Untangling the interaction between HP1α and the TERRA G-quadruplex

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Biochemistry

At Massey University, Manawatū,
New Zealand

Ruby Jean Roach

2023

Abstract

Over 50 years ago the DNA double-helix structure was solved, revealing the code in which life is written. This set of instructions equates to three billion base-pairs, amounting to two meters in length, fitting into a human nucleus just six microns across. The stable yet dynamic structure that allows for the functional organisation of the genome is chromatin: a complex formed between DNA and proteins. Chromatin is delineated into two microscopically defined compartments: the less dense gene-rich euchromatin and the structurally compact gene-poor heterochromatin. Protective heterochromatin is constitutively maintained for telomere protection, chromosome segregation, DNA repair, and suppression of transposon activity. Essential for the propagation and maintenance of heterochromatin is Heterochromatin Protein 1α (HP1α), which is comprised of a chromodomain and chromoshadow domain separated by a disordered hinge. The histone code model proposes that HP1α is recruited to regions of heterochromatin by its chromodomain-mediated recognition of silent heterochromatin mark trimethylated lysine 9 of histone H3 (H3K9me3); however, this does not account for the specificity of HP1 paralogs, location-specific recruitment, or the contribution of RNA to heterochromatin formation.

Here, Telomeric Repeat-containing RNA (TERRA), transcribed from the telomeres and shown to be involved in telomeric maintenance and stability, is investigated in its interaction with HP1α. The interaction with TERRA is proposed as a means for recruitment of HP1α to telomeres for chromosome end protection. Due to its guanine (G)-rich sequence, TERRA folds into a G-quadruplex (G4), a topology distinctly different from canonical nucleic acid structures. The interaction between HP1α and TERRA is therefore investigated to establish the means of interaction between HP1α and a G4, and to examine the HP1α specificity towards TERRA G4s. This work showed that binding to TERRA is dependent upon multiple positively charged patches within the disordered hinge of HP1α, and is also affected by mimicking N-terminal phosphorylation, which alters the structure of HP1α. While the hinge of HP1α binds to a myriad of nucleic acid structures, the globular chromoshadow domain provided the specificity for the parallel G4 topology evident in TERRA. Solution structures of HP1α in complex with TERRA show that HP1α undergoes a conformational shift, becoming less flexible. These results show that non-canonical nucleic acid structures such as those formed by TERRA may act as determinants of HP1α function, serving as signposts in the genome for formation of protective heterochromatin. This biophysical study also further implicates non-canonical structures such as G4s as formidable regulators of genomic function.

Acknowledgements

Firstly, I would like to express my gratitude for the funding that has supported me during my PhD: Massey University and the Australian Institute of Nuclear Science and Engineering.

Secondly, I would like to thank my supervisory team: Tracy Hale, Vyacheslav Filichev, Geoff Jameson, and Katy Wood. The wisdom that I have gained from all of you is indispensable.

Thirdly, I would like to thank the people in my life who were with me along the way to help me achieve this monstrous task: all of my close friendships and those that I have made at Massey and afar, my parents, and my partner, Jordan.

Finally, I would also like to take the opportunity to acknowledge the unique time in which this PhD was pursued. The world has changed immensely during this time due in part to the COVID-19 pandemic, and I will forever remember this experience and the importance of scientific truth.

Abbreviations

Å	Angstrom
AI	Artificial intelligence
ALT	Alternative lengthening of telomeres
APS	Ammonium persulfate
ATP	Adenosine triphosphate
BBB	Bead binding buffer
BCL2	B-cell lymphoma-2
BLI	Biolayer interferometry
BLM	Bloom syndrome protein
bp	Base-pair
BRCA1	Breast cancer type 1 susceptibility protein
CAF-1	Chromatin assembly factor-1
CAPITO	Circular dichroism analysis and plotting tool
CBX	Chromobox
CD	Chromodomain
cDNA	Complementary deoxyribonucleic acid
ChIP	Chromatin immunoprecipitation
CK2	Casein kinase II
cryoEM	Cryogenic electron microscopy
CSD	Chromoshadow domain
CST	CTC1–STN1–TEN1
CTCF	CCCTC-binding factor
CTE	C-terminal extension
D-loop	Displacement loop
Da	Dalton
DAPI	4',6-Diamidino-2-phenylindole
D_{max}	Maximum dimension
DMEM	Dulbecco's modified eagle medium
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EED	Embryonic ectoderm development
EMSA	Electrophoretic mobility shift assay
EZH1/2	Enhancer of zeste homolog 1/2
FENDRR	Fetal-lethal non-coding developmental regulatory RNA

Abbreviations

FISH	Fluorescence *in situ* hybridisation
fMSR	Forward major satellite-repeat RNA
FoXS	Fast X-Ray Scattering
G4	G-quadruplex
GST	Glutathione S-transferase
H3K9me3	Trimethlyated lysine 9 of histone H3
H3K27me3	Trimethlyated lysine 27 of histone H3
H4K20me3	Trimethlyated lysine 20 of histone H3
hnRNP A1	Heterogeneous nuclear ribonucleoprotein A1
HP1	Heterochromatin protein 1
HPLC	High-performance liquid chromatography
I	Intensity
i-motif	Intercalated motif
I(0)	Scattering intensity at zero angle
IB	Interaction buffer
IP	Immunoprecipitation
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
KAP1	KRAB-associated protein 1
kb	Kilobase
K_D	Equilibrium dissociation constant
k_{off}	Dissociation rate constant
k_{on}	Association rate constant
KRAB	Krüpple-associated box
LB	Lysogeny broth
LLPS	Liquid-liquid phase separation
lncRNA	Long non-coding RNA
log	Logarithm
M	Mol/L
me	Methylation
mol	Mole
MS	Mass spectrometry
MW	Molecular weight
MWCO	Molecular weight cut off

NFM	Non-fat milk
Ni-NTA	Nickel nitrilotriacetic acid
nm	Nanometre
NMR	Nuclear magnetic resonance
NSD	Normalised spatial discrepancy
nt	Nucleotide
NTE	N-terminal extension
NLS	Nuclear localisation signal
OnTEBP	*Oxytricha nova* telomeric end-binding protein
P(r)	Pairwise distance distribution function
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PcG	Polycomb group
PCR	Polymerase chain reaction
PDB	Protein data bank
PHD	Plant homeodomain
Pif1	Preimplantation factor-1
PISA	Proteins, Interfaces, Structures and Assemblies
POT1	Protection of telomeres 1
PRC	Polycomb repressive complex
PTM	Post-translational modification
q	Momentum transfer
QGRS	Quadruplex forming G-rich sequence
RAP1	Repressor-activator protein 1
RBAP46/48	Retinoblastoma protein associated proteins p46 and p48
R_g	Radius of gyration
RHAU	RNA helicase associated with AU-rich element
RITS	RNA-induced transcriptional silencing
RNA	Ribonucleic acid
RNAi	RNA interference
RPA	Replication protein A
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
SENP7L	Long SUMO-specific protease 7
siRNA	Short interfering RNA
smFRET	Förster Resonance Energy Transfer
SPR	Surface plasmon resonance
SSC	Saline Sodium Citrate
SUMO	Small Ubiquitin-like Modifier
SUV39H	Suppressor of variegation 3-9 homologue
SUZ12	Suppressor of zeste 12 homolog
Swi6	Switching-deficient 6
T-loop	Telomeric loop
TAE	Tris-acetate-EDTA
TAD	Topologically associating domain
TB	Tris-borate
TBE	Tris-borate-EDTA
TBST	Tris-buffered saline with Tween 20
TCEP	Tris(2-carboxyethyl)phosphine
TE	Tris-EDTA
TEMED	Tetramethylethylenediamine
TERC	Telomerase RNA component
TERRA	Telomeric repeat-containing RNA
TEV	Tobacco Etch Virus
TGS	Tris-glycine-SDS
TIF1	Transcription intermediary factor
TIN2	TRF1-interacting nuclear protein 2
TPP1	Telomere protection protein 1
TRF1/2	Telomere Repeat Factors 1 and 2
UV	Ultraviolet
WAXS	Wide-angle X-ray scattering
WRN	Werner syndrome ATP-dependent helicase
WT	Wildtype
Xist	X-inactive specific transcript

Nucleotide symbols

Adenine	A
Cytosine	C
Guanine	G
Thymine	T
Uracil	U

Amino acid abbreviations

Full name	Three-letter code	Single-letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In 1869 Friedrich Miescher discovered by mistake what he called "nuclein" while trying to isolate protein from white blood cells (Miescher, 1869). Seventy-five years later, Oswald Avery determined that deoxyribonucleic acid (DNA) was the "transforming principle" (Avery et al., 1944), the molecule responsible for heredity. Shortly thereafter, the crystallised DNA double-helix structure was solved by Watson and Crick (Watson et al., 1953). This revealed anti-parallel pairing of two chains of phosphate diester groups joining β-D-2'-deoxyribofuranose in a right-handed helix. Within the helix were the bases, the code in which life is written. The bases are canonically paired, adenine (A) with thymine (T), and guanine (G) with cytosine (C). This specific base-pairing was postulated to be responsible for the elusive copying mechanism of genetic material, whereby each complementary strand could function as a template for the formation of the other strand. However, the classical double-helix form is not the only kind of structure that nucleic acids can adopt.

Both DNA and subsequently transcribed ribonucleic acid (RNA) can form alternative structures depending on the sequence and conditions (Wang and Vasquez, 2014). Greater than half of the human genome is comprised of repetitive DNA sequences capable of forming these non-canonical structures, and as such, so is the 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 (Zhao et al., 2010). RNA is particularly inclined to form non-canonical nucleic acid structures when free in the cell, which have roles in regulation as long non-coding RNAs (lncRNA).

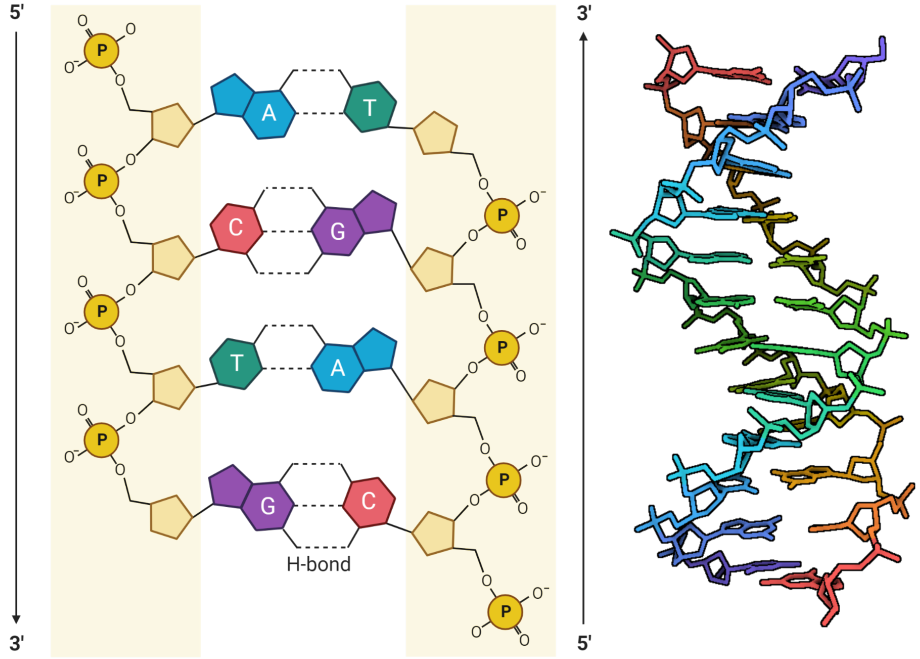

Figure 1.1: Watson-Crick base-pairing in B-DNA. Left) The anti-parallel pairing of complementary strands with typical Watson-Crick base-pairing. Right) A crystal structure of canonical B-DNA (PDB: 1BNA, Drew et al. 1981). Image created with BioRender.

1.1 Nucleic acid structures

1.1.1 DNA double-helices

DNA commonly forms the B-DNA structure, where a right-handed double-helix forms from two anti-parallel strands of DNA. The DNA strands are repeating chains of 2'-deoxyribose sugar and phosphate groups. The phosphate connects to the 5' carbon of the sugar, while a base connects to the 1' carbon to to form a nucleotide. Further nucleotides join through the 3' carbon of the sugar, creating a polymer chain of 5' to 3' links. Bases opposite one another in the double-helix hydrogen bond to form canonical A-T and G-C base-pairs (**Figure 1.1**). This structure allows for the soluble sugar-phosphate backbone to be exposed to the solvent, while the hydrophobic bases are shielded in the centre of the helix (Watson et al., 1953). This common B-DNA forms with a 2 nm diameter and 10 base-pairs per turn (**Figure 1.2**).

When the sequences are purine-rich (A and G), the A-DNA structure may be adopted (**Figure 1.2**), which is also a right-handed helix but with a larger 2.3 nm diameter and 11 base-pairs

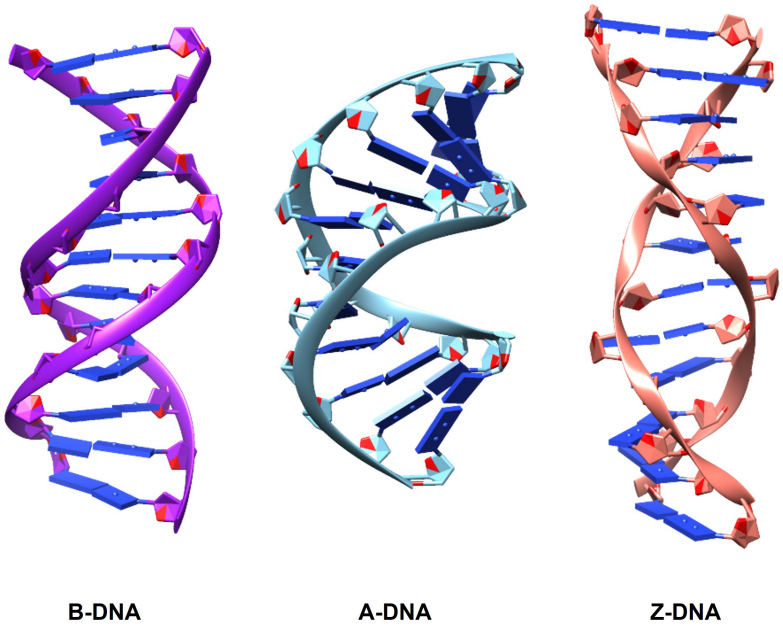

Figure 1.2: Types of DNA double-helices. B-DNA, A-DNA, and Z-DNA dodecamer double helices. B-DNA PDB: 1BNA (Drew et al., 1981), A-DNA PBD: 5MVK (Hardwick et al., 2017), Z-DNA PDB: 4OCB (Luo et al., 2014). Ribbons represent the DNA backbone, with pentagons representing the deoxyribose, and blue rectangles representing bases.

per turn. The major difference between B-DNA and A-DNA is the conformation of the sugar. Either the 2' -carbon or the 3' -carbon can be in the upwards position, resulting in a south (2' -endo) or north (3' -endo) conformation (**Figure 1.3a**). The south conformation is adopted in B-DNA, whereas the north conformation is adopted in A-DNA, resulting in reduced distance between adjacent nucleotides. This results in a more rigid structure and is also the form adopted by DNA-RNA hybrids and double-stranded RNA (Feig and Pettitt, 1998; Ivanov et al., 1974; Potaman and Sinden, 2013).

Additionally, DNA can also form a Z-DNA helix structure (**Figure 1.2**), where alternating purine and pyrimidine (eg. GCGCGC) tracts form an irregular left-handed zig-zagging helix with a diameter of only 1.8 nm but 12 base-pairs per turn (Ussery, 2002). The zig-zagging of Z-DNA is caused by G bases adopting the north conformation, as well as an alternative conformation of the glycosidic bond. This glycosidic bond connecting the 1' carbon on the sugar to the nitrogenous base can rotate to be in either an *anti* conformation, where the purine is positioned away from the sugar as in A- and B-DNA, or in a *syn* conformation, where the six-member ring of the purine is adjacent to the oxygen in the deoxyribose/ribose sugar as in Z-

DNA (**Figure 1.3b**). This conformational switch in Z-DNA results in alternating *syn* and *anti* stacking of bases, and also exposes the bases on the convex outer surface, requiring increased salt concentrations to reduce electrostatic repulsion between the phosphates (Choi and Majima, 2011; Herbert and Rich, 1999; Rich and Zhang, 2003).

1.1.2 Triplexes

Triplexes are triple-stranded structures of either DNA, RNA, or hybrid strands that can be formed when an oligonucleotide binds a duplex (**Figure 1.4a**). The binding of the third strand occurs through either Hoogsteen or reverse Hoogsteen base-pairing as opposed to Watson-Crick base-pairing (Hoogsteen, 1959). Hoogsteen base-pairs form alternative hydrogen bonds patterns (A with T or U, and C^+ with C or G, **Figure 1.5**) to typical Watson-Crick base-pairs through the rotation of a purine around the glycosidic bond (Hoogsteen, 1959, 1963). Triplex Hoogsteen base-pairing occurs with both the *anti* conformation and the less favourable *syn* conformation (Stadlbauer et al., 2014). However, despite this, triplex structures form *in vivo* shown by immunofluorescence with triplex-binding antibodies (Lee et al., 1987; Agazie et al., 1994). A lncRNA, Fetal-lethal non-coding developmental regulatory RNA (FENDRR), has been shown to form an RNA:DNA:DNA triplex at the promoters of developmental genes Foxf1 and Pitx2, thereby recruiting polycomb repressive complex 2 (PRC2) to result in silencing required for differentiation (Grote et al., 2013; Grote and Herrmann, 2013).

1.1.3 Intercalated-motifs

The intercalated (i)-motif is a four-stranded structure, where two sets of parallel double-stranded DNA intercalate in an anti-parallel fashion (**Figure 1.4c**). This occurs with cytosine-rich strands which form hemi-protonated cytosine ($C:C^+$, **Figure 1.6**) base-pairs with *anti* glycosidic bonds. Because of this protonation, acidic environments were thought to be required for their stabilisation *in vitro* (Day et al., 2014). However, i-motif structures have been shown to form at physiological pH under molecular crowding conditions (Rajendran et al., 2010), and recently were also shown to form *in vivo* in the nuclei of human cells when probed using an i-motif specific antibody (Zeraati et al., 2018). The formation of i-motif structures *in vivo* was shown to be altered by changing pH, which in turn affected the transcription of i-motif-containing promoters (Tang et al., 2020).

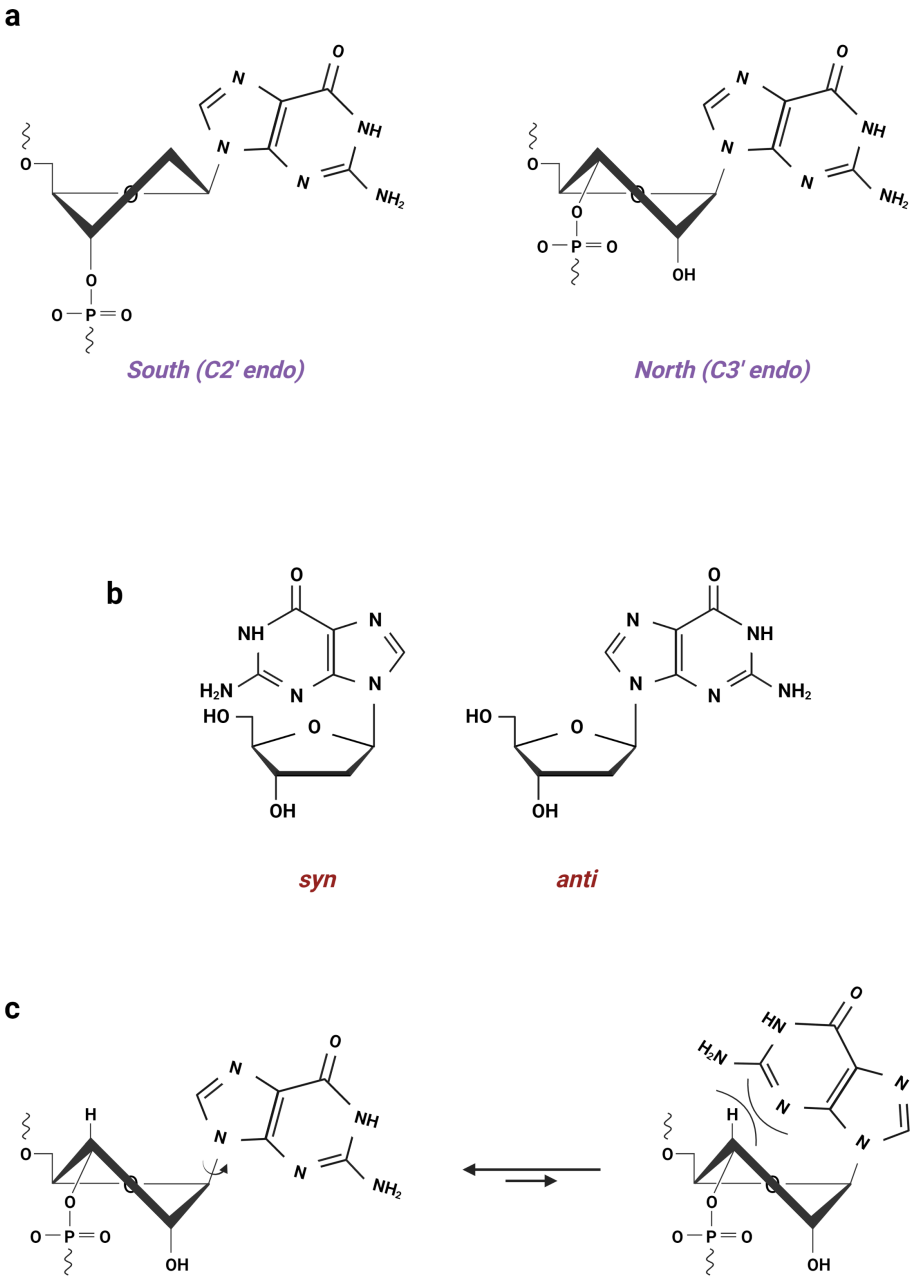

Figure 1.3: Sugar puckers and glycosidic bond angles of guanosine. a) On the left is the preferred sugar pucker for 2'-deoxyguanosine in the south conformation (C2' -endo), and on the right is the preferred sugar pucker for guanosine in the north conformation (C3' -endo). b) The *syn* and *anti* conformations of 2'-deoxyguanosine. c) In the north conformation of guanosine, the *anti* conformation is preferred due to steric clash between the H3' of the sugar and the N3 of guanine in the *syn* conformation. Image created with BioRender.

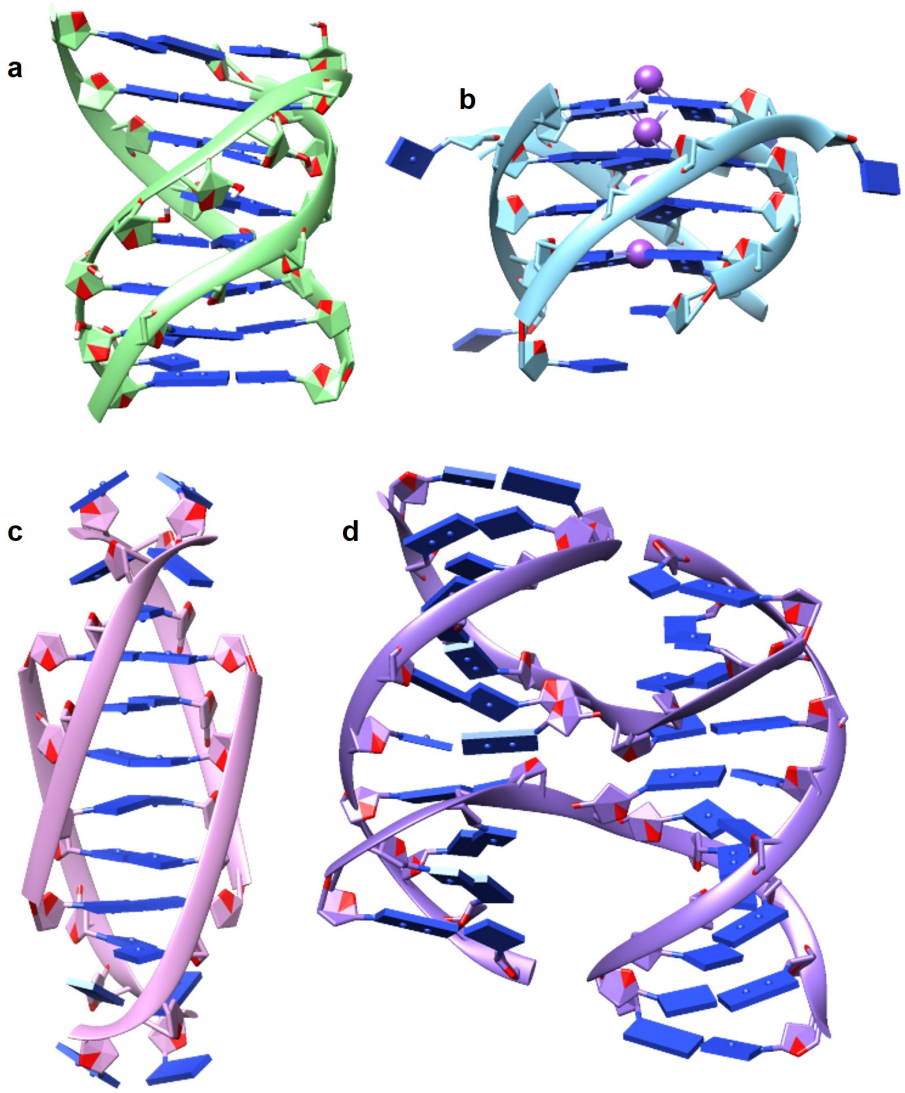

Figure 1.4: Alternative nucleic acid structures. a) A triple-helix (PDB: 1R3X, Gotfredsen et al. 1998), b) a G4 (PDB: 244D, Laughlan et al. 1994), c) an i-motif (PDB: 1YBL, Esmaili and Leroy 2005), d) and a Holliday junction (PDB: 3T8P, Mandal et al. 2011). Ribbons represent the DNA backbone, with pentagons representing the deoxyribose, and blue rectangles representing bases.

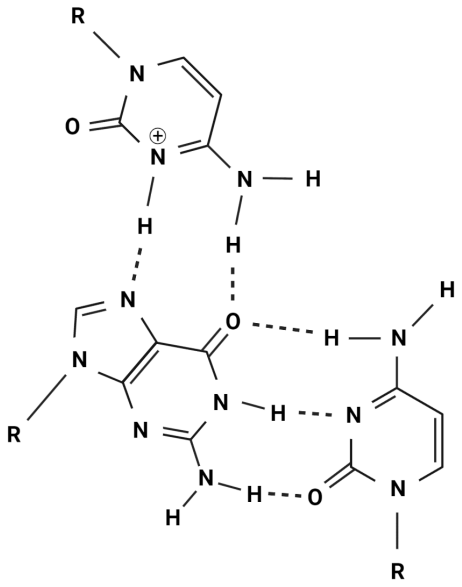

Figure 1.5: Base-pairing in a triplex. Hydrogen bonding pattern in triplex DNA, with Hoogsteen hydrogen bonds in the C^+:G base-pair, and Watson-Crick hydrogen bonds in the G:C base-pair.

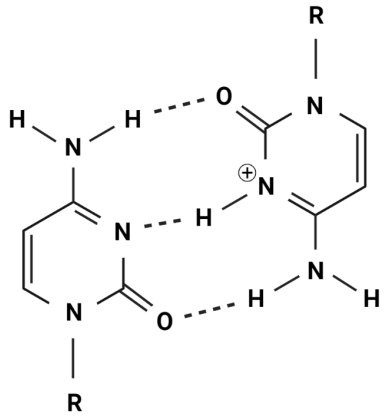

Figure 1.6: Cytosine:cytosine$^+$ base-pair. The C:C^+ base-pair forms under acidic conditions in i-motif structures.

1.1.4 Holliday junctions

Holliday junctions are four-stranded, right-handed, anti-parallel structures (**Figure 1.4d**), arranged in a cruciform-like fashion (Mandal et al., 2011). Due to the four phosphate-containing strands, charge neutralisation by cations such as Ca^{2+} are required for stabilisation of the structure (Thorpe et al., 2003). Before the structure of this junction was resolved, its organisation was hypothesised by Robin Holliday as a means of strand exchange in gene-conversion events in fungi (Holliday, 1964). Holliday junctions arise when a strand of nicked DNA invades the homologous double-stranded DNA, connecting two DNA duplexes and forming the four-stranded structure. In eukaryotes, resolution of the Holliday junction is crucial to avoid chromosome non-disjunction and aneuploidy, requiring structure-specific endonucleases such as SLX1-SLX4, MUS81-EME1, and GEN1 to resolve junctions (Wechsler et al., 2011; Wyatt et al., 2013).

1.1.5 Guanine quadruplexes

Prior to the discovery of the DNA double-helix structure, it was reported that guanine, but not any other nucleoside, formed gels when at high concentrations in solution (Steudel and Brigl, 1910). It wasn't for another 50 years when the G-tetrad structure was solved that this gelatinous behaviour was explained (Gellert et al., 1962). G-tetrads form when four guanines Hoogsteen hydrogen bond to make a planar structure (**Figure 1.7a**). When many guanines are present in a DNA or RNA strand they contribute to the formation of several G-tetrads that stack through π-π interactions. This four-stranded structure is therefore referred to as a G-quadruplex (G4, **Figure 1.4b**).

Biophysical studies on G4 structures have established specific sequence motifs which have allowed for the prediction of G4 formation using the following algorithm, $\mathbf{G}_x$-$\mathbf{N}_{1-7}$-$\mathbf{G}_x$-$\mathbf{N}_{1-7}$-$\mathbf{G}_x$-$\mathbf{N}_{1-7}$-$\mathbf{G}_x$ where **x** is 3-6 and **N** is any nucleotide (Puig Lombardi and Londoño-Vallejo, 2020). The structures of G4-forming sequences can be confirmed using circular dichroism spectroscopy (Vorlíčková et al., 2012), nuclear magnetic resonance (NMR, Adrian et al. 2012), and/or ultraviolet (UV) melting experiments (Mergny and Lacroix, 2009). Crystal structures of short G4 sequences have also been published, amounting to 286 X-ray PDB structure entries (Zok et al., 2022).

DNA or RNA strands that contain a high abundance of guanine bases can form (intramolec-

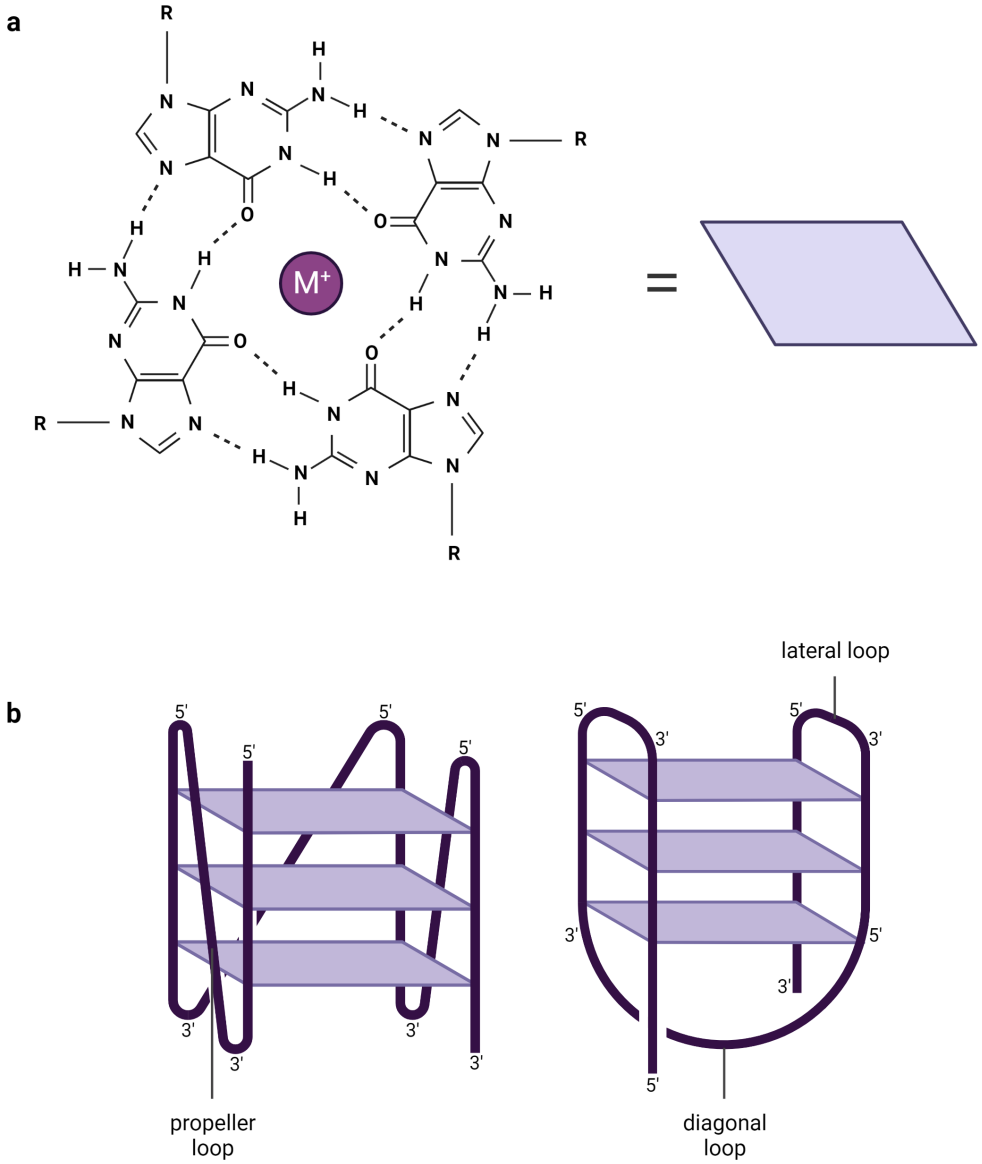

Figure 1.7: G4 structure. a) Four guanine nucleobases form a tetrad through Hoogsteen bonding, resulting in an electron-rich centre which is neutralised by a monovalent cation (M^+). b) On the left, a representation of a unimolecular parallel G4 structure with propeller loops separating G-tracts. On the right, a unimolecular anti-parallel G4 structure with two lateral loops at the top of the G4, and a diagonal loop at the bottom. Image created with BioRender.

ularly) or contribute to a G4 (intermolecularly) in many possible topologies. If strands are in the same orientation (5' to 3') then this is a parallel topology, and if one or more strands have different orientations then this is a hybrid/anti-parallel topology (**Figure 1.7b**). Variations in strand orientation, G-tract length, loop orientation, and loop length contribute to the wide array of possible G4 structures.

The way in which the G-tetrads stack is dependent on the glycosidic bond angles of the guanine bases involved. In a G4 of parallel topology, guanines adopt *anti* conformation, whereas in anti-parallel or hybrid topologies some guanines have to adopt *syn* conformation (**Figure 1.3b**) to facilitate the change in strand orientation (Masiero et al., 2010). In addition to this, the ribose sugar also has alternate conformations. For deoxyribose, the sugar ring interconverts between the north and south conformations, while for the ribose ring the north conformation is preferred (Doluca et al., 2013). Because of the steric clash between the 3'-H group of ribose and the guanine base in the *syn* conformation (**Figure 1.3c**), RNA molecules are known to exclusively form parallel G4 structures (Martadinata and Phan, 2009). This topological restriction contributes to the stability of RNA parallel G4s, resulting in structures with higher melting temperatures than their DNA counterparts (Arora and Maiti, 2009; Joachimi et al., 2009).

Loops that connect strands contributing to a G4 may be: lateral, where adjacent strands are joined on either the top or bottom of the G4, diagonal, where opposite strands join diagonally, or propeller, where strands are joined from opposite ends of the G4 such as in a completely parallel structure (**Figure 1.7b**, Burge et al. 2006). In DNA G4s, altering loop length can result in alternative structures, whereby shortening the loop length to just one nucleotide results in parallel structure, but loops of two or more allow both parallel and anti-parallel structures to form (Hazel et al., 2004). Altering loop length also affects stability, whereby addition of an extra thymidine to a loop in a G4 of four tracts of three guanines results in a decrease in the melting temperature (Guedin et al., 2010). When G4 structures are assembled with increasing numbers of consecutive guanines (eg. d(TGGGT), d(TGGGGT), and d(TGGGGGT)), this increases the number of tetrads in the G4, thereby increasing both the enthalpy of dissociation and melting temperature (Petraccone et al., 2005). To stabilise the G4, cations such as K^+ or Na^+ are required to coordinate with O6 of guanine (Burge et al., 2006). Particularly, K^+ and Sr^{2+} result in G4s with higher melting temperatures due to their atomic radii fitting within the G4 channel (Miserachs et al., 2016).

An additional layer of complexity to the G4 structure is the ability to multimerise, whereby parallel G4s can associate through stacking interactions when external tetrads are exposed to solvent (Kogut et al., 2019) and can form dimers or larger contiguous G4s as G-wires (Marsh and Henderson, 1994). Strings of anti-parallel G4s have been shown to form a "beads on a string" formation, where linker DNA between G4s is exposed to solvent but the interior of the G4 unit is dehydrated (Yu et al., 2012a).

1.1.5.1 G4 interactions and functions in the cell

Upon their discovery (Gellert et al., 1962), G4s were initially deemed an *in vitro* artefact despite being thermodynamically and kinetically stable structures that unfold slower than DNA or RNA hairpins (Lane et al., 2008). 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). Their presence suggests G4s may have novel roles in genome regulation. High-throughput sequencing-based methods have found over 700,000 potential DNA G4 structures in the human genome (Chambers et al., 2015), which led to the development of a comprehensive transcriptome-wide search engine for RNA G4s (Yu et al., 2022). Confirmation of the presence of G4s *in vivo* has since been demonstrated. Immunofluorescence studies using antibodies that recognise G4 structures showed their presence in both the cytoplasm and nuclei of cells, with their highest quantity in the S-phase (Biffi et al., 2013, 2014).

Given the abundance of G4s found in the genome, a number of G4-binding human proteins relating to telomere regulation, DNA replication, chromatin regulation, and numerous promoter functions have been categorised (Brázda et al., 2014; Meier-Stephenson, 2022; Shu et al., 2022). Meier-Stephenson (2022) separates G4-binding proteins into three categories based upon their mode of binding, either by top-stacking (hydrophobic interactions with exposed G-tetrad), loop-binding (binding with external loop nucleotides), groove-binding (between spaces of loops), or a combination of those modes (**Figure 1.8**).

The DEAH box RNA helicase associated with AU-rich element (RHAU, also known as DHX36 or G4R1) has been shown to recognise and unfold both RNA and DNA G4 structures (Vaughn et al., 2005; Heddi et al., 2015; Chen et al., 2018). RHAU G4 resolvase activity is important for

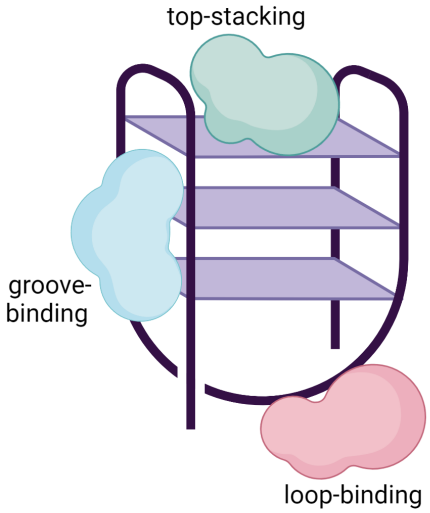

Figure 1.8: G4-binding protein types. Modes of G4 binding by proteins. Image created with BioRender.

transcriptional and post-transcriptional regulation and more specifically resolves the G4 formed by the sequence of the c-myc oncogene, which has a repressive function in the c-myc promoter (Siddiqui-Jain et al., 2002). An NMR structure of a 20 residue RHAU peptide bound to a parallel G4 formed from an 18 nt synthetic DNA construct revealed that the α-helical peptide covers the G-tetrad at one end in a top-stacking fashion (**Figure 1.9**), while electrostatic interactions are formed between the positively charged amino acids and negatively charged phosphate groups for loop-binding (PDB: 2N21, Heddi et al. 2015). Following this, a crystal structure of RHAU in complex with parallel c-myc promoter G4 yielded similar results, also showing that a single-stranded DNA tail arranged in the core of the helicase (**Figure 1.10**) aided in the unfolding of the c-myc G4 (PDB: 5VHE, Chen et al. 2018). The telomeric end-binding protein of *Oxytricha nova* (OnTEBP, analogous to human protection of telomeres 1, POT1) was shown to bind an anti-parallel telomeric DNA G4 (PDB: 1JB7, Horvath and Schultz 2001). The crystal structure showed lysine and asparagine residues interacting with the phosphate backbone, while a tyrosine protruded to bind the groove, interacting with several guanosines (**Figure 1.11**). Subsequently, it was found that human POT1 selectively binds anti-parallel G4s (Ray et al., 2014) to unfold and refold a G4 in four sequential steps shown by single-molecule Förster Resonance Energy Transfer (smFRET, Hwang et al. 2012).

In addition to the modes of binding, the targeting and functionality of G4-binding proteins are vast (Meier-Stephenson, 2022). The structural intricacies of G4s make them surprisingly

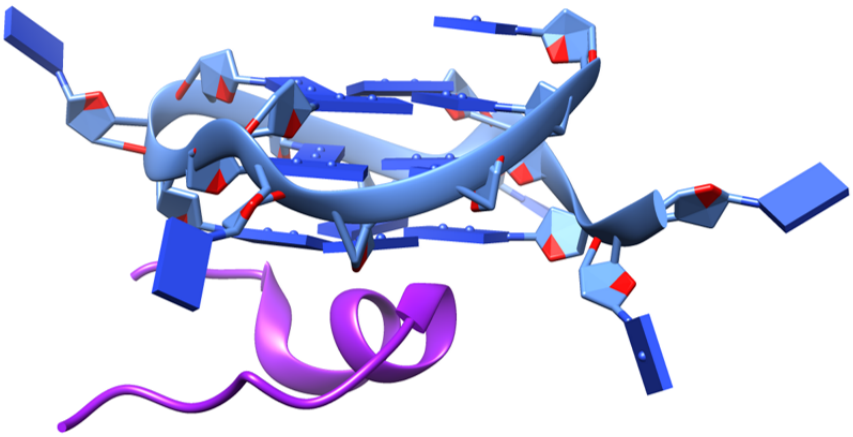

Figure 1.9: RHAU peptide interacting with a G4. An NMR structure (PDB: 2N21, Heddi et al. 2015) of a 20 amino acid RHAU peptide (purple) interacting with a parallel G4 (blue). Blue ribbons represent the DNA backbone, with pentagons representing the deoxyribose, and blue rectangles representing bases. Purple ribbons represent the protein.

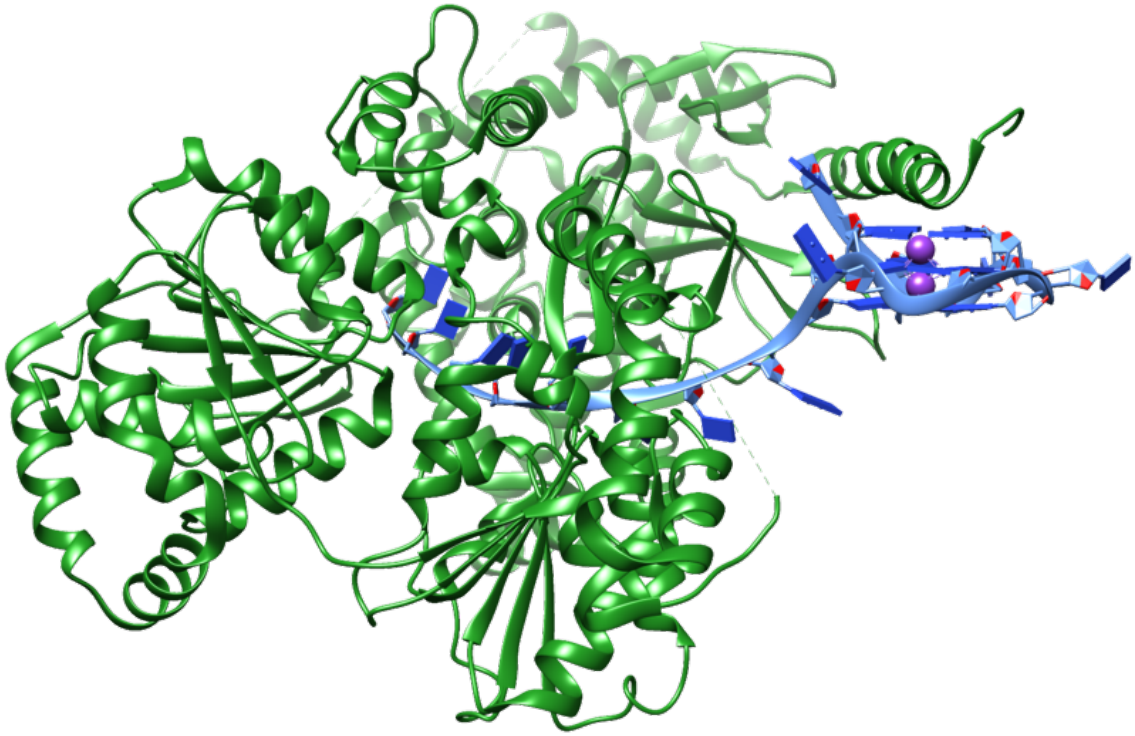

Figure 1.10: RHAU in complex with c-myc G4. A crystal structure (PDB: 5VHE, Chen et al. 2018) of RHAU (green) in complex with the c-myc G4 (blue). Blue ribbons represent the DNA backbone, with pentagons representing the deoxyribose, and blue rectangles representing bases. Green ribbons represent the protein.

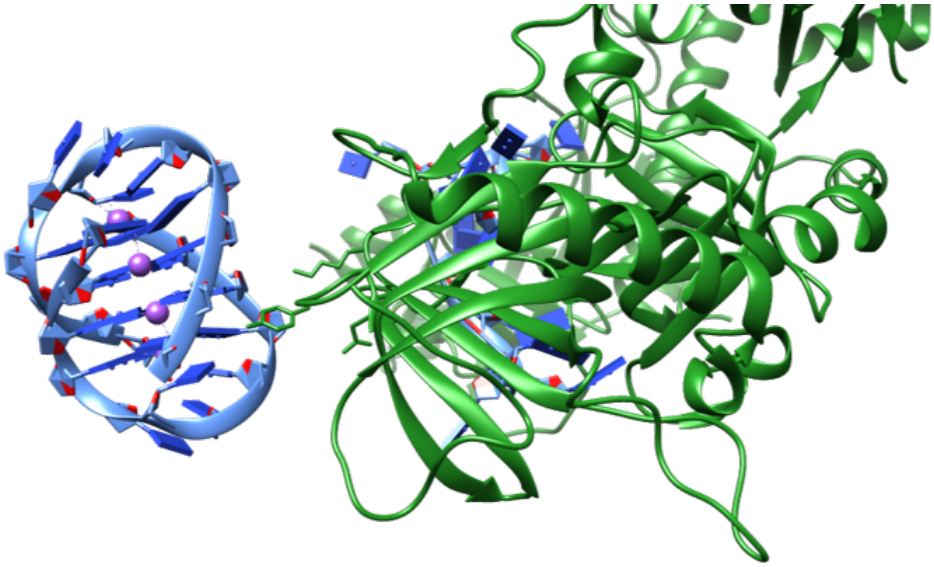

Figure 1.11: OnTEBP interacting with a telomeric DNA G4. Crystal structure (PDB: 1JB7, Horvath and Schultz 2001) of OnTEBP (green) interacting with an anti-parallel telomeric DNA G4 (blue). Blue ribbons represent the DNA backbone, with pentagons representing the deoxyribose, and blue rectangles representing bases. Green ribbons represent the protein.

heterologous and as such may be unique targets for drug design. An attractive target are the telomeric G4s, which were some of the first G4-forming sequences identified (Williamson et al., 1989; Wang and Patel, 1993).

1.2 Mammalian telomeres

Telomeres are the repetitive sequence at linear chromosome ends that act to protect the coding sequence (Olovnikov, 1973). The mammalian telomeric DNA sequence is highly conserved and consists of repeats of the sequence TTAGGG, with CCCTAA on the complementary strand (Allshire et al., 1989). At birth, human chromosomes possess approximately 10 kb of telomeric repeats (Okuda et al., 2002). Through each round of DNA replication, 50-100 bp (Harley et al., 1990) of DNA is lost 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". When telomere repeats have been lost after many successive rounds of DNA replication, either random recombination or cellular senescence occurs as a result of the DNA damage response to limit further DNA deterioration (D'Adda di Fagagna et al., 2003). This replicative senescence serves as a tumour suppressor mechanism that controls the number of divisions cells can safely

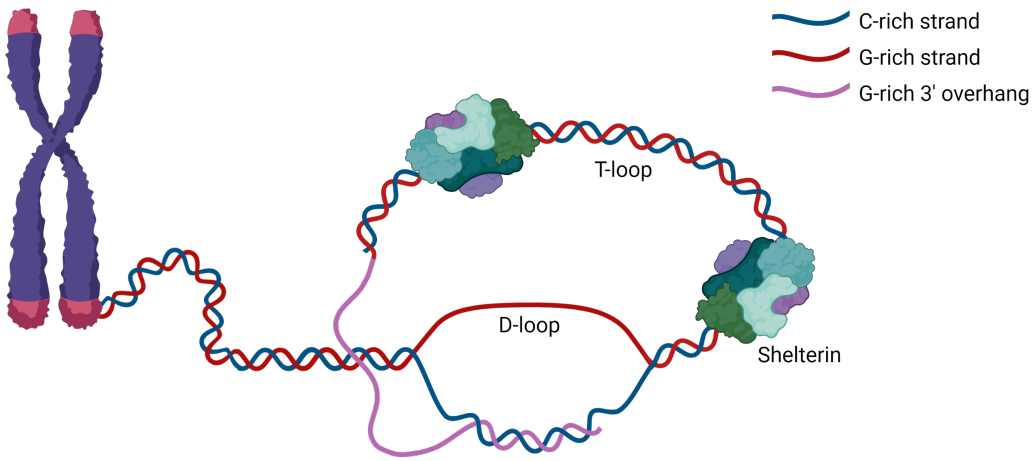

Figure 1.12: Telomere structure. Telomeres have a single-stranded 3' overhang that invades the double-stranded DNA to form a displacement (D)-loop and telomeric (T)-loop. The telomeres are protected by the shelterin complex and enriched in heterochromatin. Image created with BioRender.

undergo (Campisi and d'Adda di Fagagna, 2007). Unprotected telomeres are a source of genomic instability, putting the genome at risk of chromosome end breaks and fusions, leading to development of cancer (Murnane, 2012). Additionally, in order to gain infinite proliferative ability, cancer cells must allow for access to telomeres in order to elongate them through the expression of telomerase (Blackburn et al., 1989) by reverse transcriptase activity (Lingner et al., 1997) or recombination-mediated alternative lengthening of telomeres (ALT, Bryan et al. 1997).

The G-rich DNA strand extends further than the C-rich strand, resulting in a 3' overhang which has been shown to invade the homologous double-stranded region of the telomeres (Griffith et al., 1999). This telomeric loop (T-loop, **Figure 1.12**) prevents the end of the telomere from being recognised as a double-stranded break, protecting it from DNA damage machinery (Longhese, 2008). An additional layer of protection present at the T-loop junction is the shelterin complex, which includes POT1, Telomere Repeat Factors 1 and 2 (TRF1 and TRF2), TRF1 Interacting Nuclear protein 2 (TIN2), Repressor-Activator Protein 1 (RAP1), and Telomere Protection Protein 1 (TPP1, Diotti and Loayza 2011). These proteins act together to shelter both the double and single-stranded DNA, protecting telomere ends (De Lange, 2005). Loss of condensation leads to telomere length deregulation, disruption of telomeric silencing, and telomere recombination (Fanti et al., 1998; Blasco, 2007). Telomeres have been generally considered to be solely heterochromatic with telomeric DNA condensed constitutively; however, there is evidence that telomeres in cell lines carry chromatin marks associated with active chromatin (Cubiles et al.,

2018). Despite this, telomeric chromatin has been observed to have higher density and more compact nature than non-telomeric heterochromatin (Nikitina and Woodcock, 2004; Hübner et al., 2022; Soman et al., 2022).

1.2.1 Structures of telomeric DNA

Due to its G-rich sequence, mammalian telomeric DNA forms stable anti-parallel G4s in solution that are less favoured than double-stranded DNA but more favourable than single-stranded DNA (Lane et al., 2008; Kreig et al., 2015). These G4s are hypothesised to form in the free 3' overhang of the telomere, demonstrated by *in vitro* studies (Tang et al., 2008; Wang et al., 2011) and due to the increase in G4s during S-phase, may also form throughout the telomeres during replication (Biffi et al., 2013). Telomeric DNA containing at least four TTAGGG repeats (minimum to form an intramolecular G4) forms anti-parallel G4 structures (Wang and Patel, 1993; Xu et al., 2006; Ambrus et al., 2006; Luu et al., 2006; Phan et al., 2006, 2007; Lim et al., 2009), although telomeric DNA has been shown to form a parallel G4 topology under molecular crowding conditions (Xue et al., 2007) or in the presence of only K^+ as the stabilising cation (**Figure 1.13**, Parkinson et al. 2002).

G4 binding proteins that are associated with telomeric G4s include: POT1 as part of the shelterin complex (Horvath and Schultz, 2001; Hwang et al., 2012), Replication Protein A (RPA) involved in unfolding of G4s during replication and repair (Ray et al., 2014), human CTC1–STN1–TEN1 (CST) which also unwinds telomeric G4s (Bhattacharjee et al., 2017), Breast Cancer type 1 susceptibility protein (BRCA1) involved in DNA repair and DNA damage response (Brázda et al., 2016), heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1, Hudson et al. 2014), Preimplantation factor-1 (Pif1, Byrd and Raney 2015), Werner syndrome ATP-dependent helicase (WRN), and Bloom syndrome protein (BLM, Chatterjee et al. 2014; Sauer and Paeschke 2017). Despite lacking helicase activity, human telomerase is able to both recognise and partially unwind parallel telomeric DNA G4s, and localises to G4-containing telomeres *in vivo* (Moye et al., 2015), indicating a role of G4s in cancer immortalisation.

Complementary to the G-rich strand, the C-rich telomeric DNA repeat (CCCTAA) has been shown to form i-motif structures *in vitro* at low temperature at pH 7 (Zhou et al., 2010; Abdelhamid and Waller, 2020). As with all i-motif structures, the stable formation of telomeric i-motifs *in vivo* has been difficult to confirm. However, the recently developed i-motif structure-

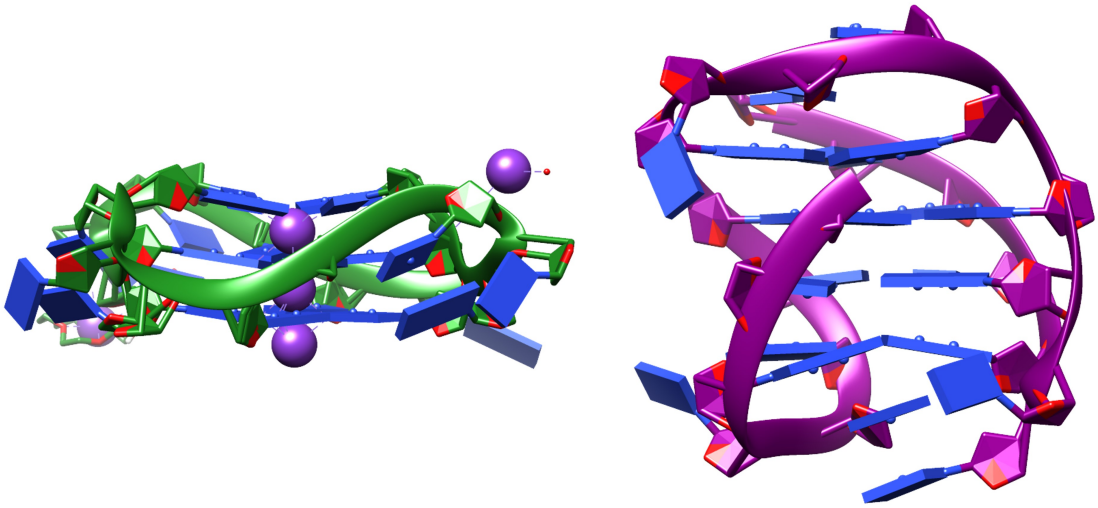

Figure 1.13: Telomeric DNA forms parallel and anti-parallel G4s. On the left is a crystal structure of a parallel G4 formed by 22 nt of telomeric DNA (PDB: 7KLP, Li et al. 2021), and on the right is an NMR structure of an anti-parallel G4 formed by 22 nt of telomeric DNA (PDB: 2KF8, Lim et al. 2009). Ribbons represent the DNA backbone, with pentagons representing the deoxyribose, blue rectangles representing bases, and purple spheres representing K^+.

specific antibody iMab, which allowed for the detection of i-motifs in the nuclei of human cells, also showed co-localisation of i-motifs with telomere-associated shelterin component TRF2 (Zeraati et al., 2018). Currently, it is unclear what implications telomeric i-motifs may have, and of particular interest is their possible formation in tandem with telomeric G4s (Wolski et al., 2018).

1.2.2 Telomeric repeat-containing RNA

Despite the heterochromatic nature of telomeres, in 2007 two research groups discovered a telomeric transcript (Azzalin et al., 2007; Schoeftner and Blasco, 2008) which is referred to as telomeric repeat-containing RNA (TERRA). Transcribed from CpG island-containing promoters in sub-telomeric regions (Nergadze et al., 2009), TERRA expression occurs from multiple chromosome ends in humans, including 1q, 2q, 7p, 9p, 10q, 13q, 15p, 17p, 17q, 18q, 20q, XpYp, and XqYq (Farnung et al., 2012; De Silanes et al., 2014; Montero et al., 2016; Feretzaki and Lingner, 2017). This repetitive lncRNA shares the G-rich nature of the telomeres with repeats of the r(UUAGGG) sequence, varying in length from 100 nt to 9000 nt, and due to the strings of

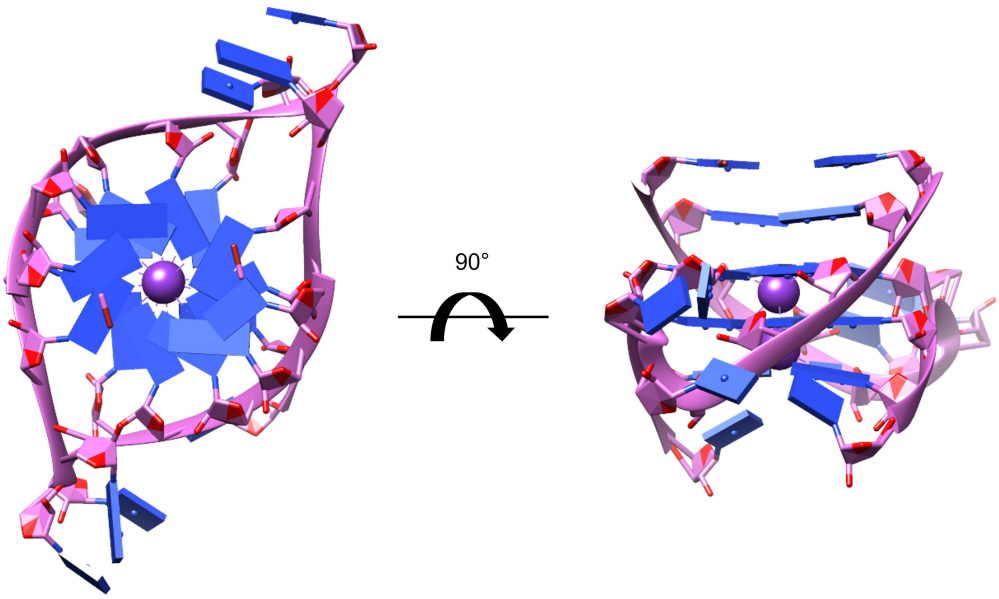

Figure 1.14: A crystal structure of TERRA. Structure of a parallel G4 formed by two 12 nt strands of TERRA (PDB: 3IBK, Collie et al. 2010). The image on the right has been rotated forward 90°.

guanines (Azzalin et al., 2007; Feretzaki and Lingner, 2017), TERRA is shown to form a parallel G4 structure (**Figure 1.14**, Li et al. 2021; Martadinata and Phan 2009; Collie et al. 2010; Bao et al. 2017).

This telomeric transcript associates with telomeres (Lai et al., 2013) but interestingly is also found at extratelomeric loci (Chu et al., 2017). TERRA transcription is regulated by the heterochromatin formation at subtelomeres by DNA methylation and histone H3 lysine 9 and histone H4 lysine 20 trimethylation marks (H3K9me3 and H4K20me3, Arnoult et al. 2012; Porro et al. 2014). TERRA expression has also been shown to be regulated by telomere length. A feedback loop is proposed, whereby long telomeres repress TERRA expression, but short telomeres allow transcription, which then brings about constitutive heterochromatin marks H3K9me3 and Heterochromatin Protein 1α (HP1α), thereby reducing expression and compacting telomeres (Arnoult et al., 2012; Montero et al., 2018). TERRA depletion also impairs telomere stability, leading to DNA damage and chromosomal fusions (Chu et al., 2017). In addition to this, TERRA has been shown to interact with a multitude of chromatin-related proteins and remodelling complexes such as HP1α and HP1β (Deng et al., 2009), histone methyltransferase Suppressor of Variegation 3(9) homologue H1 (SUV39H1, Porro et al. 2014), shelterin proteins such as TRF1

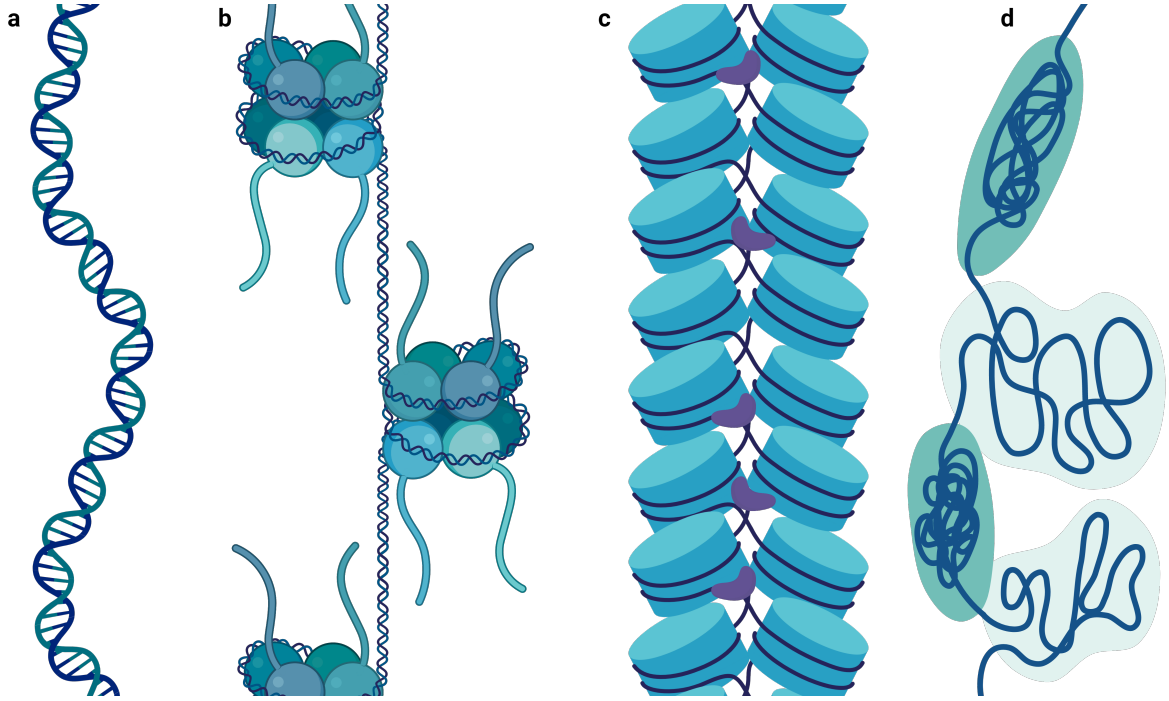

Figure 1.15: Levels of DNA folding. a) The double-stranded DNA helix. b) The 10 nm chromatin fibre structure. c) The 30 nm chromatin fibre with linker histone H1 (purple). d) TADs with compartment A (light green) and compartment B (dark green). Image created with BioRender.

and TRF2 (Deng et al., 2009), PRC2 through Enhancer of zeste homolog 2 (EZH2, Wang et al. 2017), and telomerase (Chu et al., 2017). TERRA is therefore critical for stabilising multiple interactions with chromatin components to promote the formation of protective heterochromatin to safeguard chromosome ends.

1.3 Functional organisation of the genome

The eukaryotic genome possesses an intricate and dynamic structure which needs to be tightly regulated within the nucleus. DNA needs to be densely packed, but still organised and accessible for replication, repair, and transcription. Therefore, the genome is packaged in the form of chromatin: a complex formed between DNA and proteins, and further partitioned into microscopically observed functional domains within the nucleus.

The first level of folding to achieve functional genome organisation is with the DNA double-helix. Double-stranded DNA (**Figure 1.15a**) is firstly formed into nucleosomes, where two copies of each core histone protein (H2A, H2B, H3, and H4) form the histone octamer (Luger et al., 1997),

with ~146 bp of DNA wrapping around this particle 1.65 times (Davey et al., 2002). The globular domains of the histones form the nucleosome core, while the positively charged histone tails extend outwards and neutralise the surrounding DNA (Luger et al., 1997). These nucleosomal subunits separated by 20-90 bp of linker DNA yield the 10 nm chromatin fibre structure (**Figure 1.15b**, Olins and Olins 1974). To fix nucleosomes in place, the linker histone H1 binds to the entry and exit sites of the DNA on the surface of the nucleosome core (Thoma et al., 1979) forming a chromatasome. Through the neutralisation of negative charge of the remaining linker DNA by histone H1, the 10 nm structure then folds into a 30 nm wide chromatin fibre (**Figure 1.15c**, Finch and Klug 1976). These chromatin fibres then coil and fold into loops by action of CCCTC-binding factor (CTCF) and cohesin (Haarhuis et al., 2017; Schwarzer et al., 2017) to form functionally discrete topologically associating domains (TADs) to compartmentalise active (compartment A, correlating with euchromatin) and inactive (compartment B, correlating with heterochromatin) chromatin within the nucleus (**Figure 1.15d**, Rao et al. 2014).

Chromatin is delineated into the two distinct domains of higher-order structure that can be microscopically observed: euchromatin and heterochromatin. Euchromatin is the lightly packed sub-type of chromatin, allowing for accessibility of the DNA for transcription machinery, therefore is associated with transcriptionally active genes. Heterochromatin is structurally compact, hindering access to the DNA and maintaining genomic integrity (Schneider and Grosschedl, 2007).

Variants of histones and covalent post-translational modifications (PTMs) of histones dictate the folding of chromatin. This is known as the "Histone Code" (Strahl and Allis, 2000), whereby histone tail PTMs (the "code") can be modified ("written") by acetyltransferases, methyltransferases and more, and recognised ("read") by proteins containing recognition domains such as bromodomains, chromodomains, plant homeodomain (PHD) fingers, and WD40 domains. The histone PTMs either directly influence the structure of chromatin, restricting access, or influence the further recruitment of chromatin-associated "reader" proteins (Bowman and Poirier, 2015). Acetylation of histones is generally associated with active euchromatin, whereby an acetyl group is transferred to a lysine side-chain, neutralising it, whereas methylation of histones is associated with both active and silenced heterochromatin, when lysine or arginine residues are multiply methylated, influencing binding factors such as HP1 (Bannister et al., 2001; Lachner et al., 2001) and altering nucleosome dynamics (Cosgrove et al., 2004; Cosgrove and Wolberger, 2005; Mersfelder and Parthun, 2006).

1.3.1 Heterochromatin

Compact heterochromatin is further demarcated into two sub-types: constitutive and facultative (Brown, 1966). Facultative heterochromatin is a functional, reversible form of heterochromatin whereby genes may be silenced in a temporal, or cell-specific manner, and is primarily marked by the histone PTM of trimethylated lysine 27 of histone H3 (H3K27me3, Jamieson et al. 2016) which results in repressed chromatin. Facultative heterochromatin is formed and maintained by Polycomb Repressive Complexes (PRC), which contain Polycomb group (PcG) proteins. PRC2 acts to di- or tri-methylate H3K27 through its enzymatic subunit EZH1/2, and is enhanced by PcG proteins EED (embryonic ectoderm development), SUZ12 (suppressor of zeste 12 homolog), and RBAP46 or RBAP48 (retinoblastoma protein associated proteins p46 and p48, Blackledge et al. 2015). PRC1 complexes utilise Chromobox (CBX) subunits to recognise H3K27me3 (Gao et al., 2012), further accumulating PcG proteins at target locations. The mechanism of *de novo* recruitment of PcG proteins occurs with DNA-binding transcription factors (Yu et al., 2012b; Arnold et al., 2013), CpG islands (Blackledge et al., 2014), or alternatively interactions with lncRNAs such as X inactive specific transcript (Xist) for the silencing of X chromosomes (Plath et al., 2003; Jamieson et al., 2016; Fischle et al., 2003). PRC2 preferentially binds RNA G4s and is enriched at their coding sequences, alluding to RNA-mediated PRC2 regulation at these sites (Wang et al., 2017).

Constitutive heterochromatin is maintained throughout the cell cycle and is formed in repetitive, gene-poor regions of the genome, particularly at the centromeres and telomeres. Centromeres are the central structure of the chromosomes where the kinetochore assembles and interacts with the spindle fibres. Therefore, to ensure faithful segregation during mitosis and meiosis (David et al., 2003), the centromeres and surrounding (pericentromeric) regions must be compacted in protective heterochromatin constitutively. As discussed previously (**Section 1.2**), telomeres are the protective region at linear chromosome ends (Olovnikov, 1973), and must also have their heterochromatin maintained constitutively to prevent DNA from being degraded or recognised as double-strand breaks. Constitutive heterochromatin formation is therefore critical for genomic stability within the nucleus, essential for telomere protection, chromosome segregation, DNA repair, and suppression of transposon activity, and has been referred to as "Guardian of the Genome" (Janssen et al., 2018). This domain of heterochromatin is marked by di- and trimethylation of H3K9 (Saksouk et al., 2015; Becker et al., 2016), and catalysed by the SUV39H

histone methyltransferase (Rea et al., 2000; Peters et al., 2001). It is proposed that the H3K9me3 modification allows for the enrichment of the repressive architectural protein HP1, which binds the histone H3 tail and oligomerises (Bannister et al., 2001), stabilising the chromatin structure and also acting as a scaffold for other heterochromatin elements (Kumar and Kono, 2020). SUV39H also binds this scaffold, promoting the spread of H3K9me3 into neighbouring nucleosomes through its methyltransferase activity, propagating condensed heterochromatin (Aagaard et al., 1999; Müller-Ott et al., 2014).

1.3.2 Heterochromatin Protein 1

HP1 proteins are a family of chromatin-associated non-histone proteins that propagate and maintain heterochromatin. First discovered in *Drosophila* as a dominant suppressor of position-effect variegation, HP1 expression resulted in the repression of a euchromatic gene when placed near or within heterochromatin (Eissenberg et al., 1990). A hallmark of heterochromatin, HP1 is well conserved in both sequence and function across fungi, plants, and vertebrates (Wang et al., 2000; Schoelz and Riddle, 2022). Given their roles in genome organisation, HP1 proteins function in DNA replication (Schwaiger et al., 2010), DNA repair (Goodarzi et al., 2008), transcriptional regulation (Eissenberg et al., 1990), chromosome segregation (Nonaka et al., 2002), telomere maintenance (Chow et al., 2018), and cell cycle regulation (De Lucia et al., 2005).

There are three mammalian HP1 paralogs: HP1α, HP1β, and HP1γ, encoded by the genes CBX5, CBX1, and CBX3 respectively. Despite the highly conserved amino acid sequences of the globular domains between paralogs of HP1 (**Figure 1.16a**), each differs in their localisation and specific function (Nishibuchi and Nakayama, 2014). HP1γ is distributed in euchromatin, and as such has been shown to associate with actively transcribed genes and has a role in transcriptional elongation (Vakoc et al., 2005; Lomberk et al., 2006). HP1β is localised to both heterochromatin and euchromatin, while HP1α solely localises to heterochromatin in repetitive regions (Nielsen et al., 1999; Minc et al., 1999, 2000). Each HP1 paralog is able to bind H3K9me3 (Nielsen et al., 2001), therefore, this heterochromatin-associated histone PTM is not the sole determinant of HP1 genomic localisation. This is highlighted by only partial overlap of localisation between H3K9me3 and *Drosophila* HP1a (Greil et al., 2003). The key differences between HP1 proteins lie in the less conserved disordered regions, resulting in differing binding to chromosomal proteins and/or associated nucleic acids (Schoelz and Riddle, 2022).

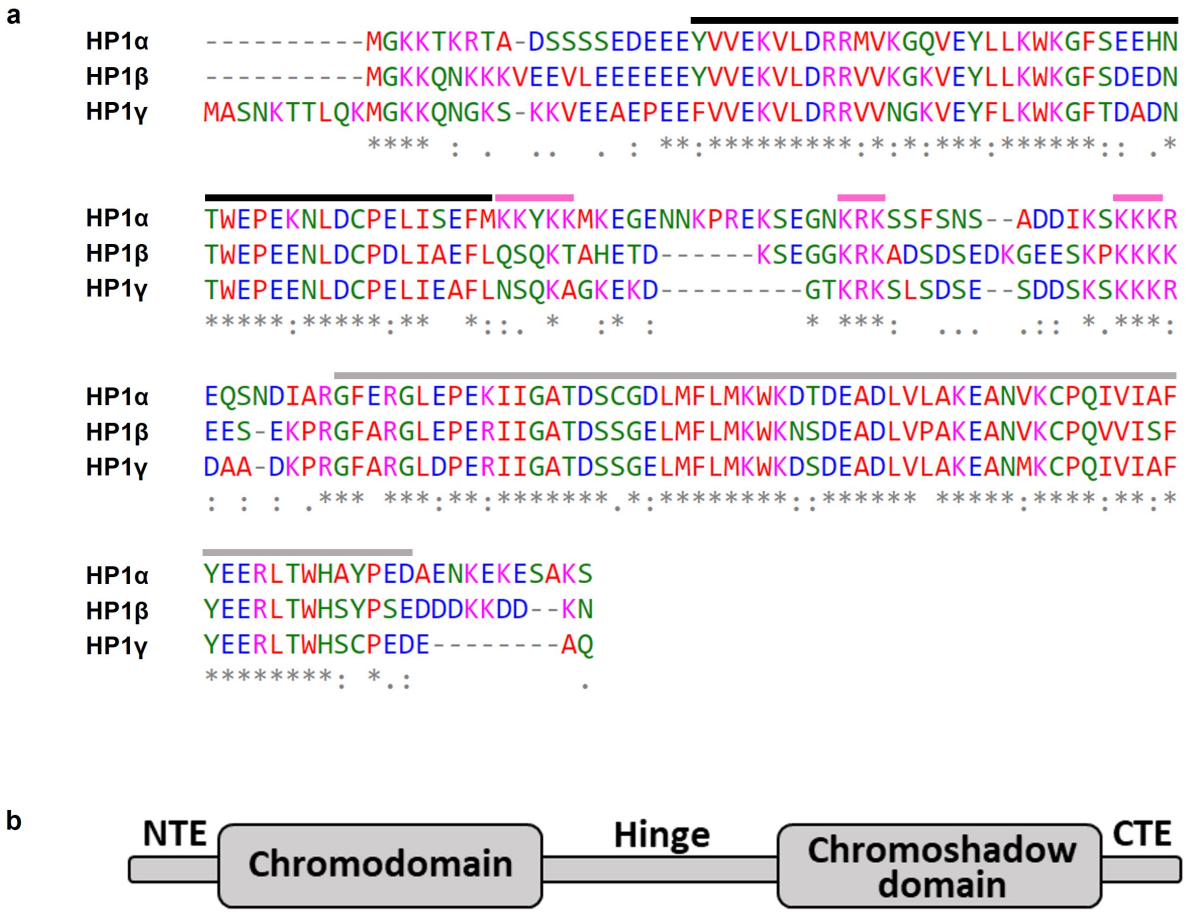

Figure 1.16: Alignment and domains of HP1 paralogs. a) Amino acid sequences of *Mus musculus* HP1α, HP1β, and HP1γ. The CD and CSD boundaries of HP1α are shown by black and grey lines, respectively. Positively charged amino acid patches in the hinge are shown by pink lines. An asterisk (*) indicates a fully conserved residue, a colon (:) indicates conservation of a residue with strongly similar properties, and a period (.) indicates conservation of a weakly similar residue. b) The globular CD and CSD of HP1α are surrounded by the intrinsically disordered NTE, hinge, and CTE.

1.3.2.1 HP1 structure

The structure of HP1 paralogs consists of two globular domains and three disordered regions surrounding them (**Figure 1.16b**). The chromodomain (CD) is a 48 residue domain that forms a globular structure consisting of three anti-parallel β-sheets and a single α-helix at the C-terminus (**Figure 1.17**, Shimojo et al. 2016). The CD is highly conserved between the mammalian paralogs with the human HP1α CD having 82% and 71% amino acid sequence identity to the CDs of HP1β and HP1γ, respectively (Canzio et al., 2014). The second globular domain of HP1 is the chromoshadow domain (CSD). As the name would suggest, the CSD shares structural conservation with the CD with three anti-parallel β-sheets, but with a helix-turn-helix at the C-terminus, and an additional α-helix at the N-terminus (**Figure 1.18**, Li et al.). Again, this globular domain is highly conserved between HP1 paralogs (human HP1α CSD has 82% and 87% amino acid sequence identity to the HP1β and HP1γ CSDs, respectively, Canzio et al. 2014). The N-terminus of HP1 consists of an unstructured extension (NTE), which contains positively and negatively charged amino acids, and ranging from 19-20 residues in mammalian HP1α and HP1β to 29 residues in HP1γ (**Figure 1.16a**). Flanked by the globular CD and CSD, the hinge of HP1 is another unstructured domain (Brasher et al., 2000), less conserved between the mammalian paralogs of HP1, but each containing at least two distinct patches of positively charged amino acids (**Figure 1.16a**). Between paralogs, the hinge ranges from 48 residues in HP1α, to 43 residues in HP1β, and 38 residues in HP1γ. Finally, the C-terminal extension (CTE) is also a short unstructured region of HP1, ranging from 11 residues in HP1α, to 9 residues in HP1β, and just 3 residues in HP1γ, and is rich in positively and negatively charged amino acids (**Figure 1.16a**).

Due to the unstructured domains, HP1 proteins are intrinsically disordered with dynamic flexibility (Munari et al., 2012; Velez et al., 2016). No full-length structures containing both the CD and CSD of any HP1 paralog have been resolved, let alone with the disordered tails.

1.3.2.2 The functional domains of HP1α

The CD is a conserved histone methyl-lysine recognition module (Ball et al., 1997; Eissenberg, 2001), and in HP1 proteins binds to H3K9me3 through hydrophobic residues (Nielsen et al., 2002; Bannister et al., 2001; Lachner et al., 2001). Due to the enrichment of both HP1α and

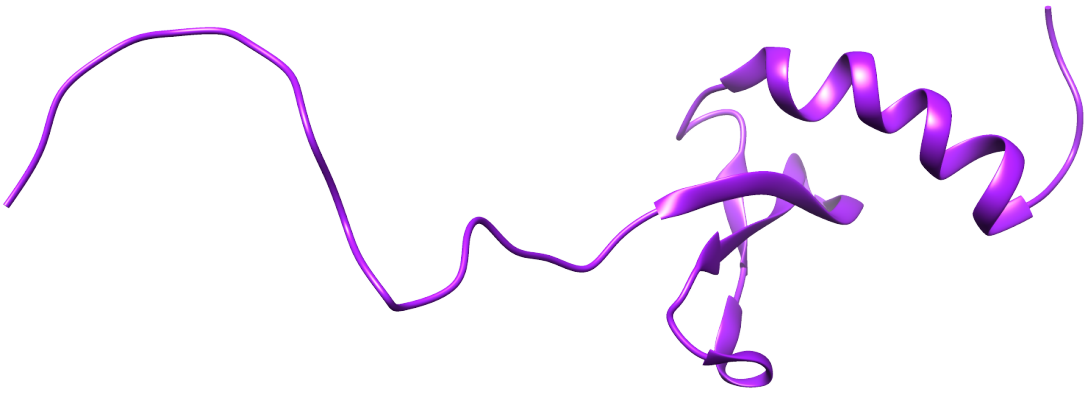

Figure 1.17: Structure of the NTE and CD of HP1. Model 1 of an NMR structure of the HP1α NTE (left) and globular CD (right). PDB:2RVL (Shimojo et al., 2016).

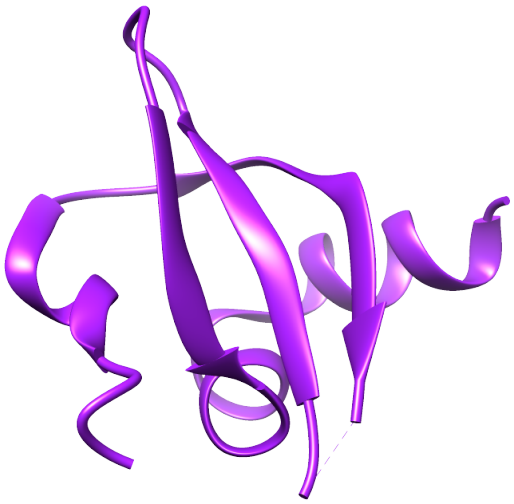

Figure 1.18: Structure of the CSD of HP1. A single chain of a crystal structure of the HP1α CSD, with the N-terminus on the left. PDB:3I3C (Li et al.).

H3K9me2/3 at constitutive heterochromatin domains, it was hypothesised that this interaction allows for the localisation of HP1α. However, in addition to partial overlap of localisation between H3K9me3 and HP1α (Greil et al., 2003), when measured *in vitro*, the HP1α without any PTM of the CD has low affinity to H3K9me3 (dissociation constant of 35 μM, Hiragami-Hamada et al. 2011), indicating that there are more factors *in vivo* that must contribute to HP1α nucleosome binding and heterochromatin localisation. Further to this, the binding of HP1α to nucleosomal arrays containing H3K9me3 was studied, showing a greater affinity of 1.16 μM, but also revealing a highly dynamic interaction between the CD and H3K9me3 with rapid HP1α rebinding dependent upon H3K9me3 density, CSD dimerisation (Kilic et al., 2015), and DNA binding by the hinge (Bryan et al., 2017).

The NTE of HP1α is a short unstructured domain, containing a tract of four consecutive serine residues unique to the HP1α paralog (**Figure 1.16a**). To aid in binding of the CD to H3K9me3, the NTE serine residues are phosphorylated by casein kinase II (CK2, Nishibuchi et al. 2014). This phosphorylation enhances the binding affinity between the HP1α CD and H3K9me3, conferring specificity towards H3K9me3 as opposed to unmethylated H3 (Hiragami-Hamada et al., 2011), while also increasing HP1α residence times on chromatin fibres (Bryan et al., 2017). In addition to this, NTE phosphorylation creates an open conformation of HP1α, where intra-HP1α contacts are relieved and the NTE and CD extend outwards (Larson et al., 2017). Phosphorylation of these NTE residues has also been shown to diminish DNA binding by HP1α (Nishibuchi et al., 2014).

The CSDs of each HP1 paralog are able to homo- and hetero-dimerise through one of the α-helices, with an additional potential dimerisation interface through a β-sheet shown by the crystal structure (Li et al., PDB: 3I3C). This dimerisation aids in the formation of higher-order multimeric complexes to assist in heterochromatin formation (Canzio et al., 2014). In addition to this, dimerisation of the CSD allows a hydrophobic binding site to form, facilitating interactions with other non-histone chromosomal proteins (**Figure 1.19**). Phage display identified a distinct pentapeptide motif (PXVXL, where X denotes any amino acid, Smothers and Henikoff 2000) that was required for interaction with the CSD dimer. This motif is found in HP1α binding partners Transcription Intermediary Factor (TIF)1-α and 1-β/Krüpple-associated box (KRAB)-associated protein 1 (KAP1, Nielsen et al. 1999), the large (p150) subunit of chromatin assembly factor 1 (CAF-1, Quivy et al. 2004), SUV39H (Aagaard et al., 1999) among many others. TIF1-β/KAP1 is a transcriptional regulator that associates with all five mammalian histone-lysine

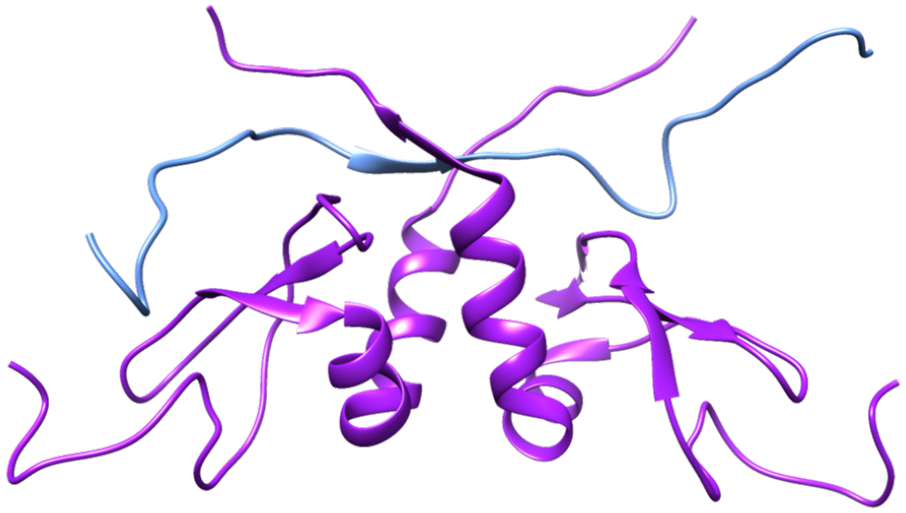

Figure 1.19: CSD dimer bound to a PXVXL-containing peptide. NMR structure of *Mus musculus* HP1β CSDs (purple) dimerising through an α-helix and binding the PXVXL motif of CAF-1 (blue, PDB: 1S4Z, Thiru et al. 2004).

methyltransferases and interacts with the HP1α CSD to ensure the maintenance and reinstatement of methylated H3K9 following replication (Jang et al., 2018). CAF-1 is a histone chaperone that ensures newly synthesised H3 and H4 are deposited into chromatin following DNA replication, and its interaction with the HP1α CSD is required for pericentromeric heterochromatin reinstatement (Quivy et al., 2008). Domain deletion mutations have shown that the HP1α CSD and hinge independently target heterochromatin loci (Smothers and Henikoff, 2001). If the CSDs of *Drosophila* HP1a and HP1c (analogous to mammalian HP1γ) are swapped, their genomic localisation is directly altered, with HP1a localising to euchromatin and HP1c localising to heterochromatin (Smothers and Henikoff, 2001), illustrating the involvement of the CSD in HP1α targeting.

Connected to the CSD, the CTE of HP1α is a short string of mainly alternating charged residues that is relatively un-studied. Despite its disorderedness, it does become more ordered when the CSD binds a PXVXL-containing peptide (Mendez et al., 2011). The HP1α CTE has also been shown to have extensive interactions with the hinge region, shown by mass spectrometry (MS, Larson et al. 2017).

Separating the CD and the CSD is the less conserved, highly hydrophilic, disordered hinge region. The hinge of HP1α contains a high concentration of positively charged amino acids (Qin

et al., 2021), with three distinct patches (**Figure 1.16a**). These positively charged amino acids within the hinge of HP1α are responsible for binding RNA (Muchardt et al., 2002; Meehan et al., 2003; Deng et al., 2009; Maison et al., 2011; Roach et al., 2020), binding nucleosomes through interaction with linker DNA (Mishima et al., 2013; Ryan and Tremethick, 2018), and driving liquid-liquid phase separation (LLPS, Her et al. 2022). While the HP1α hinge binds nucleosomal DNA non-specifically (Meehan et al., 2003), a preference has been shown for RNA over DNA; with a further preference for nuclear RNA over tRNA (Muchardt et al., 2002). Mutagenesis of these positively charged amino acids in the hinge by deletion or charge neutralisation results in loss of HP1α at heterochromatin foci (Muchardt et al., 2002). Wang et al. (2000) identified a nuclear localisation signal (NLS) in the hinge of the *Schizosaccharomyces pombe* HP1 homolog Swi6 (which binds to H3K9me3 and is involved in heterochromatin silencing, Grewal and Jia 2007), which is conserved in the mammalian HP1α, consisting of two basic amino acid patches separated by a spacer (Smothers and Henikoff, 2001; Kosugi et al., 2009). Each mammalian HP1 paralog contains at least two patches of positively charged amino acids in their hinges, but HP1α contains a unique third patch at the N-terminus of the hinge (**Figure 1.16a**). In addition to these charged patches, the HP1α hinge is also phosphorylated, which has been shown to be essential for accurate chromosome alignment and mitotic progression *in vivo* by potentially localising HP1α to kinetochores (Chakraborty et al., 2014).

HP1α has 48 documented PTMs spanning across each domain (**Figure 1.20**), which include phosphorylation, acetylation, methylation, ubiquitination, and SUMO (small ubiquitin-like modifier)-ylation (LeRoy et al., 2009). While phosphorylation of the NTE and hinge have been studied, the functional roles of acetylation and methylation of HP1α remain unaddressed, although it has been proposed that together they act as a switch to alter HP1α binding (LeRoy et al., 2009; Sales-Gil and Vagnarelli, 2020). The HP1α CSD can be ubiquitinated by Rad6, promoting its degradation to form an open chromatin state to enable DNA repair (Chen et al., 2015), while the hinge can be SUMOylated with SUMO-1, which is shown to be crucial for its targeting to pericentric heterochromatin (Maison et al., 2011).

The HP1α paralog is integral for constitutive heterochromatin maintenance and displays unique DNA compaction behaviour and LLPS for sequestration of compacted chromatin in the nucleus (Larson et al., 2017). In addition to this, HP1α has altered expression in lung (Yu et al., 2012c), colon (De Lange et al., 2001), and breast cancers, with differential expression between non-cancerous and cancerous cells (De Koning et al., 2009). HP1α expression is reduced in

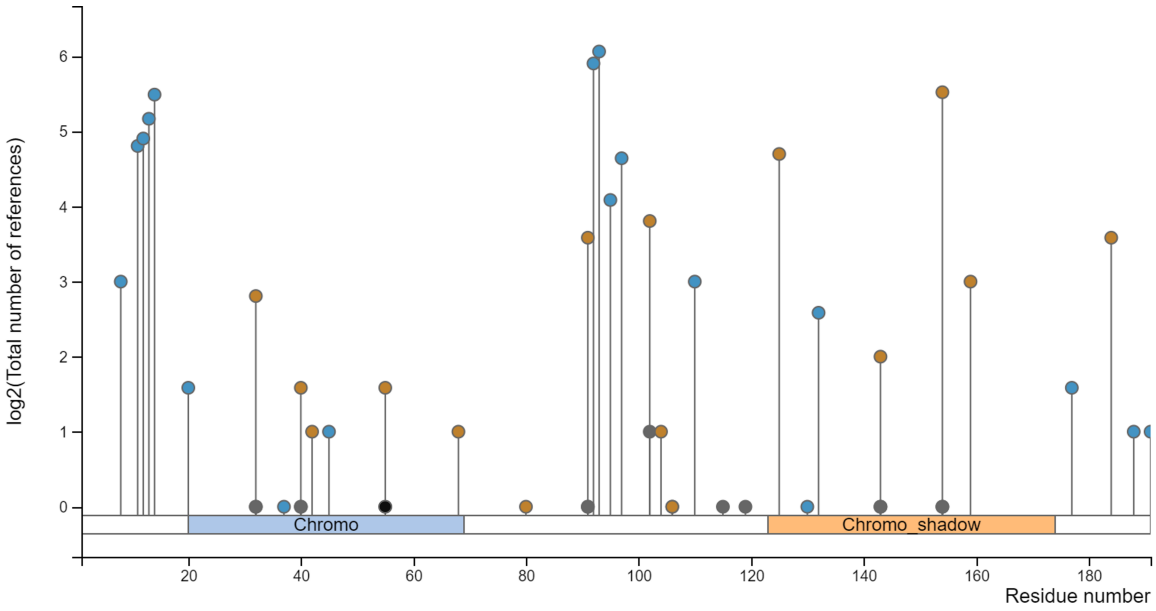

Figure 1.20: HP1α post-translational modifications. Depiction of the location of PTMs and their corresponding number of publications associated with them. Phosphorylation sites are indicated by blue dots, acetylation black dots, ubiquitination orange dots, and other PTM sites are indicated by grey dots. Graph and data obtained from PhosphoSitePlus (Hornbeck et al., 2015).

metastatic carcinoma samples (Kirschmann et al., 2000), resulting in higher invasive potential of breast cancer cells (Tretiakova et al., 2014), due in part to altered mechanical nuclear integrity (Solomon, 2019; Pradhan et al., 2021). HP1α is therefore a focus for studying constitutive heterochromatin for genomic stability.

Together, the domains of HP1α act to establish and maintain heterochromatin. The classical model dictates that the CD binds H3K9me3, and dimerisation through the CSD links together HP1α molecules either within a nucleosome or bridging dinucleosomes, acting as a scaffold for chromatin-modifying enzymes that contain the PXVXL motif (**Figure 1.21**). While HP1α and heterochromatin propagation/spreading is explained by this model, the *de novo* targeting of HP1α is not accounted for and ignores the role of the hinge and RNA in HP1α localisation.

1.4 RNA in HP1α function

In constitutive heterochromatin, one of the most well-studied models of an RNA-regulated system is RNA interference (RNAi) in *S. pombe*, whereby pericentromeric RNA is transcribed and subsequently processed by Dicer into short interfering RNA (siRNA). The nascent pericen-

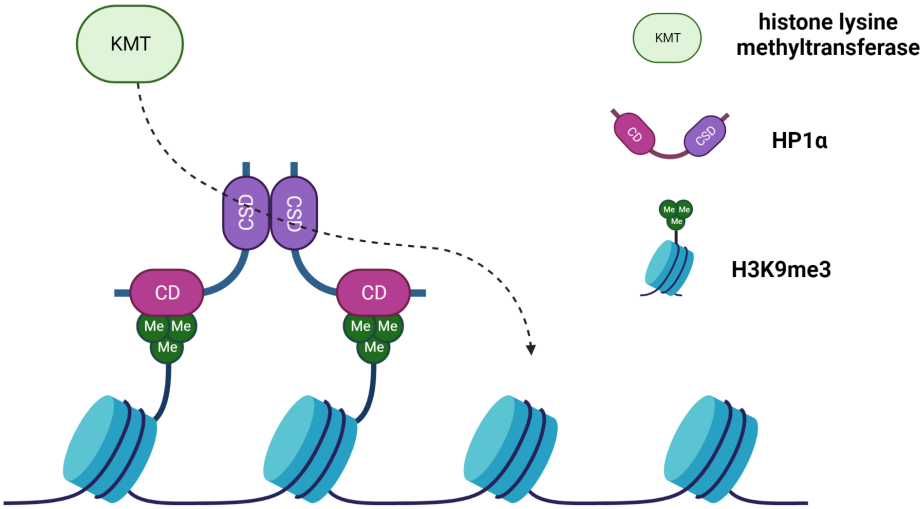

Figure 1.21: Model of heterochromatin formation by HP1α. Model of HP1α-mediated heterochromatin formation, whereby the CD binds H3K9me3 within nucleosomes, and dimerises through the CSD creating a binding site for chromatin-modifying enzymes such as histone lysine methyltransferases, which then modify adjacent nucleosomes. Image created with BioRender.

tromeric RNA and siRNA act as a scaffold for the recruitment and assembly of the RNA-induced transcriptional silencing (RITS) complex, which then recruits Clr4/SUV39 and Swi6 to result in heterochromatic silencing (Holoch and Moazed, 2015). The hinge of Swi6 binds RNA (Keller et al., 2012), and recently, an RNA aptamer was used to block hinge-nucleic acid interactions *in vivo* in *S. pombe*, which resulted in less enrichment of Swi6 at pericentromeres and more enrichment at euchromatic loci (Rani et al., 2019).

In mammalian cells, HP1α is recruited to centromeric and telomeric constitutive heterochromatin (Bosch-Presegue et al., 2017; García-Cao et al., 2004); therefore, RNA is also investigated for its role in HP1α-mediated silencing in this system. When cells are treated with RNase, HP1α loses its ability to localise to heterochromatin (Muchardt et al., 2002), showing that RNA is also essential for this process in mammals.

Within the centromeres and pericentromeric regions there are two distinct tracts of repetitive DNA referred to as the major satellite repeat (~6,000,000 bp of 234 bp units) and the minor satellite repeat (~600,000 bp of 120 bp units, Choo 1997; Guenatri et al. 2004). Transcription of major and minor satellite repeats (Rudert et al., 1995; Lu and Gilbert, 2007) has been shown to aid in the regulation and structure of centromeric heterochromatin (Liu et al., 2015). Mai-

son et al. (2011) found that HP1α when SUMOylated interacted with the forward orientation of major satellite-repeat RNA (fMSR) through the hinge region. The SUMOylation of HP1α was shown to occur at six lysine residues in the hinge (K69, K71, K72, K84, K104, and K106) and proved to be critical for the initial targeting of HP1α to pericentromeric heterochromatin. Therefore the hinge-RNA interaction is essential for *de novo* localisation of HP1α to the centromeres, while subsequent H3K9me3 marks deposited by SUV39H1 are required to stabilise the HP1α interaction with heterochromatin (Maison et al., 2011).

At the telomeres, TERRA is transcribed and involved in telomere maintenance activities (Bettin et al., 2019). When investigating the interaction of TERRA with heterochromatin components at telomeres, Deng et al. (2009) showed an interaction between HP1α and also HP1β with TERRA by immunoprecipitation, suggesting that TERRA associates with the heterochromatin maintenance assemblies at chromosome ends, and hypothesised that this interaction may be important for HP1-mediated heterochromatin formation at telomeres. For over a decade, the interaction between HP1α and TERRA had not been investigated further. Recently, the interaction between TERRA and HP1α was elaborated upon prior to the work in this thesis. It was found that HP1α bound with high affinity to TERRA (at lengths of 22 nt, 45 nt and 96 nt) and this interaction was dependent on a lysine patch (104-106KKK) in the hinge. Interestingly, HP1α also interacted with a range of other parallel G4s including DNA from the promoter regions of the c-myc proto-oncogene and the anti-apoptotic B-cell lymphoma-2 (BCL2) gene. However, HP1α did not bind to the anti-parallel G4 structure formed by the G-rich telomeric DNA, and other anti-parallel G4s either did not bind at all or bound with a similar or slower association rate and a faster dissociation rate (Roach, 2019). *In situ* experiments performed with mouse NIH3T3 fibroblasts showed that anti-parallel G4s did not compete for chromatin-bound HP1α, whereas incubation with TERRA or parallel DNA G4s did compete for HP1α binding (Roach et al., 2020). This indicates that the secondary structure of the nucleic acid provides specificity for the interactions of HP1α.

1.5 Hypothesis and objectives

The targeting of essential chromatin regulator HP1α to regions of constitutive heterochromatin at centromeres and telomeres is imperative for genomic stability. RNA is shown to be an indispensable part of this process, and the RNA transcribed from the telomeres, TERRA, has been shown to interact with HP1α. Therefore, to target HP1α to telomeric heterochromatin, it is hypothesised that HP1α has a distinct mechanism of recognition for the TERRA parallel G4 assembly. The recognition of specific nucleic acid secondary structures by HP1α may provide specificity for its functions in the cell.

The present study will investigate the interaction of HP1α with the TERRA G4 with the following aims and objectives:

1. Establish the means of interaction between HP1α and TERRA
    (a) Determine if TERRA sequence or structure contributes to HP1α binding
    (b) Ascertain amino acids responsible for the HP1α-TERRA interaction
2. Examine HP1α specificity towards TERRA
    (a) Establish the specificity of HP1α for structured nucleic acids
    (b) Determine the structural motifs of HP1α responsible for specificity for TERRA
3. Investigate structures of HP1α and TERRA
    (a) Determine the structures of HP1α and TERRA individually in solution
    (b) Examine changes in HP1α structure upon binding TERRA

Chapter 2

Materials and Methods

2.1 Oligonucleotides

Unless otherwise stated, oligonucleotides used in experiments in this study were obtained from Integrated DNA Technologies, purified by standard desalting by the manufacturer. Sequences are listed in **Table A.1**. A 20% native polyacrylamide gel of TERRA45 (and TERRA96) is shown in **Figure A.1**.

Oligonucleotides were folded in interaction buffer (IB1, 100 mM KCl, 50 mM NaCl, and 20 mM NaH_2PO_4, pH 8) or LiCl IB (100 mM LiCl, 70 mM Tris base, pH 8) by heating at 90 °C for 5 minutes, then cooling gradually to room temperature.

Primers used for mutagenesis were also obtained from Integrated DNA Technologies, purified by high-performance liquid chromatography (HPLC).

2.2 Mutagenesis of HP1α

2.2.1 Site-directed mutagenesis

The coding regions for mouse HP1α were previously subcloned from pGEX2T (Hale et al., 2006) into the vector pET24a for expression. A comparison of the amino acid sequences of human and mouse HP1α is shown in **Figure A.2**.

Where substitutions were required, the QuikChange II Site-Directed Mutagenesis Kit (Agilent) was used for mutagenesis of HP1α. Polymerase chain reaction (PCR) mixes were made containing 1x QuikChange reaction buffer, 1 µL of QuikChange dNTP mix, 50 ng of pET24a his-HP1α vector, 125 ng of forward primer, 125 ng of reverse primer, 1 µL of PfuUltra HF DNA polymerase (2.5 U/µL), and water to 50 µL. Primers used are shown in **Table A.2**.

Cycling parameters for the QuikChange II Site-Directed Mutagenesis consisted of an initial 95 °C step for 30 seconds, then cycles of 95 °C for 30 seconds, 55 °C for 1 minute, and 68 °C for 6 minutes. The number of cycles was determined as per **Table 2.1**.

Table 2.1: Cycle number determination for site-directed mutagenesis PCR

Type of mutation	Number of cycles
Point mutations	12
Single amino acid changes	16
Multiple amino acid changes or insertions or deletions	18

Following PCR, 1 µL of QuikChange DpnI restriction enzyme (10 U/µL) was added to the reaction and incubated at 37 °C for 1 hour to digest parental DNA. The presence of undigested PCR product was then assessed by agarose gel and staining with ethidium bromide as per **Section 2.3**.

Of the digested PCR reaction, 1-4 µL was used to transform 50 µL of XL1-Blue supercompetent cells (Agilent), which were incubated on ice for 30 minutes, heat shocked at 42 °C for 45 seconds, and incubated on ice for a further 2 minutes. Sterile Super Optimal Broth was then added before incubating cells at 37 °C for 1 hour. Cells were centrifuged at 3,000 ×g for 1 minute, and the pellet resuspended in 100 µL before having 50 µL aliquots spread onto agar plates containing the appropriate antibiotic and incubated overnight at 37 °C.

2.2.2 Inverse PCR

For mutations that required deletions of portions of HP1α, inverse PCR was performed. PCR reactions were made containing 20 ng of pET24a his-HP1α vector, 0.3 µM each of forward and reverse primers, 1x KAPA HiFi buffer, 0.3 mM KAPA dNTP mix, and 1 unit of KAPA Hifi Hotstart polymerase in 50 µL. Primers used are shown in **Table A.3**.

Cycling parameters for inverse PCR consisted of an initial 95 °C step for 3 minutes, then cycles of 98 °C for 20 seconds, 60 °C for 15 seconds, and 72 °C for 6 minutes for 25 cycles. PCR products of correct size were confirmed by agarose gel electrophoresis and staining with ethidium bromide as per **Section 2.3**. Following this, PCR reactions were subjected to the QIAquick PCR purification kit and eluted by the addition of 30 µL of elution buffer. BlueJuice (Invitrogen) was then added to a final concentration of 1x, and the reaction was electrophoresed by agarose gel. Bands of the correct size were then excised using a scalpel under UV light, subjected to the QIAquick Gel Purification kit (Qiagen), and again eluted with 30 µL of elution buffer.

A kinase reaction was then performed with the addition of 10 mM adenosine triphosphate (ATP, Thermo Fisher), 20 Units of T4 Polynucleotide Kinase (Thermo Fisher), and 1x of the accompanying reaction buffer A with water up to 40 µL. This mixture was incubated at 37 °C for 20 minutes before being heat inactivated by incubation at 75 °C for 10 minutes. The kinase reaction was then subjected to the QIAquick PCR Purification kit (Qiagen), and eluted with 30 µL of elution buffer.

Blunt end ligation reactions were then set up containing 100 ng of the phosphorylated vector, 1 µL of T4 DNA Ligase (NEB), 1x of the accompanying ligase buffer, and water up to 10 µL. This reaction was incubated at room temperature for 2 hours before being heat inactivated by incubation at 65 °C for 10 minutes. Of the ligation reaction, 5 µL was then used to transform DH5α cells as previously described (**Section 2.2.1**).

2.2.3 Plasmid purification

For both site-directed mutagenesis and inverse PCR, individual transformed colonies were used to innoculate 5 mL cultures of Lysogeny Broth (LB) containing 50 µg/mL kanamycin and incubated overnight at 37 °C with shaking at 200 rpm. Cultures were then subjected to plasmid DNA purification using the QIAprep Spin Miniprep Kit (Qiagen) as per the manufacturer's instructions and eluted with 30 µL of elution buffer.

The quality and quantity of DNA were assessed using a DeNovix DS-11 FX Microvolume Spectrophotometer by measuring absorbance at 260 nm. Correct mutagenesis products were confirmed by sequencing performed at the Massey Genome Service using T7 promoter and terminator primers.

2.3 Agarose gel electrophoresis

For cloning purposes, agarose gels were prepared with 1% agarose (w/v) dissolved in 1x Tris-acetate-ethylenediaminetetraacetic acid (EDTA) buffer (TAE, 40 mM Tris base, 20 mM acetic acid, 1 mM EDTA) by microwaving with occasional swirling. Ethidium bromide was then added to the agarose solution at a final concentration of 0.5 µg/mL.

Agarose gels were electrophoresed with 1x TAE with 0.5 µg/mL ethidium bromide at 100 volts for 90 minutes and subsequently imaged under UV light using a Uvidoc HD6 (Uvitech Cambridge).

2.4 Polyacrylamide gel electrophoresis

2.4.1 SDS-PAGE

For sodium dodecyl sulfate (SDS) Polyacrylamide Gel Electrophoresis (PAGE), a 12% 29:1 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% 29:1 acrylamide/bis-acrylamide, 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.

Loading buffer was added to a final concentration of 1x (62.5 mM Tris pH 6.8, 1.67% SDS, 7.5% glycerol, 0.01% bromophenol blue, 35.75 mM 2-Mercaptoethanol) to each sample, then heated to 99 °C for 5 minutes prior to loading. 3 µL of Precision Plus Protein Dual Color Standards ladder (Bio-Rad) was loaded alongside samples where marker is shown. Electrophoresis was performed in 1x Tris-glycine-SDS buffer (TGS, 25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3) at a constant 35 mA until the dye front reached the bottom of the gel.

Gels were fixed in a solution of 40% ethanol and 10% acetic acid in Milli-Q water for 15 minutes, and then stained with QC Colloidal Coomassie Stain (Bio-Rad) for 10-20 hours, and de-stained in Milli-Q water until satisfactory contrast was obtained. Gels were then imaged on a GelDoc under white light.

2.4.2 Electrophoretic mobility shift assay

Binding reactions were performed in a final volume of 20 µL using 10 pmol of 3' TAMRA-labelled TERRA45 and 20-fold or 40-fold molar excess of protein in a binding buffer containing 50 mM Tris/HCl pH 7.5, 100 mM KCl, and 10 mM Tris(2-carboxyethyl)phosphine (TCEP). Unlabelled competitor oligonucleotides were added in excess (as indicated) over TAM-TERRA45, and LightShift Poly dI-dC (Thermo Scientific) was added to a final concentration of 50 ng/µL. Reactions were incubated at room temperature for 1 hour before the addition of glycerol to a final concentration of 7% prior to loading into the gel.

Native polyacrylamide gel Electrophoretic Mobility Shift Assays (EMSAs) were performed using polyacrylamide gel prepared with 4% 19:1 acrylamide/bisacrylamide, 0.1% APS, 0.1% TEMED, with 0.5x Tris-borate-EDTA buffer (TBE, 22.5 mM Tris-borate, 0.5 mM EDTA) with addition of 50 mM KCl, and electrophoresed in 0.5x TBE for 1 hour 15 minutes at 140 volts. Agarose EMSAs were prepared with 1% agarose (w/v) dissolved in 0.5x Tris-borate buffer (TB, 45 mM Tris base, 45 mM boric acid) and electrophoresed in 0.5x TB at 200 volts for 25 minutes.

Gels were then imaged using an FLA-5000 Fluorescent Image Analyser (Fujifilm) at 532 nm, at a voltage of 400 volts.

2.5 Protein expression and purification

To 25 µL aliquots of competent BL21 DE3 cells, 100 ng of the pET24a vector (containing HP1α cDNA with an N-terminal hexahistidine (his)-tag and TEV site) were added. The cells were incubated on ice for 30 minutes, heat shocked at 42 °C for 45 seconds, and incubated on ice for a further 2 minutes. Sterile Super Optimal Broth was then added before incubating cells at 37 °C for 1 hour, and then plated on LB agar with 50 µg/mL of kanamycin to be incubated overnight at 37 °C. A single colony was then used to innoculate 5 mL of LB with 50 µg/mL of kanamycin, and this culture was incubated overnight at 37 °C. Of this overnight culture, 1 mL was then used to innoculate 50 mL of LB with 50 µg/mL of kanamycin, which was grown at 30 °C with shaking at 180 rpm until the optical density at 600 nm reached an absorbance of 0.4. After chilling cultures at 4 °C for 15 minutes, Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.4 mM to induce protein expression overnight at 22 °C.

Cultures were then centrifuged at 4,000 ×g for 10 minutes at 4 °C. The supernatant was decanted, and cells were resuspended in wash buffer (10 mM Tris HCl pH 8, 100 mM NaCl). Resuspended cells were then centrifuged at 3,000 ×g 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 ULTRA EDTA free (Roche), 1 mM Tris HCl pH 8, 1 mg/mL lysozyme). After two freeze-thaws, 10 µL of TURBO DNase (Invitrogen) was added and incubated at 4 °C for 30 minutes, then the lysate was centrifuged at 10,000 rpm for 10 minutes. The supernatant was then filtered using a 0.45 µm filter, and imidazole was added to a final concentration of 20 mM.

The lysate was loaded onto a column containing 10 mL of HisPur nickel nitrilotriacetic acid (Ni-NTA) resin (Thermo Scientific) in a column attached to a ÅKTA Prime Plus (GE Healthcare) instrument at a flow rate of 0.5 mL/min. The column was firstly washed with equilibration buffer (20 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, pH 7.5), and then washed with wash buffer (20 mM sodium phosphate, 300 mM NaCl, 50 mM imidazole, pH 7.5), then finally elution was performed with elution buffer (20 mM sodium phosphate, 300 mM NaCl, 300 mM imidazole, pH 7.5).

Selected elution fractions were pooled and concentrated using Vivaspin 20 (5 kDa MWCO) ultrafiltration devices (GE Healthcare) before Size Exclusion Chromatography (SEC) was performed with a Superdex 200 10/300 GL column (GE Healthcare) attached to an NGC Chromatography System (Bio-Rad). 250 µL of the eluate was loaded into a 500 µL sample loop, and injected onto the column at 0.3 mL/min with IB1. Fractions were then concentrated using Vivaspin 500 (5 kDa MWCO) ultrafiltration devices.

Polyacrylamide gels of HP1α WT and mutants are shown in the appendix (**Figures A.3-A.31**).

2.6 Multiple sequence alignment

Multiple sequence alignment of amino acid sequences of HP1 paralogs was performed using the MAFFT (Multiple Alignment using Fast Fourier Transform) sequence alignment tool, version 1.2.4, with default parameters (Katoh et al., 2002).

2.7 Circular dichroism spectroscopy

Folded oligonucleotides were prepared at 5 µM in IB1 and measured in a 1 mm path-length cell. Proteins were prepared at 1.5 mg/mL in IB1 and measured in a 0.1 mm path length cell for far UV measurements, and in a 1 mm cell for near UV measurements. Circular dichroism spectra were recorded using a Chirascan (Applied Photophysics Ltd) instrument. Scans were gathered at standard sensitivity, data pitch 1 nm, continuous scanning mode, response 0.25 s, and bandwidth 1 nm. IB1 was scanned as a reference sample and its spectrum was subtracted from the average of three scans for each sample. Each spectrum was smoothed by averaging neighbour points using software provided by Applied Photophysics Ltd.

2.8 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 IB1 or LiCl IB. 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 (1 minute) was performed in 250 µL of IB1 within a black opaque 0.5 mL tube, then 4 µL of 100 µg/mL (or molar equivalents of mutant) his-tagged protein was incubated with the Ni-NTA tip for 5 minutes. The biosensor tip was then washed in 250 µL of IB1 for 1 minute to remove any protein not bound to the tip. The tip was then incubated with 4 µL of 500 nM or 2 µM folded oligonucleotide for 5 minutes. Finally, dissociation was measured by washing in 250 µL of IB1 for 5 minutes.

Reference experiments, where IB1 was added in the place of the 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 the dissociation of protein from the tip. No oligonucleotide was shown to bind the Ni-NTA biosensor tip in the absence of his-tagged protein. BLItz Pro 1.2 software was used for curve fitting and kinetics calculations using a 1:1 model. The error values for each BLI experiment represent the standard error of the fit.

2.9 Peptide pulldown

A slurry of 20 µL of Streptavidin agarose beads (69203, Millipore) or 12 µL Dynabeads MyOne Streptavidin C1 beads (65001, Invitrogen) were washed three times in Phosphate-Buffered Saline (PBS, 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) by centrifuging at 500 ×g, removing supernatant, and resuspending in PBS. Beads were then resuspended to their original volume in Bead Binding Buffer (BBB, 100 mM KCl, 50 mM NaCl, and 20 mM NaH_2PO_4, 1% Triton X-100, pH 8) and separated into 25 µL aliquots in 1.5 mL tubes.

To each aliquot of beads, 25 µL of 10 µg/mL of biotinylated 21 amino acid Histone H3 peptide (Novus Biologicals) or Histone H3 Trimethyl Lys9 peptide (Novus Biologicals) were added and incubated for 10 minutes. 300 pmol of biotin was then added and incubated for a further 10 minutes.

In a separate tube, 70 pmol of HP1α alone or with 70 pmol of TERRA45 or mutTERRA45 in BBB were incubated at room temperature for 1 hour. These samples were then added to the bead-peptide-biotin mix and incubated on an IKA VXR Vibrax at a setting of 200 for 1 hour at room temperature.

Mixtures were then washed twice in 500 µL of BBB by centrifugation, before resuspending the beads in 20 µL of 2x SDS-PAGE dye. The total reaction was then loaded onto a 12% SDS polyacrylamide gel (see **Section 2.4.1**). This gel was then electrophoresed at constant 35 mA until the dye front reached the bottom of the gel and subjected to western blotting (**Section 2.10**).

2.10 Western blotting

Proteins in SDS-PAGE gels were transferred to a 0.2 µm nitrocellulose membrane using Towbin buffer (25 mM Tris, 192 mM glycine, 20% methanol) at 60 mA at 4 °C for 2 hours. The membrane was stained with Ponceau S staining solution for 5 minutes to check for protein transfer, washed in Milli-Q water, and then blocked in 5% non-fat milk (NFM, Anchor) in 1x Tris-buffered saline plus Tween-20 (TBST, 20 mM Tris, 150 mM NaCl, pH 7.6, 0.1% Tween-20) for 1 hour at room temperature, with rocking at 30 rpm.

The membrane was then incubated with primary antibody (see **Table 2.2**) diluted in 5% NFM 1x TBST overnight at 4 °C rocking at 30 rpm. The membrane was then rinsed and washed four times for 5 minutes on an orbital shaker at 50 rpm in 1x TBST.

Table 2.2: Antibodies used in western blotting experiments

Antibody name	Catalogue number	Manufacturer	Host species	Dilution
HP1α antibody	2616	Cell Signalling	Rabbit, polyclonal	1:1,000
Rabbit IgG, HRP-linked whole Ab	NA943	Cytiva	Donkey	1:10,000

The membrane was then incubated in secondary antibody (**Table 2.2**) diluted in 5% NFM 1x TBST for 1 hour at room temperature, with rocking at 30 rpm. The membrane was again rinsed and washed four times for 5 minutes on an orbital shaker at 50 rpm in 1x TBST.

The blot was then developed using ECL western blotting detection reagent (Amersham) for 5 minutes and imaged using the chemiluminescence function on an Azure 600. Band densitometry was performed using ImageLab 6.1.

2.11 Mammalian cell culturing

Cell lines used in this study were NIH3T3 and MDA-MB-231. NIH3T3 cells were grown in Gibco Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% Gibco Bovine Serum with addition of 1% Gibco Penicillin-Streptomycin. MDA-MB-231 cells were grown in DMEM supplemented with 10% Gibco Fetal Bovine Serum and 1% Gibco Penicillin-Streptomycin.

Both cell lines were routinely maintained in 75 cm^2 flasks at 37 °C with 5% CO_2 and passaged once every three days. Media was removed, and cells were washed twice with the addition of 10 mL PBS, then detached by addition of 1.5 mL Trypsin-EDTA (Gibco, 0.05%). Trypsinisation was then halted by the addition of 8.5 mL of appropriate medium, before being passaged at a suitable dilution into a new flask.

2.12 Immunofluorescence

Prior to seeding cells, round glass coverslips were treated with 20 µg/mL of fibronectin in 6 well plates. To each well, cells at a density of 0.4 x10^5 cells/mL were added and incubated at 37 °C with 5% CO_2 overnight.

To 200 µL of Gibco Opti-MEM Reduced Serum Media, 1 µg of pcDNA3.1 mHP1α-FLAG WT or mutant and 1 µg of empty vector (pcDNA3.1) was added. 3 µL of X-tremeGENE HP DNA Transfection Reagent (Roche) was then added to each tube, mixed gently, and incubated for 15 minutes at room temperature. Transfection mixtures were then added dropwise to wells of cells and incubated for 48 hours.

Coverslips were then washed twice for 3 minutes each with PBS++ (PBS + 0.9 mM calcium chloride + 0.5 mM magnesium chloride). Cells were then fixed by incubation with 4% paraformaldehyde in PBS++ for 15 minutes, followed by two washes with PBS++. Cells were permeabilised with PBS containing 0.5% Triton X-100 for 5 minutes and washed twice with PBS.

Individual coverslips were then transferred to a 24-well plate, and incubated with blocking buffer (PBS with 5% Bovine Serum Albumin, 0.5% Tween-20) for 30 minutes at room temperature, rocking at 30 rpm. Blocking buffer was aspirated, and primary antibodies (see **Table 2.3**) diluted in fresh blocking buffer were added to coverslips and incubated overnight at 4 °C.

Table 2.3: Antibodies used in immunofluorescence experiments

Antibody name	Catalogue number	Manufacturer	Host species	Dilution
Monoclonal ANTI-FLAG M2	F1804	Sigma-Aldrich	Mouse, monoclonal	1:500
Anti-Histone H3 (tri methyl K9)	ab8898	Abcam	Rabbit, polyclonal	1:500
Anti-Mouse IgG H&L (Alexa Fluor 647)	ab150115	Abcam	Goat	1:500
Anti-Rabbit IgG H&L (Alexa Fluor 555)	ab150078	Abcam	Goat	1:500

Coverslips were then washed three times for 5 minutes each with 0.1% Triton X-100 in PBS, and incubated with secondary antibody (**Table 2.3**) diluted in blocking buffer for 1 hour at room temperature. Following aspiration of the secondary antibody, coverslips were washed three times for 5 minutes each with 0.1% Triton X-100 in PBS, and once for 5 minutes with PBS.

Coverslips were then incubated with 2% paraformaldehyde in PBS for 15 minutes at room temperature, and finally washed twice for 5 minutes each with PBS. After rinsing in water, coverslips were dried and mounted face-down onto a drop of SlowFade Diamond Antifade Mountant with 4',6-diamidino-2-phenylindole (DAPI, Invitrogen) and sealed using clear Super Lustrous Nail Enamel (Revlon).

Immunofluorescence images were captured using a ZEISS LSM 900 with Airyscan 2 confocal microscope, and images were processed and exported using Fiji (Image J).

2.13 Immunoprecipitation

MDA-MB-231 cells were seeded at 1.5 x10^5 cells/mL onto 150 mm dishes and incubated at 37 °C with 5% CO_2 overnight. To 3 mL of Gibco Opti-MEM Reduced Serum Media, 30 µg of pcDNA3.1 mHP1α-FLAG WT or mutant or empty vector (pcDNA3.1) was added. 60 µL of X-tremeGENE HP DNA Transfection Reagent (Roche) was then added to each tube, mixed gently, and incubated for 15 minutes at room temperature. Transfection mixtures were then added dropwise to wells of cells and incubated for 48 hours.

Following removal of media and two washes with PBS, 2 mL of ice-cold RIPA buffer (50 mM Tris-HCl pH 7.4, 1% IGEPAL, 0.5% Sodium deoxycholate, 0.05% SDS, 1 mM EDTA, 150 mM NaCl, 1x cOmplete Mini Protease Inhibitor Cocktail (Roche), 50 U/mL SUPERase.In RNase Inhibitor (Invitrogen)) was added to dishes, and incubated at 4 °C for 20 minutes. Cells were then scraped from dishes and sonicated on ice at 1% amplitude for 5 minutes (pulse on 10 s, pulse off 5 s) and then clarified by centrifugation at 12,000 rpm for 10 minutes at 4 °C.

Aliquots of 40 µL of Anti-FLAG M2 Magnetic Beads (M8823, Millipore) were equilibrated with TBS before the application of cell lysates. Beads were incubated with lysates for 3 hours at 4 °C with end-over-end rotation, and the supernatant was removed by using a 16-Tube SureBeads Magnetic Rack (Bio-Rad). Beads were subsequently resuspended in and washed twice with 500 µL of RIPA buffer, then once with RIPA-/- (50 mM Tris-HCl pH 7.4, 1% IGEPAL, 0.5% Sodium deoxycholate, 0.05% SDS, 1 mM EDTA, 150 mM NaCl), before being suspended in 100 µL of RIPA-/-. 5% of this sample was set aside to be subjected to western blotting alongside 20 µL of cell lysate according to **Section 2.10**.

Recombinant Proteinase K (AM2646, Invitrogen) was added to bead mixtures at a final concentration of 250 µg/mL and incubated at 50 °C for 45 minutes. Samples were centrifuged at 7,500 ×g for 10 minutes, and the supernatant was then transferred to a new tube.

RNA was extracted from the supernatant using TRIzol Reagent (Invitrogen). To the supernatant, 1 mL of TRIzol was added and incubated for 5 minutes, followed by 0.2 mL of chloroform. The solution was thoroughly mixed and incubated for 3 minutes before being centrifuged at 12,000 ×g for 15 minutes at 4 °C. The upper aqueous phase was then transferred to a new tube containing 2 µL of 15 mg/mL GlycoBlue Coprecipitant (Invitrogen). 0.5 mL of isopropanol was added, mixed, and the mixture was then incubated for 10 minutes, and subsequently centrifuged at 12,000 ×g for 10 minutes at 4 °C. Supernatant was discarded, and the pellet resuspended in 1 mL of 75% ethanol, vortexed, and centrifuged at 7,500 ×g for 5 minutes at 4 °C. Again, the supernatant was discarded, and the pellet was allowed to air-dry before resolubilisation in 50 µL of Tris-EDTA (TE) buffer, and heated at 60 °C for 15 minutes.

2.14 Northern dot-blot

Three pieces of Bio-Dot SF filter paper (Bio-Rad) and one piece of Zeta-Probe membrane (Bio-Rad) were soaked in sterile distilled water before being assembled onto a Bio-Dot SF Microfiltration Apparatus (Bio-Rad).

To each well, 200 µL of ice-cold dot-blot buffer (10 mM NaOH, 1 mM EDTA) was added and then gently removed using vacuum. RNA samples were made up to 200 µL using dot-blot buffer (1 pmol of TERRA45 used as a positive control), then applied to the membrane with gentle vacuum, and subsequently washed with 200 µL of dot-blot buffer.

The membrane was removed from the apparatus, rinsed in 2x saline-sodium citrate (SSC) with 0.1% SDS, and RNA was fixed to the membrane by crosslinking at 120,000 microjoules/cm^2. The membrane was placed inside a hybridisation tube and equilibrated with hybridisation solution (0.5 M Na_2HPO_4 pH 7.2, 7% SDS, 1 mM EDTA) while rotating for 5 minutes at 65 °C. 100 pmol of probe ($(CCCTAA)_6$-biotin, Integrated DNA Technologies) was then added to 20 mL of hybridisation solution, and incubated with the membrane overnight at 65 °C.

Following hybridisation, the membrane was washed twice for 10 minutes in 1x SSC 0.1% SDS,

and subsequently twice for 30 minutes in 0.1x SSC 0.1% SDS, each time at 65 °C. The membrane was then rinsed in wash buffer (0.1 M maleic acid pH 7.5, 0.15 M NaCl) before being incubated in blocking solution (0.1 M maleic acid pH 7.5, 0.15 M NaCl, 1% BSA) for 30 minutes with rocking at 30 rpm at room temperature. The membrane was then incubated with Streptavidin HRP antibody (ab7403, Abcam) diluted 1:20,000 in blocking solution for 1 hour with rocking at 30 rpm at room temperature. The membrane was then rinsed in TBS before being developed using ECL western blotting detection reagent (Amersham) for 5 minutes and imaged using the chemiluminescence function on a Bio-Rad ChemiDoc.

2.15 Isothermal titration calorimetry

Prior to performing isothermal titration calorimetry (ITC), dialysis of protein or nucleic acid was performed to ensure accurate buffer matching. Samples were dialysed in 50 µL dialysis buttons (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. Water was held in the reference cell, and the cell temperature was set to 25 °C. The initial delay was set to 60 seconds, and a stirring speed of 1000 rpm. There was an initial injection of 0.4 µL for each experiment, and spacing between injections was 180 seconds with a 5 second filter period. The subsequent 18 injections were 2 µL. DNA or RNA was held in the syringe, injected into the sample cell containing protein. Concentrations of protein and DNA/RNA are specified for each experiment in **Figure A.39** and **Figure A.40**.

Analysis was performed using MicroCal Analysis software, under the ITC200 setting. Anomalous data were removed, including the first 0.4 µL injection in all cases. A one-to-one curve was then fit after three iterations to give output data.

2.16 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 and a Dectris Pilatus3-2M detector. A wavelength of 1.0332 Å (12 keV radiation) and detector distance of 1575 mm were implemented.

For SEC-SAXS, the sample was degassed and then 40 µL of 5 mg/mL sample was injected and eluted at 0.2 mL/min. Data were collected using a co-flow setup to mitigate radiation damage, with SEC using a Superdex 200 Increase 5/150 GL column, eluting with IB1 with 5% glycerol. For static sample mode, degassed samples were measured at 100 µM in IB1.

SEC-SAXS data were processed using CHROMIXS (Panjkovich and Svergun, 2017) to manually select the appropriate SEC peak, and subtract eluted buffer. Static mode data were processed and buffer subtracted using scatterBrain. Analysis was then performed using PRIMUS (Konarev et al., 2003) for Guinier and distance distribution analysis. The DAMMIF user interface was used for *ab initio* shape determination (Franke and Svergun, 2009) with manual selection of parameters. Models were determined using slow mode, Ångstrom angular scale, and 10 repetitions. Models are displayed using Chimera 1.16.

Chapter 3

Characterising the interaction between HP1α and TERRA

3.1 Introduction

HP1α preferentially interacts with parallel G4s including TERRA (Roach et al., 2020); however, the mechanism of binding by HP1α is unknown. Because of its G-rich sequence, TERRA readily forms G4s under native conditions, and RNA G4s such as TERRA are exclusively parallel due to the preferred *anti* glycosidic bond conformation of guanine. Therefore, it is unknown if HP1α would still bind TERRA if it were prevented from forming this parallel structure. To this end, hindering TERRA G4 formation through the use of mutation and buffer conditions is necessary to determine if the parallel G4 structure is imperative for HP1α binding. While there is some redundancy in the HP1 family, the functionality and localisation of the HP1α paralog is distinct (Nishibuchi and Nakayama, 2014), and as such its binding to TERRA is also pronounced (Roach et al., 2020), while HP1γ does not bind and HP1β binds TERRA to a lesser degree (Deng et al., 2009). Therefore, the mode of binding to parallel G4s by HP1α must be mediated by motifs unique to HP1α that are yet to be determined. To investigate the effects on HP1α binding to TERRA, multiple patches of basic residues in the hinge and also sites of phosphorylation that alter the charge and function of HP1α are scrutinised.

In analysing potential TERRA G4 binding sites within the hinge of HP1α, there are three distinct patches of nucleic acid binding, basic amino acids (**Figure 3.1**). Near the C-terminus of the

hinge is the 104-106KKK patch. Mutation of this positively charged patch to alanine residues (referred to as "3KA" in some publications) results in decreased binding to TERRA (Roach et al., 2020), reduced propensity for HP1α to bind nucleosomal DNA (Mishima et al., 2013; Ryan and Tremethick, 2018) and also decreases RNA binding ability (Meehan et al., 2003) and causes mis-localisation to heterochromatin (Muchardt et al., 2002). In the centre of the HP1α hinge is another patch of basic amino acids, 89-91KRK, which when mutated also reduces the propensity for HP1α to bind nucleosomal DNA (Mishima et al., 2013). The 89-91KRK patch is also part of the NLS within the hinge of each HP1 paralog, that stretches to also include the 104-106KKK patch (Smothers and Henikoff, 2001). Finally, unique to the HP1α paralog, on the boundary between the CD and hinge is another basic patch consisting of two pairs of lysine residues separated by a tyrosine residue (68-72KKYKK). Despite being a differentiating factor between HP1 paralogs, the 68-72KKYKK patch is relatively unstudied, but in conjunction with the other two positively charged patches in the hinge, the 68-72KKYKK patch is hypothesised to mediate hinge-DNA interactions to drive LLPS of HP1α (Her et al., 2022). Each of the three basic patches in the hinge is therefore of interest to determine their involvement in interactions with G4s.

In the hinge region of HP1α multiple residues are also phosphorylated, and of these hinge PTMs, phosphorylation of S93 has been found to be associated with metaphase, and S97 is also commonly phosphorylated with an unknown role (Hiragami-Hamada et al., 2011). Phosphorylation of HP1α hinge serine residues ensures accurate chromosome alignment and mitotic progression *in vivo* by potentially localising HP1α to kinetochores (Chakraborty et al., 2014). Therefore, due to their presence in the hinge and proximity to basic patches, the effect of phosphorylation of these residues on binding to TERRA is investigated by the use of phosphomimetics. In addition to this, the unstructured NTE of HP1α contains a tract of four consecutive serine residues that are unique to the HP1α paralog and when phosphorylated enhance the binding affinity between the HP1α CD and H3K9me3, and also confer specificity towards H3K9me3 as opposed to unmethylated H3K9 (Hiragami-Hamada et al., 2011). Phosphorylation of these unique residues has also been shown to diminish DNA binding by HP1α (Nishibuchi et al., 2014). Therefore, its effect on TERRA binding is also investigated with the use of phosphomimetics and deletion mutants.

The HP1α CD is responsible for binding H3K9me3 in heterochromatin, and valine residues within the CD have been shown to be pivotal for this function. V22 in particular, when mutated to

Figure 3.1: Alignment of the hinge regions of HP1 paralogs. Amino acid sequences of the HP1α, HP1β, and HP1γ hinge regions. An asterisk (*) indicates a fully conserved residue, a colon (:) indicates conservation of a residue with strongly similar properties, and a period (.) indicates conservation of a weakly similar residue. Pink lines indicate the three basic patches in the HP1α hinge, with the residue numbers shown above. For whole protein sequence alignment, see **Figure 1.16**.

methionine has been shown to completely abolish HP1α CD binding to H3K9me3 (Bannister et al., 2001), and the crystal structure shows that H3K9me3 binds to HP1α through a conserved groove containing V22 (Kaustov et al., 2011). Therefore, the interaction with TERRA is assessed for its ability to interfere with CD-H3K9me3 binding for its function in heterochromatin, as it may alter the flexibility and nucleic acid binding ability of the HP1α hinge.

3.2 Results

3.2.1 TERRA parallel G4 formation is required for HP1α binding

Several strategies were implemented to determine if parallel G4 formation is required for interaction with HP1α and if the interaction is sequence specific to TERRA. Firstly, the TERRA sequence is altered to inhibit stable parallel G4 formation. Circular dichroism spectroscopy is used to determine the topology of TERRA, and binding to HP1α is assessed by BLI and EMSA *in vitro*. Secondly, to determine the nucleic acid sequence specificity of the TERRA-HP1α interaction, the parallel G4 formation of TERRA was inhibited by the omission of G4-stabilising ions.

3.2.1.1 HP1α recognises the structure of parallel G4 TERRA

To confirm the interaction of HP1α with RNA G4s of parallel topology, initially a 45 nt telomeric repeat-containing RNA (TERRA45, **Table A.1**) was folded in interaction buffer (IB1, see **Section 2.1**), which is comprised of 100 mM KCl, 50 mM NaCl, and 20 mM NaH_2PO_4 (at

pH 8). Both the K^+ and Na^+ within this buffer stabilise G4s. To confirm appropriate folding, circular dichroism spectroscopy was used (**Section 2.7**). Circular dichroism spectroscopy measures the difference in absorption of circularly polarised light by molecules containing a chiral chromophore. In the case of G4s, guanine is a chromophore that is linked to a chiral deoxyribose or ribose. The homopolar stacking of G-tetrads, which is evident in parallel G4s, causes a positive circular dichroism signal at 265 nm and a negative circular dichroism signal at 240 nm. Any heteropolar stacking of G-tetrads in an anti-parallel topology causes a positive circular dichroism signal around 290 nm, and a negative circular dichroism signal at 265 nm (Karsisiotis et al., 2011).

Circular dichroism spectroscopy of TERRA45 folded in IB1 showed that it indeed formed a parallel G4 structure evidenced by characteristic positive ellipticities at 210 nm and 265 nm and a negative ellipticity at 240 nm (**Figure 3.2a**). The telomeric DNA coding strand is also G-rich, possessing the same sequence as TERRA, only with deoxyribose instead of ribose, and thymine instead of uracil (gTEL45, **Table A.1**), therefore was also tested for its ability to form G4s by circular dichroism spectroscopy. This showed that gTEL45 also formed a G4 structure. However, due to positive maxima at both 265 nm and 295 nm (**Figure 3.2a**) the structure is deemed "hybrid" (Karsisiotis et al., 2011), with both homo- and heteropolar stacking. A higher maximum at 295 nm corresponds to preferential anti-parallel G4 formation.

To further probe the dependency of the parallel G4 structure for interaction with HP1α, guanine-to-uracil mutations were made in the centre of the second, third, fifth, and seventh G-tracts of TERRA45 to result in mutTERRA45 (**Table A.1**), and its topology was assessed by circular dichroism spectroscopy. Compared to TERRA45, the circular dichroism spectrum of mutTERRA45 showed a decrease in positive circular dichroism signal at 265 nm, and a decrease in negative circular dichroism signal at 240 nm (**Figure 3.2a**), amounting to approximately half of the intensity of the TERRA45 peaks, indicating that mutTERRA45 contained transient/unstable parallel G4 structures.

In order to examine how parallel G4 formation affected the ability of HP1α to bind the oligonucleotide, they were tested using BLI (**Section 2.8**). 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, in this case Ni-NTA for binding his-tagged HP1α. The other surface is used as a reference, and the difference in interfer-

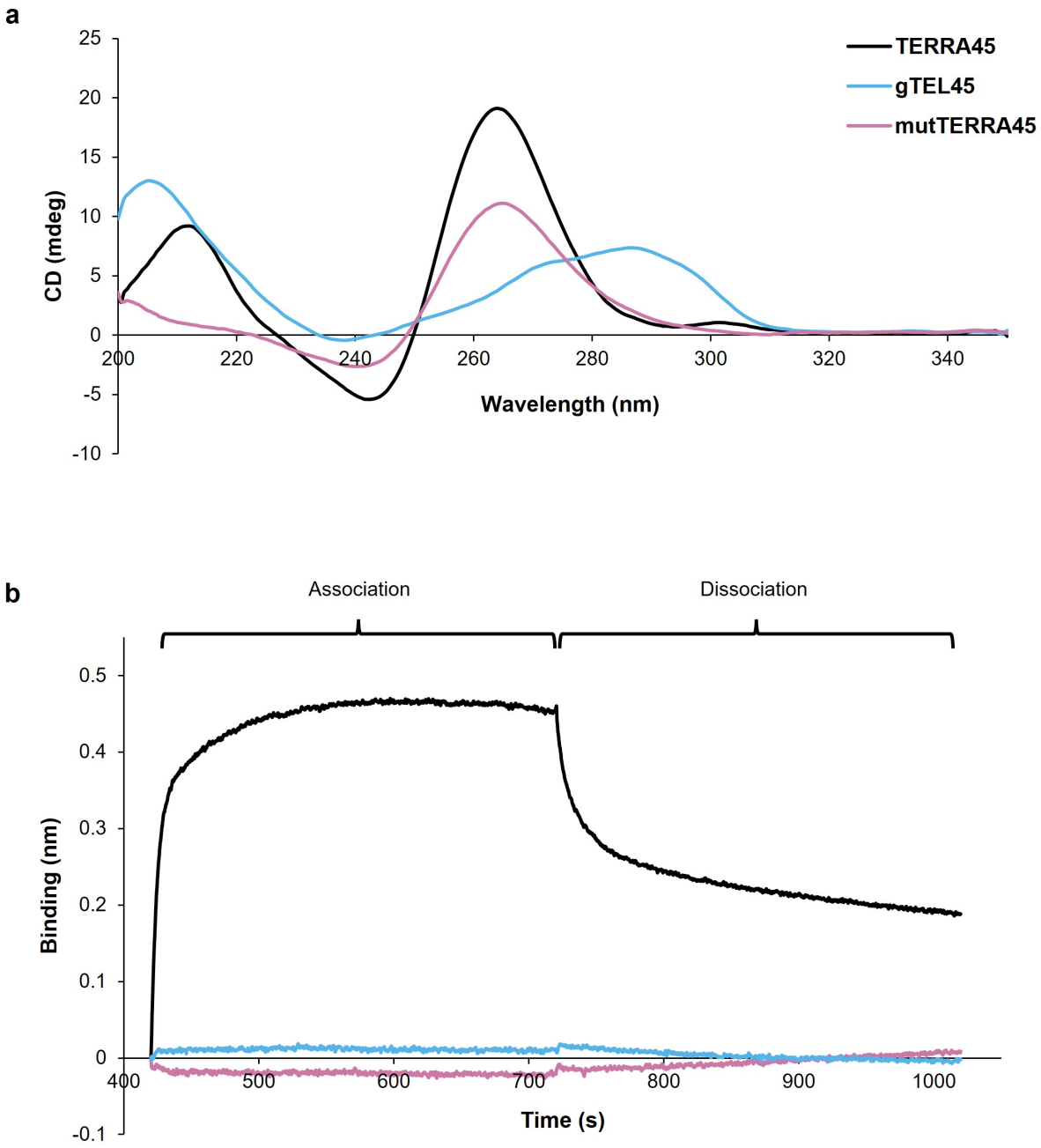

Figure 3.2: HP1α binds to parallel G4s. a) Circular dichroism spectra of TERRA45, gTEL45, and mutTERRA45 at 10 µM. b) BLI measurements showing association and dissociation steps of TERRA45, gTEL45, and mutTERRA45 to HP1α. Oligonucleotides were used at a concentration of 500 nM in this BLI assay and HP1α at 100 µg/mL. Measurements were performed at room temperature.

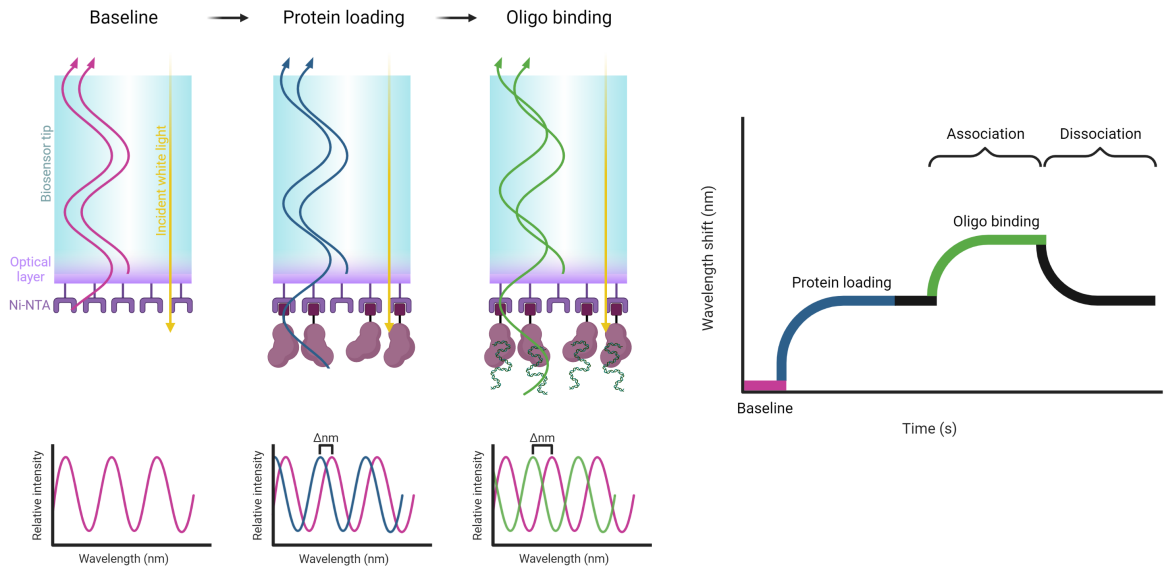

Figure 3.3: Principle of biolayer interferometry. Biolayer interferometry (BLI) uses biosensor tips that contain an optical reference layer and a biocompatible matrix coated in molecules for analyte immobilisation (Ni-NTA for binding his-tagged protein). Incident white light is reflected from both the optical reference layer and the layer containing immobilised analyte. Changes in the thickness of the analyte layer are measured, which depends upon the size and affinity of interactors. Image created with BioRender.

ence patterns of white light measured between the sample and reference is measured in nm and used to measure association and dissociation of molecules bound to the protein-coated biosensor which can then be calculated into affinity constants (**Figure 3.3**).

Recombinantly expressed and purified his-HP1α (**Section 2.5**) was bound to the Ni-NTA tip, washed, and then TERRA45 was added to test its association with HP1α, and subsequently washed with IB1 to measure its dissociation. Addition of TERRA45 to immobilised HP1α resulted in high association, and then subsequent dissociation upon washing (**Figure 3.2b**) in strong agreement with previous data showing the binding of HP1α to TERRA45 (Roach et al., 2020). BLI was then performed using the hybrid/anti-parallel gTEL45, which did not show any association with HP1α (**Figure 3.2b**). BLI was also performed using mutTERRA45, which also showed little to no binding to immobilised HP1α (**Figure 3.2b**), demonstrating the importance of stable parallel G4 structure in binding to HP1α.

To further confirm the BLI results that showed a preference of HP1α for stable parallel G4 structures, an EMSA was performed (**Section 2.4.2**). HP1α was incubated with TERRA45 conjugated with a TAMRA fluorophore (TAM-TERRA45), electrophoresed on a 4% native

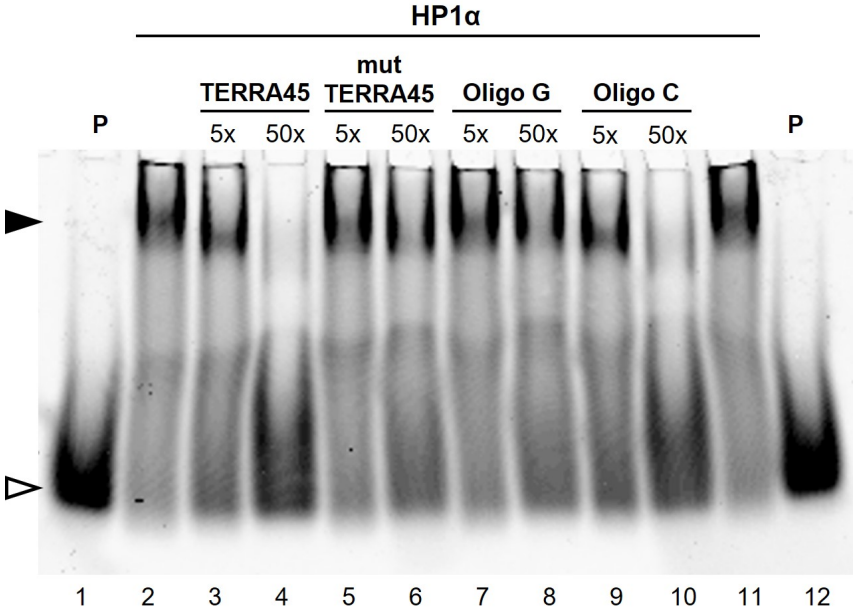

Figure 3.4: Parallel G4s compete for HP1α binding in EMSA. TAM-TERRA45 binding to HP1α (20-fold molar excess) in the absence (lanes 2 and 11) or presence of 5- or 50-fold molar excess of indicated oligonucleotide. "P" indicates TAM-TERRA45 probe only, open arrowhead denotes unbound probe, closed arrowhead denotes complex. Assay was performed at 4 °C and electrophoresed on a 4% native polyacrylamide gel.

polyacrylamide gel, and fluorescence of TAMRA was detected. As shown in **Figure 3.4**, the migration of TAM-TERRA45 through the gel was impeded by addition of HP1α, indicating an interaction (lanes 2 and 11). To test if mutTERRA45 could compete for HP1α binding, 5x and 50x molar excess of competing unlabelled mutTERRA45 was added to the HP1α-TAM-TERRA45 reaction. The mutTERRA45 was unable to compete for HP1α binding (lanes 5 and 6), while increasing concentrations of unlabelled TERRA45 decreased HP1α-TAM-TERRA45 complex formation (lanes 3 and 4, **Figure 3.4**). Oligo G (**Table A.1**), a DNA oligonucleotide which adopts a hybrid/anti-parallel G4 topology (**Figure 3.5**), was also used as a competitor against TAM-TERRA45 but was also unable to compete for binding with HP1α (lanes 7 and 8). Oligo C (**Table A.1**), a DNA oligonucleotide that forms a parallel G4 topology (**Figure 3.5**), did compete with TAM-TERRA45 for HP1α binding (lanes 9 and 10), confirming that HP1α binds to G4s of parallel topology irrespective of nucleic acid type. This shows that mutTERRA45 which is hindered in its ability to form parallel G4s is not bound by HP1α, showing the structural specificity for parallel G4s that HP1α possesses.

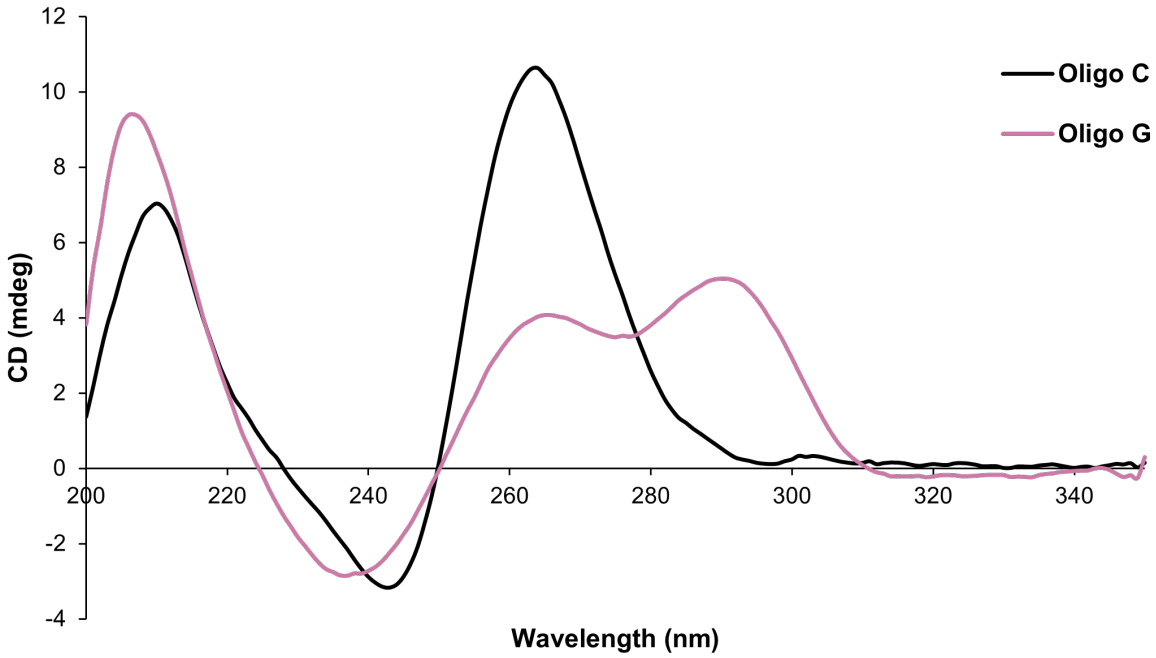

Figure 3.5: Oligo C and Oligo G form parallel and anti-parallel G4s, respectively. Circular dichroism spectra of Oligo C and Oligo G. Measurements were performed at room temperature.

3.2.1.2 Determining if HP1α recognises the sequence of TERRA

To demonstrate that the interaction between HP1α and TERRA45 is not sequence-specific but is structure-specific, TERRA45 was folded in a buffer which lacks K^+ or Na^+ that allow for the formation of G4s through the neutralisation of the negatively charged core of a G-tetrad, thus preventing formation of a G4. Lithium chloride buffer (LiCl IB, 100 mM LiCl, 70 mM Tris base, pH 8) was used as Li^+ is unable to stabilise G4 formation due to an insufficient atomic radius (Włodarczyk et al., 2005; Kim et al., 2013; Miserachs et al., 2016). Circular dichroism spectroscopy confirmed that LiCl IB hindered G4 formation. Shown in **Figure 3.6a**, peaks at 210 nm, 240 nm, and 265 nm all decreased in size, indicating that most of the TERRA45 was unfolded. BLI was then performed to test the interaction of this partially folded TERRA45 with HP1α. Association of TERRA45 in LiCl IB to HP1α was reduced compared to TERRA45 in IB1 (**Figure 3.6b**), which showed that due to the decrease in parallel G4 structure in TERRA45, it bound HP1α to a lesser degree. This confirms that HP1α binding to TERRA is dependent on parallel G4 structure formation, and not solely on the sequence of the oligonucleotide.

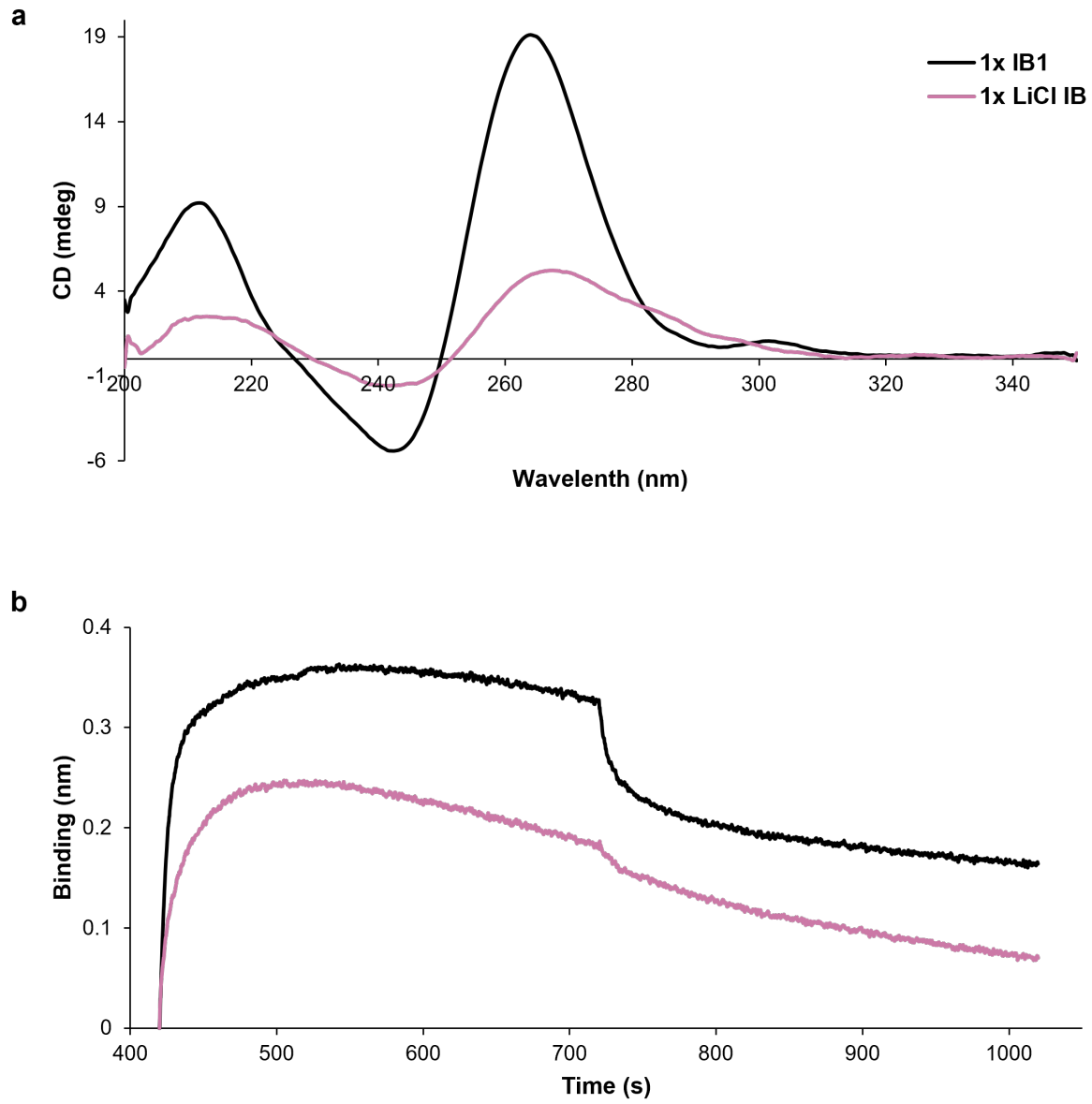

Figure 3.6: Li^+ buffer disrupts G4 formation and inhibits HP1α binding. a) Circular dichroism spectra of TERRA45 folded in either K^+/Na^+-containing buffer (IB1) or Li^+-containing buffer (LiCl IB). b) BLI measurements comparing HP1α binding to TERRA45 in either IB1 or LiCl IB. Oligonucleotides were used at a concentration of 500 nM in this BLI assay and HP1α at 100 µg/mL.

3.2.2 The preeminent interaction of HP1α with TERRA among HP1 paralogs

Previous BLI results showed that HP1α was the only HP1 paralog to bind TERRA (Roach et al., 2020), while Deng et al. (2009) also showed HP1β binding. To examine differences in HP1 paralogs, circular dichroism spectroscopy was performed. The amides in the polypeptide backbone of proteins are chiral chromophores, and when present as secondary structural elements cause characteristic circular dichroism spectra. Folding of α-helices cause negative circular dichroism signal at 220 nm and 210 nm, and a slight positive circular dichroism signal at 190 nm (Holzwarth and Doty, 1965). Anti-parallel β-sheets also cause negative circular dichroism signal at 220 nm and a slight positive circular dichroism signal at 195 nm (Greenfield and Fasman, 1969). A disordered state of folding also produces a circular dichroism spectrum, with slight positive circular dichroism signal at 210 nm and negative circular dichroism signal at 195 nm (Venyaminov et al., 1993).

Far UV (185 nm to 245 nm) circular dichroism spectra showed that all HP1 paralogs had peaks corresponding to α-helices (negative peaks at 222 nm and 207 nm and a positive peak at 193 nm) and β-sheets (negative peak at 218 nm and positive peak at 195 nm, Greenfield 2006), showing that the recombinantly expressed proteins were folded and not entirely disordered, with similar secondary structure formation in their globular domains (**Figure 3.7a**). Near UV (250 nm to 340 nm) circular dichroism spectra showed that HP1α had a large decrease in signal associated with phenylalanine residues (255 nm to 270 nm) compared with the other paralogs (**Figure 3.7b**). HP1α has seven phenylalanine residues, while HP1β and HP1γ have six and eight respectively. This difference in exposure of phenylalanine residues may indicate a tertiary structure that is unique to HP1α. In support of this is also the difference in signal associated with tryptophan residues (290 nm to 310 nm), which are conserved within the CD and CSD of each of the HP1 paralogs.

To further probe interactions of HP1 paralogs with TERRA, an EMSA was performed. When HP1α and TAM-TERRA45 were incubated together and electrophoresed on a native polyacrylamide gel, TAM-TERRA45 mobility was retarded (**Figure 3.8**), indicating that HP1α bound to TAM-TERRA45, while HP1β did not impede the migration of TAM-TERRA45 probe. Interestingly, addition of HP1γ resulted in a small amount of complex with TAM-TERRA45, with an increase in free probe compared with HP1α, indicating it is interacting with TERRA but potentially with reduced affinity.

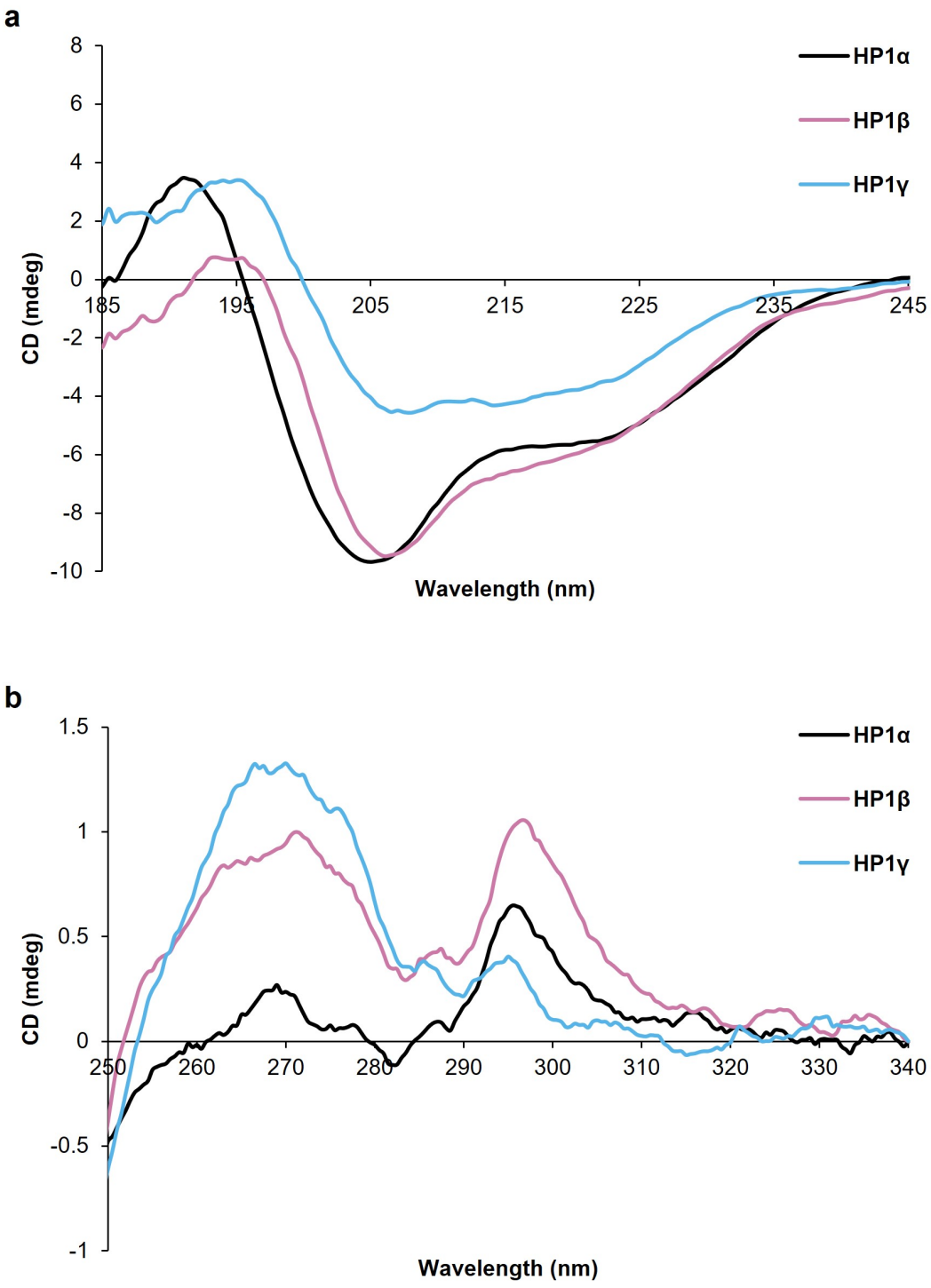

Figure 3.7: Circular dichroism spectra of HP1 paralogs. a) Far UV and, b) near UV circular dichroism spectra of HP1 paralogs. For far UV, 1.5 mg/mL of protein was measured in a 0.1 mm cuvette in 1xIB. For near UV, 1.5 mg/mL protein was measured in a 1 mm cuvette in 1xIB.

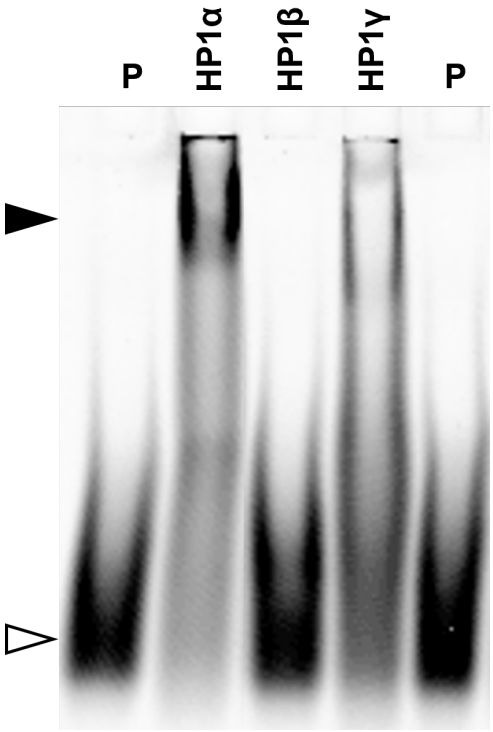

Figure 3.8: Electrophoretic mobility shift assay of HP1 paralogs with TAM-TERRA45. TAM-TERRA45 in the absence (P) or presence of a 20-fold molar excess of the indicated HP1 protein. Open arrowhead denotes unbound TAM-TERRA45 probe, closed arrowhead denotes complex. This image has been edited to remove a lane. The unaltered image can be found in the appendix (**Figure A.32**), as well as a repeat of this experiment (**Figure A.33**).

3.2.3 Multiple charged patches in the hinge of HP1α are required to bind TERRA

The intrinsically disordered, hydrophilic HP1α hinge harbours three distinct patches of positive charge, two of which are common among the HP1 paralogs (**Figure 3.1**). Using site-directed mutagenesis (**Section 2.2.1**), HP1α mutants were created to eliminate the charge of each of the three patches (by substitution with alanine), alone and in combination with one another (**Figure 3.9**) to determine their contribution to HP1α binding to TERRA45. Following mutagenesis, HP1α mutants A1, A2, A3, A1;2, A1;3, A2;3, and A1;2;3 were recombinantly expressed and purified for analysis (**Section 2.5**).

To ensure BLI analysis provided detectable results using HP1α mutants that may bind TERRA45 more weakly, the concentration of TERRA45 used herein was increased from 500 nM to 2 µM. **Figure 3.10** shows that each patch of positively charged amino acids in the hinge (either patch 1, 2, or 3) mutated to alanine residues (A1, A2, or A3, **Figure 3.9**) resulted in markedly reduced binding to TERRA45.

To further assess the interactions of HP1α WT and mutants with TERRA, the equilibrium dissociation constant (K_D) is calculated using the ratio of association (k_{on}) and dissociation (k_{off}) constants using a 1:1 binding model of the BLI data, yielding a value that 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).

BLI analysis showed the K_D for the interaction between HP1α WT and TERRA45 was 83 nM, with a k_{on} of 6.3 $\times 10^4$ 1/Ms and a k_{off} of 5.1 $\times 10^{-3}$ 1/s (**Table 3.1**). Assessment of the BLI data of the hinge patch mutants showed the K_D increased 4-fold, 67-fold, and 14-fold for mutants A1, A2, and A3, respectively (**Table 3.1**), suggesting the importance of each of the three patches.

To determine if all are required, combinations of two or all three patches in the hinge were mutated together (A1;2, A1;3, A2;3, or A1;2;3), BLI showed binding to TERRA45 was completely abolished (**Figure 3.10**, **Table 3.1**). An EMSA was also performed, where the mobility of TAM-TERRA45 was assessed in the presence or absence of HP1α WT or the hinge patch mutants. The EMSA showed no detectable complex formation with the hinge patch mutants

Figure 3.9: Charged patches in the hinge of HP1α mutated. The amino acid sequence of the hinge of HP1α. Mutated patches are numbered and underlined with their substituted residues shown above the sequence.

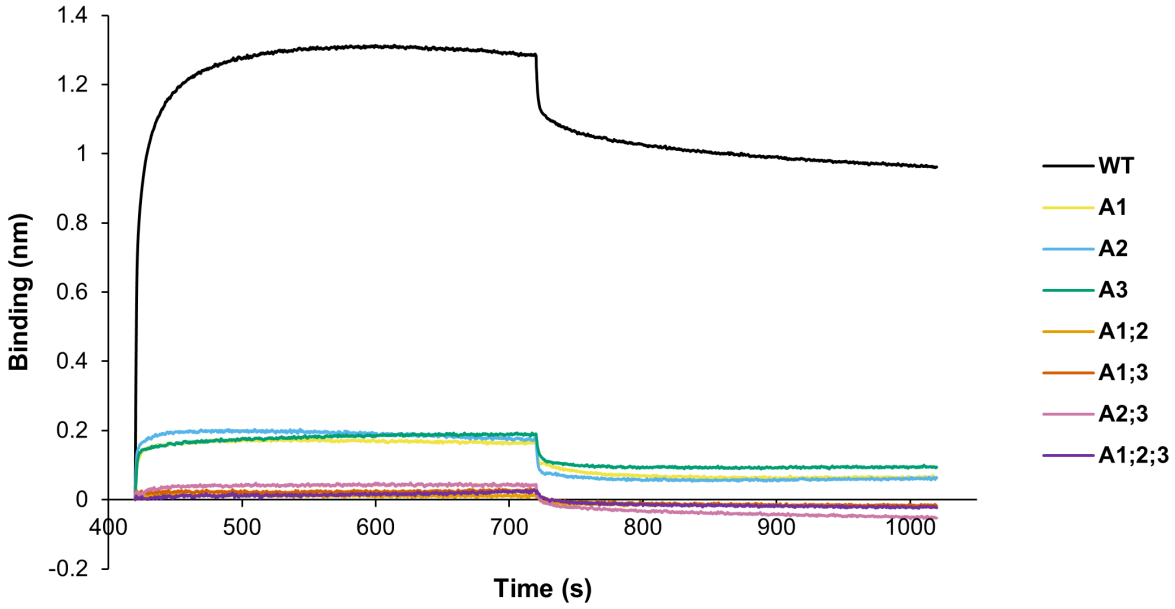

Figure 3.10: Binding of HP1α hinge patch mutants to TERRA45. BLI analysis of immobilised HP1α WT or hinge patch mutants binding to TERRA45. Oligonucleotide was used at a concentration of 2 µM in this BLI assay and HP1α WT and mutants at 100 µg/mL. Measurements were performed at room temperature.

Table 3.1: Binding affinities for HP1α WT or hinge patch mutants with TERRA45 measured by BLI

HP1α variant	K_D **(nM)**	K_D **error**	k_{on}**/1000 (1/Ms)**	k_{on} **error /1000**	k_{off}**×1000 (1/s)**	k_{off} **error ×1000**
WT	83	5	62.8	1.6	5.1	0.3
A1	339	9	141	3	47.7	0.7
A2	5550	1300	56	13	309	11
A3	1180	110	67	6	79	3
A1;2	No binding					
A1;3	No binding					
A2;3	No binding					
A1;2;3	No binding					

apart from A3 (104-106KKK>AAA), which showed a small amount of complex formation with TAM-TERRA45 (**Figure 3.11**, also shown in **Figure A.32**). These results indicate that HP1α requires at least two of these positively charged patches in the hinge to even weakly bind parallel G4s such as TERRA45.

To confirm that the markedly weaker binding affinity between HP1α hinge patch mutants and TERRA45 was due to changes in the charge of the hinge and not due to misfolding of the protein, circular dichroism was performed to assess secondary structure. Circular dichroism spectra showed that the hinge patch mutants' secondary structure folded similarly to the WT HP1α (**Figure 3.12**), suggesting that the changes in charge of the hinge do not affect the secondary structure of HP1α. Therefore, the interaction with TERRA45 through the hinge of HP1α is primarily mediated by electrostatic interactions.

3.2.3.1 Altering the HP1α hinge positive charge to stabilise the interaction with TERRA

Positively charged basic amino acids play important roles in protein binding with nucleic acids through electrostatic interactions. The three charged patches within the hinge of HP1α are predominantly comprised of the basic amino acid lysine; however, the guanidinium group of the basic arginine side-chain enables a larger number of possible electrostatic interactions compared with the one direction of interaction allowed by the side-chain of lysine (Borders Jr et al., 1994; Donald et al., 2011; Sokalingam et al., 2012). Therefore, to increase the stability of the HP1α hinge-TERRA45 interaction, the three lysine residues at 104-106 were mutated to arginine

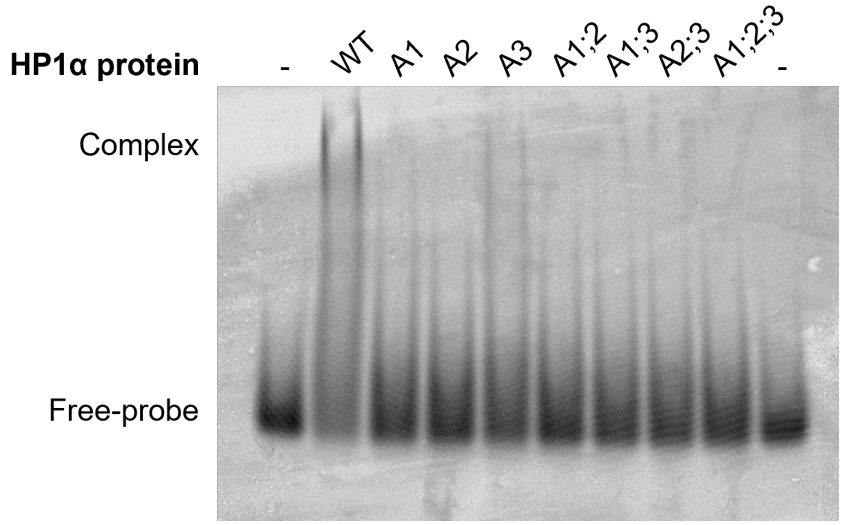

Figure 3.11: Electrophoretic mobility shift assay of HP1α WT and hinge patch mutants with TAM-TERRA45. EMSA with TAM-TERRA45 in the absence (-) or presence of a 20-fold excess of WT HP1α or hinge patch mutant. A repeat of this experiment can be found in the appendix (**Figure A.34**).

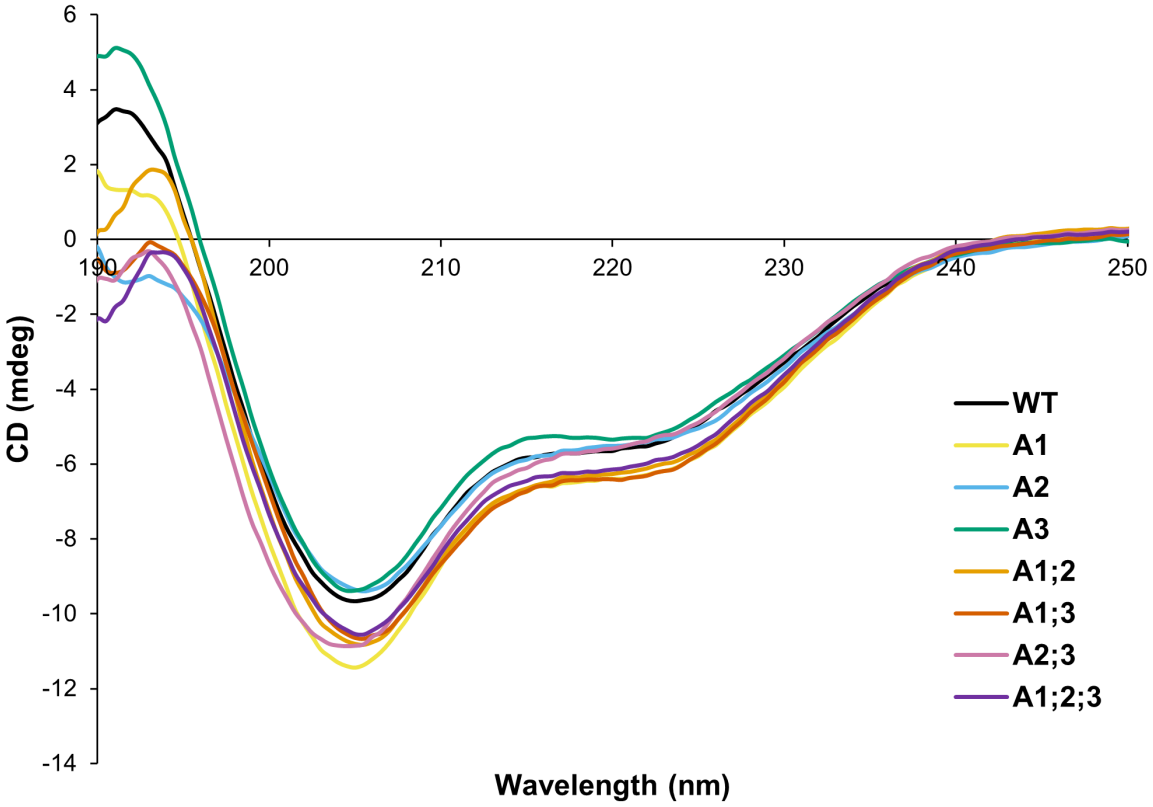

Figure 3.12: Circular dichroism spectra of hinge patch mutants. Far UV circular dichroism spectra of HP1α hinge patch mutants. 1.5 mg/mL of each protein was measured in a 0.1 mm cuvette in 1xIB at room temperature.

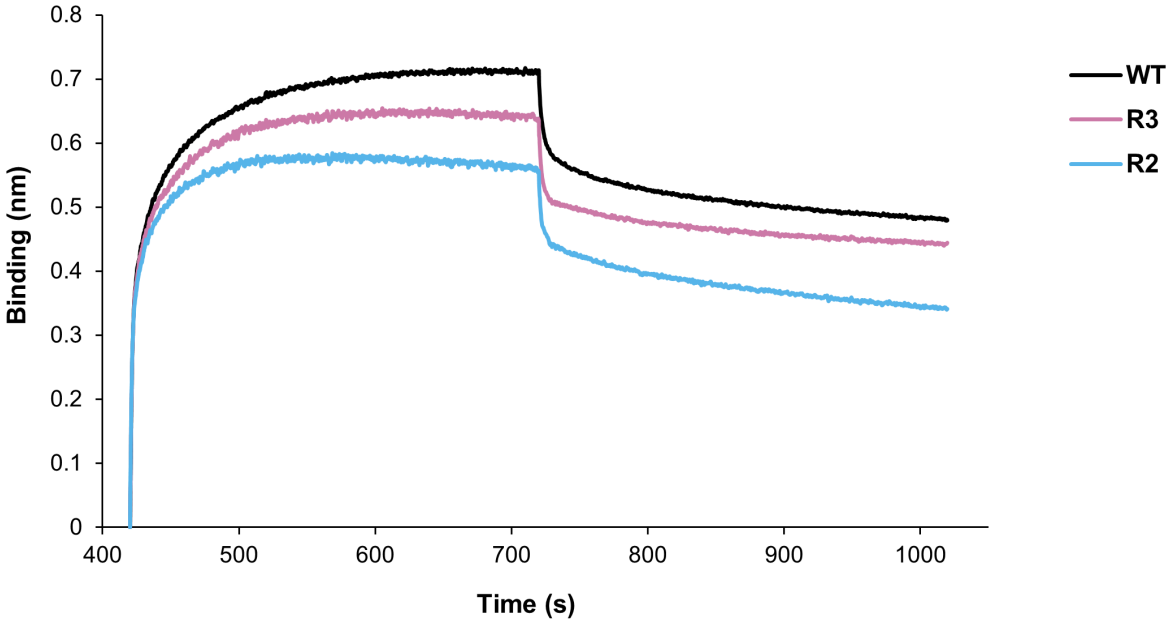

Figure 3.13: BLI analysis of supercharged mutants with TERRA45. Binding of TERRA45 to HP1α R2 and R3 supercharged mutants measured by BLI.

Table 3.2: Binding affinities for HP1α R2 and R3 mutants with TERRA45 measured by BLI

HP1α variant	K_D **(nM)**	K_D **error**	k_{on}**/1000 (1/Ms)**	k_{on} **error /1000**	k_{off}**×1000 (1/s)**	k_{off} **error ×1000**
WT	83	5	62.8	1.6	5.1	0.3
R2	299	14	43.0	1.4	12.8	0.4
R3	203	5	64.9	1.0	13.1	0.3

residues (104-106KKK>RRR) to create the R3 mutant, and the two lysine residues in the KRK motif at residues 89 and 91 were also mutated to arginine residues (89-91KRK>RRR) to create the R2 mutant. These mutants were then tested for their binding ability with TERRA45 by BLI. Both the R2 and R3 mutants bound TERRA45 similarly, although less strongly than WT HP1α (**Figure 3.13**). The R2 and R3 mutants had comparable k_{on} rates, but higher k_{off} rates contributed to a weaker K_D (**Table 3.2**). Therefore, the substitution for more distributed charge in arginine residues in place of lysine residues in two individual patches did not contribute to greater affinity of HP1α for TERRA45.

3.2.3.2 Nuclear localisation of the HP1α triple hinge patch mutant

Given the hinge is involved in HP1α localisation to H3K9me3 marked heterochromatin *in vivo*, NIH3T3 immortalised mouse cells were transfected with plasmids expressing FLAG-tagged HP1α WT or the triple hinge mutant A1;2;3. After 48 hours, cells were processed for immunofluorescence (**Section 2.12**) by using anti-FLAG to detect HP1α, an H3K9me3 antibody, and staining with DAPI to visualise DNA. FLAG-HP1α WT showed localisation to DAPI stained foci representing regions of heterochromatin (**Figure 3.14**). However, FLAG-HP1α A1;2;3 also localised to DAPI foci (**Figure 3.15**). There was also no difference in DAPI foci or H3K9me3 staining (of which nuclear penetration was poor), indicating little difference in the heterochromatin foci. Lack of discernible difference between HP1α WT and A1;2;3 localisation may be due to endogenous expression of HP1α in NIH3T3 cells, aiding in the localisation of FLAG-tagged HP1α A1;2;3.

Since endogenous HP1α may be aiding HP1α A1;2;3 localisation, this experiment was repeated in MDA-MB-231 metastatic mammary adenocarcinoma cells previously shown to have low levels of HP1α expression (Norwood et al., 2006). MDA-MB-231 cells were transfected with plasmids expressing FLAG-tagged HP1α WT or A1;2;3 and prepared for immunofluorescence as above. Both HP1α WT and A1;2;3 showed similar diffuse patterning in the nucleus, with a small number of punctate spots (**Figure 3.16** & **3.17**). In comparing the localisation of HP1α in MDA-MB-231 cells, again, there was little distinction, but a weak trend of an increased number of HP1α foci in the FLAG-HP1α A1;2;3 mutant compared to WT may be gleaned. There were also no distinct changes in DAPI staining or H3K9me3 localisation.

To determine if the A1;2;3 mutations affected HP1α binding to TERRA *in vivo*, preliminary experiments were performed where MDA-MB-231 cells were transfected with plasmids expressing FLAG-tagged HP1α WT or A1;2;3. FLAG-tagged proteins were successfully immunoprecipitated (IP) using anti-FLAG agarose beads (**Figure 3.18a**, **Section 2.13**), and the associated RNA was purified and dot-blotted. To detect TERRA transcripts, a complementary C-rich $(CCCTAA)_6$ probe was used for hybridisation (**Section 2.14**). The initial attempt at this experiment was unsuccessful in detecting the HP1α-TERRA interaction (**Figure 3.18b**) and further work will need to be done to optimise this.

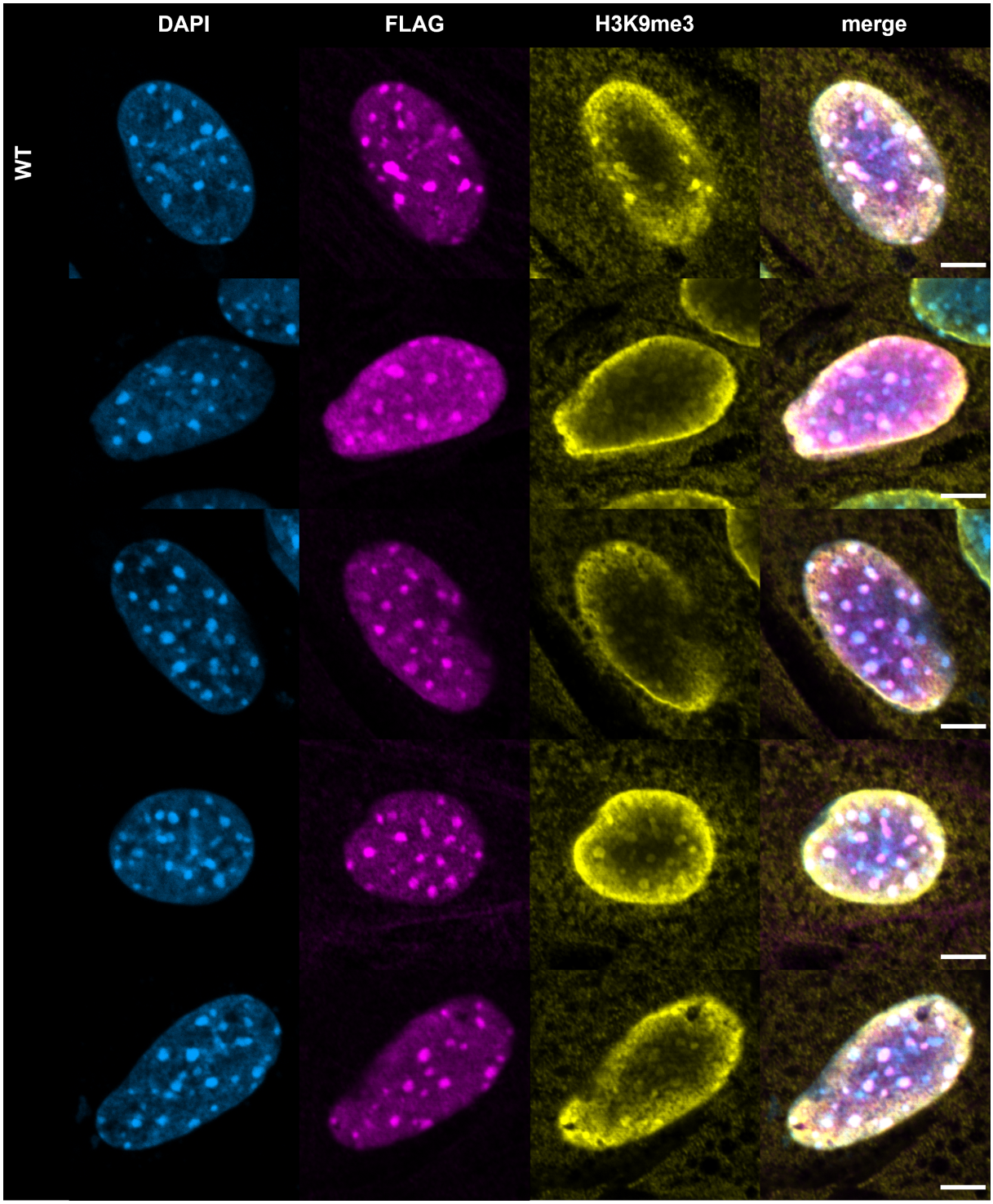

Figure 3.14: FLAG-HP1α WT localisation in NIH3T3 cells. Confocal microscope immunofluorescence images of NIH3T3 cells, with DAPI stained blue, FLAG-HP1α WT shown in magenta, and H3K9me3 shown in yellow. Scale bar is 5 µm.

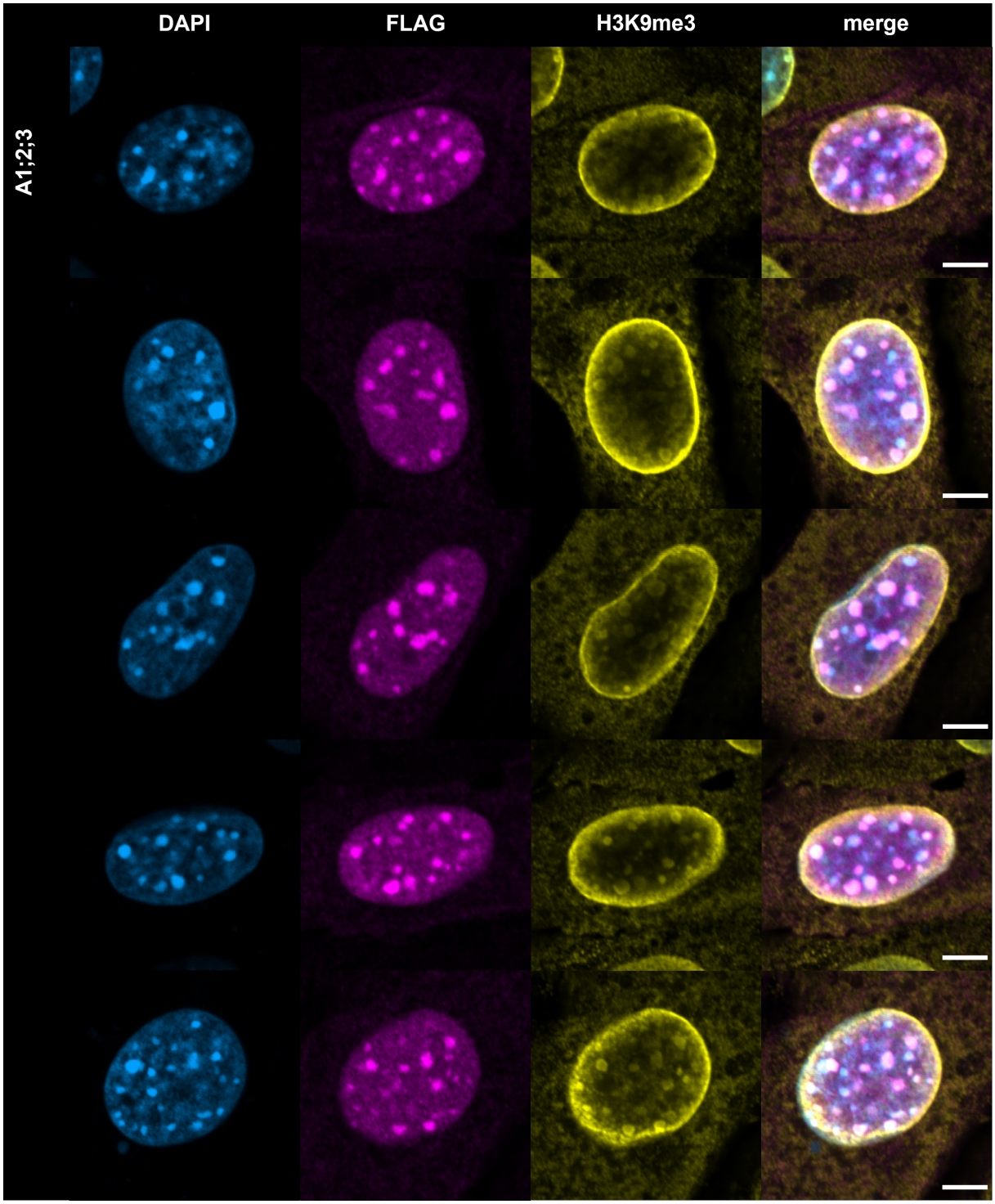

Figure 3.15: FLAG-HP1α A1;2;3 localisation in NIH3T3 cells. Confocal microscope immunofluorescence images of NIH3T3 cells, with DAPI stained blue, FLAG-HP1α A1;2;3 shown in magenta, and H3K9me3 shown in yellow. Scale bar is 5 µm.

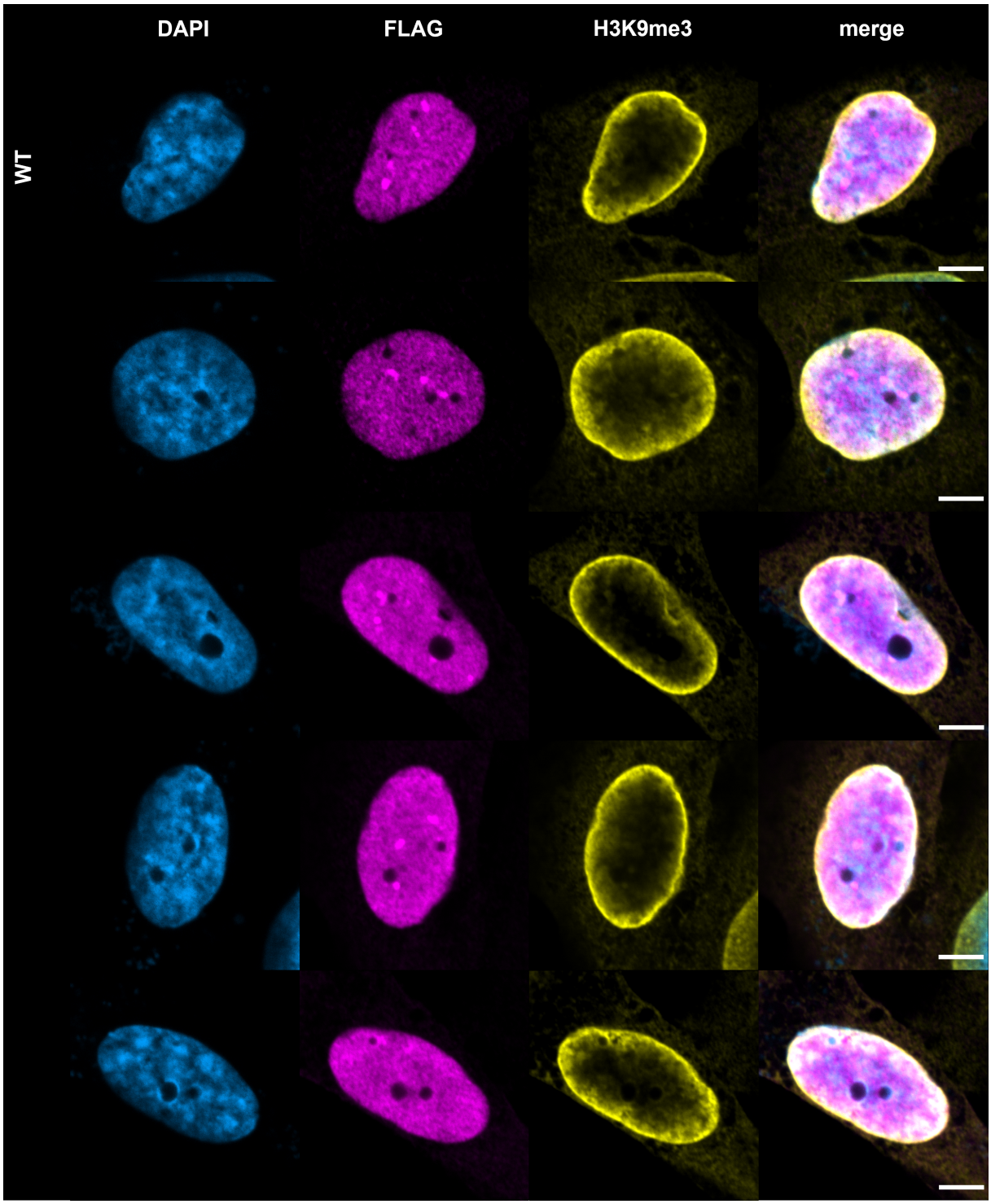

Figure 3.16: FLAG-HP1α WT localisation in MDA-MB-231 cells. Confocal microscope immunofluorescence images of MDA-MB-231 cells, with DAPI stained blue, FLAG-HP1α WT shown in magenta, and H3K9me3 shown in yellow. Scale bar is 5 µm.

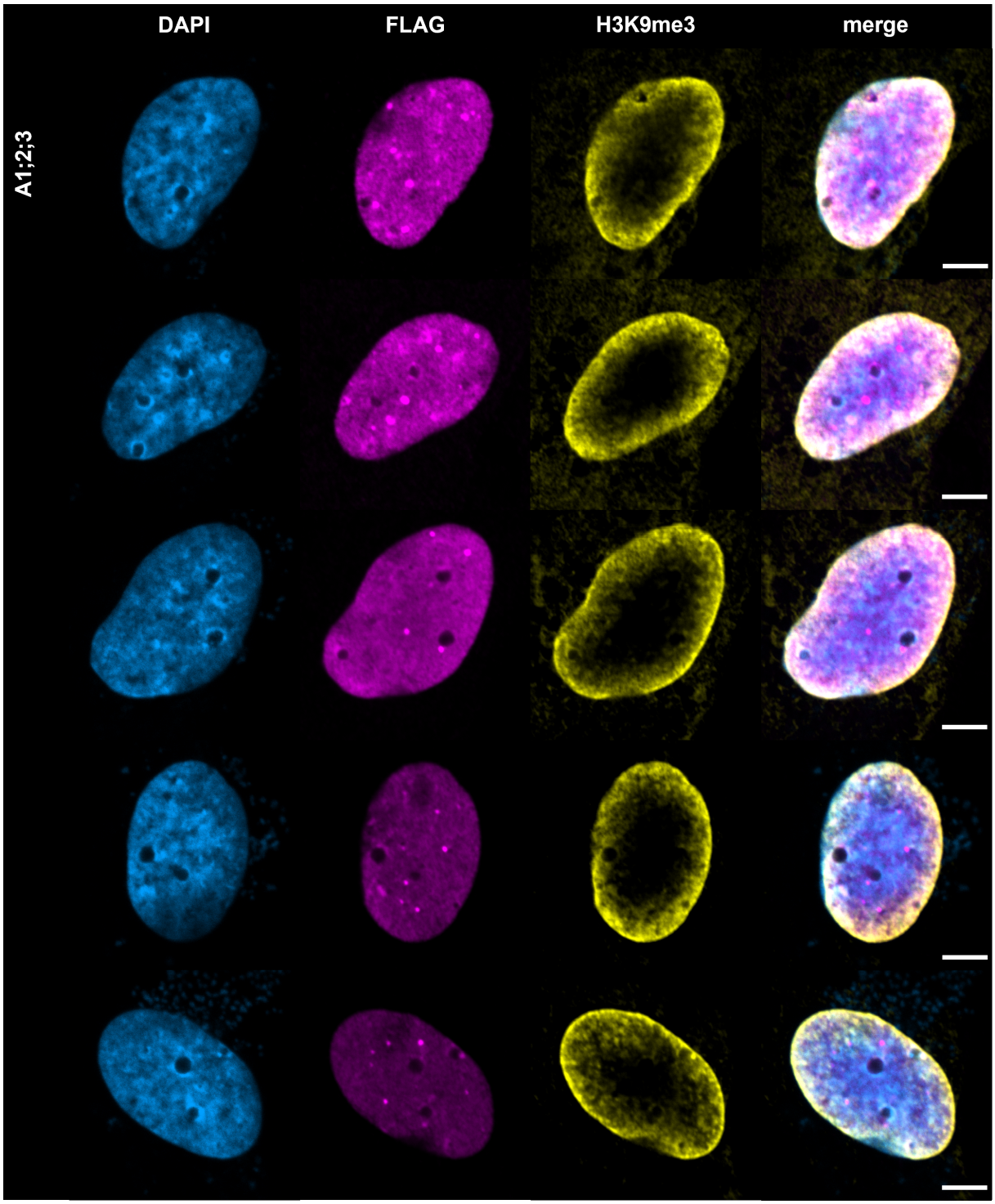

Figure 3.17: FLAG-HP1α A1;2;3 localisation in MDA-MB-231 cells. Confocal microscope immunofluorescence images of MDA-MB-231 cells, with DAPI stained blue, FLAG-HP1α A1;2;3 shown in magenta, and H3K9me3 shown in yellow. Scale bar is 5 µm.

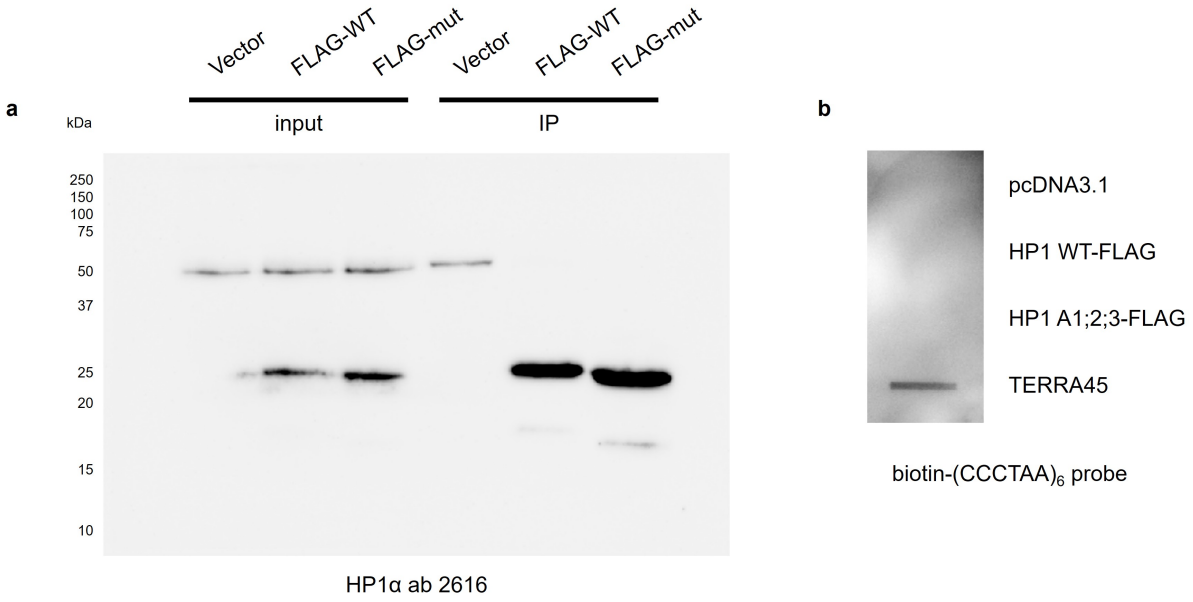

Figure 3.18: Northern dot-blot to test the *in vivo* interaction between HP1α WT and A1;2;3 with TERRA. a) MDA-MB-231 cells were transfected with FLAG-tagged HP1α or A1;2;3 or empty vector (pcDNA3.1). Western blot of input samples (before IP) and RNA IP samples (5% input) were subjected to SDS-PAGE, transferred to a nitrocellulose membrane, and subjected to western blotting using HP1α antibody. b) Samples were then subjected to RNA IP with anti-FLAG beads. IP RNA (or TERRA45 control) was dot-blotted and hybridised with biotinylated $(CCCTAA)_6$ probe, and streptavidin HRP antibody was used to detect chemiluminescence.

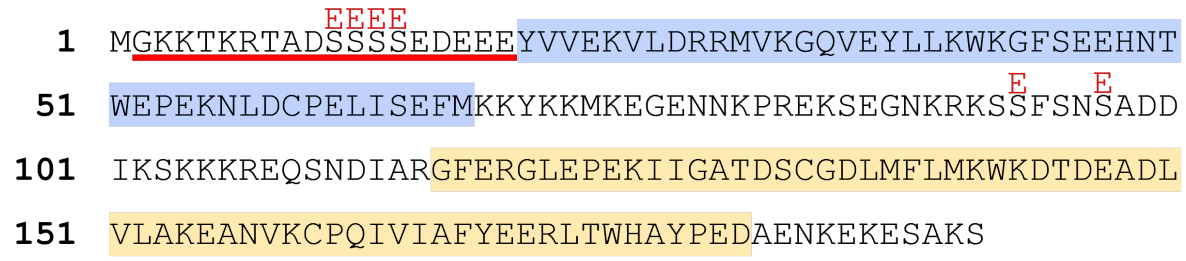

Figure 3.19: Mutants of HP1α to assess the effect of phosphorylation. Serines were mutated to glutamic acid residues, indicated by the red "E" above the amino acid sequence of HP1α to produce the S11-14E, S93E, and S97E mutants. The amino acids deleted in the ΔNTE mutant are underlined in red. The CD is highlighted in blue and the CSD is highlighted in yellow.

3.2.4 The effect of HP1α phosphomimetics on the interaction with TERRA

Phosphorylation of HP1α not only alters the charge of the protein, but can also change its conformation and function (Larson et al., 2017; Hiragami-Hamada et al., 2011). To test if phosphorylation affects the interaction of HP1α with TERRA, HP1α phosphorylation mimics were created to determine their effect on binding of TERRA.

3.2.4.1 HP1α hinge residue phosphomimetics reduce affinity for TERRA

The charge in the HP1α hinge is a strong determinant of G4 binding. Given that phosphorylation of the HP1α hinge has been observed, to determine if these PTMs regulate G4 binding, the phosphorylation sites in the hinge of HP1α were also investigated for their effect on TERRA binding. Residues S93 and S97 were mutated to glutamic acid residues (S93E and S97E, **Figure 3.19**) by site-directed mutagenesis (**Section 2.2.1**) to mimic the negative charge of a phosphorylated serine residue. BLI was performed to test their binding to TERRA45, and, as shown in **Figure 3.20** and **Table 3.3**, both S93E and S97E indeed had less affinity for TERRA45 compared with WT HP1α (but similar affinity to TERRA45 as each other), suggesting phosphorylation of the HP1α hinge may regulate this interaction.

3.2.4.2 HP1α N-terminal extension phosphomimetics reduce affinity for TERRA

Phosphorylation of NTE serine residues is unique to the HP1α paralog, causing a more open conformation and aiding in generating higher-order oligomers in chromatin (Larson et al., 2017).

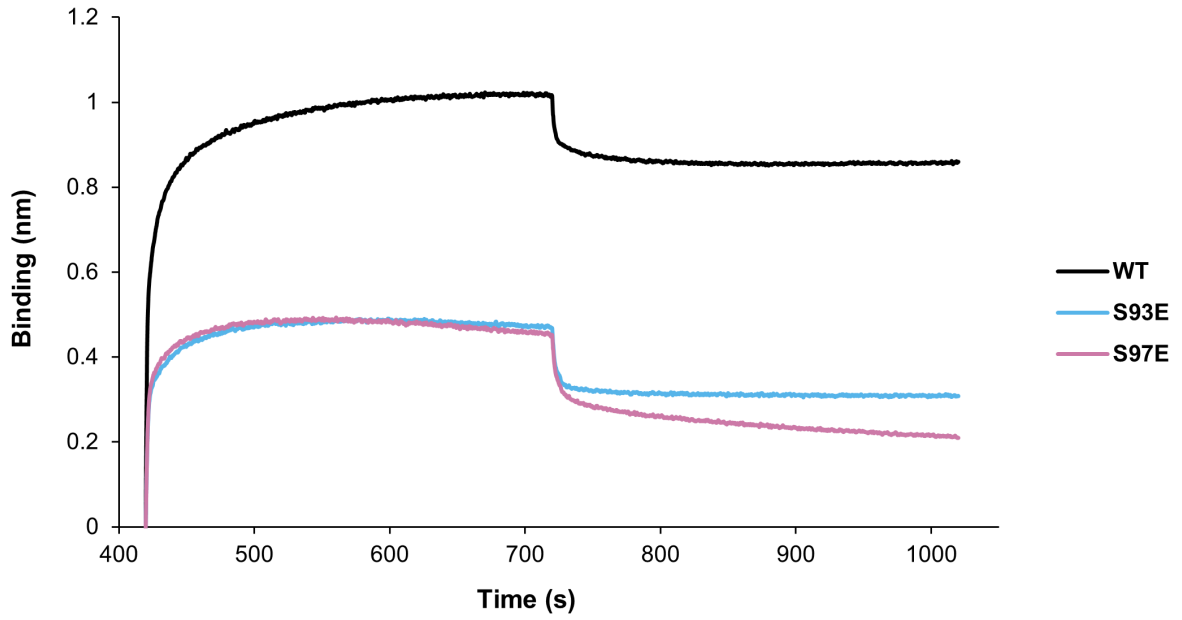

Figure 3.20: BLI analysis of HP1α WT or S93E or S97E with TERRA45. BLI analysis of immobilised HP1α WT or hinge phosphorylation mimics (S93E or S97E) binding to TERRA45.

Table 3.3: Binding affinities for HP1α phosphorylation mutants and ΔNTE with TERRA45 measured by BLI

HP1α variant	K_D **(nM)**	K_D **error**	k_{on}**/1000 (1/Ms)**	k_{on} **error /1000**	k_{off}**×1000 (1/s)**	k_{off} **error ×1000**
WT	83	5	62.8	1.6	5.1	0.3
S93E	490	30	45	2	21.9	0.8
S97E	480	20	81	2	39.1	0.9
S11-14E	407	16	101	3	41.1	1.2
ΔNTE	560	30	51	3	28.7	0.9

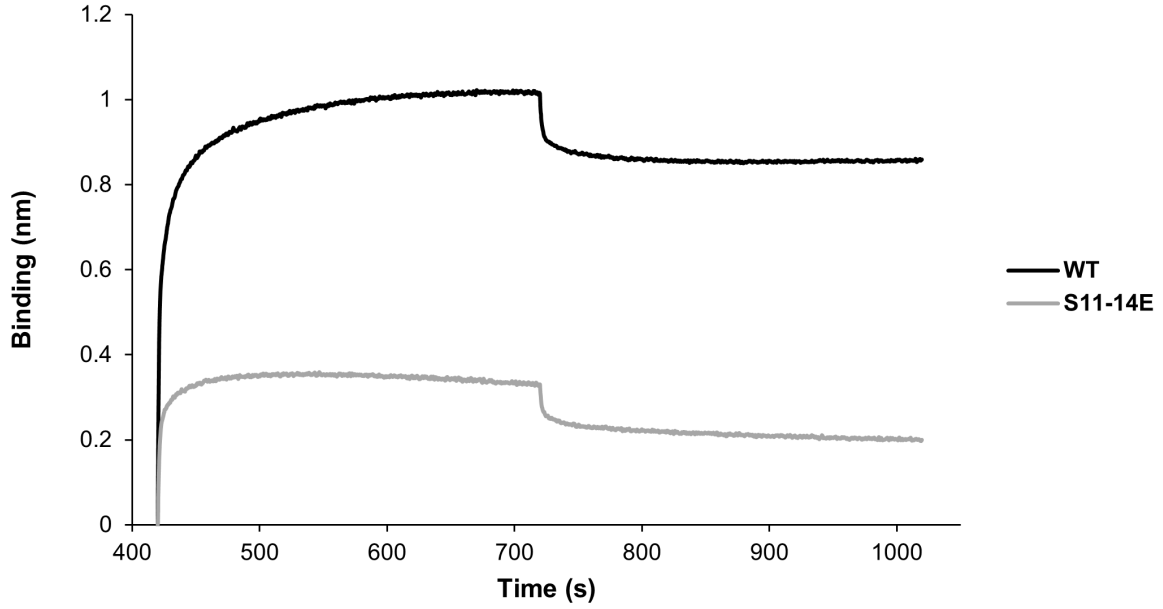

Figure 3.21: BLI analysis of HP1α WT or S11-14E with TERRA45. BLI analysis of immobilised HP1α WT or N-terminal phosphorylation mimic S11-14E binding to TERRA45.

To test if NTE phosphorylation affects G4 binding, residues 11-14 were mutated to glutamic acid (S11-14E, **Figure 3.19**) to mimic phosphorylated serine residues (Maciejewski et al., 1995).

BLI analysis with TERRA45 showed the S11-14E K_D was 407 nM (**Figure 3.21**, **Table 3.3**), indicating a weaker interaction of the mutant with faster k_{on} and much faster k_{off} rates compared to the HP1α WT-TERRA interaction. The decreased binding displayed by the HP1α S11-14E mutant may be due to the overall increase in negative charge of the protein, resulting in electrostatic repulsion with the negative phosphates within the RNA, or alternatively could be due to the previously shown conformational change in HP1α caused by NTE phosphorylation (Larson et al., 2017).

To further investigate the role of the NTE, given that the NTE phosphorylation mimic decreases binding of HP1α to TERRA, a mutant was created lacking the NTE (ΔNTE, **Figure 3.19**) by inverse PCR (**Section 2.2.2**), expressed, purified (**Section 2.5**), and tested for its binding with TERRA45 by BLI. The ΔNTE mutant also showed a decrease in binding to TERRA45 compared to WT HP1α (**Figure 3.22**, **Table 3.3**) mostly due to an increased k_{off}, indicating that the unphosphorylated NTE may facilitate G4 binding.

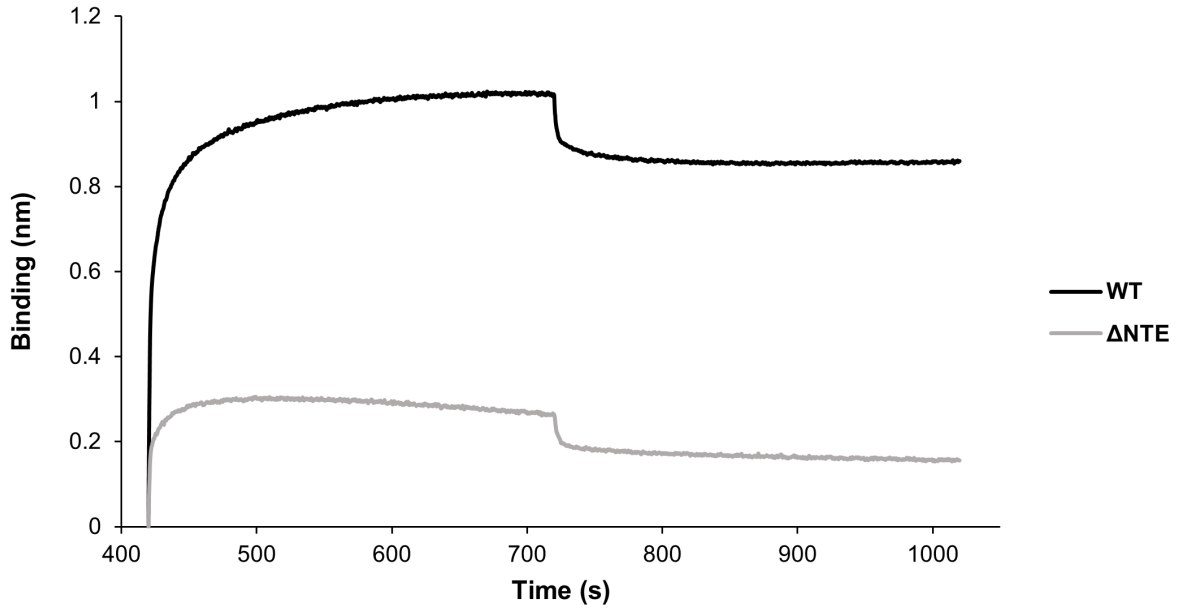

Figure 3.22: BLI analysis of HP1α WT or ΔNTE with TERRA45. BLI analysis of immobilised HP1α WT or ΔNTE binding to TERRA45

3.2.5 The interaction of forward major satellite repeat RNA with unmodified HP1α

In the RNA-mediated targeting of HP1α to centromeric heterochromatin, fMSR RNA was integral to this process when HP1α was SUMOylated (Maison et al., 2011). Therefore, preliminary work has been done to determine if WT HP1α binds to fMSR RNA.

To test whether HP1α binds this hairpin-forming RNA, an oligonucleotide of comparative length to TERRA45 must be selected. The Vfold2D (Xu et al., 2014) prediction tool can be used to assess possible RNA hairpin formation in the fMSR sequence (**Table A.1**), which found a structure rich in potential hairpins (**Figure A.35**). The hairpin from nucleotide 27 to nucleotide 70 is predicted to form with a free energy of -9.49 kcal/mol and was selected for analysis. Circular dichroism spectroscopy was performed on the fMSR oligonucleotide, which showed a spectrum similar to the hairpin-forming tRNA (**Figure 3.23a**). BLI was then performed to test the binding of HP1α WT to this 44 nt fMSR hairpin. Interestingly, minimal binding was shown by the full-length unmodified HP1α to the short fragment of fMSR RNA (**Figure 3.23b**). This is likely due to the lack of PTMs on this recombinantly expressed HP1α, which was determined to be SUMOylated within the hinge *in vivo* when targeted to centromeres by fMSR RNA. This

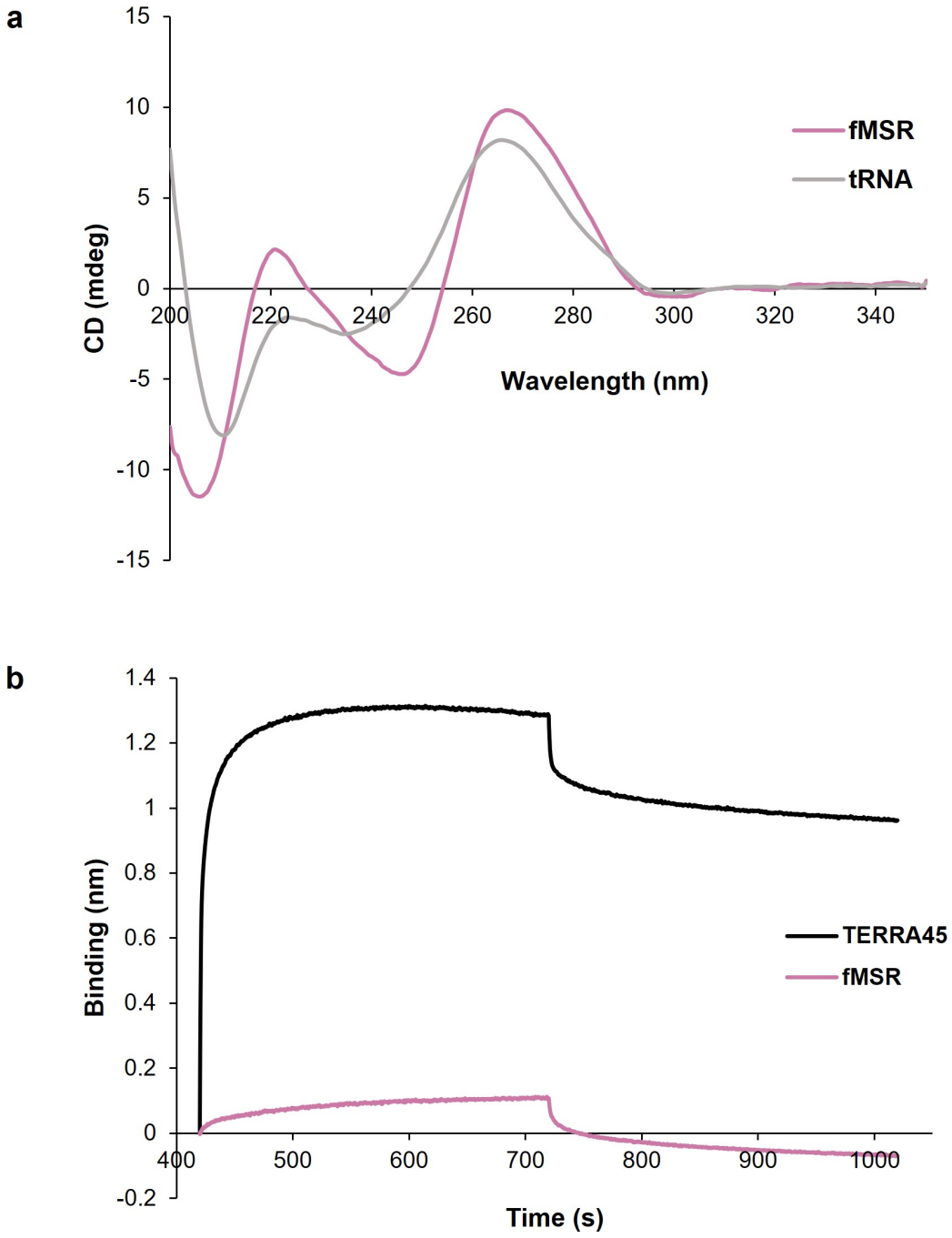

Figure 3.23: Unmodified HP1α binding the fMSR hairpin. a) Circular dichroism spectra of 44 nt fMSR and tRNA show similar RNA hairpin peaks. b) BLI measurements of HP1α WT with 44 nt fMSR. HP1α WT plus TERRA45 has been included for comparison. A prediction of the fMSR structure can be found in **Figure A.35**.

indicates there may be an alternative mechanism of binding by HP1α to fMSR, requiring hinge SUMOlaytion, which should be tested in future experiments.

3.2.6 The effect of H3K9me3 binding to HP1α on the interaction with TERRA

Given the important role of the valine 22 residue in the hydrophobic interaction of the HP1α CD in binding H3K9me3 in heterochromatin (Bannister et al., 2001; Kaustov et al., 2011), the HP1α V22M mutant was initially produced for downstream experiments including nucleosomal arrays. Due to time constraints, these experiments were unable to be completed. However, the HP1α V22M mutant was indeed tested for its interaction with TERRA by BLI. This showed that HP1α V22M bound to TERRA45 similarly to WT (**Figure 3.24**, **Table 3.4**), indicating that a mutation affecting CD-H3K9me3 binding does not substantially alter the binding ability of HP1α to TERRA, and the CD is likely not required for TERRA binding.

To assess if TERRA alters the interaction between HP1α and H3K9me3, a peptide pulldown was performed where biotinylated H3K9me3 or unmodified H3 peptide (of the N-terminal tail) was used to bind HP1α in the presence or absence of TERRA45 or mutTERRA45 (**Section 2.9**). Initially, with the use of magnetic streptavidin beads, western blot analysis of pulldown samples showed that HP1α in the absence of RNA bound to H3K9me3, and when TERRA45 was added to the interaction, HP1α was indeed still able to bind H3K9me3 (**Figure A.36**). However, there were significant issues with HP1α binding to the beads independently of biotinylated peptide. Therefore, the same experiment was repeated with the use of agarose streptavidin beads (**Section 2.9**). Western blot analysis of this experiment yielded notable improvements in background HP1α bead binding, and again showed that HP1α was able to bind H3K9me3 in the presence of TERRA45, although a slight decrease in HP1α binding is observed (**Figure A.37**).

To further mitigate background HP1α bead binding, 300 pmol of biotin was added to block streptavidin binding sites on the agarose beads following peptide binding (**Section 2.9**). Western blot analysis of these pulldown samples showed that HP1α in the absence of RNA bound to H3K9me3, and to a lesser extent, also bound unmodified H3. When TERRA45 was added to the interaction, HP1α was indeed still able to bind H3K9me3, while HP1α binding to unmodified H3 decreased compared to when no TERRA45 was added (**Figure 3.25**). Together, these results suggest that addition of TERRA45 does not impede HP1α binding to H3K9me3, but more work should be done to understand the effect of TERRA on the HP1α-H3K9me3 interaction.

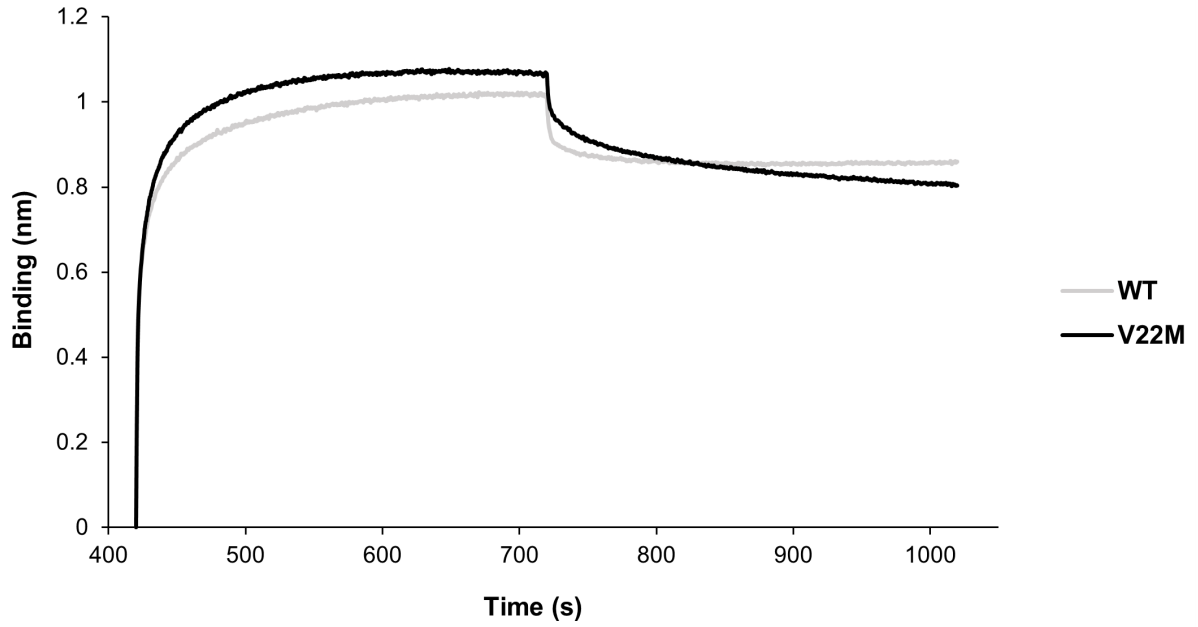

Figure 3.24: BLI analysis of HP1α WT or V22M with TERRA45. Binding of TERRA45 to HP1α V22M mutant measured by BLI.

Table 3.4: Binding affinities for HP1α V22M mutant with TERRA45 measured by BLI

HP1α variant	K_D **(nM)**	K_D **error**	k_{on}**/1000 (1/Ms)**	k_{on} **error /1000**	k_{off}**×1000 (1/s)**	k_{off} **error ×1000**
WT	83	5	62.8	1.6	5.1	0.3
V22M	98	3	61.9	0.7	6.08	0.17

3.3 Discussion

Altering TERRA's parallel G4 formation through mutation and buffer conditions caused HP1α to bind more weakly to destabilised G4 structures. Despite the modest decrease in characteristic parallel G4 circular dichroism signal, binding of HP1α to mutated TERRA was undetectable under the conditions used in BLI in **Figure 3.2b**. However, a small amount of competition for HP1α binding is observed when 50x excess of mutated TERRA is added for the EMSA. Binding of HP1α to mutant TERRA at higher concentrations will be revisited in **Chapter 4**. Removal of sodium and potassium monovalent cations from the buffer solution resulted in a partial loss of HP1α binding to TERRA. Li^+-based buffers have been shown to substantially decrease G4 formation in DNA and RNA (Miserachs et al., 2016). While removal of monovalent cations

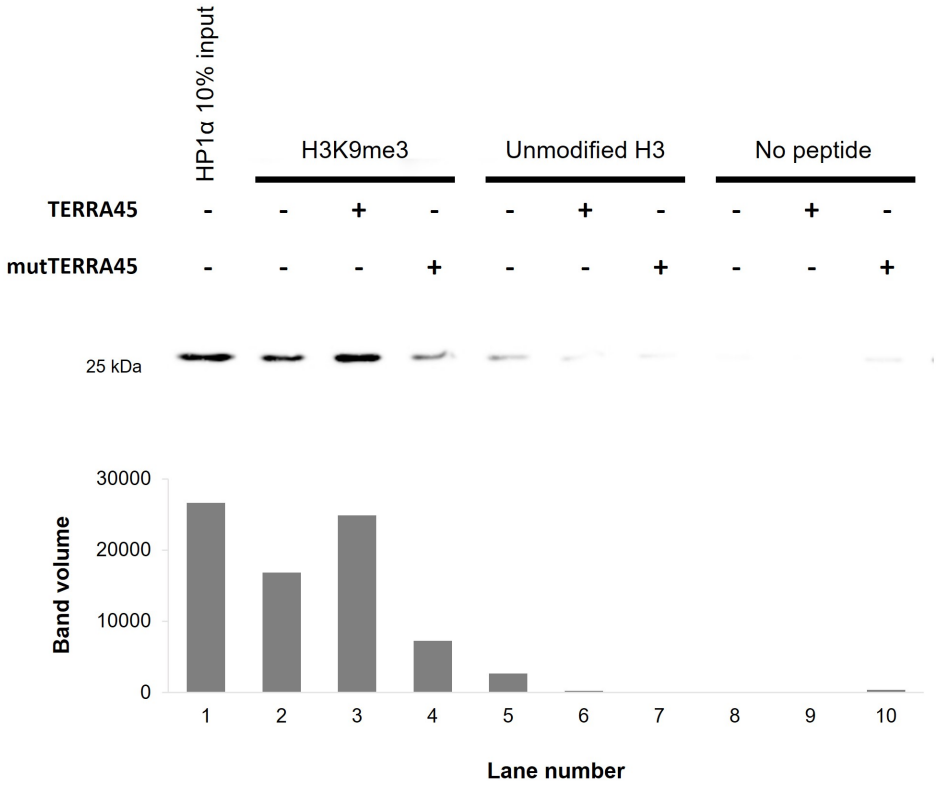

Figure 3.25: Peptide pulldown assay of HP1α binding to histone H3 with TERRA. Biotinylated H3K9me3 or unmodified H3 peptide was used to pull down HP1α in the presence or absence of TERRA45 or mutTERRA45, and samples were subjected to SDS-PAGE and western blotting using HP1α antibody (2616, Cell Signaling Technology). Band densitometry is shown below. Agarose streptavidin beads were used for this assay, with addition of 300 pmol biotin for blocking (**Section 2.9**).

from TERRA suggests HP1α binding is structure-specific and not sequence-specific, binding is not totally inhibited, likely due to the observation that parallel RNA G4 formation can also be detected when measuring circular dichroism spectra in water alone (Miserachs et al., 2016). The decrease in HP1α binding to destabilised G4s shows that this interaction is not sequence-specific, but structurally specific for the TERRA parallel G4 assembly.

By EMSA, HP1α was shown to bind fluorescently-tagged TERRA and impede its mobility to a much greater extent than HP1β or HP1γ, although some complex formation was indeed shown by HP1γ. Previous binding assays performed using BLI also showed a small amount of binding between TERRA and HP1γ (Roach et al., 2020), and this contradicted results published by Deng et al. (2009), who performed RNA IP of endogenous TERRA, and subsequently probed for each HP1 paralog and showed binding between HP1α and TERRA, but also HP1β and TERRA. HP1β is more likely to be present at heterochromatin in the nucleus and therefore come into contact

with endogenous TERRA. However, HP1γ does also function at human telomeres through an interaction with the shelterin complex (Canudas et al., 2011), of which TERRA also binds to components (Deng et al., 2009). This discrepancy between binding of HP1γ to TERRA could be an *in vitro* artefact in this case due to its relatively disordered structure compared with the other HP1 paralogs (**Figure 3.7a**). However, both HP1β and HP1γ could potentially bind weakly to TERRA due to each possessing two positively charged patches in their hinges (**Figure 3.1**). More work *in vivo* should be done to conclude the HP1 paralog specificity for TERRA.

Previously it was shown that mutation of 104-106KKK in the hinge of HP1α abrogated its binding to TERRA (Roach et al., 2020). Here, it was shown that three patches of positively charged residues in the HP1α hinge, one of which is unique to the HP1α paralog (68-72KKYKK), are each invaluable for binding TERRA (**Figure 3.10**). BLI showed that two out of three patches have weak binding but one or less is not sufficient for binding TERRA. However, by EMSA only the A3 mutant (104-106KKK>AAA) showed some complex formation with TAM-TERRA45. The reason for this discrepancy in EMSA experiments is unknown (consistent across repeated assays), as the A3 mutant did not show the lowest k_{off} rate or the highest k_{on} rate by BLI (**Table 3.1**), nor did the secondary structure differ from WT HP1α (**Figure 3.12**). The C-terminus of the hinge where the 104-106 lysines are located may be somewhat inhibited from binding TERRA at times due to the proximity of the CSD. Circular dichroism spectroscopy showed the secondary structure of HP1α is unchanged when hinge lysine residues are mutated to uncharged alanines; therefore, the interaction with TERRA is primarily mediated by the positive charge in this intrinsically disordered domain through electrostatic forces. To further investigate hinge charge, mutants were created to disperse the positive charge on two individual charged patches by substituting lysine residues for arginine residues where the positive charge is more distributed. Lysine is a basic amino acid that has a side-chain with a pKa of 10.67, whereby the amine group on the side-chain will be protonated at a pH lower than its pKa, so carries a positive charge at physiological pH (∼7.4). Arginine has a side-chain containing a guanidino group, therefore because of the multiple nitrogen atoms, the positive charge present at physiological pH is more distributed and results in arginine having a higher pKa for its side-chain. The R2 and R3 mutations increase the pKa of the hinge side-chains, dispersing the positive charge; however, the net charge on the amino acids and indeed the isoelectric point of HP1α is unchanged. Alternative methods to increase the positive charge of the patches within the HP1α hinge should be explored to further probe the dependency of positive charge on binding

to TERRA.

Following the *in vitro* work described on the lack of binding between the HP1α A1;2;3 mutant and TERRA45, IP and immunofluorescence experiments were performed to explore the effect of this interaction *in vivo*. Mutation of all three charged patches within the hinge did not substantially alter the localisation of HP1α to heterochromatin in either NIH3T3 or MDA-MB-231 cells. Others have shown that mutation of 104-106KKK>AAA alone results in mis-localisation of HP1α to heterochromatin (Muchardt et al., 2002). The immunofluorescence experiments performed by Muchardt et al. (2002) also used the NIH3T3 cell line. However, these experiments used recombinantly expressed Glutathione S-transferase (GST)-tagged HP1α incubated over fixed cells. These overlay assays may not recapitulate the true localisation of endogenously expressed HP1α. In addition to this, the 26 kDa GST-tag could exacerbate the negative effects of the resulting HP1α 3KA mutant. This illustrates that the charged patches within the hinge are not the sole determinant of HP1α localisation to heterochromatin in the present experiments. Both the hinge and the CSD have been shown to be integral to nucleosome binding and H3K9me3 specificity by HP1α (Meehan et al., 2003; Mishima et al., 2013) and they also are required for localisation to heterochromatin (Smothers and Henikoff, 2001), so it is of great interest to further study the CSD in the function of HP1α. To visualise the interaction between HP1α and TERRA, this should be more directly assessed by techniques such as RNA-fluorescence *in situ* hybridisation (FISH), as the detection of the HP1α A1;2;3 hinge mutant by itself (**Figures 3.14-3.17**) does not necessarily demonstrate TERRA-HP1α function here. Preliminary IP experiments under conditions tested here were not able to illustrate the effect of the A1;2;3 hinge mutation *in vivo*, and indeed were unsuccessful in showing the interaction between WT HP1α and TERRA (**Figure 3.18**). The interaction that occurs between WT HP1α and TERRA45 has fast k_{on} and k_{off} rates *in vitro* (**Table 3.3**), which may make it difficult to analyse by the methods utilised here. However, using the same strategy and probe, Deng et al. (2009) were able to detect TERRA bound to both HP1α and HP1β *in vivo* by dot-blot using the HCT116 human colorectal carcinoma cell line. The MDA-MB-231 cell line was used in this work as it has been shown to have strong expression of TERRA (González-Vasconcellos et al., 2022); however, conditions for this *in vivo* work need to be further optimised, possibly using crosslinking to allow for analysis of HP1α binding to TERRA.

In order to determine the effect that phosphorylation of HP1α might have, phosphomimetics were used, substituting serine for glutamic acid to create an amino acid with a similar negative

charge and side-chain length to a phosphorylated serine. In all cases where these substitutions were made, binding of HP1α to TERRA decreased (**Figure 3.21**, **Figure 3.20**). Due to the negative charge possessed by both the side-chain of glutamic acid (or the phosphate group on a phosphoserine) and the phosphate backbone of RNA, this may cause repulsion between HP1α and TERRA. Certainly in the case of HP1α S11-14E, upon mutation, the N-terminus of HP1α now contains a tract of 10 uninterrupted amino acids with negative side-chains (**Figure 3.19**) which may explain why this mutant bound to TERRA less than the hinge phosphorylation mutants. Deletion of the NTE which harbours the N-terminal phosphorylation site also results in decreased binding of HP1α to TERRA (**Figure 3.22**). This decrease in binding may be due to the positively charged amino acids in the NTE, of which there are four prior to S11-14 (**Figure 3.19**), and substitutions of these basic amino acids to neutral residues should be created to test this possibility. Larson et al. (2017) showed by SAXS that NTE phosphorylation results in a more open conformation of HP1α which also potentially contributes to decreased TERRA binding. Phosphomimetics were used here to ensure that only specific sites were modified, as CK2 phosphorylates residues S11-14 but also S97 (Hiragami-Hamada et al., 2011). CK2-mediated phosphorylation of HP1α results in increased affinity for H3K9me3 over unmodified histone H3 (Nishibuchi et al., 2014) and ensures proper heterochromatin localisation of HP1α (Hiragami-Hamada et al., 2011), thereby modulating HP1α function. Phosphorylation may be an alternative method to localise HP1α to heterochromatin in the absence of G4 or during phases of the cell cycle when transcribed RNA G4s may be less prevalent, such as during metaphase in which S93 phosphorylation is associated (Hiragami-Hamada et al., 2011). Given these results showing the effect of phosphomimetics, it is of interest to replicate these findings using the analogous phosphorylated HP1α counterparts, as some have found that glutamic acid substitutions have not been sufficient to recapitulate the effect of phosphorylation of HP1α on LLPS (Larson et al., 2017), which may in turn affect the DNA binding behaviour of HP1α.

For future studies to determine the effect of disrupting HP1α-H3K9me3 binding, the V22M mutant was tested and found to have a negligible effect on the binding between HP1α and TERRA (**Figure 3.24**). Similarly, TERRA did not impede the binding of HP1α to H3K9me3, which indicates that TERRA may be able to function with HP1α in nucleosome binding activities for heterochromatin formation. Further investigations into the effect of TERRA on HP1α binding chromatin should be tested using nucleosomal arrays to determine if TERRA G4 binding alters the affinity of HP1α for chromatin or influences the local heterochromatin structure.

Chapter 4

Investigating the domains of HP1α that provide specificity for G4s

4.1 Introduction

Binding of HP1α to TERRA is dependent upon multiple charged patches within the hinge of HP1α critical for this electrostatic binding (**Chapter 3**). In other protein-G4 interactions, electrostatic loop binding is frequently paired with either top-stacking or groove-binding interactions (Meier-Stephenson, 2022), such as in the case of the parallel c-myc G4 where an RHAU α-helix stacks with a G-tetrad while electrostatic interactions are formed for loop-binding (Chen et al., 2018), or with the anti-parallel telomeric DNA G4 where positively charged residues of OnTEBP interact with the phosphate backbone and a tyrosine protrudes into the groove (Horvath and Schultz, 2001). Therefore, HP1α electrostatic binding through the hinge may also be paired with additional stabilising interactions with TERRA.

The hinge of HP1α is an intrinsically disordered domain (Qin et al., 2021), therefore despite its pivotal role in binding through electrostatic forces, may lack structural features that allow for the recognition of particular nucleic acid structures. Therefore, the neighbouring globular CD and CSD of HP1α should be investigated for their roles in TERRA G4 recognition. Interplay between domains of HP1α has previously been shown to influence its function in heterochromatin. Meehan et al. (2003) showed that the CD alone does not bind to chromatin or DNA *in vitro*, despite being able to bind H3K9me3, and this binding to chromatin requires the hinge. Mishima

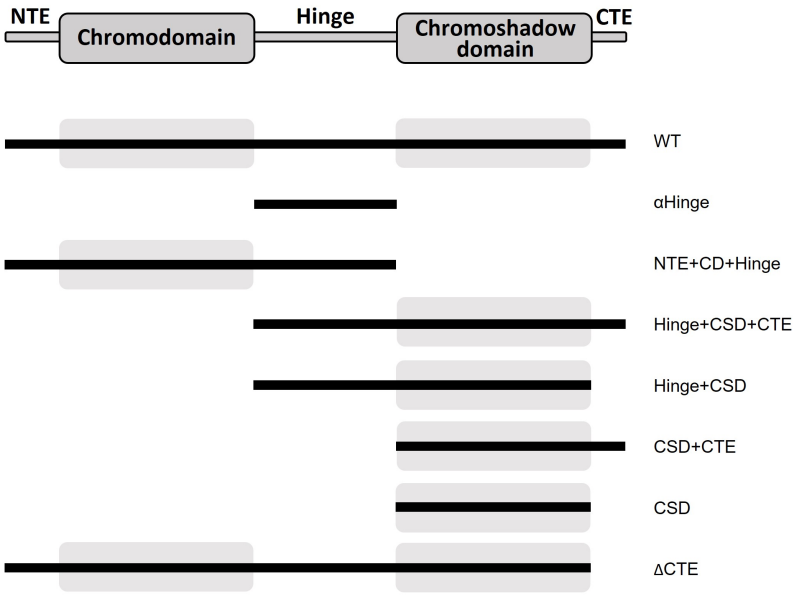

Figure 4.1: Deletion mutants of HP1α. Representation of deletion mutants of HP1α used in this chapter, corresponding to the domains of HP1α shown directly above.

et al. (2013) showed that the hinge and CSD cooperatively contribute to the binding of HP1α to nucleosomes, and that the CSD is required for co-localisation with H3K9me3 *in vivo*. Therefore, the neighbouring domains of the hinge may also contribute to its function in G4 binding.

To determine the mechanism behind the recognition and selectivity imparted by HP1α, the domains of HP1α were tested for their involvement in nucleic acid binding specificity for a range of different structured DNA and RNA.

4.2 Results

4.2.1 The HP1α hinge exhibits non-specific binding to nucleic acids

The clusters of positively charged amino acids predominantly responsible for binding TERRA are within the unstructured hinge of HP1α (**Section 3.2.3**). Therefore, to test the ability of the HP1α hinge to bind nucleic acids, an HP1α construct containing only the hinge region (plus his-tag) was created (**Figure 4.1**) by inverse PCR (**Section 2.2.2**) and recombinantly expressed and purified (**Section 2.5**). The HP1α hinge-only protein (αHinge) was then used to investigate if this intrinsically disordered peptide retains the specificity displayed by full-length HP1α.

A range of nucleic acid structures were tested for their interaction with the αHinge. Firstly, as previously discussed in **Section 3.2.1**, TERRA45 (parallel RNA G4), mutTERRA45 (mutated TERRA45, partial parallel G4 formation), and gTEL45 (G-rich telomeric DNA, anti-parallel G4) were analysed, in addition to cTEL45, dsTEL45, tRNA, and thy45 (**Table A.1**). The i-motif-forming DNA oligonucleotide, cTEL45, is complementary to gTEL45. Double-stranded dsTEL45 is produced by annealing the G- and C-rich telomeric DNA strands together. The structured RNA, tRNA, folds into a characteristic "cloverleaf" structure of three hairpin loops. Finally, to test an unstructured single-stranded DNA oligonucleotide, a string of 45 thymidine nucleotides, thy45, was also used for analysis.

To analyse the interactions of these oligonucleotides with HP1α WT and αHinge, BLI was used (**Section 2.8**). Shown in **Figure 4.2a** and **Table 4.1** the αHinge bound with high affinity (17.3 nM) to TERRA45, over 4-fold stronger than the interaction between WT HP1α and TERRA45 (82.7 nM, **Table 3.3**). Surprisingly, the αHinge also bound to mutTERRA45, gTEL45, cTEL45, dsTEL45, tRNA, and thy45. Despite the seemingly indiscriminate ability to bind, the αHinge did still show some preference for TERRA45.

The set of nucleic acids were then tested for their binding to WT HP1α by BLI for comparison. Shown in **Figure 4.2b** and **Table 4.1**, HP1α bound with high affinity to TERRA45, weakly to mutTERRA45, and did not show binding to gTEL45, cTEL45, dsTEL45, tRNA, or thy45, reiterating the high specificity of HP1α towards parallel G4s.

To analyse the topologies of the structures that the αHinge binds, the nucleic acids were subjected to circular dichroism spectroscopy (**Section 2.7**). The circular dichroism spectrum of cTEL45 showed very weak i-motif formation (**Figure 4.3**), where positive ellipticity at 285 nm and negative ellipticity at 255 nm would indicate stable i-motif structure (Manzini et al., 1994; Zeraati et al., 2018). This lack of strong i-motif structure formation is likely due to the slightly basic pH of the buffer used in these experiments (pH 8, **Section 2.1**). The dsDNA circular dichroism spectrum showed typical features of B-form double-stranded DNA (**Figure 4.3**), with maxima around 260 nm and 280 nm, and a minimum at 245 nm (Vorlíčková et al., 2012). The thy45 circular dichroism spectrum showed relatively low peaks (**Figure 4.3**), with a minimum at 250 nm and a maximum at 270 nm which corresponds with typical single-stranded DNA (Steely Jr et al., 1986). The tRNA showed a characteristic circular dichroism spectrum of the folded cloverleaf structure (**Figure 4.3**), with a maximum at 262 nm and a small negative ellipticity

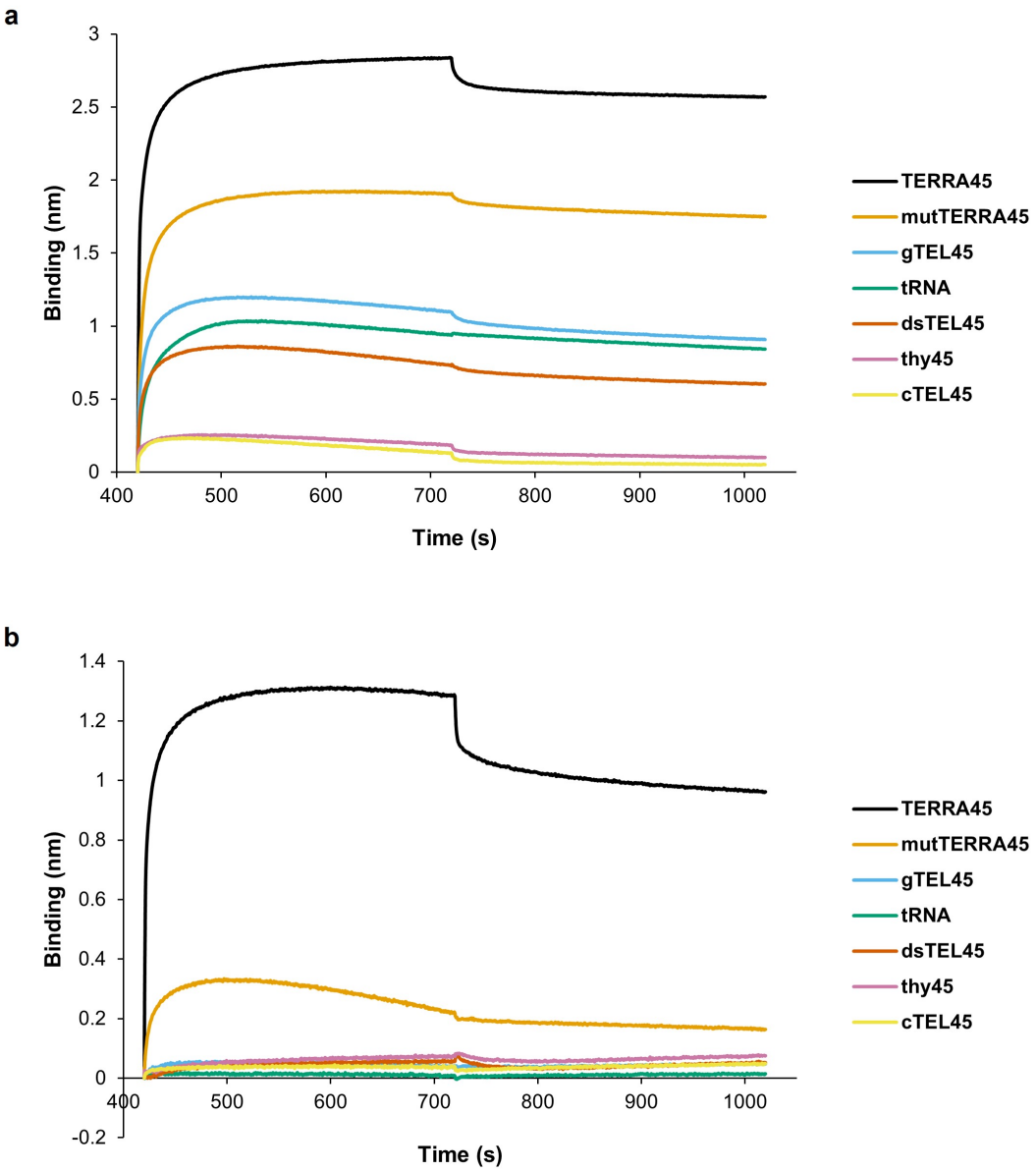

Figure 4.2: αHinge and full-length HP1α interactions with nucleic acids assessed by BLI. a) BLI measurements of αHinge and b) HP1α WT with addition of oligonucleotides listed in the legend above. Note that the concentration of oligonucleotide used in these experiments compared to **Section 3.2.1** has been increased from 0.5 µM to 2 µM to ensure detectable results with weak binders.

Table 4.1: Binding affinities for the αHinge and HP1α WT with various nucleic acids measured by BLI

Protein	**Nucleic acid**	K_D **(nM)**	K_D **error**	k_{on}**/1000 (1/Ms)**	k_{on} **error /1000**	k_{off}**×1000 (1/s)**	k_{off} **error ×1000**
αHinge	TERRA45	17.3	1.1	129	1.8	2.23	0.13
	mutTERRA45	16.0	0.5	87.3	0.5	1.40	0.05
	gTEL45	68.3	1.9	83.8	0.9	5.72	0.15
	tRNA	22.9	0.6	50.6	0.5	1.16	0.03
	dsTEL45	119	5	86.4	1.6	10.3	0.4
	thy45	1030	120	92	10	95	6
	cTEL45	10000	11000	26	30	250	30
WT	TERRA45	83	5	62.8	1.6	5.1	0.3
	mutTERRA45	1600	600	32	11	51	4
	gTEL45	No binding					
	tRNA	No binding					
	dsTEL45	No binding					
	thy45	No binding					
	cTEL45	No binding					

around 235 nm (Blum et al., 1972). Therefore, the αHinge tended to bind more strongly to the nucleic acids that formed stable secondary structures, with unstructured thy45 and cTEL45 binding in the micromolar range, while TERRA45, mutTERRA45, gTEL45, dsTEL45, and tRNA bound with nanomolar affinity (**Table 4.1**).

4.2.2 The chromoshadow domain of HP1α confers specificity for nucleic acid binding

To determine if another domain is involved in conferring selectivity, HP1α domain deletion mutants were created by inverse PCR (**Figure 4.1**). These proteins were then expressed, purified, and tested for their interaction with the suite of nucleic acids as above by BLI. The NTE+CD+Hinge, lacking the C-terminal CSD and CTE, bound similarly indiscriminately as the αHinge, binding to TERRA45, mutTERRA45, gTEL45, dsTEL45, and tRNA (**Figure 4.4**, **Table 4.2**). However, the Hinge+CSD+CTE mutant, lacking the NTE and CD, retained the specificity shown by the full-length HP1α, binding with high affinity to TERRA45 and moderately bound mutTERRA45, but also bound weakly to gTEL45 (**Figure 4.5**, **Table 4.3**).

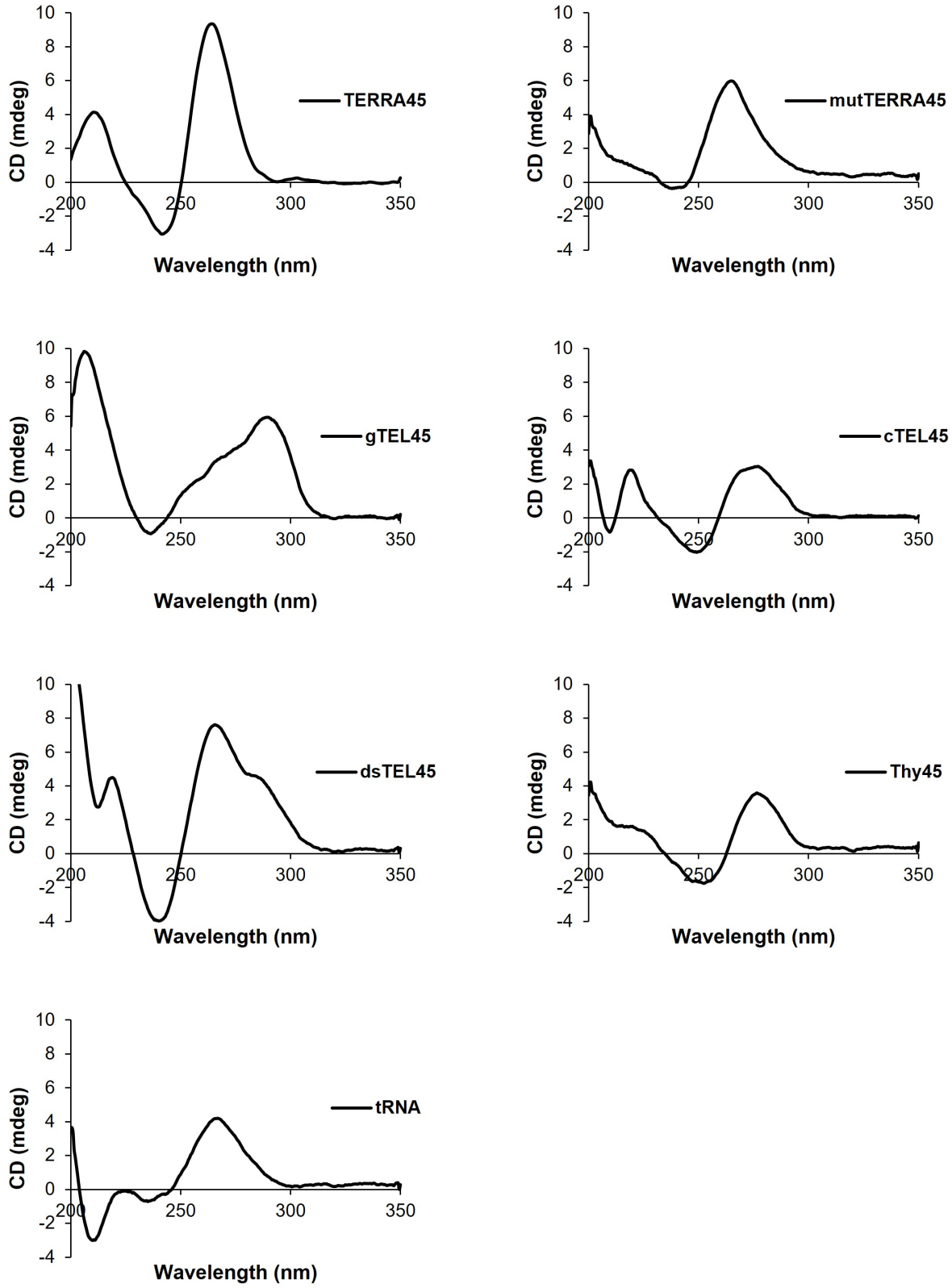

Figure 4.3: Circular dichroism spectra of oligonucleotides. Circular dichroism spectra of TERRA45, mutTERRA45, gTEL45, cTEL45, dsTEL45, tRNA, and thy45 recorded at a concentration of 5 µM in IB1 in a 1 mm cuvette. Measurements performed at room temperature.

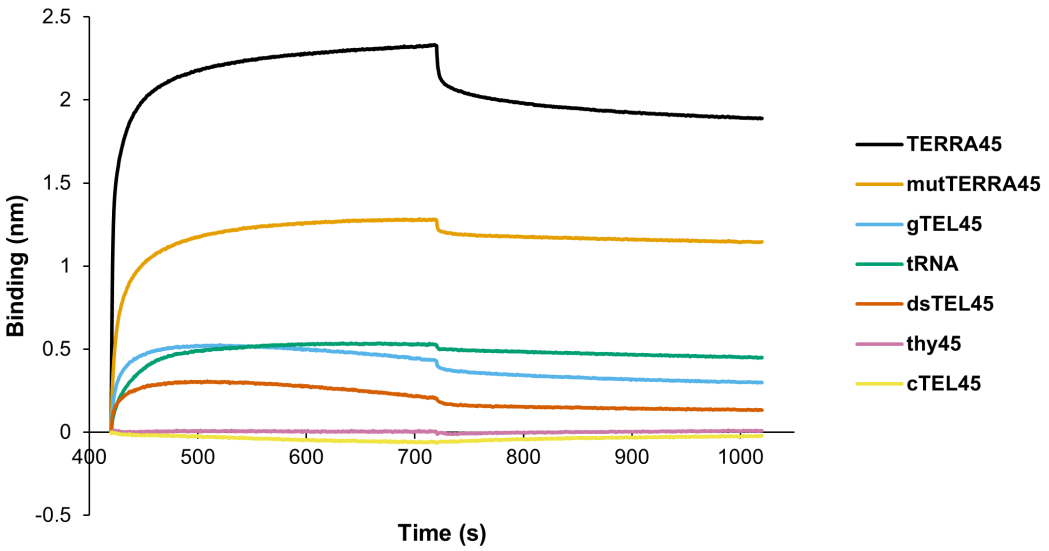

Figure 4.4: HP1α NTE+CD+Hinge interactions with nucleic acids assessed by BLI. BLI measurements of HP1α NTE+CD+Hinge with addition of oligonucleotides listed in the legend above.

Table 4.2: Binding affinities for HP1α NTE+CD+Hinge with various nucleic acids measured by BLI

Nucleic acid	K_D **(nM)**	K_D **error**	k_{on}**/1000 (1/Ms)**	k_{on} **error /1000**	k_{off}**×1000 (1/s)**	k_{off} **error ×1000**
TERRA45	54	3	84.8	1.3	4.6	0.2
mutTERRA45	17.9	0.8	52.8	0.6	0.94	0.04
gTEL45	324	16	60.1	1.9	19.5	0.7
tRNA	44.2	0.7	29.56	0.11	1.31	0.02
dsTEL45	1900	500	32	9	62	4
thy45	No binding					
cTEL45	No binding					

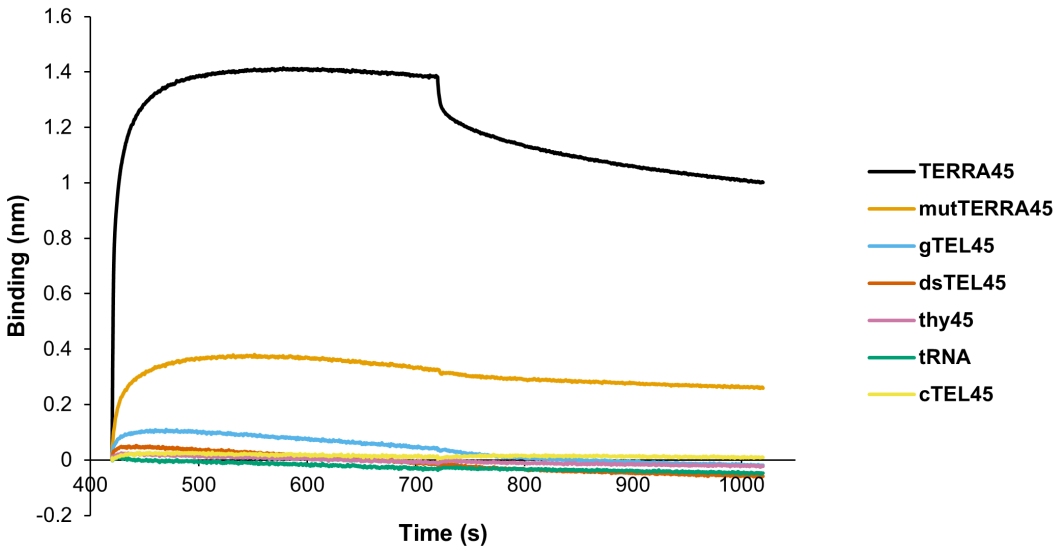

Figure 4.5: HP1α Hinge+CSD+CTE interactions with nucleic acids assessed by BLI. BLI measurements of HP1α Hinge+CSD+CTE with addition of oligonucleotides listed in the legend above.

Table 4.3: Binding affinities for HP1α Hinge+CSD+CTE with various nucleic acids measured by BLI

Nucleic acid	K_D **(nM)**	K_D **error**	k_{on}**/1000 (1/Ms)**	k_{on} **error /1000**	k_{off}**×1000 (1/s)**	k_{off} **error ×1000**
TERRA45	43.8	0.8	85.8	0.9	3.75	0.06
mutTERRA45	130	4	50.7	0.7	6.6	0.2
gTEL45	380	80	210	40	78	5
tRNA	No binding					
dsTEL45	No binding					
thy45	No binding					
cTEL45	No binding					

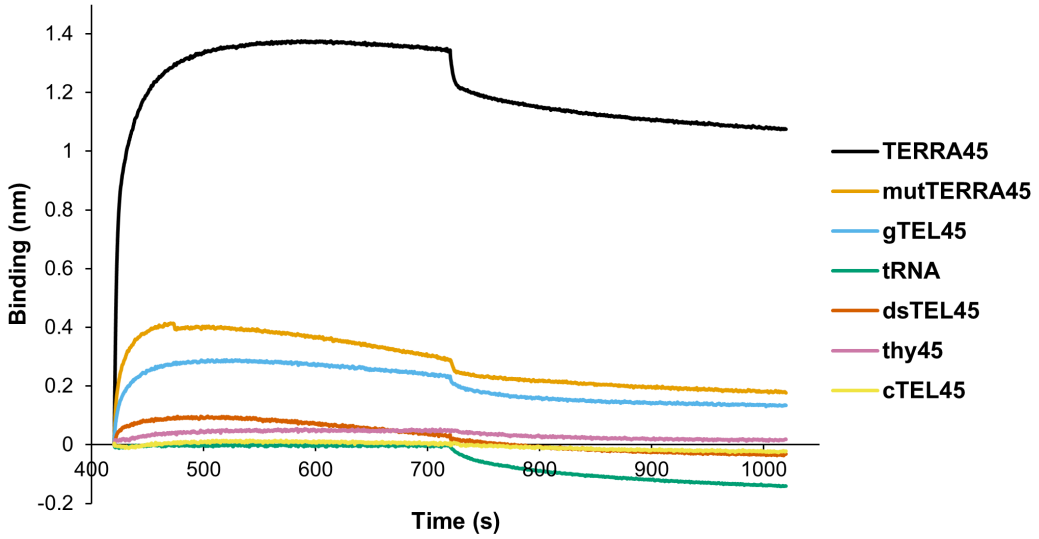

Figure 4.6: HP1α Hinge+CSD interactions with nucleic acids assessed by BLI. BLI measurements of HP1α αHinge with addition of oligonucleotides listed in the legend above.

Table 4.4: Binding affinities for HP1α Hinge+CSD with various nucleic acids measured by BLI

Nucleic acid	K_D **(nM)**	K_D **error**	k_{on}**/1000 (1/Ms)**	k_{on} **error /1000**	k_{off}**×1000 (1/s)**	k_{off} **error ×1000**
TERRA45	87	3	68.8	0.8	6.00	0.17
mutTERRA45	2600	1000	28	11	73	5
gTEL45	398	13	48.2	1.2	19.2	0.5
tRNA	No binding					
dsTEL45	3800	1600	39	16	149	15
thy45	No binding					
cTEL45	No binding					

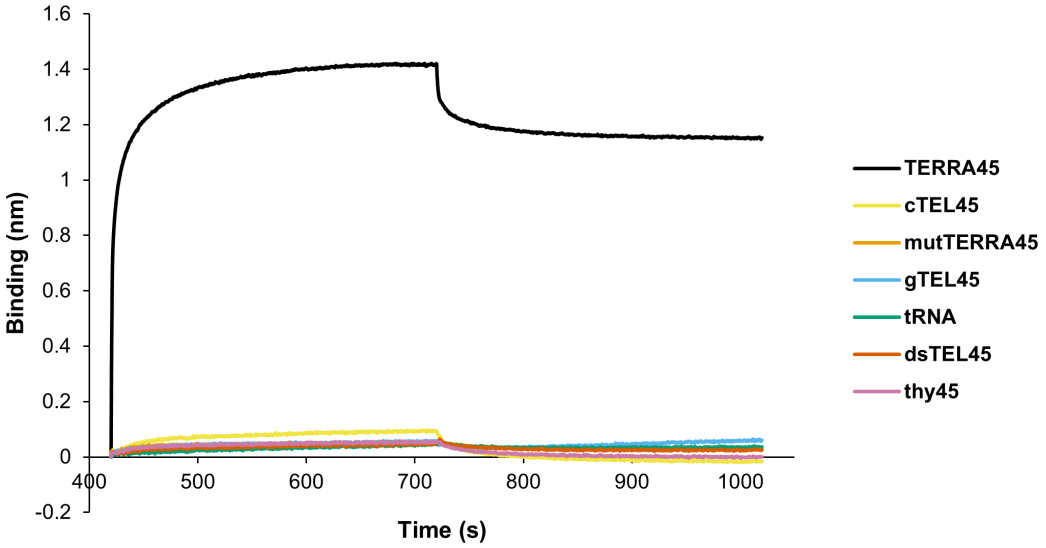

Figure 4.7: HP1α ΔCTE interactions with nucleic acids assessed by BLI. BLI measurements of HP1α ΔCTE with addition of oligonucleotides listed in the legend above.

Table 4.5: Binding affinities for HP1α ΔCTE with various nucleic acids measured by BLI

Nucleic acid	K_D **(nM)**	K_D **error**	k_{on}**/1000 (1/Ms)**	k_{on} **error /1000**	k_{off}**×1000 (1/s)**	k_{off} **error ×1000**
TERRA45	90	5	73.0	1.7	6.6	0.3
mutTERRA45	No binding					
gTEL45	No binding					
tRNA	No binding					
dsTEL45	No binding					
thy45	No binding					
cTEL45	No binding					

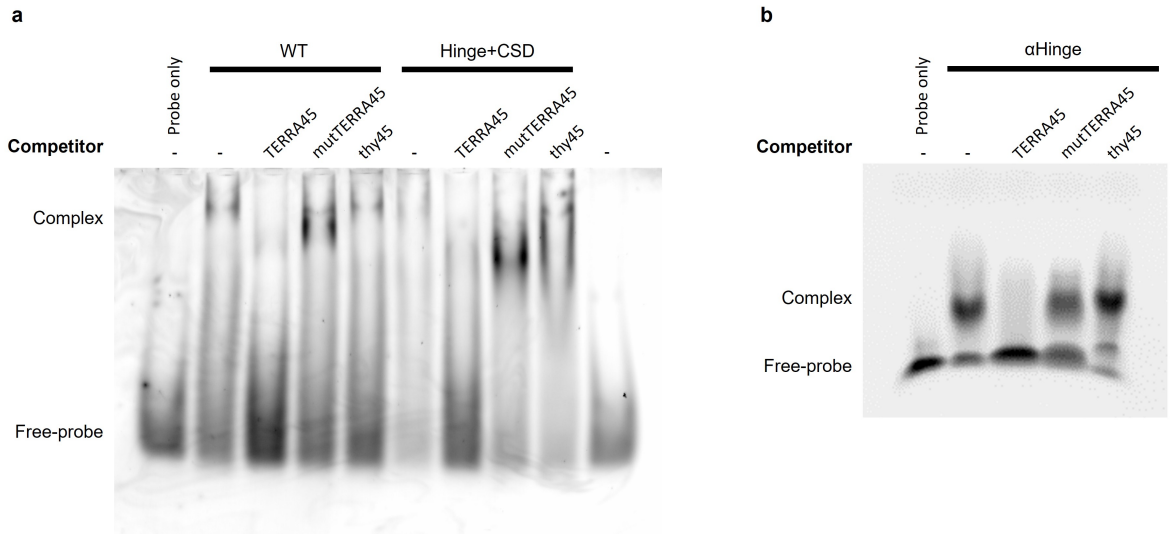

Figure 4.8: Competition EMSAs of HP1α deletion mutants. (A) Polyacrylamide gel EMSA competition assay with TAM-TERRA45 with addition of HP1α WT and Hinge+CSD (40 fold) and absence (-) or presence of molar excess (100 fold) of the listed oligonucleotide. (B) Agarose gel EMSA competition assay with TAM-TERRA45 with addition of HP1α αHinge and absence (-) or presence of molar excess of the listed oligonucleotide.

This suggests the region of HP1α conferring specificity is within either the CSD or the CTE.

To determine the contribution of the CTE in the selectivity of HP1α, mutants were created lacking the CTE (ΔCTE and Hinge+CSD, **Figure 4.1**). BLI analysis showed that for both Hinge+CSD and ΔCTE, specificity was retained, whereby the Hinge+CSD bound similarly to TERRA45 and mutTERRA45 as WT HP1α, but also had weak binding to gTEL45 and micromolar affnity to dsTEL45, while the ΔCTE mutant only bound to TERRA45 (**Figure 4.7**, **Table 4.5**, **Figure 4.6**, **Table 4.4**). These results were then corroborated by EMSA (**Section 2.4.2**), where HP1α WT, Hinge+CSD, or αHinge was incubated with TAM-TERRA45, and electrophoresed on a native polyacrylamide gel or agarose gel for the αHinge due to its low molecular weight. TERRA45, mutTERRA45, and thy45 were added in 100-fold molar excess to test the ability of each nucleic acid to compete for binding to the protein-TAM-TERRA45 complex. HP1α WT shifted the TAM-TERRA45 probe, and the complex formation was then diminished upon addition of TERRA45. Neither mutTERRA45 nor thy45 were able to compete for binding. The same pattern of competition was shown for Hinge+CSD (**Figure 4.8a**), showing that this protein retains the specificity of WT HP1α. The αHinge also shifted the TAM-TERRA45 probe and was competed using TERRA45, but showed some competition by mutTERRA45 (there is less shifted probe and more free probe compared with the Hinge+TAM-

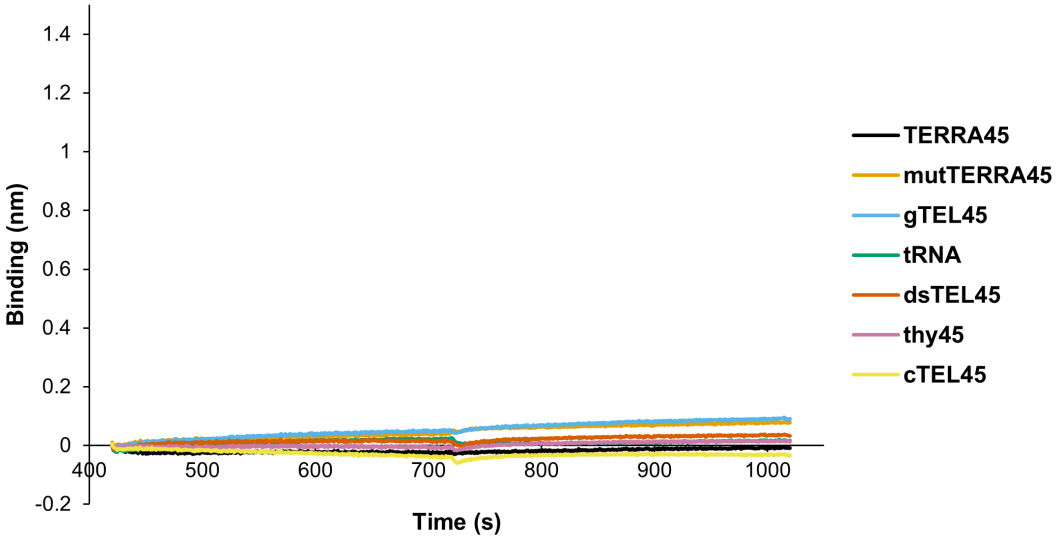

Figure 4.9: HP1α CSD interactions with nucleic acids assessed by BLI. BLI measurements of HP1α CSD with addition of oligonucleotides listed in the legend above.

TERRA45 and Hinge+TAM-TERRA45+thy45 lanes), and no competition from thy45 (**Figure 4.8b**), showing that some specificity has indeed been lost.

Because the CSD was determined to be the primary contributor to HP1α specificity, constructs were created deleting the hinge from the Hinge+CSD (CSD) and Hinge+CSD+CTE (CSD+CTE, **Figure 4.1**). These proteins were then tested for their interaction with the set of oligonucleotides. BLI showed that neither the CSD nor the CSD+CTE was capable of binding any of the oligonucleotides in the absence of the hinge region (**Figure 4.9**, **Figure 4.10**). Therefore, the CSD is responsible for imparting specificity for TERRA, but is not able to interact with nucleic acids by itself and requires the hinge for this purpose.

4.2.3 Dimerisation of HP1α is not required for binding nucleic acids

As the HP1α CSD is responsible for dimerisation, to determine if this is required for selective nucleic acid binding by HP1α, a mutant unable to dimerise was created where a tyrosine residue at position 168 within an alpha helix in the CSD (**Figure 4.11**) was substituted with a glutamic acid residue (Y168E, Brasher et al. 2000).

The lack of dimerisation of HP1α Y168E was confirmed by SEC (**Figure 4.12**) and the molecular weight was determined by SAXS (further explored in **Chapter 5**), estimating the Y168E molec-

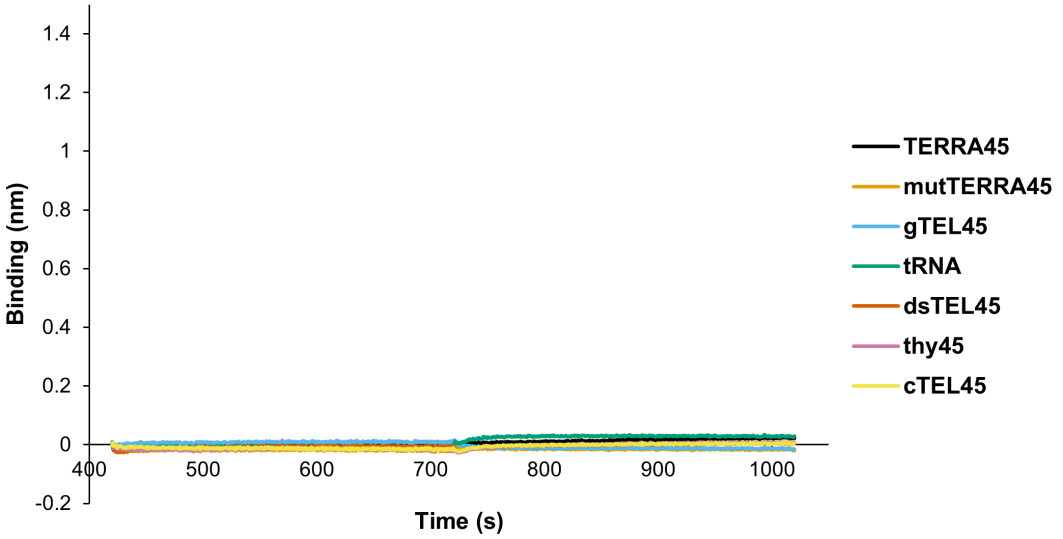

Figure 4.10: HP1α CSD+CTE interactions with nucleic acids assessed by BLI. BLI measurements of HP1α CSD+CTE with addition of oligonucleotides listed in the legend above.

ular weight to be 22.3 kDa, and estimating the WT molecular weight to be 58.6 kDa (**Table 5.2**). HP1α Y168E was then tested for its interaction with nucleic acids by BLI. This monomeric HP1α mutant retained the specificity shown by WT dimeric HP1α, where Y168E bound strongly to TERRA45, weakly to mutTERRA45, and had no detectable binding to gTEL45, tRNA, dsTEL45, thy45, or cTEL45 (**Figure 4.13**, **Table 4.6**). This indicates that dimerisation is neither required for binding to TERRA45 nor involved in the specificity for binding parallel G4s.

The crystal structure of the HP1α CSD shows a potential second dimerisation interface in addition to the previously demonstrated α-helix interface (**Figure 4.11**). This additional interface is a β-sheet, and mutation of a leucine residue (L139) within this β-sheet has been shown to disrupt CSD specificity for peptides (Mendez et al., 2013). This leucine was therefore mutated to an aspartic acid residue (L139D) to disrupt hydrophobic interactions, and as shown in **Figure 4.14**, the L139D mutant showed similar binding affinity and specificity as displayed by the WT HP1α, indicating this interface is not involved in binding or specificity for parallel G4s.

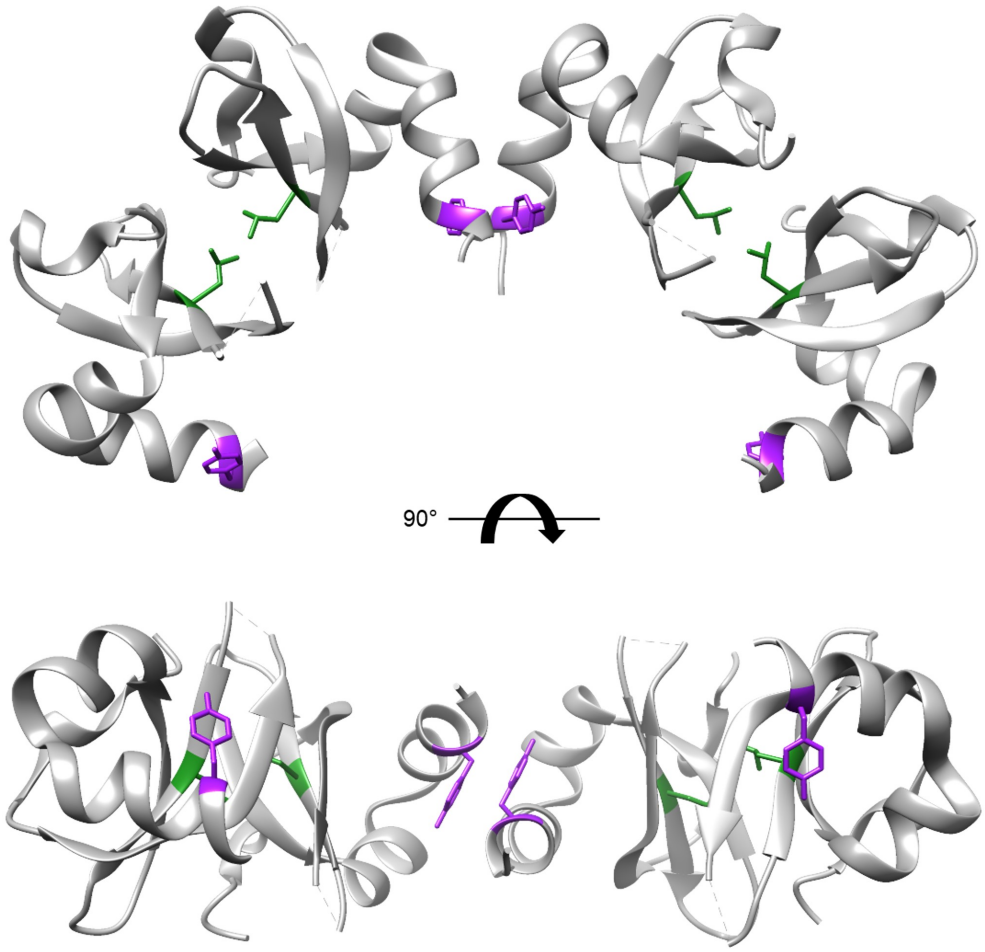

Figure 4.11: Interfaces between four HP1α CSDs. Crystal structure of the tetrameric CSD of HP1α, PDB:3I3C. Y168 is shown in purple, and L139 is shown in green, showing hydrophobic interactions between CSDs can occur between α-helices at the C-terminus, but also between β-sheets at the N-terminus. The bottom structure has been rotated 90° backwards from the top structure.

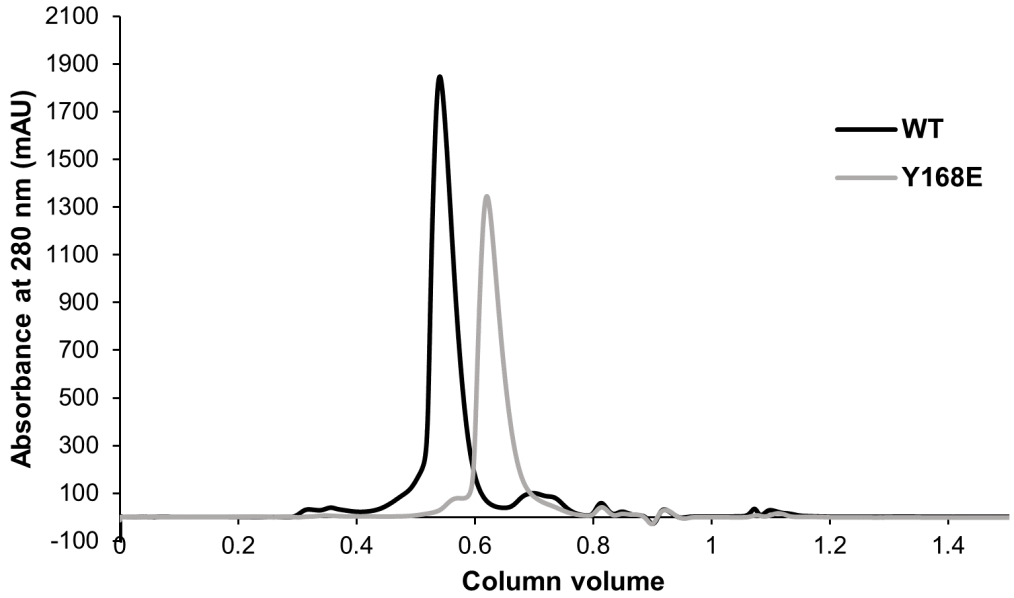

Figure 4.12: Size-exclusion chromatography of HP1α WT and Y168E. SEC using a Superdex 200 Increase 10/300 column of HP1α WT and Y168E showing relative size differences.

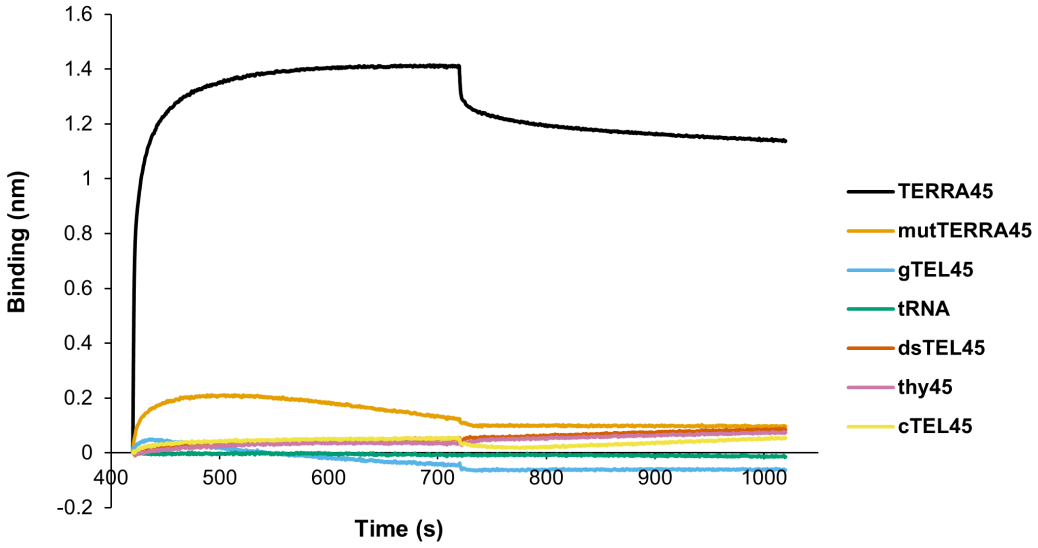

Figure 4.13: HP1α Y168E interactions with nucleic acids assessed by BLI. BLI measurements of HP1α Y168E with addition of oligonucleotides listed in the legend above.

Table 4.6: Binding affinities for HP1α Y168E with various nucleic acids measured by BLI

Nucleic acid	K_D **(nM)**	K_D **error**	k_{on}**/1000 (1/Ms)**	k_{on} **error /1000**	k_{off}**×1000 (1/s)**	k_{off} **error ×1000**
TERRA45	73	3	80.1	1.2	5.8	0.2
mutTERRA45	3200	2000	23	17	74	8
gTEL45	No binding					
tRNA	No binding					
dsTEL45	No binding					
thy45	No binding					
cTEL45	No binding					

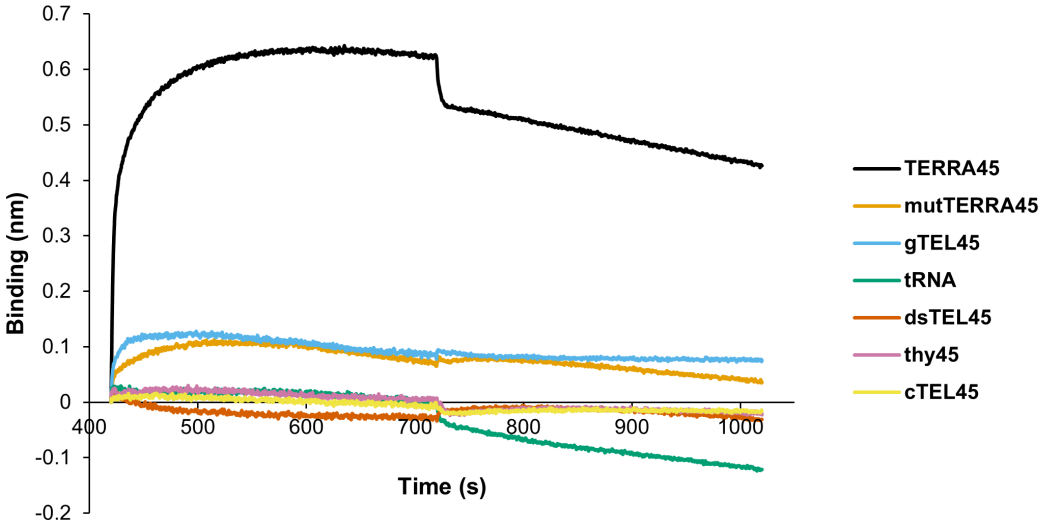

Figure 4.14: HP1α L139D interactions with nucleic acids assessed by BLI. BLI measurements of HP1α L139D with addition of oligonucleotides listed in the legend above.

Table 4.7: Binding affinities for HP1α L139D with various nucleic acids measured by BLI

Nucleic acid	K_D **(nM)**	K_D **error**	k_{on}**/1000 (1/Ms)**	k_{on} **error /1000**	k_{off}**×1000 (1/s)**	k_{off} **error ×1000**
TERRA45	69.4	1.7	49.3	0.6	3.43	0.08
mutTERRA45	610	60	18.9	1.7	11.6	0.5
gTEL45	146	13	156	8	22.8	1.7
tRNA	No binding					
dsTEL45	No binding					
thy45	No binding					
cTEL45	No binding					

4.2.4 A helix interface within the chromoshadow domain of HP1α confers specificity towards parallel G4s

As Y168 within the second helix of the helix-turn-helix motif of the HP1α CSD did not alter nucleic acid binding or specificity, other interfaces were investigated. Among PDB structures present for parallel G4-interacting proteins, a common motif is the hydrophobic interaction between an α-helix in the protein or peptide with the exposed face of the parallel G-tetrad (Heddi et al., 2015; Chen et al., 2018; Meier-Stephenson, 2022). Y168 of HP1α is within the second helix of the helix-turn-helix motif, the interface for dimerisation of the HP1α CSD (**Figure 4.11**), but the first helix of the helix-turn-helix is of particular interest due to its position relative to

the hinge in the AlphaFold model of HP1α (AF-Q61686-F1, **Figure A.38**).

To test the possibility that a helix-tetrad contact may be stabilising the interaction between HP1α and parallel G4s, this interface was disrupted. Residues K154, E155, and V158 were mutated due to their positions on the outer surface of the helix in the globular CSD (**Figure 4.15**). The K154A and E155V mutants were produced to eliminate the positive and negative charges of the side-chains of these amino acids. The V158D mutant was made to introduce a negative charge in this area of hydrophobicity, while the V158G mutant was made to potentially hinder α-helix formation. The V158W mutant was produced to introduce a large amino acid side-chain to potentially sterically hinder the nucleic acid. Their interactions with nucleic acids were then assessed by BLI. To varying degrees, each helix mutant (K154A, E155V, V158D, V158G, and V158W) bound to TERRA45, mutTERRA45, and also gTEL45, but not to tRNA, dsTEL45, thy45, or cTEL45, displaying partial specificity (**Figure 4.16**, **Table 4.8**). The helix mutants with the greatest decrease in specificity towards TERRA45 were E155V and V158W, both showing nanomolar affinity for the hybrid/anti-parallel G4 gTEL45 (E155V and V158W bound with 222 nM and 602 nM, respectively). This increase in association between these helix mutants with either anti-parallel G4 or partially folded parallel G4 shows that specificity for parallel G4s has been lost upon disruption of the stabilising force of an α-helix within the CSD of HP1α.

To assess whether this change in specificity was due to a change in the helicity in the CSD of HP1α upon mutation, circular dichroism spectra were collected of the K154A, E155V, V158D, V158G, and V158W mutants. The circular dichroism spectra indicate no substantial difference in secondary structure between WT HP1α and the helix mutants (**Figure 4.17**), although a small increase in negative ellipticity was shown with the V158W and V158G mutants (negative peaks at 222 nm and 207 nm). This shows that the global secondary structure of HP1α was not substantially altered upon mutation of the helix within the CSD; therefore, the amino acid changes and not the secondary structure itself are responsible for nucleic acid specificity changes.

4.3 Discussion

It is widely reported that the hinge of HP1α binds DNA and RNA both *in vitro* and *in vivo* (Muchardt et al., 2002; Meehan et al., 2003; Maison et al., 2011; Mishima et al., 2013; Ryan and

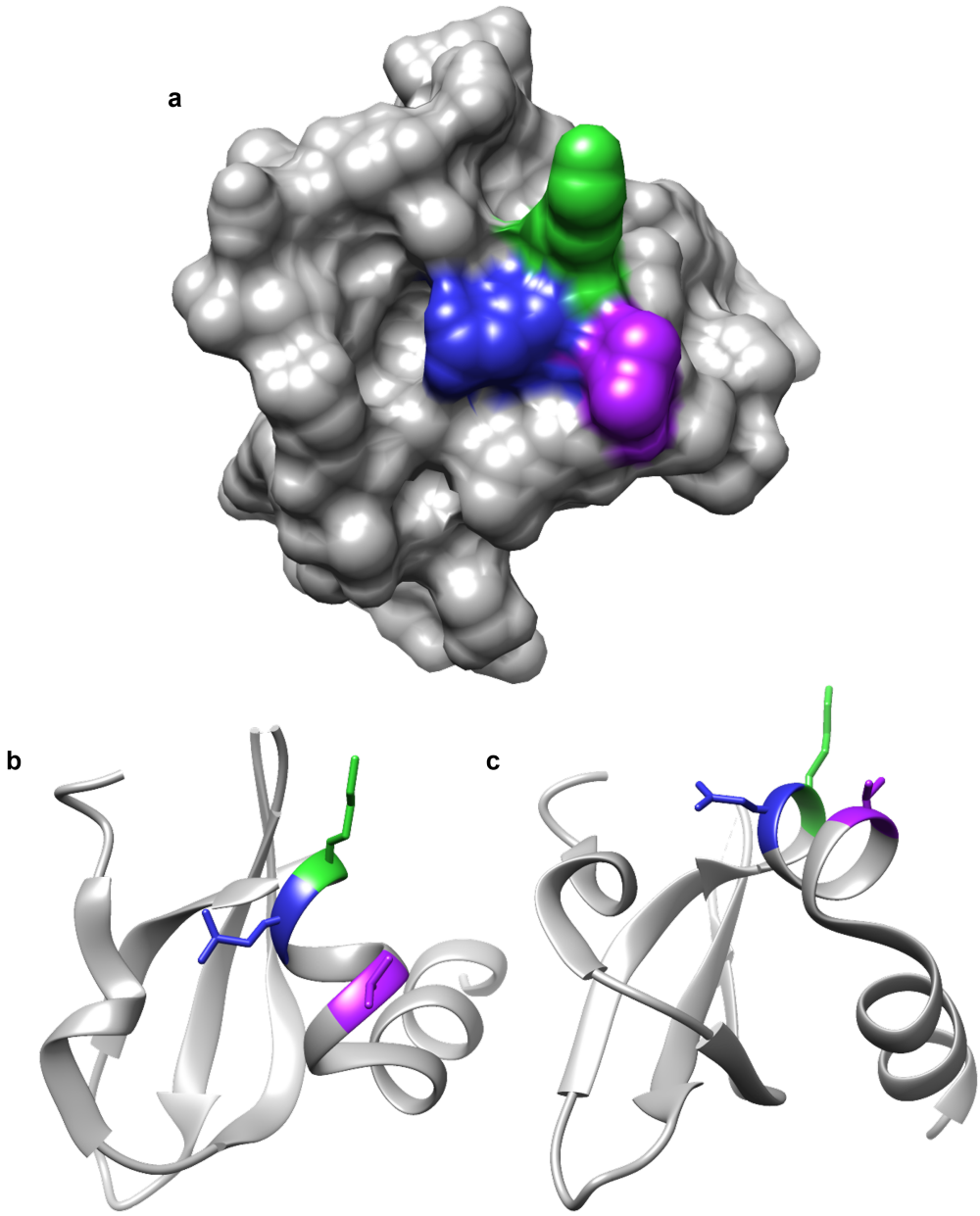

Figure 4.15: Residues on the surface of a helix of the CSD. Model of one chain of the crystal structure of the HP1α CSD (PDB:3I3C) with K154 coloured in green, E155 in blue, and V158 in purple. a) The surface of 3I3C. b) The same angle as in a), but shown as a ribbon. c) The same as in b), but rotated 90°.

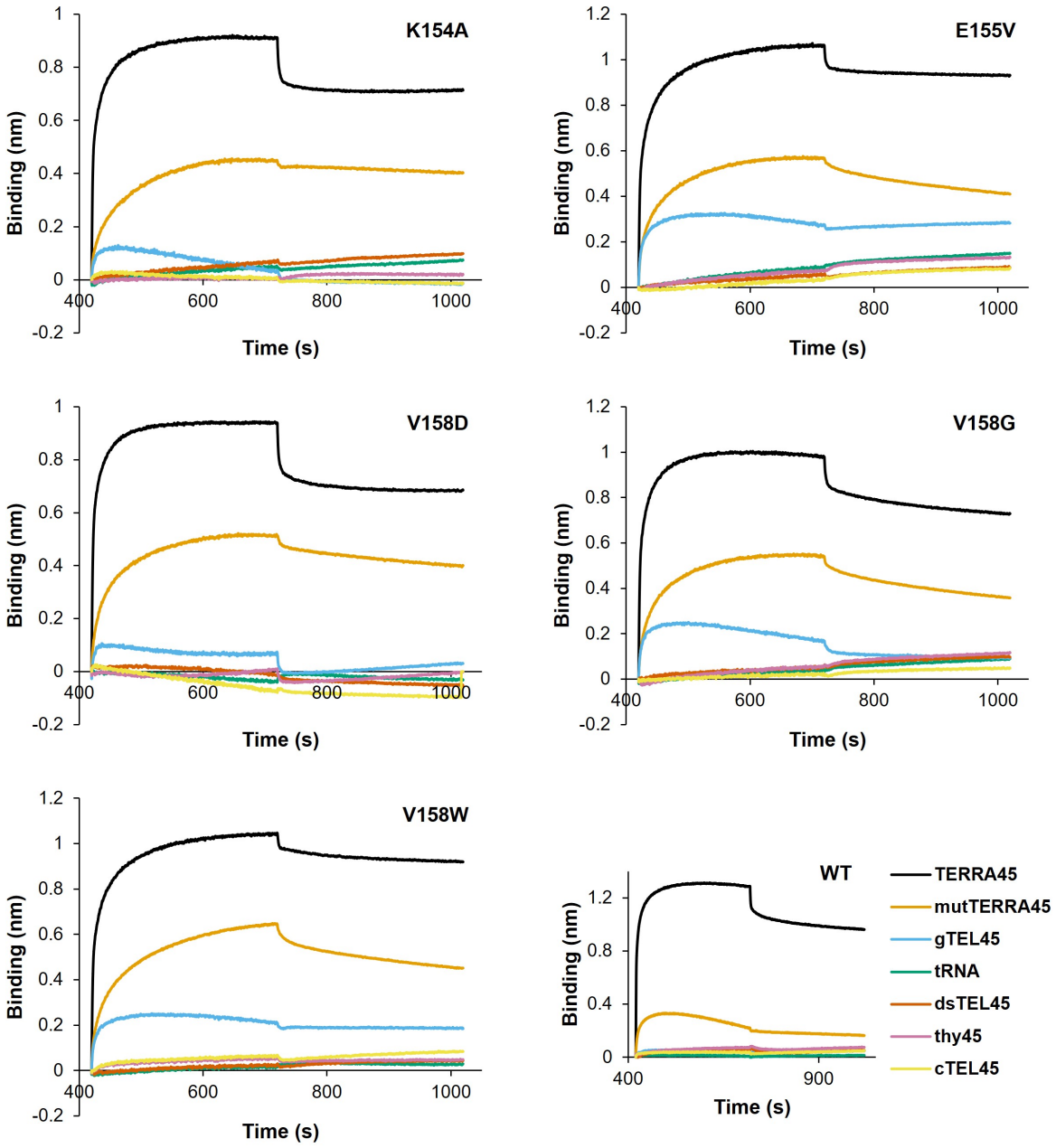

Figure 4.16: HP1α helix mutant interactions with nucleic acids assessed by BLI. BLI measurements of HP1α helix mutants K154A, E155V, V158D, V158G, and V158W with addition of oligonucleotides listed in the legend above. WT HP1α has been included for reference.

Table 4.8: Binding affinities for HP1α helix mutants with various nucleic acids measured by BLI

Protein	**Nucleic acid**	K_D **(nM)**	K_D **error**	k_{on}**/1000 (1/Ms)**	k_{on} **error /1000**	k_{off}**×1000 (1/s)**	k_{off} **error ×1000**
K154A	TERRA45	201	11	50.5	1.9	10.1	0.4
	mutTERRA45	57.3	1.9	13.1	0.2	0.75	0.02
	gTEL45	200000	8000000	3	90	620	130
	tRNA	No binding					
	dsTEL45	No binding					
	thy45	No binding					
	cTEL45	No binding					
E155V	TERRA45	28.7	1.7	51.6	1.0	1.48	0.08
	mutTERRA45	142	3	16.07	0.14	2.28	0.04
	gTEL45	220	30	39	5	8.5	0.5
	tRNA	No binding					
	dsTEL45	No binding					
	thy45	No binding					
	cTEL45	No binding					
V158D	TERRA45	157	5	115	3	17.9	0.5
	mutTERRA45	123	3	16.9	0.2	2.08	0.04
	gTEL45	280000	140000	13	60	380	50
	tRNA	No binding					
	dsTEL45	No binding					
	thy45	No binding					
	cTEL45	No binding					
V158G	TERRA45	49	3	48.9	1.4	2.40	0.15
	mutTERRA45	189	3	16.51	0.13	3.13	0.04
	gTEL45	5000	4000	21	15	108	8
	tRNA	No binding					
	dsTEL45	No binding					
	thy45	No binding					
	cTEL45	No binding					
V158W	TERRA45	21.7	1.3	55.1	0.8	1.19	0.07
	mutTERRA45	215	6	11.1	0.2	2.39	0.05
	gTEL45	600	100	31	5	18.4	0.8
	tRNA	No binding					
	dsTEL45	No binding					
	thy45	No binding					
	cTEL45	No binding					

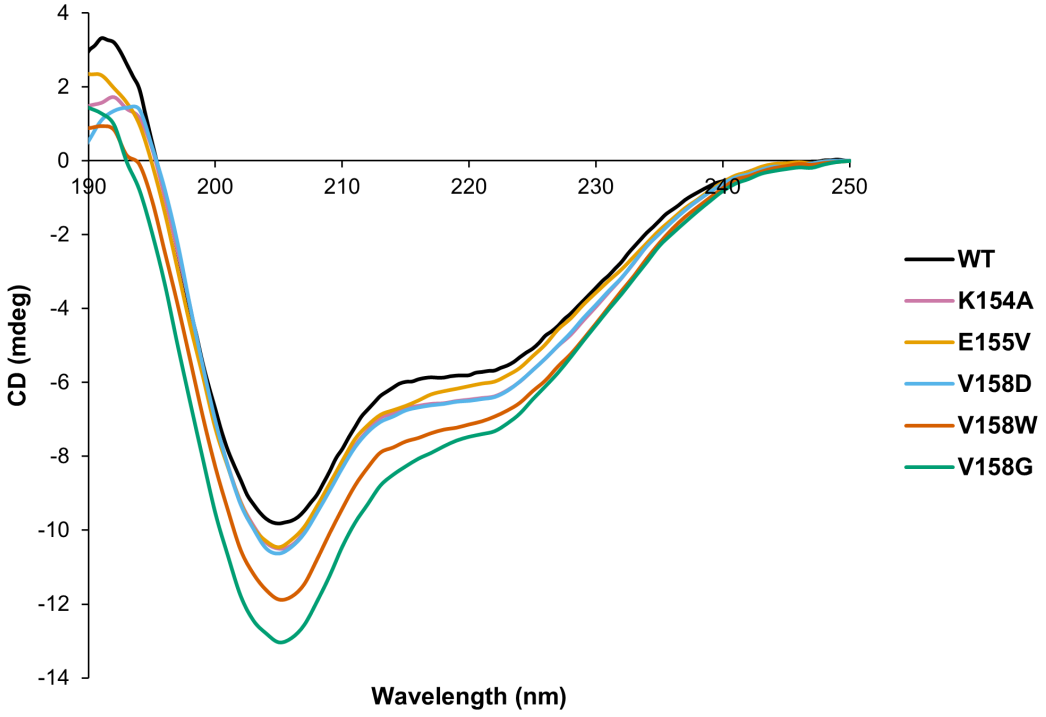

Figure 4.17: Circular dichroism spectra of HP1α CSD helix mutants. Far UV circular dichroism spectra of HP1α WT, K154A, E155V, V158D, V158G, and V158W measured at 1.5 mg/mL in a 0.1 mm cuvette in 1xIB.

Tremethick, 2018), but the specificity of HP1α binding nucleic acids through the unstructured hinge has not been previously explored. Despite the selectivity for parallel G4s shown here, HP1α has previously been shown to bind nucleosomal and linker DNA in chromatin studies (Mishima et al., 2013; Ryan and Tremethick, 2018). In a previous study, HP1α was shown to display increased affinity based upon the length of TERRA (Roach et al., 2020), with a 96 nt TERRA displaying a stronger 30-fold decrease in K_D over the 45 nt TERRA. Therefore, the binding to DNA in nucleosomal studies may be due to the larger length of oligonucleotides used compared to the short nucleic acids tested here. In the present study, oligonucleotides were intentionally synthesised to be similar in molecular weight due to the importance of size in BLI assays (**Figure 3.3**). Future work should include interactions of HP1α with longer DNA for comparison. When purified without neighbouring domains, the hinge of HP1α was sufficient to bind TERRA, however, the specificity towards parallel G4s was lost. The hinge of HP1α is intrinsically disordered due to the enrichment of polar and charged residues and lack of hydrophobic residues (Qin et al., 2021), which allows for uninhibited electrostatic interactions with the negative phosphate backbones of nucleic acids (Ottoz and Berchowitz, 2020). However, some specificity of the HP1α hinge for structured nucleic acids did remain with only this domain. The oligonu-

cleotides that bound more strongly to the αHinge were parallel G4 TERRA45, partially formed parallel G4 mutTERRA45, hybrid/anti-parallel G4 gTEL45, and hairpin-containing tRNA. The less structured oligonucleotides, single-stranded DNA thy45 and weak i-motif cTEL45 (complete formation of the structured i-motif of cTEL45 was not possible under these conditions at pH 8 due to the requirement of an acidic environment for the cytosine–cytosine$^+$ bond to occur, Gehring et al. 1993) did not bind with as high affinity to the αHinge (**Figure 4.3**, **Figure 4.2a**, **Table 4.1**). This indicates that while there is some non-selectivity displayed, the HP1α hinge is still somewhat preferential in its binding to nucleic acids, with its intrinsic disorder providing an adaptable interface which can be honed with the addition of a neighbouring globular domain.

In determining the region of HP1α that contributes structural specificity towards binding nucleic acids, the CSD was observed to confer this property. The CSD provides recognition for the PXVXL pentapeptide motif (Smothers and Henikoff, 2000; Mendez et al., 2013) and also homo- and heterodimerisation (Cowieson et al., 2000; Nielsen et al., 2001). In addition to these functions, the CSD has also been shown to be involved in targeting HP1α to heterochromatin foci (Smothers and Henikoff, 2001). The CSD shows distinct partner specificity for the PXVXL motif-containing peptides that it binds (Mendez et al., 2013), shown to be mediated by a leucine residue (in *Drosophila melanogaster* HP1a) that corresponds with L139, which was tested here for its specificity for nucleic acids upon mutation to aspartic acid. This mutation did not affect the binding ability to TERRA or nucleic acid specificity by HP1α. This shows that the β-sheet mediated recognition of the PXVXL motif is not involved nor required for nucleic acid binding or specificity. When investigating the function of dimerisation in nucleic acid binding, monomeric HP1α retained both its ability to bind parallel G4s, and also its distinct selectivity for them, showing that dimerisation through the CSD is also not necessary nor involved in G4 binding and nucleic acid specificity. Since the CSD showed no binding towards any nucleic acid, this indicates it was acting as a stabilising or docking structure for parallel G4s when interacting via the charge of the HP1α hinge. In heterochromatin, CSD dimerisation is required to prolong HP1α retention on chromatin (Kilic et al., 2015) and provide a scaffold for PXVXL-containing proteins (Kumar and Kono, 2020). It is unknown if the interaction with TERRA through stabilisation with the HP1α CSD affects dimerisation or PXVXL binding and should be investigated further to determine the effect of this interaction in a chromatin context.

To map the residues that may be stabilising the interaction with a parallel G4, it was hypothesised that it may be through an α-helix in the CSD, given the available protein-parallel G4

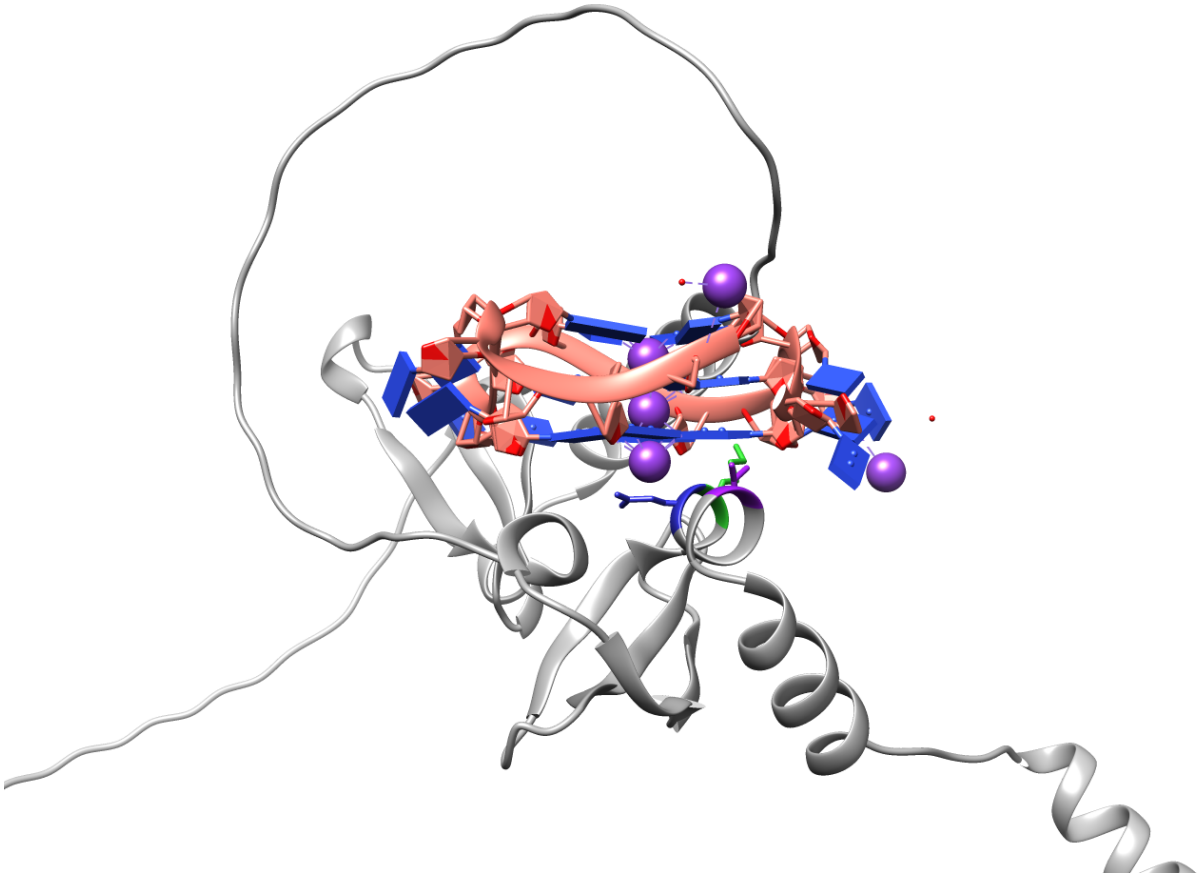

Figure 4.18: Model of CSD helix stabilisation with a parallel G4. HP1α AlphaFold structure (AF-Q61686-F1) with K154 in green, E155 in blue, and V158 in purple, with the crystal structure of a 22 nt K^+ stabilised TERRA molecule (PDB:7KLP, orange, K^+ atoms shown in purple). The unstructured hinge of HP1α (above) is free to interact with the negatively charged phosphate backbone of the RNA.

structures that show an α-helix interacting with a G-tetrad face such as RHAU (PDB:2N21, Heddi et al. 2015, PDB:5VHE, Chen et al. 2018) and RAP1 (PDB:6LDM, Traczyk et al. 2020). The first helix in the helix-turn-helix motif of the CSD was mutated, and while none of the mutations substantially affected binding to TERRA, they each had an impact on nucleic acid specificity (**Figure 4.16**, **Table 4.8**). The helix mutants shown to display the greatest decrease in specificity towards parallel G4s (where binding to gTEL45 occurred with stronger affinity) were E155V and V158W. In the case of E155V, the ability to maintain specificity toward parallel G4s therefore may be due to the electrostatic interaction between the negatively charged glutamic acid within the helix of the HP1α CSD and the positively charged ion in the centre of the G-tetrad (K^+ or Na^+), the interaction of which is disrupted upon mutation to an uncharged valine. Future work to mutate this glutamic acid residue to a positively charged amino acid such as lysine in order to replace the cation in the G-tetrad and further stabilise the CSD-G4 interaction would be of interest. In the case of V158W, the ability of the HP1α CSD to maintain specificity may also be due to the hydrophobic interaction between valine and the face of the G-tetrad, which is disrupted when mutated to tryptophan due to the large side-chain of this amino acid. In addition to these factors which contribute to the CSD stabilising the interaction with a parallel G4 tetrad face, the presence of the globular CSD may also impose a physical hindrance towards nucleic acids that are able to electrostatically interact with the hinge but do not possess the optimal structure to interact when docked by the CSD (**Figure 4.18**).

In delving into the interaction between HP1α mutants with TERRA G4s and other oligonucleotides, BLI was used for the calculation of binding affinities. ITC measurements have also been attempted numerous times with both HP1α WT and αHinge with TERRA and other parallel DNA G4s, however, with extremely low changes in heat of the samples (**Figure A.39**, **Figure A.40**, and Roach 2019). ITC measures the change in enthalpy (ΔH) caused by titration of components of an interaction, which can then be used to calculate the K_D. Spontaneous interactions with negative Gibbs free energy (ΔG) occur due to both ΔH and entropy (ΔS), which can be expressed as: $\Delta G = \Delta H - T\Delta S$ (where T is temperature in Kelvin, K). Therefore, even though an interaction may occur with sufficient negative ΔG, this may not necessarily mean that a large ΔH is detectable by ITC if the interaction is primarily driven by entropic changes. Hence, the HP1α-TERRA interaction may be more entropic despite the electrostatic interactions of the hinge, which may be due to the hydrophobic interactions of the CSD which would exclude and release associated water molecules into the solvent. The binding of TERRA

causing a change in conformation of HP1α may also exclude water and contribute to ΔS.

The results in this chapter reiterate the distinct selectivity of HP1α for the parallel G4 structure and show that a short helix within the CSD confers specificity through its interaction with the G-tetrad when binding through the charged hinge. The HP1α CSD has essential roles in localisation and heterochromatin function (Nielsen et al., 1999; Aagaard et al., 1999; Smothers and Henikoff, 2001; Quivy et al., 2004, 2008; Canzio et al., 2014), and the contribution of the CSD to nucleic acid specificity may be an important contributing factor to these functions.

Chapter 5

Small-angle X-ray scattering structural analysis of HP1α and TERRA

5.1 Introduction

Limited structural data for HP1α exist (Kumar and Kono, 2020), despite the importance of HP1α in heterochromatin propagation and maintenance, and its role in diverse chromosomal processes (Janssen et al., 2018; Strom et al., 2021). Atomic resolution structures of several HP1 globular CD and CSD units alone or in complex with H3K9me3 or PXVXL peptide have indeed been published (Brasher et al., 2000; Nielsen et al., 2002; Thiru et al., 2004; Huang et al., 2006; Kaustov et al., 2011; Kang et al., 2011; Ruan et al., 2012; Shimojo et al., 2016; Liu et al., 2017; Arora et al., 2019). However, no full-length HP1 structures have been resolved given the inability to crystallise such a dynamic and flexible protein (Canzio et al., 2014). In order to obtain structural information on HP1α alone and in complex with a G4, small-angle X-ray scattering (SAXS) can be used, with the advantage of studying the structure of these macromolecules in solution.

Raw scattering data is obtained by subjecting samples in solution to monochromatic X-rays, which are then scattered at small angles onto a detector to create a two-dimensional image

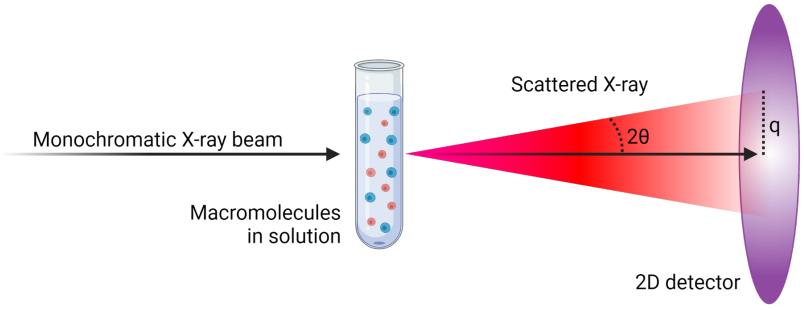

Figure 5.1: Schematic representation of small-angle X-ray scattering. Monochromatic X-rays are scattered by the sample in solution at small angles onto a detector to create a two-dimensional image. Image created with BioRender.

(**Figure 5.1**). This scattering image is then transformed into a one-dimensional SAXS profile of scattering intensity (I) versus the scattering vector/momentum transfer (q), which can be used to obtain biophysical parameters such as the radius of gyration (R_g), maximum dimension (D_{max}), and molecular weight (MW). SAXS data can then be converted to three-dimensional low-resolution molecular envelopes for fitting of structures. In addition to this, the recent release of artificial-intelligence (AI)-derived three-dimensional protein structure prediction by AlphaFold (Jumper et al., 2021; Evans et al., 2022) also aids in resolving the structures formed by HP1α.

Herein, structural data and models are obtained for both monomeric and dimeric HP1α, the hinge alone of HP1α, the NTE phosphorylation mimic, hinge charge mutant, and deletions of the NTE and CTE, along with parallel G4 TERRA and mutated TERRA. The structural information obtained on these molecules and complexes aid in gaining insight into the structure of HP1α and how these mutations and PTMs affect both the structure of HP1α and the binding to TERRA.

5.2 Results

5.2.1 Modelling of HP1α

To determine low-resolution structures of HP1α molecules, SAXS was utilised. These experiments were performed under reducing conditions due to the inter- and intra-molecular disulfide bonds that occur in HP1α molecules (**Figure A.41**, **Figure A.42**).

To initially investigate the structure of HP1α to determine its mode of binding to TERRA, SAXS was performed on the αHinge mutant of HP1α, previously shown to bind TERRA45 as well as other nucleic acids (**Section 4.2.1**). Firstly, raw two-dimensional scattering data were reduced to a one-dimensional curve and relevant frames of data are averaged and buffer subtracted using scatterBrain Analysis v2.82 (**Section 2.16**). At this point, the data were plotted on a log(I) versus log(q) graph using PRIMUS (Konarev et al., 2003). A downwards deviation at very low q would indicate intermolecular interactions (repulsion), whereas an upwards deviation would indicate aggregation of the sample (Jacques and Trewhella, 2010; Kikhney and Svergun, 2015). The αHinge SAXS data gave a relatively flat curve at low q indicating a monodisperse sample (**Figure 5.2a**).

Data are then presented using PRIMUS as I × q^2 vs q, known as a Kratky plot, which was used to assess the overall shape of the contents of the αHinge sample. Highly folded globular proteins or oligonucleotides will exhibit a bell-shaped peak that returns to baseline at high q values. When proteins are flexible or unfolded, the plot will not converge with the q axis at high q (Kikhney and Svergun, 2015). SAXS data of αHinge showed it was, as expected, highly flexible (**Figure 5.2b**).

The low q portion of the scattering curve, known as the Guinier region, is used to evaluate monodispersity of the sample, and is also used to measure the R_g which is the average of squared centre-of-mass distances in the molecule weighted by the scattering length density. The scattering intensity at zero angle (I(0)) can also be obtained, which is a measure of the overall size of the molecule. The Guinier plot (ln(I) vs q^2) of SAXS data obtained of the αHinge is shown in **Figure A.43**, which yielded an R_g of 23.54 Å and an I(0) of 0.017 cm^{-1} (**Table 5.1**) using the Guinier distribution analysis in PRIMUS.

An indirect inverse Fourier transformation of scattering data was then performed using PRIMUS, which results in the pairwise distribution function ($P(r)$) curve. The $P(r)$ curve represents a distribution of all of the interatomic distances in the sample. Therefore, the D_{max} of the molecule measured end-to-end can be determined from the $P(r)$ curve as it descends to 0 smoothly along the X-axis. $P(r)$ analysis of the αHinge (**Figure A.44**) showed an R_g value of 25.1 Å, relatively in agreement with the Guinier analysis, and a D_{max} of 94.4 Å (**Table 5.1**).

The $P(r)$ analysis also yields the Porod volume, which is the volume of the scattering particle in $Å^3$. The Porod volume can then be used to estimate molecular weight by dividing this value by

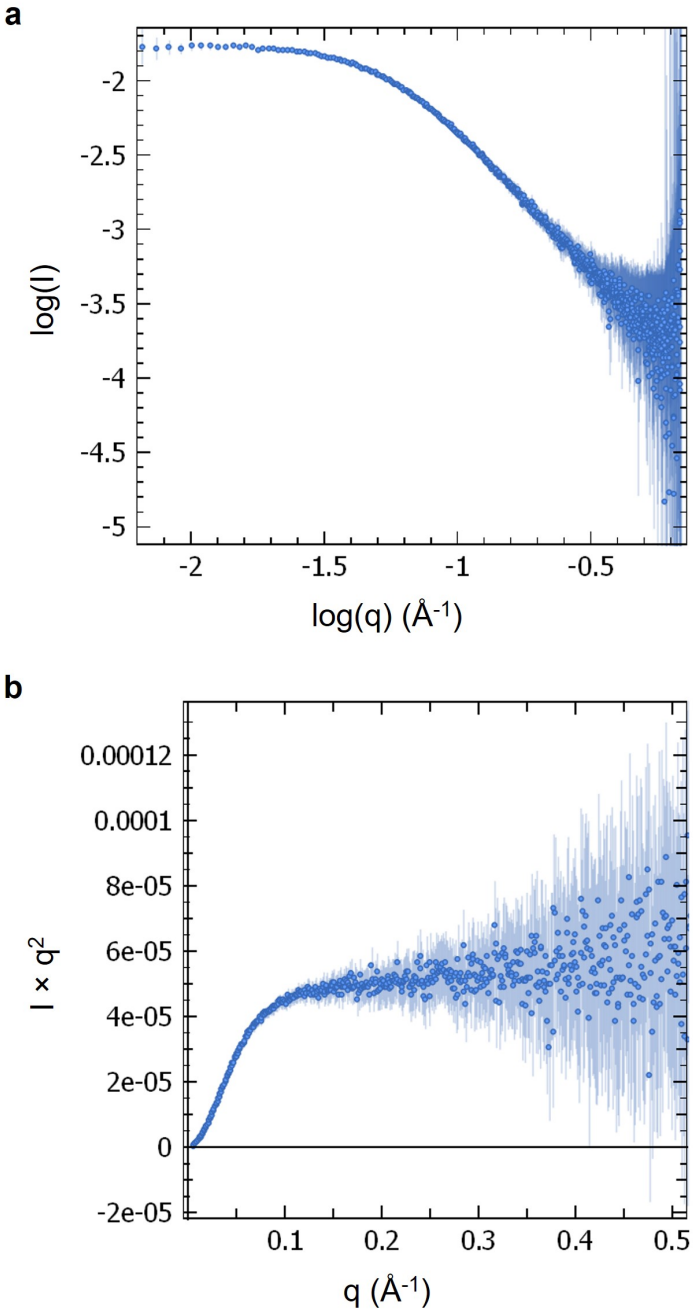

Figure 5.2: SAXS curves of αHinge. Static mode SAXS profile of αHinge presented as a) log of I versus log of q, and b) as a Kratky plot. Experiment conducted using 2.5 mg/mL of αHinge in IB1 plus 10 mM TCEP at 22°C.

Table 5.1: SAXS structural parameters of αHinge

Structural parameter	**αHinge**
Guinier analysis	
R_g (Å)	23.5 ±0.3
$q{\cdot}R_g$ range	0.28-1.29
I(0) (cm^{-1})	0.017 ±0.00012
Fidelity	1.00
P(r) analysis	
R_g (Å)	25.1
D_{max} (Å)	94.4
q range	0.0127-0.3397
Quality estimate	0.65
Porod volume ($Å^3$)	13,021
Porod volume ×0.6 = MW (Da)	7,813
Expected MW (Da)	8,241
Ratio to expected MW	0.95

1.66 $Å^3{\cdot}g{\cdot}cm^{-3}{\cdot}Da^{-1}$ (based on average protein density), or inversely multiplying by 0.6 (Rambo and Tainer, 2011; Barzak et al., 2021). The Porod volume of the αHinge converted to molecular weight was 7,813 Da (**Table 5.1**), which is similar to the molecular weight of one unit of αHinge (8,241 Da); therefore, the αHinge is assumed to be monomeric.

To create a three-dimensional *ab initio* shape from the one-dimensional αHinge SAXS data, DAMMIN bead models were used. Starting from a random seed within a sphere determined by the D_{max}, 10 models were created by DAMMIN to fit the data and aligned (an example is shown in **Figure A.45**). Given the monomeric nature of the αHinge, $P1$ symmetry was applied to this modelling. The 10 models of αHinge superimposed are shown in **Figure 5.3**, which shows an elongated kinked shape much like what would be expected from a short peptide. DAMMIN gives a χ^2 value as a measure of the differences between the experimental scattering curve and the scattering curves generated by the atomic models, with lower values indicating a better fit. The final χ^2 against the raw data of the αHinge models is 0.1204. DAMMIN also calculates the normalised spatial discrepancy (NSD) which is a measure of the agreement between each of the three-dimensional models, with a value less than 1 indicating similarity (Kozin and Svergun, 2001). The mean value of the NSD for the αHinge models is 0.987, indicating sufficient agreement or self-consistency among the generated models.

To determine how the αHinge polypeptide might fit into this molecular envelope, AlphaFold

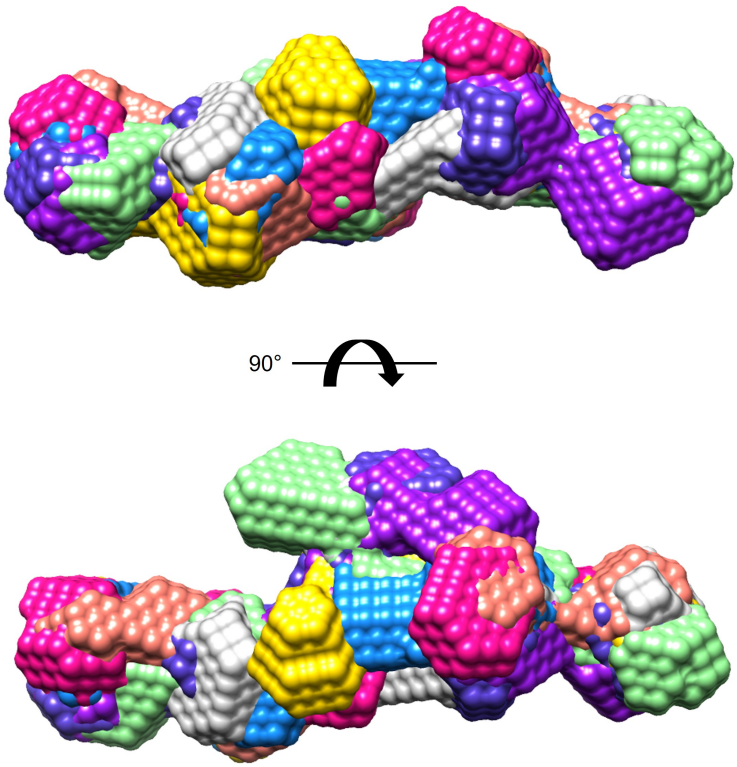

Figure 5.3: Computed models of αHinge SAXS data. DAMMIN-generated models of αHinge SAXS data based on $P1$ symmetry. Each of 10 models is shown in a different colour. Ensemble resolution 25 Å, final χ^2 against raw data 0.1204, mean value of NSD 0.987. The lower image is rotated forward 90°.

(Jumper et al., 2021) was utilised to predict an atomic resolution structure based on the amino acid sequence of the αHinge, which showed a chain mostly devoid of secondary structure apart from a short α-helix. The AlphaFold model was then manually aligned with the DAMMIN model (**Figure 5.4**) which visually appears to be a good fit.

To empirically assess the fit of the AlphaFold model to the SAXS data collected, Fast X-Ray Scattering (FoXS) was used to computationally generate a SAXS profile of the PDB file and fit it to the experimental profile. Two parameters control the fit of the profile, which are c_1 and c_2. The parameter c_1 controls the excluded volume of the molecule with a default value of 1, which increases or decreases based on scaling of the atomic radius of the molecule. Parameter c_2 controls the density of the hydration layer around the molecule, with a default value of 0, reaching up to 4 where 4 water molecules may surround 1 solute atom. Both c_1 and c_2 are adjusted to give the best fit to the experimental data, resulting in the lowest possible χ^2 value. The fit of the AlphaFold-generated PDB file to the αHinge SAXS data yielded close-to-optimal c_1 and c_2 values of 0.99 and -0.11, respectively, and a χ^2 value of 2.30 (**Figure 5.5**), indicating that the structure may indeed be representative of the SAXS data collected. It should be noted that SAXS-generated molecular envelopes could be representing multiple extended conformations of the protein (particularly those which are intrinsically disordered), whereas an AlphaFold model represents a single static structure. Therefore, the Kratky plots generated from a single conformation of an AlphaFold model can be misleading.

Circular dichroism spectroscopy was performed on the αHinge mutant, and showed a strong negative ellipticity at 200 nm (**Figure 5.6**) that corresponds to random coils (Venyaminov et al., 1993), further indicating that the AlphaFold-generated structure of the hinge reflects experimental data. The circular dichroism analysis and plotting tool (CAPITO, Wiedemann et al. 2013) estimates secondary structure content based upon a circular dichroism spectrum. When the circular dichroism spectrum of the αHinge was analysed, CAPITO predicted helix structure at a proportion of 0.01, β-strand structure at a proportion of 0.19, and irregular structure at a proportion of 0.72. Therefore, the αHinge is likely primarily disordered, with low likelihood of helix formation based upon the circular dichroism data.

To further explore the structure of full-length HP1α, the monomeric form of HP1α is investigated. As discussed in **Section 4.2.3**, mutation of a tyrosine at position 168 in an α-helix of the CSD to a glutamic acid residue (Y168E) results in HP1α being reduced to its monomeric

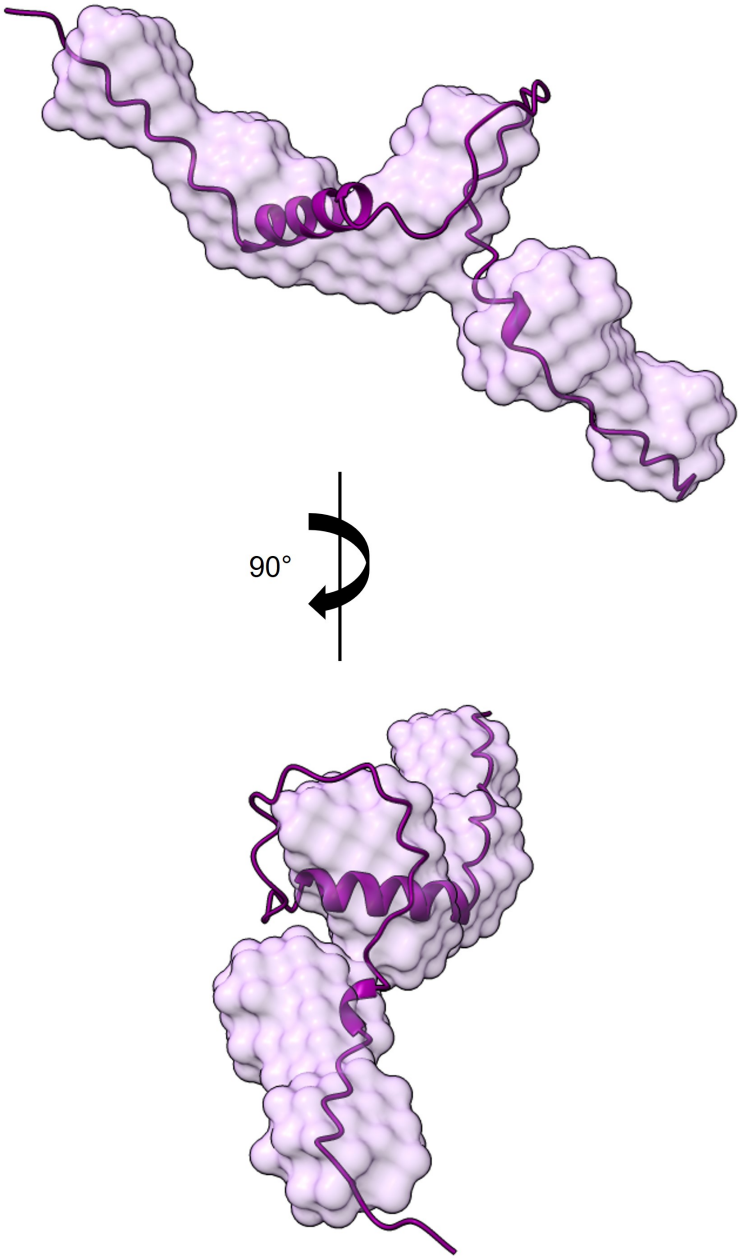

Figure 5.4: AlphaFold model overlay with DAMMIN-generated αHinge model. AlphaFold-generated model of the αHinge polypeptide sequence (purple) with a DAMMIN-generated model of αHinge (pink) from SAXS data. The N-terminus is on the left, and the C-terminus on the right. The lower image is rotated left 90°.

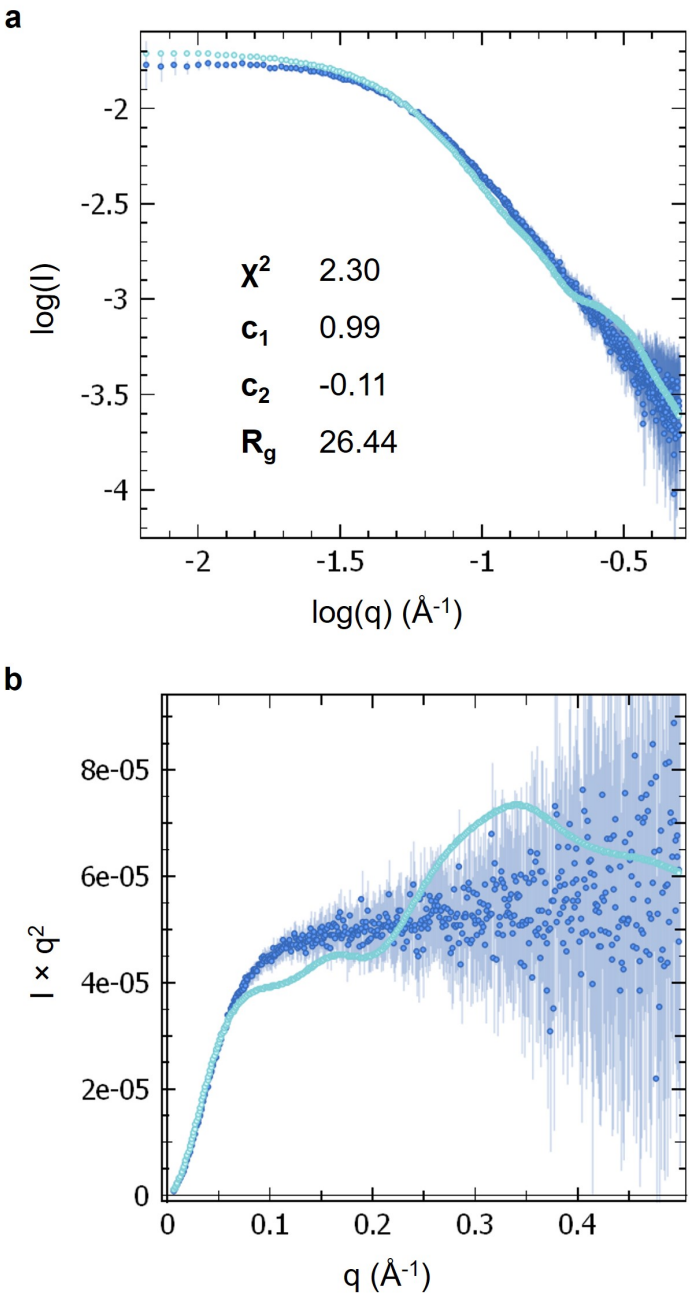

Figure 5.5: Fit of FoXS-generated SAXS profile of the αHinge AlphaFold polypeptide to αHinge SAXS data. FoXS-generated SAXS curve of the αHinge AlphaFold polypeptide shown in light blue, and SAXS data of αHinge shown in dark blue. The data are presented as a) log of I versus log of q, and b) as a Kratky plot.

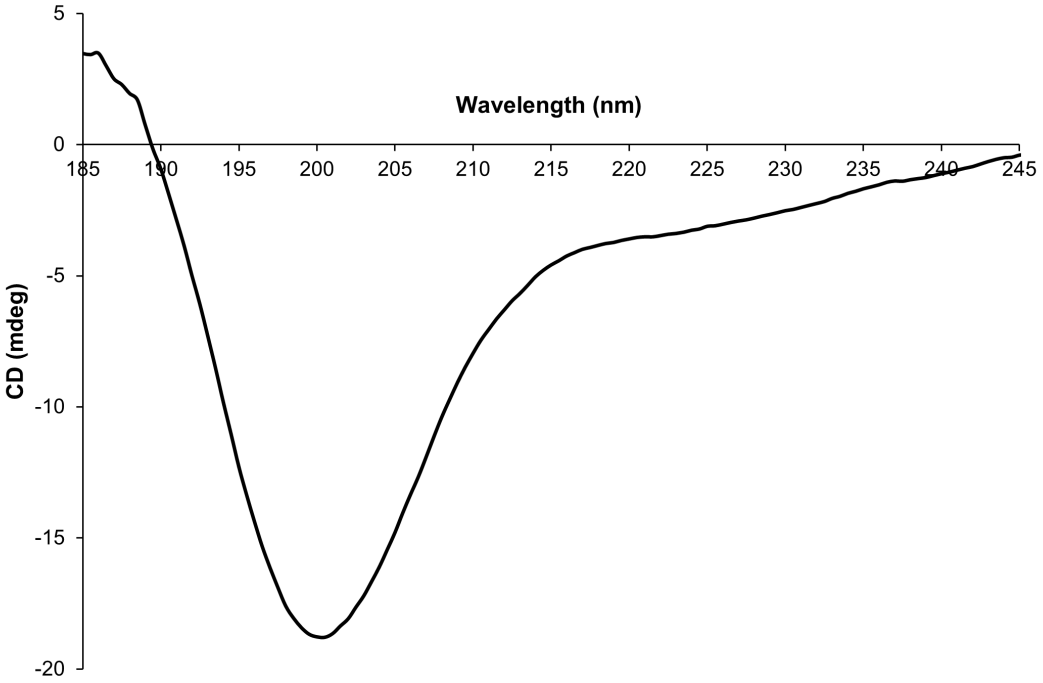

Figure 5.6: Circular dichroism spectrum of αHinge. Far UV circular dichroism spectrum of 1.5 mg/mL αHinge in IB1 in a 0.1 mm cuvette at room temperature.

form. Therefore, SAXS was performed on the HP1α Y168E mutant. For this larger protein, to ensure a monodisperse sample is measured, SAXS was used in combination with size-exclusion chromatography (SEC-SAXS) to analyse HP1α Y168E (**Figure A.46**). The scattering curves yielded monodisperse flexible shapes (**Figure 5.7**), and the Guinier analysis showed that HP1α Y168E had an R_g of 31.98 Å (**Figure A.47**), which shows some evidence of aggregation at very low q. $P(r)$ analysis showed a similar R_g of 30.57 Å and a D_{max} of 91.12 Å (**Table 5.2**, **Figure A.48**). Conversion of the Porod volume to molecular weight indicated that the HP1α Y168E mutant was indeed monomeric at 22,345.86 Da, compared to the expected molecular weight of 24,978.9 Da. A three-dimensional model of monomeric HP1α Y168E was generated using DAMMIN with $P1$ symmetry, and the 10 models are shown superimposed in **Figure 5.8**, showing a relatively flat shape with an additional protruding part.

Given the lack of experimentally derived structures of full-length HP1α, the AlphaFold database was used to give a predicted atomic resolution structure to fit the DAMMIN-generated model. AlphaFold model AF-Q61686-F1 was fitted onto one of the *ab initio* models manually (**Figure 5.9**). The globular CD and CSD were assumed to be within the major part of the model. The intrinsically disordered hinge (loop between the CD and CSD) also shares this space, and the

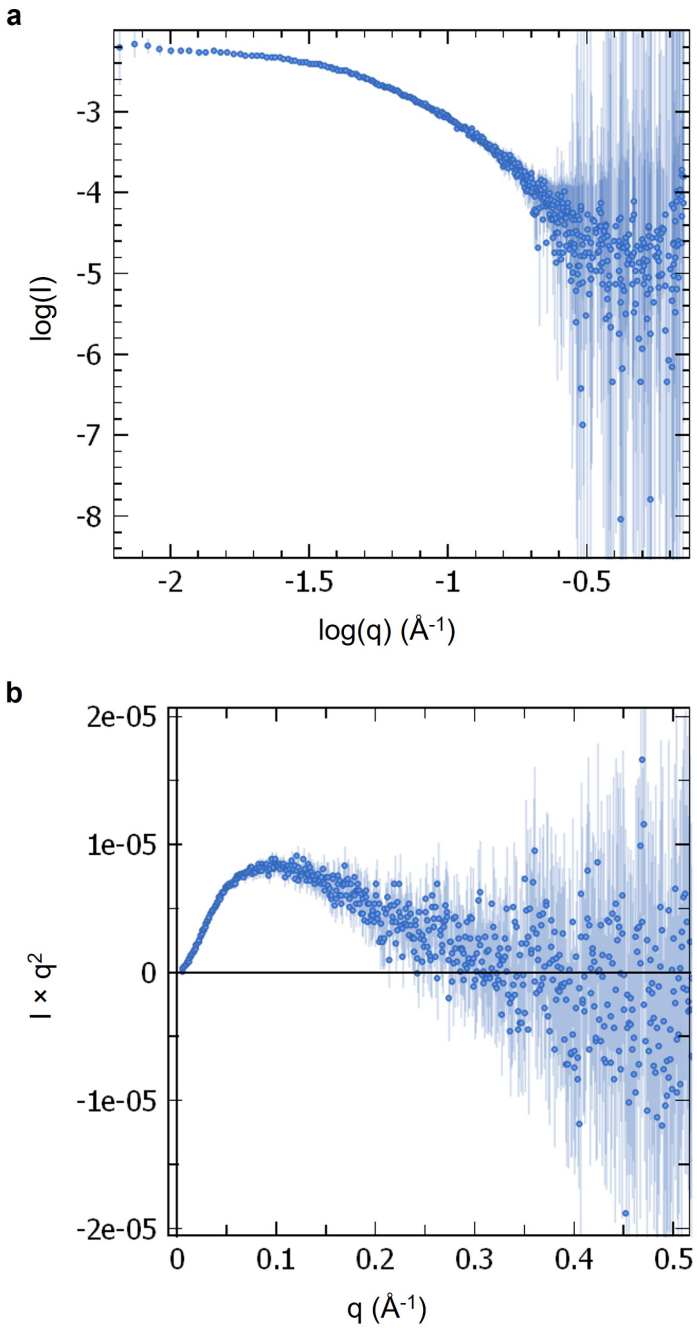

Figure 5.7: SAXS curves of SEC-separated HP1α Y168E. SAXS profile of SEC-separated HP1α Y168E presented as a) log of I versus log of q, and b) as a Kratky plot. Experiment conducted using 5 mg/mL of HP1α Y168E in IB1 plus 10 mM TCEP at 22°C.

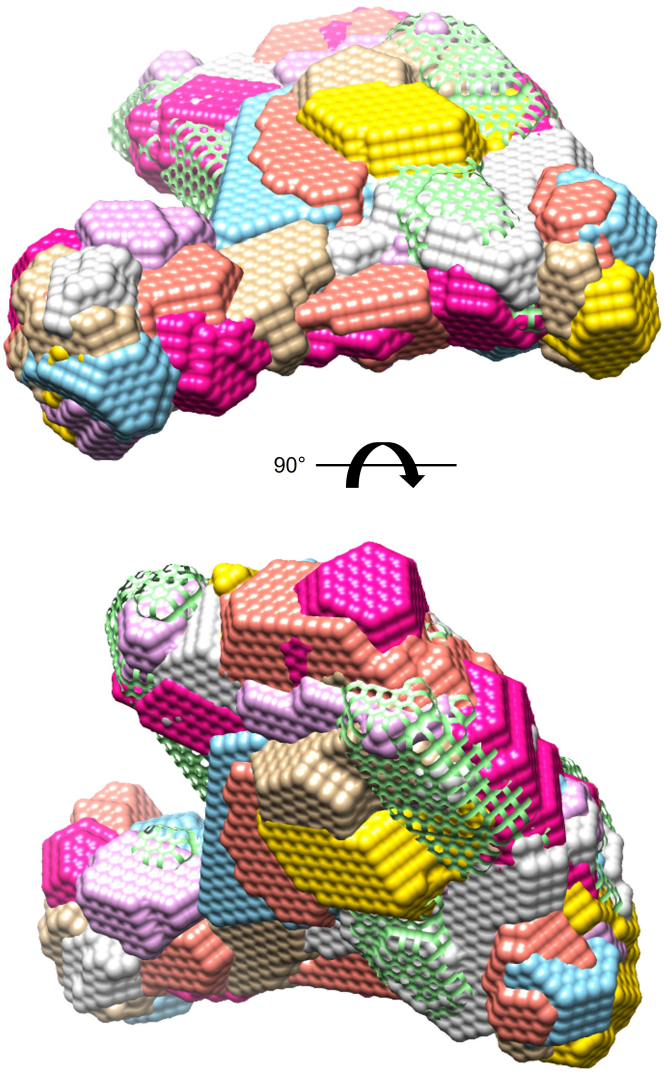

Figure 5.8: Computed models of HP1α Y168E SAXS data. DAMMIN-generated models of monomeric HP1α Y168E SAXS data based on $P1$ symmetry. Each of 10 models is shown in a different colour. Ensemble resolution 25 Å, final χ^2 against raw data 0.6717, mean value of NSD 1.223. The lower image is rotated forwards 90°.

NTE (unstructured tail) may occupy the protruding section from the shape. The CTE (shown here as an α-helical structure) does not fit within the shape, but due to the flexibility between the CSD and the CTE, may fold inwards. The fit also allows the known α-helix dimerisation interface of the CSD (Brasher et al., 2000) to be exposed on the exterior of the model. It should be noted that the AlphaFold model of monomeric HP1α has high confidence in the structures of the globular CD and CSD, but low and very low confidence for the disordered NTE, hinge, and CTE (**Figure A.38**). In addition to this, the model shows an apparent interface between hydrophobic residues in the CD and CSD. When analysed by PDBePISA (Proteins, Interfaces, Structures and Assemblies, Krissinel and Henrick 2007), this shows that an interface forms with a ΔG of -3.6 kcal/mol, indicating that this interface binding is weakly favourable.

To assess the fit of the AlphaFold model to the SAXS data collected on monomeric HP1α, FoXS was used to generate a SAXS profile of AF-Q61686-F1 and fit it to the HP1α Y168E experimental profile. The fit is shown in **Figure 5.10**, which gives a χ^2 value of 1.41 with c_1 and c_2 values well within acceptable limits, which indicates that again the experimental data may indeed be representative of the predicted AlphaFold structure of monomeric HP1α, even though the fit to the calculated envelope and Kratky plot is not ideal.

To further examine HP1α, SEC-SAXS was also performed on the WT protein (**Figure A.49**). The slight downwards deviation at the beginning of the plot shown in **Figure 5.11a** indicates disaggregation/intermolecular interactions between HP1α molecules, which could be due to the high concentration of the solution (5 mg/mL). The overall shape of the contents of the HP1α sample were determined by observation of the Kratky plot, which showed that WT HP1α was highly flexible (**Figure 5.11b**), even more so than the Y168E mutant shown by the lower $I \times q^2$ maximum. Previous circular dichroism spectra showed the presence of α-helices and β-sheets in the structure of HP1α (**Figure 3.7b**). Therefore, the structured CD and CSD are likely large contributors to the partial structure shown in the Kratky plot, and the partial unstructured nature of the molecule as shown by the q^2-weighted intensity not returning to 0 at higher q values may be attributed to the flexible linker hinge between the globular CD and CSD, and also flexibility in the dimer association.

The Guinier plot of SAXS data obtained of WT HP1α is shown in **Figure A.50**, which yielded an R_g of 34.9 Å and an I(0) of 0.004 cm^{-1} (**Table 5.2**). $P(r)$ analysis of WT HP1α (**Figure A.51**) showed an R_g value of 37.8 Å, relatively in agreement with the Guinier analysis, and a

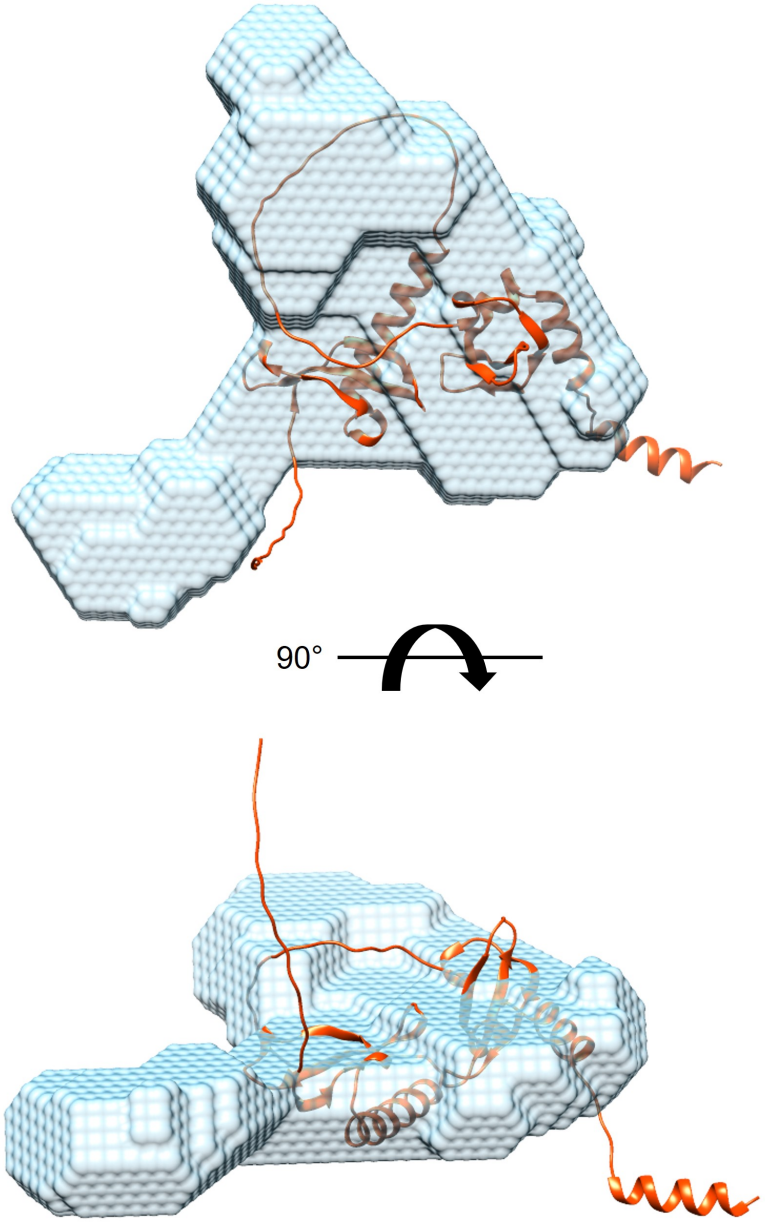

Figure 5.9: AlphaFold model overlay with DAMMIN-generated HP1α monomer model. AlphaFold model AF-Q61686-F1 (red) with a DAMMIN-generated model of HP1α Y168E (blue) from SAXS data. The N-terminus of AF-Q61686-F1 is on the left, and the C-terminus on the right. The lower image is rotated backwards 90°.

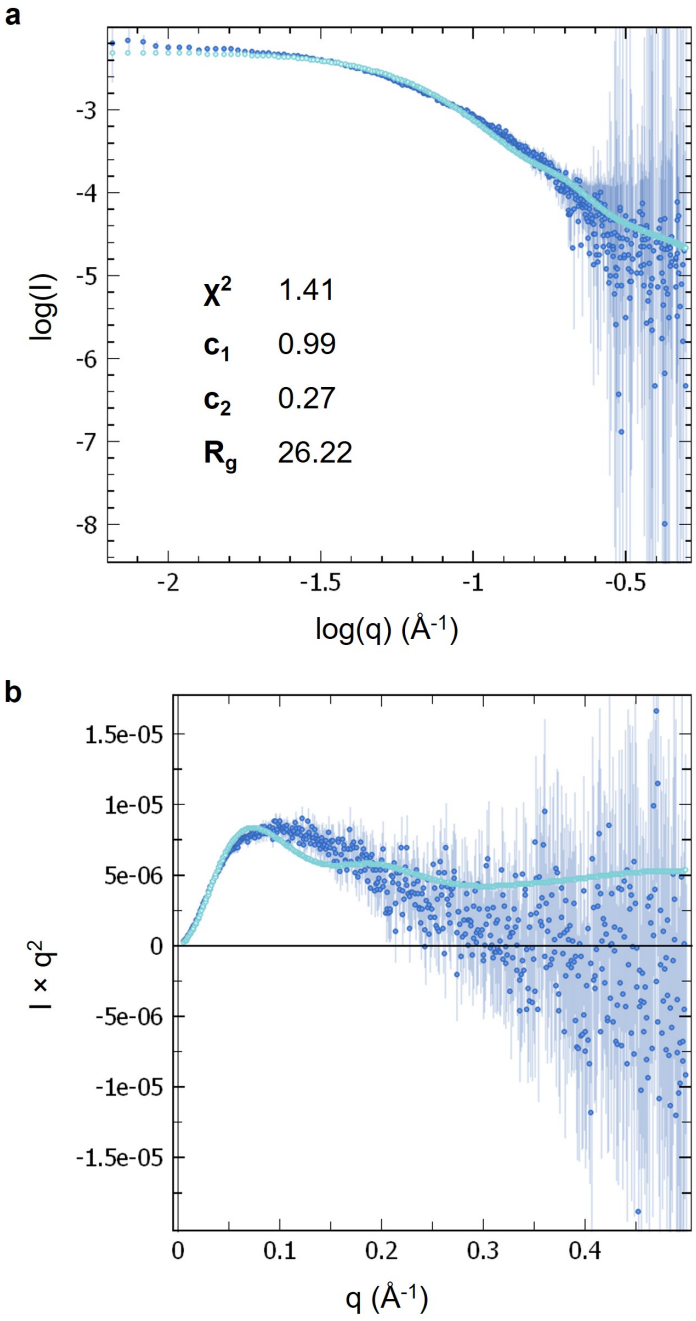

Figure 5.10: Fit of FoXS-generated SAXS profile of AF-Q61686-F1 to HP1α Y168E SEC-SAXS data. FoXS-generated SAXS curve of AF-Q61686-F1 shown in light blue, and SEC-SAXS data of HP1α Y168E shown in dark blue. The data are presented as a) log of I versus log of q, and b) as a Kratky plot.

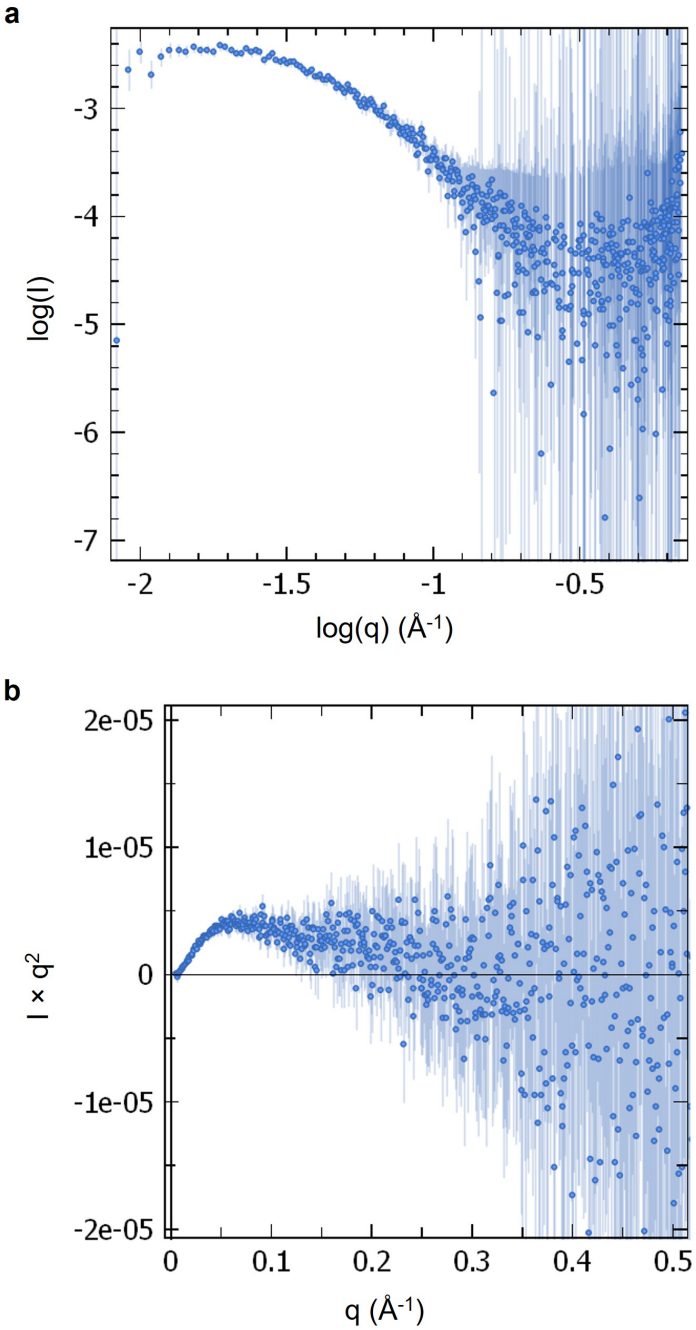

Figure 5.11: SAXS curves of SEC-separated HP1α WT. SAXS profile of SEC-separated HP1α WT presented as a) log of I versus log of q, and b) as a Kratky plot. Experiment conducted using 5 mg/mL of HP1α in IB1 plus 10 mM TCEP at 22°C.

Table 5.2: SEC-SAXS structural parameters of HP1α WT and Y168E

Structural parameter	**HP1α WT**	**HP1α Y168E**
Guinier analysis		
R_g (Å)	34.9 ±1.9	32.0 ±0.6
$q{\cdot}R_g$ range	0.44-1.30	0.41-1.30
I(0) (cm^{-1})	0.004 ±0.00014	0.0056 ±0.000065
Fidelity	0.98	1.00
P(r) analysis		
R_g (Å)	37.8	30.6
D_{max} (Å)	116.7	91
q range	0.0171-0.3004	0.0136-0.3004
Quality estimate	0.74	0.63
Porod volume ($Å^3$)	97,646	37,243
Porod volume ×0.6 = MW (Da)	58,587	22,346
Expected MW (Da)	50,026 (dimer)	24,979
Ratio to expected MW	1.17	0.89

D_{max} of 116.68 Å (**Table 5.2**). The Porod volume of HP1α WT converted to molecular weight was 58,587.3 Da (**Table 5.2**), which is similar to the molecular weight of two units of HP1α (25,012.96 Da ×2 = 50,025.92 Da). Therefore, WT HP1α is assumed to be dimeric.

DAMMIN was used to create a three-dimensional *ab initio* shape from the one-dimensional HP1α SAXS data, and given the dimeric nature of HP1α, $P2$ symmetry was applied to this modelling. The 10 models of HP1α superimposed are shown in **Figure 5.12**, which again shows a relatively flat oval shape with two thin protrusions from opposite sides.

To determine how the AlphaFold model is arranged in dimeric HP1α, the monomeric Y168E model from **Figure 5.9** was fitted onto the model of WT HP1α (**Figure 5.13**). This arrangement allowed for the dimerisation of two HP1α molecules through the previously determined dimerisation interface, in which a pair of α-helices of the CSDs interact such that the hinge region of each monomer faces in opposite directions and is in proximity to the CTE of the neighbouring molecule. However, the crystal structure of the CSD shows dimerisation in a *cis* conformation (PDB: 3I3C), whereas here (in the solution state) it is shown in a *trans* conformation. AlphaFold mulitimer (Evans et al., 2022) was used to predict the structure of two chains of HP1α (**Figure 5.14**), and indeed showed dimerisation in the *cis* conformation as per the crystal structure of the CSD.

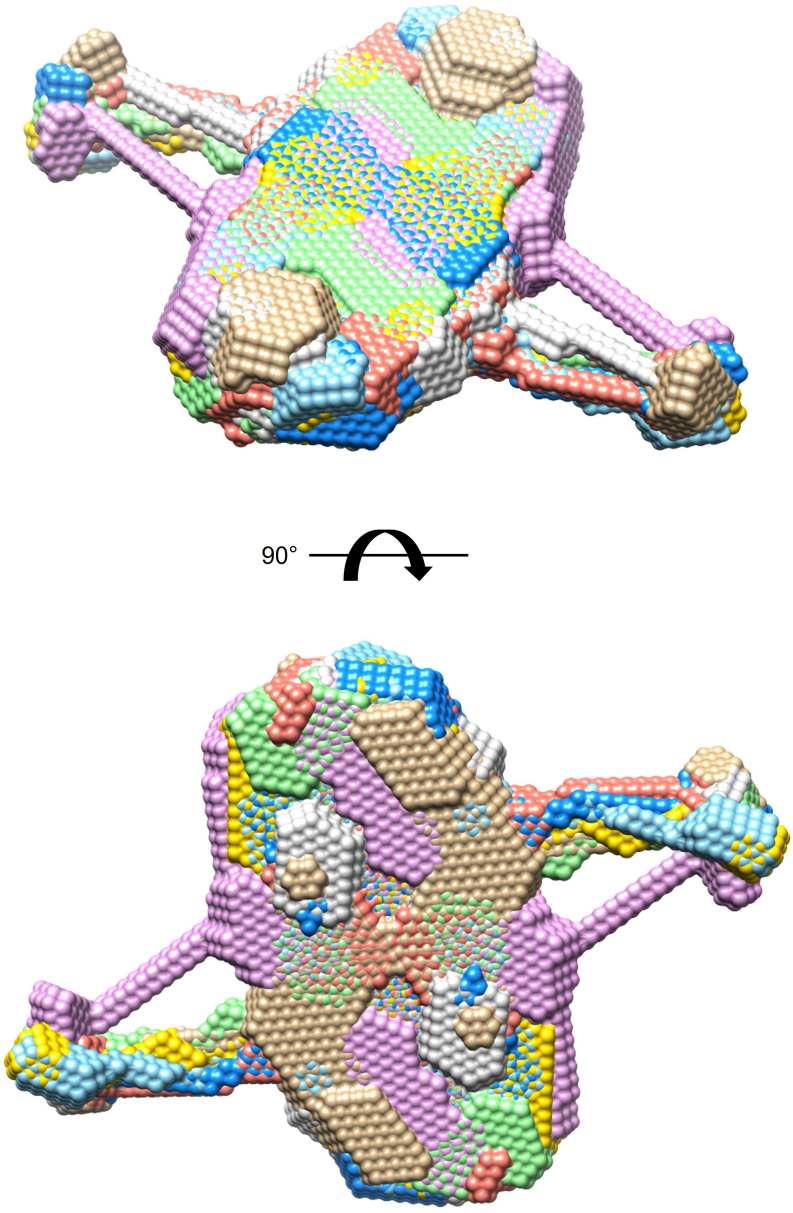

Figure 5.12: Computed models of HP1α WT SEC-SAXS data. DAMMIN-generated models of dimeric HP1α WT SAXS data based on $P2$ symmetry. Each of 10 models is shown in a different colour. Ensemble resolution 25 Å, final χ^2 against raw data 0.2833, mean value of NSD 1.137. The lower image is rotated backwards 90°.

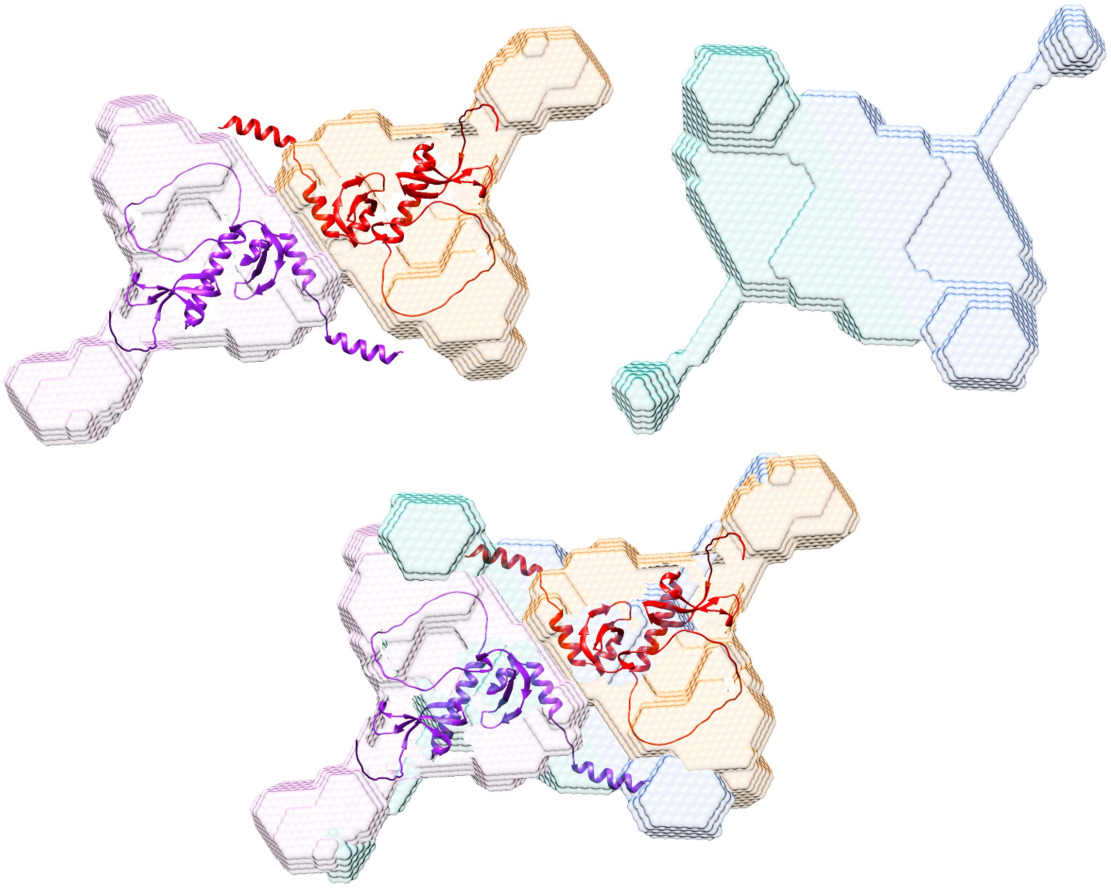

Figure 5.13: HP1α monomer fit onto DAMMIN-generated dimer model. The monomer fit from **Figure 5.9** is shown in duplicate here (with one chain in purple and another in orange) in *trans* conformation and overlaid onto a DAMMIN-generated model of dimeric HP1α WT (shown in green/blue). Note: the AlphaFold predicted structure of HP1α (AF-Q61686-F1) does not include the six histidine residues and TEV site on the N-terminus.

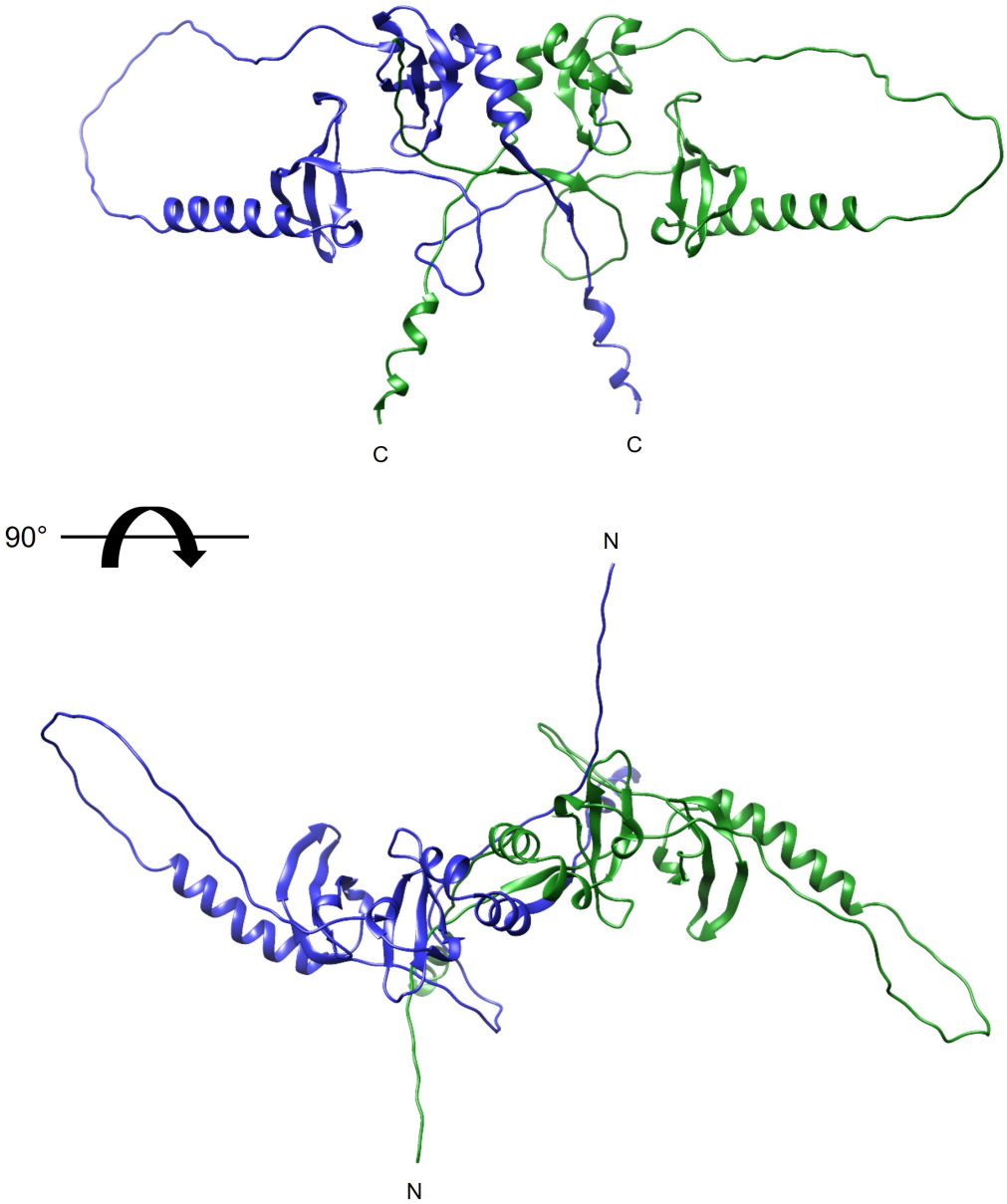

Figure 5.14: AlphaFold AI-generated dimer model of his-HP1α. AlphaFold-Multimer model of two his-HP1α molecules in *cis* conformation, shown in blue and green. The N-termini are unstructured, while the C-termini fold into α-helices. The lower image is rotated forwards 90°.

Table 5.3: SAXS structural parameters of HP1α WT and Y168E measured in static mode

Structural parameter	**HP1α WT**	**HP1α Y168E**
Guinier analysis		
R_g (Å)	47.1 ±0.7	33.9 ±0.3
$q{\cdot}R_g$ range	0.41-1.10	0.40-1.05
I(0) (cm^{-1})	0.096 ±0.00069	0.08 ±0.00036
Fidelity	1.00	1.00
P(r) analysis		
R_g (Å)	47.6	35.9
D_{max} (Å)	185.3	145.9
q range	0.0086-0.3001	0.0127-0.3004
Quality estimate	0.66	0.61
Porod volume ($Å^3$)	109,594	31,342
Porod volume ×0.6 = MW (Da)	65,756	18,805
Expected MW (Da)	50,026 (dimer)	24,979
Ratio to expected MW	1.31	0.75

FoXS was used to compare the fit of the *cis* and *trans* dimer conformations to the experimental data of WT HP1α (**Figure 5.15**), which surprisingly showed equal χ^2 values of 0.35. However, the slight differences in R_g contributed to different c_2 values, with the *trans* conformation displaying a higher value. Therefore, the dimerisation conformation is yet to be explicitly determined; however, the *trans* model brings together the hinge and CTE of neighbouring molecules, which have been shown to interact by mass spectrometry (Larson et al., 2017).

When studying the complex of HP1α and a G4, SAXS was performed in static mode (without SEC) due to difficulty analysing the complex when being separated by SEC, likely due to the fast association and dissociation rates of the interaction (**Table 3.1**), leading to separation of components when travelling through the column. Therefore, HP1α WT and Y168E were also studied in static mode in order to be comparable to complex static mode SAXS experiments. The SAXS data of both HP1α WT and Y168E appeared to have similarly flexible domains as evidenced in very similar Kratky plots; however, the double log plot of the WT data showed some aggregation (**Figure 5.16**, **Figure 5.17**). The R_g values given by both Guinier and $P(r)$ analysis of HP1α Y168E were similar, and only slightly larger than values given by SEC-SAXS; however, the D_{max} was markedly increased from 91.1 to 145.9 Å in the static sample (**Figure A.52**, **Figure A.53**, **Table 5.3**). Both the R_g and D_{max} of HP1α were larger in the static sample (**Figure A.54**, **Figure A.55**, **Table 5.3**), likely due to aggregation.

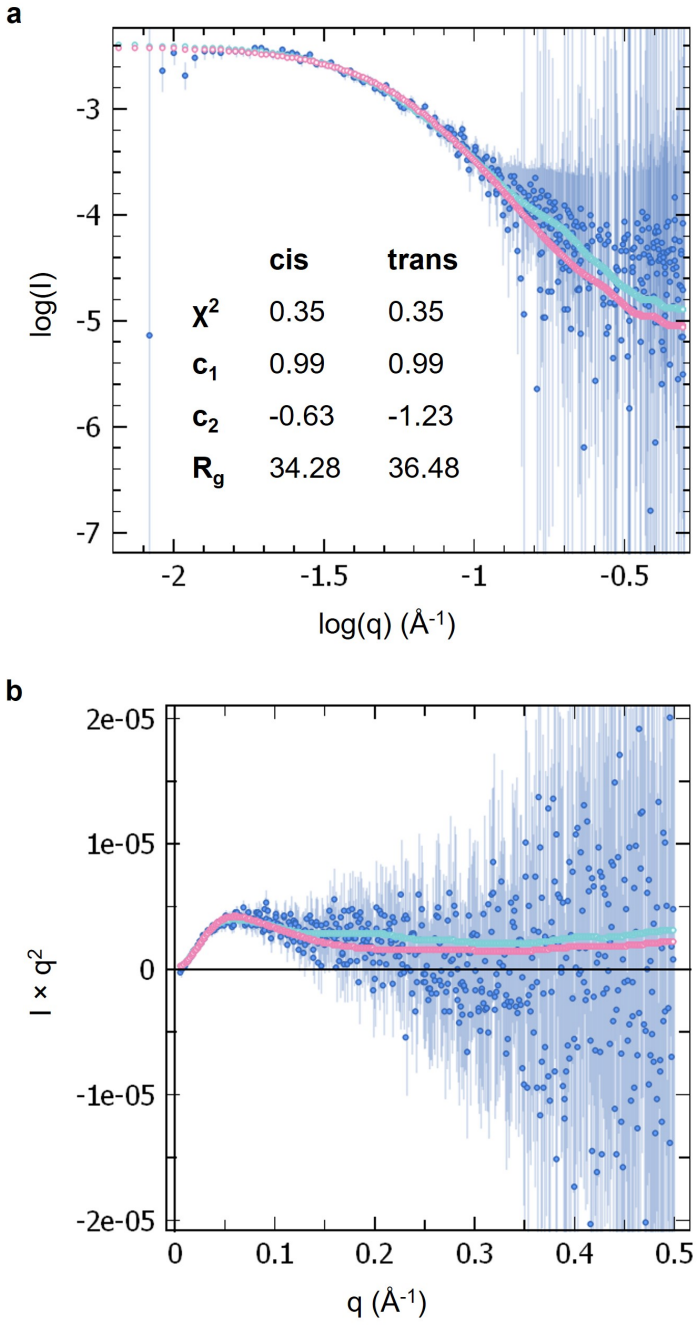

Figure 5.15: Fit of FoXS-generated SAXS profiles of HP1α dimer conformations to HP1α WT SEC-SAXS data. FoXS-generated SAXS curves of cis-HP1α dimer (pink) shown in **Figure 5.14**, and trans-HP1α dimer (light blue) shown in **Figure 5.13** to HP1α WT SEC-SAXS data (dark blue). The data are presented as a) log of I versus log of q, and b) as a Kratky plot.

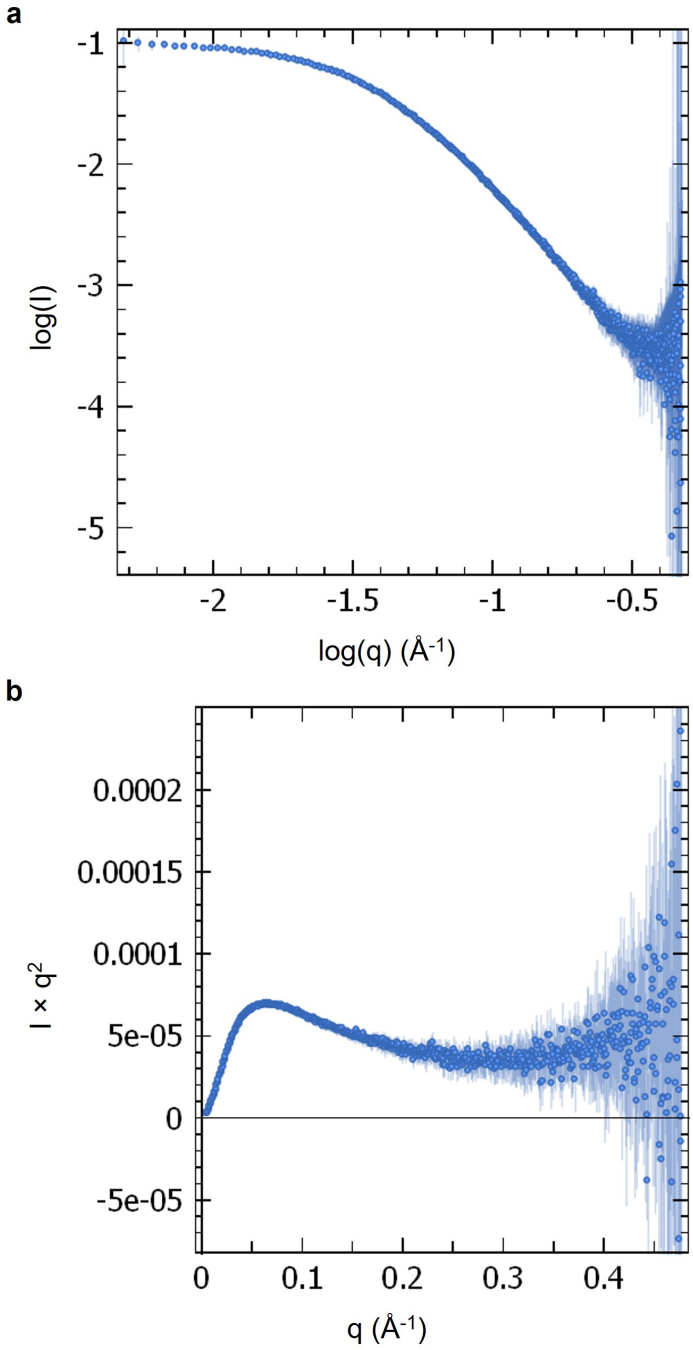

Figure 5.16: SAXS curves of HP1α WT. SAXS profiles of HP1α WT measured in static mode presented as a) log of I versus log of q, and b) as a Kratky plot. Experiment conducted using 2.5 mg/mL of HP1α in IB1 plus 10 mM TCEP at 22°C.

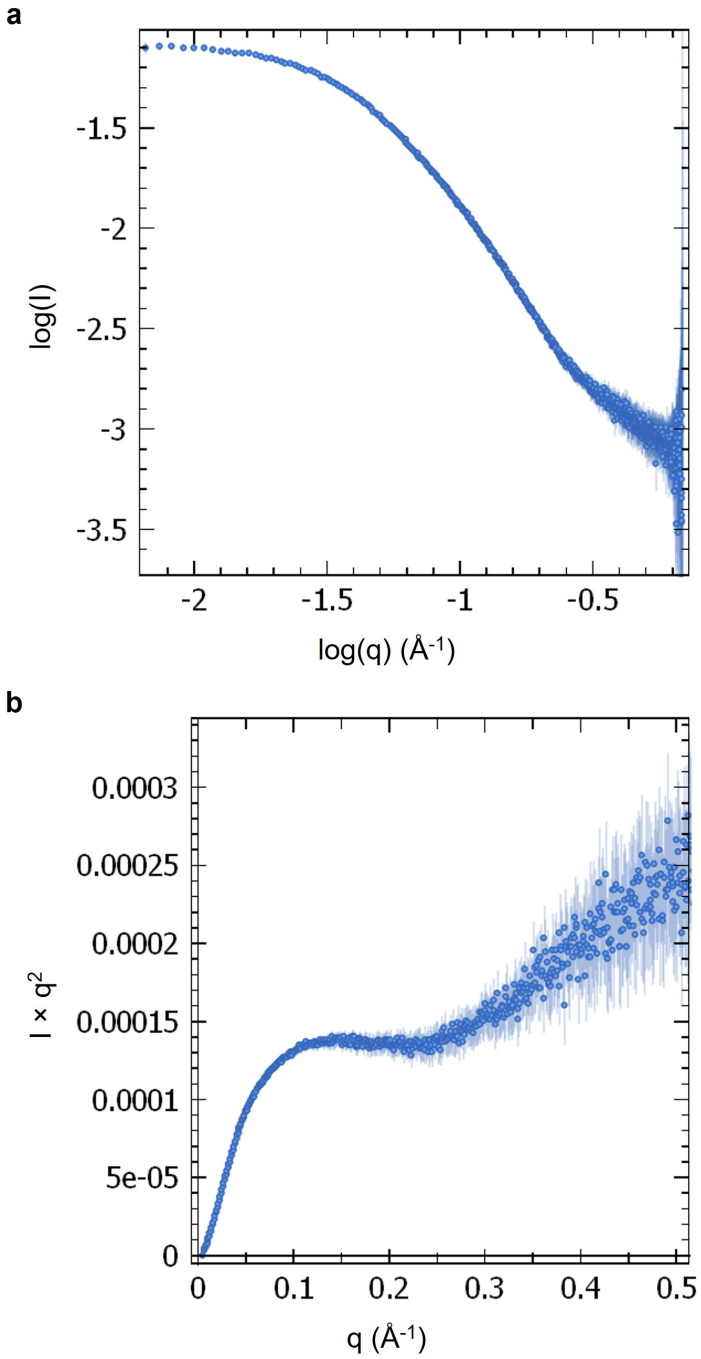

Figure 5.17: SAXS curves of HP1α Y168E. SAXS profiles of HP1α Y168E measured in static mode presented as a) log of I versus log of q, and b) as a Kratky plot. Experiment conducted using 2.5 mg/mL of HP1α Y168E in IB1 plus 10 mM TCEP at 22°C.

While the HP1α Y168E static SAXS data did not have a good fit to the monomeric AlphaFold model (χ^2 value of 29.25, **Figure A.56**), a comparison of the fit of the dimerisation models to the static WT HP1α data showed a better fit with the *trans* model over the *cis* model with χ^2 values of 4.98 and 13.27, respectively (**Figure 5.18**).

5.2.1.1 HP1α mutant SAXS analysis

To analyse how mutations of HP1α affect the structure of the protein and in turn its binding to TERRA, SAXS was performed on mutants of HP1α characterised in previous chapters. Scrutinised are the S11-14E mutant, whereby four serine residues in the NTE are mutated to glutamic acid residues to mimic N-terminal phosphorylation that increases specificity for H3K9me3; the ΔNTE and ΔCTE mutants, where either the disordered NTE or CTE has been deleted; and the A2;3 (89-91KRK>AAA;104-106KKK>AAA) mutant, where two of the positively charged patches in the hinge are mutated to uncharged alanine residues, resulting in decreased nucleic acid binding ability. SAXS curves of HP1α WT, S11-14E, ΔNTE, ΔCTE, and A2;3 are shown in **Figure 5.19**, where the Kratky plot shows that each of the mutants have flexibly linked domains. Guinier analysis showed that S11-14E and ΔCTE had similar R_g values to WT, whereas ΔNTE and A2;3 had slightly smaller R_g values of 43.2 Å and 42.7 Å respectively (**Table 5.4**, **Figure A.57**, **Figure A.59**, **Figure A.61**, **Figure A.63**). $P(r)$ analysis showed that S11-14E had a substantially larger D_{max} than WT of 206.7 Å, and ΔNTE had a smaller D_{max} of 169.5 Å (**Table 5.4**, **Figure A.58**, **Figure A.60**, **Figure A.62**, **Figure A.64**).

The larger D_{max} of the HP1α phosphorylation mimic S11-14E indicates a potentially different structure to WT HP1α. Therefore, to assess the structure of HP1α S11-14E, *ab initio* shape determination using DAMMIN was performed. The molecular envelope of S11-14E was indeed more extended than WT HP1α (**Figure 5.20**), which may contribute to the decrease in binding shown by this mutant to TERRA45 (**Figure 3.21**, **Table 3.3**).

5.2.2 Modelling of TERRA

To elucidate structural information of TERRA45, SEC-SAXS was performed (**Figure A.65**). SAXS profiles indicated that TERRA45 formed a well-ordered structure shown by the return to baseline at high q values in the Kratky plot (**Figure 5.21**). Guinier analysis showed that

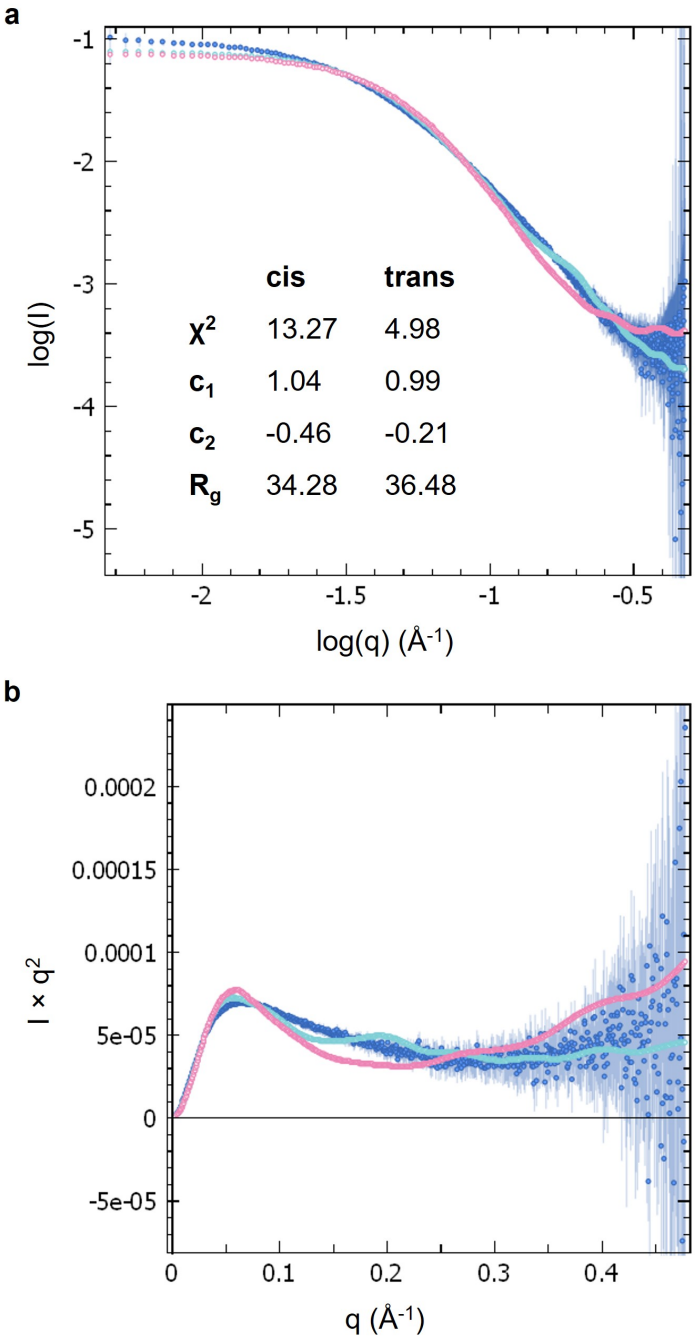

Figure 5.18: Fit of FoXS-generated SAXS profiles of HP1α dimer conformations to HP1α WT static SAXS data. FoXS-generated SAXS curves of cis-HP1α dimer (pink) and trans-HP1α dimer (light blue) to HP1α WT static SAXS data (dark blue). The data are presented as a) log of I versus log of q, and b) as a Kratky plot.

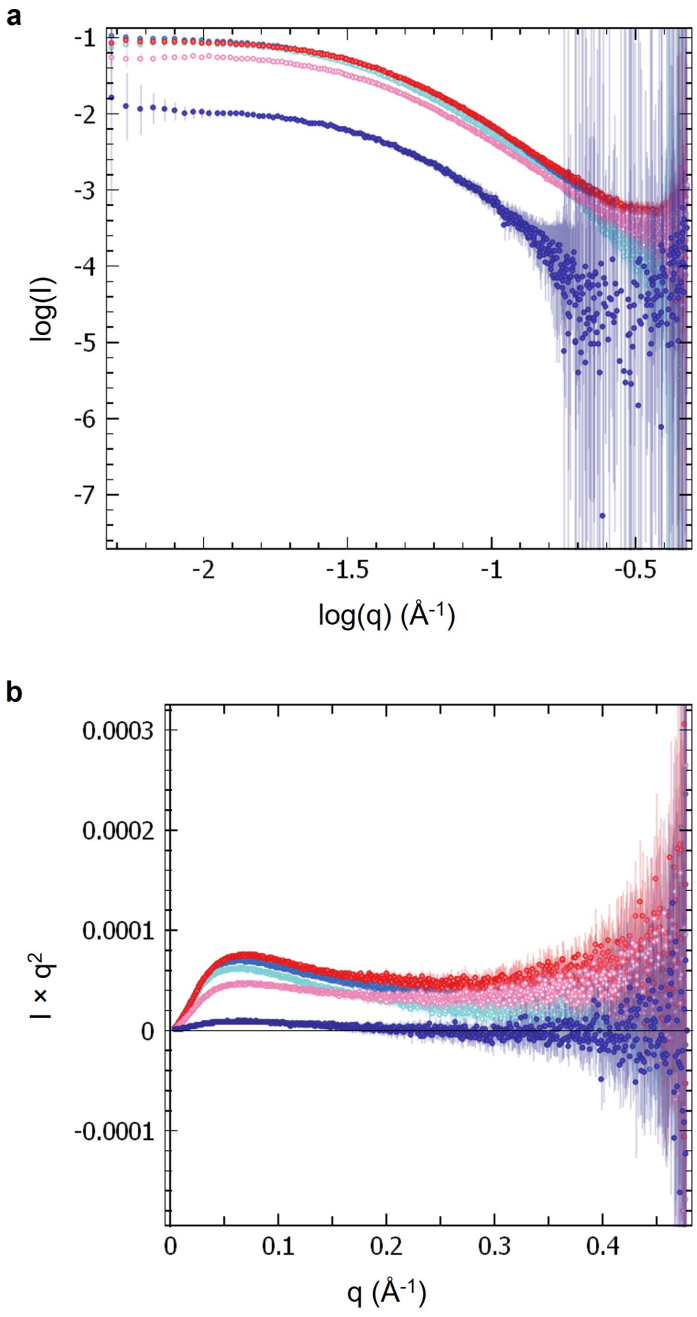

Figure 5.19: SAXS curves of HP1α mutants. Static mode SAXS profiles of HP1α WT (dark blue), S11-14E (light blue), ΔNTE (red), ΔCTE (pink), and A2;3 (purple) presented as a) log of I versus log of q, and b) as a Kratky plot. Experiments conducted using 2.5 mg/mL of HP1α WT, S11-14E, and ΔNTE, 1.7 mg/mL ΔCTE, and 0.4 mg/mL A2;3 in IB1 plus 10 mM TCEP at 22°C.

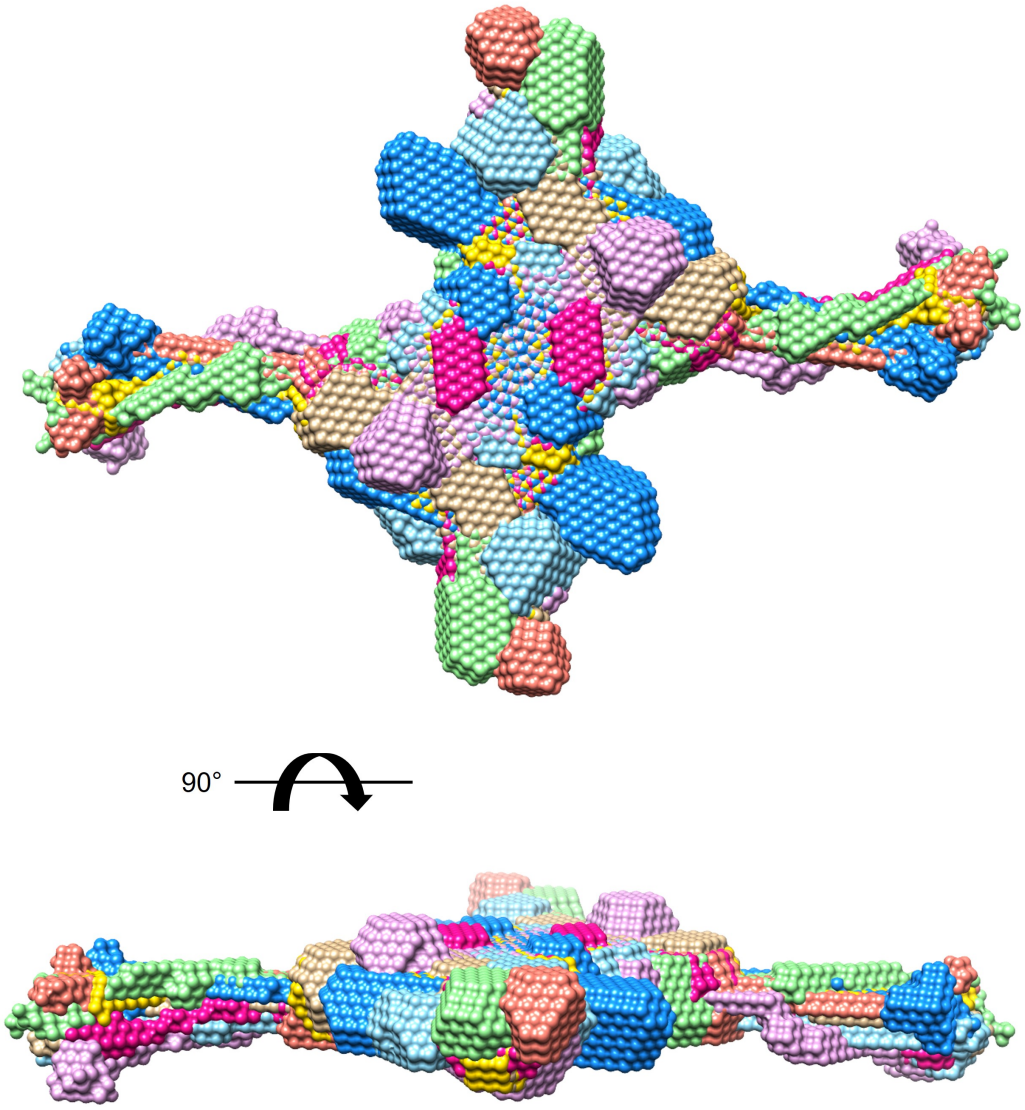

Figure 5.20: Computed models of HP1α S11-14E SAXS data. DAMMIN-generated models of HP1α S11-14E SAXS data based on $P2$ symmetry. Each of 10 models is shown in a different colour. Ensemble resolution 50 Å, final χ^2 against raw data 0.1567, mean value of NSD 1.749. The lower image is rotated backwards 90°.

Table 5.4: SAXS structural parameters of HP1α WT and mutants

Structural parameter	**WT**	**S11-14E**	**ΔNTE**	**ΔCTE**	**A2;3**
Guinier analysis					
R_g (Å)	47.1 ±0.7	46.4 ±0.6	43.2 ±0.6	46.4 ±1.3	42.7 ±1.5
$q{\cdot}R_g$ range	0.41-1.10	0.46-1.17	0.26-1.09	0.49-1.02	0.34-1.29
I(0) (cm^{-1})	0.096 ±0.00069	0.089 ±0.00062	0.09 ±0.00057	0.06 ±0.00074	0.011 ±0.00029
Fidelity	1.00	1.00	1.00	1.00	1.00
P(r) analysis					
R_g (Å)	47.6	48.9	44.3	46.1	46.8
D_{max} (Å)	185.3	206.7	169.5	175.6	182.0
q range	0.0086-0.3001	0.0099-0.3001	0.0061-0.3001	0.0105-0.3001	0.0080-0.2001
Quality estimate	0.66	0.64	0.70	0.63	0.55
Porod volume ($Å^3$)	109,594	122,701	117,123	94,999	129,945
Porod volume ×0.6 (Da)	65,756	73,621	70,274	56,999	77,967

TERRA45 had an R_g of 16.4 Å, which was agreeable with the R_g determined by the $P(r)$ plot which was 16.7 Å, with a D_{max} of 62.1 Å (**Figure A.66**, **Figure A.67**). The Porod volume was 20,404 $Å^3$ (**Table 5.5**), which converted to MW is 12,242 Da, which is similar to the expected MW of a single TERRA45 species (14,813.8 Da), and may be giving a lower value than expected due to the compact structure.

Ab initio modelling was then performed on TERRA45 using DAMMIN, which yielded the structures shown in **Figure 5.22**. These shapes with compact centres and parts extending outwards are unexpected, due to the shapes of crystal structures of parallel G4s which have been published, which show a more regular confined geometric shape.

To determine how this space might be filled with the structure of TERRA45, the crystal structure of a 22 nt telomeric G4 (PDB: 7KLP) was utilised. This short sequence forms a three-tetrad parallel G4 from four G-tracts; therefore, two of these structures were used to represent the eight G-tracts of TERRA45. These structures were then overlaid with one of the DAMMIN structures shown in **Figure 5.22**, where if positioned in an offset fashion, filled the space of the model (**Figure 5.23**). To determine if this offset arrangement fit the SAXS data of TERRA45, FoXS was used. Also included for comparison is an aligned formation of two three-tetrad G4s, illustrated in **Figure 5.24**. FoXS showed that the offset structure had a similar calculated R_g value compared to that derived from the SEC-SAXS data and a lower χ^2 value, indicating a better fit, while the aligned structure had a greater χ^2 value due to the large c_2 value (controlling

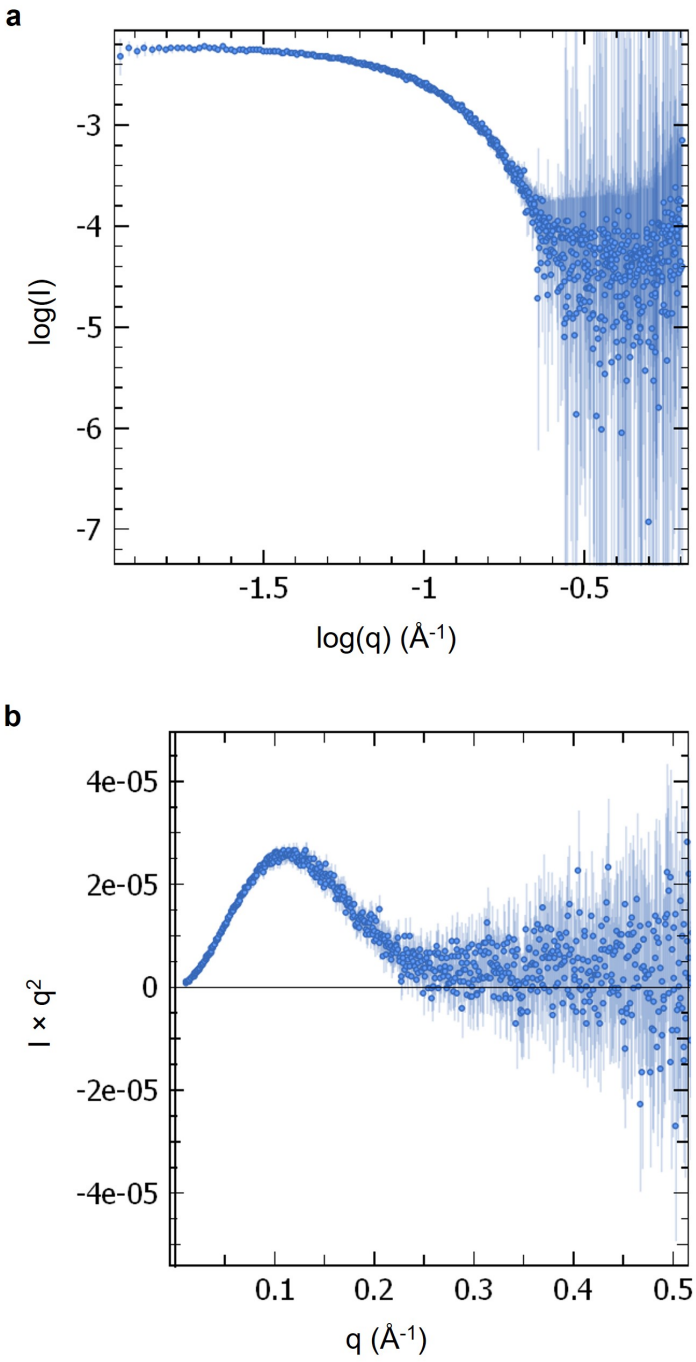

Figure 5.21: SAXS curve of TERRA45. SAXS profiles of SEC-separated TERRA45 presented as a) log of I versus log of q, and b) as a Kratky plot. Experiment conducted using 200 µM TERRA45 in IB1 at 22°C.

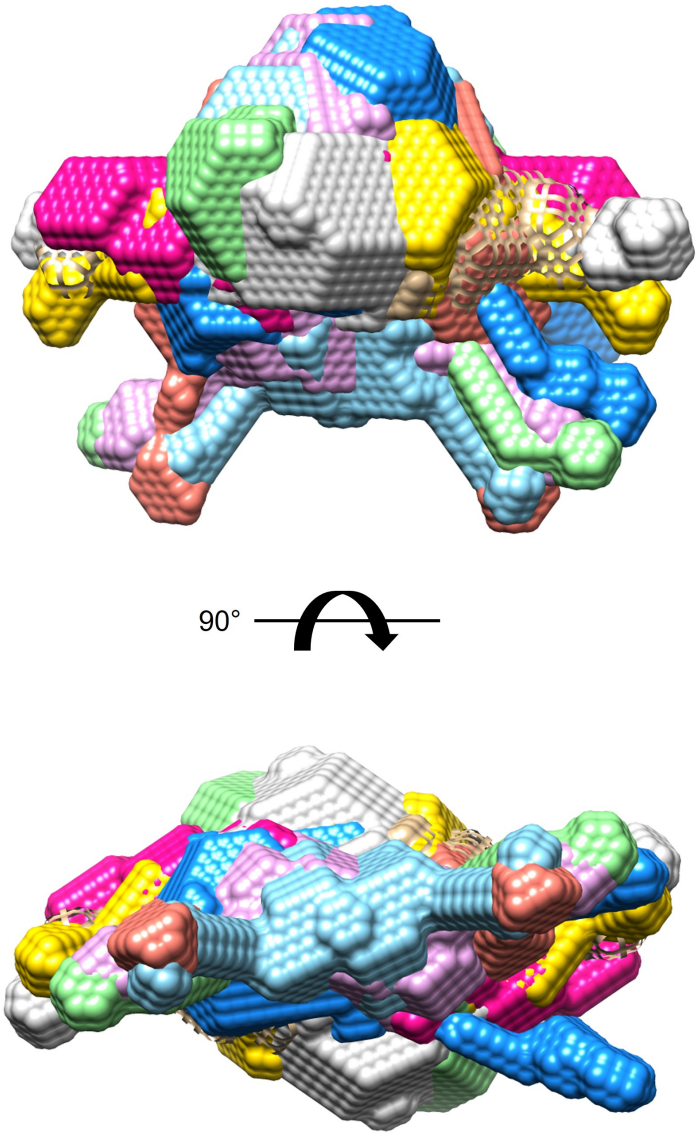

Figure 5.22: Computed models of TERRA45 SEC-SAXS data. DAMMIN-generated models of SEC-separated TERRA45 SAXS data based on $P2$ symmetry. Each of 10 models is shown in a different colour. The bottom image is rotated backwards 90°. Ensemble resolution 33 Å, final χ^2 against raw data 0.3126, mean value of NSD 1.648, indicating diversity of structures generated.

Table 5.5: SEC-SAXS structural parameters of TERRA45

Structural parameter	**TERRA45**
Guinier analysis	
R_g (Å)	16.4 ±0.2
$q{\cdot}R_g$ range	0.25-1.29
I(0) (cm^{-1})	0.0058 ±0.000043
Fidelity	1.00
P(r) analysis	
R_g (Å)	16.7
D_{max} (Å)	62.1
q range	0.0150-0.4863
Quality estimate	0.76
Porod volume ($Å^3$)	20,404
Porod volume ×0.6 = MW (Da)	12,242
Expected MW (Da)	14,814
Ratio to expected MW	0.83

the density of the hydration layer) at its maximum (4.00), and a notably smaller R_g (**Figure 5.25**).

TERRA45 as well as mutTERRA45 were also studied in static mode in order to be comparable to complex static mode SAXS experiments. Both TERRA45 and mutTERRA45 showed pronounced aggregation, shown by the upturn at low q (**Figure 5.26a**), but both also possess highly ordered structures shown by the shapes of the Kratky plots (**Figure 5.26b**). Guinier analysis (**Figure A.68**) showed that TERRA45 measured in static mode had an increased R_g of 23.9 Å (**Table 5.6**), compared with the SEC-separated TERRA45 R_g of 16.4 Å (**Table 5.5**). MutTERRA45 showed a similarly large R_g of 23.4 Å (**Figure A.70**, **Table 5.6**). $P(r)$ analysis (**Figure A.69**, **Figure A.71**) gave R_g values of 26.7 Å and 26.1 Å for TERRA45 and mutTERRA45, respectively, and larger D_{max} values of 113.6 Å for TERRA45 and 112.1 Å for mutTERRA45 (**Table 5.6**). The larger R_g and D_{max} values for these data compared to SEC-SAXS data are likely due to the aggregation present in static mode SAXS measurements. However, the molecular weight estimations from the Porod volumes still indicated they were primarily monomeric (**Table 5.6**).

Models of TERRA45 and mutTERRA45 were then generated using DAMMIN, which each showed similar structures to the SEC-SAXS TERRA45 data, again with a compact centre and thin parts extending outwards (**Figure 5.27**, **Figure 5.28**).

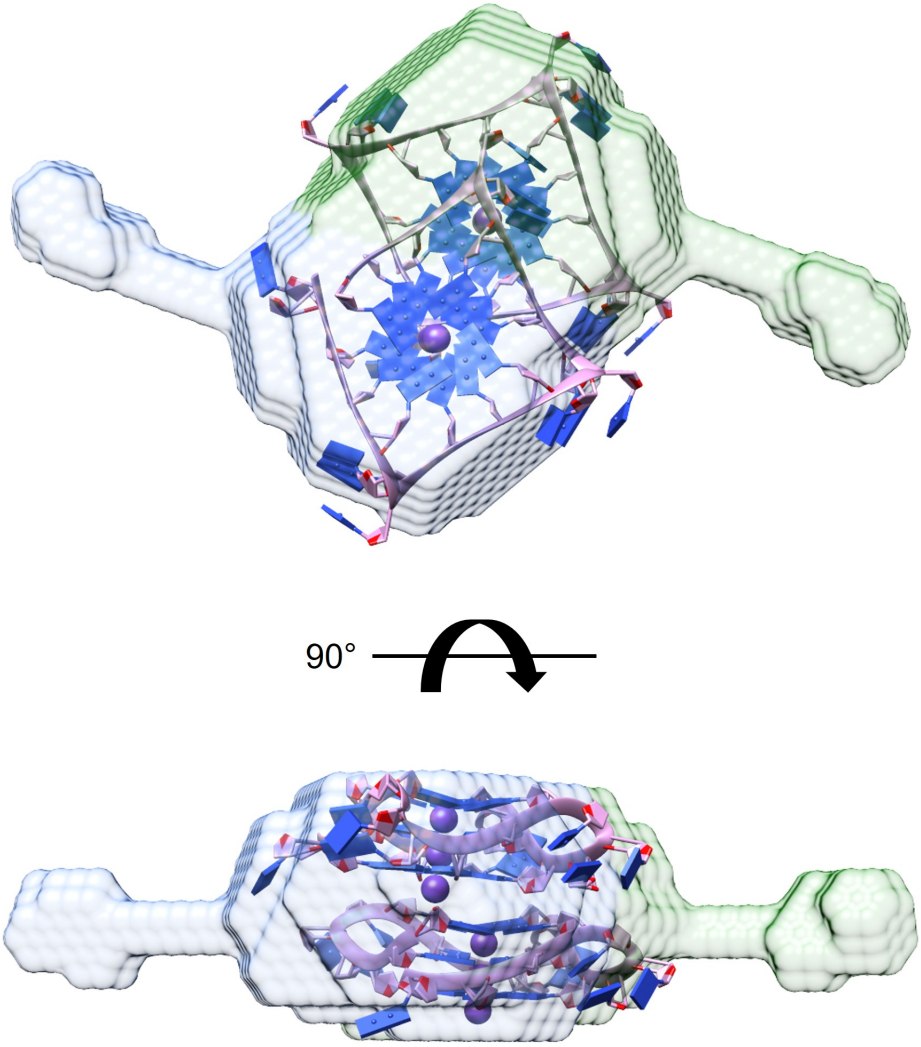

Figure 5.23: Crystal structures of TERRA overlaid with DAMMIN-generated TERRA45 model. Two copies of the crystal structure of a three-tetrad, parallel, K^+-stabilized TERRA G-quadruplex (pink, PDB: 7KLP) with a DAMMIN-generated model of SEC-separated TERRA45 from SAXS data (blue and green). The lower image is rotated backwards 90°.

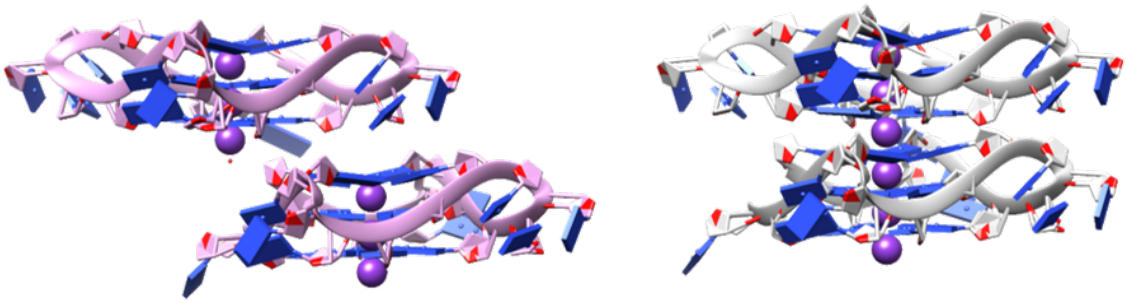

Figure 5.24: Comparison of aligned and offset structures of two TERRA G4s. Offset structure of two 7KLP models from **Figure 5.23** shown on the left, and aligned models shown on the right.

Table 5.6: SAXS structural parameters TERRA45 and mutTERRA45 (static mode)

Structural parameter	**TERRA45**	**mutTERRA45**
Guinier analysis		
R_g (Å)	23.9 ±0.2	23.4 ±0.3
$q{\cdot}R_g$ range	0.54-1.11	0.74-1.16
I(0) (cm^{-1})	0.067 ±0.00028	0.064 ±0.00045
Fidelity	1.00	1.00
P(r) analysis		
R_g (Å)	26.7	26.1
D_{max} (Å)	113.6	112.1
q range	0.0227-0.3345	0.0316-0.3415
Quality estimate	0.62	0.62
Porod volume ($Å^3$)	30,258	24,949
Porod volume ×0.6 = MW (Da)	18,115	14,969
Expected MW (Da)	14,814	14,658
Ratio to expected MW	1.23	1.02

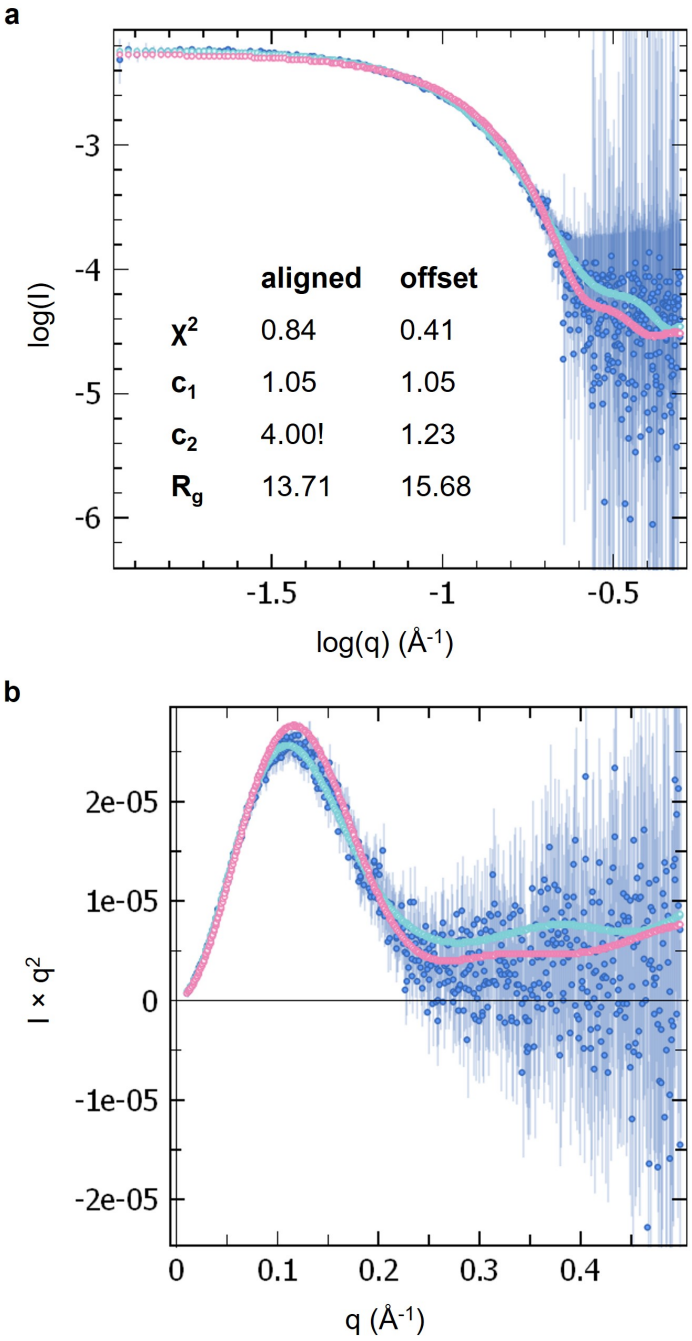

Figure 5.25: Fit of FoXS-generated SAXS profiles of aligned and offset TERRA G4s to TERRA45 SAXS data. FoXS-generated SAXS curves of the aligned (pink) and offset (light blue) 7KLP models with SAXS data of SEC-separated TERRA45 shown in dark blue. The data are presented as a) log of I versus log of q, and b) as a Kratky plot.

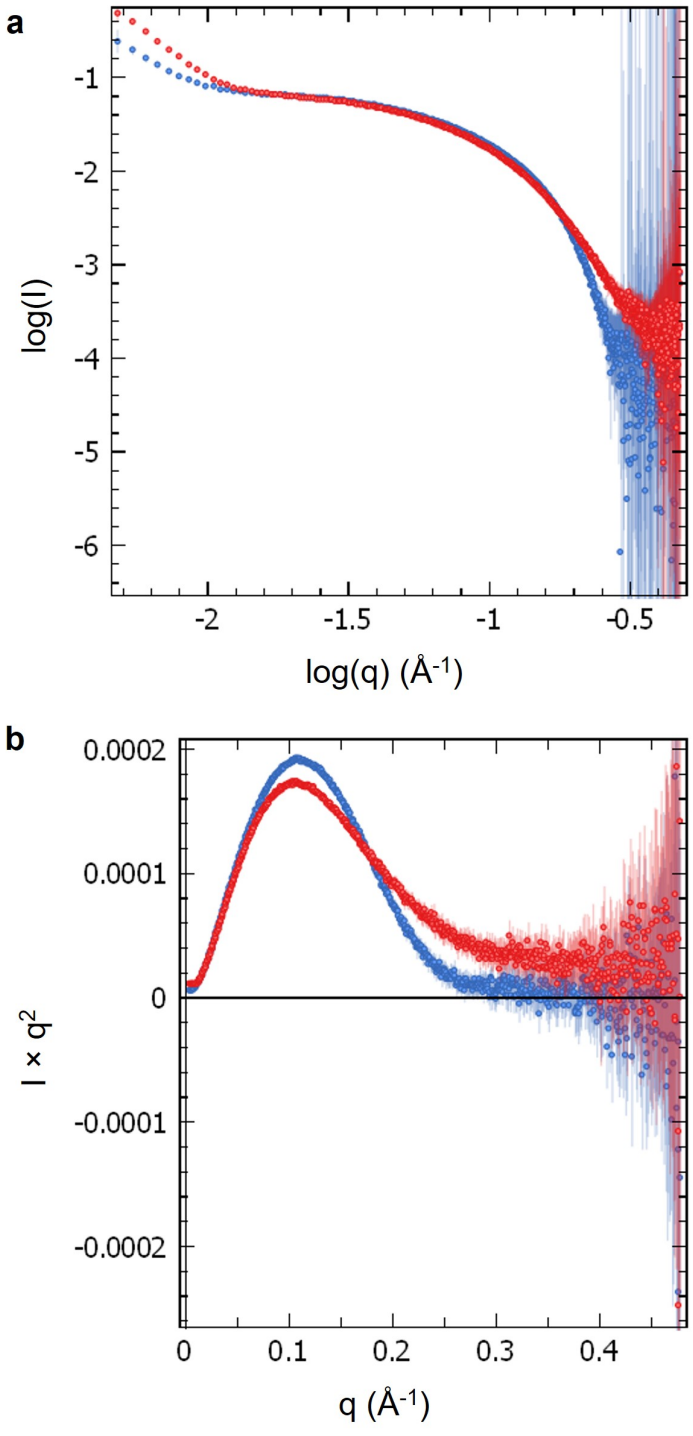

Figure 5.26: SAXS curves of TERRA45 and mutTERRA45. Static mode SAXS profiles of TERRA45 (blue) and mutTERRA45 (red) presented as a) log of I versus log of q, and b) as a Kratky plot. Experiment conducted using 100 µM oligonucleotide in IB1 plus 10 mM TCEP at 22°C.

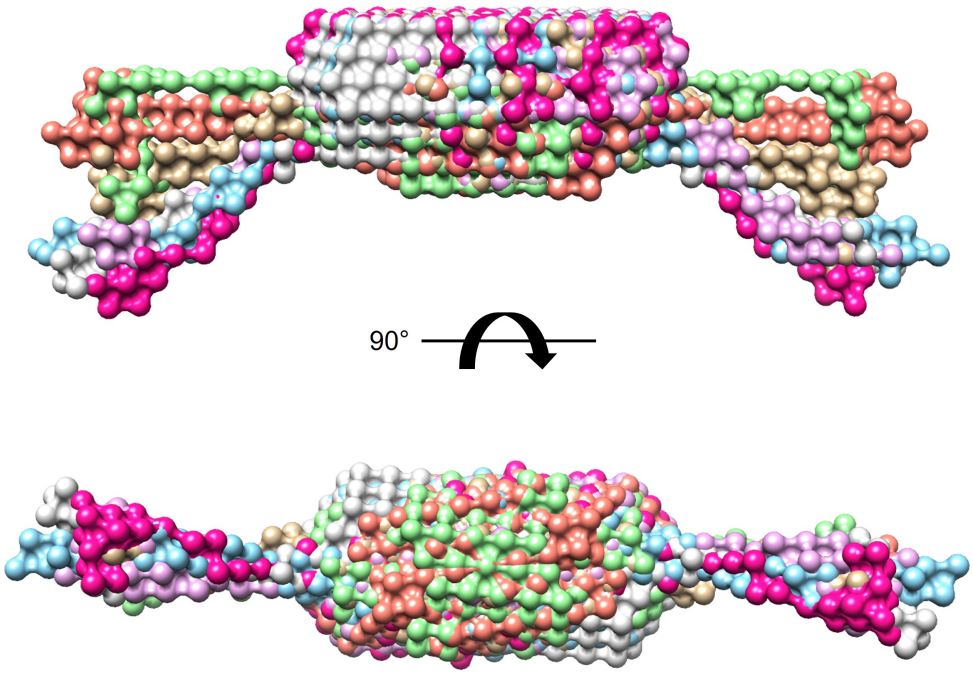

Figure 5.27: Computed models of TERRA45 static mode SAXS data. DAMMIN-generated models of static mode TERRA45 SAXS data based on $P2$ symmetry. Each of 9 models is shown in a different colour (1 model removed due to being more than 2 standard deviations from the mean NSD). The lower image is rotated backwards 90°. Ensemble resolution 33 Å, final χ^2 against raw data 0.3097, mean value of NSD 1.313.

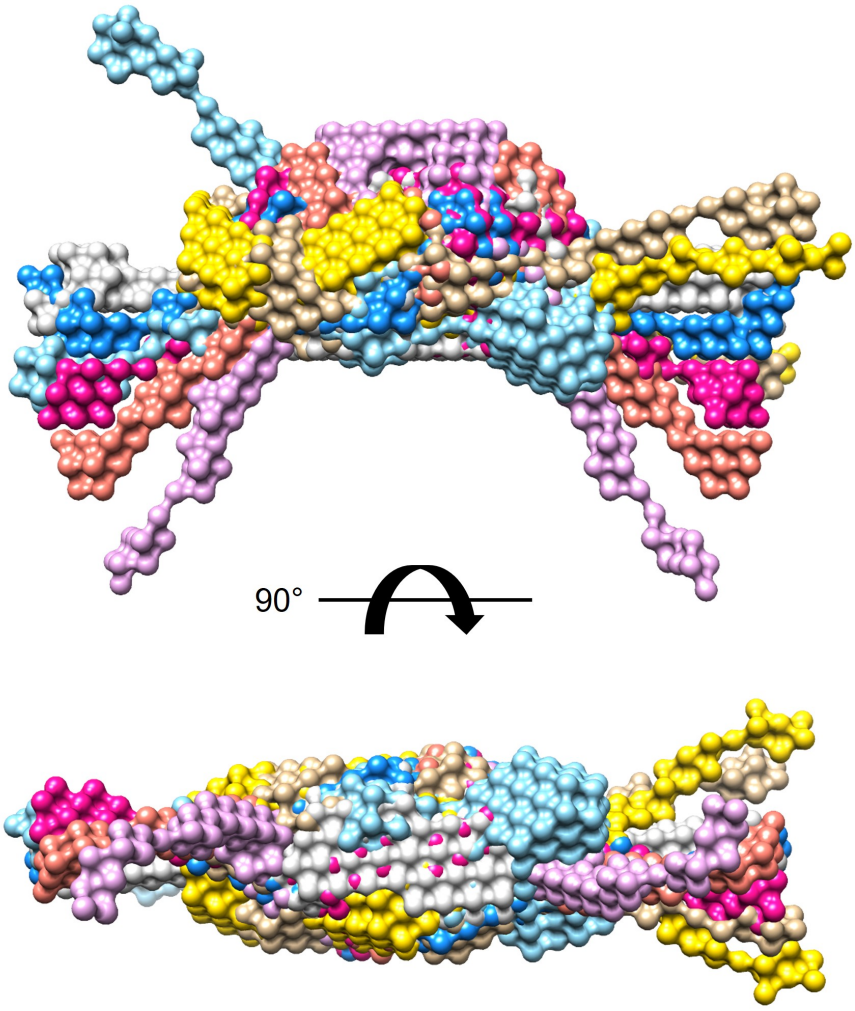

Figure 5.28: Computed models of mutTERRA45 static mode SAXS data. DAMMIN-generated models of static mode mutTERRA45 SAXS data based on $P2$ symmetry. Each of 10 models is shown in a different colour. The lower image is rotated backwards 90°. Ensemble resolution 33 Å, final χ^2 against raw data 0.2191, mean value of NSD 1.561.

5.2.3 Analysis of HP1α-G4 complexes by SAXS

When binding between molecules occurs with a high dissociation rate, as is the case in the binding of HP1α and TERRA45, SEC causes dissociation of the complex by shifting the equilibrium as the molecules are separated (Ignatev et al., 2007). Therefore, SEC-SAXS was not able to be utilised for analysing complexes between HP1α and TERRA45 as SEC yielded separate components of the complex.

5.2.3.1 The αHinge-TERRA complex

To first analyse the complex of HP1α and G4, the αHinge mutant was measured by SAXS with an equimolar addition of TERRA45. SAXS showed that the complex of αHinge with TERRA45 was less flexible than each of the components alone, shown by a shift to the left of the Kratky plot (**Figure 5.29**, αHinge shown here at the same molar concentration as TERRA45, a higher concentration of αHinge used in previous section for analysis) which indicates a potential conformational change. Analysis of the structural parameters showed an R_g of 28.9 Å (**Figure A.72**) and a D_{max} of 133.3 Å for the complex (**Table 5.7**, **Figure A.73**). Using the Porod volume to determine the molecular weight showed that the complex is likely at a 1:1 ratio, where the Porod volume estimate is 21,852 Da, and the expected molecular weight is 23,055 Da (8,214 Da + 14,814 Da, **Table 5.7**).

To produce a molecular envelope of the complex between the HP1α hinge and TERRA45, DAMMIN was used. *Ab initio* modelling showed an elongated shape, where one end held the majority of the volume (**Figure 5.30**). To determine how and where TERRA45 binds the hinge of HP1α, various models of αHinge plus the offset TERRA45 structure (**Figure 5.24**) were created and assessed for their similarity to the SAXS data of the complex. Each combination of αHinge polypeptide with TERRA45 yielded high χ^2 values (**Figure 5.31**), indicating that none of the models represented the SAXS data collected. The first model (i) uses the αHinge AlphaFold model with offset TERRA45 at one end of the molecule, to mimic the elongated shape of the model with more mass on one end. This model (i) yielded the best fit due to having the largest R_g of 29.1 Å; however, it had a c_2 value at its maximum value of 4.00. The second model (ii) arranged the offset TERRA45 in the centre of the αHinge AlphaFold model, allowing for access of charged patches in the hinge; however, it had a χ^2 value of 58.0, an R_g of only 21.1

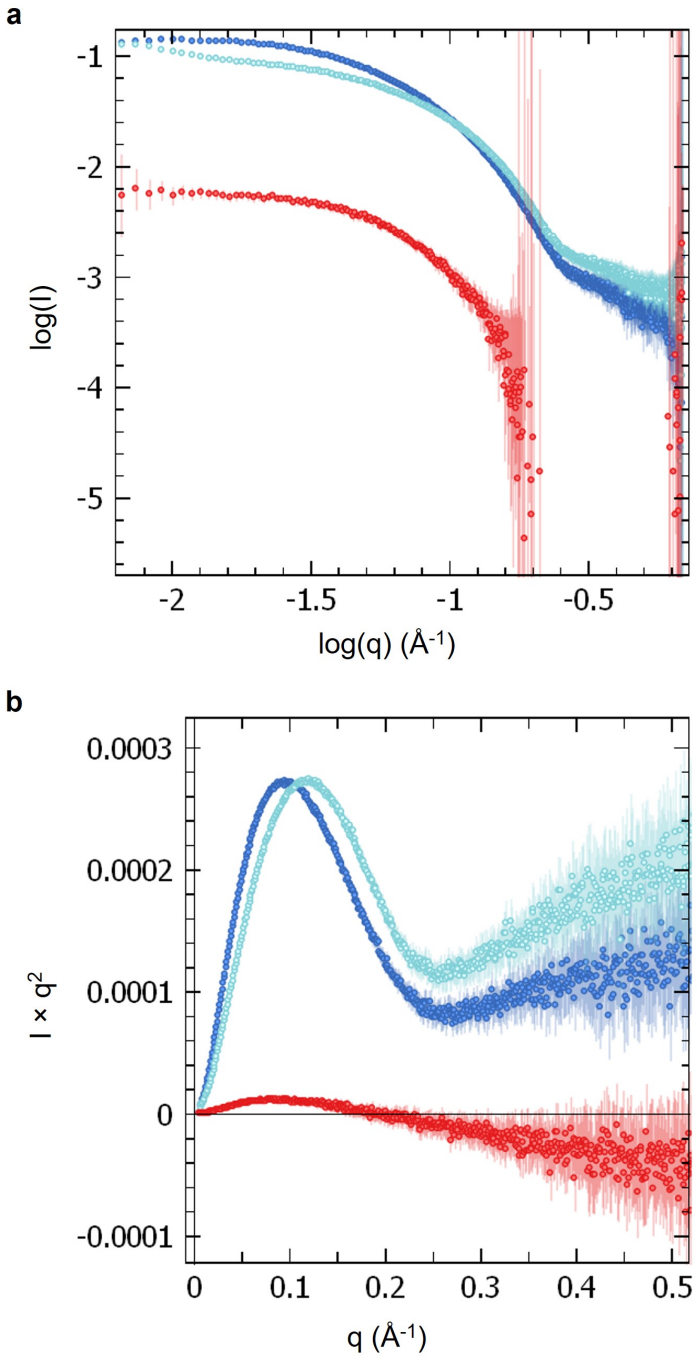

Figure 5.29: SAXS curves of αHinge plus TERRA45. SAXS profiles of αHinge (red), TERRA45 (light blue), and a complex of the two (dark blue) presented as a) log of I versus log of q, and b) as a Kratky plot. Experiments conducted using 1.0 mg/mL (120 µM) of αHinge and/or 120 µM TERRA45 in IB1 plus 10 mM TCEP at 22°C.

Table 5.7: SAXS structural parameters of the αHinge-TERRA45 complex

Structural parameter	**αHinge**	**TERRA45**	**Complex**
Guinier analysis			
R_g (Å)	23.5 ±0.3	23.9 ±0.2	28.9 ±0.2
$q{\cdot}R_g$ range	0.28-1.29	0.54-1.11	0.47-1.02
I(0) (cm^{-1})	0.017 ±0.00012	0.067 ±0.00028	0.14 ±0.00046
Fidelity	1.00	1.00	1.00
P(r) analysis			
R_g (Å)	25.1	26.7	30.9
D_{max} (Å)	94.4	113.6	133.3
q range	0.0127-0.3397	0.0227-0.3345	0.0171-0.3004
Quality estimate	0.65	0.62	0.62
Porod volume ($Å^3$)	13,021	30,258	36,420
Porod volume ×0.6 = MW (Da)	7,813	18,155	21,852
Expected MW (Da)	8,241	14,814	23,055 (1:1)
Ratio to expected MW	0.95	1.23	0.95

Å, and a c_2 value of 4.00. The third model (iii) arranged the offset TERRA45 at one end of the αHinge AlphaFold model, however at an alternative rotation to the first model. This model (iii) had a χ^2 value of 72.3, an R_g of 22.8 Å, and a c_2 value of 4.00. Due to the poor fit of these structures to the SAXS data, and the evidence that there may be a conformational change in the complex, an AlphaFold generated model of the αHinge with an alternatively folded structure was produced. This bent conformation (iv) was then arranged with the offset TERRA45 model to align charged patches with the phosphate backbone. Despite the good spacial fit and alignment of the TERRA45 with the bent hinge structure, FoXS determined that this structure did not represent the SAXS data, with the poorest χ^2 value of 97.98 (**Figure 5.31**). It is likely, then, that the hinge adopts a more extended conformation for binding TERRA45. It should also be noted that the αHinge does not bind with high specificity for only G4s, so the structure it forms in absence of the CSD may not be representative of how it behaves in the whole HP1α protein.

5.2.3.2 The HP1α monomer-TERRA complex

To explore the binding between TERRA45 and the entire HP1α protein, the monomeric Y168E mutant was utilised, and its complex with TERRA45 analysed. SAXS data showed similar results to the αHinge plus TERRA45, whereby addition of TERRA45 caused a shift to the left in the Kratky plot and a curve that returns nearly to the baseline (**Figure 5.33**), indicating a de-

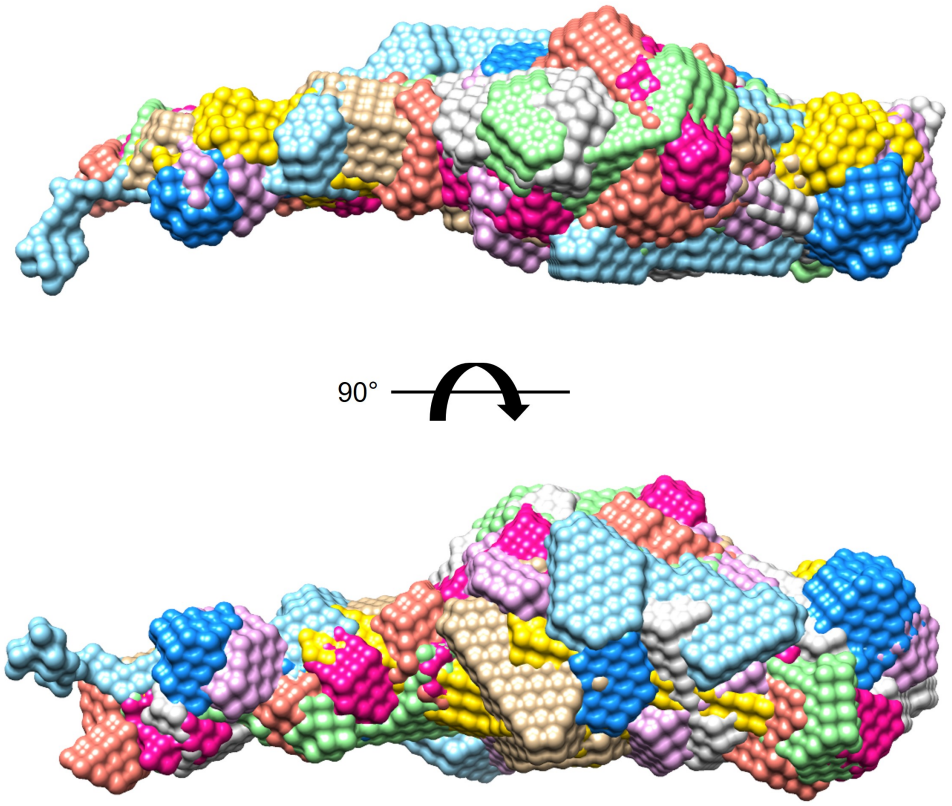

Figure 5.30: Computed models of αHinge+TERRA45 complex SAXS data. DAMMIN-generated models of αHinge plus TERRA45 SAXS data based on $P1$ symmetry. Each of 10 models is shown in a different colour. The bottom image is rotated backwards 90°. Ensemble resolution 33 Å, final χ^2 against raw data 0.1790, mean value of NSD 0.992.

Figure 5.31: αHinge AlphaFold models with offset TERRA45 structures. AlphaFold-generated model of the αHinge polypeptide sequence (purple) with the offset duplicate 7KLP crystal structures to represent TERRA45 (purple). Amino acids corresponding to the charged patches in the hinge analysed in **Chapter 3** are highlighted in red (68-72KKYKK, 89-91KRK, 104-106KKK). Images i), ii), and iii) were created using the his-αHinge sequence, and iv) was created using the his-αHinge sequence with mutations 89-91KRK>AAA and 104-106KKK>AAA.

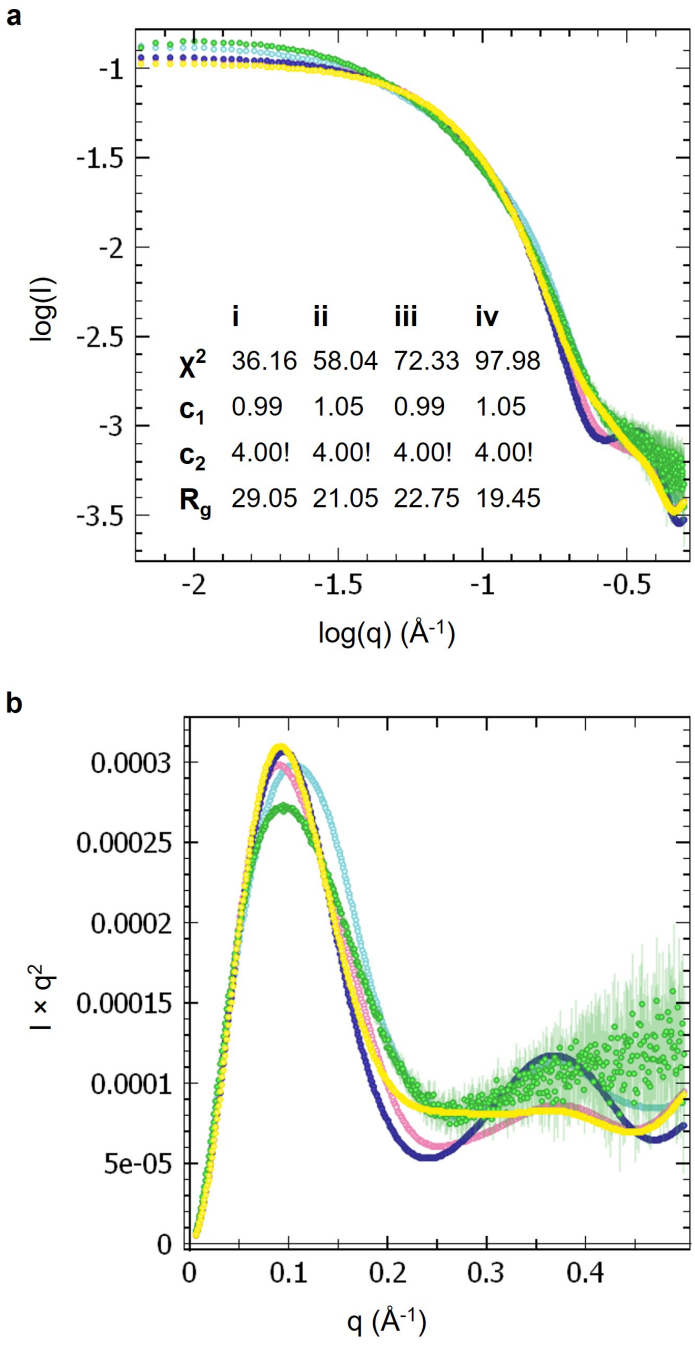

Figure 5.32: Fit of FoXS-generated SAXS profiles of αHinge AlphaFold models with offset TERRA45 structures to SAXS data of the complex. FoXS-generated SAXS curves of the structures shown in **Figure 5.31**, where i) is shown in blue, ii) is shown in pink, iii) is shown in purple, iv) is shown in yellow, and the SAXS data of αHinge plus TERRA45 is shown in green. The data are presented as a) log of I versus log of q, and b) as a Kratky plot.

Table 5.8: SAXS structural parameters of the HP1α Y168E-TERRA45 complex

Structural parameter	**HP1α Y168E**	**TERRA45**	**Complex**
Guinier analysis			
R_g (Å)	32.0 ±0.6	23.9 ±0.2	36.6 ±0.2
$q{\cdot}R_g$ range	0.41-1.30	0.54-1.11	0.72-1.17
I(0) (cm^{-1})	0.0056 ±0.000065	0.067 ±0.00028	0.30 ±0.001
Fidelity	1.00	1.00	0.97
P(r) analysis			
R_g (Å)	30.6	26.7	38.9
D_{max} (Å)	91.1	113.6	151.0
q range	0.0136-0.3004	0.0227-0.3345	0.0206-0.3004
Quality estimate	0.63	0.62	0.65
Porod volume ($Å^3$)	37,243	30,258	71,794
Porod volume ×0.6 = MW (Da)	22,346	18,155	43,076
Expected MW (Da)	24,979	14,814	39,793 (1:1)
Ratio to expected MW	0.89	1.23	1.08

crease in flexibility of the complex and a more compact structure. The HP1α Y168E-TERRA45 complex had an R_g of 36.6 Å determined by Guinier analysis (**Figure A.74**), although there is evidence of aggregation. A D_{max} of 151.0 Å was determined by $P(r)$ analysis (**Table 5.8**, **Figure A.75**). The estimated molecular weight based on the Porod volume was 43,076 Da, which is similar to the expected molecular weight of a 1:1 complex of monomeric HP1α Y168E with TERRA45 (24,979 Da + 14,814 Da = 39,793 Da).

Ab initio modelling was then performed on the complex using DAMMIN, which yielded the structures shown in **Figure 5.34**. To understand how TERRA45 and HP1α fit into this model, FoXS was firstly used to assess the binding model previously proposed in **Chapter 4** (**Figure 4.18**), but with the G4 replaced with the offset model. This combination did not visually fit the molecular envelope generated (**Figure 5.35**), and FoXS also determined it was a poor fit (**Figure 5.36**).

Due to the shift of the Kratky plot with addition of TERRA45 to both αHinge and Y168E, it is likely that a change in conformation of the whole HP1α protein occurs, therefore using the structure of unbound HP1α may not be correct. It should be noted that addition of HP1α does not influence the parallel G4 topology of TERRA45 (**Figure A.76**), nor the anti-parallel G4 topology (**Figure A.77**), as shown by circular dichroism spectroscopy experiments.

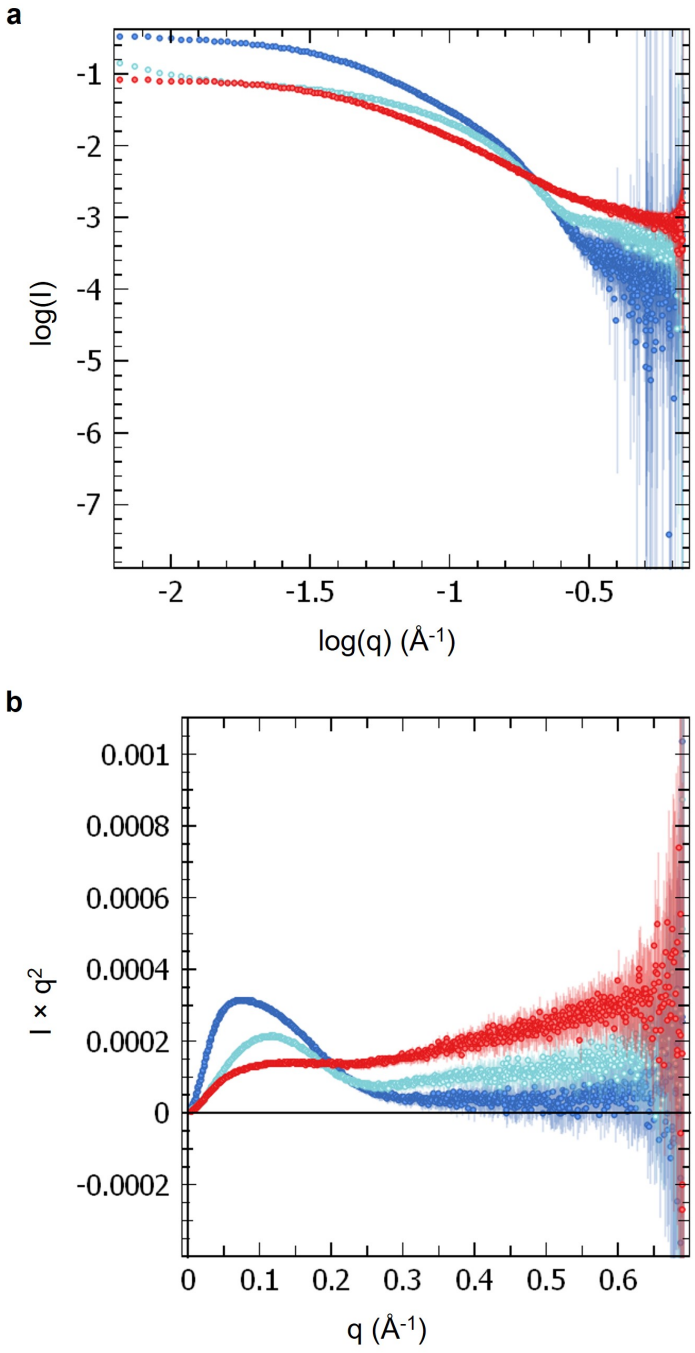

Figure 5.33: SAXS curves of HP1α Y168E plus TERRA45. SAXS profiles of HP1α Y168E (red), TERRA45 (light blue), and a complex of the two (dark blue) presented as a) log of I versus log of q, and b) as a Kratky plot. Experiments conducted using 2.5 mg/mL (100 µM) of HP1α Y168E and/or 100 µM TERRA45 in IB1 plus 10 mM TCEP at 22°C.

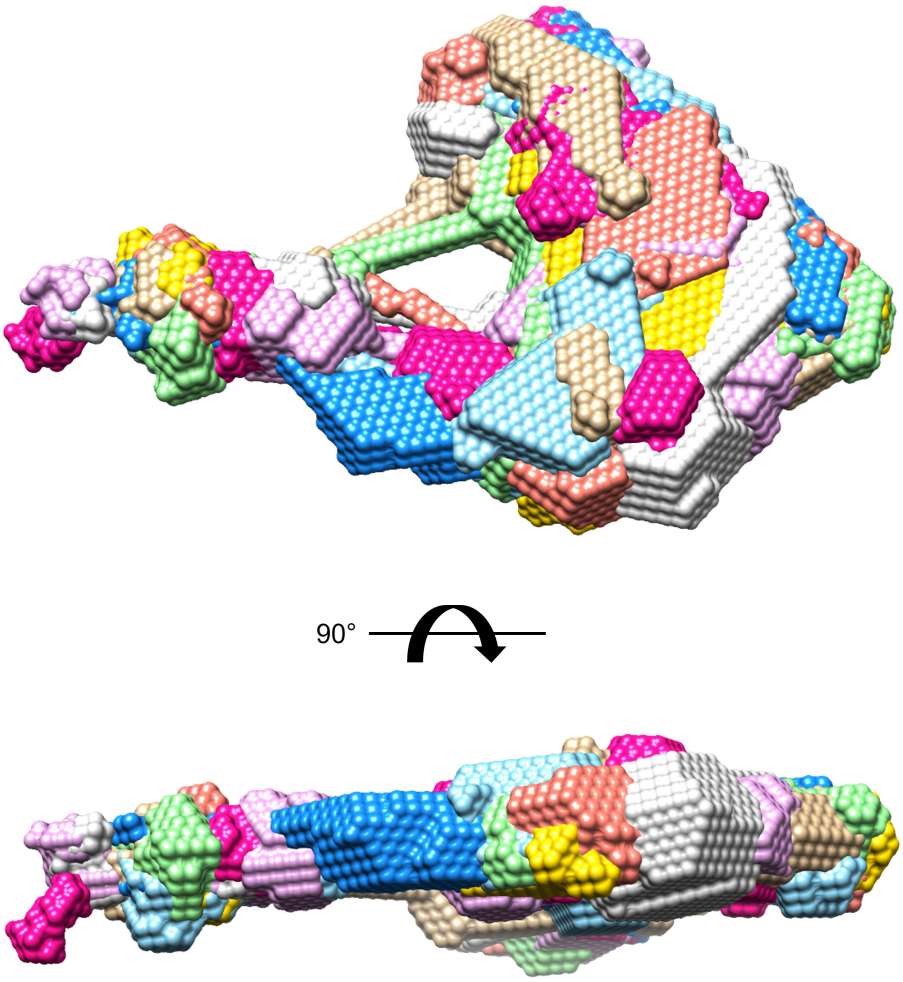

Figure 5.34: Computed models of HP1α Y168E-TERRA45 complex SAXS data. DAMMIN-generated models of monomeric HP1α Y168E-TERRA45 SAXS data of the complex based on $P1$ symmetry. Each of 10 models is shown in a different colour. Ensemble resolution 33 Å, final χ^2 against raw data 0.1459, mean value of NSD 1.216.

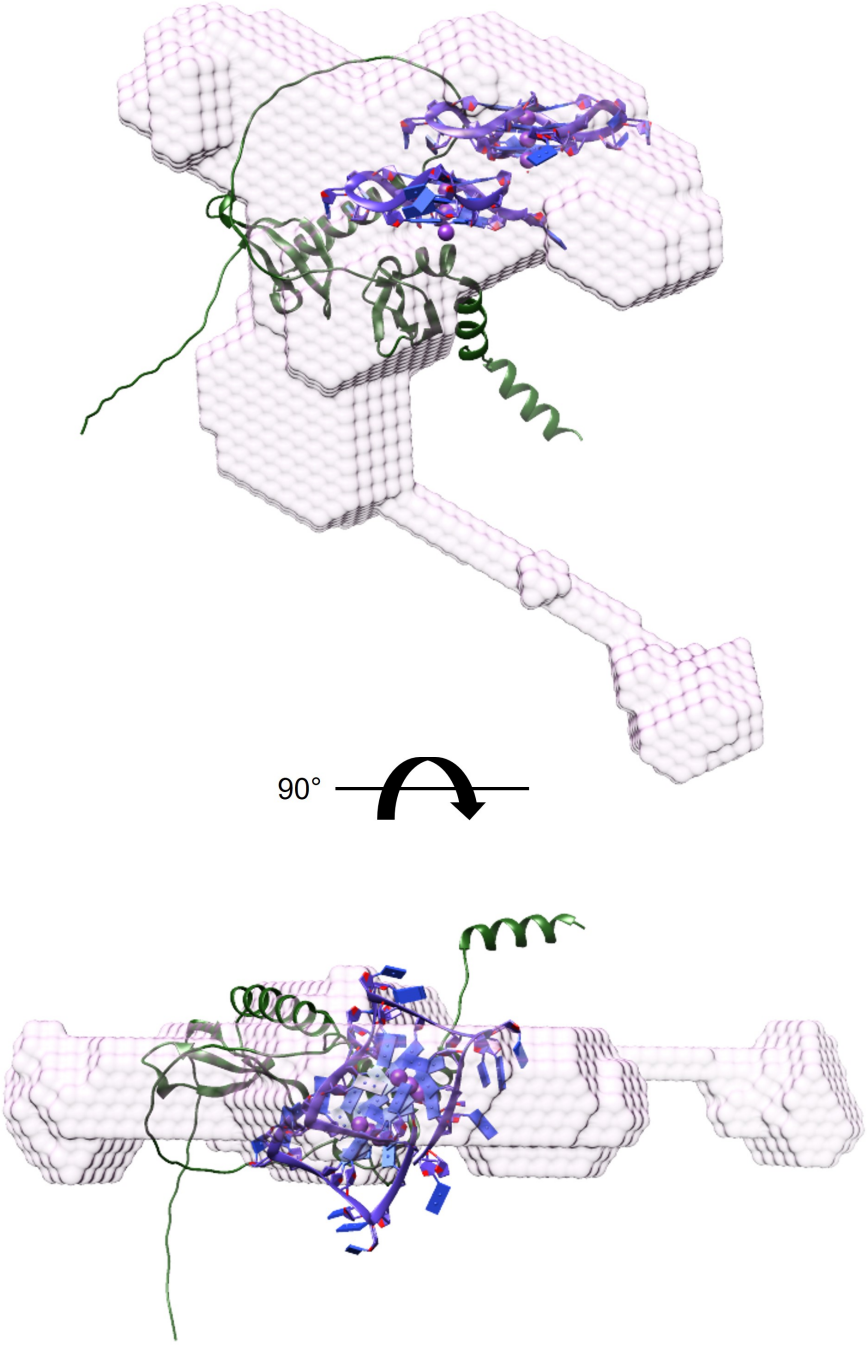

Figure 5.35: HP1α monomer AlphaFold model with offset TERRA45 structure. AlphaFold model AF-Q61686-F1 (green) and offset TERRA G4 (purple) with a DAMMIN-generated model of the HP1α Y168E-TERRA45 complex (pink) from SAXS data. The N-terminus of AF-Q61686-F1 is on the left, and the C-terminus on the right. The lower image is rotated forward 90°.

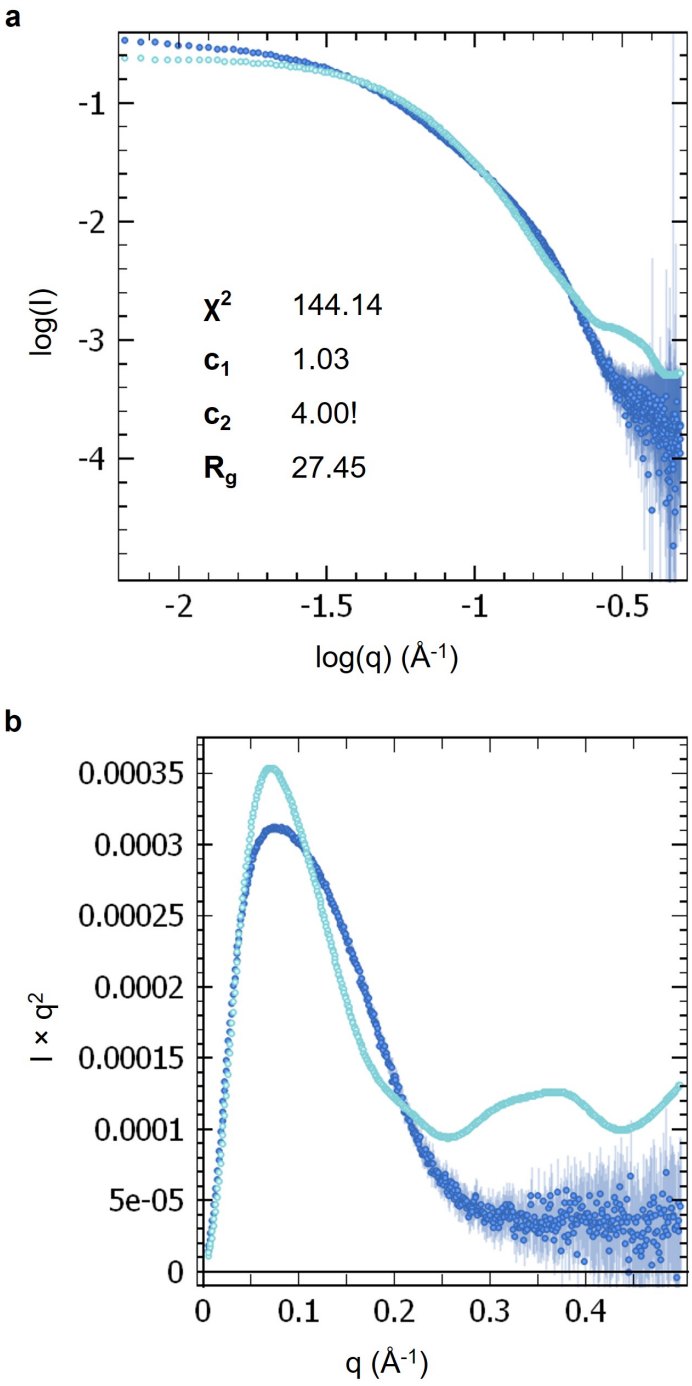

Figure 5.36: Fit of FoXS-generated SAXS profile of AF-Q61686-F1 with offset TERRA45 to complex SAXS data. FoXS-generated SAXS curves of monomeric HP1α AlphaFold model AF-Q61686-F1 plus offset TERRA45 structure (light blue) to SAXS data collected of the HP1α Y168E-TERRA45 complex (dark blue). The data are presented as a) log of I versus log of q, and b) as a Kratky plot.

To attempt to fit a more elongated HP1α monomer structure to the DAMMIN-generated complex model, AlphaFold was used to generate structures of components of his-HP1α based on the polypeptide sequence. The NTE+CD, hinge, and CSD+CTE were created and fit onto the DAMMIN-generated HP1α Y168E-TERRA45 envelope, along with the offset TERRA45 structure. Three different fits were created and tested for their similarity to the SAXS profile using FoXS. The first (**Figure 5.37i**) of the structures arranged the NTE and CD so that the NTE may extend to fill the space of the long protrusion, linked the CD to the hinge, and arranged the CSD and offset TERRA45 structure in a similar manner to **Figure 4.18**, with one of the helices in the CSD contacting the G-tetrad. This arrangement (i) also allowed for charged patches in the hinge to contact the negatively charged RNA backbone. However, FoXS determined this was a bad fit with a χ^2 value of 60.0, a c_2 of 4.00, and an R_g of only 30.9 Å (**Figure 5.38**). To increase the R_g to aid in fitting, the second arrangement (**Figure 5.37ii**) unfolded the NTE by splitting the structure in two, allowing it to fill the outward protrusion. FoXS determined this was also a bad fit, with an increased χ^2 of 76.8, a c_2 of 4.00, and a larger R_g of 37.6 Å (**Figure 5.38**). A third arrangement (**Figure 5.37iii**) was tested, where the TERRA45-CSD was rotated to fill the upper portion of the envelope, still allowing for contact with the charged hinge patches. FoXS also determined this to be a poor fit, with the worst χ^2 of 109.9, c_2 of 4.00, and an R_g of 39.8 Å (**Figure 5.38**).

5.2.3.3 The HP1α dimer-TERRA complex

The HP1α-TERRA45 complex was also analysed, and the SAXS profile again showed a shift in the Kratky plot upon addition of TERRA45, but not upon addition of mutTERRA45 (**Figure 5.39**), indicating that binding of TERRA45 causes this shift, not just addition of a highly structured oligonucleotide. Guinier analysis of HP1α plus mutTERRA45 gave an R_g of 48.1 Å (**Figure A.80**), and $P(r)$ analysis gave a similar R_g of 48.1 Å and a D_{max} of 184.9 Å (**Table 5.9**, **Figure A.81**). The Porod volume of HP1α and mutTERRA45 was shown to be 70,013 Å^3, which when converted to MW gives 42,008 Da. Given the lower k_{on} and higher k_{off} values for binding between HP1α and mutTERRA45 (**Figure 4.2b**, **Table 4.1**), the mixture between the two may result in free components rather than a complex. If the weighted average of HP1α and mutTERRA45 masses are calculated ((50,026×50,026 + 14,658×14,658) / (50,026+14,658)), the result of 42,011 Da is similar to the estimated molecular weight derived from the Porod volume of 42,008 Da. Therefore, the contents of this sample is likely to be free HP1α and mutTERRA45,

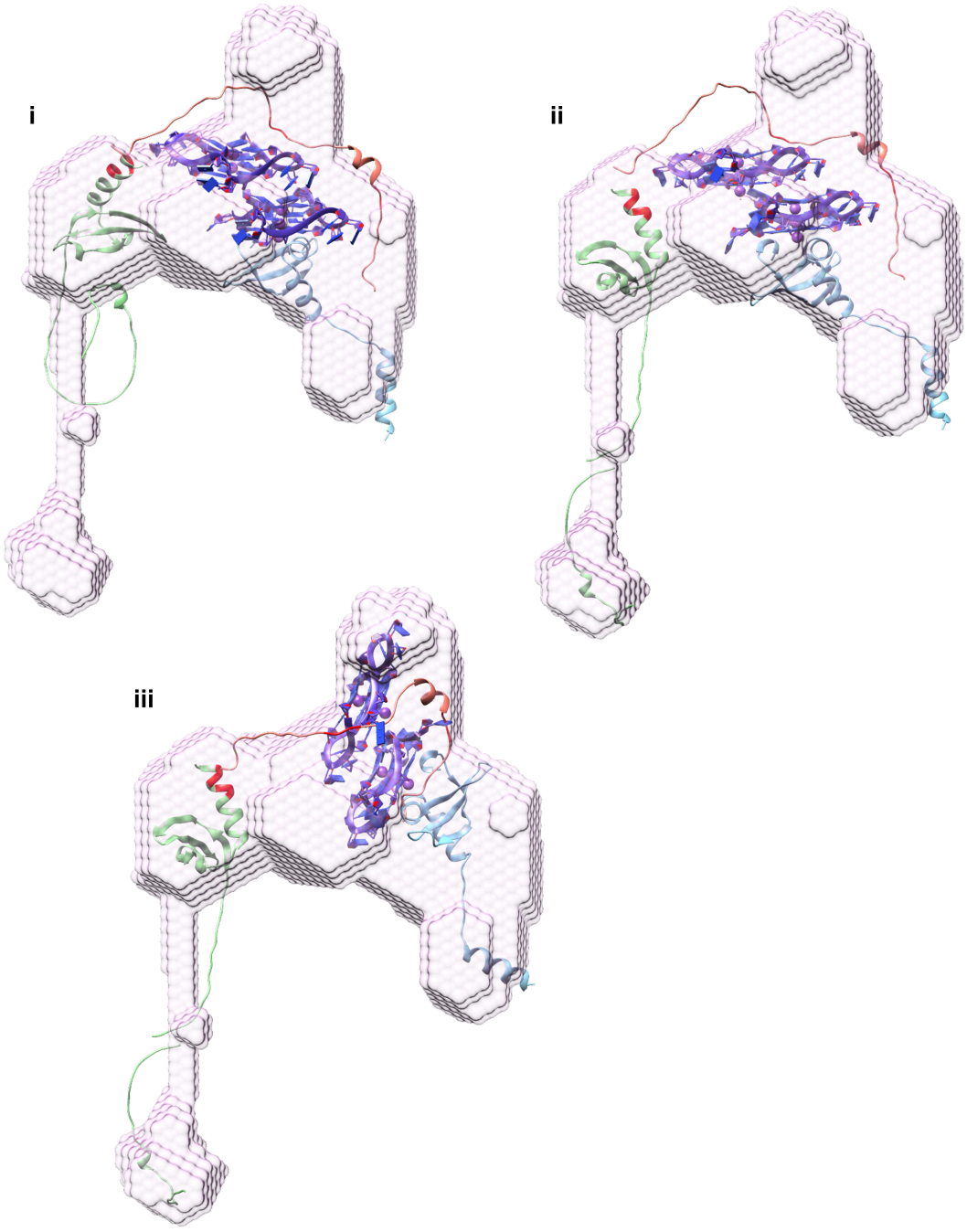

Figure 5.37: Separated HP1α domain structures with offset TERRA45 structure fit onto HP1α Y168E-TERRA45 envelope. AlphaFold-generated NTE+CD (green), hinge (orange), and CSD+CTE (blue) with 68-72KKYKK, 89-91KRK, and 104-106KKK highlighted in red. The offset TERRA45 model is shown in purple, and the DAMMIN-generated model of the HP1α Y168E-TERRA45 complex from SAXS data is shown in pink.

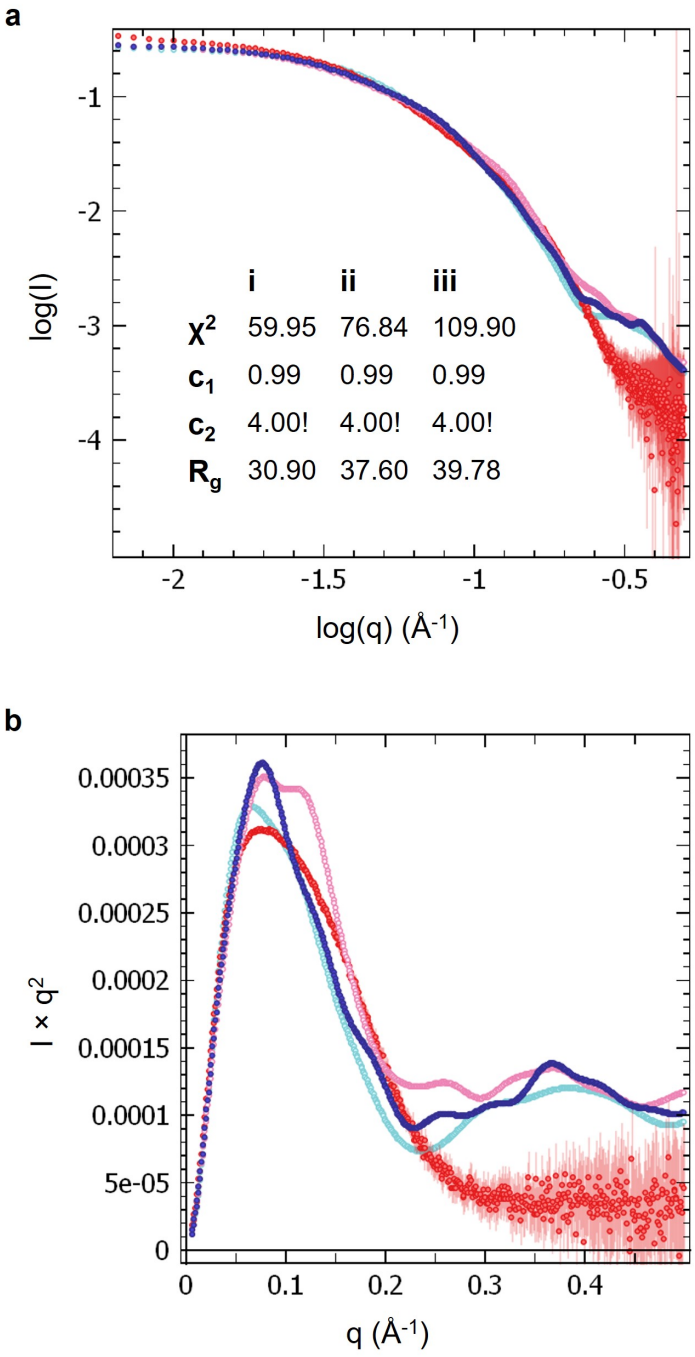

Figure 5.38: Fit of FoXS-generated SAXS profiles of separated HP1α domain structures and TERRA45 with HP1α Y168E complex SAXS data. FoXS-generated SAXS curves of structures in **Figure 5.37** plus offset TERRA45 structure to SAXS data collected of the HP1α Y168E-TERRA45 complex (red). Fit **i** is shown in blue, **ii** in purple, and **iii** in pink. The data are presented as a) log of I versus log of q, and b) as a Kratky plot.

Table 5.9: SAXS structural parameters HP1α with either TERRA45 or mutTERRA45

Structural parameter	**HP1α + TERRA45**	**HP1α + mutTERRA45**
Guinier analysis		
R_g (Å)	60.7 ±0.9	48.1 ±0.8
$q{\cdot}R_g$ range	0.79-1.14	0.60-1.00
I(0) (cm^{-1})	0.44 ±0.0044	0.24 ±0.0019
Fidelity	1.00	1.00
P(r) analysis		
R_g (Å)	69.4	48.1
D_{max} (Å)	293.9	184.9
q range	0.0131-0.3001	0.0125-0.3001
Quality estimate	0.46	0.46
Porod volume ($Å^3$)	207,772	70,013
Porod volume ×0.6 = MW (Da)	124,663	42,008

rather than a complex.

Guinier analysis of the HP1α-TERRA45 sample gave an R_g of 60.7 Å (**Figure A.78**), and $P(r)$ analysis gave a slightly larger R_g of 69.4 Å and a large D_{max} of 293.9 Å (**Table 5.9**, **Figure A.79**). The Porod volume was found to be 207,772 $Å^3$, which when multiplied by 0.6 to give MW results in 124,663 Da (**Table 5.9**). A 2:1 ratio of HP1α (dimer) to TERRA45 would yield an expected MW of 64,840 Da. Therefore, this ratio is unlikely to be representative of the complex formed in this sample. To examine the possible ratios of HP1α to TERRA45, different ratios and their expected molecular weights were calculated, and subsequently compared to the Porod volume-determined MW (**Table 5.10**). Within 15% of the expected molecular weight are the ratios of HP1α to TERRA45 of 2:4, 2:5, 2:6, 4:1, 4:2, and 4:3.

Because of the ambiguity in determining the ratio of HP1α to TERRA45, *ab initio* modelling with DAMMIN was performed with both $P1$ and $P2$ symmetry. Each set of structures gave elongated shapes, however the χ^2 value for the structure produced using $P1$ symmetry was slightly lower at 0.3229 compared with 0.4129 for $P2$ symmetry (**Figure 5.40**, **Figure 5.41**).

Due to the apparent conformational change undergone by HP1α when binding TERRA45 and the challenges faced fitting just the αHinge-TERRA45 complex to the model, determining the structure of the HP1α-TERRA45 complex remains difficult, and more discriminatory methods such as small-angle neutron scattering (SANS) contrast matching should be utilised.

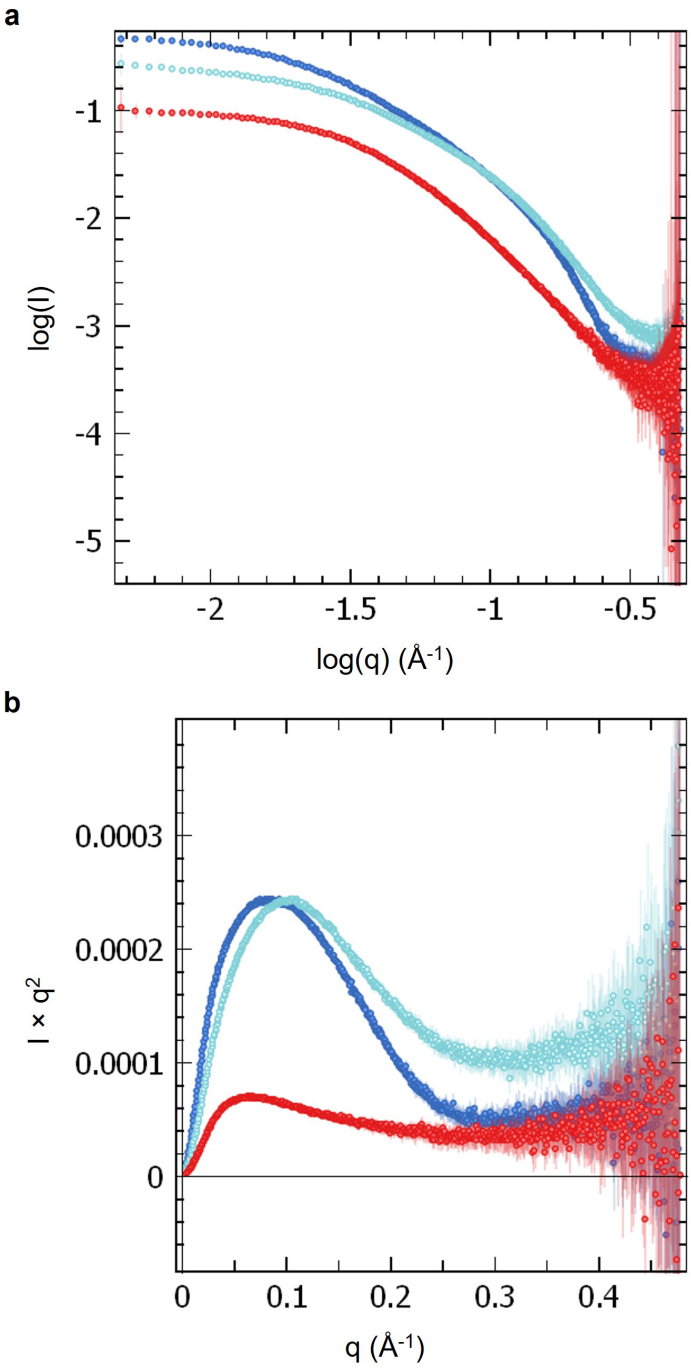

Figure 5.39: SAXS curves of HP1α WT plus TERRA45 and mutTERRA45. SAXS profiles of HP1α WT (red), HP1α plus TERRA45 (light blue), and HP1α plus mutTERRA45 (dark blue) presented as a) log of I versus log of q, and b) as a Kratky plot. Experiments conducted using 2.5 mg/mL (100 µM) of HP1α and/or 100 µM oligonucleotide in IB1 plus 10 mM TCEP at 22°C.

Table 5.10: Expected molecular weights of the HP1α-TERRA45 complex based on different molar ratios

Ratio of HP1α to TERRA45	Expected MW	Porod volume MW ratio to expected MW
2:1	64,839.72	1.92
2:2	79,653.52	1.57
2:3	94,467.32	1.32
2:4	109,281.12	1.14
2:5	124,094.92	1.00
2:6	138,908.72	0.90
4:1	114,865.64	1.09
4:2	129,679.44	0.96
4:3	144,493.24	0.86
4:4	159,307.04	0.78

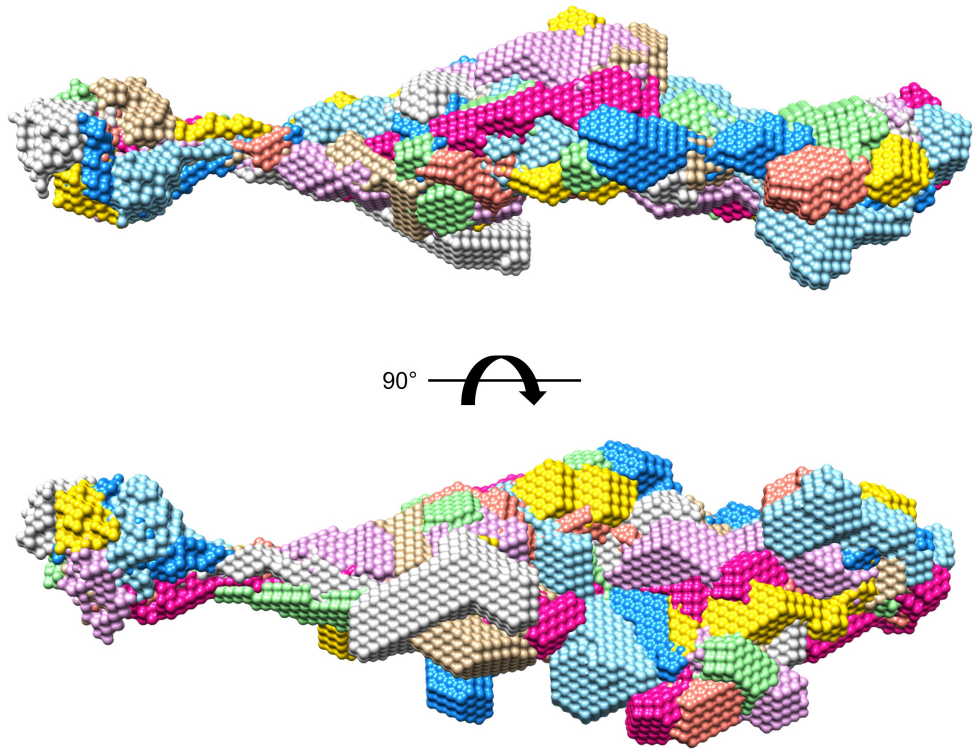

Figure 5.40: Computed models of HP1α-TERRA45 complex SAXS data with *P*1 symmetry. DAMMIN-generated models of HP1α WT-TERRA45 complex SAXS data based on *P*1 symmetry. Each of 10 models is shown in a different colour. Ensemble resolution 50 Å, final χ^2 against raw data 0.3229, mean value of NSD 1.596. The lower image is rotated forward 90°.

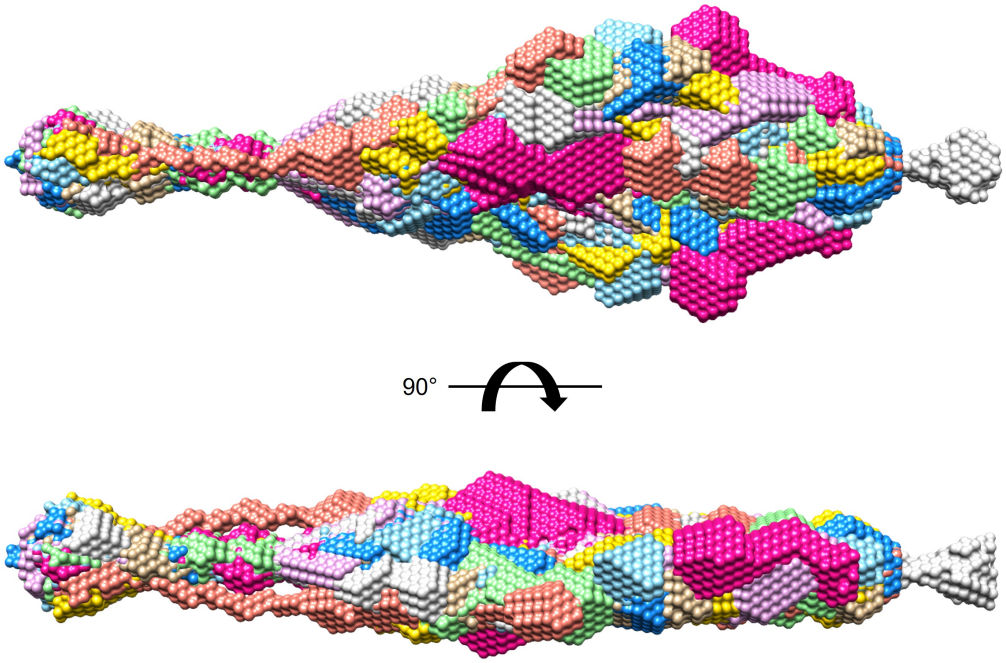

Figure 5.41: Computed models of HP1α-TERRA45 complex SAXS data with *P*2 symmetry. DAMMIN-generated models of HP1α WT-TERRA45 complex SAXS data based on *P*2 symmetry. Each of 10 models is shown in a different colour. Ensemble resolution 33 Å, final χ^2 against raw data 0.4129, mean value of NSD 1.257. The lower image is rotated forward 90°.

Table 5.11: SEC-SAXS structural parameters of HP1α Hinge+CSD+CTE plus TERRA45

Structural parameter	**Hinge+CSD+CTE plus TERRA45**
Guinier analysis	
R_g (Å)	55.8 ±1.1
$q{\cdot}R_g$ range	0.46-0.85
I(0) (cm^{-1})	0.33 ±0.0022
Fidelity	1.00
P(r) analysis	
R_g (Å)	58.4
D_{max} (Å)	260.7
q range	0.0092-0.3004
Quality estimate	0.60
Porod volume ($Å^3$)	107,949
Porod volume ×0.6 = MW (Da)	64,769
Expected MW (Da)	63,375 (2:2)
Ratio to expected MW	1.02

SAXS was also performed on TERRA45 with addition of HP1α Hinge+CSD+CTE mutant, which despite lacking the N-terminal CD and NTE, still binds and has specificity for TERRA45. Unfortunately, data were unable to be collected on the HP1α Hinge+CSD+CTE mutant alone due to its instability in solution at the concentrations measured. SAXS curves showed similar shapes to the HP1α WT plus TERRA45 samples (**Figure 5.42**), with a Guinier derived R_g of 55.8 Å (**Figure A.82**). $P(r)$ analysis showed a similar R_g of 58.4 Å and a D_{max} of 260.7 Å (**Figure A.83**, **Table 5.11**), both predictably smaller than the WT plus TERRA45 sample (**Table 5.9**). The molecular weight determined by the Porod volume was 64,769 Da, which is very similar to the predicted molecular weight of a Hinge+CSD+CTE dimer plus two TERRA45 molecules (16,874 Da ×2 + 14,814 Da ×2 = 63,375 Da, **Table 5.11**), indicating that the HP1α Hinge+CSD+CTE dimerises and each binds one TERRA45 unit.

Ab initio shape determination using DAMMIN was performed with both $P1$ and $P2$ symmetry, each of which showed the elongated shape of HP1α Hinge+CSD+CTE (**Figure 5.43**, **Figure 5.44**) and appeared to be similar to the HP1α WT plus TERRA45 shapes, indicating that the Hinge, CSD, and CTE of HP1α were similarly arranged and a large contributor to the shape of the complex. Again, due to the challenges faced fitting just the αHinge-TERRA45 complex to the model, fitting of structures to the Hinge+CSD+CTE plus TERRA45 model that yielded satisfactory χ^2 values was unavailing.

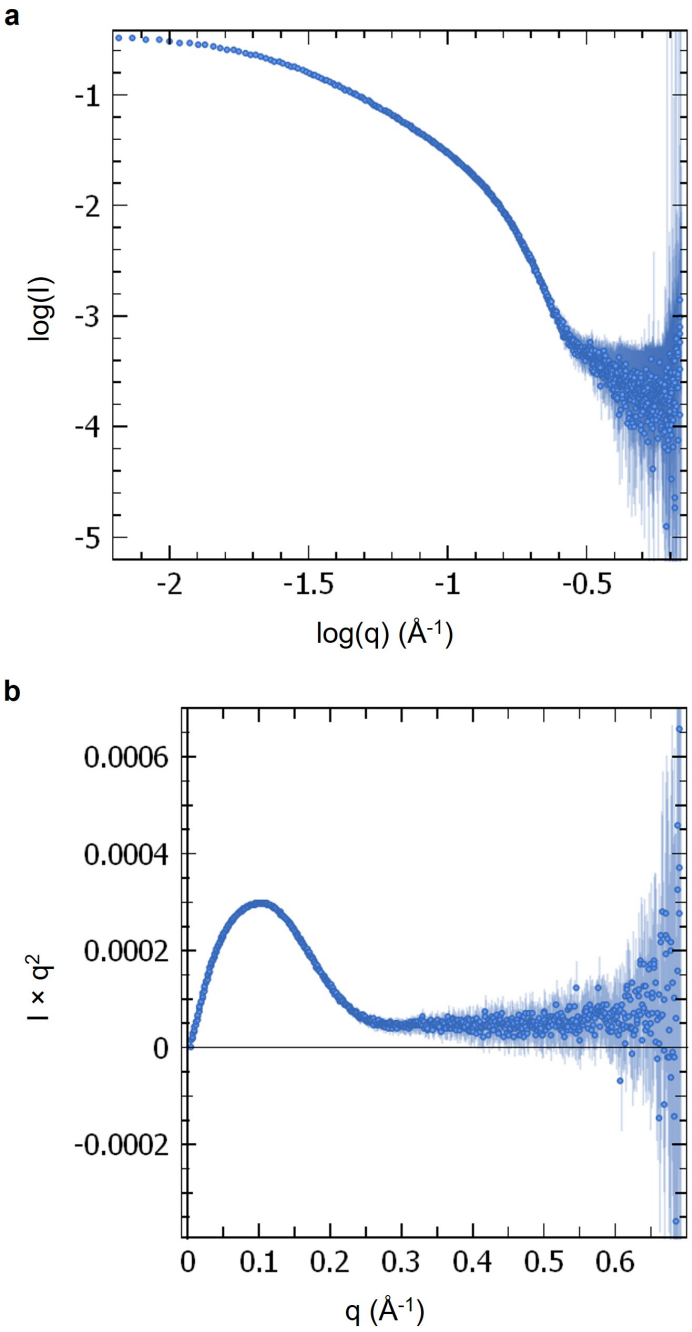

Figure 5.42: SAXS curves of HP1α Hinge+CSD+CTE plus TERRA45. SAXS profile of HP1α Hinge+CSD+CTE plus TERRA45 presented as a) log of I versus log of q, and b) as a Kratky plot. Experiments conducted using 2 mg/mL (120 µM) of HP1α Hinge+CSD+CTE and/or 120 µM TERRA45 in IB1 plus 10 mM TCEP at 22°C.

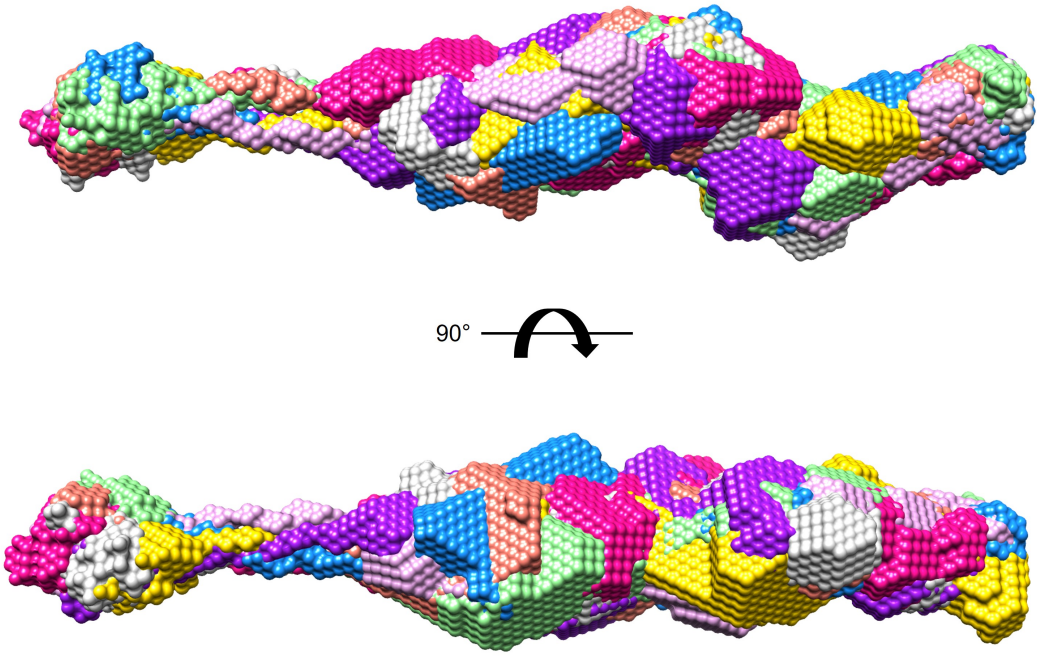

Figure 5.43: Computed models of HP1α Hinge+CSD+CTE-TERRA45 complex SAXS data with *P*1 symmetry. DAMMIN-generated models of HP1α Hinge+CSD+CTE-TERRA45 complex SAXS data based on $P1$ symmetry. Each of 10 models is shown in a different colour. Ensemble resolution 33 Å, final χ^2 against raw data 0.1982, mean value of NSD 1.123. The lower image is rotated backward 90°.

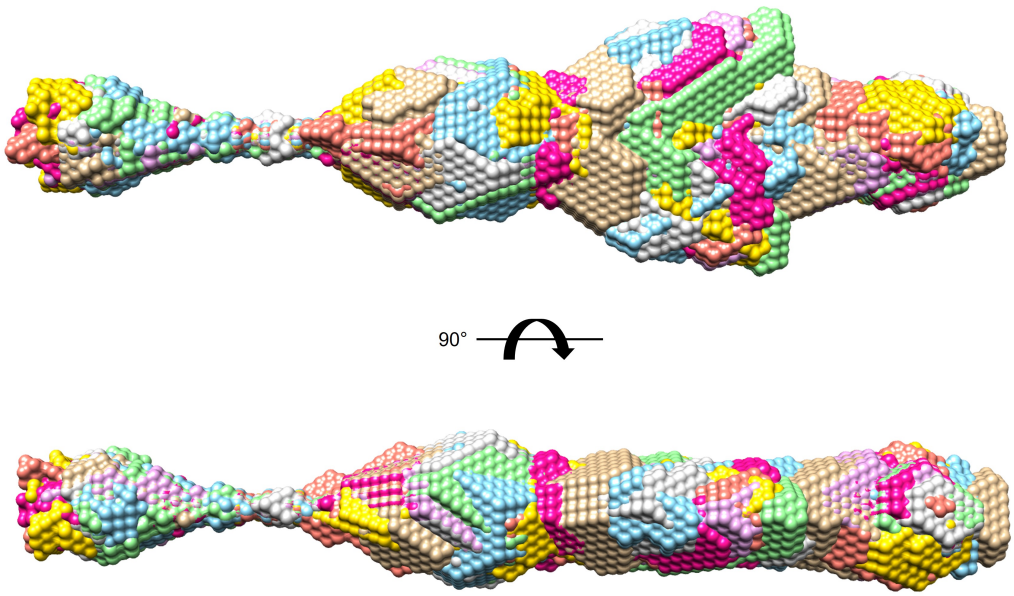

Figure 5.44: Computed models of HP1α Hinge+CSD+CTE-TERRA45 complex SAXS data with *P*2 symmetry. DAMMIN-generated models of HP1α Hinge+CSD+CTE-TERRA45 complex SAXS data based on $P2$ symmetry. Each of 10 models is shown in a different colour. Ensemble resolution 25 Å, final χ^2 against raw data 0.3523, mean value of NSD 0.749. The lower image is rotated backward 90°.

5.3 Discussion

High-resolution experimentally derived structures of HP1 proteins are unavailable, and therefore there is gap in knowledge to link the structure of HP1α to its crucial functions. Here, low-resolution structures have been generated from experimental data and compared with structure predictions of atomic resolution. The use of both the AlphaFold database and polypeptide prediction tool proved to be accurate to the solution-state SAXS data of HP1α in both monomeric and dimeric form, as well as the hinge alone.

The hinge domains of each HP1 protein are intrinsically disordered, and as such have not been resolved experimentally, with little structural information available at all. The molecular envelope determined by SAXS data of the HP1α hinge was in good agreement with the AlphaFold-generated structure (**Figure 5.4**, **Figure 5.5**), and also the circular dichroism spectrum (**Figure 5.6**), which showed that the αHinge is an extended, mostly unstructured polypeptide with a short α-helix. The presence of the α-helix in the hinge of HP1α was not supported by circular dichroism data. However, molecular dynamics simulations of the HP1α hinge do show transient helix formation (McIvor and Mercadante, personal communication, 2022). *In silico* molecular dynamics simulations that were validated by electron microscopy of the HP1γ hinge have shown a lack of secondary structure, and both elongated and kinked structures form, which highlights the flexibility of this intrinsically disordered domain (Velez et al., 2016). However, the differences in length and sequence of HP1α and HP1γ may make them incomparable (**Figure 3.1**). The data shown here of the hinge of HP1α further solidifies it as an intrinsically disordered domain, likely responsible for the conformational heterogeneity and dynamic nature of the entire protein. However, the function of the possible helix in the αHinge in nucleic acid binding and HP1α function should be explored further.

HP1α dimerises through one of the α-helices in its CSD, which has been shown by the crystal structure of the CSD (PDB: 3I3C, Li et al.), and also experimentally by mutation and molecular weight estimation by SEC (Brasher et al., 2000). This was confirmed by the monomeric HP1α Y168E, the molecular weight of which was determined by SAXS (**Table 5.2**). SAXS also showed this was a relatively flexible protein by Kratky plot, owing to the flexible hinge. The structure of monomeric HP1α from the AlphaFold database proved to be a good fit with the structure of HP1α Y168E generated from SAXS data (**Figure 5.9**, **Figure 5.10**), although the AlphaFold

structure has low confidence in the position and structure of the hinge and NTE (**Figure A.38**). Unique to the AlphaFold-predicted HP1α structure is the close distance between the CD and CSD, which are normally depicted in models as separate extended entities (Kumar and Kono, 2020). The CD and CSD are arranged in such as way that the β-sheets of each of the globular domains form an interface with one another (**Figure A.38**), that has a negative but low ΔG value for formation. In terms of β-sheet interactions in HP1α, the crystal structure of the CSD (PDB: 3I3C) contains an additional dimerisation surface on the β-sheet side of the CSD (**Figure 4.15**). Mass spectrometry of HP1α crosslinked by BS3 (bis(sulfosuccinimidyl)suberate) followed by trypsin digestion has shown extensive links within the protein, and indeed one between the CD and CSD (Larson et al., 2017). In addition to this, the *Schizosaccharomyces pombe* HP1 homologue Swi6 has been shown to tetramerise via CSD-CSD interactions as well as CD-CD interactions (Canzio et al., 2011). Due to the high amino acid sequence and structure conservation between the CD and CSD of HP1α (Aasland and Stewart, 1995), a weak interaction between β-sheets of CD and CSD domains intramolecularly may indeed occur as in the AlphaFold model, when not bound to nucleosomes.

The dimeric structure of WT HP1α is proposed here in *trans* and *cis* conformations, with monomers in contact with one another through an α-helix. While the DAMMIN-generated molecular envelope visually agrees with the *trans* conformation, comparison of *cis* and *trans* structures using FoXS determined that both similarly match the SAXS data of dimeric HP1α (**Figure 5.15**). In support of the *trans* conformation, mass spectrometry data has shown the most extensive contacts within HP1α occur between the hinge and CTE (Larson et al., 2017), which is possible in the *trans* conformation proposed here. Of the limited structural data, a low-resolution structure of HP1α bridging a dinucleosome complex has been determined using cryogenic electron microscopy (cryo-EM, Machida et al. 2018). However, the low resolution does not allow for any assumptions to be made on the mode of dimerisation. It is conceivable that each dimerisation conformation is possible *in vivo*, with different conformations applicable depending on binding partner or nucleosome arrangement, such as when binding H3K9me3 within the same nucleosome or between nucleosomes.

The D_{max} of SEC-SAXS derived HP1α measured here as 117 Å (**Table 5.2**) is in relative agreement with previous SAXS data of HP1α published by Larson et al. (2017) as 129 Å. Larson et al. (2017) also showed that phosphorylation of the NTE (S11-14) by CK2 results in an extended conformation which promotes HP1α oligomerisation. NTE-phosphorylated HP1α was

also measured by SAXS by Larson et al. (2017), which yielded a D_{max} of 222 Å, comparable to the data generated here on the HP1α NTE phosphorylation mimic S11-14E of 207 Å (**Table 5.4**). As discussed in **Chapter 3**, the HP1α NTE phosphorylation mimic S11-14E causes decreased binding to TERRA which may be influenced by the charge of this mutant, but could also be influenced by this extended conformation of HP1α affecting the G4 binding/docking site.

In addition to the lack of full-length structures available of HP1α, there are few crystal and NMR structures of G4s larger than one G4 unit, despite the relevance of longer multimeric G4s to biological functions (Kolesnikova and Curtis, 2019). To represent TERRA45, two copies of the crystal structure of a 22 nt intramolecular parallel G4 formed by telomeric DNA (PDB: 7KLP, Li et al. 2021) is used, as it shares the G-rich sequence of TERRA. It was determined that an offset structure of two G4s better fit the SAXS-derived molecular envelope of TERRA45 than a two-G4 structure in which they are perfectly aligned (**Figure 5.25**). Computer simulations of parallel G4 conformations have shown that various stacking assemblies are possible and energetically favourable (Kogut et al., 2019). Recent SAXS data (Monsen et al., 2021) collected on longer lengths of anti-parallel telomeric DNA are consistent with the previously reported "beads on a string" model of G4 multimers (Yu et al., 2012a; Kar et al., 2018). However, Monsen et al. (2021) showed that when the G4 "beads" were present as two G4 units or more, they lose the compact Kratky profile of a single compact G4. In the case of TERRA45, this two-G4 structure still maintains the compact Kratky profile (**Figure 5.21**). Therefore, TERRA45 is unlikely to conform to the "beads on a string" structure at this length, where no stacking between G4 units would occur. An unexpected feature of the DAMMIN-generated molecular envelope of TERRA45 is the thin protrusions from the sides (**Figure 5.22**), which could be due to the presence of transient unfolding or alternate folding causing free strands or longer loops. The possible offset structure of TERRA45 was also somewhat unexpected. A possible change in alignment of TERRA from the offset topology to the aligned topology should be investigated to determine if this was a contributing factor to the difficulty in fitting the offset structure to complex models. Interestingly, the SAXS data of mutTERRA45 was extremely similar to that measured of TERRA45 (**Figure 5.26**). Circular dichroism also shows only very subtle differences between the two (**Figure 3.2a**). However, the affinity displayed by HP1α for TERRA45 over mutTERRA45 is very pronounced (**Figure 4.2b**). Given the similarity of the SAXS-generated molecular envelopes of TERRA45 and mutTERRA45, the differences in binding between the two may in fact be due to the presence of the guanine base in the backbone

of the G4 in addition to the parallel topology.

When complexes were formed between TERRA45 and HP1α, a more pronounced peak and a shift of the peak maximum to the left in the Kratky plot occurs (**Figure 5.29**, **Figure 5.33**, **Figure 5.39**), which indicates a less flexible complex compared to its counterparts. It may also be indicative of the elongated complexes which are also shown by the D_{max} values and DAMMIN-generated shapes. The HP1α NTE-phosphorylation mimic also showed a slightly more elongated shape compared to unphosphorylated HP1α (**Figure 5.20**), agreeing with the proposed "open" conformation of N-terminally phosphorylated HP1α (Larson et al., 2017), which increases its affinity and specificity for H3K9me3 within nucleosomes (Nishibuchi et al., 2014). Addition of G4 to form an elongated HP1α may be an alternative method of aiding HP1α to bind H3K9me3 with more specificity to effectively bridge nucleosomes in heterochromatin. Certainly, as shown in **Section 3.2.6**, the addition of TERRA45 appeared to result in a greater amount of HP1α binding to H3K9me3 peptide, which should be investigated further.

Due to the apparent conformational change of HP1α when complexed with TERRA45, fitting of atomic structures was unable to be ameliorated. Further experimentation is needed to assess the structure of HP1α in complex, which may be achieved by the use of SANS contrast matching. Contrast matching takes advantage of the different neutron scattering length densities of different molecules such as proteins, nucleic acids, carbohydrates, and lipids, due to their atomic compositions. The scattering length density of each of these molecules can be matched by altering the ratio of H_2O to D_2O to effectively make a molecule or a component have the same neutron scattering properties as the solvent. Therefore, at different concentrations of D_2O, TERRA can be contrast matched, allowing for analysis of HP1α structure alone when bound in the complex. Work has been done to obtain these data, however, due to the high concentrations of protein and nucleic acid required for SANS, optimisation is still required to combat aggregation and precipitation of samples, particularly when at high ratios of D_2O.

Chapter 6

Conclusions and future directions

6.1 Summary of results

This study scrutinised the interaction between heterochromatin regulator HP1α and parallel G4 TERRA, establishing the means of interaction, exploring the specificity of HP1α for the TERRA G4, and investigating their structures in solution.

Despite the sequence conservation among HP1 paralogs, high-affinity binding by HP1α to TERRA occurred, with no binding shown by HP1β, and less binding shown by HP1γ. One of the unique features of HP1α is the string of serine residues within its NTE that are phosphorylated to enhance the binding between the CD and H3K9me3. Mutations to mimic NTE phosphorylation decreased the ability of HP1α to bind TERRA, as did phosphorylation mimic mutations made in the hinge. Two previously determined nucleic acid binding patches within the hinge, upon mutation, also reduced TERRA binding, and an additional charged patch unique to HP1α had a similar effect. Mutation of two or more patches completely abolished binding to TERRA, while mutation of all three hinge charge patches *in vivo* did not drastically alter heterochromatin foci. Despite this, mutation of an H3K9me3 binding site did not alter HP1α-TERRA binding, and TERRA did not negatively affect the binding of HP1α to H3K9me3. When the G4 structure of TERRA was altered through mutation or omission of G4-stabilising cations, binding to HP1α was reduced. When parallel and anti-parallel DNA G4s were tested, the preference of HP1α for parallel G4s remained, showing that the structure of the parallel G4 was integral for binding. HP1α also did not bind double-stranded DNA, i-motif DNA, tRNA, or unstructured DNA. How-

ever, the disordered hinge of HP1α alone bound to each nucleic acid structure tested, lacking the specificity of the full-length protein. The globular CSD was found to provide specificity to the HP1α hinge-TERRA interaction through a hydrophobic α-helix docking mechanism. Low-resolution molecular envelopes of full-length HP1α and the hinge were determined by SAXS, and it was found that addition of TERRA results in a more ordered complex.

6.2 HP1α charge in G4 binding

The positively charged patches within the hinge of HP1α were shown to be imperative for binding to TERRA, and indeed have been shown to be required for DNA and RNA binding. Binding to TERRA was largely by the HP1α paralog, with minimal complex formation with HP1γ, while Deng et al. (2009) also showed some binding to TERRA by HP1β along with HP1α. Certainly, the lack of the charged patch (68-72KKYKK) at the N-termini of the hinges of HP1β and HP1γ (**Figure 3.1**) likely contributes to the lack of or decrease in G4 binding ability, while the HP1α hinge possesses the highest isoelectric point (Qin et al., 2021) and is the longest in sequence, allowing for more electrostatic interactions to take place with nucleic acids. An HP1α mutant with two of these charged patches mutated (A2;3 - 89-91KRK>AAA;104-106KKK>AAA) resulted in a structure with a smaller radius of gyration, which may also be a contributing factor to G4 binding. To determine if these factors in the hinge contribute to the unique binding of HP1α to TERRA, the hinges of each paralog, and indeed other HP1 variants, can be exchanged and tested for binding.

In this study, the mouse HP1α is used, which has 97.4% identity and 99.0% similarity (Waterman and Eggert, 1987) to human HP1α (**Figure A.2**), while HP1β and HP1γ share 100% identity between human and mouse amino acid sequences. HP1 or HP1-like proteins are present in fungi, plants, invertebrates, and vertebrates (Schoelz and Riddle, 2022). These eukaryotes also possess G-rich telomeres, with the TTAGGG repeat common across vertebrates (Meyne et al., 1989). It is therefore of interest to explore if the same HP1-TERRA G4 binding is present across taxa with other HP1 variants, which would indicate its importance in heterochromatin function. *S. pombe* Swi6 and *Drosophila* HP1a should also be investigated for their propensity to bind G4s, as they also possess at least three positively charged patches in their hinge domains (**Figure 6.1**). However, the telomeric repeats of *S. pombe* and *Drosophila* differ significantly from vertebrates (Louis, 2002; Sepsiova et al., 2016), so the propensity for these HP1 variants to bind G4s may

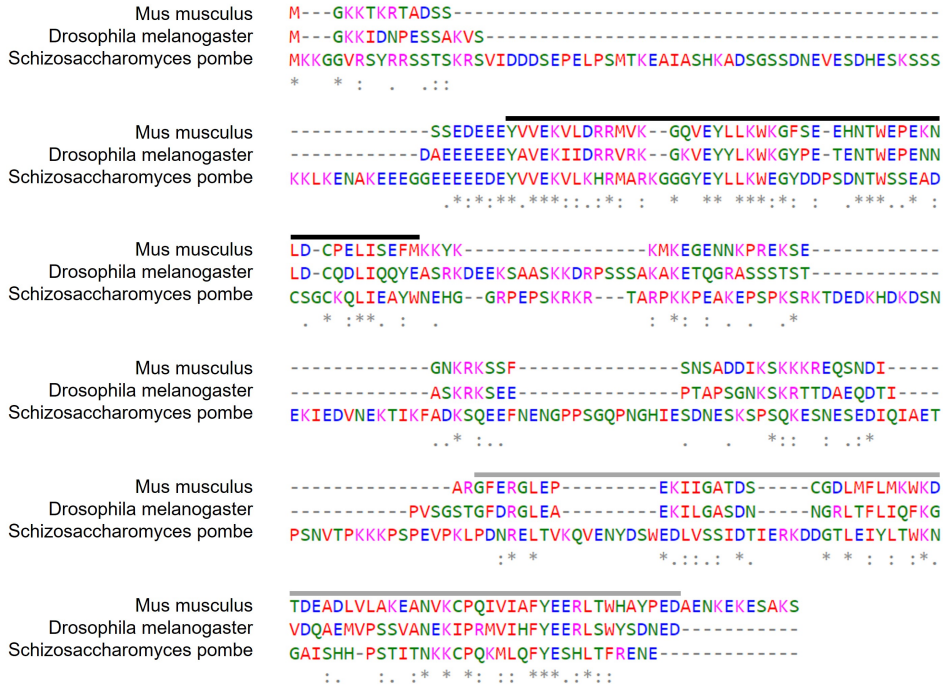

Figure 6.1: Alignment of HP1 variants. Amino acid sequences of *Mus musculus* HP1α, *Drosophila melanogaster* HP1a, and *Schizosaccharomyces pombe* Swi6. The CD and CSD boundaries of HP1α are shown by black and grey lines, respectively. An asterisk (*) indicates a fully conserved residue, a colon (:) indicates conservation of a residue with strongly similar properties, and a period (.) indicates conservation of a weakly similar residue.

highlight other roles for HP1-G4 binding in these species.

Interestingly, the charged hinge was not the only domain of HP1α to influence G4 binding. Deletion of the NTE of HP1α also resulted in a weakened affinity for TERRA. The NTE of HP1α also contains positively charged amino acids (**Figure 3.19**), the elimination of which (by deletion of the NTE) may be the cause of the decrease in binding. The phosphorylation of proximal NTE serine residues may undermine this effect by introducing an opposing negative charge, also resulting in a decrease in TERRA binding by HP1α.

6.3 Nucleic acid structure-specific binding by HP1α

The parallel G4-specific binding of HP1α to TERRA alludes to a mechanism of telomeric heterochromatin-specific recruitment of HP1α through the hinge and stabilisation by the CSD. This binding mechanism is similar to that of RNA helicase RHAU and parallel G4s, whereby

Table 6.1: Putative QGRS sequence found in MSR forward DNA and RNA.

Position	Length (nt)	Sequence	G-score
72 (DNA)	24	GGCGAGGAAAACUGAAAAAGGUGG	10

an α-helix covers the G-tetrad at one end, while electrostatic interactions are formed between the positively charged amino acids and negatively charged phosphate groups for loop-binding (**Figure 1.9**, Heddi et al. 2015). This top-stacking mode of interaction explains the parallel G4 specificity that both RHAU and HP1α possess due to this G-tetrad accessibility when no lateral or diagonal loops are present.

Although the binding to short nucleic acid sequences in this study was shown to be highly specific for the parallel G4 TERRA, HP1α has been shown to bind additional nucleic acids (Muchardt et al., 2002; Meehan et al., 2003; Maison et al., 2011; Mishima et al., 2013; Ryan and Tremethick, 2018). Other nucleic acid structures which HP1α binds could be responsible for its recruitment elsewhere. For example, HP1α is targeted to centromeric heterochromatin by fMSR RNA, specifically, when HP1α is SUMOylated (Maison et al., 2011). Interestingly, HP1α only showed minimal binding to a short hairpin from the fMSR. While it is likely that SUMOylation of the HP1α hinge is required for this binding, perhaps the full-length fMSR sequence contains additional structural moieties responsible for HP1α binding. The Quadruplex forming G-Rich Sequence (QGRS) Mapper (Kikin et al., 2006) predicted a single putative 24 nt G4 with a low G-score of 10 within the 234 nt fMSR RNA sequence (**Table 6.1**). The higher the score, the higher the likelihood to form a stable G4, with a maximum possible score of 105. To determine if this sequence may form this unlikely (due to the large gap/loop between the second and third G-tracts) G4 or another possible structure, this 24 nt fragment should be assessed for nucleic acid structure formation before its binding to HP1α is evaluated. However, this interaction between HP1α and fMSR for centromeric localisation may be exclusively dependent upon the SUMOylation of the HP1α hinge.

6.4 HP1α post-translational modifications

In addition to determining the effect of SUMOylation on the binding of HP1α to fMSR, the effect of SUMOylation on the HP1α hinge should also be investigated for its influence on binding to

Figure 6.2: SUMOylation sites in the HP1α hinge. SUMOylation sites determined by MS by Maison et al. (2011) are shown in blue, with patches mutated in this study numbered and underlined with their substituted residues shown above the sequence.

TERRA by performing *in vitro* SUMOylation (Werner et al., 2009). If SUMOylation prevented the binding of the HP1α hinge to TERRA, this could indicate alternative HP1α RNA-mediated targeting mechanisms at the centromeres and telomeres, providing the specificity for centromeric targeting. This G4-binding inhibition by SUMOylation may be likely given the SUMOylation sites are within two of the three charged patches within the HP1α hinge (**Figure 6.2**, Maison et al. 2011), and the size of the SUMO-1 modification (shown in comparison to the HP1α AlphaFold model, **Figure 6.3**) may sterically inhibit G4 binding. Contradictory to this theory, Lin et al. (2016) showed that when SUMOylation of HP1α is globally increased (by knockdown of long SUMO-specific protease 7, SENP7L) in the MCF7 breast cancer cell line, HP1α binding to TERRA is enhanced. In addition to this, RNAse H treatment reduced SUMO-HP1α enrichment at both centromeres and telomeres (Lin et al., 2016). This highlights the need to determine the ability of SUMOylated HP1α to interact with TERRA. It should also be determined if SUMOylated HP1α binds and possesses specificity towards fMSR structures under these conditions. SUMOylation of HP1α may cause a change of conformation of the protein, which may in turn affect which structure of nucleic acid it binds, therefore circular dichroism spectroscopy and SAXS can be used to assess this.

HP1α is not only SUMOylated but has 48 documented PTMs spanning across each domain (**Figure 1.20**). These PTMs, particularly those in the hinge and CSD, should also be investigated for their effect on TERRA G4 binding to determine if they are able to regulate this interaction. Specifically, mutation of K154 in the HP1α CSD affects specificity for TERRA but has been shown to be ubiquitinated, SUMOylated, or dimethylated (Sales-Gil and Vagnarelli, 2020). Therefore, the SUMOylation of K154 may also affect TERRA binding and/or specificity due to the imposition by the SUMO-1 structure (**Figure 6.3**). The role of SUMOylation and dimethylation of this residue are unknown; however, ubiquitination by the Rad6 E2 ubiquitin-conjugating enzyme promotes the degradation of HP1α to enable DNA repair (Chen et al., 2015).

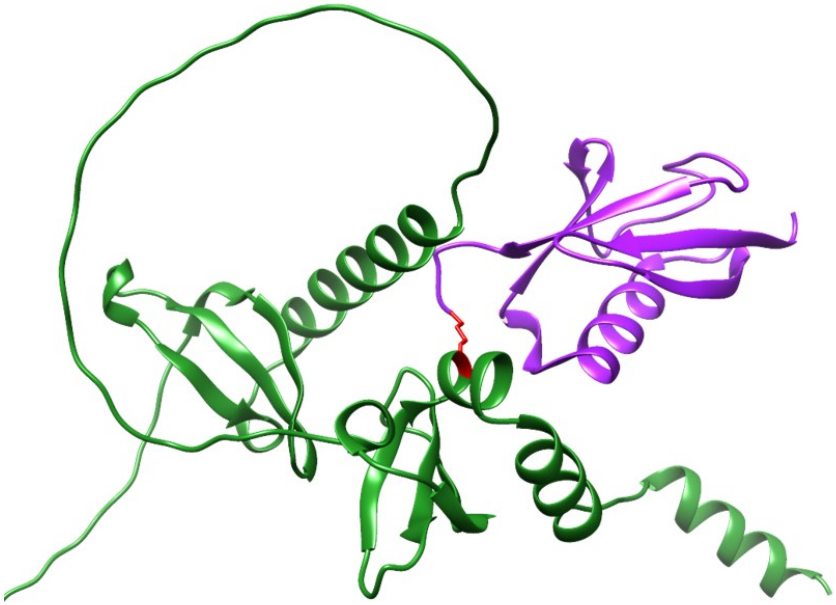

Figure 6.3: SUMO-1 shown in comparison to HP1α. SUMO-1 (PDB:2VRR, Knipscheer et al. 2008) is shown in purple alongside the AlphaFold model of HP1α (AF-Q61686-F1) shown in green, with K154 highlighted in red. The N-terminus of HP1α extends to the left, out of frame.

6.5 N-terminal phosphorylation effects on HP1α function

The N-terminal phosphorylation mimic of HP1α was partially inhibited from binding to TERRA, potentially due to negative charge and the extended conformation that it adopts. N-terminal phosphorylation may again be an alternative mechanism of HP1α recruitment to different regions of heterochromatin, whereby it is proposed that N-terminally phosphorylated HP1α is recruited to regions already rich in H3K9me3 without requiring TERRA or fMSR RNA for its *de novo* targeting (**Figure 6.4**). Moreover, addition of TERRA to HP1α was also shown to result in an elongated complex, which may mimic the extended conformation adopted by N-terminally phosphorylated HP1α (Larson et al., 2017) for H3K9me3 binding, which may aid in HP1α binding to nucleosomes in telomeric heterochromatin. To test these different recruitment models, nucleosomal arrays created with unmodified histones or H3K9me3 in the presence and absence of TERRA or fMSR RNA can be assessed (Ryan and Tremethick, 2018) to determine affinity of HP1α for nucleosomes and also the mobility/compactness of nucleosomal arrays. To further scrutinise the model proposed in **Figure 6.4**, further *in vivo* experiments can be performed. Hiragami-Hamada et al. (2011) showed that unphosphorylatable HP1α exhibits defective heterochromatin localisation *in vivo*, while Maison et al. (2011) showed that fMSR RNA localised

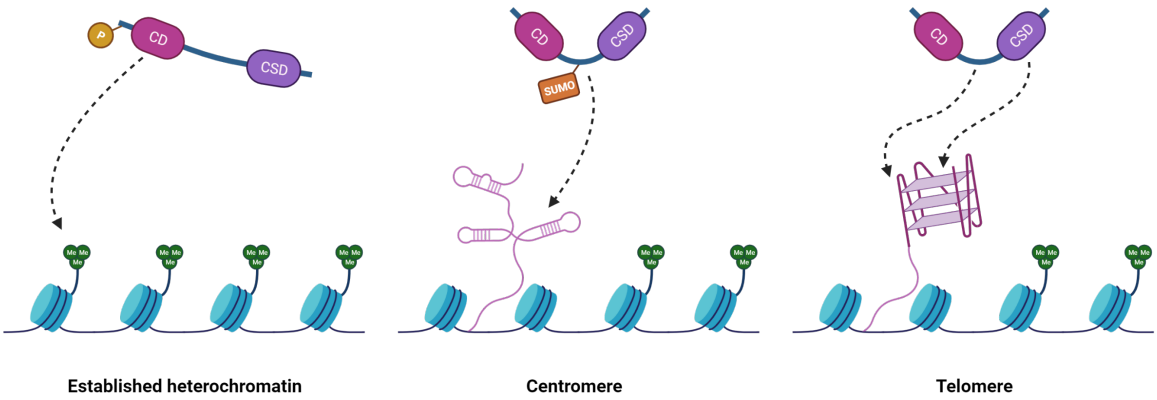

Figure 6.4: Model of HP1α heterochromatin localisation. In the model proposed, extended N-terminally phosphorylated HP1α is poised to localise to regions of already established heterochromatin through the CD-H3K9me3 interaction. When HP1α is SUMOylated, it binds through its hinge to fMSR RNA to localise to centromeres. Unmodified HP1α binds to the TERRA G4 through its hinge with CSD recognition to localise to telomeres.

to heterochromatin at centromeres, that SUMOylated HP1α interacts with this RNA, and that the hinge and SUMOylation were required for HP1α targeting to pericentric heterochromatin. For assessing the importance of SUMOylation on the hinge of HP1α for localisation to the centromeres, FLAG-tagged mutants of HP1α can be created and transfected. For example, HP1α K86R, which is unable to be SUMOylated on this residue but still maintains the positive charge, can be expressed *in vivo* and assessed for localisation, paired with FISH for fMSR RNA. Additionally, chromatin immunoprecipitation sequencing (ChIP-seq) could be utilised to find the sequences that HP1α mutants (that cannot be phosphorylated on the NTE or SUMOylated on the hinge) localise to. This can be paired with the use of siRNA for centromeric and telomeric RNA to see differences in abundance of centromeric and telomeric sequences immunoprecipitated with WT HP1α.

N-terminal phosphorylation of HP1α and nucleic acid binding promotes LLPS, which may mediate heterochromatin formation through sequestration of chromatin components into functional phase-separated droplets (Larson et al., 2017). Phase-separation-driven heterochromatin compartmentalisation is a recent hypothesis, but Strom et al. (2017) showed that *Drosophila* HP1a displays phase-separation properties in early *Drosophila* embryos. Phase-separation is unique to the HP1α paralog in mammals (Larson et al., 2017) and found to be dependent upon the basic amino acids in the HP1α hinge (Keenen et al., 2021; Qin et al., 2021). Interestingly, RNA

G4s themselves have also been shown to trigger phase separation as nucleation points both *in vitro* and *in vivo* (Fay et al., 2017; Zhang et al., 2019). It is therefore of interest to determine if the addition of TERRA or other parallel G4s results in the increased propensity of HP1α to form phase-separated droplets. This can be tested by light microscopy of HP1α-G4 mixtures *in vitro*. However, HP1α-mediated LLPS is a quickly evolving field (Hansen et al., 2021; Singh and Newman, 2022), and while phase separation has been demonstrated in *Drosophila* embryos, unlike mammalian HP1α, *Drosophila* HP1a LLPS does not require N-terminal phosphorylation (Strom et al., 2017). Therefore, more work needs to be done to address the *in vivo* existence and relevance of these properties of mammalian HP1α.

6.6 HP1α structure and conformational changes

High-resolution structures of HP1α alone and in complex with a G4 remain elusive. The structure of HP1α in complex with TERRA may provide insights into parallel G4-binding proteins that allow for the design and targeting of G4 structures for drug applications (Shu et al., 2022). For example, a core-extended naphthalene diimide (binds G4s with high affinity and stabilises through stacking interactions with G-tetrads) molecule, simply referred to as "**1**", greatly stabilises G4s within human oncogene promoters in melanoma cells. More specifically, treatment with ligand "**1**" down-regulated the KIT proto-oncogene and the anti-apoptotic BCL2 gene due to stabilising the G4s formed within the promoter DNA (Recagni et al., 2019). This shows that ligands designed to target G4s can be used to alter the expression of oncogenic genes.

The structure of HP1α is also important to gain insight into its behaviour and stabilising properties within the heterochromatin structure. In determining solution structures, due to an apparent conformational change in HP1α, complexes with TERRA were unable to be resolved. In order to analyse components of complexes in solution, contrast matching can be used in SANS (as discussed in **Chapter 5**). Deuteration of HP1α itself (which has been performed by the National Deuteration Facility, Australia) also allows for addition of auxiliary proteins to the complex which will have a different match point to the deuterated protein. For example, SANS can be performed to analyse the structure of HP1α in complex with both a G4 and histone H3 to determine its behaviour when bound in heterochromatin.

To further analyse HP1α and HP1α-G4 complexes, X-ray crystallography, cryo-EM, and NMR

should be used to determine atomic resolution structures to begin targeting these assemblies. Crystal trials have been performed on HP1α alone and in complex with TERRA, but unfortunately no crystals were obtained, likely due to the flexibility and conformational heterogeneity of HP1α. NMR of ^{15}N labelled HP1α full-length or Hinge+CSD+CTE with and without G4 will confirm the apparent conformational change of HP1α upon binding TERRA, and may also aid in determining a high-resolution structure of HP1α. It will also be of interest to study the structure of TERRA45 in NMR, as structures of multimeric TERRA G4s have not been experimentally determined. A recent study was able to determine low-resolution structures of HP1α-bound dinucleosomes by cryo-EM to assess how HP1α bridges nucleosomes (Machida et al., 2018). This experiment should be expanded upon with the addition of TERRA to determine how this RNA G4 influences HP1α dinucleosome binding and bridging to ascertain if RNA G4s can stabilise chromatin by binding HP1α.

In lieu of an atomic resolution structure of HP1α and TERRA, a model has been created using the structural and *in vitro* interaction data. The proposed model of this HP1α interaction with TERRA (**Figure 6.5**), whereby the disordered hinge is available to interact with the phosphate backbone of TERRA and the first CSD helix stabilises the face of the G4, allows for other essential HP1α functions to occur. The CD is free to interact with H3K9me3, while the CSD may still be able to bind the PXVXL motif. Conformational changes including movement of the disordered regions of HP1α and changes in the positions of the CD and CSD relative to each other may occur, and future experiments should shed light on these possibilities.

6.7 G4s in heterochromatin

Another question that remains is whether G4s are able to form within heterochromatin, either with the association of RNA G4s or the formation of a G4 within the complexed DNA strands. *In vivo* work with the G4-binding BG4 antibody (which detects both DNA and RNA G4s) has shown foci within nuclei and cytoplasm of human cells, some of which are present throughout the cell cycle in the nucleus, but in their highest abundance during S-phase (Biffi et al., 2013, 2014). These foci were present throughout DAPI-stained nuclei, with no apparent enrichment in euchromatin (Biffi et al., 2013). However, recent papers describing the detection of G4s by BG4 ChIP or immunofluorescence show that G4 folding is coupled to euchromatin marks and RNA polymerase II localisation (Shen et al., 2021; Komůrková et al., 2021). Contradictory

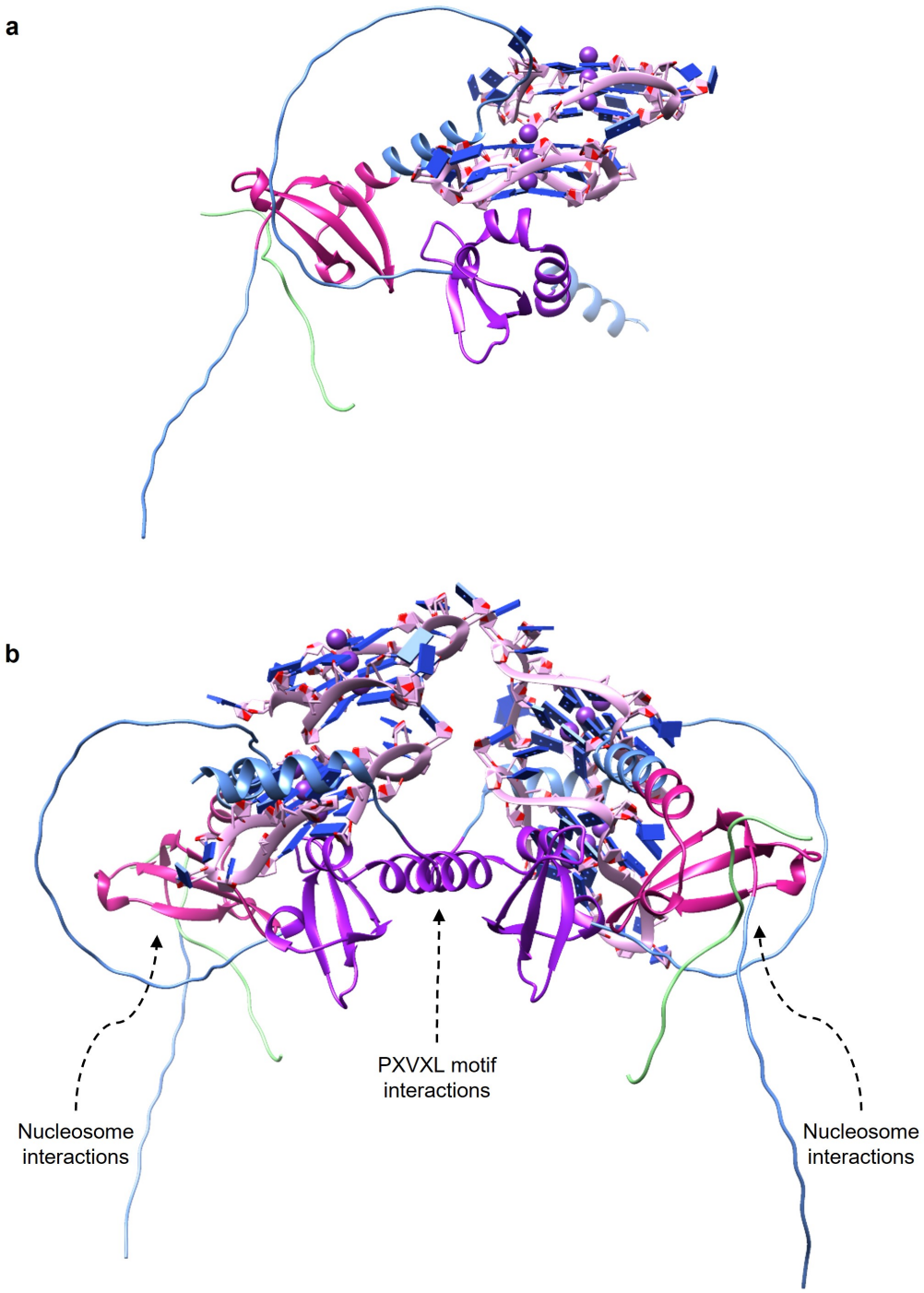

Figure 6.5: Proposed model of TERRA G4s interacting with HP1α. a) The Alphafold model of HP1α (AF-Q61686-F1) in blue with the CD coloured in magenta and the CSD in purple. The offset model of TERRA (2 × PDB:7KLP) is shown in light pink, contacting the CSD with the hinge available for interaction. An H3K9me3 peptide (PDB:2RVN, Shimojo et al. 2016) is shown in green, interacting with the CD of HP1α. b) The above proposed model displayed with HP1α in the *trans* dimer conformation, showing the CSDs available for PXVXL motif interactions.

to this is also the co-localisation of constitutive heterochromatin mark H3K9me2 with G4s (Komůrková et al., 2021), while Zyner et al. (2022) showed that G4s are associated with bivalent marks (H3K4me3/H3K27me3) on promoters in stem cells. Although heterochromatin tightly complexes nucleosomes, linker DNA may still be able to form G4s. Further work will need to be done to assess G4 localisation; however, the global view of G4s should be reassessed to focus on the roles and localisation of each G4 independently. It will be essential to identify if HP1α interacts and co-localises with G4s *in vivo*, although alternatively to BG4, RNA-FISH may be used to detect TERRA. Additionally, *in situ* proximity ligation assays can also be used to determine if HP1α is in direct proximity to TERRA or other G4s. Mutants of HP1α that affect G4 binding and specificity should be assessed to determine their effect on co-localisation with TERRA at telomeres.

6.8 G4 structures for genomic regulation

The interaction between HP1α and TERRA and the previous results showing the binding of HP1α to parallel DNA G4s (Roach et al., 2020) raises the possibility that HP1α could interact with G4s in promoters to control their expression. Therefore, how HP1α and DNA parallel G4s function in both euchromatin and heterochromatin *in vivo* should be investigated to determine if HP1α G4 binding affects silencing of genes. Shown in previous work with HP1α, although a distinct preference for the parallel G4 topology over the anti-parallel topology is displayed, HP1α does not bind each parallel G4 with the same affinity (Roach et al., 2020). Differences in the loop base composition and length, number of tetrads, and multimerisation of G4 units all add complexity to the G4 structure. Therefore, the binding of HP1α to different parallel G4s may have different consequences *in vivo*. In terms of parallel G4 DNA in promoter regions, HP1α should be assessed for its effect on gene silencing through heterochromatin via *ex vivo* luciferase assays. Parallel G4-forming promoter sequences can be used, and HP1α WT and hinge mutants can be transfected into cells to determine their effect on transcription.

An interesting case to study the effect of HP1α G4 DNA binding is with the BCL2 gene. The Bcl-2 protein inhibits apoptosis, and as such, the over-expression of the BCL2 gene is utilised by cancers to evade programmed cell death (Yip and Reed, 2008). Upstream from the BCL2 P1 promoter is a G-rich sequence shown to form multiple G4s, the putative purine-rich 39 nt (Pu39) G4 and an additional 28 nt P1G4 (Onel et al., 2016). When each of these G4-forming

sequences are cloned upstream of a luciferase reporter, expression from the BCL2 promoter can be assessed upon mutation of G-tracts. When the Pu39 G4 is mutated expression decreases, while when P1G4 is mutated expression increases, indicating that each G4 structure is affecting expression differently (Onel et al., 2016). These structures and their binding capacity for HP1α should be investigated, and the effect of HP1α binding to these promoter regions on expression should be assessed to determine if HP1α can affect gene expression by binding G4s.

In terms of the effect of HP1α binding to TERRA, further work *in vivo* is required to elucidate the effect of this interaction on telomeric heterochromatin function. When investigating TERRA *in vivo*, Deng et al. (2009) used a siRNA to deplete endogenous TERRA to assess the effect on telomere aberrations and associated proteins at telomeres. Depletion of TERRA by siRNA could also be used to assess the effect on HP1α localisation to telomeres and co-localisation with telomere-associated shelterin components. It is hypothesised that TERRA depletion would cause a decrease in HP1α at telomeres. The effects of permanent localisation of HP1α to telomeric heterochromatin via a molecular tethering mechanism has been described (Chow et al., 2018), whereby HP1α is artificially enriched at telomeres *in vivo* through the fusion of shelterin component TRF1. The manufactured enrichment of HP1α-TRF1 at telomeres resulted in locally increased H3K9me3 (relative to total H3). In the presence of the telomerase RNA template (TERC) when TRF1 was expressed alone, telomeres of the cancer cells were lengthened, while HP1α-TRF1 attenuated telomere extension. Arnoult et al. (2012) proposed a model whereby TERRA repeat abundance (TERRA length) regulates HP1α abundance at telomeres through both direct recruitment of HP1α, and also by providing additional H3K9me3 sites by recruitment of methyltransferases and/or deacetylases (as TERRA and telomeric H3K9me3 marks are strongly correlated through the cell cycle). Due to the evidence shown by Maison et al. (2011) involving pericentromeric heterochromatin, Arnoult et al. (2012) postulated that HP1α enrichment at telomeres is not due solely to the increased H3K9me3 abundance, but that HP1α recruitment requires the presence of lncRNA before HP1α anchors to H3K9me3. They hypothesised that TERRA and HP1α operate in a negative feedback loop in which TERRA transcription causes HP1α and H3K9me3 enrichment at telomeres, thereby repressing TERRA transcription and balancing heterochromatin formation (Arnoult et al., 2012). The evidence presented in this thesis adds to this model, with the G4 recognition of TERRA by HP1α causing the recruitment to this specific compartment of heterochromatin. To test this model, through the use of HP1α mutants that either don't bind G4s (hinge patch mutants) or lack specificity for G4s (CSD

helix mutants), the loss of telomeric HP1α may be assessed by immunofluorescence or possibly proximity ligation assays with telomeric components such as TRF1.

6.9 Concluding remarks

Taken together, these results show that HP1α is a highly specific parallel G4 binding protein, and its interaction with TERRA is mediated through interaction with the hinge and docking with the CSD to form an ordered complex that can still bind H3K9me3. The meticulous binding of HP1α to TERRA may provide RNA-mediated *de novo* targeting of HP1α to telomeric heterochromatin for protection of chromosome ends, and provides insight into a distinct mechanism of non-canonical nucleic acid binding.

Bibliography

Aagaard, L., Laible, G., Selenko, P., Schmid, M., Dorn, R., Schotta, G., Kuhfittig, S., Wolf, A., Lebersorger, A., Singh, P. B., et al. (1999). Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31. *The EMBO journal*, 18(7):1923–1938.

Aasland, R. and Stewart, F. (1995). The chromo shadow domain, a second chromo domain in heterochromatin-binding protein 1, HP1. *Nucleic acids research*, 23(16):3168–3173.

Abdelhamid, M. A. and Waller, Z. A. (2020). Tricky topology: persistence of folded human telomeric i-motif DNA at ambient temperature and neutral pH. *Frontiers in chemistry*, 8:40.

Adrian, M., Heddi, B., and Phan, A. T. (2012). NMR spectroscopy of G-quadruplexes. *Methods*, 57(1):11–24.

Agazie, Y., Lee, J., and Burkholder, G. (1994). Characterization of a new monoclonal antibody to triplex DNA and immunofluorescent staining of mammalian chromosomes. *Journal of Biological Chemistry*, 269(9):7019–7023.

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.

Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R. A., and Yang, D. (2006). Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic acids research*, 34(9):2723–2735.

Arnold, P., Schöler, A., Pachkov, M., Balwierz, P. J., Jørgensen, H., Stadler, M. B., van Nimwegen, E., and Schübeler, D. (2013). Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting. *Genome research*, 23(1):60–73.

Arnoult, N., Van Beneden, A., and Decottignies, A. (2012). Telomere length regulates TERRA levels through increased trimethylation of telomeric H3K9 and HP1α. *Nature structural & molecular biology*, 19(9):948–956.

Arora, A. and Maiti, S. (2009). Differential biophysical behavior of human telomeric RNA and DNA quadruplex. *The Journal of Physical Chemistry B*, 113(30):10515–10520.

Arora, S., Horne, W. S., and Islam, K. (2019). Engineering methyllysine writers and readers for allele-specific regulation of protein–protein interactions. *Journal of the American Chemical Society*, 141(39):15466–15470.

Avery, O., MacLeod, C., and McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type III. *J. of Experimental Medicine*, 79(2):137–158.

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.

Ball, L. J., Murzina, N. V., Broadhurst, R. W., Raine, A. R., Archer, S. J., Stott, F. J., Murzin, A. G., Singh, P. B., Domaille, P. J., and Laue, E. D. (1997). Structure of the chromatin binding (chromo) domain from mouse modifier protein 1. *The EMBO journal*, 16(9):2473–2481.

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.

Bao, H.-L., Ishizuka, T., Sakamoto, T., Fujimoto, K., Uechi, T., Kenmochi, N., and Xu, Y. (2017). Characterization of human telomere RNA G-quadruplex structures in vitro and in living cells using 19F NMR spectroscopy. *Nucleic acids research*, 45(9):5501–5511.

Barzak, F. M., Ryan, T. M., Kvach, M. V., Kurup, H. M., Aihara, H., Harris, R. S., Filichev, V. V., Harjes, E., and Jameson, G. B. (2021). Small-angle X-ray scattering models of APOBEC3B catalytic domain in a complex with a single-stranded DNA inhibitor. *Viruses*, 13(2):290.

Becker, J. S., Nicetto, D., and Zaret, K. S. (2016). H3K9me3-dependent heterochromatin: barrier to cell fate changes. *Trends in Genetics*, 32(1):29–41.

Bettin, N., Oss Pegorar, C., and Cusanelli, E. (2019). The emerging roles of TERRA in telomere maintenance and genome stability. *Cells*, 8(3):246.

Bhattacharjee, A., Wang, Y., Diao, J., and Price, C. M. (2017). Dynamic DNA binding, junction recognition and G4 melting activity underlie the telomeric and genome-wide roles of human CST. *Nucleic acids research*, 45(21):12311–12324.

Biffi, G., Di Antonio, M., Tannahill, D., and Balasubramanian, S. (2014). Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nature chemistry*, 6(1):75.

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.

Blackledge, N. P., Farcas, A. M., Kondo, T., King, H. W., McGouran, J. F., Hanssen, L. L., Ito, S., Cooper, S., Kondo, K., Koseki, Y., et al. (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell*, 157(6):1445–1459.

Blackledge, N. P., Rose, N. R., and Klose, R. J. (2015). Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nature reviews Molecular cell biology*, 16(11):643–649.

Blasco, M. A. (2007). The epigenetic regulation of mammalian telomeres. *Nature Reviews Genetics*, 8(4):299–309.

Blum, A. D., Uhlenbeck, O. C., and Tinoco Jr, I. (1972). Circular dichroism study of nine species of transfer ribonucleic acid. *Biochemistry*, 11(17):3248–3256.

Borders Jr, C., Broadwater, J. A., Bekeny, P. A., Salmon, J. E., Lee, A. S., Eldridge, A. M., and Pett, V. B. (1994). A structural role for arginine in proteins: multiple hydrogen bonds to backbone carbonyl oxygens. *Protein Science*, 3(4):541–548.

Bosch-Presegue, L., Raurell-Vila, H., Thackray, J. K., González, J., Casal, C., Kane-Goldsmith, N., Vizoso, M., Brown, J. P., Gómez, A., Ausió, J., et al. (2017). Mammalian HP1 isoforms have specific roles in heterochromatin structure and organization. *Cell reports*, 21(8):2048–2057.

Bowman, G. D. and Poirier, M. G. (2015). Post-translational modifications of histones that influence nucleosome dynamics. *Chemical reviews*, 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. C., and Fojta, M. (2014). DNA and RNA quadruplex-binding proteins. *International journal of molecular sciences*, 15(10):17493–17517.

Brázda, V., Hároníková, L., Liao, J. C., Fridrichová, H., and Jagelská, E. B. (2016). Strong preference of BRCA1 protein to topologically constrained non-B DNA structures. *BMC molecular biology*, 17(1):1–9.

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

Bryan, L. C., Weilandt, D. R., Bachmann, A. L., Kilic, S., Lechner, C. C., Odermatt, P. D., Fantner, G. E., Georgeon, S., Hantschel, O., Hatzimanikatis, V., et al. (2017). Single-molecule kinetic analysis of HP1-chromatin binding reveals a dynamic network of histone modification and DNA interactions. *Nucleic acids research*, 45(18):10504–10517.

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. *Nature medicine*, 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.

Byrd, A. K. and Raney, K. D. (2015). A parallel quadruplex DNA is bound tightly but unfolded slowly by pif1 helicase. *Journal of Biological Chemistry*, 290(10):6482–6494.

Campisi, J. and d'Adda di Fagagna, F. (2007). Cellular senescence: when bad things happen to good cells. *Nature reviews Molecular cell biology*, 8(9):729–740.

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., Chang, E. Y., Shankar, S., Kuchenbecker, K. M., Simon, M. D., Madhani, H. D., Narlikar, G. J., and Al-Sady, B. (2011). Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. *Molecular cell*, 41(1):67–81.

Canzio, D., Larson, A., and Narlikar, G. J. (2014). Mechanisms of functional promiscuity by HP1 proteins. *Trends in cell biology*, 24(6):377–386.

Chakraborty, A., Prasanth, K. V., and Prasanth, S. G. (2014). Dynamic phosphorylation of HP1α regulates mitotic progression in human cells. *Nature communications*, 5(1):1–14.

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):877–881.

Chatterjee, S., Zagelbaum, J., Savitsky, P., Sturzenegger, A., Huttner, D., Janscak, P., Hickson, I. D., Gileadi, O., and Rothenberg, E. (2014). Mechanistic insight into the interaction of BLM helicase with intra-strand G-quadruplex structures. *Nature communications*, 5(1):1–12.

Chen, M. C., Tippana, R., Demeshkina, N. A., Murat, P., Balasubramanian, S., Myong, S., and Ferré-D'Amaré, A. R. (2018). Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature*, 558(7710):465–469.

Chen, S., Wang, C., Sun, L., Wang, D.-L., Chen, L., Huang, Z., Yang, Q., Gao, J., Yang, X.-B., Chang, J.-F., et al. (2015). RAD6 promotes homologous recombination repair by activating the autophagy-mediated degradation of heterochromatin protein HP1. *Molecular and Cellular Biology*, 35(2):406–416.

Choi, J. and Majima, T. (2011). Conformational changes of non-B DNA. *Chemical Society Reviews*, 40(12):5893–5909.

Choo, K. A. (1997). *The centromere*, volume 320. Oxford University Press Oxford.

Chow, T. T., Shi, X., Wei, J.-H., Guan, J., Stadler, G., Huang, B., and Blackburn, E. H. (2018). Local enrichment of HP1alpha at telomeres alters their structure and regulation of telomere protection. *Nature communications*, 9(1):1–13.

Chu, H.-P., Cifuentes-Rojas, C., Kesner, B., Aeby, E., Lee, H.-g., Wei, C., Oh, H. J., Boukhali, M., Haas, W., and Lee, J. T. (2017). TERRA RNA antagonizes ATRX and protects telomeres. *Cell*, 170(1):86–101.

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 research*, 38(16):5569–5580.

Consortium, I. H. G. S. et al. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409:860–921.

Cosgrove, M. S., Boeke, J. D., and Wolberger, C. (2004). Regulated nucleosome mobility and the histone code. *Nature structural & molecular biology*, 11(11):1037–1043.

Cosgrove, M. S. and Wolberger, C. (2005). How does the histone code work? *Biochemistry and Cell Biology*, 83(4):468–476.

Cowieson, N. P., Partridge, J. F., Allshire, R. C., and McLaughlin, P. J. (2000). Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis. *Current Biology*, 10(9):517–525.

Cubiles, M. D., Barroso, S., Vaquero-Sedas, M. I., Enguix, A., Aguilera, A., and Vega-Palas, M. A. (2018). Epigenetic features of human telomeres. *Nucleic acids research*, 46(5):2347–2355.

D'Adda di Fagagna, F., 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):194–198.

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.

Day, H. A., Pavlou, P., and Waller, Z. A. (2014). i-Motif DNA: structure, stability and targeting with ligands. *Bioorganic & medicinal chemistry*, 22(16):4407–4418.

De Koning, L., Savignoni, A., Boumendil, C., Rehman, H., Asselain, B., Sastre-Garau, X., and Almouzni, G. (2009). Heterochromatin protein 1α: a hallmark of cell proliferation relevant to clinical oncology. *EMBO molecular medicine*, 1(3):178–191.

De Lange, R., Burtscher, H., Jarsch, M., and Weidle, U. H. (2001). Identification of metastasis-associated genes by transcriptional profiling of metastatic versus non-metastatic colon cancer cell lines. *Anticancer research*, 21(4A):2329–2339.

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

De Lucia, F., Ni, J.-Q., Vaillant, C., and Sun, F.-L. (2005). HP1 modulates the transcription of cell-cycle regulators in Drosophila melanogaster. *Nucleic acids research*, 33(9):2852–2858.

De Silanes, I. L., Grana, O., De Bonis, M. L., Dominguez, O., Pisano, D. G., and Blasco, M. A. (2014). Identification of TERRA locus unveils a telomere protection role through association to nearly all chromosomes. *Nature communications*, 5(1):1–13.

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.

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

Doluca, O., Withers, J. M., and Filichev, V. V. (2013). Molecular engineering of guanine-rich sequences: Z-DNA, DNA triplexes, and G-quadruplexes. *Chemical reviews*, 113(5):3044–3083.

Donald, J. E., Kulp, D. W., and DeGrado, W. F. (2011). Salt bridges: Geometrically specific, designable interactions. *Proteins: Structure, Function, and Bioinformatics*, 79(3):898–915.

Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K., and Dickerson, R. E. (1981). Structure of a B-DNA dodecamer: conformation and dynamics. *Proceedings of the National Academy of Sciences*, 78(4):2179–2183.

Eissenberg, J. C. (2001). Molecular biology of the chromo domain: an ancient chromatin module comes of age. *Gene*, 275(1):19–29.

Eissenberg, J. C., James, T. C., Foster-Hartnett, D. M., Hartnett, T., Ngan, V., and Elgin, S. (1990). Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in drosophila melanogaster. *Proceedings of the National Academy of Sciences*, 87(24):9923–9927.

Esmaili, N. and Leroy, J. L. (2005). i-motif solution structure and dynamics of the d(AACCCC) and d(CCCCAA) tetrahymena telomeric repeats. *Nucleic acids research*, 33(1):213–224.

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al. (2022). Protein complex prediction with AlphaFold-Multimer. *BioRxiv*, pages 2021–10.

Fanti, L., Giovinazzo, G., Berloco, M., and Pimpinelli, S. (1998). The heterochromatin protein 1 prevents telomere fusions in Drosophila. *Molecular cell*, 2(5):527–538.

Farnung, B. O., Brun, C. M., Arora, R., Lorenzi, L. E., and Azzalin, C. M. (2012). Telomerase efficiently elongates highly transcribing telomeres in human cancer cells. *PLoS One*, 7(4):e35714.

Fay, M. M., Anderson, P. J., and Ivanov, P. (2017). ALS/FTD-associated C9ORF72 repeat RNA promotes phase transitions in vitro and in cells. *Cell reports*, 21(12):3573–3584.

Feig, M. and Pettitt, B. M. (1998). A molecular simulation picture of DNA hydration around A-and B-DNA. *Biopolymers: Original Research on Biomolecules*, 48(4):199–209.

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

Finch, J. and Klug, A. (1976). Solenoidal model for superstructure in chromatin. *Proceedings of the National Academy of Sciences*, 73(6):1897–1901.

Fischle, W., Wang, Y., Jacobs, S. A., Kim, Y., Allis, C. D., and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes & development*, 17(15):1870–1881.

Franke, D. and Svergun, D. I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of applied crystallography*, 42(2):342–346.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Molecular cell*, 45(3):344–356.

García-Cao, M., O'Sullivan, R., Peters, A. H., Jenuwein, T., and Blasco, M. A. (2004). Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. *Nature genetics*, 36(1):94–99.

Gehring, K., Leroy, J.-L., and Guéron, M. (1993). A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature*, 363(6429):561–565.

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.

González-Vasconcellos, I., Cobos-Fernández, M. A., Atkinson, M. J., Fernandez-Piqueras, J., and Santos, J. (2022). Quantifying telomeric lncRNAs using PNA-labelled RNA-Flow FISH (RNA-Flow). *Communications biology*, 5(1):1–7.

Goodarzi, A. A., Noon, A. T., Deckbar, D., Ziv, Y., Shiloh, Y., Löbrich, M., and Jeggo, P. A. (2008). ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. *Molecular cell*, 31(2):167–177.

Gotfredsen, C. H., Schultze, P., and Feigon, J. (1998). Solution structure of an intramolecular pyrimidine- purine-pyrimidine triplex containing an RNA third strand. *Journal of the American Chemical Society*, 120(18):4281–4289.

Greenfield, N. J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols*, 1(6):2876.

Greenfield, N. J. and Fasman, G. D. (1969). Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry*, 8(10):4108–4116.

Greil, F., Van Der Kraan, I., Delrow, J., Smothers, J. F., De Wit, E., Bussemaker, H. J., Van Driel, R., Henikoff, S., and Van Steensel, B. (2003). Distinct HP1 and Su(var)3-9 complexes bind to sets of developmentally coexpressed genes depending on chromosomal location. *Genes & development*, 17(22):2825–2838.

Grewal, S. I. and Jia, S. (2007). Heterochromatin revisited. *Nature Reviews Genetics*, 8(1):35–46.

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.

Grote, P. and Herrmann, B. G. (2013). The long non-coding RNA Fendrr links epigenetic control mechanisms to gene regulatory networks in mammalian embryogenesis. *RNA biology*, 10(10):1579–1585.

Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., Macura, K., Bläss, G., Kellis, M., Werber, M., et al. (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Developmental cell*, 24(2):206–214.

Guedin, A., Gros, J., Alberti, P., and Mergny, J.-L. (2010). How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic acids research*, 38(21):7858–7868.

Guenatri, M., Bailly, D., Maison, C., and Almouzni, G. (2004). Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *The Journal of cell biology*, 166(4):493–505.

Haarhuis, J. H., van der Weide, R. H., Blomen, V. A., Yáñez-Cuna, J. O., Amendola, M., van Ruiten, M. S., Krijger, P. H., Teunissen, H., Medema, R. H., van Steensel, B., et al. (2017). The cohesin release factor WAPL restricts chromatin loop extension. *Cell*, 169(4):693–707.

Hale, T. K., Contreras, A., Morrison, A. J., and Herrera, R. E. (2006). Phosphorylation of the linker histone H1 by CDK regulates its binding to HP1α. *Molecular cell*, 22(5):693–699.

Hansen, J. C., Maeshima, K., and Hendzel, M. J. (2021). The solid and liquid states of chromatin. *Epigenetics & Chromatin*, 14(1):1–33.

Hardwick, J. S., Ptchelkine, D., El-Sagheer, A. H., Tear, I., Singleton, D., Phillips, S. E., Lane, A. N., and Brown, T. (2017). 5-Formylcytosine does not change the global structure of DNA. *Nature structural & molecular biology*, 24(6):544–552.

Harley, C. B., Futcher, A. B., and Greider, C. W. (1990). Telomeres shorten during ageing of human fibroblasts. *Nature*, 345(6274):458–460.

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.

Heddi, B., Cheong, V. V., Martadinata, H., and Phan, A. T. (2015). Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide–quadruplex complex. *Proceedings of the National Academy of Sciences*, 112(31):9608–9613.

Her, C., Phan, T. M., Jovic, N., Kapoor, U., Ackermann, B. E., Rizuan, A., Kim, Y., Mittal, J., and Debelouchina, G. T. (2022). Molecular interactions underlying the phase separation of HP1α: Role of phosphorylation, ligand and nucleic acid binding. *bioRxiv*.

Herbert, A. and Rich, A. (1999). Left-handed Z-DNA: structure and function. *Genetica*, 106(1):37–47.

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.

Holliday, R. (1964). A mechanism for gene conversion in fungi. *Genetics Research*, 5(2):282–304.

Holoch, D. and Moazed, D. (2015). RNA-mediated epigenetic regulation of gene expression. *Nature Reviews Genetics*, 16(2):71–84.

Holzwarth, G. and Doty, P. (1965). The ultraviolet circular dichroism of polypeptides. *Journal of the American Chemical Society*, 87(2):218–228.

Hoogsteen, K. (1959). The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. *Acta crystallographica*, 12(10):822–823.

Hoogsteen, K. (1963). The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallographica*, 16(9):907–916.

Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., and Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic acids research*, 43(D1):D512–D520.

Horvath, M. P. and Schultz, S. C. (2001). DNA G-quartets in a 1.86 Å resolution structure of an Oxytricha nova telomeric protein-DNA complex. *Journal of molecular biology*, 310(2):367–377.

Huang, Y., Myers, M. P., and Xu, R.-M. (2006). Crystal structure of the HP1-EMSY complex reveals an unusual mode of HP1 binding. *Structure*, 14(4):703–712.

Hübner, B., von Otter, E., Ahsan, B., Wee, M. L., Henriksson, S., Ludwig, A., and Sandin, S. (2022). Ultrastructure and nuclear architecture of telomeric chromatin revealed by correlative light and electron microscopy. *Nucleic Acids Research*, 50(9):5047–5063.

Hudson, J. S., Ding, L., Le, V., Lewis, E., and Graves, D. (2014). Recognition and binding of human telomeric G-quadruplex DNA by unfolding protein 1. *Biochemistry*, 53(20):3347–3356.

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

Hwang, H., Buncher, N., Opresko, P. L., and Myong, S. (2012). POT1-TPP1 regulates telomeric overhang structural dynamics. *Structure*, 20(11):1872–1880.

Ignatev, A., Piatkov, K., Pylypenko, O., and Rak, A. (2007). A size filtration approach to purify low affinity complexes for crystallization. *Journal of structural biology*, 159(1):154–157.

Ivanov, V. I., Minchenkova, L. E., Minyat, E. E., Frank-Kamenetskii, M. D., and Schyolkina, A. K. (1974). The B to A transition of DNA in solution. *Journal of molecular biology*, 87(4):817–833.

Jacques, D. A. and Trewhella, J. (2010). Small-angle scattering for structural biology - expanding the frontier while avoiding the pitfalls. *Protein science*, 19(4):642–657.

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 research*, 26(1):97–107.

Jang, S. M., Kauzlaric, A., Quivy, J.-P., Pontis, J., Rauwel, B., Coluccio, A., Offner, S., Duc, J., Turelli, P., Almouzni, G., et al. (2018). KAP1 facilitates reinstatement of heterochromatin after DNA replication. *Nucleic acids research*, 46(17):8788–8802.

Janssen, A., Colmenares, S. U., and Karpen, G. H. (2018). Heterochromatin: guardian of the genome. *Annual review of cell and developmental biology*, 34(1):265–288.

Joachimi, A., Benz, A., and Hartig, J. S. (2009). A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorganic & medicinal chemistry*, 17(19):6811–6815.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589.

Kang, J., Chaudhary, J., Dong, H., Kim, S., Brautigam, C. A., and Yu, H. (2011). Mitotic centromeric targeting of HP1 and its binding to Sgo1 are dispensable for sister-chromatid cohesion in human cells. *Molecular biology of the cell*, 22(8):1181–1190.

Kar, A., Jones, N., Arat, N. Ö., Fishel, R., and Griffith, J. D. (2018). Long repeating (TTAGGG)n single-stranded DNA self-condenses into compact beaded filaments stabilized by G-quadruplex formation. *Journal of Biological Chemistry*, 293(24):9473–9485.

Karsisiotis, A. I., Hessari, N. M., Novellino, E., Spada, G. P., Randazzo, A., and Webba da Silva, M. (2011). Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angewandte Chemie International Edition*, 50(45):10645–10648.

Katoh, K., Misawa, K., Kuma, K.-i., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research*, 30(14):3059–3066.

Kaustov, L., Ouyang, H., Amaya, M., Lemak, A., Nady, N., Duan, S., Wasney, G. A., Li, Z., Vedadi, M., Schapira, M., et al. (2011). Recognition and specificity determinants of the human cbx chromodomains. *Journal of Biological Chemistry*, 286(1):521–529.

Keenen, M. M., Brown, D., Brennan, L. D., Renger, R., Khoo, H., Carlson, C. R., Huang, B., Grill, S. W., Narlikar, G. J., and Redding, S. (2021). HP1 proteins compact DNA into mechanically and positionally stable phase separated domains. *Elife*, 10:e64563.

Keller, C., Adaixo, R., Stunnenberg, R., Woolcock, K. J., Hiller, S., and Bühler, M. (2012). HP1Swi6 mediates the recognition and destruction of heterochromatic RNA transcripts. *Molecular cell*, 47(2):215–227.

Kikhney, A. G. and Svergun, D. I. (2015). A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS letters*, 589(19):2570–2577.

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(1):1–11.

Kim, B. G., Shek, Y. L., and Chalikian, T. V. (2013). Polyelectrolyte effects in G-quadruplexes. *Biophysical Chemistry*, 184:95–100.

Kirschmann, D. A., Lininger, R. A., Gardner, L. M., Seftor, E. A., Odero, V. A., Ainsztein, A. M., Earnshaw, W. C., Wallrath, L. L., and Hendrix, M. J. (2000). Down-regulation of HP1Hsα expression is associated with the metastatic phenotype in breast cancer. *Cancer research*, 60(13):3359–3363.

Knipscheer, P., Flotho, A., Klug, H., Olsen, J. V., van Dijk, W. J., Fish, A., Johnson, E. S., Mann, M., Sixma, T. K., and Pichler, A. (2008). Ubc9 sumoylation regulates SUMO target discrimination. *Molecular cell*, 31(3):371–382.

Kogut, M., Kleist, C., and Czub, J. (2019). Why do G-quadruplexes dimerize through the 5'-ends? Driving forces for G4 DNA dimerization examined in atomic detail. *PLoS computational biology*, 15(9):e1007383.

Kolesnikova, S. and Curtis, E. A. (2019). Structure and function of multimeric G-quadruplexes. *Molecules*, 24(17):3074.

Komůrková, D., Svobodová Kovaříková, A., and Bártová, E. (2021). G-quadruplex structures colocalize with transcription factories and nuclear speckles surrounded by acetylated and dimethylated histones h3. *International Journal of Molecular Sciences*, 22(4):1995.

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.

Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M., and Yanagawa, H. (2009). Six classes of nuclear localization signals specific to different binding grooves of importin α. *Journal of Biological Chemistry*, 284(1):478–485.

Kozin, M. B. and Svergun, D. I. (2001). Automated matching of high-and low-resolution structural models. *Journal of applied crystallography*, 34(1):33–41.

Kreig, A., Calvert, J., Sanoica, J., Cullum, E., Tipanna, R., and Myong, S. (2015). G-quadruplex formation in double strand DNA probed by NMM and CV fluorescence. *Nucleic acids research*, 43(16):7961–7970.

Krissinel, E. and Henrick, K. (2007). Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute. *J Mol Biol*, 372:774–797.

Kumar, A. and Kono, H. (2020). Heterochromatin protein 1 (HP1): interactions with itself and chromatin components. *Biophysical Reviews*, pages 1–14.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature*, 410(6824):116–120.

Lai, L.-T., Lee, P. J., and Zhang, L.-F. (2013). Immunofluorescence protects RNA signals in simultaneous RNA–DNA FISH. *Experimental cell research*, 319(3):46–55.

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):236–240.

Laughlan, G., Murchie, A. I., Norman, D. G., Moore, M. H., Moody, P. C., Lilley, D. M., and Luisi, B. (1994). The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science*, 265(5171):520–524.

Lee, J. S., Burkholder, G. D., Latimer, L. J., Haug, B. L., and Braun, R. P. (1987). A monoclonal antibody to triplex DNA binds to eucaryotic chromosomes. *Nucleic acids research*, 15(3):1047–1061.

LeRoy, G., Weston, J. T., Zee, B. M., Young, N. L., Plazas-Mayorca, M. D., and Garcia, B. A. (2009). Heterochromatin protein 1 is extensively decorated with histone code-like post-translational modifications. *Molecular & Cellular Proteomics*, 8(11):2432–2442.

Li, K., Yatsunyk, L., and Neidle, S. (2021). Water spines and networks in G-quadruplex structures. *Nucleic acids research*, 49(1):519–528.

Li, Z., Li, Y., Amaya, M. F., Edwards, A. M., Arrowsmith, C. H., Weigelt, J., Bountra, C., Bochkarev, A., Min, J., and Ouyang, H. (to be published). Crystal Structural of CBX5 Chromo Shadow Domain. *To be published*.

Lim, K. W., Amrane, S., Bouaziz, S., Xu, W., Mu, Y., Patel, D. J., Luu, K. N., and Phan, A. T. (2009). Structure of the human telomere in K+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *Journal of the American Chemical Society*, 131(12):4301–4309.

Lin, F.-M., Kumar, S., Ren, J., Karami, S., Bahnassy, S., Li, Y., Zheng, X., Wang, J., and Bawa-Khalfe, T. (2016). SUMOylation of HP1α supports association with ncRNA to define responsiveness of breast cancer cells to chemotherapy. *Oncotarget*, 7(21):30336.

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.

Liu, H., Qu, Q., Warrington, R., Rice, A., Cheng, N., and Yu, H. (2015). Mitotic transcription installs Sgo1 at centromeres to coordinate chromosome segregation. *Molecular cell*, 59(3):426–436.

Liu, Y., Qin, S., Lei, M., Tempel, W., Zhang, Y., Loppnau, P., Li, Y., and Min, J. (2017). Peptide recognition by heterochromatin protein 1 (HP1) chromoshadow domains revisited: Plasticity in the pseudosymmetric histone binding site of human HP1. *Journal of Biological Chemistry*, 292(14):5655–5664.

Lomberk, G., Bensi, D., Fernandez-Zapico, M. E., and Urrutia, R. (2006). Evidence for the existence of an HP1-mediated subcode within the histone code. *Nature cell biology*, 8(4):407–415.

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

Louis, E. J. (2002). Are drosophila telomeres an exception or the rule? *Genome biology*, 3(10):reviews0007–1.

Lu, J. and Gilbert, D. M. (2007). Proliferation-dependent and cell cycle–regulated transcription of mouse pericentric heterochromatin. *The Journal of cell biology*, 179(3):411–421.

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):251–260.

Luo, Z., Dauter, M., and Dauter, Z. (2014). Phosphates in the Z-DNA dodecamer are flexible, but their P-SAD signal is sufficient for structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 70(7):1790–1800.

Luu, K. N., Phan, A. T., Kuryavyi, V., Lacroix, L., and Patel, D. J. (2006). Structure of the human telomere in K+ solution: an intramolecular (3+ 1) G-quadruplex scaffold. *Journal of the American Chemical Society*, 128(30):9963–9970.

Machida, S., Takizawa, Y., Ishimaru, M., Sugita, Y., Sekine, S., Nakayama, J.-i., Wolf, M., and Kurumizaka, H. (2018). Structural basis of heterochromatin formation by human HP1. *Molecular cell*, 69(3):385–397.

Maciejewski, P. M., Peterson, F. C., Anderson, P. J., and Brooks, C. L. (1995). Mutation of serine 90 to glutamic acid mimics phosphorylation of bovine prolactin. *Journal of Biological Chemistry*, 270(46):27661–27665.

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.

Mandal, P., Venkadesh, S., and Gautham, N. (2011). Structure of d(CGGGTACCCG)4 as a four-way Holliday junction. *Acta Crystallographica Section F*, 67(12):1506–1510.

Manzini, G., Yathindra, N., and Xodo, L. (1994). Evidence for intramolecularly folded i-DNA structures in biologically relevant CCC-repeat sequences. *Nucleic acids research*, 22(22):4634–4640.

Marsh, T. C. and Henderson, E. (1994). G-wires: self-assembly of a telomeric oligonucleotide, d(GGGGTTGGGG), into large superstructures. *Biochemistry*, 33(35):10718–10724.

Martadinata, H. and Phan, A. T. (2009). Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K+ solution. *Journal of the American Chemical Society*, 131(7):2570–2578.

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.

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-Stephenson, V. (2022). G4-quadruplex-binding proteins: review and insights into selectivity. *Biophysical Reviews*, pages 1–20.

Mendez, D. L., Kim, D., Chruszcz, M., Stephens, G. E., Minor, W., Khorasanizadeh, S., and Elgin, S. C. (2011). The HP1a disordered C terminus and chromo shadow domain cooperate to select target peptide partners. *Chembiochem*, 12(7):1084–1096.

Mendez, D. L., Mandt, R. E., and Elgin, S. C. (2013). Heterochromatin Protein 1a (HP1a) partner specificity is determined by critical amino acids in the chromo shadow domain and C-terminal extension. *Journal of Biological Chemistry*, 288(31):22315–22323.

Mergny, J.-L. and Lacroix, L. (2009). UV melting of G-quadruplexes. *Current protocols in nucleic acid chemistry*, 37(1):17–1.

Mersfelder, E. L. and Parthun, M. R. (2006). The tale beyond the tail: histone core domain modifications and the regulation of chromatin structure. *Nucleic acids research*, 34(9):2653–2662.

Meyne, J., Ratliff, R. L., and MoYzIs, R. K. (1989). Conservation of the human telomere sequence (TTAGGG)n among vertebrates. *Proceedings of the National Academy of Sciences*, 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 wissenschaftlichen Briefwechsel von F. Miescher*, 1:33–38.

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.

Miserachs, H. G., Donghi, D., Börner, R., Johannsen, S., and Sigel, R. K. (2016). Distinct differences in metal ion specificity of RNA and DNA G-quadruplexes. *JBIC Journal of Biological Inorganic Chemistry*, 21(8):975–986.

Mishima, Y., Watanabe, M., Kawakami, T., Jayasinghe, C. D., Otani, J., Kikugawa, Y., Shirakawa, M., Kimura, H., Nishimura, O., Aimoto, S., et al. (2013). Hinge and chromoshadow of HP1α participate in recognition of K9 methylated histone H3 in nucleosomes. *Journal of molecular biology*, 425(1):54–70.

Monsen, R. C., Chakravarthy, S., Dean, W. L., Chaires, J. B., and Trent, J. O. (2021). The solution structures of higher-order human telomere G-quadruplex multimers. *Nucleic acids research*, 49(3):1749–1768.

Montero, J. J., Lopez de Silanes, I., Grana, O., and Blasco, M. A. (2016). Telomeric RNAs are essential to maintain telomeres. *Nature communications*, 7(1):1–13.

Montero, J. J., López-Silanes, I., Megías, D., F Fraga, M., Castells-García, Á., and Blasco, M. A. (2018). TERRA recruitment of polycomb to telomeres is essential for histone trymethylation marks at telomeric heterochromatin. *Nature communications*, 9(1):1–14.

Moye, A. L., Porter, K. C., Cohen, S. B., Phan, T., Zyner, K. G., Sasaki, N., Lovrecz, G. O., Beck, J. L., and Bryan, T. M. (2015). Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nature communications*, 6(1):1–12.

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.

Müller-Ott, K., Erdel, F., Matveeva, A., Mallm, J.-P., Rademacher, A., Hahn, M., Bauer, C., Zhang, Q., Kaltofen, S., Schotta, G., et al. (2014). Specificity, propagation, and memory of pericentric heterochromatin. *Molecular systems biology*, 10(8):746.

Munari, F., Soeroes, S., Zenn, H. M., Schomburg, A., Kost, N., Schröder, S., Klingberg, R., Rezaei-Ghaleh, N., Stützer, A., Gelato, K. A., et al. (2012). Methylation of lysine 9 in histone H3 directs alternative modes of highly dynamic interaction of heterochromatin protein hHP1β with the nucleosome. *Journal of Biological Chemistry*, 287(40):33756–33765.

Murnane, J. P. (2012). Telomere dysfunction and chromosome instability. *Mutation research/Fundamental and molecular mechanisms of mutagenesis*, 730(1-2):28–36.

Nergadze, S. G., Farnung, B. O., Wischnewski, H., Khoriauli, L., Vitelli, V., Chawla, R., Giulotto, E., and Azzalin, C. M. (2009). CpG-island promoters drive transcription of human telomeres. *Rna*, 15(12):2186–2194.

Nielsen, A. L., Ortiz, J. A., You, J., Oulad-Abdelghani, M., Khechumian, R., Gansmuller, A., Chambon, P., and Losson, R. (1999). Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. *The EMBO Journal*, 18(22):6385–6395.

Nielsen, A. L., Oulad-Abdelghani, M., Ortiz, J. A., Remboutsika, E., Chambon, P., and Losson, R. (2001). Heterochromatin formation in mammalian cells: interaction between histones and HP1 proteins. *Molecular cell*, 7(4):729–739.

Nielsen, P. R., Nietlispach, D., Mott, H. R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A. G., Murzina, N. V., and Laue, E. D. (2002). Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature*, 416(6876):103–107.

Nikitina, T. and Woodcock, C. L. (2004). Closed chromatin loops at the ends of chromosomes. *The Journal of cell biology*, 166(2):161–165.

Nishibuchi, G., Machida, S., Osakabe, A., Murakoshi, H., Hiragami-Hamada, K., Nakagawa, R., Fischle, W., Nishimura, Y., Kurumizaka, H., Tagami, H., et al. (2014). N-terminal phosphorylation of HP1α increases its nucleosome-binding specificity. *Nucleic acids research*, 42(20):12498–12511.

Nishibuchi, G. and Nakayama, J.-i. (2014). Biochemical and structural properties of heterochromatin protein 1: understanding its role in chromatin assembly. *The Journal of Biochemistry*, 156(1):11–20.

Nonaka, N., Kitajima, T., Yokobayashi, S., Xiao, G., Yamamoto, M., Grewal, S. I., and Watanabe, Y. (2002). Recruitment of cohesin to heterochromatic regions by Swi6/HP1 in fission yeast. *Nature cell biology*, 4(1):89–93.

Norwood, L. E., Moss, T. J., Margaryan, N. V., Cook, S. L., Wright, L., Seftor, E. A., Hendrix, M. J., Kirschmann, D. A., and Wallrath, L. L. (2006). A requirement for dimerization of HP1Hsα in suppression of breast cancer invasion. *Journal of Biological Chemistry*, 281(27):18668–18676.

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*, 59(2):598–605.

Okuda, K., Bardeguez, A., Gardner, J. P., Rodriguez, P., Ganesh, V., Kimura, M., Skurnick, J., Awad, G., and Aviv, A. (2002). Telomere length in the newborn. *Pediatric research*, 52(3):377–381.

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.

Onel, B., Carver, M., Wu, G., Timonina, D., Kalarn, S., Larriva, M., and Yang, D. (2016). A new G-quadruplex with hairpin loop immediately upstream of the human BCL2 P1 promoter modulates transcription. *Journal of the American Chemical Society*, 138(8):2563–2570.

Ottoz, D. S. and Berchowitz, L. E. (2020). The role of disorder in RNA binding affinity and specificity. *Open Biology*, 10(12):200328.

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.

Parkinson, G. N., Lee, M. P., and Neidle, S. (2002). Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, 417(6891):876–880.

Peters, A. H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schöfer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., et al. (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell*, 107(3):323–337.

Petraccone, L., Erra, E., Duro, I., Esposito, V., Randazzo, A., Mayol, L., Mattia, C. A., Barone, G., and Giancola, C. (2005). Relative stability of quadruplexes containing different number of G-tetrads. *Nucleosides, Nucleotides and Nucleic Acids*, 24(5-7):757–760.

Phan, A. T., Kuryavyi, V., Luu, K. N., and Patel, D. J. (2007). Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K+ solution. *Nucleic acids research*, 35(19):6517–6525.

Phan, A. T., Luu, K. N., and Patel, D. J. (2006). Different loop arrangements of intramolecular human telomeric (3+ 1) G-quadruplexes in K+ solution. *Nucleic acids research*, 34(19):5715–5719.

Plath, K., Fang, J., Mlynarczyk-Evans, S. K., Cao, R., Worringer, K. A., Wang, H., de la Cruz, C. C., Otte, A. P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. *Science*, 300(5616):131–135.

Porro, A., Feuerhahn, S., Delafontaine, J., Riethman, H., Rougemont, J., and Lingner, J. (2014). Functional characterization of the TERRA transcriptome at damaged telomeres. *Nature communications*, 5(1):1–13.

Potaman, V. N. and Sinden, R. R. (2013). DNA: alternative conformations and biology. In *Madame Curie Bioscience Database [Internet]*. Landes Bioscience.

Pradhan, S., Solomon, R., Gangotra, A., Yakubov, G. E., Willmott, G. R., Whitby, C. P., Hale, T. K., and Williams, M. A. (2021). Depletion of HP1α alters the mechanical properties of MCF7 nuclei. *Biophysical Journal*, 120(13):2631–2643.

Puig Lombardi, E. and Londoño-Vallejo, A. (2020). A guide to computational methods for G-quadruplex prediction. *Nucleic acids research*, 48(1):1–15.

Qin, W., Stengl, A., Ugur, E., Leidescher, S., Ryan, J., Cardoso, M. C., and Leonhardt, H. (2021). HP1β carries an acidic linker domain and requires H3K9me3 for phase separation. *Nucleus*, 12(1):44–57.

Quivy, J., Roche, D., Kirschner, D., Tagami, H., Nakatani, Y., and Almouzni, G. (2004). A CAF-1 dependent pool of HP1 during heterochromatin duplication. *The EMBO journal*, 23(17):3516–3526.

Quivy, J.-P., Gérard, A., Cook, A. J., Roche, D., and Almouzni, G. (2008). The HP1–p150/CAF-1 interaction is required for pericentric heterochromatin replication and S-phase progression in mouse cells. *Nature structural & molecular biology*, 15(9):972–979.

Rajendran, A., Nakano, S.-i., and Sugimoto, N. (2010). Molecular crowding of the cosolutes induces an intramolecular i-motif structure of triplet repeat DNA oligomers at neutral pH. *Chemical communications*, 46(8):1299–1301.

Rambo, R. P. and Tainer, J. A. (2011). Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers*, 95(8):559–571.

Rani, R., Yaseen, A. M., Malwade, A., and Sevilimedu, A. (2019). An RNA aptamer to HP1/Swi6 facilitates heterochromatin formation at an ectopic locus in S. pombe. *RNA biology*, 16(6):742–753.

Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7):1665–1680.

Ray, S., Bandaria, J. N., Qureshi, M. H., Yildiz, A., and Balci, H. (2014). G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proceedings of the National Academy of Sciences*, 111(8):2990–2995.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., et al. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 406(6796):593–599.

Recagni, M., Tassinari, M., Doria, F., Cimino-Reale, G., Zaffaroni, N., Freccero, M., Folini, M., and Richter, S. N. (2019). The oncogenic signaling pathways in BRAF-Mutant melanoma cells are modulated by naphthalene Diimide-Like G-Quadruplex ligands. *Cells*, 8(10):1274.

Rich, A. and Zhang, S. (2003). Z-DNA: the long road to biological function. *Nature Reviews Genetics*, 4(7):566–572.

Roach, R. J. (2019). 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.*

Roach, R. J., Garavís, M., González, C., Jameson, G. B., Filichev, V. V., and Hale, T. K. (2020). Heterochromatin protein 1α interacts with parallel RNA and DNA G-quadruplexes. *Nucleic Acids Research*, 48(2):682–693.

Ruan, J., Ouyang, H., Amaya, M. F., Ravichandran, M., Loppnau, P., Min, J., and Zang, J. (2012). Structural basis of the chromodomain of Cbx3 bound to methylated peptides from histone h1 and G9a. *PloS one*, 7(4):e35376.

Rudert, F., Bronner, S., Garnier, J.-W., and Dolle, P. (1995). Transcripts from opposite strands of γ satellite DNA are differentially expressed during mouse development. *Mammalian Genome*, 6(2):76–83.

Ryan, D. P. and Tremethick, D. J. (2018). The interplay between H2A.Z and H3K9 methylation in regulating HP1α binding to linker histone-containing chromatin. *Nucleic acids research*, 46(18):9353–9366.

Saksouk, N., Simboeck, E., and Déjardin, J. (2015). Constitutive heterochromatin formation and transcription in mammals. *Epigenetics & chromatin*, 8(1):3.

Sales-Gil, R. and Vagnarelli, P. (2020). How HP1 post-translational modifications regulate heterochromatin formation and maintenance. *Cells*, 9(6):1460.

Sauer, M. and Paeschke, K. (2017). G-quadruplex unwinding helicases and their function in vivo. *Biochemical Society Transactions*, 45(5):1173–1182.

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):228–236.

Schoelz, J. M. and Riddle, N. C. (2022). Functions of HP1 proteins in transcriptional regulation. *Epigenetics & Chromatin*, 15(1):1–15.

Schwaiger, M., Kohler, H., Oakeley, E. J., Stadler, M. B., and Schübeler, D. (2010). Heterochromatin protein 1 (HP1) modulates replication timing of the Drosophila genome. *Genome research*, 20(6):771–780.

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N. A., Huber, W., Haering, C. H., Mirny, L., et al. (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, 551(7678):51–56.

Sepsiova, R., Necasova, I., Willcox, S., Prochazkova, K., Gorilak, P., Nosek, J., Hofr, C., Griffith, J. D., and Tomaska, L. (2016). Evolution of telomeres in Schizosaccharomyces pombe and its possible relationship to the diversification of telomere binding proteins. *PLoS One*, 11(4):e0154225.

Shen, J., Varshney, D., Simeone, A., Zhang, X., Adhikari, S., Tannahill, D., and Balasubramanian, S. (2021). Promoter G-quadruplex folding precedes transcription and is controlled by chromatin. *Genome biology*, 22(1):1–14.

Shimojo, H., Kawaguchi, A., Oda, T., Hashiguchi, N., Omori, S., Moritsugu, K., Kidera, A., Hiragami-Hamada, K., Nakayama, J.-i., Sato, M., et al. (2016). Extended string-like binding of the phosphorylated HP1α N-terminal tail to the lysine 9-methylated histone H3 tail. *Scientific reports*, 6(1):1–15.

Shu, H., Zhang, R., Xiao, K., Yang, J., and Sun, X. (2022). G-Quadruplex-Binding Proteins: Promising Targets for Drug Design. *Biomolecules*, 12(5):648.

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.

Singh, P. B. and Newman, A. G. (2022). HP1-Driven Micro-Phase Separation of Heterochromatin-Like Domains/Complexes. *Epigenetics Insights*, 15:25168657221109766.

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

Smothers, J. F. and Henikoff, S. (2001). The hinge and chromo shadow domain impart distinct targeting of HP1-like proteins. *Molecular and cellular biology*, 21(7):2555–2569.

Sokalingam, S., Raghunathan, G., Soundrarajan, N., and Lee, S.-G. (2012). A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. *PloS one*, 7(7):e40410.

Solomon, R. (2019). *Loss of HP1α alters nuclear integrity to promote cellular invasion: a thesis presented in partial fulfilment of the requirements for the degree of Master of Science in Biochemistry at Massey University, Manawatū, New Zealand*. PhD thesis, Massey University.

Soman, A., Korolev, N., and Nordenskiöld, L. (2022). Telomeric chromatin structure. *Current opinion in structural biology*, 77:102492.

Stadlbauer, P., Trantírek, L., Cheatham III, T. E., Koča, J., and Šponer, J. (2014). Triplex intermediates in folding of human telomeric quadruplexes probed by microsecond-scale molecular dynamics simulations. *Biochimie*, 105:22–35.

Steely Jr, H. T., Gray, D. M., and Ratliff, R. L. (1986). CD of homopolymer DNA-RNA hybrid duplexes and triplexes containing AT or AU base pairs. *Nucleic acids research*, 14(24):10071–10090.

Steudel, H. and Brigl, P. (1910). Über die Guanylsäure aus der Pankreasdrüse. II. Mitteilung.

Strahl, B. D. and Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, 403(6765):41–45.

Strom, A. R., Biggs, R. J., Banigan, E. J., Wang, X., Chiu, K., Herman, C., Collado, J., Yue, F., Politz, J. C. R., Tait, L. J., et al. (2021). HP1α is a chromatin crosslinker that controls nuclear and mitotic chromosome mechanics. *Elife*, 10:e63972.

Strom, A. R., Emelyanov, A. V., Mir, M., Fyodorov, D. V., Darzacq, X., and Karpen, G. H. (2017). Phase separation drives heterochromatin domain formation. *Nature*, 547(7662):241–245.

Tang, J., Kan, Z.-y., Yao, Y., Wang, Q., Hao, Y.-h., and Tan, Z. (2008). G-quadruplex preferentially forms at the very 3' end of vertebrate telomeric DNA. *Nucleic acids research*, 36(4):1200–1208.

Tang, W., Niu, K., Yu, G., Jin, Y., Zhang, X., Peng, Y., Chen, S., Deng, H., Li, S., Wang, J., et al. (2020). In vivo visualization of the i-motif DNA secondary structure in the Bombyx mori testis. *Epigenetics & Chromatin*, 13(1):1–12.

Thiru, A., Nietlispach, D., Mott, H. R., Okuwaki, M., Lyon, D., Nielsen, P. R., Hirshberg, M., Verreault, A., Murzina, N. V., and Laue, E. D. (2004). Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. *The EMBO journal*, 23(3):489–499.

Thoma, F., Koller, T., and Klug, A. (1979). Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *The Journal of cell biology*, 83(2):403–427.

Thorpe, J. H., Gale, B. C., Teixeira, S. C., and Cardin, C. J. (2003). Conformational and hydration effects of site-selective sodium, calcium and strontium ion binding to the DNA Holliday junction structure d(TCGGTACCGA)4. *Journal of molecular biology*, 327(1):97–109.

Traczyk, A., Liew, C. W., Gill, D. J., and Rhodes, D. (2020). Structural basis of G-quadruplex DNA recognition by the yeast telomeric protein Rap1. *Nucleic acids research*, 48(8):4562–4571.

Tretiakova, M. S., Bond, S. D., Wheeler, D., Contreras, A., Kocherginsky, M., Kroll, T. G., and Hale, T. K. (2014). Heterochromatin protein 1 expression is reduced in human thyroid malignancy. *Laboratory Investigation*, 94(7):788–795.

Ussery, D. W. (2002). DNA Structure: A-, B-and Z-DNA Helix Families. *Encyclopedia of life sciences*, pages 1–7.

Vakoc, C. R., Mandat, S. A., Olenchock, B. A., and Blobel, G. A. (2005). Histone H3 lysine 9 methylation and HP1γ are associated with transcription elongation through mammalian chromatin. *Molecular cell*, 19(3):381–391.

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. *Journal of Biological Chemistry*, 280(46):38117–38120.

Velez, G., Lin, M., Christensen, T., Faubion, W. A., Lomberk, G., and Urrutia, R. (2016). Evidence supporting a critical contribution of intrinsically disordered regions to the biochemical behavior of full-length human HP1γ. *Journal of molecular modeling*, 22(1):1–17.

Venyaminov, S. Y., Baikalov, I., Shen, Z. M., Wu, C.-S. C., and Yang, J. (1993). Circular dichroic analysis of denatured proteins: inclusion of denatured proteins in the reference set. *Analytical biochemistry*, 214(1):17–24.

Vorlíčková, M., Kejnovská, I., Bednářová, K., Renčiuk, 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., Horsley, D., Brown, N. R., Cowell, I. G., and Singh, P. B. (2000). Conservation of heterochromatin protein 1 function. *Molecular and cellular biology*, 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, Q., Liu, J., Chen, Z., Zheng, K., Chen, C., Hao, Y., and Tan, Z. (2011). G-quadruplex formation at the 3' end of telomere DNA inhibits its extension by telomerase, polymerase and unwinding by helicase. *Nucleic acids research*, 39(14):6229–6237.

Wang, X., Goodrich, K. J., Gooding, A. R., Naeem, H., Archer, S., Paucek, R. D., Youmans, D. T., Cech, T. R., and Davidovich, C. (2017). Targeting of polycomb repressive complex 2 to RNA by short repeats of consecutive guanines. *Molecular cell*, 65(6):1056–1067.

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.

Waterman, M. S. and Eggert, M. (1987). A new algorithm for best subsequence alignments with application to tRNA-rRNA comparisons. *Journal of molecular biology*, 197(4):723–728.

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

Wechsler, T., Newman, S., and West, S. C. (2011). Aberrant chromosome morphology in human cells defective for Holliday junction resolution. *Nature*, 471(7340):642–646.

Werner, A., Moutty, M.-C., Möller, U., and Melchior, F. (2009). Performing in vitro sumoylation reactions using recombinant enzymes. *SUMO Protocols*, pages 187–199.

Wiedemann, C., Bellstedt, P., and Görlach, M. (2013). CAPITO—a web server-based analysis and plotting tool for circular dichroism data. *Bioinformatics*, 29(14):1750–1757.

Williamson, J. R., Raghuraman, M., and Cech, T. R. (1989). Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*, 59(5):871–880.

Włodarczyk, A., Grzybowski, P., Patkowski, A., and Dobek, A. (2005). Effect of ions on the polymorphism, effective charge, and stability of human telomeric DNA. Photon correlation spectroscopy and circular dichroism studies. *The Journal of Physical Chemistry B*, 109(8):3594–3605.

Wolski, P., Nieszporek, K., and Panczyk, T. (2018). G-quadruplex and i-motif structures within the telomeric DNA duplex. A molecular dynamics analysis of protonation states as factors affecting their stability. *The Journal of Physical Chemistry B*, 123(2):468–479.

Wyatt, H. D., Sarbajna, S., Matos, J., and West, S. C. (2013). Coordinated actions of SLX1-SLX4 and MUS81-EME1 for Holliday junction resolution in human cells. *Molecular cell*, 52(2):234–247.

Xu, X., Zhao, P., and Chen, S.-J. (2014). Vfold: a web server for RNA structure and folding thermodynamics prediction. *PloS one*, 9(9):e107504.

Xu, Y., Noguchi, Y., and Sugiyama, H. (2006). The new models of the human telomere d[AGGG(TTAGGG)3] in K+ solution. *Bioorganic & medicinal chemistry*, 14(16):5584–5591.

Xue, Y., Kan, Z.-y., Wang, Q., Yao, Y., Liu, J., Hao, Y.-h., and Tan, Z. (2007). Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K+ solution under molecular crowding condition. *Journal of the American Chemical Society*, 129(36):11185–11191.

Yip, K. and Reed, J. (2008). Bcl-2 family proteins and cancer. *Oncogene*, 27(50):6398–6406.

Yu, H., Gu, X., Nakano, S., Miyoshi, D., and Sugimoto, N. (2012a). Beads-on-a-string structure of long telomeric DNAs under molecular crowding conditions. *Journal of the American Chemical Society*, 134(49):20060–20069.

Yu, H., Qi, Y., Yang, B., Yang, X., and Ding, Y. (2022). G4Atlas: a comprehensive transcriptome-wide G-quadruplex database. *Nucleic Acids Research*.

Yu, M., Mazor, T., Huang, H., Huang, H.-T., Kathrein, K. L., Woo, A. J., Chouinard, C. R., Labadorf, A., Akie, T. E., Moran, T. B., et al. (2012b). Direct recruitment of polycomb repressive complex 1 to chromatin by core binding transcription factors. *Molecular cell*, 45(3):330–343.

Yu, Y.-H., Chiou, G.-Y., Huang, P.-I., Lo, W.-L., Wang, C.-Y., Lu, K.-H., Yu, C.-C., Alterovitz, G., Huang, W.-C., Lo, J.-F., et al. (2012c). Network biology of tumor stem-like cells identified a regulatory role of CBX5 in lung cancer. *Scientific reports*, 2(1):1–9.

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):631–637.

Zhang, Y., Yang, M., Duncan, S., Yang, X., Abdelhamid, M. A., Huang, L., Zhang, H., Benfey, P. N., Waller, Z. A., and Ding, Y. (2019). G-quadruplex structures trigger RNA phase separation. *Nucleic Acids Research*, 47(22):11746–11754.

Zhao, J., Bacolla, A., Wang, G., and Vasquez, K. M. (2010). Non-B DNA structure-induced genetic instability and evolution. *Cellular and molecular life sciences*, 67(1):43–62.

Zhou, J., Wei, C., Jia, G., Wang, X., Feng, Z., and Li, C. (2010). Formation of i-motif structure at neutral and slightly alkaline pH. *Molecular BioSystems*, 6(3):580–586.

Zok, T., Kraszewska, N., Miskiewicz, J., Pielacinska, P., Zurkowski, M., and Szachniuk, M. (2022). ONQUADRO: a database of experimentally determined quadruplex structures. *Nucleic Acids Research*, 50(D1):D253–D258.

Zyner, K. G., Simeone, A., Flynn, S. M., Doyle, C., Marsico, G., Adhikari, S., Portella, G., Tannahill, D., and Balasubramanian, S. (2022). G-quadruplex DNA structures in human stem cells and differentiation. *Nature Communications*, 13(1):142.

Appendix A

Appendix

Table A.1: Sequence and length of oligonucleotides

Name	Sequence	Length (nt)
TERRA45	rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG	45
gTEL45	GGG TTA GGG TTA GGG TTA GGG TTA GGG TTA GGG TTA GGG	45
mutTERRA45	rGrGrG rUrUrA rGrUrG rUrUrA rGrUrG rUrUrA rGrGrG rUrUrA rGrUrG rUrUrA rGrGrG rUrUrA rGrUrG rUrUrA rGrGrG	45
Oligo G	TGG GTT AGG GTT AGG GTT AGG GTG GGT TAG GGT TAG GGT TAG GGT	45
Oligo C	TGG GTT AGG GTG GGT TAG GGT GGG TTA GGG TGG GTT AGG GT	41
fMSR	rGrGrA rCrCrU rGrGrA rArUrA rUrGrG rCrGrA rGrArA rArArC rUrGrA rArArA rUrCrA rCrGrG rArArA rArUrG rArGrA rArArU rArCrA rCrArC rUrUrU rArGrG rArCrG rUrGrA rArArU rArUrG rGrCrG rArGrG rArArA rArCrU rGrArA rArArA rGrGrU rGrGrA rArArA rUrUrU rArGrA rArArU rGrUrC rCrArC rUrGrU rArGrG rArCrG rUrGrG rArArU rArUrG rGrCrA rArGrA rArArA rCrUrG rArArA rArUrC rArUrG rGrArA rArArU rGrArG rArArA rCrArU rCrCrA rCrUrU rGrArC rGrArC rUrUrG rArArA rArArU rGrArC rGrArA rArUrC rArCrU rArArA rArArA rCrGrU rGrArA rArArA rUrGrA rGrArA rArUrG rCrArC rArCrU rGrArA	234
fMSR hairpin	rArArA rArUrC rArCrG rGrArA rArArU rGrArG rArArA rUrArC rArCrA rCrUrU rUrArG rGrArC rGrUrG rArArA rUrA	44
cTEL45	CCC TAA CCC TAA CCC TAA CCC TAA CCC TAA CCC TAA CCC	45
thy45	TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT	45
src16	GGG CTG CTG GGA CTG GTT GGG AGG GCA CTC TGG GG	35

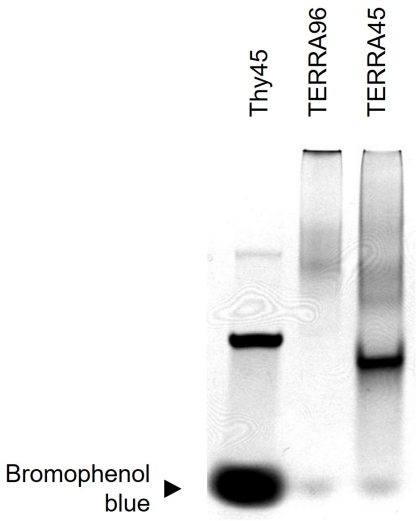

Figure A.1: Native polyacrylamide gel of TERRA oligonucleotides. Thy45, TERRA96, and TERRA45 oligonucleotides separated by native 20% PAGE, stained using 0.35% Stains-All (Merck).

```
Homo sapiens  MGKKTKRTADSSSSEDEEEYVVEKVLDRRVVKGQVEYLLKWKGFSEEHNTWEPEKNLDCP  60
Mus musculus  MGKKTKRTADSSSSEDEEEYVVEKVLDRRMVKGQVEYLLKWKGFSEEHNTWEPEKNLDCP  60
              ************************:***************************

Homo sapiens  ELISEFMKKYKKMKEGENNKPREKSESNKRKSNFSNSADDIKSKKKREQSNDIARGFERG  120
Mus musculus  ELISEFMKKYKKMKEGENNKPREKSEGNKRKSSFSNSADDIKSKKKREQSNDIARGFERG  120
              ************************.*****.*************************

Homo sapiens  LEPEKIIGATDSCGDLMFLMKWKGTDEADLVLAKEANVKCPQIVIAFYEERLTWHAYPED  180
Mus musculus  LEPEKIIGATDSCGDLMFLMKWKDTDEADLVLAKEANVKCPQIVIAFYEERLTWHAYPED  180
              ***********************.**********************************

Homo sapiens  AENKEKETAKS   191
Mus musculus  AENKEKESAKS   191
              *******:***
```

Figure A.2: Alignment of human and mouse HP1α. Aligned amino acid sequences of *Homo sapiens* and *Mus musculus* HP1α. An asterisk (*) indicates a fully conserved residue, a colon (:) indicates conservation of a residue with strongly similar properties, and a period (.) indicates conservation of a weakly similar residue.

Table A.2: Primers used for HP1α site-directed mutagenesis

Mutation	Direction	Sequence
68-72KKYKK>AAA	forward	CTA ATT TCT GAG TTT ATG GCA GCG TAT GCG GCG ATG AAG GAG GGT GAA AAC AAT AAG CCC
	reverse	GGG CTT ATT GTT TTC ACC CTC CTT CAT CGC CGC ATA CGC TGC CAT AAA CTC AGA AAT TAG
89-91KRK>AAA	forward	GCG CTG TTG GAG AAA CTG GAT GCC GCC GCG TTT CCT TCT GAT TTC TC
	reverse	GAG AAA TCA GAA GGA AAC GCG GCG GCA TCC AGT TTC TCC AAC AGC GC
104-106KKK>AAA	forward	CGA TAT CAT TGC TTT GCT CTC TCG CTG CTG CAG ATT TAA TAT CAT CAG C
	reverse	GCT GAT GAT ATT AAA TCT GCA GCA GCG AGA GAG CAA AGC AAT GAT ATC G
89-91KRK>RRR	forward	CAG AAG GAA ACA GGA GGA GAT CCA GTT TCT CC
	reverse	GGA GAA ACT GGA TCT CCT CCT GTT TCC TTC TG
104-106KKK>RRR	forward	GCG CTG ATG ATA TTA AAT CTA GAA GAA GGA GAG AGC AAA GC
	reverse	GCT TTG CTC TCT CCT TCT TCT AGA TTT AAT ATC ATC AGC GC
S93E	forward	CAG CGC TGT TGG AGA ACT CGG ATT TCC TCT TGT TTC CTT C
	reverse	GAA GGA AAC AAG AGG AAA TCC GAG TTC TCC AAC AGC GCT G
S97E	forward	GAT TTA ATA TCA TCA GCC TCG TTG GAG AAA CTG G
	reverse	CCA GTT TCT CCA ACG AGG CTG ATG ATA TTA AAT C
S11-14E	forward	CCA CAT ATT CCT CCT CAT CCT CTT CCT CCT CCT CGT CGG CTG TCC TCT TGG
	reverse	CCA AGA GGA CAG CCG ACG AGG AGG AGG AAG AGG ATG AGG AGG AAT ATG TGG
V22M	forward	GTC CAA CAC CTT TTC CAT CAC ATA TTC CTC CTC ATC CTC
	reverse	GAG GAT GAG GAG GAA TAT GTG ATG GAA AAG GTG TTG GAC
Y168E	forward	CCA CAG ATT GTG ATA GCA TTT GAG GAA GAG AGA CTG ACG TGG CAC GC
	reverse	GCG TGC CAC GTC AGT CTC TCT TCC TCA AAT GCT ATC ACA ATC TGT GG
L139D	forward	CTG TGT CTT TCC ATT TCA TAT CGA ACA TTA AGT CAC CGC
	reverse	GCG GTG ACT TAA TGT TCG ATA TGA AAT GGA AAG ACA CAG
K154A	forward	GCT GAC CTG GTT CTT GCA GCA GAA GCT AAC GTG AAG
	reverse	CTT CAC GTT AGC TTC TGC TGC AAG AAC CAG GTC AGC
E155V	forward	CCT GGT TCT TGC AAA AGT AGC TAA CGT GAA GTG TCC
	reverse	GGA CAC TTC ACG TTA GCT ACT TTT GCA AGA ACC AGG
V158D	forward	GCA AAA GAA GCT AAC GAT AAG TGT CCA CAG ATT GTG
	reverse	CAC AAT CTG TGG ACA CTT ATC GTT AGC TTC TTT TGC
V158G	forward	GCA AAA GAA GCT AAC GGG AAG TGT CCA CAG ATT GTG
	reverse	CAC AAT CTG TGG ACA CTT CCC GTT AGC TTC TTT TGC
V158W	forward	GCA AAA GAA GCT AAC TGG AAG TGT CCA CAG ATT GTG
	reverse	CAC AAT CTG TGG ACA CTT CCA GTT AGC TTC TTT TGC

Table A.3: Primers used for HP1α inverse PCR deletions

Deletion	**Direction**	**Sequence**
NTE	forward	TAT GTG GTG GAA AAG GTG TTG GAC AGG CGC
	reverse	GGA TCC CAT GGC GCC CTG AAA ATA CAG G
NTE+CD	forward	AAA AAG TAT AAG AAG ATG AAG GAG GGT GAA AAC AAT AAG CCC
	reverse	GCC CTG AAA ATA CAG GTT TTC GGT CGT TGG
CSD+CTE	forward	TAA AAG CTT GCG GCC GCA CTC
	reverse	CCG AGC GAT ATC ATT GCT TTG CTC TCT C
CTE	forward	TAA AAG CTT GCG GCC GCA CTC GAG C
	reverse	ATC CTC TGG ATA TGC GTG CCA CGT CAG
Hinge	forward	GGC TTT GAG AGA GGA CTG GAA CCA GAA AAG ATC ATC G
	reverse	GCC CTG AAA ATA CAG GTT TTC GGT CGT TGG

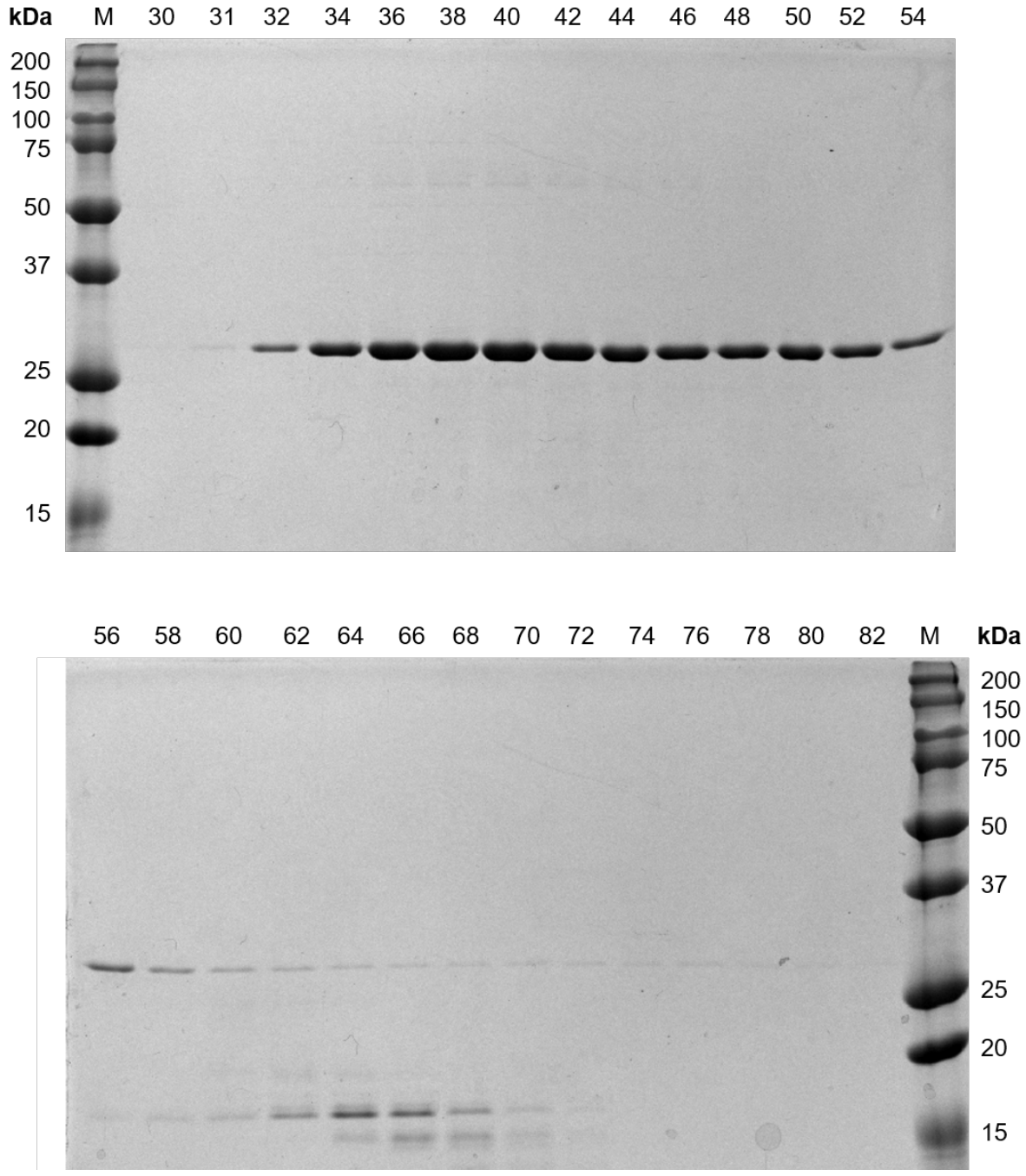

Figure A.3: Size exclusion chromatography purification of HP1α. Fractions of 2 µL from SEC with a Superdex 200 10/300 GL column electrophoresed on a 12% SDS-PAGE, stained with Coomassie. Fractions 34 to 56 were pooled.

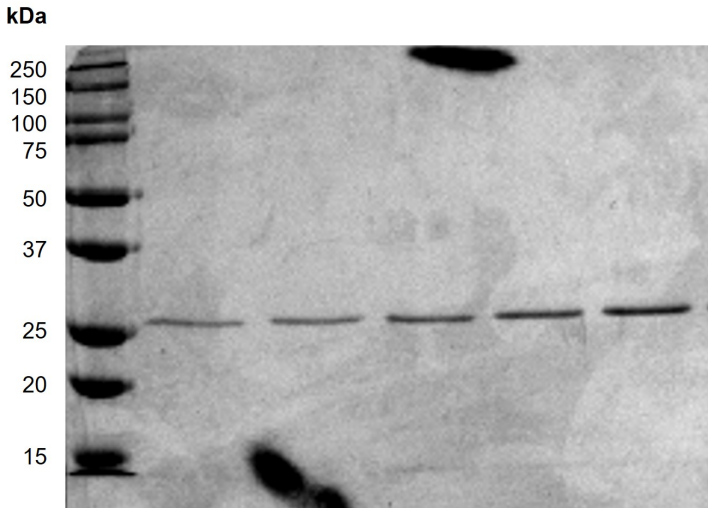

Figure A.4: Purified HP1α A1 protein. Fractions of HP1α 68-72KKYKK>AAYAA (A1) electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

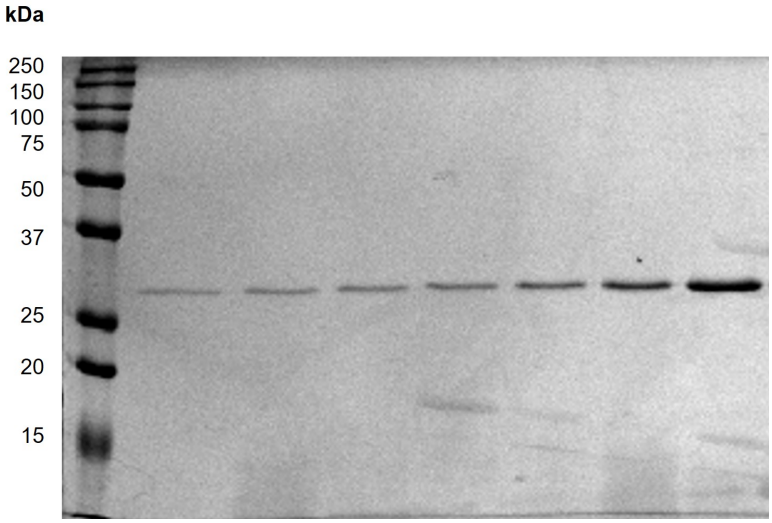

Figure A.5: Purified HP1α A2 protein. Fractions of HP1α 89-91KRK>AAA (A2) electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

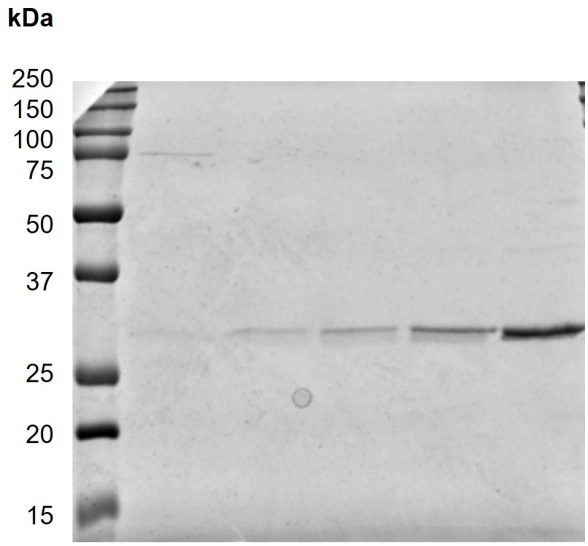

Figure A.6: Purified HP1α A3 protein. Fractions of HP1α 104-106KKK>AAA (A3) electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

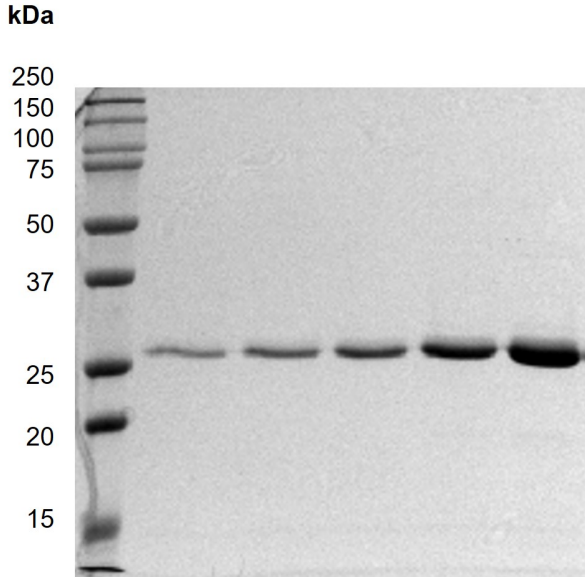

Figure A.7: Purified HP1α A1;2 protein. Fractions of HP1α 68-72KKYKK>AAYAA;89-91KRK>AAA (A1;2) electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

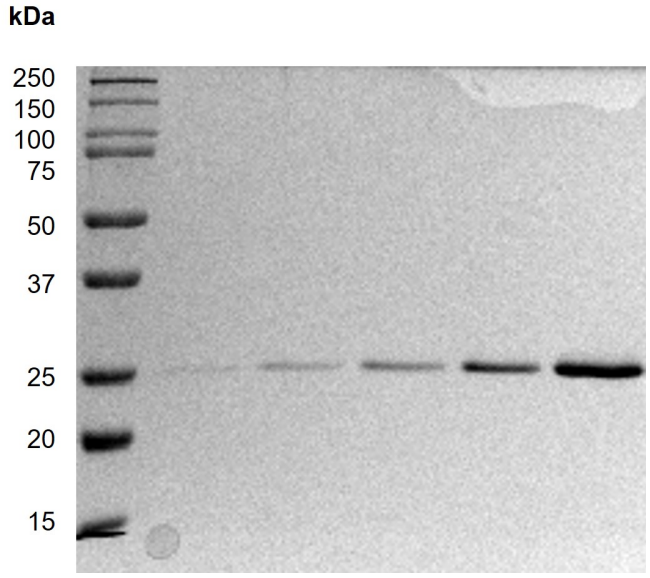

Figure A.8: Purified HP1α A1;3 protein. Fractions of HP1α 68-72KKYKK>AAYAA;104-106KKK>AAA (A1;3) electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

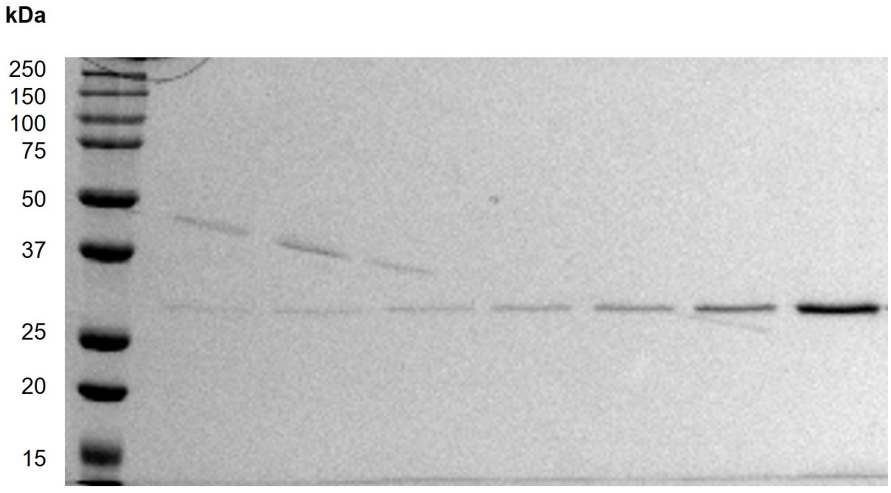

Figure A.9: Purified HP1α A2;3 protein. Fractions of HP1α 89-91KRK>AAA;104-106KKK>AAA (A2;3) electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

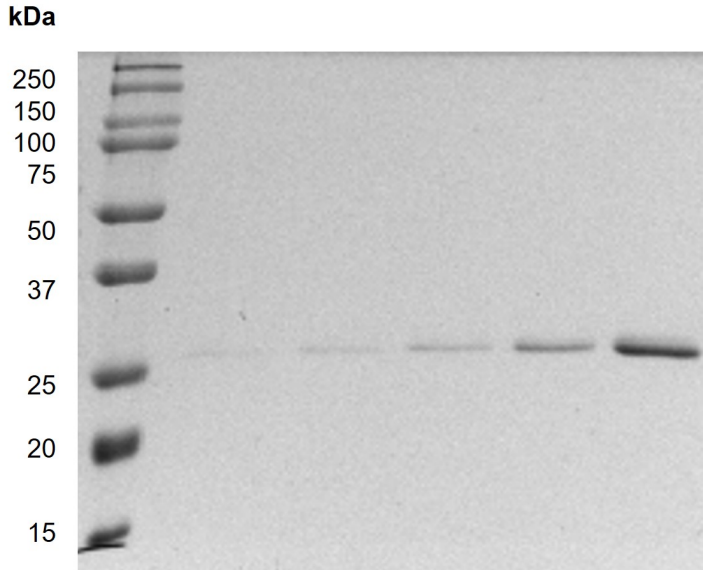

Figure A.10: Purified HP1α A1;2;3 protein. Fractions of HP1α 68-72KKYKK>AAYAA;89-91KRK>AAA;104-106KKK>AAA (A1;2;3) electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

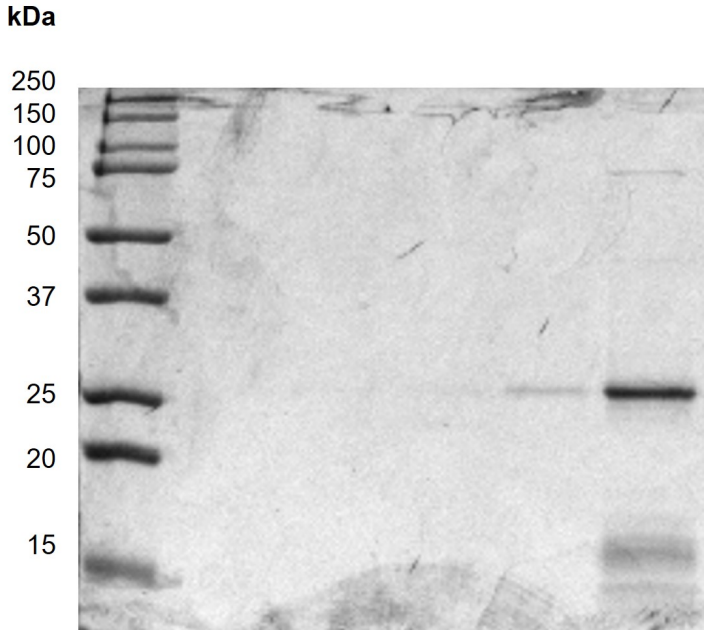

Figure A.11: Purified HP1α R2 protein. Fractions of HP1α 89-91KRK>RRR (R2) electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue. Cloning, expression, and purification of this mutant were performed by Elizabeth V. Chernysheva.

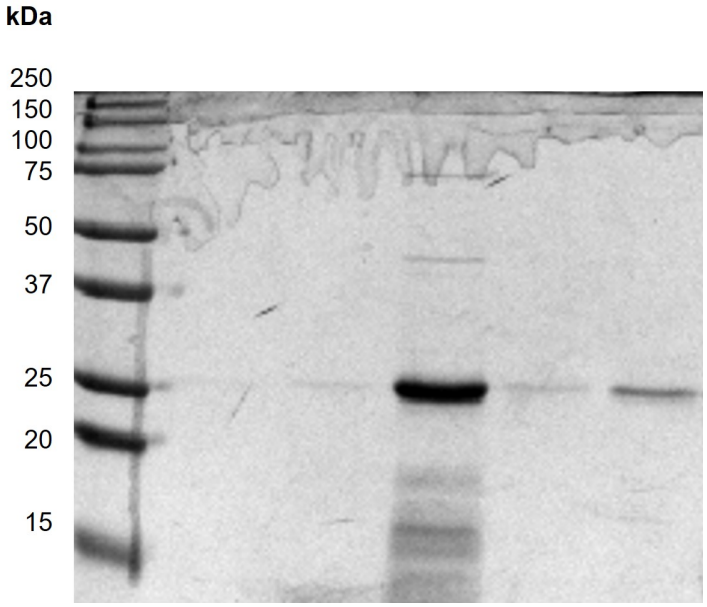

Figure A.12: Purified HP1α R3 protein. Fractions of HP1α 104-106KKK>RRR (R3) electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue. Cloning, expression, and purification of this mutant were performed by Elizabeth V. Chernysheva.

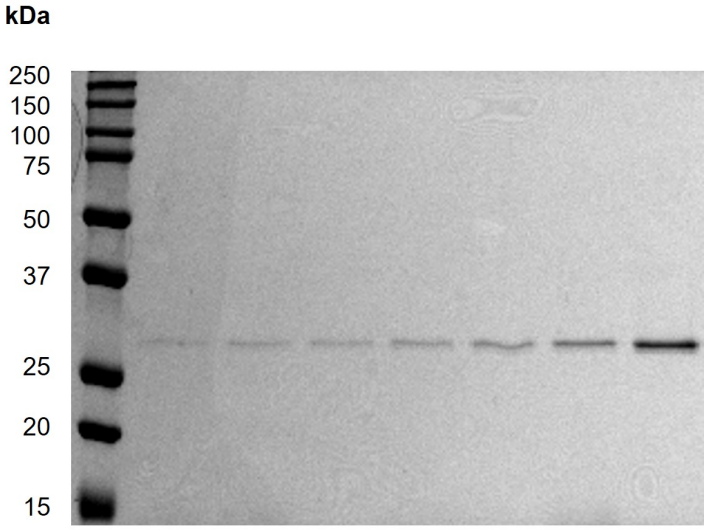

Figure A.13: Purified HP1α S93E protein. Fractions of HP1α S93E electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

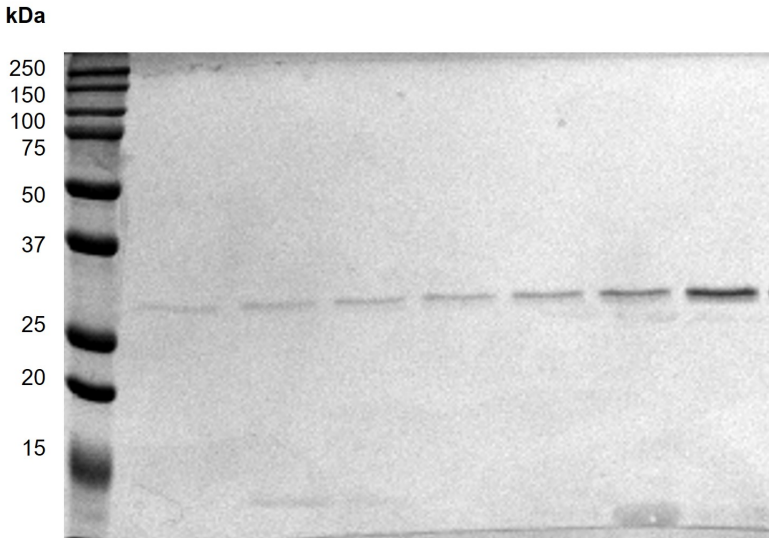

Figure A.14: Purified HP1α S97E protein. Fractions of HP1α S97E electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

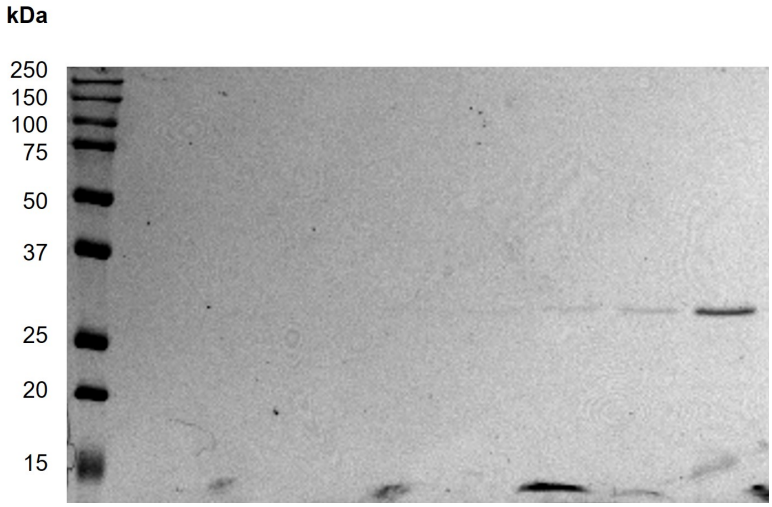

Figure A.15: Purified HP1α S11-14E protein. Fractions of HP1α S11-14E electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

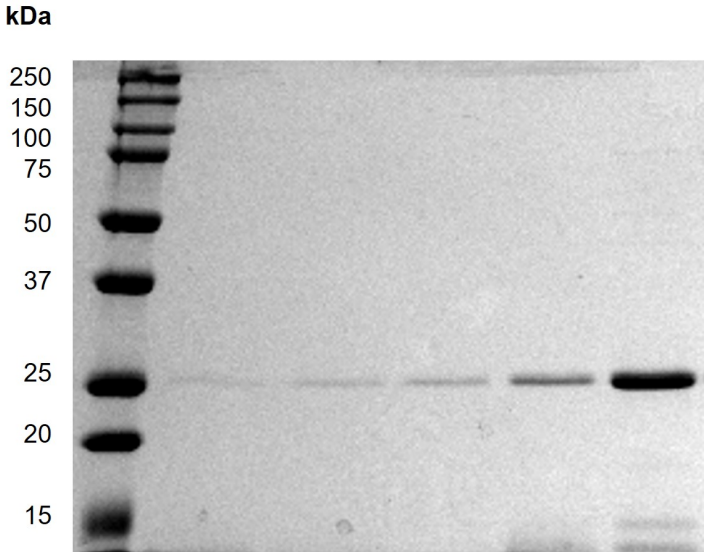

Figure A.16: Purified HP1α ΔNTE protein. Fractions of HP1α ΔNTE electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

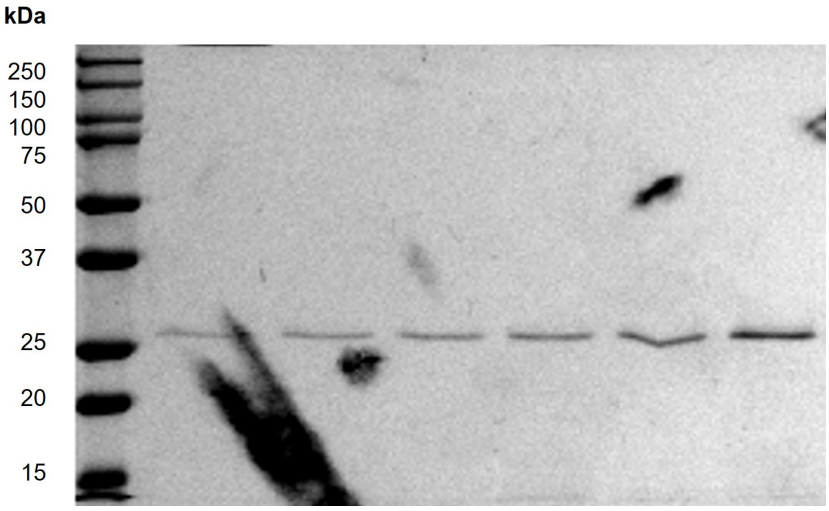

Figure A.17: Purified HP1α V22M protein. Fractions of HP1α V22M electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

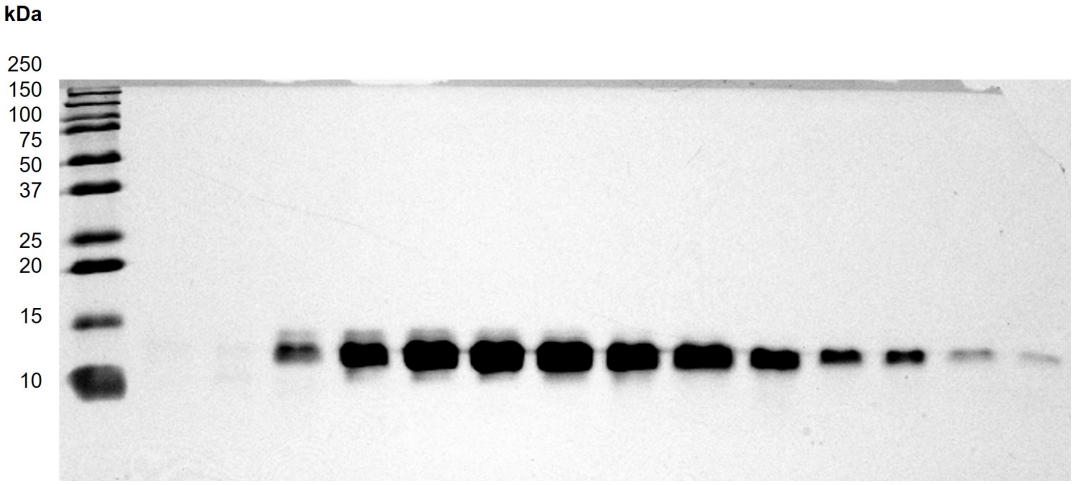

Figure A.18: Purified αHinge protein. Fractions of αHinge electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

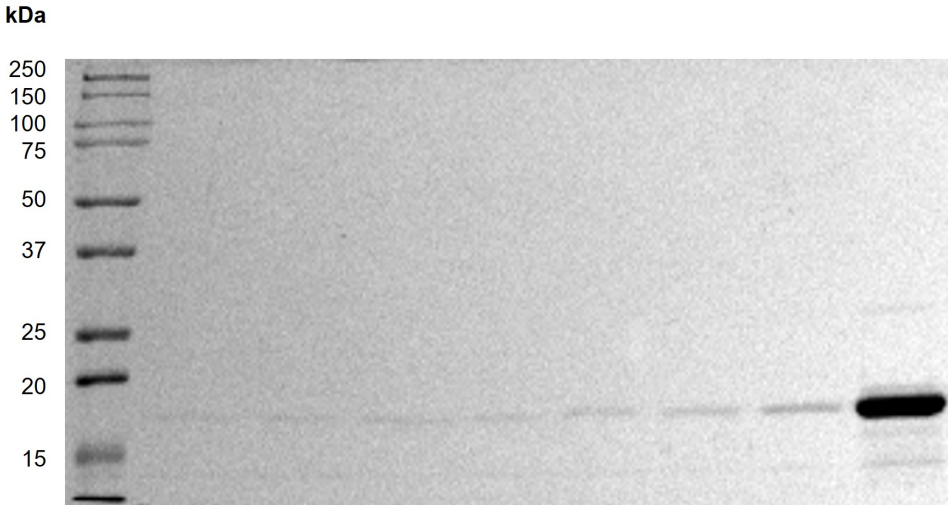

Figure A.19: Purified HP1α NTE+CD+Hinge protein. Fractions of HP1α NTE+CD+Hinge electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

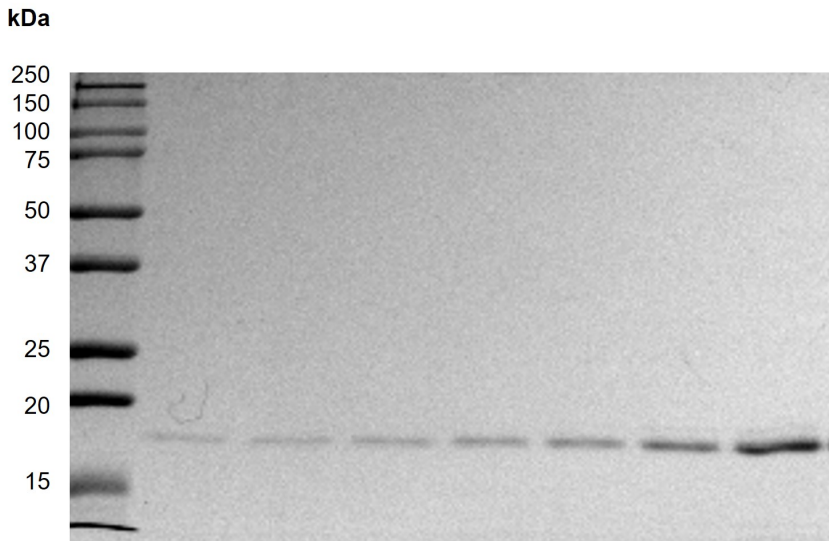

Figure A.20: Purified HP1α Hinge+CSD+CTE protein. Fractions of HP1α Hinge+CSD+CTE electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

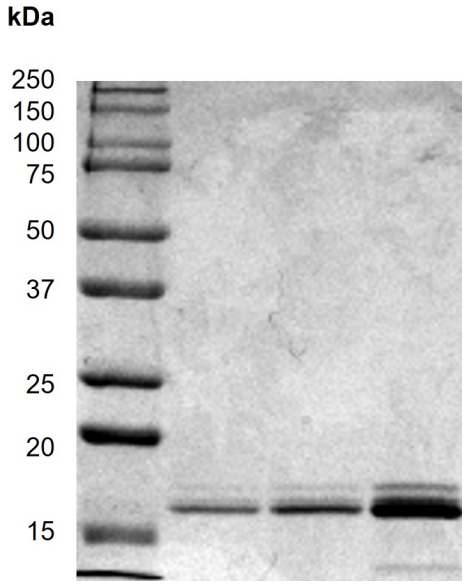

Figure A.21: Purified HP1α Hinge+CSD protein. Fractions of HP1α Hinge+CSD electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

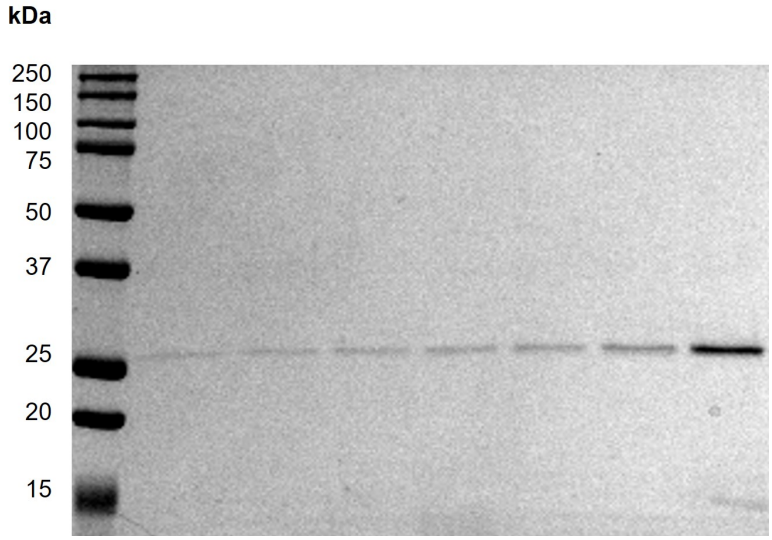

Figure A.22: Purified HP1α ΔCTE protein. Fractions of HP1α ΔCTE electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

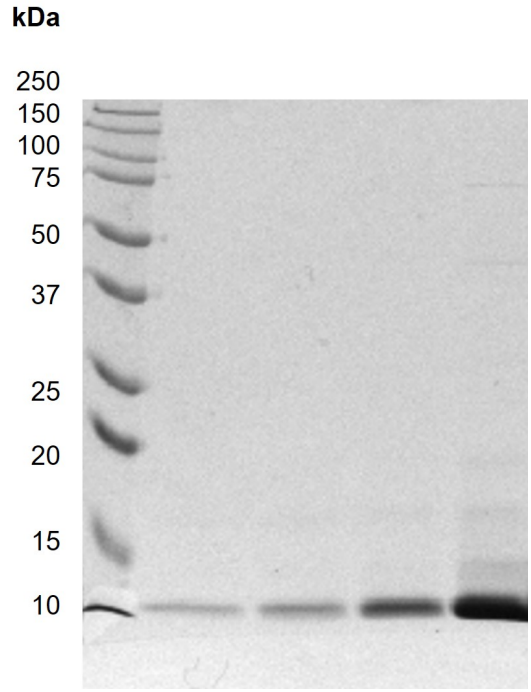

Figure A.23: Purified HP1α CSD protein. Fractions of HP1α CSD electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

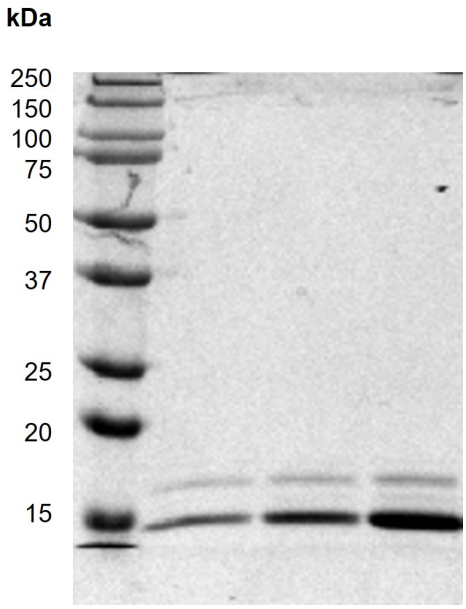

Figure A.24: Purified HP1α CSD+CTE protein. Fractions of HP1α CSD+CTE electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

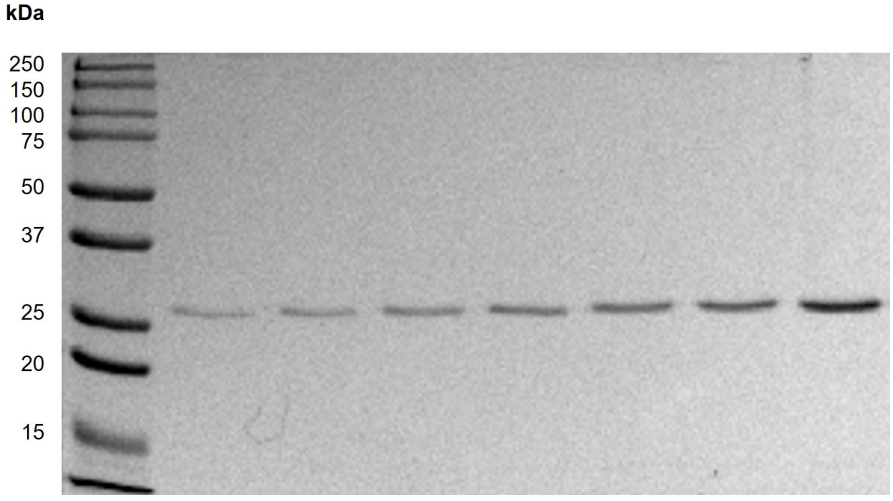

Figure A.25: Purified HP1α Y168E protein. Fractions of HP1α Y168E electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

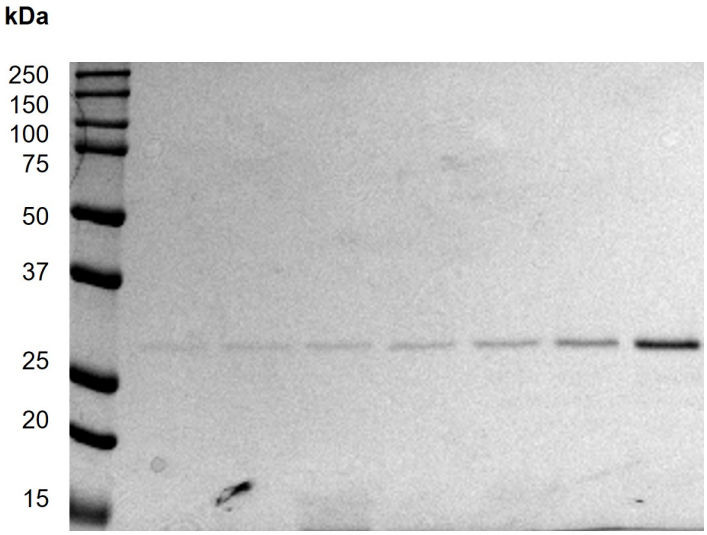

Figure A.26: Purified HP1α L139D protein. Fractions of HP1α L139D electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

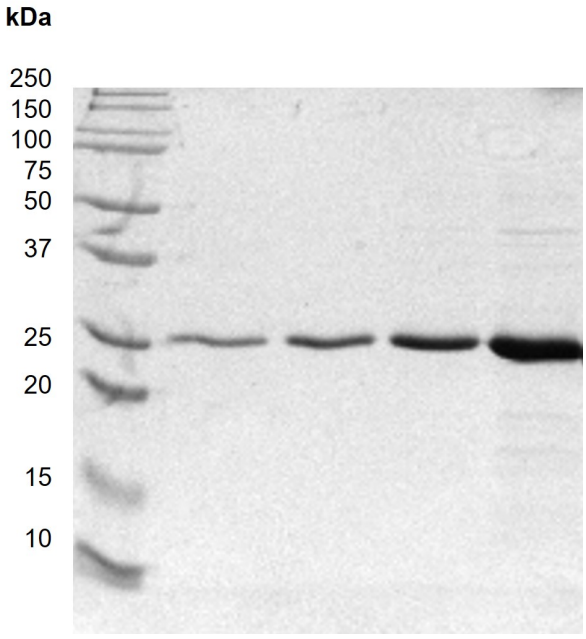

Figure A.27: Purified HP1α K154A protein. Fractions of HP1α K154A electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

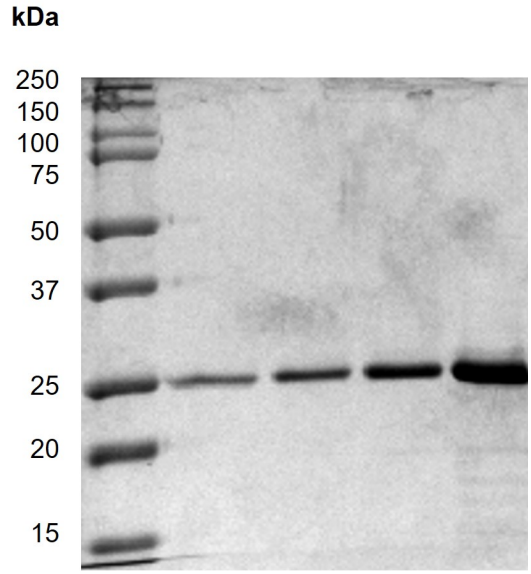

Figure A.28: Purified HP1α E155V protein. Fractions of HP1α E155V electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

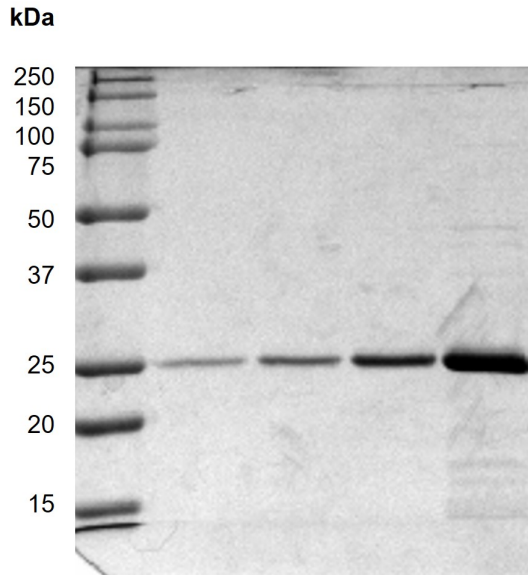

Figure A.29: Purified HP1α V158D protein. Fractions of HP1α V158D electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

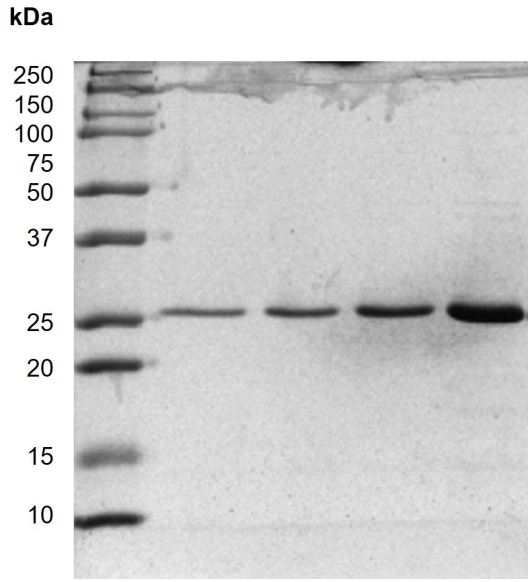

Figure A.30: Purified HP1α V158G protein. Fractions of HP1α V158G electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

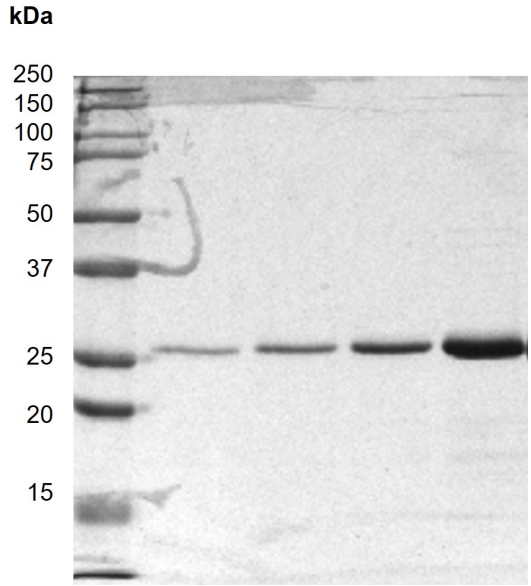

Figure A.31: Purified HP1α V158W protein. Fractions of HP1α V158W electrophoresed on a 12% SDS-PAGE, stained with Coomassie blue.

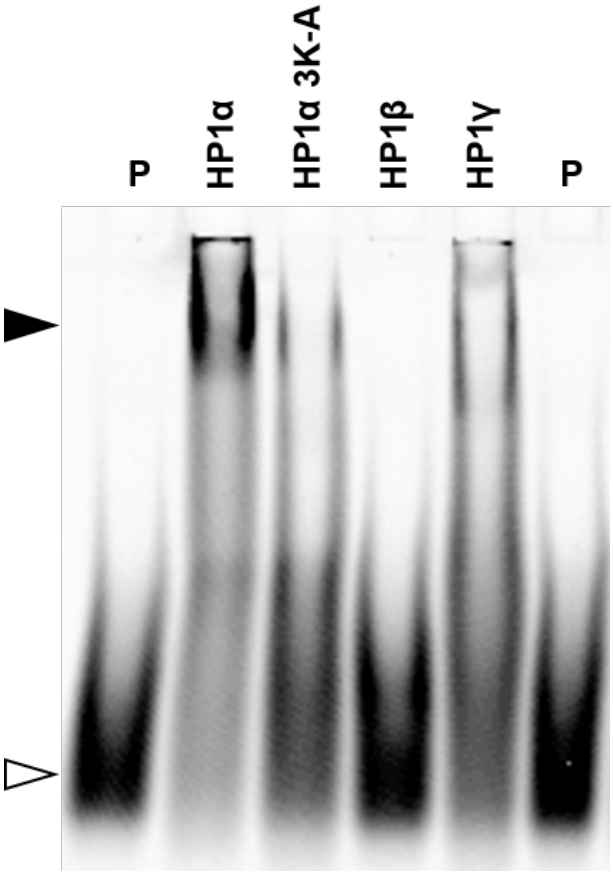

Figure A.32: Electrophoretic mobility shift assay of HP1 paralogs with TAM-TERRA45 (full image). TAM-TERRA45 in the absence (P) or presence of a 20-fold molar excess of the indicated HP1 protein. Open arrowhead denotes unbound TAM-TERRA45 probe, closed arrowhead denotes complex.

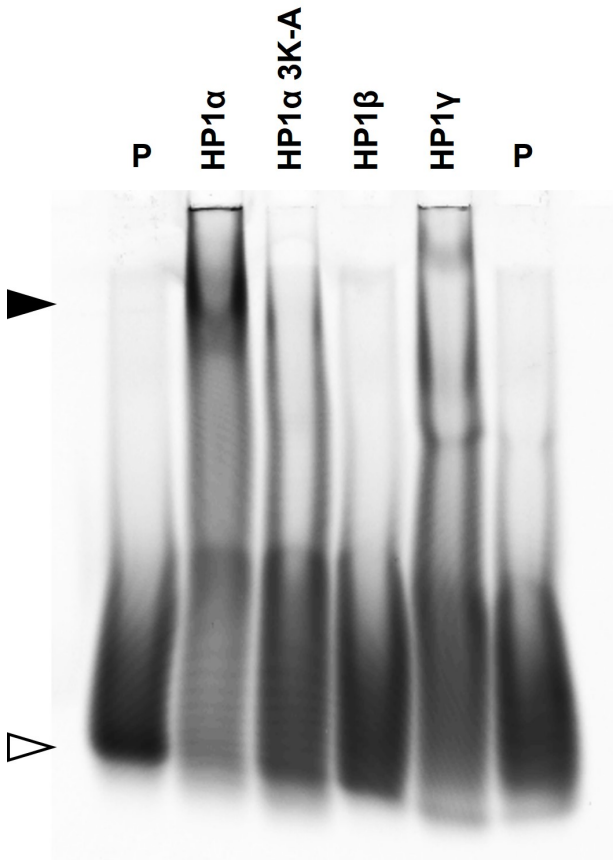

Figure A.33: Electrophoretic mobility shift assay of HP1 paralogs with TAM-TERRA45 (repeat). TAM-TERRA45 in the absence (P) or presence of a 20-fold molar excess of the indicated HP1 protein. Open arrowhead denotes unbound TAM-TERRA45 probe, closed arrowhead denotes complex.

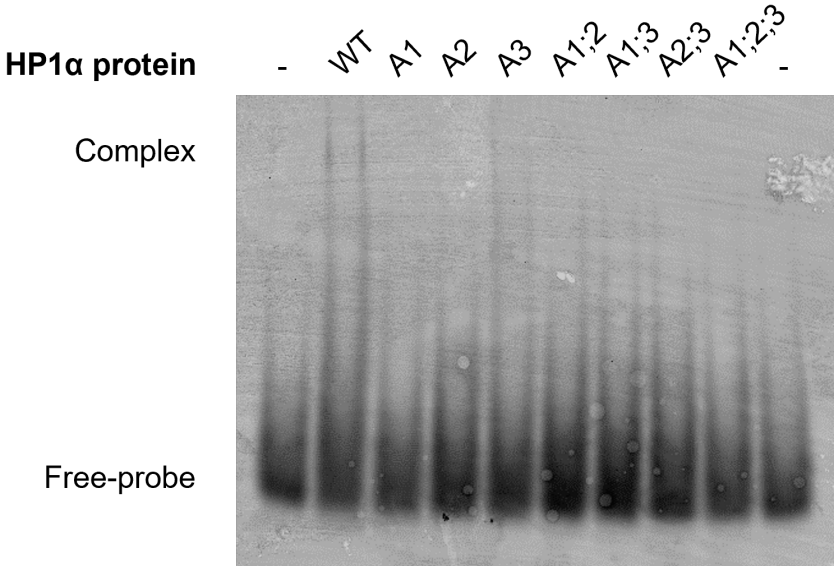

Figure A.34: Electrophoretic mobility shift assay of HP1α WT and hinge patch mutants with TAM-TERRA45 (repeat). EMSA with TAM-TERRA45 in the absence (-) or presence of a 20-fold excess of WT HP1α or hinge patch mutant.

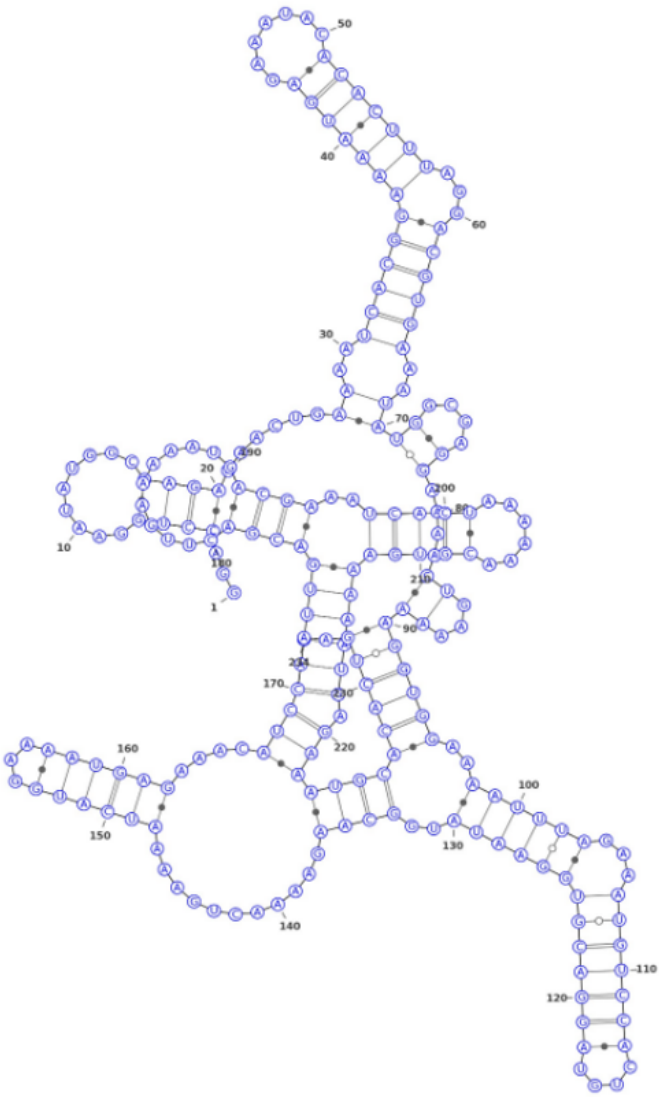

Figure A.35: Potential hairpin formation by the forward MSR RNA sequence. To test whether the αHinge binds fMSR RNA, an oligonucleotide of comparative length to TERRA45 must be selected. The Vfold2D (Xu et al., 2014) prediction tool can be used to assess possible RNA hairpin formation in the fMSR sequence (**Table A.1**), which found a structure rich in potential hairpins. The hairpin from nucleotide 27 to nucleotide 70 (44 nt) is predicted to form with a free energy of -9.49 kcal/mol and was selected for analysis.

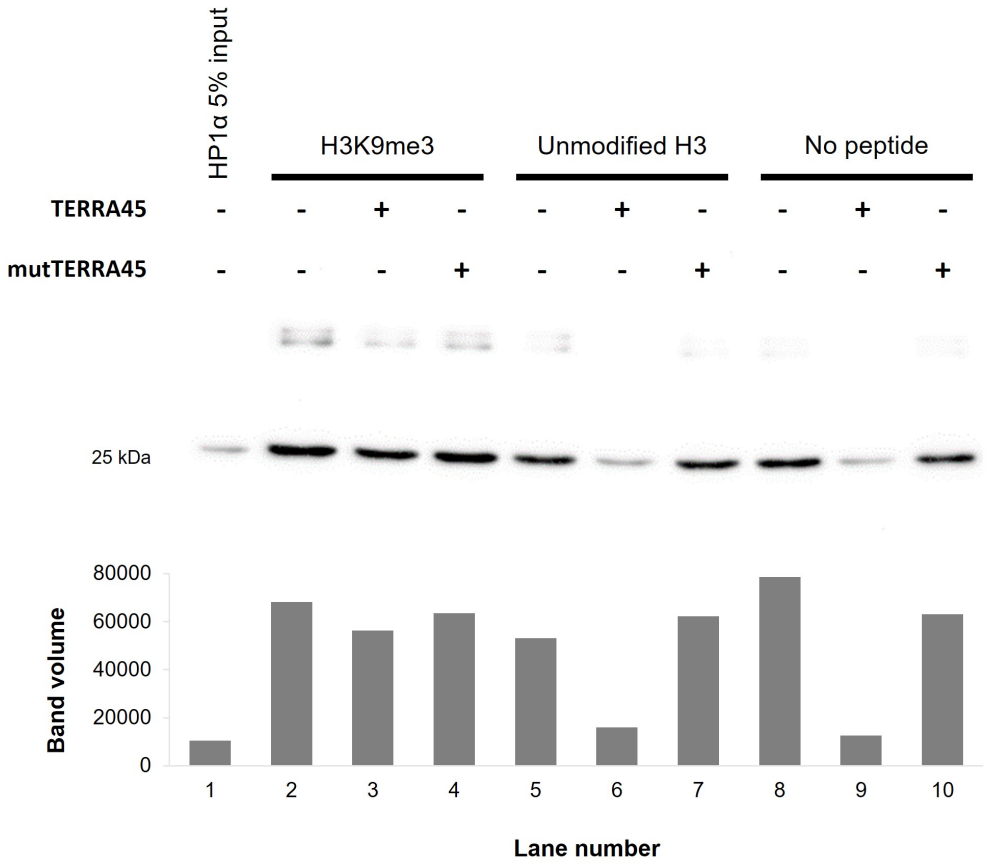

Figure A.36: Peptide pulldown assay of HP1α binding to histone H3 with TERRA using magnetic streptavidin beads. Biotinylated H3K9me3 or unmodified H3 peptide was used to pull down HP1α in the presence or absence of TERRA45 or mutTERRA45, and samples were subjected to SDS-PAGE and western blotting using HP1α antibody (2616, Cell Signaling Technology). Band densitometry is shown below.

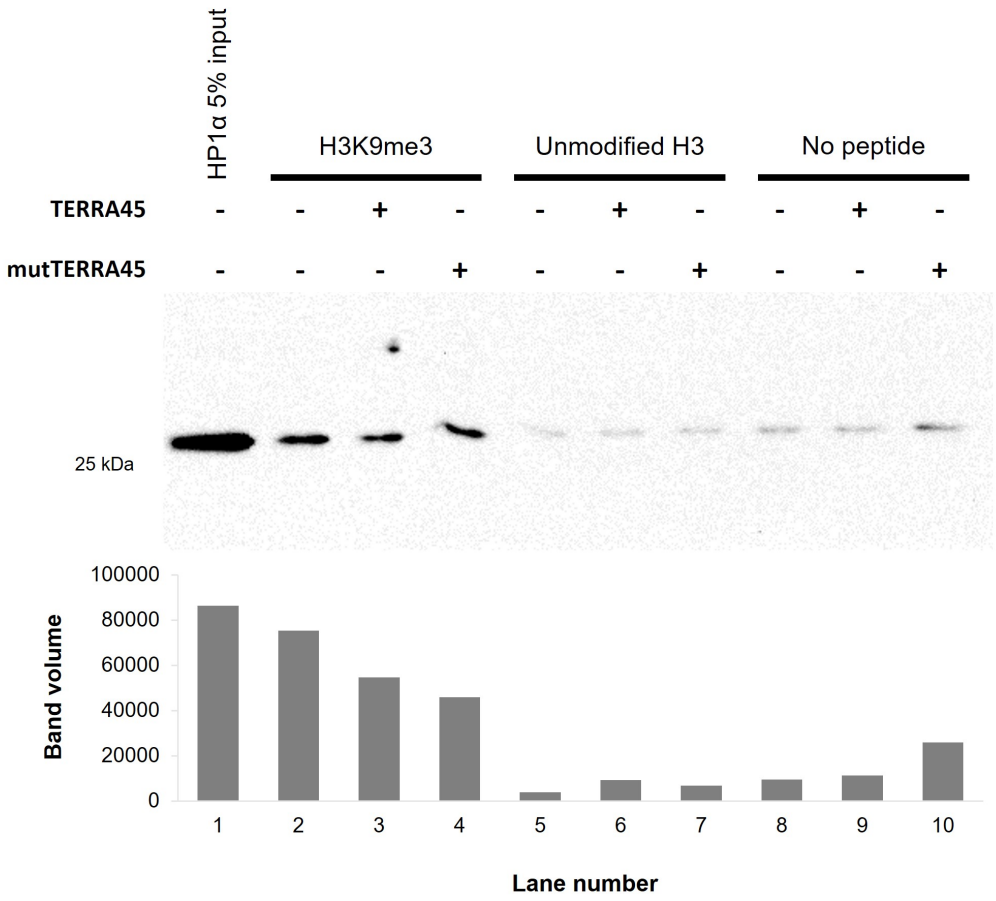

Figure A.37: Peptide pulldown assay of HP1α binding to histone H3 with TERRA using agarose streptavidin beads. Biotinylated H3K9me3 or unmodified H3 peptide was used to pull down HP1α in the presence or absence of TERRA45 or mutTERRA45, and samples were subjected to SDS-PAGE and western blotting using HP1α antibody (2616, Cell Signaling Technology). Band densitometry is shown below.

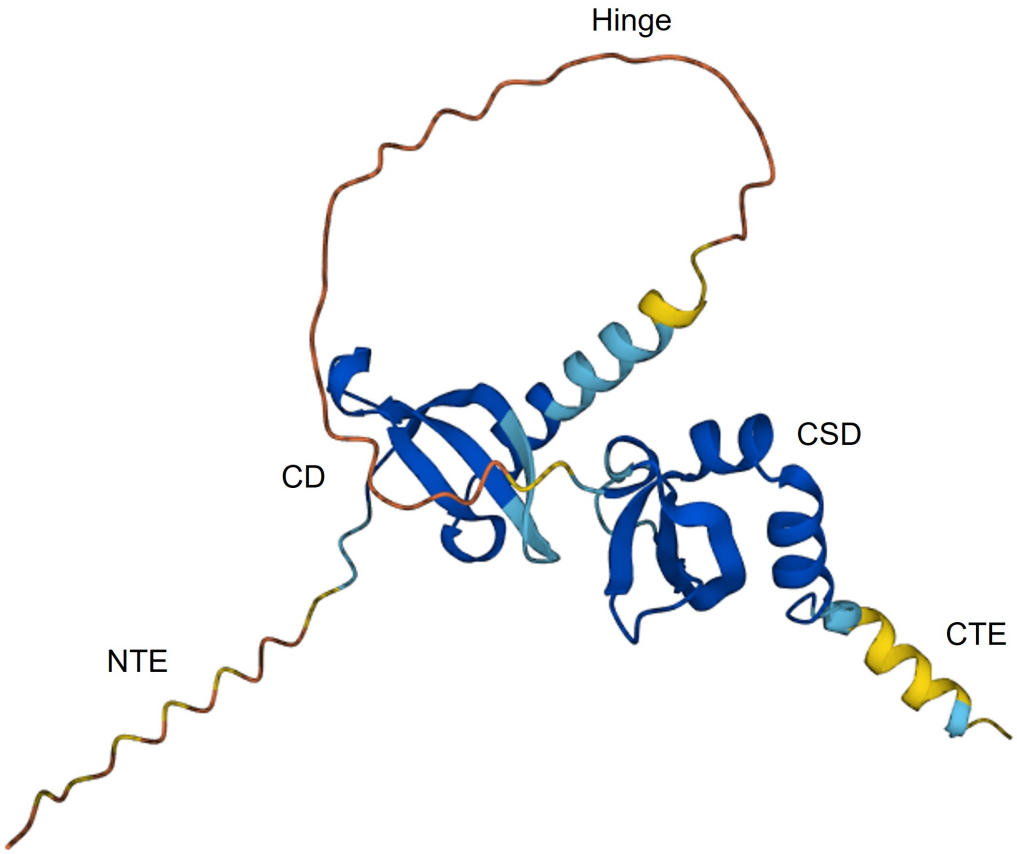

Figure A.38: AlphaFold computed structure model of HP1α. AlphaFold model AF-Q61686F1 of *Mus musculus* Chromobox protein homolog 5 (CBX5, HP1α), shown with the N-terminus on the left, and C-terminus on the right. Predicted Local Distance Difference Test (pLDDT) scores of each amino acid are shown in different colours, with very high confidence (pLDDT > 90) shown in dark blue, moderate confidence (90 > pLDDT > 70) shown in light blue, low confidence (70 > pLDDT > 50) shown in yellow, and very low confidence (pLDDT < 50) shown in orange.

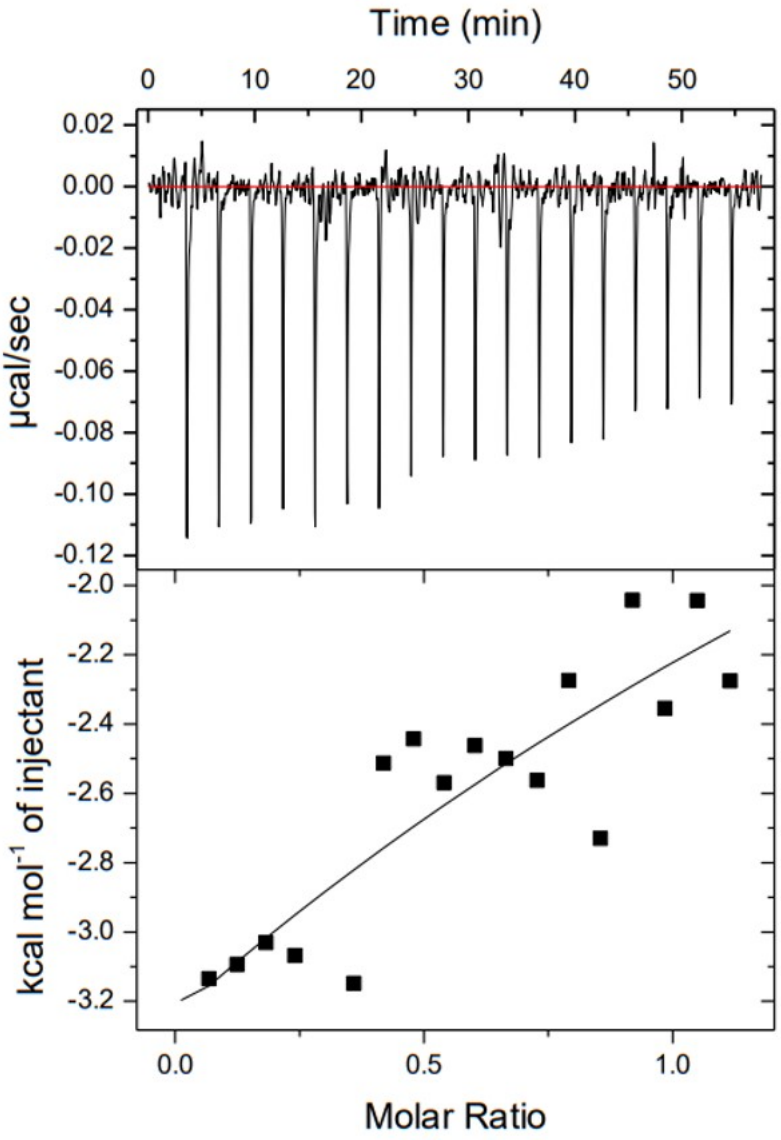

Figure A.39: ITC attempt using WT HP1α and Oligo C. ITC of WT HP1α (30 µM in cell) with Oligo C (170 µM in syringe). The top graph displays raw data after the integration baseline has been subtracted. The bottom graph displays the normalised integration data. Lack of data at low molar ratio of Oligo C : HP1α limits interpretation.

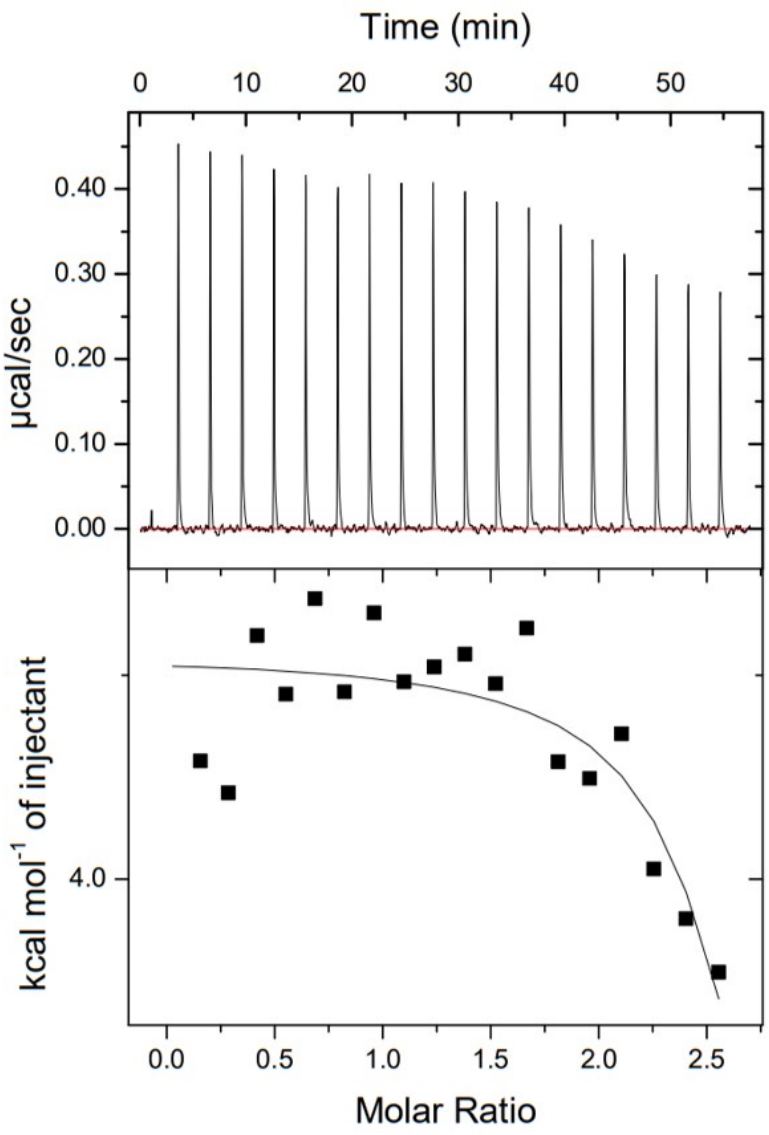

Figure A.40: ITC attempt using αHinge and Oligo C. ITC of αHinge (30 µM in cell) with Oligo C (333 µM in syringe). The top graph displays raw data after the integration baseline has been subtracted. The bottom graph displays the normalised integration data. Lack of data at high molar ratio of Oligo C : HP1α limits interpretation.

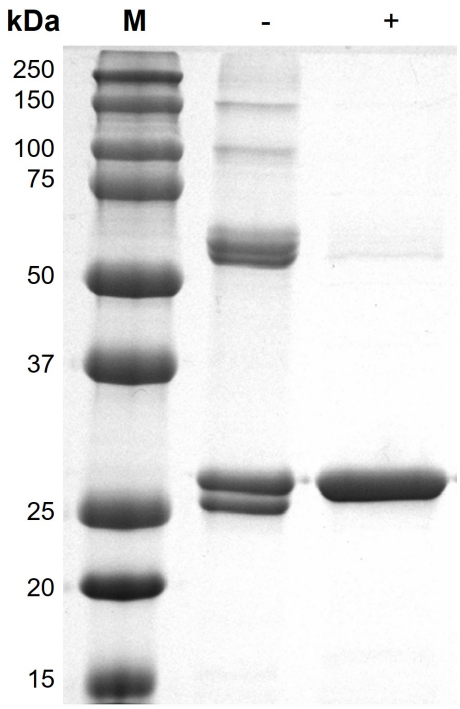

Figure A.41: SDS-PAGE of HP1α in presence and absence of 2-mercaptoethanol. HP1α contains three cysteine residues and therefore may be able to form disulfide bonds. In order to determine if either of the HP1α proteins contained disulfide bonds, an SDS-PAGE was performed, but with no reducing agent. The gel showed that in the absence of reducing agent, HP1α displayed bands of high molecular weight around 60 kDa, 100 kDa, and 150 kDa. In addition to this, two distinct bands were observed around 25 kDa. This indicated that not only does HP1α form disulfide bonds intermolecularly (accounting for higher molecular weight bands), but also intramolecularly, to create two species of slightly differing mobilities of around the same molecular weight.

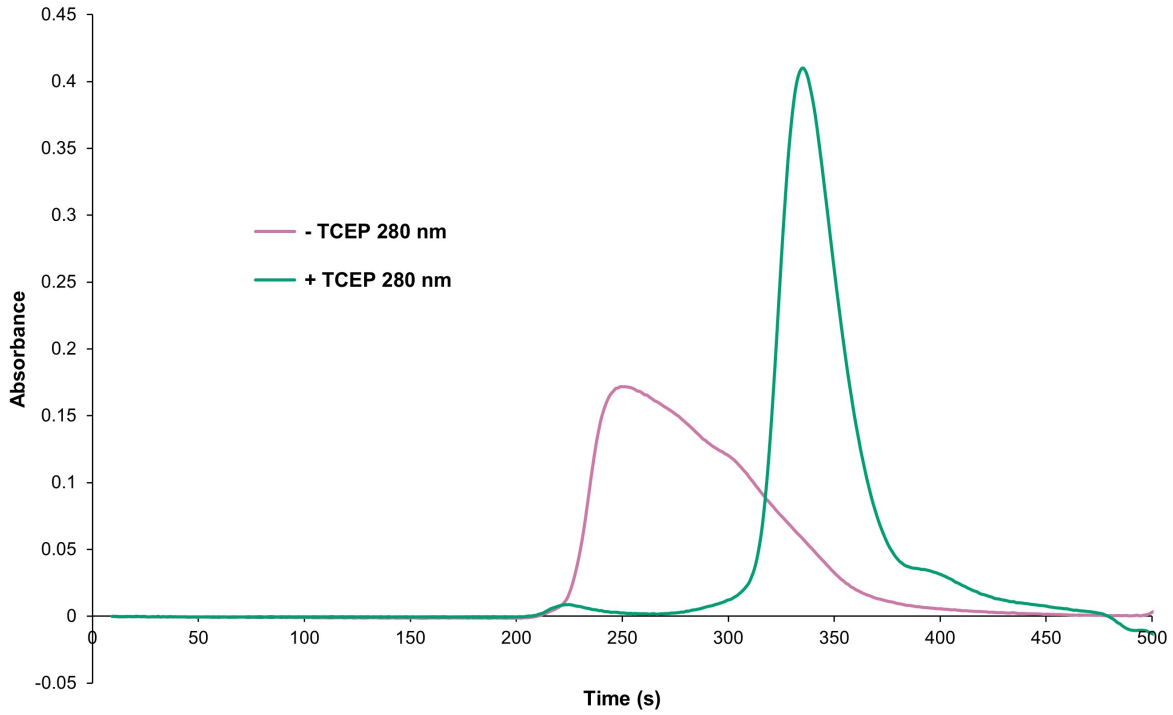

Figure A.42: Size exclusion chromatography of HP1α in the presence and absence of TCEP. The absorbance at 280 nm from the flow-through of the size-exclusion column showed a drastic shift in HP1α mobility in the presence and absence of TCEP. Addition of TCEP resulted in a more discrete peak at 280 nm, at a later retention time indicating a smaller size of molecule.

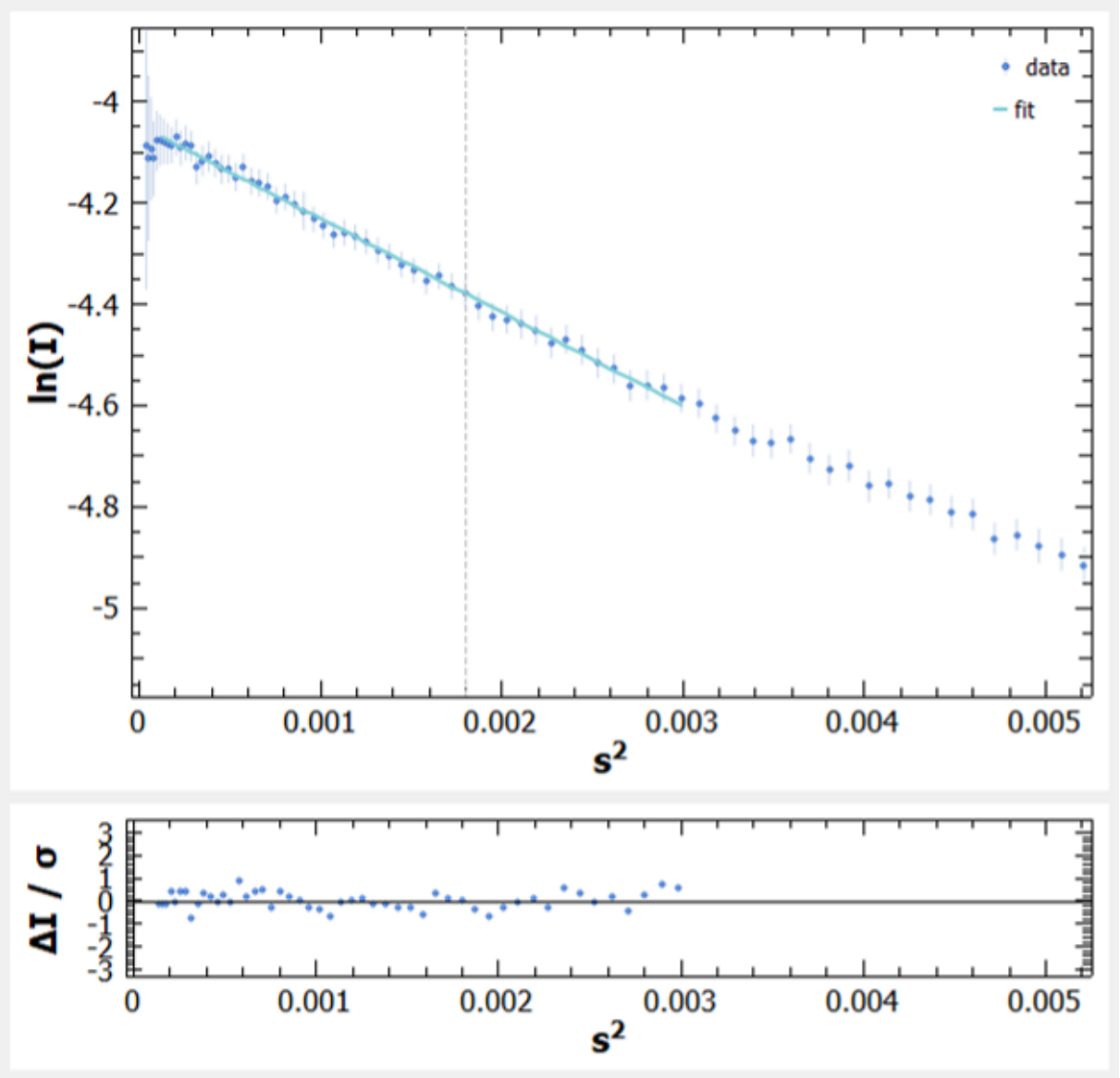

Figure A.43: Guinier fit of HP1α Hinge-only SAXS data. Linear Guinier fit to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

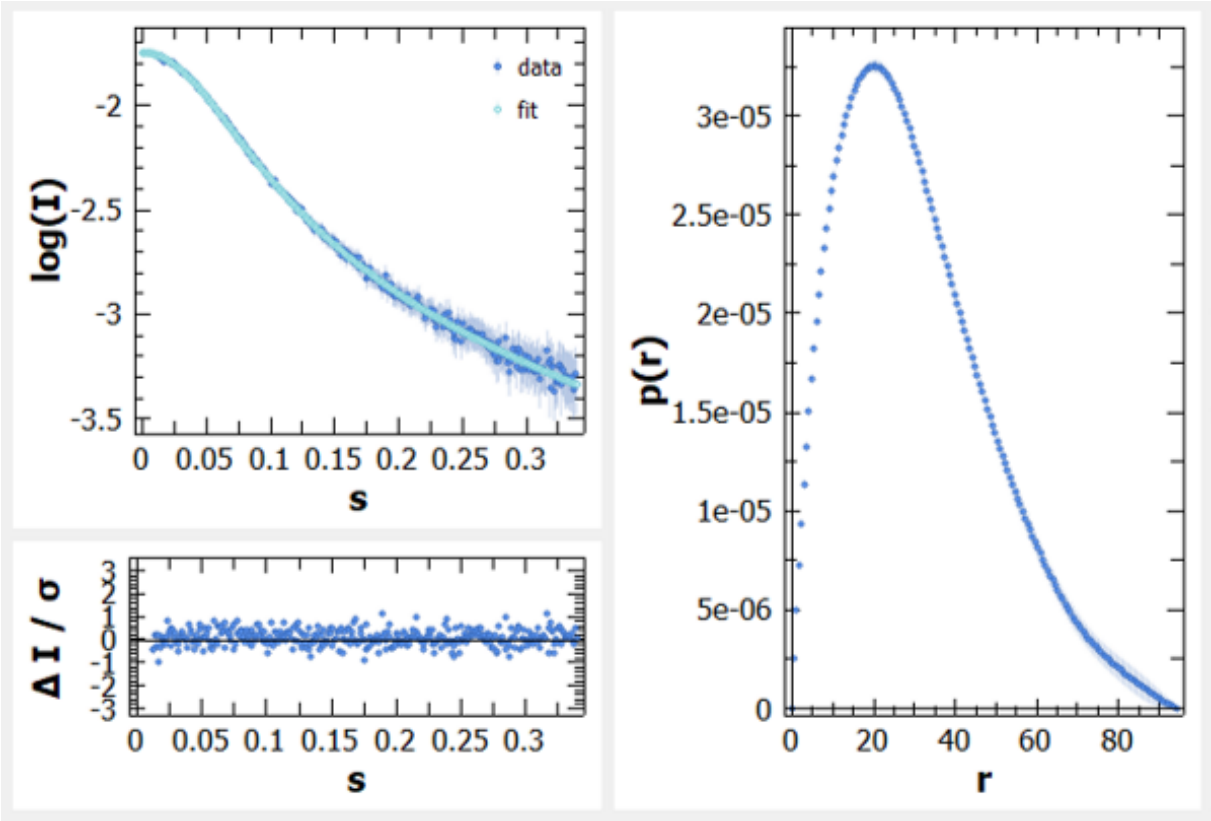

Figure A.44: *P(r)* plot of HP1α Hinge-only SAXS data. *P(r)* versus r profile from HP1α Hinge-only SAXS data. A fit of the back-transformed *P(r)* curve to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

Figure A.45: Example of bead model iterations. Eight DAMMIN-generated models of HP1α Y168E are shown in different colours, which are then aligned and overlaid. Of the ten generated models, two have been automatically excluded due to being more than two standard deviations from the mean NSD.

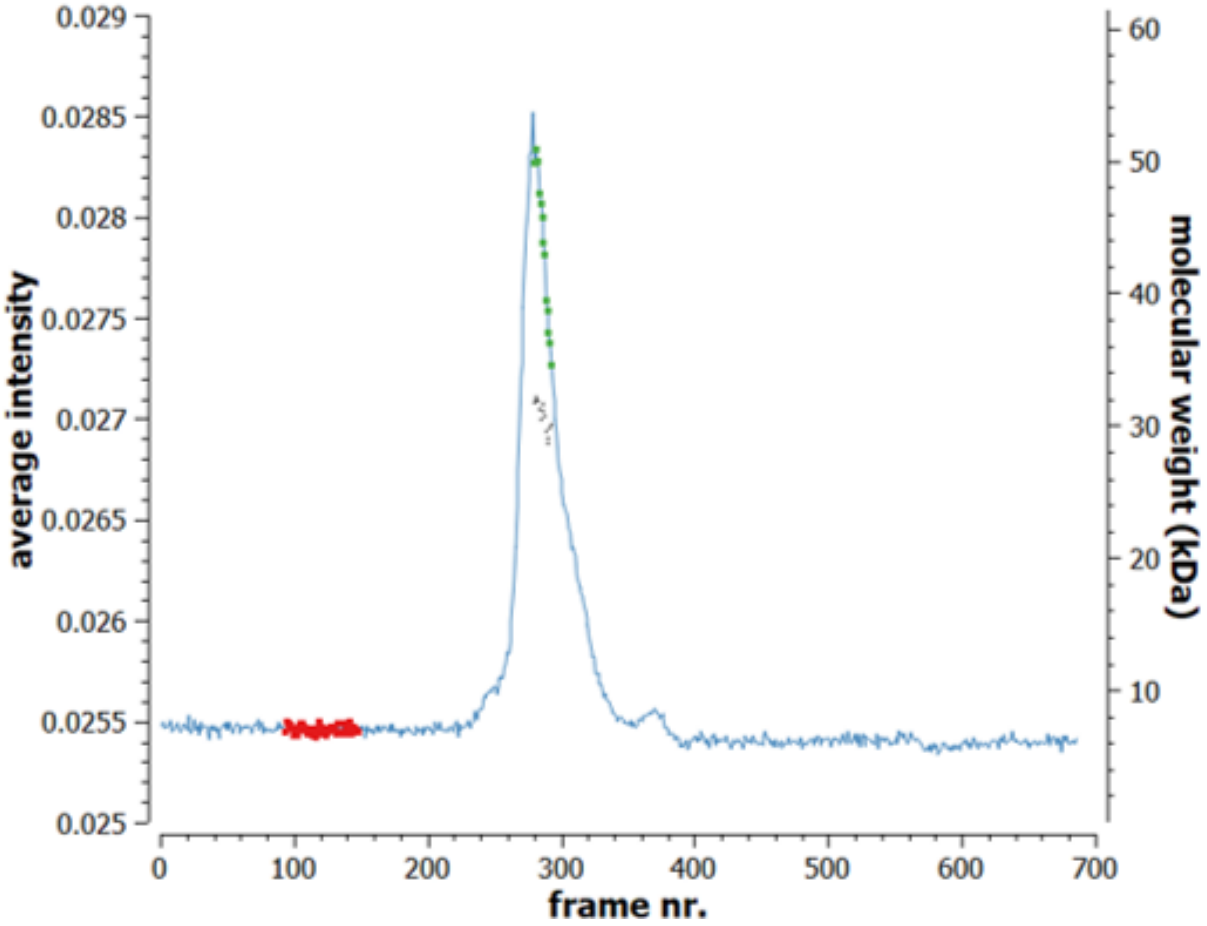

Figure A.46: Data selection of HP1α Y168E from SEC-SAXS. SAXS data intensity (blue) is plotted against frame number as HP1α Y168E sample is eluted from the SEC column. The selected frames from the peak of SAXS intensity are shown in green, with buffer frames for subtraction shown in red. Molecular weight is estimated (black dots) from these data points to show that the sample from which the data are collected is relatively homogeneous.

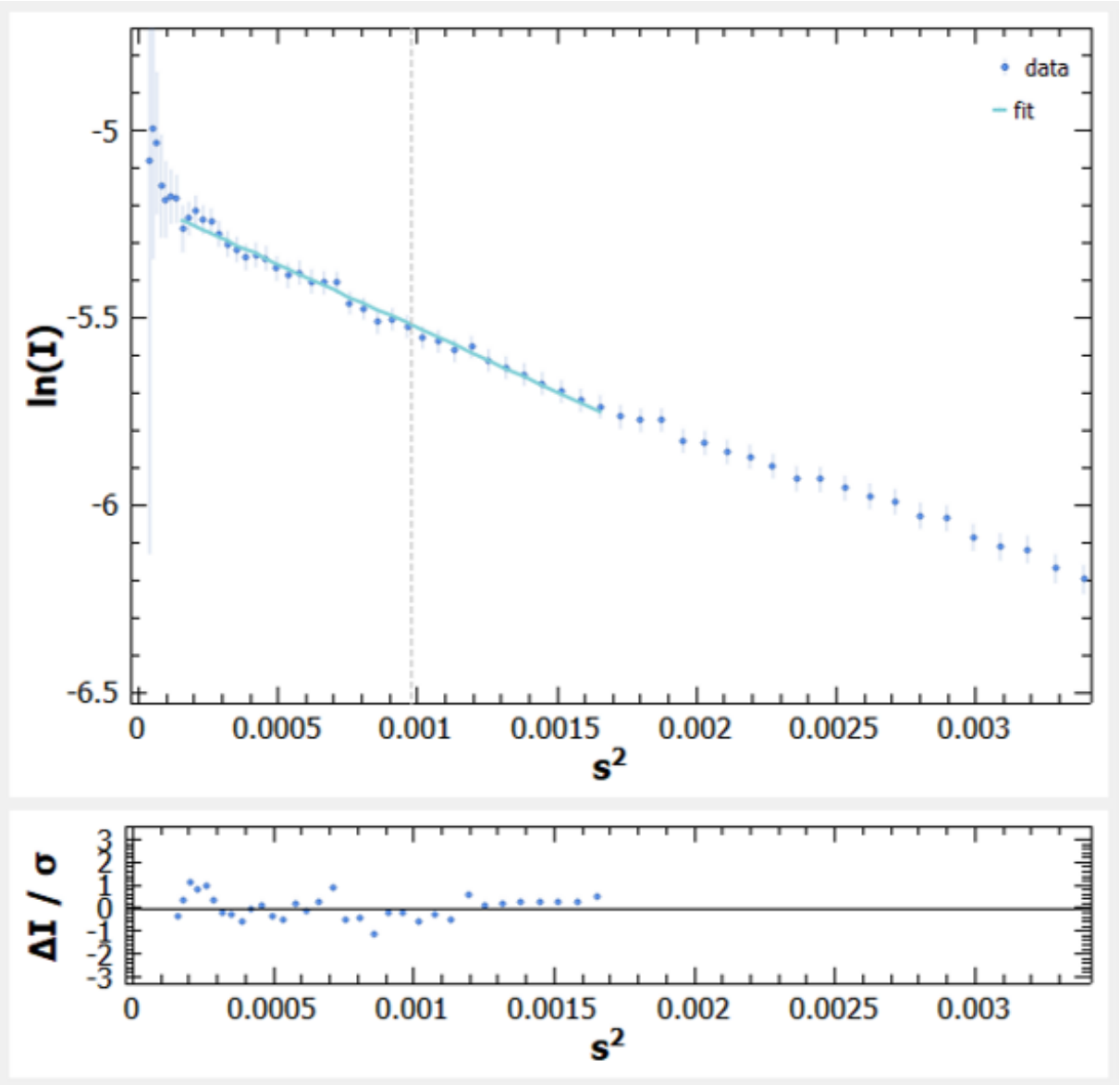

Figure A.47: Guinier fit of HP1α Y168E SEC-SAXS data. Linear Guinier fit to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

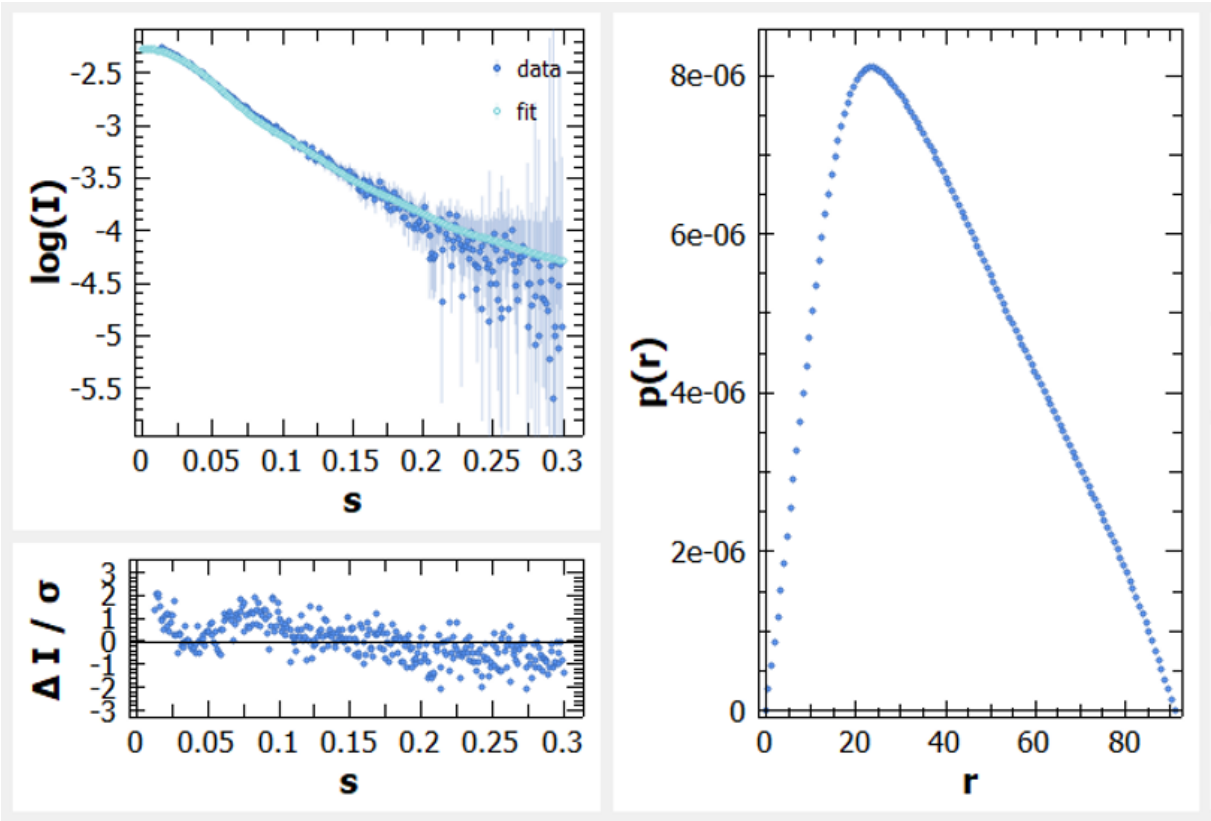

Figure A.48: *P(r)* plot of HP1α Y168E SEC-SAXS data. *P(r)* versus r profile from HP1α Y168E SEC-SAXS data. A fit of the back-transformed *P(r)* curve to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

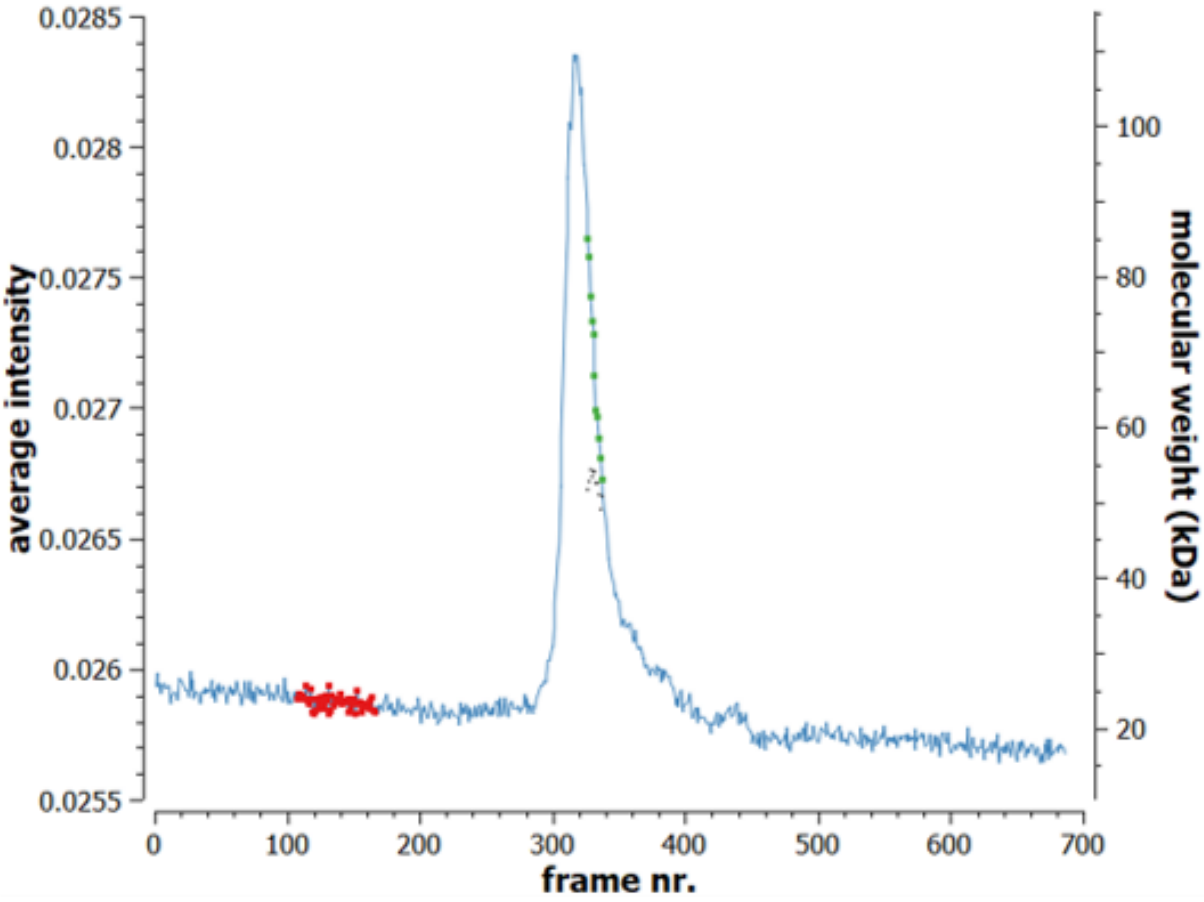

Figure A.49: Data selection of WT HP1α from SEC-SAXS. SAXS data intensity (blue) is plotted against frame number as HP1α WT sample is eluted from the SEC column. The selected frames from the peak of SAXS intensity are shown in green, with buffer frames for subtraction shown in red. Molecular weight is estimated (black dots) from these data points to show that the sample from which the data are collected is relatively homogeneous.

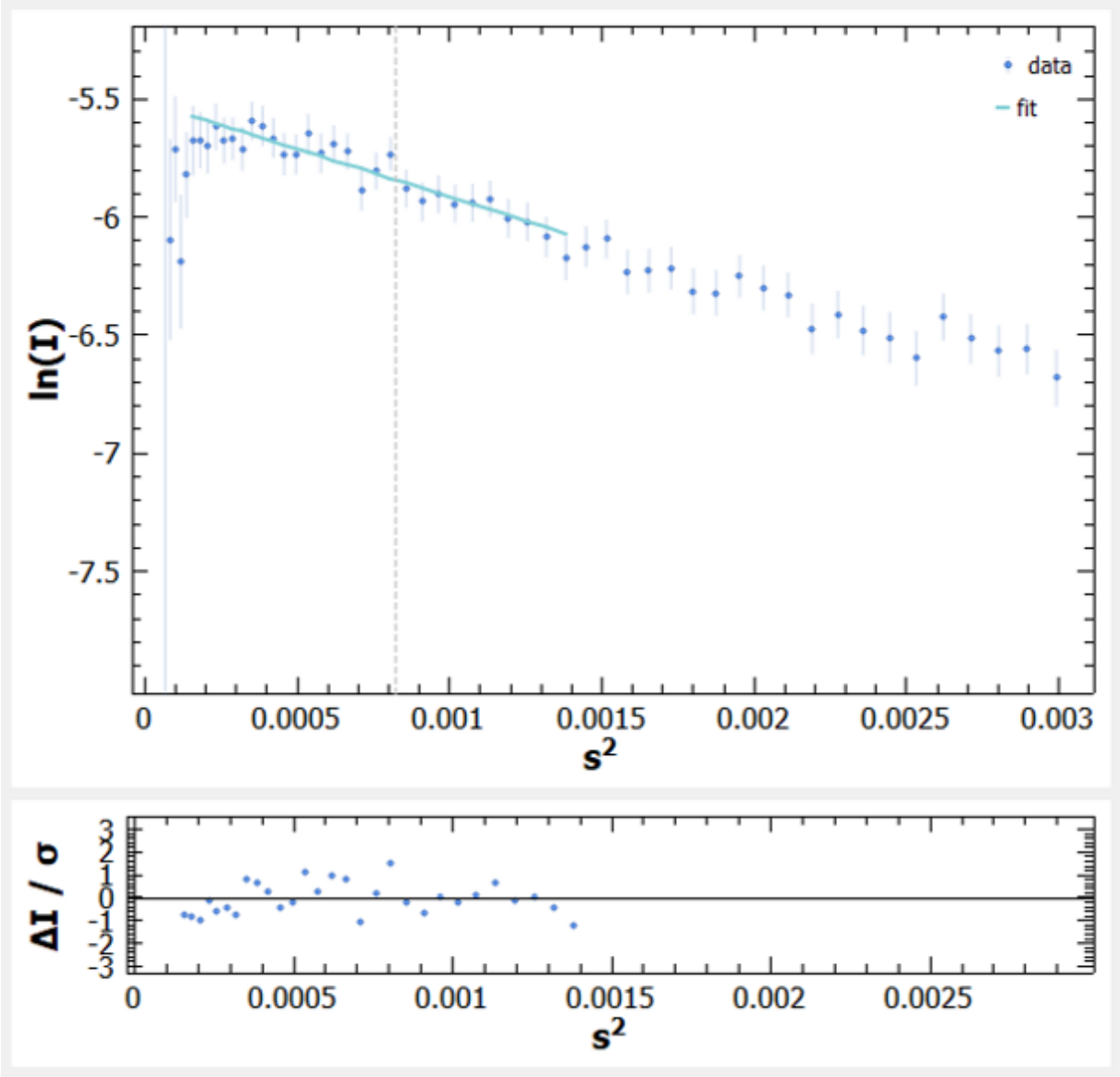

Figure A.50: Guinier fit of HP1α WT SEC-SAXS data. Linear Guinier fit to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

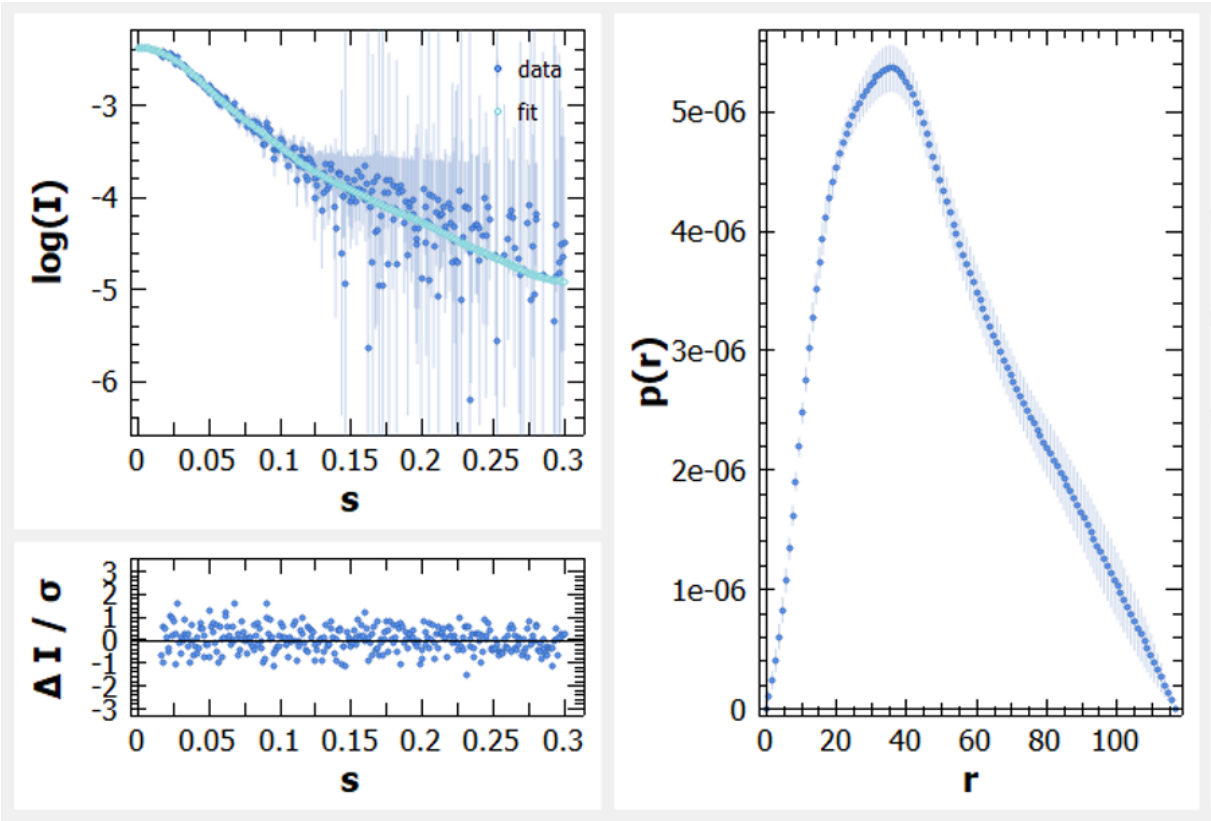

Figure A.51: *P(r)* plot of HP1α WT SEC-SAXS data. *P(r)* versus r profile from HP1α WT SEC-SAXS data. A fit of the back-transformed *P(r)* curve to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

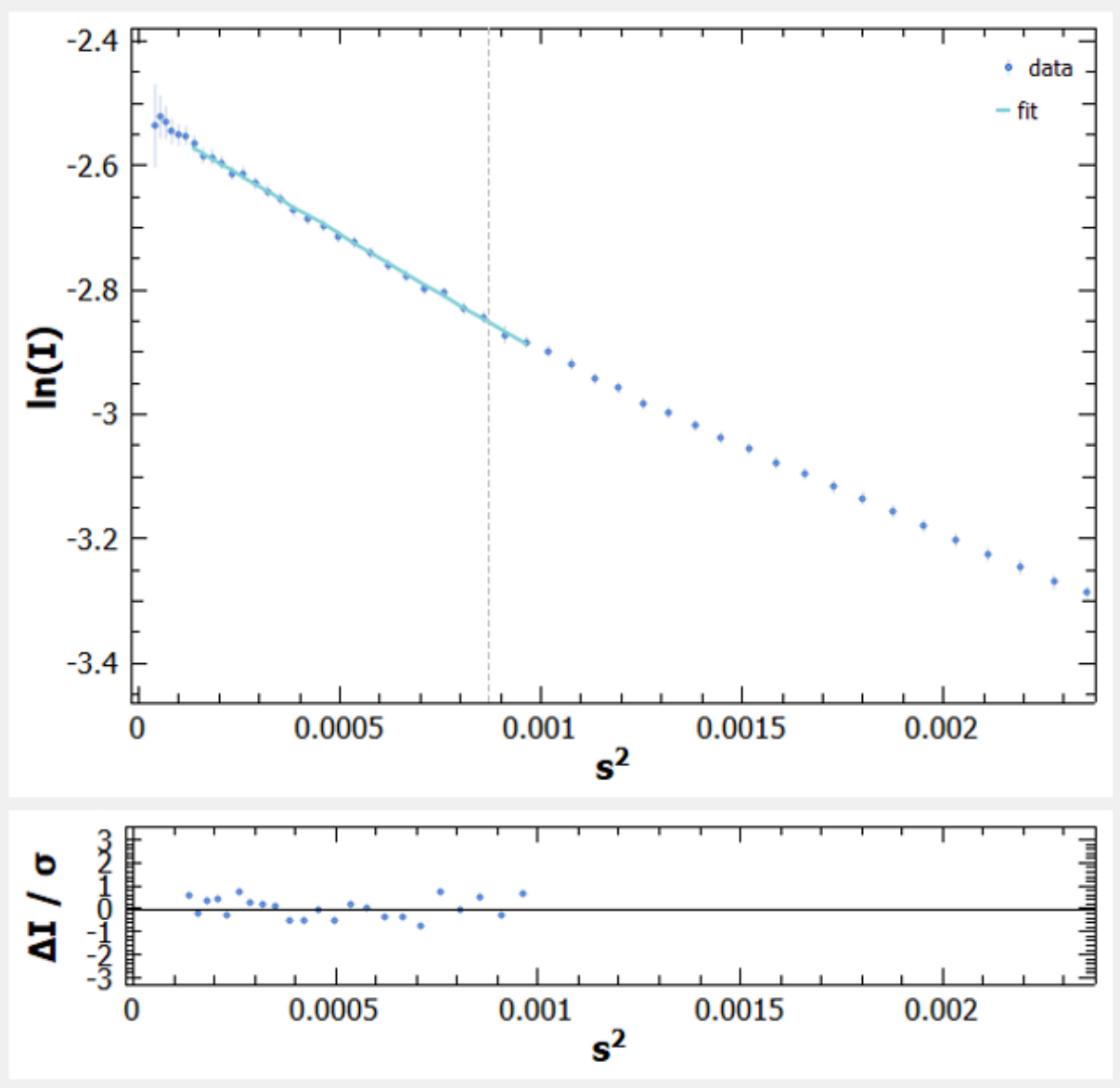

Figure A.52: Guinier fit of HP1α Y168E static mode SAXS data. Linear Guinier fit to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

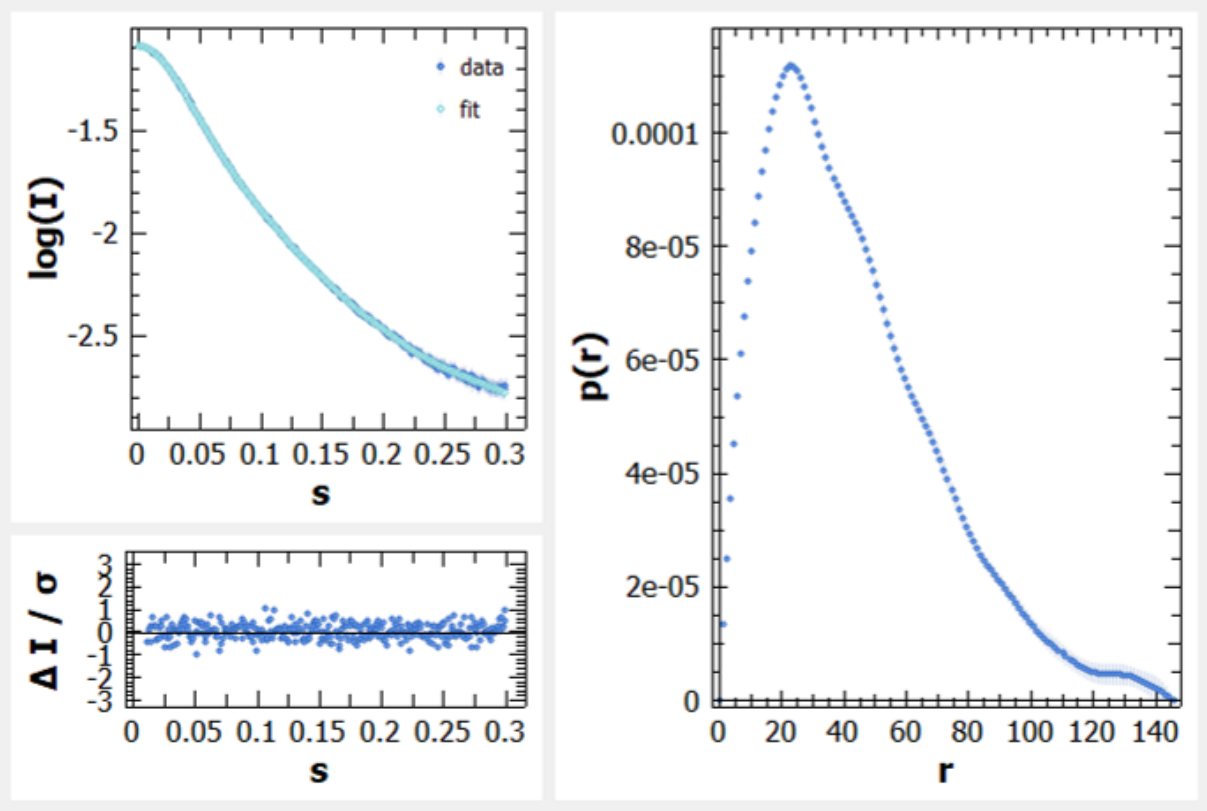

Figure A.53: *P(r)* plot of HP1α Y168E static mode SAXS data. *P(r)* versus r profile from HP1α Y168E SAXS data. A fit of the back-transformed *P(r)* curve to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

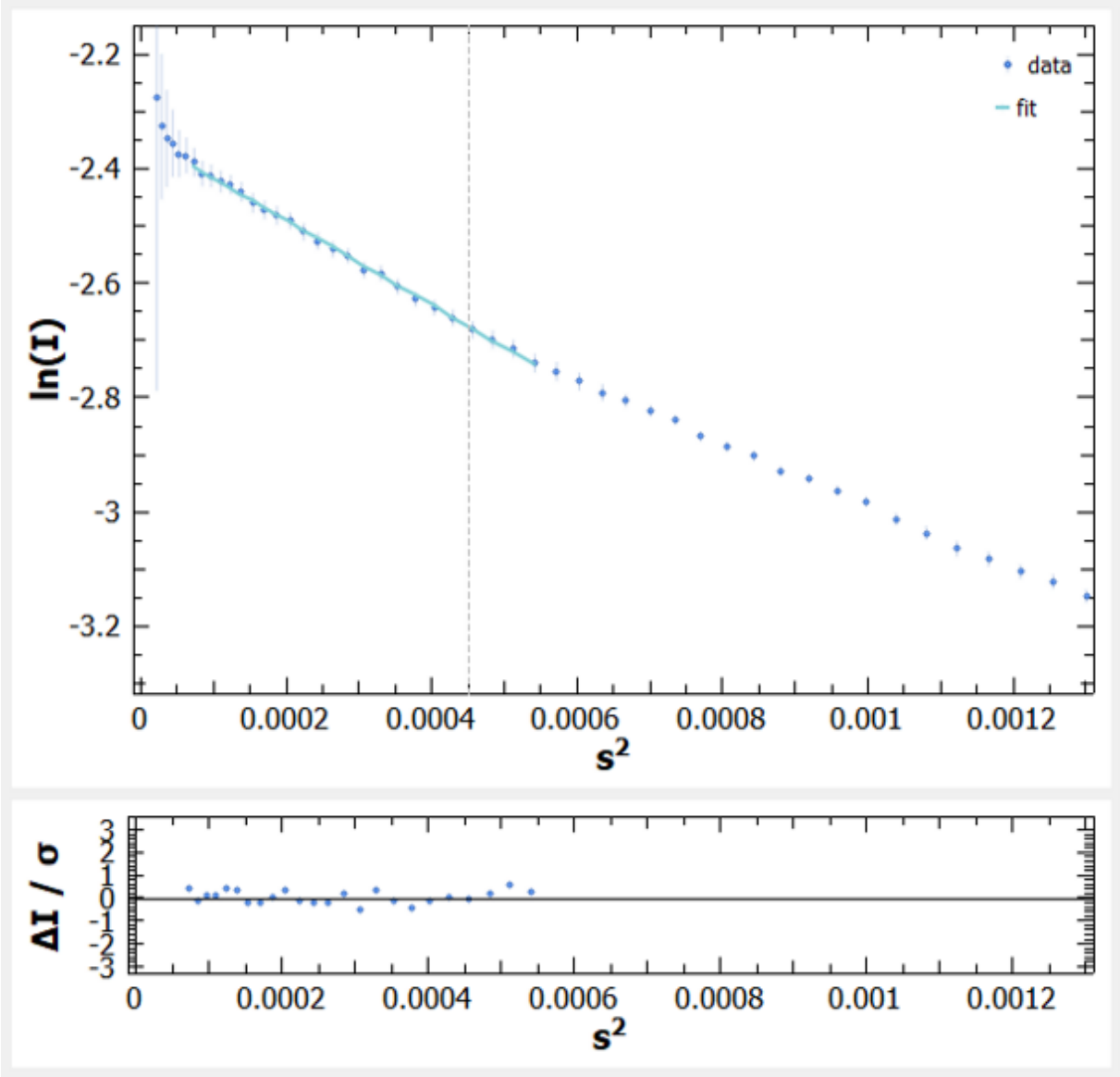

Figure A.54: Guinier fit of HP1α WT static mode SAXS data. Linear Guinier fit to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

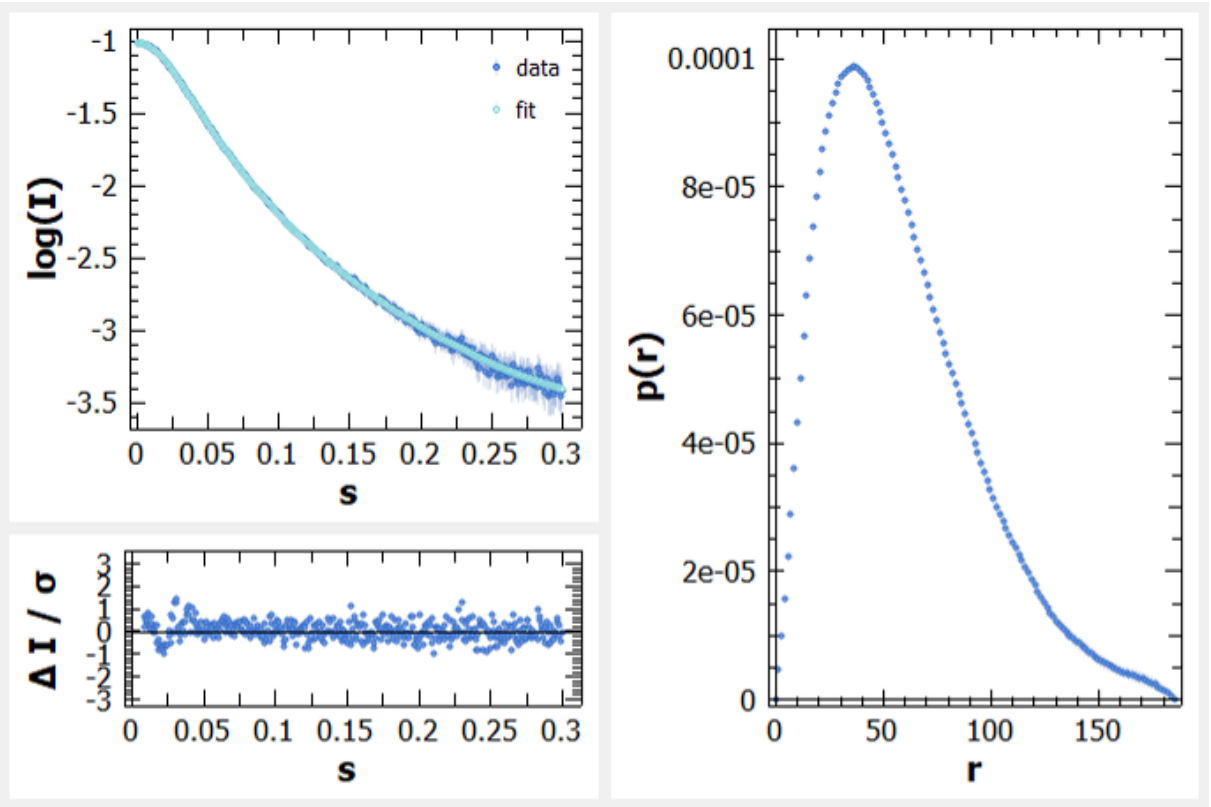

Figure A.55: *P(r)* plot of HP1α WT static mode SAXS data. *P(r)* versus r profile from HP1α WT SAXS data. A fit of the back-transformed *P(r)* curve to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

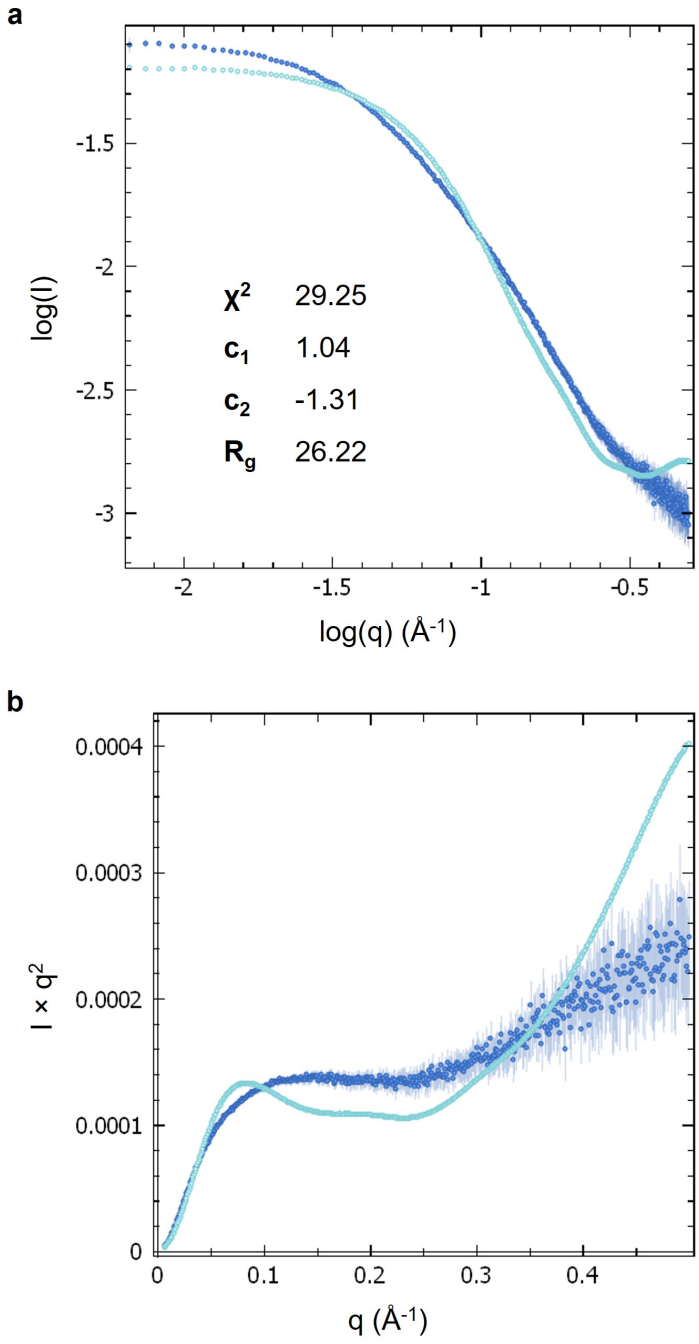

Figure A.56: Fit of FoXS-generated SAXS profile of AF-Q61686-F1 to HP1α Y168E static SAXS data. FoXS-generated SAXS curve of AF-Q61686-F1 shown in light blue, and static SAXS data of HP1α Y168E shown in dark blue. The data are presented as a) log of I versus log of q, and b) as a Kratky plot.

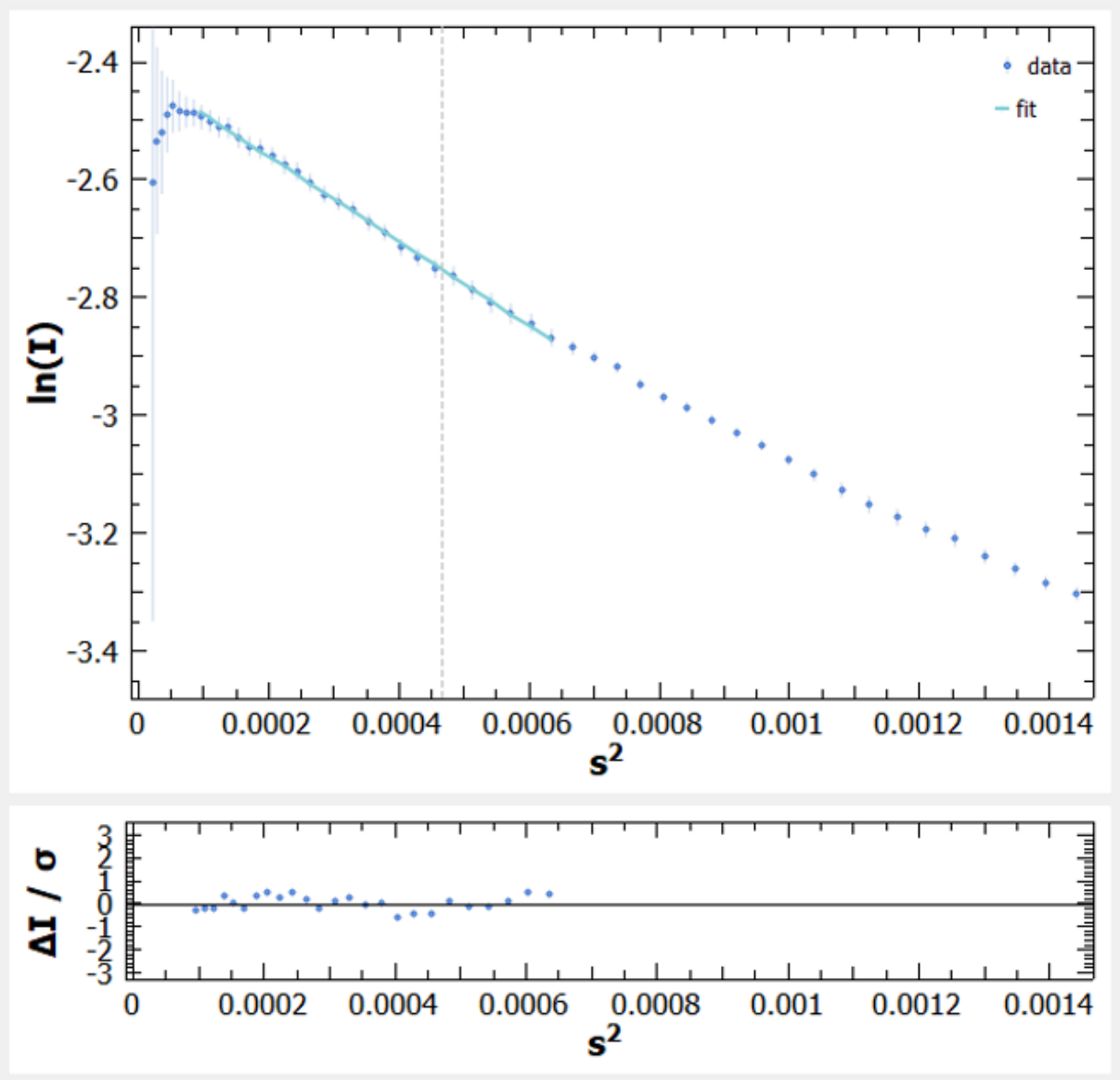

Figure A.57: Guinier fit of HP1α S11-14E SAXS data. Linear Guinier fit to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

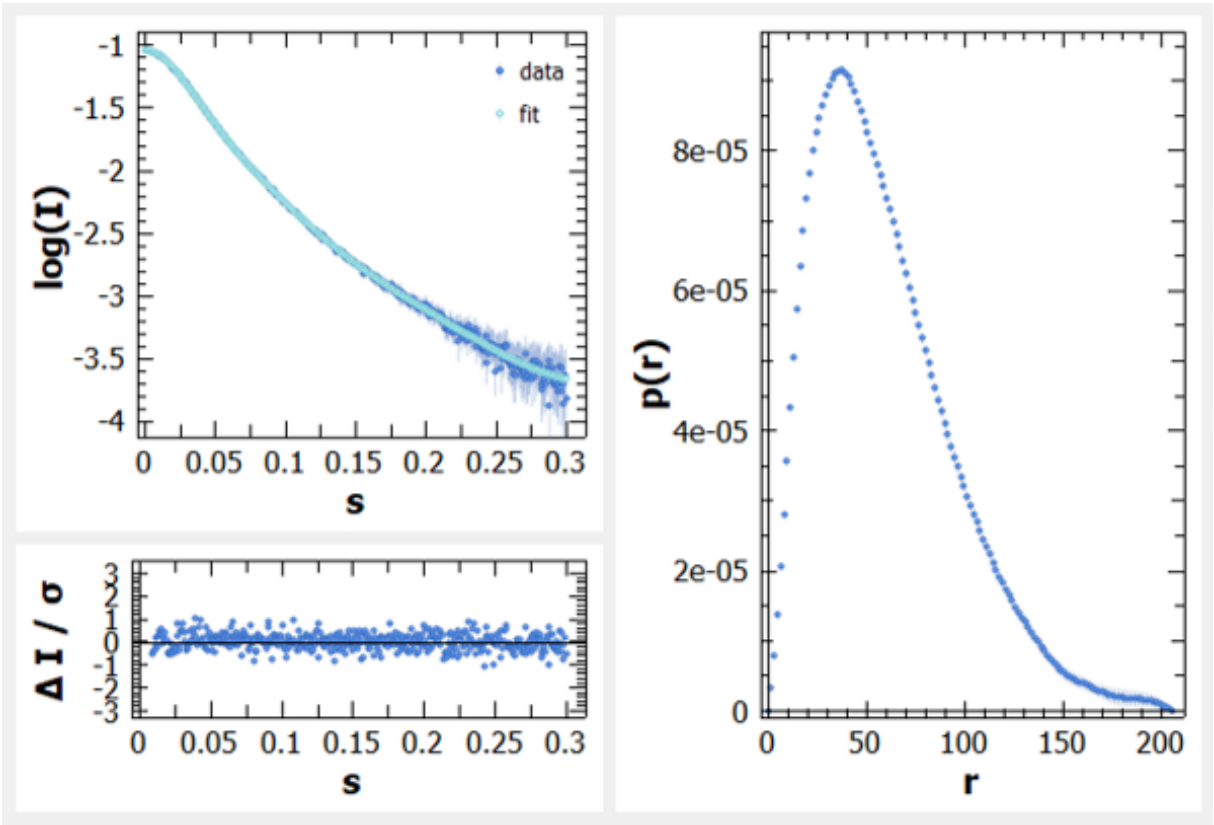

Figure A.58: *P(r)* plot of HP1α S11-14E SAXS data. *P(r)* versus r profile from HP1α S11-14E SAXS data. A fit of the back-transformed *P(r)* curve to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

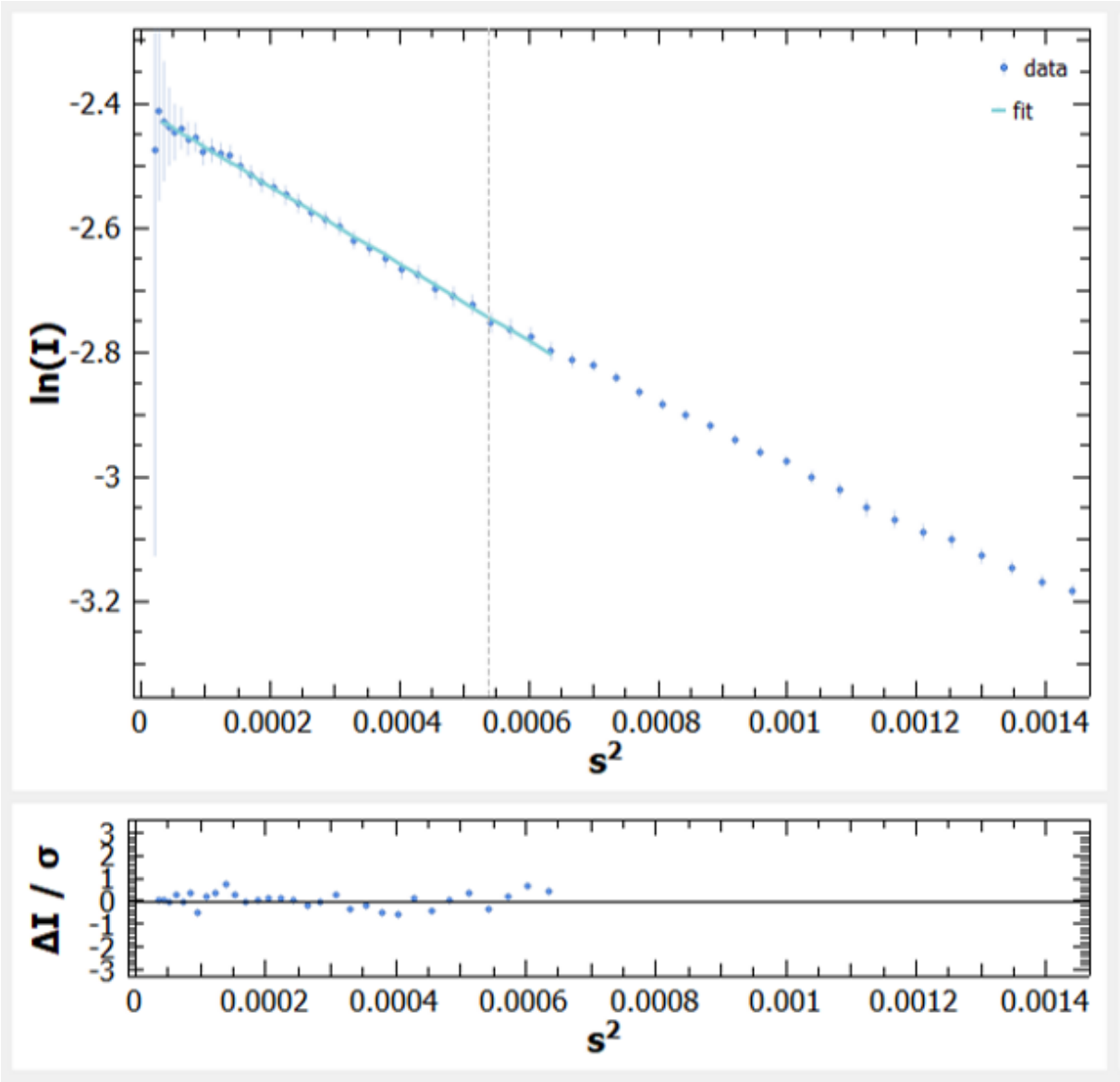

Figure A.59: Guinier fit of HP1α ΔNTE SAXS data. Linear Guinier fit to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

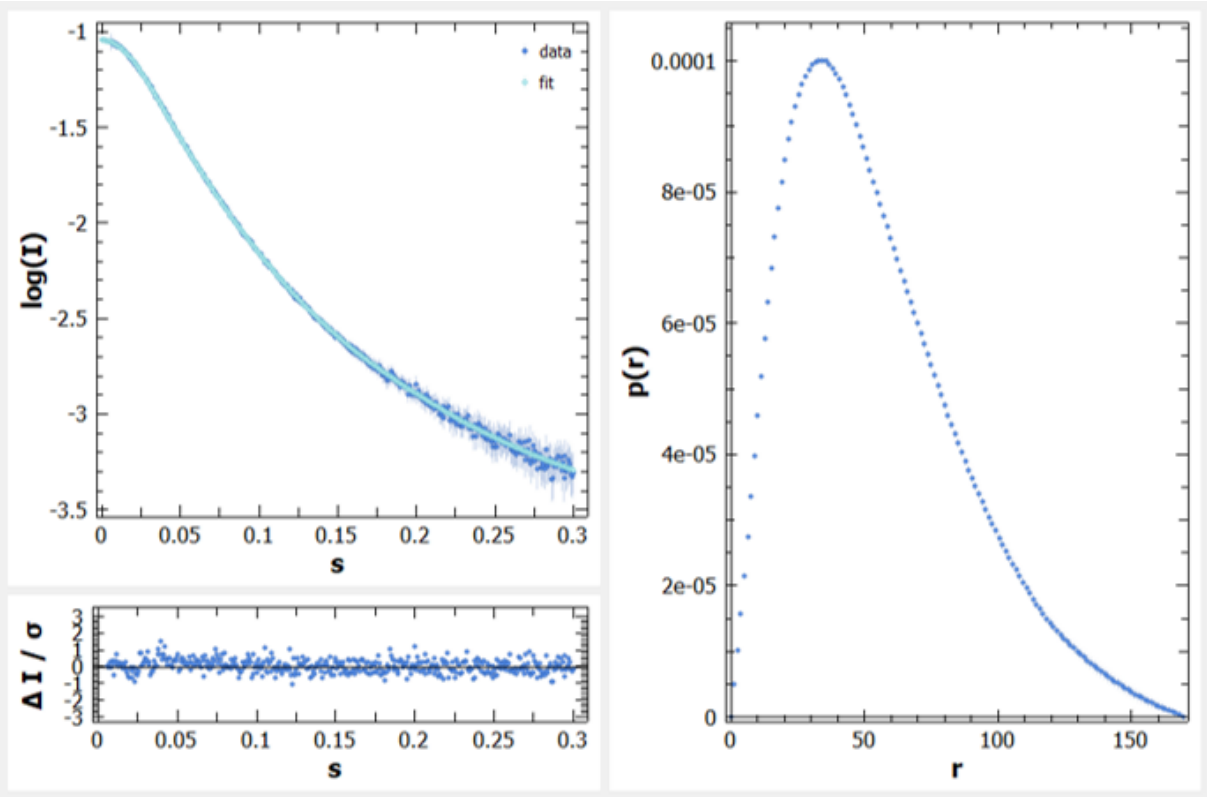

Figure A.60: *P(r)* plot of HP1α ΔNTE SAXS data. *P(r)* versus r profile from HP1α ΔNTE SAXS data. A fit of the back-transformed *P(r)* curve to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

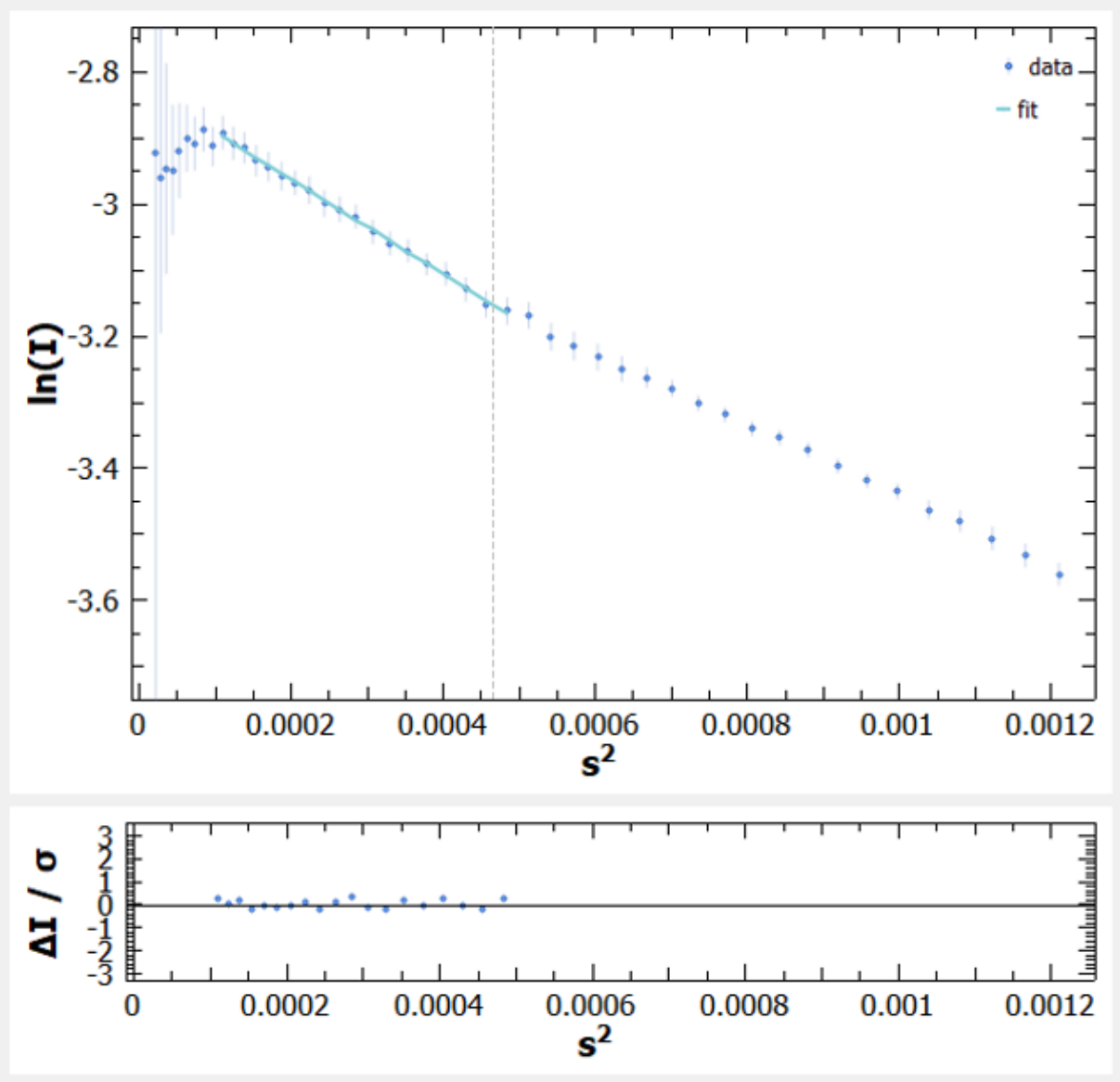

Figure A.61: Guinier fit of HP1α ΔCTE SAXS data. Linear Guinier fit to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

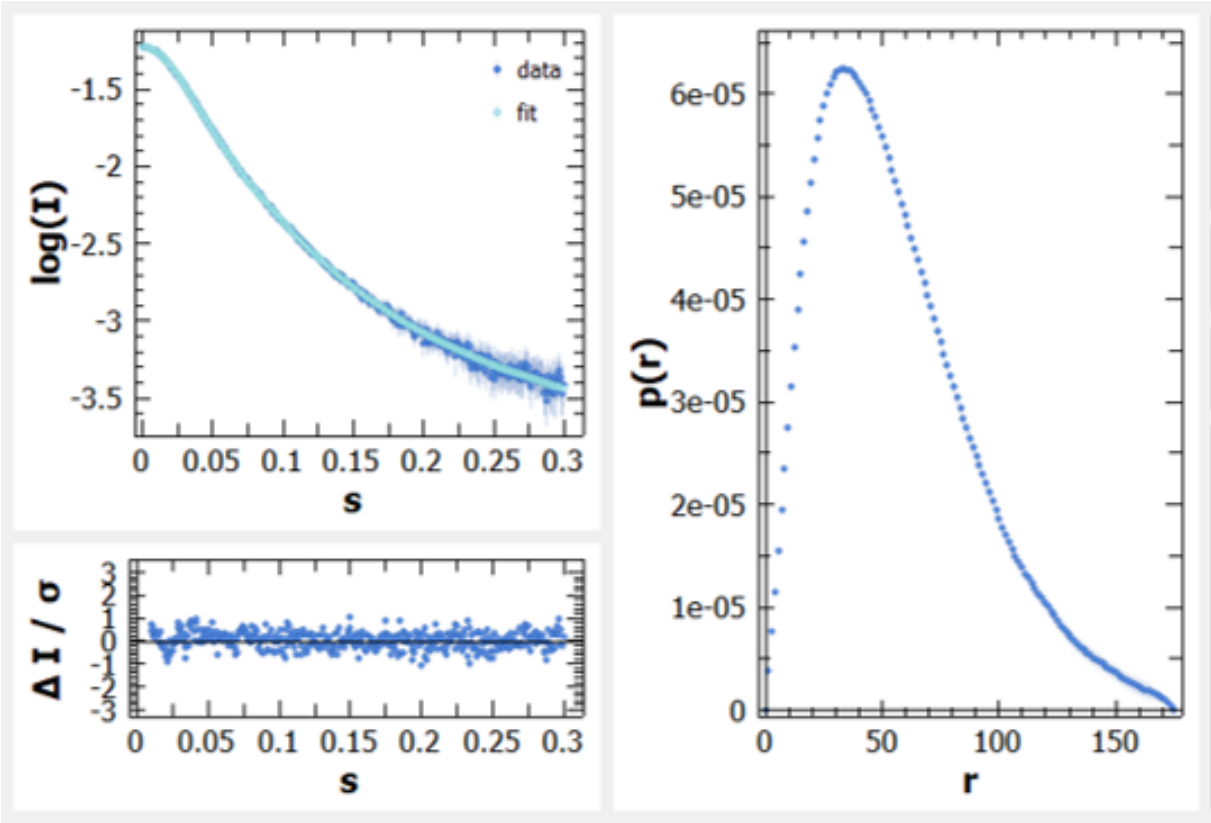

Figure A.62: *P(r)* plot of HP1α ΔCTE SAXS data. *P(r)* versus r profile from HP1α ΔCTE SAXS data. A fit of the back-transformed *P(r)* curve to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

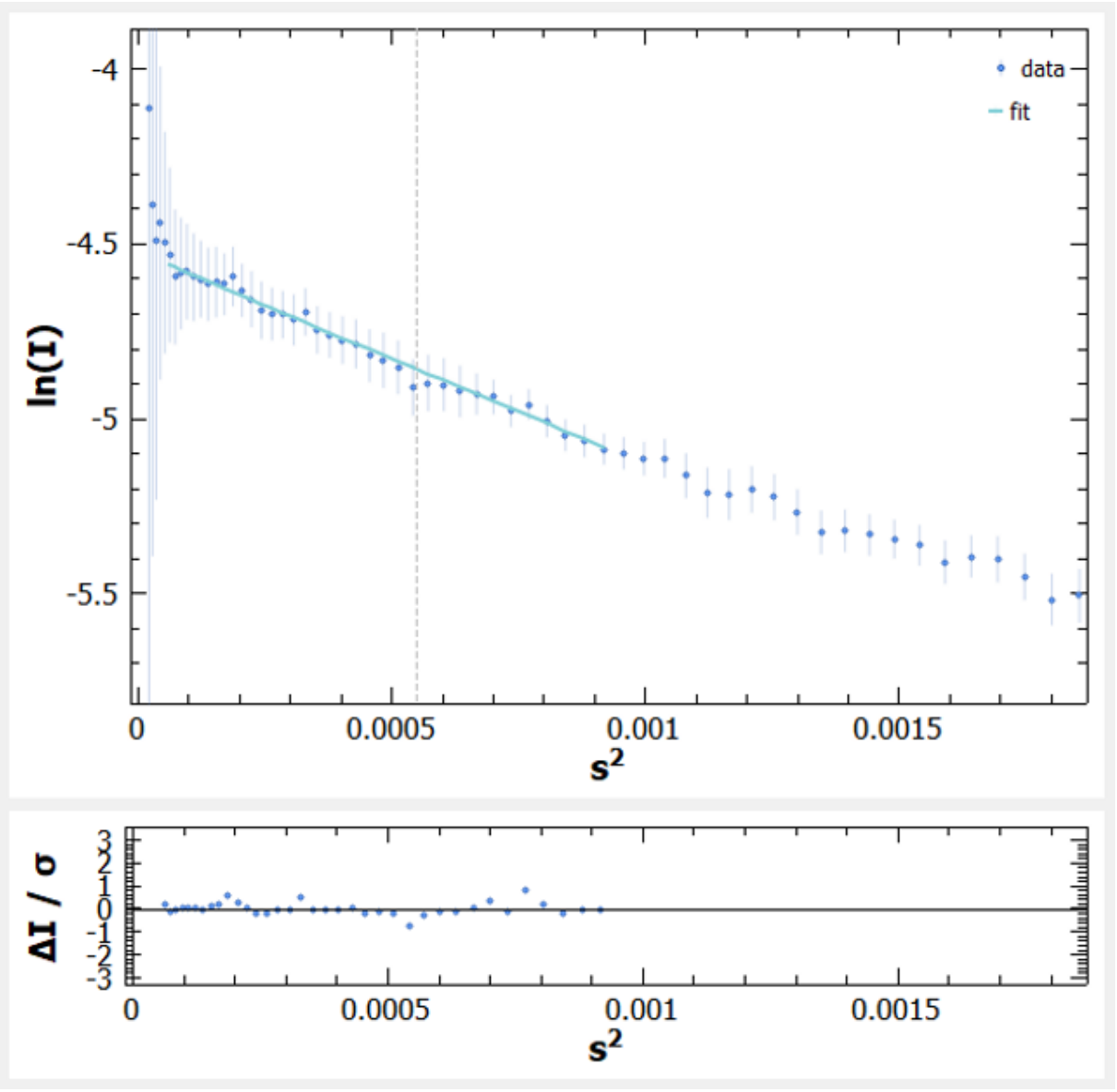

Figure A.63: Guinier fit of HP1α A2;3 SAXS data. Linear Guinier fit to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

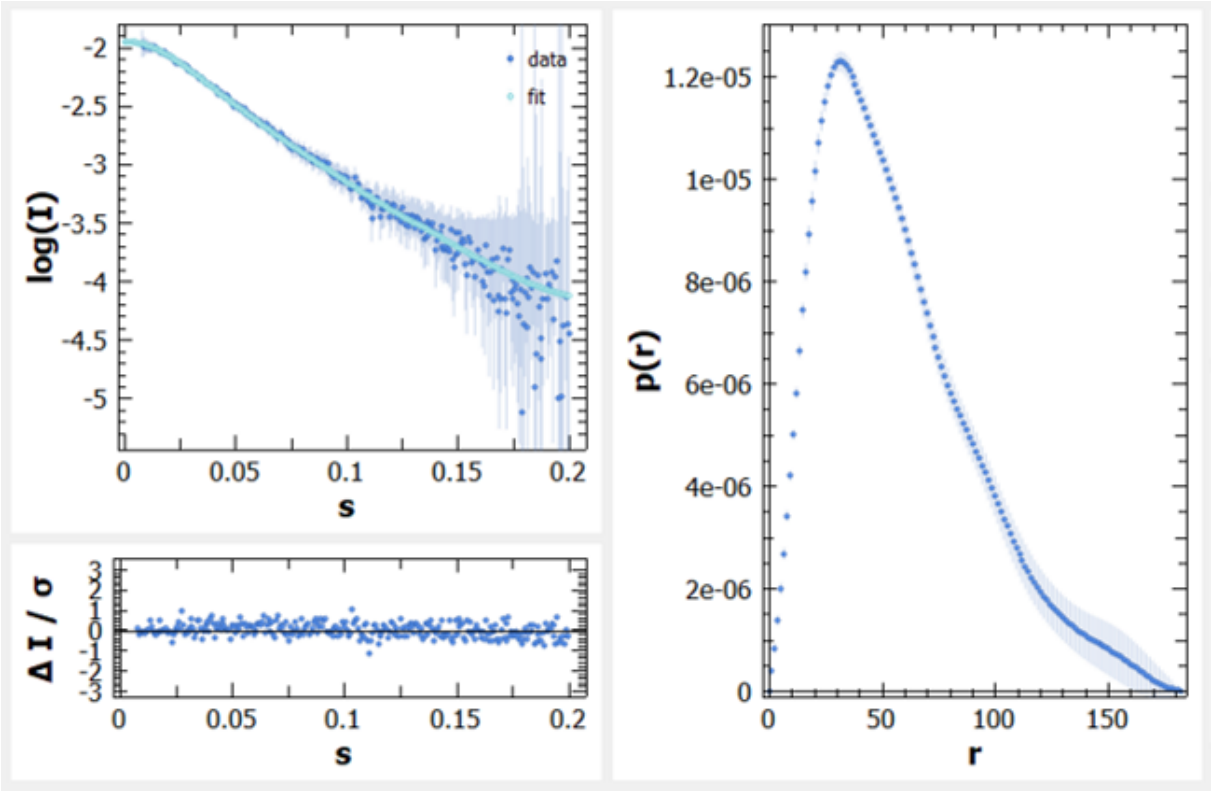

Figure A.64: *P(r)* plot of HP1α A2;3 SAXS data. *P(r)* versus r profile from HP1α A2;3 SAXS data. A fit of the back-transformed *P(r)* curve to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

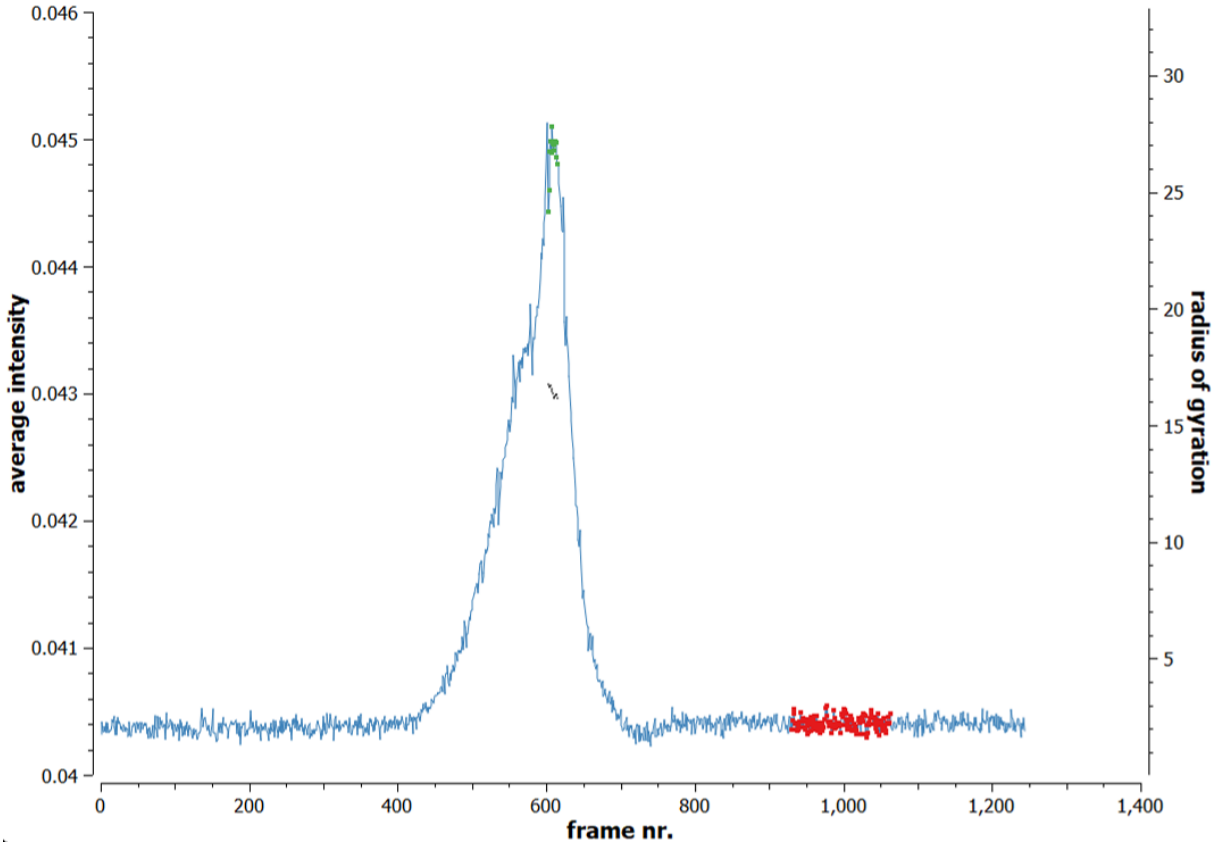

Figure A.65: Data selection of TERRA45 from SEC-SAXS. SAXS data intensity (blue) is plotted against frame number as TERRA45 sample is eluted from the SEC column. The selected frames from the peak of SAXS intensity are shown in green, with buffer frames for subtraction shown in red. R_g is estimated (black dots) from these data points to show that the samples from which the data are collected is relatively homogeneous.

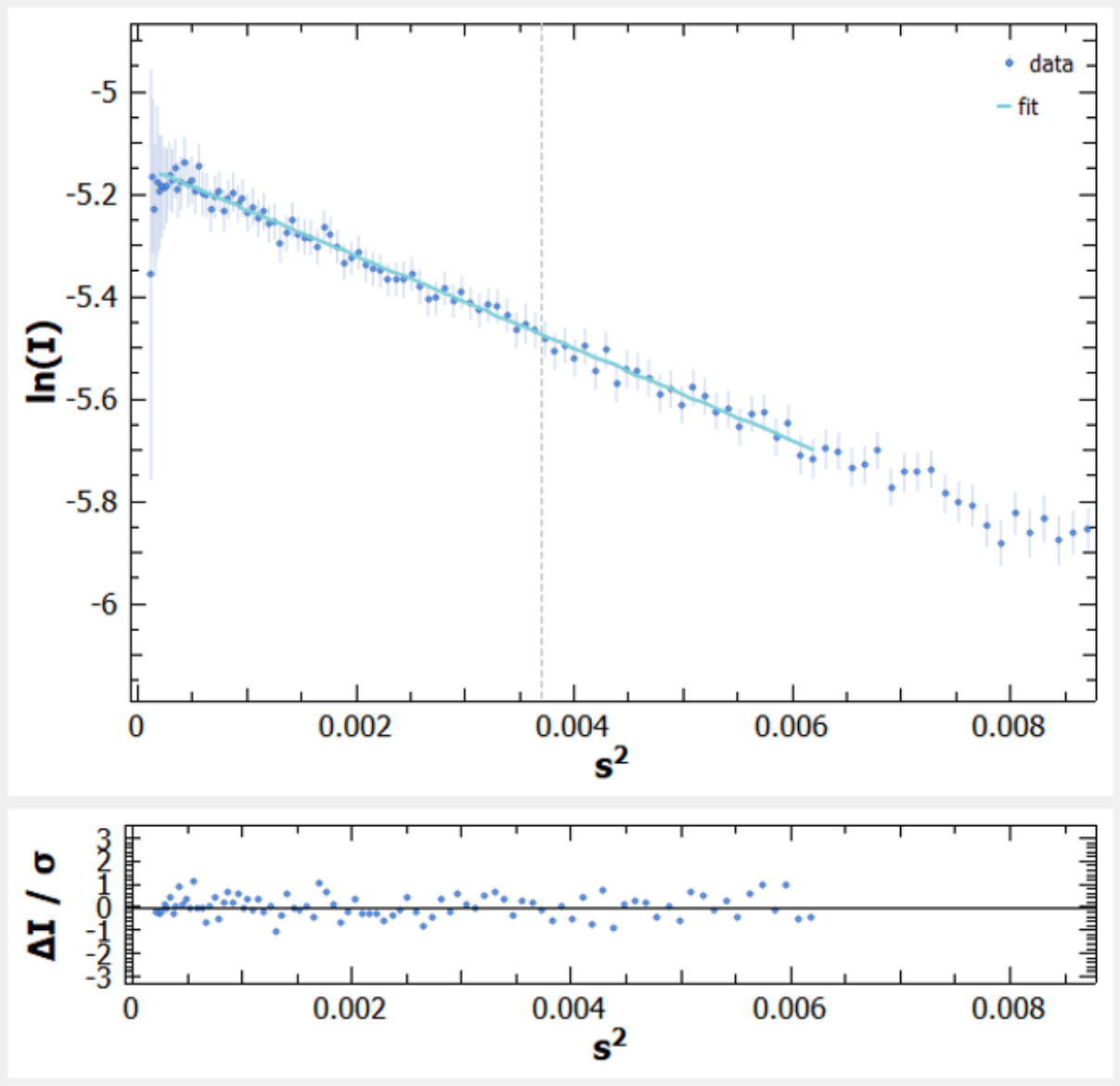

Figure A.66: Guinier fit of TERRA45 SEC-SAXS data. Linear Guinier fit to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

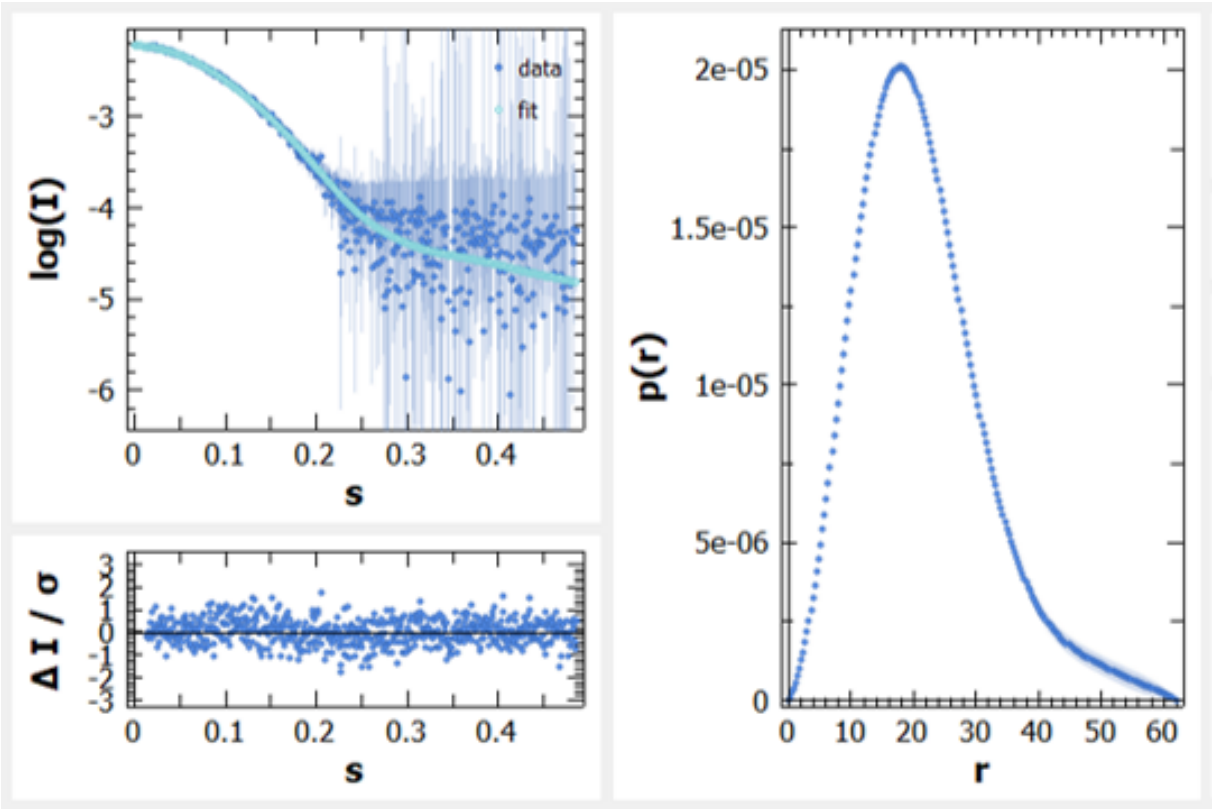

Figure A.67: *P(r)* plot of TERRA45 SEC-SAXS data. *P(r)* versus r profile from TERRA45 SEC-SAXS data. A fit of the back-transformed *P(r)* curve to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

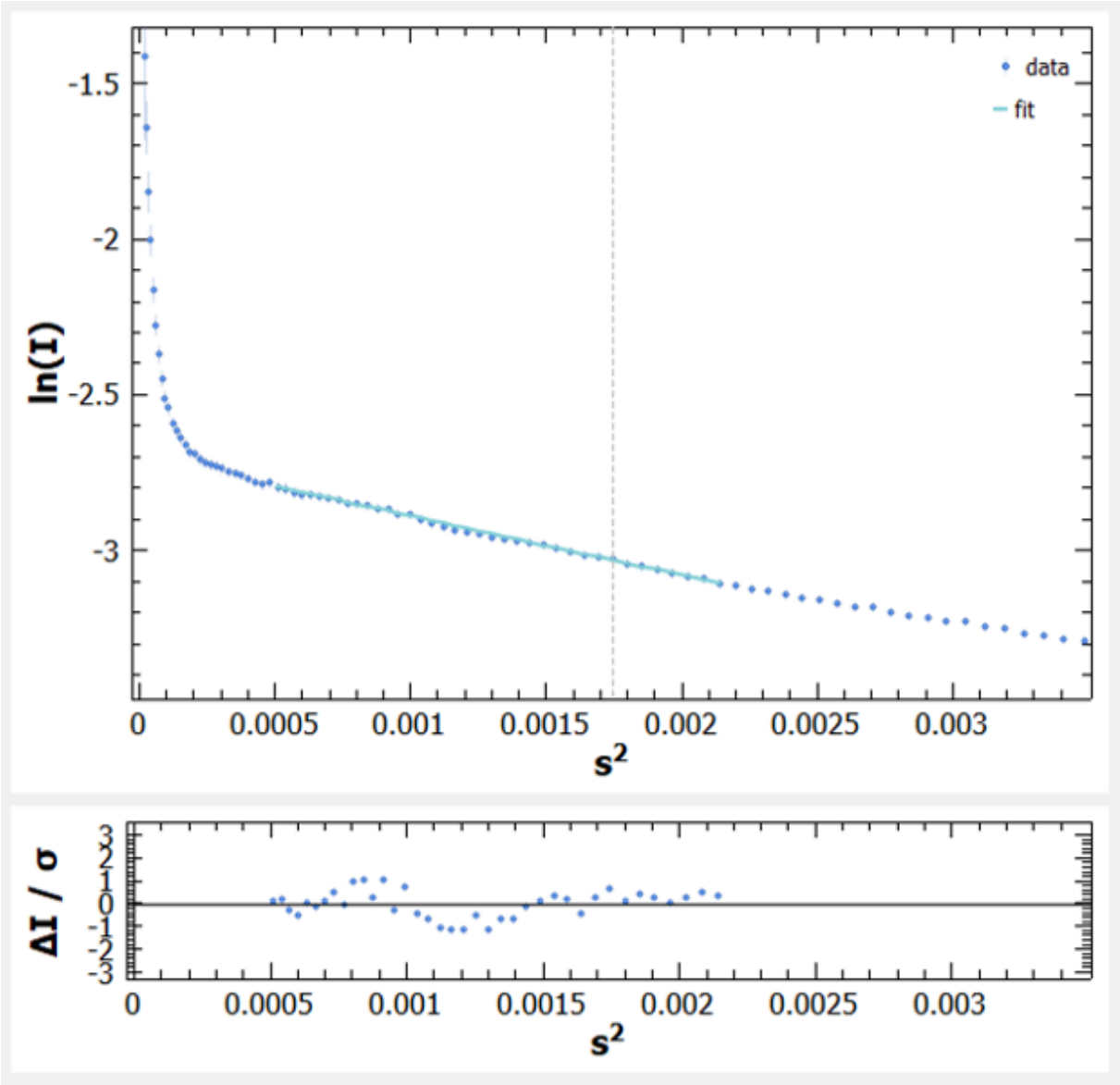

Figure A.68: Guinier fit of TERRA45 static mode SAXS data. Linear Guinier fit to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

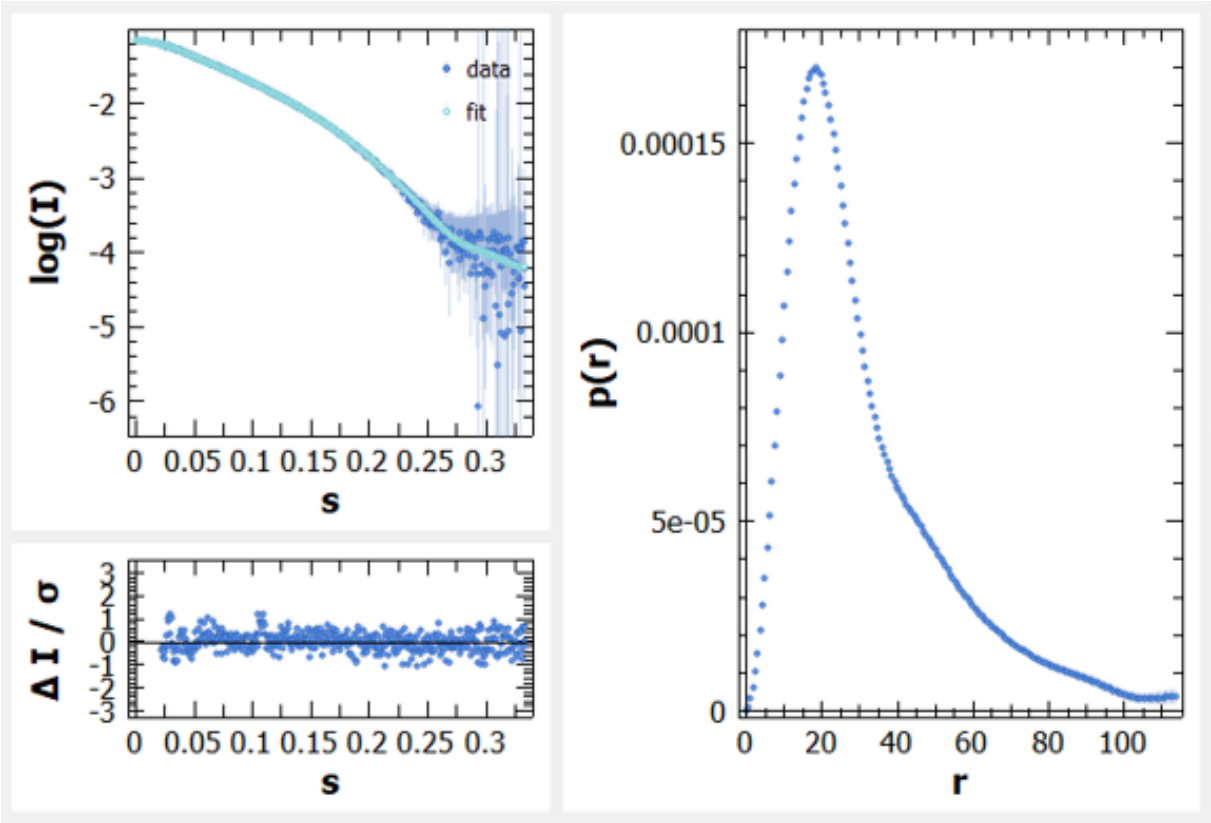

Figure A.69: *P(r)* plot of TERRA45 static mode SAXS data. *P(r)* versus r profile from TERRA45 SAXS data. A fit of the back-transformed *P(r)* curve to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

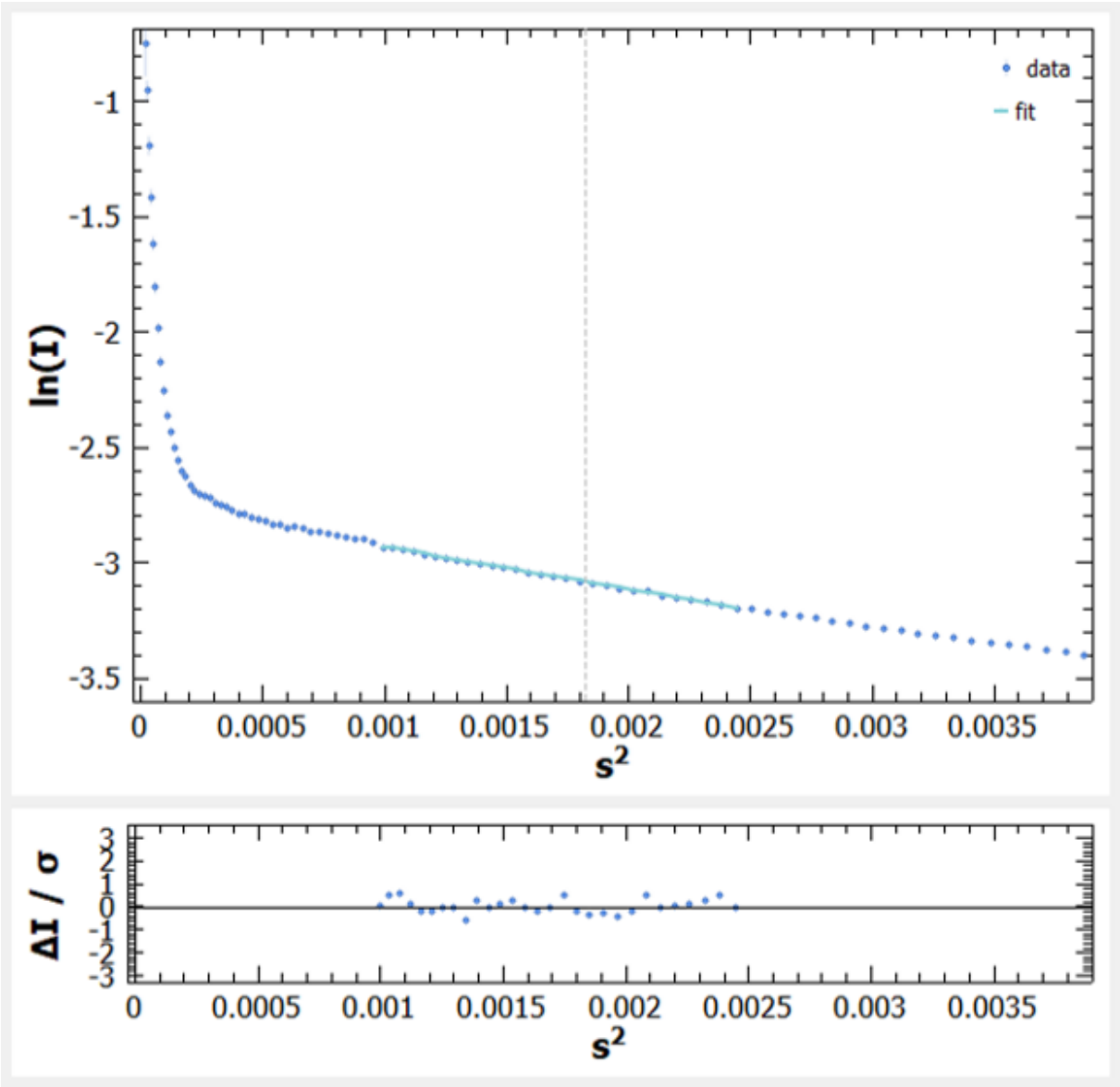

Figure A.70: Guinier fit of mutTERRA45 static mode SAXS data. Linear Guinier fit to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

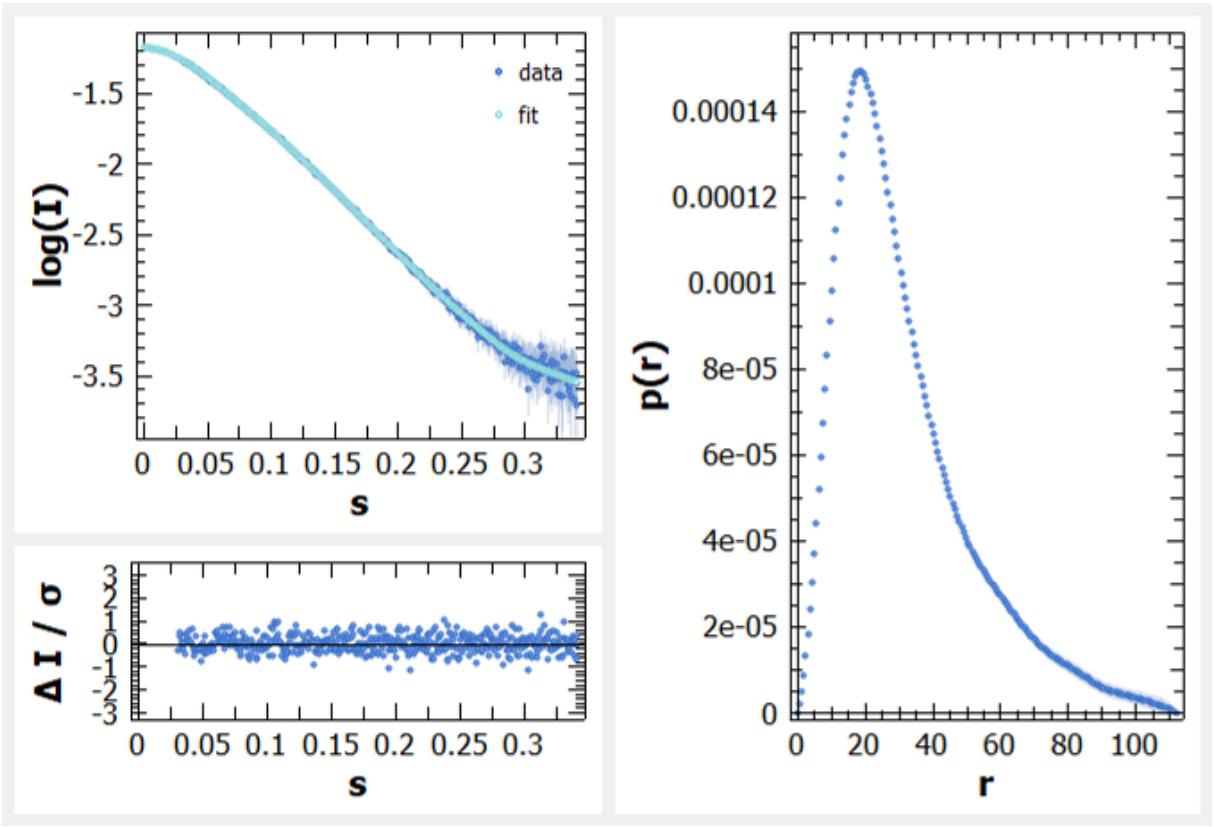

Figure A.71: *P(r)* plot of mutTERRA45 static mode SAXS data. *P(r)* versus r profile from mutTERRA45 SAXS data. A fit of the back-transformed *P(r)* curve to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

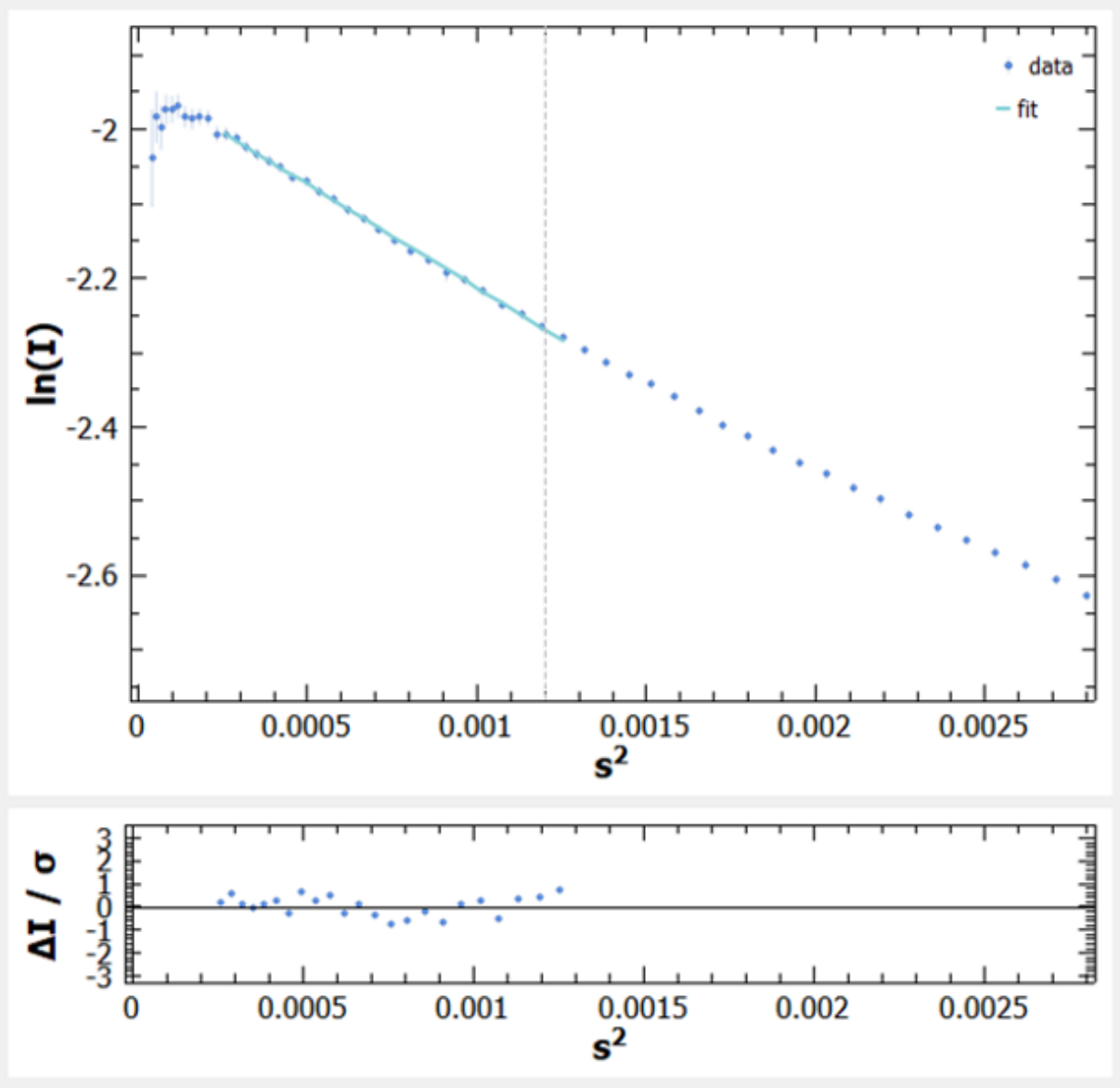

Figure A.72: Guinier fit of HP1α Hinge-only plus TERRA45 SAXS data. Linear Guinier fit to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

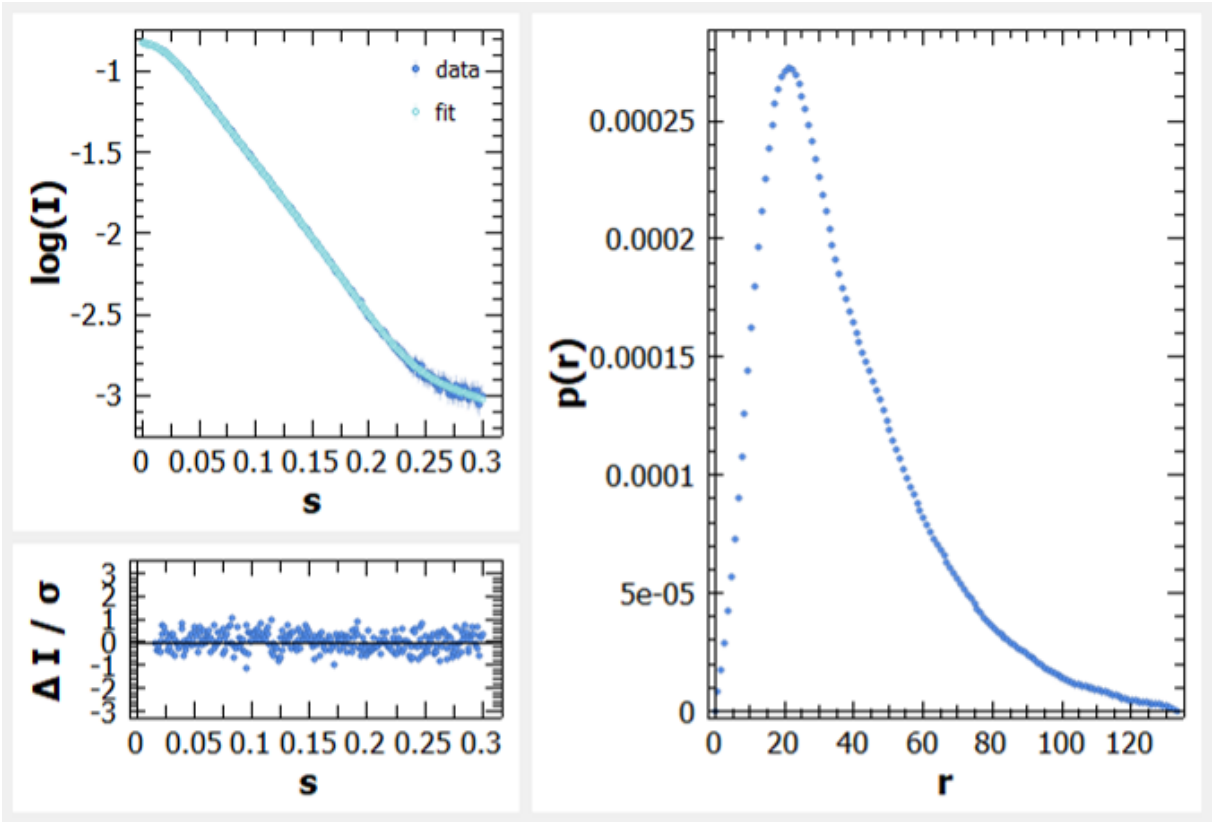

Figure A.73: *P(r)* plot of HP1α Hinge-only plus TERRA45 SAXS data. *P(r)* versus r profile from HP1α Hinge-only plus TERRA45 SAXS data. A fit of the back-transformed *P(r)* curve to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

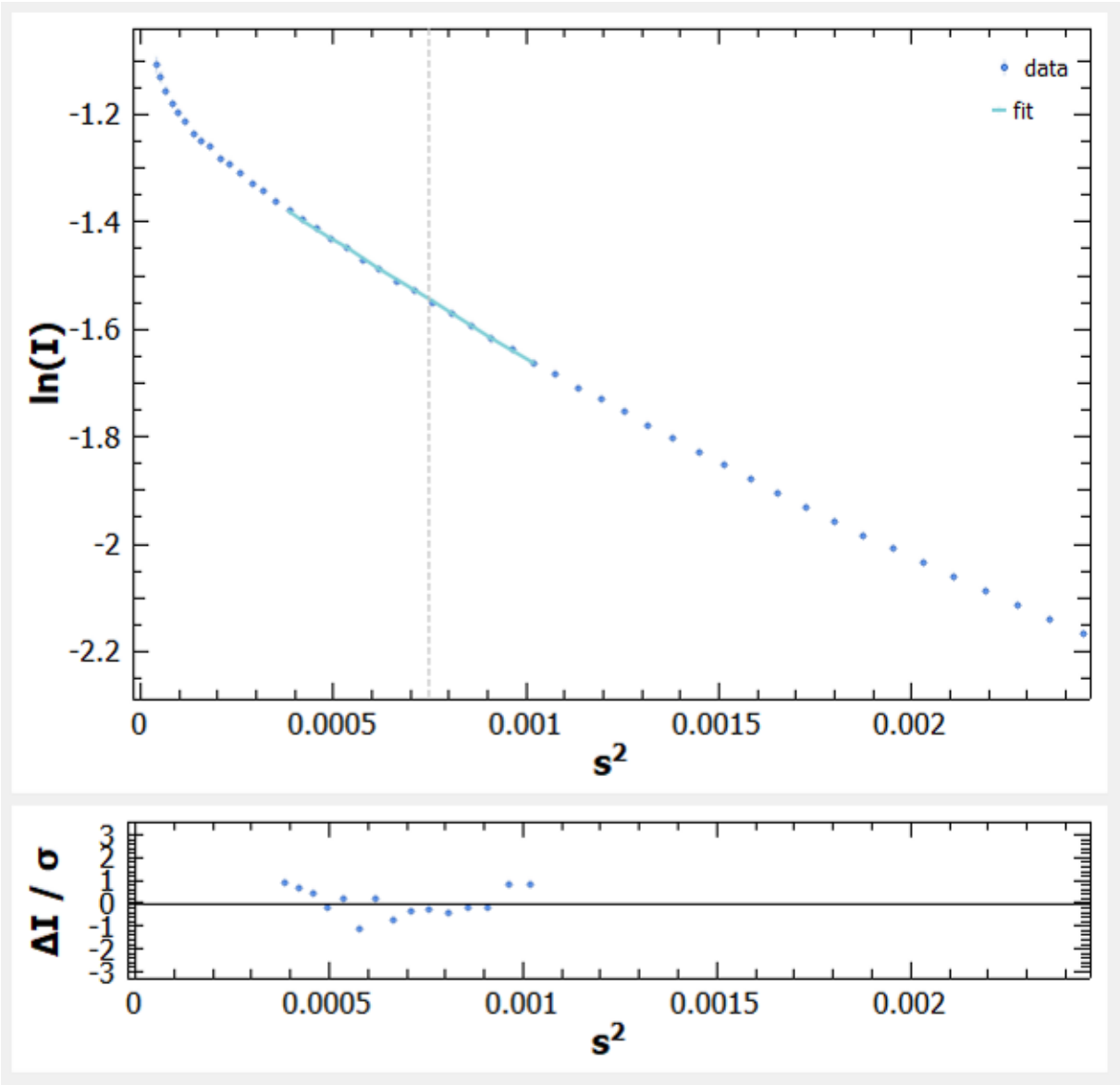

Figure A.74: Guinier fit of HP1α Y168E plus TERRA45 SAXS data. Linear Guinier fit to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

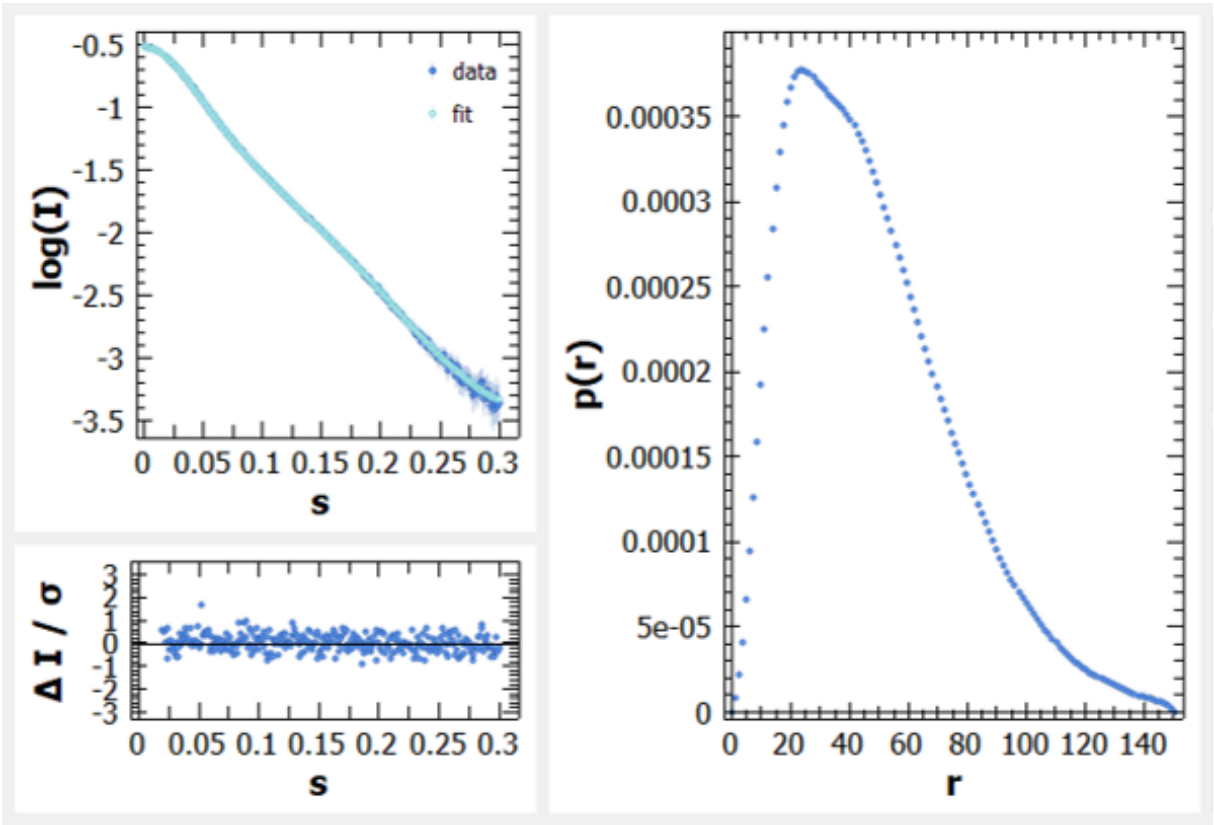

Figure A.75: *P(r)* plot of HP1α Y168E plus TERRA45 SAXS data. *P(r)* versus r profile from HP1α Y168E plus TERRA45 SAXS data. A fit of the back-transformed *P(r)* curve to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

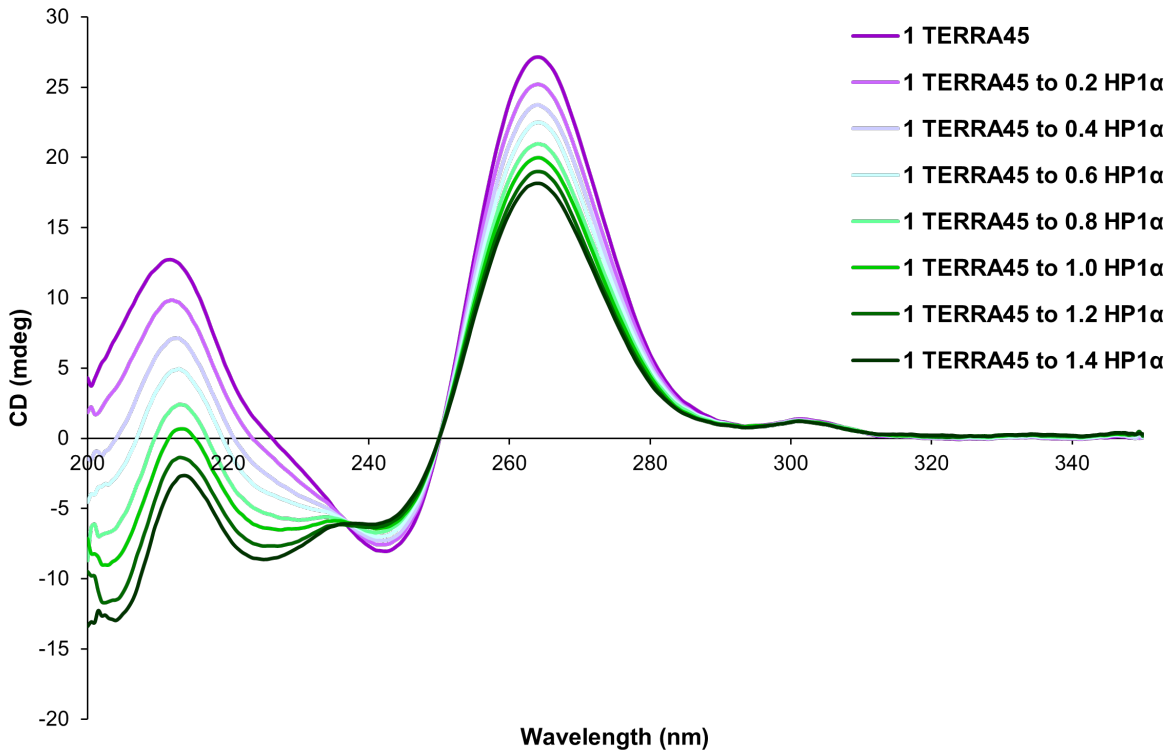

Figure A.76: HP1α binding does not alter topology of parallel G4 TERRA45. Circular dichroism spectra of TERRA45 with 0.2 mol (to 1 mol TERRA45) increment additions of HP1α. The decrease in circular dichroism signal at 265 nm with increasing concentration of HP1α is due to dilution of the sample.

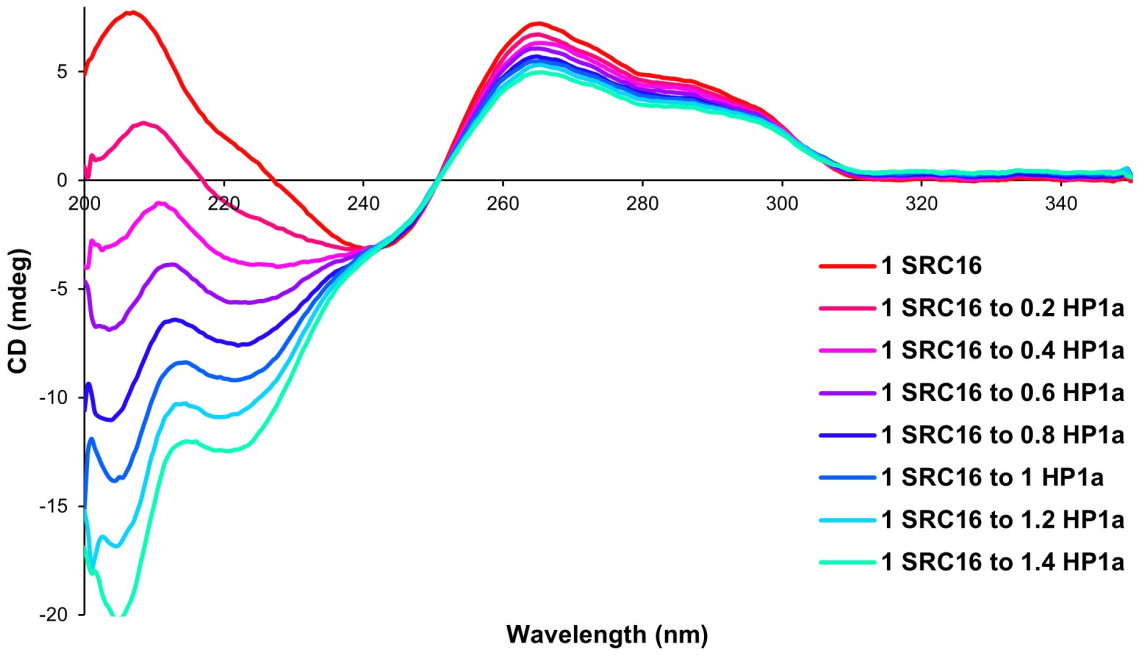

Figure A.77: HP1α binding does not alter topology of anti-parallel G4 SRC16. Circular dichroism spectra of SCR16 with 0.2 mol (to 1 mol SRC16) increment additions of HP1α. The decrease in circular dichroism signal at 265 nm with increasing concentration of HP1α is due to dilution of the sample.

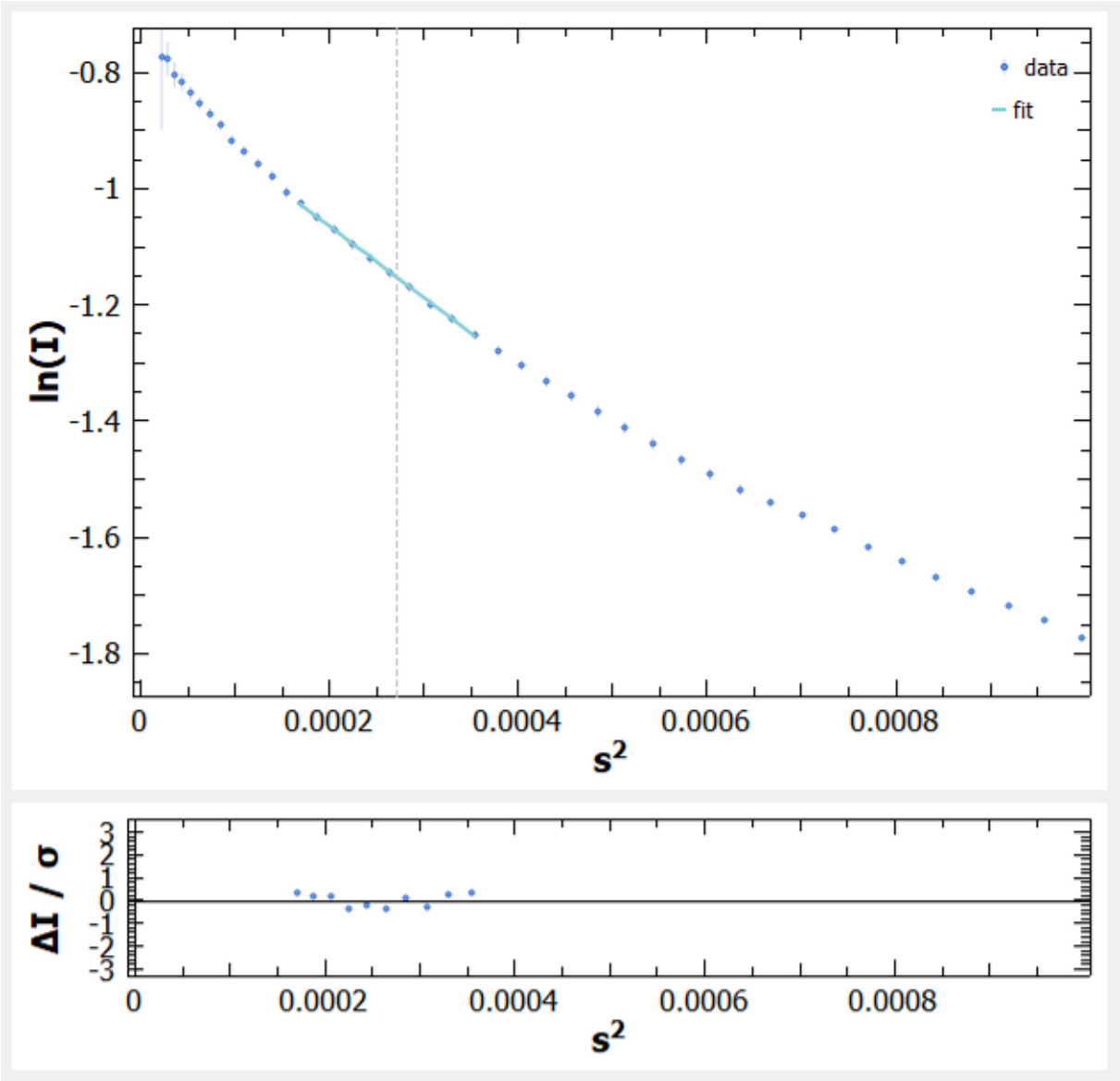

Figure A.78: Guinier fit of HP1α WT plus TERRA45 SAXS data. Linear Guinier fit to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

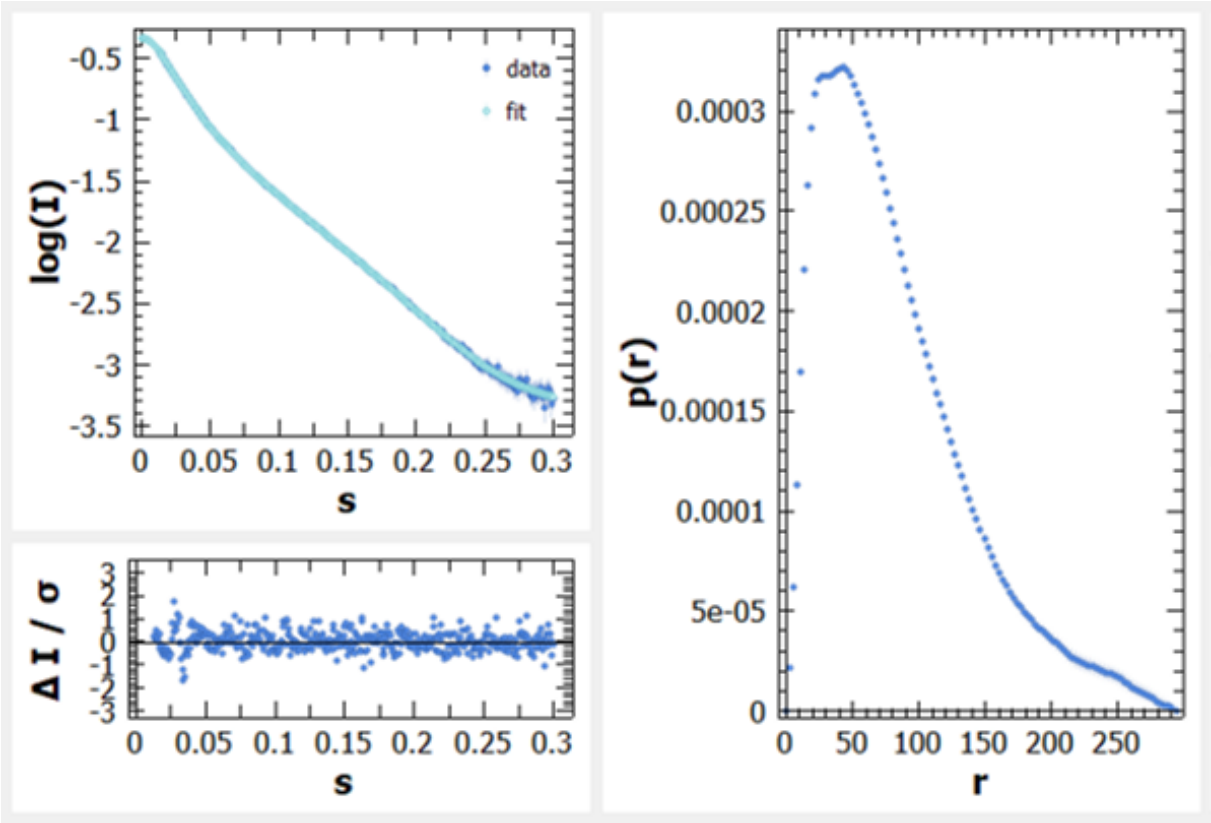

Figure A.79: *P(r)* plot of HP1α WT plus TERRA45 SAXS data. *P(r)* versus r profile from HP1α WT plus TERRA45 SAXS data. A fit of the back-transformed *P(r)* curve to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

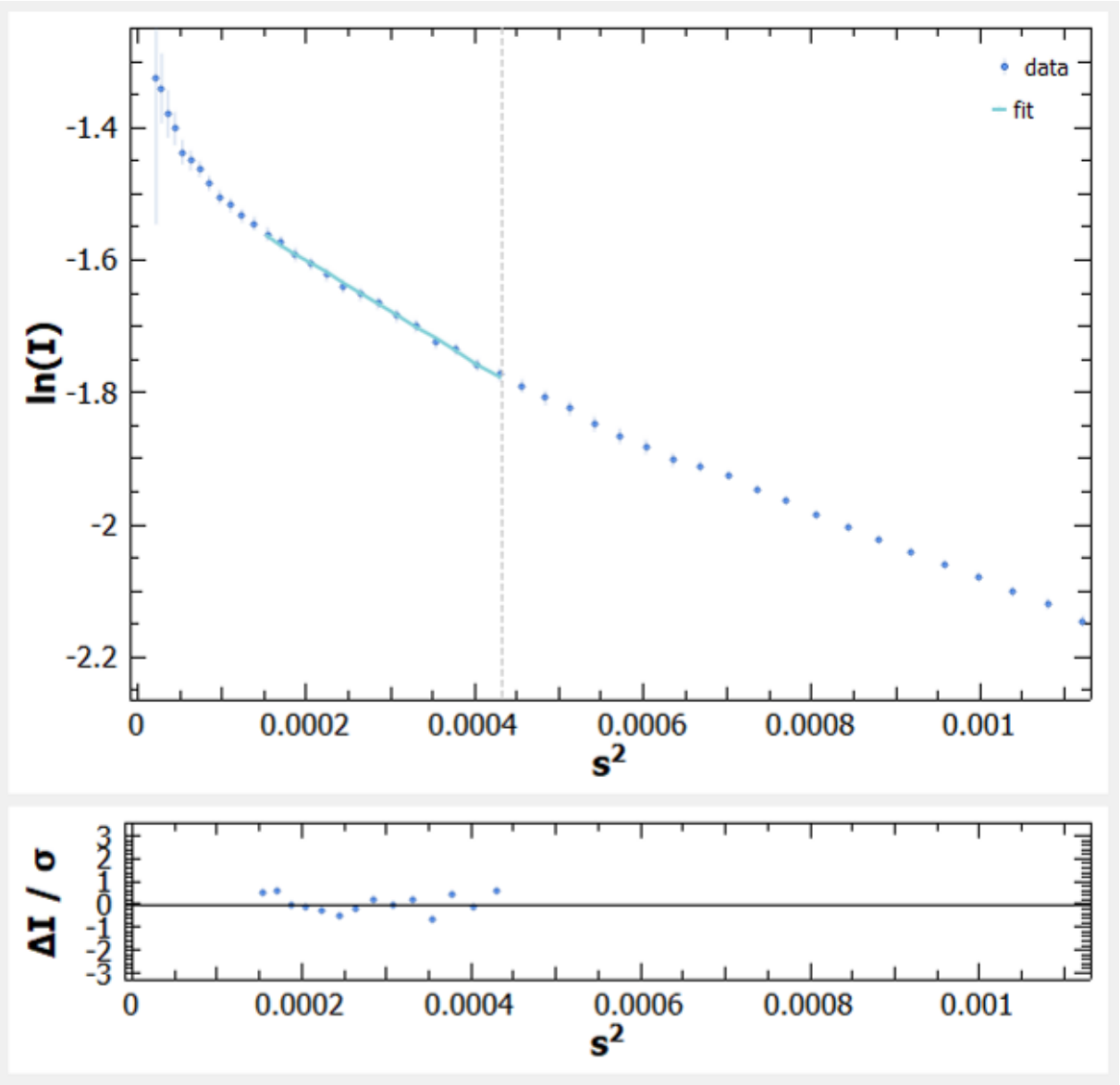

Figure A.80: Guinier fit of HP1α WT plus mutTERRA45 SAXS data. Linear Guinier fit to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

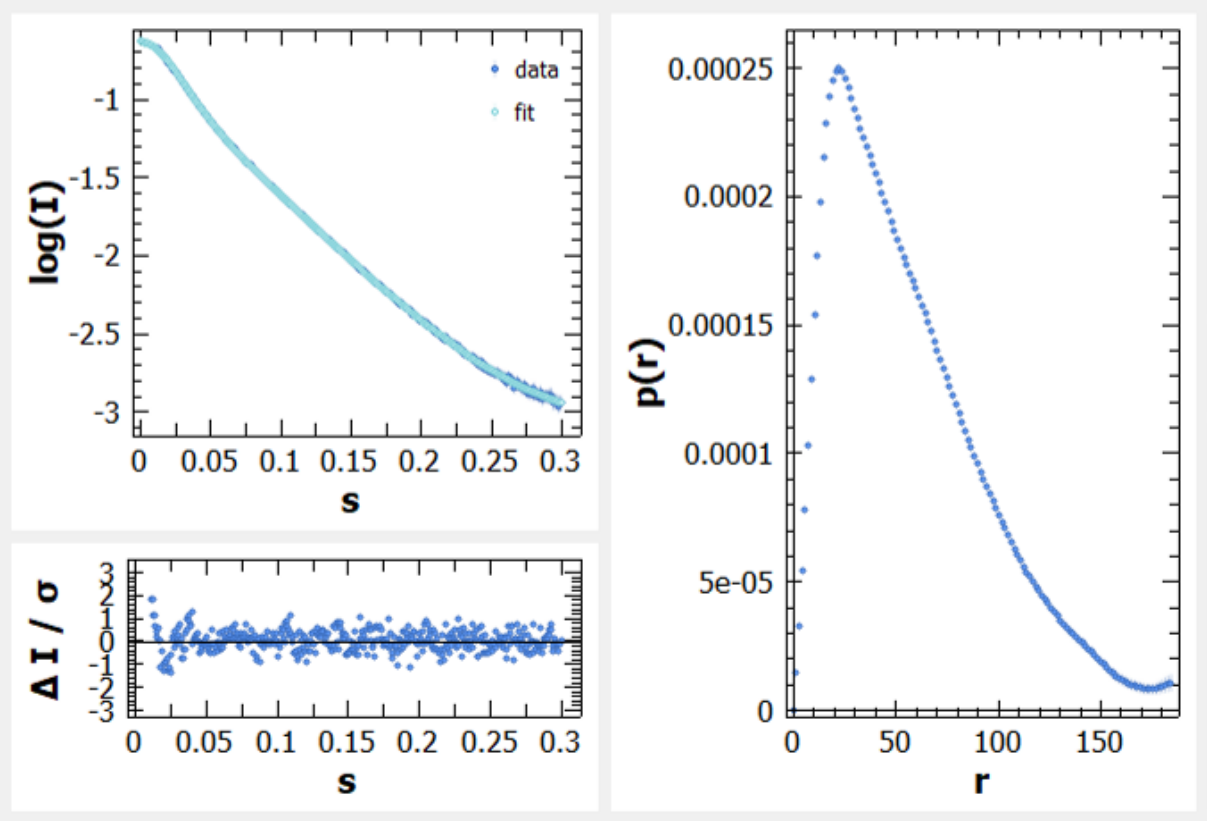

Figure A.81: *P(r)* plot of HP1α WT plus mutTERRA45 SAXS data. *P(r)* versus r profile from HP1α WT plus mutTERRA45 SAXS data. A fit of the back-transformed *P(r)* curve to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

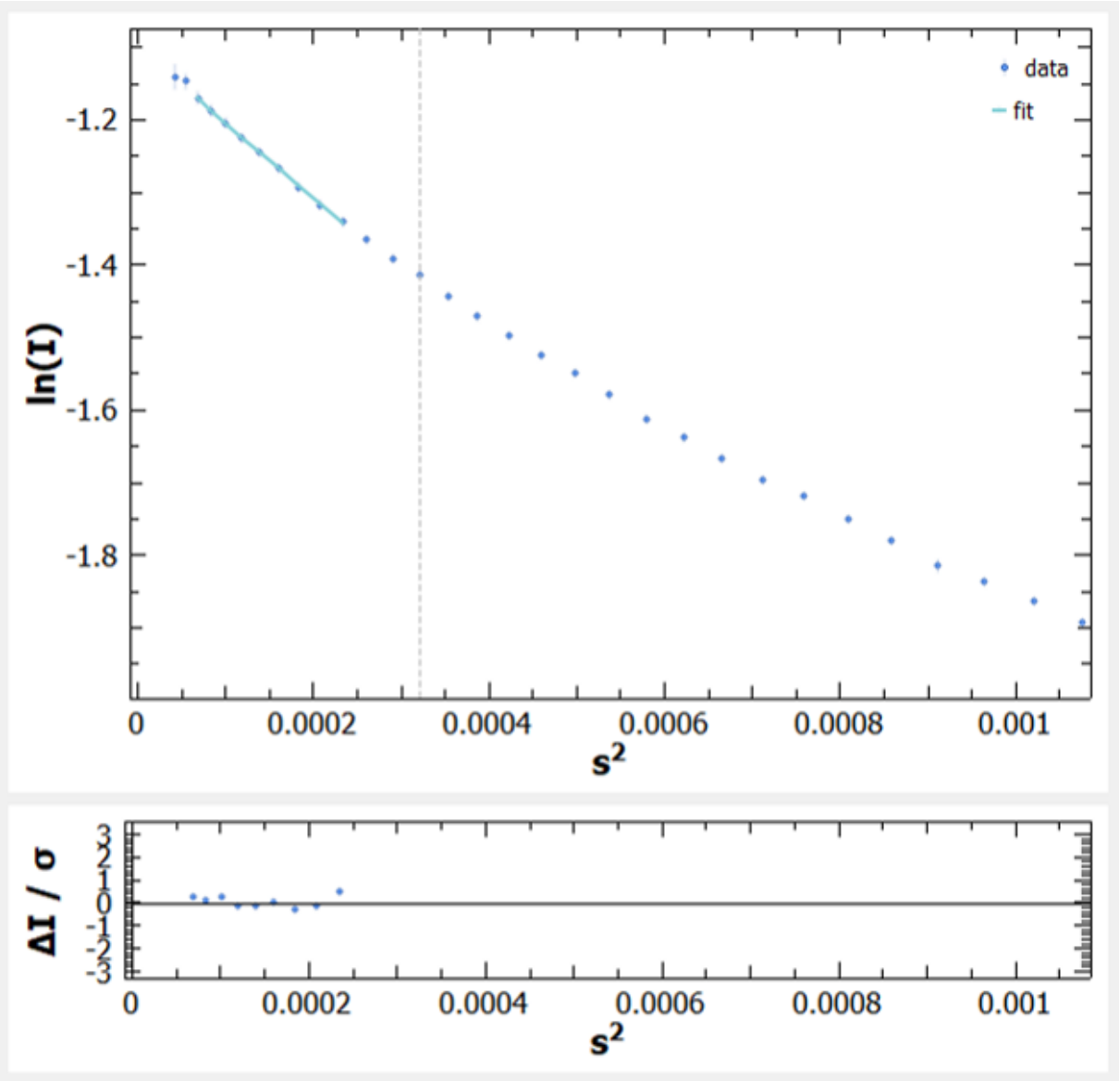

Figure A.82: Guinier fit of HP1α Hinge+CSD+CTE plus TERRA45 SAXS data. Linear Guinier fit to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

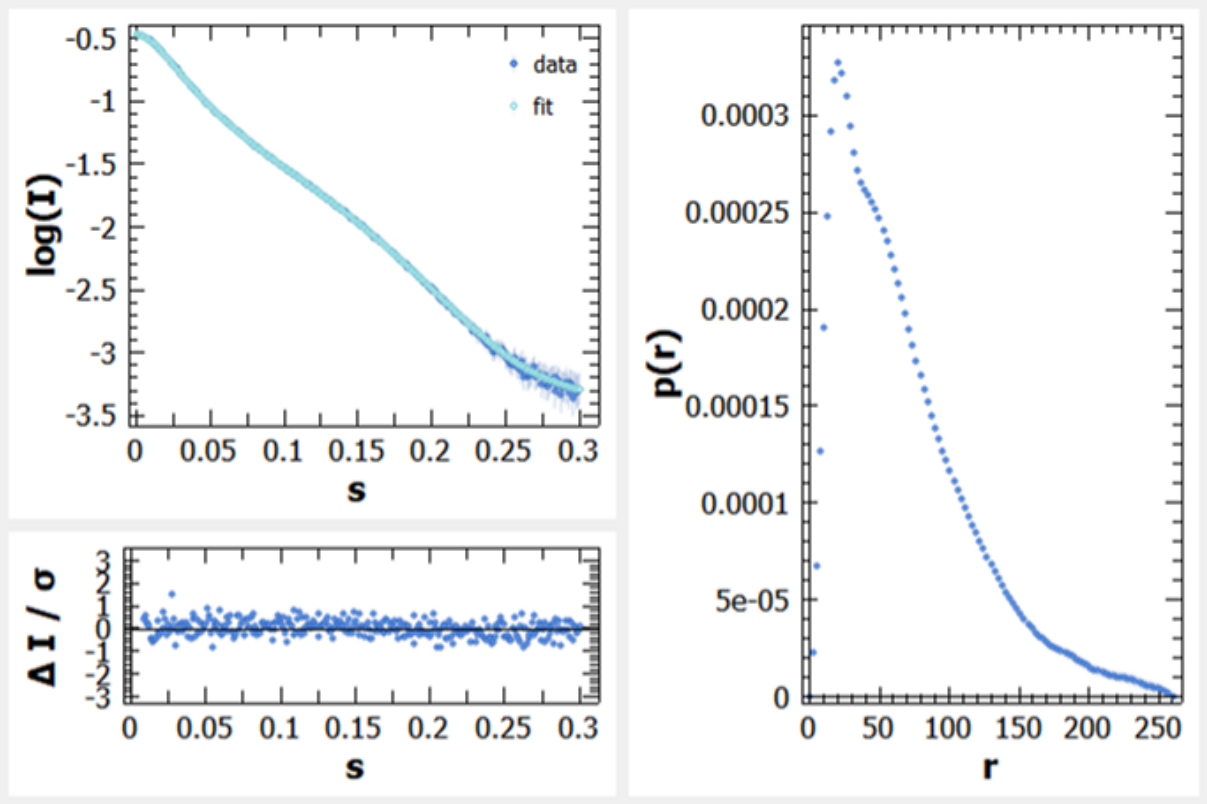

Figure A.83: *P(r)* plot of HP1α Hinge+CSD+CTE plus TERRA45 SAXS data. *P(r)* versus r profile from HP1α Hinge+CSD+CTE plus TERRA45 SAXS data. A fit of the back-transformed *P(r)* curve to the data is shown on the left with a plot of residuals underneath, where "s" represents the scattering vector/momentum transfer (q).

Fin