DNA sequence and the posttranslational modifications of histones H3 and H4, as well as other chromatin features like DNA cytosine methylation. It appears that a yet unknown 'chromatin remodeling code' exists that targets CRC to genome sites, where they change nucleosome positions. In this manner, CRCs can establish distinct local chromatin structures, rather than solely being unspecific nucleosome-moving entities that make chromatin more 'fluid'. In addition, our unpublished modeling work points to another linkage between nucleosome positions and chromatin structure: Already by changing a single nucleosome position the higher order folding of the nucleosome chain can be dramatically altered. Thus, local nucleosome positions modulate access to the linker DNA via changing the chromatin fiber conformation and, at the same time, can also promote or impede long-range interactions between proteins bound at distant sites.

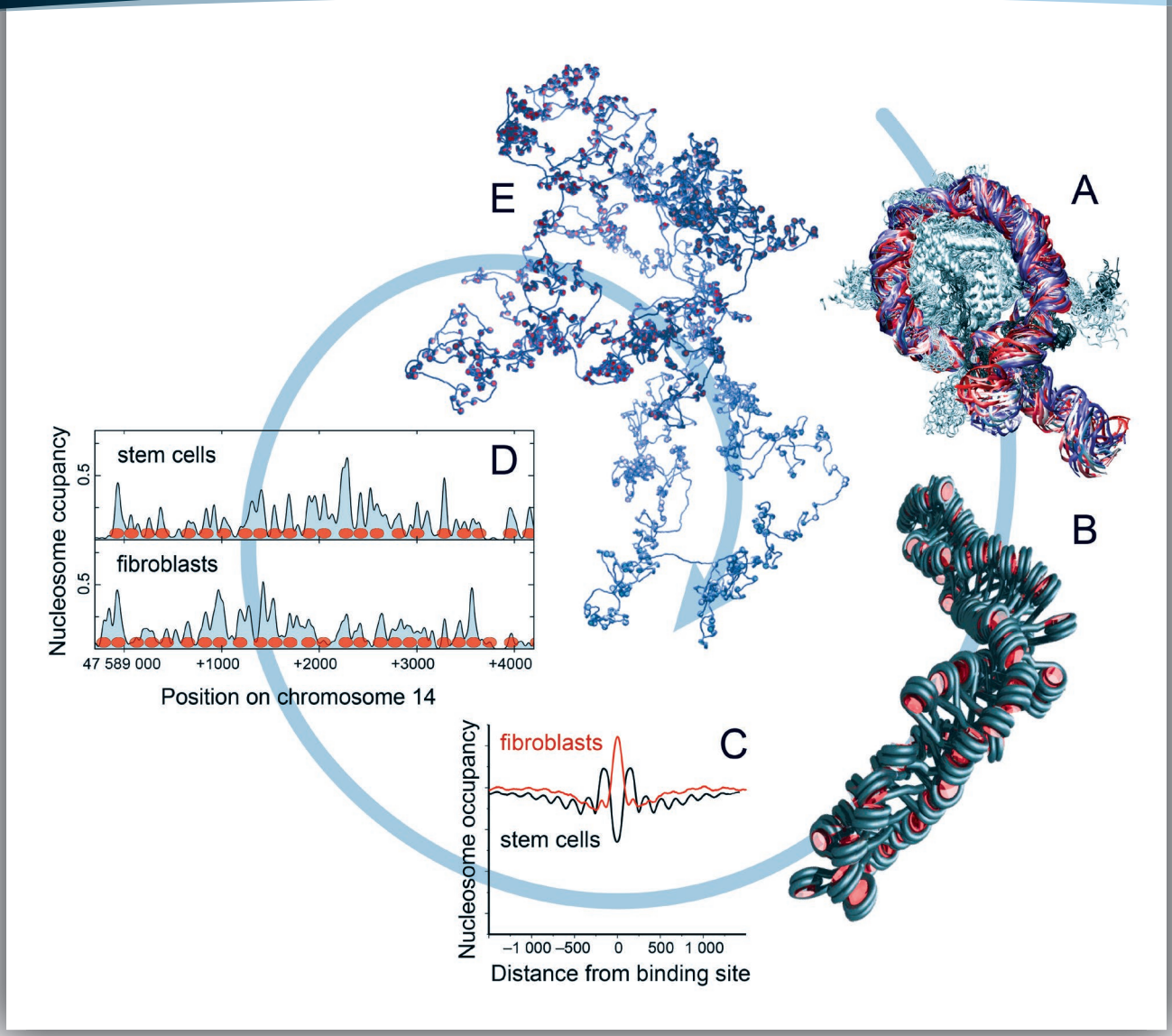



**Figure 3: Iterations between modeling and experimental studies dissect the relation between nucleosomes and DNA accessibility for transcription factor binding**

**A)** By molecular dynamics simulation access of nucleosomal DNA was studied with respect to the competitive DNA binding of the histone protein core and transcription factors (Ettig *et al.*, 2011; Teif *et al.*, 2010). In the image nucleosome conformations are overlayed in 0.2 nanosecond time intervals. The DNA is color coded with increasing simulation time from red to white to blue. The core histone proteins are shown in white. Already during the very short simulation time period of 2 nanoseconds, the nucleosome conformation is very dynamic.

**B)** Monte Carlo simulations of nucleosome chain folding provide information on access to linker DNA in chromatin fibers [reviewed in (Rippe *et al.*, 2012)].

**C)** Experimentally mapped nucleosome occupancy profiles at binding sites of the CTCF transcription factor (Teif *et al.*, 2012). In embryonic stem cells, bound CTCF introduces a regular spacing of nucleosomes in the region flanking its binding site. In differentiated fibroblast cells, nucleosomes occupy some of the sites that had CTCF bound in stem cells.

**D)** Unique nucleosome positions were extracted from experimental data sets for subsequent modeling of chromatin at specific genomic loci (Schöpflin *et al.*, 2013).

**E)** Monte Carlo simulation of chromatin organization at the 200 kb large SAMD4 gene locus based on experimental nucleosome position maps. (Source: Robert Schöpflin)

Given the importance of nucleosome positioning for genome access and readout, another exciting question is about deregulation of nucleosome positioning in cancer cells. This issue is addressed within the work of the consortium *CancerEpiSys* (www.CancerEpiSys.org) in the BMBF CancerSys program. Nucleosome positions and epigenetic modifications are mapped in primary tumor cells from patients with chronic lymphocytic leukemia to identify patterns of deregulated chromatin organization that can be related to the cancer disease state.

Based on the results obtained so far, it is emerging that integrating the various epigenetic signals attached to the genome with changes of DNA access provides valuable information for novel personalized medicine approaches. Thus, including nucleosome positions becomes an important part of quantitative descriptions that predict how cells select their active genetic program and that rationalize how this process is deregulated in cancer cells.

## The research project in brief:

The project on nucleosome and chromatin modeling as well as the experimental mapping of nucleosome positions was part of the three-year BMBF funded consortium project *EpiGenSys* – System Biological Determination of the Epigenomic Structure-Function Relation (www.EpiGenSys.org) within the European ERASysBioPlus initiative in the EU FP7 ERA-NET Plus program (grant numbers 0315712A to KR and 0315712C to GW).

In *EpiGenSys* (epi-)genomic structure-function relationships between chromatin and transcription were dissected and modeled with a multi-scale approach. The work integrated the single nucleosome structure, folding of the nucleosome chain into higher order fibers and the larges scale 3D architecture of the genome. Within the consortium, the groups of KR and GW conducted modeling work on the single nucleosome level with respect to accessing nucleosomal DNA (Ettig *et al.*, 2011; Teif *et al.*, 2010), developed coarse grained models for the nucleosome chain by Monte Carlo simulations [reviewed in (Rippe *et al.*, 2012)], mapped nucleosome positions experimentally by deep sequencing (Teif *et al.*, 2012) and developed bioinformatical approaches to link experimental data and modeling (Schöpflin *et al.*, 2013).

## References:


Ettig, R., Kepper, N., Stehr, R., Wedemann, G., und Rippe, K. (2011). Dissecting DNA-histone interactions in the nucleosome by molecular dynamics simulations of DNA unwrapping. Biophys J *101*, 1999 – 2008.

Rippe, K., Stehr, R., und Wedemann, G. (2012). Monte Carlo simulations of nucleosome chains to identify factors that control DNA compaction and access. Innovations in Biomolecular Modeling and Simulations, T. Schlick, ed. (Cambridge: Royal Society of Chemistry), 198 – 235.

Schöpflin, R., Teif, V. B., Müller, O., Weinberg, C., Rippe, K., und Wedemann, G. (2013). Modeling nucleosome position distributions from experimental nucleosome positioning maps. Bioinformatics *29*, 2380 – 2386.

Teif, V. B., Ettig, R., und Rippe, K. (2010). A lattice model for transcription factor access to nucleosomal DNA. Biophys J *99*, 2597 – 2607.

Teif, V. B., Vainshtein, Y., Caudron-Herger, M., Mallm, J.-P., Marth, C., Höfer, T., und Rippe, K. (2012). Genome-wide nucleosome positioning during embryonic stem cell development. Nat Struct Mol Biol *19*, 1185 – 1192.


## Contact:


**Karsten Rippe**
German Cancer Research Center
(DKFZ Heidelberg) & BioQuant
(Heidelberg University)
Head of Research Group
Genome Organisation & Function
Karsten.Rippe@dkfz.de
http://malone.bioquant.uni-heidelberg.de



**Gero Wedemann**
University of Applied Sciences Stralsund
Head of Institute for Applied
Computer Science
Head of Research Group
Competence Center Bioinformatics
gero.wedemann@fh-stralsund.de
www.bioinformatics.fh-stralsund.de