CHARACTERISING A BIOLOGICALLY RELEVANT PROTEIN-G4 INTERACTION: HP1α AND TERRA

A thesis presented in partial fulfilment of the requirements for the degree of
Master of Science
in
Genetics
at Massey University, Palmerston North,
New Zealand.

Ruby Jean Roach

2019

Abstract

Our genetic material is intricately folded and protected through the formation of a compact nucleoprotein complex, termed heterochromatin. In addition to controlling the expression of genes, heterochromatin formation is important for the structural integrity of our genome, specifically for the centromeres, the central attachment point of our chromosomes, and also the telomeres, the ends of our chromosomes. The way in which the heterochromatin in these areas is formed and maintained is though the recruitment and binding of the pivotal chromatin regulator, Heterochromatin Protein 1α (HP1α). The current model that explains how and why HP1α is recruited to, and maintained at, regions of constitutive heterochromatin is simple: HP1α binds post-translational modifications on histones (eg. H3K9me3). However, this binding is not high affinity, therefore may not be the sole determinant in HP1α localisation. At the centromeres, it has been shown that a long non-coding RNA transcribed from the pericentromeres is responsible for recruiting HP1α to this crucial region. At the telomeres, it is proposed that the long non-coding RNA transcribed from the telomeric DNA is responsible for this same purpose. Because of its guanine-rich sequence, it is able to form a non-canonical nucleic acid structure, the G-quadruplex, which may provide the specificity for heterochromatin formation at telomeres. This TElomeric Repeat-containing RNA (TERRA) has been implicated in telomeric elongation pathways, relating it to the immortalisation of cancer cells. It was found that HP1α can specifically recognise the parallel topology of TERRA, binding with high affinity through HP1α's unstructured hinge. HP1α was also shown to bind other G-quadruplexes of parallel topology, specifically DNA present in promoter and regulatory regions of many proto-oncogenes, implicating HP1α in potential G-quadruplex-dependent gene expression regulatory mechanisms. The interaction shown here between HP1α and TERRA shows a novel mechanism of telomeric heterochromatin formation, providing crucial insights into telomere maintenance and health in human cells, and furthering our understanding of the role of G-quadruplexes in the epigenome.

Acknowledgements

I would firstly like to thank the establishments that supported me in the form of scholarships: Lovell & Berys Clark, Graduate Women Manawatu, and Massey University. Thank you for allowing me to survive as a student.

I would also like to thank the Palmerston North Medical Research Foundation for funding this research.

To my supervisors, I would like to say thank you for giving me the freedom to pursue my interests throughout the project, and for the moral support and encouragement.

To all the people around the science towers that provided me with entertainment, excellent banter, and well needed distractions, I am very grateful.

To all of those who have supported and encouraged me throughout the entirety of my study, thank you for keeping me sane.

Abbreviations

A	Adenine
ac	Acetylation
ALT	Alternative lengthening of telomeres
ap	Anti-parallel
APS	Ammonium persulfate
ATRX	α-thalassemia/mental-retardation-syndrome-X-linked
BLI	Biolayer interferometry
bp	Base-pair
C	Concentration
C	Cytosine
Cal	Calorie
CBX	Chromobox
CD	Circular dichroism
ChIP	Chromatin immunoprecipitation
CSD	Chromoshadow domain
CTE	C-terminal extension
D-loop	Displacement loop
Da	Dalton
DMS	Dimethylsulphate
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
EDTA	Ethylenediaminetetraacetic acid
G	Guanine
G4	G-quadruplex
HDAC	Histone deacetylase
His	Histidine
HMT	Histone methyltransferase
HP1	Heterochromatin protein 1
I	Intensity
i-motif	Intercalated motif
IB	Interaction buffer
IMAC	Immobilized metal ion affinity chromatography
IP	Immunoprecipitation
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
K_a	Association constant
K_d	Dissociation constant
K_D	Equilibrium constant
LB	Lysongeny broth
log	Logarithm
lncRNA	Long non-coding RNA
M	Mol/L
me	Methylation

Abbreviations continued

mol	Mole
mRNA	Messenger RNA
MS	Mass spectrometry
MW	Molecular weight
MWCO	Molecular weight cut off
nt	Nucleotide
NTE	N-terminal extension
p	Parallel
PAGE	Polyacrylamide gel electrophoresis
PML	Promyelocytic leukaemia
POT1	Protection Of Telomeres 1
Psi	Pounds per square inch
PTM	Post-translational modification
q	Momentum transfer
QGRS	Quadruplex forming G-rich sequences
RAP1	Repressor/Activator Protein 1
Rf	Retention factor
R_g	Radius of gyration
RHAU	RNA Helicase Associated with AU-rich Element
RNA	Ribonucleic acid
rpm	Revolutions per minute
s	Seconds
SANS	Small-angle neutron scattering
SAXS	Small-angle X-ray scattering
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
ssDNA	Single-stranded DNA
T	Thymine
T-loop	Telomere loop
TEMED	Tetramethylethylenediamine
TERRA	Telomeric repeat-containing RNA
TIN2	TRF1 interacting nuclear protein 2
TPP1	Tripeptidyl Peptidase 1
TRF1	Telomere Repeat Factor 1
TRF2	Telomere Repeat Factor 2
tRNA	Transfer RNA
U	Uracil
WAXS	Wide-angle X-ray scattering
WT	Wildtype

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The molecule of life was discovered by accident in the mid-nineteenth century. Termed "nuclein", the scientific community was skeptical that something so seemingly simple compared with complex proteins could be the elusive inherited material (Miescher, 1869). Many years after this discovery, microscopy and staining techniques had advanced dramatically, allowing for the visualisation of the nuclear inherited material. Due to the stain and shape of the material, in 1888 the name "chromosome" was used to describe the unique structures of nuclein formed in nuclei (in Greek: chromos for colour, soma for body, Rowley (1991)). Alongside this, the five bases of nucleic acid (adenine (A), cytosine (C), guanine (G), thymine (T), and uracil (U)) were discovered (Levene and Bass, 1931), Mendelian genetics arose (Mendel, 1865), and finally, the double-helix structure of deoxyribonucleic acid (DNA) was proposed (Watson et al., 1953).

Watson and Crick elucidated the DNA helix structure, famously stating:

"It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."

This anti-parallel pairing of phosphate diester groups joining β-D-deoxyribofuranose, with hydrogen bonded bases in the centre of the helix was revolutionary. In the canonical double-helix structure of B-DNA described by Watson and Crick, adenine is paired with thymine, and guanine is paired with cytosine.

However, more than 60 years later, we are only now finding out about the plethora

of non-canonical nucleic acid structures formed in the cell.

1.1 Non-canonical nucleic acid structures

DNA sequences, specifically those of repetitive nature, can adopt a vast number of alternative structures that do not conform to typical B-DNA. These include guanine-quadruplexes (G4), intercalated (i)-motifs, triplexes, hairpins, Holliday junctions, and DNA:RNA hybrids, where nucleotides may not pair in a typical Watson-Crick manner (Wang and Vasquez, 2014). Greater than half of the human genome is comprised of repetitive DNA sequences capable of forming these non-B-DNA structures, as well as RNA transcribed from these repetitive regions (Consortium et al., 2001). DNA is most likely to form these structures during replication or transcription, where DNA strands are no longer confined in the duplex. Due to its single-stranded nature, RNA is particularly inclined to form non-canonical nucleic acid structures, which have roles in regulation as long non-coding RNA (lncRNA).

1.1.1 G-quadruplexes

A G4 is formed when four guanine nucleobases hydrogen bond together to form a tetrad. These tetrads can then stack on top of one another forming a tetraplex structure. G4s and other non-canonical structures form through Hoogsteen base-pairing, which is the bonding of the atoms in bases that do not typically form hydrogen bonds in Watson-Crick base-pairing. Given the oxygen-rich centre of the G-tetrad, monovalent cations such as K^+ are required to stabilise G4s (Burge et al., 2006).

In vitro, G4s have been shown to be more thermodynamically stable, and slower to unfold than their DNA or RNA hairpin counterparts (Lane et al., 2008). DNA and RNA G4s have numerous possible conformations, the sequence of which determine the final structure. The strands containing the guanines that contribute to the structure can be from the same molecule (intramolecular), or from many (intermolecular). The strands can be going in the same direction (parallel conformation, **Figure 1.1a**), or one or more can go in the opposite direction (anti-parallel conformation, **Figure 1.1b**). Interestingly however, RNA molecules have

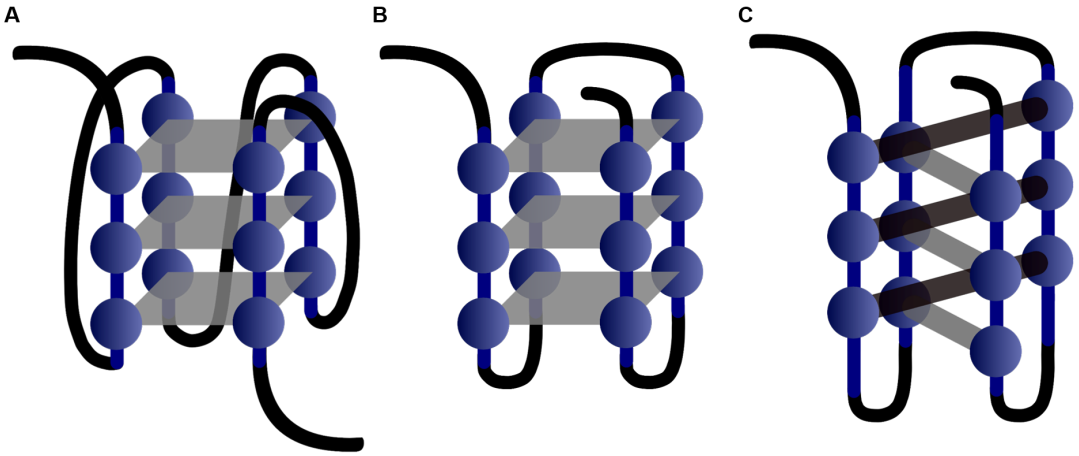

Figure 1.1: Non-canonical nucleic acid formations. a) A representation of a unimolecular parallel G4 structure, with the circles representing guanine nucleotides. b) A representation of a unimolecular anti-parallel G4 structure, with the circles representing guanine nucleotides. c) A representation of a unimolecular i-motif structure, with circles representing cytosine nucleotides (Kaushik et al., 2016).

been known to exclusively form parallel G4s (Fay et al., 2017), the reason for this is not yet fully elucidated.

The way in which the G4 tetrads stack, affecting parallel or anti-parallel topology, is dependent on the glycosidic bond angles of the guanine bases involved. The glycosidic bond connecting the 1' carbon on the deoxyribose to the nitrogenous base can be rotated to be in either a *syn* conformation, where the six member ring of the purine is adjacent to the oxygen in the deoxyribose sugar (unfavourable), or in an *anti* conformation, where the 6 member ring of the purine is positioned away from the sugar. When the glycosidic bonds are all in the *anti* conformation, this leads to parallel topology. When glycosidic bond angles alternate between *syn* and *anti* positions, this leads to anti-parallel conformations (Masiero et al., 2010). In addition to these variations involving G-tracts, nucleotides not participating in G-tetrad formation also contribute to the structure. The length of the loops outside of the G4 play a part in determining parallel or anti-parallel conformation as well. The shorter the loops, the more thermodynamically stable the parallel conformation becomes (Hazel et al., 2004). Loops are either lateral, where adjacent strands are joined on either the top or bottom of the G4 (**Figure 1.1b**), diagonal, where opposite strands join diagonally on either the top or bottom of the G4, or propeller, where strands are joined from opposite ends of

the G4 (Ogloblina et al., 2015), **Figure 1.1a**). Together, these structural intricacies provide unique qualities to each G4 molecule, allowing potential for highly targeted ligand binding.

Since G4s were discovered in 1962 (Gellert et al., 1962), many researchers have questioned the biological relevance of such structures, and deemed them an *in vitro* artefact. Possible G4 forming DNA sequences have been found throughout the human genome in promoters (including those of oncogenes, Siddiqui-Jain et al. (2002)), 5' and 3' untranslated regions, splice sites (Huppert and Balasubramanian, 2005), and in their largest abundance in repetitive sequences such as telomeres (Wang and Patel, 1993), suggesting G4s may have novel roles in genome regulation. High-resolution sequencing-based methods have found over 700,000 DNA G4 structures in the human genome (Chambers et al., 2015). These G4-forming sequences can be extracted, and depending on their length, assessed by dimethyl sulfate (DMS) footprinting, in order to isolate G4 structures unable to be degraded (Sun and Hurley, 2010), or circular dichroism (CD) spectroscopy in order to identify G4 structure by a characteristic CD profile (Randazzo et al., 2012). Another piece of supporting evidence is the fact that a number of human proteins including nucleases, helicases, and resolvases have been shown to bind G4s (Brázda et al., 2014). Further to this, antibodies raised against G4 structures have been able to detect possible G4 formation in immunofluorescence assays (Biffi et al., 2013).

As stated by Aaron Klug following the discovery of G4s, **"If G-quadruplexes form so readily *in vitro*, Nature will have found a way of using them *in vivo*".**

1.1.2 i-Motifs

An i-motif is a type of non-canonical nucleic acid structure whereby cytosine nucleobases can pair with one another, forming an intercalated, anti-parallel, tetrameric structure (**Figure 1.1c**). Cytosine base-pairs must be hemi-protonated ($C{:}C^+$) in order to form Hoogsteen hydrogen bonds between N3 of each pyrimidine (Gehring et al., 1993). Therefore, i-motifs were commonly thought to only be able to form under acidic conditions. However, there is evidence that they are capable of forming at neutral pH and at room temperature, and have also been shown to form in the nuclei of human cells by indirect immunofluorescence

staining using an i-motif-recognising antibody (Zeraati et al., 2018).

Of great interest, is the notion that when a G4 forms from duplex DNA, the opposite cytosine-rich strand is in a favourable position to form an i-motif at the same time. This raises the possibility that proteins binding to each of these non-canonical structures may work together in their genomic function.

1.2 Chromatin

Chromosomes are formed through the intricate folding of DNA complexed with nuclear proteins, which must be organised and protected within the three dimensional space of the nucleus. The way in which this occurs is though the formation of chromatin: the nucleoprotein complex. Two copies of each core histone, H2A, H2B, H3, and H4, together form a histone octamer, which is referred to as the nucleosome (Luger et al., 1997). The DNA double-helix then wraps around a nucleosome, approximately 1.65 times with 146 bp of DNA (Davey et al., 2002), and is referred to as the chromatosome. An array of these structures are then referred to colloquially as "beads on a string" due to the large globular protein and linker DNA between the nucleosome particles. The chromatosomes then fold to form a 30 nm fibre secured by linker histone H1, which is then further organised into distinct loops and domains in interphase (Dorigo et al., 2004; Olins and Olins, 1974).

This microscopically defined higher order structure is then partitioned into two distinct types of chromatin: euchromatin and heterochromatin. Euchromatin is transcriptionally active, whereas heterochromatin is more structurally compact and transcriptionally repressed, organising the genome into functional regions for appropriate gene expression patterns, as well as genomic integrity (Schneider and Grosschedl, 2007).

Active and inactive chromatin is characterised by the post-translational modifications (PTMs) present on the histones in each region. Each histone has many possible PTMs which alter the compaction and therefore gene expression in the area in which the modification is deposited. Chromatin modification through PTM is done through a change in the charge of the histone due to the addition of a PTM, and/or a protein binding to the PTM to then alter chromatin structure. Many studies have focused on the paradigm of the Histone code, the bipartite

PTMs on lysine 9 of H3 (H3K9), and their effects on chromatin structure and expression. H3K9 acetylation (H3K9ac) has been shown to highly correlate with active promoters, as well as mono-methylation of H3K9 (H3K9me1). However, di- and tri-methylation of H3K9 (H3K9me2/3) are strongly associated with silenced genes (Bowman and Poirier, 2015).

1.2.1 Heterochromatin

Heterochromatin formation is a fundamental architectural component of eukaryotic chromosomes, constitutively maintained around the centromeres and telomeres. Heterochromatin is further partitioned into two sub-types: facultative and constitutive heterochromatin (Brown, 1966).

Facultative heterochromatin, associated with H3K27me3, is a functional, reversible form of heterochromatin whereby genes may be silenced in a temporal, or cell specific manner (Jamieson et al., 2016). Constitutive heterochromatin, associated with H3K9me3, is a state of permanent compaction throughout the cell cycle, present at the same loci throughout cell types, and formed in repetitive regions of the genome, including the centromeres and telomeres (Becker et al., 2016).

The pivotal protein enriched in heterochromatin that binds the aforementioned H3K9me3 PTM in order to form and maintain constitutive heterochromatin is none other than Heterochromatin Protein 1α (HP1α) (Wongtawan et al., 2011).

1.2.2 Heterochromatin Protein 1α

Referred to as a hallmark of heterochromatin, the importance of HP1 has been suggested by the conservation of sequence and function through evolutionary time, present in almost all studied eukaryotic organisms from fission yeast, to plants, and to animals (Wang et al., 2000). As a non-histone chromosomal protein, HP1 establishes and maintains higher order heterochromatin structure.

The HP1 family of proteins have a distinct tripartite structure. This firstly consists of an N-terminal chromodomain. The chromodomain's primary function

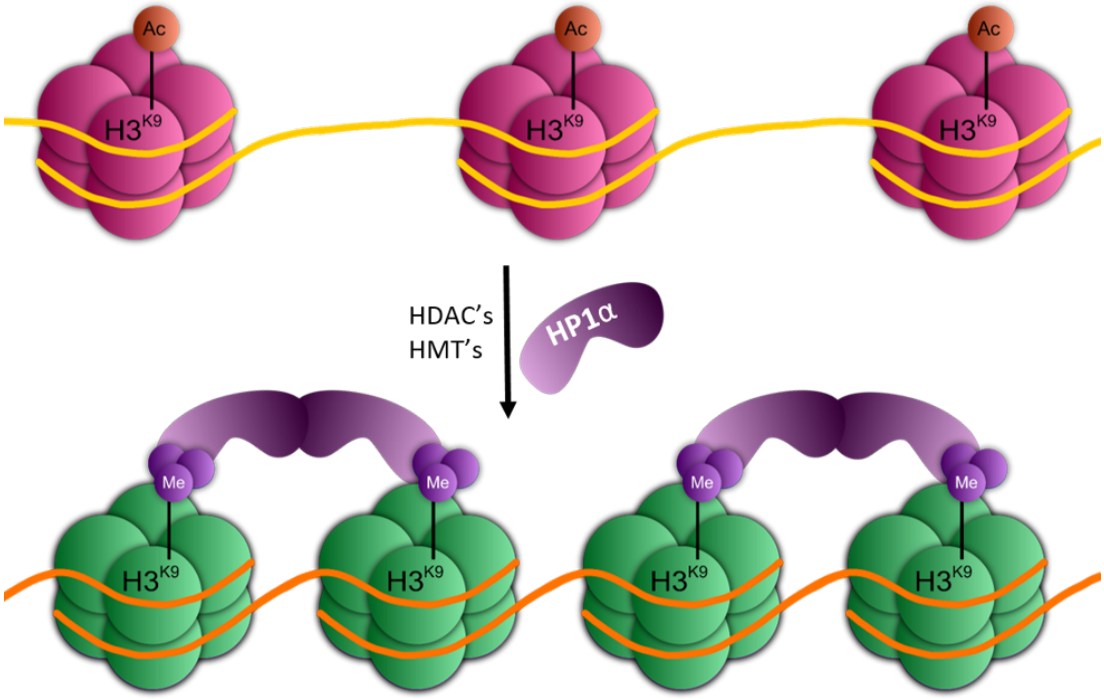

Figure 1.2: Simple model of heterochromatin formation. Nucleosomes bearing H3K9ac are deacetylated by histone deacetylases (HDACs), and subsequently methylated by histone methyltransferases (HMTs) to make H3K9me3. HP1α can then bind H3K9me3 through its chromodomain, and dimerise through its CSD to form heterochromatin (Kwon and Workman, 2011).

is to bind H3K9me2/3 in order to bind HP1 to histones and therefore, chromatin (Bannister et al., 2001). Secondly, is the C-terminal chromoshadow domain (CSD). The CSD is an interface for not only dimerisation between HP1 proteins (**Figure 1.2**), but also as a binding site allowing for recruitment of nuclear chromatin modifying factors which contain the PxVxL motif (Smothers and Henikoff, 2000). Finally, connecting the chromodomain and CSD is the hinge of HP1 (Meehan et al., 2003).

In addition to this now well characterised tripartite structure, are the N- and C-terminal extensions (NTE and CTE), before the chromodomain begins, and following the CSD. Little is known about potential functionality of these areas of HP1, apart from their roles in supporting chromodomain and CSD functionality (Canzio et al., 2014).

In mammals, three HP1 paralogs have roles in chromatin maintenance activities. The three HP1 proteins are HP1α, HP1β, and HP1γ, encoded by the genes CBX5, CBX1, and CBX3 respectively, each contributing to the structure of chromatin, as well as gene expression patterns. HP1α and HP1β localise to constitutive heterochromatin, whereas HP1γ is distributed both in heterochromatin and euchromatin (Minc et al., 1999, 2000). Given the sequence similarities between the HP1 paralogs, the molecular basis of their distinct localisation and functionality remains generally unknown.

HP1α is the primary HP1 paralog associated with constitutive heterochromatin, therefore is a focus when studying heterochromatin formation. The paradigm of the Histone code is a widely accepted model which proposes that the transcriptional activity of chromatin domains is in part regulated by the PTMs present on histone tails (Jenuwein and Allis, 2001). The multivalency of HP1α binding to H3K9me3 through its chromodomain, and its dimerisation and binding of other proteins through its CSD has been shown to accelerate chromatin retention of silent marks (Kilic et al., 2015). This emphasises the importance of the chromodomain and CSD of HP1α in its binding to both chromatin and other effectors on the chromatin to mediate interactions and heterochromatin formation. Contrary to the focus on the chromodomain's interaction with H3K9me3, this bond is not high affinity, with a dissociation constant of 35 μM, measured by fluorescence anisotropy (Hiragami-Hamada et al., 2011). Due to the weak interaction between HP1α and H3K9me3, further recruitment mechanisms or stabilising factors are required in order to maintain HP1α presence and therefore constitutive

heterochromatin, at pericentric and telomeric regions.

In HP1α, between the chromodomain and the CSD is the hinge region. This region was long ignored, as it was considered an unstructured hinge, just linking the two conserved domains. Finally, in 2002, a group postulated that the chromodomain was not solely responsible for HP1α targeting to heterochromatin, given that H3K9me3 is present throughout the genome, but HP1α is primarily localised to centromeres and telomeres. Upon treating cells with RNase, it was found that HP1α localisation was lost. Therefore HP1 may be able to recognise, and be recruited by, an RNA component (Maison et al., 2002). Muchardt et al. (2002) were able to show that HP1α is actually an RNA-binding protein, in which the hinge binds RNA and contributes to the localisation of HP1α. Further to this, three particular consecutive lysine residues (K104, K105 and K106) in the hinge domain have now been shown to be responsible for the RNA binding ability of HP1α.

Constitutive heterochromatin is maintained around centromeres, the central region of the chromosome where sister chromatids are attached, therefore condensation of pericentric regions is important to maintain faithful segregation of chromosomes during mitosis (David et al., 2003). As one would expect, HP1α is enriched in pericentric heterochromatin. Upon investigation, it was found that HP1α, when sumoylated on its hinge, interacts with lncRNA transcribed from major satellite repeats within pericentric and centric heterochromatin. Researchers proposed that this lncRNA was required to target HP1α specifically to pericentric heterochromatin, and this sumoylation of the hinge domain allowed for the chromodomain to then bind H3K9me3, forming heterochromatin (Maison et al., 2011).

Given the finding that a repetitive lncRNA transcribed from pericentric regions is required for heterochromatin formation through an interaction with HP1α around centromeres, a similar mechanism may be present at the telomeres.

1.3 Telomeres

Telomeres are the protective buffering region at linear chromosome ends (Olovnikov, 1973), preventing double-stranded DNA (dsDNA) from being degraded or recognised as double-strand breaks. The highly conserved telomeric DNA sequence in

humans consists of repeats of the sequence (TTAGGG)n (Allshire et al., 1989). This sequence is highly conserved among vertebrates, sharing a common ancestor over 400 million years ago (Meyne et al., 1989).

Through each round of cell division, a small, approximately 50-100 bp (Harley et al., 1990) section of DNA is lost during DNA replication due to the removal of the RNA primer of the last Okazaki fragment on the lagging strand (Okazaki et al., 1968), and is referred to as the end replication problem. After many rounds of DNA replication, the telomeres become shorter, and this is thought to be not only a marker for the age of an organism and it's Hayflick limit (Hayflick and Moorhead, 1961), but also contribute to old-age related diseases like cancer (Panossian et al., 2003). Once the telomeric sequence has been greatly reduced (in cultured human cells, after 50-70 cell divisions (Hayflick and Moorhead, 1961)) the cell will either elicit a DNA damage response (DDR) resulting in random recombination, or will cease to divide and place itself into a senescent state to reduce any further DNA damage (di Fagagna et al., 2003).

The function of telomeres is extremely crucial for longevity, therefore the cell has multiple strategies to prevent their decay. The highly conserved, repetitive, telomeric DNA sequence has a single-stranded 3' overhang, allowing for unique structural formations at the chromosome ends due to strand invasion. The G-rich 3' overhang is able to bind with the C-rich strand of previously duplex telomeric DNA to form a telomere loop (T-loop) structure. The original G-rich sequence may then form a displacement loop (D-loop) (**Figure 1.3**) and be replaced by the 3' overhang sequence (Griffith et al., 1999). These loop structures hide the loose telomere ends, thereby preventing them from being recognised by the cell as double-strand breaks (Longhese, 2008).

Another protective layer in addition to loop structures on the ends of chromosomes is the shelterin complex, bound to telomeric DNA to prevent telomere degradation and recognition as double-strand breaks (De Lange, 2005). Shelterin is a multimeric complex, consisting of six subunits. Together these subunits recognise and bind the TTAGGG repeat. Telomere Repeat Factor 1 and 2 (TRF1 and TRF2) recognise and bind the telomeric repeat in duplex DNA and then recruit Protection Of Telomeres 1 (POT1), TRF1 interacting nuclear protein 2 (TIN2), Repressor/Activator Protein 1 (RAP1), and Tripeptidyl Peptidase 1 (TPP1). POT1 recognises the telomeric repeat sequence in single stranded DNA, so facilitates T-loop formation, with the shelterin complex present at the junction (Diotti

Figure 1.3: Telomeres form T- and D-loops at chromosome ends via a 3' overhang (Griffith et al., 1999).

and Loayza, 2011).

In order to maintain the compact structure of telomeres, they are enriched in RNA, H3K9me3, and also HP1α. Drosophila HP1 was found to play a crucial role in telomere function, where it was found to control capping, elongation, and silencing of telomeres (Perrini et al., 2004). HP1γ is also localised to telomeres and binds the shelterin component TIN2 where it acts to maintain telomere-telomere cohesion (Canudas et al., 2011).

1.3.1 Telomeric repeat-containing RNA

As telomeres are a heterochromatic domain, this implies that they would be transcriptionally silent. However, in 2007-2008, two groups of researchers both happened upon the discovery of transcription at mammalian telomeres (Azzalin et al., 2007; Schoeftner and Blasco, 2008), with TElomeric Repeat-containing RNA (TERRA) identified as the transcript. This molecule, transcribed from sub-telomeric regions through to the telomeric tract varies in length (100 bp to 9000 bp, Azzalin et al. (2007); Feretzaki and Lingner (2017)), and little is known about its regulation and stability *in vivo*.

Deng et al. (2009) proposed that TERRA was required for the structural stability of telomeres, as well as heterochromatin formation. Shown with RNA affinity purification, and subsequent liquid chromatography-tandem mass spectrometry (MS) and western blotting, the TERRA sequence bound a myriad of RNA binding proteins, including TRF1 and TRF2. Interactions with shelterin components

were confirmed *in vivo* with RNA immunoprecipitation (RNA-IP) assays. The researchers also found using RNA-IP that TERRA was able to bind HP1α and modified histone H3, which suggests a role of TERRA in regulating heterochromatin formation.

Due to the G-rich sequence of telomeric repeats, the transcript, TERRA, is also G-rich, allowing for the formation of G4s. TERRA sequences of just 10 nt long up to 96 nt have been shown *in vitro* to form G4s. TERRA G4s form with three tetrads, due to the repeat containing three consecutive guanines, which stack together, and as TERRA is an RNA molecule, exclusively form parallel G4 topology.

TERRA has been shown to interact with proteins such as TRF2 (Deng et al., 2009), involved in telomere protection in the Shelterin complex, and also with RNA helicase associated with AU-rich element (RHAU) (Meier et al., 2013), an RNA helicase, which has been established as a major cellular quadruplex resolvase (Creacy et al., 2008; Vaughn et al., 2005). This RHAU interaction with TERRA shows that the G4 structure of the telomeric transcript is actually recognised and acted upon by a cellular protein, showing how a G4 structure can be utilised in the cell.

TERRA expression has been shown to be regulated by telomere length and presence of H3K9me3 (Arnoult et al., 2012). This alludes to a telomeric feedback mechanism, whereby long telomeres repress TERRA expression, but short telomeres allow transcription of TERRA, which then brings about H3K9me3 PTM and subsequently HP1α, condensation, and therefore reduced expression and compact telomeres.

1.3.2 Telomere elongation

To avoid the DNA damage response elicited upon telomere degradation, *in special circumstances* the enzyme telomerase is able to elongate the telomeric DNA (Blackburn et al., 1989). Telomerase (also known as terminal transferase) is an RNA dependent DNA polymerase enzyme, which has a guide RNA molecule complementary to the repeated sequence in telomeres. This directs the enzyme to reverse transcribe DNA from this RNA sequence to extend short telomeres (Lingner et al., 1997). Telomerase is not present in most cells, and in humans is

only expressed in embryonic cells, activated lymphocytes, male germ line cells, and adult stem cell populations (Cong et al., 2002). Telomerase has been found to be expressed in cancer cells to evade growth-inhibitory factors and apoptosis due to short telomeres (Wright et al., 1996). Cancer cells activate the expression of telomerase to ensure infinite proliferative ability, therefore the strict regulation of telomerase is of the utmost importance.

An alternate way cells increase telomere length is through alternative lengthening of telomeres (ALT) (Bryan et al., 1997), where TERRA is involved in lengthening of telomeres. The method for ALT has not yet been fully elucidated, particularly the elongation step, however it is widely agreed that recombination plays a large part. With two telomeres of similar length, one may form a loop structure to allow for unequal recombination, resulting in one short and one long telomere. Following this, one method proposed by Cesare and Reddel (2010) is the non-random separation of telomere sister chromatids. After recombination, all the longer telomere sister chromatids are segregated from the shorter ones. This results in one daughter cell with greater proliferative ability due to longer telomeres, and the other daughter cell with diminished proliferative capacity and a decreased chance of survival, which may result in evolved cancer cell populations. Cesare and Reddel also put forward a second possible mechanism, which may not be mutually exclusive of the first. They proposed that after recombination, telomeres could be synthesised from existing telomeric DNA sequences (rather than the RNA template used by telomerase), whether it be from adjacent chromosomes or the same telomere via T-loop formation.

The ALT mechanism of telomere elongation is utilised in very aggressive tumours, such as glioblastoma multiform, the most prevalent form of brain tumour in adults, which typically has a very poor prognosis (Hakin-Smith et al., 2003). High TERRA levels are a hallmark of ALT-positive cancer cells; therefore, it is likely that this lncRNA has a role in the ALT process. It has been proposed that TERRA facilitates the recombination process in ALT by forming RNA:DNA hybrids with the C-rich strand of the telomere to allow for DNA synthesis (Arora and Azzalin, 2015). In a recent analysis of 31 cancer types and their telomere lengths, in around 15% of cases, elongated telomeres were found along with increased TERRA (Barthel et al., 2017). Other studies have shown a correlation between α-thalassemia/mental-retardation-syndrome-X-linked (ATRX), involved in chromatin remodelling, and TERRA levels (Flynn et al., 2015), where ATRX

knockdown results in loss of cell cycle dependent regulation of TERRA and therefore increased TERRA and recombination mediated ALT. HP1α has also been shown to mediate the formation of PML (promyelocytic leukaemia) bodies in the nucleus, where ALT takes place in cancer cells (Jiang et al., 2011).

Since the discovery of TERRA, evidence linking it to telomere maintenance in cancer is substantial, therefore this lncRNA is clearly paramount to the regulation of telomeres, particularly immortalisation in cancer cells.

1.4 HP1α-TERRA

TERRA and HP1α have been shown to interact (Deng et al., 2009), and TERRA functions in heterochromatin formation at telomeres. The hinge region of HP1α, shown to have RNA binding capability, may well be responsible for the interaction between HP1α and TERRA, allowing for the targeting of HP1α to telomeres. The ability of TERRA to form G4s may provide the specificity required in order to target HP1α specifically to telomeres, setting it aside from the RNA recruitment mechanism at the centromeres dependent on sumoylation.

This TERRA association with chromatin modifier HP1α has implications associated with critically short telomeres, as well as elongation of telomeres in cancer. TERRA is transcribed from the very region that it may be responsible for condensing, suggesting a negative feedback mechanism (Arnoult et al., 2012), where long telomeres produce long TERRA molecules which repress their own transcription by increased heterochromatin formation. HP1α interacts with TERRA molecules, thus providing a mechanism for this feedback loop. A longer TERRA has more G4s, therefore more HP1α binding sites, so recruitment of more HP1α, thus condensing the telomeric heterochromatin and repressing transcription.

1.5 Objectives

HP1α, a pivotal chromatin regulator, and TERRA, a telomeric transcript involved in many tumourigenic processes, have been shown to interact. Given the importance of heterochromatin formation and therefore localisation of HP1α to telomeres, a function and mechanism for the interactions between HP1α and TERRA needs to be explored.

A model where TERRA and HP1α interact to form and maintain telomeres through protein recognition of a specific G4 is pivotal to understanding telomere regulation, and also to genome organisation as a whole, showing the relevance of non-canonical nucleic structures in epigenetic regulation.

This work seeks to understand how HP1α recognises and interacts with G4s, specifically TERRA, necessary for the targeting of HP1α to a region of constitutive heterochromatin.

Hypothesis: HP1α binding to TERRA is structure dependant, whereby HP1α recognises the G4 structure formed by TERRA and binds via the unstructured hinge domain.

The aims of this investigation are as follows:

- does HP1α bind TERRA?
- whether the lysine-rich hinge domain is involved in this interaction,
- if HP1α binds endogenous non-canonical nucleic acid structures,
- and the structural intricacies of the HP1α-TERRA interaction.

Chapter 2

Materials and Methods

2.1 Interaction Buffers

The buffer used in experiments throughout was 1x 100 mM KCl, 50 mM NaCl, and 20 mM NaH_2PO_4, at pH 8, and will be referred to as Interaction Buffer (IB). This buffer was also prepared at pH 7, and will be referred to as IB7.

A stock of 10x IB or IB7 was prepared (1 M KCl, 500 mM NaCl, 200 mM NaH_2PO_4) and filter sterilised using a 0.2 μm filter. To prepare 1x solution, 10x stock was diluted with milli-Q H_2O 10-fold.

2.2 Oligonucleotides

A 96 bp telomeric repeat-containing RNA (TERRA) 96 suspended in 100 mM NaCl and 10 mM NaH_2PO_4 was a gift from Carlos Gonzalez (Instituto de Qumica Fsica Rocasolano, Madrid, Spain), and purified according to McKenna et al. (2007). All other oligonucleotides were obtained from Integrated DNA Technologies, purified by standard desalting by the manufacturer. Sequences are listed in **Table 2.1**.

Table 2.1: Sequence and length of oligonucleotides

Name	Sequence	Length (nt)
TERRA 45 (RNA)	rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG	45
TERRA 96 (RNA)	rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA	96
d-TERRA 45	GGG TTA GGG TTA GGG TTA GGG TTA GGG TTA GGG TTA GGG TTA GGG	45
Thy45	TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT	45
Oligo A	TGG GTG GGT GGG TGG GTG GGT GGG TGG GTG GGT	33
Oligo B	TGG GTT GGG TTG GGT TGG GTT GGG TTG GGT TGG GTT GGG T	40
Oligo C	TGG GTT AGG GTG GGT TAG GGT	21
Oligo D	TGG GTG GGT TAG GGT GGG T	19
Oligo E	TGG GTG GGT TAG GGT TAG GGT	21
Oligo F	TGG GTT AGG GTT AGG GTG GGT	21
Oligo G	TGG GTT AGG GTT AGG GTT AGG GT	23
Oligo H	TGG GTT AGG GTG GGT GGG T	19
Oligo I	TGG GTG GGT GGG TTA GGG T	19
Oligo 2C	TGG GTT AGG GTG GGT TAG GGT GGG TTA GGG TGG GTT AGG GT	41
Oligo 2D	TGG GTG GGT TAG GGT GGG TGG GTG GGT TAG GGT GGG T	37
Oligo 2E	TGG GTG GGT TAG GGT TAG GGT GGG TGG GTT AGG GTT AGG GT	41
Oligo 2F	TGG GTT AGG GTT AGG GTG GGT GGG TTA GGG TTA GGG TGG GT	41
Oligo 2G	TGG GTT AGG GTT AGG GTT AGG GTG GGT TAG GGT TAG GGT TAG GGT	45
bcl2	AGG GGC GGG CGC GGG AGG AAG GGG GCG GGA GCG GGG CTG	39
cmyc_2	GGG TGG GTA GGG TGG G	16
ckit2	GGG CGG GCG CCA GGG AGG GG	20
KRAS32_2	AGG GCG GTG TGG GAA GAG GGA AGA GGG GGA GG	32
SRC2	GGG GCA GCT GGG TCG CTC GGG GAA CGG GG	29
SRC3	GGG AGG GAG GGC TGG GGG	18
SRC14	GGG TGG GCT GAC GGG GCA GAA TGG GAG GG	29
SRC16	GGG CTG CTG GGA CTG GTT GGG AGG GCA CTC TGG GG	35
SRC19	GGG TCC CCT GGG CCT GGG CGG GAG CGG GGA GGG G	34
Carlos LL3rep	TCG TTC CGT TTT TCG TTC CGT TTT TCG TTC CGT TTT TCG TTC CGT	45
DAP i-motif	CCC CCG CCC CCG CCC CCG CCC CCG CCC CC	29
HIF-1alfa i-motif	CGC GCT CCC GCC CCC TCT CCC CTC CCC GCG C	31
PDGF-A	CCG CGC CCC TCC CCC GCC CCC GCC CCC GCC CCC CCC CCC CC	41
JAZF1 i-motif	CCC CCC CCG CCC CCG CCC CCG CCC TCC CCC C	31
BCL-2 i-motif	CAG CCC CGC TCC CGC CCC CTT CCT CCC GCG CCC GCC CCT	39
C-telomere 45	CCC TAA CCC TAA CCC TAA CCC TAA CCC TAA CCC TAA CCC TAA CCC	45

Oligonucleotides (apart from TERRA 96) were folded in 1x IB or IB7 (for the formation of i-motifs and dsDNA) by heating at 90 °C for 5 minutes, then cooled gradually to room temperature.

2.3 Biolayer Interferometry

Biolayer Interferometry (BLI) was performed on the BLItz system (ForteBio), using Ni-NTA biosensor tips. Tips were hydrated for at least 10 minutes prior to each experiment in 250 μL of IB (or IB7 if i-motifs or dsDNA were being tested). The Advanced Kinetics mode of the BLItz Pro 1.2 software was used to monitor the binding of each species at the various steps. Shaking of the tip was set to 2200 rpm (default for kinetics experiments).

The initial baseline step (60 s) was performed in 250 μL of 1xIB within a black opaque 0.5 mL Eppendorf tube, then 4 μL of his-tagged protein (in a drop holder) was incubated with the Ni-NTA tip for 5 minutes. His- Heterochromatin Protein 1α (HP1α) was used at 100 μg/mL, other paralogs of his-tagged HP1 were matched relative to this level of binding (3.5 nm, HP1β at 80 μg/mL, HP1γ at 70 μg/mL, and HP1α-3KA at 140 μg/mL). The tip was then washed in 250 μL of 1xIB (60 s, in a black opaque 0.5 mL Eppendorf tube), to remove any protein not bound to the tip with high affinity. The tip was then incubated with 4 μL of 500 nM folded oligonucleotide (for 5 minutes in drop holder) for a typical binding experiment. For titrations, the concentration of oligonucleotide was varied. Finally, dissociation was measured by washing in 250 μL of 1xIB in a black 0.5 mL Eppendorf tube for 5 minutes.

Reference experiments, where 1xIB was added in the place of oligonucleotide, were also performed. The reference for each protein was then subtracted from subsequent experiments in order to isolate the specific response from adding ligand, accounting for dissociation of protein from the tip. No oligonucleotide was shown to bind the Ni-NTA biosensor tip in the absence of his-tagged protein (data not shown).

BLItz Pro 1.2 software was used for curve fitting and K_D calculations. As artefacts can be produced when starting and stopping each experiment to change between drop holder and Eppendorf tube, the data was processed to begin each step where the previous one had left off. For example, the dissociation data was set to begin at exactly the amount of binding at the end of the association step. When titrations were performed to elucidate K_D, a global curve fit was used to take all binding curves and their concentrations into account.

2.4 Protein sequence alignment

Alignment was performed using the CLUSTAL Omega Multiple Sequence Alignment tool, version 1.2.4, with default parameters (Sievers et al., 2011).

2.5 Circular Dichroism

Circular Dichroism (CD) spectroscopy was performed on the Chirascan (Applied Photophysics Ltd) instrument using 5-10 μM RNA or DNA (in order to produce maxima above 10 CD mdeg), diluted in 1xIB. Measurements were performed on 250 μL of oligonucleotide inside of a quartz 1 mm cuvette.

Scans were gathered from 200 nm to 350 nm with 1 nm data pitch, on continuous scanning mode, with a response of 0.25 seconds and bandwidth 1 nm. 1xIB was scanned prior to each experiment, and subtracted from each scan. Three scans were made for each oligonucleotide, averaged, and then smoothed by averaging 10 neighbour points. Chirascan v.4.4.1 was used for operation of the Chirascan instrument, as well as data processing described.

2.6 Polyacrylamide Gel Electrophoresis

2.6.1 SDS-PAGE

For sodium dodecyl sulfate (SDS) Polyacrylamide Gel Electrophoresis (PAGE), a 12% acrylamide/bisacrylamide resolving gel was made containing 375 mM Tris HCl pH 8.8, 0.1% SDS, 0.1% ammonium persulfate (APS), and 0.1% tetramethylethylenediamine (TEMED). A stacking gel of 5% acrylamide/bisacrylamide, 125 mM Tris HCl pH 6.8, 0.1% SDS, 0.1% APS, and 0.1% TEMED was also made and set atop the resolving gel.

6x loading buffer (62.5 mM Tris pH 6.8, 1.67% SDS, 7.5% glycerol, 0.01% bromophenol blue, 35.75 mM 2-Mercaptoethanol) was added to a final concentration of 1x to each sample, then heated to 99 °C for 5 minutes prior to loading. Electrophoresis was performed with 12% resolving gel below 5% stacking gel in 1x

TGS buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3) at a constant 35 mAmp until the dye front reached the bottom of the gel.

Unless otherwise specified, the gel was stained with 0.1% Coomassie in 30% methanol and 10% acetic acid, washed with H_2O, and de-stained with 20% methanol and 10% acetic acid until sufficient de-staining had occurred. The gel was then imaged on a Gel-Doc instrument.

2.6.2 Native PAGE of Oligonucleotides

A 20% acrylamide/bisacrylamide native polyacrylamide gel was made containing 1x TBE buffer (45 mM Tris-borate, 1 mM EDTA) with addition of salts contained in the IB (100 mM KCl, 50 mM NaCl, 20 mM NaH_2PO_4), referred to as TBE+salts, 0.1% APS, 0.1% TEMED, and H_2O.

To 15 μL of 10 μM oligonucleotide, 3 μL of 0.0025% bromophenol blue in 10% glycerol and 1x IB was added, and then the total 18 μL of each mixture was loaded. The gel was electrophoresed at 4 °C in 0.5x TBE+salts at 5 Watts per gel until the dye front had travelled the majority of the way down the gel. To combat overheating and acidity of the buffer during the electrophoresis, the buffer was removed from the apparatus, mixed, and added back every hour. The gel was rinsed in H_2O, and then stained in 0.35% Stains-All in 50% formamide for 15 minutes in the dark. The gel was then washed in H_2O three times and de-stained in H_2O, no longer in the dark, until sufficient de-staining had occurred. The gel was then imaged on the Gel-Doc under white light.

2.6.3 Native PAGE of Proteins and Complexes

A native polyacrylamide gel was made the as the oligonucleotide native PAGE (**Section 2.6.2**), but with only 5% acrylamide/bisacrylamide for protein separation.

To 10 μL of 0.5 μg/μL protein (5 μg), 0.0025% bromophenol blue in 10% glycerol and 1x IB was added, and then loaded. The gel was performed as previously described in **Section 2.6.2**. The gel was rinsed in H_2O, fixed in fixation solution (40% ethanol, 10% acetic acid), washed again in H_2O, and stained with QC

Colloidal Coomassie Stain (Bio-Rad) overnight (approximately 18 hours). De-staining was then performed by soaking the gel in H_2O, and then imaged on the Gel-Doc.

2.7 Protein Cross-linking

Proteins were prepared at a range of concentrations, to a volume of 20 μL. Formaldehyde was added to a final concentration of 1%, and incubated with protein for 5 minutes at 37 °C. A non-cross-linked control was performed by adding 1xIB instead of formaldehyde, but otherwise treated the same as all other samples. The cross-linking properties of formaldehyde were then quenched by addition of tris HCl pH 8 to a final concentration of 100 mM. 6x loading buffer (62.5 mM tris pH 6.8, 1.67% SDS, 7.5% glycerol, 0.01% bromophenol blue, 35.75 mM 2-mercaptoethanol) was then added (to a final concentration of 1x) to denature protein and to aid in loading. Because high temperatures can reverse cross-links, denaturing with loading buffer was performed at 65 °C rather than 99 °C, because this has been shown to sufficiently denature protein while minimising cross-link reversal (Klockenbusch and Kast, 2010). Samples were then loaded and separated by SDS-PAGE (**Section 2.6.1**). The gel was then stained with QC Colloidal Coomassie Stain as described previously, and imaged using the Gel Doc.

2.7.1 Protein Size Estimation

To estimate the sizes of cross-linked products, the protein standard marker was first measured from the beginning of the resolving gel to each band. The retention factor (Rf) for each of these measurements was then calculated by dividing by the distance from the beginning of the resolving gel to the dye front. The log of the molecular weight of each of the bands was calculated and subsequently plotted against the Rf of each of the bands to produce a standard curve.

The gel this analysis was performed on had a marker on each side of the gel, therefore each of these was measured and plotted as above. Although similar, the right-hand-side marker was chosen to proceed with estimating molecular weights, due to the higher R-value of the linear trendline plotted. The 250 kDa, 150 kDa,

and 100 kDa bands did not fit the linear relationship, so were excluded from trendline equation calculations. However, this means that weights could only be measured up to 75 kDa accurately.

Further to this, the gel appeared to have warped during the electrophoresis process, where proteins in the centre of the gel were less mobile than those at the left and right edges, which affected protein size estimation.

HP1α protein bands were measured, divided by the dye front length to give Rf values for each band. Using the equation given by the linear trendline, x(Rf) was used to solve for y(log MW), and the inverse log of this value was calculated to give the molecular weight estimate of the protein.

2.8 Isothermal Titration Calorimetry

Prior to performing isothermal titration calorimetry (ITC), dialysis of protein or nucleic acid was performed to ensure accurate buffer matching, reducing the heat of dilution. 50 μL of HP1α or 50 μL of TERRA 45 was dialysed in a 50 μL dialysis button (Hampton Research) with a 500-1000 Da MWCO membrane cut to size secured with an O-ring. Dialysis was performed in 100 mL of buffer at 4 °C with stirring over 48 hours, with one buffer change after 24 hours.

The MicroCal iTC$_{200}$ isothermal titration calorimeter instrument was used, along with its associated software.

Water was held in the reference cell, and the cell temperature was set to 25 °C. The reference power was 3 μcal/second, with an initial delay of 60 seconds, and a stirring speed of 750 rpm. There was an initial injection of 0.5 μL for each experiment, and spacing between injections was 120 seconds with a 5 second filter period. RNA was held in the syringe, injected into the sample cell containing protein. Concentrations of protein and RNA, and number and volume of injections are discussed in **Section 4.3**.

Prior to each experiment, the heat of dilution was measured by an identical experiment with buffer in the sample cell in place of protein (reference).

Analysis was performed using MicroCal Analysis software, under the ITC$_{200}$ setting. Firstly, reference experiments were subtracted from their respective ITC

experiments. Anomalous data was removed, including the first 0.5 μL injection in all cases. A one-to-one curve was then fit after three iterations to give output data.

2.9 Small-Angle X-ray Scattering

Small-angle X-ray scattering (SAXS) data were measured at the Australian Synchrotron on the SAXS/WAXS beamline using a sheath flow environment. A wavelength of 1.0332 $\mathring{A}$ (12 keV radiation) and detector distance of 1575 mm was implemented.

Sample was degassed, and then 50 μL was injected and eluted at 0.2 mL/min. Data was collected using a co-flow setup to mitigate radiation damage, with size exclusion chromatography coupled with SAXS (SEC-SAXS), using an Agilent AdvanceBio SEC, 2.7 μm, 300 $\mathring{A}$, 7.8 300 mm, eluting with 1xIB with 5% glycerol.

Data was processed using CHROMIXS (Panjkovich and Svergun, 2017) to manually select the appropriate SEC peak, and subtract eluted buffer. Analysis was then performed using PRIMUS (Konarev et al., 2003), as part of the ATSAS data analysis package.

2.10 Protein Expression and Purification

Mus musculus HP1α, HP1α-3KA, HP1β, and HP1γ cDNA (human HP1α protein sequence has 98% similarity to mouse, HP1β and HP1γ have 100% protein sequence identity to mouse) were cloned into pPROEX HTb with BamHI and HindIII using standard cloning techniques.

2.10.1 Expression

2.10.1.1 Plasmids

Cloning of the protein amino acid sequences was done previously for this project, using standard cloning techniques. This plasmid was chosen to add a hexahistidine tag to the protein, to aid in purification and also for use with the Ni-NTA biosensors.

This plasmid contains an ampicillin resistance gene for selection purposes, and also the lac promoter to control induction with the addition of IPTG (**Figure A.1**).

2.10.1.2 Host

The strain used for expression was Rosetta 2(DE3) competent cells, which are a derivative of the BL21 strain of *Escherichia coli*. These cells are made to enhance the expression of eukaryotic proteins by containing codons rarely used in *E. coli*, by supplying tRNAs for 7 extra codons (AGA, AGG, AUA, CUA, GGA, CCC, and CGG). The tRNAs are present on a plasmid that also carries a chloramphenicol resistance gene, aiding in selection. Previous work was done that showed Rosetta 2 cells were best for expression of HP1 protein from the pPROEX HTb vector (done by Andy Hollings).

2.10.1.3 Growth

Firstly, an aliquot of Rosetta 2 cells were thawed on ice, 1 μL of 100 ng/μL of vector was added, and the mixture was incubated on ice for 30 minutes, at 42 °C for 45 seconds, and again on ice for 2 minutes. 500 μL of Super Optimal Broth was then added, and this was then incubated at 37 °C for 1 hour.

50 μL of this was then spread onto an LB agar plate with 34 μg/mL chloramphenicol for the Rosetta 2 plasmid containing tRNAs, and also 100 μg/mL ampicillin to select for cells that contain the pPROEX HTb plasmid and therefore had ampicillin resistance.

The agar plate was then incubated overnight at 37 °C. Following this, a single

transformed colony was transferred to 10 mL of LB containing 34 μg/mL chloramphenicol and 100 μg/mL carbenicillin in a 50 mL falcon tube. Carbenicillin was used instead of ampicillin because it is more resistant to the beta-lactamase enzymes present in *E. coli* that degrade these types of antibiotics, and therefore won't be degraded over long time periods such as overnight incubations. This 10 mL culture was then incubated again overnight at 37 °C, with shaking at 200 rpm. The following day, the tube was centrifuged at 3000 xg for 5 minutes. The supernatant was discarded, and the pellet resuspended in 10 mL of fresh LB.

2.10.1.4 Induction

10 mL of the resuspended pellet was used to innoculate each 1 L of media (LB + 100 μg/mL carbenicillin and 34 μg/mL chloramphenicol) in a 5 L flask. This 1 L of media and inoculant was incubated at 30 °C, shaking at 180 rpm, until the OD_{600} had reached 0.4. The culture was then chilled for 15 minutes on ice before being induced by addition of 0.4 mM IPTG, and then incubated overnight at 22 °C, shaking at 180 rpm.

2.10.1.5 Harvest

Overnight induced cultures were centrifuged at 3000 xg for 10 minutes at 4 °C, supernatant was removed, and cells were resuspended in wash buffer (10 mM Tris HCl pH 8, 100 mM NaCl). This was then centrifuged at 3000 xg for 10 minutes at 4 °C. Supernatant was again removed, and cells were resuspended in lysis buffer (20 mM NaH_2PO_4, 50 mM NaCl, complete EDTA free, 1 mM Tris HCl pH 8, 1 mg/mL lysozyme).

2.10.2 Purification

2.10.2.1 Lysis

Cells suspended in lysis buffer firstly went through two freeze-thaw cycles at -80 °C. 100 μg/mL DNase was then added before subjecting this solution to 4000 psi using a hydraulic press, twice.

Lysate was then centrifuged at 3000 xg for 10 minutes, and the supernatant removed and filtered through a 0.2 μm filter.

2.10.2.2 Affinity Chromatography

Affinity chromatography was performed with 10 mL of Ni-NTA resin in a column attached to the *Å*KTA Prime Plus (GE Healthcare).

Imidazole was added to the lysate to a final concentration of 20 mM. Lysate was loaded onto the column at a flow rate of 1 mL/min, and washed firstly with lysis buffer (20 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, pH 7.4), and then washed with wash buffer (20 mM sodium phosphate, 300 mM NaCl, 50 mM imidazole, pH 7.4). Elution was then performed with elution buffer (20 mM sodium phosphate, 300 mM NaCl, 300 mM imidazole, pH 7.4)

2.10.2.3 Size Exclusion Chromatography

Size Exclusion Chromatography (SEC) was performed with a Superdex 200 10/300 GL column attached to the *Å*KTA Explorer. Eluate from affinity chromatography was concentrated and desalted using Vivaspin 20 (5 kDa MWCO) ultrafiltration devices (GE Healthcare, UK).

250 μL of eluate was loaded into a 500 μL sample loop, and and injected onto the column at 0.3 mL/min with 1xIB. Fractions were then concentrated using Vivaspin 20 (5 kDa MWCO) ultrafiltration devices.

An SDS-PAGE of SEC fractions showing purity of HP1α is shown in **Figure A.2**. The protein achieving this purity was used in Native PAGE, ITC, and SAXS.

Chapter 3

Characterising the specificity of the HP1α-TERRA interaction

To mimic the *in vivo* structure of transcribed telomeric repeat-containing RNA (TERRA) molecules, sequences should contain a bare minimum of eight telomeric repeats, with longer molecules being more representative of the true structural arrangement (Collie et al., 2010). Therefore, to recapitulate the structure of the G-quadruplexes formed by longer *in vivo* TERRA molecules in terms of intramolecular interactions between multiple G-quadruplexes, a TERRA molecule, termed TERRA 96 due to its 16 repeats of the 6 nt $(UUAGGG)_n$ sequence, was transcribed, purified, and kindly gifted for this purpose by Carlos Gonzales (Instituto de Qumica Fsica Rocasolano, Madrid, Spain).

Biolayer Interferometry (BLI) was performed to measure the association of Heterochromatin Protein 1α (HP1α) with RNA and DNA molecules. This technique measures the differences in incident white light through two surfaces within a biosensor tip. One surface is designed to immobilise molecules, as it is coated with a bio-compatible matrix chosen for the molecule of choice. The other surface is used as a reference (**Figure 3.1a**). The difference measured between the sample and reference light is measured in nm and used as a measure of binding (**Figure 3.1b**).

To measure the greatest amount of binding, it is best if the smaller molecule is immobilised on the tip, in order to detect the addition of the larger molecule that

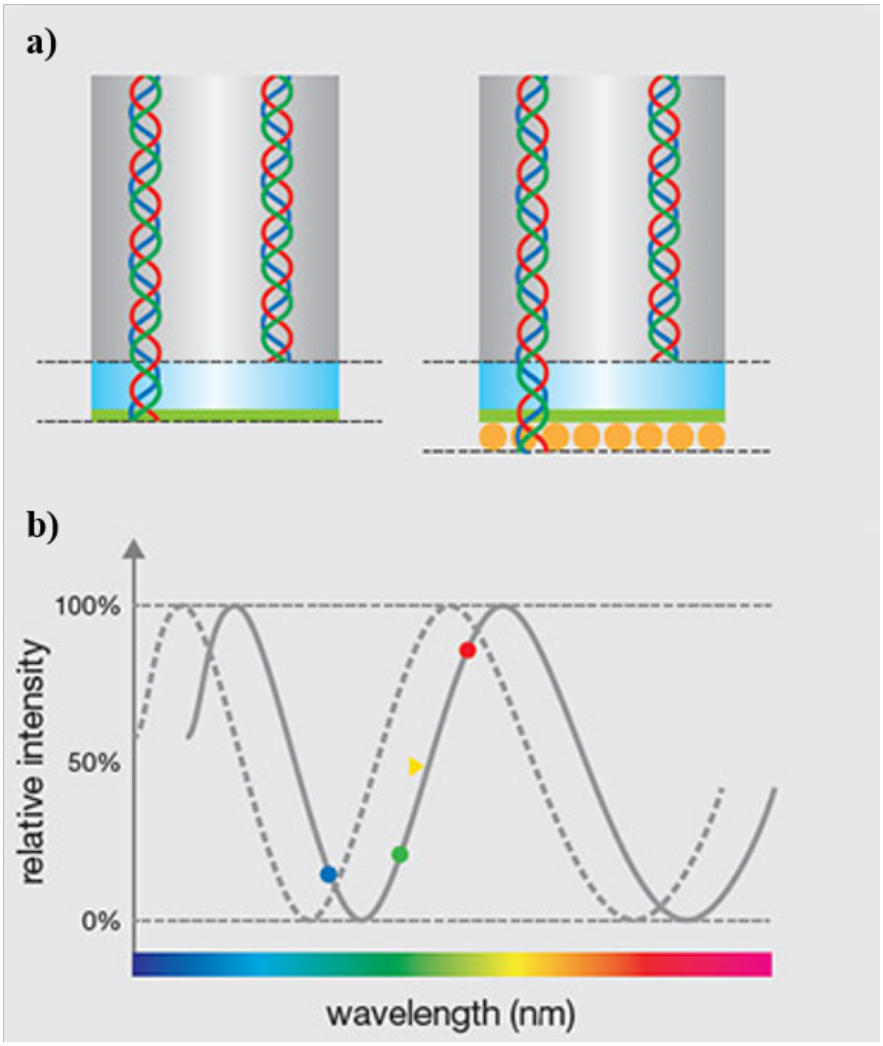

Figure 3.1: BLI technology, a) showing a representation on the inside of a tip in absence and in presence of sample molecules (orange), which increases the length that the light travels through, b) shift of the interferometric profile due to binding of sample, and therefore a change in the wavelength of light (Janzon, 2013).

it may interact with. This was originally attempted with biotinylated oligonucleotides immobilised on a streptavidin tip, which was then immersed in HP1α. However, it was found that HP1α was also able to bind to the biosensor tips. Therefore, it was decided to immobilise his-tagged HP1α onto Ni-NTA tips and measure the association of unlabelled oligonucleotides.

Circular Dichroism (CD) spectroscopy was used to assess the particular structures formed by the RNA and DNA molecules used in this study. This technique measures the difference in absorption of circularly polarised light by a molecule containing a chiral chromophore. The circularly polarised light can be rotated left or right at any given wavelength. The degree which circularly polarised light has been rotated to the right is subtracted from that which has been rotated to the left at each wavelength, and therefore certain structures have characteristic minima and maxima over a spectrum that pertain to its conformation.

In the case of G4s, guanine is a chromophore, that is linked to a chiral deoxyribose. The homopolar stacking of G-tetrads, which is evident in G4-DNAs and G4-RNAs of parallel topology with *syn* glycosidic bonds, causes positive CD signal at around 260 nm, and negative CD signal at around 240 nm. Any heteropolar stacking of G-tetrads caused by an anti-parallel topology causes a positive CD signal around 290 nm, and a negative CD signal around 260 nm. This allows for the characterisation of G4s that are either in a parallel conformation, where all strands are in the same orientation, or of an anti-parallel topology.

This chapter will explore, using the techniques explained above, the binding of HP1 protein to non-canonical nucleic acid structures by BLI, and conclusions will be drawn about the specificity of those interactions determined by the structure of each nucleic acid using CD spectroscopy.

3.1 HP1α binds TERRA through its lysine-rich hinge

Given that the HP1 proteins were purified using a hexahistidine tag, they are compatible with the Ni-NTA (NTA) biosensors available. As described in **Section 2.3**, his-tagged HP1α was immobilised on a Ni-NTA biosensor, and the tip was then immersed in 500 nM TERRA 96, and subsequently washed in 1xIB.

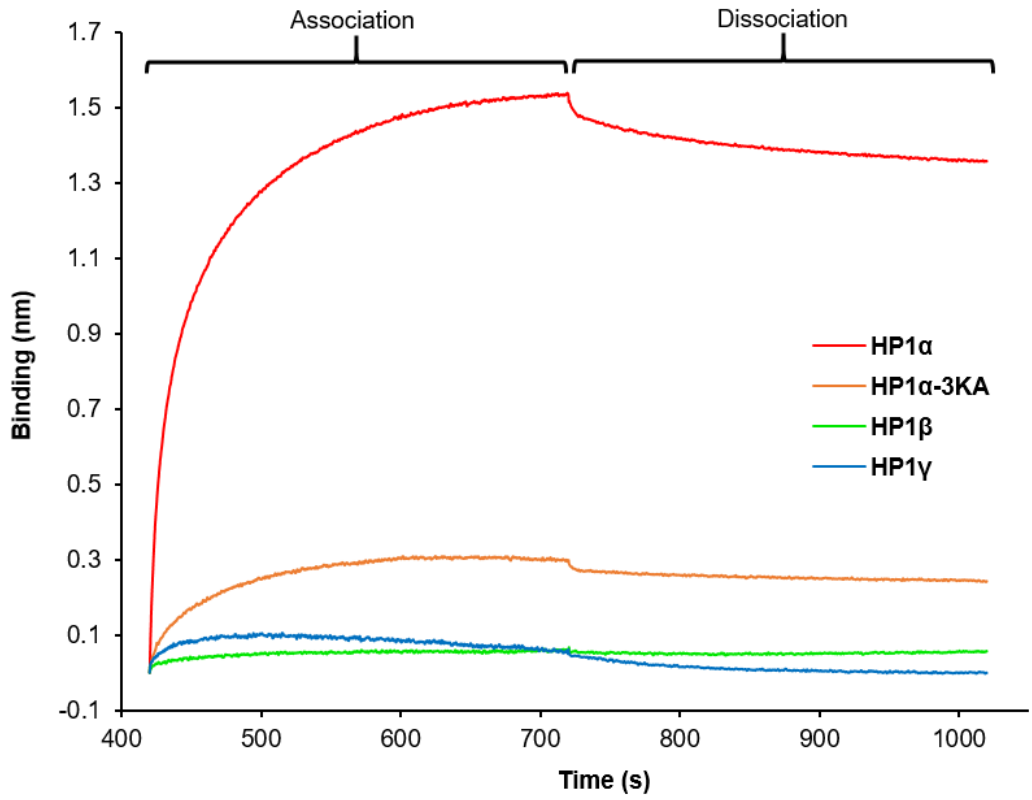

Figure 3.2: Binding of HP1 paralogs with TERRA 96. BLI performed with each HP1 paralog or mutant immobilised on a biosensor, with addition of TERRA 96, and subsequent washing in 1x IB to show the respective association and dissociation of TERRA 96 to the protein.

BLI showed a high response between HP1α and TERRA 96, indicating association of the RNA with the immobilised protein. (**Figure 3.2**). This confirmed the results shown by Deng et al. (2009), where HP1α binds to endogenous TERRA (using RNA ChIP and RNA IP). The RNA IP experiment also showed binding of HP1β with endogenous TERRA, therefore BLI was performed with all paralogs of his-tagged HP1 immobilised on the tip. TERRA bound neither of the other two paralogs of HP1 substantially. However, TERRA 96 bound to HP1γ with a slightly higher response than TERRA 96 binding to HP1β.

The hinge domain of HP1α is an RNA binding region. An RNA binding mutant, HP1α-3KA, with three lysine residues substituted for alanine (residues 104-106) in the hinge domain (**Figure 3.3**), has been demonstrated to lack RNA binding capability (Muchardt et al., 2002). Therefore, HP1α-3KA was immobilised on a biosensor tip, and tested for its interaction with TERRA 96.

BLI with HP1α-3KA immobilised, immersed in TERRA 96 showed a small amount

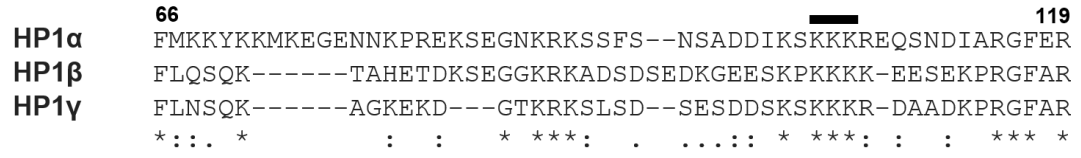

Figure 3.3: HP1 paralog hinge alignment. Alignment of the hinge region (between the chromodomain and chromoshadow domain of each paralog) using CLUSTAL O(1.2.4) multiple sequence alignment (Sievers et al., 2011). The numbers above the sequence indicate the amino acid position of the first and last hinge residue in the HP1α amino acid sequence. The black line indicates the three lysine residues in HP1α that were exchanged for alanine residues in the HP1α-3KA mutant. An "*" (asterisk) indicates positions which have a single, fully conserved residue. A ":" (colon) symbol indicates conservation between groups of residues with strongly similar properties. A "." (period) indicates conservation between groups of weakly similar residues.

of binding. This however was significantly less binding than when WT HP1α was immobilised on the tip and immersed in TERRA 96 (**Figure 3.2**). The abrogated binding by TERRA 96 to HP1α-3KA suggests the lysine patch in the hinge domain is essential for this interaction.

Due to the limited supply of TERRA 96, further analysis was carried out with a smaller, commercially synthesised RNA molecule. Termed, TERRA 45, this 45 nt RNA ($G_3(UUAGGG)_7$) has half the number of G-tracts of TERRA 96 (**Table 2.1**).

The same experiment as in **Figure 3.2** was performed with TERRA 45, where each paralog, as well as the HP1α-3KA mutant was immobilised and tested for binding by immersion in TERRA 45 using BLI (**Figure 3.4**). This, as expected, yielded similar results to the TERRA 96 binding to HP1 paralogs and mutant (**Figure 3.2**), with high binding response of TERRA 45 to HP1α, abrogated binding to HP1α-3KA, and low binding to HP1β and HP1γ.

3.2 Structure-dependent binding of HP1α

The sequences of TERRA used in BLI experiments contain G-tracts that would be expected to form G4s. For TERRA 96, this is known, as CD had been done prior by whom it was produced (Garavis et al., 2013). CD spectroscopy of TERRA 96 showed that it indeed formed a G4 structure. Due to characteristic positive

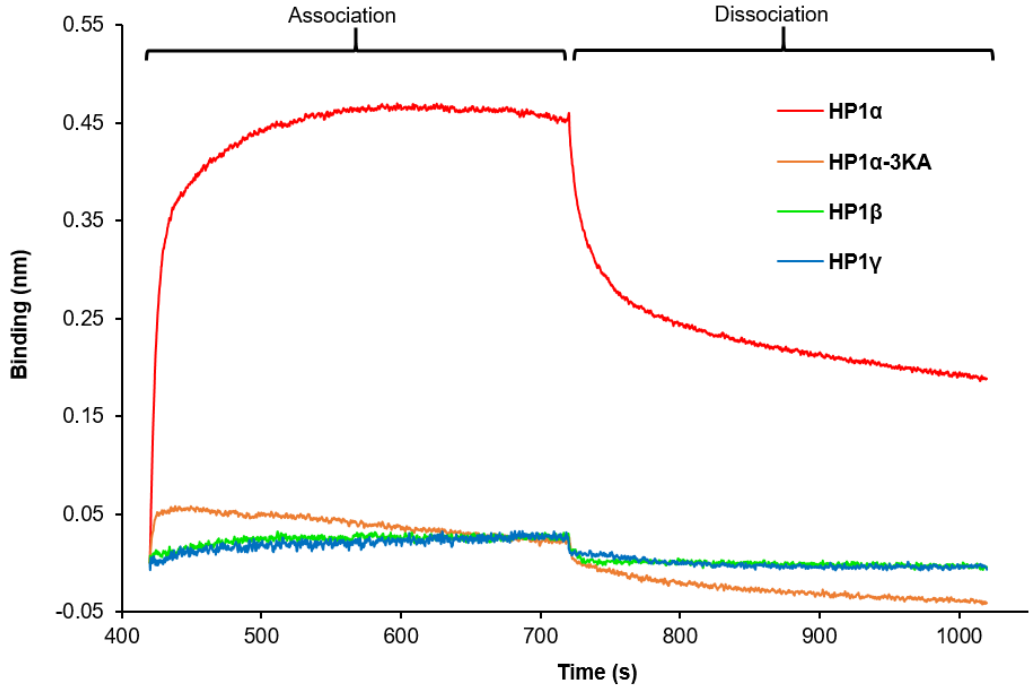

Figure 3.4: Binding of HP1 paralogs with TERRA 45. BLI performed with each HP1 paralog or mutant immobilised on a biosensor, with addition of TERRA 45, and subsequent washing in 1x IB to show the respective association and dissociation of TERRA 45 with each protein.

ellipticities at 210 nm and 265 nm and a peak with negative ellipticity at 240 nm (group I according to classification proposed by Karsisiotis et al. (2011)) this is consistent with parallel G4 topology. CD spectroscopy was also performed on the commercially produced TERRA 45, which showed virtually identical ellipticities (**Figure 3.5**), showing that TERRA 45 also forms parallel G4 topology with homopolar stacking of G-tetrads.

Since TERRA is a G-rich transcript that can form G4s, the coding strand from the telomeres of which it is produced from is also G-rich. Therefore, a 45 nt DNA molecule with the same sequence as TERRA 45, only with deoxyribose instead of ribose, and thymine instead of uracil, was produced commercially and termed d-TERRA 45 (**Table 2.1**).

Due to its G-rich sequence, d-TERRA 45 would be expected to form G4s, therefore CD spectroscopy was performed to test this. The CD spectrum showed that d-TERRA 45 indeed formed a G4 structure (**Figure 3.6**). However, due to maxima at 265 nm and 295 nm, the structure is deemed as "hybrid", with a higher maximum at 295 nm corresponding to heteropolar stacking between guanines,

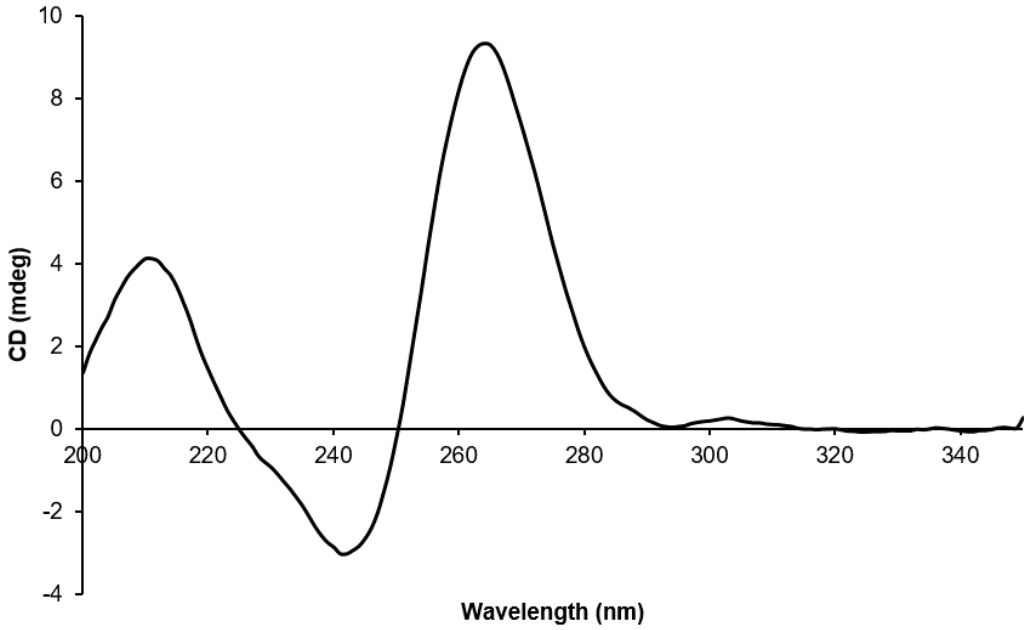

Figure 3.5: CD spectrum of TERRA 45. Spectrum of a 5 μM sample of TERRA 45 shows a maximum at approximately 265 nm, and a minimum at 240 nm, indicating a parallel G4 topology.

which is realised in anti-parallel G4s. As per the classification by Karsisiotis et al. (2011), it is a group II anti-parallel G4, with just one propeller (parallel-type) loop.

As a control for general DNA binding, a 45 nt string of thymidines, Thy45, was also produced. CD spectroscopy of Thy45 showed a spectrum of a typical (Vorlickova et al., 2012) unstructured ssDNA (**Figure 3.6**).

Binding between HP1 proteins and d-TERRA 45 was then tested by BLI as previously described. Addition of d-TERRA 45 to immobilised HP1 paralogs or mutant showed effectively no binding (**Figure 3.7**). No significant interaction was detected between HP1α and Thy45.

In vivo, d-TERRA 45, effectively the G-rich strand of telomeric DNA, may not be in its G4 form, but in a typical double-helix structure with the C-rich strand of the telomere. Therefore, d-TERRA 45 was annealed to its C-rich counterpart (C-telomere 45, **Table 2.1**) to form a 45 bp duplex to model the telomeric DNA.

Firstly, to confirm that the DNA had annealed in a double-helix structure, CD spectroscopy was performed. This showed typical features of a dsDNA (Vorlickova

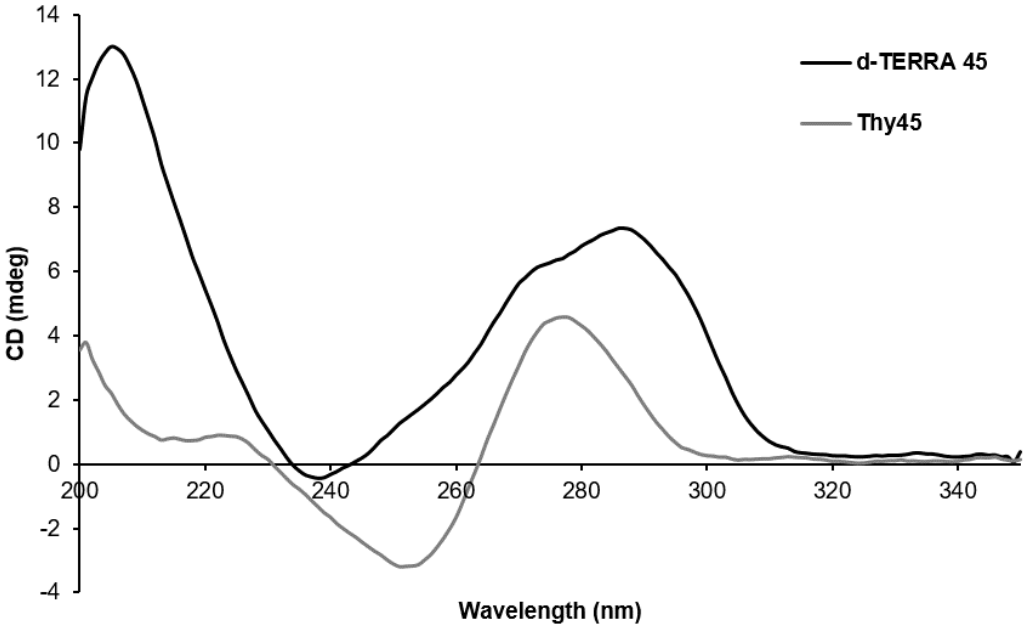

Figure 3.6: CD spectra of d-TERRA 45 and Thy45. Spectra of a 5 μM sample of d-TERRA 45, and a 5 μM sample of Thy45.

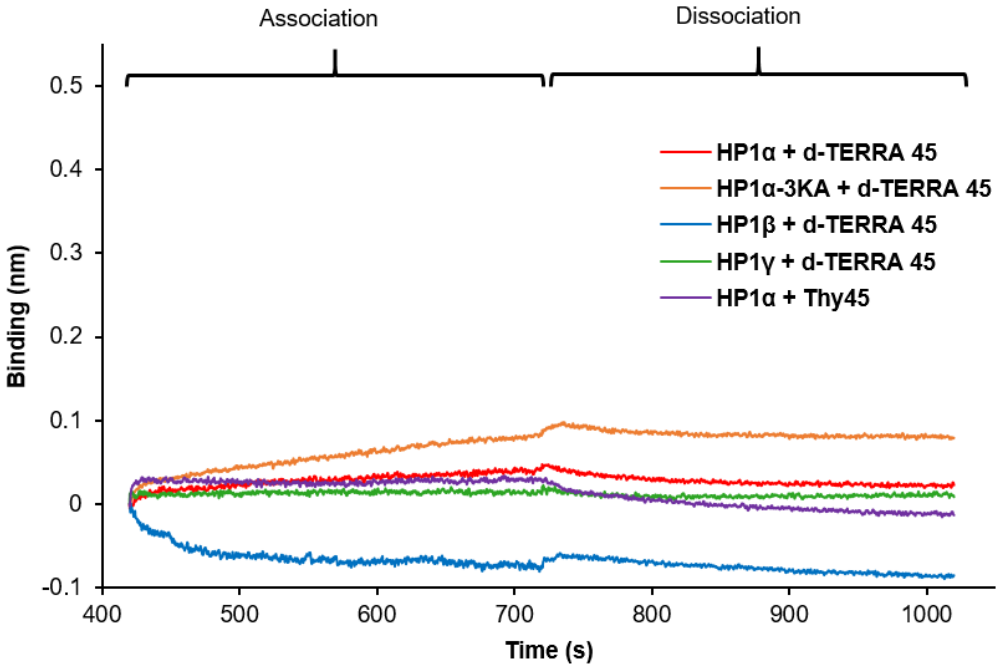

Figure 3.7: BLI spectroscopy of d-TERRA 45 and Thy45. Assessing the binding between each HP1 paralog or mutant, bound to the tip, with addition of d-TERRA 45, or with HP1α immobilised on the tip with addition of Thy45.

et al., 2012), confirming duplex formation (**Figure 3.8a**).

BLI was then performed to test whether HP1α would bind this 45 bp telomeric repeat duplex. With HP1α immobilised and then immersed in annealed d-TERRA 45 and C-telomere 45, BLI showed no significant binding (**Figure 3.8b**).

TERRA 45 is an RNA G4 with parallel topology which binds HP1α, and d-TERRA 45 is a DNA G4 with anti-parallel topology that does not bind HP1α. To test whether HP1α binding ability is simply a preference for RNA over DNA, as HP1α did not bind the other DNA molecules tested, an RNA control sample was also tested.

Given that TERRA 45 and TERRA 96 are highly ordered RNA molecules, HP1α was tested for its binding with the highly ordered RNA molecule, tRNA. Firstly, to confirm that the RNA was indeed folded into the typical tRNA structure, CD was performed, affirming that it had the same ellipticities (**Figure 3.9a**) as published CD spectra of tRNA (Blum, C. Uhlenbeck, & Tinoco, 1972). By BLI, with HP1α immobilised on the tip, and immersed in tRNA, it was shown that tRNA did not bind to HP1α (**Figure 3.9b**).

To test whether this dramatic shift in binding between TERRA 45 and d-TERRA 45 was due to the conformational differences in G4 formation, several DNA oligonucleotides derived from the 45-mer telomeric repeat in which the length of loop regions varied were investigated (**Table 2.1**). It is generally accepted that shortening of loops from three to one nucleotide can force G4s to adopt a parallel rather than anti-parallel topology (Mateus WebbadaSilva 2009, Hazel, 2004), therefore the loops between G-tracts of d-TERRA 45 were altered and then their conformations tested by CD spectroscopy.

The d-TERRA 45 sequence was altered to shorten the loops (**ATT**GGG), removing either A (**TT**GGG) (Oligo B) or AT (**T**GGG) (Oligo A), to attempt to force the DNA into a parallel topology. Shorter variations of d-TERRA 45 were chosen with various loops shortened to just T, or remaining as ATT in different positions to find exactly which loops contributed to parallel topology (**Table 2.1**).

Both Oligo A and Oligo B exhibited CD spectra of G4s in the parallel topology (**Figure 3.10**) due to the shortening of the loop of the G-quadruplex. Oligos D, H, and I also formed parallel G-quadruplexes, whereas Oligos E and G formed anti-parallel conformations. Oligos D, H, and I all have sequences with just

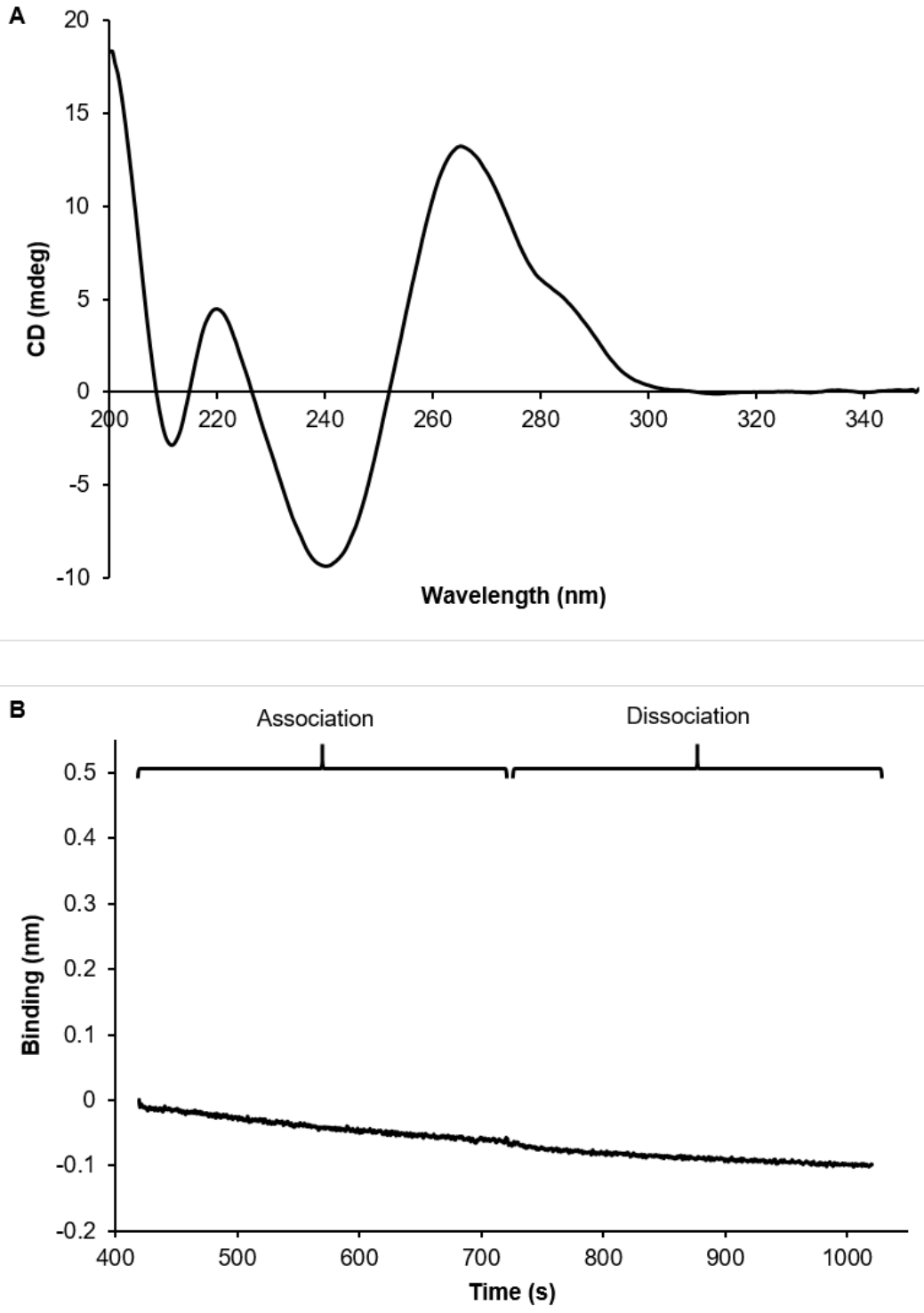

Figure 3.8: CD and BLI of telomere duplex 45. a) CD spectrum of a 45 bp duplex of the telomeric repeat. b) BLI measurements assessing the binding of 45 bp duplex telomeric DNA to HP1α bound to the tip.

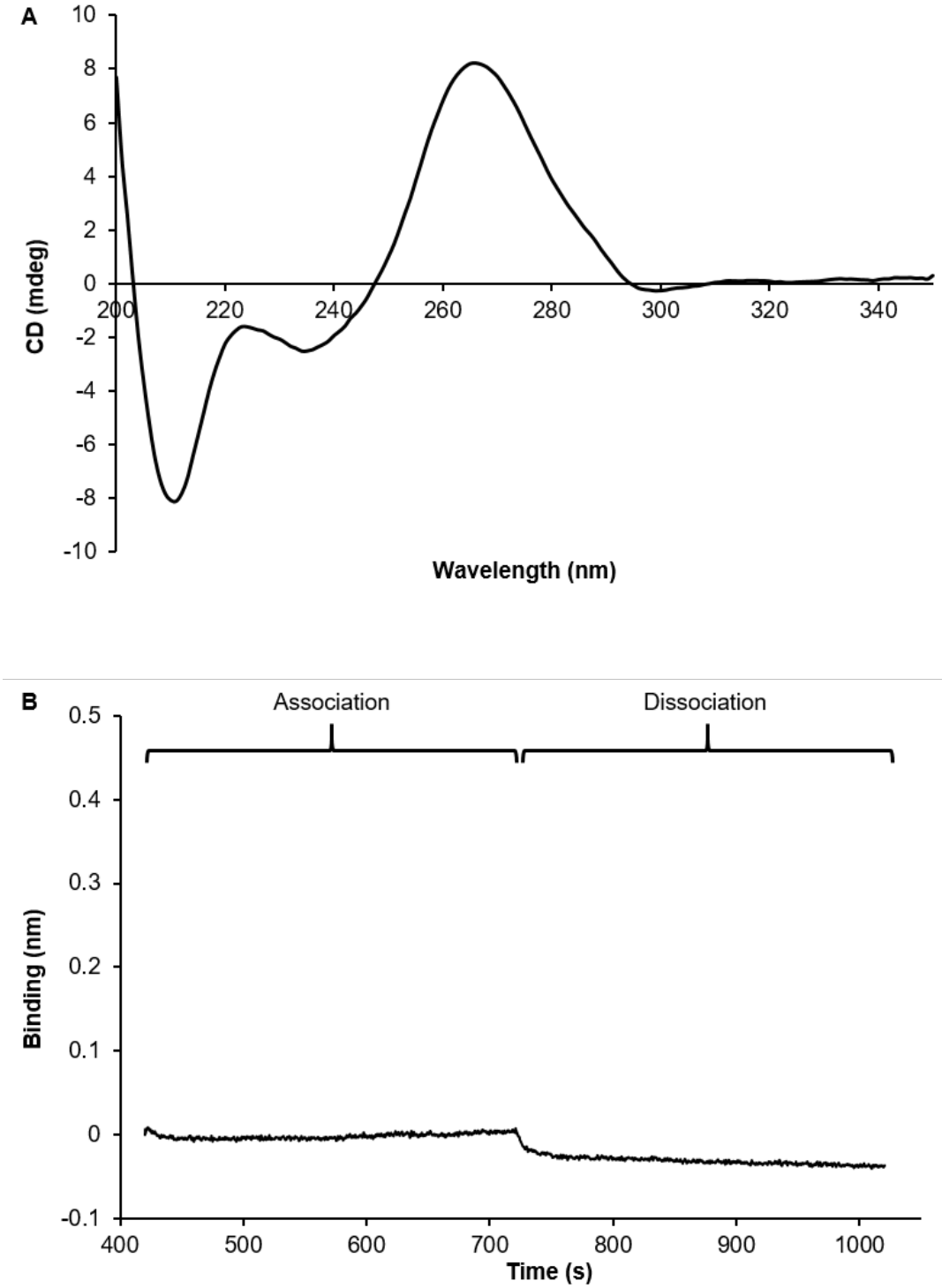

Figure 3.9: CD and BLI of tRNA. a) CD spectrum of a 5 μM sample of tRNA. b) BLI measurements assessing the binding of tRNA to HP1α bound to the tip.

one TTA loop, showing that just two of the three loops must be shortened for parallel G-quadruplex formation. Oligos E and G both had at least two TTA repeats sequentially from the 3' end of the oligonucleotide, which prevented them from forming parallel topologies. CD spectra of Oligos C and F showed a positive peak at 260 nm, as well as a positive peak at 295 nm, though at a lower intensity (**Figure 3.10**), which suggests presence of G-tetrads with heteropolar stacking of guanines.

BLI was then performed with immobilised HP1α on the tip, and immersion in each aforementioned oligonucleotide to test for binding.

Given the differences in length between Oligo A and Oligo B compared with Oligos C-I, and the dependency on a molecule's size to interfere with light in BLI, these oligonucleotides were found to be not comparable. None of the shorter Oligos (C-I) produced a high BLI signal (**Figure A.4**), indicating that either there was no binding between HP1α and these oligonucleotides, or that these oligonucleotides are too small to produce a significant signal upon binding.

Selected oligonucleotides were elongated, with their original sequence repeated twice in order to provide DNA of sufficient molecular weight for BLI analysis. Oligo C and Oligo D were chosen to add parallel G4s to the dataset, and Oligo E, Oligo F, and Oligo G were chosen to add anti-parallel G4s. CD was then performed on the elongated oligonucleotides (denoted by addition of a "2" to the name of the oligonucleotide (**Figure 3.10**)). Oligo 2C and Oligo 2D formed parallel G4 conformations as expected. Oligo 2G also formed the expected anti-parallel conformation. However, Oligo 2E and Oligo 2F both formed parallel G4s. This could be due to each of them possessing two short loops containing just one thymidine in the middle of their sequences, where Oligo 2G only has one short loop.

With HP1α immobilised on the tip, it was then immersed in each of the elongated oligonucleotides, along with Oligo A and Oligo B, and tested for their binding with HP1α using BLI. All of these oligonucleotides, bar Oligo 2G, showed binding with HP1α, while the only anti-parallel oligonucleotide did not show any significant binding (**Figure 3.12**). Of note, the response to Oligo B binding to HP1α exceeded that of TERRA 45 (**Figure 3.4**) at the same concentrations, and of similar DNA length.

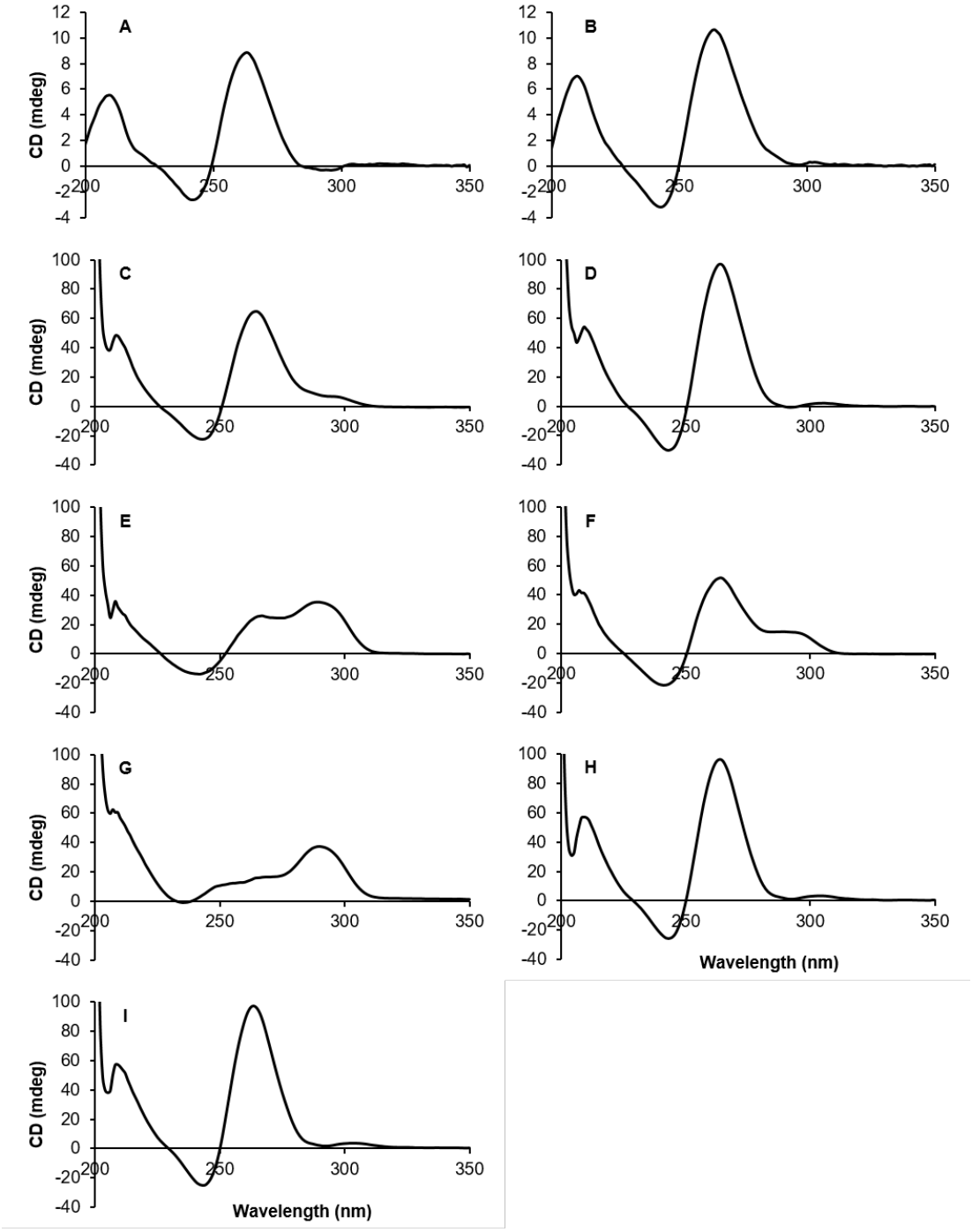

Figure 3.10: CD spectroscopy of altered d-TERRA sequences. CD spectra of oligonucleotides A-I, with Oligo A and Oligo B at 5 μM and oligonucleotides C-I at 10 μM.

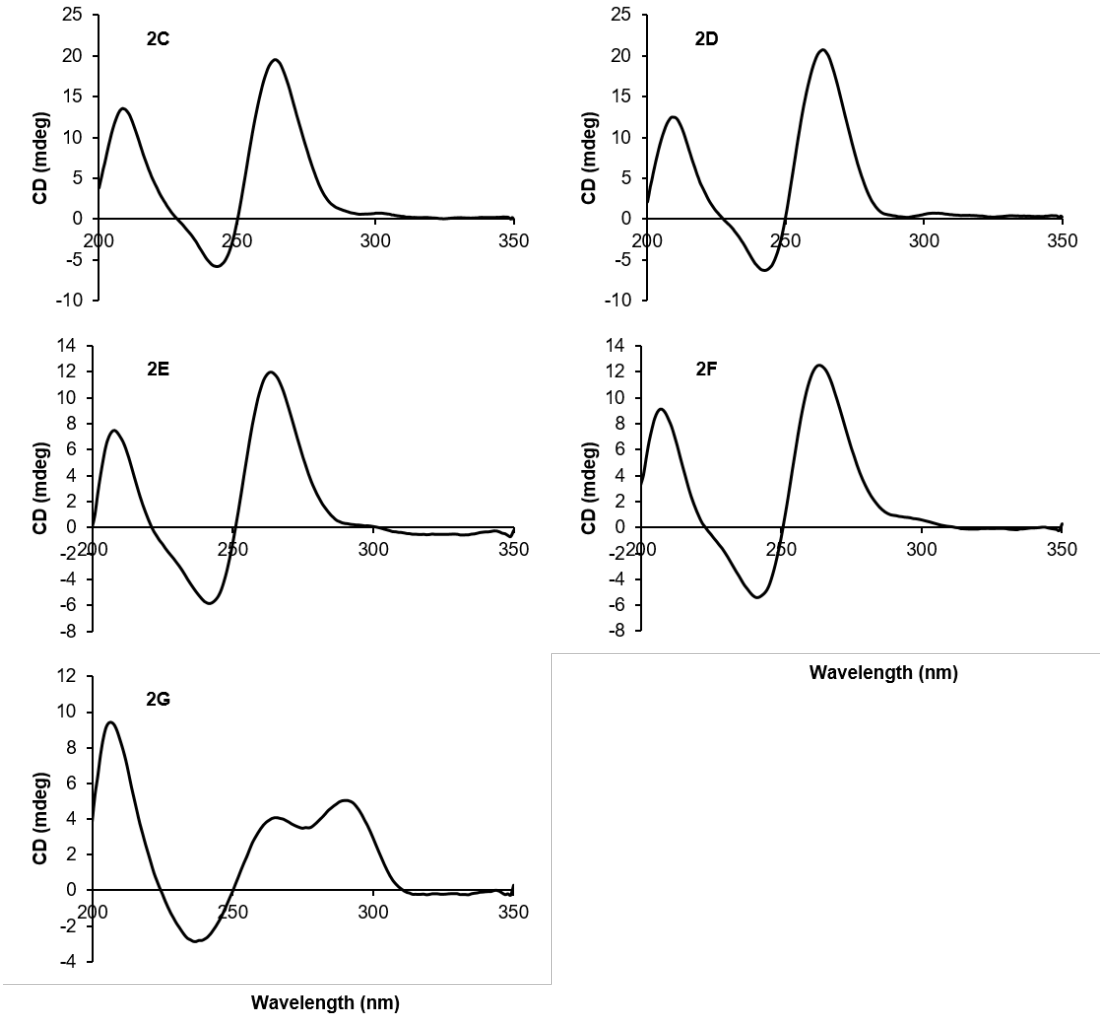

Figure 3.11: CD spectra of elongated altered d-TERRA sequences. CD spectra of oligonucleotides 2C-2G at 10 μM.

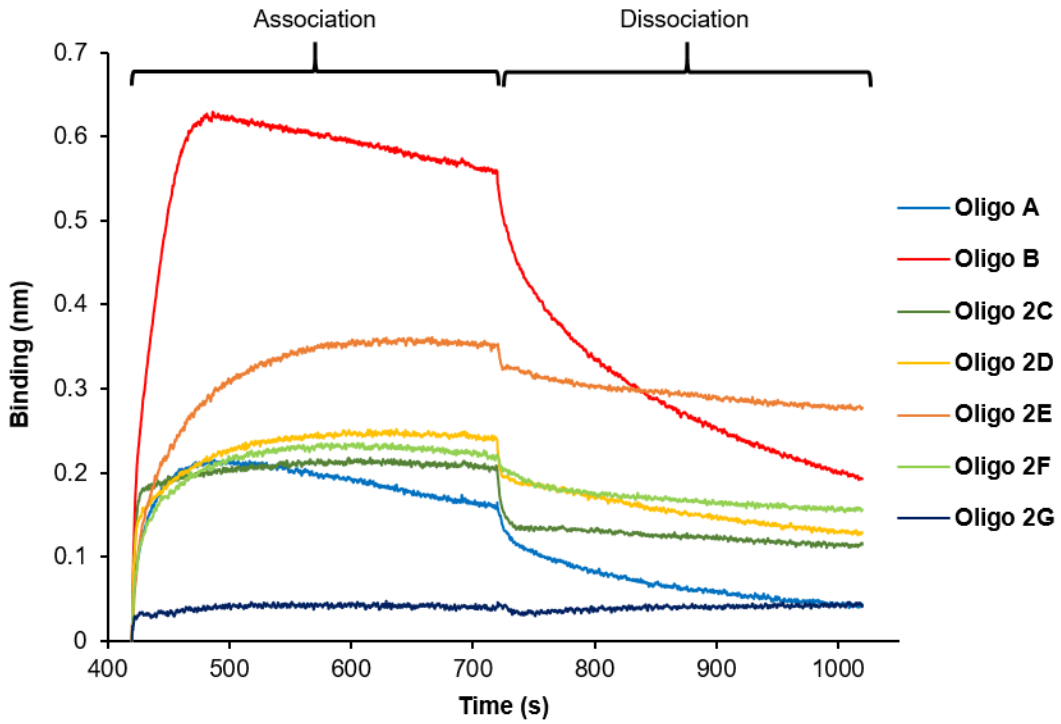

Figure 3.12: BLI of elongated altered d-TERRA sequences. BLI performed with HP1α bound to the tip, with addition of each elongated altered d-TERRA 45 sequence.

3.3 HP1α binds endogenous promoter G4s

Throughout the genome, there are countless G-rich sequences that are capable of forming G4s. Many of these sequences are found in promoter regions of genes, suggesting a role in regulation of gene expression.

To test if HP1α binds non-canonical nucleic acid structures which are present endogenously, a selection of well characterised G-rich promoters and regulatory sequences were investigated.

Firstly, their conformation was confirmed by CD to test if they indeed formed the G4s, and if they did, which G4 conformation they were in (confirming topologies stated in previous publications). All of these endogenous G4s except two were shown to have parallel conformation, with SRC2 and SRC16 conforming to mostly anti-parallel topologies (**Figure 3.13**).

All of these oligonucleotides were then tested for their interaction with HP1α, and each showed some interaction with HP1α to varying degrees (**Figure 3.14**). The

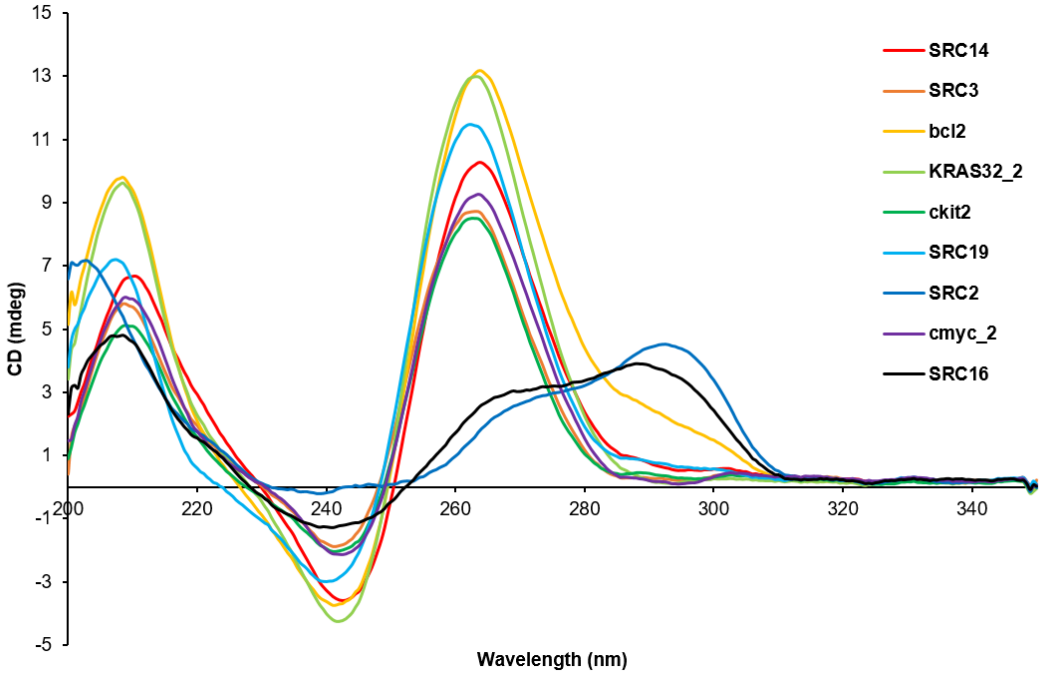

Figure 3.13: CD spectroscopy of endogenous G4s. CD spectra of sequences from the human genome, known to form G4s, at 10 μM.

three endogenous oligonucleotides with the lowest measured binding with HP1α were SRC16, SRC2, and cmyc_2. Of these three, two of them are the only anti-parallel G4s of the selection, showing that the anti-parallel conformation is very unfavourable to HP1α. The other low binding oligonucleotide was cmyc_2, which has the parallel G4 topology. However, given that BLI binding signal is dependant on amount of light being interfered with, the size of the interfering molecule is a key factor in this. The cmyc_2 sequence was the shortest of all endogenous G4 oligonucleotides to be tested (16 bp), and therefore could be expected to produce a lower signal. Further to this, despite the signal of low intensity, the binding of cmyc_2 to HP1α was the most rapid of the endogenous G4s. The BLI signal increases very sharply compared to the other oligonucleotides, with a binding pattern similar to TERRA 45's response curve.

Given this strong preference by HP1α to interact with non-canonical DNA or RNA G4s, another similar type of non-canonical DNA structure was investigated. i-Motifs form four stranded nucleic acid structures too, by the binding of pairs of cytosines (**Figure 1.1**). Therefore, they are typically formed in C-rich sequences. Also, given that guanine and cytosine are paired together in DNA, if a G4 was formed on one strand, an i-motif may form on the other.

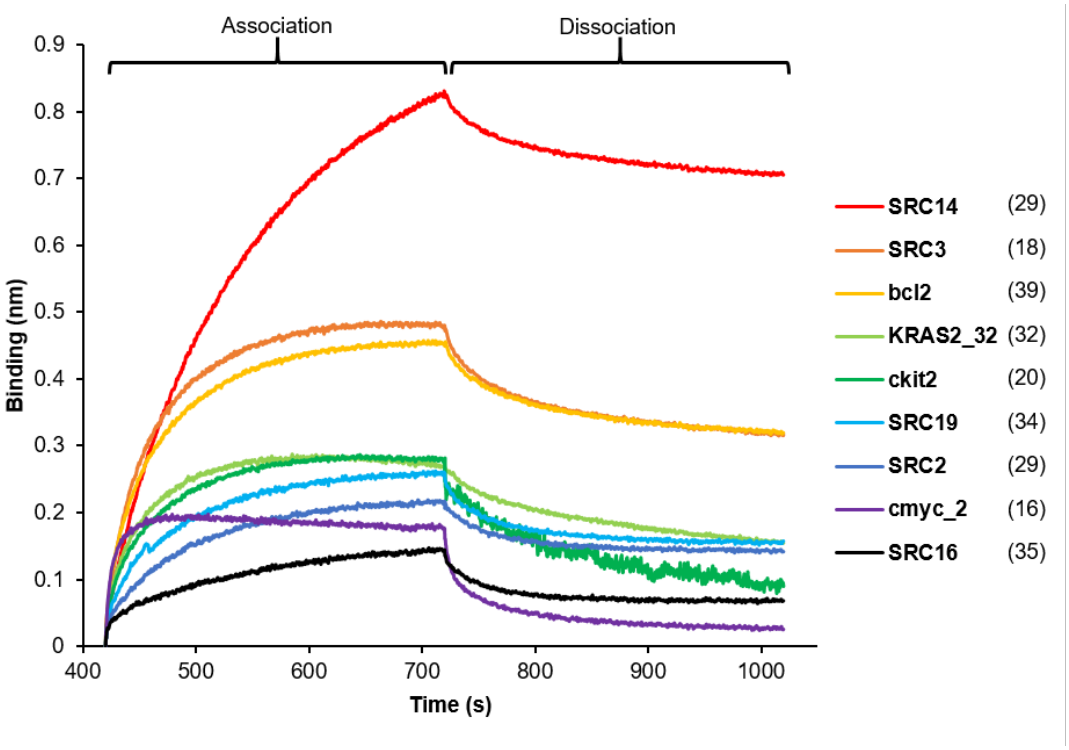

Figure 3.14: BLI of endogenous G4s. Binding measured between G4 forming sequences from the human genome, and HP1α bound to the tip, by BLI. Numbers adjacent to sequence names indicate their length in nucleotides.

Therefore, the C-rich strand of telomeric DNA (C-telomere 45, used as part of the telomere duplex molecule tested), along with other published, biologically relevant i-motif forming sequences were explored (**Table 2.1**).

The folding of i-motifs was first investigated by CD spectroscopy (after first being annealed in 1xIB7, the buffer used throughout this work but at pH 7 rather than pH 8) to confirm their folding, which were all shown to form i-motifs (**Figure 3.15**). Following this, they were tested for their interaction with HP1α by BLI, by immobilising HP1α on the tip, and immersing in each subsequent i-motif oligonucleotide. BLI showed no interaction between HP1α and any of the i-motifs (**Figure 3.16**), which were of similar lengths to the elongated and endogenous G4s tested previously. This shows that HP1α has a strong preference for a specific non-canonical nucleic acid structure that it binds to.

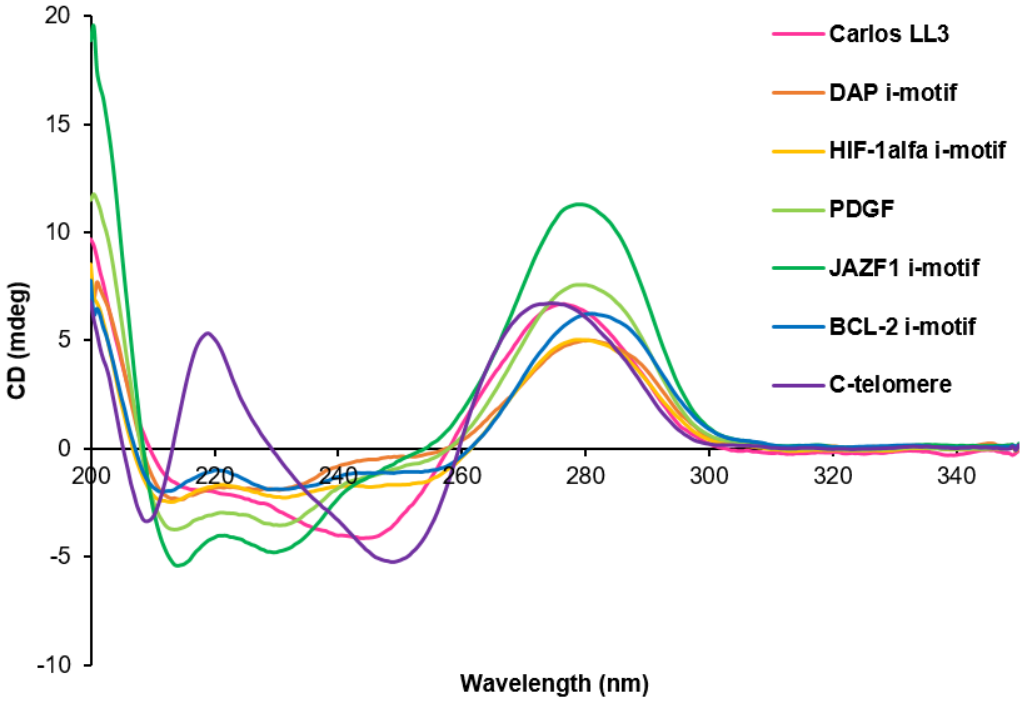

Figure 3.15: CD spectroscopy of i-motifs. CD spectra of sequences from the human genome, known to form i-motifs, at 10 μM.

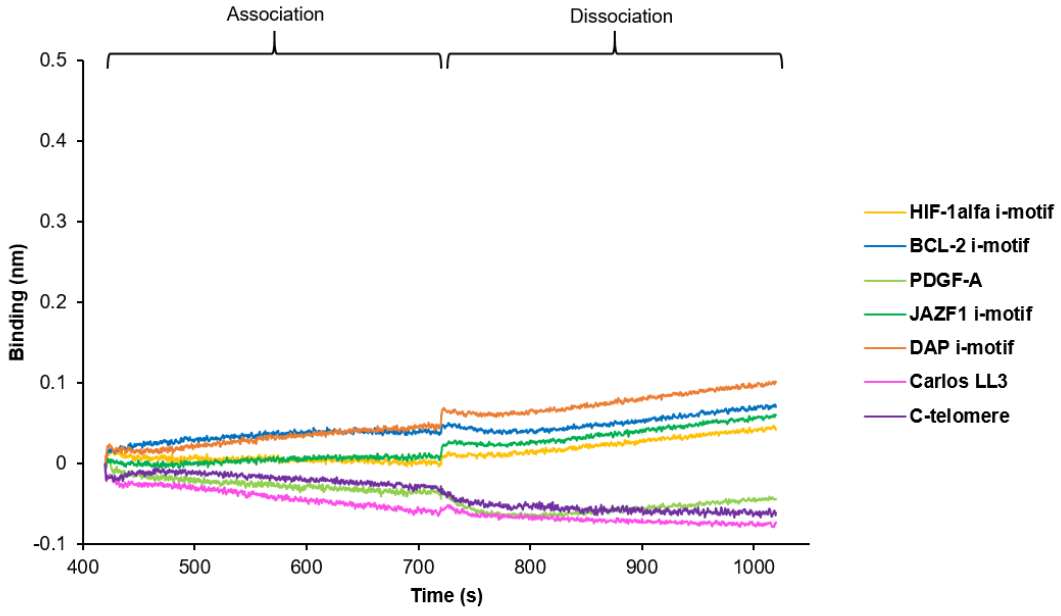

Figure 3.16: BLI of i-motifs. BLI performed with HP1α immobilised on the tip, and addition of i-motif-forming sequences from the human genome.

3.4 Discussion

These results show the outstanding specificity of the HP1α protein, and the strong dependency on only three residues within the hinge of HP1α for this binding to G4s.

Firstly, TERRA 96 and TERRA 45, both parallel G4s, bind only to the HP1α paralog. This is a contradiction to the findings of Deng et al. (2009), who show binding to both HP1α and HP1β through RNA immunoprecipitation of endogenous TERRA. Although slightly more TERRA is immunoprecipitated with HP1α than with HP1β, the amount immunoprecipitated with HP1β is significant. The discrepancy between these results could firstly be due to the difference in endogenous TERRA versus commercially produced (TERRA 45) or *in vitro* transcribed molecules (TERRA 96). Endogenous TERRA would be of many varying lengths, and may have complex folding and intermolecular G4 formations. The buffer used in each of the respective experiments may have affected folding of the G4s, and also may have caused different charges to be present in the protein, causing a potentially lower binding affinity between HP1β and TERRA. Secondly, the experiments performed above, being *in vitro*, lack the complex mixture of molecules present in a cell. Therefore, given that HP1β bound TERRA in *in vivo* experiments (Deng et al., 2009), this interaction may require additional cofactors that aren't present in *in vitro* BLI experiments, in which only purified components were tested. Contrary to this, the lack of other molecules present in these experiments highlights the strength and importance of the binding between HP1α and TERRA.

A small amount of binding was shown between HP1γ and TERRA 96 in the experiments performed above, which again contradict Deng et al. (2009), but again, the differences could be due to *in vitro* versus *in vivo* experiments. HP1γ does function at human telomeres, and is able to bind a component of the shelterin complex which protects telomere ends (Canudas et al., 2011). TERRA has also been shown to interact with shelterin components, TRF1 and TRF2 (Deng et al., 2009), which may indicate a weak interaction between HP1γ and TERRA as part of this complex.

Each of the TERRA molecules, when tested for binding with the HP1α hinge mutant, HP1α-3KA, showed a severely abrogated response. This shows the charge of the lysine patch in the hinge domain is essential for this interaction. This is to

Table 3.1: QGRS sequences found in Widom 601 DNA

Position	Length	QGRS	G-Score
29	28	GGAGACTAGGGAGTAATCCCCTTGGCGG	9
29	28	GGAGACTAGGGAGTAATCCCCTTGGCGG	10
38	30	GGAGTAATCCCCTTGGCGGTTAAAACGCGG	10
52	18	GGCGGTTAAAACGCGGGG	12
52	19	**GGCGGTTAAAACGCGGGGG**	13
52	19	GGCGGTTAAAACGCGGGGG	11

be as expected, given the affinity of protonated amino acids to bind negatively charged phosphates in DNA or RNA. What is surprising is that the same lysine-rich patches are present in both HP1β and HP1γ (**Figure 3.3**). This implies that there are other regions of the HP1α hinge contributing to the interaction, or other domains of HP1α interacting with TERRA.

HP1α has been shown to interact with isolated DNA, used for reconstituting nucleosomal arrays (Nishibuchi et al., 2014). This DNA is derived from the Widom 601 sequence, and is commonly used in nucleosomal arrays because of the high affinity of this DNA for nucleosomes. A 147 bp DNA fragment of the Widom sequence was tested for putative Quadruplex forming G-Rich Sequences (QGRS) using QRS Mapper (Kikin et al., 2006). The reverse strand was shown to have six putative QGRSs, with G-scores ranging from 9-13. The G-score is a measure of propensity to form G4s, the higher the score, the higher the likelihood to form a stable G4, with a maximum possible score of 105 (for the extremely G-rich sequence $(G_5T)_3G_5$). Despite the low scores, these sequences may still be able to form G4s, and this could account for the binding by HP1α to this isolated DNA sequence. In order to test this theory, the DNA may need to be tested by CD spectroscopy to show if it had indeed formed G4s. If testing long strands of DNA in which many G4s may form, DMS footprinting may need to be used. DMS methylates guanine on the N7 position, however, this N7 atom becomes a hydrogen acceptor when in a G4, but not in normal Watson-Crick base-pairing. Therefore, when this methylated DNA is subjected to piperidine treatment to produce double-strand breaks at methylated nucleotides, G4s are preserved and shown by Polyacrylamide Gel Electrophoresis (PAGE) as a "footprint" of absent bands on the gel (Sun and Hurley, 2010).

Of course, parallel G4s may not be the singular DNA structure that HP1α binds to, but despite a general acceptance of a DNA binding domain there is very little evidence in the literature, other than the binding to the Widom sequence, that

shows that HP1α binds DNA in mammalian cells.

Interestingly, there were DNA G4s that appeared to bind HP1α more than TERRA 45. Oligo B and SRC14 had a much higher binding response in BLI with HP1α than TERRA 45, despite being shorter in length. SRC3 and bcl2 also bound HP1α to a higher degree than TERRA 45. What they all have in common, of course is that they are of parallel G4 topology. However, there are many more facets to G4 structure. Other distinctions include dimerisation between G4s, and whether these form through interactions at the bases of G4s, or whether this is because of bi, tri, or tetra-molecular interactions within the G4 itself. This will be explored in **Section 4.1**.

The biologically relevant DNA G4 sequences used in this work were chosen because they have been previously shown to form G4s, and also because of the pivotal roles these G4s may have in the regulation of normal and/or cancerous cells.

The bcl2 sequence used in this study is a 39 nt long fragment derived from a GC-rich region upstream of the P1 promoter of the gene that codes for bcl-2 (B-cell lymphoma 2). This region has been shown to be "critically involved" in regulating expression of the BCL2 gene (Agrawal et al., 2014). The bcl-2 protein is a pivotal component of the apoptotic pathway, where cells enter programmed cell death. This is extremely important in order to prevent cancerous growth, making BCL2 a tumour suppressor gene. The binding of HP1α to the G4 formed by the BCL2 upstream sequence could influence the expression of this gene. In the context of treating cancer, if HP1α or another molecule designed to target bcl-2 expression could up-regulate production of bcl-2 protein in cancerous cells, this may provide a way to kill tumours.

C-myc is part of the MYC family of genes, which are transcription factors involved in the regulation of many genes, some of which are responsible for cell proliferation. They are therefore proto-oncogenes, capable of allowing over-proliferation of cancer cells. As with bcl2, HP1α or other molecules designed to bind these G4s could be used to target and regulate the genes that contain specific G4 structures in their regulatory regions, such as c-myc, relevant to cancer.

C-kit is a receptor tyrosine kinase involved in intracellular signalling, resulting in its dysregulation being an oncogenic driver of cancer progression. Given its propensity for forming G4s (Diveshkumar et al., 2016), again, this is a possible

mechanism of G4-dependant oncogenesis, and potential G4-targeted drug design.

Yet another oncogene, KRAS (Kirsten Rat Sarcoma), has been investigated in this work. The G4 sequence is within the promoter of the KRAS gene (Cogoi and Xodo, 2006), potentially regulating transcription of p21RAS, crucial to the Raf-MAP kinase pathway, which transmits signals that influence cell growth and apoptosis.

Within the SRC proto-oncogene are numerous G-rich sequences that have been shown to form G4s (Rodriguez et al., 2012). A selection of these previously studied sequences were chosen for this work, the locations within the gene and strand of which are shown in **Table A.1**. The work that characterised these SRC G4s found that a G4-binding molecule, pyridostatin, modulated the expression of SRC and subsequently src protein abundance, affecting the cellular motility of cancer cells. This shows that molecules capable of binding G4s affect the expression *in vivo*, with a measurable response in cancerous phenotype. This same method could be followed, modulating the HP1α abundance and/or binding capabilities (expressing HP1α-3KA), and testing the affects this might have on cancer cell phenotype.

As discussed in **Section 3.2**, the maxima at 265-270 nm and at 285-290 nm and minima at 240 nm indicate a group II G4 topology (Karsisiotis et al., 2011), which are actually "mostly" anti-parallel structures. Group II G4s possess one propeller loop (which parallel G4s have), meaning parts of the G4-forming sequence that link together G-tracts are on the outside of the G4 structure, adjacent to the quartets (not above or below like lateral and diagonal loops). If the recognition by HP1α to parallel G4s is through binding to these propeller loops, then it is possible that group II G4s may exhibit similar binding patterns if they have compatible sequences present in propeller loops despite their anti-parallel structure. Further investigation is needed to determine what HP1α is specifically recognising in parallel G4 structures, and whether this is binding to sequence-specific motifs in loops, or whether it is the overall structure of the G4 that is most important.

The i-motifs used in this work were chosen because of their biological relevance, and the mild conditions under which they can form. The i-motifs chosen were all used with IB7 in order to facilitate the formation of i-motifs at pH 7 (rather than at pH 8 in IB). This pH was also chosen because of its cellular relevance, showing how these structures may form *in vivo*.

One of the i-motifs used in this study (Carlos LL3rep) actually forms a "mini i-motif" in which two pairs of hemi-protonated cytosines bind the same way as in an i-motif, but they are then flanked by two tetrads which result from the association of G:C or G:T base pairs, similar to G4 formation (Mir et al., 2017).

Formation of this structure at 20 °C did not show complete formation of the mini i-motif structure, however when CD spectroscopy was performed at 5 °C (**Figure A.5**), the spectrum showed the same ellipticities as previously published (Mir et al., 2017). Given that BLI was performed at room temperature, it is possible that the mini i-motif was not fully formed, and to test if HP1α does bind this structure, perhaps BLI needs to be performed at 5 °C too, or have the mini i-motif structure stabilised in some other way.

Upon close inspection of the CD spectra of these i-motif-forming oligonucleotides, their maxima and minima are not as distinct as would be expected for i-motif formation. A maximum around 285 nm would be expected, however, under the conditions used, this peak with positive ellipticity is shifted to the left in most cases, forming a peak around 275-285 nm. This indicates that the i-motifs may be slightly unfolded, however, BLI analysis and CD have been performed on the C-rich telomeric DNA strand at varying pH (Zeraati, 2018). This analysis showed the same pH shift from 275 nm (at pH 7), to a more folded form at 287 nm (at pH 6). The binding between this telomeric i-motif and an i-motif-recognising antibody was then assessed using BLI, and it was found that the pH affected the binding, but not drastically (association at 10.5 nm, to association at 9 nm). The BLI experiments in the aforementioned publication were performed using a oligonucleotide concentration of 200 nM, whereas the BLI experiments carried out in this work were performed at 500 nM (same volume). If there was slight binding between the fully formed i-motif structures and HP1α, if it were significant, HP1α would still likely have bound the folded form, which would be more biologically relevant.

Another interesting note, is the similarity in structure between anti-parallel G4s and i-motifs. They both have strands that arrange themselves in opposite directions, leading to diagonal or lateral loops at the top or bottom of the structures, whereas parallel G4s possess propeller loops, on the sides of the G4 structure (**Figure 1.1**).

Overall, these results show the intricate specificity of HP1α for a subset of G4

structures, and a need to further explore the roles of non-canonical G4 structures *in vivo*.

Chapter 4

Elucidating structural properties of HP1α and TERRA 45

To elucidate the structural nuances of the Heterochromatin Protein 1α (HP1α)-telomeric repeat-containing RNA (TERRA) 45 interaction, firstly, the oligomeric state of the RNA and proteins used in **Chapter 3** were investigated.

Oligomeric qualities were investigated using native Polyacrylamide Gel Electrophoresis (PAGE) to separate oligonucleotides, proteins, or complexes while maintaining secondary structure. Sodium Dodecyl Sulfate (SDS)-PAGE was also used in order to estimate the molecular weight of denatured cross-linked protein species to compare with native PAGE results.

Following this, binding affinity was measured using Biolayer Interferometry (BLI), and corroboration of these results was attempted using Isothermal Titration Calorimetry (ITC). ITC was used to measure thermodynamic properties of the interaction. This can be done by adding small, equal amounts of TERRA 45 into a cell containing HP1α, and measuring the heat being liberated or absorbed as a result of non-covalent bond redistribution upon binding with each injection of TERRA 45. The heat change is measured by the amount of power required to keep the sample cell at the same temperature as the reference cell. This information can then be used to calculate the molar ratio of the interaction, the binding constant of the interaction, and also the enthalpy and entropy of the interaction.

Measuring the strength of the binding between HP1α and TERRA 45 gives context to the importance of this interaction, especially in comparison to other interactions that are thought to maintain constitutive heterochromatin, such as the interaction between the HP1α chromodomain and H3K9me3.

Small-Angel X-ray Scattering (SAXS) was used in this work in order to investigate sizes and flexibility of HP1α and TERRA 45. SAXS has the advantage of examining the structure of macromolecules in solution. This provides a better understanding of the structure of the molecule *in vivo*, as crystalline structures may not form in the same way in the cell as when packed and crystallised into an ordered structure. SAXS determines these structural properties through detecting the elastic scattering behaviour of X-rays, diverted at small-angles by electrons of the molecule of interest. This can then be processed to show averaged particle sizes, shapes, distribution, and surface-to-volume ratios in order to model the compact or unstructured nature of HP1α and/or TERRA 45.

This chapter will seek to determine the affinity of the HP1α-TERRA interaction, the oligomeric state of HP1α and TERRA before and after the interaction takes place, and the structural changes that take place upon binding.

4.1 G-quadruplex oligomerisation

Following on from assessing G4 topology by CD spectroscopy, the ability of G4s to form higher order structures was assessed by PAGE under native conditions, preserving G4 structure and oligomerisation ability.

Native PAGE showed that many of the oligonucleotides are present in multiple forms, perhaps forming many combinations of inter- and intramolecular G4 structures (**Figure 4.1**). Of note, some of the oligonucleotides that bind HP1α more strongly (as shown by BLI) have lower mobility (higher bands) than others. For example, Oligo B, cmyc2, SRC14, and TERRA 45 all have a strong binding, and also all appear to form higher order structures that many of the other oligonucleotides do not (repeated native PAGE showing clearer TERRA 45 monomer and dimer species is shown in **Figure A.6**).

There is a stark difference between Oligo A and Oligo B, given that their loop length only differs by one nucleotide. They both possess the parallel G4 topology,

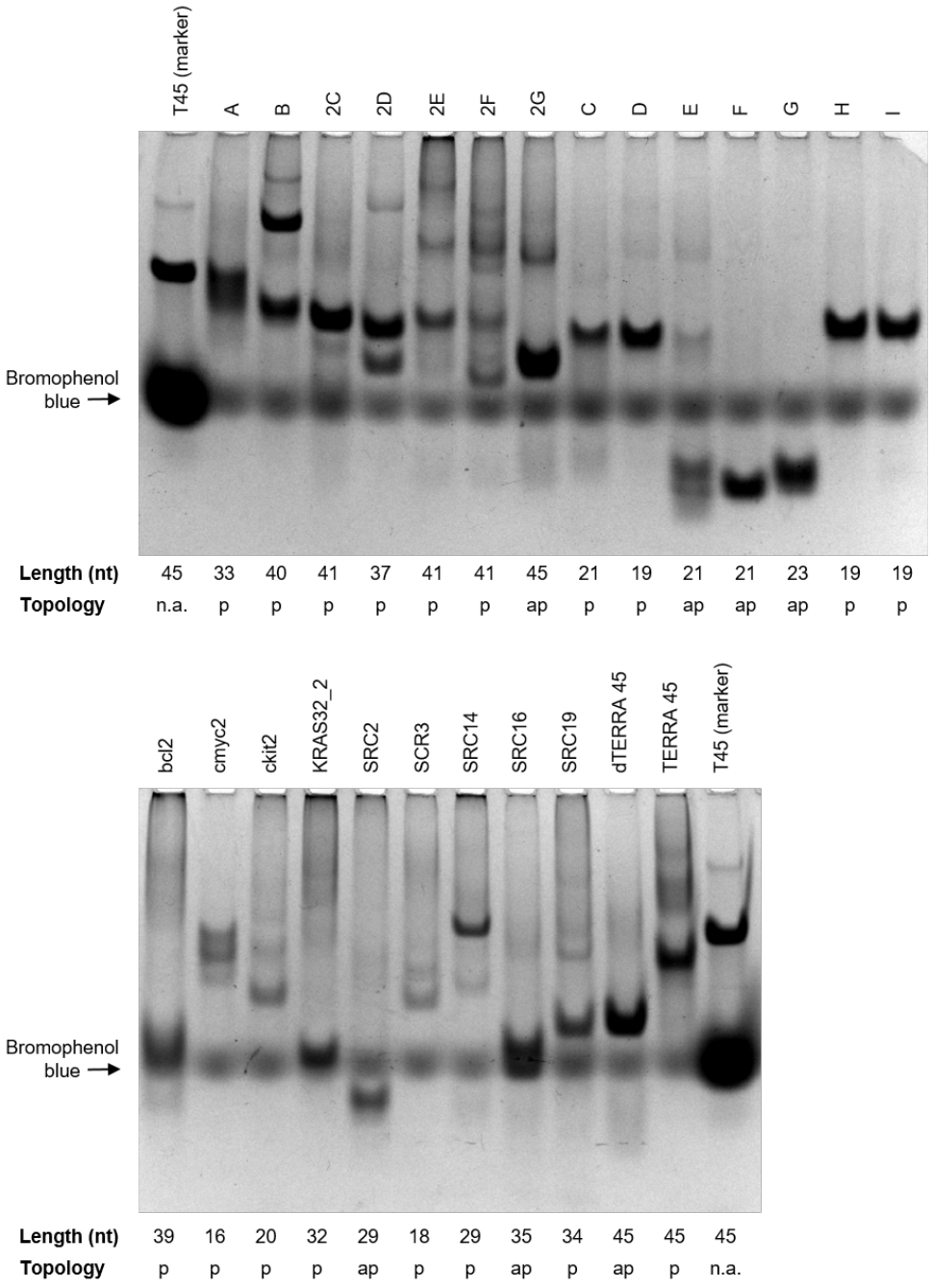

Figure 4.1: Native PAGE of G4-forming oligonucleotides. A native 20% PAGE was performed and stained with Stains-All, with Thy45 (labelled T45) used as a reference for mobility of unstructured DNA, and to visualise bromophenol blue mobility as the gel was electrophoresed. Topology refers to whether the oligonucleotide formed parallel G4s (p), anti-parallel G4s (ap), or did not form G4s (n.a.).

due to the loop reduction from anti-parallel d-TERRA 45. However, Oligo B has a band of very low mobility, showing that the loops of the G4 may contribute to the oligomerisation ability of the molecule, not just the topology.

Another observation from this experiment was that the anti-parallel G4s had higher mobility than their parallel counterparts. For example, of the short d-TERRA 45 variants (oligonucleotides C-I), the three that formed anti-parallel topology (oligonucleotides E, F, and G) also migrated substantially farther through the PAGE than the others. This stark difference is also present in the comparison of TERRA 45 to its anti-parallel DNA counterpart, where d-TERRA 45 has a much higher mobility, despite being the same number of nucleotides, and therefore very similar in molecular weight.

4.2 Protein oligomerisation and mobility

When purifying HP1α by SEC, the absorbance at 214 nm and 280 nm of the primary peak on the chromatogram was at a lower volume than expected (being released from the SEC column sooner) for a protein with a molecular weight of 25 kDa (**Figure A.7**). Using a Superdex 75 10/300 GL SEC column, an estimate of molecular weight at 8.75 mL would be approximately 100 kDa.

In order to determine whether HP1α was in a dimeric, tetrameric, or other oligomeric state under the conditions being used for experiments in this work, firstly, a SDS-PAGE was performed.

In order to determine the oligomeric state of HP1α, one must be able to determine the size of the species formed by the purified protein. In its native form, the structure of a protein can dramatically affect its mobility through gel, either in a size exclusion column matrix or native polyacrylamide gel. Therefore, a cross-linking experiment was performed in order to bind together the proteins immediately adjacent to each other, and an SDS-PAGE was performed to estimate the molecular weights of the denatured cross-linked protein.

This technique was firstly performed on varying concentrations of HP1α in order to optimise the amount of protein needed to visualise all paralogs of HP1 (**Figure 4.2**). Sizes of protein standards were measured and graphed in order to determine sizes of HP1α bands (**Figure 4.3**). Estimations showed that the primary band

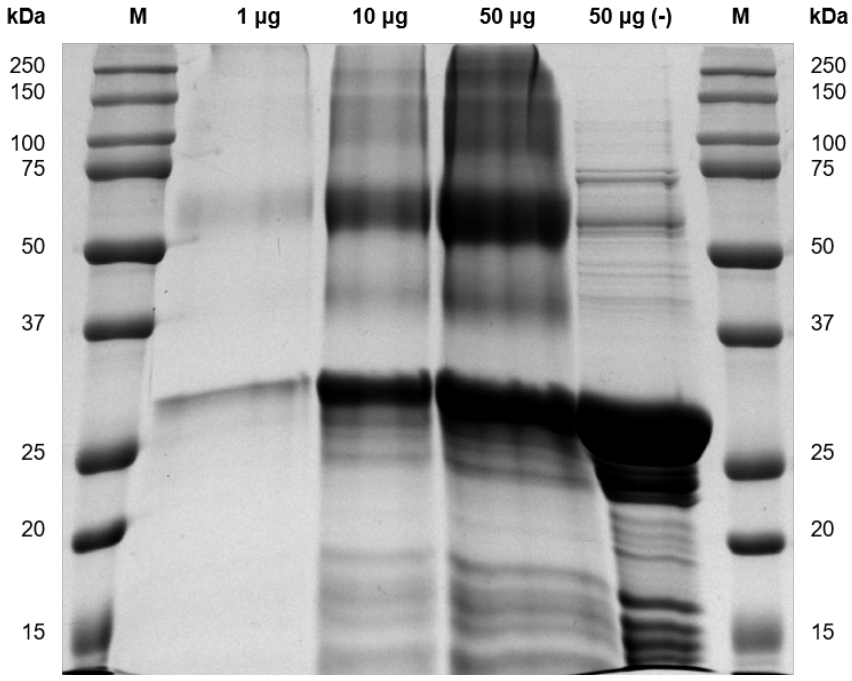

Figure 4.2: Cross-linked HP1α at varying concentrations separated by SDS-PAGE. Differing amounts of HP1α were cross-linked by formaldehyde, and separated by SDS-PAGE. The lane denoted with (-) indicates the control sample in which no formaldehyde was added to HP1α. M indicates marker lanes, with molecular weight of the protein standards adjacent.

was between 28.4 kDa and 32.8 kDa. Given that these values are overestimated due to the warping of the gel, this fits with the expected weight of monomeric his-tagged HP1α at 25.4 kDa (see **Section 2.7.1**). Another higher distinctive band was analysed, and its estimated molecular weight was 56.8 kDa, which is approximately double the expected molecular weight of HP1α, suggesting that this other dominant species was a dimer. There were also bands higher than this dimer species, showing that HP1α is forming higher oligomeric states. However, due to the inaccuracy of measuring bands above 75 kDa on this gel, their sizes cannot be estimated.

This experiment was then repeated with HP1α-3KA, HP1β, and HP1γ alongside HP1α. HP1α-3KA showed similar abundances of bands to HP1α, whereas HP1β and HP1γ showed that far fewer higher order structures were formed (**Figure 4.4**).

In order to confirm the presence of species expected due to the cross-linking experiment above, native PAGE was performed on all of the HP1 proteins. This showed that each WT paralog of HP1 existed as singular species (**Figure 4.5**),

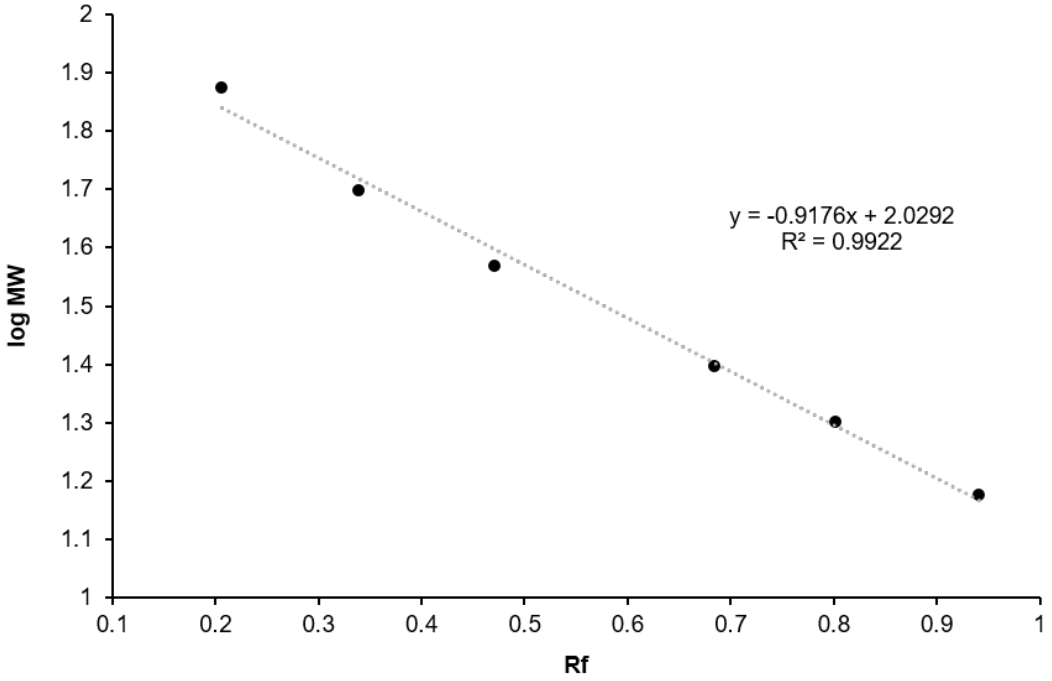

Figure 4.3: Standard curve for estimating cross-linked HP1α molecular weight. The logarithm of each of the molecular weight standards was calculated and plotted against the retention factor (Rf) of each of the standards bands. The slope of the line (displayed on the graph) was then used to calculate the molecular weight of HP1α using its Rf.

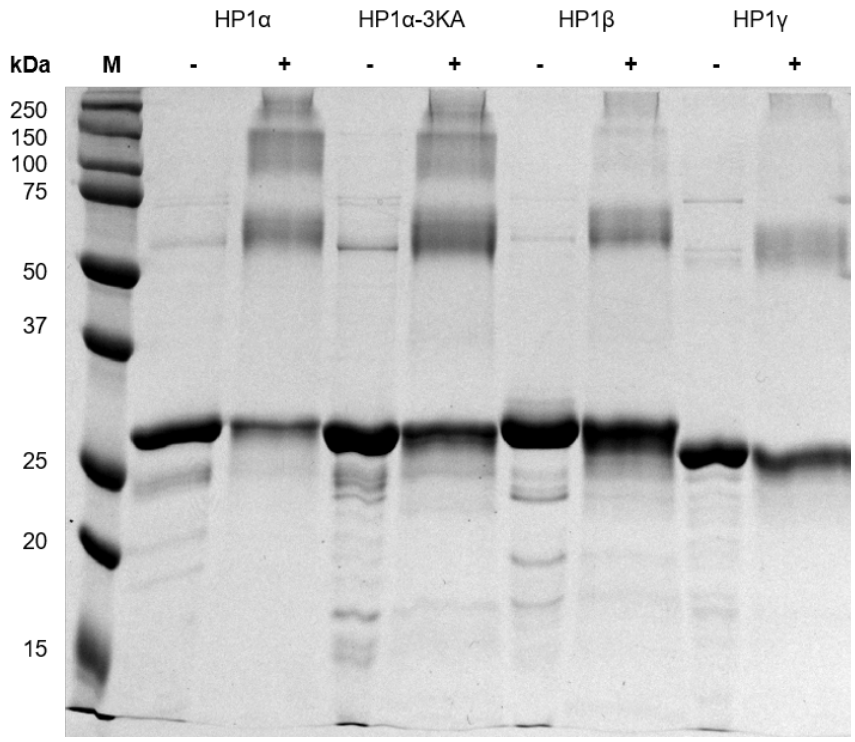

Figure 4.4: Cross-linked HP1 paralogs separated by SDS-PAGE. HP1 paralogs, HP1α, HP1β, and HP1γ, along with mutant HP1α-3KA, were cross-linked with formaldehyde, and separated by SDS-PAGE. M indicates marker lanes, with molecular weight of the protein standards adjacent, and "+" indicates addition of formaldehyde. This experiment was performed with a less pure sample preparation than that of other experiments.

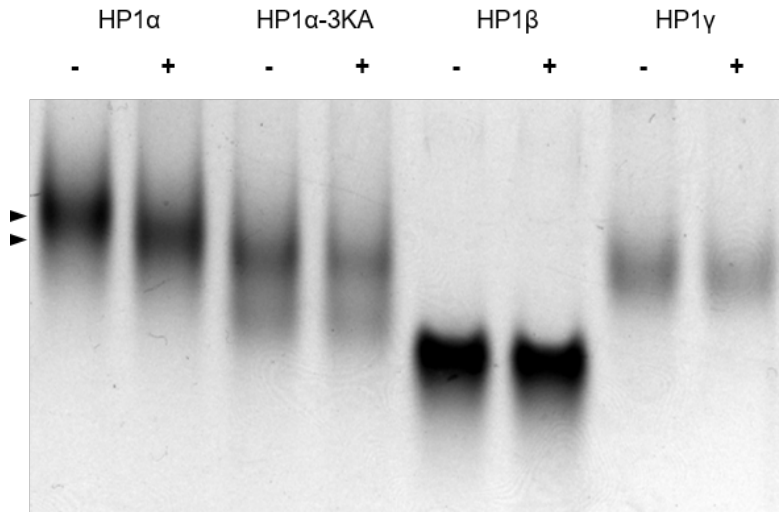

Figure 4.5: Native protein and complex mobility. PAGE performed under non-denaturing conditions of each HP1 paralog, and HP1α mutant, HP1α-3KA. Addition of TERRA 45 is indicated by a plus "+" symbol. Arrows indicate different mobilities of HP1α on its own (top), and HP1α with addition of TERRA 45 (bottom).

but due to the conditions of a native gel, it is not possible to estimate size accurately with this, and therefore it is difficult to deduce what oligomeric state the proteins are in. The mutant, HP1α-3KA however, showed two distinct species separated by the gel.

In order to see if addition and/or binding of TERRA 45 to HP1α had an effect on mobility of the protein, the RNA was added to each of the HP1 proteins. Surprisingly, the complex formed between HP1α and TERRA 45 had a higher mobility through the native gel than the protein alone (experiment repeat shown in **Figure A.8**). This implies that either the protein or the RNA (or both) change conformation upon binding, becoming more compact, increasing mobility of the complex. None of the other HP1 proteins had a change in mobility upon addition of TERRA 45, which is expected given that they have not been shown to bind strongly.

4.3 Affinity of the HP1α-TERRA interaction

To obtain a measurement for the kinetic properties of binding between HP1α and TERRA 45, immobilised HP1α was immersed in eight different concentrations

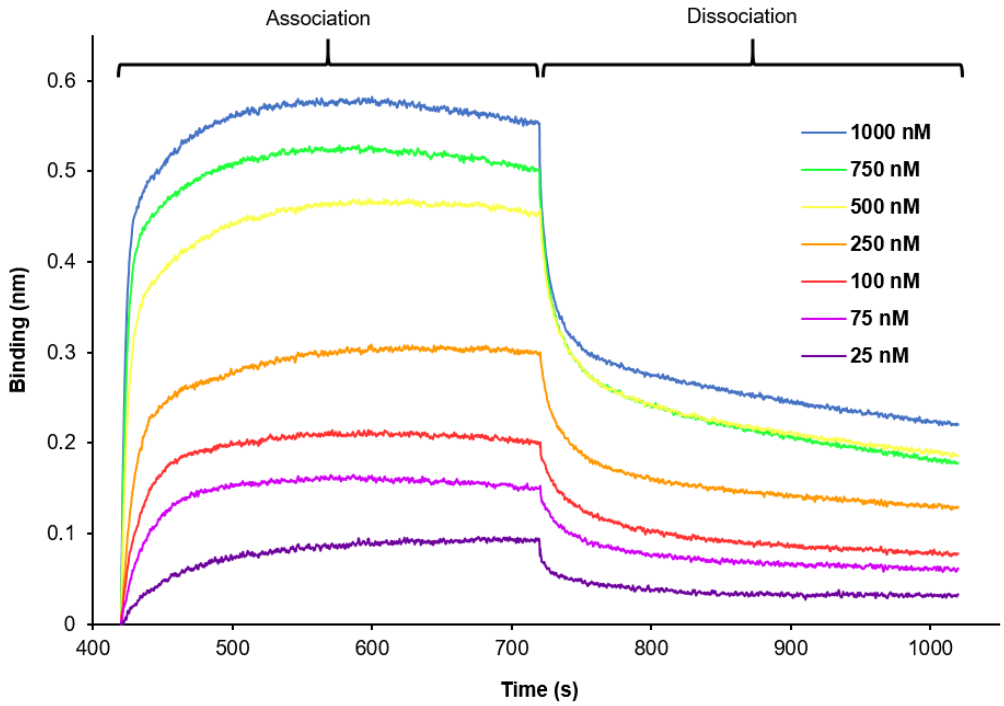

Figure 4.6: Titrating TERRA 45 with HP1α. BLI performed with HP1α immobilised on the tip, and addition of varying concentrations from 25 nM to 1000 nM of TERRA 45. These binding curves were used to calculate the equilibrium constant of the interaction.

of TERRA 45, and a dissociation equilibrium constant (K_D) was obtained using the multiple binding curves (global fit) from BLI with the BLItz Pro 1.2 software (**Figure 4.6**).

The K_D for the interaction between HP1α and TERRA 45 was in the nanomolar range (Association constant, k_a = 2.189 x 10^5 (+/- 7.836 x 10^3) $s^{-1}M^{-1}$; dissociation constant, k_d = 1.629 x 10^{-2} (+/- 4.446 x 10^{-4}) s^{-1}; equilibrium constant, K_D **= 74 nM**), showing that there was significant binding between HP1α and TERRA 45.

The experiment above was also performed with TERRA 96 in place of TERRA 45 (although with fewer data points, **Figure A.9**). The global fit of multiple binding curves produced another nanomolar K_D (Association constant, k_a = 1.593 x 10^5 (+/- 8.64 x 10^3) $s^{-1}M^{-1}$; dissociation constant, k_d = 3.96 x 10^{-4} (+/- 3.911 x 10^{-5}) s^{-1}; equilibrium constant, K_D **= 2.5 nM**), showing a stronger binding to TERRA 96 than to TERRA 45.

Although BLI is capable of outputting reasonably accurate KD data (Laigre

et al., 2018), it does not provide a complete picture of the thermodynamics of the interaction. ITC, instead of using changes in light to infer an interaction between species, uses changes in temperature. The change in temperature is directly linked to the binding or dissociation of molecules.

Following dialysis of TERRA 45 and HP1α, ITC was performed by injecting 2 μL (19 times) of 225 μM TERRA 45 into 280 μL of 23.7 μM HP1α. This yielded results shown in **Figure 4.7**, where it appears that within the first 2 μL injection, 50% of the binding sites on HP1α have already been bound by TERRA 45.

This yielded data showing that the interaction had a binding constant, K, of 1 x 10^5 M^{-1}. This is the reciprocal of the equilibrium constant, K_D, calculated by the BLItz software (see **Figure A.10**). Therefore, for comparison with BLI results, the reciprocal of K was calculated to give a K_D of 9.35 x 10^{-6} M. This value is unreliable, due to the incomplete sigmoidal curve present in this ITC attempt, the slope of which is used to calculate the binding constant, therefore ITC was repeated.

With attempt 2, HP1α concentration was increased (to 33 μM), and 1 μL injections (39 times) were used instead of 2 μL injections, so that the amount of TERRA 45 added each injection would be half of the amount as in attempt 1.

Attempt 2 also appeared to be unsuccessful in yielding a complete sigmoidal curve (**Figure 4.8**), again showing apparent saturation of 50% of the binding sites on HP1α.

ITC was attempted a third time, again, increasing the concentration of HP1α (to 44 μM), and keeping with 1 μL injections, decreasing the concentration of TERRA 45 to 109 μM.

ITC attempt 3 was also unsuccessful, this time yielding nothing resembling a sigmoidal curve (**Figure 4.9**), and low concentration of TERRA 45 compromised the signal to noise ratio.

It was then concluded that, due to the apparent complex nature of the HP1α-TERRA 45 interaction, given the mixed species present in TERRA 45, perhaps an accurate ITC measurement cannot be made until more is known about the specificity of the binding of HP1α to isolated monomer and dimer.

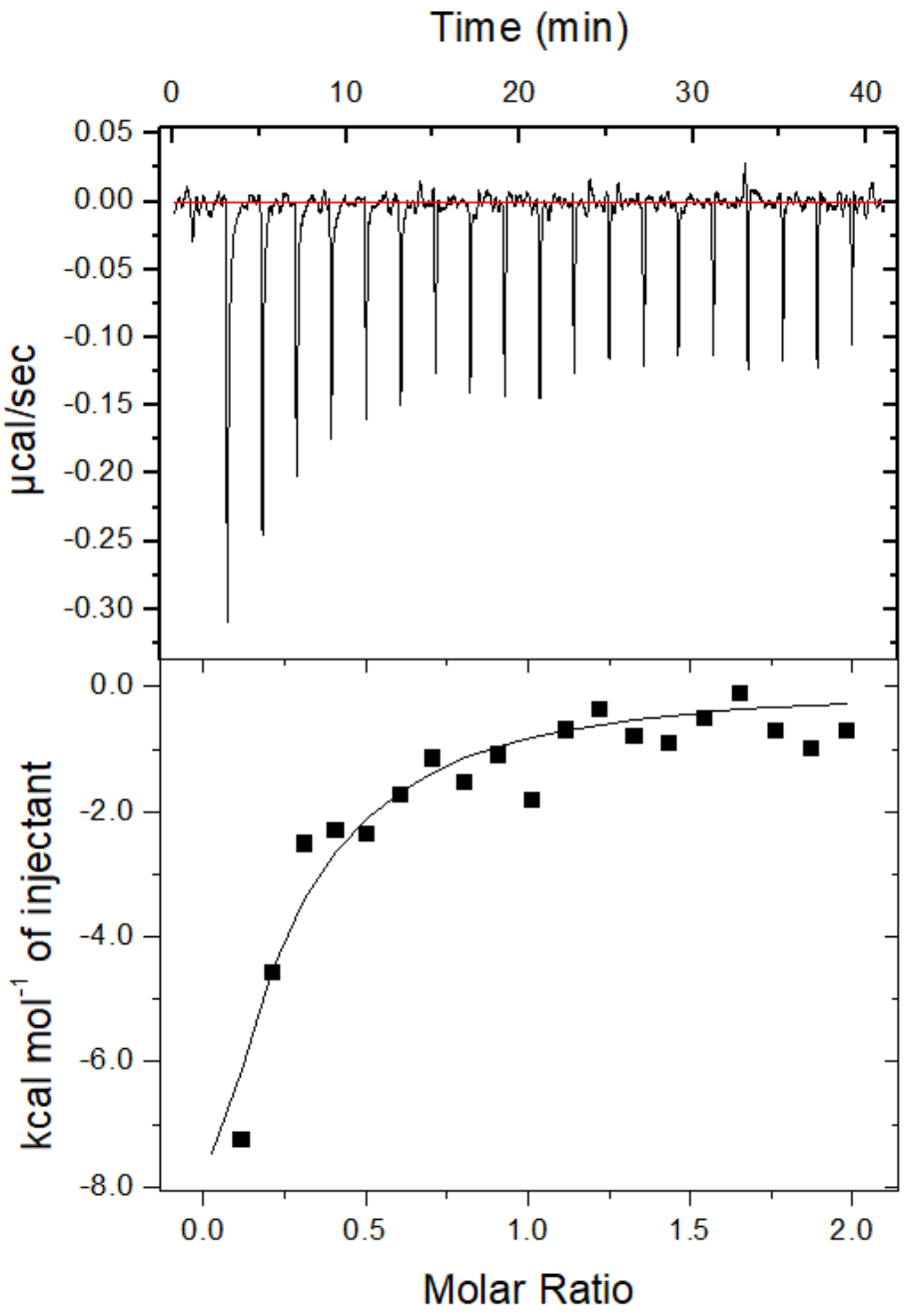

Figure 4.7: ITC attempt 1. ITC of HP1α and TERRA 45. The top graph displays raw data after the integration baseline has been subtracted. The bottom graph displays the normalised integration data. Lack of data at low molar ratio of TERRA 45 : HP1α limits interpretation to ΔH <0, and ΔS <0.

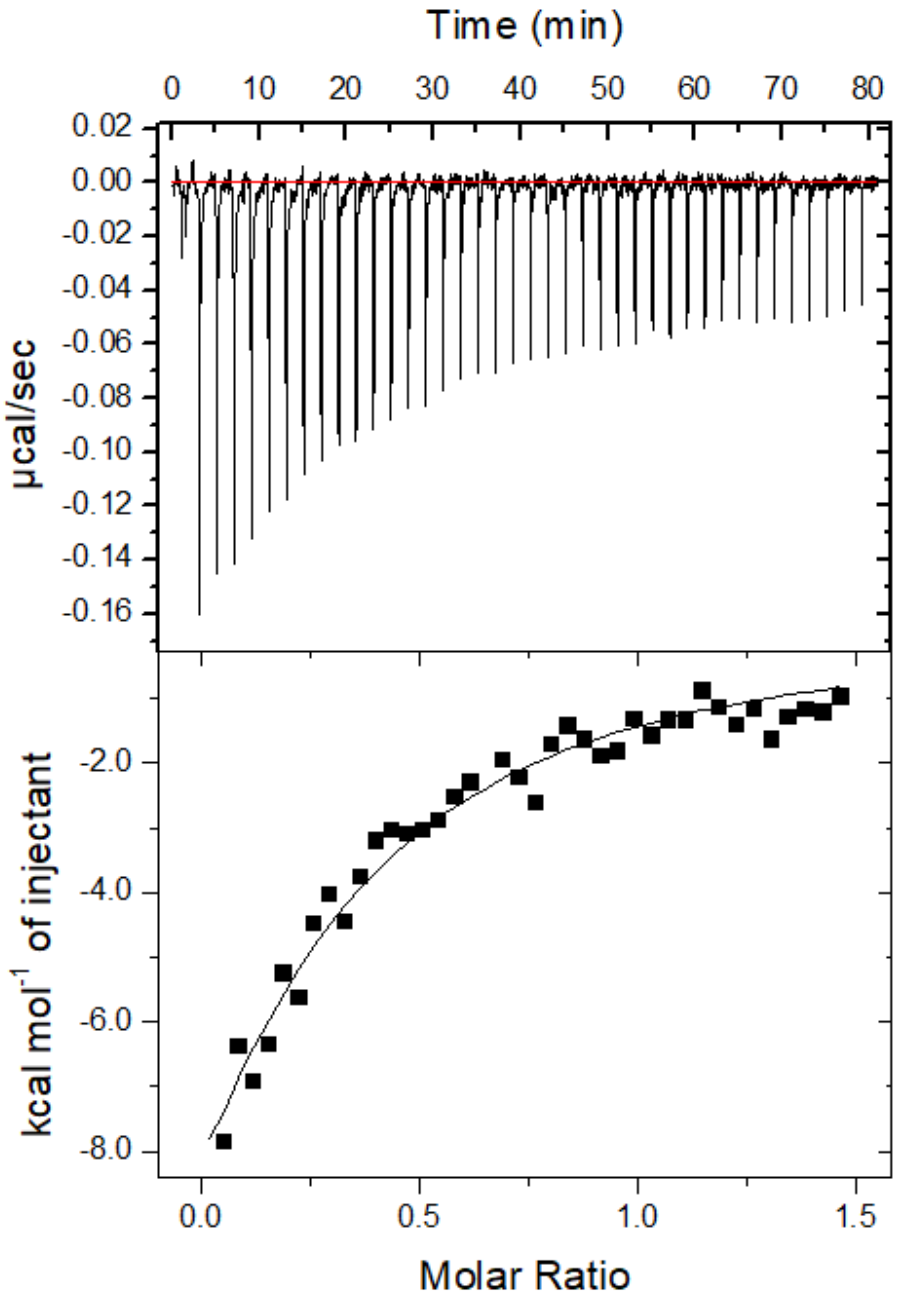

Figure 4.8: ITC attempt 2. ITC of HP1α and TERRA 45, with increased HP1α concentration, and decreased TERRA 45 injection volume. The top graph displays raw data after the integration baseline has been subtracted. The bottom graph displays the normalised integration data. Lack of data at low molar ratio of TERRA 45 : HP1α limits interpretation to ΔH <0, and ΔS <0.

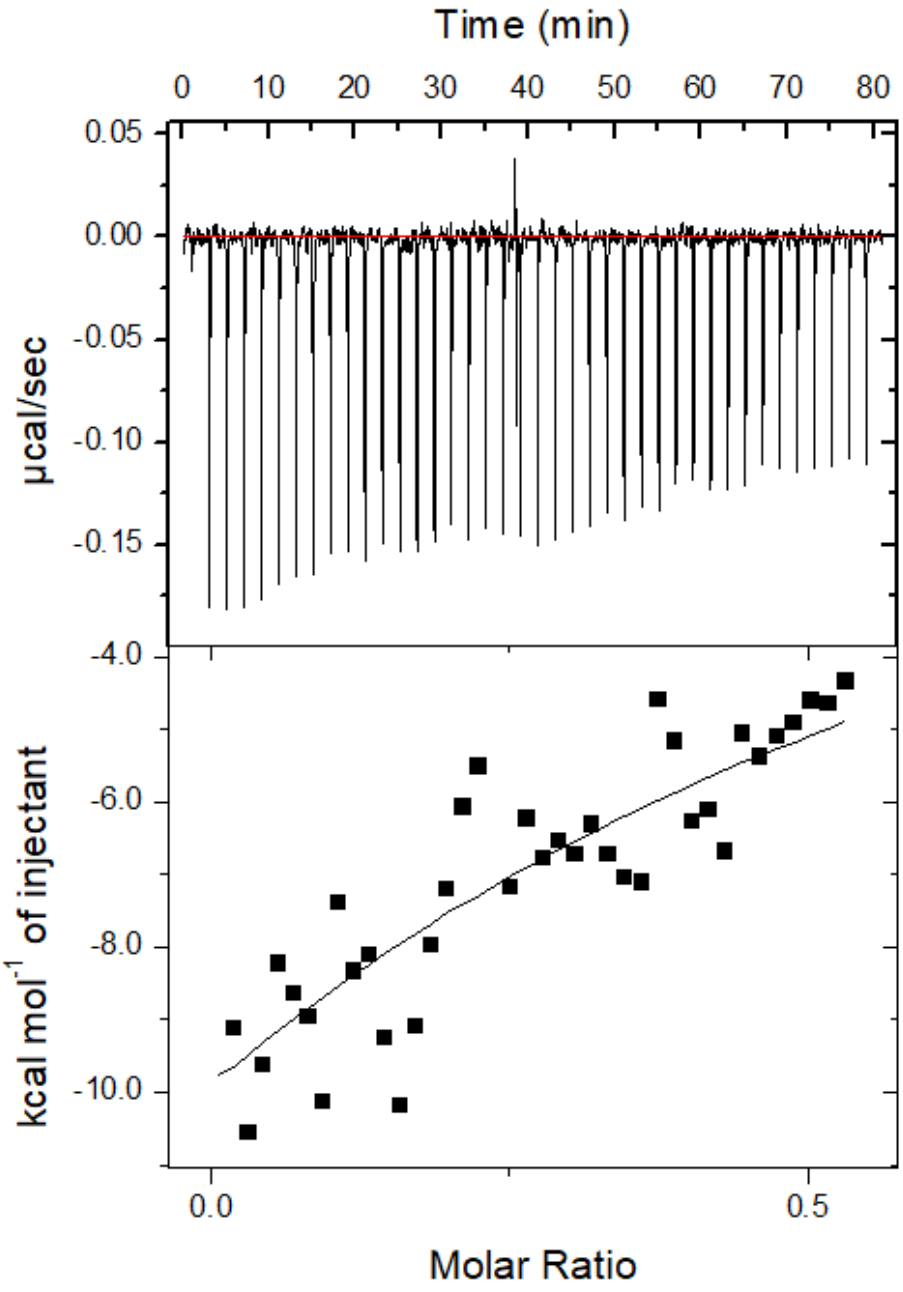

Figure 4.9: ITC attempt 3. ITC of HP1α and TERRA 45 with increased HP1α concentration, and decreased TERRA 45 concentration. The top graph displays raw data after the integration baseline has been subtracted. The bottom graph displays the normalised integration data. Lack of data at low molar ratio of TERRA 45 : HP1α limits interpretation to ΔH <0, and ΔS <0.

4.4 Small-angle X-ray scattering of HP1α and TERRA 45

SAXS was performed on each of the individual components of the interaction, HP1α and TERRA 45, and then also the complex of which they form. SAXS was also performed on HP1α-3KA in order to compare structural differences between the WT and mutant.

As shown by the double log graph (**Figure 4.10**), and the Kratky plot (**Figure 4.11**), TERRA 45 is in a more compact, ordered state. This is illustrated by the more horizontal line in the double log graph, and the parabolic curve shown in the Kratky plot.

By stark contrast, both HP1α and HP1α-3KA show the exact opposite, with a flexible, disordered state shown in the double log and Kratky plots.

The complex of HP1α and TERRA 45, shown on the Kratky plot, does appear to be slightly ordered, whereby the end of the curve does come back to around 0 (whereas the HP1α and HP1α-3KA curves do not return to 0). This may suggest a more compact structure of the complex compared with just HP1α alone.

TERRA 45 showed a small radius of gyration (R_g) value (**Table 4.1**), which is the quadratic mean of distances to the centre of the mass weighted by the contrast of electron density. This means that the mass of TERRA 45 takes up a relatively small space, meaning it is in a compact structure.

HP1α and HP1α-3KA each showed large R_g values relative to their molecular weights, with HP1α-3KA possessing a slightly more compact structure shown by the smaller R_g value.

The R_g of the complex was approximately the sum of the R_g values of TERRA 45 and HP1α.

Molecular weight estimations showed that the predominant species analysed of TERRA 45 was approximately 12,025 Da (**Table 4.2**), similar to the expected monomeric molecular weight of 14,814 Da. HP1α and HP1α-3KA molecular weight estimates were exactly the same at 53,150 kDa, approximately double the size of the monomeric form so most likely dimeric protein. The R_g values calculated from the distance distribution analysis also match up with the R_g

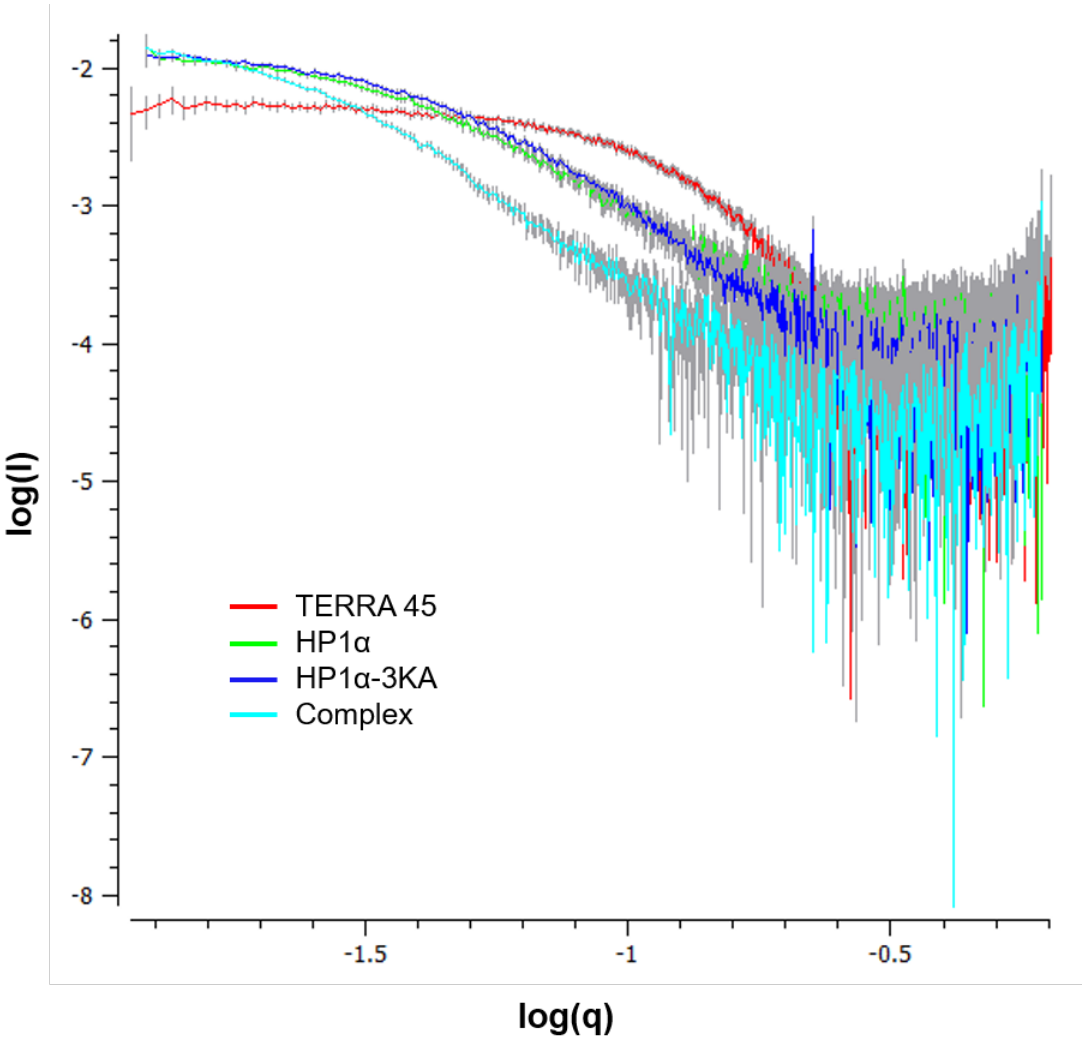

Figure 4.10: Raw scattering data plotted in log form. Raw SEC-SAXS scattering data of TERRA 45, HP1α, HP1α-3KA, and the complex between TERRA 45 and HP1α, plotted as the log of I (intensity), versus the log of q (momentum transfer).

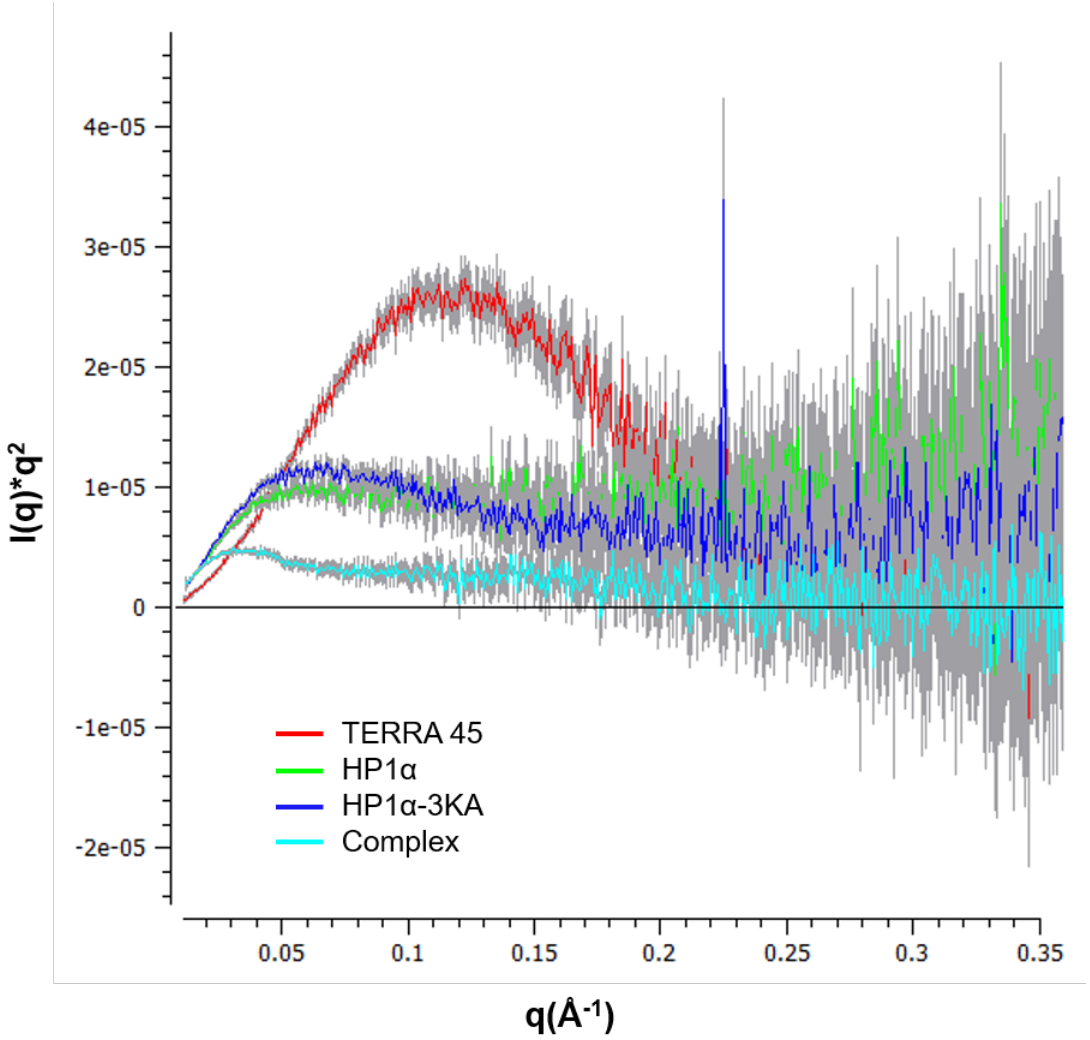

Figure 4.11: Raw scattering data presented as a Kratky plot. Raw SEC-SAXS scattering data of TERRA 45, HP1α, HP1α-3KA, and the complex between TERRA 45 and HP1α, plotted as a Kratky plot.

Table 4.1: SAXS data derived from Guinier plots (shown in appendix)

Name	C **(**μ**M)**	R_g **(+/-)**	I**(0)**	**Guinier range**
TERRA 45	200	15.43 +/- 0.27	0.0054 +/- 5.6e-05	5-100
HP1α	200	41.65 +/- 1.37	0.013 +/- 0.00029	2-27
HP1α-3KA	200	38.97 +/- 0.91	0.013 +/- 0.00021	1-30
Complex	100	64.75 +/- 3.12	0.016 +/- 0.00066	1-12

Table 4.2: SAXS data derived from molecular weight and distance distribution analysis

Name	**MW estimate**	R_g	D_{max}	**Porod**	**Porod volume**
TERRA 45	12025	15.66	56.95 (5-480)	16600	922.18 (5-861)
HP1α	53150	44.47	164 (2-450)	141000	1200 (2-826)
HP1α-3KA	53150	42.48	162.05 (4-480)	118000	2830 (1-826)
Complex	326450	70.74	244.10 (2-350)	735000	2710 (1-826)

values calculated by Guinier plot.

The molecular weight estimate of the complex was 326,450 Da, much larger than expected. This could mean that the complex has a more complicated stoichiometry than anticipated, or this structure is a large aggregate of protein or complexes. The R_g value is also agreeable with the R_g value found by Guinier analysis.

4.5 Discussion

When purifying HP1α for downstream applications, it was noted that on the SEC chromatogram (**Figure A.7**) the predominant peak (later confirmed to be HP1α by western blotting, **Figure A.15**) had a higher molecular weight estimate than expected. In the literature, many assume HP1α to be a dimer (Hinde et al., 2015), capable of homo-dimerisation with itself, or hetero-dimerisation with other HP1 paralogs (Ye et al., 1997). Others have stated the possibility of HP1 proteins forming higher order structures dependent on their PTMs (Brasher et al., 2000).

In the cross-linking experiments with HP1 proteins, higher bands on the gel indicated higher order structures such as trimers or tetramers being formed (**Figure 4.2**, **Figure 4.4**). Formaldehyde was used as the cross-linking agent in this

case, which is commonly used in immunoprecipitation experiments pertaining to chromatin interactions. Formaldehyde is used because of its small size, forming links between molecules only approximately 2 $\mathring{A}$ apart. However, its behaviour is highly dictated by its chemical environment, including pH and temperature. It is possible that the interactions shown in these experiments could be non-specific cross-linking of closely associating proteins, regardless of actual interaction between them.

It was thought that the native PAGE of the HP1 proteins might support the results shown in the cross-linking experiments; however, they somewhat contradicted one another. Each of the WT HP1 paralogs showed a singular species of indeterminate molecular weight (**Figure 4.5**). This supports what was stated above, in that the cross-links formed may have been proteins in close proximity rather than actually interacting together. Interestingly though, the HP1α-3KA protein formed two distinct species. The fact that changing only three amino acids, also shown to be the determinant of nucleic acid binding ability, also changed the ability of the protein to solely form the higher order structure is significant.

The higher band formed by HP1α-3KA seems more similar to the migration of the HP1α WT protein, but it also has a band of higher migration. This indicates that the band which migrates slower is perhaps a dimer, while the band which migrates faster may be a monomer. As the higher band of HP1α-3KA may be a dimer, given it has a similar mobility to the singular HP1α band, this may also be a dimer, assuming that the HP1α-3KA bands are not dimer and tetramer instead.

Recent work has shown that the apparent dimerisation of HP1α is not solely through the CSD as previously thought. Larson et al. (2017) showed by cross-linking MS of HP1α that the majority of links between HP1α homodimers was actually through the C-terminal extension binding with residues in the **hinge** of HP1α. Upon close inspection of their data, these links include the lysine residues at positions around 104-106, which have been modified in the HP1α-3KA mutant.

The stunted ability of HP1α-3KA to form the higher order structure that WT HP1α can may actually be a determinant of G4 binding ability, given the split of the protein at the higher order state and lower order state, and the incomplete abolishment of binding between HP1α-3KA and TERRA.

It is also noteworthy that HP1β has vastly higher mobility through the gel compared with the other HP1 paralogs, despite their similarity in molecular weight and sequence. It appears to migrate at a similar rate to the lower band of HP1α-3KA, indicating that it may only be present as a monomer, supporting the notion that it may be dimerisation ability that dictates HP1 binding to G4s. In contrast to this, is that HP1γ migrates similarly to HP1α, possibly forming a dimer. The amino acid sequences of the HP1 paralogs used in this work are shown in **Figure A.3**, which shows that the HP1γ sequence, compared to the others, has large gaps where HP1α does not, including a six residue region in the hinge. It would be interesting to see if these significant sequence differences are the regions which are also partly responsible for the binding capabilities of HP1α to G4s.

The gel electrophoresis experiments performed in this work gave an inconclusive result as to exactly what state each of the HP1 proteins was in (however it is likely to be the dimer/monomer scenario discussed above), and the large molecular weight estimates of HP1α could be due to its intrinsically disordered state and flexibility. Furthermore, these *in vitro* experiments may not be fully representative of how HP1α acts inside of the cell, showing the need for further investigation into the state of HP1α with *in vivo* experiments.

Native PAGE was also performed on the G4s used in this work (**Figure 4.1**). As stated in **Section 4.1**, the bands of lower mobility (higher position on the gel) correlated with binding ability to HP1α, and bands of higher mobility (lower position on the gel) were of anti-parallel G4 topology, which HP1α does not bind to significantly.

In vivo, TERRA has been shown to exist at lengths up to 9000 bp (Azzalin et al., 2007), possessing assumably very complex intramolecular G4 interactions. It is possible that the parallel higher order structures shown on the gel that bind HP1α are more similar to the complex structure of TERRA in the cell, accounting for HP1α's preference for these oligonucleotides. More work needs to be done to elucidate the *in vivo* structures of G4s shown to interact with proteins like HP1α, because in isolated *in vitro* experiments, the behaviour and state of the G4s they interact with are much more complex within the cell. For example, G4s can be formed with four strands from different molecules, and/or can be DNA:RNA hybrids (Zhang et al., 2014). The complexity of the G4s formed in the cell would be extremely hard to replicate, again, calling for more interaction data sought from *in vivo* experiments.

With the results obtained from the native PAGE in this study, the determination of molecular weights is not reliable, due to the secondary structure influence. There is a need to implement methods such as mass spectrometry (MS) in order to accurately measure the native mass and dimerisation ability of the protein and oligonucleotides.

The BLItz Pro software calculates K_D values by fitting binding and dissociation curves to the data provided, and using these curves, calculates the association rate constant (k_a) and dissociation rate constant (k_d) of the interaction. The k_a and k_d are then used to calculate the K_D, which indicates the concentration of oligonucleotide at which 50% of protein binding sites are occupied at equilibrium, or the concentration at which the number of protein molecules with oligonucleotide bound equals the number of protein molecules without oligonucleotide bound. In these calculations, only the concentration of oligonucleotide is taken into account, and the binding is assumed to be 1:1.

The binding strength between HP1α and TERRA 45 was shown to be 74 nM measured by BLI (**Figure 4.6**), meaning that only 74 nM of RNA was required to occupy 50% of the binding sites of protein bound (4 μL of 100 μg/mL protein) to the tip. This nanomolar range K_D is much stronger than the binding shown between the chromodomain of HP1α and H3K9me3, which suggests a very high significance of this interaction. For a long time, the simple model of heterochromatin formation has been that HP1α binds H3K9me3, recruiting methyltransferases to methylate more H3K9, resulting in spreading of condensed heterochromatin. This model is basic, and like with many things that occur in a living cell, the reality is that it's very complex. The finding that HP1α binds TERRA, a G4 present in a region of constitutive heterochromatin and is transcribed from, and present at telomeres, strongly suggests a role for G4-dependent recruitment of HP1α to regions of heterochromatin.

Due to BLI technology being a very niche and novel technique and the simplicity of the binding model that it computes, the significance of this interaction needs to be confirmed using an alternative technique, which is why ITC was used in an attempt to measure the thermodynamic properties of the interaction.

For an accurate K_D to be determined using ITC, a typical sigmoidal curve is used to estimate the point at which half of the binding sites on the protein are occupied by RNA. In order to accurately estimate this point, it is best if the

whole sigmoidal curve is visible in order to measure the slope and inflection point. This could not be achieved during these attempts to test the HP1α-TERRA 45 interaction (**Figure 4.7**, **Figure 4.8**, **Figure 4.9**).

Wiseman et al. (1989) give an equation to calculate a c-value (sample concentration, affinity-related parameter), which should be between 1 and 1000 (preferably 10 to 500) for optimal ITC results. The c-values given by the HP1α concentrations and K_D (calculated by BLI) assuming a 1:1 interaction are 338, 471, and 628 for attempts 1, 2, and 3, respectively. These are all within the recommended c-value range, which would typically give reliable results, assuming the concentration of RNA is high enough. Therefore, this would indicate that the HP1α-TERRA 45 interaction is more complex than originally thought, thwarting ITC attempts.

Given the above results investigating the oligomerisation of HP1α, which is most likely a dimer, and TERRA 45, which is present as both a dimer and a monomer, this interaction may be too complex to study using ITC. This notion is also supported by the large molecular weight estimate of the complex found by SAXS (**Table 4.2**). The dimer and monomer of TERRA 45 either need to be isolated and stabilised in those forms to show which species HP1α binds to, or substrates mimicking these monomers and dimers need to be constructed for this purpose. Only then will ITC be able to determine the thermodynamic properties of the interaction, where each species is homogeneous.

The SAXS data obtained for these experiments was produced using SEC-SAXS, whereby sample is exposed to the X-ray beam following elution by the Co-flow system, separating species based on native size. Because of this, data must first be selected from a chromatogram, separating out the appropriate peak to study. In each case, data were selected of the predominant peak, with stable R_g values (example shown in **Figure A.16**). In cases such as the complex formation, protein and complex being eluted at similar times by SEC made it difficult to isolate signal originating from the complex alone, leading to possibly skewed results. In future, a column capable of separating out these distinct species will need to be taken to and fitted onto the SEC-SAXS system in order to properly isolate complex signal.

The SAXS data obtained on TERRA 45 show it is a relatively small, very ordered structure (**Figure 4.11**, **Table 4.1**). This is as expected for a molecule formed

into a G4, which thermodynamically is very stable.

SAXS performed on HP1α and HP1α-3KA showed very disordered structures, of which HP1α-3KA may be slightly more ordered. This was also expected, due to literature alluding to an "unstructured hinge domain", and the total lack of crystal structures of HP1α (only the chromodomain and CSD have been crystallised).

In future SAXS experiments, to more accurately determine the oligomeric state, and also the formation of aggregates, a dilution series of HP1α should be performed. This could also determine if there are other significant concentration effects, such as inter-particle interactions.

Previous SAXS data has been obtained on HP1α at a similar concentration, which showed that a WT HP1α dimer had a D_{max} of 130 $\mathring{A}$, however, when phosphorylated on its N-terminal extension had a D_{max} of 220 $\mathring{A}$. The D_{max} obtained in this work was 164 $\mathring{A}$ (**Table 4.2**), in between the values obtained by Larson et al. (2017). Again, this discrepancy could be due to PTMs deposited on HP1α during its bacterial expression, highlighting the need for MS to be done to determine presence of PTMs.

The SAXS data measured on the complex was at a lower concentration to that of its components alone. Therefore, the signal obtained was weaker than that of HP1α and TERRA 45. If performed at a higher concentration, this may have made it easier to determine more accurately how ordered the complex appeared to be on the Kratky plot.

Overall, these results show the complexity of the HP1α-TERRA interaction, the challenges that are faced in a unique interaction with many possible states of binding partners. This shows the requirement for further *in vitro* work to elucidate the mechanism by which HP1α and TERRA bind, and also *in vivo* work, putting this research at the forefront of discovering just how involved G4 structures are in the genome.

Chapter 5

Conclusions and Future Work

Firstly, this work showed that the binding to parallel G-quadruplex (G4) telomeric repeat-containing RNA (TERRA) molecules was exclusive to Heterochromatin Protein 1α (HP1α), the primary HP1 paralog associated with constitutive heterochromatin formation. Upon mutation of three residues in the hinge region of HP1α from lysines to alanines, which have been shown to bind RNA and subsequently target HP1 to centromeres (Maison et al., 2011), binding to TERRA was extremely abrogated.

It was found that HP1α did not bind telomeric DNA, which formed anti-parallel G4 topology. Upon modifying the sequence of the telomeric DNA in order to change its topology, a strong preference for only parallel G4 structures was shown by HP1α.

HP1α was also able to bind a range of DNA G4s, with preference for parallel topology, that are formed in regulatory regions of genes associated with cancer, and yet HP1α did not show any binding to i-motif structures of endogenous sequence, tRNA, duplex DNA, or unstructured DNA.

It was then found that the binding between HP1α and TERRA 45 was 74 nM, and binding between HP1α and TERRA 96 was 2.5 nM, both of these values showing strong binding, however, this was unable to be confirmed by isothermal titration calorimetry (ITC).

This work showed a pattern of binding between HP1α and parallel G4 oligonucleotides capable of forming higher order structures, and also showed that the

complex formed between HP1α and TERRA 45 has higher mobility, and is possibly more rigid and compact than just HP1α alone.

This characterisation of the HP1α-TERRA interaction highlights its intricacy and specificity, owing to its importance in heterochromatin formation. This work has led to a new model of heterochromatin formation by lncRNA recruitment at the telomeres (**Figure 5.1**), whereby TERRA is localised to telomeres, and forms G4s that provide a specific docking site for HP1α to bind through its unstructured hinge region. This provides unique specificity to heterochromatin formation at telomeres, a vulnerable region of the genome.

The next steps in this research include the following. Performing mass spectrometry (MS) under native conditions with whole protein would confirm the oligomeric state of HP1α. MS of digested HP1α would also determine if bacterial expression had deposited any PTMs, potentially affecting the binding between HP1α and TERRA. Further to this, other PTMs from the literature, for example phosphorylation (on N-terminal serine residues, Hiragami-Hamada et al. (2011)) and sumoylation (on the hinge, (Maison et al., 2011) of specific residues would need to be pursued in order to test if these have any effect on the HP1α-TERRA interaction, potentially altering the targeting of HP1α if binding to TERRA is abrogated by them.

Confirmation of the effect of the HP1α-TERRA interaction on chromatin condensation then needs to be done through the studies of nucleosomal arrays. HP1α and TERRA could be tested for their requirement for the formation of heterochromatin through scrutinising nucleosomal arrays through analytical ultracentrifugation, SAXS, and SANS. The compaction of the nucleosomal array can be assessed using SAXS and SANS, and the density of the array can be assessed using analytical ultracentrifugation.

Following all of these *in vitro* techniques, the consequences of this interaction need to be confirmed *in vivo*. This could firstly be done using induced expression of tagged HP1α or HP1α-3KA (or other mutants) in mammalian cell lines. A ChIP assay could be performed, whereby tagged HP1α and associated RNA or DNA will be immunoprecipitated, and co-immunoprecipitated G4s can be detected using commercially available G4 antibodies. Further to this, the structure and proteins present on telomeres can be assessed by confocal microscopy (using antibodies for HP1α and other proteins of interest) using the cell lines discussed

above. Telomere structure can be assessed in WT and mutant HP1α cell lines, and co-localisation of TERRA or other telomeric associated molecules may be assessed using immunofluorescence.

The work discussed in this thesis, together with the experiments mentioned above, provide a methodology to study the role of TERRA and HP1α in the formation and protection of telomeric heterochromatin. This interaction is pivotal to understanding telomere protection and regulation, while also bringing the function of non-canonical nucleic acid structures in the genome into light.

High affinity binding between a pivotal chromatin regulator and a G4 molecule has potential in drug design, relating to G4s associated with cancer, such as the endogenous oligonucleotides used in this work. The interaction between HP1α and TERRA is also a large target for drug design, given the roles of TERRA in telomerase function and ALT in regulating telomere length in cancer cells.

The central dogma of molecular biology dictates that DNA and subsequent transcription into RNA are the only functions of nucleic acids, however, following the discoveries of non-coding RNAs in gene regulation, the roles of these highly structured RNAs are now an extra level to our understanding of gene regulation and the epigenome. And now, this work can further transform our understanding of biology, showing formerly overlooked non-canonical nucleic acid structures are potentially crucial to the function our genome.

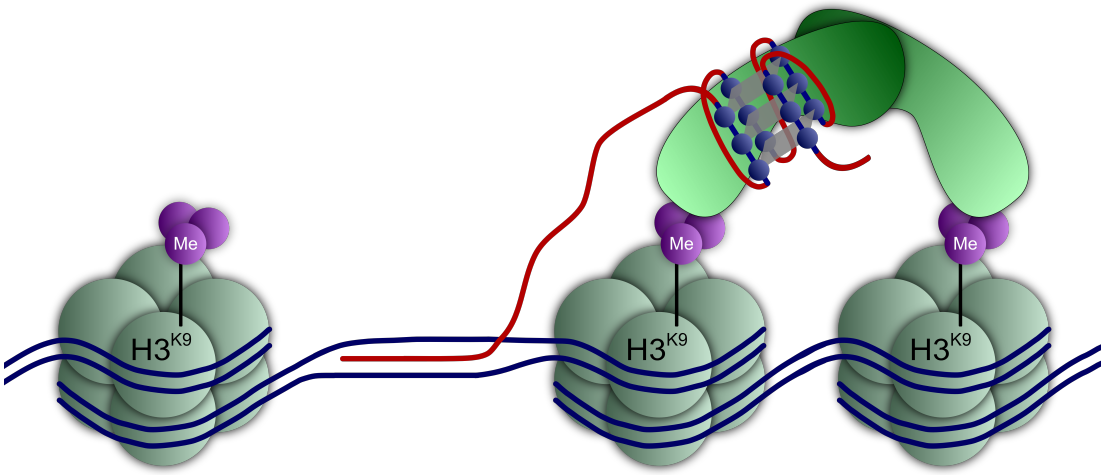

Figure 5.1: Model of TERRA-dependent targeting of HP1α to telomeres. TERRA is present at the telomeres, either through its nascent transcription, or the formation of DNA:RNA hybrids with the telomeric repeats. The formation of G4s by TERRA allows for HP1α to be recruited, then binding H3K9me3 and subsequently dimerising in order to compact chromatin. Blue strands indicate DNA, red indicates RNA.

Bibliography

Agrawal, P., Lin, C., Mathad, R. I., Carver, M., and Yang, D. (2014). The major g-quadruplex formed in the human bcl-2 proximal promoter adopts a parallel structure with a 13-nt loop in k+ solution. *Journal of the American Chemical Society*, 136(5):1750–1753.

Allshire, R. C., Dempster, M., and Hastie, N. D. (1989). Human telomeres contain at least three types of g–rich repeat distributed non-randomly. *Nucleic acids research*, 17(12):4611–4627.

Arnoult, N., Van Beneden, A., and Decottignies, A. (2012). Telomere length regulates TERRA levels through increased trimethylation of telomeric H3K9 and HP1. *Nat. Struct. Mol. Biol.*, 19(9):948–956.

Arora, R. and Azzalin, C. M. (2015). Telomere elongation chooses TERRA ALTernatives. *RNA Biol*, 12(9):938–941.

Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E., and Lingner, J. (2007). Telomeric repeat–containing rna and rna surveillance factors at mammalian chromosome ends. *Science*, 318(5851):798–801.

Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*, 410(6824):120–124.

Barthel, F. P., Wei, W., Tang, M., Martinez-Ledesma, E., Hu, X., Amin, S. B., Akdemir, K. C., Seth, S., Song, X., Wang, Q., Lichtenberg, T., Hu, J., Zhang, J., Zheng, S., and Verhaak, R. G. (2017). Systematic analysis of telomere length and somatic alterations in 31 cancer types. *Nat. Genet.*, 49(3):349–357.

Becker, J. S., Nicetto, D., and Zaret, K. S. (2016). H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends Genet.*, 32(1):29–41.

Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013). Quantitative visualization of dna g-quadruplex structures in human cells. *Nature chemistry*, 5(3):182.

Blackburn, E. H., Greider, C. W., Henderson, E., Lee, M. S., Shampay, J., and Shippen-Lentz, D. (1989). Recognition and elongation of telomeres by telomerase. *Genome*, 31(2):553–560.

Bowman, G. D. and Poirier, M. G. (2015). Post-translational modifications of histones that influence nucleosome dynamics. *Chem. Rev.*, 115(6):2274–2295.

Brasher, S. V., Smith, B. O., Fogh, R. H., Nietlispach, D., Thiru, A., Nielsen, P. R., Broadhurst, R. W., Ball, L. J., Murzina, N. V., and Laue, E. D. (2000). The structure of mouse hp1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer. *The EMBO journal*, 19(7):1587–1597.

Brázda, V., Hároníková, L., Liao, J., and Fojta, M. (2014). Dna and rna quadruplex-binding proteins. *International journal of molecular sciences*, 15(10):17493–17517.

Brown, S. W. (1966). Heterochromatin. *Science*, 151(3709):417–425.

Bryan, T. M., Englezou, A., Dalla-Pozza, L., Dunham, M. A., and Reddel, R. R. (1997). Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nat. Med.*, 3(11):1271–1274.

Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., and Neidle, S. (2006). Quadruplex dna: sequence, topology and structure. *Nucleic acids research*, 34(19):5402–5415.

Canudas, S., Houghtaling, B. R., Bhanot, M., Sasa, G., Savage, S. A., Bertuch, A. A., and Smith, S. (2011). A role for heterochromatin protein 1γ at human telomeres. *Genes & development*, 25(17):1807–1819.

Canzio, D., Larson, A., and Narlikar, G. J. (2014). Mechanisms of functional promiscuity by hp1 proteins. *Trends in cell biology*, 24(6):377–386.

Cesare, A. J. and Reddel, R. R. (2010). Alternative lengthening of telomeres: models, mechanisms and implications. *Nat. Rev. Genet.*, 11(5):319–330.

Chambers, V. S., Marsico, G., Boutell, J. M., Di Antonio, M., Smith, G. P., and Balasubramanian, S. (2015). High-throughput sequencing of dna g-quadruplex structures in the human genome. *Nature biotechnology*, 33(8):877.

Cogoi, S. and Xodo, L. E. (2006). G-quadruplex formation within the promoter of the kras proto-oncogene and its effect on transcription. *Nucleic acids research*, 34(9):2536–2549.

Collie, G. W., Haider, S. M., Neidle, S., and Parkinson, G. N. (2010). A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res.*, 38(16):5569–5580.

Cong, Y.-S., Wright, W. E., and Shay, J. W. (2002). Human telomerase and its regulation. *Microbiol. Mol. Biol. Rev.*, 66(3):407–425.

Consortium, I. H. G. S. et al. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860.

Creacy, S. D., Routh, E. D., Iwamoto, F., Nagamine, Y., Akman, S. A., and Vaughn, J. P. (2008). G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates. *J. Biol. Chem.*, 283(50):34626–34634.

Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W., and Richmond, T. J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 å resolution. *Journal of molecular biology*, 319(5):1097–1113.

David, G., Turner, G. M., Yao, Y., Protopopov, A., and DePinho, R. A. (2003). msin3-associated protein, msds3, is essential for pericentric heterochromatin formation and chromosome segregation in mammalian cells. *Genes & development*, 17(19):2396–2405.

De Lange, T. (2005). Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes & development*, 19(18):2100–2110.

Deng, Z., Norseen, J., Wiedmer, A., Riethman, H., and Lieberman, P. M. (2009). Terra rna binding to trf2 facilitates heterochromatin formation and orc recruitment at telomeres. *Molecular cell*, 35(4):403–413.

di Fagagna, F. d., Reaper, P. M., Clay-Farrace, L., Fiegler, H., Carr, P., von Zglinicki, T., Saretzki, G., Carter, N. P., and Jackson, S. P. (2003). A dna damage checkpoint response in telomere-initiated senescence. *Nature*, 426(6963):194.

Diotti, R. and Loayza, D. (2011). Shelterin complex and associated factors at human telomeres. *Nucleus*, 2(2):119–135.

Diveshkumar, K. V., Sakrikar, S., Rosu, F., Harikrishna, S., Gabelica, V., and Pradeepkumar, P. (2016). Specific stabilization of c-myc and c-kit g-quadruplex dna structures by indolylmethyleneindanone scaffolds. *Biochemistry*, 55(25):3571–3585.

Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R. R., and Richmond, T. J. (2004). Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science*, 306(5701):1571–1573.

Fay, M. M., Lyons, S. M., and Ivanov, P. (2017). Rna g-quadruplexes in biology: principles and molecular mechanisms. *Journal of molecular biology*, 429(14):2127–2147.

Feretzaki, M. and Lingner, J. (2017). A practical qpcr approach to detect terra, the elusive telomeric repeat-containing rna. *Methods*, 114:39–45.

Flynn, R. L., Cox, K. E., Jeitany, M., Wakimoto, H., Bryll, A. R., Ganem, N. J., Bersani, F., Pineda, J. R., Suva, M. L., Benes, C. H., Haber, D. A., Boussin, F. D., and Zou, L. (2015). Alternative lengthening of telomeres renders cancer cells hypersensitive to ATR inhibitors. *Science*, 347(6219):273–277.

Garavis, M., Bocanegra, R., Herrero-Galan, E., Gonzalez, C., Villasante, A., and Arias-Gonzalez, J. R. (2013). Mechanical unfolding of long human telomeric RNA (TERRA). *Chem. Commun. (Camb.)*, 49(57):6397–6399.

Gehring, K., Leroy, J.-L., and Guéron, M. (1993). A tetrameric dna structure with protonated cytosine-cytosine base pairs. *Nature*, 363(6429):561.

Gellert, M., Lipsett, M. N., and Davies, D. R. (1962). Helix formation by guanylic acid. *Proceedings of the National Academy of Sciences of the United States of America*, 48(12):2013.

Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H., and De Lange, T. (1999). Mammalian telomeres end in a large duplex loop. *Cell*, 97(4):503–514.

Hakin-Smith, V., Jellinek, D. A., Levy, D., Carroll, T., Teo, M., Timperley, W. R., McKay, M. J., Reddel, R. R., and Royds, J. A. (2003). Alternative lengthening

of telomeres and survival in patients with glioblastoma multiforme. *Lancet*, 361(9360):836–838.

Harley, C. B., Futcher, A. B., and Greider, C. W. (1990). Telomeres shorten during ageing of human fibroblasts. *Nature*, 345(6274):458.

Hayflick, L. and Moorhead, P. S. (1961). The serial cultivation of human diploid cell strains. *Experimental cell research*, 25(3):585–621.

Hazel, P., Huppert, J., Balasubramanian, S., and Neidle, S. (2004). Loop-length-dependent folding of g-quadruplexes. *Journal of the American Chemical Society*, 126(50):16405–16415.

Hinde, E., Cardarelli, F., and Gratton, E. (2015). Spatiotemporal regulation of heterochromatin protein 1-alpha oligomerization and dynamics in live cells. *Scientific reports*, 5:12001.

Hiragami-Hamada, K., Shinmyozu, K., Hamada, D., Tatsu, Y., Uegaki, K., Fujiwara, S., and Nakayama, J.-i. (2011). N-terminal phosphorylation of hp1α promotes its chromatin binding. *Molecular and cellular biology*, 31(6):1186–1200.

Huppert, J. L. and Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. *Nucleic acids research*, 33(9):2908–2916.

Jamieson, K., Wiles, E. T., McNaught, K. J., Sidoli, S., Leggett, N., Shao, Y., Garcia, B. A., and Selker, E. U. (2016). Loss of HP1 causes depletion of H3K27me3 from facultative heterochromatin and gain of H3K27me2 at constitutive heterochromatin. *Genome Res.*, 26(1):97–107.

Janzon, J. (2013). Bio-layer interferometry (bli). *Bioprocess Vision*, 5.

Jenuwein, T. and Allis, C. D. (2001). Translating the histone code. *Science*, 293(5532):1074–1080.

Jiang, W.-Q., Nguyen, A., Cao, Y., Chang, A. C.-M., and Reddel, R. R. (2011). Hp1-mediated formation of alternative lengthening of telomeres-associated pml bodies requires hira but not asf1a. *PLoS One*, 6(2):e17036.

Karsisiotis, A. I., Hessari, N. M., Novellino, E., Spada, G. P., Randazzo, A., and da Silva, M. W. (2011). Topological characterization of nucleic acid g-quadruplexes by uv absorption and circular dichroism. *Angewandte Chemie*, 123(45):10833–10836.

Kaushik, M., Kaushik, S., Roy, K., Singh, A., Mahendru, S., Kumar, M., Chaudhary, S., Ahmed, S., and Kukreti, S. (2016). A bouquet of dna structures: Emerging diversity. *Biochemistry and biophysics reports*, 5:388–395.

Kikin, O., D'Antonio, L., and Bagga, P. S. (2006). Qgrs mapper: a web-based server for predicting g-quadruplexes in nucleotide sequences. *Nucleic acids research*, 34(suppl_2):W676–W682.

Kilic, S., Bachmann, A. L., Bryan, L. C., and Fierz, B. (2015). Multivalency governs hp1α association dynamics with the silent chromatin state. *Nature communications*, 6:7313.

Klockenbusch, C. and Kast, J. (2010). Optimization of formaldehyde cross-linking for protein interaction analysis of non-tagged integrin beta1. *J. Biomed. Biotechnol.*, 2010:927585.

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H., and Svergun, D. I. (2003). Primus: a windows pc-based system for small-angle scattering data analysis. *Journal of applied crystallography*, 36(5):1277–1282.

Kwon, S. H. and Workman, J. L. (2011). The changing faces of hp1: from heterochromatin formation and gene silencing to euchromatic gene expression. *Bioessays*, 33(4):280–289.

Laigre, E., Goyard, D., Tiertant, C., Dejeu, J., and Renaudet, O. (2018). The study of multivalent carbohydrateprotein interactions by bio-layer interferometry. *Organic and Biomolecular Chemistry*, 16.

Lane, A. N., Chaires, J. B., Gray, R. D., and Trent, J. O. (2008). Stability and kinetics of g-quadruplex structures. *Nucleic acids research*, 36(17):5482–5515.

Larson, A. G., Elnatan, D., Keenen, M. M., Trnka, M. J., Johnston, J. B., Burlingame, A. L., Agard, D. A., Redding, S., and Narlikar, G. J. (2017). Liquid droplet formation by hp1α suggests a role for phase separation in heterochromatin. *Nature*, 547(7662):236.

Levene, P. A. and Bass, L. W. (1931). *Nucleic acids*. Chemical Catalog Company, Inc.; New York.

Lingner, J., Hughes, T. R., Shevchenko, A., Mann, M., Lundblad, V., and Cech, T. R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science*, 276(5312):561–567.

Longhese, M. P. (2008). Dna damage response at functional and dysfunctional telomeres. *Genes & development*, 22(2):125–140.

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 å resolution. *Nature*, 389(6648):251.

Maison, C., Bailly, D., Peters, A. H., Quivy, J.-P., Roche, D., Taddei, A., Lachner, M., Jenuwein, T., and Almouzni, G. (2002). Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an rna component. *Nature genetics*, 30(3):329.

Maison, C., Bailly, D., Roche, D., de Oca, R. M., Probst, A. V., Vassias, I., Dingli, F., Lombard, B., Loew, D., Quivy, J.-P., et al. (2011). Sumoylation promotes de novo targeting of hp1α to pericentric heterochromatin. *Nature genetics*, 43(3):220.

Masiero, S., Trotta, R., Pieraccini, S., De Tito, S., Perone, R., Randazzo, A., and Spada, G. P. (2010). A non-empirical chromophoric interpretation of cd spectra of dna g-quadruplex structures. *Organic & biomolecular chemistry*, 8(12):2683–2692.

McKenna, S. A., Kim, I., Puglisi, E. V., Lindhout, D. A., Aitken, C. E., Marshall, R. A., and Puglisi, J. D. (2007). Purification and characterization of transcribed rnas using gel filtration chromatography. *Nature protocols*, 2(12):3270.

Meehan, R. R., Kao, C.-F., and Pennings, S. (2003). Hp1 binding to native chromatin in vitro is determined by the hinge region and not by the chromodomain. *The EMBO journal*, 22(12):3164–3174.

Meier, M., Patel, T. R., Booy, E. P., Marushchak, O., Okun, N., Deo, S., Howard, R., McEleney, K., Harding, S. E., Stetefeld, J., and McKenna, S. A. (2013). Binding of G-quadruplexes to the N-terminal recognition domain of the RNA helicase associated with AU-rich element (RHAU). *J. Biol. Chem.*, 288(49):35014–35027.

Mendel, G. (1865). Versuche uber pflanzen-hybriden. *Vorgelegt in den Sitzungen.*

Meyne, J., Ratliff, R. L., and Moyzis, R. K. (1989). Conservation of the human telomere sequence (TTAGGG)n among vertebrates. *Proc. Natl. Acad. Sci. U.S.A.*, 86(18):7049–7053.

Miescher, F. (1869). Letter i; to wilhelm his; tübingen, february 26th, 1869. *Die Histochemischen und Physiologischen Arbeiten von Friedrich Miescher–Aus dem sissenschaft–lichen Briefwechsel von F. Miescher*, 1:33–8.

Minc, E., Allory, Y., Worman, H. J., Courvalin, J.-C., and Buendia, B. (1999). Localization and phosphorylation of hp1 proteins during the cell cycle in mammalian cells. *Chromosoma*, 108(4):220–234.

Minc, E., Courvalin, J.-C., and Buendia, B. (2000). Hp1γ associates with euchromatin and heterochromatin in mammalian nuclei and chromosomes. *Cytogenetic and Genome Research*, 90(3-4):279–284.

Mir, B., Serrano, I., Buitrago, D., Orozco, M., Escaja, N., and Gonzalez, C. (2017). Prevalent sequences in the human genome can form mini i-motif structures at physiological ph. *Journal of the American Chemical Society*, 139(40):13985–13988.

Muchardt, C., Guillemé, M., Seeler, J.-S., Trouche, D., Dejean, A., and Yaniv, M. (2002). Coordinated methyl and rna binding is required for heterochromatin localization of mammalian hp1α. *EMBO reports*, 3(10):975–981.

Nishibuchi, G., Machida, S., Osakabe, A., Murakoshi, H., Hiragami-Hamada, K., Nakagawa, R., Fischle, W., Nishimura, Y., Kurumizaka, H., Tagami, H., and Nakayama, J. (2014). N-terminal phosphorylation of HP1 increases its nucleosome-binding specificity. *Nucleic Acids Res.*, 42(20):12498–12511.

Ogloblina, A., Bannikova, V., Khristich, A., Oretskaya, T., Yakubovskaya, M., and Dolinnaya, N. (2015). Parallel g-quadruplexes formed by guanine-rich microsatellite repeats inhibit human topoisomerase i. *Biochemistry (Moscow)*, 80(8):1026–1038.

Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K., and Sugino, A. (1968). Mechanism of dna chain growth. i. possible discontinuity and unusual secondary structure of newly synthesized chains. *Proceedings of the National Academy of Sciences of the United States of America*, 59(2):598.

Olins, A. L. and Olins, D. E. (1974). Spheroid chromatin units (ν bodies). *Science*, 183(4122):330–332.

Olovnikov, A. M. (1973). A theory of marginotomy: the incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. *Journal of theoretical biology*, 41(1):181–190.

Panjkovich, A. and Svergun, D. I. (2017). Chromixs: automatic and interactive analysis of chromatography-coupled small-angle x-ray scattering data. *Bioinformatics*, 34(11):1944–1946.

Panossian, L., Porter, V., Valenzuela, H., Zhu, X., Reback, E., Masterman, D., Cummings, J., and Effros, R. (2003). Telomere shortening in t cells correlates with alzheimers disease status. *Neurobiology of aging*, 24(1):77–84.

Perrini, B., Piacentini, L., Fanti, L., Altieri, F., Chichiarelli, S., Berloco, M., Turano, C., Ferraro, A., and Pimpinelli, S. (2004). Hp1 controls telomere capping, telomere elongation, and telomere silencing by two different mechanisms in drosophila. *Molecular cell*, 15(3):467–476.

Randazzo, A., Spada, G. P., and da Silva, M. W. (2012). Circular dichroism of quadruplex structures. In *Quadruplex nucleic acids*, pages 67–86. Springer.

Rodriguez, R., Miller, K. M., Forment, J. V., Bradshaw, C. R., Nikan, M., Britton, S., Oelschlaegel, T., Xhemalce, B., Balasubramanian, S., and Jackson, S. P. (2012). Small-molecule–induced dna damage identifies alternative dna structures in human genes. *Nature chemical biology*, 8(3):301.

Rowley, J. D. (1991). Cytogenetics: past, present and future. In *Molecular foundations of oncology*. Williams and Wilkins, Baltimore, MD.

Schneider, R. and Grosschedl, R. (2007). Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes & development*, 21(23):3027–3043.

Schoeftner, S. and Blasco, M. A. (2008). Developmentally regulated transcription of mammalian telomeres by dna-dependent rna polymerase ii. *Nature cell biology*, 10(2):228.

Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., and Hurley, L. H. (2002). Direct evidence for a g-quadruplex in a promoter region and its targeting with a small molecule to repress c-myc transcription. *Proceedings of the National Academy of Sciences*, 99(18):11593–11598.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular systems biology*, 7(1):539.

Smothers, J. F. and Henikoff, S. (2000). The HP1 chromo shadow domain binds a consensus peptide pentamer. *Curr. Biol.*, 10(1):27–30.

Sun, D. and Hurley, L. H. (2010). Biochemical techniques for the characterization of g-quadruplex structures: Emsa, dms footprinting, and dna polymerase stop assay. In *G-Quadruplex DNA*, pages 65–79. Springer.

Vaughn, J. P., Creacy, S. D., Routh, E. D., Joyner-Butt, C., Jenkins, G. S., Pauli, S., Nagamine, Y., and Akman, S. A. (2005). The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J. Biol. Chem.*, 280(46):38117–38120.

Vorlickova, M., Kejnovska, I., Bednarova, K., Renciuk, D., and Kypr, J. (2012). Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes. *Chirality*, 24(9):691–698.

Wang, G., Ma, A., Chow, C. M., Horsley, D., Brown, N. R., Cowell, I. G., and Singh, P. B. (2000). Conservation of heterochromatin protein 1 function. *Mol. Cell. Biol.*, 20(18):6970–6983.

Wang, G. and Vasquez, K. M. (2014). Impact of alternative dna structures on dna damage, dna repair, and genetic instability. *DNA repair*, 19:143–151.

Wang, Y. and Patel, D. J. (1993). Solution structure of the human telomeric repeat d [ag3 (t2ag3) 3] g-tetraplex. *Structure*, 1(4):263–282.

Watson, J. D., Crick, F. H., et al. (1953). Molecular structure of nucleic acids. *Nature*, 171(4356):737–738.

Wiseman, T., Williston, S., Brandts, J. F., and Lin, L.-N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Analytical biochemistry*, 179(1):131–137.

Wongtawan, T., Taylor, J. E., Lawson, K. A., Wilmut, I., and Pennings, S. (2011). Histone H4K20me3 and HP1 are late heterochromatin markers in development, but present in undifferentiated embryonic stem cells. *J. Cell. Sci.*, 124(Pt 11):1878–1890.

Wright, W. E., Piatyszek, M. A., Rainey, W. E., Byrd, W., and Shay, J. W. (1996). Telomerase activity in human germline and embryonic tissues and cells. *Dev. Genet.*, 18(2):173–179.

Ye, Q., Callebaut, I., Pezhman, A., Courvalin, J.-C., and Worman, H. J. (1997). Domain-specific interactions of human hp1-type chromodomain proteins and inner nuclear membrane protein lbr. *Journal of Biological Chemistry*, 272(23):14983–14989.

Zeraati, M., Langley, D. B., Schofield, P., Moye, A. L., Rouet, R., Hughes, W. E., Bryan, T. M., Dinger, M. E., and Christ, D. (2018). I-motif dna structures are formed in the nuclei of human cells. *Nature chemistry*, 10(6):631.

Zhang, J.-y., Zheng, K.-w., Xiao, S., Hao, Y.-h., and Tan, Z. (2014). Mechanism and manipulation of dna: Rna hybrid g-quadruplex formation in transcription of g-rich dna. *Journal of the American Chemical Society*, 136(4):1381–1390.

Appendix A

Appendix

Table A.1: SRC G4 locations

Sequence name	Strand	Position in SRC gene
SRC2	+	628-657
SRC3	+	1348-1366
SRC14	-	2573-2602
SRC16	-	5451-5486
SRC19	-	8871-8905

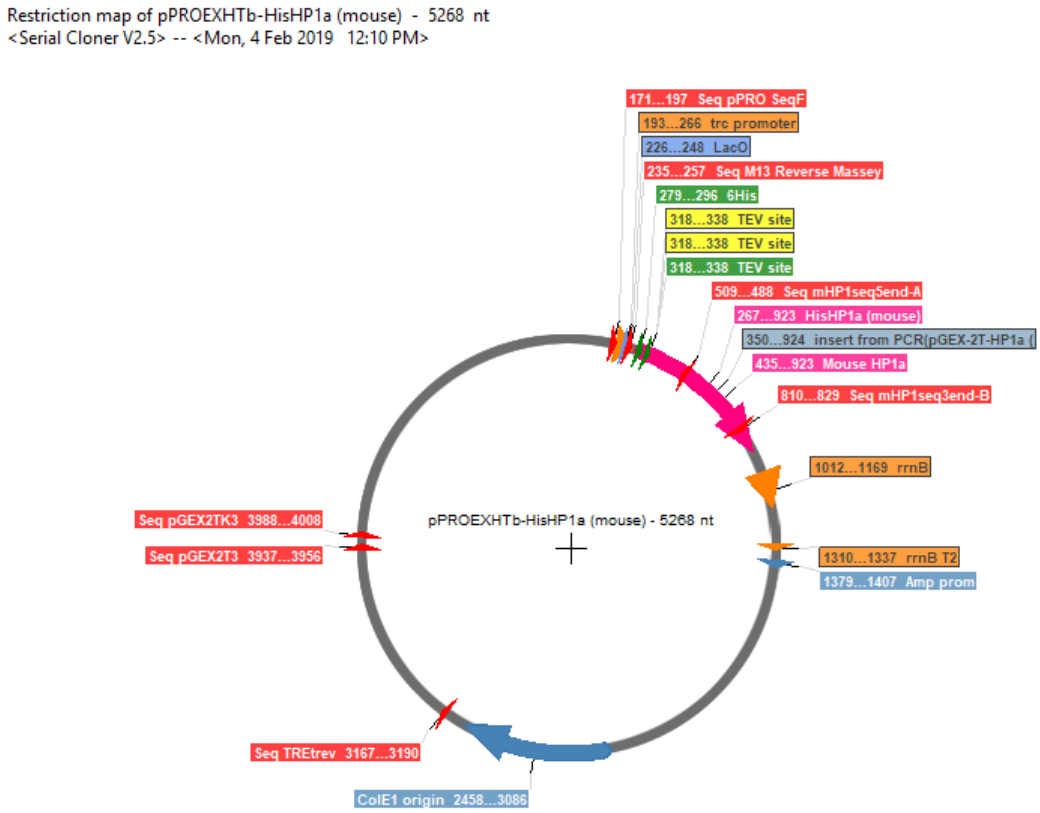

Figure A.1: Expression plasmid for HP1 proteins. Vector use for expression of HP1 isoform and mutant proteins. Shown is the HP1α expression plasmid, which contains the cDNA from *Mus musculus* HP1a. All HP1 protein used in this body of work has been produced with this plasmid, with variations in the cDNA sequence corresponding to each HP1 protein.

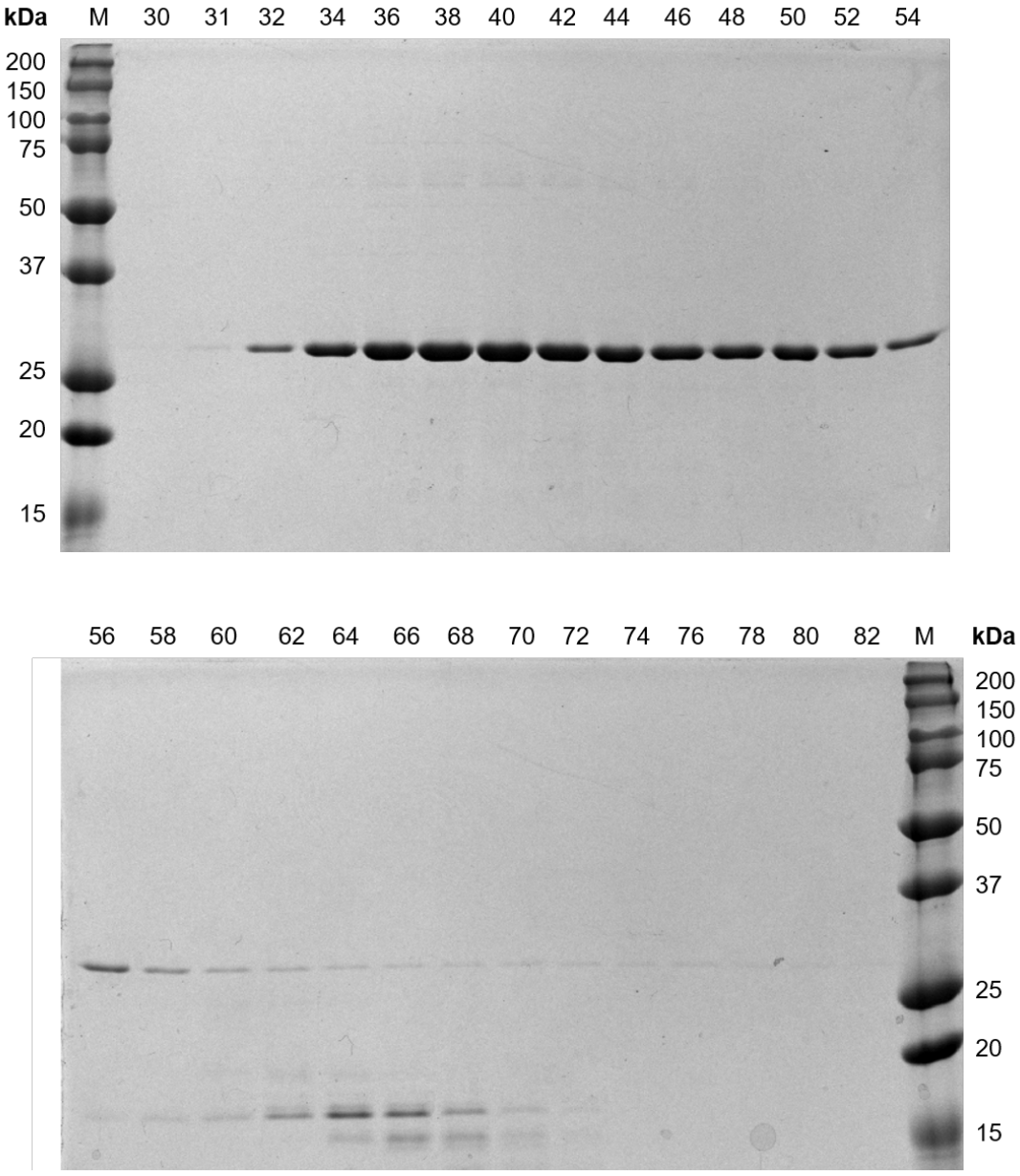

Figure A.2: SEC purification. Fractions of 2 μL from SEC with a Superdex 200 10/300 GL column electrophoresed on a 12% SDS-PAGE, stained with Coomassie. Fractions 34 to 56 were pooled.

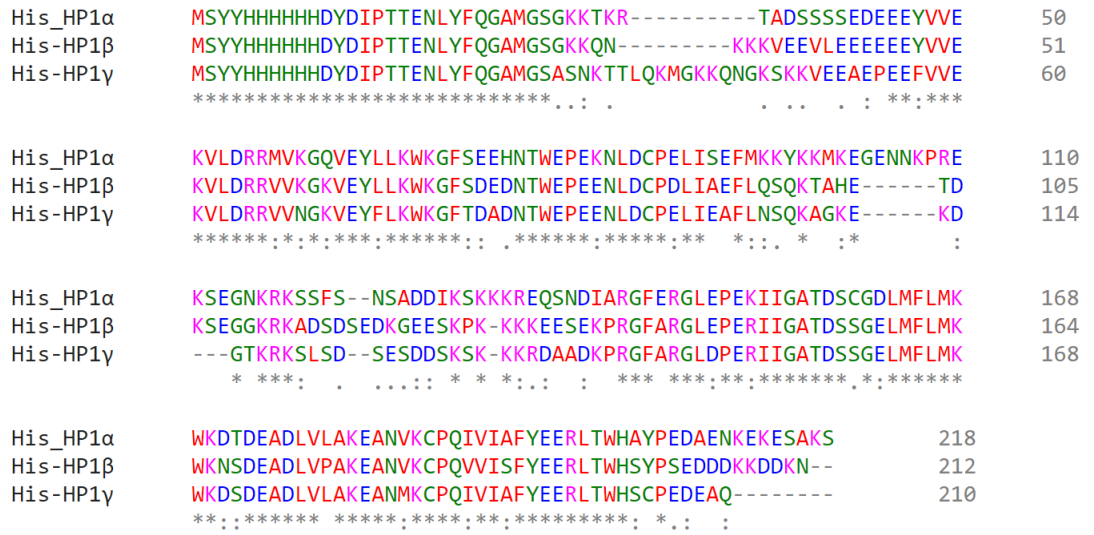

Figure A.3: His-HP1 amino acid sequences. Alignment of the his-tagged HP1 protein amino acid sequences used in this research, aligned using CLUSTAL O(1.2.4) multiple sequence alignment (Sievers et al., 2011). The numbers above the sequence indicate the amino acid position of the first and last hinge residue in the HP1α amino acid sequence. The black line indicates the three lysine residues in HP1α that were exchanged for alanine residues in the HP1α-3KA mutant. An "*" (asterisk) indicates positions which have a single, fully conserved residue. A ":" (colon) symbol indicates conservation between groups of residues with strongly similar properties. A "." (period) indicates conservation between groups of weakly similar residues.

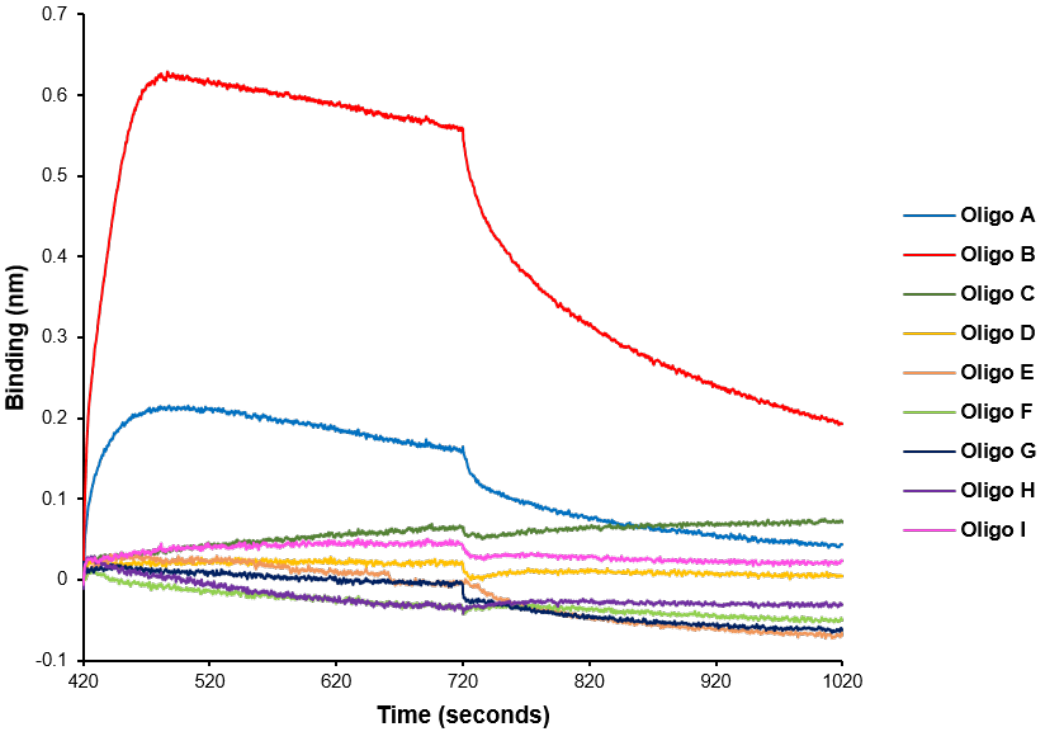

Figure A.4: Binding of HP1α to Oligos A-I. BLI performed with HP1α immobilised on a biosensor, with addition of Oligos A-I. Shorter oligonucleotides (C-I), do not produce significant signal, indicating that they may be too small to be detected by BLI.

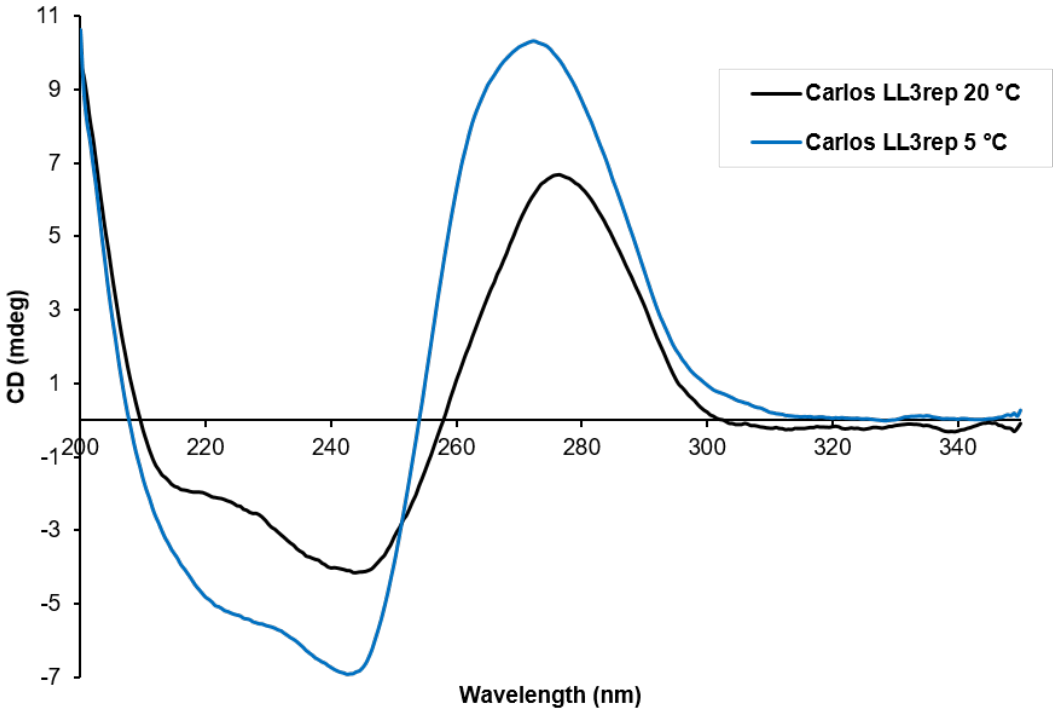

Figure A.5: CD spectra of CarlosLL3 at differing temperatures. CD performed on the CarlosLL3 i-motif oligonucleotide at 20 °C and 5 °C, at 10 μM.

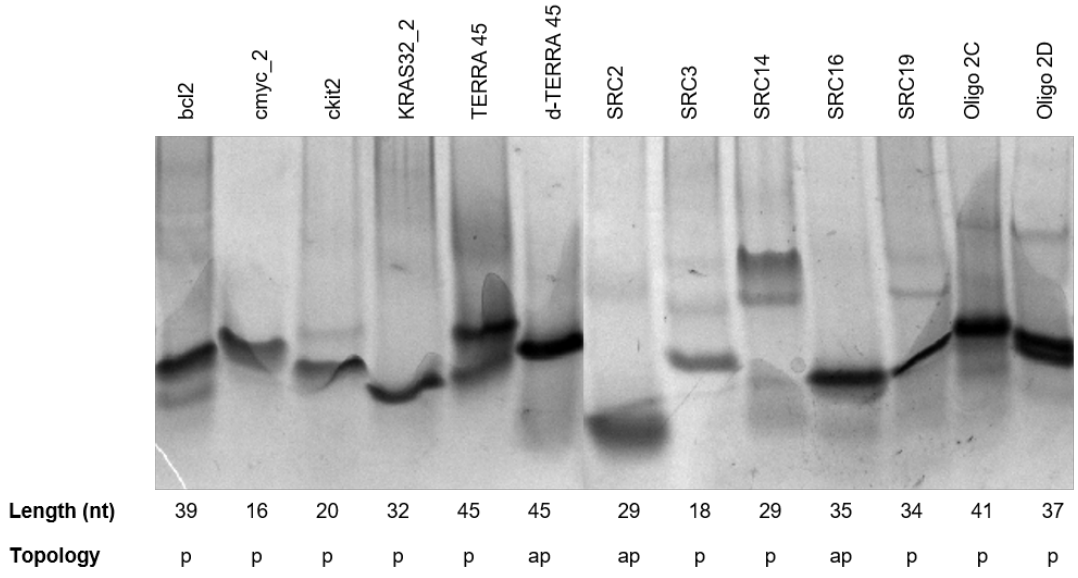

Figure A.6: Repeat of native oligonucleotide experiment. A native PAGE performed on endogenous G4s, TERRA 45, d-TERRA 45, Oligo 2C, and Oligo 2D. Of note, are the two distinctive species formed by TERRA 45.

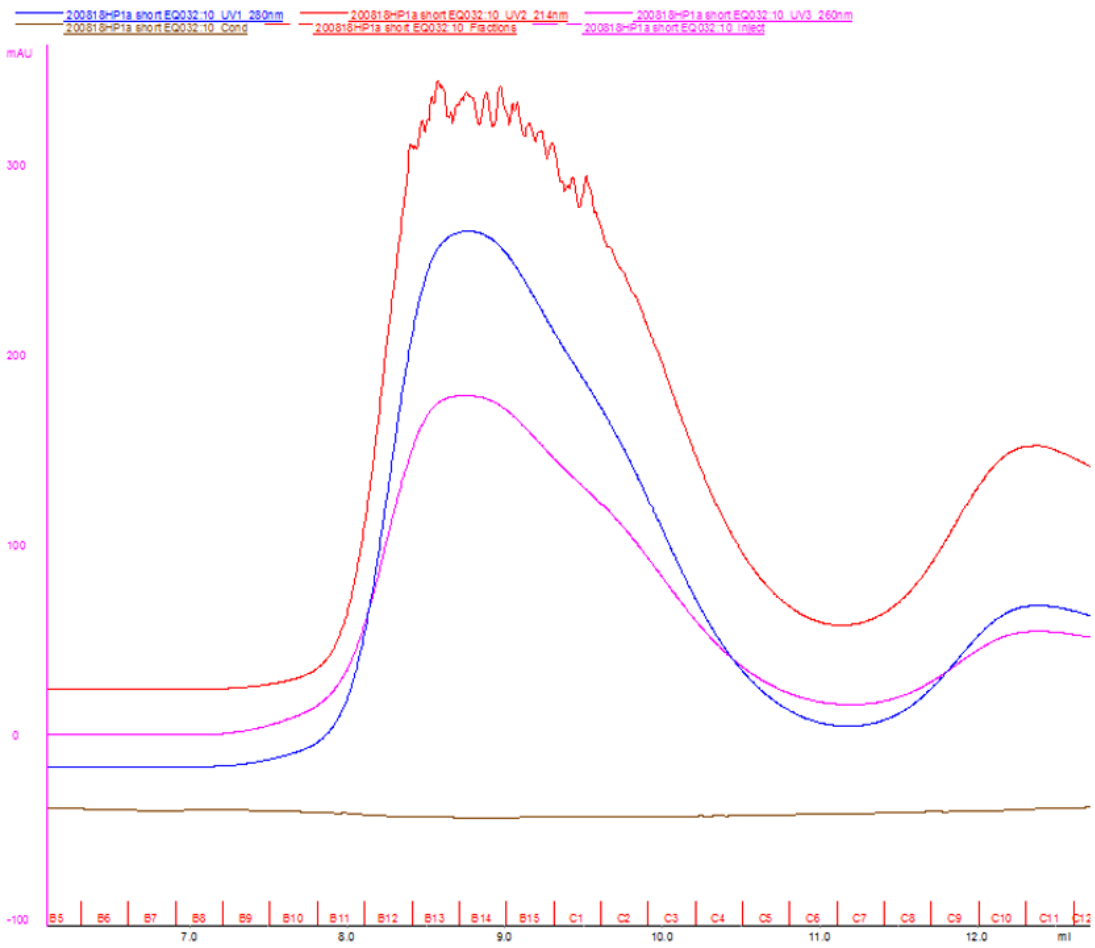

Figure A.7: HP1α chromatogram following size exclusion chromatography. Chromatogram showing HP1α elution from a Superdex 75 10/300 GL column.

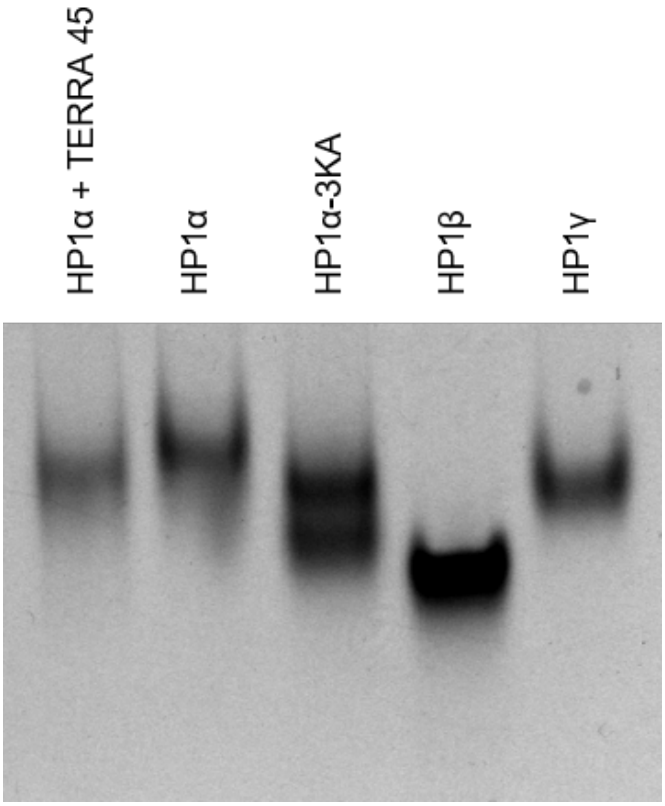

Figure A.8: Repeat of native protein experiment. Native PAGE performed on HP1 isoforms and mutant protein, with addition of TERRA 45 to HP1α.

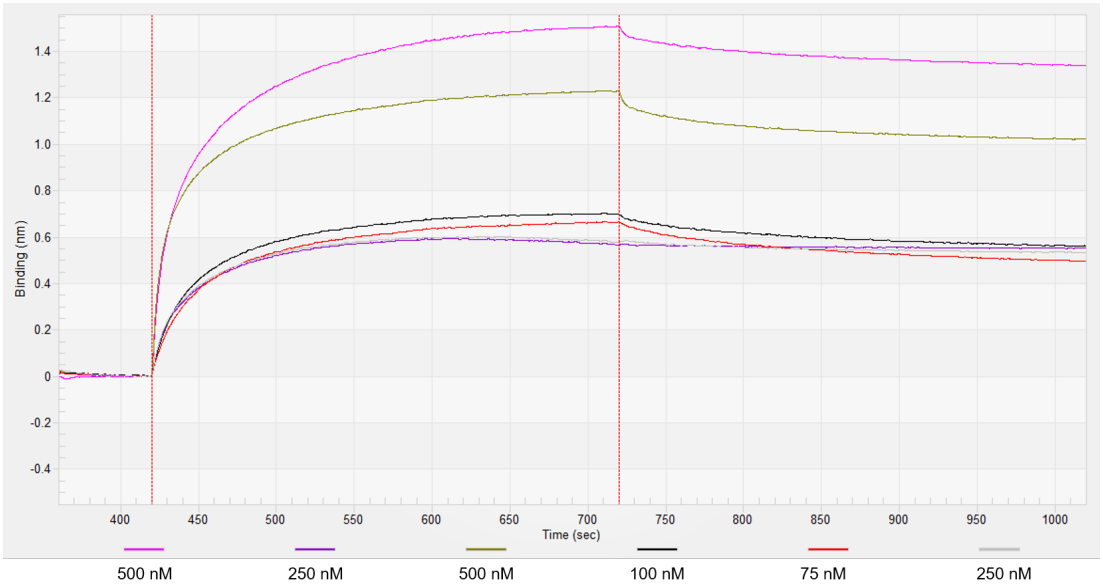

Figure A.9: Titrating TERRA 96 with HP1α. BLI performed with HP1α immobilised on the tip, and addition of varying concentrations of TERRA 96. These binding curves were used to calculate the equilibrium constant of the interaction. This data is not reliable, due to the differences in binding at the same concentration of TERRA 96, and not a clear trend in concentration versus binding.

a

$$\frac{[R][L]}{[RL]} = \frac{k_{\text{off}}}{k_{\text{on}}} = K_{\text{d}}$$

b

$$\frac{[RL]}{[R][L]} = \frac{k_{\text{on}}}{k_{\text{off}}} = K$$

Figure A.10: Equations showing how K_D and K are calculated. a) The equilibrium constant, K_D (in this equation shown as K_d) is calculated by k_{off} (k_d in BLItz analysis) divided by k_{on} (k_a in BLItz analysis), as performed by the BLItz software, in units M. b) The binding constant, K, calculated as shown, the opposite of as used in BLItz analysis, yielding units of M^{-1}. "R" refers to the receptor concentration, and "L" refers to the ligand.

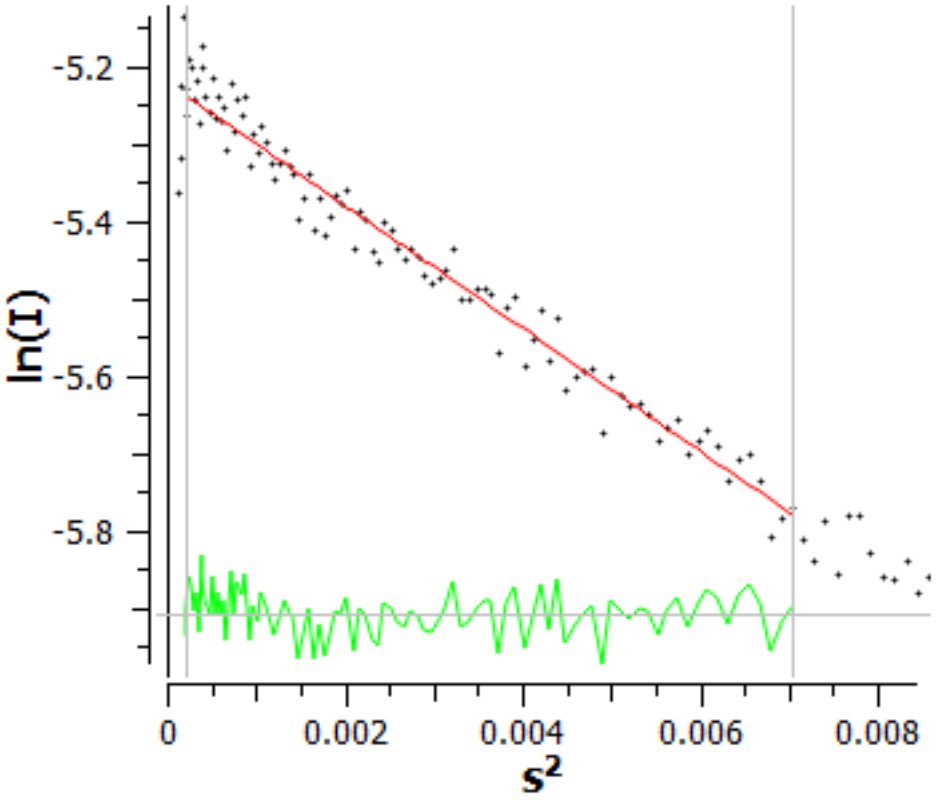

Figure A.11: Guinier plot of TERRA 45. The red line indicates a line of best fit (which is then used to calculate R_g values shown in (**Table 4.1**), and the green line indicates the position of data points above and below the line of best fit.

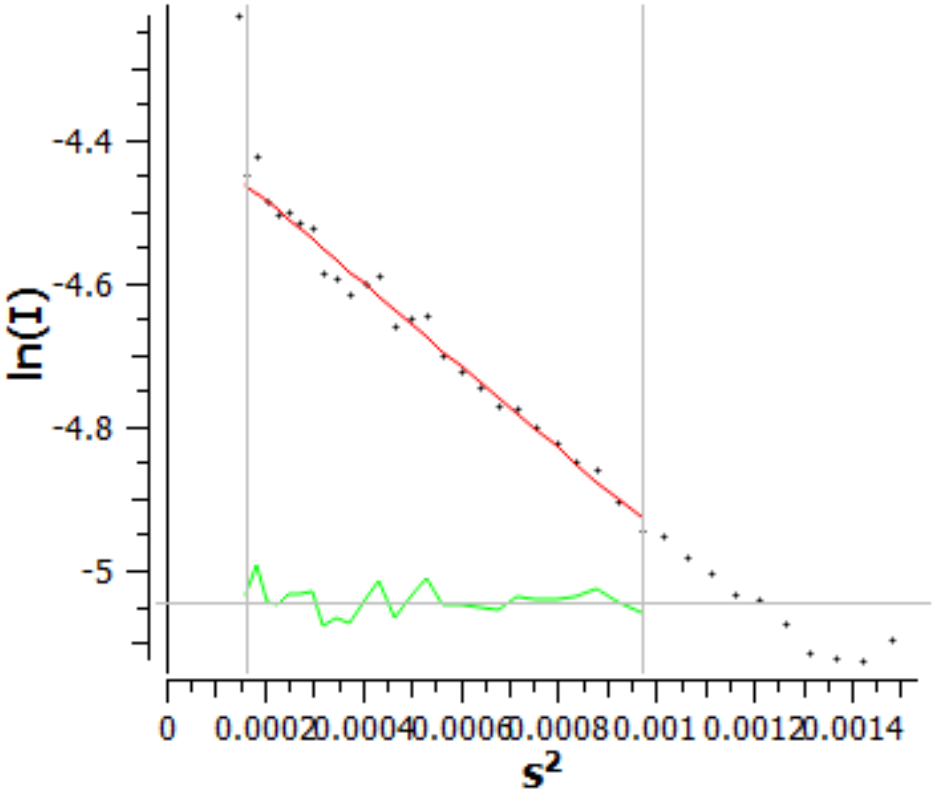

Figure A.12: Guinier plot of HP1α. The red line indicates a line of best fit (which is then used to calculate R_g values shown in (**Table 4.1**), and the green line indicates the position of data points above and below the line of best fit.

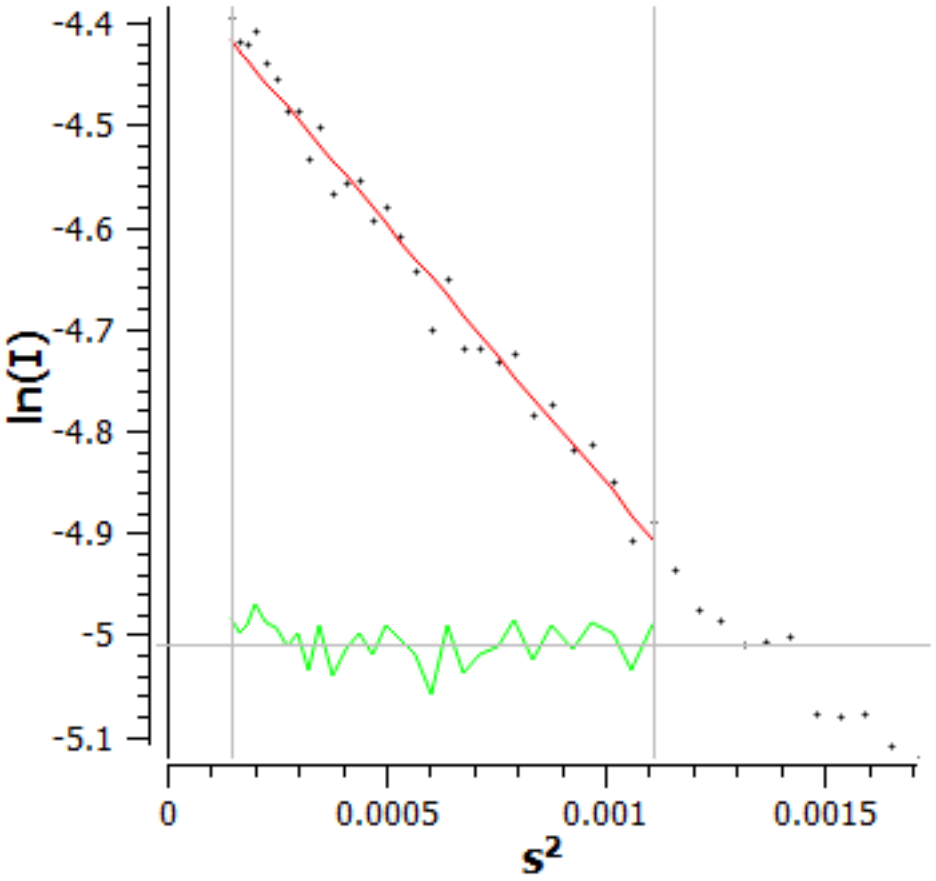

Figure A.13: Guinier plot of HP1α-3KA. The red line indicates a line of best fit (which is then used to calculate R_g values shown in (**Table 4.1**), and the green line indicates the position of data points above and below the line of best fit.

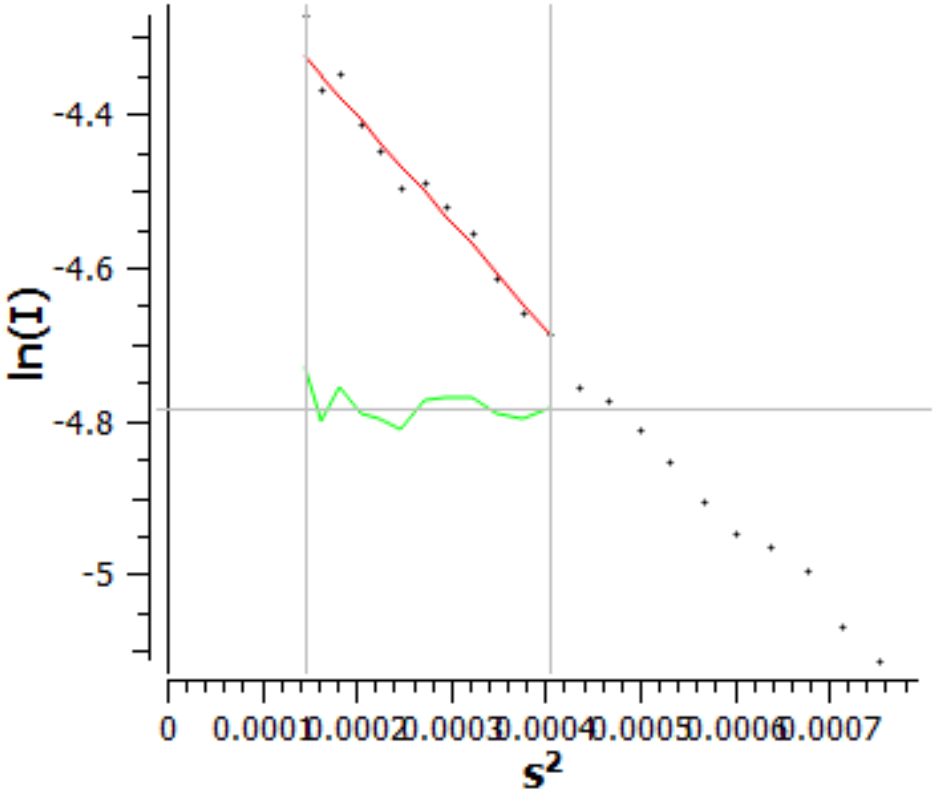

Figure A.14: Guinier plot of complex. The red line indicates a line of best fit (which is then used to calculate R_g values shown in (**Table 4.1**), and the green line indicates the position of data points above and below the line of best fit.

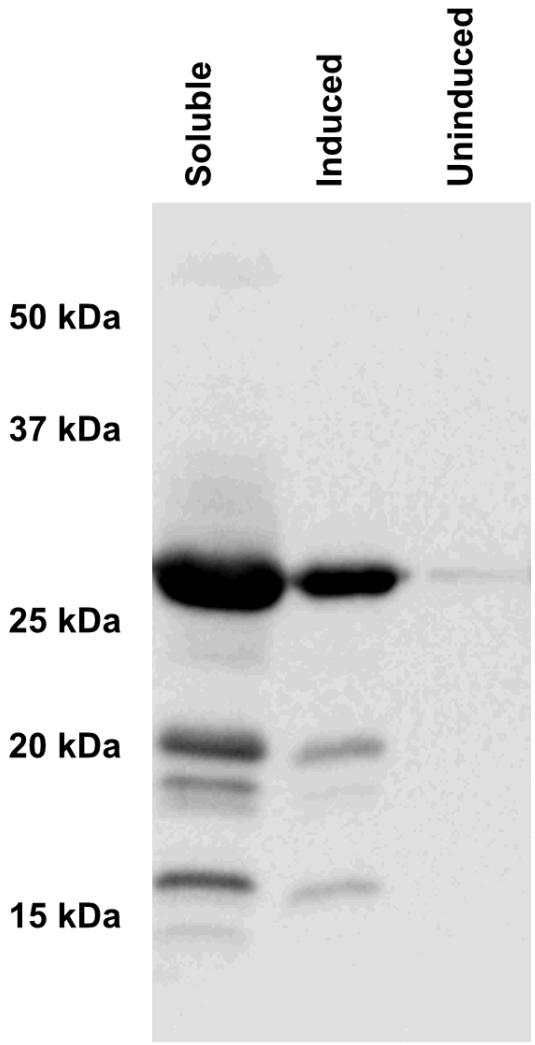

Figure A.15: Western blot of HP1α. Western blot showing the soluble fraction of an *E. coli* cell lysate, as well as induced (IPTG addition) and uninduced cell lysates in order to confirm the protein expressed was indeed HP1α, using HP1α antibody (#2616, Cell Signaling Technology, USA). Western blotting was performed by Imogen Dumville.

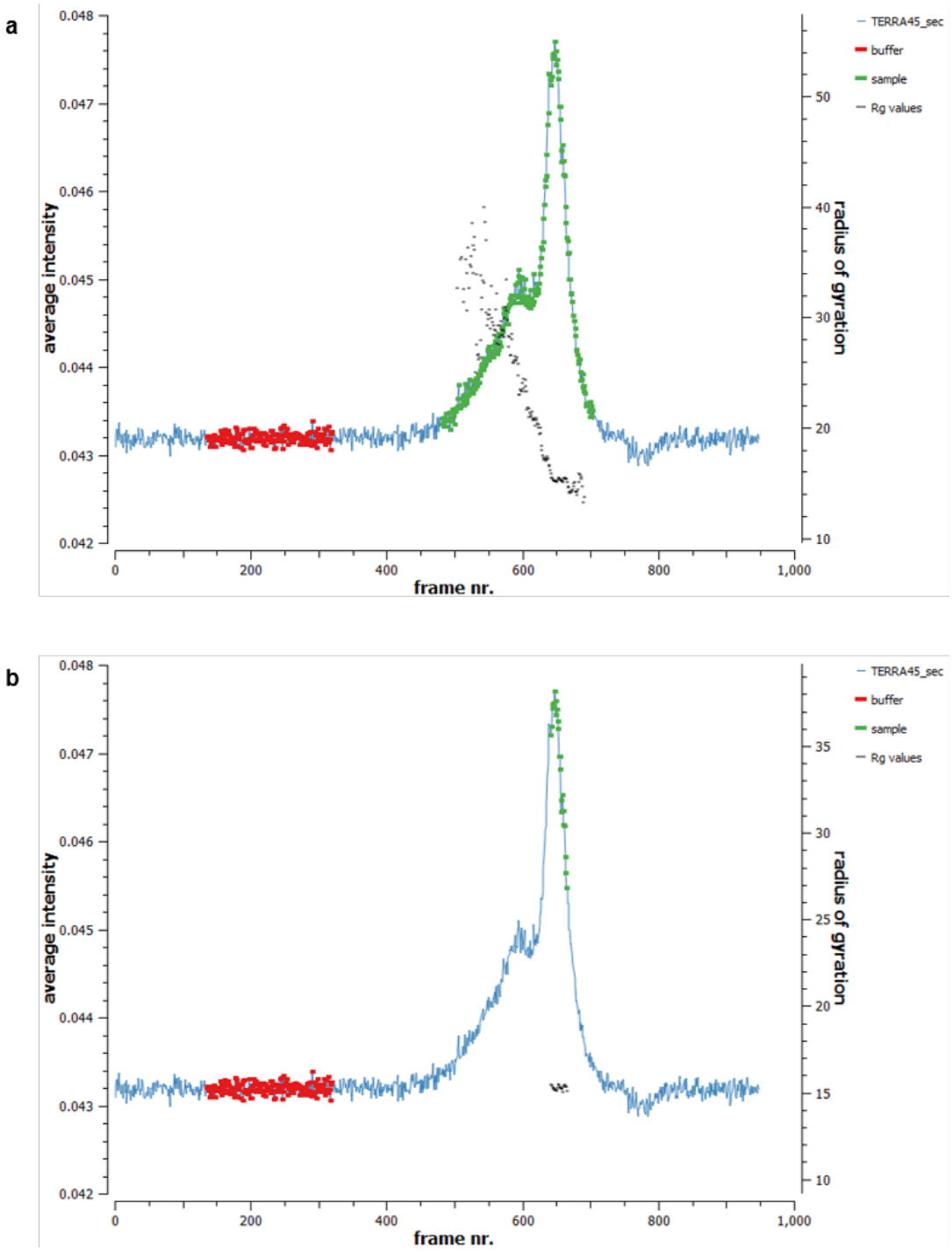

Figure A.16: Selection and subtraction of TERRA 45 using Chromixs for SEC-SAXS data processing. Chromatogram showing the average intensity of SAXS signal as TERRA 45 is eluted using SEC. a) Before data has been processed, showing widely differing R_g values. b) Data selection, showing a range of R_g values which are constant in the area of the chromatogram that corresponds to TERRA 45 elution.