3 An Equilibrium Two-Component Gas Model

While histones are the most prevalent binding agent along the genome (for yeast as well as other eukaryotes), other proteins, such as transcription factors (TF), are also present and compete for binding sites along the DNA. Transcription factor binding, for example, has been shown to be correlated with the nucleosome free region (NFR) in the promoter regions, not only due to direct competition for binding sites, but also partly due to their recruitment of remodellers [54, 55]. The distribution of DNA-binding proteins in the promoter region has been shown to play an important regulatory role in transcription initiation[28, 56, 57, 20, 58, 59]. Furthermore, it has been demonstrated that changing cell conditions triggers changes in the binding pattern near the promoter regions[60]. The interplay between transcription and chromatin structure in the promoter region has even been proposed as a means of designing synthetic gene circuits[61].

In this chapter we extend the one-dimensional model of competitive site-binding into a two-component lattice gas model to reflect the interaction of smaller, usually specifically binding, TFs. To motivate this work, experimental data of shifted nucleosome profiles was gathered from published data and compared to TF binding positions in the course of Ref. [62]. The competitive binding of TFs and nucleosomes is particularly relevant in the context of our SoNG model from the previous chapter since it is known that transient unwrapping from the end of the nucleosome can allow for TF binding to positions that would otherwise be occluded in the HaNG model[40, 63, 45].

It has also been shown that multiple TFs whose binding sites overlap with the same nucleosome bind cooperatively, as they work together to displace a single nucleosome [64]. Calculations presented below suggest that similar effective cooperativity can even exist between transcription factors whose binding sites are separated by several intervening nucleosomes. As we will see below, shifted nucleosome patterns under the influence of one TF can affect the availability of sites for others. We refer to this interplay as nucleosome-mediated cooperativity between TFs and provide a quantification (first presented in Ref. [65]) of this effect.

3.1 Equilibrium Calculations

To incorporate TFs into a theoretical model of lattice binding, it is necessary first to construct an interaction potential between the two species, just as the nucleosome-nucleosome interaction potential was used in the previous chapter. Previous studies of TF-nucleosome interaction have assumed hard-exclusion in DNA binding (analogous to our HaNG model)[54]. It has been demonstrated experimentally, however, that site exposure (and availability for TF-binding) of nucleosomal DNA due to transient unwrapping decreases incrementally from

the outer edge of the nucleosome inward[45]. Thus, for TFs binding competitively with an adjacent nucleosome, we propose a progressive unwrapping penalty that mirrors the underlying physics of the nucleosome-nucleosome SoNG interaction from the previous chapter. Transient unwrapping allows for exposure of certain sites along the DNA with the energetic penalty of broken chemical bonds. One major difference from the SoNG interaction, however, is that we assume TFs are sufficiently tightly bound as to prevent unwrapping; thus, the neighbor interaction involves summing over the set of unwrapping states from only a single side.[1] The effective interaction potential $u(\Delta x)$ between a TF of size m bp whose nearest side is separated from a nucleosome dyad by Δx bp is determined by the sum of the Boltzmann factors of possible binding states of the nucleosome:

$$u(\Delta x) = \begin{cases} -\ln\left[\frac{\sum_{i=0}^{\Delta x-1} e^{i\varepsilon}}{\sum_{j=0}^{w} e^{j\varepsilon}}\right] & \text{if } \Delta x \leq w \\ 0 & \text{if } \Delta x > w \end{cases} \tag{3.1}$$

where w, again, is the length of DNA on one side of the dyad that can bind to the histone surface. There are strong parallels with with Eq. 2.5 from Chapter 2, and Fig. 3.1 shows that the two potentials are similar in shape, but with u acting at half the range.

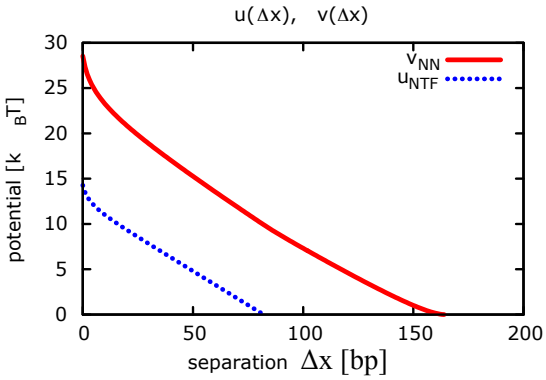

Figure 3.1: Neighbor interaction potential at a distance of Δx bp between two nucleosomes $v_{\mathrm{NN}}(\Delta x)$(red), and a Nucleosome with the nearest edge of a TF $u_{\mathrm{NTF}}(\Delta x)$(blue). To a good approximation, u_{NTF} is similar to v_{NN}, but with half the range.

It is quite straightforward to adapt the Gillespie algorithm of Chapter 2 to include particles of another type, and all of the same principles of reversible adsorption and detailed balance still apply –provided the additional energetic terms described above are included into the global energy change assumed in Eq. 2.17. Notably, however, the TFs are highly specific in their binding behavior, which is reflected in a position-dependent binding energy $\mu_{\mathrm{TF}}(x)$, as opposed to nucleosomes which bind non-specifically with energy μ_{N}. As in Chapter 2, $\mu_{\mathrm{TF}}(x)$ determines the binding rate $r_{+\mathrm{TF}} = e^{\mu_{\mathrm{TF}}(x) - u(\Delta x_l) - u(\Delta x_r)}$ in conjunction with any energetic interactions $u(\Delta x_l)$ and $u(\Delta x_r)$ with the left and right neighbors respectively. For simplicity, we take $\mu_{\mathrm{TF}}(x) \to -\infty$ at non-specific binding positions.

[1] There is, however, no reason to think that the binding energy of nucleosomes per bp ε would differ from the value determined in Chapter 2 via least-squares regression and from experimental results in Ref. [44].

To illustrate the effects of localized occlusion on the nearby nucleosome distribution pattern, we take a system with a fixed nucleosome at position $x = 0$, and measure the effect of adding a TF specific binding site at the downstream position $r = 360$, shown in Fig. 3.2 with $\mu_{\mathrm{TF}}(360) = \mu_{\mathrm{N}} = 9k_BT$. Here, the red data illustrate nucleosome density –in the absence of any TFs– similar to what was seen in the previous chapter, while the blue data show the resulting shift in this pattern due to competition with a single TF.

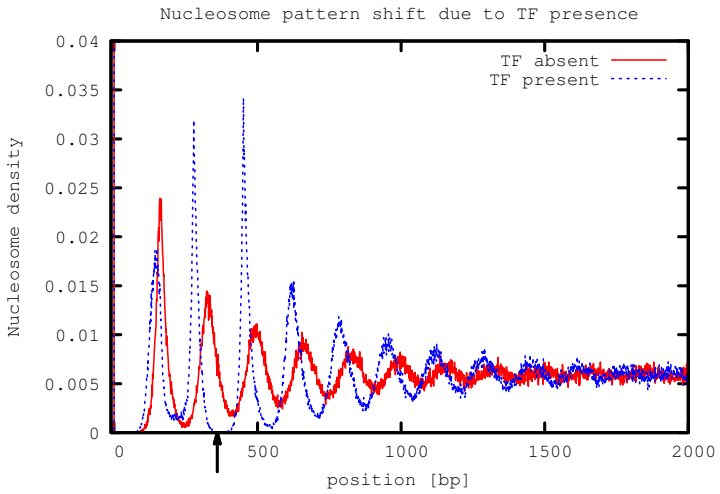

Figure 3.2: Time-averaged nucleosome positioning from stochastic kinetic simulations, assuming a fixed nucleosome at $x = 0$ as well as a TF of size $m = 10$ that binds specifically at $x = 360$ with energetic preference $\mu_{\mathrm{TF}}(360) = \mu_{\mathrm{N}} = 9k_BT$ (the TF binding position is indicated by the black arrow). The difference between the red trace (TFs removed) and the blue dashed trace (where TFs bind preferentially) illustrates the resulting shift in nucleosome patterns from the binding of the TF which, effectively, serves as an additional barrier.

As one can see from Fig. 3.2, the nucleosome density pattern is visibly perturbed over a significant length due to the TF. The new oscillating density pattern includes a new set of loci available for TF binding since the set of sites found in linker regions or peripheral nucleosome (susceptible to exposure from transient unwrapping) has shifted.

3.1.1 Nucleosomal Energetic Profile in the Promoter Region

TFs are, however, not the only factors that bind preferentially at certain target points. Until now we have assumed uniform binding affinity adjacent to a fixed +1 reference nucleosome that acts as a boundary for particles downstream. In Chapter 4 we will explicitly consider the 'landscape' of nucleosome binding energetics. For the moment, however, as a crude approximation to nucleosome binding patterns in the promoter region, we simply assign a position dependent binding affinity μ_{N} with x to capture the global energetics of binding throughout the entire NFR and +1 positioning region, rather than a fixed +1 barrier.

A very simplified profile is shown in Fig. 3.3(a); here we assume the NFR, an area of low binding affinity, has a size of $h = 200$ bp consistent with experimental estimates[66]. The

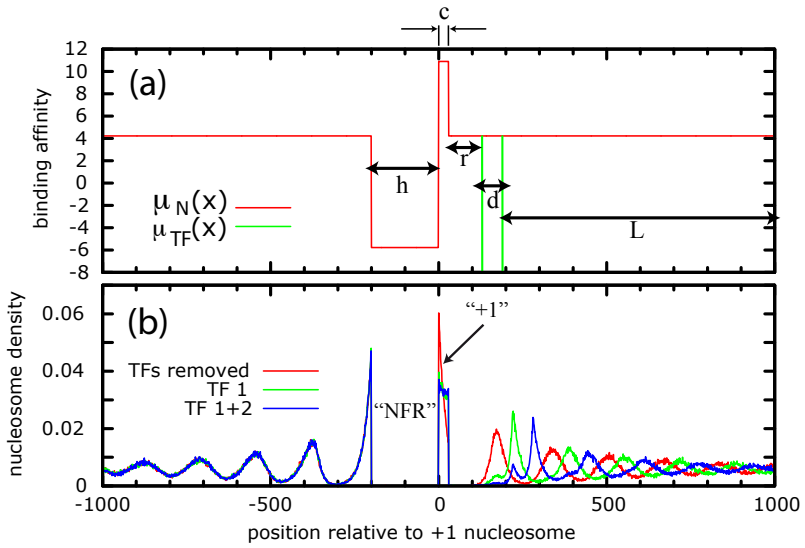

Figure 3.3: (a) sketch of simple profiles for the binding affinity of nucleosomes μ_{N} (red) and of TFs μ_{TF} (green). (b) Nucleosome density pattern in the absence of TFs (red), with a single TF at $x = c+r$(green), and with TFs binding at both $x = c+r$ and $x = c+r+d$ positions indicated in (a).

width of the high-affinity +1 positioning peak was taken to be $c = 30$ bp, consistent with typical convolution ranges of the +1 peak from yeast data in Ref. [36]. Figure 3.3(b) provides the density profile throughout the entire promoter region using this simple position-dependent nucleosome binding affinity. The figure is highly idealized (for example, the near complete elimination of histones from the NFR is exaggerated) but is intended to show the versatility of the numeric approach. Again, the influence on nucleosome density by a transcription factor at $x = 130$ can be seen in the green data. With a second TF at $x = 190$; the resulting blue data shows an even further displaced +2 peak. The model used here assumes no strong 'positioning' of nucleosomes away from the promoter region, and so displacement of nucleosomes by occlusion with TFs is incremental. Nevertheless, the curves in Fig. 3.3 suggest that two or more TFs can increase their effectiveness in displacing a nucleosome by working in concert.

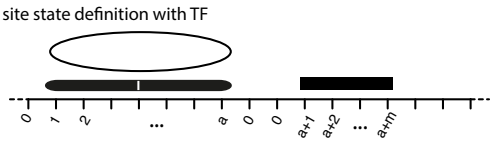

Figure 3.4: Schematic of site state definition analogous to Fig. 2.3 including states corresponding to transcription factor binding at right. Recall that state i is represented in the matrix of Eq. 3.2 by row $i+1$.

3.2 The Two-Factor Transfer Matrix

In addition to the stochastic simulations in the previous section, the displacement of nucleosomes due to competition with TFs can be observed by adapting the transfer matrix approach from Chapter 2. To do so, we define an additional set of m states corresponding to binding positions along a transcription factor m bp in length; this state convention is illustrated in Fig. 3.4. The corresponding transfer matrix T_{NTF} (the subscript 'NTF' denotes 'nucleosomes + transcription factors') for these states is $a+m+1$ in length. As in Chapter 2, we assume w bp are subject to unwrapping in either direction from the nucleosome dyad and a footprint length $a = 2w+1$. T_{NTF} contains the matrix $\tilde{T}_{\mathrm{NN}}$ (containing exclusively nucleosome-nucleosome interactions) given in Eq. 2.9 as a submatrix in the upper left corner:

$$\tilde{T} = \begin{bmatrix} & & & & & & 1 & 0 & \dots & & 0 \\ & & & & & & 0 & & & & \\ & & & \left[\tilde{T}_{\mathrm{NN}}\right] & & & \vdots & \vdots & & & \vdots \\ & & & & & & e^{\mu_N - u(1)} & & & & \\ & & & & & & e^{\mu_N - u(2)} & & & & \\ & & & & & & \vdots & & & & \\ & & & & & & e^{\mu_N - u(w)} & & & & \\ & & & & & & 1 & 0 & \dots & & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & e^{\mu_{TF}} & 0 & \dots & 0 \\ & & & & & & 0 & 0 & 1 & & 0 \\ \vdots & & & & & & & & \ddots & \ddots & \\ 0 & 0 & 0 & 0 & & 0 & & & & 0 & 1 \\ 1 & 1 & e^{-u(w)} & e^{-u(w-1)} & \dots & e^{-u(1)} & 0 & & \dots & & 0 \end{bmatrix} \tag{3.2}$$

where μ_{TF}, and μ_{N} denote the binding energy of TFs and nucleosomes respectively. As mentioned above, TFs target certain sequences for binding quite specifically, implying strong position dependence in the elements of Eq. 3.2 involving μ_{TF}. For this reason, we define a matrix $\tilde{T}_{\mathrm{TF}}$ for specific-binding TF loci where $\mu_{\mathrm{TF}} = \mu_{\mathrm{N}} = 4$, and take the limit $\mu_{\mathrm{TF}} \to \infty$ elsewhere. As in Section 3.1.1, we can incorporate position-dependent features of nucleosome formation by varying μ_N with position, and we define the matrices $\tilde{T}_{\mathrm{NFR}}$, $\tilde{T}_{+1}$, and $\tilde{T}_{\mathrm{NS}}$ to denote the transfer matrices for the nucleosome free region, the +1 binding region, and the non-specific region far away from the promoter respectively (see Fig. 3.3(a)). The partition sum for a system of size L including the full promoter region is then

$$Z_{\mathrm{NTF}} = \mathrm{Tr}\left\{\tilde{T}^{h}_{\mathrm{NFR}} \cdot \tilde{T}^{c}_{+1} \cdot \tilde{T}^{r}_{\mathrm{NS}} \cdot \tilde{T}_{\mathrm{TF}} \cdot \tilde{T}^{d}_{\mathrm{NS}} \cdot \tilde{T}_{\mathrm{TF}} \cdot \tilde{T}^{L}_{\mathrm{NS}}\right\}. \tag{3.3}$$

By using Eq. 3.3 in conjunction with projection matrices and taking the large L limit just as in the derivation of Eq. 2.15, state probabilities for various positions can be calculated; equivalent calculations are provided in full in Ref. [65], but are omitted here to avoid repetition. One benefit of this method is that it allows for efficient calculation of cooperativity in TF binding.

3.3 Transcription Factor Cooperativity

Fig. 3.3 illustrates the combined effect of multiple TFs to cooperatively displace or evict a nucleosome; naturally, this would suggest that the presence of one TF can facilitate the binding of other TFs in the vicinity. Cooperativity between TFs has been observed experimentally *in vitro* [67], and *in vivo* [68], and has been studied theoretically in Ref. [69].

Making use of the formalism from the latter of these studies, two arbitrary sites for TFs a and b are assigned statistical binding weights q_a and q_b respectively. In the absence of any binding at b, the probability that site a is occupied is $\tilde{p}_a = q_a/(1+q_a)$ (and likewise for protein b in the absence of binding by a, $\tilde{p}_b = q_b/(1+q_b)$.) More generally, however, when accounting for possible binding of other proteins and an unknown interaction, the probability of binding at site a, p_a, the probability of binding at site b, p_b, and the probability of binding at both positions simultaneously $p_{a,b}$ are given respectively as[69]

$$\begin{aligned} p_a &= \frac{q_a(1+\omega\cdot q_b)}{1+q_a+q_b+\omega\cdot q_a\cdot q_b} \\ p_b &= \frac{q_b(1+\omega\cdot q_a)}{1+q_a+q_b+\omega\cdot q_a\cdot q_b} \\ p_{a,b} &= \frac{\omega\cdot q_a\cdot q_b}{1+q_a+q_b+\omega\cdot q_a\cdot q_b}. \end{aligned} \tag{3.4}$$

Eq. 3.4 makes use of the dimensionless cooperativity ω which, on a log scale, captures the degree to which the presence of one TF promotes (or perhaps even inhibits) the binding of the other. ω can be solved for as

$$\omega = \frac{(p_a + p_b - p_{a,b} - 1)\, p_{a,b}}{p_a p_{a,b} + p_b p_{a,b} - p_a p_b - p_{a,b}^2}. \tag{3.5}$$

For $\omega > 1$ the two TFs bind cooperatively, while values of $\omega < 1$ imply antagonistic binding. The coupling constant $J_{1,2}$ between the two is then defined as $\omega_{1,2} = e^{-J_{1,2}}$

Obviously this cooperativity will depend on the proximity of loci a and b as well as their relative binding strength and the prevailing pattern of nucleosomes due to other factors. Notably, a and b need not necessarily be directly adjacent to one another. As noted in

Fig.'s 3.2 and 3.3 the presence of one TF on the DNA shifts the pattern of nucleosome density distribution, potentially uncovering or occluding the loci of secondary TFs even when separated by several nucleosomes. Long-range cooperativity between TFs *in vivo* has been pointed out by Vashee *et al*, though via a somewhat different mechanism[70] from what has been presented above.

For concreteness, we return to the example used in Fig. 3.2 except with a nucleosome fixed at $x = 0$ ($c = 0$), and TFs with $m = 5$ located at r and $r + d$ bp downstream. Fig. 3.5 charts the cooperativity $\ln(\omega)$ from Eq. 3.5 between TF a at r and TF b at $r + d$ using the equilibrium calculation based on Eq. 3.3.

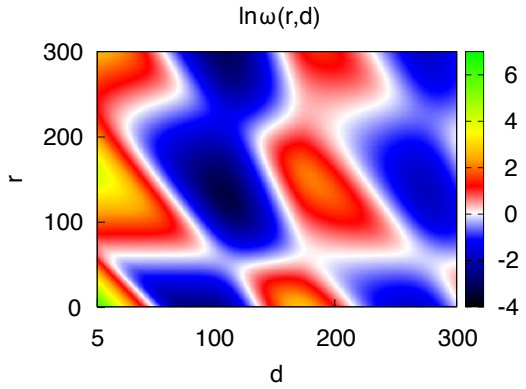

Figure 3.5: Heat map of cooperativity ω from Eq. 3.5 between two TFs separated at a distance d from each other with the first a distance r from the +1 positioned nucleosome. Reproduced from Ref. [65] with permission.

3.4 Experimental Observations

So far, this chapter has been purely theoretical. Our hope, however, is to compare these predictions with biological data. Before describing our efforts in doing so, a small amount of context is necessary. In the yeast *Saccharomyces cerevisiae*, fermentation is the major means of energy production (even in the presence of oxygen). When glucose becomes scarce, a switch is made to respiration using ethanol as a carbon source, resulting in massive reprogramming of gene expression[71, 72] that can hopefully be exploited to analyze changes in chromatin structure.

To attempt to connect the above observations to experimental data, Ref. [62] undertook a review of data on the *Saccharomyces cerevisiae* genome under such changing growth conditions. Available data on nucleosome occupancy[73], as well as expression level[74] of yeast in environments rich in glucose, galactose, and ethanol respectively were analyzed. Particular attention was paid to occupancy in the promoter regions of genes that underwent the greatest changes in expression. The report identified changes in nucleosome occupancy in the promoter region of genes with the strongest relative change in expression and the position of binding sites for TFs related to the above metabolic switch.

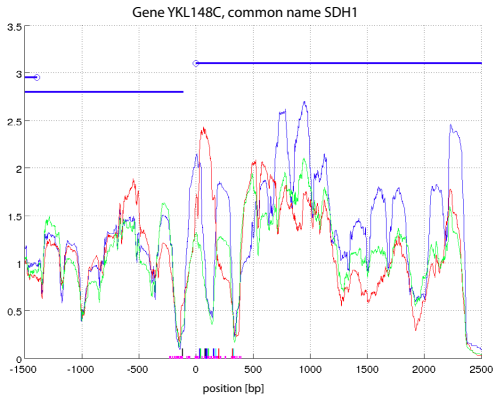

Figure 3.6: Shift in the +1 nucleosome positioning in the promoter region under the influence of TF binding. The blue line above denotes the transcript region with the circle representing the TSS. Occupancy in glucose is shown in blue, in ethanol in red and in galactose in green. TF binding sites are marked at the bottom with HAP1/2/3/4/5 in blue, GAL4 in green, MIG1/2/3 in orange, NHP10 in black, ADR1 in red, and MSN2/4 in cyan. Note the particularly dense concentration of HAP binding sites in the +1 region overlapping with the +1 nucleosome. Reproduced from Ref. [62] with permission.

Many examples of shifted nucleosome patterns in the vicinity of TF loci are listed in Ref.[62], and changes in nucleosome occupancy and TF binding activity showed some consistency with the above picture of competitive binding, however a global rule remained elusive. Several exceptions were observed. For example, the HAP4 TF is induced during the shift from fermentative to respiratory metabolism[75, 71]. Fig. 3.6 shows occupancy changes in the promoter region of the gene YKL148C, with particularly dense HAP binding sites indicated at the bottom of the figure. The shift in nucleosome occupancy at the +1 nucleosome position with ethanol (red) is clearly visible. However, unexpectedly, in the ethanol medium the +1 nucleosome is shifted *towards* the binding sites of the enriched HAP TFs, the opposite of what would be expected from the model of binding occlusion presented earlier.

As a possible explanation for this, homologs of the *S. cerevisiae* Hap(2-5) proteins Hap(A-E) in *Aspergillus nidulans*, have demonstrated bending of DNA similar to that in nucleosomes. They also have the potential for "mutual substitution in nucleosomes and [...] interactions with the histones H3 and H4 in mixed tetrasomes."[76] If such *nucleosome-like* proteins are formed by combining HAP proteins with H3 and H4 histones, then it is possible that this would account for observed occupancy shifts toward overlap with HAP TFs as illustrated in Fig. 3.6. This is, however, quite speculative. At any rate, our observations suggest that a general picture of TF-nucleosome interactions along DNA requires a more subtle description than the simple competitive interaction potential that was posited in Eq. 3.1.