Statistical–mechanical lattice models for protein–DNA binding in chromatin

Vladimir B Teif\textsuperscript{1,2} and Karsten Rippe\textsuperscript{1}

\textsuperscript{1} Research Group Genome Organization and Function, Deutsches Krebsforschungszentrum and BioQuant, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany
\textsuperscript{2} Institute of Bioorganic Chemistry, Belarus National Academy of Sciences, Kuprevich 5/2, 220141, Minsk, Belarus

E-mail: Vladimir.Teif@bioquant.uni-heidelberg.de

Received 4 February 2010
Published 30 September 2010
Online at stacks.iop.org/JPhysCM/22/414105

Abstract

Statistical–mechanical lattice models for protein–DNA binding are well established as a method to describe complex ligand binding equilibria measured \textit{in vitro} with purified DNA and protein components. Recently, a new field of applications has opened up for this approach since it has become possible to experimentally quantify genome-wide protein occupancies in relation to the DNA sequence. In particular, the organization of the eukaryotic genome by histone proteins into a nucleoprotein complex termed chromatin has been recognized as a key parameter that controls the access of transcription factors to the DNA sequence. New approaches have to be developed to derive statistical–mechanical lattice descriptions of chromatin-associated protein–DNA interactions. Here, we present the theoretical framework for lattice models of histone–DNA interactions in chromatin and investigate the (competitive) DNA binding of other chromosomal proteins and transcription factors. The results have a number of applications for quantitative models for the regulation of gene expression.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Recent breakthroughs in high-throughput DNA sequencing technologies as an approach to identify protein-associated DNA fragments have created the unprecedented situation that genome-wide experimental data are being accumulated faster than the development of the corresponding biophysical models for gene regulation (Jiang and Pugh 2009, Segal and Widom 2009a, Radman-Livaja and Rando 2009, Cairns 2009). In eukaryotes, transcription factors (TFs) do not interact with free DNA, but rather compete with complex multi-component assemblies of histones and other chromosomal proteins for DNA binding sites. The elementary unit of chromatin is the nucleosome. It consists of a histone octamer complex containing two copies of each histone H2A, H2B, H3 and H4, around which 146–147 base pairs of DNA are wrapped in 1.67 turns. Nucleosomes are connected via the intervening linker DNA and the chain of nucleosomes folds into a chromatin fiber. Both the position of the nucleosomes and the fiber organization determine DNA accessibility by TFs and RNA polymerases (Boeger \textit{et al} 2008, Kim and O’Shea 2008, Lam \textit{et al} 2008, Petesch and Lis 2008, Gilmour 2009). Depending on the context, nucleosomes can inhibit (Whitehouse \textit{et al} 2007, Henikoff 2008) or facilitate (Zhao \textit{et al} 2001) transcription factor binding. Nucleosome positions are controlled by three major contributions: First, the intrinsic binding affinity of the histone octamer depends on the DNA sequence (Thåström \textit{et al} 1999, Segal \textit{et al} 2006, Whitehouse \textit{et al} 2007, Whitehouse and Tsukiyama 2006, Ioshikhes \textit{et al} 2006, Peckham \textit{et al} 2007). Second, the nucleosome can be displaced or recruited by the competitive or cooperative binding of other protein factors (Wasson and Hartemink 2009, Workman and Kingston 1992, Morozov \textit{et al} 2009). Third, the nucleosome may be actively translocated by ATP-dependent remodeling complexes (Teif and Rippe 2009, Zhang \textit{et al} 2009, Schnitzler 2008, Radman-Livaja and Rando 2009, Whitehouse and Tsukiyama 2006, Hartley and Madhani 2009, Rippe \textit{et al} 2007).

Most current theoretical descriptions assume that gene regulatory events are governed by mass-action laws and can be approximated by a thermodynamic equilibrium description (Bintu \textit{et al} 2005a, 2005b). Accordingly, DNA occupancies of transcription factors and RNA polymerase are related to gene expression probabilities as functions of protein concentrations.
(Teif and Bohinc 2010, Teif 2010). Several approaches have been reported to predict gene expression from the DNA sequence using these concepts with promising results, but currently the contribution of the chromatin organization is not accounted for (Beer and Tavazoie 2004, Segal et al. 2008). Since DNA is characterized by the sequence of its constituting base pairs, it can be represented by a one-dimensional lattice. The complexity of the associated protein interaction network in chromatin is reflected in different types of models. These are usually constructed to investigate the relation between protein composition and DNA sequence, localization of proteins and the biological activity of chromatin in terms of gene expression. The most basic description would comprise only the linker DNA and the histone octamer occupying binding sites of 147 bp of the DNA chain in the nucleosome. The nucleosome position can be changed by energy-dependent molecular motors referred to as chromatin remodelers (see Teif and Rippe 2009 and references therein). The models might consider the existence of subnucleosomal particles in which not all eight histones are present, the linker histones H1 that can organize an additional 20 bp of the linker DNA at the nucleosome, and/or the exchange of canonical core histones with variants (reviewed in Rippe et al. 2008). In addition to histones, other chromosomal proteins such as HMG and heterochromatin protein 1 (HP1) may assemble along the DNA lattice and act as important architectural components for the folding of the chain. Finally, gene specific transcription factors may display competing or cooperative binding with the above chromatin protein components at regulatory DNA regions such as promoters or enhancers. Thus, depending on the questions to be addressed, specific lattice models need to be devised. Here we define a theoretical framework that can be adapted to account for all the above scenarios, and then apply it to the interdependence of histone and TF binding to DNA in the context of gene regulation.

2. Lattice models for DNA–protein binding

2.1. General considerations

Chromatin has a complex three-dimensional organization. The DNA is wrapped around the histone octamer core and the resulting nucleosome chain can fold into a variety of conformations (Kepper et al. 2008). However, all protein binding events can be assigned to the linear sequence of DNA base pairs, which allows construction of a one-dimensional representation of the binding sites. The classical principles of calculation of macromolecule binding to a one-dimensional DNA lattice were formulated several decades ago (Poland 1979, Hill 1985). In mathematics, this field is known as ‘sequential adsorption’ as well as ‘car parking’ problems (Evans 1993). In computational biology such models are formulated with the help of Markov chains (e.g. hidden Markov models, HMMs), which go back to the works of the mathematician Andrei Markov (Markov 1907). In biophysics they are usually known as Ising models, historically arising from the work of Ernst Ising on the theory of ferromagnetism (Ising 1925). The general concept is to divide the system into the elementary units that are associated with different states (e.g. bound/unbound). The states of the whole system are then calculated as different combinations of states of elemental units. The elementary DNA units are usually taken as nucleotides A, T, C and G that represent one DNA strand in the double helix with base pairs A=T, T=A, C=G and G=C. Other choices of the elementary lattice unit are also possible as discussed below.

Statistical–mechanical lattice models are usually based on the assumptions that proteins bind DNA reversibly, that binding affinities are determined by the DNA and protein sequence, and that protein activities are proportional to their concentrations. The binding probabilities calculated for the lattice models are then equal to the fraction of molecules in a given state within the statistical ensemble consisting of many identical systems. Within a statistical ensemble, each state of the system is characterized by a weight \( \exp(-\Delta G_i/k_B T) \), where \( \Delta G_i \) is the energy difference of this state with respect to some reference state. In particular, when protein binding to the DNA is considered, the unbound state is usually set as a reference state with unit weight. Each bound state is then characterized by a weight \( K_{ng} c_n \), where \( K_{ng} \) is the binding constant for the protein of type \( g \) to the DNA site \( n \), and \( c_n \) is the concentration of free \( g \)-type protein in solution. The linear dependence on concentration follows from the dilute solution assumption that the protein activity is proportional to the concentration. (As said above, this assumption may not hold in chromatin, which may in principle lead to nonlinear weights—a possibility which has not been tested yet.) The partition function \( Z \) is equal to the sum of the weights of all possible states:

\[
Z = \sum_i e^{-\Delta G_i/k_B T}. \tag{1}
\]

The probability \( P_g \) of a given state \( x \) is the ratio of the weight of a given state and the partition function \( Z \):

\[
P_g = \frac{e^{-\Delta G_g/k_B T}}{Z}. \tag{2}
\]

If a weight \( K_x \) corresponding to the state \( x \) enters the partition function linearly, the following expression holds for the probability of this state \( P_x \):

\[
P_x = \frac{\partial Z}{\partial K_x} \times \frac{K_x}{Z}. \tag{3}
\]

In particular, the probability of binding of a protein of type \( g \) to a DNA site \( n \) is determined by the derivative of the partition function by the corresponding binding constant \( K_{ng} \) (Teif 2007):

\[
P_{ng} = \frac{\partial Z}{\partial K_{ng}} \times \frac{K_{ng}}{Z}. \tag{4}
\]

One-dimensional models may be formulated and solved mathematically using different algorithms, which are more or less computationally effective. Several methods of constructing lattice models of DNA–protein binding have been developed in the past. They may be roughly divided into four classes (table 1): direct combinatorial methods (Zasedatelev et al. 1971, McGhee and von Hippel 1974, Tsuchiya and Szabo 1979, McGhee and von Hippel 1974).
Table 1. Basic features of DNA–protein binding and lattice approaches.

| Feature/method                  | Combinatorial method | Generating functions | Transfer matrices | Recurrent relations | Graphical representation |
|---------------------------------|----------------------|----------------------|-------------------|--------------------|--------------------------|
| Sequence specificity            | −                    | +                    | +                 | +                  |                          |
| Overlapping binding sites       | +                    | −                    | −                 | +                  |                          |
| Multiprotein competition        | −                    | −                    | −                 | −                  |                          |
| Contact interactions            | +                    | +                    | +                 | +                  |                          |
| Long-range interactions         | +                    | +                    | +                 | +                  |                          |
| Multilayer binding              | −                    | +                    | +                 | −                  |                          |
| Short DNA loops                 | −                    | −                    | −                 | +                  |                          |
| Long DNA loops (e.g. promoter–enhancer) | −                    | −                    | −                 | −                  |                          |

2.2. Combinatorial method

The combinatorial method uses binomial formulae to derive analytical expressions for the numbers of possible rearrangements of proteins along the DNA (McGhee and von Hippel 1974, Zasedatelev et al 1971). For example, a protein which covers m base pairs upon binding to the DNA may adopt $(N - m + 1)$ positions along the DNA double helix of length $N$. This yields $[(N - k \times (m - 1))]/(N - m \times k)!/k!$ possible rearrangements of $k$ proteins. For a nonspecific binding to infinitely long DNA ($N \gg n$), the degree of protein binding $\nu = k/N$ is linked to its bulk concentration $c_0$, binding site size $m$ and binding constant $K$ according to the following relation (McGhee and von Hippel 1974):

$$\frac{\nu}{c_0} = K (1 - \nu) \left( \frac{1 - mv}{1 - (m - 1)\nu} \right)^{m-1}.$$  \hspace{1cm} (5)

Exact analytical solutions such as the McGhee–von Hippel formulae are widely used for the description of in vitro binding experiments. More specific and more general solutions also exist, which consider cooperative interactions between proteins (Zasedatelev et al 1971), polarity of protein–protein interactions (Nechipurenko et al 1979, Wolfe and Meehan 1992), competitions between specific and nonspecific binding on a short DNA oligomer (Teif et al 2001), electrostatic interactions between the molecules upon binding to the DNA (Rouzina and Bloomfield 1997), ligand-induced double-helix melting (Lando 1994), DNA condensation induced by binding of multivalent counterions (Maltsev et al 2006, Teif 2005) and binding of flexible branched oligopolymers (Horsky 2008, Nishio and Shimizu 2005).

2.3. Generating function method

The generating function method is based on the idea of characterizing the system by a mathematical expression called the generating function instead of the partition function (Lifson 1964, Schellman 1974, Kong 2001). The generating function is defined by

$$G(c_0, x) = \sum_{n=0}^{\infty} Z_n(c_0) x^n.$$  \hspace{1cm} (6)
Here $G(c_0, x)$ is the generating function of the system characterized by the partition functions $Z_N(c_0)$, where $n$ is the number of DNA-bound proteins in a given configuration, and $c_0$ stands for all input concentrations (if there is only one protein type, then $c_0$ has the same meaning of a free protein concentration as in equations (1)–(5)). Variable $x$ has no direct physical meaning. The idea is that during the algebraic manipulations $x$ should vanish and infinite series converge to well defined analytical limits. In the case of small non-interacting ligands binding to one unit of an infinite DNA lattice, the generating function is given by equation (7) (Kong 2001)

$$G(c_0, x) = \frac{1}{1 - (1 + K c_0)x} = \sum_{n=0}^{\infty} (1 + K c_0)^n x^n. \quad (7)$$

This yields a trivial solution for the partition function with $Z = (1 + K c_0)^n$. If a protein covers $m$ lattice units upon binding to the infinitely long DNA, the generating function method results in McGhee–von Hippel expression (equation (5)). This method is also applicable to more complex modes of binding. For example, it successfully allows treating the ‘piggy-back’ binding problem, when a protein bound to the DNA can bind other proteins on its back (Chen et al. 1986). At first, the generating function approach seemed to be inapplicable for the case of long-range interactions (Chen 1987). Later studies demonstrated that it can be applied to long-range cooperativity (Kong 2006, Lando and Nechipurenko 2008). However, it is difficult to implement if more than one type of large proteins exists in the system (Chen 1990).

2.4. Recurrent relation method

The idea of the recurrent relation method is that a partition function $Z_N$ of a DNA lattice of length $N$ can be expressed in terms of the partition function of a smaller lattice (Gurskii and Zasedatelev 1978). In particular, the following recurrent expressions hold for non-cooperative protein binding (Nechipurenko et al. 2005):

$$Z_N = Z_{N-1} + K_{N-m+1}c_0 Z_{N-m}; \quad Z_N = 1 \text{ if } N < m; \quad Z_m = K c_0 + 1. \quad (8)$$

For nonspecific protein binding to a DNA of infinite length, equations (8) converge to the McGhee–von Hippel expression (equation (5)). More general recurrent relations including equations (8) as a specific case have been derived for sequence-specific binding with long-range interactions by Gurskii and Zasedatelev (1978). Their algorithm is restricted to a maximum protein–protein interaction distance less than the protein length, which is reasonable in the case of a nucleosome chain. Many dynamic programming approaches in computational biology use an approach similar to that of Gurskii and Zasedatelev: the partition function is derived from a recursive series of simpler calculations. Recent examples include the COMPETE algorithm (Wasson and Hartemink 2009), which provides a fast way to calculate genome-wide binding of several protein species at the cost of restricting the model only to competitive binding, the GOMER algorithm (Granek and Clarke 2005), which allows treating long-range interactions between a protein and a DNA promoter but not between two DNA-bound proteins, and the algorithm developed by Segal and co-workers for competitive binding of TFs and nucleosomes (Raveh-Sadka et al. 2009, Segal et al. 2006), as well as algorithms to describe protein–protein interactions (Laurila et al. 2009, Lubliner and Segal 2009).

2.5. Transfer matrix method

The transfer matrix formalism was initially used to obtain analytical expressions for simple homopolymer systems (Hill 1957). However, it allows treating DNA sequence specificity as well (Crothers 1968, Gurskii et al. 1972). The method is based on the construction of transfer matrices (weight matrices) for each DNA unit. Each transfer matrix element $Q_a(i, j)$ contains the probabilities of all binding events (Gurskii and Zasedatelev 1978). Their algorithm is restricted to a maximum protein–protein interaction distance less than the protein length, which is reasonable in the case of a nucleosome chain. Many dynamic programming approaches in computational biology use an approach similar to that of Gurskii and Zasedatelev: the partition function is derived from a recursive series of simpler calculations. Recent examples include the COMPETE algorithm (Wasson and Hartemink 2009), which provides a fast way to calculate genome-wide binding of several protein species at the cost of restricting the model only to competitive binding, the GOMER algorithm (Granek and Clarke 2005), which allows treating long-range interactions between a protein and a DNA promoter but not between two DNA-bound proteins, and the algorithm developed by Segal and co-workers for competitive binding of TFs and nucleosomes (Raveh-Sadka et al. 2009, Segal et al. 2006), as well as algorithms to describe protein–protein interactions (Laurila et al. 2009, Lubliner and Segal 2009).

| State # (unit $i \setminus$ unit $i+1$) | 1 | 2 | 3 | ... | $m$ | $m + 1$ | $m + 2$ | $m + 3$ |
|--------------------------------------|---|---|---|-----|----|-------|-------|-------|
| 1 (bound)                           | $c_0 K_0$ | ... |
| 2 (bound)                           | 1 | ... |
| 3 (bound)                           | ... | ... |
| ...                                 | ... | ... | ... | ... | ... | ... | ... | ... |
| $m - 1$ (bound)                     | ... | ... | ... | ... | ... | ... | ... | ... |
| $m$ (bound)                         | 1 | ... | ... | ... | 1 | 1 |
| $m + 1$ (free left end)             | 1 | ... | 1 | 1 |
| $m + 2$ (free right end)            | ... | 1|
| $m + 3$ (free inside)               | 1 | ... | 1 |

Figure 1. A transfer matrix for a model with a single protein type reversibly binding DNA.
Previous studies have shown that the transfer matrix method allows treatment of sequence specificity (Crothers 1968), long-range interactions between bound proteins (Chen 1987), binding of many protein types (Akhrem et al 1985, Chen 2004), double-helical asymmetry of the DNA (Di Cera and Kong 1996), multilayer protein assembly associated with DNA looping (Teif 2007) and the competition with nucleosomes (Teif and Rippe 2009). In addition, this method allows an exact solution of complicated boundary conditions, which is nontrivial at short regulatory DNA regions (Teif 2010).

3. Lattice models for nucleosome arrangements along the DNA

One of the first lattice models explicitly taking into account nucleosome distributions along the DNA was constructed by Kornberg and Stryer (1988). Consider a long array of nucleosomes on the DNA. The number of nucleosomes is fixed, but their positions are variable. Combinatorial laws predict that nucleosomes, which do not have any DNA sequence preferences, are periodically arranged close to the boundaries of the DNA segment. Such boundary effects are not specific to nucleosomes and apply to any DNA–ligand binding (Epstein 1978, Di Cera and Phillipson 1996, Flyvbjerg et al 2006). Recent experimental studies (Mavrich et al 2008a, Kharchenko et al 2008, Cuddapah et al 2009, Milani et al 2009) and theoretical considerations (Vaillant et al 2009, Teif and Rippe 2009) confirm that this effect is indeed important in nucleosome positioning. For example, a region immediately upstream of the transcription start site at transcriptionally active promoters is usually depleted. It is not known whether this depletion is due to nucleosome-excluding sequences or the assembled transcription preinitiation complex. This depleted region forms a barrier, which determines the oscillatory nucleosome positioning pattern decaying at increasing distances (Mavrich et al 2008a, Kharchenko et al 2008). An even stronger barrier is imposed by the insulator binding protein CTCF, which can position about 10 nucleosomes (Cuddapah et al 2009). Twenty years ago, Kornberg and Stryer proposed that the boundary effects are the first order effect, while ‘the preferential binding of histones to certain sequences is a second order effect, whose influence upon neighboring nucleosomes is then of third order’ (Kornberg and Stryer 1988). Today the relative contributions of these effects are still under discussion (Jiang and Pugh 2009, Segal and Widom 2009a, Radman-Livaja and Rando 2009, Cairns 2009), but the importance of statistical positioning is being recognized (Mavrich et al 2008a, 2008b, Vaillant et al 2009).

In 1990, a purely entropic model of Kornberg and Stryer was extended to allow a variable number of nucleosomes on the DNA determined by the thermodynamic equilibrium as in the conventional DNA–ligand binding (Nechipurenko and Vol’kenshtein 1986, Nechipurenko 1988, Iovanovich and Nechipurenko 1990). This has led to the introduction of energetic parameters—the nucleosome binding constant and the effective nucleosome concentration. In that study, the parameters were chosen to obtain an average nucleosome–nucleosome distance equal to the linker length fixed in the model of Kornberg and Stryer. This approach was extended by the introduction of sequence-dependent nucleosome affinities, as is now actively used by several groups (Teif and Rippe 2009, Segal et al 2006, Schwab et al 2008, Chevereau et al 2009, Ioshikhes et al 2006) (figure 2).

In addition to the boundary- and sequence-determined positioning, the concept of concentration-dependent nucleosome switches has been put forward. Changes in the concentration of histones (or activity of histone remodeling complexes) can lead to abrupt switching between different patterns of the nucleosome arrangement (Schwab et al 2008, Teif and Rippe 2009). Bistable patterns of nucleosome positioning have indeed been reported for many intragenic regions (Chevereau et al 2009, Vaillant et al 2009). Concentration-dependent nucleosome switches are only possible if the total number of nucleosomes on the DNA is not fixed. In general, one should be aware that there are two principally different statistical–mechanical approaches to describe nucleosome arrangement on a one-dimensional DNA lattice. The first approach is based on the Ising model, as exemplified by the McGhee–von Hippel formalism for ligand–DNA binding (McGhee and von Hippel 1974). In this model, the concentration of ligands in the solution is fixed, while the number of ligands on the DNA is not fixed, and is determined by the thermodynamic equilibrium. The second approach is based on the so-called Tonks gas model as applied to DNA–ligand binding (Woodbury 1981). This model is based on the assumption that a fixed number of ligands is freely sliding along the DNA. The latter assumption was used by Kornberg and Stryer (1988), while other works have used the assumption of a variable number of nucleosomes (Iovanovich and Nechipurenko 1990, Schwab et al 2008, Segal et al 2006). In the limit of saturating concentrations (as is the case for the nucleosomes) these two approaches yield significantly different results (Woodbury 1981, Teif and Rippe 2009).

3.1. Interaction energies and concentrations

Binding energy and concentration are the two parameters central for lattice models of nucleosome chains. One way to determine nucleosome formation energies at physiological ionic strength is via a stepwise dilution of nucleosomes down to the concentrations at which the histone octamer
dissociates from the DNA. An apparent dissociation constant can be determined by assuming equilibrium between the complete nucleosome and the DNA/subnucleosomal species (Rippe et al 2008), which strongly depends on the ionic properties of the solution. From the early studies a value of $K_d = 3 \text{nM (}$Δ$G = 11.5 \text{kcal mol}^{-1}$) was derived for histone binding to bulk mouse DNA in a 0.15 M NaCl buffer (Cotton and Hamkalo 1981). A related approach was used to study nucleosome positioning at a strong positioning sequence from 5S rDNA, which gave $K_d = 0.2 \text{nM (}$Δ$G = 13.2 \text{kcal mol}^{-1}$) in the presence of 0.15 M NaCl (Gottesfeld and Luger 2001). However, it is noted that these types of experiments do not yield the true equilibrium dissociation constant, since free histones tend to aggregate irreversibly when dissociated from the nucleosome. In vivo nucleosome assembly via a complex of histones with histone chaperones such as NAP1 and Asf1 prevents unspecific interactions and aggregation. NAP1 binds the H2A·H2B dimer and the H3·H4 histone with the same affinity of approximately 10 kcal mol$^{-1}$ (Rippe et al 2008). Histones can be released from NAP1 via competitive binding to the DNA. It is estimated that the energy of binding of the last H2A·H2B dimer to a histone hexasome to form the nucleosome is around $−1.5 \text{kcal mol}^{-1}$ in the presence of NAP1, while it is around $−11.7 \text{kcal mol}^{-1}$ in the absence of histone chaperones (Mazurkiewicz et al 2006, Rippe et al 2008). The uncertainty of the absolute energies of histone–DNA interactions is an important issue, which is usually neglected in current computational approaches. These operate with relative changes of the binding affinities of the nucleosome core for different DNA sequences. This energy difference can vary from zero to $−2.4 \text{kcal mol}^{-1}$ depending on the DNA sequence in vivo, and up to $−4.1 \text{kcal mol}^{-1}$ for artificial octamer binding sequences in vitro (Thäström et al 2004).

Although the importance of the genomic sequence to nucleosome positioning was hypothesized quite a long time ago (Trifonov 1980), systematic calculations of nucleosome positioning on the genomic DNA started only recently, when a large enough pool of nucleosome binding affinities was collected (Ioshikhes et al 2006, Segal et al 2006). As a result of such studies, not only were nucleosome sites at known positions recovered, but also the probabilities of finding nucleosomes at many unknown positions were predicted and verified later. Several algorithms have been developed to predict nucleosome positioning from the DNA sequence (Segal et al 2006, Tolstorukov et al 2008, Gабдак и др. 2010, Ioshikhes et al 2006, Morozov et al 2009). Since the experimental nucleosomal DNA pool used as a computer-training set is increasing very fast, the brute force bioinformatics approach works even without knowing physical details about molecular interactions underlying observed nucleosome distributions. However, not all nucleosome sites can be predicted from the DNA sequence (Stein et al 2009, Zhang et al 2009), and for those which can be predicted in vivo occupancies are not always recovered (Morozov et al 2009, Teif and Rippe 2009). This might be explained by the fact that experimental nucleosome patterns are obtained from whole-genome sequencing of DNA isolated by in situ digestion with micrococcocal nuclease in the environment of the cell nucleus. Therefore, they also reflect to some extent contributions from the higher order chromatin structures, competing proteins and chromatin remodeling complexes (Teif and Rippe 2009).

Using high-resolution imaging techniques such as atomic force microscopy (AFM), it is possible to track individual nucleosomes on a given DNA in vitro (Bash et al 2001). A lattice model to analyze experimental nucleosome positioning data was developed by Solis and coauthors (Solis et al 2004). The authors determined the number of nucleosomes reconstituted by salt dialysis on DNA segments of known length containing multiple 5S rDNA sequences. In addition to energetic and entropic effects considered in the previous models, this model introduced nucleosome–nucleosome interactions as virial coefficients in the frame of the perturbation theory. From this, the energy of the histone octamer and nucleosome–nucleosome interactions were estimated. It appeared that both parameters were dependent on the acetylation state of the nucleosomes. For all studied DNA templates, histone acetylation decreased the affinity of core particles to the DNA and eliminated nucleosome–nucleosome attractions (Solis et al 2007). The energies of nucleosome–nucleosome interaction were estimated as $−0.25 \text{kcal mol}^{-1}$ (attraction) for unacetylated nucleosomes and $+0.15 \text{kcal mol}^{-1}$ (repulsion) for acetylated nucleosomes. This may partly explain why acetylated nucleosomes are often found at actively transcribed regions—it is easier to remove them. However, it is not clear what is the relevance of this effect in comparison with the action of acetylated histone tails as recruitment tags for remodelers (Choi and Howe 2009). In addition to tail-mediated nucleosome interactions, there are other levels of interactions mediated by linker histones H1 and stacking interactions of the nucleosomes regularly packed in the chromatin fiber (Krutchuk et al 2009). Single-molecule experiments using optical tweezers suggest a total energy of about 2 kcal mol$^{-1}$ for each nucleosome–nucleosome contact (Cui and Bustamante 2000).

Another important parameter for the calculation of the histone octamer occupancy is the effective concentration (or activity) of the histone octamer in the environment of the nucleus. As already pointed out, the stepwise assembly of histone dimer particles mediated by histone chaperones precludes a direct measurement of this parameter. The average nucleosome repeat length, NRL, is about 200 bp in human cells (other organisms are different (van Holde 1989)). One nucleosome covers $m = 147$ bp. Therefore the average degree of DNA coverage is given as $m/NRL \approx 0.75$, which results in a degree of binding $\nu = 0.75/147 \approx 0.0051$ (nucleosomes per bp). Substituting these values of $m$ and $\nu$ in the McGhee–von Hippel equation for nonspecific protein–DNA binding (equation (5)), we get $K_{c0} \approx 0.4$. The absolute value of the concentration $c_0$ depends on the absolute value of the binding constant $K$, which is very sensitive to the experimental conditions, as described above. On the other hand, the value of the product $K_{c0}$ gives an idea of the nucleosome behavior in all competitive binding processes. Current bioinformatics software usually treats $c_0$ and $K_{c0}$ as unknown scaling parameters, equal to unity by default (Kaplan et al 2009).
According to our estimate above, this default approximation is surprisingly not far from reality. However, when additional features such as the competition with transcription factors, activity of histone chaperones and chromatin remodelers need to be considered (as follows later) one should be more careful with the choice of parameters $K$ and $c_0$ for a more faithful representation of specific experimentally studied systems.

3.2. Partial assembly of nucleosome particles

It is well known that the nucleosome is a dynamic structure, which has several intermediate states. It can ‘breathe’ by partially unwrapping the DNA or exchanging histones with the solution. This allows for a transient access of transcription factors to intra- and internucleosomal DNA with the solution. This allows for a transient access of transcription factors to intra- and internucleosomal DNA with the solution. This allows for a transient access of transcription factors to intra- and internucleosomal DNA with the solution. This allows for a transient access of transcription factors to intra- and internucleosomal DNA with the solution. This allows for a transient access of transcription factors to intra- and internucleosomal DNA with the solution. This allows for a transient access of transcription factors to intra- and internucleosomal DNA.

The nucleosome particle has a histone octamer core that is organized into four dimers (two H2A · H2B and two H3 · H4). Dissociation of these dimers from the nucleosome leads to the formation of subnucleosomal particles (Zlatanova et al 2009, Gansen et al 2009). Thus, one has to be aware that the histone octamer is not a single entity and is not a usual ligand.

Nucleosome dissociation can be described by a lattice model as depicted in figure 3, in which four ligand types are introduced. Although only two types of physically distinguishable subnucleosomal particles H2A·H2B and H3·H4 exist, we need $f = 4$ ligand types to characterize the $f$-mer assembly model (Teif 2007). For example, a combination of subunits 1–2–3–4 characterizes a standard nucleosome, while combinations 1–2–3 and 2–3–4 characterize a hexasome (a nucleosome missing one H2A · H2B subunit), and a combination 2–3 characterizes a tetrasome (a nucleosome lacking two H2A · H2B subunits). A new nucleosome can start after an existing one, so the combination 4–1 is allowed. The combinations of subnucleosomal subunits other than the order of 1–2–3–4 are prohibited. A transcription factor TF is defined as a fifth ligand type. The ligand can precede or follow a nucleosome, so the contacts 5–1 and 1–5 are allowed. To make the nucleosome a preferred assembly unit, the contacts 1–2, 2–3 and 3–4 are given high statistical weights, while the other allowed contacts are characterized by unit weights. Due to the fact that subunits 2 and 3 have two cooperative contacts, they represent the most stable part of the nucleosome core in comparison with subunits 1 and 4. Obviously such a simple model cannot capture all complex interactions between the nucleosome subunits, but it allows prediction of the main feature: during the competitive binding, it is easier to displace either subunit 1 or 4, denoted H2A · H2B. This feature is in accordance with available experimental data (Mazurkiewicz et al 2006, Zlatanova et al 2009).

3.3. Nucleosome sliding and ATP-dependent chromatin remodelers

In addition to simple dissociation–reassociation reactions, the nucleosome can slide along the DNA without being disassembled. Nucleosome sliding may happen spontaneously with a low probability (Pennings et al 1991), or be directed by energy-consuming molecular motors, so-called chromatin remodelers (Whitehouse and Tsukiyama 2006). Nucleosome sliding is achieved by breaking one or several out of ~14 contacts keeping the DNA and the histone core together in the nucleosome (Längst and Becker 2004). For example, partial unwrapping of a small segment of the intranucleosomal DNA (e.g. 10–50 bp) can form a loop, which is subsequently propagated around the histone octamer protein core and thus repositions the nucleosome (Schiessel et al 2001, Längst and Becker 2004, Chou 2007, Mollazadeh-Beidokhti et al 2009). Nucleosomes can also invade territories occupied by other nucleosomes (Engebloom et al 2009), overlapping with each other and changing the canonical nucleosome structure (Sims et al 2008).

Chromatin remodelers deviate from the concept of standard ‘ligands’ even further. Unlike usual proteins binding DNA at a thermodynamic equilibrium, remodelers are energy-dependent molecular motors. While their binding/unbinding to the nucleosomes might be still described by simple equilibrium models, the final nucleosome repositioning seems to be more complex, depending on the context, epigenetic modifications and DNA sequence (Rippe et al 2007). A systematic description of the remodeler action has been proposed based on the concept of remodeler rules (Teif and Rippe 2009) (figure 4).

Initial nucleosome positioning in the absence of remodelers can be calculated using the standard lattice model in figure 2.
assuming that nucleosome positions are determined only by the intrinsic affinities. The effect of chromatin remodelers can be then calculated by iteratively redistributing the nucleosomes according to the remodeler rules until a new steady state is reached. The rule is a probability for a given remodeler to move a given nucleosome from position \( n \) by a step \( s \) to a new position \( n + s \) or \( n - s \). Three major classes of remodeler activities have been identified from this analysis: (i) the establishment of regular nucleosome spacing in the vicinity of a strong positioning signal acting as a boundary, (ii) the enrichment/depletion of nucleosomes through amplification of intrinsic DNA sequence encoded signals, and (iii) the removal of nucleosomes from high-affinity binding sites. The computational study has shown that both the remodeler action and intrinsic nucleosome affinities are important for nucleosome positioning in vivo (Teif and Rippe 2009). A next step should be to derive systematically such rules from the analysis of in vitro single nucleosome experiments and/or genome-wide sequencing. First attempts to solve this task have already started (Le et al 2009).

3.4. Linker histones and other architectural proteins

Transcription factors and architectural proteins can be also considered as ligands, which can reversibly bind and unbind DNA. In a system of one non-interacting DNA binder covering \( m \) DNA units, each DNA unit may be in \( \sim m \) states (figure 1). In a system of \( f \) types of interacting DNA binders, each DNA unit may be in \( \sim \sum (m_g + V_g) \) states, where \( m_g \) is the length of a protein of type \( g \), \( V_g \) is the maximum interaction distance for \( g \)-type proteins, \( g = 1, \ldots, f \) (Teif 2007). In this model the histone octamer can be considered as one ligand type covering \( m = 147 \) bp, linker histones as another ligand type covering \( \sim 20 \) bp, and non-histone chromosomal proteins or transcription factors as additional ligand types. H1 and typical transcription factors cover 10–20 bp, while Pol II covers around 30–50 bp. Different thermodynamic features could be assigned to a certain type of nucleosome that for example contains the H2A.Z histone variant instead of the canonical H2A histone.

For a description of linker histone binding a lattice model was developed, in which each nucleosome provided a single site for nonspecific H1 binding (Ishii 2000) (figure 5(A)). This model predicted that cooperative H1–H1 interactions could lead to the propagation of chromatin ‘opening’ characterized by the loss of H1 histones. The correlation length, defined as a distance over which the presence or absence of H1 proteins affects the binding of new linker histones, was reported to be up to several hundreds of nucleosomes if contact H1–H1 interactions were characterized by the energy 7.5 kcal mol\(^{-1}\), and up to several dozens of nucleosomes for the energy of 5 kcal mol\(^{-1}\). Analogous models were constructed for a highly abundant non-histone chromatin protein HMGB1, which can compete with H1 and cooperatively assemble on the DNA (Teif et al 2002). However, later experimental data did not support such strong H1–H1 interactions (Mamoon et al 2005).

Figure 5. Lattice models to incorporate linker histone proteins. (A) Each nucleosome represents one binding site for H1; H1 linker histones interact with each other by contact cooperativity. (B) H1 histones do not interact with each other; the cooperativity of binding arises due to different binding constants for H1-saturated and H1-depleted chromatin, which represent different phases.

4. Transcription factor binding to chromatin

Although a nucleosome can be partially unwrapped to allow TF binding (Shlyakhtenko et al 2009, Poirier et al 2009), nucleosomes and TF binding to the DNA may be mutually exclusive (Svaren et al 1994). In the latter case, a nucleosome pre-assembled at a given site cannot be displaced by TF unless chaperones or remodelers mediate the translocation of cooperativity, while it can be described by the two-state model with the H1 binding constant for H1-saturated ‘closed’ chromatin domains being ten times higher than for H1-depleted ‘open’ domains (Mamoon et al 2005) (figure 5(B)). In analogy, a two-state model with different affinities for TF binding to the free and nucleosomal DNA was considered recently (Mirny 2009). As in the case with H1, this model yields cooperative titration curves that are similar to the two-state models for ligand-induced DNA aggregation (Saroff 1991), conformational transitions (Poland 2001) and condensation (Teif 2005).

Additional lattice models are required at higher levels of chromatin compaction. For example, nucleosomal DNA is further compacted with the help of HP1 proteins, each nucleosome representing a binding site recognized by HP1 dimers. While the exact stoichiometry of these interactions is not known, experiments demonstrate the existence of at least three classes of binding sites with different HP1 affinities in chromatin. These differences might reflect the methylation status of histone H3 lysine 9 and/or binding of a HP1 dimer to one or two nucleosomes (Müller et al 2009). Because of the three-dimensional nature of the chromatin fiber, the two nucleosomes connected by an HP1 dimer are close to each other in 3D, but are not the nearest neighbors along the 1D DNA lattice (figure 6). Thus binding/unbinding of HP1 might significantly alter the compaction of the chromatin fiber. Indeed, different concentrations of HP1 have been found in heterochromatin and euchromatin.
of the histone octamer to a different position or evict the nucleosome. Also, the reverse situation might occur: genomic regions preferentially occluded by transcription factors or Pol II exclude nucleosomes by kinetic rather than equilibrium competition scenarios. Experimentally, it is found that the promoters of transcriptionally active genes are usually depleted of nucleosomes, while the promoters of transcriptionally inactive genes are occluded by the nucleosomes and being remodelled in the process of gene activation (Schones et al. 2008). Many promoters possess both a nucleosome-exclusion signal (such as the rigid poly(dA·dT) repeats) and a TF- or RNAP-positioning signal such as a TATA box (Zhang et al. 2009). Thus, it is not clear whether the nucleosome is depleted upstream of TSS at the active promoters due to the inherently low affinity of the histone octamer to the underlying DNA sequence (Segal et al. 2006), or due to the displacement by transcription factors (Morozov et al. 2009) combined with the action of chromatin remodelers (Teif and Rippe 2009).

4.1. TF binding to the nucleosomal DNA

The lattice models shown in figures 2(A) and 3 allow treatment of two types of TF–nucleosome competition. In the simplest case the nucleosome behaves as a single entity (figure 2(A)). In the improved model depicted in figure 3, the nucleosome is allowed to partially disassemble into smaller particles and TFs are allowed to penetrate into the nucleosome. The latter process is energetically unfavorable, but possible. Figure 7 shows the results of calculations according to these two models. In the absence of nucleosomes, TF is strongly bound to its specific binding site; adding nonspecifically binding nucleosomes decreases the apparent TF binding constant by several orders of magnitude. Interestingly, when the ligand is allowed to penetrate inside the nucleosome, the binding curve is shifted to increase the apparent TF binding constant. Thus splitting the nucleosome into subnucleosomal particles facilitates TF access to the nucleosomal DNA. In addition, nucleosomal DNA can also partially unwrap from the histone octamer facilitating access of transcription factors. A corresponding lattice model that accounts for this process is described in detail elsewhere (Teif et al. 2010).

4.2. TF–TF cooperativity induced by nucleosomes

Another insightful concept for TF–nucleosome competition is the so-called ‘collaborative competition’ (Adams and Workman 1995, Polach and Widom 1996). Consider the nonspecific binding of the histone octamer to the DNA in the absence of nucleosome–nucleosome interactions competing with a sequence-specific binding of transcription factors. Let there be two TF binding sites near each other that are both covered by a nucleosome (figure 8(A)). Once the DNA segment is freed from the nucleosome by one TF, it is easier for the second TF to bind. Because multiple regulatory DNA sites often occur within a nucleosome-length distance, this cooperativity is to be expected in vivo, and has indeed been observed experimentally (Adams and Workman 1995, Miller and Widom 2003). It might also be explained by the recruitment of specific remodelers rather than by a simple mechanical competition (Hebbar and Archer 2007). Figure 8(A) shows an illustrative map of binding for a DNA region of 3100 bp containing two specific TF binding sites separated by 50 bp. Here, it is assumed that nucleosomes nonspecifically bind the DNA at a high concentration: \( K(NCP) \cdot [NCP] = 100 \). The nucleosome binding periodicity predicted for this system by figure 8(B) is due to the boundary conditions (nucleosome hanging out from this region is prohibited). The periodic nucleosome pattern is disturbed in the vicinity of TF binding sites. This plot shows that for a given set of parameters TFs always partially displace nucleosomes, creating a nucleosome-depleted region.

In order to quantify the nucleosome-induced collaborative effect in terms of the standard McGhee–von Hippel type cooperativity, the logarithmic binding curves \( v(\ln[TF]) \) and the Scatchard curves \( (\nu/\nu[TF]) \) as a function of \( \nu \) have been calculated in figures 8(C) and (D). Three situations have been considered. In the first case TF binds the
Figure 8. Collaborative competition between transcription factors and nucleosomes. (A) A scheme of the system: the histone octamer is considered as a single entity, \( m_{\text{nuc}} = 147 \) bp; transcription factors are characterized by the length \( m_{\text{TF}} = 10 \) bp; there are two sequence-specific TF binding sites near each other. \( c_0(\text{nuc}) = 10^{-7} \) M, \( K(\text{nuc}) = 10^9 \) M\(^{-1}\), \( K(\text{TF}, n \neq 1000, 1050) = 10^7 \) M\(^{-1}\), \( K(\text{TF}, n = 1000, 1050) = 10^{10} \) M\(^{-1}\), \( c_0(\text{TF}) = 10^{-9} \) M\(^{-1}\). (C), (D) TF binding sites are separated by 10 bp. TF binding in the presence of nucleosomes (1) is compared to the non-cooperative binding in the absence of nucleosomes (2) and the binding characterized by the McGhee-von Hippel cooperativity (\( w = 10 \)) in the absence of nucleosomes (3).

DNA in the absence of nucleosomes, which results in a non-cooperative curve. The absence of cooperativity is manifested by the linear Scatchard plot in figure 8(D). In the second case, nucleosomes are added to the system, which decreases the effective TF binding constant, makes the binding curve in figure 8(C) slightly steeper, but has almost no detectable curvature in the Scatchard curve in figure 8(D). For comparison, when TF binding is characterized by the standard McGhee–von Hippel cooperativity with \( w = 10 \) (lower-middle range of typical TF–TF interactions), the resulting Scatchard curve shows a significant curvature characteristic for the binding cooperativity. Thus for a chosen set of parameters, nucleosome-assisted collaborative competition leads only to a moderate cooperativity in comparison with usual protein–protein interactions.

TF–nucleosome competition covers a wide range of effects beyond the two examples above. They can all be described by the corresponding lattice models. Consistent with this, recent papers from several groups consider nucleosomes as competitors to TF binding (Morozov et al 2009, Teif and Rippe 2009, Raveh-Sadka et al 2009, Wasson and Hartemink 2009), while another possibility is to consider nucleosome occupancies as static corrections to TF affinities (Gordan et al 2009). However, the view of nucleosomes as dynamic regulators of gene expression in chromatin allows larger flexibility.

5. Conclusions and perspectives

We have described a general theoretical framework for applying DNA lattice models to gene regulation in chromatin, including several new models. Different aspects of the nucleosome particle, architectural proteins, chromatin remodeler activity and transcription factor binding can all be described using appropriate lattice models. It is hoped that the approach described here will facilitate the construction of lattice models for other specific problems of protein–DNA interaction in the context of chromatin. One exciting aspect that needs to be addressed by further developments is the presence of covalent epigenetic modifications. These post-translational modifications of the DNA (by methylation) and of the histones (by acetylation, methylation, phosphorylation etc) modulate the binding affinity of chromatin ligands. By serving as anchors for architectural proteins and transcription factors they define the gene expression programs of the different cell tissues in higher eukaryotes in a manner that is preserved through cell division (Probst et al 2009). To some extent, epigenetic modifications have already been taken into account in the models discussed above as a correction to the sequence- and context-dependent binding constants \( K(n, g) \). By assigning additional states to the elementary unit of the lattice model the modified/unmodified DNA/histone can be described in addition to the bound/unbound states. However,
including the propagation of epigenetic pattern would require more extensions to existing lattice models and new approaches. Several approaches of this type have been considered recently, albeit in the absence of explicit considerations of protein concentrations and binding affinities (Sneppen et al 2008, Dodd et al 2007, David-Rus et al 2009). Thus, future theoretical and experimental advancement will hopefully fill this gap to yield DNA lattice models that are appropriate for systematic biophysical description of gene regulation in chromatin.

Acknowledgments

We thank Ramona Ettig and Fabian Erdel for comments. This work was supported by a fellowship of the German CellNetworks Cluster of Excellence (EXC81) and by a BRFFI grant B06M-60 to VT and by DFG grant Ri 1283/8-1 to KR.

References

Adams C C and Workman J L 1995 Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative Mol. Cell Biol. 15 1405–21
Akhrem A A, Fridman A S and Lando D 1985 Effect of ligand interaction with the boundaries between different forms of DNA on intramolecular transitions Dokl. Akad. Nauk SSSR 284 212–6
Anderson J D, Thåström A and Widom J 2002 Spontaneous access of proteins to buried nucleosomal DNA target sites occurs via a mechanism that is distinct from nucleosome translocation Mol. Cell Biol. 22 7147–57
Anderson J D and Widom J 2000 Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites J. Mol. Biol. 296 970–87
Bash R C, Yodh J, Lyubchenko Y, Woodbury N and Lohr D 2001 Population analysis of subsaturated 172–12 nucleosomal arrays by atomic force microscopy detects nonrandom behavior that is favored by histone acetylation and short repeat length J. Biol. Chem. 276 48362–70
Beer M A and Tavazoie S 2004 Predicting gene expression from sequence Cell 117 185–98
Bintu L, Buchler N E, Garcia H G, Gerland U, Hwa T, Kondev J, Kuhlman T and Phillips R 2005a Transcriptional regulation by the numbers: applications Curr. Opin. Genetics Dev. 15 125–35
Bintu L, Buchler N E, Garcia H G, Gerland U, Hwa T, Kondev J and Phillips R 2005b Transcriptional regulation by the numbers: models Curr. Opin. Genetics Dev. 15 116–24
Boeger H, Griesenbeck J and Kornberg R D 2008 Nucleosome retention and the stochastic nature of promoter chromatin remodeling for transcription Cell 133 716–26
Buccheri A, Kapitza K and Thoma F 2006 Rapid accessibility of nucleosomal DNA in yeast on a second timescale EMBO J. 25 3123–32
Cairns B R 2009 The logic of chromatin architecture and remodeling at promoters Nature 461 193–8
Chen Y 1987 Binding of n-mers to one-dimensional lattices with longer than close-contact interactions Biophys. Chem. 27 59–65
Chen Y, Maxwell A and Westerhoff H V 1986 Co-operativity and enzymatic activity in polymer-activated enzymes. A one-dimensional piggy-back binding model and its application to the DNA-dependent ATPase of DNA gyrase J. Mol. Biol. 190 201–14
Chen Y D 1990 A general secular equation for cooperative binding of n-mer ligands to a one-dimensional lattice Biopolymers 30 1113–21
Chen Y D 2004 Multiple binding of ligands to a linear biopolymer Methods Enzymol. 379 145–52
Chevereau G, Palmeira L, Thernes C, Arnedo A and Vaillant C 2009 Thermodynamics of intragenic nucleosome ordering Phys. Rev. Lett. 103 188103
Choi J K and Howe L J 2009 Histone acetylation: truth of consequences? Biochim. Cell Biol. 87 139–50
Chou T 2007 Peeling and sliding in nucleosome repositioning Phys. Rev. Lett. 99 058105
Cotton R W and Hamalko B A 1981 Nucleosome dissociation at physiological ionic strengths Nucleic Acids Res. 9 445–57
Crothers D M 1968 Calculation of binding isothemrs for heterogeneous polymers Biopolymers 6 575–84
Cuddapah S, Jothi R, Schones D E, Roh T-Y, Cui K and Zhao K 2009 Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains Genome Res. 19 24–32
Cui Y and Bustamante C 2000 Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure Proc. Natl Acad. Sci. USA 97 127–32
David-Rus D, Mokhodaphyay S, Lebowitz J L and Sengupta A M 2009 Inheritance of epigenetic chromatin silencing J. Theor. Biol. 258 112–20
Di Cera E and Kong Y 1996 Theory of multivalent binding in one and two-dimensional lattices Biophys. Chem. 61 107–24
Di Cera E and Phillipson P E 1996 Map analysis of ligand binding to a linear lattice Biophys. Chem. 61 125–9
Dodd I B, Micheelsen M A, Sneppen K and Thon G 2007 Theoretical analysis of epigenetic cell memory by nucleosome modification Cell 129 813–22
Engelholm M, de Jager M, Flaus A, Brenk R, van Noort J and Owen-Hughes T 2009 Nucleosomes can invade DNA territories occupied by their neighbors Nat. Struct. Mol. Biol. 16 151–8
Epstein I R 1978 Cooperative and noncooperative binding of large ligands to a finite one-dimensional lattice. A model for ligand–oligonucleotide interactions Biophys. Chem. 8 327–39
Evans J W 1993 Random and cooperative sequential adsorption Rev. Mod. Phys. 65 1281–329
Flyvbjerg H, Keath S A and Dryden D T 2006 Strong physical constraints on sequence-specific target location by proteins on DNA molecules Nucleic Acids Res. 34 2550–7
Gabdank I, Barash D and Trifonov E N 2010 FineStr: a web server for single-base-resolution nucleosome positioning Bioinformatics 26 845–6
Ganssen A, Valeri A, Hauger B, Felekyan S, Kalinin S, Toth K, Langowski J and Seidel C A 2009 Nucleosome disassembly intermediates characterized by single-molecule FRET Proc. Natl Acad. Sci. USA 106 15308–13
Gilmour D 2009 Promoter proximal pausing on genes in metazoans Chromosoma 118 1–10
Gjoda R, Hartemink A J and Bulyk M L 2009 Distinguishing direct versus indirect transcription factor–DNA interactions Genome Res. 19 2090–100
Gottesfeld J M and Luger K 2001 15308–13
Vaillant C 2009 Thermodynamics of intragenic nucleosome placement Phys. Rev. Lett. 103 188103
Gurvits G V and Zasedatelev A S 1978 Precise relationships for calculating the binding of regulatory proteins and other lattice ligands in double-stranded polynucleotides Biofizika 23 932–46
Gurvits G V, Zasedatelev A S and Vol’kenshtein M V 1972 Theory of one-dimensional absorption. II. Adsorption of small molecules on a heteropolymer Mol. Biol. 6 385–93
Hartley P D and Madhani H D 2009 Mechanisms that specify promoter nucleosome location and identity Cell 137 445–58
Hartley P D and Madhani H D 2009 Mechanisms that specify promoter nucleosome location and identity Cell 137 445–58
Peckham H E, Thurman R E, Fu Y, Stamatoyannopoulos J A, Noble W S, Struhl K and Weng Z 2007 Nucleosome positioning signals in genomic DNA Genome Res. 17 1170–7

Pennings S, Meersseman G and Bradbury E M 1991 Mobility of positioned nucleosomes on SS rDNA J. Mol. Biol. 220 101–10

Petersch S J and Liss J T 2008 Rapid, transcription-independent loss of nucleosomes over a large chromatin domain at Hsp70 loci Cell 134 74–84

Poirier M G, Bussie L, Langowski J and Widom J 2008 Spontaneous access to DNA target sites in folded chromatin fibers J. Mol. Biol. 379 772–86

Poirier M G, Oh E, Tims H S and Widom J 2009 Dynamics and function of compact nucleosome arrays Nat. Struct. Mol. Biol. 16 938–44

Polach K J and Widom J 1996 A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites J. Mol. Biol. 258 800–12

Poland D 1979 Cooperative Equilibration in Physical Biochemistry (Oxford: Clarendon)

Poland D 2001 Ligand binding distributions in nucleic acids Biopolymers 58 477–90

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

Radman-Livaja M and Rando O J 2009 Nucleosome positioning: how is it established, and why does it matter? Dev. Biol. 339 258–66

Raveh-Sadka T, Levo M and Segal E 2009 Incorporating nucleosomes into thermodynamic models of transcription regulation Genome Res. 19 1480–96

Rippe K, Mazurkiewicz J and Kepper N 2008 DNA Interactions With Polymers and Surfactants ed R S Dias and B Lindman (London: Wiley) pp 135–72

Rippe K, Schrader A, Riede P, Strohner R, Lehmann E and Rippe K, Schrader A, Riede P, Strohner R, Lehmann E and

Längst G 2007 DNA sequence–and conformation-directed positioning of nucleosomes by chromatin-remodeling complexes Proc. Natl Acad. Sci. USA 104 15635–40

Rouzina I and Bloomfield V A 2007 Competitive electrostatic binding of charged ligands to polyelectrolytes: practical approach using the non-linear Poisson–Boltzmann equation Biophys. Chem. 64 139–55

Saroff H A 1991 Ligand-dependent aggregation and cooperativity: a critique Biochemistry 30 10085–90

Schellman J A 1974 Cooperative multisite binding to DNA Nrr. J. Chem. 12 219–38

Schiessel H, Widom J, Bruinsma R F and Gelbart W M 2001 Polymer reposition and nucleosome repositioning Phys. Rev. Lett. 86 6144–7

Schnitzler G R 2008 Control of nucleosome positions by DNA sequence Cell Biochem. Biophys. 51 67–80

Schones D E, Cui K, Cuddapah S, Roh T Y, Barski A, Wang Z, Wei G and Zhao K 2008 Dynamic regulation of nucleosome positioning in the human genome Cell 132 887–98

Schwab D J, Bruinsma R F, Rudnick J and Widom J 2008 Nucleosome switches Phys. Rev. Lett. 100 228105

Segal E, Fondufe-Mittendorf Y, Chen L, Thaström A, Field Y, Moore I K, Wang J P and Widom J 2006 A genomic code for nucleosome positioning Nature 442 772–8

Segal E, Raveh-Sadka T, Schroeder M, Unnerstall U and Groudine M 2008 Predicting expression patterns from regulatory sequence in Drosophila segmentation Nature 451 535–40

Segal E and Widom J 2009a From DNA sequence to transcriptional behaviour: a quantitative approach Nat. Rev. Genetics 10 443–56

Segal E and Widom J 2009b Poly(A(ΔA)dT) tracts: major determinants of nucleosome organization Curr. Opin. Struct. Biol. 19 65–71

Shlyakhtenko L S, Lushnikov A Y and Lyubchenko Y L 2009 Dynamics of nucleosomes revealed by time-lapse atomic force microscopy Biochemistry 48 7842–8

Sims H I, Baughman C B and Schnitzler G R 2008 Human SWI/SNF directs sequence-specific chromatin changes on promoter nucleosomes Nucleic Acids Res. 36 6118–31

Snepen K, Micheelsen M A and Dodd I B 2008 Ultrasensitive gene regulation by positive feedback loops in nucleosome modification Mol. Syst. Biol. 4 182

Solis F J, Bash R, Wang H, Yodh J, Lindsay S A and Lohr D 2007 Properties of nucleosomes in acetylated mouse mammary tumor virus versus SS arrays Biochemistry 46 5623–34

Solis F J, Bash R, Yodh J, Lindsay S M and Lohr D 2004 A statistical thermodynamic model applied to experimental AFM population and location data is able to quantify DNA-histone binding strength and internucleosomal interaction differences between acetylated and unacyetylated nucleosomal arrays Biophys. J. 87 3372–87

Stein A, Takasuka T E and Collins C K 2009 Are nucleosome positions in vivo primarily determined by histone–DNA sequence preferences? Nucleic Acids Res. 38 709–19

Svaren J, Klebanow E, Sealy L and Chalkley R 1994 Analysis of the competition between nucleosome formation and transcription factor binding J. Biol. Chem. 269 9335–44

Teif V B and Bohinc K 2010 Condenced DNA: condensing the concepts Prog. Biophys. Mol. Biol. (Published online 16 July 2010) doi:10.1016/j.pbiomolbio.2010.1007.1002

Teif V B 2005 Ligand-induced DNA condensation: choosing the model Biophys. J. 89 2574–87

Teif V B, Etting R and Rippe K 2010 A lattice model for transcription factor access to nucleosomal DNA Biophys. J. at press doi:10.1016/j.bpj.2010.08.019

Teif V B 2007 General transfer matrix formalism to calculate DNA–protein–drug binding in gene regulation: application to Octα operator of phage λ Nucleic Acids Res. 35 e80

Teif V B 2010 Predicting gene-regulation functions: lessons from temperate bacteriophages Biophys. J. 98 1247–56

Teif V B, Harries D, Lando Y and Ben-Shaul A 2008 Matrix formalism for site-specific binding of unstructured proteins to multicomponent lipid membranes J. Peptide Sci. 14 368–73

Teif V B and Rippe K 2009 Predicting nucleosome positions on the DNA: combining intrinsic sequence preferences and remodeler activities Nucleic Acids Res. 37 5641–55

Teif V B, Vorob’ev V I and Lando D Y 2002 A single DNA chemical modification by cisplatin may cause long-range changes in protein binding to DNA. Results of computer modeling Eur. J. Human Genetics 10 (suppl. 1) 104

Thaström A, Lowary P T, Widlund H R, Cao H, Kubista M and Widom J 1999 Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences J. Mol. Biol. 288 213–29

Thaström A, Lowary P T and Widom J 2004 Measurement of histone–DNA interaction free energy in nucleosomes Methods 33 33–44

Tolstorukov M Y, Choudhary V, Olson W K, Zhurkin V B and Park P J 2008 nuScore: a web-interface for nucleosome positioning predictions Bioinformatics 24 1456–8

Trifonov E N 1980 Sequence-dependent conformational anisotropy of chromatin DNA Nucleic Acids Res. 8 4041–53

Tsudikov O V, Holbrook J A, Shkel I A and Record M T Jr 2001 Analytic binding isotherms describing competitive interactions of a protein ligand with specific and nonspecific sites on the same DNA oligomer Biophys. J. 81 1960–9

Tsuchiya T and Szabo A 1982 Cooperative binding of n-mers with steric hindrance to finite and infinite one-dimensional lattices Biopolymers 21 979–84

Vaillant C, Palmeira L, Chevereau G, Audit B, d’Aubenton-Carafa Y, Thyme C and Arneodo A 2009 A novel strategy of transcription regulation by intragenic nucleosome ordering Genome Res. 20 59–67
van Holde K E 1989 *Chromatin* (New York: Springer)
Wasson T and Hartemink A J 2009 An ensemble model of competitive multi-factor binding of the genome *Genome Res.* **19** 2101–12
Whitehouse I, Rando O J, Delrow J and Tsukiyama T 2007 Chromatin remodelling at promoters suppresses antisense transcription *Nature* **450** 1031–5
Whitehouse I and Tsukiyama T 2006 Antagonistic forces that position nucleosomes *in vivo* *Nat. Struct. Mol. Biol.* **13** 633–40
Wolfe A R and Meehan T 1992 Use of binding site neighbor-effect parameters to evaluate the interactions between adjacent ligands on a linear lattice: effects on ligand–lattice association *J. Mol. Biol.* **223** 1063–87
Woodbury C P Jr 1981 Free-sliding ligands: an alternative model of DNA–protein interactions *Biopolymers* **20** 2225–41
Workman J L and Kingston R E 1992 Nucleosome core displacement *in vitro* via a metastable transcription factor-nucleosome complex *Science* **258** 1780–4
Zasedatelev A S, Gurskii G V and Vol’kenshtein M V 1971 Theory of one-dimensional adsorption. I. Adsorption of small molecules on a homopolymer *Mol. Biol.* **5** 194–8
Zhang Y, Moqtaderi Z, Rattner B P, Euskirchen G, Snyder M, Kadonaga J T, Liu X S and Struhl K 2009 Intrinsic histone–DNA interactions are not the major determinant of nucleosome positions *in vivo* *Nat. Struct. Mol. Biol.* **16** 847–52
Zhao X, Pendergrast P S and Hernandez N 2001 A positioned nucleosome on the human U6 promoter allows recruitment of SNAPc by the Oct-1 POU domain *Mol. Cell* **7** 539–49
Zlatanova J, Bishop T C, Victor J M, Jackson V and van Holde K 2009 The nucleosome family: dynamic and growing *Structure* **17** 160–71