An exact theory of histone-DNA adsorption and wrapping

TOM CHOU

Dept. of Biomathematics, UCLA - Los Angeles, CA 90095-1766

(received 6 January 2003; accepted in final form 27 March 2003)

PACS. 05.20.-y – Classical statistical mechanics.
PACS. 68.43.De – Statistical mechanics of adsorbates.
PACS. 87.14.Gg – DNA, RNA.

Abstract. – We find exact solutions to a one-dimensional (1D) interacting particle theory and apply the results to the adsorption and wrapping of polymers (such as DNA) around protein particles (such as histones). Each adsorbed protein is represented by a Tonks gas particle. The length of each particle is a degree of freedom that represents the degree of DNA wrapping around each histone. Thermodynamic quantities are computed as functions of wrapping energy, adsorbed histone density, and bulk histone concentration (or chemical potential); their experimental signatures are also discussed. Histone density is found to undergo a two-stage adsorption process as a function of chemical potential, while the mean coverage by high affinity proteins exhibits a maximum as a function of the chemical potential. However, fluctuations in the coverage are concurrently maximal. Histone-histone correlation functions are also computed and exhibit rich two length scale behavior.

Introduction. – Molecular adsorption onto 1D substrates occur in many biological settings such as protein adsorption on DNA/RNA [1, 2] and microtubules [3, 4]. Motivated by the microscopic details of such systems [4], we study histone-polymer binding and wrapping by solving a 1D theory of interacting particles with dynamically varying particle lengths. Histones are a class of hockey-puck shaped oligomer proteins about 10nm in diameter that wrap DNA. As shown in Fig. 1a, DNA can wrap around the edge of each histone protein complex at most 1.7 times, about 146 base pairs (bp) [5, 6]. DNA-histone complexes (nucleosomes) play vital roles in compacting DNA and regulating nucleic acid processing by mediating the accessibility by other regulatory proteins [1, 7, 8]. Acetylation levels of specific subgroups on the histone protein can affect their binding affinity, thereby providing a control mechanism for association and dissociation. When DNA needs to be processed (e.g. replicated), histones must move out of the way to allow other processing proteins access. Proposed mechanisms for histone dynamics include detachment/reattachment and sliding [7, 8]. Modeling the qualitative aspects of the histone binding isotherm on a one-dimensional substrate can potentially aid our understanding of DNA-histone and histone-histone interactions and their association/dissociation processes in biology.

We present a first attempt at describing histone isotherms using an exact statistical mechanical theory. The system depicted in Fig. 1 is mapped onto a generalized Tonks-Takahashi gas [9, 10] of one dimensional particles of dynamically varying length. Base pairs that come into contact with each histone protein defines a footprint which we associate with particle lengths. We consider the collective behavior of the particles mediated only by the mutual exclusion of their footprints along a tensionless DNA substrate. We also neglect the non-nearest-neighbor
nucleosome-nucleosome interactions arising from their relative, three-dimensional conformations [11]. Since DNA is very stiff, the statistical mechanics of DNA segments linking adjacent bound histones can be self-consistently neglected. This is a consequence of the fact that the histone size is smaller than the persistence length of DNA. When histones are spaced within a DNA persistence length of each other, the entropic contribution from the stiff linker DNA is negligible. On the other hand, binding of two histones far apart from each other does not appreciably reduce the phase space available to the linker polymer. Therefore, at large histone separations, the difference in entropy between a linker segment with and without histones bound to its “ends” is also negligible. Thus, neglect of linker DNA statistics is self-consistent with our assumption that nearest neighbor histone-histone exclusion interactions arise only from overlap of their footprints. It is also reasonable to expect that direct enthalpic binding energies will dominate typical entropic effects. Linker DNA entropy might be important only when histones are much larger than the persistence length and can appreciably affect the number of configurations accessible to the linking DNA segment.

**Tonks Gas Model.** – Consider a 1D collection of \( N \) particles labeled by the positions \( x_i \) of their left-most edges and confined within length \( L \). The first particle is fixed at \( x_0 = 0 \). The minimum size \( a \) of each particle defines the infinite hard core repulsive interaction between adjacent particles such that \( |x_i - x_{i+1}| \geq a \) and \( Na \leq L \). Each particle \( i \) also carries an additional internal degree of freedom \( \ell_i \) which corresponds to its length. The length \( \ell_i \) will describe the footprint of histone \( i \) on the DNA substrate (Fig. 1a). The incremental enthalpy of extending \( \ell_i \) by unit length at position \( x \) is \( \varepsilon(x) \). Although we have neglected the linker DNA entropy, the local entropy loss of the DNA segment that binds to the histone \( is \) included in \( \varepsilon \), along with the direct binding and bending energies. Under physiological conditions, typical values of \( \varepsilon \) are in the range \( \sim 0.2 k_B T / \text{bp} \) [6,8]. Since thermodynamic properties depend only on the interactions among particles, we compute the configurational partition function

\[
Z(N, L) = \int_0^{L_0} dy_{N-1} \cdots \int_0^{y_2} dy_1 \prod_{i=0}^{N-1} f(y_{i+1}, y_i),
\]

where \( y_i = x_i - ia \), \( y_N = L_0 = L - Na \), and

\[
f(y_{i+1}, y_i) = \int_0^{\ell_i} d\ell_i \exp \left[ - \int_0^{\ell_i} \varepsilon(x_i + \ell') d\ell' \right],
\]
where $\ell^*_i = \ell^*_i(y_i, y_{i+1})$ is the maximum possible wrapping around particle $i$ and $\varepsilon$ is a mesoscopic free energy for unit length extension. Although binding experiments have shown sequence-dependence histone-DNA affinities [6], we will assume only non-sequence specific interactions and uniform extension enthalpies ($\varepsilon = \varepsilon_0$) whence $f(y_{i+1}, y_i) = f(y_{i+1} - y_i)$. The integration limits in (1) reflect the particles’ impenetrability and renders $Z(N, L)$ convolutions of the functions $f(y_{i+1} - y_i)$. Upon using Laplace transforms in the variable $L_0$,

$$Z(N, L) = \int_{\gamma - i\infty}^{\gamma + i\infty} \tilde{f}^N(s)e^{sL_0} \frac{ds}{2\pi i},$$  \hspace{1cm} (3)$$

where $\tilde{f}(s)$ is the Laplace transform of $f$ and $\gamma \in \mathcal{R}$ is greater than the real parts of all singularities of the integrand. From (3), one can readily find exact thermodynamic quantities such as moments of the particle lengths $\ell_i$, $\langle \ell^n \rangle = (-1)^n N^{-1} \partial_{\varepsilon_0} \ln Z$, positions $\langle y^n_i \rangle$, and particle separations $\langle (y_{i+1} - y_i)^n \rangle$.

Although these quantities are readily computed using specific functions $\tilde{f}(s)$, the resulting sums typically involve numerical manipulation of extremely large numbers, especially for large $N$. Thus, it is also useful to derive from (3), using steepest descents, the leading order large $N$ asymptotic approximation $Z(N \to \infty, N/L = \rho) \sim e^{N(s^*(1/\rho - a) + \ln \tilde{f}(s^*))}$ [10], where $s^*$ is the saddle point defined by largest real root of

$$\left. \frac{1}{\tilde{f}(s^*)} \frac{\partial \tilde{f}(s)}{\partial s} \right|_{s = s^*} + \left( \frac{1}{\rho} - a \right) = 0. \hspace{1cm} (4)$$

Thermodynamic limits of, for example, the moments of the mean particle lengths become

$$\langle \ell^n \rangle \sim \left. \frac{(-1)^n \partial^n \tilde{f}(s)}{\tilde{f}(s^*)} \right|_{s = s^*}. \hspace{1cm} (5)$$

In the histone winding problem (Fig. 1) the unit of length will be a single nucleic acid base pair (bp). The total length $\ell_i$ of each particle corresponds to the arc-length of polymer that is wrapped around, and in direct contact with, a histone particle. The hard core cutoff $a$ depends on details of the three-dimensional arrangement of adjacent histones, and is roughly (or slightly smaller than) the diameter of a histone particle. Only for very specific phased orientations of canted histones along DNA can the histones be spaced less than about $a \approx 20$bp [1, 5, 6, 11]. The finite width of the histone limits “winding-in” lengths $\ell_i$ to either the distance to the start of an adjacent particle, $x_{i+1} - x_i - a = y_{i+1} - y_i$, or to $w \approx 146$bp, the maximum winding length corresponding to 1.7 loops around a histone particle. For example, particles zero and one in Fig. 1a have only one base pair of contact ($\ell_0 = \ell_1 = 0$), particles two and three are fully wound in ($\ell_2 = \ell_3 \approx 146$), while particle four is partially wound in ($0 < \ell_4 < w$). Upon imposing the physical limits on the particle lengths,

$$\ell^*_i = \begin{cases} x_{i+1} - x_i - a & x_{i+1} - x_i - a < w \\ w & x_{i+1} - x_i - a > w, \end{cases} \hspace{1cm} (6)$$

in the integration limit in (2), we find

$$\tilde{f}(s) = \frac{1 - e^{-(s+\varepsilon_0)w}}{s(s+\varepsilon_0)}. \hspace{1cm} (7)$$

With this form of $\tilde{f}(s)$, Eq. (3) takes the form,
\[ Z(N, L) = \sum_{k=0}^{N} a_k \int_{\gamma-i\infty}^{\gamma+i\infty} e^{-k\varepsilon_0 w} e^{s(L_0 - kw)} \frac{ds}{s^N (s + \varepsilon_0)^N} \frac{ds}{2\pi i}, \]

where \( a_k \equiv (-1)^k \binom{N}{k} \). For \( \gamma > \max\{0, -\varepsilon_0\} \) and \((L - Na - kw) > 0\), we close the contour in the left-hand \( s \)-plane. For \( L \) and \( k \) such that \((L - Na - kw) < 0\), convergence demands that we close the contour in the right \( s \)-half-plane. Since there are no poles to the right of \( \gamma \), terms with \((L - Na - kw) < 0\) correspond to configurations with more particles in \( L \) than is possible, and do not contribute to the partition function. Therefore, we need only sum (8) to \( k = k^* = \min\{\text{int}\ [(L - Na)/w], N\} \) to obtain the exact expression

\[ Z(N \geq 1, L) = \frac{(-1)^N N}{N! z_0^{2N-1}} \sum_{p,k=0}^{N-1,k^*} b_p (L - Na - kw)^p e_0^p \left[ e^{-\varepsilon_0 (L - Na)} - (-1)^p e^{-k\varepsilon_0 w} \right], \]

where \( b_p \equiv (2(N - 1) - p)!/(p!(N - 1 - p)!) \). Numerically, the asymptotic approximation (5) is typically accurate to within \( \sim 3\% \) of the exact result provided \( N \gtrsim 15 \).

Results and Discussion. – Fig. 2a plots the saddle solution \( s^* \) found from (4). In the fixed \( N \) ensemble, the probability distribution for particle \( i + 1 \) to be at position \( x_{i+1} \) given that particle \( i \) is at position \( x_i \) is readily computed in the asymptotic limit, \( g^{(2)}(x_{i+1} - x_i|N) \sim Z(N, L - x)/\int_0^L Z(N, L - x) dx = s^* e^{-s^* (x-a)} \), where \( L \gg \Lambda \gg 1/s^* \) defines an irrelevant length \( \Lambda \). This result implies that the adjacent particle is statistically confined to within \( x \lesssim 1/s^* \). At low number densities, adjacent histones are spaced far apart and \( s^* \sim \rho/(1 - \rho a) \). For attractive lengthening interactions (\( \varepsilon_0 \ll 0 \)), a sharp increase in \( s^* \) occurs near \( \rho \sim 1/(w + a) \) signaling a partial confinement of hard rod particles of roughly size \( w + a \). At extremely high densities, \( \rho \sim a^{-1} \), particles are compressed at the expense of unwinding, and \( s^* \) increases further as \( s^* \sim 2\rho/(1 - \rho a) - \varepsilon_0/2 \). Fig. 2b shows the mean winding-in length (normalized by \( w \)), found from (5). At low densities and strong attractive binding, the maximal winding in length \( w^{-1}(\ell) \sim 1 \) is approached, while at high densities, the winding-in length is restricted by nearest neighbors and \( \langle \ell \rangle \approx 1/\rho \).

Histone-DNA affinity and competitive binding experiments however, are performed by exposing DNA to fixed bulk histone concentrations [6, 14]. When the mean bound histone number is determined by the bulk histone chemical potential (thus not necessarily large), we employ the grand partition function found using the exact expression (3) or (9) in \( \Xi(\lambda, L) \equiv \sum_{N=1}^{N^*} \lambda^N Z(N, L) \). The fugacity \( \lambda \equiv e^{\mu - \varepsilon_0} \) takes into account the bulk histone chemical potential, and the binding energy \( \varepsilon_0 \) of a single base pair. \( N^* = \text{int}\{L/a\} \) is the maximum number of particles that can fit into length \( L \). In the \( L/a = \infty \), \( N^* = \infty \) limit, we find

\[ \Xi(\lambda, L \to \infty) = \oint \frac{\lambda e^{-s a} \tilde{f}(s) e^{s L}}{1 - \lambda e^{-s a} \tilde{f}(s)} \frac{ds}{2\pi i}, \]

which can be evaluated using (7) to yield \( \Xi(\lambda, L \to \infty) = \langle \rho \rangle e^{s^* L} \), where the mean density \( \langle \rho \rangle \equiv L^{-1} \partial_\mu \ln \Xi \) is explicitly

\[ \langle \rho \rangle = \frac{s_+ (s_+ + \varepsilon_0)}{2s_+ + \varepsilon_0 + \lambda e^{-s a} (a - (w + a) e^{-(s_+ + \varepsilon_0) w})}, \]

\( s_+ \) being the largest real root of \( 1 - \lambda \tilde{f}(s_+) e^{-s_+ a} = 0 \). The mean fraction of DNA base pairs covered by contacts with histones is then found from
The mean density and coverage are plotted in Figs. 2c,d. For $\mu \to -\infty$, $\langle \rho \rangle \sim s_+ \sim e^{-\varepsilon_0 + \mu (1 - e^{-\varepsilon_0 w})}/\varepsilon_0 \to 0$ since this limit corresponds to infinitely dilute bulk histone concentration. Densities of adsorbed particles increase with bulk histone chemical potential. For higher affinity histones, these increases occur earlier (smaller values of $\mu$). When the substrate is nearly covered with fully wound-in histones ($\langle \rho \rangle a \approx a/(w + a) \approx 0.12)$, a plateau arises where increases in $\mu$ are unable to inject more histones. Only at very large $\mu$ can the histone density increase at the expense of unwinding until the density approaches maximal packing at $\langle \rho(\mu \to \infty) \rangle \sim 1/a - 2/(\mu a) + O(\mu^{-2}\ln \mu)$. This two-stage behavior is seen only for high affinity ($\varepsilon_0 \ll 0$) histones. The transition from monotonic to the two-stage density behavior may be observable by tuning the nonspecific histone binding energies $\varepsilon_0$ by e.g. varying ionic
DNA accessibility mediated by histone positioning is thought to be a major determinant of nucleic acid processing by other regulatory proteins [1, 6–8]. The total coverage \( \langle \theta \rangle \), defined as the mean fraction of base pairs in contact with histones, is shown in Fig. 2d. For \( \mu \to -\infty \), there are few particles present to adsorb onto the DNA substrate and \( \langle \theta \rangle \sim e^{-\varepsilon_0 + \mu} \to 0 \). An increase in mean coverage results from increasing \( \mu \) and the number of bound histones. The theoretical maximum coverage \( w/(w+a) \approx 0.88 \) (thin dashed line) corresponds to close packing of fully wound-in particles and is approached only for high affinity histones. The minimum uncovered fraction \( a/(w+a) \) results from the linker DNA of minimum length \( a \sim 20 \) joining two adjacent nucleosomes. If \( \mu \) is further increased, the particles squeeze on each other until they unwrap to the point where only a single base pair remains in contact for each histone. In this limit, the mean coverage \( \langle \theta(\mu \to \infty) \rangle \sim 1/a + (1 - 2/\mu)/\mu + O(\mu^{-2} \ln \mu) \approx 0.05 \) while the histones are spaced at their maximal densities \( \langle \rho \rangle \approx 1/a \). The variance in coverage, \( \text{var}(\theta) = L^{-1} \partial(L^{-1} \partial \Xi/\partial \varepsilon_0 + \Xi)/\partial \varepsilon_0 \), gives a standard deviation in coverage proportional to the mean coverage, \( \sqrt{\text{var}(\theta)} = \sqrt{2(\theta)} \). Despite the high coverage at intermediate chemical potentials, fluctuations in this regime are also maximal, suggesting a dynamically controlled DNA accessibility mediated by histone-histone interactions.

Fig. 3 – The correlation \( g^{(2)}(x|\mu) \) exhibits properties of both length scales \( a \) and \( w \) depending on \( \varepsilon_0 \) and \( \mu \). For high affinity and low densities (e.g. \( \varepsilon_0 w = -30; \mu = -5 \)), the density distribution is similar to that of a hard rod Tonks gas of length \( a + w \). Upon increasing the density (\( \varepsilon_0 w = -30; \mu = -2 \)), features associated with both length scales arise as partial unwinding occurs, exposing the hard core repulsion of size \( a \). Conditions under which particles weakly wind in (\( \varepsilon_0 w = -10 \)) exhibit behavior attributed to hard rods of length \( a \). However, even at lower affinities (\( \varepsilon_0 w = -8 \)), behavior approximating that of rods of length \( w + a \) can be recovered if densities are made sufficiently low (\( \mu = -15 \)).

In the grand ensemble, the relevant correlation function is \( \langle \rho \rangle g^{(2)}(x|\mu) \equiv \Xi(\lambda, x)\Xi(\lambda, L-x)/\Xi(\lambda, L) \) which describes the probability distribution that given a particle at the origin, any other histone exists at \( x \). When \( x \gg 1/\langle \rho \rangle \) and \( N^* < \infty \), the exact expression (9) must be used to compute \( \Xi(\lambda, x) \). Fig. 3 shows \( g^{(2)}(x|\mu) \) computed for various values of \( \varepsilon_0, \mu \). Finite particle size anticorrelations give rise to oscillations at two length scales (see Fig. 3 caption) depending on affinity \( \varepsilon_0 \) and density \( \langle \rho(\mu) \rangle \).

The proposed model considers only the adjacent histone exclusion interactions mediated by nonoverlapping footprints and neglects sequence specific and nucleosome-nucleosome interactions arising in compact, 3D chromatin structure. Nonetheless, our theory can be solved
exactly with uniform wrapping energies $\varepsilon_0$ to give reasonable results for winding-in lengths $\langle \ell \rangle$ and histone-histone correlations $g^{(2)}(x|N)$ and $g^{(2)}(x|\mu)$. The model predicts a two-state adsorption process and a maximum in the DNA coverage fraction $\langle \theta \rangle$ as a function of bulk histone concentration and binding affinity. However, fluctuations in the coverage are also concurrently maximal during the peak in mean coverage, indicating that thermal effects can nonetheless provide dynamic access to DNA that is on average highly covered. These predictions can be tested experimentally by varying bulk histone coverage and ionic strength (although this would also affect DNA bendability and $\varepsilon_0$) [6, 12], provided large scale 3D structures do not arise and non-nearest-neighbor interactions remain irrelevant. Recent AFM images of equilibrium histone loading on intermediate-length DNA [15] have justified the nearest-neighbor interactions and have also demonstrated correlation functions qualitatively similar to those shown in Fig. 3. An observed sharp adsorption function (cf. Fig. 2c) is also consistent with the existence of density-induced partial unwrapping.

In other well-known 1D models that describe DNA melting [16], the corresponding particles are melted DNA bubbles that do not interact (can coalesce) and that do not have the maximum length $w$ unique to the histone wrapping problem. Our model can be readily generalized to include the effects of linker DNA twist, externally applied tension [17], and relative histone orientation [11] on the cut-off $a$. Specific sequences, and their effects on local bendability, twistability, and histone affinity has also been found to be important in vitro [6] and can be treated within a similar framework, although completely generalizing our model to include specific sequences [6,14] (spatial dependence of $\varepsilon(x)$), would require computational approaches.

***

This work was supported by the NSF through grant DMS-0206733.

REFERENCES