Title

Insights into the structural based mechanism of nucleosome array dynamic and their phase separation using individual-molecule cryo-electron tomography (ET)

Author

Zhang, Meng

Publication Date

2022

Insights into the structural based mechanism of nucleosome array dynamic and their phase separation using individual-molecule cryo-electron tomography (ET)

Meng Zhang

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Applied Science and Technology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Carlos Bustamante, Chair
Professor Evangelina Nogales de la Morena
Professor Mohammad Mofrad

Spring 2022

Insights into the structural based mechanism of nucleosome array dynamic and phase separation using individual-molecule cryo-electron tomography (ET)

Abstract

Insights into the structural based mechanism of nucleosome array dynamic and their phase separation using “individual-molecule” cryo-electron tomography (ET)

by

Meng Zhang

Doctor of Philosophy in Applied Science and Technology

University of California, Berkeley

Professor Carlos Bustamante, Chair

DNA is well protected and organized inside the living cell. The ‘architecture’ of the genetic material not only provides a “hard disk” for the storage of “gene files”, but also zips the meters-long DNA string into a micron size spatial volume that enables the positioning, exporting, and copying of the “files” at a high frequency and accuracy. Therefore, how DNA compacts itself and fits within a confined cellular volume and what mechanisms cells use to regulate the accessibility of the genetic material are fundamental questions that remain at the forefront of scientific research. As a high-resolution imaging tool, electron microscopy (EM) studies of the genetic material enabled the first glimpse of the ultrastructure of chromatin in mid-twentieth century (1970). Due to the limitation of sample preparation and 3D reconstruction techniques, these early studies were confined to a 2-dimensional (2D) characterization of chemically fixed or stained chromatin samples. In the last decades, along with the maturation of the cryo-EM single-particle averaging (SPA) reconstruction approach, it became possible to determine the high-resolution 3-dimensional (3D) structure of medium and large protein molecules in vitro. However, the flexible character and highly dynamic nature of nucleic acid polymers makes it difficult to take advantage of SPA to increase the signal-to-noise ratio of the resulting averaged image. As an alternative 3D structure determination method, cryo-electron tomography (cryo-ET) has been used to specifically solve the structure, at intermediate resolution, of biological samples that present large heterogeneity, from large-scale tissue structures, to subcellular organelles, to dynamic protein assemblies.

Therefore, finding experimental conditions and optimizing cryo-ET 3D reconstruction workflows that improve the resolution and recognition of the target samples can pave the way to a better understanding of the detailed structure of genetic material organization.

For the eukaryotic system, the genetic material is organized into chromosomes, with each of them being composed of long, intertwined chromatin fibers that consists of repeated nucleoprotein units called nucleosomes (Fig. 1.1). This hierarchical, step-wise organization was assumed to be a key step in the regulation of gene expression in response to the stage of the cell cycle and environmental changes. The robust detection of the conformations of fibers is still a difficult task for structural biologists. Similarly, defining the transitions between the different packaging states has remained elusive to biophysicists. In this dissertation, I will first describe our efforts to push the boundaries of cryo-ET reconstruction. In order to achieve higher resolution in ET reconstructions, cryo-ET adopted a similar notion to the ensemble averaging as SPA, an approach called "sub-tomogram averaging". Its workflow avoids the requirement of identifying millions of particles from 2D images by averaging fewer 3D particles from tomograms to obtain higher resolution. However, when investigating samples with large structural dynamics, the conformational heterogeneity still represents a challenge. Ideally, the reconstruction can be done just with the information collected from one particle, if images were clear enough. To explore the feasibility of this ultimate goal, and to better understand what its limiting factors are, I decided to use what I call "individual particle" cryo-ET reconstruction. In the process, I have performed a systematic characterization of a refined workflow that includes sample preparation, data collection, image processing, 3D reconstruction and refinement of mono-nucleosome samples with long flanking DNA arms as a test sample, in order to achieve an improved reconstruction. The reconstruction achieved via this "individual particle" cryo-ET technique indeed provides enough resolution to describe the DNA arm flexibility and 3D dynamics.

To address the question of chromatin fiber organization and dynamics, we adopt a "bottom up" strategy that begins by characterizing a simpler system from *in vitro* reconstituted nucleosome arrays under low salt condition to simulate the open or transcriptionally-active state of chromatin (euchromatin). Since conformational transitions in chromatin depend on how flexible the "joints" that connect the nucleosome units within the array are, mapping out the dynamics of those "joints" turns out to be critical for understanding the intrinsic structural regulatory mechanisms of chromatin fibers. As will be shown here, having established the "individual-particle" cryo-ET methodology, we were able to investigate the effect of different factors (including array length, ionic strength, or the involvement of protein condensation factors such as linker

histone H1) on nucleosome array dynamics. By quantitatively analyzing the various array structures, we extracted useful statistics and identified key parameters that determine array conformation. Resampling of the statistics allowed us to describe and reconstruct a longer nucleosome array fiber structure *in silico*, which provides a possible explanation for some of the controversy surrounding the existence of various types of "30-nm" fiber structures *in vivo*.

Throughout the cell cycle, chromatin can undergo a transition from its open or active state into a highly condensed or transcriptionally silenced state (heterochromatin or mitotic chromatin). In recent years, evidence has been obtained that rather than being governed by a hierarchical packing process, the transition between the lightly packed fibers and the condensed chromatin is mediated via a reversible thermodynamic process that has been termed "liquid-liquid phase separation". The interaction between the intrinsically disordered regions (IDR) of histones and DNA are thought to be mainly responsible for this transition. Though liquid-liquid phase separation has been described in numerous fluorescence microscopy studies of proteins possessing intrinsically disordered regions (IDRs), as well as RNAs, *in vitro* studies specifically addressing the liquid-liquid phase separation that has been proposed for chromatin are much fewer, and questions regarding the detailed physical mechanism and organization of the nucleosome units within "condensates" have remained unanswered. Accordingly, I set up to exploit the capabilities of cryo-ET for the study of the early stage formation of condensates from a homogenous nucleosomal phase. Our success in using high-quality 3D reconstruction to capture the condensates at various stages of their formation and with sufficient resolution to distinguish individual nucleosome features, has enabled us to analyze, for the first time, the nucleosome distribution, organization, and orientation within the condensates. The characterization of both the interior and the overall geometry of the condensates provided us with new insights into the mechanisms of how nascent condensates emerge from the solution. Overall, this advanced "individual-particle" methodology applied to samples at different times during the condensation process, allowed us to perform biophysical measurements to describe the chromatin dynamics during liquid-liquid phase separation. In a broader sense, our approach opens a number of future research directions to address, in a similar fashion, other heterogeneous biological systems.

Dedication

To mom and dad...

Acknowledgments

Firstly, I would like to thank my two co-advisors, Prof. Carlos Bustamante and Dr. Gang (Gary) Ren for their support and mentorship over the years during my PhD studies. They allow me the valuable opportunity to work on a cutting-edge project, which combines the shared interest and expertise from both groups.

Before I started my PhD studies, I was fascinated by the "single-molecule techniques", which for me opened the door of the microscopic world inspection that made me decide to become a joint student between the two groups. From Carlos at UC Berkeley, I learned the general physical principles of single-molecule manipulation, especially using the optical tweezers, which granted me a deeper understanding of the molecules' biophysical property by "touching" them, as well as allowed me to participate in the following exciting nucleosomes unwrapping studies as one of my collaborative side projects. I still remember Carlos's scientific philosophy, drawing from the movie "Blowup- 1966" showing us that "illusion" and "subjectivity" are the worst enemies to science, which later became a constant vigilance for me to think about my experiment results and conclusions. Gary, a brilliant scientist in Lawrence Berkeley National Lab (LBNL), introduced me to the "individual-particle" imaging fields and provided me with a thorough background training to become a structural biologist. His creativity and curiosity inspired me to perform adventurous attempts on unknown subjects. His great passion and energetic attitude to perform scientific studies set up a good example for me. I am very grateful for working with two such great scientists and having them to be my mentors.

During the first year of PhD studies, I faced the competitiveness of taking classes with other UCB students. Fortunately, Omar Herrera team-up with me and we helped each other and progressed together. Thanks to Cesar Diaz who first trained me in biochemistry for a half year when I started to work on my own projects, and later his mastery with nucleosome reconstitution was a crucial factor for the design and execution of several related projects. I would like to express my gratitude to Prof. Eva Nogales, who provided valuable advice and practical solutions to various imaging problems, such as the application of streptavidin crystal grid to capture the nucleosome condensates and the improvements to the methodology. Also, special thanks are due to my LBL colleague Jianfang Liu who is a close friend. We not only talk about science together but also share

the emotions of being abroad.

Working in both groups gave me a wonderful cohort of friends for life. Thanks to my colleagues from both labs in no particular order: Enze Cheng, Francesca Burgos, Alan Shaw, Omar Herrera, Alfredo Flores, John Van Patten, Alex Tong, King Cada, Hossein Amiri, Shannon Yan, Honglu Zhang, Wenxia Lin, Mina Sun, Rohit Satija, Juan Pablo Castillo, Varsha Desai, Lisa Alexander, Luis Ramírez, Yumeng Hao, Errol Watson, Guillermo Chacaltana, Bibiana Onoa, Alfredo Florez, Robert Sosa, Lingli Kong, Dongsheng Lei, Xiaobo Zhai, Changyu Huang, Ledet Getachew, and Natalie Thung.

At the end, I would like to acknowledge my parents, Guiying Li and Tian Zhang, who are always there supporting me despite the long distance. As the only child from them, being separated for such a long time from them made me realize how important and valuable they are in my life. And finally, thanks to you, the reader who downloaded to read this long document.

Contents

List of Figures

Chapter 1
Introduction

"Almost all aspects of life are engineered at the molecular level, and without understanding molecules we can only have a very sketchy understanding of life itself.

— Francis Crick

As pointed out by Francis Crick, biological systems are designed and engineered at the molecular level whose microscopic operation underlie cellular functions and provide a solid foundation for the living state. In my dissertation, I set to understand and describe the structural and dynamical organization of chromatin at the single-molecule level. In this chapter (Chapter 1), I describe the current understanding of chromatin transition and the various models proposed for the organization of chromatin structures. Also, I described different methodologies used to perform single-molecule structural studies, including cryo-electron tomography (cryo-ET) and why I chose this technique as a major approach to address chromatin's structural dynamics. In Chapter 2, I give a brief overview of the factors required to improve the individual particle cryo-ET 3D reconstruction workflow and demonstrate its use to achieve improved results. In Chapter 3, using the reconstituted nucleosome arrays as an example, we successfully applied the workflow to reconstruct various types of nucleosome array models under a series of experimental conditions. Through a systematic characterization of the arrays' dynamics, I propose a conformational regulatory mechanism for the arrays. In Chapter 4, I describe how we extend the capability of individual particle cryo-ET methodology to study a more challenging system, i.e. condensed nucleosomal aggregates, and shed some light on the mechanism chromatin phase transition. In Chapter 5, I summarize the pros and cons of the method, which may pave the way for the establishment of future research on similar heterogeneous systems.

1.1 Structural organization of chromatin

Understanding of how eukaryotic genomes compact DNA materials in a highly organized structure, while maintaining its functionality in orchestrating cellular activity is one of the fundamental problems in cell biology. Genome defines the complete hereditary nucleic acid material possessed by a living species, confined within each of their cells. Despite the fact that the primary linear sequence of the genome is completely decodable under the blooming of the modern high-throughput sequencing techniques, this one-dimensional (1D) information is not sufficient to elucidate the regulatory functions that control the cellular homeostasis and cell division. Indeed, these functions are closely related to the genome secondary and tertiary structural organization, which determine when and where the genetic information becomes accessible in connection to specific cellular events. For example, the two most crucial nuclear activities, transcription and replication, were found not to occur simultaneously everywhere within the cell nucleus (Cook, 1999). Different compartments of discrete foci or territories appearing in varying number (~150) and sizes (half micron to several microns) along the cell cycle (Nakayasu and Berezney, 1989; Nakamura et al., 1986) have been reported for chromatin inside the cell. It is thought that these discrete compartments correspond to transcription and replication "factories". Therefore, "What is the structural organization of the genome material?" and "how does this organization modulate gene activities?" has become the major question to be answered in modern structural biology.

1.1.1 Chromatin primary and secondary structure

The characterization of genome structure started nearly a century ago since the concept of chromatin territories was proposed based on observations done through the optical microscope (Boveri, 1909). The early electron microscopy study of chromatin isolated from eukaryotic nuclei led to the discovery of its linear fiber arrangement (with a 70 nm diameter), in which each of the fiber was composed of numerous tandem thread particles also described as 'beads-on-a-string' (Olins and Olins, 1974). Shortly after that, these threaded particles, named nucleosomes (composed of two copies of histones H2A, H2B, H3, and H4 forming an octamer core) were found involve a DNA segment wrapped 1.67 turns in a left-handed helical fashion around the core to form a nucleoprotein structure with dimensions of 110 x 110 x 57 Å and a helical pitch of 28 Å (Finch et al., 1977). The resulting nucleosome structure is regarded as the first level of hierarchical packing of the genome. Several physical properties of the nucleosome, including the periodicity of DNA endonuclease digestion (Hewish and Burgoyne, 1973), the overabundance of its distribution within nucleus, thus competing with transcription factors for the binding of DNA (Chari et al., 2019), and its

shielding function, protecting eukaryotic DNA from damage (Brambilla et al., 2020), make the nucleosome a physical "barrier" that controls the accessibility of the genetic material. The "beads-on-a-string" model, with nucleosome particles separated by 20-60 base pairs of DNA linkers becomes the foothold for studying the secondary folding of the chromatin.

The secondary structure of chromatin describes the folding of the beads-on-a-string motif in 3D space. This process is not only governed by the intrinsic mechanical property of 1-dimensional (1D) fibers that enables the spontaneous folding into a "programed" architecture, but also involves the participation of numerous condensation related protein factors, such as linker histone H1 (Wolffe, 1998), HP1 (Eissenberg et al., 1990), cohesins (Michaelis et al., 1997) among others and that takes place within the confined nuclear space. Recent studies showed that regular transcription factors can co-condensate with DNA to bring far away regions of DNA into close proximity (Quail et al., 2021; Renger et al., 2022). Since the diameter of a nucleosome core particle is approximately 10 nm, any fiber structure beyond the 10 nm width is considered a higher order chromatin fiber. Unfortunately, due to the complexity of chromatin structure and the limitations of experimental techniques (early studies of chromatin secondary structures were conducted using atomic force microscopy and transmission electron microscopy), no consensus about the higher order structure of chromatin has been attained during the last 30 years.

1.1.2 Evolution of miscellaneous chromatin higher order structure models

The difficulty to arrive at a general model of chromatin organization may come from the following reasons. On one hand, the genetic material confined within the small nucleus are highly condensed and intertwined. It is difficult to clearly distinguish the trajectory or pathway of the linear DNA along its sequence in a 3D volume given a limited spatial resolution. On the other hand, due to the heterogeneity and dynamics of the genetic material, no two chromatin structures are exactly the same, let alone the complications resulting from experiments performed on different cell lines and species, and chromatin material imaged and fixed at various stages of cell cycles. Moreover, the different imaging and sequencing methods require different treatment of the genetic material, such as staining, freezing or chemical fixation, which also leads to diverse outcomes and inconsistent conclusions.

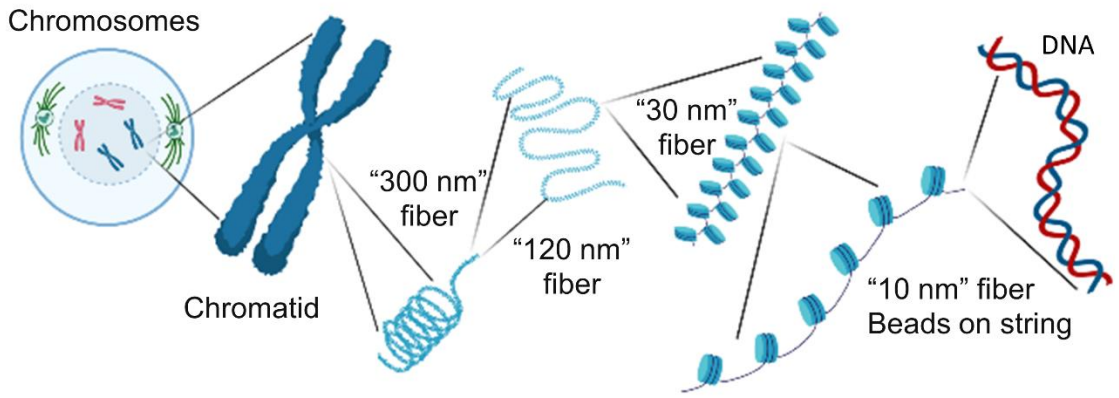


Figure 1.1 Schematic illustration of the historical model of hierarchical chromatin structure organization in a condense mitotic state. The figure was generated using BioRender.

The early discrepancy of high order structure of chromatin characterization arose from the concept of chromosome territory. However, when electron microscopy was used as the major approach for chromatin inspection in the late twentieth century, evidence of the existence of chromatin filaments of different diameters (Wolffe, 1998; Rattner and Lin, 1985; Finch and Klug, 1976a; Worcel et al., 1981; Widom et al., 1985; Widom, 1989; Woodcock et al., 1984a; Bednar et al., 1998), led to the abandonment of chromatin territory ideas. The imaging characterization of various nucleosome fibers lead to the proposal of the hierarchical folding model, which is descried in most textbooks nowadays (Figure 1.1). According to this model, formation of the higher order chromatin structure begins with the packaging of the "10-nm" nucleosome string into various forms of "30-nm" fiber arrangements (Song et al., 2014; Garcia-Saez et al., 2018; Swedlow and Hirano, 2003; Robinson et al., 2006). These "30-nm" fibers can further twist and intertwine, yielding the ~120 nm and 300 nm diameter super-helical fibers (Ou et al., 2017; Kireeva et al., 2004; Dehghani et al., 2005) and finally achieve the ~600 nm diameter fibers of the chromatid. (Daban, 2015a, 2014a). Embracing the view of the hierarchical fiber structures, microscopy studies at the same time and later began to shift their research direction toward answering more detailed questions in understanding what is the nucleosome arrangement within each type of filament, and especially within the initially formed "30-nm" fiber.

Again, due to the complexity and heterogeneity of genetic materials, multiple models of the "30-nm" fiber have been proposed to elucidate their internal nucleosome orientation and connection based on different observations of local chromatin fiber structures. These models include: i) the one-start solenoid structure (Robinson et al.,

2006; Finch and Klug, 1976b; Ohno et al., 2018), which describes a left-handed one-start helical structure that aligns 5.4 nucleosomes per helical turn. Nucleosomes from two adjacent turns interdigitate with each other forming a 33 nm diameter fiber; ii) the two-start helical/twisted-ribbon or two-start zig-zag model (Woodcock et al., 1984b) (Dorigo et al., 2004), where two parallel stacks of nucleosomes connected by DNA linkers form a super helical structure of ~30 nm diameter (The diameter of the helix depends on the length of the DNA linker length). High resolution 3D structures have provided some support for this model in vitro. For instance, x-ray crystallography revealed the zig-zag stacking conformation of tetra-nucleosomes connected by straight 20–base pair (bp) linker DNA in the absence of H1 (Luger et al., 1997a). Recently a cryo-EM SPA study showed a double helical structure of 30-nm chromatin fiber formed by tetra-nucleosomes with the presence of linker histone H1 (Song et al., 2014). Beside these two most popular structures, other models include iii) crossed-linker model (Williams et al., 1986; Staynov, 1983), where nucleosomes are non-sequentially arranged inside the helix and the DNA linkers connecting two nucleosomes transverse the central axis of the fibers; iv) the triple helix model (Makarov et al., 1985), which proposes that the fiber unit is composed of three piles of nucleosomes in a triangular geometry; v) the untwisted two-start helix (Garcia-Saez et al., 2018) proposes a similar structure as the two-start helical model but without twisting the fiber into a helical organization, but as a ladder-like structure.

Despite having higher spatial resolving capabilities, early electron microscopy (EM) imaging could not distinguish domains within the chromatin structures (Monneron and Bernhard, 1969) (this task proved to be possible for some specialized mitotic chromosomes, but generally difficult to detect fibers in somatic interphase nuclei). Therefore, other methods, such as fluorescence labeling or DNA sequencing techniques, which allow the identification of specific genomic regions has gained more attention by providing alternative ways to revisit the question of the genomic material organization. Methods including 3C classic, 4C, 5C and Hi-C (Lieberman-Aiden et al., 2009) make use of DNA crosslinking of regions found in close proximity in the nucleus. Digestion and subsequent processing of the cross-linked DNA through multiple steps of next-generation sequencing give the exact relative position of the cross-linked DNA sequences. Fluorescent labeling methods, including the early time single-molecule fluorescence in situ hybridization (smFISH) (Femino et al., 1998), recently developed DNA FISH (Beliveau et al., 2015) and RNA FISH (Chen et al., 2015) combined with super-resolution light microscopy, super-resolution stochastic optical reconstruction microscopy (STORM) (Bintu et al., 2018), and living cell imaging (Guan et al., 2017) techniques yields both the spatial and sequence information of the chromatin at intermediate resolution, which has became attractive tools to interrogate the physical arrangement of chromatin structure, connecting distinct organization principles with

different epigenetic domains.

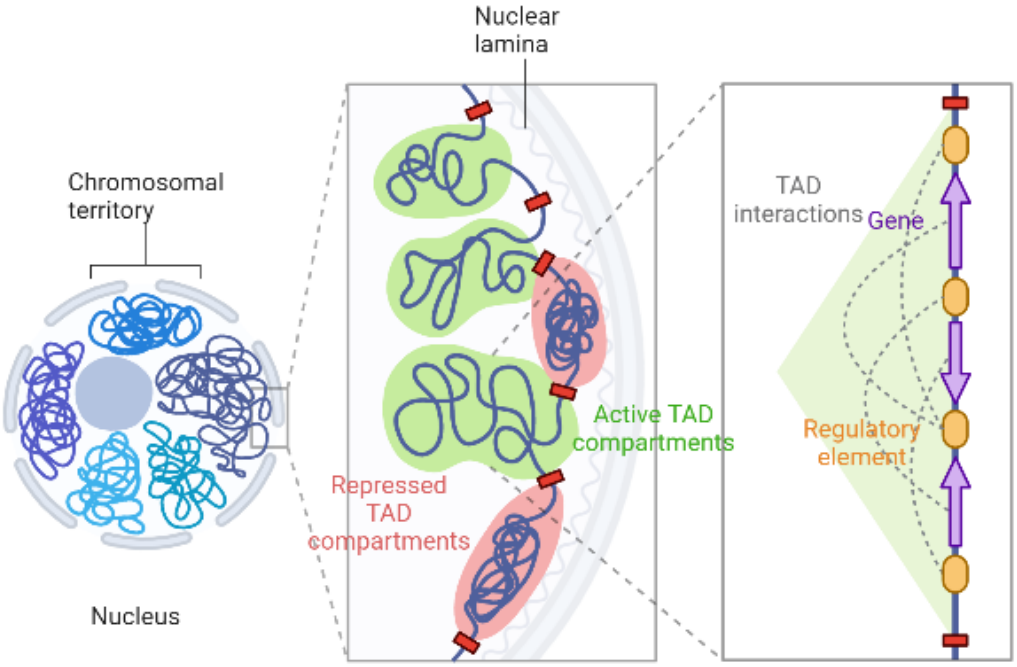


Figure 1.2 Schematic illustration of the current hierarchical interphase chromatin structure organization. Chromosome territories define discrete spaces within the nucleus. Topologically associating domains (TADs) are formed by loop extrusion and can be characterized by active and repressive histone modifications, as shown in the Hi-C heatmap illustration. The figure was generated using BioRender.

With these methods, a large amount of compelling evidence showing the existence of the chromatin domains emerged, revitalizing the concept of "chromatin territory" that had been abandoned decades ago (Dixon et al., 2016). The "territory" concept explains the hierarchical organization of chromatin in a completely new manner in contrast with the fiber-packaging models (Figure 1.2). The newly defined hierarchy model is composed of at least four levels (Krumm and Duan, 2019). i) Chromatin loops that link DNA promoters and their enhancers are closely related to gene activity and represent the low-level of genetic material organization. There are approximately 10,000 loops that can be identified in the Hi-C heatmap obtained from nuclear cross-linking assays (Rao et al., 2014). ii) Some larger, specific regions of chromosomes located either close to the nucleolus (nucleolus associate domain) or to the nuclear periphery region (lamina associated domain) have been reported and are thought to facilitate genome attachment and contribute to the genome arrangement and distribution inside the nucleus (Guelen et al., 2008). These domains are usually 0.1-10 megabases in size and represent regions of low gene expression. iii) Chromatin topologically associated domains (TAD, also called A/B compartments) refers to genome regions with constrained dynamics and distributed in the heterochromatin region. These regions

were also identified as megabase sized local domains that are highly self-interacting and conserved among species (Dixon et al., 2012). Many conventional cellular insulators participate in the boundary formation of these domains, as indicated by the binding of the associated zinc-finger-containing protein CTCF at the periphery of these regions. iv) Chromosome territory (CT) defines the highest global level of chromatin organization in nuclear space. Chromosomes have been observed with each of them occupying a distinct territory during the interconversion between interphase, prophase and metaphase and their orders are generally maintained (Guelen et al., 2008).

The solid evidence about genome organization as amorphous "territories" in the last decades is a major drive of the current EM studies of chromatin fibers which seek to attain a biologically relevant context that will make it possible to discriminate between the two types of "hierarchical" organizations proposed to connect structure and function. To reduce the artifact caused by chemical fixation or drying during sample preparation, cryo-EM combined with focused ion beam (FIB) milling technique were used to directly image native chromatin structures in cell sections at nanoscale resolution. Experiments interrogating chromatin structure inside cells in this manner have failed to consistently detect the "30-nm" fibers so far. (Eltsov et al., 2008; Maeshima et al., 2010a; Maeshima et al., 2016a). Instead, multiple in situ cryo-sectioning studies indicate that chromatin in the cell is generally composed of granular chains 5-24 nm in diameter regardless of the cell being in interphase or in a mitotic phase (Ou et al., 2017; Eltsov et al., 2008; Cai et al., 2018a; Scheffer et al., 2011a). These results suggest that the genome packaging might be merely due to the entanglement of DNA material at nucleosomal level without forming regular higher order fiber structures. According to this view, the "30 nm" fiber structure with a regular helical symmetry may represent a rare state of chromatin that can only be accessed at low salt condition (Maeshima et al., 2010b), suggesting that the fiber structures observed by conventional EM might be an artifact of sample preparation (Eltsov et al., 2014). The homogeneous granule-like distribution of nucleosome particles within the nucleus suggests that chromatin structure should be viewed as an amorphous "melted-polymer" at interphase.

1.1.3 Nucleosome structural dynamics

In polymer physics, the word "melt" refers to a state in which nucleosomes lose their "memory" by interacting non-preferentially with surrounding nucleosome particles regardless of which fibers these belongs to (Eltsov et al., 2008). In this case, the nucleosome fibers no longer adopt a well-defined secondary structure. Instead, each nucleosome array can populate a huge dynamic phase space. The dynamic properties of the nucleosome arrays then reflect:

i) The intra-nucleosome dynamics, which result from the "breathing" motion of the nucleosome core particle (Polach and Widom, 1995; Winogradoff and Aksimentiev, 2019) induced by the thermo motion of the particle, and involving the transient unwrapping and re-wrapping of the DNA from the histone octamer. Evidence of the DNA unwrapping transition can be found in studies of restriction enzyme accessibility to the DNA target site located at the inner wrap of nucleosome (Anderson et al., 2002a) and in studies of nucleosome translocation with or without the involvement of remodelers (Lai and Pugh, 2017). The electrostatic interaction between the positively charge histone and negatively charged DNA mainly contribute to the stability of the nucleosome core particle. Therefore altering the nucleosome surface charge via histone modification (Anderson et al., 2001; Simon et al., 2011) and changing DNA bending rigidity through DNA sequence (Ioshikhes et al., 2006) are basic cellular tools for regulating intra-nucleosome dynamic and nucleosome positioning.

ii) The inter-nucleosomal dynamics related to the varying spatial relationships between nucleosomes topologically constrained within a "beads-on-a-string" polymer chain. One can view it as a similar question of interrogating peptide chain dynamics, however, the difference is that more rotational constraints are imposed on each nucleosome "atom" and fewer bending constraints are applied to the DNA linker "bond". In detail, the freedom of the nucleosome array polymer is not only depending on the entry-exist position of DNA linker on nucleosome surface and its angles relative to the nucleosome discoidal plane, but also the length and curvature of the DNA linkers and the induced nucleosomal rotation at the given linker length. Together these parameters define the relative orientation between two consecutive nucleosome particles, and ultimately the transient shape of the whole nucleosomal array or fiber. Taken together, it is not hard to understand why given the large degree of freedom and uncertainties of each nucleosome, a consensus folding model is missing, especially for long arrays.

In vitro study that seek to shift the nucleosome array global dynamics by changing the ionic environment result in a dramatic change of array geometry from the "melted-polymer", to the "30-nm" fiber and to lager condensed aggregates as revealed by characteristic length in small-angle X-ray scattering (SAXS) experiments (Maeshima et al., 2016b). These in vitro experiments may simulate the cellular process of recruitment of transcriptional factors to the array, creating a lightly packed fiber organization (as most all of these factors are carrying positively net charge (Brendel and Karlin, 1989). However, the detail structures and regulatory mechanisms involved are still missing for this latest model. In chapter 3, a systematic characterization of nucleosome array was

performed in order to bridge the gap between nucleosome array structural dynamics and theoretical models.

1.1.4 Nucleosome phase separation

To integrate the above notions of "melted-polymers" and "chromatin territory" into a chromatin structure model, a transition process, named liquid-liquid phase separation (LLPS) has been proposed recently, which not only incorporates the polymorphism of nucleosome fibers (i.e. the liquid behavior of genomic materials) but can also explain the compartmentalization of chromatin (Maeshima et al., 2016c; Zhang and Kutateladze, 2019). Phase separation describes a physical process in which a solute material in a supersaturate solution spontaneously segregates into two stably coexisting phases: a dense phase and a dilute phase. It has been shown that the liquid-liquid phase separation phenomenon is a common phenomenon inside the cell, which is responsible for the generation of many membrane-less 'organelles', including P granules, Cajal body, nucleus speckle, etc. (Marnik and Updike, 2019). It is believed that the spontaneous self-assembling of these liquid-like condensates help in the sorting of cellular components and is being regarded as a previously unrecognized mechanism for triggering cellular events and performing local enzymatic reactions.

The phase separation process has been shown to occur with proteins that carry low complexity domains or intrinsic disorder regions (a long strand of unfolded peptide chains on their N- or C-terminal). The weak multivalent contacts including nonspecific hydrophobic and electrostatic interactions among these proteins and nucleic acid are believe to play a key role in phase separation (Li et al., 2012). Significantly, histones H2B, H3, and H4, contain a long-disordered region on their N-terminal (~30-40 amino acids), H2A on its C terminal for (~30 amino acids), and H1 and HP1 are completely disordered for their entire sequences. (Iwasaki et al., 2013; Turner et al., 2018). Fluorescence microscopy experiments have shown that the full-tail dodecamer nucleosome arrays undergo liquid-like phase separation under physiological salt conditions, whereas their tail-less counterparts fail to do so under similar conditions (Gibson et al., 2019). Moreover, many proteins involve in transcription including the polymerase, transcriptional factors and co-activators form liquid-like condensates in vivo, which suggest an essential role of liquid-liquid phase separation in the regulation of transcriptional activity. Therefore, understanding the self-interacting properties of nucleosome arrays and their participation in liquid-liquid phase separation is critical to decode the spatial and temporal regulatory mechanism of gene expression. In chapter 4, a cryo-ET structural study on the early stages of nucleosome phase separation process

was performed on a streptavidin crystal surface. A two-step transition mechanism was proposed to describe the generation of two different types of liquid-like condensates.

1.2 Single particle imaging method to study biomolecular dynamics

Studies pertaining to cellular processes and chemical events began in the 17th century, a time when scientists were still "blind" to the molecules responsible for the biological activity, requiring a great amount of effort and intelligence to make scientific progress. More recently, though conventional biochemistry studies achieved numerous significant breakthroughs from the molecular ensembles used in bulk experiments, important structural and dynamical information of the molecules involved in a given process was found missing. The dynamic character of proteins and nucleic acid dictates their functions and portrays their "personalities" (Pacini et al., 2021). Snapshots of proteins obtained through X-ray crystallography, for example, reveal unique structures that are often used as blueprints to study their structure-function relationships. However, molecules are dynamical entities involving in their activity both equilibrium fluctuations, as well as non-equilibrium departures, as is the case of biological motors, which convert chemical into mechanical energy. This information is often averaged out in the ensemble methods of conventional biochemistry, rendering them insufficient to fully characterize the activity of these molecules.

1.2.1 Single particle techniques

The advent of single-molecule techniques in biophysics represents an important step in our efforts to characterize biological processes, either through the use of force spectroscopy (manipulation) methods (Bustamante et al., 2000) or single molecule detection (Ha and Tinnefeld, 2012) (Murata and Wolf, 2018) of individual molecules. Single-molecule techniques include: i) Force-based manipulation such as magnetic tweezers, optical tweezers, and atomic force microscope-actuators (AFM). These methods can be used either to exert forces on molecules or to detect the forces generated in the course of their reactions, in real time and in aqueous environment. Force spectroscopy methods are ideally suited to explore the mechanical properties of molecules and thus are widely used in studies of motor proteins and protein folding. ii) Fluorescence-based detection methods, including super-resolution fluorescence microscopy and single-molecule fluorescence resonance energy transfer microscopy (FRET) are convenient means of observing one molecule at a time both in vitro and in vivo. Single-molecule fluorescence is currently the subject of intense interest for biological imaging, through the labeling of biomolecules such as proteins and

nucleotides to study dynamical biological processes that cannot easily be followed in bulk because of the impossibility of synchronizing molecular activity. iii) Structural-based single particle imaging methods, such as X-ray crystallography and cryo-electron microscopy (cryo-EM) single particle analysis (SPA) that use theoretical algorithmic calculations to retrieve 3D structural information from diffraction patterns or 2D images, respectively. It's worth noting that although the analysis yields one structure and it is thus called "single-particle analysis", the 3D reconstruction is obtained by averaging thousands to tens of thousands of particles that are assumed to share a single structure. The assumption is valid for a protein that is structurally rigid for which it is possible to attain atomic resolution.

Each of the above methods has its own pros and cons, for example the single molecule force-spectroscopy methods and single molecule fluorescence microscopy are ideal to characterize short-lived or transient states of a molecule while allowing the collection of dynamic information in real time. However, for some complex systems, the signal is convoluted and can become very difficult to analyze without the corresponding structural information at hand. Conversely, single-particle reconstruction methods provide the structural information of molecule in a great detail. However, X-ray crystallography requires molecules to be in the solid state, packed closely together with a defined pattern, and single-particle cryo-EM are mainly used to solve the most populated states of the protein by ruling out the unalike. These methods usually fail to reconstruct heterogeneous systems, such as lipoproteins or flexible polymer molecules, such as chromatin fibers, which limits their ability to study dynamic changes of the biomolecules.

1.2.2 Individual particle cryo-ET

More recently, cryo-ET approach provides impressive structural details of protein structures and cellular organizations. Cryo-ET is an alternative approach to SPA that does not require the averaging signals from thousands to millions different particles. Instead, 2D images are collected from a series of tilted viewing angles from the same object, and the alignment of the tilt series followed by image back projection produce a faithful representation of the original 3D object (Figure 1.3). Indeed, the methodology of ET reconstruction was established even before the single-particle analysis, and has been applied to the study of large biomolecules, such as bacteriophage T4 (De Rosier and Klug, 1968) and tobacco mosaic virus (Hart, 1968). However, due to the resolution limitation at the moment, this method was mainly used for the 3D reconstruction of radiation-insensitive hard materials. Thanks to important technological developments in instrumentation, computational power, reconstruction algorithms, and cell sectioning techniques (e.g. focused ion beam (FIB) milling), ET has now become a dominate method

for exploring heterogeneous molecular structures, which has long been the ambition for the application of EM in cell biology.

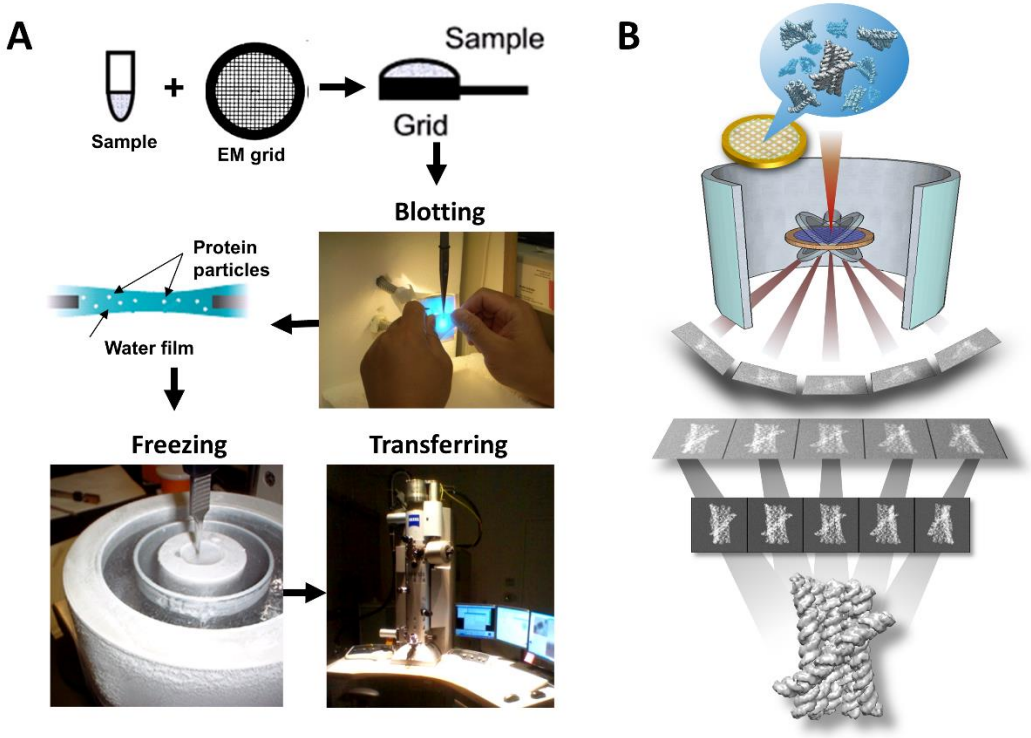


Figure 1.3 Workflow of cryo-ET and image acquisition and computational reconstruction principles. A) cryo-ET sample preparation. B) Schematic of single-tilt axis data acquisition. The sample grid is imaged in the microscope by tilting it over a typical range of ± 60° with small tilt increments. The image collected at each tilt angle is the projection of the 3D electron density of the specimen onto a 2D plane, which holds all information about the specimen. The projection images at different angles are projected back into the volume to be reconstructed.

Therefore, the notion of individual-particle cryo-ET refers to the independent 3D reconstruction of each particle without any further averaging operations. Though it might not be comparable to single particle analysis in terms of resolution, the structural and dynamical properties of the particle can be studied at the same time. The distribution of the molecular properties of heterogenous systems can be obtained from many snapshots of different molecules. The method is, therefore, ideal to characterize systems that have large structural variability and flexibility.

1.2.3 Individual particle ET 3D reconstruction method

Dr. Ren, who has long been devoted to pushing the boundary of individual particle ET 3D reconstruction has used this approach to solve many challenging problems such as the heterogeneous structures of IgG antibody (Jay et al., 2018) and high-density lipoprotein (Yu et al., 2016). This methodology uses a robust algorithm, Focused ET Reconstruction (FETR) (Lei et al., 2018a), which can tolerate small tilt-errors and accurately determine two translational parameters for each tilt image. FETR is an iterative refinement procedure that includes a series of automatically generated dynamic filters and masks to enhance the convergence of the reconstruction. To limit the adverse effect of tilt-errors, the size of reconstructed ET images are reduced by capturing only a single-instance conformation of a protein particle, instead of large whole-micrograph images through conventional methods that contain dozens of averaged protein structures. The most interesting feature of this method is that the 3D reconstruction of each individually targeted molecule is free of conformational dynamics and heterogeneity, proving that the 3D structure can be treated as a snapshot of the dynamic structure of the macromolecule. By comparing these "snapshot" structures, this method permits the study of macromolecular structural dynamics (Lei et al., 2018a; Ren et al., 2010; Zhang et al., 2015a; Zhang et al., 2015b; Lu et al., 2016). This approach was been named individual-particle electron tomography (IPET).

1.2.4 Application of individual particle ET reconstruction on DNA-related material

Advances in biophysical research hinges in part on improved characterization of structural features of biomolecules. DNA, the core of the "central dogma" not only serves as a media of storage and transmission of genetic information, but also provides a building block for the assembly of organic–inorganic hybrid nanostructures, making it a compatible and programmable substrate for molecular bioengineering (Li et al., 2013), drug delivery systems (Wang et al., 2018), and other interdisciplinary applications (Zheng et al., 2006; Alivisatos et al., 1996; Elghanian et al., 1997; Jones et al., 2010; Cigler et al., 2010; Zhang et al., 2013). Although the fundamental structural features of DNA have been thoroughly described and confirmed through various imaging techniques (i.e., x-ray crystallography, NMR, etc.), the local DNA structural fluctuations related with its mechanical property hinder the capability of inducing a fully controllable dynamic behavior in designed structures consisting of protein-DNA complex or solely of dsDNA (DNA nanoparticles).

The IPET methodology has been applied to the study of flexible DNA complexes.

The first one we discuss here is DNA origami. DNA origami fabrication is a powerful technique, which allows for the construction of nanoscale objects with geometric complexity. In 2006, Rothemund introduced a method in which a long single strand of DNA is folded into arbitrary shapes by hybridizing with programed short oligonucleotides (Rothemund, 2006). Recently, Ren et al studied the 3D structural dynamics of DNA origami using IPET (Lei et al., 2018a). Initial analysis of the DNA origami Bennett linkages with an electron microscope showed that the particles were "quadrilateral" in shape with a square-like "open" and a compacted, bundle-like "closed" conformation, which had not been previously reported. Negative staining images showed four arms each composed of four visible layers with a curved shape suggesting that the structures are highly flexible. The survey of the cryo-EM micrographs confirmed the quadrilateral-shaped particles ranging from "open" to "closed" conformations. IPET 3D reconstruction was performed on the cryo-ET tilt series of the positive-stained Bennett linkages sample. The reconstructed map confirmed the overall quadrilateral shape as well as the square-like open and compacted closed confirmation of the DNA origami Bennett linkages. The IPET 3Ds also allowed for the identification of the spatial location and orientation of each arm of the Bennett linkages, which pointed out to potential approaches for improving the control of Bennett linkage conformation for future applications.

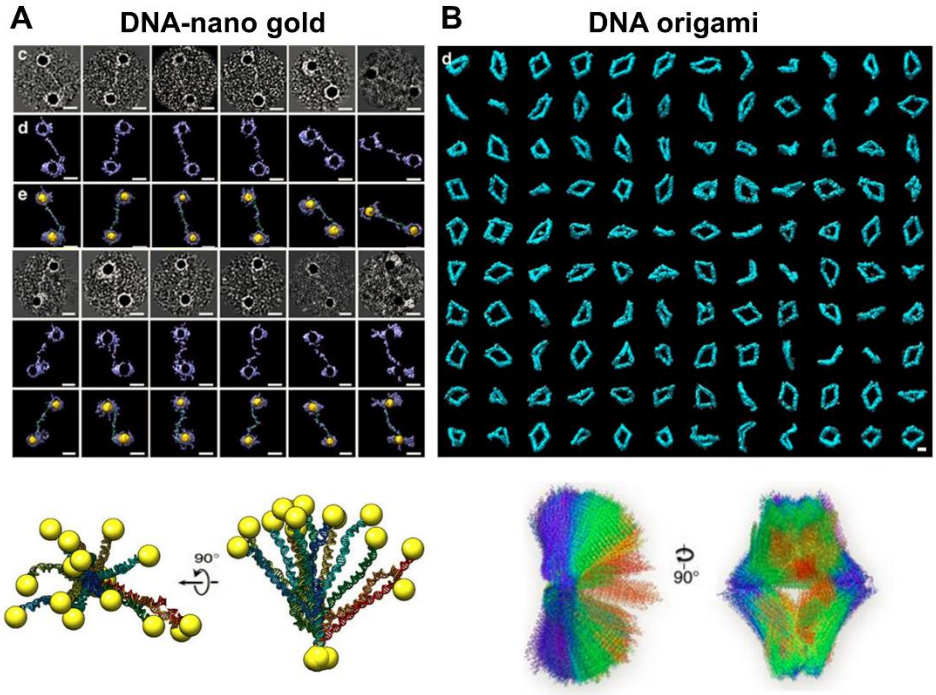


Figure 1.4 Application of IPET on DNA-related materials. A) Representative particle views, 3D reconstruction, and modeling of 84-bp DNA-nano gold conjugates. The DNA segment dynamics allow calculation of DNA bending energy. B) Representative particle

views, 3D reconstruction, and modeling of DNA origami Bennett linkages revealing the structural dynamics of the construct (Adapted from zhang *et.al.* and Lei *et.al.*).

Another application IPET has been the characterization of DNA-nano gold conjugates. They had previously been characterized by TEM 2D imaging (Buckhout-White et al., 2012) and by atomic force microscopy (AFM) (Prado-Gotor and Grueso, 2011). Further advances in characterization were obtained from their imaging in liquid chamber (Keskin et al., 2015) or in 2D crystal array (Zheng et al., 2006). Despite many previous attempts to reconstruct DNA-nanogold conjugates in 3D, the results failed to reveal each individual macromolecule's structure accurately. IPET, cryo-EM, and optimized negative staining (OpNS) techniques were more recently employed to image and reconstruct hybridized small DNA-nanogold conjugates in three dimensions, and subsequently used to calculate their bending energies (Zhang et al., 2016a). Cryo-EM and OpNS micrographs showed that pairs of polygonal-shaped nanogold particles were in proximity with each other, strongly linked by a fiber-shaped density corresponding to dsDNA. Furthermore, higher-contrast NS images of the nanogold particles showed thicker bridging densities and lengths comparable to those of full-length dsDNA. IPET reconstruction was used in conjunction with cryo-EM to obtain the 3D structure of the dsDNA-nanogold conjugates. By flexibly docking a standard structure of 84-bp DNA with the 3D maps using molecular dynamics simulations, 14 conformations of dsDNA were established. Average bending energies computed using local bending angles from IPET 3D reconstructed models yielded DNA bending rigidities three times smaller than previously calculated. The 3D reconstruction of the 52 kDa 84-bp dsDNA using IPET technique showed a promising method of visualizing and reconstructing in 3D individual small and asymmetric macromolecules previously not possible using other conventional methods.

Chapter 2
Development of advanced individual particle cryo-electron tomography reconstruction workflow

The three-dimensional structural determination of asymmetric and flexible protein and nucleic acid molecules by electron tomography (ET) provide a unique tool for understanding their functional dynamics. In Chapter 1 we reviewed the main ideas, workflow and applications of IPET reconstruction refinement methodology. As shown there, IPET has become a robust strategy to study flexible DNA molecules and their related materials, especially when forming a large complex structure (origami) or when being prepared as negative stained samples. The molecular size and particle image contrast or its signal-to-noise ratio (SNR) are factors that can critically determine the quality of the reconstruction. Since our current goal is to study chromatin dynamics by examining the local dynamics of its individual units in a native environment (vitreous ice), both of the above criteria cannot be fully met meet in this case, which greatly reduces the capability of the current reconstruction method. In this chapter, I evaluate how different factors affect the result of individual particle reconstruction of DNA and nucleosomes within the cryo-EM specimen. Based on these results, some improvements or alternative approaches that aim to increase the image contrast, fill the missing wedge data, and optimize data collection were adopted and integrated into the IPET work flow during the reconstructions.

2.1 Determination of critical factors limiting the quality of the cryo-ET individual particle reconstruction

Due to the radiation sensitivity of the soft biological sample, the total e^- dose used for collecting a snap-shot image of the molecule is usually in the range of 100-200 $e^-/Å^2$, an empirical value that represents a compromise between maximum image contrast and radiation damage of particles embedded in vitreous ice (Figure 2.1).

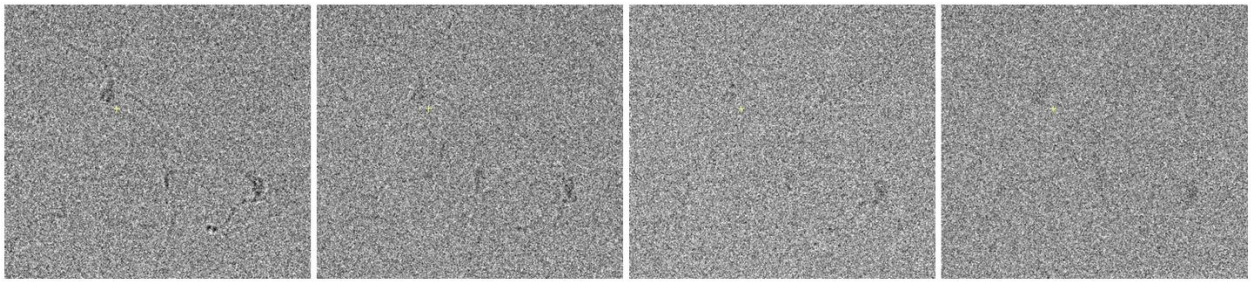


Figure 2.1 Radiation damage of a mono-nucleosome particles. Four consecutive exposures of nucleosome particles in vitreous ice. The electron dose used for each image is ~100 $e^-/Å^2$. Particles gradually become blurred after each time of electron beam exposure.

Instead of a single exposure to the beam, tomogram data collection requires imaging the same object dozens of times. By sharing the total dose among the images collected at different tilt angles, tomography data usually renders a poor image contrast (the dose for each tilt image is limited to value of 2.5-5 $e^-/Å^2$, assuming 40 images are collected). Because previous applications of IPET enable to clearly resolve the 3D pathway of individual DNA molecules by using the high contrast derived from negatively stained samples, here we try to both qualitatively and quantitively evaluate whether a similar level of reconstruction can be achieved when using low-dose cryo-ET tilt series. However, experimentally this study is complicated by the fact that we can never know whether the low contrast image-induced misalignment or the low contrast image itself are the cause of the failure of reconstruction. In this case, we used the simulated mono-nucleosome images obtained from the back projections of the crystal structure model (1AOI, with extended DNA arms) to perform the analysis, a strategy that permitted the separation of the above two factors.

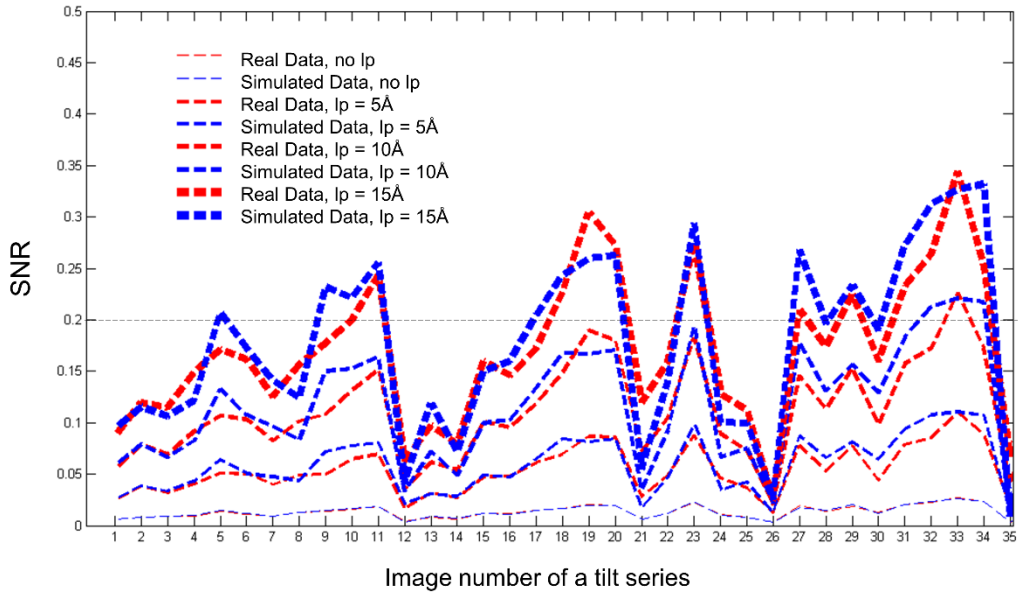


Figure 2.2 Measurement of the average SNR of cryo-ET images. The red dashed lines indicate the SNR measured from the cryo-ET tilt series (35 images). The blue dashed lines indicate the SNR of simulated tilt series (35 images), with particles containing the same level of gaussian noise as the real data. Low-pass filters from 0-15 Å, in steps of 5 Å, were applied to both image series to boost the SNR.

Before generating these simulated projection images at each tilt angle, we first need to determine the SNR from experimental data before applying it to simulated images. The SNR can be defined as (Is–Ib)/Nb, in which Is was the average intensity of a particle and Ib was the average intensity beside the particle. The standard deviation of noise, Nb, was calculated from background area outside the particle. By measuring the Is of nucleosome particle and the Ib and Nb of the surrounding ice area, the calculated SNR from the raw image of the cryo-ET tilts series (in range of ± 51° in step of 3°, corresponding to 35 images) showed values in a range of 0.01 to 0.02 (Figure 2.3). By normalizing the projection images of the model to the mean level of Is and applying the same mean and standard deviation of the gaussian noise as Ib and Nb, respectively to the normalized image, the simulated ET tilt series was generated.

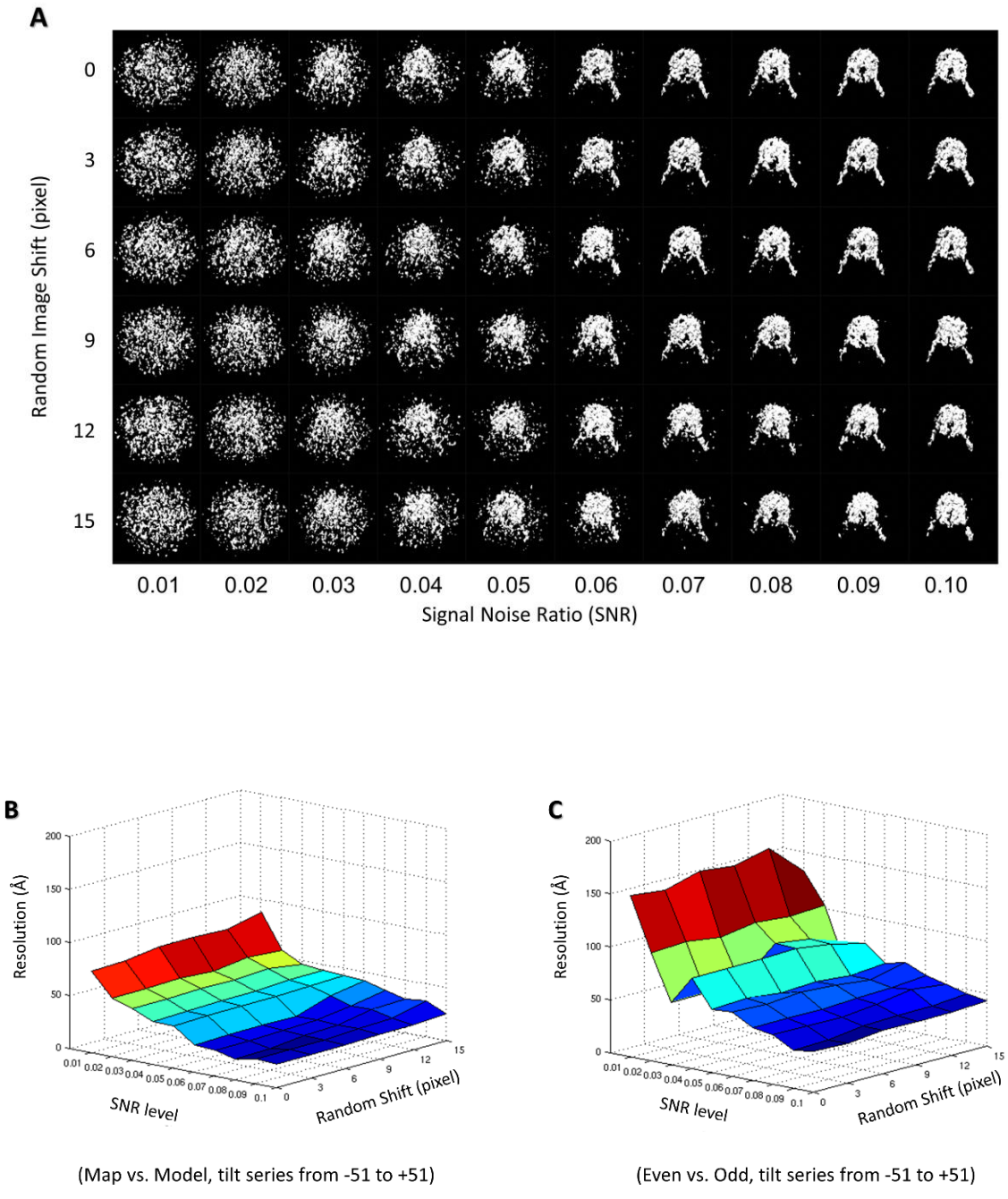
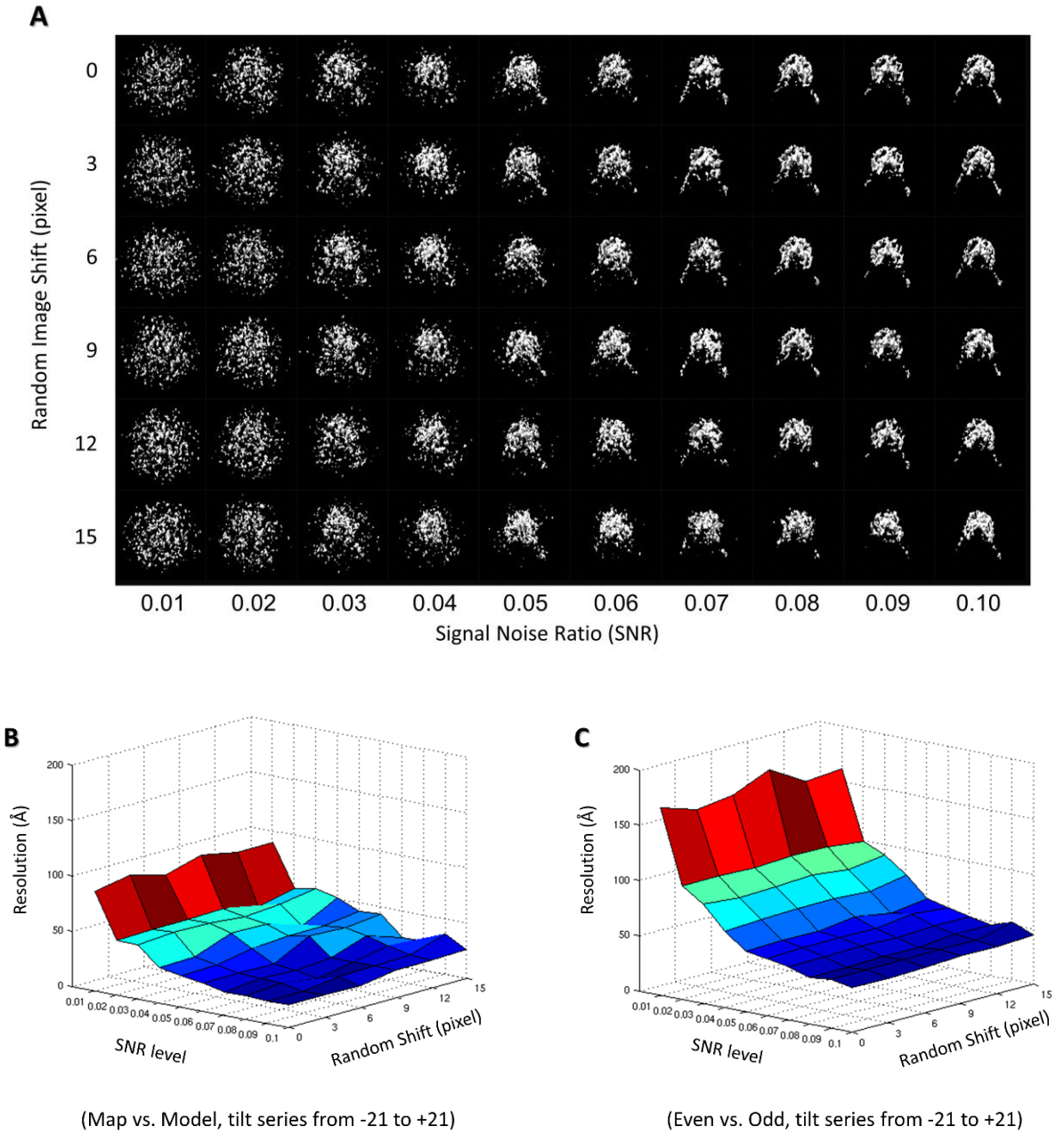


Figure 2.3 3D reconstruction on a simulated tilt series at varying SNR levels and alignment accuracies. A) Matrix of 3D density map reconstructions of mono-nucleosome particles from simulated tilt series with SNR ranging from 0.01 to 0.1 and image shift errors ranging from 0 to 15 pixels. The reconstructions were performed using a tilt series from -51° to +51° in steps of 3°. B and C) Evaluation of the reconstruction resolution by calculating the FSC between the reconstructed map and the crystal

structure, and between maps that were reconstructed from even and odd number images, respectively.

Given the angle information of the projection images and the pre-calculated noise level, the 3D reconstruction of the simulated tilt series image at different signal-to-noise level (from 0.01 to 0.1 by changing the value of Ib, where Is while Nb remained the same) and image alignment error (from 0 to 15 pixel of random image shift) was established. As it shown in Figure. 2.3A, a noticeable 3D reconstruction quality difference was observed along the horizontal direction, corresponding to the change of tilt series SNR. At 0.01 SNR, the reconstructions are fuzzy. The overall shape of the nucleosome particle cannot be distinguished from the background. When the SNR reaches a level of 0.05, we start to be able to distinguish the discoidal-shape of the nucleosome core particle, though the two DNA arms are still blurry. While, at SNR of 0.1, both of the nucleosome core and DNA arms standout from the background. On the other hand, the difference induced by the image alignment error is not very dramatic compared to the SNR variation, as there is no significant structure quality transition along the vertical direction. Based on the image pixel size of 0.74 Å/pix, a maximum of random image shift of 15-pixel in our simulation (corresponded to a 1.11 nm image alignment error) is under the coverage of the IPET alignment accuracy. Therefore, the major limitation to achieve a better 3D reconstruction for the individual particle cryo-ET depends on the SNR of each tilt image.

To qualitatively confirm the above observation, the reconstruction resolutions were estimated by calculating the Fourier Shell Correlation (FSC) between the map and the crystal structural model (Figure 2.3B), and between the maps reconstructed from even tilt images and odd tilt images (Figure 2.3C). The resolution analysis showed that the increased SNR induced a ~60 Å and 100 Å improvements on reconstruction resolution (from ~20 to ~80 Å and from ~50 to ~150 Å, evaluated by the two schemes mentioned above, respectively). In contrast, the reduction of alignment error only induced a minor change of ~12 Å on average for both evaluation methods. Therefore, increasing the SNR of each tilt image should be considered as the main objective to improve IPET performance in cryo-ET reconstruction. To do so, one simple approach is to distribute the same amount of total electron dose to fewer images. To test the capability of this method, a similar reconstruction simulation was conducted on the same image sets but within a narrower tilt angle range (from ± 51° to ± 21°, Figure 2.4). The result showed that though the resolutions of partial tilt images reconstruction were only decreased for 5~10 Å compared to the ± 51° reconstruction results, the DNA displayed a faint density and its helical pitch patterns almost disappeared. Therefore, alternative approaches are required to improve the SNR for each image without changing their electron doses.



(Map vs. Model, tilt series from -21 to +21)

(Even vs. Odd, tilt series from -21 to +21)

Figure 2.4 3D reconstruction from a simulated partial tilt series at varying SNR level and alignment accuracy. Matrix of 3D density map reconstructions of mono-nucleosome particles from simulated tilt series with SNR ranging from 0.01 to 0.1 and image shift error ranging from 0 to 15 pixels. The reconstructions were performed using a tilt series from -21° to +21° with steps of 3°. B and C) Evaluation of the reconstruction resolution by calculating the FSC between the reconstructed map and the crystal

structure, and between the maps that were reconstructed from even and odd number images, respectively.

2.2 Implementation of deep learning-based denoising approach

As a conventional method, the low-pass filter is a simple and effective approach for improving the SNR of an image by removing the high frequency noise from Fourier space. Considering that a better reconstruction was obtained when SNR is around 0.1 in the simulations, which is consistent with the fact that negative staining-ET tilt series usually have SNR above 0.2, to obtain a similar level of SNR of 0.1-0.2 for cryo-ET tilt series, at least a 15 Å lowpass filter cutoff needs to be applied to the raw images (Figure 2.2). Since deep-learning base denoising methods have received much attention for its application to fluorescence and electron microscopy, here we tested if this method can provide superior results compared to the 15 Å low-pass filter by losing less of high frequency signals and by gaining the weight of low frequency signals.

In this new approach of deep learning based-image denoising or restoration, the Noise2Noise training strategy (Lehtinen et al., 2018) is the most applicable method to electron microscopy images, since it requires no ground truth of an image which is the general case for EM imaging. Traditionally, a network needs to be trained on pairs of images, *i.e.* a noisy image composed of true signal and noise (s + n) and a noise free ground-truth image (s). In contrast, the network training in a Noise2Noise regime only requires two noisy images that contain independent zero-mean noise ($s + n_1$ and $s + n_2$). Because for ET data collection, a long exposure is generally composed of 8-10 short exposures, *i.e.* frames, the averaged image ($s + n_1$) from all even frames and the averaged image ($s + n_2$) from all odd frames meet the above requirement, which facilitate the implementation of the Noise2Noise method by using the CARE network (Buchholz et al., 2019; Weigert et al., 2018) (Figure 2.5).

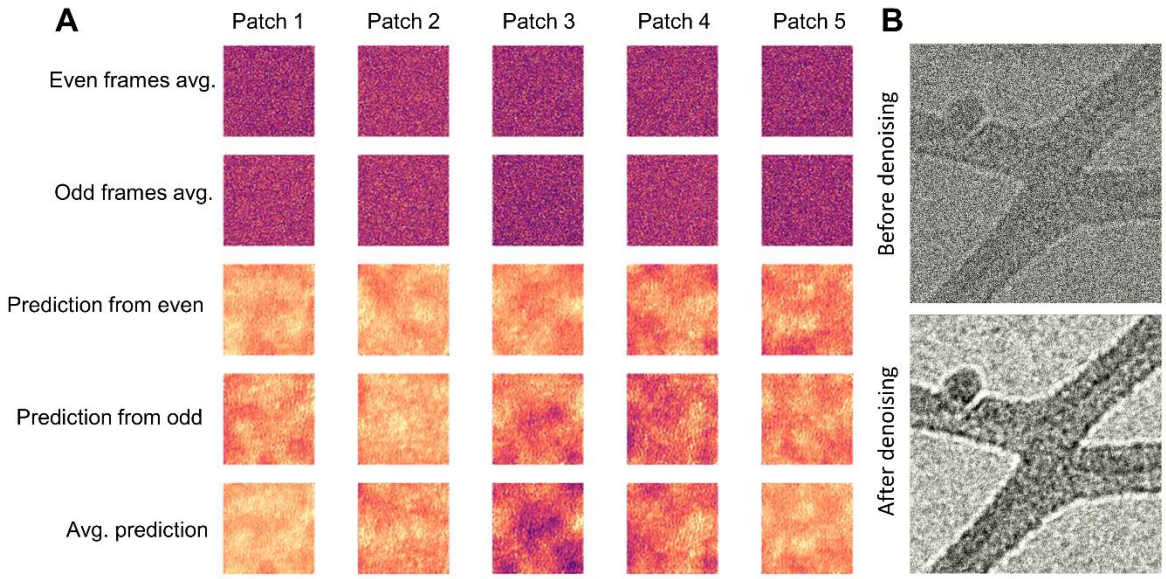


Figure 2.5 Application of deep learning-based denoising to low-dose cryo-ET raw images. A) Construction and application of a CARE network demonstrated with representative image patches (128x128 pixels) containing different features. Prepared representative image pair that averaged all even frame patches (first row, first column) and all odd frame patches (second row, first column) on the same object. Another ~12,000 image pairs were prepared in a similar way (four more are shown here, column 2 through 5) and used as training source and target, respectively, to construct the neuron network. The constructed network was applied back to the averaged even and odd images, yielding the prediction of the denoised even (third row) and odd (fourth row) image pairs. The average of the denoise image pairs gives the final prediction of the target feature (fifth row). B) Image comparison before and after the application of deep learning-based denoising of a larger image area (1200x1200 pixels).

Above result (Figure 2.5) showed a significant contrast enhancement for the low dose ET image, which enabled a level of SNR improvement (0.16) similar to the 15 Å low-pass filter. To understand the benefit of the deep learning based-method compared to conventional low-pass filters, we compare their results in a frequency regime (Figure 2.6). The power spectrum of the image (Figure 2.6) showed that the low contrast of the raw image is due to the weak intensity of its low frequency signal in the Fourier space. In contrast, both the low-pass filter and deep learning-based denoising method remarkably increase the low frequency signal intensity of the image, but at different cost of the high frequency signal intensity. Indeed, while the 15 Å gaussian low-pass filter loses its high

frequency signal intensity at ~11 Å, the image binning followed by deep learning-based denoising and its inverse operation lose their high frequency intensity at ~13 Å and ~7 Å, respectively. As it is seen from Figure 2.6, the suppression of the high-resolution information yields a coarse-grained texture of the images. Given the same contrast enhancing capability, the operation of applying deep learning-based denoising on intact raw images is a superior strategy for the denoising workflow.

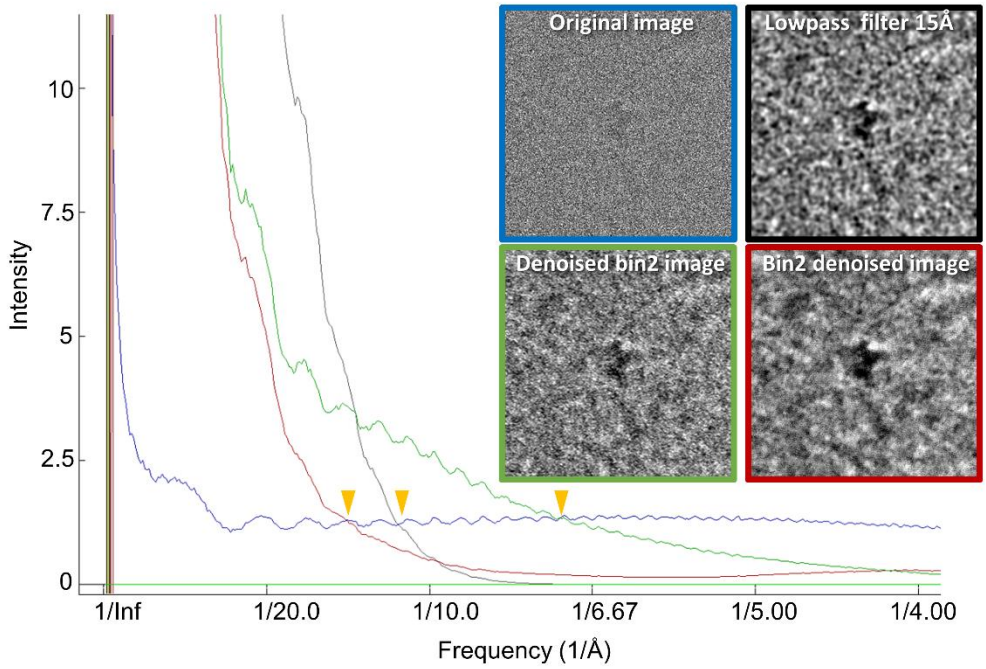


Figure 2.6 Comparison of different contrast-enhanced schemes in Fourier space. The line plot displays the power spectrum of four representative image (inset panel) for a same object with different denoising schemes, including the raw image after a median shrink (bin= 2) (blue), a 15 Å gaussian low-pass filter followed by a median shrink (bin= 2) (black), a deep learning based-denoising followed by a median shrink (bin= 2) (green), and image binning (2) followed by deep learning based-denoising (red). Yellow arrows indicate the points where the high frequency signal intensity began to be repressed compared to the raw image.

The 8 Å-resolution imaging defines the lower end threshold for entering the atom level inspection of the molecular structure that is theoretically not achievable by ET techniques. Therefore, a smarter strategy is to trade off this part of information in exchange for low frequency signal intensity by further applying low-pass filter on denoised images (Figure 2.7). By varying the strength of low-pass filtering values from 1

to 15 Å, the 5 Å the gaussian low-pass filter reached the targeted 8 Å cutoff, which doubled the low frequency signal intensity measured at 50 Å (Figure 2.8). This initial image contrast enhancement workflow on raw images was applied to all of the following cryo-ET reconstructions. Further resolution evaluation of nucleosome array particles reconstructions will be discussed in Chapters 3 and 4.

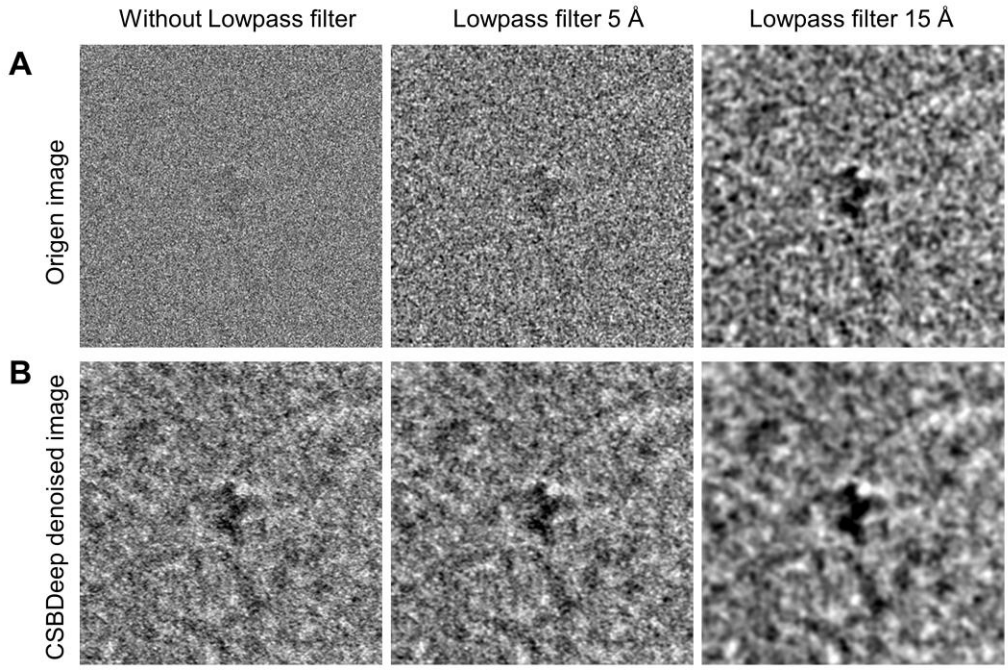


Figure 2.7 Combined use of denoising stratagems. A) Representative particle images after applying different level of low-pass filters. B) Representative particle images after sequentially applying deep learning based-denoising and different levels of low-pass filters.

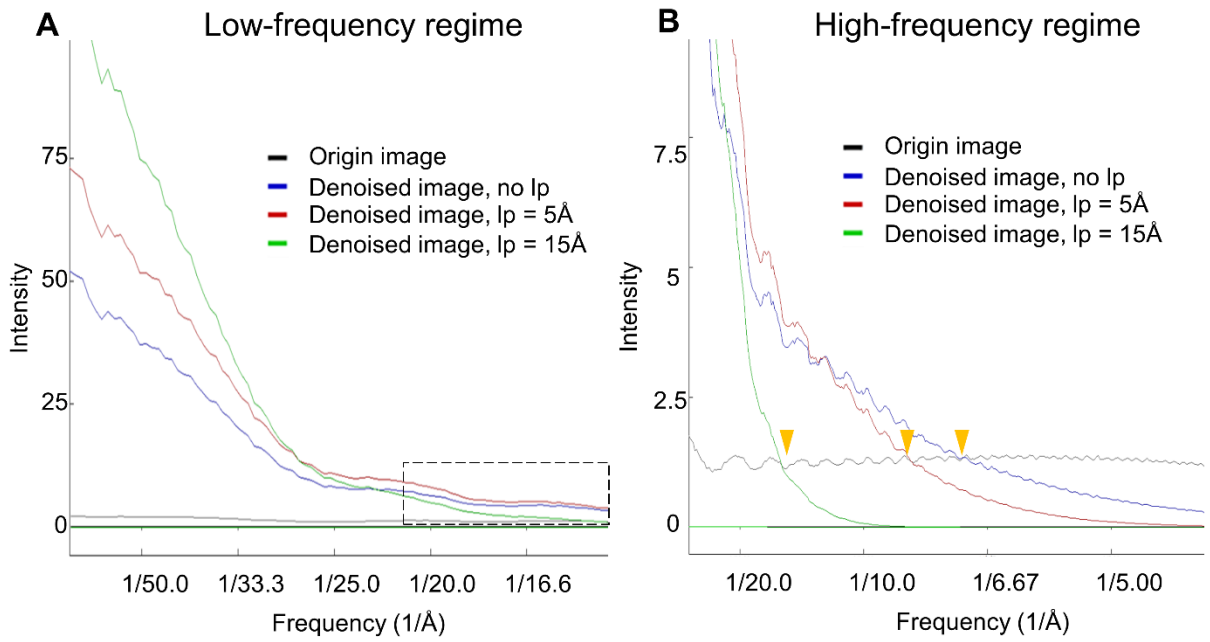


Figure 2.8 Comparison of different contrast-enhanced schemes in Fourier space. A) Low frequency signal enhancement via combined use of denoising and low-pass filtering methods. B) Preservation of the high frequency signal with combined use of denoising and low-pass filtering methods. Yellow arrows indicate the points where the high frequency signal intensity began to be repressed compared to the raw image.

2.3 Image contrast transfer function (CTF) correction

The contrast transfer function (CTF) is a mathematical description of how aberrations caused by the magnetic lens modify the image of an object, which is a partially correctable "artifact" of the transmission electron microscope. Based on the previous result of denoising, the deep learning-based method largely preserved the CTF feature of the image. As shown in Figure 2.6, the peaks of the wave-like blue and green curves consistently matched each other, but these features disappeared in the red and black curves. Therefore, before applying further the low-pass filter to the images, the CTF correction should be performed on the deep learning-based denoised tilt series with the same defocus value measured from the raw tilt series. To facilitate the defocus value determination for each image in the tilt series, a carbon area lying perpendicular to the tomogram tilt axis was deliberately included within the images during the data collection (Figure 2.9A, red boxes). Due to the fact that different tilting angles generate different defocus gradient along the direction perpendicular to the tilt axis, instead of measuring a single defocus value center at the tilt axis, the defocus gradient in each tilt image was estimated by sampling three different areas and is subsequently used for the

strip-wise CTF correction along the vertical tilt axis.

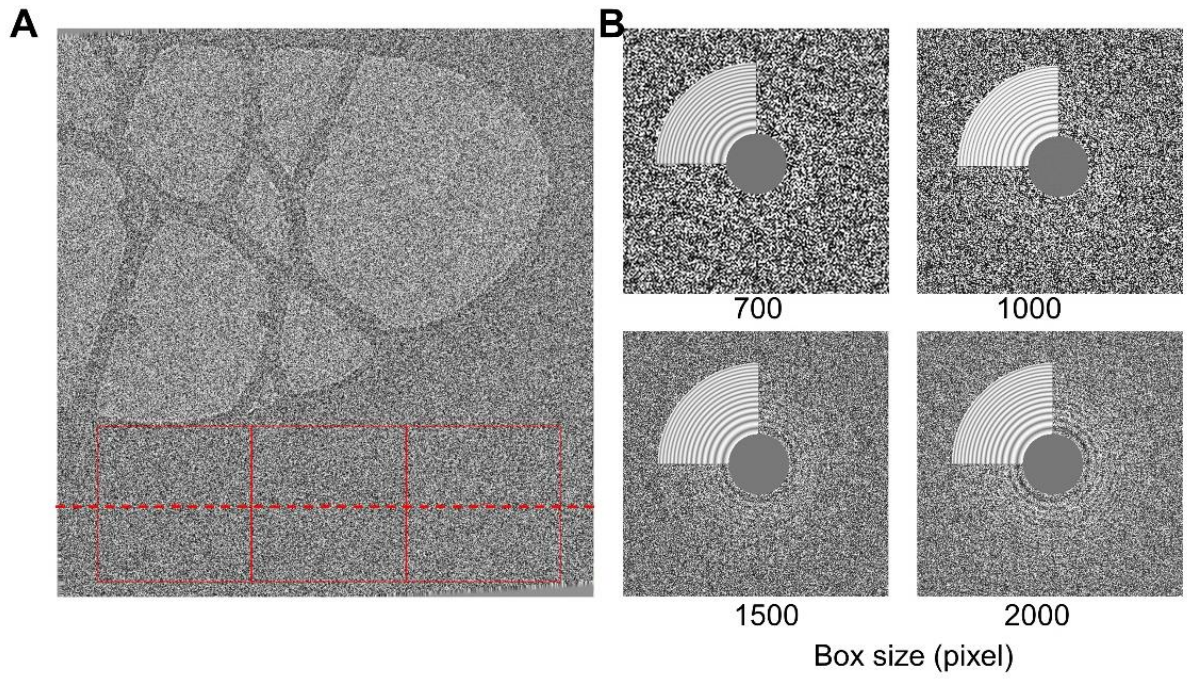


Figure 2.9 CTF determination for cryo-ET tilt series. A) A specific carbon area used to determine the defocus value of a sample. Three sampling areas (red boxes) were selected to calculate the defocus gradient along the red dashed line. B) The CTF fitting was determined from sampling the area with different box sizes of 700, 1000, 1500, 2000 pixels in Fourier space.

During the CTF correction of each image strip, two adjustable parameters, w1 and w2 (Fernandez et al., 2006) are used to defines the minimum value of CTF from which restoration by inverse filter is to be applied (w1) and the maximum value of CTF below which the standard Wiener filter is to be applied (w2). A value of 0.7 and 0.25 were used for w1 and w2 respectively, in order to generate corrected envelope as flat as possible (Figure 2.10). The denoise image used the same parameter determined from the corresponding raw image.

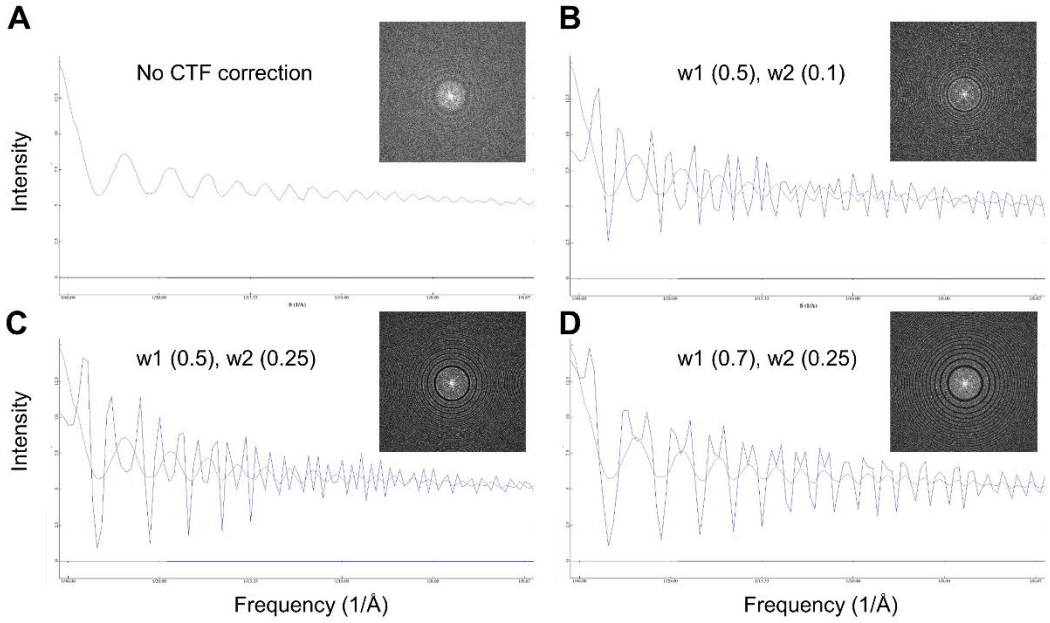


Figure 2.10 Optimization of CTF correction parameters for cryo-ET tilt series. Comparison between the original CTF curve (A) and the corrected CTF curves with different values of w1 and w2 (B though D) using a representative zero tilt image.

2.4 Improvement of the mechanical control of automated ET data collection

In addition to increase the image contrast, a stable and versatile data acquisition scheme, such as the "dose-symmetric tilt -scheme" (which begins at zero-tilt and then alternates between positive and negative tilt angles), as well as high-speed data collection, etc., have been proved to provide additional advantages for tomogram reconstruction. Therefore, tomographic data collection requires a precise mechanical control of the specimen goniometer under high magnification (approximately 80,000–160,000×). It is worth noting that during the whole process of data collection, the small off-tracking events are generally corrected by the "beam tilt/shift" function via adjusting the strength of the magnetic lens. The accumulation of beam tilt/shift used to compensate for the error from the mechanical control tends to downgrade the beam coherence. Accordingly, to maintain the beam coherence, a full mechanical control method for automating ET data acquisition without using beam tilt/shift will be an ideal stratagem. This method was developed by Liu et.al., which aims to minimize the error of the target object center during the tilting process through a closed-loop proportional-integral (PI) control algorithm (Liu et al., 2016). The method has been validated and

implemented on the data collection for the negative staining sample of nucleosome array data under 160,000× magnification (Figure 2.11).

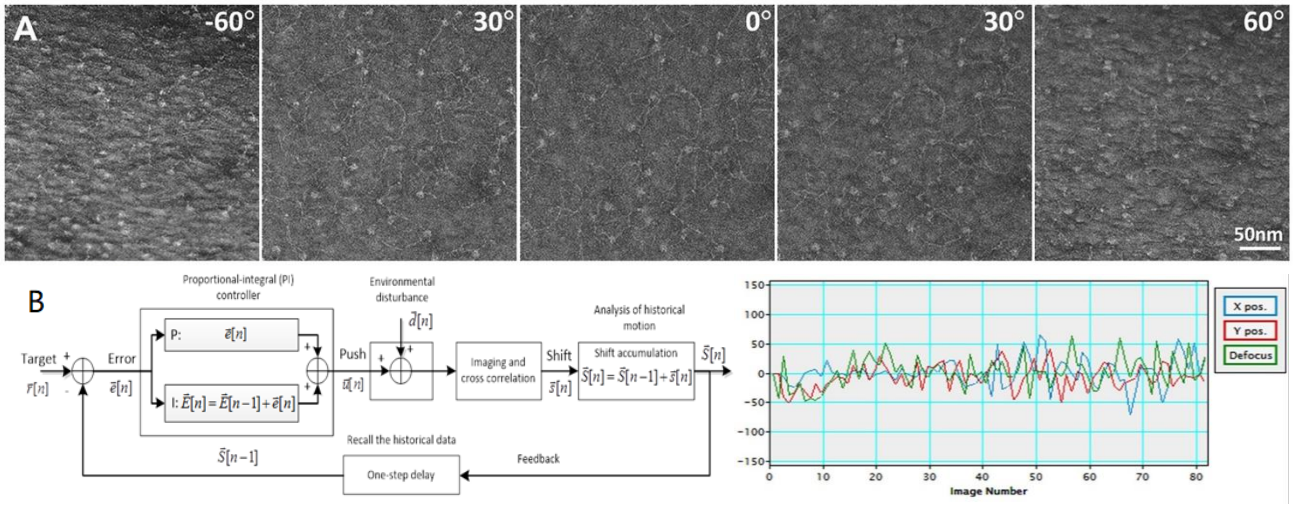


Figure 2.11 Validation of data collection stability on negative-stained nucleosome sample at 160,000x magnification. A) Images from a tilt series of a negatively stained nucleosome-DNA sample collected from −60° to +60° with 1.5° steps, at 160,000× magnification, and defocus of ~400 nm. B) Mechanical control feedback system design and tomographic data collection. This series took 1.5 h to collect, with X/Y mean absolute positional errors and a defocus error of 17.8, 17.6, and 21.6 nm, respectively, with standard deviations of 24.6, 21.1 and 26.9 nm, respectively.

2.5 Implementation of the missing wedge correction

The mechanical tilt limitation of tomographic data acquisition (Figure 2.12A) causes an incomplete spatial sampling in Fourier space (Figure 2.12B) that appears as an elongation and as an anisotropic resolution in a real-space 3D reconstruction (Figure 2.12C), which is called the "missing-wedge" artifact. To reduce the missing wedge artifact, we reported a post-processing method by filling the missing-wedge zone. This low-tilt tomographic reconstruction (LoTToR) method contains a model-free iteration process under a set of constraints in real and reciprocal spaces. A proof of concept is conducted by using the LoTToR on a phantom, i.e., a simulated 3D reconstruction from a low-tilt series of images, including that within a tilt range of ±45° (Zhai et al., 2020) (Figure 2.12C-E); the method was validated by using experimental data collected from mono-nucleosome negative staining samples (Figure 2.12F-L). A significantly reduced missing-wedge artifact verifies the capability of LoTToR and suggests a new tool to support the future studies of macromolecular dynamics, fluctuation and chemical

activity from the viewpoint of single-molecule 3D structure determination.

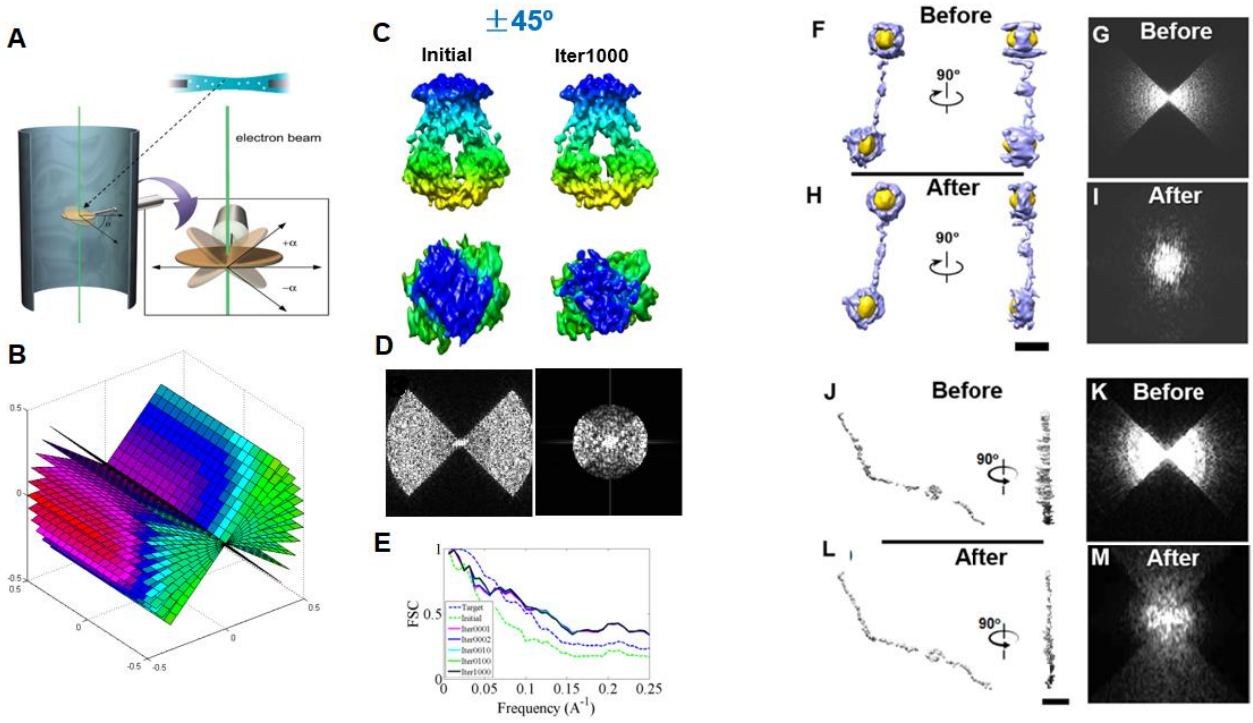


Figure 2.12 Missing-wedge correction of the NS single-molecule 3D reconstructions. A) Schematic representation of mechanical tilt limitations during ET data collection. B) The incomplete spatial sampling in Fourier space causing a "missing-wedge" zone. C) Two perpendicular views of the simulated density map for an individual protein particle (PDB id, 2ONK, −45° to +45° in steps of 1.5°) before and after 1,000 iterations of missing-wedge correction. D) Fourier transforms of X-Z projections of the initial and final 3D maps. E) Structural comparison of the corrected maps with the target crystal structure, evaluated by FSC. F–I) Missing-wedge correction on a 3D map of an individual complex of DNA-nanogold conjugate imaged by NS ET and reconstructed by IPET from a portion of the tilt series (from −45° to +45° in steps of 1.5°). Two perpendicular views and Fourier transforms of the X-Z projections of the 3D maps before and after correction are shown. J–M) Missing-wedge correction on a 3D map of an individual nucleosome particle (DNA-histone complex) imaged by NS ET and reconstructed by IPET from a portion of the tilt series (from −45° to +45° in steps of 1.5°). Two perpendicular views and Fourier transforms of X-Z projections of the 3D maps before and after correction are shown. Scale bars: 5 nm in A and 20 nm in H and L.

Chapter 3
Single-molecule 3D structures reveal nucleosome array dynamics and regulation

The following work is part of a manuscript in preparation: Zhang, M.; Díaz-Celis, C.; Liu, J.; Bustamante, C.; Ren, G. "Single-molecule 3D structures reveal nucleosome array dynamics and regulation".

3.1 Abstract

The subtle shift of microscopic equilibrium of nucleosome array unit arrangement forecasts the transition of higher-order chromatin structures, serving as regulatory mechanism for eukaryotic transcription. However, due to the structural flexibility of the nucleosome array, the mechanism of how their dynamic structures change in response to external stimuli remain elusive. Here we advanced the cryogenic electron tomography (cryo-ET) technique to determine single-molecule three-dimensional (3D) structures of mono-, di-, tri- and tetra-nucleosome array and the tetra-nucleosome with increased ionic strength and in the presence of linker histone H1. Through comparing ~40 structures of each type of nucleosome array, we found all arrays displayed an irregular arrangement instead of a symmetric structure. Quantitative analyses of the spatial orientation between the DNA linker and nucleosome core particle (NCP) showed that the change of the DNA linker angle against the NCP dictates the overall conformation of nucleosome arrays, which further governs the chromatin morphology. The finding that ~30% of foreign DNA binds to NCP in the presence of H1 suggests an alternative function of H1 in constructing nucleosomal networks. Insights into the dynamic change of the DNA linker angle against NCPs uncovers the molecular machinery that regulates eukaryotic gene expression.

3.2 Introduction

To elucidate how gene expressions occur in accordance with chromatin structural transitions, multiple models have been proposed to explain the conversion between the transcriptionally active "10-nm" chromatin fiber (Maeshima et al., 2014a) and its inactive condensed form (Strickfaden et al., 2020). The "10-nm" fiber consists of tandem repeats (Olins and Olins, 1974) of nucleosome core particles (NCPs) (Luger et al., 1997a) separated by 10-60 bp DNA linkers, while each NCP is composed of 147 bp of DNA wrapped ~1.7 turns around the histone octamer. The "10-nm" fiber has been observed to condense into various forms of "30-nm" fibers including one-start solenoid (Robinson et al., 2006; Finch and Klug, 1976b; Ohno et al., 2018), two-start helical ribbons (Woodcock et al., 1984b; Dorigo et al., 2004), untwisted two-start helix (Garcia-Saez et al., 2018), and crossed-linker (Williams et al., 1986). Therefore further hierarchical folding and intertwining of the "30-nm" fibers were historically considered as key intermediates of constructing the higher-order chromatin structures (Ou et al., 2017; Kireeva et al., 2004; Dehghani et al., 2005) and chromatids (Daban, 2015b, 2014b). However, the absence of the "30-nm" fiber *in vivo* (Eltsov et al., 2008; Maeshima et al., 2010a; Maeshima et al., 2016c; Liu et al., 2016) challenges this hypothesis. In contrast, a new model, liquid-liquid phase separation (LLPS) has been proposed to explain chromatin compartmentalization and its liquid-like condensation behavior both *in vitro* and *in vivo* (Gibson et al., 2019; Strom et al., 2017). Regardless of these models, the chromatin transition between its active and inactive form must be achieved through the conformational change of each array unit in response to stimulus (Farr et al., 2021a), such as ionic strength, temperature, and protein factors, including heterochromatin protein 1 (HP1) and linker histone H1 (Maeshima et al., 2016c; Gibson et al., 2019; Hihara et al., 2012; Sanulli et al., 2019). However, due to the enormous intrinsic structural dynamics and flexibility, insights into the "10-nm" fiber conformational changes require gathering enough numbers of unique array structures by sampling single molecule instances.

Recently, we extended the single-molecule reconstruction capability of cryo-electron tomography (cryo-ET) to capture 3D structural snap-shots of various conformations of nucleosome arrays that contain different numbers of NCP. Since tetra-nucleosome array is the basic structural unit (Ding et al., 2021) capable of forming higher-order chromatin structures including the helical fiber (Song et al., 2014; Schalch et al., 2005a) and phase separated condensates (Zhang et al., 2021), it has been used as our prototype to interrogate array conformational changes under different ionic strength and in the presence of H1. The observed changes in the spatial orientation between DNA linkers and NCP defines the "10-nm fiber" structure and dynamics, which regulate the large-scale chromatin morphology, density, and therefore activity.

3.3 Results

3.3.1 Imaging of a flexible nucleosome array particle

To snap-shot 3D structures of an individual nucleosome array requires evidencing the capability to directly image each NCP and linker DNA within a single array particle. Due to the small physical diameter and flexibility, 3D reconstruction of a single DNA (without averaging) was very challenging (Marini et al., 2015; Irobalieva et al., 2015; Beel et al., 2021). Here, we used both cryo-ET (Figure 3.1) and negative-stain ET (NS-ET) (Zhang et al., 2010; Rames et al., 2014a; Zhang et al., 2011) (Figure 3.2 and 3.3) to test our local imaging capability of DNA and NCP within mono-nucleosome arrays assembled from different lengths of DNA templates (227-bp, 447-bp, 598-bp and 646-bp). In these arrays, the entry DNA arm was designed longer than the exit DNA arm. Survey micrographs and representative particle images showed a characteristic "bead-on-string" structure for the mono-nucleosome array where i) the discoidal-shaped NCP and DNA arms were measured with a dimension of ~10 × 10 × 4 nm (Figure 3.1a and b) and width of ~2 nm (Figure 3.2), respectively; ii) the DNA arm displayed with ~2 nm helical pitch corresponding to its major groove (Figure 3.3a-d); iii) near 2 turns of DNA wrapped around the central histone octamer (Figure 3.1c and Figure 3.3e-h); and vi) a low-density pore (~1 nm) presented near the center of the nucleosomal disk (Figure 3.1c, and Figure 3.3a-d). These features were consistent with the NCP crystal structure (PDB entry 1AOI) (Luger et al., 1997a) and matched our designed DNA arm lengths, thus evidencing our imaging capability on nucleosomes as the basis for the 3D reconstructions as follows.

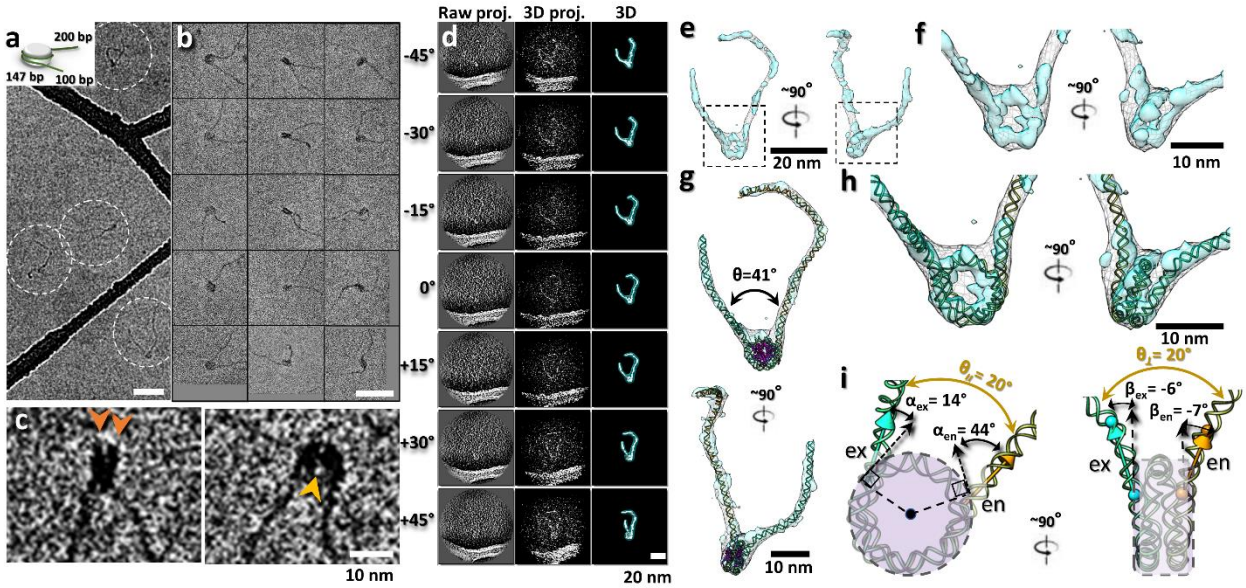


Figure 3.1 Cryo-EM images and 3D reconstruction of an individual mono-nucleosome particle. a) Survey cryo-EM micrograph of reconstituted mono-nucleosomes designed

with 200-bp entry and 100-bp exit DNA arms. b) Zoomed-in phase plate images of representative mono-nucleosome particles with various conformations. C) Zoomed-in phase plate image of NCP regions of two representative particles. Structural features, including near two turns of wrapped DNA and the low-density center of the NCP, are indicated by orange and yellow arrows, respectively. d) 3D reconstruction process of an individual particle. Seven representative projections before (left column) and after refinement (middle column) are compared to the final 3D density map (right column) at corresponding tilting angles. e) Two perpendicular views of a full particle 3D reconstruction displayed with two different contour levels. f) Representative NCP 3D reconstructions containing the same detailed feature as shown in c. g) Density map superimposed on its flexibly fitted model. θ corresponds to the angle between the two DNA arms. h) Zoomed-in view of the fitted NCP model showing the projection of its θ angle onto planes parallel or perpendicular to the NCP discoidal plane. i) The orientation of linker DNA against the NCP was measured as the wrapping angle α, and the bending angle β, for both the entry and exit DNA arms.

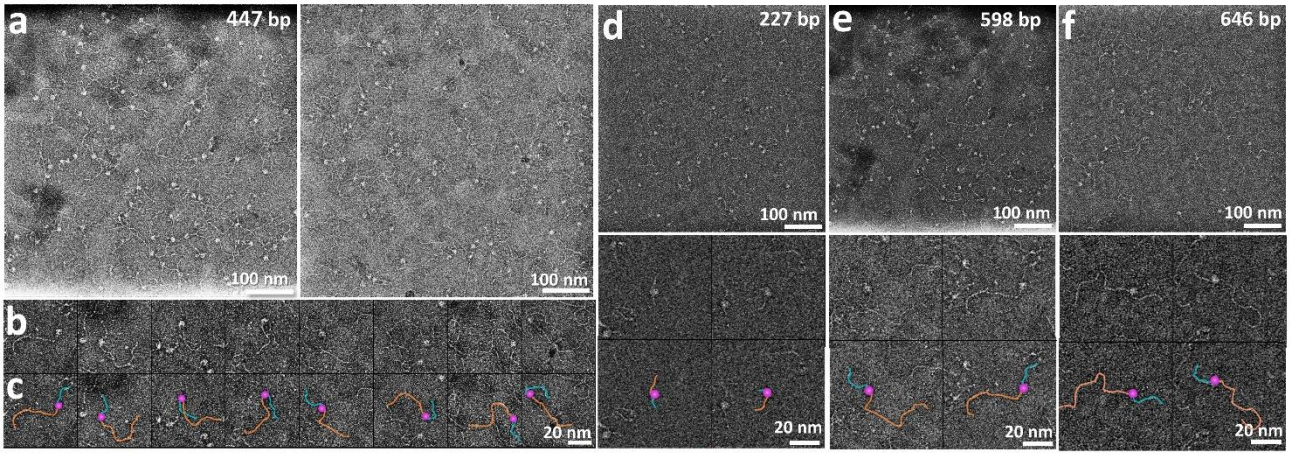


Figure 3.2 Morphology of mono-nucleosomes imaged by NS EM. a) Two representative micrographs of mono-nucleosomes, which were reconstituted by assembling histone octamers on a 447-bp DNA template with 200 bp and 100 bp entry and exit flanking DNA arms, respectively. b) Eight representative particles, and c) their corresponding morphologies. The entry and exit DNA arms and nucleosome core particles (NCPs) are indicated in orange, cyan and purple colors, respectively. d, e and f) Representative micrographs of reconstituted mono-nucleosomes with various lengths of DNA templates (227 bp, 598 bp and 646 bp). Two representative particles with corresponding morphologies are presented at the bottom.

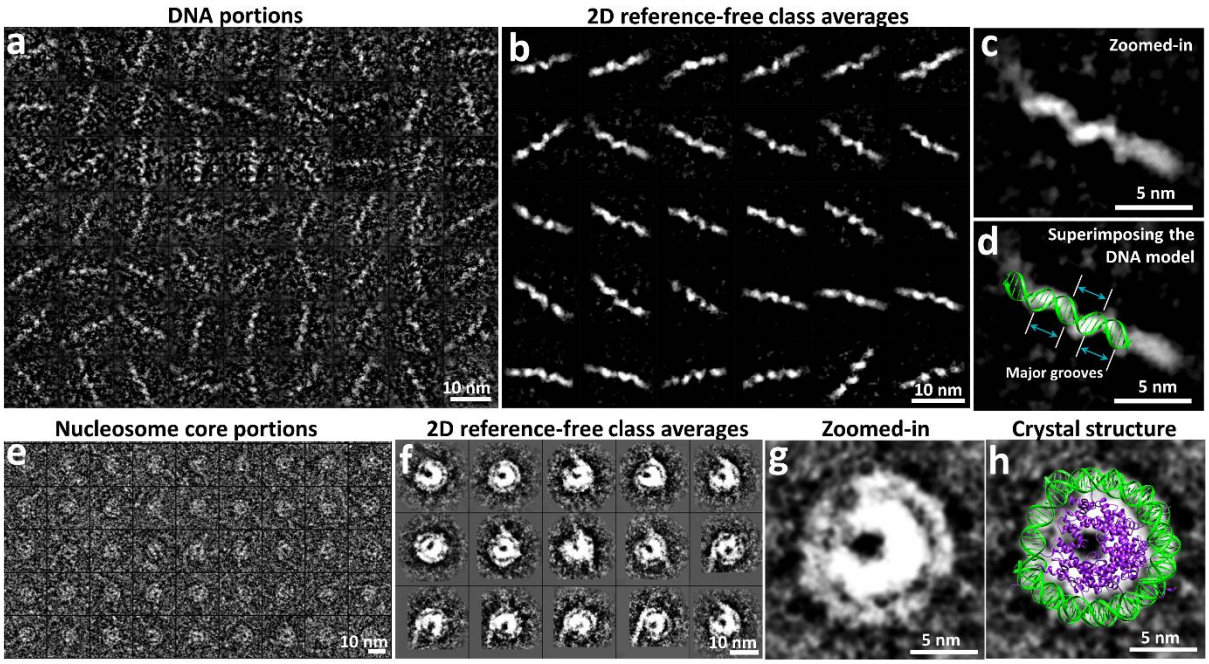


Figure 3.3 NS EM images and class averages of NCPs and DNA arm segments for 447-bp DNA mono-nucleosome. a) 63 representative images of segments of DNA arms. b) 20 representative reference-free class averages of DNA segments. c) Zoomed-in image of a representative average. d) The average super-imposed on a standard structure of DNA in which the major groove is indicated by cyan arrows. e) 45 representative images of NCPs. f) 15 representative reference-free class averages of NCPs, g) Zoomed-in image of a representative class average, h) Super-imposition of the image in g on the crystal structure of the NCP (PDB entry: 1AOI)(Luger et al., 1997a).

3.3.2 3D structure of an individual mono-nucleosome array

To achieve a snap-shot 3D structure from an individual array particle, the mono-nucleosomes with 447-bp DNA were *ab initio* reconstructed by both cryo-ET (Figure 3.1d) and NS-ET (Figure 3.4). Using the method of individual-particle electron tomography (IPET) (Zhang and Ren, 2012) combined with image denoising (Buchholz et al., 2019) and low-tilt tomographic reconstruction (LoTToR) (Zhai et al., 2020) techniques, the 3D reconstructions at resolutions up to 26 Å (Figure 3.5 and 3.6) confirmed the "bead-on-string" structure, and provided sufficient structural details, such as the discoidal shape of NCP, the approximately two turns of DNA wrapping with defined entry and exit positions, as well as a measurable DNA arm length (Figure 3.1e and f). These features allow us to obtain a structural model from each particle by first flexibly fitting two strand of DNA models into arms' density and rigid-body docking the NCP crystal structure

into the discoidal density, and then followed by merging their overlapped DNA portion under a chemical structure constrain by energy minimization (Phillips et al., 2020) (Figure 3.1g and h). Considering the models from NS-ET 3D reconstruction were flattened due to the constraint of the thin layer of stain, only the cryo-EM models were structurally characterized.

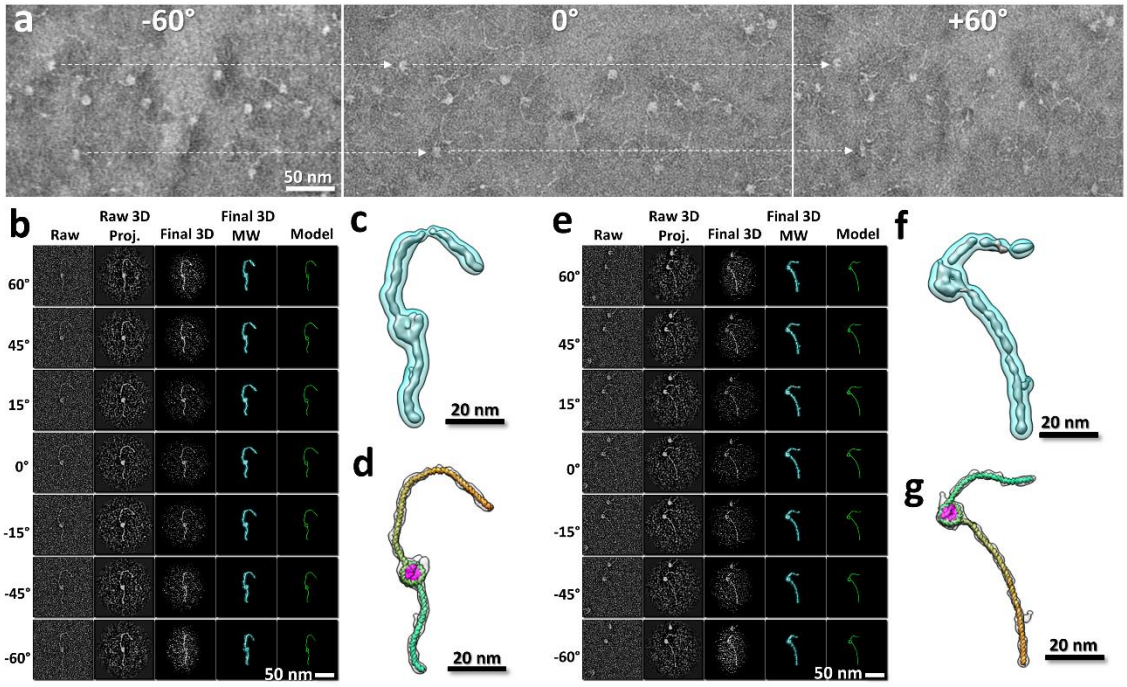


Figure 3.4 3D reconstruction of an individual particle of 447-bp DNA mono-nucleosome by IPET. a) Three images from a tilt series of NS sample imaged by ET. The arrows indicate the same two particles imaged from different tilt angles. b) IPET 3D reconstruction of a NCP. Seven representative tilt images of an individual particle displayed after CTF correction (first column). By aligning tilt images to a common center for 3D reconstruction via an iterative refinement, the raw 3D projections and reconstruction (before particle shaped masking, second and third columns), the final 3D reconstruction with missing wedge correction (after particle shaped masking, fourth column) and the flexibly fitted model (fifth column) are displayed at the corresponding tilt angles. c) Zoomed-in image of the final 3D density map (low-pass filtered to 35 Å) displayed at two contour levels using Chimera (with 'hide dust' applied). d) High contour level density map (displayed in gray color) superimposed on a flexibly fitted model. The histone entry and exit DNA arms are colored in pink, orange and cyan, respectively. e-g) IPET 3D reconstruction of another individual particle with its fitted model.

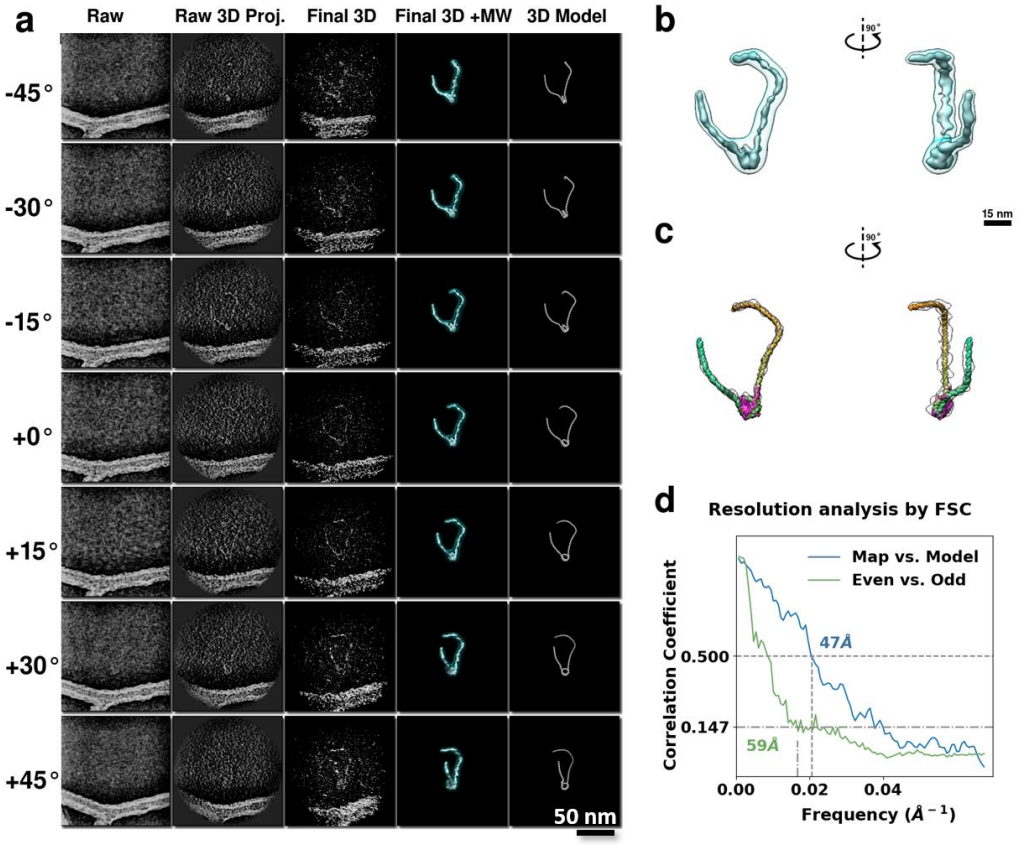


Figure 3.5 Cryo-ET 3D reconstruction of an individual mono-nucleosome particle by IPET. a) IPET 3D reconstruction of a representative NCP particle. Seven representative tilt images of an individual particle after CTF correction are shown (first column). By aligning tilt images to a common center for 3D reconstruction via an iterative refinement, the raw 3D projections and reconstruction (before particle-shaped masking, second and third columns), the final 3D reconstruction with missing wedge correction (after particle shaped masking, fourth column) and the flexibly fitted model (fifth column) are displayed at the corresponding tilt angles. b) Zoomed-in final 3D density map (shown at two contour levels in two orthogonal views), c) The corresponding high contour level map superimposed on the flexibly fitted model. d) Resolution evaluation of the final 3D density map by two criteria, *i.e.* Fourier shell correlation (FSC) between maps independently reconstructed from the even and odd index of the tilted series, and FSC between the final 3D map against the fitted structure model. The resolution for the both criteria were evaluated at two frequencies, where the FSC curve fell to 0.5 (Liao and Frank, 2010) and 0.147 (Rosenthal and Henderson, 2003).

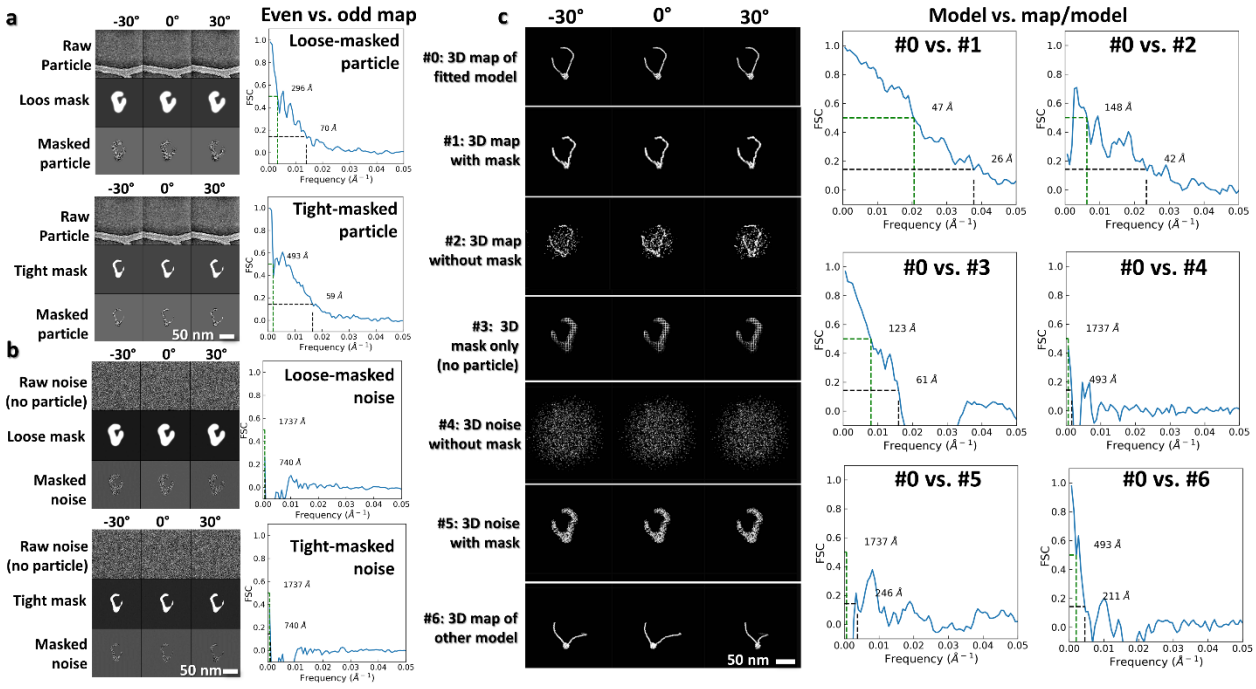


Figure 3.6 Analyses of cryo-ET 3D reconstruction resolution. a) Resolution comparison showing that the reconstructions were not sensitive to the size of mask used. A loose (top panel) and tight (bottom panel) particle-shaped mask were applied on the same aligned images before using the even and odd index of tilt series for 3D reconstructions and resolution estimation. b) Resolution comparison showing that masks alone were not the source of high resolution. The same masks were applied on noise images for 3D reconstructions and resolution estimation. c) Resolution evaluation by comparing the FSCs between the fitted model density (#0) with different maps (#1-6), including the final 3D reconstruction after masking (#1), the final 3D reconstruction before masking (#2), the 3D map of the mask (#3), the 3D reconstruction from the noise images (#4), the 3D reconstruction of the masked noise images (#5), and the 3D density map of another particle's fitted model (#6). #1,3,5 used the same tight particle-shaped mask as in a.

To quantitatively characterize this cryo-ET 3D structure, we measured the spatial orientation of the DNA against the NCP from the following four aspects: i) Wrapping length: the length of DNA wrapped around the histone core within the NCP (Figure 3.1h); ii) Arm/linker angle θ: the angle formed between two DNA arm segments extended 20-bp from the NCP surface (as indicated by two vectors in Figure 3.1g), and iii) the two orthogonal projections of the angle θ, *i.e.* $\boldsymbol{\theta}_{\parallel}$ and $\boldsymbol{\theta}_{\perp}$, against the discoidal plane of the NCP (Figure 3.1i); iv) Wrapping angle α and bending angle β: the angle of each DNA arm vector (measured from 20-bp DNA segment) relative to the NCP orientation (Figure 3.1i). The measurement of these parameters provides a basis for characterizing the dynamics of the mono-nucleosome 3D structure as follows.

3.3.3 Dynamics of mono-nucleosome 3D structure

To illuminate the dynamics of the mono-nucleosome 3D structure, a total of 47 density maps were reconstructed by cryo-ET (Figure 3.7 and Extended Data Video 1), followed by modelling and structural characterization as described above (Figure 3.7b). The statistical analysis showed that the lengths of DNA wrapping to the histone octamer are varied. Though a similar event has been observed within few distinct classes from cryo-EM single-particle averaging analysis (Bilokapic et al., 2018a), the individual particle ET reconstruction revealed continuous unwrapping states, where the entry and exit side DNA were unwrapped for 5 ± 5-bp (mean vs. standard derivation) and 11 ± 13-bp, respectively (Figure 3.7c and d) compared to the crystal structure (PDB ID: 1AOI (Luger et al., 1997a)), which agrees well with the MD simulation (Winogradoff and Aksimentiev, 2019). This unwrapping dynamics is related to the reported nucleosome "breathing" (Polach and Widom, 1995). Our observation that the NCP exit side DNA is more dynamic than its entry side in terms of unwrapping is consistent with the results from single-molecule fluorescence resonance energy transfer (FRET) study (Ngo et al., 2015), which attribute the cause of such difference to the asymmetric sequence of 601 DNA template (the exit side DNA is less flexible leading to a decreased tolerance in binding with the histone octamer).

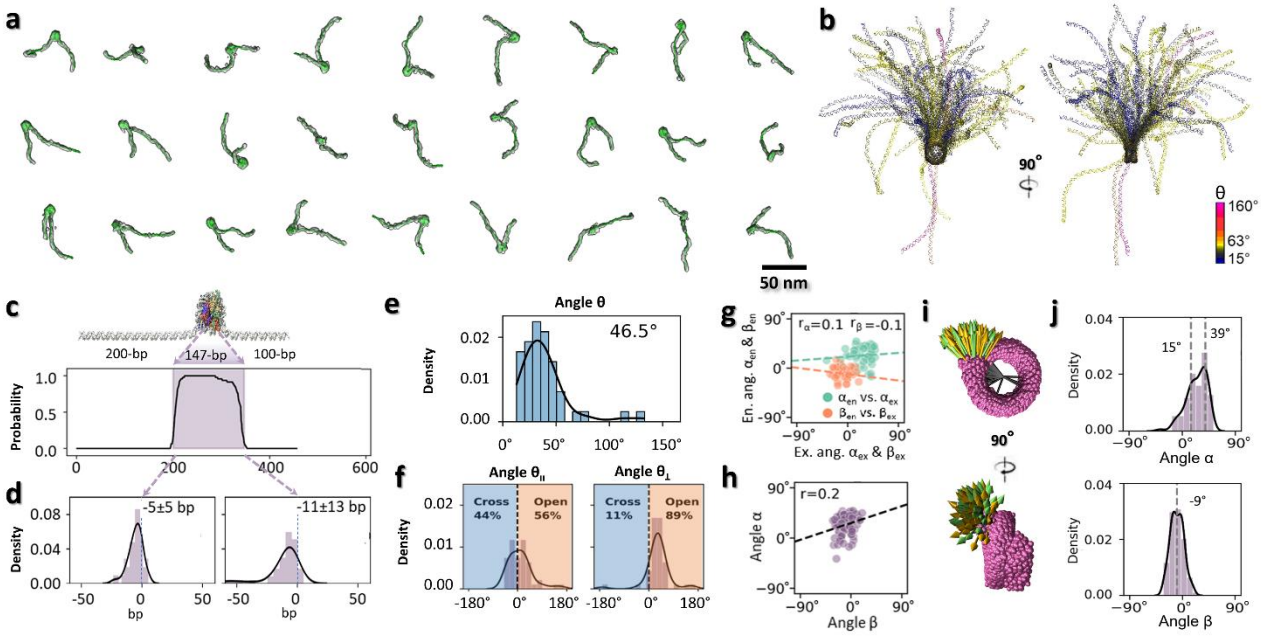


Figure 3.7 3D structure and dynamics of mono-nucleosome particles. a) 30 representative density maps reconstructed by cryo-ET from targeted mono-nucleosome

particles shown with fitted models. b) 47 super-imposed models after alignment based on their NCP portion, shown in three orthogonal views. Models color-encoded by the angle θ between their two DNA arms. c) Histone octamer positioning along the 447 bp DNA template represent as the probability of finding each DNA base pair in contact with the octamer surface measured from 47 models. The designed 601-region on the DNA template is highlighted by a filled magenta box. d) Histograms of the measured entry and exit DNA unwrapping position that deviate from the design. e) Histograms of the angle θ and f) its two planer projections $\boldsymbol{\theta}_{\parallel}$ and $\boldsymbol{\theta}_{\perp}$. g) Scatter plot showing weak correlation between the entry and exit DNA arms on the same NCP, as reflected by the small r-value between α_{en} *and* α_{ex} and between β_{en} and β_{ex}. h) Scatter plot showing weak correlation between wrapping angle α and bending angle β measured from the same side of the NCP. i) Super-imposed vectors of the entry and exit DNA linkers shown in orthogonal views. j) Histogram of the wrapping angle α and the bending angle β distribution measured from both sides of each NCP, respectively. Scale bar is 50 nm.

Besides, quantitative measurement was performed to characterize the relationship among DNA arms/linkers and NCP. The θ angle is distributed in a wide range from 0° to 140° (with a peak population at ~39°, and a mean ± std of 46° ± 27°), suggesting the two DNA arms were relatively dynamic against each other (Figure 3.7e). The distributions of $\boldsymbol{\theta}_{\parallel}$ and $\boldsymbol{\theta}_{\perp}$ showed their peak populations at ~9° (10° ± 44°) and ~32° (32° ± 45°), respectively suggesting the two DNA arms were relative parallel along NCP discoidal plane, but modestly bent away from NCP toward the plane's normal direction (Figure 3.1). Considering a similar distribution and weak correlation were observed between the entry and exit α angles (Figure 3.7g), the α_{en} *and* α_{ex} were merged as a single α. The same result goes for the β angles, suggesting the two DNA arms can move independently within each of their "territory". Also, a weak correlation was found between the α and β angles, implying that there is no preferential "swing trajectory" of linker DNA on the NCP surface (Figure 3.7h and i). The α angles displayed a left-skewed distribution in a range of ~± 50° with major and minor peaks at ~39° (39° ± 5°) and ~15° (14° ± 16°) respectively (Figure 3.7j), suggesting two possible low-energy states of the DNA linker along the NCP plane. The β analysis showed a near normal distribution in a range of ~± 30° with a peak population at ~ -9° (-9° ± 11°) (Figure 3.7j), suggesting a smaller dynamic of the DNA linker along the normal direction of the NCP plane.

To vividly display the thermodynamics of the mono-nucleosome array, above 47 snapshot structures obtained in solution were pair-wisely aligned by minimizing their root-mean-square deviations (RMSD), followed by sorting their structural similarity via hierarchical classification (Pedregosa et al., 2011) (Figure 3.8a). The intermediate morphing between two consecutive ordered structures were interpolated by target

molecular dynamics (TMD) simulations, which displayed the quasi-thermodynamics transition of mono-nucleosome array in solution from the most populated states to the rarest states (Extended Data Video 2). The above analyses evidenced our capability in characterizing the nucleosome array dynamics based on the cryo-ET 3D reconstruction of each individual nucleosome particle, as a proof-of-concept for following studies on the dynamic changes of nucleosome arrays under different conditions.

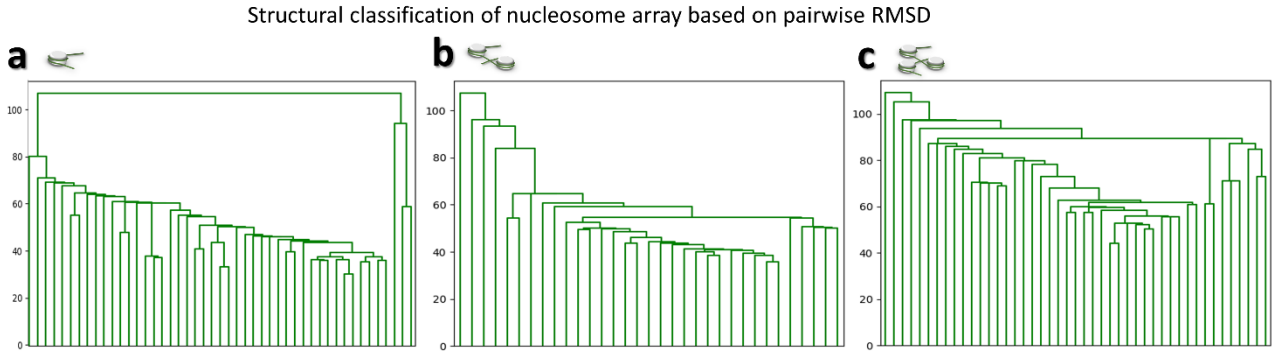


Figure 3.8 Hierarchical clustering analysis of the array structures. Dendrograms of the a) mono-, b) di-, and c) tri-nucleosome arrays were constructed based on hierarchical classification of the RMSD value between each pair of array structures. Array pairs with smaller RMSD (i.e. similar structure) are shown at lower branches of the tree. The morphing among nucleosome array structures followed the order of dendrograms starting from the lowest branch.

3.3.4 NCP number effects on the dynamics of nucleosome array 3D structure

To investigate whether nucleosome arrays containing more than one NCP share the same intrinsic structural dynamic as mono-nucleosomes, we repeated the above cryo-ET experiment by using multi-nucleosome constructs, including di-, tri- and tetra-nucleosome. These arrays were designed in a similar configuration as mono-nucleosomes with 200-bp entry and 100-bp exit distal DNA arms, except that 40-bp DNA linkers were inserted between NCPs. We reconstructed a total of 33, 45 and 31 density maps of the di-, tri-, and tetra-nucleosomes, respectively, followed by modelling and quantitative analysis as above (Figure 3.9 through 3.11 and Extended Data Video 3 through 5). These multi-nucleosome array structures displayed an overall asymmetric zig-zag architecture (Figure 3.12a and b), which is consistent with the described nucleosome fiber conformations found within oligo-nucleosome sample (Grigoryev et al., 1999), the cell nucleus sections (Cai et al., 2018a; Ng and Gan, 2020) and decondensed mitotic chromosomes (Beel et al., 2021), but was controversial to the symmetric structure, such as the twisted double-helix revealed by cryo-EM single-particle

averaging method (Song et al., 2014) and crystallography (Schalch et al., 2005b).

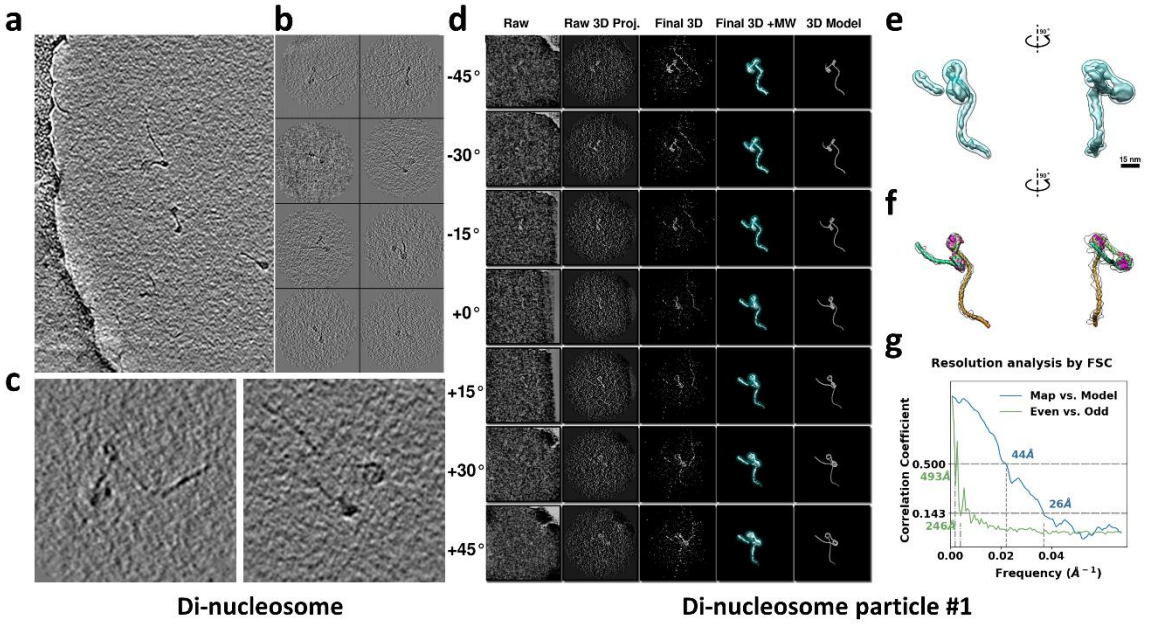


Figure 3.9 Cryo-EM images and 3D reconstruction of an individual di-nucleosome particle. a) Representative 2D slice (5nm thickness) of a reconstructed cryo-EM tilt series of di-nucleosome arrays designed with 200-bp entry and 100-bp exit DNA arms and linker length between nucleosomes of 40 bp. b) Selected representative di-nucleosome array particles of various conformations. c) Zoomed-in image two representative particles. d) 3D reconstruction process of an individual particle. Seven representative projections before (left column) and after refinement (middle column) are compared to the final 3D density map (right column) at corresponding tilting angles. e) Orthogonal views of a full particle 3D reconstruction displayed at two contour levels. f) Fitted model and map of the particle. The DNA is colored as a gradient from its entry side (orange) to its exit side (cyan). The histones are colored in pink. g) Resolution estimation of the reconstruction by using Fourier Shell Correlation (FSC).

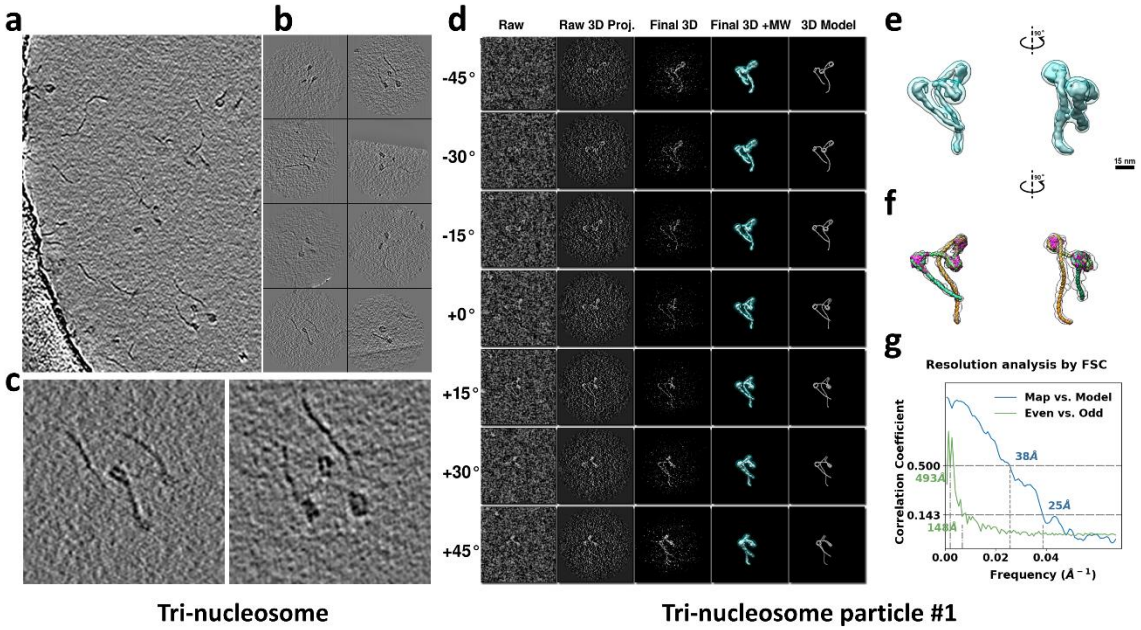


Figure 3.10 Cryo-EM images and 3D reconstruction of an individual tri-nucleosome particle. a) Representative 2D slice (5nm thickness) of the reconstructed cryo-tomogram a of tri-nucleosome array designed with 200-bp entry and 100-bp exit DNA arms and linker length between nucleosomes of 40 bp. b) Selected representative tri-nucleosome array particles of various conformations. c) Zoomed-in image of two representative particles. d) 3D reconstruction process of an individual particle. Seven representative projections before (left column) and after refinement (middle column) are compared to the final 3D density map (right column) at the corresponding tilting angles. e) Orthogonal views of a tri-nucleosome particle 3D reconstruction displayed with two contour levels, f) Fitted model and map of the particle. The DNA is colored as a gradient from its entry side (orange) to its exit side (cyan). The histones are colored in pink. g) Resolution estimation of the reconstruction by using the Fourier Shell Correlation (FSC) coefficient.

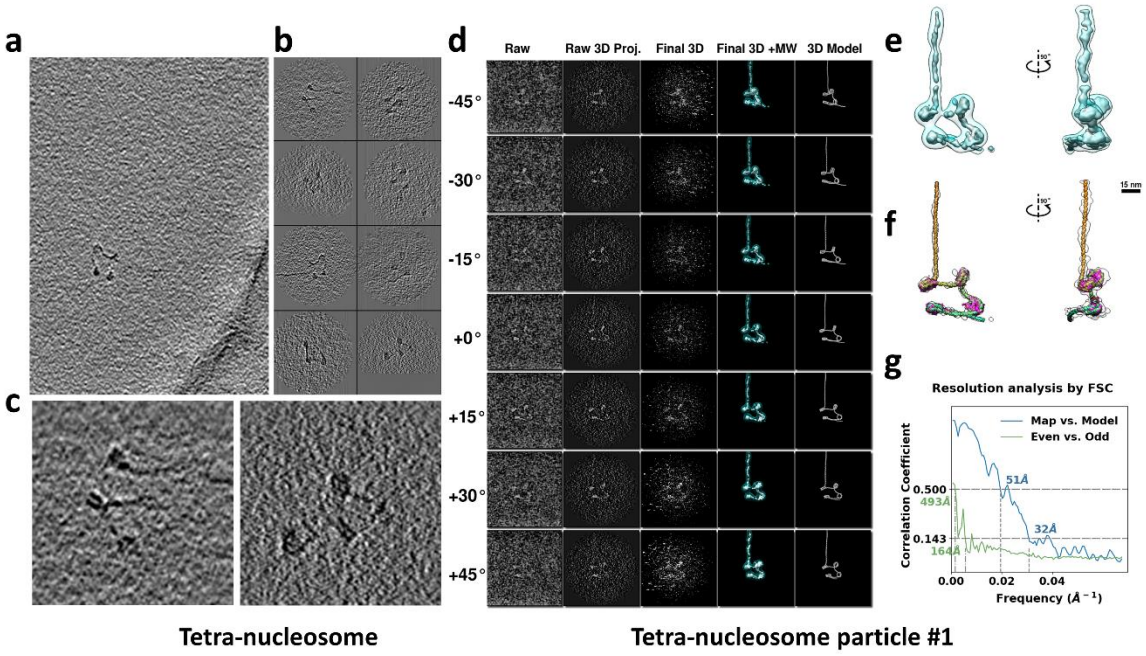


Figure 3.11 Cryo-EM images and 3D reconstruction of an individual tetra-nucleosome particle. a) Representative 2D slice (5nm thickness) of the reconstructed cryo-tomogram a of tetra-nucleosome array designed with 200-bp entry and 100-bp exit DNA arms and linker length between nucleosomes of 40 bp. b) Selected representative tetra-nucleosome array particles of various conformations. c) Zoomed-in image of two representative particles. d) 3D reconstruction process of an individual particle. Seven representative projections before (left column) and after refinement (middle column) are compared to the final 3D density map (right column) at the corresponding tilting angles. e) Orthogonal views of a tetra-nucleosome particle 3D reconstruction displayed with two contour levels, f) Fitted model and map of the particle. The DNA is colored as a gradient from its entry side (orange) to its exit side (cyan). The histones are colored in pink. g) Resolution estimation of the reconstruction by using the Fourier Shell Correlation (FSC) coefficient.

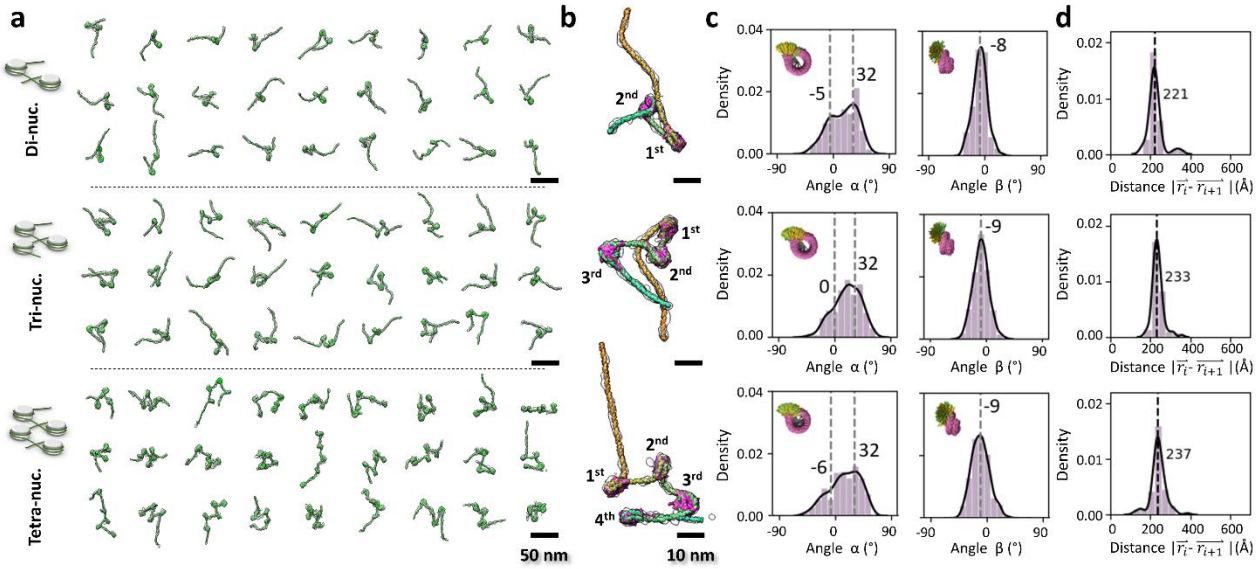


Figure 3.12 3D structures illustrating the dynamics of di-, tri- and tetra-nucleosome array particles. a) 30 representative cryo-ET density maps from individual particle reconstructions of di-, tri-, tetra-nucleosomes. Each of the density maps was super-imposed with its flexibly fitted model. b) Zoomed-in image of one representative map with fitted model. c) Histograms of the wrapping angle α and the bending angles β. d) Histogram of the core-to-core distance measured between I, I+1 NCP. Scale bars are 50 nm in a and 10 nm in b.

Quantitative characterization confirmed the average DNA wrapping lengths on the NCP entry-sides of di-, tri-, and tetra-nucleosome were shorter than that of crystal structure by -4 ± 5-bp, -7 ± 7-bp, and -4 ± 8-bp, respectively, and that wrapping lengths on the exit-side were shorter by -8 ± 10-bp, -10 ± 11-bp, and -14 ± 12-bp, respectively (Figure 3.13). The similar unwrapping levels comparing to the mono-nucleosome (-5 ± 5-bp and -11 ± 13-bp for the entry exit side, respectively) indicate an intrinsic similarity of structural dynamics corresponding to the asymmetric breathing motion of nucleosomes (Winogradoff and Aksimentiev, 2019; Bilokapic et al., 2018a; Ngo et al., 2015).

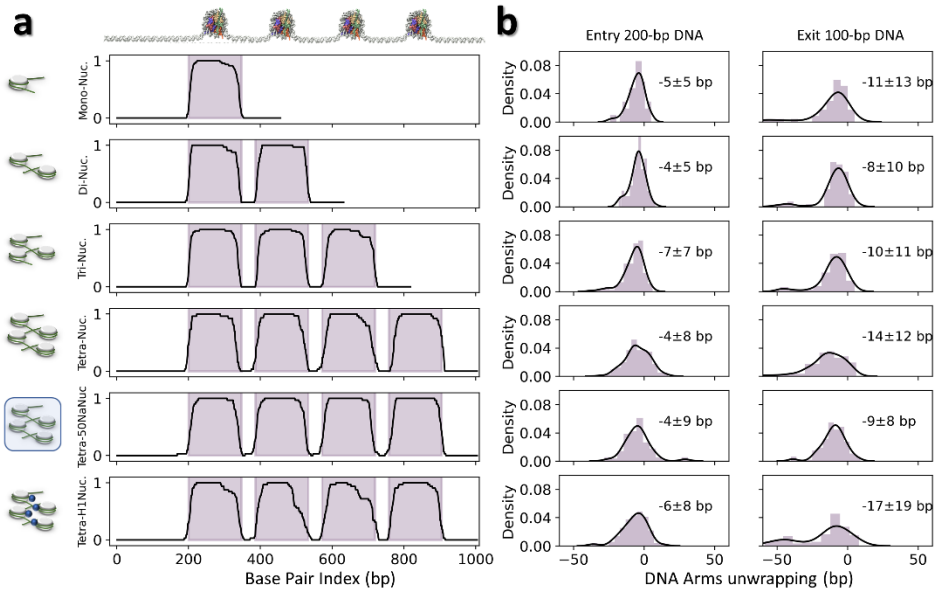


Figure 3.13 Analysis of wrapping DNA distribution and DNA arm stiffness. a) Probability of finding a base-pair in contact with the histone octamer along the template DNA. The designed 601-regions on the DNA template are highlighted by filled magenta boxes. b) Histograms of the NCP unwrapping length measured from the entry and exit sides of the NCP. c) Estimation of the entry and exit DNA arms' persistence length by measuring the autocorrelation decay along the DNA segments fitted with an exponential decay function.

Surprisingly, analysis of the θ angle among multi-nucleosome arrays (Figure 3.14f) showed that the peak of the di-nucleosome (~55°) was the most distinguished from the mono-nucleosome (~39°). While further increasing the array NCP number diminishes this peak difference to the mono-nucleosome, as it is shown for tri- (~47°) and tetra-nucleosome (~41°). In addition, analysis of $\theta_{\parallel}$ angle magnified above θ peak separation in a wider-angle range of ~33° (from -40° to -21° and -7° for di-, tri, and tetra, respectively), while the $\theta_{\perp}$ angle distributions showed their peaks within a narrower-angle range of ~3° (from -38° to -37° and -35°, respectively) (Figure 3.14f). This non-monotonic peak shifting clued additional factors that control the arm angle of NCPs, such as an "edge effect." The NCPs at two distal ends of the nucleosome array showed different $\theta_{\parallel}$ angle distributions compared to the middle NCPs (Figure 3.14g). Di-nucleosome array contained 100% (*i.e.* 2 of 2) of edging NCPs (distal-end NCPs), while tri- and tetra-nucleosome array only contained ~67% (*i.e.* 2 of 3 NCPs) and 50% (*i.e.* 2 of 4 NCPs) of edging NCPs. The 100% edging NCPs in the di-nucleosome induced a special

conformation for its two distal arms arrangement, yielding a noticeable smaller dynamic space (Figure 3.15) and causing ~83% of DNA linkers to adopt a closed-arm conformation (Figure 3.14f). This unique conformation of the di-nucleosome may be responsible for its in vivo recognition, such as its binding to the gene expression repressor, Isw1a (Yamada et al., 2011). A higher binding affinity of Isw1a to the di-nucleosome can be achieved by increasing the distal DNA arm length regardless of nucleosome histone modification (Bhardwaj et al., 2020). The α angles for multi-nucleosome arrays showed a negatively skewed distribution with two peaks, i.e. a major peak at ~32° similar to that of mono-nucleosomes (39°), and a minor peak at an angle between −6° and 0° that was distinguished from mono-nucleosomes (15°). On the other hand, the β angle displayed a similar distribution among multi-nucleosomes (peak at −9°-−8°, Figure 3.12c). The minor peak of α changing from positive to negative values along with the increment of NCP number indicated that the open-arm conformation in mono-nucleosome can be regulated into a close-arm conformation when the two repulsive (negatively charged) DNA arms are partially being neutralized in binding with other histone octamers (Gebala et al., 2019; Manning, 2003).

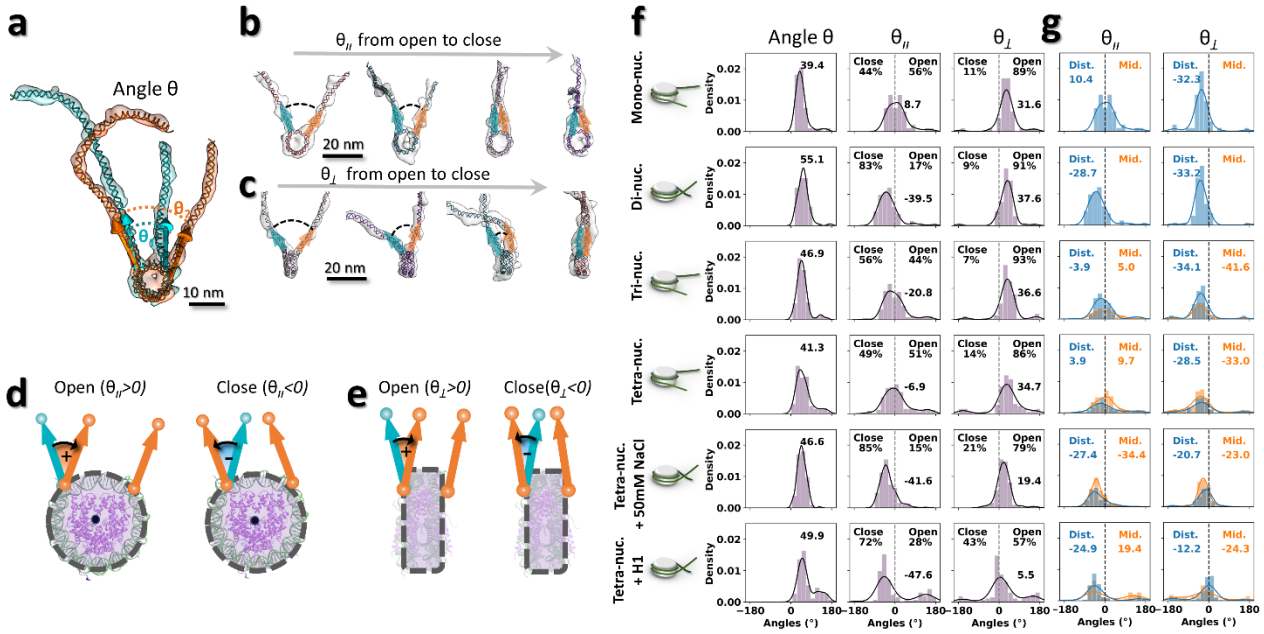


Figure 3.14 Distribution of the angle θ formed by the entry and exit DNA arms. a) Two representative 3D reconstructions showing the angle θ formed between their two DNA arms exiting the NCPs. b and c) Four representative 3D reconstructions showing the DNA arms dynamics in orthogonal views (planes parallel and perpendicular to the NCP discoidal plane, respectively). θ∥ and θ⊥ were defined as the projections of the θ angle onto the corresponding planes. d) Definition of the open-closed arm conformation

of an NCP by the sign of its θ∥ fraction. E) Definition of the open-closed arm conformation of an NCP by the sign of its θ⊥ fraction. f) Distribution of angle θ, θ∥ and θ⊥ measured from all NCP units of different nucleosome arrays. g) Differential distribution of angle θ∥ and θ⊥ from distal NCPs (blue color) and intermediate NCPs (orange color) within the array. Schematics on the left column showing the preferred DNA arm conformations for the NCPs.

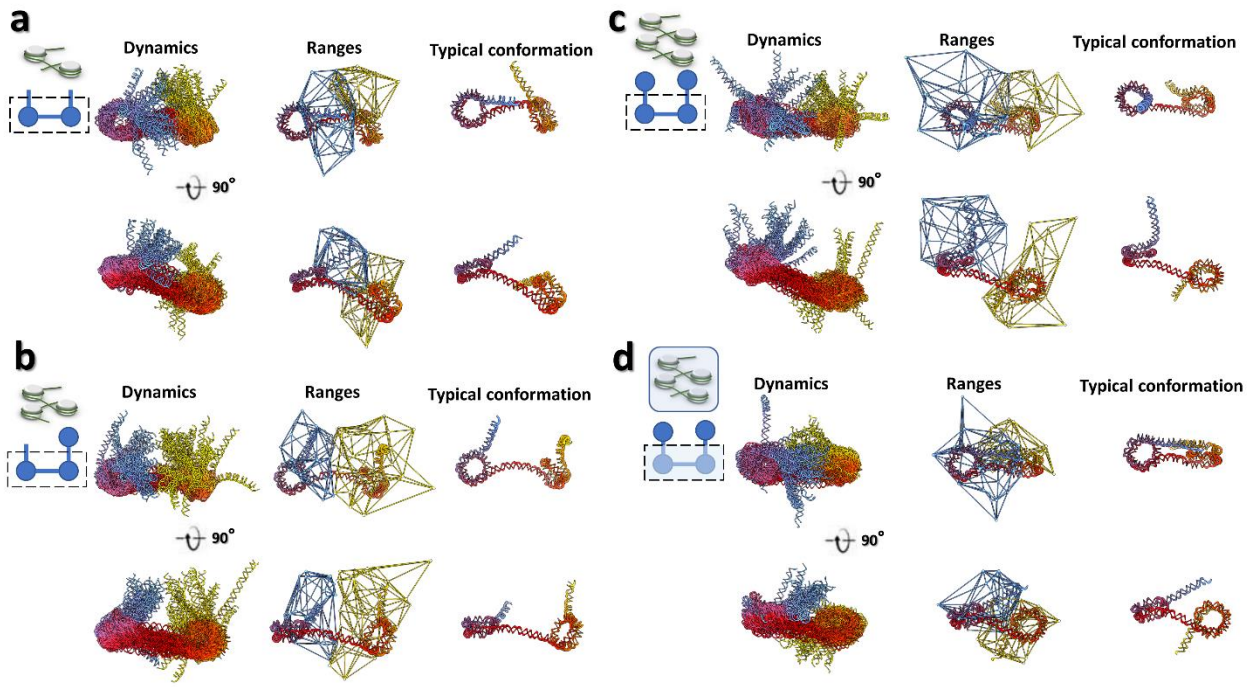


Figure 3.15 Mapping of the distal DNA linker dynamics space and typical conformations of di-nucleosome components from di-, tri-, and tetra-nucleosome arrays. a, b, c, and d) Di-nucleosome components (40-147-40-147-40) were extracted from di-, tri, tetra-nucleosome arrays at 5 mM Na^+ and tetra-nucleosome arrays observed in 50 mM Na^+ and aligned together by minimizing their NCPs RMSD. The DNA linker/arm distribution (first columns) and their dynamic ranges (second columns) showed that the di-nucleosome at 5 mM Na^+ and tetra-nucleosome array in 50 mM Na^+ were restricted within a narrower 3D space comparing to tri- and tetra-nucleosome at 5 mM Na^+. The typical conformations (third columns) from each array categories showed that the two NCP discoidal planes are more perpendicular to each other in conditions a and d comparing to conditions b and c. The DNA were rainbow colored with blue, red and yellow along the direction from the entry to the exist side of the array.

Further analyses of the distance between two consecutive NCPs (*i vs. i+1*) showed a similar distribution among multi-nucleosome arrays (Figure 3.12d), suggesting the NCP spacing was independent of the NCP number in an array. The analysis of dihedral

angle between consecutive NCPs showed no preferred orientation between their discoidal planes except for di-nucleosomes, which favored a perpendicular arrangement for its two NCPs (Figure 3.16). This singularity of the di-nucleosome supports its above unique distributions in angles $\theta_{\parallel}$ and α, and in agreement with the crystal structure conformation in binding with Isw1a (Yamada et al., 2011). To vividly display the thermodynamics of multi-nucleosomes in solution, the movies of di- and tri-nucleosomes were also created as the mono-nucleosome (Extended Data Video 2).

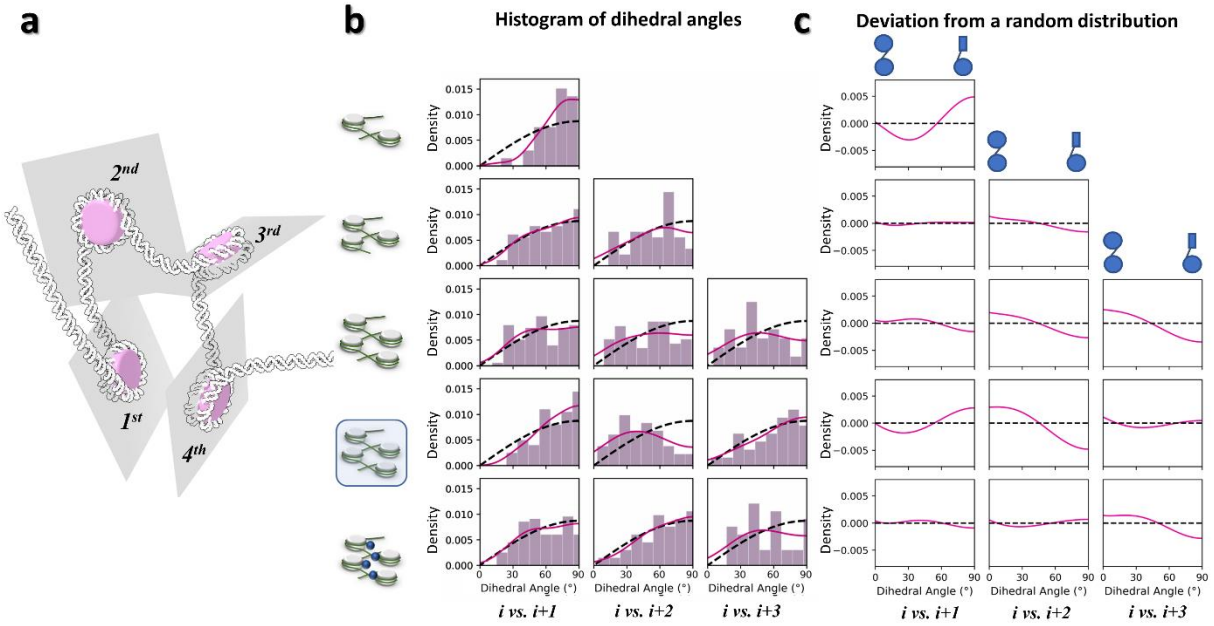


Figure 3.16 Distribution of the intra-nucleosome array NCP dihedral angles. a) Schematics defining the NCP central discoidal planes within the array. b) Histograms of the distribution of the dihedral angles between each pair of NCP discoidal planes. The measured angle distributions (in pink) were compared with a sin function (in black, presenting the angle distribution of a randomly rotated plane against a fixed plane). c) The deviations of experimental measurements from the sin function show the preferred dihedral angles found between each pair of NCPs within the arrays.

3.3.5 Salt effect on the dynamics of tetra-nucleosome array 3D structure

To investigate how external stimuli such as ionic strength or linker histone H1 regulate the structural dynamics of nucleosome arrays, tetra-nucleosome as a basic regulatory unit for large-scale chromatin structures (Ding et al., 2021) was used as a prototype to probe the array structural changes. We repeated the above cryo-ET experiment by increasing Na^+ concentration from 5 mM to 50 mM (Figure 3.17) and

reconstructed a total of 34 tetra-nucleosome array density maps with flexible fitted models (Figure 3.18 and Extend Data Video 6). These structures again confirmed the asymmetric zig-zag arrangement of the array but presented in a more compacted shape in contrast with the low-salt structures.

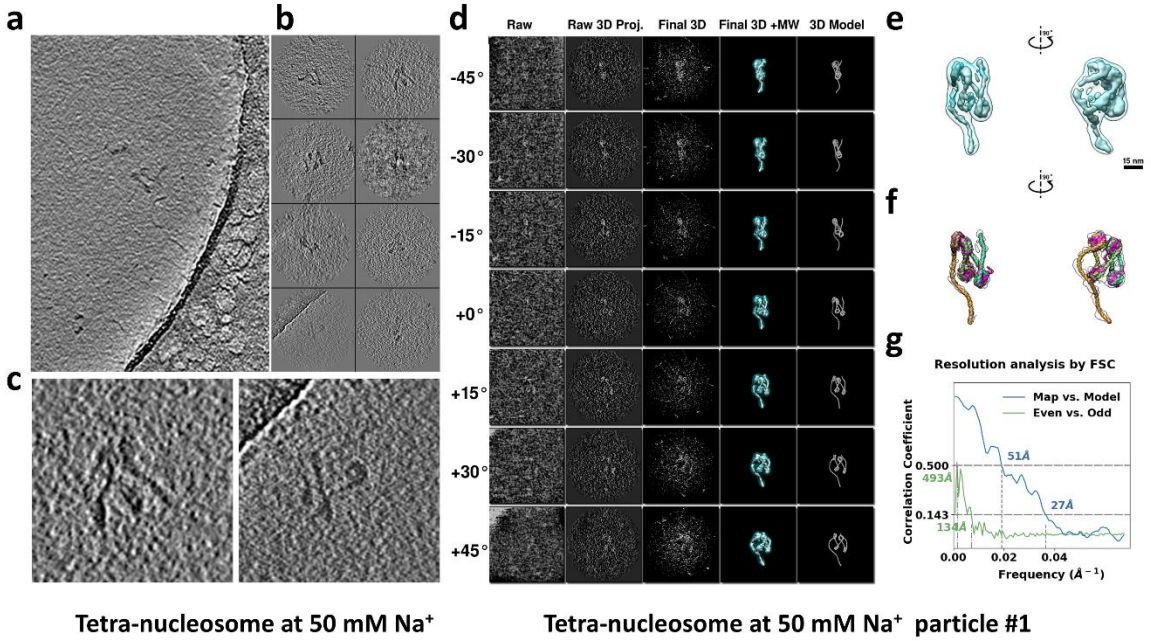


Figure 3.17 Cryo-EM images and 3D reconstruction of an individual tetra-nucleosome particle in the 50 mM Na^+ condition. a) Representative 2D slice (5nm thickness) of the. tomogram of tetra-nucleosome arrays designed with 200-bp entry and 100-bp exit DNA arms and linker length between nucleosomes of 40 bp in the presence of 50 mM Na^+. b) Selected representative tetra-nucleosome array particles with various conformations. c) Zoomed-in image of two representative particles. d) 3D reconstruction process for an individual particle. Seven representative projections before (left column) and after refinement (middle column) are compared to the final 3D density map (right column) at corresponding tilt angles. e) Orthogonal views of a tetra-nucleosome reconstruction displayed at two contour levels. f) Fitted model and map of the particle. The DNA is colored as a gradient from its entry side (orange) to its exit side (cyan). The histones are colored in pink. g) Resolution estimation of the reconstruction by using Fourier Shell Correlation (FSC).

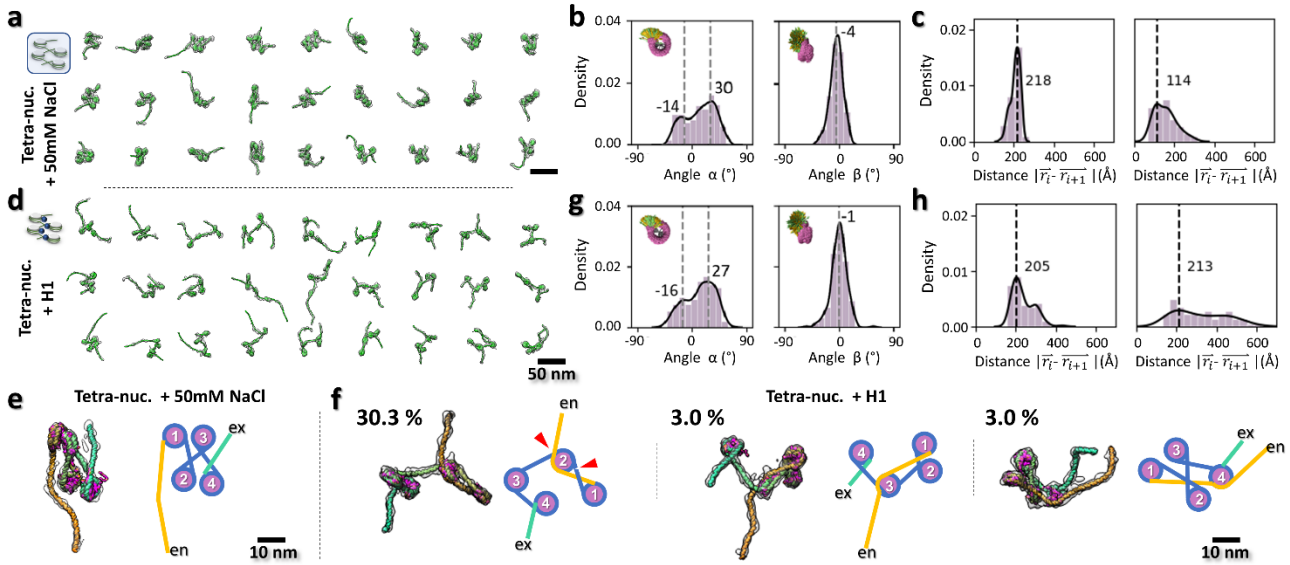


Figure 3.18 3D structures and dynamics of tetra-nucleosome arrays in 50 mM Na^+ and in the presence of H1. a) 27 representative cryo-ET density maps from individual tetra-nucleosomes prepared with higher ionic strength (50 mM Na+). Each map was superimposed with a flexibly fitted model. b) Histogram distributions of the wrapping angle α and the bending angle β. c) Histograms showing the distributions of the core-to-core distance between i, i+1 NCPs and i, i+2 NCPs. d) 27 representative density maps of tetra-nucleosome arrays in the presence of linker histone H1, super-imposed with the flitted models. e and f) Zoomed-in view of 3D maps and models comparing the conformational changes of tetranucleosome in response to higher ionic strength and H1, respectively. The schematic on the right side of each panel shows the paths of the DNA in tetranucleosomes under corresponding conditions. The NCP, entry-, intermediate-, and exit-DNA are colored in purple, orange, blue, and green, respectively. Red arrows indicated the possible H1 binding sites for the foreign DNA-invaded NCP. g and h) Histograms of α and β angle measurements and core-to-core distances among tetranucleosome NCPs in the presence of H1, respectively. Scale bars are 50 nm in a and d, 10 nm in e.

Quantitative characterization showed that comparing to the tetra-nucleosome array at low-salt condition, increasing the ionic strength retains a similar NCP unwrapping level at the entry-side, *i.e.* -4 ± 9-bp vs. -4 ± 8-bp, but slightly stabilizing the exit-side (-9 ± 8-bp vs. -14 ± 12-bp) (Figure 3.13b). Although the θ angle (Figure 3.14f) showed a similar distribution with an insignificant change of peak position from 41° to 47° after increasing the ionic strength, its two perpendicular fractions $\theta_{\parallel}$ and $\theta_{\perp}$ displayed

a distinct peak shift (changed from -7° to -42° and 35° to 19°, respectively) yielding a high percentage of closed-arm conformations (Figure 3.14f). Analysis of α angle distribution showed that, although its major peak remained similar, its minor peak clearly became more apparent and shifted from -6° to -14° (Figure 3.18b) suggesting the DNA linker on the NCP surface underwent a remarkable conformational transition. This transition may be caused by the stronger electrostatic screening effect in higher ionic strength buffer (Blank and Becker, 1995), which largely reduced the repulsive force between the two DNA linkers and brought them closer to each other. Consistently, less DNA linkers were observed bending out from the NCP discoidal plane (β changing from -9° to -4°, Figure 3.18b). Further NCP geometry analysis displayed a slight (~19 Å, from 237 Å to 218 Å), a dramatic (~151 Å, from 265 Å to 114 Å), and a significant (~119 Å, from 388 Å to 269 Å) distance reduction between consecutive NCPs (*i vs. i+1*), every-other NCPs (*i vs. i+2*), and first and fourth NCPs (*i vs. i+3*), respectively (Figure 3.18c and 3.19). The decrease in distance among all NCP pairs confirmed and quantified the above observed array compaction (Figure 3.18a). Concomitantly, in response to a higher ionic strength environment, the NCP dihedral angle analysis revealed that the orientation between consecutive NCPs became more perpendicular, while the orientation between every-other NCPs became increasingly parallel (Figure 3.16).

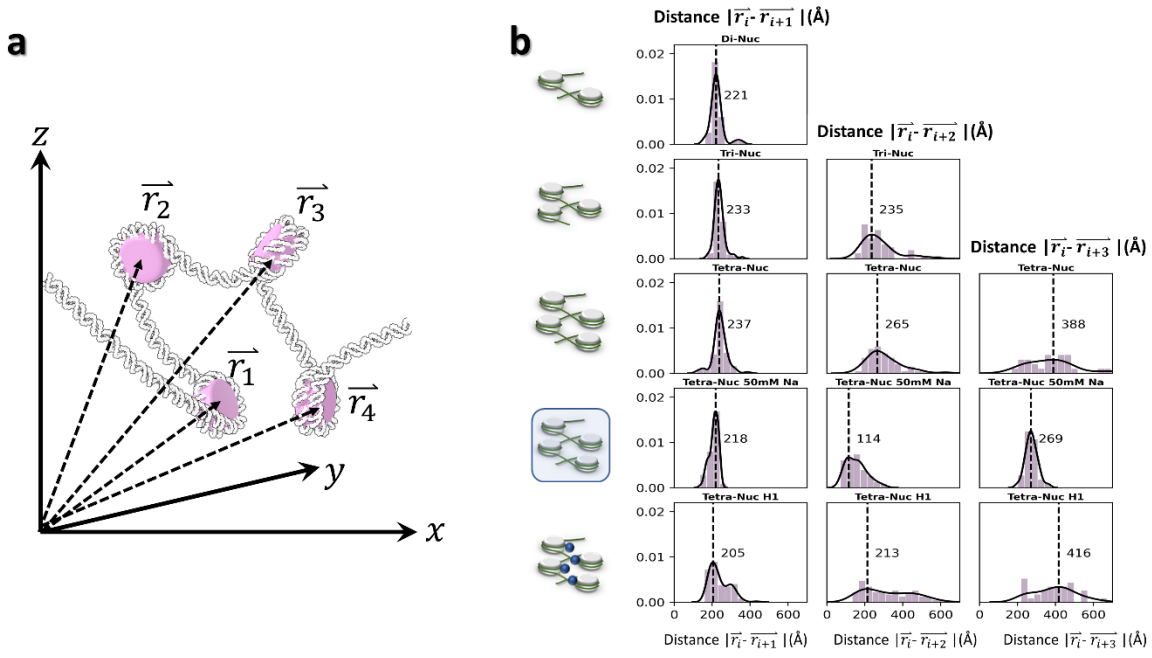


Figure 3.19 Distribution of the intra-nucleosome array NCP distances. a) Schematics of the vector, $\vec{r_i}$ starting from the origin and pointing to the center of mass of each NCP. b) Histograms of the measured distances between each pair of NCP centers within the array.

3.3.6 H1 effect on the dynamics of tetra-nucleosome array 3D structure

To investigate how the presence of the linker histone H1 regulates the nucleosome array structural dynamics, we repeated the above experiment by incubating the tetra-nucleosome with H1 (in a molar ratio of 1:4) under the same low-salt condition (5 mM Na^+). A total of 33 cryo-ET density maps with their corresponding fitted models were constructed (Figure 3.20 and Extend Data Video 7). Once again, these arrays displayed a general asymmetric zig-zag conformation (Figure 3.18d), however we did not observe similar compaction as that described above under 50 mM Na^+ (Figure 3.18e); instead, ~40% particles displayed a new conformation, in which the 200-bp DNA arm cross-linked to one of the NCP within the array by replacing its originally wrapped DNA on the exit side (Figure 3.18f). This cross-linking may also occur among nucleosome arrays causing the formation of condensates (Figure 3.21), which indicates an alternative function of H1 in strengthening the inter-arrays network (Turner et al., 2018).

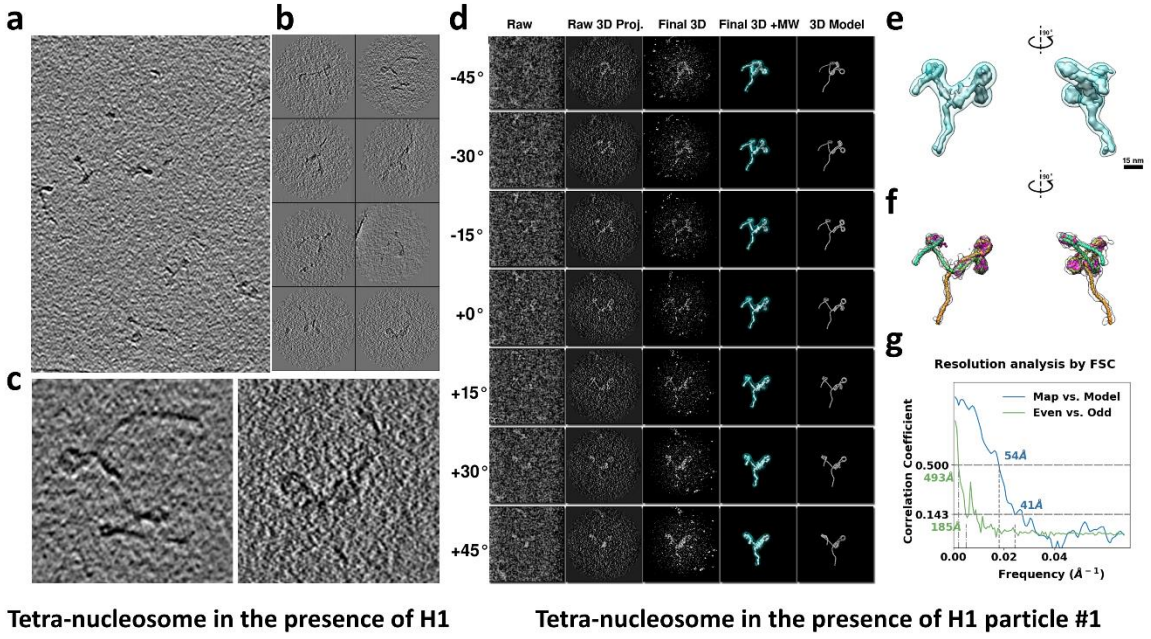


Figure 3.20 Cryo-EM images and 3D reconstruction of an individual tetra-nucleosome particle in the presence of H1. a) Representative 2D slice (5nm thickness) of a tomogram of tetra-nucleosome arrays designed with 200-bp entry and 100-bp exit DNA arms and linker length between nucleosomes of 40 bp in the presence of H1. b) Selected representative tetra-nucleosome array particles with various conformations. c) Zoomed-in image of two representative particles. d) 3D reconstruction process of an individual particle. Seven representative projections before (left column) and after refinement

(middle column) are compared to the final 3D density map (right column) at the corresponding tilt angles. e) Orthogonal views of a tetra-nucleosome 3D reconstruction displayed with two contour levels, f) Fitted model and map of the particle. The DNA was colored as a gradient from its entry side (orange) to its exit side (cyan). The histones were colored in pink. g) Resolution estimation of the reconstruction by using the Fourier Shell Correlation (FSC) coefficient.

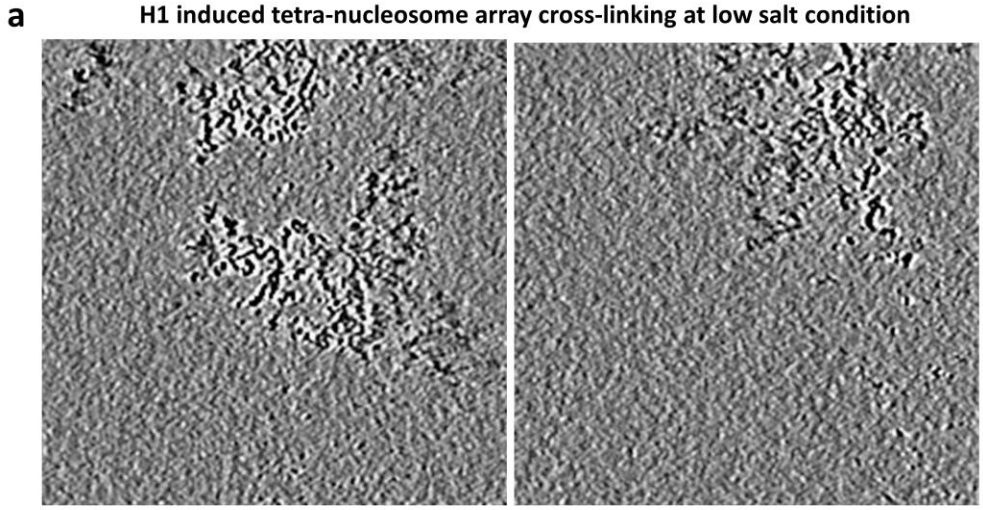


Figure 3.21 H1 induced partial tetra-nucleosome array cross-linking in low salt conditions. a) Cryo-EM 2D slice (5nm thickness) of a tomogram of tetra-nucleosome arrays in the presence of 5 mM Na^+ and H1.

NCP unwrapping analysis of tetra-nucleosomes in the presence of H1 showed a similar distribution to that of absent H1 on its the entry-side (-6 ± 8-bp vs. -6 ± 8-bp). However, a significant increase in unwrapping dynamics was observed on the exit-side (-17 ± 19-bp *vs.* -14 ± 12-bp). Moreover, the emergence of a prominent second peak at ~-45-bp Å (Figure 3.13b) indicates that the exit-side DNA was entirely unwrapped from the H2A-H2B dimer (that covers 40-bp of DNA), which may be due to its relatively weak binding to DNA (Li and Wang, 2012) and spontaneous site exposure capability (Anderson et al., 2002b; Anderson and Widom, 2000).

In the presence of H1, though the θ distribution (Figure 3.14f) showed only a slight peak position difference (from 41° to 50°), its fraction $\theta_{\|}$ revealed a remarkable shift in major peak position (from -7° to -48°), similar to the change caused by 50 mM Na^+ (from -7° to -42°). This similarity implies that high ionic strength and H1 are cooperative partners in driving nucleosome into a "closed-arms" conformation via different

approaches, *i.e.* electrostatic screening *vs.* introducing a physical "lock" (Garcia-Saez et al., 2018; Bednar et al., 2017). However, different from 50 mM Na^+ condition, a minor peak of $\theta_{\parallel}$ emerged at 136° suggesting that ~1/4 (28%) of NCPs within the array also adopted "widely-opened arms" conformation at the same time. This result is consistent with the fact that one of the NCP severely unwrapped from the H2A-H2B surface causing its two DNA links point to opposite directions. Therefore, H1 can only exert the "lock" function on part (~3/4) of NCPs within the tetra-nucleosome array, while inducing more DNA unwrapping for the rest (~1/4) at the low salt condition. Similarly, DNA linkers on the same NCP became more parallel to each other (from 35° to 6°) by viewing the $\theta_{\perp}$ distribution from the side of NCP, suggesting H1 is more efficient in driving DNA linkers back to the NCP discoidal plane comparing to 50 mM Na^+ condition (from 35° to 19°). The second peak of $\theta_{\perp}$ at 152° again confirmed the two co-existing conformations of NCP units. The similar distributions of α and β to that of 50 mM Na^+ condition implied a similar intrinsic change and a functional synergy (Figure 3.18g and Figure 3.22). The cause of the alternative conformation of tetra-nucleosome arrays may be not only due to the fact that the severe NCP unwrapping was compensated by the "foreign DNA" rewrapping, but also that the newly formed DNA junctions provided two possible H1 binding sites for one NCP, which could further lower the energy level of the system (Figure 3.18f, red arrow).

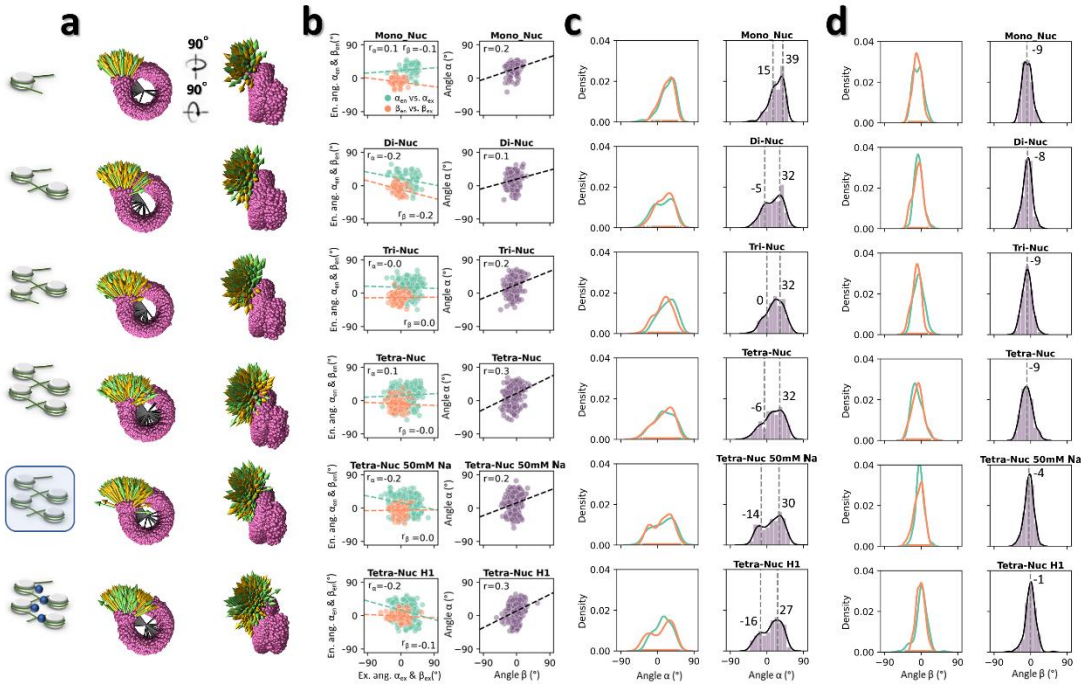


Figure 3.22 Distributions of wrapping angle α and bending angle β. a) Superposition of the DNA vectors that define the entry and exit linker DNA directions (yellow and cyan arrows, respectively, originating from the NCPs. The wrapped DNA within NCPs

is shown in magenta. b) Correlation analysis between angles formed by the DNA linkers relative to the NCP. The correlation between the two wrapping angles (measured between the entry and exit DNA linkers on the same NCP, αen vs. αex, shown in cyan) and between the two bending angles (measured from the same pair of DNA linkers, βen vs. βex, shown in orange) were calculated from the Pearson test r-value (left column). This correlation coefficient was also calculated between the wrapping angle and bending angle for each NCP (showed in purple, right column). c) Distribution of entry and exit wrapping angles (αen and αex shown in cyan and orange, respectively, left column) and their pooling (right column). d) Distribution of entry and exit bending angles (βen and βex shown in cyan and orange, respectively, left column) and their pooling (right column).

Characterization of NCP pairwise distance showed the array was more disordered than others. Although the major distance of two consecutive NCPs was decreased in the presence of H1 as that by 50 mM Na^+ (205 Å vs. 218 Å, respectively), the distribution range was increased significantly by H1 (Figure 3.18h). The orientation analysis showed the NCPs became more randomly oriented relative to each other than in other conditions (Figure 3.16). This increase in disorder of nucleosome array conformations favors the amorphous phase separation model over the symmetric super-helical model, although H1 under low-salt condition is not sufficient to complete the phase transition process by itself.

3.3.7 DNA linker orientation regulates the structure and dynamics of chromatin

Single-molecule 3D structures enable us to quantify the intrinsic conformational changes of spatial orientation of DNA linkers against NCP in response to the environment. To display how this intrinsic conformational change mechanically regulates chromatin conformation and dynamics for further understanding of its genetic regulation functions, the hecta-nucleosome arrays were computationally assembled based on the above experimentally measured angle and unwrapping length distributions under corresponding salt conditions. The simulated structures at low-salt (Figure 3.23a) presented curvy irregular fibers in a length and width peaking at 561 nm and 77 nm, respectively (Extended Figure 3.24a). In contrast, the structures at higher-salt (Figure 3.23b) also showed fiber-shaped conformations but with a much smaller length and width, peaking at 395 nm and 38 nm, respectively (Figure 3.24a), which were similar to the experimentally observed irregular/disorganized 30 nm fibers (Scheffer et al., 2011a; Konig et al., 2007). The comparison of the fiber persistence length in low- and high-salt condition (31.7 nm vs 75.3nm, Figure 3.24b) may explain why the low salt fiber largely loses its 30-nm characteristic: a small persistence length corresponding to a larger

flexibility causes the fiber to become easily entangled and intertwined. Notably, the internal symmetry was absent from either of the above cases.

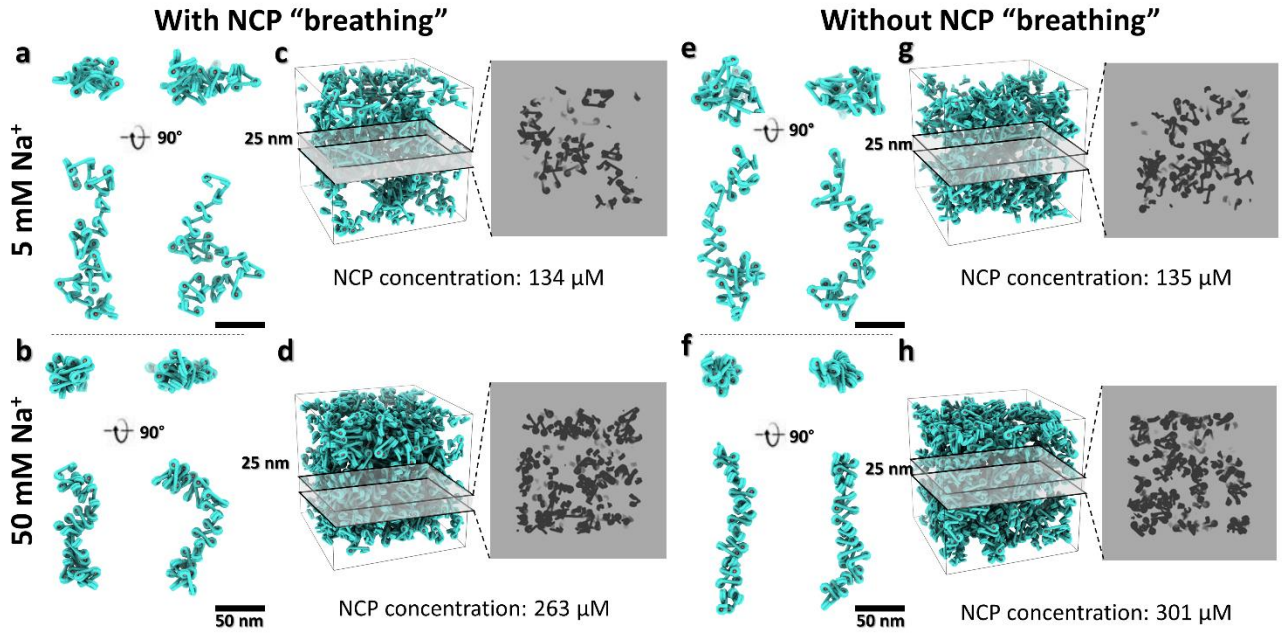


Figure 3.23 Hypothesis of chromatin regulation mechanism by changing linker DNA angles. a and b) In silico assembly of a long chromatin fiber by connecting 100 NCPs (only the first 30 NCPs were showed) following the α, β angle distribution and the NCP unwrapping distribution acquired from the experimental statistics of the tetranucleosome arrays in 5 mM and 50 mM Na^+ conditions, respectively. c and d) Simulation of a chromatin domain organization constructed from sequential fitting of 100 of the fibers from b and c, respectively, within a 200 nm radius cubic volume. The nearest inter-array distances were confined by the *i* and *i+3* core-to-core distance distribution measured from the tetranucleosome arrays for the corresponding conditions. The right bottom panel shows a central slice of 25 nm thickness from the cubic volume. e and f) In silico assembly of chromatin fibers containing 100 NCPs using only the experimental α, β angle distribution without considering the NCP "breathing" in low- and high-salt conditions, respectively. NCP unwrapping were confined to a fixed mean value of the distribution. g and h) Simulation of a chromatin domain organization and central slice representation after eliminating NCP "breathing" at low- and high-salt, respectively. Scale bars are 50 nm.

Considering the chromatins appear as condensates in heterochromatin and euchromatin domains (Strickfaden et al., 2020), we further evaluated the NCP density in condensates formed by simulated fibers in two extreme cases. In the first case, the

condensates were seamlessly assembled from array fibers. In this case, the density of the condensate is equal to the average density of the chromatin fibers. The averaged fiber NCP densities at low- and high-salt were 266 μM and 341 μM, respectively (Figure 3.24c). In the second case, all fibers were treated as rigid-bodies and randomly docked into a 200 nm cube at a maximized number. In this case, the NCP densities were 134 μM and 263 μM for low- and high salt, respectively (Figure 3.23c and d). The calculated NCP density from the simulated condensates in both cases were close to the experimentally measured density range *in vivo* (80-520 μM) (Hihara et al., 2012) and *in vitro* (~340 μM) (Gibson et al., 2019), suggesting the extracted angle dynamics parameters from the microscopic point of view are key factors in controlling the macroscopic chromatin condensation and geometry. Interestingly, by slicing through the simulated chromatin cube along its z-dimension with a 25 nm thickness slab, the fiber pattern was diminished. It implied that the absence of the 30-nm fibers in cryo-EM imaging (Eltsov et al., 2008; McDowall et al., 1986) or cryo-ET cell sectioning slices (Cai et al., 2018a; Cai et al., 2018b) may be due to following additive effects: such as, the intertwining of the fiber evidenced by its small persistence length and large standard deviation of its width, the confinement of the fiber within a restrict volume, and the low probability of capturing fibers happens to lie on plane of cryo-ET cell section slices (especially when the slice thickness approaching 30 nm (Ou et al., 2017; Cai et al., 2018a)).

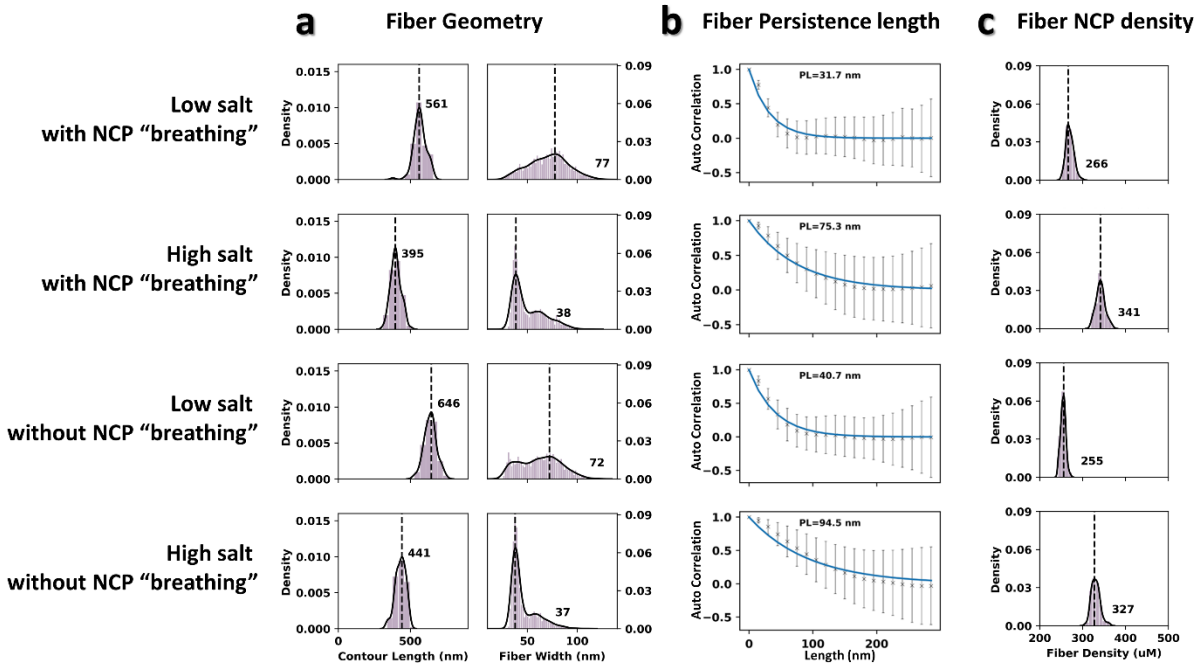


Figure 3.24 Distribution of the fiber length, width, persistence length, and NCP density of the simulated nucleosome array containing 100 NCPs. a) Histograms of the measured width and length of the simulated nucleosome array fiber. b) Stiffness analysis

of the simulated nucleosome array represented by the persistence length of the fiber. c) Histograms of the NCP concentration distribution measured by dividing the number of NCP by the corresponding fiber volume. The statistics were measured from 100 simulated fibers.

The NCP "breathing" enables restriction enzyme access to the inner wrapped DNA target site (Anderson et al., 2002b) or causes nucleosome translocation without chromatin remodeler (Lai and Pugh, 2017). To investigate how NCP site exposure dynamics influence chromatin morphology, we repeated the above simulation by removing the DNA wrapping dynamics (*i.e.* fixed the unwrapping length at its mean value of the distribution). The simulation showed the peaks of fiber length and width were 646 nm and 72 nm, respectively, which were shifted to 441 nm and 37 nm at 50 mM Na^+ condition (Figure 3.24a). After abolishing the "breathing", the 30nm fiber turned into a less flexible and straighter conformation (Figure 3.23e and f) (with a persistence length of 40.7 nm and 94.5 nm for low- and high-salt models, respectively Figure 3.24b) without changing the zig-zag organization. The reassembled condensates also showed the breathing did not induce an obvious change of the cube NCP density (135 μM and 301 μM, respectively Figure 3.23g and h) suggesting the breathing did not contribute to the enzyme accessibility through increasing the gap among condensates. Instead, the array DNA linker angles play a major role to define the density of condensates, which regulate the morphology of chromatin and thus control gene activity.

3.4 Discussion and Conclusion

The structural dynamics of nucleosomes is of fundamental importance to improve our understanding of gene regulation and DNA replication machinery. However, the structural analysis of nucleosome array is extremely challenging given its heterogeneity and the large-scale and multiple dimensional conformational changes in response to chemical and biological stimuli. This represents an obstacle in selecting a homogenous population in terms of structural identity for averaging. We took advantage of cryo-ET to obtain the single-molecule 3D structure from each individual nucleosome array under its native conditions. We found the small angle change of the DNA linker against NCP plays a key role in regulation of large-scale chromatin conformational dynamics. We expect this approach could be a powerful tool for identifying macromolecular structures in action within a cell in the future.

3.5 Materials and Methods

3.5.1 Histone octamers purification

Recombinant wild-type Xenopus laevis histone octamers were expressed in E. coli cells and purified as previously described (Dyer et al., 2004). The isolated histone octamers were stored in 10 mM Tris, pH 7.6, 1.6 M NaCl, 1 mM EDTA, 1 mM DTT, 20% glycerol at −80°C.

3.5.2 DNA templates

The DNA templates for mono-, di-, tri- and tetra-nucleosomes were cloned and purified as described (Luger et al., 1999). The templates were designed with inserting 147 bp of nucleosome position 601 sequence into a 200 bp upstream DNA (entry DNA) and a 100 bp down steam DNA (exit DNA). For the DNA templates of di-, tri- and tetra-nucleosome, additional 40 bp linkers DNA were inserted between neighboring 147 bp of nucleosome position 601 sequences (wrapping DNA). The sequences for each component were as follows:

147 bp of nucleosome position 601 sequence：
CTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCAC
CGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTA
CTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCCTGT

200 bp entry (upstream) DNA sequence：
CGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTA
TGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTTGC
ATGCCTGCAGGTCCGGGATCCTAATGACCAAGGAAAGCATGATTCTTCACACC
GAGTTCATCCCTTATGTGATGGACCCTATACGCGGCCGCC

100 bp exit (downstream) DNA sequence：
GCATGTATTGAACAGCGACCTTGCCGGTGCCAGTCGGATAGTGTTCCGAGCTC
CCACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCGCCCTATAGT

40 bp linker DNA sequence:
GCATGTATTGGAAGACTACAGTACCCTATACGCGGCCGCC

3.5.3 Nucleosome array assembling

Nucleosome arrays were assembled using the salt dialysis method modified from what was previously described (Dyer et al., 2004). In brief, the reconstitution reaction mixture with octamers and 601 based DNA templates was dialyzed over 16 h at 4 °C in TEN buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 2 M NaCl), through slowly adding

in TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) to lower the concentration of NaCl from 2 M to 0.6 M during the dilution. For histone H1 incorporation, an equal molar amount of histone H1 (relative to mono-nucleosomes) was added and further dialyzed in TE buffer with 0.6 M NaCl for 3 h, followed by a final dialysis step in HE buffer (10 mM HEPES, pH 8.0, 0.1 mM EDTA) for 4 h. The stoichiometry of histone octamer binding to the DNA template and H1 binding to nucleosome was used for EM images.

3.5.4 TEM specimen preparation

The cryo-EM specimens were prepared following the procedure described before (Zhang and Ren, 2012). Briefly, an aliquot (~ 3 μl) of nucleosome array sample ~400 nM was placed onto the 200 mesh Quantifoil copper grid (Q210CR-06, Electron Microscopy Sciences) that had been glow-discharged for ~15 s by (PELCO easiGlow™ Glow Discharge Cleaning System) for 15 seconds. After incubating for ~ 10 sec, the grid was flash-frozen in liquid ethane at ~ 90% humidity and 4 °C with a Leica EM GP rapid-plunging device (Leica, Buffalo Grove, IL, USA) after being blotted with filter paper with a controlled blotting time (2 s). The flash-frozen grids were transferred into liquid nitrogen for storage.

The NS EM specimens of nucleosome array sample were prepared using the optimized negative-staining protocol (OpNS) as described (Rames et al., 2014a). In brief, the samples were diluted to ~20 nM with sample buffer. An aliquot (~4 μL) of diluted sample was placed on an ultra-thin carbon-coated 200-mesh copper grid (CF200-Cu-UL, Electron Microscopy Sciences, Hatfield, PA, USA) that had been glow-discharged for ~15 s. After ~1 min incubation, the excess solution on the grid was blotted with filter paper. The grid was then washed with water and stained with 1% (w/v) uranyl formate (UF) before air-drying with nitrogen.

3.5.5 TEM data acquisition

Cryo-EM specimens were screened by a Titan Krios (FEI) transmission electron microscope operated at 300 kV high tension with a Gatan energy filter. The untitled cryo-EM micrographs were acquired by using Volta phase plate under defocus at ~2.5 μm using a Gatan K2 Summit direct electron detection camera under a magnification of ~81 kx (each pixel of the micrographs corresponds to ~0.9 Å in specimens). The particles with a box size of 800 × 800 pixel will be used to display the morphology. Cryo-EM tilt image series of the samples were collected from -60° to +60° at 3° increments on a Titan Krios G2 TEM equipped with a Gatan energy filter and a K2 Summit direct electron detection camera operated under 300kV high tension. During data acquisition, the SerialEM (Schorb et al., 2019) software was used to automatically track the specimen and maintain defocus at ~2.5 – 3.0 μm. The acquired tilt image series at magnification of ~50 kx (each

pixel corresponds to 1.46 Å) represents a total dose of ~150 - 200 $e^-/Å^2$. For each tilt angle, a total number of 8-10 frames was collected under the exposure of 0.25 s per frame.

OpNS specimens were screened by using a Zeiss Libra 120 Plus TEM (Carl Zeiss NTS) operated at 120 kV high tension with a 10-20 eV energy filter. The OpNS micrographs were acquired under defocus at ~0.6 μm using a Gatan UltraScan 4K × 4K CCD under a magnification of 125 kx and 160 kx (each pixel of the micrographs corresponds to ~0.94 Å and ~0.74 Å in specimens respectively). Tilt image series of mono-nucleosome sample were collected from -60° to +60° in 3° increments using a Zeiss Libra 120 Plus TEM (Carl Zeiss NTS) equipped with an in-column energy filter and a Gatan UltraScan 4K X 4K CCD. During data acquisition, the Gatan tomography module (Gatan Inc., Pleasanton, CA, USA) operated in Advanced Tomography mode was used to track the specimen and maintain defocus at ~2.0 μm. The acquired tilt image series at magnification of 50 kx (each pixel corresponds to 0.24 nm) represents a total dose of ~60 e-/Å2.

3.5.6 Image preprocessing

For each tilt image, the motion of the frames was corrected by MotionCor2(Zheng et al., 2017). To reduce the image noise, we followed a machine learning method (NOISE2NOISE, T2T method) as described (Buchholz et al., 2019). The contrast transfer function (CTF) of denoised tilt series was examined by using ctffind3 software (Mindell and Grigorieff, 2003) and the phase and amplitude were corrected by using the ''TF CTS'' command in SPIDER(Frank et al., 1996) software, GCTF(Zhang, 2016) and TOMOCTF (Fernandez et al., 2006) after the X-ray speckles were removed. The denoised tilt series were aligned by using IMOD (Kremer et al., 1996). By using boxer from EMAN software (Ludtke et al., 1999), a box of 256 × 256 pixels was used to select the particles of nucleosome from the ~113 micrographs imaged under 125 k× magnification, while a box of 200 × 200 pixels was used to select the segments of the upstream (or entry) and downstream (or exit) DNAs from the ~150 micrographs imaged under 160 k× magnification. All particles were masked using a round mask generated from SPIDER software after a Gaussian high-pass filtering. The reference-free class averages of particles were obtained by using refine2d (EMAN software) based on 30,540 particles of DNA segment and 13,029 particles of NCPs.

3.5.7 Individual particle electron tomography (IPET) 3D reconstruction

The tilt series of each targeted particles were semi-automatically tracked and then windowed in square windows of ~ 1,000 × 1,000-pixel size using IPET software (Zhang and Ren, 2012), and finally binned four to five times to reduce computation time in

subsequent reconstructions. Following the pipeline of IPET reconstruction (Zhang and Ren, 2012), a local tilt series images containing a single nucleosome array particle was extracted from the IMOD aligned full-size tilt series. This allowed us to perform "focused" 3D reconstruction, in which the local refined alignment was more robust to eliminate large image distortion, tilt-axis variation with respect to tilt angle, and tilt angle offset. Briefly, an ab initio 3D density map was directly back-projected in Fourier space and served as the initial model. The refinement was then iteratively invoked to translationally align each tilted particle image to the computed projection. During the refinement, automatically generated Gaussian low-pass filters, soft-boundary circular masks and particle-shaped soft-boundary masks were sequentially applied to the tilt images and references to increase the alignment accuracy (Zhang and Ren, 2012). An improved model was then reconstructed based upon the refined alignment at the end of each refinement iteration. The 3D map was then reconstructed by back-projection of filtered and masked particle tilt series. The final back projection was performed in Fourier space without weighting. To reduce the artifact caused by the limited tilt angle range, the final 3D map was submitted for a published missing-wedge correction method, named low-tilt tomographic 3D reconstruction method (LoTToR) (Zhai et al., 2020), in which the low resolution mask used was generated by the Model-Based Iterative Reconstruction (MBIR) method (Yan et al., 2019). All final density maps were low-pass filtered to 4.5 nm using EMAN (Ludtke et al., 1999) and displayed using UCSF Chimera (Pettersen et al., 2004).

3.5.8 Estimation of the reconstruction resolution

The resolution for the reconstructions were estimated by two methods. i) Data-to-Data based analysis: the Fourier Shell Correlation (FSC) were calculated between two independently reconstructed 3D maps, that each of the map was based on one-halves of the tilt-series (split by even and odd tilt index) after particle alignment refinement of the IPET (Zhang and Ren, 2012). The frequencies at which the FSC curve first falls to values of 0.5 and 0.147 were used to represent the reconstruction resolution. Notably, the resolution estimated by this method could be severely under-estimated since the reconstruction from one half of tilt-series significantly reduced the quality of the map compared to the final reconstruction. ii) Data-to-Model based analysis: the FSC curve between the final IPET reconstruction and the density map converted from the corresponding fitting model was calculated. The frequencies at which the FSC curve fell below 0.5 and 0.147 were used to estimate the resolution. The density map of the fitting model was generated by pdb2mrc in EMAN software (Ludtke et al., 1999).

3.5.9 Modeling the structure of nucleosome arrays

To build a model for the reconstructed map of each individual nucleosome array

particle, the pathways for the two flanking DNA arms were initially traced by sampling a group of 3D points located at the high-density loci of the map followed by sorting their order into a points list. Then by fitting the crystal structure of Xenopus laevis NCP into the discoidal-shaped high-density region, the DNA pathway with the NCP were also defined and converted to a list of 3D points. These points series from different models were merged together after removing adjacent clashes and then fitted with a smooth quadratic Bezier curve followed by conversion into DNA model by using GraphiteLifeExplorer (Hornus et al., 2013). The total length of the DNA used to thread the model matched with our designed DNA construct, i.e. 456, 632, 818, 1008 for mono-, di-, tri- and tetra-nucleosome arrays, respectively. Notice that for some of the maps, small portion of DNA densities (~10%) was missing near the middle portion of DNA arms (which may be due the orientation of DNA segments aligned near-perfectly perpendicular to the beam direction, resulting the lowest image contrast of the DNA). Fortunately, those missing portions were small and did not prevent the DNA model fitting, which can be circumvented by interpolating surrounding density at those loci. To further refine the model, the Molecular Dynamics Flexible Fitting (MDFF) was applied to energy minimize the model under a force gradient created by the electron density map.

3.5.10 Evaluating the fitting models

To evaluate the self-consistency of above fitted models, the correlation between the measured DNA linker length between two consecutive NCPs (named as $\boldsymbol{L(n)}$) and the estimated DNA unwrapping level between the same NCPs were calculated. The $\boldsymbol{L(n)}$ was acquired from the experimental data, in which distance between the DNA exit position of the nth NCP and the entry position of the (n+1)th NCP from the density map were measured. The DNA unwrapping level was estimated from the fitted model, in which the unwrapping angles from the exit side of nth NCP and entry side of the (n+1)th NCP were added together ($\boldsymbol{\theta(n) = \theta_{ex}(n) + \theta_{en}(n+1)}$). The Pearson correlation coefficient, R, was calculated from the linear regression fitting of the unwrapping length $\boldsymbol{L(n)}$ against the angle $\boldsymbol{\theta(n)}$ using the statsmodels package in Python.

3.5.11 Defining the entry and exit linker DNA origins on the NCP

By using the fitted nucleosome array models, the entry and exit DNA arm origins on the NCP can be estimated with the following procedure. The DNA portion of the fitted nucleosome array model was first converted into a list of 3D points (series m) by averaging the coordinates of C1 atoms of each base-pair. After aligning the nucleosome crystal structure to the fitted array model at its nth NCP region, a points list (series n) along the wrapped DNA from the nucleosome crystal structure was also generated by the same method. By comparing the two points lists, the overlapped region on the fitted

DNA model can be identified if any points from the series n were found within its 8 Å radius of series m. The 8 Å criterion was chosen based on that, when two DNA helical centers separated away from each other for more than one third of the DNA diameter (~24 Å), the separation of two aligned DNAs can be distinguished. The base pair indexes at the two distal ends of this overlapped region were used to define the entry and exit DNA linker arm origin along the array. The total number of the base-pairs between the entry and exit position were used to calculate the length of wrapped DNA on the histone surface.

3.5.12 Measuring the NCP wrapping dynamics on histone surface

By measuring the position and length of the wrapped DNA on each histone surface, the DNA unwrapping footprint along the array can be quantified as follows. For a fitted array model, a binary score was assigned to each DNA base pair along the DNA sequence depending on its contact with histone octamer ("1" for contact and "0" for non-contact). By averaging the scores of the ith base pair from different fitted array models, the mean score distributions along the DNA template (in an order from 1-n th base pair) for a specific array category were calculated. This score distribution reflects the probability of finding DNA unwrapping events at some of the designed wrapping regions that contains the 146-bp widom 601 position sequence (Lowary and Widom, 1998). The distribution occupancy on the entry and exit side of each NCP reflects the asymmetry property of the DNA sequence in binding to histone octamer and the dynamic relationship between DNA unwrapping and equilibrium assembling, such as transition states among the tetrasome, hexsome and full octasome (Rychkov et al., 2017).

3.5.13 Defining the entry and exit linker DNA vectors on the NCP

Due to the fact that the 20-bp DNA segment (~6.8 nm) is relatively stiff compared to the persistence length of DNA ~50 nm, the base pairs resided at the entry side DNA arm origin and its 20-bp upstream were used to define the start and end-point of a vector, respectively (based on their helical center coordinates). This vector was used to represent the entry DNA arm pointing direction. Similarity, the base pairs resided at the exit side DNA arm origin and its 20-bp downstream was used to define the exit DNA arm vector.

3.5.14 Measuring the angle θ, wrapping angle α and bending angle β of the DNA arm vectors

The angles of the DNA linker arms that extended from the discoidal-shaped NCP surface were measured by the following steps. i) Defining the X-, Y-, and Z-axes of each NCP model. The Z-axis was defined along the helical axis of the wrapping DNA

measured from its rotational symmetry. The Y-axis was defined by the dyad axis of the NCP. The cross-point of these two axes was used as the center of NCP and defined the X-axis, where it simultaneously passed the NCP center and was perpendicular to both Z- and Y-axis. ii) Defining the relative DNA arm angle θ. θ measured the angle between the previously defined entry and exit DNA linker arm vectors on the same NCP. Because DNA arm conformational states ("open and close") must be defined relative to the NCP, the projections of the θ angle on the NCP discoidal plane (X-Y plane) and its perpendicular plane (Y-Z plane) were measured as θ∥ and θ⊥, respectively. iii) Defining the in-plane wrapping angle α of an NCP arm vector. α calculated the angle formed by two vectors within the X-Y plane. One vector is the projection of the NCP linker arm vector on the X-Y plane, and the other vector is defined by the tangential direction of the discoidal-shaped projection of NCP on X-Y plane, which crosses the origin of the corresponding NCP linker arm. iv) Defining the out-of-plane bending angle β of an NCP arm vector. β calculated the angle formed by the NCP arm vector and the X-Y plane. The measured angle distribution was fitted with either one gaussian or two gaussians with the sklearn.mixture.Gaussian Mixture package.

3.5.15 Measuring the intra-array NCP core-core distances and plane-plane angles

To quantitatively define the spatial relationship among the NCPs from different types of nucleosome arrays, the core-core distances and plane-plane angles between each pair of NCPs were measured. The core-core distance, $\boldsymbol{D}(\boldsymbol{n}, \boldsymbol{n}+\boldsymbol{m})$, measured the distance from the center of the nth to the (n+m)*th* NCPs. The core-core angle, $\boldsymbol{\varphi}(\boldsymbol{n}, \boldsymbol{n}+\boldsymbol{m})$, measured the dihedral angle between the discoidal planes (X-Y planes) of the nth and (n+m)*th* NCPs. The histograms were fitted by a Kernel Density Estimation (KDE) function (Silverman, 1978). The dihedral angle distributions were compared with a sine function, which represents a distribution of orientations of two random planes (the angle θ between their normals), which in turn are equivalent to the probability of finding a point on the unit hemisphere contained in a differential ring-shape area: $P(\theta)d\theta = 2\pi(Rsin(\theta))Rd\theta/2\pi R^2 = sin(\theta)d\theta$.

3.5.16 Reconstituting the chromatin-like higher order structure

The distribution of the wrapping and bending angles relative to the NCP were used to build longer in silico nucleosome array fibers. Each longer array fiber containing 100 NCPs was generated by randomly connecting nucleosome model units. To be specific, due to the large number of atoms within the NCP (>12k) and the fact that the length of 40-bp DNA (~14nm) is much smaller than the persistence length of DNA (~50 nm), NCP units were coarse grained where the histone octamer and linker DNA arms

were treated as spheres and straight lines, respectively. Based on the distribution of the measured mean and standard deviation of the bending and wrapping angles (i.e. α and β) for the tetra-nucleosome array at low salt and high salt condition, two vectors in a 20-bp length followed the same angle distributions relative to the NCP were randomly generated and used to represent the entry and exit DNA linker arms by using VMD software (Humphrey et al., 1996). By repeating this process, a pool of NCP units with various linker arm conformation but followed the same spatial distribution to the experimental data were prepared. To assemble the NCPs unit into a fiber, randomly selected NCPs from the pool were sequentially connected together based on following procedures: i) The exit side 20-bp linker for the ith NCP were linearly connected to the entry side 20-bp linker for the i+1th NCP, ii) The 40-bp linker causing ~70° left-handed DNA rotation has been considered (Levitt, 1978), in which the i+1th NCP unit was rotated along the entry side linker vector after connecting to its previous NCP. iii) if the i+1th NCP were identified to clash with any of the previous NCPs, a new NCP will be randomly selected from the pool and resembled onto the ith NCP by repeating step i through iii. In order to simulate more realistic dynamics of the nucleosome fibers, the unwrapping events were also considered into the system. The origin of the entry and exit DNA linker vectors on the helical NCP track were also varied based on the experimentally measured distribution when constructing the pool of the NCP units. To reconstitute the chromatin-like higher order structures, a pool of 100 of the above constructed fibers was prepared by using low- and high-salt angles and unwrapping parameters. These fibers were then sequentially fitted into a spherical volume density in a diameter of 200 nm with a minimal distance equal to the distance between the i and i+2 NCPs measured at corresponding conditions. In brief, these fibers were converted into low-passed filtered density maps with a specific threshold cutoff value, which determined the mean width of the fiber. The produced fiber density maps were randomly fitted into the spherical volume by using Chimera sequential fitting function without clashing. The same rotation and translation of the maps were imposed on the corresponding fiber models, in which the central cubic volume in 20 nm length was cropped out for displaying the nucleosome organization of the chromatin-like region.

3.5.17 Classification of the nucleosome array conformations

To identify some of the low energy states that are populated with more similar array conformations, nucleosome arrays were classified based on their structural similarity. For each type of the array, 30 to 40 conformations were obtained and subjected to a pair-wise alignment through minimization of the root-mean-square deviation (RMSD) between each pair of models. Based on the values of the constructed distance matrix evaluated by RMSD, the array conformations were sorted and classified with an optimization algorithm that minimized the tree spanning (method option =

'single') using the hierarchical clustering in Scikit-Learn of Python package.(Bar-Joseph et al., 2001) The final result was displayed in a dendrogram produced by the scipy.cluster.hierarchy package.

3.5.18 Visualization of the structure dynamics of the nucleosome array

The pseudo dynamics of nucleosome array was simulated by morphing through array conformations in an order followed by the hierarchical clustering. The morphing begins with the structures at the lowest branch of the dendrogram and stops at the highest branch. By using Targeted Molecular Dynamics simulation (TMD) of the NAMD2, the coarse-grained structure was steered from one conformation toward to another under amber SIRAH force field. The moving forces used in the TMD were pre-calculated based on the distances of the corresponding CG model atoms. After 1,000,000 to 4,000,000 steps (corresponding to a total time of 20 ns to 80 ns, with each step of 20 fs) at 298 K temperature within the implicit solvent, the simulation was terminated when the real-time RMSD fell below 3 Å. Among these morphing structure pairs, ~10% of them were eliminated due to the DNA intertwining with each other. The morphing movie was also displayed with the same order as above from the most populated structure states to the rarest states.

3.5.19 Stiffness analysis of DNA and in-silico nucleosome array fiber

The persistence length of free DNA was estimated from the calculation of the autocorrelation decay along the two distal DNA arms of the array model. By using the polymer.PersistenceLength.run script from MDAnalysis (Michaud-Agrawal et al., 2011), the chain is approximated as a series of discrete segments of equal length along the contour length of DNA. The segment length (r) was defined by the distance between two weight centers of backbone atoms of two consecutive base pairs. The persistence length was obtained from the orientation correlation of segments along the DNA, which are expressed as a scalar product of two segment vectors separated by a contour length of L and are predicted from the random coil statistics as an exponential decay governed by the persistence (Cifra, 2004). Similarity, to estimate the persistence length of longer arrays containing 100 NCPs, the NCP pathway of the array was converted into smoothly connected fixed length line segments (15 nm). The persistence lengths estimated for the distal DNA arms and nucleosome array fibers were averaged form ~30, and 100 models, respectively.

type	Mono-Nuc	Di-Nuc	Tri-Nuc	Tetra-Nuc	Tetra-Nuc-50	Tetra-Nuc-H1
θ	46.5 ± 27.0 °	55.0 ± 25.0 °	51.9 ± 27.6 °	57.9 ± 35.2 °	48.2 ± 19.6 °	69.7 ± 39.8 °
θ ∥	10.4 ± 44.1 °	-28.7 ± 45.3 °	-0.9 ± 49.1 °	7.4 ± 58.3 °	-31.3 ± 35.2 °	-2.8 ± 81.3 °
θ ⊥	32.3 ± 44.7 °	33.2 ± 38.2 °	36.6 ± 48.5 °	31.2 ± 67.8 °	22.0 ± 33.8 °	18.3 ± 76.6 °

type	position	Mono-Nuc	Di-Nuc	Tri-Nuc	Tetra-Nuc	Tetra-Nuc-50	Tetra-Nuc-H1
θ	Distal	46.5 ± 27.0 °	55.0 ± 25.0 °	51.0 ± 28.9 °	59.4 ± 37.5 °	49.3 ± 22.4 °	60.0 ± 39.1 °
	Interior			53.7 ± 25.0 °	57.5 ± 33.5 °	47.4 ± 17.2 °	79.4 ± 38.5 °
θ ∥	Distal	10.4 ± 44.1 °	-28.7 ± 45.3 °	-3.9 ± 49.2 °	3.9 ± 63.3 °	-27.4 ± 42.9 °	-24.9 ± 68.4 °
	Interior			5.0 ± 49.0 °	9.7 ± 55.1 °	-34.4 ± 27.7 °	19.4 ± 87.4 °
θ ⊥	Distal	32.3 ± 44.7 °	33.2 ± 38.2 °	34.1 ± 50.6 °	28.5 ± 68.5 °	20.7 ± 40.4 °	12.2 ± 65.2 °
	Interior			41.6 ± 44.0 °	33.0 ± 67.6 °	23.0 ± 27.7 °	24.3 ± 86.6 °

side	Mono-Nuc	Di-Nuc	Tri-Nuc	Tetra-Nuc	Tetra-Nuc-50	Tetra-Nuc-H1
Long -arm DNA	-5 ± 5 bp	-4 ± 5 bp	-7 ± 7 bp	-4 ± 8 bp	-4 ± 9 bp	-6 ± 8 bp
Short-arm DNA	-11 ± 13 bp	-8 ± 10 bp	-10 ± 11 bp	-14 ± 12 bp	-9 ± 8 bp	-17 ± 19 bp

Group	twoGauss	Mono-Nuc	Di-Nuc	Tri-Nuc	Tetra-Nuc	Tetra-Nuc-50	Tetra-Nuc-H1
Alpha	0	39.0 ± 5.4 °	31.8 ± 11.4 °	32.2 ± 13.9 °	31.8 ± 14.0 °	-13.6 ± 14.4 °	26.6 ± 13.6 °
Alpha	1	14.8 ± 15.8 °	-5.0 ± 14.0 °	-0.1 ± 16.2 °	-6.4 ± 18.5 °	29.7 ± 14.1 °	-15.7 ± 13.6 °
Beta	0	-9.0 ± 11.2 °	-7.6 ± 11.3 °	-8.8 ± 12.3 °	-9.4 ± 14.1 °	-4.4 ± 11.6 °	-0.8 ± 13.0 °

Group	twoGauss	Di-Nuc	Tri-Nuc-Nuc	Tetra-Nuc	Tetra-Nuc 50	Tetra-Nuc- H1
i, i+1	0	224.9 ± 34.6 Å	237.3 ± 27.5 Å	240.9 ± 37.9 Å	208.7 ± 23.7 Å	203.8 ± 19.9 Å
i, i+1	1					293.8 ± 41.2 Å
i, i+2	0		274.2 ± 89.0 Å	301.7 ± 89.6 Å	155.1 ± 54.1 Å	204.9 ± 40.8 Å
i, i+2	1					416.4 ± 90.1 Å
i, i+3	0			372.3 ± 121.3 Å	273.4 ± 31.0 Å	394.0 ± 106.7 Å

Table 1 Statistics of the array dynamics. The measured mean and standard deviation of DNA arm/linker angle θ and its two orthogonal projections $\boldsymbol{\theta}_{\parallel}$ and $\boldsymbol{\theta}_{\perp}$ (first panel), the distal and interior NCPs angle θ, $\boldsymbol{\theta}_{\parallel}$ and $\boldsymbol{\theta}_{\perp}$ (second panel), NCP unwrapping length (third panel), NCP wrapping angle α and bending angle β (fourth panel), and NCP core-to-core distance (fifth panel).

Chapter 4
Molecular organization of the early stages of nucleosome phase separation visualized by cryo-electron tomography

The following work is an adapted version of the submitted work: Zhang, M.[*]; Díaz-Celis, C.; Onoa, B.; Cañari- Chumpitaz, C.; Requejo, K.; Liu, J.; Vien M.; Nogales, E.; Ren, G.; Bustamante, C. "Molecular Organization of the Early Stages of Nucleosome Phase Separation Visualized by Cryo-Electron Tomography" *Molecular Cell* **2021**.

([*] denotes equal contribution)

4.1 Abstract

It has been proposed that the intrinsic property of nucleosome arrays to undergo liquid-liquid phase separation (LLPS) in vitro is responsible for chromatin domain organization in vivo. However, understanding nucleosomal LLPS has been hindered by the challenge to characterize the structure of resulting heterogeneous condensates. We used cryo-electron tomography and deep learning-based 3D reconstruction/segmentation to determine the molecular organization of condensates at various stages of LLPS. We show that nucleosomal LLPS involves a two-step process: a spinodal decomposition process yielding irregular condensates, followed by their unfavorable conversion into more compact, spherical nuclei that grow into larger spherical aggregates through accretion of spinodal material or by fusion with other spherical condensates. Histone H1 catalyzes more than 10-fold the spinodal-to-spherical conversion. We propose that this transition involves exposure of nucleosome hydrophobic surfaces resulting in modified inter-nucleosome interactions. These results suggest a physical mechanism by which chromatin may transition from interphase to metaphase structures.

4.2 Introduction

The genome of all eukaryotic cells is organized into chromatin, a large protein-DNA complex whose structure regulates access to processes such as transcription, replication, and DNA repair. The nucleosome is the basic structural unit of chromatin consisting of 146 bp of DNA wrapped 1.65 turns around a histone octameric core (Luger et al., 1997b). A diverse set of post-translational modifications of histones and chromatin associated proteins modulate the structural dynamics of chromatin (Bowman and Poirier, 2015), which form two distinct subdomains in the cell nucleus: a lightly packed euchromatin state, associated with regions of active transcription, and a condensed heterochromatin state, associated with transcriptionally repressed regions.

Two models have been proposed to explain how chromatin transitions from its transcriptionally active to its inactive form. One model identifies the euchromatin regions with the 10-nm fiber, also referred to as "beads-on-a-string". This fiber can fold in a compact and helical structure displaying a 30 nm diameter (Song et al., 2014; Robinson et al., 2006; Finch and Klug, 1976b; Widom and Klug, 1985) that has been proposed to correspond to the heterochromatin domains. In this hierarchical folding model, chromatin condensation occurs through various super-helical structural intermediates formed at different stages of the cell cycle (Belmont and Bruce, 1994). However, this model has been challenged, since *in vivo* studies have failed to reveal regular 30-nm diameter helical fibers (Ou et al., 2017; Eltsov et al., 2008; Cai et al., 2018a; Maeshima et al., 2014b; Maeshima et al., 2014c; Razin and Gavrilov, 2014; Eltsov et al., 2018; Nishino et al., 2012).

An alternative model for chromatin organization has been proposed recently based on the intrinsic property of chromatin to form liquid-like droplets by liquid-liquid phase separation (LLPS), both *in vitro* and *in vivo* (Maeshima et al., 2016b; Gibson et al., 2019; Strom et al., 2017; Sanulli et al., 2019). LLPS is a physical process that underlies the generation of spatially separated membrane-less domains inside the cell (Hubstenberger et al., 2017; Boeynaems et al., 2018; Banani et al., 2017; Sanulli and G, 2020; Brangwynne et al., 2009; Erdel and Rippe, 2018; Hyman et al., 2014). These condensed phases have been described as LLPS because they appeared under the optical microscope as liquid droplets (Boeynaems et al., 2018) that coexist with a dilute phase. In this new model, phase transition of chromatin into droplets drives compartmentalization and organization of long-lasting multiphase systems (Palikyras and Papantonis, 2019; Shakya et al., 2020), which has been posited to control chromatin accessibility by the cellular machinery (Palikyras and Papantonis, 2019; Shin et al., 2018; Wright et al., 2019). *In vitro*, nucleosome phase transition is regulated by nucleosome concentration, nucleosome array length, DNA linker spacing, ionic strength, and chromatin-associated

proteins such as heterochromatin protein 1 (HP1) and linker histone H1 (Maeshima et al., 2016b; Gibson et al., 2019; Sanulli et al., 2019). However, LLPS studies have been circumscribed mainly to the use of fluorescence microscopy, which is limited to study the late-stages of the phase separation process when the condensates have grown to micron size dimensions. A number of questions remain unanswered about the phase separation: first, what is the physical process through which macroscopic domains arise from the molecular components? Do phase separation and formation of the condensates proceed through a classical mechanism of nucleation and growth? Moreover, if condensates are to function as compartments to impede or facilitate specific enzymatic reactions, chromatin droplets must be able to switch between loosely and tightly packed structures (Bancaud et al., 2009; Imai et al., 2017). What is then the arrangement of nucleosomes inside the condensates capable of this dynamic? Is the spatial distribution homogeneous or do channels and/or chambers exist that permit molecular diffusion?

Addressing the above questions requires a technique capable of imaging each nucleosome within the condensates; yet, given their heterogeneous nature, structural averaging methods such as X-ray crystallography or single-particle reconstruction cannot be used. However, the improvement of cryo-Electron Tomography (cryo-ET) as a single-molecule 3D imaging technique (Zhang and Ren, 2012; Ercius et al., 2015; Chen et al., 2017; Narayan and Subramaniam, 2015) has made it possible to obtain 3D structures at molecular resolution (Ou et al., 2017; Lei et al., 2018b; Zhang et al., 2016b; Jordan et al., 2018). Here, we have used cryo-ET reconstruction coupled to deep learning-based denoising and segmentation processes to perform a systematic study of the earliest stages of nucleosomal phase separation *in vitro* using reconstituted tetranucleosome arrays. Tetranucleosome arrays have been identified as the minimal unit capable of generating large-scale chromatin structures under conditions that induce phase separation (Song et al., 2014; Ding et al., 2021; Schalch et al., 2005a).

We find that the initial stages of phase separation take place through a two-step condensation process. The first step resembles a process of spinodal decomposition in which a nucleosome-rich phase composed of irregularly shaped, loosely packed condensates emerges throughout the medium in physiological salt. In the second step, we see the emergence, within the spinodal condensates (SpnCs), of spherical nuclei (as small as ~35 nm), which appeared to grow into larger spherical condensates (SphCs) through the accretion of nearby spinodal material or through fusion with other SphCs. The spherical shape of these aggregates indicates that their formation is accompanied by an initial unfavorable surface free energy that must be minimized relative to their favorable volume energy. Moreover, the presence of H1 linker histone does not prevent the formation of the initial SpnC but catalyzes its transition into SphC more than 10-fold.

4.3 Thermodynamics of liquid phase separation

4.3.1 Part A: Theory of liquid phase separation

Phase separation is a non-equilibrium thermodynamic process in which a uniformly mixed system can lower its free energy (*G*) by segregating into two or more phases with distinct compositions that reach chemical equilibrium. The conditions at which distinct phases occur and coexist at equilibrium are described in a phase diagram. How do these conditions fulfill minimization of free energy (G) and satisfy the criteria for phase equilibrium? For a binary system exhibiting phase separation, the Gibbs free energy of a two-component solution as a function of composition exhibits a region of downward concavity or a negative curvature $\left(\frac{\partial^2 G_{solution}}{\partial {x_B}^2} < 0\right)$ in which the mixed system is unstable (Figure 4.1A). In this instability region, the single-phase system spontaneously undergoes demixing into two phases α and β, whose compositions must fulfill the criterion for phase equilibrium, i.e., that the chemical potential of each component of the binary mixture is the same in both phases (See equation 1, part B). Since the chemical potential of component *i*, μ_i, is the first derivative of the free energy with respect to the concentration of component *i*, n_i, the equilibrium condition implies that $\boldsymbol{\mu_i}^{\boldsymbol{\alpha}} = \frac{\partial G}{\partial n_i} = \boldsymbol{\mu_i}^{\boldsymbol{\beta}}$, which can be obtained graphically at every temperature by finding the concentrations of both components for which the free energy has a common tangent (Figure 4.1A) (Clerc and Cleary, 1995). The common tangent defines the points of the binodal curve, (Figure 4.1B) which, in turn, defines the end point composition of phase separation, which may occur by different mechanisms.

i. Spinodal Decomposition vs Nucleation and Growth

At compositions where $\left(\frac{\partial^2 G_{solution}}{\partial {x_B}^2} < 0\right)$, the mixed or single-phase system is unstable to both small and large fluctuations in composition. In this case, phase separation takes place without having to cross an energetic barrier and, therefore, it will occur throughout the entire system in a process known as spinodal decomposition (Figure 4.1B). A detailed description of the equations involved in spinodal decomposition first formulated by Cahn and Hilliard (Cahn and Hilliard, 1958, 1959) is presented in the part B.

On the other hand, in the region in which the free energy as a function of composition has a positive curvature $\left(\frac{\partial^2 G_{solution}}{\partial {x_B}^2} > 0\right)$ (Figure 4.1A), the mixture is said to be metastable; in this region, phase separation only happens in discrete locations throughout the mixture where some rare, large fluctuations in composition

spontaneously occur, because the system must overcome an energetic barrier. These localized regions, which appear sparsely in the mixture, are called nucleation sites (Clerc and Cleary, 1995; Schmelzer et al., 2004). Classical Nucleation Theory (CNT) describes phase separation in terms of a process involving the formation of a thermodynamically unfavorable phase separated 'nucleus' and an increasingly thermodynamically favorable process of 'growth'. In CNT, the processes of nucleation and growth involve the interplay between the cost of an interfacial surface energy, or surface tension, and a favorable volume energy resulting from molecular interactions in the segregated phase that permits the nucleus to grow by the addition of more material to the new phase. (A general description of the parameters involved in the classical nucleation theory is presented in the Theory Section S1). However, in complex biological systems, such as the phase separation of lysozymes during crystallization, the CNT does not predict correctly the experimentally measured crystallization rates, which are ten orders of magnitude higher than the values predicted by CNT (Erdemir et al., 2009; Vekilov, 2010; Loh et al., 2017a). In these cases, non-classical nucleation models have been proposed, in which an additional step involving, for example, spinodal decomposition occurs before nucleation. For instance, a three-step mechanism has been proposed for the solidification and crystallization of gold nanocrystals based on liquid cell TEM studies (Loh et al., 2017b), which predicts that spinodal decomposition occurs before solidification and crystallization. Specifically, first, gold-rich spinodal structures form that later condense into amorphous nanoclusters, which crystallize into nuclei that can support nanoparticle growth because of their stable size (Loh et al., 2017a; Ji et al., 2007; Pong et al., 2007).

4.3.2 Part B: Principal equations of liquid phase transition

In a multicomponent system of two or more phases, the phase equilibrium criterion is that the chemical potential of a given component *i* must be equal in all phases where *i* is present:

$$\boldsymbol{\mu_i^{\alpha}} = \boldsymbol{\mu_i^{\beta}} = \boldsymbol{\mu_i^{\gamma}} \ldots \text{with } \boldsymbol{\mu_i} = \left(\frac{\partial G}{\partial n_i}\right)_{T,P,n_{j\neq i}} \qquad (1)$$

where $\boldsymbol{\mu_i}$ represents the chemical potential of the *i*th component and $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \ldots$ represent the different phases where *i* is present at equilibrium (Clerc and Cleary, 1995). The compositions of those two phases $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ (expressed as the mole fraction of component B in these phases in figure 4.1B) are fixed at a given temperature but change as a function temperature. The collection of points defining the composition as a function of temperature at which these two phases coexist defines the binodal curve in the phase diagram (Figure 4.1B) (Clerc and Cleary, 1995; Alberti et al., 2019; Shin and Brangwynne, 2017).

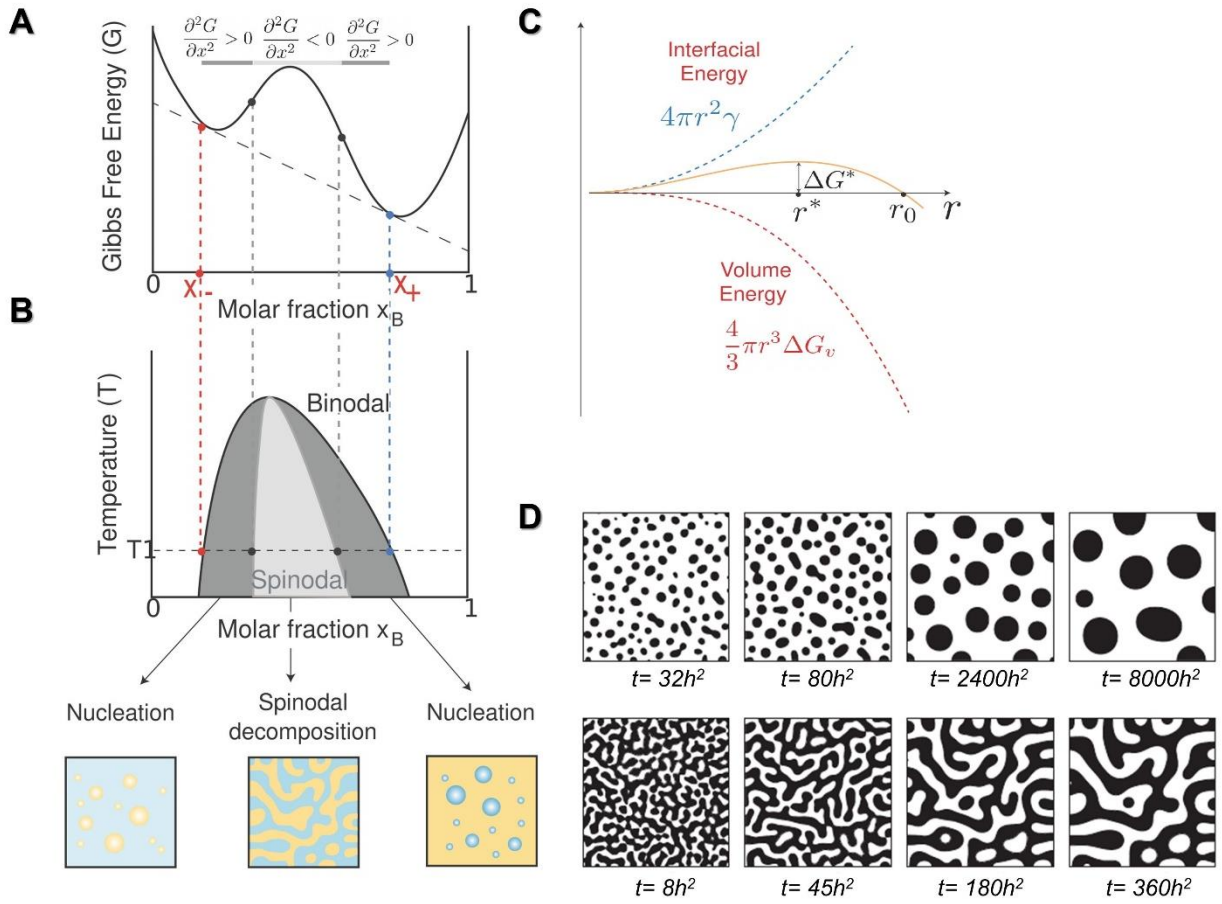


Figure 4.1 Schematics of phase transition mechanisms. A) Gibbs' free energy as a function of molar fraction for a system that shows phase separation for a composition range $X_- < X_B < X_+$. The straight dashed line is the only common tangent. B) The light gray color area is the instability region in which the system undergoes demixing through spinodal decomposition. In the area in between the binodal and spinodal, the system demixes via nucleation and growth (Clerc and Cleary, 1995; Alberti et al., 2019). C) Gibbs free energy diagram for nucleation indicating the presence of a critical nucleus (r^*) and a nucleation barrier ($\boldsymbol{\Delta G^*}$) to overcome for growth to occur (Karthika et al., 2016). D) Examples of spinodal decomposition predicted by the CH equation with different values of average composition and double-well potential function ($\bar{\boldsymbol{g}}(\boldsymbol{\phi}) = 0.25(\boldsymbol{\phi^2 - 1})^2$), where $\boldsymbol{\phi}$ is the composition of the system. The timescale h is a parameter used in the simulation and has the value of spatial step (similar to the area splitted in a grid). For this particular simulation, h= 0.03 (Image was adapted from (Kim et al., 2016)).

i. Classical Nucleation Theory

Nucleation is the process by which a different thermodynamic phase with low free energy is formed from a parent phase with high free energy. The most common theoretical model to describe nucleation is the classical nucleation theory (CNT) that was formulated to explain the condensation of vapor into a liquid and that can be used also in liquid-solid equilibrium systems. In homogeneous nucleation, large supersaturations are usually needed to initiate the process because there are no preferential nucleation sites and the nuclei tend to form with equally low probability anywhere in the original mix (Karthika et al., 2016; Thanh et al., 2014).

For a spherical nucleus with radius *r*, the Gibbs free energy change $\boldsymbol{\Delta G(r)}$ comprises the contributions of the volume free energy and the interfacial or surface energy cost of creating a nucleus inside the other phase (Equation 2; Figure 4.1C),

$$\boldsymbol{\Delta G(r)} = \frac{\boldsymbol{4\pi}}{\boldsymbol{3}}\boldsymbol{r^3\Delta G_v} + \boldsymbol{4\pi r^2\gamma} \quad (2)$$

where $\boldsymbol{\gamma}$ is the surface tension around the nucleus and $\boldsymbol{\Delta G_v}$ is the free energy per unit volume gained from molecular interactions. The unfavorable interfacial energy dominates at small nucleus sizes *r*, while the favorable volumetric contribution to the nucleation barrier dominates at large nucleus sizes. The critical size or radius (*r**) for the nucleus corresponds to the minimum size beyond which the particle can be present in solution without being redissolved (Figure 4.1C). From equation (2), $\boldsymbol{r^*} = -\frac{\boldsymbol{2\gamma}}{\boldsymbol{\Delta G_v}}$. As shown in figure 4.1C, $\boldsymbol{\Delta G^*}$ indicates the nucleation barrier for the appearance of the separated phase. Clusters with values below *r** tend to redissolve and the ones with values above *r** tend to grow, thus, beyond *r**, $\boldsymbol{\Delta G(r)}$ decreases with increasing *r* and at $\boldsymbol{r_0} = -\frac{\boldsymbol{3\gamma}}{\boldsymbol{\Delta G_v}}$, $\boldsymbol{\Delta G(r)} = \boldsymbol{0}$. For $\boldsymbol{r > r_0}$, $\boldsymbol{\Delta G(r)}$ is negative and particle growth leads to the formation of a new phase. Thus, phase separation from these nuclei can proceed through a process of growth that takes place when the surface energy price of creating a discrete phase in the medium is overcome by the energy gained inside the nucleus by the addition of more material to the phase. Several authors have estimated, using experimental data and simulations, that the size of the critical nucleus is in the range of ~10-1,000 molecules. For instance, Yau and Vekilov reported that the size of the critical nucleus of protein apoferritin is about 40 nm in aqueous solution (Karthika et al., 2016; Yau and Vekilov, 2001).

ii. Spinodal Decomposition and the Cahn-Hilliard Equation

To evaluate the thermodynamics and kinetics underlying spinodal decomposition,

it is necessary to use a theoretical framework that incorporates fluctuations in the composition field $\boldsymbol{c(x,t)}$. The Cahn-Hilliard (CH) model considers spatial inhomogeneities by adding a correction to the free energy function of a homogeneous medium. Based on the Ginzburg-Landau free energy theory (Cahn and Hilliard, 1958, 1959; Lee et al., 2014), the total free energy of a volume *V* of an isotropic system with a non-uniform composition is given by equation (3):

$$\boldsymbol{G[c(x,t)] = N_V \int [\bar{g}\big(c(x,t)\big) + \frac{1}{2}\kappa\big(\nabla c(x,t)\big)^2]dV} \quad (3)$$

where $\boldsymbol{N_V}$ is the number of molecules per unit volume, $\boldsymbol{\bar{g}(c(x,t))}$ is the free energy per molecule of the homogenous system,$\boldsymbol{\nabla c(x,t)}$ is the local composition gradient, and $\boldsymbol{\kappa}$ is a parameter that controls the free energy cost of variations in concentration. Thus, the free energy of a volume of a non-uniform solution can be expressed as the sum of the free energy that this volume would have in a homogenous solution, and a gradient energy that is a function of the local composition. In this theoretical framework, the chemical potential is redefined as the functional derivative of *G* (equation (3)):

$$\boldsymbol{\mu \equiv \frac{\delta G}{\delta c} = \bar{g}'(c(x,t)) - \kappa\nabla^2 c(x,t)} \quad (4)$$

where $\boldsymbol{\bar{g}'}$ is the first derivative of the free energy per molecule with respect to concentration. During phase separation, fluctuations in the concentration of components will lead to the increase of concentrations of certain components at the expense of depletions on other regions. Therefore, a flux of components will be produced. The flux of components in mixture ($\boldsymbol{J}$) can be described by an alternative formulation of Fick's first law considering the variations in chemical potential (μ) (equation (5)):

$$\boldsymbol{J \equiv -M\nabla\mu} \quad (5)$$

where *M* plays a role of the mobility (similar to the diffusion constant D used in Fick's first law) defined as an interface parameter that indicates a measure of the transport kinetics across the interface of the phases (Kim et al., 2016). Using a continuity equation (conservation of mass):

$$\boldsymbol{\frac{\partial c(x,t)}{\partial t} = -\nabla \cdot J} \quad (6)$$

and by replacing equation (5) in the flux term ($\boldsymbol{J}$), we obtain:

$$\boldsymbol{\frac{\partial c(x,t)}{\partial t} = M\nabla^2\mu} \quad (7)$$

The replacement of equation (4) in equation (7) generates the Cahn-Hilliard equation, which describes the temporal evolution of the concentration field $\boldsymbol{c(x,t)}$ during spinodal decomposition:

$$\frac{\partial c(x,t)}{\partial t} = \nabla \cdot [M\nabla\left(\bar{g}'\left(c(x,t)\right) - \kappa\nabla^2 c(x,t)\right)] \quad (8)$$

The CH equation is nonlinear and usually requires numerical approximations such as finite element, finite difference, or spectral methods to solve it. Spectral methods are common methods involving the use of fast Fourier transforms to solve differential equations. Computational simulations on the evolution of spinodal decomposition depicted by the CH equation are shown in figure 4.1D (Kim et al., 2016; Lee et al., 2014).

Some general implications can be obtained by evaluating particular conditions. For instance, for the double-well potential Gibbs free energy function (as the one depicted in figure 4.1A), the concentration profile on the interface between the phases is diffuse and has a sigmoidal shape. In particular, for the one-dimensional case, it has the form:

$$c(x) = tanh(\frac{x}{l})$$

with $l = \Delta c\sqrt{\frac{\kappa}{\Delta g}}$ representing the interfacial width, where $\Delta c = x_+ - x_-$, the width of the miscibility gap and where $\Delta g \geq kT$. By evaluating how the CH equations responds to small fluctuations in concentration of form $\delta c = Acos(kx - \omega t)$ where k is given by $2\pi/\lambda$ and where λ is the wavelength of the concentration perturbation (as in a Fourier series), it can be shown that the critical wavelength has the following expression:

$$\lambda_{max} = \frac{2\pi}{k_{max}} = 2\pi\sqrt{\frac{2\kappa}{-g''(c_0)}} \quad (9)$$

where g'' is the second derivative of the free energy with respect to concentration and the critical or maximum wavelength (λ_{max}) is the most unstable wavelength in which random fluctuations decompose into characteristic patterns (Figure 4.1D). There are two important conclusions from equation (9): i) λ_{max} only has a physical meaning when $g''(c_0)<0$ (a requirement for spinodal decomposition as discussed previously), and ii) λ_{max} gives a sense of the scale of the fluctuations and the size of domains formed during spinodal decomposition.

When one phase is at higher concentration, the Cahn–Hilliard equation predicts the process of Ostwald ripening, which is a thermodynamically-driven spontaneous process where the lower concentration phase forms spherical particles, and the smaller particles are absorbed through diffusion into the larger ones. The radii of these particles grow in time as in $t^{1/3}$(Lifshitz–Slyozov law) and the volume grows linearly with time. This process is favorable because the smaller particles are less energetically stable due to their high surface to volume ratio meaning high surface energy (Kim et al., 2016; Lee et al., 2014).

4.4 Results

4.4.1 Observation of the early stages of condensation by cryo-electron tomography

To study chromatin LLPS, we used tetranucleosome arrays obtained from the assembly of a DNA fragment that contains four 601-nucleosome positioning DNA sequences separated by 40 bp linkers with recombinant histone octamers (*Xenopus laevis*) (Figure 4.2A). Nucleosome phase transition time-course experiments started with the dilution of the tetranucleosomes from a low ionic strength buffer into a physiological salt concentration buffer, at a final concentration of 30 nM (See Methods). To establish the optimal protocol of deposition and time required for condensate formation, samples incubated for different times were initially subjected to optimized negative staining (OpNS) (Rames et al., 2014b) and examined by EM.

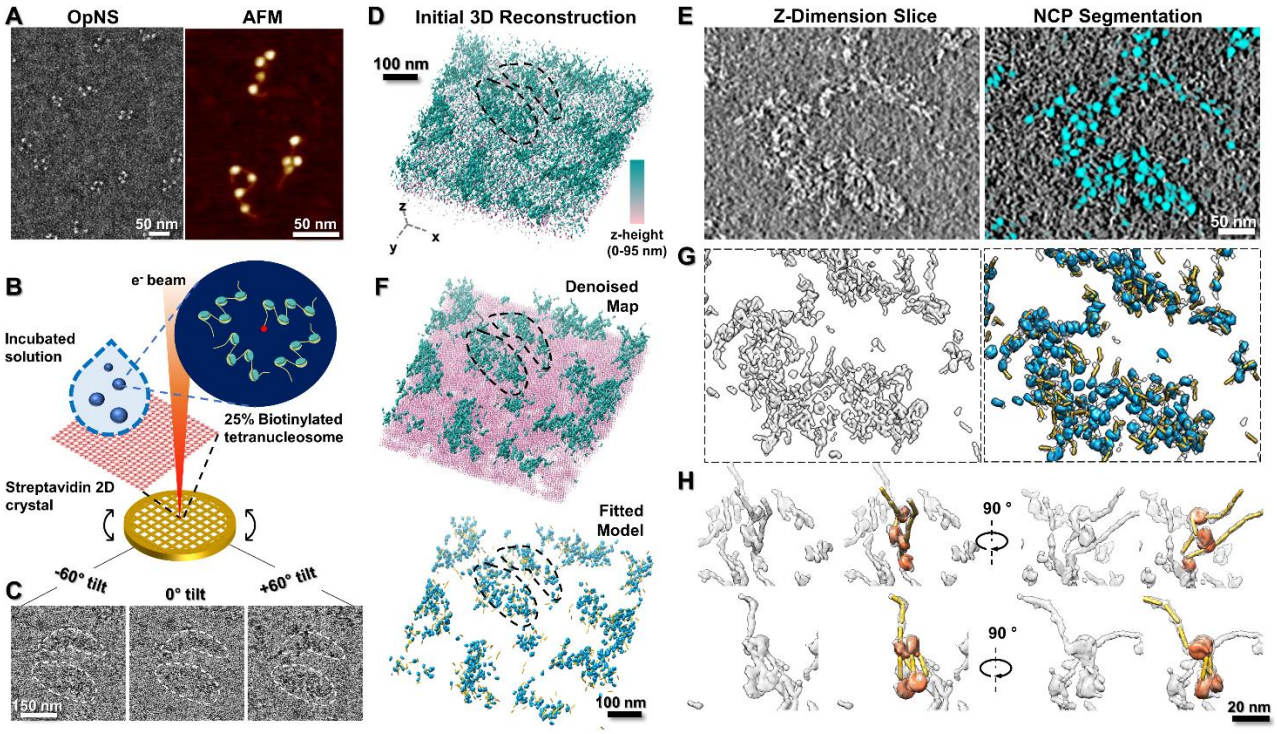


Figure 4.2 Cryo-ET workflow of deposition, reconstruction, and modeling of tetranucleosome phase condensation. A) Optimized negative staining EM and atomic force microscopy images of tetranucleosomes. B) Deposition of 30 nM (25% biotinylated; red sphere) tetranucleosome incubation solution onto the SA-crystal. C) Representative tilt series after deep learning-based denoising. The white dashed contours indicate the tracking of two clusters of condensates. D) Initial 3D reconstruction of the denoised tilt series after alignment, with z-dimensional height encoded in color. E) Central slice of the initial map (thickness of 1.17 nm, left) and corresponding deep learning-based segmentation showing predicted NCPs (cyan, right). F) Missing wedge-corrected final denoised map showing NCPs (cyan) and SA-crystal (pink) density. This final map was used to generate the fitted model (bottom). Blue and yellow depict NCPs and the 40 bp linker DNA, respectively. G) Comparison of the final 3D map (left) and the fitted model (right) of the tracking area in C-F. H) Local areas showing resolved tetranucleosomes in extended (top panels) or stacked (bottom panels) conformations.

At the earliest incubation and deposition time at 2 min, 20°C, and in physiological salt buffer (150 mM NaCl and 5 mM $MgCl_2$), we observed distinct nucleosome condensates adopting irregular shapes (Figure 4.3A and B). These condensates were not observed in low salt buffer (1.5 mM NaCl and no $MgCl_2$; Figure 4.3C). Interestingly, after 10 min of incubation, small globular condensates (~35-40 nm in diameter) displaying higher nucleosomal density, appeared sparsely distributed among the irregularly

shaped nucleosome structures (Figure 4.3A and D). The shapes of these denser condensates were also not uniform. Some of them were rounded, while others were elongated and composed of a dense, rounded condensate connected with an irregular region (Figure 4.3D and E, white arrow).

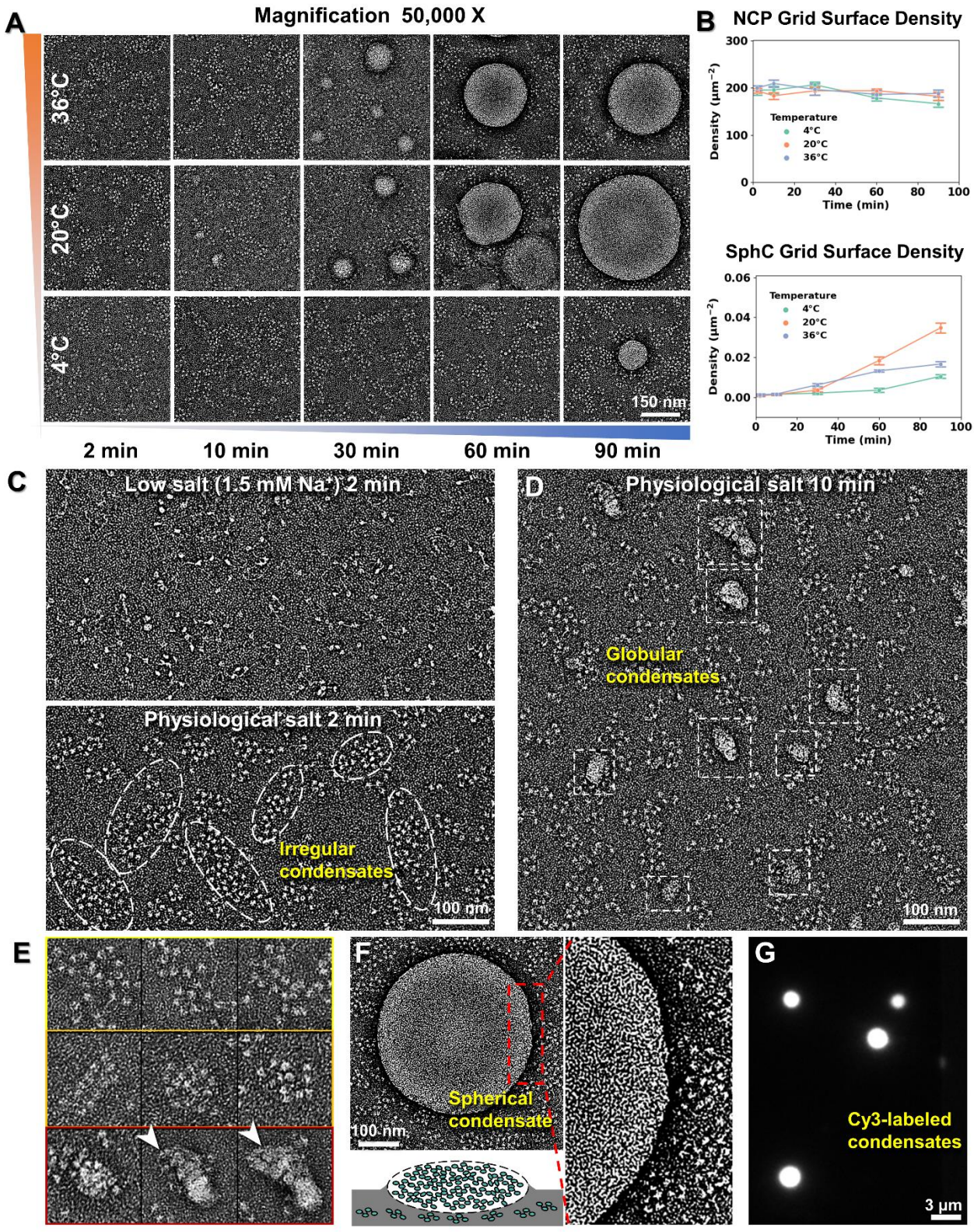


Figure 4.3 Tetranucleosome phase transition visualized by OpNS and epifluorescence imaging. A) Representative time series of 30 nM tetranucleosome in physiological salt

(20 mM HEPES-KOH pH 7.5; 150 mM NaCl; 5 mM MgCl2; and 1 mM DTT) at 4°C, 20°C, and 36°C imaged at magnifications of 50,000 X. (B) Statistics of nucleosome number density (number of nucleosomes within irregular condensates divided by the corresponding imaging area) and SphC size distribution. Data was obtained from 12-16 12,500 X images and statistics are presented as mean ± SEM. (C) Tetranucleosomes visualized by OpNS in low salt (1.5 mM Na+, top) and physiological salt (bottom) after 10 min incubation at 20°C. Dashed-line ellipsoids enclose condensates displaying an irregular shape and loosely packed nucleosomes. (D) Tetranucleosome condensates formed in physiological salt buffer after 10 min incubation at 20°C. Dashed-line boxes mark globular condensates displaying tighter nucleosomal packaging. (E) Representative images of nucleosome clusters in physiological salt showing different degrees of condensation from loosely packed structures to tightly packed condensates (top to bottom). Box size= 100 nm. (F) OpNS image of a SphC (left top panel) stained after 90 min, whose edge has been enlarged (right panel) to see the tightly packed nucleosomal pattern on its surface. The high density of the SphC prevents visualization of its interior and the irregular condensates below it as depicted in schematic (bottom panel). (G) Fluorescence microscopy imaging of Cy3-labeled nucleosome liquid droplets formed in physiological salt after 30 min incubation at 20°C.

After 30 min of incubation, we observed larger spherical condensates (SphCs > 100 nm diameter), which tended to increase in number and size with longer incubation times (Figure 4.3A and B). Compared to the surrounding irregular condensates, the SphCs appeared composed of tightly packed nucleosome core particles (NCPs) (Figure 4.3F). Statistical analysis of the SphC grid surface density (Figure 4.3B) showed that their formation rate is faster at 20°C, followed by 36°C and then 4°C. Epifluorescence microscopy images of samples in the same conditions (Figure 4.3G) showed fusion events (Figure 4.4A, top) indicating that the condensates observed by OpNS behave like liquid droplets previously described (Gibson et al., 2019). The liquid-like behavior of the tetranucleosome condensates was further confirmed by fluorescence recovery after photobleaching (FRAP) experiments ($t_{1/2}$ = 1 min) (Figure 4.4A, bottom), which showed faster diffusion of these tetramer arrays compared to those previously described for dodecamers ($t_{1/2}$ = 6 min) (Gibson et al., 2019). To confirm that the two-step process of condensation is a property of nucleosome arrays, independently of their length, we repeated the same incubation time series described above but with dodecamer arrays (Figure 4.4B). These samples displayed the same two-step condensation process, with an initial formation of irregular arrays, followed by the appearance of SphCs (Figure 4.4C). The formation of SphCs observed with dodecamer arrays and the concomitant depletion of the irregular condensates was ~10-fold faster than with the tetranucleosome arrays.

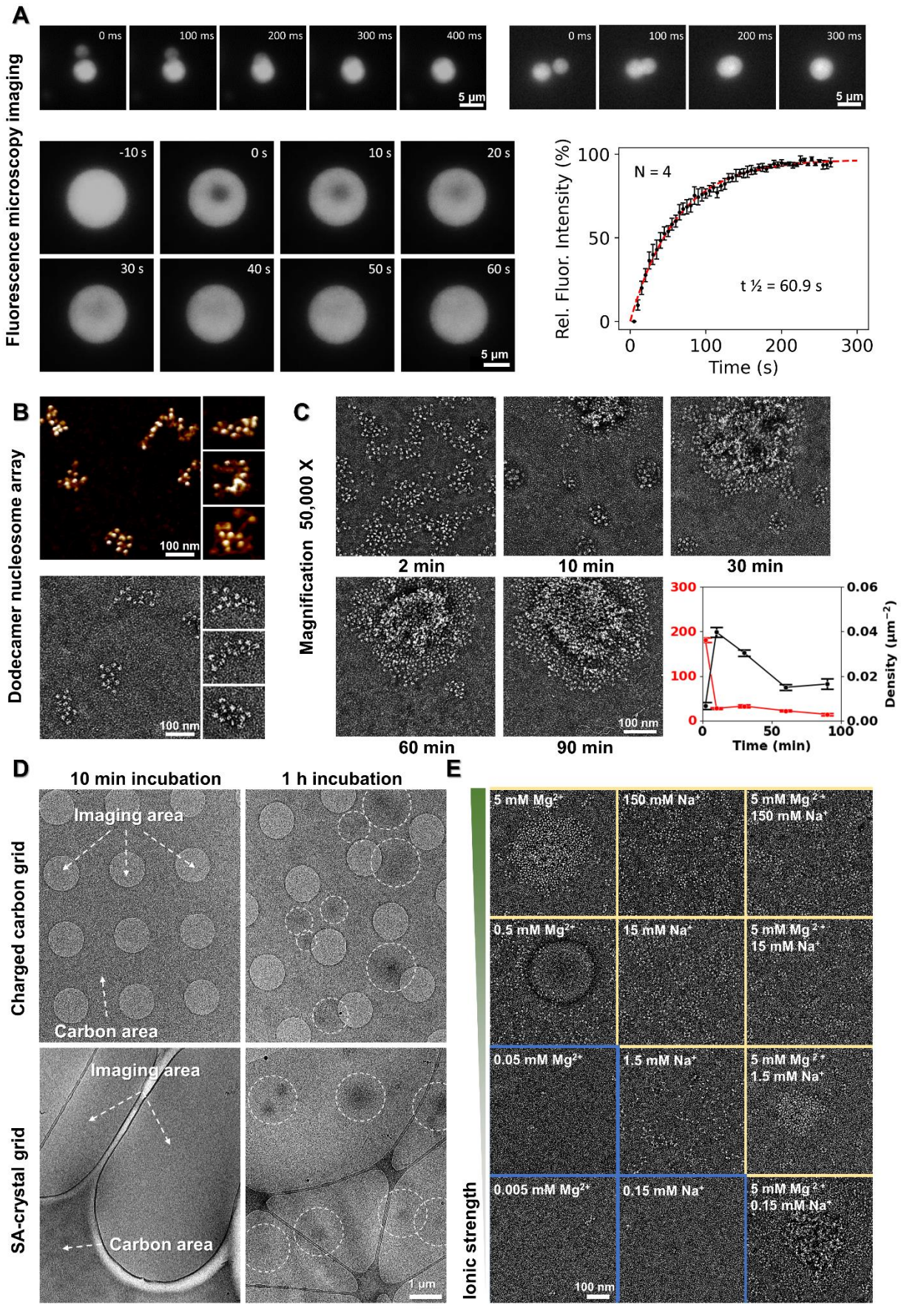

A
Fluorescence microscopy imaging
0 ms
100 ms
200 ms
300 ms
400 ms
5 µm
0 ms
100 ms
200 ms
300 ms
5 µm
-10 s
0 s
10 s
20 s
30 s
40 s
50 s
60 s
5 µm
Rel. Fluor. Intensity (%)
N = 4
t ½ = 60.9 s
Time (s)
B
Dodecamer nucleosome array
100 nm
100 nm
C
Magnification 50,000 X
2 min
10 min
30 min
60 min
90 min
100 nm
Density (µm⁻²)
Time (min)
D
10 min incubation
1 h incubation
Charged carbon grid
Imaging area
Carbon area
SA-crystal grid
Imaging area
Carbon area
1 µm
E
Ionic strength
5 mM Mg2+
150 mM Na+
5 mM Mg 2+
150 mM Na+
0.5 mM Mg2+
15 mM Na+
5 mM Mg 2+
15 mM Na+
0.05 mM Mg2+
1.5 mM Na+
5 mM Mg 2+
1.5 mM Na+
0.005 mM Mg2+
0.15 mM Na+
5 mM Mg 2+
0.15 mM Na+
100 nm

Figure 4.4 Measurement of condensates inner diffusion coefficient, size variation, grid surface density, and morphology. A) Fluorescence microscopy imaging of fusion events (top panel) and fluorescence recovery after partial photobleaching (bottom left) of Cy3-labeled tetranucleosome droplets formed in physiological salt after 1 h incubation at 20°C. The nucleosome diffusion coefficient within the droplet was determined by measuring the averaged rate of change of relative fluorescence intensity (bottom right). (B) AFM and OpNS images of dodecamer nucleosome array and representative particle views (box size 100 nm). (C) Time series of 10 nM dodecamer nucleosome array in physiological salt at 20°C imaged at magnifications of 50,000 X. Statistics of nucleosome and SphC number density shown in red and black, respectively. Data were obtained from 12-16 12,500 X images and statistics are presented as mean ± SEM. (D) Cryo-EM sample of tetranucleosomes prepared using the conventional glow discharged carbon quantifoil grid (left panel) and the non-charged SA-crystal lacey grid (right panel) after 2 min and 60 min incubation. Dashed line circles indicate large-scale condensates favored the binding to the carbon (non-imaging) area of conventional grid (left). However, similar size condensates were observed in the vitreous ice (imaging) area using a SA-crystal grid (right). (E) Effect of Mg2+ concentration on condensate formation (20 mM HEPES-KOH pH 7.5; 5-0.005 mM MgCl2; and 1 mM DTT, left column). Effect of Na+ concentration on condensate formation (20 mM HEPES-KOH pH 7.5; 150-0.15 mM NaCl; 1 mM EDTA; 1 mM DTT, middle column). Effect of Na+ concentration on condensate formation at constant Mg2+ concentration (20 mM HEPES-KOH pH 7.5; 150-0.15 mM NaCl; 5 mM MgCl2; and 1 mM DTT, right column). Low ionic strength buffer decreases nucleosome adsorption to the grid surface (Blue section) in contrast with high ionic strength buffer (yellow section). In conditions E, tetranucleosome samples were incubated 10 min at 20°C before OpNS.

We next used cryo-ET to visualize the NCP arrangement in the irregular condensates and the SphCs, and to capture their evolution during the early stages of condensation at the molecular level. To facilitate binding of the condensates to the grid surface, we used a 2D streptavidin (SA) crystal grown over the cryo-EM grids (Han et al., 2016) and tetranucleosome arrays in which 25% harbored a biotin molecule at one DNA end (Figure 4.2B). The presence of similar large condensates on grids prepared with and without the SA crystal after long incubation times in physiological salt indicated that the condensates form in solution, and that the SA crystal simply facilitated their attachment to the grid (Figure 4.4D).

To capture the rapidly formed, irregular condensates at the earliest imaging time, we collected tomographic data (Figure 4.2C) of the phase transition reaction at 2 min in low-salt (15 mM Na^+), and at 2 min and 10 min in physiological salt buffer (150 mM Na^+

and 5 mM Mg^{2+}). Given the minimum on-grid incubation time of 2 min for biotin binding to streptavidin (Han et al., 2016), we used the low salt condition to approximate the earliest (0 min) incubation times of the physiological salt concentration.

We designed a cryo-ET workflow to determine the spatial organization and coordinates of individual NCPs, both free and within each condensate (Figure 4.2C-H). To test our procedure, we used a sample incubated for 10 min at 4°C in physiological salt. By using deep learning-based denoising of each tilt image (Buchholz et al., 2019; Weigert et al., 2018) to enhance the low-dose image contrast (Figure 4.2C), coupled with a focused refinement strategy of individual-particle electron tomography (IPET) (Zhang and Ren, 2012), we obtained initial 3D density maps in which it was possible to identify NCP clusters displaying a low level of condensation (Figure 4.2D and Figure 4.5A). The z-dimensional 2D slice of the 3D map showed the irregular arrangement of NCPs within the condensates (Figure 4.2E, left panel).

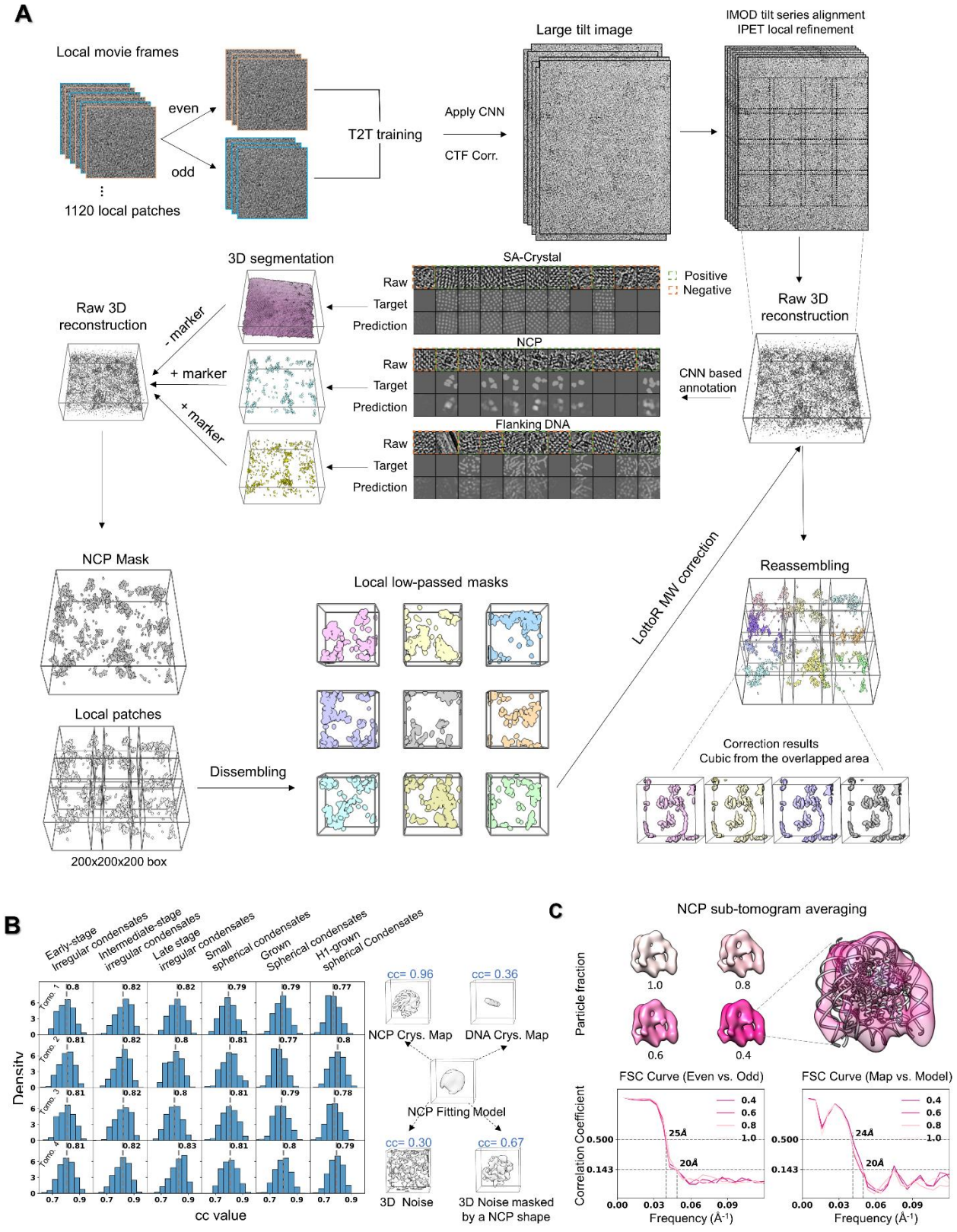


Figure 4.5 Condensates tomogram reconstruction work flow and validation of

identified NCPs. A) The tomographic 3D reconstruction work flow of the tilt series. The denoised tilt series using the TOMO2TOMO (T2T) method were initially aligned and locally refined by individual particle electron tomography (IPET) (top row). Annotation of the 3D reconstructed tilt series (middle row). Resulted NCP and DNA segments were used as markers to select nucleosome array density in raw map. Local low-tilt tomographic reconstruction (LoTToR) based missing wedge (MW) correction after dissembling 3D map into patches (bottom row). MW corrected patches were resembled by aligning the overlap area. B) Histograms showing the map-model fitting score distributions of all identified NCPs for four representative cryo-ET reconstructions (rows) from different types of condensates (columns) (left panel). The fitting scores were calculated by measuring the cross-correlation between the local density map and the corresponding model. Evaluation of the cross-correlation coefficient between NCP fitting model and reference models (right panel). A score of 0.96, 0.36, 0.30, and 0.67 were obtained for the NCP crystal structure map, DNA crystal structure map (40 bp), random noise map, and random noise enclosed by an NCP-shaped map, respectively. C) Quality evaluation of the identified NCPs by sub-tomogram averaging method. The density maps averaged from different fractions (1.0, 0.8, 0.6 and 0.4) of NCP particles and model fitting (Top). The map resolution were estimated by measuring the Fourier shell correlation (FSC) between two independently determined half-maps and between map and crystal structure 1AOI (bottom).

After annotating the initial 3D map using deep learning-based global segmentation (Chen et al., 2019) (Figure 4.2E, right panel and Figure 4.5A), we produced a coarse 3D map labelling all NCP positions (Extend Data Video 8). To regain the molecular details of NCPs, the initial and the coarse labelling 3D maps were superimposed, and the later was used as marker to select surrounding connected map density within the former. The resulting maps were then low-pass filtered and served as a mask for local missing wedge correction (Zhai et al., 2020). This procedure yielded the final denoised 3D map depicting NCPs isolated or in condensates, as well as their flanking DNA (Figure 4.2F, top panel and Figure 4.5A).

Next, flanking DNA and NCP models (PDB id: 1AOI) were docked into the final 3D map (Figure 4.2F, bottom panel) via a weighted iterative search algorithm (see Methods). We evaluated the result of this fitting by both calculating the cross-correlation (cc) score between the local map density and the model, and by performing sub-tomogram averaging. On one hand, we obtained an averaged cc value of ~0.8 ± 0.1 (Figure 4.5B; first row, third column), well above the value of ~0.3 obtained for a control (see Methods). On the other, the averaged 3D maps obtained had a resolution of 20-25 Å (estimated at Fourier shell correlation cutoff at 0.5 and 0.143), which is sufficient to

identify clearly the ~1.7 turns of DNA wrapped around the histone core (Figure 4.5C). We concluded that our identification of NCPs within the tomogram is robust. The final model and 3D map show the distribution of individual NCPs and DNA components (Figure 4.2G; Extend Data Video 9). Two representative 3D zoom-in views of individual tetranucleosome arrays show NCPs and flanking DNA arranged in extended and stacked conformations (Figure 4.2H; top and bottom panel, respectively). Finally, the models provide the coordinates of individual NCPs.

4.4.2 The early stage of nucleosomal condensation occurs by spinodal decomposition

To quantitatively analyze the early condensates, we calculated the specific space-coordinates of the center of mass of individual NCPs using cryo-ET as described above. We grouped the NCPs' space coordinates using a density-based clustering non-parametric algorithm (DBSCAN) (Ester et al., 1996). With this algorithm, NCPs surrounded by a minimum of twelve NCPs neighbors (self-included) within 27-29 nm (see Methods) were classified as NCP clusters or condensates (Figure 4.6A, B, and C). Each cluster was assigned a different color with black chosen to specifically depict "free" NCPs, i.e., those that could not be associated with any cluster. At 2 min, 4°C, and 15 mM Na^+, most NCPs appeared dispersed over the imaging area, and only a few of them were grouped into small clusters (Figure 4.6A). In contrast, at 2 min of reaction in physiological salt, most NCPs were grouped into larger, irregularly shaped condensates (Figure 4.6B) that continued to grow after 10 min of reaction (Figure 4.6C). We propose that the sparse distribution of condensates in 15 mM Na^+ likely mirrors the earliest stages (0 min) of nucleosome condensation at physiological salt that are otherwise not directly accessible given the earliest possible observation time of 2 min. Thus, the structures observed under this condition will be used as a proxy for the initial state.

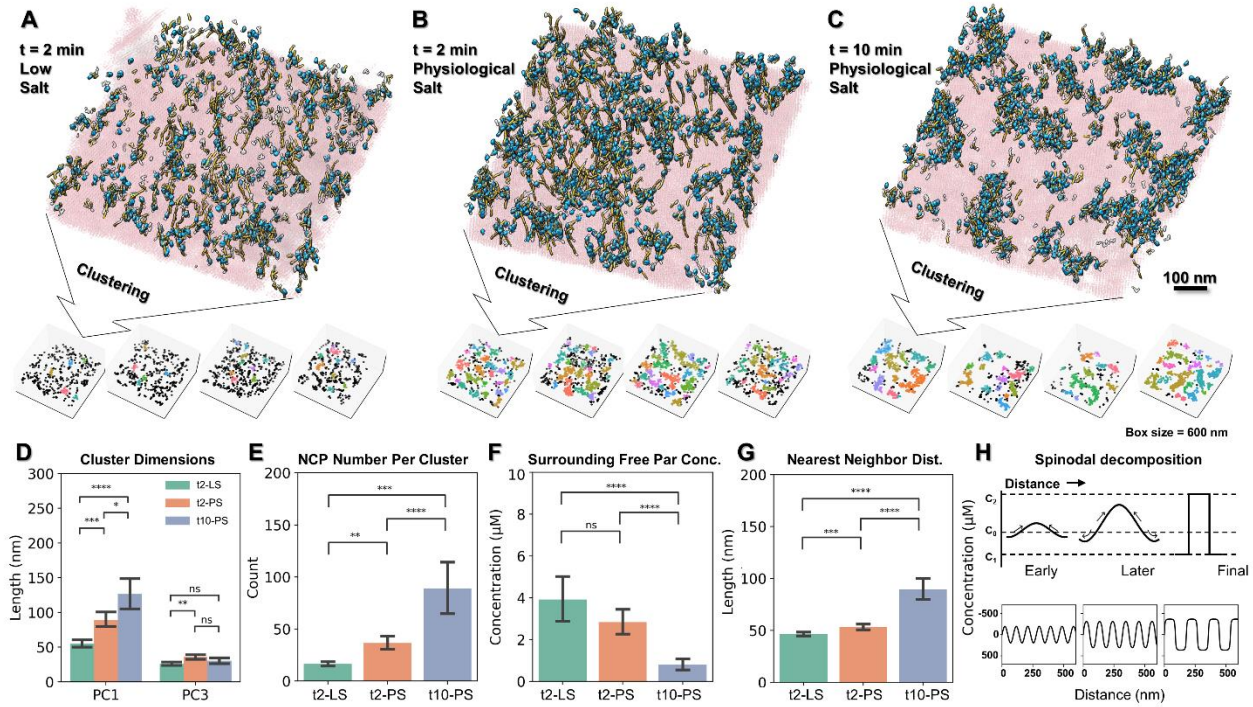


Figure 4.6 Time evolution of irregular condensates. Final denoised maps (thickness of ~90-130 nm) with fitted models displaying irregular tetranucleosome condensates obtained in low and high salt at 4°C. A; top) 2 min incubation at low salt (15 mM Na^+, t2-LS); B; top) 2 min at physiological salt (150 mM Na^+ and 5 mM Mg^{2+}, t2-PS); and C; top) 10 min at physiological salt (t10-PS). SA-crystal and nucleosome densities are depicted in pink and transparent gray, respectively. Fitted NCPs and flanking DNA colored in cyan and yellow, respectively. Bottom panels show NCP clusters of four representative tomograms for each incubation condition, grouped by the DBSCAM algorithm. Each color represents a unique NCP cluster and black depicts free, unclustered NCPs. D) Condensates' dimensions along PC1 and PC3 axes derived from principal component analysis. E) NCP number per condensate cluster. F) Concentration of free NCPs surrounding a condensate within a 20 nm shell. G) Nearest neighbor distance between condensates. The data are plotted as mean ± SEM and are compared using a two-tailed unpaired t-test, where $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$; and ns, not significant. Schematic representation of the spinodal decomposition process as a function of time obtained from the Cahn-Hilliard equation: top, image modified from (Findik, 2013); bottom, evolution of irregular condensates in terms of their nearest neighbor distance (wavelength), condensate internal concentration (amplitude), and surrounding free NCP concentration (sharpness) represented by a Squdel sine function.

The early condensates exhibited irregular, elongated shapes and were densely distributed all over the surface. We used Principal Component Analysis (PCA) (Hotelling, 1933) to quantitatively characterize these condensates by identifying their long (PC1) and their short (PC3) axes. The average size of the condensate's long axis after 2 min of incubation at 4°C and 15 mM Na^+ (the initial reference state) was 55.1 ± 5.2 nm and the average size of the short axis was 26.1 ± 2.4 nm. At 4°C and physiological salt, these two values increased to 89.0 ± 10.1 nm and 35.9 ± 3.0 nm after 2 min of incubation, and to 126.9 ± 19.6 nm and 31.6 ± 4.0 nm after 10 min of incubation (Figure 4.6D). This analysis indicates that clusters grow asymmetrically along their long axes, and that their short axes stay relatively stable throughout the reaction, with a diameter of ~30 nm. As a result, their eccentricity (defined here as 1 – short axis/long axis) increases from 0.51 ± 0.05 to 0.56 ± 0.03 to 0.73 ± 0.03 at initial, 2 min, and 10 min of reaction times, respectively (Figure 4.7A). The growth of the condensates at these time points was accompanied by an increase in the average number of NCPs per condensate, from 17 ± 3, to 37 ± 6, to 89 ± 26 (Figure 4.6E). Next, we quantified the depletion of isolated NCPs around the condensates as the reaction proceeded, by determining the number of isolated NCPs found within a 20 nm shell surrounding the condensate (Figure 4.6F). At the initial reference state, the concentration of NCPs within the shell was 3.9 ± 1.0 μM; in physiological salt this value decreased to 2.8 ± 0.5 μM after 2 min and to 0.8 ± 0.3 μM after 10 min of condensation reaction (Figure 4.6F). To further validate this analysis, we determined the total number of free NCPs within the imaged area and calculated their concentration in the ice slab. Consistent with the above results, the concentration of free NCPs was found to be 25.0 ± 3.5 μM, 8.9 ± 2.3 μM, and 3.9 ± 0.7 μM for the initial reference state, for the 2 min and 10 min observation times in physiological salt, respectively (Figure 4.7B). The total (free + condensed) concentration of NCPs within the ice slab throughout the reaction slightly increased as it was found to be 31.7 ± 5.4 μM, 41.0 ± 3.7 μM, and 48.4 ± 16.8 μM (Figure 4.7C) for the three time points analyzed.

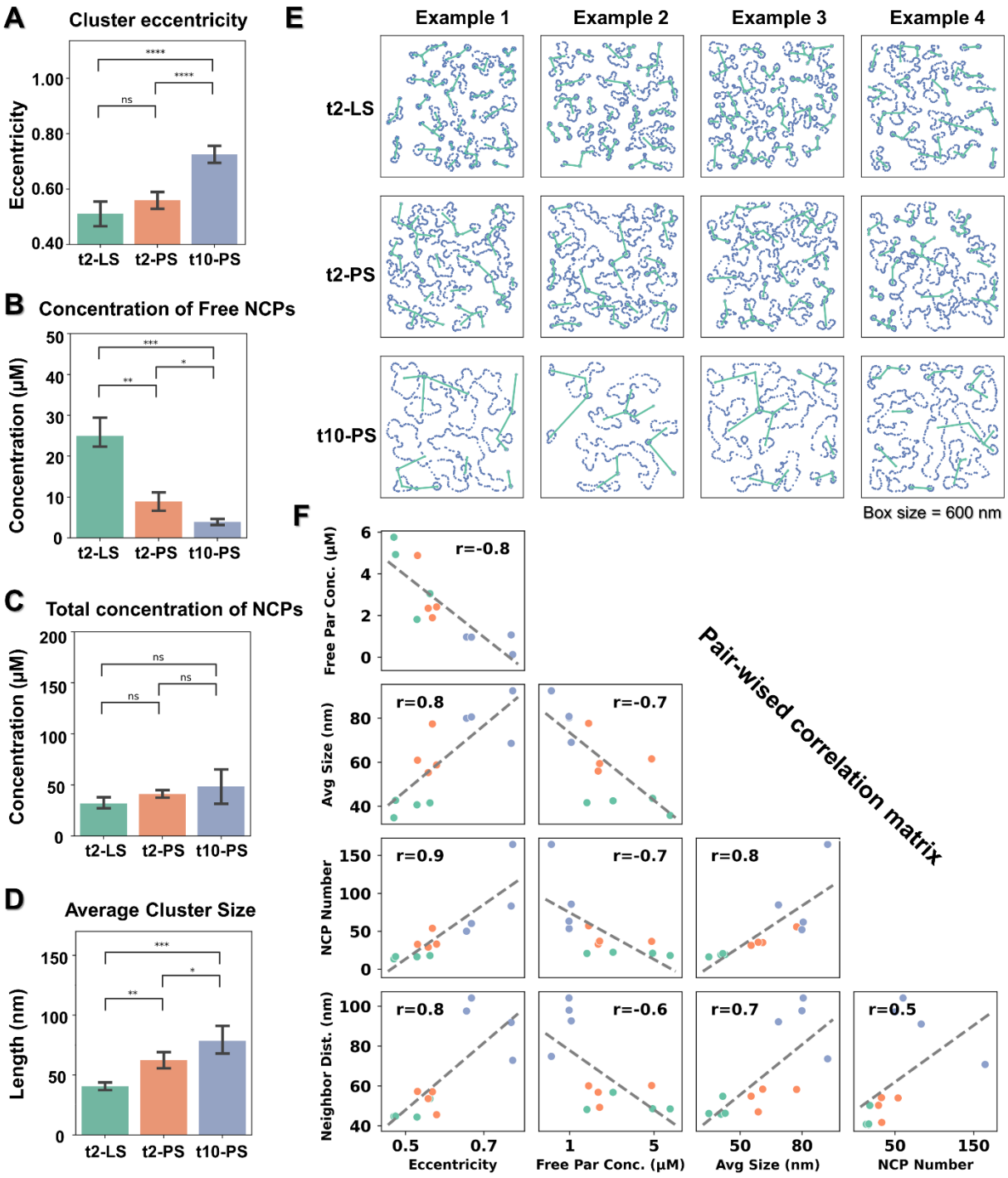


Figure 4.7 Statistical analysis of cluster size, shape, concentration, distance, and correlations for irregular condensates obtained from the cryo-ET reconstruction. A) Average dimensions of irregular condensates expressed as a mean between the long (PC1) and short (PC3) axes obtained from a Principal Component Analysis (PCA). B)

Quantitation of irregular condensates eccentricity expressed as (1 – PC3/PC1). C) Concentration of free NCPs within the ice slab (number of free NCPs divided by the ice volume of the reconstructed tomogram). D) Total NCP concentration within the ice slab (number of all identified NCPs divided by the ice volume). All data are plotted as mean ± SEM and measurements are compared using a two-tailed unpaired t-test, where $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$; ns, not significant. E) Measurement of the nearest neighbor distance among irregular condensate clusters. Identification of condensate-shaped contours (blue dots) within the z-dimension central slice for early (t2-LS), intermediate (t2-PS), and late irregular (t10-PS) condensates (top to bottom). Four examples of these three conditions are presented (columns). Nearest-neighbor condensates are connected by green lines, whose mean distances were used to represent the change of the critical wavelength, describing the evolution of spinodal decomposition in time. F) Pair-wise correlation matrix between independently determined number of NCPs per condensate, condensate size, eccentricity, neighbor distance among condensates, and concentration of free NCPs surrounding condensates. The mean value of the above five independent measurements from tomograms obtained for early, intermediate, and late stage irregular condensates are presented in green, orange, and blue points, respectively. The analysis yielded a mean Pearson correlation coefficient (r-value) of 0.73.

These observations are consistent with a mechanism of spinodal decomposition, which involves negative diffusion of solute material against a concentration gradient, gradually forming irregular condensates with defined borders. Spinodal decomposition does not involve the crossing of an energy barrier, and therefore the formation of condensates is ubiquitous and arises as a result of local, small concentration fluctuations in the medium. The growth of these spinodal condensates (SpnCs) can be described in terms of a time-dependent periodic distance or wavelength $\lambda(t)$, which gives the average separation between domains of the condensed phase, ultimately attaining a maximum value λ_{max} (Theory Section S1, Equation 9) (Emo et al., 2014). The spacing among all condensate clusters was estimated by calculating the mean nearest neighbor distance between condensates using a low-pass-filtered map (8 nm) (Figure 4.7E). The average spacing between condensates gradually increased from 46.5 ± 1.8 nm to 53.3 ± 2.8 nm, to 89.5 ± 9.8 nm at 10 min (Figure 4.6G). A pairwise correlation test among five independent measurements, i.e., condensate cluster size (Figure 4.7D), eccentricity, number of NCPs per condensate, concentration of surrounding free NCPs, and nearest neighbor distance, yielded a mean absolute r-value of 0.73 (Figure 4.7F). This high correlation confirms that SpnCs grow by the asymmetric accretion of isolated NCPs leading to the formation of elongated condensates with sharp borders according to a spinodal decomposition process (Figure 4.6H).

4.4.3 Spherical condensates arise by nucleation and growth in the later stages of phase separation

The denoised, zero-tilt images of samples prepared in physiological salt that were incubated for 10 min at room temperature revealed distinct spherically-shaped structures of varying sizes from ~40 nm (Figure 4.8A; panel I) to ~400 nm in diameter (Figure 4.8A; panel IV) that sparsely appeared among the SpnCs. These condensates seemed comparatively denser than the surrounding irregular spinodal material, suggesting that they are an early manifestation of the formation of liquid droplets. Their sparse distribution (Figure 4.9A) is consistent with the existence of an energetic barrier that must be crossed for these condensates to form.

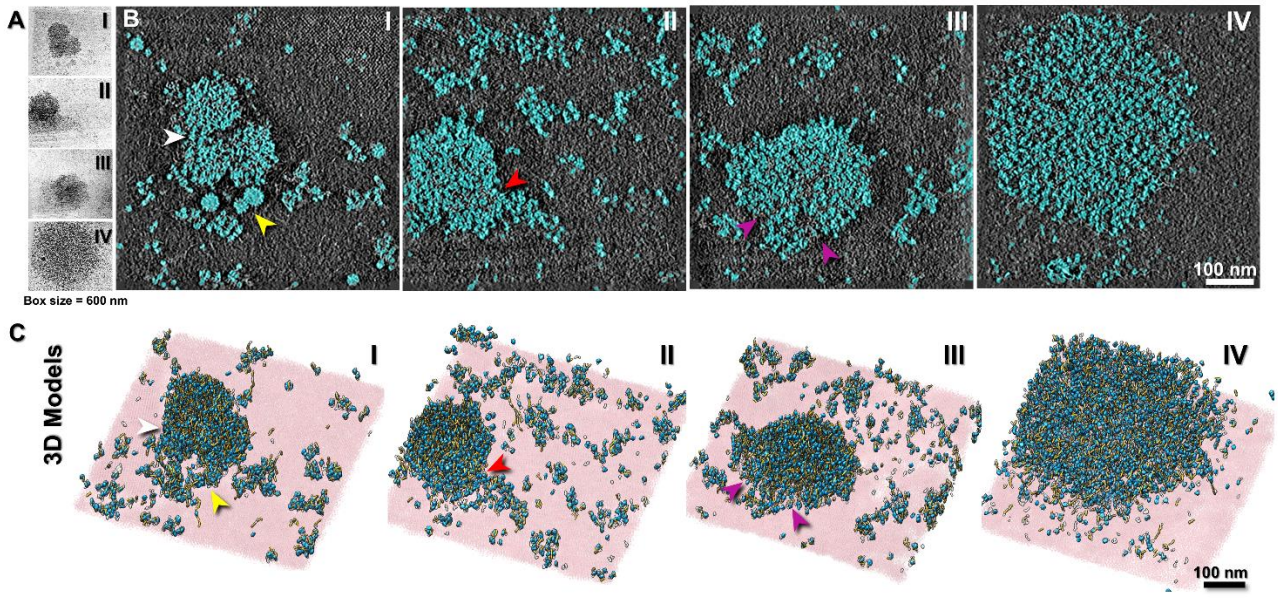


Figure 4.8 3D reconstructions of initial stages and growth of Spherical Condensates. Spinodal condensates (SpnCs) and spherical condensates (SphCs) of different sizes observed after 10 min incubation of tetranucleosome at 20°C in physiological salt. A) Enhanced cryo-ET zero-tilt images. The tomograms were organized by increasing SphCs size and labeled I-IV. B) Corresponding tomogram slices (1.17 nm thickness) depicting NCPs (cyan) after deep learning-based segmentation. C) Final 3D map of tomograms in B showing the NCP (transparent gray) and SA-crystal (pink) densities (tomograms thickness ~90-140 nm). The NCP component of the fitted models is shown in cyan and the flanking DNA segment in yellow. I-IV illustrate a possible mechanism of growth of SphCs: (I) formation of small, near-spherical condensates (~35 nm, yellow arrow) and merging of three larger SphCs (~100 nm, white arrow); (II) accretion of SpnC by a ~200

nm SphC (red arrow); (III) a ~300 nm SphC resulting from the apparent fusion (purple arrows) of smaller condensates; (IV) a larger SphC of ~400 nm result from growth processes.

To characterize the internal structure of these SphCs and to establish how it relates to that of the surrounding SpnCs, we mapped out the NCPs in areas containing both types of structures (Figure 4.8B). Because under fluorescence microscopy liquid droplets increased their size along the course of the condensation reaction, we used representative SphCs ordered by increasing size to illustrate the possible early stages of droplet formation and growth, both in segmented 2D slices and in the corresponding 3D models (Figure 4.8B and C; panels I-IV). The panels in figures 3B and 3C also depict the likely fusion between SphCs of small (panel I, yellow arrows) and intermediate sizes (panel I, white arrows). However, the images indicate that growth of these SphCs was likely to occur not only by fusion among themselves, but also by accretion of spinodal material, as reflected by the presence of irregular condensates attached to spherical ones (Figure 4.8B and C; panel II, red arrow). Panel III in Figure 4.8B and C illustrates the possible merging (purple arrow) of smaller condensates to form a larger one while maintaining its spherical morphology. Such process suggests a mechanism by which even larger SphCs are generated (~400 nm) (Figure 4.8B and C; panel IV).

The small, rounded condensates (Figure 4.8B; panel I, yellow arrows) were comparable in size to their spinodal counterparts and exhibited a near perfect spherical geometry. We propose that these small SphCs result from an early nucleation step that eventually gives rise to larger condensates. To confirm this inference, we analyzed the sample after 20 min of incubation since nucleus formation is a rare event at early times (Sear, 2014) and the likelihood of capturing them increases with longer incubation periods (Figure 4.9B). Under these conditions, we observe small SphCs (Figure 4.9C and D), either isolated or in groups of two or three (Figure 4.10A and 4B). Statistical analysis indicates that the probability of finding a group of small SphCs together is similar to that of finding them isolated (Figure 4.10C). This observation suggests that SphCs appearing in clusters all likely arose from the same common large local concentration fluctuation within a SpnC.

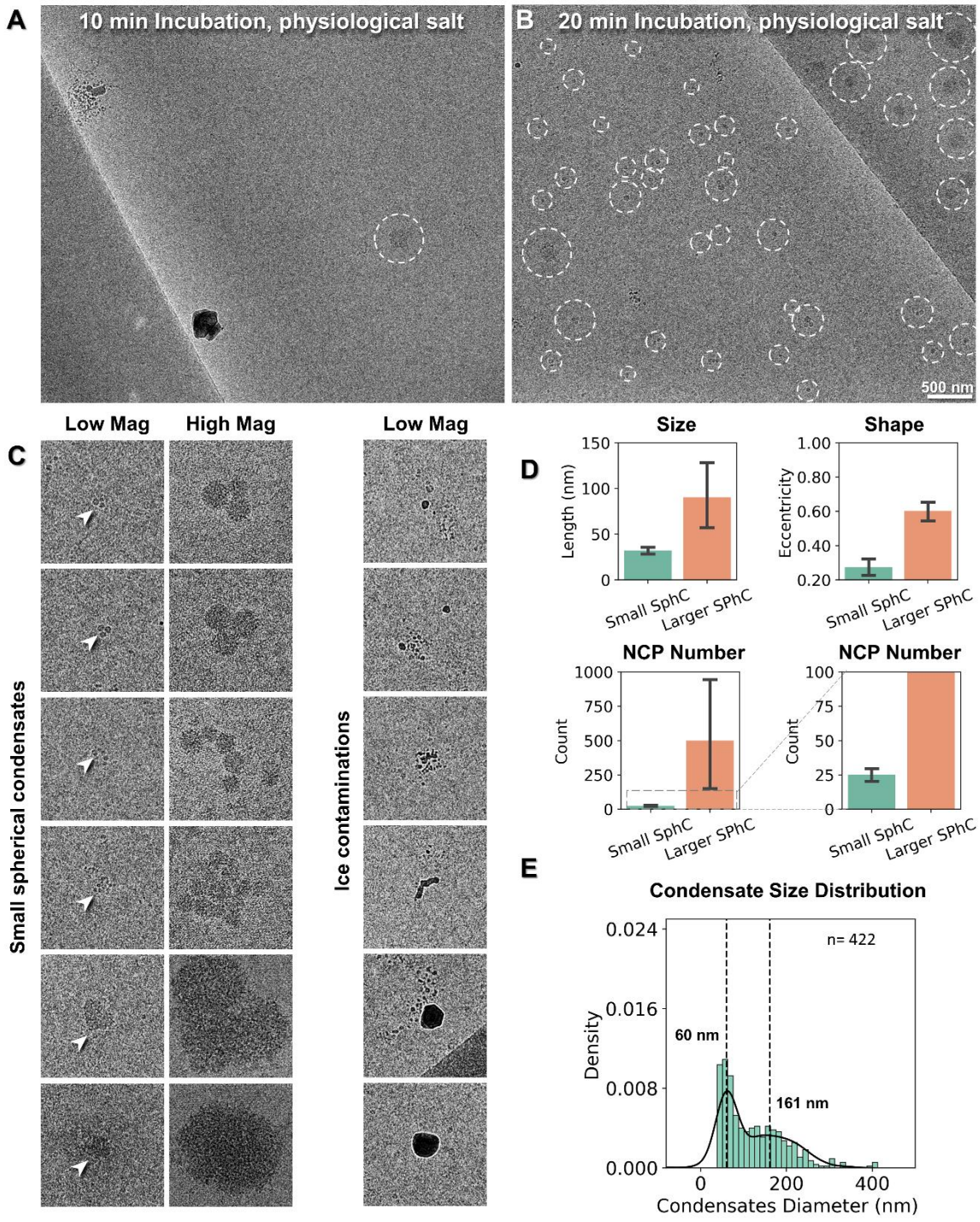


Figure 4.9 Identification of small SphCs and quantitation of their size, shape, and distribution from 2D cryo-EM images and 3D reconstructions. A) Representative cryo-EM image showing the sparsely distributed SphCs on the SA-crystal surface after 10 min incubation in physiological salt at 20°C. B) Representative cryo-EM image showing small SphCs of varying sizes under the same condition as A after 20 min incubation. C)

Representative particle views of a small and a larger bundle of attached small SphCs imaged at both low (3,600 X) and high (53,000 X) magnification (left). Control images of ice contaminants showing a high contrast and white particle fringe when compared to nucleosome condensates at similar sizes (right). Ice particles were excluded from the statistical analysis of small condensate size distribution. Box size for Low Mag= 1,000 nm; Box size for High Mag= 220 nm. D) Statistical analysis of the size (defined as the mean of the PC1 and PC3 axes after PCA analysis), eccentricity (1 – PC3/PC1 axis), and number of containing NCPs of small (n= 27) and larger (n= 7) SphCs measured from the reconstructed tomogram 3D models. E) Histogram of SphC size distribution measured from low magnification 2D cryo-EM images (B; n= 440). Data was fitted using two Gaussians distributions. Dashed lines correspond to the mean of the two distributions (60 and 161 nm).

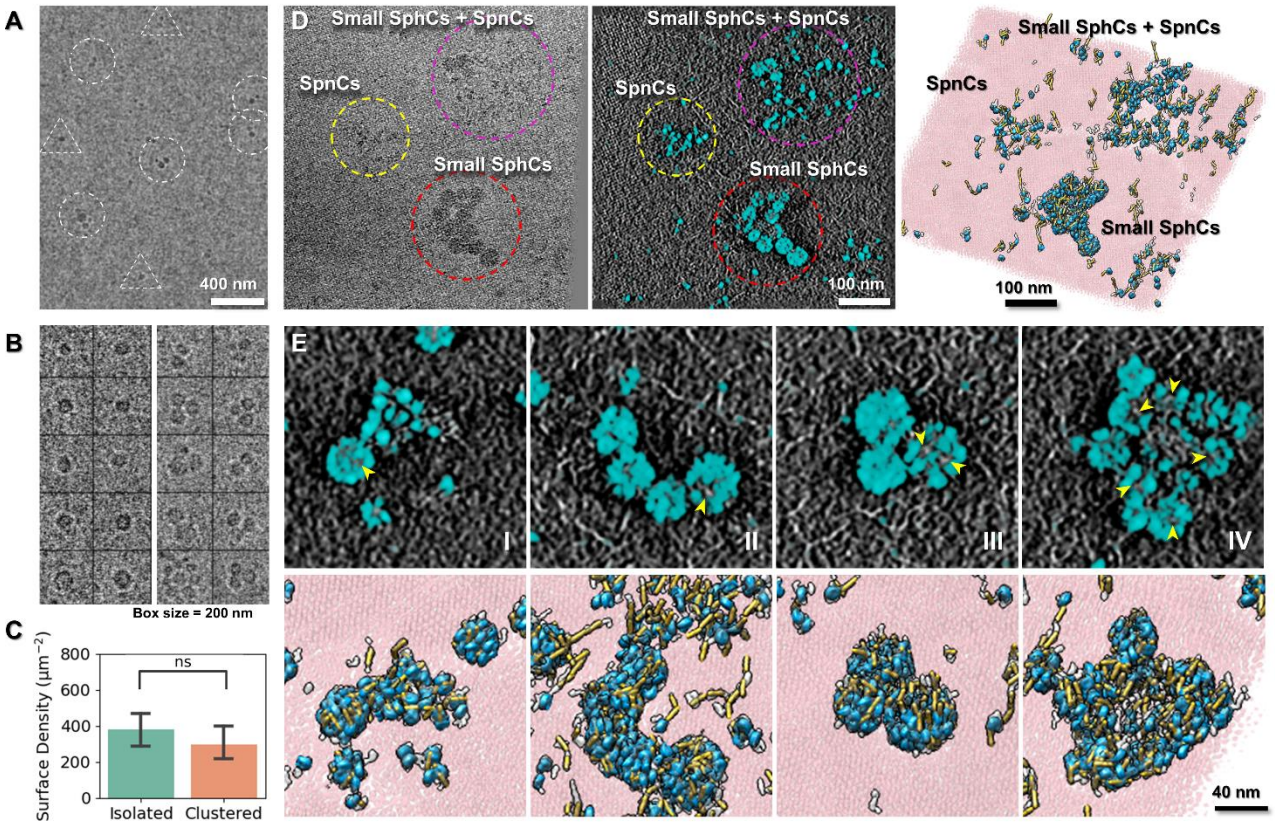


Figure 4.10 Small spherical nuclei arise from SpnCs. A) Cryo-EM image showing single (dashed triangles) and multiple (dashed circles) small SphCs formed at 20°C in physiological salt after 20 min of incubation. B) Representative particles of individual (left columns) and clustered (right columns) small SphCs. C) Statistics from 12 low-magnification images indicating a similar probability of finding condensates in clusters or in isolation. Data plotted as mean ± SEM. ns, not significant. D) High-magnification

cryo-ET zero-tilt image (left), tomogram slice (middle), and 3D model (right) of a cluster of small SphCs (red dashed circle), SpnCs (yellow dashed circles), and small SphCs in close proximity with SpnCs (magenta dashed circle). E) Various stages of the nucleation and growth process (I-IV) of the SphCs shown by representative tomographic slices (top) and their corresponding 3D maps and models from ~70 nm thickness tomograms (bottom). The states include: (I) conversion of part of a SpnC into a small nucleus; (II) formation of multiple nuclei in side-by-side contact; (III) small nuclei organized in a cluster; (IV) cluster of nuclei with smeared boundaries suggesting a process of fusion. Yellow arrows identify cavities within the small nuclei.

A zoomed-in view of the z-dimensional slices and their corresponding 3D models showed that some of the small SphCs were connected to and surrounded by irregular SpnCs (Figure 4.10D). Panel I of Figure 4.10E shows an example in which only half of the condensate was rounded and made up of 28 NCPs, while the other half, with 36 NCPs, was less organized. The total number of NCPs was similar to that of the late SpnCs (~60 NCPs) (Figure 4.6E). These hybrid structures were also seen in the OpNS experiments (Figure 4.3E, white arrow). While in early reaction SpnCs were present without SphCs, the converse was not true: SphCs were only observed among irregular condensates. Together, these observations suggest that the small SphCs emerged from the pre-formed SpnCs, and served as the nuclei that can then grow into larger SphCs, either via fusion among themselves and/or from accretion of surrounding spinodal structures.

Analysis of the size distribution of all SphCs in the 2D micrographs collected at 20 min of incubation revealed two peaks centered at diameters of 60 ± 14 nm and 161 ± 66 nm, while the smallest spherical structures observed had a diameter of ~35 nm (Figure 4.9E). We suggest that the first peak represents the formation, due to large local concentration fluctuations, of thermodynamically unstable initial nuclei. The valley between the two peaks might indicate the dimensions beyond which the initial nuclei begin to overcome the interfacial energy barrier to (irreversibly) grow into larger and more stable structures (Figure 4.9E). The probability distribution peak at 160 nm (Figure 4.10D) is likely the result of two opposing factors: one favoring the formation of larger condensates by the extra thermodynamic stability gained via the accretion of surrounding spinodal material and/or the fusion with smaller condensates, and another disfavoring further growth because of the depletion of nearby material. The smallest spherical nuclei, although tightly packed, also contained regions of randomly located cavities (Figure 4.10E; top panel, yellow arrows).

As before (Figure 4.8B), we illustrate a possible temporal evolution and growth of

nuclei by assembling representative SphCs (Figure 4.10E; panels I-IV). After the earliest stage in which small unstable nuclei are formed (panel I), these are often seen forming either linear or cluster arrangements (panels II and III); it is possible that the latter are favored over the former by the reduction in interfacial surface, a process that, upon fusion, will lead to a more favorable (smaller) surface-to-volume ratio. However, the high probability of observing clusters composed of spherical nuclei without fusion might be a manifestation of a secondary energy barrier for nucleus fusion. Larger nucleosome droplets result once this barrier is overcome (panel IV). Taken together, these observations indicate that nucleosomal phase separation occurs through a two-step process, the first involving organization of condensates by spinodal decomposition and the second in which the spinodal structures undergo re-organization into spherical, tight condensates (See Discussion).

4.4.4 H1 catalyzes the spinodal-to-spherical condensate transition

Linker histone H1 regulates chromatin compaction and packaging (Song et al., 2014; Zhou et al., 2016; Andreyeva et al., 2017). Epifluorescence microscopy revealed that H1 added in a 1:1 proportion to NCPs increased the number and size of nucleosome droplets (Figure 4.11A). To determine how H1 affects the proposed two-step mechanism, we repeated the OpNS time series imaging in the presence of H1. We found that H1 induced the formation of large, micron-size SphCs at the earliest reaction time (2 min) (Figure 4.12A-D), comparable to those observed after 60 min of incubation under the same conditions but without H1 (Figure 4.3A). Also, in the absence of H1, the surface number density of NCPs (measured from areas devoid of SphCs) decreased only by ~7% after 90 min of incubation (Figure 4.3B), whereas in the presence of H1, this parameter decreased exponentially to one-third of its initial value during the same time period (Figure 4.12E). Same incubation time series were also repeated at 4°C and 36°C in the presence of H1, which showed a similar trend but at lower rates (20°C > 36°C > 4°C, Figure 4.12E). This result is consistent with the observation that the rate of SphCs formation does not monotonically depend on the temperature of incubation, being maximum at 20°C followed by 36°C and then by 4°C without H1 (Figure 4.3B). Interestingly, a similar non-monotonic behavior with temperature has been described for tau protein condensates (Dong et al., 2021).

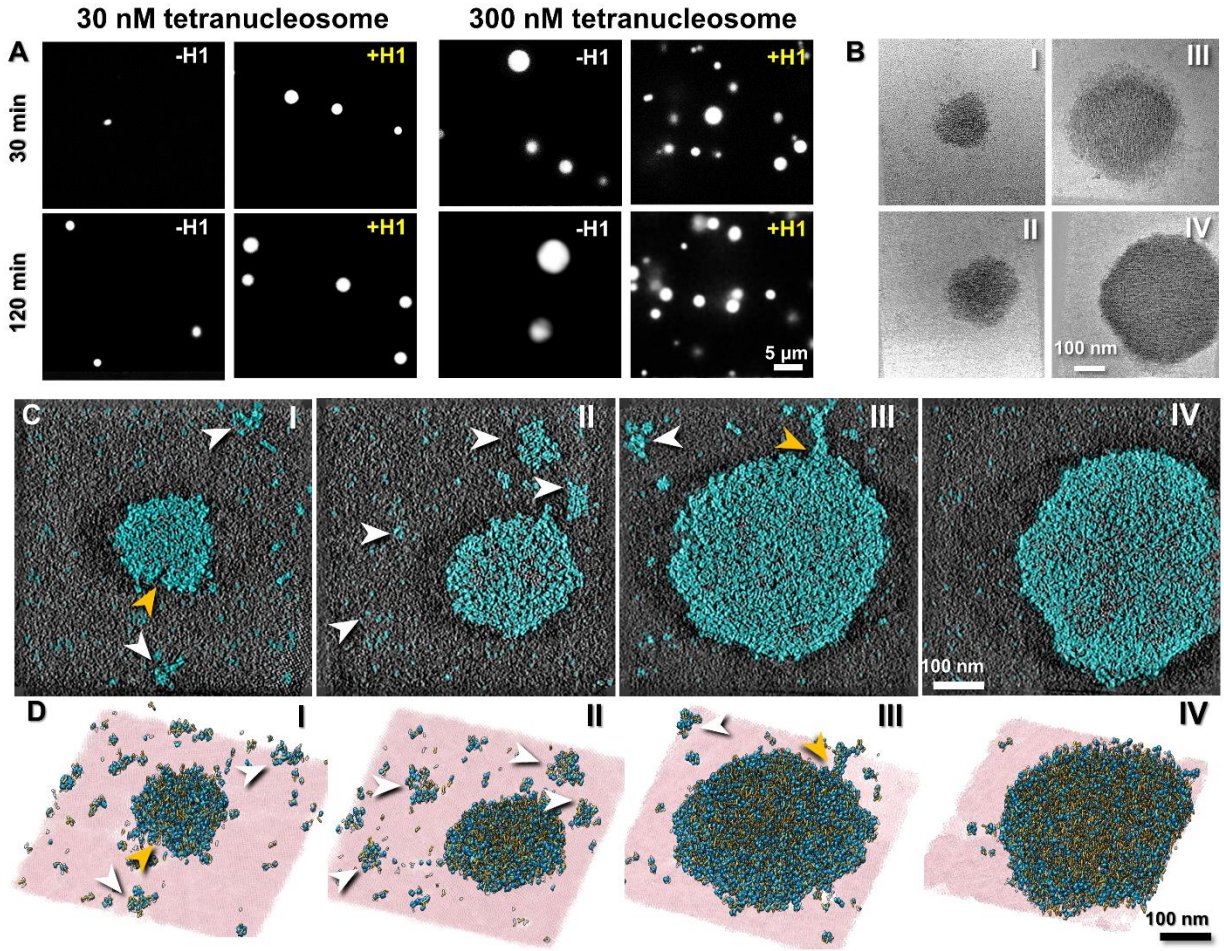


Figure 4.11 Linker histone H1 accelerates the transition from SpnCs to SphCs. A) Fluorescence microscopy images of 30 nM (left) and 300 nM (right) Cy3 labeled tetranucleosomes incubated at 20°C in physiological salt with and without H1. B) Representative zero-tilt cryo-ET images of SphCs obtained in the presence of H1 at 20°C in physiological salt after 10 min of incubation ordered by increasing size and labeled I-IV. C) Tomogram slices corresponding to B; D) 3D maps (tomograms thickness, ~90-130 nm) and models of the corresponding condensates, illustrating different stages of growth of SphCs. White arrows indicate the sparsely distributed SpnCs in the surrounding areas. Orange arrows indicate fusion events between condensates.

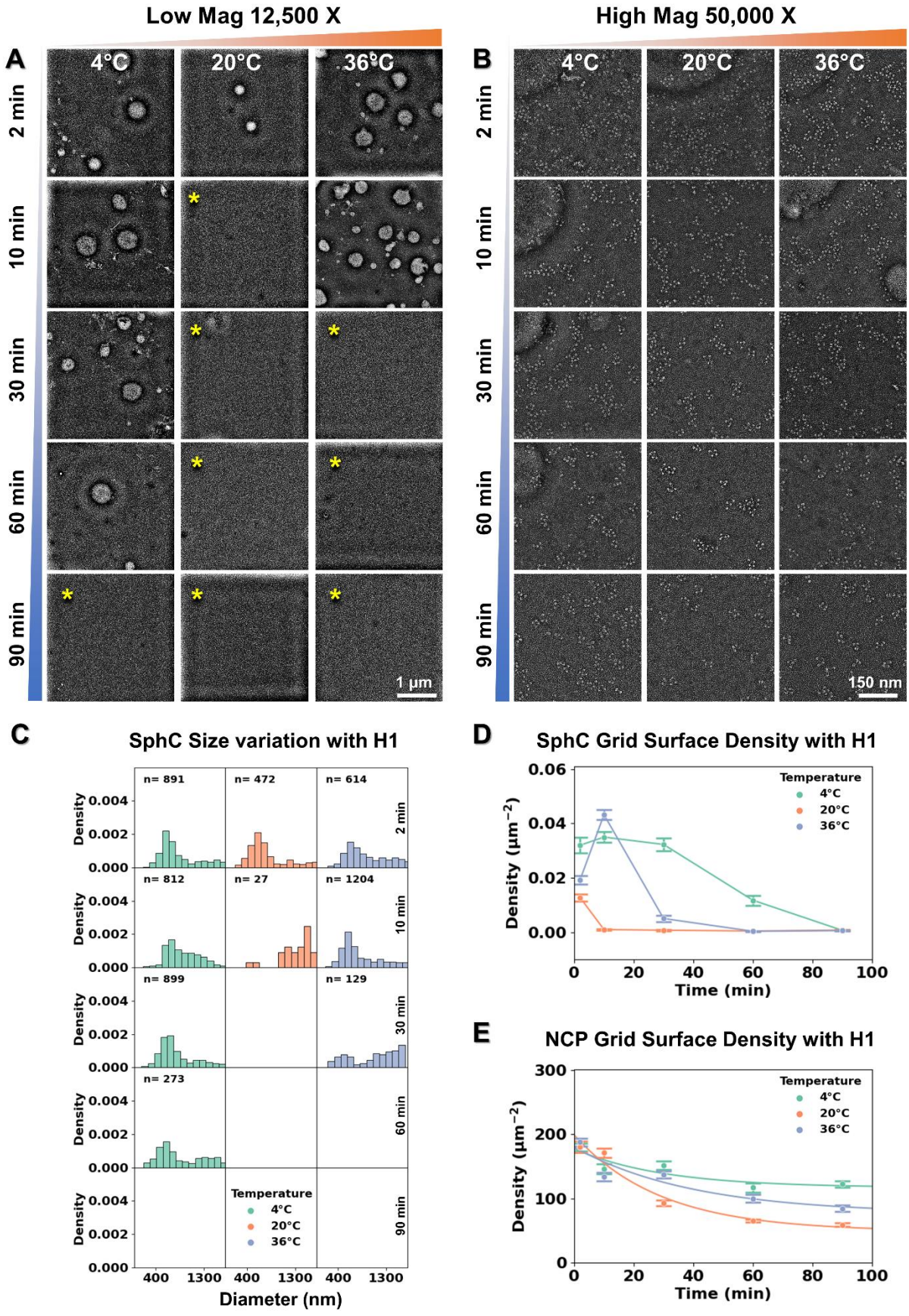

Low Mag 12,500 X
High Mag 50,000 X
A
B
4°C
20°C
36°C
2 min
10 min
30 min
60 min
90 min
1 µm
150 nm
C
SphC Size variation with H1
n= 891
n= 472
n= 614
n= 812
n= 27
n= 1204
n= 899
n= 129
n= 273
Density
Temperature
4°C
20°C
36°C
400
1300
Diameter (nm)
D
SphC Grid Surface Density with H1
Density (µm⁻²)
Time (min)
E
NCP Grid Surface Density with H1

Figure 4.12 Quantitative measurement of size variation and grid surface density of tetranucleosome condensates in the presence of H1. Representative time series of 30 nM tetranucleosome in physiological salt and 120 nM H1 at 4°C, 20°C, and 36°C imaged at magnifications of A) 12,500 X; and B) 50,000 X. C) Statistics of SphC size distribution on the carbon grid surface in the presence of H1 measured from A. D) Statistics of SphC density on the carbon grid surface in the presence of H1 (number of condensates divided by the imaging area) measured from A. E) Statistics of nucleosome density in the presence of H1 (number of nucleosomes within SpnCs over the corresponding imaging area) measured from B. Data C-E was obtained from 12-16 images and statistics are presented as mean ± SEM. Data in E was fitted with an exponential decay function. The low chance of observing SphCs on the grid surface after longer incubation (yellow asterisk in A) is a feature observed in the presence of H1, because H1 produced larger but fewer SphCs given the limited amount of nucleosome material in solution. This inference is supported by 1) the SphC grid surface number density decreased over time in D; 2) a significant amount of spinodal material was depleted after 90 min of reaction in E; 3) the disappearance of SphC at earlier time points is consistent with a faster material depletion rate (20°C > 36°C > 4°C in E); 4) due to the difficulty of binding to grid surface, no large SphCs over >1.8 μm were observed in all conditions.

Using cryo-ET, we captured SphCs formed in the presence of H1 after 10 min of incubation at 20°C in physiological salt (Figure 4.11B). Cryo-ET reconstruction of these condensates (Figure 4.11C and D) reveal that they have a tighter NCP packing (panel III) than those obtained under the same conditions in the absence of H1 (Figure 4.8A; panel IV). The central z-dimensional slice of the reconstruction and the corresponding 3D models shown in figures 5C and 5D, respectively, display SphCs that appeared to be in the process of fusing with other SphCs or accreting spinodal material (Figure 4.11C and D, yellow arrows). The sparse distribution of SpnCs in figures 5C and 5D, compared to their larger abundance around SphCs formed in the absence of H1 (Figure 4.8B and C), supports our inference that H1 only catalyzes the conversion of SpnCs into SphCs without skipping over the formation of the spinodal phase.

4.4.5 NCP organization inside the spinodal and spherical condensates

The cryo-EM images and 3D reconstructions (Figure 4.2, 4.8, and 4.11) show that the SphCs have a higher density and larger diameter than the SpnCs. In addition, the flanking DNA portions in the SphCs appear hidden inside the particle surface, compared to their SpnCs counterparts. We used the spatial coordinates of the individual NCPs to quantitatively investigate their concentration, internucleosome distance distribution, and orientations in the different types of condensates. To determine the concentration of NCPs in the condensates, we first defined a series of 3D contours with

the shape of each condensate, separated by 2 nm, and placed at the center of mass of the condensate (See details in Methods, Figure 4.13A; left panel). We then calculated the NCP concentration within contour *n* (internal volume concentration, middle panel); the difference of the number of NCPs enclosed by contours *n* and *n*-1 divided by the difference in their volumes gave the NCP concentration of shell *n* (shell concentration, right panel). Next, we average the concentrations enclosed by successive contours for all condensates of the same type (Figure 4.13B) and for each shell, starting from n > 5, as both concentrations are not well-defined when the measured volume approaches the size of a single NCP.

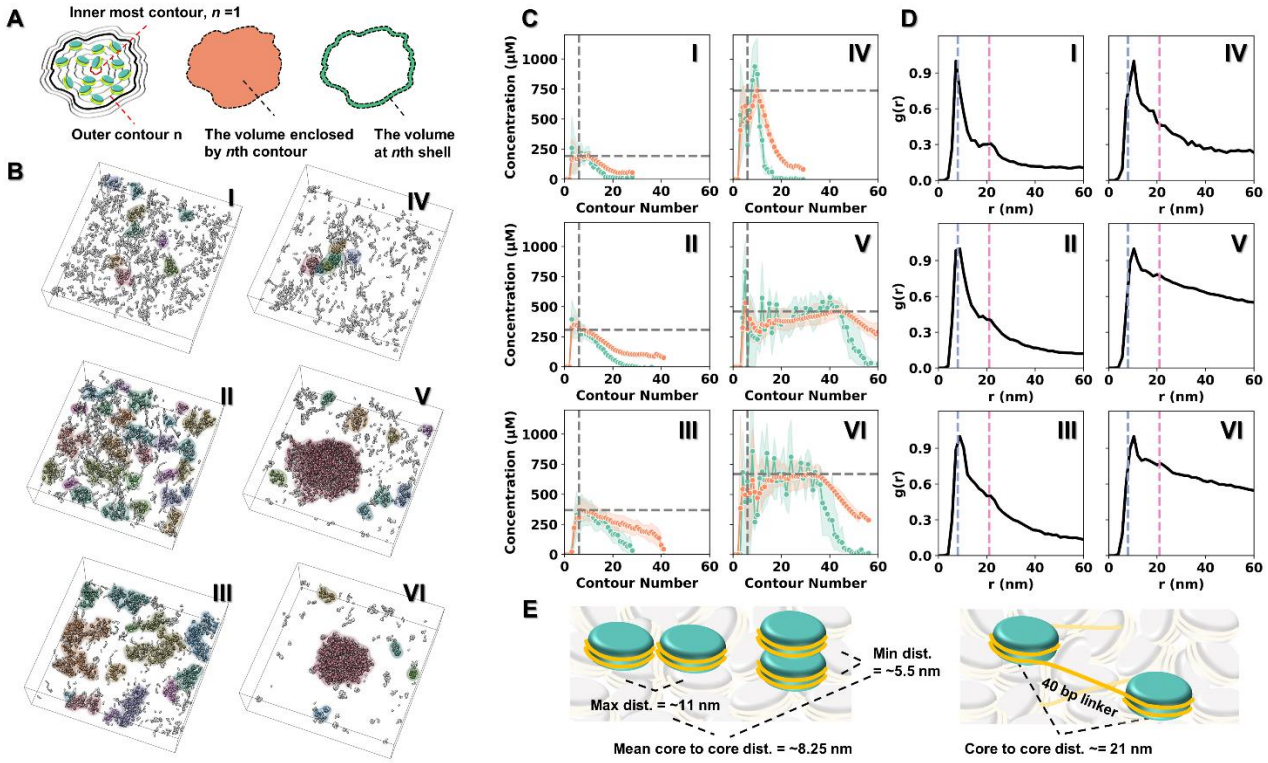


Figure 4.13 Quantitative analysis of NCP concentration and organization during different stages of condensate growth. A) Schematic depiction of a series of condensate-shaped concentric contours (2 nm apart, left). Depiction of the volume enclosed by the *n*th contour (orange, middle) and the volume delimited by contours *n* and *n*-1 (green, right). The internal volume and shell concentration are calculated as described in the text. B) The shape of the condensate was defined by 10 nm low-pass maps shown in different colors. The numbering of the maps follows the order from early (I), intermediate (II), and late stage (III) SpnCs to the small nuclei (IV) and larger SphCs without (V) and with (VI) H1. C) Quantitation of the *n*th-internal volume (orange) and *n*th-shell (green) NCP concentration corresponding to stages I through VI. The horizontal dash lines mark the highest internal volume concentration beyond *n*= 5

(vertical dash lines). D) Radial distribution function g(r) of NCPs' centers of mass for stages I through VI. The blue and pink dash lines indicate the two characteristic nucleosome core-to-core distances at 8 nm and 21 nm, respectively. E) Theoretical inter-nucleosome distances for side-to-side (11 nm) and stacked (5.5 nm) configurations (left), and intra-nucleosome distance between neighboring NCPs within an array (21 nm, right).

For contour n = 6, the shell concentrations for the initial, intermediate, and late stage SpnCs were 217 ± 43, 274 ± 36, and 356 ± 66 μM, respectively, and the corresponding internal volume concentrations were 195 ± 25, 309 ± 21, and 364 ± 47 μM (Figure 4.13C; left column). The shell concentrations for all three stages of SpnCs decreased towards the periphery of the condensates, and vanished for shell n = 17, 24, 28 for the three stages, respectively (Figure 4.13C; left column). In comparison, the initial spherical nuclei, the SphCs, and the SphCs in the presence of H1, all displayed overall higher shell concentrations than their spinodal counterparts at shell n = 6 yielding values of 286 ± 97, 339 ± 74, and 612 ± 152 μM, respectively, and correspondingly higher internal volume concentrations of 494 ± 68, 402 ± 45, and 550 ± 138 μM (Figure 4.13C; right column). Unlike the SpnCs, the initial spherical nuclei, the SphCs, and the SphCs formed in the presence of H1 displayed a slight increase in shell concentrations towards the periphery reaching peak shell concentrations of 939 ± 99, 501 ± 36, and 688 ± 65 μM, and peak internal volume concentrations of 743 ± 64, 471 ± 42, and 667 ± 62 μM, respectively (Figure 4.13C; right column). Thus, the concentration of the SphCs is ~1.3 to 3.6-fold higher than that of the SpnCs, and increases significantly in the presence of H1. The large discrepancy between the peak shell and peak volume concentrations (~200 μM) in the spherical nuclei suggests that the formation of a nucleus involves an outer shell of tightly packed NCPs with an interior somewhat lower in NCP concentration; this conclusion is supported by z-dimensional slice images of nuclei, clearly displaying these features (Figure 4.10D and 4E).

Next, we calculated radial distribution functions g(r) around each NCP and averaged them over all condensates of the same type (See Methods). This analysis reveals two characteristic peaks residing at distances of ~8 nm and ~21 nm (Figure 4.13D, dash lines). The first peak, which corresponds to the distance between nearest NCPs centers, moves from 7.5 nm to 8.1 nm, and to 9.0 nm for the early, intermediate and late SpnCs, respectively. The value of 7.5 nm can be rationalized as the distance between two NCPs averaged from their various contact conformations: the side-by-side conformation adopted by NCPs belonging to neighboring arrays and the stacking conformation of n and n+2 NCPs within the same array yielding the maximum and minimum core-to-core distance of 11 nm and 5.5 nm, respectively (Figure 4.13E; left panel). The rest of the

NCPs' contact conformations should contribute to distances that lie somewhere in between the above two extremes (Bilokapic et al., 2018b). The small shift in the peak distance as the SpnCs mature and grow might reflect a decrease in the statistical weight of intra-array n/n+2 stacking in favor of inter-array interactions. Similarly, the spherical nuclei and the grown SphCs with and without H1 also display a first peak shift toward higher values of 10.6 nm, 10.5 nm, and 10.6 nm, respectively (Figure 4.13D, blue dash line); perhaps the higher compaction in SphCs further discourage intra-array n/n+2 stacking in favor of inter-array NCPs interactions, a hypothesis supported by in silico simulations (Farr et al., 2021b). The second peak position (Figure 4.13D, pink dash line), which appears as a shoulder, matches the distance expected between n and n+1 NCPs that are connected by a 40 bp DNA linker in every array (Figure 4.13E; right panel). The prominence of this peak decreases monotonically from the early SpnCs to the grown SphCs, maybe reflecting the bending of the linker DNA within an array, and/or the increased statistical weight of neighboring NCPs belonging to different arrays as a result of the condensation process. Overall, the high-density SphCs have a slower g(r) decay indicating a higher chance to find neighbor NCPs at a close distance compared to early SpnCs, consistent with the concentration measurements.

The spherical shape of nuclei suggest that the formation of these structures involves unfavorable nucleosome surface-water interactions. Accordingly, we investigated if the orientation of NCPs on the surface of these condensates differs from those in the interior (see Methods). If the NCP disks are randomly distributed, the probability density distribution of their angular orientation should follow a cosine distribution, $P(\theta)d\theta = cos(\theta)d\theta$ (see Methods). Deviation from this distribution indicates a preferential orientation of the NCP disks either on the surface or in the interior of the condensates. The results of this analysis (Figure 4.14A-C) showed that the NCPs in the interior exhibited very small deviations from a random distribution, both in SpnCs and in SphCs with or without H1. Only NCPs on the surface of the initial spherical nuclei displayed a moderately skewed angle distribution, where the disks tended to orient more parallel to the condensates' surface. In contrast, the NCPs at the surface of the SpnCs, the SphCs, and the SphCs formed in the presence of H1, all showed an increased tendency to align perpendicular to the condensate surface ($\theta = 0$). Again, only for the initial nuclei, the NCPs at the surface tended to align parallel to the condensates' surface ($\theta = \pi/2$). Thus, the energetic and global structural differences between SpnCs and SphCs do not arise from a distinct spatial arrangement of NCPs within them. Instead, the transition between these two condensed forms must involve a change in the nature of the inter-nucleosome and nucleosome-solvent interactions (See discussion).

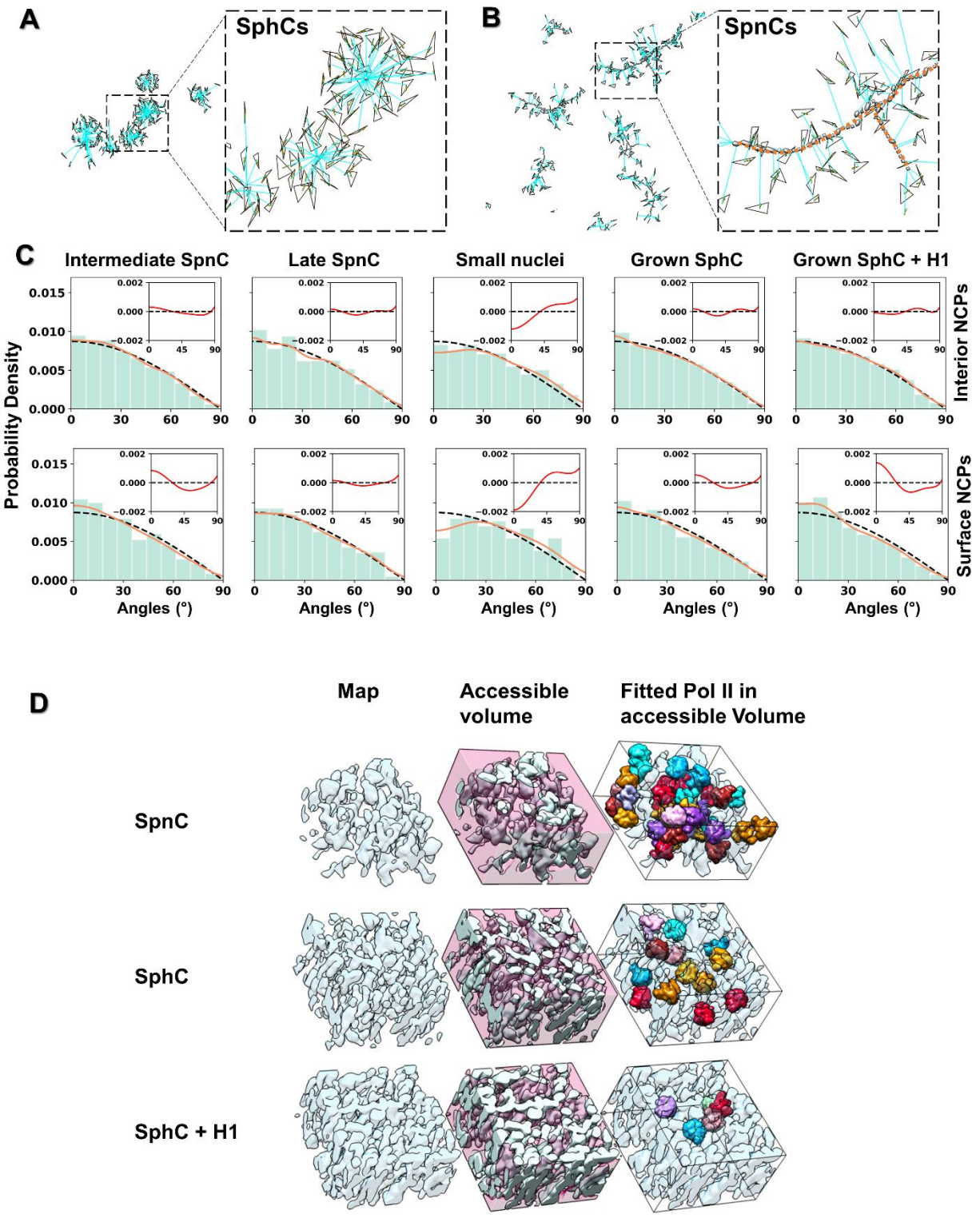


Figure 4.14 Quantitative analysis of NCP orientations and evaluation of chambers capacity for different types of condensates. A) Angle (θ) measurement of NCP disk plane (black triangle) relative to the radius of the SphC (cyan line) that crosses the center of the triangle (left panel). The zoom-in view is shown on the right panel. B) Angle (θ)

measurement of NCP disk plane relative to the radial line (cyan line, which is perpendicular to the axial axis of SpnC) that crosses the center of the triangle (left panel). The axial axis of the SpnC was determined as the skeleton of the map (orange dashed-line, on the right panel). C) Histogram of NCP disk angle distribution in the interior (top row) and on the surface (bottom row) of the condensates. The angle distribution was fitted with an 8th power polynomial (orange dashed line) and compared with a $cos(\theta)$ function (black dash line represents the angle distribution of a random rotated plane against a fixed axis). The deviation of experimental measurements from the *cos* was plotted on the top insets (red curves) indicating the preferred NCP disk orientation found within each type of condensate. D) Evaluation of chambers capacity for RNA Polymerase II (Pol II) for different types of condensates. Sampling volumes of 2.88 x 10^5 nm^3 (first column) from the final density maps of late stage SpnCs, SphCs, and SphCs in the presence of H1 (from top to bottom rows, respectively) were inverted and 3 nm low-pass filtered to show the detected empty chambers (second column). The capacity of the chambers was evaluated by quantitating the number Pol II molecules that fit within each condensate (third column, each Pol II was labeled with a random color). The number of Pol II molecules contained in the chambers from the late stage SpnCs, SphCs without and with H1, was 38,13, and 5, respectively.

4.5 Discussion and conclusion

We have used optimized cryo-ET to capture, for the first time, the 3D molecular organization of tetranucleosome arrays during the initial stages of their phase separation, when the condensates are only a few tens of nanometers in size. We have obtained snapshots of condensates at different incubation times in physiological salt, from which it has been possible to infer their likely process of growth and maturation. We show that nucleosome phase transition occurs through a two-step mechanism (Figure 7 and Extend Data Video 10). In the first step, dispersed nucleosomes (at 120 nM concentration) experience a fast system-wide condensation resulting in the ubiquitous appearance of irregularly shaped clusters. These condensates attain an NCP concentration of ~360 μM and average dimensions of (~125 nm x 65 nm x 30 nm). During their formation, they grow by enlarging their major axis, while their minor axes remain nearly constant. This step closely resembles a spinodal decomposition process in which aggregate formation occurs throughout the system without the crossing of an energy barrier. The condensates' anisotropic morphology has been described as a characteristic feature of spinodal processes at early times resulting from diffusion effects, and with isotropy regained at longer times driven by hydrodynamic effects (Datt et al., 2015). According to the Cahn-Hilliard theory of spinodal decomposition (Theory Section S1), the distance among the resulting SpnCs is determined by a "critical wavelength λ" that

describes the spatial pattern of separation between the two phases.

In the second step, the system undergoes a second condensation process in which highly packed spherical nuclei emerge locally and sparsely from the existing SpnCs. This step resembles a more traditional nucleation-and-growth process in which an initially unfavorable event (nucleation) is increasingly stabilized by a favorable volume energy through gaining NCP-NCP interactions distinct to those adopted in the SpnCs. These new inter-nucleosomal interactions must be accompanied, however, by an unfavorable surface energy with the solvent, which forces the system to adopt a spherical shape and minimize its surface-to-volume ratio. Throughout their growth, the initial small and unstable nuclei become increasingly stabilized through a process of accretion of the surrounding spinodal material, or by fusion with other spherical nuclei. We identify that the phase separation described here is a two-step process—as opposed to a simple spinodal phase that undergoes coarsening or Ostwald ripening (Ostwald, 1897), because we observed a clear separation of timescales between the formation of the two phases. The spinodal decomposition achieves a steady-state before 10 min, 2 min, and 10 min for the 4°C, 20°C, and 36°C reaction, respectively (indicated by the arrested growth of condensates' dimensions, Figure 4.3A). While the nucleation process was observed after 60 min, 2 min, and 10 min for reactions at 4°C, 20°C, and 36°C, respectively (Figure 4.3B). In contrast, a process of Ostwald ripening will occur continuously, with the formation of the matured SpnCs without a clearly defined time scale separation from the initial spinodal phase and without a nucleation step.

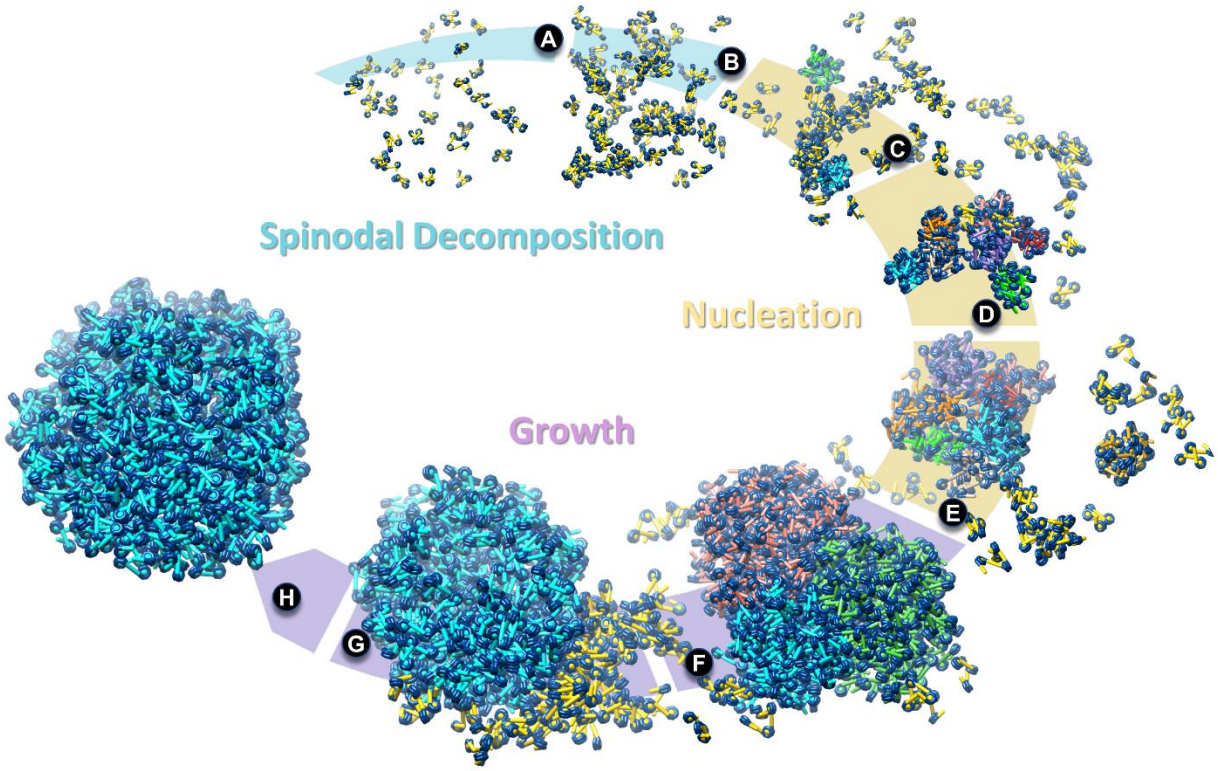


Figure 4.15 Proposed two-step mechanism of nucleosome phase separation. A) Initial, dispersed phase of tetranucleosome arrays; B) organization of tetranucleosome into irregular SpnCs; C) emergence of spherical nuclei within the spinodal phase and growth of these nuclei from accretion of the spinodal material; D) bundle of spherical nuclei; E) fusion of small nuclei bundles enables growth into larger, energetically stable SphCs; F) fusion between SphCs; G) growth of the SphCs by accretion of surrounding spinodal material; and H) large SphC corresponding to the early stages of liquid droplets. The light blue, pale brown, and purple background indicate the major steps of nucleosome condensation: spinodal decomposition, spherical nuclei formation, and growth of SphCs, respectively. H1 catalyzes step C-H. SpnCs are shown in yellow, and the SphCs are shown in multiple colors. All states are illustrated with experimentally observed condensates.

The 3D mapping of NCPs reveals that the small nuclei involve a tight packing of a large number of NCPs within a spherical volume as small as ~35 nm in diameter, attaining an NCP concentration of ~740 μM. The sparse appearance of these spherical nuclei at the beginning suggests that their formation is both kinetically and thermodynamically inaccessible from the initial solution of nucleosome arrays, and that they only arise from the pre-existing SpnCs as a result of rare, large local concentration fluctuations that are capable of overcoming a significant energy barrier. Given the

localized and rare appearance of these nuclei, it is somewhat surprising that we often observe them close together forming clusters. It is statistically unlikely that these condensates arose at widely separate loci in the medium and came together to form a bundle. Instead, it is more likely that large, localized concentration fluctuations gave rise to multiple nucleation sites at a given location. As these nuclei grow by accretion or fusion, their initial cores must rearrange so that the resulting SphCs display a decreased concentration (~470 μM).

The transition from SpnC to larger SphC is not accompanied by a significant difference in the orientation of NCPs relative to one another or to the outer surface of the condensates. Therefore, this transition with 1.3-fold increase in NCP concentration and a spherical re-organization, must involve a qualitative change in the interaction of the NCPs among themselves and with the solvent. Thus, we propose a model in which, in the spinodal process, the condensates retain significant amount of solvent, and the contacts of NCPs among themselves and with the solvent are mediated by polar interactions. These condensates do not experience a surface energy cost, as the difference between the interactions among NCPs in the interior and on the surface are not qualitatively different, only quantitatively so, as revealed by their angular orientation. During the nucleation step, in contrast, solvent must be significantly excluded, and it is likely that hydrophobic nucleosome surfaces are exposed and mediate the NCP-NCP and NCP-solvent interactions. Indeed, Swi6, an HP1-family protein in *S. pombe*, has been shown to induce LLPS of NCPs by promoting a conformational change that increases the solvent accessibility and dynamics of hydrophobic residues isoleucine, leucine and valine of the octamer core (Sanulli et al., 2019). Based on these observations, these authors have proposed that nucleosome plasticity is key for the action of Swi6 favoring nucleosomal states that promote LLPS by increasing multivalent interactions between NCPs. Other authors have also emphasized the relevance of hydrophobic residues as adhesive elements in LLPS (Alberti et al., 2019). Thus, treatment of living cells with 1,6-hexanediol (which inhibits weak hydrophobic interactions) caused a more uniformed chromatin distribution (Ulianov et al., 2021) and increased chromatin accessibility (Henikoff et al., 2020). We propose that while thermodynamically favorable in the condensates' interior, these interactions result in an unfavorable energy cost for NCPs at the surface, forcing the condensates to minimize the area exposed to the solvent by adopting spherical shapes. Interestingly, the isotropic distribution of the NCPs disks observed in the SpnCs and the larger, stabilized SphCs, shows that these condensates are not organized as a solid but as a liquid-like state.

The observation of a two-step phase separation pathway in the formation of large spherical phase droplets, as opposed to a conventional one-step (classical) nucleation

mechanism, has precedence in the literature. Examples include the phase separation observed in inorganic systems such as nanogold particles, Fe-Mn alloys, and in molecules of biological origin, such as ferritin (Loh et al., 2017b; Kwiatkowski da Silva et al., 2018; Houben et al., 2020).

Along the cell cycle, chromatin exhibits different levels of compaction. The internal NCP concentration of the irregular SpnCs adopted by the tetranucleosomes is in the range ~195-360 μM, larger than the estimated chromatin concentration at interphase (80-110 μM) (Hihara et al., 2012). However, if we take into account the characteristic wavelength λ of SpnCs, we obtain an NCP concentration similar to the interphase value. Interestingly, the average dimension of SpnCs along their short axes (~30 nm) suggests an inherent tendency of nucleosomal arrays to spontaneously self-assemble into structures of similar dimensions that have been describe as irregular (Cai et al., 2018c), short-range helical (Scheffer et al., 2011b), and regular ladder-like or regular helical (Song et al., 2014; Garcia-Saez et al., 2018; Robinson and Rhodes, 2006; Grigoryev et al., 2009). Here we found very few condensates displaying regular structure out of the great majority of irregular spinodal structures in physiological salt (Figure 4.3E; left column second row). We note, however, that the tendency of nucleosome arrays to form spinodal condensates could furnish the physical basis for a recent model of chromatin organization during interphase in which the 10-nm fibers self-interact and condense to form a polymer-melt structure (Hansen et al., 2018).

The transition from the SpnCs to SphCs involves a 1.3-fold increase in NCP concentration (~470 μM, Figure 4.13), which is in the range estimated for metaphase chromosomes (Hihara et al., 2012) and the concentrations predicted by Monte Carlo simulations (~500 μM) (Maeshima et al., 2016b). It is interesting to estimate how an increase in NCP concentration affects the ability of a molecule such as RNA Polymerase II (MW= 550 kDa) to diffuse inside the condensates, given that the flanking DNA segments of the arrays appear to be hidden inside the SphCs. Empty chamber evaluation analysis (Figure 4.14D) shows that concomitant to the increase of internal NCP concentration, there is a nearly 10-fold decrease in the number of Pol II molecules that could fit in these structures from the late stage SpnC to the SphC without H1, and then to the SphC in the presence of H1. This analysis indicates that SphCs, especially those formed in the presence of H1, could strongly repress transcription initiation, given the even larger dimensions of the transcription pre-initiation complex. In contrast, smaller molecules may still be able to diffuse inside liquid droplets. Indeed, linker DNA within the liquid droplets can be readily cleaved by micrococcal nuclease (17 kDa) (Maeshima et al., 2016b) and by Cas12 (120 kDa) (Strohkendl et al., 2021). Similarly, Tet Repressor (23 kDa) was shown to be strongly recruited to nucleosome droplets in the presence of

its operator (Gibson et al., 2019).

Formation of the early spherical nuclei requires at least 10 min of reaction, and it is accompanied by ~2-fold increase in local NCP concentration (~740 μM). This observation suggests that the transition from SpnCs to SphCs is the rate-limiting step in the formation of LLPS, making it a possible target of regulation. Indeed, we find that incubation of nucleosome arrays with H1 catalyzes the formation of SphCs at the expense of the surrounding spinodal material (Figure 4.11). It is not possible at this point to state whether or not the transition from SpnCs to SphCs takes place in vivo, with H1 and other protein factors reducing the energy barrier for the transition to take place. If that were the case, other protein factors and histone post-translational modifications could be needed to maintain an open chromatin state during interphase. For example, proteins such as FoxO1 are able to bind nucleosomes and induce chromatin opening (Hatta and Cirillo, 2007) and histone acetylation has been shown to disrupt chromatin droplets (Gibson et al., 2019). Future efforts will be directed toward determining how histone acetylation, methylation, and removal of core histone tails affect the early stages of nucleosome spinodal condensation and spherical nuclei formation.

Collectively, it is tempting to extrapolate our *in vitro* observations and propose that in the cell, chromatin can exist in an equilibrium between a loosely packed state (similar to SpnCs), which would be fully accessible for transcription, repair, and replication, and a highly packed transcriptionally repressed state (resembling the SphCs). In this proposed scenario, it would be through a tight spatial and temporal regulation of the interconversion between these two chromatin forms that cell identity and differentiation could be established, and so explain how physiological homeostasis could be attained in response to external cues.

4.6 Materials and methods

4.6.1 Purification of histones proteins

Purification of *Xenopus laevis* histones H2A, H2B, H3, and H4, was performed as previously described (Wittmeyer et al., 2004). Briefly, histone plasmids were transformed into *E. coli* BL21(DE3) and bacterial cells were grown while shaking at 37°C in LB media supplemented with ampicillin and chloramphenicol. Histone expression was induced at an OD_{600} of ~0.6 by adding Isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 1 mM. After 3 h of induction, cell cultures were pelleted, suspended in wash buffer (50 mM Tris-HCl pH 7.5; 100 mM NaCl; 1 mM EDTA; 5 mM 2-mercaptoethanol (BME); 1% Triton X-100 [w/v]; and protease inhibitors

(Roche)) and frozen at -80°C for later use. To purify the inclusion bodies, cell pellets were thawed, sonicated, and centrifuged at 25,000 x g for 30 min. Supernatant was discarded, and the pellet, corresponding to the inclusion bodies, was rinsed, suspended in wash buffer, and centrifuged three times. To remove Triton X-100, inclusion bodies were rinsed and centrifuged three times with wash buffer without Triton X-100. Inclusion bodies were solubilized with a buffer containing 20 mM Tris-HCl pH 7.5; 8 M Urea; 1 mM EDTA; 10 mM DTT, and purified by anion and cation exchange. Purification was checked by 15% SDS-PAGE and fractions containing pure histones were pooled and dialyzed against 1 L of 10 mM Tris pH 8.0, with three buffer exchanges. Histones were centrifuged to remove aggregates, concentrated by centrifugation (~10 mg/mL), lyophilized, and stored at -80°C.

4.6.2 Fluorescence labeling of H2A

For fluorescence microscopy experiments, *Xenopus laevis* histone H2A119C (The Histone source, Colorado State University) was chemically modified with the fluorescent dye Cy3 (Cyanine3 maleimide; Lumiprobe) as previously shown, with some modifications (Zhou and Narlikar, 2016). Lyophilized H2A119C was resuspended to 2 mg/mL using labeling buffer (20 mM Tris-HCL, pH 7.0; 7 M guanidine hydrochloride; 5 mM EDTA) and incubated for 2 h at room temperature. Unfolded histone was reduced by the addition of 100 mM TCEP (Sigma) and incubated for 2 h in the dark with occasional mixing. A second round of 100 mM TCEP reduction was done overnight. Cy3 maleimide was dissolved with dimethylformamide (DMF) to a final concentration of 20 mM and H2A119C was labeled by the addition of 15-fold molar excess of Cy3 maleimide (~2 mM). Labeling reaction was performed at 21°C overnight, in the dark, and with constant shaking. To remove non-conjugated Cy3 maleimide, the labeling product was concentrated by centrifugation using a Microcon of 3.5 kDa membrane cutoff. Concentrated Cy3-histone was diluted with labeling buffer and concentrated again. This procedure was repeated six times, and every other time, the Microcon filter was changed for a new one. Concentration of labeled histone and degree of Cy3 labeling was ~50%, which was calculated by measuring the absorbance at 278 nm and 555 nm and using the extinction coefficients of 4,400 Mcm^{-1} and 150,000Mcm^{-1} for H2A119C and Cy3, respectively.

4.6.3 Histone octamer reconstitution

To reconstitute the wild type histone octamer (Wittmeyer et al., 2004) and the Cy3-labeled octamer (H2A-Cy3 replaces H2A), lyophilized histones were solubilized with unfolding buffer (20 mM Tris-HCl pH 7.5; 7 M guanidine hydrochloride; 10 mM DTT), combined at a H2A:H2B:H3:H4 ratio of 1.2:1.2:1:1, and dialyzed 4 times against 1 L of refolding buffer (10 mM Tris-HCl pH 8.0; 2 M NaCl; 1 mM EDTA; 5 mM DTT), with

three buffer exchanges, for a total of 48 h. Refolded octamer was centrifuged, concentrated to ~0.5 mL, and loaded onto Superdex 200 Increase 10/300 GL (Cytiva) equilibrated with refolding buffer. Gel filtration separated octamers from aggregates, tetramers, and dimers. Fractions were checked by 15% SDS-PAGE, and fractions containing the four histones in equimolar quantities (based on Coomassie blue staining) were pooled, concentrated to ~10 mg/mL, and stored at -80˚C. Fractions containing H2A-H2B heterodimer were also pooled, concentrated to ~10 mg/mL, and stored at -80˚C.

4.6.4 Synthesis of tetranucleosome and dodecamer array DNA templates

The tetranucleosome DNA template array contains four repeats of the 601-nucleosome positioning sequence separated by a 40 bp linker length, and flanked to the left by 100 bp DNA, and to the right by 50 bp DNA (100-4W-50; left and right sides correspond to the 5′ and 3′ ends, using the 601 sequence (W) as a reference). Tetranucleosome DNA template was synthetized by ligation of four fragments containing one 601 sequence each. Each fragment was produced by PCR using the pGEM-3Z/601 plasmid (Addgene), which contains the 601 nucleosome positioning sequence. Primers containing the restriction recognition sites for enzymes BbsI and BsaI (NEB) were used to amplify the 601 sequence. BbsI and BsaI recognizes asymmetric DNA sequences and cut outside of their recognition sequence, allowing the ligation in tandem of DNA fragments, and the control over the number of repeats. In our design, the tetranucleosome is flanked by two BsaI sites, and the BbsI sites were used to ligate the four nucleosome fragments. In the first cleavage step, the four fragments were digested with BbsI and ligated at an equimolar ratio using *E. coli* DNA ligase (NEB). The longest product, corresponding to the tetranucleosome, was purified from 0.8% agarose gels, digested using the BsaI enzyme, and cloned into BsaI-restricted pGEM-3Z/601 plasmid using T4 DNA ligase (NEB). Ligation product was transformed into *E. coli* DH5α for plasmid extraction by miniprep. The product of the tetranucleosome DNA synthesis was checked by DNA sequencing. The dodecamer nucleosome array was synthetized by ligation of three tetranucleosome templates. Resulting ligation product was digested using BsaI enzyme and cloned into BsaI-restricted pGEM-3Z/601.

To produce large quantities of the tetranucleosome DNA template, the tetranucleosome plasmid was grown in *dam*$^-$/*dcm*$^-$ *E. coli* cells (NEB), and it was purified by maxiprep. Tetranucleosome DNA template was excised from the vector backbone by digestion using the BsaI enzyme. Tetranucleosome template was purified by preparative 5% acrylamide electrophoresis using the Model 491 Prep Cell (Bio-Rad) electroelution system. The dodecamer array was produced and purified in the same way as tetranucleosome DNA template, but using preparative 4% acrylamide electrophoresis.

4.6.5 Synthesis of biotinylated tetranucleosome and dodecamer arrays

To incorporate a biotin molecule to the 100-4W-50 template without modifying its length, a 70-4W-50 template was ligated with a biotinylated 30 bp dsDNA to produce the biotin-100-4W-50 DNA template. The 70-4W-50 template was designed and produced in the same way as the 100-4W-50. The 30 bp dsDNA was synthesized by the annealing of two complementary oligos (IDT), forming a 5′ biotinylated end and a 3′ end with a BsaI overhang, which is complementary to the left side of 70-4W-50. Biotin-30 bp ds oligo and 70-4W-50 were ligated using T4 ligase at a molar ratio of 20:1, respectively. Ligation reaction was inactivated with SDS loading buffer (NEB), and the biotin-100-4W-50 product was purified by preparative 5% acrylamide electrophoresis using the Model 491 Prep Cell system.

4.6.6 In vitro reconstitution of tetranucleosome and dodecamer arrays

Wild type histone octamers and tetranucleosome DNA templates (100-4W-50 and biotin-100-4W-50) were combined at a ratio of 1:1.2, respectively, in 200 μL of high-salt buffer (10 mM Tris-Cl pH 8.0; 2 M NaCl; 1 mM EDTA; 0.5 mM DTT; and 1 mM PMSF) at a final concentration of 100 ng/uL of DNA template. H2A-H2B dimer was also incorporated at a molar ratio of 0.2 compared to the octamer to promote nucleosome, rather than hexasome formation. For fluorescent nucleosome arrays, wild type octamers were doped with 25% Cy3-octamer and combined with the 100-4W-50 template. Tetranucleosome assembly solutions were dialyzed against 500 mL of high-salt buffer for 1 h at 4˚C, followed by a lineal gradient dialysis against 2 L of low-salt buffer (10 mM Tris-Cl pH 8.0; 1 mM EDTA; 0.5 mM DTT; and 1 mM PMSF) using a peristaltic pump at 0.8 mL/min and with continued stirring. A final dialysis of 3 h in 500 mL of low salt buffer was done to reduce the residual NaCl concentration, and the nucleosome reconstitution was checked by 4% acrylamide native electrophoresis.

Reconstituted arrays were loaded into 4.8 mL of 10-40% lineal sucrose gradient (20 mM HEPES-NaOH pH 7.5; 1 mM EDTA; 1 mM DTT) and centrifuged for 16 h at 38,000 rpm at 4˚C, using an ultracentrifuge Beckman Optima MAX-XP with the rotor MLS-50 (Beckmann). Gradients were fractionated into 100 μL fractions using a gradient fractionator (Brandel), and the resulting purification was checked by 4% acrylamide native electrophoresis. Fractions containing 1-2 bands were pooled, concentrated, dialyzed against 20 mM HEPES-NaOH pH 7.5; 1 mM EDTA; and 1 mM DTT, and stored at 4˚C. Tetranucleosome purification was checked by atomic force microscopy (AFM). The dodecamer arrays were synthetized and purified in the same way as tetranucleosomes, but the loading was verified by 0.8% TBE agarose gel and atomic force microscopy (AFM).

4.6.7 Atomic force microscopy of nucleosome arrays

Purified tetranucleosome and dodecamer arrays were diluted to 10 nM and crosslinked with 1% formaldehyde for 1 h at room temperature. Crosslinked sample was dialyzed against 20 mM HEPES-NaOH pH 7.5; 1 mM EDTA; 1 mM DTT, and centrifuged at 20,000 x g to remove aggregates. Crosslinked nucleosome array samples were diluted to 1 nM using 10 mM MOPS pH 7.0 and 5 mM $MgCl_2$, and 3 μL of sample were deposited and incubated for 2 min on freshly cleaved bare mica V1 (Ted Pella Inc.), after which was rinsed with Milli-Q water, and then gently dried under a stream of N2 perpendicular to the mica surface. AFM micrographs were taken with a MultiMode NanoScope 8 atomic force microscope (Bruker Co.) equipped with a vertical engagement scanner E. The samples were excited at their resonance frequency (280-350 kHz) with free amplitudes (Ao) of 2-10 nm and imaged in tapping mode using silicon cantilevers (Nanosensors). The image amplitude (set point As) and A0 ratio (As/A0) was kept at ~0.8 in a repulsive tip-sample interaction regime, and phase oscillations were no greater than ± 5 degrees. The surface was rastered following the fast scan axis (x) at rates of 2 Hz, capturing the retrace line to reconstruct the AFM micrographs. All samples were scanned at room temperature in air, at a relative humidity of 30%.

4.6.8 Tetranucleosome phase transition visualized by epifluorescence microscopy

For phase transition experiments using fluorescence microscopy, tetranucleosome stocks 100-4W-50 (300 nM) and Cy3-100-4W-50 (300 nM) were combined at a molar ratio of 3:1, respectively, in 10 mM HEPES-NaOH pH 7.5; 1 mM EDTA; and 1 mM DTT. Nucleosome phase transition was induced by diluting the tetranucleosome stocks tenfold in 50 mM Tris-HCl pH 7.5; 150 mM NaCl; 5 mM $MgCl_2$; and 1 mM DTT. Phase transition reactions were incubated at room temperature, and 4.5 μL aliquots were combined at different times with 0.5 μL of 10X GODCAT visualization buffer (50 mM Tris-HCl pH 7.5; 2 mg/mL Glucose Oxidase (SIGMA cat. no. G2133); 350 ng/mL Catalase (SIGMA cat. no. C1345); 20% Glucose; 10 mg/mL AcBSA (Thermo); 50% glycerol) and added into the PEGylated and BSA passivated slide/coverslip chamber.

To prepare the visualization chamber, glass coverslips and slides were passivated using polyethylene glycol (PEG) as previously described (Chandradoss et al., 2014). A microfluidic chamber was assembled with both the PEG-coverslip and the PEG-slide, and the channel formed between both surfaces was blocked for 30 min by the addition of 1 mg/mL of AcBSA (Thermo) in 20 mM Tris-HCl pH 7.5; 50 mM NaCl, and 1 mM EDTA, followed by rinsing with 50 mM Tris-Cl pH 7.5; 150 mM NaCl; 5 mM $MgCl_2$; and 1 mM DTT. 5 μL of the phase transition reaction combined with the GODCAT was

applied to the channel chamber and visualized using epifluorescence microscopy. Nucleosome phase transition in the presence of human H1.0 linker histone (The Histone Source, Colorado State University) was performed in the same way as explained above, but including H1 in the phase transition reaction at a concentration of 120 nM to keep a ratio of H1: nucleosome 1:1.

Epifluorescence micrographs were taken using a laboratory built wide-field fluorescence microscope. An oil-immersion 100x objective lens (Olympus 100× UPlansApo, N.A. 1.4) was used with a diode-pumped 532-nm laser (75 mW; CrystaLaser) for excitation and a red LED for brightfield illumination. Fluorescence emissions and brightfield illumination were splitted via a laboratory-built multichannel imaging system and projected side-by-side on an EMCCD (IXon EM+ 897; Andor). The illuminated area on each channel was ~50 × 25 μm with ~110 nm pixel widths.

4.6.9 Fluorescence recovery after photobleaching (FRAP) of tetranucleosomes

FRAP was performed on 300 nM H2A-Cy3 tetranucleosome droplets using a Nikon TIRF inverted scope (Nikon Eclipse Ti2, 488/560/642nm OPSL lasers) with perfect focus, a 100x N.A. 1.4 APO TIRF oil objective, Nikon TIRF/iLAS2 FRAP Module (Gataca Systems) and an EMCCD camera (Photometrics Prime 95B). The droplets were imaged every 5 s for 15 s of pre-bleach, and every 5 s for 250 s after photobleaching. Photobleaching was performed with a 561 nm laser (100% laser power; 50 μs irradiation). FRAP data was processed by the normalization method. The first-post bleach point was set to zero.

4.6.10 Nucleosome Array Phase transition visualized by OpNS-EM

Tetranucleosome stocks 100-4W-50 and biotin-100-4W-50 were combined at a molar ratio of 3:1, respectively, in 20 mM HEPES-NaOH pH 7.5; 1 mM EDTA; and 1 mM DTT. Nucleosome phase transition was induced by diluting tenfold the tetranucleosome stocks in buffers with different salt concentrations (20 mM HEPES-NaOH pH 7.5; 0-150 mM NaCl; 0-5 mM $MgCl_2$; and 1 mM DTT) at a final tetranucleosome concentration of 30 nM (This value has been used for all of the following incubation experiments. And to keep the total nucleosome concentration consistent, the dodecamer arrays were diluted to 10 nM during incubation). Phase transition reactions were incubated at 4°C, 20°C, and 36°C, and were analyzed at different incubation times by the optimized negative staining protocol (OpNS) as previously described (Rames et al., 2014b). In brief, 4 μL of the phase transition reaction was deposited on a glow-discharged ultra-thin carbon-coated 200-mesh copper grid (CF200-Cu-UL, Electron Microscopy Sciences, Hatfield, PA, USA, and Cu-200CN, Pacific Grid-Tech, San Francisco, CA, USA). After 30 seconds of incubation, phase transition solution on the grid was blotted with filter paper. The grid was stained for 10 s with 30 μL of 1% (w/v) uranyl formate (UF), UF was blotted

with filter paper, and the staining procedure was repeated two more times, followed by air-drying with nitrogen. OpNS images were acquired under a defocus of 0.5-1 μm on a Zeiss Libra 120 Plus transmission electron microscope (TEM) (Carl Zeiss NTS GmbH) equipped with a Gatan UltraScan 4 k × 4 k CCD. The TEM was set to a high-tension of 120 kV with energy filtering at 20 eV. Carbon-coated grids are usually glow-discharged in residual air to promote the absorption of biological material deposition from the aqueous solution, a process that makes the surface hydrophilic by depositing negative charges. However, the negatively charged surface does not efficiently bind the negatively charged nucleosome arrays at low ionic strength (due to weak shielding effect, see Figures 4.4E; blue section). An analysis of the effect of Na^+ and Mg^{2+} within a 10 min incubation interval indicated that in absence of Na^+, spherical condensates were formed above 0.5 mM Mg^{2+}, but irregular condensates were not observed (Figure 4.4E, left column). When tetranucleosomes were incubated in absence of Mg^{2+}, tetranucleosomes formed irregular condensates above 1.5 mM Na^+ (Figure 4.4E, middle column). When Mg^{2+} and Na^+ are both present, high concentration of Na^+ caused a delayed formation of SphCs (Figure 4.4E, right column).

4.6.11 Preparation of streptavidin 2D crystal grids and cryo-EM sample deposition

The streptavidin (SA) crystal grids were prepared as previously described (Han et al., 2016) with a few modifications. Lacey carbon grids with 5 μm holes (LC200-Au-FF, Electron Microscopy Sciences) and Quantifoil carbon grids with 7 μm holes (S 7/2, Electron Microscopy Sciences) were cleaned with 100% chloroform to remove the plastic cover. Since functionalization of carbon grids involves hydrophobic interactions, no glow discharge was applied on the grids. To form a biotinylated monolayer of lipids, 30 μL of castor oil (BRAND) was added to 5 ml of crystallization buffer (50 mM HEPES-NaOH pH 7.5; 150 mM KCl; and 5 mM EDTA) in a 50 mm plastic petri dish, followed by the addition of 1 μL of 1 mg/mL Biotinyl Cap PE (1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine-N-(cap biotinyl); Avanti). To functionalize the grids, cleaned carbon grids were put in contact with the lipid monolayer for 1 s, then washed three times with 100 μL of crystallization buffer. To add the streptavidin surface to the grid, 4 μL of 0.2 mg/mL of streptavidin (NEB) in crystallization buffer were added to the biotinylated lipid surface and incubated for 30 min in a humidity chamber at room temperature. To remove excess of SA, SA-crystal grids were washed and blocked with 30 μl of rinsing buffer (10 mM HEPES-NaOH pH 7.5; 50 mM KCl; 5 mM EDTA; and 10% trehalose) followed by blotting and air drying. A thin layer of carbon was deposited by evaporation onto the backside of the SA crystal grid to protect the crystal from being damaged during grid manipulation. SA-crystal grids were stored at room temperature in a sealed box for further use.

For sample deposition, excess trehalose was removed from SA crystal grids by rinsing the grids two times with 30 μL of crystallization buffer for 10 min. Grids were blotted at the edge with filter paper, then 4 μL of the tetranucleosome phase transition reactions (prepared as for OpNS-EM visualization) were added to the SA grids and incubated in a humidity chamber to allow for biotin and streptavidin binding. After 2 min, grids were blotted for 2 s and immediately flash frozen in liquid ethane at 100% humidity to form an amorphous ice, using a Leica EM GP rapid-plunging device (Leica, Buffalo Grove, IL, USA). Humidity chamber and the plunging device were set to the temperature used in each phase transition reaction.

4.6.12 Cryo-ET data acquisition

We found that the negatively charged nucleosome arrays did not efficiently bind to the negatively charged carbon grids. An increase ionic strength facilitates the nucleosome grid surface deposition, but also affect the nucleosome condensates formation (Figure4.5E). To avoid using a charged grid surface, we used the 2D streptavidin crystal grid (Han et al., 2016) and tetranucleosome arrays harboring a biotin molecule at one DNA end for sample deposition. The similar appearance of large condensates on grids prepared with and without the streptavidin crystal, after long incubation times indicated that the condensates form in solution and are then absorbed onto the SA-crystal (Figure 4.5D).

Frozen nucleosome samples on SA-crystal grids were subjected to tomography data collection on a FEI Titan Krios transmission electron microscope, operated at 300 kV with an energy filter at 20 eV (Gatan BioQuantum). Tilt series were acquired with SerialEM (Schorb et al., 2019) using the automated data collection function, including fine eccentricity, item realignment, and autofocus scheme. The range for tomography collection was set from -60° to +60°, with an angular increment of 3° per tilting step. The target defocus was set between 2.5-3 μm. The total electron dose for the whole tomography was calculated between 200 and 250 electrons per $Å^2$, with an adjustment of the exposure time following the 1/*cos* schemes. For each tilt angle, 8-10 frames were acquired every 0.25 s from 2 to 2.5 s of exposure using a Titan Krios Gatan K2/K3 camera with a pixel size of 1.45/1.47 Å.

4.6.13 Deep learning-based denoising

For deep learning denoising, we followed the description of the TOMO2TOMO (T2T) method (Buchholz et al., 2019). Frames of tilt angles were aligned using MotionCor2 (Zheng et al., 2017) to correct for image drift. Aligned frames were split into even and odd frame stacks and averaged separately to generate noise independent pairs

for the same object. To select representative areas of the averaged image pair, 1,120 local box path-pairs of 600x600 pixels were generated at each apparent nucleosome core particle (NCP) position. Inside each NCP position, 10 patches of 128 × 128 pixels were further randomly cropped to train a CARE network in the NOISE2NOISE regime (Weigert et al., 2018). The neural network was constructed with a U-Net of depth two, a kernel size of three, one hundred training epochs, and a linear activation function at the last layer. The loss function was evaluated with a per-pixel mean squared error (MSE). After training, the network was applied to restore the full-size image pairs at each tilt. The final image of each tilt is the result of averaging two individual restorations of even and odd images.

4.6.14 Preprocessing of cryo-ET tilt series

The denoised tilt series were pre-aligned using the IMOD software package with the patch tracking method (Kremer et al., 1996). Then, further alignment was performed on the local image patches (200 x 200 pixels) by using the focused refinement strategy in individual-particle electron tomography (IPET)(Zhang and Ren, 2012). The alignment of the image tilt axis onto its central vertical axis allowed for the determination of the Contrast Transfer Function (CTF) of the microscope, using the GCTF software package (Zhang, 2016). CTF of the center slice of each micrograph and the cosine of the tilt axis angle were used to estimate the CTF of the flanking strip bands. Both the phase and amplitude were corrected for each strip band using TomoCTF (Fernandez et al., 2006).

4.6.15 Cryo-ET 3D reconstruction and NCPs segmentation

To obtain the single-molecule 3D images, the CTF corrected tilt series were then submitted for a focused electron tomographic refinement (FETR) algorithm in IPET (Zhang and Ren, 2012) followed by 3D reconstruction using weighted Fourier back projection using the e2tomo software package (Chen et al., 2019; Tang et al., 2007a) without a further alignment step. The reconstructed maps were subjected to low-pass (abs= 0.25), normalization and median shrink filters enhancement (bin= 8; pixel size= 11.68 Å). To implement the deep learning-based segmentation method in e2tomo, areas containing representative features for network training (crystal lattice, free distal DNA, NCPs, and background noise) were boxed out from z-dimensional slices of the reconstruction. The application of the trained network to the tomograms allowed for the segmentation of the NCPs, DNA, and SA crystals from the background. NCP segmentation was used as a marker to select surrounding connected map densities (within 10 Å) using the zonsel function of Chimera software (Pettersen et al., 2004). SA crystal segmentation was used as a marker to remove the surrounding SA crystal densities (within 10 Å). To produce a soft boundary mask for the missing wedge correction, the processed density maps were low-pass filtered to 30 Å and clipped to box

of 512 x 512 x 200 pixels.

4.6.16 Missing wedge correction of 3D density map reconstruction

To reduce the tilt limitation caused the elongation artifact in the 3D reconstruction, the missing wedge correction was performed by the low-tilt tomographic reconstruction (LoTToR) method (Zhai et al., 2020), which is a model-free iterative process, under a set of constraints in the real and reciprocal spaces, aiming to fill the missing wedge zone in Fourier space. To implement the algorithm, a soft boundary 3D mask was generated using the previously denoised density map after 30 Å low-pass filtering. This 3D mask was paired with the initial 3D reconstruction to generate nine cubic tiles of 200 x 200 x 200 pixels with overlapping regions of 30 x 200 x 200 pixels as reference for realignment. These cubic tiles were subjected to iterations of the missing wedge correction until the final corrected 3D map converged with a stable Fourier shell correlation value. The final 3D density map of 512 x 512 x 200 pixels was reassembled by aligning the corresponding missing wedge corrected density cubes.

4.6.17 Modelling of the final denoised 3D electron density map

A 3D global template-matching was obtained by competitively docking the nucleosome model (PDB ID:1AOI) and a linear 40 bp DNA model into the final missing wedge corrected density map. This procedure consisted of four major steps: i) NCP segmentation maps were split into domains using the watershed method of Chimera (4 smoothing steps with a step size of 3 voxels). High-density weight centers of each domain were calculated at different map contour levels (from 10-90% of its maximum in a step of 0.5). ii) NCP and 40 bp DNA models were fitted into the weight centers with a random shift and rotation. iii) The competitive docking of positioned NCP and DNA models was determined using the sequential multi-model fitting function in Chimera (Pettersen et al., 2004). All docked models were ranked based on the cross-correlation value (CC-score) between the models and the map. Models with either a CC-score < 0.7 or exhibiting a clash event with higher-ranked models were removed from the analysis. iv) Steps ii) and iii) were repeated multiple times until a stable fitting was identified as explained below.

A stable fitting of each map domain was obtained by the process of maximizing the average CC-score and minimizing the map/model fitting residue (MaFR/MoFR). MaFR corresponds to the remaining density volume after subtracting the combined model from the density map, and MoFR corresponds to the remaining density volume after subtracting the density map from the combined model. A score to evaluate the minimization of the fitting residue compared the current fitting with the best previous fitting and was defined as the ratio between ($MaFR_{current}$ - $MaFR_{best}$)/$MaFR_{best}$ and

($MoFR_{current}$ - $MoFR_{best}$)/$MoFR_{best}$. A negative value of both ($MaFR_{current}$ - $MaFR_{previous}$) and ($MoFR_{current}$ - $MoFR_{previous}$) indicated that a better fitting was found. A score threshold of < -0.1 was set to accept the current fitting as the best fitting. The fitting loop was stopped when no better fittings were found in ten consecutives iterations. The translations and rotations of each NCP and DNA linker models were recorded for the statistical analysis described in the following sections.

A 3D particle stack was generated by cropping out identified NCPs within the tomograms. This particle stack was submitted to the e2tomo program (Chen et al., 2019) for sub-tomogram averaging in order to further validate the quality of the selected nucleosome particles. We carried out the 3D refinement routine with a single featureless discoidal-shaped initial model (niter= 3, sym= c1, and without masks), and calculated 3D averaging for different fractions of particles (pkeep= 1.0, 0.8, 0.6, and 0.4). The map resolutions were estimated by measuring the Fourier shell correlation (FSC) between two independently determined half-maps and between map and crystal structure 1AOI when the coefficient fell below 0.5 and 0.143.

4.6.18 Identification of free and condensed NCPs

The weight centers of NCPs in the fitted model were classified based on their distance from their neighbors, using the density-based spatial clustering of applications with noise algorithm (DBSCAN) (Ester et al., 1996). DBSCAN required two input parameters, *eps* and *min_samples*, where *eps* defines the maximum distance between two points to be considered as neighbors, and *min_samples* defines the minimum number of neighbors to be considered as clusters. An initial classification to identify free NCPs and condensates was performed using an *eps* value of 250 Å, which is the minimal theoretical distance between two consecutive nucleosomes in the same array (2 x 55 Å NCP (disc radius) + 40 bp DNA x 3.4 Å/bp= 250 Å), and a *min_sample* value of 8, which is, by definition, the smallest cluster composed of two tetranucleosome units. A second classification was performed based on the previous classification, using an *eps* value between 270-290 Å, and a *min_sample* value of 12. This produced a cleaner segmentation that differentiated free NCPs from NCPs grouped in condensates, representing the light and dense phase of spinodal decomposition, respectively. In the case of small SphCs that interact among themselves or with irregular condensates, a subclassification was manually performed to distinguish each SphC within the cluster.

4.6.19 Calculation of tetranucleosome condensates geometry

The shape of irregular and spherical nucleosome condensates was determined by measuring its physical dimensions and eccentricity. The condensate dimensions are represented by the length of its long and short axis. The direction of the long and short

axis was calculated using the coordinates of the center of mass of NCPs and the principal component analysis (PCA)(Pearson, 1901) which produced the vector PC1 for the long axis and the vector PC3 for the short axis. The lengths of the long and short axis were calculated as the distances between the two most separated NCPs along the PC1 and PC3 vectors. The overall size of the condensate was estimated by averaging the length of the long and short axis. The eccentricity was calculated using the formula "1- (short axis length / long axis length)" which yield a score between zero and one, where zero represents a perfect spherical shape, and one represents an elongated structure.

4.6.20 Calculation of NCP condensate boundary sharpness

To define the overall shape of condensates, a boundary mask was produced using the electron density of condensates with a 10 nm low-pass filter. To calculate the surrounding free NCPs concentration, which represents the condensate boundary sharpness, a shell of 20 nm extending from the boundary mask was established to determine the number of the surrounding free NCPs, which in turn was divided by the volume of the external 20 nm shell.

4.6.21 Determination of NCP concentration in condensates

The nucleosome ice-slab (global) concentration for the early stage of phase separation (when light phase and dense phase are formed under ~100 nm scale) was calculated by dividing the total number of NCP within the tomogram by the volume of the smallest cuboid that enclosed all NCPs. To measure the concentration of NCPs in SpnCs and SphCs, an initial condensate-shaped mask was produced based on the electron density after a 10 nm low-pass filter, and a series of 3D contours were established either by shrinking or expanding the mask in steps of 2 nm. The innermost contour was labeled as n= 1, and the next consecutive contour was labeled as n + 1. A shell defined as the 3D space between two consecutive contours was also labeled starting from inside the condensate (n= 1). To calculate the concentration of NCPs inside a specific contour c, the number of NCPs within contour n was divided by the volume enclosed by the contour. To calculate the concentration of NCPs inside a specific shell n, the number of NCPs within the shell n was divided by the volume of the shell. To avoid double counting NCPs that cross two shells or contours, the center of mass of NCP was used to determine its residing shell or contour. The mean NCP concentration was obtained by averaging the NCP concentration inside a specific contour or shell n at the same position in the same type of condensates. Since condensates vary in size, the total number of contours within a condensate is different from each other, which may cause more variance in concentration measurements for the exterior contour/shell.

4.6.22 Calculation of the pair distribution function ($g(r)$) of NCPs

within condensates

The *g(r)* function was determined using the distribution of the pairwise distances between the weight centers of NCPs. For each NCP weight center, a series of spheres separated by 1.6 nm (*dr*; small step of radius extension) were generated using a radius (*r*) ranging from 0 to 60 nm. This resulted in shells formed by two consecutive concentric spheres (i.e., the shell between *r* and *r* + *dr* away from the reference NCP weight center). The NCP density of each shell was calculated by dividing the number of NCPs within the shell by the shell volume ($4\pi r^2 \, x \, dr$), and the shell densities of different spheres with the same radius were averaged and divided by the ice-slab (global) density for normalization. However, the shell density estimation for NCPs that reside at the border of the tomography is not accurate. To correct this, a periodical boundary was applied to the *x*- and *y*-direction of NCP models, lengthening the respective dimensions of the tomogram 2-fold. If the shell surpasses the *z*-direction boundary of the tomogram, a correction factor (the ratio between the overall shell volume and the partial shell volume inside the tomogram) was multiplied with the corresponding shell density because the *z*-direction is non-extensible, due to contact with ice (top) and the SA-crystal surface (bottom). For tomograms containing SphCs, only the condensate region was selected for *g(r)* analysis, and due to the large background density variation among different condensates, the *g(r)* density was normalized by setting the highest peak equal to one.

4.6.23 Calculation of the NCPs orientation within condensates

To calculate the orientation of NCPs on the surface and in the interior of different types of condensates, each NCP within the fitted model was simplified as a triangular shape which centered on the nucleosomal discoidal plane. For SphCs, the orientation of each NCP relative to its containing condensates was measured as θ which is the angle between the triangular plane and the radius of the condensate that crosses the triangle center (i.e., the center of mass of NCP). Similarly, for irregular shape condensate that cannot define the radius, θ was measured as angle between the triangular plane and the line that is perpendicular to the spine of the condensate and crossing the triangle center. To calculate the spine of the condensate, the density map of the condensate was eroded to irregular lines by using Lee's 3D skeletonizing algorithm (Lee et al., 1994). The θ angle distribution of NCPs for each type of condensate was evaluated by comparing with a $cos(\theta)$ function, which represent a randomly distributed orientation of NCPs. The $cos(\theta)$ represents the distribution of orientations of random unit vector relative to a fixed plane, which in turn equivalent to the probability of finding a point on the unit hemisphere contained in a differential ring-shape area: $P(\theta)d\theta = 2\pi(Rcos(\theta))Rd\theta/2\pi R^2 = cos(\theta)d\theta$.

4.6.24 Key source table

REAGENT or RESOURCE	SOURCE	IDENTIFIER

Bacterial strains		
dam-/dcm- Competent E. coli	New England Biolabs	C2925I
DH5α E. coli competent cells	Macrolab, QB3-Berkeley	N/A
BL21(DE3) pLysS E. coli competent cells		
Chemicals, Peptides, and Recombinant Proteins		
Cyanine3 maleimide	Lumiprobe	21080
TCEP (Tris(2-carboxyethyl) phosphine hydrochloride)	Sigma-Aldrich	C4706
E. coli DNA Ligase	New England Biolabs	M0205
T4 DNA Ligase	New England Biolabs	M0202
BbsI-HF	New England Biolabs	R3539
BsaI-HFv2	New England Biolabs	R3733
Glucose Oxidase	Sigma-Aldrich	G2133
Catalase	Sigma-Aldrich	C9322
Human H1.0 linker histone	The Histone Source, Colorado State University	N/A
EDTA-free Protease Inhibitor Cocktail	Roche	04693159001
mPEG-Succinimidyl Valerate, MW 5,000	Laysan Bio	MPEG-SVA-5000
Gel Loading Dye, Purple with SDS	New England Biolabs	B7024
Xenopus H2A K119C	The Histone Source, Colorado State University	XH2A_K119C
Recombinant DNA		
Plasmid: pGEM-3Z/601	Addgene	26656
pGEM-3Z/601_4x601_40bplinker	This paper	N/A
Plasmid: Pet3a_H2A (Xenopus laevis)	Narlikar Lab	N/A
Plasmid: Pet3a_H2B (Xenopus laevis)	Narlikar Lab	N/A

Plasmid: Pet3a_H23 (Xenopus laevis)	Narlikar Lab	N/A
Plasmid: Pet3a_H4 (Xenopus laevis)	Narlikar Lab	N/A
Other		
Biotinyl Cap PE	Avanti	870277
streptavidin	New England Biolabs	N7021S
Superdex 200 Increase 10/300 GL	Cytiva	28990944
Model 491 Prep Cell	Bio-Rad	1702928
Density Gradient Fractionation System	Brandel	BR-186
QUANTIFOIL S7/2 300 mesh Au carbon grids	ELECTRON MICROSCOPY SCIENCES	https://www.emsdiasum.com/
Lacey LC-Au-FF 200 mesh Au carbon grids	ELECTRON MICROSCOPY SCIENCES	https://www.emsdiasum.com/
Oligonucleotides		
Forward primer for 100 bp 1st nucleosome of tetranucleosome DNA template: 5′ ATTCCAGTTGGTCTCTCAACAGCTTGCATGCCTGCAGGTCC 3′	IDT	N/A
Forward primer for 70 bp 1st nucleosome of tetranucleosome DNA template:5′ ATTCCAGTTGGTCTCTCAACATGACCAAGGAAAGCATGATTC 3′	IDT	N/A
Reverse primer for 1st nucleosome of tetranucleosome DNA template: 5′ AGTTGTACTGTAGTCTTCCAATACATGCACAGGATGTATATATCTG 3′	IDT	N/A
Forward primer for 2nd nucleosome of tetranucleosome	IDT	N/A

DNA template: 5′ ATTCCAGTTGAAGACTACAG TACCCTATACGCGGCCGCCC TGGAGAAT 3′		
Reverse primer for 2nd nucleosome of tetranucleosome DNA template: 5′ AGTTGTGTCATAGTCTTCCAA TACATGCACAGGATGTATAT ATCTG 3′	IDT	N/A
Forward primer for 3rd nucleosome of tetranucleosome DNA template: 5′ ATTCCAGTTGAAGACTATGA CACCCTATACGCGGCCGCCC TGGAGAAT 3′	IDT	N/A
Reverse primer for 3rd nucleosome of tetranucleosome DNA template: 5′ AGTTGTTGCGTAGTCTTCCAA TACATGCACAGGATGTATAT ATCTG 3′	IDT	N/A
Forward primer for 4th nucleosome of tetranucleosome DNA template: 5′ ATTCCAGTTGAAGACTACGC AACCCTATACGCGGCCGCCC TGGAGAAT 3′	IDT	N/A
Reverse primer for 4th nucleosome of tetranucleosome DNA template: 5′ AACTGGAATGGTCTCAAGGA CTCGGAACACTATCCGACTG GCAC 3′	IDT	N/A
Forward primer for 30 bp 5′ biotin ds oligo: 5′ /5Biosg/ TCTTCACACCGAGTTCATCCC TTATG 3′	IDT	N/A

Reverse primer for 30 bp 5′ biotin ds oligo: 5′ GTTGCATAAGGGATGAACTC GGTGTGAAGA 3′	IDT	N/A
Deposited Data		
Early irregular condensate, 1	This paper	EMD-24888
Early irregular condensate, 2	This paper	EMD-24901
Early irregular condensate, 3	This paper	EMD-24902
Early irregular condensate, 4	This paper	EMD-24903
Intermediate irregular condensate,1	This paper	EMD-24914
Intermediate irregular condensate, 2	This paper	EMD-24915
Intermediate irregular condensate, 3	This paper	EMD-24916
Intermediate irregular condensate, 4	This paper	EMD-24917
Late irregular condensate, 1	This paper	EMD-24918
Late irregular condensate, 2	This paper	EMD-24919
Late irregular condensate, 3	This paper	EMD-24920
Late irregular condensate, 4	This paper	EMD-24921
Small spherical condensate, 1	This paper	EMD-24923
Small spherical condensate, 2	This paper	EMD-24924
Small spherical condensate, 3	This paper	EMD-24925
Small spherical condensate, 4	This paper	EMD-24926
Gown spherical condensate, 1	This paper	EMD-24904
Gown spherical condensate, 2	This paper	EMD-24905
Gown spherical condensate, 3	This paper	EMD-24906
Gown spherical condensate, 4	This paper	EMD-24907
Gown spherical condensate, 5	This paper	EMD-24908
Gown spherical condensate with H1, 1	This paper	EMD-24909
Gown spherical condensate with H1, 2	This paper	EMD-24910
Gown spherical condensate with H1, 3	This paper	EMD-24911
Gown spherical condensate with H1, 4	This paper	EMD-24912

Gown spherical condensate with H1, 5	This paper	EMD-24913
Software and Algorithms		
IMOD	(Kremer et al., 1996)	https://bio3d.colorado.edu/imod/
Serial EM	(Schorb et al., 2019)	https://bio3d.colorado.edu/SerialEM/
MotionCor2	(Zheng et al., 2017)	https://msg.ucsf.edu/software
GCTF	(Zhang, 2016)	https://www2.mrc-lmb.cam.ac.uk/research/locally-developed-software/zhang-software/
UCSF Chimera	(Pettersen et al., 2004)	https://www.cgl.ucsf.edu/chimera/
EMAN2	(Tang et al., 2007b)	https://blake.bcm.edu/emanwiki/EMAN2
CARE	(Weigert et al., 2018)	https://csbdeep.bioimagecomputing.com/tools/care/
TomoCTF	(Fernandez et al., 2006)	https://sites.google.com/site/3demimageprocessing/tomoctf
DBSCAN	(Ester et al., 1996)	https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html
3D skeletonizing	(Lee et al., 1994)	https://scikit-image.org/docs/0.13.x/auto_examples/edges/plot_skeleton.html
IPET	(Zhang and Ren, 2012)	Available on request
LoTToR	(Zhai et al., 2020)	Available on request

Chapter 5
Conclusions

In this dissertation, entitled "Insights into the structural based mechanism of nucleosome array dynamic and their phase separation using individual-molecule cryo-electron tomography (ET)", I describe my contributions to push the boundaries of single-molecule imaging, in particular, in the field of individual particle electron tomography reconstruction. By determining the critical factor that limits the cryo-ET reconstruction capability and combining application of various denoising methods to improve the signal noise ratio of the raw tilt series, the individual particle ET can be applied to study biomolecules in the low-dose cryo-EM condition without the requirement of ensemble averaging (Chapter 2).

These improvements on the reconstruction workflow allowed us to study the flexible and heterogeneous chromatin structural organization. By using a "bottom up" strategy, we quantitatively characterized the inter-nucleosome array structural dynamics, which is required for understanding the "global" chromatin conformational changes induced by the local array thermodynamic equilibrium shift. With this method, we were able to determine that the DNA linker angle relative to its residing nucleosome core particle is the key factor during the chromatin transition process. Taking advantage of the IPET method, we were able to obtain direct measurement of the geometric statistical distributions of the fiber structures to rebuild the silicon model, which brought new insights into some of the arguments in the field, such as the existence of the "30-nm" fibers (Chapter 3).

Finally, I lay some of the groundwork for the in vitro study of nucleosome array phase separation, which describes a different picture to the historic hierarchical packing model. This work enabled us to for the first time observe the nanometer scale nucleosome condensate structure and its transition into different types of condensates. The proposed two-step transition mechanism from spinodal condensation to nucleation may have it counterparts in vivo, which need to be further validated in future experiments (Chapter 4).

I hope that these works will be useful to the scientific community in terms of either enhancing our understanding for the fundamental genome structural organization, or opening up new questions and inspiring new thoughts for future research.

Bibliography

[1] Cook, P.R. (1999). The organization of replication and transcription. Science *284*, 1790-1795.

[2] Nakayasu, H., and Berezney, R. (1989). Mapping replicational sites in the eucaryotic cell nucleus. The Journal of cell biology *108*, 1-11.

[3] Nakamura, H., Morita, T., and Sato, C. (1986). Structural organizations of replicon domains during DNA synthetic phase in the mammalian nucleus. Exp Cell Res *165*, 291-297.

[4] Boveri, T. (1909). Die Blastomerenkerne von Ascaris megalocephala und die Theorie der Chromosomenindividualitat. Archiv fur Zellforschung *3*, 181-268.

[5] Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). Science *183*, 330-332.

[6] Finch, J.T., Lutter, L.C., Rhodes, D., Brown, R.S., Rushton, B., Levitt, M., and Klug, A. (1977). Structure of nucleosome core particles of chromatin. Nature *269*, 29-36.

[7] Hewish, D.R., and Burgoyne, L.A. (1973). Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. Biochemical and biophysical research communications *52*, 504-510.

[8] Chari, S., Wilky, H., Govindan, J., and Amodeo, A.A. (2019). Histone concentration regulates the cell cycle and transcription in early development. Development *146*.

[9] Brambilla, F., Garcia-Manteiga, J.M., Monteleone, E., Hoelzen, L., Zocchi, A., Agresti, A., and Bianchi, M.E. (2020). Nucleosomes effectively shield DNA from radiation damage in living cells. Nucleic acids research *48*, 8993-9006.

[10] Wolffe, A. (1998). Chromatin: structure and function (Academic press).

[11] Eissenberg, J.C., James, T.C., Foster-Hartnett, D.M., Hartnett, T., Ngan, V., and Elgin, S.C. (1990). Mutation in a heterochromatin-specific chromosomal protein is

associated with suppression of position-effect variegation in Drosophila melanogaster. Proceedings of the National Academy of Sciences of the United States of America *87*, 9923-9927.

[12] Michaelis, C., Ciosk, R., and Nasmyth, K. (1997). Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. Cell *91*, 35-45.

[13] Quail, T., Golfier, S., Elsner, M., Ishihara, K., Murugesan, V., Renger, R., Jülicher, F., and Brugués, J. (2021). Force generation by protein–DNA co-condensation. Nature Physics *17*, 1007-1012.

[14] Renger, R., Morin, J.A., Lemaitre, R., Ruer-Gruss, M., Jülicher, F., Hermann, A., and Grill, S.W. (2022). Co-condensation of proteins with single-and double-stranded DNA. Proceedings of the National Academy of Sciences *119*, e2107871119.

[15] Rattner, J., and Lin, C. (1985). Radial loops and helical coils coexist in metaphase chromosomes. Cell *42*, 291-296.

[16] Finch, J.T., and Klug, A. (1976a). Solenoidal model for superstructure in chromatin. Proceedings of the National Academy of Sciences *73*, 1897-1901.

[17] Worcel, A., Strogatz, S., and Riley, D. (1981). Structure of chromatin and the linking number of DNA. Proceedings of the National Academy of Sciences *78*, 1461-1465.

[18] Widom, J., Finch, J.T., and Thomas, J.O. (1985). Higher-order structure of long repeat chromatin. The EMBO journal *4*, 3189-3194.

[19] Widom, J. (1989). Toward a unified model of chromatin folding. Annual review of biophysics and biophysical chemistry *18*, 365-395.

[20] Woodcock, C., Frado, L.-L., and Rattner, J. (1984a). The higher-order structure of chromatin: evidence for a helical ribbon arrangement. The Journal of cell biology *99*, 42-52.

[21] Bednar, J., Horowitz, R.A., Grigoryev, S.A., Carruthers, L.M., Hansen, J.C., Koster, A.J., and Woodcock, C.L. (1998). Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. Proceedings of the National Academy of Sciences *95*, 14173-14178.

[22] Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R.M., Zhu, P., and Li, G. (2014). Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. Science *344*, 376-380.

[23] Garcia-Saez, I., Menoni, H., Boopathi, R., Shukla, M.S., Soueidan, L., Noirclerc-Savoye, M., Le Roy, A., Skoufias, D.A., Bednar, J., Hamiche, A., *et al.* (2018). Structure of an H1-Bound 6-Nucleosome Array Reveals an Untwisted Two-Start Chromatin Fiber Conformation. Molecular cell *72*, 902-915 e907.

[24] Swedlow, J.R., and Hirano, T. (2003). The making of the mitotic chromosome: modern insights into classical questions. Molecular cell *11*, 557-569.

[25] Robinson, P.J., Fairall, L., Huynh, V.A., and Rhodes, D. (2006). EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. Proceedings of the National Academy of Sciences of the United States of America *103*, 6506-6511.

[26] Ou, H.D., Phan, S., Deerinck, T.J., Thor, A., Ellisman, M.H., and O'Shea, C.C. (2017). ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. Science *357*.

[27] Kireeva, N., Lakonishok, M., Kireev, I., Hirano, T., and Belmont, A.S. (2004). Visualization of early chromosome condensation: a hierarchical folding, axial glue model of chromosome structure. The Journal of cell biology *166*, 775-785.

[28] Dehghani, H., Dellaire, G., and Bazett-Jones, D.P. (2005). Organization of chromatin in the interphase mammalian cell. Micron *36*, 95-108.

[29] Daban, J.R. (2015a). Stacked thin layers of metaphase chromatin explain the geometry of chromosome rearrangements and banding. Sci Rep-Uk *5*.

[30] Daban, J.R. (2014a). The energy components of stacked chromatin layers explain the morphology, dimensions and mechanical properties of metaphase chromosomes. J R Soc Interface *11*.

[31] Finch, J.T., and Klug, A. (1976b). Solenoidal model for superstructure in chromatin. Proceedings of the National Academy of Sciences of the United States of America *73*, 1897-1901.

[32] Ohno, M., Priest, D.G., and Taniguchi, Y. (2018). Nucleosome-level 3D organization of the genome. Biochem Soc Trans *46*, 491-501.

[33] Woodcock, C.L., Frado, L.L., and Rattner, J.B. (1984b). The higher-order structure of chromatin: evidence for a helical ribbon arrangement. The Journal of cell biology *99*, 42-52.

[34] Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R.R., and Richmond, T.J. (2004). Nucleosome arrays reveal the two-start organization of the chromatin fiber. Science *306*, 1571-1573.

[35] Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997a). Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. Nature *389*, 251-260.

[36] Williams, S.P., Athey, B.D., Muglia, L.J., Schappe, R.S., Gough, A.H., and Langmore, J.P. (1986). Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length. Biophys J *49*, 233-248.

[37] Staynov, D. (1983). Possible nucleosome arrangements in the higher-order structure of chromatin. International Journal of Biological Macromolecules *5*, 3-9.

[38] Makarov, V., Dimitrov, S., Smirnov, V., and Pashev, I. (1985). A triple helix model for the structure of chromatin fiber. FEBS letters *181*, 357-361.

[39] Monneron, A., and Bernhard, W. (1969). Fine structural organization of the interphase nucleus in some mammalian cells. Journal of ultrastructure research *27*, 266-288.

[40] Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., *et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.

[41] Femino, A.M., Fay, F.S., Fogarty, K., and Singer, R.H. (1998). Visualization of single RNA transcripts in situ. Science *280*, 585-590.

[42] Beliveau, B.J., Boettiger, A.N., Avendaño, M.S., Jungmann, R., McCole, R.B.,

Joyce, E.F., Kim-Kiselak, C., Bantignies, F., Fonseka, C.Y., and Erceg, J. (2015). Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. Nature communications *6*, 1-13.

[43] Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. Science *348*, aaa6090.

[44] Bintu, B., Mateo, L.J., Su, J.-H., Sinnott-Armstrong, N.A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A.N., and Zhuang, X. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science *362*, eaau1783.

[45] Guan, J., Liu, H., Shi, X., Feng, S., and Huang, B. (2017). Tracking multiple genomic elements using correlative CRISPR imaging and sequential DNA FISH. Biophysical journal *112*, 1077-1084.

[46] Dixon, J.R., Gorkin, D.U., and Ren, B. (2016). Chromatin domains: the unit of chromosome organization. Molecular cell *62*, 668-680.

[47] Krumm, A., and Duan, Z. (2019). Understanding the 3D genome: emerging impacts on human disease. Paper presented at: Seminars in cell & developmental biology (Elsevier).

[48] Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., and Lander, E.S. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665-1680.

[49] Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., De Klein, A., Wessels, L., and De Laat, W. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948-951.

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

[51] Eltsov, M., Maclellan, K.M., Maeshima, K., Frangakis, A.S., and Dubochet, J. (2008). 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 of the United States of America *105*, 19732-19737.

[52] Maeshima, K., Hihara, S., and Eltsov, M. (2010a). Chromatin structure: does the 30-nm fibre exist in vivo? Curr Opin Cell Biol *22*, 291-297.

[53] Maeshima, K., Rogge, R., Tamura, S., Joti, Y., Hikima, T., Szerlong, H., Krause, C., Herman, J., Seidel, E., and DeLuca, J. (2016a). Nucleosomal arrays self-assemble into supramolecular globular structures lacking 30-nm fibers. The EMBO journal *35*, 1115-1132.

[54] Cai, S., Bock, D., Pilhofer, M., and Gan, L. (2018a). The in situ structures of mono-, di-, and trinucleosomes in human heterochromatin. Mol Biol Cell *29*, 2450-2457.

[55] Scheffer, M.P., Eltsov, M., and Frangakis, A.S. (2011a). Evidence for short-range helical order in the 30-nm chromatin fibers of erythrocyte nuclei. Proceedings of the National Academy of Sciences of the United States of America *108*, 16992-16997.

[56] Maeshima, K., Hihara, S., and Eltsov, M. (2010b). Chromatin structure: does the 30-nm fibre exist in vivo? Current opinion in cell biology *22*, 291-297.

[57] Eltsov, M., Sosnovski, S., Olins, A.L., and Olins, D.E. (2014). ELCS in ice: cryo-electron microscopy of nuclear envelope-limited chromatin sheets. Chromosoma *123*, 303-312.

[58] Polach, K.J., and Widom, J. (1995). Mechanism of Protein Access to Specific DNA-Sequences in Chromatin - a Dynamic Equilibrium-Model for Gene-Regulation. Journal of molecular biology *254*, 130-149.

[59] Winogradoff, D., and Aksimentiev, A. (2019). Molecular Mechanism of Spontaneous Nucleosome Unraveling. Journal of molecular biology *431*, 323-335.

[60] Anderson, J.D., Thastrom, A., and Widom, J. (2002a). Spontaneous access of proteins to buried nucleosomal DNA target sites occurs via a mechanism that is distinct from nucleosome translocation. Mol Cell Biol *22*, 7147-7157.

[61] Lai, W.K.M., and Pugh, B.F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. Nature reviews Molecular cell biology *18*, 548-562.

[62] Anderson, J.D., Lowary, P.T., and Widom, J. (2001). Effects of histone acetylation on the equilibrium accessibility of nucleosomal DNA target sites. Journal of molecular biology *307*, 977-985.

[63] Simon, M., North, J.A., Shimko, J., Forties, R., Ferdinand, M., Manohar, M., Zhang, M., Fishel, R., Ottesen, J.J., and Poirier, M.G. (2011). Histone fold modifications control nucleosome unwrapping and disassembly. Eur Biophys J Biophy *40*, 222-222.

[64] Ioshikhes, I.P., Albert, I., Zanton, S.J., and Pugh, B.F. (2006). Nucleosome positions predicted through comparative genomics. Nat Genet *38*, 1210-1215.

[65] Maeshima, K., Rogge, R., Tamura, S., Joti, Y., Hikima, T., Szerlong, H., Krause, C., Herman, J., Seidel, E., DeLuca, J., *et al.* (2016b). Nucleosomal arrays self-assemble into supramolecular globular structures lacking 30-nm fibers. EMBO J *35*, 1115-1132.

[66] Brendel, V., and Karlin, S. (1989). Association of charge clusters with functional domains of cellular transcription factors. Proceedings of the National Academy of Sciences of the United States of America *86*, 5698-5702.

[67] Maeshima, K., Ide, S., Hibino, K., and Sasai, M. (2016c). Liquid-like behavior of chromatin. Curr Opin Genet Dev *37*, 36-45.

[68] Zhang, Y., and Kutateladze, T.G. (2019). Liquid–liquid phase separation is an intrinsic physicochemical property of chromatin. Nature structural & molecular biology *26*, 1085-1086.

[69] Marnik, E.A., and Updike, D.L. (2019). Membraneless organelles: P granules in Caenorhabditis elegans. Traffic *20*, 373-379.

[70] Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., and Banani, S.F. (2012). Phase transitions in the assembly of multivalent signalling proteins. Nature *483*, 336-340.

[71] Iwasaki, W., Miya, Y., Horikoshi, N., Osakabe, A., Taguchi, H., Tachiwana, H., Shibata, T., Kagawa, W., and Kurumizaka, H. (2013). Contribution of histone N-terminal tails to the structure and stability of nucleosomes. FEBS open bio *3*, 363-369.

[72] Turner, A.L., Watson, M., Wilkins, O.G., Cato, L., Travers, A., Thomas, J.O., and Stott, K. (2018). Highly disordered histone H1− DNA model complexes and their

condensates. Proceedings of the National Academy of Sciences *115*, 11964-11969.

[73] Gibson, B.A., Doolittle, L.K., Schneider, M.W.G., Jensen, L.E., Gamarra, N., Henry, L., Gerlich, D.W., Redding, S., and Rosen, M.K. (2019). Organization of Chromatin by Intrinsic and Regulated Phase Separation. Cell *179*, 470-484 e421.

[74] Pacini, L., Dorantes-Gilardi, R., Vuillon, L., and Lesieur, C. (2021). Mapping Function from Dynamics: Future Challenges for Network-Based Models of Protein Structures. Front Mol Biosci *8*, 744646.

[75] Bustamante, C., Macosko, J.C., and Wuite, G.J. (2000). Grabbing the cat by the tail: manipulating molecules one by one. Nat Rev Mol Cell Bio *1*, 130-136.

[76] Ha, T., and Tinnefeld, P. (2012). Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging. Annual review of physical chemistry *63*, 595-617.

[77] Murata, K., and Wolf, M. (2018). Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. Biochim Biophys Acta Gen Subj *1862*, 324-334.

[78] De Rosier, D.J., and Klug, A. (1968). Reconstruction of three dimensional structures from electron micrographs. Nature *217*, 130-134.

[79] Hart, R.G. (1968). Electron microscopy of unstained biological material: the polytropic montage. Science *159*, 1464-1467.

[80] Jay, J.W., Bray, B., Qi, Y., Igbinigie, E., Wu, H., Li, J., and Ren, G. (2018). IgG Antibody 3D Structures and Dynamics. Antibodies (Basel) *7*.

[81] Yu, Y.D., Kuang, Y.L., Lei, D.S., Zhai, X.B., Zhang, M., Krauss, R.M., and Ren, G. (2016). Polyhedral 3D structure of human plasma very low density lipoproteins by individual particle cryo-electron tomography. Journal of lipid research *57*, 1879-1888.

[82] Lei, D.S., Marras, A.E., Liu, J.F., Huang, C.M., Zhou, L.F., Castro, C.E., Su, H.J., and Ren, G. (2018a). Three-dimensional structural dynamics of DNA origami Bennett linkages using individual-particle electron tomography. Nature communications *9*.

[83] Ren, G., Rudenko, G., Ludtke, S.J., Deisenhofer, J., Chiu, W., and Pownall, H.J. (2010). Model of human low-density lipoprotein and bound receptor based on CryoEM.

Proceedings of the National Academy of Sciences of the United States of America *107*, 1059-1064.

[84] Zhang, M., Charles, R., Tong, H., Zhang, L., Patel, M., Wang, F., Rames, M.J., Ren, A., Rye, K.A., Qiu, X., *et al.* (2015a). HDL surface lipids mediate CETP binding as revealed by electron microscopy and molecular dynamics simulation. Scientific reports *5*, 8741.

[85] Zhang, X., Zhang, L., Tong, H., Peng, B., Rames, M.J., Zhang, S., and Ren, G. (2015b). 3D Structural Fluctuation of IgG1 Antibody Revealed by Individual Particle Electron Tomography. Scientific reports *5*, 9803.

[86] Lu, Z., Reddy, M.V., Liu, J., Kalichava, A., Liu, J., Zhang, L., Chen, F., Wang, Y., Holthauzen, L.M., White, M.A., *et al.* (2016). Molecular Architecture of Contactin-associated Protein-like 2 (CNTNAP2) and Its Interaction with Contactin 2 (CNTN2). The Journal of biological chemistry *291*, 24133-24147.

[87] Li, W., Yang, Y., Yan, H., and Liu, Y. (2013). Three-input majority logic gate and multiple input logic circuit based on DNA strand displacement. Nano Lett *13*, 2980-2988.

[88] Wang, P., Rahman, M.A., Zhao, Z., Weiss, K., Zhang, C., Chen, Z., Hurwitz, S.J., Chen, Z.G., Shin, D.M., and Ke, Y. (2018). Visualization of the Cellular Uptake and Trafficking of DNA Origami Nanostructures in Cancer Cells. Journal of the American Chemical Society *140*, 2478-2484.

[89] Zheng, J., Constantinou, P.E., Micheel, C., Alivisatos, A.P., Kiehl, R.A., and Seeman, N.C. (2006). Two-dimensional nanoparticle arrays show the organizational power of robust DNA motifs. Nano letters *6*, 1502-1504.

[90] Alivisatos, A.P., Johnsson, K.P., Peng, X., Wilson, T.E., Loweth, C.J., Bruchez, M.P., Jr., and Schultz, P.G. (1996). Organization of 'nanocrystal molecules' using DNA. Nature *382*, 609-611.

[91] Elghanian, R., Storhoff, J.J., Mucic, R.C., Letsinger, R.L., and Mirkin, C.A. (1997). Selective colorimetric detection of polynucleotides based on the distance-dependent optical properties of gold nanoparticles. Science *277*, 1078-1081.

[92] Jones, M.R., Macfarlane, R.J., Lee, B., Zhang, J., Young, K.L., Senesi, A.J., and

Mirkin, C.A. (2010). DNA-nanoparticle superlattices formed from anisotropic building blocks. Nature materials *9*, 913-917.

[93] Cigler, P., Lytton-Jean, A.K., Anderson, D.G., Finn, M.G., and Park, S.Y. (2010). DNA-controlled assembly of a NaTl lattice structure from gold nanoparticles and protein nanoparticles. Nature materials *9*, 918-922.

[94] Zhang, Y., Lu, F., Yager, K.G., van der Lelie, D., and Gang, O. (2013). A general strategy for the DNA-mediated self-assembly of functional nanoparticles into heterogeneous systems. Nature nanotechnology *8*, 865-872.

[95] Rothemund, P.W. (2006). Folding DNA to create nanoscale shapes and patterns. Nature *440*, 297-302.

[96] Buckhout-White, S., Ancona, M., Oh, E., Deschamps, J.R., Stewart, M.H., Blanco-Canosa, J.B., Dawson, P.E., Goldman, E.R., and Medintz, I.L. (2012). Multimodal characterization of a linear DNA-based nanostructure. ACS Nano *6*, 1026-1043.

[97] Prado-Gotor, R., and Grueso, E. (2011). A kinetic study of the interaction of DNA with gold nanoparticles: mechanistic aspects of the interaction. Physical chemistry chemical physics : PCCP *13*, 1479-1489.

[98] Keskin, S., Besztejan, S., Kassier, G., Manz, S., Bucker, R., Riekeberg, S., Trieu, H.K., Rentmeister, A., and Miller, R.J. (2015). Visualization of multimerization and self-assembly of DNA-functionalized gold nanoparticles using in-liquid transmission electron microscopy. The journal of physical chemistry letters *6*, 4487-4492.

[99] Zhang, L., Lei, D.S., Smith, J.M., Zhang, M., Tong, H.M., Zhang, X., Lu, Z.Y., Liu, J.K., Alivisatos, A.P., and Ren, G. (2016a). Three-dimensional structural dynamics and fluctuations of DNA-nanogold conjugates by individual-particle electron tomography. Nature communications *7*.

[100] Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., and Aila, T. (2018). Noise2Noise: Learning image restoration without clean data. arXiv preprint arXiv:180304189.

[101] Buchholz, T.O., Krull, A., Shahidi, R., Pigino, G., Jekely, G., and Jug, F. (2019). Content-aware image restoration for electron microscopy. Methods Cell Biol *152*, 277-289.

[102] Weigert, M., Schmidt, U., Boothe, T., Muller, A., Dibrov, A., Jain, A., Wilhelm, B., Schmidt, D., Broaddus, C., Culley, S., *et al.* (2018). Content-aware image restoration: pushing the limits of fluorescence microscopy. Nature methods *15*, 1090-1097.

[103] Fernandez, J.J., Li, S., and Crowther, R.A. (2006). CTF determination and correction in electron cryotomography. Ultramicroscopy *106*, 587-596.

[104] Liu, J., Li, H., Zhang, L., Rames, M., Zhang, M., Yu, Y., Peng, B., Celis, C.D., Xu, A., Zou, Q., *et al.* (2016). Fully Mechanically Controlled Automated Electron Microscopic Tomography. Sci Rep *6*, 29231.

[105] Zhai, X., Lei, D., Zhang, M., Liu, J., Wu, H., Yu, Y., Zhang, L., and Ren, G. (2020). LoTToR: An Algorithm for Missing-Wedge Correction of the Low-Tilt Tomographic 3D Reconstruction of a Single-Molecule Structure. Sci Rep *10*, 10489.

[106] Maeshima, K., Imai, R., Tamura, S., and Nozaki, T. (2014a). Chromatin as dynamic 10-nm fibers. Chromosoma *123*, 225-237.

[107] Strickfaden, H., Tolsma, T.O., Sharma, A., Underhill, D.A., Hansen, J.C., and Hendzel, M.J. (2020). Condensed Chromatin Behaves like a Solid on the Mesoscale In Vitro and in Living Cells. Cell *183*, 1772-1784 e1713.

[108] Daban, J.R. (2015b). Stacked thin layers of metaphase chromatin explain the geometry of chromosome rearrangements and banding. Sci Rep *5*, 14891.

[109] Daban, J.R. (2014b). The energy components of stacked chromatin layers explain the morphology, dimensions and mechanical properties of metaphase chromosomes. J R Soc Interface *11*, 20131043.

[110] Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X., and Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. Nature *547*, 241-245.

[111] Farr, S.E., Woods, E.J., Joseph, J.A., Garaizar, A., and Collepardo-Guevara, R. (2021a). Nucleosome plasticity is a critical element of chromatin liquid-liquid phase separation and multivalent nucleosome interactions. Nature communications *12*, 2883.

[112] Hihara, S., Pack, C.G., Kaizu, K., Tani, T., Hanafusa, T., Nozaki, T., Takemoto, S.,

Yoshimi, T., Yokota, H., Imamoto, N., *et al.* (2012). Local nucleosome dynamics facilitate chromatin accessibility in living mammalian cells. Cell Rep *2*, 1645-1656.

[113] Sanulli, S., Trnka, M.J., Dharmarajan, V., Tibble, R.W., Pascal, B.D., Burlingame, A.L., Griffin, P.R., Gross, J.D., and Narlikar, G.J. (2019). HP1 reshapes nucleosome core to promote phase separation of heterochromatin. Nature *575*, 390-394.

[114] Ding, X., Lin, X., and Zhang, B. (2021). Stability and folding pathways of tetra-nucleosome from six-dimensional free energy surface. Nature communications *12*, 1091.

[115] Schalch, T., Duda, S., Sargent, D.F., and Richmond, T.J. (2005a). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature *436*, 138-141.

[116] Zhang, M., Diaz-Celis, C.D., Onoa, B., Canari-Chumpitaz, C., Requejo, K.I., Liu, J., Vien, M., Nogales, E., Ren, G., and Bustamante, C. (2021). Molecular Organization of the Early Stages of Nucleosome Phase Separation Visualized by Cryo-Electron Tomography. bioRxiv.

[117] Marini, M., Falqui, A., Moretti, M., Limongi, T., Allione, M., Genovese, A., Lopatin, S., Tirinato, L., Das, G., Torre, B., *et al.* (2015). The structure of DNA by direct imaging. Sci Adv *1*, e1500734.

[118] Irobalieva, R.N., Fogg, J.M., Catanese, D.J., Jr., Sutthibutpong, T., Chen, M., Barker, A.K., Ludtke, S.J., Harris, S.A., Schmid, M.F., Chiu, W., *et al.* (2015). Structural diversity of supercoiled DNA. Nature communications *6*, 8440.

[119] Beel, A.J., Azubel, M., Mattei, P.J., and Kornberg, R.D. (2021). Structure of mitotic chromosomes. Molecular cell *81*, 4369-4376 e4363.

[120] Zhang, L., Song, J., Newhouse, Y., Zhang, S., Weisgraber, K.H., and Ren, G. (2010). An optimized negative-staining protocol of electron microscopy for apoE4 POPC lipoprotein. J Lipid Res *51*, 1228-1236.

[121] Rames, M., Yu, Y., and Ren, G. (2014a). Optimized negative staining: a high-throughput protocol for examining small and asymmetric protein structure by electron microscopy. Journal of visualized experiments : JoVE, e51087.

[122] Zhang, L., Song, J., Cavigiolio, G., Ishida, B.Y., Zhang, S., Kane, J.P., Weisgraber, K.H., Oda, M.N., Rye, K.A., Pownall, H.J., *et al.* (2011). Morphology and structure of

lipoproteins revealed by an optimized negative-staining protocol of electron microscopy. J Lipid Res *52*, 175-184.

[123] Zhang, L., and Ren, G. (2012). IPET and FETR: experimental approach for studying molecular structure dynamics by cryo-electron tomography of a single-molecule structure. PloS one *7*, e30249.

[124] Phillips, J.C., Hardy, D.J., Maia, J.D.C., Stone, J.E., Ribeiro, J.V., Bernardi, R.C., Buch, R., Fiorin, G., Henin, J., Jiang, W., *et al.* (2020). Scalable molecular dynamics on CPU and GPU architectures with NAMD. J Chem Phys *153*, 044130.

[125] Liao, H.Y., and Frank, J. (2010). Definition and estimation of resolution in single-particle reconstructions. Structure *18*, 768-775.

[126] Rosenthal, P.B., and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. Journal of molecular biology *333*, 721-745.

[127] Bilokapic, S., Strauss, M., and Halic, M. (2018a). Histone octamer rearranges to adapt to DNA unwrapping. Nat Struct Mol Biol *25*, 101-108.

[128] Ngo, T.T., Zhang, Q., Zhou, R., Yodh, J.G., and Ha, T. (2015). Asymmetric unwrapping of nucleosomes under tension directed by DNA local flexibility. Cell *160*, 1135-1144.

[129] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research *12*, 2825-2830.

[130] Grigoryev, S.A., Bednar, J., and Woodcock, C.L. (1999). MENT, a heterochromatin protein that mediates higher order chromatin folding, is a new serpin family member. The Journal of biological chemistry *274*, 5626-5636.

[131] Ng, C.T., and Gan, L. (2020). Investigating eukaryotic cells with cryo-ET. Molecular biology of the cell *31*, 87-100.

[132] Schalch, T., Duda, S., Sargent, D.F., and Richmond, T.J. (2005b). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature *436*, 138-141.

[133] Yamada, K., Frouws, T.D., Angst, B., Fitzgerald, D.J., DeLuca, C., Schimmele, K., Sargent, D.F., and Richmond, T.J. (2011). Structure and mechanism of the chromatin remodelling factor ISW1a. Nature *472*, 448-453.

[134] Bhardwaj, S.K., Hailu, S.G., Olufemi, L., Brahma, S., Kundu, S., Hota, S.K., Persinger, J., and Bartholomew, B. (2020). Dinucleosome specificity and allosteric switch of the ISW1a ATP-dependent chromatin remodeler in transcription regulation. Nature communications *11*, 5913.

[135] Gebala, M., Johnson, S.L., Narlikar, G.J., and Herschlag, D. (2019). Ion counting demonstrates a high electrostatic field generated by the nucleosome. eLife *8*.

[136] Manning, G.S. (2003). Is a small number of charge neutralizations sufficient to bend nucleosome core DNA onto its superhelical ramp? Journal of the American Chemical Society *125*, 15087-15092.

[137] Blank, T.A., and Becker, P.B. (1995). Electrostatic mechanism of nucleosome spacing. Journal of molecular biology *252*, 305-313.

[138] Li, M., and Wang, M.D. (2012). Unzipping single DNA molecules to study nucleosome structure and dynamics. Methods in enzymology *513*, 29-58.

[139] Anderson, J.D., Thastrom, A., and Widom, J. (2002b). Spontaneous access of proteins to buried nucleosomal DNA target sites occurs via a mechanism that is distinct from nucleosome translocation. Molecular and cellular biology *22*, 7147-7157.

[140] Anderson, J.D., and Widom, J. (2000). Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. Journal of molecular biology *296*, 979-987.

[141] Bednar, J., Garcia-Saez, I., Boopathi, R., Cutter, A.R., Papai, G., Reymer, A., Syed, S.H., Lone, I.N., Tonchev, O., Crucifix, C., *et al.* (2017). Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. Molecular cell *66*, 384-397 e388.

[142] Konig, P., Braunfeld, M.B., Sedat, J.W., and Agard, D.A. (2007). The three-dimensional structure of in vitro reconstituted Xenopus laevis chromosomes by EM tomography. Chromosoma *116*, 349-372.

[143] McDowall, A.W., Smith, J.M., and Dubochet, J. (1986). Cryo-electron microscopy

of vitrified chromosomes in situ. The EMBO journal *5*, 1395-1402.

[144] Cai, S., Chen, C., Tan, Z.Y., Huang, Y., Shi, J., and Gan, L. (2018b). Cryo-ET reveals the macromolecular reorganization of S. pombe mitotic chromosomes in vivo. Proceedings of the National Academy of Sciences of the United States of America *115*, 10977-10982.

[145] Dyer, P.N., Edayathumangalam, R.S., White, C.L., Bao, Y., Chakravarthy, S., Muthurajan, U.M., and Luger, K. (2004). Reconstitution of nucleosome core particles from recombinant histones and DNA. Methods in enzymology *375*, 23-44.

[146] Luger, K., Rechsteiner, T.J., and Richmond, T.J. (1999). Preparation of nucleosome core particle from recombinant histones. Methods in enzymology *304*, 3-19.

[147] Schorb, M., Haberbosch, I., Hagen, W.J.H., Schwab, Y., and Mastronarde, D.N. (2019). Software tools for automated transmission electron microscopy. Nature methods *16*, 471-477.

[148] Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods *14*, 331-332.

[149] Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. Journal of structural biology *142*, 334-347.

[150] Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J Struct Biol *116*, 190-199.

[151] Zhang, K. (2016). Gctf: Real-time CTF determination and correction. Journal of structural biology *193*, 1-12.

[152] Kremer, J.R., Mastronarde, D.N., and McIntosh, J.R. (1996). Computer visualization of three-dimensional image data using IMOD. Journal of structural biology *116*, 71-76.

[153] Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. J Struct Biol *128*, 82-97.

[154] Yan, R., Venkatakrishnan, S.V., Liu, J., Bouman, C.A., and Jiang, W. (2019). MBIR: A cryo-ET 3D reconstruction method that effectively minimizes missing wedge artifacts and restores missing information. J Struct Biol *206*, 183-192.

[155] Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

[156] Hornus, S., Levy, B., Lariviere, D., and Fourmentin, E. (2013). Easy DNA modeling and more with GraphiteLifeExplorer. PloS one *8*, e53609.

[157] Lowary, P.T., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. Journal of molecular biology *276*, 19-42.

[158] Rychkov, G.N., Ilatovskiy, A.V., Nazarov, I.B., Shvetsov, A.V., Lebedev, D.V., Konev, A.Y., Isaev-Ivanov, V.V., and Onufriev, A.V. (2017). Partially Assembled Nucleosome Structures at Atomic Detail. Biophys J *112*, 460-472.

[159] Silverman, B.W. (1978). Weak and Strong Uniform Consistency of the Kernel Estimate of a Density and its Derivatives. . The Annals of Statistics *6*, , 177–184.

[160] Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J Mol Graph *14*, 33-38, 27-38.

[161] Levitt, M. (1978). How many base-pairs per turn does DNA have in solution and in chromatin? Some theoretical calculations. Proceedings of the National Academy of Sciences of the United States of America *75*, 640-644.

[162] Bar-Joseph, Z., Gifford, D.K., and Jaakkola, T.S. (2001). Fast optimal leaf ordering for hierarchical clustering. Bioinformatics *17 Suppl 1*, S22-29.

[163] Michaud-Agrawal, N., Denning, E.J., Woolf, T.B., and Beckstein, O. (2011). MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J Comput Chem *32*, 2319-2327.

[164] Cifra, P. (2004). Differences and limits in estimates of persistence length for semi-flexible macromolecules. Polymer *45*, 5995-6002.

[165] Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997b). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

[166] Bowman, G.D., and Poirier, M.G. (2015). Post-translational modifications of histones that influence nucleosome dynamics. Chem Rev *115*, 2274-2295.

[167] Widom, J., and Klug, A. (1985). Structure of the 300A chromatin filament: X-ray diffraction from oriented samples. Cell *43*, 207-213.

[168] Belmont, A.S., and Bruce, K. (1994). Visualization of G1 chromosomes: a folded, twisted, supercoiled chromonema model of interphase chromatid structure. The Journal of cell biology *127*, 287-302.

[169] Maeshima, K., Imai, R., Tamura, S., and Nozaki, T. (2014b). Chromatin as dynamic 10-nm fibers. Chromosoma *123*, 225-237.

[170] Maeshima, K., Imai, R., Hikima, T., and Joti, Y. (2014c). Chromatin structure revealed by X-ray scattering analysis and computational modeling. Methods *70*, 154-161.

[171] Razin, S.V., and Gavrilov, A.A. (2014). Chromatin without the 30-nm fiber: constrained disorder instead of hierarchical folding. Epigenetics *9*, 653-657.

[172] Eltsov, M., Grewe, D., Lemercier, N., Frangakis, A., Livolant, F., and Leforestier, A. (2018). Nucleosome conformational variability in solution and in interphase nuclei evidenced by cryo-electron microscopy of vitreous sections. Nucleic acids research *46*, 9189-9200.

[173] Nishino, Y., Eltsov, M., Joti, Y., Ito, K., Takata, H., Takahashi, Y., Hihara, S., Frangakis, A.S., Imamoto, N., Ishikawa, T., *et al.* (2012). Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. The EMBO journal *31*, 1644-1653.

[174] Hubstenberger, A., Courel, M., Benard, M., Souquere, S., Ernoult-Lange, M., Chouaib, R., Yi, Z., Morlot, J.B., Munier, A., Fradet, M., *et al.* (2017). P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. Molecular cell *68*, 144-157 e145.

[175] Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van Den Bosch, L., *et al.* (2018). Protein Phase Separation: A New Phase in Cell Biology. Trends Cell Biol *28*, 420-435.

[176] Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: organizers of cellular biochemistry. Nat Rev Mol Cell Biol *18*, 285-298.

[177] Sanulli, S., and G, J.N. (2020). Liquid-like interactions in heterochromatin: Implications for mechanism and regulation. Curr Opin Cell Biol *64*, 90-96.

[178] Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Julicher, F., and Hyman, A.A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science *324*, 1729-1732.

[179] Erdel, F., and Rippe, K. (2018). Formation of Chromatin Subcompartments by Phase Separation. Biophysical journal *114*, 2262-2270.

[180] Hyman, A.A., Weber, C.A., and Julicher, F. (2014). Liquid-liquid phase separation in biology. Annu Rev Cell Dev Biol *30*, 39-58.

[181] Palikyras, S., and Papantonis, A. (2019). Modes of phase separation affecting chromatin regulation. Open biology *9*, 190167.

[182] Shakya, A., Park, S., Rana, N., and King, J.T. (2020). Liquid-Liquid Phase Separation of Histone Proteins in Cells: Role in Chromatin Organization. Biophysical journal *118*, 753-764.

[183] Shin, Y., Chang, Y.C., Lee, D.S.W., Berry, J., Sanders, D.W., Ronceray, P., Wingreen, N.S., Haataja, M., and Brangwynne, C.P. (2018). Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. Cell *175*, 1481-1491 e1413.

[184] Wright, R.H.G., Le Dily, F., and Beato, M. (2019). ATP, Mg(2+), Nuclear Phase Separation, and Genome Accessibility. Trends Biochem Sci *44*, 565-574.

[185] Bancaud, A., Huet, S., Daigle, N., Mozziconacci, J., Beaudouin, J., and Ellenberg, J. (2009). Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin. The EMBO journal *28*, 3785-3798.

[186] Imai, R., Nozaki, T., Tani, T., Kaizu, K., Hibino, K., Ide, S., Tamura, S., Takahashi, K., Shribak, M., and Maeshima, K. (2017). Density imaging of heterochromatin in live cells using orientation-independent-DIC microscopy. Molecular biology of the cell *28*, 3349-3359.

[187] Ercius, P., Alaidi, O., Rames, M.J., and Ren, G. (2015). Electron Tomography: A Three-Dimensional Analytic Tool for Hard and Soft Materials Research. Advanced materials *27*, 5638-5663.

[188] Chen, M., Dai, W., Sun, S.Y., Jonasch, D., He, C.Y., Schmid, M.F., Chiu, W., and Ludtke, S.J. (2017). Convolutional neural networks for automated annotation of cellular cryo-electron tomograms. Nature methods *14*, 983-985.

[189] Narayan, K., and Subramaniam, S. (2015). Focused ion beams in biology. Nature methods *12*, 1021-1031.

[190] Lei, D., Marras, A.E., Liu, J., Huang, C.M., Zhou, L., Castro, C.E., Su, H.J., and Ren, G. (2018b). Three-dimensional structural dynamics of DNA origami Bennett linkages using individual-particle electron tomography. Nat Commun *9*, 592.

[191] Zhang, L., Lei, D., Smith, J.M., Zhang, M., Tong, H., Zhang, X., Lu, Z., Liu, J., Alivisatos, A.P., and Ren, G. (2016b). Three-dimensional structural dynamics and fluctuations of DNA-nanogold conjugates by individual-particle electron tomography. Nat Commun *7*, 11083.

[192] Jordan, M.A., Diener, D.R., Stepanek, L., and Pigino, G. (2018). The cryo-EM structure of intraflagellar transport trains reveals how dynein is inactivated to ensure unidirectional anterograde movement in cilia. Nature cell biology *20*, 1250-1255.

[193] Clerc, D.G., and Cleary, D.A. (1995). Spinodal Decomposition as an Interesting Example of the Application of Several Thermodynamic Principles. Journal of chemical education *72*, 112.

[194] Cahn, J.W., and Hilliard, J.E. (1958). Free energy of a nonuniform system. I. Interfacial free energy. The Journal of chemical physics *28*, 258-267.

[195] Cahn, J.W., and Hilliard, J.E. (1959). Free energy of a nonuniform system. III. Nucleation in a two-component incompressible fluid. The Journal of chemical physics *31*, 688-699.

[196] Schmelzer, J.W., Abyzov, A.S., and Möller, J. (2004). Nucleation versus spinodal decomposition in phase formation processes in multicomponent solutions. The Journal of chemical physics *121*, 6900-6917.

[197] Erdemir, D., Lee, A.Y., and Myerson, A.S. (2009). Nucleation of crystals from solution: classical and two-step models. Accounts of chemical research *42*, 621-629.

[198] Vekilov, P.G. (2010). The two-step mechanism of nucleation of crystals in solution. Nanoscale *2*, 2346-2357.

[199] Loh, N.D., Sen, S., Bosman, M., Tan, S.F., Zhong, J., Nijhuis, C.A., Král, P., Matsudaira, P., and Mirsaidov, U. (2017a). Multistep nucleation of nanocrystals in aqueous solution. Nature chemistry *9*, 77-82.

[200] Loh, N.D., Sen, S., Bosman, M., Tan, S.F., Zhong, J., Nijhuis, C.A., Kral, P., Matsudaira, P., and Mirsaidov, U. (2017b). Multistep nucleation of nanocrystals in aqueous solution. Nature chemistry *9*, 77-82.

[201] Ji, X., Song, X., Li, J., Bai, Y., Yang, W., and Peng, X. (2007). Size control of gold nanocrystals in citrate reduction: the third role of citrate. Journal of the American Chemical Society *129*, 13939-13948.

[202] Pong, B.-K., Elim, H.I., Chong, J.-X., Ji, W., Trout, B.L., and Lee, J.-Y. (2007). New insights on the nanoparticle growth mechanism in the citrate reduction of gold (III) salt: formation of the Au nanowire intermediate and its nonlinear optical properties. The Journal of Physical Chemistry C *111*, 6281-6287.

[203] Alberti, S., Gladfelter, A., and Mittag, T. (2019). Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. Cell *176*, 419-434.

[204] Shin, Y., and Brangwynne, C.P. (2017). Liquid phase condensation in cell physiology and disease. Science *357*.

[205] Karthika, S., Radhakrishnan, T., and Kalaichelvi, P. (2016). A review of classical and nonclassical nucleation theories. Crystal Growth & Design *16*, 6663-6681.

[206] Kim, J., Lee, S., Choi, Y., Lee, S.-M., and Jeong, D. (2016). Basic principles and

practical applications of the Cahn–Hilliard equation. Mathematical Problems in Engineering *2016*.

[207] Thanh, N.T., Maclean, N., and Mahiddine, S. (2014). Mechanisms of nucleation and growth of nanoparticles in solution. Chemical reviews *114*, 7610-7630.

[208] Yau, S.-T., and Vekilov, P.G. (2001). Direct observation of nucleus structure and nucleation pathways in apoferritin crystallization. Journal of the American Chemical Society *123*, 1080-1089.

[209] Lee, D., Huh, J.-Y., Jeong, D., Shin, J., Yun, A., and Kim, J. (2014). Physical, mathematical, and numerical derivations of the Cahn–Hilliard equation. Computational Materials Science *81*, 216-225.

[210] Rames, M., Yu, Y.D., and Ren, G. (2014b). Optimized Negative Staining: a High-throughput Protocol for Examining Small and Asymmetric Protein Structure by Electron Microscopy. Jove-J Vis Exp.

[211] Han, B.G., Watson, Z., Kang, H., Pulk, A., Downing, K.H., Cate, J., and Glaeser, R.M. (2016). Long shelf-life streptavidin support-films suitable for electron microscopy of biological macromolecules. Journal of structural biology *195*, 238-244.

[212] Chen, M., Bell, J.M., Shi, X., Sun, S.Y., Wang, Z., and Ludtke, S.J. (2019). A complete data processing workflow for cryo-ET and subtomogram averaging. Nature methods *16*, 1161-1168.

[213] Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Paper presented at: kdd.

[214] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. J Educ Psychol *24*, 417-441.

[215] Emo, J., Pareige, C., Saillet, S., Domain, C., and Pareige, P. (2014). Kinetics of secondary phase precipitation during spinodal decomposition in duplex stainless steels: A kinetic Monte Carlo model–Comparison with atom probe tomography experiments. Journal of Nuclear Materials *451*, 361-365.

[216] Sear, R.P. (2014). Quantitative studies of crystal nucleation at constant supersaturation: experimental data and models. Crystengcomm *16*, 6506-6522.

[217] Zhou, B.R., Feng, H., Ghirlando, R., Li, S., Schwieters, C.D., and Bai, Y. (2016). A Small Number of Residues Can Determine if Linker Histones Are Bound On or Off Dyad in the Chromatosome. Journal of molecular biology *428*, 3948-3959.

[218] Andreyeva, E.N., Bernardo, T.J., Kolesnikova, T.D., Lu, X., Yarinich, L.A., Bartholdy, B.A., Guo, X., Posukh, O.V., Healton, S., Willcockson, M.A., *et al.* (2017). Regulatory functions and chromatin loading dynamics of linker histone H1 during endoreplication in Drosophila. Genes Dev *31*, 603-616.

[219] Dong, X., Bera, S., Qiao, Q., Tang, Y., Lao, Z., Luo, Y., Gazit, E., and Wei, G. (2021). Liquid-Liquid Phase Separation of Tau Protein Is Encoded at the Monomeric Level. The journal of physical chemistry letters *12*, 2576-2586.

[220] Bilokapic, S., Strauss, M., and Halic, M. (2018b). Cryo-EM of nucleosome core particle interactions in trans. Scientific reports *8*, 7046.

[221] Farr, S.E., Woods, E.J., Joseph, J.A., Garaizar, A., and Collepardo-Guevara, R. (2021b). Nucleosome plasticity is a critical element of chromatin liquid–liquid phase separation and multivalent nucleosome interactions. Nature communications *12*, 1-17.

[222] Datt, C., Thampi, S.P., and Govindarajan, R. (2015). Morphological evolution of domains in spinodal decomposition. Physical review E, Statistical, nonlinear, and soft matter physics *91*, 010101.

[223] Ostwald, W. (1897). Studies on the formation and transformation of solid bodies. Zeitschrift für Physikalische Chemie *22*, 289–330.

[224] Ulianov, S.V., Velichko, A.K., Magnitov, M.D., Luzhin, A.V., Golov, A.K., Ovsyannikova, N., Kireev, II, Gavrikov, A.S., Mishin, A.S., Garaev, A.K., *et al.* (2021). Suppression of liquid-liquid phase separation by 1,6-hexanediol partially compromises the 3D genome organization in living cells. Nucleic acids research.

[225] Henikoff, S., Henikoff, J.G., Kaya-Okur, H.S., and Ahmad, K. (2020). Efficient chromatin accessibility mapping in situ by nucleosome-tethered tagmentation. eLife *9*.

[226] Kwiatkowski da Silva, A., Ponge, D., Peng, Z., Inden, G., Lu, Y., Breen, A., Gault, B., and Raabe, D. (2018). Phase nucleation through confined spinodal fluctuations at crystal defects evidenced in Fe-Mn alloys. Nature communications *9*, 1137.

[227] Houben, L., Weissman, H., Wolf, S.G., and Rybtchinski, B. (2020). A mechanism of ferritin crystallization revealed by cryo-STEM tomography. Nature *579*, 540-543.

[228] Cai, S.J., Song, Y.J., Chen, C., Shi, J., and Gan, L. (2018c). Natural chromatin is heterogeneous and self-associates in vitro. Molecular biology of the cell *29*, 1652-1663.

[229] Scheffer, M.P., Eltsov, M., and Frangakis, A.S. (2011b). Evidence for short-range helical order in the 30-nm chromatin fibers of erythrocyte nuclei. Proceedings of the National Academy of Sciences of the United States of America *108*, 16992-16997.

[230] Robinson, P.J., and Rhodes, D. (2006). Structure of the '30 nm' chromatin fibre: a key role for the linker histone. Current opinion in structural biology *16*, 336-343.

[231] Grigoryev, S.A., Arya, G., Correll, S., Woodcock, C.L., and Schlick, T. (2009). Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. Proceedings of the National Academy of Sciences of the United States of America *106*, 13317-13322.

[232] Hansen, J.C., Connolly, M., McDonald, C.J., Pan, A., Pryamkova, A., Ray, K., Seidel, E., Tamura, S., Rogge, R., and Maeshima, K. (2018). The 10-nm chromatin fiber and its relationship to interphase chromosome organization. Biochem Soc Trans *46*, 67-76.

[233] Strohkendl, I., Saifuddin, F.A., Gibson, B.A., Rosen, M.K., Russell, R., and Finkelstein, I.J. (2021). Inhibition of CRISPR-Cas12a DNA targeting by nucleosomes and chromatin. Science advances *7*.

[234] Hatta, M., and Cirillo, L.A. (2007). Chromatin opening and stable perturbation of core histone:DNA contacts by FoxO1. The Journal of biological chemistry *282*, 35583-35593.

[235] Wittmeyer, J., Saha, A., and Cairns, B. (2004). DNA translocation and nucleosome remodeling assays by the RSC chromatin remodeling complex. Methods in enzymology *377*, 322-343.

[236] Zhou, C.Y., and Narlikar, G.J. (2016). Analysis of Nucleosome Sliding by ATP-Dependent Chromatin Remodeling Enzymes. Methods in enzymology *573*, 119-135.

[237] Chandradoss, S.D., Haagsma, A.C., Lee, Y.K., Hwang, J.H., Nam, J.M., and Joo, C. (2014). Surface passivation for single-molecule protein studies. Journal of visualized experiments : JoVE.

[238] Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007a). EMAN2: an extensible image processing suite for electron microscopy. Journal of structural biology *157*, 38-46.

[239] Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin philosophical magazine and journal of science *2*, 559-572.

[240] Lee, T.-C., Kashyap, R.L., and Chu, C.-N. (1994). Building skeleton models via 3-D medial surface axis thinning algorithms. CVGIP: Graphical Models and Image Processing *56*, 462-478.

[241] Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007b). EMAN2: An extensible image processing suite for electron microscopy. Journal of structural biology *157*, 38-46.