The Pennsylvania State University

The Graduate School

Eberly College of Science

STRUCTURAL INVESTIGATION OF CENP-A CONTAINING NUCLEOSOMES UTILIZING SINGLE-MOLECULE FÖRSTER RESONANCE ENERGY TRANSFER

A Thesis in

Chemistry

by

Michael A. Sennett

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

December 2014

The Thesis of Michael A. Sennett was reviewed and approved* by the following:

Tae-Hee Lee
Associate Professor of Chemistry
Thesis Advisor

Philip Bevilacqua
Professor of Chemistry

Scott Showalter
Associate Professor of Chemistry

Ken Feldman
Professor of Chemistry-Graduate Program Chair

*Signatures are on file in the Graduate School

Abstract

Interest in how a functional kinetochore assembles at the centromere has grown in recent years as a result of our increased understanding of the chromosome. Centromere Protein A (CenP-A) is a histone H3 variant and is believed to play an essential role in defining the location of the centromere, rather than the DNA sequence.[1] Based on recent crystal structures and models of nucleosomes containing Cenp-A, it has been hypothesized that the structures of Cenp-A nucleosomes is different from that of canonical nucleosomes.[2,3] Using single-molecule Förster resonance energy transfer (smFRET), I tested the hypothesis in vitro using physiologically relevant salt conditions. smFRET measurements reveal that the structure of fluorescently labeled DNA in complex with CenP-A containing centromere nucleosomes is different from the canonical nucleosomes. This indicates a specific role for CenP-A variants, suggesting a basis for the mechanism of centromere identification and kinetochore assembly.

Table of Contents

List of Table

Compilation of FRET E and distances for each nucleosome. Y1 and Y2 indicates Yokoyama nucleosomes assembled with a fluorophore between [-33:-34] and [-43:-44] respectively. B1 and B2 indicates the same except the Bunick sequence was used. The distances and changes in distances calculated are in nm.

List of Figures

Chapter 1: Introduction

1.1 Canonical and Centromere Nucleosome Structure:

Cell division is an essential process for the development and the growth of an organism. After DNA replication, an array of DNA-protein complexes are assembled and condense to form sister chromatids. Each sister chromatid has a copy of identical genes and they are joined together at the centromere forming the chromosome.[4] The centromere is a region of the chromosome that directs the formation of a protein complex known as the kinetochore. The kinetochore links the centromere to microtubules such that when the chromatids are pulled apart each daughter cell will receive a copy of each gene. It is not yet known what marks the centromere region or how the proteins that assemble the kinetochore recognize the centromere region.[5] The centromere, as well as the entire chromosome, is composed of a fundamental repeating unit known as the nucleosome core particle (NCP).[6] However, those present at the centromere are slightly different from canonical NCPs present throughout the chromosome.[2,3] NCPs in the centromere contain two copies of the histone variant CenP-A in place of H3 proteins. It is hypothesized that the nucleosome facilitates the kinetochore proteins recognition of the centromere.[2]

Canonical NCPs consist of a histone octamer wrapped by ~147 base pairs (bp) of double stranded deoxryribonucleic acid (dsDNA), resulting in a complex that has a flattened cylindrical shape.[6] The histone octamer contains 2 copies of four different histone proteins: H2A, H2B, H3 and H4. Each of the histone proteins has a structurally conserved α1-L1-α2-L2-α3 helix-turn-helix motif. These motives are the bases of the association of two histones in H2A-H2B, H3-H4,

H3-H3, and H2B-H4. H3 with H4, as well as H2A with H2B, associate together to form heterodimers by what is known as a handshake motif, in which the palms are the second alpha helix. Two (H3-H4) heterodimers complex together to form a four helix bundle resulting in a $(H3\text{-}H4)_2$ tetramer.[6] The four-helix bundle forms at the interface between H3-H3 (the four blue helices at the top central region of the core in Figure 1A). Another four helix bundle forms at the interface of the two pairs of H2B-H4 (the two green and two red helices at the interface between the two block C shapes) and are essential in the formation of the octameric histone core in the nucleosome.

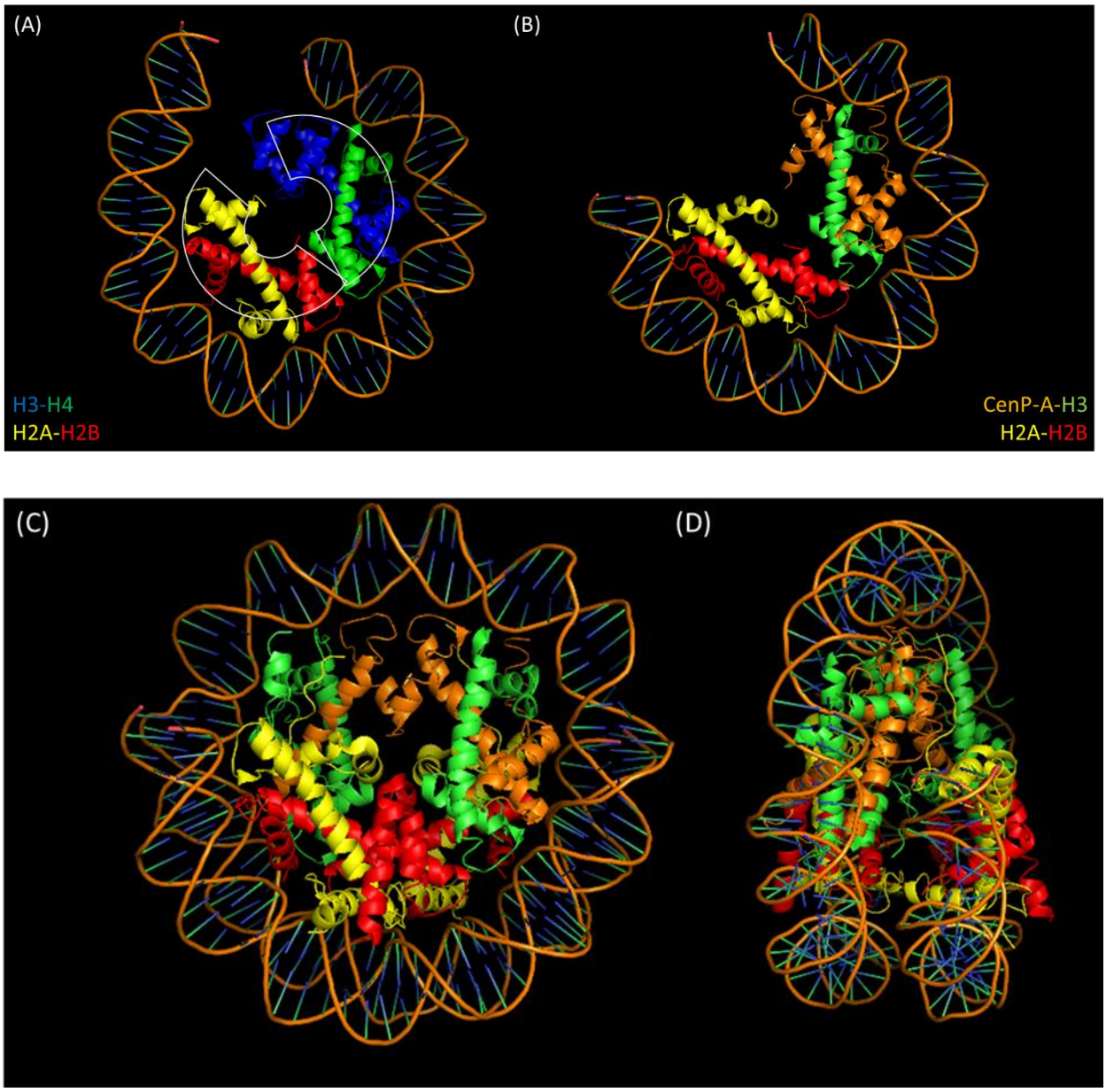

Figure 1: Crystallographic structures of the canonical (left), PDB-1AOI, and centromere nucleosomes (right), PDB-3AN2. Two block C shapes depict the two heterodimers, respectively, H2A-H2B (yellow and red) and H3-H4 (green and blue). The termini of the DNA in (B) were not present in this model possibly due to increased flexibility. In both cases the second set of four histones were removed as well as half the DNA for clarity. (C) and (D) are crystal structures of

the whole centromere nucleosome, which includes CenP-A, from the front view and side view respectively.

One set of the four histones occupies one half of the dsDNA with electrostatic interactions between positively charged histone residues and negatively charged phosphates in the minor groove of the DNA backbone.[6] Nucleosomes assembled at the centromere have the H3 subunit replaced by a structurally similar variant known as CenP-A (orange in Figure 1B) yet all interactions remain the same.[2,3]

The histone octamer structure in the nucleosome crystal models is centered on a single base pair in the palindromic dsDNA that is known as the dyad.[6] It is around this point that a pseudo two-fold axis of symmetry can be drawn. From this dyad, every tenth base pair to each end of the 147 bp dsDNA is a point of contact with histones. The DNA sequence used in the canonical nucleosome crystal model is a palindrome partially taken from a 171 bp monomer repeat derived from the human X chromosome and is known hereafter as Bunick alpha-satellite DNA. Alpha-satellite is a spatial designation, referring to a particular sequence that is found at the centromere and does not necessarily mean each sequence is identical.[1-3,7] This preferred sequence, from the Bunick group at Oakridge National Lab, allowed for increased resolution of nucleosome crystal structures because of the strength of positioning and the resultant symmetry.[7] However, another DNA alpha-satellite sequence is used in generating the CenP-A-containing nucleosome crystal model. This particular sequence, referred to as the Yokoyama sequence, is also palindromic but derived from an alpha-satellite monomer that contains a specific 17bp sequence called the centromere protein B (CenP-B) box.[8] Some alpha-satellite

monomers contain a CenP-B box that is recognized and bound by CenP-B and is of interest because it may play a role in helping to further mark the centromere as CenP-A does.

The study of nucleosomes is generally considered to fall under the nebulous cloud of genetics which is the study of how DNA sequence codes for a biological process. DNA sequence plays, in addition to histone structure, a large role in how stable the nucleosome is, as well as how well a gene is expressed or repressed.[7,9,10,1112] Therefore it was first assumed that the centromere location was determined by DNA sequence. However, the non-conserved DNA sequences present in human alpha-satellite imply that the functional kinetochore is governed, not by DNA sequence, but by another mechanism.[13,14,15] Indeed, studies of neo-centromeres have suggested that CenP-A, not alpha-satellite DNA sequence, is the minimal requirement to determine centromere position.[12,14] Chromosomes have been studied where CenP-A has been directed to another region rather than to the alpha-satellite DNA. It was observed that the functional kinetochore assembles at the region CenP-A is present and not where alpha-satellite is present. To this end, it is reasonable to hypothesize that Cenp-A nucleosome structure is different from canonical nucleosomes and this difference might be what marks the centromere region.[1,2,5,16,17]

The structure-function relationship between the centromere and CenP-A NCPs has yet to be fully resolved. The full CenP-A nucleosome crystal model, 3AN2 (Figure 1B), was solved by molecular replacement with the human canonical nucleosome structure 3AFA.[3,18] The crystal model, 3AFA, was also solved using the molecular replacement method, but with the Xenopus nucleosome crystal model 1AOI (Figure 1A).[6] There is essentially no difference between the canonical nucleosome and the centromere nucleosome structure. Therefore, contrary to my

hypothesis, this implies that the CenP-A molecule does not uniquely mark the centromere to facilitate its recognition by the kinetochore proteins. The only major difference observed was that 10bp termini on both ends of centromere nucleosome were not visible by x-ray diffraction and therefore have a higher degree of mobility.[3] However another model, one that supports my hypothesis proposed by the Black group, indicates that there is indeed a structural difference between H3 and CenP-A as shown in the tetramer structure (Figure 2A).[2]

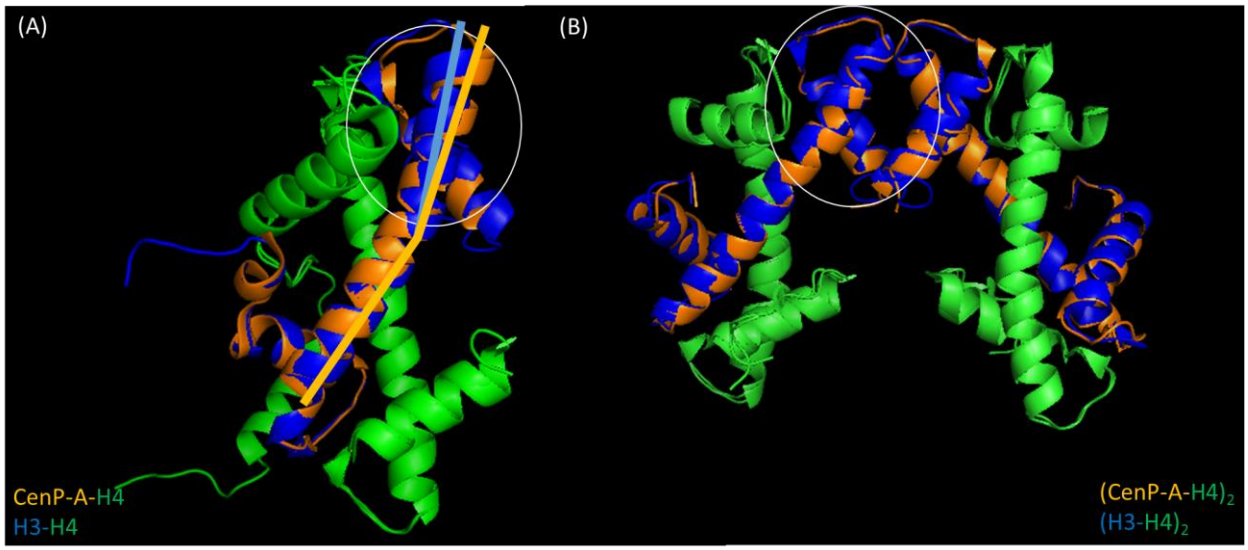

Figure 2: (A) Overlay of a (CenP-A-H4) dimer, PBD-3NQJ, crystal model proposed by the Black group and the (H3-H4), PDB-1AOI. The white circle indicates where the CenP-A rotation occurs relative to H3. (B) Overlay of the $(CenP\text{-}A\text{-}H4)_2$, PDB-3AN2, and $(H3\text{-}H4)_2$, PDB-1AOI, nucleosome crystal models. The white circle indicates where a rotation should be observed.

The structure solved by the Black group was only of the subunit $(CenP\text{-}A\text{-}H4)_2$ by itself and not the nucleosome as a whole, which was the case in the previous crystal models in Figure 1. As seen in the white circle of Figure 2A, the CenP-A dimer is rotated relative to the H3 dimer. In addition, as seen in Figure 2A, there is a kink in the second alpha helix of CenP-A relative to the second alpha helix of H3. The rotation is not observed in the overlay of the tetramer subunits,

isolated from Figure 1B, of the CenP-A nucleosome crystal model in Figure 2B. Results observed in the crystallized (CenP-A-H4) molecule show a rotation relative to (H3-H4) witnessed in Figure 2A, which is believed to result in a change of the overall nucleosome structure and is hypothesized to be a marker for other proteins to recognize functionally different nucleosomes from those present throughout the rest of the chromosome.[2,19,20] In addition to being structurally different from H3, CenP-A contains a targeting domain (CATD).[18,21] The CATD is a set of amino acids present in the L1-α2 region that is sufficient for a chaperone to distinguish CenP-A from H3. It is within the CATD region, specifically the α2 helix, that there is an increase in rigidity as shown by hydrogen-deuterium exchange experiments.[1] This rigidity in α2 is proposed to be the reason why helix rotation should be maintained once the nucleosome is assembled. Therefore it is hypothesized that adjusting the top of the CenP-A α2 helix to overlay with the H3 α2 helix would result in a narrowing of the diameter of the centromere nucleosome (Figure 3) as compared to the canonical nucleosome.[15,16]

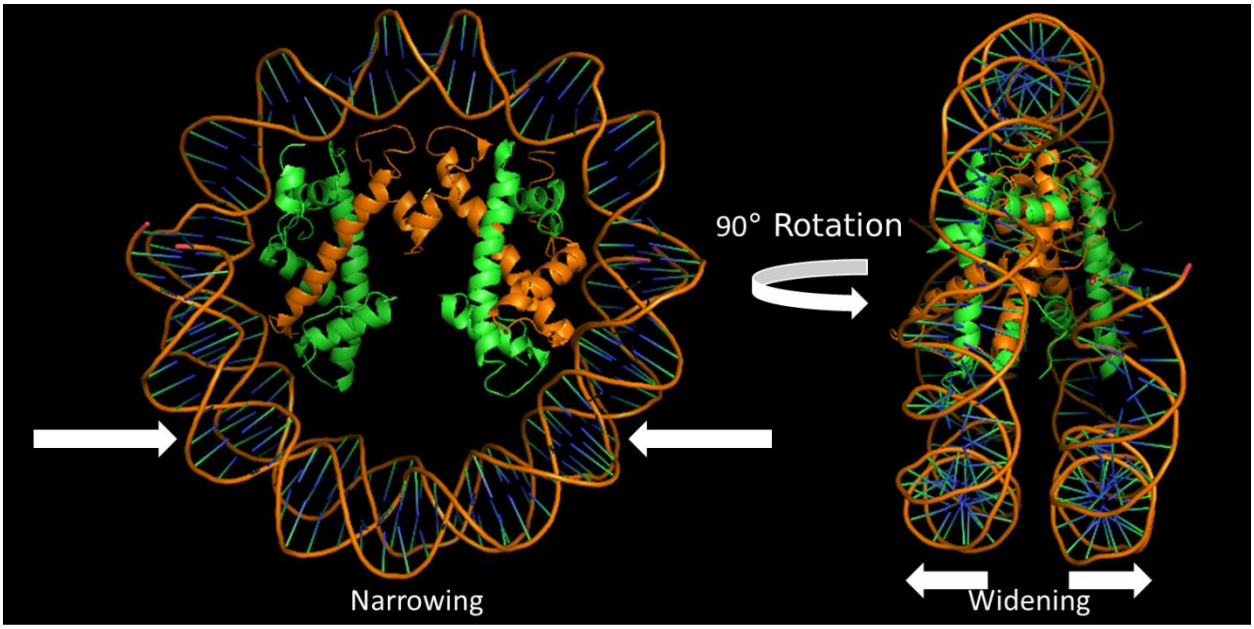

Figure 3: Expected changes in centromere nucleosome structure from canonical nucleosome structure. (H2A-H2B) dimers are not depicted, but were taken in to consideration when determining these changes in the nucleosome dimensions.

This narrowing in the diameter would result in the (H2A-H2B) dimer being pushed out and cause the nucleosome gyre to widen overall as demonstrated above.[2,15] This would be the basis for centromere nucleosome structural change relative to the canonical nucleosome structure and therefore be a unique centromere marker. These results would contradict the assertion that nucleosome assembly with CenP-A, in place of H3, does not change the overall structure and would provide a structural basis for the recognition of functionally different nucleosomes at the centromere. [2,3,15,16]

1.2 smFRET:

In order to determine if CenP-A causes a nucleosome structural change and is the epigenetic mark of the centromere, single-molecule Förster resonance energy transfer (smFRET) can be utilized to observe structural changes between Cenp-A and canonical nucleosomes. smFRET takes advantage of ratio metric changes in emission intensities of a fluorophore pair due to energy transfer from the donor to the acceptor in the nanometer distance range.[22] This range, in which FRET can be utilized is known as the Förster field and is dependent on the wavelength of the donor emission.[22,23] This field extends out a relatively small diameter from the molecule, for Cy3 the Förster field is between 1-8 nm, which is convenient for the NCP because their dimensions fall within this region.[6,22,23] The efficiency of FRET is a function of the distance between the two fluorophores, as shown below in Eqn. 1, and is sensitive to Ångstrom level changes.

$$E = \frac{R_o^6}{R_o^6 + R^6} \quad (1)$$

$$\text{Where, } R_0^6 = 0.53\left(\frac{\kappa^2 Q_D J}{\eta^4 N_{AV}}\right)$$

The characteristic distance R_0 is defined as the distance between two fluorophores where the non-radiative FRET relaxation pathway becomes equal to the fluorescence relaxation pathway, resulting in a FRET efficiency that becomes 50%. This distance is called Förster distance and is set for a given pair of fluorophores. In the case for the Cy3-Cy5 pair the Förster distance is 5.1nm. E is the FRET efficiency which depends on the dye pair and the distance between them. R is the distance between the fluorophore dye pair which determines the FRET E. κ^2 is called the orientation factor and depends on the angle between the emission dipole moment of the

donor and the absorption dipole moment of the acceptor. In order for excitation to occur the donor and acceptor must not be totally orthogonal to one another. Q_D is the quantum yield of the donor molecule and J is the overlap integral of the donor emission spectrum and the acceptor absorption spectrum. The quantum yield determines the percentage of donor's excited decay as a result of fluorescence emission and not through an energy transfer. The overlap integral is the chance that an acceptor absorbs the radiation emanating from the donor. η is the refractive index of the medium between the donor and acceptor molecule and N_{AV} is Avogadro's number. N_{AV} is a normalization such that the variable refers to a single acceptor within the near-field zone of radiation.[24] The refractive index is present due to conversion of the electromagnetic radiation to intensity.[24]

The one-dimensional changes in distances that can be observed with FRET efficiency are quite small, sub-angstrom in seconds' time scale, around the Förster distance due to the dependence of FRET efficiency on R^6. These small changes are difficult to detect using other structural probes available for large macromolecular complexes in solutions, such as small angle x-ray scattering.[25] X-ray crystallography provides a snapshot of the entire structure of a complex, which is identified as one of the most thermodynamically stable states, under the specific crystallization conditions.[26] When properly utilized to test a well-defined hypothesis, smFRET can provide a powerful and very efficient way to probe very small changes in a complex molecule in physiological solutions.

For smFRET measurements, very sensitive detectors such as electron multiplication charge coupled detectors (EMCCD) are required to record the emissions of single fluorophores.[23,27] Total internal reflection (TIR) offers a way to minimize and contain

fluorescence excitation to a thin surface layer of ~100 nm, which helps achieve a low fluorescence background level from diffusing molecules in solution.[28,29] TIR takes place when light traveling in one medium meets another medium with a refractive index different from the first one at an angle larger than the threshold angle determined by the Snell's Law.

$$\Theta_C = sin^{-1}(\frac{n_1}{n_2}) \quad (2)$$

Θ_C is the minimum angle needed to achieve TIR between two different mediums. Prior to reaching this angle light is both reflected and refracted and the ratio of the two amounts is dependent on the incident angle of the light. As the angle is increased the amount of refracted light decreases and reflected light increases until all the light is reflected.

By utilizing TIR, only a small depth of solution at the interface is illuminated because the evanescent wave that forms decays along the depth.[29]

$$I(z) = I(0)e^{-z/d} \quad (3)$$

$$z = \frac{\lambda_0}{4\pi n_2}(sin^2\Theta - sin^2\Theta_C)^2$$

Θ is the experimental incident angle, and λ_0, which is the wavelength of light used. By controlling the angle that light hits the surface molecules, a high signal-to-noise ratio is achieved by preventing any freely diffusing molecules in the solution from being excited.

The low fluorescence emission is collected by an electronically cooled EMCCD and each photon is converted into an electron and is amplified.[10,11] Photons are collected for a defined amount of time with a defined number of pixels that correspond to the recorded intensity of a single fluorophore and helps to control the signal-to-noise ratio. Increasing the acquisition time results in a greater number of photons collected from the fluorophore and therefore an

increased signal-to-noise ratio. Increasing the number of pixels results in an increased surface area for detection and therefore more photons collected from a fluorophore in less time. It is this technology that really has allowed single-molecule experiments with complex systems to become possible. smFRET measurements of nucleosomes with properly labeled fluorophores will allow observation of whether there is a narrowing of the nucleosome, widening of the gyre or both as a result of CenP-A incorporation. Here I resolve the difference between the canonical nucleosome and the centromeric nucleosome using smFRET to determine if indeed the centromere is marked by physically distinct nucleosomes.

Chapter 2: Experimental

2.1 DNA Constructs:

Two different 147bp DNA sequences were constructed from oligonucleotides ordered from Integrated DNA Technology Inc. (IDT, Iowa City, Iowa). Each of the two sequence has two different labelling positions in the following figure.

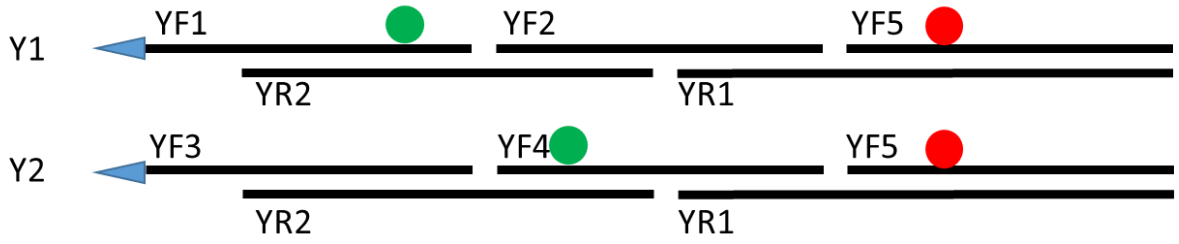

Figure 4: Y1 and Y2 are the different sequences that will be used to assemble nucleosomes. The green and red dots indicate where the internal linker is present and the position of the fluorophores. The blue triangles represent the biotin molecule.

After the individual oligonucleotides are labeled, each sequence is assembled by mixing the correct oligonucleotides in a 1:1 mole ratio in 10 mM tris(hydroxymethyl)aminoethane-HCl and 5 mM ethylenediaminetetraacetic acid (TE) buffer at pH 7.5 and annealed from 95°C to 4°C.[10,11] The first sequences, Y1 and Y2, are (Biotin-GCATGTAAGT ATCCTTCGTT GGAAACGGGA TTTCTTCAT**/iUniAmM/**T TCATGCTGG**/iUniAmM/**A CAGAAGAATT CTCAGTAACT TCTTTGTGCT GTGTGTATTC AACTCACAGA GTGGAACGTC CCTTTGCACA GA**/iUniAmM/**GCAGATTT GAAACACTCT TTTTGTAGTC GACAGAT). The individual oligonucleotide sequences for Y1 and Y2 are:

- YF1-/5Biosg/GCATGTAAGT ATCCTTCGTT GGAAACGGGA TTTCTTCAT/iUniAmM/T TCATGC
- YF2/5phos/TGGACAGAAG AATTCTCAGT AACTTCTTTG TGCTGTGTGT ATTCAACTCA CAGAGTG
- YF3-/5Biosg/GCATGTAAGT ATCCTTCGTT GGAAACGGGA TTTCTTCAT
- YF4-/5phos/TCATGCTGG/iUniAmM/A CAGAAG AATTCTCAGT AACTTCTTTG TGCTGTGTGT ATTCAACTCA CAGAGTG
- YF5-/5phos/GAACGTCCCT TTGCACAGA/iUniAmM/G CAGATTTGAA ACACTCTTTT TGTAGTCGAC AGAT
- YR1-ATCTGTTCGAC TACAAAAAGA GTGTTTCAAA TCTGCTCTGT GCAAAGGGAC GTTCCACTCT GTGAGTTGAA TACACACA
- YR2-/5phos/GCACAAAGAA GTTACTGAGA ATTCTTCTGT CCAGCATGAA ATGAAGAAAT CCCGTTTCCA ACGAAGGAT

Each of the sequences that contain (/iUniAmM/) has an internal 1° amine modification covalently linked as part of the phosphate backbone between two bases. A biotin molecule (5Biosg) is covalently linked to the 5' end of the DNA sequence, which serves as a way to tether the future nucleosome to the surface.

Each sequence after the annealing process is purified using a PCR purification kit from Qiagen Inc. (Qiagen, Germantown, MD). After purification, utilizing 400 U of commercially available T4 ligase and 1X buffer (New England Biolabs, Ipswich, MA) the oligonucleotides are covalently linked and four different 147 bp dsDNA sequences are formed. The dsDNA of the reaction mixture is purified from T4 ligase using the Qiagen PCR purification kit once again.

In order to label the oligonucleotides, the N-hydroxysuccinimide Ester (NHS) Cy3 and Cy5 fluorophores were dissolved in DMSO.[30] The dyes were added in 15x molar excess to the DNA, such that the final ratio of DMSO:buffer was 1:5. The final DNA buffer contained 0.09M sodium carbonate and 0.5M sodium chloride and reacted for 1 hour at 75°C. The DMSO helps

to facilitate the chemical interaction between the internal linker and the dye as shown in the following figure.

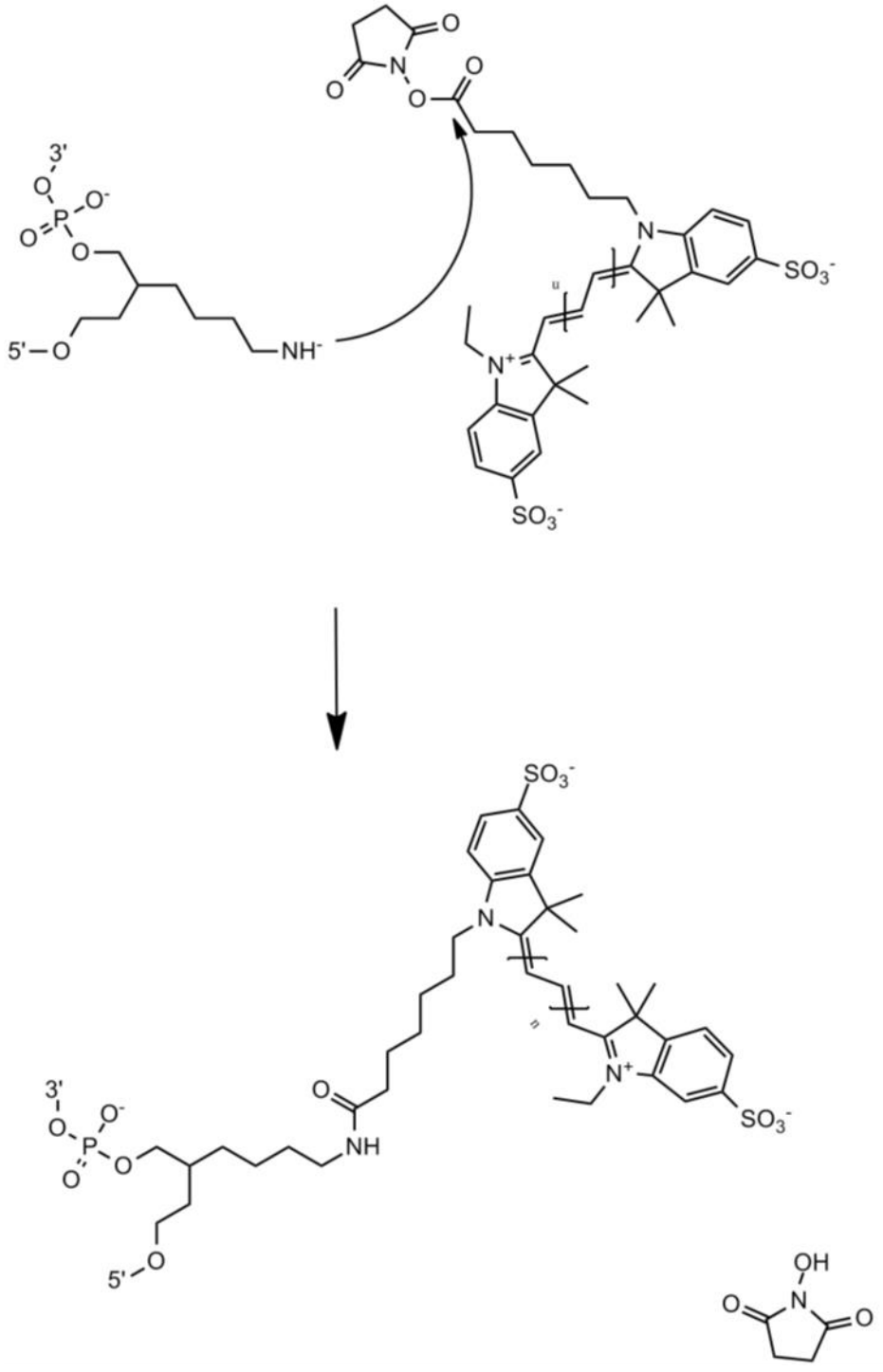

Figure 5: The internal linker, shown in the upper left corner of the figure, is placed in the phosphate backbone of an oligonucleotide and deprotonated. The value of n depends on whether or not Cy3 or Cy5 is being used.

The sodium carbonate deprotonates the amine linker so that the nucleophilic substitution occurs. The heat and the salt help prevent any secondary structure of the oligonucleotide. After the reaction is completed, some excess dye is removed by ethanol precipitation. The remaining DNA was dissolved in TE buffer. The remaining excess dye was removed by dialysis against 500 mL of TE buffer, for two hours, three times.[28]

The resultant DNA is then assembled into a nucleosome, utilizing human histone tetramers and dimers supplied by Ben Black of The University of Pennsylvania in Philadelphia, as seen below.

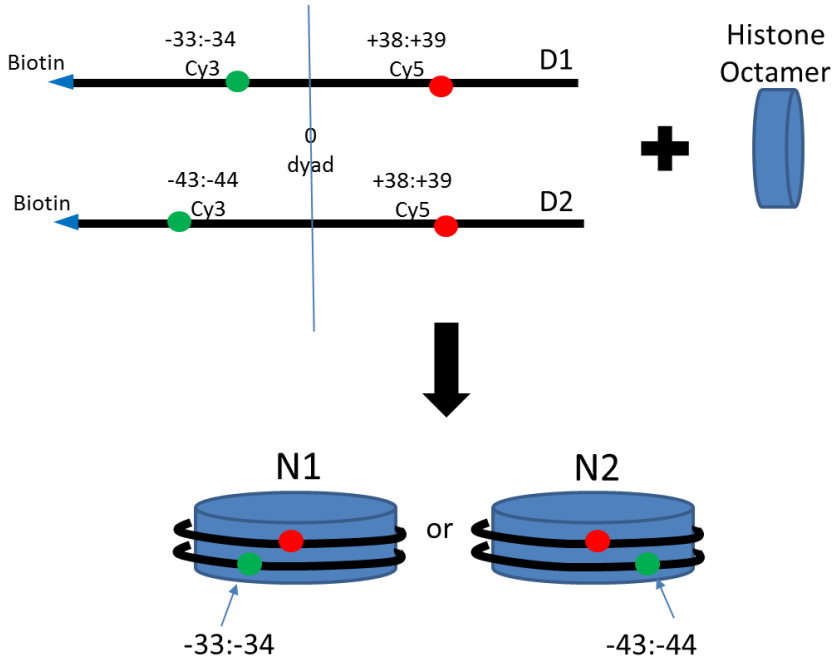

Figure 6: Canonical nucleosome and CenP-A containing nucleosome labelling scheme. D1 is the DNA strand with the donor fluorophore in between bases -33 & -34 upstream from the dyad. D2 with the donor between bases -43 & -44. Both D1 and D2 contain an acceptor fluorophore

between bases 38 & 39. N1 and N2 represent the subsequent nucleosomes formed from addition of the core histones.

The dyad serves as position zero for the insertion of the primary amine modifier with each sequence having two internal linkers. These positions, for Yokoyama sequences, were between the 38th and 39th bases upstream of the dyad and between either the 33rd and 34th or 43rd and 44th bases downstream of the dyad as seen above. This labelling scheme will allow for the changes in the nucleosome. In addition, the positions selected should have a minimal impact on nucleosome structure. These particular positions are in between the DNA-histone contacts and are pointed away from the protein complex.[3,6] Luger *et. al.* observed that these phosphates have a high degree of mobility relative to those that contact the histone. This observation coupled with the flexible linker and soluble cyanine dye allows the assumption to be made that the dyes are freely rotating.

2.2 Nucleosome assembly:

The histone tetramers and dimers used to construct the nucleosomes were received from the Ben Black group of The University of Pennsylvania. The optimal DNA to tetramer to dimer mole ratio was found by systematic addition of histone protein to a solution of 1000ng of DNA.[31] The DNA solution used varied between 150ng/µL to 200ng/µL in TE Buffer. A salt solution of 4M NaCl was added to the DNA so that the final salt concentration would be 2M NaCl so that the histone protein added would be stable.[29] TE buffer concentrated 50X was added to the DNA solution so that the final solution would contain 1X concentration of TE buffer. Water was then added, so that upon addition of histone tetramer and dimer, the final solution volume would be 25µL. The histone tetramer and dimer were the last to be added to the solution thereby reducing the chance of disassembly. Binding of histones to DNA was carried out by stepwise dialysis with each step one hour long at decreasing NaCl concentrations; 0.85 M, 0.65 M, 0.5 M, 0.35 M and 0.15 M all with constant TE buffer concentration.[29] The nucleosomes were subsequently heated to 55°C, in a PCR thermal cycler, to reduce the nucleosome positioning heterogeneity. This increases the population of histone octamers centered on the dyad of DNA. The DNA is then centrifuged on a table top centrifuge for one minute and then transferred to another microcentrifuge tube. This helps minimize any aggregates that may be present as a result of the heating process.

2.3 Experimental Nucleosome Scheme:

By utilizing the two variations in labelling of the DNA sequences two dimensional changes can be deduced despite the one dimensional FRET E changes. One assumption made in this case is that if the change is small then the change in the angle between the two fluorophores is negligible. Therefore, the changes observed are only due to horizontal and vertical changes, defined as changes in X and Y planes respectively, from the initial position of the fluorophore to the final position of the fluorophore. The three changes in nucleosome structure that are expected to occur are presented in the following figure.

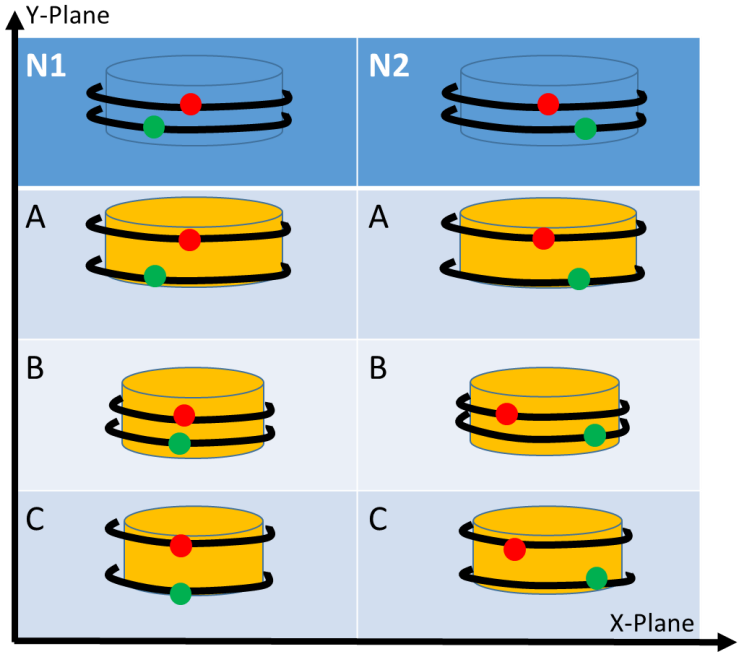

Figure 7: A cartoon reflecting the possible changes in the nucleosome structure upon incorporation of CenP-A (a-c) in place of H3 (N1 & N2). The first scenario could result in nucleosome widening (a), the second in just narrowing (b), the third case could be a combination of both (c).

N1 and N2 represent the nucleosomes assembled with $(H3\text{-}H4)_2$ and models A, B and C represent the nucleosomes assembled with $(CenP\text{-}A\text{-}H4)_2$. In the case of nucleosome gyre widening the inter-dye distance increases in the Y plane and a decrease in FRET E will be observed for both N1 and N2. In the case of narrowing, an increase in FRET E will be observed in N1 and a decrease in N2, in the X plane. In the case of both widening and narrowing, a small change in FRET E will be observed in N1 and a large change in N2.

2.4 smFRET Experimental Set-Up:

Quartz slides and glass coverslip surfaces were functionalized with an amine modifier and incubated with a solution of 100:1 PEG-COOH to Bio-PEG-COOH.[10] The quartz slides and glass coverslips are placed on top of one another and sealed with epoxy. The chamber is incubated with streptavidin protein, rinsed with buffer, incubated with NCPs and rinsed with a final buffer solution as seen in the following figure.

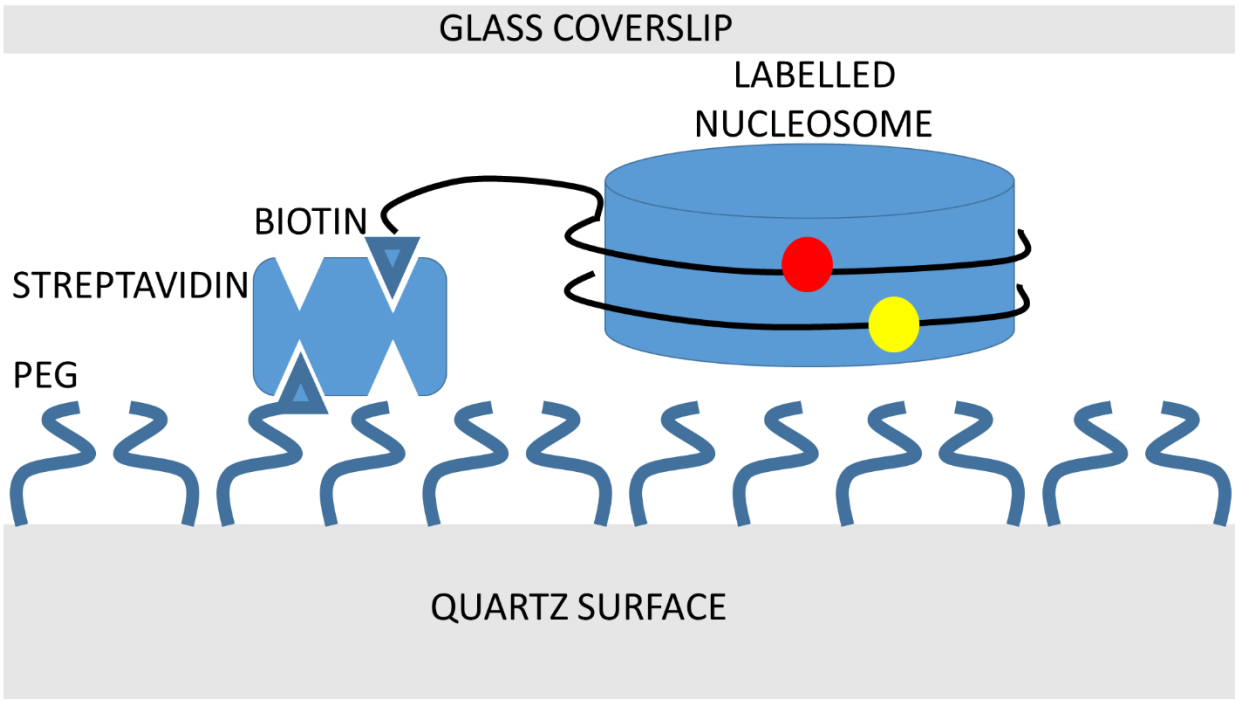

Figure 8: A cartoon illustration of a single-molecule surface. Each nucleosome assembled contains a biotin molecule at the end of the DNA so that it can be tethered to the surface.

The final buffer solution contains 4U of PCD, 0.01M PCA and 2mM trolox to increase the time before photobleaching and minimize dark states.[10] The Cy3 donors on the surface of the

chamber are illuminated using a prism coupled microscope (TE2000, Nikon, Tokyo Japan) for TIR with a green LASER at 532nm (Laser Quantum, UK) as shown in Figure 9.

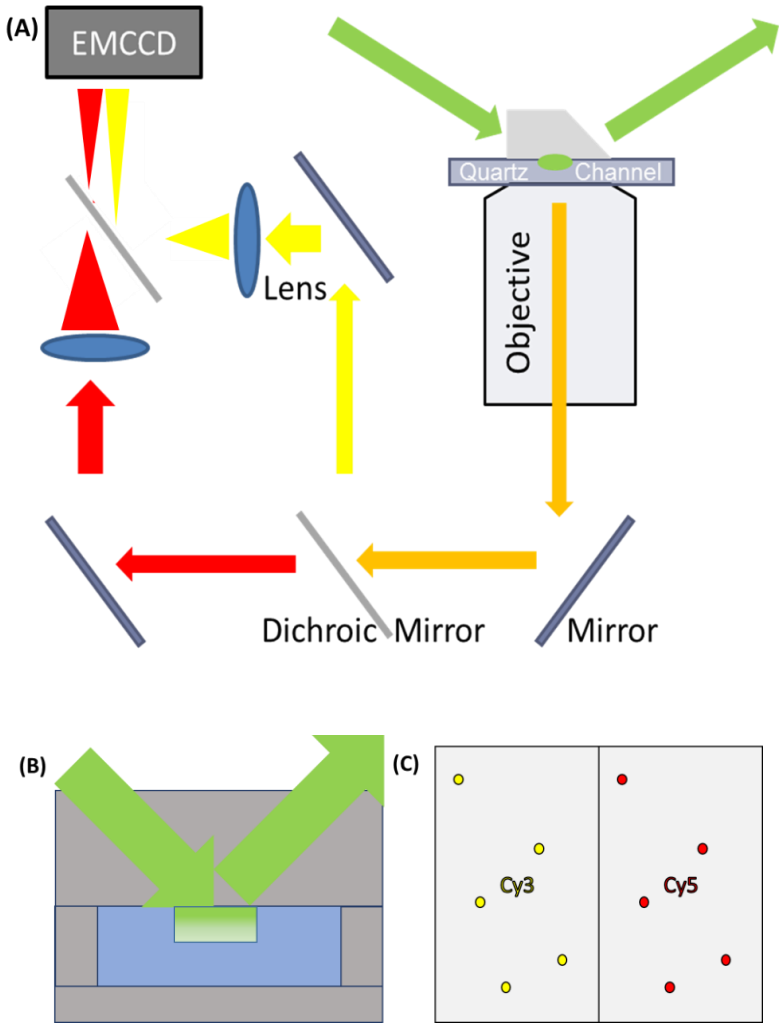

Figure 9: (A) Illustration of single molecule set-up in which green LASER light excites surface molecules and fluorophore emission is separated using a dichroic mirror. The emission is redirected to the EMCCD detector for recording. (B) A zoomed in representation of a channel where excitation occurs by way of TIR. (C) An example of what is displayed on the monitor as emission is collected on the EMCCD.

The donor and acceptor emission intensities are separated in to two spectral ranges, 550-645 nm and 645-750 nm. The emission of each fluorophore is simultaneously projected onto ½ of the EMCCD (EMCCD, iXon+897, Andor Technology, Belfast UK) and collected at a rate of 1/50 ms. As seen above in Figure 9B, TIR occurs at the glass-water interface and the intensity decays exponentially preventing any scattering or solution fluorescence. There is no reflection between prism and air because an index of refraction matching oil is coated on the prism interface of the quartz slide. The fluorophore emission for single molecules, collected by the two halves of the EMCCD in Figure 9C, is correlated to each other experimentally. This is done by fitting the peak of fluorescent beads, whose emission spectra is long enough that it is collected on both sides of the EMCCD, and relating them to one another.[10]

2.5 smFRET Data Collection and Analysis:

The intensity collected by the EMCCD during one frame, for the two fluorophores, is then plotted as a function of time. These two intensities collected create a single FRET trace as shown in the following figure.

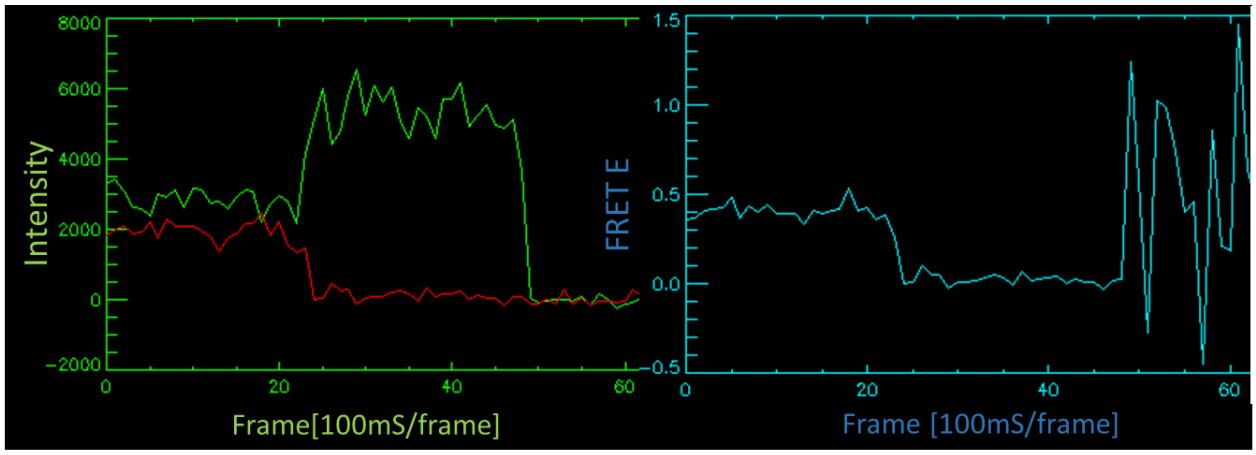

Figure 10: Sample trace for a single nucleosome complex that is labeled with a Cy3 donor fluorophore, green trace, and a Cy5 acceptor fluorophore, red trace. The intensity is plotted over the number of frames, which is 100ms/frame. The corresponding FRET E is plotted over acquisition time as well. The sudden cessation of acceptor fluorescence just after the 20th frame indicates acceptor photobleaching and the termination of FRET, which is also indicated in decrease in FRET E in the same corresponding frame. The FRET E values collected for histogram assembly occurs prior to the 20th frame.

Each data point from the FRET E trace contributes to one point in the histogram. The FRET E is the ratio of recorded fluorophore intensities as seen in Eqn. 4.

$$E = \frac{I_{Cy5}}{I_{Cy3}+I_{Cy5}} \quad (4)$$

I_{Cy5} and I_{Cy3} are the intensities of the emission of the donor and acceptor fluorophores from Figure 9. The regions where FRET occurs are manually selected, ignoring the regions where only a single fluorophore emits or neither emits, and compiled in to a histogram.[10] The regions in which a single fluorophore emits are ignored because, as seen from Eqn. 2, the very high and low FRET E values are asymptotic and do not provide reliable distance information. In addition, it is impossible to determine, like in the above case, if the nucleosome disassembled or if the Cy5 molecule photobleached. An average FRET E is determined from the compiled histogram for a population. This is used to compare two or more populations of nucleosomes together.

Once the FRET E is determined for a nucleosome population, the corresponding fluorophore pair distance is calculated using Eqn. 1. The distance vector determined for each nucleosome population can be thought of as a composition of two other vectors. The change in distances between the canonical nucleosomes and centromere nucleosomes could be a result of a narrowing and widening of the nucleosomes as described in Figure 3. If the nucleosome in Figure 3A was rotated in to the plane of the paper the widening should appear as a vertical translation between the two fluorophores on the DNA backbone, whereas the narrowing should appear as a horizontal translation between the two fluorophores on the DNA backbone.

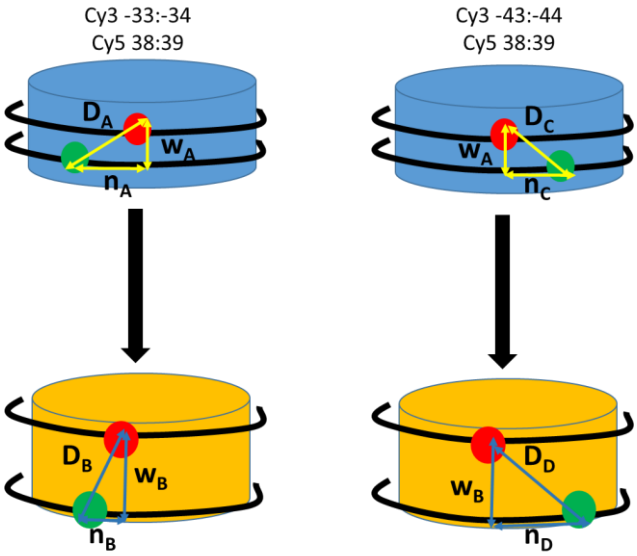

Figure 11: The two columns of nucleosomes represent the two different labelling positions. The two rows represent nucleosomes assembled with canonical H3, blue, and CenP-A, yellow. D represents the one-dimensional measurement that is calculated from smFRET measurements. w and n are the vertical and horizontal translations of the dyes as a result of the change in nucleosome structure. The w for each row is assumed to be the same, unlike the n for each nucleosome.

This is a simplification because the changes involve an angular component. However on a small enough scale, it can be ignored, which is initially assumed to be the case. The resultant deconvolution of D into its component horizontal and vertical contributions is a geometric problem where we can take advantage of the Pythagorean Theorem as shown below.

$$D_A^2 = w_A^2 + n_A^2 \quad (6.1)$$

$$D_B^2 = w_B^2 + n_B^2 \quad (6.2)$$

$$D_C^2 = w_A^2 + n_C^2 \quad (6.3)$$

$$D_D^2 = w_B^2 + n_D^2 \text{ (6.4)}$$

D is the distance between the fluorophore pair determined from the FRET E, w is the vector component due to widening and n is the vector component due to narrowing. D_A and D_C are the distances for H3 containing nucleosomes with labeling between [-33:-34] and [-43:-44] respectively. D_B and D_D are the distances for CenP-A containing nucleosomes with labeling between [-33:-34] and [-43:-44] respectively. Assuming w does not change between DA and DC or DB and DD allows the difference to be taken and the elimination of w as a variable, shown below.

$$\Delta\Delta D^2 = (D_D^2 - D_C^2) - (D_B^2 - D_A^2) = (n_D^2 - n_C^2) - (n_B^2 - n_A^2) \text{ (7)}$$

$\Delta\Delta D^2$ is the difference between the two sets of nucleosome structures. The value for the narrowing between the two sets of the nucleosomes should be equal but the last term is negative rather than positive. Therefore the equation can be condensed as shown below.

$$\Delta\Delta D^2 = 2(n_B^2 - n_A^2) = 2(n_B + n_A)(n_B - n_A) \text{ (8.1)}$$

$$\frac{\Delta\Delta D^2}{2(n_B + n_A)} = (n_B - n_A) = \Delta n \text{ (8.2)}$$

By guessing the solution to the variables n_B and n_A the difference in narrowing (Δn), between the canonical and centromere nucleosome, can be determined. Once the change in the Δn reaches a minimum, a solution is found for the change in narrowing. Using the Δn, the difference in widening (Δw) can be found in the following equations.

$$-\Delta w_1 = -w + (D_C^2 - (n_A - \Delta n)^2)^{1/2} \text{ (9.1)}$$

$$\Delta w_2 = -w + (D_D^2 - (n_B + \Delta n)^2)^{1/2} \text{ (9.2)}$$

$\Delta w_{1,2}$ are both the difference in widening determined from the two different nucleosome systems, which turn out to be equal. The ratio of the differences (N) can be compared to determine what factor contributes the most to the changes in the nucleosome structure as shown below.

$$N = \frac{\Delta n}{\Delta w_2} \text{ (10)}$$

It is not necessary to look at the ratio between the narrowing and widening because Δw_2 is equal in absolute value to Δw_1. This determines what contributes more to the nucleosome structure change, either the narrowing or widening due to CenP-A incorporation.

Chapter 3: Results

3.1 Nucleosomal DNA Constructs & Assembly:

Free DNA for nucleosome assembly and nucleosomes are shown in a 6% native polyacrylamide gel run for 42 minutes at a constant 120V. The optimum molar ratio of histone to DNA for nucleosome formation is found by varying the molar amount of histone protein to DNA as described in section 2.2.

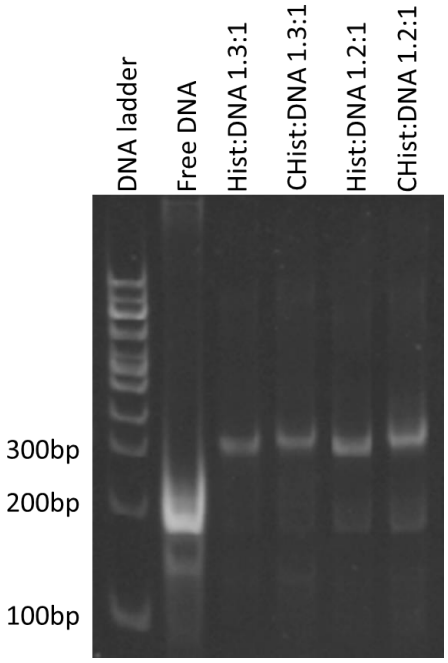

Figure 12: Native PAGE of ligated DNA and nucleosomes assembled from the DNA. The gel contains both canonical nucleosomes, with ratios labeled as Hist:DNA, and centromere nucleosomes, CHist:DNA.

The histone to DNA ratios were found to be best at the ratios, between 1.3:1 and 1.2:1, demonstrated in the above figure. CenP-A has a slightly higher molecular weight as indicated by the band location being slightly above 300bp band whereas the canonical nucleosome is right in

line with the 300 bp band. In using the Qiagen PCR kit all of the DNA that is not 147 bp long, which accounts for the DNA band closer to 100 bp, is not removed. The same conditions for nucleosome assembly were carried out for the nucleosomes with alpha-satellite DNA that contains different labeling positions and sequences.

3.2 smFRET measurements:

As mentioned in the experimental section the regions of FRET E are manually selected from an individual trace for a single nucleosome molecule. Each data point from the selected range of FRET E is used to construct the histograms with an average that most closely reflects the distances measured from the crystal structure (PDB ID: 3AN2) and is the most populous of states, as shown in the following figure.[3] Additionally, each set of nucleosome histograms were constructed with more than 80 nucleosomes.

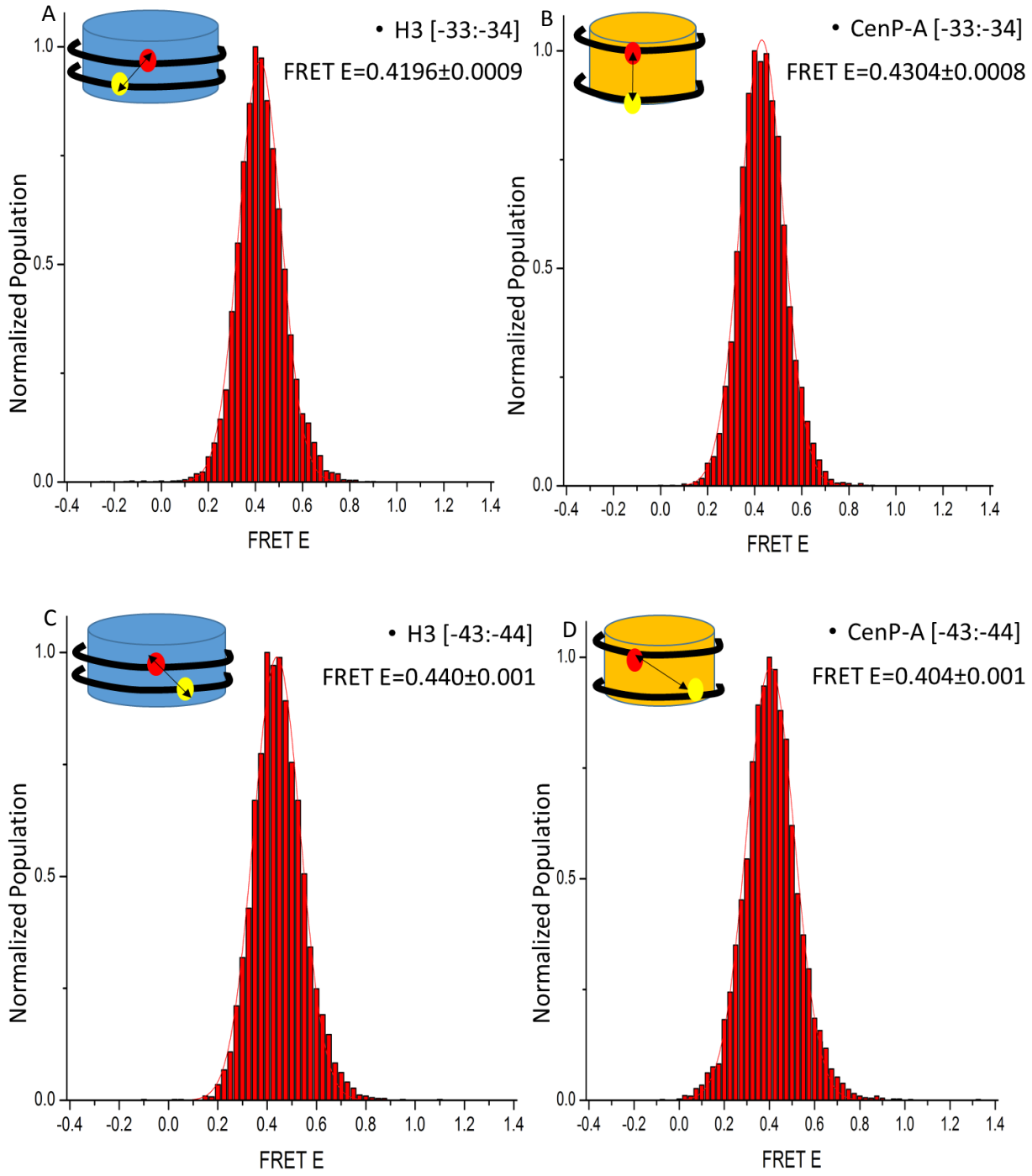

Figure 13: Yokoyama FRET E histograms for each nucleosome sample. Each contains a cartoon illustration of the histogram nucleosome. A & C are nucleosomes containing the canonical H3 histones. B & D are nucleosomes assembled with CenP-A histones in place of H3.

These experiments were done using the same DNA sequence used for the crystal structure in Figure 1. The difference between FRET E for the first set of nucleosomes is 0.011±0.001 while the change in FRET E for the second set of nucleosomes is -0.036±0.001. While the changes in structure between the nucleosome are small they are statistically significant. The absolute changes in FRET E for N1 and N2 are not equal which indicates both narrowing and widening effects are observed.

Utilizing the equations described in section 2.5, the distances for each nucleosome were determined. The results were compiled in the following table.

	FRET E			**D/[nm]**		
	H3	**CenP-A**	**ΔE**	**H3**	**CenP-A**	**ΔD**
Y1	0.4196(±.0009)	0.4304(±.0008)	0.011(±.001)	5.38	5.34	-0.04
Y2	0.440(±.001)	0.404(±.001)	-0.036(±.001)	5.31	5.44	0.13

Table 1: Compilation of FRET E and distances for each nucleosome. Y1 and Y2 indicates Yokoyama nucleosomes assembled with a fluorophore between [-33:-34] and [-43:-44] respectively. The distances and changes in distances calculated are in nm.

From the distances calculated in Table 1, the relative changes in narrowing and widening were found as a result of CenP-A incorporation into the assembled Yokoyama nucleosomes. A ratio N=4.4 for the Yokoyama assembled nucleosome set were determined as described in section 2.5. As mentioned earlier, N represents the ratio of the change in narrowing of nucleosome to the widening of the nucleosome. This value of N shows that greatest contribution to the change of dimensions for Yokoyama nucleosomes are a result of narrowing by a factor of four.

Chapter 4: Discussion

In order for the mitotic spindle to recognize the centromere, it must contain a unique mark that distinguishes itself from the rest of the chromosome. Previously, two crystallographic models predicted contradictory results that included an overall compaction of the nucleosome or no change at all.[2,3] The crystallized (CenP-A-H4)$_2$ predicted a dimensional change by way of nucleosome narrowing and widening.[2] The (CenP-A-H4)$_2$ was crystallized in HEPES buffer at an approximate pH of 7.2 and 125 mM NaCl salt concentration. The salt concentration and pH closely reflect the physiological conditions in the nucleus of a cell and it is reasonable to say that the crystallographic structure obtained is a close model of what may exist physiologically. The buffer conditions that the small angle x-ray scattering experiment was performed under contained 0.2mM phosphate buffer and 1 M NaCl, 1 mM EDTA, 8 mM DTT and 2.4% glycerol.[2] The high salt concentration is needed to keep the tetramer stable in solution in the absence of chaperones. The SAXS experiment determined that the change between (CenP-A-H4)$_2$ and (H3-H4)$_2$ should be about 1 nm in solution. The SAXS experiments corroborated the crystal model predictions and provide further indication that the centromere may be marked by a physical change.

However, the crystallized CenP-A nucleosome indicated that no such change should be observed.[3] It is possible that the population sampled by this method is not indicative of the structure in solution and is insignificant in solution. The crystallized CenP-A nucleosome was assembled using the Yokoyama alpha-satellite DNA sequence. The entire nucleosome was crystallized in a 20 mM potassium cacodylate buffer at pH 6.0 with 38-56 mM KCl and 70-75 mM $MnCl_2$. These conditions are very similar to the canonical nucleosome by which all

comparisons are drawn: 40-46 mM MnCl2, 35-40 mM KCl and 20 mM potassium cacodylate at pH 6.0. The crystal forming conditions are not physiological and the decrease in pH results in an increase in positive charge of the proteins causing tighter electrostatic binding between histones and the phosphate backbone. The presence of divalent metals, such as Mg^{2+}, at concentrations greater than 2 mM have been known to cause tighter binding in nucleosome arrays.[32] This effect could also be observed in the presence of Mn^{2+} causing single nucleosomes and nucleosome-nucleosome packing to be tighter. Therefore, it is not unreasonable to assume that part of the reason the CenP-A nucleosome resembles the canonical nucleosome so closely is a result of the method used to solve the crystal structure.

Upon nucleosome assembly by stepwise dialysis a difference in vitro is observed under more physiological conditions. It is important to note that although a change was measured, using smFRET, it is small but statistically significant. The change in narrowing is 4.4 times greater than the widening when using the Yokoyama DNA. The nucleosomes assembled with the Bunick alpha-satellite DNA sequence show a much greater change in nucleosome structure. The biggest difference is that the widening is 2 times greater than the narrowing. The change in nucleosome narrowing for the Yokoyama nucleosomes is about 1 Å and the widening about 0.25 Å the change in nucleosome narrowing. The contribution from DNA sequence to nucleosome structure could explain, in addition to the salt concentration, the very little change observed by X-Ray crystallography with the Yokoyama sequences.

The SAXS experiments and the tetramer crystal model did predict, and indeed the smFRET measurements agree, that a narrowing and widening would be observed once the nucleosome was formed in order to mark the centromere physically.[2] However, the degree to

which change was expected differed from the smFRET results. The SAXS model measured a 10 Å difference between $(CenP\text{-}A\text{-}H4)_2$ and $(H3\text{-}H4)_2$, which was not observed by smFRET. The reason that the predicted difference of SAXS and observed difference by smFRET could result from a bias due to the DNA sequence. The DNA sequence effects both positioning of the octamer as well as the stability of the nucleosome formed. The change in centromere nucleosome structure implies that the function of the nucleosome is different.

Further experiments are needed in order to discern exactly how the structure affects the function. It is posited that a change in nucleosome structure upon incorporation of CenP-A is the unique centromere mark.[5] There are two possibilities that may occur considering CenP-A marks the centromere by a unique physical change. The first is that the CenP-A containing nucleosomes may mark the centromere and in turn recruit other proteins that recognize this change and continue to help the recruitment of more CenP-A. The second is that sequential CenP-A containing nucleosomes interact differently with one another, as compared to their canonical counterparts, and it is the array that marks the centromere and functional kinetochore. In either case, the different structure dictates the different function.

References

[1] Black, BE; Foltz, D; Chakravarthy, S; Luger, K; Woods, V; Cleveland, D; *Nat*, 2004, 430, 578-582

[2] Sekulic, N; Bassett, E; Rogers, D; Black, BE; *Nat*, 2010, 467, 347-352

[3] Tachiwana, H; Kagawa, W; Shiga, A; Osakabe, T; Miya, Y; Saito, K; Hayashi-Tanaka, Y; Oda, T; Sato, M; Park, S; Kimura, H; Kurumizaka, H; *Nat*, 2011, 476, 232-237

[4] Alberts, B; Johnson, A; Lewis, J; Raff, M; Roberts, K; Walter, P; DNA and Chromosomes. *Cell,* 4th Edition; Garland Science, New York, 2002; pp 198-2011

[5] Panchenko, T; Black, BE; *Prog Mol Subcell Biol,*2009, 48, 1-32

[6] Luger, K; Mader, A; Richmond, R; Sargent, D; Richmond, T; *Nat,* 1997, 389, 251-260

[7] Harp, J; Uberbacher, E; Roberson, A; Palmer, E; Gewiess, A; Bunick, G; *Acta Cryst*, 1996, D52, 283-288

[8] Tanaka, Y; Nureki, O; Kurumizaka, H; Fukai, S; Kawaguchi, S; Ikuta, M; Iwahara, J; Ozakazi, T; Yokoyama, S; *Embo J*, 2001, 20(23), 6612-6618

[9] Widom, J; Lowary, PT; *J. Mol. Biol.*, 1998, 276, 19-42

[10] Choy, JS; Wei, S; Lee, JY; Tan, S; Chu, S; Lee, TH; *J. Am. Chem. Soc.*, 2010, 132, 1782-1783

[11] Lee, JY; Wei, S; Lee, TH; *J. Biol. Chem.*, 2011, 286, 11099-11109

[12] Sullivan, B; Karpen, G; *Nat. Struc. Mol. Biol.*, 2004, 11,1076-1083

[13] Sullivan, KF; Hechenberger, M; Masri, K; *J. Cell Biol.*, 1994, 127, 581-592

[14] Rudd, MK; Wray, G; Willard, H; *Genome Res.*, 2006, 16, 88-96

[15] Hasson, D; Alonso, A; Cheung, F; Tepperberg, J; Papenhausen, P; Engelen, J; Warburton, P; *Chromosoma*, 2011, 120, 621-632

[16] Sekulic, N; Black, BE; *Cell*, 2012, 37, 220-229

[17] Cleveland, D; Mao, Y; Sullivan, K; *Cell*, 2003, 112, 407-421

[18] Tachiwana, H; Kagawa, W; Osakabe, A; Kawaguchi, K; Shiga, T; Hayashi-Takanaka, Y; Kimura, H; Kurumizaka, H; *PNAS*, 2010, 107, 10454-10459

[19] Hu, H; Liu, Y; Wang, M; Fang, J; Huang, H; Na, Y; Li, Y; Wang, J; Yao, X; Shi, Y; Li, G; Xu, R; *Genes and Devel.*, 2011, 25, 901-906

[20] Bassett, E; Denizio, J; Barnhart-Dailey, M; Panchenko, T; Sekulic, N; Rogers, D; Foltz, D; Black, BE; *Devel. Cell*, 2012, 22, 749-762

[21] Panchenko, T; Sorenson, T; Woodcock, C; Kan, ZY; Wood, S; Resch, M; Luger, K; Englander, SW; Hansen, J; Black, BE; *PNAS*, 2011, 108, 16588-16593

[22] Van Der Meer, BW; Coker III, G; Chen, SY; Introduction, *Resonance Energy Transfer Theory and Data*, VCH Publishers, New York, 1994, pp 1-2

[23] Lakowicz, J; Introduction to Fluorescence, *Principles of Fluorescence Spectroscopy*, 3rd Edition, Springer, New York, 2006, pp 13-25

[24]Förster, T; *Ann. Phys.*, 1948, 2, 55-75

[25] Rhodes, G; Other Diffraction Methods, *Crystallography Made Crystal Clear*, 3rd Edition, Academic Press, New York, 2006, pp 211-235

[26] Rhodes, G; An Overview of Protein Crystallography, *Crystallography Made Crystal Clear*, 3rd Edition, Academic Press, New York, 2006, pp 7-30

[27] Roper Scientific Inc; PVCAM, A High-Level C Library, *PVCAM 2.6 Roper Scientific Beyond Imaging*, Roper Scientific, Trenton, 2003, pp 7-10

[28] Axelrod, D; *J. Cell Biol.*, 1981, 89, 141-145

[29] Axelrod, D; Total Internal Reflection Fluorescence Microscopy, *Optical Imaging in Microscopy*, Springer, Berlin, 2007, pp 195-236

[30] Invitrogen Molecular Probes Inc.; *Alexa Fluor Oligonucleotide Amine Labeling Kit*, Invitrogen, Paisley, 2006, MP2091-MP2094

[31] Luger, K; Rechsteiner, T; Richmond T; Expression and Purification of Recombinant Histones and Nucleosome Reconstitution, *Chromatin Protocols Methods in Molecular Biology*, Humana Press Inc., Totowa, 1999, 119, 1-16

[32] Grigoryev, S; Arya, G; Correll, S; Woodcock, C; Schlick, T; *PNAS*, 2009, 106, 13317-13322