THE UNIVERSITY OF CHICAGO

PHASE TRANSITIONS OF DNA-PROTEIN CONDENSATES DURING THE IMMUNE RESPONSE

A THESIS SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY
CLAYTON W. SEITZ

CHICAGO, ILLINOIS
SPRING 20XX

TABLE OF CONTENTS

ABSTRACT

Eukaryotic transcription is episodic, consisting of a series of transcriptional bursts, Bursty transcriptional dynamics are well-exemplified by the transient expression of pro-inflammatory guanylate binding proteins (GBPs) - a group interferon-inducible GTPases that restrict the replication of intracellular pathogens [XXX]. Classical models of gene regulation explain transcriptional bursts by invoking stochastic binding and unbinding of transcription factors, RNA polymerase and mediator proteins at enhancer or promoter sequences. However, more recent studies have pointed towards a more cooperative picture of transcriptional control where phase-separated aggregates of DNA, RNA, and proteins form higher-order structures to control gene expression. For example, both chromatin immunoprecipitation and super resolution imaging have captured the phase separation of super-enhancer-binding proteins MED1 and BRD4 in transcriptional condensates at the *Essrb* genomic locus [XXX]. Furthermore, fluorescence microscopy techniques have colocalized MED1 and BRD4 with the GBP gene cluster alongside a reduction in the degree of disorder of 3D chromatin structure in murine macrophages after infection with *Mycobacterium tuberculosis.* Taken together, these results suggest that phase separation may play a role in the reorganization of chromatin structure during trancriptional control of innate immune response genes [XXX]. Here, we hypothesize that phase separation reduces the entropy of chromatin structure in order to induce bursty gene expression. Using single molecule localization microscopy (SMLM) to obtain super-resolution images of the H2B protein, we intend to demonstrate simultaneous (i) loss of disorder in chromatin structure (ii) formation of transcriptional condensates containing MED1 and BRD4 and (iii) non-Poissonian gene expression. The following sections discuss recent the biological evidence in more detail and summarize the single molecule microscopy techniques and biophysical models we employ to study the interactions between transcriptional condensates and the chromatin scaffold.

CHAPTER 1

SINGLE MOLECULE LOCALIZATION MICROSCOPY

1.1 A generative model for single molecule localization microscopy

Most detectors used for imaging have many elements (pixels) so that we can record an image projected onto the detector by a system of lenses. In fluorescence imaging, this is usually a relay consisting of an objective lens and a tube lens to focus the image onto the camera. Due to diffraction, any point emitter, such as a single fluorescent molecule, will be registered as a diffraction limited spot. The profile of that spot is often described as a Gaussian point spread function

$$\mathrm{G}(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}} + B_0 \tag{1.1}$$

Modern cameras used in light microscopy, such as scientific complementary metal oxide semiconductor (sCMOS) cameras, are powered by the photoelectric effect. Electrons within each pixel, called photoelectrons, absorb enough energy from incoming photons to be promoted to the conduction band to give electrical current which can be detected. Integration of photoelectrons during the exposure time results in a monochrome image captured by a camera. The image of a single point particle, such as a fluorescent molecule, can be thought of as two-dimensional histogram of photon arrivals and a discretized form of the classical intensity profile $\mathrm{G}(x,y)$. The value at a pixel approaches an integral of this density over the pixel:

$$\mu_k = I_0\lambda_k = I_0 \int_{\text{pixel}} G(x,y)dxdy \tag{1.2}$$

where $I_0 = \eta N_0 \Delta t$. The parameter η is the quantum efficiency and Δt is the exposure time. N_0 represents the number of photons emitted per unit time, which may be itself a Poisson random variable; however, this is inconsequential since it is a fixed value for a single image of a fluorescent molecule. The value of the integral λ_k at a pixel k defines the fraction of I_0 photons observed at that pixel during the camera exposure. Since the 2D Gaussian in (1.1) has cylindrical symmetry, it is separable, and we can write

$$\lambda_k = \frac{1}{2\pi\sigma^2}\left(\int_{x_k}^{x_{k+1}} e^{-\frac{(x-x_0)^2}{2\sigma^2}}dx\right)\left(\int_{y_k}^{y_{k+1}} e^{-\frac{(y-y_0)^2}{2\sigma^2}}dy\right)$$

We can then express the Gaussian integrals over a pixel by making use of the following property of the error function

$$\frac{1}{\sqrt{2\pi}\sigma}\int_a^b e^{\frac{-(x-\mu)^2}{2\sigma^2}} = \frac{1}{2}\left(\operatorname{erf}\left(\frac{b-\mu}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{a-\mu}{\sqrt{2}\sigma}\right)\right)$$

Now, suppose the particle is known to be located at (x_0, y_0) and some square pixel k is centered on coordinates (x, y) and has a width a. Then λ_k at this pixel is

$$\lambda_k(x,y) = \lambda_x(x)\lambda_y(y) \tag{1.3}$$

where

$$\begin{aligned}\lambda_x(x) &= \frac{1}{2}\left(\operatorname{erf}\left(\frac{x+a/2-x_0}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{x-a/2-x_0}{\sqrt{2}\sigma}\right)\right)\\ \lambda_y(y) &= \frac{1}{2}\left(\operatorname{erf}\left(\frac{y+a/2-y_0}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{y-y/2-y_0}{\sqrt{2}\sigma}\right)\right)\end{aligned}$$

We are now prepared to write the shot-noise limited signal, which is a vector with units of phototelectrons

$$\vec{S} = [\text{Poisson}(\lambda_1), \text{Poisson}(\lambda_2), ..., \text{Poisson}(\lambda_N)] \tag{1.4}$$

However this noise model is incomplete, because detectors often suffer from dark noise, which may refer to readout noise or dark current, and contributes to a nonzero signal even in the absence of incident light. Dark current is due to statistical fluctuations in the photoelectron count due to thermal fluctuations. Readout noise is introduced by the amplifier circuit during the coversion of photoelectron charge to a voltage. Here, we use the Hamamatsu ORCA v3 CMOS camera, which is air cooled to -10C and has very low dark current - around 0.06 electrons/pixel/second - and can therefore be safely ignored for exposure times on the order of milliseconds. Readout noise has been often neglected in localization algorithms because its presence in EMCCD cameras is small enough that it can be ignored within the tolerances of the localization precision. In the case of sCMOS cameras, however, the readout noise of each pixel is significantly higher and, in addition, every pixel has its own noise and gain characteristic sometimes with dramatic pixel-to-pixel variations.

On the other hand, readout noise is not negligible, and must represented as a vector-valued random variable. It is important to note that we cannot measure the contribution by readout noise before amplification and therefore it must be expressed in units of ADU. This is in contrast to $\vec{S}$, which can be expressed in units of photoelectrons, because these statistical fluctuations can be predicted to be Poisson by quantum mechanics. The number of photoelectrons S_k is multiplied by a gain factor g_k which has units of [ADU/e^-], which generally must be measured for each pixel. Here, we will always assume that readout noise per pixel ξ_k is Gaussian with some pixel-specific offset o_k and variance σ_k^2.

We will now define a vector which represents the image we measure, in units of ADU:

$$\vec{H} = \vec{S} + \vec{\xi} \tag{1.5}$$

What we are after is the joint distribution $P(\vec{H})$. A fundamental result in probability theory is that the distribution of H_k is the convolution of the distributions of S_k and ξ_k,

$$\begin{aligned} P(H_k|\theta) &= P(S_k) \circledast P(\xi_k) && (1.6) \\ &= A \sum_{q=0}^{\infty} \frac{1}{q!} e^{-\mu_k} \mu_k^q \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(H_k - g_k q - o_k)}{2\sigma_k^2}} && (1.7) \end{aligned}$$

where $P(\xi_k) = \mathcal{N}(o_k, \sigma_k^2)$ and $P(S_k) = \text{Poisson}(g_k \mu_k)$. In practice, this expression is difficult to work with, so we look for an approximation. Notice that

$$\xi_k - o_k + \sigma_k^2 \sim \mathcal{N}(\sigma_k^2, \sigma_k^2) \approx \text{Poisson}(\sigma_k^2)$$

Since $H_k = S_k + \xi_k$, we transform $H_k' = H_k - o_k + \sigma_k^2$, which is distributed according to

$$H_k' \sim \text{Poisson}(\mu_k')$$

where $\mu_k' = g_k \mu_k + \sigma_k^2$. This result can be seen from the fact the the convolution of two Poisson distributions is also Poisson. Under this approximation, the model negative log-likelihood is

$$\begin{aligned} \ell(\vec{H}) &= -\log \prod_k \frac{e^{-(\mu_k')} \left(\mu_k'\right)^{n_k}}{n_k!} \\ &= \sum_k \log n_k! + \mu_k' - n_k \log \left(\mu_k'\right) \end{aligned}$$

1.1.1 The Cramer-Rao lower bound

A general task in Bayesian inference is to determine θ from the data under the model $\mathcal{M}_\theta$. We may then ask - does the log-likelihood ℓ vary as we vary the parameters? If the likelihood is flat, all parameter sets are equally likely and the data does not appear to carry much information about the parameters. Moreover, if ℓ has a number of bumps or inflection points, then we expect that maybe some parameter sets are more likely that others. The "bumpiness" of the likelihood surface is called the Fisher information - a fundmantal concept in information geometry. The Fisher information matrix $I(\theta)$ can be directly related to the curvature of the KL-Divergence over the parameter space

$$\begin{aligned}
\nabla^2_{\theta'} D_{KL}[\ell(H|\theta) \parallel \ell(H|\theta')] &= -\nabla_{\theta'} \int \ell(H|\theta) \nabla_{\theta'} \log \ell(H|\theta')\, dH \\
&= -\int \ell(H|\theta) \nabla^2_{\theta'} \log \ell(H|\theta')\, dH \\
&= -\mathbb{E}_\theta[\nabla^2_{\theta'} \log \ell(H|\theta')] \\
&= I(\theta)
\end{aligned}$$

We often call the Hessian matrix the *score*. The Fisher information is the result of averaging the score over the parameter space. To be clear, the score is a function of the *data*, not the parameters. It is a measure of sensitivity of the likelihood to changes in the parameters. Of course, the Fisher information matrix also depends on the parameterization chosen. Looking at (1.4), the likelihood is a hierarchical function that maps a vector space Θ to a vector space Λ to a scalar value. Formally, we define $T : \Theta \to \Lambda$ and $W : \Lambda \to \mathbb{R}$. The parameter vector $(x_0, y_0, \sigma, N_0) \in \Theta$, the Poisson rate vector $\vec{\lambda} \in \Lambda$ and $\ell \in \mathbb{R}$. To get the Hessian, we need the chain-rule for Hessian matrices.

$$\hat{H}_{(\ell,\theta)} = \hat{J}^T_{(\lambda,\theta)}\hat{H}_{(\ell,\lambda)}\hat{J}_{(\lambda,\theta)} + (J_{(\ell,\lambda)} \otimes I_n)\hat{H}_{(\lambda,\theta)}$$

In the second term of the equation, we have a Kronecker product between the Jacobian matrix of the likelihood with respect to the parameters of the hierarchical model ($J_{(L,\lambda)}$) and the $n \times n$ identity matrix (I_n), denoted as $J_{(L,\lambda)} \otimes I_n$. This Kronecker product gives a diagonal matrix where the elements along the diagonal are the elements of the vector $J_{(L,\lambda)}$. This provides a concise way of summarizing the operation:

$$((J_{(L,\lambda)} \otimes I_n)\hat{H}_{(\lambda,\theta)})_{ij} = \sum_k \hat{J}^{ij}_{(L,\lambda)}\hat{H}^{ijk}_{(\lambda,\theta)}$$

Note that we sum over the last index to get a 3 x 3 matrix, which is not directly obvious in the compact notation.

1.1.2 Vanilla Variational Bayes

We now describe a Bayesian inference method for the parameters θ from experiment and simulated data. In previous sections, we have given a a formal probabilistic description of the entire SMLM measurement process along with fundamental lower bound on parameter uncertainty. With this in hand, we can now describe practical methods for inferring parameters from data and assess their ability to reach these bounds. In our framework, at each time point a measurement is made $\vec{H}_t$, which can be used to infer a matrix of parameters $\boldsymbol{\theta}_t = (\boldsymbol{x_0}, \boldsymbol{y_0}, \boldsymbol{\sigma}, \boldsymbol{B_0}, \boldsymbol{N_0})_t$. Occasionally, we will refer to parameter matrices θ_t at each time point as latent variables which give rise to the noisy low resolution images which constitute the measured data $\vec{H}$. A complete description of the generative model in the Bayesian framework includes the prior distribution $P(\theta)$ which describes the spatial distribution and temporal dynamics of fluorophores, and the likelihood $P(\vec{H}|\theta)$ - a distribution of images generated by the microscope for a given configuration of fluorophores. According to Bayes rule, the posterior on θ can be constructed as

$$P(\theta|\vec{H}) = \frac{P(\vec{H}|\theta)P(\theta)}{\int P(\vec{H}|\theta)P(\theta)d\theta} \tag{1.8}$$

Note this likelihood was formalized in terms of the point spread function and the measurement noise of the camera in (1.7). There are several methods for finding the posterior, for example Markov Chain Monte Carlo (MCMC) or variational methods. Variational inference attempts to approximate the true posterior distribution $p(\theta|x)$ with a variational distribution $q(\theta)$, which is often a model distribution which can be easily sampled from e.g., a multivariate Gaussian. To fit the variational distribution, we can try to minimize the KL-divergence between the variational distribution and the true posterior:

$$\begin{aligned}
D_{\mathrm{KL}}(q(\theta)||p(\theta|x)) &= \underset{\theta\sim q(\theta)}{\mathbb{E}}\left(\log\frac{q(\theta)}{p(\theta|x)}\right) \\
&= \underset{\theta\sim q(\theta)}{\mathbb{E}}\left(\log\frac{q(\theta)p(x)}{p(\theta,x)}\right) \\
&= \log p(x) + \underset{\theta\sim q(\theta)}{\mathbb{E}}\left(\log\frac{q(\theta)}{p(x|\theta)p(\theta)}\right) \\
&= \log p(x) + \underset{\theta\sim q(\theta)}{\mathbb{E}}\left(\log q(\theta) - \log p(x|\theta) - \log p(\theta)\right) \\
&= \log p(x) + H(q) - \underset{\theta\sim q(\theta)}{\mathbb{E}}\left(\log p(x|\theta) + \log p(\theta)\right)
\end{aligned}$$

Clearly the KL-divergence is minimized by minimizing this expectation. It is common to define the evidence lower bound (ELBO).

$$\ell(x,\theta) = -\mathop{\mathbb{E}}_{\theta\sim q(\theta)}\left(\log q(\theta) - \log p(x|\theta) - \log p(\theta)\right) \tag{1.9}$$

which gets its name from the fact that

$$\log p(x) = D_{\mathrm{KL}}(q(\theta)||p(\theta|x)) + \ell(x,\theta) \geq \ell(x,\theta)$$

Of great practical interest are the gradients of the ELBO

$$\nabla_{\Phi}\ell(x,\theta) = -\mathop{\mathbb{E}}_{\theta\sim q(\theta)}\left(\log q(\theta) - \log p(x|\theta) - \log p(\theta)\right)$$

The quality of our parameter estimation will undoubtedly depend on our choice for the variational distribution. In some contexts, the posterior is approximated as a multivariate Gaussian, which is valid when the posterior has this relatively simple structure and higher order moments vanish. However, in general, the posterior is not Gaussian and more flexible variational distributions must be used. Deep neural networks are advantageous in this case, as they are known to be able to represent very complex high-dimensional distributions.

1.1.3 Autoencoding Variational Bayes

The use of deep neural networks to represent the posterior was a central ingredient in Welling and Teh's original paper introducing the variational autoencoder for generative modeling. Indeed, variational Bayesian inference is very closely related to generative modeling. A generative model according to Bayes rule is found by rearranging (1.8)

$$P(\mathbf{x}) = \frac{P_{\Phi}(\mathbf{x}|\theta)P_{\Omega}(\theta)}{Q_{\Psi}(\theta|\mathbf{x})}$$

As it should, this relates the generative distribution of the data $\boldsymbol{x}$ to the likelihood (which can be thought of as a measurement model), posterior which is unknown, and the prior on the parameters. Therefore, by modeling each of these distributions simultaneously in the form of a variational autoencoder, one obtains a generative model and the ability to perform parameter inference. Essentially all we need to do is recast (1.9) with the distributions being represented as neural networks Ψ and Φ.

$$\ell = -\underset{\theta \sim q_\Psi(\theta)}{\mathbb{E}} \left(\log q_\Psi(\theta) - \log p_\Phi(x|\theta) - \log p_\Omega(\theta)\right) \tag{1.10}$$

It is common to call Ψ and Φ the encoder and decoder, respectively.

1.2 Photoswitching dynamics of JF646

The central assumption underlying a Markov process, is the memoryless property

$$P(X_t|X_{t-1}, X_{t-2}, ..., X_{t-N}) = P(X_t|X_{t-1})$$

A single Markov chain is the set of states $\boldsymbol{X} = \{X_1, X_2, ..., X_N\}$. Such a set can be generated provided that $P(X_t|X_{t-1})$ is known. To capture $P(X_t|X_{t-1})$ for all possible pairs X_t and X_{t-1}, we define a square transition matrix $T \in \mathcal{R}^{N\times N}$ where $N = |\Omega|$. As such, the elements of T represent the probability of a transition from a state ω_j to ω_i in a unit time

$$T_{ij} = \Pr\left(X_t = \omega_i, \mid X_{t-1} = \omega_j\right)$$

Under these definitions, the row T_i represents the present time, and is a conditional probability distribution $P(\omega|X_{t-1} = \omega_j)$ which requires that

$$\sum_j T_{ij} = \sum_j P(X_t = \omega_j | X_{t-1} = \omega_i) = 1$$

The matrix T is not necessarily symmetric $T_{ij} \neq T_{ji}$. One should note that the columns T_j *do not* define a probability distribution $P(X_t = \omega_i | X_{t-1} = \omega_j)$ and therefore do not necessarily sum to unity. The probability $P(X_t = \omega_i | X_{t-1} = \omega_j)$ has no meaning in this context, since we have defined the rows to represent a probability of the future given the present. We simply sample $X_t \sim P(X_t = \omega_j | X_{t-1} = \omega_i)$, assign $i = j$, and repeat. It directly follows from the fundamental rules of probability, the first order dynamics for a **particular** state ω_i: $P(\omega_i, t)$ is given by

$$P(\omega_i, t + dt) = P(\omega_i, t) + \mathcal{J}_i dt \tag{1.11}$$

The net probability current $\mathcal{J}_i$ must be

$$\mathcal{J}_i = \sum_i T_{ij} P(\omega_j, t) - \sum_j T_{ij} P(\omega_i, t)$$

The first is a sum on a column and the second a sum on a row. This can be simplified further by noticing that the normalization condition implies

$$\begin{aligned} T_{ij} &= 1 - \sum_j T_{ij}(1 - \delta_{ij}) \\ &= 1 - \sum_j T_{ij} + \sum_j T_{ij}\delta_{ij} \end{aligned}$$

$$
\begin{aligned}
\mathcal{J}_i &= \sum_i T_{ij} P(\omega_j, t) - \sum_j T_{ij} P(\omega_i, t) \\
&= \sum_i \left(1 - \sum_j T_{ij} + \sum_j T_{ij} \delta_{ij} \right) P(\omega_j, t) - \sum_j T_{ij} P(\omega_i, t) \\
&= |\Omega| - |\Omega| + \sum_i \sum_j T_{ij} P(\omega_j, t) \delta_{ij} - \sum_j T_{ij} P(\omega_i, t) \\
&= \sum_j T_{ji} P(\omega_j, t) - T_{ij} P(\omega_i, t)
\end{aligned}
$$

Notice that the Kronecker delta effectively just swaps the index. Taking the limit of (1.1), we arrive at the **master equation**

$$\frac{\partial P(\omega_i)}{\partial t} = \sum_j T_{ji} P(\omega_j, t) - T_{ij} P(\omega_i, t)$$

It is common to then define an operator $\boldsymbol{W}$ s.t. $W_{ij} = T_{ij}$ and $W_{ii} = -\sum_j T_{ij}$

$$\frac{dP(\omega_i)}{dt} = \sum_j W_{ij} P(\omega_j) \rightarrow \frac{dP(\boldsymbol{\omega})}{dt} = \mathcal{J}(\boldsymbol{\omega}) = \mathbf{W} P(\boldsymbol{\omega})$$

This operator form has a solution in terms of a matrix exponential

$$P(\boldsymbol{\omega}, t) = \exp(\mathcal{J}(\boldsymbol{\omega}))$$

This matrix exponential is intractable for large $|\Omega|$. However, in the Finite State Projection algorithm, it is possible to truncate the state space $\Omega \rightarrow \tilde{\Omega}$ and obtain good estimates $\tilde{P}(\boldsymbol{\omega}, t)$ with some certificate of accuracy.

1.2.1 Lifetime of the ON and OFF states

Alongside the solution to the master equation $P_{\text{ON}}(t)$ and $P_{\text{OFF}}(t)$, we are often interested in lifetime of the ON and OFF states and how these lifetimes change according to experimental parameters. In fact, the distribution over the ON and OFF state lifetimes gives a convenient frequentist mechanism for estimating the rate constants of a photoswitching model. To see this, consider the two-state model where $\alpha_{12}(t) = k_{12}(t)dt$ and $\alpha_{21}(t) = k_{21}(t)dt$ are the propensity of transitions from on to off and off to on, respectively. Suppose the system begins in the ON state,

$$\begin{aligned} P(\text{ON} \geq t \mid \text{ON}) &= \lim_{n\to\infty} (1 - k_{12}dt)^n \\ &= \lim_{n\to\infty} \left(1 - \frac{k_{12}t}{n}\right)^n \\ &= e^{-k_{12}t} \end{aligned}$$

This is the cumulative density function for the lifetime of the ON state. Notice that we are given that the system is in the ON state at $t = 0$, and all we want to know is the lifetime of that state, so we need not consider the transition rates back to the ON state: k_{j1}. The probability density of the lifetime of the ON state is:

$$f_{\text{ON}}(t) = -\frac{dP(\text{ON} \geq t)}{dt} = k_{12}e^{-k_{12}t}$$

Therefore one possible way of inferring the rate constant k_{12} is by fitting the observed lifetime distribution with an exponential distribution with k_{12} as a free parameter. However, for the OFF state, it has been previously shown that a single rate constant may not fit the data accurately and the introduction of multiple rate constants representing multiple dark states are necessary. Common photoswitching models include three OFF states: the triplet state, dark state, and long-lived dark state and rate constants between these two states.

$$P(\mathrm{OFF} \geq t \mid \mathrm{OFF}) = P_{\mathrm{T}}(t) + P_{\mathrm{D}}(t) + P_{\mathrm{LLD}}(t)$$

where we have the system of equations for transitions within the OFF state

$$\begin{aligned}
\frac{dP_{\mathrm{T}}(t)}{dt} &= -k_{21}P_{\mathrm{T}}(t) - k_{23}P_{\mathrm{T}}(t) \\
\frac{dP_{\mathrm{D}}(t)}{dt} &= P_{\mathrm{T}}(t)k_{23} - k_{31}P_{\mathrm{D}}(t) - k_{34}P_{\mathrm{D}}(t) \\
\frac{dP_{\mathrm{LLD}}(t)}{dt} &= k_{34}P_{\mathrm{D}}(t) - k_{41}P_{\mathrm{LLD}}(t)
\end{aligned}$$

This system can be solved analytically by Laplace transformation. Similar to the ON state,

$$f_{\mathrm{OFF}}(t) = -\frac{dP(\mathrm{OFF} \geq t)}{dt} = \sum_i a_i \lambda_i e^{-\lambda_i t}$$

CHAPTER 2

PHASE TRANSITIONS OF DNA-PROTEIN CONDENSATES

Metropolis-Hastings algorithms are commonly used to sample from the configurational space of polymers, in order to generate a set of conformations that can be used to study polymer properties such as their equilibrium structure, thermodynamics, and response to external fields. This method is particularly useful when studying the equilibrium properties of polymers, where the system is in thermal equilibrium and can be described by a Boltzmann distribution. Examples of Metropolis-Hastings algorithms that have been used for polymer simulations include the pivot algorithm, the reptation algorithm, and the configurational-bias Monte Carlo algorithm.

On the other hand, Langevin dynamics simulations can be used to study the non-equilibrium dynamics of polymers, where the system is driven out of thermal equilibrium by external forces, such as shear or electric fields. Langevin dynamics simulations can also incorporate more realistic models of the environment, such as solvent effects or hydrodynamic interactions, which can affect the behavior of the polymer. This method is particularly useful when studying the dynamics of polymers in complex environments, where the system is far from equilibrium and the dynamics cannot be described by a simple Boltzmann distribution. Examples of Langevin dynamics simulations that have been used for polymer simulations include the Brownian dynamics method and the dissipative particle dynamics method.

In summary, both Metropolis-Hastings algorithms and Langevin dynamics simulations can be used to study polymer dynamics, but they are used for different purposes. Metropolis-Hastings algorithms are typically used to study the equilibrium properties of polymers, while Langevin dynamics simulations are used to study the non-equilibrium dynamics of polymers.

The main difference between solving the Langevin equation and Newton's equations of motion is the addition of the stochastic force term in the Langevin equation. This term models the effect of thermal fluctuations and can lead to a more realistic description of the dynamics of the system. In contrast, Newton's equations of motion assume a deterministic behavior of the system without any stochasticity.

Another difference is that Langevin dynamics simulations are particularly useful for simulating systems in contact with a heat bath, which is not possible with Newton's equations of motion alone. In the case of Langevin dynamics, the friction coefficient gamma models the effect of the heat bath on the system and allows for the simulation of both thermal and non-thermal forces that affect the dynamics of the system.

Appendices

.0.1 ODE model for the ensemble average

We can define the following system of ODEs for a autorepressive gene circuit

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \frac{\beta_m}{1+(p/k)^n} - \gamma_m m, \tag{1}$$

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \beta_r m - \gamma_r r, \tag{2}$$

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \beta_p r - \gamma_p p, \tag{3}$$

We can greatly reduce the number of parameters by nondimensionalizing the system. To that end, we define a characteristic time scale t_0 and characteristic "length" scales for each variable: m_0, r_0, p_0. The choice of these characteristic scale will become apparent after writing the nondimensionalized ODEs out for a general case. The derivatives transform as

$$\frac{dm}{dt} \to \frac{m_0}{t_0}\frac{dm'}{d\tau}, \quad \frac{dr}{dt} \to \frac{r_0}{t_0}\frac{dr'}{d\tau}, \quad \frac{dp}{dt} \to \frac{p_0}{t_0}\frac{dp'}{d\tau}$$

Our general system of nondimensionalized ODEs reads:

$$\begin{aligned}
\frac{\mathrm{d}m'}{\mathrm{d}\tau} &= \frac{t_0\beta_m}{m_0(1+(p_0p'/k)^n)} - \gamma_m t_0 m' \\
\frac{\mathrm{d}r'}{\mathrm{d}\tau} &= \frac{\beta_r t_0 m_0 m'}{r_0} - \gamma_r t_0 r' \\
\frac{\mathrm{d}p'}{\mathrm{d}\tau} &= \frac{\beta_p t_0 r_0 r'}{p_0} - \gamma_p t_0 p'
\end{aligned}$$

Define the characteristic scales as: $t_0 = 1/\gamma_m$, $m_0 = \gamma_m^2 k/\beta_r\beta_p$, $r_0 = \gamma_m k/\beta_p$, and $p_0 = k$. Making these substitutions, we have

$$
\begin{aligned}
\frac{\mathrm{d}m'}{\mathrm{d}\tau} &= \frac{\beta}{(1+(p')^n)} - m' \\
\frac{\mathrm{d}r'}{\mathrm{d}\tau} &= m' - \frac{\gamma_r}{\gamma_m} r' \\
\frac{\mathrm{d}p'}{\mathrm{d}\tau} &= r' - \gamma p'
\end{aligned}
$$

.0.2 Telegraph model of gene expression

We will begin by writing the the system of ODEs describing the dynamics of the first moment

$$
\begin{aligned}
\frac{\mathrm{d}m}{\mathrm{d}t} &= \beta_m(1-m) - \gamma_m m && (4) \\
\frac{\mathrm{d}r}{\mathrm{d}t} &= \beta_r m - \gamma_r r && (5) \\
& && (6)
\end{aligned}
$$

This system has 4 parameters, making it difficult to directly visualize interesting relationships between parameterization and dynamics. Therefore, we will nondimensionalize this system

$$
\begin{aligned}
\frac{\mathrm{d}m'}{\mathrm{d}t} &= \frac{t_0}{m_0}\beta_m - \frac{t_0}{m_0}\beta_m m_0 m' - \frac{t_0}{m_0}\gamma_m m_0 m' && (7) \\
\frac{\mathrm{d}r'}{\mathrm{d}t} &= \frac{t_0}{m_0}\beta_r m_0 m' - \frac{t_0}{m_0}\gamma_r r_0 r' && (8) \\
& && (9)
\end{aligned}
$$

Let $t_0 = 1/\gamma_m$, $m_0 = \beta_m/\gamma_m$, $r_0 = \beta_m\beta_r/\gamma_m^2$, $\gamma = \gamma_r/\gamma_m$, $\beta = \beta_m/\gamma_m$, and we have

$$\frac{\mathrm{d}m'}{\mathrm{d}t} = 1 - \beta m' - m' \tag{10}$$

$$\frac{\mathrm{d}r'}{\mathrm{d}t} = m' - \gamma r' \tag{11}$$

$$\tag{12}$$