Investigating the biological phase separation of CBP; a balance of opposing forces.

Katie Gelder

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

University of Sheffield

Faculty of Science

School of Biosciences

20th October 2023

Acknowledgements

First and foremost, I would like to thank my supervisor Dan Bose, for this opportunity and for the constant support and guidance. You have made me the scientist I am today, you inspire me to be the change and I am extremely grateful to have had the opportunity to work with you throughout my PhD.

I would like to thank all the members of the Bose lab, past and present, for sharing this journey with me and being my cheerleaders throughout this experience. I'd like to thank Petra Celadova, for your wisdom and adventure; because of you I will never forget to stop and smell the roses. To Nicola Carruthers, I would not have been able to complete this journey without you. You have been my rock, my emotional support and my hype queen, and I am very grateful to have shared this experience with you. Thank you to Laura Harrison, your influence and advice has inspired me to be brave, and be different. I would like to thank Archna Shah for your unwavering support and advice, and for setting a great example of what a Postdoc should be. Thank you to Grace Gilbert, for always making me smile, and keeping my going during the highs and lows. Thank you Rabiatul Adawiyah, for your kindness in this last stretch. I'd also like to thank all the masters and undergraduate students along the way, for contributing to such a wonderful journey!

Additionally, I would like to thank Peter Daniels, for the smiles, wiggles and cups of tea; you have brought endless joy during this adventure! Thank you to Katy Boswell, for being a wonderful friend and peer over the last four years, and for Koda cuddles.

Big thanks to Emma Thomson and the Thomson lab for all your help along the way. Thank you to Holly Sutherland, for the company during quiet times, and the good times throughout the years. Thank you Vincent Chan, for your friendship, sass, and constant support. Thank you to Moni Feigenbutz, Sara Ilic, and Joe Cunningham for your kindness and laughter over the years.

I would like to thank Tim Craggs for always being excited by my work, and all of the Craggs lab for the advice and assistance given throughout my PhD. I would also like to thank Stuart Wilson and the Wilson lab for providing me with support, whenever it was needed. I would like

to thank my advisors, Julie Gray and Sherif El-Khamisy, for your pastoral support and for being a safe space for advice, scientific and otherwise.

Thank you to Ali Twelvetrees and the Twelvetrees lab, for welcoming me in the lab and providing guidance and support over the course of my project, including the use of BONSAI. Special thanks to Ash Davey, for being a ray of sunshine, you instil me with confidence and have been my greatest cheerleader. Thank you for the stories, and the laughs.

An extra thanks should be given to all members of our thesis writing group: Nicola Carruthers, Ash Davey, Peter Daniels and Vincent Chan. Thank you for the TEA, the support, and making this writing journey enjoyable!

I would also like to thank the Wolfsen Light Microscopy Facility, specifically Dr Nicholas Van Hateren and Dr Darren Robinson for the technical assistance throughout this project.

A massive thank you to my family and friends, who have cheered me on throughout this whole experience. To my parents, Tinabambina and Kev, I can not thank you enough for everything you have done to get me to this point. I will be forever grateful for your support, encouragement and G&GFMF! Thank you to my sister Paige, for showing me the way and keeping me laughing with every new bit. Thank you to Aaron Shaw, for supporting me throughout this whole journey, for believing in me and pushing me to succeed. Thank you to all the friends who have supported me along the way, you know who you are.

Finally, I would like to thank my examiners, Tatyana Shelkovnikova and Alex Borodavka for taking the time to examine my thesis and viva, it is greatly appreciated.

Abstract

CREB binding protein (CBP) is a transcriptional coactivator protein which has histone acetyltransferase (HAT) activity and binds multiple transcription factors, and is therefore commonly found at enhancer regions. Enhancers are DNA regulatory elements which increase the transcription of their target genes by acting as transcription factor binding hubs. Recent evidence indicates phase separation, the formation of a condensate due to a high concentration of proteins and nucleic acids, as a mechanism of how enhancers donate transcriptional machinery to the promoter, to upregulate gene expression. Proteins commonly found at enhancer regions, including CBP, have been shown to phase separate; however, the importance of how phase separation affects protein activity and dynamics is only now becoming clear. I wanted to determine what drives CBP phase separation, and develop experimental systems to understand how phase separation affects CBP HAT activity.

I have shown that CBP forms phase separated condensates both at endogenous levels and when overexpressed. Utilising an overexpression model system I defined regions which influence the phase behaviour of CBP. These experiments suggested that the phase separation of CBP is not caused by a single region but it is the cumulative effect of different regions with opposing influences on phase separation. I further developed doxycycline inducible cell lines expressing GFP tagged CBP and selected phase separating mutants, to express near-endogenous levels of protein, and reduce concentration as a factor for influencing phase separation. Finally, I have started to validate a single molecule approach to monitor CBP dynamics and HAT activity within the context of the different phase separating mutants.

The toolkit developed within this doctoral work will broaden our knowledge on the diverse effects of phase separation on CBP HAT activity.

Declaration

I, the author, confirm that this Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means. This work has not been previously presented for an award at this, or any other, University.

Contents

List of Figures

List of Tables

Abbreviations

1,6 - HD	1,6 - Hexanediol
3C	Chromosome Conformation Capture
3D	three-dimensional
ACD	L-Acridonylalanine
AIL	Autoinhibitory Loop
ALS	Amyotrophic Lateral Sclerosis
ATXN2	Ataxin-2
BCA	Bicinchoninic Acid
BRD4	Bromodomain-containing protein 4
BSA	Bovine Syrum Albumin
BWA	Burrows-Wheeler Alignment
CBP	CREB Binding Protein
CBP-HaloTag	HEK293T - CBP-Flag-TEV-6His-HaloTag
CFID	CBP FUS Interaction Domain
ChIP	Chromatin Immunoprecipitation
ChIP-Seq	Chromatin Immunoprecipitation Sequencing
CIDER	Classification of Intrinsically Disordered Ensemble Relationships

CIP	Calf Intestinal Phosphatase
CLIP	Cross-Linking Immunoprecipitation
CLK1	cycle-like kinase 1
CREB	cAMP-Response-Element Binding protein
Csat	concentration saturation
CTCF	CCCTC-binding factor
DAPI	4,6-Diamidino-2-Phenylindole
DDX4	DEAD-box helicase 4
DMEM	Dublecco's High Glucose Modified Eagles Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
DoMs	Detection of Molecules
Dox	Doxycycline
DPBS	Dulbecco's Phosphate Buffered Saline
DSLM	Digital Scanning Light Sheet microscopy
DTT	Dithiothreitol
eRF1	eukarayotic Release Factor 1
eRNA	enhancer RNA
FACS	Fluorescence-Activated Cell Sorting
FBS	Fetal Bovine Syrum
FCR	Fraction of Charged Residues
FCS	Fluorescence Correlation Spectroscopy

FRAP	Fluorescence Recovery After Photobleaching
FRET	Fluorescence Resonance Energy Transfer
FTHH	Flag-Tev-His-HaloTag
FUS	Fused in Sarcoma
GCE	Genetic Code Expansion
GFP	Green Fluorescent Protein
HAT	Histone acetyltransferases
Hi-C	High throughput Chromosome Conformation Capture
HMT	Histone methyltransferases
HP1	Histone Protein 1
HRP	Horseradish Peroxidase
HV	High Voltage
iCLIP-Seq	improved Cross-Linking Immunoprecipitation followed by Sequencing
IDRs	Intrinsically Disordered Regions
IF	Immunofluorescence
IP	Immunoprecipitation
iRFP	near-infrared Fluorescent Protein
JF	Janelia Fluor
LB	Luria Broth
LCR	Locus Control Region
LLPS	Liquid-Liquid Phase Separation
LMF	Light Microscopy Facility

lncRNA	Long non coding RNA
MACS2	Model-based Analysis of ChIP-Seq
MED1	Mediator of RNA polymerase II transcription subunit 1
mESCs	mouse Embryonic Stem Cells
MOPs	3-(N-Morpholino) Propanesulphonic Acid
MW	Molecular Weight
NARDINI	Non-random Arrangement of Residues in Disordered Regions Inferred using Numerical Intermixing
nSBs	nuclear Stress Bodies
P/S	Penicillin Streptomycin
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
PHR	Photolyase Homology Region
PLL	Poly-L Lysine
PML	Promyelocytic Leukemia
PONDR	Predictor of Naturally Disordered Proteins
PPPS	Polymer-Polymer Phase Separation
PTM	Postranslational Modifications
rcf	relative centrifugal force
RICS	Raster Image Correlations Spectroscopy
RNAPII	RNA polymerase II
RNP	RNA protein
rpm	rounds per minute

saSPT	standard array Single Particle Tracking
SDS	Sodium Dodecyl Sulfate
SE	Superenhancer
SG	Stress Granules
SiTRaN	Sheffield Institute for Translational Nueroscience
smFRET	single molecule Fluorescence Resonance Energy Transfer
snRNPs	small nuclear Ribonucleoproteins
SPRIA	Single Particle Raster Image Analysis
sptPALM	single particle tracking Photoctivated Localisation Microscopy
SRSFs	serine/arginine-rich splicing factors
TADs	Topological Assosciated Domains
TAE	Tris-Acetate-EDTA
TBS	Tris Buffered Saline
TCO*A	{[(E)-cyclooct-2-en-1-yl]oxy}carbonyl)-L-lysine
TDP-43	TAR DNA-binding Protein 43
TIRF	Total Internal Reflection Fluorescence
tmFRET	transition metal-ion Fluorescence Resonance Energy Transfer
uaa	unnatural amino acid
WT	Wild Type
YAP	Yes Associated Protein

Chapter 1

Introduction

1.1 Introduction to biological phase separation

Over the past decade a plethora of research has been performed into the phenomena of phase separation, characterising why proteins phase separate and what effect this has on their function. The term phase separation is used to describe the spontanious separation of a mixture of macromolecules into distinct phases, one which is enriched in macromolecules while the other is depleted (Banani et al. 2017; Hyman, Weber and Jülicher 2014; Shin and Brangwynne 2017; Alberti et al. 2018; Alberti, Gladfelter and Mittag 2019). Within a cell, the highly concentrated phase is typically composed of proteins and nucleic acids, compared to a dilute environment where there is no separating membrane (Chong and Forman-Kay 2016; Monahan et al. 2017; Sabari et al. 2018; Shin and Brangwynne 2017). Condensates are analogous to membraneless organelles and are thought to be caused by weak multivalent interactions (Boeynaems et al. 2018; Plys and Kingston 2018; Gibson et al. 2019a; Lin and Chan 2017; Peng and Weber 2019). In the literature, this concentrated state is referred to by multiple names such as "clusters", "droplets", "hubs", "granules", "foci", "puncta" and "condensates" (Chong, Vernon and Forman-Kay 2018; McSwiggen et al. 2019; Lin, Currie and Rosen 2017; Monahan et al. 2017; Sabari et al. 2018; Strom et al. 2017). Within this doctoral work they will be referred to either by puncta or condensates, interchangeably.

Condensates within the cell can be found in different phase states, including gel-like, sponge-like and liquid-like states (Boeynaems et al. 2018; Plys and Kingston 2018; Gibson et al. 2019a; Lin and Chan 2017; Alberti, Gladfelter and Mittag 2019; Peng and Weber 2019). Phase separation itself can occur through different mechanisms including polymer-polymer phase separation (PPPS) which relates to the bridging of proteins and nucleic acids in a solid-type structure (Peng

and Weber 2019). Another mechanism of phase separation is referred to as liquid-liquid phase separation (LLPS), due to the condensates ability to display liquid-like properties (Boeynaems et al. 2018; Plys and Kingston 2018; Gibson et al. 2019a; Lin and Chan 2017; Peng and Weber 2019). While LLPS is typically described as the mechanism of phase separation for many biological condensates, in the last couple of years this term has been brought into question; while condensates typically display fluid-like characteristics, these dynamics can also be seen for different structures such as sponge-type structures (Alberti, Gladfelter and Mittag 2019; Taylor et al. 2019). Without proper determination of the phase of these condensates it is sensible to move away from the term LLPS; however, regardless of phase, these condensates are universally shown to behave in a fluid-like manner as opposed to more solid aggregates (Boeynaems et al. 2018; Plys and Kingston 2018; Gibson et al. 2019a; Lin and Chan 2017; Peng and Weber 2019).

Protein aggregates are abnormal insoluble structures containing proteins; this typically happens if proteins are misfolded and has been commonly linked to disease, an example being the aggregation of microtubule associated protein Tau, into neurofibrillary tangles causing neuorodegeneration, leading to Alzheimer's disease (Wegmann et al. 2018; Venko and Žerovnik 2023). Protein aggregates are non-mobile structures in contrast to condensates which are thought to behave dynamically. Phase separated condensates are typically identified through microscopic visualisation of puncta either within a cell or *in vitro*; a key identifier of whether these condensates are formed through phase separation is by visualising punta fusion and dispersion, indicating dynamic behaviour.

In understanding phase separation, a pressing question to answer is how proteins form condensates and for what purpose does this phenomena exist. Research has been performed to determine common properties of proteins that phase separate and what factors can influence condensate formation, with important factors discussed below.

1.1.1 Properties of proteins that phase separate

A common feature of proteins that phase separate is that they have intrinsic disordered regions (IDRs), which are regions that have very little secondary or tertiary structure (Larson et al. 2017; Plys and Kingston 2018; Lin and Chan 2017; Nair et al. 2019; Sabari et al. 2018; Taylor et al. 2019; Ptitsyn et al. 1990; Redfield, Smith and Dobson 1994; Shin et al. 2018). These IDRs are thought to facilitate protein function, especially in RNA binding proteins and transcription factors, allowing flexibility to bind to multiple different partners (Chong and Forman-Kay 2016; Lin, Currie and Rosen 2017; Pak et al. 2016; Peng and Kurgan 2015; Plys and Kingston 2018;

Protter et al. 2018; Xue et al. 2010). Phase separation is thought to occur between these different IDRs, where their lack in structure facilitates weak intermolecular interactions between different proteins and nucleic acids, that hold the condensate together, compared to the dilute environment (Figure 1.1).

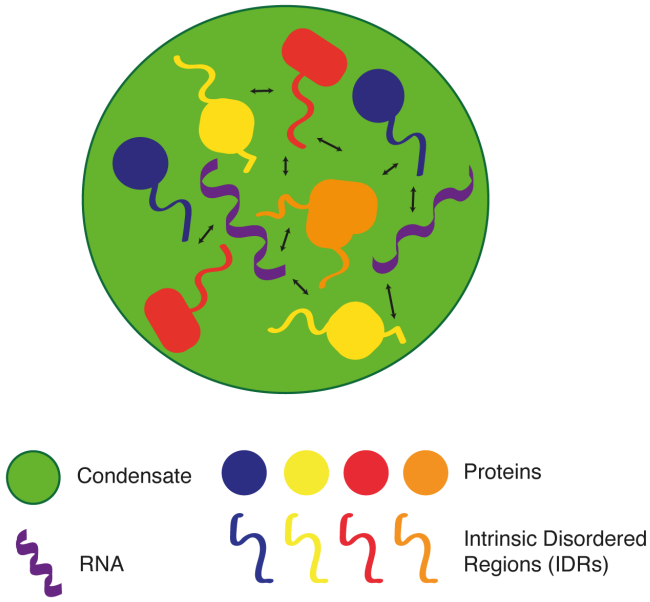

Figure 1.1: IDRs facilitate condensate formation through weak intermolecular interactions
Weak multivalent interactions (black arrows) between proteins (coloured circles) and their intrinsic disordered regions (IDRs, coloured wavy lines) facilitate the formation of a highly concentrated condensate (green circle) compared to a dilute environment, where there is no separating membrane.

These intermolecular interactions have been characterised more specifically at an amino acid level, to determine key residues and interactions which are important for this behaviour.

1.1.1.1 Amino acid residues which are important for phase separation

Disordered proteins have been characterised to have a high proportion of polar and charged amino acids, with a low proportion of hydrophobic residues (Lee et al. 2014; Rani, Baruah and Biswas 2014; Oldfield et al. 2005). These few hydrophobic residues are regularly interspersed between other residues, a feature which appears to be conserved among proteins that form condensates

(Keppel, Howard and Weis 2011; Mittag and Parker 2018; Nair et al. 2019; Schmidt, Barreau and Rohatgi 2019). Specifically, polar and aromatic amino acids are thought to drive phase separation; with two particularly important residues, tyrosine and serine, being described in the literature (Mittag and Parker 2018; Boeynaems et al. 2018; Pessina et al. 2019; Riback et al. 2017; Sabari et al. 2018; Pak et al. 2016). Tyrosine has been shown to be imperative for the phase separation of DNA/RNA-binding protein Fused in Sarcoma (FUS), and this role is conserved in other phase separating proteins (Banani et al. 2017; Lin and Chan 2017; Molliex et al. 2015). Phase separation of mediator of RNA polymerase II transcription subunit 1 (MED1) can be inhibited when key serine residues are mutated to alanine residues (Sabari et al. 2018). A similar observation was seen for p53 binding protein (53BP1), where an increase in polar residues, such as serines, was shown to facilitate condensate generation (Pessina et al. 2019).

Post translationsal modifications in phase separation Post translational modifications (PTM) of these residues are also thought to play a key role in regulating phase behaviour, with phosphorylation being highlighted as an important regulator of condensate formation (Carlson et al. 2020; Han et al. 2020; Boehning et al. 2018; Monahan et al. 2017; Lin, Currie and Rosen 2017; Murray et al. 2017). Phosphorylation of serine residues has been shown to dissolve numerous types of condensates. For example, phosphorylation of bromodomain-containing protein 4 (BRD4) was seen to diminish condensate formation, suggesting a negative regulation of condensate formation (Carlson et al. 2020; Han et al. 2020; Boehning et al. 2018; Monahan et al. 2017; Lin, Currie and Rosen 2017; Murray et al. 2017). Phosphorylation has also been shown to alter the phase of nucleocapsid (N) protein of coronavirus SARS-CoV-2, where it transformed gel-like compartments into more fluid-like condensates, which was thought to facilitate viral genome processing (Carlson et al. 2020).

Another PTM which has been shown to be important for phase separation, is the methylation of arginine residues (Wang et al. 2022a; Owen and Shewmaker 2019; Hofweber et al. 2018; Qamar et al. 2018). Methylation of arginine does not alter the positive charge of the side chain however it does effect the volume of the group and the charge distribution (Owen and Shewmaker 2019; Wang et al. 2022a). In the context of phase separation this affects the hydrophobicity of the side chain and the ability of the protein to form hydrogen bonds, which may alter a proteins ability to phase separate (Owen and Shewmaker 2019; Wang et al. 2022a). One example of this is Fused in sarcoma (FUS), where methylation of arginine has been indicated as important for regulating the recruitment of FUS into condensates (Qamar et al. 2018; Hofweber and Dormann 2019; Hofweber et al. 2018). FUS is an RNA binding protein which is commonly misregulated in disease, and has

been shown to phase separate. Evidence suggests that the hypomethylation of arginine residues in FUS strongly promotes phase separation, and further gelation (Qamar et al. 2018; Hofweber and Dormann 2019; Hofweber et al. 2018).

Acetylation is another PTM that has been suggested as a mechanism to prevent phase separation (as reviewed by Owen and Shewmaker (2019)). *In vitro* studies of dead-box RNA helicase 3 (DDX3X), have shown that acetylation of the N-terminal IDR1 can disrupt phase separation, preventing DDX3X recruitment to stress granules (SG) (Saito et al. 2019). DDX3X is a known substrate of CBP, and upon inhibition of CBPs catalytic activity, an increased volume of SGs was observed (Saito et al. 2019). Mutant DDX3X, where key lysine residues were replaces with acetyllysine mimic, glutamine, loses its capability to interact with SG proteins, preventing the maturation of SGs (Saito et al. 2019). In a similar manner, acetylation of key lysine residues on microtubule associated protein Tau, has been shown to disrupt phase separation (Ferreon et al. 2018). *In vitro* studies of Tau, showed hyperacetylation upon incubation with catalytically active p300, which displayed a reduced propensity to phase separate, preventing further aggregation (Ferreon et al. 2018). Therefore suggesting a mechanism for regulating phase separation through acetylation, with potential therapeutic links if acetylation can prevent Tau aggregation, linked to Alzheimer's disease (Ferreon et al. 2018). In both of these cases, the reduction in phase separation is thought to be caused by disruption of cation-π interactions which occur when the positive charge of the lysine residues is neutralised by the acetyl group (Ferreon et al. 2018; Saito et al. 2019; Owen and Shewmaker 2019). This further highlights the importance of PTMs on a proteins phase separating ability. For a comprehensive review on the effect of PTMs on phase separation see Owen and Shewmaker (2019).

Alongside the amino acids and PTMs which are important for regulating phase separation, the patterning of amino acids have also been described to be important for phase separation; one model of this patterning is the 'stickers-and-spacers' model. In this case, it is the pattern of stickers, which are residues which form weak interactions, and spacers, amino acids between the stickers that do not form interactions, which is important for forming the key interactions which facilitate phase separation (Zeng et al. 2020; Martin et al. 2020; Choi, Dar and Pappu 2019; Choi, Holehouse and Pappu 2020).

These residues, PTMs and patterning are important for regulating phase separation as they alter the weak intermolecular interactions that facilitate the formation and dissolution of condensates.

1.1.1.2 Forces that influence phase separation

Generally it is stated in the literature that weak multivalent interactions between proteins and nucleic acids facilitate phase separation. Multivalent interactions occur between multiple units, forming strong, reversible interactions which are usually the result of different processes occurring simultaneously (Haag 2015).

One interaction thought to be important is from hydrophobic interactions within proteins, where hydrophobic residues cause the displacement of water. Hydrophobic interactions are often perturbed when studying phase separation using 1,6 - hexanediol (1,6-HD), which causes condensates to disperse if they are formed through hydrophobic interactions (Molliex et al. 2015; Strom et al. 2017; Chong et al. 2018; Sabari et al. 2018). One example of this was shown when nuclear ER-α-dependent megadalton-scale protein complex (MegaTrans) condensates were dissolved upon the addition of 1,6 -HD, implying that hydrophobic interactions are necessary for condensate formation (Lin and Chan 2017; Nair et al. 2019).

The presence of aromatic and polar residues are also important in generating π-π and cation-π interactions (Banani et al. 2017; Chong, Vernon and Forman-Kay 2018; Nott, Craggs and Baldwin 2016; Pak et al. 2016; Mittag and Parker 2018). π-π interactions occur between two aromatic residues, where the most simple form of this interaction is generated based on attraction between the slightly positive benzene rings overcoming the repulsion between the negative electrons (Hunter and Sanders 1990; Molčanov and Kojić-Prodić 2019; Zhao and Zhang 2016). For example stacking interactions are thought to be important in the phase separation of hNRNPA1, where phenylalanine and tyrosine residues generate stacking interactions which facilitate condensate formation (Figure 1.2 A) (Molliex et al. 2015). Cation-π interactions are formed between aromatic residues and positive cations, where delocalised π electrons in the aromatic residues generate a slight negative charge that attracts positive charges, such as the side chains of arginine and lysine (Dougherty 1996; Ma and Dougherty 1997). Interactions between phenylalanine and arginine, have been shown to be important for DEAD-box helicase 4 (DDX4), forming critical interactions in the N-terminal disordered regions required for phase separation (Figure 1.2 B) (Nott, Craggs and Baldwin 2016).

Other forces can also contribute to condensates, such as hydrogen bonding and electrostatic interactions between charged amino acids (Chong et al. 2018; St George-Hyslop et al. 2018). For example in FUS, charged interactions between residues such as glutamate and arginine are thought to be responsible for long range interactions that facilitate phase separation (Figure 1.2 C) (St George-Hyslop et al. 2018).

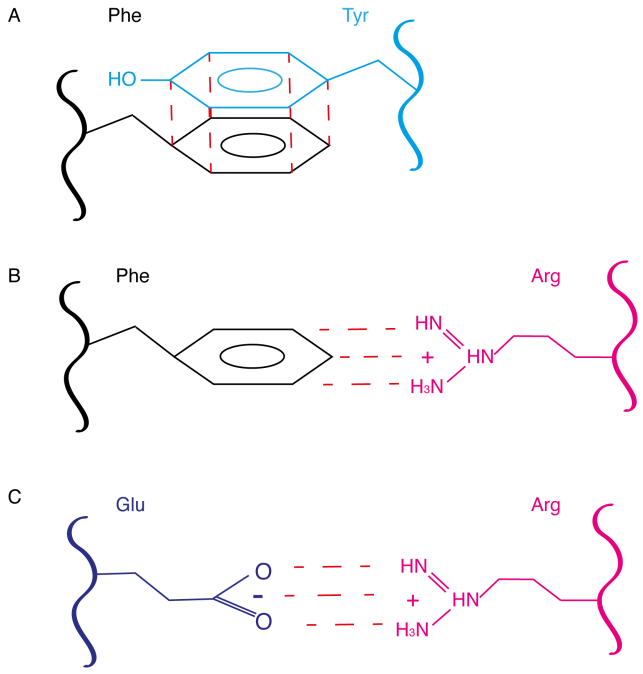

Figure 1.2: Weak interactions between amino acids that have been documented to influence phase separation
A - Example of stacking interactions by the ring structures in phenylalanine (black) and tyrosine (blue). B - Interactions generated by an aromatic residue such as phenylalanine (black) and arginine (pink). C - Example of charged interactions, the strongest of the weak forces seen to contribute to phase separation. Here the interactions can be seen to occur between glutamate (navy) and arginine (pink). Interactions in this diagram are represented by red dotted lines.

While individually these weak intermolecular interactions have been shown to affect phase separation, it is usually a combination of weak forces between different proteins and nucleic acids that holds these condensates together. The nature of the forces involved is dependant on the sequence of the components which are involved in phase separation. To demonstrate the principles of phase separation I will give an example of a well characterised case study of a protein which has been shown to phase separate in the literature, FUS.

1.1.1.3 An example protein that phase separates: Fused in Sarcoma (FUS)

Fused in sarcoma (FUS) is involved in transcription and DNA repair, where translocations in the N-terminal region are implicated in causing diseases such as Amyotrophic Lateral Sclerosis (ALS), as well as cancers (Shang and Huang 2016). FUS like many proteins form condensates with RNA, which are termed RNA protein (RNP) granules (An et al. 2022; Luo et al. 2018; Murakami et al. 2015); while RNA is important for the phase separation of FUS, the focus within this section is looking specifically at the dynamic protein behaviour within a condensate. The N-terminal region of FUS is intrinsically disordered and has a low complexity (LC) domain which is thought to be the driving force for phase separation (Figure 1.3); FUS mutants missing the LC region are unable to form condensates (Murakami et al. 2015; Patel et al. 2015). FUS containing condensates have been characterised as spherical, using digital scanning light sheet microscopy (DSLM), with liquid like properties (Patel et al. 2015). Here, a half-bleach technique was used to bleach half of the droplet area, where rapid rearrangement of FUS from the unbleached area to the bleached area was observed, indicative of a high level of mobility (Patel et al. 2015). Observations of fusion and deformation of FUS condensates under shear flow was also documented providing further evidence of the liquid-like properties of FUS (Patel et al. 2015).

The poly (ADP) ribose (PAR) binding domain in FUS is thought to be important for phase separation, due to interaction with PAR polymerase 1 (PARP1) which causes FUS accumulation at sites of DNA damage. When PARP1 was inhibited, FUS did not form phase separated condensates at these sites (Patel et al. 2015). As mentioned above, specific residues are key to phase separation, with the QGSY repeat in the LC domain and the RGG repeats in the middle and C-terminal being crucial to cation-π interactions which facilitate phase separation (St George-Hyslop et al. 2018). Interactions with other molecules can positively or negatively effect phase separation of FUS by promoting or inhibiting the formation of a condensate. For example, when FUS interacts via a critical arginine with molecular chaperone transportin 1, condensate formation was prevented (Qamar et al. 2018; St George-Hyslop et al. 2018).

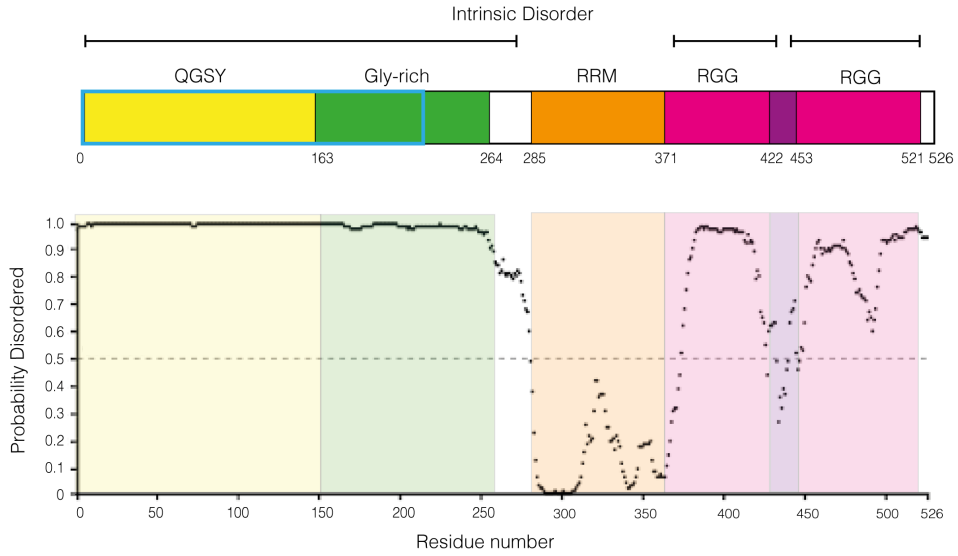

Figure 1.3: Intrinsic disorder profile of FUS
Domain diagram of fused in sarcoma (FUS) showing the glutamate - glycine - serine - tyrosine repeat region (QGSY - yellow), glycine rich region (green), RNA recognition motif (orange), arginine - glycine - glycine repeat domains (RGG - pink), zinc finger motif (purple) and the low complexity domain (blue outline). Domain diagram based on figure found at Luo et al. (2018) and Monahan et al. (2017). The disordered regions are indicated above the domain diagram, with the residue number of the domain written below. Graphical representation of disorder is also shown below with domains indicated on the graph in their respective colours, where 1.0 indicates disordered regions and 0 indicates no disorder. The graphs of intrinsic disorder were generated using DISOPRED 3, for reference see (Jones and Cozzetto 2015).

Since it is known that FUS can phase separate, and has many important roles within cells, it is thought that phase separation may be important for correct function of FUS. Proteins are often involved in the DNA damage response (DDR) which acts to maintain genome stability; phase separation of FUS has been implicated in the initiation of the DDR, as when FUS was knocked-out of cells, key proteins were not recruited to the site of DNA damage, implicating FUS in protein recruitment (Levone et al. 2020). FUS has been extensively studied due to it's role in diseases, such as ALS; the N-terminal region of FUS has been suggested to cause pathogenesis by facilitating phase separation and further hydrogel formation, where these hydrogels have the potential to nucleate pathogenic fibrous structures (Murakami et al. 2015; Patel et al. 2015; Shang and Huang 2016; St George-Hyslop et al. 2018). Disease associated mutations of FUS have been

shown to increase the rate of phase separation and subsequent gelation, which is thought to lead to a more severe disease phenotype (St George-Hyslop et al. 2018). Moreover, the liquid-like properties of FUS can be altered by phosphorylation or the use of a phosphomimetic mutant; for example using the phosphomimetic amino acid glutamate which is negatively charged (Lin, Currie and Rosen 2017; Monahan et al. 2017; Murray et al. 2017). When phosphorylated, condensates of FUS are more diffuse and have decreased toxicity (Lin, Currie and Rosen 2017; Monahan et al. 2017; Murray et al. 2017).

This highlights the importance of studying phase separation in the context of disease, with the potential to develop novel therapeutics to target proteins that phase separate.

1.2 Studying protein phase separation

Investigating phase separation involves understanding which components are found within a condensate and how this behaviour is regulated.

1.2.1 Bioinformatical tools to understand protein phase separation

A proteins ability to phase separate is often investigated first using bioinformatical tools. Protein phase separation is thought to occur through IDRs which are able to form different conformations and form weak contacts with other proteins and nucleic acids, facilitating condensate formation (Lotthammer et al. 2023; Borcherds et al. 2020; Martin and Holehouse 2020; Wei et al. 2017; Sun et al. 2019). The percentage disorder of several heterochromatin phase separating proteins has been calculated at approximately 44 %, which is comparable to other proteins that phase separate within the proteome (Guthmann, Burton and Torres-Padilla 2019; Sawyer, Sturgill and Dundr 2019). This discovery has led to the development of multiple different disorder prediction softwares to determine regions of intrinsic disorder, which implies whether or not a protein may phase separate; these include DISOPRED3, PONDR, IUPred2 and ANCHOR (Jones and Cozzetto 2015; Xue et al. 2010; Dosztányi, Mészáros and Simon 2009; Dosztányi 2018). These softwares determine the likelihood of each amino acid being disordered, from the amino acid sequence of the protein of interest, from this disorder can be mapped within the whole protein (Jones and Cozzetto 2015; Xue et al. 2010; Dosztányi, Mészáros and Simon 2009; Dosztányi 2018).

Bioinformatic tools can not only predict regions of disorder, but also patterns of amino acids which maybe important for phase separation. One tool which has been developed to do this is Classification of Intrinsically Disordered Ensemble Relationships (CIDER) a web server that ana-

lyses the primary sequence of IDRs, as well as localCIDER which is the locally installable software package (Holehouse et al. 2017). This package allows for sequence analysis of parameters such as hydrophobicity, number and spacing of charged residues which may influence phase separation. Sequence variants from this analysis provides an indication of which amino acids or pattern of amino acids maybe important for phase separation. Another software package, available as an extension of localCIDER is Non-random Arrangement of Residues in Disordered Regions Inferred using Numerical Intermixing (NARDINI) (Cohan et al. 2022). This software was made to extract sequence binary patterns within intrinsically disordered proteins, with the aim to identify sequence features which are important for phase separation (Cohan et al. 2022). Together these tools allow for the prediction of IDRs as well as specific sequence patterns which may be responsible for protein phase behaviour.

If a protein contains IDRs this indicates that it may have the ability to form phase separated condensates, from which experimental tools can be used to test this hypothesis.

1.2.2 Experimental tools to understand protein phase separation

Initial experiments to test whether proteins form phase separated condensates involve tagging the protein with a fluorophore such as GFP, expressing the tagged protein in cells and observing the pattern of expression (examples of this can be found in Nott et al. (2015), Qamar et al. (2018), Ma et al. (2021), Wang et al. (2022b) and Zhang et al. (2021)); this can be diffuse across the nucleus for example, or localised to specific regions in a more punctate pattern. Visualisation of puncta using a microscope either within a cell or *in vitro* is usually the first experiment performed when understanding whether a protein phase separates or not. If puncta are observed this may indicate that the protein of interest forms phase separated condensates; however at this stage further tests need to be performed to determine whether these clusters of protein are fluid-like, which is typical of a condensate, or more solid-like in behaviour.

One of the first experiments usually performed when investigating phase separation is to visualise puncta fusion and dispersion through microscopy. Phase separated condensates typically display fluid-like properties, as such these puncta will fuse together and disperse apart, as is typically expected of a liquid (Figure 1.4 A), this would indicate whether a protein behaves in a mobile or solid manner (Mitrea et al. 2022; Sanulli and Narlikar 2021). Alongside this, experiments such as fluorescence recovery after photobleaching (FRAP) are used to determine how dynamic a protein is (Gibson et al. 2019b; Mittag and Parker 2018; Monahan et al. 2017; Niaki et al. 2019; Nott et al. 2015; Nott, Craggs and Baldwin 2016; Riback et al. 2017; Sabari et al. 2018; Strom

et al. 2017).

In FRAP one region of interest is targeted by a laser, this irreversibly bleaches the molecules within the bleached region, the recovery of fluorescence is then monitored over time, as new molecules of protein with unbleached fluorophores move back into that space (Figure 1.4 B) (Gibson et al. 2019b; Mittag and Parker 2018; Monahan et al. 2017; Niaki et al. 2019; Nott et al. 2015; Nott, Craggs and Baldwin 2016; Riback et al. 2017; Sabari et al. 2018; Strom et al. 2017). The amount of bleaching that occurs and how quickly the fluorescence recovers is directly linked to how easily the protein can diffuse. If a protein bleaches and recovers over a timescale of seconds this is indicative of a phase separating protein (Gibson et al. 2019b; Mittag and Parker 2018; Monahan et al. 2017; Niaki et al. 2019; Nott et al. 2015; Nott, Craggs and Baldwin 2016; Riback et al. 2017; Sabari et al. 2018; Strom et al. 2017). The region to be bleached could be a whole condensate, or a smaller region within the condensate could be targetted, following a technique known as half-FRAP, where internal motion within a condensate can be observed (Muzzopappa et al. 2022). While the half-FRAP approach is useful in looking at mobility within a condensate, this technique is limited by the size of the condensate, and the minimum size the laser can bleach, in some cases the half-FRAP approach will not be feasible. Although FRAP is a technique that can identify different molecular mobilities, it is clear that a lot of emphasis has been put on using FRAP as evidence of phase separation (Mittag and Pappu 2022; Musacchio 2022). Care must be taken when drawing conclusions about phase separation based on a liquid-like phenotype from FRAP, however it is a useful approach for comparing between samples. Another technique which is also used to determine the dynamic behaviour of a protein is 1,6-Hexanediol treatment (1,6-HD).

Phase separation is thought to be facilitated through weak intermolecular interactions - these include Van der Waals, weak electrostatic interactions and hydrophobic interactions. Which of these interactions are important is specific to the protein of interest, due to the specific sequence of the IDR. Using 1,6 - HD, hydrophobic interactions can be perturbed, causing the condensate to disperse if these interactions are important for phase separation (Figure 1.4 C) (Boehning et al. 2018; Chong et al. 2018; Molliex et al. 2015; Sabari et al. 2018; Strom et al. 2017). In the literature, this experiment is typically used as an indicator of whether a protein phase separates; however, in the last couple of years the validity of using this technique has come into question, as the use of hexanediols has been shown to have a negative impact on cell physiology (Musacchio 2022; Itoh et al. 2021; Ulianov et al. 2020; Sabari et al. 2018; Kroschwald, Maharana and Simon 2017; Maharana et al. 2018; Sehgal et al. 2020). While 1,6-HD may disturb hydrophobic interactions, it also has effects on kinase and phosphatase activity, as well as organisation of the 3D genome -

which in itself my alter phase separation (Düster et al. 2021; Ulianov et al. 2020); it is therefore important to understand the caveat that a direct link between addition of 1,6 - HD and disruption of condensates is not clear. Other aliphatic diols have been shown in the literature to disrupt liquid-like condensates in the same manner as 1,6-HD without causing toxic effects (Geiger et al. 2021; Mochida and Gomyoda 1987). 1,2 - propanediol, commonly callen propylene glycol, has been used to disperse liquid-like cytoplasmic viral factories, termed viroplasms, within cells; this compound is thought to be safe to use within cells at concentrations below 5 % (v/v) (Geiger et al. 2021; Mochida and Gomyoda 1987).

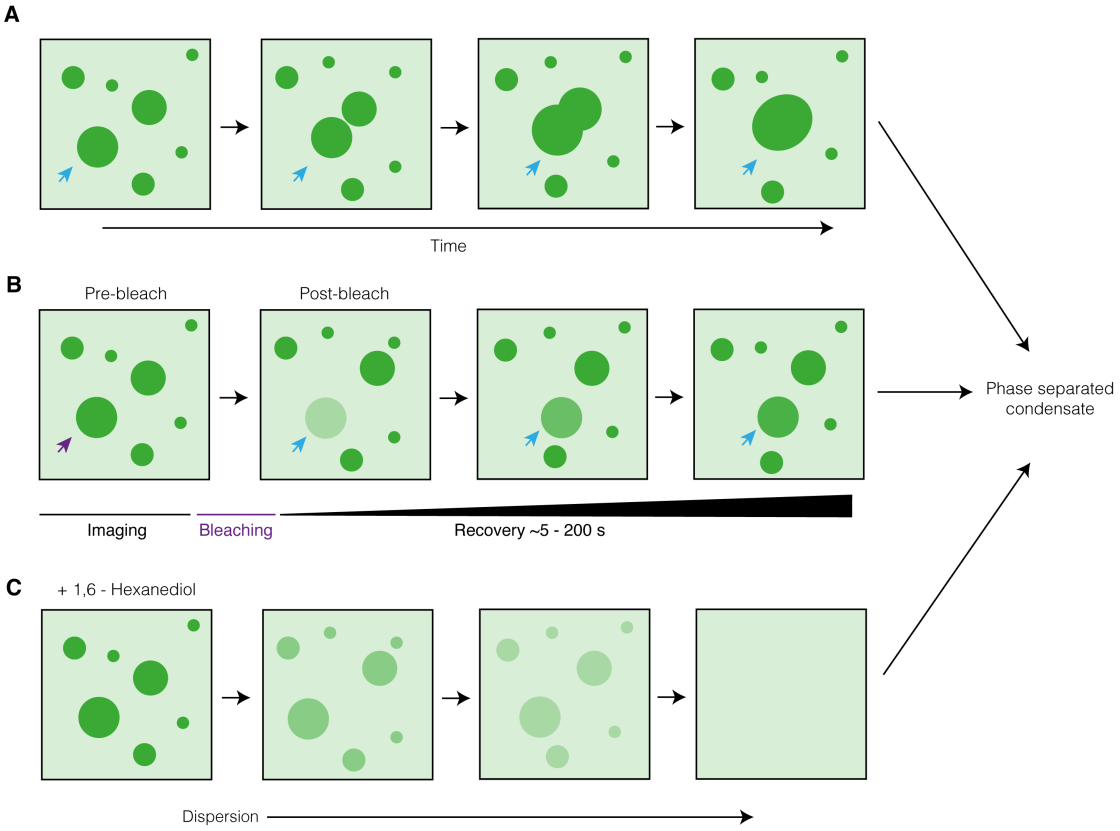

Figure 1.4: Techniques for determining dynamic behaviour of condensates
A - Illustration of condensate fusion of a GFP tagged protein that has phase separated, where condensates (green) can fuse together (blue arrow) to form larger condensates. B - GFP tagged condensates (green) that are photobleached (purple arrow), and recover fluorescence (blue arrow) over a time scale of ~ 5 - 200 seconds are thought to behave dynamically as a phase separated condensate. B - GFP tagged condensates (green) are dispersed when treated with 1,6 - hexanediol, behave in a manner consistent with a phase separated condensate.

If a protein has been shown to phase separate, using the techniques described above, investigations into characteristics of the protein sequence are usually performed to determine which regions are important for phase separation.

1.2.2.1 Understanding how specific regions within a protein are important for phase separation

Using bioinformatics tools, as described in subsection 1.2.1, specific regions within a protein can be identified which might be important for phase separation, including IDRs. Once regions of interest have been identified the influence of these regions on phase separation can be experimentally tested, looking at the sequence in isolation, and subsequently deleting these regions from the full length protein to observe how this alters the full length phase behaviour of the proteins.

Techniques such as the optoDroplet system have been developed to test the phase separation ability of IDRs independently (Shin et al. 2017; Bracha et al. 2018). This system uses blue light activation of an oligomerisation protein, to nucleate the condensation, which is fused to the IDR of interest and a fluorescent tag, to allow observation of condensate formation (Shin et al. 2017; Bracha et al. 2018). In the case of the optoDroplet system, the oligomerisation protein is a CRY2 protein from *Arabadopsis thaliana*, which is stimulated upon blue light to oligomerise (Shin et al. 2017; Basu et al. 2020). The optoDroplet assay was initially tested using IDRs from common phase separating proteins FUS, DDX4 and HNRNPA1, and has since been used by other groups to further investigate FUS, as well as proteins such as YTHDC1, a m6A reader protein, to understand how this protein affects transcriptional condensates in the context of eRNAs (Shin et al. 2017; Basu et al. 2020; Lee et al. 2021b; Ulianov et al. 2020). It has also been used in the study of transcription factor homeobox 13 (HOXD13), where an IDR and several mutants of the IDR were tested in the optoDroplet assay; WT IDR formed condensates upon blue light stimulation, suggesting this this region was capable of phase separation (Basu et al. 2020). Deletion of 7 and 15 alanine residues from this IDR region prevented the formation of condensates in the optoDroplet assay, implying that hydrophobic interactions may be important for the phase separation of this IDR and potentially full length HOXD13 (Basu et al. 2020). While a useful technique it is unknown how much influence the oligomerisation of the CRY2 domain affects the condensates observed using the optoDroplet system. It highlights the importance for controls such as using an empty CRY2 vector as a negative control to show the baseline restructuring that occurs upon blue light activation.

Recent advances have produced an updated experimental technique to test whether individual IDRs can phase separate that removes this caveat, known as the corelet system (Bracha et al. 2018; Garcia-Jove Navarro et al. 2019; Berry et al. 2015). This system utilises two constructs, instead of the one used in the optoDroplet system. The first construct is comprised of a nuclear localisation signal fused to an engineered protein iLID, as well as EGFP for visualisation, and the

human ferritin heavy chain (FTHI), while the second is made up of the IDR of interest fused to mCherry and SspB, the cognate partner to iLID. The FTHI self assembles into a 24-mer core with a width of 12 nm, which acts as a scaffold for IDR binding in the same manner as DNA or RNA is proposed to act as a scaffold for natural phase separation (Bracha et al. 2018; Garcia-Jove Navarro et al. 2019; Berry et al. 2015). The iLID domain fused to this FTHI module strongly dimerises with SspB, upon the addition of blue light, allowing a maximum of 24 copies of the IDR of interest to assemble. Like the optoDroplet, if condensates are formed over the timecourse, the IDR has the ability to phase separate. The advantage of this system is the SspB-iLID interaction has been characterised and has been shown not to directly contribute to interactions of the condensate, which relies solely on IDR-IDR interactions, opposed to the impact of the CRY2 dimerisation which is largely unknown. This mechanism also utilises two fluorescent proteins EGFP to track the cores and mCherry to visualise the IDRs, allowing distinction between the core module and IDR phase separation, and is finely tuneable by blue light activation allowing cycles of on/off condensation in response to blue light, which the optoDroplet system is not capable of (Bracha, Walls and Brangwynne 2019; Bracha et al. 2018). Assays such as the optoDroplet or more recently the corelet system highlight which regions, if any, are importanat in regulating phase behaviour either positively or negatively.

Another group used an *in vitro* approach to test IDRs that were identified using prediction software IUPred and VSL2 from EP300 (Wang et al. 2022b; Dosztányi 2018; Xue et al. 2010). Recombinant protein was purified from a bacterial expression system, producing individual IDRs which could be tested to see if they phase separate using *in vitro* experiments (Wang et al. 2022b). Using molecular corwding agent polyethylene glycol (PEG) 8000, two regions were identified which could phase separate independently, IDR3 and IDR5 (Wang et al. 2022b). A mutant construct of the full length protein was produced, where these two regions were deleted, showing a lack of phase separation of the resulting EP300 construct in cells (Wang et al. 2022b).

This method of investigation is common within the field of phase separation to determine which regions of proteins are important for phase separation, and identify any patterns which could be conserved between phase separating proteins. In understanding the relevance of phase separation it is important to understand the effect this has on protein function and the dynamics within a condensate. Testing the effect of condensates on protein function is very specific to the type of protein being investigated, and its method of catalytic activity. However, understanding the dynamics of a protein within phase separation is more generalised, and over the last decade methods of understanding this influence have been determined.

1.2.2.2 The effect of condensates on protein dynamics

In studying phase separation the mobility of proteins are often monitored through FRAP to check that the protein accumulation is caused through phase separation of a dynamic protein, as opposed to a more solid aggregation of proteins (Gibson et al. 2019b; Mittag and Parker 2018; Monahan et al. 2017; Niaki et al. 2019; Nott et al. 2015; Nott, Craggs and Baldwin 2016; Riback et al. 2017; Sabari et al. 2018; Strom et al. 2017). Techniques such as FRAP, facilitate the understanding of the dynamics within a whole population of one protein. To specifically understand the effect of phase separation on protein dynamics, several single molecule approaches have been developed which can reveal details about the effect of phase separation on individual protein molecules without averaging out over the whole population (Nasir et al. 2019; Chowdhury, Nettels and Schuler 2023; Mekonnen et al. 2023). These dynamics include monitoring the movement of the whole protein in and out of a condensate (Mekonnen et al. 2023; Miné-Hattab et al. 2021), as well as investigating the conformational dynamics of a protein within the condensate (Nasir et al. 2019; Mitrea et al. 2016; Mitrea et al. 2018); experiments to investigate these include single molecule microscopy and a technique called single molecule fluorescence resonance energy transfer (smFRET).

Imaging single molecules Single molecule imaging within a condensate allows for the calculation of diffusion coefficients of individual molecules of a protein, which allows the separation of different populations of proteins based on their diffusion pattern. This technique also enables the quantification of their movement, to work out if there is movement or confinement within particular regions within a condensate. This has been performed in relation to DNA repair foci, which are condensates that contain repair enzymes such as Rad52 (Mekonnen et al. 2023; Miné-Hattab et al. 2021). Photoactivatable tracking of single molecules was used to show that Rad52 molecules are confined within DNA repair foci and move at a diffusion rate which is 6 times faster than that of the foci itself, suggesting rapid movement of protein within the condensate (Mekonnen et al. 2023; Miné-Hattab et al. 2021); providing evidence of the dynamic nature of proteins within a condensate.

smFRET to study protein conformation dynamics Fluorescence resonance energy transfer (FRET) refers to the non-radiative transfer of energy from a high energy fluorophore to a lower energy fluorophore (Stryer and Haugland 1967; Murchie et al. 1989). This transfer in energy results in a decreased intensity of the donor and an increased intensity of the acceptor. For FRET to occur there needs to be an overlap between the donor emission and the acceptor absorbance,

and the fluorophores need to be in close proximity, between 2 - 10 nm (Figure 1.5).

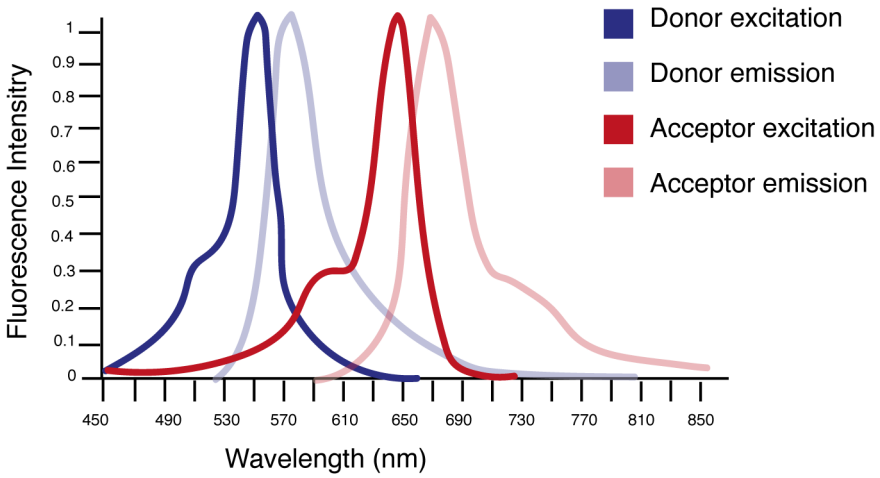

Figure 1.5: Absorption spectra of two example dyes used in smFRET experiments
For single molecule fluorescence resonance energy transfer (smFRET) to occur the donor emission spectra and the acceptor excitation spectra need to overlap. The Spectra above shows the excitation spectra (solid lines) of ATTO 550 (purple) an example donor fluorescent dye and ATTO 647N (red) an example acceptor fluorescent dye, and the emission spectra (faint lines). Any dyes of fluorescent molecules that show a similar donor/acceptor excitation and emission spectra can be used for smFRET. Figures based on images found at https://www.atto-tec.com/.

FRET efficiency (E) is calculated based on the acceptor emmission (F_a) and the donor emission (F_d) using the following equation:

$E = (F_a \,/\, F_a + F_d)$

The FRET efficiency is dependant on the distance between the two fluorophores, where these fluorophores are usually attached via flexible linkers, which means that their position is not strictly defined. Accessible volume modelling therefore needs to be done, to determine the radius around the dye where it can exist, to determine the interdye distance (R) (Hellenkamp et al. 2018). The FRET efficiency can then be calculated by looking at the relationship between the interdye distance (R) and the Forster radius (R_0), which is the distance between the dyes where the energy efficiency is 50 % (Hohlbein, Craggs and Cordes 2014).

$E = [1 + (R/R_0)^6]^{-1}$

Due to this relationship FRET can be used as a spectroscopic ruler where information can

be gathered about conformations based on the distance determined between the two fluorophores (Hohlbein, Craggs and Cordes 2014; Roy, Hohng and Ha 2008). The relationship between FRET efficiency and distance of the fluorophores is sigmoidal, when the distance between the fluorophores is small there is a high FRET efficiency, and when the distance is larger the FRET efficiency is lower (Figure 1.6) (Roy, Hohng and Ha 2008).

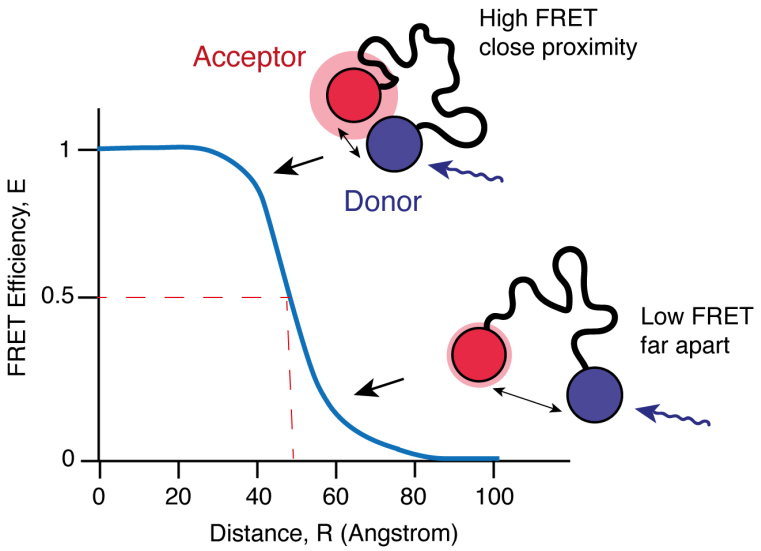

Figure 1.6: smFRET interaction between FRET efficiency and distance
The relationship between FRET efficiency and distance between donor (purple) and acceptor (red) fluorophores is sigmoidal. A high FRET efficiency corresponds to a small distance between fluorophores, when excitation of the donor (purple arrow) causes emmission from the acceptor. Whereas a low FRET efficiency represents a larger distance between the fluorophores, resulting in a reduced emission from the acceptor. To work out accurate FRET distances the R_0 between the two fluorophores needs to be calculated, which is the distance between them at which the FRET efficiency is 50 %, this is represented by the red dotted line. Figure based on one found at Roy, Hohng and Ha (2008).

FRET can be exploited to answer biological questions, for example by tagging two proteins to see if they interact (Margineanu et al. 2016). Alternatively this can be done on a single molecule level, single molecule FRET (smFRET), attaching different fluorescent dyes to the same protein or nucleic acid to monitor different FRET distances and provide information on single molecule intramolecular dynamics (Hellenkamp et al. 2018; Sasmal et al. 2016).

Nucleophosmin (NPM1) is a protein typically found within the nucleolus which binds proteins,

RNA and DNA, and has been shown to phase separate through a variety of weak interactions (Nasir et al. 2019; Mitrea et al. 2016; Mitrea et al. 2018). smFRET experiments have been used to monitor the typical conformation of NPM1 and after transitioning into a phase separated environment a shift in conformation was observed (Mitrea et al. 2016; Mitrea et al. 2018). While smFRET has not been commonly practised within phase separated condensates, it provides a unique opportunity to monitor protein conformation within a condensate; it therefore stands that if a protein changes conformation upon catalytic activation, smFRET could be used to monitor the activity of a protein both in and outside of a phase separated environment.

Phase separated condensates are difficult to investigate due to their complicated nature, with multiple different factors influencing the formation of a phase separated condensate.

1.2.3 Factors that influence phase separation

In physical terms, the formation of a condensate is a density transition that results in a highly dense phase compared to a more dilute environment, where even in the dense phase molecules can move relatively freely (Alberti and Dormann 2019; Ceballos, McDonald and Elbaum-Garfinkle 2018; Arzideh, Movagharnejad and Pirdashti 2018; Arter et al. 2022). At the cellular level, the environment dramatically affects the ability of proteins and molecules to phase separate into the dense phase. Concentration, temperature, pH, and salt concentrations are all critical factors which can affect the ability of molecules to form condensates. To simplify the system and to determine the effect of these parameters on condensates containing the protein of interest, experiments are often performed *in vitro* using recombinant protein purified from bacteria or insect cells, to remove the complicated environment of the nucleus and focus on the direct implications of the parameter being tested (Alberti et al. 2018; Venko and Žerovnik 2023; Kar et al. 2022).

These parameters and how they affect condensation are often mapped using phase diagrams, which is a graphical representation of the phase behaviour of a molecule (Figure 1.7) (Alberti and Dormann 2019). Phase diagrams map a two component system, which involves systematically changing two parameters, for example temperature and concentration, and monitoring whether a molecule is present in a dense phase or not. In changing these parameters it is possible to predict when phase separation becomes thermodynamically favourable and is denoted by the binodal line (Figure 1.7). These experiments determine what conditions are necessary for a given molecule to phase separate, and when performed on a protein of interest it can provide an insight as to whether a protein will phase separate under physiological conditions (Alberti and Dormann 2019; Ceballos, McDonald and Elbaum-Garfinkle 2018; Arzideh, Movagharnejad and Pirdashti 2018;

Arter et al. 2022).

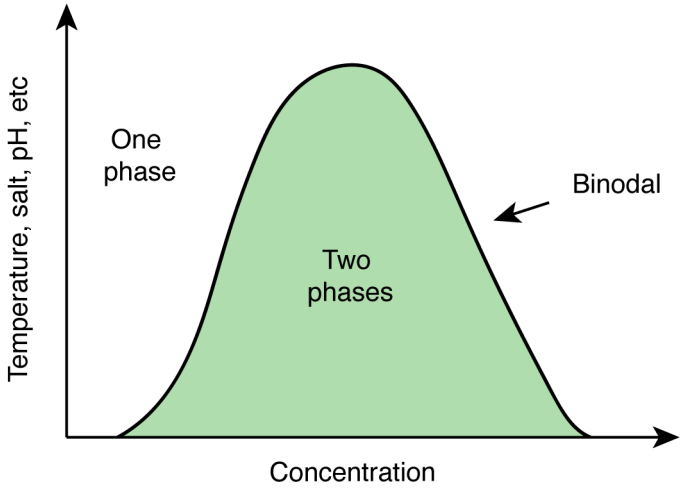

Figure 1.7: Mapping condensation through the use of a phase diagram
Phase diagrams are produced from changing two conditions for example temperature, pH, salt or concentration and monitoring the system to determine if the system is in a one phase state or a two phase state (green). The binodal seperates the two different states and is where the formation of two states become thermodynamically favourable. When a system is in a two phase state, both the dense and the dilute phase coexist. Image based on one found at Alberti and Dormann (2019).

1.2.3.1 Concentration

The most important factor that influences phase separation is the concentration of the components that cause condensate generation (Chong, Vernon and Forman-Kay 2018; Erdel and Rippe 2018; Lin et al. 2015; Tsang et al. 2019). It has been proposed that there is a threshold value, above which phase separation becomes energetically favourable, below which the formation of the dense phase wouldn't be beneficial (Erdel and Rippe 2018). Phase separation is therefore not possible in some cases where the expression of a protein is too low (Tsang et al. 2019). The effect of concentration is expected due to the supposed cause of phase separation being weak multivalent interactions (Ruthenburg et al. 2007); the more molecules involved, the more interactions there are, and the more likely condensate formation. It is therefore critical when studying phase separation to consider the concentration of your protein of interest. Many groups use overexpression to study

the phase separation of their protein of interest (Chong, Vernon and Forman-Kay 2018; Sabari et al. 2018), this dramatically increases the expression of the protein and care must therefore be taken when considering the implications of these findings on physiological proteins, when shown in isolation.

In vitro studies can be performed to determine the concentration of protein necessary for phase separation, which may indicate the physiological relevance of these findings. The concentration necessary for phase separation, also known as the concentration saturation (C_{sat}), is typically determined by centrifugation or via microscopy, using different concentrations of the protein of interest (Venko and Žerovnik 2023; Hallegger et al. 2021; Zhang et al. 2021; König et al. 2021; Alberti et al. 2018). In determining the concentration needed for phase separation, it is also possible to map phase diagrams testing other parameters including, temperature, salt, pH and ionic strength. Thus providing further information on the environmental requirements for phase separation.

1.2.3.2 Temperature and salt

Temperature is a parameter often tested within the context of phase separation. Increasing the temperature, increases the solubility of proteins, therefore at higher temperatures forming condensates becomes energetically unfavourable; similarly decreasing the temperature would increase condensation, as is seen in the literature (Kato et al. 2012; Murakami et al. 2015). Decreasing the temperature, by temperature shock has been shown to increase the ability of a protein to phase separate, as has been shown for DDX4 where condensates formed when cooled, dissolve when the temperature is increased (Nott et al. 2015; Riback et al. 2017; Chong, Vernon and Forman-Kay 2018; Lin, Currie and Rosen 2017). Salt concentration has also been shown to influence phase separation, where an increase in salt concentration has been shown to decrease phase separation of several proteins (Sabari et al. 2018; Tsang et al. 2019). Salt concentration is involved in the stabilisation of interactions that promote phase separation, therefore changing the salt concentration can influence these interactions and lead to phase separation becoming unfavourable (Garaizar and Espinosa 2021). In understanding the physiological relevance of phase separation it is important to consider whether a protein can phase separate at a cellular salt concentration.

1.2.3.3 pH and ionic strength

As for many biochemical reactions, condensate formation is also sensitive to changes in pH (Babinchak et al. 2019; Nott et al. 2015; Riback et al. 2017). Changes in both pH and ionic strength

regulate phase separation by affecting the interactions between proteins. The pH influences the charge state of amino acid residues, therefore facilitating or inhibiting the interactions that cause phase separation depending on the pH of the environment. In a similar way influencing the ionic strength affects the range and magnitude of interactions which will affect a protein's ability to phase separate; for example, increasing the ionic strength will increase the solubility and decrease a protein's ability to phase separate (Thompson et al. 2016). Due to the fluid nature of the environment in the nucleus, the effect of one process can be rescued by another. For example, decreasing the pH and increasing the salt concentration has been shown to allow phase separation of some proteins (Babinchak et al. 2019).

While phase separation has been implicated in multiple biological processes such as transcription, DNA repair, and mechanism of storage of RNAs, it is still largely unknown what effect this phenomena has on protein function.

1.3 The functional significance of condensates

Phase separation is thought to have numerous functions and take part in fundamental processes within the cell, with many common membraneless organelles displaying fluid-like properties. Condensates can function in multiple different ways including: as reaction crucibles where they sequester specific enzymes and subtrates to facilitate a particular biochemical reaction; molecular hubs at specific chromsomal loci typically to influence gene regulation, or as molecular sponges which act to sequester proteins which are not currently needed within the cell, acting to suppress function (Hirose et al. 2023).

Nuclear bodies are membraneless organelles within the nucleus which consist of multiple proteins and nucleic acids and are thought to be functionally relevant; examples of which include nuclear stress bodies, (nSBs) the nucleolus, cajal bodies and nuclear speckles (Decker et al. 2022; Smith, Hall and Lawrence 2020; Freibaum et al. 2020; Van Treeck et al. 2018; Riback et al. 2020; Shimobayashi et al. 2021). NSBs are formed in the nucleus under thermal stress conditions, and can act as reaction crucibles (Hirose et al. 2023; Ninomiya et al. 2020). In response to changes in temperature nSBs are formed to control splicing, through regulation of phosphorylation of serine/arginine-rich splicing factors (SRSFs). During the stress response cycle-like kinase 1 (CLK1) is recruited to nSBs to re-phosphorylate SRSFs and arrest splicing (Hirose et al. 2023; Ninomiya et al. 2020).

The oldest documented phase separated condensate is the nucleolus, a membraneless organelle in the nucleus which is the site of RNA synthesis and ribosome biogenesis (Ide et al. 2020; Pederson

2011; Mensah et al. 2023; Shin et al. 2018). The nucleolus itself is thought to consist of multiple phases, where the fibrillar component of the nucleolus is formed through phase separation, as it displays behaviour such as liquid droplet fusion (Ide et al. 2020; Brangwynne, Mitchison and Hyman 2011). Cajal bodies, are typically found associated with the nucleolus, and are functionally involved in RNA metabolism and formation of small nuclear ribonucleoproteins (snRNPs) (Rippe 2007; Love et al. 2017). These phase separated condensates, like many others, are thought to occur to bring enzymes and substrates together to increase the speed of individual biochemical processes (Rippe 2007; Love et al. 2017). Nuclear speckles or interchromatin granules, as they are otherwise known, were initially thought to act as a storage factory for splicing factors and RNA (as reviewed by Spector and Lamond (2011) and Galganski, Urbanek and Krzyzosiak (2017)). Recent studies have shown that the largest group of proteins found in nuclear speckles are transcription factors and chromatin remodelling proteins which facilitate regulation of gene expression, such as transcriptional hubs (as reviewed by Spector and Lamond (2011) and Galganski, Urbanek and Krzyzosiak (2017)).

In some cases, condensates are thought to act as molecular sponges or storage pools of proteins and nucleic acids, one of the most commonly documented examples of this are stress granules (Van Treeck et al. 2018; Boeynaems et al. 2019; Freibaum et al. 2020; Wang et al. 2020; Guenther et al. 2018; Molliex et al. 2015). Stress granules (SG) are formed in the cytoplasm during times of cellular stress, compartmentalising RNA and RNA binding proteins into stores, typically when translation is stopped until the stress subsides (Dewey et al. 2011; Gilks et al. 2004; Dang et al. 2006). TAR DNA-binding protein 43 (TDP-43) regulates RNA processing, and has been found within SGs, with interactions with long non coding RNAs (lncRNA) such as NEAT1 having been shown to promote condensation (Guenther et al. 2018; Wang et al. 2020; Dewey et al. 2011; Hallegger et al. 2021). TDP-43 is found in pathogenic inclusions in patients with ALS, a neurodegenerative disease; with a potential mechanism of pathology being suggested whereby misregulation of TDP-43 into SG's could lead to pathogenic protein inclusions (Guenther et al. 2018; Wang et al. 2020; Dewey et al. 2011; Hallegger et al. 2021). Proteins have also been shown to accumulate in their inactive form into condensates; proteins such as p300, may phase separate in a deactivated form, condensing protein that is not currently needed to activate transcription (Zhang et al. 2021).

Both functional and storage condensates have been documented which have specific functions; one such function is the organisation of chromatin within a cell (Conte et al. 2020; Stam, Tark-Dame and Fransz 2019; Gibson et al. 2019a; Gibson et al. 2019b; Bonev et al. 2017; Zhan et al. 2017; Dixon, Gorkin and Ren 2016; Shin et al. 2018).

1.4 DNA and Chromatin

The human genome is made up of molecules of deoxyribonucleic acid (DNA) that codes for over 20,000 proteins within the body (Crick 1970; Harrow et al. 2012). DNA is packaged within the nucleus, wrapped around a protein core; this core is an octamer which is comprised of two copies of four histone proteins: histone 3 (H3), histone 4 (H4), histone 2A (H2A) and histone 2B (H2B) (Venkatesh and Workman 2015; Kornberg 1974). These four histone proteins assemble to form the nucleosome, which together with DNA forms a structure called chromatin (reviewed by Wolffe and Guschin 2000, Stam, Tark-Dame and Fransz 2019). The majority of chromatin present within the cell is highly compact, termed heterochromatin, where little transcription occurs due to the lack of accessibility to the DNA (Figure 1.8) (reviewed by Morrison and Thakur 2021). In comparison, euchromatin is more open, with a greater distance between the nucleosomes to facilitate binding of transcriptional machinery and gene expression (Figure 1.8) (reviewed by Morrison and Thakur 2021; Venkatesh and Workman 2015).

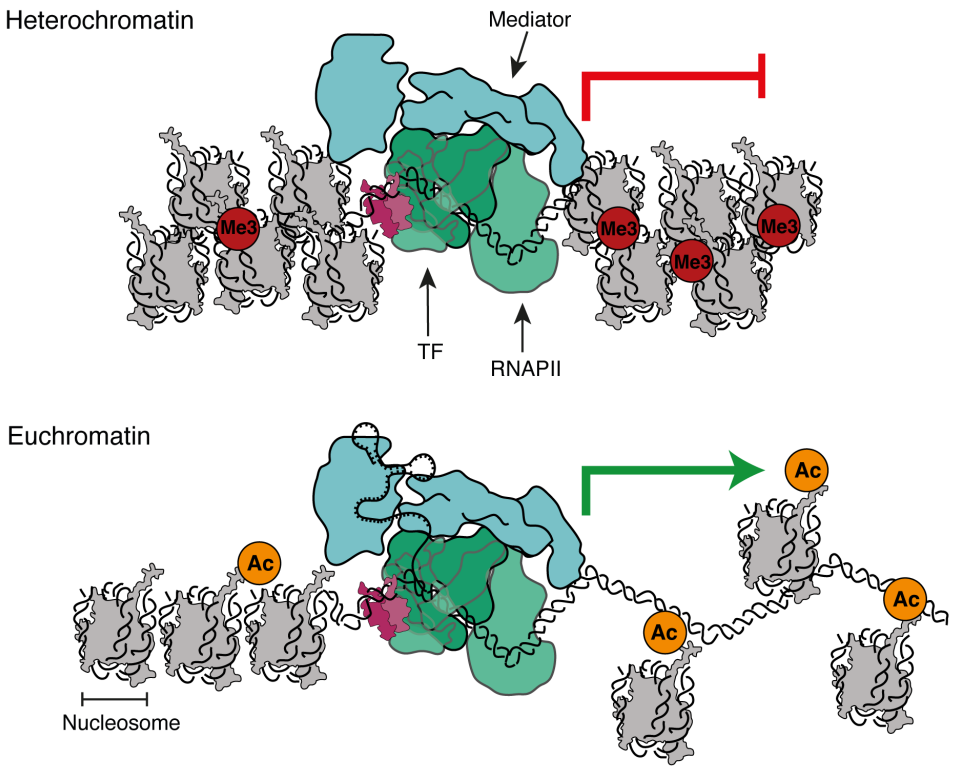

Figure 1.8: Characteristics of heterochromatin and euchromatin
Chromatin is made up of nucleosomes, which are tightly compact within heterochromatin, providing insufficient space for transcription via transcriptional regulators such as the mediator complex (blue), RNA polymerase II (RNAPII, light green) and transcription factors (TF, dark green). Heterochromatin is usually indicated by methylation marks (Me, red). In active chromatin, or euchromatin, the nucleosomes are spread further apart allowing binding of transcriptional complexes and transcription of the necessary genes. Euchromatin has modifications such as acetylation of lysine residues (Ac, orange), which is less compact due to the neutralisation of the charged amino acids by the acetylation. Image made by Dr Dan Bose and modified for this thesis.

Chromatin organisation into heterochromatin and euchromatin is extremely important for regulating gene expression, with moving between chromatin states being driven with PTMs of histones and DNA (reviewed by Luger 2003; Millán-Zambrano et al. 2022; Venkatesh and Workman 2015). These modifications that drive and maintain chromatin architecture have been very well characterised to define the gene regulatory landscape.

1.4.1 Chromatin modifications

Heterochromatin and euchromatin are characterised by different DNA and histone modifications. These regulatory marks typically involve, but are not limited to, methylation or acetylation of lysines or arginine residues, which can either activate or repress regions of chromatin based on their positions (as reviewed by Millán-Zambrano et al. 2022). Histone 3 lysine 9 tri-methylation (H3K9me3) and lysine 27 tri-methylation (H3K27me3) are both indicative of repressed heterochromatic regions, compared to histone 3 lysine 27 acetylation (H3K27ac) and lysine 122 acetylation (H3K122ac) which are representative of active euchromatic regions (Figure 1.8) (Bannister and Kouzarides 1996; Bose et al. 2017; Creyghton et al. 2010; Kalkhoven et al. 2003; Struhl 1998; Tie et al. 2016; Toth et al. 2013; Wang, Marshall and Ikura 2013; Millán-Zambrano et al. 2022; Venkatesh and Workman 2015; Jin et al. 2011). Proteins such as histone methyltransferases (HMT) and histone acetyltransferases (HAT) are involved in depositing these residues and activating or repressing chromatin. When HATs acetylate lysine residues present on chromatin, such as H3K27 and H3K122, this neutralises the positive charge of the lysine residues which destabilises nucleosome structure, creating space between the nucleosomes which allow binding of transcriptional proteins to DNA (Hebbes, Thorne and Crane-Robinson 1988; Turner 1993; Lee et al. 1993; Tropberger et al. 2013; Sungalee et al. 2021; Ogryzko et al. 1996; Bannister and Kouzarides 1996; Stasevich et al. 2014; Millán-Zambrano et al. 2022). Transcriptional complexes containing proteins such as RNA polymerase II (RNAPII), the Mediator complex and multiple transcription factors can therefore bind to the DNA sequences and have sufficient space to transcribe the DNA sequence into mRNA to initiate gene expression (as reviewed by Richter et al. (2022), Furlong and Levine (2018), Millán-Zambrano et al. (2022) and Whyte et al. (2013)).

1.4.2 Influence of phase separation on genome organisation

1.4.2.1 Maintenance of heterochromatin through phase separation

Chromatin organisation is functionally relevant in controlling regions which are transcriptionally active and silenced, the mechanism by which organisation is performed and maintained is still unclear. Evidence in recent years suggests a role of phase separation, where factors such as histone protein 1 (HP1) and BRD4 modulate condensation to change genomic organisation and facilitate function.

HP1 is associated with maintaining silenced heterochromatin, where the human homologue HP1α was shown to phase separate, with oligomerization occuring upon N-terminal phosphoryla-

tion which correlated with the spreading of heterochromatin (Larson et al. 2017). Work on HP1 homologues in *Drosophila melanogaster* and *Schizosaccharomyces pombe,* proposed a mechanism where HP1 molecules phase separate, which recruits further proteins to participate in weak multivalent interactions that drive phase separation (Sanulli et al. 2019; Strom et al. 2017). The accumulation of evidence from different species suggest a conserved mechanism of function for the HP1 protein. While in humans, HP1α has been suggested to cause chromatin organisation through phase separation, this mechanism is not conserved in HP1β and HP1γ which are also involved in genomic rearrangements.

Further evidence implicates phase separation in keeping heterochromatin in a transcriptionally silent state, that is sensitive to activating histone modifications. Acetylation of chromatin, transitioning heterochromatin into euchromatin, has been shown to cause the dissolution of condensates (Gibson et al. 2019b; Gibson et al. 2019a). After condensate dispersion the addition of BRD4, which binds acetylated lysine residues, induced the formation of a new phase separated state, which displayed distinct physical properties (Gibson et al. 2019b). Taken together, this evidence indicates a role for phase separation in maintaining chromatin organisation to facilitate genomic silencing.

1.4.2.2 Higher order genome organisation and phase separation

Chromatin is also organised in three-dimensional (3D) space into topological associated domains (TADs). Chromatin organisation has been extensively studied to understand how organisation is linked to gene regulation and function. Experimental techniques such as chromosome conformation capture (3C) technology (Hi-C) which involves chemical cross-linking and high-throughput DNA sequencing, has identified regions of chromatin that are in close physical proximity, that form loops of DNA known as TADs (Zhan et al. 2017; Szabo, Bantignies and Cavalli 2019; Bonev et al. 2017; Dixon, Gorkin and Ren 2016; Rosencrance et al. 2020). These regions are chromosome regulatory units, where genes that are co-regulated together during differentiation are located within the same TAD; enhancers and their gene promoters are restricted to the same TAD, suggesting a functional role in chromatin architecture (Figure 1.9) (Bonev et al. 2017; Zhan et al. 2017; Dixon, Gorkin and Ren 2016; Nora et al. 2012; Wang et al. 2021). TADs are constrained by cohesin rings, that cause loop extrusion, and these rings are shown to accumulate at sites of CCCTC-binding factor (CTCF) binding sites (Figure 1.9).

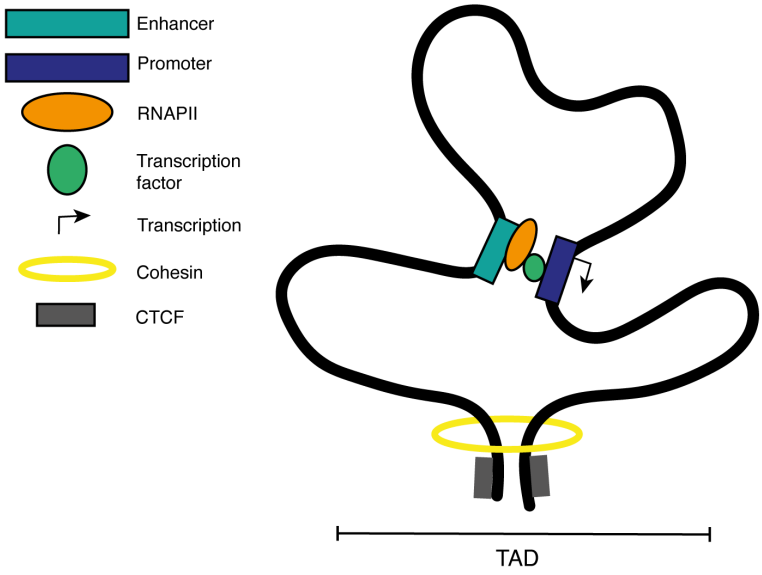

Figure 1.9: Chromatin looping within a topological associated domain (TAD)
Chromatin organisation into a TAD is characterised by cohesin looping and binding of structural protein CTCF. TADs represent boundaries where enhancers can bind to their target promoter to initiate gene expression.

Chromatin has been shown to directly interact with nuclear condensates, and due to the current interest in research of the role of phase separation in biological function, it is thought that it is a potential mechanism of chromatin organisation (Conte et al. 2020; Stam, Tark-Dame and Fransz 2019; Shin et al. 2018). At the level of TADs, OCT4 a master transcription factor for somatic reprogramming has been shown to form phase separated condensates, which when disrupted cause the reorganisation of TADs. This correlated to the redistribution of super-enhancers, altering their target genes (Wang et al. 2021). Chromatin organisation has been shown to be regulated through the process of phase separation via structural protein CTCF, which is enriched within TAD boundaries (Wei et al. 2022). Disrupting phase separation on a cellular level was shown to affect both TAD organisation and larger scale chromatin organisation, further providing evidence of the importance of this phenomena for chromatin organisation (Liu et al. 2021).

Investigations into chromosomal rearrangement have suggested a role for proteins that phase separate, incorporating chromatin into condensates to achieve genomic accessibility (Cai et al. 2019; Gibson et al. 2019b; Gibson et al. 2019a). Yes associated protein (YAP), a transcriptional

coactivator which forms condensates, has been shown to associate with accessible chromatin in the nucleus, where these condensates contain RNAPII suggesting a mechanism where YAP causes chromosomal rearrangements to facilitate active transcription in the nucleus (Cai et al. 2019).

Chromatin has been shown to be both included and excluded from phase separated condensates. One study which focussed on intrinsic disordered regions (IDRs), showed that condensates containing these IDRs were accumulating on the periphery of chromatin (Shin et al. 2018). This paper proposed a "chromatin filter" theory where regions of the genome are pulled together in the condensate and others excluded, therefore allowing the condensate to regulate genome organisation (Shin et al. 2018).

Phase separated condensates are therefore extremely important for genome organisation and function within the nucleus. Alongside genome organisation, phase separation has also been implicated in driving gene expression through transcriptional condensates, which is a proposed mechanism by which enhancers donate transcriptional machinery to the promoter to increase transcription of their target genes (as reviewed by Demmerle, Hao and Cai (2023)).

1.5 Enhancers

1.5.1 Introduction to enhancers and superenhancers

Enhancers are *cis*-regulatory elements which are typically 100 - 1000 bp in length, and increase the transcription of their target genes through interaction with their target promoter sequence (Figure 1.9) (Furlong and Levine 2018; Long, Prescott and Wysocka 2016; Claringbould and Zaugg 2021). Enhancers act independently of orientation and distance to the promoter, where the distance between enhancer and promoter can span over great distances (Furlong and Levine 2018; Long, Prescott and Wysocka 2016; Claringbould and Zaugg 2021). Despite the distance, enhancers and promoters reside within the same TAD (Figure 1.9), and their function can be spatiotemporal and cell-type specific, having important roles in development (Bulger and Groudine 1999; Smith and Shilatifard 2014; Furlong and Levine 2018; Long, Prescott and Wysocka 2016; Visel et al. 2009; Zhan et al. 2017; Dixon, Gorkin and Ren 2016). The first enhancer to be discovered was acting distally in *cis* to increase the transcription of the rabbit haemoglobin β-1 gene, since then over 400,000 enhancers have been discovered which influence gene transcription (Banerji, Rusconi and Schaffner 1981; ENCODE Project Consortium 2012). As information has increased for the functional role of enhancer sequences, so has evidence for their role in disease. Misregulation at enhancer sequences leads to numerous diseases, termed enhanceropathies; one example being

the overexpression of MYC, caused by translocation of the IgH enhancer, leading to Burkitt's lymphoma (Taub et al. 1982; Claringbould and Zaugg 2021; Smith and Shilatifard 2014). It is therefore extremely important to understand how these regulatory elements function to prevent these disease states.

Enhancers are found within euchromatin regions of the genome, and are typically defined by hypersensitivity to DNase I treatment, due to the open conformation of the chromatin (Gurumurthy et al. 2019; Felsenfeld and Groudine 2003). Active enhancers are also associated with H3K4me and H3K27ac, and act as binding hubs for a diverse variety of transcription factors and cofactors (Figure 1.10) (Plank and Dean 2014; Heinz et al. 2015; Claringbould and Zaugg 2021; Kim et al. 2010; Wang, Marshall and Ikura 2013; Creyghton et al. 2010). One such cofactor is CREB binding protein (CBP), which is a transcriptional coactivator protein that deposits acetyl residues onto chromatin leaving it in an open conformation for transcription to occur (Creyghton et al. 2010; Bannister and Kouzarides 1996; Chan and La Thangue 2001). RNA polymerase II (RNAPII) is also found at enhancer regions, which bi-directionally transcribes the DNA sequence into enhancer RNA (eRNA) molecules which have a functional role in gene transcription (Figure 1.10) (Blinka et al. 2017; Harrison and Bose 2022; Lee et al. 2021b; Sartorelli and Lauberth 2020; Mousavi et al. 2013). eRNA molecules have been shown to aid in enhancer function through multiple mechanisms, including: activating the histone acetyltransferase activity of CBP; binding transcription factors bringing them to the enhancer region, and have been suggested to act as a scaffold in the phase separation model of enhancer function, which will be discussed in the following section (Harrison and Bose 2022; Bose et al. 2017; Gurumurthy et al. 2019; Gurumurthy et al. 2021).

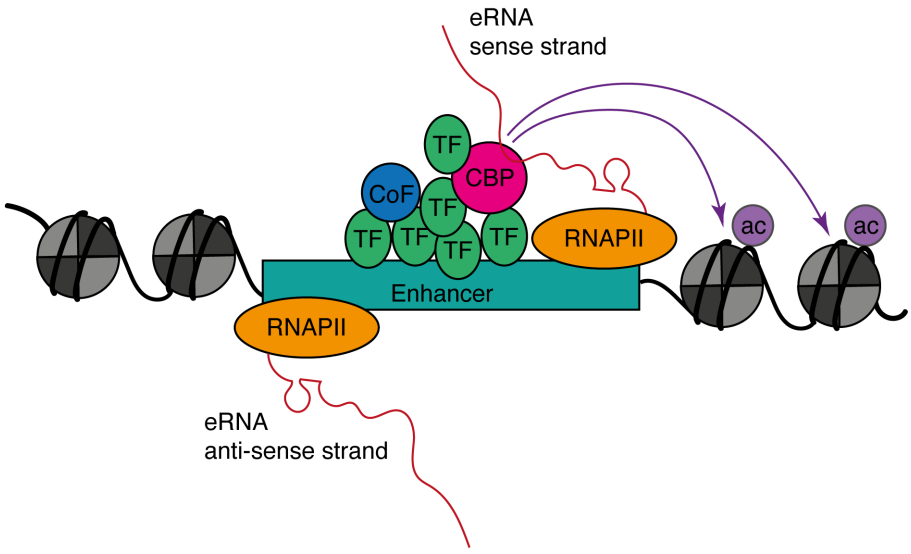

Figure 1.10: Enhancers act as transcription factor binding hubs
Enhancers (turquoise), which is present in the euchromatic regions in between nucleosomes (black and grey circle), bind multiple transcription factors (TFs, green) and cofactors (CoF, blue) including CREB binding protein (CBP, pink). At some enhancers, RNA polymerase II (RNAPII, orange) is recruited, which bidirectionally transcribes both sense and antisense enhancer RNAs (eRNAs, red). The diagram also shows two key features of CBP including its histone acetyltransferase ability, adding acetyl groups (ac, purple) to histones (purple arrows) and it's ability to bind to eRNAs. Images based on Li, Notani and Rosenfeld (2016) and Bose et al. (2017).

Beyond enhancers, there are a class of regulatory elements called superenhancers (SE) which can be clusters of enhancers, enhancers which are large in size or enhancers which are a large distance away from their target promoter, spanning over 10 Kb (Figure 1.11) (Brown et al. 2014; Hnisz et al. 2013; Whyte et al. 2013; Tang et al. 2022). SEs are typically characterised by binding of master transcription factors such as BRD4 and the mediator complex (Crump et al. 2021; Tang et al. 2022; Sabari et al. 2018; Whyte et al. 2013). 3D genome organisation, as has been previously discussed, facilitates SE interaction with their promoters, facilitated by chromatin remodelling proteins such as CBP (Bonev et al. 2017; Zhan et al. 2017; Dixon, Gorkin and Ren 2016; Nora et al. 2012; Wang et al. 2021; Kim et al. 2010; Whyte et al. 2013).

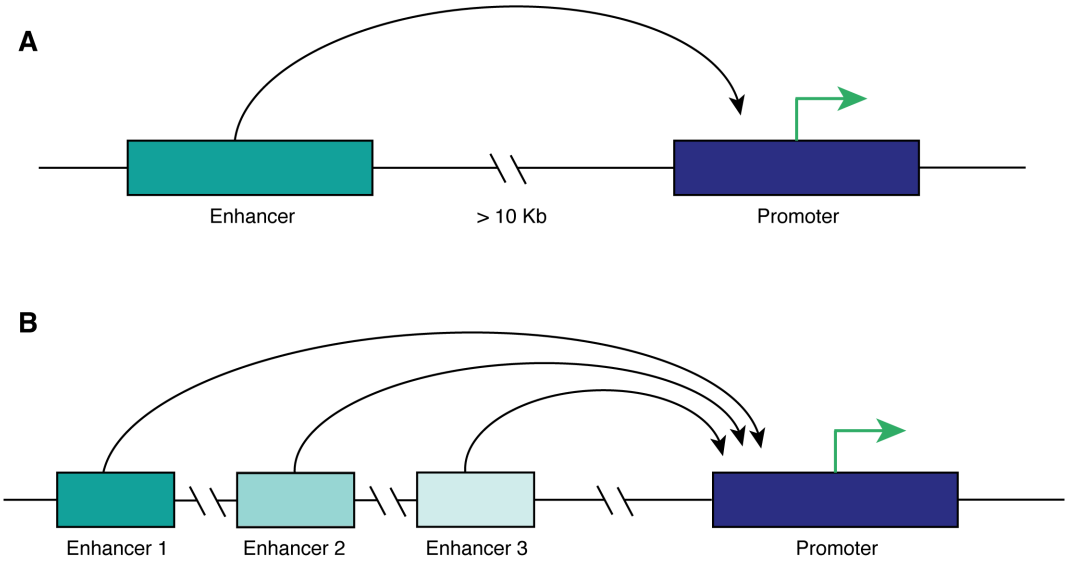

Figure 1.11: Example types of superenhancer
A - Superenhancers (turquoise) can increase gene expression (green arrow) by interaction with promoter regions (navy) of their target genes up to 10 Kb away from there target promoter. B - Another type of superenhancer involves clusters of multiple enhancers (shown here as 3 opacities of turquoise) which all contribute to amplifying gene expression (green arrow) of their target gene by interaction with the target promoter (navy).

SEs also play a role in disease, one example is of a superenhancer which is implicated in T-cell acute lymphoblastic leukemia (T-ALL), which leads to cancer of T-cell progenitors (Smith et al. 2023; Mansour et al. 2014). One example of this can be found in the Jurkat cell line, used to study T-ALL, which contains a 12 base pair insertion that generates two binding sites for the transcription factor MYB, which creates a *de-novo* superenhancer which upregulates production of T-cell acute lymphocytic leukemia protein 1 (TAL1), a key protein involved in T-ALL (Smith et al. 2023; Mansour et al. 2014).

While the importance and implications of enhancer and SEs functions are known, the mechanism with which they act is still under investigation.

1.5.2 Transcription regulation

It is thought that enhancers act as transcription factor binding hubs, to accumulate transcriptional machinery to donate to the promoter region and initiate transcription (as reviewed by Spitz and Furlong (2012) and Popay and Dixon (2022)). There are multiple proposed models as to how enhancers may reach their distal promoters to increase gene expression, these include: enhancer-promoter looping, the linking model and donation through formation of a phase separated condensate (as reviewed by Popay and Dixon (2022), Spitz and Furlong (2012), Claringbould and Zaugg (2021) and Demmerle, Hao and Cai (2023)).

1.5.2.1 Enhancer-promoter looping

One of the earliest proposed mechanisms as to how enhancers initiate gene transcription is through enhancer-promoter looping (Rao et al. 2017; Nora et al. 2017; Popay and Dixon 2022; Schwarzer et al. 2017). In this model, the enhancer is brought into close proximity of the promoter to transfer the transcriptional machinery needed to activate gene expression (Rao et al. 2017; Nora et al. 2017; Popay and Dixon 2022; Schwarzer et al. 2017). This model is facilitated by 3D genome organisation, where the formation of TADs through the cohesin complex, generates chromosomal loops that can bring the enhancer and promoter into close physical space (Figure 1.9) (Schwarzer et al. 2017; Rao et al. 2017; Zhan et al. 2017; Wang et al. 2021; Szabo, Bantignies and Cavalli 2019). A protein bridge is then formed between the enhancer and promoter by the transcriptional machinery which excludes a loop of DNA facilitating gene expression (Figure 1.12) (as reviewed by Popay and Dixon (2022), Spitz and Furlong (2012), Bulger and Groudine (1999) and Claringbould and Zaugg (2021)). This model was termed enhancer-promoter looping.

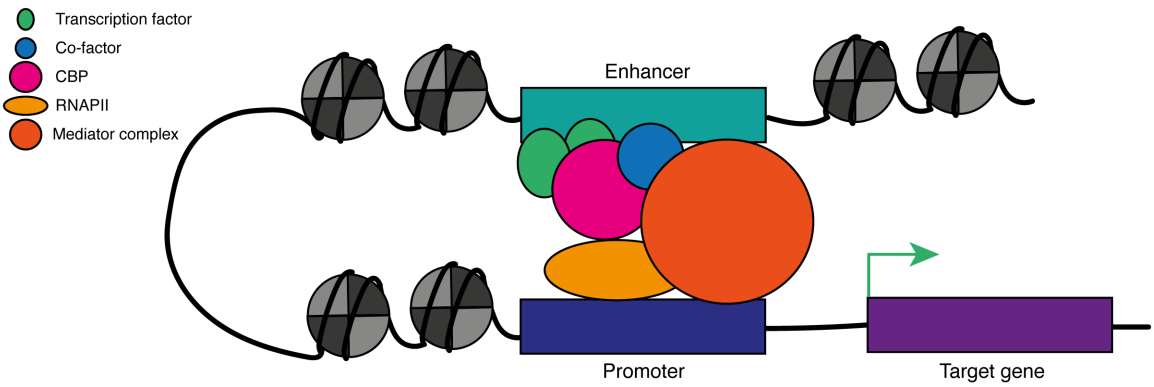

Figure 1.12: Enhancer-promoter looping as a mechanism for enhancer function
Enhancers act as transcription factor binding hubs, binding transcription factors and cofactors such as CBP and RNAPII. Due to confinement in 3D space within a TAD, enhancers and promoters are within physical proximity, a protein bridge can form to donate this transcriptional machinery which activates transcrition (green arrow) of the target gene. Based of an image from Claringbould and Zaugg (2021).

While this model of enhancer function is generally accepted, several recent studies suggest that direct contact through enhancer-promoter looping is not necessary for gene expression. It was shown that enhancer promoter looping was not occurring in the case of Sox2 in embryonic stem cells (ESCs), as transcription of Sox2 was not associated with proximity to its target enhancer, Sox2 control region (SCR) (Alexander et al. 2019). In this model, TADs play a clear role in bringing the enhancers and promoters into close physical proximity, where CTCF has been shown to be necessary to maintain TAD structure. Upon CTCF depletion no change in active and inactive genes was observed, suggesting chromatin looping through TADs is not necessary for gene activation (Nora et al. 2017). While there is evidence to suggest enhancer looping may play a role in enhancer function, evidence to the contrary implies that alternative mechanisms of enhancer function should be considered, such as enhancer linking.

1.5.2.2 Enhancer promoter connection through linking

Enhancer linking is an alternative model of enhancer function, where a protein scaffold is formed between the enhancer and the promoter along the DNA (as reviewed by Bulger and Groudine (1999) and Popay and Dixon (2022)). This was initially proposed based on the Chip protein from *Drosophila melanogaster* which has roles in enhancer function, and has been suggested as

a facilitator in communications between enhancers and promoters (Dorsett 1999; Morcillo et al. 1997). Homologues of Chip are found in humans, and have been show to bind to homeodomain DNA binding motif containing proteins (Dorsett 1999; Agulnick et al. 1996; Morcillo et al. 1997). A model was therefore proposed where Chip-related molecules bound to homeodomain proteins which anchored to chromatin, potentially via acetylated histones, to create a chain of proteins that links the enhancer to promoter and initiate gene expression (Figure 1.13)(as reviewed by Bulger and Groudine (1999) and Popay and Dixon (2022)).

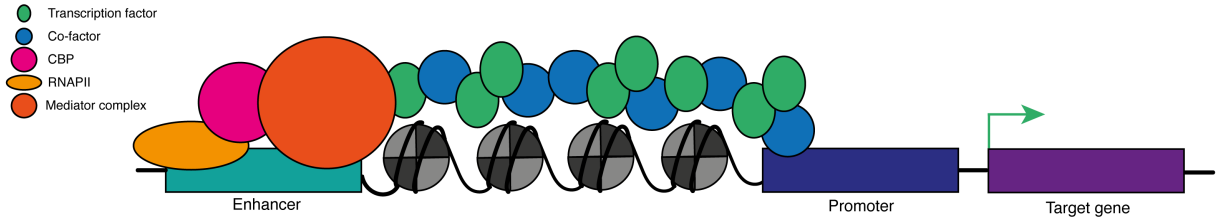

Figure 1.13: Enhancer promoter function through linking
Alternative mechanism by which a chain of proteins form, linking the enhancer and promoter of a gene of interest. The gene expression (green arrow) is then activated by signals carried along by the physical bridge. The protein bridge is shown here as containing transcription factors and cofactors for simplicity. This bridge is thought to occur between Chip-related proteins binding homodomain containing proteins which associate with DNA and potentially histone modifications. Image based on Popay and Dixon (2022).

In this model, enhancer dependent gene activation occurs in the absence of close proximal distance of the promoter and enhancer, therefore the factor of distance is much less an object than in the enhancer looping model. However, it is not yet clear how promoters would terminate the chain of Chip-related molecules to prevent extension to distal genes (as reviewed by Bulger and Groudine (1999) and Popay and Dixon (2022)). Therefore other models of enhancer function may be more feasible. More recently, a proposed mechanism of action is by phase separation.

1.5.2.3 Phase separation as a mechanism of enhancer function

Due to the high concentration of proteins and eRNA molecules present at the enhancers, phase separation has been proposed as a way in which enhancers donate transcriptional complexes to the promoter region (Demmerle, Hao and Cai 2023; Sabari, Dall'Agnese and Young 2020; Sabari et al. 2018; Plys and Kingston 2018; Banani et al. 2016; Gurumurthy et al. 2019; Gurumurthy et al. 2021; Shrinivas et al. 2019; Wei et al. 2020; Zhang et al. 2021; Ma et al. 2021). Several proteins

commonly found at enhancers regions such as BRD4, YAP, MED1, CBP and p300 have IDRs and have been shown to phase separate (Sabari et al. 2018; Zhang et al. 2021; Ma et al. 2021; Cai et al. 2019; Han et al. 2020; Wang et al. 2022b). RNAPII is commonly located at enhancer regions, and has been shown to localise to transcriptional condensates with factors such as BRD4 and MED1 (Sabari et al. 2018; Kim et al. 2010; De Santa et al. 2010). eRNA molecules transcribed by RNAPII, have been suggested to play a role in facilitating the formation of transcriptional condensates (as reviewed by Harrison and Bose (2022)). eRNAs are thought to act as a scaffold, and have been suggested to have a regulatory role in condensate formation where RNA molecules contribute to weak multivalent interactions between proteins, bringing them into proximity to form a condensate (Decker et al. 2022; Henninger et al. 2020; Shrinivas et al. 2019; Banani et al. 2016). m6A methylation of eRNAs has also been shown to be recruited to enhancers to stimulate gene activation and formation of transcriptional condensates (Figure 1.14) (Lee et al. 2021b).

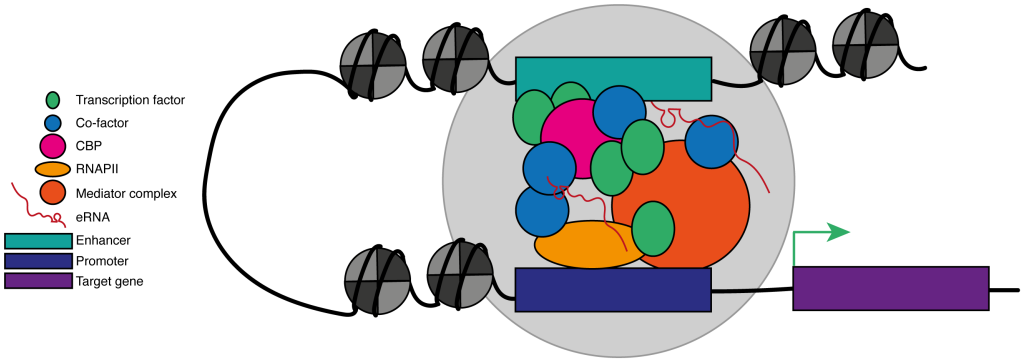

Figure 1.14: Enhancer-promoter interaction can occur through phase separation
Proteins (CBP, transcription factors, RNAPII and the mediator complex) and nucleic acids (eRNAs) accumulate at enhancer regions and form phase separated condensates (grey circle) between intrinsically disordered regions of the proteins, and the nucleic acids. The transcriptional machinery is then donated to promoter through phase separation to activate gene expression (green arrow).

Taken together this suggests a model where proteins and nucleic acids accumulate at the enhancer region via phase separation, and are then transferred to the promoter through close proximity (Demmerle, Hao and Cai 2023; Sabari, Dall'Agnese and Young 2020; Sabari et al. 2018; Plys and Kingston 2018; Banani et al. 2016; Gurumurthy et al. 2019; Gurumurthy et al. 2021; Shrinivas et al. 2019; Wei et al. 2020; Zhang et al. 2021; Ma et al. 2021). Condensates

have also been proposed as a mechanism of superenhancer function, for collaborative donation of proteins from different enhancers to the promoter region, and from enhancers which are far away in the genome (Gurumurthy et al. 2019; Gurumurthy et al. 2021; Sabari et al. 2018; Tang et al. 2022). Phase separation of superenhancers, made up of multiple different enhancers, are suggested to function through compartmentalising transcriptional machinery at cell-type specific genes via condensates, as a robust mechanism for transcriptional control (Figure 1.15) (Sabari et al. 2018).

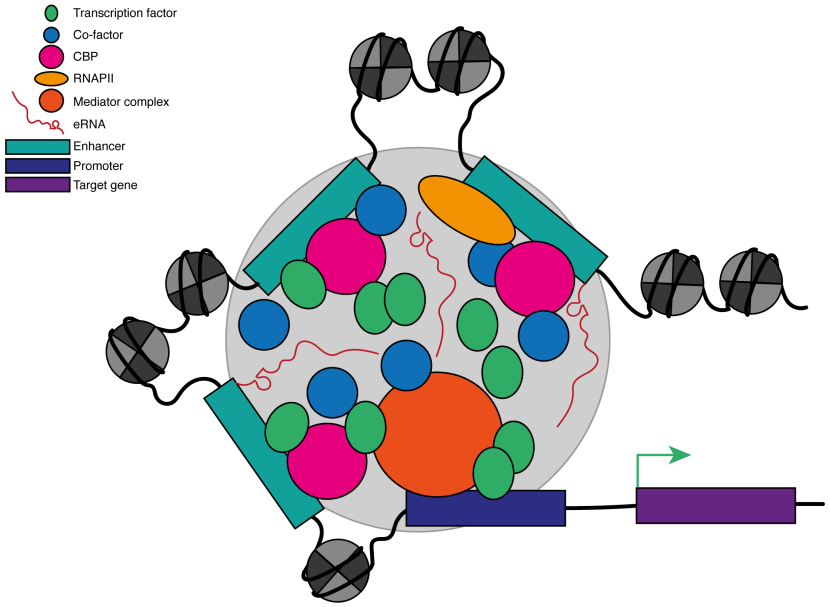

Figure 1.15: Phase separation as a mechanism for superenhancer function
Multiple enhancers, or superenhancers, can increase the gene expression (green arrow) of their target gene by forming a phase separated condensate (grey circle) to donate transcriptional machinery including CBP, transcription factors, RNAPII and the mediator complex to the promoter region. eRNAs produced from enhancer regions facilitate phase separation which is a proposed mechanism of action. Image based on Sabari et al. (2018).

Building on the model of transcriptional condensates at superenhancers, a mechanism of gene activation has been suggested which involves the transfer of elongation competent transcription complexes containing RNAPII (Gurumurthy et al. 2019; Gurumurthy et al. 2021; Gelder and Bose 2023).

RNAPII transfer model The RNAPII transfer model suggests a model for enhancer-dependent gene expression via phase separation, relying on the transient transcription of enhancers to release elongation competent, dephosphorylated transcriptional complexes, that are recruited to the promoter through phase separation (Gurumurthy et al. 2019; Gurumurthy et al. 2021; Shao et al. 2022). In this model, the Integrator complex, which has protein phosphatase and endonuclease activities, terminates transient transcription of eRNAs, which prevents RNAPII progressing into elongation (Gurumurthy et al. 2019; Stein et al. 2022). Evidence for this model was proposed when *INTS11*, a component of the integrator complex, was depleted in murin erythroleukemia (MEL) cells which showed an increase in RNAPII at locus control regions (LCR) of β-globin, and decreased RNAPII recruitment to the promoter region; where LCRs are a type of superenhancer found in transgenic mice (Gurumurthy et al. 2021). Depletion of *INTS11* in mouse ESCs showed an accumulation of unphosphorylated RNAPII at promoters with a build up of short RNA transcripts, supporting the RNAPII transfer model (Stein et al. 2022). eRNAs are thought to play a distinct role in β-globin gene expression, staying present at the site of transcription, and upon depletion a reduction in gene expression of β-globin was observed (Gurumurthy et al. 2021). RNA molecules have previously been proposed to aid in promoter function, by binding at sites of transcription initiation and recruiting RNAPII, producing short RNA molecules which are negatively charged that facilitate condensate formation (Shao et al. 2022).

It is therefore feasible that phase separated condensates form at enhancers, via eRNAs which are prematurely terminated by the Integrator complex, forming elongation competent dephosphorylated RNAPII, which is transferred to the promoters of target genes in a primed position to initiate gene expression (Gurumurthy et al. 2019; Gurumurthy et al. 2021).

Perspectives based on phase separation model of transcription regulation A plethora of evidence has been published in support of transcriptional condensates, and phase separation as a mechanism of enhancer function (Demmerle, Hao and Cai 2023; Sabari, Dall'Agnese and Young 2020; Sabari et al. 2018; Plys and Kingston 2018; Banani et al. 2016; Gurumurthy et al. 2019; Gurumurthy et al. 2021; Shrinivas et al. 2019; Wei et al. 2020; Zhang et al. 2021; Ma et al. 2021). Many proteins found at enhancers have been shown to phase separate, with this behaviour shown to effect protein function (Sabari et al. 2018; Zhang et al. 2021; Ma et al. 2021; Cai et al. 2019; Han et al. 2020; Wang et al. 2022b). One protein ubiquitously found at enhancers which has been shown to phase separate is CBP, however the driving forces behind this behaviour and the effect on function is not clear.

1.6 CBP

1.6.1 CBP is a transcriptional coactivator protein

CBP is a coactivator protein which is found at enhancers due to its ability to bind transcription factors, eRNAs and its histone acetyltransferase activity (Bannister and Kouzarides 1996; Chan and La Thangue 2001; Bose et al. 2017; Ma et al. 2021; Narita et al. 2021; Wiese et al. 2020). CBP gets its name from binding transcription factor cAMP-response-element binding protein (CREB), an interaction shown to activate transcription (Cardinaux et al. 2000). CBP contains multiple globular domains which are important for transcription factor binding, including: transcriptional adapter zinc binding motif (TAZ1) and (TAZ2); CREB binding domain (KIX); bromodomain (BRD); zinc binding domain (ZZ) and nuclear receptor coactivator binding domain (NCBD) (Figure 1.16 A) ((Dyson and Wright 2016). CBP also has a histone acetyltransferase (HAT) domain which is highly structured and is responsible for CBPs catalytic activity; it deposits acetyl groups onto transcription factors such as p53 and PolII, and histones, facilitating the opening of chromatin into euchromatin to activate transcription (Bannister and Kouzarides 1996; Thompson et al. 2004; Thompson et al. 2001; Bose et al. 2017; Tie et al. 2009; Schröder et al. 2013; Barlev et al. 2001; Ortega et al. 2018). Acetylation of RNAPII by CBP was also shown to increase release of RNAPII from enhancers and activate transcription (Schröder et al. 2013; Stasevich et al. 2014). While CBP contains many structured, functional domains, it is also over 60 % disordered; when using disorder prediction software DISOPRED3, these disordered regions are localised to both the N- and C-terminus of the protein (Figure 1.16 B) (Dyson and Wright 2016; Keppel, Howard and Weis 2011). These regions of disorder have no defined secondary structure, which afford flexibility and provide easy access to over 400 proteins which bind CBP (Bedford et al. 2010; Oldfield and Dunker 2014).

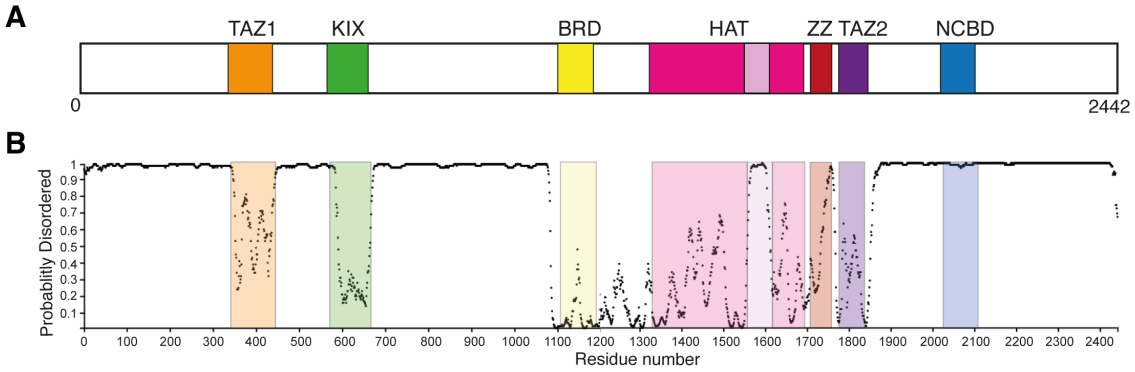

Figure 1.16: Disorder prediction map of CBP

Intrinsic disorder shown over the full length of CBP. A - domain diagram of CBP showing globular structured domains including: transcriptional adapter zinc binding motif (TAZ1, orange) and (TAZ2, purple); CREB binding domain (KIX, green); bromodomain (BRD, yellow); histone acetyltransferase domain (HAT, pink); zinc binding domain (ZZ, red) and nuclear receptor coactivator binding domain (NCBD, blue). Image based on one found at Dyson and Wright (2016). B - Graph showing probability of disorder for the amino acid residues of CBP, where a probability of 1 indicates unstructured disordered regions. Shaded colour areas represent approximate domain positions, coloured as described above. Plot B was generated using DISOPRED3 (Jones and Cozzetto 2015).

1.6.2 CBP and close paralog p300

1.6.2.1 Similarities and differences of CBP and p300

CBP and paralog p300 are both histone acetyltransferases and their names are often used interchangeably in the literature. Both have similar functions, with occupancy of both being found at enhancer regions, which is often used as a marker for active enhancers in genome-wide studies (Creyghton et al. 2010; May et al. 2011). CBP and p300 display similar properties and have some level of redundancy; both proteins share approximately 90 % sequence homology within the catalytic HAT domain, with a reduced similarity in the IDRs (Henry, Kuo and Andrews 2013; Chan and La Thangue 2001). For a full sequence alignment of CBP and p300 see appendix section 9.4. However, when both proteins where knocked-out independently in a mouse model system, the removal of both proteins was embryonic lethal (Yao et al. 1998; Kawasaki et al. 1998). It was also observed in mice, that a full complement of CBP alleles is needed for normal differentiation of hematopoietic stem cells into cells found in blood, where this wasn't the case for p300 (Kung et al.

2000). This suggests individual functions for each of these proteins, which are not compensated for by the other. It is therefore important to study both CBP and p300 separately, even though studies in one protein can be used for comparison in the other.

1.6.2.2 Phase separation of p300 and implications for CBP

Both p300 and CBP are key features of enhancers, increasing enhancer function through deposition of enhancer marks such as H3K27ac and binding transcription factors (Creyghton et al. 2010; Tie et al. 2014; May et al. 2011; Bannister and Kouzarides 1996; Chan and La Thangue 2001; Jin et al. 2011). As previously described, one mechanism of enhancer function is through phase separation. Many proteins commonly found at enhancer regions, such as RNA binding proteins and transcription factors have disordered regions which facilitate promiscuous binding to many proteins; CBP and p300 have similar patterns of disorder, with both proteins being over 60 % disordered (Dyson and Wright 2016; Chong and Forman-Kay 2016; Lin, Currie and Rosen 2017; Peng and Kurgan 2015; Plys and Kingston 2018; Protter et al. 2018; Rani, Baruah and Biswas 2014; Pak et al. 2016). Both full length CBP and p300, in the last couple of years, have been shown to phase separate, with the identification of nuclear puncta in cellular systems (Ma et al. 2021; Zhang et al. 2021; Wang et al. 2022b).

CBP itself has been described to colocalise to promyelocytic leukemia nuclear bodies (PML) which are phase separated nuclear puncta which have been suggested to have roles in transcriptional regulation (LaMorte et al. 1998; Banani et al. 2016; Boisvert et al. 2001). CBP has also been shown to be recruited to Huntingtin aggregates, by interactions of the expanded polyglutamine tract of huntingtin, binding to a small stretch of glutamines in the C-terminal region of CBP. This is thought to have toxic affects by removing CBP from sites where transcription is critical (Nucifora et al. 2001).

Specific focus on p300 has led to further investigation of these puncta to distinguish their liquid-like properties and investigate whether specific regions of intrinsic disorder within p300 are important for its phase behaviour. Using a fluorescent knock-in mouse embryonic stem cell line (mESCs), full length p300 condensates were observed that were dynamic in movement, and colocalise with CBP phase separated condensates (Ma et al. 2021). Formation of condensates containing the transactivating domains of transcription factors p65 and Sox2 showed rapid colocalisation with p300 condensates (Ma et al. 2021).

CBP and p300 are mainly conserved through their functional domains, having a greater degree of difference within the disordered regions (Chan and La Thangue 2001; Eckner et al. 1994).

Studies on p300 have investigated intrinsic disorder regions within p300 using the optoDroplet system (technique described in section 1.2.2.1) showing phase separation of the N- and C- terminal regions, with no clusters forming from IDRs closer to the centre of the protein (Ma et al. 2021). Mutations in the C-terminal region of p300, where the glutamine residues were changed to alanines (Q-to-A) showed an inability to phase separate, suggesting the importance of hydrophobic interactions in p300 phase behaviour (Ma et al. 2021). All disordered regions in p300 were examined to see if they form phase separated condensates *in vitro*; results suggest a region upstream of the HAT domain, and in the C-terminal region as being able to form condensates independently, and upon deleting both regions from full length p300, the resulting construct was not able to form condensates in cells (Wang et al. 2022b). Together these studies have shown that the IDRs of p300 are important for influencing the full length proteins phase behaviour. This data can be extrapolated to suggest that intrinsic disorder within CBP may influence its phase separating behaviour.

It is yet to be understood which regions of CBP are important for regulating its phase separating ability and how this effects its catalytic function.

1.6.3 CBP histone acetyltransferase activity

1.6.3.1 Regulating CBP HAT activity

CBP's HAT activity is important for regulating gene expression in the cell, therefore understanding how this is regulated is extremely important. The HAT domain of CBP has an active site for substrate binding. In an inactive state, a region known as the autoinhibitory loop (AIL), or the activation loop, sits within the active site to prevent substrate binding (Thompson et al. 2004; Ortega et al. 2018). Upon activation, the loop is displaced from the active site, leaving the active site free to bind substrates, activating CBPs catalytic activity (Figure 1.17) (Thompson et al. 2004; Ortega et al. 2018). The AIL is therefore extremely important in regulating CBP HAT activity.

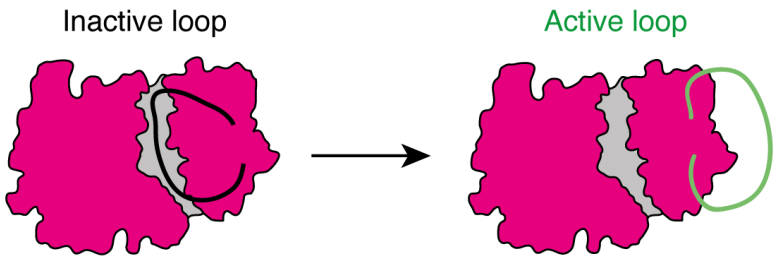

Figure 1.17: Displacement of the AIL activates CBP HAT activity
In an inactive form the AIL loop (black) sits within the active site (grey) of CBPs HAT domain (pink). Upon activation, the loop is displaced (green) from the active site leaving the active site available to bind positively charged substrates to acetylate. CBP HAT domain diagram made by Dr Dan Bose and adapted here, diagram based on the HAT domain structure published in Wang et al. (2008a).

AIL is important for regulating the catalytic activity of CBP The AIL region is positively charged and contains many lysine residues. In an inactive, autoinhibited state, the AIL binds to the negatively charged active site of CBP, and prevents binding of positively charged lysine-rich substrates. Upon activation, lysine residues within the AIL become acetylated which neutralises the positive charge of the lysines residues, the loop is then displaced from the active site, leaving CBP in an active conformation where positively charged substrates can bind (Figure 1.18) (Thompson et al. 2004; Ortega et al. 2018). In CBP, and p300, HAT activity is thought to be activated by autoacetylation from within the same molecule, however this has also been shown to occur *in trans* (Thompson et al. 2004; Ortega et al. 2018). Dimerisation of transcription factors which bind p300 have been shown to enable trans-autoacetylation of p300, where a crystal structure shows the AIL of one p300 molecule invade the active site of another, stimulating HAT activity (Ortega et al. 2018).

Another mechanism of HAT activation is through eRNA binding through the RNA binding region within the AIL, which displaces the loop and activates CBP's HAT activity (Figure 1.18) (Bose et al. 2017; Carullo et al. 2020). This was shown to have a similar catalytic activity to that of the acetylated CBP, in a mechanism that is independent of recruitment to chromatin; further allowing regulation of gene expression through acetylation of chromatin (Bose et al. 2017).

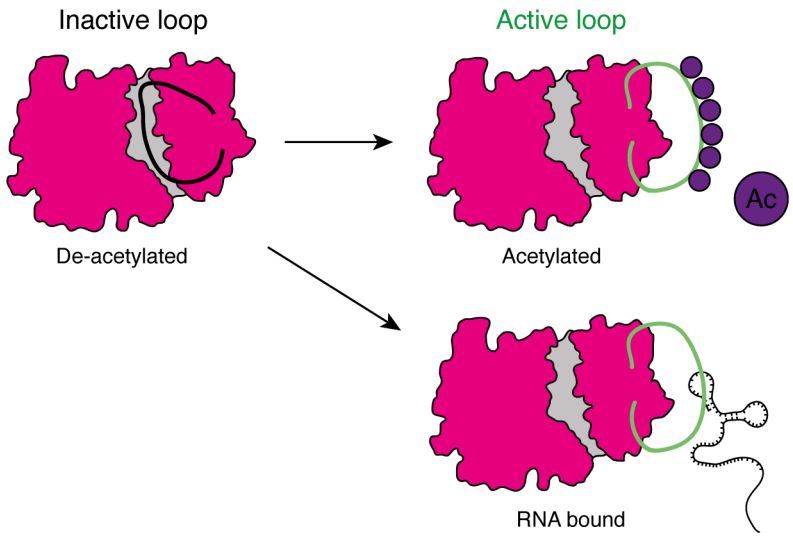

Figure 1.18: Mechanisms of CBP HAT activation
CBP in the inactive form, has the AIL (black loop) positioned within the active site in a de-acetylated form. The AIL can be activated by acetylation (purple circles, Ac) of key residues on the AIL neutralising the positive charge of the lysine residues, or through binding eRNAs to the RNA binding region within the AIL, both mechanisms displace the loop from the active site and activate HAT activity. CBP HAT domain diagram made by Dr Dan Bose and adapted here, diagram based on the HAT domain structure published in Wang et al. (2008a).

The HAT domain is important in regulating the phase separating ability of p300
Since CBP is found at enhancers which may function through transcriptional condensates, it is probable that this phenomena will have an effect on its function. p300 has been shown to phase separate, with this phenomena being shown to influence HAT activity. Co-condensation of p300 with the transactivation domain of p65 showed an increase in promoter acetylation compared to when transfected with p53, a protein which does not colocalise with p300, suggesting that co-condensation of p300 stimulates it's HAT activity (Ma et al. 2021). Upon incubation with potent HAT inhibitor A-485, a seletive, competative inhibitor of p300/CBP activity, larger p300 condensates dispersed into smaller more numerous condensates, suggesting a role of HAT activity in regulating larger condensate formation (Ma et al. 2021). p300 was indirectly shown to have active HAT activity by monitoring binding of BRD4, which binds to acetyllysine residues; taken together this work suggests that recruiting p300 to condensates through transcription factor

transactivation domains increases the HAT activity of p300 (Ma et al. 2021).

Looking specifically at the core of p300 (aa 1024 - 1830), it was shown that the HAT domain can phase separate both in HeLa cells and *in vitro* (Zhang et al. 2021). The acetylation of the core region was indicated to negatively regulate phase separation, with a deletion of the AIL causing a reduction in droplet formation compared to WT, when tested *in vitro* using molecular crowding agent PEG 8000 (Zhang et al. 2021). While this indicates the AIL is important for phase separation of p300, when the AIL was deleted from full length p300 in HeLa cells, condensates were still able to form; therefore it is not the only regulator of p300 phase separation (Zhang et al. 2021). A catalytically dead mutant of the core of p300 (D1399A) also formed condensates *in vitro*, suggesting a potential role of condensates as storing inactive p300 molecules (Zhang et al. 2021). *In vitro* condensates made by the full core complex colocalised with fluorescein (FAM) labelled double stranded DNA and histone H3 tails, suggesting an overall role for p300 condensates that contain inactive p300 that is recruited to regions of heterochromatin (Zhang et al. 2021).

While it is unclear whether the HAT domain of p300 is active within condensates, whether in cells or *in vitro*, both studies suggest that the HAT is important in regulating phase behaviour. While this implies that the HAT domain, and the AIL may be involved in CBP phase separation, this has yet to be tested.

1.6.4 Current understanding of CBP phase separation

CBP, which is known to be over 60 % disordered, has been shown to form phase separated condensates which are dynamic in movement and behave in a fluid-like manner (Ma et al. 2021; Zhang et al. 2021; Wang et al. 2022b; Dyson and Wright 2016). However no studies have been performed to define IDRs within CBP to determine if they are important for phase separation, and which ones regulate this behaviour, or further affect CBP HAT activity.

1.7 Aims of this research project

Enhancers are DNA regulatory elements which increase the transcription of their target genes by acting as transcription factor binding hubs. These enhancers increase the transcription of their target genes through donation of these transcription factors to the target promoters, one mechanism as to how this occurs is through phase separation. Proteins commonly found at the enhancer region such as BRD4, RNAPII and transcriptional co-activator protein CBP have been shown to form condensates, where these condensates behave in a fluid-like manner. CBP is a

transcriptional coactivator protein which acts to bind transcription factors, bringing them to active enhancer regions. CBP also has histone acetyltransferase activity, depositing acetyl residues onto histones and other proteins, facilitating active transcription. Due to CBPs role in gene activation, misregulation of CBP can lead to diseases such as cancer, it is therefore imperative to understand how CBP acts under normal cellular conditions, to further understand the changes that occur in a diseased state. While evidence suggests that CBP can form phase separated condensates, this behaviour is not well characterised; it is still unknown what drives CBP phase separation, and what effect this has on its localisation and activity.

This doctoral work aims to characterise CBP phase separation, the behaviour of full length protein and determine what drives this behaviour. The specific aims of this project are as follows:

1. Characterise the phase separation of endogenous CBP, including its fluid-like behaviour and develop an overexpression system in mammalian cells to fully explore full length CBP phase separation.

2. Determine which regions of CBP influence this phase separation behaviour including IDRs and the AIL, using an overexpression model system.

3. Investigate the behaviour of key phase separating mutants at endogenous concentrations, and characterise the effect these mutants have on chromatin localisation and endogenous CBP phase behaviour.

4. To optimise experimental approaches to view conformational dynamics of CBP within a condensate as an indicator of CBP activity, and characterise the dynamics of single molecules of CBP within condensates.

The work in this doctoral thesis shows that endogenous CBP forms phase separated condensates that behave in a fluid-like manner. Using an overexpression model system, regions of CBP were defined that either inhibit or promote phase separation; suggesting that the phase separation of full length CBP is not driven by one particular region but that it is an accumulation of the different regions of CBP that cause the overall phenotype. Mutants that affect the phase separation of CBP were shown to behave consistently at endogenous levels, and when overexpressed are shown to influence the phase separation of endogenous CBP. Finally, I performed the optimisation of several experiments such as smFRET and single particle imaging techniques which in the future will provide a platform to understand the effect of phase separation on protein dynamics and activity. The protein toolkit developed in this work provides a platform to expand our knowledge on the diverse effects of phase separation on CBPs HAT activity.

Chapter 2

Materials and Methods

2.1 Cloning and plasmid preparation

2.1.1 Buffers

Luria broth (LB) agar: 3.2 g of Luria broth agar low salt formula (Melford) was dissolved in 100 ml Milli Q® and autoclaved at 121^{o}C for 20 minutes.

LB broth: 2.5 g of LB broth in 100 ml MilliQ and autoclaved at 121^{o}C for 20 minutes.

1x TAE: 20 ml of 50X Tris-acetate-EDTA (TAE) buffer (Fisher) was diluted in 1 L deionised water.

2.1.2 Molecular biology kits

PCR DNA clean up kit: Monarch PCR DNA Cleanup Kit (NEB)

Gel extraction: QIAquick Gel Extraction Kit (QIAGEN),
Monarch DNA Gel Extraction Kit (NEB)

Plasmid DNA extraction (small): Monarch Plasmid Miniprep Kit (NEB)

Plasmid DNA extraction (medium): QIAGEN Plasmid Plus Midi Kit (QIAGEN)

2.1.3 Plasmids

Plasmid	Bacterial Resistance	Mammalian Resistance	Source
pAcGFP-N1	Kanamycin	-	Clontech (#632426)
pHR-mCh-Cry2WT	Ampicillin	-	Addgene (#101221)
pHR-FUSN-mCh-Cry2WT	Ampicillin	-	Addgene (#101223)
pRetroX-pTRE3G	Ampicillin	Puromycin	Clontech (#631194)
pRetroX-pTet3G	Ampicillin	Neomycin/G418	Clontech (#631192)
pCLAP 3	Ampicillin	-	Kind gift from Dr Alison Twelvetrees
pET17_K560_HaloTag_Strep	Ampicillin	-	Kind gift from Dr Alison Twelvetrees
iRFP-GFPY39TAG	Ampicillin	-	Kind gift from Edward Lemke
pcDNA3.1-tRNAPyl/NESPylRSAF	Ampicillin	-	Kind gift from Edward Lemke
pFastBac1-ACD tRNA synthetase 82RS	Ampicillin	-	Kind gift from Edward Lemke
eRF1 E55D plasmid	Ampicillin	-	Kind gift from Jason Chin

Table 2.1: Plasmids used in this doctoral work

2.1.4 General cloning protocols

In cloning individual constructs, the general protocols used throughout the cloning procedure are detailed below. Following DNA assembly all individual constructs underwent bacterial transformation and plasmid purification as detailed in the following sections.

2.1.4.1 Polymerase chain reaction

All polymerase chain reactions (PCRs) were performed using PFuUltra II fusion HS DNA Polymerase (Agilent) using 50 µl reactions - specific components are indicated in Table 2.2. PCR settings used throughout are shown in Table 2.3 where an extension temperature of 68^{o}C was used for DNA templates greater than 10 Kb in size, or 72^{o}C for DNA templates less than 10 Kb.

Reagent	**Volume (µl)**
H_2O	up to 50
10x PFuUltra reaction buffer (Agilent)	5
DMSO (Sigma)	2.5
100 µM forward primer (IDT)	0.25
100 µM reverse primer (IDT)	0.25
PFuUltra II Hi Fidelity Polymerase (Agilent)	0.5
10 mM dNTP mix (Thermo Fisher)	5
DNA template	1

Table 2.2: Components used in 50 µl PCR reaction

Cycle step	**Temperature**	**Time**	**Number of Cycles**
Initial denaturation	95^{o}C	5 min	1
Denaturation	95^{o}C	45 s	35
Annealing	$55 - X^o$C (depending on the primer)	45 s	
Extension	68/72^{o}C	30 s/Kb	
Final Extension	68/72^{o}C	15 min	1

Table 2.3: Parameters used for PCR reaction

2.1.4.2 Gel electrophoresis and gel extraction

1 % agarose gels were made using UltraPure^{TM} Agarose (Thermo Fisher) and 1X TAE, prepared with SYBR^{TM} Safe DNA gel stain (Thermo Fisher) at a concentration of 1:10,000. Once set, the gel and cassette were transferred to a gel tank and the tank was filled with 1X TAE. Samples were prepared with 6X loading dye (NEB) and loaded in the wells along with the appropriate marker (Quick-Load 1 Kb DNA ladder, QuickLoad Purpler 100 bp DNA ladder or 1 Kb plus DNA ladder, all NEB). Gels were ran at a voltage between 100 - 120 Volts for a time of 0.5 - 1.5 hours before visualisation using blue light imaging on the G:box Chemi-XRQ (Syngene).

To purify DNA from an agarose gel, typically for cloning purposes, bands were identified

using blue light dark reader transilluminator (DR-45M, Clare Chemical Research), excised and placed into an microcentrifuge tube. QIAquick Gel Extraction Kit (QIAGEN), or Monarch DNA Gel Extraction Kit (NEB) was used to purify the DNA according to manufacturers instructions, which was eluted in approximately 30 µl. DNA concentrations were determined using Qubit dsDNA Broad Range (BR) assay kit (Thermo Fisher), according to manufacturers instructions.

2.1.4.3 Bacterial transformation

Competent *Stbl3* strain (Thermo Fisher) *Escherichia coli* cells were thawed on ice for 10 minutes. 1 - 8 µl of plasmid (between 1 pg - 100 ng) was added to 50 µl of competent cells in a microcentrifuge tube and gently mixed. The cells were incubated on ice for 30 minutes, before a heat shock was performed at 42^oC for 30 seconds, followed by 2 further minutes on ice. 450 µl or 950 µl of room temperature SOC media (Melford) was then added to the tube and the cells were incubated at 37^oC with vigorous shaking (300 rounds per minute (rpm)) in a thermocycler for 0.5 - 4 hours. The competent cells were pelleted at 3000 rpm for 5 minutes in a bench top centrifuge, and either 400 µl or 900 µl of supernatant was removed and discarded - leaving 100 µl of supernatant plus pellet. The cells were resuspended in the remaining 100 µl of supernatant and plated on LB agar plates containing either 100 µg/ml ampicillin or 50 µg/ml kanamycin, using glass beads. The plates were inverted before incubation at 37^oC for 12 - 16 hours, at which point the plates were retrieved and checked for colonies.

2.1.4.4 Preparation of plasmids and sequencing

Overnight cultures of single selected colonies were made in 15 ml tubes, containing 5 ml of LB broth and either 100 µg/ml ampicillin or 50 µg/ml kanamycin. Cultures were grown for 12 - 16 hours in a shaking incubator set to 37^oC at 180 rpm. Plasmid DNA was isolated from the cultures using Monarch Plasmid Miniprep Kit (NEB), and the correct sequence was checked using either Eurofins Genomics or GENEWIZ Sanger sequencing.

To make larger quantities of plasmid, *Stbl3* cells were transformed as described above. 1 colony from the resulting transformation was added to a 250 ml conical flask containing 100 ml LB broth suplemented with 100 µg/ml ampicillin or 50 µg/ml kanamycin. The cells were incubated at 37^oC for 12 - 16 hours at 180 rpm before the plasmid was purified using QIAGEN Plasmid Plus Midi Kit (QIAGEN).

Concentration of the resulting plasmid preparations were determined using Qubit dsDNA Broad Range (BR) assay kit (Thermo Fisher).

2.1.5 Cloning CBP into pAcGFP-N1

Full length human CBP was cloned into pAcGFP-N1 (ClonTech, #632426) in a two step approach using NEBuilder HiFi DNA Assembly protocol (NEB). Primers shown in Table 2.4 were designed using NEBuilder (NEB). First, Homo Sapiens CBP transcript variant 2 which is missing exon 5 (HsCBPv2, Sino Biological), and pAcGFP-N1 was PCR amplified and gel extracted using QIAquick Gel Extraction Kit (QIAGEN). DNA assembly was performed using NEBuilder HiFi DNA Assembly reaction protocol (NEB) following the manufacturer's guidlines with the following exceptions: a 2:1 insert to vector ratio was used, and the assembly reaction was incubated at 50^oC for 30 minutes.

To complete the full length CBP, the missing exon was purchased as a custom synthesised DNA sequence (IDT), and inserted into CBP-pAcGFP-N1 following the same protocol as for step one. The resulting plasmid will be referred to as CBP GFP.

Primer Name	**Primer Sequence 5'-3'**
CBP_fwd	taccgcgggcccgggatccaATGGCTGAGAACTTGCTG
CBP_rev	cccttgctcaccatgaccggCAAGCCCTCCACAAACTTC
pAcGFP-N1_fwd	CCGGTCATGGTGAGCAAGGGCG
pAcGFP-N1_rev	TGGATCCCGGGCCCGCGG
CBP_missing _exon_fwd	ggctgggaaagcctgccaagTTGCCCATTGTGCATCTTC
CBP_missing _exon_rev	agctggagaccccaggatggTTTGTTGGTTTCGCTTGTC
pAcGFP-N1-CBP_fwd	CCATCCTGGGGTCTCCAG
pAcGFP-N1-CBP_rev	CTTGGCAGGCTTTCCCAG

Table 2.4: Primers used in cloning full length human CBP into pAcGFP-N1

2.1.6 Cloning regions of interest into the optoDroplet vector

Regions of interest (ROIs) were cloned into pHR-mCh-Cry2WT (Addgene, #101221), also known as the "optoDroplet" plasmid, which was purchased from addgene. Before inserting the regions of interest a nuclear localisation signal (NLS) was added to the optoDroplet plasmid, as CBP is nuclear localised. To achieve this a custom synthesised DNA sequence containing the NLS from SV40 (IDT) was ordered.

2.1.6.1 Cloning NLS into pHR-mCh-Cry2WT optoDroplet plasmid

Primers were designed to amplify the NLS and input flanking restriction sites that would allow DNA assembly into the optoDroplet plasmid via restriction cloning (Table 2.5). The SV40-NLS

was PCR amplified, ran on an agarose gel and gel extracted, using QIAquick Gel Extraction Kit (QIAGEN). A restriction digest was then performed on the plasmid and amplified SV40 NLS using 1 µl *NotI* and 1 µl *SbfI* restriction enzymes (both NEB) in a 10 µl reaction, following manufacturer's guidelines. The reaction was ran on an agarose gel and extracted as previously described. DNA assembly was performed using T4 DNA ligase (NEB) following manufacturers guidelines.

2.1.6.2 Cloning regions of interest into pHR-mCh-Cry2WT-NLS

The regions of interest were cloned into the pHR-mCh-Cry2WT-NLS plasmid using HiFi DNA assembly protocol (NEB), where the regions of interest were inserted upstream of the mCherry tag. The regions of interest and plasmid were amplified by PCR, where the primers were designed using NEBuilder (NEB) (Table 2.5). Amplified DNA was gel extracted using QIAquick Gel Extraction Kit (QIAGEN) and both insert and vector underwent a *DpnI* digest (NEB), following the NEB protocol before DNA assembly. Assembly reactions were set up with 50 ng of vector with a 3:1 ratio of insert to vector, and completed following the manufacturer's guidelines.

Primer Name	Primer Sequence 5'-3'
SV40_NLS_NotIRD_fwd	CATCATGCGGCCGCAACTCCCACCTGCAACAT
SV40_NLS_SbfIRD_rev	CATCATCCTGCAGGTCGACTCTAGAGTCGCGG
pHR-mCh-Cry2WT-NLS_fwd	AGTGGAGCAAGGCAGGTG
pHR-mCh-Cry2WT-NLS_rev	CGTGAGAATTCTCGAGAGCTC
IDR1_opto_fwd	agctctcgagaattctcacgGAGCTTCTACGAGGAGGC
IDR1_opto_rev	tccacctgccttgctccactGATACCAGGTCCAGTCTG
IDR3_opto_fwd	agctctcgagaattctcacgAACCAACAAACCATCCTGG
IDR3_opto_rev	tccacctgccttgctccactACCGGTGCTAGAAGGAGG
IDR4_opto_fwd	agctctcgagaattctcacgCAAAAAGAACTAGAAGAAAAACGGAGGTC
IDR4_opto_rev	tccacctgccttgctccactCGCCTGGGCCATCATGTTG
AIL_opto_fwd	agctctcgagaattctcacgGACAGGCTCACCAGTGCC
AIL_opto_rev	tccacctgccttgctccactCTTCTTCTTGTTGGCGCG
IDR6_opto_fwd	agctctcgagaattctcacgCAGCACCGCCTGCAGCAG
IDR6_opto_rev	tccacctgccttgctccactGAGAATGTTCAGCACCTGCTGTTG
IDR7_opto_fwd	agctctcgagaattctcacgGTGGCCAATCAGCCCGGC
IDR7_opto_rev	tccacctgccttgctccactCGTGTCCCCCGTGGTGTC
CFID_opto_fwd	agctctcgagaattctcacgACCCCCACACAGCAGCCC
CFID_opto_rev	tccacctgccttgctccactCAAGCCCTCCACAAACTTCTCTAGC

Table 2.5: Primers used for cloning of the optoDroplet plasmid

2.1.7 Generating CBP-GFP IDR deletion mutants

Intrinsic disorder regions (IDRs) within CBP were determined using Predictor of Naturally Disordered Proteins (PONDR) disorder prediction software, where an IDR was classified as any consecutive stretch of 50 or more amino acids that have a probability of disorder above 0.5, on a scale of 0 - 1, where a probability of 1 is disordered (Xue et al. 2010). The CBP FUS interaction domain (CFID) where CBP interacts with a common phase separation protein Fused in Sarcoma (FUS), was further chosen as an additional region of interest.

To generate IDR deletion mutants PCR amplification was performed on CBP-GFP using primers adjacent to the region of interest (Table 2.6). After PCR amplification, bands of correct size were gel extracted using QIAquick Gel Extraction Kit (QIAGEN) and the resulting DNA underwent a *DpnI* (NEB) digest following manufacturer's instructions. The ends of the PCR fragment were ligated together using T4 DNA ligase (NEB) following recommended guidelines.

Primer Name	Primer Sequence 5'-3'
ΔIDR1_fwd	TGCATGAATGCTAACTTTAACCAG
ΔIDR1_rev	CGACAGTTGTTTATGTTTGGA
ΔIDR3_fwd	GTAAGGAAAGGCTGGCACGAA
ΔIDR3_rev	TCGCTTGTCACTGGCATTTTT
ΔIDR4_fwd	CCCGCTCAGAGCCAGTTTCTG
ΔIDR4_rev	TATCTTGTAGATTTTCTCTGC
ΔIDR6_fwd	AAATCAAACCCGCAGCTAATG
ΔIDR6_rev	GATCTGCTGCTGGCGGAGCTT
ΔIDR7_fwd	CTAGAGAAGTTTGTGGAGGGC
ΔIDR7_rev	GTACTTGGCTGTGCGCTGTTT
ΔCFID_fwd	CCGGTCATGGTGAGCAAGGGC
ΔCFID_rev	CCCGGGCGGTGCTGAGGTAGG

Table 2.6: Primers used in the cloning of CBP-GFP IDR deletion mutants
Δ indicates the missing region these primers were used to amplify, with intrinsic disorder regions (IDR) and the CBP-FUS interaction domain (CFID).

2.1.8 Cloning AIL mutants into CBP-GFP and key mutants

To investigate the autoinhibitory loop (AIL) of CBP two custom synthesised DNA sequences were ordered from IDT containing a central region of CBP which had the AIL mutated - either removing the whole AIL and replacing it with a glycine serine linker Gly-Ser-Ala-Gly-Ser-Ala-Ala-Gly-Ser-Gly-Glu-Phe (GSAGSAAGSGQF), or where the lysine residues within the AIL were replaced with glycine residues. These AIL custom synthesised DNA sequences were cloned by restriction digest into vector constructs: CBP-GFP, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP.

Both custom synthesised DNA sequences and plasmids were restriction digested using 1 µl *NruI-HF* and 1 µl *PmlI* (both NEB) in 10 µl reactions following manufacturer's guidelines. The digest was run on agarose gel and gel extracted using QIAquick Gel Extraction Kit (QIAGEN). The vector fragments were dephosphorylated using Quick Calf Intestinal Phosphatase (CIP, NEB) following manufacturers guidelines, and DNA purified using PCR DNA clean up kit (NEB). The assembly reaction was performed using T4 DNA ligase (NEB), where the concentrations of vector ranged between 75 - 100 ng with a ratio of insert:vector between 2:1 and 3:1.

In the case where the construct was not correctly assembled in the initial round of cloning - a second round of cloning was undertaken where the AIL insert was generated by restriction digest of the correctly made constructs $CBP_{\Delta IDR6:\Delta AIL}$-GFP and $CBP_{\Delta IDR7:KTG}$-GFP. This generated an increased concentration of the insert and facilitated a higher insert:vector ratio in the ligation

reaction. The rest of the cloning continued as described above.

2.1.9 Cloning CBP-GFP and key mutants into p-RetroX to generate doxycyline inducible cell lines

CBP-GFP, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP were cloned into pRetroX-pTRE3G plasmid (ClonTech, #631194) using restriction digest. All plasmids were restriction digested using 2 µl *BamHI-HF* and 2 µl *NotI-HF* (both NEB) in a 20 µl reaction following the manufacturer's protocol. DNA was extracted from an agarose gel using QIAquick Gel Extraction Kit (QIAGEN). The pRetroX-pTRE3G vector was dephosphorylated using Quick CIP (NEB) following manufacturers guidelines and the DNA was purified using Monarch PCR and DNA clean up protocol (NEB). The plasmid was assembled using a T4 DNA ligase (NEB) reaction which contained 5 ng of vector and 25 ng of insert individually.

2.1.10 Generating CBP-Halotag-Flag construct with mutants for tmFRET

CBP-HaloTag-Flag was generated from the CBP-GFP construct in a two step approach. Firstly, the GFP tag was replaced with the HaloTag, followed by insertion of the Flag tag. The resulting construct was then used to make transition metal-ion Fluorescence Resonance Energy Transfer (tmFRET) mutants via restriction digest to insert dihistidine mutants and a TAG stop codon for fluorescent amino acid labelling.

2.1.10.1 Replacing GFP in CBP-GFP with the HaloTag

Both the HaloTag plasmid, pET17_K560_HaloTag_Strep which was kindly provided by Dr Alison Twelvetrees, and CBP-GFP plasmid were amplified by PCR, using primers designed using NEBuilder (NEB) (Table 2.7). The PCR was extracted from an agarose gel using QIAquick Gel Extraction Kit (QIAGEN). The DNA was *DpnI* digested (NEB), as described by the manufacturer, before assembly using NEB HiFi DNA Assembly reaction (NEB) with a 2:1 insert:vector ratio.

Primer Name	Primer Sequence 5'-3'
CBP-GFP_fwd	TGAGCGGCCGCGACTCTA
CBP-GFP_rev	GACCGGCAAGCCCTCCAC
HaloTag_fwd	ttgtggagggcttgccggtcATGGGATCCGAAATCGGTAC
HaloTag_rev	tctagagtcgcggccgctcaTTAACCGGAAATCTCCAGAG

Table 2.7: Primers used to clone HaloTag into CBP plasmid

2.1.10.2 Inserting Flag tag at the C-terminus of CBP-HaloTag

Once the CBP-HaloTag construct had been cloned, a Flag tag was inserted onto the C-terminus of this construct to allow purification from HEK293FT cells for future smFRET applications. A custom synthesised DNA sequence (IDT) of the the C-terminal region of the plasmid was used, containing the additional Flag tag sequence and restriction sites used to clone the Flag into the CBP-HaloTag construct. Restriction digest was performed using 1 µl *HpaI* and 1 µl *BssHII* (both NEB) in a 10 µl reaction, following manufacturer's guidelines, including a heat inactivation step. The DNA was extracted from an agarose gel using the QIAquick Gel Extraction Kit (QIAGEN). The CBP-HaloTag cut vector was dephosphorylated using Quick CIP (NEB), following recommended guidelines and the DNA was cleaned up using Monarch PCR and DNA clean up protocol (NEB). The DNA was assembled using T4 DNA ligase (NEB) following recommended guidelines.

2.1.10.3 Generating CBP-HaloTag-Flag tmFRET mutants

Mutants of the CBP-HaloTag-Flag construct were then generated for use in tmFRET experiments. Custom synthesised DNA sequences were generated of CBP's sequence containing nucleic acids 3656 - 5152 (from CBP's start codon), each containing a pair of mutated histidines and the mutated TAG sequence. In total, four DNA sequences were synthesised by IDT, the mutations in each DNA sequence can be found in Table 2.8. Cloning was performed by restriction digest using 1 µl *NruI* and 1 µl *PmlI* (both NEB) in a 10 µl reaction following recommended settings, including heat inactivation. The correct fragments were gel extracted using the QIAquick Gel Extraction Kit (QIAGEN). A T4 DNA ligase (NEB) reaction was performed using 100 ng of vector and 50 ng of insert (which was an insert:vector ratio of between 3:1 and 5:1).

The resulting constructs were termed CBP_{HHE^*D}-HaloTag-Flag, CBP_{HHEQ^*}-HaloTag-Flag, CBP_{NHH^*D}-HaloTag-Flag and CBP_{NHHQ^*}-HaloTag-Flag, based on the amino acid mutations within the AIL shown in Table 2.8.

Construct	Mutations
WT CBP	N1547, E1551, E1555, Q1579, D1665
CBP_{HHE*D}-HaloTag-Flag	H1547, H1551, E1555, *1579, D1665
CBP_{HHEQ*}-HaloTag-Flag	H1547, H1551, E1555, Q1579, *1665
CBP_{NHH*D}-HaloTag-Flag	N1547, H1551, H1555, *1579, D1665
CBP_{NHHQ*}-HaloTag-Flag	N1547, H1551, H1555, Q1579, *1665

Table 2.8: Mutations in CBP-HaloTag-Flag for tmFRET
Table shows the important amino acids in single letter form for CBP wild type and the 4 different synthesised DNA sequences for CBP-HaloTag-Flag. N = Asparagine, E = Glutamate, Q = Glutamine, D = Aspartate, * = TAG.

2.1.11 Adding SNAP tag into CBP-HaloTag-Flag tmFRET constructs for single molecule imaging

For single molecule FRET experiments the SNAP tag (NEB) was cloned within the sequence of CBP in two positions, upstream of the histone acetyltransferase (HAT) domain in the IDR (between nucleotide 3231-3232, from the start of the CBP sequence), and downstream of the HAT domain before the ZZ-Taz2 domain (between nucleotide 5097-5098, from the start of the CBP sequence). In the case of inserting the SNAP tag between the HAT domain and the ZZ-Taz2 domain, a glycine serine linker was added, Gly-Ser (GS), to maintain the structure of the HAT domain as the insertion site breaches the end of an alpha helix. The cloning strategy used to insert the SNAP tag was a two step approach. Firstly, the SNAP tag was inserted in both positions in the CBP_{HHE*D}-HaloTag-Flag construct, using HiFi DNA assembly (NEB). Once this was completed, restriction cloning was used to insert the SNAP tag into the remaining CBP-HaloTag-Flag constructs.

2.1.11.1 Inserting SNAP tag into CBP_{HHE*D}-HaloTag-Flag by HiFi DNA assembly

Primers were designed for the vector CBP_{HHE*D}-HaloTag-Flag and the insert SNAP tag using the NEBuilder tool (NEB). Two sets of primers were designed to insert the SNAP tag in both positions within CBP-HaloTag-Flag (Table 2.9). PCR was used to amplify the vector, CBP_{HHE*D}-HaloTag-Flag, and the SNAP tag (NEB) from pCLAP3; a custom plasmid containing FLAG-CLIPf-small linker-SNAPf-Double Strep-Tag kindly provided by Dr Alison Twelvetrees. The DNA was extracted from an agarose gel using Monarch DNA Gel Extraction Kit (NEB). HiFi DNA Assembly reaction (NEB) was performed with a 5:1 insert:vector ratio, following manufacturers guidelines.

The resulting constructs were termed $CBP_{HHE*D\ SH}$-HaloTag-Flag for the tag upstream of the HAT domain (SNAP-HAT, SH) and $CBP_{HHE*D\ HSZ}$-HaloTag-Flag for the tag down stream of the HAT domain (HAT-SNAP-ZZ, HSZ).

Primer Name	**Primer Sequence 5'-3'**
CHF_SH_fwd	TCGCAGCCGCGCAAAAAAATC
CHF_SH_rev	AGGAGATGTTGACTGAGAGGC
SNAP_fwd	cctctcagtcaacatctcctGATAAGGATTGTGAAATGAAGCG
SNAP_rev	atttttttgcgcggctgcgaACCAAGTCCAGGCTTTCC
CHF_HSZ_fwd	CAGGACCGCTTTGTCTACACC
CHF_HSZ_rev	GCCCTGGGTGTGCAGCTC
GS_SNAP_fwd	tggagctgcacacccagggcAGTGGTGATAAGGATTGTGAAATG
GS_SNAP rev	gtgtagacaaagcggtcctgACCAAGTCCAGGCTTTCC

Table 2.9: Primers to insert the SNAP tag into the CHF tmFRET constructs
Primers used in the HiFi DNA assembly of SNAP tag into two different positions within the CBP_{HHE*D}-HaloTag-Flag construct. Where CHF stands for CBP-HaloTag-Flag, SH stands for SNAP-HAT, for the first position of cloning, used in conjunction with SNAP_fwd/rev. HSZ stands for HAT_SNAP_ZZ/Taz2 which is used in conjunction with GS_SNAP.

2.1.11.2 Cloning the SNAP tag into the remaining CBP-HaloTag-Flag mutant constructs

Secondly, the SNAP tag was cloned into the remaining three CBP-HaloTag-Flag constructs using a restriction digest approach. For SH cloning restriction enzymes *AgeI-HF* and *NruI-HF* were used, and for HSZ cloning *BsrGI-HF* and *PmlI* (all NEB). Restriction digests were set up containing plasmids $CBP_{HHE*D\ SH}$-HaloTag-Flag or $CBP_{HHE*D\ HSZ}$-HaloTag-Flag with their respective enzyme pairs, while the remaining three CBP-HaloTag-Flag mutant constructs were set up with both pairs of restriction digests separately. 10 µl reactions were set up using 1 µl of each respective enzyme, following NEB manufacturer's guidelines, including a heat inactivation step. The DNA was extracted from an agarose gel using Monarch DNA Gel Extraction Kit (NEB). The vector gel extracts (CBP_{HHEQ*}-HaloTag-Flag, CBP_{NHH*D}-HaloTag-Flag and CBP_{NHHQ*}-HaloTag-Flag) were then dephosphorylated using Quick CIP following manufacturers guidelines, and Monarch PCR DNA Cleanup Kit (NEB) was performed to remove any residual enzymes or buffer. T4 DNA ligase (NEB) reaction was performed for DNA assembly with a 2:1 insert to vector ratio.

2.2 Cell culture

2.2.1 Cell lines

Cell line	Description	Source
HEK293T	Immortalised human embryonic kidney cells that express a mutant version of the SV40 large T antigen	Prof. Stuart Wilson, The University of Sheffield
GP2-293	Derived HEK293 cell line engineered to express MoMuLV Gag and Pol proteins used for viral transduction	Clontech
HEK293FT	HEK293-derived cells for optimised transfection to produce high viral titres as they are fast growing and stably express the large T antigen	Thermo Fisher
HEK293T - CBP-Flag-TEV-6His-HaloTag (CBP-HaloTag)	HEK293T cells containing endogenous CBP tagged with Flag-Tev-6His-HaloTag in both polyclonal and clonal form	Made in house by Dr Nicola Carruthers and Sophie Ball
HEK293T-Tet3G	Polyclonal HEK293T cell line containing the Tet-transactivator protein, a tetracycline-dependent transcription activator	Made in house
HEK293T-Tet3G-TRE3G-CBP-GFP (and mutants)	Clonal HEK293T cell line containing Tet-transactivator protein and doxycyline inducible CBP-GFP or mutants including, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP or $CBP_{\Delta CFID}$-GFP	Made in house

Table 2.10: Cell lines used within the doctoral work in this thesis

2.2.2 Tissue culture reagents and growth media

HEK293T, HEK293FT, CBP-HaloTag: Cytivia HyClone^{TM}Dublecco's High Glucose Modified Eagles Medium (DMEM, Thermo Fisher), supplemented with 10% Gibco^{TM} Fetal Bovine Syrum (FBS) and 1% Penicillin Streptomycin (P/S) solution (Thermo Fisher).

HEK293T-Tet3G-TRE3G-CBP-GFP (and mutants): DMEM supplemented with 10% Gibco^{TM}FBS, heat inactivated, E.U. approved, South American Origin, tetracycline-free (Thermo

Fisher) was used for viral transduction. For passaging cells this media was supplemented with 1% P/S for regular passaging. For transfected cells 2 mg/ml of G418 and 0.25 μg/ml puromycin was applied as continual selection.

Dulbecco's Phosphate Buffered Saline (DPBS), $Ca^{2+}$and Mg^{2+}free (Sigma-Aldrich) and 0.25% Trypsin dissociation reagent (Thermo Fisher) were used throughout.

Imaging media: FluoroBright DMEM (Fisher) supplemented with a final concentration of 20 nM Gibco HEPES (ThermoFisher Scientific).

2.2.3 Selection antibiotics

Neomycin: G418 Sulphate (BioVision)

Puromycin: Puromycin Dihydrochloride (Sigma-Aldrich)

2.2.4 Routine culturing procedures

All cells were cultured at 37^{o}C with 5 % CO_2.

For instances where accurate cell numbers were needed, cells were counted using Countess II (ThermoFisher) following manufacturer's instructions.

Cryopreservation of cells used cell type specific media supplemented with 10 % dimethyl sulfoxide (DMSO), to act as a cryoprotectant. 1 ml aliquots containing cells, typically at a split of 1:8 - 1:10, were frozen overnight at - 80^{o}C using Mr. Frosty^{TM} Freezing Container (Thermo Scientific) which contained isopropanol and decreased the temperature gradually, before being transferred to liquid nitrogen, at approximately - 196^{o}C, for storage. To revive cells, 1 ml aliquots were thawed rapidly at room temperature and used immediately.

2.2.5 Culturing practices of HEK293T cells

HEK293T and derived lines were thawed quickly and plated in 10 cm dishes with 9 ml of respective media. After 16 - 24 hours of culture, the media was changed to fresh media. Alternatively, thawed cells were resuspended in 4 ml of fresh media in a 15 ml tube, centrifuged at 200 relative centrifugal force (rcf) before the freezing media was removed and the cell pellet was resuspended in 10 ml of culture media and plated in 10 cm dishes. Both methods were used to remove any remnants of DMSO from the cryopreservation media. Cells were typically passaged with ratios

between 1:5 - 1:15 and split every 3 - 5 days. To passage cells, the media was aspirated and discarded before washing in 1 ml DPBS for 30 seconds. After aspirating the PBS, 1 ml of 0.25 % Trypsin was added to the dish and tipped from side to side to spread the liquid evenly over the surface. After discarding the Trypsin, the cells were then incubated at 37^oC with 5 % CO_2 for 4 minutes, before resuspending in the appropriate amount of media, for example 6 ml for a 1:6 split. After resuspending the cells, 1 ml of culture was transferred to 9 ml of media in a new 10 cm dish before swirling to mix and incubating at 37^oC with 5 % CO_2. In the event that cells were plated in another cell dish, for example after cell sorting, the amount of DPBS/Trypsin was scaled according to plate growth surface area.

2.2.6 Culturing under antibiotic selection

Antibiotic selection was applied to HEK293T-Tet3G and HEK293T-Tet3G-TRE3G-CBP-GFP and mutants to maintain integration of the gene of interest, where the gene of interest had a suitable antibioitic selection marker. pRetroX-Tet3G was under G418 (neomycin) resistance and pRetro-X-TRE3G was under puromycin resistance.

To determine which concentration to use for selection, a concentration curve of G418 was performed in HEK293T cells and of puromycin in HEK293T-Tet3G cells, which were plated in 6-well plates. 300,000 cells/well were plated and placed under antibiotic selection. Cells were split every 3 - 4 days and antibiotic media was changed, with a visual percentage of cell survival noted throughout the timecourse. Concentrations used for G418 were 8 mg/ml, 4 mg/ml, 2mg/ml, 1 mg/ml, 0.5 mg/ml and 0 mg/ml. Concentrations for puromycin were 1 µg/ml, 0.5 µg/ml, 0.25 µg/ml, 0.125 µg/ml, 0.0625 µg/ml and 0 µg/ml. The results showed concentrations of 8 mg/ml for G418 and 0.5 µg/ml for puromycin was sufficient to kill the appropriate cells over an 8 day period, and this was used as the minimum lethal dose. Maintenance concentrations of 2 mg/ml of G418 and 0.25 µg/ml puromycin were used post-selection. Antibiotics were applied to transduced cells 24 hours post transduction and replaced every 3 days, cells were passaged as needed and antibiotics replaced with the concentrations described.

2.2.7 Transfecting HEK293T cells

Cells were transfected through chemical means using both, Lipofectamine 3000 (Thermo Fisher) and LipoD293 (TEBU-Bio). For both transfection methods DMEM high glucose was used in the absence of FBS and P/S for diluting DNA and respective reagents.

2.2.7.1 Lipofectamine 3000 (Thermo Fisher)

Cells were plated 24 hours prior to transfection in either 35-mm glass bottom dishes, 6-well plates or T-75 flasks for viral transduction. Two microcentrifuge tubes were prepared, one tube containing the DNA to be transfected in DMEM high glucose (Thermo Fisher), with the addition of the p3000 reagent (Thermo Fisher). The second tube contained DMEM high glucose (Thermo Fisher) and the Lipofectamine 3000 reagent (Thermo Fisher). Exact volumes and concentrations can be found in Table 2.11. Once prepared the tubes were combined and incubated at room temperature for 5 minutes. The resulting complex was added drop-wise to the cells and incubated for 24 hours before further use.

	Reagent	**6-well/35mm dish**	**T-75 flask**
Tube 1	DMEM	150 µl	750 µl
	Lipofectamine 3000	7.5 µl	45 µl
Tube 2	DMEM	150 µl	750 µl
	Plasmid	1.5 µg	15 µg
	p3000	6 µl	30 µl

Table 2.11: Lipofectamine 3000 (Thermo Fisher) reagent volumes

2.2.7.2 LipoD293 (TEBU-Bio)

Transfection with LipoD293 (TEBU-Bio) was completed following manufacturers guidelines. Cells were plated 24 hours before transfection to be approximately 30 - 40% confluent on day of transfection. Many different dishes were used for transfection including: 8-chambered coverglasses (Ibidi, IB-80807), 35 mm glass bottom dishes (Ibidi), 6-well plates (Corning) and 10 cm dishes (Corning). On the day of transfection, the media was exchanged to DMEM containing FBS and P/S. Two tubes were initially prepared containing DMEM high glucose without FBS and P/S, tube one contained the DNA only and the second tube contained the LipoD293 reagent (TEBU-Bio). Specific volumes for each plate/dish are listed in Table 2.12. Both tubes were combined, mixed and incubated at room temperature for either 15 minutes (version I) or 10 minutes (version II) before the solution was added drop-wise to the cells and the culture was swirled to mix. Media was replaced 12 - 18 hours post-transfection. After 24 - 48 hours post-transfection cells were used in various experiments detailed below.

Culture dish	Media volume	Plasmid concentration	Diluent volume	LipoD293 volume
8-chambers	300 µl	100 ng	2 x 15 µl	0.75 µl
6-well/35-mm dish	1000 µl	1 µg	2 x 50 µl	3 µl
10-cm dish	5000 µl	5 µg	2 x 250 µl	15 µl

Table 2.12: LipoD293 (TEBU-Bio) reagent volumes

2.2.8 Viral transduction

2.2.8.1 Transfection of packaging cells

For viral transduction of the Tet3G plasmid, 1.94 x10^6 GP2-293 cells/well were plated in a 6-well plate in 2 ml of antibiotic-free, tet-free media. For viral transduction of the TRE3G constructs, cells were plated in T-75 flasks at a cell density of 7.2 x 10^6 in a volume of 12 ml. After 24 hours the cells were transfected using Lipofectamine 3000 reagent (Thermo Fisher), as described in section 2.2.7.1, components listed in Table 2.13. 15 - 17 hours post-transfection the media was changed to tet-free antibiotic free media.

	Reagent	Tet3G transduction	TRE3G transduction
Tube 1	DMEM	125 µl	750 µl
	Lipofectamine 3000	7.5 µl	45 µl
Tube 2	DMEM	125 µl	750 µl
	Plasmid	2 µg	12 µg
	pVSV-G envelope	2 µg	3 µg
	p3000	8 µl	30 µl

Table 2.13: Preparation of transfection mixtures

Plasmid used in Tet3G reaction was pRetroX-Tet3G, and in TRE3G reaction it was p-RetroX-TRE3G-CBP-GFP, pRetroX-TRE3G-CBP$_{\Delta IDR6}$-GFP, pRetroX-TRE3G-CBP$_{\Delta IDR7}$-GFP and pRetroX-TRE3G-CBP$_{\Delta CFID}$-GFP.

2.2.8.2 Transduction with concentrated virus

24 hours post-media change, the media from GP2-293 packaging cells was harvested and placed in a 15 ml tube and centrifuged at 500 rcf for 10 minutes to remove any cellular debris. The supernatant was then transferred to a 15 ml tube and combined with 1 x volume of lenti-X concentrator (Takara Bio), before being mixed and incubated overnight at 4^{o}C. The following

day the samples were centrifuged at 1,500 rcf for 45 minutes at 4^{o}C, and the supernatant was discarded and the pellet resuspended in 50 µl DPBS. If the virus was not to be used immediately, samples were flash frozen in liquid nitrogen and stored at -80^oC.

For viral transduction HEK293T cells were counted using the Countess II and plated at a seeding density of 8 x10^5 cells/well in a 6-well dish in media without antibiotics. After 24 hours, 1 ml of media from every well was removed and stored in a 1.5 ml microcentrifuge tube. The virus was combined with 8 µg/ml of polybrene (Merck Millipore) and combined with 1 ml of media. The virus media cocktail was then added to the wells drop-wise, 1 well at a time and incubated at 37^{o}C with 5 % CO_2. After 24 hours incubation, the media was exchanged for media containing antibiotics to select for correctly transduced cells. Cells were cultured until fully confluent, at which point aliquots were taken for cryopreservation and the antibiotic selection was reduced to maintenance concentration. Cells were expanded before being sorted into polyclonal and clonal lines by fluorescence activated cell sorting (FACS) and screened via imaging.

2.2.9 Fluorescence activated cell sorting

Fluorescence activated cell sorting (FACS) was carried out at the Flow Cytometry Facility in the Medical School at the University of Sheffield, where technical support was given for training purposes as well as on the day assistance. FACS was completed by Dr Nicola Carruthers, PhD student in the Bose lab. Preparation of cells and post cell culture was completed by myself.

FACS was carried out on cell lines HEK293T-Tet3G-TRE3G-CBP-GFP, HEK293T-Tet3G-TRE3G-CBP$_{\Delta IDR6}$-GFP, HEK293T-Tet3G-TRE3G-CBP$_{\Delta IDR7}$-GFP and HEK293T-Tet3G-TRE3G-CBP$_{\Delta CFID}$-GFP and on CBP-HaloTag cells to increase the percentage of transduced cells. For each cell line a bulk population that had an increased percentage of cells containing the appropriate construct and single cell clones were generated.

2.2.9.1 Preparation of cells for FAC3

Doxycycline-inducible cell lines For FACS, cell lines were used at as early a passage as possible. Cells were plated 72 hours prior to FACS in media containing selection antibiotics in 15 cm dishes to be confluent on the day of FACS. After 24 hours, the cells were induced with doxycycline (Dox) at a concentration of 2 µg/ml; 5 ml of media was removed from the dish, mixed with the appropriate volume of Dox and added back to the plate gently and incubated for 48 hours. On the day of FACS, fresh Dox was added to the cells.

For each cell line, 1 x 96-well plate was prepared for single cell clones per construct, and 1 x

6-well dish was prepared for the bulk sorts. The plates were prepared with conditioned media; tet-free antibiotic-free media, that had been harvested from healthy HEK293T-Tet3G cells, grown for 24 hours. The media was centrifuged at 400 rcf for 15 minutes, the supernatant was collected and syringe filtered using 0.45 μm filter to remove any remaining cell debris. 100 μl of conditioned media was added per well of a 96-well plate prior to FACS. For the bulk sorts conditioned media was taken in a 50 ml tube for preparation at the facility. For each cell line an appropriate control was used, containing the respective cell line in the absence of Dox which was used to help set gates for FACS.

Before FACS the cells were lifted and resuspended in 5 ml DPBS and transferred to 15 ml tubes and centrifuged at 200 rcf for 5 minutes. The supernatant was removed and the cells resuspended in 1 ml DPBS. All tubes and plates were sealed with Parafilm (Sigma) and transported to the facility on ice. Collection tubes for cell samples were provided by Flow Cytometry Core facility and did not need to be prepared ahead of time.

CBP-HaloTag endogenously tagged cell lines Early passage CBP-HaloTag cells were plated in 3 x 15 cm dishes to be confluent 48 hours post-plating. 2 hours before sorting, 2 x 15 cm dishes containing cells were labelled with 50 nM TMR for 15 minutes in culture media at 37^oC, the third dish was used as an unlabelled control. After 15 minutes the media was changed to fresh media and left at 37^oC for the ligand to wash out. Before FACS the cells were lifted, and each plate was resuspended in 5 ml of DPBS and transferred to 15 ml tubes where they were centrifuged at 200 rcf for 5 minutes. After centrifugation, the DPBS was removed and cells were resuspended in DPBS at volumes of 1 ml for the negative control, 2 ml and 4 ml respectively for the two labelled plates. The different dilutions were to optimise cell confluency for the FACS machine to maximise the number of cells that were recovered.

1 x 96-well plate was prepared with 100 μl of conditioned media and a 48-well plate was prepared with 200 μl for the bulk polyclonal cells post-FACS. Conditioned media was harvested from healthy HEK293T cells after 24 hours, centrifuged at 400 rcf for 15 minutes, before filtering and storing at 4^oC until the day of use. Plates, cells, and extra conditioned media and DPBS were parafilmed and transported to the Flow Cytometry Facility in the Medical School on ice before FACS.

2.2.9.2 GFP bulk sorts

For each cell line a bulk sort was carried out to increase the proportion of cells positive for the desired GFP expressing construct. The GFP tag was selected for using the 488 nm laser. The machine used was FACSMelody Cell Sorter (BD Biosciences) and was operated by Dr Nicola Carruthers. Gates were set using the no Dox version of each cell line, where forward scatter was compared to side scatter of single cells, to obtain a healthy population of single cells. Measuring fluorescence at 488 nm, the Dox induced cell line was used to adjust the gate to select for GFP positive cells. Once set, GFP positive samples were collected into a single tube containing 1 ml of conditioned media to aid cell survival. Cells were transported back to our TC suite on ice, where they were centrifgued at 400 rcf for 5 minutes to remove any DPBS. Depending on cell number, the cells were resuspended in either 2 ml or 0.5 ml conditioned media and plated in a 6-well or a 48-well plate respectively.

2.2.9.3 GFP single cell sort

Single cell sorting was performed by Sue Clark, member of the Flow Cytometry Facility in the Medical School at the University of Sheffield, due to issues with the plate reader on the FACSMelody Cell Sorter (BD Biosciences). In this case, gates were set as described for GFP bulk sorts and cells were sorted into a 96-well plate, where a single cell was plated in each well containing 100 µl of conditioned media. Plates were parafilmed before being transported back to our TC suite, where the parafilm was removed and the cells were left to grow at 37^{o}C with 5 % CO_2 for a minimum of 2 weeks before expanding.

2.2.9.4 TMR bulk sort

A bulk sort was completed to generate a population of cells that had an increased proportion of positive CBP-HaloTag cells. These cells were selected for using the 561 nm laser on the FACSMelody Cell Sorter (BD Biosciences) operated by Dr Nicola Carruthers. Gates were set using the no TMR CBP-HaloTag cells, where forward scatter was compared to side scatter of single cells to obtain a population of live cells. The gates were then adjusted for the plus TMR conditions to collect cells that were labelled, with a stringent gate such that positive cells accounted for 1 % of the population. The cells were collected into a single tube containing 1 ml of conditioned media to help with cell survival. The cells were transported back to our TC suite on ice, where they were spun at 400 rcf for 5 minutes to remove any DPBS. The cells were resuspended in conditioned media and then plated in one well of a 48-well plate until ready to expand.

2.2.9.5 TMR single cell

Single cell sorting was set up as described in the previous section using the FACSMelody Cell Sorter (BD Biosciences) performed by Dr Nicola Carruthers. Cells were sorted into a 96-well plate containing 100 µl of conditioned media, plates were parafilmed before being transported back to our TC suite before the parafilm was removed and the cells were left to grow at 37^{o}C with 5 % CO_2 until ready to expand.

2.2.9.6 Culturing processes following FACS

Cells sorted via FACS were closely monitored for cell growth following the procedure. For single cell clones, if there was indication that the cells had expanded they were gently resuspended to avoid clumping and to encourage cell growth. When cells appeared to be expanding, for example if the media changed colour, cells were lifted and placed in a larger vessel with a larger volume of media. The expanding process followed the route of 96-well, 48-well, 24-well, 6-well and finally a 10 cm dish. Once cells were confluent they were harvested for cryopreservation before validating clones by western blotting and imaging.

2.2.9.7 Screening cells post FACS

Positive clones for both Dox inducible cell lines and CBP-HaloTag endogenous HEK293T cells were screened via imaging using high well 8-chambered coverslips (Ibidi). Dishes were coated in poly-L lysine (PLL, Merck) at a concentration of 10 - 20 ng/ml at 37^{o}C for 1 hour in an appropriate volume. Dishes were subsequently washed twice in dH_2O before leaving to dry before cell plating.

Screening for GFP tag Cells were plated at an appropriate cell density to be approximately 80 - 100 % confluent on the day of imaging. For Dox inducible cell lines, cells were induced with 2 µg/ml of Dox for 48-hours prior to imaging, with induction maintained in the imaging media. Cells were imaged on the Nikon A1 confocal microscope using the 60 X oil objective using the 488 nm laser line. Full microscopy details can be found in section 2.3.1.

Screening for HaloTag For CBP-HaloTag cell lines, cells were labelled with 50 nM TMR for 15 minutes at 37^{o}C in full culture media, before a period of 1 - 2 hours of recovery in fresh media. Cells were subsequently washed in DPBS and placed in imaging media before screening. Nikon W1 spinning disc was used to capture images using the 100 X oil objective lense, and the 561

nm laser line with 300 ms exposure on 60 % laser power. Full microscope details can be found in section 2.3.2.

2.2.10 Fixing and mounting coverslips for overexpression experiments

2.2.10.1 Preparing coverslips

Coverslips were washed in 0.25 % acetic acid before being washed and stored in 100 % ethanol to remove any contaminants from the glass surface. Coverslips were dried to remove any residual ethanol before being placed in a suitable vessel and coated in poly-L lysine (PLL, Merck) at a concentration of 10 - 20 ng/ml for 1 hour at 37^{o}C. Coverslips were washed twice in dH_2O and were then transferred to another dish for cell plating, either a 6-well dish or 10 cm dish.

2.2.10.2 Cell culture and fixing

Cells were grown for 48 hours before fixing. If the cells were to be transfected, they were grown for 24 hours before transfection and subsequently grown for a further 24 hours before fixing. To fix, coverslips were transferred to a fresh 6-well plate and washed three times in DPBS. The cells were fixed in 2 % paraformaldehyde which was diluted from a 32 % stock (Electron Microscopy Sciences) with DPBS and had been heated to 37^{o}C. Cells were washed a further three times before continuation with either immunofluorescence or mounting the cells for storage.

2.2.10.3 Cell mounting

To mount, a drop of VECTORshield that contained 4,6-diamidino-2-phenylindole (DAPI) (Vector Laboratories) was placed on an imaging slide (VWR), one drop per coverslip, with a maximum of two coverslips per slide. Coverslips were inverted onto the mounting medium and were gently pushed onto the slide, with excess DPBS and mounting media being removed. The edges of the coverslip were then sealed using clear nail polish to help preserve the samples and left to dry at room temperature for approximately 30 minutes before storing at 4^{o}C until imaging.

2.2.11 Immunofluorescence of fixed cells

After fixing of the cells using paraformaldehyde cells were washed 3 times in DPBS. The DPBS was removed and 1 ml of blocking buffer (comprised of 5 % Bovine Syrum Albumin (BSA) and 0.2% Triton (Sigma) in DPBS) was added to each well and incubated for 1 hour at 37^{o}C. Primary antibody was prepared in blocking buffer and incubated for 1 hour at 37^{o}C, for full list of antibodies

used during immunofluorescence, see Table 2.14. After primary antibody incubation, cells were washed three times in DPBS. Secondary antibodies were diluted in blocking buffer, added to the cells and incubated at 37^{o}C for 1 hour. A further three washes was performed before mounting the coverslips onto slides as was described in section 2.2.10.3.

Antibody	Source	Cat. No.	Species	Type	Dilution
α-CBP/KAT3A/CREBBP C1	Santa Cruz	sc-7300	Mouse	Monoclonal	1:100
Recombinant α-BRD4	Abcam	ab128874	Rabbit	Monoclonal	1:200
α-Histone H3 (acetyl K27)	Abcam	ab4729	Rabbit	Polyclonal	1:500
Alexa Fluor 647 AffiniPure	Jackson Laboratories	711-605-152	Donkey Anti-Rabbit	-	1:500
Alexa Fluor 488 AffiniPure	Jackson Laboratories	715-545-150	Donkey Anti-Mouse	-	1:500

Table 2.14: Antibodies used in immunofluorescence

2.2.12 General PLL coating practices for live cell imaging

Imaging dishes including 35 mm high glass bottom μ-dishes (Ibidi, IB-81158) and 8-chambered high well cell culture slides (Ibidi, IB-80807) were coated will PLL before cell plating. 1 ml or 300 μl of PLL at a concentration of 10 - 20 ng/ml was added per dish/well respectively and incubated at 37^{o}C for 1 hour. Dishes were subsequently washed twice in dH_2O and either left to dry before cell plating or storage in dH_2O. Within this doctoral work PLL coating was performed on the day of plating.

2.2.13 Live cell culture of overexpression experiments (FRAP and 1,6-hexanediol treatments)

For live cell imaging of overexpressed CBP and mutants, cells were plated in 35 mm high glass bottom μ-dishes (Ibidi, IB-81158) which were PLL coated. Cells were plated at a concentration of 250,000 cells/well and incubated for 24 hours. Cells were transfected using LipoD293 for FRAP

experiments and Lipofectamine 3000 for 1,6-HD experiments, full details of those protocols can be found in section 2.2.7. Cells were incubated for a further 24 hours post-transfection.

Before imaging, nuclei were labelled with Hoechst 33342 (Fisher) using a 1:2000 dilution prepared in DPBS. The media on the cells was removed and an appropriate volume of diluted Hoechst was added per well and incubated in the dark, at 37^{o}C for 5 minutes. The cells were subsequently washed once in DPBS before the media was changed for imaging media. Imaging perameters are discussed in subsequent sections.

2.2.14 Plating cells for optoDroplet experiment

Cells were plated in either 35 mm high glass bottom µ-dishes (Ibidi) or 8-chambered high well cell culture slides (Ibidi) which were PLL coated before cell plating. Cells were plated at a concentration of 200,000 cells/well for 35 mm dishes, and 20,000 cells/well in 8-chambered coverslips. After 24 hours of culture, the cells were transfected using LipoD293 (TEBU-Bio) following the protocol described in section 2.2.7. Cells were cultured for a further 48 hours before media was exchanged with imaging media. In this experiment, Hoechst 33342 was not used to label the nuclei as the 405 nm laser, needed to visualise this stain, stimulated oligomerisation of the CRY2 domain of the optoDroplet construct and therefore could not be used to visualise nuclei.

2.2.15 Labelling HaloTag for imaging

2.2.15.1 Labelling CBP-HaloTag for full population imaging studies

For imaging endogenous CBP-HaloTag, cells were plated the same as was described in section 2.2.13. Briefly, cells were plated in 35-mm imaging dishes which were PLL-coated, at a concentration of 250,000 cells/well. Cells were grown for 48 hours before imaging. On the day of imaging, cells were stained with either TMR (Promega) or Janelia Fluor (JF) 549 (Promega), which are fluorophores conjugated to the HaloTag substrate. To optimise labelling, a range of different conditions were tested to optimise for signal of both dyes, these included concentrations of 50 nM, 100 nM, 250 nM, which were incubated for 15 minutes, 30 minutes or 60 minutes. From this it was determined that 50 nM of dye for 15 minutes was sufficient to selectively label the CBP-HaloTag with a reduced background, and that there was little difference between TMR (Promega) or JF 549 (Promega). Over the course of imaging, concentrations of 5 - 50 nM of stain was prepared, with both fluorophores being used for different experiments. Fluorophores were prepared in full DMEM containing FBS and P/S, exisiting media on the cells was aspirated and replaced with

media containing the fluorophore and incubated at 37^{o}C for 15 minutes. After this time the media was removed and replaced with fresh full DMEM, the cells were further incubated for 1 - 3 hours prior to imaging. Immediately before imaging the cells were Hoechst 33342 (Fisher) stained and the media was replaced with imaging media. In certain experiments HEK293T cells were used as a negative control, these cells were labelled as described for the CBP-HaloTag.

2.2.15.2 Labelling for single molecule imaging experiments

For single molecule imaging CBP-HaloTag cells (either polyclonal or clonal) were typically plated in PLL-coated 8-chamber coverglasses at a concentration of 50,000 cells per well. Cells were grown for 48-hours before labelling. For single molecule imaging cells were typically labelled to visualise the whole CBP population as well as single molecules, including use of a photoactivateable ligand or down labelling with JF 646 (Promega). To label for the whole CBP population, JF 549 (Promega) was used at a concentration of 50 nM which was incubated with the cells for 15 minutes. In the case of the photoactivateable ligand, the cells were also labelled with photoactivateable ligand 646 (Promega) at a concentration of 50 nM for 15 minutes. To achieve single molecule imaging the CBP-HaloTag cells were labelled with JF 646 (Promega) at concentrations between 0.5 nM - 5 nM and the cells were incubated for 15 minutes. Post-labelling cells were washed once in DPBS and left in imaging media for approximately 15 - 30 minutes before imaging. For single molecule imaging experiments cells were not labelled with Hoechst 33342 (Fisher).

2.2.16 Live cell culture of endogenous CBP-HaloTag with overexpressed CBP-GFP and mutant constructs

35 mm high glass bottom μ-dishes (Ibidi, IB-81158) were PLL-coated before 250,000 cells/well of endogenous CBP-HaloTag cells were plated and grown for 24 hours. After 24-hours the cells were transfected with either CBP-GFP, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP or $CBP_{\Delta CFID}$-GFP using the LipoD293 protocol, as described in section 2.2.7.2. 24 hours after transfection the cells were ready for imaging, in the case of 4-colour control images to check bleed-through of overexpressed constructs, the media was changed to imaging media, imaged as described in section 2.3.3.3, and then changed back to regular culture media before labelling and imaging. Cells were labelled with JF 649 at a final concentration of 2 nM for 15 minutes before washing in DPBS and imaging in imaging media. 35 mm dishes were labelled sequentially, immediately before imaging. Images were taken on custom TIRF microscope BONSAI, full details of the microscopy and analysis can be found in section 2.3.3.3.

2.2.17 Transfecting HEK293FT cells for smFRET experiments

Transfection protocol adapted from Zagotta *et al.* (2021) and Nikić *et al.* (2016). iRFP-GFP^{Y39TAG} (GFP positive control plasmid), pcDNA3.1-tRNA^{Pyl}/NESPylRS^{AF} (NES-PyIRS) were kind gifts from Edward Lemke, and eRF1 E55D plasmid was a kind gift from Jason Chin.

35-mm imaging dishes were PLL-coated before HEK293FT cells were plated at a concentration of 250,000 cells/well and left to grow for 24 hours. In this case, HEK293FT cells were used as they are more efficient in producing protein after transfection. After 24 hours, the cells were transfected using the LipoD293 protocol, as described in section 2.2.7.2. Cells were transfected with a total concentration of 1.05 µg of DNA which included 400 ng of NES-PyIRS (tRNA and aminoacyl tRNA synthetase), 400 ng of plasmid DNA (GFP positive control plasmid or CBP$_{HHE*D}$-HaloTag-Flag construct) and 250 ng of eukaryotic release factor 1 (eRF1) E55D plasmid. Immediately after transfection, unnatural amino acid {[(E)-cyclooct-2-en-1-yl]oxy}carbonyl)-L-lysine (TCO*A) was added drop-wise to a final concentration of 250 µM. After 8 - 10 hours incubation with TCO*A, the media was changed to remove any unused TCO*A. 24 hours after transfection the cells were prepared for imaging.

For the GFP positive plasmid test experiment cells were labelled with Hoechst 33342 (Fisher), using a 1:2000 dilution in DPBS. The media on the cells was removed and 1 ml was added per well and incubated in the dark, at 37^{o}C for 5 minutes. Cells were subsequently washed in DPBS before leaving in imaging media before imaging. If CBP-HaloTag plasmid was used, protein was labelled with JF 549 at a concentration of 50 nM for 15 minutes in the dark. The cells were then washed with DPBS before being put in imaging media for imaging. Full details of imaging can be found under the microscopy section 2.3.

2.2.18 Transfecting HEK293FT cells for tmFRET experiments

Transfection protocol adapted from Zagotta *et al.* (2021) and Nikić *et al.* (2016). iRFP-GFP^{Y39TAG} (GFP positive control plasmid), pFastBac1-ACD tRNA synthetase (82RS mutant) were kind gifts from Edward Lemke, and eRF1 E55D plasmid was a kind gift from Jason Chin.

For the initial tmFRET test experiment the protocol followed was as described in section 2.2.17 with the following changes: ACD tRNA synthetase (82RS mutant) was used instead of NES-PyIRS, with the unnatural amino acid L-Acridonylalanine (ACD). In this experiment, the ACD was prepared in 1M HEPES and was added 4 hours after transfection at a concentration of 450 µM and left on until Hoechst 33342 (Fisher) labelling and imaging.

In optimising this experiment, the above protocol was followed except the ACD was added

immediately after transfection and was removed after 6 - 8 hours. The cells were then prepared for imaging as was described in section 2.2.17.

2.2.19 Cell culture for western blotting

For western blotting cells were typically cultured in 10 cm dishes, with 1 dish per condition. Cells were split to be confluent on the day of cell harvest. For overexpression western blots, cells were transfected after 24 hours using the LipoD293 protocol, as described in section 2.2.7.2, and grown for a further 24 hours before the cells were harvested. In the case of the Dox inducible cell lines, cells were induced with the appropriate concentration of Dox after 24 hours and grown for a further 48 hours before cell harvest.

2.2.19.1 Plating cells for a hybrid western blot

For hybrid western blot - 2 circular coverslips were PLL-coated per condition, as described in section 2.2.10.1, and added to each 10 cm dish before plating cells. The cells were then plated in the dish, as usual, and if necessary the coverslips were pressed gently onto the dish to prevent further movement. On the day of harvest, the coverslips were removed from the 10 cm dish and placed in a 6-well plate, 1 well per condition. The coverslips were then fixed and mounted as described in section 2.2.10.

Details on harvesting cells for western blotting is described in section 2.4.3.

Doxycycline induction for hybrid western blot Cell lines including, HEK293T-Tet3G-TRE3G-CBP-GFP (and deletion mutants) as well as a HEK293T-Tet3G control, were plated in 10 cm dishes as described above. After 24 hours the cells were induced with Dox concentration curve, where one 10 cm dish is treated with one concentration. For CBP-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP, these concentration were: 0 ng/ml, 10 ng/ml, 25 ng/ml, 50 ng/ml, 75 ng/ml, 100 ng/ml and 125 ng/ml. For $CBP_{\Delta IDR6}$-GFP these concentrations include 0 ng/ml, 1 ng/ml, 2.5 ng/ml, 5 ng/ml, 7.5 ng/ml and 10 ng/ml. After 48 - hours of induction the cells were treated as described above.

2.2.20 Cell culture for ChIP-Seq for doxycycline inducible cell lines

For Chromatin immunoprecipitation experiments HEK293T-Tet3G-TRE3G-CBP-GFP (and deletion mutants) as well as a HEK293T-Tet3G control were plated in 10 cm dishes, 2 plates per construct. Cells were plated to be confluent after 72-hours. 24-hours after plating, the cells

were induced with Dox at near-endogenous concentration, 25 ng/ml for CBP-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP and 5 ng/ml for $CBP_{\Delta IDR6}$-GFP. 72-hours post plating the cells were cross-linked and harvested - see section 2.5 for full details.

2.3 Imaging and analysis

Imaging within this doctoral work was performed on three microscope. Two of these microscopes are housed in the Wolfsen Light Microscopy Facility (LMF) at the University of Sheffield; including the Nikon A1 confocal microscope and Nikon W1 spinning disc confocal microscope. Technical support was provided in the LMF by Dr Nicholas Van Hateren and Dr Darren Robinson who provided training and assistance over the course of this doctoral work. Further endogenous and single molecule imaging of CBP-HaloTag was performed at Sheffield Institute for Translational Nueroscience (SiTRaN) using custom built total internal reflection fluorescence (TIRF) and fluorescence recovery after photobleaching (FRAP) imaging system called BONSAI owned by Dr Alison Twelvetrees, who provided technical support for the microscope as well as assistance with planning and imaging analysis.

2.3.1 Nikon A1 confocal imaging

2.3.1.1 Microscope details

Nikon A1 confocal microscope was equipped with 60 X oil immersion objective that was used throughout this doctoral work. For live cell imaging, this microscope was equipped with a humidified temperature stage set to 37^{o}C, where a minimum of 30 minutes was given for the microscope to adjust to the increased temperature. The microscope was equipped with four laser lines including 405 nm, 488 nm sapphire laser, 560 nm saffire laser and 642 nm diode laser, specific details of laser setting used for different experiments are detailed below. Typically, a pixel dwell of 12.1/512 was used with an appropriate zoom per experiment. This microscope was used for all overexpression data, fixed cell overexpression, FRAP and 1,6 - hexanediol treatments.

2.3.1.2 Fixed cell imaging for overexpression experiments

Fixed cell imaging experiments included overexpression experiments and immunofluorescence (IF). Different experiments used different combinations of the laser lines 405 nm laser for DAPI, 488 nm for GFP and Alexa Fluor 488, 562 nm for mCherry and 642 nm for Alexa Fluor 647. Images

were taken either at the approximate central point of the Z-stack or a whole Z-stack of the entire nuclei was captured with 0.5 μm intervals.

Imaging fixed cell overexpression experiments Z-stacks were taken to capture the whole volume of the nuclei with 0.5 μm intervals for fixed cell overexpression experiments. Overexpressed constructs were visualised using the 488 nm laser line using a High Voltage (HV) of 86, no offset and a laser power of 1, DAPI was visualised using the 405 nm laser line with a HV of 100, no offset and a laser power of 1. A minimum of 3 images were captured per construct containing multiple nuclei, per biological replicate, of which there were 3. Data analysis was performed in ImageJ/Fiji and the data was visualised in R studio using R (Schindelin et al. 2012; Posit team 2023; R Core Team 2020).

Analysing fixed cell overexpression experiments A custom macro was written in ImageJ/Fiji to generate a maximum projection of every nuclei in the dataset. Then an intensity threshold was applied, where the intensity threshold was consistent within the experimental replicate to isolate the regions of interest (ROIs); in this case the condensates. The ROIs generated were saved as .roi files and the samples then underwent "analyse particles" to calculate the area, integrated density (area multiplied by the intensity of a particular condensate) and circularity of the ROI. Before running the macro, the raw images were used to determine an intensity threshold and individual Z-stacks were generated for each nuclei. The code for the custom macro can be found in appendix, section 9.1.

The data was imported into Rstudio where filtering steps and graphs were generated using R (Posit team 2023; R Core Team 2020). The initial data analysis was performed on single mutant deletion data. Initially I identified the number of nuclei which had passed the threshold for each construct, to see how many of the nuclei sampled were transfected to a high enough extent to pass threshold. From this, I then filtered the dataset. The filters determined were as follows: maximum gray value of 10 was set to exclude overexpressed nuclei; an area threshold of 0.36 μm was applied to remove regions of high intensity that were too small to represent a puncta (3 pixels), and circularity threshold of 0.25 or fewer was set to remove clumps of puncta that were not single and therefore would skew the data.

This data analysis, for the single mutant data, was performed to compare the differences in condensates produced by these deletion constructs, some of which did not form puncta. Therefore I applied a threshold based on the median number of puncta per nuclei, which determined which constructs formed condensates, compared to those that did not. A threshold of 5 puncta per nuclei

was applied, which showed that $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP did not form condensates in the same manner as the other constructs, and therefore were not included in the following data analysis to compare puncta between different deletion constructs. Samples that had a median number of puncta per nuclei above 5, were deemed phase separating constructs. They were filtered to remove any nuclei which had 5 or fewer puncta, as these nuclei were classed as non-condensate forming nuclei. After filtering, I determined the data did not fit a Gaussian distribution and therefore meant non-parametric statistical testing should be performed, which compares the median per sample compared to the mean. I therefore continued data analysis by calculating the median integrated density of puncta per nuclei, this was plotted as a violin plot and was used as our main comparison to determine whether these deletion constructs affected puncta formation compared to WT CBP. Kruskal-Wallis pairwise comparison statistical testing was performed in R. For data and a more thorough explanation of these filtering steps see section 4.4.2.

Analysing the data for the AIL deletion and KTG mutation, alongside the double mutant, experiments followed the same filtering principles described above. In this case I calculated the number of puncta per nuclei and displayed the data as a boxplot, comparing double mutant samples to the original mutation. The double mutants of $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP reduce the number of nuclei that passed filtering, as well as the number of puncta per nuclei, therefore the double mutants of these constructs also do not form puncta, and were not used in comparison of the effect of double mutants on phase separation. The remaining constructs that do form puncta CBP-GFP, $CBP_{\Delta AIL}$-GP, CBP_{KTG}-GFP, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR6:\Delta AIL}$-GFP and $CBP_{\Delta IDR6:KTG}$-GFP were then filtered to remove any nuclei that had 3 puncta or fewer. This value was based on CBP_{KTG}-GFP which had a median number of puncta per nuclei of 4. After applying a nuclei filter of 3 the data was treated as described above.

Image processing Images were typically modified in ImageJ/Fiji to change the type of image (for example from a nd2 file to a tiff file) and to include a scale bar for visual representation (Schindelin et al. 2012).

Capturing immunofluorescence images and generating line scans A Z-stack was taken for each image using 488 nm laser line to visualise endogenous CBP using HV of 90, no offset, a laser power of 2, and laser line 642 nm to visualise either BRD4 or H3K27ac using settings HV 80, no offset and a laser power of 2. To confirm the immunofluorescence was nuclear localised DAPI was imaged using the 405 nm laser line with HV 100, no offset and a laser power of 2. A minimum of 2 Z-stacks were taken containing multiple nuclei per biological replicate, of which there were 2.

To determine if CBP was colocalised with BRD4 and H3K27ac, line scans were generated in ImageJ/Fiji (Schindelin et al. 2012). To do this the nd2 file was opened and the coloured channels split into separate images. A line was drawn as a ROI and applied to both channels for CBP (488) and for BRD4/H3K27ac (642). Under the analysis tab in ImageJ/Fiji you can select "plot profile" which will produce a graph of the gray value, or intensity, of each pixel along that line. This was done for both channels and the results combined for comparison.

2.3.1.3 Live cell imaging for the optoDroplet experiment

Two laser wavelengths were used for the optoDroplet experiment 488 nm laser for the stimulation of the CRY2 domain and the 560 nm laser for capturing the mCherry for tracking puncta formation. mCherry was captured every 8 seconds with subsequent activation of the CRY2 with the 488 nm laser. The mCherry signal was captured with the 560 nm laser with a HV of 90, no offset and a laser power of 1. For dimerisation the laser settings were kept consistent at HV 0.26, offset of 0, and laser power of 0.26. Videos were taken for up to 3 minutes, with a minimum of 3 time-courses taken for each construct for each biological replicate, of which there was 3.

Analysis of optoDroplet data To gain a quantitative readout from the data collected for this experiment a custom macro was written in ImageJ/Fiji for quantifying the puncta formation ability of transfected nuclei in the mCherry channel. Briefly, the images were ran through an ImageJ plugin called Stack Registration using rigid body options, which accounts for drift in the sample (Thévenaz, Ruttimann and Unser 1998). A threshold using RenyiEntropy was applied to generate a binary mask that underwent binary processing of closing followed by opening. The parameters were then set in analyse particles, applying a size range of 50 - infinity, to generate ROIs which equated to the outline of the signal region, which was the nuclei. These ROIs were then applied to the original, raw image that underwent a multi-measure to calculate the standard deviation of the nuclei over the timecourse. Regions found in the first frame were used to monitor the change in the standard deviation over the timecourse. This raw data was collated in Google Sheets and the average fold change per construct was calculated. The fold change of each ROI was initially visualised independently to remove any nuclei which had too much drift, indicated by an erratic curve. Once outliers had been removed, the averages were generated for each biological replicate, which were collated and read into Rstudio for data visualisation (Posit team 2023; R Core Team 2020). The macro code for the optoDroplet experiment can be found in appendix, section 9.2.

Rate of change was claculated over the first 32 seconds of the time course, by generating an average fold change for each biological repeat and dividing by 32 seconds. The statistics for the rate of change were performed in GraphPad prism, where an unpaired t test was used to determine statistical significance (San Diego 2022).

Videos of optoDroplet constructs Videos of the raw data were created in ImageJ/Fiji, the raw data was cropped, converted into an RGB stack and scale bars were applied (Schindelin et al. 2012). Videos were typically saved as AVI files or GIFs with a frame rate of 4 frames per second.

2.3.1.4 Live cell imaging for FRAP

For FRAP experiments the 488 nm laser was used to visualise GFP, and perform the photobleaching. Each image was taken with the 488 nm laser with a HV of 90, no offset and a laser power of 1.44. For every video, a ROI with a Feret diameter of 0.2 μm was generated to target a particular condensate or area for photobleaching. 5 images were taken before bleaching followed by 5 cycles of bleaching using the 488 nm laser on 2 % laser power (taking 9.7 seconds). Post bleaching, images were taken every 0.5 seconds for the first 30 cycles, followed by images taken with a 1 second delay for 60 cycles, and images taken with a 3 second delay for 30 seconds. The method of changing image acquisition times was used to reduce the effect of natural photobleaching on the sample of interest over the timecourse.

Data analysis of FRAP data The raw data was loaded into ImageJ/Fiji and processed using Stack Registration to remove any drift from the sample, in attempt to keep the bleached area in the same position over the timecourse (Schindelin et al. 2012; Thévenaz, Ruttimann and Unser 1998). The intensity was calculated over the timecourse, using a ROI that covered the bleached area for each construct, ROI size was the same for each experimental replicate. Natural photobleaching over the timecourse, due to imaging, was also monitored using a ROI from another region within the nucleus, that was not near the bleaching region. Using Google Sheets, the photobleached intensity was normalised for natural photobleaching and a percentage recovery was generated by normalising to the pre-bleach intensity for each nuclei. The normalised data was imported into Rstudio for data visualisation (Posit team 2023; R Core Team 2020).

Rate constants such as time (T) = 1/2 and the percentage mobile fraction were calculated in GraphPad Prism 9 (San Diego 2022). A graph was generated using a non-linear regression model with one phase association with a variance weighting on $1/Y^2$. The mobile fraction was

calculated from the fluorescence plateau and the first post-bleach intensity of the fitted curve. The equation used was as follows: mobile fraction = ($\text{Intensity}_{\text{plateau}}$- $\text{Intensity}_{\text{post-bleach}}$)/(1 - $\text{Intensity}_{\text{post-bleach}}$). A diagrammatic representation of the quantified parameters can be found in Figure 2.1.

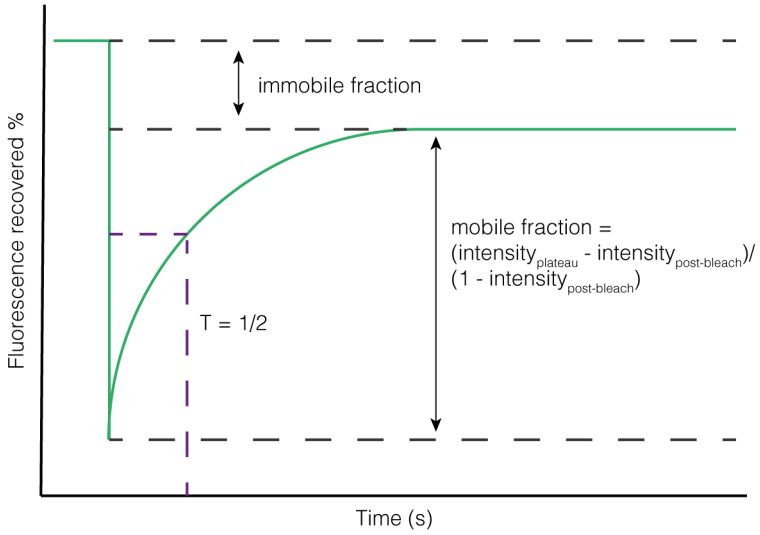

Figure 2.1: FRAP curve to demonstrate quantification of dataset
Example FRAP curve to demonstrate how quantification was calculated. Green line represents the fluorescence recovery after photobleaching (FRAP) curve. Purple dotted line represents the value taken to calculate the T = 1/2, which is the time it takes for half of the fluorescence to be recovered to the plateau. Grey lines represent the immobile fraction from pre-bleach to the plateau, top, and the values used to calculate the mobile fraction the post-bleach fluorescence and plateau value, with the equation for mobile fraction shown.

2.3.1.5 Live cell imaging for 1,6 - hexanediol treatment

For 1,6-hexanediol (1,6-HD) treatment, media was changed to 1 ml of imaging media prior to imaging. The 488 nm laser was used to visualise GFP with a HV of 90, no offset and a laser power of 1. Images were taken every 1.9 seconds for a total of 4 minutes. 1 minute after starting the timecourse 1 ml of 10 % 1,6-HD (Merck) that was diluted in sterile Milli Q H_2O, was added to reach a final concentration of 5 %, and imaging was continued for a further 3 minutes. Videos were compiled in ImageJ/Fiji to show the effect of adding 1,6-hexanediol on condensate formation (Schindelin et al. 2012).

2.3.1.6 Optimising smFRET and tmFRET protein expression using GFP control plasmid

When testing the expression of GFP for both the smFRET and tmFRET system, images were taken using the 60 X oil immersion objective. Live cells were imaged using the 488 nm laser to image GFP with a HV of 90, offset of 0 and laser power of 1. The 405 nm laser to image the nuclei via Hoechst 33342 (Fisher) with a HV of 60, offset of 0 and laser power of 1. Still images were taken at the mid-plain of a z-stack for each condition, with at least 3 images taken per condition. Images were contrast adjusted in ImageJ/Fiji to match the contrast of the negative control, which did not contain the unnatural amino acid, and the scale bar was added (Schneider, Rasband and Eliceiri 2012; Schindelin et al. 2012).

2.3.2 Nikon W1 spinning disc microscope

2.3.2.1 Microscope details

Nikon W1 spinning disc confocal microscope was equipped with a humidified temperature chamber set to 37^{o}C, and for this doctoral work a 100 X oil immersion objective (NA 1.45) was utilised, with the 50 μm pinhole wheel. This microscope was equipped with 405 nm, 488 nm, 514 nm, 560 nm and 647 nm laser lines, with a subset being used for imaging below.

2.3.2.2 CBP-HaloTag imaging

For CBP-HaloTag imaging the 405 nm laser was used to image the nuclei and 514 nm laser was used to image the TMR (Promega) labelling. The 405 nm laser was set with a laser power of 45 % with 100 ms exposure, and the 514 nm laser was set with 60 % laser power with a 300 ms exposure. Images were taken every 500 ms to account for exposure times, in the absence of binning. Videos ran for a maximum of 3 minutes, and underwent immediate processing of the Nikon denoise package. Videos were further optimised in ImageJ/Fiji and exported as videos either as AVI files or GIF with a frame rate of 4 frames per second. Captured images were further processed in ImageJ/Fiji where scale bars were added (Schindelin et al. 2012; Schneider, Rasband and Eliceiri 2012).

2.3.2.3 Optimising system for smFRET with CBP-Halotag construct

In the optimisation of smFRET protein expression CBP-HaloTag was visualised using the Nikon W1 spinning disc microscope. Images were taken using the 100 X oil immersion objective using

the 514 nm laser line with a 200 ms exposure with 40 % laser power. Images were taken at the mid-plain of the Z-stack to image in the middle of the nucleus, 3 representative images were taken for each condition. Images were initially processed using the Nikon denoise package and contrast was further adjusted in ImageJ/Fiji, adding scale bars to each image (Schindelin et al. 2012; Schneider, Rasband and Eliceiri 2012).

2.3.2.4 Imaging doxycycline induced cells

For Dox induction, imaging media was supplemented with Dox at the desired concentrations (0 ng/ml - 100 ng/ml). Laser lines 405 nm and 488 nm were used to image the Hoechst 33342 (Fisher) and therefore the nuclei, and GFP expression respectively. The 405 nm laser was set with a 100 ms exposure with 60 % laser power, and the 488 nm laser was set with 300 ms exposure time and 60 % laser power. Images were processed in ImageJ/Fiji.

2.3.3 TIRF microscope BONSAI

2.3.3.1 Microscope details

This microscope is a TIRF/FRAP imaging system (BONSAI) which was custom-built by Cairn for single-molecule scanning microscopy. This microscope is owned by Dr Alison Twelvetrees from SITraN. Specific microscope details are included below.

This microscope used a Ti-2 inverted microscope base (Nikon), with ASI automated XY stage and pieco-Z, MonoLED brightfield illuminator (Cairn). The TIRF/FRAP unit iLas scanning had Cairn dual-collimation optics, with Cairn multi-line laser bank (containing 405, 488, 532, 561 and 637 laser lines) and a Photometrics Prime 95B sCMOS camera. Specific details of laser power and exposure time are detailed for each specific subsection, all using the TRF89901-EMv2 filter set from Chroma for the TIRF Quad cube: Excitation, ZET405/488/561/640x-V2; Dichroic, ZT405/488/561/640rpc-UF2-V2 and Emission filter ZET405/488/561/640m-V2. Further emission clean up was performed with Cairn Optospin filterwheel with the following filter and laser pairs: ET445/58m for 405 nm, ET525/50m for 488 nm, ET595/44m for 561 nm and ET706/95m filters for 638 nm laser.

Images were taken using the Nikon CFI Apochromat TIRF 100 X oil objective with N.A. 1.49 and W.D. 0.12 mm, on the MetaMorph acquisition software. Selected laser lines were controlled using Modular software, where different laser lines were used based on different fluorophores, which will be detailed below. Maximum power of each laser was: 75 mW for 405, 61 mW for 488, 53 mW for 561, and 98 mW for 638. Details for laser power and exposure time are included below.

2.3.3.2 CBP-HaloTag imaging for FRAP

Endogenous CBP-HaloTag was labelled with JF 549 (Promega) as was described in section 2.2.15, therefore imaging was performed using the 561 nm laser line with exposure time of 200 ms at 10 % laser power. ROIs were drawn with a diameter of 10 to mark where the lasers will target, with several regions being targeted at one time. FRAP settings included 10 repetitions using the 405 nm laser line of weight 2 on 3 % laser power. A timecourse was taken with the following pattern: 15 time points with a 500 ms delay, followed by FRAP, followed by 60 time points with a 500 ms delay and 30 time points with a 2 second delay. Images were processed in ImageJ/Fiji where contrast was adjusted to visualise puncta, scale bars were added and false colour applied (Schindelin et al. 2012; Schneider, Rasband and Eliceiri 2012). Images were compiled in Adobe Illustrator.

Data analysis of endogenous FRAP data Data was analysed as described in subsection 2.3.1.4.

2.3.3.3 Imaging endogenous CBP-HaloTag transfected with CBP-GFP and main mutants

4 colour images were taken of transfected constructs before labelling of CBP-HaloTag to determine bleed-through of overexpressed constructs into the 646 channel prior to colocalisation experiments. Images were taken using laser lines 405 nm, 488 nm, 532 nm and 637 nm, with an exposure time of 200 ms using 10 % laser power. A 4 colour control was also taken with CBP-HaloTag labelled with JF 646 and no overexpression, for this all laser lines used the same conditions as before except the 637 nm laser had a power of 80 % to visualise the endogenous protein.

For colocalisation experiments, 2 colour multidimensional images were captured using the 488 nm laser with 10 % laser power and a 200 ms exposure to visualise GFP tagged constructs, and 637 nm laser with 80 % laser power and 200 ms exposure to view endogenously tagged HaloTag. Four images were taken per construct and further analysed in ImageJ/Fiji (Schindelin et al. 2012). A representative zoomed in region was amplified for each construct, where the background was subtracted using a rolling ball of 100 pixels. Individual channels were saved in grayscale and merged images were generated where the GFP tagged constructs were shown in green and the endogenously tagged puncta was shown in magenta. Scale bars were added for context of size showing 10 μm (Schindelin et al. 2012; Schneider, Rasband and Eliceiri 2012).

2.3.3.4 Imaging CBP-HaloTag clonal line

To confirm the clonal nature of CBP-HaloTag images were taken using the 561 nm laser of CBP-HaloTag and labelled HEK293T cells as a control. Laser power of 50 % was used with a 500 ms exposure time, where brightfield images were captured to indicate the presence of cells in the absence of using Hoechst. Images were processed in Imagej/Fiji where a rolling ball of 100 pixels was used to remove any background, and contrast was adjusted to gray values 0 - 1056, which was the range for the CBP-HaloTag clonal line (Schindelin et al. 2012).

2.3.3.5 CBP-HaloTag optimisation for single molecules imaging

Testing photoactivateable 646 for imaging single molecules of CBP-HaloTag For test experiments CBP-HaloTag polyclonal cells were labelled with 50 nM JF 549 and 50 nM photoactivateable ligand. Imaging using the 532 nm laser was done to visualise nuclei that had CBP tagged with the HaloTag using 10 % laser power with an exposure time of 100 ms. Activation of the photoactivateable ligand was done using the 405 nm laser in a similar manner as was performed for the FRAP. ROIs were drawn with a diameter of 10 to mark where the laser was targeted, two regions were drawn within nuclei, while two regions were randomly selected to monitor off-target effects. The 405 nm laser was set with 3 % laser power, for 3 repetitions, with a thickness of 1. An initial representative image was taken of the nuclei using the 532 nm laser to visualise JF 549, and a stream timecourse was taken using the 637 laser line to visualise the photoactivateable ligand with an exposure time of 100 ms. Images were processed in ImageJ/Fiji where contrast was adjusted to optimise puncta visualisation. ROIs were drawn around the nuclei in the 549 channel and applied to the 646 channel so that we could observe where the nuclei should have been, and scale bars were applied (Schindelin et al. 2012). Images were compiled in Adobe Illustrator.

Dual labelling CBP-HaloTag for single molecule imaging CBP-HaloTag cells were labelled with JF 549 and/or 646. For 4 colour control images, where only one Janelia Fluor was used to label the cells, a representative image was taken using laser lines 405 nm, 488 nm, 532 nm and 637 nm, with a laser power for each at 10 % and exposure time of 100 ms. Scale bars were applied using ImageJ/Fiji (Schindelin et al. 2012).

For dual colour imaging a rectangle ROI was drawn around the nuclei of interest, this determined the imaging region, and increased the speed that each frame could be taken on BONSAI. The 532 nm laser was used at a laser power of 10 % with 100 ms exposure to visualise whole nucleus CBP-HaloTag. For single molecules the 637 nm laser line was used, when testing a concentration

of 0.5 nM JF 646 a laser power of 50 % was used with a 100 ms exposure, compared to 5 nM where 80 % laser power was used with a 100 ms exposure. Upon optimising for single molecules 100 % laser power was used at a range of different exposure times including 20 ms, 50 ms, 100 ms and 200 ms.

Tracking for single molecule imaging Video of CBP-HaloTag with exposure time of 20 ms was used for optimisation of tracking single particles. The video was imported into ImageJ/Fiji and the contrast adjusted to a range of 108 - 117 to reduce the background and optimise detection of single particles (Schindelin et al. 2012). Two softwares were tried within the context of this doctoral work including TrackMate and DoM (Katrukha 2020; Tinevez et al. 2017).

TrackMate software was used initially as a user friendly method to detect and track particles (Tinevez et al. 2017). Two detectors were trialled here: Difference of Gaussian (DoG) particle detector, the quickest particle detector for smaller spot sizes, or Laplacian of Gaussian (LoG) particle detector. For both methods an estimated object diameter of 7 pixels was applied with a 0.2 quality threshold, which was preprocessed with median filter and sub-pixel localisation. A second quality threshold was then applied with a 0.3 cut off before tracking the particles using the LAP tracker. Tracking parameters included a frame to frame linking maximum distance of 8 pixels, with track segment gap closing with a maximum distance of 15 pixels with a 5 frame maximum gap. These parameters generated the tracks that were then used to generate plots including distance displacement for each track and confinement ratio of each track.

Alternatively, particle tracking using the Detection of Molecules (DoM) software was also attempted (Katrukha 2020). Detection of molecules and fit was selected with the following parameters: PSF standard deviation of 2, intensity threshold of 2, particle size of 660 nm, parallel threads number of 500 with 5 fitting iterations. Detected molecules were marked in an overlay and an excel spreadsheet was produced displaying features such as particle coordinates. The following could then be linked to tracks in the DoM software which linked all particles, with a maximum searching distance of 8 pixels with a maximum gap of 5 frames. Tracks were then plotted on top of the video overlay and a table of tracks was generated. This table could be used in the future for manual processing, however due to time constraints this was not performed, and the tracks were not shown in this data analysis.

2.3.4 Calculation of percentage phenotype

For several representative images shown throughout this doctoral work, a percentage phenotype was calculated. This is the percentage of either transfected or labelled, in the case of CBP-HaloTag cells, nuclei which display the phenotype that is represented in the image. This was calculated manually through counting the nuclei that display signal, and those that show the given phenotype, from which a percentage was calculated. The percentage phenotype is overlayed on the represented images shown throughout this thesis.

2.4 Western blotting

2.4.1 Buffers for western blotting

RIPA lysis buffer: 50 mM Tris HCl (Sigma) pH 8 at 4^{o}C, 100 mM NaCl (Melford), 2 mM $MgCl_2$ (Fisher), 1% Triton X-100 (Sigma), 0.1% Sodium deoxycholate (Sigma), 0.1 % Sodium Dodecyl Sulfate (SDS, Sigma) made up to the required volume with Milli Q water. On the day of running the protocol, 1 mM Dithiothreitol (DTT, Melford), 1 X Halt phosphatase inhibitor cocktail (Fisher), 10 mM sodium butyrate (Sigma) was added to the buffer. Benzonase 500 U/100 µl (Insight Biotech) or DNAse RQ1 (Promega), 1µl per pellet, was added to the pellet during cell lysis.

1X MOPS running buffer (1 Litre): 50 ml of 20X NuPAGE^{TM} 3-(N-morpholino) propanesulphonic acid (MOPs, Fisher), diluted into 950 ml of dH_2O.

1X Tris acetate SDS running buffer (1 Litre): 50 ml of 20X NuPAGE^{TM} Tris-Acetate SDS Running Buffer (ThermoFisher) was diluted into 950 ml of dH_2O.

10X TBS (1 Litre): Tris buffered saline (TBS) was made up of 24 g of Tris, 88 g of NaCl at a pH of 7.6 made up to a final volume of 1 litre.

1X TBS-T (1 Litre): 100 ml of 10X TBS was added to 900 ml of dH_2O. 1 ml of Tween 20 was added to give a final concentration of 0.1 %.

Milk Blocking solution 40 ml: 2 g of dried skimmed milk (either SLS or Marvel) was dissolved in 40 ml of TBS-T.

BSA Blocking solution 40 ml: 2 g Bovine Serum Albumin (Sigma) diluted in 40 ml TBS-T.

2.4.2 Antibodies

Antibodies used for western blotting and concentrations used can be found in Table 2.15.

Antibody	Source	Cat. No.	Species	Type	Dilution
α-CBP (D6C5)	Cell Signalling	7389	Rabbit	Monoclonal	1:1000
α-GFP	Abcam	ab290	Rabbit	Polyclonal	1:1000
α-GFP	Millipore	MAB3580	Mouse	Monoclonal	1:2500
α-GAPDH	Proteintech	60004-1	Mouse	Monoclonal	1:10,000
α-HaloTag	Promega	G9211	Mouse	Monoclonal	1:1000
HRP conjugated α-Rabbit IgG	Stratech	211-032-171	Goat	IgG light chain only	1:20,000
HRP conjugated α-Mouse IgG	Proteintech	SA00001-1	Goat	Anti-Mouse IgG heavy and light chain	1:10,000

Table 2.15: Antibodies used for western blotting throughout doctoral work

2.4.3 Cell lysis and Bicinchoninic Acid protein assay

Cells were harvested by washing in PBS for 1 minute, followed by a wash in 0.25% 1X Trypsin before incubation at 37^{o}C for 4 minutes. The cells were resuspended in 5 ml of DPBS and pelleted by centrifugation at 400 rcf for 5 minutes. The supernatant was subsequently removed and the pellet resuspended in 1 ml DPBS and transferred to a LoBind polypropylene microcentrifuge tube (Fisher Scientific). Cells were centrifuged at 400 rcf for 5 minutes, before the supernatant was removed and the cell pellet was washed with 1 ml of DPBS before centrifuging for a further 5 minutes at 400 rcf. The supernatant was once again removed and the pellet was either used directly for cell lysis or flash frozen on dry ice before storage at −80^{o}C.

On the day of the western blot, if the pellet was frozen it was thawed at room temperature, and 1 µl of DNase or Benzonase was added to either the cell pellet directly or to 50 µl of RIPA buffer. The pellet was resuspended in RIPA buffer and incubated at 37^{o}C for 20 minutes before centrifugation at 20,000 rcf for 10 minutes at 4^{o}C to remove any cellular debris. The supernatant

was transferred into a clean LoBind microcentrifuge tube and the total protein concentration was determined using Pierce Bicinchoninic Acid (BCA) Protein Assay Kit (Thermo Fisher) according to the manufacturer's instruction.

2.4.4 Gel electrophoresis

A desired concentration of total protein lysate, typically between 25 - 100 µg, was prepared with the addition of Invitrogen Novex NuPAGE LDS Sample Buffer (Fisher Scientific) and NuPAGE Sample Reducing Agent (Fisher) before denaturation of the protein at 70^oC for 10 minutes. The sample was left at room temperature while the XCell SureLock Mini-Cell Electrophoresis System (Thermo Fisher) gel tank was assembled. Different gels were used depending on the purpose of the western blot; NuPAGE^{TM}4-12% Bis-Tris Protein Gels, 1.5 mm 10 - well (ThermoFisher) were used with 1X MOPs running buffer or NuPAGE^{TM} Tris-acetate gels 3 - 8 %, 1.5 mm 10-wells (ThermoFisher) were used with 1X Tris-Acetate SDS running buffer. 500 µl of Invitrogen Novex NuPAGE^{TM} Antioxidant (Thermo Fisher) was added to the buffer reservoir in between the gel/s. Samples were loaded at approximate volumes of 30 µl, with 5 µl of sample buffer being loaded into any empty well and 5 - 10 µl of SeeBlue Plus2 (ThermoFisher) pre-stained protein standard loaded on the gel. Gels were ran at a voltage of 90 - 120 volts for a range of 90 minutes to 130 minutes.

2.4.5 Protein transfer

Protein transfer was completed using Trans-Blot Turbo Transfer System (BioRad) using a semi-dry approach with Trans-Blot Turbo Mini 0.2 µm Nitrocellulose Transfer Packs (BioRad, 1704158). Transfer was completed with the following settings, 25 Volts, 1.3 Amp for 25 minutes.

2.4.6 Membrane probing and imaging

After the transfer, the membrane was transferred to suitable container and washed twice in TBS-T before being blocked in milk blocking solution for a minimum of 40 minutes. Primary antibody was prepared in blocking solution, typically milk, or for Abcam ab290 α-GFP in BSA blocking solution, and incubated overnight or in some cases over the weekend at 4^oC on a shaking platform. After incubation, the primary antibody in blocking solution was typically saved and sodium azide at a final concentration of 0.02 % was added to preserve the solution which was stored at 4^oC and used for a maximum of 10 times. The membrane was subsequently washed three times in TBS-T, with each wash being incubated for 5 minutes on a shaking platform at room temperature.

Horseradish peroxidase (HRP)-secondary antibody was then prepared in milk blocking solution and added to the membrane, which was incubated at room temperature on the shaking platform for 30 minutes for α-Rabbit and 60 minutes for α-Mouse, before 3 further washes in TBS-T as described previously. SuperSignal^{TM} West Pico Chemiluminescent Substrate (ThermoFisher) was used as a HRP substrate and chemiluminescence was visualised using G:BOX Chemi-XRQ imager (Syngene).

2.4.7 Stripping the membrane

If the membrane was to be reprobed after imaging, the membrane was placed back in TBS-T and incubated for 10 minutes on a shaking platform. The TBS-T was removed and approximately 10 ml of Restore PLUS Western Blot Stripping Buffer (Fisher) was added to the blot, which was incubated at room temperature on a shaking platform for 20 minutes. The blot was then washed twice in dH_2O before a 10 minute incubation in TBS-T on the shaking platform. The membrane was once again blocked and the rest of the protocol was followed as described in section 2.4.6.

2.5 Chromatin immunoprecipitation sequencing

2.5.1 Buffers for ChIP-Seq

The following buffers were filtered and stored at 4^oC, with the addition of 1 X Halt protease inhibitors on the day of use, unless otherwise stated. Buffers were made by Dr Petra Celadova to a final volume of 250 ml.

ChIP lysis buffer 1: 50 mM HEPES-KOH (Simga), pH 7.5 140 mM NaCl, 1 mM EDTA, 10 % Glycerol, 0.5 % NP40, 0.25 % Triton X - 100.

ChIP lysis buffer 2: 10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA (Sigma).

ChIP lysis buffer 3: 10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1 % Na-deoxycholate, 0.5 % N-lauroylsarcosine (Sigma).

ChIP IP buffer: 10 mM Tris-HCl pH 8.0, 600 mM NaCl, 1 mM EDTA, 3 % Triton X-100.

ChIP wash buffer: 50 mM HEPES-KOH pH 7.5, 500 mM LiCl, 1 mM EDTA, 1 % NP40, 0.7 % Na-deoxycholate, 0.1 % N-lauroylsarcosine.

ChIP final wash buffer: 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 50 mM NaCl.

ChIP elution buffer: 50 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM EDTA, 1 % SDS. Buffer was filtered and stored at room temperature, without the addition of protease inhibitors.

10 X TE buffer 100 mM Tris-HCl pH 8.0, 10 mM EDTA.

2.5.2 Chromatin immunoprecipitation

2.5.2.1 Cross-linking Dox-inducible cell lines

Cells plated for chromatin immunoprecipitation (ChIP) were grown for 72-hours, with Dox induction for 48-hours as described in section 2.2.20. Media was gently aspirated from each of the plates and replaced with 5 ml of DPBS. For the plates that were cross-linked, 343.5 µl of 16 % fresh formaldehyde was added to the 5 ml on each of the respective dishes and incubated on a rotary shaker at room temperature for 10 minutes. After incubation, 250 µl of 2.5 M Glycine (Melford) was added to each of the plates to quench the crosslinking reaction, this was incubated on the rotator for 5 minutes at room temperature. The cells were then harvested by scraping, and spun at 200 rcf for 5 minutes. The supernatant was discarded and the resulting pellet was resuspended in 1 ml of ChIP lysis buffer 1 (5 cell pellet volumes) to begin cell lysis.

2.5.2.2 Cell lysis

Cells were incubated on a rotator at 4^oC for 5 minutes before cells were centrifuged at 1500 rcf for 5 minutes before discarding the supernatant. The nuclear pellet was then resuspended in 1 ml of ChIP lysis buffer 2 (5 cell pellet volumes), this caused the nucleus to swell to aid in cell lysis. The samples were then incubated on a rotator at room temperature for 10 minutes. Samples were then centrifuged at 1500 rcf for 5 minutes before the supernatant was discarded. Each pellet was then resuspended in 180 µl of ChIP lysis buffer 3 and transferred to a polycarbonate tube for sonication.

2.5.2.3 Sonication of chromatin

Sonication was performed using the Bioruptor Pico (Diagenode). To determine how long to sonicate HEK293T cells for to get optimal shearing of 150 - 300 bp fragments needed for sequencing, I performed a test experiment where I sonicated HEK293T cells for 0, 5, 8, 10, 12, and 15 cycles of 30 seconds on, 30 seconds off pulse sonication. 50 µl samples were taken at each cycle to be tested and replaced with 50 µl of ChIP lysis buffer 3. The resulting samples were loaded on a 1 % agarose gel which showed the optimal number of cycles for shearing of HEK293T cells was 12 cycles, this was used in subsequent ChIP experiments.

Samples were sonicated for 12 cycles of 30 seconds on, 30 seconds off pulse sonication. Once complete samples were briefly centrifuged to collect any samples from the top of the lid and then transferred to a pre-chilled low adhesion microcentrifuge tube. The cells were centrifuged at 20,000 rcf for 15 minutes. The supernatant was then transferred to a pre-chilled microcentrifuge tube, leaving any pelleted debris behind. The resulting concentration of overall protein was calculated using BCA assay (ThermoFisher) according to the manufacturers instructions.

2.5.2.4 Immunoprecipitation (IP)

After protein concentration was determined for each sample, immunoprecipitation (IP) reactions were set up with 900 µg of protein in a total of 300 µl, made with ChIP IP buffer. Input reactions were also made for each sample with 90 µg of protein in 50 µl of ChIP IP buffer. 4 µg of antibody was then added to the IP reactions, where α-CBP (D6C5) (Cell Signalling, 7389) was added to the HEK293T-Tet3G control cells of the maternal line, and α-GFP (Abcam, a290) was added to HEK293T-Tet3G-TRE3G-CBP-GFP and mutant constructs to pull down overexpressed protein and not endogenous CBP. IP samples were incubated overnight on a rotator at 4^oC.

Protein G Dynabeads were prepared, where 30 µl beads were used per sample, and aliquots were made in pre-chilled 1.5 ml low adhesion microcentrifuge tube. The beads were placed on a magnetic rack and supernatant was removed, followed by 3 washes of 1 ml 1 % BSA (wt/vol) in PBS. Beads were resuspended for a final time in 1 ml 1 % BSA and incubated overnight at 4^oC with rotation to block the beads and prevent unspecific binding of proteins during elution.

After incubation the microcentrifuge tubes were placed on a magnetic stand and the blocking solution was removed. Beads were resuspended in 1 ml of ChIP IP buffer and washed for a total of 3 washes. On the final wash the supernatant was discarded and the beads resuspended in the original bead volume of 30 µl of ChIP IP buffer. Microcentrifuge tubes containing the IP reaction were briefly centrifuged at 100 rcf at 4^oC to remove any sample from the lid of the tube, before

30 µl of beads were added and incubated for 90 minutes at 4^oC with rotation. Fresh low adhesion microcentrifuge tubes were prepared for each sample and incubated on ice for 15 minutes before use. The IP reaction was spun at 100 rcf at 4^oC briefly to collect any sample from the lid of the tube before being transferred to the fresh microcentrifuge tube. The original tube was centrifuged a final time to collect any residual sample and this was transferred to the new tube using a p10 pipette.

2.5.2.5 Elution

Samples were placed on a magnetic rack and supernatant was removed and discarded, beads were resuspended in 500 µl of ChIP wash buffer. The process was repeated a total of 4 times to wash away any non-specific interactions, taking care not to let the beads dry out. On the fifth wash the final resuspended sample was transferred to a new pre-chilled low adhesion microcentrifuge tube. The original tubes were centrifuged at 100 rcf at 4^oC to collect any remaining sample, this was combined with the rest of the sample using a p10 pipette. The samples were placed back on the magnetic rack and the supernatant was removed. The beads were resuspended in 1 ml of ChIP final wash buffer, before placing back on the rack and the supernatant being removed and discarded. Samples were centrifuged for 2.5 minutes at 400 rcf at 4^oC, the tubes were placed back on the rack and any remaining buffer was removed using a p10 pipette.

To elute the samples the beads were resuspended in 200 µl elution buffer and incubated for 30 minutes at 65^oC in a thermomixer containing a heated lid, with rotation of 900 rpm to keep the beads in suspension. The samples were centrifuged at 100 rcf for 1 minute to collect all liquid in the bottom of the tube and the samples were placed on the magnetic rack to collect the beads. The supernatant was then transferred to a new 1.5 ml microcentrifuge tube, this contained the IP. The original tubes were centrifuged at 2,500 rcf for 1 minute, and any remaining supernatant was removed and added to the rest of the sample using a p10 pipette.

2.5.2.6 Cross-link reversal

150 µl of elution buffer was then added to the input samples, reaching a final volume of 200 µl so that all samples have the same volume. To reverse the cross-link samples were incubated in a thermomixer at 65^oC overnight, for a minimum of 12 hours and a maximum of 18 hours to reduce the background observed.

2.5.2.7 DNA purification

To remove any RNA from the sample prior to DNA purification, 1X TE buffer was made, supplemented with 0.2 mg/ml RNaseA and 4 mM $CaCl_2$. 200 µl of this buffer was added per microcentrifuge tube, to a total of 400 µl, and incubated at 37^{o}C for 2 hours. Proteins were then removed from the samples with 0.2 mg/ml of Proteinase K, which works most efficiently with the addition of 4 mM $CaCl_2$. The reaction was incubated at 55^{o}C for 2 hours.

DNA purification was performed using phenol chloroform extraction. 400 µl of phenol chloroform (pH 8.0) was added to each sample, before samples were vortexed for 15 seconds followed by cenrifugation for 10 minuted at 20,000 rcf. Following this, the solution will have formed three layers, the upper aqueous layer containing DNA, the interphase and the lower organic fraction. The aqueous layer was transferred to a fresh microcentrifuge tube, making sure not to carry over the other layers. To each tube 50 µl of 3 M sodium acetate (Sigma) was added, to encourage precipitation, and 1 µl of glycogen, which will co-precipitate with the DNA allowing visualisation of the pellet in the tube. 2.5 volumes of cold 100 % ethanol was added to each tube and mixed, the samples were transferred to the -80^{o}C and incubated for a minimum of 1 hour to allow sufficient precipitation.

Following the incubation all centrifugation steps were performed at 20,000 rcf at 4^{o}C. Samples were centrifuged for 15 minutes, before supernatant was removed and discarded. The pellets were washed with 500 µl 70 % ethanol, centrifuged for 2 minutes before the ethanol was discarded. The pellets were washed with 500 µl 80 % ethanol, centrifuged for 2 minutes before the ethanol was discarded. The pellets were centrifuged for a further 2 minutes, and any remaining ethanol was removed using a p10 pipette. The pellets were briefly left to dry at room temperature to remove any excess ethanol, before resuspending in 30 µl of 1X TE buffer. Samples were stored at -20^{o}C until starting the library preparation.

2.5.3 Library preparation for ChIP-Seq

For ChIP-seq, samples were prepared using NEBNext Ultra^{TM}II DNA Library Prep Kit for Illumina (NEB, E7645) following manufacturers instructions.

Briefly, fragmented DNA underwent 5' phosphorylation and dA-Tailing to prepare the fragment ends for adapter ligation. Sample concentration was checked using Qubit dsDNA High Sensitivity (HS) kit (Thermo Fisher), to determine adapter dilutions. For input samples the adapters were not diluted, however a 1:25 dilution of adapters in 10 mM Tris-HCl pH 7.5, was performed for immunoprecipitation samples. A cleanup of the adapter ligation was performed using AMPure

XP beads (Beckman Coulter) for size selection following the 200 bp protocol. PCR amplification was performed with 3 cycles of amplification for input samples, and 13 cycles of amplification for immunoprecipitation samples. A cleanup of the PCR reaction was performed using AMPure XP beads, following protocol guidelines. Fragment sizes were checked before sequencing using the TapeStation (Agilent), using High Sensitivity DNA ScreenTape (Agilent). Paired-end (PE150) sequencing of libraries was performed by Novogene, using the Illumina NovaSeq6000 platform, generating 9 GB of raw data.

2.5.4 Data analysis of ChIP-Seq

ChIP-Seq data analysis was performed by Dr Mark Dunning from Sheffield Bioinformatics Core, where raw reads were aligned to the hg38 reference genome using Burrows-Wheeler Alignment (BWA) tool (Li and Durbin 2009), and peaks were called using Model-based Analysis of ChIP-Seq (MACS2) python package (Zhang et al. 2008; Di Tommaso et al. 2017), as part of the Nextflow ChIP-Seq pipeline. Further downstream analysis was performed by Dr Dan Bose.

Chapter 3

Phase separating properties of wild type CBP

3.1 Introduction

3.1.1 CBP, enhancers and condensates

CREB binding protein (CBP) is a transcriptional coactivator protein that is ubiquitously found at enhancer regions (Bose et al. 2017; Holmqvist and Mannervik 2013; Kim et al. 2010). It aids in enhancer function, by binding multiple transcription factors and RNA molecules bringing them to the enhancer, and increasing transcription (Cheng et al. 2015; Ding et al. 2018; Plank and Dean 2014; Shrinivas et al. 2019; Zhang et al. 2019). CBP has histone acetyltransferase activity, it can acetylate both transcription factors and chromatin, where acetylated histone markers such as H3K27ac and H3K122ac are representative of euchromatin, where transcription occurs (Bannister and Kouzarides 1996; Bose et al. 2017; Creyghton et al. 2010; Kalkhoven et al. 2003; Struhl 1998; Tie et al. 2016; Toth et al. 2013; Wang, Marshall and Ikura 2013; Jin et al. 2011).

Recently, phase separation has been proposed as the mechanism of how enhancers interact with their target promoter; proteins typically found at enhancers, such as MED1 and BRD4, have been shown to phase separate, and that this behaviour can influence their catalytic activity (Han et al. 2020; Hsieh et al. 2014; Sabari et al. 2018; Wang et al. 2022b; Shao et al. 2022; Shrinivas et al. 2019; Gurumurthy et al. 2019). Proteins that phase separate have been shown to have intriniscally disordered regions (IDRs), which are regions that have little secondary or tertiary structure, allowing the formation of weak interactions which facilitate the formation of phase

separated condensates (Larson et al. 2017; Lin, Currie and Rosen 2017; Nair et al. 2019; Plys and Kingston 2018; Sabari et al. 2018; Taylor et al. 2019; Mittag and Pappu 2022; Musacchio 2022; Haag 2015; Chong et al. 2018; Nott, Craggs and Baldwin 2016; Pak et al. 2016; Molliex et al. 2015; Shin et al. 2018).

CBP is found at enhancer regions and is over 60 % disordered, therefore I believe that it can form phase separated condensates, and that this behaviour may influence its catalytic activity (Dyson and Wright 2016). Using disorder prediction software, such as DISOPRED3, CBP is predicted to be largely disordered and that this disorder is distributed throughout the whole length of CBP (Figure 1.16) (Jones and Cozzetto 2015). Intrinsic disorder within CBP tends not to overlap with characterized domains such as the HAT domain or KIX, which are highly structured in order to function (Bannister and Kouzarides 1996; Chan and La Thangue 2001; Delvecchio et al. 2013; De Guzman et al. 2006; Wang et al. 2012). The overall structure of CBP can be predicted by AlphaFold, an artificial intelligence system used to predict 3D protein structure, showing structured domains such as the HAT domain surrounded by intrinsic disorder regions, which dominate the N- and C-terminus (Figure 3.1), AlphaFold prediction was performed by Dr Archna Shah (Jumper et al. 2021). Using phase separation prediction tool DeePhase, CBP is predicted a LLPS propensity score of 0.9, further indicating that CBP may form phase separated condensates (Saar et al. 2021).

A paralog of CBP, p300, has been shown to phase separate, with condensates having been shown to influence HAT activity (Ma et al. 2021; Zhang et al. 2021). p300 has been shown to be catalytically active within a condensate, suggesting a link between phase separation and function (Ma et al. 2021). CBP has also been shown to colocalise with PML bodies, which are nuclear bodies involved in transcriptional regulation, and in recent years has been suggested to form phase separated condensates in cells (LaMorte et al. 1998; Banani et al. 2016; Boisvert et al. 2001; Wang et al. 2022b; Ma et al. 2021; Zhang et al. 2021). These condensates are typically fluid-like, with puncta fusing together and diffusing apart, indicating phase separation (Boeynaems et al. 2018; Gibson et al. 2019b; Lin, Currie and Rosen 2017; Plys and Kingston 2018; Peng and Weber 2019)

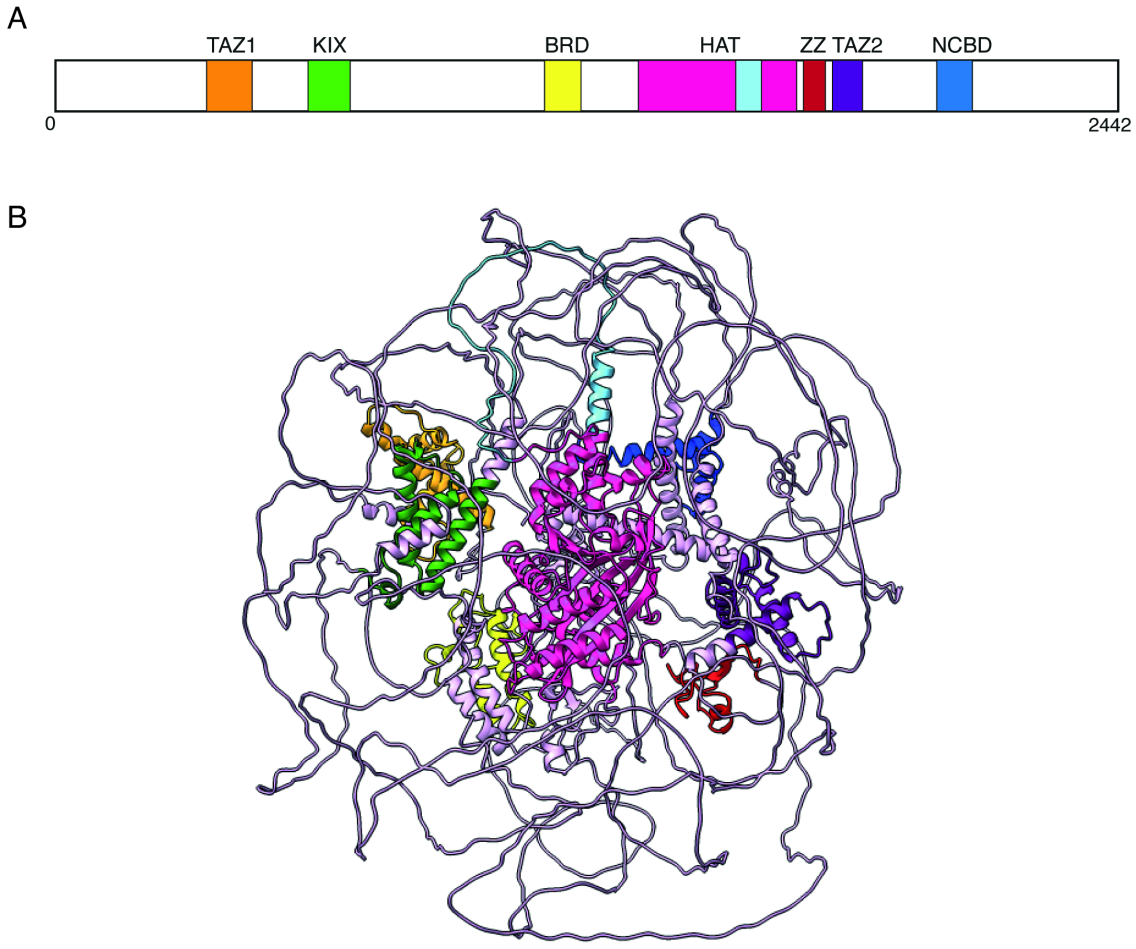

Figure 3.1: Structural prediction of CBP
Structural prediction of CBP to indicate intrinsic disorder shown over the full length of CBP. A - domain diagram of CBP showing globular structured domains including: transcriptional adapter zinc binding motif (TAZ1, orange) and (TAZ2, purple); CREB binding domain (KIX, green); bromodomain (BRD, yellow); histone acetyltransferase domain (HAT, pink); zinc binding domain (ZZ, red) and nuclear receptor coactivator binding domain (NCBD, blue). The autoinhibitory loop (AIL, light blue) within the HAT domain is also indicated on this diagram. Image based on one found at (Dyson and Wright 2016). B - Structural prediction of CBP using AlphaFold, structured domains are coloured as described in A, Jumper et al. (2021).

Understanding which proteins phase separate, and how this affects their function is very

important, especially for CBP since changes in function can alter transcriptional regulation in the cell. When CBP is either depleted or overexpressed this commonly leads to disease, such as cancer (Chen et al. 2022). For example in acute myeloid leukemia, CBP and paralog p300 increase the transcription of genes involved in DNA replication, which increases in proliferation and immortilization of cancer cells (Giotopoulos et al. 2016; Roe et al. 2015). Understanding the phase separation of CBP and how this affects its function is therefore extremely important.

3.1.2 Chapter aims

In this chapter I will describe cellular techniques used to investigate the phase separation of endogenous CBP, and if these condensates behave in a fluid-like manner, with puncta fusing together and diffusing apart. I will further develop an overexpression system to explore full length wild type (WT) CBP phase separation, and characterise this behaviour using techniques such as 1,6 - hexanediol (1,6 - HD) and fluorescence recovery after photobleaching (FRAP). Overall, providing evidence that full length CBP can form phase separated condensates both endogenously and when overexpressed.

3.2 Identifying CBP puncta by immunofluorescence

3.2.1 CBP puncta were observed by immunofluorescence of HEK293T cells

As an initial way to determine whether endogenous CBP phase separates immunofluorescence on HEK293T was performed. Using this method endogenous CBP was labelled with antibodies to get an insight into whether, in cells, CBP forms puncta. HEK293T cells were fixed using 2 % paraformaldehyde and stained with a primary antibody for CBP and a fluorescently labelled secondary antibody Alexa Fluor - 488. The results show distinct nuclear puncta for CBP (Figure 3.2). This correlates with evidence in the literature showing CBP recruitment to foci in the case of double strand breaks, and is consistent with observed p300 puncta, a paralog of CBP (Manickavinayaham et al. 2019; Zhang et al. 2021; Ma et al. 2021). There was some background observed, even after processing, however this was due to background labelling that can occur with immunofluorescence. CBP was shown to be nuclear localised, as all CBP staining was observed within the DAPI stain, this was expected as the nucleus is where transcription occurs (Figure 3.2). Immunofluorescence was performed with only one primary antibody against CBP, while this is typically done with multiple antibodies, there is only one antibody which is specific to CBP and not p300, which is the antibody I chose to use. Overall, this data suggests that CBP forms

puncta at endogenous levels. I next wanted to see if these observed puncta colocalised with other factors commonly found at enhancers such as BRD4 and histone mark H3K27ac.

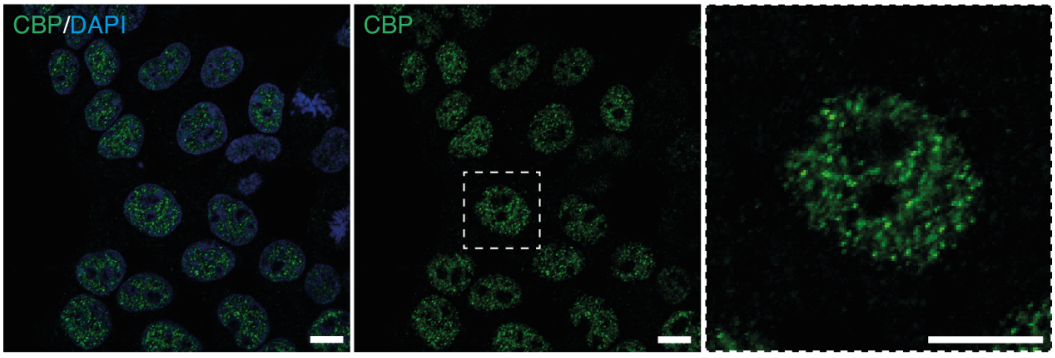

Figure 3.2: Endogenous CBP forms nuclear puncta as detected by immunofluorescence
CBP distribution was visualised in HEK293T cells using an α-CBP antibody (green). Nuclei were visualised with DAPI (blue). Zoomed region is indicated by white dashed box. Image was processed in ImageJ/Fiji background was subtracted with a rolling ball of 20 μm, max projections were generated, scale bars represent 10 μm.

3.2.2 CBP puncta colocalise with BRD4 and H3K27ac

To investigate whether CBP puncta colocalised with factors commonly found at enhancers, I chose to look at BRD4 and H3K27ac. BRD4 is involved in RNA polymerase II pause release in transcriptional initiation, which has been shown to phase separate at enhancers, and colocalises with CBP and p300 (Sabari et al. 2018; Han et al. 2020; Wang et al. 2022b; Tang et al. 2022; Arnold et al. 2021; Ma et al. 2021). Alongside BRD4 I also looked at H3K27ac, a histone mark deposited by CBP which is commonly found at enhancers, representing open chromatin (Chen et al. 2022; Sungalee et al. 2021; Tie et al. 2009; Creyghton et al. 2010; Bose et al. 2017). HEK293T cells were stained for either CBP and BRD4 or CBP and H3K27ac (Figure 3.3). Staining for CBP, BRD4 and H3K27ac was localised to the nucleus, as they were all contained within the DAPI staining, with puncta observed (Figure 3.3). This is more distinct for CBP and BRD4, potentially as a result of their phase separating ability. This could also be due to broader H3K27ac domains across the chromatin leading to a more diffuse signal, than that of CBP or BRD4. When overlayed the more distinct puncta appear to overlap and therefore colocalise, however this is not conclusive from the images alone.

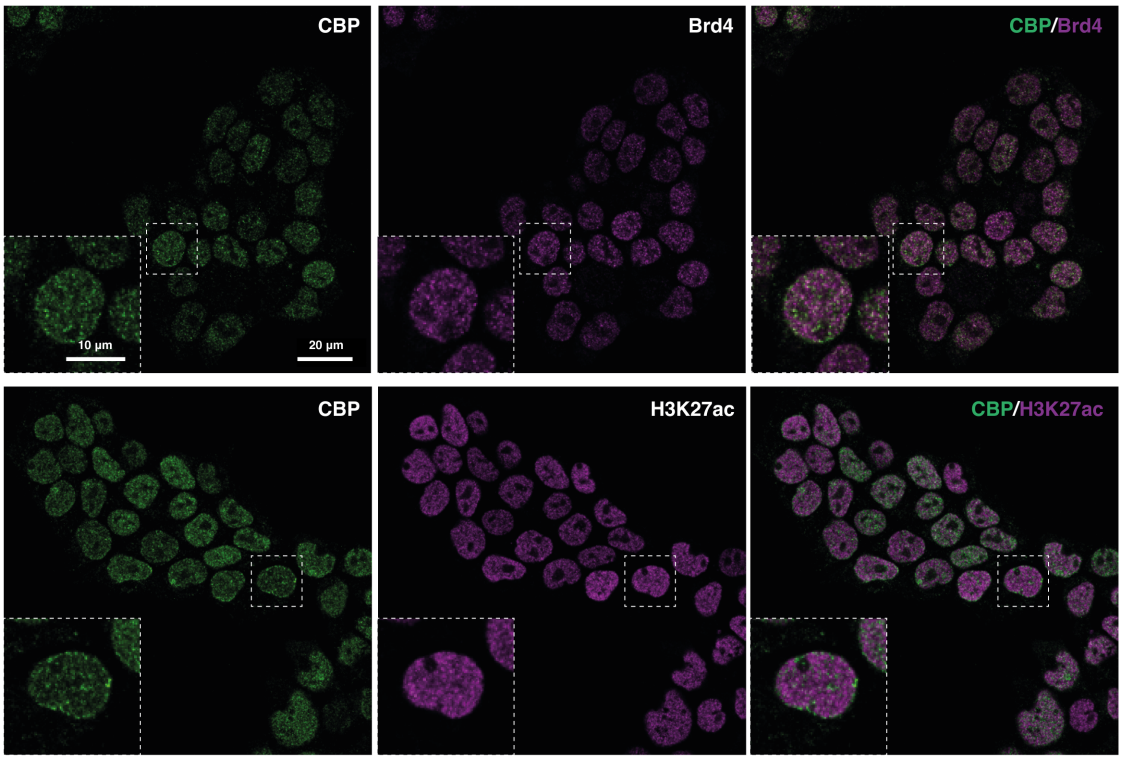

Figure 3.3: Immunofluorescence of endogenous CBP with BRD4 and H3K27ac
Colocalisation of CBP and BRD4 or CBP and H3K27ac using antibodies against α-CBP (green), α-BRD4 (magenta) and α-H3K27ac (magenta). Zoomed images are shown in the bottom left corner, indicated by the dashed white box. Images were processed in ImageJ/Fiji where the contrast was adjusted per image and scale bars added.

I did secondary antibody controls to check the level of backgroud staining; both secondary antibodies show non-distinct background staining that is not localised to the nuclei, which were shown here using DAPI (Figure 3.4). Signal observed for the secondary antibody controls were much lower than when used in combination with primary antibodies. By calculating the per pixel intensity, or gray values, in ImageJ/Fiji, the secondary antibody controls show an intensity within the range of 100 - 200 gray values, in comparison to the immunofluorescence images where for all primary antibodies a per pixel intensity within the range of 500 - 2000 gray values was observed. Suggesting low-level background staining.

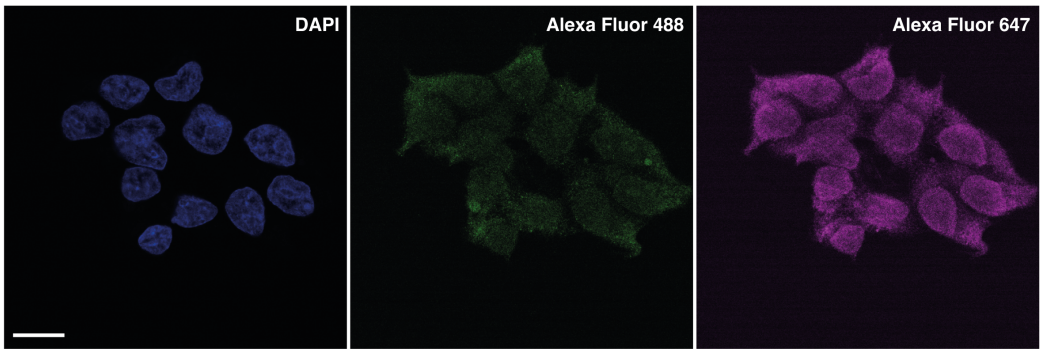

Figure 3.4: Secondary control for immunofluorescence of HEK293T cells
Visualisation of background staining in HEK293T cells, labelled with Alexa Fluor-488 (green) and Alexa Fluor-647 (magenta). Nuclei were identified using DAPI (blue). Scale bar was added in ImageJ/Fiji to represent 20 μm.

To quantify possible colocalisation of CBP with BRD4 and H3K27ac I generated line graphs in ImageJ/Fiji (Schindelin et al. 2012). This involved drawing a line on the nuclei of interest and measuring the intensity, or gray value, of the pixels along that line. The same line was applied to both channels, 488 and 642, and a graph was plotted comparing the grey value of the two channels along that line. The linescans show a general trend of overlap between CBP and both BRD4 and H3K27ac, however this was not fully consistent across the line (Figure 3.5). While there was some overlap, which could suggest CBP puncta being present at enhancers, this data was not conclusive. As it is known that CBP colocalises with BRD4, and that CBP is found to be present at the same genomic locus as H3K27ac, from the literature, I decided not to investigate this further (Sabari et al. 2018; Han et al. 2020; Wang et al. 2022b; Tang et al. 2022; Arnold et al. 2021; Chen et al. 2022; Sungalee et al. 2021; Tie et al. 2009; Creyghton et al. 2010; Bose et al. 2017).

Taken together this data has shown that CBP forms distinct clusters in HEK293T cells at endogenous concentrations. However, I was working with fixed cells, visualising only a snapshot of these cells at one timepoint. To confirm that these are phase separated condensates instead of aggregates for example, I wanted to visualise CBP in live cells. This will provide further evidence of CBP phase separation and whether these potential condensates are fluid-like. To do this I developed an overexpression model system which would allow further examination of CBP phase separation, while providing a versatile platform to manipulate this behaviour. Before beginning to test any parameters which affect CBP phase separation I first needed to establish

how CBP behaves when overexpressed and address some of the obvious caveats to overexpression, for example at these higher concentrations are these condensates still mobile.

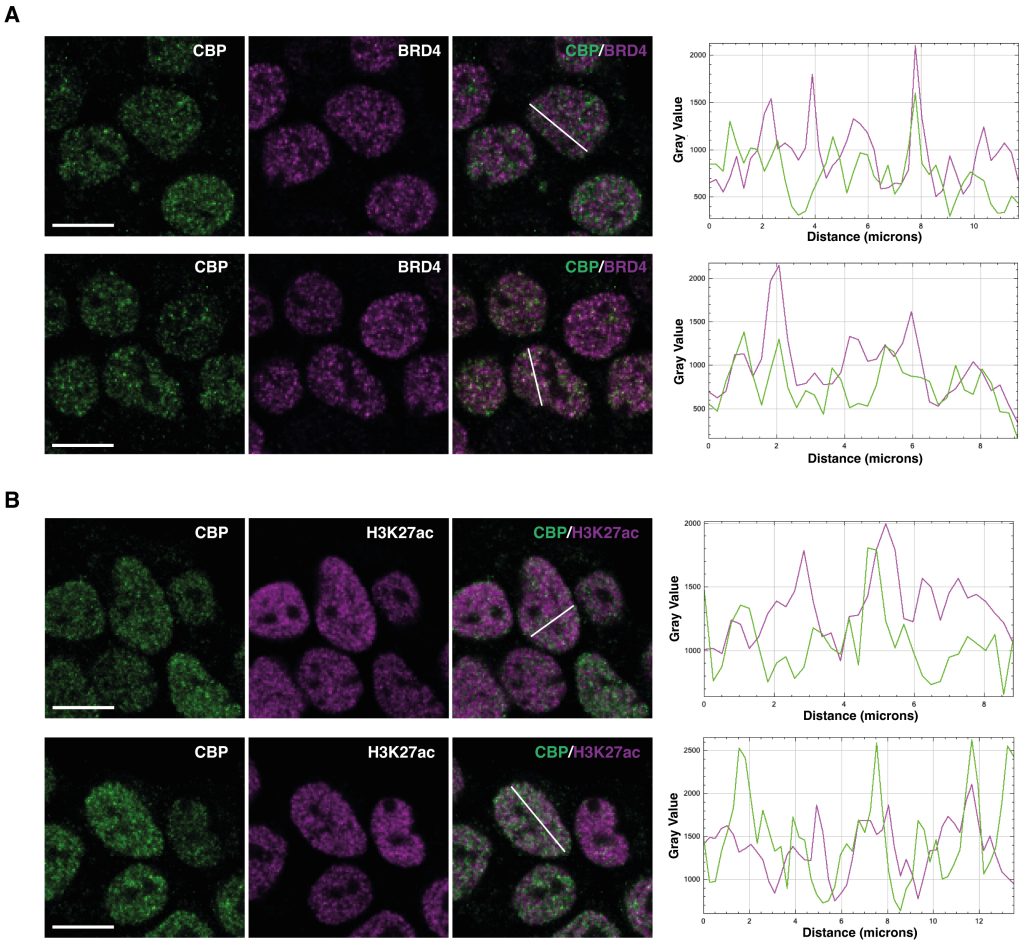

Figure 3.5: Colocalisation of CBP with BRD4 and H3K27ac as analysed by immunofluorescence Two representative nuclei and linescans for A - CBP and BRD4 and B - CBP and H3K27ac. Visualisation of puncta using antibodies against α-CBP (green), α-BRD4 (magenta) and α-H3K27ac (magenta). Linescans were performed in ImageJ/Fiji to visualise the intensity of each channel, with the line shown in white on the overlayed image. Linescans show the gray value, intensity, for each pixel on the line. Images were processed in ImageJ/Fiji where the contrast was adjusted per image and scale bars added at 10 μm.

3.3 Phase separation of overexpressed CBP-GFP

One way to study phase separation in cells is using overexpression systems, this allows for easier visualisation of a protein of interest and allows for easy manipulation of the protein sequence, as has been done multiple times in the literature (Chong et al. 2018; Sabari et al. 2018; Clemson et al. 2009). While overexpression brings into question how physiologically relevant these findings are, for the purposes of understanding the phase separating ability of CBP, overexpression is a useful tool to quickly gain answers about which regions within CBP are important for phase separation, without the lengthy processes of making inducible cell lines. I wanted to visualise CBP in human cells, therefore I cloned full length human CBP into a construct containing a C-terminal Green Fluorescent Protein (GFP) tag (pAcGFP-N1). For overexpression experiments I chose to use chemical transfection methods to generate transient overexpression of CBP-GFP.

3.3.1 CBP-GFP forms condensates when overexpressed in HEK293T cells

To understand the behaviour of CBP in a phase separated environment, I used a transient transfection system in HEK293T where CBP tagged with GFP could be overexpressed. Fixed HEK293T cells that had been transfected with CBP-GFP for 24 hours were imaged to visualise the phenotype of overexpressed CBP. Distinct nuclear CBP-GFP puncta were observed (Figure 3.6). These condensates were similar to those seen at endogenous levels, however were slightly larger and more distinct; with puncta observed by immunofluorescence being approximately 0.25 μm in size, and those observed when overexpressed approximately 0.8 μm in size. This was expected as I was using an overexpression system and it is known that phase separation is concentration dependant. This provided further evidence that CBP can phase separate, as this behaviour seen at endogenous levels was recapitulated when overexpressed. It also indicates that this overexpression system can be used as a model system for studying phase separation, with the caveat that concentration may be playing a role in increasing the propensity of CBP to phase separate.

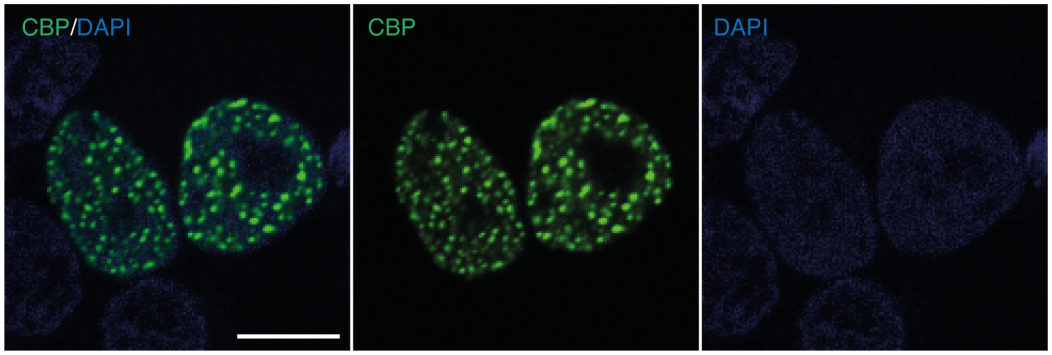

Figure 3.6: CBP-GFP forms distinct nuclear puncta when overexpressed
Representative images of fixed HEK293T cells showing nuclear puncta of CBP-GFP (green) that were transiently transfected. Nuclei were stained using DAPI (blue). Image is of the midplain slice from a Z-stack of the nucleus, images were processed in ImageJ/Fiji with a scale bar of 10 μm.

3.3.2 Understanding the mobility of overexpressed CBP-GFP using FRAP

In the previous section I used fixed cell overexpression in HEK293T cells as an indicator of phase separation. While this was a useful tool to observe the phenotype of this construct, a caveat of fixed cell studies is that this was only looking at a snapshot in time of overexpressed CBP-GFP, which means these observed puncta could be less dynamic in an aggregate as opposed to dynamic condensates, due to the increased concentration present in the system. To test how fluid-like these condensates are, I switched to using live cells so that I could perform fluorescence recovery after photobleaching (FRAP).

3.3.2.1 Using FRAP to study CBP dynamics

HEK293T cells were grown in glass bottom dishes and transfected with CBP-GFP for 24 hours before live cell imaging, where a timecourse was performed over a total of 3 minutes. Briefly, 5 frames were captured before photobleaching followed by 5 cycles of bleaching using a region of interest with a Feret diameter of 0.2 μm, and subsequent imaging to monitor the recovery. The time between capturing images after photobleaching was varied to reduce the effect of photobleaching on the sample due to repetitive use of the laser to capture the recovery images. For this experiment, images were taken every 0.5 seconds for 30 cycles, followed by every 1 second for 60 cycles and every 3 seconds for 30 cycles.

The results show distinct nuclear puncta for CBP-GFP, as has been observed in the fixed

samples. Upon photobleaching of a distinct condensate, a reduction of fluorescence intensity was observed that recovers over the timecourse (Figure 3.7, for the purposes of comparison only the first 30 seconds of imaging is shown, as this is where the majority of change occurs). Overall, this data shows that the FRAP system I had in place was functioning; CBP-GFP forms puncta that behave in a mobile manner.

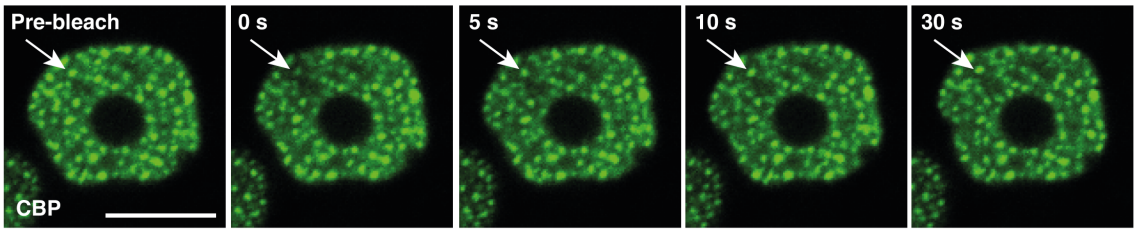

Figure 3.7: Fluorescence Recovery after photobleaching of CBP-GFP
Visualisation of fluorescence recovery after photobleaching (FRAP) of HEK293T cells transfected with CBP-GFP. White arrow indicates where the laser was targetted and therefore where the photobleaching occured over the timecourse. Images were processed in ImageJ/Fiji, with a scale bar of 10 μm.

Alongside imaging data I also wanted to quantify the fluorescence recovery of CBP-GFP. After capturing the timecourses the data was analyzed in ImageJ/Fiji, using an ImageJ plugin called Stack Registration (Thévenaz, Ruttimann and Unser 1998; Schindelin et al. 2012), which accounts for drift in the sample which can happen over the timecourse. Then a region of interest (ROI, of defined size per experimental repeat) was drawn around the area that was photobleached as the FRAP region, another ROI was then drawn in a separate area that has not been targeted by the laser to act as a reference region where background photobleaching can be monitored over the timecourse. An example image showing ROIs for the photobleach and reference region are shown in Figure 3.8. Percentage fluorescence recovered was then determind using the equation shown in Figure 3.8 A. Averages from experimental repeats were created to generate a fluorescence recovery curve of CBP-GFP (Figure 3.8).

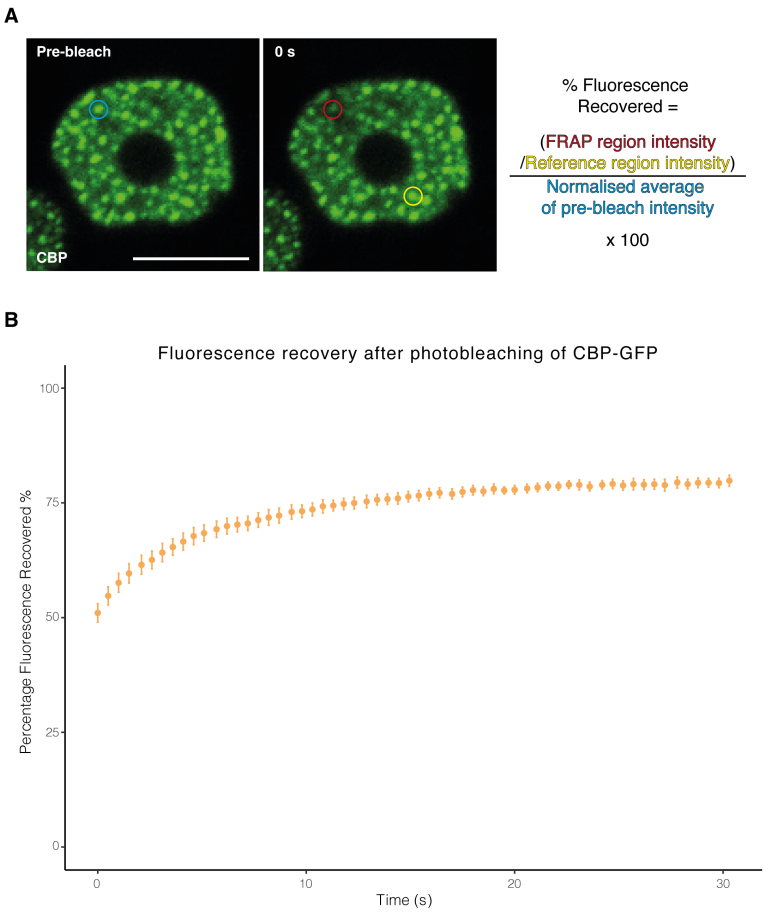

Figure 3.8: Fluorescence recovery of CBP-GFP
A - captured image of CBP-GFP taken from Figure 3.7 demonstrating the regions of interest used to calculate the percentage fluorescence recovered. Where the FRAP region is shown in red, the reference region is shown in yellow and the pre-bleached region is shown in blue. The equation used to calculate the fluorescence recovery is shown to the right. B - average percentage fluorescent recovery of CBP-GFP (orange) over 30 seconds post-bleaching. Average fluorescence recovery is shown with error bars representing the standard error of the mean. Graph was plotted using R in RStudio.

The results show CBP-GFP photobleaches to approximately 50 % from the pre-bleach intensity and recovers within approximately 20 seconds (Figure 3.8). This recovery was on a time scale of seconds and is comparable to recovery rates of other phase separating proteins in the literature (Wang et al. 2022b; Hallegger et al. 2021; Taylor et al. 2019).

This data indicates that overexpressed CBP-GFP behaves in a liquid-like manner in this model system.

3.3.3 1,6 - hexanediol causes dispersion of CBP-GFP overexpressed condensates

To further understand the liquid-like nature of CBP I also wanted to test it's susceptibility to 1,6 - hexanediol treatment. This aliphatic alcohol will disrupt hydrophobic interactions, which may be important for CBP phase separation. To test this I used overexpression of CBP-GFP, in the same manner as that described in section 3.3.2. A timecourse was taken, imaging every 1.9 seconds for 4 minutes; where 10 % 1,6 - HD was added after 1 minute to reach a final concentration of 5 %. The affect of adding 1,6-HD was then monitored over the timecourse (Figure 3.9). For CBP-GFP, puncta were observed at the start of the timecourse and after addition of 1,6-HD dispersion of condensates occurs over a time scale of 10 seconds (Figure 3.9). Overall, CBP puncta observed in overexpression experiments can be disrupted by the addition of 1,6-HD and therefore hydrophobic interactions may be involved in the formation of CBP condensates.

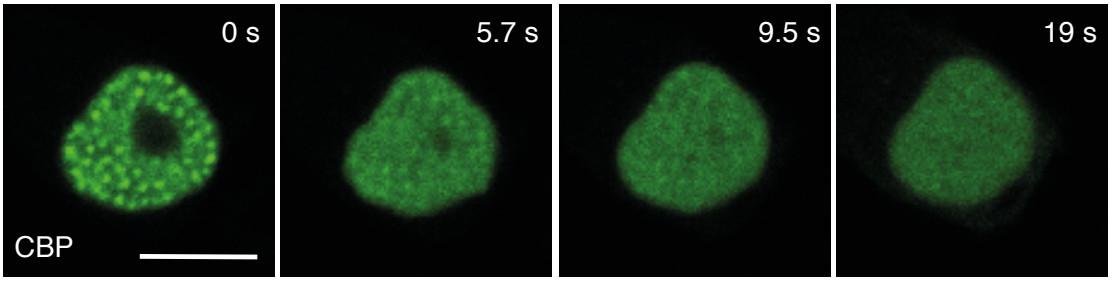

Figure 3.9: Disruption of CBP-GFP condensates upon addition of 1,6-hexanediol treatment
Representative images of the affect of adding 1,6-hexanediol to a final concentration of 5 % in HEK293T cells that are overexpressing CBP-GFP. Time of 0 seconds indicates the addition of 1,6-hexanediol. Images were processed in ImageJ/Fiji, scale bar represents 10 μm.

Together, the FRAP and 1,6-HD results show that overexpressed CBP-GFP puncta behave in a manner which indicate they are mobile phase separated condensates, which is in line with the data for the initial endogenous puncta seen by immunofluorescence. However, using overexpression to investigate the behaviour of full length CBP phase separation has the caveat that the extent to which overexpression, and concentration is driving phase separation is unknown. Next, I wanted to further confirm CBP phase separation using an endogenously tagged cell line.

3.4 Behaviour of endogenous CBP condensates

3.4.1 Endogenous CBP-HaloTag puncta behave dynamically in live cells

To further validate CBP as being able to phase separate, I wanted to visualise puncta in live cells at endogenous concentrations, to see them fusing together and dispersing apart, as is characteristic in the literature (Kar et al. 2022; Ma et al. 2021; Riback et al. 2020; Garcia-Jove Navarro et al. 2019). To do this I utilised a HEK293T cells line where endogenous CBP was tagged on the C-terminus with a sequence of tags: 1X Flag epitope (Flag, F) - Tobacco etch virus protease site (Tev, T) - 6X Histidine (His, H) - HaloTag (Halo, H): (FTHH) using the CRISPR-Cas9 system - this work was done by masters student Sophie Ball and Dr Nicola Carruthers. The validation for this cell line was performed by Dr Nicola Carruthers and can be found in the appendix, section 9.3. This cell line will be referred to as CBP-HaloTag.

The addition of these tags meant that I could fluorescently label the HaloTag with a fluorophore, in this case TMR (Promega), to visualise enodgenous CBP within live cells. Endogenously tagged cells were plated on PLL-coated 35-mm glass bottom dishes and grown for 48 hours before labelling with TMR and Hoechst. Live cell images show clear distinct nuclear puncta, as was seen in endogenous cell staining which is indicative of a protein that phase separates (Figure 3.10 A). This puncta was specific to the CBP-HaloTag line and was not an artefact of aberrant TMR staining, as in control images some fluorescence was observed but this was not nuclear and cannot be mistaken for the puncta observed in the CBP-HaloTag line (Figure 3.10 A). These puncta were mobile as they were seen to fuse together and separate (display fission) overtime, which is typically observed for a protein that phase separates (Figure 3.10 B). Using the CBP-HaloTag HEK293T cell line endogenous CBP puncta has been observed that behave in a mobile manner.

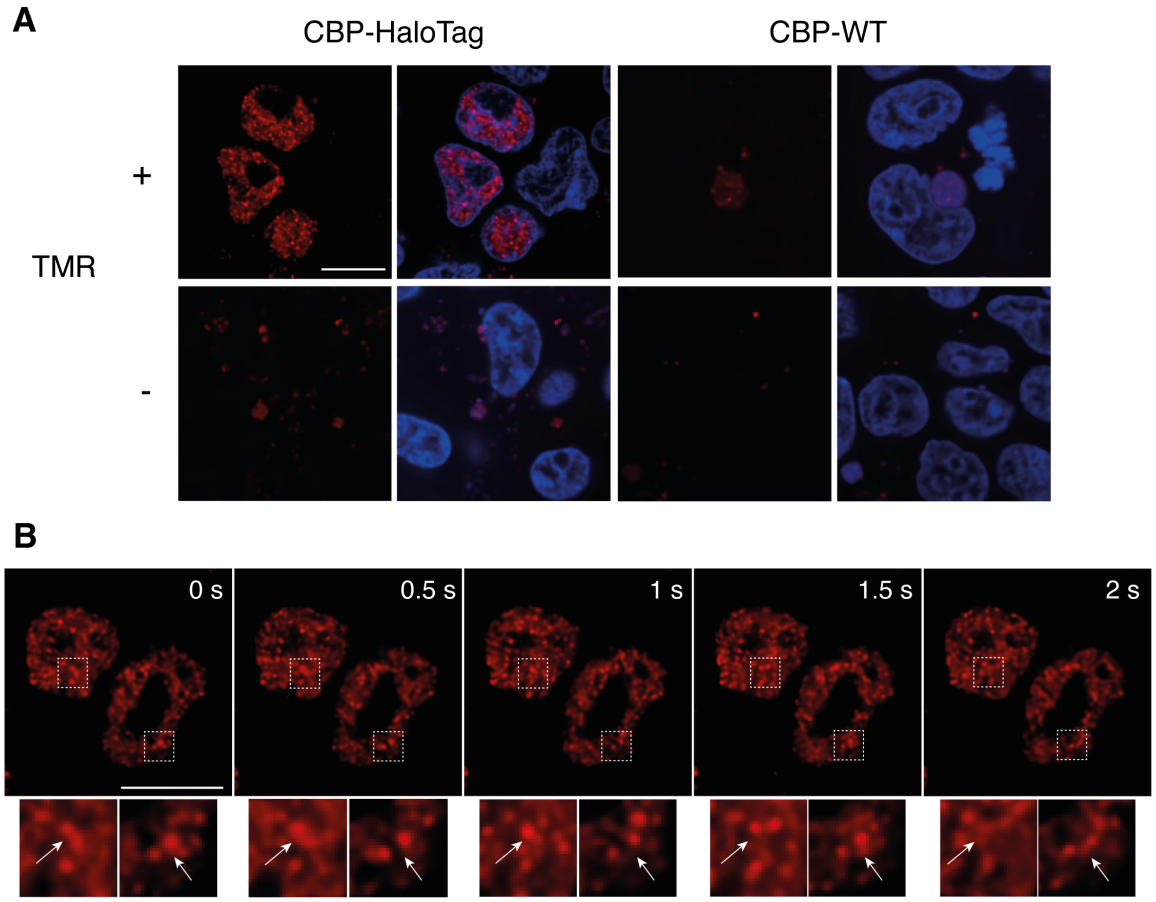

Figure 3.10: Endogenous CBP-HaloTag condensates in HEK293T live cells.
A - Representative images of CBP-HaloTag and HEK293T cells with and without TMR staining, to label the HaloTag. Nuclei were identified using Hoechst (blue). B - Representative images taken from a timecourse of CBP-HaloTag with TMR showing puncta fusion and fission, zoomed in regions are indicated by the dashed boxes, with white arrows pointing to condensates of interest. Images were preprocessed using Nikon denoise algorithem, images were procesed in ImageJ/Fiji, scale bars represent 10μm.

3.4.2 FRAP of endogenous CBP-HaloTag puncta show that the protein is mobile

To understand the behaviour of endogenous CBP puncta, I wanted to characterise their mobility. Since larger puncta move less in the nucleus, I want to determine whether they are still dynamic

in behaviour, which is typical of a phase separated protein. To do this I performed FRAP on endogenous CBP-HaloTag puncta labelled with Janelia Fluor (JF) 594 (Promega). The experiment was set up as described for section 3.4.1. ROIs were generated that defined where the laser would aim, this may be multiple places per field of view. A timecourse was taken, where in the first frame both the nuclei and the tagged CBP were visualised, to show that the signal that was observed was nuclear localised and not just abberant staining from the JF 594. Subsequent frames were taken before and after FRAP to visualise recovery; 15 frames were captured before bleaching with a 500 ms delay, followed by bleaching, subsequently 60 frames were captured with a 500 ms delay and 30 frames with a 2 second delay, to minimise the effect of photobleaching. In this experiment the FRAP settings were set to photobleach only the targetted structure, to minimise any damage to the nucleus or surrounding proteins which could have unknown consequences within the cell.

The results show large distinct nuclear puncta for CBP-HaloTag, that became photobleached after being targetted by the laser which recover over time. Images of a representative FRAP experiment on CBP-HaloTag puncta is shown in Figure 3.11 A. The recovery after photobleaching was quanitified over time using ImageJ/Fiji (Schindelin et al. 2012), and performed as described for overexpressed CBP-GFP in subsection 3.3.2.1. Plotted fluorescence recovery shows bleaching to ~ 25 % with a recovery to ~ 75 % within 60 seconds (Figure 3.11 B). One parameter often investigated is the half time or T = 1/2, which is the time it takes for half of the fluorescence to be recovered to the plateau. By fitting a curve using GraphPad Prism this was calcuated to be 11.14 seconds (San Diego 2022). This data suggests that even the larger endogenous condensates display mobile behaviour at physiological concentrations, providing further evidence that CBP forms phase separated condensates.

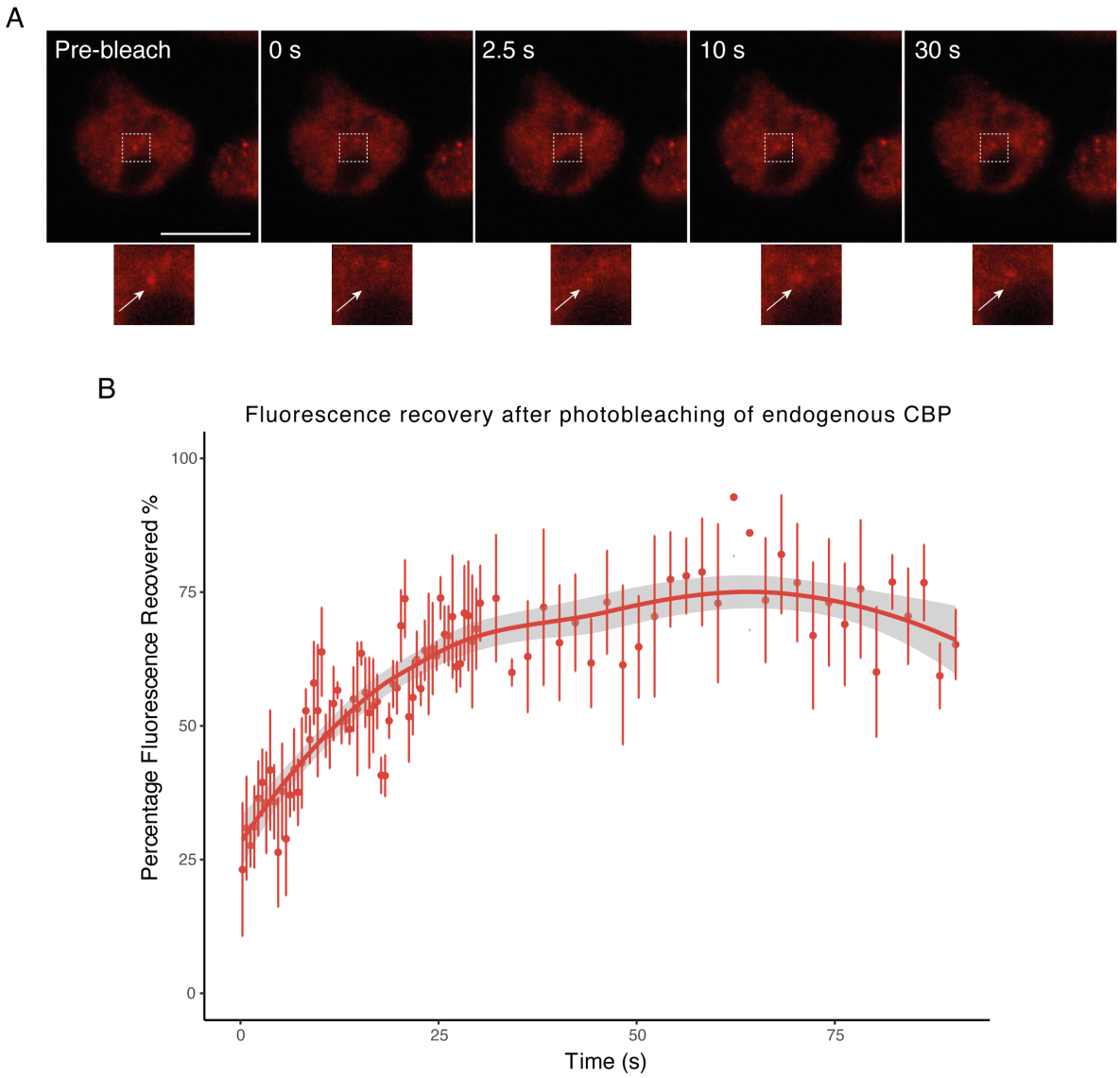

Figure 3.11: FRAP of endogenous CBP-HaloTag puncta

Fluorescence recovery after photobleaching (FRAP) of CBP-HaloTag puncta (red). A - Representative images of the CBP-HaloTag cells before bleaching and subsequently after bleaching, labelled with Janelia Fluor-549 to visualise puncta. Images were processed in ImageJ/Fiji, scale bar represents 10 μm. B - Average percentage fluorescent recovery shown over a 90 second time course. Error bars represent the standard error of the mean. Graph was plotted using R in RStudio. Number of FRAP events quantified = 4. Biological replicates = 1.

3.5 Discussion

The first question I wanted to address was whether endogenous CBP can form phase separated condensates. This hypothesis was developed as CBP is over 60 % disordered, a common feature of proteins that phase separate, and previous work on p300, a paralog of CBP, has show that it can form condensates at endogenous levels (Ma et al. 2021; Wang et al. 2022b; Zhang et al. 2021). Taken together this highly suggests this phenomena is possible in CBP. Initially, I did this via immunofluorescence of fixed cells, as this experiment is relatively quick and provided an initial indication that CBP does form endogenous puncta.

3.5.1 Using linescans to determine if fixed cell CBP puncta were part of transcriptional condensates

To determine if CBP puncta observed in fixed cells were part of transcriptional condensates, dual-colour immunofluorescence was performed between CBP, BRD4 and H3K27ac. Linescans were used to determine colocalisation of the signals, which would indicate that CBP forms puncta at enhancers. A general trend of colocalisation was observed, while not conclusive.

3.5.2 Using overexpression as a model system for CBP puncta

To further characterise CBP phase separation, an overexpression model system was developed, to allow more rigorous testing of CBPs phase separating ability. In this system, CBP was tagged with GFP for visualisation when overexpressed. The GFP tag was chosen as it is a bright fluorescent protein so is easily visualised with conventional confocal microscope and it has been well characterised and used frequently for imaging. To validate the overexpression system, HEK293T cells were transfected with CBP-GFP, where bright, distinct nuclear puncta were visualised in fixed cells, as had been observed for endogenous CBP. Puncta observed as a result of overexpression were more defined, and were visually larger than that observed at endogenous levels, which may suggest that concentration may be fuelling phase separation in this system.

From the experimental results above I have shown that CBP-GFP behaves as a mobile protein as it photobleaches and recovers over a time scale of seconds, a strong indicator commonly used in the literature that a protein phase separates (Wang et al. 2022b; Ma et al. 2021; Hallegger et al. 2021; Taylor et al. 2019). This is also corroborated by 1,6-HD results as CBP-GFP are susceptible to the addition of 5 % 1,6-HD. 1,6 - HD even at low concentrations is harsh on cells, in the experiments detailed here changes in nuclei shape and size were observed over the

timecourse of 1,6-HD treatement, indicating levels of cellular toxicity. It has also been shown to affect enzyme activity of kinases and phosphatases, as well as its ability to compromise 3D genome organisation (Düster et al. 2021; Ulianov et al. 2020). While in this assay dispersion of condensates for CBP-GFP was observed, it is extremely difficult to directly correlate the affect of 1,6-HD on weak intermolecular forces that are holding condensates together, since it clearly affects multiple different components of the cell besides phase separation. As such, very little significance on this assay is being placed in understanding the phase separation of CBP. This experiment was performed as an indicator of protein behaviour in thisoverexpression system, in line with current techniques in the literature, but no further weight is being placed on this result. While for these purposes 1,6-HD was sufficient to indicate the fluid-like nature of CBP-GFP, if future experiments are needed, propylene glycol should be considered as a less toxic alternative to indicate the fluid-like nature of the condensates of interest (Geiger et al. 2021; Mochida and Gomyoda 1987).

3.5.3 Developing an endogenously tagged cell line to characterise CBP puncta in living cells

Dr Nicola Carruthers and masters student Sophie Ball generated a HEK293T cell line where endogenous CBP was labelled with a series of tags, Flag-Tev-His-HaloTag. These tags can be used for purification of the protein from the cell lysate, but more importantly for this work it facilitated labelling the protein in live cells using the HaloTag. A plethora of ligands are commercially available to attach to the HaloTag, including fluorescence ligands such as TMR (Promega) or Janelia Fluor (Promega) ligands that excite in different wavelengths of light. The benefit of using this cell line is that it enabled the visualisation of endogenous CBP in live cells, in order to track the puncta and perform experiments to check the dynamics of CBP such as FRAP.

Tagging CBP, as with any protein may lead to changes in protein behaviour; this is equally true with this cell line as the tag itself is quite large at 297 amino acids in length. This tag has been used in multiple strategies in the past to study phase separation, and is similar in size to GFP, which is 238 amino acids long and is commonly used for protein visualisation (Ye et al. 2022; Lackner et al. 2022; Cai et al. 2019; Liu et al. 2015). To mitigate the effect of adding the tag it was placed on the C-terminus to reduce the effect on folding of CBP.

Using fluorescently labelled CBP-HaloTag, endogenous puncta were observed that behaved in a fluid-like manner, showing condensates fusing together and dissipating. To observe the mobility of these condensates, I used FRAP using custom built FRAP/TIRF microscope BONSAI which

has a more sensitive detector than that of a standard confocal micoscope. Using targeted laser power I photobleached a condensate to reduce the fluorescence intensity of that specific puncta, while disturbing the cell as little as possible. Laser power itself has been shown to cause various effects of the cellular structure, with photodamage being shown to affect nuclei adjacent to the target cell, having largely unknown cellular effects (Dobrucki, Feret and Noatynska 2007).

Using this technique disruption of the condensate structure and recovery of fluorescence of that condensate over the timecourse was observed. With recovery occuring within 60 seconds, with a half time of 11.14 seconds. FRAP has been performed for p300, where a recovery between 8 and 10 seconds was shown for endogenously tagged protein (Ma et al. 2021; Wang et al. 2022b). While the recovery for CBP in this system took slightly longer than that observed for p300, the recovery occurred on a timescale of seconds which suggests that CBP is behaving in a mobile manner, as has been described for several phase separating proteins. Although a promising method to show the mobility of endogenous CBP, the consequences of photobleaching are not well characterised and the results therefore should serve no more than an indicator of the dynamics of CBP.

3.5.4 Outlook for full length CBP phase separation

Overall, I have shown that CBP can form phase separated condensates at endogenous levels using both fixed and live cell techniques. These endogenous puncta are fluid-like in motion, as they diffuse apart and fuse together and recover after photobleaching, displaying qualities of a mobile protein. Similarly, when CBP is overexpressed in HEK293T cells, puncta were observed that were distinct and dynamic in their recovery to photobleaching, as well as apparent susceptability to 1,6-HD treatment. This overexpression system is reasonably representative of endogenous protein in its phase separation behaviour and can therefore be a model system to further investigate regions of interest within CBP, and how this affects CBP's phase separating ability.

Chapter 4

Defining intrinsic disorder regions within CBP which influence phase separation

4.1 Introduction

In the previous chapter I have established that endogenous CBP forms nuclear condensates, that behave in a mobile manner. I have also developed a model system, using overexpression, to further study the phase separation of CBP. I next wanted to investigate if there are specific regions within CBP that regulate phase separation.

4.1.1 The importance of intrinsic disorder on CBP phase separation

Intrinsic disordered regions (IDRs) have been commonly investigated as drivers of protein phase separation, however, in the last couple of years several reviews have highlighted that there importance may have been over exaggerated and the these weak forces are a gross simplification of the big picture (Mittag and Pappu 2022; Musacchio 2022). While the direct implications of IDRs on condensation and protein function are not yet understood, IDRs are still routinely investigated. As I have shown that CBP can form phase separated condensates, and that it is heavily intrinsically disordered I next wanted to understand if there are specific regions or sequences within these disordered areas that are important for the phase separation of CBP.

4.1.2 Does FUS aid in CBP phase separation

One protein that has been shown to phase separate consistently is Fused in Sarcoma (FUS), this is a protein involved in transcription and DNA repair that is commonly mutated in diseases, such as Amyotrophic Lateral Sclerosis (ALS) (Shang and Huang 2016). There are multiple FUS binding sites within CBPs sequence, with the largest region where they interact being in the C-terminal region of CBP (Birsa et al. 2020; Hernández-Candia, Pearce and Tucker 2021; Monahan et al. 2017; Wang et al. 2008b). I therefore wanted to further understand this relationship between FUS and CBP to determine if FUS's interaction with CBP is necessary for phase separation.

4.1.3 Chapter aims

In this chapter I will define several regions within CBP which may influence CBPs phase separating behaviour. I will determine if these regions are sufficient or essential for phase separation of full length CBP. Several CBP deletion mutants were then chosen to undergo fluorescence recovery after photobleaching (FRAP) and 1,6 - hexanediol (1,6-HD) treatment to further characterise these mutations.

4.2 Defining IDRs within CBP

To define specific IDRs within CBP I used the Predictor of Natural Disordered Regions (PONDR), a disorder prediction software (Xue et al. 2010). The results from this was a disorder prediction graph, showing the prediction of disorder for each amino acid residue, where a prediction of 1 is indicative of disorder, and 0 is indicative of highly structured regions (Figure 4.1 B). Specific IDRs were defined, where stretches of 50 or more amino acids that have a probability above 0.5 of being disordered were predicted IDRs, these were labelled IDR 1 - 7 (Figure 4.1 B) (Xue et al. 2010; Basu et al. 2020). The IDRs predicted from PONDR, aligned within the regions predicted as disordered using DISOPRED3 (Figure 1.16) (Jones and Cozzetto 2015), and were therefore what I have used in this doctoral work.

At this point I also included a region termed the CBP FUS interaction domain (CFID), which is the largest region where CBP binds phase separating protein FUS as described in Wang et al. (2008b), to see if this region affects CBPs phase separating ability (Birsa et al. 2020; Hernández-Candia, Pearce and Tucker 2021; Monahan et al. 2017; Wang et al. 2008b). Another region I included was the autoinhibitory loop (AIL), this loop sequence is known to be important for the activation of CBP via acetylation, and eRNA interaction and has been suggested to be important

for the phase separation of p300, a paralog of CBP. I therefore highlighted the AIL as a region of interest which I wanted to investigate further (Ma et al. 2021; Zhang et al. 2021; Bose et al. 2017). These regions, including specific domain boundaries, are mapped to the domain diagram of CBP in Figure 4.1A.

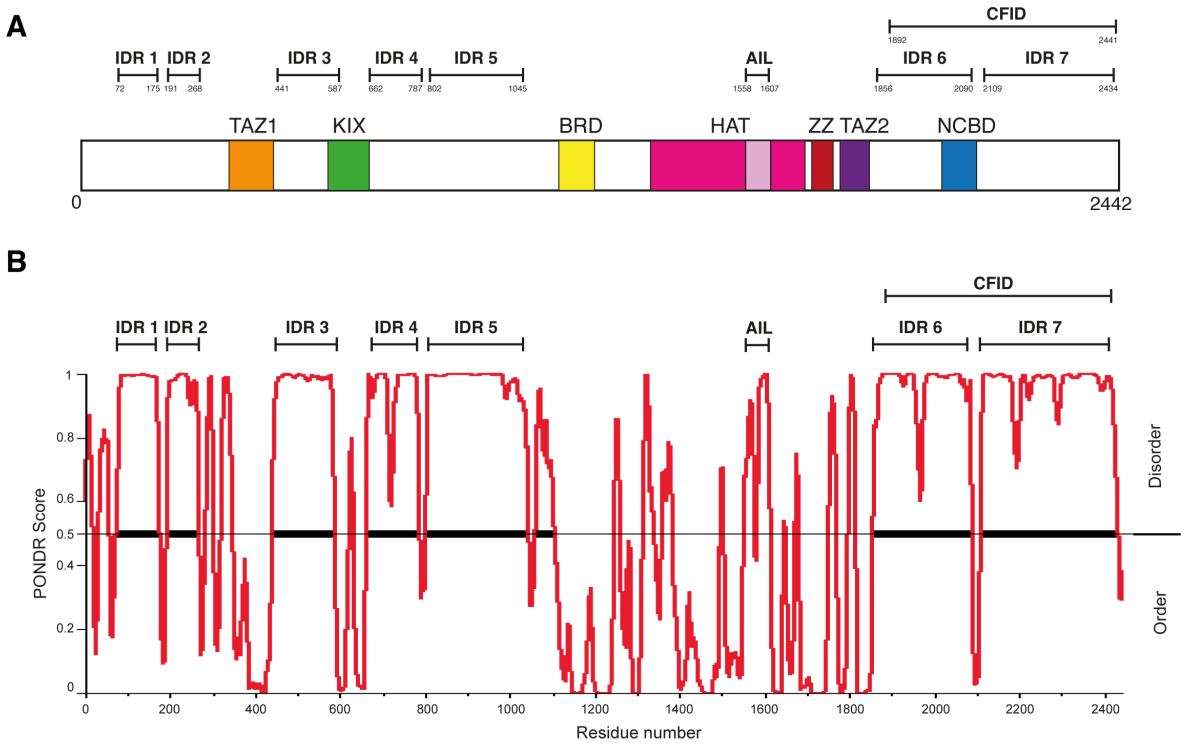

Figure 4.1: Intrinsic disorder mapping over CBP

A - Domain diagram of CBP showing globular structured domains including: transcriptional adapter zinc binding motif (TAZ1, orange) and (TAZ2, purple); CREB binding domain (KIX, green); bromodomain (BRD, yellow); histone acetyltransferase domain (HAT, pink); zinc binding domain (ZZ, red) and nuclear receptor coactivator binding domain (NCBD, blue). Image based on one found at Dyson and Wright (2016). Indicated above the domain diagram are the boundaries of intrinsic disorder regions 1 - 7, the autoinhibitory loop (AIL) and the CBP-FUS interaction domain (CFID). Amino acid domain boundaries are shown for each region, where the boundaries are the first and last amino acid of that region. B - Disorder prediction plot showing the PONDR score which represents disorder, a score of 1 represents regions of disorder across the residues within CBP. PONDR software was used to determine the IDR boundaries within this PhD thesis, with IDR regions indicated above. Plot was produced using PONDR intrinsic disorder prediction software Xue et al. (2010).

Upon defining regions of intrinsic disorder I next wanted to characterise whether these regions phase separate independently and how the removal of these regions affect full length CBP phase separation. In generating CBP-GFP deletion mutants for use in this overexpression system I was unsuccessful in producing mutants containing IDR2 and IDR5, this was due to difficulty with the highly repetitive sequence in the intrinsically disordered regions, therefore these were not included in the subsequent work.

4.3 Does CBPs IDRs phase separate independently using the optoDroplet system?

Before investigating the role ROIs play on the phase separation of CBP as a whole, I wanted to look at them individually to see whether they phase separate independently. To do this I used the optoDroplet system introduced by Shin et al. (2017), which has been used in the literature to determine whether specific IDRs phase separate, and therefore contribute to full length protein behaviour (Shin et al. 2017; Basu et al. 2020).

4.3.1 Introduction to the optoDroplet system

The optoDroplet system was first introduced by Shin *et al.* (2017) as a method to control phase transitions in a spatiotemporal manner, to look at the organisation of molecules within the cytoplasm. This system utilises a plasmid which contains a mCherry tag fused to the Photolyase Homology Region (PHR) of the CRY2 protein from *Arabidopsis thaliana*, this domain oligomerises upon stimulation by blue light and brings any accompanying sequences into close proximity, increasing their local concentration. If these sequences form phase separated condensates, puncta will be formed upon stimulation with blue light, thus providing mechanism with which to test if sequences within CBP can phase separate. Typically, it has been used to show whether specific IDRs phase separate, to imply the importance of those sequences to the phase separation of a whole protein (Shin et al. 2017; Basu et al. 2020).

The positive control contains the IDR from FUS, a phase separating protein, which has been used in previous publications, to test whether the system was working (Shin et al. 2017). The empty plasmid contains only the mCherry tag attached to the PHR domain of CRY2. To this I cloned in the nuclear localisation signal from SV40 so that the protein would be nuclear localised, as CBP is typically found in the nucleus, this plasmid was used as the negative control. All regions of interest were cloned into this construct, upstream of the mCherry tag. Each region

could therefore be individually tested to determine if they phase separate independently of CBP, in the nucleus upon blue light stimulation. Domain maps of these constructs can be found in Figure 4.2.

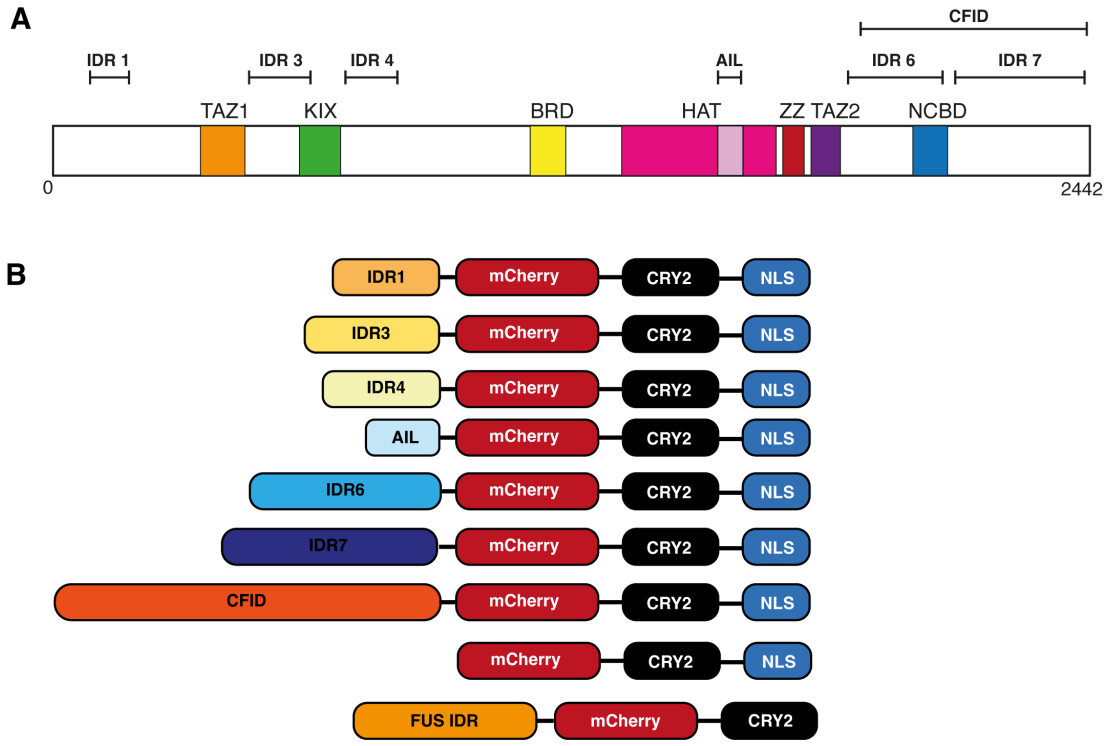

Figure 4.2: Construct maps of IDR constructs
Regions of interest used within this doctoral work and the domain map of the optoDroplet construct. A - domain map of CBP showing the regions of interest looked at within the optoDroplet work, including intrinsic disorder regions 1, 3, 4, 6, 7, the autoinhibitory loop (AIL) and the CBP FUS interaction domain (CFID). Domain maps of the essential part of the optoDroplet construct where regions of interest are fused to mCherry tag (red) followed by the Photolyase Homology Region (PHR) region of CRY2 (black) which is the light inducible dimerisation domain and a SV40 nuclear localisation signal (NLS, blue). The negative control has no region of interest attached and the positive control construct contains the FUS IDR in the absence of a nuclear localisation signal.

4.3.2 Setting up the optoDroplet assay

HEK293T cells were transiently transfected after 24 hours of cell growth with the respective plasmid; 48 hours after transfection the cells were imaged. A timecourse was taken, where images

were captured every 1.9 seconds; at each time point an image was taken first by the 594 nm laser to visualise the mCherry tag which was used to monitor condensate formation, and subsequently by the 488 nm laser used to stimulate oligomerisation of the CRY2 domain.

In setting up the assay I used the FUS-IDR plasmid as the positive control, and the empty plasmid containing the nuclear localisation signal as the negative control. After optimisation of the assay conditions the system was behaving as expected, with puncta observed for the positive control and no puncta formed with the negative control (Figure 4.3). For the positive control, puncta were observed immediately after blue light exposure over the whole cell body as there was no NLS (Figure 4.3). The positive control showed puncta after 8 seconds that become larger and brighter over the timecourse (Figure 4.3). In contrast, the negative control showed no formation of puncta, however mCherry signal was observed (Figure 4.3). These controls were used to determine the appropriate laser power and imaging conditions in each experimental repeat to test the assay conditions on the day.

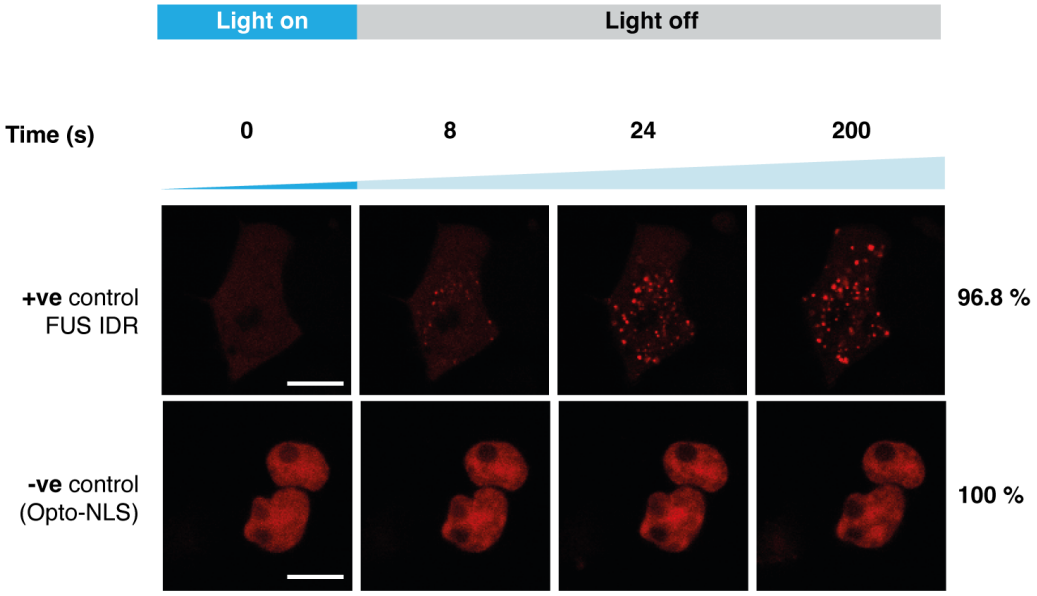

Figure 4.3: Representative images of optoDroplet controls
Visualisation of construct behaviour upon blue light exposure, of the FUS IDR positive control and the negative control which was the optoDroplet plasmid containing the nuclear localisation signal (NLS). The time of the representative images is shown above in seconds (s). Images were processed in ImageJ/Fiji with a scale bar of 10 μm. Percentage of nuclei transfected which display the represented phenotype is shown on the right hand side. 62 - 86 nuclei were imaged per sample, biological replicates = 3.

4.3.3 Investigating the phase separating properties of distinct regions of CBP

Once the system was set up, I began to test the constructs containing the regions of interest, following the same experimental procedure as described above. Controls for this experiment were shown previously in Figure 4.3. Expressed truncated protein regions IDR1, IDR3, IDR4, AIL and IDR6 stay diffuse over the timecourse, with no puncta observed, suggesting that these regions do not phase separate independently (Figure 4.4). As in the negative control some mCherry signal was observed. Puncta were observed for IDR7 and CFID over the timecourse (Figure 4.5). Condensates were observed after blue light exposure, that become bigger and brighter throughout the timecourse similar to the behaviour of FUS (Figure 4.5). The behaviour displayed by IDR7 and the CFID is what would be expected of regions that phase separate.

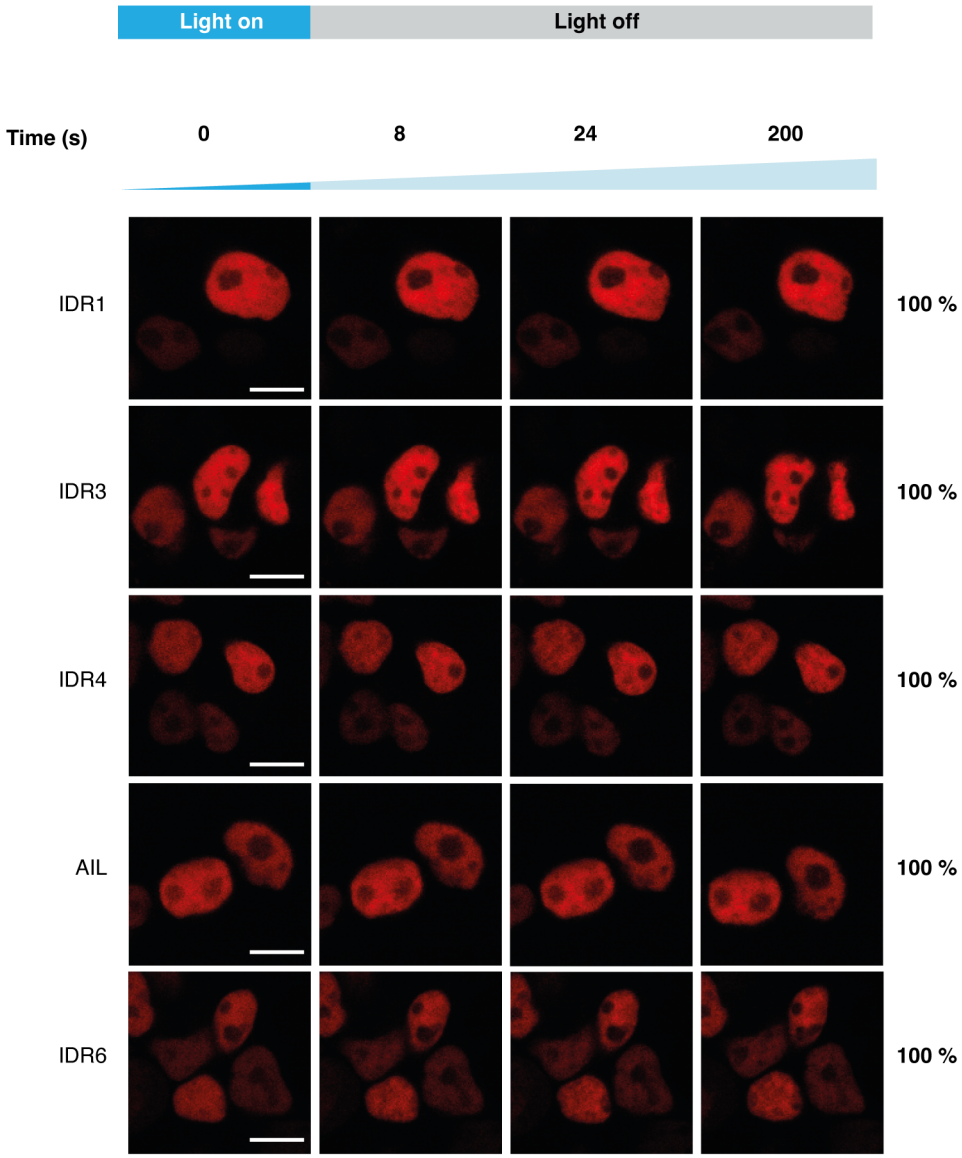

Figure 4.4: Expression of regions within CBP which do not phase separate using the optoDroplet system

Representative images of expressed regions within CBP including intrinsic disorder regions 1, 3, 4, 6 and the autoinhibitory loop (AIL) that do not form puncta upon blue light exposure over the timecourse. Time is shown above in seconds (s). Images were processed in ImageJ/Fiji with a scale bar of 10 μm. Percentage of nuclei transfected which display the represented phenotype is shown on the right hand side. 63 - 102 nuclei were imaged per sample, biological replicates = 3

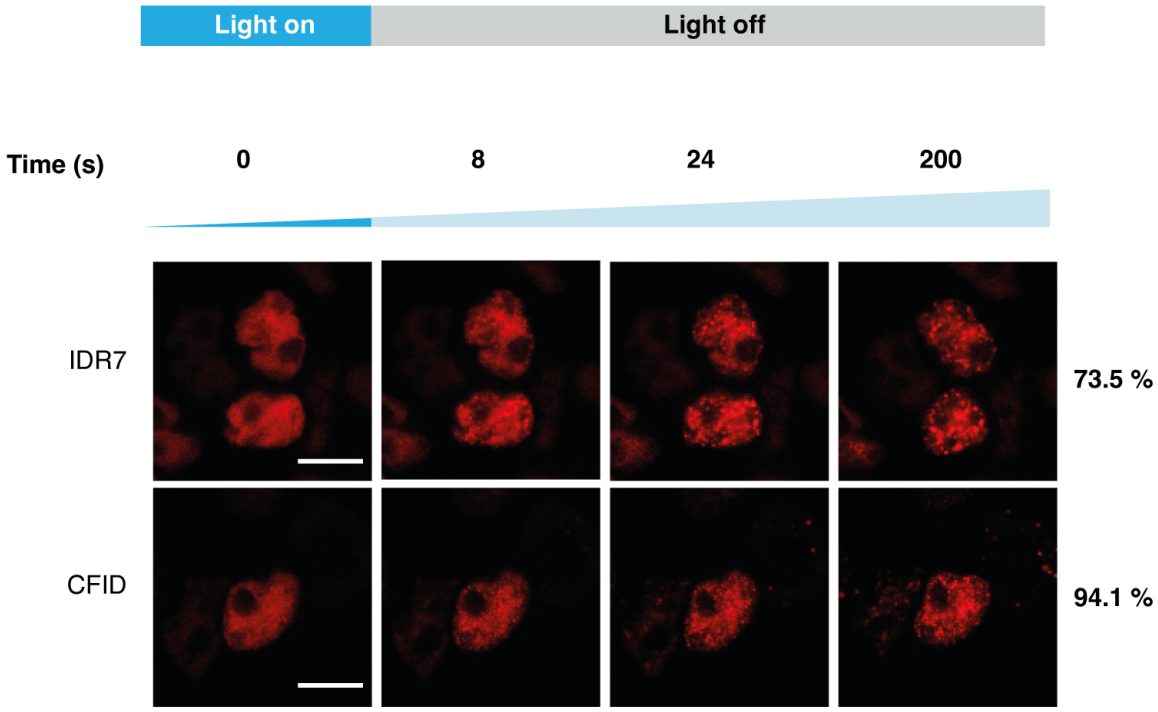

Figure 4.5: Expression of regions within CBP which form phase separated condensates using the optoDroplet system
Representative images of expressed regions within CBP including intrinsic disorder region 7 and the CBP FUS interaction domain (CFID) which form puncta upon blue light exposure over the timecourse. Time is shown above in seconds (s). Images were processed in ImageJ/Fiji with a scale bar of 10 μm. Percentage of nuclei transfected which display the represented phenotype is shown on the right hand side. 83 - 101 nuclei were imaged per sample, biological replicates = 3

4.3.4 Quantification of the optoDroplet data

In order to quantitatively compare the different phase separating behaviours of the regions of interest identified in CBP, I wanted to quantify the difference in puncta formation between the different constructs. The optoDroplet assay is not typically quantified, in one example where quantification was performed a region of interest was drawn in the area to be quantified and fluorescence intensity of that region was monitored over the timecourse (Shin et al. 2017). This is challenging when looking at nuclear proteins using live cell imaging, as the cells can move during the timecourse, causing the nuclei to drift in the imaging plane which would result in differences in

fluorescent intensity not caused by puncta formation. This is further amplified as these constructs do not enter the nucleolus which appear as dark circles in the nucleus (Figure 4.4), meaning that any drift in the sample may cause large changes in fluorescence intensity if the nucleolus enters the region being quantified. I therefore wanted a method to quantify changes in the whole nucleus over the timecourse that did not look at the raw gray values.

To do this, I used the standard deviation of pixels across the timecourse, which was measured in ImageJ/Fiji (Schindelin et al. 2012). For this analysis a threshold was applied to the first image in a timecourse to generate a region of interest that covers the whole nucleus. This region was then applied to all images in the timecourse and the standard deviation was measured within that region; where the standard deviation is the variation in gray values, or intensity, of the pixels within the nucleus. Over the timecourse the standard deviation will increase if puncta are formed, as some of the gray values will increase as the puncta are brighter and more intense, compared to the regions that have no puncta where the gray value will decrease. For this quantification I assume that the nucleolus stays relatively consistent throughout the timecourse, so that the effect of the nucleolus on the standard deviation was consistent. The standard deviation over the timecourse was then normalised to the first image, and a graph was produced comparing the fold change in standard deviation over the timecourse (Figure 4.6 A). An increase in standard deviation over the timecourse was observed with the positive control FUS, with an average plateau fold change of approximately 1.6 (Figure 4.6 A). A slight increase in standard deviation was observed over the timecourse with the negative control opto-NLS, with an average fold change of approximately 0.25 (Figure 4.6 A). Quantifying the regions of interest within CBP, both CFID and IDR7 show an increased change in standard deviation compared to the negative control, with plateaus of approximately 0.8 and 0.4 respectively, which is indicative of a puncta forming construct. In contrast, IDR1, IDR3, IDR4, AIL and the IDR6 show a similar increase in standard deviation to the negative control, with IDR6 showing the smallest change in standard deviation, this indicates that these regions do not have the propensity to form condensates independently.

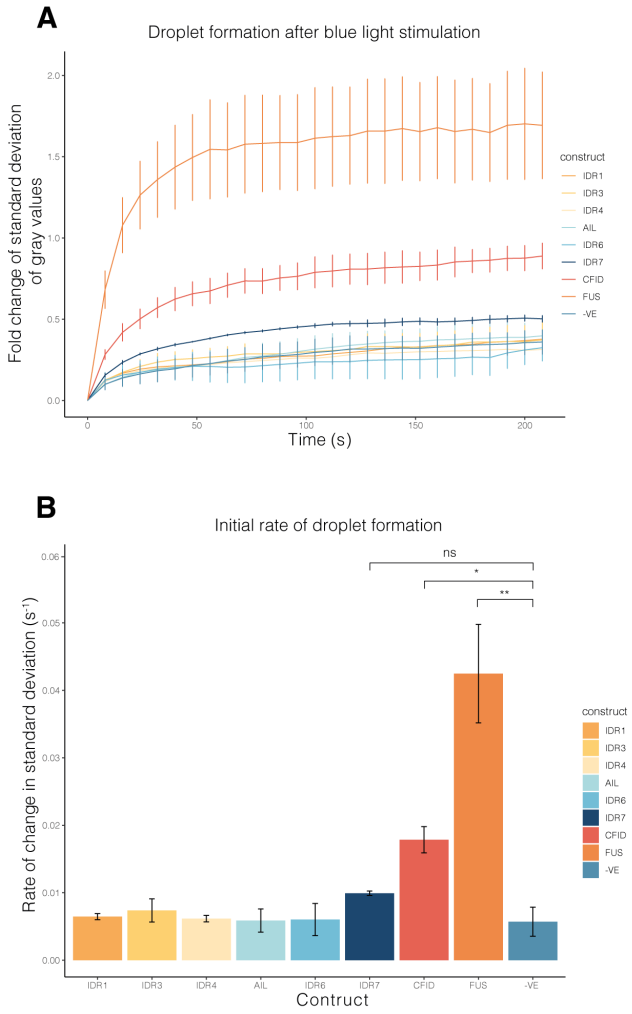

Figure 4.6: Quantification of the optoDroplet assay
A - Quantification of the fold change of standard deviation of gray values over the duration of the 4 minute timecourse, error bars are standard error of the mean (SEM). B - Bar graph to show the initial rate of change per second of standard deviation over the first 30 seconds of imaging, error bars are SEM. Statistics were performed using GraphPad Prism, where a unpaired t test was used to determine statistical significance. Stars represent significance with $p > 0.05$ = ns, $p < 0.05$ = *, $p < 0.01$ = **. Number of nuclei sampled 62 - 102 per construct. Biological replicates = 3. Graphs were produced using R in Rstudio.

From Figure 4.6 A, the majority of change in standard deviation happens in the first 30 seconds of the timecourse, therefore I wanted to look at the initial rate of puncta formation using

the rate of change of standard deviation per second, over the first 30 seconds. These results show a signficantly increased rate of change of standard deviation for FUS and the CFID (Figure 4.6 B). While visually it was observed that IDR7 formed puncta over the timecourse, and an increased rate of change of standard deviation was observed, this was not significantly increased compared to the negative control. All other constructs do not show a significantly different rate of change compared to the negative control (Figure 4.6 B).

The optoDroplet assay has identified that IDR7 and the CFID can both form puncta, and may play a role in CBP phase separation. Since the CFID is a large part of the C-terminus which encompasses most of IDR6 and all of IDR7, it is interesting that the CFID forms puncta and IDR6 does not. This suggests that IDR7 is the main driving factor in the phase separation of CBP. In contrast, the IDR1, IDR3, IDR4, AIL and IDR6 do not form puncta in the optoDroplet system and are therefore not driving factors in CBP phase separation.

Since it has been established whether these specific regions can form puncta independently of the rest of CBP, I next wanted to determine the effect that deleting these regions may have on CBP phase separation.

4.4 Do IDRs affect the phase separation of full length CBP

To understand the influence of these regions on full length CBP phase separation I wanted to investigate how CBP behaves when these regions are removed. Therefore I generated deletion mutants of CBP-GFP where they are missing these intrinsic disorder regions. For intrinsic disorder regions 1 - 7 and the CFID these regions were simply deleted, as the regions themselves have very little structure, in areas that are unlikely to affect catalytic function. In the case of the AIL, which is an unstructured loop that residues in the middle of two alpha helices in the HAT domain, more care needs to be taken in removing this sequence. Deleting this region would likely change the structure of the HAT domain, an essential part of CBP, having unknown consequences to function as well as phase separation. Therefore instead of deleting the region, I replaced it with a serine-glycine linker, Gly-Ser-Ala-Gly-Ser-Ala-Ala-Gly-Ser-Gly-Glu-Phe (GSAGSAAGSGQF), this would allow the helices to adopt their usual conformation while allowing the investigation of the importance of this region on phase separation. Domain maps and sizes of the resulting constructs can be found in Figure 4.7.

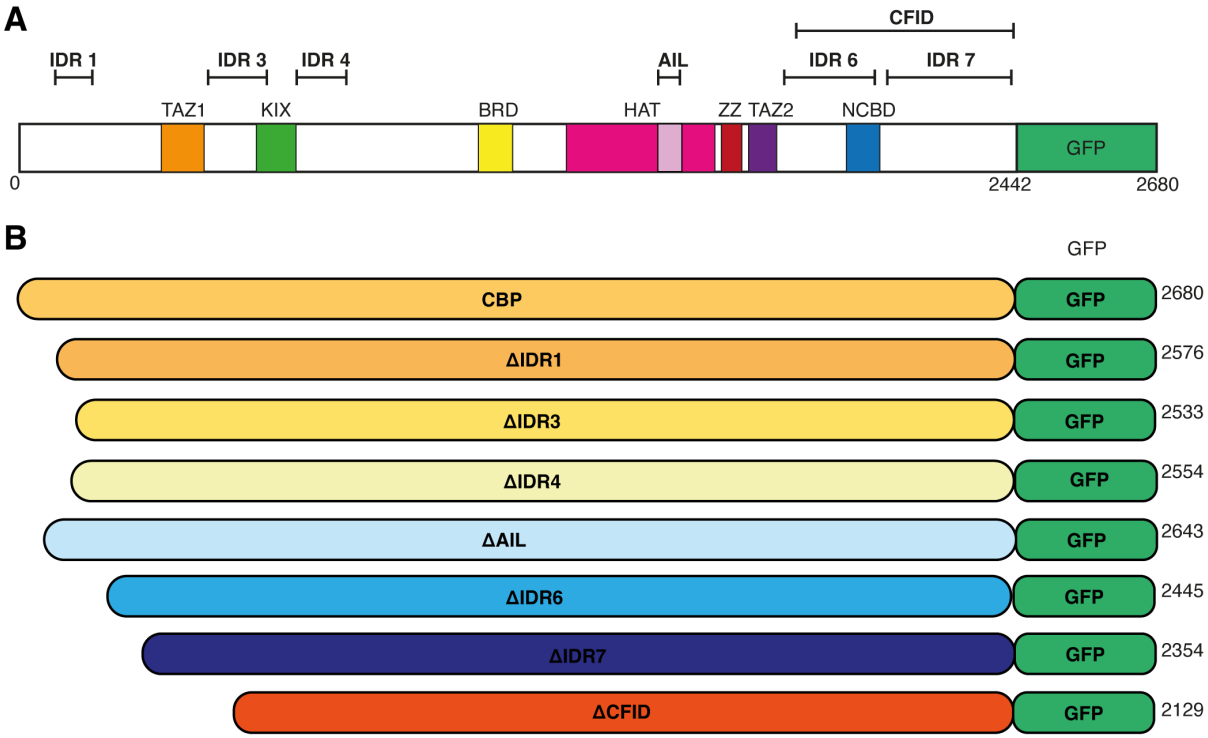

Figure 4.7: Domain maps for CBP-GFP and deletion mutants
A - Domain diagram of CBP showing globular structured domains including: transcriptional adapter zinc binding motif (TAZ1, orange) and (TAZ2, purple); CREB binding domain (KIX, green); bromodomain (BRD, yellow); histone acetyltransferase domain (HAT, pink); zinc binding domain (ZZ, red) and nuclear receptor coactivator binding domain (NCBD, blue). Image based on one found at (Dyson and Wright 2016). Image also includes boundaries of regions of interest, shown above which are being deleted for the overexpression system. B - maps of CBP-GFP, and size in amino acid of the final protein including the GFP tag, as well as deletion mutants of regions including intrinsic disorder regions 1, 3, 4, 6, 7 the autoinhibitory loop (AIL) and the CBP FUS interaction domain (CFID).

The resulting constructs were then used in the overexpression model system to determine if they from phase separated condensates. Briefly, HEK293T cells were transfected after 24 hours with GFP tagged constructs, and then fixed 24 hours post transfection. Z-stacks were taken of each construct, where a series of images taken at 0.5 μm increments through the nucleus to get a representation of the number of condensates throughout the nucleus. Maximum projections of these Z-stacks were then made in ImageJ/Fiji (Schindelin et al. 2012). As previously seen,

CBP-GFP forms distinct nuclear puncta when overexpressed (Figure 4.8). Within our mutant population there are several different phase separation behaviours observed. $CBP_{\Delta IDR1}$-GFP, $CBP_{\Delta IDR3}$-GFP, and $CBP_{\Delta IDR4}$-GFP all show a similar phenotype to wildtype with distinct nuclear puncta (Figure 4.8). Slightly larger and fewer puncta were observed with $CBP_{\Delta AIL}$-GFP and $CBP_{\Delta IDR6}$-GFP showing larger brighter puncta than CBP WT or any other mutant (Figure 4.8). In contrast, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP both show a diffuse pattern, which is not what is expected for a protein that phase separates (Figure 4.8). These results are consistent with those observed with the optoDroplet, futher suggesting that the IDR7 region is essential for phase separation.

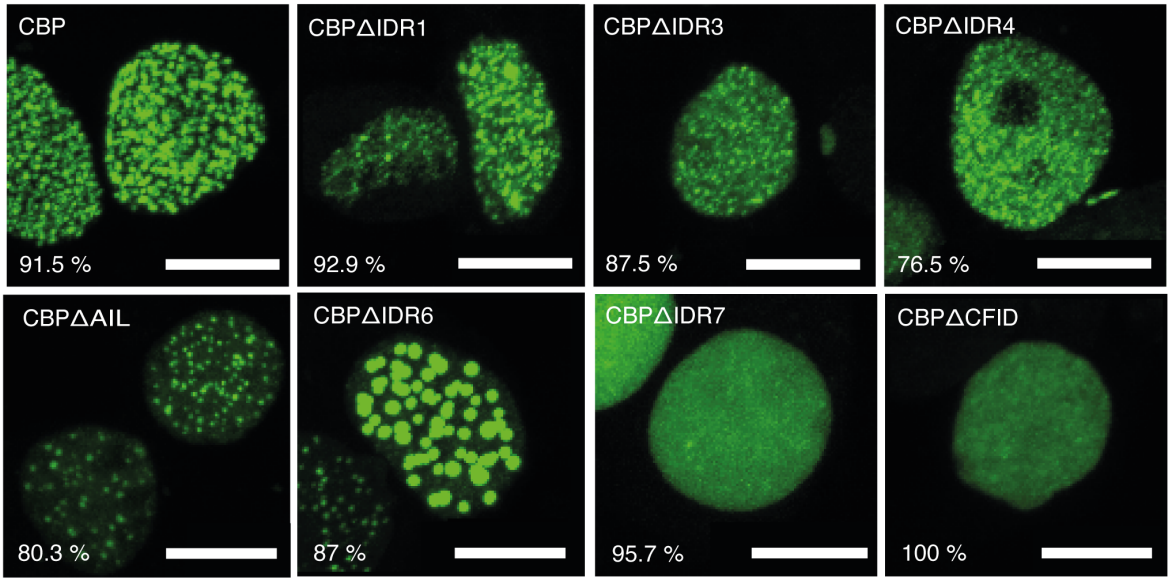

Figure 4.8: Overexpressed CBP-GFP and mutant constructs
Representative images of CBP-GFP and GFP tagged mutants in HEK293T cells. Mutated cells include intrinsic disorder regions (IDR), the autoinhibitory loop (AIL) and the CBP FUS interaction domain (CFID). Images represent a maximum projection of a Z-stack. Percentage of the transfected nuclei which display the representated phenotype are shown in the bottom left corner. Images were processed in ImageJ/Fiji, scale bar of 10 μm. 28 - 47 nuclei were sampled per construct. Biological replicates = 3.

Although I transfected the same amount of plasmid DNA for each construct, it wasn't clear whether these behaviours were a consequence of different expression levels or differences in the sequences themselves. Especially because the constructs are different sizes, transfection is more ef-

ficient with smaller constructs and therefore differences in the observed phenotype could be caused by different expression levels. It is also possible that these differences in phase separation could be a result of the different mutants having different saturation concentrations, therefore altering their ability to phase separate at these concentrations. As concentration is an extremely important parameter for phase separation, I didn't know at this stage whether increases or decreases in expression were affecting the phase separating behaviours of $CBP_{\Delta IDR6}$-GFP or $CBP_{\Delta IDR7}$-GFP for example.

4.4.1 Differences in expression have minimal effect on different CBP condensate behaviours

To determine whether the different behaviours were caused by differences in overall expression or were a property of the mutated constructs I used a hybrid western blot approach. This technique involved plating HEK293T cells in 10 cm dishes that contain 2 poly-L-lysine coated coverslips, the cells were plated for 24 hours before transfection of the plasmid DNA. 24 hours after transfection the coverslips were removed from the dish, fixed and mounted onto slides and the remaining cells harvested for western blot to check for the expression of CBP, GFP and a loading control GAPDH. This particular approach was used to compare directly the expression observed by confocal imaging and how that translates to the expression that was observed by western blot.

The coverslips from the hybrid western blot were first imaged to confirm the expression of the constructs and check that the phenotypes observed correlate with the data previously captured. Images show expression of all constructs, with phenotypes corresponding to that previously observed and described in section 4.4 (Figure 4.9 A). $CBP_{\Delta IDR1}$-GFP shows the lowest level of expression, in contrast to C-terminal mutants $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP and GFP-NLS which have a high level of expression within the large number of transfected nuclei (Figure 4.9 A). The negative control shows no GFP expression within the nucleus, with only background autofluorescence visualised in the cytoplasm.

The results from the western blot probing for CBP (Cell Signalling, 7389) showed protein expression for all samples (Figure 4.9 B). Both GFP-NLS and the negative control show endogenous expression of CBP at 300 kDa, with an increased expression of CBP for CBP-GFP and all CBP tagged mutants (Figure 4.9 B). Not all constructs show equal expression of CBP, with $CBP_{\Delta IDR1}$-GFP having lower expression than CBP-GFP WT, and $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP having particularly increased expression (Figure 4.9 B). This could be due to the difference in size of these C-terminal mutants, as they are smaller than WT this may

lead to increased expression of these particular constructs.

The same pattern was observed when probing for GFP (Abcam, ab290) for all constructs (Figure 4.9 B). No GFP was present in the GFP-NLS control and in the untransfected cells at the height of CBP, suggesting that the bands observed for the transfected constructs are the tagged proteins (Figure 4.9 B). For GFP-NLS a large band was present at 27 kDa which acts as a positive GFP antibody control (Figure 4.9 B). Normalisation was done using GAPDH (Proteintech, 60004-1).

As I was testing whether differences in phase separation behaviour were due to expression levels, the minor differences in expression between constructs was not crucial. C-terminal deletion mutants $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP were all expressed more than WT but to a similar level to each other, yet very different properties in their phase separating behaviour were observed. Two of the highly expressed constructs are $CBP_{\Delta IDR6}$-GFP and $CBP_{\Delta CFID}$-GFP, the first of which displays large condensates, where the latter behaves in a diffuse manner. With the caveat that this western blot is analysing a hetrogeneous population, it is possible to infer that differences in phase behaviour observed in this overexpression model system are not due to expression but are intrinsic properties of the constructs themselves.

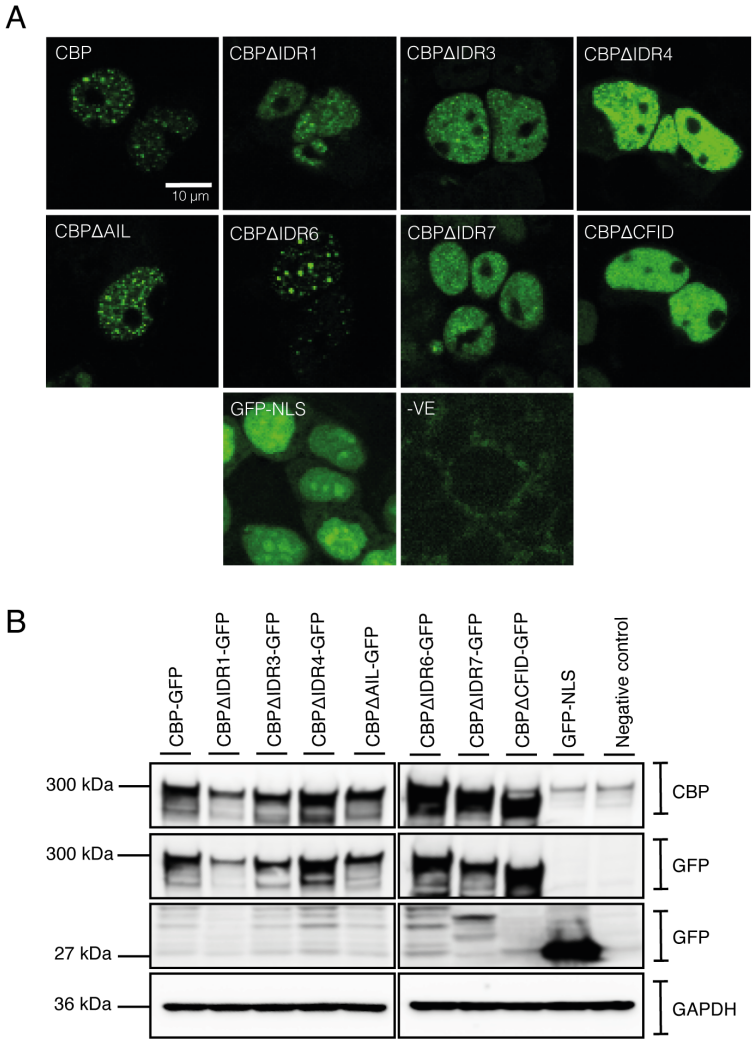

Figure 4.9: Hybrid western of overexpressed CBP-GFP and mutants
Hybrid western approach showing representative images and expression of protein by western blotting of CBP-GFP and GFP tagged mutants including intrinsic disorder regions (IDR), the autoinhibitory loop (AIL) and the CBP FUS interaction domain (CFID). Including GFP-NLS as a positive GFP control, and untransfected cells as a negative control. A - Representative images taken in the mid plane of an imaging field, from the cells used to perform a western blot. Images were processed in ImageJ/Fiji, scale bar represents 10 μm. B - Differences in expression observed by western blot. Approximate sizes of bands in kDa shown on the left, with the protein the antibody targets being shown on the right hand side. Images were processed in ImageJ/Fiji.

4.4.2 Developing a quantification pipeline for condensates formed by CBP and its mutants

In order to quantify differences in condensates, a custom ImageJ/Fiji macro was designed and applied to each nuclei within the dataset to measure regions of high intensity, indicative of condensates (Schindelin et al. 2012). Briefly, this macro makes a maximum projection from a Z-stack (compiling all of the images into one plane). From this, a threshold intensity filter was applied to identify puncta which have an intensity above that threshold. This threshold was manually set for each experimental repeat. The analyse particles function in ImageJ/Fiji then defined the regions of interest that passed the threshold and quantified each ROI for parameters such as intensity, area and circularity (Schindelin et al. 2012). The quantification was then combined for all experimental repeats and the data visualised using R in Rstudio (Posit team 2023; R Core Team 2020).

Preliminary analysis of this data revealed a number of cellular characteristics which were picked up in the ImageJ/Fiji macro that were not characteristic of individual condensates and therefore should be filtered out of the dataset. These were: nuclei which did not pass the threshold value, overexpressed nuclei, regions of high intensity that due to size cannot be a phase separated condensates, and clumps of puncta idenitified as one ROI. I therefore designed a filtering strategy to remove these datapoints, to compare the different condensate forming abilities of CBP deletion constructs (Figure 4.10).

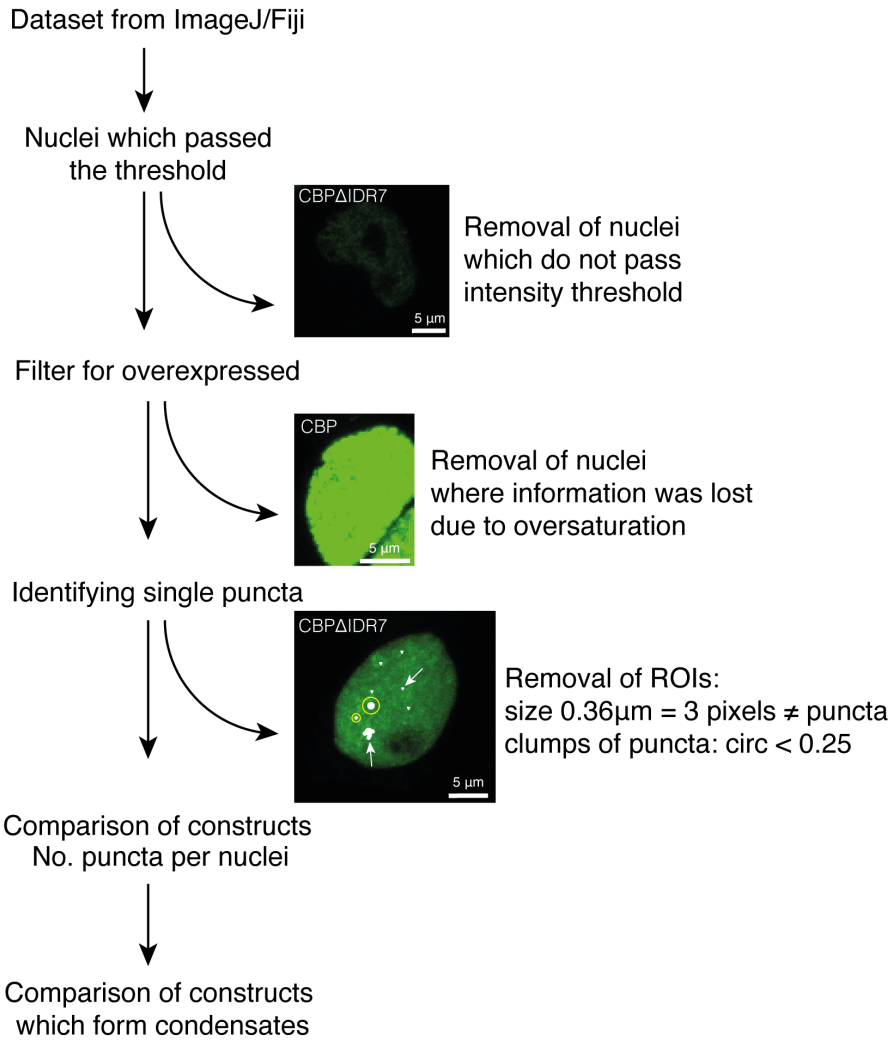

Figure 4.10: Filtering pipeline for CBP deletion constructs

Pipeline for filtering overexpression dataset to compare the different condensate behaviour observed between CBP-GFP deletion constructs. Example images are shown which represent the filtered parameters, images were processed in ImageJ/Fiji. White arrows represent 3 pixels which do not constitute a puncta and clumps of puncta that were filtered out of the dataset. Yellow circles represent puncta kept for further data analysis.

4.4.2.1 Selecting for nuclei which had passed the threshold

The first stage of filtering this dataset was to remove any nuclei that did not pass the threshold value (Figure 4.10), therefore nuclei which did not show an intensity reading once the manual threshold was applied. For all samples, over 50 % of nuclei imaged passed the threshold value, with mutants $CBP_{\Delta IDR1}$-GFP and $CBP_{\Delta CFID}$-GFP having the lowest percentage of nuclei retained after thresholding (Figure 4.11). This could indicate a lower transfection rate than the other samples, for example $CBP_{\Delta IDR1}$-GFP correlates with the western blot which showed lower expression than all other constructs, and a lower number of nuclei which passed the threshold (Figure 4.11). In comparison, for the CFID mutant, as the pattern is more diffuse, the intensity of the whole nucleus is lower than that of the puncta forming constructs; a lot of nuclei therefore didn't pass the initial threshold as they did not have any regions with high enough intensity. The remaining quantification was performed only on the nuclei that had passed the threshold value.

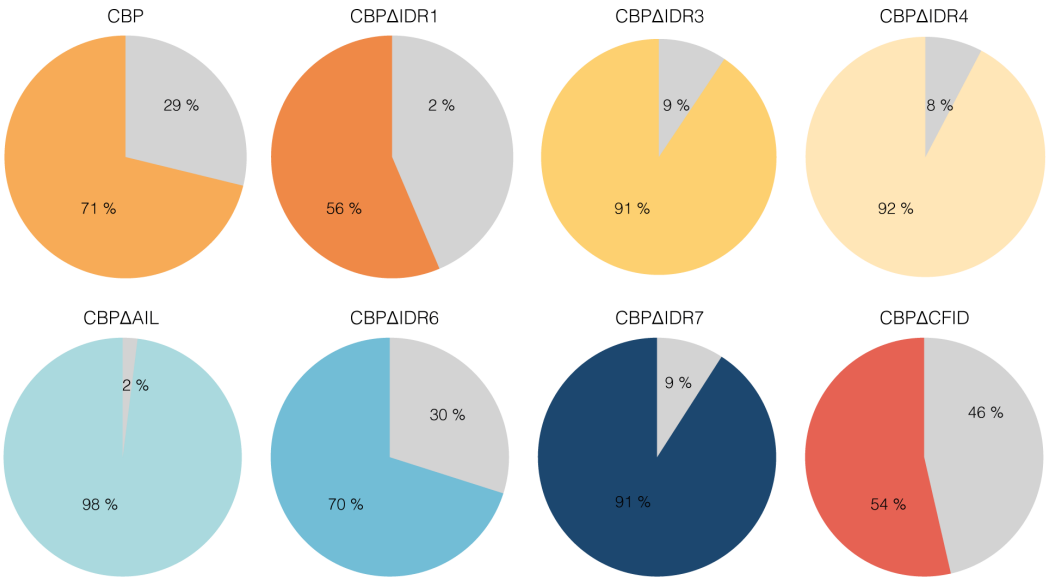

Figure 4.11: Number of nuclei that passed the threshold per sample for overexpresion of CBP-GFP deletion mutants
Proportion of nuclei sampled per construct which showed an intensity reading above the threshold, compared to nuclei which did not pass the threshold (grey) shown via piechart. Graphs were made in Rstudio using R. Number of nuclei sampled per construct ranged from 28 - 47 nuclei. Biological replicates = 3.

4.4.2.2 Applying filters to the dataset

Looking only at the transfected nuclei, I next wanted to visualise the whole dataset and further apply filters so that I could identify individual puncta (Figure 4.12 A). This data compares the area and the circularity of puncta, where each individual circle represents a puncta, and the size of the circle corresponds to the integrated density (area of the puncta multiplied by the intensity of the puncta). In this graph, the data was not separated by sample and contains all puncta from every sample. From this dataset I was able to determine cut off values to isolate only the single puncta (Figure 4.12 A). First of all, I wanted to isolate overexpressed nuclei, since the maximum gray values have been exceeded no information can be gathered from this data, I therefore set an area threshold of 10 to isolate nuclei which were overexpressed (dark purple). I also wanted to rule out regions of high intensity which do not correlate to puncta, this was especially important for the $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP which do not form phase separated condensates. From ImageJ/Fiji I calculated that each pixel was 0.12 μm in size, the smallest reading of puncta observed is 0.36 μm which was equivalent to 3 pixels. I decided that a reading of 3 pixels or fewer does represent a high region of intensity but does not qualify as a phase separated condensate (pink). As I was only interested in the individual puncta, I also wanted to rule out clumps of puncta that were being counted as one in the thresholding process, as this may skew the data based on size. To do this I applied a circularity threshold of 0.25, as puncta are typically circular I decided that any puncta which had a circularity of 0.25 or smaller were irregular in shape and therefore representative of clumps of puncta (blue) (Boeynaems et al. 2018; Alberti, Gladfelter and Mittag 2019; Banani et al. 2017; Brangwynne et al. 2009). From these applied filters it is the purple "puncta" which I used in further data analysis.

Before using this criteria I wanted to check how these filters apply to all the samples, and therefore looked just at the integrated density per sample, with these filters applied (Figure 4.12 B). This graph suggests that the filters were affecting each of our samples consistently, with every construct showing puncta that were classified as pixels, puncta, clumps or overexpressed in their expected positions (Figure 4.12 B). Therefore the isolated puncta fraction can be used for further data analysis to compare the behaviour of condensates. From this point on it is the filtered puncta dataset that has been used.

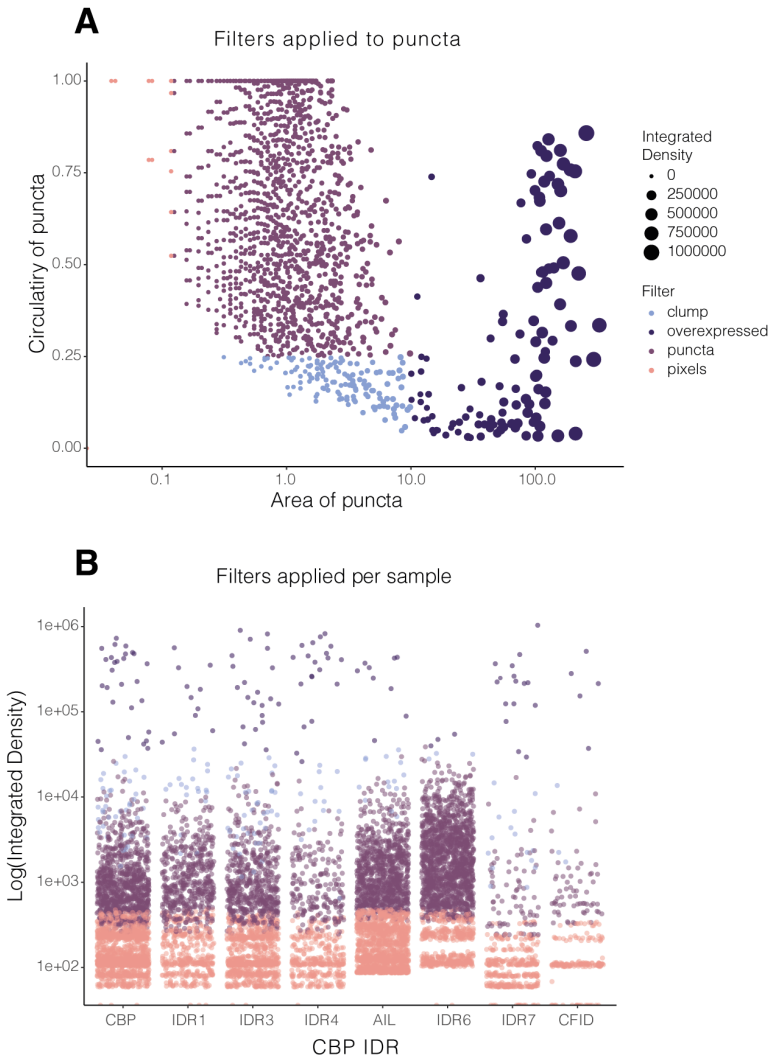

Figure 4.12: Filters applied to overexpression of CBP-GFP and mutants
A - filters applied to the whole dataset based on circularity and area of puncta. Each circle represents a puncta, where the size is correlated to the integrated density of the puncta (intensity multiplied by the area). The colours relate to the filters applied to the sample blue represents puncta which are clumped together (circularity below 0.25), dark purple represents overexpressed nuclei (area above 10), purple represents individual puncta and pink represents intensities which are three pixels or less and therefore cannot be confirmed as a puncta (area of 0.36 μm). B - filters as described in A shown per sample, using the log value of the integrated density, where each dot represents a single puncta reading. Graphs were made using R in Rstudio. Number of nuclei sampled per construct ranged from 28 - 47 nuclei. Biological replicates = 3.

4.4.2.3 Isolating constructs which form puncta

The aim of this quantification was to detect differences in puncta observed between CBP and the CBP mutants. However some of these constructs, such as $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP, don't form visible condensates in the same manner as CBP-GFP. Therefore I wanted to detect this in the data and filter to isolate only the constructs which form visibile puncta. The main differences observed between constructs that appear to form condensates and those that do not, was the number of nuclei identified to have puncta and the number of condensates identified per nucleus. I first looked at the distribution of the number of puncta within all nuclei for each sample (Figure 4.13 A). It was observed for $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP that a smaller number of nuclei have passed thresholding, and of those that remain the number of puncta per nuclei was generally lower than other construct (Figure 4.13 A). To complement this I looked at the median number of puncta per construct via boxplot (Figure 4.13 B). This data shows all constructs have a median number of puncta above 5 (indicated by the black dotted line) except for $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP (Figure 4.13 B). From this I concluded that a minimum of 5 puncta per nuclei was necessary to be classed as a condensate forming construct. Using these parameters, I decided to not include $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP in the comparison of puncta from CBP deletion constructs compared to WT. Since these constructs don't appear to form condensates in the same manner as the other CBP deletion mutants.

Having determined that deletion of IDR7 and CFID largely abolishes the ability of CBP to form puncta, the remaining constructs were also defined as a population of constructs that maintain the ability to form condensates. I continued to analyse the quantification to see how condensates produced by the different phase separating mutants compare to CBP WT.

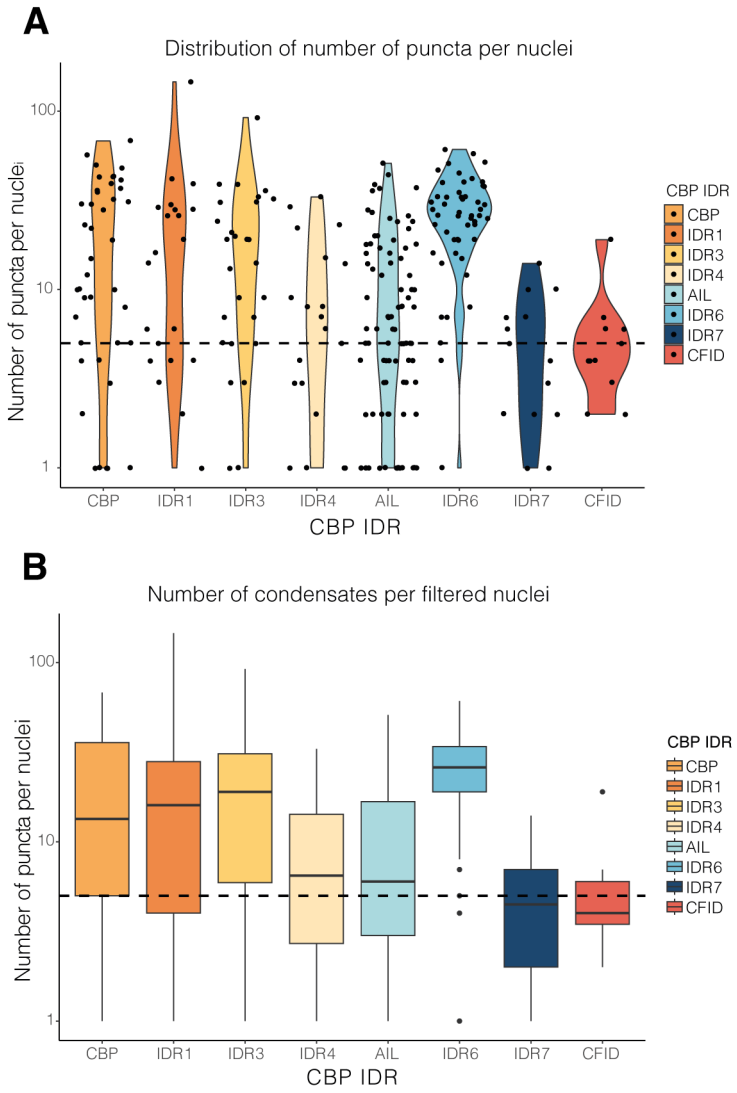

Figure 4.13: Comparing the number of puncta per nuclei from CBP deletion constructs A - violin plot to represent the distribution of number of puncta from filtered nuclei per sample. X-axis represents the sample and the Y-axis represents the number of puncta shown on a logarithmic scale, each dot represents a single nuclei. The dotted line represents the cut-off for puncta forming constructs which was 5 puncta. B - boxplot to summarise number of puncta data per sample, the median was used to determine the cut-off for puncta forming constructs, represented by the dotted line which is placed at 5 puncta. Y-axis is shown on a logarithmic scale. Graphs were made using R in Rstudio. Number of nuclei sampled per construct ranged from 28 - 47 nuclei. Biological replicates = 3.

4.4.3 Comparing the different puncta of condensate forming constructs

I wanted to compare the differences between puncta in my overexpressed constructs, to do this I used the integrated density, which is the area multiplied by the mean gray value. At this stage I had filtered out nuclei that were overexpressed, values which were too small to be puncta or that were clumped together. Constructs $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP were not included in this comparison as they do not form condensates in the same manner as WT. In the previous section I determined that nuclei that have 5 puncta or fewer were characterised as non-condensate forming, therefore I applied an extra filter step to remove any nuclei containing 5 or fewer puncta from this dataset.

The first thing that I did was to look at the distribution of the data, as I wanted to perform statistical testing between the different CBP constructs; the distribution would determine which type of statistical testing to perform. I determined that this data does not conform to a normal distribution or gaussian curve and therefore I used non-parametric tests to do statistical analysis. A non-parametric approach accounts for differences in distribution away from the normal, which would utilise the median value, of integrated density in this case, to do statistics.

At this point, each sample had a different number of nuclei, each of which had a different number of puncta with variable integrated densities. Lots of variation was observed between the different constructs, some nuclei had a larger number of puncta that were smaller, while others had fewer puncta that were evenly distributed in integrated density; I therefore needed a way to compare the constructs without the number of puncta skewing the results. To address this I decided to use the median integrated density per nuclei, as this data was not normally distributed, and I was using the median for statistical comparison. On this data I performed Kruskal-Wallis statistical testing to compare the median integrated density of puncta per nuclei between CBP WT and each of the different constructs.

The results show that there was no statistical difference between the puncta of CBP-GFP, $CBP_{\Delta IDR1}$-GFP, $CBP_{\Delta IDR3}$-GFP and $CBP_{\Delta IDR4}$-GFP, therefore the puncta produced by these constructs were similar in size and intensity to WT (Figure 4.14). In comparison, $CBP_{\Delta AIL}$-GFP and $CBP_{\Delta IDR6}$-GFP had a higher integrated density to WT, with a statistically significant p-value of 5.9e-08 (Figure 4.14). These puncta were larger and brighter than observed for the WT construct, suggesting that the AIL and the IDR6 region may be important for influencing the phase separation behaviour of full length CBP.

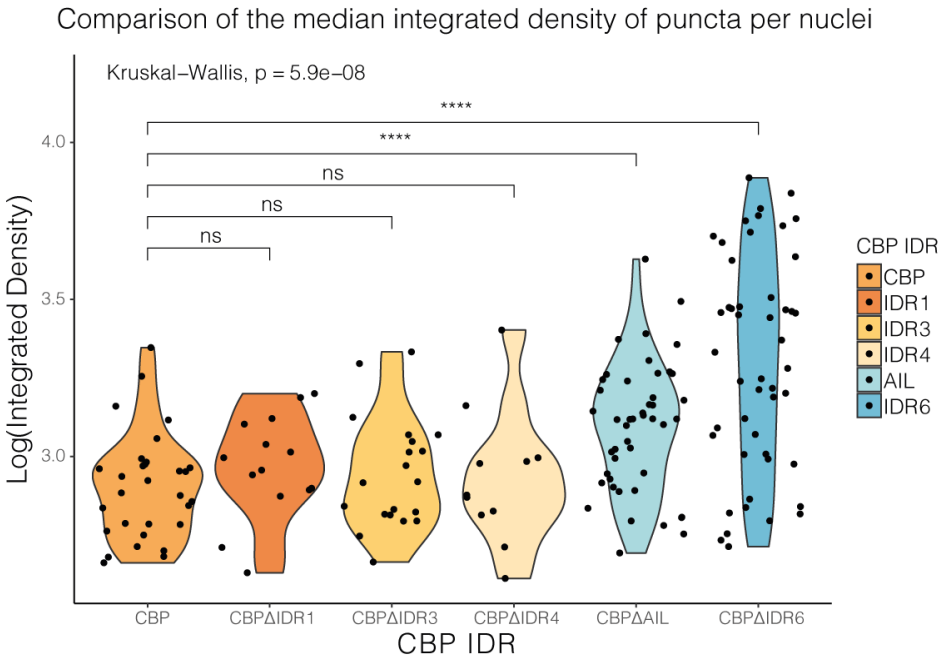

Figure 4.14: CBP and mutant puncta have different integrated densities
Violin plot comparing the median integrated density of puncta per nuclei between CBP and deletion constructs. Each dot represents the median integrated density of an individual nuclei, with the number of dots representing the number of nuclei sampled. Kruskal-Wallis statistical test was performed to compare CBP WT with each of the mutant constructs, with the violin representing the distribution. Graph was made using R in Rstudio. Stars represent significance with $p > 0.05$ = ns, $p < 0.0001$ = ****. Number of nuclei sampled per construct ranged from 28 - 47 nuclei. Biological replicates = 3.

This data highlights four key regions of interest when studying CBP phase separation; the AIL, IDR6, IDR7 and the CFID region. While the AIL and IDR6 region did not phase separate independently in the optoDroplet assay, in their absence the ability of CBP to phase separate was increased, with larger and brighter puncta than observed in the WT construct. This suggests that these regions may have some inhibitory effect on the overall ability of CBP to phase separate. In contrast, the IDR7 region does phase separate independently in the optoDroplet system, but when removed the ability of CBP to phase separate was reduced. This data implicates IDR7 as a driver of CBP phase separation. The CFID region also behaves in the same manner as IDR7, even though it is comprised of most of IDR6 and all of IDR7 which have opposing phase behaviours;

as the IDR6 region inhibits phase separation and the IDR7 promotes it.

In conclusion, important regions of CBP have been identified which affect CBP's phase separating ability. While this data was a good indication of the effects these regions have on phase separation phenotypically, these were fixed samples and therefore I was only looking at a snapshot of the cellular environment. This data provided no indication of how these mutants behave in live cells and whether they are mobile, like phase separated condensates are expected to be. Therefore I next wanted to look at how these constructs behave in live cells, using some of the cellular techniques discussed in chapter 3.

4.5 Are CBP phase separation mutants mobile?

To determine if the condensates seen above were mobile or more fixed aggregates I wanted to use live cell assays to assess the fluidity of these constructs. To do this I utilised the techniques of 1,6 - hexanediol treatment and FRAP which I have used to show if they behave in a mobile manner and if these condensates are held together using hydrophobic interactions as would be expected from a protein that phase separates. From previous experiments, I focused on CBP-GFP and $CBP_{\Delta IDR6}$-GFP, which forms larger and brighter condensates compared to WT, this will act as an extreme of CBP phase separation to check mobility. Since $CBP_{\Delta IDR1}$-GFP, $CBP_{\Delta IDR3}$-GFP and $CBP_{\Delta IDR4}$-GFP have the similar phase separating ability to the WT protein, and $CBP_{\Delta AIL}$-GFP did not show as extreme a phenotype as CBPΔIDR6-GFP, I did not test these constructs in these assays.

4.5.1 1,6 - hexanediol treatment of CBP deletion mutants

As a quick method to visualise how fluid-like these proteins were and whether they were formed through phase separation as opposed to aggregation I applied the 1,6-hexanediol method, as was previously described and tested for CBP-GFP in section 3.3.3. In this experiment I focused on CBP-GFP and $CBP_{\Delta IDR6}$-GFP, which forms larger and brighter condensates compared to WT; I also tested $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP, which are more diffuse than WT CBP, to observe how 1,6-hexanediol affects a more diffuse protein. Both CBP-GFP WT and $CBP_{\Delta IDR6}$-GFP, show disruption of phase separated puncta after the addition of 1,6-HD over the first 20 seconds (Figure 4.15). In comparison, there was no change in the phenotype of $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP after the addition of 1,6-HD (Figure 4.15). A change in shape of nuclei across all constructs was observed upon the addition of 1,6-HD.

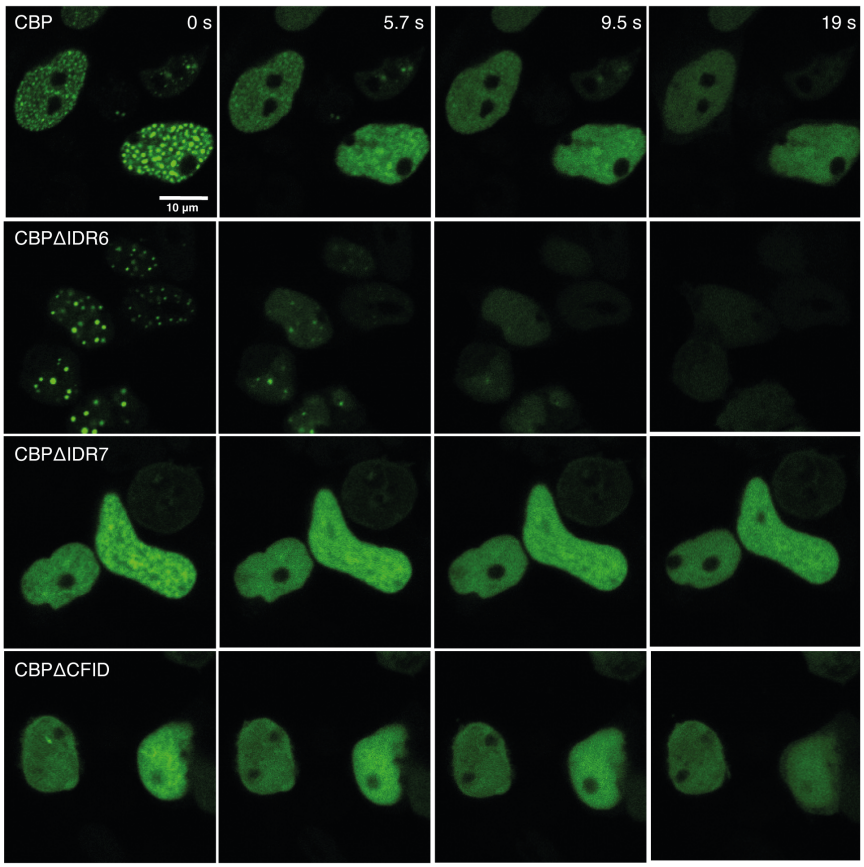

Figure 4.15: Effect of 1,6-hexanediol treatment on CBP and deletion mutants
Visualisation of the treatment of HEK293T cells expressing CBP-GFP and deletion mutants with 5 % 1,6-hexanediol. Time of 0 seconds indicates the addition of 1,6-hexanediol. Images were processed in ImageJ/Fiji, scale bar represents 10 µm. Nuclei sampled per construct ranged between 14 - 35. Biological replicates = 2.

To gain more information on the mobility of overexpressed CBP-GFP and mutants I wanted to performed FRAP to try and quantitate differences in dynamics between CBP WT and mutant constructs.

4.5.2 Assesing the mobility of CBP-GFP and CBP$_{\Delta IDR6}$-GFP by FRAP

To determine if the condensates observed by overexpression were mobile I used FRAP, as described in section 3.3.2. Since I only wanted to test the mobility of condensate forming constructs, I tested

CBP-GFP and CBP$_{\Delta IDR6}$-GFP which acts as an extreme of CBP phase separation. Visually, the data showed CBP$_{\Delta IDR6}$-GFP formed larger puncta in live cells than CBP-GFP WT, which correlates to the fixed cell data (Figure 4.16). The effect of photobleaching on CBP-GFP WT and CBP$_{\Delta IDR6}$-GFP was easily identifiable by the reduction in fluorescence intensity of the puncta, with recovery observed over the timecourse (Figure 4.16 A). While visually photobleaching and recovery was observed for both CBP-GFP WT and CBP$_{\Delta IDR6}$-GFP, I wanted to confirm this by quantifying the recovery rates.

Data for CBP-GFP is the same as was shown in Figure 3.8, but has been included here for comparison. CBP-GFP bleaches to approximately 50 % and recovers within 20 seconds; similarly, CBP$_{\Delta IDR6}$-GFP bleaches to 40 %, which was more than WT CBP but recovers over a similar time frame (Figure 4.16 B). CBP-GFP and CBP$_{\Delta IDR6}$-GFP both photobleach and recover over a timescale that is indicative of phase separated constructs (Gibson et al. 2019b; Monahan et al. 2017; Niaki et al. 2019; Ma et al. 2021; Wang et al. 2022b; Zhang et al. 2021), indicating confinement via the condensate and mobility of the protein via recovery over a 30 second time scale. This data shows that these constructs behave in a predictable manner in FRAP experiments, with a susceptibility to photobleaching that is dependent on the constructs ability to phase separate.

To compare the effect of photobleaching between CBP-GFP and CBP$_{\Delta IDR6}$-GFP, I calculated several rate constants from the FRAP recovery curve, this data is shown in Table 4.1. These values were derived from calculations of the raw data, as well as recovery curves plotted in GraphPad Prism (San Diego 2022).

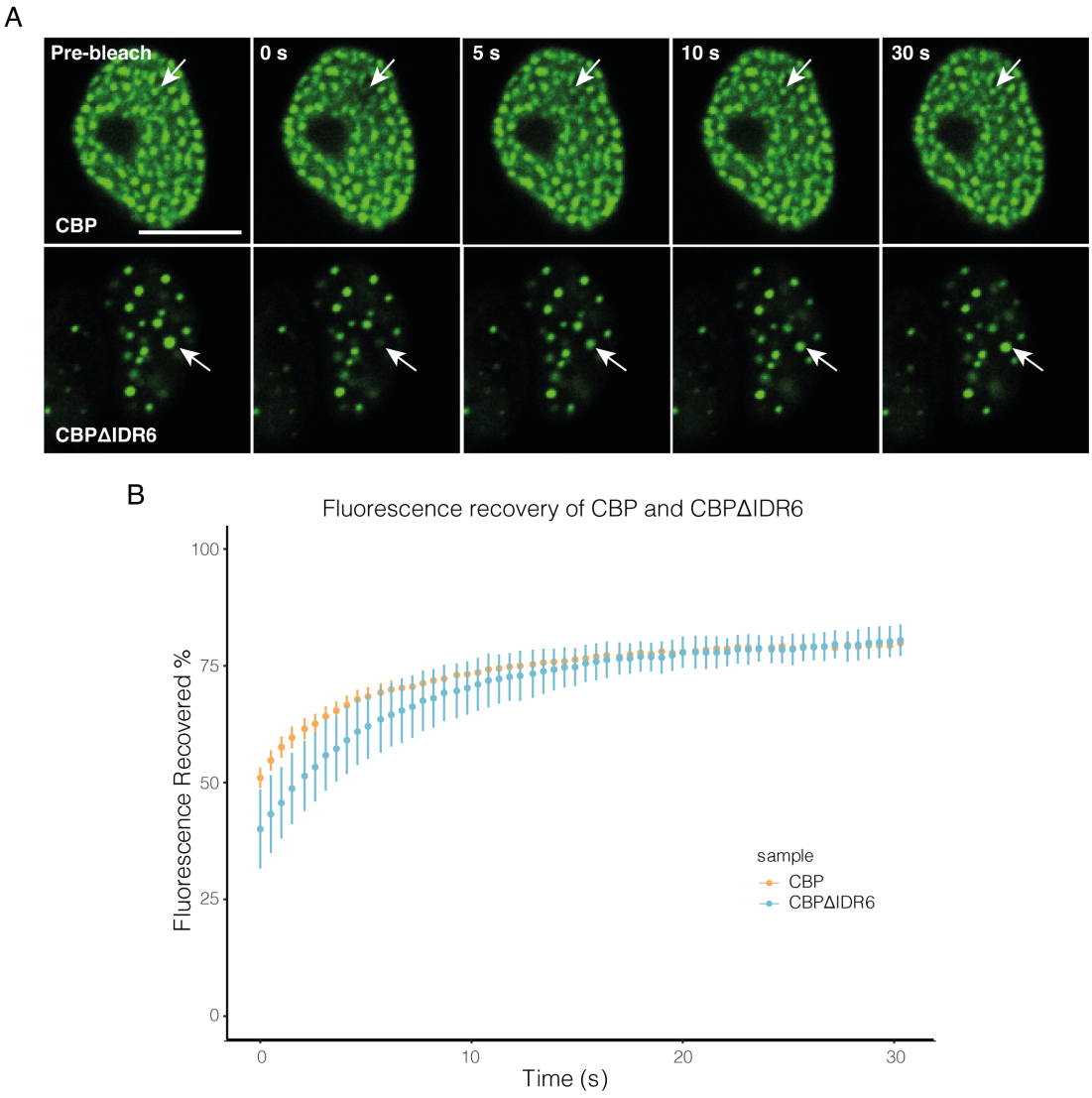

Figure 4.16: FRAP of CBP-GFP and CBP$_{\Delta IDR6}$-GFP

Fluorescence recovery after photobleaching (FRAP) of HEK293T cells transfected with CBP-GFP and CBP$_{\Delta IDR6}$-GFP. A - Representative images were taken in one plane of the nucleus, white arrow indicates where the laser was pointed for the photobleaching. Images were processed in ImageJ/Fiji, scale bar indicates 10 μm. B - Fluorescence recovery curve over 30 seconds post-bleaching. Points on the graph represent the average percentage fluorescent recovery with the error bars showing the standard error of the mean. Graph was plotted using R in Rstudio. Nuclei sampled per construct ranged between 18 - 31. Biological replicates = 3.

Sample	Initial Rate (s^{-1})	Rate Constant (s^{-1})	T = 1/2 (s)	Mobile Fraction
CBP-GFP	2.262	0.165	4.208	0.555
CBPΔIDR6-GFP	3.08	0.151	4.592	0.657

Table 4.1: Quantitative comparison of photobleaching of CBP-GFP and $CBP_{\Delta IDR6}$-GFP Quantitative data generated from the fluorescence recovery profiles generated from fluorescence recovery after photobleaching (FRAP) experiments. The initial rate was calculated over the first 10 seconds, showing the rate of recovery per second, whereas the rate constant is generated over the whole plot. Time or T = 1/2 is the time it takes for half of the fluorescence to recover in seconds. The mobile fraction refers to the proportion of molecules that undergo exchange with the FRAP region.

Initially I wanted to compare the initial rate of fluorescence recovery, referring to the rate of fluorescence recovery over the first 10 seconds. The rate for CBP WT was 2.262, which was increased when looking at $CBP_{\Delta IDR6}$-GFP to 3.08. This data was consistent with the plots observed in Figure 4.16 B. I then calculated the rate constant of the different recoveries, where a large rate constant is indicative of a fast recovery and a low rate constant is indicative of a slow recovery. Both CBP-GFP and $CBP_{\Delta IDR6}$-GFP have a similar overall rate constant, suggesting a similar recovery to photobleaching.

As well as looking at the rate of fluorescence recovery I wanted to directly compare the mobility of the different constructs. I did this in two ways; first identifying the T = 1/2, which is the time it takes half of the fluorescence to be recovered, and secondly by calculating the mobility fraction of the different constructs. Comparing the T = 1/2, a very similar pattern to the rate constants was observed. CBP-GFP has a T = 1/2 of 4.2 seconds, compared to 4.592 seconds for the large condensate forming $CBP_{\Delta IDR6}$-GFP.

One value I wanted to quantify was the mobility fraction, this is the proportion of molecules which are recoverable from the photobleached area. This was calculated using the starting intensity and the post-bleaching plateau, using the equation from Hallegger et al. (2021) which is fully detailed in the materials and methods section 2.3.1.4. CBP-GFP WT had a mobile fraction of approximately 0.555, compared to $CBP_{\Delta IDR6}$-GFP which had a higher proportion of intractable molecules at 0.657. These values were consistent with the FRAP recovery curve (Figure 4.16 B), and reflect their mobility based on their cellular phenotype.

These results suggest that CBP-GFP and $CBP_{\Delta IDR6}$-GFP both display a similar recovery

profile in response to photobleaching, with the caveat that incomplete photobleaching and recovery was observed. Together, the 1,6-hexanediol and FRAP data demonstrates that the observed puncta from CBP-GFP and $CBP_{\Delta IDR6}$-GFP display properties of phase separated condensates, as they behave in a mobile manner. This data highlights the different mutant constructs, therefore the IDRs, are influencing the phase separating behaviour of WT CBP.

4.6 Discussion

In this chapter I have explored intrinsic disorder regions within CBP to determine the importance of different regions on phase separation, revealing that it is not one region which is important for CBP phase separation but a combination of opposing forces.

4.6.1 Defining intrinsic disorder regions

For identifying regions of intrinsic disorder (IDRs) within CBP I used two different softwares DISOPRED3 and PONDR which showed the overall levels of disorder within CBP, both of which have been used in the literature to define unstructured regions (Xue et al. 2010; Jones and Cozzetto 2015; Basu et al. 2020). DISOPRED3 was used mainly to visualise an overall map of disorder, whereas PONDR looked more specifically at the prediction score per residue within the protein. For defining specific IDRs I used PONDR prediction, where stretches of 50 or more amino acids that have a probability of disorder over 50 % were defined as IDRs, this method was applied in the literature by Basu *et al.* (2020). Using this method I identified several regions, that ranged in size, and selected regions that covered the whole span of CBP, and termed them IDR1 - IDR7; with a range of sizes from 77 amino acids for IDR2 to a maximum of 325 amino acids for IDR7.

IDRs were previously investigated within p300, a paralog of CBP, using IUPred2 (Wang et al. 2022b). The prediction used within this paper defined all regions that were not structured domains intrinsic disorder regions - naming them IDR1 - 5. Comparing these to CBP some overlap is observed, with the IDRs detailed in this doctoral work being more restrictive (Figure 4.17).

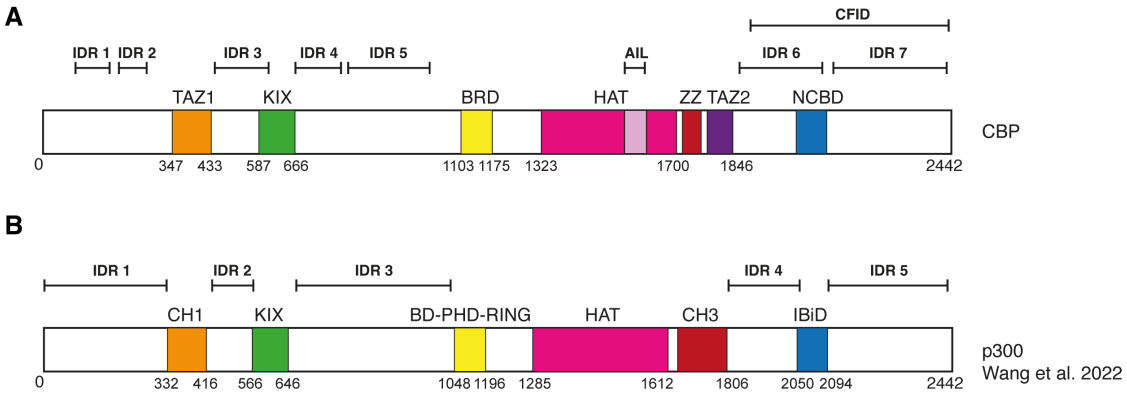

Figure 4.17: Comparison of IDR regions defined for CBP and p300 in Wang et al. 2022 Domain diagrams of CBP (A) and p300 (B). A - domain diagram of CBP including intrinsic disorder regions defined within this thesis, including the AIL region and the CFID. Image based on one found at Dyson and Wright (2016). B - domain diagram of p300 showing the IDRs described in Wang et al. (2022b). For both globular structured domains shown here include: transcriptional adapter zinc binding motif (CH1/TAZ1, orange), CREB binding domain (KIX, green), Bromodomain PHD RING domain (BD-PHD-RING, yellow), histone acetyltransferase domain (HAT, pink), CH3/TAZ2 and IRF-3 binding domain (IBiD/NCBD, blue).

Within the sequence of p300, taken from UniProt (Q09472), PONDR prediction can be performed using the same principles used for CBP, to define IDRs within p300 for comparative purposes (Figure 4.18). Both CBP and p300 show a similar profile of intrinsic disorder, highlighting both the N- and C-terminal regions as having a larger degree of disorder (Figure 4.18). For p300, 8 IDRs were identified compared to the 7 IDRs for CBP, which span similar areas of the protein. While the HAT domain is highly conserved between CBP and p300, divergence in the sequence is observed typically in these disordered regions, which may be why differences in the definition of the IDRs was observed (Chan and La Thangue 2001). This difference may account for any differences in phase behaviour in this work compared to that published for p300, which will be discussed in more detail in the following sections. For a sequence alignment of CBP and p300 see appendix, section 9.4.

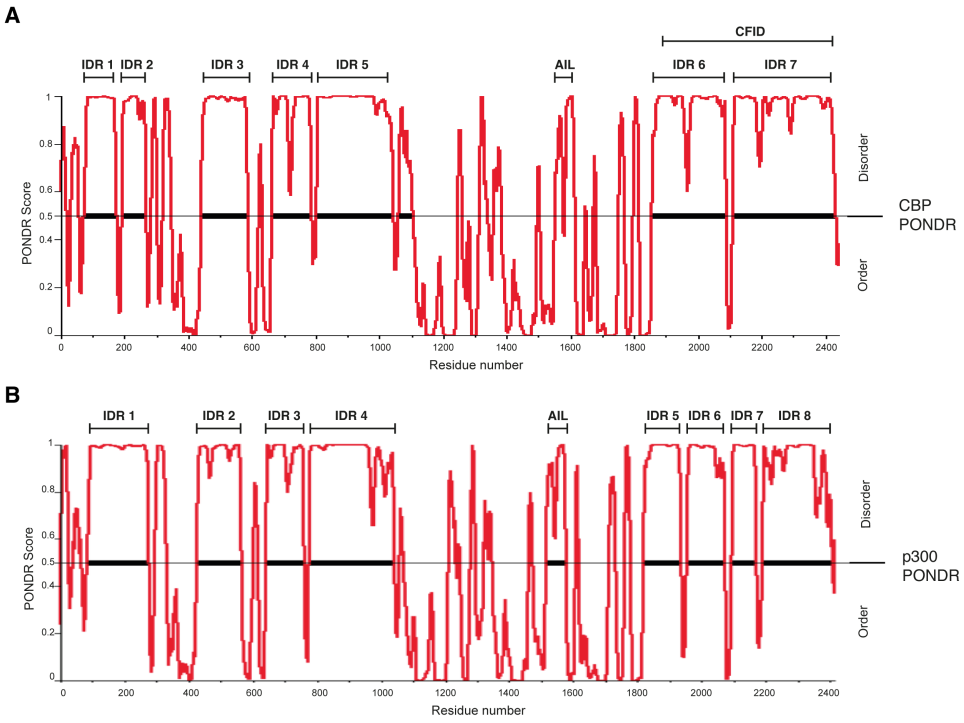

Figure 4.18: PONDR prediction comparison between CBP and p300
Disorder prediction plot showing the PONDR score, where a score of 1 represents regions of disorder across the amino acid residues within each protein. A - disorder prediction of CBP. B - disorder prediction of p300 (UniProt, Q09472). PONDR software was used to determine the IDR boundaries within this PhD thesis, with IDR regions indicated above. IDRs were defined as stretches of 50 or more amino acids with a disorder prediction of 0.5 or above, the autoinhibitory loop (AIL) is also indicated on each plot (Xue et al. 2010).

4.6.2 optoDroplet as a method to study the importance of individual intrinsic disorder regions

In this thesis the optoDroplet system was used to investigate whether specific IDRs had the propensity to phase separate. While this system provided a good indication of whether sepcific regions within CBP phase separate or not, there are several caveats to using the optoDroplet system. The oligomers produced via CRY2 are poorly characterised, and generate dispersed condensates that prevent rigorous quantification. It is also unclear to what extend the hydrophobic interactions that are keeping the CRY2 protein oligomerised contributes to the formation of these

condensates themselves (Bracha et al. 2018). During the course of this degree work, a new system was developed and published from the Brangwynne lab called the corelet system, which is a more advanced biophysical system which allows more control of phase transitions in a quantifiable manner (Bracha et al. 2018; Bracha, Walls and Brangwynne 2019). For work carried out in this thesis, the optoDroplet system was sufficient to identify regions within CBP which can phase separate. However, should further work be carried out looking into the influence of specific IDRs, where more control over the system in required, I would use the corelet system.

This method highlighted the IDR7 and the CFID region, which formed puncta under blue light stimulation, suggesting that this C-terminal region may be the driving force in CBP phase separation.

4.6.3 CBP-GFP constructs and deletion mutants

Once I had characterised the IDRs individually I wanted to investigate the effect of removing these regions on CBP phase separation. To do this I generated CBP-GFP deletion mutants, that were missing these regions.

4.6.3.1 Investigations of IDR2 and IDR5

CBP-GFP mutant constructs were generated by amplifying the CBP-GFP plasmid by PCR using primers which bordered each of the IDRs; this was unsuccessful for IDR2 and IDR5. These primer sequences were too repetitive to allow efficient amplification, as the IDRs are fixed in space I was not able to change the primer sequence to allow efficient amplification and I was unable to generate these mutants. If I wanted to revisit these mutants in the future, I could design custom synthesised DNA sequences that contain that region of CBP sequence with the IDR already missing which I could clone into CBP-GFP using a restriction digest approach. In Wang *et al.* (2022), they identified IDRs in p300 which encompass these regions; CBP's IDR2 and IDR5 are included within p300's IDR1 and IDR3 respectively (Figure 4.17). In their study, Wang et al. (2022b) showed that recombinant IDR1 alone does not form condensates *in vitro*, in contrast to the IDR3 which does form condensates (Wang et al. 2022b). They also generated a GFP tagged deletion construct $p300_{\Delta IDR3\Delta IDR5}$ which could not phase separate in live cells (Wang et al. 2022b). This suggests that in p300, the IDR3 region may be involved in regulating p300's phase behaviour (Wang et al. 2022b), in a similar manner to that of CBP's IDR7 which I have demonstrated here. Since p300's IDR3 region encompasses our IDR5 region, it may be worth investigating the effect of this region in future experiments. Deletion mutants of full length

p300 were not made for IDR1, therefore it is not clear what effect a deletion of this would have on full length p300 phase separation and therefore it is hard to extrapolate CBP's IDR2 phase behaviour. Taken together this work indicates a potential role of CBP's IDR5 on phase separation, however p300 and CBP sequence differ mainly within the IDRs, which could result in different phase separating abilities than those seen for p300. The current data presented here suggests an important regulatory role of the C-terminal region containing the IDR6, IDR7 and CFID regions which has a significant impact on the phase separation of full length CBP, therefore I decided not to investigate IDR2 and IDR5 further.

4.6.3.2 Key mutants which influence CBP phase separation

To test the effect the mutants have on CBP phase separation, I used transient overexpression in fixed cells, and generated a pipeline in ImageJ/Fiji to quantify the condensates. A limitation to using fixed cells is that there is no way to determine what state the puncta observed are in, do they behave as a solid or display liquid-like properties. This data was however useful as an indicator of important regions for CBP phase separation which can further be investigated in live cells. To quantify the puncta observed I used a macro in ImageJ/Fiji (Schindelin et al. 2012), where a manual threshold was applied to the raw images to generate the ROIs; this was determined individually for each experimental repeat. ImageJ/Fiji have multiple different preset methods which can be used to generate a threshold which would allow to standardise this procedure, however, within the constructs tested there were a range of different behaviours observed, from large condensates for $CBP_{\Delta IDR6}$-GFP to a diffuse signal for $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP (Schindelin et al. 2012). This made it extremely difficult to use a premade threshold value that could be applied to all images, and meant a manual threshold had to be applied. While this was consistent as I changed the value to account for differences in transfection and laser power seen on the day, it does introduce bias based on manual input. In analysing the dataset I filtered out the constructs which did not form puncta, if at the point of analysing the images in ImageJ/Fiji I had excluded the non-puncta forming constructs it may have been easier to pick a threshold preset that could be applied consistently to all constructs, as the variation between constructs would be less. While I transfected the same amount of DNA, concentration of the expressed protein was not taken into account in the analysis pipeline. A different approach would be to only compare nuclei which show a similar intensity profile, indicating similar protein expression. However, for constructs such $CBP_{\Delta IDR6}$-GFP which were brighter due to a higher local concentration it would be difficult to find a standardised intensity profile. Nevertheless the method chosen allowed the

quantification of all constructs, and highlighted specific mutants to investigate further, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP.

The main constructs identified were tested in live cells using 1,6-HD treatment and FRAP; as was discussed in section 3.5.2, these experiments have their weaknesses. Particularly in the 1,6-HD treatment, it was clear when looking at diffuse mutants $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP that the nuclei were extremely unhappy as they changed shape through the timecourse. This brings into question how 1,6-HD is directly inferring the presence of hydrophobic interactions if it is causing such disturbance in the cell. Even so, 1,6-HD treatment suggested that the puncta observed were formed through phase separation, as dispersal of condensates was observed.

In this doctoral work I used FRAP to further investigate the mobility of CBP and $CBP_{\Delta IDR6}$-GFP condensates. While this technique cannot directly qualify how liquid-like the puncta observed are, it can indicate protein mobility and thus provide an indication to whether they are formed through phase separation. Both CBP and $CBP_{\Delta IDR6}$-GFP behaved similary, as they photobleached and recovered within 30 seconds, which is typical of a phase separating protein.

Taken together this work has highlighted the regions of IDR6, IDR7 and CFID as having an impact on CBP phase separating behaviour.

4.6.4 Influence of CBP IDRs on phase separation compared to p300's IDRs

Over the course of this work, numerous papers have been published which investigate the phase separation of p300, a paralog of CBP, including work on specific IDRs and larger regions of interest (Ma et al. 2021; Zhang et al. 2021; Wang et al. 2022b). Larger regions of p300 have been observed to phase separate using the optoDroplet system including the N-terminal region and the C-terminal region (Ma et al. 2021). Where the N-terminal region encompasses CBPs IDR1 and IDR3, and the C-terminal region covers CBPs IDR6 and IDR7, including structured domains in this region. Wang *et al.* (2022) has also investigated more restrictive IDR regions, suggesting that in p300 IDR3 and IDR5 can form phase separated condensates using an *in vitro* droplet approach, and upon removal of IDR3 and IDR5 in cells they do not see the appearance of puncta (Wang et al. 2022b). IDRs of p300 used in Wang *et al.* (2022) are shown in comparison to CBPs IDRs in Figure 4.17. Together this highlights the importance of the IDR7 region, for which the behaviour is consistent across all studies, as it forms condensates independently, and when removed forms a diffuse pattern from the full length protein.

This work does not support findings that the N-terminal region can phase separate independently from Ma *et al.* (2021). It is likely that the differences observed here are due to differences in

the sequence of the IDR regions between CBP and p300 and the presence or absence of structured regions, for a comparison of CBP and p300 sequence see appendix, section 9.4. Although the largest differences between CBP and p300 are found within the IDRs (Chan and La Thangue 2001), it is useful to compare differences between these sequences to understand if their phase separating behaviour is conserved, along with their catalytic activity.

4.6.5 The influence of the AIL on CBP phase separation

Alongside identified IDRs the AIL was also shown to influence CBP phase separating behaviour. In the optoDroplet system the AIL did not phase separate and upon removal from full length CBP an increased propensity to phase separate was observed. This suggests that the AIL may have an inhibitory effect on CBP phase separation. To investigate the effect of the complete removal of the AIL on CBP phase separation, I removed this region and replaced it with a glycine serine linker; the AIL is at the base of two alpha helices, therefore the direct removal of the AIL would have caused significant effects on structure and function of the HAT domain, which could have adversely influenced CBPs expression, folding or phase separation ability. A similar method was applied in Delvecchio *et al.* (2013), where the AIL of p300 was replaced with a glycine serine linker, in order to maintain the structure of the HAT domain, to generate a crystal structure of the catalytic core of p300. It is possible however, that in replacing the loop with the linker I have introduced a sequence which itself increases the phase separation of CBP. Certain residues have been shown to be preferred in IDRs and proteins that phase separate, those include serine and glycine; however, the benefit of using these residues is that they are flexible and so can easily bridge the gap between the alpha helices, with the caveat being that because of this reason they are often found in IDRs (Boeynaems et al. 2018; Mittag and Parker 2018; Riback et al. 2017; Sabari et al. 2018; Pessina et al. 2019; Pak et al. 2016). In this case, the risk of introducing phase separating competent residues was outweighed by trying to keep the integrity and function of the HAT domain intact. It is also possible that the inhibitory affect of phase separation seen when the AIL is present, is due to the acetylation of key lysine residues. This will be further explored in Chapter 5.

4.6.6 The role of FUS in CBP phase separation

Initially, I identified the CFID region within CBP because of its interaction with FUS, a protein that has been routinely shown to phase separate in the literature (Wang et al. 2008b; Birsa et al. 2020; Hamad et al. 2020; Rhine et al. 2020; Qamar et al. 2018). The CFID within CBP was shown

to phase separate independently and when removed reduced CBPs ability to phase separate. This is likely a result of removing the IDR7 region, a fragment of the CFID, as opposed to removal of the FUS binding site. To test whether FUS is important in regulating CBP phase separation in the future, a FUS knockout or knockdown could be generated to see how this affects the phase separation of CBP. While an interesting avenue to pursue, the influence of FUS on CBP phase separation beyond the CFID mutant is not currently being investigated.

4.6.7 The contributions of IDR6 and IDR7 on CFID phase behaviour

What is particularly interesting about the CFID region is that it encompasses both the IDR6 and the IDR7 region and yet the CFID region displays the same behaviour as that observed for IDR7. The IDR6 region has no propensity to phase separate independently and when removed forms larger condensates, suggesting a mechanism to negatively regulate the phase separation of CBP. For IDR7 this is reversed, the IDR7 promotes phase separation of full length CBP, which when removed prevents phase separation of the full length protein. This suggests that the IDR7 region, is a dominant region for regulating phase behaviour. It may also be possible that the IDR6 region acts to balance the phase separation ability of the IDR7, acting in cooperation to form the overall phase separation ability of CBP. A similar interaction between IDRs has been described within Ataxin-2 (ATXN2), a protein found in stress granules that is commonly implicated in neurodegenerative diseases, due to the presence of a polyglutamine repeat (Boeynaems et al. 2023). The IDR2 within ATXN2 binds to RNA which facilitates compartmentalisation into aberrant granules; the IDR3 within ATXN2 serves to quench this behaviour by interacting with IDR2 and preventing RNA binding (Boeynaems et al. 2023). Multiple studies have shown IDRs with different enrichments of amino acids are capable of interacting with each other, in this case, the IDR2 and IDR3 of ATXN2 may interact together for regulation purposes, to prevent aberrant activity and neurodegeneration (Boeynaems et al. 2023; Wang et al. 2018). With reflection on CBP, it is therefore possible that the IDR6 and IDR7 exist synergistically to facilitate CBP function in a phase separated state.

The CFID region does not fully span the IDR6 region and therefore is not a double mutant, however it is interesting to see how the combination of these mutants affects phase separation. This line of questioning led to the generation of double mutants, containing these original mutants with further deletions of the AIL, to understand how different combinations of mutants can affect CBP phase separation.

Chapter 5

Further investigations of the AIL and double mutants on CBP phase separation

5.1 Introduction

In the previous section, several intrinsic disorder regions, including the IDR6 and IDR7, were identified which influence the phase separating ability of CBP. Apart from IDRs I was also interested in investigating how the autoinhibitory loop of CBP affects CBP phase separation.

5.1.1 Role of the autoinhibitory loop in CBP

The histone acetyltransferase (HAT) activity of CBP, similarly to p300, is regulated by a basic activation loop, also known as the autoinhibitory loop (AIL) (Thompson et al. 2004; Ortega et al. 2018). The AIL region spans amino acids 1559 - 1608 in the HAT domain, and contains a large number of lysine residues which are positively charged in the inactive, deacetylated state. In this state the lysines are thought to engage with the active site of CBPs HAT domain through electrostatic interactions, preventing substrate binding and leaving CBP in an inactive form. Upon HAT activation these lysines become acetylated, the interaction with the binding site is disrupted and the loop is displaced from the site, leaving the active site free for substrate binding and HAT activity. The AIL has been shown to have RNA binding capability, with eRNA binding able to stimulate HAT activity by increasing the affinity for histone substrates (Bose et al. 2017). The

AIL, and more specifically the lysine residues, are therefore incredibly important in regulating the activation of CBPs histone acetyltransferase activity (Thompson et al. 2004; Thompson et al. 2001; Liu et al. 2008; Wang et al. 2008a; Ortega et al. 2018).

The acetylation of lysine residues has been investigated within the context of p300, with conflicting results being published in the literature. The effect of acetylation of the core was also investigated using mass spectrometry, which suggested a negative correlation between acetylation and the ability of p300 to form condensates *in vitro* (Zhang et al. 2021). In contrast, full length p300 has been shown to phase separate with proteins containing transactivator domains, with co-condensation stimulating acetylation via p300 (Ma et al. 2021). Upon inhibiting p300's catalytic activity, large condensates dispersed into smaller more numerous condensates, indicating that p300s catalytic activity may provide a positive feedback loop promoting large condensate formation and maintenance (Ma et al. 2021). A summary of these papers can be found in Figure 5.1. I was therefore interested in looking at the effect the AIL region has on CBP phase separation, and more specifically the effect of lysine residues and implication of HAT activity on the phase separation of CBP.

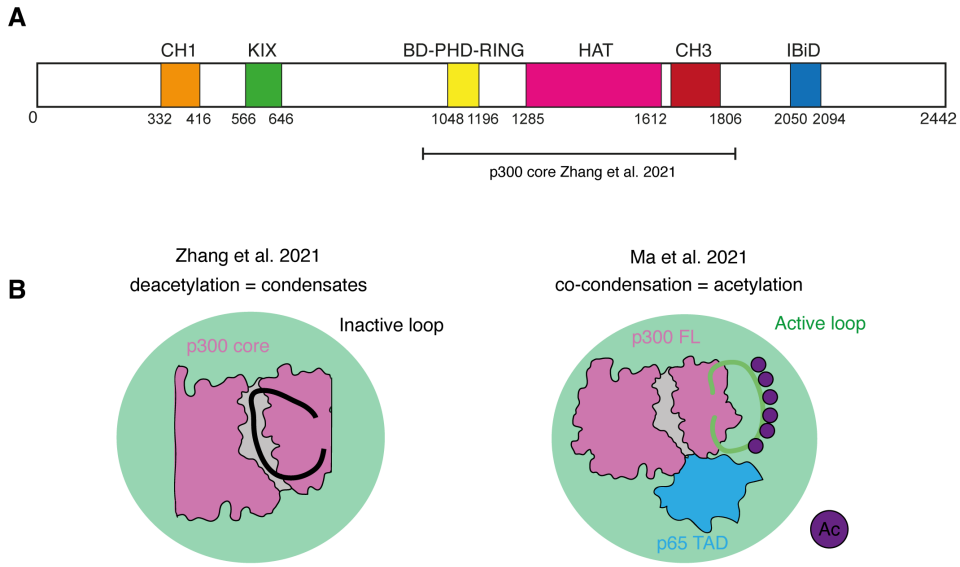

Figure 5.1: Summary of published p300 AIL data
A - domain diagram of p300 showing full length and truncated protein used in (Zhang et al. 2021). Globular domains shown here include: transcriptional adapter zinc binding motif (CH1/TAZ1, orange), CREB binding domain (KIX, green), Bromodomain PHD RING domain (BD-PHD-RING, yellow), histone acetyltransferase domain (HAT, pink), CH3/TAZ2 and IRF-3 binding domain (IBiD/NCBD, blue). B - Key finding of Zhang et al. (2021) and Ma et al. (2021) in understanding the importance of acetylation on phase separation. Zhang et al. (2021) showed deacetylation of AIL (black loop), leads to condensate formation (green circle) *in vitro*. Ma et al. (2021) showed that co-condensation between full length (FL) p300 and protein transactivator domain (TAD) of p65 (for example) increased acetylation activity of p300 indicated by the acetyl residues (purple circles) on the active AIL (green loop). Here p300 is depicted in a similar form to CBP, whereby in the diagram of Ma et al (2021) the diagram which represents full length p300 is based off CBP HAT domain diagram made by Dr Dan Bose and adapted here, diagram based on the HAT domain structure published in Wang et al. (2008a).

5.1.2 How do multiple mutations influence CBP phase behaviour

In the previous chapter, regions were identified which influence the phase behaviour of CBP, including IDR7 and the CFID which promotes phase separation and IDR6 which inhibits it. Investigating the AIL provided an opportunity to investigate how combinations of these mutations would affect the phase separation of CBP. I therefore generated mutants, of the original mutants $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP, and $CBP_{\Delta CFID}$-GFP, containing the ΔAIL or KTG

to understand if there is a hierarchy of influence of these regions on CBP phase separation.

5.1.3 Chapter aims

In this chapter, I investigate the influence of the AIL, particularly the lysines of the AIL, on CBP phase separation and generate double mutants to begin to understand how the different regions contribute to phase separation. The work in this chapter highlights that the phase separation of full length CBP is a combination of the contributions of different regions within CBP, with no single region driving CBP phase separation.

5.2 Are the lysines in the AIL important for CBP phase separation?

Alongside investigating the role of the AIL, I was also interested in understanding how acetylation of the AIL in the HAT domain could affect full length CBP phase separation. I have previously looked at the replacement of the whole AIL with a linker sequence in section 4.4, suggesting that the puncta observed when replacing the whole AIL were significantly different to WT, with larger puncta observed (Figures 4.8 and 4.14). One explanation for this is that the AIL itself may have a slightly inhibitory role on the phase separation of full length CBP. This inhibitory effect observed could also be caused by the acetylation of key lysine residues within the AIL. To further investigate this, I wanted to look specifically at the lysine residues within the region to determine if mutating sites of acetylation is sufficient to alter the phase separating ability of CBP. To test this I generated a CBP-GFP construct where the lysine residues between amino acids 1565 - 1608, a total of 13 residues, were mutated to glycine residues, which will be referred to as CBP_{KTG}-GFP (lysine-to-glycine, KTG). This mutation should prevent loop acetylation, and is likely to have an effect on the dynamics of the AIL, and its interaction with the active site due to removal of the positive charge on the lysine residues, while conserving the disorder within the loop. Domain diagrams of both AIL mutants are shown in Figure 5.2.

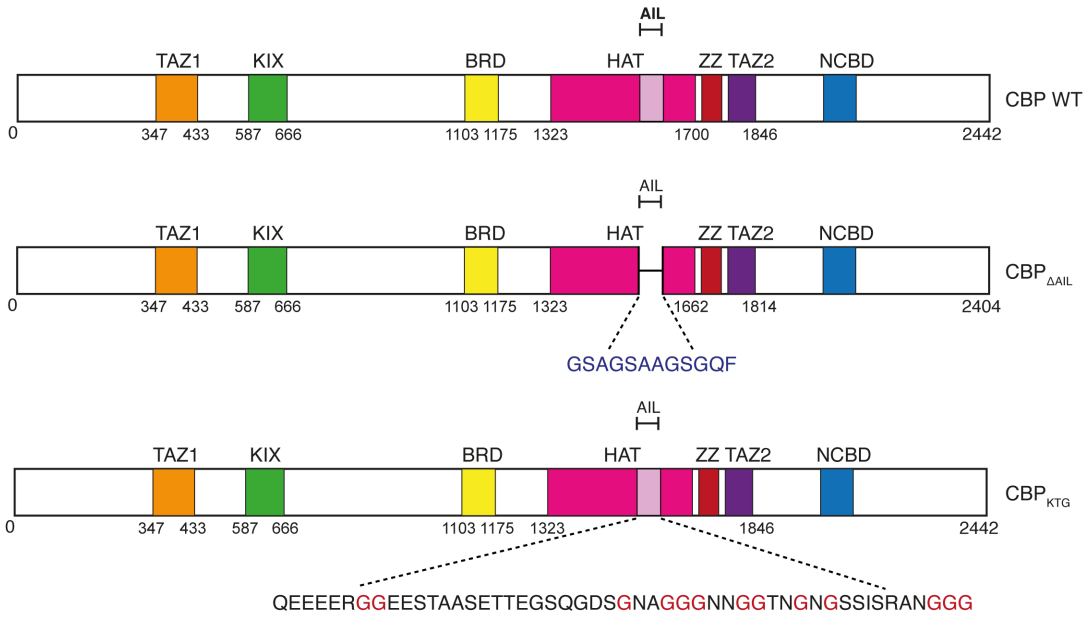

Figure 5.2: Domain diagrams of CBP-GFP, CBP$_{\Delta AIL}$-GFP and CBP$_{KTG}$-GFP mutants Diagram of CBP and two mutants CBP$_{\Delta AIL}$-GFP and CBP$_{KTG}$-GFP, where CBP$_{\Delta AIL}$-GFP shows the amino acid of the linker sequence in blue and CBP$_{KTG}$-GFP shows the amino acid sequence of the AIL with mutated lysine residues shown as red glycines (G). Globular structured domains include: transcriptional adapter zinc binding motif (TAZ1, orange) and (TAZ2, purple); CREB binding domain (KIX, green); bromodomain (BRD, yellow); histone acetyltransferase domain (HAT, pink); zinc binding domain (ZZ, red) and nuclear receptor coactivator binding domain (NCBD, blue). Image based on one found at Dyson and Wright (2016)

Previously, I had identified several regions that have a significant impact of CBP phase separation; IDR6, IDR7 and the CFID. Alongside investigating the effect of the AIL and key lysine residues on full length CBP phase separation, I also wanted to combine some of these regions to see if any phenotype is dominant over the other. I therefore decided to introduce the ΔAIL deletion and KTG mutant into key intrinsic disorder mutants CBP$_{\Delta IDR6}$-GFP, CBP$_{\Delta IDR7}$-GFP and CBP$_{\Delta CFID}$-GFP. This allowed the interrogation of the driving forces of CBP phase separation, to investigate if there is a cumulative effect. This experiment was performed as described in section 4.4.

The results for the lysine mutant, CBP$_{KTG}$-GFP, showed a similar phenotype to that of CBP$_{\Delta AIL}$-GFP mutant, with fewer, larger condensates when compared to WT (Figure 5.3).

Suggesting that the lysine residues within the AIL contribute to the phase separation phenotype of full length CBP. However, the effect of the ΔAIL and KTG mutant was not consistent within the double deletion constructs. Visually, the phenotype for each double mutant does not change from the original IDR phenotype (Figure 5.3). For example, both $CBP_{\Delta IDR6:\Delta AIL}$-GFP and $CBP_{\Delta IDR6:KTG}$-GFP mutants display a large condensate phenotype, similar to that of the $CBP_{\Delta IDR6}$-GFP phenotype (Figure 5.3). Similarly, all mutants containing ΔIDR7 and ΔCFID display a diffuse pattern regardless of the presence of either ΔAIL or KTG (Figure 5.3). As the individual $CBP_{\Delta AIL}$-GFP and CBP_{KTG}-GFP mutants form larger condensates to WT, I predicted that the diffuse constructs, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP, would form slightly more condensates with the addition of the AIL mutants (Figure 5.3). This was not observed, and suggests that the phase separation of CBP is more complex than a simple additive model. This data was then quantified as was described in section 4.4.

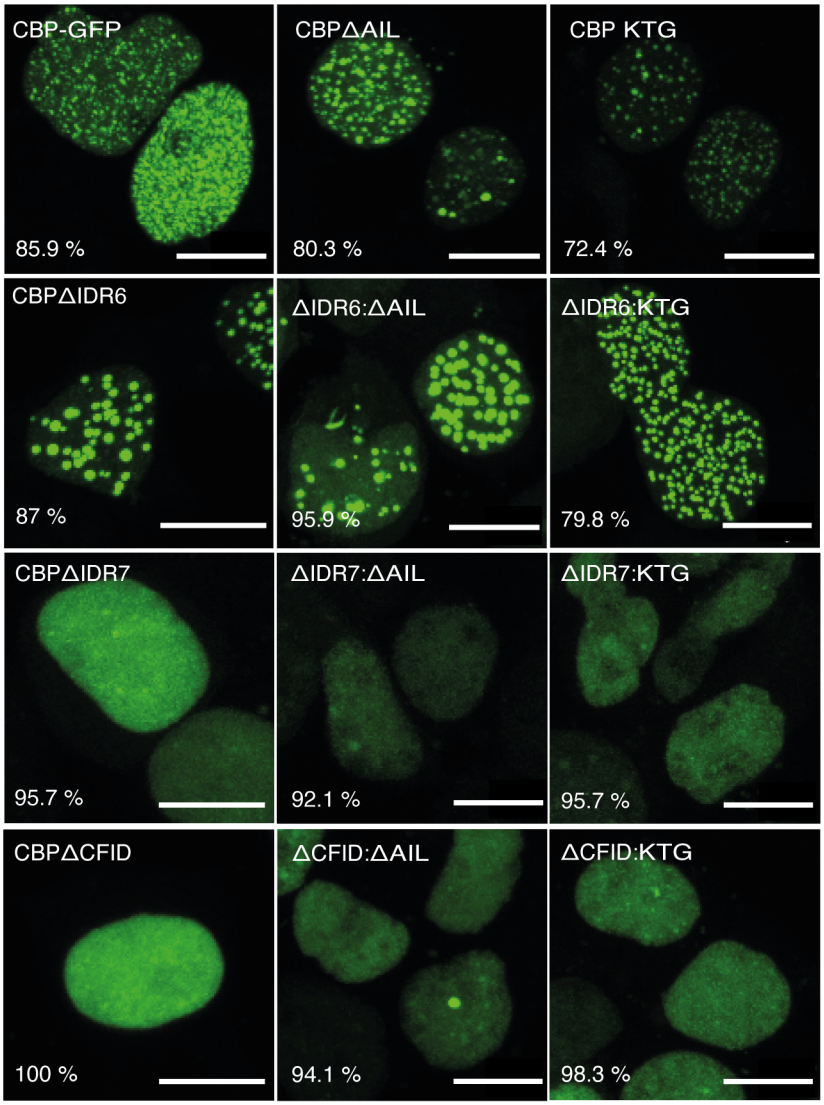

Figure 5.3: Representative images of double mutants of CBP-GFP
Maximum projections of CBP-GFP and panel of overexpressed mutants. The left hand images represent the original constructs CBP-GFP, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP, with each row containing one of the original IDR mutants. The middle column represents the double mutant containing ΔAIL and the right hand column the KTG, lysine to glycine mutation. Images were processed in ImageJ/Fiji and a scale bar of 10 μm was applied. Percentage of the transfected nuclei which display the representated phenotype is shown in the bottom left corner. 46 - 127 nuclei were sampled per construct. Biological replicates = 3.

5.2.1 Analysing puncta from AIL and double mutants of CBP

Before comparing puncta observed between these constructs, the dataset was filtered as previously described in section 4.4.2, in order to identify single puncta to compare mutant condensates to WT.

Firstly, I looked at the number of nuclei which passed the threshold per construct. A high proportion of nuclei passed the threshold for all CBP WT constructs and $CBP_{\Delta IDR6}$-GFP constructs, where greater than 60 % of nuclei passed the threshold value in ImageJ/Fiji (Figure 5.4). This was similar to that observed in previous overexpression experiments and provided confidence that the system and analysis were running as expected. Original mutants $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP had greater than 50 % of nuclei which passed the threshold, with all double mutants having a lower percentage of nuclei which showed an intensity reading (Figure 5.4). On initial inspection this may indicate that the transfection itself may have been less successful in these constructs. However, from imaging a diffuse pattern for these constructs is observed (Figure 5.3), where the intensity of this diffuse signal was not bright enough to pass the threshold value, therefore it is not that the transfection was unsuccessful with these constructs, it is that the signal for these constructs was not bright enough to be identified by ImageJ/Fiji.

I then looked at how the filters initially set up for the original dataset affect the AIL double mutant dataset, where the puncta identified is shown in purple (Figure 5.5 A). This distribution was the same as observed in the previous dataset in section 4.4, suggesting that the filters were isolating a comparable population of puncta in area, circularity and integrated density. It is clear that I have many fewer data points for $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP, both original and double mutants, compared to CBP WT and $CBP_{\Delta IDR6}$-GFP (Figure 5.5 B); this reflects the low number of nuclei which passed the threshold value, moreover it shows the difference between the condensate forming constructs and those that behave in a diffuse manner. The data was filtered to isolate the puncta fraction to be further analysed to identify the puncta forming constructs.

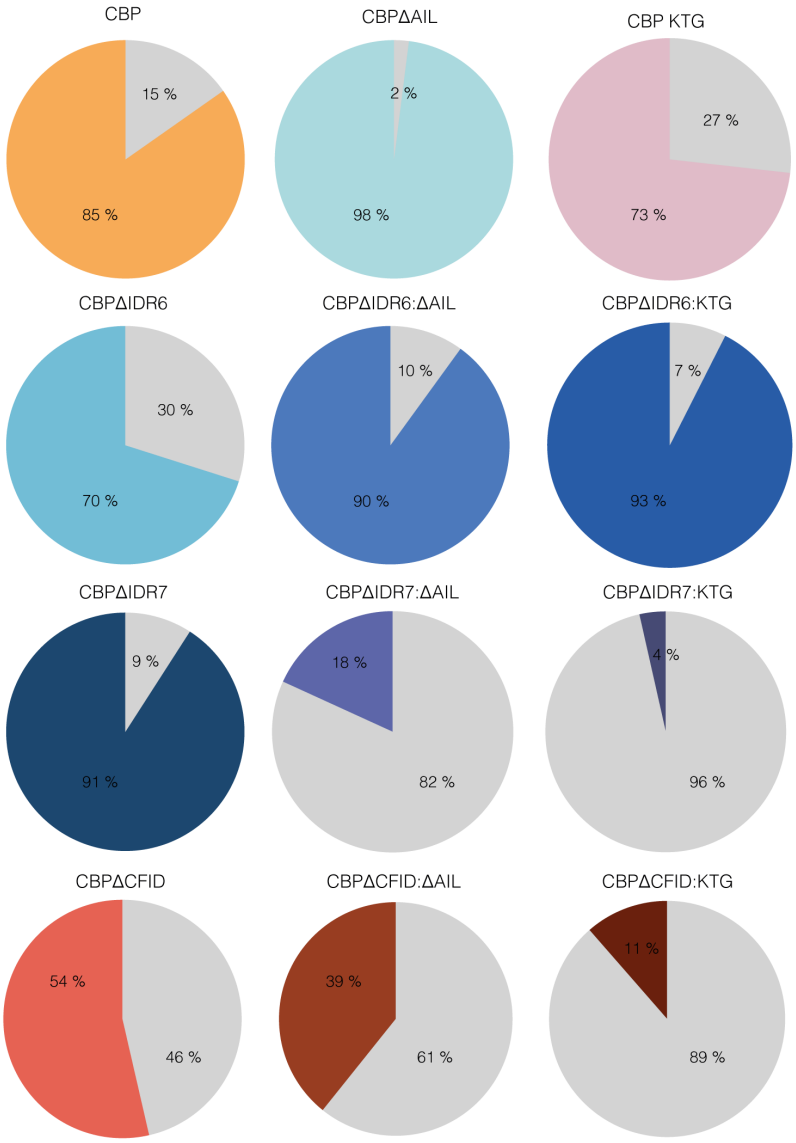

Figure 5.4: Proportion of nuclei that passed the threshold value for double mutant constructs Percentage of nuclei sampled per construct which showed an intensity reading above the threshold value, compared to the nuclei which did not (grey) shown via piechart. Sample names indicate the missing region e.g. CBPΔIDR6 is $CBP_{\Delta IDR6}$-GFP. Graphs were made in Rstudio using R. 46 - 127 nuclei were sampled per construct. Biological replicates = 3.

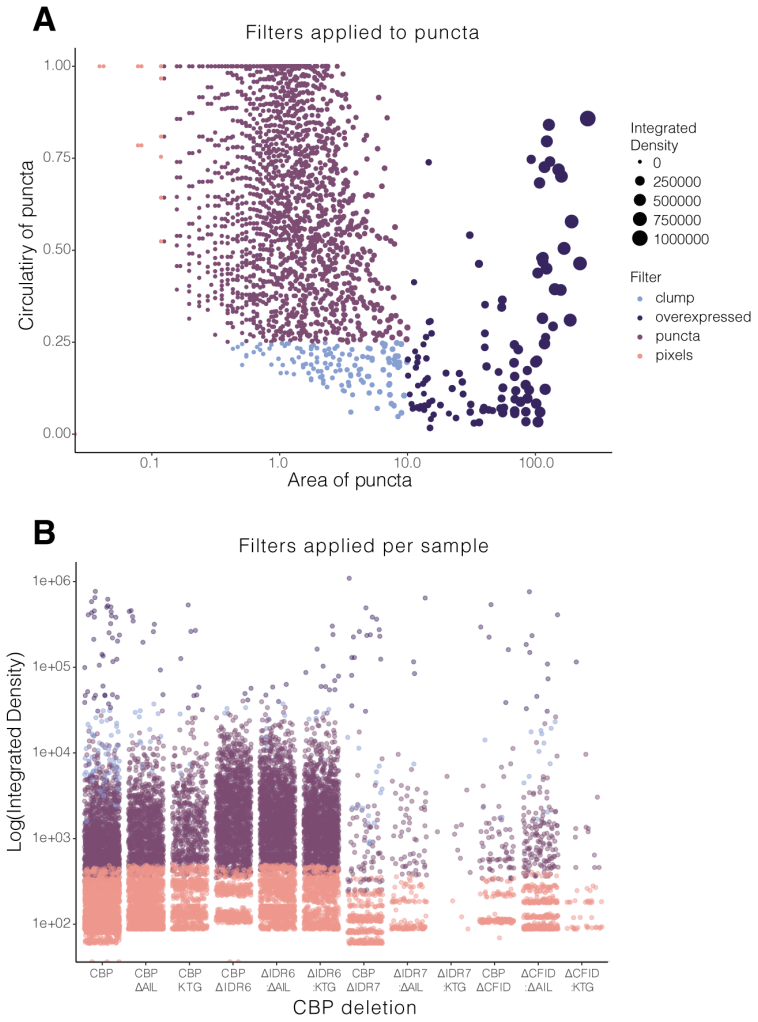

Figure 5.5: Applying the filter to double mutant overexpression dataset
A - visualisation of the whole dataset comparing the circularity of puncta and the area of the puncta. Each circle represents a puncta, where the size is correlated to the integrated density of the puncta. The colours relate to the filters applied to the sample, blue represents puncta which are clumped together (circularity below 0.25), dark purple represents overexpressed nuclei (area above 10), purple are the individual puncta and pink represents intensities which are three pixels or less and therefore cannot be confirmed as a puncta (area of 0.36 μm). B - a visual representation of the different filters used in A per sample using the log value of the integrated density per sample. Here each dot represents a single puncta reading, colours for each filter are the same as applied in A. Graphs made using R in Rstudio. 46 - 127 nuclei were sampled per construct. Biological replicates = 3.

For this experiment I wanted to look at the number of puncta per nuclei to determine the effect of ΔAIL and KTG on the puncta formed from each original construct, for example $CBP_{\Delta IDR6}$-GFP compared to $CBP_{\Delta IDR6:\Delta AIL}$-GFP, to do this I generated a boxplot grouped by each original construct (Figure 5.6). Comparing the median number of puncta per nuclei, a decrease was observed for $CBP_{\Delta AIL}$-GFP, and a further decrease for the CBP_{KTG}-GFP mutant compared to WT (Figure 5.6 A). This correlates with what was visually observed from the data, both $CBP_{\Delta AIL}$-GFP and CBP_{KTG}-GFP mutants show fewer puncta which are larger compared to the WT construct, which has smaller more numerous puncta, therefore the decrease in number of puncta per nuclei in these mutant constructs was consistent with the visual data observed (Figure 5.3). $CBP_{\Delta IDR6}$-GFP visually forms puncta and the number of puncta does not change with the presence of any additional mutation (Figure 5.3 B). For the $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP, I had previously determined that the original mutants have a low ability to form phase separated condensates. Correspondingly, the double mutants for each of these constructs reduced the median number of puncta per nuclei and therefore do not form condensates in the same manner as WT CBP (Figure 5.6 C and D respectively). This suggests that the AIL mutants do not affect the overall phenotype of the $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP; since none of these constructs formed puncta they were not included in the comparison of double mutants of CBP that do form puncta.

Since condensates were observed for the remaining constructs CBP-GFP and $CBP_{\Delta IDR6}$-GFP, I next wanted to filter these condensate forming constructs for nuclei which have too few puncta to be classed as a condensate forming nuclei. Based on the median number of puncta for CBP_{KTG}-GFP, which is 4 puncta, I filtered using a cut-off value of 3. This threshold filtered out any nuclei which have too few puncta to be classified as puncta forming nuclei, while comparing the effect of AIL deletion or KTG mutation on CBP WT and $CBP_{\Delta IDR6}$-GFP phase separation.

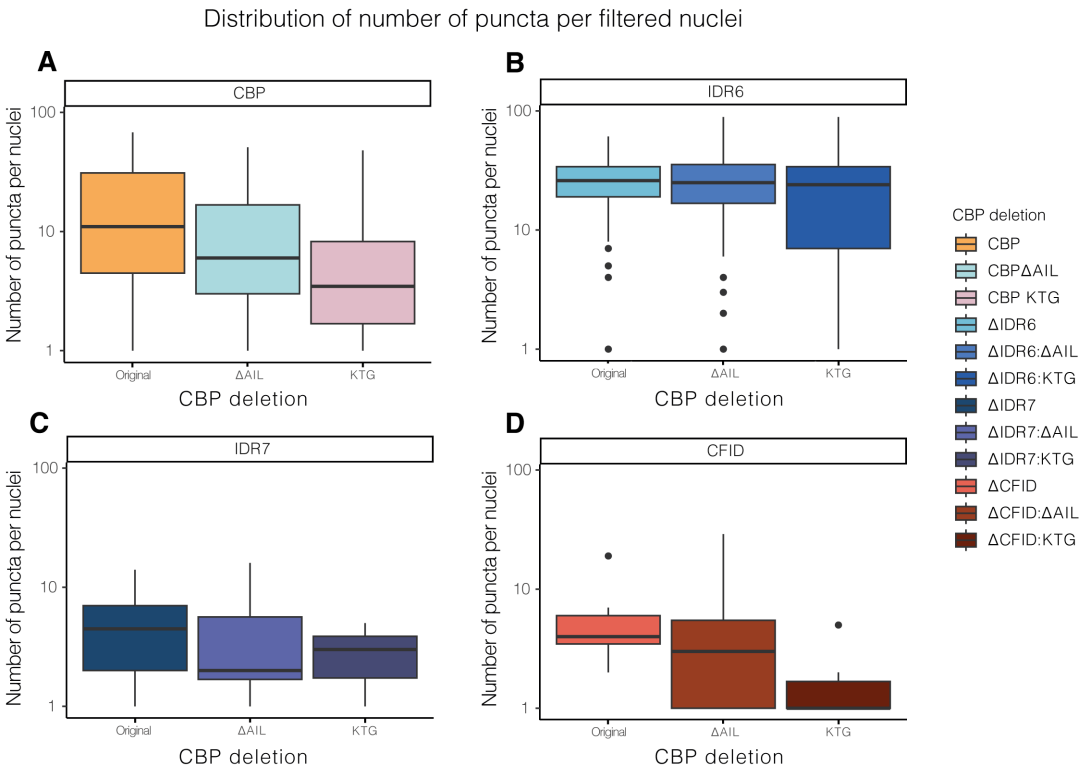

Figure 5.6: Distribution of number of puncta per nuclei for AIL and double mutant constructs Boxplot to show the distribution of number of puncta per nuclei for each construct. Original construct for each plot is: A - CBP-GFP, B - $CBP_{\Delta IDR6}$-GFP, C - $CBP_{\Delta IDR7}$-GFP and D - $CBP_{\Delta CFID}$-GFP. Plots to compare the original, ΔAIL and KTG deletion of the original construct. The boxplot represents the maximum value, upper quartile, median, lower quartile and minimum value, in that order. Graphs were made using R in Rstudio. 46 - 127 nuclei were sampled per construct. Biological replicates = 3.

After filtering out nuclei which had fewer than 4 puncta per nuclei, I compared differences in condensates between CBP WT, $CBP_{\Delta AIL}$-GFP and the CBP_{KTG}-GFP, as well as data from $CBP_{\Delta IDR6}$-GFP, and double mutants $CBP_{\Delta IDR6:\Delta AIL}$-GFP and $CBP_{\Delta IDR6:KTG}$-GFP (Figure 5.7). I therefore looked at the median integrated density of puncta per nuclei, per construct, and statistically compared the different pairs using the Kruskall-Wallis test (Figure 5.7).

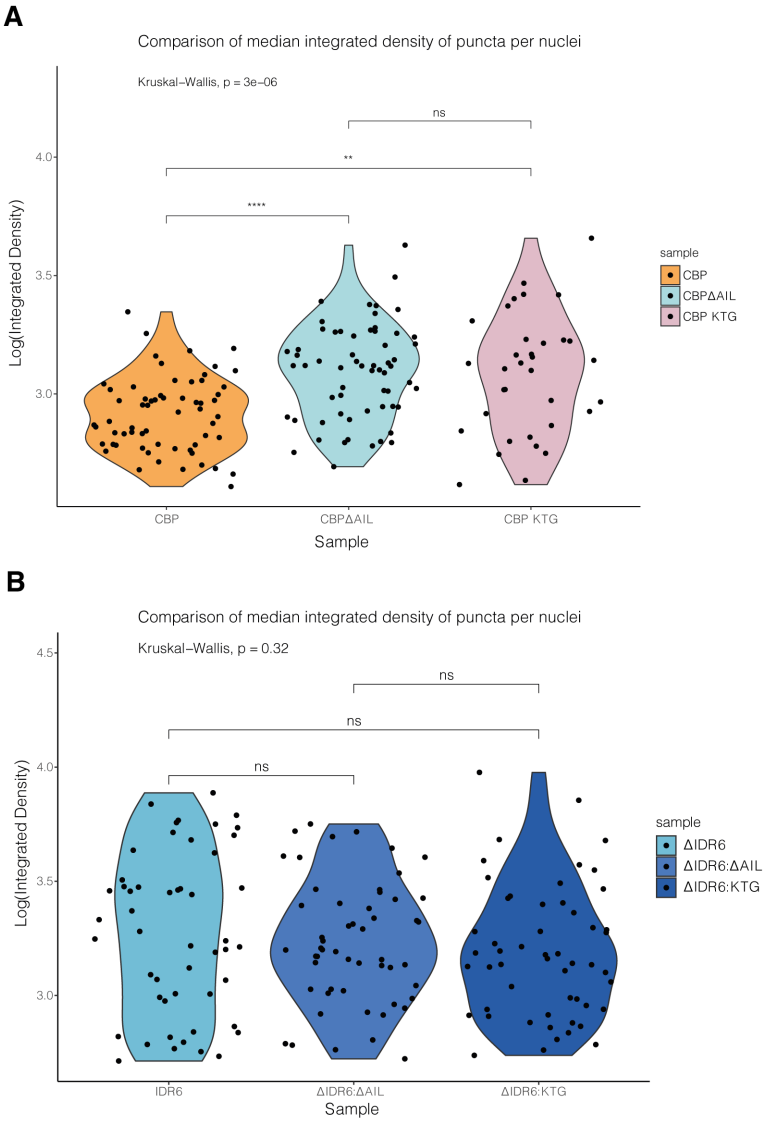

Figure 5.7: Integrated density of CBP double mutant constructs

Violin plot comparing the median integrated density of puncta per nuclei (dots) between overexpressed constructs. A - violin plot to compare CBP-GFP to two mutant constructs $CBP_{\Delta AIL}$-GFP and CBP_{KTG}-GFP. B - violin plot to compare the difference in integrated density between $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR6:\Delta AIL}$-GFP and $CBP_{\Delta IDR6:KTG}$-GFP. Stars represent significance with $p > 0.05$ = ns, $p < 0.01$ = **, and $p < 0.0001$ = ****. Graph and Kruskal-Wallis statistical testing was done using R in Rstudio. 46 - 127 nuclei were sampled per construct. Biological replicates = 3.

Both $CBP_{\Delta AIL}$-GFP and CBP_{KTG}-GFP show a significant increase in the median integrated density of puncta compared to the WT protein, confirming the visual phenotype of larger puncta (Figure 5.7 A). Of these, $CBP_{\Delta AIL}$-GFP puncta have a higher significance than that of the CBP_{KTG}-GFP when compared to WT (Figure 5.7 A). Interestingly, there is no significant difference between the $CBP_{\Delta AIL}$-GFP and the CBP_{KTG}-GFP mutant (Figure 5.7 A). This suggests that the lysine residues, or potentially the acetylation of these lysines, may be involved in regulating this phase behaviour, which is consistent with the possible role of lysine acetylation in regulating phase separation of the p300 HAT domain (Zhang et al. 2021).

As the constructs of $CBP_{\Delta AIL}$-GFP or CBP_{KTG}-GFP showed an increase in condensate formation (Figure 5.7 A), I expected that combining the mutations of ΔIDR6 with either the ΔAIL or KTG would cause large condensates to form, potentially larger than $CBP_{\Delta IDR6}$-GFP alone. However, visually the double mutants of $CBP_{\Delta IDR6}$-GFP look very similar to the single mutant. When I compared quantitatively $CBP_{\Delta IDR6}$-GFP to $CBP_{\Delta IDR6:\Delta AIL}$-GFP and $CBP_{\Delta IDR6:KTG}$-GFP no significant difference was observed between the single mutant and the double mutants, or between the double mutants themselves (Figure 5.7 B).

Altogether I have identified several regions which in isolation are important for influencing the phase separation of full length CBP, a summary of the main findings can be found in Figure 5.8. Upon mutating the AIL from CBP-GFP WT, both $CBP_{\Delta AIL}$-GFP and CBP_{KTG}-GFP form fewer puncta which were larger in size compared to WT (Figure 5.8). In these mutant constructs, the AIL was deacetylated compared to WT, which suggested a potential role of acetylation on influencing the phase separation of CBP. When the IDR6 region was deleted from full length CBP an increase in puncta size and a decrease in number of puncta was observed, this was independent of acetylation of the AIL, since both CBP WT and $CBP_{\Delta IDR6}$-GFP have an acetylated AIL (Figure 5.8). The $CBP_{\Delta IDR6:\Delta AIL}$-GFP and $CBP_{\Delta IDR6:KTG}$-GFP mutants fail to significantly alter the properties of condensates that arise from the $CBP_{\Delta IDR6}$-GFP mutant alone (Figure 5.8). This therefore implies that the contribution of the IDR6 region in negatively regulating condensate formation is greater than the role of the AIL. It could also be that the condensates observed for $CBP_{\Delta IDR6}$-GFP are the maximum size puncta can be in cells, therefore whether the AIL is contributing to condensation cannot be observed.

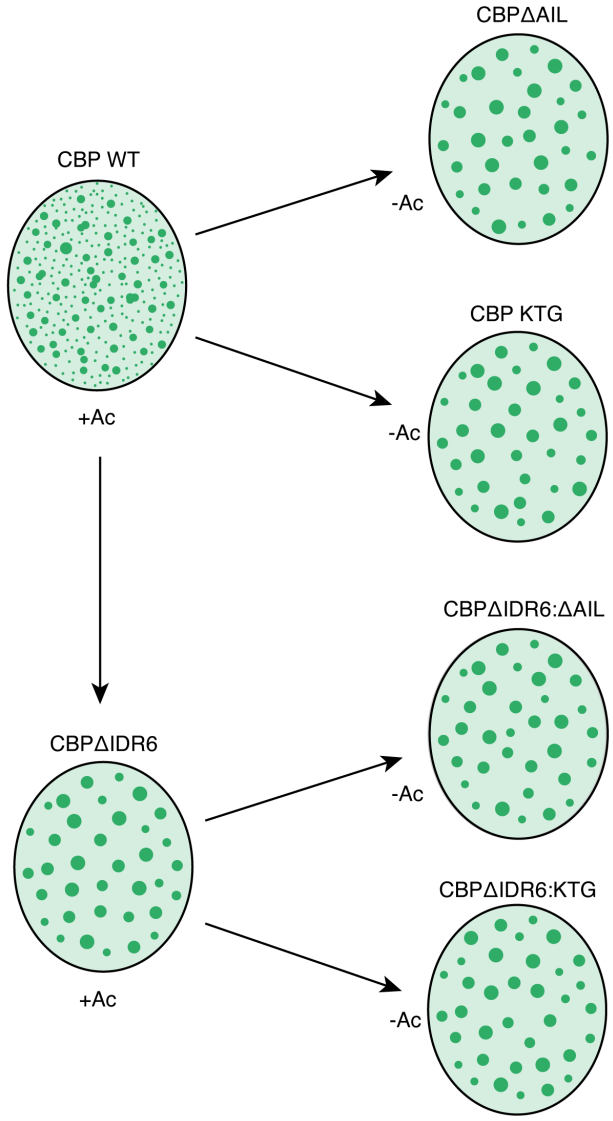

Figure 5.8: Summary of the effect of double mutants on CBP-GFP phase separation
In the full length construct, CBP-GFP has an acetylated AIL (+Ac) that forms small nuclear puncta. When the AIL is deleted, $CBP_{\Delta AIL}$-GFP, or the lysines are mutated to glycines within the AIL, CBP_{KTG}-GFP, these constructs which are deacetylated (-Ac), form fewer puncta which are larger in size. Upon deleting the IDR6 region, $CBP_{\Delta IDR6}$-GFP, forms larger puncta that is not dependent on the lack of acetylation, since the AIL is still acetylated. Both double mutants $CBP_{\Delta IDR6:\Delta AIL}$-GFP and $CBP_{\Delta IDR6:KTG}$-GFP form large puncta in a similar manner to $CBP_{\Delta IDR6}$-GFP, even though in these mutant constructs the AIL is deacetylated.

I have shown in this section that the AIL can modulate the size and intensity of CBP WT condensates, I have also demonstrated the importance of the lysine residues within the AIL for this regulation. In the absence of these positively charged residues, which are capable of being acetylated, slightly larger puncta compared to WT was observed, equivalent to the AIL deletion, suggesting an inhibitory role when present naturally in the AIL. Alternatively, removing the ability of this region to be acetylated could remove the inhibitory effect of acetylation of CBP phase separation.

Overall, the work displayed in this chapter, demonstrates that the phase separation of CBP is not dependent on one region in particular but the contributions of multiple different regions. Some of these regions may be more dominant in the effect they have on phase separation compared to others, such as the IDR7 region, but it is the combined contributions of all of these regions that allow CBP the ability to phase separate.

5.3 Discussion

5.3.1 The AIL as a factor influencing CBP phase separation

Alongside the IDR constructs I investigated the influence of the AIL on CBP phase separation. These results showed that upon removal of the AIL, or simply the lysine residues, an increase in phase separation was observed compared to WT CBP.

The role of glycine in CBP_{KTG}-GFP phase separation phenotype Investigating the AIL, large condensates were observed when the AIL was replaced by a glycine-serine linker. A similar phase separating pattern was also observed when13 lysine residues in the AIL were mutated to glycines, a mutation that increased CBP's ability to phase separate. Glycine residues were chosen as they cannot be acetylated, with the importance of the lysines in the AIL being that they can be acetylated to activate CBP's HAT activity. There is however a drastic difference in size and charge between glycine and lysine. If there was significant concern as to the role that glycines are playing in this mutant, as was discussed previously in subsection 4.6.5, I could in the future generate a new mutant which had these residues mutated to histidines. Similarly to lysines, histidines are positively charged and relatively bulky, unlike lysines they cannot be acetylated, therefore could be a suitable replacement if further investigation into the AIL is to be completed. At this time, I don't think these measures need to be taken. The data shown in this chapter, indicates that the AIL, more specifically the lysine residues themselves, play a role in regulating CBP phase

separation.

The role of acetylation of lysine residues within the AIL Both $CBP_{\Delta AIL}$-GFP and CBP_{KTG}-GFP prevent acetylation of the AIL, as they lack the important lysine residues that become acetylated during CBP activation. Both of these constructs generate larger condensates than that of the WT construct. This suggested that acetylation negatively regulates condensation of CBP, which is consistent with what was found by Zhang *et al.* (2021). It is plausible that in the absence of the negative acetyl residues larger puncta formation was observed due to increased electrostatic interactions, which are usually disrupted with acetyl-lysine, which are not disrupted by the glycine residues in these mutants. Since, in these deletion mutants there is no negative charge and therefore no contributing factors to disperse the condensates into a smaller WT phenotype. From this data, acetylation may negatively influence phase separation of full length CBP, in order to regulate CBP HAT activity.

The influence of RNA binding on AIL phase separation Another possible explanation for the influence of the AIL on full length CBP phase separation is through RNA binding. CBP has been shown to bind RNA within the AIL to activate CBP HAT activity due to the presence of an RNA binding motif within the disordered loop (Bose et al. 2017). The RNA binding region within the AIL was predicted to be amino acid residues 1561 - 1620, with the positive lysine residues binding to the negatively charged RNA (Bose et al. 2017). This RNA binding region is directly within the region manipulated in order to generate the AIL deletion and mutated lysine construct (amino acid residues 1565 - 1608). Therefore it is possible that in removing the ability of the AIL to bind RNA molecules, I have increased the phase separating ability of CBP; potentially due to the lack of negatively charged RNA molecules present in the system.

Example of this can be found in the literature, where the absence of RNA has been described to increase phase separation of NONO, a molecular scaffolding protein (Zhang et al. 2023). A mutated NONO, where the RNA binding region had been removed, therefore preventing RNA binding, was shown to phase separate more readily both *in vitro* and within cells, compared to the WT protein (Zhang et al. 2023); indicating a crucial role for RNA in the context of phase separation. This indicates a potential influence of RNA molecules on CBP phase separation. To test this theory, I could generate *in vitro* condensates which consist of recombinant CBP-GFP, both in the presence and absence of purified RNA to see the effect of RNA binding on the phase separation of full length CBP.

5.3.1.1 HAT activity within the AIL mutants

In this work both the $CBP_{\Delta AIL}$-GFP and CBP_{KTG}-GFP are thought to be transcriptionally active. In the absence of the AIL there is nothing bound in the active site preventing substrate binding, and in the CBP_{KTG}-GFP mutant there is no binding of the AIL to the active site, as this is facilitated through the positive charge of the lysine residues which have been removed. While I have not tested the HAT activity of these two constructs, similar work was completed by Bose *et al.* (2017), studying HAT activity of CBP mutants. Bose *et al.* (2017) generated a mutant where the AIL was removed, as well as a mutant where the positive charge of the lysine residues was reversed to a negative charge; both mutants were shown to be catalytically active. Structural studies of the p300 HAT domain, where the AIL was deleted and replaced with a glycine serine linker, was also shown to be catalytically active (Delvecchio et al. 2013); which further implies that these two constructs could be catalytically active within these larger condensates. However future HAT assays should be performed to test this theory. If catalytically active, this work supports part of the results from Ma *et al.* (2021), which suggests p300 is catalytically active within condensates. However, it does not support acetylation acting as an anchor to promote phase separation, since both of these mutants are deacetylated (Ma et al. 2021).

This work implies that the AIL is important for regulating the phase separating ability of CBP, and potentially regulating its HAT activity, however it is currently unclear how the two are linked, and from this current data I cannot fully depict this interaction.

5.3.2 CBP phase separation is a combination of regions with differing phase separating abilities

Within this chapter I have shown multiple different ways to begin to understand which regions affect the phase separation of CBP as a whole, including generating several double mutants of ΔAIL, KTG and main IDR mutants of CBP phase separation $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP, and $CBP_{\Delta CFID}$-GFP. I have showed that formation of double mutants containing ΔAIL or KTG, for all constructs, did not affect the overall phenotype observed. For example, constructs $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR6:\Delta AIL}$-GFP and $CBP_{\Delta IDR6:KTG}$-GFP all showed a phenotype of large condensates, both of the double mutants display the same phenotype to the original. This suggests that some regions may have a more dominant influence over CBP phase separation. This may be due to the size of the region itself - for example the IDR7 region is larger than the IDR6 and both of which are larger than the AIL mutant; it could also be the position of the region within the context of the whole protein. It may simply be that the different regions have different strengths of influence.

Overall, I have shown that the phase separation of CBP is not controlled by one region in particular, which is something commonly concluded in the literature; it is a combination of all the different regions and there phase separating ability which creates the phase separating phenotype that is observed in the WT protein.

Chapter 6

Combating the concentration caveat: generating doxycycline-inducible cell lines expressing CBP-GFP and its mutants

6.1 Introduction

In the previous chapter, I looked specifically at how regions within CBP affect its phase separating ability using an overexpression system. From this, I highlighted regions of importance including IDR6, IDR7 and the CFID which inhibit or drive CBP phase separation respectively. As was discussed briefly in section 3.5.2, studying overexpression brings into question the physiological relevance of these findings. Even though this is a tool regularly used in the literature to study phase separation, changing the concentration of protein by orders of magnitude compared to the physiological levels, may facilitate phase separation in some cases. Especially for IDR6, when removed from CBP WT larger condensates were observed, which correlates with high protein expression identified by western blotting. While the IDR6 region has influence over CBP's phase separating ability, it is unclear from the work discussed thus far whether these condensates are due to the mutant itself or aided by overexpression. Differences in expression were observed between CBP-GFP and $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP, when transiently transfected into HEK293T cells, this makes further experimentation difficult. For example, if I

was to look at differences in chromatin localisation of CBP WT and these mutants, or differences in RNA binding, it would be challenging to separate the effect of increased concentration from the effect of the mutant itself. Therefore, I decided to make inducible cells lines allowing for more specific control of the level of expression. To do this I used the Retro-X Tet-On 3G Inducible Expression System (Clontech).

6.1.1 Retro-X Tet-On 3G Inducible Expression System

To generate finely tunable cell lines containing the protein of interest I used the Retro-X Tet-On 3G Inducible Expression System (Clontech) for human cells. This system utilised a TRE3G doxycycline (Dox) inducible promoter that controls the expression of a gene of interest via the Tet-On 3G transactivator protein. In the absence of Dox the Tet-On 3G transactivator protein cannot bind the promoter region, providing a system with low background and tunable expression of the protein of interest. Upon addition of Dox the Tet-transactivator protein undergoes a conformational change facilitating binding to the promoter region, recruiting transcription factors which activates transcription of the target gene (Figure 6.1) (Gossen et al. 1995). The Tet-On 3G transactivator protein is highly sensitive to Dox and has increased activity, compared to earlier variants of the protein used in this system (Gossen et al. 1995; Baron, Gossen and Bujard 1997; Zhou et al. 2006).

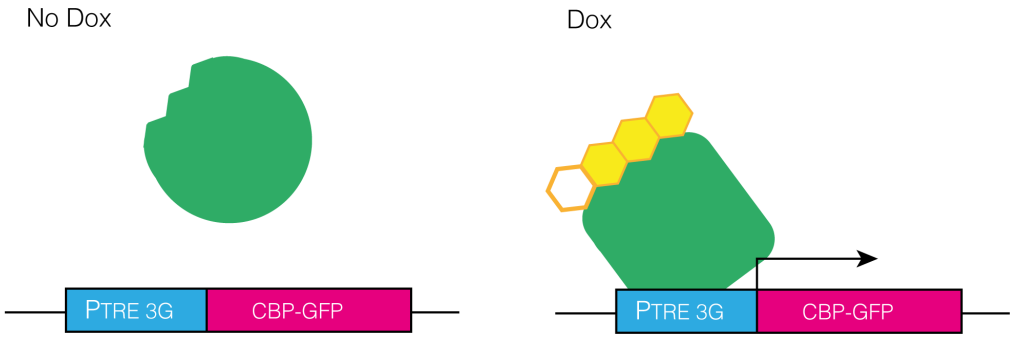

Figure 6.1: Diagramatic figure of doxycyline induction of Tet-transactivotor protein
Binding of doxycycline (Dox, yellow hexagons) triggers a conformational change of the Tet-On 3G transactivator protein (green) facilitating binding to the PTRE 3G tetracycline responsive promoter (blue). Once bound transcription of the gene of interest is activated, for example CBP-GFP (pink). Figure adapted from Retro-X Tet-On 3G Inducible Expression System User Manual (Clontech 2013; Das, Tenenbaum and Berkhout 2016).

I decided to use this method to generate inducible cell lines as it had already been established in the Bose lab and would allow the fine tuning of expression using Dox. I would then be able to express CBP-GFP and mutants at near-endogenous concentrations, to reduce the effect of concentration as a factor in downstream experiments. This system is very user friendly as both plasmids for viral transfection have antibiotic selection markers which easily facilitates selection of transduced cells - post viral transduction. I therefore generated individual cell lines containing CBP-GFP, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP under control of the Tet-On system.

Once the cell lines were established I could then express these constructs at near-endogenous concentrations and begin to understand how changes in phase separation affect CBP behaviour at much lower concentrations. For example, does the phase behaviour of these constructs affect CBP chromatin localisation; to test this I expressed these constructs at near-endogenous concentrations and performed Chromatin Immunoprecipitation Sequencing (ChIP-Seq) to determine where these constructs were localised within the genome.

Focussing on the main regions IDR6, IDR7 and CFID I wanted to investigate the sequences of these regions to understand if differences in phase separation could be attributed to the properties of amino acids that make up these sequences. For example is there an abundance or lack of polar, aromatic or hydrophobic residues which could explain the different phase separating properties of these sequences. To do this I used CIDER and localCIDER to analyse the sequence parameters of the different regions, to identify any sequences or sequence patterns which maybe important for their phase separating abilities.

6.1.2 Chapter aims

In this chapter I will focus on understanding in greater detail how IDR6, IDR7 and the CFID region effect phase separation. I will describe the process of generating and validating Dox inducible cell lines of CBP-GFP, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP to express near-endogenous concentrations of these proteins, and reduce the concentration caveat. I also begin to understand if these mutants, and therefore phase separation affect the chromatin localisation of CBP by performing ChIP-Seq. Finally, I test how overexpression of CBP-GFP and its mutants affect the phase separation of endogenous CBP, to understand if different phase behaviours can influence the endogenous protein.

6.2 Production and validation of doxycycline inducible cell lines

6.2.1 Generating doxycycline inducible lines

I generated Dox inducible cell lines including a maternal cell line HEK293T-Tet3G, and subsequent cell lines containing CBP-GFP, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP. Successful transduction was selected for using antibiotics G418, for the maternal cell line, as well as puromycin for GFP tagged cell lines. After 7 days of culture, GFP expression in the presence of Dox was observed for all cell lines. The cells were then processed by Fluorescence Activated Cell Sorting (FACS) to generate a polyclonal population of cells to enrich the population of GFP positive cells, and to select for single cells to establish clonal Dox inducible lines.

6.2.2 Validating the polyclonal populations

After bulk sorting, polyclonal cell lines were grown into 10 cm dishes and plated into 8-chamber cover glasses for screening. All polyclonal cells were plated with and without Dox to compare GFP expression and to see background levels of GFP expression in the absence of Dox (Figure 6.2). For expression of GFP tagged protein, a Dox concentration of 2 μg/ml was used, this is the upper limit of what is used for cell culture, to observe how well the system is expressing these constructs without tuning the expression. As this experiment was performed for screening purposes, the images were not processed and therefore show a higher background than some of my later images. For all cell lines induction of the protein of interest tagged with GFP, was observed upon addition of Dox, this expression was nuclear localised as viewed using Hoechst (blue) (Figure 6.2). Differences in expression were observed between cells within the same cell lines, this was to be expected as this is a polyclonal population and therefore variable amounts of transactivator and variable amounts of expression was observed (Figure 6.2). There was a higher proportion of cells that were expressing GFP in the FAC-sorted bulk population compared to the original population, therefore FACS was successful in enriching for the GFP-tagged expressing cells. All cell lines show some level of GFP expression in the no Dox control, with CBP-GFP having the highest amount of background compared to any of the mutant cell lines (Figure 6.2).

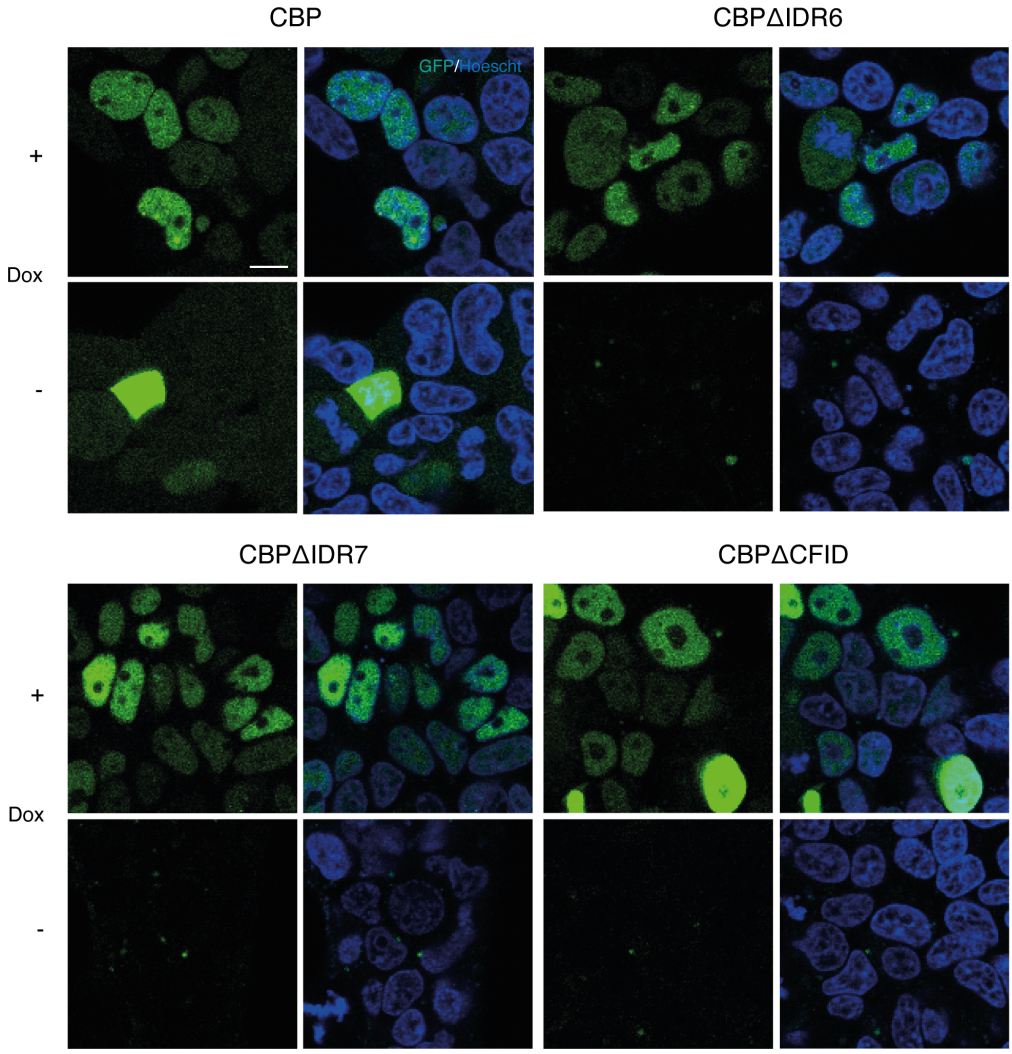

Figure 6.2: Initial screening of polyclonal populations of inducible CBP-GFP and mutants from FACS

Representative images of polyclonal cell lines, HEK293T-Tet3G-TRE3G-CBP-GFP (CBP), CBPΔIDR6-GFP (CBPΔIDR6), CBPΔIDR7-GFP (CBPΔIDR7) and CBPΔCFID-GFP (CBPΔCFID). Cells were cultured with (+) or without (-) doxycycline (Dox) for 48 hours. Nuclei were stained with Hoechst (blue) and imaged for Dox induction of GFP tagged protein (green). Images were processed in ImageJ/Fiji, scale bars represents 10 μm.

6.2.3 Screening of clonal lines

Clonal populations of all four cell lines were left to grow in their respective 96-well plate for approximately two weeks, at which point the cells were expanded sequentially until they were in 10 cm dishes. For each cell line, approximately 30 clones had survived the FACS procedure, these were screened in batches of eight in 8-chamber cover glasses at the point at which they were transferred into a 10 cm dish. The clones were screened after 48 hours in media either with or without Dox, at a concentration of 2 μg/ml.

One successful clone for each cell line was then used for further experiments (Figure 6.3). For each cell line GFP expression was observed for all cells within the population, and the tagged protein was nuclear localised when compared to the Hoechst (Figure 6.3). While there was some background GFP expression in the absence of Dox this was lower than was observed in the polyclonal population (Figure 6.3).

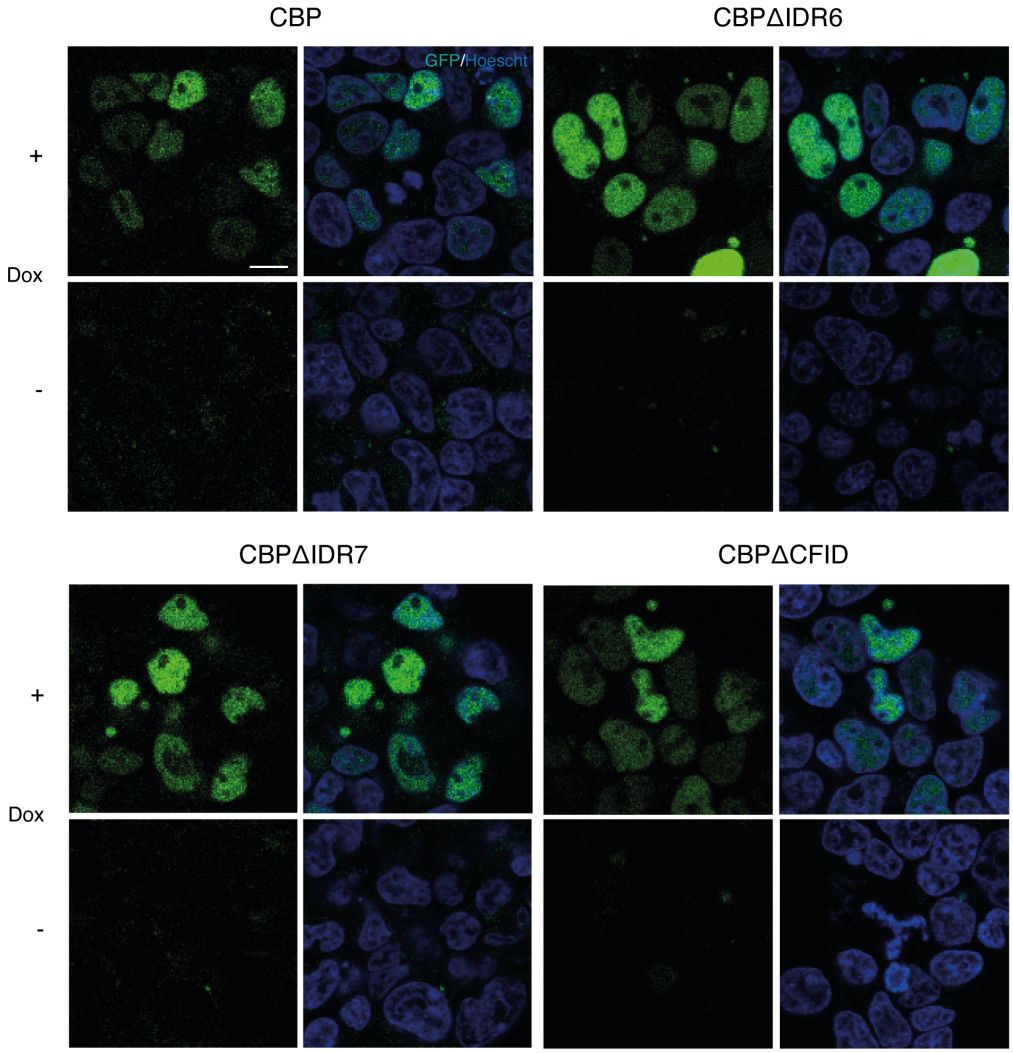

Figure 6.3: Screening of clonal lines containing CBP-GFP and mutants
Representative images of clonal lines of HEK293T-Tet3G-TRE3G-CBP-GFP (CBP-GFP), CBPΔIDR6-GFP (CBPΔIDR6), CBPΔIDR7-GFP (CBPΔIDR7) and CBPΔCFID-GFP (CBPΔCFID), cultured with (+) or without (-) doxycycline (Dox) for 48 hours. Nuclei were stained with Hoechst (blue) to check nuclear localisation of the protein and imaged for induced expression of GFP tagged protein (green). Images were cropped in ImageJ/Fiji, scale bar represents 10 μm. Images were not processed to remove background.

The main benefit of using the Dox inducible cell lines was that the expression was tunable

and provided a mechanism to visually understand how protein expression affects the phase separation of CBP and mutant constructs. Once clonal lines had been chosen, I next optimised Dox concentration to induce expression of the tagged protein to a concentration similar to that of the endogenous protein. Up to this point a concentration of 2 μg/ml of Dox was used to induce expression of GFP tagged constructs, this gave a clear indication of whether the viral transduction was successful and whether the clonal lines contain the gene of interest downstream of the Dox-inducible promoter. However, at this concentration significant overexpression of protein was observed, therefore I next wanted to perform a Dox response curve to finely tune the expression of CBP and mutants in the aforementioned cell lines.

6.2.4 Doxycycline induction curves for CBP-GFP and mutants

To determine the endogenous concentration of protein, I reduced the concentration of Dox, and performed a hybrid western blot approach, to check induction via imaging as well as by western blotting.

CBP-GFP, CBP$_{\Delta IDR7}$-GFP and CBP$_{\Delta CFID}$-GFP I performed a hybrid western blot for all cell lines, in this case coverslips were placed in the 10 cm dish where the cells were plated. A Dox concentration curve was used to produce lower than endogenous expression, endogenous expression, and overexpressed protein; the concentrations used were the following: 0 ng/ml, 10 ng/ml, 25 ng/ml, 50 ng/ml, 75 ng/ml, 100 ng/ml and 125 ng/ml. After 48 hours of Dox induction the coverslips were collected and fixed, and the cells harvested for western blot. In the case of CBP$_{\Delta IDR6}$-GFP, even at 10 ng/ml the expression of protein was saturated, I therefore repeated this experiment with different Dox concentrations - this experiment will be discussed in a subsequent section.

The results via imaging show non-specific GFP signal with 0 ng/ml of Dox, with increased expression of GFP tagged protein as the concentration of doxycycline was increased (Figure 6.4). These cell lines displayed the same phenotype as was expected, where puncta was observed for CBP-GFP and a more diffuse pattern observed for both CBP$_{\Delta IDR7}$-GFP and CBP$_{\Delta CFID}$-GFP (Figure 6.4).

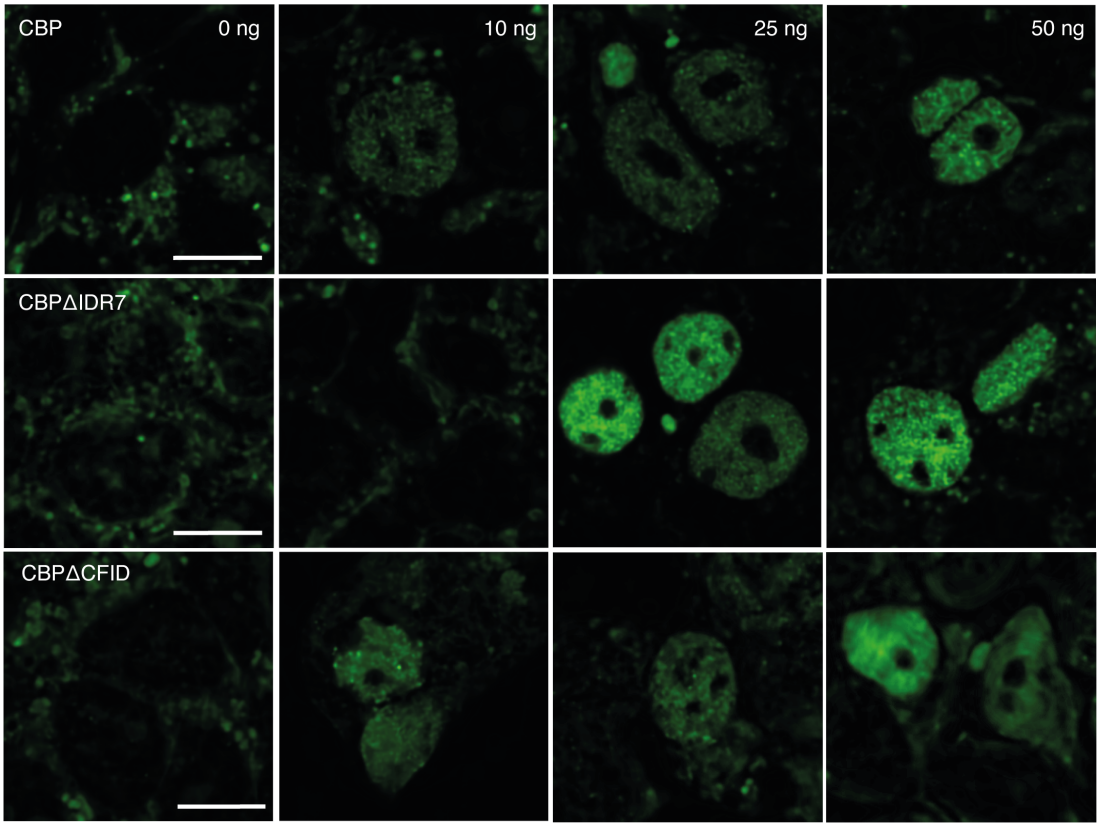

Figure 6.4: Representative images from Dox induction curve of CBP-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP
Visualisation of doxycycline (Dox) induction curve of HEK293T-Tet3G-TRE3G-CBP-GFP (CBP), HEK293T-Tet3G-TRE3G-$CBP_{\Delta IDR7}$-GFP ($CBP_{\Delta IDR7)}$and HEK293T-Tet3G-TRE3G-$CBP_{\Delta CFID}$-GFP ($CBP_{\Delta CFID}$) from hybrid western blot. Concentration of Dox is indicated in the top right image for CBP. Images were preprocessed using the Nikon denoise algorithm and processed in ImageJ/Fiji, scale bar represents 10 μm. Biological replicates = 1.

This was recapitulated by western blotting, where an increase in CBP (Cell Signalling, 7389) and GFP (Abcam, ab290) expression with increasing concentrations of Dox was observed for all cell lines (Figure 6.5). For each of the cell lines the position of the overexpressed protein on the western blot was different relative to the protein size; CBP-GFP was observed above that of the endogenous due to the presence of the GFP tag, $CBP_{\Delta IDR7}$-GFP was observed at the same height of the endogenous protein and $CBP_{\Delta CFID}$-GFP was observed below that of the endogenous as it is smaller in size (Figure 6.5, indicated by the blue arrows). For $CBP_{\Delta IDR7}$-GFP the increase

in GFP expression with Dox concentration was not as distinct as that seen for CBP-GFP, which was unexpected as when probed for CBP an increased expression was observed with a clear band present at 50 ng/ml (Figure 6.5). Since GFP expression was observed by imaging at 25 ng/ml a Dox concentration could still be chosen that produced near-endogenous expression. The blot was normalised to GAPDH (Proteintech, 60004-1). From this experiment a concentration of Dox was determined that would express the GFP tagged constructs to near or below endogenous expression, relative to the endogenous CBP expression per cell line. For CBP-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP this was determined to be 25 ng/ml (Figure 6.5). These western blots were not quantified as HRP-conjugated secondary anibodies were used, which are not quantitative due to amplification of signal. An additional challenge to quantification occurs when comparing endogenous and overexpressed protein due to poor separation of bands produced between the endogenous protein and overexpressed GFP tagged constructs.

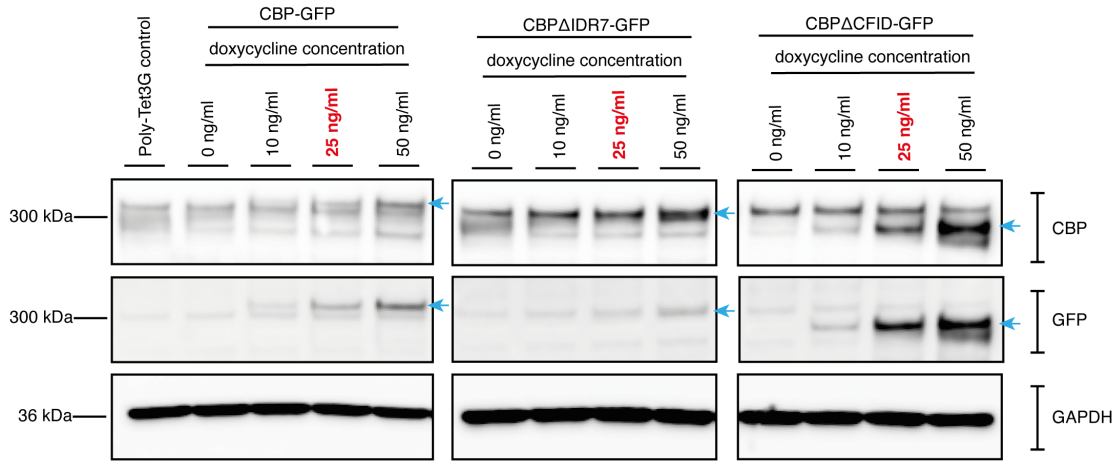

Figure 6.5: Hybrid western blot of CBP-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP Western blot of doxycycline (Dox) induction curve of HEK293T-Tet3G-TRE3G-CBP-GFP (CBP), HEK293T-Tet3G-TRE3G-$CBP_{\Delta IDR7}$-GFP ($CBP_{\Delta IDR7}$)and HEK293T-Tet3G-TRE3G-$CBP_{\Delta CFID}$-GFP ($CBP_{\Delta CFID}$) probed for CBP (upper panel), GFP (middle panel) and GAPDH (lower panel). Blue arrows indicate the band representative of induced protein, and the Dox concentration shown in red indicates the concentration chosen to give endogenous levels of expression. Approximate sizes of bands in kDa shown on the left, with the protein the antibody targets being shown on the right hand side. Images were processed in ImageJ/Fiji. Biological replicates = 1.

$CBP_{\Delta IDR6}$-GFP Since saturation was observed with $CBP_{\Delta IDR6}$-GFP at the first concentration, 10 ng/ml, used with the previous three constructs, I repeated the concentration curve, using the following Dox concentrations: 0 ng/ml, 1 ng/ml, 2.5 ng/ml, 5 ng/ml, 7.5 ng/ml and 10 ng/ml. The hybrid western blot showed non-specific GFP signal when no Dox was present, with an increase in expression of $CBP_{\Delta IDR6}$-GFP with increasing Dox concentrations (Figure 6.6 A). Puncta was observed when $CBP_{\Delta IDR6}$-GFP was expressed, which was consitent with its overexpressed phenotype (Figure 6.6 A). A similar expression pattern was observed by western blot, when probed for both CBP and GFP an increase of expression was observed for the tagged construct, where each lane had a similar GAPDH expression (Figure 6.6 B, blue arrow). $CBP_{\Delta IDR6}$-GFP is a similar size to that of the endogenous protein, therefore it was more difficult to distinguish expressed protein from the endogenous than in some of the other mutants. From this western I determined that a concentration of 5 ng/ml of Dox, for this mutant, was sufficient to give near-endogenous expression of $CBP_{\Delta IDR6}$-GFP (Figure 6.6 B).

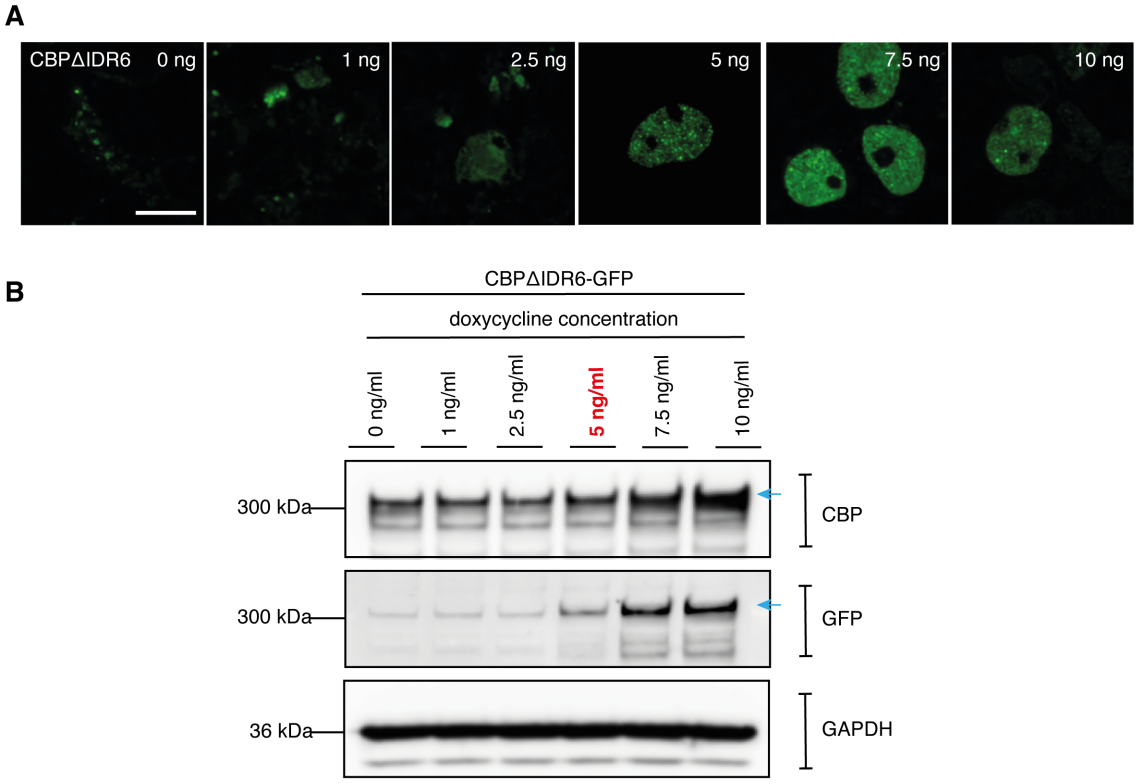

Figure 6.6: Hybrid western blot of Dox induction curve for CBP$_{\Delta IDR6}$-GFP
Doxycycline (Dox) induction curve of HEK293T-Tet3G-TRE3G-CBP$_{\Delta IDR6}$-GFP. A - Visualisation of expressed CBP$_{\Delta IDR6}$-GFP induced with a range of different doxycycline concentrations, which are shown in the top right corner of the image. Processing was performed using the Nikon denoise package, and further processing was done in ImageJ/Fiji, scale bar represents 10 μm. B - Western blot probed for using CBP antibody (upper panel), GFP antibody (middle panel) and GAPDH antibody (lower panel). The induced protein is shown by the blue arrow and the concentration chosen to give similar to endogenous expression is shown in red. Approximate sizes of bands in kDa shown on the left, with the protein the antibody targets being shown on the right hand side. Images were processed in ImageJ/Fiji. Biological replicates = 1.

From this data Dox concentrations were determined for each of the cell lines to give approximately the same concentration of the tagged protein compared to the endogenous CBP.

6.2.5 Increasing the expression of CBP-GFP through Dox induction in live cells

To supplement the hybrid western approach, I wanted to perform a live cell imaging study of CBP-GFP to observe what happens to the WT phenotype when the concentration is increased. Some background GFP was observed when no Dox was added, however this was not localised to the nucleus and therefore could be due to cellular autofluorescence (Figure 6.7). With the addition of 10 ng/ml of Dox, CBP-GFP expression was observed at a low level (Figure 6.7); from the hybrid western blot at this Dox concentration this is a lower expression of GFP tagged protein than the endogenous protein. Upon increasing the concentration of CBP-GFP puncta can be observed, which appear more distinct as expression was increased (Figure 6.7). This further confirms what was observed with the hybrid western blot, in that CBP-GFP expression is tunable with Dox concentration. While more replicates need to be performed to quantify how the increase in concentration affects the size and intensity of puncta, this data suggests that the phase separation of CBP could be influenced by concentration, supporting evidence in the literature that concentration is extremely important for studying phase separation (Chong et al. 2018; Erdel and Rippe 2018; Lin et al. 2015; Tsang et al. 2019; Sabari et al. 2018).

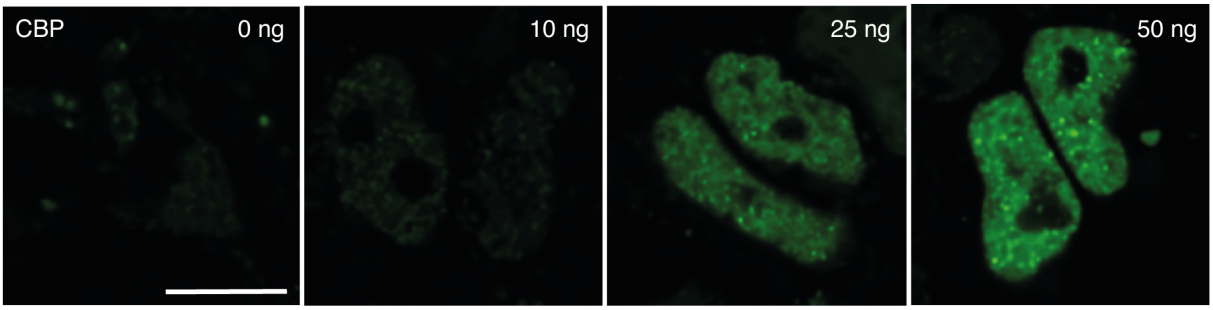

Figure 6.7: Representative images of doxycycline induction of CBP-GFP
Expression of CBP-GFP (green) in HEK293T-Tet3G-TRE3G-CBP-GFP cells induced with increasing concentrations of doxycycline (Dox), where Dox concentration is shown in the top right corner. Samples on the microscope and images were treated the same for comparative purposes. Images were processed in ImageJ/Fiji, scale bar represents 10μm. Biological replicates = 1.

This work suggests that CBP-GFP at near-endogenous levels, can form phase separated puncta in this Dox inducible cell line.

6.3 Imaging CBP and mutants at endogenous concentrations

Since I had determined concentrations of Dox which can produce near-endogenous levels of CBP-GFP and mutants in these cell lines, I next wanted to check whether the mutant constructs still behave phenotypically the same at endogenous concentrations as they did when overexpressed; to reduce the caveat of overexpression from conclusions drawn from previous overexpression work. Each cell line was cultured both with and without Dox for 48 hours, using the Dox concentrations that was calculated to give endogenous expression of constructs from section 6.2.4, before live cell imaging.

The results show that at endogenous concentrations, for each cell line, the GFP tagged construct behaves in the same manner as was observed for the overexpressed construct; for example, CBP-GFP and CBP$_{\Delta IDR6}$-GFP form puncta, while CBP$_{\Delta IDR7}$-GFP and CBP$_{\Delta CFID}$-GFP show a more diffuse pattern (Figure 6.8). This provided confidence in previous experiments, since CBPΔ_{IDR7}-GFP does not form condensates at endogenous levels, which was predicted to be one of the main contributing regions to CBP phase separation. At endogenous concentrations the puncta observed from CBP$_{\Delta IDR6}$-GFP were similar to that of CBP-GFP, with smaller puncta observed, compared to the large puncta visualised when CBP$_{\Delta IDR6}$-GFP was overexpressed (Figure 6.8). This was expected, as phase separation is heavily concentration dependent, smaller puncta observed could simply be a result of fewer molecules being expressed and were therefore available to contribute to condensate formation. For each cell line the respective negative controls shows only non-specifc GFP signal (Figure 6.8).

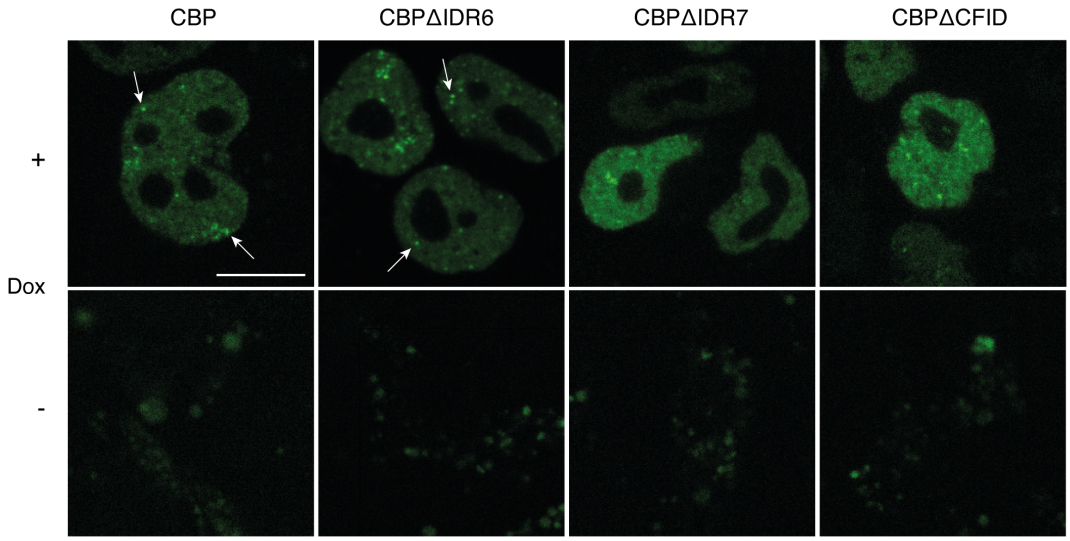

Figure 6.8: Imaging of endogenous concentrations of CBP-GFP and main mutants
Clonal doxycycline (Dox) inducible cell lines for CBP-GFP, CBP$_{\Delta IDR6}$-GFP, CBP$_{\Delta IDR7}$-GFP and CBP$_{\Delta CFID}$-GFP. Cells were cultured with (+) or without (-) Dox for 48-hours at a concentration of 25 ng/ml for CBP, CBP$_{\Delta IDR7}$-GFP and CBP$_{\Delta CFID}$-GFP and a concentration of 5 ng/ml for CBP$_{\Delta IDR6}$-GFP. Dox induction of the GFP tagged protein is shown in green, white arrows indicate observed puncta. Images were processed in ImageJ/Fiji, where background was removed with a rolling ball of 500 pixels, scale bar represents 10 μm.

Overall, this data shows that using Dox inducible GFP tagged cell lines that CBP IDR mutants behave phenotypically the same as when overexpressed.

6.4 How do mutants of CBP effect chromatin binding by ChIP-seq

This work has characterised differences in CBP condensate forming behaviour modulated by different CBP IDR's, and has demonstrated that these behavioural differences are still relevant at near-endogenous protein expression levels. Since CBP and p300 form condensates at enhancers, I next asked how these identified IDR mutants interact with chromatin. To do this, I carried out ChIP-Seq, comparing binding of CBP IDR mutants to both CBP-GFP, and endogenous CBP in the maternal cell line. As I wanted to test the effect of mutants with different phase behaviours, I chose to use CBP-GFP, CBP$_{\Delta IDR6}$-GFP and CBP$_{\Delta IDR7}$-GFP, as they form larger puncta than WT or a more diffuse pattern than WT. In this case, I selected not to test CBP$_{\Delta CFID}$-GFP, since

it displayed a similar phenotype to $CBP_{\Delta IDR7}$-GFP.

For ChIP-Seq experiments, cells were plated in 10 cm dishes and incubated either with or without Dox for 48 hours. The cell lines were cross-linked using formaldehyde, so that the protein was linked to the region of chromatin where it was bound. The cells were then harvested, sheared to generate chromatin fragments, and an immunoprecipitation experiment was done to retrieve the protein of interest and the regions of chromatin which it was bound to. Cross-linked DNA was then purified and sent for Illumina sequencing. Following sequencing, raw reads were aligned to the hg38 reference genome using Burrows-Wheeler Alignment (BWA) tool (Li and Durbin 2009), and peaks were called using Model-based Analysis of ChIP-Seq (MACS2) (Zhang et al. 2008; Van Rossum and Drake 2011), as part of the Nextflow ChIP-Seq pipeline (Done by Dr Mark Dunning and Dr Dan Bose).

This recalled 5091 peaks for endogenous CBP, 8687 peaks for CBP-GFP, 27,165 peaks for $CBP_{\Delta IDR6}$-GFP, and 8838 for $CBP_{\Delta IDR7}$-GFP (Table 6.1). This data showed that CBP-GFP produced more peaks than endogenous CBP, which could be a result of higher expression of CBP-GFP than the endogenous protein, caused by Dox induction or as a result of a more efficient immunoprecipitation (Table 6.1). A similar number of peaks were counted for $CBP_{\Delta IDR7}$-GFP compared to CBP-GFP, which suggested a similar chromatin binding profile of these constructs (Table 6.1). In contrast, $CBP_{\Delta IDR6}$-GFP showed an increased number of peaks of approximately 27,000 compared to approximately 8,600 for CBP-GFP (Table 6.1).

Sample	**Peaks called**
endogenous CBP	5091
CBP-GFP	8687
$CBP_{\Delta IDR6}$-GFP	27165
$CBP_{\Delta IDR7}$-GFP	8838

Table 6.1: Number of peaks called for ChIP-Seq
Peaks called from Burrows-Wheeler Alignment using MACS2 as part of the Nextflow ChIP-Seq pipeline (Zhang et al. 2008; Li and Durbin 2009). Work was done by Dr Mark Dunning and Dr Dan Bose.

Next, called peaks and distribution of reads was examined by visualising them in a UCSC genome browser (Kent et al. 2002). The tracks show accumulation of CBP reads and peaks around regulatory elements (ENCODE cCREs), and demonstrate overlap of CBP-GFP with sites of endogenous CBP binding (Figure 6.9 A). The tracks also suggest that more reads were present for CBP-GFP than endogenous CBP, and a further increase was observed for $CBP_{\Delta IDR6}$-GFP

(Figure 6.9 A). Both observations were expected as I induced expression of both proteins and even though I optimised for near-endogenous concentrations, I may have a higher expression level of the GFP tagged constructs compared to endogenous expression. The UCSC genome browser view also suggest a greater accumulation of reads and broader peaks for $CBP_{\Delta IDR6}$-GFP at individual loci. I therefore asked whether this was the case across all peaks (Figure 6.9 B). When comparing different lengths of peaks called, CBP-GFP displayed a similar profile to that of the endogenous protein (Figure 6.9 B). In contrast, $CBP_{\Delta IDR6}$-GFP showed a shift in peak length towards longer peaks called, and $CBP_{\Delta IDR7}$-GFP showed the opposite shift towards shorter peak lengths (Figure 6.9 B). This reflected the phase behaviour of both mutants, since $CBP_{\Delta IDR6}$-GFP forms larger condensates it could interact with more regions of chromatin increasing the number and length of peaks called; similarly, $CBP_{\Delta IDR7}$-GFP is expressed more diffusely, and is therefore not concentrated in particular regions, which could be a reason for the decrease in peak length (Figure 6.9 B).

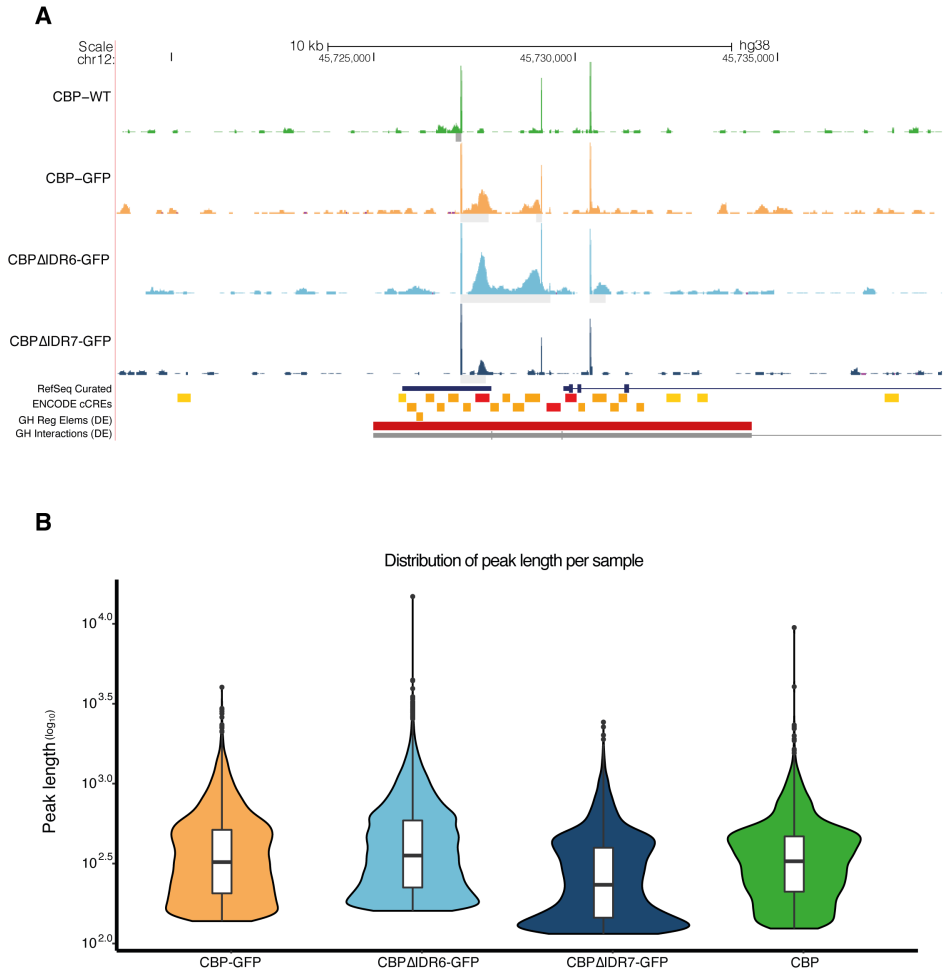

Figure 6.9: UCSC genome browser and peak distribution for ChIP-Seq

ChIP-Seq data from cell lines: polyclonal HEK293T-Tet3G endogenous CBP (green), and inducible CBP-GFP (orange), $CBP_{\Delta IDR6}$-GFP (light blue) and $CBP_{\Delta IDR7}$-GFP (navy). A - ChIP-Seq CBP peaks from UCSC genome browser at representative genome locus, chr12:45,725,000-45,735,000 . Peaks called are shown by light grey bars under each sample, individual reads are coloured according to sample. Reference annotations from UCSC genome browser are as follows: NCB1 RefSeq genes curated annotations (RefSeq Curated); ENCODE candidate Cis-Regulatory Elements combined for all cell types (ENCODE cCREs); enhancer and promoter GeneHancer Double Elite (GH Reg Elems (DE)), and Interactions between GeneHancer regulatory elements and genes Double Elite (GH Interactions (DE)) (Kent et al. 2002). B - Violin plot of peak lengths called for each sample on a Log_{10} scale, with boxplot to show maximum value, upper quartile, median, lower quartile and minimum value. Biological replicate = 1. Plot was made by Dr Mark Dunning, using MACS2 python package (Zhang et al. 2008).

Since differences in number and length of peaks called for each of the different constructs had been observed, and also accumulation of these peaks at regulatory regions, I next asked how many peaks overlap between the different constructs. Firstly, Bedtools was used to insersect endogenous CBP peaks with peaks from each of the different GFP tagged constructs (Figure 6.10). For CBP-GFP, over 57 % of peaks intersect with mapped endogenous CBP (Figure 6.10 A); a similar proportion of peaks overlapped between $CBP_{\Delta IDR7}$-GFP and endogenous CBP at 59 % (Figure 6.10 B). A higher percentage of peaks overlap between $CBP_{\Delta IDR6}$-GFP and endogenous CBP at 69 % (Figure 6.10 B), this was expected due to the increase in number of peaks and peak length for $CBP_{\Delta IDR6}$-GFP, which would increase the chance of intersection of the peaks observed between these two proteins. While over half of the peaks from theGFP tagged constructs map to peaks identified for endogenous CBP, this was only based on one experimental repeat.

Since I have shown that the GFP tagged constructs localise to endogenous CBP at common loci, I next wanted to understand how the different deletions, and different phase behaviours, affect chromatin localisation. The overlap of CBP-GFP peaks to $CBP_{\Delta IDR6}$-GFP and $CBP_{\Delta IDR7}$-GFP was then compared (Figure 6.11). The results showed a 77 % overlap in peaks called from $CBP_{\Delta IDR6}$-GFP (Figure 6.11 A), and 43 % overlap with peaks called for $CBP_{\Delta IDR7}$-GFP (Figure 6.11 B), with CBP-GFP. It was interesting that a higher percentage of peaks overlap with $CBP_{\Delta IDR6}$-GFP which forms large phase separated condensates, compared to $CBP_{\Delta IDR7}$-GFP which is more diffusely expressed within the cell. The higher proportion of peaks that overlapped between CBP-GFP and $CBP_{\Delta IDR6}$-GFP was reflective of the greater number and size of peaks observed for the mutant construct, which could be due to the increased propensity of this construct to phase separate (Figure 6.11 A). Conversely, it was interesting that only 43 % of CBP-GFP and $CBP_{\Delta IDR7}$-GFP peaks intersected, since a similar number of peaks were called for each construct, at approximately 8,700. This suggested that the removal of the IDR7 region, and its ability to phase separate, affects the chromatin localisation compared to its full length counterpart. While this implies a role for the influence of these regions and phase separation on CBP function, further experiments need to be taken in order to fully understand the implications of this data.

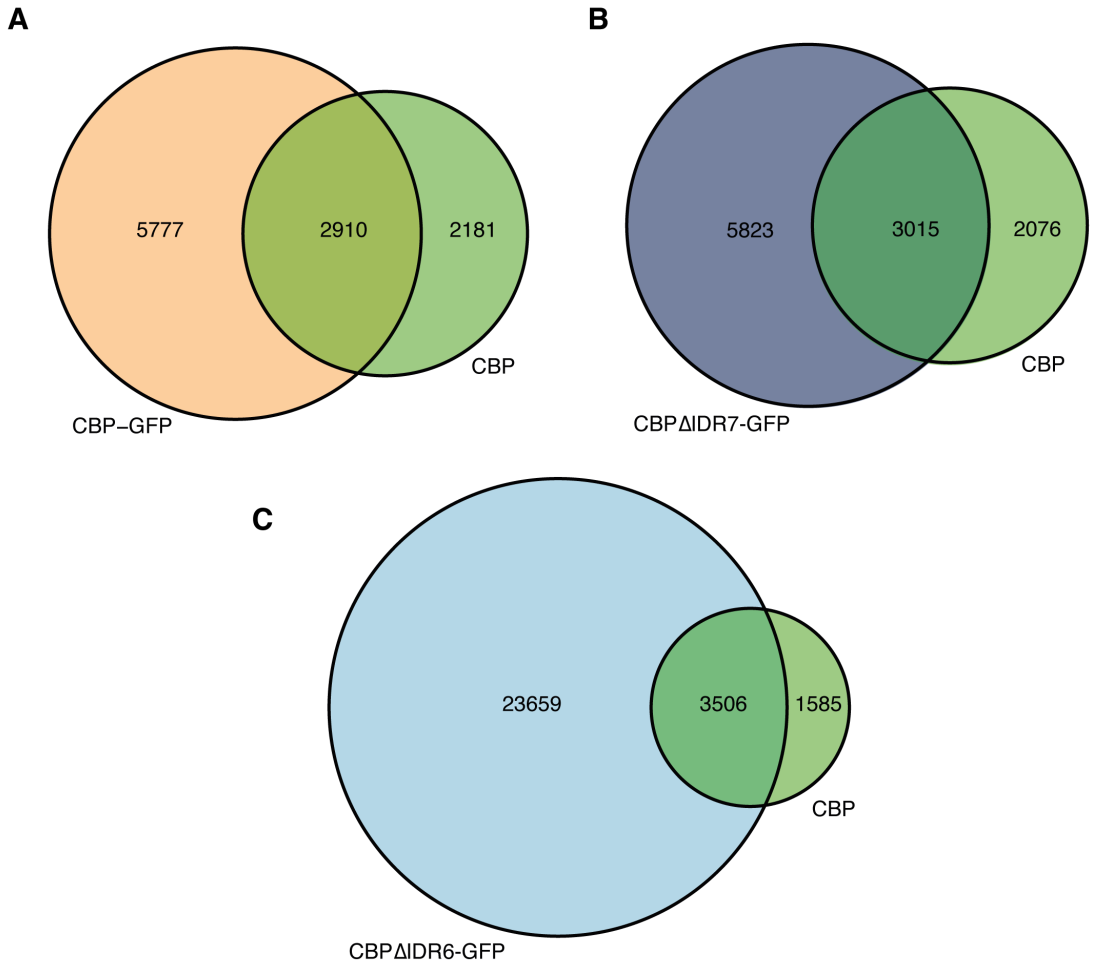

Figure 6.10: Intersecting peaks between endogenous CBP and CBP-GFP constructs
Venn diagram to show the number of intersecting ChIP-Seq peaks between endogenous CBP (green) to each of the GFP tagged constructs: A - CBP-GFP (orange); B - $CBP_{\Delta IDR6}$-GFP (light blue), and C - $CBP_{\Delta IDR7}$-GFP (navy). Peak calling was performed by Dr Mark Dunning, using MACS2 python package (Zhang et al. 2008), data analysis was performed by Dr Dan Bose. Venn Diagrams were made in Rstudio using R.

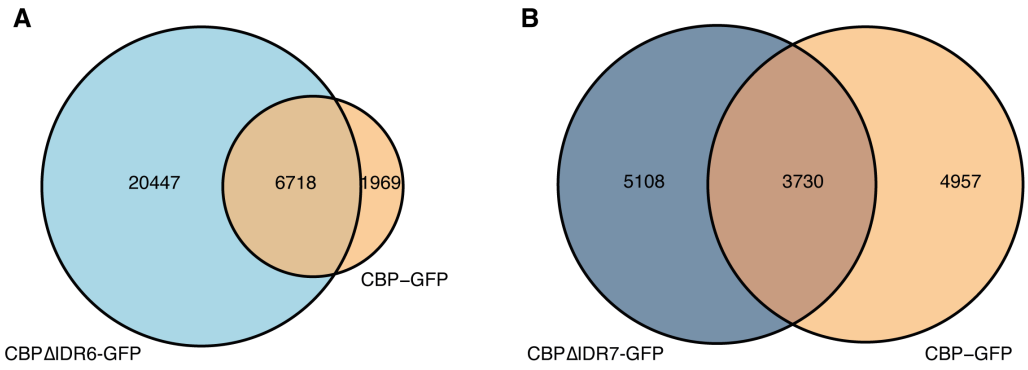

Figure 6.11: Intersecting peaks between CBP-GFP and CBP deletion constructs
Venn diagram of intersecting ChIP-Seq peaks between CBP-GFP (orange) and A - $CBP_{\Delta IDR6}$-GFP (light blue), or B - $CBP_{\Delta IDR7}$-GFP (navy). Peak calling was performed by Dr Mark Dunning, using MACS2 python package (Zhang et al. 2008), data analysis was performed by Dr Dan Bose. Venn diagrams were made in Rstudio using R.

Taken together, this work suggested that the GFP tagged constructs localised to the same chromosome locations as endogenous CBP, implying function of these tagged proteins within regulatory regions. This data also implied that mutants $CBP_{\Delta IDR6}$-GFP and $CBP_{\Delta IDR7}$-GFP influence chromosome contacts, where different phase behaviours were linked to the number and length of peaks identified. Due to time constraints only one replicate of this experiment has been analysed, a second repeat has been completed and will be combined with this data for future analysis, to provide confidence in the data shown within this doctoral work.

6.5 The importance of sequence: differences between IDR6, IDR7 and CFID

In order to understand how these important IDRs affect phase separation and chromatin binding, I wanted to investigate the amino acid sequence of these regions to identify any key residues which could explain their phase separating behaviour. Sequence has been shown, in some cases, to be important for phase separation. IDRs can have enriched populations of amino acids or patterns of amino acids that facilitate phase separation (Martin et al. 2020; Zeng et al. 2020; Weiner, Meir and Wingreen 2021; Lin, Currie and Rosen 2017; Lin and Chan 2017; Pak et al. 2016). In this work, I have shown that the IDR6 and IDR7 have opposing effects on regulating CBPs phase

separating ability. I therefore wanted to see if the differences observed in phase separation could be due to difference in the sequences of these regions. To compare sequence properties within these regions of interest Classification of Intrinsically Disordered Ensemble Relationships (CIDER) or localCIDER was used (Holehouse et al. 2017). Sequence analysis by CIDER, localCIDER and PLAAC was performed by Dr Dan Bose.

First, I looked at all of the regions of interest as well as full length CBP to compare the proportion of amino acids within each sequence which contains proline, aromatic or charged residues. These factors may influence the weak interactions that facilitate phase separation. The results for this are shown in Table 6.2. From this data analysis, the IDR6 region has a higher percentage of proline residues (21.2%), compared to the IDR7 (13.8%), CFID (16.5%) and the full length protein (11.4%). The IDR6 also has a lower number of aromatic residues (1.3%) compared to the full length protein (4.4%). It was also observed that the AIL region has a high percentage of charged residues compared to all other constructs, this is due to the high proportion of lysine residues found within the AIL that are important for HAT activity (Thompson et al. 2004; Thompson et al. 2001; Liu et al. 2008; Wang et al. 2008a; Ortega et al. 2018).

Domain	**Proline**	**Aromatic**	**Charged**
CBP-FL	0.114	0.044	0.162
IDR1	0.106	0.000	0.058
IDR3	0.136	0.014	0.041
IDR4	0.167	0.008	0.103
AIL	0.025	0.051	0.443
IDR6	0.213	0.013	0.089
IDR7	0.138	0.009	0.055
CFID	0.165	0.016	0.074

Table 6.2: Comparison of amino acid residues in regions of interests
Proportion of amino acids that are prolines, aromatic or charged in CBP-FL (full length) and the different regions of interest. Values generated using localCIDER (Holehouse et al. 2017).

Alongside this, values such the fraction of charged residues (FCR), the pattern of alternatively charged amino acids (Kappa, K) and and the pattern of proline residues in relation to charged residues (Omega, Ω) was investigated. The results are shown in Table 6.3. In looking at these values, the only difference observed was for the AIL for the FCR ratio - this was due to the large number of lysine residues needed for regulation of the HAT loop. No other differentiating sequence features were observed for any of the regions of interest.

Domain	FCR	Kappa	Omega
CBP-FL	0.162	0.203	0.240
IDR1	0.058	0.274	0.435
IDR3	0.041	0.365	0.368
IDR4	0.103	0.216	0.306
AIL	0.443	0.272	0.113
IDR6	0.089	0.187	0.161
IDR7	0.055	0.206	0.385
CFID	0.074	0.164	0.278

Table 6.3: Sequence parameters of all regions of interest
Comparison of parameters FCR, Kappa and Omega for CBP-FL (full length) and all regions of interest. Where FCR refers to the fraction of charged residues in the sequence, Kappa refers to the charge separation throughout the sequence as defined by Das and Pappu, and Omega defines the patterning of charged and proline residues within the sequence. Values generated using localCIDER (Holehouse et al. 2017).

Focussing more specifically on the C-terminal regions of interest IDR6, IDR7 and CFID, the amino acid sequence and the distribution and ratio of amino acid qualities over the sequence was identified (Figure 6.12). From the sequence of the C-terminal region alone there are multiple glutamine (Q) residues distributed throughout the IDR6 and IDR7 region, including the polyglutamine stretch of 18 amino acids in the IDR7 (Figure 6.12 A).

Using localCIDER the contribution of different amino acid classes to each overall sequence was calculated, catagorised by charged, polar, aliphatic, aromatic amino acids or proline residues (Figure 6.12 B). From this data analysis the IDR6 has a larger proportion of proline and charged amino acids and fewer polar amino acids than the IDR7 and the CFID, where as the IDR7 and CFID show a similar distribution of amino acids throughout the different categories (Figure 6.12 B). The pattern of these residues over the span of the sequence was also investigated, which highlights a higher level of polar residues in the IDR7 region compared to the IDR6 (Figure 6.13 A). By separating the CFID region into the IDR6 (Figure 6.13 B) and IDR7 (Figure 6.13 C) regions, no specific pattern unique to each was observed.

Prions are proteins regions that confer different states which may be functionally relevant, they are typically associated with disease as prions form protein aggregates which lead to neurodegeneration (Alberti and Hyman 2021; Alberti et al. 2009; Piroska et al. 2023; Wang et al. 2018; Martin et al. 2020). Using PLAAC analysis it was tested whether the C-terminal region was predicted to form a prion like domains. The IDR7 region was predicted to form one contiguous

prion-like domain in contrast to the IDR6 region which does not display this behaviour. This therefore suggests a propensity for IDR7 to form contacts that leads to protein accumulation.

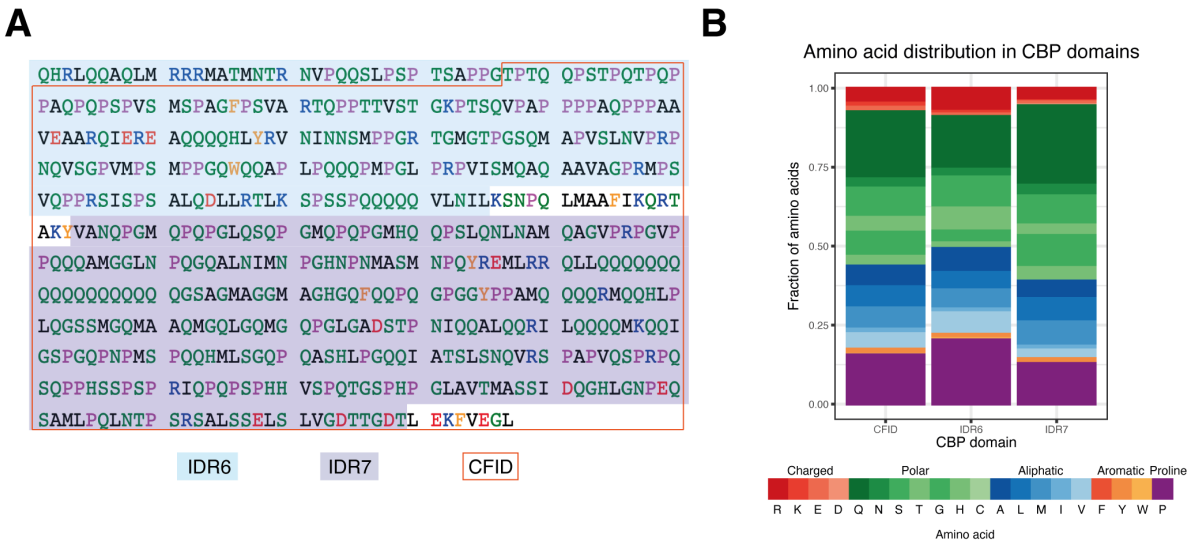

Figure 6.12: Sequence comparison of IDR6, IDR7 and CFID
A - amino acid sequence of the C-terminal region of CBP highlighting the IDR6 region in light blue, the IDR7 region in dark blue and the CFID region is outlined in an orange box. B - distribution of properties of amino acids within the IDR region including charged residues (red), polar residues (green), aliphatic residues (blue), aromatic residues (orange) or proline residues (purple). Sequence analysis was performed with the help of Dr Dan Bose using localCIDER (Holehouse et al. 2017)

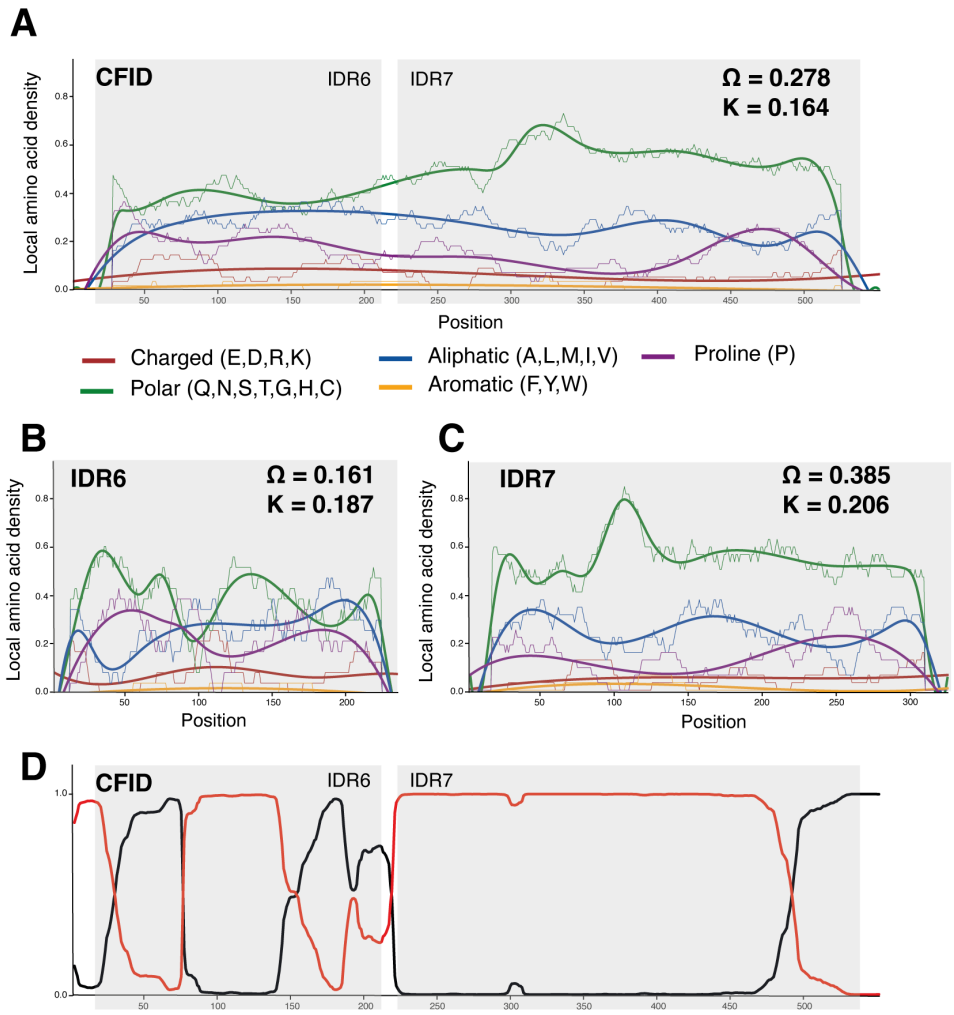

Figure 6.13: Diagramatic representation of sequence parameters of IDR6, IDR7 and CFID A/B/C - local density of amino acids displaying the property of charged residues (red), polar residues (green), aliphatic residues (blue), aromatic residues (orange) or proline residues (purple) over the length of the C terminal region for the CFID, IDR6 and IDR7 respectively. Where K refers to the Kappa value of patterning of charged residues, and Ω refers to Omega, the patterning of proline residues with respect to charged residues. Graphs were generated by Dr Dan Bose using localCIDER (Holehouse et al. 2017). D - Output of cPrD prediction algorithm using PLAAC analysis to determine the likely hood of the amino acids being part of a prion-like domain; where a probability of 1 represents a predicted prion domain. Graph was generated by Dr Dan Bose using methods described in (Alberti et al. 2009).

Taken together, the sequence components of IDR7 contain a number of sequence properties which favour condensate formation, where the IDR6 region contains sequence properties which could oppose phase separation. This provides initial insight into the influence of these regions on the phase separation of full length CBP.

6.6 Analysing the effect of overexpressed protein on CBP-HaloTag condensates

Within the previous chapters the ability of full length CBP to phase separate has been described, both when overexpressed and at endogenous concentrations. I have also demonstrated the differing phase separation abilities of CBP deletion mutants, specifically the IDR6, IDR7 and CFID deletions. Since I have characterised how CBP behaves endogenously and how these mutants behave when overexpressed, I next wanted to perform a colocalisation experiment to determine if overexpressing deletion mutants in the CBP-HaloTag cell line alters its behaviour.

Due to the efficiency of the transient transfection I used a clonal cell line of CBP-HaloTag to increase the probability of transfecting a cell which contained tagged CBP, the validation for the CBP-HaloTag clonal line is shown in section 7.4.3. To visualise colocalisation I labelled the HaloTag with JF 646 and overexpressed the GFP tagged constructs, images were taken of both the 488 channel and the 637 channel and the images overlaid to visualise colocalisation.

6.6.1 Checking for bleed-through

Before performing a colocalisation experiment, I wanted to check that overexpressed GFP would not bleed-through into the 637 channel, to provide confidence that the overlap of signal observed was due to colocalisation. CBP-HaloTag cells were transiently transfected with CBP-GFP, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP, or $CBP_{\Delta CFID}$-GFP for 24 hours before imaging in 4 channels using the 405 nm, 488 nm, 532 nm and 637 nm lasers to see if any bleed-through was observed. An untransfected well that contained only CBP-HaloTag labelled with JF 646 was also tested to check for bleed-through from this channel. When untransfected, CBP-HaloTag cells labelled with JF 646 showed signal of nuclear puncta within the 637 channel with no labelling in any other channels (Figure 6.14). For all overexpressed constructs, the signal was seen within the 488 nm channel, with each construct's respective phase behaviour, and no bleed-through into the other channels was observed (Figure 6.14). For both 532 nm and 637 nm some non-specific signal was observed, however this did not colocalise with the signal seen for GFP (Figure 6.14).

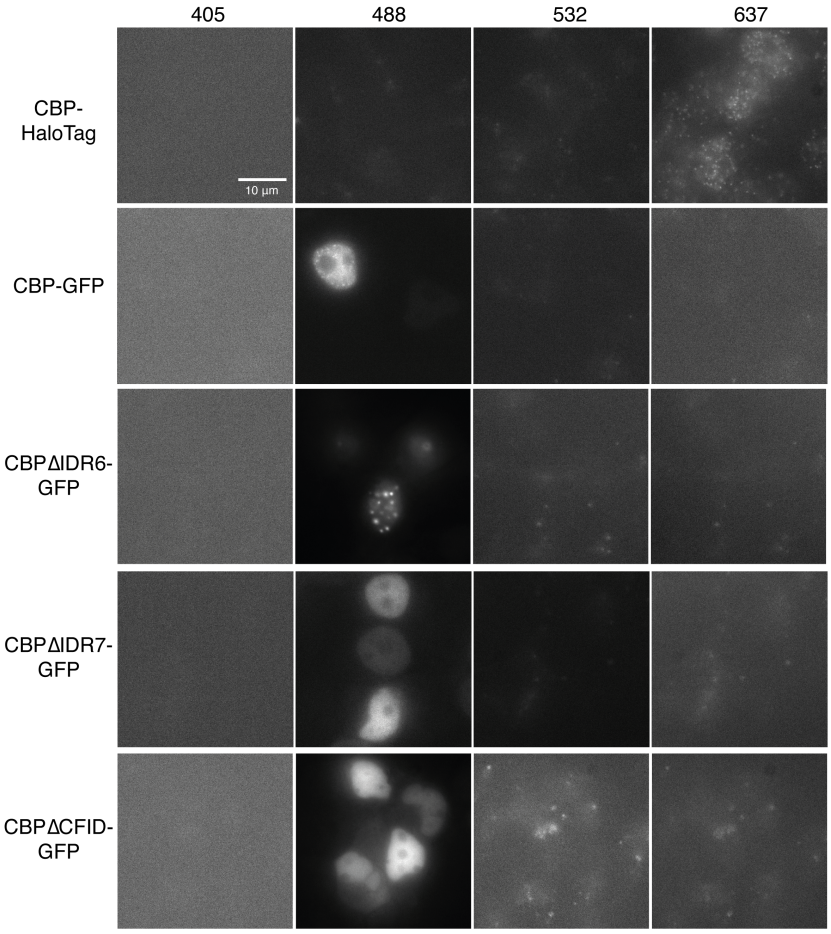

Figure 6.14: 4 colour control images for overexpression of GFP tagged constructs in clonal CBP-HaloTag cells
Representative images of HEK293T-CBP-HaloTag cell line labelled with Janelia Fluor 646 (JF 646) or transfected with CBP-GFP, CBPΔIDR6-GFP, CBPΔIDR7-GFP or CBPΔCFID-GFP. Images were taken using 4 lasers - 405 nm, 488 nm (GFP tagged constructs), 532 nm and 637 nm (endogenously labelled CBP JF 646). Images were processed in ImageJ/Fiji, scale bar represents 10 µm. Biological replicates = 1.

Since no bleed-through was observed from the overexpressed GFP constructs into the 646 channel I continued onto the colocalisation experiment.

6.6.2 Colocalisation of endogenous CBP-HaloTag with overexpressed GFP tagged constructs

After checking for bleed-through the samples were labelled individually with 2 nM of JF 646; images were captured of the overexpressed GFP tagged constructs and endogenous CBP to observe colocalisation.

Interestingly, when CBP-GFP was overexpressed, puncta were observed for both overexpressed and endogenously tagged CBP, with condensates colocalising between the channels as indicated by the red arrows (Figure 6.15). This suggests that endogenous CBP was incorporated into the puncta formed by overexpression of the CBP-GFP construct, and is consistent with the ChIP-Seq data showing localisation of CBP-GFP at the same chromosomal loci as endogenous CBP. In some nuclei there was mass overexpression of GFP tagged CBP and other constructs, which would not be imaged as the signal would be too high and not representative of the population of transfected cells; these nuclei are analogous to the overexpressed nuclei removed by filtering in section 4.4.2. Interestingly, when there was a higher amount of overexpressed protein within the cell a reduction in signal was observed for the endogenously tagged HaloTag.

Similarly to CBP-GFP, when $CBP_{\Delta IDR6}$-GFP was overexpressed, large puncta were observed in both the 488 channel and the 637 channel (Figure 6.15). The endogenous puncta were larger than those typically seen for the endogenous protein, and these puncta colocalise extremely well with the overexpressed $CBP_{\Delta IDR6}$-GFP puncta, suggesting a high level of interaction between this mutant and the endogenous protein (Figure 6.15). For both $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP no puncta formed with the overexpressed protein, and fewer endogenous puncta were observed within the nuclear volume compared to when untransfected (Figure 6.15). While the number of endogenous puncta when transfected with each construct has not been quantified, this suggests a dominant effect of the overexpressed protein on the endogenous. For example in the presence of $CBP_{\Delta IDR6}$-GFP larger endogenous condensates were observed, and when transfected with $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP a dominant negative effect was observed where there were fewer endogenous puncta. Due to time constraints only one replicate of this experiment was performed.

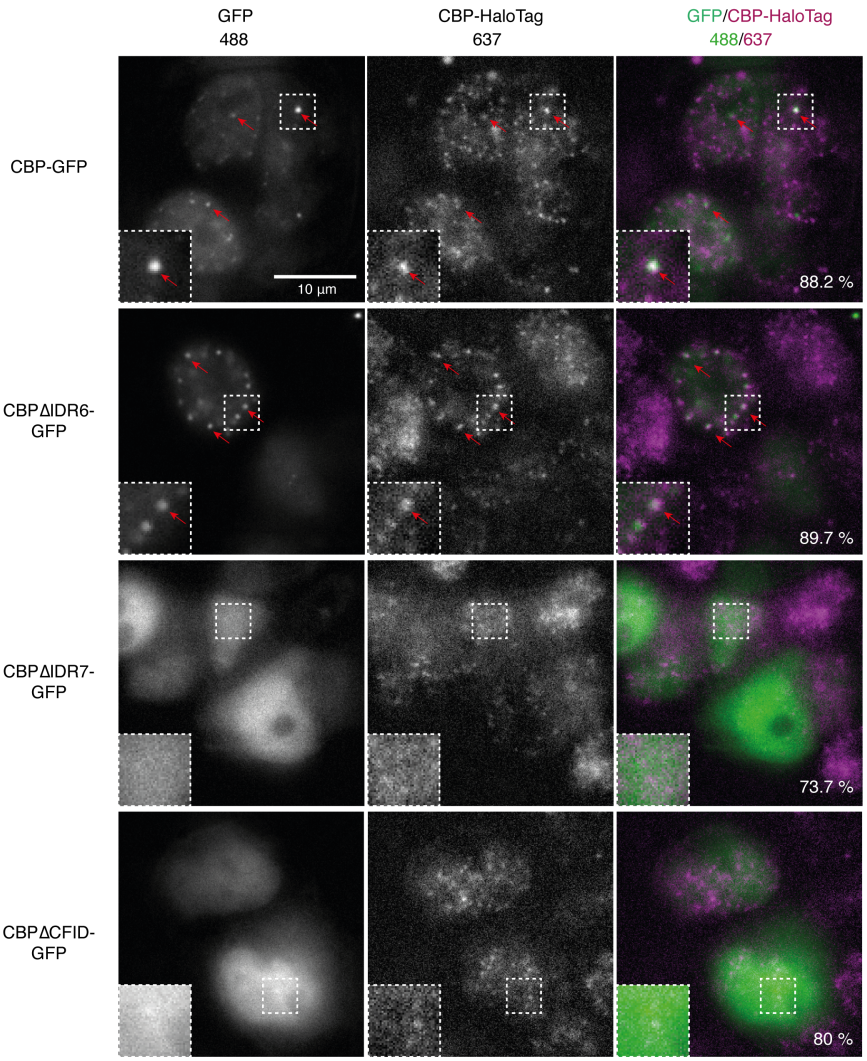

Figure 6.15: The effect of overexpression on endogenous CBP-HaloTag puncta
Representative images of CBP-HaloTag labelled with Janelia Fluor 646 and transfected with overexpressed constructs CBP-GFP, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP or $CBP_{\Delta CFID}$-GFP to visualise colocalisation of endogenous and overexpressed CBP puncta. Images of overexpressed GFP tagged constructs (green) were merged with endogenous CBP-HaloTag puncta (magenta). Examples of puncta colocalisation is shown by the red arrows, with a zoomed in region highlighted by the white dashed box. Percentage of nuclei sampled where the overexpressed protein has a dominant effect on the endogenous protein is shown on the bottom right of the combined image. Processing and colocalisation was performed in ImageJ/Fiji, where background was subtracted using a rolling ball of 100 pixels. Scale bar represents 10 µm. Number of nuclei sampled per construct ranged from 17 - 57 nuclei. Biological replicates = 1.

The data shown here suggests that endogenous CBP phase behaviour is influenced by that of the overexpressed CBP deletion mutants, the impact this has on CBPs activity is not yet clear.

6.7 Discussion

In this chapter, I generated Dox inducible cell lines for CBP-GFP, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP, $CBP_{\Delta CFID}$-GFP to express these constructs at near-endogenous levels, to investigate the effect of protein expression on phase separation of CBP and mutants. I further show how the different phase separating phenotypes affect the localisation of CBP within the genome at endogenous concentrations, removing the concentration caveat. Finally, I described some sequence factors which may influence the phase behaviours of key regions of IDR6, IDR7 and CFID, and show that when overexpressed these different deletion mutants affect the phase separation of endogenous CBP.

6.7.1 The doxycycline induction system

I generated inducible stable cell lines using the Retro-X Tet-On 3G Inducible Expression System (Clontech), which was previously established in the lab. This system was the most efficient with smaller plasmids, for example the Retro-X Tet-3G transactivator protein (approximately 8.1 Kb) plasmid transduced more efficiently than the TRE3G-CBP-GFP (approximately 14.6 Kb) plasmid, which did not transduce well and took longer to select for and recover. Polyclonal HEK293T-Tet-3G cell lines were transduced with the plasmid of interest, for example TRE3G-CBP-GFP. After FACS these cell lines, both polyclonal and clonal, differed in expression of protein in response to Dox, this is thought to be a result of using a polyclonal population of HEK293T-Tet-3G which have different levels of transactivator protein leading to the differing protein expression observed in the final experiments. I chose one clone for each cell line to perform further experiments, however I have more clones frozen down which could act as biological replicates. I did not validate other clones, by Dox concentration curve, however for future experiments having different biological replicates being clonal lines would be beneficial.

To validate the cell lines used in this doctoral work, Dox induction curves were used which measured the CBP, GFP and GAPDH expression by hybrid western blot. This experiment showed that at low Dox concentrations all cell lines were responsive to Dox in a concentration dependent manner. Using the hybrid western blot approach a concentration of Dox was determined that would express the GFP tagged protein to a similar expression to that of their individual

endogenous CBP concentration. This experiment was not quantified as a result of the use of HRP-conjugated secondary antibodies and the difficulties in separating endogenous bands compared to GFP tagged overexpressed protein. For future experiments it would be beneficial to repeat this experiment using fluorescent secondary antibodies, and quantify the dox induction to saturation for each of these cell lines. By using different fluoroscent secondary antibodies the different bands could be quantified with a greater degree of certainty, which would allow a more quantitative Dox concentration to be determined. This method did however provide a concentration for each cell line which would generate near-endogenous concentrations of protein. For CBP-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP this was 25 ng/ml, compared to $CBP_{\Delta IDR6}$-GFP which used a concentration of 5 ng/ml. The $CBP_{\Delta IDR6}$-GFP cell line was much more sensitive to Dox than the other three cell lines - this could be an intrinsic property of this mutant that it was expressed to a greater amount, or this could be due to this clone having a higher amount of transactivator protein from the polyclonal population. It is likely that this was a property of $CBP_{\Delta IDR6}$-GFP as it was also expressed more in transient transfection, as seen in previous work, section 4.4.1.

Once I had validated these cell lines, I next wanted to determine how these mutant constructs affect chromatin localisation at endogenous concentrations.

6.7.2 Phase separation affects the chromosome binding ability of CBP

In this chapter I used ChIP-Seq to investigate the localisation of induced CBP-GFP, $CBP_{\Delta IDR6}$-GFP and $CBP_{\Delta IDR7}$-GFP compared to endogenously untagged CBP. This allowed the determination of whether the GFP in CBP-GFP was affecting its localisation compared to the untagged protein, and whether the $CBP_{\Delta IDR6}$-GFP and $CBP_{\Delta IDR7}$-GFP mutants were affecting chromatin localisation.

The ChIP-Seq data showed that the GFP tagged constructs localise to similar regulatory regions in the genome as endogenous CBP. A greater number of peaks for CBP-GFP was observed compared to the endogenous protein. This could be a result of overexpression of CBP-GFP, it could also be possible that these differences were due to the antibodies used for the IP; since endogenous protein was isolated using a CBP antibody and the GFP tagged constructs were isolated using GFP. Differences in number of peaks identified could therefore be caused by different affinities of the antibody to the protein, and different efficiencies of the IP itself. It was observed that a greater number of peaks and peaks larger in size compared to CBP-GFP for $CBP_{\Delta IDR6}$-GFP, and conversely, a decrease in length of peaks for $CBP_{\Delta IDR7}$-GFP (Table 6.1, Figure 6.9).

Given the very different condensate forming behaviours of these constructs, these results suggest a role for phase separation on the chromosome binding of CBP. One caveat of this work is that in the GFP tagged cell lines there is still endogenous CBP, and therefore at this stage it cannot be confirmed whether these GFP tagged proteins bind to the same chromosomal regions as the endogenous, or whether they are just binding to the enodgenous protein at these regions. In the future, this could be tested through a knockdown of endogenous CBP in these Dox induction cell lines.

In this dataset I have compared intersecting peaks between samples, however it is difficult to quantitatively compare between ChIP-Seq datasets without using spike-in controls. If further repeats of this experiment are to be performed, a spike-in should be used, for a more thorough comparison of chromatin localisation. A second biological replicate has been performed, but not analysed at the time of submission; this analysis would allow optimization of peak calling and a more stringent analysis pipeline which would provide more confidence in these observations, and potentially a greater degree of overlap between called peaks from the different proteins.

While this implicates the importance of phase separation on CBP function, it is still unknown whether these constructs are active in their different phase separation states, and whether in removing the IDR6 and IDR7 regions the binding of RNA and other transcription factors have been affected which may be impairing function. Further studies to identify HAT activity and RNA binding will need to be done to fully understand the role phase separation plays on CBP function.

6.7.3 Influence of sequence on phase separation

All of this work has highlighted the C-terminal region of CBP as being important for influencing phase behaviour; specifically the IDR6 in inhibiting phase separation, and the IDR7 as promoting phase separation. In section 1.1.1 I discussed sequence factors which are enriched within phase separating proteins including amino acids, patterns of amino acids and interactions caused by these specific residues. I investigated these factors within the C-terminal region to determine if factors in the amino acid sequence can account for the difference in phase behaviour of these different regions.

6.7.3.1 Increased proline residues and decreased aromatic residues could prevent IDR6 phase separation

From the localCIDER analysis the IDR6 region has a higher proportion of proline residues compared to other regions, with fewer polar, aromatic and charged amino acids compared to WT. The IDR6 region itself does not phase separate, and when deleted increases the ability of CBP to phase separate. Proline residues are hydrophobic in nature, and are typically not enriched within IDRs, and are usually regularly dispersed between other residues in proteins that phase separate (Lee et al. 2014; Oldfield et al. 2005; Rani, Baruah and Biswas 2014; Keppel, Howard and Weis 2011; Mittag and Parker 2018; Nair et al. 2019). A high proportion of proline residues, and reduced amount or charged and aromatic residues could account for the inability of IDR6 to phase separate, since these ratios of amino acids do not favour weak interactions.

6.7.3.2 Increased polar and aromatic amino acids could account for IDR7s ability to phase separate

The IDR7 region has a higher proportion of polar amino acids and contains a polyglutamine stretch of 18 amino acids. Glutamine (Q) residues have been shown to be important for phase separation in the context of FUS, contributing to the long range charged interactions (St George-Hyslop et al. 2018; Wang et al. 2018). A high proportion of polar amino acids are thought to drive phase separation, particularly in residues such as serine and tyrosine, as IDR7 has an increased number of these residues, including the polyQ region it is possible that this sequence property is why IDR7 can form phase separated condensates (Boeynaems et al. 2018; Mittag and Parker 2018; Pak et al. 2016; Pessina et al. 2019; Riback et al. 2017; Sabari et al. 2018). It is hard to determine how influential this polyQ region is on the phase behaviour of the IDR7 or the full length protein; previous attempts to generate a polyQ deletion construct was unsuccessful due to the repetative nature of the IDR7 sequence. To investigate this region in the future, a custom synthesised DNA sequence could be designed containing the sequence for the IDR7 region with the polyQ deleted, which could be subsequently cloned into CBP-GFP using a restriction digest approach to see what affect this has on phase separation.

Alongside localCIDER, PLAAC analysis was performed to determine if the C-terminal region can form a prion-like domain, where prions are proteins that can accumulate in different states, similar to phase separation. This is typically linked with proteins that are implicated in neurodegenerative diseases, such as Tau in Alzheimer's disease, or FUS in Amyotrophic Lateral Sclerosis (Alberti and Hyman 2021; Alberti et al. 2009; Piroska et al. 2023; Wang et al. 2018; Martin

et al. 2020; Zbinden et al. 2020; Boyko and Surewicz 2022). This analysis showed that IDR7 is predicted as a large prion-like domain where the IDR6 does not have this ability. Proteins containing prion-like domain have been found to drive phase separation in some cases (Sprunger and Jackrel 2021; Harrison and Shorter 2017), further suggesting that the sequence of these regions is important for its differing phase behaviour.

From analysing the sequence of this C-terminal region the IDR7 region has many sequence factors which favour phase separation, while in contrast the sequence of IDR6 contains elements which are thought to negatively regulate condensate formation, this may explain the difference in phase behaviours observed when deleting these regions in full length CBP.

6.7.4 Colocalisation of endogenous CBP-HaloTag with overexpressed CBP-GFP and mutant constructs

In this chapter I have shown that there is a dominant effect on endogenous CBP phase separation when CBP-GFP and mutant constructs are overexpressed. In the case of $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP this is a dominant negative effect. It was also observed that nuclei with higher expression of overexpressed protein, those that would be too highly expressed to be imaged, have a decrease in signal from the endogenously tagged protein. This could be because the cell is regulating endogenous protein to account for the increased expression of CBP from the overexpression. Alternatively, this could be because endogenously tagged CBP is being sequestered within the overexpressed protein, preventing effecient labelling. To test the efficiency of CBP-HaloTag labelling, in-gel fluorescence could be performed both in the presence and absence of overexpressed protein. In combination with this, protein levels could be checked via western blotting, to identify whether endogenously tagged CBP is being down regulated in the presence of overexpressed protein.

Overall, I have shown that endogenous CBP phase behaviour was influenced by that of the overexpressed protein, in a manner that mirrors the behaviour of the GFP tagged construct. What cannot be determined from colocalisation experiments is if this change in phase behaviour affects CBP activity or dynamics within a cellular context.

6.7.5 CBP phase separation

In the last four chapters, I have shown that CBP can form phase separated condensates at endogenous levels and when overexpressed, with the results being recapitulated when concentration is reduced to approximately that of the endogenous. I have identified regions within CBP which

influence its phase separating behaviour and have characterised this within a cellular context. I have also observed that these phase separating mutants affect the phenotype of endogenously tagged CBP, but it is yet to be understood how this may affect CBP function, including CBPs HAT activity. One way in which I have begun to investigate this is by using single molecule approaches within cells, which have the potential to be studied *in vitro*.

Chapter 7

Investigating CBP phase separation using single molecule approaches

7.1 Introduction

I have shown that CBP can form phase separated condensates and that mutations in CBPs C-terminal region generates differences in CBPs phase separation behaviour, where the C-terminal region is not known to be involved in CBPs HAT activity. While condensates have been shown to affect the function of several proteins, it is not yet clear how phase separation affects the HAT activity of CBP.

The paralog of CBP, p300, has been investigated to understand the implications of phase separation on its HAT activity, but the results remain unclear. Ma *et al.* (2021), suggested that co-condensation of p300 with several transcription factors enhances p300 HAT activity, in contrast to Zhang *el al.* (2021) who implied that condensation of p300 occludes the active site and therefore reduces HAT activity. These two papers highlight two potential theories as to why condensates form with proteins such as p300 and CBP, the first being transcriptional condensates bringing together proteins and nucleic acid to facilitate transcription, which is a well characterised hypothesis in the literature (Ma et al. 2021; Sabari et al. 2018; Henninger et al. 2020; Wei et al. 2020; Basu et al. 2020). The second involves sequestering proteins into condensates that act as a store of proteins when transcription is not needed, an example of this type of condensate is a stress granule, where during times of stress, proteins that are not essential are compartmentalised into condensates (Boeynaems et al. 2023; Freibaum et al. 2020; Van Treeck et al. 2018; McGurk et al. 2018; Molliex et al. 2015; Riback et al. 2017).

I therefore next wanted to investigate how phase separation affects the overall dynamics of CBP and the activity of CBP's HAT domain, providing critical information on the functional relevance of the phenomena of phase separation.

7.1.1 Studying CBP HAT activity within condensates

To understand how condensates affect CBP HAT activity I needed a way to measure CBP HAT activity, typically this is done *in vitro* using HAT assays (Thompson et al. 2004; Bose et al. 2017). For these experiments, I wanted a method to investigate CBP HAT activity within the context of condensates, both inside and outside of cells, which is not possible utilising typical HAT assay format. I decided to approach looking at HAT activity not from a biochemical perspective but by measuring the conformational movements of CBP itself, which occur during HAT activation. The catalytic core of CBP has been well documented, with crystal structures being published of the core of both CBP and p300 (Delvecchio et al. 2013; Zhang et al. 2021; Park et al. 2017). Within the core, the HAT domain is not completely structured, the disordered AIL region has been shown to be dynamic in movement as well as being functionally relevant. When CBPs HAT domain is inactive the AIL sits within the active site, and upon activation the loop is displaced from the active site to facilitate acetylation; I decided to use this movement in AIL to monitor CBP's HAT activity by tracking this conformational change (Bose et al. 2017; Thompson et al. 2004; Zhang et al. 2021).

Two methods were designed to monitor this change in conformation using techniques called single molecule fluorescence resonance energy transfer (smFRET) and transition metal-ion fluorescence resonance energy transfer (tmFRET). Using these FRET based approaches will allow monitoring of changes in intramolecular conformations that occur during HAT activation. Using the smFRET approach, the AIL and a region either within the active site of CBP or within neighbouring domains will be labelled with fluorescent dyes, the distance between these regions can then be monitored using FRET. This distance can be used to determine whether CBP is in an active conformation, with a large distance between the dyes when the AIL is displaced, or an inactive conformation with a short distance between the dyes when the AIL is within the active site. Extreme care must be taken to label the AIL and catalytic core to prevent interfering with CBPs HAT activity. Genetic codon expansion technology will be used to label the AIL with a non-canonical amino acid which is either fluorescent or has the capability to be fluorescently labelled via click chemistry, to cause the least amount of disruption within the active site (Nikić et al. 2015; Nikić et al. 2016; Schmied et al. 2014; Zagotta et al. 2016; Zagotta et al. 2021). Us-

ing this system, intramolecular conformational changes of CBP can be monitored under normal conditions and within the context of the different phase separating mutants, to implicate whether phase separation affects the HAT activity of CBP.

7.1.2 Understanding how condensates affect CBP dynamics

Alongside understanding the effect of phase separation on CBP HAT activity I wanted to determine how the nuclear dynamics of CBP change within the environment of phase separation, which in turn should provide information on the role of phase separation on CBP function. One way to understand protein dynamics within condensates is by tracking single particles of protein in and out of phase separated environments (Gao and Walter 2023; Kent et al. 2020). The benefit of single particle imaging is that you can record information based on dynamics of CBP which would be lost by looking at the bulk population, as the information would be averaged out over the different dynamics observed.

To understand the dynamics of CBP I used the endogenously tagged CBP-HaloTag cell line, labelling the HaloTag with Janelia Fluor dyes to image single particles. Once single particles had been visualised over time, I began to develop a pipeline to track these single particles to determine the natural behaviour of endogenously tagged CBP. I demonstrated in chapter 6 that endogenous CBP could interact with condensates formed by IDR mutants. Once the diffusional behaviour of CBP dynamics has been defined I can overexpress CBP-GFP and the main phase separating mutants to understand if the dynamics of endogenous CBP changes when I express $CBP_{\Delta IDR6}$-GFP which forms larger condensates, or $CBP_{\Delta IDR7}$-GFP which is expressed in a diffuse manner.

Taken together this would provide evidence on the effect of phase separation on CBP movement and function.

7.1.3 Chapter aims

In this chapter initial optimisation and control experiments are described to investigate how phase separation affects the activity and dynamics of CBP. This includes establishing genetic code expansion (GCE) experiments to produce mutated CBP-HaloTag for both smFRET and tmFRET experiments, as well as optimisation of labelling and tracking of endogenously tagged cell line CBP-HaloTag to understand the dynamics of endogenous CBP.

The data described within this chapter is only preliminary, and is provided to give an insight into future work. As such, it would be premature to quantify the data within this chapter as the experiments described are still in optimisation.

7.2 SmFRET

Firstly, I wanted to design and set up an experiment to investigate CBP HAT activity by monitoring conformational changes in the HAT domain, to do that I wanted to establish smFRET.

7.2.1 Introduction to the smFRET approach

smFRET is an incredibly sensitive approach, which will allow the determination of intramolecular movements of CBP molecules, which will indicate CBP HAT activity. Thus providing a method to determine whether CBP activity is changed compared to normal conditions when inside different phase separated states, including phase separation mutants which form larger condensates or those that are expressed more diffusely.

To do this CBP will be labelled in multiple places with two different fluorophores, one which will act as the donor and the other an acceptor (Figure 7.1). The donor fluorophore will be excited with laser, the energy from which will be transferred to the acceptor molecule through a non-radiative process that leads to acceptor excitation, if they are in close proximity (Hochreiter, Garcia and Schmid 2015). The photons detected are used to calculate the FRET efficiency, otherwise known as E. Using the FRET efficiency the distance between the two fluorophores can be deteremined to gain insight into the conformation of CBP. A broader introduction to smFRET can be found in subsection 1.2.2.2.

In order to monitor the conformational dynamics of the CBP HAT domain I am going to label the AIL loop and input another tag either side of the CBP core. The AIL was chosen due to the displacement which occurs upon HAT activation (Bose et al. 2017; Thompson et al. 2004). By labelling the AIL and another region in CBP the change in distance which occurs when the AIL changes conformation during HAT activity should be detected by smFRET (Figure 7.1). Alternatively, dyes could be incorporated into the RING or bromodomain of the CBP core, both of which are thought to undergo approximately 20 - 30 Å movements during CBP activation.

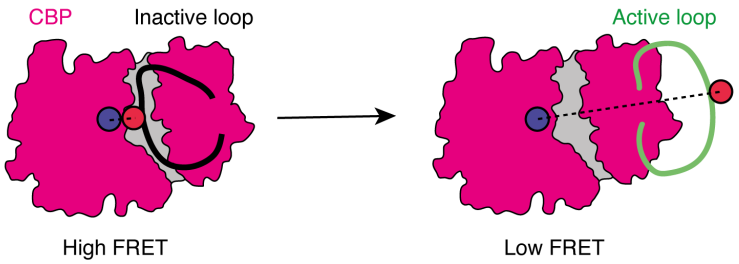

Figure 7.1: SmFRET of CBP to look at HAT activity
Diagram of CBP (pink) in both active and inactive forms. When CBP is active the AIL (black) is present within the active site of CBP (grey) the donor and acceptor fluorophores are close together which will give a high FRET signal. When CBP is active the AIL (green) is dislodged from the active site (grey) which increases the distance between donor and acceptor and would produce a low FRET signal. Here the diagram of CBP represents the full length protein, and is based off a CBP HAT domain diagram made by Dr Dan Bose, based on the HAT domain structure published in Wang et al. (2008a).

Extreme care must be taken to label the AIL region as any bulky dyes or fluorophores could disrupt the natural position of the AIL within the active site in its inactive form, and prevent success of this experiment. To label the AIL I have chosen to use the amber codon (TAG) suppression system, where an unnatural amino acid (uaa) replaces either a glutamine or an asparagine residue in the AIL, both of which are bulky amino acid side chains, therefore the insertion of the uaa should cause the least amount of disruption to the HAT domain (Schmied et al. 2014; Nikić et al. 2015). The amber codon refers to the nucelotide sequence TAG, which encodes for a stop codon; this stop codon is less commonly used than the alternatives and can therefore be used for genetic codon expansion (GCE) experiments (Schmied et al. 2014; Nikić et al. 2015). Using a tRNA and aminoacyl tRNA synthetase pair, which is orthogonal to the cells machinery, an uaa can be incorporated into the protein sequence, at the site of the amber codon (TAG) (Schmied et al. 2014; Nikić et al. 2015). I can therefore use this technique to incorporate either a fluorescent amino acid or an amino acid which can be labelled with a fluorescent dye, for FRET experiments.

A total of 4 constructs were generated which can be used for both single molecules FRET (smFRET) and transition metal-ion FRET (tmFRET) (which will be discussed in later sections) that have the nucleotide sequence TAG in two different positions, this is where the uaa will be incorporated. For smFRET the uaa is {[(E)-cyclooct-2-en-1-yl]oxy}carbonyl)-L-lysine (TCO*A) which can be labelled with a fluorophore using click chemistry for FRET experiments (Nikić et al.

2016). In order to label CBP with a secondary fluorophore for smFRET a SNAP tag was inserted either side of the core of CBP, which can also be labelled by click chemistry (Figure 7.2 A) (Nikić et al. 2015; Nikić et al. 2016). AlphaFold predictions were run of both constructs, by Dr Dan Bose, which predicted no effect on structure of the core domain compared to WT (Figure 7.2 B and C) (Jumper et al. 2021).

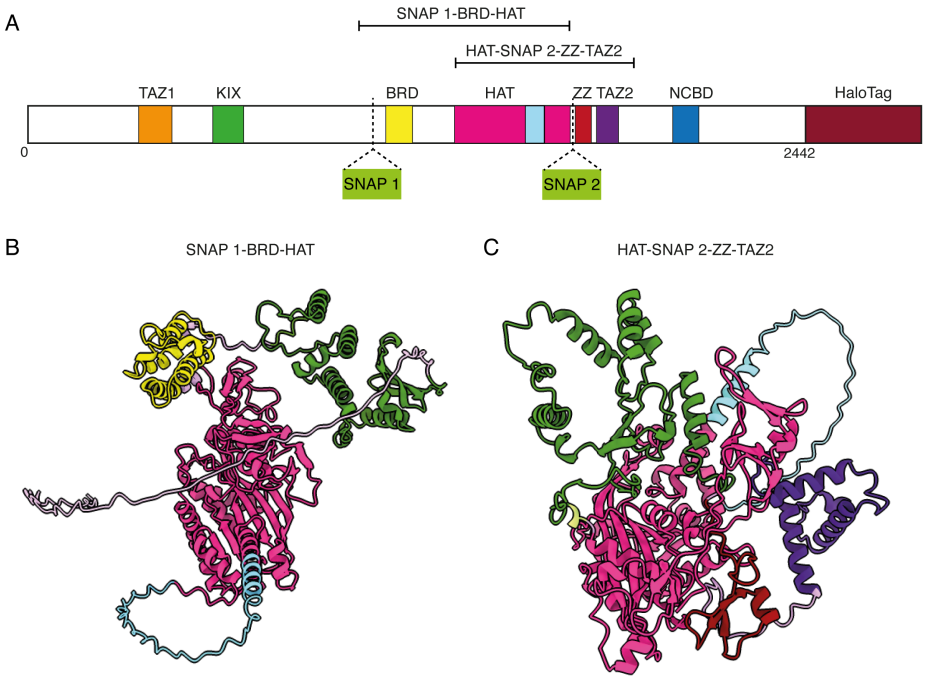

Figure 7.2: Insertion of the SNAP tag in CBP for smFRET to study HAT activity
A - Domain diagram of CBP showing globular structured domains as have been described previously. The autoinhibitory loop (AIL) is shown in light blue. For smFRET there will be an unnatural fluorescent amino acid TCO*A (orange star) mutation as well as an inserted SNAP tag either upstream or downstream of the HAT domain. The dotted lines represent the positioning of the SNAP tag (lime green) in construct 1 (SNAP 1) and construct 2 (SNAP 2). B - Structure prediction by AlphaFold of SNAP 1 - BRD-HAT or HAT - SNAP 2 - ZZ - TAZ2, to indicate the placement of the SNAP tag and the influence on structural regions Jumper et al. (2021).

These constructs have a C-terminal HaloTag and Flag tag which can be used for visualisation and purification of protein respectively. Once I have a construct which has two fluorescent labels I can begin to measure smFRET, both in cells and *in vitro,* to define the conformational changes that

occur within CBPs HAT domain under normal circumstances (an example of the labelled core and potential dynamics is shown in Figure 7.3). Since it has already been observed that endogenous CBP-HaloTag can interact with the GFP tagged constructs, I can then alter the conditions to include the previously characterised constructs of CBP-GFP, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP to see if the activity of CBPs HAT domain changes with and without a phase separating environment.

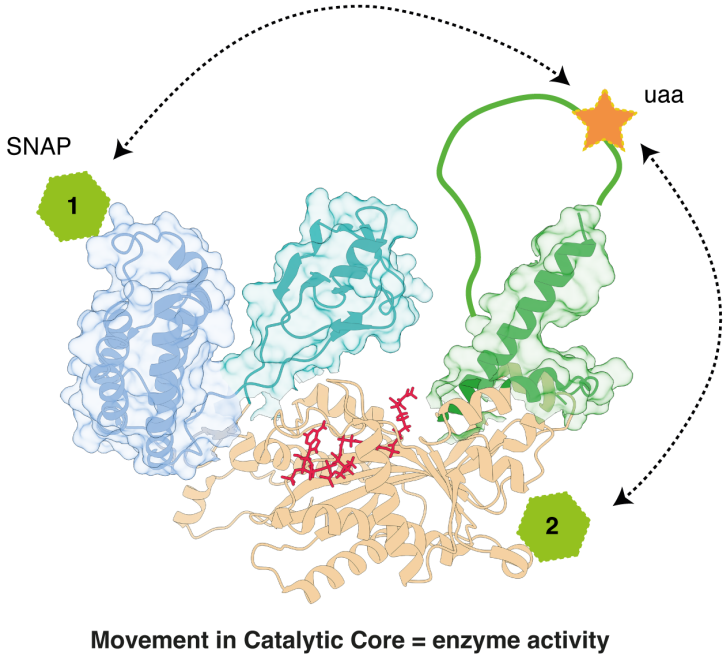

Figure 7.3: Structural reorganisation of CBP core during HAT activation
Structure prediction of the core of CBP HAT, where the bromodomain (BRD) domain is shown in blue and the autoinhibitory loop (AIL) is shown in green, structure from Delvecchio et al. (2013). Green hexagons represent SNAP tag 1 and 2, in their respective constructs, and the orange star represents fluorescent amino acid TCO*A. Positions of the SNAP tag shown in this diagram are for purposes of demonstrating smFRET, these locations are not exactly where the tag will reside. SmFRET will be calculated between one of the SNAP tags and TCO*A as demonstrated by the dotted lines.

7.2.2 Setting up the smFRET system using a GFP control plasmid

To set up this system I first needed to label the AIL region using the amber suppression system (Nikić et al. 2015; Koehler et al. 2020). Mutated CBP-HaloTag constructs were generated that contains a TAG codon where the uaa was going to be incorporated, as well as the SNAP tag for smFRET. Typically the TAG would code for a stop codon, terminating protein production, however this does not happen in this system. HEK293FT cells were transfected with the plasmid of interest containing the TAG stop codon, in the position I want to label, and the orthogonal tRNA/tRNA synthetase pair, as well as the addition of the uaa, in this case TCO*A. Upon transfection with the tRNA/tRNA synthetase pair and addition of TCO*A, the tRNA synthetase will charge the tRNA with the TCO*A. The ribosome will then read the mutant plasmid sequence through the TAG, incorporating the TCO*A into the protein sequence (Figure 7.4).

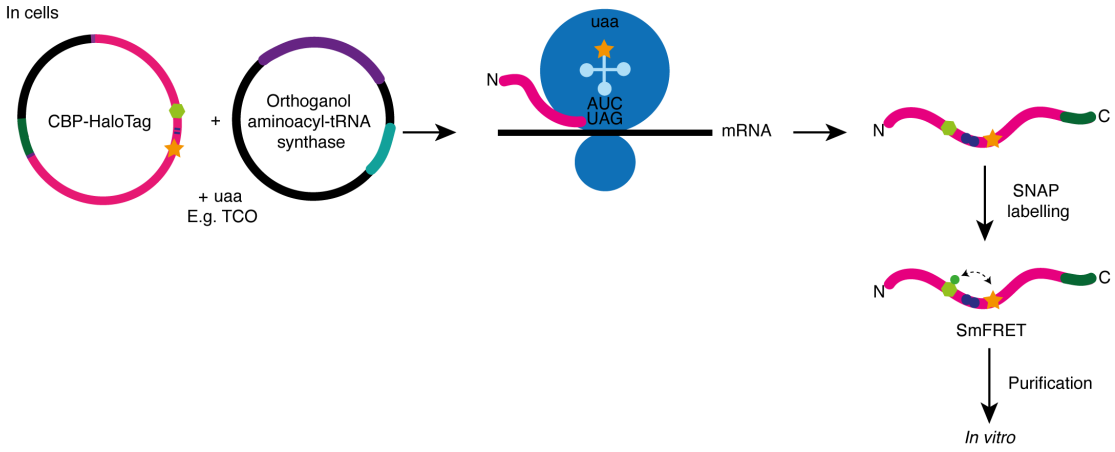

Figure 7.4: Diagram of experimental procedure of smFRET
CBP-HaloTag construct containing appropriate smFRET mutations (orange star represents TAG mutation) is transfected into HEK293FT cells with orthogonal tRNA and aminoacyl-tRNA synthetase along with unnatural amino acid (uaa) TCO*A. The ribosome (blue) reads through the CBP-HaloTag protein sequence and upon reaching the TAG, incorporates TCO*A via the tRNA which was charged with the uaa by the tRNA synthetase. The remainder of the protein is translated producing a full length protein product within the cell, containing a SNAP tag (green hexagon) and TCO*A (orange star). This can be used for smFRET once labelled with fluorophores both inside cells, or can be purified to perform smFRET experiments *in vitro*.

I therefore generate a protein that contains the uaa which can be labelled for smFRET, and also the HaloTag which can be labelled to check for protein production. Only if the TAG is

suppressed will protein be made where the HaloTag can be successfully labelled, otherwise the protein would be truncated and would not contain the HaloTag (Nikić et al. 2016). This HaloTag construct also contains a Flag-6His-TEV cleavage site which allows for purification from the cell lysate which can be used downstream to purify protein for *in vitro* smFRET experiments.

Before working directly with CBP-HaloTag I wanted to check whether the system worked, to do this I used a GFP positive control plasmid which was a kind gift from Edward Lemke and used in (Nikić et al. 2016). This GFP control plasmid is a near infrared fluorescent protein (iRFP) fused to a GFP tag that has the TAG stop codon present at tyrosine 39 (Nikić et al. 2016). If the system is working correctly, the TAG will be suppressed and GFP expression will be observed, however if no difference is observed compared to the background the system is not working as expected.

HEK293FT cells were transfected with the GFP positive control plasmid, the tRNA/tRNA synthetase pair, termed NES-PyIRS-AF as well as eRF1 E55D construct. This final construct eRF1 E55D, was a kind gift from Jason Chin, and is a mutated form of eukaryotic release factor 1 (eRF1) (Schmied et al. 2014). In its natural form eRF1 terminates transcription at the stop codon TAG, however in its mutated form it increases the efficiency of the amber codon system by allowing readthrough of the TAG stop codon without influencing the remaining stop codons (Schmied et al. 2014).

Upon addition of TCO*A an increase in GFP expression in HEK293FT cells was observed when transfected with the positive control construct, which indicates that the system was working as expected (Figure 7.5). In the absence of TCO*A, some background GFP expression was observed, which hasn't been contrast matched to the positive control sample treated with TCO*A (Figure 7.5, -TCO*A raw). This was to be expected as translation is error prone, some GFP will be expressed in the absence of TCO*A being present. The increase in expression that was observed was very promising, and since the system appeared to be working as I would expect, I decided to test the system with one of the CBP-HaloTag constructs, CBP_{HHE*D}-HaloTag-Flag (CBP-HaloTag-Flag which contains the following amino acids H1547, H1551, E1555, *1579, D1665, where * is the TAG amber codon).

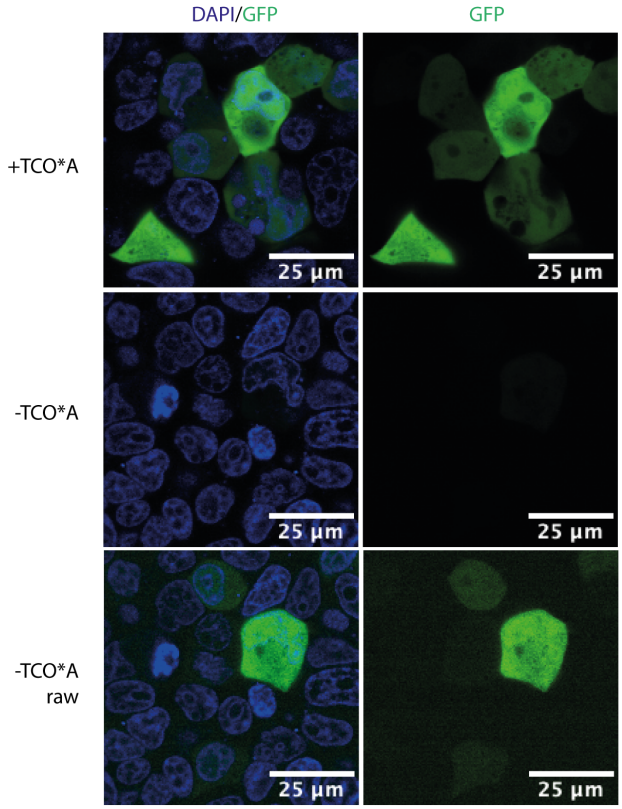

Figure 7.5: Testing TCO*A system using GFP positive control plasmid
Representative images of HEK293FT cells transfected with GFP control plasmid and orthogonal aminoacyl-tRNA synthetase with and without TCO*A. GFP expression is shown in green, with nuclei shown by staining with Hoechst. Contrast was matched for -TCO*A to compare to + TCO*A, -TCO*A raw refers to the image for -TCO*A at maximum exposure. Images were processed in ImageJ/Fiji, scale bar represents 25 µm. Biological replicates = 1.

7.2.3 Testing CBP-HaloTag

Since the system worked when using the GFP positive control construct, I decided to test the system using CBP_{HHE*D}-HaloTag-Flag. The experiment was performed as described for GFP, except that as the HaloTag isn't intrinsically fluorescent I labelled the cells with JF 549 to visualise protein expression. Using this system, bright nuclear expression of CBP-HaloTag was observed, which forms nuclear puncta in the presence of TCO*A (Figure 7.6); suggesting that the suppression of the TAG is working as expected. Similarly to the GFP positive control construct, in the

absence of TCO*A some expression of protein was observed, as can be seen when the contrast was not matched to the sample with TCO*A (Figure 7.6, -TCO*A raw). In the presence of TCO*A nuclear puncta was observed, indicating that CBP was behaving as I would expect phenotypically (Figure 7.6). As there was a clear difference in expression between our + TCO*A and - TCO*A samples, this suggested that the amber suppression system could be used to incorporate the uaa for labelling for smFRET.

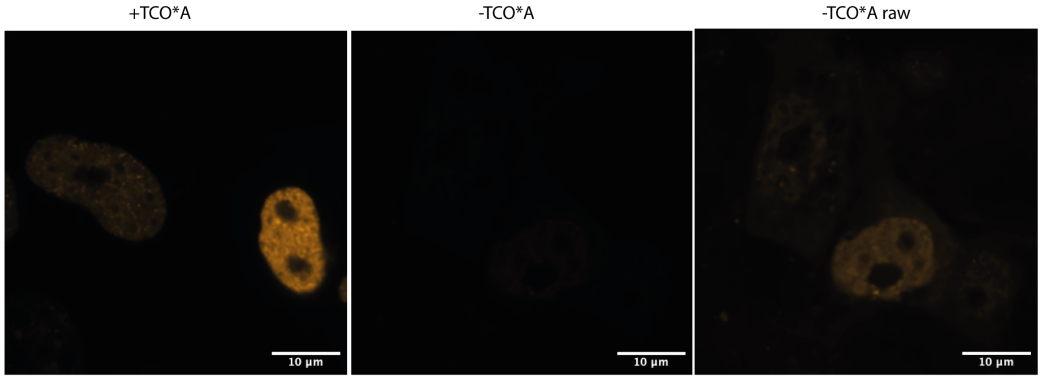

Figure 7.6: CBP-HaloTag expression using TCO*A system
Representative images of HEK293FT cells transfected with CBP_{HHE*D}-HaloTag-Flag plasmid and orthogonal aminoacyl-tRNA synthetase with and without TCO*A, HaloTag was labelled with Janelia Fluor 549 (orange). Contrast was matched for -TCO*A to compare to + TCO*A, -TCO*A raw refers to the image for -TCO*A at maximum exposure. Images were preprocessed using Nikon package denoise followed by further processing in ImageJ/Fiji, scale bar represents 10 µm. Biological replicates = 1.

Due to time constraints no other advances were made on optimising this experiment. Further work on this system is needed to produce single molecule levels of the CBP-HaloTag-Flag constructs, as well as optimisation of labelling for smFRET and purification of protein for *in vitro* work.

7.3 TmFRET

Alongside the smFRET approach to studying CBP HAT activity, I also wanted to apply a variation of the technique called transition metal-ion FRET (TmFRET).

7.3.1 Introduction to the tmFRET approach

Using smFRET a distance between 30 - 80 Å can be accurately measured, which was appropriate for the large domain conformational changes that I was looking at between the AIL and SNAP tag positions in the previous section (Zagotta et al. 2021; Roy, Hohng and Ha 2008). However this fails to account for smaller changes in conformation that are over shorter distances, less than 30 Å, which can be measured by tmFRET (Mortensen and Loland 2020). In tmFRET, instead of using a pair of fluorophores as the donor and acceptor, this uses a fluorescent amino acid as the donor and a divalent metal cation as a quencher; the cation such as nickel, copper or cobalt acts as a non-fluorescent acceptor, quenching the fluorescent signal in a distance dependant manner (Latt, Auld and Vallee 1972; S. A. Latt and Vallee 1970; Richmond et al. 2000; Sandtner, Bezanilla and Correa 2007; Dai et al. 2019; Dai et al. 2021).

This system utilises a fluorescent amino acid L-Acridonylalanine (ACD), to act as the donor fluorophore, and the metal cation is coordinated by a pair of di-histidine mutants, inserted at positions i + and i + 4 in an α-helix, which are placed within close proximity to this region. In total 4 constructs were made, which contained 2 different TAG mutation sites and 2 different pairs of di-histdine mutants; the amino acid substitutions can be found in the materials and methods in Table 2.8. In this experiment the ACD fluorescent donor is placed within the AIL, and the di-histidine mutation is present within the active site of the HAT domain. When in an inactive conformation the AIL is within the active site, and is therefore close to the metal cation which will quench the fluorescence of the amino acid (Figure 7.7). When CBPs HAT domain is active, the AIL is displaced from the active site, increasing the distance between the metal cation acceptor and the donor fluorophore, allowing the amino acid to fluoresce (Figure 7.7). Using this system, changing only 3 amino acids, HAT activity could be monitored through energy transfer. Once the system is set up I will be able to monitor CBPs behaviour under normal conditions, and in different phase separation environments.

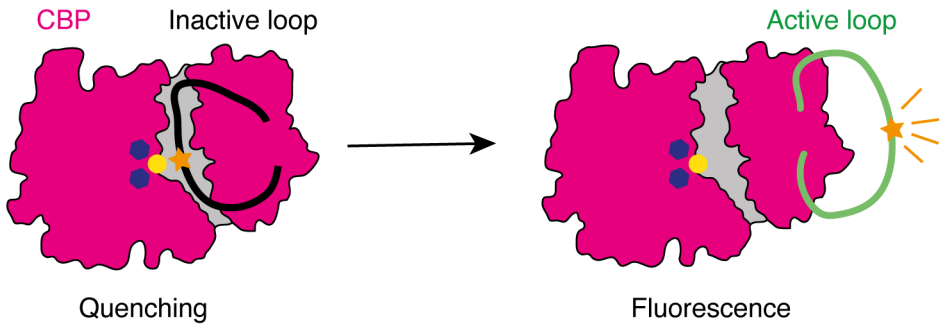

Figure 7.7: Illustration of tmFRET to study CBP HAT activity
Diagram of CBP (pink) in both active and inactive forms. Mutations in the CBP construct include TAG site for unnatural amino acid insertion (orange star), in this case ACD, and di-histidine mutations (purple), yellow circle represents a cation that is coordinated by the di-histidine mutant which facilitates quenching. In the inactive form the AIL (black) is positioned within the active site of CBP (grey) where the fluorophore ACD is in close proximity to the cation, when excited the electrons will be transferred to the cation, therefore quenching will be observed. When CBP in an active form the AIL (green) is dislodged from the active site (grey) which increases the distance between the ACD and cation, preventing quenching and allowing fluorescent emission from the ACD. Here the diagram of CBP represents the full length protein, and is based off a CBP HAT domain diagram made by Dr Dan Bose, based on the HAT domain structure published in Wang et al. (2008a).

7.3.2 Setting up the ACD system using the GFP control plasmid

For the tmFRET experiment I used the same amber codon suppression system as was used for smFRET, and was described in section 7.2.2. In this case, HEK293FT cells were transfected with the GFP positive control plasmid, the tRNA/tRNA synthetase pair (82RS mutant) as well as eRF1 E55D (Zagotta et al. 2021; Schmied et al. 2014). A slight variation in the protocol was performed as to when the amino acid ACD was added and removed, following methods in Zagotta *et al.* (2021), I added ACD 4 hours after transfection and left it in the media until imaging. Similarly to smFRET, the protein of interest is translated by host cell machinery, which incorporates the ACD amino acid at the given TAG codon, to produce a full length protein that can be used for tmFRET experiments (Figure 7.8).

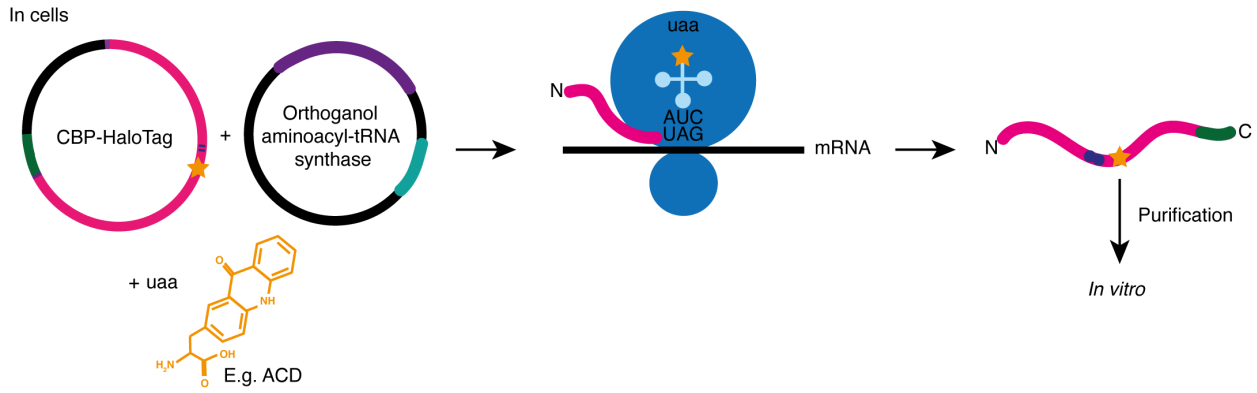

Figure 7.8: Experimental approach to tmFRET to study CBP HAT activity
CBP-HaloTag construct containing appropriate tmFRET mutations (purple stripes represent di-histidine mutation, orange star represents TAG mutation) is transfected into HEK293FT cells with orthogonal aminoacyl-tRNA synthetase along with unnatural amino acid (uaa), in this case ACD. The ribosome (blue) reads through the CBP-HaloTag sequence until the TAG sequence, at which point it incorporates ACD which has been charged to the transfected tRNA by the tRNA synthetase. The remainder of the protein is translated producing full length protein product within the cell, containing the di-histidine mutation (purple) and ACD (orange star). This can be used for tmFRET inside cells, or can be purified to perform tmFRET experiments *in vitro*.

To initially validate the system I used the GFP positive control plasmid that was previously used for smFRET. If expression of GFP was observed, the suppression system was working with the tRNA/tRNA synthetase pair and ACD amino acid. When HEK293FT cells were transfected with the GFP positive control plasmid, GFP expression was observed in cells transfected both with and without ACD, indicating that in the presence of ACD there was no more GFP expression than can be expected at a background level (Figure 7.9). No background GFP expression was observed in untransfected HEK293FT cells, which removed any concerns over background GFP expression from autofluorescence (Figure 7.9). Since no increase in expression of GFP in the presence of the amino acid was observed, it was clear that optimisation of this system was needed for this amino acid and tRNA/tRNA synthetase pair.

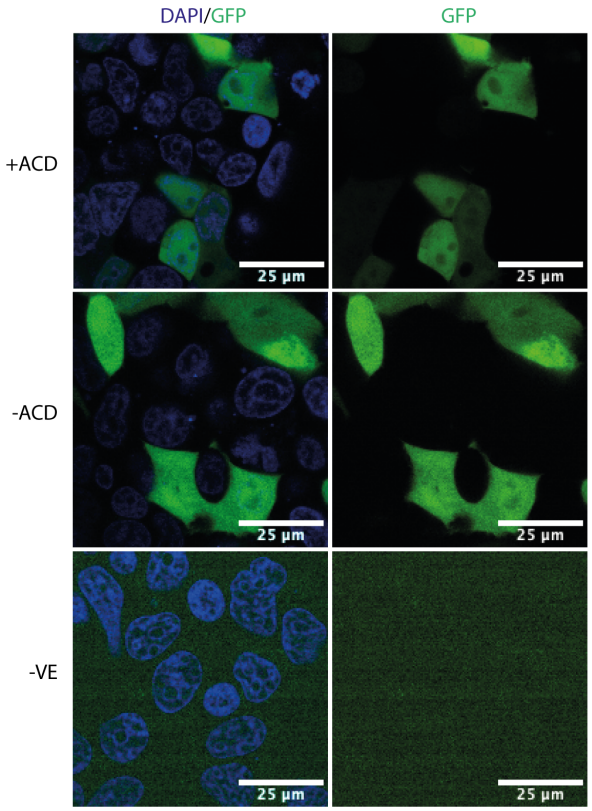

Figure 7.9: Testing ACD amino acid system with GFP positive control plasmid
Representative images of HEK293FT cells transfected with GFP control plasmid and orthogonal aminoacyl-tRNA synthetase with and without ACD. GFP expression is shown in green, with nuclei shown by staining with Hoechst in blue. Contrast was matched for -ACD to compare to + ACD. Negative control was untransfected HEK293FT cells to test for autofluorescence. Images were processed in ImageJ/Fiji, scale bar represents 25 µm. Biological replicates = 1.

In an attempt to optimise this system for the ACD, I altered the experimental approach to one which aligned to the smFRET protocol, where the ACD was added immediately after transfection and removed after 8 - 10 hours in culture. No difference between GFP expression was observed in the HEK293FT cells that were incubated with ACD compared to those without (Figure 7.10). This suggested that the system needed further optimisation to increase the production of protein that contains the amino acid of interest.

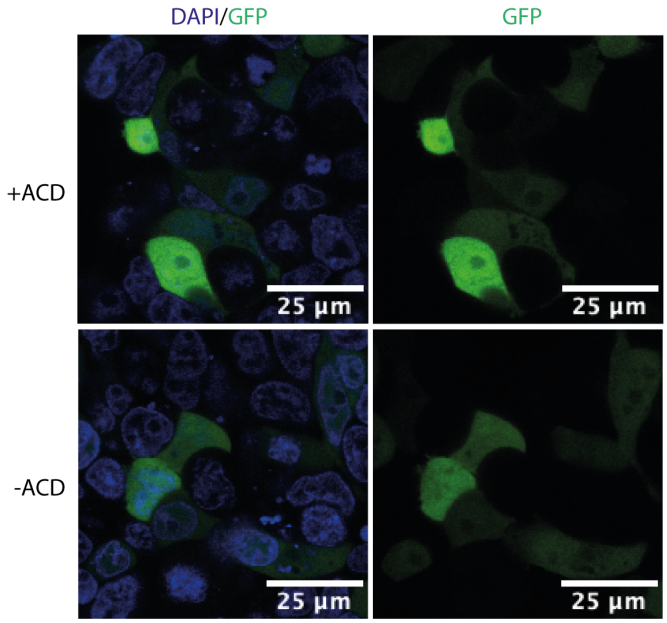

Figure 7.10: Optimisation of ACD tmFRET system with GFP positive control plasmid Representative images of HEK293FT cells transfected with GFP control plasmid and orthogonal aminoacyl-tRNA synthetase with and without ACD, with optimised ACD addition protocol. GFP expression is shown in green, with nuclei shown by staining with Hoechst in blue. Contrast was matched for -ACD to compare to + ACD. Images were processed in ImageJ/Fiji, scale bar represents 25 µm. Biological replicates = 1.

7.4 Single particle imaging of endogenous CBP-HaloTag

7.4.1 Determining an approach to study single particle dynamics

Alongside studying the conformational dynamics of CBP at single molecule levels to determine HAT activity within a condensate, I wanted to look at the nuclear dynamics of CBP and how this changes within a condensate. Using the endogenously tagged CBP-HaloTag cell line, I wanted to image single particles of CBP to monitor the dynamics of the protein on its own, to see whether the protein is fully mobile or confined within a particular region (Figure 7.11 A). Once I have calculated the usual proportion of CBP that is either mobile or confined, I want to determine if within different phase behaviours such as in the presence of overexpressed constructs CBP-GFP, $CBP_{\Delta IDR6}$-GFP, $CBP\Delta_{IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP (Figure 7.11 B), the proportion of mobile

or confined endogenous CBP changes.

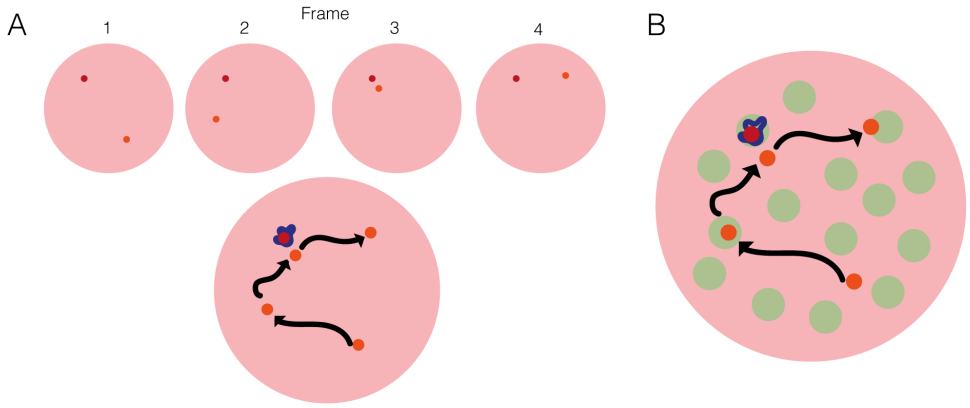

Figure 7.11: Studying single particles of CBP-HaloTag within the context of condensates Diagram demonstrating the aim of single particle tracking of two CBP-HaloTag puncta. A - One puncta is very mobile (orange), with the movement being tracked across a timecourse, as indicated by the black arrows. The other CBP puncta is confined (red) with very little movement occuring, as indicated by the navy trajectory. B - overall aim of this work is to perform single molecule experiments within the context of GFP tagged phase separation mutants, an example here is represented by green circles.

This experimental procedure involves labelling single particles of CBP, tracking the particles over a timecourse and analysing the data to determine movements of the particles. This was an iterative process which involved optimising labelling of single particles of CBP, the imaging parameters used, and testing analysis parameters to track single particles before starting the cycle over (Figure 7.12). The first step involved optimising the labelling for single particles, for example choosing compatible dyes, optimising the concentration of dye for single particle work and practising dual labelling so that I can look at the bulk population of CBP as well as track single particles (Figure 7.12). After optimisation of labelling I had to optimise the imaging for single particles - this involved optimising the exposure time and laser power so that I could see the single particles clearly, while imaging fast enough that we weren't biassing for particles which were stationary. I also had to optimise for the length of time I was imaging, so that enough particles were captured without over bleaching the sample (Figure 7.12). Subsequently, I could then optimise image analysis - processing the images to reduce the signal to noise ratio, which allowed for clearer tracking of the particles to monitor the movement of the single particles (Figure 7.12).

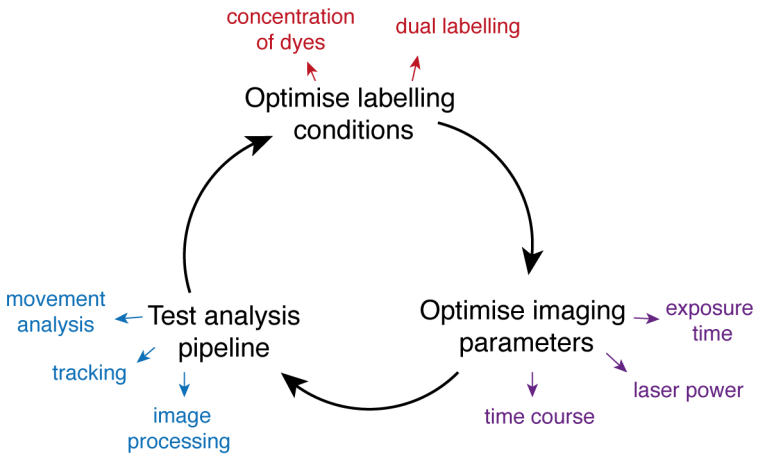

Figure 7.12: Optimisation strategy for single particle tracking
Schematic to indicate the cycle for optimising single particle tracking of endogenously labelled CBP featuring three main steps: optimising the labelling of CBP-HaloTag cells, optimising the imaging parameters and testing the analysis pipeline to track single molecules. Examples of what needs to be optimised for each subheading is indicated below each step with each step having its own colour: labelling (red), imaging (purple) and analysis (blue).

7.4.2 Sparse labelling of CBP-HaloTag

To label single molecules of CBP, I determined two strategies: the first utilised a photoactivatable ligand which could be excited with laser to visualise only a few single molecules per nuclei, which would allow for fewer optimisation steps for the labelling itself (Grimm et al. 2016). The second approach involved reducing the concentration of ligand so that only a small population of CBP molecules were labelled per cell, this required more experimental tests to determine the concentration to be used. An alternative strategy would be label the whole population and photobleach the sample to the point that only a few labelled single particles remained, while a valid approach this was not tested during this doctoral work.

7.4.2.1 Using a photoactivatable ligand

In order to visualise single particles, I first tested the photoactivatable ligand as this would eliminate the need to optimise labelling concentrations of the ligand. Polyclonal CBP-HaloTag cells were labelled with JF 549 to observe which nuclei contained CBP-HaloTag, and photoactivatable

HaloTag ligand 646 to observe single particles. In this case blue laser light was targeted to ROIs to stimulate activation of the photoactivatable ligand. Several ROIs were targeted both within the nuclei of CBP-HaloTag positive cells and randomly placed within the cellular environment to test for off-target effects (Figure 7.13). Upon blue light stimulation fluorescence was observed, but not within the nucleus (Figure 7.13). This indicates nonspecific binding and activation of the photoactivatable ligand (Figure 7.13). This ligand was very lipophilic and therefore it is possible that this ligand preferentially bound to the membrane causing unspecific labelling when activated (Dr Alison Twelvetrees personal communication). Since no activation was observed of particles within the nuclear volume, I decided not to proceed with this method of labelling.

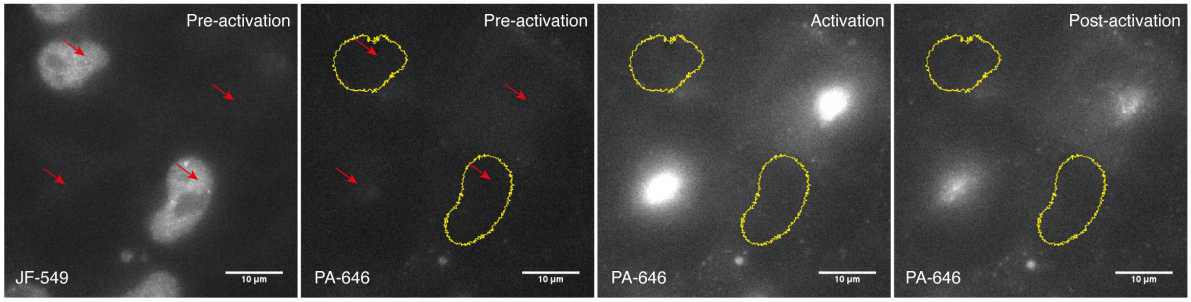

Figure 7.13: Photoactivatable Janelia Fluor is not targetable to the nucleus to study single molecule CBP-HaloTag
Representative images of CBP-HaloTag polyclonal cell lines imaged in bulk with Janelia Fluor 549 (JF-549) and with a photoactivatable ligand at 646 (PA-646). Four regions of interest were generated to target the laser power of the 405 nm laser as indicated by the red arrows, two regions targeted to puncta within the nucleus (indicated by yellow outline) and two regions targeted unspecific regions to test for nonspecfic activation. Images were taken pre-activation in both channels 549 and 646 followed by images in the 646 during photoactivation and post-activation, where the post-activation image was taken approximately 4.6 seconds after activation. Images were processed in ImageJ/Fiji, scale bars represent 10 µm. Biological replicates = 1.

The alternative option was to reduce the concentration of the Janelia Fluor label so that only a few single particles could be visualised at one time. To do this I would label the cells with JF 549 to find the nuclei that had CBP-HaloTag within the polyclonal population, and also label for single particles with a lower concentration of JF 646. Before attempting this labelling procedure I wanted to determine if both of these ligands could be used without getting bleed-through between the two channels.

7.4.2.2 Checking bleed-through of double labelled CBP-HaloTag cells

For the initial set-up I wanted to check whether I could image in the 549 channel and 646 channel without getting any signal from either of the Janelia Fluors in the opposing channel, known as bleed-through. To test this, I plated polyclonal CBP-HaloTag cells in 8-chamber coverslips and labelled for either JF 549 or JF 646 and Hoechst to visualise the nuclei. With both channels labelled separately, representative images were taken in 4 different channels using lasers 405 nm, 488 nm, 532 nm and 637 nm laser lines. For the JF 549 signal was observed for the nuclei using the 405 nm laser and nuclear CBP puncta using the 532 nm laser, as would be expected (Figure 7.14). Both other channels only show background fluorescence that was not indicative of bleed-through, as it did not correspond to the signal observed in the 532 channel (Figure 7.14). Similarly, for JF 646 nuclear stain was observed using the 405 nm laser, and CBP puncta using the 637 nm laser but no specific signal in either of the two other channels (Figure 7.14). Since no bleed-through was observed using the Janelia Fluor ligands on endogenous CBP-HaloTag I decided to proceed with optimising labelling for single particle imaging.

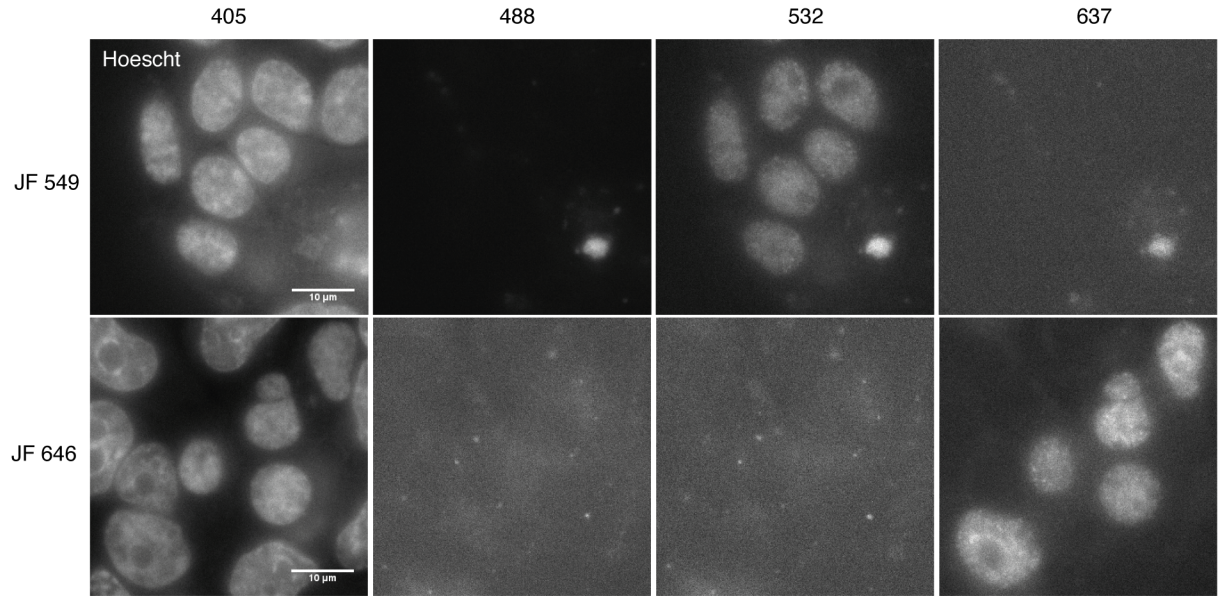

Figure 7.14: 4 colour images to investigate bleedthrough of signal between the channels
Representative images of HEK293T-CBP-HaloTag cell line labelled with Hoechst and either Janelia Fluor 549 (JF 549) or Janelia Fluor 646 (JF 646). Images were processed ImageJ/Fiji, scale bar represents 10 µm. Biological replicates = 1.

7.4.2.3 Determining a concentration of Janelia Fluor that labels single particles

In future experiments, I want to overexpress GFP tagged constructs, and monitor the dynamics of single particles of CBP-HaloTag, therefore the greater distance between the wavelengths of the fluorophores, the less chance of interference from each of the respective channels with the other, therefore I chose to use JF 646 for single particle imaging. In this experiment, I used polyclonal CBP-HaloTag cells therefore I needed to label with JF 549 to label the bulk population of CBP-HaloTag to determine which nuclei contained the tagged CBP, and also label with JF 646 to observe single particles. In order to optimise the concentration of JF 646 for single particle imaging I tested a range of different concentrations from 0.5 nM - 5 nM.

Polyclonal CBP-HaloTag HEK293T cells were plated in 8 - chambered cover glasses and grown for 48 hours before labelling with 50 nM of JF 549 and differing concentrations of JF 646. The labelled cells were imaged for bulk CBP-HaloTag labelling and a timecourse was taken of JF 646 to capture single molecules over time, where images were taken with a 100 ms exposure and captured approximately every 140 ms. The first concentration tested was 0.5 nM of JF 646, this was based on previous single molecules experiments in Dr Alison Twelvetrees Lab.

The timecourse was processed in ImageJ/Fiji where ROIs were generated of the nuclei from the 549 channel, and applied to the 646 channel to identify the nuclear regions that contain CBP-HaloTag; this separated single particles of CBP in the nucleus (red arrows) compared to any background staining that could be observed (blue arrows) (Figure 7.15) (Schindelin et al. 2012). The results showed bulk labelling of CBP-HaloTag using JF 549, and some single particles labelled with JF 646 that move in and out of the imaging plane (Figure 7.15 A). Some non-nuclear background labelling was observed. Unlike single particles of CBP, these stay stationary throughout the timecourse (Figure 7.15 A). At a concentration of 0.5 nM, JF 646 labelled only a few molecules within the 5 nuclei that were imaged, where many nuclei showed little to no labelling, this concentration was therefore too low to use for single particle tracking of CBP, therefore I decided to try a higher concentration (Figure 7.15 A).

The next concentration I tried was 5 nM of JF 646, this increased the number of single molecules observed as well as increasing the background labelling outside of the nucleus (Figure 7.15 B). Single particles of CBP were observed over the timecourse, where some move rapidly in the imaging plane and others stay stationary (Figure 7.15 B). Approximately 5 - 15 single molecules were identified per nucleus, this was optimal for single molecule tracking as I wanted enough molecules to track, while having few enough so that the single molecules can be distinguished from each other (Figure 7.15 B). I therefore decided to use 5 nM of JF 646 to capture videos of

CBP-HaloTag to optimise image analysis and tracking of single molecules.

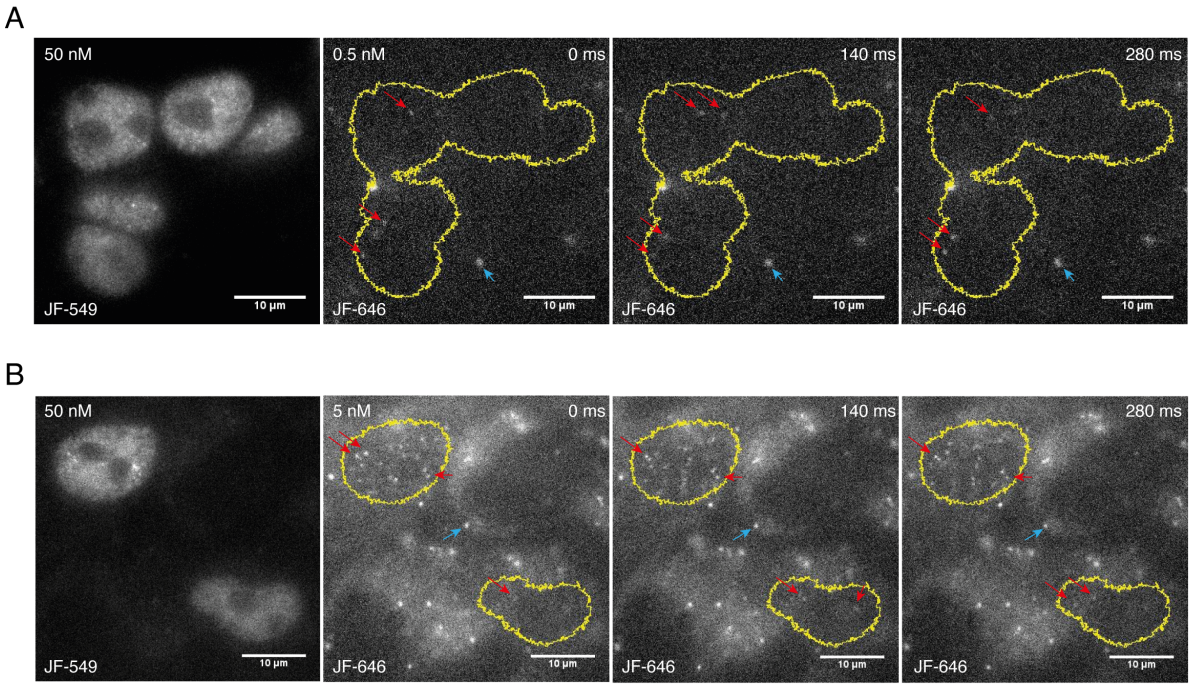

Figure 7.15: Testing concentrations of Janelia Fluor 646 to label single molecules of CBP-HaloTag Representative images of endogenously tagged CBP-HaloTag that is bulk labelled with 50 nM Janelia Fluor 549 and labelled for single molecules using Janelia Fluor 646 at concentrations of 0.5 nM (A) and 5 nM (B). Three images were taken from each timecourse, where the images are consecutive and time between images is shown in the top right corner in milliseconds. The outline of the nuclei from the 549 image was generated in ImageJ/Fiji and applied to the time-course (yellow outline) to distinguish nuclear CBP single molecules (red arrows) compared to background staining (blue arrows). Images were processed in ImageJ/Fiji, scale bar represents 10 µm. Biological replicates = 1.

7.4.3 Generating a clonal CBP-HaloTag cell line for single particles imaging

Up until this point single particle optimization was performed using the polyclonal CBP-HaloTag HEK293T cell line. This is suboptimal for future experiments where single particles will be tracked within the context of overexpressed GFP tagged constructs, as transfection is not 100 % efficient, therefore the chance of transfecting a cell that also contains CBP-HaloTag is significantly reduced in a polyclonal cell line. Therefore I decided to generate a clonal cell line which would be beneficial

for future experiments, and remove the need for bulk labelling with JF 549.

Clonal cell lines were generated of CBP-HaloTag by single cell sorting the polyclonal cell line using FACS (this was done with the help of Dr Nicola Carruthers). The cells were labelled with JF 549, resuspended and sorted into single cells based on fluorescence. Single cells were sorted into individual wells in a 96-well plate and left to expand over a 2 week period before the cultures were grown up and screened. When clones were expanded into a 12-well dish they were also plated into an 8-chamber coverglass and stained with TMR for screening. Two separate rounds of FACS and clonal screening was performed, from which one clone showed positive for the HaloTag. Tagged CBP was observed in every nuclei with the prescence of puncta, when labelled with JF 549 (Figure 7.16). As a negative control HEK293T cells were labelled with JF 549, where only non-specific signal was observed (Figure 7.16). Brightfield images are shown below to indicate the presence of cells in each field of view (Figure 7.16). While one clone was all that was necessary for future experiments, I had hoped to find more clones in the screening procedure so that I would have reproducible biological replicates. This unfortunately was not the case. It was decided that in the course of this doctoral work, no more screening would be performed.

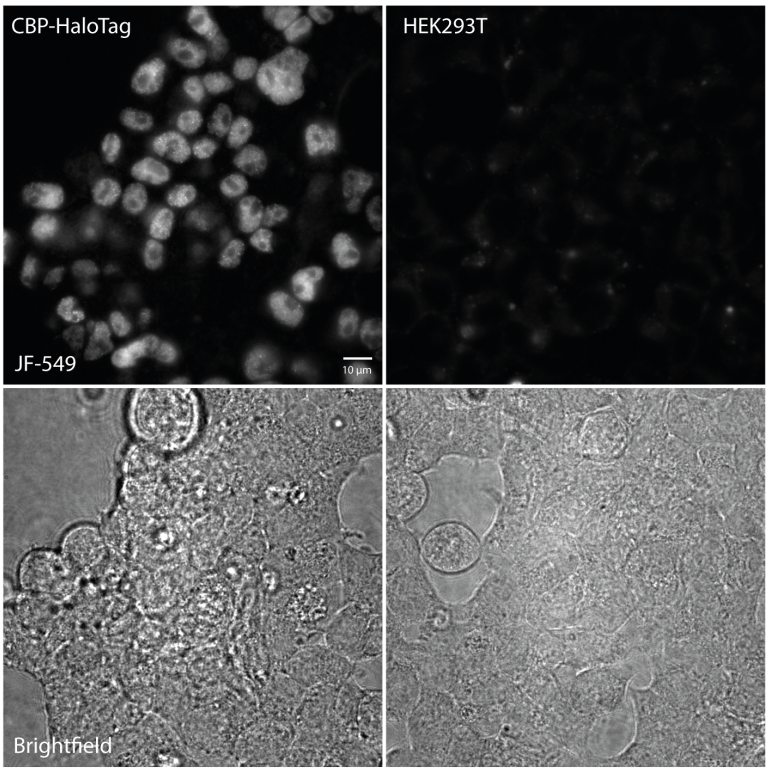

Figure 7.16: Representative image of CBP-HaloTag and HEK293T cells labelled with JF 549 Representative images of CBP-HaloTag clonal line and HEK293T cells when labelled with Janelia Fluor 549. Brightfield images indicate the presence of cells in each field of view. Image was taken in the mid-plane of the Z-stack, therefore not all nuclei are in focus. Images were processed in ImageJ/Fiji, where for background was subtracted from fluorescent images using a rolling ball of 100 pixels. Scale bar represents 10 µm.

Although by imaging I had observed nuclear CBP-HaloTag puncta I also wanted to validate this cell line by western blot, probing for the HaloTag (Promega, G9211), CBP (Cell Signalling, 7389) and GAPDH (Proteintech, 60004-1). The results showed the presence of CBP in all cell lines, with a noticeable increase in molecular weight (MW) in the clonal line due to the addition of the C-terminal tags including the HaloTag (Figure 7.17). In the polyclonal population this upward shift was not observed, this was probably due to the low proportion of cells that carry the HaloTag, from previous imaging experiments this was estimated to be approximately 1 % (Figure 7.17). When probing for the HaloTag a clear band was observed for the clonal population, but no band in the HEK293T cells or in the polyclonal population (Figure 7.17). In the clonal cell

line there is no band present at the height of endogenous CBP, which suggests that this cell line may be homozygous, however this conclusion has yet to be further investigated (Figure 7.17). There also appears to be a reduction in expression of CBP in the clonal and polyclonal cell line compared to the HEK293T cells, suggesting potential down regulation in these cell lines (Figure 7.17). This could be a result of the addition of the tags to the C-terminal region of CBP. This result further validated the presence of the HaloTag in the clonal line, from which future single particle imaging experiments can be performed.

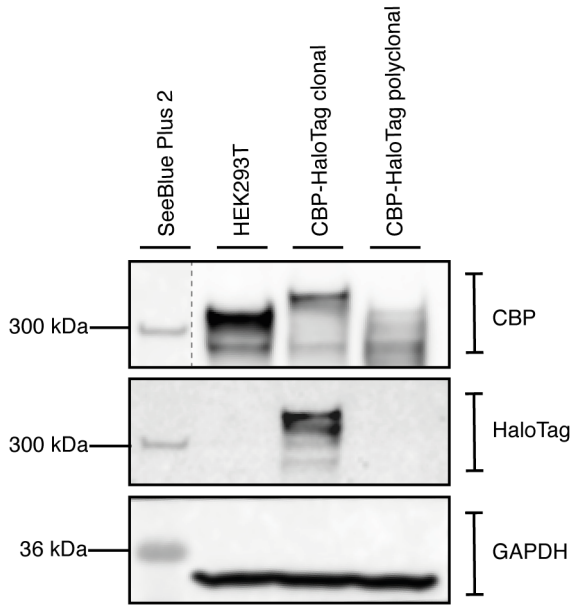

Figure 7.17: Western Blot validation for CBP-HaloTag clonal cell line
Validation of CBP-HaloTag clonal line compared to HEK293T WT cells and original CBP-HaloTag polyclonal population. Western blot was probed for CBP (upper panel), HaloTag (middle panel) and GAPDH (lower panel) for each of the different cell lines, where GAPDH was used as a loading control. Estimated protein sizes in kDa from the SeeBlue Plus 2 Protein ladder, are shown on the left, with the protein the antibody targets being shown on the right hand side. For the CBP westen blott the ladder was added separately, indicated by the grey dotted line.

7.4.4 Tracking single particles

Since I had established the clonal CBP-HaloTag line and determined a concentration of Janelia Fluor to label single particles, I next performed some timecourses to capture single particles of

CBP for tracking analysis. While I had managed to determine some initial conditions to image single particles of CBP I had yet to test whether these videos were suitable for tracking. The first stage of particle tracking is to identify the particles from the background noise in every frame in the timecourse, the background is particularly high in single particle imaging due to the low signal received from imaging only a few labelled particles (Figure 7.18 frame 1). Most tracking software have parameters to identify particles, such as the average diameter in pixels that is expected of a single particle, this helps select particles from the background. Once the particles have been identified, analysis is performed between consecutive frames to determine whether the particles in frame 1 are the same as in frame 2 and 3 and so on through the movie. This is usually done by applying distance parameters, such as a single particle can travel 15 pixels and still be considered the same particle (Figure 7.18). Alongside monitoring the movement of individual particles, in some cases particles may move within close proximity to each other, which will increase confusion for particle tracking softwares, such as in frame 3 for particles A and B in the example below (Figure 7.18). Depending on the resolution of the images and the proximity of the particles, tracking may still be possible, however in some cases this may be too difficult for analysis pipelines to contend with and therefore may need to be excluded from the dataset. Once tracks have been made between the particles identified in different frames, quantification can be performed. This quantification includes average track length of particles and the confinement ratio, which is the net displacement of the particle divided by the total distance travelled, therefore a confinement value close to 0 is a particle that has not moved far from its starting point, such as particle C in this example (Figure 7.18). These parameters will allow the determination of the proportion of CBP molecules which are fully mobile compared to those that are static throughout the given timecourse.

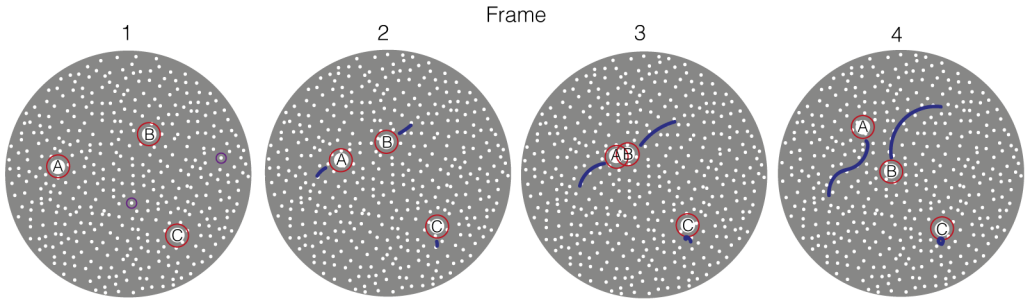

Figure 7.18: Diagram of single particle tracking
Example timecourse of three single particles (A, B and C) within the nucleus that are being tracked over four imaging frames. Initially the single particles need to be identified (red circles) in a way that differentiates them from the background noise (purple circles). These particles need to be identified in all frames, in each subsequent frame from frame 1 analysis needs to be performed that determines if the particles in frame 2 are the same ones as in frame 1, they can then be assigned a reference such as particle A. Due to the free movement of the particles parameters also need to be considered for determining the difference between single particles that move close together such as for particles A and B in frame 3. Once this has been completed and the particles have been sufficiently labelled, tracks can be generated for each particle (blue lines) and from this quantification can occur to determine molecule dynamics.

7.4.4.1 Capturing images to track

I used the CBP-HaloTag clonal cell line and dual labelled with JF 549 at a concentration of 50 nM and Janelia Flour 646 at a concentration of 2 nM to image single particles for tracking. A concentration of 2 nM was used for single particle labelling based on imaging on the day, as labelling of 5 nM (the previous chosen concentration) labelled too many particles. In this case, I imaged both the whole population of CBP as well as single particles to define the nuclear region as I did not use Hoechst. A single image was captured with the 532 nm laser to look at whole nuclei CBP labelling, followed by capturing a timecourse of single particles using the 637 nm laser at 100 % laser power with different exposure times, including: 5 ms, 10 ms, 20 ms and 50 ms. For single particle imaging, I wanted to image as fast as possible to capture the movements of as many CBP particles as possible. Longer exposure times bias the sample towards CBP particles that don't move, as they will be in the same position in each frame compared to faster molecules which may have moved in and out of the imaging volume during the time it takes to capture one image. Compared to previous imaging where I used the full field of view to image CBP-HaloTag, in this case I created a region of interest that isolated only a few nuclei, this reduces the time

it takes for each image to be captured. Since I was still optimising this experiment, the ROI generated was a slightly different size for each exposure time.

Reducing the exposure time showed an increase in background noise, and at 5 ms the particles were very difficult to see and therefore couldn't be used for tracking (data was not shown). At 10 ms, the particles were much clearer to see, with more brightness and definition, that improved with each increase in exposure time (Figure 7.19). Single particles viewed at 50 ms were very bright and easy to view, however there were lots of molecules observed, some very close together which would cause difficulty when isolating individual particles (Figure 7.19). From this experiment I chose to use the 20 ms exposure video to test particle tracking softwares, as it had a suitable signal to noise ratio with a reasonable amount of single molecules to track.

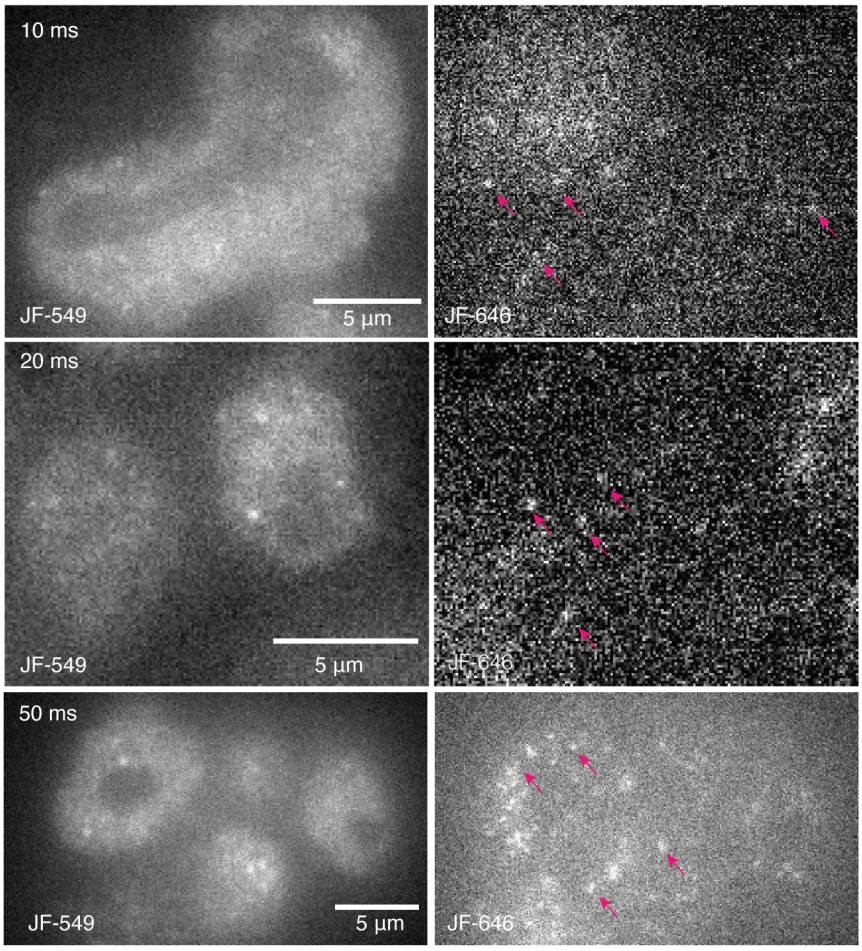

Figure 7.19: CBP-HaloTag single particle videos for tracking using different exposure times CBP-HaloTag cells imaged for whole nuclei CBP using Janelia Fluor 549 (JF 549) and single particles of CBP using Janelia Fluor 646 (JF 646). Optimisation of exposure time to visualise single particles at times of 10 ms, 20 ms and 50 ms. Representative images of both channels are shown here, where pink arrows indicate single particles within a single frame. A single frame was taken using the 532 nm laser for whole nucleus CBP and a timecourse was taken using the 637 nm laser for single molecules of CBP using the above exposure times. Images processed and videos made in Image/Fiji, scale bar represents 10 µm. Biological replicates = 1. Videos of the above timecourses more clearly show the differences in single particles between the conditions, for full videos see: https://figshare.com/s/cbe1ea9ebf3fdb7da05d.

7.4.4.2 Single particle tracking

In this work, I will discuss the use of Detection of Molecules (DoMs) which is used to identify single particles, as well as TrackMate which is a user-friendly plugin for identifying and tracking single particles, and performing pre-set analysis parameters such as track length and confinement ratio, both plugins are available in ImageJ/Fiji (Schindelin et al. 2012; Tinevez et al. 2017; Katrukha 2020). I also explored the use of python tracking software quot and standard array single particle tracking (saSPT) that were designed for super resolution single particle tracking of photoactivated localisation microscopy (sptPALM), however due to time constraints the optimisation of this software was not performed (Heckert et al. 2022).

The timecourse of clonal CBP-HaloTag where an exposure time of 20 ms was used, was imported into ImageJ/Fiji and the contrast was adjusted to increase the signal to noise ratio (Schindelin et al. 2012). Using different softwares I identified the single particles in each frame. In TrackMate, I used the Laplacian of Gaussian (LoG) detector which applies a LoG segmentation to the image and is best used for spot sizes between 5 and 20 pixels, as well as the Differential of Gaussian (DoG) detector, which applies of LoG filter by the difference of 2 Gaussians, which is optimimal for small spot sizes (Figure 7.20) (Tinevez et al. 2017). Both need input information such as average size of a particle in pixels; this was calculated to be 7 pixels, by drawing a line on top of a representative particle and counting the number of pixels along that line. Other parameters such as quality threshold and application of a preprocessed median filter and sub pixel localisation were applied through an iterative process to find the optimum settings to identify single particles, while reducing false positives. Once I had decided on a set of conditions for the particle detector I next tried to optimise the particle tracking - for this I used the particles tracked using the DoG detector. Again, an iterative process was performed to optimise the tracking of particles. Parameters for tracking were set, such as how far particles can move between frames to still be classed as the same particle (8 pixels), as well as parameters for gap closing such as how many frames until the particle was a different particle (5 frames). Applying these parameters generated tracks which were overlaid on the video to check the accuracy of the track plotting (Figure 7.20). Tracks were then quantified to calculate the total distance of the track over the timecourse, and how far the particle had moved. Another similar parameter, the confinement ratio, which is the net displacement of the particle divided by the distance, indicates whether a single particle was stationary or more mobile (Figure 7.20). The data allows the calculation of the percentage of particles that are mobile compared to stationary, and what proportion of molecules can move a certain distance; this final quantification was not performed.

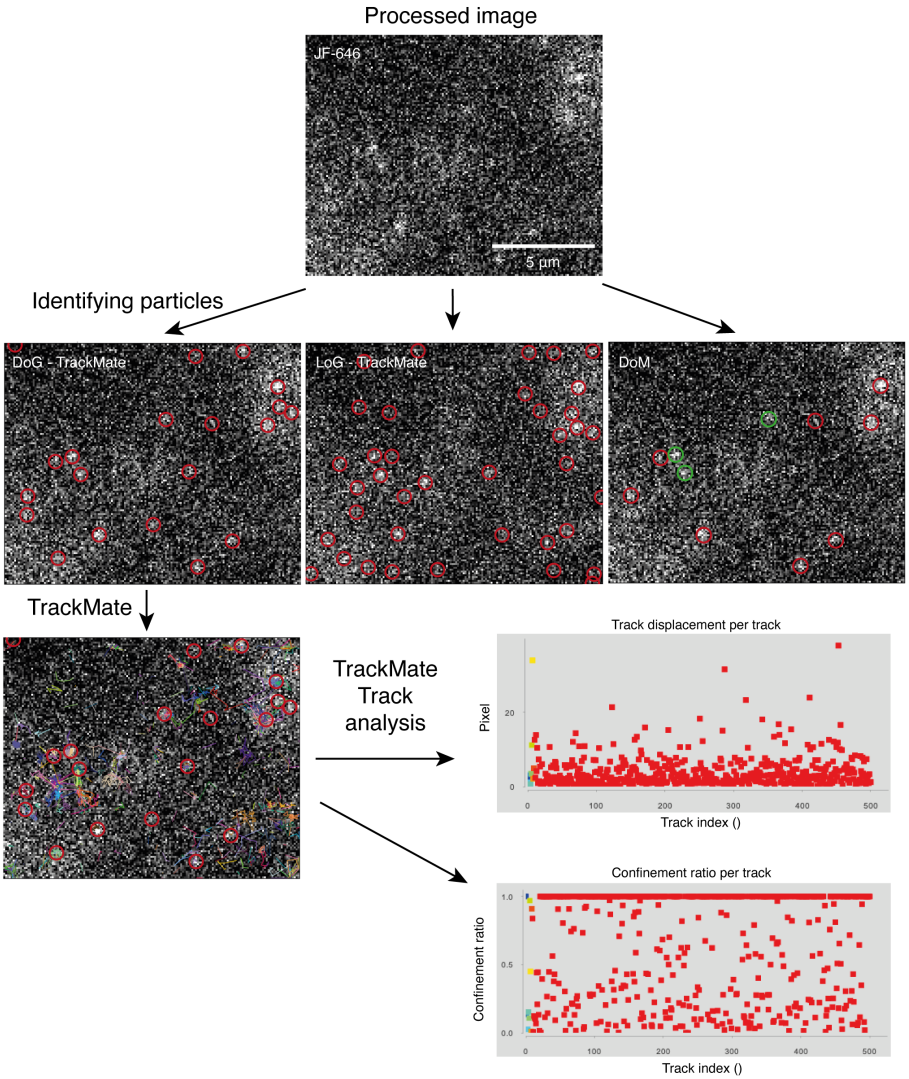

Figure 7.20: Optimization of different tracking softwares
Tracking optimization workflow of single particles of CBP-HaloTag labelled with Janelia Flour 646 (JF 646) using an exposure time of 20 ms. Single particles in ImageJ/Fiji were tracked using either TrackMate software or Detection of Molecules (DoM). DoM applies a quality threshold, particles which pass that threshold are shown in green, particles that do not pass that threshold are shown in red. Using TrackMate two different detectors were tested to track particles: Difference of Gaussian particle detection (DoG) or Laplacian of Gaussian particle detection (LoG). From DoG tracked particles, tracks were generated, an example is shown here, where individual tracks are coloured. Quantification of the tracks can be performed within ImageJ/Fiji, two examples are shown here (Tinevez et al. 2017; Katrukha 2020).

I also tried using DoM to detect particles and plot tracks (Katrukha 2020). Analogous to TrackMate several parameters were applied, such as particle size (660 nm) and number of iterations to detect the particles (5). Unlike TrackMate, DoM applies a quality threshold where it determines the confidence in a particle; if the detected particle passed the quality threshold they were shown in green, and if they did not they were shown in red (Figure 7.20). Using this software you can also link the particles to generate tracks, applying parameters such as maximum searching distance (8 pixels) and maximum gap to be classed as the same particle (5 frames). Like TrackMate, the determined tracks were overlaid on top of the video and a table was produced of the raw information concerning the tracks. DoM does not produce any quantification based on the tracks, this needs to be manually computed downstream. Due to time constraints this was not attempted.

Using initial testing conditions I have managed to detect single particles of CBP-HaloTag and track them over a timecourse, generating quantitative values to begin to understand the working dynamics of single particles of CBP-HaloTag. No further testing was performed in optimisation of single particle tracking within the time frame of this doctoral work.

7.5 Discussion

In this chapter I have begun to validate systems which will allow the determination of how phase separation affects CBP HAT activity and nuclear dynamics in future experiments.

7.5.1 Amber suppression technology for FRET approaches

To monitor conformational changes in the CBP HAT domain through a FRET based approach, I needed a method to label CBP with the least amount of disruption to the protein sequence. I chose to do this using the amber codon suppression system, the name of which refers to the use of the amber stop codon TAG (Koehler et al. 2020). Using this system, only one amino acid in the AIL was changed, with the unnatural amino acids having a similar size to that of the bulkier side chains such as asparagine or glutamine, therefore the effect of inserting this amino acid should be minimal (Koehler et al. 2020; Nikić et al. 2015; Schmied et al. 2014). This was of high priority in this work since the mutation is within the AIL of CBP, which upon mutation may cause unknown functional consequences. This prevented the tagging of the AIL with conventional approaches, such as using a tag which can be labelled, such as the SNAP tag, which I have used for adding the second fluorophore for smFRET. The main disadvantage to using this system is that high levels of orthogonal tRNA are required to suppress the amber codon, which may interfere with host cell

function, having unknown consequences (Koehler et al. 2020). In recent years, advances have been made to increase the efficiency of the system such as using mutant eRF1, which helps mitigate this caveat (Schmied et al. 2014). This system has been used in the literature to measure distances between proteins in the plasma membrane for example, allowing for speculation of rearrangements of proteins relative to the plasma membrane (Zagotta et al. 2016). This system was therefore a clear choice for labelling CBP with the hopes of mapping conformational dynamics, and CBP HAT activity using FRET based approaches.

7.5.2 Using SNAP tag for smFRET

For smFRET I generated constructs that contain a TAG codon for uaa incorporation and labelling for smFRET, alongside using a SNAP tag which facilitates labelling with a secondary fluorophore. The SNAP tag itself is versatile and is highly compatible with labelling, and allows conjugations to many different chemical groups, providing options for fluorescent labelling in the smFRET approach (Merlo et al. 2021). The SNAP tag itself is 181 amino acids in length, which is small relative to CBP (approximately 2442 amino acids), however this is still a large protein to insert into the protein coding sequence of CBP. Since I need CBP to be functionally active, care was taken on the placement of the tag. To mitigate the potential functional effects of altering confirmations of CBP by adding such a large tag, the sites were chosen outside of the catalytic core region. The first site was within the intrinsically disordered region before the bromodomain, and the second was downstream of the HAT domain. AlphaFold predictions were run of both constructs, by Dr Dan Bose, which predicted no effect on structure of the core domain compared to WT (Figure 7.2) (Jumper et al. 2021). This experimental approach was therefore designed to optimise labelling options and reduce functional effects of labelling CBP for future smFRET and tmFRET approaches.

7.5.3 Future work for smFRET

In this work, I detailed initial validation experiments for using the amber codon suppression system to produce overexpressed CBP-HaloTag constructs containing amino acid TCO*A and a SNAP tag for future smFRET experiments. I successfully managed to express CBP_{HHDE*D}-HaloTag-Flag in HEK293FT cells, which was an extremely promising start to looking at enzymatic activity of CBP using a single molecule approach. However, this was the first step of many needed to get this system functioning. The next immediate experiment will be optimising amino acid concentration or concentration of DNA for transfection to produce only single particles of this construct. Along-

side optimising for single particles, I need to optimise labelling of the TCO*A amino acid and SNAP tag, with different fluorophores, with enough efficiency that a high proportion of molecules will be labelled with both fluorophores. Firstly, I will optimise labelling of the TCO*A and SNAP independently and then begin to try dual labelling. This may be more challenging for the in cell approach compared to after protein purification, as there are more variables to consider within live cells. Once I have the correct conditions optimised I will further need to optimise imaging for FRET purposes, using custom TIRF microscope BONSAI. Alongside optimisation of the cell work I will also need to optimise purification of CBP-HaloTag-Flag constructs using the Flag tag to perform *in vitro* experiments. Once I have validated the smFRET itself, I will be able to determine if phase separation affects the HAT activity of CBP.

7.5.4 Future work for tmFRET

In this chapter, initial unsuccessful validation experiments were described for tmFRET which need further optimisation. Future experiments will include optimisation of the amber codon system to increase the expression of the GFP positive control plasmid compared to background. Initial optimization will focus on performing an ACD concentration curve to optimise the amino acid concentration to favour incorporation into GFP to increase the expression of the desired protein, containing the amino acid needed for tmFRET. Another optimization strategy includes testing different concentrations of the orthogonal tRNA used within the system, as large amounts are required for successful suppression of the amber codon (Koehler et al. 2020). Once the system is optimised several further experiments need to be performed, similar to those for the smFRET including: optimising for single particles of the protein in a cellular context; purifying protein from HEK293FT cells and optimising imaging for tmFRET.

Due to time constraints I chose not to proceed with optimisation of the tmFRET system, and focussed on single particle imaging of endogenous CBP-Halotag.

7.5.5 Single particle imaging of endogenous CBP-HaloTag

I have shown optimisation experiments to visualise single particle imaging of CBP-HaloTag including testing different labelling methods, concentrations of ligand and image optimisation. For future experiments, which investigate single particles of CBP within cells transfected with GFP tagged constructs a CBP-HaloTag clonal line was generated to increase the probability of a dual labelled nuclei. One benefit of the clonal line is that I no longer have to dual label which JFs to determine which nuclei contain the HaloTag. I did not optimise for single labelling with JFs, however

in the absence of competitive labelling, which is experienced when two dyes are labelling the same tag, a lower concentration of dye should be used, such as 0.5 nM. While the data generated with this clonal line aligns with what was observed within the polyclonal cell line, there is currently only one clonal line, and therefore one biological replicate. For future work into single particle imaging, an attempt should be made to generate more clonal cell lines with which conclusions can be drawn with a greater degree of certainty.

Alongside labelling conditions I tested different exposure times on the microscope, the shorter the exposure time the more movements of particles can be captured, however this comes at the cost of resolution. Initially I tested imaging using an exposure time of 100 ms, however this biassed the samples towards static objects as particles that move in and out of the imaging frame in 100 ms would not be captured. I therefore tried a range of different exposure times from 5 ms to 50 ms. I chose to use 20 ms since it had a sufficient signal to noise ratio, and had previously been used to image single molecules of glutamate receptors in neurons (Scheefhals, Westra and MacGillavry 2023). Exposure times below 20 ms produced particles which were harder to distinguish from the background, if it is necessary to image at these lower exposure times I would need to boost the signal. At the time, I was using 100 % laser power, however using the custom TIRF microscope BONSAI there is a possibility to double this laser power which could sufficiently increase the signal to noise ratio in these samples. For initial optimisation of tracking single particles, the example video using a 20 ms exposure was sufficient.

7.5.6 Tracking of CBP-HaloTag

In order to track single particles of CBP, softwares TrackMate and DoM were used to detect single particles. DoM was tested to detect particles and attempt to link these particles together, while this was successful, fewer examples of this have been used in the literature, and more downstream processing is needed than that of TrackMate (Katrukha 2020). TrackMate has been widely used in the literature to track molecules, both at a single particle level and for tracking single particles in a bulk population (Liu et al. 2022; Tinevez et al. 2017; Bayle et al. 2021). This platform is user friendly and auto generates values for comparison of single particles of CBP, such as distance travelled. While an example pipeline was defined in this doctoral work, this was based off one video, for future optimisation experiments I would need a greater number of example videos to optimise the settings for CBP.

Other than the softwares described above, one pipeline I wanted to investigate was using quot followed by saSPT python packages (Heckert et al. 2022). Unlike the previous examples, these

packages function mainly through coding, with certain aspects such as the particle tracking having a graphical user interface for better optimisation of the settings used to track the particles (Heckert et al. 2022). These packages were developed for single molecule tracking events and provide the quantification of diffusion coefficients of different particles which would allow the determination of different populations of CBP molecules which have different diffusion coefficients. In future experiments I want to optimise using this pipeline to track and quantify single particles of CBP. Unfortunately, time constraints did not allow for optimisation of quot and saSPT within this doctoral work.

Other methods that could be used to determine diffusion coefficients of CBP are Raster Image Correlations Spectroscopy (RICS) and Fluorescence Correlation Spectroscopy (FCS) (Yu et al. 2021; Rossow et al. 2010). RICS can estimate disffusion coefficients from microseconds to milliseconds, by analysing images captured from any conventional scanning microscope, where the images have been captured in raster format (measuring the fluorescence per pixel in a left to right fashion, starting from row one, and once complete moving to the left of row two, and so on and so forth) (Rossow et al. 2010). This can further be performed at a single particle level using Single Particle Raster Image Analysis (SPRIA) (Longfils et al. 2018). FCS is a similar technique that can measure diffusion coefficients from a nanosecond to second scale using a confocal experimental setup and data analysis pipeline, which can also be performed at a single particle level (Yu et al. 2021; Berglund and Mabuchi 2005). While this is not something I am actively pursuing, these techniques could be used in the future to support data gathered on the different diffusing populations of endogenous CBP, and how this changes in different phase separated environments.

7.5.7 Future perspectives on single particle tracking of CBP

The work discussed in this chapter introduces several key experiments to investigate how condensates affect CBP dynamics and function. Taken together this work will allow a more detailed understanding of the complex environment of phase separated condensate, and shed light on the role of these condensates being either transcriptional condensates or static repositories of protein when transcription is not needed.

Chapter 8

Discussion

8.1 Importance of understanding CBP phase separation

The work described here enables the understanding of how phase separation influences key proteins involved in gene transcription, such as CBP. I have shown that phase separation of CBP is not driven by one particular region but it is the combination of phase behaviours of different regions, which gives CBP the overall phenotype of phase separation. In generating deletion mutants which have different phase behaviours, I have developed a unique opportunity to test the activity and dynamics of CBP within the context of different phase behaviours. The protein toolkit I have developed will allow the expansion of current knowledge of how phase separation influences CBP activity, providing a greater understanding of the role phase separation plays in enhancer function.

8.2 Conclusions

8.2.1 CBP forms liquid-like condensates

This research shows that CBP can form phase separated condensates at endogenous levels by immunofluorescence (Figure 3.2) and using an endogenously tagged cell line CBP-HaloTag (Figure 3.10). Distinct nuclear puncta was observed that supports CBP condensates observed in the literature (Wang et al. 2022b; Ma et al. 2021; Zhang et al. 2021). Using CBP-GFP I developed an overexpression model system in mammalian cells, visualising nuclear puncta that behaved in a dynamic manner via fluorescence recovery after photobleaching (FRAP) and 1,6 - hexanediol (1,6-HD) treatment. I confirmed that overexpressed CBP was forming condensates that were

fluid-like and not forming solid protein aggregates.

This work characterised full length CBP phase separation using multiple experimental techniques, showing that this phenomena is reproducible and physiologically relevant. The techniques developed here including visualising endogenous CBP puncta, and establishing an overexpression model system, provided a unique opportunity to study a range of different parameters that could influence CBP phase separation. Developing an overexpression system, allowed the flexibility to manipulate the sequence of CBP and understand what effect different regions have on phase separation without altering the endogenous protein.

I therefore focused on dissecting which regions within CBP were important for driving this phase separating behaviour.

8.2.2 Identifying regions which affect CBP phase separation

8.2.2.1 IDR6, IDR7 and CFID are key regulators of CBP phase separation

Initial studies into what drives CBP phase separation focused on identifying regions of interest including intrinsic disordered regions (IDRs) and the CBP FUS Interaction Domain (CFID) and understanding their ability to phase separate. Using a combination of the optoDroplet assay and overexpression of CBP-GFP deletion mutants I determined three key C-terminal regions that influence the phase behaviour of CBP, these were the IDR6, IDR7 and the CFID; where the CFID is comprised of most of the IDR6 and the IDR7. I have shown that the IDR7 and CFID regions phase separate independently in the optoDroplet system, and when deleted from CBP-GFP they inhibit phase separation, producing a more diffuse pattern of expression. In contrast, the IDR6 region cannot form puncta independently using the optoDroplet assay, and when deleted from CBP-GFP leads to the formation of larger puncta than that observed for WT. All other IDRs did not form condensates in the optoDroplet system and and did not affect puncta formation when removed from CBP-GFP, and so they were not investigated further.

It is interesting that the CFID region encompasses all of the IDR7 region and most of the IDR6 region, and yet displays a similar phenotype to IDR7, showing diffuse expression of protein when deleted from CBP-GFP. This further implicates the importance of the C-terminal region on CBP phase separation. Together, this suggests a role for IDR7 in driving phase separation of CBP, and IDR6 in inhibiting it. Since the phenotype of CBP-GFP is distinct nuclear puncta, which is an intermediate between what was seen between the $CBP_{\Delta IDR6}$-GFP and $CBP_{\Delta IDR7}$-GFP, it is possible that the IDR6 region balances the phenotype of the IDR7 to produce that of the WT phenotype.

When studying protein phase separation, it is typically identified that one particular IDR, or low complexity (LC) region, is important for driving the phase separation of that protein; one example is the C-terminal LC domain of TDP-43 which has been shown to drive condensate formation (Babinchak et al. 2019; Conicella et al. 2016; McGurk et al. 2018). Since complex interactions between specific IDRs have been shown to be important for phase separation, care must be taken when defining these regions in order to understand their influence on phase separation. If I had chosen to define CBPs IDRs more broadly at the N- and C-terminus of CBP, I likely would have identified a similar phenotype to the CFID, implicating the C-terminal region as driving CBP phase separation; this would not reflect the complex, opposing interactions which contribute to CBPs overall phase behaviour. While this is protein specific, for CBP, defining stringent IDRs within broader disordered regions has identified specific IDRs with independent phase separating properties. It is currently unknown how the IDR6 and IDR7 region may be interacting to cause this balancing phenotype.

8.2.2.2 The importance of the AIL in CBP phase separation

I also investigated the autoinhibitory loop (AIL), as the AIL is important for regulating HAT activity, and has been shown to influence the phase behaviour of p300, I wanted to observe what effect the removal of the AIL may have on CBP phase separation. Two mutants were generated to understand how the AIL, or more specifically the lysine residues within the AIL, affect phase separation; these were termed $CBP_{\Delta AIL}$-GFP and CBP_{KTG}-GFP. In these studies, I showed that both mutants, $CBP_{\Delta AIL}$-GFP and CBP_{KTG}-GFP, cause the formation of larger phase separated condensates compared to WT; suggesting that the AIL and more specifically the lysine residues were inhibiting the phase separation of WT CBP. Therefore the AIL is an important regulator of CBP phase separation.

The AIL itself could be inhibiting phase separation, however the CBP_{KTG}-GFP mutant suggests that it is the lysine residues which are playing a key role in altering phase separation. What is still unclear is if this larger phase separation phenotype is caused by the removal of the lysine residues themselves, or is it the removal of the ability to be acetylated which leads to this behaviour. Acetylation has been suggested in the literature to inhibit phase separation, as has been discussed previously in subsection 1.1.1.1. Both mutants, $CBP_{\Delta AIL}$-GFP and CBP_{KTG}-GFP, are missing the key lysine residues within the AIL which can be acetylated to stimulate HAT activity, both of which formed larger condensates than that observed with the WT protein. Therefore it is possible that in removing the acetyl groups from this region the ability to inhibit phase separation

is also being removed, independent of the rest of the sequence. The acetyl group is negatively charged therefore it is possible that in replacing the positive charge of the lysine group with the negative charge of the acetyl group, electrostatic interactions are being disrupted which prevents larger condensates forming. This work suggests that acetylation could potentially attenuate condensate formation, supporting the studies of Zhang *et al.* (2021), Ferreon *et al.* (2018) and Saito *et al.* (2019). To test whether acetylation of the lysine residues within the AIL is responsible for inhibiting large condensate formation, I could replace the lysine residues with acetyllysine mimic glutamine, which would display a WT condensate phenotype, if this hypothesis is correct.

8.2.2.3 Phase separation of CBP is a combination of regions with differing phase separating behaviours

I also generated double deletion constructs of the main phase separating mutants, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP, and $CBP_{\Delta CFID}$-GFP that were additionally mutated with ΔAIL or the KTG. The addition of either of these mutations did not alter the phenotype of the original mutant; for example, $CBP_{\Delta IDR6}$-GFP forms large puncta which are not significantly different to $CBP_{\Delta IDR6:\Delta AIL}$-GFP or $CBP_{\Delta IDR6:KTG}$-GFP (Figure 5.7). This suggests some mutations may have a dominant effect on CBP phase separation, where the IDR6 region holds a greater influence than either of the AIL mutations, for example.

In the case of CBP, what I show here is that there is not one region which is important for this behaviour, but multiple. Importantly, I have identified regions which both inhibit and drive phase separation, where it is the accumulation of these different influences which are important for CBPs full length phase separation phenotype.

8.2.3 Implications for HAT activity within a condensate

What is still unclear is whether CBP, or any of the mutant constructs are catalytically active within phase separated condensates. Since the mutations which influence phase separation are localised to the C-terminal region, away from the HAT domain (which is responsible for CBPs catalytic activity), I suspect that these mutants are catalytically active.

So far, the closest I have come to understanding the relationship between HAT activity and phase separation is through the $CBP_{\Delta AIL}$-GFP and CBP_{KTG}-GFP mutants. While I haven't performed HAT assays on either mutant, similar mutants were made by Bose *et al.* (2017), who showed that both deletion of the AIL or the reversal of the charge of the lysine residues, from positive to negative, left CBP catalytically active. This suggests that these CBP mutants may be

catalytically active within larger condensates.

8.2.4 Doxycycline inducible cell lines

Intial work on CBP phase separatiton focussed on using an overexpression model system to determine the effect of IDRs on phase separation of CBP, which highlighted three regions: IDR6, IDR7 and CFID. Since phase separation is heavily influenced by concentration, I made cell lines containing these constructs of interest under the control of an inducible promoter to express near-endogenous levels of CBP.

8.2.4.1 IDR6, IDR7 and CFID influence CBP phase separation at endogenous concentrations

Expressing constructs CBP-GFP, CBP$_{\Delta IDR6}$-GFP, CBP$_{\Delta IDR7}$-GFP and CBP$_{\Delta CFID}$-GFP at near-endogenous concentration, all showed a consistent phenotype compared to in their overexpressed form. Both CBP$_{\Delta IDR7}$-GFP and CBP$_{\Delta CFID}$-GFP show a diffuse pattern of expression - suggesting that at physiological concentrations, the IDR7 is contributing to CBP phase separation. CBP$_{\Delta IDR6}$-GFP at endogenous concentrations still forms condensates, but they were smaller than those observed for the overexpressed protein, displaying a similar phenotype to the CBP-GFP protein. As all mutants at near-endogenous concentrations showed the same overall phenotype, compared to their overexpressed counterparts, cells expressing these mutants at physiological concentrations can be used to begin to understand how differences in phase behaviour can alter characteristics such as chromatin localisation.

8.2.4.2 Phase separation influences chromatin localisation

CBP peaks from ChIP-Seq are commonly identified at enhancer regions within the genome, along with histone modifications H3K27ac, H3K4me and RNAP II localisation (Mansour et al. 2014; Blinka et al. 2017; Visel et al. 2009). I therefore wanted to ask how these GFP tagged constructs interact with chromatin, and if this interaction changes with the IDR mutants constructs.

Using ChIP-Seq I observed that the Dox induced GFP protein variants localise to the same regulatory regions as endogenous CBP, with overlap of CBP-GFP and endogenous CBP peaks (Figure 6.9). I also observed differences in chromatin binding dependent on the phase behaviour of the different constructs. More numerous peaks that spanned larger regions of the chromatin were identified for CBP$_{\Delta IDR6}$-GFP, which forms larger condensates compared to CBP-GFP (Table 6.1, Figure 6.9). In contrast, shorter peaks were identified for CBP$_{\Delta IDR7}$-GFP which behaves in a

more diffuse manner (Table 6.1, Figure 6.9). This work indicates a role for phase separation in genome localisation of CBP, which may influence CBP function by affecting CBP binding within regulatory regions.

It will be interesting to view these results in the light of future experiments which will indicate whether CBP is catalytically active or not. Combining HAT assays with ChIP-Seq analysis will provide a greater degree of detail into the functional role of phase separation on CBP activity at enhancer regions.

8.2.5 Overexpressed CBP mutant constructs can influence endogenous CBP behaviour

To understand whether different phase separating mutants of CBP could alter the behaviour of endogenous CBP, I overexpressed CBP muatant constructs in the CBP-HaloTag cell line. The results suggests that the overexpressed protein has a dominant effect on the phase behaviour of the endogenous protein. CBP-HaloTag cells that were transfected with CBP-GFP, showed GFP puncta which colocalise with that of the endogenous protein (Figure 6.15). When transfected with $CBP_{\Delta IDR6}$-GFP, large condensates were observed for the overexpressed protein and for the endogenously tagged protein, suggesting recruitment of endogenous CBP into larger condensates (Figure 6.15). Mutants $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP displayed a dominant negative effect, where fewer endogenous puncta were observed in nuclei that were transfected (Figure 6.15). While this was based on one experimental repeat, and has yet to be quantified, this suggests that endogenous CBP is influenced by the different phase behaviours of the CBP constructs.

Since this experiment used captured images taken from live cells of CBP-HaloTag, it did not provide information on how overexpression of GFP tagged constructs changed the dynamics of endogenous CBP, for example within larger condensates. To visualise the movement of CBP within condensates I could capture a timecourse to see how endogenous CBP dynamically behaves when transfected with CBP-GFP and other constructs. However, this would only provide an average view of the dynamics of all labelled molecules within the condensate. To generate more specific data which can be quantified to determine dynamics within a condensate, I would need to image single particles of CBP-HaloTag.

8.2.6 Single molecule CBP experiments

Since I have characterised the phenotypic behaviour of full length CBP phase separation, the next goal was to determine how phase separation influences CBP nuclear dynamics and HAT

activity. To tackle this complicated question, I turned to using single particle approaches. During the course of this doctoral work, I started to validate systems which I can use for this purpose. I began to establish codon expansion experiments to monitor conformational changes, and therefore HAT activity, within a condensate. I also imaged single molecules of CBP-HaloTag, to determine how nuclear dynamics of full length CBP changes within different phase separated environments.

The work decribed within this doctoral thesis is exteremly preliminary and as such no conclusions can yet be made on how phase separations affects CBP activity and dynamics.

8.3 Future experiments

In characterising the phase separating ability of CBP, and identifying regions which are important for phase separation, I have developed a protein toolkit which can be exploited to further understand how phase separation influences CBP activity, and the environment at enhancers.

8.3.1 Using the doxycycline inducible cell line to determine if phase separation promotes recruitment of different eRNA molecules

eRNA molecules are transcribed from the enhancer DNA sequence by RNAPII, and have been shown to activate CBP HAT activity (Blinka et al. 2017; Harrison and Bose 2022; Lee et al. 2021b; Sartorelli and Lauberth 2020; Mousavi et al. 2013; Bose et al. 2017). Many RNA molecules, including eRNAs, have been shown to form condensates with proteins, acting as a scaffold and aiding in weak multivalent interactions which hold the condensate together (Decker et al. 2022; Henninger et al. 2020; Shrinivas et al. 2019; Banani et al. 2016). It would therefore be interesting to understand if eRNAs are bound to CBP within a phase separated condensate, and whether this changes within the context of different phase separated states.

Using the doxycycline inducible system I could interrogate the different eRNA population bound to CBP-GFP and the different phase separating mutants, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP using improved individual nucleotide crosslinking immunoprecipitation followed by next generation sequencing (iCLIP-Seq). Crosslinking immunoprecipitation (CLIP) is used to provide information on protein-RNA interactions, where iCLIP provides an updated method which allows quicker execution and transcript enrichment (Lee et al. 2021a). To do this I would induce expression of these constructs to endogenous levels, and subsequently irradiate the cells using Ultra Violet (UV) radiation so that RNA bound to any of the CBP constructs are covalently linked together. I would then isolate the protein of interest via a pull down (using either

GFP or CBP, depending on the cell line) from cell lysate, reverse the cross-linking and degrade any residual protein and DNA, isolating the RNA population bound to the different constructs. High throughput sequencing of these RNAs would identify any differences in sequence of eRNA molecules that are bound to CBP.

Since eRNAs have been shown to activate CBPs HAT activity, it would be interesting to see if between the different mutants, which display different phase behaviours, there are differences in the populations of eRNAs that are bound to CBP. Coupled with HAT assays this would provide an overall picture of how CBP activity is influenced by eRNA binding and phase behaviour.

8.3.2 Investigating how phase separation affects CBP HAT activity

In this work so far, I have yet to determine whether CBP or any of the deletion constructs are catalytically active. While in the process of determining a single particle approach to tackle this question, this can be done *in vitro*. In the future, this will be investigated using the GFP tagged proteins. Purification of overexpressed GFP tagged constructs from HEK293FT cells, will be performed using a GFP pull down experiment. Catalytic activity of these constructs can then be investigated either by western blot to look for H3K27ac, and other indicators of HAT activity, or using a fluorescence based HAT assay (Abcam). Together this will provide information on how the different mutant constructs, and therefore different phase behaviours, affect HAT activity.

8.3.3 Using smFRET and tmFRET to study CBP conformational dynamics

Understanding HAT activity within condensates is challenging and usually involves correlation between images that show condensates, and biochemical tests to check for acetylation or experiments *in vitro* to suggest protein activity; an indirect approach such as this is described above in section 8.3.2. While this would imply HAT activity within condensates, a direct approach would be to monitor HAT activity within a condensate via imaging; this is what I aim to do by single molecule fluorescence resonance energy transfer (smFRET) and transition metal-ion fluorescence resonance energy transfer (tmFRET) experiments detailed within this doctoral work.

Using genetic codon expansion I generated CBP-HaloTag-Flag mutant constructs that can be fluorescently labelled to track conformations of the HAT domain of CBP. Conformations of CBP can be mapped via fluorescence using FRET, where FRET efficiency can be used to determine the distance between donor and acceptor fluorophores, providing information on CBP conformation dynamics. These dynamics can be correlated with HAT activity due to changes in conformation within the HAT domain that happens upon activation, with relation to the displacement of the

AIL. Several experiments still need to be performed, detailed in section 7.5, before I can measure FRET and conformational dynamics.

Once I have optimised this system, I can express these constructs in the Dox inducible cell lines of CBP-GFP, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP. This will allow the understanding of whether there are any differences in conformational dynamics, of the overexpressed CBP-HaloTag-Flag constructs within cells, in the context of the different phase separating mutants, thus implying if there are differences in HAT activity. I can subsequently purify the CBP-HaloTag-Flag mutant constructs via a Flag pull down experiment and test conformation dynamics *in vitro*. I can also purify CBP-GFP, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP using the GFP tag, to seed into the *in vitro* experiments to test how conformations change outside of a cellular environment, under different phase conditions. These two complementary experiments would provide a thorough analysis of the connection between phase separation and activity.

Alongside understanding the role of phase separation on CBP activity, I also wanted to understand if changes in dynamics occur within different phase separated states.

8.3.4 Understanding CBP dynamics within phase separated condensates

I began to set up experiments to label single particles of CBP-HaloTag, then image these particles at a rate at which can be tracked. I observed single particles of CBP-HaloTag which move both dynamically and in a more static fashion within the nucleus, from which I began to optimise tracking to further quantify this behaviour. This work is in its infancy and needs optimization before quantification can be performed. Future experiments involve capturing more videos of single particles of CBP-HaloTag and optimising tracking to quantify diffusion coefficients from different populations of CBP (full description in section 7.5.5).

The overall aim of this work was to label CBP-HaloTag cells for single particle imaging and overexpress CBP-GFP, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP to see if the dynamics of endogenous CBP changes within a diffuse phase separated environment compared to regular WT puncta or larger phase separated condensates. This builds on initial work performed on colocalisation as described in section 8.2.5. Overall, this work will show the direct impact of phase separation on the dynamic behaviour of CBP.

8.3.5 *In vitro* experiments

The work detailed in this thesis focused on utilising cellular techniques to investigate CBP condensates, which is not uncommon in the field of phase separation. However, many studies utilise *in vitro* techniques to determine the specific conditions needed for phase separation. In the absence of other cellular components, determining the concentration required for phase separation, and the effect of temperature, salt and pH is much easier.

8.3.5.1 Challenges of expressing CBP recombinantly

Recombinant protein is used to perform *in vitro* experiments, which is typically purified from bacterial or insect cells (Airenne et al. 2013; Lesch et al. 2008; Langlais et al. 2007; Alberti et al. 2018). Expressing CBP recombinantly is extremely challenging due to its large size and its disordered nature. It is not feasible to express CBP, being approximately 300 kDa in size, in a bacterial expression system, therefore an insect cell system has to be used (Langlais et al. 2007). Insect cells are capable of expressing protein of larger sizes, which can be introduced to the cells via transfection using a lentiviral or baculovirus system (Lesch et al. 2008; Airenne et al. 2013). The baculovirus system is better equipped to express proteins of larger size, which is needed to express CBP (Lesch et al. 2008; Airenne et al. 2013). Using a baculovirus expression system, Dr Archna Shah, has managed to purify CBP-GFP, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP. These were cloned into a pFastBac plasmid, which is a baculovirus expression plasmid. However, expression of such proteins was challenging and the concentration of protein produced was too low for *in vitro* experimentation. To perform *in vitro* experiments in the future I would need to optimise protein expression and concentration, one way I could try to optimise this is through cloning a codon optimised version of CBP into the pFastBac plasmid, which should increase the expression of protein from the insect cells. Once I have purified recombinant protein of CBP-GFP, $CBP_{\Delta IDR6}$-GFP, $CBP_{\Delta IDR7}$-GFP and $CBP_{\Delta CFID}$-GFP, I can begin to study CBP phase separation *in vitro*.

8.3.5.2 *In vitro* experiments for recombinant CBP

To complement the work done in this doctoral thesis, *in vitro* experiments would provide a greater degree of knowledge on how CBP phase separates. Using *in vitro* experiments the effect of concentration of protein, and factors such as temperature, salt and pH on the phase separating ability of CBP could be tested.

One of the most important factors to determine is the concentration of protein necessary for phase separation; this will indicate if phase separation is biologically relevant, if the concentration is similar to the physiological concentration of protein within a cell. One way to determine the concentration necessary for phase separation, also called the saturation concentration, is performed by centrifuging a solution containing different concentrations of protein (Hallegger et al. 2021). The centrifugation separates out the different phase states, and allows the determination of concentration needed to phase separate the protein of interest (Hallegger et al. 2021).

More commonly, this is performed via microscopy, where a solution containing the protein of interest is generated to see if it forms condensates on a coverslip (Venko and Žerovnik 2023; Hallegger et al. 2021; Zhang et al. 2021; König et al. 2021; Alberti et al. 2018). These experiments are sometimes performed with molecular crowding agents such as polyethylene glycol (PEG) or Dextran, which mimics the molecular crowding which occurs within cells and aids in condensate formation (Venko and Žerovnik 2023; Hallegger et al. 2021; Zhang et al. 2021; König et al. 2021). Solutions containing different concentrations of protein can then be tested to determine the minimum protein concentration necessary for phase separation. Within these solutions different salt concentrations, pH's and the temperature can be changed to define a set of optimal conditions for CBP phase separation. Droplets formed *in vitro* can also undergo FRAP and 1,6-HD treatment to measure the dynamics of protein in solution; this helps to determine whether the droplets formed are protein aggregates or more fluid-like.

Given their artificial nature, *in vitro* studies are not vital to understand phase separation of CBP, however it would provide further evidence on the specific set of criteria that is needed for CBP phase separation, contributing to the data already generated within this docotoral work.

8.4 Concluding remarks

The work presented in this thesis characterises the phase separating ability of CBP, at both endogenous levels and when overexpressed, and defines regions important for this behaviour. The phase separation of CBP is driven not by one particular region but by a combination of different regions which display differing phase behaviours, that together define the overall phase separating phenotype of CBP. In identifying CBP mutants that have differing phase separating abilities, I have developed a platform with which to investigate the biological relevance of phase separation. This toolkit provides the opportunity to understand how phase separation affects CBP HAT activity and enhancer function.

Bibliography

Agulnick, A D et al. (Nov. 1996). "Interactions of the LIM-domain-binding factor Ldb1 with LIM homeodomain proteins". en. In: *Nature* 384.6606, pp. 270–272.

Airenne, Kari J et al. (Apr. 2013). "Baculovirus: an Insect-derived Vector for Diverse Gene Transfer Applications". In: *Mol. Ther.* 21.4, pp. 739–749.

Alberti, Simon and Dorothee Dormann (Dec. 2019). "Liquid-Liquid Phase Separation in Disease". en. In: *Annu. Rev. Genet.* 53, pp. 171–194.

Alberti, Simon, Amy Gladfelter and Tanja Mittag (Jan. 2019). "Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates". en. In: *Cell* 176.3, pp. 419–434.

Alberti, Simon and Anthony A Hyman (Jan. 2021). "Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing". en. In: *Nat. Rev. Mol. Cell Biol.*

Alberti, Simon et al. (Apr. 2009). "A systematic survey identifies prions and illuminates sequence features of prionogenic proteins". en. In: *Cell* 137.1, pp. 146–158.

Alberti, Simon et al. (Nov. 2018). "A User's Guide for Phase Separation Assays with Purified Proteins". en. In: *J. Mol. Biol.* 430.23, pp. 4806–4820.

Alexander, Jeffrey M et al. (May 2019). "Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity". In: *Elife* 8, e41769.

An, Haiyan et al. (Jan. 2022). "ALS-linked cytoplasmic FUS assemblies are compositionally different from physiological stress granules and sequester hnRNPA3, a novel modifier of FUS toxicity". en. In: *Neurobiol. Dis.* 162, p. 105585.

Arnold, Mirjam et al. (Sept. 2021). "A BRD4-mediated elongation control point primes transcribing RNA polymerase II for 3'-processing and termination". en. In: *Mol. Cell* 81.17, 3589–3603.e13.

Arter, William E et al. (2022). "Biomolecular condensate phase diagrams with a combinatorial microdroplet platform". In: *Nat. Commun.* 13.1, p. 7845.

Arzideh, Seyyed Mohammad, Kamyar Movagharnejad and Mohsen Pirdashti (Aug. 2018). "Influence of the Temperature, Type of Salt, and Alcohol on Phase Diagrams of 2-Propanol + Inorganic Salt Aqueous Two-Phase Systems: Experimental Determination and Correlation". In: *J. Chem. Eng. Data* 63.8, pp. 2813–2824.

Babinchak, W Michael et al. (Apr. 2019). "The role of liquid-liquid phase separation in aggregation of the TDP-43 low-complexity domain". en. In: *J. Biol. Chem.* 294.16, pp. 6306–6317.

Banani, Salman F et al. (July 2016). "Compositional Control of Phase-Separated Cellular Bodies". en. In: *Cell* 166.3, pp. 651–663.

Banani, Salman F et al. (May 2017). "Biomolecular condensates: organizers of cellular biochemistry". en. In: *Nat. Rev. Mol. Cell Biol.* 18.5, pp. 285–298.

Banerji, J, S Rusconi and W Schaffner (Dec. 1981). "Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences". en. In: *Cell* 27.2 Pt 1, pp. 299–308.

Bannister, A J and T Kouzarides (Dec. 1996). "The CBP co-activator is a histone acetyltransferase". In: *Nature* 384.6610, pp. 641–643.

Barlev, N A et al. (Dec. 2001). "Acetylation of p53 activates transcription through recruitment of coactivators/histone acetyltransferases". en. In: *Mol. Cell* 8.6, pp. 1243–1254.

Baron, U, M Gossen and H Bujard (July 1997). "Tetracycline-controlled transcription in eukaryotes: novel transactivators with graded transactivation potential". en. In: *Nucleic Acids Res.* 25.14, pp. 2723–2729.

Basu, Shaon et al. (May 2020). "Unblending of Transcriptional Condensates in Human Repeat Expansion Disease". In: *Cell*.

Bayle, Vincent et al. (Mar. 2021). "Single-particle tracking photoactivated localization microscopy of membrane proteins in living plant tissues". en. In: *Nat. Protoc.* 16.3, pp. 1600–1628.

Bedford, David C et al. (Jan. 2010). "Target gene context influences the transcriptional requirement for the KAT3 family of CBP and p300 histone acetyltransferases". en. In: *Epigenetics* 5.1, pp. 9–15.

Berglund, Andrew and Hideo Mabuchi (Oct. 2005). "Tracking-FCS: Fluorescence correlation spectroscopy of individual particles". en. In: *Opt. Express* 13.20, pp. 8069–8082.

Berry, Joel et al. (Sept. 2015). "RNA transcription modulates phase transition-driven nuclear body assembly". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 112.38, E5237–45.

Birsa, Nicol et al. (Sept. 2020). "FUS-ALS mutants alter FMRP phase separation equilibrium and impair protein translation". en.

Blinka, Steven et al. (2017). "Identification of Transcribed Enhancers by Genome-Wide Chromatin Immunoprecipitation Sequencing". en. In: *Methods Mol. Biol.* 1468, pp. 91–109.

Boehning, Marc et al. (Sept. 2018). "RNA polymerase II clustering through carboxy-terminal domain phase separation". en. In: *Nat. Struct. Mol. Biol.* 25.9, pp. 833–840.

Boeynaems, Steven et al. (June 2018). "Protein Phase Separation: A New Phase in Cell Biology". en. In: *Trends Cell Biol.* 28.6, pp. 420–435.

Boeynaems, Steven et al. (Apr. 2019). "Spontaneous driving forces give rise to protein-RNA condensates with coexisting phases and complex material properties". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 116.16, pp. 7889–7898.

Boeynaems, Steven et al. (June 2023). "Poly(A)-binding protein is an ataxin-2 chaperone that regulates biomolecular condensates". en. In: *Mol. Cell*.

Boisvert, F M et al. (Mar. 2001). "The transcription coactivator CBP is a dynamic component of the promyelocytic leukemia nuclear body". en. In: *J. Cell Biol.* 152.5, pp. 1099–1106.

Bonev, Boyan et al. (Oct. 2017). "Multiscale 3D Genome Rewiring during Mouse Neural Development". en. In: *Cell* 171.3, 557–572.e24.

Borcherds, Wade et al. (Aug. 2020). "How do intrinsically disordered protein regions encode a driving force for liquid-liquid phase separation?" In: arXiv: `2008.02917 [q-bio.BM]`.

Bose, Daniel A et al. (Jan. 2017). "RNA Binding to CBP Stimulates Histone Acetylation and Transcription". en. In: *Cell* 168.1-2, 135–149.e22.

Boyko, Solomiia and Witold K Surewicz (Feb. 2022). "Tau liquid-liquid phase separation in neurodegenerative diseases". en. In: *Trends Cell Biol.*

Bracha, Dan, Mackenzie T Walls and Clifford P Brangwynne (Dec. 2019). "Probing and engineering liquid-phase organelles". en. In: *Nat. Biotechnol.* 37.12, pp. 1435–1445.

Bracha, Dan et al. (Nov. 2018). "Mapping Local and Global Liquid Phase Behavior in Living Cells Using Photo-Oligomerizable Seeds". en. In: *Cell* 175.6, 1467–1480.e13.

Brangwynne, Clifford P, Timothy J Mitchison and Anthony A Hyman (Mar. 2011). "Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 108.11, pp. 4334–4339.

Brangwynne, Clifford P et al. (June 2009). "Germline P granules are liquid droplets that localize by controlled dissolution/condensation". en. In: *Science* 324.5935, pp. 1729–1732.

Brown, Jonathan D et al. (Oct. 2014). "NF-κB directs dynamic super enhancer formation in inflammation and atherogenesis". en. In: *Mol. Cell* 56.2, pp. 219–231.

Bulger, M and M Groudine (Oct. 1999). "Looping versus linking: toward a model for long-distance gene activation". en. In: *Genes Dev.* 13.19, pp. 2465–2477.

Cai, Danfeng et al. (Dec. 2019). "Phase separation of YAP reorganizes genome topology for long-term YAP target gene expression". en. In: *Nat. Cell Biol.* 21.12, pp. 1578–1589.

Cardinaux, J R et al. (Mar. 2000). "Recruitment of CREB binding protein is sufficient for CREB-mediated gene activation". en. In: *Mol. Cell. Biol.* 20.5, pp. 1546–1552.

Carlson, Christopher R et al. (Nov. 2020). "Phosphoregulation of Phase Separation by the SARS-CoV-2 N Protein Suggests a Biophysical Basis for its Dual Functions". en. In: *Mol. Cell.*

Carullo, Nancy V N et al. (Sept. 2020). "Enhancer RNAs predict enhancer-gene regulatory links and are critical for enhancer function in neuronal systems". en. In: *Nucleic Acids Res.* 48.17, pp. 9550–9570.

Ceballos, Alfredo Vidal, Charles J McDonald and Shana Elbaum-Garfinkle (Nov. 2018). "Methods and Strategies to Quantify Phase Separation of Disordered Proteins". en. In: *Methods Enzymol.* 611, pp. 31–50.

Chan, H M and N B La Thangue (July 2001). "p300/CBP proteins: HATs for transcriptional bridges and scaffolds". en. In: *J. Cell Sci.* 114.Pt 13, pp. 2363–2373.

Chen, Qingjuan et al. (June 2022). "Histone acetyltransferases CBP/p300 in tumorigenesis and CBP/p300 inhibitors as promising novel anticancer agents". en. In: *Theranostics* 12.11, pp. 4935–4948.

Cheng, Jen-Hao et al. (July 2015). "Genome-wide analysis of enhancer RNA in gene regulation across 12 mouse tissues". en. In: *Sci. Rep.* 5, p. 12648.

Choi, Jeong-Mo, Furqan Dar and Rohit V Pappu (Oct. 2019). "LASSI: A lattice model for simulating phase transitions of multivalent proteins". en. In: *PLoS Comput. Biol.* 15.10, e1007028.

Choi, Jeong-Mo, Alex S Holehouse and Rohit V Pappu (Jan. 2020). "Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions". en. In: *Annu. Rev. Biophys.*

Chong, P Andrew and Julie D Forman-Kay (Dec. 2016). "Liquid-liquid phase separation in cellular signaling systems". en. In: *Curr. Opin. Struct. Biol.* 41, pp. 180–186.

Chong, P Andrew, Robert M Vernon and Julie D Forman-Kay (Nov. 2018). "RGG/RG Motif Regions in RNA Binding and Phase Separation". en. In: *J. Mol. Biol.* 430.23, pp. 4650–4665.

Chong, Shasha et al. (July 2018). "Imaging dynamic and selective low-complexity domain interactions that control gene transcription". In: *Science* 361.6400, eaar2555.

Chowdhury, Aritra, Daniel Nettels and Benjamin Schuler (May 2023). "Interaction Dynamics of Intrinsically Disordered Proteins from Single-Molecule Spectroscopy". In: *Annu. Rev. Biophys.* 52.1, pp. 433–462.

Claringbould, Annique and Judith B Zaugg (Nov. 2021). "Enhancers in disease: molecular basis and emerging treatment strategies". en. In: *Trends Mol. Med.* 27.11, pp. 1060–1073.

Clemson, Christine M et al. (Mar. 2009). "An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles". en. In: *Mol. Cell* 33.6, pp. 717–726.

Clontech (2013). *Retro-X Tet-On 3G Inducible Expression System User Manual.*

Cohan, Megan C et al. (Jan. 2022). "Uncovering Non-random Binary Patterns Within Sequences of Intrinsically Disordered Proteins". en. In: *J. Mol. Biol.* 434.2, p. 167373.

Conicella, Alexander E et al. (Sept. 2016). "ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain". en. In: *Structure* 24.9, pp. 1537–1549.

Conte, Mattia et al. (July 2020). "Polymer physics indicates chromatin folding variability across single-cells results from state degeneracy in phase separation". en. In: *Nat. Commun.* 11.1, p. 3289.

Creyghton, Menno P et al. (Dec. 2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 107.50, pp. 21931–21936.

Crick, Francis (Aug. 1970). "Central Dogma of Molecular Biology". en. In: *Nature* 227.5258, pp. 561–563.

Crump, Nicholas T et al. (Jan. 2021). "BET inhibition disrupts transcription but retains enhancer-promoter contact". en. In: *Nat. Commun.* 12.1, p. 223.

Dai, Gucan et al. (Aug. 2019). "The HCN channel voltage sensor undergoes a large downward motion during hyperpolarization". en. In: *Nat. Struct. Mol. Biol.* 26.8, pp. 686–694.

— (May 2021). "Electromechanical coupling mechanism for activation and inactivation of an HCN channel". en. In: *Nat. Commun.* 12.1, p. 2802.

Dang, Yongjun et al. (Oct. 2006). "Eukaryotic initiation factor 2alpha-independent pathway of stress granule induction by the natural product pateamine A". en. In: *J. Biol. Chem.* 281.43, pp. 32870–32878.

Das, Atze T, Liliane Tenenbaum and Ben Berkhout (2016). "Tet-On Systems For Doxycycline-inducible Gene Expression". en. In: *Curr. Gene Ther.* 16.3, pp. 156–167.

De Guzman, Roberto N et al. (Feb. 2006). "Structural basis for cooperative transcription factor binding to the CBP coactivator". In: *J. Mol. Biol.* 355.5, pp. 1005–1013.

De Santa, Francesca et al. (May 2010). "A large fraction of extragenic RNA pol II transcription sites overlap enhancers". en. In: *PLoS Biol.* 8.5, e1000384.

Decker, Carolyn J et al. (Mar. 2022). "RNA is required for the integrity of multiple nuclear and cytoplasmic membrane-less RNP granules". en. In: *EMBO J.*, e110137.

Delvecchio, Manuela et al. (Aug. 2013). "Structure of the p300 catalytic core and implications for chromatin targeting and HAT regulation". In: *Nat. Struct. Mol. Biol.*

Demmerle, Justin, Siyuan Hao and Danfeng Cai (Dec. 2023). "Transcriptional condensates and phase separation: condensing information across scales and mechanisms". en. In: *Nucleus* 14.1, p. 2213551.

Dewey, Colleen M et al. (Mar. 2011). "TDP-43 is directed to stress granules by sorbitol, a novel physiological osmotic and oxidative stressor". en. In: *Mol. Cell. Biol.* 31.5, pp. 1098–1108.

Di Tommaso, Paolo et al. (Apr. 2017). "Nextflow enables reproducible computational workflows". en. In: *Nat. Biotechnol.* 35.4, pp. 316–319.

Ding, Mengting et al. (June 2018). "Enhancer RNAs (eRNAs): New Insights into Gene Transcription and Disease Treatment". en. In: *J. Cancer* 9.13, pp. 2334–2340.

Dixon, Jesse R, David U Gorkin and Bing Ren (June 2016). "Chromatin Domains: The Unit of Chromosome Organization". en. In: *Mol. Cell* 62.5, pp. 668–680.

Dobrucki, Jurek W, Dorota Feret and Anna Noatynska (Sept. 2007). "Scattering of exciting light by live cells in fluorescence confocal imaging: phototoxic effects and relevance for FRAP studies". en. In: *Biophys. J.* 93.5, pp. 1778–1786.

Dorsett, D (Oct. 1999). "Distant liaisons: long-range enhancer-promoter interactions in Drosophila". en. In: *Curr. Opin. Genet. Dev.* 9.5, pp. 505–514.

Dosztányi, Zsuzsanna (Jan. 2018). "Prediction of protein disorder based on IUPred". en. In: *Protein Sci.* 27.1, pp. 331–340.

Dosztányi, Zsuzsanna, Bálint Mészáros and István Simon (Oct. 2009). "ANCHOR: web server for predicting protein binding regions in disordered proteins". en. In: *Bioinformatics* 25.20, pp. 2745–2746.

Dougherty, Dennis A (Jan. 1996). "Cation-π Interactions in Chemistry and Biology: A New View of Benzene, Phe, Tyr, and Trp". In: *Science* 271.5246, pp. 163–168.

Düster, Robert et al. (Jan. 2021). "1,6-Hexanediol, commonly used to dissolve liquid-liquid phase separated condensates, directly impairs kinase and phosphatase activities". en. In: *J. Biol. Chem.* 296, p. 100260.

Dyson, H Jane and Peter E Wright (Mar. 2016). "Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-binding Protein (CBP) and p300". en. In: *J. Biol. Chem.* 291.13, pp. 6714–6722.

Eckner, R et al. (1994). "The adenovirus E1A-associated 300-kD protein exhibits properties of a transcriptional coactivator and belongs to an evolutionarily conserved family". en. In: *Cold Spring Harb. Symp. Quant. Biol.* 59, pp. 85–95.

ENCODE Project Consortium (Sept. 2012). "An integrated encyclopedia of DNA elements in the human genome". en. In: *Nature* 489.7414, pp. 57–74.

Erdel, Fabian and Karsten Rippe (May 2018). "Formation of Chromatin Subcompartments by Phase Separation". en. In: *Biophys. J.* 114.10, pp. 2262–2270.

Felsenfeld, Gary and Mark Groudine (Jan. 2003). "Controlling the double helix". en. In: *Nature* 421.6921, pp. 448–453.

Ferreon, Josephine C et al. (May 2018). "Acetylation Disfavors Tau Phase Separation". en. In: *Int. J. Mol. Sci.* 19.5, p. 1360.

Freibaum, Brian D et al. (Sept. 2020). "High fidelity reconstitution of stress granules and nucleoli in mammalian cellular lysate". en.

Furlong, Eileen E M and Michael Levine (Sept. 2018). "Developmental enhancers and chromosome topology". en. In: *Science* 361.6409, pp. 1341–1345.

Galganski, Lukasz, Martyna O Urbanek and Wlodzimierz J Krzyzosiak (Oct. 2017). "Nuclear speckles: molecular organization, biological function and role in disease". en. In: *Nucleic Acids Res.* 45.18, pp. 10350–10368.

Gao, Guoming and Nils G Walter (Sept. 2023). "Critical Assessment of Condensate Boundaries in Dual-Color Single Particle Tracking". en. In: *J. Phys. Chem. B.*

Garaizar, Adiran and Jorge R Espinosa (Sept. 2021). "Salt dependent phase behavior of intrinsically disordered proteins from a coarse-grained model with explicit water and ions". en. In: *J. Chem. Phys.* 155.12, p. 125103.

Garcia-Jove Navarro, Marina et al. (July 2019). "RNA is a critical element for the sizing and the composition of phase-separated RNA–protein condensates". In: *Nat. Commun.* 10.1, p. 3230.

Geiger, Florian et al. (Nov. 2021). "Liquid-liquid phase separation underpins the formation of replication factories in rotaviruses". en. In: *EMBO J.* 40.21, e107711.

Gelder, Katie and Daniel Bose (Oct. 2023). "Enhancers, phase separation and the RNA polymerase II transfer model". en. In: *Bioessays* 45.10, e2300128.

Gibson, B A et al. (Jan. 2019a). "Organization and Regulation of Chromatin by Liquid-Liquid Phase Separation". en.

Gibson, Bryan A et al. (Oct. 2019b). "Organization of Chromatin by Intrinsic and Regulated Phase Separation". en. In: *Cell* 179.2, 470–484.e21.

Gilks, Natalie et al. (Dec. 2004). "Stress granule assembly is mediated by prion-like aggregation of TIA-1". en. In: *Mol. Biol. Cell* 15.12, pp. 5383–5398.

Giotopoulos, G et al. (Jan. 2016). "The epigenetic regulators CBP and p300 facilitate leukemogenesis and represent therapeutic targets in acute myeloid leukemia". en. In: *Oncogene* 35.3, pp. 279–289.

Gossen, M et al. (June 1995). “Transcriptional activation by tetracyclines in mammalian cells”. en. In: *Science* 268.5218, pp. 1766–1769.

Grimm, Jonathan B et al. (Aug. 2016). “Bright photoactivatable fluorophores for single-molecule imaging”. en.

Guenther, Elizabeth L et al. (June 2018). “Atomic structures of TDP-43 LCD segments and insights into reversible or pathogenic aggregation”. en. In: *Nat. Struct. Mol. Biol.* 25.6, pp. 463–471.

Gurumurthy, Aishwarya et al. (Jan. 2019). “Phase Separation and Transcription Regulation: Are Super-Enhancers and Locus Control Regions Primary Sites of Transcription Complex Assembly?” en. In: *Bioessays* 41.1, e1800164.

Gurumurthy, Aishwarya et al. (Feb. 2021). “Super-enhancer mediated regulation of adult β-globin gene expression: the role of eRNA and Integrator”. en. In: *Nucleic Acids Res.* 49.3, pp. 1383–1396.

Guthmann, Manuel, Adam Burton and Maria-Elena Torres-Padilla (Dec. 2019). “Expression and phase separation potential of heterochromatin proteins during early mouse development”. en. In: *EMBO Rep.* 20.12, e47952.

Haag, Rainer (May 2015). “Multivalency as a chemical organization and action principle”. en. In: *Beilstein J. Org. Chem.* 11, pp. 848–849.

Hallegger, Martina et al. (Aug. 2021). “TDP-43 condensation properties specify its RNA-binding and regulatory repertoire”. In: *Cell*.

Hamad, Nesreen et al. (July 2020). “Direct visualization of the conformational change of FUS/TLS upon binding to promoter-associated non-coding RNA”. en. In: *Chem. Commun.*

Han, Xinye et al. (Apr. 2020). “Roles of the BRD4 short isoform in phase separation and active gene transcription”. en. In: *Nat. Struct. Mol. Biol.* 27.4, pp. 333–341.

Harrison, Alice Ford and James Shorter (Apr. 2017). “RNA-binding proteins with prion-like domains in health and disease”. en. In: *Biochem. J* 474.8, pp. 1417–1438.

Harrison, Laura J and Daniel Bose (Aug. 2022). “Enhancer RNAs step forward: new insights into enhancer function”. en. In: *Development* 149.16.

Harrow, Jennifer et al. (Sept. 2012). “GENCODE: the reference human genome annotation for The ENCODE Project”. en. In: *Genome Res.* 22.9, pp. 1760–1774.

Hebbes, Tim R, Alan W Thorne and C Crane-Robinson (1988). “A direct link between core histone acetylation and transcriptionally active chromatin”. In: *Plan. Perspect.* 1395, p. 1402.

Heckert, Alec et al. (Sept. 2022). “Recovering mixtures of fast-diffusing states from short single-particle trajectories”. en. In: *Elife* 11, e70169.

Heinz, Sven et al. (Mar. 2015). "The selection and function of cell type-specific enhancers". en. In: *Nat. Rev. Mol. Cell Biol.* 16.3, pp. 144–154.

Hellenkamp, Björn et al. (Sept. 2018). "Precision and accuracy of single-molecule FRET measurements-a multi-laboratory benchmark study". en. In: *Nat. Methods* 15.9, pp. 669–676.

Henninger, Jonathan E et al. (Dec. 2020). "RNA-Mediated Feedback Control of Transcriptional Condensates". In: *Cell*.

Henry, Ryan A, Yin-Ming Kuo and Andrew J Andrews (Aug. 2013). "Differences in specificity and selectivity between CBP and p300 acetylation of histone H3 and H3/H4". en. In: *Biochemistry* 52.34, pp. 5746–5759.

Hernández-Candia, Carmen N, Sarah Pearce and Chandra L Tucker (Mar. 2021). "A modular tool to query and inducibly disrupt biomolecular condensates". en. In: *Nat. Commun.* 12.1, pp. 1–13.

Hirose, Tetsuro et al. (Apr. 2023). "A guide to membraneless organelles and their various roles in gene regulation". en. In: *Nat. Rev. Mol. Cell Biol.* 24.4, pp. 288–304.

Hnisz, Denes et al. (Nov. 2013). "Super-enhancers in the control of cell identity and disease". en. In: *Cell* 155.4, pp. 934–947.

Hochreiter, Bernhard, Alan Pardo Garcia and Johannes A Schmid (Oct. 2015). "Fluorescent proteins as genetically encoded FRET biosensors in life sciences". en. In: *Sensors* 15.10, pp. 26281–26314.

Hofweber, Mario and Dorothee Dormann (May 2019). "Friend or foe-Post-translational modifications as regulators of phase separation and RNP granule dynamics". en. In: *J. Biol. Chem.* 294.18, pp. 7137–7150.

Hofweber, Mario et al. (Apr. 2018). "Phase Separation of FUS Is Suppressed by Its Nuclear Import Receptor and Arginine Methylation". en. In: *Cell* 173.3, 706–719.e13.

Hohlbein, Johannes, Timothy D Craggs and Thorben Cordes (Feb. 2014). "Alternating-laser excitation: single-molecule FRET and beyond". en. In: *Chem. Soc. Rev.* 43.4, pp. 1156–1171.

Holehouse, Alex S et al. (Jan. 2017). "CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins". en. In: *Biophys. J.* 112.1, pp. 16–21.

Holmqvist, Per-Henrik and Mattias Mannervik (Jan. 2013). "Genomic occupancy of the transcriptional co-activators p300 and CBP". In: *Transcription* 4.1, pp. 18–23.

Hsieh, Chen-Lin et al. (May 2014). "Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 111.20, pp. 7319–7324.

Hunter, Christopher A and Jeremy K M Sanders (July 1990). "The nature of .pi.-.pi. interactions". In: *J. Am. Chem. Soc.* 112.14, pp. 5525–5534.

Hyman, Anthony A, Christoph A Weber and Frank Jülicher (2014). "Liquid-liquid phase separation in biology". en. In: *Annu. Rev. Cell Dev. Biol.* 30, pp. 39–58.

Ide, Satoru et al. (Oct. 2020). "Transcriptional suppression of ribosomal DNA with phase separation". en. In: *Science Advances* 6.42, eabb5953.

Itoh, Yuji et al. (Apr. 2021). "1,6-hexanediol rapidly immobilizes and condenses chromatin in living human cells". en. In: *Life Sci. Alliance* 4.4, e202001005.

Jin, Qihuang et al. (Jan. 2011). "Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation". en. In: *EMBO J.* 30.2, pp. 249–262.

Jones, David T and Domenico Cozzetto (Mar. 2015). "DISOPRED3: precise disordered region predictions with annotated protein-binding activity". en. In: *Bioinformatics* 31.6, pp. 857–863.

Jumper, John et al. (Aug. 2021). "Highly accurate protein structure prediction with AlphaFold". en. In: *Nature* 596.7873, pp. 583–589.

Kalkhoven, Eric et al. (Feb. 2003). "Loss of CBP acetyltransferase activity by PHD finger mutations in Rubinstein-Taybi syndrome". In: *Hum. Mol. Genet.* 12.4, pp. 441–450.

Kar, Mrityunjoy et al. (Feb. 2022). "Phase separating RNA binding proteins form heterogeneous distributions of clusters in subsaturated solutions". en.

Kato, Masato et al. (May 2012). "Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels". en. In: *Cell* 149.4, pp. 753–767.

Katrukha, E (2020). *Detection of Molecules (DoM) plugin for ImageJ.*

Kawasaki, H et al. (May 1998). "Distinct roles of the co-activators p300 and CBP in retinoic-acid-induced F9-cell differentiation". en. In: *Nature* 393.6682, pp. 284–289.

Kent, Samantha et al. (Oct. 2020). "Phase-Separated Transcriptional Condensates Accelerate Target-Search Process Revealed by Live-Cell Single-Molecule Imaging". en. In: *Cell Rep.* 33.2, p. 108248.

Kent, W James et al. (June 2002). "The human genome browser at UCSC". en. In: *Genome Res.* 12.6, pp. 996–1006.

Keppel, Theodore R, Brent A Howard and David D Weis (Oct. 2011). "Mapping unstructured regions and synergistic folding in intrinsically disordered proteins with amide H/D exchange mass spectrometry". In: *Biochemistry* 50.40, pp. 8722–8732.

Kim, Tae-Kyung et al. (May 2010). "Widespread transcription at neuronal activity-regulated enhancers". In: *Nature* 465.7295, pp. 182–187.

Koehler, Christine et al. (Nov. 2020). "Inducible genetic code expansion in eukaryotes". en. In: *Chembiochem* 21.22, pp. 3216–3219.

König, Iwo et al. (May 2021). "Impact of In-Cell and In-Vitro Crowding on the Conformations and Dynamics of an Intrinsically Disordered Protein". en. In: *Angew. Chem. Int. Ed Engl.* 60.19, pp. 10724–10729.

Kornberg, Roger D (May 1974). "Chromatin Structure: A Repeating Unit of Histones and DNA". In: *Science* 184.4139, pp. 868–871.

Kroschwald, Sonja, Shovamayee Maharana and Alberti Simon (May 2017). "Hexanediol: a chemical probe to investigate the material properties of membrane-less compartments". In: *Matters*.

Kung, A L et al. (Feb. 2000). "Gene dose-dependent control of hematopoiesis and hematologic tumor suppression by CBP". en. In: *Genes Dev.* 14.3, pp. 272–277.

Lackner, Rachel M et al. (Aug. 2022). "A General Strategy for the Design and Evaluation of Heterobifunctional Tools: Applications to Protein Localization and Phase Separation". en. In: *Chembiochem* 23.16, e202200209.

LaMorte, V J et al. (Apr. 1998). "Localization of nascent RNA and CREB binding protein with the PML-containing nuclear body". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 95.9, pp. 4991–4996.

Langlais, Claudia et al. (Oct. 2007). "A systematic approach for testing expression of human full-length proteins in cell-free expression systems". en. In: *BMC Biotechnol.* 7, p. 64.

Larson, Adam G et al. (July 2017). "Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin". en. In: *Nature* 547.7662, pp. 236–240.

Latt, S A, t D S Auld and B L Vallee (1972). "Distance Measurements at the Active Site of Carboxypeptidase A during Catalysist". In: *Biochemistry* 11.16, pp. 3015–3022.

Lee, D Y et al. (Jan. 1993). "A positive role for histone acetylation in transcription factor access to nucleosomal DNA". en. In: *Cell* 72.1, pp. 73–84.

Lee, Flora C Y et al. (Aug. 2021a). "An improved iCLIP protocol". en.

Lee, Joo-Hyung et al. (Aug. 2021b). "Enhancer RNA m6A methylation facilitates transcriptional condensate formation and gene activation". In: *Mol. Cell*.

Lee, Robin van der et al. (July 2014). "Classification of Intrinsically Disordered Regions and Proteins". In: *Chem. Rev.* 114.13, pp. 6589–6631.

Lesch, H P et al. (Sept. 2008). "Generation of lentivirus vectors using recombinant baculoviruses". en. In: *Gene Ther.* 15.18, pp. 1280–1286.

Levone, Brunno R et al. (Aug. 2020). "FUS-dependent liquid-liquid phase separation is an early event in double-strand break repair". en.

Li, Heng and Richard Durbin (July 2009). "Fast and accurate short read alignment with Burrows-Wheeler transform". en. In: *Bioinformatics* 25.14, pp. 1754–1760.

Li, Wenbo, Dimple Notani and Michael G Rosenfeld (Apr. 2016). "Enhancers as non-coding RNA transcription units: recent insights and future perspectives". en. In: *Nat. Rev. Genet.* 17.4, pp. 207–223.

Lin, Yi-Hsuan and Hue Sun Chan (May 2017). "Phase Separation and Single-Chain Compactness of Charged Disordered Proteins Are Strongly Correlated". en. In: *Biophys. J.* 112.10, pp. 2043–2046.

Lin, Yuan, Simon L Currie and Michael K Rosen (Nov. 2017). "Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs". en. In: *J. Biol. Chem.* 292.46, pp. 19110–19120.

Lin, Yuan et al. (Oct. 2015). "Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins". en. In: *Mol. Cell* 60.2, pp. 208–219.

Liu, Shu-Su et al. (July 2015). "Structural plasticity of green fluorescent protein to amino acid deletions and fluorescence rescue by folding-enhancing mutations". en. In: *BMC Biochem.* 16, p. 17.

Liu, Xin et al. (Feb. 2008). "The structural basis of protein acetylation by the p300/CBP transcriptional coactivator". en. In: *Nature* 451.7180, pp. 846–850.

Liu, Xinyi et al. (Aug. 2021). "Time-dependent effect of 1,6-hexanediol on biomolecular condensates and 3D chromatin organization". en. In: *Genome Biol.* 22.1, p. 230.

Liu, Yongyang et al. (Oct. 2022). "Single-particle fluorescence tracking combined with TrackMate assay reveals highly heterogeneous and discontinuous lysosomal transport in freely orientated axons". en. In: *Biotechnol. J.* 17.10, e2200006.

Long, Hannah K, Sara L Prescott and Joanna Wysocka (Nov. 2016). "Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution". en. In: *Cell* 167.5, pp. 1170–1187.

Longfils, M et al. (Mar. 2018). "Single particle raster image analysis of diffusion for particle mixtures". en. In: *J. Microsc.* 269.3, pp. 269–281.

Lotthammer, Jeffrey M et al. (May 2023). "Direct Prediction of Intrinsically Disordered Protein Conformational Properties From Sequence". en.

Love, Andrew J et al. (June 2017). "Cajal bodies and their role in plant stress and disease responses". en. In: *RNA Biol.* 14.6, pp. 779–790.

Luger, Karolin (Apr. 2003). "Structure and dynamic behavior of nucleosomes". en. In: *Curr. Opin. Genet. Dev.* 13.2, pp. 127–135.

Luo, Feng et al. (Apr. 2018). "Atomic structures of FUS LC domain segments reveal bases for reversible amyloid fibril formation". en. In: *Nat. Struct. Mol. Biol.* 25.4, pp. 341–346.

Ma, Jennifer C and Dennis A Dougherty (Aug. 1997). "The Cationminus signpi Interaction". en. In: *Chem. Rev.* 97.5, pp. 1303–1324.

Ma, Liang et al. (Feb. 2021). "Co-condensation between transcription factor and coactivator p300 modulates transcriptional bursting kinetics". en. In: *Mol. Cell*.

Maharana, Shovamayee et al. (May 2018). "RNA buffers the phase separation behavior of prion-like RNA binding proteins". en. In: *Science* 360.6391, pp. 918–921.

Manickavinayaham, Swarnalatha et al. (Oct. 2019). "E2F1 acetylation directs p300/CBP-mediated histone acetylation at DNA double-strand breaks to facilitate repair". en. In: *Nat. Commun.* 10.1, p. 4951.

Mansour, Marc R et al. (Dec. 2014). "Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element". en. In: *Science* 346.6215, pp. 1373–1377.

Margineanu, Anca et al. (June 2016). "Screening for protein-protein interactions using Förster resonance energy transfer (FRET) and fluorescence lifetime imaging microscopy (FLIM)". en. In: *Sci. Rep.* 6, p. 28186.

Martin, Erik W and Alex S Holehouse (Oct. 2020). "Intrinsically disordered protein regions and phase separation: sequence determinants of assembly or lack thereof". en. In: *Emerg Top Life Sci*.

Martin, Erik W et al. (Feb. 2020). "Valence and patterning of aromatic residues determine the phase behavior of prion-like domains". en. In: *Science* 367.6478, pp. 694–699.

May, Dalit et al. (Dec. 2011). "Large-scale discovery of enhancers from human heart tissue". en. In: *Nat. Genet.* 44.1, pp. 89–93.

McGurk, Leeanne et al. (Sept. 2018). "Poly(ADP-Ribose) Prevents Pathological Phase Separation of TDP-43 by Promoting Liquid Demixing and Stress Granule Localization". en. In: *Mol. Cell* 71.5, 703–717.e9.

McSwiggen, David T et al. (2019). *Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences*.

Mekonnen, Gemechu et al. (Oct. 2023). "Advanced imaging techniques for studying protein phase separation in living cells and at single-molecule level". In: *Curr. Opin. Chem. Biol.* 76, p. 102371.

Mensah, Martin A et al. (Feb. 2023). "Aberrant phase separation and nucleolar dysfunction in rare genetic diseases". en. In: *Nature* 614.7948, pp. 564–571.

Merlo, Rosa et al. (Dec. 2021). "The SNAP-tag technology revised: an effective chemo-enzymatic approach by using a universal azide-based substrate". en. In: *J. Enzyme Inhib. Med. Chem.* 36.1, pp. 85–97.

Millán-Zambrano, Gonzalo et al. (Sept. 2022). "Histone post-translational modifications - cause and consequence of genome function". en. In: *Nat. Rev. Genet.* 23.9, pp. 563–580.

Miné-Hattab, Judith et al. (Feb. 2021). "Single molecule microscopy reveals key physical features of repair foci in living cells". In: *Elife* 10, e60577.

Mitrea, Diana M et al. (Feb. 2016). "Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA". en. In: *Elife* 5.

Mitrea, Diana M et al. (Feb. 2018). "Self-interaction of NPM1 modulates multiple mechanisms of liquid-liquid phase separation". en. In: *Nat. Commun.* 9.1, p. 842.

Mitrea, Diana M et al. (Aug. 2022). "Modulating biomolecular condensates: a novel approach to drug discovery". en. In: *Nat. Rev. Drug Discov.* 21.11, pp. 841–862.

Mittag, Tanja and Rohit V Pappu (June 2022). "A conceptual framework for understanding phase separation and addressing open questions and challenges". en. In: *Mol. Cell* 82.12, pp. 2201–2214.

Mittag, Tanja and Roy Parker (Nov. 2018). "Multiple Modes of Protein-Protein Interactions Promote RNP Granule Assembly". en. In: *J. Mol. Biol.* 430.23, pp. 4636–4649.

Mochida, Kyo and Manabu Gomyoda (Jan. 1987). "Toxicity of ethylene glycol, diethylene glycol, and propylene glycol to human cells in culture". In: *Bull. Environ. Contam. Toxicol.* 38.1, pp. 151–153.

Molčanov, Krešimir and Biserka Kojić-Prodić (Mar. 2019). "Towards understanding π-stacking interactions between non-aromatic rings". en. In: *IUCrJ* 6.Pt 2, pp. 156–166.

Molliex, Amandine et al. (Sept. 2015). "Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization". en. In: *Cell* 163.1, pp. 123–133.

Monahan, Zachary et al. (Oct. 2017). "Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity". en. In: *EMBO J.* 36.20, pp. 2951–2967.

Morcillo, P et al. (Oct. 1997). "Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in Drosophila". en. In: *Genes Dev.* 11.20, pp. 2729–2740.

Morrison, Olivia and Jitendra Thakur (June 2021). "Molecular Complexes at Euchromatin, Heterochromatin and Centromeric Chromatin". en. In: *Int. J. Mol. Sci.* 22.13.

Mortensen, Jonas S and Claus J Loland (2020). "Short-Range Distance Measurement by Transition Metal Ion FRET". en. In: *Methods Mol. Biol.* 2168, pp. 299–311.

Mousavi, Kambiz et al. (Sept. 2013). "eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci". en. In: *Mol. Cell* 51.5, pp. 606–617.

Murakami, Tetsuro et al. (Nov. 2015). "ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function". en. In: *Neuron* 88.4, pp. 678–690.

Murchie, A I et al. (Oct. 1989). "Fluorescence energy transfer shows that the four-way DNA junction is a right-handed cross of antiparallel molecules". en. In: *Nature* 341.6244, pp. 763–766.

Murray, Dylan T et al. (Oct. 2017). "Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains". en. In: *Cell* 171.3, 615–627.e16.

Musacchio, Andrea (Mar. 2022). "On the role of phase separation in the biogenesis of membraneless compartments". en. In: *EMBO J.* 41.5, e109952.

Muzzopappa, Fernando et al. (Dec. 2022). "Detecting and quantifying liquid–liquid phase separation in living cells by model-free calibrated half-bleaching". en. In: *Nat. Commun.* 13.1, pp. 1–15.

Nair, Sreejith J et al. (Mar. 2019). "Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly". In: *Nat. Struct. Mol. Biol.* 26.3, pp. 193–203.

Narita, Takeo et al. (Mar. 2021). "Enhancers are activated by p300/CBP activity-dependent PIC assembly, RNAPII recruitment, and pause release". In: *Mol. Cell*.

Nasir, Irem et al. (Oct. 2019). "Single-molecule fluorescence studies of intrinsically disordered proteins and liquid phase separation". In: *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1867.10, pp. 980–987.

Niaki, Amirhossein Ghanbari et al. (Oct. 2019). "Loss of Dynamic RNA Interaction and Aberrant Phase Separation Induced by Two Distinct Types of ALS/FTD-Linked FUS Mutations". In: *Mol. Cell*.

Nikić, Ivana et al. (May 2015). "Labeling proteins on live mammalian cells using click chemistry". en. In: *Nat. Protoc.* 10.5, pp. 780–791.

Nikić, Ivana et al. (Dec. 2016). "Debugging eukaryotic genetic code expansion for site-specific click-PAINT super-resolution microscopy". en. In: *Angew. Chem. Int. Ed Engl.* 55.52, pp. 16172–16176.

Ninomiya, Kensuke et al. (Feb. 2020). "LncRNA-dependent nuclear stress bodies promote intron retention through SR protein phosphorylation". en. In: *EMBO J.* 39.3, e102729.

Nora, Elphège P et al. (Apr. 2012). "Spatial partitioning of the regulatory landscape of the X-inactivation centre". en. In: *Nature* 485.7398, pp. 381–385.

Nora, Elphège P et al. (May 2017). "Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization". en. In: *Cell* 169.5, 930–944.e22.

Nott, Timothy J, Timothy D Craggs and Andrew J Baldwin (June 2016). "Membraneless organelles can melt nucleic acid duplexes and act as biomolecular filters". In: *Nat. Chem.* 8.6, pp. 569–575.

Nott, Timothy J et al. (Mar. 2015). "Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles". In: *Mol. Cell* 57.5, pp. 936–947.

Nucifora Jr, F C et al. (Mar. 2001). "Interference by huntingtin and atrophin-1 with cbp-mediated transcription leading to cellular toxicity". en. In: *Science* 291.5512, pp. 2423–2428.

Ogryzko, V V et al. (Nov. 1996). "The transcriptional coactivators p300 and CBP are histone acetyltransferases". en. In: *Cell* 87.5, pp. 953–959.

Oldfield, Christopher J and A Keith Dunker (Mar. 2014). "Intrinsically disordered proteins and intrinsically disordered protein regions". en. In: *Annu. Rev. Biochem.* 83, pp. 553–584.

Oldfield, Christopher J et al. (Feb. 2005). "Comparing and combining predictors of mostly disordered proteins". en. In: *Biochemistry* 44.6, pp. 1989–2000.

Ortega, Esther et al. (Oct. 2018). "Transcription factor dimerization activates the p300 acetyltransferase". en. In: *Nature* 562.7728, pp. 538–544.

Owen, Izzy and Frank Shewmaker (Nov. 2019). "The Role of Post-Translational Modifications in the Phase Transitions of Intrinsically Disordered Proteins". en. In: *Int. J. Mol. Sci.* 20.21.

Pak, Chi W et al. (July 2016). "Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein". en. In: *Mol. Cell* 63.1, pp. 72–85.

Park, Sangho et al. (July 2017). "Role of the CBP catalytic core in intramolecular SUMOylation and control of histone H3 acetylation". In: *Proceedings of the National Academy of Sciences* 114.27, E5335–E5342.

Patel, Avinash et al. (Aug. 2015). "A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation". en. In: *Cell* 162.5, pp. 1066–1077.

Pederson, Thoru (Mar. 2011). "The nucleolus". en. In: *Cold Spring Harb. Perspect. Biol.* 3.3.

Peng, A and Stephanie C Weber (Nov. 2019). "Evidence for and against Liquid-Liquid Phase Separation in the Nucleus". en. In: *Non-Coding RNA* 5.4, p. 50.

Peng, Zhenling and Lukasz Kurgan (Oct. 2015). "High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder". en. In: *Nucleic Acids Res.* 43.18, e121.

Pessina, Fabio et al. (Sept. 2019). "Functional transcription promoters at DNA double-strand breaks mediate RNA-driven phase separation of damage-response factors". en. In: *Nat. Cell Biol.*

Piroska, Leonard et al. (Aug. 2023). "α-Synuclein liquid condensates fuel fibrillar α-synuclein growth". en. In: *Sci Adv* 9.33, eadg5663.

Plank, Jennifer L and Ann Dean (July 2014). "Enhancer function: mechanistic and genome-wide insights come together". en. In: *Mol. Cell* 55.1, pp. 5–14.

Plys, Aaron J and Robert E Kingston (July 2018). "Dynamic condensates activate transcription". In: *Science* 361.6400, pp. 329–330.

Popay, Tessa M and Jesse R Dixon (Aug. 2022). "Coming full circle: On the origin and evolution of the looping model for enhancer–promoter communication". en. In: *J. Biol. Chem.* 298.8.

Posit team (2023). *RStudio: Integrated Development Environment for R.*

Protter, David S W et al. (Feb. 2018). "Intrinsically Disordered Regions Can Contribute Promiscuous Interactions to RNP Granule Assembly". en. In: *Cell Rep.* 22.6, pp. 1401–1412.

Ptitsyn, O B et al. (Mar. 1990). "Evidence for a molten globule state as a general intermediate in protein folding". en. In: *FEBS Lett.* 262.1, pp. 20–24.

Qamar, Seema et al. (Apr. 2018). "FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-π Interactions". en. In: *Cell* 173.3, 720–734.e15.

R Core Team (2020). *R: A Language and Environment for Statistical Computing.*

Rani, Pooja, Anupaul Baruah and Parbati Biswas (Oct. 2014). "Does lack of secondary structure imply intrinsic disorder in proteins? A sequence analysis". en. In: *Biochim. Biophys. Acta* 1844.10, pp. 1827–1834.

Rao, Suhas S P et al. (Oct. 2017). "Cohesin Loss Eliminates All Loop Domains". en. In: *Cell* 171.2, 305–320.e24.

Redfield, C, R A Smith and C M Dobson (Jan. 1994). "Structural characterization of a highly-ordered 'molten globule' at low pH". en. In: *Nat. Struct. Biol.* 1.1, pp. 23–29.

Rhine, Kevin et al. (Apr. 2020). "ALS/FTLD-Linked Mutations in Glycine Residues of FUS Lead to Immiscibility with Wild-Type FUS".

Riback, Joshua A et al. (Mar. 2017). "Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response". en. In: *Cell* 168.6, 1028–1040.e19.

Riback, Joshua A et al. (May 2020). "Composition-dependent thermodynamics of intracellular phase separation". In: *Nature.*

Richmond, T A et al. (Feb. 2000). "Engineered metal binding sites on green fluorescence protein". en. In: *Biochem. Biophys. Res. Commun.* 268.2, pp. 462–465.

Richter, William F et al. (Nov. 2022). "The Mediator complex as a master regulator of transcription by RNA polymerase II". en. In: *Nat. Rev. Mol. Cell Biol.* 23.11, pp. 732–749.

Rippe, Karsten (Oct. 2007). "Dynamic organization of the cell nucleus". en. In: *Curr. Opin. Genet. Dev.* 17.5, pp. 373–380.

Roe, Jae-Seok et al. (June 2015). "BET Bromodomain Inhibition Suppresses the Function of Hematopoietic Transcription Factors in Acute Myeloid Leukemia". en. In: *Mol. Cell* 58.6, pp. 1028–1039.

Rosencrance, Celeste D et al. (Apr. 2020). "Chromatin Hyperacetylation Impacts Chromosome Folding by Forming a Nuclear Subcompartment". en. In: *Mol. Cell* 78.1, 112–126.e12.

Rossow, Molly J et al. (Nov. 2010). "Raster image correlation spectroscopy in live cells". en. In: *Nat. Protoc.* 5.11, pp. 1761–1774.

Roy, Rahul, Sungchul Hohng and Taekjip Ha (June 2008). "A practical guide to single-molecule FRET". en. In: *Nat. Methods* 5.6, pp. 507–516.

Ruthenburg, Alexander J et al. (Dec. 2007). "Multivalent engagement of chromatin modifications by linked binding modules". In: *Nat. Rev. Mol. Cell Biol.* 8.12, pp. 983–994.

S. A. Latt, D S Auld and B L Vallee (Nov. 1970). "Surveyor Substrates: Energy-Transfer Gauges of Active Center Topography during Catalysis". In: *Proc. Natl. Acad. Sci. U. S. A.* 67.3, pp. 1383–1389.

Saar, Kadi L et al. (2021). "Learning the molecular grammar of protein condensates from sequence determinants and embeddings". In: *Proceedings of the National Academy of Sciences* 118.15, e2019053118.

Sabari, Benjamin R, Alessandra Dall'Agnese and Richard A Young (July 2020). "Biomolecular Condensates in the Nucleus". en. In: *Trends Biochem. Sci.*

Sabari, Benjamin R et al. (July 2018). "Coactivator condensation at super-enhancers links phase separation and gene control". In: *Science* 361.6400, eaar3958.

Saito, Makoto et al. (Jan. 2019). "Acetylation of intrinsically disordered regions regulates phase separation". en. In: *Nat. Chem. Biol.* 15.1, pp. 51–61.

San Diego, California Usa (Aug. 2022). *GraphPad Prism 9.0 Software*.

Sandtner, Walter, Francisco Bezanilla and Ana M Correa (Nov. 2007). "In vivo measurement of intramolecular distances using genetically encoded reporters". en. In: *Biophys. J.* 93.9, pp. L45–7.

Sanulli, S et al. (Oct. 2019). "HP1 reshapes nucleosome core to promote heterochromatin phase separation". en. In: *Nature*.

Sanulli, Serena and Geeta J Narlikar (May 2021). "Generation and Biochemical Characterization of Phase-Separated Droplets Formed by Nucleic Acid Binding Proteins: Using HP1 as a Model System". en. In: *Curr Protoc* 1.5, e109.

Sartorelli, Vittorio and Shannon M Lauberth (June 2020). "Enhancer RNAs are an important regulatory layer of the epigenome". en. In: *Nat. Struct. Mol. Biol.* 27.6, pp. 521–528.

Sasmal, Dibyendu K et al. (Dec. 2016). "Single-molecule fluorescence resonance energy transfer in molecular biology". en. In: *Nanoscale* 8.48, pp. 19928–19944.

Sawyer, Iain A, David Sturgill and Miroslav Dundr (Mar. 2019). "Membraneless nuclear organelles and the search for phases within phases". en. In: *Wiley Interdiscip. Rev. RNA* 10.2, e1514.

Scheefhals, Nicky, Manon Westra and Harold D MacGillavry (Jan. 2023). "mGluR5 is transiently confined in perisynaptic nanodomains to shape synaptic function". en. In: *Nat. Commun.* 14.1, p. 244.

Schindelin, Johannes et al. (June 2012). "Fiji: an open-source platform for biological-image analysis". en. In: *Nat. Methods* 9.7, pp. 676–682.

Schmidt, Hermann Broder, Ariana Barreau and Rajat Rohatgi (Oct. 2019). "Phase separation-deficient TDP43 remains functional in splicing". en. In: *Nat. Commun.* 10.1, p. 4890.

Schmied, Wolfgang H et al. (Nov. 2014). "Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1". en. In: *J. Am. Chem. Soc.* 136.44, pp. 15577–15583.

Schneider, Caroline A, Wayne S Rasband and Kevin W Eliceiri (July 2012). "NIH Image to ImageJ: 25 years of image analysis". en. In: *Nat. Methods* 9.7, pp. 671–675.

Schröder, Sebastian et al. (Nov. 2013). "Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells". en. In: *Mol. Cell* 52.3, pp. 314–324.

Schwarzer, Wibke et al. (Nov. 2017). "Two independent modes of chromatin organization revealed by cohesin removal". en. In: *Nature* 551.7678, pp. 51–56.

Sehgal, Pravin B et al. (Mar. 2020). "Biomolecular condensates in cell biology and virology: phase-separated membraneless organelles (MLOs)". en. In: *Anal. Biochem.*, p. 113691.

Shang, Yulei and Eric J Huang (Sept. 2016). "Mechanisms of FUS mutations in familial amyotrophic lateral sclerosis". en. In: *Brain Res.* 1647, pp. 65–78.

Shao, Wen et al. (Jan. 2022). "Phase separation of RNA-binding protein promotes polymerase binding and transcription". en. In: *Nat. Chem. Biol.* 18.1, pp. 70–80.

Shimobayashi, Shunsuke F et al. (Nov. 2021). "Nucleation landscape of biomolecular condensates". en. In: *Nature* 599.7885, pp. 503–506.

Shin, Yongdae and Clifford P Brangwynne (Sept. 2017). "Liquid phase condensation in cell physiology and disease". en. In: *Science* 357.6357.

Shin, Yongdae et al. (Jan. 2017). "Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets". en. In: *Cell* 168.1-2, 159–171.e14.

Shin, Yongdae et al. (Nov. 2018). "Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome". en. In: *Cell* 175.6, 1481–1491.e13.

Shrinivas, Krishna et al. (Aug. 2019). "Enhancer Features that Drive Formation of Transcriptional Condensates". en. In: *Mol. Cell* 75.3, 549–561.e7.

Smith, Charlotte et al. (May 2023). "TAL1 activation in T-cell acute lymphoblastic leukemia: a novel oncogenic 3' neo-enhancer". en. In: *Haematologica* 108.5, pp. 1259–1271.

Smith, Edwin and Ali Shilatifard (Mar. 2014). "Enhancer biology and enhanceropathies". en. In: *Nat. Struct. Mol. Biol.* 21.3, pp. 210–219.

Smith, Kelly P, Lisa L Hall and Jeanne B Lawrence (June 2020). "Nuclear hubs built on RNAs and clustered organization of the genome". en. In: *Curr. Opin. Cell Biol.* 64, pp. 67–76.

Spector, David L and Angus I Lamond (Feb. 2011). "Nuclear speckles". en. In: *Cold Spring Harb. Perspect. Biol.* 3.2.

Spitz, François and Eileen E M Furlong (Sept. 2012). "Transcription factors: from enhancer binding to developmental control". en. In: *Nat. Rev. Genet.* 13.9, pp. 613–626.

Sprunger, Macy L and Meredith E Jackrel (July 2021). "Prion-Like Proteins in Phase Separation and Their Link to Disease". en. In: *Biomolecules* 11.7.

St George-Hyslop, Peter et al. (Aug. 2018). "The physiological and pathological biophysics of phase separation and gelation of RNA binding proteins in amyotrophic lateral sclerosis and fronto-temporal lobar degeneration". en. In: *Brain Res.* 1693.Pt A, pp. 11–23.

Stam, Maike, Mariliis Tark-Dame and Paul Fransz (Apr. 2019). "3D genome organization: a role for phase separation and loop extrusion?" en. In: *Curr. Opin. Plant Biol.* 48, pp. 36–46.

Stasevich, Timothy J et al. (Dec. 2014). "Regulation of RNA polymerase II activation by histone acetylation in single living cells". en. In: *Nature* 516.7530, pp. 272–275.

Stein, Chad B et al. (Nov. 2022). "Integrator endonuclease drives promoter-proximal termination at all RNA polymerase II-transcribed loci". en. In: *Mol. Cell* 82.22, 4232–4245.e11.

Strom, Amy R et al. (July 2017). "Phase separation drives heterochromatin domain formation". en. In: *Nature* 547.7662, pp. 241–245.

Struhl, K (Mar. 1998). "Histone acetylation and transcriptional regulatory mechanisms". en. In: *Genes Dev.* 12.5, pp. 599–606.

Stryer, L and R P Haugland (Aug. 1967). "Energy transfer: a spectroscopic ruler". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 58.2, pp. 719–726.

Sun, Tanlin et al. (Nov. 2019). "Prediction of liquid-liquid phase separation proteins using machine learning". en.

Sungalee, Stephanie et al. (May 2021). "Histone acetylation dynamics modulates chromatin conformation and allele-specific interactions at oncogenic loci". en. In: *Nat. Genet.* 53.5, pp. 650–662.

Szabo, Quentin, Frédéric Bantignies and Giacomo Cavalli (Apr. 2019). "Principles of genome folding into topologically associating domains". en. In: *Sci Adv* 5.4, eaaw1668.

Tang, Seng Chuan et al. (June 2022). "Super-Enhancers, Phase-Separated Condensates, and 3D Genome Organization in Cancer". en. In: *Cancers* 14.12.

Taub, R et al. (Dec. 1982). "Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 79.24, pp. 7837–7841.

Taylor, Nicole O et al. (Aug. 2019). "Quantifying Dynamics in Phase-Separated Condensates Using Fluorescence Recovery after Photobleaching". en. In: *Biophys. J.*

Thévenaz, P, U E Ruttimann and M Unser (1998). "A pyramid approach to subpixel registration based on intensity". en. In: *IEEE Trans. Image Process.* 7.1, pp. 27–41.

Thompson, P R et al. (Sept. 2001). "Transcriptional coactivator protein p300. Kinetic characterization of its histone acetyltransferase activity". In: *J. Biol. Chem.* 276.36, pp. 33721–33729.

Thompson, Paul R et al. (Apr. 2004). "Regulation of the p300 HAT domain via a novel activation loop". en. In: *Nat. Struct. Mol. Biol.* 11.4, pp. 308–315.

Thompson Jr, Ronald W et al. (Nov. 2016). "Evaluation of effects of pH and ionic strength on colloidal stability of IgG solutions by PEG-induced liquid-liquid phase separation". en. In: *J. Chem. Phys.* 145.18, p. 185101.

Tie, Feng et al. (Sept. 2009). "CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing". In: *Development* 136.18, pp. 3131–3141.

Tie, Feng et al. (Mar. 2014). "Trithorax monomethylates histone H3K4 and interacts directly with CBP to promote H3K27 acetylation and antagonize Polycomb silencing". In: *Development* 141.5, pp. 1129–1139.

Tie, Feng et al. (Feb. 2016). "Polycomb inhibits histone acetylation by CBP by binding directly to its catalytic domain". In: *Proceedings of the National Academy of Sciences* 113.6, E744–53.

Tinevez, Jean-Yves et al. (Feb. 2017). "TrackMate: An open and extensible platform for single-particle tracking". en. In: *Methods* 115, pp. 80–90.

Toth, Zsolt et al. (Dec. 2013). "Biphasic euchromatin-to-heterochromatin transition on the KSHV genome following de novo infection". en. In: *PLoS Pathog.* 9.12, e1003813.

Tropberger, Philipp et al. (Feb. 2013). "Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer". en. In: *Cell* 152.4, pp. 859–872.

Tsang, Brian et al. (Mar. 2019). "Phosphoregulated FMRP phase separation models activity-dependent translation through bidirectional control of mRNA granule formation". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 116.10, pp. 4218–4227.

Turner, Bryan M (Oct. 1993). "Decoding the Nucleosome Minireview". In: *Cell* 75.

Ulianov, Sergey V et al. (May 2020). "Suppression of liquid-liquid phase separation by 1,6-hexanediol partially compromises the 3D genome organization in living cells". en.

UniProt Consortium (Jan. 2023). "UniProt: the Universal Protein Knowledgebase in 2023". en. In: *Nucleic Acids Res.* 51.D1, pp. D523–D531.

Van Rossum, Guido and Fred Drake (Mar. 2011). *The python language reference manual*. Bristol, England: Network Theory.

Van Treeck, Briana et al. (Mar. 2018). "RNA self-assembly contributes to stress granule formation and defining the stress granule transcriptome". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 115.11, pp. 2734–2739.

Venkatesh, Swaminathan and Jerry L Workman (Mar. 2015). "Histone exchange, chromatin structure and the regulation of transcription". en. In: *Nat. Rev. Mol. Cell Biol.* 16.3, pp. 178–189.

Venko, Katja and Eva Žerovnik (Aug. 2023). "Protein Condensates and Protein Aggregates: In Vitro, in the Cell, and In Silico". en. In: *Front. Biosci.* 28.8, p. 183.

Visel, Axel et al. (Feb. 2009). "ChIP-seq accurately predicts tissue-specific activity of enhancers". en. In: *Nature* 457.7231, pp. 854–858.

Wang, Chen et al. (July 2020). "Stress Induces Dynamic, Cytotoxicity-Antagonizing TDP-43 Nuclear Bodies via Paraspeckle LncRNA NEAT1-Mediated Liquid-Liquid Phase Separation". en. In: *Mol. Cell*.

Wang, Feng, Christopher B Marshall and Mitsuhiko Ikura (Nov. 2013). "Transcriptional/epigenetic regulator CBP/p300 in tumorigenesis: structural and functional versatility in target recognition". In: *Cell. Mol. Life Sci.* 70.21, pp. 3989–4008.

Wang, Feng et al. (Apr. 2012). "Structures of KIX domain of CBP in complex with two FOXO3a transactivation domains reveal promiscuity and plasticity in coactivator recruitment". In: *Proc. Natl. Acad. Sci. U. S. A.* 109.16, pp. 6078–6083.

Wang, Jia et al. (Oct. 2021). "Phase separation of OCT4 controls TAD reorganization to promote cell fate transitions". en. In: *Cell Stem Cell* 28.10, 1868–1883.e11.

Wang, Jie et al. (July 2018). "A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins". en. In: *Cell* 174.3, 688–699.e16.

Wang, Ling et al. (Dec. 2008a). "Structure and chemistry of the p300/CBP and Rtt109 histone acetyltransferases: implications for histone acetyltransferase evolution and function". en. In: *Curr. Opin. Struct. Biol.* 18.6, pp. 741–747.

Wang, Qi et al. (Oct. 2022a). "Global profiling of arginine dimethylation in regulating protein phase separation by a steric effect-based chemical-enrichment method". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 119.43, e2205255119.

Wang, Wenmeng et al. (May 2022b). "A histidine cluster determines YY1-compartmentalized coactivators and chromatin elements in phase-separated enhancer clusters". en. In: *Nucleic Acids Res.* 50.9, pp. 4917–4937.

Wang, Xiangting et al. (July 2008b). "Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription". en. In: *Nature* 454.7200, pp. 126–130.

Wegmann, Susanne et al. (Apr. 2018). "Tau protein liquid-liquid phase separation can initiate tau aggregation". en. In: *EMBO J.* 37.7.

Wei, Chao et al. (Aug. 2022). "CTCF organizes inter-A compartment interactions through RYBP-dependent phase separation". en. In: *Cell Res.* 32.8, pp. 744–760.

Wei, Ming-Tzo et al. (Nov. 2017). "Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles". en. In: *Nat. Chem.* 9.11, pp. 1118–1125.

Wei, Ming-Tzo et al. (Sept. 2020). "Nucleated transcriptional condensates amplify gene expression". In: *Nat. Cell Biol.*

Weiner, Benjamin G, Yigal Meir and Ned S Wingreen (Mar. 2021). "Sequence dependence of biomolecular phase separation". en.

Whyte, Warren A et al. (Apr. 2013). "Master transcription factors and mediator establish super-enhancers at key cell identity genes". en. In: *Cell* 153.2, pp. 307–319.

Wiese, Maria et al. (Aug. 2020). "Combined treatment with CBP and BET inhibitors reverses inadvertent activation of detrimental super enhancer programs in DIPG cells". en. In: *Cell Death Dis.* 11.8, p. 673.

Wolffe, A P and D Guschin (Apr. 2000). "Review: chromatin structural features and targets that regulate transcription". en. In: *J. Struct. Biol.* 129.2-3, pp. 102–122.

Xue, Bin et al. (Apr. 2010). "PONDR-FIT: a meta-predictor of intrinsically disordered amino acids". en. In: *Biochim. Biophys. Acta* 1804.4, pp. 996–1010.

Yao, Tso-Pang et al. (May 1998). "Gene Dosage–Dependent Embryonic Development and Proliferation Defects in Mice Lacking the Transcriptional Integrator p300". en. In: *Cell* 93.3, pp. 361–372.

Ye, Songtao et al. (Feb. 2022). "Visualizing the Multistep Process of Protein Aggregation in Live Cells". en. In: *Acc. Chem. Res.* 55.3, pp. 381–390.

Yu, Lan et al. (2021). "A Comprehensive Review of Fluorescence Correlation Spectroscopy". In: *Frontiers in Physics* 9.

Zagotta, William N et al. (Feb. 2016). "Measuring distances between TRPV1 and the plasma membrane using a noncanonical amino acid and transition metal ion FRET". en. In: *J. Gen. Physiol.* 147.2, pp. 201–216.

Zagotta, William N et al. (Oct. 2021). "An improved fluorescent noncanonical amino acid for measuring conformational distributions using time-resolved transition metal ion FRET". In: *Elife* 10, e70236.

Zbinden, Aurélie et al. (Oct. 2020). "Phase Separation and Neurodegenerative Diseases: A Disturbance in the Force". In: *Dev. Cell* 55.1, pp. 45–68.

Zeng, Xiangze et al. (June 2020). "Connecting Coil-to-Globule Transitions to Full Phase Diagrams for Intrinsically Disordered Proteins". en. In: *Biophys. J.*

Zhan, Yinxiu et al. (Mar. 2017). "Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes". en. In: *Genome Res.* 27.3, pp. 479–490.

Zhang, Song et al. (Feb. 2023). "NONO enhances mRNA processing of super-enhancer-associated GATA2 and HAND2 genes in neuroblastoma". en. In: *EMBO Rep.* 24.2, e54977.

Zhang, Yi et al. (July 2021). "Nuclear condensates of p300 formed though the structured catalytic core can act as a storage pool of p300 with reduced HAT activity". en. In: *Nat. Commun.* 12.1, p. 4618.

Zhang, Yong et al. (Sept. 2008). "Model-based analysis of ChIP-Seq (MACS)". en. In: *Genome Biol.* 9.9, R137.

Zhang, Zhao et al. (Oct. 2019). "Transcriptional landscape and clinical utility of enhancer RNAs for eRNA-targeted therapy in cancer". en. In: *Nat. Commun.* 10.1, p. 4562.

Zhao, Rundong and Rui-Qin Zhang (Sept. 2016). "A new insight into π-π stacking involving remarkable orbital interactions". en. In: *Phys. Chem. Chem. Phys.* 18.36, pp. 25452–25457.

Zhou, X et al. (Oct. 2006). "Optimization of the Tet-On system for regulated gene expression through viral evolution". en. In: *Gene Ther.* 13.19, pp. 1382–1390.

Chapter 9

Appendix

9.1 Macro to analyze puncta from overexpression experiments

```
/* script function: open image of an individual nuclei, apply a threshold, pre-set identify ROIs then analyze the ROIs for integrated density, area and circularity. The ROIs are saved within a directory and the results are then saved within a directory. The images in the list are pre cut to be individual nuclei and named accordingly e.g. "CBP-GFP_F1_NUC1_" */
dir1 = getDirectory("Choose a source directory");
dir2 = getDirectory("Choose a destination directory for puncta roi");
dir3 = getDirectory("Choose a destination directory for puncta results");
//For loop to open all images within the source directory
list = getFileList(dir1)
setBatchMode(true);
for(i=0; i<lengthOf(list); i++){
filename = list[i];
open(dir1 + list[i]);
//First generate a max projection from the Z-stack
run("Z Project...", "projection=[Max Intensity]");
//setting the threshold, the first value in the setThreshold is what we want to change dependent on the dataset, this identified the ROIs
setAutoThreshold("Default dark");
setThreshold(2700, 65535);
//set the measurements that we want to save including integrated density, area and circularity
run("Set Measurements...", "area mean standard modal min shape integrated area_fraction add redirect=None decimal=3");
//analyze the particles
run("Analyze Particles...", "display clear add");
//save the roi zipset
roiManager("Save", dir2 + list[i] + '.zip');
//Save the measurements
saveAs("Results", dir3 + list[i] + 'Results.csv'); /
//close down the window once complete
run("Close");
}
```

Table 9.1: Macro to record the integrated density of puncta within overexpression experiments Custom ImageJ/Fiji macro to define puncta as ROIs and analyze parameters such as integrated density, area and circularity of ROIs (Schindelin et al. 2012). Comments are shown in green and code is shown in black.

9.2 Macro to analyze standard deviation in optodroplet experiment

```
/* script function: open a video from the optodroplet section, apply stack reg to account for drift in the sample. apply threshold to define the nucleus, convert to mask and open and close mask to fill any holes within the nucleus ROI. select the values we want to measure in the analyze particle tab and then select the nuclei ROIs we want to measure. next we want to select the initial video, apply the ROIs and measure the ROIs across the timecourse. the ROIs and results are both saved in separate directories of the users choosing. */
//choose the directory of the origional videos and the directories of the results
dir1 = getDirectory("Choose a source directory");
dir2 = getDirectory("Choose a destination directory for results");
dir3 = getDirectory("Choose a destination directory for rois");
//this will set the macro to run through the files in directory 1 using a for loop
list = getFileList(dir1)
setBatchMode(true);
for(i=0; i<lengthOf(list); i++){
filename = list[i];
open(dir1 + list[i]);
//applying stack reg
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
setOption("ScaleConversions", true);
run("StackReg ", "transformation=[Rigid Body]");
//duplicate image to generate the ROI mask
run("Duplicate...", "duplicate");
//running the threshold on the duplicate to define nucleus
setAutoThreshold("RenyiEntropy dark");
run("Convert to Mask", "method=RenyiEntropy background=Dark calculate black");
//then close and open applied to the stack to fill any holes in the mask
run("Close-", "stack");
run("Open", "stack");
//select measurements that we want to take
run("Set Measurements...", "area mean standard modal min shape integrated median area_fraction stack display add redirect=None
decimal=3");
//selecting the ROIs - going to use an analyze particles for this part
run("Analyze Particles...", "size=50-Infinity display clear include overlay add composite stack");
//select image that hasn't been processed
selectWindow(list[i]);
//measure the ROIs over the timecourse
roiManager("Multi Measure");
//save the ROIs
roiManager("Save", dir3 + list[i] + '.zip');
//save the results
saveAs("Results", dir2 + list[i] + 'Results.csv');
//close the window
run("Close");
}
```

Table 9.2: optoDroplet macro to calculate change in standard deviation per nuclei
Custom ImageJ/Fiji macro to calculate the standard deviation of the nuclei in the optoDroplet assay over the timecourse (Schindelin et al. 2012).. Comments are shown in green and code is shown in black.

9.3 Validation of the CBP-HaloTag cell line

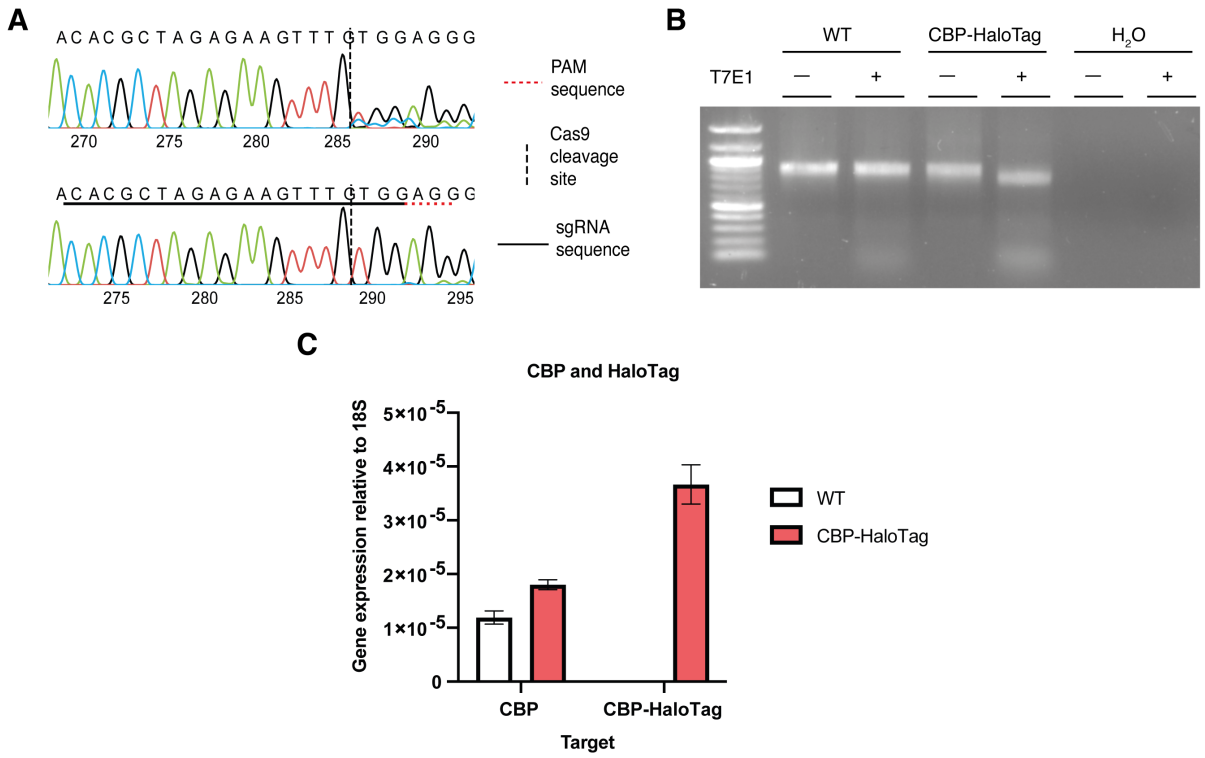

Figure 9.1: Validation of endogenous tagging of CBP in HEK293T cells
A - comparison of tagged CBP-HaloTag cells compared to WT Sanger sequencing traces. Upper trace shows the tagged allele with mutated PAM site after the Cas9 cleavage site, indicating successful insertion of the tagged sequence. The lower trace shows the WT sequence with correct PAM and 3'UTR. B - PCR products of WT cells, CBP-HaloTagged cells or water with and without T7E1. 3 bands are present in the CBP-HaloTag with T7E1 in comarison to WT, indicating insertion of the tag. C - RT-qPCR showing CBP-HaloTag expression in WT cell lines compared to CBP-HaloTag, expression is normalised to 18S. Graph was made in Graphpad Prism, where error bars denote SEM, n=3. This work and analysis was conducted by Dr Nicola Carruthers.

9.4 Sequence alignment of CBP and p300

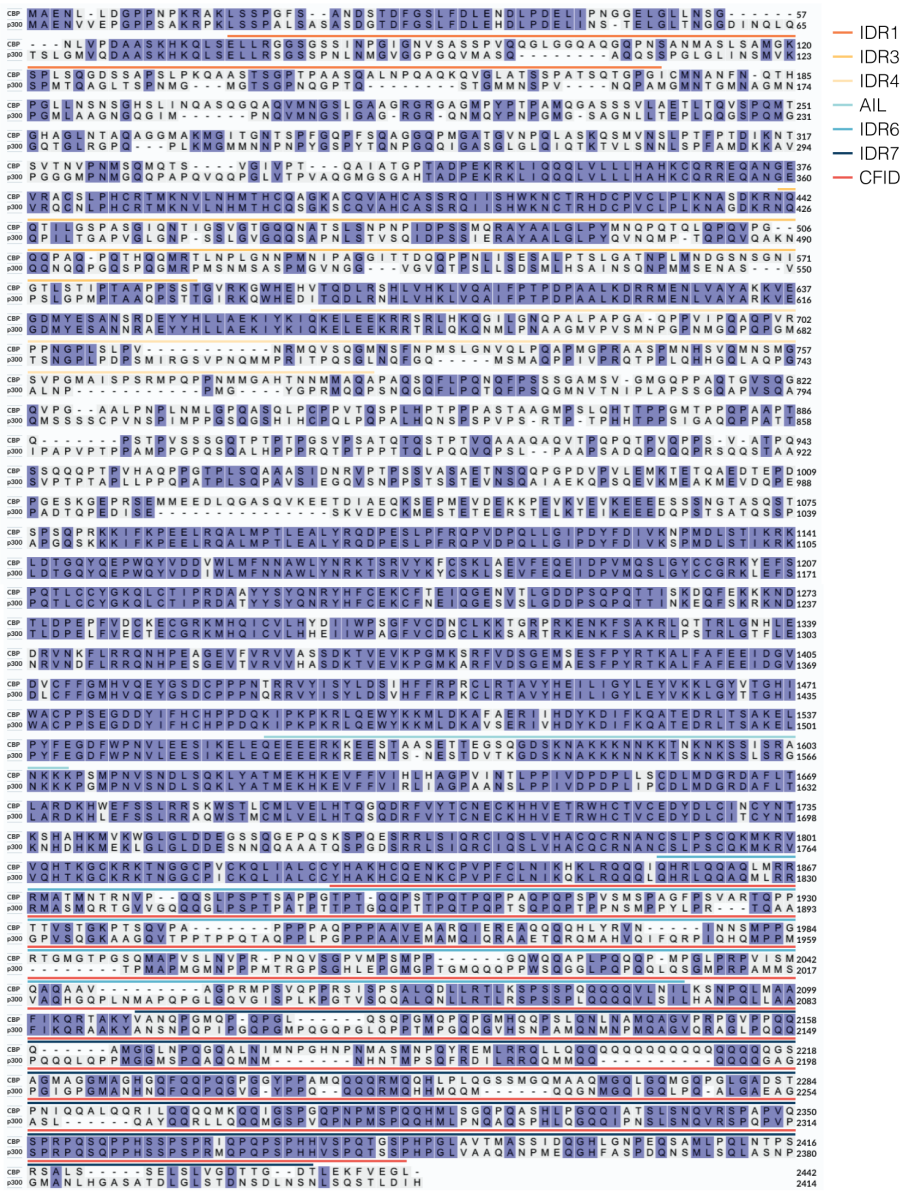

Figure 9.2: Sequence alignment of CBP and p300

Sequence alignment of amino acid sequence of CBP (UniProt, Q92793) and p300 (UniProt, Q09472) using UniProt Align, which uses Clustal Omega alignment software (UniProt Consortium 2023). Amino acids which are the same for both CBP and p300 are highlighted in purple. CBP IDRs defined in this doctoral work are indicated above the sequence of CBP.