Dissertation

submitted to the

Combined Faculty of Natural Sciences and Mathematics of the

Ruperto Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

M.Sc. Madhura De

Born in: Kolkata, India

Oral examination: 23rd September 2021

The effects of the linker-DNA on linker histone positioning: a single-pair FRET study from mono- to trichromatosomes

Referees

Prof. Dr. Rebecca C. Wade

Prof. Dr. Karsten Rippe

Summary

The nucleosome consists of a core complex of two copies each of four histone proteins wrapped by about 1.65 turns of DNA. The DNA arms entering and leaving the core are known as linker-DNA (L-DNA) arms. The linker histone, H1, associates with the nucleosome at the region bounded by the two ends of the DNA leaving the core. The nucleosome in conjunction with the H1 forms the chromatosome, the smallest repeating unit of the chromatin. How does H1 associate with the nucleosome? Are there any contributions from the L-DNA stretches flanking the core to H1 association? Does the length or the sequence of the L-DNA affect the way H1 associates with the nucleosome? Does the binding mode of H1 affect the higher-order structuring of the chromatin? The work presented in this thesis aims to address these questions.

Chromatosomes were reconstituted from core histone octamers (*Xenopus laevis*), linker histone of subtype H1.0b (*X. laevis*), and 226 bp DNA containing the strongly positioning Widom 601 sequence. The linker histone was labelled with the fluorophore Alexa 488 on the globular domain or the upstream end of the C-terminal tail domain (CTD), and the DNA was labelled with Alexa 594 on either one or the other L-DNA arm. Single-pair FRET (Förster Resonance Energy Transfer) spectroscopy was used to measure the proximity of the globular domain or the CTD to one or the other L-DNA. First, it was determined how the length and the sequences of the L-DNA flanking the H1 affect the location of the H1 on the nucleosome. It was observed that the structured globular domain of the H1 was proximal to L-DNA arms that contained an A-tract consisting of eleven contiguous adenines. However, the globular domain was not proximal to either a purely GC tract or a mixed sequence having a 64% AT content. The fluorophore on the disordered CTD was equidistant from both the L-DNAs, despite sequence variations. Distances extracted from the single-pair FRET measurements were used to build models of A-tract-containing nucleosomes. It was observed that the globular domain of the H1 associated in an on-dyad fashion on the nucleosomes, and two conserved arginine residues on the globular domain of H1 were proximal to the characteristically narrow minor groove of the A-tract. To experimentally check whether the A-tract recognition was mediated via the minor grooves as observed in the models, or via hydrophobic interactions with the thymine methyl groups, a nucleosome was reconstituted with two A-tracts, one complementary to thymine and the other complementary to methyl-group-lacking deoxy-uridine. The globular domain showed similar proximity to both the A-tracts, proving the role of the A-tract minor groove in its recognition.

Finally, I studied how linker histones compact two types of trinucleosomes containing A-tracts flanking the first and the third nucleosome either on the inner or the outer L-DNA. The extent of compaction was measured by single-pair FRET between the two inner L-DNAs, one joining the first and second, and the other joining the second and third nucleosomes. Based on the mononucleosome models, linker histones associating with the first and the third nucleosome are expected to be oriented towards the A-tracts on the outer or the inner L-DNAs. Both the trinucleosome types were highly compacted in the presence of linker histones. However, there was no difference in the extent of compaction between the two types of trinucleosomes.

Overall, this thesis shows a new mode of DNA sequence recognition by the linker histone that may affect the compaction of AT- and A-tract-rich heterochromatin. However, trinucleosome compaction by linker histones was not observed to be affected by the location of A-tracts.

Zusammenfassung

Das Nukleosom setzt sich aus einem Kernkomplex bestehend aus jeweils zwei Kopien der vier Histonproteine zusammen, um die etwa 1,65 Umdrehungen DNS (Desoxyribonukleinsäure, engl. Deoxyribonucleic acid) aufgewickelt sind. Das Linker histone H1 assoziiert an das Nukleosom an der Schnittstelle der, vom Kernkomplex überhängenden DNS. Nukleosom und H1 bilden dabei das Chromatosom, die kleinste, sich wiederholende Einheit des Chromatins. Wie bindet H1 an das Nukleosom? Sind die Abschnitte der DNS, die den Kernkomplex flankieren (Linker-DNS/L-DNS) funktional entscheidend für die Bindung an H1? Beeinflussen Länge oder Sequenz der L-DNS die Art der Bindung von H1 an das Nukleosom? Wenn dem so ist: wie würde die Art der Bindung von H1 sich dann auf die übergeordneten Strukturen des Chromatins auswirken? Die vorliegende Arbeit befasst sich mit der Beantwortung dieser Fragen.

Chromatosomen wurden aus den Kernhistooktameren (aus *Xenopus laevis*), dem Linker Histon (Subtyp: H1.0b (aus *X. laevis*)), sowie einer 226 bp langen DNS rekonstruiert. Das Linkerhiston wurde mit dem Fluorophor Alexa488 an der globulären Domäne oder vor der C-terminalen Armdomäne (engl. C-terminal tail domain; CTD) markiert. Die DNS, welche die Histon-positionierende Widom 601 Sequenz enthält, wurde im Bereich der L-DNS einseitig mit Alexa584 markiert. Einzelpaar-FRET (Förster Resonanzenergietransfer) Spektroskopie wurde verwendet um die Nähe der globulären Domäne, beziehungsweise der CTD zur L-DNS zu bestimmen. Dabei wurde zunächst untersucht, inwiefern Länge und Sequenz der L-DNS die Lokalisierung von H1 am Nukleosom beeinflussen. Es wurde gezeigt, dass die globuläre Domäne von H1 proximal zu einer Ansammlung von 11 konsekutiven Adeninen, des A-Trakts, positioniert. Die globuläre Domäne hat sich jedoch niemals proximal zu einem reinen GC-Abschnitt, oder einer gemischten Sequenz mit 64% AT-Anteil angeordnet. Das Fluorophor der ungeordneten CTD befand sich in gleicher Distanz zu beiden L-DNS-Abschnitten, unabhängig von Sequenzvartiationen. Anhand der mittels Einzelpaar-FRET ermittelten Distanzen wurden Modelle des von A-Trakt enthaltenden Nukleosomen erstellt.

Es wurde beobachtet, dass die globuläre Domäne von H1 sich auf der Dyade am Nukleosom positioniert und, dass zwei konservierte Arginin-Residuen auf der globulären H1-Domäne proximal der kleinen Furche des A-Traktes liegen. Um zu zeigen, dass die Erkennung des A-Traktes nicht über hydrophobe Wechselwirkungen mit den Thymin-Methylgruppen kontrolliert wird, wurde ein Nukleosom mit zwei A-Trakten generiert, die eine zum A-Trakt komplementäres Thymin und ein deoxy-Uridin ohne Methylrest enthielten. Die globuläre Domäne

positioniert sich hierbei nahe den beiden A-Trakten, was die Bedeutung der schmalen Furche des A-Traktes in der Erkennung der globulären Domäne unterstreicht.

Abschließend wurde untersucht, inwiefern die linker histone zwei Arten von trinukleosom-enthaltenden A-Trakten komprimieren, welche das erste und dritte Nukleosom entweder in der Inneren oder Äußeren L-DNS flankieren. Das Ausmaß der Kompression wurde über Einzelpaar-FRET zwischen den beiden inneren L-DNS gemessen. Eine L-DNS verbindet dabei das erste und zweite Nukleosom, die Zweite das zweite und dritte Nukleosom. Basierend auf den Modellen der Einzelnukleosomen, sollten sich die Linker histone des ersten und dritten Nukleosoms in Richtung der A-Trakte an den äußeren oder inneren L-DNS ausrichten.

Beide Arten von Trinukleosom waren in Anwesenheit des Linker histones maximal komprimiert. Es konnte dabei keine Varianz im Ausmaß der Kompression zwischen den Arten des Trinukleosoms quantifiziert werden.

Zusammenfassend zeigt diese Dissertation eines neuen Modus der DNS-Sequenzerkennung durch das Linker histone, der die Komprimierung von AT und A-Trakt-reichem Heterochromatin beeinflusst. Die Komprimierung von Trinukleosomen ist dabei jedoch unabhängig von der Anordnung der A-Trakte.

To Dadu, for teaching me to think fearlessly,
And Duda, for teaching me to imagine fearlessly.

Contents

1		**Introduction**	**13**
	1.1	Chromatin: 'particles on a string'	13
	1.2	Nucleosome & Chromatosome	14
	1.3	The sequence of DNA associated with the nucleosome	17
	1.4	Chromatin packaging: hierarchical folding or not?	18
	1.5	Linker histones and cancer	19
	1.6	Plan of the thesis	20
2		**Using FRET as a molecular ruler**	**22**
	2.1	Förster/Fluorescence Resonance Energy Transfer	22
	2.2	Calculating Förster distance	24
	2.3	Single-pair FRET	28
	2.4	Calculating inter-fluorophore distance	34
3		**Materials and methods: experiment and modeling**	**35**
	3.1	DNA preparation for monochromatosomes	36
	3.2	DNA preparation for trichromatosomes	37
	3.3	Linker histone purification	39
	3.4	Linker histone labelling	44
	3.5	Reconstituting mono and trichromatosomes	46
	3.6	Modeling monochromatosomes	48
	3.7	Computing theoretical inter-fluorophore distances	50
4		**Role of the linker-DNA on the positioning of the linker histone**	**53**
	4.1	Introduction	53
	4.2	Lengthening L-DNA from 24 to 40 bp does not affect the orientation of the LH	56
	4.3	Neither the gH nor the CTD have any preference for AT-rich sequence over GC-tracts	59
	4.4	L-DNA flank sequences do not affect the CTD	60
	4.5	gH shows a higher preference for A-tracts	62
	4.6	Concluding remarks	63
5		**How does the linker histone recognize A-tracts?**	**64**
	5.1	Introduction	64
	5.2	Conserved arginines in H1.0 subtypes recognize A-tracts	65
	5.3	gH shows the same preference for both A-tracts paired to thymines as well as to deoxyuridines	69
	5.4	Concluding remarks	72

6 Effect of A-tract dependent LH orientation on trichromatosome compaction **73**
6.1 Introduction 73
6.2 Linker histones compact A-far and A-near trichromatosomes similarly 75
6.3 Orientation of neighbouring LH does not affect trichromatosome compaction 76
6.4 How does the linker histone aid trichromatosome compaction? 78
6.5 Concluding remarks 80

7 Conclusions & Outlook **81**
7.1 Contribution of L-DNA to the location of LH: Length or sequence? 81
7.2 Direct and indirect recognition of AT-rich DNA by the linker histone 82
7.3 Implications in heterochromatin compaction 84
7.4 Implications in higher-order structure: a preliminary understanding and future perspectives 84
7.5 Opening a Pandora's box 85

Bibliography **87**

Appendices **95**
A Sequences 95
B Reagents and compositions 97
C Instrumentation 101

Supplementary Information **102**
Chapter 2 102
Chapter 3 103
Chapter 4 103
Chapter 5 120
Chapter 6 124

Acknowledgements **133**

Lists of figures and tables

	List of figures	Pg. No
1.1	Illustration of a generic eukaryotic cell.	13
1.2	Chromatosome and linker histone.	15
1.3	Modes of association of the LH and its implication in higher-order structure.	16
1.4	How is the chromatin packaged?	19
1.5	Plan of the thesis	20
2.1	Förster Resonance Energy Transfer.	22
2.2	Calculating spectral overlap.	26
2.3	Donor and Acceptor fluorophores described as dipoles.	27
2.4	SpFRET setup	29
2.5	Overlap detection by single-laser excitation and ALEX-spFRET.	31
2.6	Histogram of proximity ratios fitted to multiple Gaussian peaks.	32
2.7	γ-factor calculation by ALEX-FRET.	34
2.8	Extracting distance from proximity ratio	34
3.1	HPLC purified 226 and 212 bp DNA for monochromatosome reconstitution.	36
3.2	Preparing 600 bp DNA for trichromatosome reconstitution.	37
3.3	Digested and undigested fragments.	39
3.4	Presence of insert in plasmid.	41
3.5	IPTG induction test.	42
3.6	Labelled linker histone.	45
3.7	Workflow for monochromatosome reconstitution by salt dialysis.	46
3.8	Checking reconstitution of mono and trichromatosomes.	48
3.9	Modeling of AG and GA chromatosome ensembles.	49
3.10	Computing theoretical inter-fluorophore distance	50
4.1	Experimental designs for monochromatosome study.	53
4.2	L-DNA length does not affect the gH location.	56
4.3	L-DNA length does not affect the CTD location.	58
4.4	gH and CTD does not prefer GC-tracts over AT-rich mixed DNA sequences.	59
4.5	L-DNA sequence does not affect the CTD location.	60
4.6	gH shows a higher FRET for A-tracts.	62

List of figures

5.1 DNA interacting zones on the globular domain of LH. 64

5.2 AG and GA chromatosomes 65

5.3 AG and GA chromatosomes: alternate models. 67

5.4 Further evidence of minor-groove based recognition of A-tracts. 70

6.1 Experimental design for the trichromatosome study. 74

6.2 Single-pair FRET observations for trichromatosomes. 75

6.3 Modelling the highly compact population II of trichromatosomes 77

6.4 Geometric parameters obtained from the model. 77

6.5 Positioning of the gHs in the stacked nucleosomes 1 and 3 in the trichromatosome models. 79

List of tables

2.1 Absorption and emission maxima for flurorophores 25

2.2 Parameters calculated for R_0 estimation, assuming a κ^2 value of 2/3 28

2.3 R_0 values obtained from literature for comparison 28

3.1 Decription of chromatosome constructs. 35

3.2 Fluorophore parameters used in the FPS software. 51

4.1 Chromatosome construct information chart. 55

4.2 Between the gH and minus and plus L-DNA arms for constructs AsyMG, AsInvMG and MG. 56

4.3 Between the CTD and minus and plus L-DNA arms 57

4.4 Between the gH or CTD and minus and plus L-DNA arms in GM construct. 59

4.5 Between the CTD and minus and plus L-DNA arms 61

4.6 Between the gH and minus and plus L-DNA arms 62

5.1 Comparison of inter-fluorophore distances calculated from proximity ratio plots and models 66

5.2 Distances (in Å) computed from alternate model ensembles (Figure 5.3A and B) not fitting with the experiment 68

5.3 Distances (in Å) computed from alternate off-dyad models 68

5.4 TU construct studied. 70

5.5 Experimental FRET efficiency of construct TU 71

6 Geometric parameters obtained from the model. 78

7 Zones A, B and C of different LH isoforms across organisms. 83

Abbreviations

1xTE	1x 10mM Tris 0.1mM Ethylenediaminetetraacetic acid (EDTA)
AFM	Atomic Force Microscopy
Cryo-EM	cryogenic electron microscopy
CTD	C-terminal domain
EE	Entry-exit site
EtBr	Ethidium Bromide
FPLC	Fast Protein Liquid Chromatography
FPS	FRET Positioning and Screening
FRAP	Fluorescence recovery after photobleaching
FWHM	full width at half maximum
gH	globular domain
HPLC	High Performance Liquid Chromatography
IPTG	Isopropyl b-D-1-thiogalactopyranoside
L-DNA	Linker DNA
LH	Linker histone
Ncp	Nucleosome core particle
NMR	Nuclear Magnetic Resonance spectroscopy
NTD	N-terminal domain
OD	Optical density
PAGE	poly-acrylamide gel electrophoresis
PMSF	Phenylmethanesulfonylfluoride
SDS	Sodium dodecyl-sulfate
Sp-FRET	Single pair – Fluorescence Resonance Energy Transfer
TCEP	Tris (2-carboxyethyl) phosphine
TEMED	Tetramethylenediamine
XRC	X-Ray Crystallography

“And now, Harry, let us step out into the night and pursue that flighty temptress, adventure.”
– Albus Dumbledore, Harry Potter and the Half-Blood Prince

Chapter 1

Introduction

1.1 Chromatin: 'particles on a string'

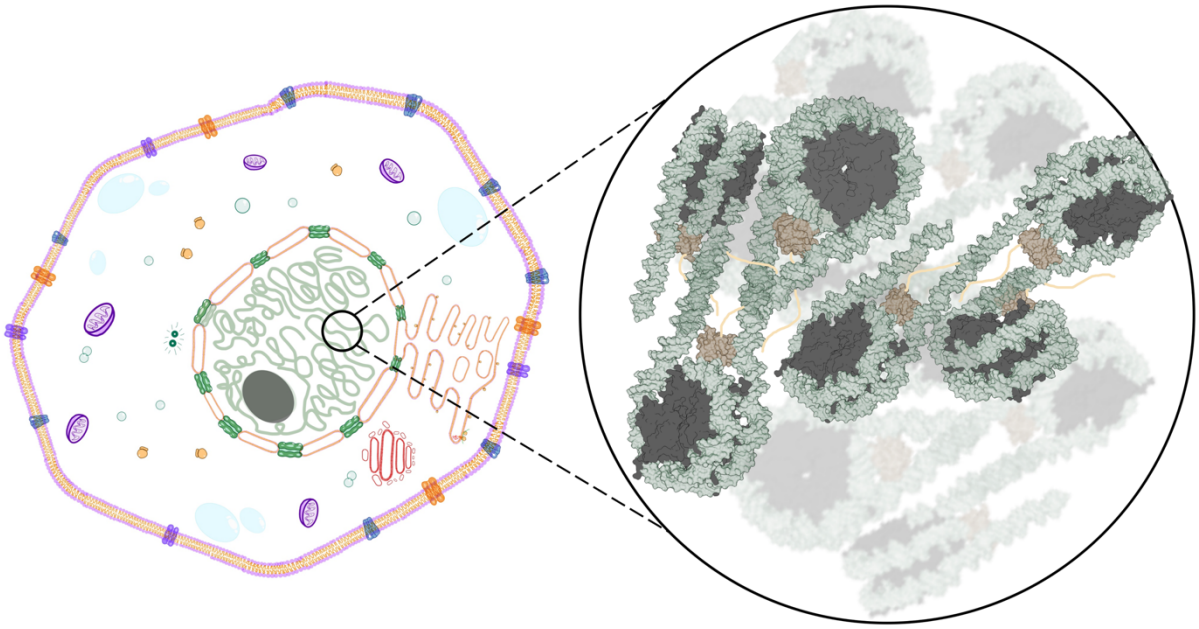


Figure 1.1: *Illustration of a generic eukaryotic cell. The chromatin (in pale-green) is in the nucleus. The close-up reveals that the chromatin has a 'bead on a string' structure.*

In the 17th century an entire world that was previously unseen came into view with the invention of the microscope. While studying cork slices using his compound microscope, Robert Hooke (1635-1703) observed a 'honey-comb'-like structure that he called 'cells' (1). Around the same time, Antonie van Leeuwenhoek (1632-1723) observed single-celled organisms, that he called 'animalcules'. The following century marked some of the most important landmarks leading to the foundations of cell biology. In 1831, Robert Brown (1773-1858), studying orchid tissues, observed and named the 'nucleus' of the cell (for a historical overview see reference (2)). The race began to understand its contents.

Nearly two hundred years later the precise arrangement of the contents of the cell nucleus, or the lack thereof, remains a puzzle to this day.

In the late 19th century, Walther Flemming observed 'fibrous' structures within the cell nucleus that could be stained by aniline dyes (3–6). He termed them 'chromatin' or stainable. In

the next few years he observed 'threadlike metamorphosis of the nucleus' during cell division (6), coined the term, and formally described 'mitosis'.

A hundred years later, Donald and Ada Olins ruptured the nuclei of rat thymus cells and chicken erythrocytes and stained the chromatin spilling out from it using phosphotungstic acid (7). Using electron microscopy, they observed structures resembling strings of pearls, calling them 'beads' or 'particles on a string'. They coined the term 'ν bodies' to describe these beads and estimated their diameters to be about 6 to 8 nm. In 1975, Oudet showed that these beads are repeating units and likely the monomers of the chromatin, and coined the term 'nucleosome' to describe the bead (8).

1.2 Nucleosome & Chromatosome

The chemical composition of the chromatin garnered a great deal of attention since the beginning of the 20th century. With the development and improvement of protein purification techniques, highly positively charged histone proteins were purified and their fractions estimated by numerous groups from the 1920s (9) right up to the late 60s (9–13).

In addition to the four types of histone fractions purified, an additional lysine-rich fraction was obtained (9, 11) as well. Termed as the 'fraction 1', or F1 (14), this protein fraction was found to comprise about one fifth of the total histone content of the cell (15). An important observation noted by Mirsky *et al.* (1968) (16) showed that this lysine-rich histone cross-linked chromatin fibers during metaphase and interphase, to condense the chromatin.

How do all the histone proteins and DNA come together to build the bead on the string?

In the mid-70s, Roger Kornberg (17, 18) described some possible models of the ultrastructure of the nucleosome, based on results obtained from sedimentation coefficient studies. He proposed that the nucleosomes are composed of two copies of four histone proteins, forming an octamer, and that the chromatin comprises of a chain of these octamers. Digestion by micrococcal nuclease enabled isolation of single nucleosomes for structural studies (19, 20). By performing micrococcal nuclease digestion assays, Simpson noted (21) that one F1 protein associated with 166 bp of DNA, and the histone octamer. He coined the term 'chromatosome' to describe this particle. In the same work, he reported that the DNA ends diverged in the absence of F1 and proposed that F1 associates on the outside of the nucleosome. Further work confirmed the stoichiometries of the various histone proteins (named H2A, H2B, H3, H4, and H1) in the chromatosome (22). The first x-ray crystallographic measurements of the nucleosome (at 7 Å resolution from Richmond *et al.* (1984) (23) and at 2.8 Å from Luger *et al.* (1997) (24)) ultimately solved its structure. This structure showed that two copies each of the histones H2A, H2B, H3 and H4 formed the core octamer, around which about 146 bp of DNA wrapped in 1.65 turns of left-handed superhelix. Also called nucleosome core particle (ncp), its diameter was estimated to be 11 nm. In this thesis, the word chromatosome will be used to mean nucleosome associated with the LH. In case of absence of the LH, it will be referred to as the nucleosome (e.g., LH associated with the nucleosome).

But how does the lysine-rich F1, or H1 (linker histone (LH)) associate?

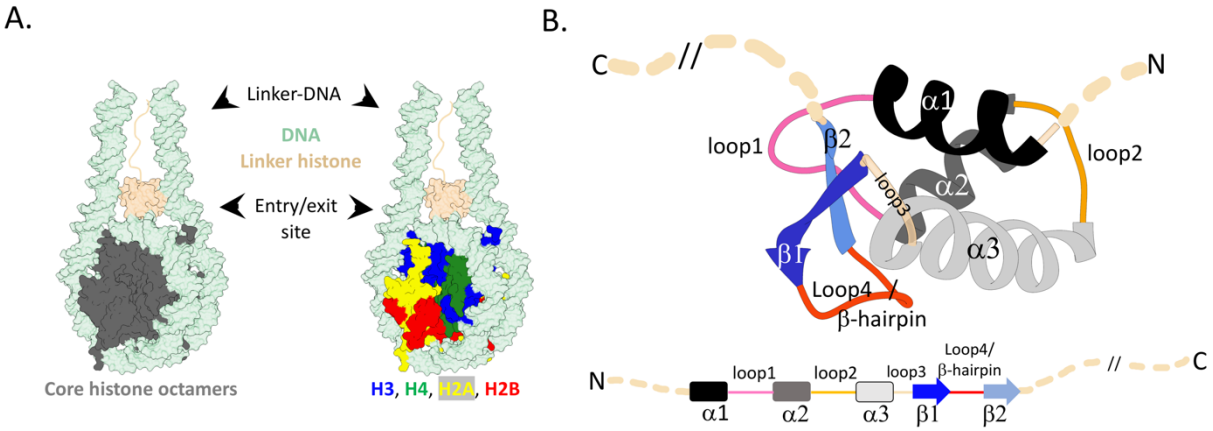


Figure 1.2: *Chromatosome and linker histone.* ***A.*** *The chromatosome. The DNA is shown in pale green. The grey core histone octamer (left) is comprised (right) of the core histones H3 (blue), H4 (green), H2A (yellow), H2B (red). Linker histone (LH) is shown in beige. Globular domain is represented as a surface.* ***B.*** *The linker histone. The globular domain is shown in ribbon representation (top) and in a linear representation (bottom). Three alpha helices are marked as α1 (black), α2 (dark grey), and α3 (light grey). Two beta strands are depicted as β1 (dark blue) and β2 (light blue). There are 4 'loops' connecting the helices and sheets: 1 (pink), 2 (orange), 3 (beige), 4 (red). Loop 4 is a β-hairpin motif connecting the two β-sheets. The intrinsically disordered N and C-terminal domains (NTD and CTD) are denoted as dotted beige lines. // denotes stretches of the CTD not shown in the picture.*

The structure of the linker histone (LH), or H1 (the avian variant is called H5 in case of *Gallus gallus*), in isolation (25) shows a structured globular (gH) domain having a conserved sequence of about 80 amino acids. It is a winged-helix comprising of three alpha helical stretches connected by loops (Figure 1.2B), followed by two beta-strands, connected by a turn, forming the beta-hairpin motif. This crystal structure lacks about 50% of the LH which are the unstructured N-terminal domain of 20 amino acids and the unstructured C-terminal domain of ca. 100 amino acids (Figure 1.2B). Both Simpson's work (21) and subsequent cryo-Electron Microscopic (cryo-EM) studies (26) suggested that the LH associating on the outside of the nucleosome, brought the two incoming and outgoing DNA strands (termed as linker-DNA or L-DNA) close together by charge neutralisation, forming a stem-like structure.

Since X-ray crystallography is unable to capture intrinsically disordered regions, the first structures of the chromatosome involved only the gH domain (27, 28) of the LH. Very recently, by complementing X-ray crystallographic measurements with cryo-EM, the CTD was observed to associate with either one or both L-DNA (29, 30).

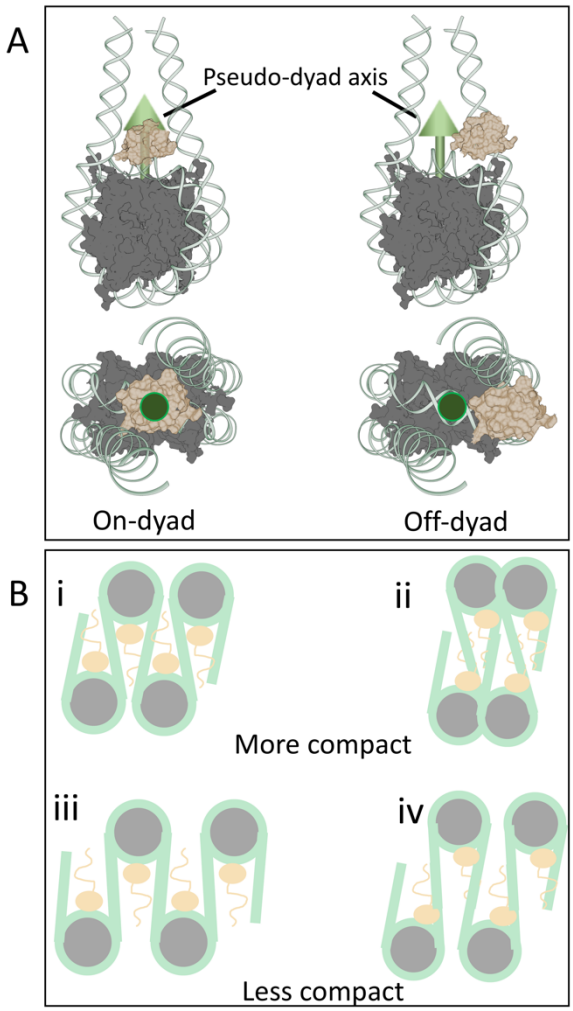


Figure 1.3: *Modes of association of the LH and its implication in higher-order structure.* ***A.*** *Schematic of on-dyad and off-dyad association of the gH domain of the LH on the nucleosome. Core histones denoted in grey surface representation. LH represented as beige surface. DNA in pale green ribbon. Green arrow denotes the pseudo-dyad axis.* ***B. (i)*** *and* ***(iii)*** *shows on-dyad association and* ***(ii)*** *and* ***(iv)*** *off-dyad. Top: How either on-dyad mode* ***(i)*** *or off-dyad mode* ***(ii)*** *can increase compaction in higher-order structures. Bottom: How either on-dyad mode* ***(iii)*** *or off-dyad mode* ***(iv)*** *can decrease compaction in higher-order structures.*

Structures of the chromatosome show that the gH domain is associated with the DNA wrapped around the core octamer at a region bound by the incoming and outgoing Linker-DNA (L-DNA) (Figure 1.2A). If the gH is associated with the pseudo-dyad axis passing through the center of this stretch of DNA, the gH is said to be associated 'on-dyad' (27–30) (Figure 1.3A left). This mode of association has been observed in X-ray crystal and cryo-EM structures of mononucleosomes. If the gH is associated with the DNA between this dyad axis and one of the L-DNA arms, the mode of association is termed as 'off-dyad' (Figure 1.3A right). This mode of association has been observed by employing Nuclear Magnetic Resonance spectroscopy on mononucleosomes associated with *Drosophila melanogaster* H1 (31). This mode of association has also been suggested in chromatosomal arrays observed using cryo-EM (32). There is more than one isoform of the LH in every organism, each differing slightly from the other in its amino acid sequences. Experiments using LHs of two different isoforms from two different organisms (*Gallus*

gallus H5 and *Drosophila melanogaster* H1) suggested that these two modes of association of the gH domain are driven by its amino acid sequence (31).

The two modes of association of the LH have been suggested to play an important role in the compaction of the chromatin at larger scales (Figure 1.3B). Zhou *et al.* (2015) (27) used analytical ultracentrifugation measurements to suggest that an on-dyad bound LH would lead to a higher compaction (Figure 1.3B (i)) and off-dyad bound LH would lead to a lower compaction of chromatosome arrays (Figure 1.3B (iv)). This suggestion is in contrast with the suggestion from Perišić *et al.* (2019) (33) whose mesoscale modeling suggested that off-dyad bound LH would result in the highest compaction of chromatosome arrays (Figure 1.3B (ii)). Evidence for the latter proposition, i.e., an off-dyad bound LH leads to maximum compaction of the chromatin, also came from cryo-EM studies (32). These contradictory observations make it crucial to study how the positioning of the LH could affect the compaction of the chromatin.

As mentioned before, certain amino acid residues on the gH domain determine how the gH will associate with the nucleosome. What about the sequence of DNA associated with the nucleosome? Do certain DNA sequences play any role in the positioning of the LH?

1.3 The sequence of DNA associated with the nucleosome

The thirteen year long, multi-national Human Genome Project resulted in the sequencing of the entire human chromatin. This led to the observation that the ca. 30,000 protein coding genes in the human genome (34) make up only up to 1 to 2% of the total DNA present in the chromatin, suggesting that apart from encoding proteins, the DNA sequence also has other functions.

The sequence of DNA affect its physical parameters such as the pitch (35), minor and major groove widths (36, 37), the bendability and flexibility of the DNA (38, 39). This in turn affects biologically relevant protein-DNA interactions (37, 40). In chapters 4 and 5 of the thesis, we will focus mainly on the sequence-dependent structural properties of the DNA and how the LH can exploit this to recognize certain DNA sequences.

To form a nucleosome the DNA has to curve around the core octameric histone proteins. The capacity of a strand of DNA to bend depends upon its sequence. GC-rich DNA can curve more than the more rigid AT-rich DNA (41, 42), making GC-rich favorable for incorporation into the NCP. In contrast, rigid A-tracts (series of adenines (43)) are excluded from the core or NCP. In contrast to this, the LH has been observed to have a preference for AT-rich DNA and A-tracts in particular (44–50), and no preference for GC-rich DNA (30, 47, 51). Is this an adaptation to enable nucleosome stability in GC-rich regions (52)? Entry-exit sites (EE site) bounding the NCP (Figure 1.2A) show a high AT-content (49). Does this prevent nucleosome sliding and additionally promote LH binding as proposed by Cui and Zhurkin (2009) (49). In this thesis, we have explored how DNA sequences, specifically A-tracts on the L-DNA may affect the positioning of the LH. To do this we used a strongly positioning DNA sequence: the 147 bp 'Widom 601' sequence (53). This sequence strongly prevents the sliding of the core histone octamer, and is thus widely used in structural studies (29, 30). We have modified the L-DNA arm right above the entry-exit site to include AT-rich DNA or A-tracts to observe its effect on the LH.

But how is the DNA, amounting to 2 meters per cell, is packaged?

1.4 Chromatin packaging: hierarchical folding or not?

The remarkable feat of packing a 2 meters (54) long DNA into a volume a million times smaller is achieved by neutralising the negative charges on the DNA backbone by the positively charged histone proteins. But how exactly this packaging takes place is a matter of much debate. Although the work presented in this thesis mainly concerns studies on a single chromatosome, Chapter 6 describes a study on trichromatosomes where we have modeled a zig-zag arrangement of the three chromatosomes. Therefore, I will briefly describe the study of chromatin packaging in this section.

In the late 70s, Finch and Klug (55) obtained electron micrographs of the chromatin extracted from cell nuclei. They observed a condensed filamentous structure with a diameter of 30 nm. They also observed that in the absence of LH, the condensed 30 nm fibre unspooled into an extended chain-like form (55, 56). They proposed that this 30 nm nucleofilament was composed of nucleosomes arranged in a 1-start solenoid and that the central bore of this solenoid was lined with LHs (Figure 1.4A top). Evidence for this one-start solenoidal folding also came from electron microscopic (57) and force spectroscopic studies (58) on reconstituted nucleosomal arrays.

A second type of higher-order folding was suggested in which nucleosomes adjacent to each other when arranged linearly were opposite to each other when arranged around a central bore. This was the two-start or zig-zag model of the 30 nm fibre (Figure 1.4A bottom). This type of higher-order folding was first suggested from cryo-EM images of purified chromatin obtained by Bednar *et al.* (1998) (26) and in in silico modeling studies (59, 60). The crystal structure of an in vitro reconstituted tetranucleosome (61) showed for the first time a zig-zag arrangement of the four nucleosomes. Further evidences of zig-zag folding or a 2-start arrangement also came from cryo-EM studies (32, 62).

So far, the 30 nm fibre has only been observed in vitro either in chromatin extracted or purified from cell nuclei or in reconstituted chromatosomal and nucleosomal arrays. Although in vivo fragmentation studies suggested 2-start folding of the chromatin in vivo (63, 64), further evidence of hierarchical folding of the chromatin in vivo is still largely lacking (65, 66).

How is the chromatin arranged in vivo? It was suggested by Grosberg *et al.* (1993) (67) that the chromatin existed as a fractal globule in the nucleus. Very nicely reviewed by Leonid Mirny (2011) (68), this model is described as a 'hierarchy of crumples' (69). In vivo imaging studies (70) and Hi-C mapping (71, 72) have shown that the chromatin indeed shows a fractal form.

A highly advanced electron microscopy tomography method (73) was used to obtain images of the nucleus in vivo. It did not reveal any 30 nm fibre. The chromatin was observed to be disordered with variably dense regions: likely the transcriptionally inactive, denser heterochromatin, and the transcriptionally active euchromatin of less density.

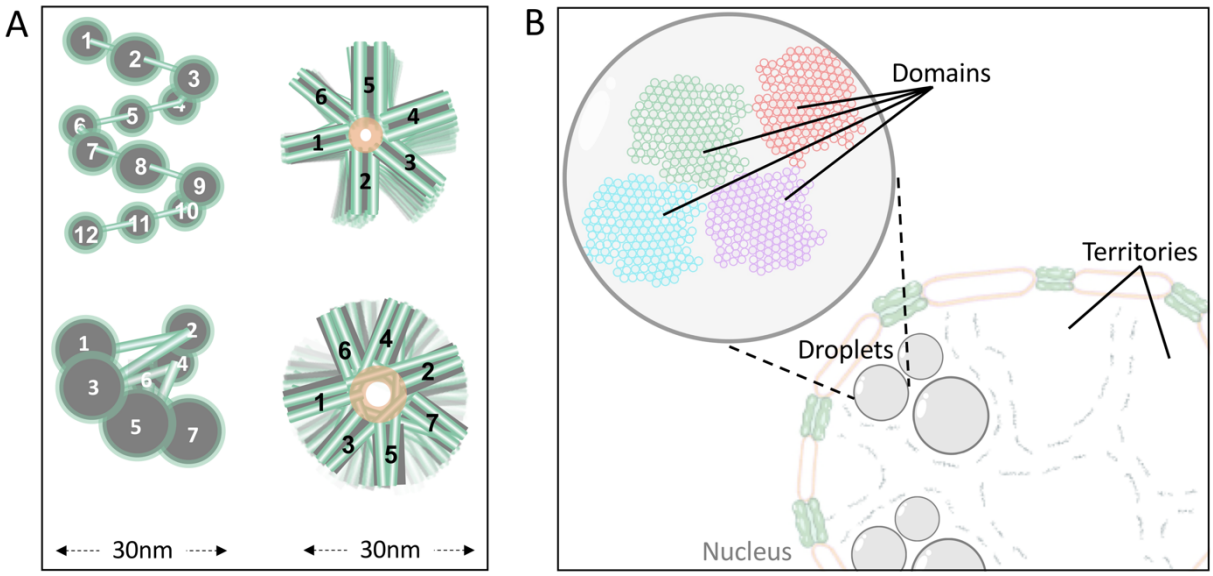


Figure 1.4: *How is the chromatin packaged?* ***A.*** *Hierarchical packaging of the chromatin into 30nm fibre: Top: one-start solenoidal model, Bottom: two-start zig-zag model. Left shows side views and right shows top views.* ***B.*** *Clustering of stretches of nucleosomes (coloured circles) into domains, domains into phase separated droplets, and finally into chromosome territories in the nucleus.*

Recently, the possible role of phase separation in chromatin organisation (74–78) (Figure 1.4B) has taken the chromatin world by storm. The underlying principle of this phenomenon is that chromatin binding proteins of different types can induce segregation of parts of the chromatin they are acting upon into phase separated droplets. Thus, by this model, proteins acting on the chromatin drive its compartmentalization. The LH has been observed to promote (77, 78) phase separation of the chromatin both in vivo and in vitro. L-DNA lengths have also been implicated in regulating the properties of phase separated droplets (77).

1.5 Linker histones and cancer

The LH has been implicated in numerous diseases including cancer, Alzheimer's disease (79) and aberrant immune response (reviewed in Ye *et al* (2017) (80)). However, the role of the LH in the pathology of any disease is not universal but differs between different subtypes of the LH (80). Different disease types that involve the LH are either caused by mutations on the LH or an aberrant expression that leads to low or high levels of the LH in the diseased cell (80). The cellular levels of LH often serve as biomarkers for cancer prognosis (81–84).

A recent work by Torres *et al.* (2016) (85) explored a possible mechanism by which low levels of LH of the subtype H1.0 mediate cancer progression. In this work they show that an absence of the H1.0 LH selectively destabilises nucleosomes present in AT-rich regions of the genome, but not the ones present in the GC-rich regions of the genome. This is because of the inability of AT-rich DNA to bend and wrap around the NCP resulting in the instability of nucleosomes found in AT-rich regions of the chromatin, in the absence of H1.0. In the absence of nucleosomes and H1.0 to silence these portions of the chromatin, oncogenes that are present in these regions are upregulated. This work clearly showed how the interplay between AT-rich DNA

and H1.0 can result in cancer progression.

Their observation that LHs of H1.0 subtype can selectively stabilise nucleosomes in AT-rich regions is of no less importance. This sheds some light on the biological function of H1.0. How does the LHs recognize AT-rich DNA, to execute this function? This question is dealt with in this thesis (Chapter 4 and 5).

1.6 Plan of the thesis

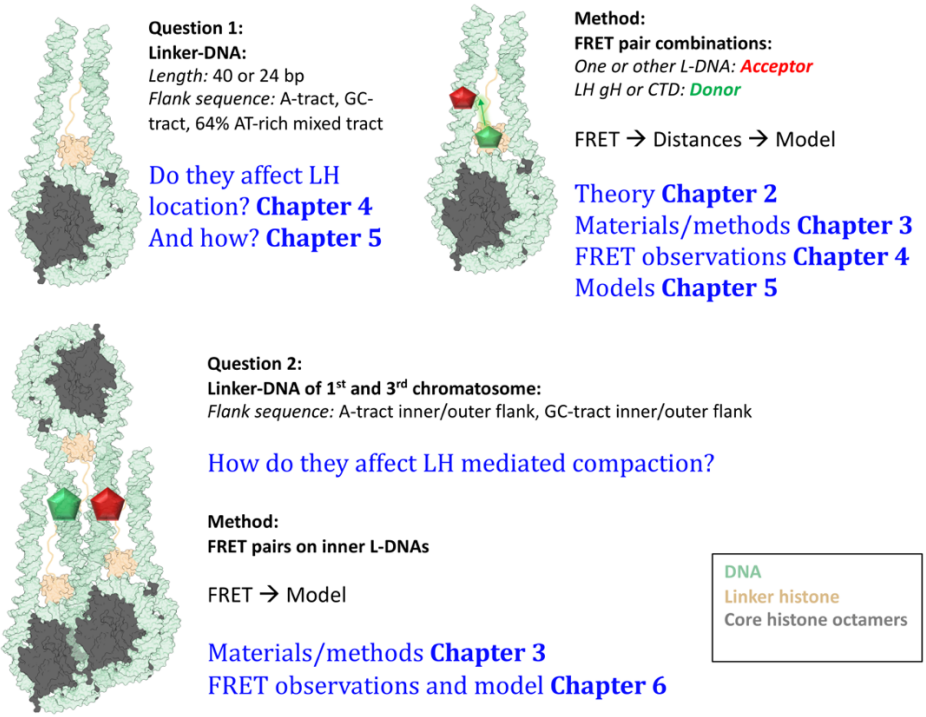


Figure 1.5: *Plan of the thesis.*

The main aims of the thesis (Figure 1.5):

a. To observe possible effects of certain L-DNA sequences, specifically A-tracts, on the gH and the C-terminal tail domain of the LH.
b. To find out how LHs can recognize A-tracts, and
c. To find out if different placements of A-tracts in trichromatosomal arrays affect its compaction by the LH.

For these purpose, LH of subtype H1.0 (*Xenopus laevis* H1.0b) had been used. Using single-pair Förster/Fluorescence Resonance Energy Transfer, the effect of pure GC-tracts, A-tracts (continuous stretch of adenines), and mixed sequences having 64% AT-content, on the positioning of the H1.0b on mononucleosomes. Computational modeling based on the FRET results helped us uncover a mechanism by which LHs can recognize A-tracts. It also showed how A-tracts present on the L-DNA flanking the globular domain of the LH can reorient the globular domain. But how does this A-tract induced gH domain reorientation affect higher-order structures? A preliminary study was designed to this end. Single-pair FRET measurements were performed on trichromatosomes labelled on the two inner L-DNAs. These trichromatosomes were designed to contain A-tracts either on the outer or inner L-DNA arms. The compaction of the two types of trichromatosomes were measured in the presence of LHs. Single-pair FRET and modeling showed that despite orientational effects of the LH induced by A-tracts, these structural changes do not affect the compaction of the trichromatosomes. Chapter 7 deals with the conclusion and future perspectives of the thesis.

Chapter 2

Using FRET as a molecular ruler

2.1 Förster/Fluorescence Resonance Energy Transfer

In the early 1920s, the Perrins observed a decrease in the polarization of fluorescent molecules in solution. This led them to formulate a model (86) where the fluorophores were treated as oscillating electric dipoles. According to this model, if the fluorophores were close enough, the energy could be transferred from one fluorophore to the other non-radiatively. Building upon the works of the Perrins, Theodor Förster (87) gave the first mathematical description of this non-radiative energy transfer between closely spaced fluorescent molecules in solution.

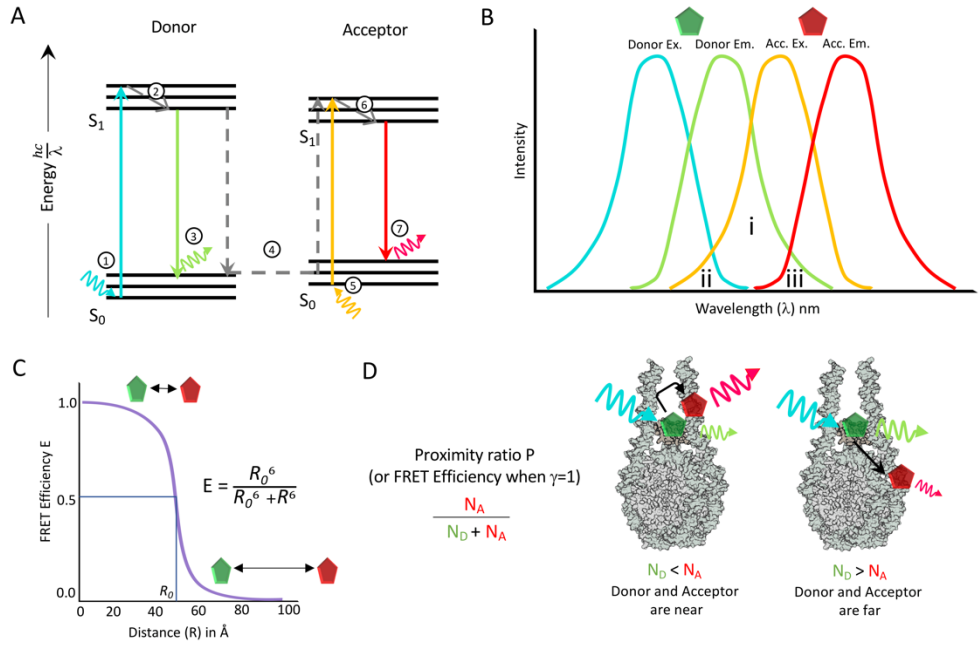


Figure 2.1: *Förster Resonance Energy Transfer.* ***A.*** *Jablonski diagram showing absorption (1 and 5), emission (3 and 7) and FRET (4) between donor and acceptor fluorophores.* ***B.*** *Absorption and emission spectra of donor (blue and green plots) and acceptor (orange and red plots) showing three types of spectral overlaps (i-iii) explained in the text.* ***C.*** *Relationship between FRET efficiency and distance* ***D.*** *How FRET can be used as a molecular ruler to measure distance between fluorophores attached to a macromolecular complex. E = FRET Efficiency, R_0 = Förster distance, R = inter-fluorophore distance, P = Proximity ratio, γ = gamma factor or detection factor, N_D = Number of photons of donor fluorophore (represented by green pentagon) and N_A = number of photons of acceptor fluorophore (represented by red pentagon).*

This non-radiative, distance dependent energy transfer between two fluorophores came to be termed as Fluorescence or Förster Resonance Energy Transfer (FRET). The distance dependence of this phenomenon has made it a widely used tool in structural biology (88).

The phenomenon of absorption and emission of radiation and non-radiative energy transfer can be described by a standard Jablonski diagram (88) as shown in Figure 2.1A. The singlet ground state levels of the donor and the acceptor fluorophores have a lower energy and are denoted as S_0, and the higher energy singlet excited states of the fluorophores are denoted as S_1. When a photon of higher energy and lower wavelength (depicted by blue-green wavy arrow, Figure 2.1A) strikes the donor fluorophore, originally in the ground state, the fluorophore absorbs the photon and begins to oscillate with a higher energy (event ①, Figure 2.1A), thereby getting 'elevated' to the S_1 level. Independently, the acceptor fluorophore also can absorb a photon (event ⑤, Figure 2.1A) having a lesser energy than the blue-green photon used to excite the donor fluorophore. Fluorophores in the excited singlet S_1 level undergo vibrational relaxation and lose energy non-radiatively (events ② for donor fluorophore and ⑥ for acceptor, Figure 2.1A). When the fluorophores de-excite to their ground singlet state, they emit a photon that has a lower energy, and higher wavelength than that absorbed by that fluorophore (event ③ green arrow for donor and event ⑦ and red arrow for acceptor, Figure 2.1A). This photon emission upon transitioning back to the ground singlet state from the excited singlet state is fluorescence and its lifetime is in the order of 10ns (88).

In Figure 2.1B, the excitation (or absorption) and emission spectra of the donor and the acceptor fluorophores are illustrated. The photon absorbed by the donor has the lowest wavelength (blue-green peak) and the highest energy. The wavelengths of the photons emitted by the donor, absorbed by the acceptor and emitted by the acceptor progressively increases, and the energy decreases. The emission spectra of the donor (green, Figure 2.1B) sufficiently overlaps with the acceptor absorption spectra (orange), with this overlap region denoted by the number (i) in Figure 2.1B. This overlap between the donor emission and the acceptor absorption is essential for resonance energy transfer to take place when the donor and the acceptor are separated by a small distance. Event ④ on Figure 2.1A depicts this non-radiative transfer of energy by dipole-dipole coupling from the donor fluorophore oscillating in the excited singlet state to the acceptor fluorophore at the ground state. This elevates the acceptor fluorophore to its excited singlet state. The subsequent de-excitation of the acceptor back to its ground state occur radiatively by emission of the red photon (event ⑦).

The efficiency of the energy transfer is inversely proportional to the sixth-power of the distance between the two fluorophores. As shown in Figure 2.1C, smaller the inter-fluorophore distance R, higher the FRET efficiency and vice-versa. At a certain distance R_0, known as the Förster distance, the transfer efficiency is half of its maximum. The relationship between the transfer efficiency E and the inter-fluorophore distance R is given by the formula:

$$E = \frac{R_0^6}{R_0^6 + R^6} \qquad \text{(a)}$$

The Förster distance for standard FRET (donor-acceptor) pairs fall in the range between 30 to 60Å, making it highly useful for measuring intra-molecular distances in biological

macromolecular complexes (88). The efficacy of FRET to measure distances in biological molecules was first demonstrated on poly-L-proline chains by Stryer and Haugland (1967) (89) who called this technique a 'spectroscopic ruler'. Complementing high resolution techniques to study macromolecular structures, such as X-ray crystallography, FRET quickly gained popularity due to its suitability to studying macromolecular dynamics.

The nucleosome has a diameter of 110Å. By placing suitable FRET pairs on different parts of the nucleosome, as illustrated in Figure 2.1D, distances can be efficiently measured with a very high resolution. FRET began to be widely used in the study of nucleosome structure and dynamics since the early 2000s (90–94).

The term proximity ratio is introduced in the section 'Single-pair FRET' later in this chapter. Two other spectral overlaps shown in Figure 2.1B (ii and iii) are also explained in that section.

2.2 Calculating Förster distance

The Förster distance for a standard FRET (donor-acceptor) pair is known for fluorophores in water. However, in protein and DNA containing samples samples, different buffers and presence of proteins and DNA in the solution change the refractive index of the solvent and the quantum yield of the fluorophore. This can significantly change the R_0, thus warranting its measurement in the sample under study (95).

The R_0 of a FRET pair is calculated via a multi-step procedure that involved bulk measurements of absorption and fluorescence. The formula used to calculate Förster distance (in Å) of a FRET pair is (96):

$$R_0 = 0.211\left(\kappa^2 . n^{-4} . Q_{DO} . J(\lambda)\right)^{1/6} \quad \text{(b)}$$

Here, κ^2 is the dipole orientation factor, n is the refractive index of the solvent, Q_{DO} is the quantum yield of donor fluorophore (in our experiments, Alexa 488), in samples containing only the donor fluorophore, also referred to as 'donor-only' samples. $J(\lambda)$ refers to the overlap integral calculated from the spectral overlap region (Figure 2.1B overlap i) between the emission of the donor fluorophore in a donor-only sample and the absorption of the acceptor fluorophore (Alexa 594 in our experiments) in an acceptor-only sample (97). Each of the parameters of equation (b) are separately described below.

The refractive index of the sample buffer is assumed to be 1.4 (88). For calculating each of the parameters, bulk absorption and fluorescent spectroscopy were performed. The instruments for measuring absorption and fluorescence are detailed in Appendix C.

2.2.1 Quantum yield estimation

To obtain the Förster distance of the Alexa 488-Alexa 594 fluorophore pair in the samples, the quantum yield of Alexa 488 or donor fluorophore (Q_{DO}) was measured in a chromatosome containing only Alexa 488 label (donor-only).

The quantum yield of a fluorophore is defined as the ratio of the number of photons emitted by the fluorophore to the number of photons absorbed (88). In order to calculate the quantum yield of a fluorophore, the absorption and emission spectra of a reference fluorophore are measured with the same instrument setting. The reference is chosen so as to have a known quantum yield, and the same absorption and emission maxima as the fluorophore under study. The equation for measuring the quantum yield of Alexa 488 in the donor-only sample is given below:

$$Q_{DO} = Q_{ref} \frac{{n_{DO}}^2}{{n_{ref}}^2} \frac{F_{DO}}{OD_{DO}} \frac{OD_{ref}}{F_{ref}} \qquad (c)$$

Here Q_{DO} is the quantum yield of Alexa 488 in the donor-only sample, Q_{ref} is the quantum yield of the reference, which in this case was fluorescein isothiocyanate (FITC) in 0.1M NaOH. The Q_{ref} is known to be 0.95 (88). The refractive index of the sample buffer (n_{DO}) was assumed to be 1.4 (88), while that of 0.1M NaOH (n_{ref}) was assumed to be 1.33 (98). F_{DO} is the integrated fluorescence intensity of Alexa 488 and OD_{DO} referred to the absorption maxima of Alexa 488. Similarly, the integrated fluorescence integrity and the absorption maxima for the reference are denoted by F_{ref} and OD_{ref}.

A wavelength range from 220 to 750 nm (UV-visible) was used to obtain absorption spectra of Alexa 488 and fluorescein isothiocyanate. For obtaining emission spectra, a wavelength range of 500 to 750 nm was scanned.

Table 2.1: *Absorption and emission maxima for flurorophores*

	Absorption maxima (nm)	Emission maxima (nm)
Donor: Alexa 488	494	517
FITC (reference for donor)	495	518
Acceptor: Alexa 594	590	617

2.2.2. Spectral overlap estimation

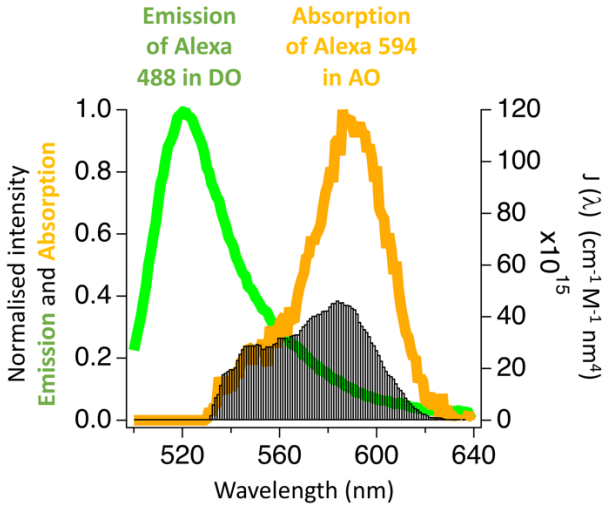


Figure 2.2: *Calculating spectral overlap. Plot showing fluorescence emission spectra of Alexa 488 in a donor–only sample (green trace), the absorption spectra of Alexa 594 in an acceptor-only sample (orange trace), and the spectral overlap integral J (λ) in black bars. The emission and absorption spectra are height-normalised. The left axis denotes the height normalised intensities of emission and absorbance. The right axis denotes the overlap integral.*

The spectral overlap between the donor emission and the acceptor absorption (Figure 2.1B overlap (i)) is what determines resonance energy transfer. The spectral overlap integral J(λ) is calculated as follows (88):

$$J(\lambda) = \frac{\int_0^\infty F_{DO}(\lambda).\varepsilon_{AO}(\lambda).\lambda^4 d\lambda}{\int_0^\infty F_{DO}(\lambda) d\lambda} \qquad \text{(d)}$$

In this equation $F_{DO}(\lambda)$ refers to the fluorescence intensity of the donor fluorophore in the donor only chromatosome (donor label on the LH) in the wavelength range 500 to 640 nm, ε_{AO} refers to the molar extinction coefficient of the acceptor in the acceptor-only chromatosome (97) in the same wavelength range of 500 to 640 nm. ε or molar extinction coefficient is related to the absorbance (OD) by the following equation: OD = ε x molar concentration x path length (cm). Its unit is $M^{-1}cm^{-1}$. The wavelength λ has the unit of nm. F_{DO} has no unit. The unit of J (λ) or overlap integral is $cm^{-1}M^{-1}nm^4$. The integral of the product of fluorescence intensity, molar extinction coefficient and wavelength in the range of 500 to 640 nm (equation (d) numerator) is divided by the integrated fluorescence intensity of the donor in that wavelength range (equation (d) denominator). This is to normalise the area of the emission or fluorescence spectrum to unity (88).

2.2.3 The dipole orientation factor κ²

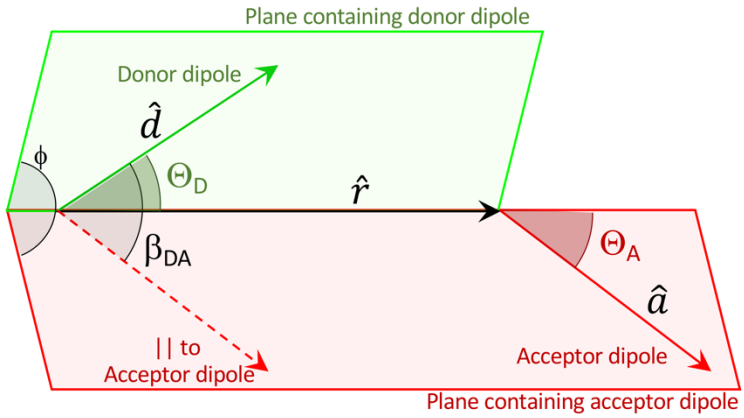


Figure 2.3: *Donor and Acceptor fluorophores described as dipoles. The unit vectors $\hat{d}$ and $\hat{a}$ represent the donor and acceptor dipoles. $\hat{r}$ is the unit vector in the direction from donor to acceptor. Planes containing the vectors $\hat{d}$ and $\hat{a}$ are denoted as donor and acceptor plane respectively. ϕ = angle between donor and acceptor plane, Θ_D = angle between $\hat{d}$ and $\hat{r}$, Θ_A = angle between $\hat{a}$ and $\hat{r}$, β_{DA} = angle between $\hat{d}$ and $\hat{a}$. Illustration adapted from references (96, 99).*

The dipole orientation factor κ^2 in equation (b) describes the relative orientation of the donor dipole moment with respect to the acceptor dipole moment (88, 99). Its value ranges from 0 to 4, with 0 denoting randomly oriented dipoles, 1 denoting dipoles that are parallel to each other, and 4 denoting dipoles that are aligned in a head to tail orientation (88). The κ^2 is defined by the following equation (96, 99, 100):

$$\kappa^2 = (\cos\beta_{DA} - 3\cos\Theta_D\cos\Theta_A)^2 \qquad \text{(e)}$$

with

$$\cos\beta_{DA} = \cos\Theta_A\cos\Theta_D + \sin\Theta_A\sin\Theta_D\cos\phi_{DA} \qquad \text{(f)}$$

The angles in the equations (e) and (f) are explained graphically in Figure 2.3.

To obtain the R_0 and consequently, inter-fluorophore distances (R), the value of the dipole orientation factor was assumed to be 2/3. This average value of the dipole orientation factor denotes that the orientations of the donor and acceptor dipoles become random due to rotational diffusion before energy transfer can occur. This value is standardly used in the calculation of the Förster distance (88, 101).

2.2.4 Obtaining the Förster distance

After each parameter of equation (b) was measured and calculated separately, the Förster distance R_0 for the Alexa 488-Alexa 594 fluorophore pair was calculated for the AG and the GA chromatosomes. The κ^2 was assumed to be 2/3 and the refractive index, 1.4 (96). These values are tabulated below:

Table 2.2: *Parameters calculated for R_0 estimation, assuming a κ^2 value of 2/3*

Construct	**Labeling specifications**	**QY_{DO}**	**κ^2** (assumed)	**J (λ)** ($cm^{-1}M^{-1}nm^4$)	**Refractive index** (assumed)	**R_0 (Å)**
AG	Alexa 488–C5-maleimide (LH T77C) Alexa 594–C6-dT (L-DNA at +94)	0.88	2/3	1.46×10^{15}	1.4	52.0
GA	Alexa 488–C5-maleimide (LH T77C) Alexa 594–C6-dT (L-DNA at +94)	0.88	2/3	1.45×10^{15}	1.4	52.0

The R_0 of the FRET pairs on the chromatosomes were calculated to be 52Å. This value was compared to two other values of Förster distance for Alexa 488-594 pair obtained by Cristóvão *et al.* (2012) (102) and by Reinartz *et al.* (2018) (103) (see Table 2.3). The apparent discrepancy between the R_0 value used in this thesis and that obtained from these works is caused by the assumption of the refractive indices. Reconverting the reported R_0 values for a refractive index of 1.4 resulted in values close to 52Å.

Table 2.3: *R_0 values obtained from literature for comparison*

Article	**Sample type**	**solvent**	**FRET pair**	**QY_{DO}**	**κ^2**	**Refractive index (RI)**	**R_0 (Å) Reported by the authors**	**R_0 (Å) converted (for RI=1.4)**
Cristóvão *et. al.* NAR 2012 (102)	DNA	water	Alexa 488–C6-dT Alexa 594–C6-dT	0.60	2/3	1.33	53.2	52.3
Reinartz *et.al.* J. Chem. Phys. 2018 (16)	Protein	water	Alexa 488–C5-maleimide Alexa 594–C5-maleimide	0.92	2/3	1.33	54.0	53.0
		GdmCl*				1.465	51.0	52.0

*GdmCl: Guanidium hydrochloride

2.3 Single-pair FRET

Bulk or ensemble FRET experiments yield fluorescence signals that are averaged over a wide range of conformations of macromolecules. Individual conformations cannot be detected by this procedure. Single-molecule (or single-pair) FRET (104, 105) was developed to measure conformational heterogeneity present in the sample (106, 107). Different approaches to perform single-pair FRET are very nicely described in the work of Roy *et al*. (2008) (106). The work presented in this thesis has been performed on a diffusion-based spFRET setup.

2.3.1 Single and Dual-laser excitation

Instrumentations and in-house developed softwares (FRETtchen and AlexEval) for the in-house single-pair FRET setup are detailed in the dissertations of Dr. Alexander Gansen (single-laser excitation) (108) and Dr. Kathrin Lehmann (single- and dual-laser excitation) (109). The AlexEval software developed by Dr. Sebastian Isbaner can be found at this link: https://github.com/sisbaner/AlexEval. Instrumentation details are added in Appendix C. In this thesis the concepts of proximity ratio, stoichiometry and the detection factor are detailed, leading up to the extraction of distance information.

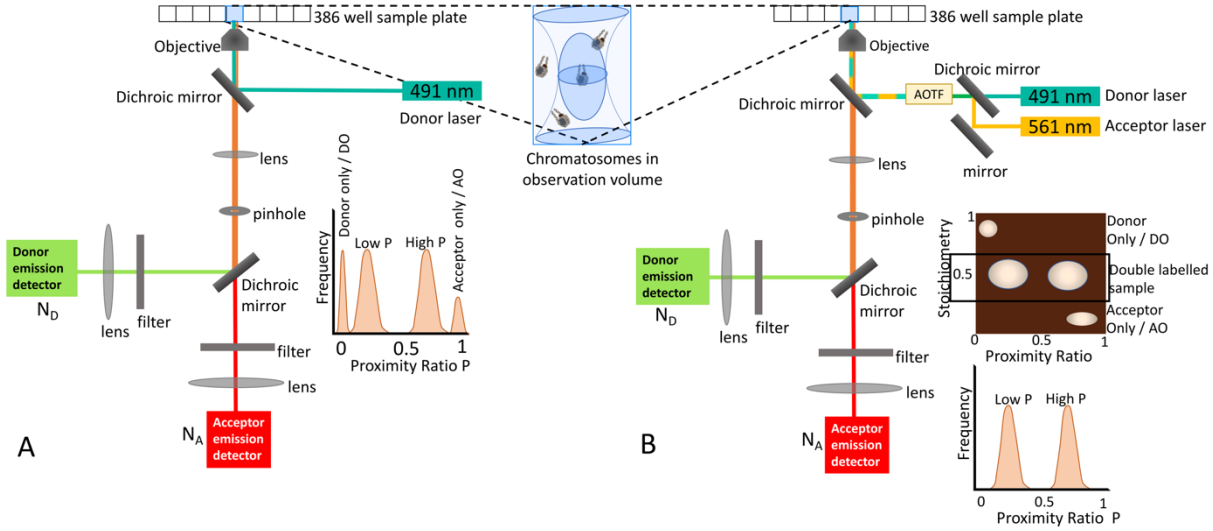


Figure 2.4: *SpFRET setup.* ***A.*** *Setup of single-laser excitation FRE .* ***B.*** *Setup of dual excitation ALEX-spFRET.*

The in-house setup for single-pair FRET built by Dr. Gansen contained a confocal microscope that focused the excitation laser beam to a point that resulted in an observation volume in the range of femtoliters. To ensure detection at single-molecule or single FRET-pair levels, the labelled macromolecule, i.e., chromatosomes or chromatosomes, were diluted to picomolar concentrations for the measurement. The sample buffer (Appendix B) with a physiological pH of 7.5 contained 1xTE, an NaCl concentration of 25 mM for standard measurements or up to 750 mM in salt-series experiments as shown in chapter 6. It also contained ascorbic acid of a concentration of 0.8 mM to scavenge free radicals generated by photobleaching of the fluorophores (92). About 0.01 mM of the surfactant nonidet-P40 was added to the buffer to maintain the stability of the chromatosomes (93). In order to improve the stability of the chromatosomes further unlabelled chromatosomes were added at a greater concentration than the labelled chromatosome. The concentration of labelled chromatosomes was 50-100 pM and that of the unlabelled, 200-250 pM. The total concentration of chromatosomes was 300 pM.

In this setup, labelled and unlabelled chromatosomes freely diffuse in and out of the focus as described graphically in Figure 2.4. The observational volume gets illuminated by either the 'donor' laser having a wavelength of 491nm (Figure 2.4A) in the single-laser excitation mode. In ALEX (Alternating Laser Excitation) (95, 110) both the donor laser and the 'acceptor' laser having a wavelength of 561 nm (Figure 2.4B) were alternately illuminated at intervals of 100 µs with 20 µs pauses. The laser power at the objective was maintained at 40 µW.

Labelled chromatosomes diffusing through the ellipsoidal observation volume fluoresce when excited by the laser. Resonance energy transfer also occur depending on how close the fluorophores are to each other. The emitted photons are detected by the donor and the acceptor avalanche photodiodes. Here, instead of FRET efficiency, the term proximity ratio comes to play. Proximity ratio is directly proportional to the FRET efficiency. It is the ratio of the number of photons detected in the acceptor channel (N_A) to the total number of photons detected in the donor (N_D) and the acceptor channels (N_A). The equation describing proximity ratio in terms of the number of photons in the acceptor detection channels vs. the total is:

$$P = \frac{N_A}{N_D + N_A} \qquad (g)$$

2.3.2 Detecting unwanted spectral overlaps

The spectral overlap between donor emission and acceptor absorbance is the requirement for resonance energy transfer to occur. But there are two other overlap regions between the emission and absorbance spectra of a FRET pair.

The first one is termed 'direct excitation of acceptor' (95, 108). This is an overlap between the donor excitation and the acceptor excitation spectra (denoted as (ii) in Figure 2.5A). The second one, termed 'crosstalk' (95, 108, 111) (denoted as (iii) in Figure 2.5A) is an overlap between the donor emission and the acceptor emission. These two overlaps can be detected in FRET histograms obtained from both single and dual laser excitation (ALEX) spFRET, as illustrated in Figure 2.5.

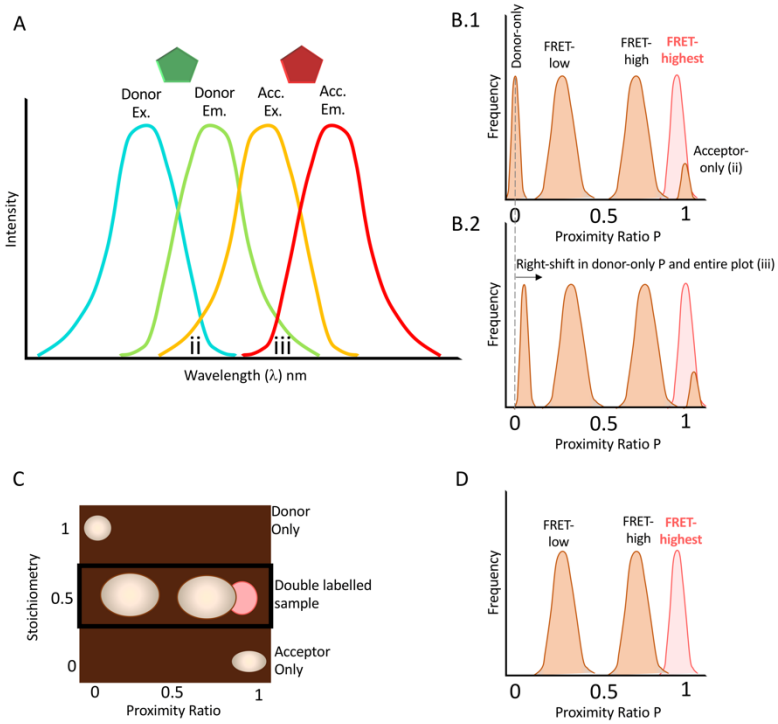


Figure 2.5: *Overlap detection by single-laser excitation and ALEX-spFRET.* ***A.*** *Spectral overlap regions in the excitation and emission spectra of donor and acceptor fluorophore denoted as (ii) and (iii)* ***B.1*** *Schematic of a proximity ratio histogram plot obtained in single-laser excitation spFRET* ***B.2*** *Same plot as B.1. However spectral overlap (iii) causes the plot to shift to the right.* ***C.*** *Schematic of a plot obtained from ALEX-spFRET showing both stoichiometry and proximity ratio. Donor only and Acceptor only peaks have different stoichiometry.* ***D.*** *A part selected in plot C. (black rectangle) replotted. A stoichiometry range of 0.25 to 0.75 and the entire proximity ratio range from 0 to 1 is plotted showing absence of donor-only and acceptor-only peaks that were present in the histograms obtained from single-laser excitation in B.1 and B.2.*

In single-laser excitation, the sample is excited only by the donor laser (Figure 2.4A). Ideally only the donor fluorophore should fluoresce. In presence of an acceptor fluorophore in close proximity, energy transfer can occur and the acceptor fluorophore can fluoresce too. So ideally photons emitted by the donor in absence of the acceptor fluorophore would have a proximity ratio of 0, because N_A or number of photons in the acceptor channel would be 0 (refer to equation (j)). In the presence of an acceptor fluorophore, the acceptor fluorophore would be able to fluoresce only after resonance energy transfer from the donor fluorophore. However, if an overlap of type (ii) (Figure 2.5A) exists between donor and acceptor excitation, a small fraction of the acceptor fluorophore would be able to become 'directly excited' by the donor laser and thus would fluoresce even in absence of resonance energy transfer (95). This direct excitation of the acceptor in the absence of a donor would result in a proximity ratio of 1 (equation (j)). This shows up as the small peak to the extreme right in Figure 2.5 B.1. Typically, acceptor labelled nucleosomes where the donor labelled LH has dislodged, or acceptor labelled free DNA would represent this small peak. How do we know that this small peak represents the 'acceptor-only' fraction due to direct excitation and is not a high FRET peak due to very close proximity of the donor and acceptor? This can be concluded by the height of the peak. If the fluorophores are indeed very close to each other they would be so in a substantial fraction of the total double-labelled population, if not the whole, resulting in a taller peak (represented by a pink peak termed 'FRET-highest' in the schematic B.1 in Figure 2.5). On the contrary, overlap (ii) involves only a small fraction of the donor excitation spectrum. This means that only a very small percentage of acceptor-only samples would get directly excited, resulting in a much smaller peak. Examples of very high FRET peak (not acceptor-only) can be found in chapter 6 for the trichromatosome study.

The second type of overlap, numbered (iii) involves the emission spectra of the donor and the emission spectra of the acceptor (Figure 2.5A). This indicates that the donor can emit partly in the acceptor emission range and can therefore be picked up in the acceptor detector channel. This type of overlap, or 'crosstalk' (95, 108, 111) would mean that even in the complete absense of acceptor, the donor-only peak would have a proximity ratio greater than 0 (Figure 2.5 B.2). This crosstalk shifts the entire plot to the right and requires correction (111).

In ALEX-spFRET (Alternating Laser Excitation), (95, 110, 112), the output consists of not just the proximity ratio but the stoichiometry (Figure 2.5C) of the fluorophores. The stoichiometry is defined as the relative proportion of the donor and the acceptor fluorophores. Thus, donor-only, double labelled and acceptor-only species are well separated in terms of stoichiometry (Figure 2.5C). The stoichiometries of all double-labelled FRET populations, including the pink 'FRET-highest', are 0.5 whereas the acceptor-only have a stoichiometry of 0 and the donor-only, a stoichiometry of 1. A stoichiometry range of 0.25 to 0.75 are selected to capture the double-

labelled population and exclude the donor and acceptor only populations. This chosen region can then be plotted (Figure 2.5D).

In the ALEX setup the donor and the acceptor lasers are alternately allowed to excite the sample volume. The sequence is 100μs donor excitation - 20μs gap - 100μs acceptor excitation - 20μs gap and so on. The events that occur during this sequence, and consequently, the formulae for calculating stoichiometry and proximity ratio are described as follows (95, 111):
'N' refers to photon counts for a single FRET pair. The subscript D_{ex} or A_{ex} refers to excitation by donor or acceptor. Similarly, the superscript, D_{em} or A_{em} refers to emission by donor or acceptor.

N_{Dex}^{Dem} = Donor emission during donor excitation

N_{Dex}^{Aem} = Acceptor emission during donor excitation: occurs due to FRET and direct excitation of acceptor (overlap (ii)). Overlap (iii) or emission by donor in the acceptor emission range can also contribute to this.

N_{Aex}^{Aem} = Acceptor emission during acceptor excitation. Here, despite the overlap (ii) between donor and acceptor excitations, the donor is not excited. This is because the acceptor laser excites the sample with a wavelength higher than that covered by the overlap region.

The proximity ratio can be calculated by the following formula (95, 111):

$$P = \frac{N_{Dex}^{Aem}}{N_{Dex}^{Aem} + N_{Dex}^{Dem}} \quad \text{(h)}$$

Histograms were plotted from the proximity ratios obtained from a measurement, comprising of 50 bins with a bin width of 0.02. These plots were fitted to multiple Gaussian peaks in Matlab. The mean proximity ratio of the peak major population peak, the full width at half maximum (FWHM) and the relative area under the peak are reported (Supplementary Information Chapter 4 and 6).

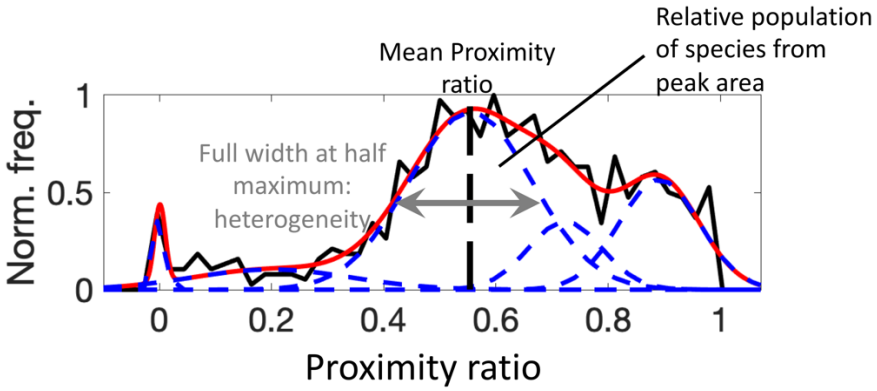


Figure 2.6: *Histogram of proximity ratios fitted to multiple Gaussian peaks. The grey arrow denotes the full width at half maximum (FWHM). The dashed black line denotes mean proximity ratio of the peak. The relative area under the peak denotes the population of the particular species represented by the peak.*

The full-width at half maximum describes the heterogeneity of a species represented by the peak or the dynamics of the macromolecule that happens faster than the observation time

(113). The stoichiometry or the relative proportions of the fluorophores can be calculated by the following formula (95, 111):

$$S = \frac{N_{Dex}^{Aem} + \gamma N_{Dex}^{Dem}}{\gamma N_{Dex}^{Dem} + N_{Dex}^{Aem} + N_{Aex}^{Aem}} \quad \text{(i)}$$

Here γ refers to the instrument detection factor that will be described in the next section.

2.3.3 γ–factor correction

The relation between experimentally derived proximity ratio P and the FRET efficiency E hinges on the detection factor γ (95, 108) .

$$E = \frac{P}{\gamma - P(\gamma - 1)} \quad \text{(j)}$$

Where,

$$\gamma = \frac{Q_A \eta_A}{Q_D \eta_D} \quad \text{(k)}$$

Here $Q_{A \text{ or } D}$ refers to quantum yield of acceptor and donor in the sample measured and $\eta_{A \text{ or } D}$ refers to the detection efficiency in the acceptor and donor channel. Thus, the detection factor depends both upon the detection efficiencies and the alignment of the detectors and the optics, and on the chemical environment of the sample solvent that governs the quantum yield of the fluorophore.

Calculation of the γ factor is described in Lee *et al.* (2005) (95). Two FRET standards or DNA oligomers having Alexa 488 (green pentagon Figure 2.6A) and 594 (red pentagon Figure 2.6A) separated by 10 bp or 21 bp were measured prior to each measurement day to calculate the γ factor.

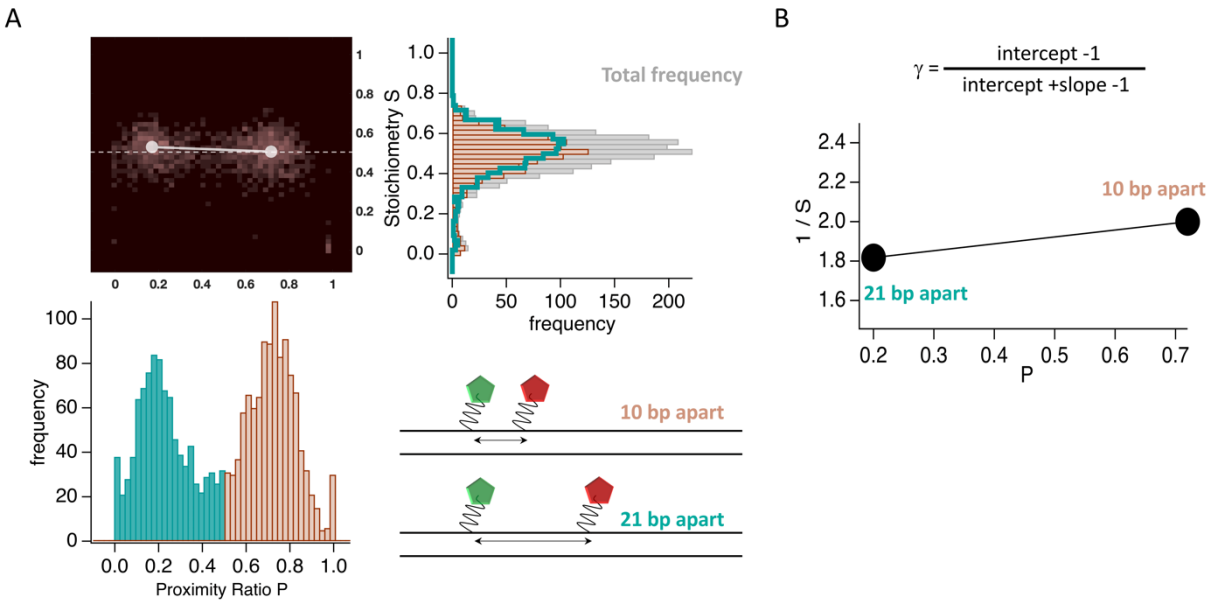


Figure 2.7: *γ-factor calculation by ALEX-FRET.* ***A.*** *Top left shows a proximity ratio (x-axis) vs. stoichiometry (y-axis) plot obtained from ALEX-spFRET measurement of FRET standards (oligonucleotides having Alexa 488 and 594, 10 bp apart and 21 bp apart) mixed in a 1:1 ratio. The mean stoichiometries of the two peaks, low FRET for 21 bp apart and high FRET for 10 bp apart are on a slope. The γ-factor can be found out by the equation (95) in* ***B.***

This calculation (95, 108) is based on observations that due to instrument misalignment, the double labelled sample peaks that should ideally fall on the stoichiometry 0.5 line (Figure 2.6A) fall on a slope instead. By exploiting this, the γ factor can be found out on each measurement day. The γ factors calculated for each measurement day are tabulated in Supplementary Information under Chapter 2.

2.4 Calculating inter-fluorophore distance

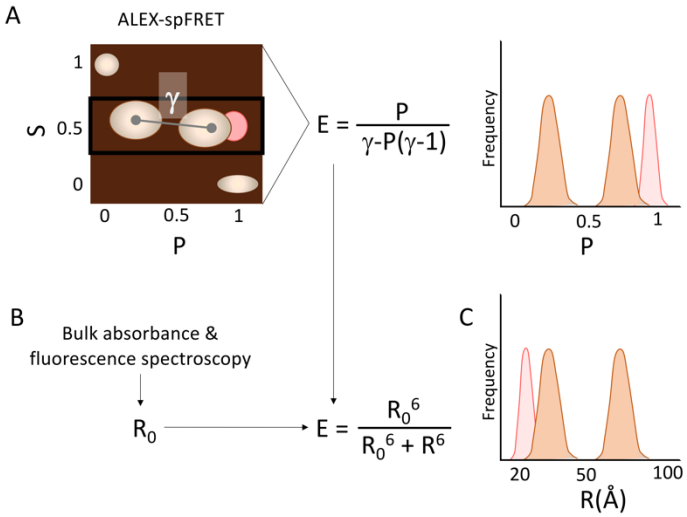


Figure 2.8: *Extracting distance from proximity ratio.* ***A.*** *ALEX-spFRET experiments yield proximity ratio P and γ factor. FRET efficiency E can be obtained. Extreme right shows the P histogram for stoichiometry range 0.25 to 0.75 (within black rectangle in S vs. P plot in* ***A****.* ***B.*** *Bulk spectroscopy to obtain Förster distance R_0. FRET efficiency and Förster distance is used to calculate distance R. Thus, in* ***C.*** *proximity ratio histogram can be converted to distance histograms. Here distance values are given arbitrarily for illustration.*

Once P and γ obtained from single-molecule experiment are used to calculate E, and R_0 is obtained from bulk experiments, inter-fluorophore distance R can be found out using equation (a). In this thesis, inter-fluorophore distances were extracted from proximity ratios obtained from spFRET measurements. These distances were used to build models as described in Chapter 3.

Chapter 3

Materials and methods: experiment and modeling

Table 3.1: Decription of chromatosome constructs. *Mono and Trichromatosome construct information chart containing type of construct (symmetric, asymmetric monochromatosome or trichromatosome), length of L-DNA, sequence of flank (for detailed sequences see Chapter 4 Table 4.1 and Appendix A1), and FRET pairs studied.*

Name of construct	**Type**	**L-DNA length**		**Flank sequences**		**FRET pairs**			
						LH-gH-Don		**LH-CTD-Don**	
		Minus	**Plus**	**Minus**	**Plus**	**-Acc**	**+Acc**	**-Acc**	**+Acc**
AsyMG	Asymmetric monochromatosome	24	40	mixed	GC-tract	✓	✓	✓	✓
AsInvMG	Asymmetric monochromatosome	40	24	mixed	GC-tract	✓	✓	✓	✓
MG	Symmetric monochromatosome	40	40	mixed	GC-tract	✓	✓	✓	✓
GM	Symmetric monochromatosome	40	40	GC-tract	mixed	✓	✓	✓	✓
AG	Symmetric monochromatosome	40	40	A(T)-tract	GC-tract	✓	✓	✓	✓
GA	Symmetric monochromatosome	40	40	GC-tract	A(T)-tract	✓	✓	✓	✓
TU	Symmetric monochromatosome	40	40	A(T)-tract	A(U)-tract	✓	✓	×	×
A-far	Trichromatosome	All 40		Nuc 1: outer: A(T)-tract Inner: GC-tract	Nuc 3: outer: A(T)-tract Inner: GC-tract	L-DNA between Nuc 1 and 2: Don L-DNA between Nuc 2 and 3: Acc			
A-near	Trichromatosome	All 40		Nuc 1: outer: GC-tract Inner: A(T)-tract	Nuc 3: outer: GC-tract Inner: A(T)-tract				

3.1 DNA preparation for monochromatosomes

Monochromatosomes were reconstituted from either 212 bp (asymmetric chromatosomes) or 226 bp (symmetric chromatosomes) DNA. Fluorescent labels attached to specific thymine residues on the primer (IBA Lifesciences, Göttingen, Germany) via a C6 linker were incorporated to the DNA by PCR. All the DNA constructs contained the highly positioning Widom 601 (53) sequence 40bp from one (asymmetric chromatosome) or both (symmetric chromatosome) the DNA ends. The asymmetric DNA (AsyMG and AsInvMG) and the symmetric DNA MG have the same sequence. These DNA constructs were amplified from the pGEM-3z/601 vector (53). For the other constructs, 226 bp DNA templates were bought from Biolegio B.V. (Nijmegen, Netherlands). Entire sequences of the constructs and the primers are given in Appendix A1.

Typically, 1ml PCR reaction mix comprised of 10ng of 226bp DNA template (Bioloegio B.V.), 4μg of each of the forward and reverse primers (IBA Lifesciences), and Taq-polymerase containing master mix (Thermo Scientific). 4ml of PCR reactions were carried out for each DNA construct.

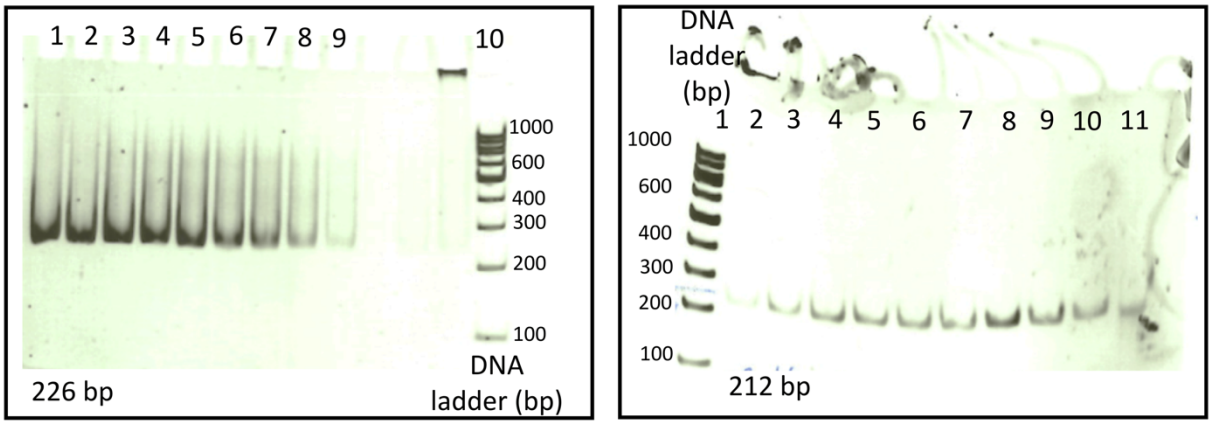


Figure 3.1: *HPLC purified 226 and 212bp DNA for monochromatosome reconstitution. Left: 8% polyacrylamide gel showing labelled 226 bp DNA for symmetric chromatosomes. Right: 8% polyacrylamide gel showing labelled 212 bp DNA for asymmetric chromatosomes. 8% gels are run at 150V (15V/cm) for 1 hour. Both the gels were stained with EtBr and visualised under UV light. The numbers represent the fractions from HPLC that are pooled and concentrated.*

DNA were amplified and labelled using PCR and were purified by Nathalie Schwarz using Gen-Pak FAX HPLC (Waters). 8% polyacrylamide gel electrophoresis were performed to check the size of the final product (Figure 3.1). The HPLC fractions were pooled together and further passed through DNA purification columns NAP-10 and NAP-5 and the resultant solution was concentrated using UNIVAPO vacuum concentrators (Progen Scientific).

3.2 DNA preparation for trichromatosomes

Trichromatosomes were reconstituted from 600 bp DNA containing three Widom 601 sequence separated by 40bp L-DNA stretches. The flank sequences of chromatosome 1 and 3 contain the A-tract either on the outer L-DNA arms (A-far) or on the inner L-DNA arms (A-near) (see Appendix A for full sequences). The final 600bp long DNA was built from three separate DNA fragments. To ensure that the fragments were ligated in the correct orientation, a Golden-Gate assembly (114, 115) was performed *in vitro*. This procedure involves a type IIs restriction enzyme BsaI-HF-v2, that recognizes a non-palindromic DNA sequence 5'-GGTCTC-3' and cleaves the DNA five base pairs away from the recognition site, producing a four base pair sticky end. The advantage of a non-palindromic recognition site is that the enzyme can cleave the DNA only on one side of the recognition site. The advantage of cleaving a few base pairs away from the recognition site is that unique sticky ends could be generated. These two features of BsaI-HF-v2 ensure that the DNA fragments are joined in a specific orientation (Figure 3.2).

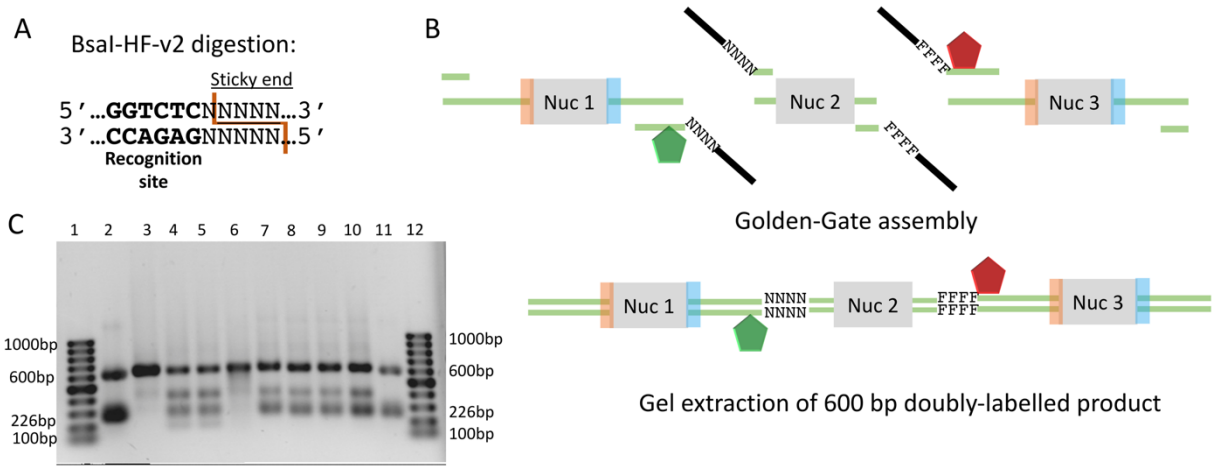


Figure 3.2: *Preparing 600 bp DNA for trichromatosome reconstitution. A. shows mode of action of BsaI-HF-v2. B. Golden-Gate assembly for preparing 600 bp DNA. Green: DNA. Grey: Widom 601 ncp DNA sequence termed as Nuc 1, 2, and 3. Black line denotes recognition site for BsaI-HF-v2. NNNN and FFFF are unique 4bp sticky ends. The orange and blue rectangles on either side of Nuc 1 Widom 601 and Nuc 3 Widom 601 denote flank regions in general. The colour coding is done to denote a specific orientation of ligation. Top: Nuc 1, 2 and 3 are separately amplified by PCR from labelled or unlablled primers. Bottom: 600bp DNA after the digestion and ligation steps of Golden-gate assembly. C. 1% agarose gel run at 116 V for 1h. Lanes 1 and 12 are 1 kbp ladders. Lanes 2 and 11 are a mix of two marker DNA of lengths 600 bp and 226 bp. Lane 3 shows 600 bp A-far DNA. Lanes 4 and 5 are the reaction mix for A-far preparation, showing a mix of 600 bp DNA and incompletely ligated products of lengths ~400bp, ~200bp and unligated 150bp fragment. Lane 6 shows A-near DNA. Lanes 7-10 shows reaction mix for A-near preparation with completely ligated product (600bp) and similarly incompletely ligated (~400 and ~200bp) or unligated (~150bp) products.*

The modified in-vitro Golden-Gate assembly protocol was developed jointly with Gabriele Müller. The assembly protocol is as follows:

- *PCR of fragments 1, 2 and 3:* To amplify fragments 1, 2, and 3, and simultaneously, to incorporate BsaI-HF-v2 recognition site and fluorescent labels, PCRs of individual fragments were performed separately. The PCR reaction mix comprised of 20ng/ml templates, 4μg/ml of each of the forward and reverse primers (sequence in Appendix A). On incorporation of the recognition sequences plus the overhangs, the PCR products were of length 242 bp (fragments 1 and 3) and 172 bp (fragment 2). The PCR products were cleaned up using the Macherey-Nagel PCR clean-up protocol (NucleoSpin Gel and PCR cleanup).
- *Digestion of fragments:* Each fragment was separately digested using 1.5 to 3.0 units of BsaI-HF-v2 (New England Biolabs)/μg DNA/20 μl of reaction volume. The digestion reaction was carried out at 37°C for a minimum of 18 hours to a maximum of 36 hours. After the reaction, the products were cleaned up using Macherey-Nagel protocol (NucleoSpin Gel and PCR cleanup). Digestion was checked by running 8% polyacrylamide gels for 1.5 hours at 120 V. Digested products had lower molecular weights than undigested products and migrated faster. Incomplete digestion was suggested by the presence of ladders (Figure 3.3) instead of one low molecular weight band, and re-digestion was carried out if necessary.
- *Ligation of fragments:* All three well digested fragments were added together in the ratio of 1.00:0.65:1.25 by molecular weight. Bulk ligation was performed using 500 units of T4 ligase (New England Biolabs)/μg DNA/300 μl of reaction volume. Ligation was carried out under 15°C for a minimum of 24 hours. Ligation products were checked by 1% agarose gel run at 116 V for 1 hour (Figure 3.2).

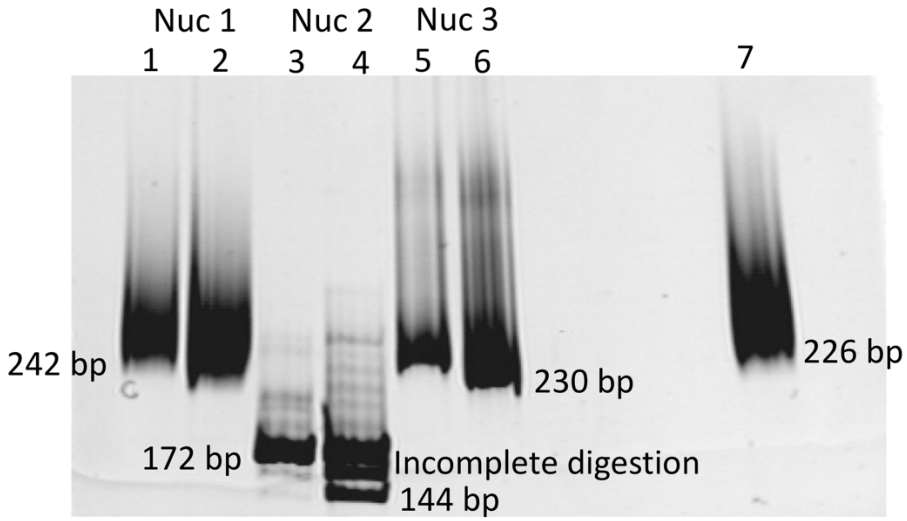


Figure 3.3: *Digested and undigested fragments. 8% polyacrylamide gel run at 120 V for 1.5h. Lanes 1 and 2 show fragment Nuc 1 undigested (Lane 1) and digested (Lane 2). Lanes 3 and 4 show fragment Nuc 2 undigested (Lane 3) and digested (Lane 4). Lane 4 shows a ladder of incompletely digested DNA. Lanes 5 and 6 show fragment Nuc 3 undigested (Lane 5) and digested (Lane 6). Lane 7 shows a 226 bp DNA molecular weight marker.*

The 600 bp fragment was gel-extracted from the mix of unligated and incompletely ligated products (Figure 3.2C), and Macherey Nagel gel-extraction cleanup (NucleoSpin Gel and PCR cleanup) was performed. Since the DNA sample was to be studied by fluorescence, EtBr wsas not used to stain the bands prior to gel extraction. Instead, Typhoon (Typhoon 9400 scanner, GE Healthcare) scanning of the gels were done to determine the 600 bp band that was to be excised from the gel. DNA was quality controlled by UV-Vis spectroscopy, to check for particulates. Particulates were removed by centrifugation steps to finally yield a clean product suitable for spectroscopy.

3.3 Linker histone purification

Core histone (full-length *X. laevis* recombinant core histones) used for chromatosome reconstitution were purchased from Planet Protein (Colorado State University, USA). The core histone octamer reconstitution has been described previously (92, 116) and was done by Nathalie Schwarz. In this section, the purification of LHs will be discussed. Full length *Xenopus laevis* H1.0b LH was used to reconstitute mono and trichromatosomes. For monochromatosomes, the full-length LH was either labelled on the gH domain (T77C) or at the CTD (G101C). For trichromatosomes, unlabelled wild-type H1.0b was used. Prof. Jeffrey Hayes and Dr. Amber Cutter (University of Rochester Medical Center) kindly gifted us the purified mutant G101C protein, and the plasmids for the T77C mutant and the wild-type. Plasmid amplification and protein expression steps were carried out together with Maria Mildenberger and Gabriele Müller. Nathalie Schwarz operated the FPLC.

3.3.1 Purifying H1.0b T77C mutant

Plasmid amplification:

The plasmid containing the *X. laevis* H1.0b T77C mutant gene was a pET-3d vector (containing the gene for β-lactamase, AmpR, that confers resistance to carbenicillin) that was transformed to *E. coli* XL-1 Blue competent cells for amplification. Transformation steps were:

- To 10 µl of cells on ice, 1µl of plasmid dissolved in PCR grade ribonuclease free water was added.
- 20 minutes of incubation in ice was followed by a heat shock for 3 minutes at 37°C. After this the cells were kept for 1min in ice.
- 400 µl SOC medium (Thermo Fischer Scientific) was added to the cells and the cells were kept for a further 1 minute in ice.
- Cells were incubated at 37°C for 1 hour

LB (Luria-Bertani) agar plates with carbenicillin (50 µg/ml) were prepared while the cells were being incubated. After incubation of the cells, about 100 µl of cells were plated per plate. The plates were incubated for 37°C overnight.

A single colony was selected from the incubated plates and transferred to 10 ml liquid LB medium containing carbenicillin (50 µg/ml). The medium was mixed well to ensure proper distribution of the cells in the medium. The cells were incubated at 37°C overnight, shaking at 170 rpm.

Restriction digestion to check the presence of H1.0b gene:

The amplified plasmid was purified following standard protocol (Macherey-Nagel GmbH) and subjected to restriction digestion to check for the presence of the insert. Restriction enzymes and buffer were purchased from Thermo Fischer Scientific.

Type of reaction	**Enzymes**	**Amount of plasmid DNA (µg)**	**Enzyme units**
Single digestion	BamHI	10	10
Double digestion	BamHI and NcoI	10	10 each

The digestion reaction was carried under 37°C for overnight. Digested products were visualised by 1.2% agarose gel electrophoresis. The purified plasmid was also sequenced to check for the presence of the correct sequence of the inserts.

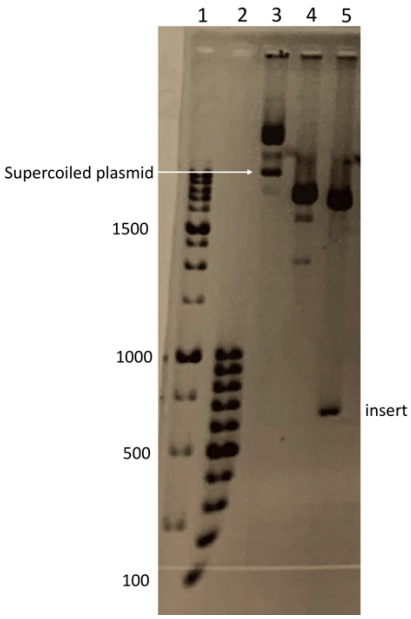


Figure 3.4: *Presence of insert in plasmid. 1.2% agarose gel run at 120V for 1h. Lane 1: 2kbp ladder, Lane 2: 1kbp ladder, Lane 3: undigested plasmid showing supercoiled plasmid. Lane 4: single digestion. Lane 5: double digestion showing plasmid free of insert (slower migrating band on top) and insert (faster migrating) of a size of ~600bp.*

Protein expression:

Purified plasmid was used to transform protein expressing *E. coli* BL21(DE3) cells using the same protocol used to transform XL-1 Blue cells. To check for protein expression an IPTG induction test was performed as advised in the histone purification protocol by Luger, Rechsteiner and Richmond (1999) (117) on an aliquot of the culture, using 0.2 mM to 0.4 mM of IPTG and protein expression was compared with a control uninduced culture. Cell growth was measured at an OD (optical density) of 595 nm. IPTG induction test was carried out at different time points between OD 0.35 and 0.5. Aliquots from induced cultures were centrifuged at 12000 rpm for 2 minutes and 15% SDS-PAGE was performed to check for protein expression.

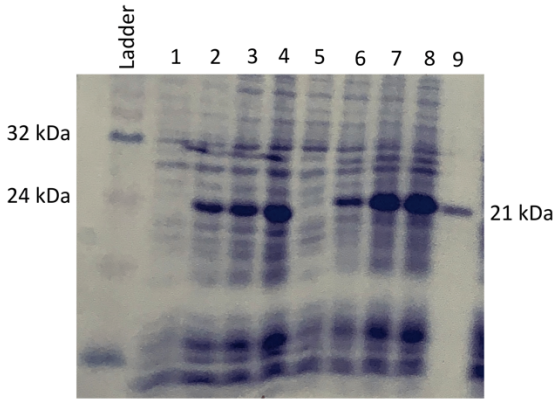


Figure 3.5: *IPTG induction test. 15% SDS-PAGE run for 1 hours at 150 V (21.4 V/cm). Lane 1 and 5 are controls without IPTG treatment. Lane 2-4 0.4 mM IPTG was added. Lanes 6-8 0.2 mM IPTG was added. Lane 9: 21 kDa Mus musculus H1.0 LH. 2-4 and 6-8: the IPTG was added at different times during the growth, between* OD_{595} *0.35 to 0.5.*

After standardising the amount of IPTG to be added, 0.2 mM of IPTG was added to bulk culture at OD (at 595 nm) 0.5, for protein expression and purification in bulk.

Protein purification:

After bulk IPTG induction, the culture was pelleted down and the bacterial pellet was homogenised by sonication (Branson Sonifier 250) in lysis buffer (composition in Appendix B). Using Vivaspin 2 columns (Sartorius, 10 kDa molecular weight cutoff), the supernatant was concentrated and its buffer was exchanged to unfolding buffer (composition in Appendix B). After this step, the concentrated supernatant was first subjected to size-exclusion chromatography using Sephacryl S-200 (109) that was pre-equilibrated with SAU-1000 buffer. The flow rate was 3 ml per minute.

The fractions eluted were pooled together and using Vivaspin 20 (Sartorius, 10 kDa molecular weight cutoff), was concentrated and buffer exchanged to SAU-50 (composition in Appendix B). After buffer exchange the protein solution was subjected to cation–exchange chromatography using Mono-S-HR 10/10 FPLC (Amershan GP250, Pharmacia) (109) column that was equilibrated with SAU-50 and SAU-1000 buffers prior to passing the protein solution. Protein fractions were eluted with SAU-1000. Fractions were pooled together, concentrated and refolded by exchanging the buffer with 2M NaCl 1xTE, pH 7.5.

3.3.2 Purifying H1.0b wild-type

Prof. Jeffrey Hayes and Dr. Amber Cutter kindly gifted us with a H1.0b mutant that had a cysteine residue instead of a lysine at the 195th position. The K195C was mutated back to its original wild-type, containing lysine instead of cysteine. This protein was purified to reconstitute unlabeled chromatosomes and trichromatosomes. This work was done jointly with Gabriele

Müller.

Mutagenesis of K195C to K:

Primers were designed to contain restriction sites for NdeI and NotI and the codon for lysine at position 195 (see Appendix A for sequences). The software Serial Cloner 2.6.1 was used to design the primers.

PCR was performed to amplify the insert from the original plasmid. The amplified insert contained the lysine codon. The PCR product was extracted from a 1% agarose gel using the Macherey-Nagel protocol. The gel extracted insert was digested simultaneously with NdeI and NotI to generate sticky ends. Simultaneously, amplified pET-17b vector (Novagen) were purified by standard protocol (Macherey-Nagel GmbH) and digested with NdeI and NotI to generate sticky ends. To remove the original smaller fragment generated by the digestion, the larger cut plasmid was gel extracted following Macherey-Nagel protocol. While gel-extracting the cut plasmid or the C195K insert, EtBr staining was not performed. Instead, markers of the same size adjacent to the main well were stained with Etbr, the exact band positions were marked and the corresponding area was cut out from the unstained gel. The cut pET-17b vector and the insert were ligated with T4 ligase overnight under 15°C. The 20 µl ligation reaction comprised of: 50 ng cut plasmid, 250 ng insert, 0.5 µl T4 ligase (New England Biolabs M0202). The ligation reaction was carried out overnight at 4°C.

Transforming pET-17b C195K for plasmid amplification:

The newly ligated plasmid was transformed to *E. coli* XL-10-Gold (Agilent) cells following the protocol:

- To the cells on ice, the plasmid mix was pipetted. Incubation was for 30 minutes in ice.
- Heat shock for 40 seconds at 42°C.
- Cells were again placed back in ice for 1 minute.
- 300µl of SOC medium was added.
- Incubation for 1 hour at 37°C.
- Plating was done on carbenicillin containing plates (50 µg/ml).
- Incubation for overnight under 37°C.

Colony was transferred to 10 ml carbenicillin (50 µg/ml) containing Luria-Bertani medium and incubated at 37°C overnight. Amplified plasmids were purified (Macherey-Nagel) and were subjected to restriction digestion with NdeI and NotI enzymes, to check for the presence of insert. Additionally, sequencing was performed to check for the presence of insert and to ascertain presence of the C195K mutation.

Protein expression:

The purified plasmid was transformed to protein expressing *E. coli* BL21(DE3) following standard transformation protocol described earlier. An IPTG induction test was performed by Dr. Kathrin Lehmann, as described for the T77C mutant, prior to bulk protein expression.

Protein purification:

Protein extraction and purification protocol was followed in the same way as described for the T77C LH. The FPLC was operated by Nathalie Schwarz.

3.4 Linker histone labelling

The LH was labelled with the donor fluorophore Alexa 488 C5 maleimide on either the gH domain at residue 77 or at the C-terminal domain at residue 101. The corresponding residues, threonine at 77 and glycine at 101 were mutated to cysteine for the labelling.

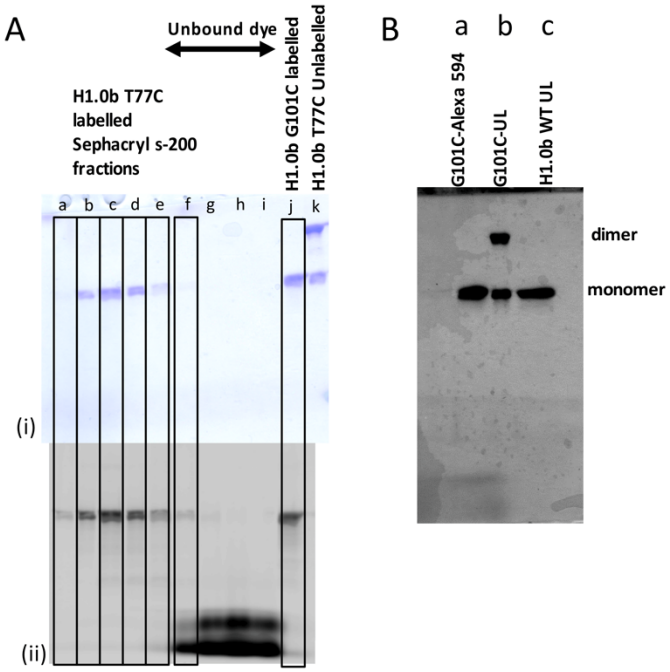


Figure 3.6: *Labelled linker histone. A and B show 15% SDS-PAGE showing labelled T77C and G101C mutants and the wild type H1.0b. The gel shown on A. was run at 150 V (21.4 V/cm) for 40 minutes. Both Ai and Aii gel are the same and show the T77C mutant that was passed through sephacryl S200 after labelling. Ai shows Coomassie blue staining of the gel that stains both labelled (a to f and j) and unlabeled proteins (lane k). Aii show the same gel prior to Coomassie blue staining, and imaged by Typhoon 9400 (GE Healthcare) scanner (excitation 488 nm, emission 500 to 540 nm). In Aii only labelled protein (a to f and j) and unbound fluorophores are seen (g to i). B. the gel was run at 150 V (21.4 V/cm) for 1 hour and was Coomassie stained. This shows the G101C mutant labelled (lane a), unlabeled (lane b) and the wild type protein on lane c. Ai Lane k and B lane b showing unlabeled proteins show the presence of two bands. The higher molecular weight band is formed by LH dimerization due to the free cysteine residue. Due to absence of cysteines on the wild-type, it shows up as one band.*

For labelling, purified cysteine-containing protein was diluted to about 0.4 µg/µl using labelling buffer (composition in Appendix B) containing the reducing agent TCEP at a molar concentration 10 times higher than that of the protein to prevent formation of protein-protein dimers via disulphide linkages. The donor fluorophore Alexa 488 C5 maleimide (Thermo Fischer Scientific) was dissolved in N'N'-dimethyl formamide at a stock concentration of 10 mM. Aliquots of this fluorophore was added to the unfolded protein shaking at room temperature, in three batches every hour, to achieve a final molar concentration 10 times that of the protein. The entire reaction was carried out in the dark to prevent photodegradation of the fluorophore. At the end the mix was stored at 4°C overnight. The next day, the maleimide reaction was stopped by adding L-cysteine at a molar concentration 10 times that of the protein. To remove unbound fluorophore, the labelled protein was dialysed into SAU-1000 buffer (Appendix B for composition) at 4°C overnight. Then it was passed through a sephacryl S200 column. The labelled protein was then refolded by multiple dialysis steps in 2M NaCl 1xTE buffer (pH 7.5). The dialysis step removed the unfolding buffer and also residual unbound fluorophores. The removal of free fluorophores was ascertained by performing an SDS-PAGE (Figure 3.6).

3.5 Reconstituting mono and trichromatosomes:

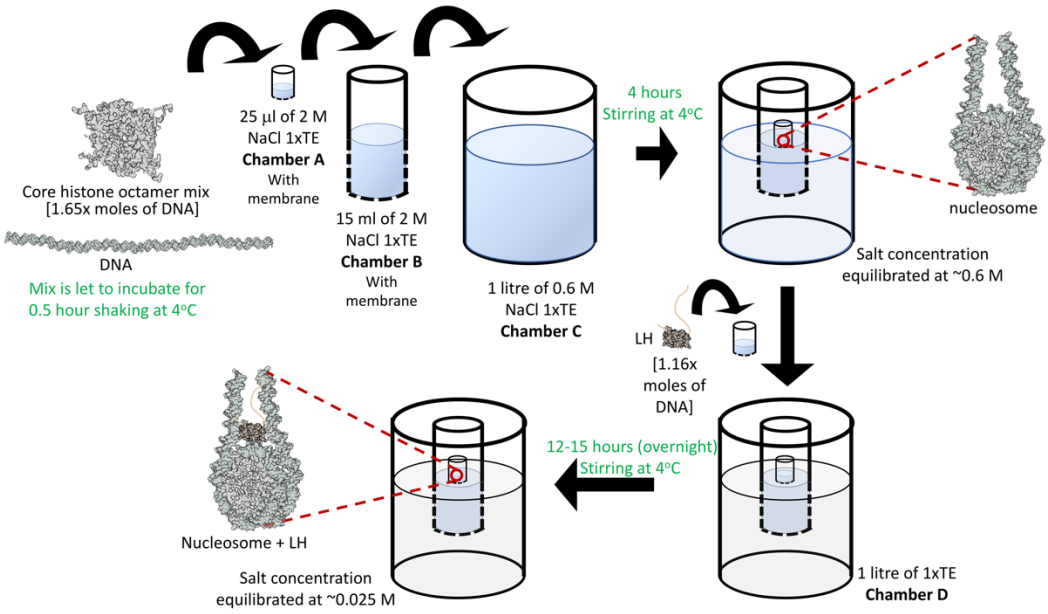


Figure 3.7: *Workflow for monochromatosome reconstitution by salt dialysis. Dashed lines on chamber A and B indicate dialysis membranes.*

Monochromatosomes and trichromatosomes were reconstituted from Widom 601 containing labelled or unlabelled DNA, full-length *Xenopus laevis* core histone octamers and the full-length, labelled or unlabelled *Xenopus laevis* linker histone of the H1.0b subtype. A two-step dialysis protocol was followed to gradually lower the concentration of the NaCl in the solution thus enabling the reconstitution.

For monochromatosome reconstitution core octamers were added at a molar concentration 1.65 times that of DNA, whereas the LH (labelled or unlabeled) was added at a molar concentration of 1.6 times that of DNA (116, 118, 119). For trichromatosome reconstitution core octamers were added at a molar concentration of 8 times that of DNA whereas unlabeled LH was added at a molar concentration of 3.2 times that of DNA. Aside from the molar ratios of protein to DNA, the protocol for reconstituting both mono and trichromatosomes were the same, and described graphically in Figure 3.7. DNA and core histone octamers were added in the specific ratio with 2M NaCl 1xTE buffer (pH 7.5) and the contents were allowed to mix by gently rocking back and forth at 4°C for half an hour. This mix, having a volume of 25µl, was then transferred to chamber A, a Slide-A-Lyzer MINI Dialysis unit (Thermo Fischer) having a molecular weight cutoff of 7kDa. Using a floatation device, the Chamber A was placed into a larger dialysis bag, chamber B, containing 15ml of 2M NaCl 1xTE (pH 7.5). The chamber B was placed into a 1 litre 1xTE (pH

7.5) buffer containing a NaCl concentration of 0.6M (chamber C). A magnetic stirrer was placed in the chamber C to stir the contents gently. Dialysis was performed at 4°C for 4 hours. At this stage the LH was added to the mix in chamber A, after diluting the LH to reduce its original salt concentration from 2M to 0.6M. The chamber A and B was placed on 1 litre of 1xTE (pH 7.5) buffer (chamber D), containing no NaCl. Dialysis with gentle stirring was performed for 12 to 15 hours at 4°C. The juncture between the first and the second dialysis steps differ for trichromatosome reconstitution. For trichromatosome reconstitution the first dialysis step was carried out for greater than 7 hours in 4°C accompanied by slow stirring. A slow lowering of the NaCl was to ensure the proper reconstitution of the three octamers with the 600bp DNA. The second dialysis step was the same as that of the monochromatosomes.

Reconstitution was checked by polyacrylamide gel electrophoresis to estimate chromatosomes and free DNA. For double labelled monochromatosomes, spFRET-ALEX was performed to check for the amount of 1:1 donor:acceptor labelled samples (119).

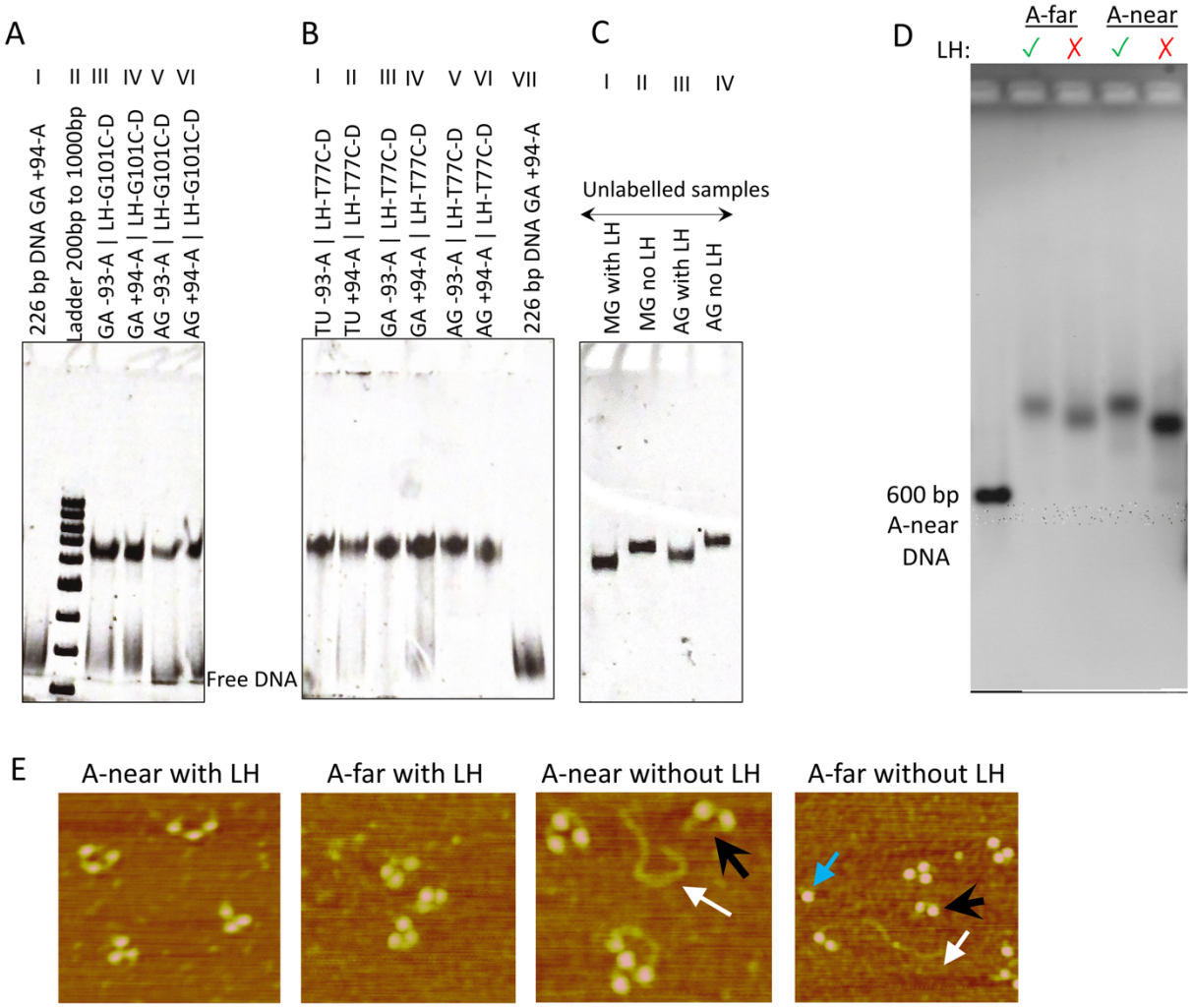


Figure 3.8: *Checking reconstitution of mono and trichromatosomes. A B and C shows native 6% polyacrylamide gels run at 70V for 1.5 hours and stained with EtBr. The high molecular weight bands correspond to monochromatosomes while the low molecular weight bands correspond to free DNA. A. Left to right: 226 bp singly-labelled DNA, ladder 200 to 1000bp, and GA and AG constructs with CTD-labelled LH. B. Left to right: TU, GA and AG constructs with gH labelled LH, and singly labelled 226 bp DNA. C. shows reconstituted samples with and without LH. D. 1% agarose gel with trichromatosomes A-far and A-near with LH (green tick) and without LH (cross). E. AFM images obtained by Martin Würtz showing trichromatosomes. White arrow in A-near and A-far (without LH) denotes free DNA, black arrow denotes dinucleosomes (no LH in these samples) and blue arrow denotes mononucleosome (A-far without LH sample).*

Figure 3.8C shows reconstituted samples with and without LH. The monochromatosome bands with LH migrate faster than the ones without LH. This is because, despite the added molecular weight, the LH compacts the monochromatosome to enable it to migrate faster through the pores of the gel. Figure 3.8D shows a 1% agarose gel with trichromatosomes. The trichromatosomes migrate slower on addition of the LH (green tick).

Atomic force microscopic images (Figure 3.8E) of trichromatosomes were obtained by Martin Würtz (for original images see Supplementary Information Chapter 3). The methodology of AFM sample preparation and image acquisition using NanoScope V (Digital Instruments) and NanoScope software (version 7.13), are detailed in the work by Würtz *et al.* (118). The various species observed in the AFM images are free DNA (Figure 3.8E white arrow) and mono (blue arrow), di (black arrow), trichromatosomes or trinucleosomes. Fully reconstituted products show up as three core histone octamers.

3.6 Modeling monochromatosomes

Starting structures of the chromatosomes were modeled in Chimera (120) from the X-ray crystallographic structure 5NL0 (29) and the cryo-EM structure 7K5X (30). The reason these two structures were used was because of the similarity in the experimental system studied. Both this structures, as well as our experimental system comprised of *X. laevis* full length core histone octamers, and *X. laevis* full-length LH of the H1.0b subtype. These two structures differ mostly in the extent of L-DNA arm opening. These two PDB structures were first modified using Chimera. Modifications first involved stripping the structures of additional proteins, such as the scFv20 in 7K5X, keeping intact the core histone octamers, and the DNA. The gH too were removed and saved as separate PDBs. After this first step, the 'swapna' command in UCSF Chimera (120) was used to mutate the DNA to match the AG and the GA constructs. For experiments, chromatosomes reconstituted with 226 bp of DNA comprised of 40bp L-DNA arms with the fluorophores attached at the midpoints of the arms, i.e., 20bp from the entry/exit sites and 20bp from the ends. The PDB structures 5NL0 and 7K5X both have 23bp L-DNA arms. The fluorophore attachment positions at 20bp from the entry/exit sites were 3 bp from the ends in these structures. Since no positional

information was obtained for the 20 bp stretch of DNA above the fluorophores in the AG and GA construct, the L-DNA arms in the models were not further extended.

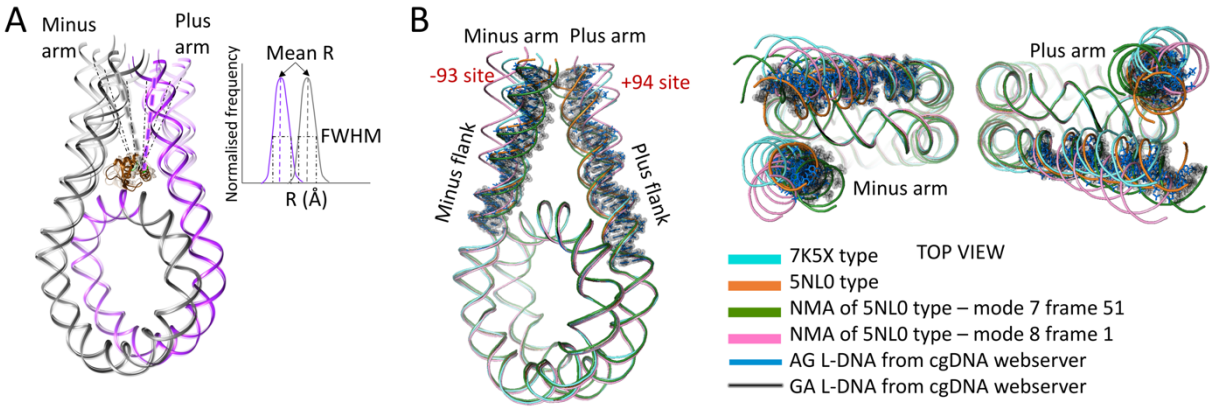


Figure 3.9: *Modeling of AG and GA chromatosome ensembles. A. full-width at half maxima (FWHM) of the distance peaks, above the shot noise limit (113), may reflect L-DNA dynamics. B. L-DNA conformations sampled by using two different starting structures (5NL0 (29) and 7K5X(30)) and performing normal mode analysis on the 5NL0 structure (mode 7 frame 51 and mode 8 frame 1). The L-DNA arms of constructs AG and GA modeled using cgDNA webserver (121) were overlayed onto the four different conformers. (i) shows view from the side and (ii) shows view from the top showing that the DNA bending induced by A-tracts can be sampled by modeling the four different L-DNA arm opening conformations.*

After mutation of the DNA, the two PDBs were protonated in Chimera. Keeping the core histone octamers fixed, the DNA of both the structures were minimised in Chimera using the ff99bsc0 force field (122). This was done to remove possible clashes arising from mutating the DNA. The minimisation protocol was as follows:

- Steepest descent steps = 100. Step size = 0.02Å
- Conjugate gradient steps = 10. Step size = 0.02Å

After the minimisation the mutated DNA was visually inspected for presence of kinks or clashes. Further inspection using the webserver 3DNA (123) was performed to check for presence of any A-DNA step. The AG and GA models contained 91% B-DNA after minimisation.

The minimised AG and GA constructs of the 5NL0 type were then subjected to elastic network normal mode analysis using the elNémo webserver (124) using an amplitude range (DQMIN to DQMAX) of -100 to +100 with an increment (DQSTEP) of 20 steps. This was performed to sample further arm-opening conformations (125, 126) that can lead to broadening of the proximity ratio peak above the shot noise level (113). Mode 7 frame 51 and mode 8 frame 1 was selected because they showed maximum arm deviations. This resulted in four different conformations of nucleosomes (Figure 3.9), each with a different extent of arm-opening. For the TU construct only the 5NL0-type conformer was mutated in Chimera and L-DNA dynamics was not sampled.

A-tracts have been observed to induce bending in the DNA (127, 128). To take sequence dependent bending into account, the cgDNA webserver (121) was used to model the two L-DNA

arms of AG and GA, using paramset+ 1 (129). New nucleosomes were built from one of the models, by removing the old L-DNAs and attaching the new L-DNAs built on cgDNA. Figure 3.9B shows that sequence dependent DNA bending (blue and grey L-DNA) can be sampled by the four different arm-opening conformations I modeled.

The gH was separately mutated in Chimera using the 'swapaa' command, to contain a cysteine residue instead of the threonine at position 77. After mutation, the gH was minimised separately using the ff14SB (130) force field in Chimera (120). Minimisation protocol was same as that for nucleosomes. The minimised gH was then replaced onto the four different nucleosome conformations (5NL0-type, 7K5X-type, NMA of 5NL0-mode 7 frame 51 type and the mode 8 frame 1) in different orientations and positions, generating an ensemble of structures, with differently oriented or positioned gH per nucleosome arm-opening conformation. Trichromatosome modeling is detailed in Chapter 6.

3.7 Computing theoretical inter-fluorophore distances

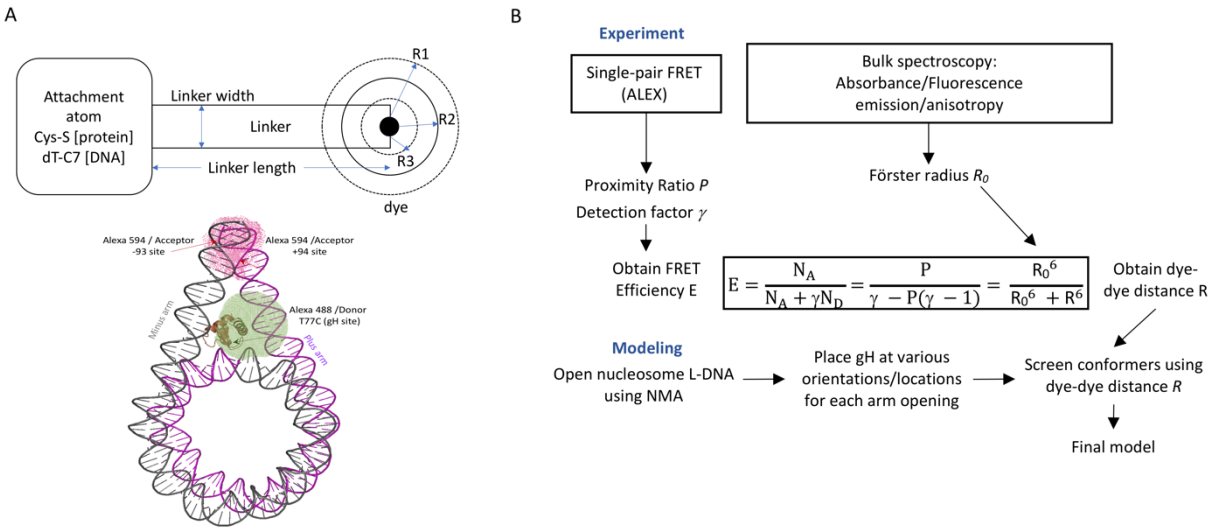


Figure 3.10: *Computing theoretical inter-fluorophore distance. A. Top: shows a schematic of the accessible volume simulation of the fluorophores in the software FRET Positioning and Screening or FPS (131, 132). Bottom shows one chromatosome with the fluorophores generated from FPS. Alexa 488 accessible volume sphere is shown in green while that for Alexa 594 is shown in red. B. General workflow from single-pair FRET to modeling as described in this thesis.*

The ensemble of nucleosome structures generated were screened by using the software FRET Positioning and Screening (FPS) (131). The distances derived from experiments (see Chapter 2 for methodology and Figure 3.10 B for workflow) and the Förster distance calculated (see Chapter 2) were used as the input. In this software, the fluorophores Alexa 488 and 594 are represented by accessible volume spheres (131, 132). These spheres, represented in green and red in Figure 3.10 encompass all the allowed and equally probable positions that the fluorophore

can sample without encountering any steric clash. The inter-fluorophore distance is thus represented in three ways (131, 132):

R_{mp} = distance between the mean positions of donor and acceptor fluorophore. This parameter is solely obtained from the model and is independent of experimentally derived parameters like R_0.

$<R_{DA}>$ = overall average of distances between each allowed positions of donor and acceptor

$<R_{DA}>_E$ = distance derived from FRET experiments. To theoretically calculate this, the Förster distance is required.

For screening of the ensembles, the input distances are thus designated as $<R_{DA}>_E$ and screening is carried out. In the input file for FPS, apart from R_0 and experimentally derived distances, the atom number where the fluorophore is attached (dT-C7 atom for DNA and cysteine sulphur for protein), and the fluorophore parameters are also given as inputs. The fluorophore parameters are graphically shown in Figure 3.10A (adapted from the works of Kalinin (131) and Sindbert (132)) and tabulated in Table 3.2 below.

The R1 to 3 represent three different fluorophore radii from the central atom of the fluorophore. Linker length is the length between attachment atom on DNA or protein to the central atom on the fluorophore and the linker width is the width of this linker arm.

Table 3.2: *Fluorophore parameters used in the FPS software. A graphic description of these parameters is given in Figure 3.10A. A. All values are in Å.*

Fluorophore	Linker length	Linker width	R1	R2	R3
Donor: Alexa 488-C5-maleimide-Cys (S atom) [protein label] (FPS 1.1 software (131))	20.5	4.5	5.0	4.5	1.5
Acceptor: Alexa 594-C6-dT (C7 atom) [DNA label] (101)	20.0	4.5	8.1	3.2	2.6

Computed distances were compared to experimentally derived distances. In Chapter 5 the modeling results are described. While only one frame from the ensemble of conformers represents the mean distance, the peak width of the distance histogram represents arm flexibility. For the TU construct, since distances were not experimentally measured, the FRET efficiency was computed using FPS.

Chapter 4

Role of the linker-DNA on the positioning of the linker histone

4. 1 Introduction

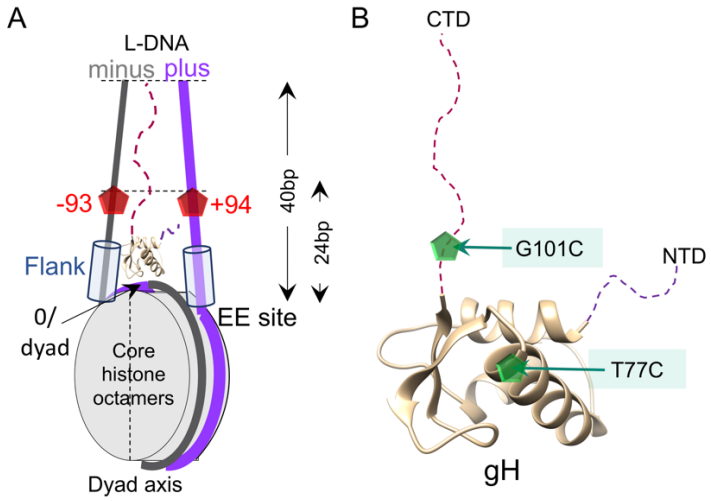


Figure 4.1: *Experimental designs for monochromatosome study.* ***A.*** *Schematic of a monochromatosome. Grey disc represents full-length core histone octamers. The DNA wrapped around is denoted in grey and purple. The central base pair is numbered 0 and termed as the dyad base pair. DNA on the grey arm is numbered with negative values and DNA on the purple arm is numbered with positive values. L-DNA length could be 40bp or 24bp. The chromatosome is labelled with Alexa 594 acceptor fluorophore either on the grey minus, or purple plus arm, at positions -93 or +94. Acceptor fluorophore labelling positions are denoted by red pentagons.* ***B.*** *Schematic of the full-length linker histone (LH) showing the ~80 amino acid globular domain (gH), the 100 amino acid long, intrinsically disordered CTD and the shorter, disordered NTD (both represented by dashed lines). The LH is labelled with donor fluorophore, Alexa 488 (green arrows denoting labelling positions) either on the gH at the position T77C, or on the tail, at the position G101C.*

About 147 bp of DNA wraps around the core histone octamers in 1.65 turns, forming the nucleosome core particle (ncp). The two DNA strands emerging out of the core are called the linker-DNA (L-DNA) arms. The junctions at which the DNA 'enters' or 'exits' the core is termed as the 'entry-exit' (EE) sites (Figure 4.1A). These L-DNA stretches link one nucleosome with its neighbouring nucleosomes, and further on, to subsequently build up the chromatin polymer.

The LH associates with the nucleosomal DNA in-between the two emerging L-DNA arms, forming the chromatosome (21). Depending upon whether it associates in an on- or off-dyad

mode, the gH domain can interact with both or one of the L-DNA arms. The highly positively charged C-terminal domain of the LH neutralizes the negative charges of the L-DNA when the LH is associated with the nucleosome. The CTD has been observed to electrostatically interact with either one or both of the L-DNA arms(29, 133–137).

L-DNA length

The nucleosome free L-DNA stretches vary in length from about 10 to 100 bp in different organisms, (138)(139), with an average of 41 bp as observed in HeLa cells (140)(141). The length of the L-DNA and its uniformity over an array of nucleosomes *in vitro* play a key role in determining whether the array will be organised into a one-start or two-start 30nm fibre, or have a disordered arrangement (57, 58, 133, 142–144). The length of the L-DNA has a major implication on LH binding. As summarised in the review by Woodcock *et. al* (145), the length of the L-DNA correlates with the number of LHs bound to the nucleosome. Less than one LH per nucleosome is found *in vivo* for L-DNA lengths of ca. 25 bp, whereas for lengths of ca. 10 bp or less, LHs appear to not bind to nucleosomes. This was experimentally demonstrated by White *et. al* (146) on monochromatosomes (mononucleosome with LH) having L-DNAs of different lengths, concluding that the LH needs at least one L-DNA having a minimum of 11 bp of length, to associate with the nucleosome.

L-DNA sequence

The sequence of the DNA determines its intrinsic curvature (147) which further determines whether such a sequence can wrap around the core histone octamer or not (53, 148, 149). Sequences that favour core histone octamer binding are termed as 'positioning' sequences, of which the 147 bp '601' (53) is a very strongly positioning sequence. While GC-rich DNA, due their flexibility, favour core octamer binding (42), AT-rich DNA on the other hand disfavour core octamer binding due to their rigidity (150–152). Thus, AT-rich DNA, mostly excluded from the core octamer, are largely found in the linker-DNA regions (148). AT-rich L-DNA sequences play a major role in structuring higher-order chromatin by introducing bends (128), and also may have a major implication in the binding of the LH as proposed by Cui *et al.* (49). Contrary to the core histone octamers, the LH has been observed to have an affinity towards AT-rich sequences (43–48). However, the possible role that the sequence of the L-DNA may have on the LH location or orientation, is largely overlooked. In this chapter we will focus on whether (a) the length or (b) the sequence of L-DNA plays a role in the positioning of the LH.

For the first purpose, we designed an asymmetric chromatosome (AsyMG, see chapter 3, Table 3.1 for a description of the constructs, and Appendix A for sequences) with a 24bp and a 40bp L-DNA, and a symmetric chromatosome in which both the L-DNA were 40bp long (MG). As a control, we designed another asymmetric chromatosome in which the lengths of the two L-DNA arms were swapped (AsInvMG), such that the formerly short 24bp arm was extended to 40bp and the formerly long 40bp arm was shortened to 24bp.

To address the second question, we designed symmetric chromatosomes, each having two 40bp long L-DNA arms. The chromatosomes differed from each other only on the L-DNA, at

regions we term as flanks. These are just above the EE site, and extend for about 11bp (Figure 4.1A). The DNA sequences of the two flank regions differ between the constructs, and are either an A-tract, or a purely GC tract, or a mixed DNA sequence having 64% AT content. Chapter 3 details the DNA preparation protocol.

Each chromatosome is labelled with one Alexa 594 acceptor fluorophore, either on the minus (denoted in grey, Figure 4.1A) L-DNA arm, or the plus (purple) L-DNA arm, at about 20 bp from the two EE sites. To detect the position of the LH, the LH was labelled either on the gH domain at residue 77 (T77C site, Figure 4.1B), or on the CTD at residue 101 (G101C site), with Alexa 488 donor fluorophore. Proximity ratio was measured between the gH domain and either of the L-DNA arms or the CTD and either of the L-DNA arms.

Table 4.1: *Chromatosome construct information chart. The length of the minus and plus L-DNA are given, along with the 11bp flank sequences. For TU construct, see chapter 5.*

Construct	Minus L-DNA	Plus L-DNA
AsyMG	Length: 24 bp 5’ – ATACATGCACA – 3’ 3’ – TATGTACGTGT – 5’	Length: 40 bp 5’ - GGGCGGCCGCG – 3’ 3’ - CCCGCCGGCGC – 5’
AsInvMG	Length: 40 bp 5’ – ATACATGCACA – 3’ 3’ – TATGTACGTGT – 5’	Length: 24 bp 5’ - GGGCGGCCGCG – 3’ 3’ - CCCGCCGGCGC – 5’
MG	Length: 40 bp 5’ – ATACATGCACA – 3’ 3’ – TATGTACGTGT – 5’	Length: 40 bp 5’ - GGGCGGCCGCG – 3’ 3’ - CCCGCCGGCGC – 5’
GM	Length: 40 bp 5’ – CGCGGCCGCC – 3’ 3’ - GCGCCGGCGG – 5’	Length: 40 bp 5’ –AGGCATGTAT -3’ 3’ - TCCGTACATA -5’
AG	Length: 40 bp 5’-AAAAAAAAAAA-3’ 3’- TTTTTTTTTTT -5’	Length: 40 bp 5’ – GGGCGGCCGCG – 3’ 3’ - CCCGCCGGCGC – 5’
GA	Length: 40 bp 5’ – CGCGGCCGCCC – 3’ 3’ - GCGCCGGCGGG – 5’	Length: 40 bp 5’ – TTTTTTTTTTT – 3’ 3’ - AAAAAAAAAAA – 5’

4.2 Lengthening L-DNA from 24 to 40 bp does not affect the orientation of the LH

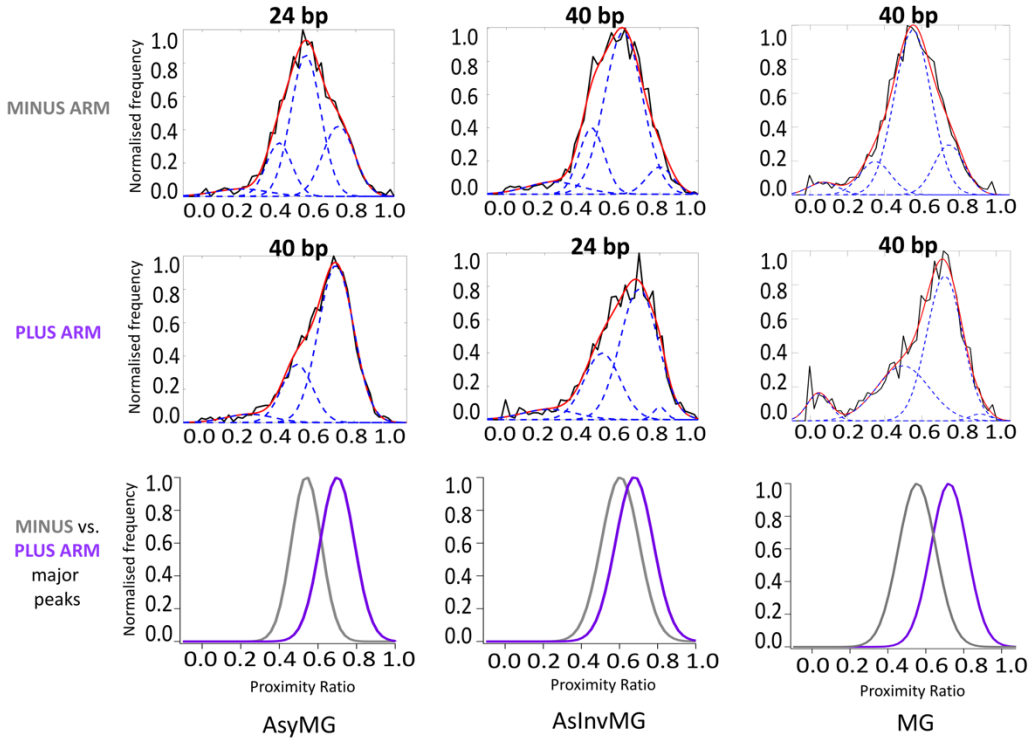


Figure 4.2: *L-DNA length does not affect the gH location. Top and middle rows show average proximity ratio (P) histogram between minus arm and the gH (top) or plus arm and the gH (bottom), with the fits. The black trace is the data averaged from replicates. Blue dashed lines are the subpopulations obtained on fitting the histogram. The red trace is the summation of the fits. L-DNA specifications, such as the length are denoted on the top of each plot. The bottom row shows just the major population peaks (height normalised for comparison) from the fitted data, colour coded to represent the minus arm (grey) or the plus arm (purple). From left to right, construct AsyMG, AsInvMG and MG are shown.*

Table 4.2: *Between the gH and minus and plus L-DNA arms for constructs AsyMG, AsInvMG and MG. Standard deviations calculated from the mean proximity ratio of the major peaks from replicates (See Supplementary Information under Chapter 4 for replicates) are given as errors. The relative population of the major peak is represented as percentages and its range is obtained from replicates.*

construct	Minus label	Plus label
AsyMG	0.54 ± 0.01	0.70 ± 0.01
Relative population (%)	48.6-70.0	61.0-72.0
AsInvMG	0.61 ± 0.01	0.680 ± 0.006
Relative population (%)	62.5-68.0	55.1-60.0
MG	0.55 ± 0.01	0.72 ± 0.02
Relative population (%)	53.0-74.0	55.0-67.2

To investigate whether the length of the L-DNA had any effect on the location or orientation of the LH, two asymmetric chromatosomes were constructed. In both the cases one of the L-DNA arm was 24bp and the other, 40bp. In AsyMG, the minus arm was the short arm. In AsInvMG, the lengths were inversed, i.e., the minus arm was extended to 40bp and the plus arm (denoted in dashed purple line) was shortened. The MG construct was symmetric, with two 40bp L-DNA arms (Table 4.1). LH labelled on the gH domain or the CTD domain (site T77C or site G101C) was added to the nucleosomes labelled 20bp from the EE site, on one or the other L-DNA arm, during reconstitution (Chapter 3 details sample preparation). The proximity ratio histograms obtained, were fitted into multiple Gaussian peaks, with each peak corresponding to a particular species. The mean proximity ratios of the dominant peaks are tabulated in Table 4.2.

Overlaying the proximity ratio histograms obtained between the gH and the minus (grey), and plus (purple) L-DNA arms shows that the gH is closer to the 40bp plus L-DNA than to the 24bp minus L-DNA in AsyMG chromatosome (Figure 4.2 first column). The preference of LH for a longer L-DNA arm could be due to the energetic favourability of associating electrostatically with a longer arm with respect to a shorter one. However, on testing with AsInvMG, we found that the gH was not similarly closer to the now 40bp minus arm, with respect to the now shortened plus arm (Figure 4.2 second column). In fact, it was still slightly closer to the shortened plus arm (purple plot at the bottom of second column) than with the elongated minus arm, although this difference is slight (ΔP_{mean}=0.07, see Table 4.2). On extending both the arms to 40bp (MG construct, Figure 4.2 third column), we expected the gH to prefer both the arms equally and the proximity ratio histograms to overlap. But the proximity ratio histogram of the gH with the plus arm is once again shifted to the right. Thus, the 40bp plus arm of the AsyMG shows a similar proximity ratio profile as that of the 40bp plus arm of the MG construct, and the 24bp plus arm of the AsInvMG construct (Table 4.2).

This shows that the gH prefers the plus arm irrespective of its length.

Table 4.3: *Between the CTD and minus and plus L-DNA arms*

construct	Minus label	Plus label
AsyMG	0.700 ± 0.006	0.63 ± 0.01
Relative population (%)	72.2-80.0	62.2-68.6
AsInvMG	0.69 ± 0.02	0.66 ± 0.01
Relative population (%)	76.0-86.4	58.0-72.0
MG	0.690 ± 0.006	0.67 ± 0.01
Relative population (%)	68.0-84.0	64.0-83.0

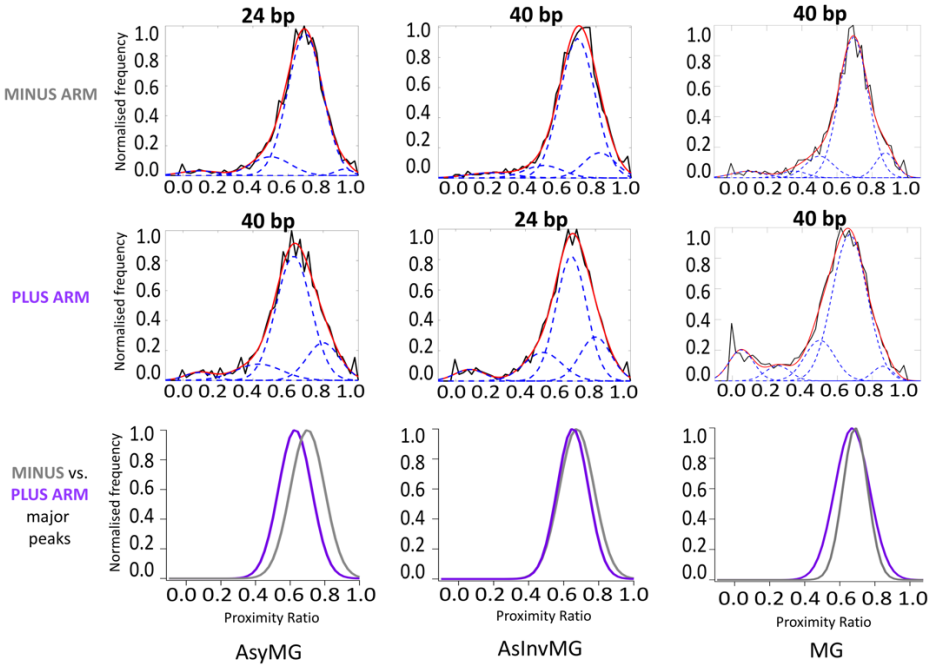


Figure 4.3: *L-DNA length does not affect the CTD location. Top and middle rows show average proximity ratio (P) histogram between minus arm and the CTD (top) or plus arm and the CTD (bottom), with subpopulations. The black trace is the data averaged from replicates. Blue dashed lines are the subpopulations obtained on fitting the histogram. The red trace is the summation of the subpopulations. L-DNA specifications, such as the length are denoted on the top of each plot. The bottom row shows just the major population peaks (height normalised for comparison) from the fitted data, colour coded to represent the minus arm (grey) or the plus arm (purple). From left to right, construct AsyMG, AsInvMG and MG are shown.*

As seen in Figure 4.3, in all the constructs, the proximity ratio profiles of the tail fluorophore (site G101C on CTD) with the minus and plus arms nearly overlap. The ΔP_{mean} between the minus and plus arms are 0.03 to 0.04. This shows that the tail fluorophore is nearly equidistant from the minus and the plus arms.

4.3 Neither the gH nor the CTD have any preference for AT-rich sequence over GC-tracts

Table 4.4: *Between the gH or CTD and minus and plus L-DNA arms in GM construct.*

GM construct	Minus label	Plus label
gH	0.57 ± 0.02	0.60 ± 0.03
Relative population (%)	58.0-68.0	58.0-68.0
CTD	0.68 ± 0.02	0.70 ± 0.01
Relative population (%)	54.3-71.3	74.0-79.0

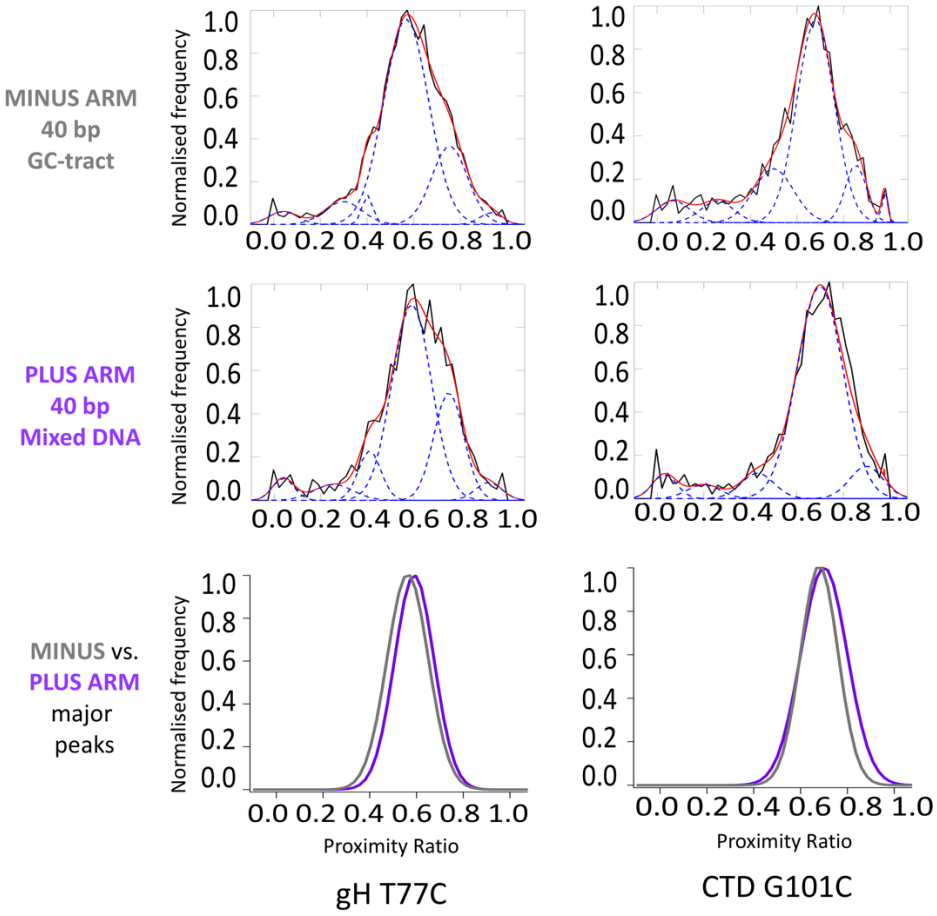


Figure 4.4: *gH and CTD does not prefer GC-tracts over AT-rich mixed DNA sequences. Proximity ratio histograms from the GM construct is shown. Left column shows proximity ratio histograms between minus arm (top) and the gH, and between plus arm (middle) and the gH. Right column shows proximity ratio histograms between the L-DNA arms and the CTD. Averaged proximity ratio histograms on the top and the middle show the averaged data (black), subpopulations (blue dashed lines), and the summation of subpopulations (red trace). L-DNA specifications, such as the sequence (mixed/GC-tract) are mentioned on the extreme left. The bottom row shows just the major population peaks (height normalised for comparison) from the fitted data, colour coded to represent the minus arm (grey) or the plus arm (purple).*

The Widom 601 sequence used for the constructs contains on the plus arm just outside the ncp, a purely GC stretch of about 11bp (Table 4.1). On the minus arm, the same length of DNA emerging from the ncp has a mixed sequence with a pyrimidine-purine base step. This mixed sequence is 64% AT and 36% GC. To make sure that the observed high proximity ratio peaks in the plus arm plots (Figure 4.2) was not caused by any sequence preference of gH for GC-rich stretches, we swapped the two 'flank' sequences on the plus and minus arm for the MG construct, resulting in a new construct, GM, where the minus arm has the pure GC stretch and the plus arm has the mixed 64% AT-rich stretch. Figure 4.4 first column shows the proximity ratio histograms for the minus and plus L-DNA with the gH and second column shows that with the CTD label. Each of the plots show a dominant population, and minus and plus arm plots are nearly overlayed both in the case of the gH fluorophore as well as the CTD-fluorophore. This shows that the gH as well as the CTD fluorophore sit equidistant from the two L-DNA arms. Thus, neither the gH fluorophore, nor the CTD fluorophore show any preference for either the pure GC stretch, or the 64% AT-rich mixed DNA stretch.

4.4 L-DNA flank sequences do not affect the CTD

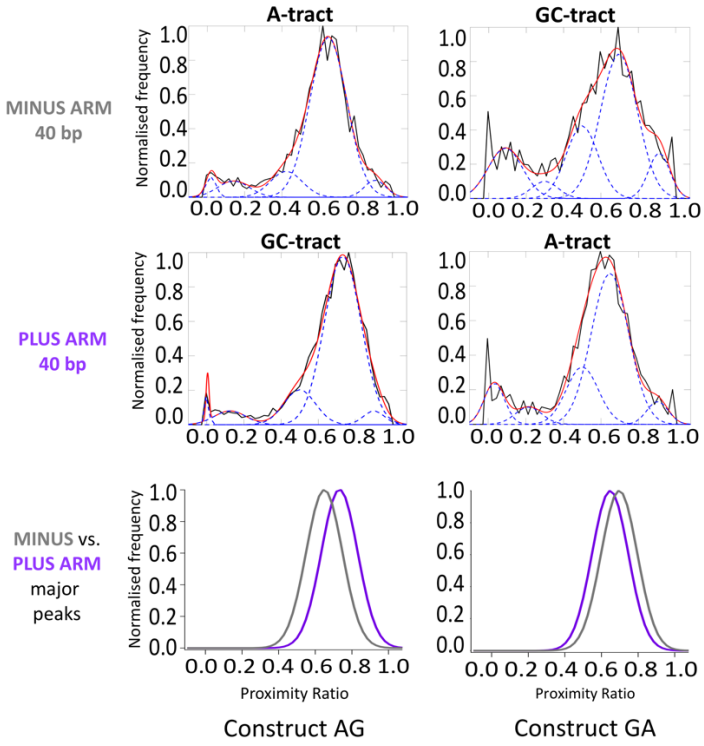


Figure 4.5: *L-DNA sequence does not affect the CTD location. Proximity ratio histograms from the AG construct (left) and the GA construct (right) is shown. Top row show proximity ratio histograms between minus arm and the CTD, and between plus arm (middle) and the CTD. Averaged proximity ratio histograms on the top and the middle show the averaged data (black), subpopulations (blue dashed lines), and the summation of subpopulations (red trace). L-DNA specifications, such as the sequence (A-/GC-tract) are mentioned on the tops of the averaged proximity ratio plots. The bottom row shows just the major population peaks (height normalised for comparison) from the fitted data, colour coded to represent the minus arm (grey) or the plus arm (purple).*

Table 4.5: *Between the CTD and minus and plus L-DNA arms*

constructs	Minus label	Plus label
AG	0.65 ± 0.01	0.73 ± 0.01
Relative population (%)	60.0-81.0	70.2-84.0
GA	0.70 ± 0.02	0.65 ± 0.02
Relative population (%)	37.0-57.6	55.0-81.0

The asymmetric constructs and the symmetric constructs MG and GM revealed that the LH (both gH and CTD) showed no preference for the pure GC stretch. This was expected from previous studies showing that the LH does not have any affinity for GC-rich sequences (47, 51, 134). But contrary to earlier studies showing that the LH has affinity towards AT-rich sequences (44, 45), our experiments showed no preference of the gH or the CTD for the 64% AT-rich region. New constructs were built from the DNA templates of MG and GM, replacing the 11bp 64% AT/36% GC mixed sequence flanks with an A-tract, i.e., a series of 11 adenines paired to 11 thymines on the complementary strand. The pure GC-tract flank on the other L-DNAs were left unchanged. These new constructs are the AG and GA (detailed in Table 3.1 in Chapter 3 and Table 4.1). The effect of the A-tracts on the CTD was first investigated. As we see in Figure 4.5, first column, the CTD fluorophore is closer to the GC-tract containing plus L-DNA arm in AG. When the flank sequences are swapped in the GA construct, the proximity ratio profiles of the minus and plus L-DNA arms with the CTD fluorophore nearly overlap. This states that for the GA construct, the CTD fluorophore is nearly equidistant from both the arms.

Thus, the high FRET population exhibited by the GC-tract plus arm in AG is not reproduced by the GC-tract minus arm in GA construct, suggesting that this could just be a non-specific orientational effect of the fluorophore. We next study if the A-tracts have any effect on the gH domain.

4.5 gH shows a higher preference for A-tracts

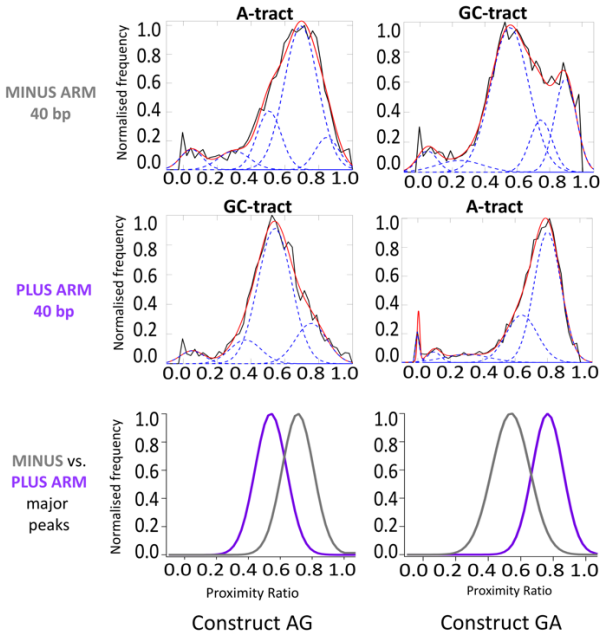


Figure 4.6: *gH shows a higher FRET for A-tracts. Proximity ratio histograms from the AG construct (left) and the GA construct (right) is shown. Top row show proximity ratio histograms between minus arm and the gH, and between plus arm (middle) and the gH. Averaged proximity ratio histograms on the top and the middle show the averaged data (black), subpopulations (blue dashed lines), and the summation of subpopulations (red trace). L-DNA specifications, such as the sequence (A-/GC-tract) are mentioned on the tops of the averaged proximity ratio plots. The bottom row shows just the major population peaks (height normalised for comparison) from the fitted data, colour coded to represent the minus arm (grey) or the plus arm (purple).*

Table 4.6: *Between the gH and minus and plus L-DNA arms*

construct	Minus label	Plus label
AG	0.710 ± 0.005	0.540 ± 0.005
Relative population (%)	51.0-69.0	61.0-67.0
GA	0.55 ± 0.01 [I] 0.89 ± 0.02 [II]	0.80 ± 0.01
Relative population (%)	55.0-61.0 [I] 19.0-24.0 [II]	57.6-83.0

LH labelled at site T77C or the gH domain was used to reconstitute AG and GA chromatosomes to study the possible effect of A-tracts on the gH domain. We clearly observed that the gH fluorophore showed higher proximity ratio with L-DNA arms that contained A-tracts, such as the minus arm of the AG and the plus arm of the GA constructs (Figure 4.6 left and right). Each of the proximity ratio histograms of AG chromatosome (Figure 4.6 left) show a dominant population. The ΔP between the plus or minus arm labels and the gH is nearly 0.2 units. This translates to a ΔR of 5Å (chapter 5 table 5.1 lists distance values). This is quite a significant difference and could be a result of different positioning of the gH with respect to the plus and

minus L-DNA arms.

In the GA chromatosome, the GC-tract containing minus L-DNA arm shows two populations (Figure 4.6 right). However, the low FRET peak with a mean proximity ratio of 0.58 (Table 4.6) shows up as the major population and the high FRET peak (P = 0.89) shows up as 20% of the total population. The plus arm also shows a dominant and one minor population, with the minor peak varying between 13 to 38% of the total population in the replicates.

The consistent high proximity ratio shown by the gH fluorophore for the A-tracts as opposed to the GC-tracts reflect earlier observations suggesting that the LH has a preference for not just AT-rich DNA but also to A-tract regions having adenines continuously for over 4 to 6 base pairs (43, 44, 46–48, 128).

To find out whether the observed differences in the proximity ratio histograms between the plus and the minus arms of AG and GA constructs and the gH is due to a change in the orientation of the gH or its binding mode, we modeled the constructs AG and GA with the gH domain. Using the software FRET Positioning and Screening (FPS) (131), we modeled the inter-fluorophore distances as obtained from our FRET experiments, to explain our observations. This is detailed in chapter 5. Modeling details are written in chapter 2 and 3.

4.6 Concluding remarks

From our observations we conclude the following aspects of DNA dependent LH binding. Although the gH domain shows a higher proximity ratio for the longer, 40 bp plus arm in AsyMG construct, a shortening in its length to 24 bp (AsInvMG) does not affect the gH domain. The gH domain does not show a higher proximity ratio for the minus arm elongated to 40 bp.

The CTD fluorophore is also unaffected by the differences in L-DNA length and lies nearly equidistant from both the L-DNA arms. However, the CTD-fluorophore is just 6 residues away from the gH domain and cannot account for the nearly 100 residue, disordered tail. Thus, we cannot conclude how the rest of the tail may associate with the L-DNA arms.

The sequence of the L-DNA arms affects the proximity of the gH domain, but once again, not the CTD fluorophore. The gH shows an apparent proximity towards GC-tracts when the opposite arm contained a mixed DNA sequence that was 64% AT-rich. However, the proximity towards GC-tract was ruled out by inversing the flank sequences. By replacing the 64% AT-rich flanks with A-tracts, we observed that the gH showed proximity towards the A-tract sequences. According to the Modeling studies from Cui and Zhurkin (49), and later, from Öztürk *et al.* (153), the gH should interact with the thymine methyl groups hydrophobically. If this is indeed the only basis of recognition of AT-rich DNA by the LH, then why does our observation with 64% AT-rich mixed DNA (construct MG and GM) show otherwise? To target this problem, we modeled the AG and GA chromatosomes. Additionally, we studied whether the LH recognizes A-tracts solely by hydrophobic interactions. These studies are detailed in Chapter 5.

Chapter 5

How does the linker histone recognize A-tracts?

5.1 Introduction

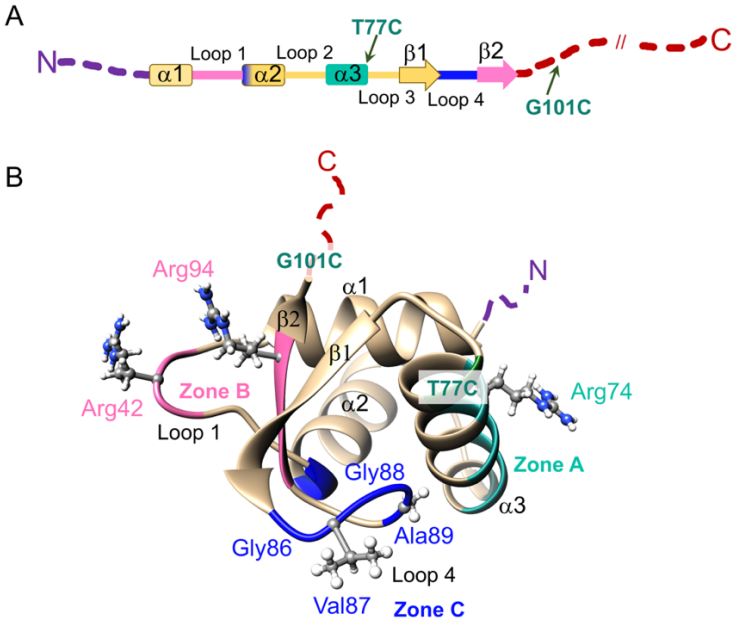


Figure 5.1: *DNA interacting zones on the globular domain of LH.* ***A.*** *Linear representation of the full-length LH. Alpha helices α1, α2, α3, are represented as cylinders and beta sheets β1 and β2 is represented by arrows. Labelling positions, T77C (gH-fluorophore) and G101C (CTD-fluorophore) are denoted in green. The different DNA interaction zones are coloured as: A in sea-green, B in pink, and C in blue.* ***B.*** *Folded full-length linker histone showing the globular domain (gH), the 100 amino acid long, intrinsically disordered CTD (represented by maroon dashed line), and the shorter, disordered NTD (purple dashed line). DNA interacting zones with certain residues are shown: Zone A or α3 helix with the R74 residue, Zone B comprising of loop 1 and β2 with arginines 42 and 94, and Zone C beta-hairpin loop containing the hydrophobic 'GVGA' motif.*

A large number of isoforms of the LH exist in different organisms. While the general tripartite structure of the LH remains unchanged, and the amino acid sequence is largely conserved, especially in the gH domain, there are subtler changes in the amino acid sequence in-between the subtypes. These subtle changes can bring about very large effects on how the LH associates with the nucleosome, as was demonstrated by Bai and colleagues (31) using LH mutated at certain residues that determined on vs. off-dyad binding. Earlier FRAP and mutagenesis studies of the LH also clearly established the importance of certain amino acid residues on the binding of the LH to the nucleosome (154–156).

High resolution x-ray crystallographic studies, complemented with advancements in cryo-electron microscopy complemented these studies to establish a comprehensive understanding of

how the amino acid residues aid in the association of the LH. XRC and cryoEM structures (27, 29, 134, 157) (PDB ID 5NL0, 4QLC, 5WCU, 7K5X) of on-dyad positioned gH domains show three DNA binding domains on the LH (Figure 5.1B). The zone A on the α3 helix and the zone B on the loop 1 and β2 have been observed to associate with the two L-DNAs while the zone C has been observed to associate with the dyad DNA (25, 27, 29, 134, 157). The zone C, contains a stretch of hydrophobic 'GVGA' motif. Although this motif has largely been observed to associate with the ncp DNA, it was proposed that this hydrophobic region could interact with thymine methyl groups, thereby identifying AT-rich region that are found mostly at the entry-exit sites of nucleosomes (49, 50).

In chapter 4 we observed that the gH consistently showed high FRET for A-tracts but not so for a mixed DNA sequence that was 64% AT-rich. In this chapter we address how the LH can identify A-tracts, why cannot it identify any other mixed sequence, and whether it is just the zone C hydrophobic region involved in the A-tract recognition or are there any other mechanisms involved.

5.2 Conserved arginines in H1.0 subtypes recognize A-tracts

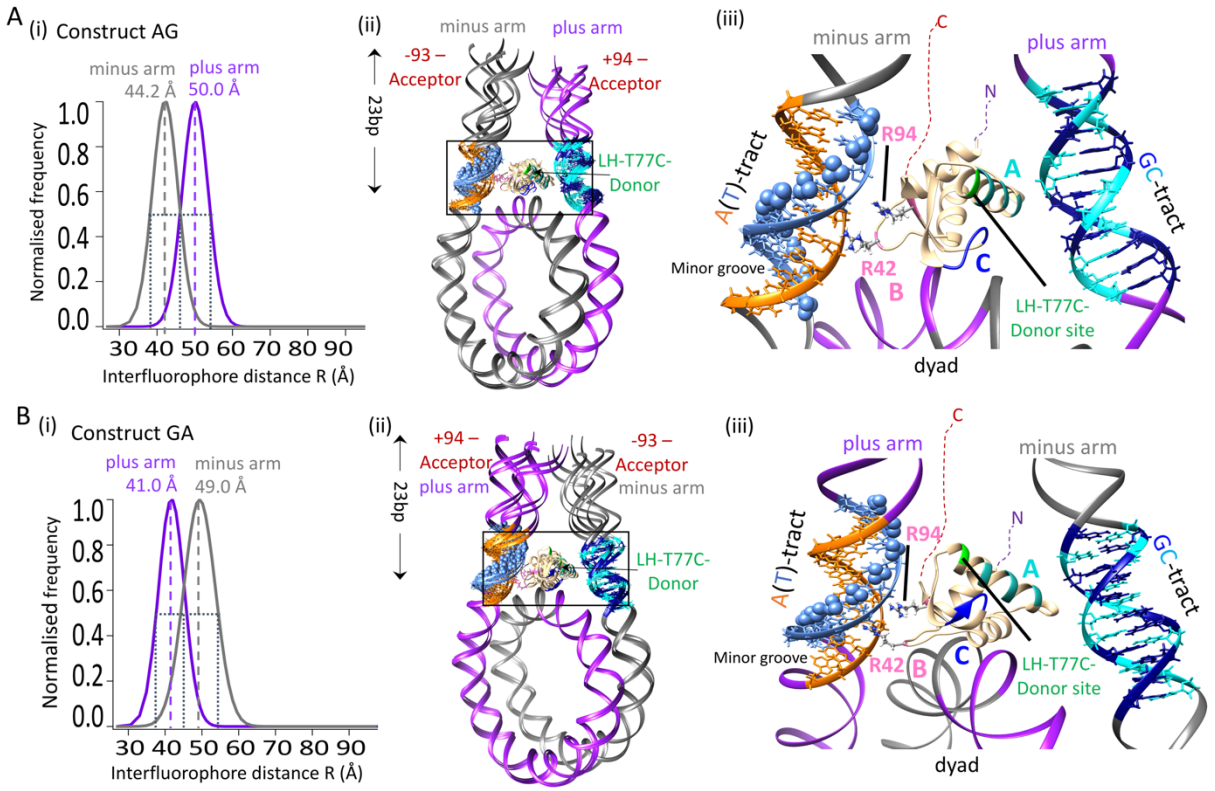


Figure 5.2: *AG* ***(A.)*** *and GA* ***(B.)*** *chromatosomes: distance histograms (i) and FRET-restrained models (ii and iii, zoomed in). Four different L-DNA arm opening (**Aii** and **Bii**) were modeled to account for the L-DNA dynamics shown by the peak width at half maxima (**Ai** and **Bi**). Both the models show that the zone B of the gH is oriented towards and proximal to the A-tract minor grooves. The DNA is colour coded as in the plots, i.e., minus arm is grey and plus arm is purple. The flank sequences are shown as sticks and are colour coded according to the type of base: blue for thymine, orange for adenine, deep blue for guanine, cyan for cytosine. Thymine methyl groups are shown as blue spheres. The gH is coloured according to that shown in Figure 5.1B.*

Table 5.1: *Comparison of* inter-*fluorophore distances calculated from proximity ratio plots (see Supplementary Information under Chapter 5) (standard deviation obtained from replicates) and model shown in Figure 5.2. Values in bold are the computed distances from the model shown in 5.2Aiii and Biii. All values are in Å.*

	AG (-93) gH-T77C	AG (+94) gH-T77C	GA (-93) gH-T77C	GA (+94) gH-T77C
Experiment R_{mean} (replicates) ± standard deviation of R_{mean}	44.2±0.84	50.0±0.24	49.0±0.74	41.0±0.74
Experiment half-width at half maxima	6.0	6.0	7.0	6.0
Model: 5NL0-type	**44.6** 46.6 47.7	**51.1** 53.3 54.3	47.4 54.4	42.5 48.5
Model: 7K5X-type	48.2 50.3 51.6	54.4 57.3 56.6	44.9 53.0	42.8 50.0
Model: NMA-mode 7 frame 51	43.5 45.0 46.3	54.0 56.2 57.0	**50.2** 58.0	**40.6** 46.0
Model: NMA-mode 8 frame 1	50.2 52.0 54.1	53.6 56.5 56.0	50.0 57.1	46.0 53.0

The extreme left column of Figure 5.2 A and B show only the major population peaks of the distance histograms that are obtained from the proximity ratio histograms. Detailed workflow of how the inter-fluorophore distances were obtained from the proximity ratio plots are given in chapter 2 and 3. The right column shows the AG (A) and GA (B) chromatosomes modeled using Chimera (120) and FPS (131). For the GA chromatosome, only the major population was modeled.

To account for the peak width of the major populations in the distance plots, L-DNA opening dynamics were captured by using chromatosome structures having two different types of L-DNA arm conformations (5NL0 (29) and 7K5X (30)). Two more arm conformations were generated by subjecting the 5NL0 to normal mode analysis. Chimera (120) was used to place the gH in different orientations on the nucleosomes, generating different configurations of chromatosomes of each arm-opening type. Using FPS, the entire ensemble of conformations was screened with the experimental distances as input. Table 5.1 shows that the models conform well to our experimental results.

We observe that for both the AG and GA models, the gH associates in an on-dyad binding mode. Since we use a LH of H1.0 subtype, this binding mode is expected, as evidenced in numerous previous studies (27, 29, 134, 157, 158). To prove this further, some alternate off-dyad conformations were also modeled (Figure 5.3C and D). In these off-dyad models the gH was positioned in a way as to make the Zone C proximal to the major groove (49, 50) of the A-tract on the minus arm of AG (in grey, Figure 5.3C) or the plus arm of GA (in purple, Figure 5.3D). The computed inter-fluorophore distances did not conform with our experimentally derived distances (Table 5.3), suggesting that our results do not show an off-dyad binding mode.

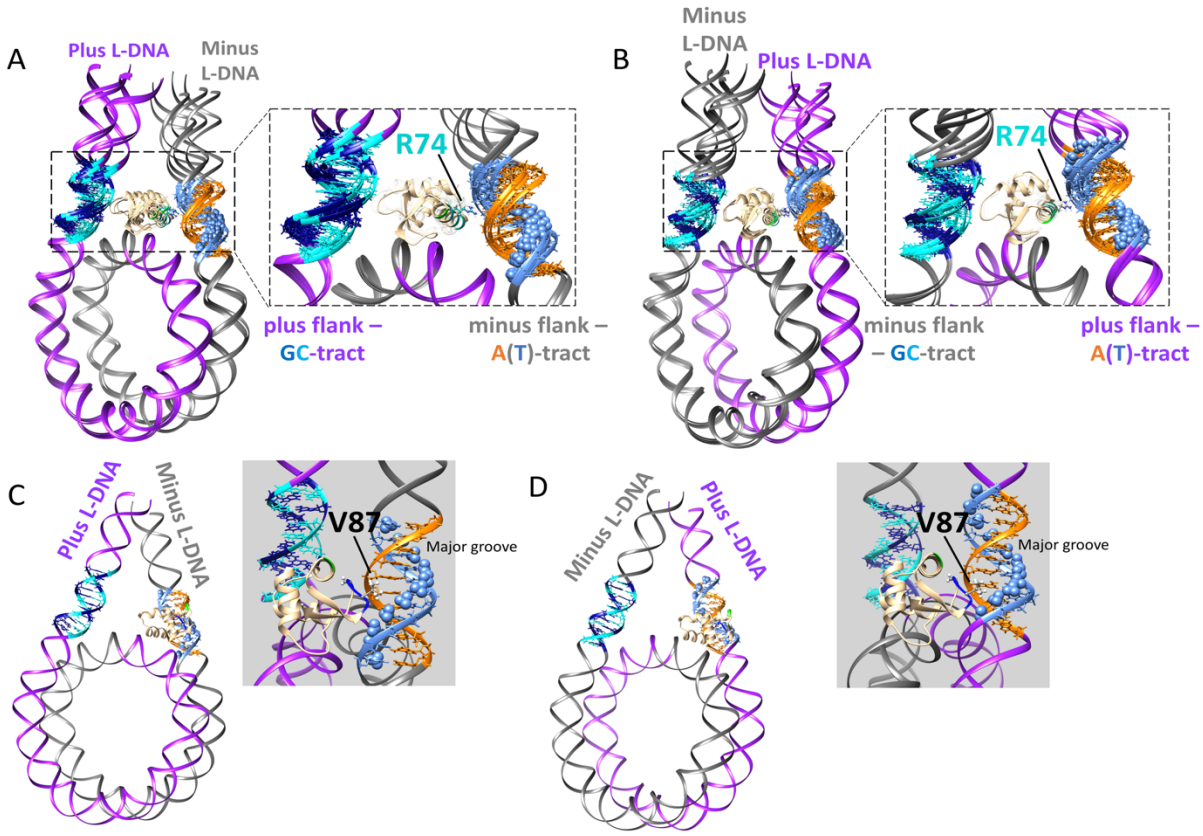


Figure 5.3: *AG and GA chromatosomes: alternate models.* ***A.*** *AG and* ***B.*** *GA AG(i) and GA (ii) with the Zone A (R74) of the gH placed proximal to the A-tract minor groove, and the Zone B facing the opposite L-DNA flank.* ***C.*** *AG and* ***D****. GA with the gH placed off-dyad, proximal to A-tract, and the Zone C poised to interact with the thymine methyl groups [blue spheres] on the major groove.*

Table 5.2: *Distances (in Å) computed from alternate model ensembles (Figure 5.3A and B) not fitting with the experiment.*

construct	L-DNA conformation	Minus label	Plus label
AG [zone A R74-A(T)-tract]	5NL0-type	60.0±1.0	54.0±1.0
	7K5X-type	59.0±1.0	56.0±0.6
	Mode 7 frame 51	63.0±1.1	51.0±1.0
	Mode 8 frame 1	62.5±1.6	59.4±1.0
GA [zone A R74-A(T)-tract]	5NL0-type	52.0±1.6	59.0±1.4
	7K5X-type	56.0±1.7	61.0±1.4
	Mode 7 frame 51	49.2±1.5	61.4±1.3
	Mode 8 frame 1	59.0±1.8	61.0±1.6

Table 5.3: *Distances (in Å) computed from alternate off-dyad models (Figure 5.3C and D).*

AG [gH off-dyad, A-tract minor groove]	56.0 with minus label
GA [gH off-dyad, A-tract minor groove]	57.0 with plus label

The AG and GA models in Figure 5.2 show that although the gH domain binds on-dyad it has a specific orientation in both. The zone B (in pink, Figure 5.2Aiii and Biii) of the gH is proximal to the minor groove of the A-tracts. But what if this orientation is an artefact brought about by repulsion between the negatively charged Alexa 488 fluorophore on the α3 helix and the L-DNA backbone that forces the fluorophore attachment site into the gap between the two L-DNAs, on the plane of the chromatosome, and turns the Zone B towards the A-tracts and not the other L-DNA interacting Zone A. To rule this out we thus modeled another set of alternatives, turning the gH 180° opposite to that shown in Figure 5.2A and B, to position the Zone A proximal to the A-tract. In this pose too the fluorophore is free to rotate, being mostly situated in the gap between the two L-DNAs, suggesting that our results are not biased by the labelling position of the gH fluorophore. These two alternate models are shown in Figure 5.3A and B. The computed inter-fluorophore distances in these models do not conform with our experimentally derived distances (table 5.2). This proves that the orientation of gH that satisfies our experimental results is a significant and possibly a biologically relevant observation and is not an artefact.

DNA sequences are recognized by proteins either directly, via interactions with the bases at the major groove, or indirectly, exploiting the sequence induced geometric parameters of the DNA (159). One such indirect sequence recognition mechanism involves the minor groove. A-tract regions have the smallest minor groove widths, while GC-tracts have the widest (160). The narrowing of the minor groove brings the two sugar-phosphate backbone in close proximity, increasing the negative charge density in A-tracts and making it ideal for recognition by positive amino acid residues. Among the positive amino acid residues that can interact electrostatically with the highly negative narrow minor groove of A-tracts, arginines are observed to be ideally suited (43, 159, 161, 162). This could be because the side chain of arginine contains a guanidium group that can form multiple hydrogen bonds with the DNA, more than that formed by the side chain of lysine (159, 161, 163, 164). Additionally, Rohs *et al.* (159) came up with an alternate

theory of arginine-minor groove interactions that does not involve hydrogen bonding, stating that the energy penalty of removing the charged guanidium group of arginines from water is less than that required to remove the amino group of lysine.

It is therefore not surprising that the zone B that reproducibly orients towards the minor groove of the A-tracts contain two highly conserved arginine residues (R42 and R94) positioned in a pincer-like motif (Figure 5.2). All H1.0 subtypes contain these two arginines in the zone B region (Figure 5.1 and Chapter 7 Table 7). The arginine on the loop 1 region (R42) is conserved in all LH subtypes across organisms. These two arginine residues have been observed to have a major effect in the LH binding to the nucleosome, as demonstrated by Brown *et al. in-vivo*, using FRAP (154), and subsequently also by Öztürk *et al. in silico* (165). It was suggested by Zhou et al (2021) (30) that the Zone A arginine could also identify At-rich DNA. However, our alternate model with the Zone A proximal to the A-tract show that this is not the case.

The fact that the on-dyad associating H1.0b interacts with the A-tract with the minor groove explains why it could not recognize the 64% AT-rich mixed DNA flanks on the asymmetric, MG, and GM constructs (chapter 4). The minor grooves of a mixed DNA sequence are not as narrow as A-tracts (160) and therefore has less negative charge density. This rendered it unrecognizable by the arginines on zone B.

To test our theory that A-tracts are recognized via their minor grooves and not the methyl groups, we reconstituted a final construct (TU) containing A-tracts base paired to thymines as well as A-tracts base paired to deoxy-uridines. This is detailed in the next section.

5.3 gH shows the same preference for both A-tracts paired to thymines as well as to deoxyuridines

A possible mode of recognition of AT-rich sequences by the LH was speculated by Cui and Zhurkin in 2009 (49) and was then shown *in silico* by Öztürk *et al.* (50). In this mode the LH associates off-dyad with the nucleosome. The non-polar residues on the Zone C interact hydrophobically with the thymine methyl groups that protrude on the major groove of the L-DNA which can be accessed by the gH when it is associating in an off-dyad mode. Since our LH, H1.0b has been demonstrated to associate with an on-dyad binding mode (27, 29, 134, 157), this could be a reason why the gH showed no preference for the mixed, 64% AT-rich flank on the asymmetric constructs and on the symmetric MG and GM chromatosomes (chapter 4). On replacing the 64% AT-rich mixed flank with an A-tract, the gH domain showed a consistent preference towards the A-tract in the AG and GA chromatosomes (chapter 4). The question here is, if the LH is indeed associating in an on-dyad fashion, how can it interact with the thymine methyl groups on the major groove of the A-tract. Does this mean that there could be another mode of interaction of AT-rich sequences, depending on the mode of association of the gH?

Table 5.4: *TU construct studied.*

Construct	Minus L-DNA	Plus L-DNA
TU	Length: 40 bp 5'-AAAAAAAAAAA-3' 3'- TTTTTTTTTTT -5'	Length: 40 bp 5'-AAAAAAAAAAA-3' 3'- UUUUUUUUUUU -5'

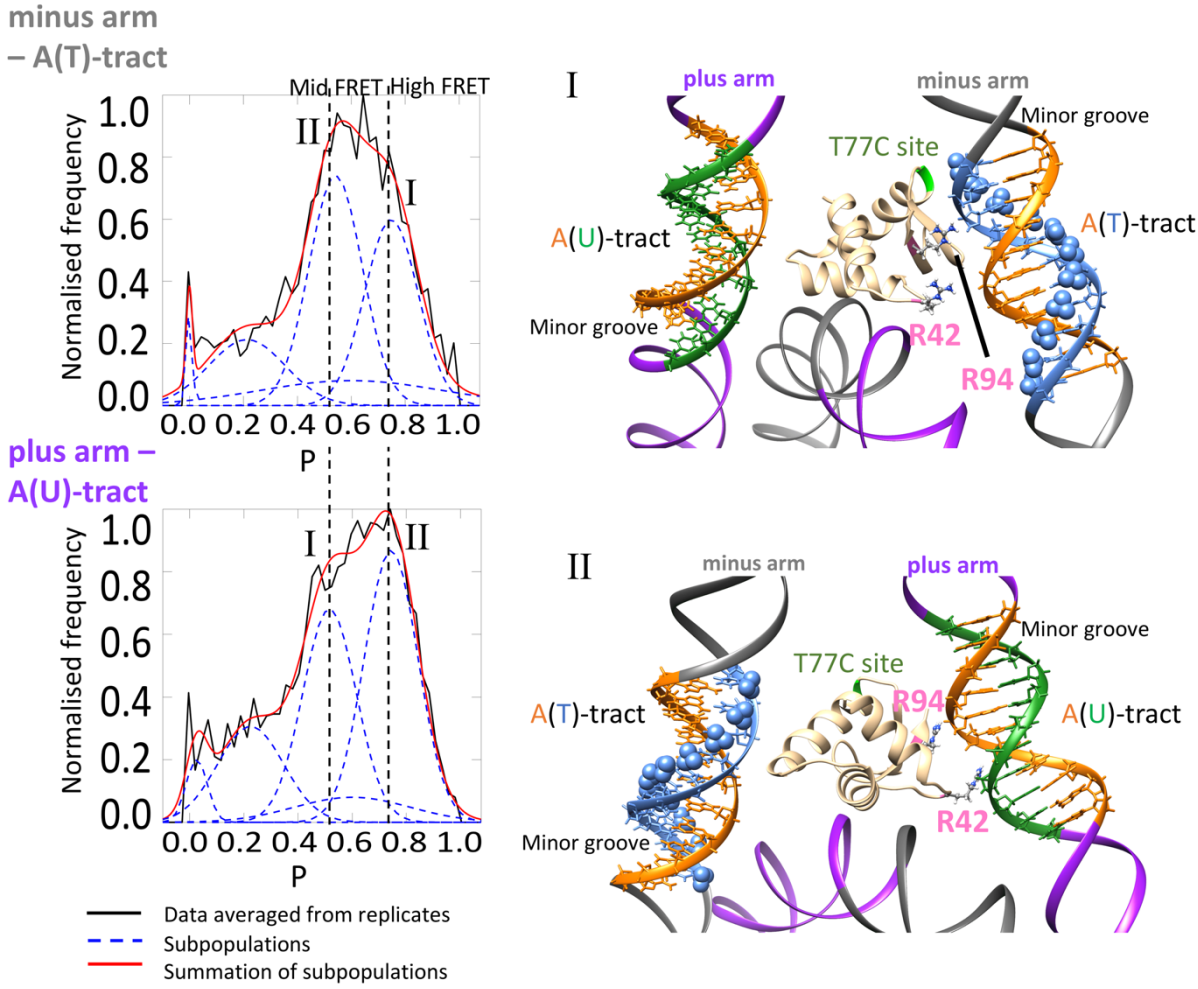


Figure 5.4: *Further evidence of minor-groove based recognition of A-tracts. On the left are proximity ratio histograms between the gH and the L-DNA arms in the TU chromatosomes, with their fitted plots. On the right are the two possible models of the TU chromatosome that correspond to the subpopulations seen in both the proximity ratio histograms. Shown on the top is subpopulation I which shows higher FRET with A(T)-tract (complementary to thymine) and lower FRET with the A(U)-tract (complementary to deoxyuridine). On the bottom is subpopulation II which shows lower FRET with A(T)-tract and higher FRET with A(U)-tract.*

Table 5.5: *Experimental FRET efficiency of construct TU (standard deviations of mean P obtained from replicates, see Supplementary Information Chapter 5), compared to computed FRET efficiencies*

construct	Minus label	Plus label
TU gH label	0.54 ± 0.03 [II] 0.75 ± 0.03 [I]	0.52 ± 0.03 [I] 0.74 ± 0.02 [II]
Relative population (%)	28.0-42.0 [II] 30.0-36.0 [I]	30.0-40.0 [I] 38.0-50.0 [II]
Computed P Model I	0.77	0.53
Computed P Model II	0.55	0.72

This problem prompted us to investigate whether the gH was indeed interacting with the methyl group of thymine. To do this, we modified the AG construct by replacing the GC-tract on the plus L-DNA with another A-tract, but this one complementary to a stretch of 11 deoxyuridines. In this chromatosome, the minus arm has a continuous stretch of thymine methyl groups lining the major groove but the deoxy-uridine containing plus arm lacks the methyl groups.

If the gH domain recognized A-tracts or AT-rich region by the thymine methyl groups then we would have observed higher FRET for the minus arm and a lower FRET for the plus arm. Figure 5.4 first column shows that it is not so. The proximity ratio profiles between the gH domain and the plus (in purple) and minus (in grey) arms exactly overlap. However, these chromatosome peaks (P 0.3 to 0.9) for both the minus and plus arms are broad and can be fitted into two populations modeled in the second column of Figure 5.4. In the proximity ratio plot between the A(T)-tract minus arm and the gH, the high FRET peak corresponds to a population where the gH is oriented towards the A(T)-tract. This peak is numbered I. In the proximity ratio plot between the A(U)-tract plus arm and the gH, this subpopulation corresponds to the intermediate FRET peak (Figure 5.4 first column bottom). Similarly, the subpopulation of chromatosomes with the gH oriented towards the A(U)-tract would have a higher FRET peak in the proximity ratio plot between A(U)-tract and gH and an intermediate FRET peak in the proximity ratio plot between A(T)-tract and gH. These subpopulations are numbered II. The chromatosome models in the second column show the corresponding subpopulations. Model I show the TU chromatosome with the Zone B (R42/94) of the on-dyad positioned gH oriented towards the minor groove of the A(T)-tract (adenines in orange, thymines in blue with the methyl group in sphere representation). Model II (Figure 5.4 second column bottom) show that the Zone B of the gH is oriented towards the minor groove of the A(U)-tract (deoxy-uridines represented in green). Using the software FPS (131) we computed the theoretical FRET efficiency between the minus and plus arms and the gH label for both the models, and compared them with the mean proximity ratio of the two subpopulations I and II in each proximity ratio plots. Table 5.5 show that the computed FRET efficiencies between the arms and the gH in both the models are very similar to the mean proximity ratio of the peaks I and II in both the proximity ratio plots.

This shows that an on-dyad positioned gH does not distinguish flank sequences based on their functional groups (direct sequence recognition) but rather identifies DNA sequences via their effects on the physical properties of the DNA, such as the narrowing of the minor groove by A-tracts. Furthermore, both the thymine and the uridine containing tracts have similar minor groove widths, as despite the lack of methyl group, uridines in DNA have similar geometric

parameters as thymine (166).

5.4 Concluding remarks

Arginines are known to play a very important role in the recognition of A-tracts. Proteins such as the Integration Host Factor (159), protamines (48), and the core histone octamers are known to employ arginines to contact the DNA at the minor groove (159). In fact, in-vivo nucleosomal sequences have been observed to alternate in GC-rich stretches and short 3bp A-tract regions. In parallel, structural studies have shed light on how arginines from the core histone can contact these short A-tract regions (159, 162). Thus, it is not surprising that the LH too, widely observed to have an affinity for AT-rich regions and A-tracts, will employ a similar mechanism. But previously it was also speculated and computationally observed that the gH can recognize thymine methyl groups exposed on the major groove, directly, via hydrophobic interactions (49, 50). We suggest, that, more than one mechanism of recognizing and interacting with AT-rich DNA are found in LHs. If the LH associates on-dyad, it can access the A-tract via its minor groove and employ the arginines at zone B, or at zone A (as suggested by Zhou et al (2021) (30)) to recognize the A-tract indirectly via its narrow minor groove. If it associates in an off-dyad fashion, it can access the thymine methyl groups lining the major groove of the L-DNA via its hydrophobic Zone C. This could be an adaptation necessary for multiple isoforms of the LH to associate with and condense AT-rich constitutive heterochromatin (167, 168).

But does the L-DNA sequence-induced orientation of LH and its neighbours, affect higher-order structures? We address this question in chapter 6 using a trichromatosome setup having A-tracts at specific sites on the L-DNA arms.

Chapter 6

Effect of A-tract dependent LH orientation on trichromatosome compaction

6.1 Introduction

The location of the LH on the nucleosome has implications for the structuring of higher-order chromatin. Experimental and computational studies have so far focused on how the two modes of association of the gH domain (on-dyad or off-dyad) with the nucleosome affect the compaction of nucleosome arrays. While cryo-EM studies (133) on tetranucleosomes showed that an off-dyad binding mode of the gH leads to a higher compaction of chromatin, analytical ultracentrifugation studies (27) have showed that this leads to a less compacted chromatin, whereas an on-dyad binding mode leads to a more compacted chromatin. Computational studies (33) on the other hand pointed out that an off-dyad binding mode can lead from minimal to maximum compaction, while an on-dyad binding leads to an intermediate compaction of the chromatin.

The various effects that the LH positioning have on chromatin compaction may have to do with how the gH domain is oriented on the nucleosome, and consequently, how neighbouring LHs are oriented with respect to each other. The LH has a tripartite structure, comprising of the disordered NTD and the CTD and the structured globular domain (gH). The structured gH domain along with the NTD is referred to as 'head', and the CTD, as 'tail'. Song et al. (133) proposed that when LHs on neighbouring chromatosomes face each other via the same domains, i.e., a 'head-to-head' or 'tail-to-tail' configuration, it leads to twisting and further compaction of a zig-zagged chromatin fibre. Building up on this observation, and upon their own observation that the CTD associates with one L-DNA arm, conferring an asymmetry to the nucleosome, Bednar and colleagues hypothesized that this asymmetry could have major implications for the stability, and the charge and mass distributions of higher-order structures (29).

In chapter 5 we report that in monochromatosomes, the orientation of the gH is affected by the flank sequences on the two L-DNAs. In presence of A-tracts on the L-DNA flanks, the gH is oriented in a way as to suitably position two conserved arginine residues (Zone B, denoted in pink, Figure 6.1B, detailed models in Chapter 5) proximal to the narrow minor groove of the A-tract. What effect can this A-tract induced reorientation of the gH domain have on the compaction of chromatosomal arrays? To explore this, a study was performed on trichromatosomes, in which, by placing A-tracts on outer or inner L-DNAs, we aimed to orient the gH domains of the LHs on the 1st and 3rd chromatosomes accordingly.

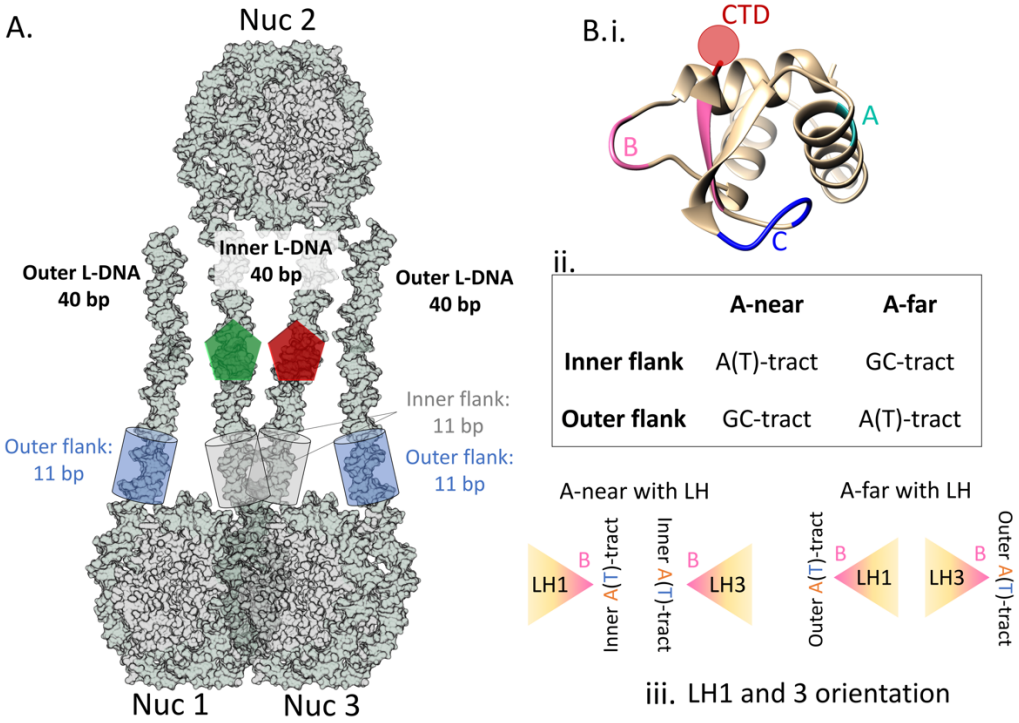


Figure 6.1: *Experimental design for the trichromatosome study.* ***A.*** *A general schematic of doubly-labelled (Alexa 488-green, Alexa 594-red) trinucleosome.* ***B.i.*** *The gH showing the three DNA interacting zones A, B and C. The juncture at which the CTD starts is marked with a red circle.* ***ii.*** *Table with information about the flank sequences of outer and inner L-DNA arms of constructs A-near and A-far.* ***iii.*** *The gH domain with the pink zone B is represented as a triangle, with the Zone B pointing towards A-tracts.*

Two DNA constructs of 600 bp were designed, named A-near (for A-tract near) and A-far (for A-tract far). They comprised of three Widom 601 positioning sequences separated by 40 bp L-DNAs. L-DNAs between the nucleosomes (or chromatosomes) 1 and 2, and 2 and 3 (named Nuc 1, Nuc 2 and Nuc 3) were termed 'inner' L-DNAs and the ones upstream of nucleosome 1 and downstream of nucleosome 3 were termed 'outer' L-DNAs. The A-near construct had A-tracts on the inner L-DNA flanks and GC-tracts on the outer L-DNA flanks of Nuc 1 and Nuc 3 (Figure 6.1Bii). The A-far construct had A-tracts on the outer L-DNA flanks and GC-tracts on the inner L-DNA flanks of Nuc 1 and Nuc 3. The fluorophores Alexa 488 and 594 were placed at the midpoints of the two 40bp, inner L-DNA arms. Chapter 3 contains detailed description of the DNA preparation, trichromatosome reconstitution and quality control via AFM and agarose gel electrophoresis. Appendix A contains the DNA sequences.

6.2 Linker histones compact A-far and A-near trichromatosomes similarly

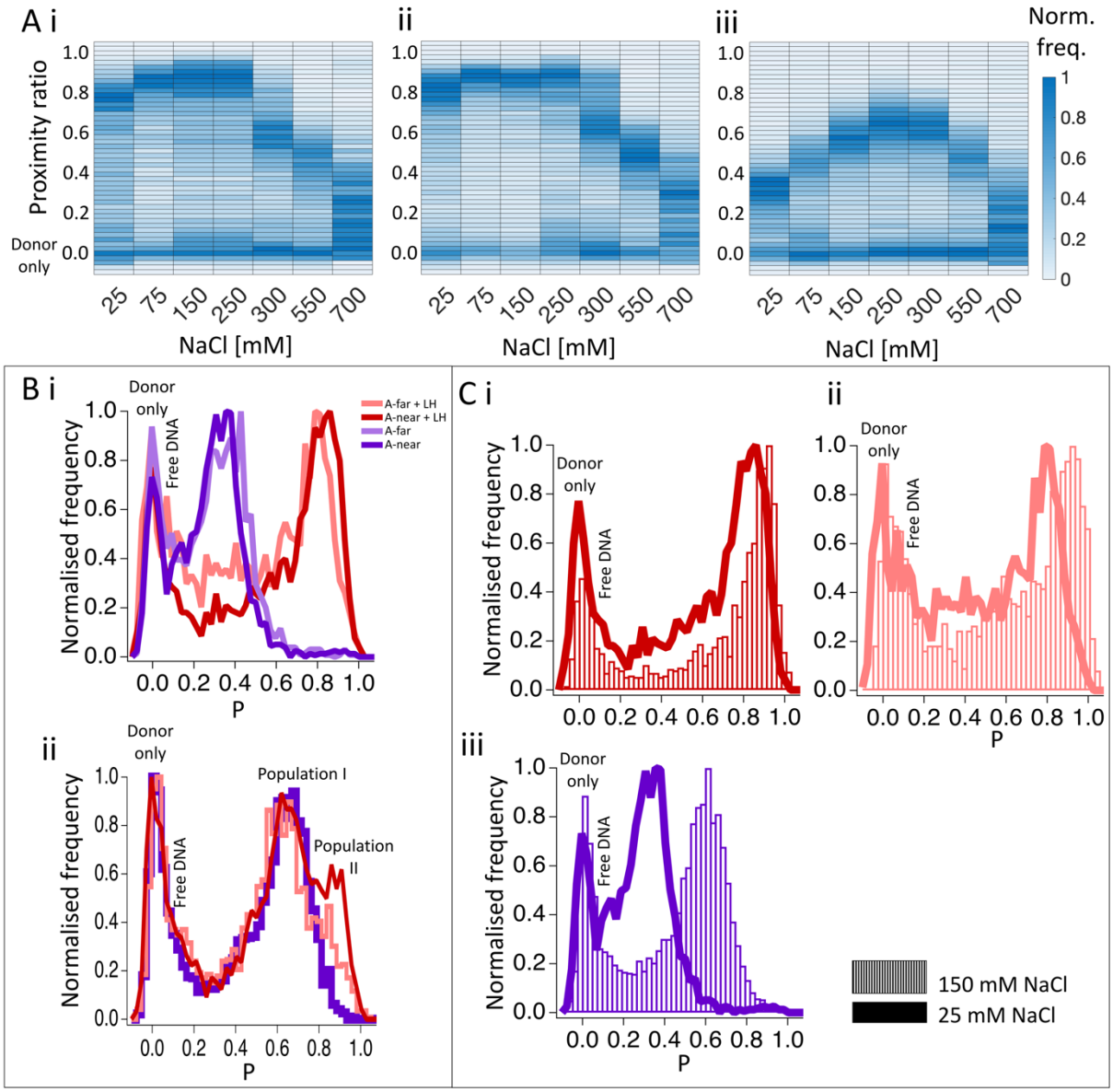


Figure 6.2: *Single-pair FRET observations for trichromatosomes.* ***A****. Salt-induced trichromatosome dissociation assay. (i) A-far with LH, (ii) A-near with LH, (iii) A-near without LH. Proximity ratio is on y-axis and NaCl concentration (mM) on x-axis. The colour intensity represents the normalised frequency.* ***B.*** *and* ***C.*** *Proximity ratio (P) plots at individual salts.* ***B(i)****. All the four constructs, A-far with and without LH, A-near with and without LH at 25 mM NaCl 1xTE.* ***(ii)*** *A-near with and without LH, and A-far with LH at 300 mM NaCl 1xTE.* ***C.*** *Comparison between proximity ratio plots at 25 mM NaCl (bold line) and 150 mM NaCl (bars) for* ***(i)*** *A-near with LH* ***(ii)*** *A-far with LH and* ***(iii)*** *A-near without LH. Colour: Red: A-near with LH, Pink: A-far with LH, Purple: A-near without LH, Mauve: A-far without LH (Figure Bi).*

Salt assays were performed on the A-far and A-near constructs with and without LH to assess how compaction of the trichromatosomes varied with salt concentration. The samples were incubated in 1xTE buffers (pH 7.5) containing 25, 75, 150, 250, 300, 550, and 700 mM NaCl for 1 hour prior to spFRET measurements using single-laser excitation. Figure 6.2A shows heatmap representations of the salt assay. At 25 mM NaCl (Figure 6.2 B(i)), the proximity ratio of

the trichromatosomes with the LH are higher (at ca. 0.82) than for the sample without LH (at ca. 0.35). A further increase in salt upto 250mM leads to further compaction of the trichromatosomes. This salt-induced increase in compaction (increase in proximity ratio) is less prominent in the LH containing samples than the ones without LH (Figure 6.2C). This compaction is brought about by shielding of the negative charges of the DNA by the positively charged sodium ions.

Upon fitting the proximity ratio plots (for fitted data, please see Supplementary Information Chapter 6B), it is observed that the prominent high FRET peak (proximity ratio 0.6 to 1.0) present in LH containing samples, can be fitted by two populations, denoted as Population I, having a lower proximity ratio and Population II having the highest proximity ratio. At 25 mM NaCl 1xTE, we observe that the population II show a higher mean proximity ratio in the A-near 'with LH' sample as compared to the A-far with LH. But the change in proximity ratio is negligible at ca. 0.05.

The difference in the salt-induced dissociation between A-far and A-near 'with LH' samples is negligible for the population II in the high-FRET range (Supplementary Information Chapter 6B II). This species coexists with the population I for salt concentrations 25 mM to about 250 mM, and decreases sharply at 300mM (Supplementary Information Chapter 6 Figure B II).

At 300 mM, the proximity ratio of the population I in LH-containing samples become equal to the major population peak of the samples without LH (0.61 to 0.65, see Supplementary Information B5, 300 mM). This species in the A-near and A-far 'with LH' samples may have lost the LH. The LH-containing population II still exists at this salt (Figure 6.2B(ii)). This high FRET species corresponds to about 20.4% of the total trichromatosome population in A-far with LH and 25% for A-near with LH. The remaining 75% to 80% of the total trichromatosome (population I), peak loses the LH.

The presence of linker histone compacts the trichromatosomes (169), resulting in a compact population II that is observable from 25 mM up to 250 mM NaCl 1xTE. However, the placement of the A-tracts on the inner (A-near) vs. outer (A-far) L-DNA arms does not affect the compaction or stability of the trichromatosomes.

6.3 Orientation of neighbouring LH does not affect trichromatosome compaction

Following up from our observations described in chapter 5, we expect that the zone B of the gH domains of the LHs associated with nucleosomes 1 and 3, will be oriented towards the A-tracts, which are present on the outer L-DNA (A-far) or inner L-DNA (A-near). From the spFRET observations we conclude that the orientation of the gH domain of the LHs on 1 and 3 does not affect the compaction or stability of the two trichromatosome constructs. The highly compact Population II was modeled using the software UCSF Chimera (120) and FPS (131), using a P value of 0.8 (SI Chapter 6 BII) between the mid-points of the inner L-DNA arms, at 25 mM. Normal mode analysis using the elNémo webserver (124) was performed on the starting structure 5NL0 (29) to obtain a range of arm opening conformations. The frame at which the computed proximity ratio was 0.8, was chosen, and this was assigned as the middle nucleosome, Nuc 2. Using Chimera, two other copies of 5NL0 were placed to build the Nuc 1 and 3. The DNA sequence of the structure was

mutated to that of A-far or A-near using the "Swapna" command in UCSF Chimera (120). The gH domain of the LH was placed facing the A-tracts on the Nuc 1 and 3, in accordance to the monochromatosome AG and GA models in Chapter 5.

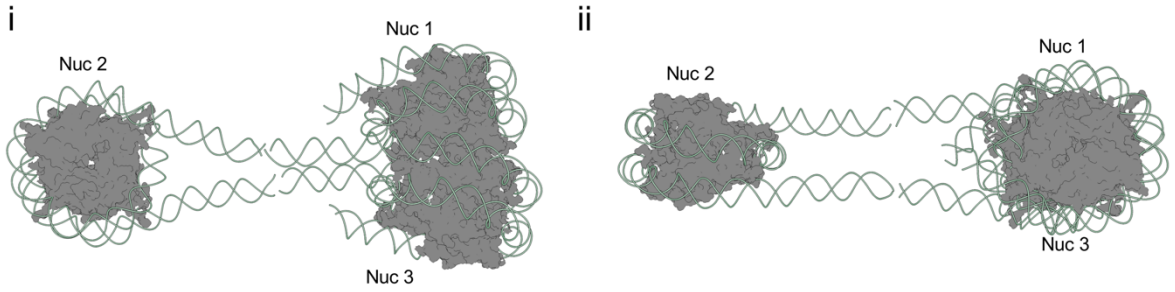


Figure 6.3: *Modelling the highly compact population II of trichromatosomes.* ***(i).*** *and* ***(ii).*** *show two perpendicular orientations of the trinucleosome structure. The outer L-DNA and the gH domains are removed in this representation for clarity; the core histones of nucleosomes (Nuc) 1-3 are shown in grey surface representation.*

Apart from the experimental proximity ratio constraint, two other considerations had to be kept in mind while modeling. One is the optimal stacking (when compared to PDBs 6HKT (144), 6L49 (170)) of nucleosomes 1 and 3, and secondly, that the inner L-DNA arms were not bent at a non-physiological angle. We find that the nucleosome 2 is twisted nearly orthogonally with respect to the nucleosomes 1 and 3 (Figure 6.3). Geometric parameters of the trichromatosome models were calculated using Chimera (120) (Figure 6.4 and Table 6). The three core histone octamers (Nuc 1, 2, and 3) were defined by the octamer planes in pink.

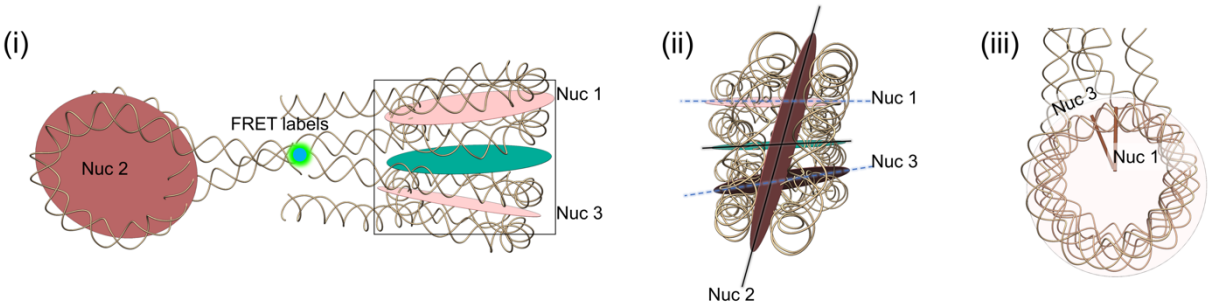


Figure 6.4: *Geometric parameters obtained from the model.* ***(i).*** *Pink discs represent the three core octamers, the green disc denote the stacking plane. The fluorescent dot denotes the site on the inner L-DNA where the FRET occurs between the fluorophores.* ***(ii).*** *Viewed from the Nuc 2 end. Nuc 1 and Nuc 3 planes are shown as dashed blue lines. Nuc 2 and stacking plane as black continuous line.* ***(iii).*** *The dyad axes of Nuc 1 and 3 are rotated with respect to each other.*

Table 6: *Geometric parameters obtained from the model.*

Angles	
octamer planes: Nuc 1 and Nuc 3 (pink. (i) and dashed blue lines (ii))	20.0°
Stacking plane (green) and Nuc 2 (ii)	72.6°
Dyad axis 1 and dyad axis 3 (iii)	28.1°

The angle between the planes Nuc 1 and Nuc 3 (angle between the blue dashed lines shown on Figure 6.4 (ii)) was calculated to be 20°. A stacking plane in green was defined in between the octamer planes 1 and 3, using the residues of the core histones of 1 and 3 that are facing each other: Arg 134 of the H3 histones of 1 and 3, Asp 24 of the H4 histones, Asn 68 of the H2A histones and Ser 120 of the H2B histones. This plane was defined to calculate the screw angle (Figure 6.4 (ii), angle between black lines) between it and the Nuc 2 octamer plane. This angle was about 73°. To measure the rotation of the plane Nuc 3 with respect to Nuc 1, dyad axes were defined in chromatosomes 1 and 3 using the dyad base pairs and the Arg 134 residue on the two copies of H3 in each octamer. The Nuc 3 is rotated by 28° with respect to Nuc 1.

Assuming nearly straight inner L-DNAs, our model of the trichromatosomes resemble the standard 2-start, zig-zagged conformations seen in cryo-EM, X-ray crystallographic, and computational studies of nucleosomal arrays (61, 143, 144, 170). As seen in our spFRET results presence of LH is essential for bringing the inner L-DNA arms, and hence the chromatosomes 1 and 3 in close proximity. However, the gH domain orientations anticipated from the AG and GA monochromatosome models (chapter 5) do not affect this compaction. The most likely reason for the lack of change in compaction could be that our LH subtype, the H1.0b, associates in an on-dyad fashion (29, 30, 158). Differences in compaction are shown mostly by LHs associated in an on- vs. off-dyad fashion (27, 33). The question now is, how does the LH mediate the compaction of trichromatosomes.

6.4 How does the linker histone aid trichromatosome compaction?

Biochemical assays and bulk-spectroscopic studies in the 90s (171, 172) showed that the LH can self-associate via its gH and C-terminal domains, in the presence of DNA, and this dimerization could have a major effect on higher-order structures of chromatin. Cryo-EM studies by Song *et al.* (133) on dodecamers showed that the 'head' domain (comprising of the NTD and gH) of off-dyad bound LHs, interacted directly with each other.

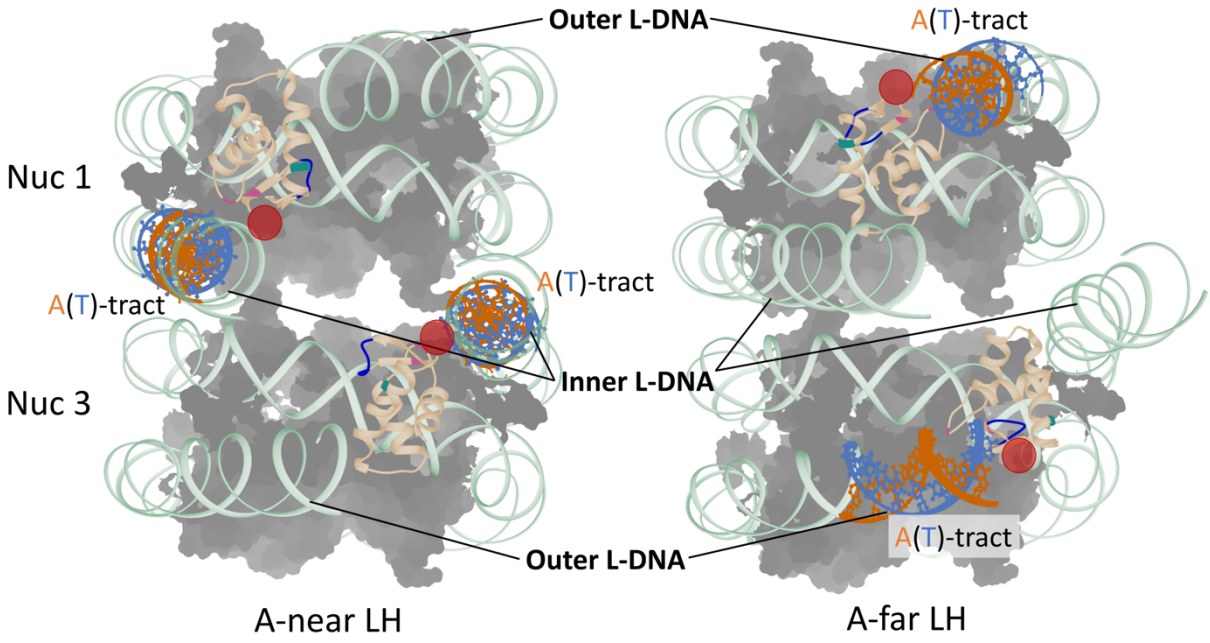


Figure 6.5: *Positioning of the gHs in the stacked nucleosomes 1 and 3 in the trichromatosome models. The stacked chromatosomes 1 and 3 in* ***Left:*** *A-near and* ***Right:*** *A-far constructs. The gH is coloured according to the scheme presented in Figure 6.1, i.e., zone A in sea-green, zone B in pink and zone C (containing the hydrophobic domain) in blue. The beginning of the CTD is represented by a red dot, as in Figure 6.1. Only the A-tract flank is shown in atomic detail, on either the inner (A-near) or the outer (A-far) L-DNA. Adenine is in orange and thymine in blue.*

In the model corresponding to population II, LH of Nuc 1 and 3 face each other via the same domains in both the A-near (Figure 6.4 left) and the A-far (Figure 6.4 right) constructs. Considering the gH+NTD as the 'head' and the CTD as the 'tail' (29, 133), the A-near construct has a tail-to-tail and the A-far construct has a head-to-head configuration. However, as we see in our model, the gH of LHs 1 and 3 are not in close proximity for either hydrophobic interactions (Zone C-Zone C) to take place or salt-bridges to form. Thus, this model of population II does not support the idea of direct gH interactions in our constructs.

We next consider the CTD. We did not explicitly model the CTDs of the LHs 1 and 3, and in chapter 5 we describe that the CTD fluorophore at 101st residue is nearly equidistant from the A-tract and GC-tract containing L-DNAs. However, this one fluorophore position cannot account for the exact location of these 100 amino acids long intrinsically disordered domain. Previous experimental and computational studies have shown the CTD, owing to being highly positively charged, to be locally structured in the presence of nucleosomes (135, 136) and with nucleosomal arrays (137). Cryo-EM studies on mononucleosomes have shown that the CTD of the H1.0 subtype associates with one of the L-DNA arms (29) or both (30). Cryo-EM studies by Song *et al.* (133) on nucleosomal arrays have shown that the CTD of off-dyad bound gH domains, associates with both the L-DNAs.

In our constructs, the starting point of the C-terminus (residue 95 of the gH, denoted as a red dot) of the 1st and 3rd gH face towards each other in A-near (Figure 6.4 left) and away from each other in A-far (Figure 6.4 right). If the CTD associates with one of the L-DNAs, it may associate with the outer L-DNAs in A-far and the inner L-DNAs in A-near. In this case, charge neutralisation

of the inner L-DNAs in A-near would be expected to result in its greater compaction that A-far. This is not the case. If the CTD associates with both the L-DNAs as suggested by Zhou *et al.* (2021) (30), then, irrespective of the orientation of the gH, the CTD would neutralise charges of both the outer and the inner L-DNA arms. In such a case, both the A-near and A-far constructs would be expected to be compacted similarly. Thus, such an arrangement of the CTDs would be compatible with our observations.

6.5 Concluding remarks

We observe no change in the extents of compaction of the two constructs studied. The differing orientations of the 1st and 3rd LH towards A-tracts on the inner (A-near) or outer (A-far) L-DNA does not affect the inner L-DNA arm-arm distance. However, a limitation of this study is that we studied only one subtype of linker histone, the H1.0b, which has been observed to bind on-dyad (29, 30). Another limitation in this study is that we do not incorporate A-tracts to the 2nd chromatosome (Nuc 2). Both the L-DNA flanks have a mixed sequence. Thus, we expect the gH of the 2nd LH to be oriented non-specifically.

This study was carried out for trichromatosomes, as minimal systems for studying chromatin compaction. We cannot extrapolate the results to longer chromatin fibres, of varied L-DNA lengths, or to in-vivo conditions. The possible role that A-tract induced orientational differences of the gH can have on larger nucleosomal arrays deserve further investigation.

Chapter 7

Conclusions & Outlook

Since the first formal description of the chromatosome in the late 70s (21), the linker histone has been somewhat of an enigma. Despite nearly fifty years of both experimental and computational studies, numerous questions regarding this elusive, 'forgotten' (173) histone still remain. Geometrically, the nucleosome (i.e., minus the LH) has an axis of symmetry also known as the pseudodyad axis. The DNA sequence around the core, the PTMs on the histone tails (174), etc. confers asymmetry to the nucleosome. On forming a chromatosome, i.e. with the association of the LH (on-/off-dyad (27–32)), an additionally layer of asymmetry is conferred. This has been speculated to impart an asymmetry in higher-order structures of the chromatin (27, 29) and likewise affect its compaction. But what determines the mode of association of the LH on the nucleosome? Is it only the amino acid sequence variations (31) in different LH isoforms or LH PTMs (163) that determine this, or does the L-DNA sequence also contribute to the observed modes of LH association?

There has been an observed preference of the LH for AT-rich sequences (44, 45, 50, 125) and A-tracts (46–48). What are the structural and consequently biological implications of this sequence specificity? As a corollary, how exactly does the LH recognize AT-rich DNA stretches and A-tracts? This thesis has aimed to answer some of these questions.

7.1 Contribution of L-DNA to the location of LH: Length or sequence?

In Chapter 4 we first studied the proximity of the globular domain to variably long L-DNAs. Although the globular domain showed a higher FRET peak for the 40 bp long L-DNA arm as opposed to the shorter 24 bp arm, it did not show the same effect when the arm lengths were swapped, or when both the arms were elongated to 40 bp. This was expected as the binding of the LH to the nucleosome has been observed to be affected by L-DNA lengths less than 10 bp (146). But the globular domain consistently showed higher FRET peaks for the plus arm, irrespective of its length.

This led us to inspect the sequence of the L-DNA arms at the sites flanking the globular domain of the LH. The pGEM-3z/601 plasmid containing the strongly positioning 601 sequence (53) has a purely GC stretch on one side of the NCP sequence. The other side has a mixed DNA sequence with a 64% AT content. It was expected, based on previous studies (44–51, 125) that the globular domain would show a higher FRET peak for the AT-rich flank, but it was not so. Instead, it showed a higher FRET peak for the L-DNA that contained the purely GC flank. When the 64% AT-rich flank was replaced by an A-tract, the globular domain showed reproducibly high FRET peaks with the A-tract, even upon swapping the flank sequences of the two L-DNA arms. This shows us that L-DNA sequences at the sites flanking the globular domain can affect its location or orientation.

7.2 Direct and indirect recognition of AT-rich DNA by the linker histone

Recognition of AT-rich DNA by the LH has so far been attributed either to the CTD (48) or a hydrophobic region of the globular domain (49, 50). The latter (we term as Zone C), containing a Gly-Val-Gly-Ala motif in H1.0 isoforms (Table 7) had been previously observed *in silico* (50, 125) to mediate recognition of AT-rich DNA via hydrophobic interactions with thymine methyl groups. Why did we then observe a selective preference of the globular domain for A-tracts and not for the 64% AT-rich sequence? The answer lies in the mode of association of the LH.

Thymine methyl groups are exposed on the major groove of the DNA that can be accessed by a LH positioned in an off-dyad fashion (50). In Chapter 5 our FRET-restrained nucleosome models show that the globular domain of the LH is situated on-dyad and in the canonical fashion, in accordance to the previous structural and computational studies (27–31). At this position, the major groove of the flanking L-DNAs is out of reach for the hydrophobic Zone C region of the LH. This can be a possible reason why the LH does not show high FRET for the 64% AT-rich flank. But how does it recognize A-tracts?

A characteristic feature of A-tracts is the narrowing of its minor groove (43, 159, 162). Proteins that can recognize A-tracts (48, 159) typically do so by employing arginine residues or an arginine-containing, conserved 'AT-hook' motif (175, 176) comprising of the sequence: Pro-Arg-Gly-Arg-Pro (175). The arginine residues mediate ionic interactions with the highly negative A-tract minor groove (159, 162). Amongst the positively charged amino acid residues lysines and arginines, stripping off the hydration shell of the guanidium group in arginines, is energetically more favorable than the amino group of lysine (159, 162).

In our models of AG and GA constructs (Chapter 5), as well as in previous structural (27, 29–31) and computational models (158) of the on-dyad bound linker histones, the minor grooves of the L-DNA flanks are proximal to the globular domain of the LH. Our models show that the on-dyad bound globular domain has a specific orientation with respect to the two L-DNA arms. A particular region that we designate as Zone B, comprising of the loop 1 and the β2 domain, is situated proximal to the A-tract flank in the AG construct. In the GA construct where the flank sequences are swapped between the two L-DNA arms, the globular domain is reoriented accordingly. Two arginines in this Zone B protrude into the minor groove of the A-tract in both AG and the GA constructs. These two residues, R42 on the loop 1 and R94 on β2 are highly conserved in LH isoforms across organisms, and form a 'pincer'-like motif, that we suggest has similar functionality as the AT-hook motif, that is, recognition of A-tracts via its narrow minor groove. Both these conserved arginine residues have been previously mutated to alanines (154, 165, 177) or to glutamic acid (156) to show, using both experimental (154, 156, 177)and computational techniques (165) that such mutants bind to DNA with less affinity.

Although we observe that the Zone B mediates A-tract recognition, recent cryo-EM structures by Zhou *et al.* (2021) (30) showed that a conserved arginine residue on the α3 helix (R74) was proximal to the minor groove of the 64% AT-rich flank. Further inspection into the sequence of the α3 helix in LH isoforms across organisms (Table 7.1) show that there is indeed one arginine in the α3 helix although its location is not conserved.

Table 7: *Zones A, B and C of different LH isoforms across organisms. Boxed is the isoform used in the work presented in this thesis.*

Organism	LH isoforms	helix α3 [Zone A]	Loop 1 [Zone B]	β2 [Zone B]	beta-hairpin loop [Zone C]
Xenopus laevis	H1A	VDKNNS**R**LKLALK	KE**R**SG	SFKL	GSGA
	H1B	NS**R**LKLALKALVTK	KE**R**SG	SFKL	GSGA
	H1C	VDKNNS**R**LKLALK	KE**R**GG	SFKL	GSGA
	H1.0-A	DSQIKLSIK**R**LV	KS**R**SG	SF**R**L	***GVGA***
	H1.0-B	DSQIKLSIK**R**LV	KS**R**SG	SF**R**L	***GVGA***
Homo sapiens	H1.1	NNS**R**IKLGIKSLVS	KE**R**GG	SFKL	GTGA
	H1.2	NNS**R**IKLGLKSLVS	KE**R**SG	SFKL	GTGA
	H1.3	NNS**R**IKLGLKSLVS	KE**R**SG	SFKL	GTGA
	H1.4	NNS**R**IKLGLKSLVS	KE**R**SG	SFKL	GTGA
	H1.5	NNS**R**IKLGLKSLVS	KE**R**NG	SFKL	GTGA
	H1.6/H1t	NNS**R**IKLSLKSLV	QE**R**VG	SFKL	GTGA
	H1.7	KSG**R**HEAP**R**GQA	THKGL	YF**R**V	
	H1.8/H1oo	**R**FKYLLKQALATGM**RR**	EQ**RR**G	SFKL	ARGA
	H1.9	AYHFK**R**VLKGLV	TCKYV	SFTL	GTCK
	H1.10	G**R**TYLKYSIKALVQN	GE**R**NG	SFKL	GTGA
	H1.0	DSQIKLSIK**R**LV	KN**R**AG	SF**R**L	***GVGA***
Mus musculus	H1t	NS**R**IKLALK**R**LVN	QE**R**AG	SFKL	GTGA
	H1.0	DSQIKLSIK**R**LV	KN**R**AG	SF**R**L	***GVGA***
Gallus gallus	H5	DLQIKLSI**RR**LL	KS**R**GG	SF**R**L	***GVGA***
Drosophila melanogaster	H1	LAPFIKKYLKSAVV	KE**R**GG	SFKL	GKGA

We decided to obtain further evidence on whether the on-dyad bound linker histones mediate recognition of A-tracts via hydrophobic interactions with thymine methyl groups. In Chapter 5 I have described our FRET observations from the TU construct that contained two A-tracts, one complementary to methyl-group containing thymine and the other complementary to deoxy-uridine lacking the methyl-group. Our FRET-restrained modeling shows the possibility of two species of chromatosomes: one where the Zone B of the globular domain is oriented towards the thymine tract and one where the Zone B is oriented towards the deoxy-uridine tract. Replacing thymines with deoxy-uridines would not affect the width of the minor groove (166) of the A-tract, making both the A-tracts indistinguishable when recognized via the minor groove.

This observation further proved that on-dyad positioned linker histones recognize DNA sequences such as A-tracts indirectly, by sequence induced changes in the physical parameters of the DNA, such as the narrowing of the minor groove. Off-dyad positioned linker histones, being able to access the major groove can directly recognize DNA sequences via interactions with the functional groups of the bases, such as the methyl group of thymines.

Taking our observations together with previous works (30, 49, 50, 125) we conclude that there are multiple ways for the LH to recognize AT-rich DNA. The mode of association of the LH on the nucleosome determines its mode of recognition of AT-rich DNA: either directly via the

major groove or indirectly via the minor groove.

The biological significance of the linker histone's ability to recognize AT-rich DNA can be two-fold: (a) the compaction of heterochromatin and (b) in the higher-order structuring of the chromatin.

7.3 Implications in heterochromatin compaction

Constitutive heterochromatin are highly condensed and unexpressed parts of the chromatin that are rich in satellite sequences having a high AT content (47). A-tracts having a continuous stretch of four to six adenines (43) make up the consensus sequences of Scaffold/matrix associated regions (S/MAR) (46). The observed affinity of LH for AT-rich sequences and A-tracts could be directly correlated with observations of the LH having high occupancy in S/MAR or AT-rich heterochromatin (167, 168).

7.4 Implications in higher-order structure: a preliminary understanding and future perspectives

The widely observed on- (29, 30, 158) and off-dyad (31, 32) modes of association of the LH on nucleososomes led to speculations as to how these two binding modes may affect the chromatin at larger scales (27, 32, 33). At the heart of these suggestions lie the question: how can linker histones on neighbouring chromatosomes compact higher-order chromatin structures. Do globular domains of the off-dyad positioned linker histones self-associate (32, 171, 172) to mediate compaction or is this mediated by the ca. 100 amino acid long CTD (136, 178, 179) that enable compaction by neutralising the negative charges of the linker-DNA?

We suggest a possible way to study the role of the linker histone in structuring chromatin at larger scales by exploiting its preference for AT-rich sequences, specifically A-tracts. Chapter 6 of this thesis describes a work in this direction. Here we show a modified *in-vitro* Golden-Gate assembly protocol (described in Chapter 3) that can be used to build arrays of Widom 601 sequences with A-tracts at user-defined positions. Ideally, linker histones of both types, i.e., on-dyad binder and off-dyad binder should be used to reconstitute nucleosome arrays containing A-tracts, to properly elucidate the role of linker histone in higher-order structuring. Here we show the possible contribution of the linker histone of subtype H1.0b on the compaction of trinucleosomes.

A-tracts were placed either on the inner or on the outer L-DNA arms flanking nucleosomes 1 and 3, generating two types of trinucleosomes. According to our AG and GA models in Chapter 5, in the first type of trinucleosome, the hydrophobic Zone C on the LH globular domains on nucleosomes 1 and 3 would have faced each other and in the second type, away from each other.

Although both the trinucleosome types are greatly compacted in presence of linker histones, we observe no change in the extent of compaction between these two types of trinucleosomes. Modeling the trinucleosome using the inner L-DNA-L-DNA distance shows us that

on-dyad positioned globular domains on nucleosomes 1 and 3 are not close enough to enable self-association. Thus, apart from other factors such as nucleosome-nucleosome stacking, one of the contributors to the compaction of the trinucleosomes could be the C-terminal domains of the three linker histones. We observed that the upstream end of the CTD (position G101, Chapter 4) is equidistant from both the L-DNA arms. Recent cryo-EM studies by Zhou *et al.* (2021) (30) show that the CTD of H1.0 subtypes interact similarly with both the L-DNA arms. The CTD has also been shown to be essential in the compaction of nucleosome arrays (128), by neutralising the negative charge on the L-DNAs. Thus, overall, we show that if positioned on-dyad, orientational changes of the gH domain does not affect compaction. The CTD that associates with both the L-DNA arms similarly can compact the two trinucleosome types in a similar fashion, irrespective of the orientational differences in the gH domain.

7.5 Opening a Pandora's box

Overall, the work presented in this thesis has shown the influence of DNA sequence on linker histone positioning and a new mode of DNA sequence recognition by the linker histone. As a follow up, we have demonstrated how A-tract recognition by linker histone can be exploited to study the effect of linker histone in the structuring of chromatin in larger scales.

But newer questions and avenues arise from this work that can be addressed in the future. For instance:

- What if the conserved arginines (42 and 94) are mutated to lysines? Being positively charged, I would expect these mutants to compact nucleosomes. But would A-tract recognition be affected? How would this affect heterochromatin compaction *in vivo*?
- Arginines 42 and 94 can be mutated to see if arginine 74 at Zone A can mediate A-tract recognition. This arginine too can be mutated to lysine, to remove indirect A-tract recognition capability of the LH. In this case the expected observation would be that the gH, if positioned on-dyad, would not show high FRET peaks for A-tracts.
- The position of the single arginine residue in Zone A (R74) is not conserved (Table 7). How would this affect recognition of A-tracts? To do this, LH of different subtypes should be used and the Zone B arginines should be mutated so as to bias A-tract recognition to Zone A.
- Preference towards AT-rich DNA had been previously attributed to the CTD (48). This issue needs to be further resolved by performing similar sp-FRET experiments as ours, on A-tract containing nucleosomes and CTD-lacking LH. Additionally, the CTD should be labelled at multiple positions to ascertain whether any part of the CTD has a preference towards AT-rich DNA.
- To further support our spFRET based modeling, cryo-EM studies are necessary to directly visualise A-tract induced orientational changes of the globular domain.

- A major limitation of this study was that we ignored the contribution of core histone tails. It is necessary to systematically study the contribution of core histone tails to A-tract recognition by the LH. Is there any competition between the two?
- How would off-dyad bound linker histones affect the compaction of trinucleosomes? Would the globular domains self-associate if they are proximal enough? Using A-tracts and an off-dyad binding LH, this question can be addressed.

These avenues can be explored in order to gain a complete understanding of how L-DNA sequences, linker histones of different isoforms, core histone tails, etc., together help in modulating the compaction of the chromatin.

Bibliography

1. Society, N.G. 2019. History of the Cell: Discovering the Cell. *National Geographic Society*.
2. Ford, B.J. 2014. Living Images from the Birth of Microscopy. *Microscopy Today*. 22:18–23.
3. Flemming, W. 1882. Zellsubstanz, Kern und Zelltheilung. Leipzig: F.C.W. Vogel.
4. Flemming, W. 1879. Beitrage zur Kenntniss der Zelle und ihrer Lebenserscheinungen. *Archiv f. mikrosk. Anatomie*. 16:302–436.
5. Hardy, P.A., and H. Zacharias. 2009. Walther Flemming on histology in medicine 1878: A newly discovered letter to his father. *Annals of Anatomy - Anatomischer Anzeiger*. 191:171–185.
6. Paweletz, N. 2001. Walther Flemming: pioneer of mitosis research. *Nat Rev Mol Cell Biol*. 2:72–75.
7. Olins, A.L., and D.E. Olins. 1974. Spheroid Chromatin Units (ν Bodies). *Science*. 183:330–332.
8. Oudet, P., M. Gross-Bellard, and P. Chambon. 1975. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell*. 4:281–300.
9. Luck, J.M., P.S. Rasmussen, K. Satake, and A.N. Tsvetikov. 1958. Further studies on the fractionation of calf thymus histone. *J Biol Chem*. 233:1407–1414.
10. Neelin, J.M., P.X. Callahan, D.C. Lamb, and K. Murray. 1964. The histones of chicken erythrocyte nuclei. *Can. J. Biochem.* 42:1743–1752.
11. Hnilica, L.S., H.A. Kappler, and V.S. Hnilica. 1965. Biosynthesis of Histones and Acidic Nuclear Proteins under Different Conditions of Growth. *Science*. 150:1470–1472.
12. de Nooij, E.H., and H.G.K. Westenbrink. 1962. Isolation of a homogeneous lysine-rich histone from calf thymus. *Biochimica et Biophysica Acta*. 62:608–609.
13. Dick, C., and E.W. Johns. 1969. The biosynthesis of the five main histone fractions of rat thymus. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*. 174:380–386.
14. Johns, E.W., D.M.P. Phillips, P. Simson, and J.A.V. Butler. 1960. Improved fractionations of arginine-rich histones from calf thymus. *Biochem J*. 77:631–636.
15. Olins, D.E. 1969. Interaction of lysine-rich histones and DNA. *Journal of Molecular Biology*. 43:439–460.
16. Mirsky, A.E., C.J. Burdick, E.H. Davidson, and V.C. Littau. 1968. The role of lysine-rich histone in the maintenance of chromatin structure in metaphase chromosomes. *Proc Natl Acad Sci U S A*. 61:592–597.
17. Kornberg, R.D. 1974. Chromatin Structure: A Repeating Unit of Histones and DNA. *Science*. 184:868–871.
18. Thomas, J.O., and R.D. Kornberg. 1975. An octamer of histones in chromatin and free in solution. *Proc Natl Acad Sci U S A*. 72:2626–2630.
19. Olins, A.L., M.B. Senior, and D.E. Olins. 1976. Ultrastructural features of chromatin nu bodies. *J Cell Biol*. 68:787–793.
20. Olins, D.E., and A.L. Olins. 1978. Nucleosomes: the structural quantum in chromosomes. *Am Sci*. 66:704–711.
21. Simpson, R.T. 1978. Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry*. 17:5524–5531.
22. Bates, D.L., and J.O. Thomas. 1981. Histories H1 and H5: one or two molecules per nucleosome? *Nucleic Acids Research*. 9:5883–5894.
23. Richmond, T.J., J.T. Finch, B. Rushton, D. Rhodes, and A. Klug. 1984. Structure of the nucleosome core particle at 7 A resolution. *Nature*. 311:532–537.
24. Luger, K., A.W. Mäder, R.K. Richmond, D.F. Sargent, and T.J. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. 389:251–260.
25. Ramakrishnan, V., J.T. Finch, V. Graziano, P.L. Lee, and R.M. Sweet. 1993. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature*. 362:219–223.
26. Bednar, J., R.A. Horowitz, S.A. Grigoryev, L.M. Carruthers, J.C. Hansen, A.J. Koster, and C.L. Woodcock. 1998. Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *PNAS*. 95:14173–14178.
27. Zhou, B.-R., J. Jiang, H. Feng, R. Ghirlando, T.S. Xiao, and Y. Bai. 2015. Structural Mechanisms of Nucleosome Recognition by Linker Histones. *Mol Cell*. 59:628–638.
28. Zhou, B.-R., J. Jiang, R. Ghirlando, D. Norouzi, K.N. Sathish Yadav, H. Feng, R. Wang, P. Zhang, V.

Zhurkin, and Y. Bai. 2018. Revisit of Reconstituted 30-nm Nucleosome Arrays Reveals an Ensemble of Dynamic Structures. *J Mol Biol*. 430:3093–3110.

29. Bednar, J., I. Garcia-Saez, R. Boopathi, A.R. Cutter, G. Papai, A. Reymer, S.H. Syed, I.N. Lone, O. Tonchev, C. Crucifix, H. Menoni, C. Papin, D.A. Skoufias, H. Kurumizaka, R. Lavery, A. Hamiche, J.J. Hayes, P. Schultz, D. Angelov, C. Petosa, and S. Dimitrov. 2017. Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. *Mol Cell*. 66:384-397.e8.
30. Zhou, B.-R., H. Feng, S. Kale, T. Fox, H. Khant, N. de Val, R. Ghirlando, A.R. Panchenko, and Y. Bai. 2021. Distinct Structures and Dynamics of Chromatosomes with Different Human Linker Histone Isoforms. *Molecular Cell*. 81:166-182.e6.
31. Zhou, B.-R., H. Feng, R. Ghirlando, S. Li, C.D. Schwieters, and Y. Bai. 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.
32. Song, F., P. Chen, D. Sun, M. Wang, L. Dong, D. Liang, R.-M. Xu, P. Zhu, and G. Li. 2014. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science*. 344:376–380.
33. Perišić, O., S. Portillo-Ledesma, and T. Schlick. 2019. Sensitive effect of linker histone binding mode and subtype on chromatin condensation. *Nucleic Acids Research*. 47:4948–4957.
34. International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature*. 409:860–921.
35. Trifonov, E.N., and J.L. Sussman. 1980. The pitch of chromatin DNA is reflected in its nucleotide sequence. *PNAS*. 77:3816–3820.
36. Wang, S., M. Munde, S. Wang, and W.D. Wilson. 2011. Minor Groove to Major Groove, an Unusual DNA Sequence-Dependent Change in Bend Directionality by a Distamycin Dimer. *Biochemistry*. 50:7674–7683.
37. Oguey, C., N. Foloppe, and B. Hartmann. 2010. Understanding the Sequence-Dependence of DNA Groove Dimensions: Implications for DNA Interactions. *PLoS ONE*. 5:e15931.
38. Ortiz, V., and J.J. de Pablo. 2011. Molecular Origins of DNA Flexibility: Sequence Effects on Conformational and Mechanical Properties. *Phys Rev Lett*. 106:238107.
39. Rettig, M., M.W. Germann, S. Wang, and W.D. Wilson. 2013. Molecular Basis For Sequence-Dependent Induced DNA Bending. *Chembiochem*. 14:323–331.
40. Steffen, N.R., S.D. Murphy, L. Tolleri, G.W. Hatfield, and R.H. Lathrop. 2002. DNA sequence and structure: direct and indirect recognition in protein-DNA binding. *Bioinformatics*. 18 Suppl 1:S22-30.
41. Vinogradov, A.E. 2003. DNA helix: the importance of being GC-rich. *Nucleic Acids Research*. 31:1838–1844.
42. Tillo, D., and T.R. Hughes. 2009. G+C content dominates intrinsic nucleosome occupancy. *BMC Bioinformatics*. 10:442.
43. Haran, T.E., and U. Mohanty. 2009. The unique structure of A-tracts and intrinsic DNA bending. *Quart. Rev. Biophys.* 42:41–81.
44. Zlatanova, J., and J. Yaneva. 1991. Histone H1-DNA Interactions and Their Relation to Chromatin Structure and Function. *DNA and Cell Biology*. 10:239–248.
45. An, W., S.H. Leuba, K. van Holde, and J. Zlatanova. 1998. Linker histone protects linker DNA on only one side of the core particle and in a sequence-dependent manner. *Proc Natl Acad Sci USA*. 95:3396.
46. Käs, E., Izaurralde, E., & Laemmli, U. K. 1989. Specific inhibition of DNA Binding to nuclear scaffolds and histone H1 by distamycin. *Journal of Molecular Biology*. 210:587–599.
47. Cao, K., N. Lailler, Y. Zhang, A. Kumar, K. Uppal, Z. Liu, E.K. Lee, H. Wu, M. Medrzycki, C. Pan, P.-Y. Ho, G.P. Cooper, X. Dong, C. Bock, E.E. Bouhassira, and Y. Fan. 2013. High-Resolution Mapping of H1 Linker Histone Variants in Embryonic Stem Cells. *PLoS Genet*. 9:e1003417.
48. Roque, A. 2004. The preferential binding of histone H1 to DNA scaffold-associated regions is determined by its C-terminal domain. *Nucleic Acids Research*. 32:6111–6119.
49. Cui, F., and V.B. Zhurkin. 2009. Distinctive sequence patterns in metazoan and yeast nucleosomes: Implications for linker histone binding to AT-rich and methylated DNA. *Nucleic Acids Research*. 37:2818–2829.
50. Ozturk, M.A., G.V. Pachov, R.C. Wade, and V. Cojocaru. 2016. Conformational selection and dynamic adaptation upon linker histone binding to the nucleosome. *Nucleic Acids Res*. 44:6599–6613.
51. Tomaszewski, R. 1997. The AT-rich flanks of the oocyte-type 5S RNA gene of Xenopus laevis act as a strong local signal for histone H1-mediated chromatin reorganization in vitro. *Nucleic Acids Research*. 25:458–466.
52. Torres, C.M., A. Biran, M.J. Burney, H. Patel, T. Henser-Brownhill, A.-H.S. Cohen, Y. Li, R. Ben-Hamo, E. Nye, B. Spencer-Dene, P. Chakravarty, S. Efroni, N. Matthews, T. Misteli, E. Meshorer, and P. Scaffidi.

2016. The linker histone H1.0 generates epigenetic and functional intratumor heterogeneity. *Science*. 353:aaf1644–aaf1644.
53. Lowary, P.T., and J. Widom. 1998. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of Molecular Biology*. 276:19–42.
54. Piovesan, A., M.C. Pelleri, F. Antonaros, P. Strippoli, M. Caracausi, and L. Vitale. 2019. On the length, weight and GC content of the human genome. *BMC Research Notes*. 12:106.
55. Finch, J.T., and A. Klug. 1976. Solenoidal model for superstructure in chromatin. *Proc Natl Acad Sci U S A*. 73:1897–1901.
56. Thoma, F., T. Koller, and A. Klug. 1979. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *Journal of Cell Biology*. 83:403–427.
57. Robinson, P.J., and D. Rhodes. 2006. Structure of the '30nm' chromatin fibre: A key role for the linker histone. *Current Opinion in Structural Biology*. 16:336–343.
58. Kruithof, M., F.-T. Chien, A. Routh, C. Logie, D. Rhodes, and J. van Noort. 2009. Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber. *Nature Structural & Molecular Biology*. 16:534–540.
59. Wedemann, G., and J. Langowski. 2002. Computer Simulation of the 30-Nanometer Chromatin Fiber. *Biophysical Journal*. 82:2847–2859.
60. Martino, J.A., V. Katritch, and W.K. Olson. 1999. Influence of nucleosome structure on the three-dimensional folding of idealized minichromosomes. *Structure*. 7:1009–1022.
61. Schalch, T., S. Duda, D.F. Sargent, and T.J. Richmond. 2005. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*. 436:138–141.
62. Tremethick, D.J. 2007. Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber. *Cell*. 128:651–654.
63. Rydberg, B., W.R. Holley, I.S. Mian, and A. Chatterjee. 1998. Chromatin conformation in living cells: support for a zig-zag model of the 30 nm chromatin fiber11Edited by T. Richmond. *Journal of Molecular Biology*. 284:71–84.
64. Risca, V.I., S.K. Denny, A.F. Straight, and W.J. Greenleaf. 2017. Variable chromatin structure revealed by in situ spatially correlated DNA cleavage mapping. *Nature*. 541:237–241.
65. Fussner, E., R.W. Ching, and D.P. Bazett-Jones. 2011. Living without 30 nm chromatin fibers. *Trends in Biochemical Sciences*. 36:1–6.
66. Maeshima, K., S. Hihara, and M. Eltsov. 2010. Chromatin structure: does the 30-nm fibre exist in vivo? *Curr Opin Cell Biol*. 22:291–297.
67. Grosberg, A., Y. Rabin, S. Havlin, and A. Neer. 1993. Crumpled Globule Model of the Three-Dimensional Structure of DNA. *EPL*. 23:373–378.
68. Mirny, L.A. 2011. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res*. 19:37–51.
69. Grosberg, A.Yu., S.K. Nechaev, and E.I. Shakhnovich. 1988. The role of topological constraints in the kinetics of collapse of macromolecules. *Journal de Physique*. 49:2095–2100.
70. Bancaud, A., S. Huet, N. Daigle, J. Mozziconacci, J. Beaudouin, and J. Ellenberg. 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.
71. Lieberman-Aiden, E., N.L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B.R. Lajoie, P.J. Sabo, M.O. Dorschner, R. Sandstrom, B. Bernstein, M.A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L.A. Mirny, E.S. Lander, and J. Dekker. 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*. 326:289.
72. Dekker, J., A.S. Belmont, M. Guttman, V.O. Leshyk, J.T. Lis, S. Lomvardas, L.A. Mirny, C.C. O'Shea, P.J. Park, B. Ren, J.C.R. Politz, J. Shendure, and S. Zhong. 2017. The 4D nucleome project. *Nature*. 549:219–226.
73. Ou, H.D., S. Phan, T.J. Deerinck, A. Thor, M.H. Ellisman, and C.C. O'Shea. 2017. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science*. 357:eaag0025.
74. Narlikar, G.J. 2020. Phase-separation in chromatin organization. *J Biosci*. 45:5.
75. Erdel, F., and K. Rippe. 2018. Formation of Chromatin Subcompartments by Phase Separation. *Biophysical Journal*. 114:2262–2270.
76. Zhang, Y., and T.G. Kutateladze. 2019. Liquid–liquid phase separation is an intrinsic physicochemical property of chromatin. *Nat Struct Mol Biol*. 26:1085–1086.
77. Gibson, B.A., L.K. Doolittle, M.W.G. Schneider, L.E. Jensen, N. Gamarra, L. Henry, D.W. Gerlich, S. Redding, and M.K. Rosen. 2019. Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell*. 179:470-484.e21.
78. Shakya, A., S. Park, N. Rana, and J.T. King. 2020. Liquid-Liquid Phase Separation of Histone Proteins

in Cells: Role in Chromatin Organization. *Biophysical Journal*. 118:753–764.
79. Duce, J.A., D.P. Smith, R.E. Blake, P.J. Crouch, Q.-X. Li, C.L. Masters, and I.A. Trounce. 2006. Linker histone H1 binds to disease associated amyloid-like fibrils. *J Mol Biol*. 361:493–505.
80. Ye, X., C. Feng, T. Gao, G. Mu, W. Zhu, and Y. Yang. 2017. Linker Histone in Diseases. *Int J Biol Sci*. 13:1008–1018.
81. Medrzycki, M., Y. Zhang, J.F. McDonald, and Y. Fan. 2012. Profiling of linker histone variants in ovarian cancer. *Front Biosci (Landmark Ed)*. 17:396–406.
82. Sepsa, A., G. Levidou, A. Gargalionis, C. Adamopoulos, A. Spyropoulou, G. Dalagiorgou, I. Thymara, E. Boviatsis, M.S. Themistocleous, K. Petraki, G. Vrettakos, V. Samaras, A. Zisakis, E. Patsouris, C. Piperi, and P. Korkolopoulou. 2015. Emerging Role of Linker Histone Variant H1x as a Biomarker with Prognostic Value in Astrocytic Gliomas. A Multivariate Analysis including Trimethylation of H3K9 and H4K20. *PLOS ONE*. 10:e0115101.
83. Bauden, M., T. Kristl, A. Sasor, B. Andersson, G. Marko-Varga, R. Andersson, and D. Ansari. 2017. Histone profiling reveals the H1.3 histone variant as a prognostic biomarker for pancreatic ductal adenocarcinoma. *BMC Cancer*. 17:810.
84. Behrends, M., and O. Engmann. 2020. Linker histone H1.5 is an underestimated factor in differentiation and carcinogenesis. *Environmental Epigenetics*. 6.
85. Torres, C.M., A. Biran, M.J. Burney, H. Patel, T. Henser-Brownhill, A.-H.S. Cohen, Y. Li, R. Ben-Hamo, E. Nye, B. Spencer-Dene, P. Chakravarty, S. Efroni, N. Matthews, T. Misteli, E. Meshorer, and P. Scaffidi. 2016. The linker histone H1.0 generates epigenetic and functional intratumor heterogeneity. *Science*. 353.
86. Clegg, R.M. 2006. The History of Fret. In: Geddes CD, JR Lakowicz, editors. Reviews in Fluorescence 2006. Boston, MA: Springer US. pp. 1–45.
87. Förster, Th. 1948. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Annalen der Physik*. 437:55–75.
88. Lakowicz, J.R. 2006. Principles of Fluorescence Spectroscopy. 3rd ed. Springer, Boston, MA.
89. Stryer, L., and R.P. Haugland. 1967. Energy transfer: a spectroscopic ruler. *Proc Natl Acad Sci U S A*. 58:719–726.
90. Tóth, K., N. Brun, and J. Langowski. 2001. Trajectory of Nucleosomal Linker DNA Studied by Fluorescence Resonance Energy Transfer. *Biochemistry*. 40:6921–6928.
91. Tóth, K., N. Brun, and J. Langowski. 2006. Chromatin Compaction at the Mononucleosome Level. *Biochemistry*. 45:1591–1598.
92. Gansen, A., F. Hauger, K. Tóth, and J. Langowski. 2007. Single-pair fluorescence resonance energy transfer of nucleosomes in free diffusion: Optimizing stability and resolution of subpopulations. *Analytical Biochemistry*. 368:193–204.
93. Koopmans, W.J.A., A. Brehm, C. Logie, T. Schmidt, and J. van Noort. 2007. Single-Pair FRET Microscopy Reveals Mononucleosome Dynamics. *J Fluoresc*. 17:785–795.
94. Gansen, A., K. Tóth, N. Schwarz, and J. Langowski. 2009. Structural Variability of Nucleosomes Detected by Single-Pair Förster Resonance Energy Transfer: Histone Acetylation, Sequence Variation, and Salt Effects. *J. Phys. Chem. B*. 113:2604–2613.
95. Lee, N.K., A.N. Kapanidis, Y. Wang, X. Michalet, J. Mukhopadhyay, R.H. Ebright, and S. Weiss. 2005. Accurate FRET Measurements within Single Diffusing Biomolecules Using Alternating-Laser Excitation. *Biophysical Journal*. 88:2939–2953.
96. Principles of Fluorescence Spectroscopy | Joseph R. Lakowicz | Springer. .
97. Haugland, R.P., J. Yguerabide, and L. Stryer. 1969. Dependence of the kinetics of singlet-singlet energy transfer on spectral overlap. *Proc Natl Acad Sci U S A*. 63:23–30.
98. Yang, Z.-C., M. Wang, A.M. Yong, S.Y. Wong, X.-H. Zhang, H. Tan, A.Y. Chang, X. Li, and J. Wang. 2011. Intrinsically fluorescent carbon dots with tunable emission derived from hydrothermal treatment of glucose in the presence of monopotassium phosphate. *Chem. Commun.* 47:11615–11617.
99. Ivanov, V., M. Li, and K. Mizuuchi. 2009. Impact of emission anisotropy on fluorescence spectroscopy and FRET distance measurements. *Biophys J*. 97:922–929.
100. Dale, R.E., J. Eisinger, and W.E. Blumberg. 1979. The orientational freedom of molecular probes. The orientation factor in intramolecular energy transfer. *Biophys J*. 26:161–193.
101. Hellenkamp, B., S. Schmid, O. Doroshenko, O. Opanasyuk, R. Kühnemuth, S. Rezaei Adariani, B. Ambrose, M. Aznauryan, A. Barth, V. Birkedal, M.E. Bowen, H. Chen, T. Cordes, T. Eilert, C. Fijen, C. Gebhardt, M. Götz, G. Gouridis, E. Gratton, T. Ha, P. Hao, C.A. Hanke, A. Hartmann, J. Hendrix, L.L. Hildebrandt, V. Hirschfeld, J. Hohlbein, B. Hua, C.G. Hübner, E. Kallis, A.N. Kapanidis, J.-Y. Kim, G. Krainer, D.C. Lamb, N.K. Lee, E.A. Lemke, B. Levesque, M. Levitus, J.J. McCann, N. Naredi-Rainer, D. Nettels, T. Ngo, R. Qiu, N.C. Robb, C. Röcker, H. Sanabria, M. Schlierf, T. Schröder, B. Schuler, H. Seidel, L. Streit, J. Thurn, P. Tinnefeld, S. Tyagi, N. Vandenberk, A.M. Vera, K.R. Weninger, B. Wünsch, I.S.

Yanez-Orozco, J. Michaelis, C.A.M. Seidel, T.D. Craggs, and T. Hugel. 2018. Precision and accuracy of single-molecule FRET measurements—a multi-laboratory benchmark study. *Nat Methods*. 15:669–676.

102. Cristóvão, M., E. Sisamakis, M.M. Hingorani, A.D. Marx, C.P. Jung, P.J. Rothwell, C.A.M. Seidel, and P. Friedhoff. 2012. Single-molecule multiparameter fluorescence spectroscopy reveals directional MutS binding to mismatched bases in DNA. *Nucleic Acids Research*. 40:5448–5464.
103. Reinartz, I., C. Sinner, D. Nettels, B. Stucki-Buchli, F. Stockmar, P.T. Panek, C.R. Jacob, G.U. Nienhaus, B. Schuler, and A. Schug. 2018. Simulation of FRET dyes allows quantitative comparison against experimental data. *The Journal of Chemical Physics*. 148:123321.
104. Ha, T., T. Enderle, D.F. Ogletree, D.S. Chemla, P.R. Selvin, and S. Weiss. 1996. Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *PNAS*. 93:6264–6268.
105. Deniz, A.A., M. Dahan, J.R. Grunwell, T. Ha, A.E. Faulhaber, D.S. Chemla, S. Weiss, and P.G. Schultz. 1999. Single-pair fluorescence resonance energy transfer on freely diffusing molecules: Observation of Förster distance dependence and subpopulations. *PNAS*. 96:3670–3675.
106. Roy, R., S. Hohng, and T. Ha. 2008. A Practical Guide to Single Molecule FRET. *Nat Methods*. 5:507–516.
107. Gansen, A., A.R. Hieb, V. Böhm, K. Tóth, and J. Langowski. 2013. Closing the Gap between Single Molecule and Bulk FRET Analysis of Nucleosomes. *PLOS ONE*. 8:e57018.
108. Gansen, A. 2007. Chromatin at the Nanolevel - Development of a single molecule FRET experiment and analysis of the structure and stability of individual nucleosomes. .
109. Lehmann, K. 2016. Fluoreszenz-Einzelmolekül-Studien zur Dynamik von Mononukleosomen. .
110. Kapanidis, A.N., N.K. Lee, T.A. Laurence, S. Doose, E. Margeat, and S. Weiss. 2004. Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *PNAS*. 101:8936–8941.
111. Lehmann, Kathrin. 2016. Fluoreszenz-Einzelmolekül-Studien zur Dynamik von Mononukleosomen. Heidelberg University.
112. Kapanidis, A.N., T.A. Laurence, N.K. Lee, E. Margeat, X. Kong, and S. Weiss. 2005. Alternating-laser excitation of single molecules. *Acc Chem Res*. 38:523–533.
113. Nir, E., X. Michalet, K.M. Hamadani, T.A. Laurence, D. Neuhauser, Y. Kovchegov, and S. Weiss. 2006. Shot-Noise Limited Single-Molecule FRET Histograms: Comparison between Theory and Experiments. *J. Phys. Chem. B*. 110:22103–22124.
114. Engler, C., R. Gruetzner, R. Kandzia, and S. Marillonnet. 2009. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLoS One*. 4:e5553.
115. Engler, C., R. Kandzia, and S. Marillonnet. 2008. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS One*. 3:e3647.
116. Lehmann, K., R. Zhang, N. Schwarz, A. Gansen, N. Mücke, J. Langowski, and K. Toth. 2017. Effects of charge-modifying mutations in histone H2A α3-domain on nucleosome stability assessed by single-pair FRET and MD simulations. *Sci Rep*. 7:13303.
117. Luger, K., T.J. Rechsteiner, and T.J. Richmond. 1999. Preparation of nucleosome core particle from recombinant histones. In: Methods in Enzymology. Elsevier. pp. 3–19.
118. Würtz, M., D. Aumiller, L. Gundelwein, P. Jung, C. Schütz, K. Lehmann, K. Tóth, and K. Rohr. 2019. DNA accessibility of chromatosomes quantified by automated image analysis of AFM data. *Sci Rep*. 9:12788.
119. De, M., M.A. Öztürk, S. Isbaner, K. Tóth, and R.C. Wade. 2021. DNA sequence-dependent positioning of the linker histone in a nucleosome: a single-pair FRET study. *Biophysical Journal*. https://doi.org/10.1016/j.bpj.2021.07.012
120. Pettersen, E.F., T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, and T.E. Ferrin. 2004. UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*. 25:1605–1612.
121. De Bruin, L., and J.H. Maddocks. 2018. cgDNAweb: a web interface to the cgDNA sequence-dependent coarse-grain model of double-stranded DNA. *Nucleic Acids Research*. 46:W5–W10.
122. Pérez, A., I. Marchán, D. Svozil, J. Sponer, T.E. Cheatham 3rd, C.A. Laughton, and M. Orozco. 2007. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J*. 92:3817–3829.
123. Li, S., W.K. Olson, and X.-J. Lu. 2019. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Research*. 47:W26–W34.
124. Suhre, K., and Y.-H. Sanejouand. 2004. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Research*. 32:W610–W614.

125. Pachov, G.V., R.R. Gabdoulline, and R.C. Wade. 2011. On the structure and dynamics of the complex of the nucleosome and the linker histone. *Nucleic Acids Research*. 39:5255–5263.
126. Bahar, I., T.R. Lezon, A. Bakan, and I.H. Shrivastava. 2010. Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins. *Chem. Rev.* 110:1463–1497.
127. Hizver, J., H. Rozenberg, F. Frolow, D. Rabinovich, and Z. Shakked. 2001. DNA bending by an adenine–thymine tract and its role in gene regulation. *Proc Natl Acad Sci USA*. 98:8490.
128. Buckwalter, J.M., D. Norouzi, A. Harutyunyan, V.B. Zhurkin, and S.A. Grigoryev. 2017. Regulation of chromatin folding by conformational variations of nucleosome linker DNA. *Nucleic Acids Research*. 45:9372–9387.
129. A sequence-dependent coarse-grain model of B-DNA with explicit description of bases and phosphate groups parametrised from large scale Molecular Dynamics simulations. 262.
130. Maier, J.A., C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, and C. Simmerling. 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* 11:3696–3713.
131. Kalinin, S., T. Peulen, S. Sindbert, P.J. Rothwell, S. Berger, T. Restle, R.S. Goody, H. Gohlke, and C.A.M. Seidel. 2012. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nature Methods*. 9:1218–1225.
132. Sindbert, S., S. Kalinin, H. Nguyen, A. Kienzler, L. Clima, W. Bannwarth, B. Appel, S. Müller, and C.A.M. Seidel. 2011. Accurate Distance Determination of Nucleic Acids via Förster Resonance Energy Transfer: Implications of Dye Linker Length and Rigidity. *J. Am. Chem. Soc.* 133:2463–2480.
133. Song, F., P. Chen, D. Sun, M. Wang, L. Dong, D. Liang, R.-M. Xu, P. Zhu, and G. Li. 2014. Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units. *Science*. 344:376.
134. Zhou, B.-R., H. Feng, S. Kale, T. Fox, H. Khant, N. de Val, R. Ghirlando, A.R. Panchenko, and Y. Bai. 2020. Distinct Structures and Dynamics of Chromatosomes with Different Human Linker Histone Isoforms. *Molecular Cell*. S1097276520307723.
135. Luque, A., R. Collepardo-Guevara, S. Grigoryev, and T. Schlick. 2014. Dynamic condensation of linker histone C-terminal domain regulates chromatin structure. *Nucleic Acids Res*. 42:7553–7560.
136. Caterino, T.L., and J.J. Hayes. 2011. Structure of the H1 C-terminal domain and function in chromatin condensationThis paper is one of a selection of papers published in a Special Issue entitled 31st Annual International Asilomar Chromatin and Chromosomes Conference, and has undergone the Journal's usual peer review process. *Biochem. Cell Biol.* 89:35–44.
137. Fang, H., S. Wei, T.-H. Lee, and J.J. Hayes. 2016. Chromatin structure-dependent conformations of the H1 CTD. *Nucleic Acids Res*. gkw586.
138. Grigoryev, S.A., and M. Schubert. 2019. Unraveling the multiplex folding of nucleosome chains in higher order chromatin. *Essays in Biochemistry*. 63:109–121.
139. van Holde, Kensal E. Chromatin (Springer, New York, 1989). Springer, New York, NY.
140. Compton, J.L., M. Bellard, and P. Chambon. 1976. Biochemical evidence of variability in the DNA repeat length in the chromatin of higher eukaryotes. *Proc Natl Acad Sci USA*. 73:4382.
141. Perišić, O., R. Collepardo-Guevara, and T. Schlick. 2010. Modeling Studies of Chromatin Fiber Structure as a Function of DNA Linker Length. *Journal of Molecular Biology*. 403:777–802.
142. Routh, A., S. Sandin, and D. Rhodes. 2008. Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proc Natl Acad Sci USA*. 105:8872.
143. Wu, C., and A. Travers. 2019. Modelling and DNA topology of compact 2-start and 1-start chromatin fibres. *Nucleic Acids Research*. 47:9902–9924.
144. Garcia-Saez, I., H. Menoni, R. Boopathi, M.S. Shukla, L. Soueidan, M. Noirclerc-Savoye, A. Le Roy, D.A. Skoufias, J. Bednar, A. Hamiche, D. Angelov, C. Petosa, and S. Dimitrov. 2018. Structure of an H1-Bound 6-Nucleosome Array Reveals an Untwisted Two-Start Chromatin Fiber Conformation. *Molecular Cell*. 72:902-915.e7.
145. Woodcock, C.L., A.I. Skoultchi, and Y. Fan. 2006. Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Research*. 14:17–25.
146. White, A.E., A.R. Hieb, and K. Luger. 2016. A quantitative investigation of linker histone interactions with nucleosomes and chromatin. *Scientific Reports*. 6:19122.
147. Hagerman, P.J. 1990. Sequence-directed curvature of dna. *Annu. Rev. Biochem.* 59:755–781.
148. Satchwell, S.C., H.R. Drew, and A.A. Travers. 1986. Sequence periodicities in chicken nucleosome core DNA. *Journal of Molecular Biology*. 191:659–675.
149. Crothers, D.M. 1998. DNA curvature and deformation in protein–DNA complexes: A step in the right direction. *Proc Natl Acad Sci USA*. 95:15163.
150. Kaplan, N., I.K. Moore, Y. Fondufe-Mittendorf, A.J. Gossett, D. Tillo, Y. Field, E.M. LeProust, T.R. Hughes, J.D. Lieb, J. Widom, and E. Segal. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*. 458:362–366.

151. Segal, E., and J. Widom. 2009. Poly(dA:dT) tracts: major determinants of nucleosome organization. *Current Opinion in Structural Biology*. 19:65–71.
152. Kunkel, G.R., and H.G. Martinson. 1981. Nucleosomes will not form on double-stranded RNA or over poly(dA)-poly(dT) tracts in recombinant DNA. *Nucleic Acids Research*. 9:6869–6888.
153. Öztürk, M.A., G.V. Pachov, R.C. Wade, and V. Cojocaru. 2016. Conformational selection and dynamic adaptation upon linker histone binding to the nucleosome. *Nucleic Acids Res*. 44:6599–6613.
154. Brown, D.T., T. Izard, and T. Misteli. 2006. Mapping the interaction surface of linker histone H10 with the nucleosome of native chromatin in vivo. *Nature Structural & Molecular Biology*. 13:250–255.
155. George, E.M., T. Izard, S.D. Anderson, and D.T. Brown. 2010. Nucleosome Interaction Surface of Linker Histone H1c Is Distinct from That of H1 ^{0}. *J. Biol. Chem.* 285:20891–20896.
156. Goytisolo, F.A., S.E. Gerchman, X. Yu, C. Rees, V. Graziano, V. Ramakrishnan, and J.O. Thomas. 1996. Identification of two DNA-binding sites on the globular domain of histone H5. *EMBO J*. 15:3421–3429.
157. Zhou, B.-R., J. Jiang, R. Ghirlando, D. Norouzi, K.N.S. Yadav, H. Feng, R. Wang, P. Zhang, V. Zhurkin, and Y. Bai. 2019. Revisit of reconstituted 30-nm nucleosome arrays reveals an ensemble of dynamic structures. 30.
158. Woods, D.C., and J. Wereszczynski. 2020. Elucidating the influence of linker histone variants on chromatosome dynamics and energetics. *Nucleic Acids Research*. 48:3591–3604.
159. Rohs, R., S.M. West, A. Sosinsky, P. Liu, R.S. Mann, and B. Honig. 2009. The role of DNA shape in protein–DNA recognition. *Nature*. 461:1248–1253.
160. Hud, N.V., and J. Plavec. 2003. A unified model for the origin of DNA sequence-directed curvature. *Biopolymers*. 69:144–158.
161. Luscombe, N.M., R.A. Laskowski, and J.M. Thornton. 2001. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res*. 29:2860–2874.
162. West, S.M., R. Rohs, R.S. Mann, and B. Honig. 2010. Electrostatic Interactions between Arginines and the Minor Groove in the Nucleosome. *Journal of Biomolecular Structure and Dynamics*. 27:861–866.
163. Öztürk, M.A., V. Cojocaru, and R.C. Wade. 2018. Dependence of Chromatosome Structure on Linker Histone Sequence and Posttranslational Modification. *Biophysical Journal*. 114:2363–2375.
164. Bewley, C.A., A.M. Gronenborn, and G.M. Clore. 1998. Minor groove-binding architectural proteins: Structure, Function, and DNA Recognition. *Annu. Rev. Biophys. Biomol. Struct.* 27:105–131.
165. Öztürk, M.A., and R.C. Wade. 2020. Computation of FRAP recovery times for linker histone – chromatin binding on the basis of Brownian dynamics simulations. *Biochimica et Biophysica Acta (BBA) - General Subjects*. 1864:129653.
166. Fadda, E., and R. Pomès. 2011. On the molecular basis of uracil recognition in DNA: comparative study of T-A versus U-A structure, dynamics and open base pair kinetics. *Nucleic Acids Research*. 39:767–780.
167. Haaf, T., and M. Schmid. 2000. Experimental condensation inhibition in constitutive and facultative heterochromatin of mammalian chromosomes. *Cytogenetic and Genome Research*. 91:113–123.
168. Jamieson, K., E.T. Wiles, K.J. McNaught, S. Sidoli, N. Leggett, Y. Shao, B.A. Garcia, and E.U. Selker. 2016. Loss of HP1 causes depletion of H3K27me3 from facultative heterochromatin and gain of H3K27me2 at constitutive heterochromatin. *Genome Res.* 26:97–107.
169. Bussiek, M., K. Tóth, N. Schwarz, and J. Langowski. 2006. Trinucleosome Compaction Studied by Fluorescence Energy Transfer and Scanning Force Microscopy. *Biochemistry*. 45:10838–10846.
170. Takizawa, Y., C.-H. Ho, H. Tachiwana, H. Matsunami, W. Kobayashi, M. Suzuki, Y. Arimura, T. Hori, T. Fukagawa, M.D. Ohi, M. Wolf, and H. Kurumizaka. 2020. Cryo-EM Structures of Centromeric Tri-nucleosomes Containing a Central CENP-A Nucleosome. *Structure*. 28:44-53.e4.
171. Maman, J.D., T.D. Yager, and J. Allan. 1994. Self-association of the globular domain of histone H5. *Biochemistry*. 33:1300–1310.
172. Carter, G.J., and K. van Holde. 1998. Self-Association of Linker Histone H5 and of Its Globular Domain: Evidence for Specific Self-Contacts. *Biochemistry*. 37:12477–12488.
173. Brockers, K., and R. Schneider. 2019. Histone H1, the forgotten histone. *Epigenomics*. 11:363–366.
174. Bowman, G.D., and M.G. Poirier. 2015. Post-Translational Modifications of Histones That Influence Nucleosome Dynamics. *Chem Rev*. 115:2274–2295.
175. Aravind, L., and D. Landsman. 1998. AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Research*. 26:4413–4421.
176. Rodríguez, J., J. Mosquera, J. R. Couceiro, M. Eugenio Vázquez, and J. L. Mascareñas. 2015. The AT-Hook motif as a versatile minor groove anchor for promoting DNA binding of transcription factor fragments. *Chemical Science*. 6:4767–4771.
177. Duggan, M.M., and J.O. Thomas. 2000. Two DNA-binding sites on the globular domain of histone H5

are required for binding to both bulk and 5 S reconstituted nucleosomes. *J Mol Biol*. 304:21–33.

178. Allan, J., T. Mitchell, N. Harborne, L. Bohm, and C. Crane-Robinson. 1986. Roles of H1 domains in determining higher order chromatin structure and H1 location. *Journal of Molecular Biology*. 187:591–601.

179. Clausell, J., N. Happel, T.K. Hale, D. Doenecke, and M. Beato. 2009. Histone H1 Subtypes Differentially Modulate Chromatin Condensation without Preventing ATP-Dependent Remodeling by SWI/SNF or NURF. *PLOS ONE*. 4:e0007243.

Appendices

Appendix A: Sequences

1. **DNA sequences:** Forward strand are shown. All strands run from 5' to 3'. The dyad base is numbered 0 and depicted in bold. Bases in red denote the position of Alexa 594 labelling. Labelling is always on the thymine base, T -93 (first red T) on forward strand and T +94 on reverse strand (the corresponding A +94, coloured red, is on forward strand).

a. **Monochromatosome:** For TU construct, the base (dT or dU) the A-tract is complementary to is given within square brackets and depicted in bold.

Asy MG

GGCAATGTCGCTGTTCAATACATGCACAGGATGTATATATCTGACACGTGCCTGGAGACTAGGGAGTA
ATCCCCTTGGCGGTTAAAACGCGGGGGACA**G(0)**CGCGTACGTGCGTTTAAGCGGTGCTAGAGCTGTC
TACGACCAATTGAGCGGCCTCGGCACCGGGATTCTCCAGGGCGGCCGCGTATAGGGTCCATCACATAA
GGGATGAACTC

As InvMG

ATCCGACTGGCACCGGCAATGTCGCTGTTCAATACATGCACAGGATGTATATATCTGACACGTGCCTG
GAGACTAGGGAGTAATCCCCTTGGCGGTTAAAACGCGGGGGACA**G(0)**CGCGTACGTGCGTTTAAGCG
GTGCTAGAGCTGTCTACGACCAATTGAGCGGCCTCGGCACCGGGATTCTCCAGGGCGGCCGCGTATAG
GGTCCATCACA

MG

ATCCGACTGGCACCGGCAATGTCGCTGTTCAATACATGCACAGGATGTATATATCTGACACGTGCCTG
GAGACTAGGGAGTAATCCCCTTGGCGGTTAAAACGCGGGGGACA**G(0)**CGCGTACGTGCGTTTAAGCG
GTGCTAGAGCTGTCTACGACCAATTGAGCGGCCTCGGCACCGGGATTCTCCAGGGCGGCCGCGTATAG
GGTCCATCACATAAGGGATGAACTC

GM

ATCCGACTGGCACCGGCAATGTCGCTGTTCGCGGCCGCCACAGGATGTATATATCTGACACGTGCCTG
GAGACTAGGGAGTAATCCCCTTGGCGGTTAAAACGCGGGGGACA**G(0)**CGCGTACGTGCGTTTAAGCG
GTGCTAGAGCTGTCTACGACCAATTGAGCGGCCTCGGCACCGGGATTCTCCAGGCATGTATTGTATAG
GGTCCATCACATAAGGGATGAACTC

AG

ATCCGACTGGCACCGGCAATGTCGCTGTTCCAAAAAAAAAAAGGATGTATATATCTGACACGTGCCTG
GAGACTAGGGAGTAATCCCCTTGGCGGTTAAAACGCGGGGGACA**G(0)**CGCGTACGTGCGTTTAAGCG
GTGCTAGAGCTGTCTACGACCAATTGAGCGGCCTCGGCACCGGGATTCTCCAGGGCGGCCGCGTATAG
GGTCCATCACATAAGGGATGAACTC

GA

ATCCGACTGGCACCGGCAATGTCGCTGTTCACGCGGCCGCCCTGATGTATATATCTGACACGTGCCTG
GAGACTAGGGAGTAATCCCCTTGGCGGTTAAAACGCGGGGGACA**G(0)**CGCGTACGTGCGTTTAAGCG
GTGCTAGAGCTGTCTACGACCAATTGAGCGGCCTCGGCACCGGGATTCTCCCTTTTTTTTTTTGGTAG
GGTCCATCACATAAGGGATGAACTC

TU

ATCCGACTGGCACCGGCAATGTCGCTGTTCCAAAAAAAAAAA**[dT]**GGATGTATATATCTGACACGTG
CCTGGAGACTAGGGAGTAATCCCCTTGGCGGTTAAAACGCGGGGGACA**G(0)**CGCGTACGTGCGTTTA
AGCGGTGCTAGAGCTGTCTACGACCAATTGAGCGGCCTCGGCACCGGGATTCTCCCAAAAAAAAAAA**[**
dU]GGTAGGGTCCATCACATAAGGGATGAACTC

b. Trichromatosome: Bases labelled with Alexa 488 are depicted in green and bases labelled with Alexa 594 are depicted in red. Labelling is always at thymine. Here in the forward strand, the Adenine corresponding to the thymine on the opposite strand, is depicted in green.

A-far

ATCCGACTGGCACCGGCAATGTCGCTGTTCCAAAAAAAAAAAGGATGTATATATCTGACACGTGCCTGGAGACTAGG
GAGTAATCCCCTTGGCGGTTAAAACGCGGGGGACAGCGCGTACGTGCGTTTAAGCGGTGCTAGAGCTGTCTACGACC
AATTGAGCGGCCTCGGCACCGGGATTCTCCAGGGCGGCCGCGTATAGGGTCC**A**TCACATAAGGGATGAACTCACAGG
ATGTATATATCTGACACGTGCCTGGAGACTAGGGAGTAATCCCCTTGGCGGTTAAAACGCGGGGGACAGCGCGTACG
TGCGTTTAAGCGGTGCTAGAGCTGTCTACGACCAATTGAGCGGCCTCGGCACCGGGATTCTCCAGGATCCGACTGGC
ACCGGCAA**T**GTCGCTGTTTACGCGGCCGCCCTGATGTATATATCTGACACGTGCCTGGAGACTAGGGAGTAATCCCC
TTGGCGGTTAAAACGCGGGGGACAGCGCGTACGTGCGTTTAAGCGGTGCTAGAGCTGTCTACGACCAATTGAGCGGC
CTCGGCACCGGGATTCTCCCTTTTTTTTTTTGGTAGGGTCCATCACATAAGGGATGAACTC

A-near

ATCCGACTGGCACCGGCAATGTCGCTGTTTACGCGGCCGCCCTGATGTATATATCTGACACGTGCCTGGAGACTAGG
GAGTAATCCCCTTGGCGGTTAAAACGCGGGGGACAGCGCGTACGTGCGTTTAAGCGGTGCTAGAGCTGTCTACGACC
AATTGAGCGGCCTCGGCACCGGGATTCTCCCTTTTTTTTTTTGGTAGGGTCC**A**TCACATAAGGGATGAACTCACAGG
ATGTATATATCTGACACGTGCCTGGAGACTAGGGAGTAATCCCCTTGGCGGTTAAAACGCGGGGGACAGCGCGTACG
TGCGTTTAAGCGGTGCTAGAGCTGTCTACGACCAATTGAGCGGCCTCGGCACCGGGATTCTCCAGGATCCGACTGGC
ACCGGCAA**T**GTCGCTGTTCCAAAAAAAAAAAGGATGTATATATCTGACACGTGCCTGGAGACTAGGGAGTAATCCCC
TTGGCGGTTAAAACGCGGGGGACAGCGCGTACGTGCGTTTAAGCGGTGCTAGAGCTGTCTACGACCAATTGAGCGGC
CTCGGCACCGGGATTCTCCAGGGCGGCCGCGTATAGGGTCCATCACATAAGGGATGAACTC

2. Linker histone sequence: Sequence for the full-length *Xenopus laevis* H1.0b used. Labelling (Alexa 488) positions are depicted in green.

MAENSAATPAAKPKRSKALKKSTDHPKYSDMILAAVQAEKSRSGSSRQSIQKYIKNHYKVGENADSQIKLSIKRLV
T^{77}**C**SGALKQTKGVGASGSFRLAKADE**G**^{101}**C**KKPAKKPKKEIKKAVSPKKVAKPKKAAKSPAKAKKPKVAEKKVKKV
AKKKPAPSPKKAKKTKTVKAKPVRATKVKKAKPSKPKAKASPKKSGRKK

3. Primer sequences for DNA: Both forward and reverse strand primers are tabulated below. All strands run from 5' to 3'. Residue for Alexa 594 or Acc (red) and Alexa 488 or Don (green) labells are in bold and coloured accordingly. For trichromatosome primers, sequences in italics denote overhangs that allow the enzyme BsaI-HF-v2 to dock properly, and sequence in bold denotes BsaI-HF-v2 recognition sites. '|' denotes where the BsaI-HF-v2 cuts and underlined sequences are sticky ends after digestion with BsaI-HF-v2.

DNA:
Mochromatosomes

Gen-Forward	ATCCGACTGGCACCGGCAA**T**GTCGCTG
Gen-Reverse	GAGTTCATCCCTTATGTGA**T**GGACCCT

Gen (general) primers are for AG, GA, MG, GM, TU (forward), asyMG (Reverse) and asInvMG (Forward).

For TU: Reverse	GAGTTCATCCCTTATGTGATGGACCCTACCUUUUUUUUUUUCC
For asyMG: Forward	GGCAA**T**GTCGCTGTTCAATACATGCACAG
For asInvMG: Reverse-Don	TGTGA**T**GGACCCTATACGCGGCCGCCCTG

Trichromatosomes	Nuc1 forward = Gen-Forward and Nuc3 reverse = Gen-Reverse
Nuc1 Reverse-Don	*CCATC***GGTCTC**A<u>CTGT</u>|GAGTTCATCCCTTATGTGA**T**GGACCC
Nuc2 Forward	*ATCGA***GGTCTC**T|<u>ACAG</u>GATGTATATATCTGACACGTGCC
Nuc2 Reverse	*CCATC***GGTCTC**A<u>CCTG</u>|GAGAATCCCGGTGCC
Nuc3 Forward-Acc	*GTCAT***GGTCTC**T|<u>CAGG</u>ATCCGACTGGCACCGGCAA**T**GTCGC

4. **Primer sequences for wild-type LH (K195C to K195):** Site in orange is the recognition sequence for restriction enzyme NdeI. Site in green is recognition sequence for NotI. The AAG codon on reverse primer codes for lysine.

Forward	TATTATCATATGGCAGAGAATTCAGCCGCTACTCC
Reverse	AGCCCAAAGAAATCTGGACGGAAGAAGTAATGGCGGCCGCTATATTAT

Appendix B: Reagents and compositions

1. Protein purification

Lysis buffer	**pH 7.5**
Tris	50 mM
KCl	100 mM
Ethylene diamine tetraacetic acid (EDTA)	100 µM
Phenylmethanesulfonylfluoride (PMSF)	100 µM
Nonidet-P40	0.1% volume/volume

Unfolding buffer	**pH 7.5**
Guanidium hydrochloride	7 M
Tris	20 mM

SAU-50 buffer	**pH 5.2**
Deionised urea	7 M
Sodium acetate	20 mM
EDTA	100 µM
KCl	0.05 M

SAU-1000 buffer	**pH 5.2**
Deionised urea	7 M
Sodium acetate	20 mM
EDTA	100 µM
KCl	1 M

Refolding buffer	**pH 7.5**
NaCl	2 M
1xTE: Tris	10 mM
EDTA	100 µM

LB medium

Difco LB-Agar (Lennox) heated to 45°C until molten.

Carbenicillin	50 µg/ml

Add 100ml molten LB agar per plate.

Isopropyl-β-D-thiogalactopyranoside	1 M
Prepare in double distilled water	
Store at -20°C	

2. Protein labelling

Labelling buffer	**pH 7.5**
Guanidium hydrochloride	7 M
dithiothreitol	10 mM
Tris	20 mM
Tris-(2-carboxyethyl) phosphine	Final concentration 10x number of moles of protein

Buffer for unbound dye removal	**pH 5.2**
urea	7 M
Sodium acetate	20 mM
EDTA	100 μM
KCl	1 M

3. Labelled DNA purification

HPLC Buffer A	**pH 7.5**
Tris	250 mM

HPLC Buffer B	**pH 7.5**
Tris	250 mM
NaCl	1 M

4. Reconstituting chromatosomes (mono and tri)

Dialysis 1: Chamber B buffer	**pH 7.5**
NaCl	2 M
Tris	0.01 M
EDTA	100 μM

Dialysis 1: Chamber C buffer	**pH 7.5**
NaCl	0.6 M
Tris	0.01 M
EDTA	100 μM

Dialysis 2: Chamber D buffer	**pH 7.5**
Tris	0.01 M
EDTA	100 μM

5. Buffer for single-pair FRET measurements pH 7.5

NaCl	250 mM [standard] upto 0.7 M [salt series]
Tris	10 mM
EDTA	100 μM
Nonidet-P40	10 μM
Ascorbic acid	800 μM

6. Electrophoresis

a. Native non-denaturing gels

10x TAE buffer	**pH 7.5**
Tris	0.4 M
Glacial acetic acid	0.2 M
EDTA	0.01 M

10x TBE buffer	**pH 7.5**
Tris	0.89 M
Boric acid	0.89 M
EDTA	0.02 M

1% agarose gel (trichromatosome)	
agarose	1 g
TAE buffer 1x ditution	1 litre
1x TAE buffer used as running buffer	

6% polyacrylamide gel (monochromatosome)	1 gel
Acrylamide:bisacrylamide 60:1	2 ml
TBE buffer 10x	1 ml
Double distilled water	7 ml
Ammonium persulfate 10%	200 μl
TEMED	3 μl
1xTBE buffer used as running buffer	

8% polyacrylamide gel	1 gel
Acrylamide:bisacrylamide 30%	4 ml
TBE buffer 5x	4 ml
Double distilled water	6.7 ml
Ammonium persulfate 10%	300 μl
TEMED	3 μl
1xTBE used as running buffer	

Loading buffer for non-denaturing gels (colourless)	
Glycerine	99.9%
TBE buffer	5x
Double distilled water	

Non-denaturing or native gels were visualised in Typhoon scanner unstained. For visualisation under UV, they were stained with ethidium bromide.

b. Denaturing gels

15% SDS-PAGE

Lämmli buffer	**pH 8.8**
Tris	1.5 M
Double distilled water	1 litre
Sodium dodecyl-sulfate (SDS)	0.4%

Lämmli buffer	**pH 6.8**
Tris	0.5 M
Double distilled water	1 litre
SDS	0.4%

Running buffer for SDS-PAGE	**1x**
Tris	0.025 M
Glycine	0.25 M
SDS	1%

Resolving gel 15%	5 gels
Acrylamide:bisacrylamide 40% (37.5:1)	10.8 ml
Double distilled water	9.4 ml
Lämmli buffer pH 8.8 (with SDS, see above)	7 ml
Ammonium persulfate 10%	100 µl
TEMED	23 µl

Stacking gel	5 gels
Acrylamide:bisacrylamide 40% (37.5:1)	1.1 ml
Double distilled water	10 ml
Lämmli buffer pH 6.8 (with SDS, see above)	4 ml
Ammonium persulfate 10%	150 µl
TEMED	15 µl

Loading buffer 2x (dilution 1x to be used per well)	
Tris pH 6.8	0.125 M
Glycerin	20%
SDS	2%
Bromophenol blue	0.1%

Staining with coomassie blue dye.

Coomassie Blue staining for denaturing gels	
Staining	
Acetic acid	10%
methanol	20%
ethanol	42.5%
Brilliant Blue R250	0.2%
Double distilled water	

Destaining	
Isopropanol	20%
Acetic acid	7.5%

Appendix C: Instrumentation

Absorption spectroscopy	Cary 4E spectrometer	Varian, Mulgrave, Australia
Fluorescence spectroscopy	SLM-AMINCO 8100 fluorescence spectrometer	SLM, Urbana, IL, USA
Typhoon scanning	Typhoon 9400	GE Healthcare, USA
Single-pair FRET spectroscopy:		
Micro chamber plates	Sensoplate plus	Greiner Bio-One GmbH
Lasers:		
491nm	Cobolt Calypso	Hübner Photonics, Germany
561 nm	Cobolt Jive	Hübner Photonics, Germany
Acousto-optic tunable filter	AOTFnC-VIS-TN	AA Optoelectronics, France
Dichroic mirror	600DCXR	Omega Optical, Brattleboro, USA
Filters:		
Donor channel	520df40	Omega Optical, Brattleboro, USA
Acceptor channel	610ALP	Omega Optical, Brattleboro, USA
Avalanche photodiodes	SPAD-AQ-14	PerkinElmer Optoelectronics, Boston, Massachusetts, USA
Protein/DNA synthesis, purification:		
PCR	T3 Thermocycler	Biometra
Incubator	WTC	Binder
HPLC	Crystal 200 High Pressure system	Unicam
	Column for DNA purification	Gen-Pak-Fax, Waters
FPLC	Amershan GP250	Pharmacia
Gels:		
For polyacrylamide gels	Mini Protean	Bio-rad
Agarose gels	EC370 M	E-C Apparatus Corporation
SDS gels	CTI, Gelkammer TV21	CTI GmbH
Gel dock	UV BioDocAnalyse	Biometra

Supplementary Information

Supplementary information is organised chapter-wise.

For the monochromatosome studies, additional data are in the repository: https://doi.org/10.5281/zenodo.4686909. In this repository the following datasets can be found: raw data for single-pair FRET (.t3r), processed data for single-pair FRET (.xlsx, .csv), model ensembles (.pdb) and AV representations of dyes (.xyz). The fitted proximity ratio histograms for constructs AsyMG and AsInvMG are given in this section, subsection Chapter 4.

For trichromatosome studies, additional data are in the repository:
https://doi.org/10.5281/zenodo.5090768. In this repository are models of A-far and A-near (.pdb), raw single-pair FRET data for salts (.t3r). AFM pictures are added to this supplementary under the section Chapter 3.

The AlexEval software, developed by Dr. Sebastian Isbaner, can be found in this link: https://github.com/sisbaner/AlexEval.

Chapter 2

γ factor calculated for separate measurement days:

Day of measurement	P [10bp apart]	S [10bp apart]	P [21bp apart]	S [21bp apart]	γ	R: 10 bp apart (Å)	R: 21 bp apart (Å)
1	0.70	0.45	0.20	0.46	0.93	46.0	66.0
2	0.67	0.51	0.20	0.51	1.03	47.0	66.0
3	0.70	0.50	0.17	0.50	0.95	45.0	67.0
4	0.70	0.50	0.16	0.50	0.90	45.0	68.0
5	0.67	0.51	0.16	0.52	0.92	46.0	68.0
6	0.68	0.50	0.15	0.51	0.88	45.0	68.0
7	0.63	0.52	0.18	0.54	0.80	46.0	65.0
8	0.70	0.51	0.18	0.48	1.23	46.0	69.0

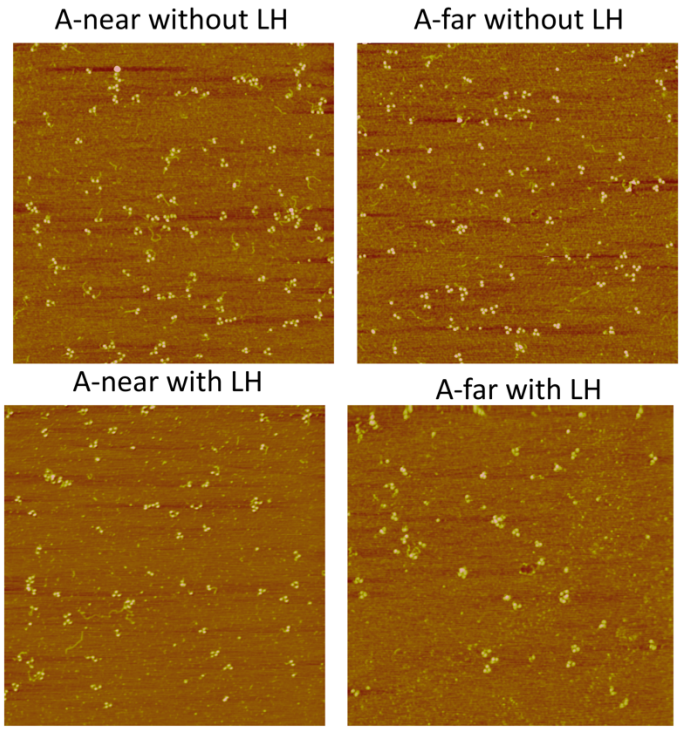
To check whether the γ factor estimation was correct, distances for the FRET standards were calculated (tabulated above) using the observed P and the (calculated using γ) FRET efficiency E. Using the software FRET Positioning and Screening (131), the R_0-independent inter-fluorophore distances (R_{mp} and $\langle R_{DA} \rangle$, these terms are explained in Chapter 3) were computed from the models to compare with the experimentally extracted distances.

Computed R_{mp} for FRET standard '10bp apart' = 44.5Å, $\langle R_{DA} \rangle$ = 47.0Å

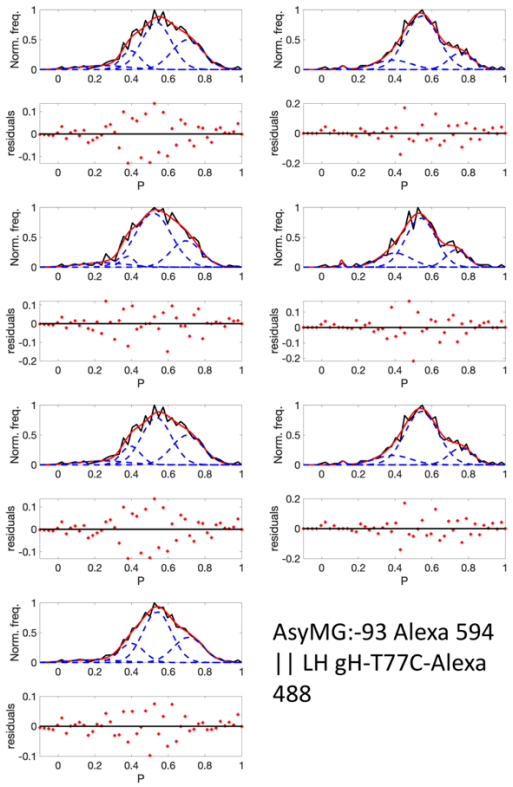
Computed R_{mp} for FRET standard '21bp apart' = 67.5Å, $\langle R_{DA} \rangle$ = 69.5Å. The experimental distances fall within this range for both the FRET standards.

Chapter 3

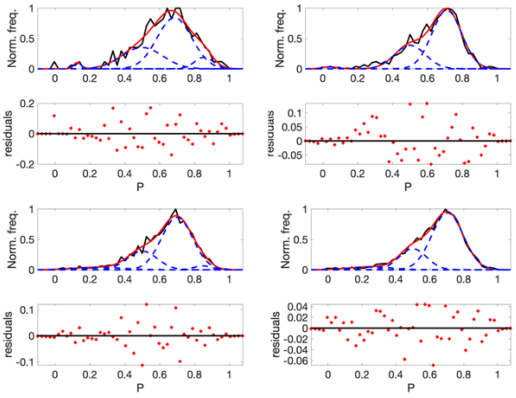
AFM pictures acquired by Martin Würtz for trichromatosomes:



Chapter 4

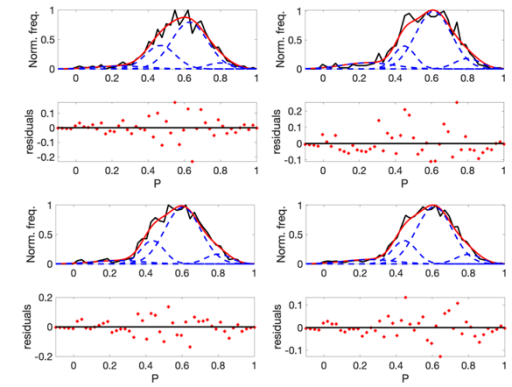


AsyMG:+94 Alexa 594 || LH gH-T77C-Alexa 488

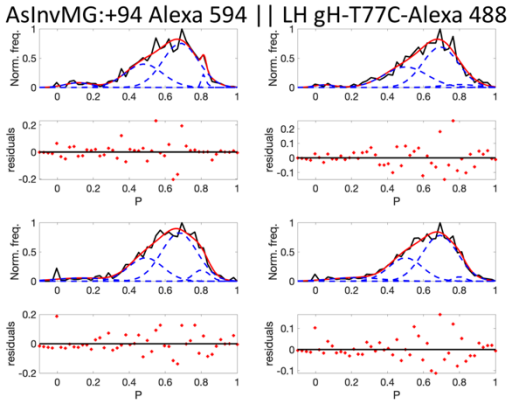


	AsyMG (-) GH-T77C	AsyMG (+) GH-T77C
P_{mean} of peak(s) of replicates	0.53 0.54 0.52 0.54 0.53	0.68 0.71 0.70 0.70
Standard deviation of P_{mean} from replicates	±0.01	±0.01
FWHM of peak(s) of replicates	0.27 0.30 0.33 0.27 0.27	0.33 0.28 0.29 0.31
% population	48.6 70.0 63.0 63.1 48.6	61.0 70.0 68.6 72.0

AsInvMG:-93 Alexa 594 || LH gH-T77C-Alexa 488

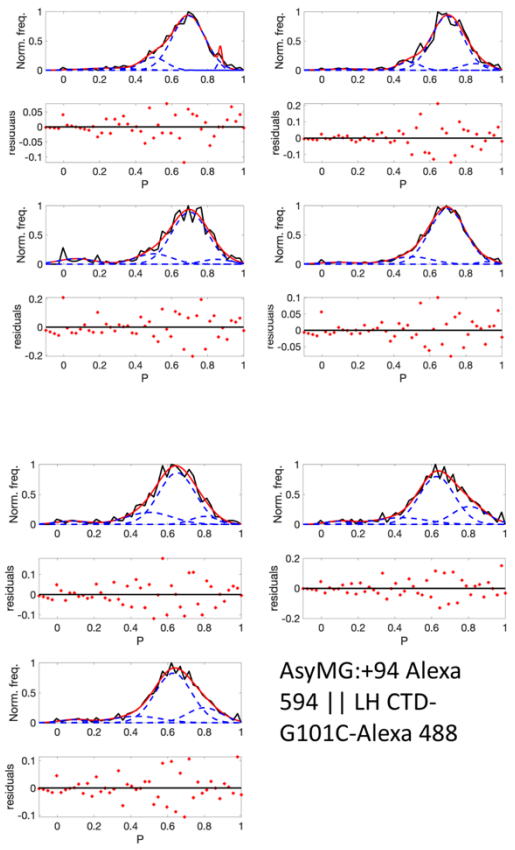


AsInvMG:+94 Alexa 594 || LH gH-T77C-Alexa 488



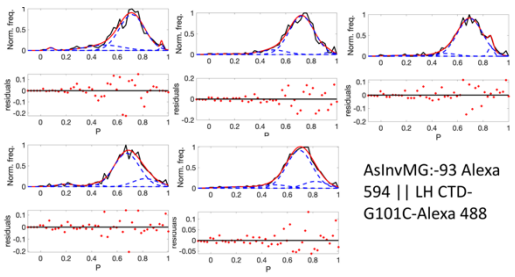
	AsInvMG (-) GH-T77C	AsInvMG (+) GH-T77C
P_{mean} of peak(s) of replicates	0.63 0.61 0.60 0.61	0.68 0.69 0.68 0.68
Standard deviation of P_{mean} from replicates	±0.01	±0.006
FWHM of peak(s) of replicates	0.33 0.33 0.33 0.33	0.33 0.33 0.33 0.33
% population	62.5 66.0 68.0 67.0	59.6 58.0 60.0 55.1

AsyMG:-93 Alexa 594 || LH CTD-G101C-Alexa 488

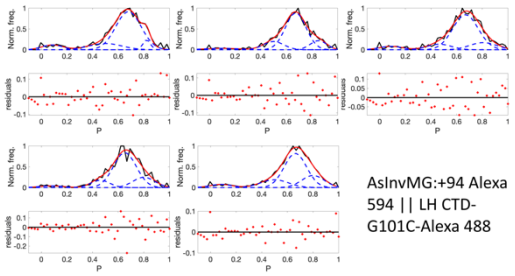


AsyMG:+94 Alexa 594 || LH CTD-G101C-Alexa 488

	AsyMG (-) CTD-G101C	AsyMG (+) CTD-G101C
P_{mean} of peak(s) of replicates	0.69 0.70 0.70 0.69	0.65 0.63 0.63
Standard deviation of P_{mean} from replicates	±0.006	±0.01
FWHM of peak(s) of replicates	0.30 0.30 0.33 0.33	0.33 0.32 0.33
% population	79.0 80.0 72.2 82.6	68.6 62.2 66.5



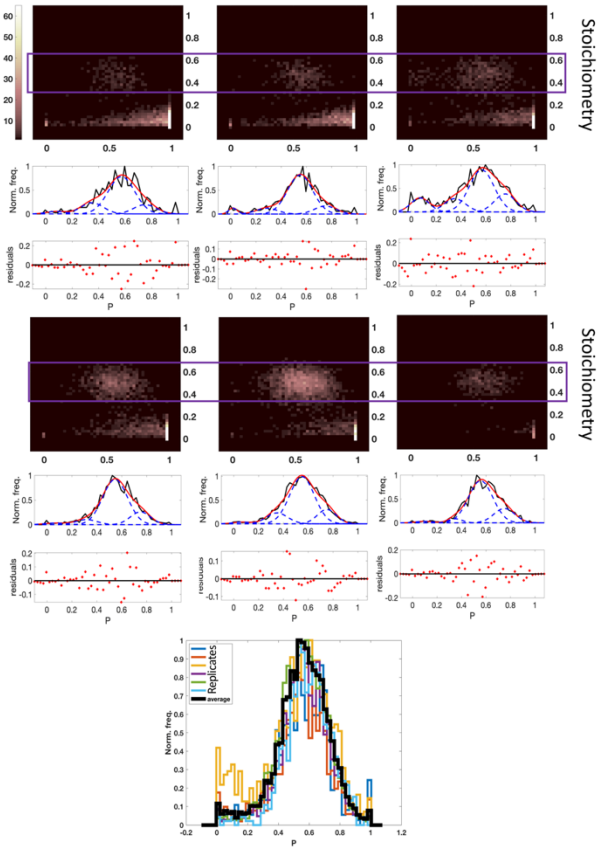
AsInvMG:-93 Alexa 594 || LH CTD-G101C-Alexa 488



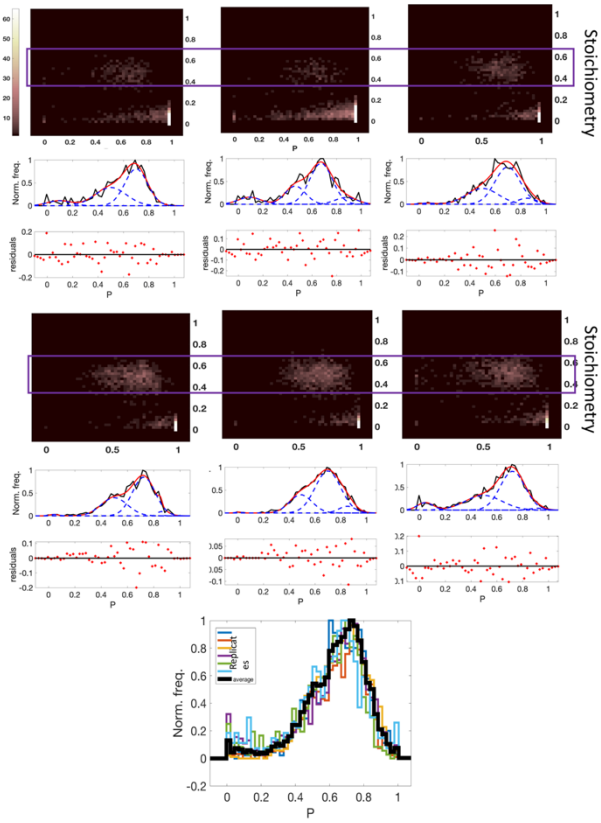
AsInvMG:+94 Alexa 594 || LH CTD-G101C-Alexa 488

	AsInvMG (-) CTD-G101C	AsInvMG (+) CTD-G101C
P_{mean} of peak(s) of replicates	0.71 0.72 0.70 0.67 0.70	0.68 0.67 0.67 0.65 0.66
Standard deviation of P_{mean} from replicates	±0.02	±0.01
FWHM of peak(s) of replicates	0.33 0.33 0.33 0.33 0.33	0.33 0.26 0.33 0.30 0.30
% population	80.0 85.4 86.4 76.0 76.2	72.0 67.3 68.7 66.4 58.0

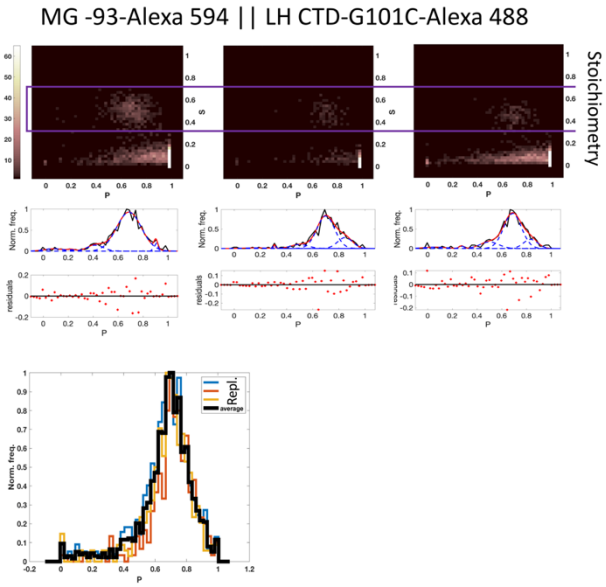
MG: -93-Alexa 594 || LH gH-T77C Alexa 488

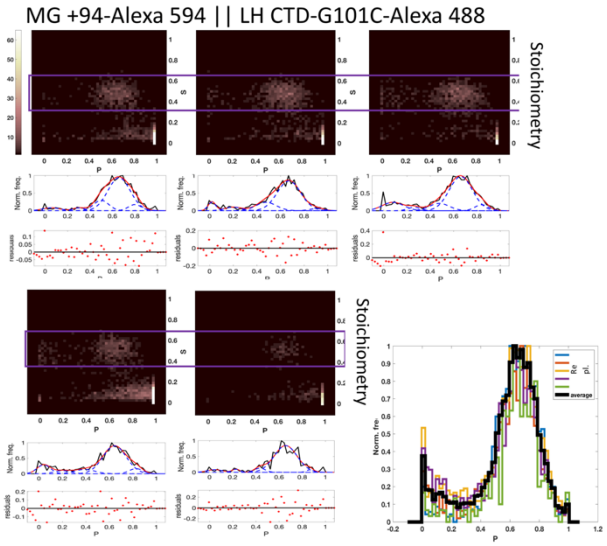


MG +94-Alexa 594 || LH gH T77C-Alexa 488



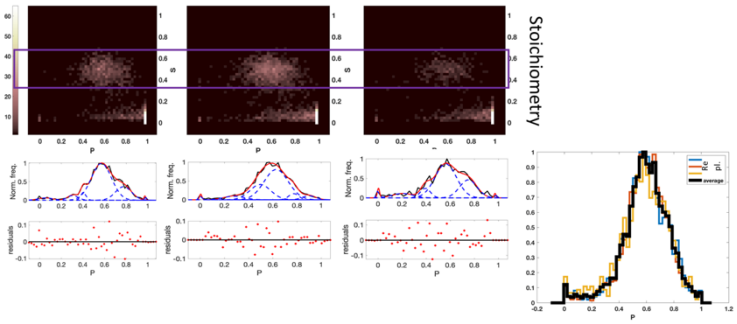
	MG (-) GH-T77C	MG (+) GH-T77C
P_{mean} of peak(s) of replicates	0.57 0.55 0.57 0.55 0.55 0.56	0.70 0.67 0.70 0.72 0.70 0.72
Standard deviation of P_{mean} from replicates	±0.01	±0.02
P_{mean} of peak in averaged plot	0.55	0.72
FWHM of peak(s) of replicates	0.33 0.33 0.33 0.33 0.33 0.33	0.30 0.30 0.33 0.29 0.33 0.32
FWHM of peak in averaged plot	0.33	0.33
% population: replicates	66.0 74.0 53.0 73.0 69.0 73.0	55.0 58.0 60.0 63.0 67.2 58.5
% population: averaged plot	67.0	63.0



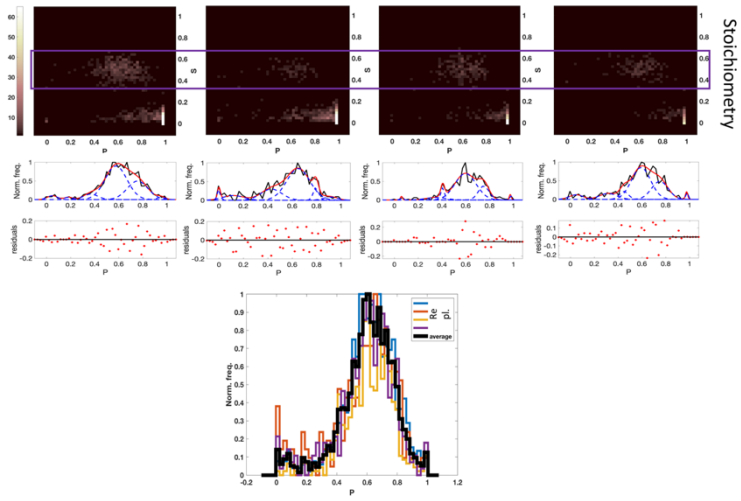


	MG (-) CTD-G101C	MG (+) CTD-G101C
P_{mean} of peak(s) of replicates	0.69 0.70 0.69	0.66 0.67 0.66 0.64 0.66
Standard deviation of P_{mean} from replicates	±0.006	±0.01
P_{mean} of peak in averaged plot	0.69	0.67
FWHM of peak(s) of replicates	0.33 0.22 0.24	0.33 0.33 0.33 0.33 0.33
FWHM of peak in averaged plot	0.27	0.33
% population: replicates	84.0 69.0 68.0	70.0 68.0 64.0 67.0 83.0
% population: averaged plot	73.0	66.0

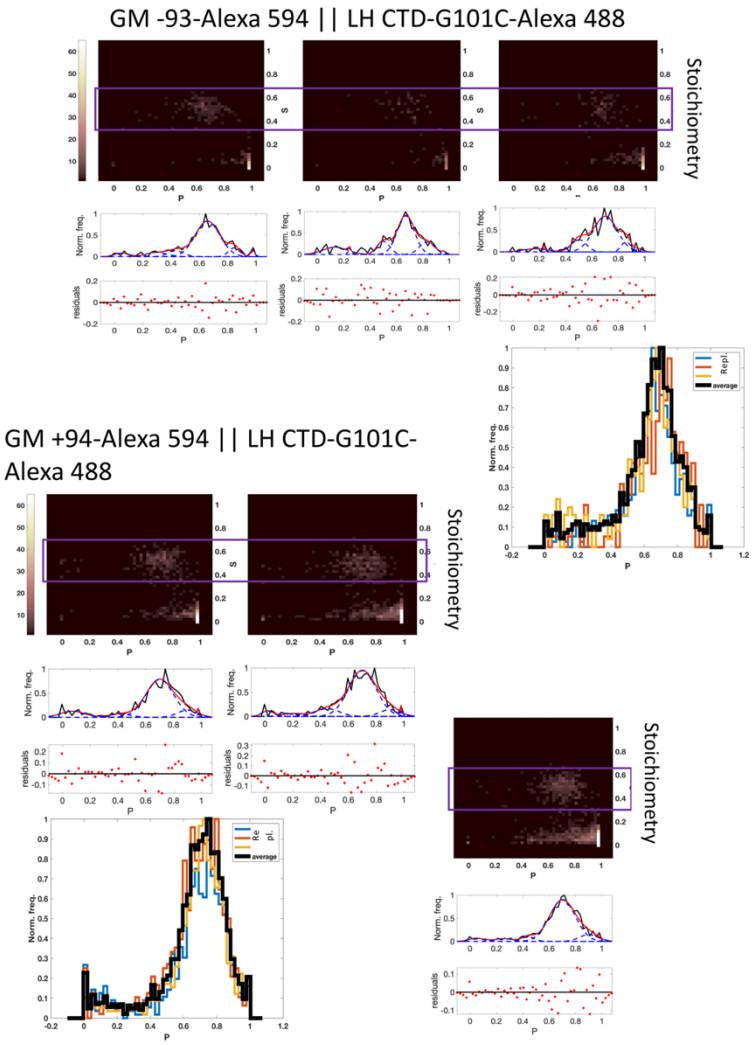
GM -93-Alexa 594 || LH gH-T77C-Alexa 488



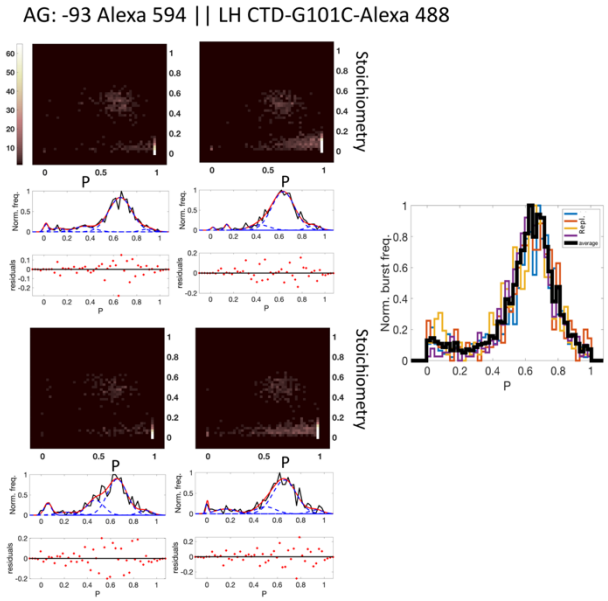
GM +94-Alexa 594 || LH gH-T77C-Alexa 488

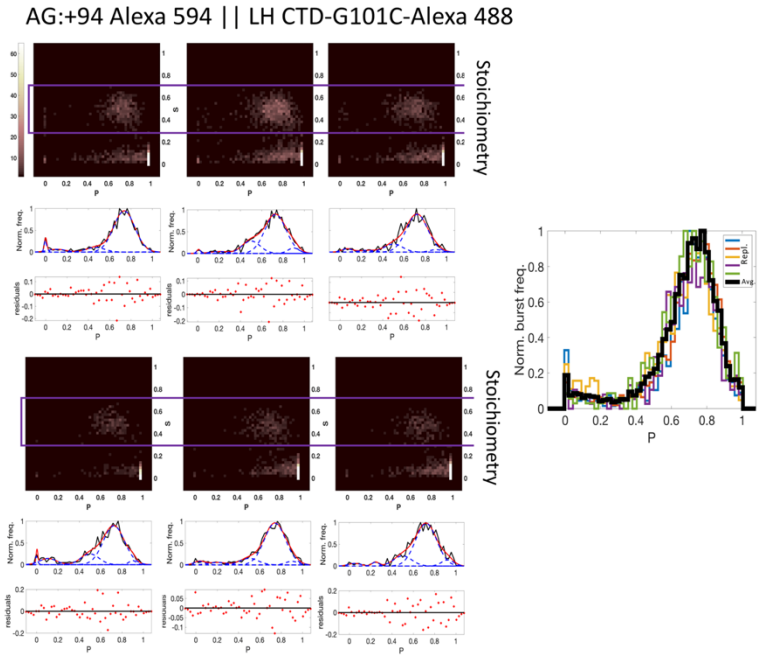


	GM (-) GH-T77C	GM (+) GH-T77C
P_{mean} of peak(s) of replicates	0.57 0.60 0.56	0.58 0.65 0.60 0.60
Standard deviation of P_{mean} from replicates	±0.02	±0.03
P_{mean} of peak in averaged plot	0.57	0.60
FWHM of peak(s) of replicates	0.30 0.33 0.30	0.30 0.32 0.30 0.26
FWHM of peak in averaged plot	0.32	0.28
% population: replicates	68.0 60.0 58.0	58.0 61.0 68.0 52.3
% population: averaged plot	67.0	57.0

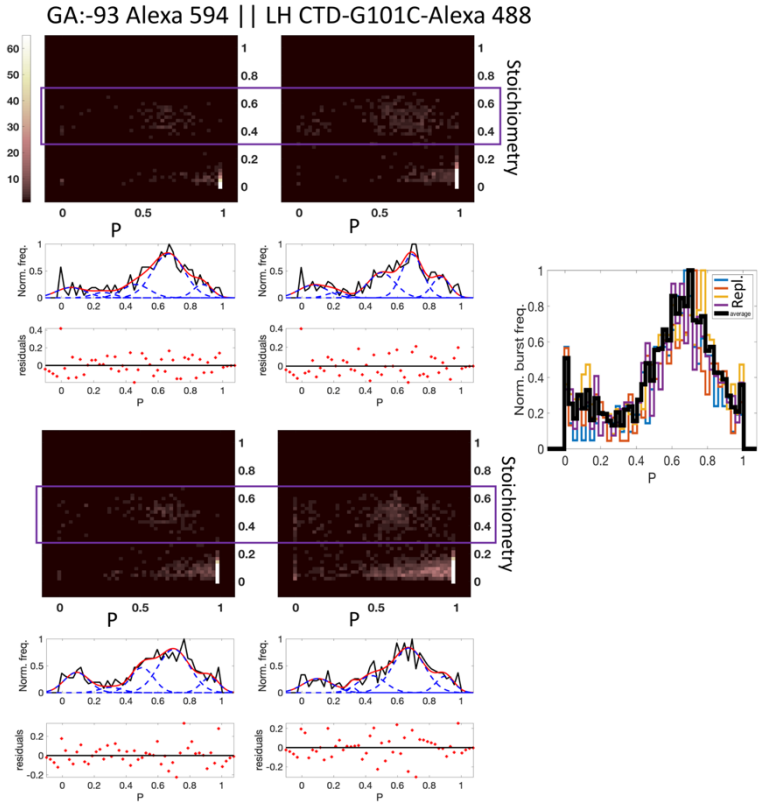

GM -93-Alexa 594 || LH CTD-G101C-Alexa 488
Stoichiometry
Norm. freq.
residuals
P
GM +94-Alexa 594 || LH CTD-G101C-Alexa 488
Stoichiometry
Norm. freq.
residuals
P
Repl.
average

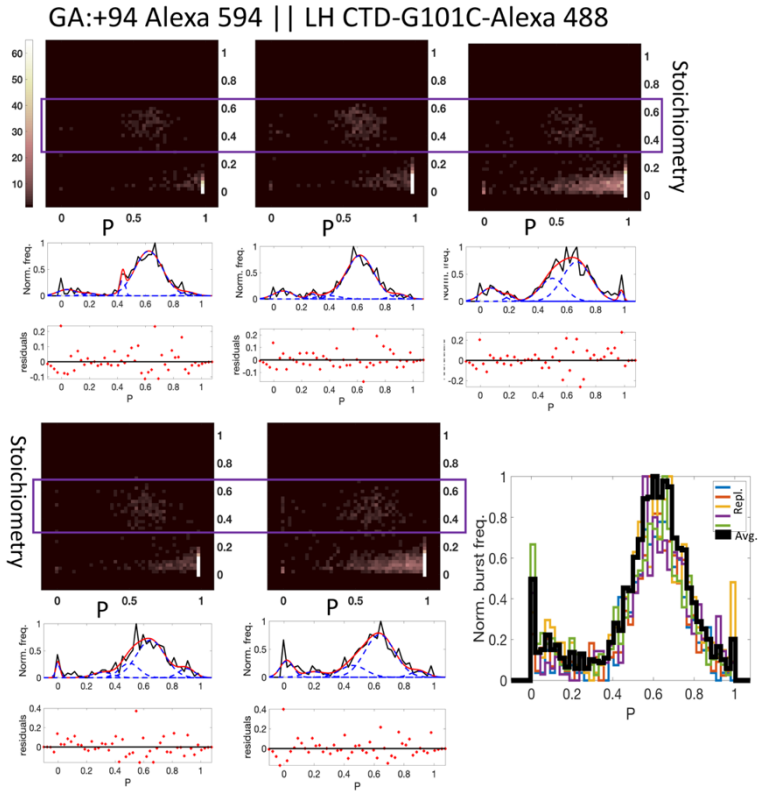
	GM (-) CTD-G101C	GM (+) CTD-G101C
P_{mean} of peak(s) of replicates	0.66 0.70 0.67	0.69 0.70 0.68
Standard deviation of P_{mean} from replicates	±0.02	±0.01
P_{mean} of peak in averaged plot	0.68	0.70
FWHM of peak(s) of replicates	0.28 0.30 0.21	0.33 0.33 0.33
FWHM of peak in averaged plot	0.26	0.33
% population: replicates	71.3 69.0 54.3	74.0 74.0 79.0
% population: averaged plot	60.0	77.0



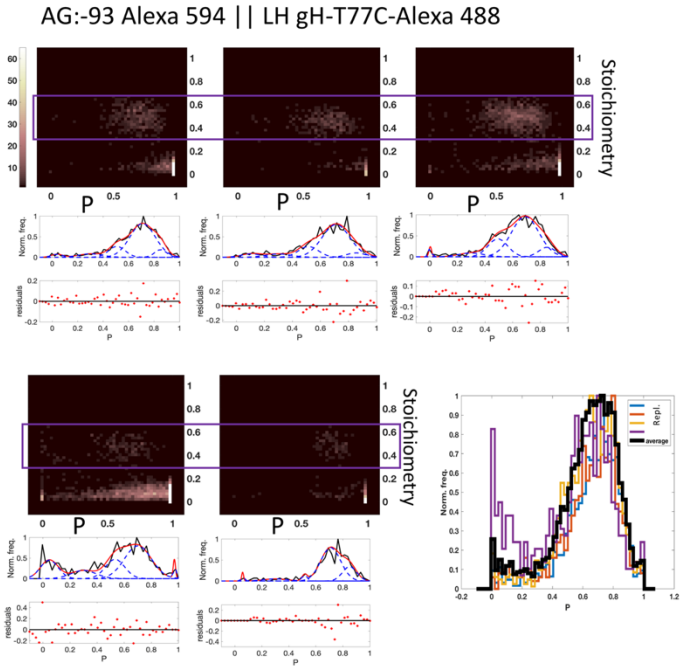


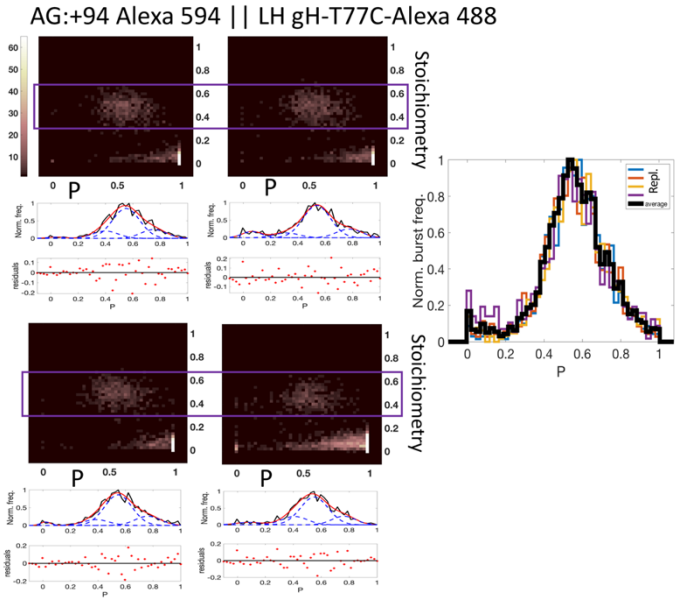
	AG (-) CTD-G101C	AG (+) CTD-G101C
P_{mean} of peak(s) of replicates	0.66 0.66 0.66 0.64	0.74 0.74 0.72 0.72 0.73 0.72
Standard deviation of P_{mean} from replicates	±0.01	±0.01
P_{mean} of peak in averaged plot	0.65	0.73
FWHM of peak(s) of replicates	0.33 0.33 0.28 0.33	0.33 0.33 0.33 0.33 0.33 0.33
FWHM of peak in averaged plot	0.33	0.33
% population: replicates	78.0 71.0 60.0 81.0	84.0 78.0 70.0 77.0 71.0 74.0
% population: averaged plot	76.0	76.0

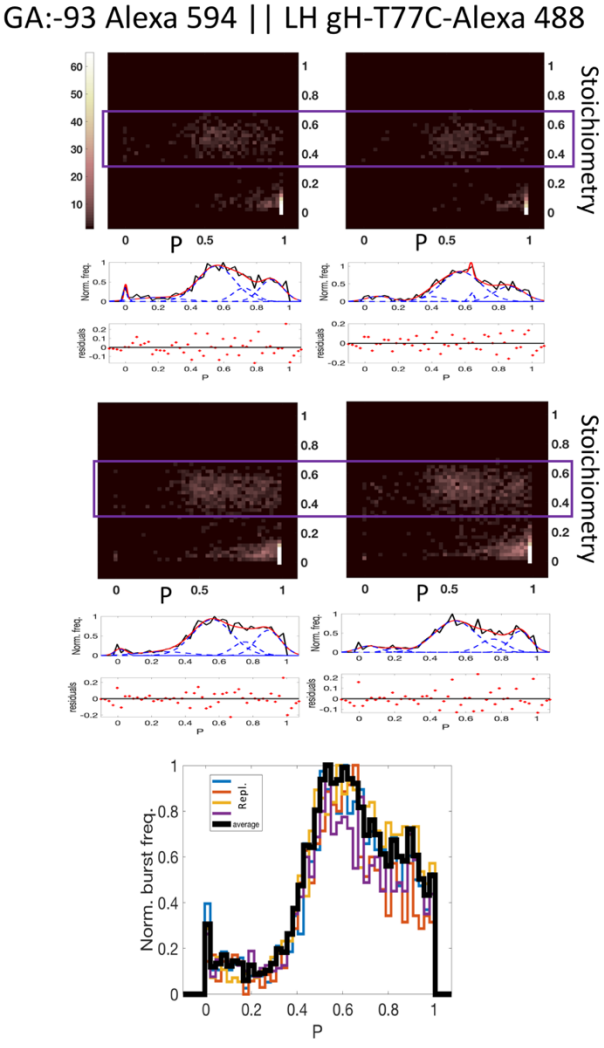

GA:-93 Alexa 594 || LH CTD-G101C-Alexa 488
Stoichiometry
P
Norm. freq.
residuals
Norm. burst freq.
Repl.
average

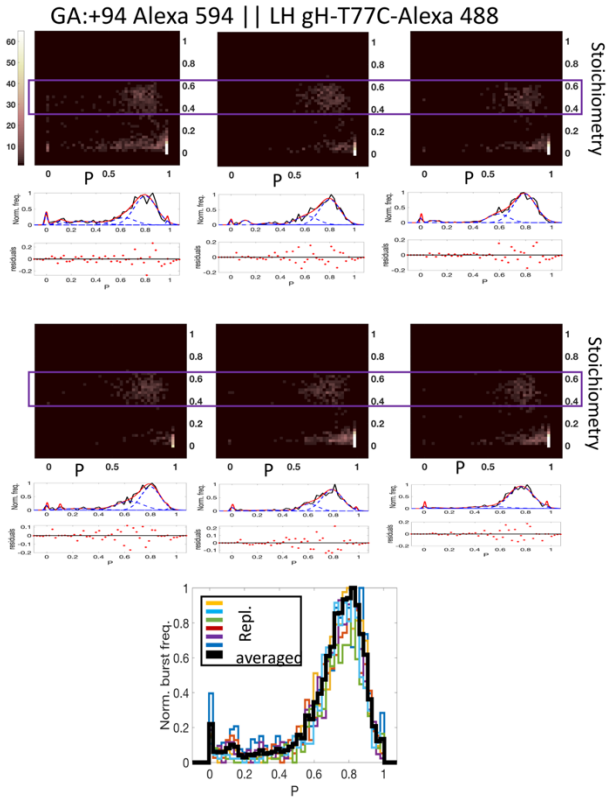

GA:+94 Alexa 594 || LH CTD-G101C-Alexa 488
Stoichiometry
P
Norm. freq.
residuals
Norm. burst freq.
Repl.
Avg.

	GA (-) CTD-G101C	GA (+) CTD-G101C
P_{mean} of peak(s) of replicates	0.66 0.50 0.70 0.70 0.67	0.62 0.62 0.67 0.65 0.64
Standard deviation of P_{mean} from replicates	±0.02	±0.02
P_{mean} of peak in averaged plot	0.70	0.65
FWHM of peak(s) of replicates	0.33 0.28 0.22 0.33 0.33	0.33 0.33 0.33 0.33 0.33
FWHM of peak in averaged plot	0.33	0.33
% population: replicates	57.6 28.0 37.0 48.0 54.0	81.0 78.0 55.0 64.0 63.0
% population: averaged plot	49.5	61.0




AG:+94 Alexa 594 || LH gH-T77C-Alexa 488
Stoichiometry
Norm. burst freq.
Repl.
average
P
Norm. freq.
residuals


GA:-93 Alexa 594 || LH gH-T77C-Alexa 488
Stoichiometry
Norm. burst freq.
Repl.
average
P
Norm. freq.
residuals



	AG (-) GH-T77C	AG (+) GH-T77C	GA (-) GH-T77C		GA (+) GH-T77C
P_{mean} of peak(s) of replicates	0.70 0.71 0.71 0.70 0.71	0.55 0.55 0.55 0.54	major 0.55 0.57 0.56 0.54	minor 0.89 0.86 0.89 0.90	0.80 0.80 0.79 0.80 0.79 0.77
Standard deviation of P_{mean} from replicates	±0.005	±0.005	±0.01	±0.02	±0.01
P_{mean} of peak in averaged plot	0.71	0.54	0.55		0.80
FWHM of peak(s) of replicates	0.33 0.33 0.33 0.33 0.24	0.33 0.33 0.33 0.33	major 0.37 0.39 0.39 0.39	minor 0.24 0.29 0.24 0.19	0.30 0.27 0.28 0.25 0.27 0.32
FWHM of peak in averaged plot	0.33	0.33	0.39	0.22	0.27
% population: replicates	69.0 67.0 56.0 51.0 68.0	65.0 61.0 67.0 66.0	major 55.0 61.0 54.0 57.0	minor 22.0 23.0 24.0 19.0	64.0 66.0 74.0 57.6 71.0 83.0
% population: averaged plot	58.0	66.0	57.0	21.0	62.0

Chapter 5

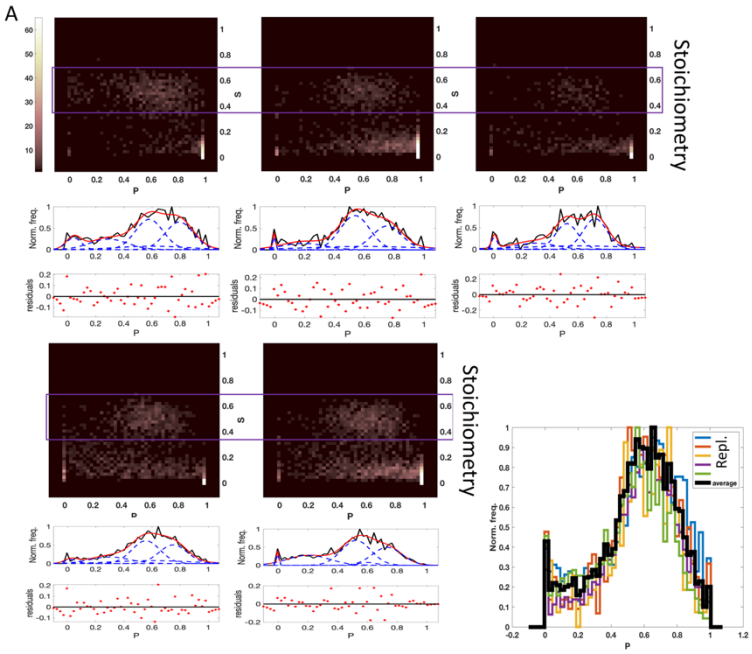


Figure SI_5.1: TU -93 Alexa 594 || LH-gH T77C-Alexa 488

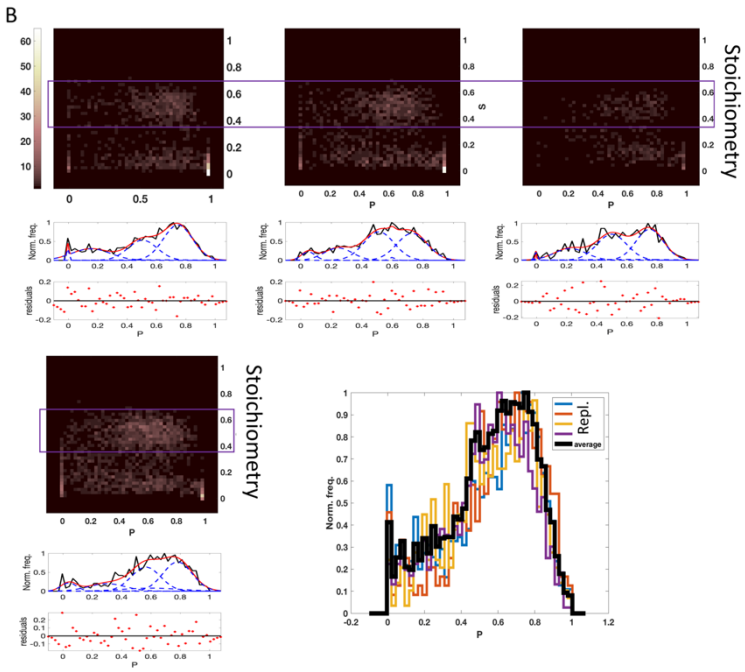


Figure SI_5.2: TU +94 Alexa 594 || LH-gH T77C-Alexa 488

	TU (-) GH-T77C		TU (+) GH-T77C	
P_{mean} of peak(s) of replicates	0.58 0.54 0.52 0.56 0.52	0.79 0.75 0.72 0.75 0.73	0.51 0.57 0.50 0.51	0.75 0.78 0.75 0.73
Standard deviation of P_{mean} from replicates	±0.03	±0.03	±0.03	±0.02
P_{mean} of peak in averaged plot	0.54	0.75	0.52	0.74
FWHM of peak(s) of replicates	0.33 0.33 0.26 0.33 0.33	0.33 0.33 0.28 0.33 0.33	0.33 0.33 0.33 0.33	0.33 0.33 0.30 0.33
FWHM of peak in averaged plot	0.33	0.33	0.33	0.33
% population: replicates	35.0 42.0 28.0 37.0 41.5	32.0 30.0 36.0 31.0 30.4	30.0 33.0 38.0 40.0	50.0 40.0 41.0 38.0
% population: averaged plot	38.0	31.0	31.0	40.0

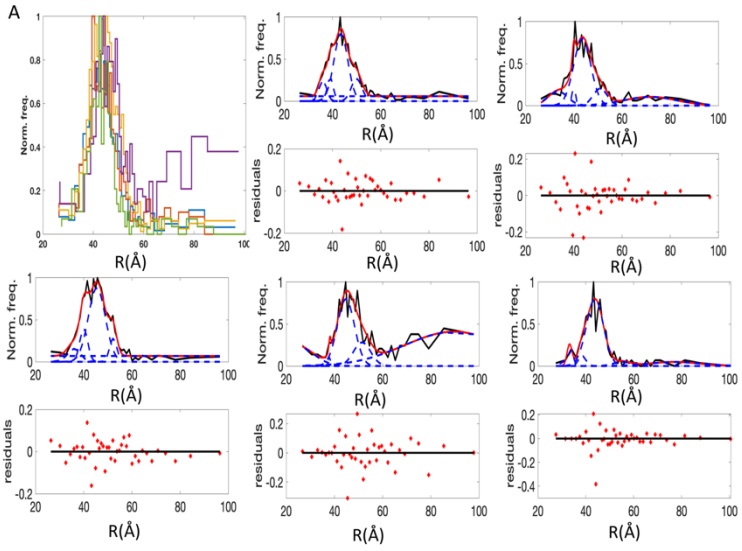


Figure SI_5.3: Distance histograms AG -93 Alexa 594 || LH-gH T77C-Alexa 488

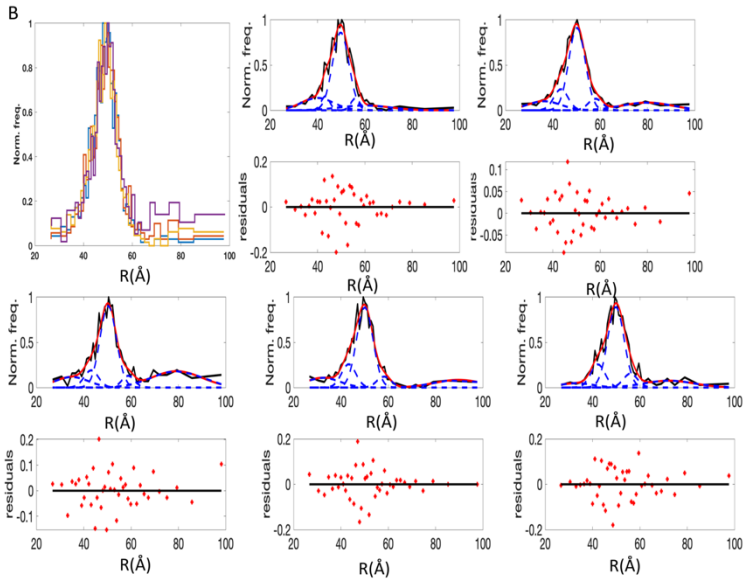


Figure SI_5.4: Distance histograms AG +94 Alexa 594 || LH-gH T77C-Alexa 488

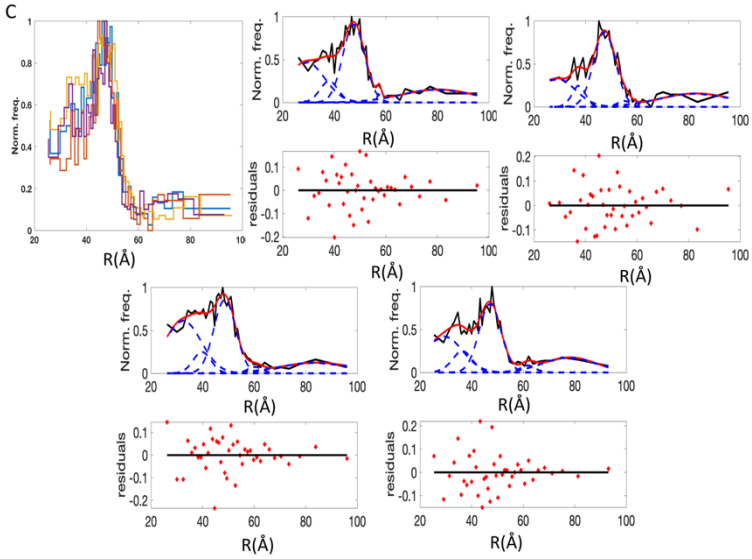


Figure SI_5.5: Distance histograms GA -93 Alexa 594 || LH-gH T77C-Alexa 488

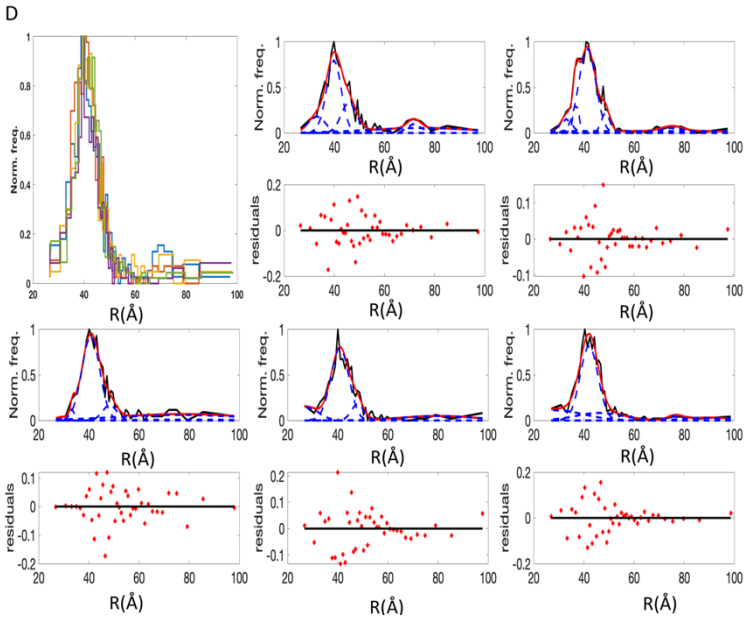


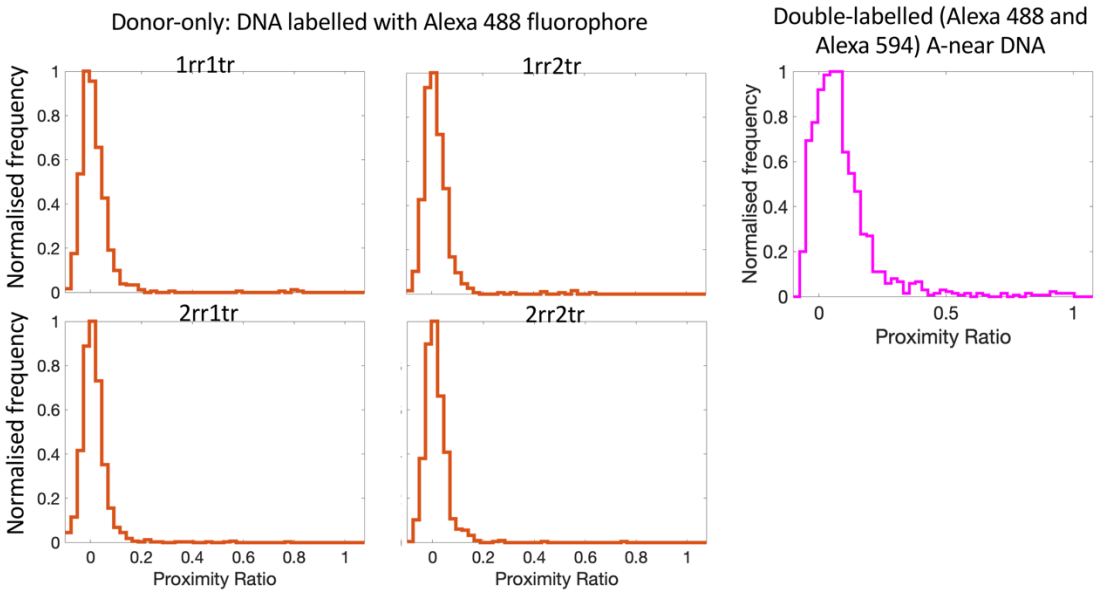
Figure SI_5.6: Distance histograms GA +94 Alexa 594 || LH-gH T77C-Alexa 488

Constructs:	AG (-) GH-T77C	AG (+)GH- T77C	GA (-) GH-T77C	GA (+) GH-T771C
Mean R of peak(s) of replicates (Å) and [γ value]	43.5 [0.95] 43.5 [0.95] 45.4 [0.96] 44.7 [1.03] 43.8 [1.23]	49.7 [0.90] 50.0 [0.90] 50.0 [0.90] 50.3 [0.93]	49.0 [0.92] 48.6 [0.92] 50.0 [0.95] 48.3 [0.80]	40.0 [0.88] 41.3 [0.88] 40.7 [0.92] 40.8 [0.91] 42.0 [0.95]
Standard deviation of mean R (Å) from replicates	±0.84	±0.24	±0.74	±0.74
Average of Mean R (Å) from replicates	44.2	50.0	49.0	41.0
Full-width at half maximum (Å) of peak(s) of replicates	9.6 12.0 12.0 12.0 12.0	11.5 11.5 11.5 11.5	14.0 14.0 14.0 14.0	10.0 11.0 12.0 12.0 12.0

Chapter 6

All FRET measurements were performed using single-laser excitation (491nm - Cobolt Calypso, Hübner Photonics).

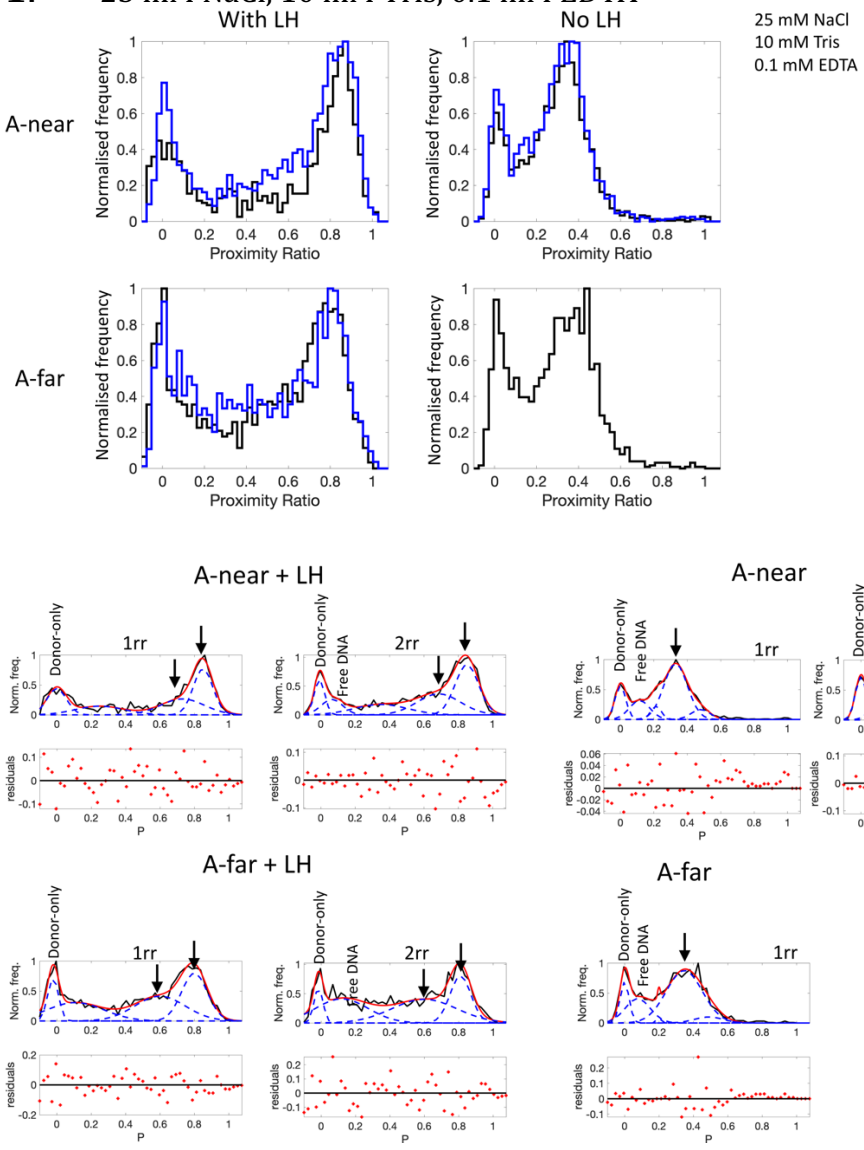
A. Donor-only single-pair FRET



B. Salt-series

I.

1. 25 mM NaCl, 10 mM Tris, 0.1 mM EDTA



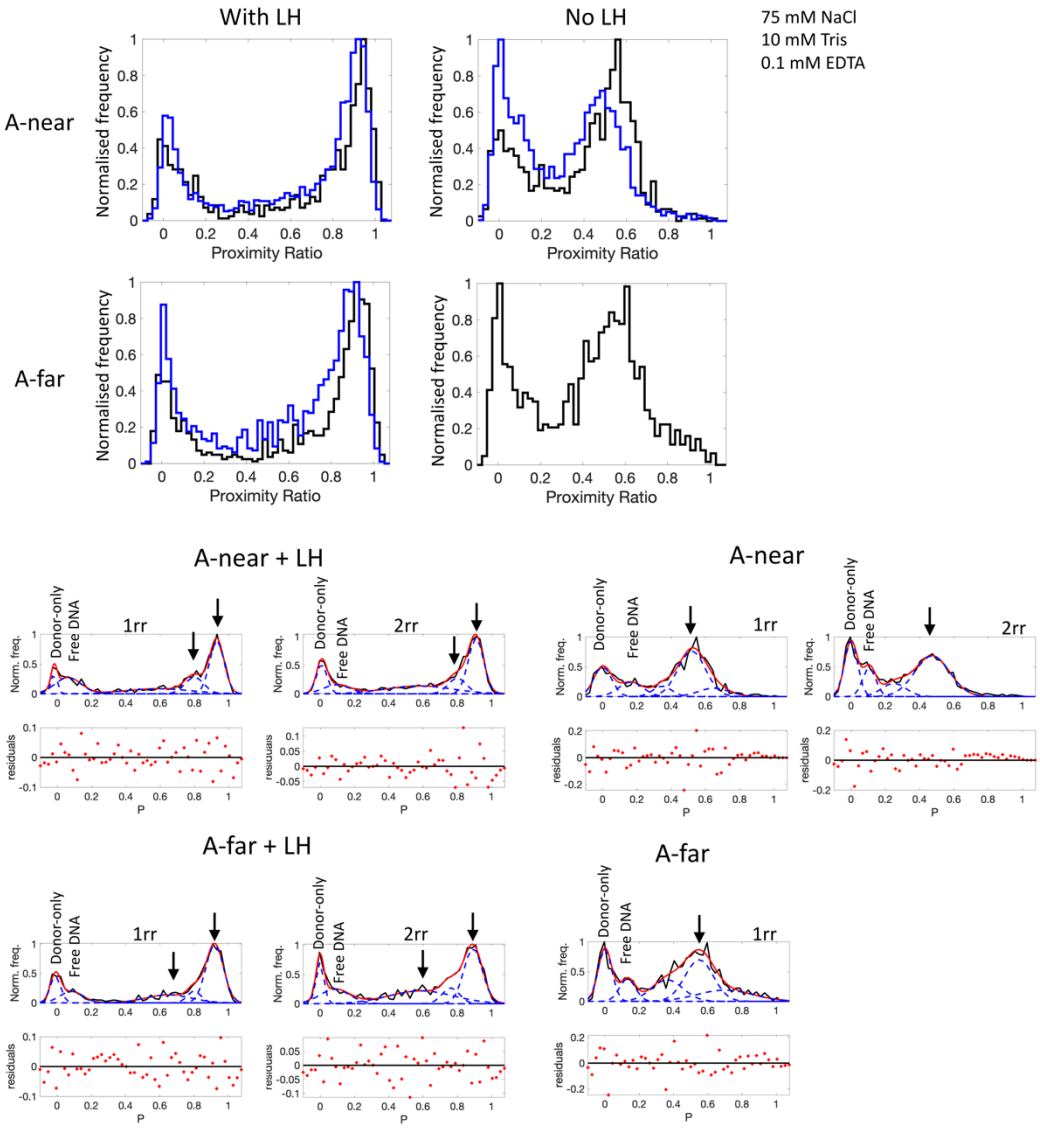
25 mM NaCl, 10 mM Tris, 0.1 mM EDTA 1st replicate (1rr)

	Mean P	% population	FWHM
A-far + LH	0.60, 0.80	31.3, 31.0	0.48, 0.25
A-near + LH	0.73, 0.85	27.5, 35.6	0.40, 0.20
A-far	0.35	61.6	0.35
A-near	0.33	55.2	0.24

25 mM NaCl, 10 mM Tris, 0.1 mM EDTA 2nd replicate (2rr)

	Mean P	% population	FWHM
A-far + LH	0.60, 0.81	37.3, 24.0	0.63, 0.20
A-near + LH	0.69, 0.85	27.2, 35.3	0.40, 0.20
A-near	0.34	49.5	0.24

2. 75 mM NaCl, 10 mM Tris, 0.1 mM EDTA



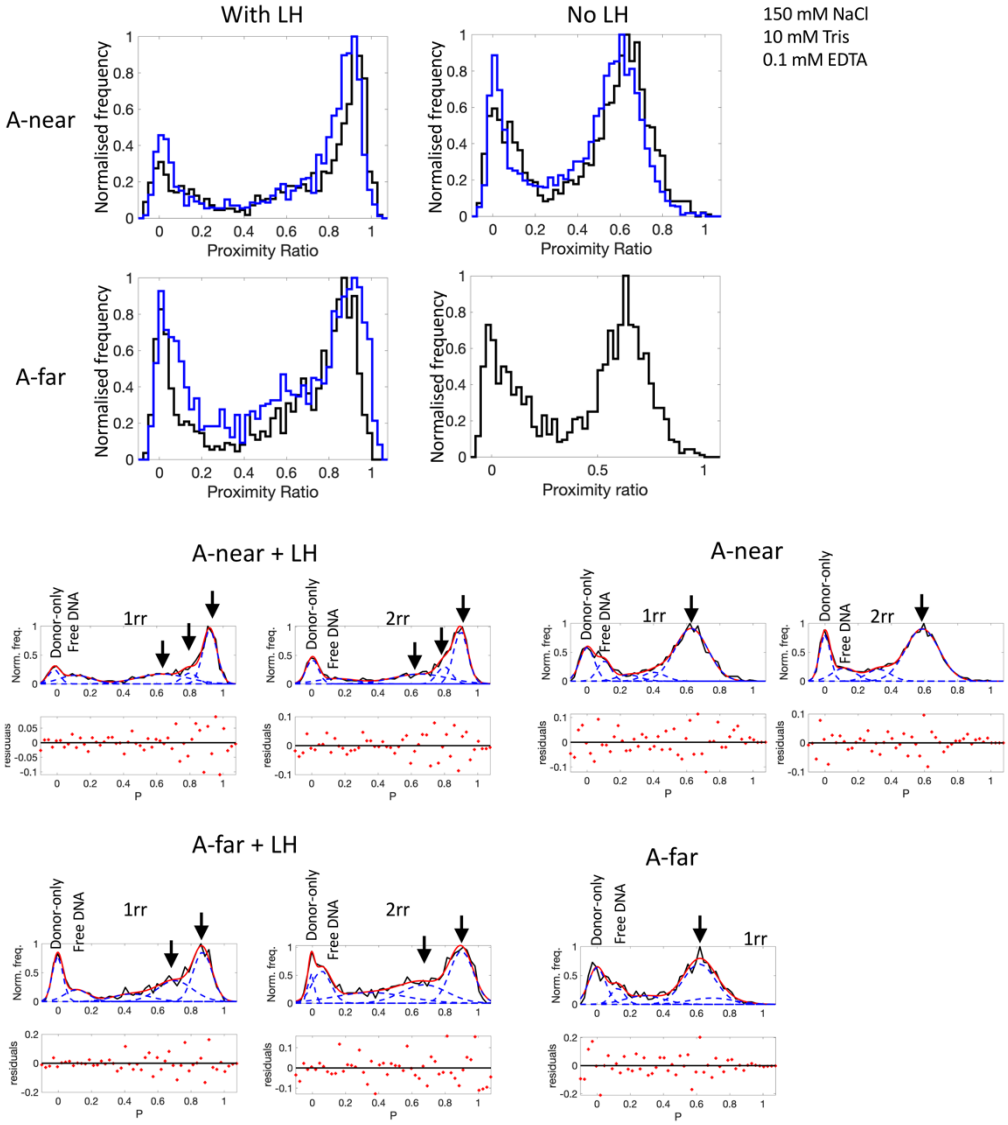
75 mM NaCl, 10 mM Tris, 0.1 mM EDTA 1st replicate

	Mean P	% population	FWHM
A-far + LH	0.65, 0.92	17.4, 49.1	0.44, 0.17
A-near + LH	0.80, 0.93	15.8, 39.4	0.20, 0.14
A-far	0.55	33.2	0.27
A-near	0.52	48.5	0.27

75 mM NaCl, 10 mM Tris, 0.1 mM EDTA 2nd replicate

	Mean P	% population	FWHM
A-far + LH	0.60, 0.90	29.0, 35.6	0.63, 0.20
A-near + LH	0.80, 0.91	12.8, 38.2	0.17, 0.15
A-near	0.47	50.0	0.34

3. 150 mM NaCl, 10 mM Tris, 0.1 mM EDTA



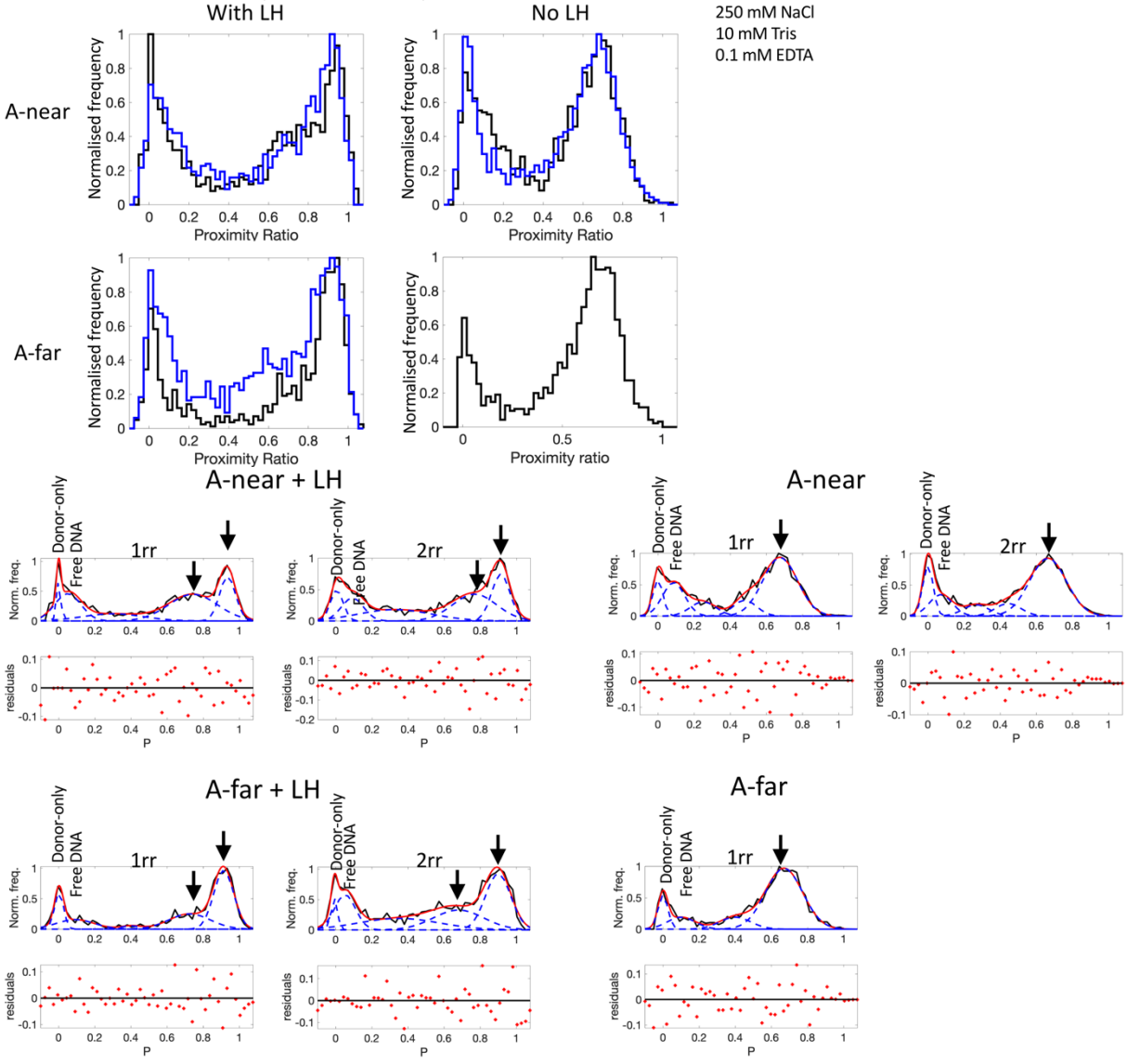
150 mM NaCl, 10 mM Tris, 0.1 mM EDTA 1st replicate

	Mean P	% population	FWHM
A-far + LH	0.70, 0.87	27.0, 34.4	0.35, 0.20
A-near + LH	0.63, 0.80, 0.92	26.7, 8.9, 41.0	0.51, 0.15, 0.14
A-far	0.62	47.2	0.32
A-near	0.63	61.0	0.33

150 mM NaCl, 10 mM Tris, 0.1 mM EDTA 2nd replicate

	Mean P	% population	FWHM
A-far + LH	0.67, 0.90	23.1, 33.8	0.45, 0.23
A-near + LH	0.65, 0.80, 0.90	26.2, 12.3, 36.7	0.54, 0.15, 0.14
A-near	0.60	62.0	0.34

4. 250 mM NaCl, 10 mM Tris, 0.1 mM EDTA



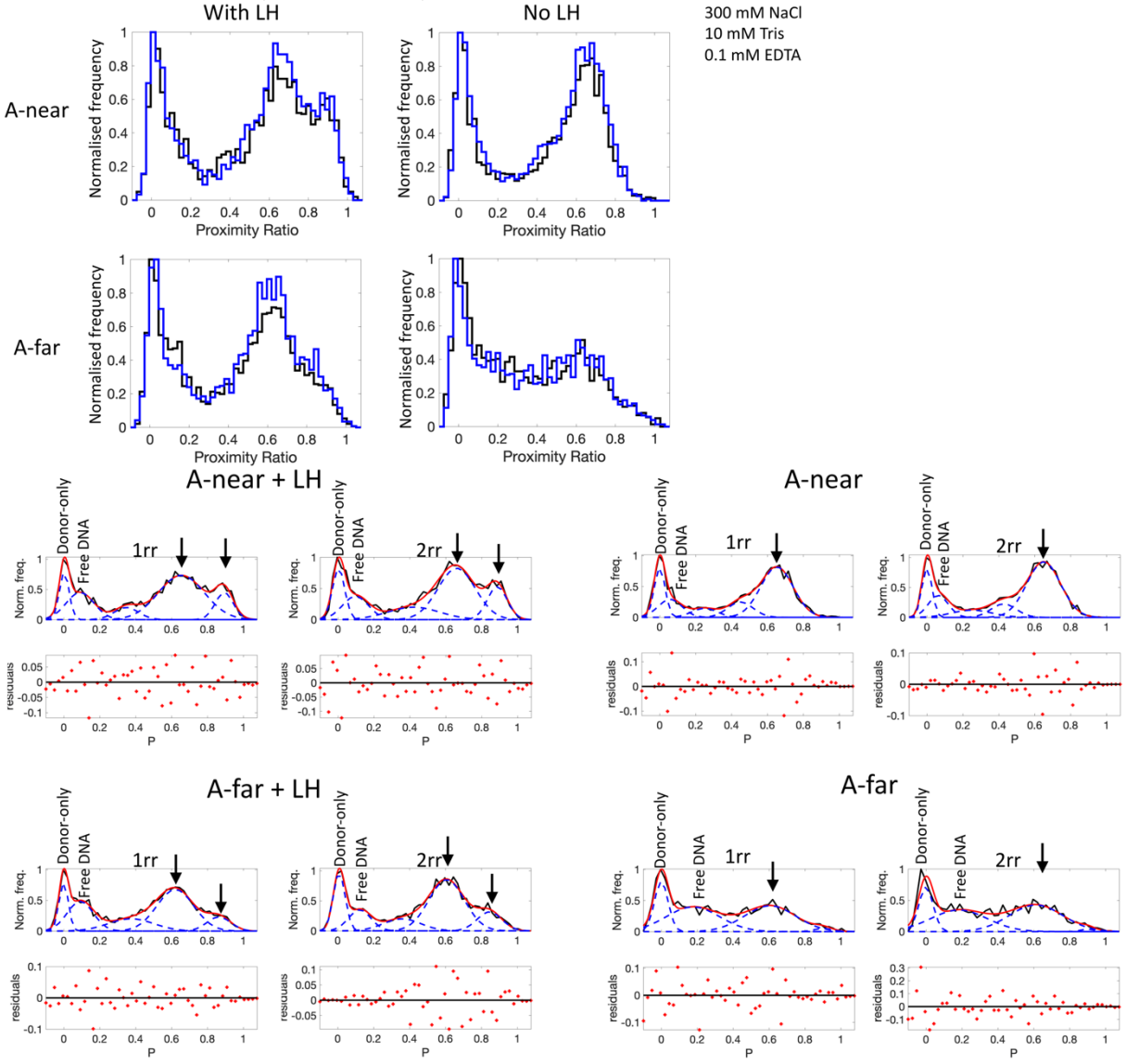
250 mM NaCl, 10 mM Tris, 0.1 mM EDTA 1st replicate

	Mean P	% population	FWHM
A-far + LH	0.72, 0.91	24.6, 45.8	0.38, 0.20
A-near + LH	0.74, 0.93	39.2, 21.1	0.45, 0.14
A-far	0.67	68.5	0.34
A-near	0.68	62.5	0.32

250 mM NaCl, 10 mM Tris, 0.1 mM EDTA 2nd replicate

	Mean P	% population	FWHM
A-far + LH	0.68, 0.90	23.3, 33.0	0.46, 0.23
A-near + LH	0.77, 0.91	31.2, 22.7	0.40, 0.17
A-near	0.66	56.6	0.34

5. 300 mM NaCl, 10 mM Tris, 0.1 mM EDTA



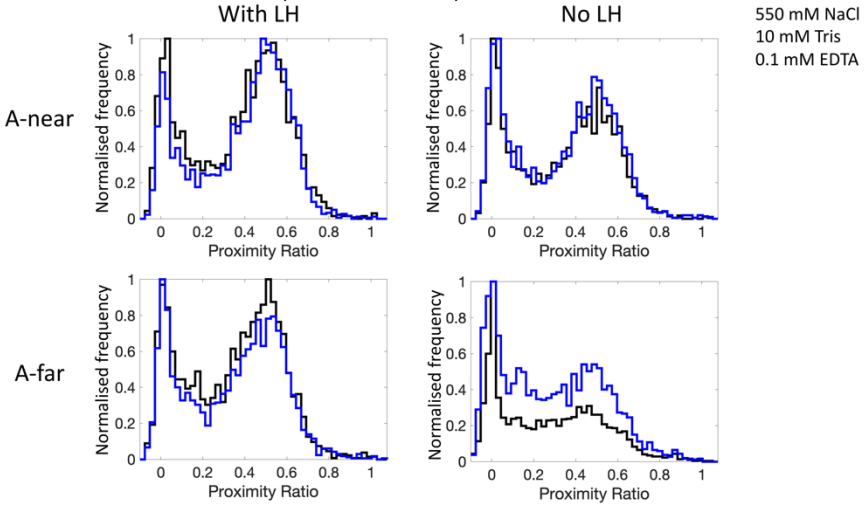
300 mM NaCl, 10 mM Tris, 0.1 mM EDTA 1st replicate

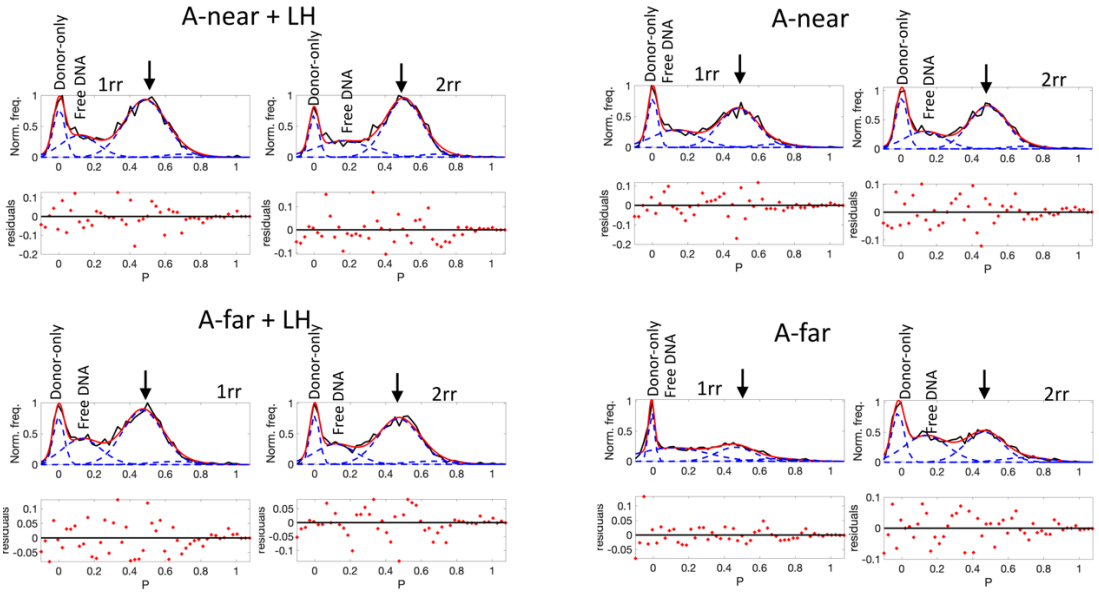
	Mean P	% population	FWHM
A-far + LH	0.62, 0.86	42.1, 9.2	0.34, 0.22
A-near + LH	0.65, 0.89	50.5, 12.1	0.43, 0.17
A-far	0.61	39.2	0.47
A-near	0.65	51.0	0.30

300 mM NaCl, 10 mM Tris, 0.1 mM EDTA 2nd replicate

	Mean P	% population	FWHM
A-far + LH	0.61, 0.84	46.7, 12.0	0.33, 0.23
A-near + LH	0.66, 0.89	43.2, 14.7	0.35, 0.20
A-far	0.61	43.1	0.50
A-near	0.65	54.4	0.31

6. 550 mM NaCl, 10 mM Tris, 0.1 mM EDTA





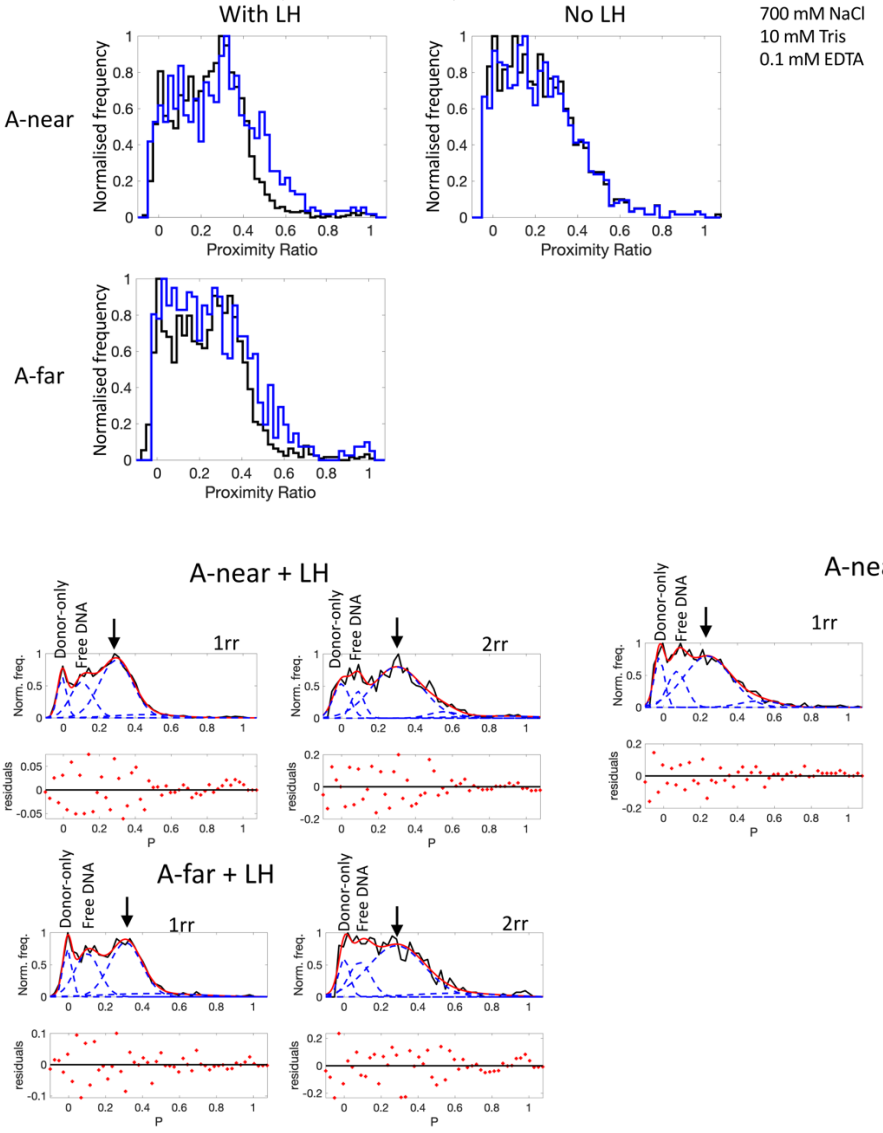
550 mM NaCl, 10 mM Tris, 0.1 mM EDTA 1st replicate

	Mean P	% population	FWHM
A-far + LH	0.47	57.0	0.38
A-near + LH	0.49	62.2	0.40
A-far	0.46	28.2	0.35
A-near	0.48	51.0	0.35

550 mM NaCl, 10 mM Tris, 0.1 mM EDTA 2nd replicate

	Mean P	% population	FWHM
A-far + LH	0.47	57.2	0.39
A-near + LH	0.50	61.1	0.34
A-far	0.46	36.6	0.35
A-near	0.49	54.0	0.35

7. 700 mM NaCl, 10 mM Tris, 0.1 mM EDTA



700 mM NaCl, 10 mM Tris, 0.1 mM EDTA 1st replicate

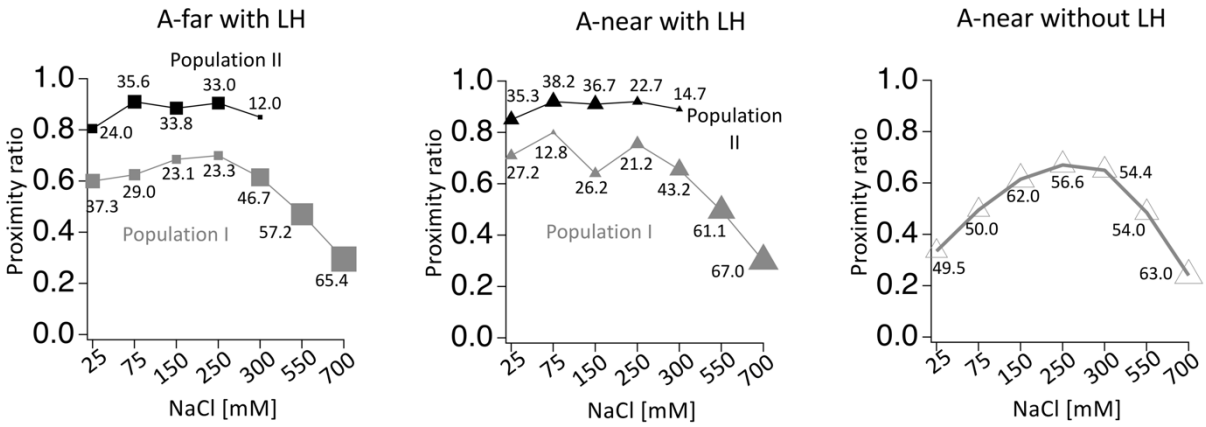
	Mean P	% population	FWHM
A-far + LH	0.31	49.2	0.30
A-near + LH	0.30	58.0	0.31
A-near	0.24	62.8	0.42

700 mM NaCl, 10 mM Tris, 0.1 mM EDTA 2nd replicate

	Mean P	% population	FWHM
A-far + LH	0.28	65.4	0.50
A-near + LH	0.30	67.0	0.45
A-near	0.24	63.0	0.41

II. From 2nd replicate of constructs A-near with/without LH, A-far with LH: trends in % population of FRET peaks

Each peak is associated with the relative population (in %) of that peak, and the marker size is scaled according to the peak population.



Acknowledgements

And now the time has come to look back, and to say thanks.
First and foremost, to late Professor Jörg Langowski, who I never got a chance to thank in person. I am grateful to him for giving me the opportunity to work at the Biophysics of Macromolecules Division, for inspiring me to (as he put it) 'go to the lab and get my hands dirty', for encouraging me with tremendous enthusiasm to work on linker histones. This drive helped and motivated me. Thank you.
Dr. Katalin Tóth, for steering the ship through the storm. Thank you for your strength and your unwavering enthusiasm that kept the lab afloat.
Professor Rebecca Wade, who took up an enormous responsibility of taking up a student from a very different field. Thank you for immense support, patience, and continuous help through these years. Rebecca and Kati, thanks for helping me mould and finish the work, despite tough circumstances.
Professor Jeffrey Hayes (Uni. Rochester Medical Center) and Dr. Amber Cutter so very kindly provided us with the linker histone protein and plasmids to work on. We are very grateful, thank you.
Professor John van Noort (Uni. Leiden), for the tremendously helpful advice of using primers to incorporate uridines. Without this step, a lot of the work would have remained incomplete and perhaps unanswered. Thank you!
To my thesis committee since 2017: Prof. Dirk-Peter Herten, Prof. Jeroen Krijgsveld, Prof. Karsten Rippe, your suggestions and tips really helped me finish the experiments in time. Thank you!
Division Biophysics of Macromolecules: I never wanted the lab to close! My thanks to Dr. Alexander Gansen (Alex!) for helping whenever spFRET troubled us! Kathrin, for your supervision and friendship and for the memories we shared at San Francisco ;) and in Heidelberg. Gabi, Maria, Nathalie, the real superheroes behind the screens, it had been wonderful to work with you all! I have made all of you quite crazy, numerous times, but thank you so much for always obliging, I learnt a lot from you all! Gabi, thanks so much for checking the thesis too!
Norbert, the time-keeper and motivator for any student trying to catch some sleep in the laser room. Thanks for your daily support and optimism! Christina, thanks, for your optimism and our fun discussions during lunch time. B040 gave me wonderful officemates: Ruihan, I fondly remember the times we shared, and even, for a short while, Don and Ada Olins, with whom I shared fruitful discussions back in '17. For a newcomer chromatin enthusiast, this was nothing less than a dream!
Lucas, thanks for being the perfect opponent for our day-to-day dad joke sessions! Thanks for sharpening my wit (or otherwise), for your take on cubism (!!) and for Peter the Plant! Martin, thanks for being a great collaborator for the AFM and for being a wonderful friend! My students,

Sarah and Angga, for not just being such great students yourselves, but for giving me the opportunity to explore my new found love for teaching!
After B040 dissolved, to my much pleasant surprise, I found another family at MCM!
Mehmet, thanks for being such a great collaborator and friend! Thanks for always patiently listening to my rants, and sometimes, my jokes. It is wonderful, working with you. Ariane, thanks for all the wonderful time and for being such a role model! Stefan, for always being there to support, Daria, Goutam, Christina, Alexandros for being such great office mates! Abraham, Jonathan, Manuel, Bosco, for our exciting discussions in science and life, and now Giulia & Giulia (and Alfie) for upholding the Friday coffee tradition! Thank you all for making the pandemic tolerable!
The last nearly five years or so had been mostly tremendous upheavals. Not to mention, a global pandemic. But I've always had help.
Krzysztof, thank you for being such a good mentor and a great friend during my early years here in Europe. Thanks for supporting me through some pretty rough times!
Irisita, what a pleasure it has been to have met you (Science March of '17)! Thanks so much for those passionate discussions about science, life, freedom, and politics! I can't thank you enough for giving me the opportunity to raise a voice at the '19s Women's March!
Sankar dadu, thank you for all the exciting conversations we had on science, back at your place, and for your unfaltering support!
My closest friends here, three strong women: Anja, Jagoda, Damaris: you three had been with me through my darkest phases to the light! I think of you three as fireflies (Anja, not glow-worms!), precious pin pricks of light when the night is blind. I feel blessed and to have met you three! Thanks for being my strength and for always being there!
And finally, my support system back home! To my school-mates, friends for decades now, thanks for the unending, decade-long conversations! It bridges distances of thousands of kilometres!
Rdi, and Bda, two of the strongest people I know, thanks for supporting me through my first year in Germany, and for your bravery that I draw strength from! My pesky younger cousins, for your patience (and often appreciable retaliations) at my relentless meme-making. Its great to have a laugh, especially during the current times!
And now, to my family.
Rubin, we chased dreams through the dusty streets of Calcutta, and then we chased the Northern Lights through the minus twenty something Arctic nights! Thank you for being my best mate in science and travels and life.
Didi, thank you for patiently waiting for me. And for constantly cheering me onward! Baba and Ma for making me fall in love with science in the first place, and for your relentless support all the time and anytime I need it. And Ma, maybe I should quote this line from R. Thakur ☺ :

"যেই ভাবে, মা যে হয় মাটি তার
ভালো লাগে আরবার
পৃথিবীর কোণটি।"