NORTHWESTERN UNIVERSITY

Characterizing the Fundamental Chromatin Structure and Function in a Realistic Nuclear Environment

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Biomedical Engineering

By

Ranya Virk

EVANSTON, ILLINOIS

August 2022

ABSTRACT

Characterizing the Fundamental Chromatin Structure and Function in a Realistic Nuclear Environment

Ranya Virk

Chromatin is the biological material that packages our genetic information. In humans, 2 meters of linear DNA is compacted into an approximately 6 μm nucleus. Our DNA is transcribed into RNA, which is then translated into proteins. Cellular phenotype, the composite of all cellular functions, is defined by the entire transcriptome and proteome of the cell. Thus, the organization of chromatin, which controls accessibility of DNA to transcriptional machinery, helps to dictate cellular function. Recent developments in technology, ranging from high-throughput sequencing, to super-resolution optical imaging and electron microscopies, to increases in computing power, have greatly expanded our understanding of chromatin structure and function.

It is not just the genome, the sequence of A's, T's, C's, and G's, that influences cellular phenotype. DNA wraps around histone octamers to form nucleosomes, which are strung together by linker DNA to form the 'beads-on-a-string' chromatin fiber. The epigenome - including chemical modifications such as DNA methylation and histone tail

acetylation/methylation - modulates cell function by controlling local chromatin structure and accessibility. Epigenetic marks can recruit chromatin readers, which are capable of nucleosome remodeling and activating transcription or compacting nucleosomes and repressing transcription. In contrast to the previously accepted textbook view of a highly ordered 30 nm fiber, chromatin also exhibits a highly disordered structure at the level of the primary fiber. Higher-order chromatin domains have also been observed at the level of 100's of kilobase pairs to megabase pairs. The existence of these dynamic structures has been shown to modulate transcriptional efficiency by influencing four-dimensional enhancer-promoter contacts and the spreading of epigenetic marks.

Despite these recent advances in our understanding, there remain several key open questions in the chromatin field that this thesis aims to address. Chapter 2 identifies and investigates the fundamental units of chromatin folding. Chapter 2 begins by employing a unique combination of high-resolution electron microscopy imaging and polymer physics-based analysis techniques to characterize the morphological and functional properties of higher-order chromatin packing domains. Next, in Chapter 2 we uncover fundamental organizational principles of the genome using nanoimaging and chromosome conformation capture experimental approaches to validate and better understand a statistical model of chromatin structure. Altogether, we demonstrate that the topology of chromatin can be represented by branching, tree-like network structures and that statistical rearrangements in connectivity and mass density distribution occur upon heat stress. Chapter 3 then transitions into the functional implications of the fundamental statistical chromatin organization identified in Chapter 2. The Chromatin Packing Macromolecular Crowding

(CPMC) model, which combines a kinetic model of transcription with the statistical packing descriptors of chromatin packing domains, is able to faithfully predict the phenotypic plasticity of cancer cells. The initial model is then extended to predict cancer cell death in response to cytotoxic chemotherapy treatment. Altogether, the initial chromatin packing state of cells is shown to directly influence cellular adaptability to external stressors. Chapter 4 then focuses on developing and employing a molecular theoretical approach to characterize the effects of the physicochemical intranuclear environment, including bulk ions, pH, and density, on the structure and charge of DNA-like and chromatin-like systems. We identify bridging of multivalent cations as an important mechanism for both neutralizing the strongly negative charges of DNA-phosphates and increasing compaction of DNA-phosphate loops. We extend our approach to investigate the effects of the physicochemical environment on individual nucleosomes and 8-mer nucleosome arrays, and determine the importance of chromatin density on the effects of the monovalent electrolyte environment on chromatin structure. Finally, Chapter 5 provides a summary of this thesis work and an outlook for future directions. Overall, this thesis combines physics-based modeling, nanoimaging, and sequencing-based molecular approaches to better understand fundamental mechanisms underlying chromatin structure and function in a realistic nuclear environment.

Acknowledgements

Before introducing my thesis work, I would first like to thank the many amazing people in my life for their mentorship, patience, and wisdom throughout the past five years. A commonly used cliché is that it takes a village to raise a child, but I believe this also applies to getting someone successfully through a PhD.

I would like to begin by thanking both of my advisors, Vadim Backman and Igal Szleifer. Vadim recruited me to Northwestern with the promise of engaging in cutting-edge, high impact research and truly excited me about his lab's concept of macrogenomic engineering, and the research I have been able to do in his group has lived up to these expectations. I would like to thank Vadim for his unwavering support during the trials and tribulations of manuscript and grant writing, for the long and many brainstorming sessions we have had together, for his patience when I didn't understand a complex concept immediately or when we came to a disagreement about how something should be approached, and for fully supporting and believing in me and my scientific abilities. I am truly inspired by Vadim's dedication to his research and his desire for the science to have an actual impact in people's lives. Vadim has also provided me with so many opportunities for my professional and scientific development. He has always encouraged me to attend academic conferences and has pushed to increase my visibility in the scientific community, even from my first year in graduate school. One such opportunity was an invitation to a recent faculty retreat that he organizes annually in Telluride, Colorado.

There, I was the only trainee and was able to get facetime with leaders in the chromatin field. I would like to thank Igal for his dedicated support and outstanding mentorship throughout my PhD. Igal has helped me hone my intellectual curiosity and improve my research process. He always asks me to explain my rationale for why I am doing something before I dive right into things. Igal has also taught me to not take everything at face value and to always question your science instead of believing that everything you have published is unwavering fact. He has also patiently sat through many of my imperfect practice presentations and provided invaluable feedback, the most recent of which he was recovering from COVID. His constant reminder that we need to combine "experiment and theory" in order to understand complex biological systems like chromatin will truly stick with me throughout my scientific career. As a team, both of my advisors have helped me to accomplish and, most importantly, to grow as a scientific researcher throughout my PhD training.

Vadim and Igal have both also created a warm and welcoming environment for science in both of their groups. I would like to thank all past and present members of the Backman and Szleifer group. Specifically, I would like to thank Jane Frederick and Vasundhara Agrawal who have been with me in Vadim's group since my first year. They have been amazing scientific colleagues - we have spent countless hours discussing scientific findings and ideas for new projects - and very supportive friends. Also, from the Backman group, I would like to thank Yue Li for many brainstorming and cafe writing sessions related to manuscripts, which turned into her mentoring me about life. I would like to thank Wenli Wu and Greta Wodarcyk for their countless hours of mentorship during the first two years of my PhD. I would like to thank Maria Proenca, the Backman group lab manager,

and Benjamin Keane and Jayms Peterson for their administrative and emotional support throughout my PhD. Everything from scheduling meetings to figuring out conference logistics to organizing DEI initiatives within the lab was made so much easier and more pleasant through their diligent efforts. I would also like to specifically thank (in no particular order) Emily Pujadas, Adam Eshein, Nicholas Anthony, Scott Gladstein, David VanDerway, Aya Eid, Cody Dunton, Wingshun Li, Surbhi Jain, and Luay Almassalha.

From the Szleifer group, I would like to sincerely thank Dr. Rikkert Nap for his instruction and collaboration throughout my PhD. He has taught me an incredible amount, ranging from statistical mechanics to electrostatics to going over the Molecular Theory approach and code in great detail. He has patiently answered all of my questions, re-explained concepts I have had trouble understanding, and helped me think through results and future directions for my projects. Rikkert and the rest of the Szleifer group have also sat through many of my practice presentations, which have substantially improved throughout my PhD due in no small part to their detailed and thoughtful feedback. I would like to thank Marcelo for many useful discussions ranging from implementing biased Molecular Dynamics simulations to combining experiments and theory to define fundamental chromatin structures with his Self-Returning Excluded Volume model. I would also like to thank Kai Huang, Anne Shim, Estefania Gonzalez Solveyra, and Sung Hyun Park for their mentorship and countless inspiring scientific discussions.

I would like to thank all present and past collaborators. Allen Taflove, may you rest in peace, your belief in and support of our research was truly inspiring. I miss our weekly updates for our joint supercomputer grant. Talking to you about science and life was always a pleasure. Thank you to Juan de Pablo and, especially, to his graduate student,

Aria Coroar, for our fruitful Molecular Theory collaboration involving many hours of thoughtful discussions. And thank you to my thesis committee members Professors Ting Wu, Zhe Ji, and Alexandra Zidovska for their feedback and encouragement.

Thank you to my parents for patiently listening to me when things were difficult and for always pushing for me to advocate for myself. Thank you to my cousins, Melina and Alexandra, for their help throughout the graduate school application process. Both of you truly inspired me to pursue a PhD and your help through it all has been significant. Thank you to all of my family members who have believed in me and supported me in ways large and small.

I would also like to thank the many friends I have made in graduate school. Thank you to my first roommates (and first friends) Abigail Smith and Maggie Swerdloff. None of us knew that the other two random people we met on a Northwestern housing site would lead to real, lasting friendships. You have both been staunch supporters of me not working *all* the time and have organized countless adventures for me, from Caribbean festivals in Toronto to my first backpacking trip. Thank you to my dear friends Sebastian Rodriguez and Linda Pei for basically getting me through Covid and my PhD simultaneously through Facetimes, daily walks, memes that I don't understand, and dinners on your apartment floor (before you got furniture). Thank you to my cohort-mate and friend Reem Rashid for the countless hours of social gatherings and support, despite my chronic lateness. Even though we are technically not related, despite everyone thinking we were sisters during the Northwestern interview weekend, I still consider you basically a family member. I would also like to thank countless other graduate school friends who have empowered and supported me through the years including Mena Whelan for opening up her kitchen and

home to me during times both stressful and fun, Nilou Sarvian for her exuberant and positive personality, and Bram Diamond, Kunal Shah, Carolyn Ramirez, Yena Lee, and Zach Ladwig for being genuinely kind and supportive human beings (and thank you again to Abby Smith again for introducing me to most of these amazing people). I will treasure the memories and remarkable support I have received during my PhD.

List of abbreviations

1CPN - 1-Cylinder-Per-Nucleosome

3C - Chromosome Conformation Capture

3SPN - 3-Sites-Per-Nucleotide

4C - Circularized Chromosome Conformation Capture

5C - Chromosome Conformation Capture Carbon Copy

ACF - Auto-Correlation Function

BD - Brownian Dynamics

bp - base pair (of DNA)

CDA - Chromatin-Dependent Adaptability (model)

CPMC - Chromatin Packing Macromolecular Crowding (transcription model)

ChromEM - Chromatin Electron Microscopy

ChromEMT/TEM - Chromatin Transmission Electron Microscopy

ChromSTEM - Chromatin Scanning Transmission Electron Microscopy

ChIP-seq - Chromatin Immunoprecipitation Sequencing

CPT - Chromatin Protective Therapy

CTCF - CCCTC-binding factor

CVC - Chromatin Volume Concentration

D - Chromatin packing scaling

Hi-C - High-throughput Chromosome Conformation Capture

LJ - Lennard-Jones (potential)

MC - Monte Carlo

MD - Molecular Dynamics

MT - Molecular Theory

PD - Packing Domain

PWS - Partial Wave Spectroscopic (microscopy)

RNA Pol II - RNA Polymerase II

RNA-seq - RNA sequencing

scRNA-seq - Single-Cell RNA Sequencing

Se - Sensitivity

SR - Super-Resolution (microscopy)

SRRW - Self-Returning Random Walk (statistical model)

SR-EV - Self Returning-Excluded Volume (polymer model)

STORM - Stochastic Optical Reconstruction Microscopy

TAD - Topologically Associating Domain

TF - Transcription Factor

t-SNE - t-Distributed Stochastic Neighbor Embedding

Table of Contents

List of Tables

List of Figures

regression (B) analyses for BJ cells treated with DXM for 0, 16, and 32 hours. For (A), we assume the linear regression fit used to calculate contact probability scaling follows a normal distribution $\mathcal{N}(\mu_s, \sigma_s)$ where the mean contact probability scaling, μ_s is the slope of the regression and standard deviation, σ_s is the root-mean-square error (RMSE) of the residuals. Contact probability scaling is significantly different between control, and 16 hour and 32 hour DXM treated BJ cells ($P < 0.001$). (C-D) PWS analysis of BJ cells at 0, 16, and 32 hour time points for untreated cells (C) and DXM treated cells (D). (C) Chromatin packing scaling D measured in untreated live BJ cells shows no observable change in D over a 16 hour period and only a small decrease in D over a 32 hour period that is not statistically significant ($P > 0.05$). (D) Using double-sided student-t-test, $P < 0.005$ for DXM treated cells at 16 and 32 hour time points. 150

2.14 Measuring chromatin packing scaling and contact probability scaling alterations induced by dexamethasone (DXM) treatment in A549 cancer cells. (A&B) Characterization of A549 chromatin structure with and without DXM treatment. From left to right: TEM images of chromatin structure with ChromEM staining, scale bar: 1μm. PWS map of chromatin packing scaling, scale bar: 10 μm. Qualitatively, both ChromTEM and PWS images show that DXM treatment homogenizes chromatin packing. (C) Hi-C contact map of human chromosome 1 rendered with 5 kbp resolution for the control and DXM treated

A549 cells. (D) ACF analysis of ChromTEM images of A549 cells. The average ACF of the control group (blue) is significantly different from the average ACF of the treated group (red). The shaded regions represent standard errors. D was calculated from the PD regime (50 nm to 100 nm) by performing a linear regression fit to the ACF in the log-log scale. (E) Contact probability analysis performed on whole-genome intrachromosomal Hi-C contact data. Contact probability scaling (s) was calculated from a linear regression fit (dotted line) of the contact probability curve in the log-log scale between genomic distance $10^{4.4}$ and $10^{5.5}$ bp. (F-H) Chromatin packing scaling alterations induced by DXM treatment measured using ACF analysis of TEM images and PWS and changes in contact probability scaling of Hi-C contact data. Across the platform, consistent changes were observed in chromatin packing scaling upon treatment. Using double-sided student-t-test, $P = 0.051$ for ChromTEM, $P < 0.005$ for PWS. (I) Comparing distributions of contact probability scaling for A549 cells calculated from analysis of Hi-C contact matrices. We assume the linear regression fit used to calculate contact probability scaling follows a normal distribution. Contact probability scaling is significantly different between control, and 12 hour DXM treated A549 cells ($P < 0.001$). (J) Chromatin packing scaling D measured in untreated live A549 cells at 0 and 12 hours shows no observable change in D at time points comparable to

3.1 The Chromatin Packing Macromolecular Crowding (CPMC) model integrates molecular and physical regulators of transcription. The regulators influencing transcription reactions can be generally divided into two categories: (A-D) molecular regulators (k_m, K_D, and $[C]_{tot}$) and (E-H) physical regulators (D, $\phi_{in,0}$, and N_d). (A) The CPMC model describes transcription as a series of diffusion-limited chemical reactions. *Ex vivo*, expression depends on (B) concentration of transcriptional reactants $[C]_{tot}$ (TFs (green), RNA Pol II (yellow)), (C) RNA Pol II elongation rate, k_m, and (D) the disassociation rate of RNA Pol II from the transcription start site (TSS) K_D. (E) (Left) In addition to the molecular determinants, transcription reactions are influenced by the highly dense and complex nuclear environment. The concentration of the main crowder within the nucleus, chromatin, can be measured by ChromTEM. As an example, a ChromTEM image of a nucleus from an A549 lung adenocarcinoma cell is shown. (Right) ChromTEM measurements of CVC demonstrates that chromatin density varies throughout the nucleus. (F) Representative PWS image of an A549 cell demonstrating the existence of chromatin PDs as regions of elevated chromatin packing scaling, which vary throughout the nucleus. (G) A PD with a higher D (right) has a more heterogeneous density distribution and a greater accessible surface area compared to a PD with a lower D (left). (h) N_d is the genomic size (in bp) of a chromatin packing domain and can range from less than 100 kbp to several Mbp.

probability of survival. Cells with higher average D (right, D_b) have increased rates of change in gene expression induced by an exogenous stressor by a factor δ relative to the changes in lower-D cells (left, D_a). For the higher-D cells, this may increase the probability of the cell remaining viable by reaching a critical threshold of expression of pro-acclimation genes compared to the lower-D cell which is unable to meet this threshold. (B&C) The fraction of high-D cells in a cell culture increases after treatment with paclitaxel for 48 hours, suggesting that cells with higher D are more likely to survive exposure to a cytotoxic chemotherapeutic agent. (B) The percentage of cells having D above the top quartile of a control cell population (y-axis) increases in cells that survive treatment with paclitaxel for 48 hours. For both conditions, each dot represents percentage of high-D cells in one replicate for a total number of $N = 5$ replicates per condition. (C) Combination treatment with the D-lowering celecoxib agent and then paclitaxel for 48 hours (Combo) results in increased elimination of cancer cells compared with untreated controls and cells treated with paclitaxel (PAC) alone. (D) CPMC model predictions of the relative transcriptional malleability coefficient δ for initially lowly-expressed (blue spline) and highly-expressed genes (red spline). Here, $D_a = 2.3$ and $D_b = 2.5$, which is comparable to experimentally observed differences in celecoxib-treated versus untreated A2780 cells. (E) scRNA-seq on A2780 cells was performed to compare transcriptional profiles of

control A2780 cells (high-D population) and cells treated with 75 μM of a D-lowering agent celecoxib (low-D population) and their response to treatment with 5 nM paclitaxel (stressor) for 16 hours. Initially lowly-expressed and initially highly-expressed genes are defined based on control expression levels. Genes are grouped based on their quantile of $log_2(E_{PAC}/E_{control})$ and the mean and standard errors of each quantile for initially lowly-expressed genes (blue dots, 300 genes/data point) and initially highly-expressed genes (red dots, 100 genes/data point) are plotted. (F) Gene ontology analysis identified biological processes that are most significantly involved in the response to 48 hour paclitaxel treatment. Upregulated genes were defined as those with at least 2× increase in expression. (G) D-facilitated upregulation (δ) of the stress-response genes identified by the GO analysis (red points, 150 genes/data point) was similar to that for all upregulated genes (blue points, 650 genes/data point). 202

3.4 Chromatin packing scaling regulates intercellular transcriptional heterogeneity of cancer cells. (A-E) 3D projections of scRNA-seq data (TPM values of 8,275 expressed genes) onto reduced t-SNE space for 5 conditions: (A) control cells ($N = 46$), (B) cells treated with 5 nM paclitaxel for 16 hours (1 6hr PAC, $N = 55$), (C) 5 nM paclitaxel for 48 hours (48 hr PAC, $N = 5$), (D) 75 μm celecoxib for 16 hours (16 hr CBX, $N = 62$), (E) and combination of 75 μM celecoxib and 5 nM paclitaxel for 16 hours (16 hr Combo, $N = 59$). The size of the cluster

indicates the transcriptional heterogeneity within the population of surviving cells for each condition. (F) The radius of genomic space R_c (the radius of clusters through A-E) increases as a function of D, which was measured by live-cell PWS microscopy at each time point prior to sequencing. Cells treated with paclitaxel (higher D) have greater transcriptional heterogeneity, especially when compared to cells treated with the D-lowering celecoxib agent. Likewise, the CPMC model (red curve, right side y-axis) shows that intercellular transcriptional heterogeneity increases with D. Error bars represent the standard error of D calculated from PWS measurements (x-axis) and R_c (y-axis) for each condition. (G) Relative expression of high-D versus low-D cells in response to paclitaxel treatment for genes associated with DNA repair pathways, which are upregulated in 48 hour paclitaxel-treated cells. For each condition (Control, 16 hr PAC, 2 hr CBX, 16 hr Combo), TPM values of these genes (48 in total) were averaged within each cell. Next, expression of paclitaxel-stimulated cells was normalized by the average of the corresponding unstimulated population. The resulting intercellular distribution of relative expression levels is shown. Dashed lines represent mean relative expression. Solid red and blue arrows represent the standard deviation of distributions $E_{PAC}/E_{Control}$ and E_{CBX}/E_{Combo}, respectively. For these stress response genes, cells with a higher initial D versus cells with a lower initial D had an increase in transcriptional malleability ($\uparrow \delta$) as well as a higher intercellular

scaling have both an increased average and spread of upregulation of potentially critical genes for cell survival. Thus, high-D cells (in red) will have a higher probability of upregulating key gene(s) beyond a threshold, x_{crit}, before a critical cell decision timepoint. Consequently, these high-D cells will have a lower death probability, θ, which will increase cell survival. Conversely, low-D cells (in blue) might not be able to upregulate genes beyond the critical threshold within the same amount of time, and will have a higher death probability. 216

3.7 CDA parameters influence death probability of cells under cytotoxic stress. (A) Cells with higher $\langle D \rangle$ have increased transcriptional malleability k and heterogeneity s compared to reference cells with $D_a = 2.6$. (B) Death probability compared to $\langle D \rangle$ follows a general sigmoidal relationship that is dependent on the critical upregulation rate of certain gene(s), x_{crit}. Higher x_{crit} results in increased death probability, θ, for cells with the same $\langle D \rangle$. (C) Initial relative expression of genes $ln(E_i/\bar{E}_i)$ influences the sigmoidal behavior of the $\langle D \rangle$ versus death probability relationship. Genes with lower initial relative expression have a steeper sigmoidal curve. (D) Agreement was found between the death probability curve predicted by the CDA model and the experimental death probability values calculated for cell clusters with varying initial $\langle D \rangle$ after a 48 hour oxaliplatin treatment of HCT116 cells. After optimization, fitting parameters were determined to be $x_{crit} = 6.9$ and $\beta_a = 19.1$. 221

A2780.m248 cells ($N = 259$), 5-fluorouracil treated A2780.m248 cells ($N = 100$, $P = 3.9 \times 10^{-3}$), paclitaxel treated A2780.m248 cells ($N = 45$, $P = 4.7 \times 10^{-6}$), and oxaliplatin treated A2780.m248 cells ($N = 85$,$P = 1.5 \times 10^{-18}$); control HCT116 cells ($N = 262$), and oxaliplatin treated HCT116 cells ($N = 289$, $P = 1.7 \times 10^{-35}$); control MDA-MB-231 cells ($N = 128$), 5-fluorouracil treated MDA-MB-231 cells ($N = 81$,$P = 4.1 \times 10^{-2}$), oxaliplatin treated MDA-MB-231 cells ($N = 59$, $P = 2.8 \times 10-5$), and paclitaxel treated MDA-MB-231 cells ($N = 36$,$P = 4.7 \times 10^{-5}$); control MES-SA cells ($N = 265$), docetaxel treated MES-SA cells ($N = 194$, $P = 2.0 \times 10^{-2}$), and gemcitabine treated MES-SA cells ($N = 101$, $P = 4.0 \times 10^{-13}$); control MES-SA.MX2 cells ($N = 203$), gemcitabine treated MES-SA.MX2 cells ($N = 103$, $P = 7.3 \times 10^{-6}$), and docetaxel treated MES-SA cells ($N = 106$, $P = 1.7 \times 10^{-8}$). Significance was determined using Student's t-test with unpaired, unequal variance on the average nuclear D of the treated group against the control group within each cell line ($^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$). (D) Representative PWS microscopy images of control and chemoevasive cells for each chemotherapy tested. Image pseudocolor is D, with brighter red corresponding to higher D values. Scale bars are 15 μm. A2780, A2780.m248, HCT116, MDA-MB-231, MES-SA, and MES-SA.MX2 were treated for 48 hours with 5-fluorouracil, paclitaxel, oxaliplatin, 5-fluorouracil, gemcitabine, and docetaxel respectively as representatives of all cell line and chemotherapy combinations. 239

3.12 Chromatin packing scaling D increases with selective resistance to chemotherapy. (A) Representative PWS microscopy images of ovarian carcinoma A2780 wild-type (WT), and TP53-mutant A2780.m273 (M273), A2780.m175 (M175), and A2780.m248 (M248) cells. Arrows indicate representative nuclei. Scale bars, 10 mum. Pseudo-color: D. (B) Under normal growth conditions, D remained similar to the WT in the A2780.m273 subclone and increased in the A2780.m248 subclone ($P = 1.0 \times 10^{-59}$) relative to the WT A2780 cells. (C) Analysis of TCGA data for high-grade serious epithelial ovarian carcinoma patients revealed a strong correlation between median survival (as reported by TCGA) and D (as measured by PWS). (D) Representative PWS microscopy images of leiomyosarcoma MES-SA and mitoxantrone resistant MES-SA/MX2 derivative (MX2) cells. Arrows indicate representative nuclei. Scale bars, 10 mum. Pseudo-color: D. (E) Under normal growth conditions, D was increased in the MES-SA.MX2 chemoresistant subclone ($P = 3.1 \times 10^{-30}$) compared to the sensitive MES-SA subclone. Significance was determined using Student's t-test with unpaired, unequal variance on the average nuclear D of the mutant subclone against the WT within each cell line ($^{***}P < 0.001$). $N = 1877$ A2780, $N = 309$ M273, $N = 237$ M175, $N = 1321$ M248, $N = 836$ MES-SA, and $N = 558$ MX2 cells. 242

3.13 Increased transcriptional malleability for higher D cells is a generalizable phenomenon. The transcriptional malleability coefficient $\delta = \frac{E_{2,b}/E_{1,b}}{E_{2,a}/E_{1,a}}$

was determined from additional bulk RNA-seq experiments on A2780 cells and TP53 mutated clone A2780.m248 cells along with propranolol, another D-lowering compound. PWS measurements showed a 2% decrease in D in A2780 cells after propranolol treatment for 16 hours and a $\sim$5% decrease in D in m248 cells treated separately with celecoxib and then propranolol for 16 hours. (A) Transcriptional malleability in A2780 cells treated with propranolol to lower D. All treatment conditions include: control, 16 hour propranolol, 16 hour paclitaxel, and 16 hours paclitaxel plus celecoxib. (B&C) δ tested in m248 cells treated with (B) celecoxib and (C) propranolol as D-lowering compounds for 16 hours. All treatment conditions include control, 16 hours celecoxib/propranolol, 16 hours paclitaxel, 16 hours paclitaxel plus celecoxib/propranolol. All results are based on the expression data at $t = 16$ hours. Error bars represent the standard error of δ for all genes within each quantile. There are three biological replicates for every condition. 243

3.14 Transcriptional heterogeneity is increased in high-D cells. (A) Spread of pairwise Euclidean distance was calculated between cells in each condition for genes associated with DNA repair pathways that are upregulated in 48 hour paclitaxel treated cells. (B) Coefficient of variation (COV) across treatment populations of genes grouped by control expression levels normalized by control COV. Genes were first binned into groups of $\sim$100 genes (80 quantiles total) each based on relative control expression, which are assumed to be exposed to roughly

group ($^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$). (B) Representative PWS microscopy images of control and CPT treated cells. Image pseudocolor is D, with brighter red corresponding to higher D values. Scale bars are 15 μm. A2780 cells were treated for 30 minutes with celecoxib and digoxin. (C) A 30-minute treatment with select CPT compounds decreases population-wide D in the nine different cell lines tested. Violin plots correspond to control A2780.m248 cells ($N = 125$), celecoxib-treated A2780.m248 cells ($N = 36$, $P = 9.3 \times 10^{-3}$), valproic acid-treated A2780.m248 cells ($N = 51$, $P = 4.9 \times 10^{-8}$), and digoxin-treated A2780.m248 cells ($N = 91$, $P = 1.5 \times 10^{-8}$); control AsPC-1 cells ($N = 107$), and 9-ING-41-treated AsPC-1 cells ($N = 111$, $P = 1.1 \times 10^{-14}$); control HCT116 cells ($N = 64$), 9-ING-41-treated HCT116 cells ($N = 85$, $P = 9.0 \times 10^{-10}$), aspirin-treated HCT116 cells ($N = 130$, $P = 9.0 \times 10^{-23}$), and celecoxib-treated HCT116 cells ($N = 75$, $P = 5.79 \times 10^{-35}$); control L3.6pl cells ($N = 150$), and 9-ING-41-treated L3.6pl cells ($N = 163$, $P = 1.6 \times 10^{-19}$); control M9K cells ($N = 191$), and 9-ING-41-treated M9K cells ($N = 224$, $P = 1.1 \times 10^{-27}$); control MDA-MB-231 cells ($N = 89$), celecoxib-treated MDA-MB-231 cells ($N = 86$, $P = 5.6 \times 10^{-4}$), and digoxin-treated MDA-MB-231 cells ($N = 83$, $P = 1.2 \times 10^{-10}$); control MES-SA cells ($N = 314$), digoxin-treated MES-SA cells ($N = 342$, $P = 1.7 \times 10^{-5}$), and celecoxib-treated MES-SA cells ($N = 275$, $P = 1.5 \times 10^{-29}$); control MES-SA.MX2 cells ($N = 227$), digoxin-treated MES-SA.MX2

cells ($N = 252$, $P = 1.3 \times 10^{-30}$), and celecoxib-treated MES-SA cells ($N = 216$, $P = 8.8 \times 10^{-55}$); control OVCAR-8 cells ($N = 65$), and 9-ING-41-treated OVCAR-8 cells ($N = 82$, $P = 4.2 \times 10^{-7}$). Significance was determined using Student's t-test with unpaired, unequal variance on the average nuclear D of the treated group against the control group within each cell line ($^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$). 248

3.17 Inhibition of cancer cells by chemotherapy increases upon addition of "strong" CPTs. To examine the relationship between cell death and initial chromatin packing scaling experimentally, we focused on three complementary methods for assessing cell survival – cell coverage analysis using transmission microscopy, which provided information on total population viability, automated cell counting paired with dead and apoptotic cell stains, and flow cytometry measurement of live cell caspase 3/7 staining, which captured the relative induction of apoptosis. Altogether, these measures provided quantifiable information on cell inhibition as well as percentage viability for each population. Mild CPTs, which cause a relatively smaller decrease in D (valproic acid and aspirin) are less effective at improving chemotherapeutic efficacy than "strong" CPTs (celecoxib and digoxin). Error bars are standard error. Significance was determined using Student's t-test with unpaired, unequal variance on the inhibition of the co-treated group against the chemotherapy treated group within each cell line ($^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$). 249

exposure to chemotherapy, allowing them to survive chemotherapeutic stress. **Chapter 4** characterizes the effects of the physicochemical environment on DNA-like and chromatin-like systems. (Left) The ability of divalent Mg^{2+} (green cations) to form ion bridges between two phosphates allows for increased charge neutralization and compaction of these loop systems. These effects are dependent on average system density, determined by grafting density σ. (Right) The bulk monovalent electrolyte environment, pH, and system density influence the charge and structural parameters, including bond angle, nucleosome spacing, and radius of gyration, of chromatin. 331

CHAPTER 1

Introduction

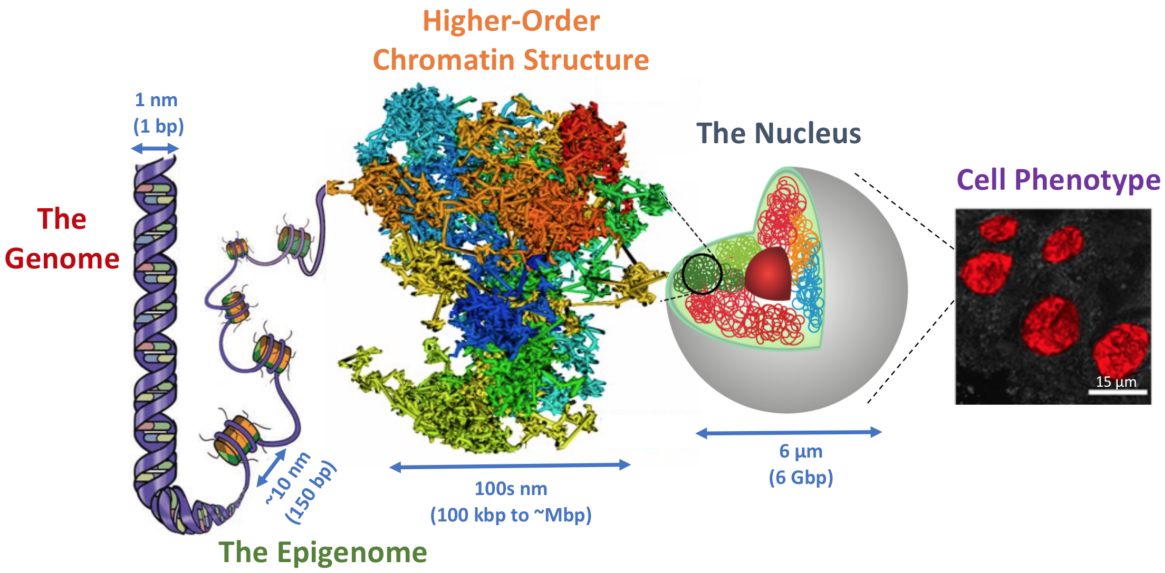

Figure 1.1. Chromatin structure relates to function at multiple length scales from the genome (Chapter 1.1.1), to the epigenome (Chapter 1.1.3), to higher-order chromatin structures (Chapter 1.1.4). Adapted from [120, 70].

1.1.1. DNA as the Heritable Material

The discovery of DNA as a heritable material spanned an entire century. As the "Father of Genetics", Gregor Mendel discovered fundamental laws of inheritance, now referred to as Mendelian inheritance, through experiments performed on pea plants in the 1850s and 1860s. In 1869, Friedrich Miescher was the first to isolate DNA, which he termed "nuclein", from lymphoid white blood cells. In 1881, Albrecht Kossel identified the four types of nucleic acids that constitute the building blocks of DNA: adenine (A), thymine (T), cytosine (C), and guanine (G). Throughout the late 19^{th} and early 20^{th} centuries, prominent scientists such as Walther Flemming, Oswald Avery, Barbara McClintock, and

Erwin Chargaff discovered key principles that proved DNA as the material of genetic inheritance and that this genetic code was passed onto daughter cells during the process of mitosis, or cellular division. Finally, in the 1950s, James Watson, Francis Crick, and Rosalind Franklin first visualized the three-dimensional structure of DNA using X-ray crystallography and determined that DNA is a double-helix, where nucleotides where stabilized by a sugar-phosphate backbone [293]. The importance of DNA was so solidified in the collective mentality of scientists that, in the 1990s and early 2000s an international research consortium, called the Human Genome Project, was formed to identify, map, and sequence all genes within the human genome [154].

1.1.2. A General Overview of Disordered Chromatin Structure and Function

However, unlike the bacterial genome, the eukaryotic genome does not simply exist as naked DNA in the nucleus. In human cells, 2 meters of DNA is packed into an approximately 6 μm nucleus as chromatin. Unlike proteins, chromatin is a dynamic, fluid-like material without well-defined primary, secondary, and tertiary structures [180, 308, 139]. In 1997, Karolin Luger discovered the crystal structure of the nucleosome, the fundamental unit of chromatin [174]. To form the nucleosome, 147 base pairs (bp) of DNA wrap around a core histone octamer composed of dimers of H2A, H2B, H3, and H4 proteins, each with their own intrinsically disordered histone tails [174]. The core particle adopts a squat cylindrical shape, with a diameter and a height of approximately 11 nm and 5.5 nm, respectively [210]. Nucleosomes are connected by linker DNA, which altogether form the 10 nm chromatin fiber [149]. Central to the textbook view of chromatin packing is that the 10 nm 'beads-on-a-string' fiber assembles into the 30 nm fiber, which further folds

into 120 nm chromonema that organize into 300 to 700 nm chromatids, and ultimately, mitotic chromosomes [26, 25, 81, 27]. In interphase nuclei, chromosomes themselves are spatially partitioned into chromosome territories [54].

However, the key tenant of this view, the 30 nm fiber, has been challenged by an abundance of recent evidence. Various studies using cryo-electron microscopy, small-angle X-ray scattering, electron spectroscopy imaging, and super-resolution (SR) microscopies failed to observe 30 nm fibers in both interphase chromatin and mitotic chromosomes in numerous cell lines [75, 131, 91, 240, 212]. For example, Ricci et al. observed the existence of heterogeneous nucleosome 'clutches' at the level of the primary fiber, the size of which depends on the local epigenetic state of chromatin and the cell type [240]. Recently, a combination of DNA-specific staining (ChromEM) and multi-tilt electron tomography (ChromEMT) observed *in situ* that the chromatin folds into disordered fibers that have diameters between 5 to 24 nm during both interphase and mitosis (Fig. 1.2) [212]. Additionally, ChromEMT demonstrated that the nucleus is highly crowded, with chromatin volume concentrations (CVC) ranging from 12-52% in interphase cells and $>$ 40% in mitotic cells [212]. Altogether, these studies suggest that the interphase chromatin and mitotic chromosome organization is constructed by 10 nm fibers without folding into ordered 30 nm fibers [237, 179]. In this new paradigm, the 10 nm fibers condense into highly disordered and interdigitated states, which may be constantly moving and rearranging at the local level [180, 111]. Label-free Partial Wave Spectroscopic (PWS) microscopy has also demonstrated large mass density heterogeneities between length-scales of 20-300 nm throughout the entire nucleus [3], further proving that chromatin folding is disordered and heterogeneous.

In general, the chromatin organization may influence critical cellular processes, ranging from gene transcription to DNA replication and repair [52, 146, 233]. This thesis will focus specifically on the inter-relationship between chromatin structure and transcription. Transcription of most protein-coding genes is enacted by RNA polymerase II (RNA Pol II), which binds to the promoter region of the gene, along with initiating factors such as transcription factors (TFs), to form a pre-initiation complex [52]. Promoters themselves are activated by TFs, which recruit proteins that modulate promoter accessibility, for example by creating nucleosome-depleted regions to increase transcription of genes in that area [85, 280]. Enhancers are genomic elements located far away from promoters on the linear DNA sequence that can influence the activation of promoters when the enhancer is in close 3D proximity to its cognate promoter [52]. Thus, transcription processes can remodel local chromatin organization, but the four-dimensional structure of chromatin can also influence enhancer-promoter contacts and facilitate gene expression.

1.1.3. The Epigenome: Heredity on Top of the Genome

Chromatin writers and erasers chemically modify the epigenome while chromatin readers interpret the epigenetic code and reorganize the local chromatin structure. The term epigenetics refers to modifications outside of the DNA sequence itself that contribute to a heritable phenotype. These include DNA methylation, covalent modifications to histone tails, including acetylation, methylation, ubiquitylation, phosphorylation, and sumoylation, as well as RNA such as long noncoding RNAs (lncRNAs) and micro RNAs (miRNAs), all of which are fully detailed in several key reviews such as [306, 14].

DNA methylation occurs primarily at CpG sites, DNA regions with sequential stretches of CG repeats, and approximately 70-80% of CpG sites in the mammalian genome have methylated DNA [103, 162]. Functions of DNA methylation include gene repression when promoters and enhancers are heavily methylated [294], gene activation when gene bodies are methylated [305], repression of transposons, genomic imprinting [80], and specific DNA methylation patterns are even associated with aging [117]. Chromatin writers include DNMT (DNA methyltransferase) proteins, which maintain DNA methylation, and chromatin erasers include TET proteins, which remove DNA methylation [103]. Chromatin readers include methyl-CpG-binding domain (MBD) proteins, which interact with nucleosome remodelling and histone deacetylase complexes [199].

Histone tail modifications are generally characterized into those associated with transcriptionally active euchromatin and others which are associated with transcriptionally inactive heterochromatin. Lysine acetylation (e.g., H4K16ac, H3K27ac) neutralizes the positive charge of histone tail lysines, which disrupts electrostatic interactions between histones and DNA [14]. Chromatin writers include histone acetyltransferases (HATs) such as the CBP/p300 family and erasers include histone deacetylases (HDACs) [14, 306]. Chromatin readers of acetylation marks have bromodomains, including the ATP-dependent SWI/SNF nucleosome remodelling complexes which increase DNA accessibility for active transcription [112]. Methylation of lysine and arginine histone tail residues can be associated with either euchromatin or heterochromatin, depending on the exact mark itself. H3K9 di-, and tri-methylation are associated with constitutive heterochromatin, defined by stable repression in key structural regions such as centromeres and telomeres [14].

The SUV39H1 histone lysine methyltransferase chromatin reader enacts H3K9 methylation and heterochromatin protein 1 (HP1) reads H3K9 methylation and oligomerizes in regions with high H3K9me3 density to repress transcription [238, 16]. H3K27 di- and tri-methylation are associated with facultative heterochromatin, which is associated with developmental transcriptional repression and is known as the more "plastic" of the two heterochromatin types. Polycomb Repressive Complexes (PRCs) are both readers and writers of H3K27 methylation. The EZH2 component of PRC2 catalyzes H3K27 di- and tri-methylation and PRC1 mediates transcriptional repression via histone ubiquitylation and chromatin compaction [43, 292]. Histone tail marks such as H3K4 and H3K36 methylation are associated with active transcription [258, 15]. Chromatin erasers of methylation include lysine-specific demethylases and JMJD2 which specifically demethylates trimethylated lysines [14].

Overall, the balance of the chemical modifications comprising the epigenome contributes to the a spectrum of states ranging from open chromatin regions with highly active transcription to compacted regions that are transcriptionally repressed. One notable phenomenon which the epigenome may contribute to is phase separation, a physical process that creates two distinct phases from a homogeneous mixture. For example, first discovered by the Narliker group, oligomerization of HP1 proteins associated with constitutive heterochromatin (H3K9 methylation) have been shown to contribute to phase-separated condensates [155, 253]. Additionally, super-enhancers, regions of the mammalian genome that contain multiple enhancers and are associated with large swaths of

H3K27 acetylation, are thought to form phase-separated condensates of active transcription by cooperatively recruiting transcriptional regulators such as the Mediator complex [250, 116].

By definition, the epigenome contributes to heritability of cellular phenotype. A hierarchy of gene expression patterns are faithfully reestablished after mitosis to ensure the maintenance of cell identity. Transcriptional memory propagation across generations may occur through mechanisms such as mitotic bookmarking [216]. Specifically, bookmarking TFs remain bound to condensed chromosomes and allow gene expression to occur throughout mitosis, potentially helping to reestablish transcription patterns following cell division [216, 215]. In addition, both active and repressive histone modifications are preserved throughout the cell cycle, although genomic regions with larger, more repressive heterochromatin marks are more stably inherited [215, 58].

In addition, the epigenome is highly responsive to external stimulation. Changes in extracellular pH regulate histone acetylation levels that, in turn, help to buffer intracellular pH [186]. Cells exposed to mechanical stress increase cytoplasmic-to-nuclear shuttling of histone deacetylases, which lowers histone acetylation levels and increases chromatin compaction [2]. The epigenetic inheritance of trauma has huge societal implications, including the inter-generational propagation of internalized racism. Differential epigenetic profiles, including DNA methylation of promoters related to stress-response genes, have been discovered in the progeny of children diagnosed with post-traumatic stress disorder from the Holocaust as well as mouse models of childhood trauma [302, 87].

1.1.4. Higher-Order Chromatin Structure

Several recent developments, allowing for both high-throughput sequencing and SR imaging, have helped to uncover higher-order structural elements of chromatin above the level of the epigenome.

High-throughput chromosome conformation capture (Hi-C), the modern version of 3C, 4C, and 5C approaches, measures the two-dimensional chromatin connectivity throughout the entire genome [168]. As described in the seminal 2009 paper by Lieberman-Aiden and colleagues, the process of Hi-C involves [168]:

(1) Crosslinking DNA that is in close 3D proximity,

(2) Cutting the genome with restriction enzymes,

(3) Biotinylating the ends of both DNA fragments,

(4) Ligating the ends of fragments together,

(5) Purifying and shearing all DNA,

(6) Pulling down biotinylated fragments, and

(7) Using paired-end sequencing to determine genomic regions that are in direct contact.

Alignment of the sequenced reads to your genome of interest results in a 2D contact map of all chromatin-chromatin contacts within the genome. Several sequencing-based techniques, most notably Chromatin ImmunoPrecipitation Sequencing (ChIP-seq), allow for characterization of 1D epigenetic marks throughout the genome, which can be co-localized with Hi-C maps to determine how the epigenome modulates chromatin structure [128]. *In situ* Hi-C improved the resolution of Hi-C by almost an order of magnitude, from just under 1 megabase pair (Mbp) to 10's of kilobase pairs (kbp) [232]. Micro-C

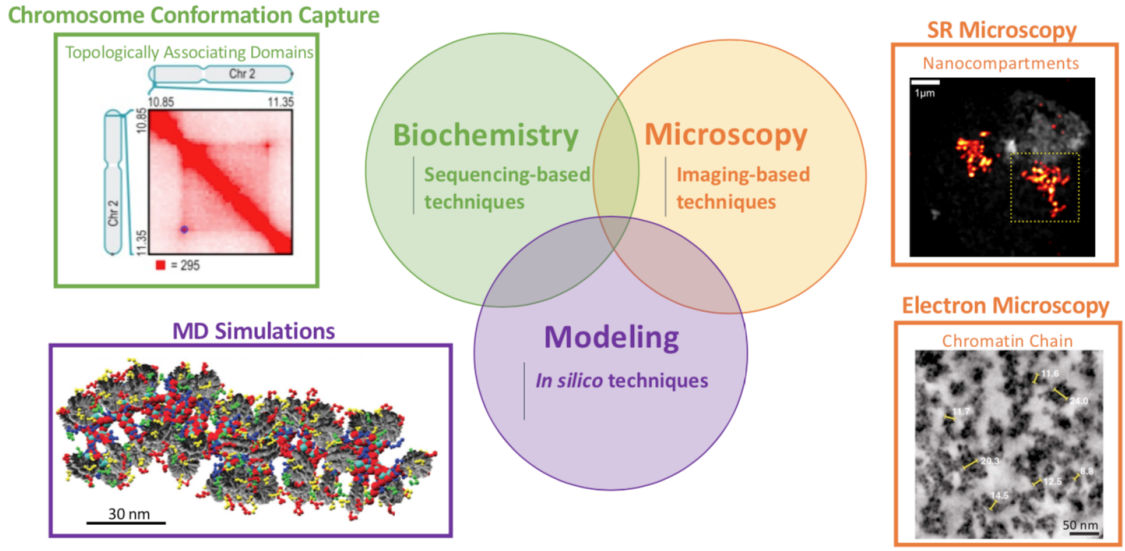

Figure 1.2. Summary of techniques to interrogate chromatin structure. Recent advances in high-throughput sequencing, super-resolution (SR) optical and electron microscopies, and computational modeling have uncovered several key tenants underlying supranucleosomal chromatin structure. (Green) Topologically Associating Domains (TADs) are key structural features that are distinctly visible on 2D High-throughput Chromosome Conformation Capture (Hi-C) contact maps [68, 262, 232]. (Orange) SR optical microscopy techniques, which label TADs and observe their three-dimensional structure, observed that heterochromatic TADs form compacted nanocompartments [275]. DNA-specific electron microscopy techniques show that the primary chromatin fiber is disordered [212]. (Purple) Computational simulations can be employed to elucidate principle mechanisms underlying chromatin structure. Shown are mesoscale simulations from the Schlick group investigating how linker histone H1, which is not part of the canonical nucleosome core particle, and divalent cations both influence internucleosome interactions [104].

can interrogate chromatin-chromatin contacts at the single nucleosome (∼100s bp) level, but is less sensitive to long-range contacts and more costly than traditional Hi-C methods [119].

Several key structural elements are observable from Hi-C contact maps. Topologically associating domains (TADs) are regions of genomic size ∼100s kbp with a high frequencies of self-self chromatin contacts (Fig. 1.2) [68, 262, 232, 274]. TADs are thought to form through dynamic biophysical processes, including loop extrusion [251]. During loop extrusion, cohesin actively extrudes chromatin and CCCTC-binding factor (CTCF) proteins are bound to specific genomic regions and act as a stopping signal for the cohesin proteins [90]. The existence of TAD cliques, clusters of TADs which interact frequently between TADs, have also been demonstrated, and such cliques are enriched in heterochromatin marks [170]. A and B compartments are larger than TADs (on the order of several Mbp) and are determined from eigenvector decomposition of the 2D Hi-C contact matrix to categorize preferential associations within versus between compartments [168]. 'A' compartments are associated with transcriptionally active euchromatic marks and 'B' compartments are associated with transcriptionally inactive heterochromatic marks [168]. There is an inherent competition between biophysical processes driving TAD formation versus compartmentalization, as depletion of cohesin proteins involved in loop extrusion strengthens compartmentalization [206, 208]. Topological constraints are enacted by architectural proteins such as cohesin to form TADs. Chromatin can also be physically tethered to the nuclear lamina by lamin proteins to form lamin associated domains (LADs) [118]. LADs are associated with the transcriptionally inactive Hi-C B compartments and are enriched in constitutive heterochromatic marks (H3K9me2/3) [118]. As Hi-C detects chromatin contacts from millions of nuclei, this only provides a population-average snapshot of higher-order chromatin structures. Notably, single-cell Hi-C methods have demonstrated the potential existence of TADs in individual nuclei, although a high degree

of intercellular heterogeneity in TAD distribution has been reported [198, 281]. Despite the advent of single-cell Hi-C methods, Hi-C itself is still a two-dimensional technique that detects chromatin organization at one fixed timepoint.

Recently, the development of super-resolution (SR) microscopies, have allowed for interrogation of chromatin structure down to the level of tens of nanometers. This resolution is well below the diffraction limit of light, which is around 200-300 nm for spectra in the visible range. Such microscopies include stochastic optical reconstruction microscopy (STORM) and photoactivated localization microscopy (PALM). Paired with labeling methods, fluorescence in situ hybridization-based (FISH-based) labeling methods such as Oligopaint [24], this allows for the study of 3D chromatin organization of specific genomic regions. Several types of domains have been observed using such microscopy techniques, including punctate "chromomeres" in both interphase and mitotic cells and replication domains with coherent motion [101, 207].

Multiple independent studies have also reported the existence of TAD-like chromatin domains using SR microscopies [275, 32, 194]. Repressed, heterochromatic TADs form discrete and compacted nanocompartments, which are interspersed with more loosely packed transcriptionally active genomic regions (Fig. 1.2) [275, 194]. Employing high-throughput optical mapping, Finn et al. demonstrated that there is extensive variability in TAD structures between individual cells, pointing to the hypothesis that TADs may only be aggregate behaviors observed at the level of cell population [82]. Upon cohesin removal, domain-like structures remain with similar nanoscale topography, but are randomly localized throughout the genome [32, 194]. Therefore, the specificity, not the existence, of three-dimensional domains may be determined by biophysical mechanisms

that maintain population-average TAD structures. Additionally, Gabriele et al. recently observed the dynamic process of loop extrusion, which contributes to TAD formation, in live cells [92]. This study revealed that the fully looped states, where cohesins are halted by CTCF proteins, only occurs approximately 3% of the time, while 92% of the time loops exist in partially extruded states [92]. Altogether, this points to extensive heterogeneity in the structure of TADs in individual cells due to highly dynamic processes such as loop extrusion.

There is a complex, bidirectional link between TADs and transcription. As reviewed in [286], TADs may influence transcription by: (1) insulating promoters from enhancers located in neighboring TADs, (2) reducing the search space and increasing the probability of promoters and enhancers within the same TAD to find each other [161], and (3) TAD boundaries may act as a barrier to the spreading of euchromatic and heterochromatic epigenetic marks. However, disruption of TAD structures by genetic manipulation of CTCF sites and depletion of loop extrusion-associated proteins do not always have significant effects on gene expression [296, 261]. A very recent result from Kane et al. points to the complexity of this regulation by demonstrating that cohesin, not CTCF, is required for enhancer-promoter activation, but only for enhancers and promoters > 100 kbp apart on the linear DNA sequence [136]. Conversely, transcription also regulates TADs themselves. TAD boundaries are enriched in active genes and these boundaries are weakened by inhibition of transcription [284]. Additionally, negative supercoiling induced by transcription is a key regulator of loop extrusion and chromatin density distribution within loops [204]. Overall, TADs seems to have functional importance but cannot be thought of as the unequivocal fundamental unit of transcription.

1.2. Chromatin and Complex Diseases

Waddington's landscape, a term proposed by Conrad Waddington in 1957, describes an epigenetic landscape where ridges and valleys denote paths that a stem cell can take on its way to differentiation. In general, the epigenome and 4D chromatin structure itself have been implicated in driving cellular differentiation and lineage commitment [225, 50, 22, 224, 33]. Thus, it comes as no surprise that chromatin processes can be dysregulated in the disease state, resulting in complex diseases such as cancer, neurological disorders, and have recently even been implicated in post-traumatic stress disorder [277, 218, 1].

Flavahan et al. posit that all of the "hallmarks of cancer", a concept introduced by Hanahan and Weinberg [109, 110], can be explained by aberrations in chromatin and associated epigenetic mechanisms [84]. Extending the idea of Waddington's landscape, they argue that chromatin insults can result in overly permissive, or "plastic", epigenetic landscapes that could allow for transitions to more malignant cell states [84]. In malignancy, genetic, metabolic, or environmental insults may aberrantly remodel chromatin structure across a hierarchy of length-scales.

The epigenetic landscape of chromatin is altered in many cancer states. Gain-of-function mutations of Polycomb repressor EZH2 block differentiation in prostate, lung, and skin cancers [28, 29, 144]. Global DNA hypomethylation and local promoter/enhancer hypermethylation are common among multiple cancer types, including Acute Myeloid Leukemia [145, 135, 185]. However, Acute Lymphoblastic Leukemia exhibits CpG island hypermethylation and minimal global loss of methylation, indicating the existence of a non-canonical regulation of methylation in other cancer types [113]. Kretzmer et al. recently determined that aberrant DNA methylation occurs early on in malignant cell

state transitions and persists stably, even upon chemotherapy treatment, indicating a potential role of DNA methylation in disease onset [153].

Nucleosome remodeling and three-dimensional chromatin organization are also altered in cancer, which can be linked to epigenetic changes. Exome-wide sequencing studies demonstrate that the ATP-dependent chromatin remodeler, SWI/SNF, is mutated in $> 20\%$ of all cancers, and several of these mutations are oncogenic drivers [133, 12]. The increased activity of TET DNA demethylase, caused by oncometabolites, disrupts CTCF binding, resulting in insulator dysfunction and oncogene activation in gliomas [83]. A and B compartment switching and mixing are accompanied by compartment-specific hypomethylation [307, 129].

The chromatin state has also been heavily implicated in cancer progression and chemoevasion. Drug-tolerant persisters (DTPs), small subpopulations of cancer cells which transiently commit to a slow-cycling drug-resistant state through non-genetic mechanisms, can evade chemotherapy through distinct chromatin-mediated pathways [264, 235]. In addition, clinical studies across multiple cancer types identified alterations in chromatin packing as predictors of cancer progression [134, 246], and survival time of late-stage cancer patients undergoing therapeutic treatments [288]. Although much progress has been made, understanding how alterations in the 4D chromatin structure contribute to disease remains an actively investigated question.

1.3. nanoscale Chromatin Imaging and Analysis (nano-ChIA) Platform

The nano-ChIA platform in the Backman group aims to quantify chromatin organization at broad spatial and temporal scales and to relate this structure to transcription (Fig. 1.3) [167]. Because no individual technique can accomplish this feat, it is necessary to develop a multi-modal platform combining complementary techniques. Such a platform should have high resolution across the entire nucleus with dynamic, live-cell imaging capabilities and analysis methodologies to link these results to genome connectivity and the localization of critical molecular factors. To meet these requirements, we have developed the nanoscale chromatin imaging and analysis (nano-ChIA) platform, which incorporates chromatin scanning transmission electron microscopy (ChromSTEM), chromatin transmission electron microscopy (ChromTEM), Partial Wave Spectroscopic (PWS) microscopy, and Stochastic Optical Reconstruction Microscopy (STORM). Each facet of nano-ChIA interrogates distinct aspects of chromatin architecture as outlined in Table 1.1. Consolidating these modalities, nano-ChIA is a fully quantitative nanoscale imaging platform that complements the genomic information provided by chromatin conformation capture and other sequencing-based techniques. By bridging high-resolution imaging of chromatin structure and molecular processes with high-throughput, label-free analysis of chromatin dynamics in live cells across time scales spanning from minutes to hours, nano-ChIA has the potential to provide insights into crucial questions in 4D genomics.

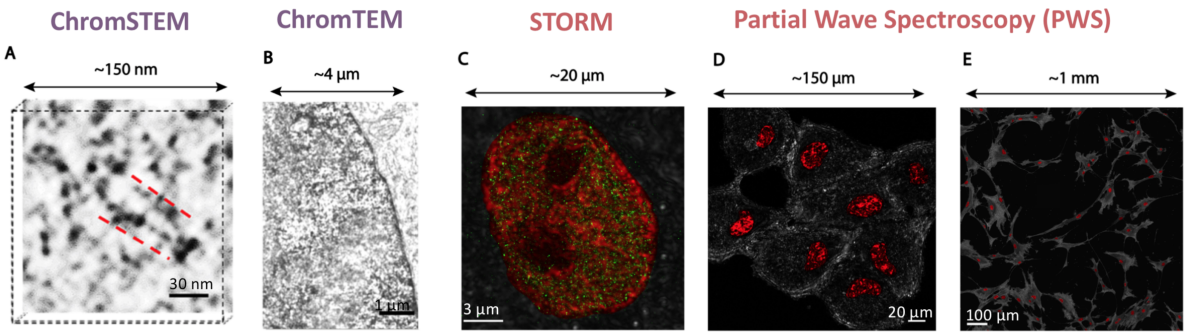

Figure 1.3. nano-ChIA platform. (A) ChromSTEM HAADF tomography characterizes the 3D chromatin structure of a lung adenocarcinoma A549 cell (contrast inverted). The inverted image contrast is inversely proportional to the local DNA density. As the electrons encounter a higher density of DNA along their trajectory, the image contrast appears darker. Individual nucleosomes and linker DNA are resolved at 2 nm spatial resolution. Scale bar: 30 nm. (B) ChromTEM imaging of a BJ cell nucleus on a 50 nm resin section prepared by ChromEM staining. Similar to ChromSTEM, ChromTEM also maps the DNA distribution, but the image contrast follows Beer's law. Scale bar: 1 μm. (C) Coregistered PWS and STORM imaging of chromatin packing scaling (D, red pseudocolor) and active RNA Pol II (green) of an ovarian carcinoma M248 cell nucleus. Scale bar: 3 μm. (D and E) Label-free PWS images of live A549 cells, including both (D) one field of view where chromatin packing variations within nuclei are visible (scale bar: 20 μm) and (E) a 9x9, stitched together, image to demonstrate the ability of PWS to visualize chromatin packing structure of cell populations (scale bar: 100 μm). The pseudocolor represents the chromatin packing scaling inside the cell nuclei.

1.3.1. ChromSTEM/ChromTEM

At the smallest length scales, the nano-ChIA platform combines DNA-specific labeling (ChromEM) with high-angle annular dark-field (HAADF) imaging in STEM (ChromSTEM)and TEM imaging (ChromTEM). Specifically, ChromSTEM, an adaptation of the pioneering work demonstrated by Ou et al. [212], is able to reconstruct chromatin ultrastructure of a thick nuclear cross section at sub-6-nm resolution (Fig. 1.3A) with the

Technique	Contrast	Spatial/ Temporal Resolution	Field of View/ Throughput	Information
ChromSTEM	Chromatin	<6 nm; Fixed timepoint	2 μm x 2 μm x 300 nm; Several cells	3D chromatin density distribution and packing scaling
ChromTEM	Chromatin	6-10 nm; Fixed timepoint	150 μm x 150 μm x 50 nm; Tens of cells	2D chromatin density distribution and packing scaling
STORM	Molecular	<20 nm; Fixed timepoint	10 μm x 10 μm; Several cells	Spatial distribution of macromolecules of interest
PWS	Senses statistical chromatin properties	Sensitivity: 20 - 300 nm; ∼5s	150 μm x 150 μm; 100s to 1000s of cells	Chromatin mass density variations and packing scaling

Table 1.1. Descriptions of imaging modalities in nano-ChIA.

potential to image the entire nucleus by serial sectioning [268]. As ChromSTEM is not high throughput, the platform uses ChromTEM to gain statistical power. By imaging ultrathin (∼50 nm) cross sections with a larger field of view, ChromTEM extends the yield of ChromSTEM from a fraction of a cell nucleus to multiple entire cell nuclei. Although not a 3D technique, ChromTEM provides faster, pseudo-2D quantification of chromatin packing structure from the cross sections of the nucleus at 6 to 10 nm lateral resolution (Fig. 1.3B).

The Click-EM staining procedure was adapted from [212]. Briefly, after fixation, samples are stained by DRAQ5 and then submerged in 3-5'-diaminobenzidine (DAB)

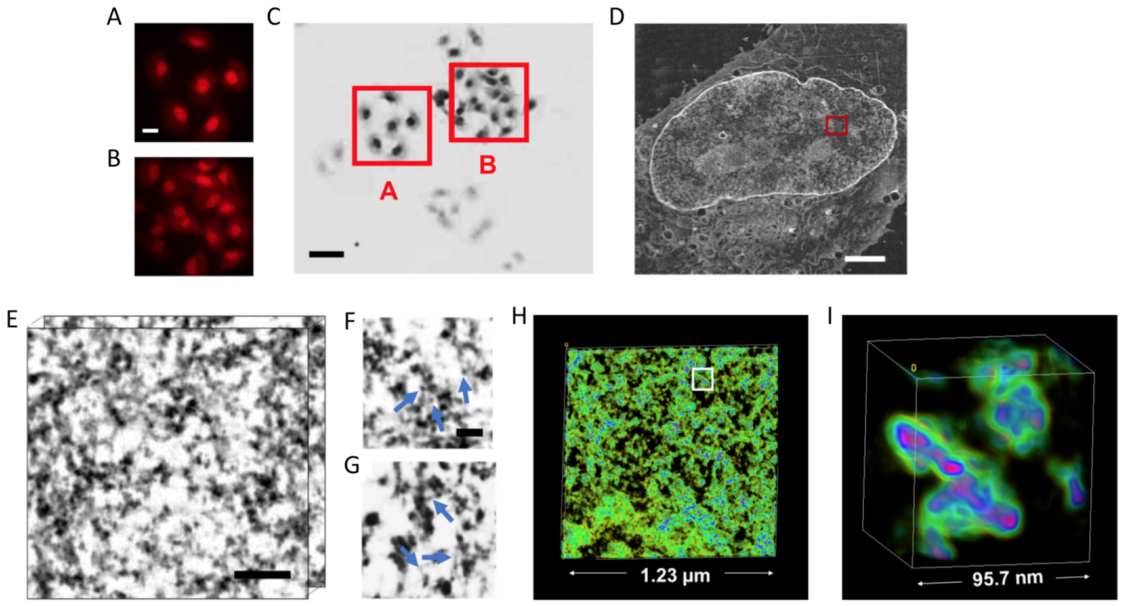

Figure 1.4. ChromSTEM tomography reconstruction of chromatin in an A549 cell. (A-B) The DRAQ5 photo-oxidation process takes 7 min for each region of interest. Scale bar: 10 μm. (C) The DRAQ5-labeled regions were more intensely stained than the nearby regions (red squares; the letter corresponds to the regions in the left panels). Scale bar: 20 μm. (D) STEM image of a 100 nm thick section of an A549 cell in HAADF mode. Scale bar: 2 μm. (E) 3D tomography of the A549 chromatin. Scale bar: 120 nm. (F-G) The fine structure of the chromatin fiber: Nucleosomes (blue arrows in F), linker DNA (blue arrows in G) Scale bar: 30 nm. (H-I) 3D rendering of the chromatin organization, the pseudo-color was based on the intensity of the tomograms. (J) A magnified view of the region labeled by a white square in I. Pink and green regions represent high and low DNA density regions, respectively.

solution. After photobleaching, which induces DAB polymerization onto DRAQ5, chromatin is stained with the heavy metal osmium tetroxide. DAB polymerization increases the contrast and specificity of DNA staining compared to other macromolecules. Ethanol dehydration is performed before resin embedding and sections of varying thicknesses are prepared using an ultramicrotome and diamond knife. After resin embedding, the labeled

regions can be identified based on image contrast in bright field optical micrographs: the photo-oxidized cells appeared significantly darker than the non-photobleached cells (Fig. 1.4A-C). For ChromSTEM, dual-tilt STEM tomography in HAADF mode was performed for part of the nucleus on a ∼100 nm resin section. Within the same tomogram, there are large variations in DNA contrast, potentially indicating the coexistence of euchromatic and heterochromatic regions (red box in Fig. 1.4D). Only interior sections of the nucleus were analyzed, as images with peripheral chromatin also included signal from the nuclear envelope that is difficult to segment from the DNA signal. Unlike the near-binary image contrast from the conventional EM staining and analysis methods [212], ChromSTEM provides continuous variations of the DNA contrast inside the nucleus. The final tomogram (Fig. 1.4E) has a nominal voxel size of 2.9 nm, with clearly resolved nucleosomes (Fig. 1.4F) and linker DNA (Fig. 1.4G). A rendering of the 3D volume of the chromatin is shown in Fig. 1.4H&I.

1.3.2. PWS Microscopy

PWS microscopy is used for for label-free, real-time imaging of statistical chromatin packing across thousands of cells (Fig. 1.3D&E). PWS directly measures variations in spectral light interference. Such variations are caused by light scattering due to refractive index variations from heterogeneities in chromatin density. This interference signal is then processed to characterize the shape of the auto-correlation function (ACF) of chromatin density within the coherence length (∼1 μm in depth) in either fixed or live cells. The mass density ACF determines the chromatin packing scaling (D) in each voxel as detailed in [74]. Chromatin packing scaling is a statistical property derived from polymer physics

that characterizes the relationship between the mass of chromatin (M) and the 3D space the chromatin occupies (r) as $M \propto r^D$ (Fig. 1.5). Although the spatial resolution is ∼250 nm, PWS is sensitive to structural length scales between 20 and 300 nm [46].

The PWS optical instrument is built on a Leica commercial inverted microscope using a CCD camera coupled to an LCTF to perform hyperspectral imaging. Spectrally resolved images are collected between 500 and 700 nm with 2 nm steps. Broadband illumination is provided by an Xcite-120 light-emitting diode lamp. For live-cell measurements, cells are imaged live and maintained under physiological conditions (5% CO_2 and 37°C) via a stagetop incubator.

To investigate the molecular functionality relevant to chromatin structure, nano-ChIA coregisters STORM and PWS. This allows for visualization of chromatin packing structure with respect to the spatial distribution of functionally important macromolecules, such as actively elongating RNA Pol II(Fig. 1.3C). All STORM images have an average localization precision below 20 nm.

1.4. Computational Modeling of Chromatin and Transcription

1.4.1. A General Overview of Modeling in the Chromatin Space

Computational modeling of biological systems can uncover fundamental mechanisms that can explain experimentally observable phenomena. As discussed by Moller et al., general modeling categories can fall into "top-down" or "bottom-up" approaches [196]. Briefly, "top-down" modeling employs data-driven approaches to integrate data from experimental methods such as Hi-C, ChIP-seq, and SR microscopies to learn about four-dimensional chromatin structure. Both Monte Carlo (MC) and Molecular Dynamics (MD) approaches can be employed. Many data-driven approaches have been developed to obtain three-dimensional information about the chromatin polymer from 2D contact maps generated by Hi-C experiments, including how chromatin conformations in individual cells contribute to structural ensembles at the population level [245, 190, 283, 65, 222]. Computational methodologies developed by Michele di Pierro are also able to integrate Hi-C data with epigenetic information from 1D ChIP-seq tracks and 3D super-resolution microscopy data [65, 205] and are even able to reproduce subdiffusive dynamics and spatially coherent motion observed by optical experiments [64, 308, 173].

"Bottom-up" modeling encompasses polymer models that are built from the first principles of physics to determine fundamental rules underlying chromatin structure and function. The computational models presented in this thesis work can be categorized as "bottom-up" models. Such models can be theory-based or simulation-based.

Theory-based models oftentimes characterize the statistical properties of polymers using scaling laws. For example, they determine how the number of monomers or, equivalently, the polymer mass for homopolymers, scales with the physical space the polymer occupies [63]. The scaling laws of a homopolymer chain, where all monomers interact in the same way, depend on the balance of the free energy of monomer-monomer compared to monomer-solvent interactions. Under dilute, equilibrium conditions, such homopolymers are expected to exhibit mass scaling behavior characterized by a length-scale invariant power-law relationship between the mass (M) and the size of the polymer r: $M \propto r^D$, where D is the mass scaling coefficient, or the packing scaling, of the polymer (Fig. 1.5).

For example, the fractal globule model first proposed by Alexander Grosberg in 1988 has a $D = 3$ [105]. The fractal globule is a collapsed polymer where topological constraints result in a hierarchical organization of non-entangled structures. The first Hi-C studies were compatible with chromatin being organized as a fractal globule [193, 168]. The contact probability scaling is the power-law scaling relationship (s) between contact probability (P) and linear genomic distance (N), i.e., $P \propto N^{-s}$. Experimental analysis observed $s = 1$ that was predicted by the fractal globule model. Additionally, the non-entangled nature of the fractal globule made it an attractive model to explain how transcription and replication processes were able to access a non-dilute chromatin system. However, higher-resolution *in situ* Hi-C demonstrated that contact probability scaling is actually smaller within TADs and larger for longer-range contacts [251], thus disproving the fractal globule model as one that fully describes statistical chromatin structure.

Simulation-based models can be sub-categorized into atomistic, mesoscale, and more course-grained simulations. Atomistic simulations explicitly represent all atoms, and all

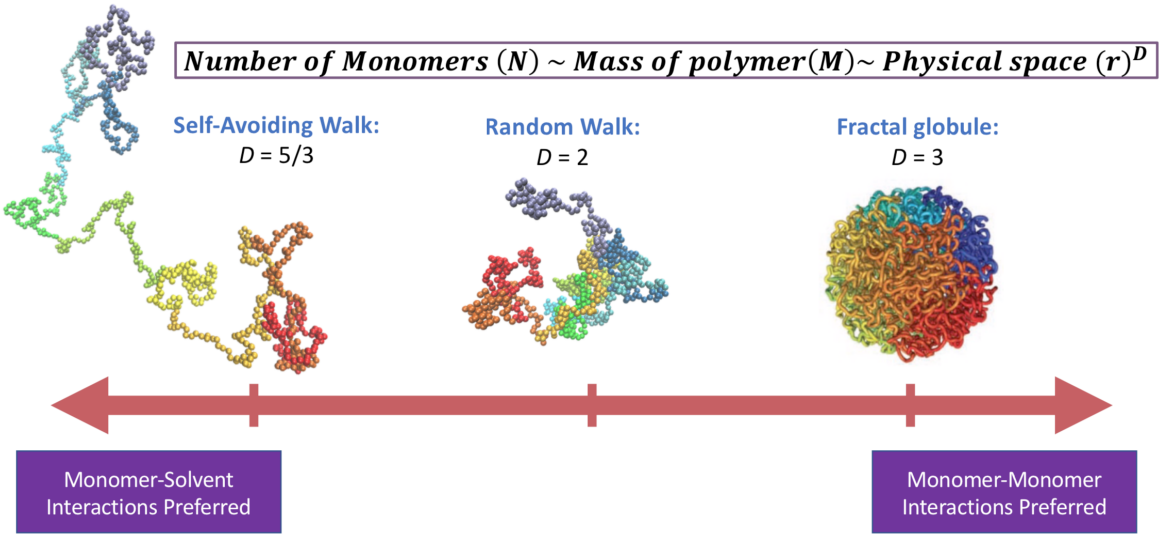

Figure 1.5. Power-law scaling behavior of homopolymers. The mass scaling, or packing scaling, coefficient determines how the mass of the polymer (M) or number of monomers (N) scales with the 3D space the polymer occupies (r). In a good solvent (i.e., monomer-solvent interactions are preferred), $D = 5/3$ and the polymer adopts a swollen self-avoiding walk. When a polymer's self-interaction and interaction with the encompassing solvent are equally preferred, as in the case of a random walk, $D = 2$. When monomer-monomer interactions are heavily preferred, the polymer collapses. A special case of a $D = 3$ collapsed globule is the fractal globule [193].

pairwise interactions between atoms, in the polymer system. These simulations are highly detailed, and are therefore highly computationally expensive. This is especially for larger and denser systems, such as the dynamic chromatin structure in the nucleus. Thus, most atomistic simulations are performed to understand the fundamental organization and dynamics of smaller structures such as the nucleosome [183, 297]. Even with the advent of supercomputing, which has tremendously increased the computational power available for simulations, the largest massively parallelized atomistic simulations have been performed only for the GATA4 gene locus [132]. Many mesoscale models of chromatin structure have

been developed to increase the length- and time-scales accessible to simulation (Fig. 1.2). Mesoscale simulations coarse-grain interactions from atomistic-level simulations and thus reduce the level of detail within the system. For example, the de Pablo group has developed the three-sites-per-nucleotide (3SPN) model of DNA and subsequent "1-cylinder-per-nucleosome" (1CPN) model of chromatin with different levels of coarse-graining for each mesoscale model [88, 159]. Such models were then employed to study, for example, how changes in the length of linker DNA can alter the free energy landscape of chromatin structure [159]. The Schlick group's nucleosome-resolution mesoscale model was recently combined with SR microscopy experiments to determine direct contributions of structural changes, such as nucleosome positioning, and chemical changes, such as histone tail acetylation, to experimentally observed changes in nucleosome clutch organization throughout differentiation [227].

At the most coarse-grained scale, chromatin can be represented using a "beads-on-a-string" approach, where interactions between monomers are treated even more generically. Such models are important for studying biological processes at larger length-scales, up to entire nuclei, and time-scales, including across multiple cell divisions. Macpherson et al. determined how preferential binding of HP1 to H3K9me3 heterochromatin results in phase separation only for stretches of H3K9me3 marks above a certain length threshold [178]. This model was then extended to explain how constitutive heterochromatin methylation is reliably maintained over generations ($\sim$days) [252]. Additionally, the mechanistic loop extrusion model combined with differential interactions between A and B compartments has been able to explain several key experimentally observable phenomena, including the effects of loop extrusion-related genetic perturbations on Hi-C maps [90, 206]. This model

was then employed used to determine the minimal number of interactions necessary to reproduce conventional nuclear organization. For example, heterochromatin compartmentalization and lamin interactions with heterochromatin, but not euchromatin interactions, are necessary for conventional nuclear organization with heterochromatin at the periphery and euchromatin in the nuclear interior [77].

However, all simulation-based models discussed in this section require many different inputs to parameterize interactions. The Self-Returning Random Walk (SRRW) statistical model of chromatin structure was first developed by Kai Huang in the Szleifer group [120]. With just one parameter, the SRRW model was able to faithfully represent and reconcile several distinct aspects of chromatin structure: a high degree of mass-density heterogeneity [3, 212], high contact frequency within chromatin domains (i.e., TADs) [232, 274], and a hierarchical folding structure [281]. This thesis will extend results from the SRRW study to help uncover fundamental folding mechanisms of chromatin that are agnostic to specific biophysical mechanisms (see Chapter 2.3).

1.4.2. A Discussion of the Representation of Electrostatics and Charge in Simulations

DNA is a strong polyelectrolyte due to the chemical behavior of the phosphate in its sugar-phosphate backbone [35]. When DNA wraps around histones to form nucleosomes, the basic histone amino acid residues only partially neutralize the negative charge of DNA-phosphates [95]. Due to the strongly negative charge density of chromatin, the physicochemical environment, including the electrolyte environment, pH, and chromatin density, are expected to influence chromatin structure. Experiments have demonstrated

the effects of the ionic environment on chromatin, ranging from DNA persistence length (i.e., elasticity) [20] to compaction of chromatin domains [309]. Thus, a faithful representation of electrostatics is crucial from a modeling perspective in order to better understand the relationship between the intranuclear physicochemical environment and chromatin.

Here, let us discuss the representation of electrostatics in simulation-based approaches for DNA and chromatin. Although simulations have been integral to our mechanistic understanding of how the physicochemical environment modulates DNA and chromatin, it is important to discuss the deficiencies of how these models represent electrostatic interactions between charged groups in these systems. Atomistic MD simulations provide a more detailed representation of electrostatics by accounting for Coulombic interactions between charged particles at the atomistic scale. Such simulations provide detailed information regarding, for example, ion-DNA localization, but can have varying results due to the inherent difficulty in implementing correct force fields to represent ion-phosphate and ion-ion interactions [151, 44, 226]. Recent advances in experimental methods have facilitated a more complete characterization of the ionic atmosphere surrounding DNA, including Small Angle X-ray Scattering (SAXS) [60] and, more recently, Inductively Coupled Plasma Mass Spectrometry (ICP-MS) [94, 93, 95] and have been used to better parameterize the force-fields of MD simulations [303, 93]. However, even with experimentally-informed parameterization, such approaches may still incorrectly predict trends which are not observed experimentally. For example, modeling of DNA nucleotides by Panteva et al. incorrectly predicted a size dependence of cation occupancy of DNA that was subsequently disproven by ICP-MS measurements [219, 93]. Acid-based equilibrium, the

protonation and deprotonation of the phosphate acid, and ion condensation, the binding of positively charged cations to negatively charged phosphates and consequent charge neutralization, are both dynamic chemical reactions. Although approaches such as λ-integration have been developed to incorporate variable pH in atomistic simulations [69], it remains computationally intractable to concomitantly account for explicit pH and ion condensation reactions. Additionally, atomistic simulations are computationally expensive to converge and simulating larger and denser systems that have more relevance to intranuclear conditions becomes even more computationally intractable.

On the other hand, several well-developed mesoscale models, which effectively coarse-grain atomistic force fields, are able to reproduce experimental observations, such as the decrease in persistence length of dsDNA and chromatin with increasing salt concentration [271, 88, 255, 150, 159]. Yet, the question of how to represent electrostatics becomes more complex with such coarse-grained models. Nonlinear Poisson-Boltzmann is a mean-field approximation that works relatively well if there are no significant short-range electrostatic correlations, especially since Poisson-Boltzmann considers ions as point charges without excluded volume. Debye-Hückel is a linearized approximation of Poisson-Boltzmann and is only valid for relatively weak electrostatic fields. Thus, for denser systems with multivalent ions, Debye-Hückel approximations are not as correct. Beard, et al. have developed an innovative approach which minimizes the error between Poisson Boltzmann and discrete Debye-Hückel charges and can be applied to mesoscale modeling of a chromatin system [21]. However, this approach still contains the pitfalls of Poisson Boltzmann. Finally, ion condensation has been accounted for in coarse-grained simulations by using Manning condensation, which predicts a charge of ≈ -0.6 for a strong polyelectrolyte such as

DNA. However, this theoretical approach assumes cylindrical geometry of the polymer and, again, becomes less accurate for multivalent cations. Most coarse-grained models that represent DNA employ Debye-Hückel, sometimes with Manning condensation to account for charge neutralization of phosphates by counterions [88]. Additionally, many coarse-grained simulations are performed under implicit solvent conditions, and thus do not explicitly account for the excluded volume effects of counterions. All coarse-grained computational methodologies discussed thus far assume both fixed charges and fixed pH throughout the entire simulation. Thus, so far, no simulation-based approach that models DNA and chromatin, especially in denser, non-dilute systems, has properly accounted the effects of charge regulation.

1.4.3. An Introduction to Polymer Brush Theory and Molecular Theory

Previous theoretical approaches that have been employed to study brushes of strong polyelectrolytes may also be relevant to studying DNA and chromatin systems. In general, theoretical treatment of polymer brushes involves expressing the system free energy as a functional of the density of all species involved, including solvent, ions, and monomers. Different theories employ distinct methods to represent polymer conformations and their resulting conformational entropy. Self-Consistent Field (SCF) theories represent polymers as Gaussian chains and Scheutjens-Fleer Self-Consistent Field (SF-SCF) theory can be thought of as a version of SCF with segment distributions discretized into a lattice [122, 123]. Such SCF approaches have explicitly considered charge regulation of polyelectrolyte brushes.

In this thesis, we employ a molecular theoretical approach, henceforth referred to a Molecular Theory (MT), to investigate the effects of the physicochemical environment on DNA-like and chromatin-like systems. MT employs a similar approach to the discretized SF-SCF, but free energy minimization determines the probability distribution of input polymer conformations instead of assuming Gaussian statistics. Thus, unlike SF-SCF, MT: (1) is able to represent more complex polyelectrolyte systems that deviate from Gaussian statistics and (2) explicitly considers chain connectivity.

MT is a statistical thermodynamical theory that accounts for complex molecular-level interactions among the elementary units of the system, including chemical reactions and electrostatics [201, 100]. MT predictions have been found to agree with experimental observations for relevant biological systems, including polyelectrolyte brushes and nuclear pore complexes as well as phenomena such as protein adsorption of polymer brush surfaces [189, 254, 299, 278, 269, 239, 291, 156].

An overview of the MT approach is represented in Fig. 1.6. This theoretical mean-field approach links the structural state of the polymer system with its chemical/charged state and solves for equilibrium properties by explicitly accounting for the charge, volume, and conformations of all constituent molecules. The MT calculations take as input environmental conditions, the size and charge of all molecular species, the free energies of all chemical reactions, and a representative set of chain conformations. MT outputs any thermodynamic or structural quantity of interest, including the system free energy, the polymer density distribution, and the distribution of solvent, ions, and electrostatic potential. Importantly, the MT outputs the probability of all input conformations, which

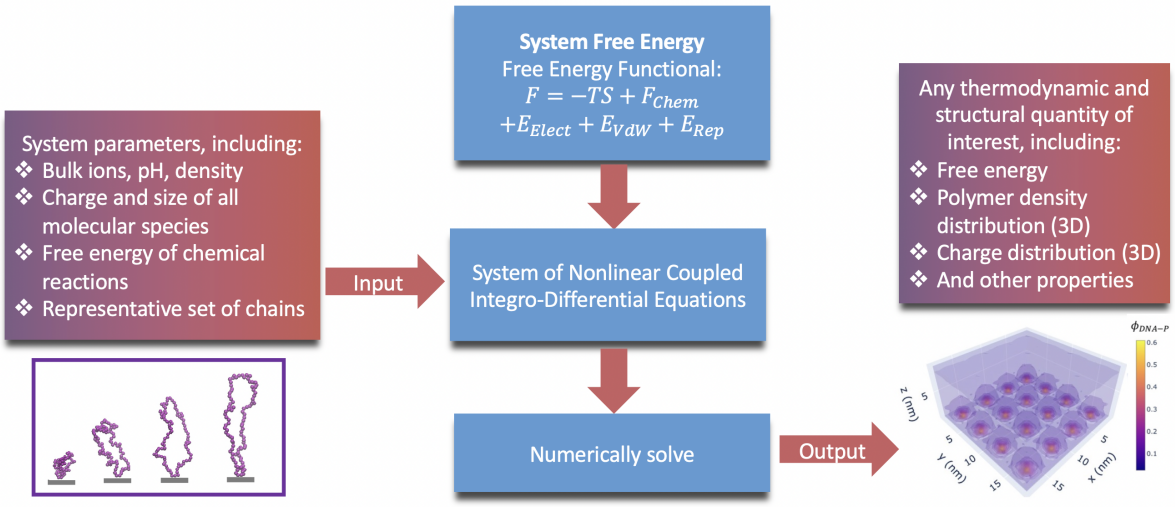

Figure 1.6. Overview of Molecular Theory (MT) approach for DNA-like system, described in greater details in Chapter 4.6.1. Inputs to MT include environment conditions (e.g., bulk ions, pH, DNA density), charge and size of all molecular species, free energy of all chemical reactions, and a representative set of chains. Shown above in red are example conformations of loop chains ranging from more collapsed to more extended. MT relies on a free energy functional which takes into account the system entropy (TS), chemical potential (F_{chem}), electrostatics (E_{elect}), Van der Waals interactions (E_{VdW}), and steric repulsions (E_{rep}). This free energy functional is then decomposed into a system of nonlinear coupled integro-differential equations, which are numerically solved using optimization. Outputs of the theory include any thermodynamic and structural quantity of interest, including the three-dimensional distribution of polymer (e.g., DNA) volume fraction.

directly contributes to the 3D distribution of polymer density, electrostatics, etc. Unlike simulation-based methods, with MT it is computationally feasible to: (1) perform parameter scans over a wide range of environmental conditions, including for non-dilute systems, and (2) have a relatively accurate representation of system electrostatics and charge.

1.4.4. Modeling of Transcription Under Crowded Conditions

Macromolecular crowding was first described by Allen Minton in 1981 [192]. Crowders are macromolecules that exclude volume and thus influence adjacent chemical reactions, but do not directly participate in them. The intranuclear environment is a highly crowded environment, with chromatin volume concentration (CVC) ranging from 12-52% in interphase nuclei [212], along with currently unknown concentrations of mobile crowders, such as transcriptional machinery. As 98% of chromatin is non-coding and is only transcribed at negligible levels, chromatin density can be considered as the major intranuclear crowder (Fig. 1.7A). Hiroki Matsuda in the Szleifer group developed a model of gene expression in which transcription is considered as a network of crowding-dependent chemical reactions [184]. The model employs MC simulations to determine the free energy of binding and Brownian Dynamics (BD) simulations to determine diffusion rates of transcriptional reactants, which include RNA Pol II and TFs [184]. Matsuda et al. demonstrate that gene expression is a non-monotonic function of crowding [184]. As crowding increases from zero, attractive depletion interactions increase the binding free energy of transcriptional reactants, thus increasing transcription rates (Fig. 1.7B). However, after a certain critical crowding concentration, the reduction in diffusion of reactants decreases gene expression (Fig. 1.7B). The exact relationship between crowding and transcription is dependent on molecular factors of gene expression, including the concentrations of transcriptional reactants (Fig. 1.7B). Interestingly, the "optimal" crowding conditions occur within physiological average chromatin density, as determined by ChromEM experiments [212, 167].

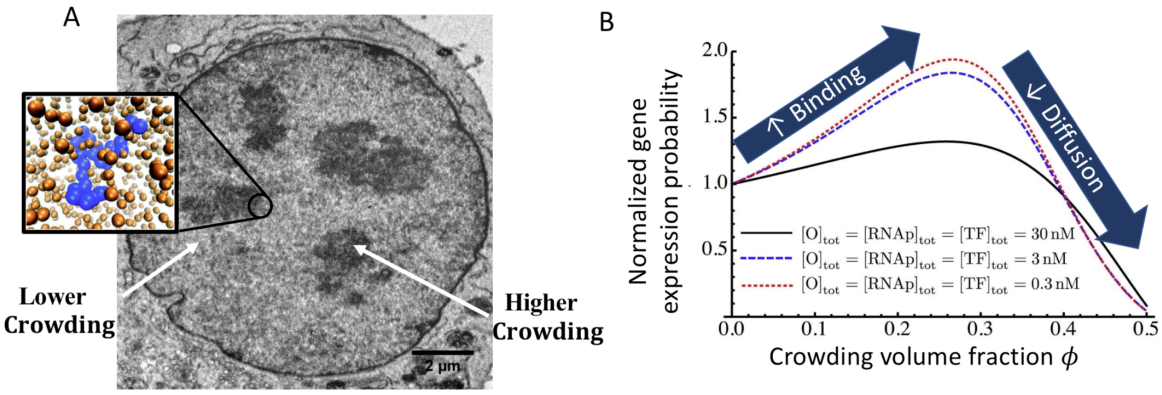

Figure 1.7. Macromolecular crowding influences gene transcription. (A) The nucleus is a highly crowded and heterogeneous environment with chromatin density as the major crowder. ChromTEM image from [288] shows areas of low crowding/chromatin density and high crowding/chromatin density in the same nucleus. (Inset) A hypothetical gene in blue is surrounded by chromatin density and mobile crowders (brown spheres). (B) Non-monotic effect of crowding volume fraction on gene expression depends on molecular regulators of transcription including concentrations of promoters (O), RNA Pol II, and TFs [184].

1.5. Scope of Dissertation

Overall, this thesis will employ a quantitative and biophysical approach to address how molecular-level interactions in a realistic nuclear environment contribute to the regulation of chromatin structure and function. More specifically, this dissertation will focus on answering three key questions in the field of chromatin research.

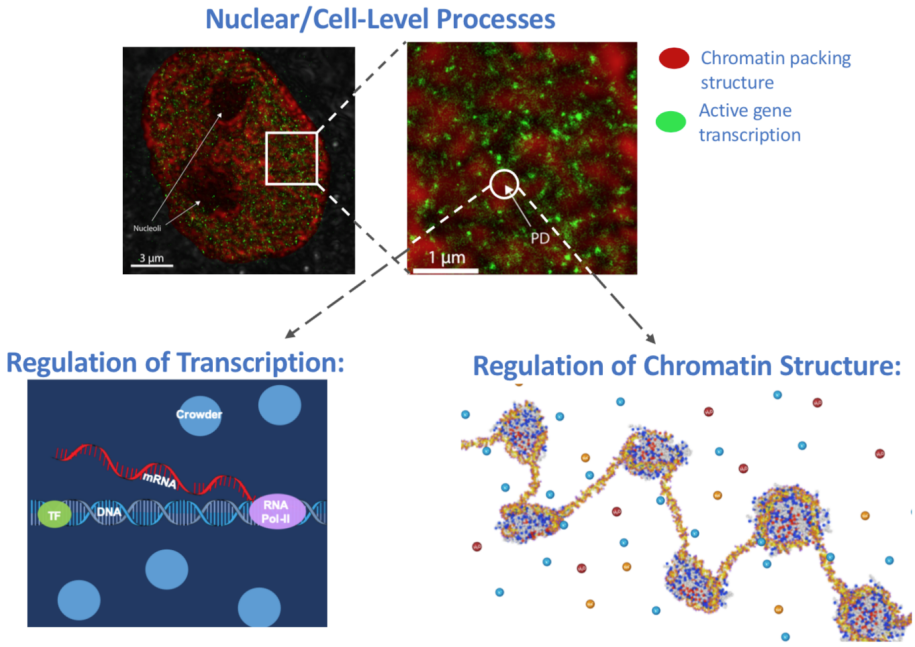

Figure 1.8. Characterizing the chromatin structure and function in a realistic nuclear environment. Nanoimaging techniques and polymer physics-based analysis identify packing domains (PDs) as key functional units of chromatin organization, as their morphological properties have direct consequences for transcription (Chapter 2). Overall, molecular-level, physics-based interactions are demonstrated to influence nuclear- and cell-level processes. Specifically, macromolecular crowding modulates transcription and phenotypic plasticity through statistical chromatin packing behavior (Chapter 3) and the physicochemical intranuclear environment influences DNA and chromatin structure (Chapter 4).

Firstly, what are the fundamental units of higher-order chromatin structure at the single-cell level? Structural units, such as TADs, have previously been identified by a combination of Hi-C and SR microscopy methods. However, their exact functional importance and how the chromatin conformation in individual cells contributes to population-level behavior is still under debate, as discussed in Chapter 1.1.4. Employing a combination of statistical modeling, chromosome conformation capture, and optical and electron microscopies introduced in Chapter 1.3, Chapter 2 will characterize key principles of supranucleosomal chromatin organization through the lens of polymer physics.

Next, we address how exactly the supranucleosomal chromatin structure modulates *genome-wide* transcription processes. Chapter 1.2 identifies specific aberrations in chromatin structure that are observed in the complex diseases such as cancer. However, due to the inherent complexities of nuclear processes, there does not yet exist a universal, length-scale-independent relationship between chromatin structure and the disease state. Chapter 3 will define a framework for how statistical chromatin packing of packing domains contributes to large-scale gene expression patterns through macromolecular crowding-mediated effects. Our modeling can even predict cellular adaptability to external stressors, encompassing the chemoevasion potential of cancer cells.

Finally, as chromatin is a highly charged polyelectrolyte, Chapter 4 will address the question of exactly how all aspects of the physicochemical environment, including bulk ions, pH, and chromatin density influence chromatin structure. The intranuclear environment is highly crowded, and single-molecule experiments and simulations performed under dilute conditions to interrogate these effects are not representative of a realistic intranuclear environment. Additionally, it is difficult to experimentally determine the

specific effects of the physicochemical environment on chromatin *in vitro*, as altering the pH and electrolyte environment influence enzymatic reactions, signalling pathways, and other important cellular processes. It is also computationally difficult to properly model electrostatic interactions in such dense systems, as discussed in Chapter 1.4.2. Thus, Chapter 4 will introduce a molecular theoretical modeling approach to characterize the influence of the physicochemical environment on both the charge and structure of DNA-like and chromatin-like systems in physiologically dense environments.

1.6. List of Key Contributors by Chapter

The chapters of this dissertation are structured around the key questions that are being addressed. Each chapter is composed of multiple independent journal articles (as detailed in Chapter 1.7) and the main contributors to each are listed below.

Chapter 2 can be broken down into two sections. The first section of Chapter 2 contains work published in [167, 166]. For the experimental work, Yue Li and Vasundhara Agrawal mainly performed and analyzed ChromSTEM and ChromTEM experiments, with help from Wing Shun Li, and Adam Eshein and David VanDerway performed and analyzed PWS experiments. All ChromSTEM work was performed in collaboration with Professor Vinayak Dravid. The second section introduces and validates the Self-Returning Random Walk (SRRW) as published in [120] and the Self Returning-Excluded Volume (SR-EV) model, which is currently in preparation. Kai Huang developed the SRRW model and Marcelo Carignano developed the SR-EV model of chromatin, with help from Rikkert Nap and Anne Shim. Vasundhara Agrawal performed Hi-C and PWS heat shock experiments.

Chapter 3 is composed of one published manuscript [288], with Wenli Wu and Luay Almassalha as co-first authors, and one manuscript currently in preparation with Jane Frederick as co-first author. Wenli Wu and Luay Almassalha first conceptualized the Chromatin Packing Macromolecular Crowding (CPMC) model along with Vadim Backman and Igal Szleifer. The initial iterations of this model were published in [6, 5]. Luay Almassalha, Greta Bauer, Jane Frederick, and David VanDerway performed experiments to validate and understand the CPMC and Chromatin-Dependent Adaptability (CDA) models. Adam Eshein and Jane Frederick performed STORM and PWS experiments to

determine the relationship between packing scaling and gene expression as published in [167].

Chapter 4 is composed of two separate manuscripts currently in preparation. Rikkert Nap helped with theory and software development as well as analysis for the DNA-like and chromatin-like systems. Marcelo Carignano helped with biased MD simulations of DNA-like loops. The chromatin work was performed in collaboration with Juan de Pablo's group with help from Aria Coraor, who generated the 8-mer nucleosome chains via 1CPN simulations which were used as input for the theory.

Vadim Backman and Igal Szleifer are senior authors on all manuscript and helped with conceptualization of models and experimental analyses.

1.7. Publications

1.7.1. Peer-Reviewed Articles

(1) Y. Li*, V. Agrawal*, **R. K. A. Virk***, E. Roth, W. Li, A. Eshein, J. Frederick, K. Huang, L. Almassalha, R. Bleher, M. A. Carignano, I. Szleifer, V. P. Dravid, and V. Backman. "Analysis of three-dimensional chromatin packing domains by chromatin scanning transmission electron microscopy (ChromSTEM)", *Scientific Reports*, 12(1):12198, 2022.

(2) S. Yadav*, **R. K. A. Virk**, C. H. Chung, M. B. Eduardo, D. VanDerway, D. Chen, K. Burdett, H. Gao, Z. Zeng, M. Ranjan, G. Cottone, X. Xuei, S. Chandrasekaran, V. Backman, R. Chatterton, S. A. Khan, and S. E. Clare. "Exposure of mammary cells to lipid activates gene expression changes associated with ER-negative breast cancer via chromatin remodeling", *npj Breast Cancer*, 8(59), 2022.

(3) Y. Li*, A. Eshein*, **R. K. A. Virk***, A. Eid, W. Wu, J. Frederick, D. VanDerway, S. Gladstein, K. Huang, A. R. Shim, N. M. Anthony, G. M. Bauer, X. Zhou, V. Agrawal, E. M. Pujadas, S. Jain, G. Esteve, J. E. Chandler, T. Nguyen, R. Bleher, J. J. de Pablo, I. Szleifer, V. P. Dravid, L. M. Almassalha, and V. Backman. "Nanoscale Chromatin Imaging and Analysis (nano-ChIA) platform bridges 4-D chromatin organization with molecular function", *Science Advances*, 7(1):eabe4310, 2021.

(4) A. Eid*, A. Eshein, Y. Li, **R. Virk***, D. VanDerway, D. Zhang, A. Taflove and V. Backman. "Characterizing chromatin packing scaling in whole nuclei using interferometric microscopy", *Optics Letters* 45(17): 4810-4813, 2020.

(5) K. Huang*, Y. Li, A. R. Shim, **R. K. A. Virk***, V. Agrawal, A. Eshein, R. J. Nap, L. M. Almassalha, V. Backman, and I. Szleifer. "Physical and data structure of 3D genome", *Science Advances*, 6(2): eaay4055.

(6) **R. K. A. Virk***, W. Wu*, L. M. Almassalha*, G. M. Bauer, Y. Li, D. VanDerway, J. Frederick, D. Zhang, A. Eshein, H. K. Roy, I. Szleifer and V. Backman. "Disordered chromatin packing regulates phenotypic plasticity", *Science Advances* 6(2): eaax6232, 2020.

* These authors contributed equally to this work

1.7.2. Manuscripts Under Review

(1) X. Wang*, V. Agrawal*, Y. Li, P. A. Patel, **R. K. A. Virk**, J. Frederick, V. Backman, and G. Ameer. "Biophysical Modulation of Chromatin via Nuclear Deformation Promotes Osteogenic Differentiation of Mesenchymal Stem Cells", *Nature Biomedical Engineering* (*Under Review*).

* These authors contributed equally to this work

1.7.3. Manuscripts In Preparation

(1) J. Frederick*, **R. K. A. Virk***, L. M. Almassalha*, G. M. Bauer, L. Cherkezyan, D. VanDerway, A. Starwarz, S. Gladstein, J E. Chandler, V. Agrawal, A. Kendra, B. L. Seagle, S. Morochnik, A. Ugolkov, D. D. Billadeau, G. Ameer, T. V.

O'Halloran, A. P. Mazar, I. Szleifer, S. Shahabi, and V. Backman. "Leveraging chromatin scaling to target chemoevasion in vivo".

(2) **R. K. A. Virk***, R. Nap*, M. A. Carignano, A. Coraor, P. C. Gonzalez, J. de Pablo, V. Backman, and I. Szleifer. "The importance of charge regulation for DNA-like loops in a physiological intranuclear environment".

(3) **R. K. A. Virk***, R. Nap*, A. Coraor, J. J. de Pablo, V. Backman, and I. Szleifer. "Influence of the physicochemical intranuclear environment on charge regulation of chromatin".

(4) M. A. Carignano*, **R. K. A. Virk**, V. Agrawal, J. J. de Pablo, V. Backman, and I. Szleifer. "Self returning-excluded volume polymer captures statistical rearrangements in chromatin domains upon heat shock".

* These authors contributed equally to this work

1.8. Conference Presentations

1.8.1. Invited Talks

(1) **R. K. A. Virk**. "Leveraging Chromatin Packing to Modulate Phenotypic Plasticity and Chemoevasion Potential of Cancer Cells", Northwestern's Lurie Cancer Center Symposium; Chicago, IL; 6/2022.

(2) **R. K. A. Virk**. "Leveraging Chromatin Packing to Modulate Phenotypic Plasticity and Chemoevasion Potential of Cancer Cells", Chromatin Structure and Function Gordon Research Seminar; Castelldefels, Spain; 5/2022.

(3) **R. K. A. Virk**. "Charge Regulation in the Nucleus: Effects of the Physicochemical Environment on DNA and Chromatin." Telluride Workshop on Physical Genomics and Transcriptional Engineering; Telluride, CO; 2/2022.

(4) **R. K. A. Virk**. "Nanoscale Chromatin Imaging and Analysis (nano-ChIA) Platform Bridges 4D Chromatin Organization with Molecular Function." 3rd Annual Simpson Querrey Institute for Epigenetics Symposium "Epigenetics, Metabolism, and Human Disease"; Chicago, IL; 10/2021.

(5) **R. K. A. Virk**. "The importance of charge regulation for DNA-like loops in a physiological intranuclear environment." Biophysical Society Multiscale Genome Organization Webinar; Virtual; 8/2021.

(6) **R. K. A. Virk**. "Nanoscale Chromatin Imaging and Analysis (nano-ChIA) Platform Bridges 4D Chromatin Organization with Molecular Function." NYU Chromatin Club; Virtual; 4/2021.

(7) V. Backman and **R. K. A. Virk**. "Understanding the 'Operating System' of the Genome: The Structure and Function of Chromatin Packing." Broad Institute; Virtual; 6/2020.

(8) **R. K. A. Virk**. "Physical Organization of Chromatin Regulates Phenotypic Plasticity", Physical Sciences in Oncology Annual Investigators Meeting; Minneapolis, MN; 9/2019.

1.8.2. Poster Presentations

(1) **R. K. A. Virk**, R. J. Nap, M. A. Carignano, A. Coraor, J. de Pablo, V. Backman, and I. Szleifer "Uncovering the Impact of Charge Regulation on DNA and Chromatin", Chromatin Structure and Function Gordon Research Conference; Castelldefels, Spain; 6/2022.

(2) **R. K. A. Virk**, R. Nap, V. Backman, and I. Szleifer. "The importance of charge regulation for DNA-like loops in a physiological intranuclear environment", NSF Emerging Frontiers in Research and Innovation Teams Workshop; Virtual; 8/2021.

(3) **R. K. A. Virk**, V. Agrawal, A. Coraor, J. de Pablo, and V. Backman. "Heat shock as a model to study the relationship between statistical chromatin packing and transcriptional response", NSF Emerging Frontiers in Research and Innovation Teams Workshop; Virtual; 8/2021.

(4) **R. K. A. Virk**, W. Wu, L. M. Almassalha, G. M. Bauer, Y. Li, D. VanDerway, J. Frederick, D. Zhang, A. Eshein, I. Szleifer, and V. Backman. "Disordered

Chromatin Packing Regulates Ensemble Gene Expression and Phenotypic Plasticity", Biophysical Society Annual Meeting; San Diego, CA; 2/2020.

(5) **R. K. A. Virk**, W. Wu, L. M. Almassalha, G. M. Bauer, Y. Li, D. VanDerway, J. Frederick, A. Eshein, H. K. Roy, I. Szleifer, and V. Backman. "Physical Organization of Chromatin Regulates Phenotypic Plasticity", Physical Sciences in Oncology Annual Investigators Meeting; Minneapolis, MN; 9/2019.

(6) **R. K. A. Virk**, G. M. Bauer, L. M. Almassalha, W. Wu, S. Gladstein, I. Szleifer, and V. Backman. "A Macrogenomic Engineering Approach to Understanding and Manipulating the Chromatin Packing Code." Chromatin Structure and Function Gordon Research Conference; Newry, ME; 7/2018.

CHAPTER 2

Higher-Order Organization of Chromatin into Hierarchically Folded Packing Domains

2.1. Introduction

Dynamic, three-dimensional chromatin organization plays an important role in regulating a vast number of cellular processes, including cell type–specific gene expression and lineage commitment [225, 50, 22]. Large-scale alterations in chromatin structure are associated with cancer, numerous neurological and autoimmune disorders, and other complex diseases [277, 218]. However, the precise conformation of chromatin remains contested. The basic units of chromatin is the nucleosome, formed by 147 bp of DNA wrapping around histone proteins. Nucleosomes are connected by linker DNA to form a "beads-on-a-string" chromatin fiber. Previously, the primary 10 nm fiber was thought to aggregate into a thicker 30 nm chromatin fiber, but this textbook view has been challenged by several recent studies [131, 75]. One such work employed a novel imaging technique, chromatin electron microscopy tomography (ChromEMT), to interrogate chromatin ultrastructure down to the level of single nucleosomes [212]. Using ChromEMT, Ou et al. discovered that DNA and nucleosomes assemble into disordered chains, with diameters varying between 5 and 24 nm, which themselves pack at various densities within the nucleus [212].

Parallel to microscopy-based techniques such as ChromEMT, chromosome conformation capture–based (3C-based) methods have provided key insights into higher-order chromatin structures by linking chromatin topology with genomic information [257]. Specifically, Hi-C measures pairwise chromatin-chromatin contact frequencies throughout the entire genome to capture average chromatin connectivity over millions of cells [168]. 3C and related methods (4C, 5C, Hi-C) have revealed that the eukaryotic genome is partitioned into topologically associating domains (TADs) at the scale of several 100's of kilobases (kbp) and smaller loop domains, or sub-TADs [68, 262, 232, 281]. Notably,

single-cell Hi-C methods have demonstrated the potential existence of TADs in individual nuclei, although a high degree of intercellular heterogeneity in TAD distribution has been reported [198]. The potential functional importance of TADs is also implied by their heritability. Higher-order cell-type–specific structures, such as TADs, are lost during mitosis and reestablished along with a lineage-specific replication timing program in the early G1 phase of the cell cycle [66].

The recent development of SR microscopies, including STORM and PALM, in combination with FISH-based labeling techniques, has allowed for investigation of higher order chromatin structures down to a resolution of ∼10s of nm, almost at the level of the primary chromatin fiber. 'Chromomeres', punctate chromatin particles around 200-300 nm in diameter, have been observed in both interphase chromatin and mitotic chromosomes using stimulated emission depletion (STED) microscopy [101]. A recent study employing live-cell photo-activated localization microscopy (PALM) in mammalian cells determined that nucleosomes are arranged into physically compact chromatin domains with a 160 nm diameter [207]. The dynamics of these chromatin domains were correlated with those of replication domains, which range in diameter between 110-150 nm [207, 57, 300]. Also in mammalian cells, 3D-structured illumination microscopy (SIM) imaging discovered chromatin domain clusters (CDCs) of around 120 to 150 nm in diameter in which the chromatin compaction increases radially from the periphery towards the CDC core [53]. Recently, high-resolution imaging experiments have visualized the three-dimensional organization of heterochromatic TADs identified by Hi-C contact maps as compacted domains in single cells, providing a link between the nanoscopic spatial structures and genomic domains [275, 32]. Altogether these higher-order chromatin structures potentially play an

important role in DNA-based processes, such as transcription, replication, and repair, and perhaps extend to complex processes, such as aging and diseases such as cancer [84, 163, 298, 195, 175].

Mechanistic "bottom-up" polymer models (see 1.4.1), including the loop extrusion [251, 90] and the Strings&Binders models [18, 31, 47], have been proposed to better understand physical mechanisms underlying chromatin organization. The loop extrusion model assumes dynamic loop formation is generated by the active process of CTCF-cohesin-mediated loop extrusion. However, although TAD structures are significantly altered after depletion of CTCF, which supposedly provides genomic specificity for TADs, interactions persist at the A/B compartment level and within former TADs [206]. Multiplexed-FISH microscopy also demonstrates that TAD boundaries can occur with nonzero probability at any genomic locus and, after cohesin depletion, domains are still observed at the single-cell level [32]. Chromatin structure also involves the interplay between multiple folding mechanisms such as supercoiling [152, 202, 51], phase separation [155, 253, 250, 96], molecular binding [30], crowding effects [141], as well as loop extrusion [251, 90], all under the feedback control of transcription to be responsive to external stimuli. Altogether, this demonstrates the intricacy of biological phenomena at play which are difficult to fully represent using mechanistic modeling. An ideal minimal model of chromatin structure would recapitulate the major experimental observations with a small number of adjustable parameters and computational complexity, while establishing the existence of fundamental principles of genome organization. Such a model would not have the burden of accounting for all the physical interactions and biological mechanisms that are far from fully understood.

Despite advances in -omics, SR microscopies, and modeling techniques there remain several key open questions in the chromatin field, such as:

(1) What are the functional units of chromatin?

(2) What is the hierarchy of chromatin folding hidden in the disordered morphology?

(3) What is the connection between chromatin conformation, gene loci connectivity, and transcription?

(4) What mechanisms contribute to the formation and maintenance of higher-order chromatin structure in live cells?

(5) What are the inner workings of higher-order chromatin domains at the single-cell level?

Among these questions, perhaps the most fundamental one is whether there are abstract, yet universal, folding principles of our genomic code independent of the known molecular and mechanistic complexity.

This chapter begins with the discovery of chromatin packing domains (PDs). We utilized ChromSTEM (see Chapter 1.3) for an in-depth quantification of 3D chromatin conformation down to 6 nm resolution, allowing for characterization of higher-order chromatin structure almost at the level of the DNA base pair. Employing mass scaling analysis on ChromSTEM mass density tomograms, we observed that chromatin forms spatially well-defined higher-order domains around 80 nm in radius. Within domains, chromatin exhibits a polymeric fractal-like behavior and a radially decreasing mass-density from the center to the periphery. Unlike other nanoimaging and analysis techniques, we demonstrate that our unique combination of a high-resolution imaging technique with polymer physics-based analysis enables us to (i) investigate the chromatin conformation within

PDs and (ii) quantify statistical descriptors of chromatin structure that are relevant to transcription. We observe that PDs have heterogeneous morphological properties, even within the same cell line, underlying the potential role of statistical chromatin packing in regulating gene expression within mammalian nuclei. We also observed PDs in live-cells using our label-free PWS modality (see Chapter 1.3) and determined that the packing behavior within these domains is both heritable across cell division and related to statistical genome connectivity behavior, as determined by Hi-C.

Next, we show that the Self-Returning Random Walk (SRRW) statistical model and the Self Returning-Excluded Volume (SR-EV) polymer model are able to reconcile distinct properties of chromatin structure. This necessitates modularizing the three-dimensional genome into tree-like data structures where the functional modules are connected by an isolated by open backbone. Using live-cell PWS microscopy and Hi-C, we then demonstrate that the SR-EV polymer model is able to represent statistical rearrangements in chromatin structure that occur upon heat stress. Altogether, our results also suggest that α, the single parameter of the SR-EV model, is related to the temperature history of the system due to rearrangements in connectivity and mass density distribution from gradations of the heat shock response. Our multiscale theoretical and experimental results suggest the existence of higher-order universal folding principles for the disordered chromatin fiber to avoid entanglement and fulfill its biological functions.

2.2. Chromatin Forms Packing Domains (PDs) with Heterogeneous Packing Properties

2.2.1. ChromSTEM Uncovers Power-Law PD Mass Scaling Regime in A549 and BJ Nuclei

The mass scaling behavior of homopolymers under dilute and equilibrium conditions conditions is well-defined for simple polymer models (Fig. 1.5). Mass scaling behavior is characterized by a power-law relationship at certain length-scales between the mass (M) of the polymer and the 3D size r the polymer occupies: $M \propto r^D$, where D is the mass scaling coefficient, or the packing scaling.

Chromatin itself exists as a heteropolymer. Its monomers, i.e., nucleosomes, possess varying biochemical properties in the form of chemical modifications - including epigenetic marks - and physical constraints - including CTCF-cohesin- or transcription-dependent looping, interactions with nuclear lamins, and phase separation driven by chromatin-associated proteins such as HP1 [208, 77, 253, 178]. Therefore, at any given point in time, chromatin conformation is determined by different, and potentially competing forces, altogether resulting in a dynamic, non-equilibrium system. Additionally, chromatin occupies a significant volume fraction within the nucleus. As a result, the intranuclear chromatin environment is both non-equilbrium and non-dilute, and thus the rules of polymer physics do not guarantee that the genome-wide structure of chromatin can be described using the same power-law packing scaling relationship.

Thus, in the nucleus, there may be separate regimes or length-scales in which chromatin exhibits different mass scaling behavior. For example, (1) the primary 10 nm

chromatin fiber may exhibit a unique intra-fiber scaling compared to larger length-scales, (2) higher-order chromatin domains could exist where power-law packing scaling behavior is observed within certain regimes, and (3) for length scales above the size of the individual domains there could be additional structured organization of domains or a random distribution of spatially uncorrelated domains.

To elucidate the chromatin structure within the cell nucleus, we investigated the mass scaling behavior of the continuous signal of DNA contrast obtained from ChromSTEM imaging (Fig. 2.1A-C). Image acquisition was performed as follows. To first locate the cell of interest, we collected an image of the nucleus at lower magnification (∼2kX). Next, we collected tilt series ChromSTEM images in the HAADF mode by selecting a random intranuclear region of interest at higher magnification (∼90 kX) that is located away from the nucleoli and the nuclear envelope (Fig. 1.6D&E, Fig. 2.1A-C). The resultant 3D tomogram has a high contrast signal emerging specifically from chromatin. In the analysis, we consider the disordered chromatin fiber with heterogeneous diameter to be the building block of higher-order structures [212]. Practically, the 3D mass scaling relationship is defined as how the total amount of chromatin (M) enclosed within a volume $V = 4/3\pi r^3$ changes with its radius r. The 2D case can be described as a cross-sectional slice of the 3D system. In this case, M is the amount of chromatin enclosed within an area $A = \pi r^2$. The derivative of the area results in the perimeter, which represents the 1D case. Therefore, in the 1D scenario, M is the amount of chromatin positioned on the circumference of a circle $P = 2\pi r$, which we refer to as "ring mass scaling". We calculated the ring, 2D, and 3D mass scaling by performing linear regression analysis on the log-log mass scaling curves for the given dimensions. The law of additivity of fractal codimensions approximates

the conversion of chromatin packing scaling between different dimensions [282], and we confirmed from our calculations that the 3D mass scaling exponent can be estimated using the 2D and ring mass scaling (Fig. 2.7).

As ChromSTEM only provides a snapshot of the chromatin conformation at a single time point, we randomly sampled different regions within the field of view and calculated the mean mass scaling to capture the statistical behavior. We performed mass scaling analysis on tomograms from both A549 and BJ cells. For four A549 cells with a total volume of 1.16 μm^3 resolved at a voxel size of 2.0 to 2.9 nm, we obtained the mass scaling curves for all three dimensions (Fig. 2.1D&E). A total volume of 0.09 μm^3 was reconstructed from three BJ cells at a nominal voxel size of 1.8 to 2 nm and both the 1D and 2D mass scaling analysis were performed (Fig. 2.7C). To identify length scales where a single packing scaling exponent can sufficiently describe the mass scaling behavior, and to determine average packing scaling within this regime, we evaluated the derivative of the log-log scale of the 3D and 2D mass scaling curves as a function of r. The slope, D_{log} was defined as the linear regression fit to the log-log mass scaling curves. In this linear regression fit, D_{log} should be equivalent to the packing scaling, D, within the power-law scaling regime. Power-law scaling occurs when the length scales associated with D_{log} extend over at least one order of magnitude. From our 3D mass scaling analysis on A549 cells, we observed a power-law mass scaling regime extending from 2 nm to 60 nm with a fitting parameter of $D_{log} = 2.82 \pm 0.01$ (Fig. 2.1D&E, blue dashed line). We refer to the region where power-law mass scaling occurs with one chromatin packing scaling exponent as the PD regime. From ∼60 to 90 nm, a gradual increase in D_{log} to about 2.92 ± 0.02 was observed, which we refer to as the supradomain regime. However, because the maximum

section thickness of A549 tomograms was 180 nm, our 3D analysis was unable to reliably evaluate mass scaling behavior above 90 nm. Additionally, we did not perform the 3D mass scaling analysis for BJ cells, as the thickness of the reconstructed section of BJ cells was smaller than 70 nm, and thus the 3D mass scaling curve would only extend up to 35 nm.

Due to the intrinsic length-scale limitations of 3D mass scaling determined by sample thickness, we next performed the mass scaling analysis at different dimensions for both A549 and BJ cells. Employing the law of additivity of fractal codimensions [282], we calculated the 3D mass scaling exponent from 2D and 1D mass scaling curves as: $D_{3D} = D_{2D} + 1$ and $D_{3D} = D_{1D} + 2$ (Fig. 2.7B). For both A549 and BJ cells, we first evaluated the slope of the 2D mass scaling curve in the log-log scale along its entire length using a 12 nm sliding window. By estimating the local slope for small ranges of r along the entire length of the 2D mass scaling curves, two distinct regimes were identified. The first regime extended up to r ~55 nm, followed by a gradual increase in the local log-log derivative towards a value of $D_{log} \sim 3$. Similar to the 3D mass scaling analysis, for A549 cells (Fig. 2.1E), we then obtained the slope of linear regression, $D_{log} = 2.74 \pm 0.01$ for 2 nm $< r <$ 55 nm (blue dashed line). Above these length scales (r ~60 nm), the slope continuously increases until it approaches 3 for $r >$ 145 nm (red dashed line) up to 200 nm. Similarly, for BJ cells (Fig. 2.7B), the fitting parameter for the linear regression was estimated to be $D_{log} = 2.78 \pm 0.01$ (blue dashed line) for 2 nm $< r <$ 50 nm, and D_{log} approaches 3 (red dashed line) for $r >$100 nm. The shift from the PD regime with similar packing scaling behavior (2 nm $< r <$ 50-60 nm) to the supra-domain regime where $D_{log} \sim 3$ ($r >$

100-145 nm) is continuous, as opposed to a sharp, biphasic transition. The implications of this result will be discussed later.

Therefore, both the average 3D and 2D mass scaling analyses suggest that, for length scales up to 60 nm, chromatin packs into domains which statistically exhibit internal mass scaling behavior and can be described by one average packing scaling exponent. As this behavior was observed in two genetically distinct cell lines, it seems to be a relatively conserved phenomenon in mammalian cells.

Besides the two regimes determined by 2D and 3D mass scaling, the ring mass scaling curve exhibits a third regime from 2 nm $< r <$ 10 nm for both cell lines (Fig. 2.7C-D). This can be interpreted as the chromatin fiber regime. The upper length scale of 10 nm agrees with the upper limit of the primary chromatin fiber size (24 nm maximum diameter) [212]. However, this regime was not identified on the mass scaling curves of higher dimensions, possibly due to limited tomography resolution.

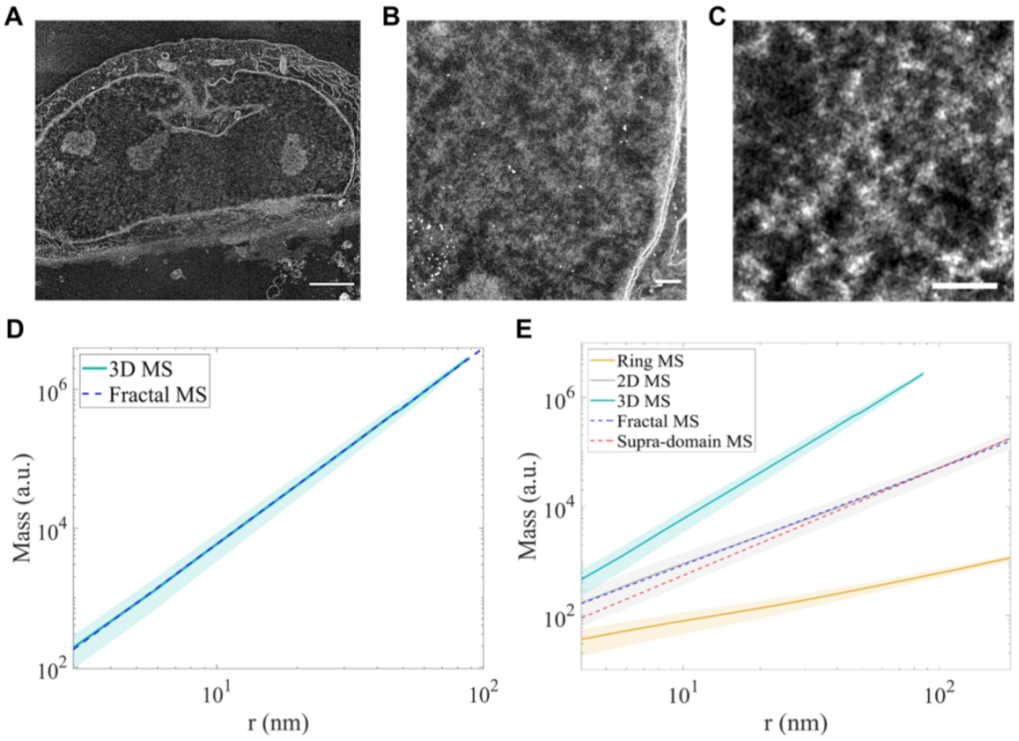

Figure 2.1. Mass scaling analysis uncovers PD regime with power-law mass scaling behavior. (A) STEM HAADF image of a 150 nm section of a BJ cell nucleus for tomography reconstruction. Scale bar: 2 μm. (B) A magnified view of the chromatin and the nuclear periphery of the same cell in (A) with gold fiducial markers. The intensity variation of the image shows that the chromatin packs at different densities throughout. Scale bar: 200 nm. (C) A virtual 2D slice of the chromatin of a BJ cell after tomography reconstruction. Scale bar: 100 nm. The mass scaling analysis was performed on the gray scale tomograms (D-E) The average mass scaling curves from analysis of four A549 cells in (D) 3D and (E) Ring, 2D, and 3D. The mass scaling was conducted for the entire grayscale tomogram and the average mass scaling curve for all the centers weighted by the chromatin density values was computed. 3D mass scaling curve exhibits power-law behavior with a single scaling coefficient up to $r = 60$ nm. Two regimes of mass scaling with different packing scaling D can be identified. In the 2D cases for both A549 cells and BJ cells, the mass scaling curve starts with a packing scaling with $D_{log} < 3$ (blue dashed line) and smoothly transitions to values close to $D_{log} = 3$ (red dashed line).

2.2.2. Quantifying PD Size and Chromatin Packing Behavior at the Domain Boundary of Individual Domains

Our previous analysis averaged the mass scaling behavior from all PDs analyzed within a given field of view. Next, we wanted to better characterize the mass scaling behavior of individual PDs. Here, we outline the criteria to define PD boundaries, which involves analyzing both mass scaling behavior and radial chromatin density. To begin this more detailed analysis, we first identified the "domain center region" of each PD using grayscale ChromSTEM z-stacks with local chromatin density information (Fig. 2.2A-C, Fig. 2.8). From the spatial distribution of 3D chromatin density distribution (Fig. 2.2A), we applied Gaussian filtering and local contrast enhancement before segmentation to create a map of local maxima from the enhanced chromatin contrast tomogram (green areas in Fig. 2.2B). The centers of these local chromatin intensity maps were then identified (Fig. 2.2C). For each PD, we resampled the mass scaling curves with centers inside the "domain center region" (Fig. 2.2D) and determined mass scaling behavior from these "domain centers" up to $r = 400$ nm for A549 cells and $r = 200$ nm for BJ cells due to differences in section thickness between samples. For each individual domain, the average mass scaling curve exhibits a power-law scaling regime with similar chromatin packing scaling and, at larger length-scales, a gradual deviation from the initial power-law behavior (Fig. 2.2E). We performed linear regression on the 2D mass scaling curve and obtained a slope, $D_{log} = 2.56 \pm 0.02$ for r up to 100 nm (Fig. 2.2E, blue dashed line). This power-law scaling relationship can model the mass scaling curve with less than 5% error within the given fitting range, while a more significant divergence is observed beyond r =110 nm (Fig. 2.2E, red asterisk). Therefore, the mass scaling behavior for a single PD demonstrates

that the smaller length scales have a packing scaling $D < 3$ and that, as r increases up to around 100 nm, there is a sharp transition to the supra-domain regime with $D_{log} = 3$.

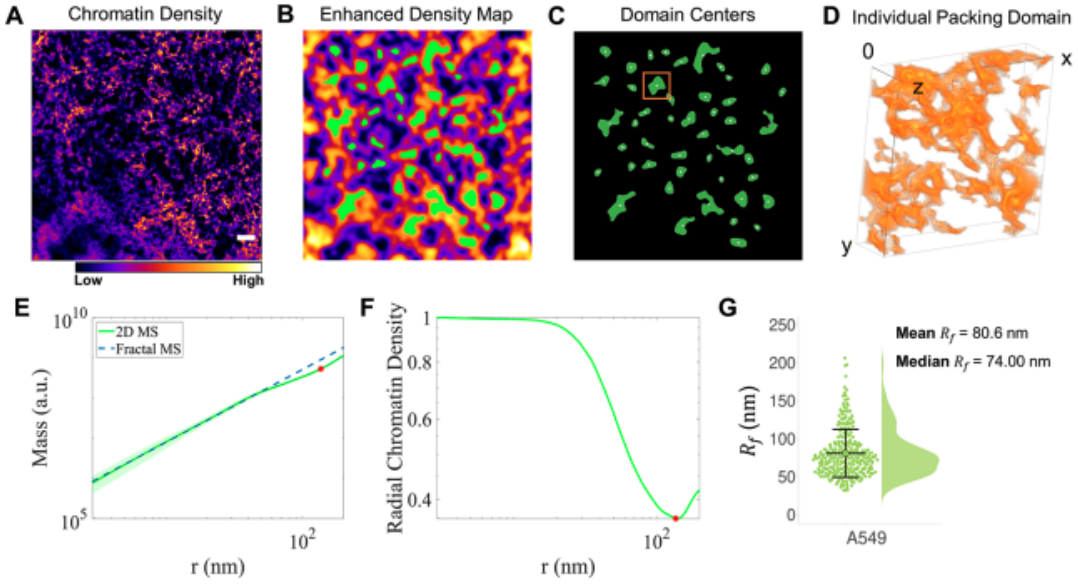

Figure 2.2. Quantifying PD size and chromatin packing behavior at PD boundaries. (A) ChromSTEM grayscale tomogram for one field of view of an A549 cell. The color bar represents chromatin mass density. Scale bar: 200 nm. (B&C) Local chromatin maxima map estimated from an enhanced chromatin density projection was utilized to find chromatin PD centers shown in C. (D) 3D rendering of the surface of chromatin density in a region containing the PD of interest (orange square in C). (D) The average 2D mass scaling curve of the chromatin within the region of interest (orange square in B&C). (E) For one PD, the mass scaling curve is resampled from all loci within the PD center identified in C. The mass scaling analysis was conducted starting from the PD center. The mass scaling curve starts with $D < 3$ (blue dashed line) and transitions to values closer to $D = 3$ (beyond the red asterisk). (F) Radial distribution of chromatin density for the same PD. The radial CVC initially decreases slowly within the PD regime. As the length scale approaches the PD boundary (red asterisk), the radial CVC rapidly dips which is followed by a recovery, potentially due to the presence of other domains at those length scales. (G) The distribution of R_f, the PD radius, for A549 cells.

Additionally, we determined the radial distribution of chromatin density, or chromatin volume concentration (CVC), to characterize changes from the "domain center region" to the periphery of individual PDs (Fig. 2.2F). We observed three key trends in the radial CVC at different distances from the PD center: 1) a relatively flat, slowly decreasing curve near the PD center, 2) a rapidly decreasing curve at a moderate distance from the PD center, and 3) an increasing curve at even larger distances. This third trend is likely caused by the inclusion of chromatin from other nearby PDs. The transition point from rapid decrease to increase in radial CVC (red asterisk in Fig. 2.2F) is consistent with the transition point in the mass scaling curve from power-law mass scaling behavior to $D_{log} = 3$ (red asterisk in Fig. 2.2E), and both transitions indicate the edge of the analyzed PD.

Next, we estimated the median genomic size of A549 and BJ domains. Assuming that the highest intensity in the tomograms represents 100% dehydrated DNA (density = 2 g/cm^3) and the average molecular weight for a nucleotide is 325 Da, we calculated the highest mass (m) per voxel (dr = 2 nm) to be ∼15 bp. We further calculated the median genomic size of A549 and BJ PDs to be 207 kilobase pairs (kbp) and 82 kbp, respectively, based on the relation $M = m\left(\frac{R_f}{dr}\right)^D$.

Here, we interpret R_f as the length scale where the chromatin mass scaling behavior no longer follows a power-law relationship, or where a single packing scaling coefficient is not sufficient to explain the packing behavior. However, this view does not indicate that each PD is spherical with radius R_f. We further investigated the shape of the PD boundary by calculating the 2D asphericity (A_s) of the chromatin enclosed by the PD boundary [276, 249]. Considering a 2-dimensional ellipse, $A_s = \frac{(a^2-b^2)^2}{(a^2+b^2)^2}$, where a and b

are the semiaxes of the ellipse. Here, A_s can take on values from 0 to 1, depending on the ratio $\frac{a}{b}$. For the case $a = b$, $A_s = 0$ indicates an isotropic or spherical configuration. In the limit, $a \gg b$, $A_s = 1$ indicates a linear or stretched configuration. To avoid edge effects, we only considered PDs that are entirely within the field of view. We estimated the average of A_s to be 0.446 ± 0.04 from 280 PDs for A549 cells and 0.458 ± 0.05 from 104 PDs for BJ cells, respectively (Fig. 2.9E). Altogether, analysis of individual PDs from two different cell lines demonstrates that chromatin fibers are packed into anisotropic PDs of variable sizes.

2.2.3. Heterogeneous Morphological Properties of Chromatin PDs

Statistical descriptors of PDs, including chromatin packing scaling, average chromatin density, and size of domains were previously determined to be physical regulators of transcription through crowding-mediated effects (see Chapter 1.4.4, Chapter 3) [288]. Thus, characterizing the distribution of statistical properties that control chromatin density distribution can help decode the complex chromatin structure-function relationship.

First, for individual PDs, we obtained the distribution of packing scaling D, with a mean value of 2.619 ± 0.010 for A549 cells (Fig. 2.3A), and $D = 2.622 \pm 0.012$ for BJ cells (Fig. 2.10A-C), both with relatively wide distributions. For the same PDs, we determined the average CVC per domain to quantify chromatin compaction. For each pixel, a CVC=0 signifies there is no chromatin density within the pixel and a CVC=1 signifies that the entire pixel volume is filled by chromatin density. Similar to the anisotropy analysis, we excluded the PDs at the edge of the field of view. We obtained a median CVC of 0.37 for A549 cells (IQR: 32-45%; Fig. 2.3B), and 0.33 for BJ cells (IQR: 24-61% Fig. 2.10D). As with domain size, we again observed large heterogeneity of average chromatin density and packing scaling between PDs within the same cell line.

For a polymer that exhibits power-law mass scaling behavior within a certain regime, such as chromatin within PDs, the relationship between mass density (i.e., CVC) and packing scaling should follow the relationship $CVC = \frac{N_f V_{pix}}{V_f} = A\left(\frac{R_f}{R_{min}}\right)^{D-3} \propto AR_{eff}^{D-3}$ [125]. Here, the total mass of chromatin within a PD $N_f V_{pix} = A\left(\frac{R_f}{R_{min}}\right)^{D}$ is measured as the product of the number of pixels within the PD that contain chromatin, N_f, and the resolution or smallest unit of chromatin measured by ChromSTEM, V_{pix}. R_f and V_f are the PD size and total volume of all pixels within the domain, R_{min} is the radius

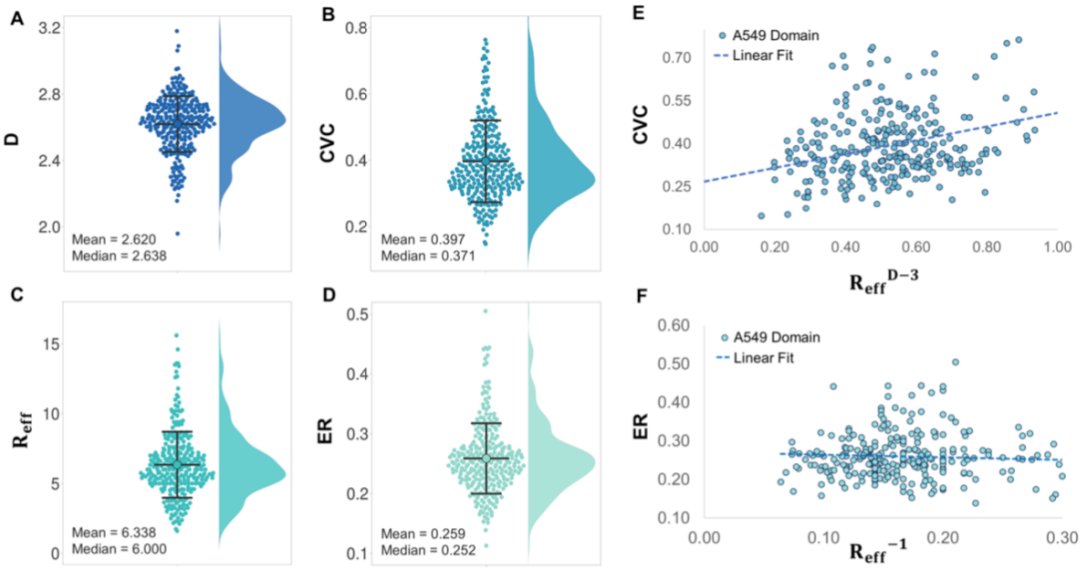

Figure 2.3. Characterizing morphological properties of chromatin PDs in A549 cells. A total of 280 A549 cell PDs were analyzed (A) Chromatin packing scaling D distribution was calculated for A549 cells. The mean of the wide distribution is equal to 2.57. (B) CVC distribution per PD. We observed the CVC distribution ranges from 0.15 to 0.92 with a mean value of 0.40 for A549 cells. (C) Effective domain size R_{eff} for A549 cells. The effective domain size is the ratio between domain size R_f and domain fiber size R_{min}. For A549 domains, the median R_{min} is 11.6 nm. (D) Exposure Ratio (ER) is defined as the fraction of chromatin voxels on the surface of the interchromatin voids compared to the entire volume of the PDs and represents the surface accessibility of PDs to transcriptional machinery. For A549 PDs, the ER ranges from 0.11 to 0.50 with a mean value of 0.25. (E) A moderate correlation between domain CVC and D has been observed for A549 cells, with $R^2 = 0.32$. (F) ER is negatively correlated with inverse effective domain size with the weak linear coefficient for A549, with R^2 =0.1

of the elementary unit of the chromatin fiber, and $R_{eff} = \frac{R_f}{R_{min}}$ is the effective PD size (Fig. 2.3C). A is the packing efficiency factor of the fundamental chromatin fiber within the domain. A chromatin domain with $A = 1$ specifies that each concentric layer of the domain is packed in the most efficient manner, and the mass-density distribution of

chromatin is fully designated by the PD packing scaling. Here, we assume that the packing efficiency within the chromatin fiber, the primary building block is 1. In other words, the entire volume of the fiber is completely filled by chromatin. Similar to R_f, R_{min} can be estimated from the limits of the first regime of the ring mass scaling curve. We define R_{min} as the upper bound of the chromatin fiber regime, or the spatial separation that significantly deviates from the mass scaling behavior within the chromatin fiber. Next, we investigated the relationship between average density (CVC), effective size (R_{eff}^{D-3}), and chromatin packing scaling (D) within PDs across a population of isogenic cells. Our goal was to determine whether there was a universal relationship between CVC and D which could be described by one packing efficiency factor (A) and one fiber size (R_{min}). In general, we observed a positive correlation between CVC and D for A549 cells (linear regression $R^2 = 0.328$) (Fig. 2.3E). From the ring mass scaling analysis, we also observed that R_{min} was not significantly different across PDs, between cells within the same cell lines, and even between the two cell lines. This relationship suggests that the chromatin fiber size may be constant, even across genetically different cells. However, the packing efficiency factor A is PD-specific as there is a wide spread of CVC-D relationships that cannot be described by just one A from one linear regression fit (Fig. 2.3E). Average A for each cell line was evaluated from the regression of CVC on R_{eff}^{D-3}.

As the boundaries of TADs and chromatin PDs are enriched in active transcription processes [284], we next studied how the probability of chromatin being exposed to transcriptional machinery on the PD surface changes across PDs for both cell lines. Here, we define an exposure ratio (ER) as the fraction of voxels containing chromatin on the surface of the PD compared to the total number of pixels encompassing the PD volume.

The surface here exclusively refers to the internal surface created by the interchromatin voids within PDs. This metric evaluates the surface area to volume ratio of a PD. Without changing the genomic size of a PD, an increase in ER for a given chromatin PD would indicate an increase in the chromatin PD surface, which could increase the amount of surface chromatin that is accessible to transcription processes. First, we define A_{sp} as the surface packing efficiency, i.e., the prefactor in the scaling relationship $S_f = A_{sp} S_{min} \left(\frac{R_f}{R_{min}} \right)^{D-1}$ where S_f is the total surface area of the PD and S_{min} is the surface area of the elementary unit of the chromatin fiber, measured as the number of pixels. For each PD, ER can then be estimated by the following relation: $ER = \frac{S_f / S_{min}}{M_f / M_{min}} = A_{ER} R_{eff}^{-1}$, where M_f is the total mass of a PD, M_{min} is the mass of the elementary unit of the chromatin fiber, and $A_{ER} = \frac{A_{sp}}{A}$ is the exposure ratio efficiency factor and represents the ratio between the packing efficiency at the PD surface, A_{sp}, compared to the packing efficiency of the entire PD, A. There is a relatively large variability in the distributions of effective domain size, R_{eff}, (Fig. 2.3C, Fig. 2.10E) and the exposure ratios (ERs) of PDs within each cell line (Fig. 2.3D, Fig. 2.10F). Next, we investigated if A_{ER} is constant for all PDs within the same cell line. We performed linear regression analysis to better characterize the relationship between the inverse effective domain size and ER at the PD level. We observed a weak negative association between the ER and R_{eff}^{-1} for A549 cells (Fig. 2.3F). This suggests that the exposure ratio is very weakly dependent on effective PD size and that the exposure ratio efficiency factor is very small, although it varies between PDs.

Altogether, these results demonstrate that PDs have unique morphological properties that are transcriptionally relevant, including average density, packing scaling, packing efficiency, and exposure ratios, that are heterogeneous within the same cell line.

2.2.4. PDs are Observable in Live Cells and are Heritable across Cell Division

As ChromSTEM has a limited field of view and requires chemical fixation, we used PWS microscopy to inspect the chromatin packing scaling distribution across the entire nucleus and confirm the presence of PDs in live cells. As previously mentioned, PWS microscopy measures chromatin density fluctuations. Chromatin packing scaling, D, can be calculated from these measured fluctuations as described in [74]. PWS analysis also identified spatially separable chromatin PDs characterized by similar D values within each PD. Specifically, the average chromatin packing scaling determined from the ChromSTEM D map (Fig. 2.4) differs from the average D values measured with PWS microscopy (Fig. 2.4) by only ~6%. In summary, by combining the high spatial resolution of ChromSTEM in fixed cells and live-cell imaging capabilities of PWS, we have identified the existence of spatially separable supranucleosomal chromatin PD structures (Fig. 2.4).

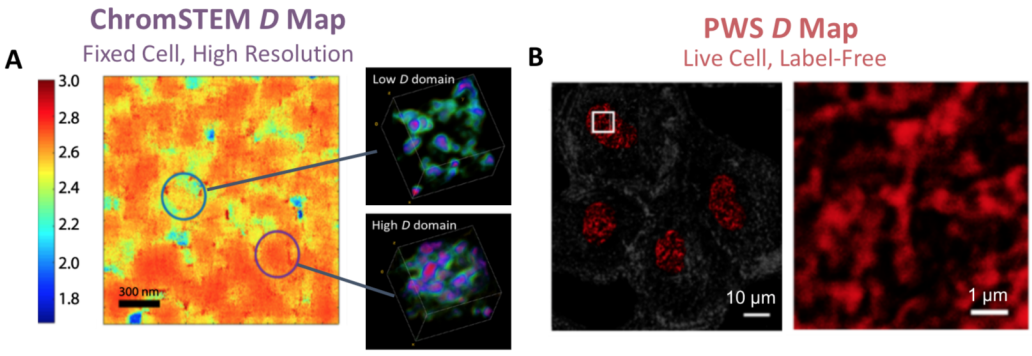

Figure 2.4. PDs are observable structures in fixed and live cells. (A) (Left) Chromatin packing scaling (D) map of a ChromSTEM A549 cell tomogram. (Right) Two PDs with different Ds, highlighted in left map by blue and purple circles, have distinct supranucleosomal packing configurations. DNA concentration increases from green to red. (B) (Left) PWS D map of several A549 cells with nuclei shown in red. (Right) corresponds to inset.

Next, we employed live-cell PWS microscopy to investigate whether chromatin packing behavior is transferred between parent and progeny cells across cell division. With its high-throughput, label-free, and live-cell imaging capabilities, PWS microscopy is uniquely suited for this task. PWS imaging was performed on HCT116 colon cancer cells over the course of 20 hours, over which several cell division events were observed. We then used these measurements to determine how the spatial distribution of the chromatin packing scaling evolves over a long period of time (hours) and quantified the time-dependent fluctuations of chromatin packing scaling across the entire nucleus for parent and progeny cells (Fig. 2.11,2.12). Overall, we determined that PD structure, specifically the spatial distribution of chromatin packing scaling and the temporal fluctuations of average chromatin packing scaling, are heritable through the process of cell division (Fig. 2.11,2.12).

2.2.5. Relationship between Chromatin Packing Behavior and Genome Connectivity

The packing behavior of a mass-fractal polymer is expected to influence the probability distribution of contacts between non-neighboring monomers. We thus wanted to test whether this intuition was the case within chromatin PDs. Contact probability scaling is an important statistical property of chromatin that represents overall chromatin connectivity and can be measured using chromosome conformation capture techniques such as Hi-C. Prior studies have revealed the critical role of such contact properties in transcription regulation, with implications for enhancer-promoter interactions [286]. Returning to simple homopolymeric systems as a conceptual example, the probability of contact (P) between two monomers of N distance apart on the linear chain can follow a power-law scaling relation: $P \propto N^{-s}$, where s is the contact probability scaling exponent. Recent advances in Hi-C have demonstrated that no single power-law scaling exponent can describe chromatin organization throughout the entire nucleus, and several studies have used analyses of genome-wide contact probability scaling behavior to disprove the previously popular fractal globule model of chromatin [251].

A chromatin polymer with a higher D has a lower rate of decrease of CVC as a function of distance from the PD core. Thus, intuitively, a higher D should be associated with higher contact frequencies between distant genomic loci. This would translate into a lower contact probability scaling s. Multiple homopolymer models show an inverse relationship between these two properties, with $s = 3/2$ for a random walk with $D = 2$ and $s = 1$ for a fractal globule with $D = 3$ [193]. Both cases are in agreement with a more general

relationship: $s = 3/D$. Does this inverse relationship still hold for more complex models of chromatin structure, or is it only relevant for these simpler polymer models?

Halverson et al. [108] reached the same functional relationship between s and D by using a mean-field argument. However, this initial derivation assumed that two monomers separated by genomic distance N will have a uniform probability of being at spatial distance r apart anywhere within volume $R_g^3 \sim N^{3/D}$. This assumption is not true for all cases. Good solvent conditions are one such limiting case, which result in polymer swelling, causing repulsions between non-neighboring monomers that would break down the mean-field assumption. Halverson et al. [108] provide additional scaling analysis, demonstrating that the contact probability scaling exponent does not depend solely on D. These results suggest the more complex chromatin polymer might follow a general inverse relationship between these two scaling properties, but the exact functional form has yet to be established.

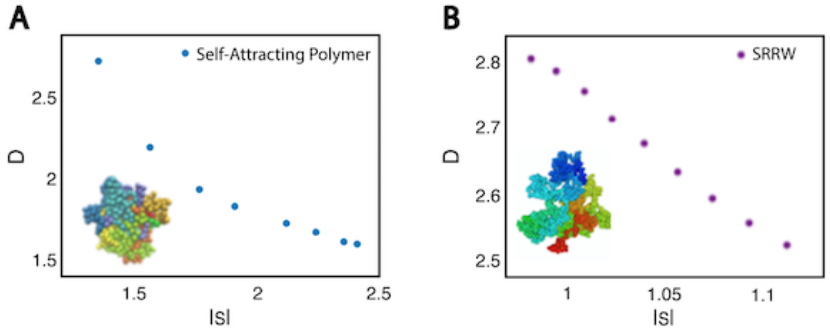

Figure 2.5. Relationship between s and D as determined by modeling. (A and B) A general inverse relationship between s and D is demonstrated using (A) self-attracting polymer and (B) SRRW simulations, although the exact relationship is model-dependent.

Because no existing model can faithfully capture all aspects of chromatin structure, we sought to test this hypothesis by implementing two distinct computational models of

chromatin. The models we use here are not expected to be an exhaustive set but instead were used as test beds to ascertain whether the inverse relationship between s and D was likely to be a model-independent property. First, we ran BD simulations of a basic homopolymer model under varying solvent conditions to represent chromatin structure within PDs for different intranuclear environments. We introduced effective attractions between monomers using the Lennard-Jones (LJ) potential, which physiologically represents the solvent quality of the polymer solution. We tuned the attractive potential between monomers to generate polymers ranging from a swollen self-avoiding walk under good solvent conditions to a collapsed globule under poor solvent conditions. This tuning allowed us to modulate two measurable statistical polymeric properties, D and s, and investigate their relationship. For these simulations, each monomer represented one nucleosome ($\sim$147 bp DNA), and the entire polymer chain contained 1000 monomers.

In addition, we used the self-returning random walk (SRRW) model (discussed further in Chapter 2.3) to investigate this relationship. We varied the SRRW folding parameter to tune statistical chromatin conformation [120]. SRRW conformations were generated as in [120] with each step size representing 2 kbp of DNA ($\sim$10 nucleosomes).

For both models, we determined D and s by performing a linear regression on their respective power-law scaling relations. The linear regression was performed for the genomic range from 20 to 200 kbp, which is within the predicted genomic size of chromatin PDs. Although the two chromatin models resulted in two distinct functional forms of $s(D)$, as would be expected, both models demonstrated an inverse relationship between these two statistical parameters (Fig. 2.5). After computationally establishing a more

chromatin-specific inverse relationship between packing behavior and polymer connectivity, we wanted to investigate whether this property can be observed *in vitro*.

To test this hypothesis experimentally, we used the nano-ChIA platform to measure changes in chromatin packing scaling, D, upon external stimulation, which we compared with changes in contact probability scaling, s, determined by Hi-C analysis. Dexamethasone (DXM) treatment has previously been demonstrated to alter large-scale genome connectivity [67]. Analysis of publicly available Hi-C data revealed that s increases upon 32 hours of DXM treatment in differentiated BJ fibroblast cells treated with 100 nM DXM (Fig. 2.6 A-D), which we hypothesized would result in an inverse change in chromatin packing scaling.

Thus, we first used ChromTEM (Fig. 2.6 E-H) to measure statistical changes in D before and after treatment with DXM in fixed cells. Unlike ChromSTEM, which resolves the exact 3D structure, ChromTEM images the projection of a thin cross section (50 nm) of chromatin. To calculate chromatin packing scaling D from ChromTEM data, we performed auto-correlation function (ACF) analysis (Fig. 2.6G). The ACF of the spatial variations of the density of a polymer, such as chromatin, can be derived from its mass scaling relationship and is thus used to measure D: $ACF(r) \propto \frac{dM(r)}{dV} \propto r^{D-3}$. For an infinite, continuous, and random structure, the 2D ACF can be considered identical to the 3D ACF of the original 3D structure with high accuracy. For a finite fractal structure, we demonstrated numerically that 2D ACF is more accurate at determining D, compared to mass scaling, for thinner 50 nm ChromTEM sections [167].

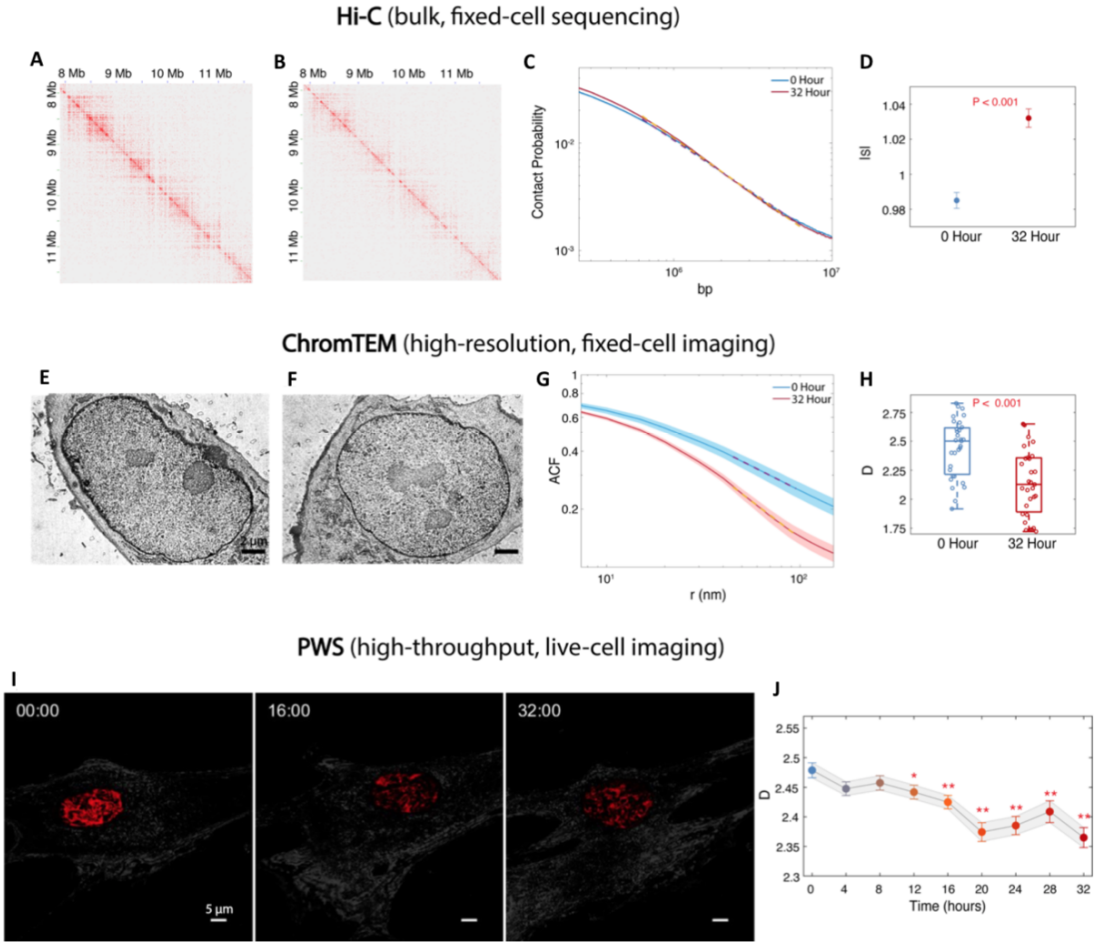
Hi-C (bulk, fixed-cell sequencing)
A
8 Mb
9 Mb
10 Mb
11 Mb
B
C
0 Hour
32 Hour
Contact Probability
bp
D
P < 0.001
ISI
0 Hour
32 Hour
ChromTEM (high-resolution, fixed-cell imaging)
E
F
G
ACF
r (nm)
H
P < 0.001
D
PWS (high-throughput, live-cell imaging)
I
00:00
16:00
32:00
5 μm
J
D
Time (hours)

Figure 2.6. Relationship between s and D as determined by experiments. (A&B) Hi-C contact maps for differentiated BJ fibroblasts treated with DXM treatment for (A) 0 hours and (B) 32 hours. (C) Intrachromosomal contact probability plotted against genomic distance in log-log scale. (D) s for BJ cells treated with DXM for 0 and 32 hours. The linear regression fit was performed on contact probability versus genomic distance between $10^{5.8}$ and $10^{6.8}$ bp. (E&F) ChromTEM images of BJ cells (E) without and (F) with DXM treatment for 32 hours. (G) The average ACF of chromatin mass density for untreated cells (blue) significantly differs from that of treated cells (red). D was measured inside the fractal PD (50 to 100 nm) by a linear regression fit of the ACF in log-log scale. (H) Using ChromTEM ACF analysis on fixed cells, an increase in D was observed after the 32-hour DXM treatment ($N = 31$ cells per condition; $P < 0.001$). (I & J) Live-cell PWS analysis of BJ cells treated with DXM. (I) PWS images of BJ cells with DXM treatment at 0, 16, and 32 hour time points. (J) Time course PWS measurements showed a significant decrease in D for all time points after 12 hours ($N > 67$ cells; $^{*}P < 0.05$ and $^{**}P < 0.001$) compared to the 0 hour time point.

In agreement with our modeling results, we observed inverse changes in D and s at the level of individual cells upon DXM treatment, as measured by ChromTEM (2.6 E-H). Next, we confirmed that the change in D as measured by ChromTEM in fixed cells was comparable to the D measured in live cells using PWS microscopy, and tracked changes in chromatin packing scaling for cells treated with DXM every 4 hours for 32 hours (2.6 I&J). We found that the relationship between chromatin packing scaling and contact probability scaling was consistent for ChromTEM and PWS microscopy measurements compared to bulk Hi-C methods. Notably, the absolute values of chromatin packing scaling measured by higher-resolution ChromTEM on fixed cells and PWS microscopy on live cells under the control and treated conditions were in good agreement (2.6 H&J). In addition, using PWS microscopy, we saw an intermediate but still significant decrease in D for the midway 16 hour time point, corresponding to an intermediate increase in s determined by further Hi-C analysis at this 16 hour time point (Fig. 2.13). To further test

the inverse relationship between D and s, we performed additional ChromTEM and PWS experiments on A549 cells treated with DXM for 0 and 12 hours and compared the results to publicly available Hi-C results under the same treatment conditions [71]. Again, we observed a decrease in D after DXM treatment and the same inverse relationship between D and s (Fig. 2.14).

Overall, these results suggest that genome connectivity is inversely related to the packing scaling behavior of chromatin within PDs. Thus, measuring chromatin packing behavior within PDs could provide information regarding the overall statistical connectivity of gene promoters and enhancers within a given PD.

2.2.6. Conclusions and Discussion

Consolidating results from electron and PWS microscopies allowed us to uncover the existence of chromatin PDs, identified by their power-law mass scaling behavior, in both fixed and live cells.

First, by employing ChromSTEM on two genetically different cell lines, both chemically fixed A549 cells (cancer) and BJ cells (non-cancer), we were able to quantify chromatin packing *in vitro* down to the level of the primary chromatin fiber. Importantly, we studied these cell lines to distinguish basic principles behind chromatin packing that are generally cell line-invariant. By analyzing the mass-scaling behavior of the chromatin polymer, we observed spatially separable, and geometrically anisotropic, packing domains ∼80 nm in radius averaging across both cell lines (Fig. 2.2). We also demonstrated that PDs exist in live cells using our label-free PWS microscopy technique (Fig. 2.4).

The mass scaling within the packing domains follows a power-law relationship with $D < 3$, indicating that chromatin packs into PDs that have radially arranged layers with decreasing chromatin density from the domain center to the periphery. This "core-shell" structure supports earlier experimental work using SR microscopy at a coarser spatial resolution [53, 55]. At the same time, PDs are not completely isolated from each other without any chromatin density in between, as CVC values are always above 0. From these observations, it is reasonable to suggest chromatin is organized into complex, porous PD structures which are connected by less dense chromatin fibers. The porosity of PDs could provide additional surface area, potentially promoting diffusion and targeted search mechanisms for nuclear processes such as transcription. Outside of PDs, the packing scaling increases to 3 after crossing the PD boundary. A packing scaling of $D_{log} = 3$

potentially indicates a random distribution of multiple domains with respect to each other and, importantly, does not substantiate the existence of higher-order packing structures above the level of PDs.

Interestingly, the previous ChromEMT study did not observe any higher-order chromatin structures above the level of the primary fiber [212], which is incongruous with other EM and optical microscopy studies. The size of the PDs observed using ChromSTEM (∼160 nm diameter) is consistent with previous observations of higher-order chromatin domains, including 'chromomeres' (∼200-300 nm) [101], replication domains (∼110-160 nm)[207, 57, 300], and domains associated with TADs (∼200-300 nm) [194]. However, the PD structures observed using ChromSTEM are obtained at a much higher resolution than previous SR optical techniques. Additionally, instead of utilizing conventional TEM tomography as reported in the ChromEMT approach [212], ChromSTEM utilizes quantitative STEM HAADF imaging. Unlike TEM signal, STEM HAADF signal is approximately linearly proportional to the chromatin concentration and therefore enables a more in-depth characterization of higher-order chromatin structures at the single cell level. This allows for highly accurate characterization of the packing of the primary chromatin fiber down to sub-6-nm resolution for sections of the nucleus $2\,\mu m \times 2\,\mu m \times 300\,nm$.

Previous imaging studies have investigated domains by either labeling specific genomic regions, including replication domains [207, 57, 300] and Hi-C-identified TADs [275, 194] or by delineating boundaries based on absolute chromatin density distribution or coherent motion [101, 207, 55, 194]. On the other hand, ChromSTEM PDs were identified by a polymer physics-based mass scaling analysis. Thus, the boundaries were not known a priori as in the other labeling experiments. Additionally, the ChromSTEM technique

enables the quantification of chromatin structure down to the level of the DNA base pair, resulting in a highly detailed characterization of 3D chromatin conformation within PDs.

Additionally, the statistical packing behavior of PDs has more direct functional implications than domains distinguished by chromatin density or motion. Functionally important properties of the packing domains, including average chromatin density, domain size, packing scaling, surface exposure ratio, and packing efficiency, which are all potential regulators of crucial nuclear processes (see Chapter 3) [288, 167] can be determined from ChromSTEM analysis. Such analyses can thus help us to understand the implications of chromatin structure on gene expression and vice-versa.

Notably, we also determined that contact probability scaling and chromatin packing scaling within these PDs follow an inverse relationship through computational modeling (Fig. 2.5) and experimental cross-validation with Hi-C and ChromTEM and PWS nanoimaging methods (Fig. 2.6,2.13,2.14). This observation adds to the functional relevance of packing domains, as the packing scaling within PDs could relate to enhancer-promoter connectivity probabilities.

Our previous experiments on isogenic cell lines have demonstrated D as a crucial modulator of transcriptional plasticity (see Chapter 3) [288]. The large range in D values observed in this study within the same cell line implies that genes may be localized into PDs with different D values depending on how responsive the gene must be to external stimuli, and that this could be potentially co-opted in the cancer cell state for chemoevasion purposes. Furthermore, for PDs of both A549 and BJ cell lines, we observed a diverse range of average chromatin densities, domain sizes, asphericities, and exposure ratios, all of which could also impact transcription rate (Fig. 2.3). Due to the power-law mass

scaling observed in PDs, some of these morphological properties are interrelated from a polymer physics perspective. From ChromSTEM data, although they are positively correlated, we observed chromatin density and packing scaling within PDs cannot be described by a universal relationship. We also observed a similarly complex relationship between the exposure ratio, the probability of a chromatin segment to be on the domain surface, and PD size. The heterogenous morphological properties of domains could potentially play a role in regulating gene activities by controlling the size of proteins and other macromolecular complexes that can navigate through this network, thus influencing material transportation and gene accessibility.

Higher-order chromatin structure changes significantly throughout the cell cycle. Mitotic chromosomes lose their cell type–specific organization and gene expression profiles, yet both are reestablished upon mitotic exit [203]. This poses the question of whether chromatin organization can be preserved over generations of cells, and in what sequence the higher-order chromatin structures are reestablished. Unfortunately, nanoimaging techniques such as ChromSTEM and biochemical methods such as chromosome conformation capture can provide only snapshots of chromatin organization, as chemical fixation is involved. Using our live-cell, label-free PWS modality, we uncovered a strong correlation between the chromatin packing scaling of progeny cells, which is also correlated with that of the progenitor cell (Fig. 2.11, Fig. 2.12). For the same progenitor cells, we observed significant synchronization of the redistribution of chromatin packing immediately after cell division.

The spatiotemporal coherence of D among progenitor and progeny cells is indicative of a heritable chromatin packing structure. This raises the question of what molecular mechanisms contribute to the reestablishment of higher-order chromatin structure across cell division. Although the molecular mechanisms of PD formation remain to be elucidated, most of the putative determinants are potentially heritable. The expression of ion channels, which are direct regulators of the intranuclear physicochemical environment and thus influence chromatin structure (see Chapter 4), is genetically and epigenetically conserved across cell division. In particular, dysregulated expression and function of ion channels have been associated with the propagation of cancer phenotypes [223]. The CTCF-cohesin complex has been shown to play a crucial role in maintaining coherent, cell type–specific, and heritable TAD boundaries [274]. Furthermore, transcriptional memory propagation occurs through mechanisms such as mitotic bookmarking [215]. In addition, both active and repressive histone modifications are preserved throughout the cell cycle [215], although heterochromatin marks have been recently shown to be more stably inherited in regions above critical densities of these marks [58]. Future investigations elucidating the contribution of these potential mechanisms to the heritability of supranucleosomal chromatin organization may, in turn, provide insights into the phenotype of disease states such as cancer.

The major limitations of ChromSTEM/TEM include chemical fixation, low throughput due to electron tomography, and the inability to obtain locus-specific information. Therefore, ChromSTEM findings are not directly comparable to discoveries obtained from sequencing-based techniques such as Hi-C or locus-based imaging methods such as Fluorescence In Situ Hybridization (FISH). Additionally, ChromSTEM involves reagents

such as DRAQ5, DAB, and osmium for DNA-specific labeling that may alter nucleotide structure, however the reagents are added after chemical fixation to minimize the effect. Dehydration and resin embedding are also known to create unavoidable volume changes.

Another caveat of our nano-ChIA platform is that the conversion from PWS signal to D assumes that the mass density distribution of chromatin has a given ACF. Although the mass density ACF has been confirmed by ChromTEM [74], this does not mandate that this is correct for all cell lines under all treatment conditions and ChromSTEM/TEM analysis should be performed before making these assumptions. Additionally, PWS directly measures mass density fluctuations in the nucleus, which is composed of other macromolecules besides chromatin. Mobile crowders might also contribute to these fluctuations, but we assume variations in chromatin mass density dominate PWS signal. This is reasonable given the large density occupied by chromatin within the nucleus. Additionally, the similar behavior of chromatin packing for BJ and A549 cells with and without dexamethasone treatment observed for PWS microscopy and ChromSTEM/TEM approaches provides confidence that these assumptions are fairly accurate (Fig. 2.6, Fig. 2.14). Despite its limitations, we believe that the nano-ChIA imaging platform and the associated analysis methods presented in this thesis work should become an important tool for understanding the 3D structure and function of chromatin.

Future experiments are necessary to elucidate the molecular basis of these PDs, as it is currently unknown whether they are the same as other domain-like structures that are observable with different methodologies, such as TADs. First, we would like to point out that several properties of PDs are similar to those of sub-TADs and TADs. From ChromSTEM data, we estimated the average genomic size of PDs to be 207 kbp in A549

cells and 82 kbp in BJ fibroblasts, which is within the range of typical TAD and sub-TAD sizes [68, 274]. However, we assumed that the highest DNA intensity in ChromSTEM data represents pure, dehydrated DNA, which is likely to be an overestimation. A more accurate evaluation requires additional calibration experiments to link ChromSTEM image contrast to the total mass of DNA at different pixel sizes. In addition, PDs are heritable through the process of cell division, as are TADs. Finally, the genome connectivity behavior within TADs is potentially related to the 3D conformation of the chromatin chain within PDs. However, a definitive link between these structures has yet to be established.

In the future, co-registering ChromSTEM with 3D SR techniques which enable labeling of markers for heterochromatin and euchromatin will help improve our understanding of the relationships between the physical structure of chromatin within PDs, epigenetic modifications, and transcription. Future work should also focus on developing novel locus-specific labeling methods that are compatible with ChromSTEM sample preparation and imaging, and colocalizing chromatin morphological and genetic information for a greater number of cells. For example, labeling TADs identified by Hi-C experiments on the same cell lines and co-registering them with ChromSTEM could help to elucidate the relationship between PDs and TADs. This would require the development of non-FISH-based labeling methods which are not reliant on DNA denaturation, a process that disrupts the endogenous chromatin packing structure at the nanoscale level. Additionally, such studies could help to uncover a domain-specific relationship between contact probability scaling and chromatin packing scaling. Altogether, these experiments would help to better understand the functional consequences of observed PD organization.

2.2.7. Supplementary Material

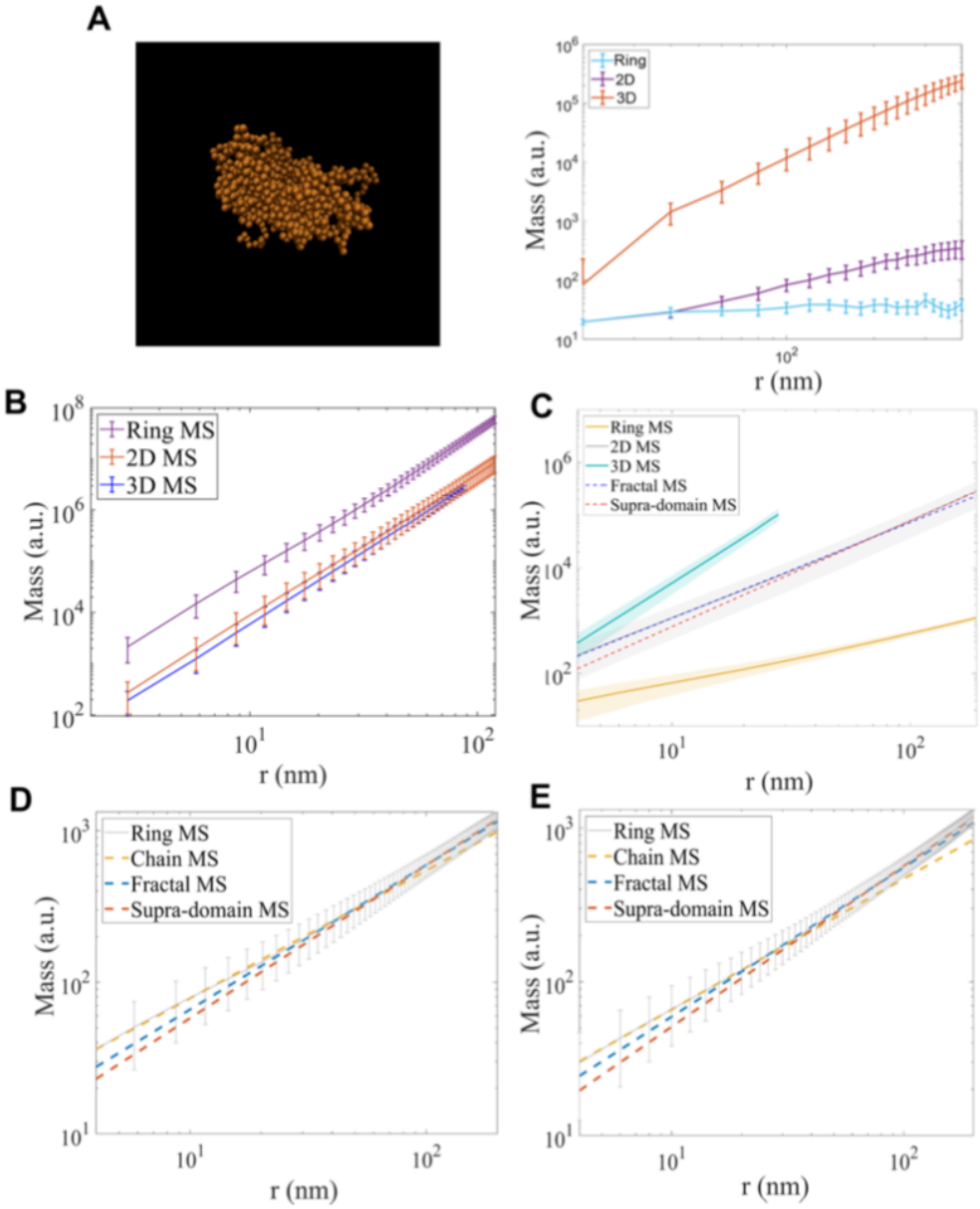

Figure 2.7. Mass scaling analysis at different dimensions. (A) Rendering of a self-attracting homopolymer with $D = 2.37$ (left) which was estimated from 3D, 2D, and Ring mass scaling (right). 3D mass scaling exponent can be approximated using the relations $D_{3D} = D_{2D} + 1$ and $D_{3D} = D_{1D} + 2$, with standard errors of the mean of 0.023 and 0.019 respectively. (B) A549 mass scaling curves plotted as 3D mass scaling, 2D mass scaling + r, Ring mass scaling + 2r in the log-log scale. The equivalent slope for the 3D mass scaling regime extending from 2 - 100 nm indicates that the 3D mass scaling exponent can be derived from 2D and ring mass scaling exponents. (C) The average mass scaling curves determined from different dimensions of the amalgam of ChromSTEM tomograms from three BJ cells. (D) The ring mass scaling curve for A549 cells has three regimes: 1. Chain mass scaling with slope, $D = 2.88 \pm 0.2$ fitted from $r = 2$ - 10 nm (yellow dashed line); 2. Domain mass scaling (blue dashed line); 3. Supra-domain mass scaling with slope, $D = 3.01 \pm 0.01$ fitted from $r = 145$ - 200 nm (red dashed line). (E) The ring mass scaling curve for BJ cells also shows three regimes: 1. Chain mass scaling with slope, $D = 2.85 \pm 0.04$ fitted from $r = 2$ - 10 nm (yellow dashed line); 2. Domain mass scaling (blue dashed line) ; 3. Supradomain mass scaling with slope, $D = 3.04 \pm 0.003$ fitted from $r = 100$ - 140 nm (red dashed line).

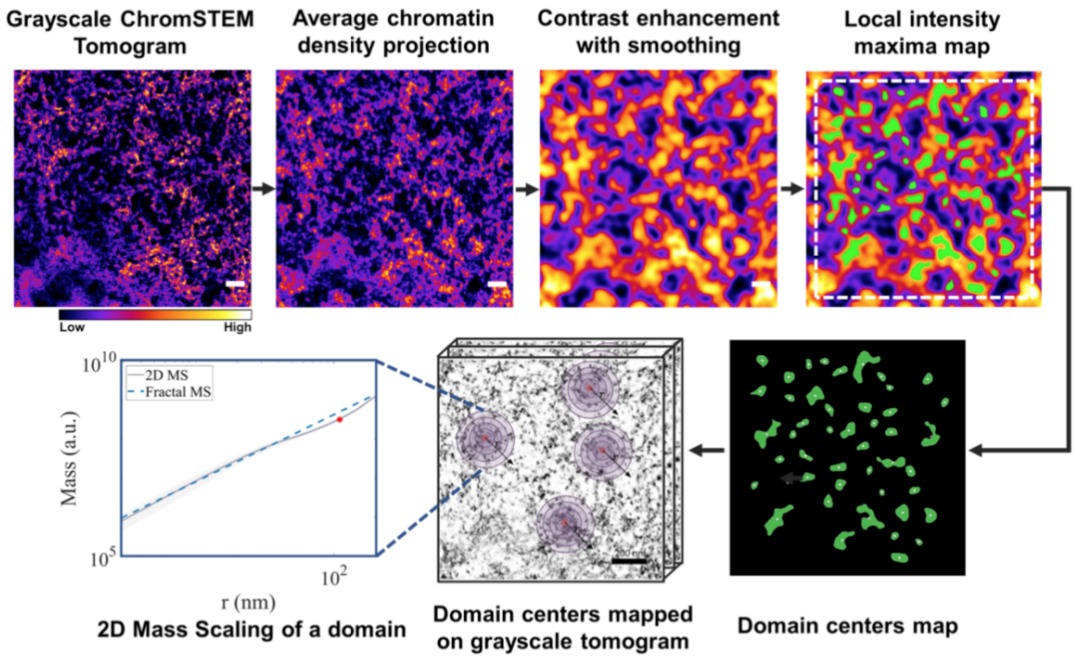

Figure 2.8. Chromatin mass density analysis to identify PD centers. The average z-projection for the grayscale tomograms was evaluated to obtain a map of the average chromatin mass density distribution using ChromSTEM-HAADF intensity. Then we applied Gaussian filtering with radius = 5 pixels followed by CLAHE contrast enhancement with a block size of 120 pixels in FIJI. We identified the local maxima for unbiased segmentation of chromatin PDs. We then identified the center of mass pixels for each segmented PD. To obtain the mass scaling curve for a single PD, we first sampled multiple mass scaling curves starting from the nonzero pixels within the PD centers, defined as the 15 pixel x 15 pixel window surrounding the center pixel. We then used the average mass scaling curve for that PD for subsequent analysis. Scale bar: 200 nm.

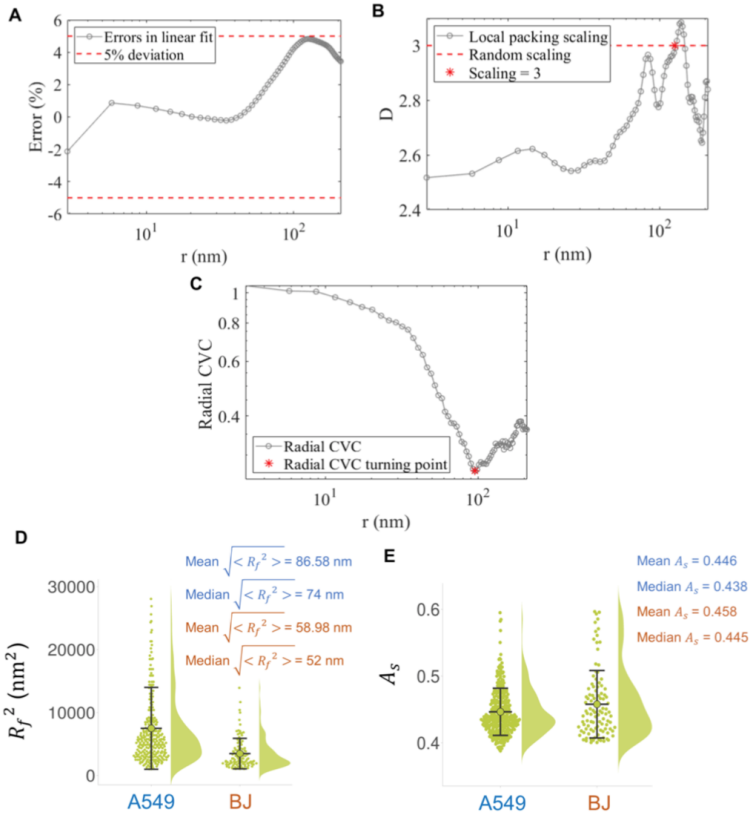

Figure 2.9. Determining PD boundaries from the mass scaling behavior. Beyond a given length scale, r (nm), the 2D mass scaling curve deviates from a power-law mass scaling. We performed three types of analyses (A-C) to determine the boundary of PDs, which are denoted as the smallest r of the three analyses, if it exists. (A) Mass scaling curve deviates from the initial power-law mass scaling calculated from small length scales within PD center region by 5%. (B) Local packing scaling D_{log} reaches 3 at r = 102 nm. (D) The absolute value of the second derivative of the logarithm of the mass scaling curve is greater than 2, indicating a divergence from the power law. Here, all length scales follow under this error margin. (C) The radial CVC starts to increase. The radial CVC decreases initially, then increases at $r = 95.7$ nm for this domain. In this case, comparing (A-C), we determined the domain size R_f =95.7 nm. (D) The distribution of R_f^2, the square of the radius of the PD, for A549 (blue) and BJ (orange) cells. (E) The distribution of A_s, the asphericity of the chromatin density distribution within the identified PDs, for A549 (blue) and BJ (orange) cells.

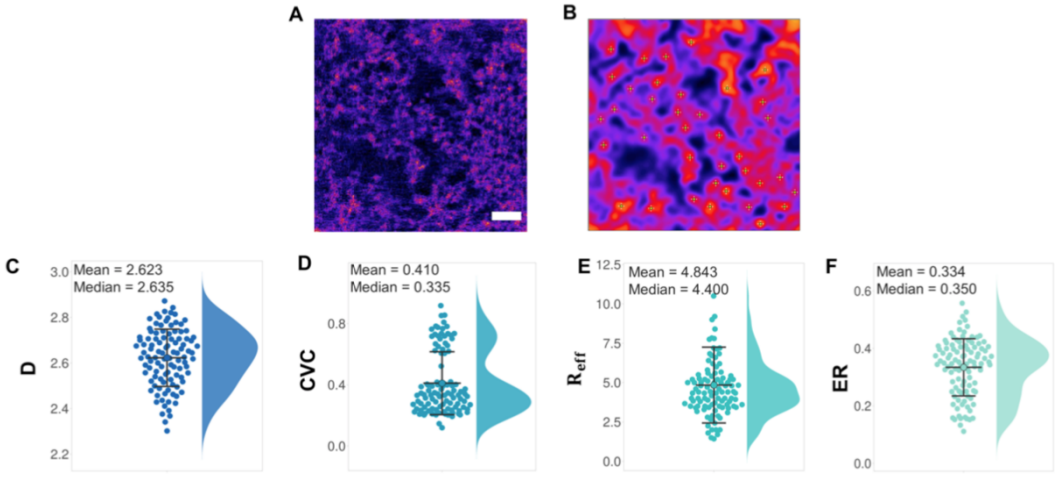

Figure 2.10. Characterizing morphological properties of chromatin PDs in BJ cells. The grayscale BJ cell tomogram (A) was utilized to estimate the (B) chromatin PD centers. Scale bar: 100 nm. (C) Chromatin packing scaling D distribution with a mean equal to 2.62 was evaluated for BJ cells. (D) CVC distribution ranges from 0.12 to 0.92 with a median value of 0.34. (E) Mean effective domain size R_{eff} was 4.84, and median R_{min} is 10.8 nm. (F) Exposure Ratio (ER) ranges from 0.11 to 0.56 with a mean value of 0.35.

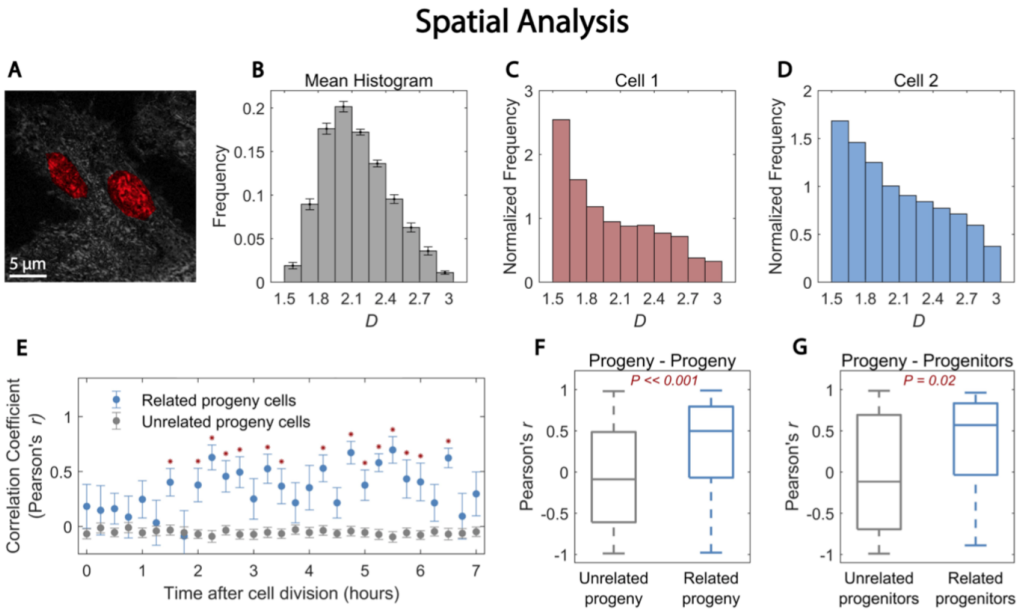

Figure 2.11. Time-resolved PWS imaging of HCT116 cells determines spatial heritability of chromatin packing scaling for $N = 10$ progenitor cells and $N = 20$ progeny cells. (A) PWS D map of two progeny cells originating from the same progenitor. (B) Average spatial D distribution of all cells imaged 5 hours after cell division. (C and D) Histogram ratio of the spatial D distribution for each individual progeny cell from (A) normalized by the average histogram of all cells at that time point from (B). (E) After cell division, the normalized histograms of paired progeny cells are more highly correlated with each other than with unrelated progeny cells at the same time point ($P < 0.05$). (F) Across all time points, normalized histograms of paired progeny cells are more significantly correlated compared to those of unrelated progeny. (G) Comparing all progeny cells 3 hours after division to all progenitors 3 hours before division shows that progeny cells have a higher correlation with their "parent" than with unrelated progenitors ($P = 0.021$).

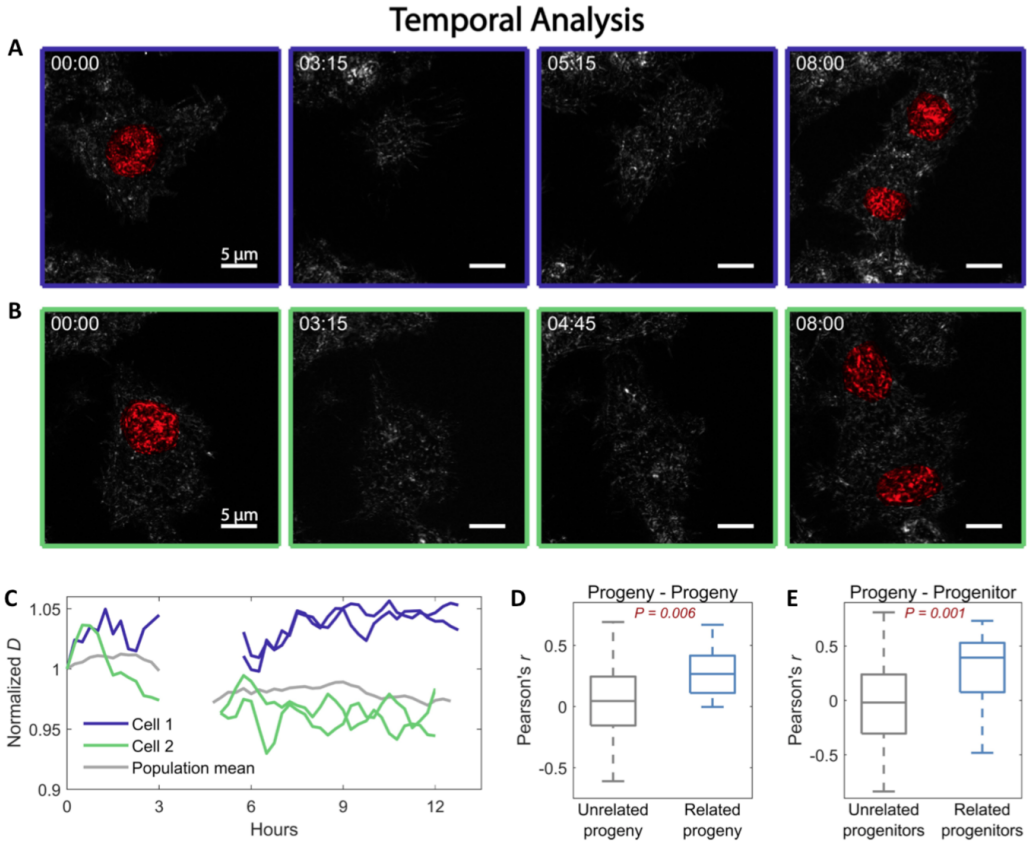

Figure 2.12. Time-resolved PWS imaging of HCT116 cells determines temporal heritability of chromatin packing scaling for $N = 10$ progenitor cells and $N = 20$ progeny cells. (A & B) PWS D maps at four time points before, during, and after cell division. During cell division, nuclei exit the objective's depth of field by lifting off the glass and return to the glass when they have finished dividing. (C) Average nuclear D tracked over time from cells in (A) and (B). After ∼5 hours, both cells have finished dividing, and their progeny cells were tracked for an additional ∼7 hours. (D) D of progeny cells is more strongly correlated with that of their paired progeny than with other unrelated cells ($P < 0.001$). (E) Progeny cells are more correlated with their parent progenitor cells than with other unrelated cells.

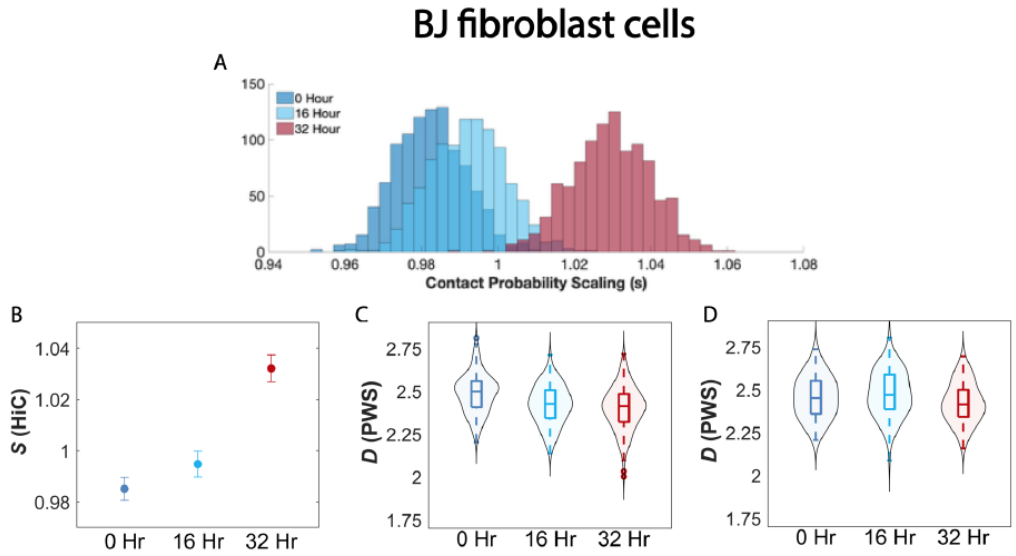

Figure 2.13. Measuring chromatin packing scaling and contact probability scaling alterations induced by dexamethasone (DXM) treatment in BJ differentiated fibroblasts. (A&B) Contact probability scaling analysis as analyzed by moving-window linear regression (A) and classical linear regression (B) analyses for BJ cells treated with DXM for 0, 16, and 32 hours. For (A), we assume the linear regression fit used to calculate contact probability scaling follows a normal distribution $\mathcal{N}(\mu_s, \sigma_s)$ where the mean contact probability scaling, μ_s is the slope of the regression and standard deviation, σ_s is the root-mean-square error (RMSE) of the residuals. Contact probability scaling is significantly different between control, and 16 hour and 32 hour DXM treated BJ cells ($P < 0.001$). (C-D) PWS analysis of BJ cells at 0, 16, and 32 hour time points for untreated cells (C) and DXM treated cells (D). (C) Chromatin packing scaling D measured in untreated live BJ cells shows no observable change in D over a 16 hour period and only a small decrease in D over a 32 hour period that is not statistically significant ($P > 0.05$). (D) Using double-sided student-t-test, $P < 0.005$ for DXM treated cells at 16 and 32 hour time points.

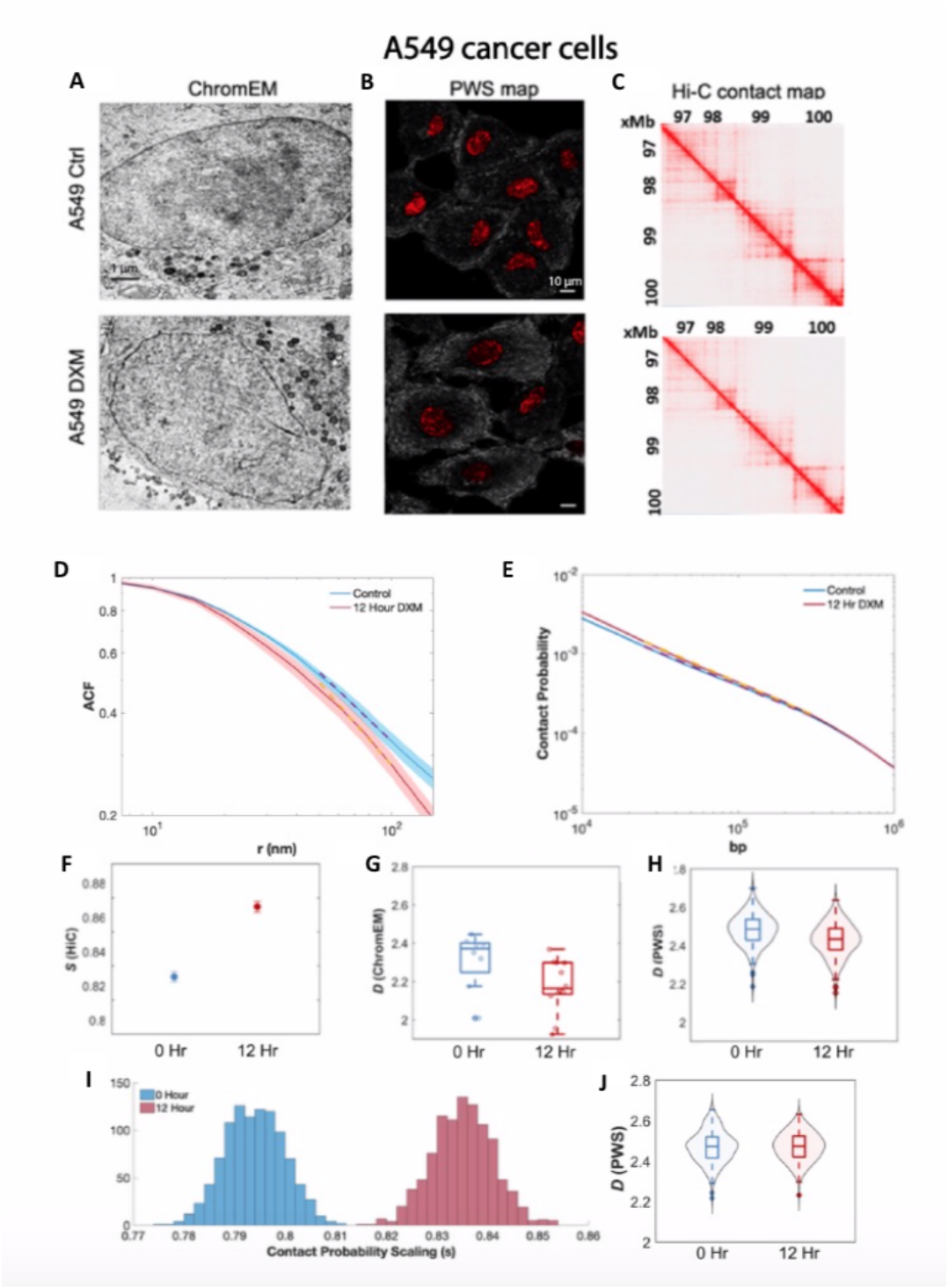
A549 cancer cells
A
ChromEM
B
PWS map
C
Hi-C contact map
A549 Ctrl
A549 DXM
1 μm
10 μm
xMb
97 98 99 100
D
Control
12 Hour DXM
ACF
r (nm)
E
Control
12 Hr DXM
Contact Probability
bp
F
S (HiC)
0 Hr
12 Hr
G
D (ChromEM)
H
D (PWS)
I
0 Hour
12 Hour
Contact Probability Scaling (s)
J
D (PWS)

Figure 2.14. Measuring chromatin packing scaling and contact probability scaling alterations induced by dexamethasone (DXM) treatment in A549 cancer cells. (A&B) Characterization of A549 chromatin structure with and without DXM treatment. From left to right: TEM images of chromatin structure with ChromEM staining, scale bar: 1μm. PWS map of chromatin packing scaling, scale bar: 10 μm. Qualitatively, both ChromTEM and PWS images show that DXM treatment homogenizes chromatin packing. (C) Hi-C contact map of human chromosome 1 rendered with 5 kbp resolution for the control and DXM treated A549 cells. (D) ACF analysis of ChromTEM images of A549 cells. The average ACF of the control group (blue) is significantly different from the average ACF of the treated group (red). The shaded regions represent standard errors. D was calculated from the PD regime (50 nm to 100 nm) by performing a linear regression fit to the ACF in the log-log scale. (E) Contact probability analysis performed on whole-genome intrachromosomal Hi-C contact data. Contact probability scaling (s) was calculated from a linear regression fit (dotted line) of the contact probability curve in the log-log scale between genomic distance $10^{4.4}$ and $10^{5.5}$ bp. (F-H) Chromatin packing scaling alterations induced by DXM treatment measured using ACF analysis of TEM images and PWS and changes in contact probability scaling of Hi-C contact data. Across the platform, consistent changes were observed in chromatin packing scaling upon treatment. Using double-sided student-t-test, $P = 0.051$ for ChromTEM, $P < 0.005$ for PWS. (I) Comparing distributions of contact probability scaling for A549 cells calculated from analysis of Hi-C contact matrices. We assume the linear regression fit used to calculate contact probability scaling follows a normal distribution. Contact probability scaling is significantly different between control, and 12 hour DXM treated A549 cells ($P < 0.001$). (J) Chromatin packing scaling D measured in untreated live A549 cells at 0 and 12 hours shows no observable change in D at time points comparable to DXM treatment, which showed a statistically significant decrease in D after 12 hours of treatment (H).

2.3. Self Returning-Excluded Volume Model Uncovers Fundamental Principles of Chromatin Organization

2.3.1. The Self-Returning Random Walk (SRRW) Statistical Model

Kai Huang developed the Self-Returning Random Walk (SRRW) model, which was initially published in Ref. [120]. The SRRW is essentially a random walk with specific rules introduced to capture statistical features of chromatin organization as revealed by experiments. At each step in the SRRW generation there are two possibilities: (1) Perform a forward jump or (2) Return over the previous step to the previous position. The probability distribution function controlling this decision is

$$P_R(U_0) = 1 - P_F(U_0) = \frac{U_0^{-\alpha}}{\alpha}. \tag{2.1}$$

Here, P_R and P_F represent the p.d.f. for a return or a forward step, respectively. U_0 is the length of the previous step along the backbone. The folding parameter $\alpha > 1$ controls the number of returns. If the SRRW does not continue with a return step, it must continue with a forward jump. The new forward jump is chosen with an arbitrary direction and with a length U_1 given by the following p.d.f.

$$P_J(U_1 > 1) = (\alpha + 1)U_1^{-(\alpha+2)}. \tag{2.2}$$

There is a minimum size for the forward jumps that also defines the unit of length in the model. To avoid unrealistically long steps, we introduce a local cutoff to discard jumps with lengths lying on the higher 0.1% of the probability distribution Eq.(2.2). Additionally, to incorporate the confinement effect on chromatin induced by a finite nuclear

size, we impose a global cutoff of 2α times the local cutoff. This global cutoff is applied during the generation of the conformation, and is measured from the center of mass of the already-generated steps. Notice that the SRRW is fully defined by the single folding parameter α. The SRRW conformations can be analyzed in terms of a backbone from which tree-like structures branch away. For $\alpha \gtrsim 1$, the resulting SRRWs have a short backbone with tree structures involving many segments. For large α, the SRRWs are essentially a linear random walk with the backbone comprising the great majority of the segments and very few branching trees.

Notably, the statistical SRRW model is able to reproduce several key experimentally observed features of chromatin structure. These include: (1) hierarchical organization due to its representation as a tree-like data structure, (2) heterogeneous mass density distribution due to its heterogeneous step sizes, (3) different mass scaling regimes within versus outside of domains as observed using ChromSTEM experiments, and (4) different contact probability scaling behavior within versus outside of domains, as observed by Hi-C experiments, caused by an increase in contact frequency within tree domains from the self-returning steps (Fig. 2.21).

2.3.2. The Self Returning-Excluded Volume (SR-EV) Polymer Model

By construction, since the SRRW includes returns over the previous steps, it contains a large number of overlaps. For example, for $\alpha = 1.15$ the number of returns in a conformation with a total number of steps $N = 50,000$ is approximately $23,500$. Therefore, as a representation of a physical system, such as chromatin, the SRRW has two important drawbacks: (1) the conformations violate the principle of excluded volume and (2) it is not a folded linear chain. The schemes represented in Fig. 2.15 exemplify these two points. In order to recover these two physical properties, Marcelo Carignano extended the SRRW statistical model to develop the Self Returning-Excluded Volume (SR-EV) polymer model. In this new method, the overlapping points are transformed into connected clusters of beads that explicitly represent a linear chain (see Chapter 2.3.6.1). This procedure affects the SRRW conformation only at small length-scales, and maintains the overall shape and mass distribution of the original conformation, as exemplified by Fig. 2.15.

At this point it is convenient to introduce the unit conversion to map both models to a genome system. For the SR-EV model we keep the dimensions introduced originally for the SRRW. The minimum step size represents 30 nm, and each step includes 2000 bp. In Fig. 2.22 we show the mean quadratic distance $\sqrt{\langle R^2(s)\rangle}$ between two monomers separated by a genomic distance s. The plots are obtained as a population average over sets of 1000 conformations. The plot shows that the difference between the two models is reflected only at small genomic distance separations. This is indeed the expected result, since the SR-EV was designed to maintain the overall structure of the conformation while only affecting the overlaps. Since the overlaps are mainly due to the return steps, this

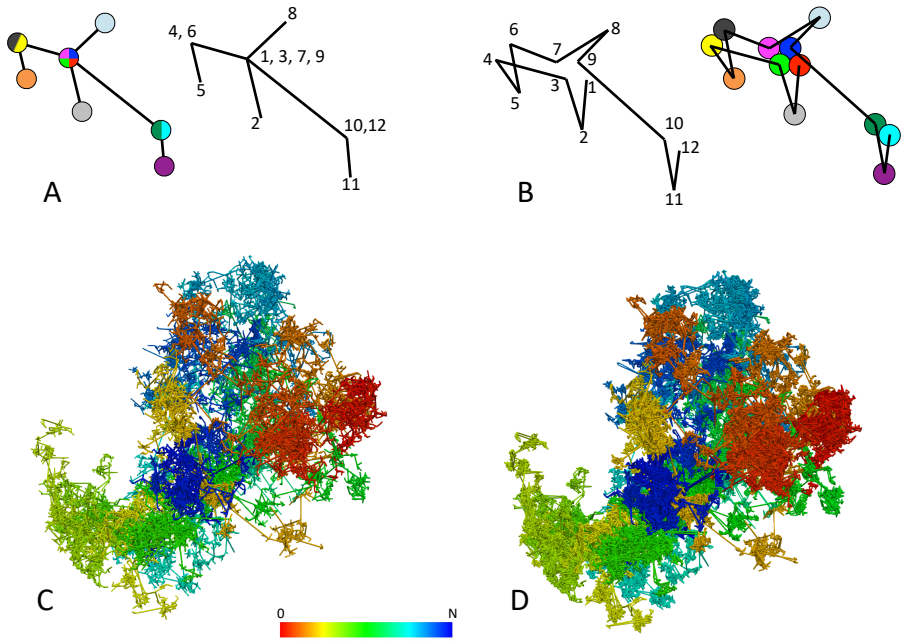

Figure 2.15. SR-EV removes overlaps introduced by self-returning steps of SRRW. (A) Schematic representation of the first 12 beads (left) and their connecting bonds (right) of an SRRW conformation of 50,000 steps and $\alpha = 1.15$. Circles with more than one color represent overlapping points. (B) Equivalent connectivity (left) and beads (right) of the corresponding SR-EV. (C) Rendering of the full SRRW conformation and (D) its SR-EV counterpart. The color bar indicates the connectivity from the start to the end of the conformation.

effect is significant only at small genomic distance separations. Disregarding this small discrepancy between the models, we find that, for $\alpha = 1.15$ (1.10), $\sqrt{\langle R^2(s)\rangle} \propto s^{\nu}$ with $\nu = 0.38$ (0.36) for $10^5 < s < 5 \times 10^6$. From polymer theory, we know that ν measures the balance of effective attractions and repulsions between the monomers of a polymer chain. The exponent that we find suggests that chromatin has a stronger effective attraction between monomers than an ideal polymer chain in a Theta solvent ($\nu = 0.5$), a direct consequence of the frequent folding. For larger genomic distance separations, the

models reflects the effect of the global cut-off, introduced to account for a well-defined chromosomal territory.

2.3.3. Heat Shock as an Experimental Model of Environmental Stress

Our goal here is to develop a physical model of chromatin that is able to reproduce the experimental results relative to statistical contact probability and mass density distribution with a minimal set of parameters. Besides the goal of fully understanding how chromatin is organized under standard conditions, we are also interested in understanding how nuclear chromatin structure responds to environmental stressors, and if the fundamental organization we observe under normal conditions persists. For this study, we concentrate our attention on heat shock, where cells are incubated at a higher temperature for a defined period of time. Heat shock is a stress response that is evolutionarily well-conserved and enacted by heat shock proteins, which help to mitigate cellular damage induced by temperature elevations by promoting protein homeostasis, among other cellular functions [182, 169, 241]. The heat shock response induces massive transcriptional changes with distinct temporal profiles. The early heat shock response increases production of heat shock proteins within minutes, while upregulation and, predominantly, downregulation of thousands of genes occurs at 1 hour timepoints [241, 182]. Large-scale changes in chromatin connectivity have been observed using Hi-C and HiChIP techniques in both Drosophila and human embryonic stem cells (hESCs) [164, 176]. Rearrangements in chromatin positioning of heat shock-related genes such as HSP70 also influence changes in gene expression [138]. Altogether, we expect that inducing the heat shock response will alter the internal nuclear organization and, in particular, the statistical properties of chromatin.

2.3.4. SR-EV Model Captures Statistical Chromatin Changes Upon Heat Shock

The generation of a sample of SR-EV conformations allows for the calculation of individual properties as well as the estimation of ensemble averages, facilitating a meaningful comparison with experimental results. Our ChromTEM analysis has determined that the mass density distribution of chromatin can be somewhat accurately described by a power-law ACF at certain length-scales [74, 167]. Thus, converting PWS signal to a mass scaling coefficient relies on the assumption that chromatin organizes, at experimentally relevant length scales, as a self-similar mass fractal. The SR-EV conformations enable the direct calculation of the mass scaling behavior. However, to compare experimental results with SR-EV predictions we must keep in mind the key differences between modeling and the experimental setting to avoid artifacts. The SR-EV conformations are created in free space with no periodic boundary conditions or hard, perfectly defined confinement. Consequently, the outer boundary region of the SR-EV is not well-defined and the radial density vanishes at the end of the chromosome, instead of transitioning to the territory of the neighboring chromosomes as occurs in the nucleus [54]. In order to prevent spurious edge effects, we began the calculation of the mass density ACF by determining the center of mass of the given conformation, r_{CM}. Next, we consider a sphere centered at r_{CM} with radius $r_{in} = 360$ nm. All SR-EV beads contained in this inner sphere were used as a reference point to calculate the commutative mass scaling behavior. We use all monomers of the SR-EV conformation as valid second points for the calculation, but we analyze the mass scaling behavior only up to $r_{out} = 450$ nm. Beyond that, the system density shows substantial decay and the results are not comparable to the experiments.

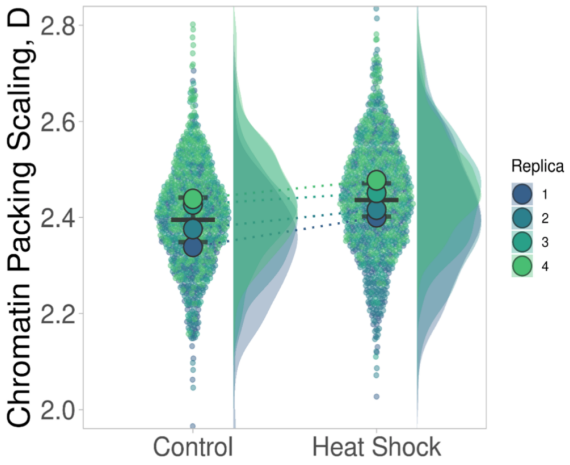

Figure 2.16. PWS microscopy reveals an increase in chromatin packing scaling, D upon heat shock for 1 hour at 42°C compared to controls incubated at 37°C. Each color in the violin plot represents a separate experiment. A total of 4 experiments, each with a number of cells greater than 167, were averaged to obtain a mean D of 2.4 for control and 2.44 for the heat shock conditions.

Here, we compare changes in statistical chromatin structure upon heat shock observed by experiments to the statistics of SR-EV conformations with different folding parameters (i.e., α's). We performed heat shock experiments on HCT116 colon cancer cells by incubating cells at 42 °C for 1 hour. This timepoint was chosen because it was consistent with changes in chromatin connectivity demonstrated by Lyu et al. in hESCs [176], and was after a 30 minute timepoint where chromatin connectivity changes were not observed by Ray et al. in K562 human lymphoblast cells [236]. Using PWS microscopy, we measured chromatin packing scaling for the same cell populations before and after exposure to heat shock. To determine population-wide changes in chromatin packing scaling, D, we averaged chromatin packing scaling for entire cell nuclei and then determined the distributions of average nuclear D for HCT116 populations exposed to normal

temperature (control) and heat stress conditions (Fig. 2.16). PWS experiments demonstrate an increase in population-average chromatin packing scaling from 2.40 to 2.44 for control and heat shock, respectively. Thus, there is an increment of approximately 1.5% in the chromatin packing scaling for the heat shocked samples compared to controls.

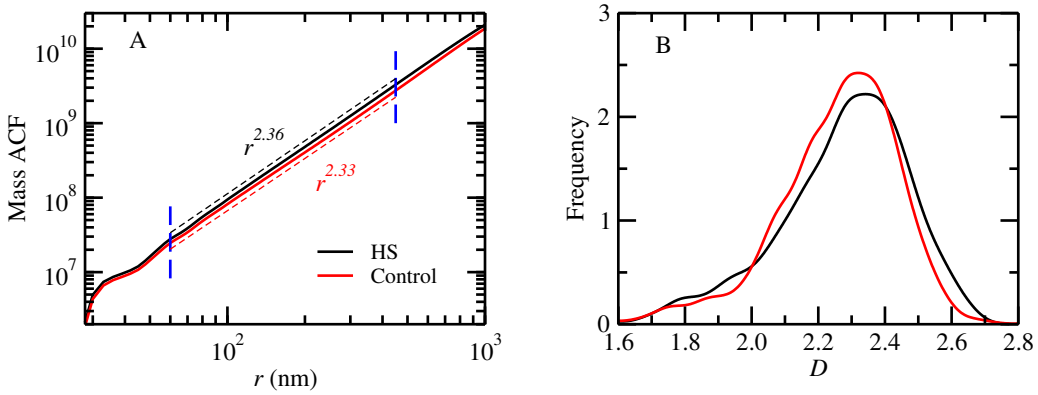

Figure 2.17. SR-EV mass scaling analysis matches PWS heat shock experiments. (A) Average mass scaling for two SR-EV populations with different α's, each with 1000 conformations. The red line, representing the control case, corresponds to $\alpha = 1.15$. The black line, representing the heat shock case, corresponds to $\alpha = 1.10$. The vertical blue lines indicate the boundaries of the power-law fitting, from 60 to 450 nm, which are displayed in dashed lines in the corresponding color next to the raw results. (B) Distribution of D values obtained from the mass scaling analysis from individual conformations in the population.

To compare with experiments, we generated sets of SR-EV conformations with different folding parameter α. In Fig. 2.17A we show our ensemble average for the mass scaling behavior corresponding to two samples, one using $\alpha = 1.15$ that best represents the control case, and a second case using $\alpha = 1.10$ that best represents the heat shock condition. Notably, we can see that there is a power-law scaling behavior spanning from 60 to 450 nm, which is similar to length scales that PWS microscopy is sensitive to (20-300 nm) [46, 3]. The exponent slightly increases from 2.33 for $\alpha = 1.15$ (control) to 2.36

for $\alpha = 1.10$ (heat shock). This increase of approximately 1.3% is comparable to that observed by experimental results. It is important to emphasize that this agreement is obtained using the sample averages for the SR-EV model. Additionally, the mass density ACF of individual conformations allows for the calculation of D for each conformation, similar to the PWS experimental procedure for individual nuclei in a given cell population. In Fig. 2.17B, we show the distribution of D values for each conformation obtained by assuming a power-law scaling regime in the same range used for the average calculations (60-450 nm). The results display a considerable spread, as in the experimental case. Nevertheless, the distribution for $\alpha = 1.10$ is slightly shifted towards higher D values than that corresponding to $\alpha = 1.15$, in full agreement with experimental findings.

Another measurement that probes the organization of chromatin is how the contact frequency between two genomic loci decays as a function of their genomic separation. Previous Hi-C experiments on human embryonic stem cells (hESCs) determined an increase in long-range contact frequencies upon heat shock [176], a trend which was also observed in the original SRRW model [120].

Here, we wanted to investigate changes in the statistical contact probabilities of our new SR-EV model between normal temperature (control) and heat shock conditions for both short and long genomic separations. In our chromatin model, the contact probability is easily estimated from the knowledge of the coordinates of the individual SR-EV conformations, and this is simply averaged over the complete sample of 1000 conformations to obtain ensemble behavior. In Fig. 2.18 we demonstrate the SR-EV contact probability results for both conditions, along with the experimental results obtained from averaging over all chromosomes. Both experiments and theory predict a crossing between the two

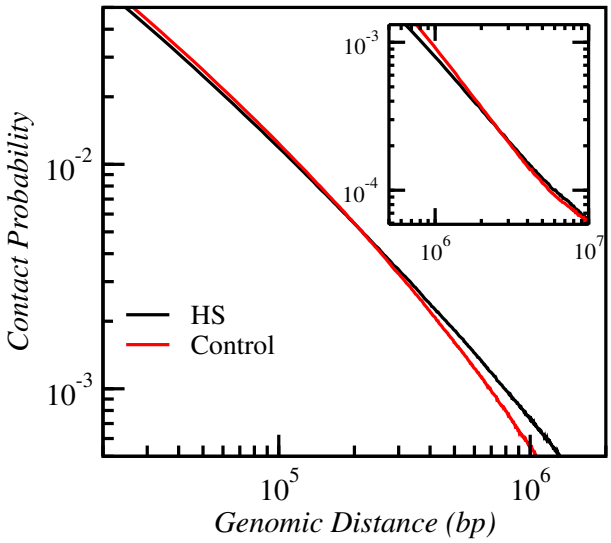

Figure 2.18. Contact probability calculated from the same two populations as Fig. 2.17. The insert shows the experimental results obtained from Hi-C experiments on HCT116 cells. Both experimental findings and theory display the same behavior, with a crossing that shows more contacts for the control sample at smaller genomic separations and more contacts for the heat shock sample at larger genomic separations.

curves at intermediate genomic separations. The crossing occurs at genomic distances of 2×10^6 base pairs for the experimental case, and at 2×10^5 for the theoretical results. No exact quantitative agreement should be expected, as the model considers only one chromosome (total genomic length 100,000 Mbp) and the experiments combine Hi-C results from all chromosomes. Importantly, SR-EV results reproduce experimental observations that the control sample has more contacts than the heat shock for separations smaller than the crossing, and that heat shock has more contacts for larger genomic separations. Also, there is a larger difference between conditions for contact probabilities corresponding to larger compared to smaller genomic separations.

One intrinsic characteristic of the SR-EV conformations is the formation of loops, directly related to the probability of return steps controlled by Eq. (2.1). Large loops may

contain smaller loops, forming a hierarchical branched structure that stems out of the open chromatin backbone. The steps in these loops are not part of the backbone. Therefore, since the total length of the model chromosome is fixed, the length of the backbone is shorter than the total length of the chromatin model. The number of steps involved in the backbone is a direct measure of the number of branches in the SR-EV conformation. Hi-C experiments reveal the formation of TADs as an important structural feature of contact maps. While the sequence-specificity of TADS can depend on biophysical mechanisms such as CTCF-cohesin looping, the existence of loops is cohesin-independent and can be thought of as statistical features of chromatin organization [261, 32].

Using TopDom for TAD annotation [265], Hi-C experiments reveal that the total number of TADs genome-wide decreases after the heat shock treatment. There are 7963 TADs annotated in the control sample, and only 7484 after heat shock, an approximately 6% decrease. Concomitantly, average TAD size increases from 342 to 363 kbp, an overall $\sim$6% increase. Although our SR-EV model does not contain any parameters relating to specific architectural proteins involved in TAD and loop formation, it has an intrinsic statistical rule for the formation of loops. In Fig. 2.19 we show the distributions of the backbone size for our two populations generated with different *alpha*'s. For the control case, $\alpha = 1.15$, the mean value for the number of steps in the backbone is 2486. For the heat shock case, $\alpha = 1.10$, the corresponding mean value is 1364. As backbone size and tree domain size are inversely correlated in the SR-EV model, this implies that the number of steps within tree domains increases in the heat shock case, which is consistent with Hi-C experiments. The question that arises from these observations is why are there

less contacts within loops in the heat shock than in the control case, provided that the tree domains are less populated in the latter case.

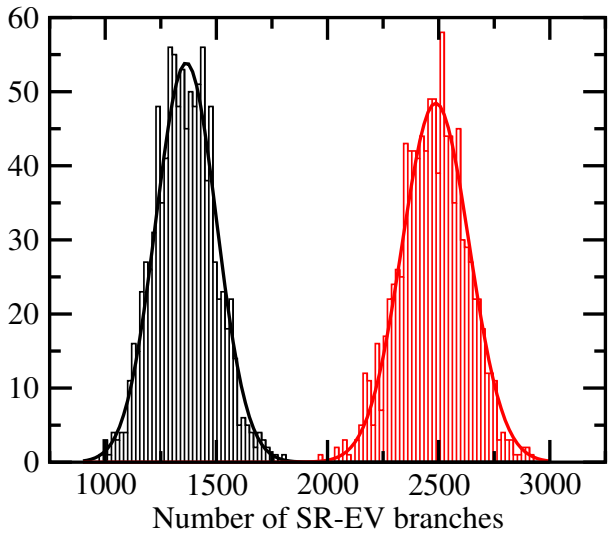

Figure 2.19. Distributions of the number of branches in the SR-EV populations for the control (red) and heat shock (black) cases show a decrease in the number of SR-EV branches, proportional to the number of TADs, in the heat shock condition.

By comparing with the SRRW results that do not include excluded volume, it is reasonable to conclude that the decrease in contact probability at shorter genomic distances in the heat shock case results from the separation of the overlapping beads by introducing the excluded volume effects in the SR-EV model. On the other hand, for the control case, the clusters of beads are more numerous than in the heat shock case and, consequently, the contact probability is larger at smaller to medium genomic distances. For larger genomic distances, the contact probability is dominated by the inter-cluster interactions. Since the heat shock case is associated with large clusters that have considerable surface areas, they favor a higher inter-cluster contact probability than in the control case.

To visualize what those changes connote for actual chromatin conformations, we show in Fig. 2.20 renderings of example conformations for $\alpha = 1.15$ (control) and $\alpha = 1.10$ (heat shock). In a simple representation, where SR-EV steps are colored by their location along the chain, the differences in chromatin rearrangements do not stand out. However, if we represent the backbone chain as a linear molecule (in red) and the branches as spheres (in grey), we can appreciate a clear difference in the granularity between the two conditions. The volume of the spheres is proportional to the number of steps contained in the corresponding branch. There are fewer branches containing many more steps in the heat shock case compared to the control case.

The p.d.f. for the length of the jump steps also sheds light on the origin of the observed changes. The first moment of Eq. (2.2) is 1.87 for the control case, and 1.91 for the heat shock one. This difference, although small, reflects that less dense branching structures are observed as a smaller contact probability for the heat shock compared to the control case at shorter genomic separations. For larger genomic separations, since the branching structures (spheres) are larger for heat shock, the inter-branch contacts are more likely than in the control case.

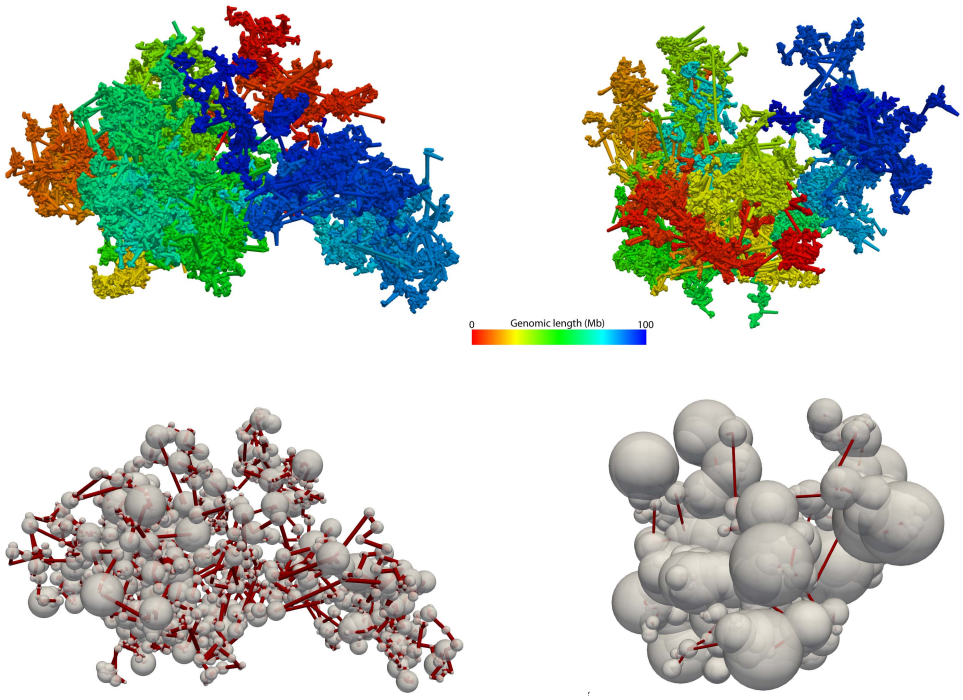

Figure 2.20. (A&B) Representation of a typical conformation for (A) $\alpha = 1.15$ (control) and (B) $\alpha = 1.10$ (heat shock). (C&D) SR-EV backbone in red, and branches represented as grey spheres, the volume of which corresponds to the size of tree domains, for (C) control and (D) heat shock cases.

2.3.5. Conclusions and Discussion

Here, statistical differences between the organization of chromatin under normal conditions and after heat shock have been investigated by experimental methods and compared to a novel theoretical model. The study includes bulk Hi-C, which measures alterations in genome-wide connectivity for millions of cells, and PWS microscopy, which determines differences in nuclear mass density distribution for hundreds to thousands of cells at the single-cell level. The theoretical results were obtained with the novel Self Returning-Excluded Volume (SR-EV) polymer model, where ensemble averages are determined from the aggregate behavior of 1000 conformations.

PWS analysis is based on the assumption that the mass density distribution of chromatin can be represented by a mass fractal model, implying a power-law behavior for the chromatin mass density ACF. The experiments use this assumption to interpret the output of the PWS signal in a more biologically meaningful way, producing a spatial map of chromatin packing scaling, D, for each acquired image. We have performed an analogous computational experiment using the SR-EV model. We found that, in general, individual conformations do not produce a mass density ACF that can be well described by any one power-law coefficient for each conformation, as there is a heterogeneous distribution of packing scaling coefficients at the single-conformation level. However, the ensemble average mass scaling behavior can be described by power-law scaling. Moreover, the histogram of all D values derived from the individual SR-EV conformations produces a similar result as the average nuclear D distribution determined by PWS microscopy experiments. Thus, PWS microscopy and the SR-EV analysis produce a consistent description in relation to the expected chromatin mass density ACF. Importantly,

SR-EV provides a robust explanation for how heterogeneous mass scaling behavior at the single conformation levels can produce ensemble averages that can be described with one chromatin packing scaling exponent. This is an underlying assumption of our interpretation of population-wide PWS microscopy measurements. Additionally, although ChromSTEM/TEM results demonstrate similar mass scaling behavior as observed with SR-EV [**?** 167, 166], there are intrinsic limitations of sample size for these measurements and they are currently unable to reconstruct exact trajectories of chromatin conformation. Thus, unlike ChromSTEM/TEM measurements, we can use SR-EV to determine changes in both statistical mass density distribution and genome connectivity.

The SR-EV depends on a single parameter α. Chromatin under normal conditions is well represented by SR-EV using $\alpha = 1.15$, while chromatin after the heat shock treatment corresponds to $\alpha = 1.10$. The SR-EV model is a heuristic model, aimed to provide a statistical description of chromatin structure. Thus, the choice of these particular values for α is supported by the comparison with the experimental results. Nevertheless, our findings support the idea that α is related to the temperature or thermal history of the system. The SR-EV model allows for a graphical 3D representation of the reorganization of connected chromatin structure triggered by the heat shock treatment, which is unattainable by experiments alone. The branching patterns emerging from the SR-EV backbone are different for each condition. The number of branches is smaller in the case of the heat shock than in the control case, but the size of the branching structures are larger for the heat shock case. The interplay between the effects of excluded volume interactions and the number of elements per branch (e.g., size of tree domains) results in a

higher probability of contacts at shorter genomic separations for the control case, and the opposite for larger genomics distances, all of which were validated by Hi-C experiments.

The fact that the salient chromatin features can be explained by an abstract folding algorithm described by the SR-EV statistics indicates the existence of universal principles of chromatin architecture that are mechanism-independent. Our results suggest a global coupling between different chromatin properties including domain hierarchy, domain size distribution, backbone openness, packing heterogeneity, and genomic interactions. As all of these properties can be modulated with just one folding parameter, α, this indicates a substantial dimensionality reduction of chromatin folding. This folding picture also stresses the importance of understanding the 3D genome from a data structure point of view on top of polymer physics.

The existence of hierarchical tree domains are hypothesized to be the result of a concert of biophysical mechanisms which potentially create branching structures, including DNA supercoiling [152, 202, 51] and CTCF-cohesin-mediated loop extrusion [251, 90]. Future studies employing the SR-EV polymer model could include dynamical simulations which investigate mechanisms contributing to chromatin organization, including the phenomena of phase separation, the influence of architectural proteins that occupy physical space and could constrain positions of domain structures, and the study of crowding-mediated effects of transcription as demonstrated by [184, 288] (see Chapter 3). Such computations could be paired with experiments that perturb specific mechanisms and measure changes in chromatin structure and dynamics to better uncover key mechanisms determining fundamental chromatin structure.

2.3.6. Supplementary Material

2.3.6.1. SR-EV Energy Minimization Procedure Removes Overlaps. There are several potential methods to remove the overlap. Here, we describe one of those possibilities. After generating an SRRW conformation, we perform the following steps:

(1) We create a force field where each point is transformed to a Lenard-Jones particle with $\sigma = 1$ and $\epsilon = 1$.

(2) Each SRRW step is considered as a harmonic bond with equilibrium distance equal to the original SRRW step size.

(3) Each pair of consecutive steps is considered as an angular potential with the equilibrium angle equal to the original angle of the SRRW conformation.

(4) We perform an energy minimization using the new force field and the original SRRW conformation as reference for position restraints.

(5) We stop the minimization once all particles are separated by at least 0.9σ.

2.3.6.2. Supplementary Figures.

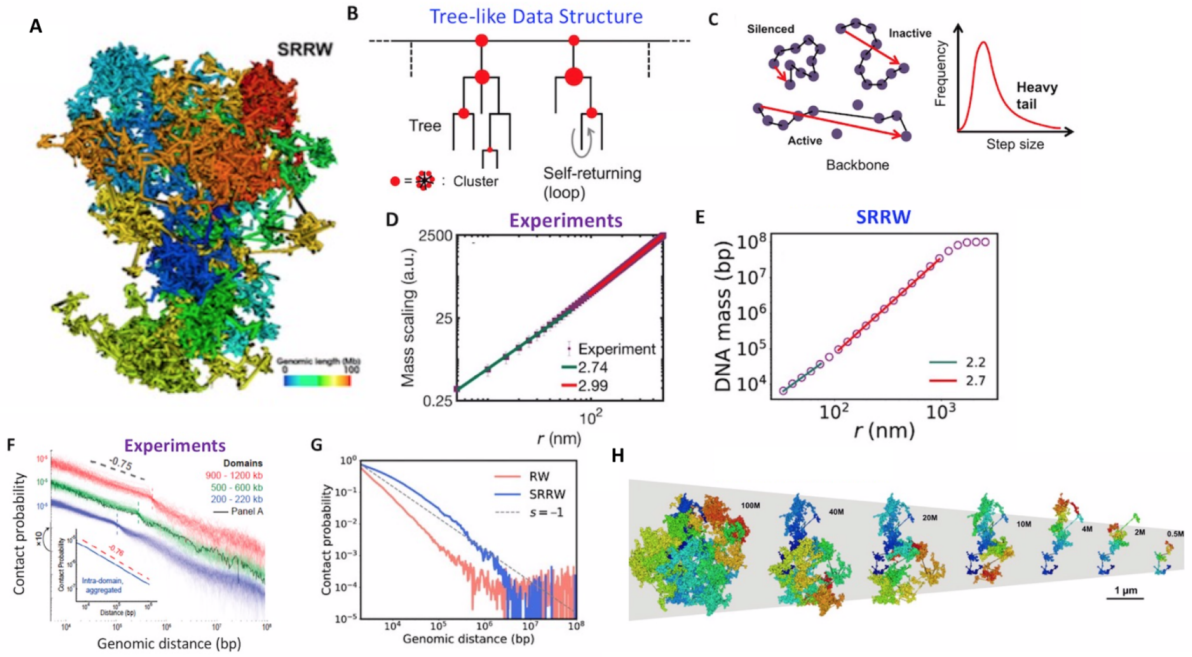

Figure 2.21. Self-Returning Random Walk (SRRW) reproduces key experimental observations of chromatin structure with just one folding parameter. (A) Example rendering of SRRW with different colors representing different genomic regions along the linear sequence. (B) The SRRW's topological architecture featuring random trees connected by an open backbone. Tree nodes are formed by frequent self-returning of short steps. (C) Coarse-graining diverse epigenetic states at the nanoscale into a wide distribution of step sizes. One step approximately maps to 10 nucleosomes, or 2 kbp of DNA. The balls represent histones, and the lines represent DNA. The arrows represent the coarse-grained steps in the SRRW procedure. (D) Chromatin mass scaling as determined by ChromSTEM analysis. a.u., arbitrary units. (E) SRRW mass scaling of the modeled chromatin sampled over 1000 SRRW trajectories. (F) Hi-C experiments from [251] determined that contact probability scaling within TADs has a lower absolute contact probability scaling $|s| < 1$ and then transitions to a higher contact probability scaling at larger length scales. (G) Contact probability scaling of SRRW compared to Random Walk (RW). As a guide to the eye, the dashed line shows power-law scaling, with exponent $s = -1$. The SRRW reproduces the two scaling regimes observed by experiments in (F). (H) Structures of the modeled chromatin at different genomic scales demonstrate a hierarchical organization of SRRW tree domains.

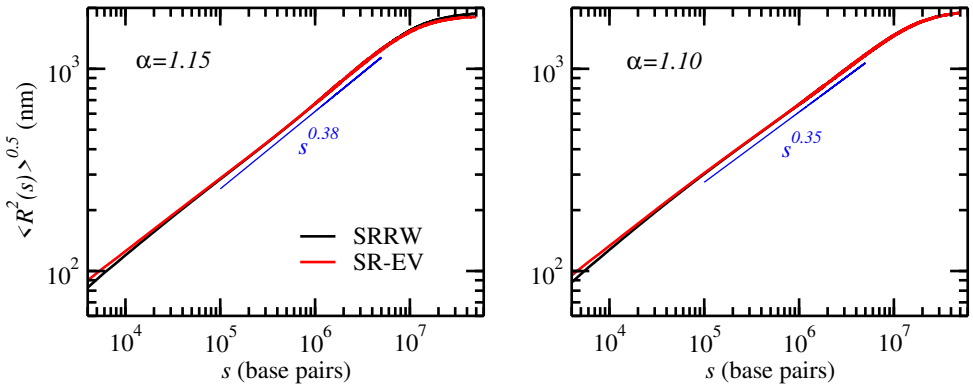

Figure 2.22. Mean quadratic end-to-end distance as a function of the genomic separation s for the SRRW and SR-EV models and for $\alpha = 1.15$ and $\alpha = 1.10$, as indicated. The results are averages over ensembles composed of 1000 individual conformations. The scaling behavior shows a strong effective attraction resulting from the frequent folding of the SR-EV polymer which describes chromatin structure.

CHAPTER 3

Chromatin Packing Behavior Determines Phenotypic Plasticity and Responsiveness to Chemotherapy

3.1. Introduction

Most perturbations a eukaryotic cell experiences occur at non-replicative time scales. These perturbations are remarkably varied, range in intensity, and can be completely distinct from previously encountered stimuli. Examples exist throughout the human body, including within the skin, the alimentary tract, the immune system, the respiratory tract, the reproductive system, and in malignancy. Consider the epithelial lining of the digestive and respiratory systems. While both systems are constantly renewing their lining, the majority of functional cells within these tissues persist for days to weeks after replication. During their lifespan, these cells are exposed to a wide range of nutrients and toxicants that necessitate modification of gene expression to carry on basic cellular functions across these variable conditions, including nutrient absorption, regulating ionic homeostasis, maintaining a sufficient mucosal barrier, excreting waste products, and secreting immunoglobulins.

No better example may exist than malignancy, as tumor cells are remarkably adept at acclimating to a broad spectrum of cytotoxic chemotherapies and radiation exposure, while evading detection from the myriad tools present within the immune system. Furthermore, in the field of carcinogenesis, the distinction between healthy tissue and malignancy becomes blurred as seemingly normal cells acquire cancer-like traits [9, 289, 19]. These capabilities evoke a critical question – how do individual cells acclimate to fluctuating or completely novel conditions? Likewise, how do collections of cells, such as an organ or a tumor mass, acclimate in aggregate to a heterogeneous, rapidly evolving environment?

One widely explored mechanism to respond to such varied conditions is to have a level of predetermined functionalization: intermixing specialized cells within an organ

to carry out specific roles. Beyond establishing pre-coordinated responses, an intriguing possibility is for cells and cell populations to have an encoded level of phenotypic plasticity in order to acclimate to novel conditions in real time [234, 270]. In the context of multicellular systems, the level of phenotypic plasticity encoded would be a product of cellular malleability, the functional responsiveness of cells toward end-stable states upon external stimulation, and the level of intercellular heterogeneity, the diversity of states that are observed within the same population at a given time.

Transcriptional malleability is related to cancer cell survival in response to chemotherapy as well as the functional responsiveness of immune cells to microbes [5]. Recent work using single-cell RNA and DNA sequencing technologies has also demonstrated that transcriptional heterogeneity is characteristic of chemoevasive tumors [38, 272, 263, 158]. Furthermore, the cancer state is associated with considerable structural [147, 23], epigenetic [73, 8], and mutational heterogeneity [157, 73] – all of which have been demonstrated to be independently linked to chemotherapeutic resistance, metastasis, survival and resilience in multiple cancer models. The basis of chemoevasion in cancer chemotherapy models remains poorly understood due to a lack of methods that can image and detect relevant changes in rare chemoevasive subpopulations. Indeed, the efficacy of sequential or combination treatments of some solid tumors with conventional chemotherapies produces diminishing returns for each round of therapy [126]. Newer immunotherapies target a priori identifiable transformations to recruit the patient's own immune surveillance for clearance. However, the likelihood of efficacy for both chemotherapy and immunotherapy depends in part on the intrinsic heterogeneity within the tumor, which itself increases as a result of cytotoxic intervention [187, 188]. Despite the evidence

that this tumor heterogeneity is a key determinant of chemotherapeutic efficacy, there are no existing strategies that definitively decrease intratumor heterogeneity across all tumor types.

Both the malleability and heterogeneity of gene expression within cell populations could result from the physical organization of chromatin. Although aberrant chromatin remodeling across multiple length-scales has been correlated with changes in gene expression profiles related to malignancy (see Chapter 1.2), to date no one aspect of chromatin structure has been able to predict both aspects of phenotypic plasticity that relate to chemoevasion potential. This chapter aims to uncover a genome-wide regulator of phenotypic plasticity that is able to predict responsiveness to cytotoxic stressors, such as chemotherapy treatment in cancer cells. The underlying hypothesis of this work is that large-scale gene expression patterns, which influence phenotypic outcomes, are influenced by the statistical distribution of chromatin packing into PDs (as characterized in Chapter 2) due to crowding-mediated effects. Briefly, macromolecular crowding influences the efficiency and kinetics of transcription reactions (see Chapter 1.4.4) and chromatin density is the major crowder in the nucleus [5, 4, 6].

First, to test the hypothesis that the statistical packing of chromatin structure is a regulator of both transcriptional malleability and intercellular heterogeneity, we developed the chromatin packing macromolecular crowding (CPMC) model. CPMC describes transcription as a series of chemical reactions occurring in a heterogeneous, crowded environment. Pairing the CPMC model with single-cell RNA sequencing (scRNA-seq), ChromTEM and PWS microscopy, we demonstrate that sensitivity of transcriptional

changes to chromatin packing scaling is dependent on three physical descriptors of chromatin PDs: domain size, average chromatin density, and, especially, chromatin packing scaling. Additionally, average D of PDs determines both the level of transcriptional malleability and heterogeneity in cancer cell populations exposed to chemotherapy.

Next, we extend the CPMC model to study the role of statistical chromatin conformation in chemoevasion that is independent of the underlying tumor model. We term this new model the Chromatin-Dependent Adaptability (CDA) model. Here, we show that chromatin packing scaling is tightly paired with adaptability and chemoevasion potential of cancer cells. Leveraging this information and the capability of live-cell PWS microscopy to measure changes in chromatin packing within chromatin PDs, we identify compounds that decrease average chromatin packing scaling of cell populations and increase cancer cell death upon co-treatment with chemotherapy.

3.2. Chromatin Packing Macromolecular Crowding (CPMC) Model of Transcription

The CPMC model considers transcription in dilute, *ex vivo* conditions as a series of diffusion-limited chemical reactions that use DNA, TFs, and RNA Pol II to produce mRNA (Fig. 3.1A). The total production of mRNA in these conditions will depend on the concentration of reactants ($[C]_{tot}$; Fig. 3.1B), the rate of polymerase elongation (k_m; Fig. 3.1C), and the dissociation rates of TFs and Pol II from DNA (K_D: Fig. 3.1D). These molecular factors are well-studied regulators of gene expression *in vitro*. For example, at the scale of nuclear compartments, the formation and dissipation of TADs can alter local TF concentrations [56]. In addition, post-translational histone modifications alter nucleosomal stability, thereby influencing the rate of Pol II elongation [287]. Other post-translational modifications of RNA Pol II itself independently control polymerase activity [48]. Furthermore, gene motifs determine binding affinities of Pol II and TFs, resulting in varied dissociation constants of these molecules from their respective target genes [130].

Compared to *ex vivo* conditions, the eukaryotic nucleus is a highly crowded, heterogeneous environment (Fig. 3.1E). To model transcription reactions within such an environment requires consideration of the length scales involved. At the smallest scale (within ∼20 nm of a gene, i.e., the "transcription interaction volume"), macromolecular crowding (ϕ_{in}) influences transcription by affecting the mobility of transcriptional reactants and the dissociation rate of these molecules from DNA [184, 197, 142]. In addition, the accessible surface area of chromatin determines the number of DNA binding sites available to transcriptional reactants. The probability of a gene promoter to be available for transcription depends on its local accessible surface area. At these small length scales,

transcription can be modeled as a network of chemical reactions involving TFs, RNA Pol II, and DNA. TFs bind to their respective DNA binding sites and recruit RNA Pol II to gene promoters, which, in turn, bind DNA. These series of reactions result in intermediary transcription complexes that stochastically transcribe genes into mRNA. Each reaction coefficient depends on local crowding effects, which can be calculated using BD and MC simulations. Gene expression for particular crowding conditions is calculated by solving the steady-state network of equations that models these transcription reactions [184, 142] (see Chapter 1.4.4).

Notably, the most prevalent macromolecular crowder in the nucleus is chromatin. Thus, local chromatin density within the interaction volume of a gene should have a profound effect on transcription processes. Recent electron microscopy studies have shown that chromatin packing density is highly heterogeneous across the genome. Some genes have interaction volumes with exceedingly high densities (CVC$\sim$0.5), while others may be positioned in regions of the nucleus with CVC as low as $\sim$0.1 to 0.2 [212]. One approach to study the effect of local crowding on transcription in cells would be to experimentally measure the local density of chromatin near every gene using electron microscopy and pair these measurements with *in situ* mRNA levels. This, however, is beyond existing technical capabilities, and an alternate approach is needed.

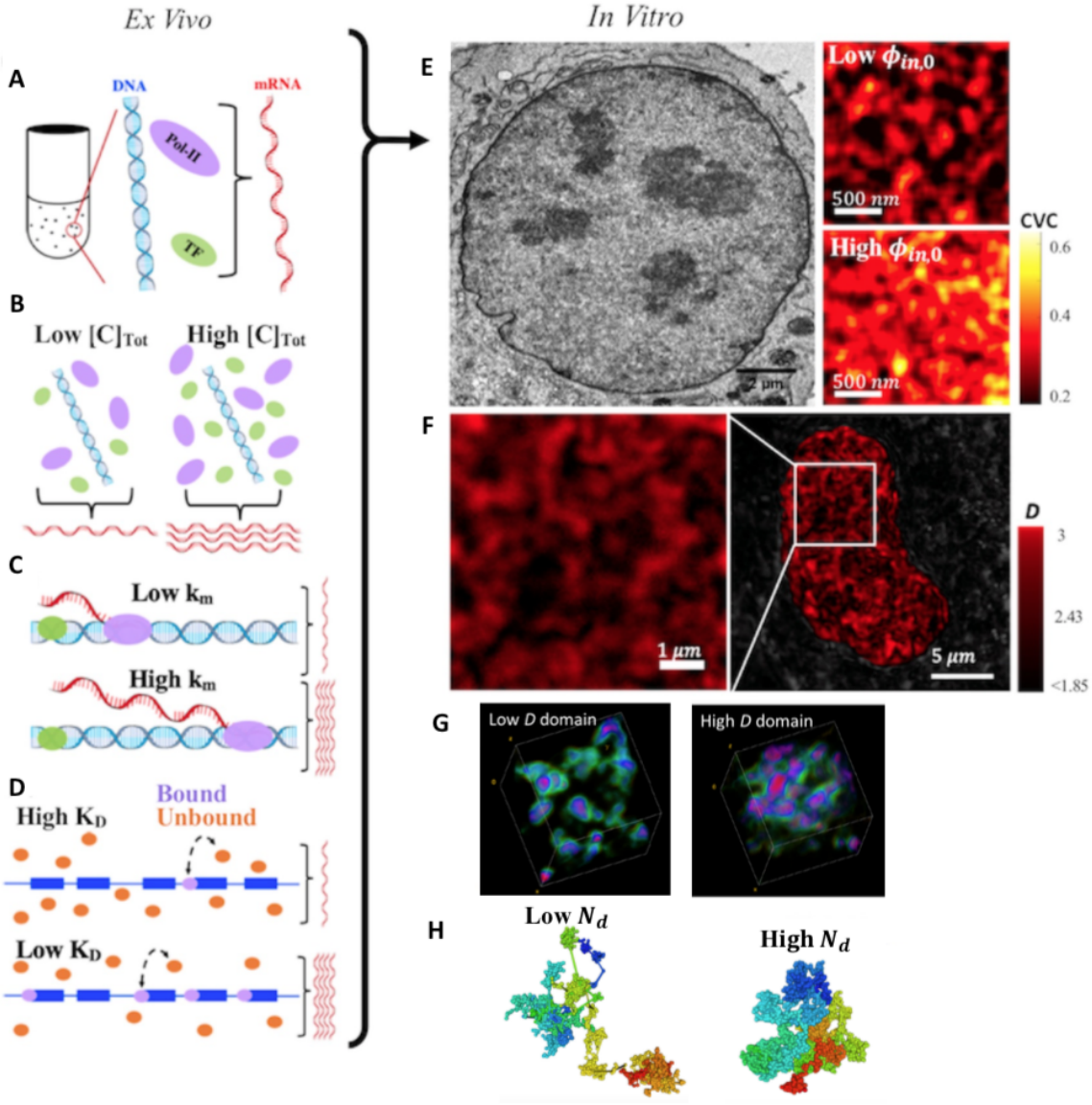
Ex Vivo
In Vitro
A
DNA
mRNA
Pol-II
TF
B
Low [C]Tot
High [C]Tot
C
Low km
High km
D
Bound
Unbound
High KD
Low KD
E
Low $\phi_{in,0}$
500 nm
High $\phi_{in,0}$
500 nm
2 μm
CVC
0.6
0.4
0.2
F
1 μm
5 μm
D
3
2.43
<1.85
G
Low D domain
High D domain
H
Low N_d
High N_d

Figure 3.1. The Chromatin Packing Macromolecular Crowding (CPMC) model integrates molecular and physical regulators of transcription. The regulators influencing transcription reactions can be generally divided into two categories: (A-D) molecular regulators (k_m, K_D, and $[C]_{tot}$) and (E-H) physical regulators (D, $\phi_{in,0}$, and N_d). (A) The CPMC model describes transcription as a series of diffusion-limited chemical reactions. *Ex vivo*, expression depends on (B) concentration of transcriptional reactants $[C]_{tot}$ (TFs (green), RNA Pol II (yellow)), (C) RNA Pol II elongation rate, k_m, and (D) the disassociation rate of RNA Pol II from the transcription start site (TSS) K_D. (E) (Left) In addition to the molecular determinants, transcription reactions are influenced by the highly dense and complex nuclear environment. The concentration of the main crowder within the nucleus, chromatin, can be measured by ChromTEM. As an example, a ChromTEM image of a nucleus from an A549 lung adenocarcinoma cell is shown. (Right) ChromTEM measurements of CVC demonstrates that chromatin density varies throughout the nucleus. (F) Representative PWS image of an A549 cell demonstrating the existence of chromatin PDs as regions of elevated chromatin packing scaling, which vary throughout the nucleus. (G) A PD with a higher D (right) has a more heterogeneous density distribution and a greater accessible surface area compared to a PD with a lower D (left). (h) N_d is the genomic size (in bp) of a chromatin packing domain and can range from less than 100 kbp to several Mbp. PDs are illustrated by color coding with each color representing a separate domain.

Instead of experimentally mapping gene expression to locus-specific crowding conditions, the CPMC model probabilistically samples the polymeric properties of chromatin in order to approximate transcriptional output of an ensemble of genes under similar molecular and varying physical conditions [5, 125, 13]. A combination of molecular factors influences the relative initial expression levels of these genes [184]. In this thesis, we focus on how physical regulators further modulate transcription reactions to produce a final observed transcription rate. The model considers chromatin to be a disordered heteropolymer that is heterogeneously packed in three-dimensional space. The statistical packing of the chromatin polymer determines the volume fraction occupied by chromatin

(CVC, the number of nucleotides acting together as a grouped polymeric entity (N_d), and the packing scaling behavior (D) of these polymeric entities. N_d can be considered as the number of nucleotides that are contained within a subset of the chromatin polymer that has self-similar, power-law scaling properties. As described in Chapter 1.4.1, the power-law scaling behavior describes the relationship between the length of a given segment of the chromatin polymer (e.g. the number of nucleotides, N) and the size (r) of the physical space occupied by the segment, $N \propto r^D$ for $N \leq N_d$.

We have been able to visualize the existence of PDs with power-law mass scaling behavior using ChromSTEM (Chapter 2.2, Fig. 3.1E) and PWS (Chapter 2.2, Fig. 3.1F) which are ∼100-200 nm in diameter with genomic sizes on the order of 100s kbp. PDs are visualized as globular regions of higher D. The CPMC model considers a gene's interaction volume to be located within these PDs. Accordingly, the local environment of a gene's interaction volume is determined by the encompassing PD, each of which may have its own average nuclear crowding density ($\phi_{in,0}$) (Fig. 3.1E), chromatin packing scaling D (Fig. 3.1F&G), and genomic size (N_d) (Fig. 3.1H). In addition, gene length (L) partially influences the size of the interaction volume of a given gene, affecting the range of crowding conditions the gene is probabilistically exposed to. The CPMC model employs these physical regulators of chromatin, measurable by experimental nanoimaging techniques, to approximate distributions of mass density and accessibility of chromatin, all to determine transcription for each gene throughout the entire nucleus, a feat which is currently experimentally infeasible [6].

The expected expression rate of a gene *in vitro* is the product of the steady-state mRNA transcription rate of that gene (ϵ) and the probability of the gene to be on the

accessible surface of the chromatin polymer (p_g). Steady-state expression rate is a function of molecular features surrounding the gene of interest ($\vec{m}$; e.g., concentrations of RNA Pol II, TFs, and promoters) (Fig. 3.1B-D) in the context of local physical conditions (Fig. 3.1E-H) [5, 143, 184, 197]. The probability of gene accessibility contributes to the likelihood of a gene to interact with transcriptional reactants (TFs and RNA Pol II) *in vitro* [86]. It is beyond technical capabilities to measure all molecular and physical parameters of the model for specific genes at the single-cell level. Thus, we explore how a given ensemble of genes with similar molecular features $\vec{m}$ grouped by their initial expression or associated with similar biological pathways as defined by gene ontologies would respond to changes in average measurable physical conditions.

Specifically, we study how average nuclear crowding density, $\phi_{in,0}$, average chromatin packing scaling, D, and genomic size of a PD, N_d, change the behavior of global transcription processes. It is critical to stress that the CPMC model does not assume that the chromatin polymer has the same power-law scaling behavior or constant density throughout the entire nucleus, but that using population averages is instead an approximation due to existing experimental limitations. The model can further be extended to consider each PD has its own chromatin packing scaling D as technological capabilities to co-register chromatin packing, molecular, and genomic properties advance (see Chapter 3.5 for an experimental analysis of how local packing behavior influences gene expression). Finally, in this model, nuclear crowding density within each interaction volume, ϕ_{in}, is assumed to be constant relative to the time-scale of transcription, which occurs on the order of minutes, in line with recent live-cell PWS imaging studies of chromatin mobility [97].

Given these considerations, in a population of cells, each gene will be exposed to different crowding densities ϕ_{in}. Each ϕ_{in} will be sampled from the probability distribution function $f(\phi_{in})$, which is assumed to follow a normal distribution with mean $\phi_{in,0}$ and variance $\sigma^2_{\phi_{in}} \approx \phi_{in,0}\big(1-\phi_{in,0}\big)\big(r_{min}/r_{in}\big)^{3-D}$ where r_{min} is the radius of the elementary unit of chromatin (e.g., one DNA bp) and r_{in} is the radius of the transcription interaction volume [5]. Due to the mass-fractal nature of chromatin, $r_{in} = r^0_{in} + L^{1/D} r_{min}$ for a gene of length L, where r^0_{in} is the radius of the interaction volume for a single base pair and is approximated from previous MC simulations of crowding effects [5, 184]. Thus, the expected range of crowding densities each gene is exposed to is dependent on the statistical properties of the PD where the gene is located, including D and $\phi_{in,0}$, and is further influenced by the length L of the gene.

The transcription rate ϵ itself is assumed to depend on molecular features $\vec{m}$ as well as on local crowding density ϕ_{in} surrounding the gene. We calculate all expression rates under the assumption that molecular features $\vec{m}$ remain constant throughout the population, with physiologically relevant values used in previous MC and BD crowding simulations (Table 3.1) [184]. This gives rise to the form of $\bar{\epsilon}$, the average expression rate for an ensemble of genes that share a given $\vec{m}$ as:

$$\bar{\epsilon} = \int \epsilon\big(\vec{m}, \phi_{in}\big) f\big(\phi_{in}\big) d\phi_{in} \tag{3.1}$$

Likewise, a fractal model of chromatin PDs allows the CPMC model to calculate the probability of a unit of DNA, such as a gene promoter, to be on the accessible surface

of chromatin, p_g [125]:

$$p_g = N_d^{-1/D} \tag{3.2}$$

Finally, merging accessibility with steady-state expression rate for a group of genes with similar molecular regulators of transcription, the ensemble expression rate is:

$$E = \bar{\epsilon} \cdot p_g \tag{3.3}$$

To quantitatively analyze the effect of D on gene expression, and compare CPMC model predictions with experiments, we calculate the sensitivity of gene expression as a function of D. Sensitivity (Se) is the measurement of how a dependent variable, such as gene expression, will change as a function of a perturbation to an independent variable, such as D. In Chapter 3.8.2, we determine that Se_D is much more sensitive to changes in D and molecular factors compared to both Se_{N_d} and $Se_{\phi_{in,0}}$ by at least one order of magnitude. Thus, the rest of the study will focus on determining cell death solely using Se_D, which we will refer to as Se.

Se of expression rate for any group of genes to changes in D is defined as:

$$Se = \frac{\partial ln(E)}{\partial ln(D)}\Big|_{E=E_i, D=D_i} \tag{3.4}$$

where E_i is the initial average expression rate of the group of genes sharing similar molecular features $\vec{m}$ and gene length L, and D_i is the initial average chromatin packing scaling before external perturbation.

A positive Se for a given group of genes indicates that an increase in D, on average, enhances their collective expression rate. Importantly, the CPMC model predicts the output of transcription reactions that occur within the nucleus. Assuming that the half-life of mRNA transcripts is dictated by cytoplasmic conditions, structural changes in chromatin that alter D are not considered to alter the degradation rate of mRNA. Thus, Se should be directly related to the number of transcripts produced for any group of genes in the nucleus.

To solve Eq. 3.4, we utilized a Taylor series approximation of $\bar{\epsilon}$ around $\phi_{in,0}$:

$$\bar{\epsilon} \approx \epsilon\big(\vec{m}, \phi_{in,0}\big) + \frac{1}{2}\sigma^2_{\phi_{in}} \frac{\partial^2 \epsilon\big(\vec{m}, \phi_{in}\big)}{\partial^2 \phi_{in}}\Big|_{\phi_{in}=\phi_{in,0}} \tag{3.5}$$

where $\bar{\epsilon} \approx \epsilon\big(\vec{m}, \phi_{in,0}\big)$ is a non-monotonic function of ϕ_{in} due the competing effects of crowding on depletion interactions and molecular diffusion, and $\frac{\partial^2 \epsilon\left(\vec{m},\phi_{in}\right)}{\partial^2 \phi_{in}}\Big|_{\phi_{in}=\phi_{in,0}} \approx -\sqrt{\epsilon\big(\vec{m}, \phi_{in,0}\big)\kappa}$ quantifies gene expression as a function of crowding within a transcription interaction volume. Expression rate κ is derived from a steady-state solution of rate equations that model transcription and whose crowding-dependent rates were determined from BD and MC simulations as described previously [5]. Although, in principle, the exact form of $\frac{\partial^2 \epsilon\left(\vec{m},\phi_{in}\right)}{\partial^2 \phi_{in}}\Big|_{\phi_{in}=\phi_{in,0}}$ as a function of $\epsilon\big(\vec{m}, \phi_{in,0}\big)$ may depend on which component of $\vec{m}$ is being varied, i.e. $\kappa = \kappa(\vec{m})$, in practice κ is only weakly dependent on $\vec{m}$. In other words, $\frac{\partial^2 \epsilon\left(\vec{m},\phi_{in}\right)}{\partial^2 \phi_{in}}\Big|_{\phi_{in}=\phi_{in,0}}$ depends on $\vec{m}$ primarily through $\epsilon\big(\vec{m}, \phi_{in,0}\big)$, with the average expression rate as the "common dominator" of multiple molecular factors. Thus, predictions of the CPMC model regarding the effects of physical regulators on ensemble gene expression should be robust to changes in molecular factors.

Combining Eq. 3.1,3.2,3.3,3.4,3.5 the Se of expression rate becomes:

(3.6)

$$Se \approx \frac{1}{D_i} lnN_d - \frac{\kappa}{8\bar{\epsilon}}(\sigma^2_{\phi_{in}})^2 \left(1 + \sqrt{1 + \frac{16}{(\sigma^2_{\phi_{in}})^2}\frac{\bar{\epsilon}}{\kappa}}\right)\left[D_i ln\left(\frac{r_{in}}{r_{min}}\right) + \frac{3 - D_i}{D_i}\frac{r_{min}}{r_{in}} L^{1/D_i} lnL\right]$$

3.3. Statistical Descriptors of Chromatin PDs Regulate Sensitivity of Gene Expression to Chromatin Packing Scaling

To first test the CPMC model predictions *in vitro*, we employed live-cell PWS microscopy to measure D (Fig. 3.2A&B) and ChromTEM to measure $\phi_{in,0}$ (Fig. 3.2C&D) paired with mRNA microarrays, RNA-Seq, and single-cell RNA sequencing (scRNA-seq) to measure gene expression of cell populations under different conditions. Specifically, average D of cell populations was calculated by first averaging D values from PWS measurements within each cell nucleus and then averaging these measurements over the entire cell population for each treatment condition. Utilizing ChromTEM, average chromatin density was measured within each nucleus. As $\phi_{in,0}$ represents the crowding contributions from all nuclear macromolecules, we added to CVC measured by ChromTEM an additional 0.05 volume concentration contribution to account for nuclear mobile crowders. In addition, we utilized publicly available DNA sequencing information to obtain gene length and Hi-C data to approximate N_d from TAD sizes. While N_d might not necessarily represent the organization observed in TADs (see Chapter 2.2.6), TAD size was utilized as an approximate measure of N_d as these domains have been shown to obey power-law scaling for contact probability [251]. Combining these methods, we then tested the CPMC model's predictions of Se of gene expression against *in vitro* measurements for each identified physical regulator of gene expression.

To test the role of initial D_i, we performed an RNAi knockdown of the chromatin remodeling enzyme, Arid-1a (A-KD) in human colon carcinoma HT-29 cells, which resulted in a lower D_i compared to wild-type (WT) cells [6]. Next, we measured changes

in D in serum starved WT and A-KD HT-29 cells before and 30 minutes after stimulation with 10% fetal bovine serum, 100 nM epidermal growth factor (EGF), and 100 nM phorbol 12-myristate 13-acetate (PMA) [5]. In parallel, we measured gene expression for these conditions at 5 hours utilizing mRNA microarrays. Genes were grouped for WT and A-KD cells separately based on their relative initial expression during serum starvation, and the experimentally measured sensitivity $\Delta ln(E)/\Delta ln(D)$ was calculated for each group of genes. Here, $\phi_{in,0}$ $\sim$0.39 was approximated by dividing chromosome copy number by nuclear volume. As predicted by the CPMC model, experimental measurements of the Se of gene expression show a transcriptional divergence due to D as a function of initial expression in HT-29 cells. Here, transcriptional divergence denotes the difference between the upregulation of highly-expressed genes and the downregulation of lowly-expressed genes. In addition, we found that D_i predominantly changes the responsiveness of initially lowly-expressed genes (Fig. 3.2E&F).

These results indicate that populations of cells with a higher D would have a higher level of transcriptional divergence than low-D cells. Cancer cells across a variety of malignancies have been shown to have elevated chromatin packing scaling as biomarkers of early precancerous changes [248, 98, 247, 59] and as predictors of cancer progression [134, 246]. Functionally, this suggests that D can act as a means to optimize transcriptional response as is explored in subsequent sections.

Next, we tested the effect of average nuclear crowding density, $\phi_{in,0}$, on gene expression sensitivity to changes in D. ChromTEM was employed to measure average chromatin density for both human lung adenocarcinoma A549 cells and differentiated BJ fibroblast cells, which had mean CVCs of 0.35 and 0.30, respectively (Fig. 3.2C&D, distribution

of CVC values for each nucleus are shown in Fig. 3.10). Approximating for additional crowding effects induced by non-chromatin mobile crowders, estimates of $\phi_{in,0}$ were 0.40 in A549 and 0.35 in BJ cells. Each cell line was treated with 100 nM dexamethasone (DXM) to modulate D, which was measured by PWS microscopy. Gene expression of both cell lines with and without DXM treatment was measured by RNA-seq. Sensitivity of gene expression was measured as described above for each cell line. Interestingly, the CPMC model predicts cells with a lower $\phi_{in,0}$ have an attenuated bidirectional Se, an effect confirmed experimentally in the lower chromatin density BJ cells (Fig. 3.2G). In contrast, the higher chromatin density A549 cells (Fig. 3.2H) have a higher transcriptional divergence.

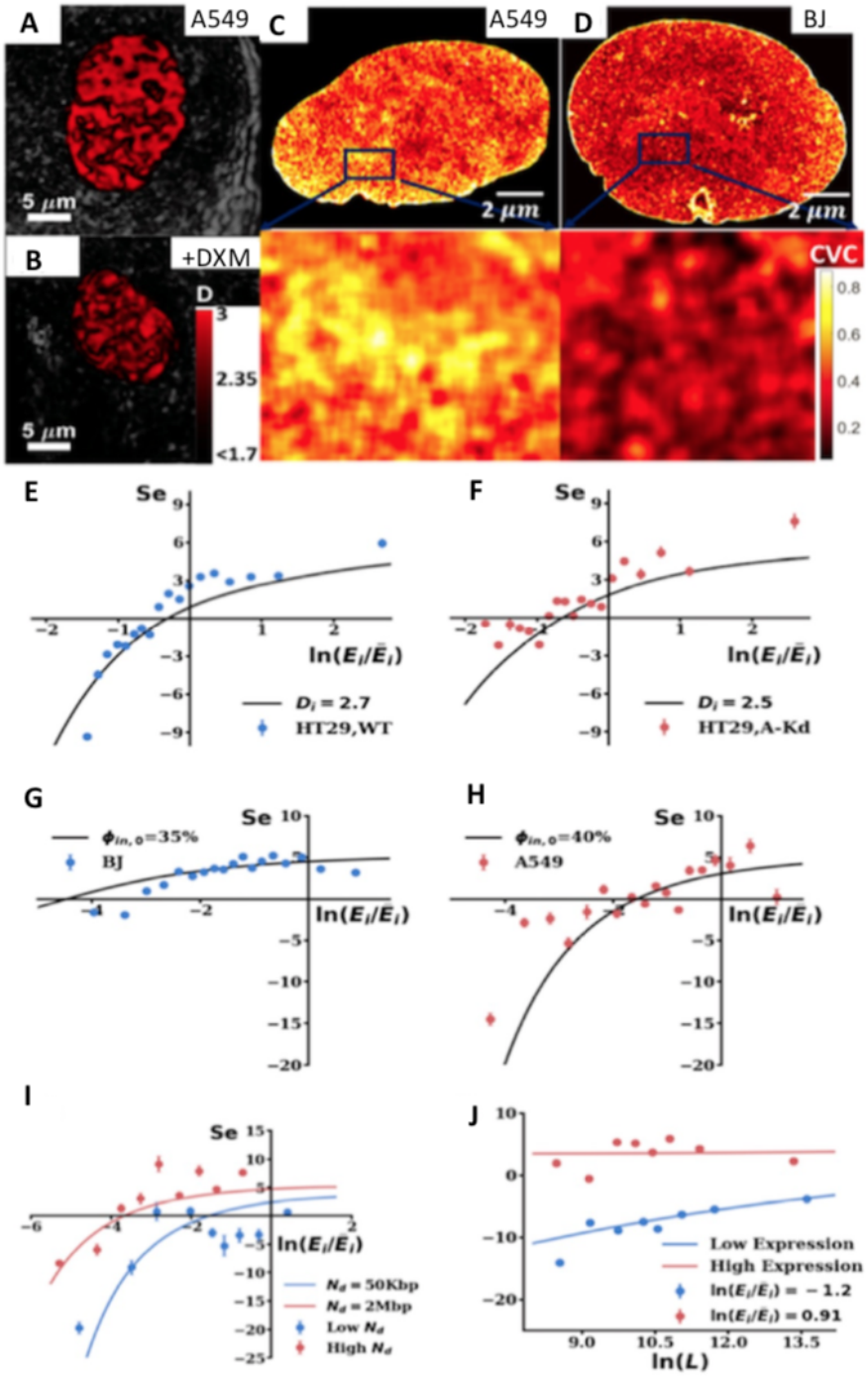
A
A549
5 μm
B
+DXM
D
3
2.35
<1.7
5 μm
C
A549
2 μm
D
BJ
2 μm
CVC
0.8
0.6
0.4
0.2
E
Se
ln(Ei/Ēi)
Di = 2.7
HT29,WT
F
Se
ln(Ei/Ēi)
Di = 2.5
HT29,A-Kd
G
Se
ϕin,0=35%
BJ
ln(Ei/Ei)
H
Se
ϕin,0=40%
A549
ln(Ei/Ei)
I
Se
ln(Ei/Ei)
Nd = 50Kbp
Nd = 2Mbp
Low Nd
High Nd
J
Low Expression
High Expression
ln(Ei/Ei) = −1.2
ln(Ei/Ei) = 0.91
ln(L)

Figure 3.2. Comparison of the CPMC model with experimental measurements of gene expression as a function of physical regulators D_i, N_d, $\phi_{in,0}$, and gene length L. (A&B) Representative live-cell PWS microscopy images of nuclear D distributions scaled between 2.56 and 2.66 for (A) control and (B) 12 hour dexamethasone-treated lung adenocarcinoma A549 cells. Brighter red corresponds to higher D regions. (C&D) Representative heat maps of CVC values from analysis of ChromTEM images of cell nuclei from (C) A549 cancer cells and (D) BJ fibroblasts. Representative magnified regions from each nucleus demonstrate average CVC=0.35 in A549 cell compared to 0.30 in BJ cells. (E-J) Comparison between the CPMC model (solid lines) and experimentally measured (points) sensitivity of gene expression to D (Se, y-axis) as a function of the initial relative gene expression ($ln(E_i/\bar{E}_i)$, x-axis). (E) Cell populations with a higher initial $D_i = 2.7$ (wild-type HT-29 cells) have a bidirectional Se curve that becomes attenuated if D_i is lowered to 2.5 (shRNA knockdown Arid-1a HT-29 cells) (F). Each point represents the average of 100 genes. Changes in D were induced by cell treatment with 10% fetal bovine serum, 100 nM epidermal growth factor (EGF), and 100 nM phorbol 12-myristate 13-acetate (PMA). The CPMC model was able to explain 86% of the variance of the experimental data for wild-type HT-29 cells and 51% of the variance for Arid-1a HT-29 cells. (G) Se in cells with a lower average nuclear density (BJ cells, $\phi_{in,0} = 0.35$; each point corresponds to 300 genes; explained variance (EV) = 59%) is attenuated in comparison to that of cells with a higher nuclear density (H) (A549 cells; $\phi_{in,0} = 0.40$; 100 genes per point; EV = 74%). (I) Genes located within larger domains (N_d ~2 Mbp, 12 genes/point, EV = 56%) have a lower initial expression, but have a positive Se to changes in D in comparison to genes localized within smaller domains (N_d ~50 kbp, 12 genes/point, EV = 37%). The change in D was induced in A549 cells by treatment with 100 nM of dexamethasone. N_d was approximated based on the corresponding TAD size: 2 Mbp TADs for the high N_d group of genes vs. 50 kbp TADs for the low N_d genes. TAD size was measured using the Arrowhead function from the Juicer Tools to analyze Hi-C data [72]. (J) Comparison between the CPMC model (solid line) with experimental results (points, 60 genes/point) in HT-29 cells showing the effect of gene length, L, (x-axis) on Se (y-axis). In agreement with the model, shorter, initially lowly-expressed genes (blue curve, points, EV = 67%) are disproportionally repressed by an incremental increase in D compared to longer genes (high expression, red curve, points). Error bars represent standard error from 4 biological replicates.

This suggests that cells with smaller nuclear volume, such as immune cells, or cells with increased chromosome copy number, such as malignant cancer cells, would be predisposed to produce a more pronounced bidirectional response in gene expression to stimuli compared to cells with lower chromatin density. These results demonstrate that the gene expression sensitivity to chromatin packing scaling for cells with higher $\phi_{in,0}$ results in an increased transcriptional divergence between initially highly- and lowly-expressed genes.

We then tested the roles of domain size, N_d on Se. From our model, N_d determines the probability of genes being on an exposed surface to allow transcription reactions to occur, a relationship which depends non-linearly on D (Eq. 3.2). Consequently, the CPMC model predicts that (1) genes in larger packing domains (e.g., $N_d > 2$ Mbp) would be relatively under-expressed in comparison to those within smaller N_d domains (e.g., $N_d < 50$ kbp) and (2) genes within larger N_d domains would be more likely to become enhanced as a function of increasing D ($+Se$). To test these predictions experimentally, we utilized the Arrowhead function in Juicer tools to measure TAD sizes from Hi-C data of untreated and DXM treated A549 cells [67]. As the dissociation and formation of TADs has previously been shown to alter gene expression, for our analysis we only selected TADs that were unaltered with DXM treatment. The top 20% largest ($\sim$2 Mbp) and bottom 30% smallest ($\sim$50 kbp) of these TADs were chosen to produce gene groups with roughly equal sizes ($\sim$130 genes in each group). Using RNA-seq to measure gene expression and PWS microscopy to measure the change in D before and after DXM treatment, we analyzed the sensitivity of expression of genes localized to smaller 50 kbp TADs compared to those localized in larger 2 Mbp TADs (Fig. 3.2I). As predicted from the CPMC

model, *in vitro* results demonstrate that genes within larger TADs have an overall higher sensitivity to changes in D (Fig. 3.2I), while simultaneously having lower initial expression compared to those within smaller TADs. Consequently, these findings suggest a regulatory role of spatially confining genes into structures with self-similar statistical organization, such as those found in TADs, in determining the probability of a gene being accessible to transcription processes. Given that there exists significant intercellular variability in TADs [82], this would suggest yet another mechanism that cells can use to regulate their functional diversity within a population.

Finally, we tested the role of gene length, L, on the sensitivity of two fold lowly-expressed (low) and two fold highly-expressed (high) genes in the serum starved WT HT-29 cells described above. Using the built-in Mathematica function, GenomeData, to obtain sequence length of genes, the sensitivity of gene expression to D was then calculated as a function of their length. The model predicts shorter genes have a smaller interaction volume, increasing the variance of crowding conditions these genes are exposed to. Consequently, an increase in D should further increase fluctuations in crowding concentrations surrounding these shorter genes, causing initially lowly-expressed genes to further reduce their expression in proportion to decreasing L. However, genes with an initially higher expression level will be relatively unaffected by changes in L due to more optimal molecular characteristics (e.g., high TF and RNA Pol II concentrations) and initial crowding conditions these genes are exposed to. In line with the CPMC model, our experimental microarray data demonstrates that shorter, initially lowly-expressed genes become disproportionately downregulated as a function of increasing D, whereas length minimally influences initially highly-expressed genes (Fig. 3.2J).

3.4. Chromatin Packing Scaling Regulates Phenotypic Plasticity

A major implication of the CPMC model is the role physical chromatin structure plays in shaping gene expression. Thus, the model could provide a mechanistic link between two aspects of phenotypic plasticity of cell populations: transcriptional malleability and intercellular transcriptional heterogeneity. In this case, we can consider transcriptional malleability to be the average change in expression of a gene in response to an external stimulus, while transcriptional heterogeneity can be thought of as the range in expression levels of each gene across a cell population. While there is likely to be increased complexity that results from cell to cell variations in average chromatin density and D, here we test how heterogeneity and malleability are influenced by average chromatin packing scaling within cell populations. An ideal testbed for this mechanistic integration is cancer. Multiple lines of evidence have shown that chromatin structure is nearly universally transformed in malignancy [263, 45, 148] (see Chapter 1.2). Since (1) elevated D is a hallmark of malignancy [23], (2) there is an emergent role of intercellular heterogeneity in determining chemotherapeutic responsiveness and (3) cancer cells rapidly alter their gene expression to overcome cytotoxic stressors [214], we hypothesized that cancer cells could leverage physical transformation within the nucleus to gain survival advantages. Therefore, we wanted to test if cells could utilize chromatin packing scaling as a regulator of both transcriptional malleability and heterogeneity to achieve a rapid response to external stressors.

3.4.1. Chemotherapy Treatment Selects for Cells with Increased Chromatin Packing Scaling

We first tested whether chemotherapy treatment of cancer cells resulted in a pre-selection of high D cells. Using PWS, we measured changes in D for A2780 cells before and after chemotherapy treatment with 5 nM paclitaxel (PAC) for 48 hours. We also monitored cell coverage, which represents survival of a cell population. Here, we define high-D cells as those that fall within the top 25^{th} percentile of D in the A280 cell population prior to paclitaxel treatment. We then measured the percentage of high-D cells at 48 hours after paclitaxel treatment and observed an increase in the percentage of high-D cells in paclitaxel-treated cells compared to the control population (Fig. 3.3B). In combination with coverage measurements, which demonstrated significant cell death after 48 hours of paclitaxel treatment, our results indicate that high-D cells have an increased survival rate when exposed to chemotherapy (Fig. 3.3B&C).

To ensure that this trend can be extended beyond ovarian cancer and taxel treatment, we then performed additional PWS analysis. Over the course of a 48 hour chemotherapeutic intervention, the average population D of colon cancer HCT116 cells treated with oxaliplatin increased, while D in control cells remained at a similar level (Fig. 3.11A). We also observed an increase in D upon chemotherapy treatment in multiple cancer cell lines with three classes of chemotherapy drugs: DNA intercalators (oxaliplatin), microtubule assembly inhibitors (paclitaxel and docetaxel), and nucleotide analogs (5-fluorouracil and gemcitabine) (Fig. 3.11C&D). We additionally tested two paired cell line models previously described to confer chemotherapeutic resistance by different mechanisms and determined that the high D state is maintained once a cancer cell line has

become stably chemoresistant (see Chapter 3.8.4). Overall, this demonstrates that a higher chromatin packing scaling allows cancer cells to better evade a stressor across multiple cell lines.

However, these population level studies did not elucidate whether this increase in D was due to preferential killing of cells with low-D or a transformation of low-D cells to a high-D state. To answer ths, we tracked clusters of HCT116 through a 48 hour oxaliplatin treatment to determine how initial D changes due to chemotherapy. We observe a clear trend in that most of the surviving clusters had a higher D prior to treatment and that cell clusters with initially lower D that evaded chemotherapy experienced the largest overall increase in D from chemotherapy treatment (Fig. 3.11B). Thus, in order to survive chemotherapeutic stress, cells must increase their average D.

3.4.2. Transcriptional Divergence and Malleability

According to the CPMC model, the dependence of transcriptional malleability on D results from the observed asymmetric response of upregulated and downregulated genes to changes in D (Fig. 3.2), which we denote as transcriptional divergence. Here, we focus on changes in gene expression caused by an external stimulus, specifically chemotherapeutic stress. Chemotherapeutic induction of apoptosis has been shown to depend on the rate of change in expression of critical genes (e.g., p53) and not their steady-state levels alone [214]. Accordingly, mechanisms which increase the rate of upregulation of these critical genes would facilitate the development of cellular resilience to stressors. Consider two populations of cells that have a baseline difference in their average initial D, D_i. These two populations are then exposed to the same exogenous stressor and a series of stress

signaling pathways are activated in an attempt to overcome the perturbation. The cells' survival now depends, in part, on the increased expression of these genes within a critical time frame. The CPMC model predicts that the population of cells with initially higher D will be more likely to upregulate these critical genes (Fig. 3.3A).

To quantify the effect of initial D on transcriptional responsiveness, let $mRNA_{1,a}$ be the initial expression (i.e., the number of mRNA transcripts) for a given gene in cell a with initial chromatin packing state D_a. At time point $t = 0$, a stimulus produces an increase in the gene's rate of expression from $E_{1,a}$ to $E_{2,a}$. Without loss of generality, we assume that the expression rate $E_{2,a}$ remains stable over time. In other words that a steady-state expression has been reached poststimulation. For a non-steady-state treatment see Chapter 3.6. We also assume that the rate of cytoplasmic mRNA degradation, ν, remains constant after stimulation. The relative change in expression at time t is $\big(mRNA_a(t) - mRNA_{1,a}\big)/mRNA_{1,a} = \big(E_{2,a}/E_{1,a} - 1\big)\big(1 - e^{-\nu t}\big)$, where $mRNA_{1,a} = E_{1,a}/\nu$ is the pre-stimulation steady-state expression. This relative change in expression increases with the ratio $E_{2,a}/E_{1,a}$, which is itself a function of both molecular features and the chromatin packing state surrounding the gene. This can be illustrated by comparing the response of an individual gene to an exogenous stressor in two cells, a and b. Let the same gene in both cells be associated with similar molecular features $[\vec{m}_{i,a} = \vec{m}_{i,b}, i = 1, 2]$, but different chromatin packing states D_a and D_b, with $D_b > D_a$. From Eq.3.4, $\frac{dE}{E} = \frac{Se(D)}{D}dD$, it follows that:

$$E_{i,b} = E_{i,a}\, exp\left[\int_{D_a}^{D_b} \frac{Se_i(D')}{D'}dD'\right],\ i = 1, 2 \tag{3.7}$$

where $Se_i(D)$ is the sensitivity of expression state $E_{i,a}$.

Under these conditions, the effect of D on relative changes in transcription in cell b compared to cell a would be defined as:

$$\delta = \frac{\left(E_{2,b}/E_{1,b}\right)}{\left(E_{2,a}/E_{1,a}\right)} = exp\left[\int_{D_a}^{D_b} \frac{Se_2(D') - Se_1(D')}{D'} dD'\right] \tag{3.8}$$

Within the physiological range of transcription, Se is an increasing function of E (Fig. 3.2) and, as $E_2 > E_1$ for both cells, $\delta > 1$. Consequently, the same stimulus will result in enhanced upregulation of the same gene in cell b with higher D_b compared to cell a, driven by the differences in chromatin packing scaling between the two cells. This effect is expected to be particularly pronounced for initially lowly-expressed genes with $Se_1 < 0$ that undergo a significant amplification ($Se_2 > 0$) upon stimulation. We see that δ is directly related to the transcriptional divergence and the shape of the function $Se(E)$ (Fig. 3.2). A faster rise of Se as a function of E results in a higher transcriptional malleability, δ. For cells a and b with small enough differences between D_a and D_b, $\delta = 1 + (Se_2 - Se_1)(D_b - D_a)/D_a$. This implies that factors that tend to increase transcriptional divergence, including higher D, higher chromatin density $\phi_{in,0}$, and smaller N_d, would be expected to result in a higher transcriptional malleability.

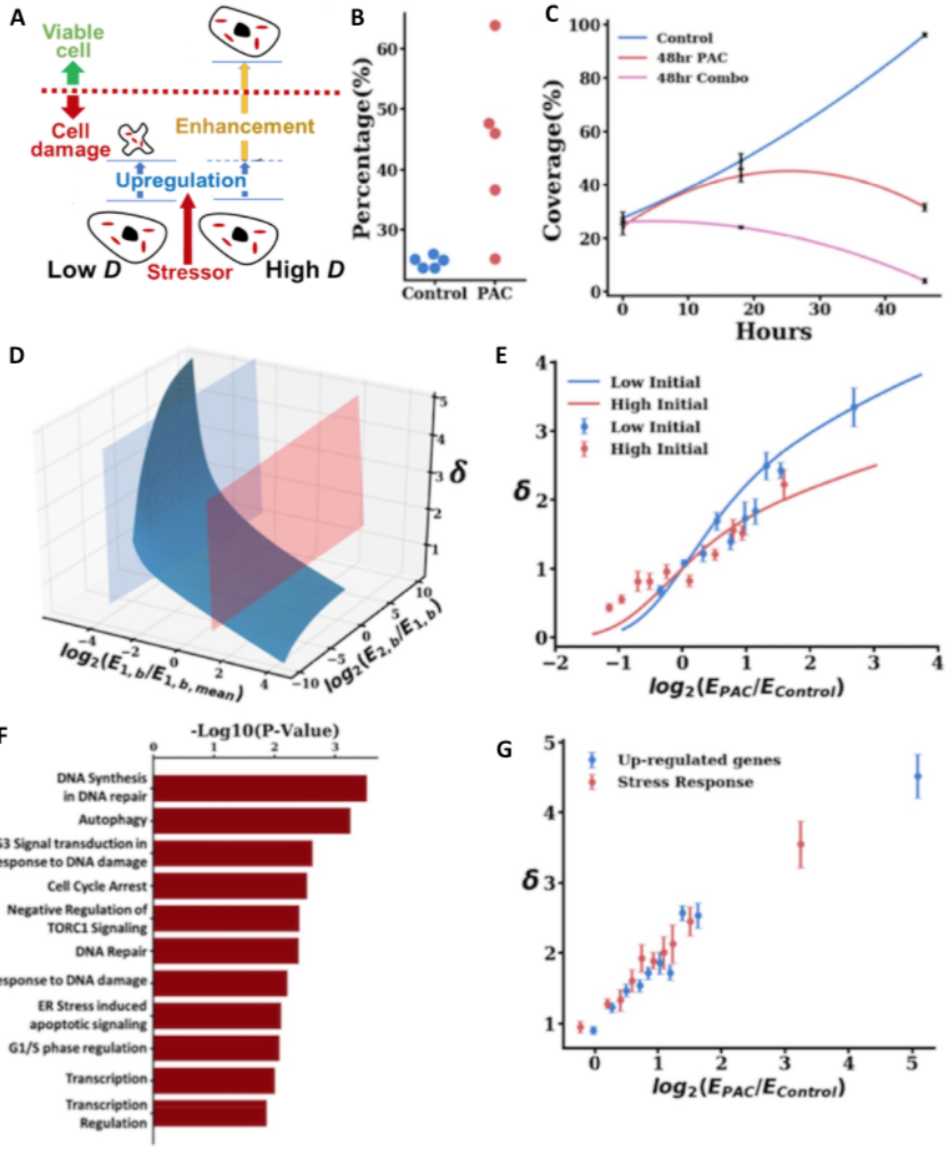
A
Viable cell
Cell damage
Enhancement
Upregulation
Low D
Stressor
High D
B
Percentage(%)
Control
PAC
C
Control
48hr PAC
48hr Combo
Coverage(%)
Hours
D
δ
log2(E1, b/E1, b, mean)
log2(E2, b/E1, b)
E
Low Initial
High Initial
Low Initial
High Initial
δ
log2(EPAC/EControl)
F
-Log10(P-Value)
DNA Synthesis in DNA repair
Autophagy
P53 Signal transduction in response to DNA damage
Cell Cycle Arrest
Negative Regulation of TORC1 Signaling
DNA Repair
Response to DNA damage
ER Stress induced apoptotic signaling
G1/S phase regulation
Transcription
Transcription Regulation
G
Up-regulated genes
Stress Response
δ
log2(EPAC/EControl)

Figure 3.3. Chromatin packing scaling increases the transcriptional malleability of cancer cells. (A) In response to a stressor, such as a chemotherapeutic agent, cells with a higher level of transcriptional malleability may have the ability to respond faster, which may lead to an increased probability of survival. Cells with higher average D (right, D_b) have increased rates of change in gene expression induced by an exogenous stressor by a factor δ relative to the changes in lower-D cells (left, D_a). For the higher-D cells, this may increase the probability of the cell remaining viable by reaching a critical threshold of expression of pro-acclimation genes compared to the lower-D cell which is unable to meet this threshold. (B&C) The fraction of high-D cells in a cell culture increases after treatment with paclitaxel for 48 hours, suggesting that cells with higher D are more likely to survive exposure to a cytotoxic chemotherapeutic agent. (B) The percentage of cells having D above the top quartile of a control cell population (y-axis) increases in cells that survive treatment with paclitaxel for 48 hours. For both conditions, each dot represents percentage of high-D cells in one replicate for a total number of $N = 5$ replicates per condition. (C) Combination treatment with the D-lowering celecoxib agent and then paclitaxel for 48 hours (Combo) results in increased elimination of cancer cells compared with untreated controls and cells treated with paclitaxel (PAC) alone. (D) CPMC model predictions of the relative transcriptional malleability coefficient δ for initially lowly-expressed (blue spline) and highly-expressed genes (red spline). Here, $D_a = 2.3$ and $D_b = 2.5$, which is comparable to experimentally observed differences in celecoxib-treated versus untreated A2780 cells. (E) scRNA-seq on A2780 cells was performed to compare transcriptional profiles of control A2780 cells (high-D population) and cells treated with 75 μM of a D-lowering agent celecoxib (low-D population) and their response to treatment with 5 nM paclitaxel (stressor) for 16 hours. Initially lowly-expressed and initially highly-expressed genes are defined based on control expression levels. Genes are grouped based on their quantile of $log_2(E_{PAC}/E_{control})$ and the mean and standard errors of each quantile for initially lowly-expressed genes (blue dots, 300 genes/data point) and initially highly-expressed genes (red dots, 100 genes/data point) are plotted. (F) Gene ontology analysis identified biological processes that are most significantly involved in the response to 48 hour paclitaxel treatment. Upregulated genes were defined as those with at least 2$\times$ increase in expression. (G) D-facilitated upregulation (δ) of the stress-response genes identified by the GO analysis (red points, 150 genes/data point) was similar to that for all upregulated genes (blue points, 650 genes/data point).

The functional significance of the transcriptional malleability coefficient δ is twofold. First, for genes that are highly amplified post-stimulation ($E_2/E_1 \gg 1$) the relative increase in transcription at any given time after the stimulation is proportional to δ:

$$\frac{[mRNA]_b(t) - [mRNA]_{1,b}}{[mRNA]_{1,b}} \approx \frac{\delta[mRNA]_a(t) - [mRNA]_{1,a}}{[mRNA]_{1,a}} \tag{3.9}$$

Second, the time τ required to reach a given level of expression E_2 is dependent on D and is inversely proportional to δ, $\tau_b/\tau_a \approx \delta^{-1}$. This conclusion is applicable to genes that are both upregulated as well as those that are downregulated in response to a stimulus, an effect that might be especially consequential if decisions regarding cell fates must be made within a limited time period after the introduction of the stressor [214].

To experimentally explore the relationship between D and transcriptional malleability, we employed scRNA-seq to track gene expression changes of A2780 ovarian adenocarcinoma cells in response to treatment conditions that modulate chromatin packing scaling, which was assessed using live-cell PWS microscopy. We then compared the transcriptional malleability of populations of cells with different initial D values. As a model system, we relied on chemically-induced modulation of D. To reduce D, we treated A2780 cells with 75 μM celecoxib, a nonsteroidal anti-inflammatory agent, for 16 hours. Previously, we have determined that in A2780 cells celecoxib reduces D by at least 8% compared to untreated cells within 30 minutes of treatment [5]. As a model of high-D cells, we used untreated A2780 cells. Both celecoxib-treated cells (low-D) and untreated A2780 cells (high-D) were then exposed to a chemotherapeutic agent for 16 or 48 hours. scRNA-seq was conducted using Illumina NextSeq500. Raw reads were aligned, mapped

and used to calculate transcripts per million (TPM) for each condition using bowtie2 and RSEM.

Inputting the experimentally observed difference in D into the CPMC model, we estimated $\delta > 4$ for initially lowly-expressed genes that become activated (Fig. 3.3D, blue manifold) and a smaller increase in δ for initially highly-expressed genes that are upregulated in response to stimulation (Fig. 3.3D, red manifold). We then tested if these predicted trends are observed experimentally using scRNA-seq. Importantly, the crucial window for response to chemotherapy frequently is thought to occur within 24 hours [214, 102]. Thus, we compared changes in gene expression for both high-D and low-D A2780 cell populations after paclitaxel treatment for 16 hours. In agreement with the CPMC model, the stimulation of initially lowly-expressed genes by chemotherapy treatment in the initially high-D population (upregulation of expression rate from control rate $E_{1,b}$ to 16 hr paclitaxel-treated rate $E_{2,b}$) was much higher than that in the low-D population (from celecoxib-treated rate $E_{1,a}$ to 16 hr combo rate $E_{2,a}$), resulting in $\delta \sim 4$ (Fig. 3.3E). This signifies that genes with initially lower expression in the pre-stimulated state are upregulated 4× more in the high-D compared to the low-D A2780 cell populations. Likewise, a similar but mitigated effect was observed in initially highly-expressed genes (Fig. 3.3E), in strong agreement with the model predictions.

Next, we tested whether these trends were independent of cell line and compound. We performed parallel experiments using propranolol as a D-lowering agent in A2780 cells and celecoxib and propranolol to decrease D in more malignant TP53 mutant A2780 (M248) cells. These additional conditions demonstrated a similar effect of initial D on transcriptional malleability in response to paclitaxel stimulation of high-D compared to

low-D cell populations (Fig. 3.13). Finally, we tested if the observed effect of chromatin packing scaling influences genes specifically involved in functionally-relevant stress response pathways. We first identified differentially expressed genes that, on average, increased their expression at least two-fold in A2780 cells treated with paclitaxel for 48 hours compared to control cells. Gene ontology analysis of these upregulated genes demonstrated the activation of multiple stress response pathways by paclitaxel treatment, including DNA repair, autophagy, cell cycle arrest, and apoptosis ($P < 0.05$, Fig. 3.3F). The effect of D on the activation of these established stress response genes was consistent with the activation observed for all upregulated genes, with δ as high as $\sim$4 (Fig. 3.3G).

3.4.3. Chromatin Packing Scaling is Associated with Intercellular Transcriptional Heterogeneity

Another key aspect of phenotypic plasticity that can be modulated by the disordered packing of chromatin is transcriptional heterogeneity, or the range of expression levels across genes exposed to similar molecular conditions. The CPMC model predicts that transcriptional heterogeneity increases as a function of D due to increased variations in both packing density, $\sigma^2_{\phi_{in}}$, and gene accessibility (p_g). To quantify this effect from the CPMC model, the variance in ϵ across any given cell population, Var_ϵ, is [5]:

$$Var_\epsilon \approx \frac{1}{2}\Big(\frac{\partial^2 \epsilon(\vec{m}, \phi_{in})}{\partial^2 \phi_{in}}\Big|_{\phi_{in}=\phi_{in,0}}\Big)^2 \sigma^4_{\phi_{in}} \tag{3.10}$$

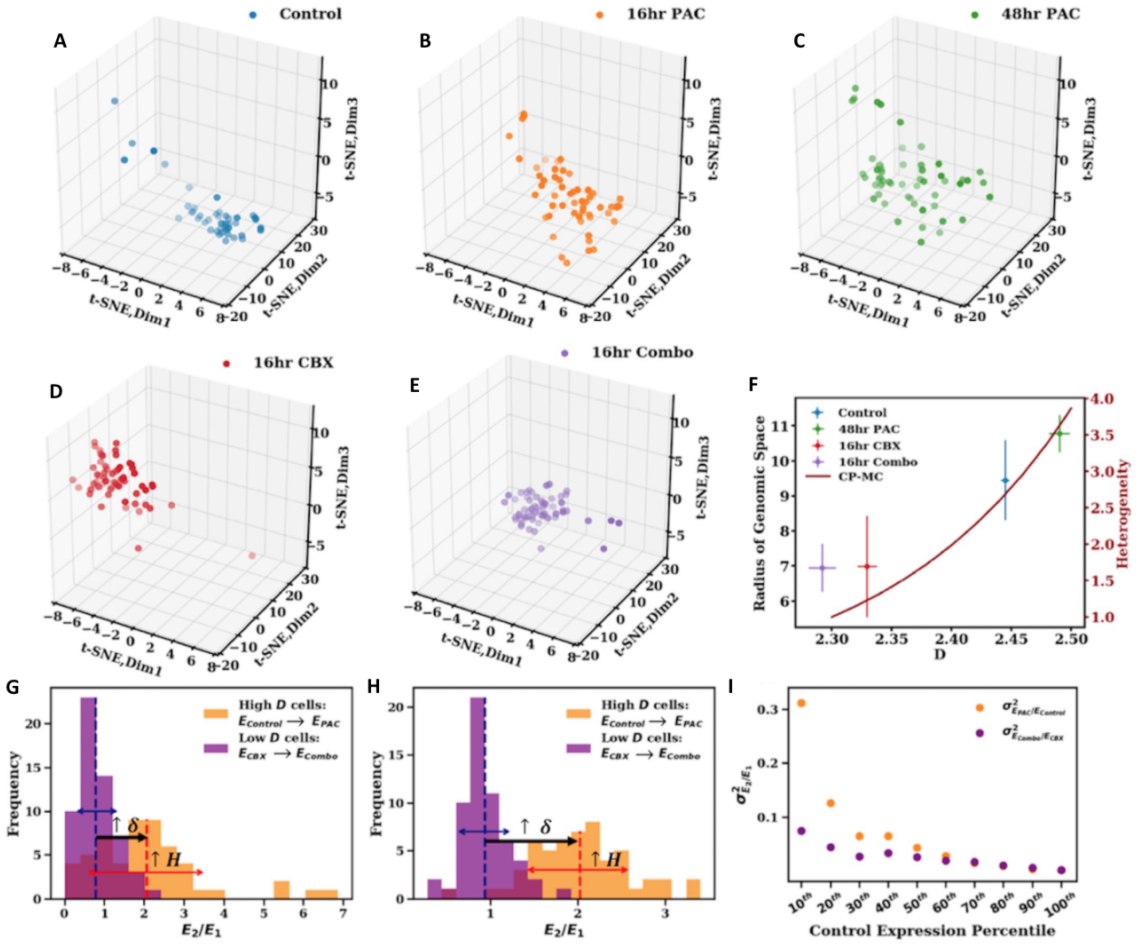
A
Control
B
16hr PAC
C
48hr PAC
D
16hr CBX
E
16hr Combo
t-SNE,Dim1
t-SNE,Dim2
t-SNE,Dim3
F
Control
48hr PAC
16hr CBX
16hr Combo
CP-MC
Radius of Genomic Space
Heterogeneity
D
G
H
High D cells:
$E_{Control} \rightarrow E_{PAC}$
Low D cells:
$E_{CBX} \rightarrow E_{Combo}$
Frequency
$\uparrow \delta$
$\uparrow H$
E_2/E_1
I
$\sigma^2_{E_2/E_1}$
$\sigma^2_{E_{PAC}/E_{Control}}$
$\sigma^2_{E_{Combo}/E_{CBX}}$
Control Expression Percentile

Figure 3.4. Chromatin packing scaling regulates intercellular transcriptional heterogeneity of cancer cells. (A-E) 3D projections of scRNA-seq data (TPM values of 8,275 expressed genes) onto reduced t-SNE space for 5 conditions: (A) control cells ($N = 46$), (B) cells treated with 5 nM paclitaxel for 16 hours (1 6hr PAC, $N = 55$), (C) 5 nM paclitaxel for 48 hours (48 hr PAC, $N = 5$), (D) 75 μm celecoxib for 16 hours (16 hr CBX, $N = 62$), (E) and combination of 75 μM celecoxib and 5 nM paclitaxel for 16 hours (16 hr Combo, $N = 59$). The size of the cluster indicates the transcriptional heterogeneity within the population of surviving cells for each condition. (F) The radius of genomic space R_c (the radius of clusters through A-E) increases as a function of D, which was measured by live-cell PWS microscopy at each time point prior to sequencing. Cells treated with paclitaxel (higher D) have greater transcriptional heterogeneity, especially when compared to cells treated with the D-lowering celecoxib agent. Likewise, the CPMC model (red curve, right side y-axis) shows that intercellular transcriptional heterogeneity increases with D. Error bars represent the standard error of D calculated from PWS measurements (x-axis) and R_c (y-axis) for each condition. (G) Relative expression of high-D versus low-D cells in response to paclitaxel treatment for genes associated with DNA repair pathways, which are upregulated in 48 hour paclitaxel-treated cells. For each condition (Control, 16 hr PAC, 2 hr CBX, 16 hr Combo), TPM values of these genes (48 in total) were averaged within each cell. Next, expression of paclitaxel-stimulated cells was normalized by the average of the corresponding unstimulated population. The resulting intercellular distribution of relative expression levels is shown. Dashed lines represent mean relative expression. Solid red and blue arrows represent the standard deviation of distributions $E_{PAC}/E_{Control}$ and E_{CBX}/E_{Combo}, respectively. For these stress response genes, cells with a higher initial D versus cells with a lower initial D had an increase in transcriptional malleability ($\uparrow \delta$) as well as a higher intercellular transcriptional heterogeneity ($\uparrow H$). (H) Distribution of the relative expression of genes, as described in (G), in the lowest quantile (10^{th} percentile) of control expression levels (839 in total). (I) Variance (σ^2) of intercellular distribution of relative expression for each percentile of control expression levels. Initially lowly-expressed genes show an increased effect of chromatin packing scaling on increasing intercellular transcriptional heterogeneity in response to paclitaxel stimulation compared to that of initially highly-expressed genes in higher quantiles.

Consequently, the transcriptional heterogeneity, or the standard deviation of steady-state expression rate E in Eq. 3.3, becomes:

$$H(D) = p_g \cdot \sqrt{Var_\epsilon} \approx \frac{\sqrt{2}}{8} p_g \cdot \left(\sigma^2_{\phi_{in}}\right)^2 \kappa \left(1 + \sqrt{1 + \frac{16}{\left(\sigma^2_{\phi_{in}}\right)^2}} \frac{\bar{\epsilon}}{\kappa}\right) \tag{3.11}$$

and the coefficient of variation (the ratio of the standard deviation to the mean expression) is

$$COV(D) = \frac{\sqrt{2}}{8} \left(\sigma^2_{\phi_{in}}\right)^2 \frac{\kappa}{\bar{\epsilon}} \left(1 + \sqrt{1 + \frac{16}{\left(\sigma^2_{\phi_{in}}\right)^2}} \frac{\bar{\epsilon}}{\kappa}\right) \tag{3.12}$$

Both H and COV increase with D.

To investigate the association between D and intercellular transcriptional heterogeneity, using our scRNA-seq data we quantified the spread in transcriptional states across each treatment condition. Focusing on overall transcriptional differences between cells within the same condition provides better validation to the model than analyzing the spread of all observed genes. Thus, we first used t-Distributed Stochastic Neighbor Embedding (t-SNE) combined with principal component analysis (PCA) to reduce the dimensionality of the system on all cells simultaneously [285]. The dimensionality reduction mapped each cell onto a three-dimensional projection. Distances between cells in 3D space represent overall differences in transcriptional states, as has been described by van der Maaten and Hinton [285]. Intercellular transcriptional heterogeneity for each cell population was quantified by the average radius of the cluster of cells, $R_c = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(r_i - r_{mean}\right)^2}$ where r_i is the position of each cell in the reduced spaced, N is the total number of cells in each treatment group, and $r_{mean} = \frac{1}{N}\sum_{i=1}^{N} r_i$. Intuitively, R_c

can be thought of as the radius of relative genomic space. Consistent with predictions of the CPMC model, we found that transcriptional heterogeneity, as measured by the radius of genomic space, increases with D in response to paclitaxel treatment, which pre-selects for high D cells, as shown in Fig. 3.3. Notably, after 48 hours of paclitaxel treatment, the population of surviving cells had both higher D and increased transcriptional heterogeneity compared to control cells (Fig. 3.4A-C&F). In contrast, celecoxib treatment reduces average D of cancer cell populations. Accordingly, cells treated with celecoxib for 16 hours had a lower transcriptional heterogeneity compared to control cells. In addition, when these celecoxib-primed cells with initially lower D were treated with paclitaxel for 16 hours, they had a decreased transcriptional heterogeneity compared to paclitaxel-treated control cells (Fig. 3.4D-F). Additional analyses quantifying the Euclidean distance between expression of stress response-associated DNA repair genes upregulated in 48 hour paclitaxel treatment as well as the coefficient of variation of expression between cells in the same treatment condition also demonstrate that cell populations with higher D also have higher transcriptional heterogeneity (Fig. 3.14).

3.5. nano-ChIA Validates CPMC in Individual Nuclei

In addition, we employed the STORM-PWS modality of our nano-ChIA platform to investigate the relationship between chromatin packing behavior and molecular transcriptional events in individual nuclei (see Chapter 1.3). Here, we use the CPMC model to predict how chromatin packing scaling, D, modulates active gene transcription. As D of a PD increases, the model predicts an increase in the accessible surface area of chromatin, p_g, within the PD. In addition, an increase in D increases the variance of crowding conditions, $\sigma^2_{\phi_{in}}$, to which the genes within the PD are exposed. As a result of the competing effects of accessible surface and distribution of crowding conditions, at lower D, gene expression will increase with D up to an inflection point. Above this critical value of D, the range of crowding conditions to which the genes are exposed becomes suboptimal. Thus, after a certain critical D value, the transcriptional output is predicted to decrease. The shape of this nonmonotonic relationship between D and gene expression is dependent on the molecular and physical regulators of transcription defined in Chapter 3.2. For example, higher concentrations of transcriptional reactants increase ensemble expression rates across all D values (Fig. 3.5A). In addition, these more favorable molecular conditions shift the critical D inflection point to higher values (Fig. 3.5A).

To test the predicted relationship between chromatin structure and gene expression experimentally, we used STORM-PWS to localize regions of active gene transcription by labeling actively elongating RNA Pol II with the phospho-Ser2 antibody and imaging with STORM, and then measuring the surrounding chromatin packing scaling with PWS microscopy (Fig. 3.5 C&D). CPMC model predictions of transcription rates were in excellent agreement with the *in situ* experimental STORM-PWS findings across multiple cell

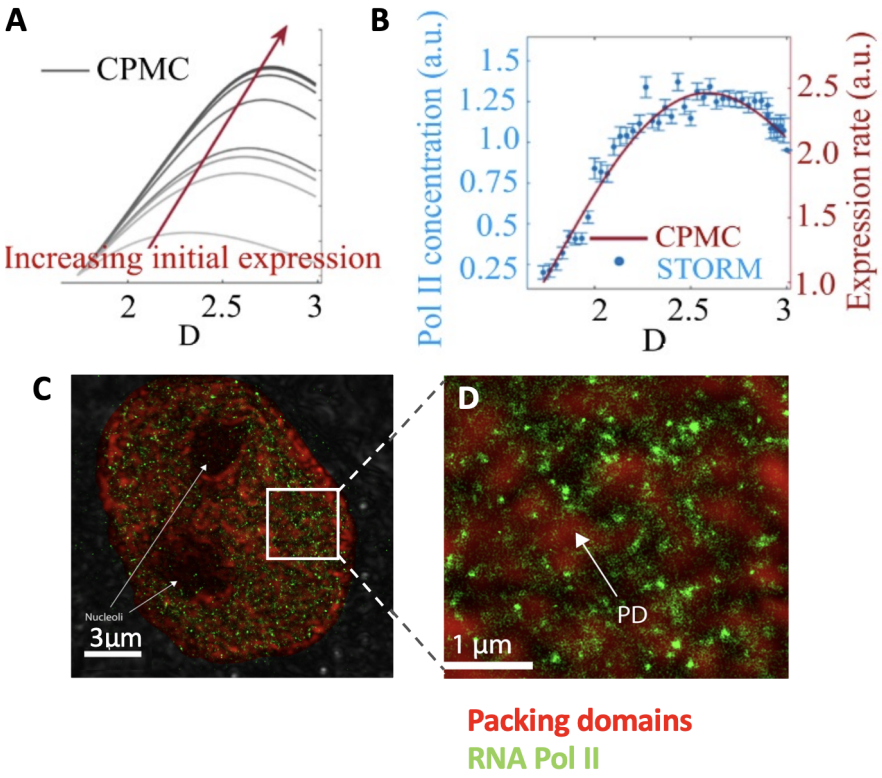

Figure 3.5. STORM-PWS imaging validates the relationship predicted by CPMC between local chromatin packing behavior and active transcription. (A) Multiple realizations of the CPMC model with varying molecular conditions, specifically different concentrations of RNA Pol II, TFs, and promoters. The modeling predicts that in all cases, the surrounding chromatin packing scaling has a nonmonotonic relationship with gene expression. (B) The relationship between D and the local concentration of active RNA Pol II which correlates with gene expression level ($N = 4$ cells) compared with one realization of the CPMC model. (C) STORM image of an M248 ovarian cancer cell with labeled active RNA Pol II (green) overlaid on top of chromatin packing scaling D map measured by PWS (red). (D) Magnified view of the white square in (C).

lines, demonstrating a consistent nonmonotonic relationship between chromatin packing behavior (D) and transcription (Fig. 3.5B) [45].

3.6. Chromatin-Dependent Adaptability (CDA) Model

3.6.1. Model Description

Now that we had established that (1) chromatin packing scaling was a strong predictor of two components of phenotypic plasticity, transcriptional malleability and intercellular heterogeneity, and (2) D predictably increases upon exposure to numerous types of chemotherapeutic agents and across multiple cancer lines (see Chapter 3.8.4), as a next step, we wanted to directly incorporate the effects of initial chromatin state on observed cell phenotype. This led us to develop a model that mechanistically explores whether average changes in population-level D directly contribute to cancer cell survival under cytotoxic stress.

For this purpose, we extended our non-steady-state CPMC model to predict how the chromatin-mediated changes in expression of certain critical genes contribute to cell survival upon exposure to a novel cytotoxic stressor. We term our new model the Chromatin-Dependent Adaptability (CDA) model. A schematic of the CDA model can be found in Fig. 3.6.

Consider that a cell is exposed to an exogenous stressor and that the survival of the cell is dependent on the change in expression of gene(s) within key biological pathway(s) before a critical timepoint. This framework of understanding cell survival is relevant for the field of cancer where, for example, the change in expression of p53 within a certain amount of time after chemotherapeutic treatment determines whether cancer cells undergo apoptosis [214]. Additionally, in the field of ecology the "plasticity-first" theory of evolution stipulates that environmentally-initiated phenotypic plasticity of a

natural population can facilitate evolutionary adaptation to a given stressor [160, 243]. This definition of phenotypic plasticity encompasses [160, 243]:

(1) The idea of a reaction norm, or transcriptional changes that contribute to fitness of a phenotype in response to a stressor (i.e., transcriptional malleability).
(2) The fact that stressors uncovering cryptic genetic variation (i.e., transcriptional heterogeneity) usually has negligible effects on phenotypic variation, except under atypical conditions.

Such a framework has been successfully applied to modern ecological problems, such as how corals survive rapidly changing environmental conditions [243].

The CDA model predicts that if a cell population is exposed to a potentially cytotoxic stressor, that the population's survival probability depends on the population's average packing behavior, which modulates transcriptional changes in response to the stressor. To note, we are not considering that all cells will survive all stressors if they upregulate or downregulate one specific gene or genes in a pathway. This is merely a statistical model that evaluates cell survival in the context of a dynamic response to a stressor that is preprogrammed by the average initial chromatin packing state, which encompasses crowding conditions of all genes within the cell.

For any given gene(s) of interest, the number of transcripts will be up- or down-regulated by $x = N_2/N_1$, where N_1 and N_2 are the number of mRNA transcripts before and after stress exposure, respectively. Without loss of generality, let us consider upregulation with the understanding that gene downregulation can be treated with a similar formalism. Let us assume that $PDF(x)$ follows a log-normal distribution. This is a reasonable

assumption considering that, in general, gene expression (i.e., N_1 and N_2) follows a log-normal distribution and $ln(x) = ln(N_2) - ln(N_1)$.

$$PDF(x) \approx \frac{1}{s\sqrt{2\pi}x} exp\left(-\frac{ln(x/m)^2}{2s^2}\right) \tag{3.13}$$

For $s \ll 1$, $ln(m) \approx \mu$, where μ is the mean and s approximates the coefficient of variation (COV) of the upregulation of transcripts.

Let us now assume that a cell's decision to survive a cytotoxic stressor within a critical time period is dependent on the upregulation of certain stress response genes above a threshold level x_{crit}. Thus, the probability of cell survival can be considered as a step function of x/x_{crit}: if $x > x_{crit}$ the cell survives, and if $x < x_{crit}$ the cell dies. The resulting probability of cell death, θ, can then be defined as the cumulative distribution function of this threshold x_{crit} and:

$$\theta(x_{crit}) = CDF(x_{crit}) = \frac{1}{2} erfc\left(\frac{ln(m/x_{crit})}{\sqrt{2}s}\right) \tag{3.14}$$

Next, we approximate the log-normal CDF as the Hill equation, which is simpler to approach numerically

$$\theta(x_{crit}) \approx \frac{1}{1 + (k/x_{crit})^n} \tag{3.15}$$

where $k \approx m$ is the malleability and $n \approx \frac{3/\sqrt{\pi}}{s}$ is the inverse COV/heterogeneity.

If we compare the upregulation of cell a ($\langle D \rangle = D_a$) to that of cell b ($\langle D \rangle = D_b$), then we can calculate the death probability of cell b as:

$$\theta(x_{crit}) \approx \frac{1}{1 + (k_a \gamma_k / x_{crit})^{\frac{n_a}{\gamma_n}}} \tag{3.16}$$

where $\gamma_k = k_b/k_a$ and $\gamma_n = n_b/n_a$. Here, γ_k represents the ratio of the mean upregulation $N_{2a,b}/N_{1a,b}$ of cell b and cell a and γ_n represents the ratio of the COV of the upregulation of the two cells.

Next, we must link the upregulation of critical gene(s) and cell survival with the initial chromatin packing state of the cell using our CPMC model (see Chapter 3.2). In the context of CPMC, k represents the transcriptional malleability, or the average change in gene expression upon stimulation, and s represents the transcriptional heterogeneity, or the range of transcriptional states accessible by a cell population. Both of these metrics of phenotypic plasticity increase with D (see Chapter 3.4). Again, note that when we consider the initial chromatin packing state of a cell we take the average of all descriptors over all PDs of the cell. If a cell has a higher average D, for example, then this will increase the probability of all genes within the cell to be localized to a PD that has a higher average D.

Note, for steady-state conditions the number of mRNA transcripts (N) is not time-dependent, i.e., $\frac{\partial N}{\partial t} = 0$. In steady-state, CPMC can predict N by determining expression rate E and then using the relation $E = N\nu$ for mRNA degradation rate ν [215]. However, under extreme cytotoxic stressors, such as treatment with chemotherapeutic agents, we are not likely to be dealing with steady-state conditions. Here, we develop a formalism for non-steady-state CPMC. Let a cell population be exposed to such a stressor at time

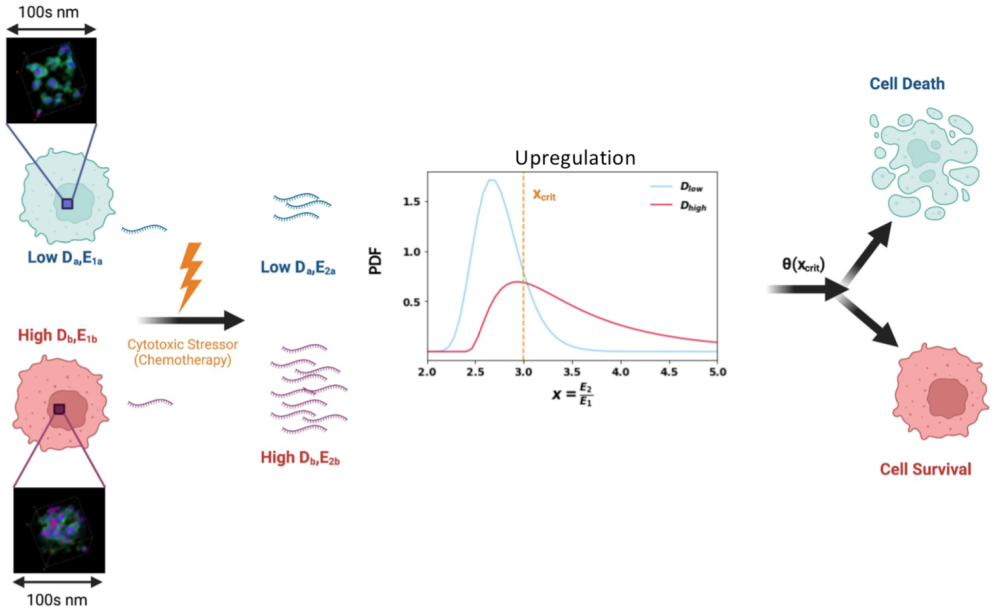

Figure 3.6. Chromatin-Dependent Adaptability (CDA) model predicts cell survival to cytotoxic stressors from the initial chromatin packing behavior of cells. CDA model predicts that cells with higher average chromatin packing scaling have both an increased average and spread of upregulation of potentially critical genes for cell survival. Thus, high-D cells (in red) will have a higher probability of upregulating key gene(s) beyond a threshold, x_{crit}, before a critical cell decision timepoint. Consequently, these high-D cells will have a lower death probability, θ, which will increase cell survival. Conversely, low-D cells (in blue) might not be able to upregulate genes beyond the critical threshold within the same amount of time, and will have a higher death probability.

$t = 0$. The number of transcripts for a subsequent timepoint t can be determined from the relation

$$\frac{\partial N}{\partial t} = \frac{\partial E}{\partial t}(1 - e^{-t/\tau}) \tag{3.17}$$

where τ is the mRNA elimination time constant. τ is calculated from the mRNA half-life, $\tau_{1/2}$ using the relation $\tau = \frac{1}{ln2}\tau_{1/2}$. We consider $\tau_{1/2} = 10$ hours [301]. We can approximate

$\frac{\partial N}{\partial t} \approx \frac{N_2(t)-N_1}{N_1}$ as the average change in the number of transcripts and $\frac{\partial E}{\partial t} \approx \frac{E_2(t)-E_1}{E_1}$ as the average change in the mRNA production rate. Notice here that both mRNA transcript number $N_2(t)$ and expression rate $E_2(t)$ after stress exposure are both functions of time after treatment t.

Defining $\beta = \frac{E_2(t)}{E_1}$ and $k = \frac{N_2(t)}{N_1}$, we can determine the transcriptional malleability by:

$$k(t) = 1 + (\beta - 1)(1 - e^{-t/\tau}) \tag{3.18}$$

In relation to the death probability, θ, we consider that the cell makes a survival decision after exposure to a cytotoxic stressor at time $t = T_{crit}$.

If we know β_a, i.e., the average upregulation of the expression rate for cell a, then we can predict β_b with CPMC using the relation

$$\frac{\beta_b}{\beta_a} = \exp\left(\int_{D_a}^{D_b} Se(\beta_a E_1, D') - Se(E_1, D')\,\frac{dD'}{D'}\right) \tag{3.19}$$

Here, Se represents the sensitivity of gene expression to average chromatin packing scaling D with initial expression E as described in Chapter 3.2. Thus, we predict the upregulation of expression rate in cell b based on the differences in average D values between cell b and cell a. We then can then determine $k_b = k_a\gamma_k$, which we need to calculate $\theta(x_{crit})$ (see Eq. 3.16), by plugging β_b determined from Eq. 3.19 into Eq. 3.18.

Next, we determine the time-dependent changes of transcriptional heterogeneity for cell b using the relations:

$$s_b(t) = COV\left[x = \frac{N_2(t)}{N_1}\right] = \frac{\sqrt{\sigma^2_{\bar{\epsilon}_{2,b}/\bar{\epsilon}_{1,b}}}}{k(\frac{\bar{\epsilon}_{2,b}}{\bar{\epsilon}_{1,b}})} \tag{3.20}$$

$$\sqrt{\sigma^2_{\bar{\epsilon}_{2,b}/\bar{\epsilon}_{1,b}}} = \frac{\bar{\epsilon}_{2,b}}{\bar{\epsilon}_{1,b}}\sqrt{COV[\bar{\epsilon}_{1,b}]^2 + COV[\bar{\epsilon}_{2,b}]^2} \tag{3.21}$$

$$s_b(t) = \frac{\bar{\epsilon}_{2,b}}{\bar{\epsilon}_{1,b}} \frac{1 - e^{-t/\tau}}{1 + \left(\frac{\bar{\epsilon}_{2,b}}{\bar{\epsilon}_{1,b}} - 1\right)(1 - e^{-t/\tau})} \sqrt{COV[\bar{\epsilon}_{1,b}]^2 + COV[\bar{\epsilon}_{2,b}]^2} \tag{3.22}$$

where $COV[\bar{\epsilon}_{1,2}] = \sqrt{2}G(\bar{\epsilon}_{1,2})$ increases with average initial D of the cell population. Here, $\bar{\epsilon}_{1,2}$ is the expression rate of a given gene per unit of DNA pre and post-stimulus, which is defined by the following equations:

$$\bar{\epsilon}_{1,b} = \bar{\epsilon}_{1,a}\gamma(D_a, D_b, \bar{\epsilon}_{1,a}) \tag{3.23}$$

$$\bar{\epsilon}_{2,b} = \bar{\epsilon}_{1,a}\beta_a\gamma(D_a, D_b, \bar{\epsilon}_{1,a}\beta_a) \tag{3.24}$$

$$\gamma(\bar{\epsilon}) = \exp\left(\int_{D_a}^{D_b} Se(\bar{\epsilon}, D')\,\frac{dD'}{D'}\right) \tag{3.25}$$

where $\gamma(\bar{\epsilon}) = \frac{\bar{\epsilon}_{i,b}}{\bar{\epsilon}_{i,a}}$ represents the average change in expression rate for cell a compared to cell b before ($i = 1$) and after ($i = 2$) treatment with a cytotoxic stressor and is calculated using the CPMC sensitivity Eq. 3.6. Substituting Eqs. 3.21,3.22,3.23,3.24,3.25 into Eq. 3.20, we can then calculate $n_b = \frac{3/sqrt\pi}{s_b} = \frac{n_a}{\gamma_n}$, which is necessary to calculate $\theta(x_{crit})$ in Eq. 3.16.

3.6.2. Chromatin Packing Scaling Influences Death Probability of Cancer Cells Treated with Chemotherapy

Thus, we can predict the death probability of cell b if we know $beta_a$, the upregulation rate of cell a, and D_a and D_b. Additionally, we must estimate the relative initial expression of the given upregulated gene(s), $ln(E_i/\bar{E}_i)$, the threshold of upregulation for cell survival x_{crit}, and the critical decision timepoint T_{crit} as well as other relevant parameters from the CPMC model including average PD size, N_d, and average PD density, $\phi_{in,0}$. Here, we perform all calculations with $N_d = 210$ kbp, the median value for A549 PDs analyzed by ChromSTEM (see Chapter 2.2), and $\phi_{in,0} = 0.4$, where we added the median $CVC = 0.37$ of A549 PDs to an estimated volume concentration of 0.03 for mobile crowders.

First, we investigated how the different model parameters contribute to overall cell death probability. Both malleability, k, and heterogeneity, s, increase with average nuclear $\langle D \rangle$ (Fig. 3.7A) as also demonstrated by the shift in the log-normal distribution for gene expression upregulation (Fig. 3.6). For a set critical threshold, this means that for higher $\langle D \rangle$ the overall distribution will be more likely to fall above x_{crit} (Fig. 3.6). This is especially the case for genes with low expression levels before stress exposure, which the CPMC model predicts will be upregulated to a greater extent than genes with an initial higher expression. There is a general sigmoidal behavior between $\langle D \rangle$ and cell death probability. For a specific case of $\beta_a = 10$ and $x_{crit} = 10$, compared to a reference $D_a = 2.6$, lower $\langle D \rangle \approx 2.2-2.5$ has a death probability $\theta \approx 1$, which decreases sigmoidally until it plateaus again between $\langle D \rangle \approx 2.7-2.8$ (Fig. 3.7B). As the threshold x_{crit} changes, there is a shift in the sigmoidal distribution (Fig. 3.7B).

Note that the sigmoidal relationship is also dependent on the initial relative expression of the critical gene(s). Genes with lower initial expression have a steeper sigmoidal behavior, resulting in lower cell death for higher $\langle D \rangle$ values, compared to genes with higher initial expression, which have death probabilities that are much less sensitive to $\langle D \rangle$ (Fig. 3.7C).

The critical timepoint for the cell death decision, T_{crit}, is another important parameter in the CDA model. Cells usually exhibit signs of apoptosis between $5 - 10$ hours after treatment with chemotherapy, although there is significant heterogeneity in apoptosis induction both within and across cell lines that is also dependent on dosage of cytotoxic chemotherapeutic treatment [99, 127, 242]. Additionally, Mielgo et al. demonstrated that paclitaxel treatment activates the proapoptic response within 6 hours, which is sufficient for cell death induction by caspase 8 at later timepoints [191]. This indicates that induction of caspases is a sufficient predictor of commitment to cell death. Here, we use $T_{crit} = 7$ hours. Fig. 3.15 demonstrates the effects of decision time on the relationship between $\langle D \rangle$ and death probability. We calculated these death probability curves within the range of $T_{crit} = 5 - 10$ hours and for T_{crit} =24 hours, which represents an upper limit for the cell death decision. Increasing T_{crit} shifts the sigmoidal curve so that cells with lower $\langle D \rangle$ have a decreased death probability.

The CDA model as derived can be applied to any potentially cytotoxic stressor. Now, let us evaluate how this model performs for the specific case of cancer cells exposed to chemotherapeutic agents. Specifically, we will compare model predictions to experimental data for HCT116 cells which have been treated with oxaliplatin for 48 hours. Briefly, an optimization procedure over β_a and x_{crit} was performed to produce a best fit

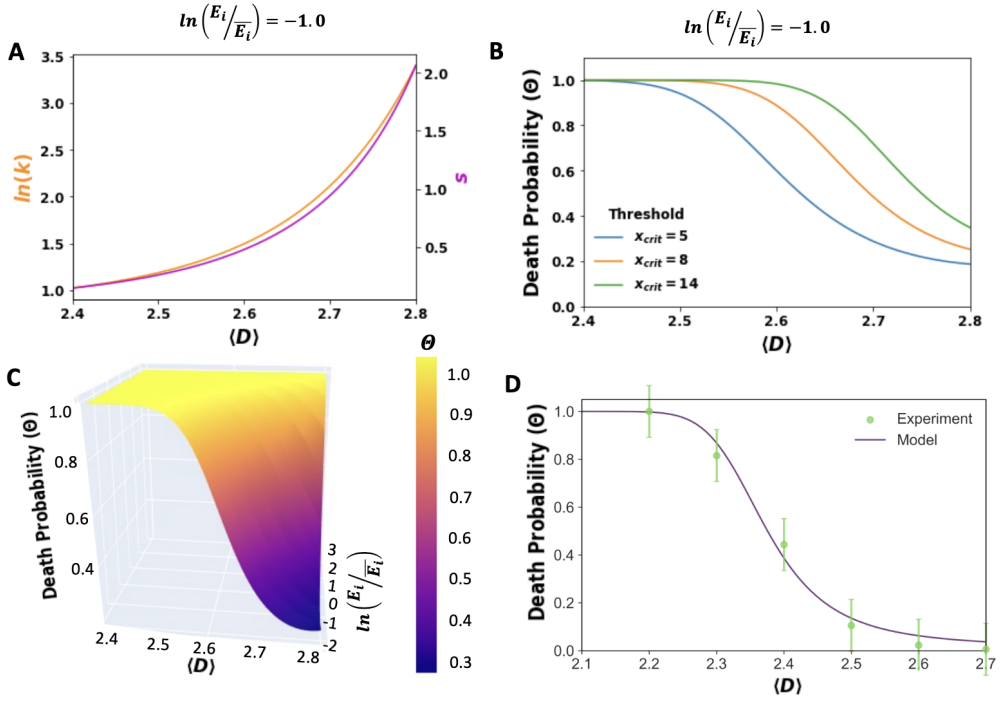

Figure 3.7. CDA parameters influence death probability of cells under cytotoxic stress. (A) Cells with higher $\langle D \rangle$ have increased transcriptional malleability k and heterogeneity s compared to reference cells with $D_a = 2.6$. (B) Death probability compared to $\langle D \rangle$ follows a general sigmoidal relationship that is dependent on the critical upregulation rate of certain gene(s), x_{crit}. Higher x_{crit} results in increased death probability, θ, for cells with the same $\langle D \rangle$. (C) Initial relative expression of genes $ln(E_i/\bar{E}_i)$ influences the sigmoidal behavior of the $\langle D \rangle$ versus death probability relationship. Genes with lower initial relative expression have a steeper sigmoidal curve. (D) Agreement was found between the death probability curve predicted by the CDA model and the experimental death probability values calculated for cell clusters with varying initial $\langle D \rangle$ after a 48 hour oxaliplatin treatment of HCT116 cells. After optimization, fitting parameters were determined to be $x_{crit} = 6.9$ and $\beta_a = 19.1$.

for the model to experimental data for these specific experimental conditions. For cells within the same cell population, the model predicts a higher probability of cell death for cells with lower $\langle D \rangle$ values due to their inability to alter expression levels above a

critical level during chemotherapeutic treatment, which matches well with experiments (Fig. 3.7D). Experiments were performed by measuring initial $\langle D \rangle$ of cell clusters using PWS microscopy and then tracking cell survival over time. Overall, this demonstrates the ability of the CDA model to capture the relationship between initial chromatin packing state, $\langle D \rangle$, of a cancer cell and cell death probability upon exposure to chemotherapeutic stress.

3.6.3. Evaluating Effects of Combination Treatment with D-lowering Agents on Adaptability and Chemoevasion Potential of Cancer Cells

3.6.3.1. Identifying Candidate Chromatin Protective Therapies (CPTs) Using PWS Microscopy. In response to cytotoxic stress, cellular fitness was positively associated with increased average chromatin packing scaling of cancer cells. We next investigated the effects of candidate chromatin-modifying drugs to determine whether shifting the population-wide distribution of average nuclear packing scaling to lower D values would increase chemotherapeutic efficacy. We term such compounds Chromatin Protective Therapies (CPTs) [5]. As cytotoxic stress response occurs within a matter of hours after chemotherapy treatment *in vitro*, we selected for agents that would decrease cellular fitness within this decision window. Specifically, we looked for CPT drugs that decrease $\langle D \rangle$ within very short time scales ($<$ 1 hour).

We found that modulating the epigenetic state of chromatin, though processes such as deacetylation and methylation, decreased the average D distribution (Fig. 3.16A). However, even high doses of such acetylation and methylation regulators resulted in only modest changes in D compared to compounds that alter the physicochemical intranuclear environment. As chromatin is a negatively charged polymer, due to the phosphate in the sugar-phosphate DNA backbone, altering the ionic environment is expected to modulate DNA-histone and nucleosome-nucleosome interactions, and could thus alter nuclear-wide chromatin structure (see Chapter 4). Here, we tested compounds such as non-steroidal anti-inflammatory drug celecoxib, which has previously been identified to inhibit Na^+, K^+, and Ca^{2+} channels [89], and cardiac glycoside digoxin, which suppresses the Na^+/K^+-ATPase ion pump and reduces intracellular $[K^+]$ as a side-effect

of its activity. Overall, we determined that these drugs decreased population-wide D to a greater extent than epigenetic modifiers (Fig. 3.16A&B). From our initial screen, we selected a handful of compounds that acted through potentially different mechanisms to confer a large decrease in D. We then further characterized and tested these compounds in multiple cancer cell lines (see Chapter 3.8.7). Although CPTs modulate D in a cell line-specific manner, there exist several "strong" CPTs, including celecoxib and digoxin, that decreased average D for cell populations across multiple cancer cell lines (Fig. 3.16).

Next, we wanted to evaluate the increased effectiveness of chemotherapy upon pre-treatment with compounds we identified as "strong" CPTs. To compare the effects of chromatin modulation on chemosensitivity, we analyzed the differential effects of "strong" CPTs, celecoxib and digoxin, as well as weaker CPTs, valproic acid (VPA) and aspirin. Under normal growth conditions, untreated A2780 cells rapidly grew into colonies and covered much of the imaging field. As expected, 48 hour mono-treatment with paclitaxel resulted in about 60% cellular inhibition over the same growth period as the controls (Fig. 3.17). Notably, combination treatment of paclitaxel with celecoxib or digoxin, both of which rapidly decreased D, greatly enhanced the efficacy of chemotherapeutic intervention with clearance approaching 100% (Fig. 3.16 & 3.17). Importantly, the degree of measured decreases in D by the tested agents strongly correlated with their increased chemotherapeutic efficacy, although efficacy varied depending on the cell line (Fig. 3.17 & 3.17). For all cell lines and chemotherapy agents tested, the co-treatment with the chemotherapy drug and at least one "strong" CPT agent resulted in nearly 100% cancer cell elimination, whereas relatively less elimination occurred for cancer cells co-treated with weaker CPTs that did not modulate D as significantly.

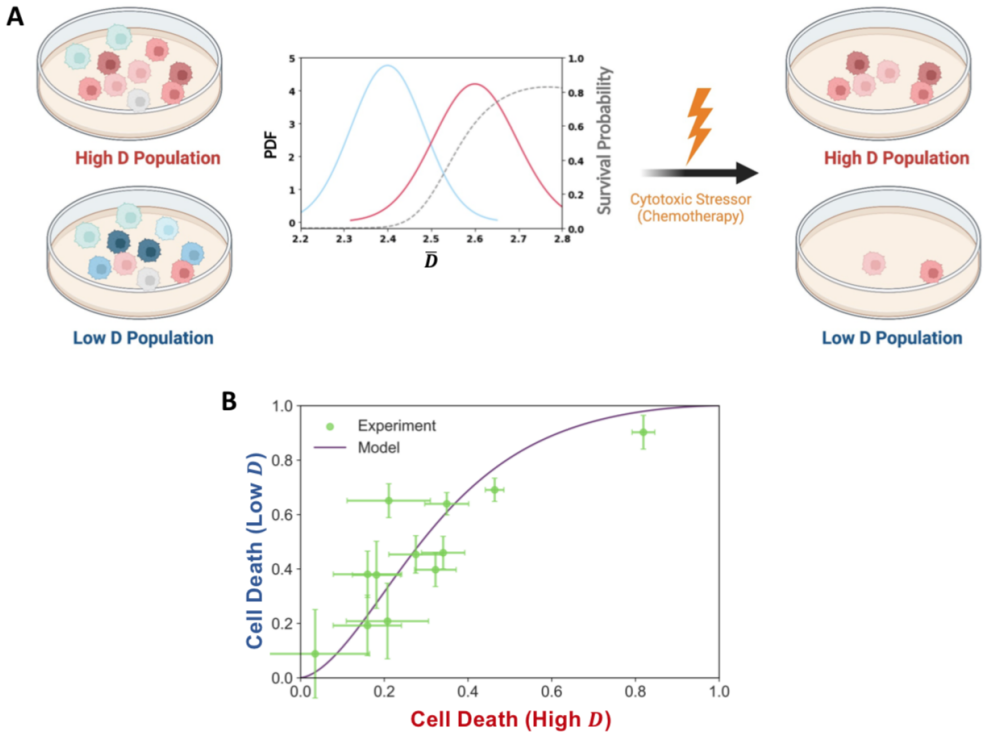

Figure 3.8. Lowering population-wide D with "strong" CPTs increases cell death upon treatment with cytotoxic chemotherapy. (A) CDA model predicts that higher D populations have more cells above the survival probability threshold (1-θ) and thus more cells from the high-D population will survive compared to the low-D population. (B) CDA model predictions are validated by experiments assessing survival of A2780 cancer cells upon exposure to different concentrations of chemotherapy (high-D population) compared to A2780 cells first treated with the CPT celecoxib and then exposed to the same concentrations of chemotherapy (low-D population). Fitting parameter for optimization $\beta_a = 9$.

3.6.3.2. Strong CPTs Increase Cell Death Upon Combination Treatment Compared to Treatment with Chemotherapy Alone. Finally, we wanted to directly evaluate cell death upon chemotherapy treatment of a lower D cancer cell population, treated with the "strong" CPT celecoxib, compared to the same untreated cell

population with higher D. The CDA model predicts that more cells from the higher D population will survive compared to the lower D population due to their ability to upregulate certain gene(s) within a critical timeframe (Fig. 3.8A). To generate experimental data with variable cancer cell death, we treated A2780 ovarian cancer cells with different concentrations of paclitaxel. Model predictions were generated by optimizing the fit to experimental cell death data for two populations with experimentally determined D distributions. Cell death for the high-D population was varied between 0 and 1 and then used to predict cell death for the low-D populations. Model predictions match experimental data of cell death from A2780 cells treated with only paclitaxel (High D) compared with cells treated with combination paclitaxel and celecoxib treatment (Low D) (Fig. 3.8B). The almost exponential increase of the cell death curve demonstrates that combination therapy with a CPT can improve chemotherapeutic efficacy, even in a limiting case of moderate cell death when cells are treated with smaller doses of chemotherapy (Fig. 3.8B). Altogether, this demonstrates that modulating the initial chromatin packing state of cancer cells influences their adaptability and can be harnessed to potentially increase cancer cell death in response to chemotherapy treatment.

3.7. Conclusions and Discussion

In this thesis work, we combined computational modeling with Hi-C, scRNA-seq, and imaging modalities from our nano-ChIA platform (see Chapter 1.3) to demonstrate the role of the disordered chromatin polymer on regulating intercellular transcriptional heterogeneity and transcriptional malleability which, in turn, influences adaptability to external stressors. Based on predictions from the CPMC model, which were verified experimentally, the chromatin packing within PDs affects gene expression through three key physical regulators: average density $\phi_{in,0}$, genomic size N_d, and packing scaling D (Fig. 3.1&3.2).

We demonstrate, both computationally and experimentally, that a crucial role of chromatin packing is to determine the level of phenotypic plasticity within a cell population. In particular, the scaling of chromatin packing, D, modulates both the transcriptional malleability (Fig. 3.3), and the level of intercellular transcriptional heterogeneity (Fig. 3.4). This effect is further regulated by other physical properties of chromatin. A higher average crowding density within the nucleus suppresses the expression of initially lowly-expressed genes as D increases (Fig. 3.2G&H). The modulatory effects of N_d are two-fold. Genes localized to domains with a larger N_d are more suppressed than those localized to domains with smaller N_d owing to the reduced accessibility to TFs and RNA Pol II. However, as D increases, the expression of genes associated with large N_d is disproportionately enhanced (Fig. 3.2I). Overall, higher D, higher average chromatin density, and lower N_d increase both transcriptional malleability and heterogeneity, with D having a much larger effect on transcriptional divergence compared to the other two chromatin packing properties (see Chapter 3.8.2). Using STORM-PWS colocalization, we

even observed that the chromatin packing scaling of a PD influences the extent of active transcription within the PD at the level of individual nuclei (Fig. 3.5).

Additionally, we established a link between the chemoevasion potential of cancer cells and their initial chromatin packing state (Fig. 3.6&3.7). Given that increased chromatin packing scaling was associated with chemoresistance (Fig. 3.13), we then explored if decreasing chromatin packing scaling could itself be a drugable target. To decrease D, we identified several pathways, including those that governed the chemical state of chromatin by altering epigenetic modifications and those that modified the physicochemical nanoenvironment (Fig. 3.16). Remarkably, we observed that drugs that modified the physicochemical environment (e.g., celecoxib and digoxin) had stronger effects on the chromatin packing scaling of cancer cell populations, across multiple cells lines, compared to drugs that modified the epigenetic state of chromatin. Additionally, the chemoadjuvant efficacy of these agents, as determined by the inhibition when co-treated with chemotherapy, is highly correlated with the degree to which they reduced D (Fig. 3.17). Finally, we determined that lowering the average D of cancer cell populations increases cell death upon exposure to the same dose of chemotherapy as higher D populations (Fig. 3.8).

Most of the existing anti-cancer drugs act via cytotoxic induction mechanisms. This cytotoxicity might be induced through a variety of pathways, whether it is through direct DNA damage (e.g., intercalating agents), the disruption of other cellular structures (e.g., inhibition of microtubule depolymerization), or the activation of the immune system attacking the tumor cells. Despite the wide range of molecular targets of these cytotoxic therapies, the primary reason why anti-cancer drugs fail is that cancer cells eventually develop resistance to almost all chemotherapeutic drugs. Drug resistance can occur

through a variety of mechanisms, including reduced drug accumulation and/or increased drug export, alterations in drug targets and signaling transduction molecules, repair of drug-induced DNA damage, and evasion of apoptosis [217]. Furthermore, experimental evidence has shown that tumor heterogeneity is a critical factor in primary drug-resistance (intrinsic resistance) as well as the emergence of new stress-induced drug-resistant clones (acquired resistance) [217, 273, 40, 39]. New gene mutations are not always necessary for drug resistance, and a change in the expression of existing genes due to transcriptional diversity [263] or transcriptional malleability may influence the ability of cancer cells to directly evade apoptosis [214] or activate compensatory pathways [234].

The fact that: (1) elevated D is a hallmark of cancer cells [9, 289, 19], (2) treating cells with chemotherapy selects for cancer cells with higher D (Fig. 3.3B&C, Fig. 3.12), and (3) D is associated with increased phenotypic plasticity in cancer cells which is (4) reduced upon co-treatment with a D-lowering drug agent, indicates a potentially crucial role of chromatin packing behavior in malignancy. Collectively, these findings indicate a means to identify the likelihood of chemoresistance to occur and to develop a novel class of adjuvant compounds that act at the level of chromatin packing scaling. This approach could lead to the creation of a low-cost personalized therapeutics, via screening with live-cell PWS microscopy, that limits the adaptability of each patient's tumor cells to assist in chemotherapeutic clearance. From a therapeutic standpoint, while mutations are difficult to remove from a cell population, this work suggests that limiting cancer cell evolution might be possible pharmacologically by lowering the chromatin packing scaling of cancer cell populations.

Although not explored in this work, there are several additional implications of these results on the understanding of multicellular fitness in the context of cell biology. For example, the localization of genes into higher-order domains has been demonstrated to be a conserved, albeit heterogeneous, process that can be disrupted in cancer [274, 82, 84]. In the context of chromatin PDs (see Chapter 2), our modeling predicts that cells would benefit from localizing genes into larger PDs that are intended to be suppressed at baseline but need rapid amplification if conditions change. Likewise, crowding density could be adjusted by cells either as a preprogrammed response by changing nuclear volume or incidentally from the retention of an extra chromosome during replication. Consequently, this could be a mechanism linking nuclear size and density (e.g. hyperchromasia) with differential gene expression. Interestingly, nuclear size, hyperchromasia, and abnormal nuclear texture are some of the most ubiquitous histological markers of neoplasia, although their etiology and functional consequences have been poorly understood [49].

In addition, this work may have implications on an open question in chromatin biology regarding the importance of non-coding DNA. Several roles of non-coding DNA have been illuminated, including the production of non-coding RNA and the distribution of transcriptional regulatory motifs such as enhancers and insulators, both of which are linked with cancer [17, 42]. In light of this work, and in relation to previously suggested hypotheses of the role of macromolecular crowding on gene expression, one of the evolutionary functions of non-coding DNA could be derived from its ability to exclude volume. Consequently, non-coding DNA might be a critical component within the genome to determine phenotypic plasticity as it contains the ability to modulate chemical transcription reactions by influencing their free-energy and the diffusion of transcriptional reactants.

At present, experimental validation of the CPMC model relies on the measurement of average PD structural properties. We are currently working on developing and implementing paired gene-tracking techniques, specifically CRISPR-Sirius [177], along with live-cell PWS microscopy and SR imaging of molecular factors. Such an imaging platform would enable a more precise characterization of how packing behavior at the level of individual PDs contributes to transcriptional malleability and intercellular heterogeneity. Although currently beyond technical abilities, relocalizing certain critical stress response genes to PDs with different packing behavior would prove the direct relationship between PD organization and responsiveness to cytotoxic stress.

Finally, one could also consider how D plays a role in the adaptability of cancer cells throughout carcinogenesis, which depends on cells overcoming aberrations in metabolism, inhospitable microenvironments, inadequate vascular supply, immune surveillance, and acclimation to distal tissue environments during metastasis. As it could take multiple replicative generations to develop a new useful mutation within a population for each of these processes, cancer cells could leverage the physical properties of chromatin packing to increase their transcriptional plasticity and acclimate to these conditions over a faster time scale. Thus, it may be worth investigating, for example, whether cancer cells with elevated D are better able to survive an immune response and acclimate to distant tissue sites during metastasis by performing imaging studies using nano-ChIA on patient samples.

3.8. Supplementary Material

3.8.1. Supplementary Tables

Parameters	Description	Value
$[C_{tot}]$	Total concentration of transcription complexes	[0.035 μM,350 μM]
K_D	Dissociation rate of Pol II in the absence of crowders	1 nm
k_m	Transcription rate of Pol II in the absence of crowders	1 $\mu M^{-1}s^{-1}$
r_{min}	Lower length scale of chromatin self-similarity	1 nm
L	Average gene length in bp	6 kbp
r_{in}^0	Radius of interaction volume for single bp	15 nm
N_d	Genomic size of chromatin PD	Average for all cell types: $\sim$1 Mbp Low N_d in A549 cells: 50 kbp High N_d in A549 cells: 2 Mbp
$\phi_{in,0}$	Average crowding density	HT-29 cells: 0.39 A549 cells: 0.40 v/v BJ cells: 0.31 v/v A2780 cells: 0.39 v/v
D_i	Initial chromatin packing scaling	Wild-type HT-29 cells: 2.7 HT-29 Arid-1a KD cells: 2.5 A549 cells: 2.66 BJ cells: 2.66 A2780 cells: 2.5

Table 3.1. Descriptions and values of CPMC model parameters. All other parameters for the crowding model are the same as in [184].

3.8.2. Analysis of Sensitivity Equations for D, N_d, and $\phi_{in,0}$

CPMC predicts the steady-state change in the rate of expression of a cell population exposed to a stressor as a function of the population's initial chromatin packing state. Thus, all three physical regulators of transcription determine transcriptional responsiveness, i.e. E_2/E_1. We can quantify the sensitivity of the responsiveness to each of these factors for initial chromatin packing state $D_i, N_d, \phi_{in,0}$ using the following analytical relations. Here, we denote the dependence of sensitivity on average expression rate and accessible surface as $Se_{\bar{\epsilon}}$ and Se_{p_g}, respectively.

Let us begin by defining the sensitivity of gene expression to D, Se_D:

$$Se_{\bar{\epsilon},D} = \frac{\partial ln\bar{\epsilon}}{\partial lnD} \approx -G(\bar{\epsilon})\left[D_i ln\left(\frac{r_{in}}{r_{min}}\right) + \frac{3-D_i}{D_i}\frac{r_{min}}{r_{in}}L^{1/D_i}lnL\right] \tag{3.26}$$

$$Se_{p_g,D} \approx \frac{\partial lnp_g}{\partial lnD} = \frac{1}{D_i}lnN_d \tag{3.27}$$

$$Se_D = Se_{\bar{\epsilon},D} + Se_{p_g,D} \tag{3.28}$$

where r_{min} is the radius of the elementary unit of chromatin (i.e. the DNA bp), r_{in} is the radius of the transcriptional interaction volume in, and L is the length of the gene being transcribed. For further simplification of equations, $G(\bar{\epsilon}) = \frac{\kappa}{8\bar{\epsilon}}(\sigma^2_{\phi_{in}})^2\left(1+\sqrt{1+\frac{16}{(\sigma^2_{\phi_{in}})^2}\frac{\bar{\epsilon}}{\kappa}}\right)$. Physiologically, κ exceeds physiologically relevant ranges of transcription and can be considered as the critical rate of expression such that for $\bar{\epsilon} < \kappa$ crowding has a significant effect on gene transcription. Overall all molecular factors $\vec{m}$, $\bar{\kappa} \approx 33.6\mu M/s$, while average $\bar{\epsilon} = 1\mu M/s$. Note that, in most cases $\frac{16}{\sigma^2_{\phi_{in}}} \gg 1$, so $G(\bar{\epsilon}) \approx \frac{1}{2}\sigma^2_{\phi_{in}}\sqrt{\frac{\kappa}{\bar{\epsilon}}}$. Thus, $G(\bar{\epsilon})$ is positively correlated with $\sigma^2_{\phi_{in}}$, which in turn increases with D, and is negatively correlated with $\bar{\epsilon}$.

Next, for domain size N_d:

$$Se_{\bar{\epsilon},N_d} = \frac{\partial ln\bar{\epsilon}}{\partial lnN_d} \approx -G(\bar{\epsilon})\left(1 - \frac{3}{D_i}\right)\left[\frac{1-2\phi_c}{1-\phi_c}\right] \tag{3.29}$$

$$Se_{p_g,N_d} \approx \frac{\partial lnp_g}{\partial lnN_d} = -\frac{1}{D_i} \tag{3.30}$$

$$Se_{N_d} = Se_{\bar{\epsilon},N_d} + Se_{p_g,N_d} \tag{3.31}$$

where $\phi_c = \phi_{in,0}\left(\frac{N_d}{\phi_{in,0}}\right)^{1-\frac{3}{D_i}}$.

And finally, for average crowding $\phi_{in,0}$:

$$Se_{\bar{\epsilon},\phi_{in,0}} = \frac{\partial ln\bar{\epsilon}}{\partial ln\phi_{in,0}} = -G(\bar{\epsilon})\left(\frac{3}{D_i} - 1\right)\left[\frac{1-2\phi_c}{1-\phi_c} + \frac{\phi_{in,0}}{L}\frac{r_{min}}{r_{in}}\right] \tag{3.32}$$

$$Se_{p_g,\phi_{in,0}} = \frac{\partial lnp_g}{\partial ln\phi_{in,0}} \approx \frac{1}{D_i} \tag{3.33}$$

$$Se_{\phi_{in,0}} = Se_{\bar{\epsilon},\phi_{in,0}} + Se_{p_g,\phi_{in,0}} \tag{3.34}$$

Next, we analyzed the effects of initial chromatin packing state on these sensitivity equations within physiological ranges for varying molecular factors. We determined the 25^{th}, 50^{th}, and 75^{th} percentiles of D, N_d, and $\phi_{in,0}$ from ChromSTEM packing domain analysis of unstimulated A549 lung adenocarcinoma cells (Fig. 3.9). Our analysis demonstrates that Se_D is much more sensitive to changes in D and molecular factors compared to both Se_{N_d} and $Se_{\phi_{in,0}}$ by 1 to 2 orders of magnitude (see Chapter 3.8.2).

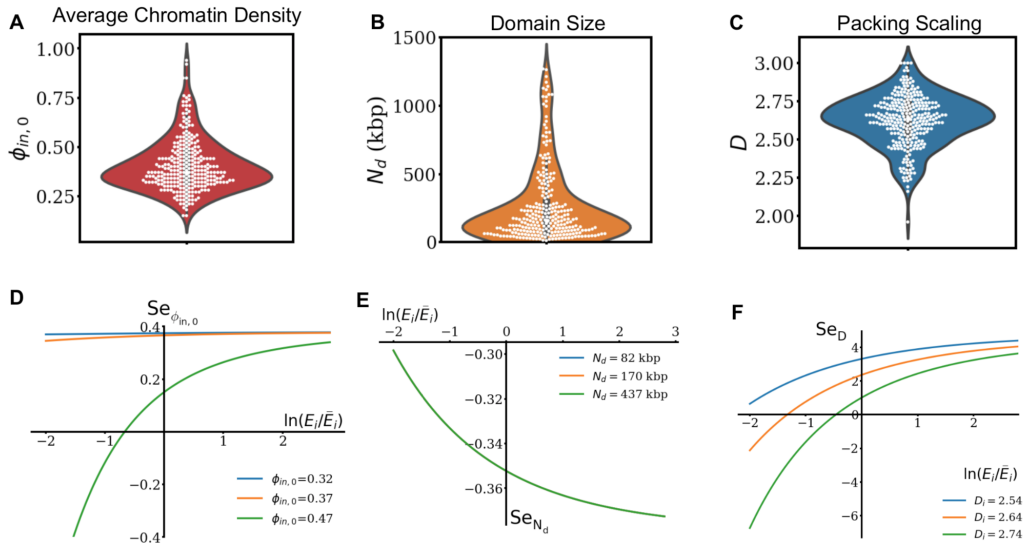

Figure 3.9. Sensitivity of gene expression is greatest for chromatin packing scaling compared to average density and genomic size of domains. (A-C) Statistical properties of PDs from A549 cells as determined by ChromSTEM analysis: (A) average PD chromatin density, $\phi_{in,0}$, (B) genomic size of PD, N_d, in kbp, and (C) chromatin packing scaling D. (E-G) Sensitivity of gene expression to (E) $\phi_{in,0}$, (F) N_d, and (G) D versus initial relative expression $ln(E_i)/ln(\bar{E}_i$ as determined by the CPMC model. The three different values for each sensitivity curve were determined from the 25^{th}, 50^{th}, and 75^{th} percentiles of PD properties determined from (A-C). Note that the range of sensitivity values for initial lowly expressed versus initial highly expressed genes occurs over a much greater range for Se_D compared to $Se_{\phi_{in,0}}$ and Se_{N_d}.

3.8.3. ChromTEM Analysis of A549 and BJ cells

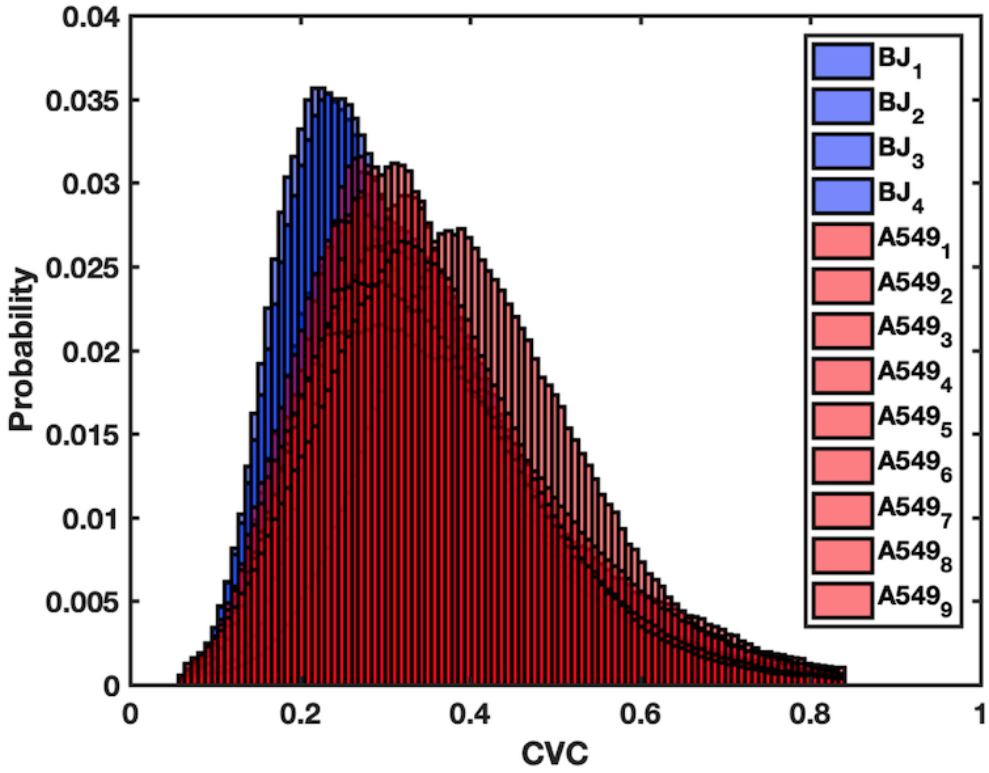

Figure 3.10. CVC distributions of A549 and BJ cells as measured by ChromTEM. Analysis of ChromTEM CVC values across N=4 replicates of differentiated BJ fibroblast nuclei and N=9 replicates of A549 lung adenocarcinoma nuclei. A549 nuclei have a pooled CVC average of 0.35 while BJ nuclei have a pooled CVC average of 0.30. These values represent the chromatin contribution to $\phi_{in,0}$.

3.8.4. Chromatin Packing Scaling Increases in Chemoevasive and Chemoresistant Cancer Cells

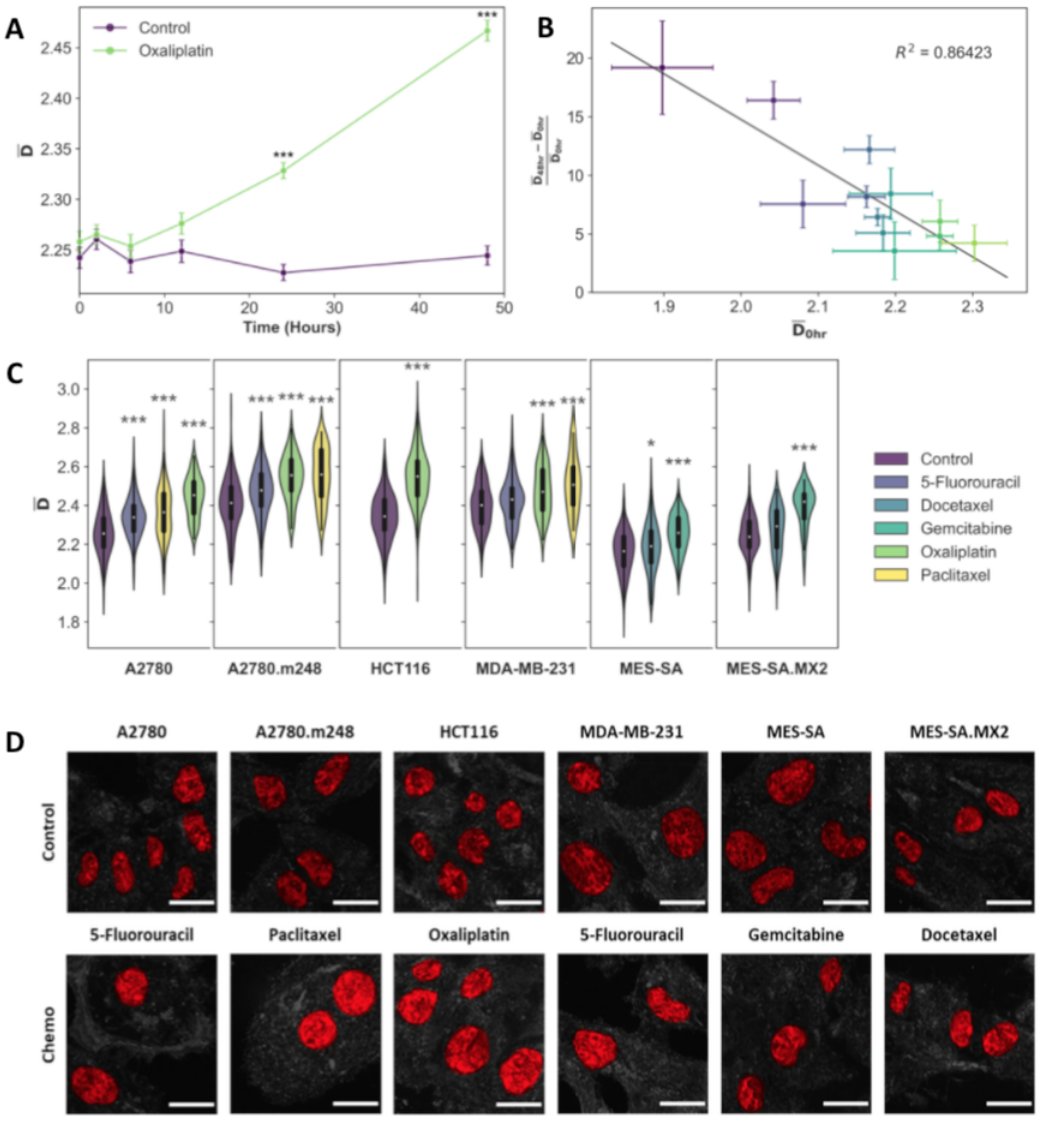

Figure 3.11. Chromatin packing scaling D increases in chemoevasive cells. (A) Average population of surviving cells steadily increases over the course of chemotherapeutic intervention with oxaliplatin in HCT116 cells. Error bars are standard error. (B) The relative increase in D of HCT116 cell clusters treated with oxaliplatin for 48 hours depends largely on the initial D of the cluster prior to treatment, with low D clusters experiencing the most pronounced change in D. Each point represents one cluster and error bars are standard error. (C) Exposure of cancer cell lines to standard concentrations of chemotherapeutic agents for 48 hours results in a shift in the population distribution of D to higher values in the surviving cells. Cells were treated with previously reported concentrations of these chemotherapeutics based on which chemotherapies are currently used as standard of care for a given malignancy. Violins correspond to control A2780 cells ($N = 332$), paclitaxel treated A2780 cells ($N = 99$, $P = 1.1 \times 10^{-3}$), 5-fluorouracil treated A2780 cells ($N = 147$, $P = 9.9 \times 10^{-21}$), and oxaliplatin treated A2780 cells ($N = 101$, $P = 2.6 \times 10^{-35}$); control A2780.m248 cells ($N = 259$), 5-fluorouracil treated A2780.m248 cells ($N = 100$, $P = 3.9 \times 10^{-3}$), paclitaxel treated A2780.m248 cells ($N = 45$, $P = 4.7 \times 10^{-6}$), and oxaliplatin treated A2780.m248 cells ($N = 85$,$P = 1.5 \times 10^{-18}$); control HCT116 cells ($N = 262$), and oxaliplatin treated HCT116 cells ($N = 289$, $P = 1.7 \times 10^{-35}$); control MDA-MB-231 cells ($N = 128$), 5-fluorouracil treated MDA-MB-231 cells ($N = 81$,$P = 4.1 \times 10^{-2}$), oxaliplatin treated MDA-MB-231 cells ($N = 59$, $P = 2.8 \times 10-5$), and paclitaxel treated MDA-MB-231 cells ($N = 36$,$P = 4.7 \times 10^{-5}$); control MES-SA cells ($N = 265$), docetaxel treated MES-SA cells ($N = 194$, $P = 2.0 \times 10^{-2}$), and gemcitabine treated MES-SA cells ($N = 101$, $P = 4.0 \times 10^{-13}$); control MES-SA.MX2 cells ($N = 203$), gemcitabine treated MES-SA.MX2 cells ($N = 103$, $P = 7.3 \times 10^{-6}$), and docetaxel treated MES-SA cells ($N = 106$, $P = 1.7 \times 10^{-8}$). Significance was determined using Student's t-test with unpaired, unequal variance on the average nuclear D of the treated group against the control group within each cell line ($^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$). (D) Representative PWS microscopy images of control and chemoevasive cells for each chemotherapy tested. Image pseudocolor is D, with brighter red corresponding to higher D values. Scale bars are 15 μm. A2780, A2780.m248, HCT116, MDA-MB-231, MES-SA, and MES-SA.MX2 were treated for 48 hours with 5-fluorouracil, paclitaxel, oxaliplatin, 5-fluorouracil, gemcitabine, and docetaxel respectively as representatives of all cell line and chemotherapy combinations.

For a mutationally induced model, we utilized ovarian A2780 wild-type (WT) cells with mutations to the TP53 DNA binding domain: A2780 TP53.m273 (A2780.m273), A2780 TP53.m175 (A2780.m175), and A2780 TP53.m248 (A2780.m248) mutant cells. TP53 mutations are present in over 95% of high-grade serous epithelial ovarian carcinoma (HGS EOC) and missense mutations at these codons (R273, R175, and R248) are the most predominant in HGS EOC patients. Additionally, these are gain-of-function mutations resulting in interactions between TP53 and the cohesin complex protein, Rad21, suggesting that they would be associated with alterations in chromatin structure. Under normal growth conditions, D was significantly higher in the A2780.m248 and A2780.m175 subclones than the A2780 WT line while D in the A2780.m273 subclone matched the WT line (Fig. 3.12A&B). When median survival time for HGS EOC patients with these mutations was compared using The Cancer Genome Atlas (TCGA) data, we found a strong correlation between median survival and D for each hot-spot mutation (Fig. 3.12C). Patients with the M273 mutation had the longest median survival (84.1 months) while patients with the M248 had the worst median survival (33.6 months) and those with the M175 mutation had an intermediate survival (62.1 months). While all these mutations produce a gain of function interaction with Rad21, their effect on chromatin packing scaling was inversely associated with the median patient survival duration.

Next, we explored whether inductive resistance in cell line models was likewise associated with increased chromatin packing density scaling D. To test this relationship, we utilized a uterine leiomyosarcoma line MES-SA with a mitoxantrone resistant MES-SA/MX2 subclone (MES-SA.MX2), which displays constitutive atypical topoisomerase II and MDR1 activity. Similar to the TP53 mutationally induced resistance models,

we observed that D was significantly increased in the MES-SA.MX2 resistant subclone compared to the MES-SA WT cell line (Fig. 3.14). As these results indicated that increased chromatin packing scaling correlates with patient prognosis and chemotherapeutic resistance, we hypothesized that chemotherapeutic intervention would confer a selection advantage on the population resulting in the surviving, chemoevasive population displaying increased D relative to the initial population that is maintained in cells with stable chemoresistance. Consequently, these findings confirmed that a convergence between both chromatin packing density scaling D and chemoevasion was cancer cell line and drug independent with cytotoxic intervention selecting for an increasingly heterogeneous chromatin packing state.

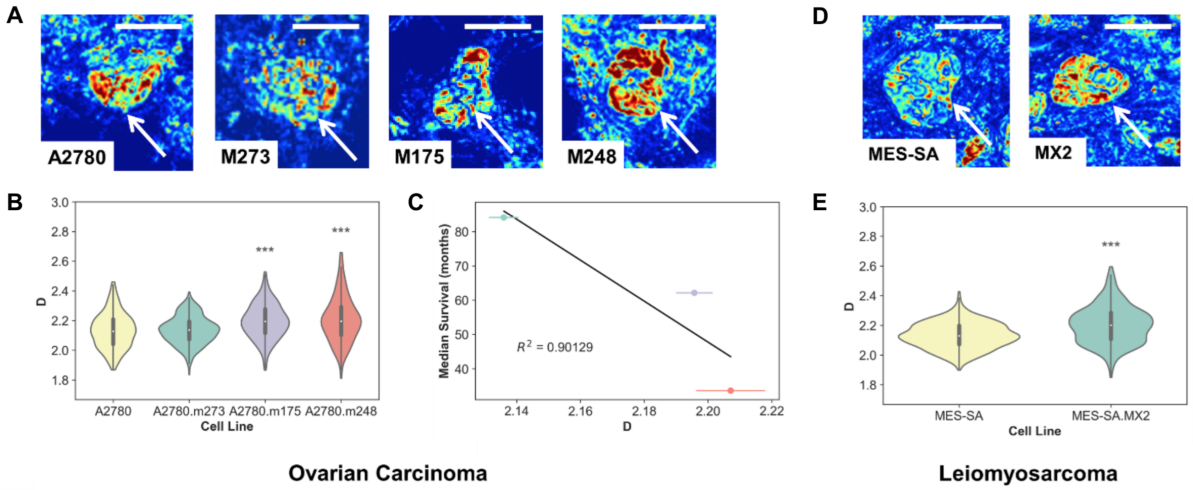

Figure 3.12. Chromatin packing scaling D increases with selective resistance to chemotherapy. (A) Representative PWS microscopy images of ovarian carcinoma A2780 wild-type (WT), and TP53-mutant A2780.m273 (M273), A2780.m175 (M175), and A2780.m248 (M248) cells. Arrows indicate representative nuclei. Scale bars, 10 mum. Pseudo-color: D. (B) Under normal growth conditions, D remained similar to the WT in the A2780.m273 subclone and increased in the A2780.m248 subclone ($P = 1.0 \times 10^{-59}$) relative to the WT A2780 cells. (C) Analysis of TCGA data for high-grade serious epithelial ovarian carcinoma patients revealed a strong correlation between median survival (as reported by TCGA) and D (as measured by PWS). (D) Representative PWS microscopy images of leiomyosarcoma MES-SA and mitoxantrone resistant MES-SA/MX2 derivative (MX2) cells. Arrows indicate representative nuclei. Scale bars, 10 mum. Pseudo-color: D. (E) Under normal growth conditions, D was increased in the MES-SA.MX2 chemoresistant subclone ($P = 3.1 \times 10^{-30}$) compared to the sensitive MES-SA subclone. Significance was determined using Student's t-test with unpaired, unequal variance on the average nuclear D of the mutant subclone against the WT within each cell line ($^{***}P < 0.001$). $N = 1877$ A2780, $N = 309$ M273, $N = 237$ M175, $N = 1321$ M248, $N = 836$ MES-SA, and $N = 558$ MX2 cells.

3.8.5. Supplementary Analysis for Transcriptional Malleability and Heterogeneity

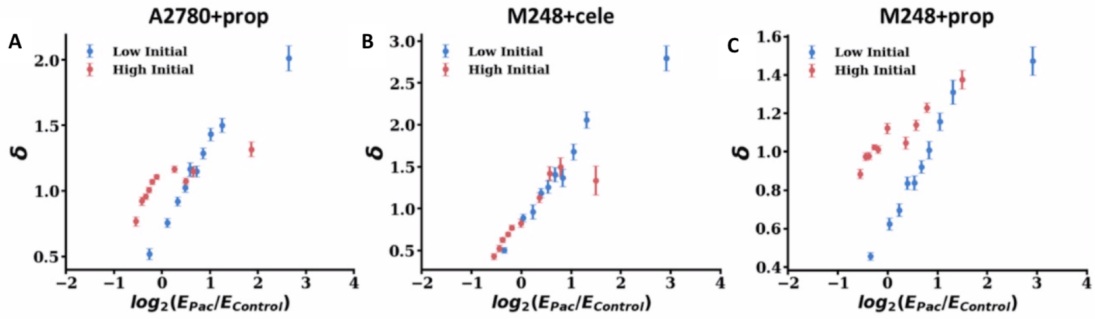

Figure 3.13. Increased transcriptional malleability for higher D cells is a generalizable phenomenon. The transcriptional malleability coefficient $\delta = \frac{E_{2,b}/E_{1,b}}{E_{2,a}/E_{1,a}}$ was determined from additional bulk RNA-seq experiments on A2780 cells and TP53 mutated clone A2780.m248 cells along with propranolol, another D-lowering compound. PWS measurements showed a 2% decrease in D in A2780 cells after propranolol treatment for 16 hours and a $\sim$5% decrease in D in m248 cells treated separately with celecoxib and then propranolol for 16 hours. (A) Transcriptional malleability in A2780 cells treated with propranolol to lower D. All treatment conditions include: control, 16 hour propranolol, 16 hour paclitaxel, and 16 hours paclitaxel plus celecoxib. (B&C) δ tested in m248 cells treated with (B) celecoxib and (C) propranolol as D-lowering compounds for 16 hours. All treatment conditions include control, 16 hours celecoxib/propranolol, 16 hours paclitaxel, 16 hours paclitaxel plus celecoxib/propranolol. All results are based on the expression data at $t = 16$ hours. Error bars represent the standard error of δ for all genes within each quantile. There are three biological replicates for every condition.

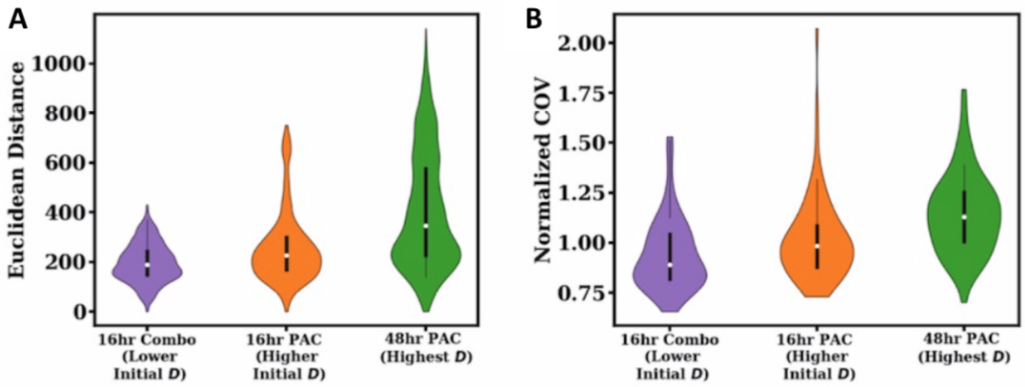

Figure 3.14. Transcriptional heterogeneity is increased in high-D cells. (A) Spread of pairwise Euclidean distance was calculated between cells in each condition for genes associated with DNA repair pathways that are upregulated in 48 hour paclitaxel treated cells. (B) Coefficient of variation (COV) across treatment populations of genes grouped by control expression levels normalized by control COV. Genes were first binned into groups of $\sim$100 genes (80 quantiles total) each based on relative control expression, which are assumed to be exposed to roughly similar molecular regulators of transcription. The expression of these genes was averaged within each cell. $COV_j = \sigma^2_{E_i}/\mu_{E_i}$ was calculated over all average expression levels of cells in treatment condition i for genes in control expression quantile j and each non-control condition was normalized to COV calculated for each bin in the control condition.

3.8.6. Effect of Decision Time on CDA Model

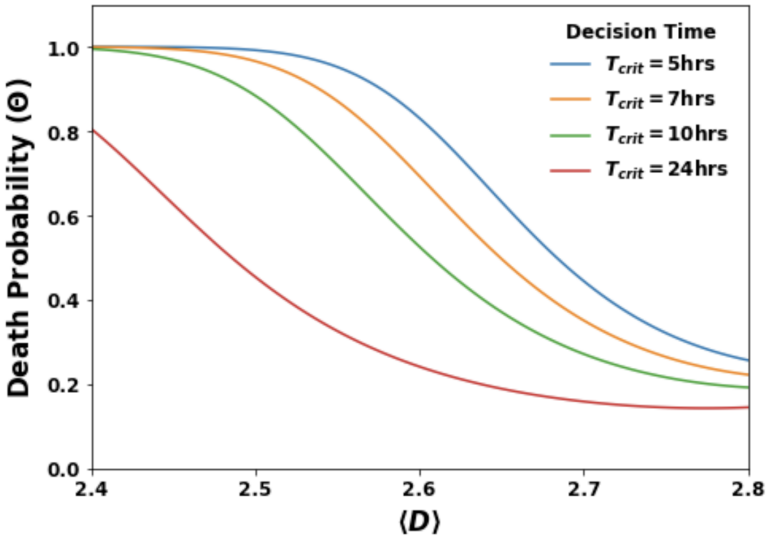

Figure 3.15. Decision time T_{crit} influences relationship between average packing scaling and, $\langle D \rangle$, and death probability, θ. Decision times were varied over commonly observed times to observe signs of apoptosis ($5 - 10$ hours) and up to 24 hours.

3.8.7. Identifying CPTs Across Multiple Cell Lines and Treatment Conditions

To determine the specificity of compound mechanisms of action for a given cell line, we tested five potential CPT agents of varying strength that could alter either histone modifications or the nuclear ionic content on nine additional cancer cell lines: ovarian cancer (A2780.m248, and OVCAR-8), pancreatic cancer (AsPC-1, and L3.6pl), colon cancer (HCT116), mesothelioma (M9K), breast cancer (MDA-MB-231), and leiomyosarcoma (MES-SA, and MES-SA.MX2) (Fig. 3.16C). For drugs impacting histone modifications, we tested valproic acid (VPA), which was identified in our initial screen as a moderate strength CPT that could potentially act through its role as an HDAC inhibitor, as well as 9-ING-41, which is a selective GSK-3 inhibitor. 9-ING-41 was chosen because chemotherapy treatment increased variations in GSK-3β expression in MDA-MB-231 cells and GSK-3β preferentially localizes to the nucleus in human cancer cells. Additionally, GSK-3β has previously been shown to play an important role in histone modifications involved in NFκB regulation [213]. For drugs that potentially alter nuclear ionic content through ion channel inhibition, the two strongest CPTs in the initial screen, celecoxib and digoxin, were chosen. We observed varying decreases in D in as little as 30 minutes (Fig. 3.16C). Of note, each compound had a slightly different effect in the ten cell lines. VPA, which was a moderate CPT in A2780, had a stronger effect on A2780.m248 cells than celecoxib. Digoxin, which was the strongest CPT in A2780, A2780.m248, and MDA-MB-231 cells, was not as effective at decreasing D compared to celecoxib in MES-SA and MES-SA.MX2 cells.

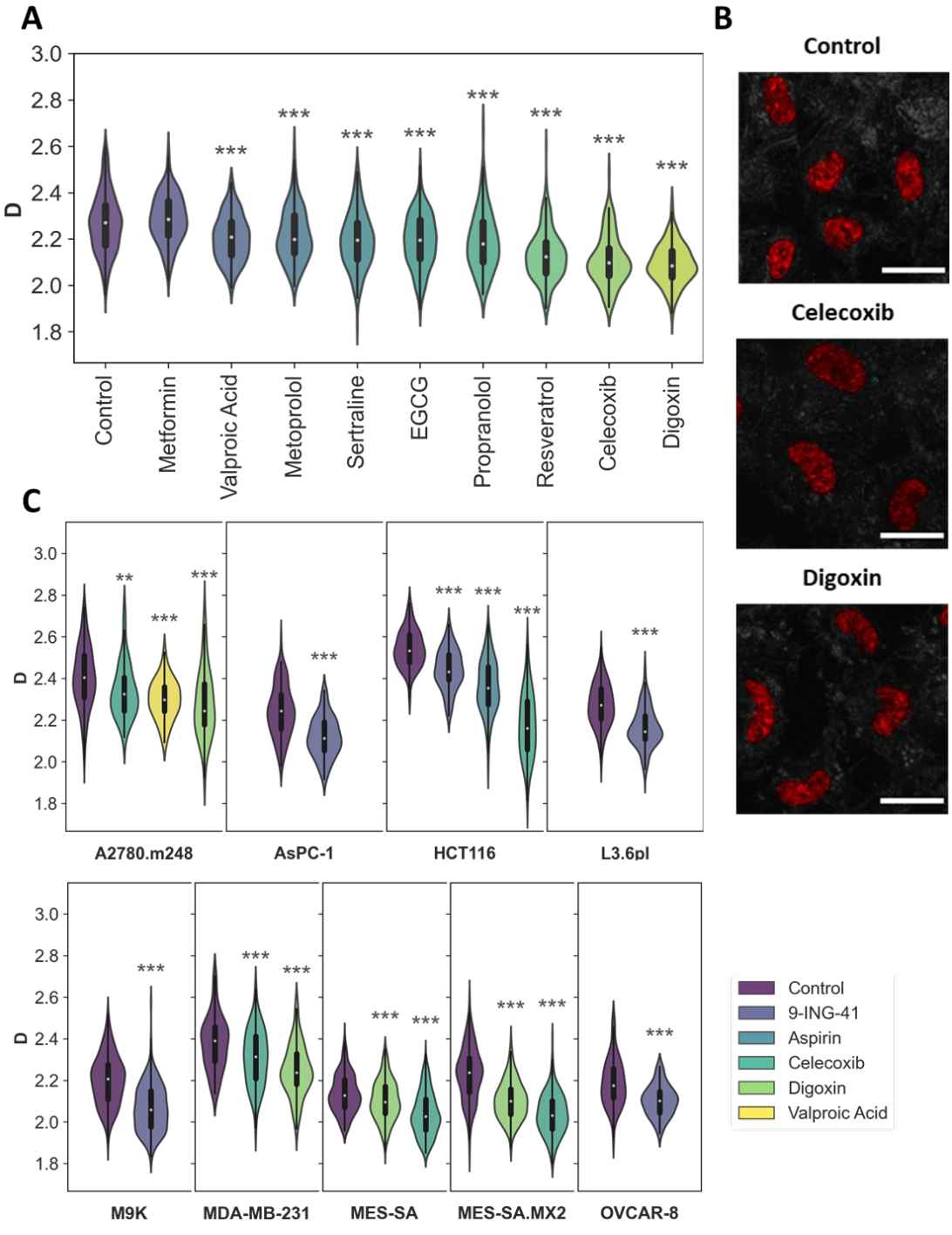
A
B
C
D
Control
Metformin
Valproic Acid
Metoprolol
Sertraline
EGCG
Propranolol
Resveratrol
Celecoxib
Digoxin
Celecoxib
Digoxin
A2780.m248
AsPC-1
HCT116
L3.6pl
M9K
MDA-MB-231
MES-SA
MES-SA.MX2
OVCAR-8
Control
9-ING-41
Aspirin
Celecoxib
Digoxin
Valproic Acid

Figure 3.16. Strong CPT agents can decrease chromatin packing scaling D independent of cancer cell line. (A) Drug treatments on A2780 cells caused varying levels of decrease in D. Violin plots correspond to A2780 cells treated with control ($N = 360$), insulin regulator metformin ($N = 195$, $P = 7.9\times10^{-2}$), HDAC inhibitor valproic acid ($N = 234$, $P = 1.4\times10^{-11}$), beta-blocker metoprolol ($N = 156$, $P = 674\times10^{-6}$), seratonin reuptake inhibitor sertraline ($N = 157$, $P = 3.7\times10^{-10}$), anti-oxidant green tea extract EGCG ($N = 276$, $P = 4.1 \times 10^{-15}$), beta receptor agonist propranolol ($N = 111$, $P = 1.4 \times 10^{-8}$), HDAC inhibitor resveratrol ($N = 271$, $P = 1.8 \times 10^{-43}$), non-steroidal anti-inflammatory drug celecoxib ($N = 132$, $P = 7.0 \times 10^{-34}$), and cardiac glycoside digoxin ($N = 572$, $P = 8.3 \times 10^{-86}$). Significance was determined using Student's t-test with unpaired, unequal variance on the average nuclear D of each treated group against the control group ($^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$). (B) Representative PWS microscopy images of control and CPT treated cells. Image pseudocolor is D, with brighter red corresponding to higher D values. Scale bars are 15 μm. A2780 cells were treated for 30 minutes with celecoxib and digoxin. (C) A 30-minute treatment with select CPT compounds decreases population-wide D in the nine different cell lines tested. Violin plots correspond to control A2780.m248 cells ($N = 125$), celecoxib-treated A2780.m248 cells ($N = 36$, $P = 9.3\times10^{-3}$), valproic acid-treated A2780.m248 cells ($N = 51$, $P = 4.9 \times 10^{-8}$), and digoxin-treated A2780.m248 cells ($N = 91$, $P = 1.5\times10^{-8}$); control AsPC-1 cells ($N = 107$), and 9-ING-41-treated AsPC-1 cells ($N = 111$, $P = 1.1 \times 10^{-14}$); control HCT116 cells ($N = 64$), 9-ING-41-treated HCT116 cells ($N = 85$, $P = 9.0 \times 10^{-10}$), aspirin-treated HCT116 cells ($N = 130$, $P = 9.0 \times 10^{-23}$), and celecoxib-treated HCT116 cells ($N = 75$, $P = 5.79 \times 10^{-35}$); control L3.6pl cells ($N = 150$), and 9-ING-41-treated L3.6pl cells ($N = 163$, $P = 1.6 \times 10^{-19}$); control M9K cells ($N = 191$), and 9-ING-41-treated M9K cells ($N = 224$, $P = 1.1 \times 10^{-27}$); control MDA-MB-231 cells ($N = 89$), celecoxib-treated MDA-MB-231 cells ($N = 86$, $P = 5.6\times10^{-4}$), and digoxin-treated MDA-MB-231 cells ($N = 83$, $P = 1.2 \times 10^{-10}$); control MES-SA cells ($N = 314$), digoxin-treated MES-SA cells ($N = 342$, $P = 1.7 \times 10^{-5}$), and celecoxib-treated MES-SA cells ($N = 275$, $P = 1.5\times10^{-29}$); control MES-SA.MX2 cells ($N = 227$), digoxin-treated MES-SA.MX2 cells ($N = 252$, $P = 1.3 \times 10^{-30}$), and celecoxib-treated MES-SA cells ($N = 216$, $P = 8.8 \times 10^{-55}$); control OVCAR-8 cells ($N = 65$), and 9-ING-41-treated OVCAR-8 cells ($N = 82$, $P = 4.2\times10^{-7}$). Significance was determined using Student's t-test with unpaired, unequal variance on the average nuclear D of the treated group against the control group within each cell line ($^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$).

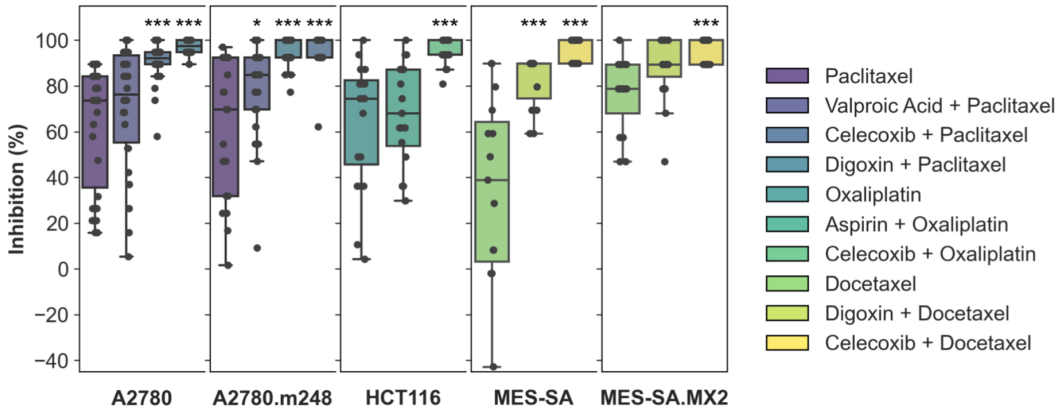

Figure 3.17. Inhibition of cancer cells by chemotherapy increases upon addition of "strong" CPTs. To examine the relationship between cell death and initial chromatin packing scaling experimentally, we focused on three complementary methods for assessing cell survival – cell coverage analysis using transmission microscopy, which provided information on total population viability, automated cell counting paired with dead and apoptotic cell stains, and flow cytometry measurement of live cell caspase 3/7 staining, which captured the relative induction of apoptosis. Altogether, these measures provided quantifiable information on cell inhibition as well as percentage viability for each population. Mild CPTs, which cause a relatively smaller decrease in D (valproic acid and aspirin) are less effective at improving chemotherapeutic efficacy than "strong" CPTs (celecoxib and digoxin). Error bars are standard error. Significance was determined using Student's t-test with unpaired, unequal variance on the inhibition of the co-treated group against the chemotherapy treated group within each cell line ($^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$).

CHAPTER 4

The Physicochemical Intranuclear Environment Modulates Structure and Charge of DNA and Chromatin

4.1. Introduction

DNA, the biological material that contains our genetic information as a unique sequence of nucleotides, is a highly charged polyelectrolyte due to the negative charge it carries from its phosphate diester backbone [35]. As a polyelectrolyte, the structure, conformations, and also charge of DNA will change depending on the bulk physicochemical environment, including the electrolyte environment, pH, and average polyelectrolyte density. While the behavior of DNA in dilute electrolyte solutions is widely studied and fairly well understood, the behavior of DNA in denser environments is far less characterized.

A prototypical and biologically relevant example is chromatin, the macromolecular assembly of DNA and histone proteins that are compacted and packed into the eukaryotic cell nucleus.[1] Chromatin is a negatively charged system, where positively charged residues of the histone proteins only partially mitigate the highly negative charge of the phosphates in the DNA backbone [95, 183]. Chromatin itself is relatively densely packed inside the cellular nucleus, with volume fractions ranging from 12% to 52% [212]. Our computational modeling has demonstrated the direct relationship between chromatin packing, transcription, and phenotypic plasticity of cancer cells (see Chapter 3). Experimentally, we have determined that drugs which influence the intracellular physicochemical environment, such as celecoxib and digoxin, have a larger influence on modulating chromatin packing behavior and increasing chemotherapeutic efficacy than other drugs, even those that modulate the epigenetic state of chromatin (Fig. 3.15,3.16). Additionally, intracellular ion concentrations can be altered in diseases such as cancer and the extracellular environment can also become acidic, potentially modulating intracellular pH [220, 228, 79, 137]. Thus,

[1]Note that DNA is also densely packed in the histone-less bacterial nucleus [11].

we want to determine key biophysical mechanisms that can be used to predictably modulate chromatin structure and function, and our hypothesis is that the physicochemical environment is one such mechanism. Here, we shall formulate a theoretical approach to appropriately describe the effects of the physicochemical environment on systems that represent DNA and chromatin in non-dilute, denser systems.

Experimental methods, including single-molecule studies in dilute systems, have determined that structural properties, ranging from DNA persistence length to compaction of chromatin at the ∼Mbp scale, are highly dependent on both the strength and composition of the bulk electrolyte environment [20, 309, 310]. Modulating Mg^{2+} concentration has even been shown by Tanase et al. to influence differentiation efficiency in mouse embryonic stem cells [120]. Additionally, recent experimental advances have facilitated a more complete characterization of the ionic atmosphere surrounding DNA, including Anomalous X-ray Scattering [60] and, more recently, Inductively Coupled Plasma Mass Spectrometry (ICP-MS) [94, 93, 95]. These techniques demonstrate a strong dependence of DNA charge screening on ion type and strength [94].

However, experiments are also unable to concomitantly measure structure and charge of DNA and chromatin systems, making it even more difficult to establish principle mechanisms of action. Additionally, such experiments are either performed in dilute, *ex vivo* systems, which are not relevant to dense physiological conditions of DNA packed in a nucleus, or *in vitro* cellular models, making it difficult to disentangle the charge screening effects of ions on DNA and chromatin from other effects due to signaling and enzymatic action. Recent work by Zinchenko et al. observe coil-globule phase transitions of λ-DNA in only monovalent electrolyte environments when external crowders are added

to solution, further pointing to the importance of properly accounting for dense intranuclear environments when considering the effects of the electrolyte environment on DNA structure [311, 309].

Flexible polyelectrolyte solutions containing both monovalent ions as well as multivalent ions have been extensively studied using theoretical and computational approaches encompassing density functional theory, Poisson-Boltzmann approaches, scaling theory, and simulations [256, 122, 37, 211, 267, 172, 114, 304]. Specifically, DNA in solution has received considerable attention. For example, both MD simulations [172] and theoretical arguments [34] have predicted collapse of extended DNA in the presence of multivalent counterions induced by ion condensation. In general, and especially for dense systems, theoretical approaches make certain assumptions about the chain-statistics. For example, they use lattice chains or assume Gaussian chain-statistics [122, 123] or assume specific three-dimensional structures of polyelectrolyte chain conformations [37, 267, 10]. On the other hand, MD simulations allow for a more explicit study of the conformations of DNA. However, the inherent tradeoff in simulations between computational feasibility and level of detail is a current barrier for a more complete understanding of charge regulation mechanisms in physiologically dense systems (see Chapter 1.4.2).

Here, we present a theoretical MT model to first study a DNA-like polyelectrolyte brush of phosphates and then a chromatin-like system. This study is performed in both dilute conditions and at higher densities, all of which include the chemical equilibrium between the protonated, deprotonated, and ion-condensed states of the chargeable DNA-phosphates and histone amino acids. Importantly, the theory does not assume the

charged state of the DNA-phosphates and amino acid residues *a priori*, but rather predicts the position-dependent state of charge. The theory is based on a molecular statistical thermodynamic approach that has previously been developed to predict thermodynamic and structural properties of end-tethered polymers and weakly ionizable polyelectrolytes [201, 100]. Predictions of the MT have been found to agree with experimental observations for relevant biological systems [189, 254, 299, 278, 269, 239, 291, 156]. Importantly, MT takes as input a representative set of polymer chains and calculations determine the probability of each conformation under specific environmental conditions.

Herein, we would like to address the effects of charge regulation on DNA-like and chromatin-like systems with varying densities. We first focus our study on end-tethered loops of phosphates, as DNA experiences many topological constraints in the nucleus, including CTCF-cohesin-induced looping, interactions with nuclear lamins, and DNA wrapping around histone proteins to form nucleosomes [251, 62, 118, 113]. Such constraints are expected to make the DNA polymer more flexible, which is usually very rigid in its double-stranded form. Additionally, we further simplify the system by only explicitly modeling the DNA-phosphates, which are the only chargeable molecule of double-stranded DNA. Related questions to be studied involve the effects of physiologically relevant monovalent versus divalent cations on these charge regulation mechanisms, and how these effects are modulated by bulk density. Specifically, we focus on the effects of the most prevalent intracellular monovalent (K^+ and Na^+) and divalent (Mg^{2+}) cations [171]. From our simpler phosphate loop system, we also determine physiologically relevant constants for phosphate-ion condensation, which are difficult to determine experimentally, before increasing the complexity of our system to chromatin. Next, we extend our MT approach

to characterize the effects of monovalent cations on the charge and structure of single nucleosomes and nucleosome arrays under varying bulk densities. To our knowledge, this is the first study that considers the variable charge of phosphates and amino acids and variable local pH to investigate the effects of the physicochemical environment on DNA and chromatin in dense systems via charge regulation mechanisms.

4.2. Theoretical Approach

4.2.1. Molecular Theory

Here, we investigate a DNA-like and a chromatin-like system. First, we model an end-tethered polyelectrolyte brush composed of DNA-phosphates in contact with an aqueous solution for both dilute and non-dilute conditions (Fig. 4.1). We then increase the complexity of the system and model a chromatin-like system of both single nucleosomes and 8-mer nucleosome arrays (Fig. 4.2).

The reservoir of our polymeric system is characterized by a given pH and contains monovalent KCl and NaCl as well as divalent $MgCl_2$ salt at given concentrations. The salts are assumed to be completely dissociated. These salts were chosen because they are the most prevalent intracellular ions [171]. The pH is adjusted by adding either HCl or NaOH to the system. The DNA-phosphates are assumed to be in one of six chemical states: deprotonated (P^-), protonated (PH) or condensed with K^+, Na^+, or Mg^{2+} counterions. The following chemical reactions are explicitly included in the theory for both the phosphate loop and chromatin-like systems

$$\mathrm{PH} \rightleftharpoons \mathrm{P^-} + \mathrm{H^+}, \tag{4.1}$$

$$\mathrm{P^-} + \mathrm{Na^+} \rightleftharpoons \mathrm{PNa}, \tag{4.2}$$

$$\mathrm{P^-} + \mathrm{K^+} \rightleftharpoons \mathrm{PK}, \tag{4.3}$$

$$\mathrm{P^-} + \mathrm{Mg^{2+}} \rightleftharpoons \mathrm{PMg^+}, \tag{4.4}$$

$$2\,\mathrm{P^-} + \mathrm{Mg^+} \rightleftharpoons \mathrm{P_2Mg}, \tag{4.5}$$

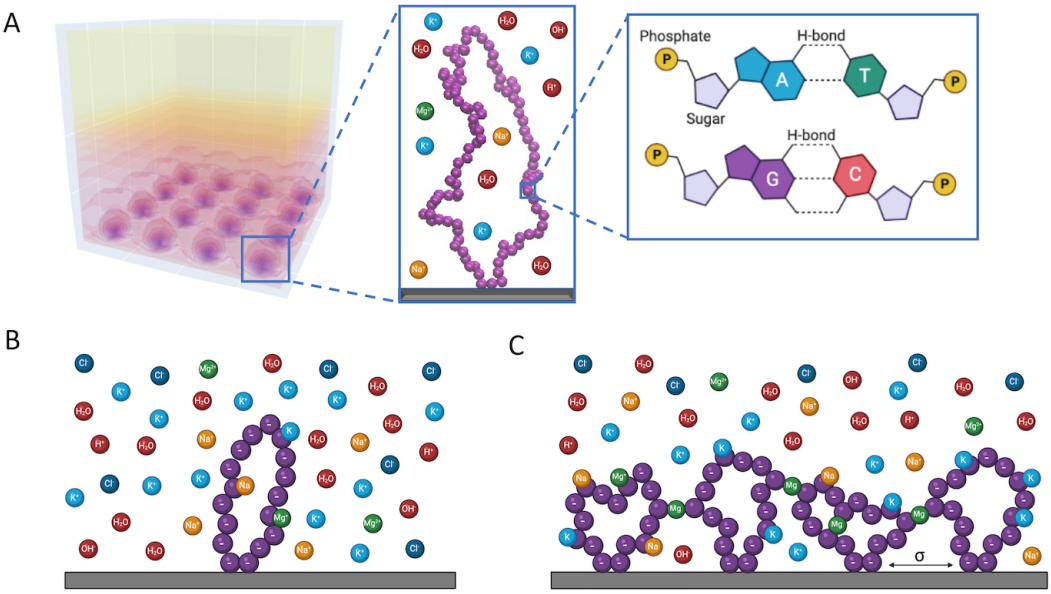

Figure 4.1. Description of phosphate loop system. (A) Representation of an end-grafted planar layer of loops, where each monomer has the chemical properties of the phosphate molecule of DNA. Phosphate loops are composed of 100 monomers. Our system explicitly contains the most prevalent intracellular or intranuclear ions, including Na^+, K^+, Mg^{2+}, and Cl^- as well as water, OH^- and H^+ to account for charge regulation effects of ions and acid-base equilibrium. We consider both (B) dilute and (C) denser systems that are more relevant to physiological DNA concentrations in the nucleus. The bulk density of the system is determined by the grafting density σ, which controls spacing between the graft-points of the loops. Note (B) & (C) are example conformations and a large range of conformations are considered for MT calculations.

The differences in the effects of cations on the charge regulation of phosphates are influenced by a combination of the explicitly defined volumes of the ions, their mechanisms of binding, and the binding strength (i.e., $\Delta G_d^{\ominus}$) of the ion condensation reactions. The condensed states for monovalent cations are denoted as PNa, PK. For Mg^{2+}, there are two condensed states representing 1:1 binding, PMg^+, and 2:1 binding, P_2Mg, of the phosphates with the divalent cations. The last reaction, the 2:1 binding, or ion bridging

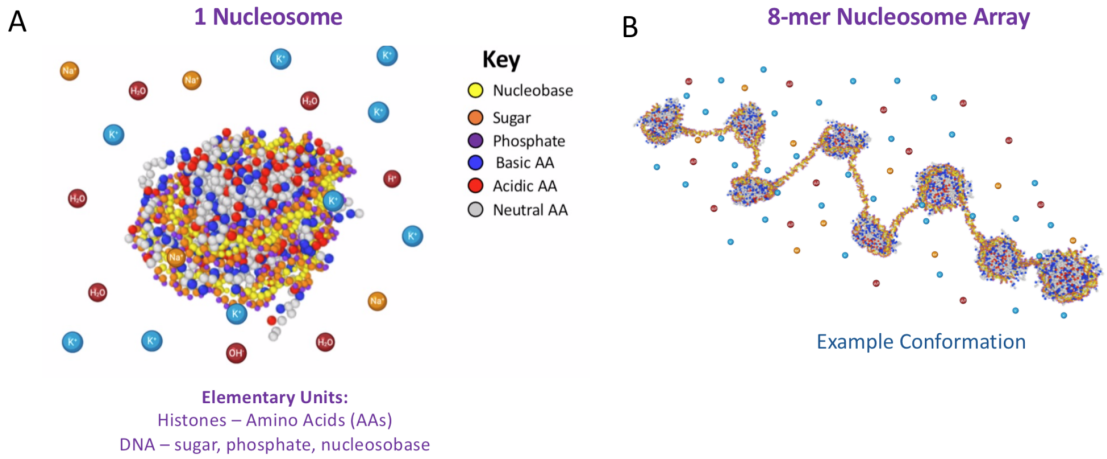

Figure 4.2. Description of chromatin-like system for (A) single nucleosomes and (B) 8-mer nucleosome arrays. Our chromatin system also explicitly contains the most prevalent intracellular ions, including Na^+, K^+, and Cl^- as well as water, OH^- and H^+ to account for charge regulation effects of ions and acid-base equilibrium. (A) Rendering of single, tailless nucleosome that is coarse-grained to the molecular level from the human 1KX5 crystal structure [61]. Elementary units of histones include basic, acidic, and neutral amino acids and basic units of DNA are represented by the 3SPN model [88], where each nucleotide is represented by one sugar, one phosphate, and one nucleobase. (B) Rendering of an example conformation for the 8-mer nucleosome chain system composed of 8 nucleosomes connected by linker DNA. The density of this system is controlled by the box size, which is an input into the MT calculations.

reaction, does not occur directly but via $P^- + PMg^+ \rightleftharpoons P_2Mg$. However, these reactions are thermodynamically equivalent. We have not considered the formation of ion pairs involving multiple divalent cations and phosphates simultaneously. Thus, monovalent cations can only bind with one phosphate, while divalent cations can form additional ion bridges with a stochiometry of two phosphates to one Mg^{2+} ion.

Additionally, for the chromatin-like systems, we include the acid-base equilibrium of chargeable amino acid residues of the histone proteins

$$\mathrm{RCOOH} \rightleftharpoons \mathrm{RCOO^-} + \mathrm{H^+}, \tag{4.6}$$

$$\mathrm{ROH} \rightleftharpoons \mathrm{RO^-} + \mathrm{H^+}, \tag{4.7}$$

$$\mathrm{RNH_2^+} \rightleftharpoons \mathrm{RNH} + \mathrm{H^+}, \tag{4.8}$$

$$\mathrm{RNH_3^+} \rightleftharpoons \mathrm{RNH_2} + \mathrm{H^+}, \tag{4.9}$$

$$\mathrm{RSH} \rightleftharpoons \mathrm{RS^-} + \mathrm{H^+}, \tag{4.10}$$

where Eq. 4.6 refers to acid-base equilibrium for aspartic acid (Asp) and glutamic acid (Glu), Eq. 4.7 refers to acid-base equilibrium for the acidic tyrosine (Tyr) residue, Eq. 4.8 refers to acid-base equilibrium for the basic arginine (Arg) and histidine (His) residues, Eq. 4.9 refers to acid-base equilibrium for the basic lysine (Lys) residue, and Eq. 4.10 refers to acid-base equilibrium for the basic cystein (Cys) residue. Explicitly incorporating these mechanisms of ion pairing provides a clearer understanding of charge regulation mechanisms which occur in DNA-like and chromatin-like systems.

Molecular Theory (MT) is a statistical-thermodynamics approach with a free energy functional that explicitly takes into account the conformations of the polymer chains and the size, shape, and charge of all molecular species in the system. The free energy describing the polyelectrolyte of interest in contact with an aqueous electrolyte solution

has a number of distinct contributions, which can be summarized as follows

$$\begin{aligned} F &= -TS_{\text{conf}} - TS_{\text{mix}} + F_{\text{chem}} \\ &\quad + F_{\text{elect}} + E_{\text{elect,solv}} + E_{\text{VdW}} + E_{\text{rep}}. \end{aligned} \tag{4.11}$$

The first contribution (S_{conf}) is related to the conformational entropy of the phosphate/chromatin chains. The second term encompasses the mixing or translational entropy of the mobile ions and solvent (S_{mix}). The next three contributions stem from the acid-base chemical equilibrium of the phosphates and amino acids and the counterion condensation of phosphates (F_{chem}), the electrostatic interaction energy (F_{elect}), and the electrostatic solvation energy of the charged ionic species ($E_{\text{elect,solv}}$). The next term encompasses the effective Van der Waals, or hydrophobic interactions, among the units of the system. For all calculations in this work, effective Van der Waals interactions are assumed to be zero as DNA and chromatin are highly charged polyelectrolytes, so hydrophobic effects are expected to be small compared to electrostatic and osmotic effects. The last term, E_{rep} accounts for the steric repulsions, or excluded volume interactions, among all molecular species.

For the sake of brevity, we discuss here only the two most salient features of the free energy: namely, the conformational entropy and the free energy contribution pertaining to the acid-base equilibrium and ion condensation. The other terms, such as the translational entropy of solvent and mobile ions as well as the electrostatic energy terms have been discussed in previous works and are briefly discussed in Chapter 4.6.1. A complete description of the free energy functional, including the explicit free energy

contributions related to electrostatics interactions and excluded volume are presented in the supporting material as well as in references [100, 200].

$-TS_{\text{conf}}$ describes the conformational entropy of the polyelectrolyte chains and is given by

$$-\frac{S_{conf}}{k_B} = \sum_g \sum_\alpha P_g(\alpha) \ln P_g(\alpha) \tag{4.12}$$

Here $P_g(\alpha)$ is the probability of chain conformation α that is end-tethered to location r_g. Since $P_g(\alpha)$ represents a probability distribution function (p.d.f.) over all conformations, $\sum_\alpha P_g(\alpha) = 1$ for all graft points. Note that there is only one fixed conformation representing the single nucleosome, but there are 100,000s of 8-nucleosome chromatin chains input into the 8-mer MT chromatin calculations and tens of millions of loop conformations input into the phosphate loop MT calculations.

$P_g(\alpha)$ is the central quantity of the theory, because once its value is established, we can compute any structural and thermodynamic quantity of interest related to the polyelectrolyte system. For example, given $P_g(\alpha)$, we can calculate the number density of phosphates for our loop system for each position in the lattice $\overrightarrow{r}$

$$\langle \rho_{DNA-P}(\overrightarrow{r}) \rangle = \sum_\alpha P(\alpha) n(\alpha; \overrightarrow{r}) v_{DNA-P} \tag{4.13}$$

for $n(\alpha; \overrightarrow{r})$ the number of elementary units (e.g., DNA-phosphates) from conformation α that can be found in volume element $[\overrightarrow{r}, \overrightarrow{r} + \overrightarrow{dr}]$ and v_{DNA-P} the volume of each elementary unit. The polyelectrolyte volume fraction can also be calculated using the relation $\langle \phi_{DNA-P}(\overrightarrow{r}) \rangle = \frac{\langle \rho_{DNA-P}(\overrightarrow{r}) \rangle}{\delta^3}$ for lattice size δ. For our phosphate loop system the

elementary units are DNA-phosphates. This formalism can be extended to our chromatin-like system where the densities (ρ), volumes (v), and volume fractions (ϕ) can be calculated for each of the different types of elementary units as detailed in Tables 4.1 & 4.2 (Fig. 4.2)).

Here, a representative set of conformations, generated using MD simulations, are input into the MT. Chain generation is described in greater detail in Chapters 4.2.2 & 4.6.2. Observe that the internal excluded volume interaction of the phosphate loops and chromatin elements are explicitly accounted for, as MD simulations were performed such that all chains are self-avoiding. The intermolecular excluded volume interactions, E_{rep}, are represented by a mean-field treatment. Specifically, we assume that the system is incompressible at every position:

$$\langle \phi_{poly}(\overrightarrow{r}) \rangle + \phi_w(\overrightarrow{r}) + \sum_k \phi_k(\overrightarrow{r}) = 1. \tag{4.14}$$

where $\phi_{poly}(\overrightarrow{r})$ represents the position-dependent volume fraction occupied by the polyelectrolyte (phosphates for the phosphate loop system or all elementary units from Tables 4.1 & 4.2 for the chromatin system), $\phi_w(\overrightarrow{r})$ represents the position-dependent volume fraction occupied by water, and $\phi_k(\overrightarrow{r})$ represents the position-dependent volume fraction occupied by mobile ions. These volume packing constraints are enforced through the introduction of the Lagrange multipliers $\pi(\overrightarrow{r})$ since these are constraints, they are formally not part of the Helmholtz free energy. Note that, although MT is a mean-field approach, the explicit inclusion of polymer conformations implies that intra-chain correlations are considered.

For the phosphate loop system, the term F_{chem}, describes the chemical free energy associated with (de)protonation of the phosphates and the ion condensation of K^+, Na^+, and Mg^{2+} ions.

$$
\begin{aligned}
(4.15)\quad \beta F_{\text{chem}} = \int \mathrm{d}^3 r \langle \rho_{DNA-P}(\overrightarrow{r}) \rangle \Big[& \\
& f_{P^-}(\overrightarrow{r})(\ln f_{P^-}(\overrightarrow{r}) + \beta \mu^{\ominus}_{P^-}) \\
& + f_{PH}(\overrightarrow{r})(\ln f_{PH}(\overrightarrow{r}) + \beta \mu^{\ominus}_{PH}) \\
& + f_{PNa}(\overrightarrow{r})(\ln f_{PNa}(\overrightarrow{r}) + \beta \mu^{\ominus}_{PNa}) \\
& + f_{PK}(\overrightarrow{r})(\ln f_{PK}(\overrightarrow{r}) + \beta \mu^{\ominus}_{PK}) \\
& + f_{PMg^+}(\overrightarrow{r})(\ln f_{PMg^+}(\overrightarrow{r}) + \beta \mu^{\ominus}_{PMg^+}) \\
& + \frac{1}{2} f_{P_2Mg}(\overrightarrow{r})(\ln \frac{1}{2} f_{P_2Mg}(\overrightarrow{r}) + \beta \mu^{\ominus}_{P_2Mg}) \\
& - \frac{1}{2}(f_{P_2Mg}(\overrightarrow{r}))(\ln \langle \rho_{DNA-P}(\overrightarrow{r}) \rangle v_w - 1) \Big] \\
& + \sum_{k \in \{H^+, OH^-, Na^+, K^+, Mg^{2+}, Cl^-\}} \beta \mu^{\ominus}_k \int \mathrm{d}^3 r \rho_k(\overrightarrow{r}).
\end{aligned}
$$

Here, $f_{P^-}(\overrightarrow{r})$ is the fraction of phosphate acid residues that are charged or deprotonated at position $\overrightarrow{r}$, $f_{PH}(\overrightarrow{r})$ is the fraction of neutral, protonated phosphate acids, and $f_{PNa}(\overrightarrow{r})$, $f_{PK}(\overrightarrow{r})$, $f_{PMg^+}(\overrightarrow{r})$ are the fraction of phosphate acids that are condensed with Na^+, K^+, and Mg^{2+}, respectively. Lastly, $f_{P_2Mg}(\overrightarrow{r})$ is the fraction of phosphate groups that form a Mg^{2+} bridge, i.e., a complex of two phosphates and one Mg^{2+} counterion. Complexes involving multiple ions and phosphates are not considered.

For the chromatin-like systems, where we have so far only considered monovalent cations, we also include the (de)protonation of amino acid residues:

(4.16)

$$\begin{aligned}
\beta F_{\text{chem}} = \int \mathrm{d}^3 r \langle \rho_{DNA-P}(\overrightarrow{r}) \rangle \Big[& f_{P^-}(\overrightarrow{r})(\ln f_{P^-}(\overrightarrow{r}) + \beta\mu^{\ominus}_{P^-}) + f_{PH}(\overrightarrow{r})(\ln f_{PH}(\overrightarrow{r}) + \beta\mu^{\ominus}_{PH}) \\
& + f_{PNa}(\overrightarrow{r})(\ln f_{PNa}(\overrightarrow{r}) + \beta\mu^{\ominus}_{PNa}) + f_{PK}(\overrightarrow{r})(\ln f_{PK}(\overrightarrow{r}) + \beta\mu^{\ominus}_{PK}) \Big] \\
+ \int \mathrm{d}^3 r \langle \rho_{AA-Asp}(\overrightarrow{r}) \rangle \Big[& f_{Asp^-}(\overrightarrow{r})(\ln f_{Asp^-}(\overrightarrow{r}) + \beta\mu^{\ominus}_{Asp^-}) + f_{AspH}(\overrightarrow{r})(\ln f_{AspH}(\overrightarrow{r}) + \beta\mu^{\ominus}_{AspH}) \Big] \\
+ \int \mathrm{d}^3 r \langle \rho_{AA-Glu}(\overrightarrow{r}) \rangle \Big[& f_{Glu^-}(\overrightarrow{r})(\ln f_{Glu^-}(\overrightarrow{r}) + \beta\mu^{\ominus}_{Glu^-}) + f_{GluH}(\overrightarrow{r})(\ln f_{GluH}(\overrightarrow{r}) + \beta\mu^{\ominus}_{GluH}) \Big] \\
+ \int \mathrm{d}^3 r \langle \rho_{AA-Tyr}(\overrightarrow{r}) \rangle \Big[& f_{Tyr^-}(\overrightarrow{r})(\ln f_{Tyr^-}(\overrightarrow{r}) + \beta\mu^{\ominus}_{Tyr^-}) + f_{TyrH}(\overrightarrow{r})(\ln f_{TyrH}(\overrightarrow{r}) + \beta\mu^{\ominus}_{TyrH}) \Big] \\
+ \int \mathrm{d}^3 r \langle \rho_{AA-Arg}(\overrightarrow{r}) \rangle \Big[& f_{Arg^-}(\overrightarrow{r})(\ln f_{Arg^-}(\overrightarrow{r}) + \beta\mu^{\ominus}_{Arg^-}) + f_{ArgH}(\overrightarrow{r})(\ln f_{ArgH}(\overrightarrow{r}) + \beta\mu^{\ominus}_{ArgH}) \Big] \\
+ \int \mathrm{d}^3 r \langle \rho_{AA-His}(\overrightarrow{r}) \rangle \Big[& f_{His^-}(\overrightarrow{r})(\ln f_{His^-}(\overrightarrow{r}) + \beta\mu^{\ominus}_{His^-}) + f_{HisH}(\overrightarrow{r})(\ln f_{HisH}(\overrightarrow{r}) + \beta\mu^{\ominus}_{HisH}) \Big] \\
+ \int \mathrm{d}^3 r \langle \rho_{AA-Lys}(\overrightarrow{r}) \rangle \Big[& f_{Lys^-}(\overrightarrow{r})(\ln f_{Lys^-}(\overrightarrow{r}) + \beta\mu^{\ominus}_{Lys^-}) + f_{LysH}(\overrightarrow{r})(\ln f_{LysH}(\overrightarrow{r}) + \beta\mu^{\ominus}_{LysH}) \Big] \\
+ \int \mathrm{d}^3 r \langle \rho_{AA-Cys}(\overrightarrow{r}) \rangle \Big[& f_{Cys^-}(\overrightarrow{r})(\ln f_{Cys^-}(\overrightarrow{r}) + \beta\mu^{\ominus}_{Cys^-}) + f_{CysH}(\overrightarrow{r})(\ln f_{CysH}(\overrightarrow{r}) + \beta\mu^{\ominus}_{CysH}) \Big] \\
& + \sum_{k \in \{H^+, OH^-, Na^+, K^+, Cl^-\}} \beta\mu^{\ominus}_k \int \mathrm{d}^3 r \rho_k(\overrightarrow{r}).
\end{aligned}$$

Here, $f_{Asp^-}(\overrightarrow{r})$, $f_{Glu^-}(\overrightarrow{r})$, $f_{Tyr^-}(\overrightarrow{r})$, $f_{Arg^-}(\overrightarrow{r})$, $f_{His^-}(\overrightarrow{r})$, $f_{Lys^-}(\overrightarrow{r})$, and $f_{Cys^-}(\overrightarrow{r})$ are the fraction of amino acid residues that are charged or deprotonated at position $\overrightarrow{r}$ and $f_{AspH}(\overrightarrow{r})$, $f_{GluH}(\overrightarrow{r})$, $f_{TyrH}(\overrightarrow{r})$, $f_{ArgH}(\overrightarrow{r})$, $f_{HisH}(\overrightarrow{r})$, $f_{LysH}(\overrightarrow{r})$, and $f_{CysH}(\overrightarrow{r})$ are the fraction of neutral, protonated amino acids.

Free energy minimization results in a system of nonlinear coupled integro-differential equations, whose unknowns are (1) the osmotic pressure, or Lagrange multipliers that enforce the incompressibility constraint E_{rep}, $\pi(\overrightarrow{r})$, and (2) the electrostatic potential, $\psi(\overrightarrow{r})$. Through discretization of the system, the integro-differential equations are transformed into a set of non-linear algebraic equations whose solution can be iteratively obtained using standard numerical methods [115]. Thus, minimization of the total free energy results in a set of equations that determine the probability of every input conformation, α. These equations self-consistently determine the probability distribution function $P_g(\alpha)$, the density profiles of all molecular species, the charged states of the phosphates, the charged states of the amino acids, and other relevant quantities (see Chapter 4.6.1).

4.2.2. Polymer Models

The MT requires, as input, a set of chain conformations that are representative of our polyelectrolyte system.

Here, we opted for a relatively simplified model of DNA. Our model incorporated topological constraints via end-tethered loops to account for biophysical mechanisms which increase DNA flexibility in the eukaryotic nucleus. The elementary units of this system are phosphates, the only chargeable molecule of the dsDNA system. To accomplish this, we represent 100 phosphate monomers as a loop with fixed end monomers (Fig. 4.1). To generate a large set of conformations we performing MD simulations using GROMACS. To obtain a representative sample of the conformational space of these end-tethered loops, ranging from completely extended to completely collapsed, it was

necessary to perform biased sampling via introduction of an external bias potential (Fig. 4.15).

For the single nucleosome conformation of the chromatin-like system, there is only one conformation input into the MT. The 1KX5 crystal structure [61], composed of 147 bp of DNA wrapped around a canonical histone protein, was coarse-grained at the molecular level using the AICG protocol [165] for amino acid residues and the 3SPN model [88] for DNA nucleotides. For the 8-mer nucleosome array system, conformations were generated using the 1CPN model of chromatin [159], which is coarse-grained at the nucleosome level. Next, molecular detail, at the level of AICG and 3SPN, was reintroduced to the 1CPN conformations (Fig. 4.16).

4.3. Divalent Cations More Effectively Modulate Structure and Charge of Phosphate Loop Systems than Monovalent Cations Due to Ion Bridging Reactions

4.3.1. Effects of Monovalent Cations on Phosphate Loops Under Dilute Conditions

We begin our study by characterizing the influence of monovalent cations on phosphate loops for a more dilute system with very low bulk density, determined by a very low grafting density between phosphate loops, σ (Fig. 4.1B). Each monomer of the loop system represents one DNA-phosphate, a strong acid with a $pK_a = 1$ (Fig. 4.1). Unless otherwise noted, all MT calculations in this text are run for bulk $pH = 7.4$, which represents physiological pH. Thus, at physiological pH the phosphates are expected to be almost completely deprotonated and, hence, there will be a large amount of electrostatic repulsions within the phosphate loop system, even for dilute systems. There are several mechanisms available to reduce the unfavorable electrostatic repulsions in the system, which are all accounted for within the MT. Polymer stretching increases the distances between charged phosphates of neighboring loops and between neighboring monomers in the same loop. This vertical stretching thus decreases the charge density of the system, consequently reducing electrostatic repulsions. However, there is a physical limit to chain stretching and this mechanism might be unable to completely mitigate electrostatic repulsions between charged nearest neighbor phosphates within the same loop. Chain stretching also results in a loss of conformational entropy of the polyelectrolyte chains and, hence, is entropically unfavourable.

Additionally, counterions can reduce electrostatic repulsions through two mechanisms. First, positively charged counterions can localize in close proximity to the negatively charged phosphates within the loop layer (Fig. 4.1B&C). This counterion confinement effect increases electrostatic screening of phosphates, and thus reduces overall electrostatic repulsions. Secondly, ions can physically bind to phosphates. This ion pairing, or ion condensation, reaction neutralizes the negatively charged phosphates by chemically altering their charged state, and thus reducing the charge density of the system (Fig. 4.1B&C). However, like polymer stretching, both ion confinement and condensation come with an entropic cost. Namely, the loss of mixing entropy of ions and water molecules. Ion condensation also results in a loss of mixing entropy, as water is replaced with confined ions. Similar to ion condensation, acid-base equilibrium can also potentially decrease the net amount of charged phosphates. Shifting the acid-base equilibrium towards its protonated state decreases the number of phosphates in the deprotonated state. This mechanism is opposed by the chemical work required to perform this chemical reaction, which is determined by the free energy of the reaction and the local proton concentration. At physiological pH, acid-base equilibrium is expected to have a relatively small contribute to charge neutralization compared to ion condensation since phosphate $pK_a = 1$. It is nevertheless taken into account in the theoretical formulation. Overall, the system will need to balance the opposing enthalphic and entropic interactions in order to minimize the free energy. Consequently, a balance between the various opposing chemical and physical interactions will determine the resulting structure and amount of charge of the phosphate loop system.

One of the parameters that influences this equilibrium is the free energy difference between bound and unbound ions, or the standard Gibbs reaction free energy, $\Delta G_d^{\ominus}$. This quantity is determined by the dissociation equilibrium constant, pK_d, of the ion condensation chemical reaction. The dissociation constants or, conversely, the binding constants of the ion condensation reactions have not been properly established in the both the experimental and computational literature. Here, to determine physiologically relevant ion binding constants, we performed calculations with parameter scans over ion binding constants. We then indirectly compared our results with experiments performed under dilute conditions to determine physiologically relevant ion binding constants, which were subsequently employed to study denser systems.

We begin by determining the effects of ion binding by only considering Na^+ binding and acid-base equilibrium in a dilute system ($\sigma = 0.0002\, nm^{-2}$) with negligible interactions between neighboring loops. Over a range of bulk $[Na^+] = 10 - 500\, mM$ concentrations, we performed a parameter scan of $\Delta G_d^{\ominus}(PNa)$ from 1 to 6 k_BT. We considered this range of binding free energies to be reasonable given our previous studies involving ion condensation of Na^+ with acetate, a slightly weaker polyelectrolyte with $pK_a = 4$ [221, 200]. All MT calculations were performed in three-dimensions. The system lattice was discretized into three-dimensional cubic cells of size $\delta = 0.65\, nm$ and the segment length of monomers was $l_{seg} = 0.34\, nm$, similar to the size of DNA-phosphates. 3D calculations were performed to assess the effects of lateral heterogeneity.

We characterized the effects of a varying electrolyte environment on both the structure and charged state of the phosphate loop layer, which are inherently linked.

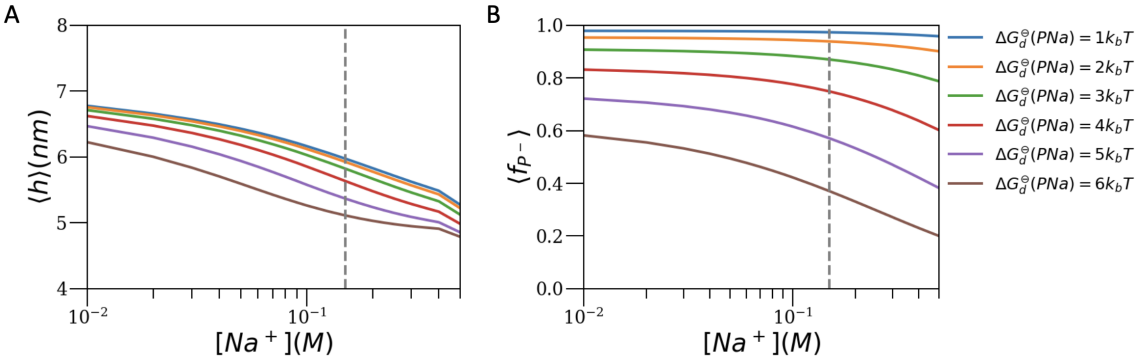

Figure 4.3. The influence of monovalent cations on structure and charge of dilute phosphate loop system depends on ion binding strength. The influence of increasing Na^+ concentration on (A) average height $\langle h \rangle$ in nm and (B) average fraction of charged phosphates, f_{P-}, is dependent on strength of ion binding $\Delta G_d^{\ominus}(PNa)$ for dilute systems of $\sigma = 0.0002\,nm^{-2}$, physiological bulk $pH = 7.4$ and no K^+ or Mg^{2+}. Physiological monovalent salt concentration of $150\,mM$ is denoted by the dotted grey line.

Specifically, we investigated these effects on the average height of the system, the extent to which the phosphate loop extends in the z-direction, and the average fraction of negatively charged phosphates. The height is defined as twice the first moment of the DNA-phosphate volume fraction:

$$\langle h \rangle = 2\langle z \rangle = 2\frac{\int d^3r z \langle \phi_{DNA-P}(\overrightarrow{r}) \rangle}{\int d^3r \langle \phi_{DNA-P}(\overrightarrow{r}) \rangle} \tag{4.17}$$

Note that $\langle \phi_{DNA-P}(\overrightarrow{r}) \rangle$ is dependent on the p.d.f. $P_g(\alpha)$ for each input conformation α. For the same grafting density, a smaller height signifies a more compact, less elongated polymer layer. The average fraction of phosphates that carry negative charges ($\langle f_{P^-} \rangle$), is

given by

$$\langle f_{P^-} \rangle = \frac{\int d^3r f_{P^-}(\overrightarrow{r}) \langle \rho_{DNA-P}(\overrightarrow{r}) \rangle}{\int d^3r \langle \rho_{DNA-P}(\overrightarrow{r}) \rangle}. \tag{4.18}$$

A lower $\langle f_{P^-} \rangle$ signifies increased charge neutralization due to ion condensation of counterions.

The height decreases slightly with both increasing bulk $[Na^+]$ and $\Delta G_d^{\ominus}(PNa)$ for our dilute system. With increasing Na^+ concentration the electrostatic repulsions between monomers in the same chain are reduced and the polymer conformations may adopt less extended states, as indicated by the slightly decreased average height of the polymer layer (Fig. 4.3A).

Additionally, increasing Na^+ concentration increases the amount of Na^+ binding. This reduces the negative charges in the system, as indicated by a decrease in the average fraction of charged phosphates. The ion-pairing also contributes to a smaller layer thickness, i.e., a shorter height. This charge neutralization effect is highly dependent on the value of $\Delta G_d^{\ominus}(PNa)$.

Larger Na^+ binding constants result in a greater decrease in charge for increasing bulk $[Na^+]$, while smaller $\Delta G_d^{\ominus}(PNa) = 1 - 2\,k_BT$ are only marginally sensitive to changes in bulk $[Na^+]$ (Fig. 4.3B). Notably, we do not observe an appreciable difference of sensitivity of height to changes in bulk $[Na^+]$ for varying $\Delta G_d^{\ominus}(PNa)$ (Fig. 4.3A), indicating the larger charge neutralization effects of greater Na^+ binding do not necessarily correlate with structural changes.

Next, we examined the effects of ion binding on the electrolyte environment of the phosphate loop system, by characterizing the counterion distribution surrounding the

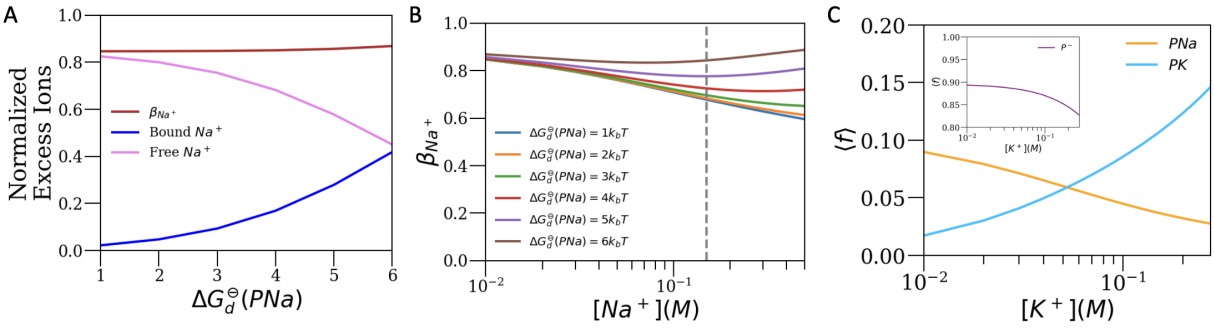

Figure 4.4. Ionic atmosphere depends on Na^+ binding strength and concentration. (A) For bulk $[Na^+] = 10\,mM$, fraction of excess ions contributed by free versus bound ions to total normalized excess ions, β_{Na^+}. Here, increasing ion binding $G_d^{\ominus}(PNa)$ results in increasing Na^+ ion condensation and decreasing free ions undergoing ion confinement. The total ion cloud encompassing the phosphate loop system remains the same. (B) β_{Na^+} is more sensitive to increasing Na^+ concentration for lower $G_d^{\ominus}(PNa)$. Physiological monovalent salt concentration of $150\,mM$ is denoted by the dotted grey line. (C) Fraction of phosphates that are bound to Na^+ f_{PNa} and bound to K^+ f_{PK} changes as K^+ concentration increases for $G_d^{\ominus}(PNa) = G_d^{\ominus}(PK) = 3\,k_BT$ and fixed bulk $[Na^+] = 50\,mM$. f_{PK} surpasses f_{PNa} around bulk $[K^+] = 50\,mM$, which is equal to bulk $[Na^+]$ in this system. *Inset* Fraction of phosphates that are negatively charged f_{P^-} decreases with increasing bulk $[K^+]$.

phosphate loop layer. Recently, ion counting of inductively coupled plasma mass spectroscopy (ICP-MS) experiments have probed the attraction of counterions to negatively charged DNA [241, 231, 306]. ICP-MS can determine the total number of excess ions observed in a system with DNA versus bulk solutions. These excess ions are composed of both condensed counterions and counterions contained in the cloud that surrounds a charged molecule or particle via ion confinement. For varying bulk ion concentrations, Gebala et al. have experimentally measured β^+, which is the normalized excess ion density, or the fraction of charge neutralization that arises from localization and condensation of positively charged counterions. Total β^+ is defined as the sum of the ion excess of all

positively charge counterions i^+,i.e., $\beta^+ = \sum_{i^+} \beta_{i^+}$ and

$$\beta_{i^+} = \frac{q_{i^+}\Gamma_{i^+}}{N_{DNA-P}|q_{DNA-P}|} \quad \text{with} \quad \Gamma_{i^+} = \int d^3r\big(\rho_{i^+}(\overrightarrow{r}) - \rho_{i^+}^{bulk}\big) \tag{4.19}$$

where Γ_{i^+} is the number of excess counterions of type i in the DNA solution compared to the bulk solution, N_{P-DNA} is the total number of DNA-phosphate molecules, and q_{P-DNA} is the number of charges per DNA-phosphate molecule.

The ion excess includes both the density of bound and unbound free ions. A value of β_{i^+} close to 0 indicates that you will find a similar concentration of counterions close to the DNA system as compared to the bulk solution. On the other hand, a β_{i^+} close to 1 indicates that counterions have a very high preference to be localized to, and even condensed with, DNA while negatively charged coions (i.e. Cl^-) are expelled from the DNA system. A high value of β^+ is indicative of strong electrostatic interactions and a more negative electrostatic potential. Note the sum of normalized excess ions over all cations i^+ and anions j^- should equal to 1 to maintain overall charge neutrality of the system, i.e. $\sum \beta^+ + \beta^- = 1$ [94, 95].

Importantly, our MT calculations can differentiate between excess counterions that are bound and condensed on DNA-phosphates and counterions that are free and experiencing ion confinement, while ion counting experiments can only measure the combined value of excess ions. For lower bulk $[Na^+] = 10\,mM$ as measured in Gebala et al. [95], we find that the total normalized excess ions, β_{Na^+}, is relatively insensitive to changes in Na^+ binding strength, $\Delta G_d^{\ominus}(PNa)$ (Fig. 4.4A). However, the normalized fraction of condensed ions substantially increased from close to 0 to 0.4 when varying $\Delta G_d^{\ominus}(PNa)$ between $1-6\,k_BT$, while normalized free excess ions decreased by a similar absolute range

(Fig. 4.4A). Additionally, the Gebala et al. report measured $\beta_{Na^+} = 0.85$ for a DNA system under similar environmental conditions as our 3D MT calculations, which varied from $\beta_{Na^+} = 0.845 - 0.868$ within the given range of binding energies $G_d^{\ominus}(PNa)$ considered (Fig. 4.3A). Given the matching of experimental results for DNA with 3D MT calculations of phosphate loops, we demonstrate our calculations for a simplified phosphate loop system are able to fairly accurately represent the ionic environment of DNA.

Calculations performed at this low salt concentration of bulk $[Na^+] = 10\,mM$ where β_{Na^+} is only slightly sensitive to changes to $G_d^{\ominus}(PNa)$ did not allow us to determine ion binding constants that would be physiological based on ion counting experiments. As a next step, we determined how ion binding strength influences sensitivity of total normalized excess ions to increasing concentrations of Na^+. Notably, for 3D calculations, we observe a variation in the behavior of excess ions, β_{Na^+} as a function of increasing bulk $[Na^+]$ that depends on $G_d^{\ominus}(PNa)$ (Fig. 4.4B). For lower $G_d^{\ominus}(PNa) = 1 - 3\,k_BT$, there is a monotonic decrease in β_{Na^+} with increasing bulk $[Na^+]$ (Fig. 4.4B). Higher $G_d^{\ominus}(PNa)$ exhibits nonmonotonic behavior, first decreasing up until $[Na^+] = 100 - 200\,mM$ and then increasing (Fig. 4.4B). For our 3D calculations, we see a similar sensitivity of β_{Na^+} to bulk $[Na^+]$ for $G_d^{\ominus}(PNa) = 3\,k_BT$ to $G_d^{\ominus}(PNa) = 6\,k_BT$ and this behavior for these lower ion binding constants is very similar to that determined by Gebala et al. via IC-PMS experiments under similar ranges of bulk $[Na^+]$ [94]. As $G_d^{\ominus}(PNa) = 3\,k_BT$ is very similar to the binding constant obtained for the ion pair of Na^+ and acetate [221, 200], and is within the physiological range determined by comparison to ion counting experiments, a value of $3\,k_BT$ for $G_d^{\ominus}(PNa)$ will be used for the rest of this study.

Finally, we wanted to determine physiological binding constants for the most prevalent intracellular monovalent cation, K^+. ICP-MS measurements demonstrated similar excess ion environments for Na^+ and K^+ under equal concentrations of both ions, despite differences in ionic size [93]. Na^+ has a radius $r_{Na^+} \approx 0.101\,nm$ and K^+ is slightly larger, with a radius of $r_{K^+} \approx 0.138\,nm$, which the MT explicitly takes into account. If Na^+ and K^+ have similar binding constants, i.e., $G_d^{\ominus}(PNa) = G_d^{\ominus}(PK) = 3\,k_BT$, then, according to ion counting observations, the bound fraction of Na^+ should be comparable to the bound fraction of K^+ when the concentrations of both ions are equal. For bulk $[Na^+] = 50\,mM$, varying bulk $[K^+] = 10 - 275\,mM$ we see a clear transition between predominance of bound fraction of Na^+ (f_{PNa}) versus bound K^+ (f_{PK}) at bulk $[K^+] = 50\,mM$, indicating our MT calculations also show no preference for Na^+ versus K^+ binding when ion binding energies for both ions are equal (Fig. 4.4C). Thus, MT calculations performed in the following sections ahve $G_d^{\ominus}(PNa) \sim G_d^{\ominus}(PK) = 3\,k_BT$, which we estimate to be physiologically relevant to our phosphate loop system.

Overall, there is a large dependence of structural (i.e., height) and charged state (i.e., average negatively charge fraction of phosphates) of the polymer system on the ion binding constant $G_d^{\ominus}(PNa)$, which also determines sensitivity to changes in Na^+ concentration, predominantly for the charged state. Here, we found that the most prevalent monovalent cations, Na^+ and K^+ have $G_d^{\ominus} \approx 3\,k_BT$, indicating relatively weak ion binding for the concentrations of monovalent cations studied under dilute conditions.

4.3.2. Charge Regulation Effects of Monovalent Cations in Dense Systems

Next, we investigated the effects of monovalent ion binding on the phosphate loop system under non-dilute conditions.

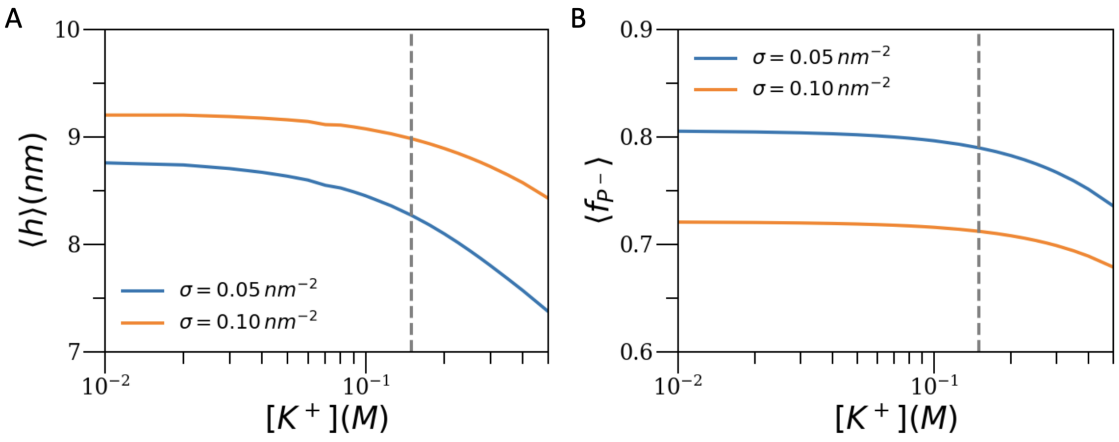

Figure 4.5. Effects of monovalent salt on average phosphate loop height and charge depends on bulk density. The effect of K^+ concentration on (A) the average height $\langle h \rangle$ in nm and (B) the average charged fraction of phosphates $\langle f_{P-} \rangle$ of the phosphate loop layer varies depending on grafting density σ, which determines the bulk density. Higher σ has a larger height and a lower charged fraction of phosphates that are both less sensitive to changes in monovalent cations. Physiological intracellular $[K^+] = 150\,mM$ is denoted by the grey dotted line.

The bulk $[K^+]$ concentration was varied from $10mM - 500mM$ to determine the effects of monovalent cations on the structure and charge of phosphate loops for a lower density ($\sigma = 0.05\,nm^{-2}$) and a higher density ($\sigma = 0.10\,nm^{-2}$) system.[2] Both average height and charged fraction of phosphates are relatively insensitive to changes in bulk $[K^+]$ up to $\approx 100\,mM$, which is close to combined physiological K^+ and Na^+ concentration of $150\,mM$ (Fig. 4.5A&B). After $[K^+] = 100\,mM$, both height and charge decrease

[2]Although the bulk densities are different both systems are non-dilute.

with increasing bulk $[K^+]$, however the exact behavior is dependent on bulk density (Fig. 4.5A&B). The higher density system ($\sigma = 0.10\,nm^{-2}$) has a larger height than the lower density system ($\sigma = 0.05\,nm^{-2}$) (Fig. 4.5A). There is increased polymer stretching for higher density systems to account for increased intra-chain interactions due to decreased distance between end-grafted loops. Additionally, the higher density system has a lower fraction of charged phosphates (Fig. 4.5B). The increase in grafting density results in a higher charge density of negatively charged phosphates that are electrostatically repelling each other. Consequently, an increased binding of K^+ to phosphates occurs to reduce the number of charged phosphates within the denser polymer layer. Notably, both the height, and especially the charged fraction, for the higher density system ($\sigma = 0.10\,nm^{-2}$) are less sensitive to changes in monovalent salt concentration compared to the lower density system ($\sigma = 0.05\,nm^{-2}$) (Fig. 4.5A&B). The fact that the average charged fraction does not change appreciably for bulk $[K^+]$ concentrations ranging over an order of magnitude indicates almost maximal K^+ binding even at the lowest $[K^+]$ (Fig. 4.5B).

These trends are consistent when varying ion binding strength of K^+, $G_d^{\ominus}(PK)$, demonstrating their general validity (Fig. 4.18). Contrary to the trends observed in our dilute system (Fig. 4.3), increases in K^+ ion binding strength result in a decreased sensitivity of height to increasing bulk $[K^+]$ (Fig. 4.2,4.18). The average fraction of charged phosphates, $\langle f_{P^-} \rangle$, is also less sensitive to increases in monovalent salt for higher ion binding strengths, especially for the higher $\sigma = 0.10\,nm^{-2}$ case (Fig. 4.18). These effects are caused by maximal K^+ binding at lower cation concentrations for these denser systems, which is augmented by increased ion binding strength. Altogether, this demonstrates that

increasing K^+ concentration beyond average physiological monovalent cation concentrations compacts phosphate loops and slightly decreases average system charge, and that this sensitivity is highly dependent on system density.

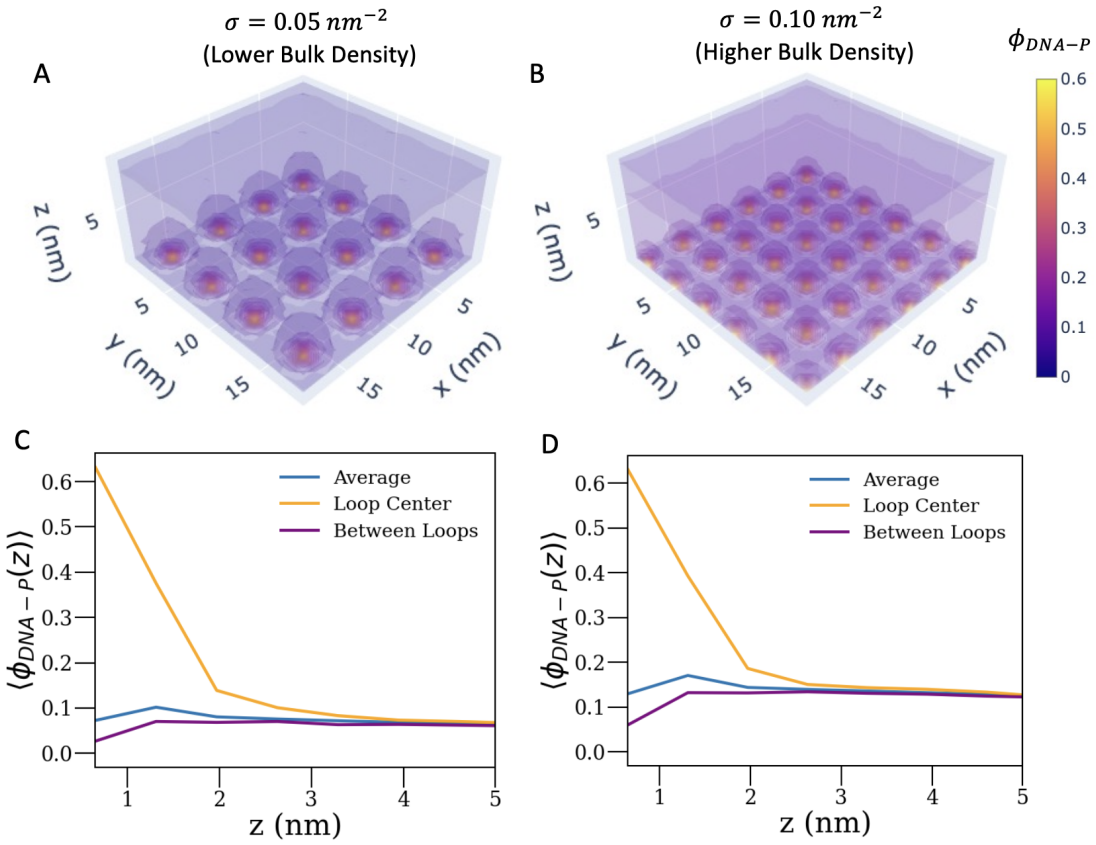

Figure 4.6. Bulk density influences phosphate density distribution for monovalent salt conditions. (A-B) 3D isosurface plot of DNA-phosphate volume fraction, $\langle\phi_{DNA-P}\rangle$ for a phosphate loop system with (A) $\sigma = 0.05\,nm^{-2}$ and (B) $\sigma = 0.10\,nm^{-2}$. Loop centers are indicated by yellow regions with higher volume fractions of DNA-phosphates. Resolution is $0.65\,nm$ x $0.65\,nm$ x $0.65\,nm$.(C&D) Average distribution of phosphate volume fraction in the z-direction up to $z = 5\,nm$ for (A) $\sigma = 0.05\,nm^{-2}$ and (B) $\sigma = 0.10\,nm^{-2}$. Lines represent the average over the entire system (Average; Blue Line), the average over the centers of all 16 graft points (Loop Center; Orange Line), and the average over "between loop" regions that are equidistant from graft points (Between Loops; Purple Line).

Next, we examined the three-dimensional distribution of relevant structural and chemical quantities under physiological monovalent salt $[K^+] = 150\,mM$ for different grafting densities. Fig. 4.6A&B) visualized the 3D distribution of DNA-phosphate volume fraction $\langle\phi_{DNA-P}\rangle$, which ranges from $0.1 - 0.6$ within the phosphate loop layer, with the highest volume fractions surrounding the graft points of phosphate loops. Visually, the figure demonstrates a distinct phosphate density distribution for the lower density ($\sigma = 0.05\,nm^{-2}$) compared to the higher density ($\sigma = 0.10\,nm^{-2}$) system, (Fig. 4.6A&B). This is further corroborated by quantitatively analyzing phosphate density distribution in different regions of the lateral plane and comparing this to the average density in the z-direction as show in Figures 4.6C&D. Loop center and between loop regions are compared with z-plane averages, as indicated by Fig. 4.18. Higher σ, indicating that phosphate loops are closer together, also results in a larger spread in phosphate density further away from the plane due to chain stretching, verifying our previous observations of increased height for higher density systems (Fig. 4.5A, Fig. 4.6). The higher density system also appears more compact in the horizontal plane, where each phosphate loop is relegated to a smaller area. Interestingly, the differences between $\langle\phi_{DNA-P}\rangle$ for the center of the loop compared to the average of the system and the area between phosphate loops is larger for the less dense system ($\sigma = 0.05\,nm^{-2}$) (Fig. 4.6C&D).

As polyelectrolyte structure is coupled with charged state, we next examined the influence of phosphate density on the three-dimensional distribution of electrostatic potential and local pH (Fig. 4.7, Fig. 4.8). MT explicitly considers H^+ ions and acid-base equilibrium and is therefore able to account for variable local (i.e., position-dependent) pH, or variable local proton concentration. As expected, an increase in phosphate loop

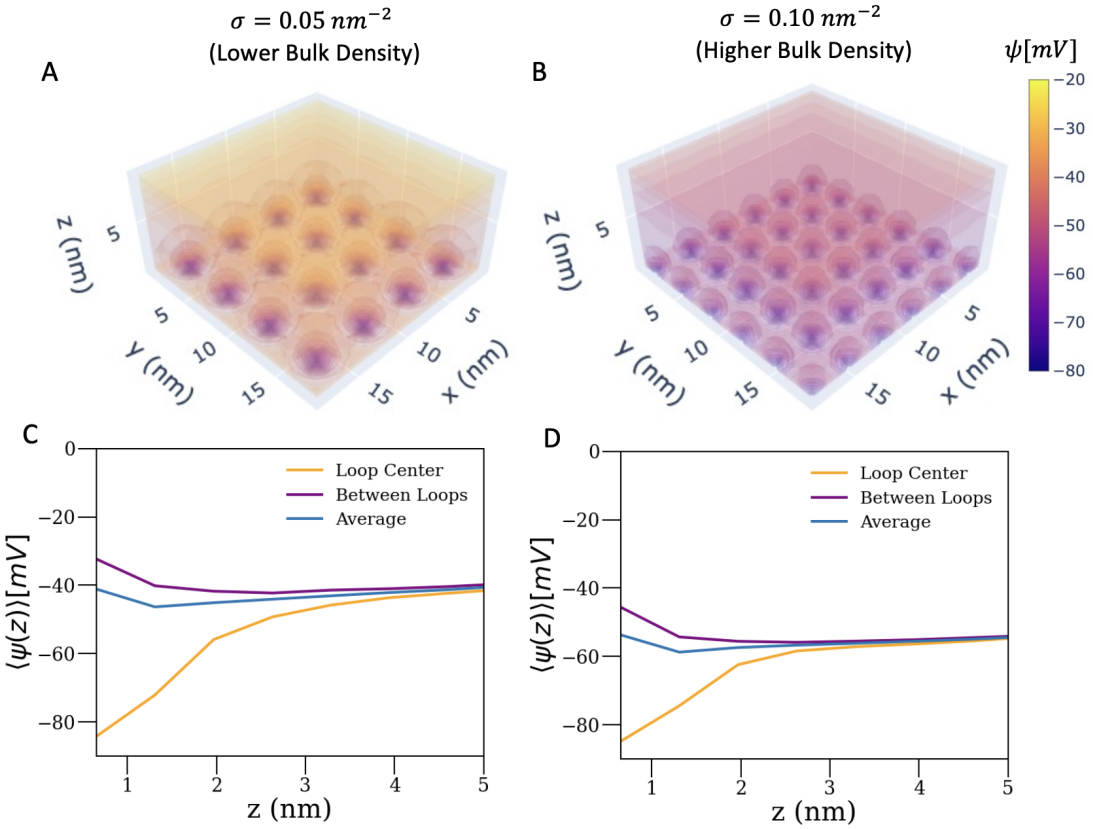

Figure 4.7. Bulk density influences electrostatic potential for monovalent salt conditions. 3D isosurface plot of electrostatic potential ψ in units of mV, for a 4x4 phosphate loop system with (A) $\sigma = 0.05\,nm^{-2}$ and (B) $\sigma = 0.10\,nm^{-2}$. Loop centers are dark purple regions with lower electrostatic potential due to the negative charges of DNA-phosphates. Resolution is $0.65\,nm$ x $0.65\,nm$ x $0.65\,nm$. (C&D) Average distribution of electrostatic potential in the z direction up to $z = 5\,nm$ for (A) $\sigma = 0.05\,nm^{-2}$ and (B) $\sigma = 0.10\,nm^{-2}$. Lines represent the average over the entire system (Average; Blue Line), the average over the centers of all 16 graft points (Loop Center; Orange Line), and the average over "between loop" regions that are equidistant from graft points (Between Loops; Purple Line). At large enough values of z, electrostatic potential will reach $\approx 0\,mV$.

grafting density (i.e., higher σ) at the surface correlates with an overall more negative electrostatic potential. Due to the increased charge density, the negative electrostatic

potential permeates further into the electrolyte solution above the phosphate loop layer for the denser system (Fig. 4.7A&B). Increasing grafting density also results in a decrease in local pH closer to the polymer layer, resulting from an increase in local $[H^+]$ to mitigate the highly negative electrostatic potential (Fig. 4.8A&B). Although there is also an osmotic component influencing local pH that increases with increasing phosphate density, the electrostatic potential has a much larger contribution to increasing this local proton concentration.

Next, we compared average values (Blue) of electrostatic potential and pH to those at the centers of loop graft points (Orange) and the areas between loops (Purple) to quantitatively assess lateral fluctuations in the xy-plane. At the z-plane closest to the graft point, the electrostatic potential decreases to $\approx -84\,mV$ within the loop centers. For lower bulk density $\sigma = 0.05\,nm^{-2}$ the average reaches $\approx -41\,mV$ and the between loop region reaches $\approx -46\,mV$, while for higher bulk density $\sigma = 0.10\,nm^{-2}$ the average reaches $\approx -54\,mV$ and the between loop region reaches $\approx -33\,mV$ (Fig. 4.7C&D). All three curves merge and go to $0\,mV$ for large enough z, the Loop Center, Between Loops, and Average curves remain different for larger z for the lower $\sigma = 0.05\,nm^{-2}$ system (Fig. 4.7C&D).

A similar trend can be seen when we examine pH (Fig. 4.8C&D). Additionally, the range of the observed decrease in local pH close to the graft points is also not predicted by averages in the lateral direction. Local pH decreases to $pH = 6.4$ for $\sigma = 0.05\,nm^{-2}$ in 3D, but averaging over xy-plane in 3D results in $pH = 6.6$ (Fig. 4.8A&C). Although these differences may appear small, they represent significant differences in local proton concentration, as $[H^+] = 10^{-pH}$.

Altogether, like the distribution of phosphate density, the electrostatic potential and local pH at loop centers and between loops diverges more for lower system densities $\sigma = 0.05\, nm^{-2}$ closer to the planar surface (i.e., $z = 0$), indicating a larger degree of lateral heterogeneity (Fig. 4.7C&D, Fig. 4.8C&D). The calculations demonstrate that systems with higher bulk density have lower average electrostatic potentials with a smaller degree of lateral fluctuations due to the fact that higher density systems have a more negative charge density (Fig. 4.7). Higher density systems have loops with highly charged phosphates that are closer together, which homogenizes the electrostatic potential compared to lower density systems that have areas between loops with less negative electrostatic potential.

Finally, we examined the effects of random compared to regular grafting patterns (Fig. 4.19). MT calculations demonstrate that, although there are large differences in DNA density distribution in the lateral plane (Fig. 4.19 A&B), average system properties, including dependence of height and average charged fraction of phosphates, are very similar between the different grafting patterns (Fig. 4.19 C&D). Thus, the trends observed from calculations performed with homogeneously grafted loops are expected to remain for heterogeneously grafted systems.

In conclusion, structural and electrostatic/chemical properties of the phosphate loop system are very much dependent on bulk density, and thus the effects of system density on charge regulation are integral to take into consideration. These properties are also highly heterogeneous in three-dimensional space, and the extent of lateral heterogeneity decreases for higher grafting densities compared to more dilute conditions.

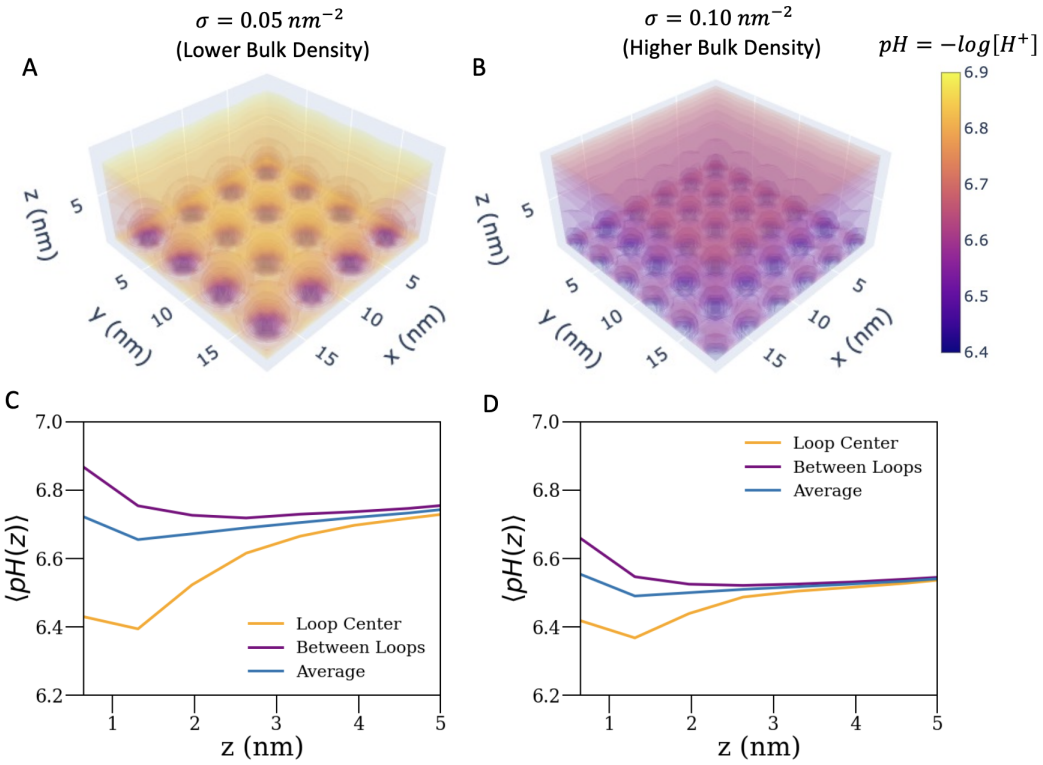

Figure 4.8. Bulk density influences local pH under monovalent salt conditions. 3D isosurface plot of local $pH(\overrightarrow{r}) = -log_{10}[H^+](\overrightarrow{r})$, for a 4x4 loop system with (A) $\sigma = 0.05\,nm^{-2}$ and (B) $\sigma = 0.10\,nm^{-2}$. Loop centers are dark purple regions with lower local pH resulting from a higher local density of H^+ attracted by the negative phosphate charges. Resolution is $0.65\,nm$ x $0.65\,nm$ x $0.65\,nm$. (C&D) Average distribution of electrostatic potential in the z direction up to $z = 5\,nm$ for (A) $\sigma = 0.05\,nm^{-2}$ and (B) $\sigma = 0.10\,nm^{-2}$. Lines represent the average over the entire system (Average; Blue Line), the average over the centers of all 16 graft points (Loop Center; Orange Line), and the average over "between loop" regions that are equidistant from graft points (Between Loops; Purple Line). At large enough values of z, pH will reach bulk levels of $pH = 7.4$.

4.3.3. Effects of Divalent Cations on Phosphate Loops Under Dilute Conditions

After characterizing the effects of monovalent cations on phosphate loops, we turned to study charge regulation of divalent cations. Due to their multivalent nature, they are expected to increase charge neutralization of strong polyelectrolyte systems. Mg^{2+} is the most prevalent intracellular multivalent cation, with total $[Mg^{2+}] \approx 10 - 20\,mM$ [244]. However, the majority of Mg^{2+} remains complexed with intracellular components, predominantly ATP and free $[Mg^{2+}] \approx 0.5\,mM$ [106].

We began our study by estimating the physiological 1:1 Mg^{2+} binding and its effects on the charge and structure of phosphate loops by performing a parameter scan over $\Delta G_d^{\ominus}(PMg^+)$. $\Delta G_d^{\ominus}(PK) = \Delta G_d^{\ominus}(PNa) = 3\,k_BT$ as determined in the previous section. A relatively low ion bridging energy of $\Delta G_d^{\ominus}(P_2Mg) = 7\,k_BT$ was chosen to better evaluate the effects of 1:1 binding alone. As our previous estimates of excess Na^+ ions from our 3D MT calculations matched well with ion counting experiments, we used results from ion counting experiments analyzing the competition between Na^+ and Mg^{2+} to estimate physiological 1:1 Mg^{2+} binding with phosphate [95]. Here, we performed MT calculations under bulk physiological $pH = 7.4$ with bulk $[NaCl] = 25\,mM$ and $[MgCl_2] = 2.5\,mM$ to compare to relevant ion counting experiments. Note that bulk $[Na^+] = 10[Mg^{2+}]$. The normalized excess ions for Na^+ decreases with increased 1:1 Mg^{2+} binding energy, $\Delta G_d^{\ominus}(PMg^+)$, indicating replacement of a predominantly Na^+ cloud with that of Mg^{2+} (Fig. 4.9A). The ionic environment surrounding phosphate loops is very sensitive to changes in 1:1 Mg^{2+} binding. A lower 1:1 binding energy, comparable to that of Na^+ and K^+ ($\Delta G_d^{\ominus}(PMg^+) = 3\,k_BT$), results in a two-fold increase in Na^+

excess ions compared to Mg^{2+}. Notably, even at these lower Mg^{2+} binding concentrations, there is still a 5x increase in bulk $[Mg^{2+}]$ compared to bulk $[Na^{+}]$ in the phosphate loop system compared to their relative concentrations in bulk solution. A higher Mg^{2+} binding ($\Delta G_d^{\ominus}(PMg^{+}) > 6\,k_BT$) results in an increased prevalence of excess Mg^{2+} compared to excess Na^{+} cations, despite Na^{+} concentration being an order of magnitude higher in the bulk solution (Fig. 4.9A). From a theory standpoint, Mg^{2+} is preferred over monovalent cations because the more highly charged Mg^{2+} cation screens electrostatics to a greater extent than monovalent cations for the same loss of translational entropy due to ion confinement and ion condensation. This effect is especially high for more dilute systems.

Ion counting measurements performed under similar environmental conditions demonstrate that the normalized excess cations (β_{i^+}), which include both free and bound ions, for Na^{+} and Mg^{2+} are equivalent, despite the concentration of Mg^{2+} being ten-fold lower than $[Na^{+}]$ in the bulk solution [95]. Normalized excess ions for Na^{+} are equal to that for Mg^{2+} around $\Delta G_d^{\ominus}(PMg^{+}) = 6\,k_BT$, making it a reasonable estimate for physiological 1:1 Mg^{2+} binding (Fig. 4.9A).

Increasing the ion bridging free energy $\Delta G_d^{\ominus}(P_2Mg)$ from $7\,k_BT$ to $11\,k_BT$ increases the fraction of phosphates in Mg^{2+} bridges, which slightly decreases the critical ion bridging coefficent $\Delta G_d^{\ominus}(PMg^{+})$, where the excess ions for Na^{+} crosses Mg^{2+}, to $\Delta G_d^{\ominus}(PMg^{+}) \approx 5\,k_BT$ (Fig. 4.20). However, under these conditions the fraction of phosphates in ion bridges f_{P_2Mg} becomes non-negligible (Fig. 4.20), and ion counting experiments with stiffer linear DNA under dilute conditions would not be expected to form significant ion bridges.

Notably, our estimates for 1:1 Mg^{2+} binding are quite similar to 1:1 binding of divalent Ca^{2+} to acetate despite having slightly different chemical properties [221, 200]. Based on extrapolation of ion dissociation measurements from acetate and formate, Sigel and Sigel [266] also estimate a 1:1 Mg^{2+} binding with phosphate of approximately $pK_d \approx 1$, which has subsequently been employed to study interactions between Mg^{2+} and nucleic acids employing atomistic MD simulations [271]. A standard Gibbs reaction free energy $\Delta G_d^{\ominus}(PMg^+) = 6\,k_BT$ is equivalent to a dissociation equilibrium constant $pK_d \approx 0.87$, which is similar to that estimated by Sigel and Sigel [266]. This again validates that our MT model of phosphate loops is able to recapitulate key features of DNA with a much simpler representation.

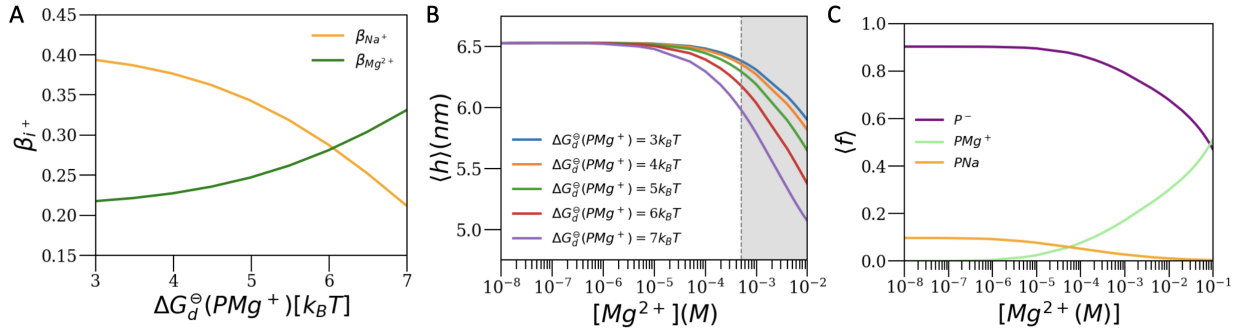

Figure 4.9. Effects of 1:1 Mg^{2+} binding on phosphate loop system under dilute conditions. All calculations are performed for bulk pH=7.4, $[Na^+] = 25\,mM$, $[Mg^{2+}] = 2.5\,mM$, $\sigma = 0.0002\,nm^{-2}$ (i.e., dilute conditions). (A) Normalized excess ions, β_{i^+}, for Na^+ and Mg^{2+} versus $\Delta G_d^{\ominus}(PMg^+)$. (B) Average height of loop layer (in nm) versus bulk $[Mg^{2+}]$ for varying $\Delta G_d^{\ominus}(PMg^+)$. Shaded grey region extends from $0.5\,mM$, the estimated intracellular free Mg^{2+} concentration, to $10\,mM$, the estimated total intracellular $[Mg^{2+}]$ [244, 106]. (C) Average fraction of phosphates with a negative charge (P^-), in a protonated state (PH), bound to 1 Mg^{2+} cation (PMg^+), or bound to 1 Na^+ cation (PNa) for $\Delta G_d^{\ominus}(PMg^+) = 6\,k_BT$ and varying bulk $[Mg^{2+}]$.

Next, in our dilute system, we investigated the effects of Mg^{2+} binding on phosphate loop structure and charge over a large range of bulk Mg^{2+} concentration $[Mg^{2+}] = 10^{-5} - 100\,mM$, to explore the effects of 1:1 Mg^{2+} binding at above and below estimated intracellular free $[Mg^{2+}] = 0.5\,mM$. All calculations were performed under physiological monovalent salt conditions to account for competition between monovalent K^+ and Na^+ and divalent Mg^{2+}. Although exact values are dependent on ion binding strength, the height of the phosphate loop system is sensitive to changes in bulk $[Mg^{2+}]$ after $\sim 0.1\,mM$ (Fig. 4.9B). Overall, Mg^{2+} reduces phosphate loop height by approximately $1.5\,nm$ between bulk $[Mg^{2+}] = 0.1 - 100\,mM$, while Na^+ reduces phosphate loop height by a similar amount over higher bulk $[Na^+] = 10 - 500\,mM$ (Fig. 4.3A, Fig. 4.9B). Height is also dependent on $\Delta G_d^{\ominus}(PMg^+)$, with larger binding strengths compacting phosphate loops at lower bulk $[Mg^{2+}]$ (Fig. 4.9B).

Finally, we investigated how varying Mg^{2+} concentration for bulk $[Na^+] = 25\,mM$ and physiological 1:1 Mg^{2+} binding influences the charged state of the phosphate loop system. Phosphates exhibit a preference for binding to Mg^{2+} compared to Na^+ even at low $[Mg^{2+}] = 0.1\,mM$, which is 250 times smaller than bulk $[Na^+] = 25\,mM$ (Fig. 4.9C). Additionally, the negatively charged fraction of phosphates decreases from ∼0.86 to ∼0.47 from bulk $[Mg^{2+}] = 0.1 - 100\,mM$ (Fig. 4.9C)). This is a much greater effect than for Na^+ under physiological $\Delta G_d^{\ominus}(PNa) = 3\,k_BT$, where the charged fraction of phosphates ranges from ∼0.91 to ∼0.79 for bulk $[Na^+] = 10 - 500\,mM$ (Fig. 4.3B, Fig. 4.9C). Additionally, the single Mg^{2+} bound fraction of phosphates (f_{PMg^+}) begins increasing from 0.0 around bulk $[Mg^{2+}] = 0.01\,mM$ and reaches ∼0.51 around bulk $[Mg^{2+}] = 100\,mM$ (Fig. 4.9C). When bulk $[Mg^{2+}]$ is equal to bulk $[Na^+] = 25\,mM$, bound Mg^{2+} fraction is $\sim$ 53x

greater than the fraction of phosphates bound with Na^+ (Fig. 4.9C). Altogether, this demonstrates that, even when only considering 1:1 Mg^{2+} binding to phosphate, Mg^{2+} is able to more efficiently compact and neutralize the charge of phosphate loops compared to monovalent cations due to its increased valency.

4.3.4. Effects of Ion Bridging on Phosphate Loops in Dense Systems

Ion bridging occurs when one multivalent cation binds and neutralizes two negatively charged phosphates. A previous MT study demonstrated that end-tethered poly(acrylic acid) brushes collapse in the presence of a critical Ca^{2+} concentration due to the formation of ion bridges between monomers [200]. Similar experimental studies by Tirrel and coworkers demonstrated experimentally and computationally that synthetic polyelectrolyte brushes can collapse or contract in electrolyte environments that contain di- and or multivalent ions. [37, 36, 124, 304] Likewise MD simulations [124] and Flory-like free energy approaches [78] also demonstrate the possibility of collapse of polyelectrolyte brushes in multivalent ion solutions. Thus, we hypothesize that Mg^{2+} will induce collapse of phosphate loops via ion bridging between phosphates, which is especially important to neutralize charges in denser systems.

Here, we determine the effects of Mg^{2+} bridging on the structure and charge of dense phosphate loop systems. First, we investigated the effects of varying bulk $[Mg^{2+}]$ and $\Delta G_d^{\ominus}(P_2Mg)$, or the ion bridging free energy, on the average height of the phosphate loop layer. Calculations were performed using the physiological binding constants for K^+, Na^+, and 1:1 Mg^{2+} binding obtained in the previous sections and for physiological bulk

$pH = 7.4$, higher bulk density ($\sigma = 0.10\,nm^{-2}$), and physiological bulk $[K^+] = 140mM$ and $[Na^+] = 10mM$.

The average height of the phosphate loop system is much more sensitive to lower bulk Mg^{2+} concentrations compared to monovalent counterions under similarly dense conditions (Fig. 4.10A, Fig. 4.5A). Additionally, the change in height is notably larger, varying by $3\,nm$ with Mg^{2+} bridging reactions (Fig. 4.10A), while varying over only approximately $1.5\,nm$ when only considering 1:1 Mg^{2+} binding alone in dilute systems (Fig. 4.9B) and $< 0.5\,nm$ for varying $[K^+]$ for similar bulk densities (Fig. 4.5A). Complete collapse of phosphate loops is observed after bulk $[Mg^{2+}] = 10\,mM$, the total estimated concentration of intracellular Mg^{2+}, for all ion bridging free energies (Fig. 4.10A). Such a collapse can be inferred by the insensitivity of the height to changes in bulk $[Mg^{2+}]$ above a certain threshold, indicating a saturation that is not seen when considering 1:1 Mg^{2+} or monovalent cation binding alone. The collapse of the phosphate loop layer is observed at lower bulk $[Mg^{2+}]$ concentrations for higher ion bridging free energies, $\Delta G_d^{\ominus}(P_2Mg)$ (Fig. 4.10A). Systems with lower average density $\sigma = 0.05\,nm^{-2}$ are more sensitive to bulk $[Mg^{2+}]$, although the range over which they are most sensitive is similar to higher density systems (Fig. 4.10, Fig. 4.22).

To determine a physiologically relevant free energy of Mg^{2+} bridging, we combine several different experimental observations for dsDNA and chromatin, which we expect to be largely influenced by the chemical behavior of phosphates. Under more dilute conditions, dsDNA self-assembly and attraction between dsDNA strands as well as intermolecular association of chromatin arrays have been shown to occur at concentrations of Mg^{2+} in the several millimolar range [172, 121, 230, 259, 260, 209]. Denser systems

are expected to increase ion bridging to account for higher charge densities. Additionally, Engelhardt and colleagues determined that heterochromatin remains compacted in isolated nuclei, e.g., a denser system, for $[Mg^{2+}] \geq 2\,mM$ [76]. Altogether, this points to the prevalence of ion bridging reactions, resulting in compaction of DNA and phosphate loops, around several mM of bulk $[Mg^{2+}]$. Thus, $\Delta G_d^{\circ}(P_2Mg) \approx 14\,k_BT$ could be estimated as physiological a physiological ion bridging coefficient as the bulk $[Mg^{2+}]$ versus height curve for this value of $\Delta G_d^{\circ}(P_2Mg)$ best represents these experimental findings (Fig. 4.10A). Similar to the effects of system density on the influence of K^+, less dense systems ($\sigma = 0.05\,nm^{-2}$) are even more sensitive to changes in bulk $[Mg^{2+}]$ for physiological ion bridging energies (Fig. 4.22A).

To better understand the effects of ion bridging on the ion cloud surrounding phosphate loops, we determined the contribution to total excess Mg^{2+} ions of the free, 1:1, and 2:1 bound Mg^{2+} fractions for the critical $[Mg^{2+}] = 2\,mM$. As the ion bridging free energy becomes stronger, both the free Mg^{2+} and the 1:1 bound Mg^{2+} fractions decrease, the 2:1 bound Mg^{2+} fraction increases and the total normalized excess Mg^{2+}, $\beta_{Mg^{2+}}$ remains fairly insensitive to changes in bulk $[Mg^{2+}]$ (Fig. 4.10B). Notably, the "tipping point" for Mg^{2+} to be predominantly found in an ion bridge with phosphates compared to in a free or 1:1 bound state occurs right before our estimate of physiological $\Delta G_d^{\circ}(P_2Mg) = 14\,k_BT$, for critical $[Mg^{2+}] = 2\,mM$ (Fig. 4.10B). Total $\beta_{Mg^{2+}}$ is higher for denser systems (Fig. 4.10B) compared to more dilute systems (Fig. 4.9A) indicating a larger Mg^{2+} cloud to compensate for an increased charge density.

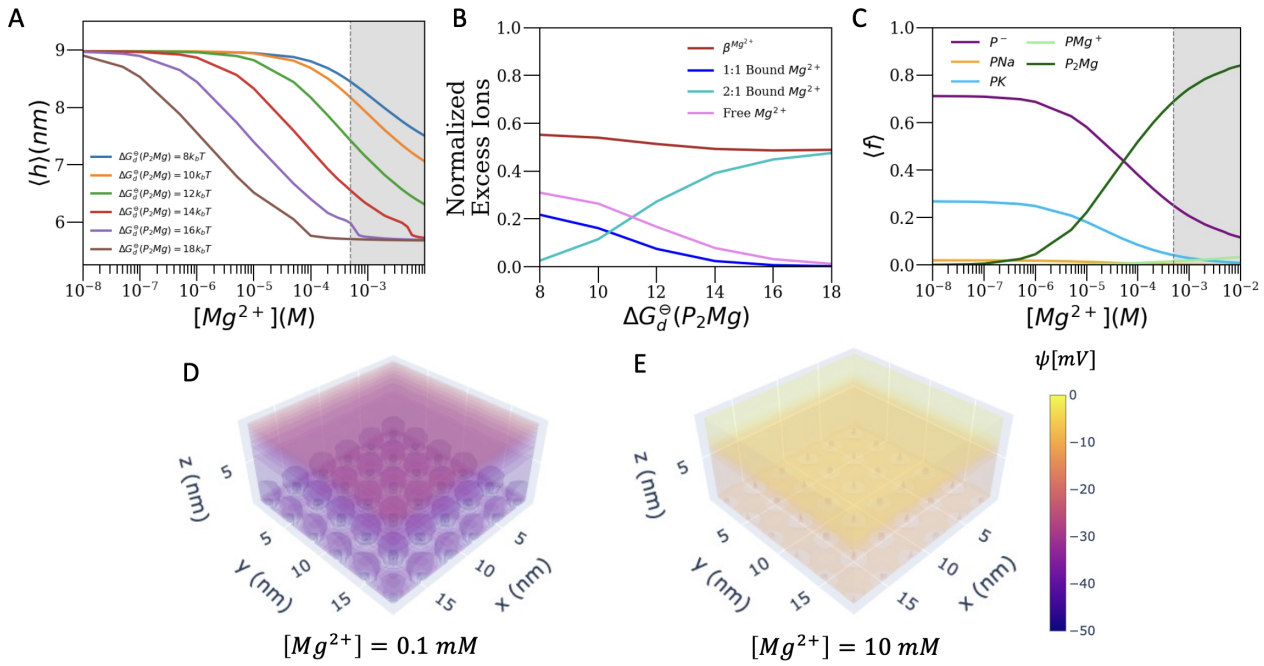

Figure 4.10. Ion bridging has a large influence on structure and charge of dense phosphate loop systems. All calculations were performed for higher bulk density ($\sigma = 0.10\,nm^{-2}$), bulk $pH = 7.4$, physiological monovalent cations (bulk $[K^+] = 140\,mM$, bulk $[Na^+] = 10\,mM$), $\Delta G_d^{\ominus}(PNa) = G_d^{\ominus}(PK) = 3\,k_BT$, and $\Delta G_d^{\ominus}(PMg^+) = 6\,k_BT$. (A) Average height (in nm) of phosphate loop layer for varying bulk $[Mg^{2+}]$. Each colored line represents different ion bridging free energies, $\Delta G_d^{\ominus}(P_2Mg)$. Higher ion bridging results in a stronger loop collapse for lower bulk $[Mg^{2+}]$. Shaded region represents physiologically relevant intracellular concentrations of free Mg^{2+}. (B) Normalized excess Mg^{2+} cations split into total, $\beta_{Mg^{2+}}$, 1:1 Mg^{2+}-phosphate pair, 2:1 Mg^{2+}-phosphate pair, and free Mg^{2+} for varying $\Delta G_d^{\ominus}(P_2Mg)$. (C) Fraction of phosphates with a negative charge (P^-), bound to Na^+ (PNa), bound to K^+ (PK), bound in a 1:1 Mg^{2+}-phosphate pair, or Mg^{2+} forming an ion bridge between two phosphates (P_2Mg) for $\Delta G_d^{\ominus}(P_2Mg) = 14\,k_BT$. (D-E) Electrostatic potential, ψ, in units of mV for (D) lower, $0.1\,mM$ and (E) upper, $10\,mM$, bounds of free Mg^{2+} with physiological ion bridging, $\Delta G_d^{\ominus}(P_2Mg) = 14\,k_BT$.

Next, we examined how the chemical state of DNA-phosphates is modulated by changes in bulk $[Mg^{2+}]$ for physiological ion bridging free energy. There is a dramatic decrease of average negatively charged phosphates, $\langle f_{P^-} \rangle$, after approximately bulk

$[Mg^{2+}] = 0.01\,mM$ (Fig. 4.10C), which occurs an entirely one order of magnitude lower than when only considering 1:1 Mg^{2+} binding in dilute systems where $\langle f_{P^-} \rangle$ attenuates around bulk $[Mg^{2+}] = 1\,mM$ (Fig. 4.9C). This corresponds to a sharp increase of Mg^{2+}-phosphate ion bridges around similar bulk Mg^{2+} concentrations (Fig. 4.10C). The largest changes in the chemical state of phosphates occur between bulk $[Mg^{2+}] = 0.01 - 1\,mM$, which correspond to the largest changes in the average height of the phosphate loop system (Fig. 4.10A&C). Increasing bulk $[Mg^{2+}]$ also substantially decreases the fraction of K^+ and Na^+ ions bound to phosphates, while 1:1 Mg^{2+}-phosphate binding increases to a much lesser extent, indicating a large preference for ion bridging compared to 1:1 binding (Fig. 4.10C). Such an increased fraction of phosphates in ion bridges compared to ion pairs with monovalent cations can also be explained by considering entropic versus electrostatic components of the free energy. Ion bridging is an energetically much more efficient way to neutralize charges than 1:1 Mg^{2+} and Na^+ and K^+ ion binding, as all ion condensation reactions reduce the translational entropy of ions. Thus, monovalent cation binding is replaced by ion bridging reactions when bulk Mg^{2+} concentrations are approximately 4 orders of magnitude smaller than total bulk monovalent cation concentrations (crossover occurs at $[Mg^{2+}] \approx 0.001\,mM$ while total monovalent concentration is $150\,mM$). For less dense systems ($\sigma = 0.05\,nm^{-2}$), although the absolute values of fractions of phosphates in certain chemical states are shifted (i.e., for lower $[Mg^{2+}]$ there is a higher fraction of negatively charged phosphates), similar to trends observed with the average height, the overall range where the charged state is most sensitive to changes in bulk $[Mg^{2+}]$ remains the same (Fig. 4.10C, Fig. 4.22B).

Finally, we examined the 3D electrostatic potential distribution for ranges of free $[Mg^{2+}]$ that may be relevant to observed intracellular fluctuations (Fig. 4.10D-E). Although difficult to determine exactly, it is estimated that total intracellular $[Mg^2+] \approx 10\,mM$, while free $[Mg^2+] \approx 0.5\,mM$, although free Mg^{2+} can vary significantly between cells and within the cell cycle [244, 106, 181]. For a lower bound of bulk $[Mg^{2+}] = 0.1\,mM$, the electrostatic potential centered around loop graft points is highly negative and even permeates above in the z direction beyond the phosphate loop layer (Fig. 4.10D). However, for an upper bound of bulk $[Mg^{2+}] = 10\,mM$, the phosphate loop system becomes nearly charge neutral, with decreased lateral fluctuations in electrostatic potential (Fig. 4.10E). Analyzing the 3D distribution of phosphate density and local pH, we see that increasing bulk $[Mg^{2+}]$ between these lower to upper bounds results in more compacted phosphate density and a higher pH closer to the bulk value of 7.4, with more diminished lateral fluctuations (Fig. 4.23).

We would like to point out that the mechanism for the observed collapse is the large free energy associated with ion bridge formation. The energy gained upon the chemical reaction of ion bridging is so substantial that the system wants to form as many ion bridges as possible, resulting in increased phosphate compaction. This free energy gain for sufficiently strong ion bridging is enough to overcome the loss of conformational entropy associated with polymer brush collapse. Concomitantly, the system decreases its charge and the electrostatic repulsions decrease, as demonstrated by Fig. 4.10C-E, further facilitating the collapse of the phosphate loop layer. Finally, MT calculations also demonstrate that system properties are much more sensitive to changes in $[Mg^{2+}]$ than changes in pH, indicating that changes in bulk pH do not have a significant charge

regulation effects for our phosphate loop system compared to the electrolyte environment and average phosphate density (Fig. 4.24), although there are significant fluctuations in local pH (Fig. 4.23C&D). Altogether, our results show that phosphate loop structure and charge are both highly sensitive to the divalent Mg^{2+} cation due to its ability to form ion bridges between phosphates.

4.4. The Bulk Monovalent Electrolyte Environment, pH, and System Density Influence Charge and Structure of Chromatin

4.4.1. Determining the Influence of Monovalent Cations on the Charged State of Single Nucleosomes

After estimating physiologically relevant ion condensation free energies for DNA-phosphates and investigating the differences between monovalent and divalent cation environments in our simpler phosphate loop system, we extended our MT approach to investigate chromatin. We began our chromatin study by characterizing the effects of the monovalent electrolyte environment on the charged state and ion cloud of single nucleosomes (Fig. 4.11A). Comparing the total charge of the yeast nucleosome (1ID3) [295] to that of the human nucleosome (1KX5) [61], we determined that the charge of both nucleosomes is sensitive to changes in $[Na^+]$, especially around physiological monovalent salt concentration of $150\,mM$ (Fig. 4.11B). However, the yeast nucleosome carries a much larger negative charge than the human nucleosome. Although there is $> 60\%$ homology between the human and yeast nucleosome core particles [295], changes in key amino acid residues, especially on the surface of the histone subunits, seem to have a non-neglibile influence on total nucleosome charge (Fig. 4.11B).

Next, we determined how the ionic environment influences the ion cloud of the human and yeast nucleosomes. Notably, we calculated a total excess Na^+ of $\beta_{Na^+} = 0.78$ for $[Na^+] = 10\,mM$ and physiological pH (Fig. 4.11C). This value is very similar to the $\beta_{Na^+} = 0.85 \pm 0.019$ determined from ion counting experiments of nucleosomes under the same environmental conditions [95], although MT calculations were performed on tailless

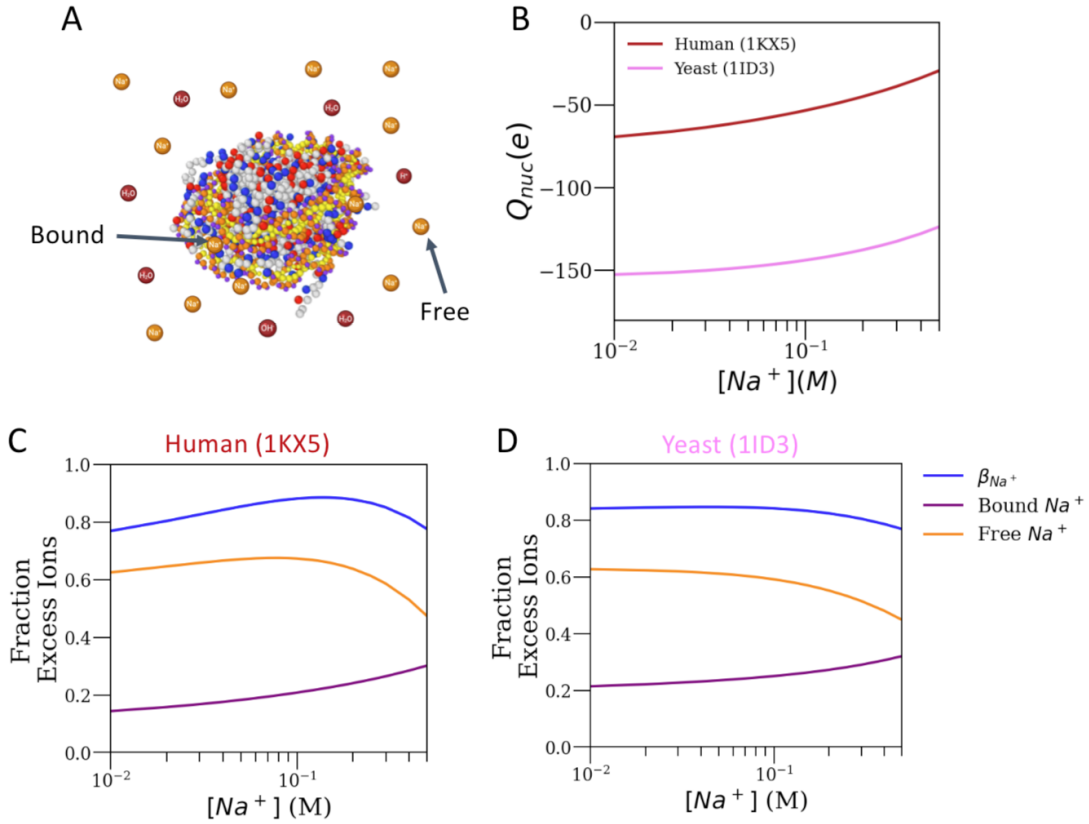

Figure 4.11. Monovalent electrolyte environment influences charge and excess ions in distinct ways for yeast versus human nucleosomes. All calculations were performed for bulk $pH = 7.4$ and bulk $[K^+] = [Mg^{2+}] = 0\,mM$. (A) Rendering of single nucleosome in solution. Na^+ counterions shield the negative electrostatic potential of the nucleosome by either binding and changing the charged state of the DNA-phosphates or being localized in close proximity to the nucleosome in the free state via ion confinement. (B) Comparing total nucleosome charge versus $[Na^+]$ for human (1KX5) and yeast (1ID3) nucleosomes. In general, yeast nucleosomes carry a more negative charge than human nucleosomes. (C&D) Total excess ions (β_{Na^+}), fraction of bound excess Na^+, and fraction of free Na^+ versus bulk $[Na^+]$ for (C) human and (D) yeast nucleosomes.

nucleosomes whereas experiments were performed on canonical nucleosomes with histone tails.

We then compared the sensitivity of total, bound, and free excess Na^+ ions to $[Na^+]$ for human and yeast nucleosomes (Fig. 4.11C&D) as well as DNA-phosphate loops (Fig. 4.25) and observed several key differences. Both the total and free excess Na^+ cations monotonically decrease with increasing $[Na^+]$ for the phosphate loop system (Fig. 4.25). However, the free and total excess Na^+ ions for the human nucleosome first increase with increasing $[Na^+]$, up until physiological monovalent cation concentration of $\sim 150\,mM$, and then decrease (Fig. 4.11C). The yeast nucleosome remains almost insensitive to changes in $[Na^+]$ up until $[Na^+] \approx 100\,mM$, and then begins to decrease as well (Fig. 4.11D). There is also increased Na^+ binding for both nucleosome systems compared to the phosphate loop system, and there is even more bound Na^+ for the yeast compared to the human nucleosome (Fig. 4.11C&D, Fig. 4.25). This observation can be explained by the fact that wrapping the DNA around histone proteins increases the charge density of DNA-phosphates, which requires extra charge neutralization via ion condensation of Na^+ counterions. The overall different behavior between phosphate loops and nucleosomes indicates the increased complexity of the chromatin system, which is composed of highly negatively charged phosphates as well as neutral, acidic, and basic amino acid residues, all of which influence how chromatin responds to changes in the electrolyte environment.

4.4.2. Bulk pH and Ion Condensation Modulate Charge of Single Nucleosome and Nucleosome Chain Chromatin Systems

Next, we investigated the effects of bulk pH on the charged state of our two chromatin systems, the single nucleosome (Fig. 4.2A) and the 8-mer nucleosome chain (Fig. 4.2B).

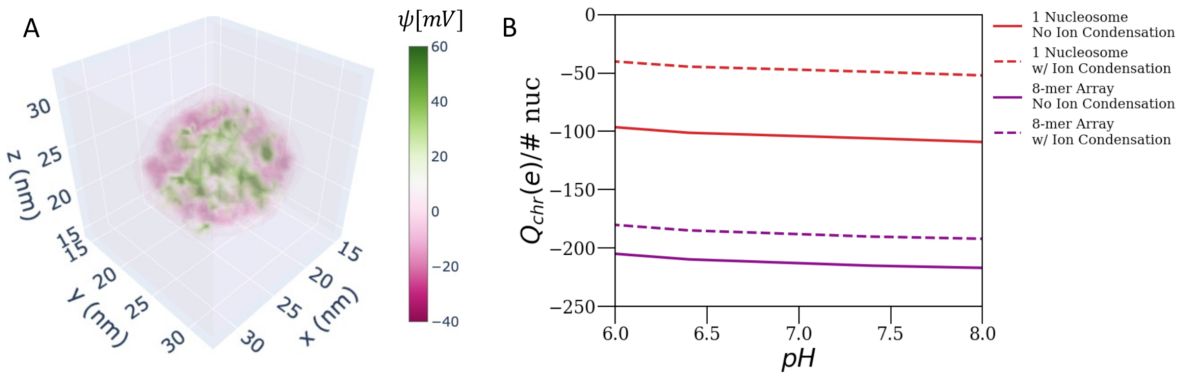

Figure 4.12. Influence of bulk pH and ion condensation on the charge of chromatin systems. All calculations were performed for $[K^+] = 140\,mM$, $[Na^+] = 10\,mM$, and $[Mg^{2+}] = 0\,mM$. (A) The three-dimensional distribution of electrostatic potential of the single nucleosome system for bulk $pH = 7.4$. Regions with more negative electrostatic potential, occupied by negatively charged phosphates and acidic amino acids, are in pink and regions with more positive electrostatic potential, occupied by basic amino acids, are in green. (B) Charge per nucleosome ($Q_{chr}(e)/\#\,nuc$) versus bulk pH for single nucleosome and 8-mer nucleosome chain systems with and without ion condensation to DNA-phosphates. Range of bulk pH encompasses approximately ±1.0 from physiological pH.

At the single nucleosome level, note that the electrostatic potential is heterogeneously distributed (Fig. 4.12A). The charged state of the system is dependent on the bulk pH due to the acid-base equilibrium of the histone protein amino acid residues (Table 4.2). However, like most biological systems, total nucleosome charge is only slightly sensitive to bulk pH around physiological pH (Fig. 4.12B), although the chromatin system is still more sensitive to changes in bulk pH for this range compared to the phosphate loop system (Fig. 4.24). Notably, the average charge per nucleosome of the 8-mer nucleosome array system is much more negative than for the single nucleosome within this pH range (Fig. 4.12B). This points to the large effects of linker DNA on the total charge of the chromatin

system, as linker DNA connects nucleosomes in the 8-mer array but is not part of the single nucleosome. Note that, for the same total charge, the charge density of wrapped DNA is higher compared to linker DNA. Additionally, ion condensation by monovalent K^+ and Na^+ cations substantially neutralizes the charge of both systems, underlying the importance of ion binding in the charge regulation of chromatin (Fig. 4.12B).

4.4.3. Monovalent Electrolyte Environment and Bulk Density Modulate Charge and Structure of Chromatin Chains

Finally, we investigated the effects of monovalent cations on our 8 nucleosome chromatin chain system for lower and higher bulk density systems (Fig. 4.13A). The average charge per nucleosome is sensitive to $[K^+]$, especially around physiological monovalent cation concentrations of $\sim 150\,mM$ (Fig. 4.13B). Higher bulk density systems have a slightly less negative average charge for lower $[K^+]$, but this difference becomes negligible around physiological monovalent concentrations (Fig. 4.13B). Note that, although the total charge for the lower and higher bulk density systems seems similar, due to osmotic pressure from surrounding chromatin chains, higher density systems are more compacted (Fig. 4.14B) and thus have a higher charge density.

The sensitivity of the counterion cloud to changes in bulk $[K^+]$ is more distinct for the two systems (Fig. 4.13C). Below physiological $[K^+] = 150\,mM$, there is a larger observable difference between lower bulk density and higher bulk density systems for the total excess K^+ counterions, (Fig. 4.13C). Remember that normalized excess ions, β_{K^+} increases when there is a larger difference between cation concentration in the bulk solution and that of the chromatin solution, and is indicative of a more negative system

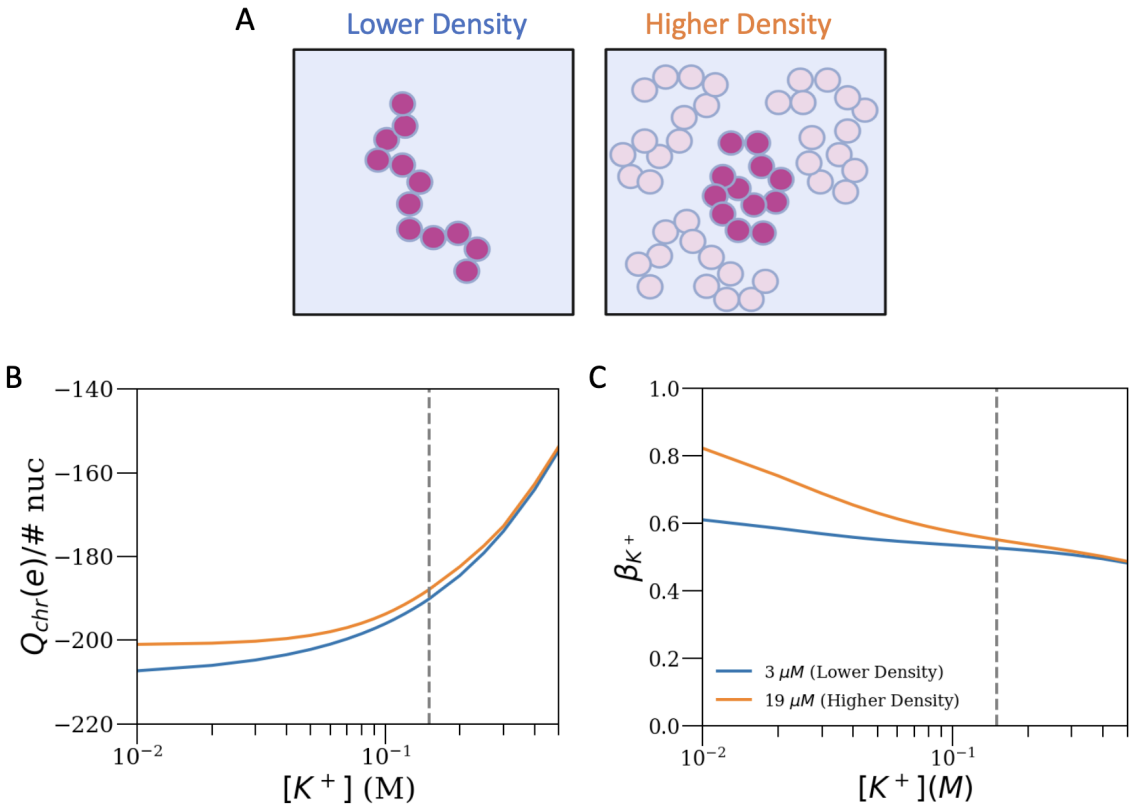

Figure 4.13. Influence of monovalent electrolyte conditions and bulk density on charged state and of 8-mer nucleosome arrays. All calculations were performed for physiological bulk $pH = 7.4$ and bulk $[Na^+] = [Mg^{2+}] = 0\,mM$. (A) Calculations were performed for lower density (3 μM) and higher density (19 μM) systems where concentration was determined by the number of chains within a given box size. (B&C) Influence of K^+ on charged state and ion cloud of 8-mer system. (B) Average charge per nucleosome and (C) total excess K^+, β_{K^+}, versus bulk $[K^+]$ for lower density (Blue) and higher density (Orange) systems. Physiological monovalent cation concentration $\sim 150\,mM$ is denoted by grey dotted line.

charge. At lower $[K^+] < 150\,mM$, the higher bulk density system has a more negative total charge per nucleosome Fig. 4.13B), and thus a highly negative charge density. Thus, a relatively larger concentration of K^+ compared to the lower bulk $[K^+]$ is localized to the counterion cloud surrounding the chromatin chain to reduce electrostatic repulsions in

the higher density system (Fig. 4.13C). As bulk K^+ concentration increases, the relative difference in excess total K^+ between the chromatin system and bulk solution decreases for the high density system (Fig. 4.13C), as the system becomes relatively more charge neutral (Fig. 4.13B). For both lower and higher bulk density systems, total excess K^+ becomes relatively insensitive to changes in $[K^+] \geq 150\,mM$.

Next, we studied the influence of monovalent cations and system density on structural parameters of the chromatin chain, such as radius of gyration, $\sqrt{\langle R_g^2 \rangle}$, the average spacing between nucleosomes, and the bond angle between neighboring nucleosomes (Fig. 4.14). Radius of gyration indicates the location of the highest concentration of mass of the chromatin chain.

A larger radius of gyration represents a more elongated chain and a smaller radius of gyration represents a more compacted chain. There are much larger differences between the radius of gyration of lower versus higher bulk density systems compared to changes induced by $[K^+]$ (Fig. 4.14B). The higher density system is more compacted, indicated by a lower radius of gyration, and is relatively insensitive to changes in $[K^+]$ (Fig. 4.14B). This observed larger-scale compaction is due to osmotic pressure from neighboring chains, which increases with system density. As the charge density also increases with bulk system density, the charge neutralization effects of increasing $[K^+]$ within this range (Fig. 4.13B) are not sufficient to further compact the higher density system. On the other hand, the lower density system is more sensitive to changes in $[K^+] < 150\,mM$ (Fig. 4.14B)

Finally, we investigated the effects of bulk system density and electrolyte environment on more local descriptors of chromatin chain structure. Specifically, we determined the pairwise spacing between neighboring nucleosomes and the bond angles between three

connected nucleosomes and averaged these properties across the chromatin chain. As observed with the radius of gyration, there are large differences in average nucleosome spacing and bond angles for lower and higher bulk system densities, with lower density systems being more sensitive to $[K^+]$ (Fig. 4.14C&D). Notably, both of these "local" structural parameters are more responsive to changes in $[K^+] < 150\,mM$ especially compared to the radius of gyration, which is a more "global" descriptor of chromatin structure (Fig. 4.14C&D). Higher bulk density systems have smaller nucleosome spacing, which further decreases with increasing $[K^+]$ (Fig. 4.14C). Again, the higher bulk density system is relatively less responsive to changes in $[K^+]$ due to its highly negative charge density. Thus, decreasing spacing between nucleosomes via charge regulation of K^+ is not as noticeable compared to the lower density system (Fig. 4.14C). Additionally, higher bulk density systems have larger bond angles (Fig. 4.14D). Increasing bond angles reduces electrostatic repulsions between neighboring nucleosomes and between highly negatively charged linker DNA for these higher density systems (Fig. 4.14D). Interestingly, increasing bulk K^+ concentration seems to compactify chromatin chains (Fig. 4.14B&C) while increasing bond angles (Fig. 4.14D), potentially unveiling a conserved mechanism for increasing chromatin compaction.

Overall, bulk K^+ has a non-negligible effect on the charge regulation of chromatin chains, which is modulated further by bulk density. Structural elements of the chromatin chain, including the radius of gyration, nucleosome spacing, and bond angles, are more sensitive to $[K^+]$ for lower density systems, although this sensitivity tapers off around physiological monovalent salt concentration $\sim 150\,mM$. This indicates that the structural

effects of K^+ are "saturated" above this concentration, although the effects on the system charge are still highly sensitive to increases in $[K^+]$.

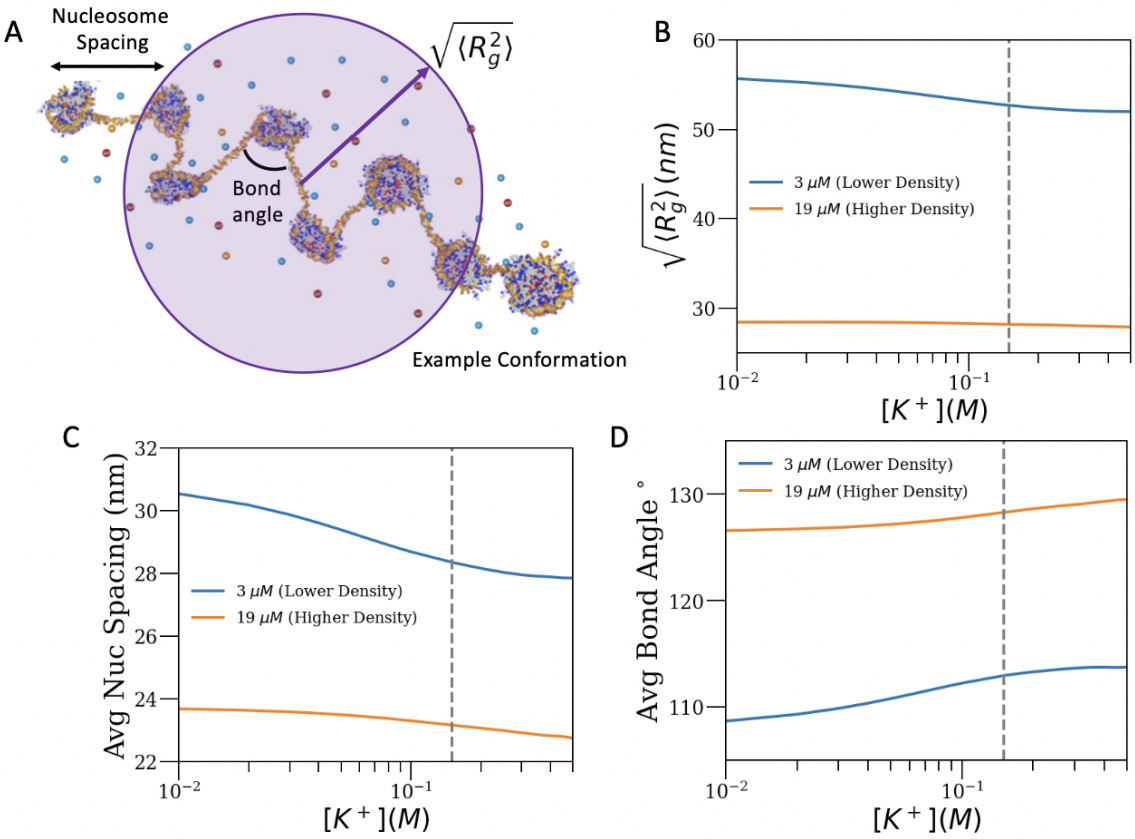

Figure 4.14. Influence of monovalent electrolyte conditions and bulk density on structure of 8-mer nucleosome arrays. Calculations were performed for lower density (3 μM) and higher density (19 μM) systems where concentration was determined by the number of chains within a given box size. All calculations were performed for physiological bulk $pH = 7.4$ and bulk $[Na^+] = [Mg^{2+}] = 0\,mM$. (A) Rendering of example conformation for 8-mer nucleosome array system with bond angle and nucleosome spacing explicitly denoted. (B-D) Influence of bulk $[K^+]$ on chromatin structure depends on bulk system density (Blue: lower bulk density, Orange: higher bulk density). (B) Radius of gyration (in nm), $\sqrt{\langle R_g^2 \rangle}$, representing the spread of the chromatin chain, (C) average nucleosome spacing (in nm) determined by averaging pairwise distance between neighboring nucleosomes (as in A), and (D) bond angle (in degrees) determined by three neighboring nucleosomes (as in A). For (B-D) physiological monovalent cation concentration $\sim 150\,mM$ is denoted by grey dotted line.

4.5. Conclusions and Discussion

In this thesis work, we have examined the charge regulation mechanisms of the bulk electrolyte environment, pH, and system density, on the charge and structure of two strong polyelectrolyte systems. To accomplish this, we employed an MT approach that is able to accurately represent the chemical, or charged, state along with the structural state of polyelectrolytes, and is thus able to investigate charge regulation in physiologically dense systems such as the nucleus. In addition, it is computationally feasible to perform parameter scans over many bulk environmental conditions using this theoretical approach.

First, we studied the behavior of a simpler polyelectrolyte brush system consisting of DNA-phosphate loops (Fig. 4.1). We characterized the effects of monovalent and divalent cations on the structure and charge of these phosphate loop systems for both dilute and dense environmental conditions. Non-dilute phosphate loop systems are sensitive to changes in Na^+ and K^+ concentrations around physiological monovalent ion concentrations (Fig. 4.3, Fig. 4.5). We find that for non-dilute systems, 20-30% of the phosphates carry condensed counterions, and this ion pairing increases with system density (Fig. 4.5). However, higher density systems are less sensitive to changes in the monovalent electrolyte environment due to the high charge density within these system.

Additionally, we see a large degree of lateral heterogeneity in phosphate density distribution, electrostatic potential, and local pH of our phosphate loop system. This is exemplified by Fig. 4.6, Fig. 4.7, and Fig. 4.8. Thus, the bulk density primarily determines the charge density and the charge regulation of our phosphate loop systems, however local fluctuations in phosphate density also play a non-negligible role.

Notably, we were able to mechanistically distinguish between the effects of divalent and monovalent cations and quantitatively determine the importance of ion bridging reactions between two phosphates and one divalent Mg^{2+} cation. Although 1:1 Mg^{2+} binding has a larger influence on the charge and thickness of phosphate loop layers than monovalent Na^{+} and K^{+} cations (Fig. 4.9), the ability of divalent ions to form ion bridges has a more profound impact on charge and thickness of phosphate loop layers (Fig. 4.10). Ion bridging for physiologically relevant intracellular ranges of Mg^{2+} concentrations result in increased charge neutralization and collapse of the phosphate loops (Fig. 4.10). Our preliminary study with the phosphate loop system allowed us to fully characterize the effects of charge regulation of the physiochemical environment on strongly negative phosphate systems before increasing the complexity of the system to represent chromatin.

We extended our MT approach for the simplified phosphate loop system to study charge regulation of the more complex chromatin-like system for both single nucleosomes and chromatin chains composed of 8 nucleosomes (Fig. 4.2). In our chromatin-like system, the elementary units for DNA were phosphates, sugars, and nucleobases and for histones were amino acids. Employing our simpler phosphate loop system, we were able to determine physiologically relevant free energies for phosphate-ion condensation reactions, which we used to study the effects of the bulk monovalent electrolyte environment, pH, and system density on chromatin. Both the charge and counterion cloud of our single nucleosome systems are sensitive to changes in monovalent cation concentration, although the exact behavior of yeast and human nucleosomes is distinct (Fig. 4.11). Additionally, the charge of our chromatin-like system is sensitive to changes in bulk pH, the charge regulation effects of monovalent ion condensation are even larger (Fig. 4.12). Our 8-mer

nucleosome array system also has a larger negative charge per nucleosome than the single nucleosome system due to the addd effects of linker DNA (Fig. 4.12). Finally, the sensitivity of the structure of chromatin-like chains to changes in $[K^+]$ is highly dependent on bulk density. The radius of gyration, average nucleosome spacing, and bond angles of higher density systems are, in general, less sensitive to $[K^+]$ for higher bulk density systems due to their increased compaction and more negative charge density (Fig. 4.13 &, Fig. 4.14). In general, "local" structural parameters (e.g., average nucleosome spacing and bond angle) are more sensitive to changes in the monovalent electrolyte environment than "global" descriptors (e.g., radius of gyration.

We would like to highlight the novelty of these computational results. Previous atomistic MD simulations have uncovered the effects of the electrolyte environment on system charge, nucleosome unwrapping, and the ion cloud surrounding chromatin [183, 297]. However, the limitations of such computationally intensive simulations prevent scanning over many environmental conditions, especially for denser system, and have mostly been performed for single nucleosomes. Less accurate coarse-grained simulations have investigated the effects of the electrolyte environment on, for example, the persistence length and the inter-nucleosome interactions of chromatin chains [159, 104]. However, the representation of electrostatics (see Chapter 1.4.2) and a more robust understanding of how system density influences the structure and charge of chromatin systems, have not been thoroughly studied. Here, we present a large-scale study on phosphate-like loops, a simple representation of DNA, and chromatin-like nucleosomes and chains which characterize the effects of charge regulation over many different environmental conditions. Overall, we quantitatively determine that bulk system density has a large effect on the

charge regulation of the electrolyte environment on DNA and chromatin structure, with substantial fluctuations in local charge and pH in three-dimensions that are not taken into consideration with coarse-grained simulation techniques.

There are also several limitations of our methodology we would like to discuss. Firstly, we are modeling a simplified representation of DNA-phosphates as loops. Each monomer only has one chargeable site to represent the chemical behavior of one phosphate and thus does not completely represent the DNA system. However, we believe the effects of divalent cations due to their ion bridging mechanisms are generalizable. For example, previous MT calculations examining the effects of Ca^{2+} concentration on a polyelectrolyte layer of poly(acrylic acid) have demonstrated that ion bridging of divalent Ca^{2+} with two acetate monomers results in collapse of the polyelectrolyte layer under specific environmental conditions [200], similar to the one observed here for the phosphate loop system. That MT study used as input linear chains generated from a rotational isomeric state (RIS) model and acetate, which is a weaker acid than phosphate [200], demonstrating that the effects of ion bridging are at least substantial for acids with a $pK_a < 4$. Additionally, previous experimental and computational studies have demonstrated the importance of divalent cations on the structure of dense brushes of strong polyelectrolytes. For example, Langevin simulations have been used to explore the effects of ions on polymer brushes [304, 107], and both simulations and Atomic Force Microscopy (AFM) experiments demonstrate lateral inhomogeneities induced by multivalent cations [304].

Interestingly, our study has found that the Mg^{2+}, (i.e., divalent) electrolyte environment has large effects on charge and structure of DNA with a "tipping point" around

estimated concentrations of free intracellular Mg^{2+} [244, 106]. Also of note is that strong ion bridging reactions are not always guaranteed, but depend on divalent ion concentration, ion bridging binding free energy, and bulk system density. By comparing MT calculations to experimental observations, we determine that ion bridging is not only strong in our phosphate loop system, but also has a very consequential influence on the overall charge and structure of the system.

An additional consideration is that MT is a mean-field approach that treats electrostatics interactions in a mean-field manner (see 4.6.1). Thus, if short-range charge-charge correlations are large enough, the electrostatic interactions are not as accurately represented by MT. To compensate for this limitation, we explicitly include short-range charge-charge correlations by representing them as ion binding reactions [100]. This is an intuitive way of reintroducing these short-range correlations into our system and has the additional advantage that it does not require detailed knowledge of ion-phosphate and ion-ion interactions. For example, the dynamics of Mg^{2+} solvation shells are estimated to be long-lived and are thus difficult to account for in atomistic simulations of DNA solution with Mg^{2+}[41].

There are several different methods to obtain the ion binding free energies of these ion condensation reactions, both experimental and computational. Experimental methods are preferred, but due to the complex chemical nature of DNA-phosphates and their strong acidic behavior, ion binding free energies between phosphate and counterions have not been robustly determined in the experimental literature.[3]. In principle, atomistic MD simulations could be used to obtain the ion-binding free energies. Observe that such

[3]Even $pKa = 1$ of phosphate is an estimate and, although this acid-base equilibrium constant has been cited in the literature [35], exact experiments have not been reported.

atomistic simulations provide detailed information regarding, for example, ion-DNA localization, but can have varying results due to the inherent difficulty in implementing correct force fields to represent ion-phosphate and ion-ion interactions [151, 44, 226]. Although recent advances in experimental methods have facilitated a more complete characterization of the ionic atmosphere surrounding DNA [60, 94, 93, 95] to better parametrize the force-fields of MD simulations [303], such approaches may still incorrectly predict trends which are not observed experimentally. For example, modeling of DNA nucleotides by Panteva et al. incorrectly predicted a size dependence of cation occupancy of DNA that was later disproven by ICP-MS measurements [219, 93]. Additionally, we have previously employed such atomistic simulation methods to determine ion binding constants of poly(acrylic acid), but find that these values are sensitive to water models and force fields [221]. To avoid these potential pitfalls, here we obtain binding free energies, which are needed as input for the MT calculations, by performing parameter scans over ion binding constants to determine physiologically relevant binding of K^+, Na^+, and Mg^{2+} counterions and then comparing to experimental observations of DNA charge [95, 94] and structure [76, 259, 260].

Future directions for our chromatin study include investigating the effects of Mg^{2+} as well as additional physiologically prevalent multivalent cations, such as trivalent Spermidine (Spd^{3+}) and tetravalent Spermine (Spm^{4+}), both of which have even higher valency than Mg^{2+}. Comparing yeast and human nucleosomes, we demonstrated that the amino acid composition of histones determines charge regulation at the level of nucleosomes. As there is crystallography data available, it could be interesting from an evolutionary perspective to characterize the charge regulation of the physicochemical environment

on the nucleosomes of other organisms, including archaea. We could also quantify the effects of histone variants (e.g., H2A.Z) on these charge regulation effects. In our chromatin studies, we have not included the histone tails, which have been shown experimentally and computationally to play an important role in salt-induced chromatin compaction [259, 7]. An important future direction would also be to add histone tails to our chromatin chain system and investigate the effects of the physicochemical environment on structure and charge for different chemical modifications (e.g., acetylation, methylation). Finally, we are currently modeling a system in equilibrium. However, chromatin, and most biological systems, are nonequilibrium systems. Thus, we could extend our equilibrium MT approach to model a nonequilibrium system where, for example, ion concentrations are in flux, using approaches previously developed in the Szleifer group [279].

4.6. Supplementary Material

4.6.1. Theoretical Approach

The $-TS_{\text{mix}}$ in the free energy (Eq. 4.11) corresponds to the mixing or translational entropy of the solvent (water) and the mobile ionic species

$$-\frac{TS_{mix}}{k_B} = \sum_k \int \mathrm{d}^3 r \rho_k(\overrightarrow{r})(\ln \rho_k(\overrightarrow{r}) v_w - 1). \tag{4.20}$$

The index k runs over all the different types of mobile species: the water molecule, cations (Na^+,K^+,Mg^{2+}), anion (Cl^-), proton (H^+), and hydroxyl ion (OH^-). The variable $\rho_k(\overrightarrow{r})$ corresponds to the number density of mobile species k and v_w is the volume of a water molecule, which is used as the unit of volume.

For the F_{chem} free energy term in Eq. 4.11, the variable $\mu_i^{\ominus}$ corresponds to the standard chemical potential molecule of type i. The first and third terms within the Eq. 4.15 describe the entropy of the deprotonated charged state (P^-) and protonated state (PH), respectively. The second and fourth terms in Eq. 4.15 correspond to the standard chemical potential of the charged and uncharged state, respectively. The subsequent terms in the equation have an identical meaning and pertain to the enthalpic and entropic contribution arising from the ion-condensation of Na^+ and K^+. The factor of half in front of $f_{P_2Mg}(\overrightarrow{r})$ occurs because the density of P_2Mg complexes is $\frac{1}{2} f_{P_2Mg}(\overrightarrow{r})\langle \rho_P(\overrightarrow{r})\rangle$ since the Mg^{2+} ion bridge consists of two phosphate. The second to last line describes the reduction in entropy associated with the reduction of the number of phosphate states as ion bridges are formed.

The F_{elect} term in the free energy functional describes the electrostatic contribution to the free energy and is given by

$$F_{\text{elect}} = \int \mathrm{d}^3 r \left[\langle \rho_q(\overrightarrow{r}) \rangle \psi(\overrightarrow{r}) + \frac{1}{2} \epsilon_0 \epsilon_w (\nabla_r \psi(\overrightarrow{r}))^2 \right]. \tag{4.21}$$

In this electrostatic functional ϵ_0 and ϵ_w correspond to the dielectric permittivity of vacuum and the dielectric constant of water, respectively. $\psi(\overrightarrow{r})$ is the electrostatic potential and $\langle \rho_q(\overrightarrow{r}) \rangle$ is the total charge density. Variation of the above functional with respect to the electrostatic potential yields the Poisson equation. Here, it should be noted that local electrostatic potential $\psi(\overrightarrow{r})$ is coupled with osmotic pressure $\pi(\overrightarrow{r})$. Additionally, the Poisson Equation, charge density and electrostatic potential are replaced by the thermodynamic averages as a consequence of the mean-field approximation. Hence, fluctuations and short range electrostatic correlations are not considered directly. However, we represent physical processes which occur due to shorter range electrostatic interactions using a chemical equilibrium approach, which provides an intuitive way of introducing electrostatic 'correlations' and short range electrostatic interactions at a mean-field level.

The total charge density is the sum of the charge density of all charged mobile ions and the charge density of the chargeable polyelectrolyte residues. For example, for the phosphate loop system

$$\langle \rho_q(\overrightarrow{r}) \rangle = \sum_i e z_i \rho_i(\overrightarrow{r}) + e[(-1) f_{P^-}(\overrightarrow{r}) + f_{PMg^+}(\overrightarrow{r})] \langle \rho_P(\overrightarrow{r}) \rangle. \tag{4.22}$$

Here, the summation runs over all charged mobile ions Na^+,K^+,Mg^{2+},Cl^-,H^+,OH^- with z_i corresponding to their valency. e is the unit of charge. The second term is the net charge number density of the phosphate residues.

The total free energy is minimized with respect to the p.d.f. ($P_g(\alpha)$) as well as the number density of all species, $\rho_i(\overrightarrow{r})$, the fraction of the different states the chargeable residues can be found, $f_k(\overrightarrow{r})$, and varied with respect to the electrostatic potential, $\psi(\overrightarrow{r})$, under the constraints of incompressibility and the fact that the system is in contact with a bath of cations, anions, protons, and hydroxyl ions. A Lagrange multiplier is introduced to enforce that the sum over all states of chargeable residues equals one to ensure conservation.

Minimization of the free energy yields the following expression for the local volume fraction of the solvent

$$\phi_w(\overrightarrow{r}) = \rho_w(\overrightarrow{r})v_w = \exp(-\beta\pi(\overrightarrow{r}))v_w, \tag{4.23}$$

while the density of the ions (for a given ion γ) reads

$$\rho_\gamma(\overrightarrow{r}) = \frac{1}{v_w}\exp\big(\beta\mu_\gamma - \beta\mu_\gamma^{\ominus} - \beta\pi(\overrightarrow{r})v_\gamma - \beta\psi(\overrightarrow{r})z_\gamma e\big). \tag{4.24}$$

For the phosphate loop system, minimization of the free energy with respect to the different states of the phosphate residue, $f_{P^-}(\overrightarrow{r})$, $f_{PH}(\overrightarrow{r})$, $f_{PNa}(\overrightarrow{r})$, and $f_{PK}(\overrightarrow{r})$

$$\frac{f_{P^-}(\overrightarrow{r})}{f_{PH}(\overrightarrow{r})} = e^{-\beta\Delta G^{\ominus}_{PH}} e^{-\beta\Delta G^{solv}_{PH}(\overrightarrow{r})} \frac{e^{-\beta\pi(\overrightarrow{r})\Delta v_{PH}}}{\rho_{H^+}(\overrightarrow{r})v_w}, \tag{4.25}$$

$$\frac{f_{P^-}(\overrightarrow{r})}{f_{PNa}(\overrightarrow{r})} = e^{-\beta\Delta G^{\ominus}_{PNa}} e^{-\beta\Delta G^{solv}_{PNa}(\overrightarrow{r})} \frac{e^{-\beta\pi(\overrightarrow{r})\Delta v_{PNa}}}{\rho_{Na^+}(\overrightarrow{r})v_w}, \tag{4.26}$$

$$\frac{f_{P^-}(\overrightarrow{r})}{f_{PK}(\overrightarrow{r})} = e^{-\beta\Delta G^{\ominus}_{PK}} e^{-\beta\Delta G^{solv}_{PK}(\overrightarrow{r})} \frac{e^{-\beta\pi(\overrightarrow{r})\Delta v_{PK}}}{\rho_{K^+}(\overrightarrow{r})v_w}, \tag{4.27}$$

The variable $\Delta G^{\ominus}_i$ is the standard reaction free energy change of either the acid-base equilibrium reaction of the acid or the dissociation equilibrium reaction of the metal-ion pairs (e.g., PNa, PK). Here, Δv_i corresponds to the difference in volume between the products and reactants. This formalism can be extended for each of the elementary chargeable units of the chromatin-like system.

4.6.2. Input Polymer Models

4.6.2.1. Phosphate Loop System.

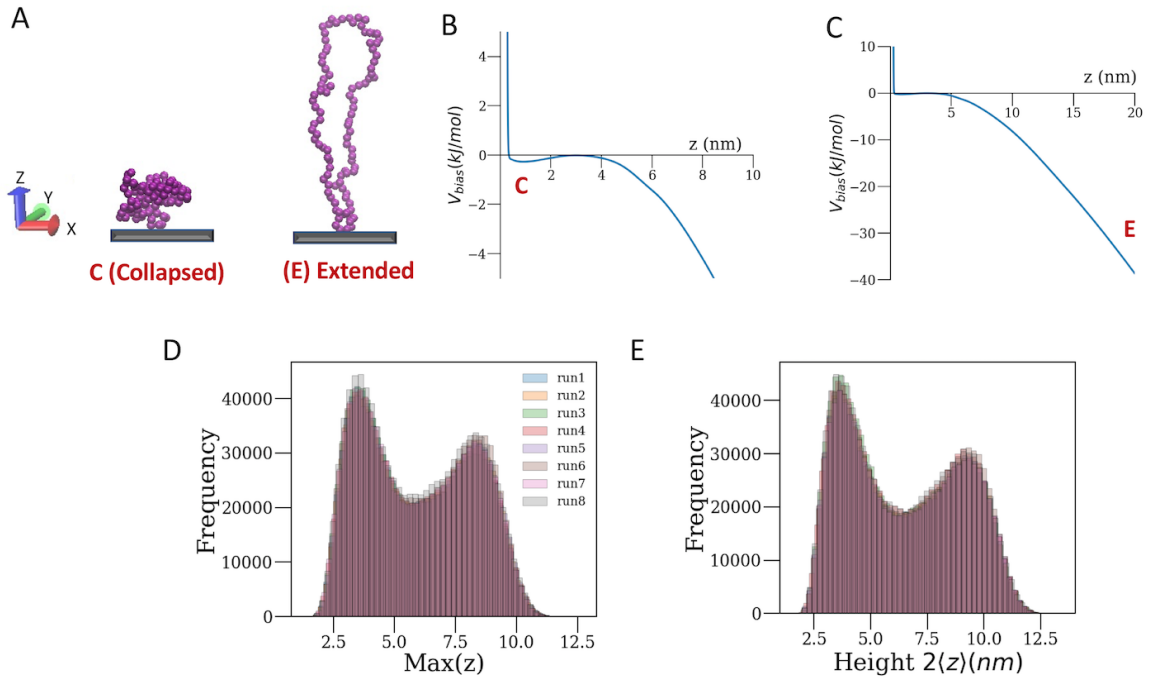

Figure 4.15. Biased sampling for chain generation of coarse-grained DNA-phosphate loops. (A) VMD renderings of examples of Collapsed (C) and Extended (E) conformations from GROMACS MD simulations performed with a bias potential. (B-C) Bias potential V_{bias} in units of kJ/mol versus z (in nm). This is a two-well potential that biases the MD trajectory towards both more collapsed (C) and more extended (E) states to increase the conformational space of the loops. (D-E) Distribution of (D) the maximum z value (in nm) and (E) height (in nm) for all conformations in all 8 production runs. Here, height is calculated as $2\langle z \rangle$.

4.6.2.2. Chromatin-like System. For the chromatin system, sizes of elementary units were determined by calculating the average minimum distances between neighboring amino acid residues in the single nucleosome system for amino acids and between phosphates, sugars, and nucleobases for a sample of 3SPN simulations. Segment lengths (l_{seg}) and chemical behavior of each nucleobase are shown in Tables 4.1 & 4.2. Volumes of each elementary unit were calculated assuming each unit is a sphere with volume $V = \frac{4}{3}\pi\left(\frac{l_{seg}}{2}\right)^3$.

Additionally, for chromatin-level MT calculations, three amino acids from the central nucleosome are fixed to the center of the box and rotations are performed around these fixed points to increase the conformational space of the 1CPN MD simulations.

DNA unit	$l_{seg}(nm)$	Chemical Behavior
Adenine (A)	0.333 nm	Neutral
Thymine (T)	0.438 nm	Neutral
Cytosine (C)	0.41 nm	Neutral
Guanine (G)	0.283 nm	Neutral
Sugar (S)	0.414 nm	Neutral
Phosphate (P)	0.3 nm	Acidic ($pK_a = 1$)

Table 4.1. Properties of DNA units of chromatin-like system.

Amino Acid Residue	$l_{seg}(nm)$	Chemical Behavior
Alanine (Ala)	0.484 nm	Neutral
Arginine (Arg)	0.484 nm	Basic ($pK_a = 12.1$)
Asparagine (Asn)	0.484 nm	Neutral
Aspartic acid (Asp)	0.484 nm	Acidic ($pK_a = 3.71$)
Cysteine (Cys)	0.484 nm	Acidic ($pK_a = 8.14$)
Glutamine (Gln)	0.484 nm	Neutral
Glutamic acid (Glu)	0.484 nm	Acidic ($pK_a = 4.15$)
Glycine (Gly)	0.484 nm	Neutral
Histidine (His)	0.484 nm	Basic ($pK_a = 6.04$)
Isoleucine (Ile)	0.484 nm	Neutral
Leucine (Leu)	0.484 nm	Neutral
Lysine (Lys)	0.484 nm	Basic ($pK_a = 10.67$)
Methionine (Met)	0.484 nm	Neutral
Phenylalanine (Phe)	0.484 nm	Neutral
Proline (Pro)	0.484 nm	Neutral
Serine (Ser)	0.484 nm	Neutral
Threonine (Thr)	0.484 nm	Neutral
Tyrosine (Tyr)	0.484 nm	Acidic ($pK_a = 10.1$)
Valine (Val)	0.484 nm	Neutral

Table 4.2. Properties of amino acid residues of chromatin-like system.

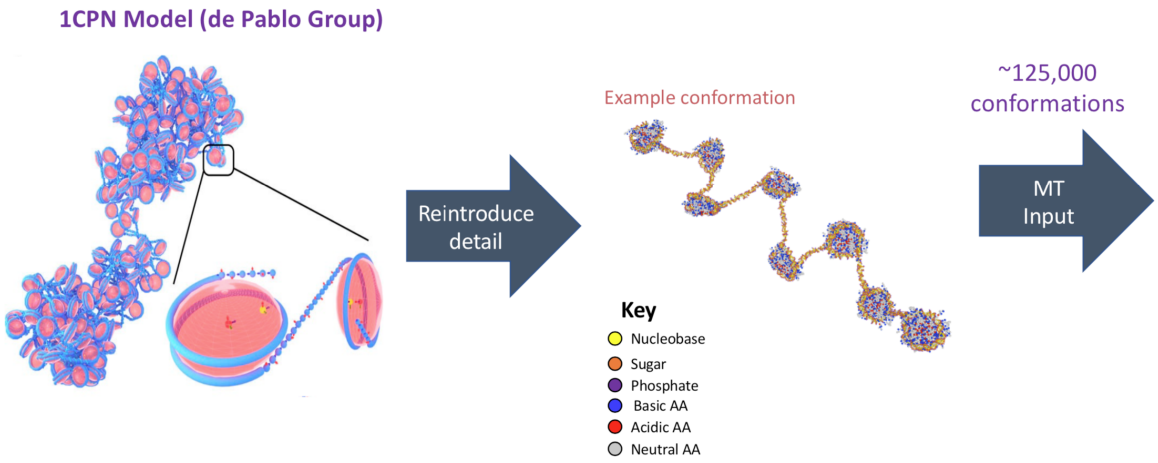

Figure 4.16. Conformations for 8 nucleosome chromatin-like system were generated by first performing 1CPN MD simulations [159] and then re-introducing molecular detail. Elementary units of DNA are sugars, phosphates, and nucleobases, as represented by the 3SPN model of DNA [88]. Elementary units of histones are amino acid residues where coarse-graining was performed using the AICG method [165]. A total of ∼125,000 such conformations were input into the chromatin-level MT of 8-mer chains.

4.6.3. Additional Supplementary Figures

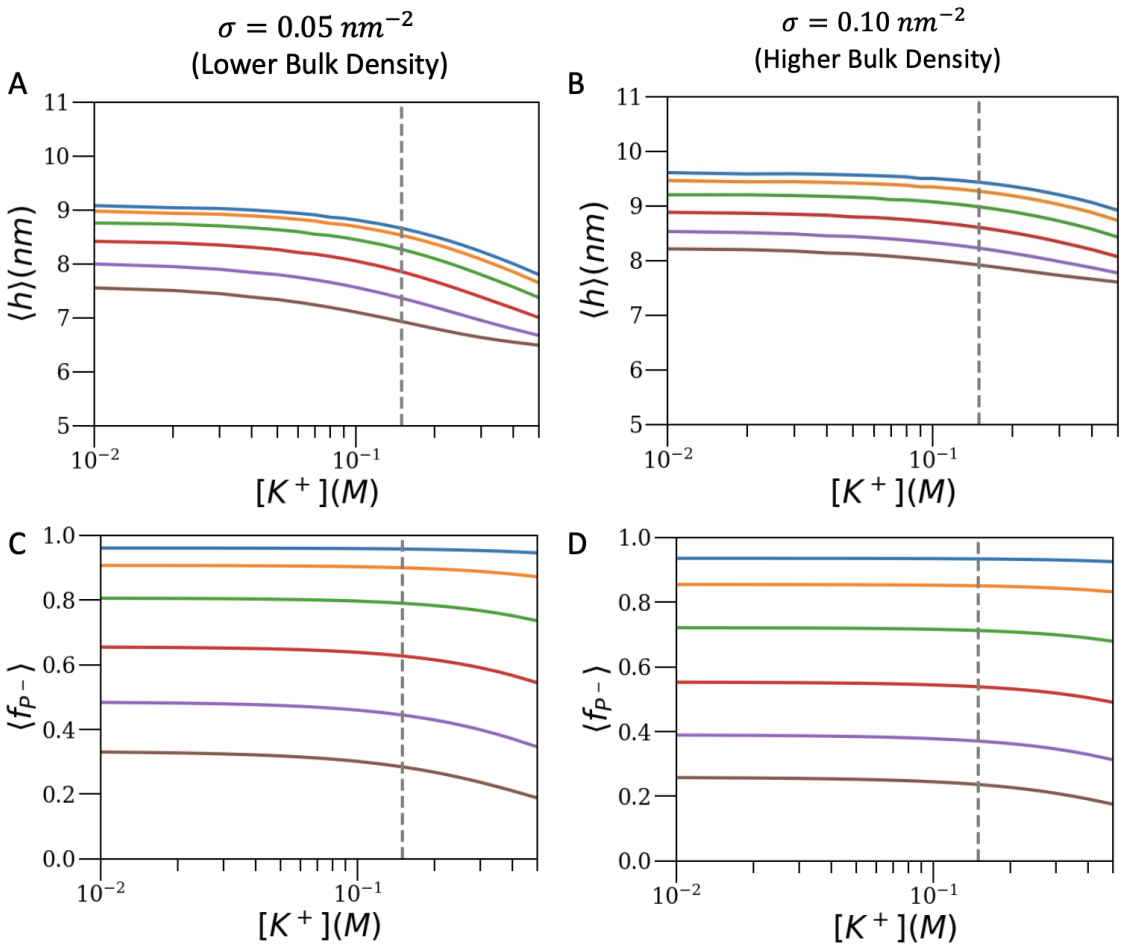

Figure 4.17. Effect of monovalent cations for dense phosphate loop systems depends on ion binding strength for lower ($\sigma = 0.05\,nm^{-2}$) and higher ($\sigma = 0.10\,nm^{-2}$) bulk densities under non-dilute conditions. Compared to dilute systems, 1D and 3D calculations are more comparable for the (A&B) height of loop system and (C&D) average charged fraction of phosphates.

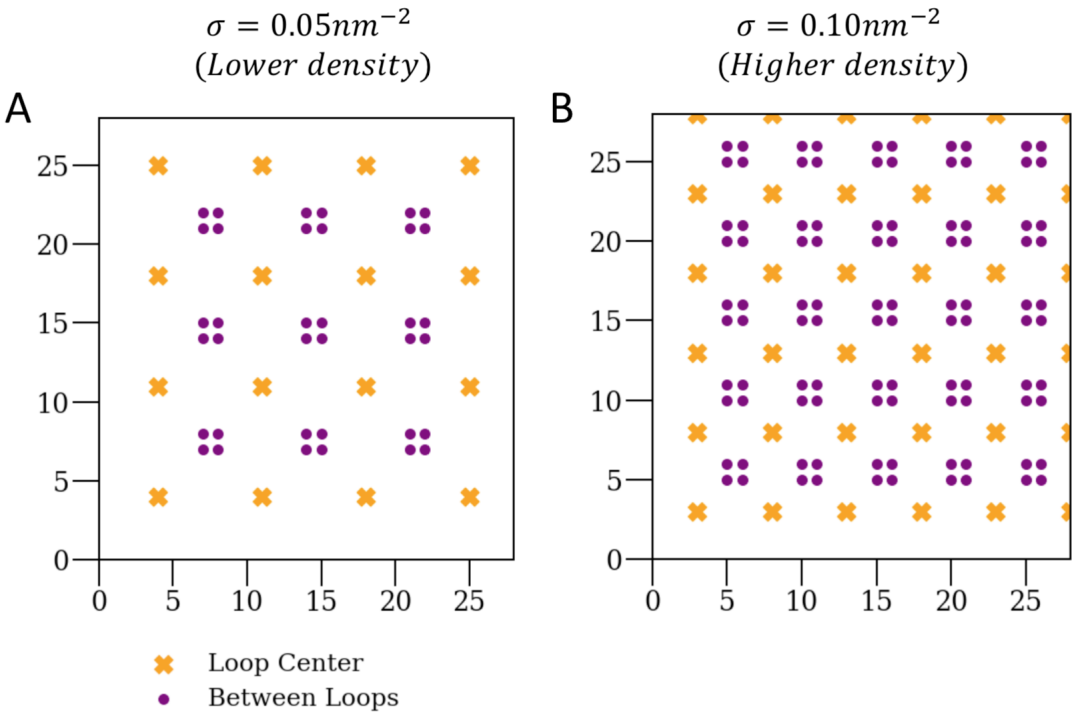

Figure 4.18. Indices for average z analysis: domain center versus between loop regions for (A) $\sigma = 0.05\,nm^{-2}$ and (B) $\sigma = 0.10\,nm^{-2}$. "Loop center" regions are represented by orange X's and "between loop" regions are represented by purple dots and are equidistant from loop centers on the planar lattice.

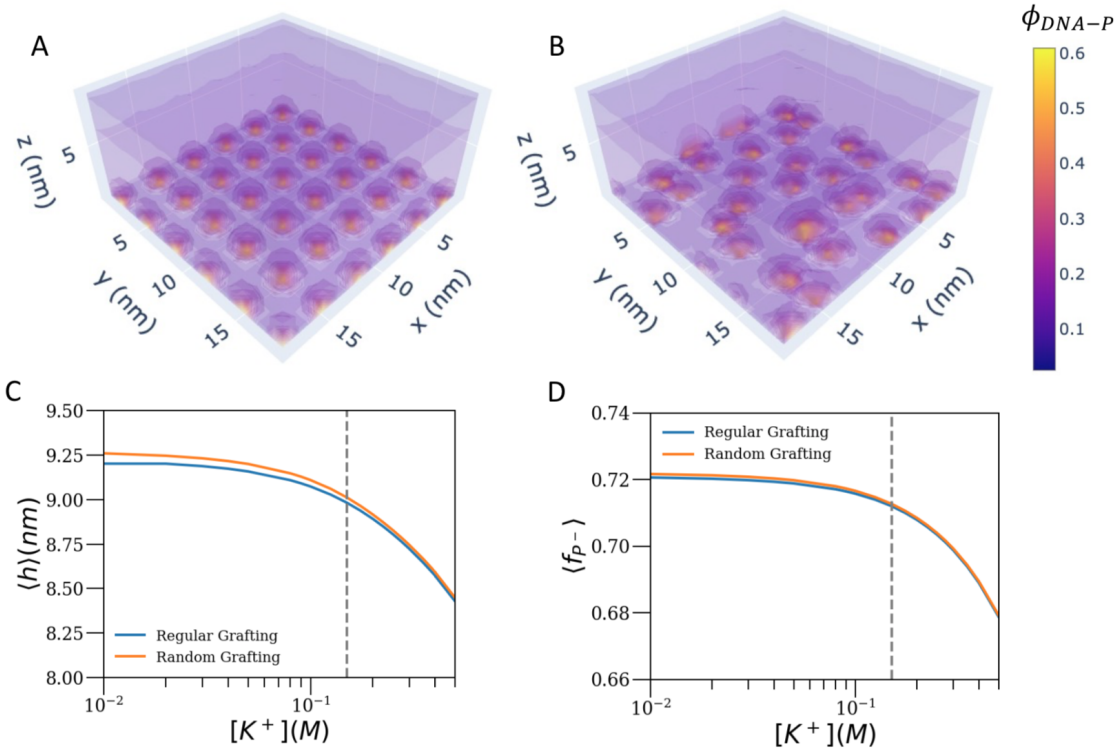

Figure 4.19. Effects of random compared to homogeneous grafting of phosphate loops. (A-B) 3D isosurface plot of DNA-phosphate volume fraction, ϕ_{DNA-P} for a phosphate loop system with (A) regular and (B) random grafting patterns for higher bulk density ($\sigma = 0.10\, nm^{-2}$), bulk $[K^+] = 150\, mM$, bulk $[Na^+] = [Mg^{2+}] = 0\, mM$ and physiological bulk $pH = 7.4$. Random grafting is achieved by shifting each homogeneously grafted loop by a random value generated from a random seed. Loop center are indicated by regions with higher phosphate volume fractions. (C) Height and (D) average charged fraction of phosphates have a similar dependence on changes in $[K^+]$ for homogeneously and randomly grafted systems indicating that, although there are differences between the two systems in the xy plane, their average properties are very similar.

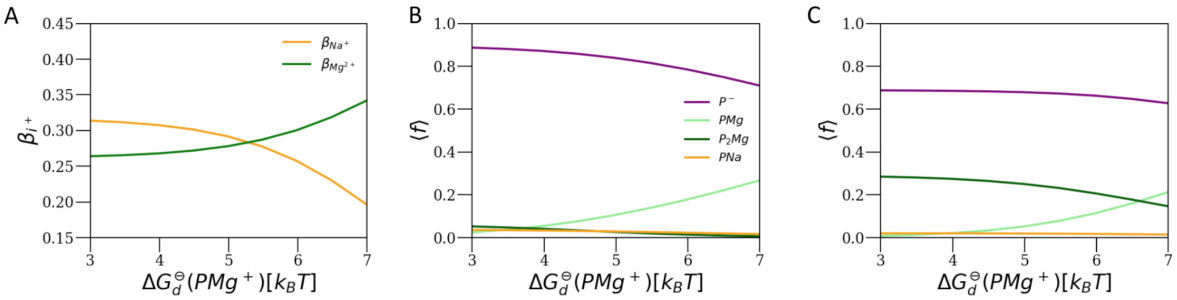

Figure 4.20. Effects of 2:1 Mg^{2+} binding on charge regulation of 1:1 Mg^{2+} binding in dilute systems for bulk $pH = 7.4$, bulk $[Na^+] = 25\,mM$, bulk $[Mg^{2+}] = 2.5\,mM$, bulk $[K^+] = 0\,mM$, and dilute bulk density ($\sigma = 0.0002\,nm^{-2}$). (A) Normalized excess ions, β_{i^+}, for Na^+ and Mg^{2+} versus $\Delta G_d^{\ominus}(PMg^+)$ for $\Delta G_d^{\ominus}(P_2Mg) = 11\,k_BT$ (higher ion bridging). (B-C) Fraction of each chemical state depends on 1:1 Mg^{2+} binding coefficient $\Delta G_d^{\ominus}(PMg^+)$ for (B) $\Delta G_d^{\ominus}(P_2Mg) = 7\,k_BT$ (lower ion bridging) and (C) $\Delta G_d^{\ominus}(P_2Mg) = 1\,k_BT$ (higher ion bridging).

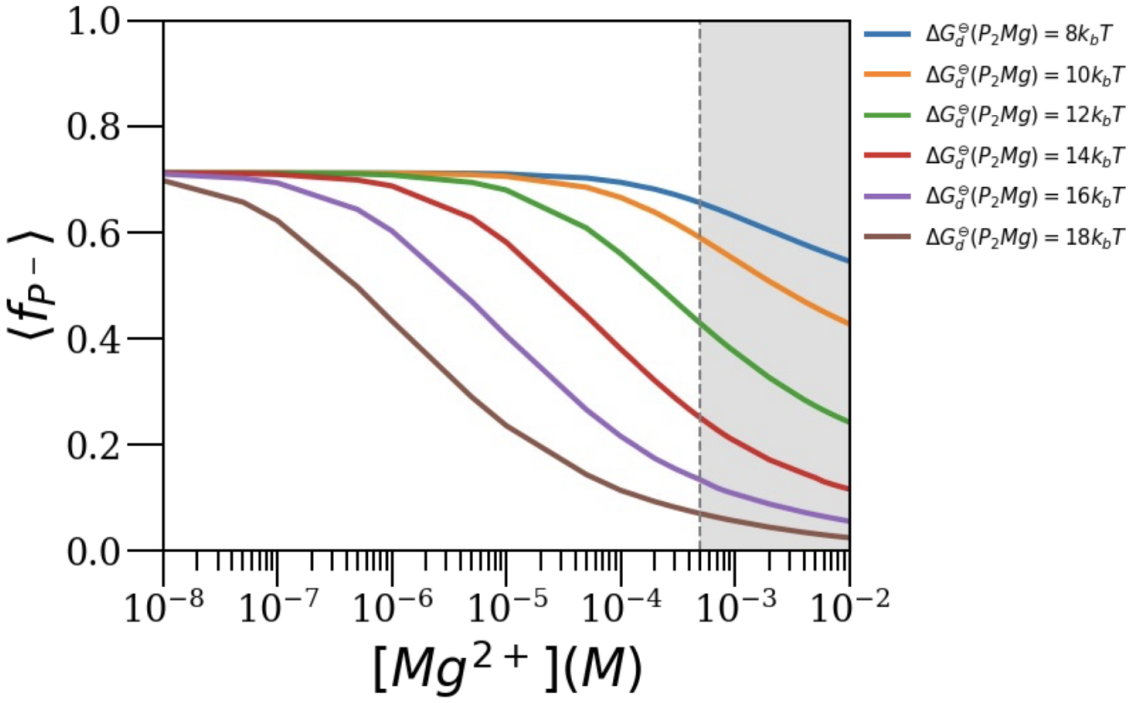

Figure 4.21. Influence of ion bridging coefficient on charged fraction for higher bulk density ($\sigma = 0.10\,nm^{-2}$), bulk $pH = 7.4$, and physiological monovalent cations (bulk $[K^+] = 140\,mM$, bulk $[Na^+] = 10\,mM$) corresponding to Fig. 4.10A. Each line represents different ion bridging coefficients, $\Delta G_d^{\ominus}(P_2Mg)$.

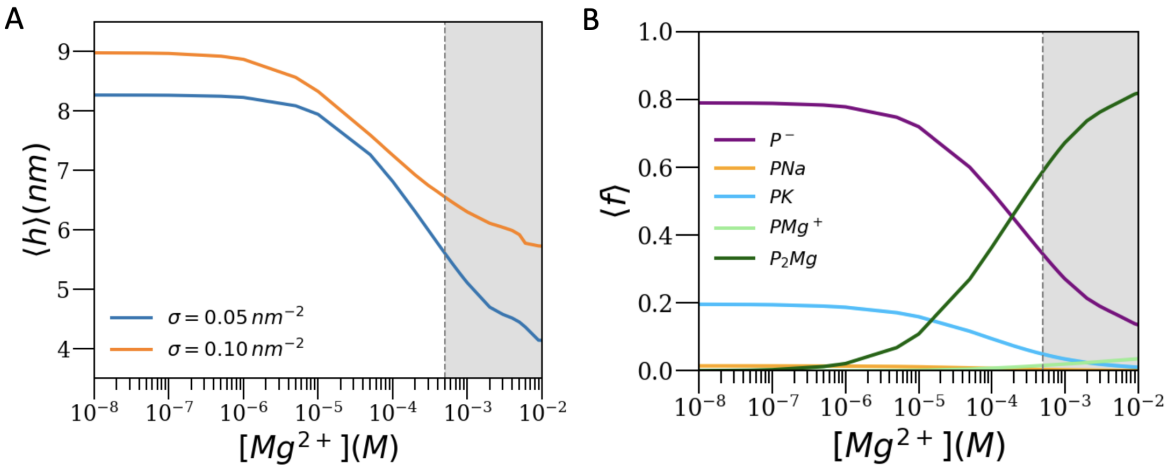

Figure 4.22. Effects of system density on charge regulation of Mg^{2+} via ion bridging. All calculations performed for higher bulk density ($\sigma = 0.10\,nm^{-2}$), bulk $pH = 7.4$, and physiological monovalent cations (bulk $[K^+] = 140\,mM$, bulk $[Na^+] = 10\,mM$). (A) Average height (in nm) versus $[Mg^{2+}]$ for lower ($\sigma = 0.05\,nm^{-2}$) compared to higher $\sigma = 0.10\,nm^{-2}$ grafting densities and physiological ion bridging $\Delta G_d^{\ominus}(P_2Mg) = 14\,k_BT$. (B) Fraction of phosphates with a negative charge (P^-), bound to Na^+ (PNa), bound to K^+ (PK), bound in 1:1 fashion to Mg^{2+} cation, or Mg^{2+} forming an ion bridge between two phosphates (P_2Mg) for $\Delta G_d^{\ominus}(P_2Mg) = 14\,k_BT$ and lower bulk density ($\sigma = 0.05\,nm^{-2}$) compared to Fig. 4.10C.

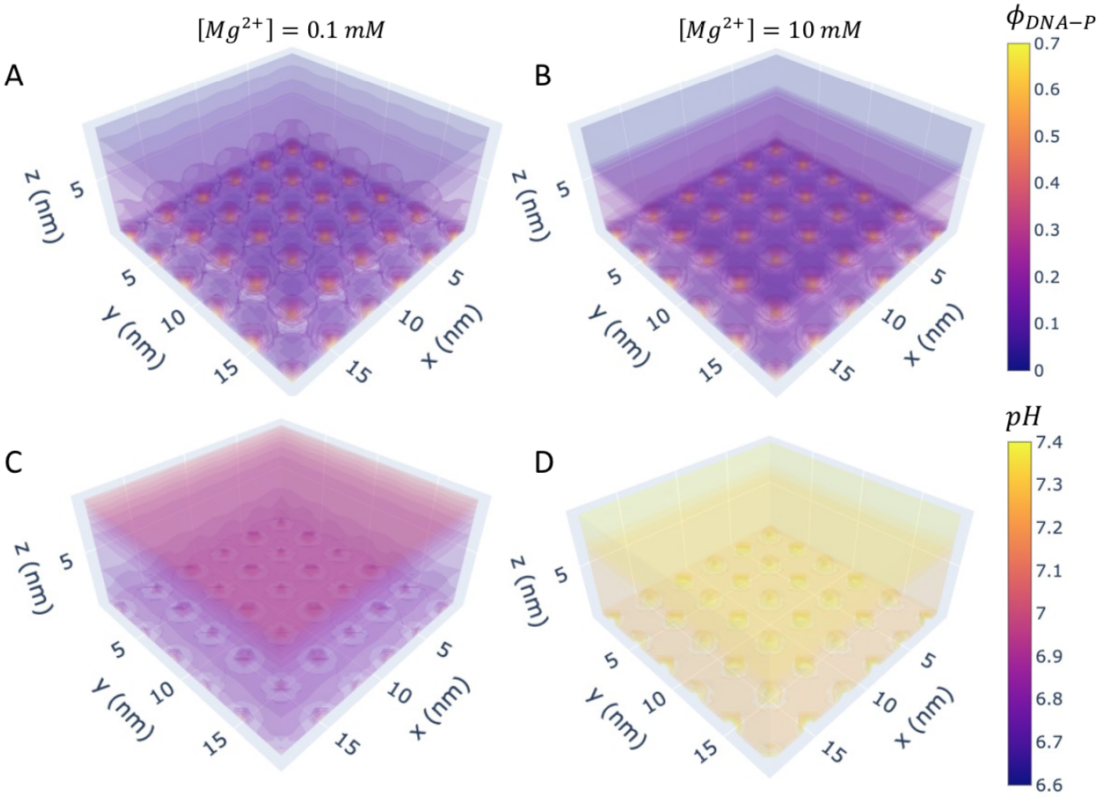

Figure 4.23. 3D distribution of DNA-phosphate density and local pH for lower and higher bounds of physiological Mg^{2+}. All calculations performed for higher bulk density ($\sigma = 0.10\,nm^{-2}$), bulk $pH = 7.4$, and physiological monovalent cations (bulk $[K^+] = 140\,mM$, bulk $[Na^+] = 10\,mM$), and physiological ion bridging $\Delta G_d^{\ominus}(P_2Mg) = 14\,k_BT$. (A&B) Distribution of phosphate volume fraction, ϕ_{DNA-P} for (A) lower bulk $[Mg^{2+}] = 0.1\,mM$ and (B) higher bulk $[Mg^{2+}] = 10\,mM$. As bulk $[Mg^{2+}]$ increases, the phosphate density distribution becomes compacted and more localized closer to the planar surface (i.e., towards smaller z values). (C&D) Local pH distribution for (C) lower bulk $[Mg^{2+}] = 0.1\,mM$ and (D) higher bulk $[Mg^{2+}] = 10\,mM$. As $[Mg^{2+}]$ increases, pH becomes more homogenous and closer to bulk pH values (i.e. lower local $[H^+]$).

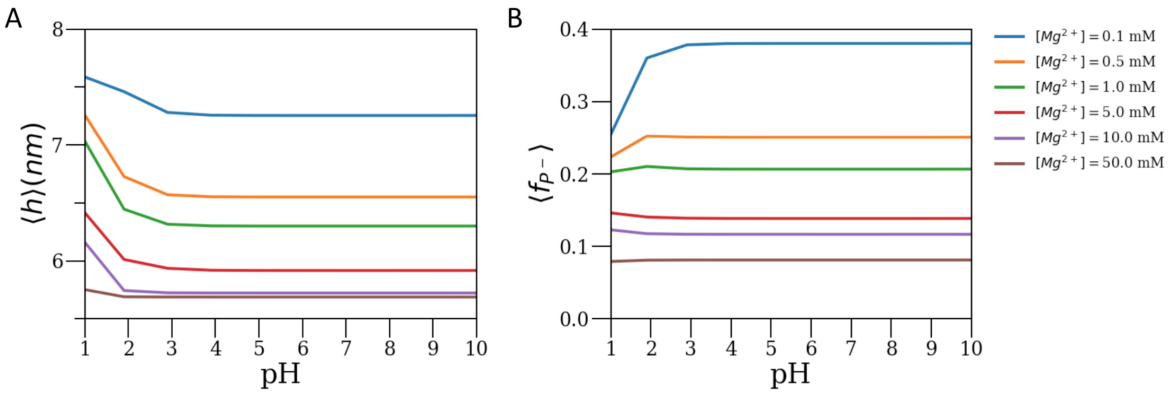

Figure 4.24. Dependence of phosphate loop structure and charge on bulk pH compared to bulk $[Mg^{2+}]$. (A) Average height (in nm) and (B) average charged fraction of phosphates for bulk pH ranging from 1 to 10 with varying bulk $[Mg^{2+}]$. Although these properties are sensitive to bulk pH at very low pH, they remain stable above $pH \approx 2$.

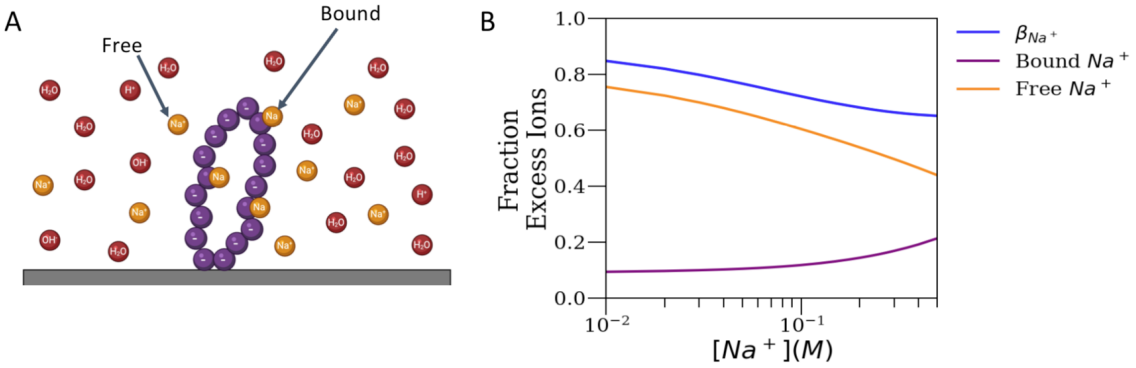

Figure 4.25. Excess ionic environment of dilute phosphate loop system under monovalent cation conditions for dilute bulk density ($\sigma = 0.0002\, nm^{-2}$), bulk $[K^+] = [Mg^{2+}] = 0\, mM$ and bulk $pH = 7.4$. (A) Representation of dilute phosphate loop system in a solvent with Na^+ cations that shield negative phosphate charges in both the free and the bound state. (B) Total excess ions (β_{Na^+}), fraction of bound excess Na^+, and fraction of free Na^+ for phosphate loop system.

CHAPTER 5

Summary of Thesis Work and Future Outlook

5.1. Major Conclusions

Employing a combination of nanoimaging modalities from our nano-ChIA platform (Chapter 1.3), genome mapping and RNA sequencing techniques, and several computational models, this thesis aims to identify and characterize several key aspects of chromatin structure and function in a realistic nuclear environment (Fig. 5.1).

In Chapter 2, we begin by employing high-resolution ChromSTEM and live-cell PWS microscopy to identify the existence of chromatin packing domains (PDs) (Fig. 2.1, Fig. 2.8, Fig 2.4), fundamental units of higher-order chromatin structure identified with polymer physics-based mass scaling analysis (Fig. 1.5). These PDs are spatially separable and have heterogeneous statistical packing properties, including chromatin packing scaling (D), average chromatin volume concentration (CVC), and size (R_f), and distinct morphological properties related to accessibility of transcriptional machinery (Fig. 2.2, Fig. 2.3). The existence of PDs and characterization of their properties was determined in both A549 lung adenocarcinoma cells and differentiated BJ fibroblasts (Fig. 2.9, Fig. 2.10) indicating they are a generalizable higher-order structure of chromatin organization in human cells.

The packing scaling of PDs was also determined to have several functional implications. The spatial distribution of D and its nuclear average over time are heritable across cell division (Fig. 2.12, Fig. 2.13). There is also a general inverse relationship between contact probability scaling and chromatin packing scaling, indicating a potential connection between a statistical descriptor of genome connectivity and chromatin packing behavior (Fig. 2.5, Fig. 2.6, Fig. 2.13, Fig. 2.14).

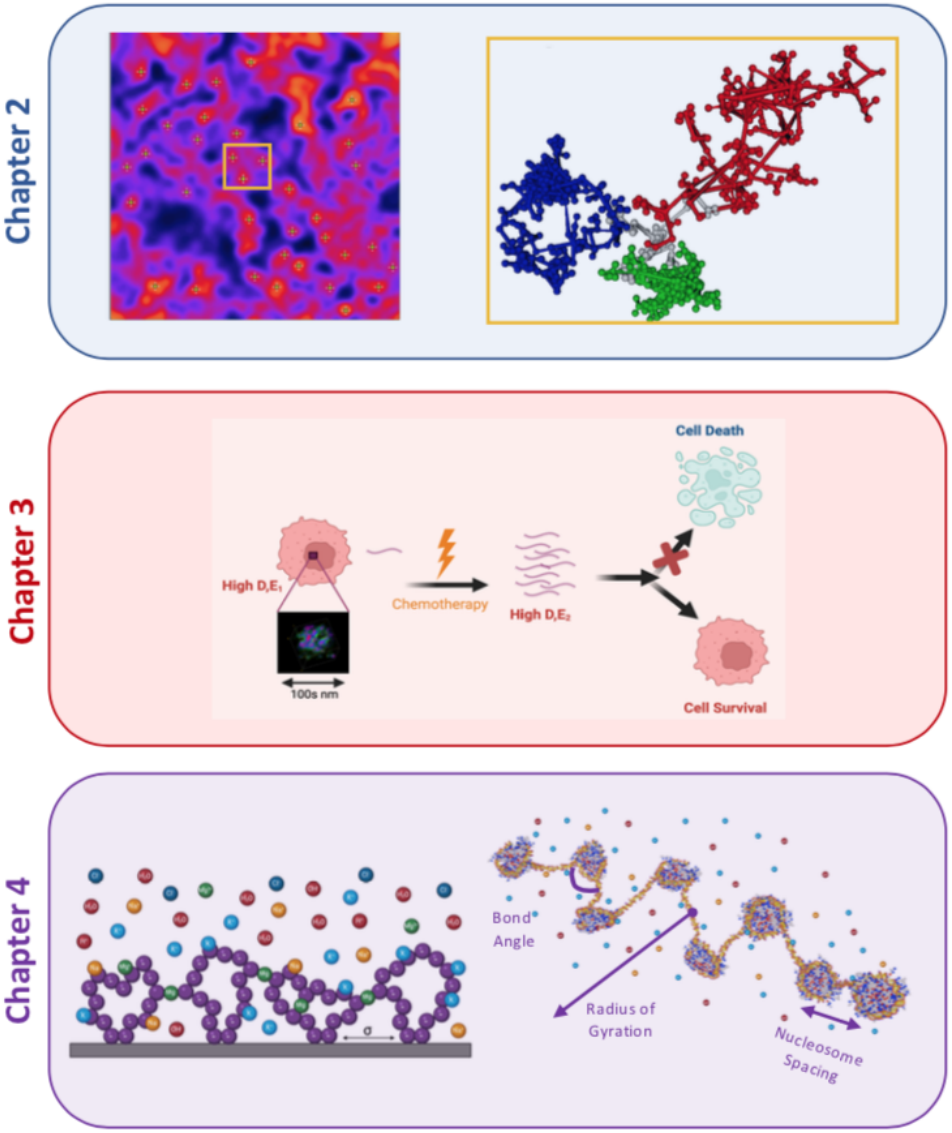
Chapter 2
Chapter 3
Cell Death
High D,E1
100s nm
Chemotherapy
High D,E2
Cell Survival
Chapter 4
Bond Angle
Radius of Gyration
Nucleosome Spacing

Figure 5.1. Summary of thesis work to determine underlying principles of chromatin structure and function. **Chapter 2** identifies the existence of chromatin PDs and that statistical chromatin organization can be described by the SR-EV model of chromatin. (Left) ChromSTEM tomogram of chromatin mass density distribution with PD centers marked with crosses. (Right) SR-EV-like rendering of chromatin PDs (red, green, blue) each with distinct statistical packing properties separated by low density chromatin (in grey) representing the SR-EV backbone. **Chapter 3** determines that the initial chromatin packing state of cells determines their ability to survive cytotoxic stressors. Specifically, high-D cancer cells are more likely to upregulate critical stress response genes upon exposure to chemotherapy, allowing them to survive chemotherapeutic stress. **Chapter 4** characterizes the effects of the physicochemical environment on DNA-like and chromatin-like systems. (Left) The ability of divalent Mg^{2+} (green cations) to form ion bridges between two phosphates allows for increased charge neutralization and compaction of these loop systems. These effects are dependent on average system density, determined by grafting density σ. (Right) The bulk monovalent electrolyte environment, pH, and system density influence the charge and structural parameters, including bond angle, nucleosome spacing, and radius of gyration, of chromatin.

Next, in Chapter 2, we introduce the Self-Returning Excluded Volume (SR-EV) polymer model, an extension of the Self-Returning Random Walk (SRRW) statistical model. Like the SRRW model, the SR-EV has heterogeneous step sizes to represent the heterogeneous mass density distribution of chromatin (Fig. 2.21). SRRW/SR-EV statistics are determined by the folding parameter, α, which modulates the number of self-returning steps to increase self-connectivity within domains (Fig. 2.21). Tree-like domains are interconnected by a sparsely packed backbone structure (Fig. 2.21). Unlike SRRW, the SR-EV model satisfies the principle of excluded volume and is a linear folded chain (Fig. 2.15).

We determined that statistical changes induced by the heat shock response, as measured by PWS microscopy and Hi-C, were recapitulated by changing just the α folding parameter of SR-EV. Such relevant properties include: (1) an increase in average chromatin packing scaling (Fig. 2.16, Fig. 2.17A), (2) a decrease in short-range and an increase in long-range contacts (Fig. 2.18), and (3) a decrease in number of domains (Fig. 2.19). In addition, the SR-EV was able to represent the heterogeneity of chromatin packing at the level of individual conformations (Fig. 2.17B). Altogether, the morphology and branching structure of domains is re-organized in the heat shock case as represented by Fig. 2.20. The fact that these large-scale statistical changes in chromatin packing and connectivity upon heat stress can be represented by a simple model, with only one tuning parameter, indicates that the representation of chromatin as an SR-EV polymer is a generalizable feature.

In Chapter 3, we developed a computational model, the chromatin packing macromolecular crowding (CPMC) model, that considers the crowding-mediated effects (see Chapter 1.4.4) of statistical chromatin PD structure (Fig. 3.1), thus modeling transcription in the context of a realistically crowded nuclear environment. Specifically, we determined the influence of PD descriptors chromatin packing scaling (D), genomic size (N_d, related to R_f), and average chromatin density ($\phi_{in,0}$, related to CVC) on large-scale gene expression patterns (Fig. 3.2). Chromatin packing scaling was identified as a strong determinant of phenotypic plasticity of cancer cells in response to chemotherapy, which was validated using nanoimaging and scRNA-seq techniques. Here, phenotypic plasticity encompasses transcriptional malleability (the average change in gene expression) and

intercellular heterogeneity (the range of functional states in a cell population) (Fig. 3.3, Fig. 3.4, Fig. 3.13, Fig. 3.14).

We then developed the Chromatin-Dependent Adaptability (CDA) model, an extension of CPMC, to predict cell survival in response to cytotoxic stressors from the initial chromatin packing state. Employing our CDA modeling, PWS imaging, and cell viability measurements, we determined that cancer cells with PDs that have higher average D are able to better upregulate genes beyond a potential critical threshold in order to survive chemotherapy (Fig. 3.6, Fig. 3.7). Indeed, chemotherapy was found to select for cells with increased average D (Fig. 3.3B&C, Fig. 3.11, Fig. 3.12) and lowering the average D of cell populations increased the effectiveness of chemotherapy to induce cell death, as predicted by CDA (Fig. 3.8, Fig. 3.16, Fig. 3.17). Notably, drug agents that are thought to alter the physicochemical environment of the cell by modulating ion channel activity had a larger effect on decreasing average D and increasing effectiveness of chemotherapy compared to other drugs, even ones that influenced epigenetics (Fig. 3.16, Fig. 3.17). As a next step, we wanted to determine principle biophysical mechanisms that predictably alter chromatin packing in order to be able to modulate cellular adaptability.

In Chapter 4, we developed a Molecular Theory (MT) approach to characterize the effects of the physicochemical environment, including bulk chromatin density, ions, and pH, on the structure and charge of chromatin. We began by modeling a simplified phosphate loop system to represent the chemical behavior of DNA (Fig. 4.1). Due to their ability to form ion bridges between two phosphates, Mg^{2+} has a much larger influence on physical compaction and charge neutralization of dense phosphate loop systems compared to monovalent cations, and these effects occur within physiologically relevant ranges of

free intracellular $[Mg^{2+}]$ (Fig. 4.5, Fig. 4.10, Fig. 4.23). The charge regulation effects of the electrolyte environment are also strongly influenced by average system density, with strong local fluctuations in phosphate density, electrostatic potential, and pH (Fig. 4.5, Fig. 4.6, Fig. 4.7, Fig. 4.8, Fig. 4.18, Fig. 4.19, Fig. 4.21, Fig. 4.22). Physiologically relevant ion condensation free energies for K^+, Na^+, and Mg^{2+} were determined by performing parameter scans for our phosphate loop system and comparing MT calculations to experimental observations (Fig. 4.3, Fig. 4.4, Fig. 4.9, Fig. 4.10).

Next, we expanded our MT approach to model a chromatin-like system. Comparing the behavior of yeast and human nucleosomes to our phosphate loop system, we demonstrated that amino acid composition modulates the sensitivity of the nucleosome charge and ionic environment to changes in monovalent salt concentration (Fig. 4.11, Fig. 4.25). Additionally, the electrostatic potential of chromatin is more sensitive to bulk pH fluctuations than the DNA-like phosphate loop system due to the acid-base equilibrium of acidic and basic amino acid residues, although ion condensation reactions have larger charge regulation effects (Fig. 4.12). The structure and charge of chromatin chains is modulated by a combination of bulk system density and the bulk electrolyte environment (Fig. 4.13, Fig. 4.14). Local structural parameters, including nucleosome spacing and bond angle, are still sensitive to changes in bulk monovalent ion concentrations even for higher density systems, although this sensitivity is diminished for concentrations $> 150\,mM$, i.e., the physiological concentration of monovalent cations (Fig. 4.14).

5.2. Future Work

Results from this thesis work open up several key questions. Firstly, what are the major mechanisms underlying PD formation, maintenance, and heritability. We have observed that inhibition of transcriptional elongation significantly alters PD structures and their average packing scaling [167]. Thus, transcription itself could play a role in modulating chromatin packing structure. More extensive studies employing genetic manipulation, such as auxin-inducible degron (AID) systems [206] and CRISPR interference (CRISPRi) [229], could be employed in the future to test which biophysical mechanisms are most important for PD structural stability. CRISPRi is a high-throughput platform, so many pathways related to transcription (e.g., RNA Pol II subunits, topoisomerase, Mediator complex), loop extrusion (e.g., CTCF and cohesin subunits), and other relevant pathways could be targeted at once. Here, PWS microscopy could be employed as a phenotypic readout to determine which biophysical mechanisms have the largest influence on PD structure.

After these mechanisms have been identified, the more tunable AID system could degrade specific target proteins and be used for more complex studies. For example, the heritability analysis performed in Fig. 2.11 & Fig. 2.12 could be employed to investigate how perturbing PD structure influences heritability. Specifically, we would investigate whether the perturbed structure caused by genetic manipulation is still inherited across cell division or if PD structure between parent and progeny cells is no longer correlated. Such studies could help to discern fundamental mechanisms that contribute to the heritability of PD packing behavior. Additionally, ChromSTEM experiments combined with the polymer physics-based analysis described in Chapter 2 could further probe changes in

statistical and morphological PD structure upon genetic perturbation. Finally, Hi-C experiments could be performed to assess how the perturbation influences TAD structures, and then compare these to changes observed in PD structures using PWS microscopy and ChromSTEM.

Additionally, modifications to the SR-EV model could help probe which biophysical mechanisms are most necessary for domain maintenance. We could introduce forces by adding external proteins which can reversibly modify chromatin either chemically (e.g., epigenetic modifiers such as HDACs) or topologically (e.g., nuclear lamins, CTCF-cohesin proteins). MD simulations could be performed until equilibrium is reached to determine the minimal amount of biophysical mechanisms necessary to maintain statistical behavior of tree domain structures over time.

Another question stemming from this thesis work involves the contributions of molecular (e.g., RNA Pol II and TF concentrations) compared to physical regulators of transcription (e.g., chromatin packing scaling) on transcriptional plasticity, specifically in response to chemotherapy. Currently, our CPMC and CDA models both assume that molecular regulators of transcription do not change significantly after stimulation with cytotoxic stressors such as chemotherapy. The heat shock response discussed in Chapter 2.3.3 could be an interesting model to answer this question. This stress response increases D of cancer cells, thus modulating the initial chromatin packing state, and has well-defined molecular mechanisms of action. The transcriptional response of heat shock is regulated by the TF heat shock factor 1 (HSF1) [231], which would be expected to alter the molecular regulators of transcription upon heat shock. Interestingly, HSF1 is also implicated in chemoresistance [140]. We could potentially expose WT and HSF1-KD cells

to heat stress and then expose these sensitized cell populations to cytotoxic chemotherapy. Next, we would experimentally determine the cell survival benefit of pre-acclimation to heat stress, which would alter the average chromatin packing behavior before exposure to chemotherapy. These experiments would be performed for WT cells, which, upon heat stress, would increase molecular regulators of HSF1 and stress response genes potentially related to chemoevasion, and HSF1-KD cells. We can then compare these experimental observations to CPMC predictions to evaluate differences in cell survival to cytotoxic stress that can be attributed to changes in the initial chromatin packing state of cell populations.

If we combined these heat experiments with CRISPR-Sirius, we could also track heat shock-related genes and determine if they become localized to PDs with higher D after repeated exposure to heat shock. Such an observation would indicate that increasing the packing scaling surrounding stress response genes is a cellular stress response mechanism.

SR-EV could also be employed to better answer this complex question. If we explicitly add transcriptional machinery to the SR-EV model, which is able to represent statistical chromatin structure in both control and heat shock states, then we could predict transcriptional responsiveness to external stimulation by combining SR-EV simulations with the CPMC model. We could then deconvolute if experimentally observed changes in transcription, as assessed by RNA-seq, could be explained mostly be the rearrangements in chromatin structure due to heat shock, or if we also need to account for changes in molecular regulators.

Finally, we would like to characterize the effects of the physicochemical environment on statistical chromatin packing within PDs. Preliminary *in vitro* experiments are being performed to investigate the influence of the electrolyte environment on chromatin packing in live cells. Our MT results on phosphate loop systems point to the increased effects of Mg^{2+}, the most prevalent intracellular divalent cation, on DNA structure compared with monovalent cations K^+ and Na^+. We are currently performing experiments with Mg^{2+} chelators, monitoring changes in ion concentrations using fluorescent dyes, and tracking changes in chromatin packing using live-cell PWS microscopy. Chelation is performed using BAPTA, a widely available divalent cation chelator, and APDAP, which has a much higher affinity for Mg^{2+} compared to Ca^{2+} and other divalent cations [290]. Changes in chromatin packing will be compared to changes observed upon K^+ perturbation with valinomycin, a K^+ ionophore, to experimentally quantify the differences in modulation of chromatin packing by monovalent versus divalent cations. Future experiments could involve more precisely tracking changes in ion concentrations after cation chelation using mass spectrometry-based techniques (i.e., ICP-MS), conducting Chrom-STEM experiments before and after modulation of the electrolyte environment to determine corresponding changes in chromatin packing with higher resolution, and measuring consequent changes in transcription after K^+ and Mg^{2+} chelation using RNA-seq.

As there is a fundamental limit to the system size we would feasibly be able to study with molecular-level MT, we could investigate the effects of the physicochemical environment on larger-scale chromatin packing due to the by developing a multi-scale MT approach. First, we could determine the total charge of multiple nucleosomes under different environmental conditions using our current chromatin-level MT approach with

molecular-level coarse-graining. Next, we could reduce the detail of our system by coarse-graining our input chains at the level of multiple nucleosomes. For example, we could generate different conformation sets of SR-EV polymers with different folding parameters and determine which α is the most stable (e.g., has the lowest system free energy) under certain environmental conditions. The size, packing scaling, and average density of SR-EV domains will change with varying α [120], making it an interesting parameter to study with our MT approach. We could even explicitly include transcription reactions into our MT representation of the chromatin system and combine this with our crowding-based model of transcription to model how changes in the physicochemical environment directly contribute to changes in gene expression via crowding-mediated effects.

Along with advancing experimental techniques and computational methodologies, our knowledge of the chromatin is constantly evolving. The chromatin system has many degrees of freedom due to the overall complexity of nuclear processes acting on it. Interdisciplinary approaches, involving close collaborations from experts in different fields, are necessary to help disentangle this complexity. Experiments should inform theory and theory should, in turn, inform experiments to help us better understand fundamental rules that control chromatin structure and function in a realistic nuclear environment.

References

[1] D. C. Alexander, T. Corman, M. Mendoza, A. Glass, T. Belity, R. R. Campbell, J. Han, A. A. Keiser, J. Winkler, M. A. Wood, T. Kim, B. A. Garcia, H. Cohen, P. Mews, G. Egervari, and S. L. Berger. Targeting acetyl-coa metabolism attenuates the formation of fear memories through reduced activity-dependent histone acetylation. *bioRxiv*, page 2022.05.22.492937, 2022.

[2] F. Alisafaei, S. Jokhun Doorgesh, G. V. Shivashankar, and B. Shenoy Vivek. Regulation of nuclear architecture, mechanics, and nucleocytoplasmic shuttling of epigenetic factors by cell geometric constraints. *Proceedings of the National Academy of Sciences*, 116(27):13200–13209, 2019.

[3] L. M. Almassalha, G. M. Bauer, J. E. Chandler, S. Gladstein, L. Cherkezyan, Y. Stypula-Cyrus, S. Weinberg, D. Zhang, P. Thusgaard Ruhoff, H. K. Roy, H. Subramanian, N. S. Chandel, I. Szleifer, and V. Backman. Label-free imaging of the native, living cellular nanoarchitecture using partial-wave spectroscopic microscopy. *Proc Natl Acad Sci U S A*, 113(42):E6372–e6381, 2016.

[4] L. M. Almassalha, G. M. Bauer, J. E. Chandler, S. Gladstein, I. Szleifer, H. K. Roy, and V. Backman. The greater genomic landscape: The heterogeneous evolution of cancer. *Cancer Res*, 76(19):5605–5609, 2016.

[5] L. M. Almassalha, G. M. Bauer, W. Wu, L. Cherkezyan, D. Zhang, A. Kendra, S. Gladstein, J. E. Chandler, D. VanDerway, B.-L. L. Seagle, A. Ugolkov, D. D. Billadeau, T. V. O'Halloran, A. P. Mazar, H. K. Roy, I. Szleifer, S. Shahabi, and V. Backman. Macrogenomic engineering via modulation of the scaling of chromatin packing density. *Nature Biomedical Engineering*, 1(11):902–913, 2017.

[6] L. M. Almassalha, A. Tiwari, P. T. Ruhoff, Y. Stypula-Cyrus, L. Cherkezyan, H. Matsuda, M. A. Dela Cruz, J. E. Chandler, C. White, C. Maneval, H. Subramanian, I. Szleifer, H. K. Roy, and V. Backman. The global relationship between chromatin physical topology, fractal structure, and gene expression. *Scientific Reports*, 7(1):41061, 2017.

[7] G. Arya and T. Schlick. A tale of tails: how histone tails mediate chromatin compaction in different salt and linker histone environments. *The journal of physical chemistry. A*, 113(16):4045–4059, 2009.

[8] M. J. Aryee, W. Liu, J. C. Engelmann, P. Nuhn, M. Gurel, M. C. Haffner, D. Esopi, R. A. Irizarry, R. H. Getzenberg, W. G. Nelson, J. Luo, J. Xu, W. B. Isaacs, G. S. Bova, and S. Yegnasubramanian. Dna methylation alterations exhibit intraindividual stability and interindividual heterogeneity in prostate cancer metastases. *Sci Transl Med*, 5(169):169ra10, 2013.

[9] V. Backman and H. K. Roy. Advances in biophotonics detection of field carcinogenesis for colon cancer risk stratification. *Journal of Cancer*, 4(3):251–261, 2013.

[10] N. Badi and J.-F. Lutz. Sequence control in polymer synthesis. *Chemical Society Reviews*, 38(12):3383–3390, 2009.

[11] A. Badrinarayanan, T. B. K. Le, and M. T. Laub. Bacterial chromosome organization and segregation. *Annual Review of Cell and Developmental Biology*, 31(1):171–199, 2015.

[12] M. H. Bailey, C. Tokheim, E. Porta-Pardo, S. Sengupta, D. Bertrand, A. Weerasinghe, A. Colaprico, M. C. Wendl, J. Kim, B. Reardon, P. K. Ng, K. J. Jeong, S. Cao, Z. Wang, J. Gao, Q. Gao, F. Wang, E. M. Liu, L. Mularoni, C. Rubio-Perez, N. Nagarajan, I. Cortés-Ciriano, D. C. Zhou, W. W. Liang, J. M. Hess, V. D. Yellapantula, D. Tamborero, A. Gonzalez-Perez, C. Suphavilai, J. Y. Ko, E. Khurana, P. J. Park, E. M. Van Allen, H. Liang, M. S. Lawrence, A. Godzik, N. Lopez-Bigas, J. Stuart, D. Wheeler, G. Getz, K. Chen, A. J. Lazar, G. B. Mills, R. Karchin, and L. Ding. Comprehensive characterization of cancer driver genes and mutations. *Cell*, 173(2):371–385.e18, 2018.

[13] A. Bancaud, C. Lavelle, S. Huet, and J. Ellenberg. A fractal model for nuclear organization: current evidence and biological implications. *Nucleic Acids Res*, 40(18):8783–92, 2012.

[14] A. J. Bannister and T. Kouzarides. Regulation of chromatin by histone modifications. *Cell Research*, 21(3):381–395, 2011.

[15] A. J. Bannister, R. Schneider, F. A. Myers, A. W. Thorne, C. Crane-Robinson, and T. Kouzarides. Spatial distribution of di- and tri-methyl lysine 36 of histone h3 at active genes. *J Biol Chem*, 280(18):17732–6, 2005.

[16] A. J. Bannister, P. Zegerman, J. F. Partridge, E. A. Miska, J. O. Thomas, R. C. Allshire, and T. Kouzarides. Selective recognition of methylated lysine 9 on histone h3 by the hp1 chromo domain. *Nature*, 410(6824):120–124, 2001.

[17] I. Barbieri and T. Kouzarides. Role of rna modifications in cancer. *Nat Rev Cancer*, 20(6):303–322, 2020.

[18] M. Barbieri, M. Chotalia, J. Fraser, L.-M. Lavitas, J. Dostie, A. Pombo, and M. Nicodemi. Complexity of chromatin folding is captured by the strings and binders switch model. *Proceedings of the National Academy of Sciences*, 109(40):16173–16178, 2012.

[19] G. M. Bauer, Y. Stypula-Cyrus, H. Subramanian, L. Cherkezyan, P. Viswanathan, D. Zhang, R. Iyengar, S. Bagalkar, J. Derbas, T. Graff, S. Gladstein, L. M. Almassalha, J. E. Chandler, H. K. Roy, and V. Backman. The transformation of the nuclear nanoarchitecture in human field carcinogenesis. *Future Sci OA*, 3(3):Fso206, 2017.

[20] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante. Ionic effects on the elasticity of single dna molecules. *Proceedings of the National Academy of Sciences*, 94(12):6185, 1997.

[21] D. A. Beard and T. Schlick. Modeling salt-mediated electrostatics of macromolecules: the discrete surface charge optimization algorithm and its application to the nucleosome. *Biopolymers*, 58(1):106–15, 2001.

[22] J. S. Becker, R. L. McCarthy, S. Sidoli, G. Donahue, K. E. Kaeding, Z. He, S. Lin, B. A. Garcia, and K. S. Zaret. Genomic and proteomic resolution of heterochromatin and its restriction of alternate fate genes. *Mol Cell*, 68(6):1023–1037.e15, 2017.

[23] V. Bedin, R. L. Adam, B. C. de Sá, G. Landman, and K. Metze. Fractal dimension of chromatin is an independent prognostic factor for survival in melanoma. *BMC Cancer*, 10:260, 2010.

[24] J. Beliveau Brian, F. Joyce Eric, N. Apostolopoulos, F. Yilmaz, Y. Fonseka Chamith, B. McCole Ruth, Y. Chang, B. Li Jin, N. Senaratne Tharanga, R. Williams Benjamin, J.-M. Rouillard, and C.-t. Wu. Versatile design and synthesis platform for visualizing genomes with oligopaint fish probes. *Proceedings of the National Academy of Sciences*, 109(52):21301–21306, 2012.

[25] A. S. Belmont and K. Bruce. Visualization of g1 chromosomes: a folded, twisted, supercoiled chromonema model of interphase chromatid structure. *J Cell Biol*,

127(2):287–302, 1994.

[26] A. S. Belmont, S. Dietzel, A. C. Nye, Y. G. Strukov, and T. Tumbar. Large-scale chromatin structure and function. *Curr Opin Cell Biol*, 11(3):307–11, 1999.

[27] A. S. Belmont, J. W. Sedat, and D. A. Agard. A three-dimensional approach to mitotic chromosome structure: evidence for a complex hierarchical organization. *J Cell Biol*, 105(1):77–92, 1987.

[28] H. Beltran, D. Prandi, J. M. Mosquera, M. Benelli, L. Puca, J. Cyrta, C. Marotz, E. Giannopoulou, B. V. S. K. Chakravarthi, S. Varambally, S. A. Tomlins, D. M. Nanus, S. T. Tagawa, E. M. Van Allen, O. Elemento, A. Sboner, L. A. Garraway, M. A. Rubin, and F. Demichelis. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nature Medicine*, 22(3):298–305, 2016.

[29] H. Beltran, D. S. Rickman, K. Park, S. S. Chae, A. Sboner, T. Y. MacDonald, Y. Wang, K. L. Sheikh, S. Terry, S. T. Tagawa, R. Dhir, J. B. Nelson, A. de la Taille, Y. Allory, M. B. Gerstein, S. Perner, K. J. Pienta, A. M. Chinnaiyan, Y. Wang, C. C. Collins, M. E. Gleave, F. Demichelis, D. M. Nanus, and M. A. Rubin. Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. *Cancer Discov*, 1(6):487–95, 2011.

[30] S. Bianco, A. M. Chiariello, C. Annunziatella, A. Esposito, and M. Nicodemi. Predicting chromatin architecture from models of polymer physics. *Chromosome Res*, 25(1):25–34, 2017.

[31] S. Bianco, D. G. Lupiáñez, A. M. Chiariello, C. Annunziatella, K. Kraft, R. Schöpflin, L. Wittler, G. Andrey, M. Vingron, A. Pombo, S. Mundlos, and M. Nicodemi. Polymer physics predicts the effects of structural variants on chromatin architecture. *Nature Genetics*, 50(5):662–667, 2018.

[32] B. Bintu, L. J. Mateo, J. H. Su, N. A. Sinnott-Armstrong, M. Parker, S. Kinrot, K. Yamaya, A. N. Boettiger, and X. Zhuang. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362(6413), 2018.

[33] M. A. Blanco, D. B. Sykes, L. Gu, M. Wu, R. Petroni, R. Karnik, M. Wawer, J. Rico, H. Li, W. D. Jacobus, A. Jambhekar, S. Cheloufi, A. Meissner, K. Hochedlinger, D. T. Scadden, and Y. Shi. Chromatin-state barriers enforce an irreversible mammalian cell fate decision. *Cell Reports*, 37(6):109967, 2021.

[34] V. A. Bloomfield. Dna condensation by multivalent cations. *Biopolymers*, 44 3:269–82, 1997.

[35] V. A. Bloomfield, D. M. Crothers, I. Tinoco, P. A. Killman, J. E. Hearst, D. E. Wemmer, and D. H. Turner. *Nucleic Acids: Structure, Properties, and Functions*. University Science Books, 2000.

[36] B. Brettmann, N. Laugel, N. Hoffmann, P. Pincus, and M. Tirrell. Bridging contributions to polyelectrolyte brush collapse in multivalent salt solutions. *Journal of Polymer Science Part A: Polymer Chemistry*, 54:n/a–n/a, 2015.

[37] B. Brettmann, P. Pincus, and M. Tirrell. Lateral structure formation in polyelectrolyte brushes induced by multivalent ions. *Macromolecules*, 50(3):1225–1235, 2017.

[38] R. A. Burrell, N. McGranahan, J. Bartek, and C. Swanton. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*, 501(7467):338–345, 2013.

[39] R. A. Burrell, N. McGranahan, J. Bartek, and C. Swanton. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*, 501(7467):338–345, 2013.

[40] R. A. Burrell and C. Swanton. Tumour heterogeneity and the evolution of polyclonal drug resistance. *Mol Oncol*, 8(6):1095–111, 2014.

[41] K. M. Callahan, N. N. Casillas-Ituarte, M. Roeselová, H. C. Allen, and D. J. Tobias. Solvation of magnesium dication: Molecular dynamics simulation and vibrational spectroscopic study of magnesium chloride in aqueous solutions. *The Journal of Physical Chemistry A*, 114(15):5141–5148, 2010.

[42] K. Cao and A. Shilatifard. *Enhancers in Cancer: Genetic and Epigenetic Deregulation*, pages 559–568. Academic Press, Oxford, 2019.

[43] R. Cao, L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R. S. Jones, and Y. Zhang. Role of histone h3 lysine 27 methylation in polycomb-group silencing. *Science*, 298(5595):1039–43, 2002.

[44] Y. Cheng, N. Korolev, and L. Nordenskiöld. Similarities and differences in interaction of k+ and na+ with condensed ordered dna. a molecular dynamics computer simulation study. *Nucleic acids research*, 34(2):686–696, 2006.

[45] L. Cherkezyan, Y. Stypula-Cyrus, H. Subramanian, C. White, M. Dela Cruz, R. K. Wali, M. J. Goldberg, L. K. Bianchi, H. K. Roy, and V. Backman. Nanoscale changes in chromatin organization represent the initial steps of tumorigenesis: a transmission electron microscopy study. *BMC Cancer*, 14(1):189, 2014.

[46] L. Cherkezyan, D. Zhang, H. Subramanian, I. Capoglu, A. Taflove, and V. Backman. Review of interferometric spectroscopy of scattered light for the quantification of subdiffractional structure of biomaterials. *J Biomed Opt*, 22(3):30901, 2017.

[47] A. M. Chiariello, S. Bianco, A. Esposito, L. Fiorillo, M. Conte, E. Irani, F. Musella, A. Abraham, A. Prisco, and M. Nicodemi. Physical mechanisms of chromatin spatial organization. *Febs j*, 289(5):1180–1190, 2022.

[48] K. Chiba, J. Yamamoto, Y. Yamaguchi, and H. Handa. Promoter-proximal pausing and its release: molecular mechanisms and physiological functions. *Exp Cell Res*, 316(17):2723–30, 2010.

[49] K.-H. Chow, R. E. Factor, and K. S. Ullman. The nuclear envelope environment and its cancer connections. *Nature Reviews Cancer*, 12(3):196–209, 2012.

[50] E. J. Clowney, M. A. LeGros, C. P. Mosley, F. G. Clowney, E. C. Markenskoff-Papadimitriou, M. Myllys, G. Barnea, C. A. Larabell, and S. Lomvardas. Nuclear aggregation of olfactory receptor genes governs their monogenic expression. *Cell*, 151(4):724–737, 2012.

[51] S. Corless and N. Gilbert. Effects of dna supercoiling on chromatin architecture. *Biophysical Reviews*, 8(3):245–258, 2016.

[52] P. Cramer. Organization and regulation of gene transcription. *Nature*, 573(7772):45–54, 2019.

[53] M. Cremer, V. J. Schmid, F. Kraus, Y. Markaki, I. Hellmann, A. Maiser, H. Leonhardt, S. John, J. Stamatoyannopoulos, and T. Cremer. Initial high-resolution microscopic mapping of active and inactive regulatory sequences proves non-random 3d arrangements in chromatin domain clusters. *Epigenetics & Chromatin*, 10(1):39, 2017.

[54] T. Cremer and C. Cremer. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*, 2(4):292–301, 2001.

[55] T. Cremer, M. Cremer, B. Hübner, A. Silahtaroglu, M. Hendzel, C. Lanctôt, H. Strickfaden, and C. Cremer. The interchromatin compartment participates in the structural and functional organization of the cell nucleus. *Bioessays*, 42(2):e1900132,

2020.

[56] W. Criscione Steven, M. De Cecco, B. Siranosian, Y. Zhang, A. Kreiling Jill, M. Sedivy John, and N. Neretti. Reorganization of chromosome architecture in replicative cellular senescence. *Science Advances*, 2(2):e1500882, 2016.

[57] Z. Cseresnyes, U. Schwarz, and C. M. Green. Analysis of replication factories in human cells by super-resolution light microscopy. *BMC Cell Biology*, 10(1):88, 2009.

[58] A. R. Cutter DiPiazza, N. Taneja, J. Dhakshnamoorthy, D. Wheeler, S. Holla, and S. I. S. Grewal. Spreading and epigenetic inheritance of heterochromatin require a critical density of histone h3 lysine 9 tri-methylation. *Proc Natl Acad Sci U S A*, 118(22), 2021.

[59] D. Damania, H. K. Roy, D. Kunte, J. A. Hurteau, H. Subramanian, L. Cherkezyan, N. Krosnjar, M. Shah, and V. Backman. Insights into the field carcinogenesis of ovarian cancer based on the nanocytology of endocervical and endometrial epithelial cells. *Int J Cancer*, 133(5):1143–52, 2013.

[60] R. Das, T. T. Mills, L. W. Kwok, G. S. Maskel, I. S. Millett, S. Doniach, K. D. Finkelstein, D. Herschlag, and L. Pollack. Counterion distribution around dna probed by solution x-ray scattering. *Physical Review Letters*, 90(18):188103, 2003.

[61] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 å resolution. *J Mol Biol*, 319(5):1097–113, 2002.

[62] I. F. Davidson, B. Bauer, D. Goetz, W. Tang, G. Wutz, and J.-M. Peters. Dna loop extrusion by human cohesin. *Science*, 366(6471):1338, 2019.

[63] P. G. de Gennes and T. A. Witten. Scaling concepts in polymer physics. *Physics Today*, 33(6):51–54, 1980.

[64] M. Di Pierro, A. Potoyan Davit, G. Wolynes Peter, and N. Onuchic José. Anomalous diffusion, spatial coherence, and viscoelasticity from the energy landscape of human chromosomes. *Proceedings of the National Academy of Sciences*, 115(30):7753–7758, 2018.

[65] M. Di Pierro, B. Zhang, L. Aiden Erez, G. Wolynes Peter, and N. Onuchic José. Transferable model for chromosome architecture. *Proceedings of the National Academy of Sciences*, 113(43):12168–12173, 2016.

[66] V. Dileep, F. Ay, J. Sima, D. L. Vera, W. S. Noble, and D. M. Gilbert. Topologically associating domains and their long-range contacts are established during early g1 coincident with the establishment of the replication-timing program. *Genome research*, 25(8):1104–1113, 2015.

[67] A. M. D'Ippolito, I. C. McDowell, A. Barrera, L. K. Hong, S. M. Leichter, L. C. Bartelt, C. M. Vockley, W. H. Majoros, A. Safi, L. Song, C. A. Gersbach, G. E. Crawford, and T. E. Reddy. Pre-established chromatin interactions mediate the genomic response to glucocorticoids. *Cell Syst*, 7(2):146–160.e7, 2018.

[68] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–380, 2012.

[69] S. Donnini, F. Tegeler, G. Groenhof, and H. Grubmüller. Constant ph molecular dynamics in explicit solvent with λ-dynamics. *Journal of Chemical Theory and Computation*, 7(6):1962–1978, 2011.

[70] E. S. Doğan and C. Liu. Three-dimensional chromatin packing and positioning of plant genomes. *Nature Plants*, 4(8):521–529, 2018.

[71] I. Dunham, A. Kundaje, S. F. Aldred, P. J. Collins, C. A. Davis, F. Doyle, C. B. Epstein, S. Frietze, J. Harrow, R. Kaul, J. Khatun, B. R. Lajoie, S. G. Landt, B.-K. Lee, F. Pauli, K. R. Rosenbloom, P. Sabo, A. Safi, A. Sanyal, N. Shoresh, J. M. Simon, L. Song, N. D. Trinklein, R. C. Altshuler, E. Birney, J. B. Brown, C. Cheng, S. Djebali, X. Dong, I. Dunham, J. Ernst, T. S. Furey, M. Gerstein, B. Giardine, M. Greven, R. C. Hardison, R. S. Harris, J. Herrero, M. M. Hoffman, S. Iyer, M. Kellis, J. Khatun, P. Kheradpour, A. Kundaje, T. Lassmann, Q. Li, X. Lin, G. K. Marinov, A. Merkel, A. Mortazavi, S. C. J. Parker, T. E. Reddy, J. Rozowsky, F. Schlesinger, R. E. Thurman, J. Wang, L. D. Ward, T. W. Whitfield, S. P. Wilder, W. Wu, H. S. Xi, K. Y. Yip, J. Zhuang, B. E. Bernstein, E. Birney, I. Dunham, E. D. Green, C. Gunter, M. Snyder, M. J. Pazin, R. F. Lowdon, L. A. L. Dillon, L. B. Adams, C. J. Kelly, J. Zhang, J. R. Wexler, E. D. Green, P. J. Good, E. A. Feingold, B. E. Bernstein, E. Birney, G. E. Crawford, J. Dekker, L. Elnitski, P. J. Farnham, M. Gerstein, M. C. Giddings, T. R. Gingeras, E. D. Green, R. Guigó, R. C. Hardison, T. J. Hubbard, M. Kellis, W. J. Kent, J. D. Lieb, E. H. Margulies, R. M. Myers, M. Snyder, J. A. Stamatoyannopoulos, S. A. Tenenbaum, et al. An integrated encyclopedia of dna elements in the human genome. *Nature*, 489(7414):57–74, 2012.

[72] N. C. Durand, M. S. Shamim, I. Machol, S. S. P. Rao, M. H. Huntley, E. S. Lander, and E. L. Aiden. Juicer provides a one-click system for analyzing loop-resolution hi-c experiments. *Cell Systems*, 3(1):95–98, 2016.

[73] H. Easwaran, H. C. Tsai, and S. B. Baylin. Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. *Mol Cell*, 54(5):716–27, 2014.

[74] A. Eid, A. Eshein, Y. Li, R. Virk, D. Van Derway, D. Zhang, A. Taflove, and V. Backman. Characterizing chromatin packing scaling in whole nuclei using interferometric microscopy. *Optics Letters*, 45(17):4810–4813, 2020.

[75] M. Eltsov, M. MacLellan Kirsty, K. Maeshima, S. Frangakis Achilleas, and J. Dubochet. Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. *Proceedings of the National Academy of Sciences*, 105(50):19732–19737, 2008.

[76] M. Engelhardt. Condensation of chromatin in situ by cation-dependent charge shielding and aggregation. *Biochem Biophys Res Commun*, 324(4):1210–4, 2004.

[77] M. Falk, Y. Feodorova, N. Naumova, M. Imakaev, B. R. Lajoie, H. Leonhardt, B. Joffe, J. Dekker, G. Fudenberg, I. Solovei, and L. A. Mirny. Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature*, 570(7761):395–399, 2019.

[78] R. Farina, N. Laugel, P. Pincus, and M. Tirrell. Brushes of strong polyelectrolytes in mixed mono- and tri-valent ionic media at fixed total ionic strengths. *Soft Matter*, 9(44):10458–10472, 2013.

[79] A. Felipe, R. Vicente, N. Villalonga, M. Roura-Ferrer, R. Martínez-Mármol, L. Solé, J. C. Ferreres, and E. Condom. Potassium channels: New targets in cancer therapy. *Cancer Detection and Prevention*, 30(4):375–385, 2006.

[80] A. C. Ferguson-Smith. Genomic imprinting: the emergence of an epigenetic paradigm. *Nat Rev Genet*, 12(8):565–75, 2011.

[81] J. T. Finch and A. Klug. Solenoidal model for superstructure in chromatin. *Proc Natl Acad Sci U S A*, 73(6):1897–901, 1976.

[82] E. H. Finn, G. Pegoraro, H. B. Brandão, A. L. Valton, M. E. Oomen, J. Dekker, L. Mirny, and T. Misteli. Extensive heterogeneity and intrinsic variation in spatial genome organization. *Cell*, 176(6):1502–1515.e10, 2019.

[83] W. A. Flavahan, Y. Drier, B. B. Liau, S. M. Gillespie, A. S. Venteicher, A. O. Stemmer-Rachamimov, M. L. Suvà, and B. E. Bernstein. Insulator dysfunction and oncogene activation in idh mutant gliomas. *Nature*, 529(7584):110–114, 2016.

[84] W. A. Flavahan, E. Gaskell, and B. E. Bernstein. Epigenetic plasticity and the hallmarks of cancer. *Science*, 357(6348), 2017.

[85] Y. W. Fong, C. Cattoglio, T. Yamaguchi, and R. Tjian. Transcriptional regulation by coactivators in embryonic stem cells. *Trends Cell Biol*, 22(6):292–8, 2012.

[86] C. Francastel, D. Schübeler, D. I. K. Martin, and M. Groudine. Nuclear compartmentalization and gene activity. *Nature Reviews Molecular Cell Biology*, 1(2):137–143, 2000.

[87] T. B. Franklin, H. Russig, I. C. Weiss, J. Gräff, N. Linder, A. Michalon, S. Vizi, and I. M. Mansuy. Epigenetic transmission of the impact of early stress across generations. *Biol Psychiatry*, 68(5):408–15, 2010.

[88] G. S. Freeman, D. M. Hinckley, J. P. Lequieu, J. K. Whitmer, and J. J. de Pablo. Coarse-grained modeling of dna curvature. *The Journal of Chemical Physics*, 141(16):165103, 2014.

[89] R. V. Frolov and S. Singh. Celecoxib and ion channels: A story of unexpected discoveries. *European Journal of Pharmacology*, 730:61–71, 2014.

[90] G. Fudenberg, N. Abdennur, M. Imakaev, A. Goloborodko, and L. A. Mirny. Emerging evidence of chromosome folding by loop extrusion. *Cold Spring Harbor symposia on quantitative biology*, 82:45–55, 2017.

[91] E. Fussner, M. Strauss, U. Djuric, R. Li, K. Ahmed, M. Hart, J. Ellis, and D. P. Bazett-Jones. Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres. *EMBO Rep*, 13(11):992–6, 2012.

[92] M. Gabriele, B. Brandão Hugo, S. Grosse-Holz, A. Jha, M. Dailey Gina, C. Cattoglio, S. Hsieh Tsung-Han, L. Mirny, C. Zechner, and S. Hansen Anders. Dynamics of ctcf- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science*, 376(6592):496–501, 2022.

[93] M. Gebala, S. Bonilla, N. Bisaria, and D. Herschlag. Does cation size affect occupancy and electrostatic screening of the nucleic acid ion atmosphere? *J Am Chem Soc*, 138(34):10925–34, 2016.

[94] M. Gebala, G. M. Giambaşu, J. Lipfert, N. Bisaria, S. Bonilla, G. Li, D. M. York, and D. Herschlag. Cation–anion interactions within the nucleic acid ion atmosphere revealed by ion counting. *Journal of the American Chemical Society*, 137(46):14705–14715, 2015.

[95] M. Gebala, S. L. Johnson, G. J. Narlikar, and D. Herschlag. Ion counting demonstrates a high electrostatic field generated by the nucleosome. *Elife*, 8, 2019.

[96] B. A. Gibson, L. K. Doolittle, M. W. G. Schneider, L. E. Jensen, N. Gamarra, L. Henry, D. W. Gerlich, S. Redding, and M. K. Rosen. Organization of chromatin by intrinsic and regulated phase separation. *Cell*, 179(2):470–484.e21, 2019.

[97] S. Gladstein, L. M. Almassalha, L. Cherkezyan, J. E. Chandler, A. Eshein, A. Eid, D. Zhang, W. Wu, G. M. Bauer, A. D. Stephens, S. Morochnik, H. Subramanian, J. F. Marko, G. A. Ameer, I. Szleifer, and V. Backman. Multimodal interference-based imaging of nanoscale structure and macromolecular motion uncovers uv induced cellular paroxysm. *Nature Communications*, 10(1):1652, 2019.

[98] S. Gladstein, D. Damania, L. M. Almassalha, L. T. Smith, V. Gupta, H. Subramanian, D. K. Rex, H. K. Roy, and V. Backman. Correlating colorectal cancer risk with field carcinogenesis progression using partial wave spectroscopic microscopy. *Cancer Med*, 7(5):2109–2120, 2018.

[99] J. C. Goldstein, N. J. Waterhouse, P. Juin, G. I. Evan, and D. R. Green. The coordinate release of cytochrome c during apoptosis is rapid, complete and kinetically invariant. *Nat Cell Biol*, 2(3):156–62, 2000.

[100] E. Gonzalez Solveyra, R. J. Nap, K. Huang, and I. Szleifer. Theoretical modeling of chemical equilibrium in weak polyelectrolyte layers on curved nanosystems. *Polymers*, 12(10), 2020.

[101] T. J. Gould, K. Tóth, N. Mücke, J. Langowski, A. S. Hakusui, A. L. Olins, and D. E. Olins. Defining the epichromatin epitope. *Nucleus*, 8(6):625–640, 2017.

[102] A. E. Granada, A. Jiménez, J. Stewart-Ornstein, N. Blüthgen, S. Reber, A. Jambhekar, and G. Lahav. The effects of proliferation status and cell cycle phase on the responses of single cells to chemotherapy. *Mol Biol Cell*, 31(8):845–857, 2020.

[103] M. V. C. Greenberg and D. Bourc'his. The diverse roles of dna methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology*, 20(10):590–607, 2019.

[104] S. A. Grigoryev, G. Arya, S. Correll, C. L. Woodcock, and T. Schlick. Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *Proceedings of the National Academy of Sciences*, 106(32):13317, 2009.

[105] A. Grosberg, S. K. Nechaev, and E. Shakhnovich. The role of topological constraints in the kinetics of collapse of macromolecules. *Journal de Physique*, 49, 1988.

[106] R. D. Grubbs. Intracellular magnesium and magnesium buffering. *Biometals*, 15(3):251–259, 2002.

[107] V. Guptha and P.-Y. Hsiao. Polyelectrolyte brushes in monovalent and multivalent salt solutions. *Polymer*, 55, 2014.

[108] J. D. Halverson, J. Smrek, K. Kremer, and A. Y. Grosberg. From a melt of rings to chromosome territories: the role of topological constraints in genome folding. *Rep Prog Phys*, 77(2):022601, 2014.

[109] D. Hanahan and R. Weinberg. Hallmarks of cancer: The next generation. *Cell*, 144(5):646–674, 2011.

[110] D. Hanahan and R. A. Weinberg. The hallmarks of cancer. *Cell*, 100(1):57–70, 2000.

[111] J. C. Hansen, M. Connolly, C. J. McDonald, A. Pan, A. Pryamkova, K. Ray, E. Seidel, S. Tamura, R. Rogge, and K. Maeshima. The 10-nm chromatin fiber and its relationship to interphase chromosome organization. *Biochemical Society transactions*, 46(1):67–76, 2018.

[112] A. H. Hassan, P. Prochasson, K. E. Neely, S. C. Galasinski, M. Chandy, M. J. Carrozza, and J. L. Workman. Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell*, 111(3):369–79, 2002.

[113] S. Hetzel, A. L. Mattei, H. Kretzmer, C. Qu, X. Chen, Y. Fan, G. Wu, K. G. Roberts, S. Luger, M. Litzow, J. Rowe, E. Paietta, W. Stock, E. R. Mardis, R. K. Wilson, J. R. Downing, C. G. Mullighan, and A. Meissner. Acute lymphoblastic leukemia displays a distinct highly methylated genome. *Nature Cancer*, 2022.

[114] D. M. Hinckley and J. J. de Pablo. Coarse-grained ions for nucleic acid modeling. *Journal of Chemical Theory and Computation*, 11(11):5436–5446, 2015.

[115] A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker, and C. S. Woodward. Sundials: Suite of nonlinear and differential/algebraic equation solvers. *ACM Trans. Math. Softw.*, 31(3):363–396, 2005.

[116] D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, and P. A. Sharp. A phase separation model for transcriptional control. *Cell*, 169(1):13–23, 2017.

[117] S. Horvath and K. Raj. Dna methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics*, 19(6):371–384, 2018.

[118] V. E. Hoskins, K. Smith, and K. L. Reddy. The shifting shape of genomes: dynamics of heterochromatin interactions at the nuclear lamina. *Current Opinion in Genetics & Development*, 67:163–173, 2021.

[119] T. S. Hsieh, C. Cattoglio, E. Slobodyanyuk, A. S. Hansen, O. J. Rando, R. Tjian, and X. Darzacq. Resolving the 3d landscape of transcription-linked mammalian chromatin folding. *Mol Cell*, 78(3):539–553.e8, 2020.

[120] K. Huang, Y. Li, A. R. Shim, R. K. A. Virk, V. Agrawal, A. Eshein, R. J. Nap, L. M. Almassalha, V. Backman, and I. Szleifer. Physical and data structure of 3d genome. *Science Advances*, 6(2):eaay4055, 2020.

[121] S. Inoue, S. Sugiyama, A. A. Travers, and T. Ohyama. Self-assembly of double-stranded dna molecules at nanomolar concentrations. *Biochemistry*, 46(1):164–71, 2007.

[122] R. Israëls, F. A. M. Leermakers, and G. J. Fleer. On the theory of grafted weak polyacids. *Macromolecules*, 27(11):3087–3093, 1994.

[123] R. Israëls, F. A. M. Leermakers, G. J. Fleer, and E. B. Zhulina. Charged polymeric brushes: Structure and scaling relations. *Macromolecules*, 27(12):3249–3261, 1994.

[124] N. E. Jackson, B. K. Brettmann, V. Vishwanath, M. Tirrell, and J. J. de Pablo. Comparing solvophobic and multivalent induced collapse in polyelectrolyte brushes. *ACS Macro Letters*, 6(2):155–160, 2017.

[125] S. Janke. Fractals in molecular biophysics. t. gregory dewey. *The Quarterly Review of Biology*, 74(4):512–512, 1999.

[126] G. C. Jayson, E. C. Kohn, H. C. Kitchener, and J. A. Ledermann. Ovarian cancer. *The Lancet*, 384(9951):1376–1388, 2014.

[127] R. Jessel, S. Haertel, C. Socaciu, S. Tykhonova, and H. A. Diehl. Kinetics of apoptotic markers in exogeneously induced apoptosis of el4 cells. *J Cell Mol Med*, 6(1):82–92, 2002.

[128] D. S. Johnson, A. Mortazavi, R. M. Myers, and B. Wold. Genome-wide mapping of in vivo protein-dna interactions. *Science*, 316(5830):1497–502, 2007.

[129] S. E. Johnstone, A. Reyes, Y. Qi, C. Adriaens, E. Hegazi, K. Pelka, J. H. Chen, L. S. Zou, Y. Drier, V. Hecht, N. Shoresh, M. K. Selig, C. A. Lareau, S. Iyer, S. C. Nguyen, E. F. Joyce, N. Hacohen, R. A. Irizarry, B. Zhang, M. J. Aryee, and B. E. Bernstein. Large-scale topological changes restrain malignant progression in

colorectal cancer. *Cell*, 182(6):1474–1489.e23, 2020.

[130] R. Jothi, S. Cuddapah, A. Barski, K. Cui, and K. Zhao. Genome-wide identification of in vivo protein-dna binding sites from chip-seq data. *Nucleic acids research*, 36(16):5221–5231, 2008.

[131] Y. Joti, T. Hikima, Y. Nishino, F. Kamada, S. Hihara, H. Takata, T. Ishikawa, and K. Maeshima. Chromosomes without a 30-nm chromatin fiber. *Nucleus*, 3(5):404–10, 2012.

[132] J. Jung, W. Nishima, M. Daniels, G. Bascom, C. Kobayashi, A. Adedoyin, M. Wall, A. Lappala, D. Phillips, W. Fischer, C.-S. Tung, T. Schlick, Y. Sugita, and K. Y. Sanbonmatsu. Scaling molecular dynamics beyond 100,000 processor cores for large-scale biophysical simulations. *Journal of computational chemistry*, 40(21):1919–1930, 2019.

[133] C. Kadoch and G. R. Crabtree. Mammalian swi/snf chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci Adv*, 1(5):e1500447, 2015.

[134] R. S. Kalman, A. Stawarz, D. Nunes, D. Zhang, M. A. Dela Cruz, A. Mohanty, H. Subramanian, V. Backman, and H. K. Roy. Biophotonic detection of high order chromatin alterations in field carcinogenesis predicts risk of future hepatocellular carcinoma: A pilot study. *PLoS One*, 13(5):e0197427, 2018.

[135] C. Kandoth, M. D. McLellan, F. Vandin, K. Ye, B. Niu, C. Lu, M. Xie, Q. Zhang, J. F. McMichael, M. A. Wyczalkowski, M. D. M. Leiserson, C. A. Miller, J. S. Welch, M. J. Walter, M. C. Wendl, T. J. Ley, R. K. Wilson, B. J. Raphael, and L. Ding. Mutational landscape and significance across 12 major cancer types. *Nature*, 502(7471):333–339, 2013.

[136] L. Kane, I. Williamson, I. M. Flyamer, Y. Kumar, R. E. Hill, L. A. Lettice, and W. A. Bickmore. Cohesin is required for long-range enhancer action. *bioRxiv*, page 2021.06.24.449812, 2021.

[137] Y. Kato, S. Ozawa, C. Miyamoto, Y. Maehata, A. Suzuki, T. Maeda, and Y. Baba. Acidic extracellular microenvironment and cancer. *Cancer cell international*, 13(1):89–89, 2013.

[138] N. Khanna, Y. Hu, and A. S. Belmont. Hsp70 transgene directed motion to nuclear speckles facilitates heat shock activation. *Curr Biol*, 24(10):1138–44, 2014.

[139] N. Khanna, Y. Zhang, J. S. Lucas, O. K. Dudko, and C. Murre. Chromosome dynamics near the sol-gel phase transition dictate the timing of remote genomic interactions. *Nature Communications*, 10(1):2771, 2019.

[140] J. A. Kim, S. Lee, D. E. Kim, M. Kim, B. M. Kwon, and D. C. Han. Fisetin, a dietary flavonoid, induces apoptosis of cancer cells by inhibiting hsf1 activity through blocking its binding to the hsp70 promoter. *Carcinogenesis*, 36(6):696–706, 2015.

[141] J. S. Kim, V. Backman, and I. Szleifer. Crowding-induced structural alterations of random-loop chromosome model. *Phys Rev Lett*, 106(16):168102, 2011.

[142] J. S. Kim and I. Szleifer. Crowding-induced formation and structural alteration of nuclear compartments: insights from computer simulations. *Int Rev Cell Mol Biol*, 307:73–108, 2014.

[143] J. S. Kim and A. Yethiraj. Crowding effects on protein association: Effect of interactions between crowding agents. *The Journal of Physical Chemistry B*, 115(2):347–353, 2011.

[144] K. H. Kim and C. W. M. Roberts. Targeting ezh2 in cancer. *Nature Medicine*, 22(2):128–134, 2016.

[145] M. S. Kim, Y. R. Kim, N. J. Yoo, and S. H. Lee. Mutational analysis of dnmt3a gene in acute leukemias and common solid cancers. *Apmis*, 121(2):85–94, 2013.

[146] N. Klein Kyle, A. Zhao Peiyao, X. Lyu, T. Sasaki, A. Bartlett Daniel, M. Singh Amar, I. Tasan, M. Zhang, P. Watts Lotte, S.-i. Hiraga, T. Natsume, X. Zhou, T. Baslan, D. Leung, T. Kanemaki Masato, D. Donaldson Anne, H. Zhao, S. Dalton, G. Corces Victor, and M. Gilbert David. Replication timing maintains the global epigenetic state in human cells. *Science*, 372(6540):371–378, 2021.

[147] A. Kleppe, F. Albregtsen, L. Vlatkovic, M. Pradhan, B. Nielsen, T. S. Hveem, H. A. Askautrud, G. B. Kristensen, A. Nesbakken, J. Trovik, H. Wæhre, I. Tomlinson, N. A. Shepherd, M. Novelli, D. J. Kerr, and H. E. Danielsen. Chromatin organisation and cancer prognosis: a pan-cancer study. *The Lancet Oncology*, 19(3):356–369, 2018.

[148] V. J. A. Konda, L. Cherkezyan, H. Subramanian, K. Wroblewski, D. Damania, V. Becker, M. H. R. Gonzalez, A. Koons, M. Goldberg, M. K. Ferguson, I. Waxman, H. K. Roy, and V. Backman. Nanoscale markers of esophageal field carcinogenesis: potential implications for esophageal cancer screening. *Endoscopy*, 45(12):983–988,

2013.

[149] R. D. Kornberg and Y. Lorch. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell*, 98(3):285–94, 1999.

[150] N. Korolev, D. Luo, A. P. Lyubartsev, and L. Nordenskiöld. A coarse-grained dna model parameterized from atomistic simulations by inverse monte carlo. *Polymers*, 6(6), 2014.

[151] N. Korolev, A. P. Lyubartsev, A. Laaksonen, and L. Nordenskiöld. On the competition between water, sodium ions, and spermine in binding to dna: a molecular dynamics computer simulation study. *Biophys J*, 82(6):2860–75, 2002.

[152] F. Kouzine, A. Gupta, L. Baranello, D. Wojtowicz, K. Ben-Aissa, J. Liu, T. M. Przytycka, and D. Levens. Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nature Structural & Molecular Biology*, 20(3):396–403, 2013.

[153] H. Kretzmer, A. Biran, N. Purroy, C. K. Lemvigh, K. Clement, M. Gruber, H. Gu, L. Rassenti, A. W. Mohammad, C. Lesnick, S. L. Slager, E. Braggio, T. D. Shanafelt, N. E. Kay, S. M. Fernandes, J. R. Brown, L. Wang, S. Li, K. J. Livak, D. S. Neuberg, S. Klages, B. Timmermann, T. J. Kipps, E. Campo, A. Gnirke, C. J. Wu, and A. Meissner. Preneoplastic alterations define cll dna methylome and persist through disease progression and therapy. *Blood Cancer Discovery*, 2(1):54–69, 2021.

[154] E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.-F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, 2001.

[155] A. G. Larson, D. Elnatan, M. M. Keenen, M. J. Trnka, J. B. Johnston, A. L. Burlingame, D. A. Agard, S. Redding, and G. J. Narlikar. Liquid droplet formation by hp1α suggests a role for phase separation in heterochromatin. *Nature*, 547(7662):236–240, 2017.

[156] K. H. A. Lau, C. Ren, S. H. Park, I. Szleifer, and P. B. Messersmith. An experimental-theoretical analysis of protein adsorption on peptidomimetic polymer brushes. *Langmuir*, 28(4):2288–2298, 2012.

[157] D. A. Lawson, K. Kessenbrock, R. T. Davis, N. Pervolarakis, and Z. Werb. Tumour heterogeneity and metastasis at single-cell resolution. *Nature Cell Biology*, 20(12):1349–1360, 2018.

[158] M. C. Lee, F. J. Lopez-Diaz, S. Y. Khan, M. A. Tariq, Y. Dayn, C. J. Vaske, A. J. Radenbaugh, H. J. Kim, B. M. Emerson, and N. Pourmand. Single-cell analyses of transcriptional heterogeneity during drug tolerance transition in cancer cells by rna sequencing. *Proc Natl Acad Sci U S A*, 111(44):E4726–35, 2014.

[159] J. Lequieu, A. Córdoba, J. Moller, and J. J. de Pablo. 1cpn: A coarse-grained multi-scale model of chromatin. *The Journal of Chemical Physics*, 150(21):215102, 2019.

[160] N. A. Levis and D. W. Pfennig. Evaluating 'plasticity-first' evolution in nature: Key criteria and empirical approaches. *Trends in Ecology & Evolution*, 31(7):563–574, 2016.

[161] M. Levo, J. Raimundo, X. Y. Bing, Z. Sisco, P. J. Batut, S. Ryabichko, T. Gregor, and M. S. Levine. Transcriptional coupling of distant regulatory genes in living embryos. *Nature*, 605(7911):754–760, 2022.

[162] E. Li and Y. Zhang. Dna methylation in mammals. *Cold Spring Harb Perspect Biol*, 6(5):a019133, 2014.

[163] G. Li and D. Reinberg. Chromatin higher-order structures and gene regulation. *Curr Opin Genet Dev*, 21(2):175–86, 2011.

[164] L. Li, X. Lyu, C. Hou, N. Takenaka, H. Nguyen, C.-T. Ong, C. Cubeñas-Potts, M. Hu, E. Lei, G. Bosco, Z. Qin, and V. Corces. Widespread rearrangement of 3d chromatin organization underlies polycomb-mediated stress-induced silencing. *Molecular cell*, 58, 2015.

[165] W. Li, G. Wolynes Peter, and S. Takada. Frustration, specific sequence dependence, and nonlinearity in large-amplitude fluctuations of allosteric proteins. *Proceedings*

of the National Academy of Sciences, 108(9):3504–3509, 2011.

[166] Y. Li, V. Agrawal, R. K. A. Virk, E. Roth, W. S. Li, A. Eshein, J. Frederick, K. Huang, L. Almassalha, R. Bleher, M. A. Carignano, I. Szleifer, V. P. Dravid, and V. Backman. Analysis of three-dimensional chromatin packing domains by chromatin scanning transmission electron microscopy (chromstem). *Scientific Reports*, 12(1):12198, 2022.

[167] Y. Li, A. Eshein, R. K. A. Virk, A. Eid, W. Wu, J. Frederick, D. VanDerway, S. Gladstein, K. Huang, A. R. Shim, N. M. Anthony, G. M. Bauer, X. Zhou, V. Agrawal, E. M. Pujadas, S. Jain, G. Esteve, J. E. Chandler, T.-Q. Nguyen, R. Bleher, J. J. de Pablo, I. Szleifer, V. P. Dravid, L. M. Almassalha, and V. Backman. Nanoscale chromatin imaging and analysis platform bridges 4d chromatin organization with molecular function. *Science Advances*, 7(1):eabe4310, 2021.

[168] E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289, 2009.

[169] S. Lindquist. The heat-shock response. *Annual Review of Biochemistry*, 55(1):1151–1191, 1986.

[170] T. M. Liyakat Ali, A. Brunet, P. Collas, and J. Paulsen. Tad cliques predict key features of chromatin organization. *BMC Genomics*, 22(1):499, 2021.

[171] H. F. Lodish, A. Berk, C. Kaiser, M. Krieger, M. P. Scott, A. Bretscher, H. L. Ploegh, and P. T. Matsudaira. *Molecular Cell Biology*. W.H. Freeman, 6 edition, 2008.

[172] B. Luan and A. Aksimentiev. Dna attraction in monovalent and divalent electrolytes. *Journal of the American Chemical Society*, 130(47):15754–15755, 2008.

[173] J. S. Lucas, Y. Zhang, O. K. Dudko, and C. Murre. 3d trajectories adopted by coding and regulatory dna elements: first-passage times for genomic interactions. *Cell*, 158(2):339–352, 2014.

[174] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648):251–260, 1997.

[175] D. G. Lupiáñez, M. Spielmann, and S. Mundlos. Breaking tads: How alterations of chromatin domains result in disease. *Trends Genet*, 32(4):225–237, 2016.

[176] X. Lyu, M. J. Rowley, and V. G. Corces. Architectural proteins and pluripotency factors cooperate to orchestrate the transcriptional response of hescs to temperature stress. *Mol Cell*, 71(6):940–955.e7, 2018.

[177] H. Ma, L.-C. Tu, A. Naseri, Y.-C. Chung, D. Grunwald, S. Zhang, and T. Pederson. Crispr-sirius: Rna scaffolds for signal amplification in genome imaging. *Nature Methods*, 15(11):928–931, 2018.

[178] Q. MacPherson, B. Beltran, and A. J. Spakowitz. Bottom–up modeling of chromatin segregation due to epigenetic modifications. *Proceedings of the National Academy of Sciences*, 115(50):12739, 2018.

[179] K. Maeshima, S. Ide, and M. Babokhov. Dynamic chromatin organization without the 30-nm fiber. *Current Opinion in Cell Biology*, 58:95–104, 2019.

[180] K. Maeshima, S. Ide, K. Hibino, and M. Sasai. Liquid-like behavior of chromatin. *Current Opinion in Genetics & Development*, 37:36–45, 2016.

[181] K. Maeshima, T. Matsuda, Y. Shindo, H. Imamura, S. Tamura, R. Imai, S. Kawakami, R. Nagashima, T. Soga, H. Noji, K. Oka, and T. Nagai. A transient rise in free mg(2+) ions released from atp-mg hydrolysis contributes to mitotic chromosome condensation. *Curr Biol*, 28(3):444–451.e6, 2018.

[182] D. B. Mahat, H. H. Salamanca, F. M. Duarte, C. G. Danko, and J. T. Lis. Mammalian heat shock response and mechanisms underlying its genome-wide transcriptional regulation. *Molecular Cell*, 62(1):63–78, 2016.

[183] C. K. Materese, A. Savelyev, and G. A. Papoian. Counterion atmosphere and hydration patterns near a nucleosome core particle. *Journal of the American Chemical Society*, 131(41):15005–15013, 2009.

[184] H. Matsuda, G. G. Putzel, V. Backman, and I. Szleifer. Macromolecular crowding as a regulator of gene transcription. *Biophys J*, 106(8):1801–10, 2014.

[185] A. L. Mattei, N. Bailly, and A. Meissner. Dna methylation: a historical perspective. *Trends in Genetics*, 38(7):676–707, 2022.

[186] M. A. McBrian, I. S. Behbahan, R. Ferrari, T. Su, T.-W. Huang, K. Li, C. S. Hong, H. R. Christofk, M. Vogelauer, D. B. Seligson, and S. K. Kurdistani. Histone acetylation regulates intracellular ph. *Molecular cell*, 49(2):310–321, 2013.

[187] N. McGranahan, A. J. Furness, R. Rosenthal, S. Ramskov, R. Lyngaa, S. K. Saini, M. Jamal-Hanjani, G. A. Wilson, N. J. Birkbak, C. T. Hiley, T. B. Watkins, S. Shafi, N. Murugaesu, R. Mitter, A. U. Akarca, J. Linares, T. Marafioti, J. Y. Henry, E. M. Van Allen, D. Miao, B. Schilling, D. Schadendorf, L. A. Garraway, V. Makarov, N. A. Rizvi, A. Snyder, M. D. Hellmann, T. Merghoub, J. D. Wolchok, S. A. Shukla, C. J. Wu, K. S. Peggs, T. A. Chan, S. R. Hadrup, S. A. Quezada, and C. Swanton. Clonal neoantigens elicit t cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science*, 351(6280):1463–9, 2016.

[188] N. McGranahan and C. Swanton. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell*, 27(1):15–26, 2015.

[189] T. McPherson, A. Kidane, I. Szleifer, and K. Park. Prevention of protein adsorption by tethered peo layers: Experiments and single chain mean field analysis. *Langmuir*, 14:176–186, 1998.

[190] D. Meluzzi and G. Arya. Recovering ensembles of chromatin conformations from contact probabilities. *Nucleic Acids Res*, 41(1):63–75, 2013.

[191] A. Mielgo, V. A. Torres, K. Clair, S. Barbero, and D. G. Stupack. Paclitaxel promotes a caspase 8-mediated apoptosis through death effector domain association with microtubules. *Oncogene*, 28(40):3551–62, 2009.

[192] A. P. Minton and J. Wilf. Effect of macromolecular crowding upon the structure and function of an enzyme: glyceraldehyde-3-phosphate dehydrogenase. *Biochemistry*, 20(17):4821–4826, 1981.

[193] L. A. Mirny. The fractal globule as a model of chromatin architecture in the cell. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 19(1):37–51, 2011.

[194] E. Miron, R. Oldenkamp, M. Brown Jill, M. S. Pinto David, C. S. Xu, R. Faria Ana, A. Shaban Haitham, D. P. Rhodes James, C. Innocent, S. de Ornellas, F. Hess Harald, V. Buckle, and L. Schermelleh. Chromatin arranges in chains of mesoscale domains with nanoscale functional topography independent of cohesin. *Science Advances*, 6(39):eaba8811, 2020.

[195] T. Misteli. Higher-order genome organization in human disease. *Cold Spring Harbor perspectives in biology*, 2(8):a000794–a000794, 2010.

[196] J. Moller and J. J. de Pablo. Bottom-up meets top-down: The crossroads of multiscale chromatin modeling. *Biophys J*, 118(9):2057–2065, 2020.

[197] M. J. Morelli, R. J. Allen, and P. R. t. Wolde. Effects of macromolecular crowding on genetic networks. *Biophysical journal*, 101(12):2882–2891, 2011.

[198] T. Nagano, Y. Lubling, T. J. Stevens, S. Schoenfelder, E. Yaffe, W. Dean, E. D. Laue, A. Tanay, and P. Fraser. Single-cell hi-c reveals cell-to-cell variability in chromosome structure. *Nature*, 502(7469):59–64, 2013.

[199] X. Nan, H. H. Ng, C. A. Johnson, C. D. Laherty, B. M. Turner, R. N. Eisenman, and A. Bird. Transcriptional repression by the methyl-cpg-binding protein mecp2 involves a histone deacetylase complex. *Nature*, 393(6683):386–9, 1998.

[200] R. J. Nap, S. H. Park, and I. Szleifer. Competitive calcium ion binding to end-tethered weak polyelectrolytes. *Soft Matter*, 14(12):2365–2378, 2018.

[201] R. J. Nap*, M. Tagliazucchi*, E. Gonzalez Solveyra, C.-l. Ren, M. J. Uline, and I. Szleifer. *Modeling of Chemical Equilibria in Polymer and Polyelectrolyte Brushes*, pages 161–221. Wiley, 2017.

[202] C. Naughton, N. Avlonitis, S. Corless, J. G. Prendergast, I. K. Mati, P. P. Eijk, S. L. Cockroft, M. Bradley, B. Ylstra, and N. Gilbert. Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nature Structural & Molecular Biology*, 20(3):387–395, 2013.

[203] N. Naumova, M. Imakaev, G. Fudenberg, Y. Zhan, R. Lajoie Bryan, A. Mirny Leonid, and J. Dekker. Organization of the mitotic chromosome. *Science*, 342(6161):948–953, 2013.

[204] M. V. Neguembor, L. Martin, A. Castells-Garcìa, P. A. Gòmez-Garcìa, C. Vicario, D. Carnevali, J. AlHaj Abed, A. Granados, R. Sebastian-Perez, F. Sottile, J. Solon, C. T. Wu, M. Lakadamyali, and M. P. Cosma. Transcription-mediated supercoiling regulates genome folding and loop formation. *Mol Cell*, 81(15):3065–3081.e12, 2021.

[205] G. Nir, I. Farabella, C. Pérez Estrada, C. G. Ebeling, B. J. Beliveau, H. M. Sasaki, S. D. Lee, S. C. Nguyen, R. B. McCole, S. Chattoraj, J. Erceg, J. AlHaj Abed, N. M. C. Martins, H. Q. Nguyen, M. A. Hannan, S. Russell, N. C. Durand, S. S. P. Rao, J. Y. Kishi, P. Soler-Vila, M. Di Pierro, J. N. Onuchic, S. P. Callahan, J. M. Schreiner, J. A. Stuckey, P. Yin, E. L. Aiden, M. A. Marti-Renom, and C. T. Wu. Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLoS Genet*, 14(12):e1007872, 2018.

[206] E. P. Nora, A. Goloborodko, A. L. Valton, J. H. Gibcus, A. Uebersohn, N. Abdennur, J. Dekker, L. A. Mirny, and B. G. Bruneau. Targeted degradation of ctcf decouples

local insulation of chromosome domains from genomic compartmentalization. *Cell*, 169(5):930–944.e22, 2017.

[207] T. Nozaki, R. Imai, M. Tanbo, R. Nagashima, S. Tamura, T. Tani, Y. Joti, M. Tomita, K. Hibino, M. T. Kanemaki, K. S. Wendt, Y. Okada, T. Nagai, and K. Maeshima. Dynamic organization of chromatin domains revealed by super-resolution live-cell imaging. *Mol Cell*, 67(2):282–293.e7, 2017.

[208] J. Nuebler, G. Fudenberg, M. Imakaev, N. Abdennur, and A. Mirny Leonid. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proceedings of the National Academy of Sciences*, 115(29):E6697–E6706, 2018.

[209] T. Ohyama. New aspects of magnesium function: A key regulator in nucleosome self-assembly, chromatin folding and phase separation. *International journal of molecular sciences*, 20(17):4232, 2019.

[210] D. E. Olins and A. L. Olins. Chromatin history: our view from the bridge. *Nature Reviews Molecular Cell Biology*, 4(10):809–814, 2003.

[211] M. Olvera de la Cruz, L. Belloni, M. Delsanti, J. P. Dalbiez, O. Spalla, and M. Drifford. Precipitation of highly charged polyelectrolyte solutions in the presence of multivalent salts. *The Journal of Chemical Physics*, 103(13):5781–5791, 1995.

[212] D. Ou Horng, S. Phan, J. Deerinck Thomas, A. Thor, H. Ellisman Mark, and C. O'Shea Clodagh. Chromemt: Visualizing 3d chromatin structure and compaction in interphase and mitotic cells. *Science*, 357(6349):eaag0025, 2017.

[213] A. V. Ougolkov, N. D. Bone, M. E. Fernandez-Zapico, N. E. Kay, and D. D. Billadeau. Inhibition of glycogen synthase kinase-3 activity leads to epigenetic silencing of nuclear factor kappab target genes and induction of apoptosis in chronic lymphocytic leukemia b cells. *Blood*, 110(2):735–42, 2007.

[214] A. L. Paek, J. C. Liu, A. Loewer, W. C. Forrester, and G. Lahav. Cell-to-cell variation in p53 dynamics leads to fractional killing. *Cell*, 165(3):631–42, 2016.

[215] K. C. Palozola, J. Lerner, and K. S. Zaret. A changing paradigm of transcriptional memory propagation through mitosis. *Nature Reviews Molecular Cell Biology*, 20(1):55–64, 2019.

[216] C. Palozola Katherine, G. Donahue, H. Liu, R. Grant Gregory, S. Becker Justin, A. Cote, H. Yu, A. Raj, and S. Zaret Kenneth. Mitotic transcription and waves of gene reactivation during mitotic exit. *Science*, 358(6359):119–122, 2017.

[217] S. T. Pan, Z. L. Li, Z. X. He, J. X. Qiu, and S. F. Zhou. Molecular mechanisms for tumour resistance to chemotherapy. *Clin Exp Pharmacol Physiol*, 43(8):723–37, 2016.

[218] P. Panikker, S. J. Xu, H. Zhang, J. Sarthi, M. Beaver, A. Sheth, S. Akhter, and F. Elefant. Restoring tip60 hat/hdac2 balance in the neurodegenerative brain relieves epigenetic transcriptional repression and reinstates cognition. *J Neurosci*, 38(19):4569–4583, 2018.

[219] M. T. Panteva, G. M. Giambaşu, and D. M. York. Force field for mg2+, mn2+, zn2+, and cd2+ ions that have balanced interactions with nucleic acids. *The Journal of Physical Chemistry B*, 119(50):15460–15470, 2015.

[220] L. A. Pardo, C. Contreras-Jurado, M. Zientkowska, F. Alves, and W. Stühmer. Role of voltage-gated potassium channels in cancer. *J Membr Biol*, 205(3):115–24, 2005.

[221] S. Park, R. Nap, and I. Szleifer. Association free energies of metal cations with mesylate and acetate in brine calculated via molecular dynamics simulation. *arXiv*, 2018.

[222] J. Paulsen, M. Sekelja, A. R. Oldenburg, A. Barateau, N. Briand, E. Delbarre, A. Shah, A. L. Sørensen, C. Vigouroux, B. Buendia, and P. Collas. Chrom3d: three-dimensional genome modeling from hi-c and nuclear lamin-genome contacts. *Genome Biology*, 18(1):21, 2017.

[223] S. F. Pedersen and C. Stock. Ion channels and transporters in cancer: pathophysiology, regulation, and clinical potential. *Cancer Res*, 73(6):1658–61, 2013.

[224] C. Pessoa Rodrigues, S. Herman Josip, B. Herquel, K. Valsecchi Claudia Isabelle, T. Stehle, D. Grün, and A. Akhtar. Temporal expression of mof acetyltransferase primes transcription factor networks for erythroid fate. *Science Advances*, 6(21):eaaz4815, 2020.

[225] J. Phillips-Cremins, M. G. Sauria, A. Sanyal, T. Gerasimova, B. Lajoie, J. K. Bell, C.-T. Ong, T. Hookway, C. Guo, Y. Sun, M. Bland, W. Wagstaff, S. Dalton, T. McDevitt, R. Sen, J. Dekker, J. Taylor, and V. Corces. Architectural protein subclasses shape 3d organization of genomes during lineage commitment. *Cell*, 153(6):1281–1295, 2013.

[226] S. Y. Ponomarev, K. M. Thayer, and D. L. Beveridge. Ion motions in molecular dynamics simulations on dna. *Proceedings of the National Academy of Sciences of the United States of America*, 101(41):14771, 2004.

[227] S. Portillo-Ledesma, L. H. Tsao, M. Wagley, M. Lakadamyali, M. P. Cosma, and T. Schlick. Nucleosome clutches are regulated by chromatin internal parameters. *J Mol Biol*, 433(6):166701, 2021.

[228] N. Prevarskaya, R. Skryma, and Y. Shuba. Ion channels in cancer: Are cancer hallmarks oncochannelopathies? *Physiological Reviews*, 98(2):559–621, 2018.

[229] L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, and W. A. Lim. Repurposing crispr as an rna-guided platform for sequence-specific control of gene expression. *Cell*, 152(5):1173–83, 2013.

[230] X. Qiu, K. Andresen, L. W. Kwok, J. S. Lamb, H. Y. Park, and L. Pollack. Inter-dna attraction mediated by divalent counterions. *Physical Review Letters*, 99(3):038104, 2007.

[231] M. Åkerfelt, R. I. Morimoto, and L. Sistonen. Heat shock factors: integrators of cell stress, development and lifespan. *Nature Reviews Molecular Cell Biology*, 11(8):545–555, 2010.

[232] S. P. Rao, M. Huntley, N. Durand, E. Stamenova, I. Bochkov, J. Robinson, A. Sanborn, I. Machol, A. Omer, E. Lander, and E. Aiden. A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7):1665–1680, 2014.

[233] G. Raschellà, G. Melino, and M. Malewicz. New factors in mammalian dna repair—the chromatin connection. *Oncogene*, 36(33):4673–4681, 2017.

[234] P. Rathert, M. Roth, T. Neumann, F. Muerdter, J.-S. Roe, M. Muhar, S. Deswal, S. Cerny-Reiterer, B. Peter, J. Jude, T. Hoffmann, L. M. Boryń, E. Axelsson, N. Schweifer, U. Tontsch-Grunt, L. E. Dow, D. Gianni, M. Pearson, P. Valent, A. Stark, N. Kraut, C. R. Vakoc, and J. Zuber. Transcriptional plasticity promotes primary and acquired resistance to bet inhibition. *Nature*, 525(7570):543–547, 2015.

[235] D. Ravindran Menon, S. Das, C. Krepler, A. Vultur, B. Rinner, S. Schauer, K. Kashofer, K. Wagner, G. Zhang, E. Bonyadi Rad, N. K. Haass, H. P. Soyer, B. Gabrielli, R. Somasundaram, G. Hoefler, M. Herlyn, and H. Schaider. A stress-induced early innate response causes multidrug tolerance in melanoma. *Oncogene*, 34(34):4448–4459, 2015.

[236] J. Ray, P. R. Munn, A. Vihervaara, J. J. Lewis, A. Ozer, C. G. Danko, and J. T. Lis. Chromatin conformation remains stable upon extensive transcriptional changes driven by heat shock. *Proc Natl Acad Sci U S A*, 116(39):19431–19439, 2019.

[237] S. V. Razin and A. A. Gavrilov. Chromatin without the 30-nm fiber: constrained disorder instead of hierarchical folding. *Epigenetics*, 9(5):653–7, 2014.

[238] S. Rea, F. Eisenhaber, D. O'Carroll, B. D. Strahl, Z.-W. Sun, M. Schmid, S. Opravil, K. Mechtler, C. P. Ponting, C. D. Allis, and T. Jenuwein. Regulation of chromatin structure by site-specific histone h3 methyltransferases. *Nature*, 406(6796):593–599, 2000.

[239] C.-l. Ren, D. Carvajal, K. R. Shull, and I. Szleifer. Streptavidin-biotin binding in the presence of a polymer spacer. a theoretical description. *Langmuir*, 25:12283–12292, 2009.

[240] M. A. Ricci, C. Manzo, M. F. García-Parajo, M. Lakadamyali, and M. P. Cosma. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell*, 160(6):1145–1158, 2015.

[241] K. Richter, M. Haslbeck, and J. Buchner. The heat shock response: life on the verge of death. *Mol Cell*, 40(2):253–66, 2010.

[242] T. L. Riss and R. A. Moravec. Use of multiple assay endpoints to investigate the effects of incubation time, dose of toxin, and plating density in cell-based cytotoxicity assays. *Assay Drug Dev Technol*, 2(1):51–62, 2004.

[243] H. E. Rivera, H. E. Aichelman, J. E. Fifer, N. G. Kriefall, D. M. Wuitchik, S. J. S. Wuitchik, and S. W. Davies. A framework for understanding gene expression plasticity and its influence on stress tolerance. *Mol Ecol*, 30(6):1381–1397, 2021.

[244] A. M. Romani. Cellular magnesium homeostasis. *Archives of biochemistry and biophysics*, 512(1):1–23, 2011.

[245] M. Rousseau, J. Fraser, M. A. Ferraiuolo, J. Dostie, and M. Blanchette. Three-dimensional modeling of chromatin structure from interaction frequency data using markov chain monte carlo sampling. *BMC Bioinformatics*, 12(1):414, 2011.

[246] H. K. Roy, C. B. Brendler, H. Subramanian, D. Zhang, C. Maneval, J. Chandler, L. Bowen, K. L. Kaul, B. T. Helfand, C.-H. Wang, M. Quinn, J. Petkewicz, M. Paterakos, and V. Backman. Nanocytological field carcinogenesis detection to mitigate overdiagnosis of prostate cancer: A proof of concept study. *PLOS ONE*, 10(2):e0115999, 2015.

[247] H. K. Roy, D. P. Damania, M. DelaCruz, D. P. Kunte, H. Subramanian, S. E. Crawford, A. K. Tiwari, R. K. Wali, and V. Backman. Nano-architectural alterations in mucus layer fecal colonocytes in field carcinogenesis: potential for screening.

Cancer Prev Res (Phila), 6(10):1111–9, 2013.

[248] H. K. Roy, H. Subramanian, D. Damania, T. A. Hensing, W. N. Rom, H. I. Pass, D. Ray, J. D. Rogers, A. Bogojevic, M. Shah, T. Kuzniar, P. Pradhan, and V. Backman. Optical detection of buccal epithelial nanoarchitectural alterations in patients harboring lung cancer: implications for screening. *Cancer Res*, 70(20):7748–54, 2010.

[249] J. Rudnick and G. Gaspari. The aspherity of random walks. *Journal of Physics A: Mathematical and General*, 19(4):L191–L193, 1986.

[250] B. R. Sabari, A. Dall'Agnese, A. Boija, I. A. Klein, E. L. Coffey, K. Shrinivas, B. J. Abraham, N. M. Hannett, A. V. Zamudio, J. C. Manteiga, C. H. Li, Y. E. Guo, D. S. Day, J. Schuijers, E. Vasile, S. Malik, D. Hnisz, T. I. Lee, I. I. Cisse, R. G. Roeder, P. A. Sharp, A. K. Chakraborty, and R. A. Young. Coactivator condensation at super-enhancers links phase separation and gene control. *Science*, 361(6400):eaar3958, 2018.

[251] A. L. Sanborn, S. S. P. Rao, S.-C. Huang, N. C. Durand, M. H. Huntley, A. I. Jewett, I. D. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, K. P. Geeting, A. Gnirke, A. Melnikov, D. McKenna, E. K. Stamenova, E. S. Lander, and E. L. Aiden. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences*, 112(47):E6456, 2015.

[252] H. Sandholtz Sarah, Q. MacPherson, and J. Spakowitz Andrew. Physical modeling of the heritability and maintenance of epigenetic modifications. *Proceedings of the National Academy of Sciences*, 117(34):20423–20429, 2020.

[253] S. Sanulli, M. J. Trnka, V. Dharmarajan, R. W. Tibble, B. D. Pascal, A. L. Burlingame, P. R. Griffin, J. D. Gross, and G. J. Narlikar. Hp1 reshapes nucleosome core to promote phase separation of heterochromatin. *Nature*, 575(7782):390–394, 2019.

[254] J. Satulovsky, M. A. Carignano, and I. Szleifer. Kinetic and thermodynamic control of protein adsorption. *Proc. Natl. Acad. Sci. U.S.A.*, 97(15):9037–9041, 2000.

[255] A. Savelyev and G. A. Papoian. Chemically accurate coarse graining of double-stranded dna. *Proceedings of the National Academy of Sciences*, 107(47):20340, 2010.

[256] H. Schiessel and P. Pincus. Counterion-condensation-induced collapse of highly charged polyelectrolytes. *Macromolecules*, 31(22):7953–7959, 1998.

[257] A. D. Schmitt, M. Hu, and B. Ren. Genome-wide mapping and analysis of chromosome architecture. *Nature Reviews Molecular Cell Biology*, 17(12):743–755, 2016.

[258] R. Schneider, A. J. Bannister, F. A. Myers, A. W. Thorne, C. Crane-Robinson, and T. Kouzarides. Histone h3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat Cell Biol*, 6(1):73–7, 2004.

[259] P. M. Schwarz, A. Felthauser, T. M. Fletcher, and J. C. Hansen. Reversible oligonucleosome self-association: dependence on divalent cations and core histone tail domains. *Biochemistry*, 35(13):4009–15, 1996.

[260] P. M. Schwarz and J. C. Hansen. Formation and stability of higher order chromatin structures. contributions of the histone octamer. *Journal of Biological Chemistry*, 269(23):16284–16289, 1994.

[261] W. Schwarzer, N. Abdennur, A. Goloborodko, A. Pekowska, G. Fudenberg, Y. Loe-Mie, N. A. Fonseca, W. Huber, C. H. Haering, L. Mirny, and F. Spitz. Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, 551(7678):51–56, 2017.

[262] T. Sexton, E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc, M. Hoichman, H. Parrinello, A. Tanay, and G. Cavalli. Three-dimensional folding and functional organization principles of the drosophila genome. *Cell*, 148(3):458–72, 2012.

[263] S. M. Shaffer, M. C. Dunagin, S. R. Torborg, E. A. Torre, B. Emert, C. Krepler, M. Beqiri, K. Sproesser, P. A. Brafford, M. Xiao, E. Eggan, I. N. Anastopoulos, C. A. Vargas-Garcia, A. Singh, K. L. Nathanson, M. Herlyn, and A. Raj. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature*, 546(7658):431–435, 2017.

[264] S. V. Sharma, D. Y. Lee, B. Li, M. P. Quinlan, F. Takahashi, S. Maheswaran, U. McDermott, N. Azizian, L. Zou, M. A. Fischbach, K.-K. Wong, K. Brandstetter, B. Wittner, S. Ramaswamy, M. Classon, and J. Settleman. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell*, 141(1):69–80, 2010.

[265] H. Shin, Y. Shi, C. Dai, H. Tjong, K. Gong, F. Alber, and X. J. Zhou. Topdom: an efficient and deterministic method for identifying topological domains in genomes. *Nucleic Acids Research*, 44(7):e70–e70, 2016.

[266] R. K. O. Sigel and H. Sigel. A stability concept for metal ion coordination to single-stranded nucleic acids and affinities of individual sites. *Accounts of Chemical Research*, 43(7):974–984, 2010.

[267] F. J. Solis and M. Olvera de la Cruz. Collapse of flexible polyelectrolytes in multivalent salt solutions. *The Journal of Chemical Physics*, 112(4):2030–2035, 2000.

[268] A. A. Sousa, A. A. Azari, G. Zhang, and R. D. Leapman. Dual-axis electron tomography of biological specimens: Extending the limits of specimen thickness with bright-field stem imaging. *J Struct Biol*, 174(1):107–14, 2011.

[269] A. R. Statz, J. Kuang, C. Ren, A. E. Barron, I. Szleifer, and P. B. Messersmith. Experimental and theoretical investigation of chain length and surface coverage on fouling of surface grafted polypeptoids. *Biointerphases*, 4(2):FA22–FA32, 2009.

[270] S. Stern, T. Dror, E. Stolovicki, N. Brenner, and E. Braun. Genome-wide transcriptional plasticity underlies cellular adaptation to novel challenge. *Molecular systems biology*, 3:106–106, 2007.

[271] T. Sun, V. Minhas, N. Korolev, A. Mirzoev, A. P. Lyubartsev, and L. Nordenskiöld. Bottom-up coarse-grained modeling of dna. *Frontiers in molecular biosciences*, 8:645527–645527, 2021.

[272] C. Swanton. Intratumor heterogeneity: evolution through space and time. *Cancer Res*, 72(19):4875–82, 2012.

[273] C. Swanton. Intratumor heterogeneity: evolution through space and time. *Cancer Res*, 72(19):4875–82, 2012.

[274] Q. Szabo, F. Bantignies, and G. Cavalli. Principles of genome folding into topologically associating domains. *Sci Adv*, 5(4):eaaw1668, 2019.

[275] Q. Szabo, D. Jost, J. M. Chang, D. I. Cattoni, G. L. Papadopoulos, B. Bonev, T. Sexton, J. Gurgo, C. Jacquier, M. Nollmann, F. Bantignies, and G. Cavalli. Tads are 3d structural units of higher-order chromosome organization in drosophila. *Sci Adv*, 4(2):eaar8082, 2018.

[276] I. Szleifer. A new mean-field theory for dilute polymer solutions: Phase diagram, conformational behavior and interfacial properties. *The Journal of Chemical Physics*, 92(11):6940–6952, 1990.

[277] P. C. Taberlay, J. Achinger-Kawecka, A. T. Lun, F. A. Buske, K. Sabir, C. M. Gould, E. Zotenko, S. A. Bert, K. A. Giles, D. C. Bauer, G. K. Smyth, C. Stirzaker, S. I. O'Donoghue, and S. J. Clark. Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. *Genome Res*, 26(6):719–31, 2016.

[278] M. Tagliazucchi, E. J. Calvo, and I. Szleifer. A molecular theory of chemically modified electrodes with self-assembled redox polyelectrolyte thin films: Reversible cyclic voltammetry. *Electrochim. Acta*, 53(23):6740–6752, 2008.

[279] M. Tagliazucchi, Y. Rabin, and I. Szleifer. Ion transport and molecular organization are coupled in polyelectrolyte-modified nanopores. *Journal of the American Chemical Society*, 133(44):17753–17763, 2011.

[280] P. B. Talbert, M. P. Meers, and S. Henikoff. Old cogs, new tricks: the evolution of gene expression in a chromatin context. *Nat Rev Genet*, 20(5):283–297, 2019.

[281] L. Tan, D. Xing, C.-H. Chang, H. Li, and X. S. Xie. Three-dimensional genome structures of single diploid human cells. *Science (New York, N.Y.)*, 361(6405):924–928, 2018.

[282] A. Thill, S. Veerapaneni, B. Simon, M. Wiesner, J. Y. Bottero, and D. Snidaro. Determination of structure of aggregates by confocal scanning laser microscopy. *J Colloid Interface Sci*, 204(2):357–62, 1998.

[283] H. Tjong, W. Li, R. Kalhor, C. Dai, S. Hao, K. Gong, Y. Zhou, H. Li, X. J. Zhou, M. A. Le Gros, C. A. Larabell, L. Chen, and F. Alber. Population-based 3d genome structure analysis reveals driving forces in spatial genome organization. *Proc Natl Acad Sci U S A*, 113(12):E1663–72, 2016.

[284] S. V. Ulianov, E. E. Khrameeva, A. A. Gavrilov, I. M. Flyamer, P. Kos, E. A. Mikhaleva, A. A. Penin, M. D. Logacheva, M. V. Imakaev, A. Chertovich, M. S. Gelfand, Y. Y. Shevelyov, and S. V. Razin. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res*, 26(1):70–84, 2016.

[285] L. van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning*, 9(86):2579–2605, 2008.

[286] B. van Steensel and E. E. M. Furlong. The role of transcription in shaping the spatial organization of the genome. *Nature reviews. Molecular cell biology*, 20(6):327–337, 2019.

[287] S. Venkatesh and J. L. Workman. Histone exchange, chromatin structure and the regulation of transcription. *Nature Reviews Molecular Cell Biology*, 16(3):178–189, 2015.

[288] K. A. Virk Ranya, W. Wu, M. Almassalha Luay, M. Bauer Greta, Y. Li, D. VanDerway, J. Frederick, D. Zhang, A. Eshein, K. Roy Hemant, I. Szleifer, and V. Backman. Disordered chromatin packing regulates phenotypic plasticity. *Science Advances*, 6(2):eaax6232, 2020.

[289] R. K. Wali, N. Momi, M. Dela Cruz, A. H. Calderwood, Y. Stypula-Cyrus, L. Almassalha, A. Chhaparia, C. R. Weber, A. Radosevich, A. K. Tiwari, B. Latif, V. Backman, and H. K. Roy. Higher order chromatin modulator cohesin sa1 is an early biomarker for colon carcinogenesis: Race-specific implications. *Cancer Prev Res (Phila)*, 9(11):844–854, 2016.

[290] E. R. H. Walter, M. A. Fox, D. Parker, and J. A. G. Williams. Enhanced selectivity for mg2+ with a phosphinate-based chelate: Apdap versus aptra. *Dalton Transactions*, 47(6):1879–1887, 2018.

[291] D. Wang, R. J. Nap, I. Lagzi, B. Kowalczyk, S. Han, B. A. Grzybowski, and I. Szleifer. How and why nanoparticle's curvature regulates the apparent pka of the coating ligands. *J Am Chem Soc*, 133(7):2192–7, 2011.

[292] H. Wang, L. Wang, H. Erdjument-Bromage, M. Vidal, P. Tempst, R. S. Jones, and Y. Zhang. Role of histone h2a ubiquitination in polycomb silencing. *Nature*, 431(7010):873–8, 2004.

[293] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, 1953.

[294] M. Weber, I. Hellmann, M. B. Stadler, L. Ramos, S. Pääbo, M. Rebhan, and D. Schübeler. Distribution, silencing potential and evolutionary impact of promoter dna methylation in the human genome. *Nat Genet*, 39(4):457–66, 2007.

[295] C. L. White, R. K. Suto, and K. Luger. Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *Embo j*, 20(18):5207–18, 2001.

[296] I. Williamson, L. Kane, P. S. Devenney, I. M. Flyamer, E. Anderson, F. Kilanowski, R. E. Hill, W. A. Bickmore, and L. A. Lettice. Developmentally regulated shh expression is robust to tad perturbations. *Development*, 146(19), 2019.

[297] D. Winogradoff and A. Aksimentiev. Molecular mechanism of spontaneous nucleosome unraveling. *Journal of molecular biology*, 431(2):323–335, 2019.

[298] J. Wood and S. Helfand. Chromatin structure and transposable elements in organismal aging. *Frontiers in Genetics*, 4, 2013.

[299] T. Wu, J. Genzer, P. Gong, I. Szleifer, P. Vlček, and V. Šubr. *Behavior of surface-anchored poly(acrylic acid) brushes with grafting density gradients on solid substrates*, pages 287–315. Wiley-VCH, Weinheim, Germany, 2004.

[300] W. Xiang, M. J. Roberti, J.-K. Hériché, S. Huet, S. Alexander, and J. Ellenberg. Correlative live and super-resolution imaging reveals the dynamic structure of replication domains. *Journal of Cell Biology*, 217(6):1973–1984, 2018.

[301] E. Yang, E. van Nimwegen, M. Zavolan, N. Rajewsky, M. Schroeder, M. Magnasco, and J. Darnell, J. E. Decay rates of human mrnas: correlation with functional characteristics and sequence attributes. *Genome Res*, 13(8):1863–72, 2003.

[302] R. Yehuda, N. P. Daskalakis, L. M. Bierer, H. N. Bader, T. Klengel, F. Holsboer, and E. B. Binder. Holocaust exposure induced intergenerational effects on ¡em¿fkbp5¡/em¿ methylation. *Biological Psychiatry*, 80(5):372–380, 2016.

[303] J. Yoo and A. Aksimentiev. Improved parametrization of li+, na+, k+, and mg2+ ions for all-atom molecular dynamics simulations of nucleic acid systems. *The Journal of Physical Chemistry Letters*, 3(1):45–50, 2012.

[304] J. Yu, N. E. Jackson, X. Xu, B. K. Brettmann, M. Ruths, J. J. d. Pablo, and M. Tirrell. Multivalent ions induce lateral structural inhomogeneities in polyelectrolyte brushes. *Science Advances*, 3(12):eaao1497, 2017.

[305] A. Zemach, I. E. McDaniel, P. Silva, and D. Zilberman. Genome-wide evolutionary analysis of eukaryotic dna methylation. *Science*, 328(5980):916–9, 2010.

[306] V. W. Zhou, A. Goren, and B. E. Bernstein. Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet*, 12(1):7–18, 2011.

[307] Y. Zhou, D. L. Gerrard, J. Wang, T. Li, Y. Yang, A. J. Fritz, M. Rajendran, X. Fu, G. Stein, R. Schiff, S. Lin, S. Frietze, and V. X. Jin. Temporal dynamic reorganization of 3d chromatin architecture in hormone-induced breast cancer and endocrine resistance. *Nature Communications*, 10(1):1522, 2019.

[308] A. Zidovska, A. Weitz David, and J. Mitchison Timothy. Micron-scale coherence in interphase chromatin dynamics. *Proceedings of the National Academy of Sciences*,

110(39):15555–15560, 2013.

[309] A. Zinchenko, N. V. Berezhnoy, Q. Chen, and L. Nordenskiöld. Compaction of single-molecule megabase-long chromatin under the influence of macromolecular crowding. *Biophys J*, 114(10):2326–2335, 2018.

[310] A. Zinchenko, N. V. Berezhnoy, S. Wang, W. M. Rosencrans, N. Korolev, J. R. C. van der Maarel, and L. Nordenskiöld. Single-molecule compaction of megabase-long chromatin molecules by multivalent cations. *Nucleic Acids Res*, 46(2):635–649, 2018.

[311] A. Zinchenko, Q. Chen, N. V. Berezhnoy, S. Wang, and L. Nordenskiöld. Compaction and self-association of megabase-sized chromatin are induced by anionic protein crowding. *Soft Matter*, 16(18):4366–4372, 2020.