Investigating the Molecular Basis of Menke Hennekam Syndrome: A Structural and Functional analysis of the TAZ2 & IDR4 domains of CREB Binding Protein.

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

Courtney Walsh

2023

Abstract

Menke Hennekam syndrome (MHS) is a rare genetic disorder characterised by intellectual disability and very specific physical abnormalities. MHS occurs as a result of de novo mutations to the cAMP response element binding protein binding protein (CREB-BP or CBP) gene. CBP is a master gene regulator protein, containing both intrinsic histone acetyltransferase activity and transcriptional coactivator activity. MHS associated mutations are located within exons 30 and 31 of the CBP gene, which encode the TAZ2 and IDR4 domains. The role which the TAZ2 and IDR4 domains play in the gene regulation carried out by CBP are yet to be fully understood. Previous research has established DNA binding activity by both TAZ2 and IDR4 domains, in addition to a lack of HAT specificity in the absence of the TAZ2 domain. It is suggested that the TAZ2 domain and neighbouring residues of the IDR4, may be involved in a regulatory interaction with DNA targets of the HAT domain. This study focused on uncovering more about the molecular basis of MHS by comparing the behaviour of MHS associated mutant CBP fragments of the TAZ2 and IDR4 domains with wild type fragments. Comparisons of structure, stability and DNA binding activity were carried out utilising circular dichroism spectroscopy, electrophoretic shift mobility assays and biolayer interferometry assays. The results produced throughout these experiments displayed that the MHS associated TAZ2-IDR4 mutant fragment contained DNA binding activity similar to the wild type TAZ2-IDR4 CBP fragments, during both EMSA and BLItz DNA binding assays. Structural investigations by circular dichroism spectroscopy of the MHS associated CBP mutants revealed a lack of native folding by the mutant protein fragments, which were prone to aggregation throughout expression and purification.

Acknowledgments

I would firstly like to thank my supervisor, Dr. Andrew Sutherland-Smith for being the best supervisor I could ask for. Thank you for your patience and endless support throughout this project. The opportunity to have received guidance and tutelage from such an accomplished mind is something I am very grateful for.

To my friends within the SNS Jennifer and Ruby, thank you for the countless hours spent listening to my problems and the invaluable advice and encouragement you have provided me. I am so lucky to have made such great friends who I look up to so fondly.

To Trevor Loo, thank you for all of your time spent answering my questions and helping me with equipment. Your breadth of knowledge and passion is inspiring. Past, present and future X-lab students are very lucky to have you.

To my family and dearest friends, Laura, Steph, Richelle and August thank you for your endless support throughout the entirety of my studies. I am so grateful to have such amazing friends and family.

Abbreviations

ABC	Ammonium Bicarbonate
ACN	Acetonitrile
Amp	Ampicillin
BCA	Bicinchoninic acid
BLItz	Bio-layer interferometry technology
BRD	Bromodomain
BSA	Bovine serum albumin
cAMP	Cyclic adenosine monophosphate
CD	Circular Dichroism
CH2	Cysteine histidine rich domain
CHARGe	Coloboma, heart disease, atresia of the choanae, retarded growth and mental development, genital anomalies, and ear malformations and hearing loss
CHOPs	Cognitive impairment, coarse facial features, heart defects, obesity, lung (pulmonary) involvement, short stature, and skeletal abnormalities.
CREB-BP/CBP	Cyclic adenosine monophosphate response element binding protein binding protein
CREB	Cyclic adenosine monophosphate response element binding protein
CV	Column volumes
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
EtBr	Ethidium Bromide
EtOH	Ethanol
Fs	Forward Standard (primer)

Ft	Forward tail (primer)
GoF	Gain of function
HAT	Histone Acetyltransferase
IDR4	Intrinsically Disordered Region #4
IDT	Integrated DNA technologies
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl-b-D-1-thiogalactopyranoside
KIX	Kinase inducible interacting domain
LB	Luria Bertani
LS MS	Liquid chromatography mass spectrometry
MH	Menke Hennekam
MHS	Menke Hennekam Syndrome
NCBD	Nuclear receptor coactivator binding domain
OD	Optical Density
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PTM	Post translational modification
Rs	Reverse standard (primer)
Rt	Reverse tail (primer)
RTS	Rubenstein Taybi Syndrome
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SLIM	Site directed ligase independent mutagenesis
TAE	Tris Acetate Ethylenediaminetetraacetic acid
TAZ1	Transcriptional Adapter Zinc Finger #1
TAZ2	Transcriptional Adapter Zinc Finger #2
TBE	Tris Borate Ethylenediaminetetraacetic acid
TCEP	Tris(2-carboxyethyl) phosphine
TF	Transcription Factor
UV	Ultraviolet
WT	Wild type
ZZ	ZZ type domain

Table of Contents

Abstract ..... *iii*

Acknowledgments ..... *v*

Abbreviations ..... *vi*

List of Tables ..... *viii*

List of Figures ..... *x*

Chapter 1. Introduction. *13*

1.1 Gene Regulation **2**

1.1.1 Chromatin 2

1.1.2 Gene Regulation in Human Development 3

1.2 CREB-BP: A Master Genetic Regulator **3**

1.2.1 CBP: A Transcriptional Coactivator 5

1.2.2 CBP: A Histone Acetyltransferase 5

1.3 Rubenstein Taybi Syndrome **6**

1.4 An Introduction to Menke Hennekam Syndrome **7**

1.5 The Zinc Finger Domain, TAZ2 ..... **9**

1.6 Intrinsically Disordered Region 4 ..... **12**

1.7 Clinical Clues **12**

1.8 Hypotheses, Aims and Objectives of this study. **14**

1.8.1 Hypotheses 14

1.8.2 Aims and Objectives 15

1.8.3 Introduction to Experimental CBP Protein Fragments 15

Chapter 2: Materials & Methods. 21

2.1 General Methods 22

2.1.1 Autoclaving 22

2.1.2 Water 22

2.1.3 Filter Sterilisation 22

2.1.4 Sonication 22

2.2 Microbiology 22

2.2.1 Bacterial Strains 22

2.2.2 Antibiotics 22

2.2.3 Media 23

2.2.4 Culture Optical Density 23

2.2.5 Preparation of Chemically Competent *E. coli* 23

2.2.6 Transformation & Starter Culture 24

2.2.7 Plasmid Purification 25

2.2.8 Culture Controls 25

2.3 DNA Methods 26

2.3.1 pET 15b Plasmid Stock Preparation 26

2.3.2 DNA Quantification 26

2.3.3 DNA Agarose Gel Electrophoresis 27

2.3.4 Polymerase Chain Reaction 28

2.3.5 Site Directed Ligase Independent Mutagenesis 29

2.3.6 DNA Sequencing 33

2.4 Protein Methods 33

2.4.1 Expression Culture 33

2.4.2 Cell Lysis 34

2.4.3 Protein Purification 35

2.4.4 SDS PAGE 36

2.4.5 Buffer Exchange 38

2.4.6 Concentrating Samples 39

2.4.7 Protein Quantification 39

2.4.8 In Gel Digestion for Mass Spectrometry 40

Chapter 3: Expression and Purification of TAZ2/IDR4 CBP Fragments..... 43

3. CBP Fragment Expression & Purification Optimisation 44

3.1 Features of an Expression System 45

3.1.1 Host Organism 45

3.1.2 Plasmid & Coding Sequence 46

3.1.3 Reducing Agent 47

3.1.4 Protease Inhibitor 48

3.1.5 C4A Mutants 48

3.2 Expression Optimisation Trials 50

3.2.1 Incubation Trials 50

3.2.2 $ZnCl_2$ Trials 54

3.2.3 Glycerol Stock Trials 57

3.2.4 Purification Buffer Trial 58

3.2.5 Sample Concentration & Buffer Exchange 61

Chapter 4: Structural Analysis of TAZ2/IDR4 CBP Fragments. 65

4.1 Circular Dichroism 66

4.1.1 CD Methodology 68

4.1.2 CD Secondary Structure Analysis of CBP Fragments 68

4.1.3 CD Tertiary Structure Analysis of CBP Fragments 72

Chapter 5: DNA Binding Analysis of TAZ2/IDR4 CBP Fragments. 79

5. TAZ2 DNA Binding Analysis 80

5.1 Electrophoretic Mobility Shift Assays (EMSA) 81

5.1.1 EMSA Methodology 82

5.1.2 EMSA Results 83

5.2 Bio-layer Interferometry (BLItz) 88

5.2.1 BLItz Procedure 89

5.2.2 CBP Fragment DNA Binding BLItz Results 90

Chapter 6: Discussion and Future Directions. 97

6.1 Discussion of Experimental Results 98

6.2 Future Directions 105

6.2.1 Expression and Purification Protocol 105

6.2.2 Structural and Stability Studies 106

6.2.3 DNA Binding Activity Studies 106

6.2.4 Other CBP Functional Analysis 107

6.3 Final Summary 107

Bibliography 109

Appendix 119

SLIM Primers 119

CBP Protein Sequences 120

EMSA and BLItz DNA Probes 122

Web Resources 122

List of Tables

Table 1	Overview of CBP fragments used throughout this study with descriptions of contents, length and mutations present.	18
Table 2	Summary of CBP WT fragments.	26
Table 3	PCR master mix.	28
Table 4	Colony PCR thermocycle protocol.	29
Table 5	CBP mutants produced using SLIM overview.	29
Table 6	SLIM PCR master mix.	30
Table 7	SLIM PCR amplification protocol.	30
Table 8	SLIM PCR hybridisation protocol.	31
Table 9	Polyacrylamide resolving gel.	37
Table 10	Polyacrylamide stacking gel.	37
Table 11	SDS PAGE running buffer.	38
Table 12	Colloidal Coomassie stock.	38
Table 13	10x Bradford reagent stock.	40
Table 14	Overview of CBP WT fragments.	49
Table 15	Incubation trial SDS PAGE band intensity values.	53
Table 16	Tm (ºC) values of CBP fragments which produce unfolding spectra in the far UV.	76
Table 17	1x TBE buffer recipe.	83
Table 18	CBP fragment EMSA result summary.	86
Table 19	Dissociation (K_D) values produced by BLItz assays.	93

List of Figures

Figure 1	Compilation of CBP showing the solved domain fragment structures.	4
Figure 2	Rubenstein Taybi patient.	6
Figure 3	Menke Hennekam patients.	8
Figure 4	Overview of RTS and MHS patient mutations located in the CBP protein.	9
Figure 5	Structure of the TAZ2 domain.	10
Figure 6	Overlay of crystallography results of TAZ2 interacting with different helical binding partners.	11
Figure 7	Facial feature analysis of MHS, RTS and Dup16p13.3 patients.	13
Figure 8	SLIM methodology overview.	32
Figure 9	pET15b Plasmid.	46
Figure 10	Incubation trial SDS PAGE 37ºC 3-hour incubation post IPTG induction.	52
Figure 11	Incubation trial SDS PAGE 20ºC 14-hour incubation post IPTG induction.	51
Figure 12	SDS PAGE Analysis Zn^{2+} Trial.	55
Figure 13	CBP TAZ2-IDR4 fragment Zn^{2+} SDS PAGE.	56
Figure 14	CBP TAZ2-IDR4 C4A fragment Zn^{2+} SDS PAGE.	57
Figure 15	IMAC buffer trial SDS PAGE.	61
Figure 16	Characteristic CD spectra for different protein secondary structures.	67
Figure 17	Circular Dichroism Far UV Spectrum of CBP fragments at room temperature.	69
Figure 18	CBP TAZ2 C4A fragment thermal unfolding far UV spectra.	73
Figure 19	CBP TAZ2-IDR4 fragment thermal unfolding far UV spectra.	74
Figure 20	CBP TAZ2-IDR4 C4A fragment thermal unfolding far UV spectra.	74
Figure 21	CBP TAZ2 (I1776T) mutant fragment thermal unfolding far UV spectra.	75

Figure 22	CBP IDR4 (R1868W) mutant fragment thermal unfolding far UV spectra.	75
Figure 23	EMSA carried out with 50% GC probe and CBP fragments.	83
Figure 24	EMSA carried out with 60% GC probe and CBP fragments.	84
Figure 25	EMSA carried out with gTEL45 probe and CBP fragments.	85
Figure 26	BLItz TAZ2 C4A DNA probe association and dissociation.	90
Figure 27	BLItz TAZ2-IDR4 DNA probe association and dissociation.	91
Figure 28	BLItz TAZ2-IDR4 C4A DNA probe association and dissociation.	91
Figure 29	BLItz TAZ2 mutant (I1776T) DNA probe association and dissociation.	92
Figure 30	BLItz IDR4 mutant (R1868W) DNA probe association and dissociation.	92
Figure 31	Alphafold prediction software structure prediction of TAZ2-IDR4 C4A CBP fragment.	101

Chapter 1. Introduction.

1.1 Gene Regulation

The human genome contains over 3.2 billion deoxyribonucleic acid (DNA) base pairs (Spector et al. 2003). Each and every cell within the human body (with a few exceptions) contains two full copies of this repetitive, four-letter code which is estimated to total 2 meters in length (O'Donnell et al., 2013). Not all regions of our huge, complex genome are required at all times for expression (Latchman, 2007). The diverse cell-types, with specific functions, that make up different human tissues require particular parts of the genetic blue print (Felsenfeld, 2014; Latchman, 2007). The expression of our enormous genome is highly regulated through a multitude of mechanisms by an array of key protein players, all of whom must function in a coordinated fashion (Latchman, 2007).

1.1.1 Chromatin

As a means for storage, eukaryotic DNA is wound upon histone proteins into chromatin, which then at a higher order of folding forms our chromosomes (Woodcock & Dimitrov, 2001). This condensed storage system also has an important secondary purpose. The nuclear organisation system also limits gene expression machinery access to different regions of the DNA (Cunliffe, 2003; Spector, 2003). Altering the density of chromatin packing is one of the ways in which the expression of our genes can be regulated (Li & Reinberg, 2011).

Regions of the genome which are required for expression, are less densely packed within euchromatin. Alternatively, heterochromatin refers to segments of genes which are inaccessible by transcription machinery and have silenced expression (Strålfors & Ekwall, 2006). Different chromatin forms arise through post translational modifications (PTMs) to either the histone proteins or the DNA (Biel et al., 2005). This alters the affinity which the two components have for each other, and therefore results in tightening or weakening of the complex (Gelato & Fischle, 2008; Igo-Kemenes et al., 1982). The addition or removal of PTMs, is done by gene regulator proteins. These gene regulator proteins have the task of meticulously identifying target regions of DNA or histones within the genome, and making specific modifications in order to influence gene expression (Gelato &

Fischle, 2008; Li & Reinberg, 2011). This is a role with little room for error (Izumi, 2016; Tanaka et al., 1997).

1.1.2 Gene Regulation in Human Development

The process of our development from a single totipotent cell through to an entire, complex multicellular being, is dependent on the regulation of gene expression (Reik, 2007). The fundamental difference between a differentiated cell with a determined tissue fate and an embryonic pluripotent cell, are the patterns of gene expression occurring (Latchman, 2007). As cell differentiation occurs during embryonic development, the genes which are expressed in a cell to maintain its totipotency are repressed by packing the DNA into heterochromatin (Theunissen & Jaenisch, 2017). Simultaneously, genes which are essential for the differentiated cell gain increased accessibility through remodelling into euchromatin (Boland et al., 2014; Theunissen & Jaenisch, 2017). Mutations which occur to gene regulator proteins responsible for chromatin remodelling throughout development, give rise to a range of developmental or malformation diseases such as; Cornelia de Lange, CHOPS, Coffin-Siris, CHARGE, Wiedemann-Steiner, Kabuki, Young-Simpson and Say-Barber syndromes (Izumi, 2016).

1.2 CREB-BP: A Master Genetic Regulator

A vital transcriptional regulator protein in humans, is the cyclic adenosine monophosphate (cAMP) regulatory element binding protein binding protein, also known as CREBBP or CBP (Goldman et al., 1997; Tanaka et al., 1997). Human CBP is composed of 2,442 amino acids, 60% of which compose intrinsically disordered regions that link the several structured domains together (figure 1) (Dyson & Wright, 2016; Ogryzko et al., 1996; Sobulo et al., 1997).

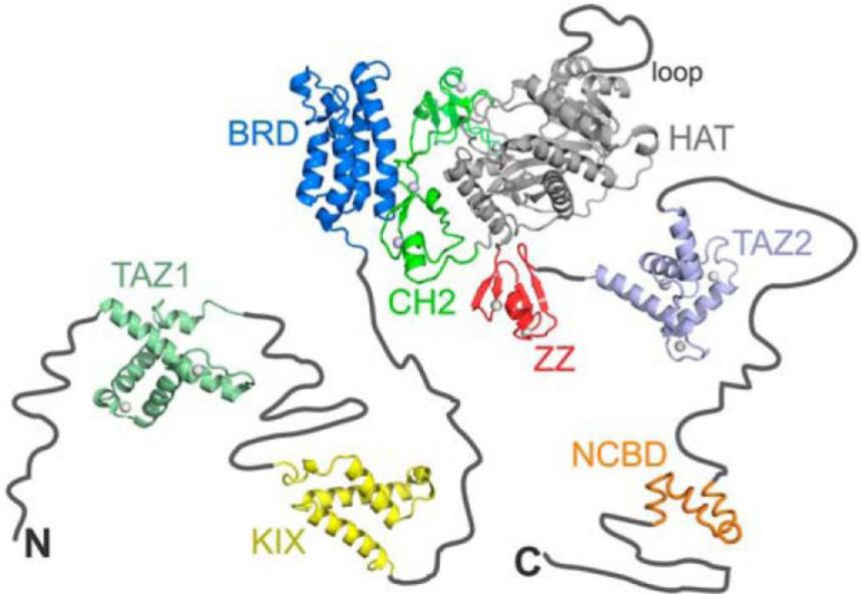

Figure 1 Compilation of CBP showing the solved domain fragment structure. Structured domains are coloured regions, intrinsically disordered linkers are black. Zinc finger domains include TAZ1 (teal), TAZ2 (purple) and ZZ (red). KIX domain (yellow) is a transcription factor binding site. Other domains include the bromodomain (BRD) (blue), cysteine histidine rich domain (CH2 or PHD/RING domain) (green), HAT (grey) and nuclear receptor coactivator binding domain (NCBD) (orange). The IDR4 region (unlabelled in illustration), is the black linker located between the TAZ2 domain (purple) and the NCBD domain (orange). Sourced from an open access article, permission for reuse not required (Dyson & Wright, 2016).

CBP is called a master transcriptional regulator, due to its ability to influence the expression of a wide array of gene targets (Janknecht & Hunter, 1996). In fact, CBP and its paralog p300 have been found at the transcription start sites of over 16,000 human genes (Ramos et al., 2010). CBP is able to influence gene expression through at least two known mechanisms (Bannister & Kouzarides, 1996; Janknecht & Hunter, 1996; Ogryzko et al., 1996).

1.2.1 CBP: A Transcriptional Coactivator

The first mechanism involves CBP acting as a transcriptional coactivator protein, binding to cAMP response element protein (CREB) (Cardinaux et al., 2000). Elevated cellular cAMP levels result in activation of protein kinase A, which is then able to phosphorylate serine residue 133 of CREB (Chawla et al., 1998). Tyrosine residue 658 of the CBP KIX domain (figure 1) can then form hydrogen bonds with the phosphorylated serine of CREB, which is mounted upon cAMP response element DNA (Radhakrishnan et al., 1997). CREB bound CBP is then able to recruit transcription machinery in order to initiate transcription (Kwok et al., 1994).

1.2.2 CBP: A Histone Acetyltransferase

The second mechanism by which CBP can regulate gene expression is through its histone acetyltransferase domain (HAT) (Bannister & Kouzarides, 1996). The centrally located HAT domain of CBP (figure 1) is able to remodel chromatin though acetylating lysine residues on target histone proteins (Bannister & Kouzarides, 1996). Lysine tail acetylation by the HAT domain reduces the affinity which the histone proteins have for the negatively charged DNA. This increases the access which transcription machinery has to target DNA (Ogryzko et al., 1996).

Master transcriptional regulators such as CBP are critical for normal growth and development, fine tuning gene expression by existing at the intersection of numerous regulatory pathways (Izumi, 2016; Smith et al., 2004; Tanaka et al., 1997). Smith et al. (2004) refers to CBP as a “molecular interpreter” with the ability to understand the encrypted language of the human genome. Mutations which alter the structure and function of CBP, result in significant consequences seen at not only a molecular level, but beyond (Banka et al., 2019; Menke et al., 2018; Menke et al., 2016; Sima et al., 2022).

1.3 Rubenstein Taybi Syndrome

In 1963, Rubenstein and Taybi investigated a small group of unrelated individuals with intellectual disabilities and striking physical similarities. Typical characteristics include short stature, broad toes, big thumbs, grimacing smile, arched palate, beaked nose, long eyelashes and arched brows, broad nasal bridge, talon cusps, downward sloping palpebral fissures and mild micrognathia (figure 2) (Hennekam, 2006; van Genderen et al., 2000). The extremely particular dysmorphology (figure 2) suggested the individuals may possess the same “syndrome” (Rubinstein & Taybi, 1963). Initially, the standard of genetic investigation was not yet developed enough to accurately identify any chromosomal abnormalities which may be at the root of the condition (Rubinstein & Taybi, 1963).

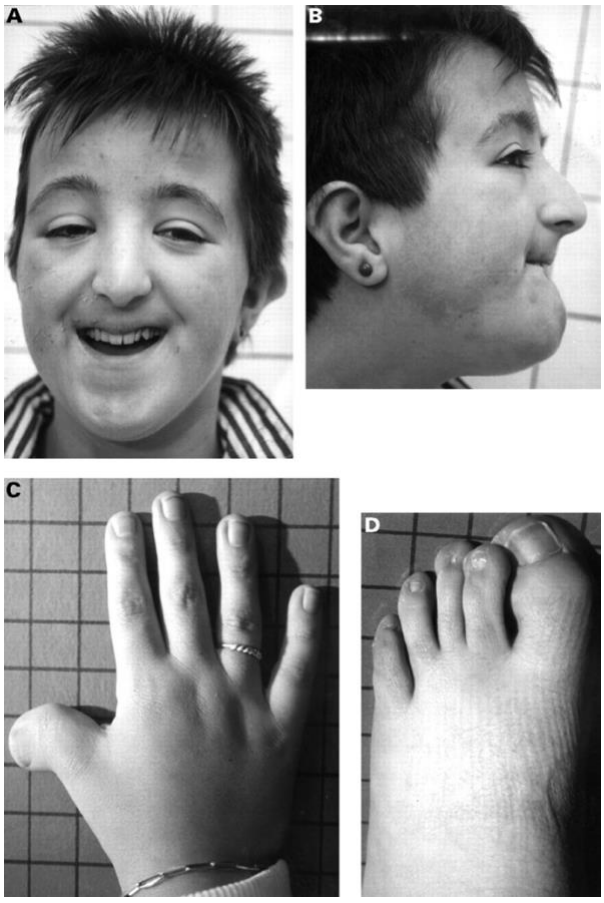

Figure 2 Rubenstein Taybi Patient (A) Front on profile and (B) side profile of characteristic facial features displayed in a 13-year-old RTS female including beaked nose, broad nasal ridge, micrognathia and downward sloping palpebral fissures. (C) Characteristic RTS hand with disproportionally large thumb and (D) feet with broad toes (same patient). Figure reproduced with permission (van Genderen et al., 2000).

Imaizumi and Kuroki (1991) were the first to report on the genetic basis of Rubenstein Taybi Syndrome (RTS). They carried out cytogenetic studies on a single patient and found a reciprocal translocation between chromosomes 16 and 2. Shortly after this Petrif et al. (1995) carried out further investigations and confirmed that mutations to chromosomal location 16p13.3 resulted in RTS. Petrif et al. (1995) also established that the protein product of the chromosomal location of interest was CBP, and that the loss of just one functional copy of *CBP* (haploinsufficiency) resulted in dysregulated gene expression.

1.4 An Introduction to Menke Hennekam Syndrome

History repeated itself in 2016 when Menke et al. (2016). investigated a group of unrelated individuals all displaying similar physical features and intellectual impairment to one another, eventually categorised as having a condition named Menke Hennekam Syndrome (MHS). Exome sequencing was carried out on the patients, focusing on genes which are typically associated with intellectual impairment when mutations arise. Due to the previous characterisation of RTS, this meant the *CBP* gene was included in the analysis. Sanger sequencing results revealed that the patients contained a range of different *CBP* mutations, all within exons 30 and 31. However the phenotype of these MHS individuals did not fit that of typical RTS (Menke et al., 2016). Common characteristics included autistic behaviour, microcephaly, telecanthus, depressed nasal ridge, short nose, up slanted palpebral fissures, anteverted nares, short columella, long philtrum and typical fingers and toes (figure 3) (Menke et al., 2018; Menke et al., 2016).

Figure 3 Menke Hennekam patients front on profile displaying characteristic facial features including short nose, depressed nasal ridge, telecanthus, anteverted nares, short columella and long philtrum and upward sloping palpebral fissures. Numbering matches mutations listed in original paper Menke et al. (2018). Image reproduced with permission.

At this time, it was noted that RTS associated mutations were located all throughout the *CBP* gene, notably except within exons 30 and 31 (figure 4) (Menke et al., 2016). The distinct phenotype seen for MHS patients, compared to RTS in addition to the striking location of MHS mutations only in exons 30 and 31, suggests that a different molecular mechanism for MHS is at the root of this newly distinguished syndrome. The protein products of *CBP* exons 30 and 31, are the TAZ2 and IDR4 domains (figure 4). The functions of the TAZ2 and IDR4 domains are still currently under investigation, and the molecular basis of the syndrome, termed Menke Hennekam syndrome (MHS), is still undetermined.

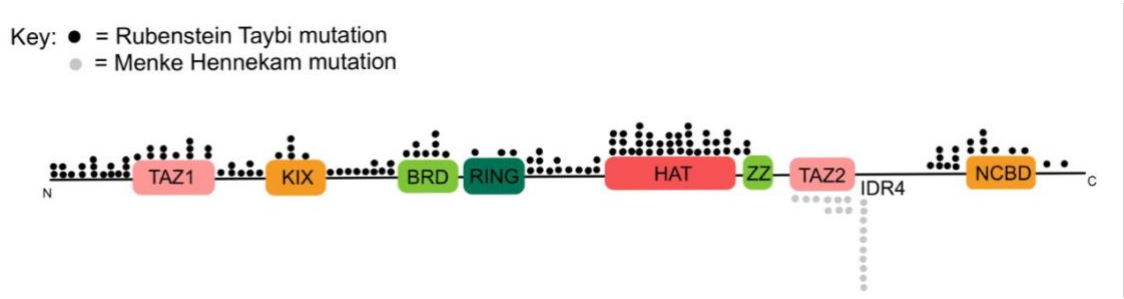

Figure 4 Overview of RTS and MHS patient mutations located in the CBP protein. CBP domains are represented as coloured blocks attached by disordered linker regions (black lines), with the recorded mutations seen in RTS patients (black spots) and MHS patients (grey spots) illustrated. RTS mutations are present throughout CBP entirely except in the TAZ2 (pink) and N-terminal region of IDR4 (black line). In contrast, MHS mutations are found solely within TAZ2 (pink) and the N-terminus of IDR4 (black line) regions (Menke et al. 2016). Image created using GoodNotes.

1.5 The Zinc Finger Domain, TAZ2

TAZ2, short for transcriptional adapter zinc finger 2 is the second of two zinc finger domains within CBP (figure 1, purple) (Dyson & Wright, 2016). The domain is composed of CBP residues 1763-1855, which fold into four alpha helices labelled α1-α4 (figure 5) (Dyson & Wright, 2016; Menke et al., 2018). The main characteristic of zinc finger domains is the interaction between cysteine and histidine residues of the protein and zinc ions in order for stabilised folding into a finger like structure, which is able to straddle the major groove of DNA to form sequence specific interactions with base pairs (Laity et al., 2001; Wolfe et al., 2000). Folding of the TAZ2 domain in particular, is dependent on the interaction between three zinc atoms and the HCCC type motifs of the domain (De Guzman et al., 2000). The HCCC regions are composed of one histidine and three cysteine residues (specific residues in appendix), which reside spread out across the loops and at the edges of the helices.

Figure 5 Structure of TAZ2 domain. α1 (blue), α2 (green), α3 (yellow) and α4 (red). Teal spheres represent bound Zinc atoms. Image generated using Chimera X.

The alpha helical structures within the TAZ2 domain pack together to create a hydrophobic binding core which is flanked by positively charged arginine and lysine residues (De Guzman et al., 2000). This is a known binding site for various transcription factors such as p53, ETF2, STAT1, E1A, E2A, and proteins from the CCAAT enhancer binding protein family (Dyson & Wright, 2016; H. Feng et al., 2009; Ferreon et al., 2009; Miller et al., 2009; Wojciak et al., 2009). Sequence specific DNA or protein binding is another attribute frequently seen among zinc finger domains, which is why they are so common within gene regulator proteins (Klug, 1999). However, many of the TAZ2 binding partners share minimal sequence homology (Miller et al., 2009). This suggests another mechanism must therefore be in place by which the TAZ2 domain accurately identifies binding targets.

H. Feng et al. (2009) investigated interactions between TAZ2 and p53. They displayed the formation of a transient amphipathic helical structure within p53 upon TAZ2 binding. Other research such as that by Ferreon et al. (2009) and Wojciak et al. (2009) investigating E1A-TAZ2 and STAT1-TAZ2 interactions respectively, have also displayed a similar pattern. This pattern being that the TAZ2 binding targets possess amphipathic or hydrophobic alpha helices, for at least the duration of TAZ2 binding. This suggests the hydrophobic binding groove

of TAZ2 best fits helical binding partners, and binding selectivity is potentially based upon hydrophobicity. This suggestion is supported further by the work reviewed by Dyson and Wright (2016) of investigations into these intrinsically disordered protein domains forming transient helices upon binding with the TAZ2 domain. The crystallography studies display multiple hydrophobic, helical binding partners interacting with the TAZ2 hydrophobic core (figure 6).

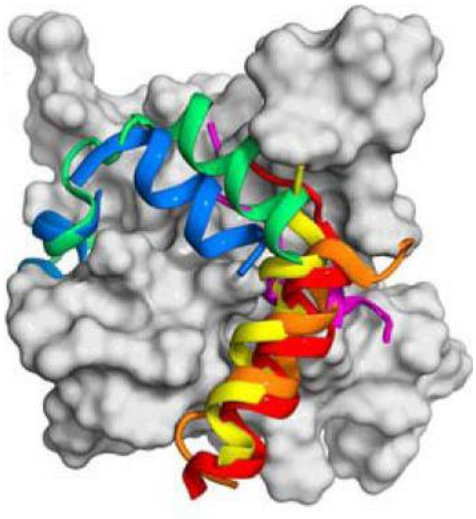

Figure 6 Overlay of crystallography results of TAZ2 (grey) interacting with different binding partners, all helical and hydrophobic/amphipathic. STAT1 (green), C/EBP (blue), p53 AD1 (orange), p53 AD2 (magenta), EDTA1 (red), Free neighbouring TAZ2 α4 (yellow). Sourced from an open access article, permission for reuse not required (Dyson & Wright, 2016).

New and still to be formally published (available on Biorxiv) research by Sheahan et al. (2020) displays an additional role potentially played by TAZ2 in mediating gene expression. Deletion studies carried out with just the HAT catalytic core (BRD (bromodomain), RING/CH2 (atypical ring domain), HAT and ZZ (ZZ type zinc finger) domains (figure 1)) of CBP, failed to acetylate target histones accurately. This implies an additional domain must be involved in the process of precisely identifying target histones. With the addition of the TAZ2 domain (figure 1, purple) to the catalytic core, precise and robust histone acetylation was seen (Sheahan et al., 2020). Sheahan et al. (2020) also carried-out TAZ2 DNA binding assays in order to further investigate the novel DNA binding feature of the domain. The results displayed sequence independent DNA binding by TAZ2. Further investigations are required, but it may be that DNA binding partners,

which have a net negative charge, form interactions with the positive arginine and lysine residues which surround the TAZ2 binding site.

1.6 Intrinsically Disordered Region 4

Over half of the residues within CBP make up intrinsically disordered regions (IDRs) (Dyson & Wright, 2016). IDR4, is the fourth IDR within CBP and is over 150 residues long (CBP 1855-2020) (figure 1) (Piai et al., 2016). Intrinsically disordered regions are said to be common among gene regulator proteins with many binding partners (Dyson & Wright, 2016). It is the intrinsic disorder and conformational flexibility of these regions which allows for binding conformity to a range of different target sites (Contreras-Martos et al., 2017; Minezaki et al., 2006). Therefore, it is understandable that mutations within an IDR which may bring about undesirable structure or prevent the formation of any transient structures from forming, may be detrimental to function (Seera & Nagarajaram, 2022).

Although conformational adaptability is a key attribute of IDRs, it is understood that they may also acquire transient structures upon interaction with targets. NMR studies by Piai et al. (2016) aimed to uncover more about the said transient structures of the IDR4 within CBP using NMR based studies. They found that the IDR4 was inclined to form two alpha helical structures, the function of which is still unknown. The two helices formed with residues 1852-1875 and 1951-1978 (residues 2-25 and 101-128 of the IDR4). An unignorable coincidence of the helical structure forming within residues 1852-1875 is that this is the same region in which all recorded MHS associated IDR4 mutations reside (Sima et al., 2022). Further investigations are required to determine the function of this apparently critical reg

1.7 Clinical Clues

Clinical studies conducted on MHS patients also provide some interesting insight towards uncovering the molecular basis of MHS. Menke et al. (2018) presented a facial feature analysis utilising 3D photogrammetric image (figure 7). They

compared an individual with a *CBP* gene duplication (Dup16p13.3), a Menke Hennekam syndrome patient (C11), and 13 Rubenstein Taybi syndrome patients (Mean RSTS; 13 RTS patient features averaged). All individuals were of the same age, sex and ethnicity in attempt to reduce any non-relevant variables. The results displayed extreme similarities between the inverted MHS C11 patient and the averaged RTS individuals. MHS patients presenting as “clinically opposite” to RTS individuals, suggests that the molecular basis of the two syndromes may also be opposing (Menke et al., 2018).

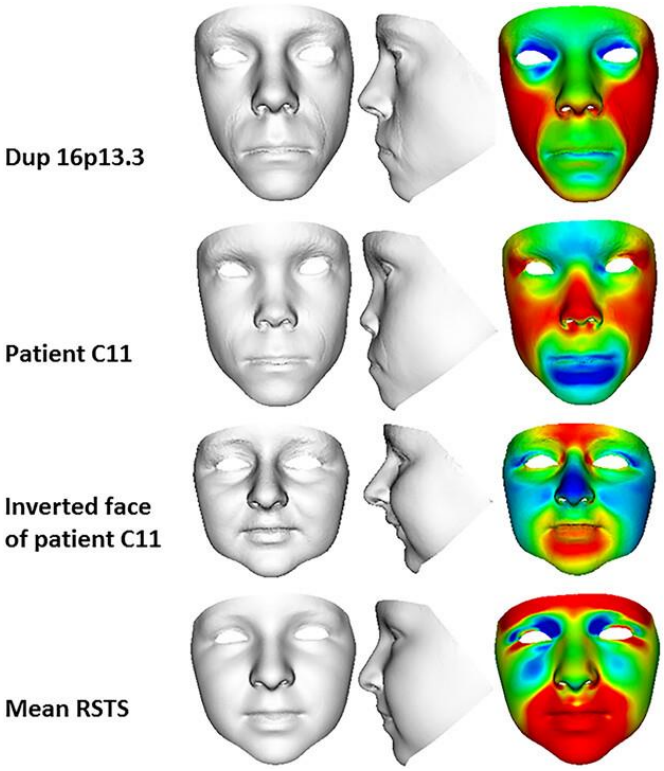

Figure 7 Facial feature analysis of MHS, RTS and Dup16p13.3 patients by Menke et al. 2018. Dup16p13.3 is an individual with a duplication of the CBP gene. Patient C11 is an MHS individual. Mean RSTS is a representation of the average of 13 RTS patient features. Red represents compressed features, where blue represents features with expansion compared to typical individuals. Colours in between represent features in between these extremes. Patients included for analysis were matched by sex, age and ethnicity. Image reproduced with permission (Menke et al., 2018).

1.8 Hypotheses, Aims and Objectives of this study.

Piecing the molecular and clinical clues together, we can begin to theorise what the molecular basis of the syndrome may include. Given that mutations which cause a loss of typical CBP function (LoF) result in RTS, it is suggested that MHS may be a result of mutations which cause a gain of *CBP* function (GoF) (Menke et al., 2018). The difference in phenotype between RTS and MHS, combined with the absence of RTS mutations in exons 30-31 (TAZ2-IDR4) suggests that MHS is not caused by a LoF mechanism. Additionally in support of this idea are the physical facial similarities visible between the MHS patients and the Dup16p13.3 patient (figure 7). An increase in CBP activity due to gain of function (GoF) mutations, which is the proposed cause of MHS, would at a cellular level resemble a duplication of the *CBP* gene.

If the molecular basis of MHS is indeed the result of an aberrant gain in *CBP* function, the clustering of the MHS patient mutations in the TAZ2 domain/IDR4 regions only (figure 4), suggests that in wild type (WT) CBP the TAZ2/IDR4 region is involved in a regulatory interaction. When this regulatory mechanism by TAZ2 and IDR4 is altered by mutations, the other domains within CBP (such as the HAT domain) under the influence of any abnormal interaction will be hyperactive, mimicking the effects of duplicated *CBP*. This suggestion is further supported by the work discussed earlier by Sheahan et al. (2020) which displayed CBP off-target histone acetylation in the absence of the TAZ2 domain. The full nature of any regulatory activity of TAZ2-IDR4 remains unclear and requires further investigation.

1.8.1 Hypotheses

We hypothesise that the TAZ2 and IDR4 domains interact with DNA through the positive arginine and lysine residues (De Guzman et al., 2000) which border the central hydrophobic binding pocket of the TAZ2 domain. The work conducted by Sheahan et al. (2020) suggests that this interaction may regulate activity of the HAT domain. We also hypothesise that the N-terminal, helical forming region of the IDR4 (CBP 1855-1876) (Piai et al., 2016) has some involvement in this

regulatory interaction and the DNA binding activity by the TAZ2 domain will differ in the presence and absence of this region. This hypothesis is supported by the majority of MHS associated mutations being located within this region of the protein (1855-1876). We hypothesise that MHS associated mutations may alter either the structure, the stability, the key characteristics or a combination of the above of the TAZ2-IDR4 region of CBP, which are essential for native function.

1.8.2 Aims and Objectives

1. To construct and optimise expression and purification protocols for a range of TAZ2/IDR4 fragments of CBP. This will be done using a bacterial recombinant expression system and a combination of liquid chromatography techniques. This work is presented in chapter 3.

2. To determine if MHS-associated mutations alter the structure or stability of TAZ2/IDR4 fragments. This will be investigated using circular dichroism (CD) spectroscopy in both the near and far ultraviolet (UV) regions, in order to investigate any large-scale differences about both secondary and tertiary protein structure. This work is presented in chapter 4.

3. To further investigate TAZ2 DNA binding and if TAZ2 DNA binding activity is altered in the presence of the N-terminal, helical forming region of the IDR4. Additionally, to investigate if and how mutations within the TAZ2 and IDR4 domains impact TAZ2 DNA binding. This will be done using a combination of electrophoretic mobility shift assays (EMSAs) and BLItz assays. This work is presented in chapter 5.

1.8.3 Introduction to Experimental CBP Protein Fragments

In order to investigate our hypotheses, a number of different CBP fragments were utilised. All fragments included within this study are listed in table 1. To determine the impact of the IDR4 N-terminal, helical forming region on the DNA binding activity of the TAZ2 domain, wild type (WT) CBP fragments containing only the TAZ2 domain (TAZ2, CBP 1763-1855) and the TAZ2 domain plus the N-terminal

region of the IDR4 (TAZ2-IDR4, CBP 1763-1876) were produced. These fragments were also used for structural and stability studies to investigate the changes (if any) brought about by MHS associated mutations.

In addition to these WT fragments, C4A versions of these fragments were also produced. The C4A substitution fragments were first introduced by De Guzman et al. (2000). During their TAZ2 domain fragment studies, they established the cysteine residues involved in binding Zn^{2+} for structural stability. Upon substitution of four other, non Zn^{2+} cysteine binding residues with alanine residues, structure of the domain remained native. The substitution of cysteine residues, as long as structure or function is not impaired is favourable because the expression and purification of a protein fragment with less cysteine residues is typically more successful. For this reason, C4A versions of the TAZ2 and TAZ2-IDR4 fragments were also ordered, in hopes that expression and purification of these fragments may provide more soluble fragment for experimental use, while behaving in an experimental setting as a WT. These fragments are referred to throughout this study as TAZ2 C4A (CBP 1763-1855) and TAZ2-IDR4 C4A (CBP 1763-1876). Plasmids containing the CBP fragment coding sequence were ordered from Integrated DNA Technologies (IDT).

Menke Hennekam associated mutants included in this study were picked from previous clinical papers and represent the CBP activity occurring in real patients. One additional mutant, Ile1776Thr was included in order to investigate the importance of the hydrophobic binding core, without substituting a Zn^{2+} binding residue. Mutations were introduced using the site directed ligase independent mutagenesis (SLIM) methodology and the plasmid containing the coding sequence for fragment 4 as the template. Therefore, all mutants contain the full TAZ2 domain and the N-terminal region of the IDR4 (CBP 1763-1876), in addition to the C4A mutations.

The first mutant (TAZ2 Ile1776Thr mutant) contains the substitution of the hydrophobic TAZ2 core isoleucine residue 1776 for a polar threonine residue.

The second mutant (Cys1819Phe) contains the substitution of the Zn^{2+} binding cysteine residue 1819 within the TAZ2 domain for a bulky, hydrophobic phenylalanine residue. This mutation is present within patient 4 in the clinical study by Menke et al. (2016).

The third mutant (His1829Asp) contains the substitution of the Zn^{2+} binding histidine residue 1829 within the TAZ2 domain for a negative aspartic acid residue. This mutation is present within patient 14 in the clinical study by (Menke et al., 2018). This mutant was unable to be produced, despite multiple attempts with SLIM methodology. Provided this mutant was likely to behave similarly to the Cys1819Phe and Cys1838Tyr, further work to produce this mutant was not prioritised and therefore this mutant was not included in any experimental work.

The fourth mutant (Cys1838Tyr) contains the substitution of the Zn^{2+} binding cysteine residue 1838 within the TAZ2 domain for a bulky, hydrophobic tyrosine residue. This mutation is present within patient 6 in the clinical study by Menke et al. (2016).

The fifth and final mutant (Arg1868Trp) contains the substitution of the positive arginine residue 1868 within the IDR4 for a bulky, hydrophobic tryptophan residue. This mutation is present within patients 17, 18 and 19 in the clinical study by Menke et al. (2018). Very similar substitution mutations of arginine 1867 and 1868 are also present in patient 20 from Menke et al. (2018), patients 7,8 9 and 10 in Menke et al. (2016) and patient 2 in Banka et al. (2019).

Table 1 Overview of CBP fragments used throughout this study with descriptions of contents, length and mutations present. All mutants were produced using site directed ligase independent mutagenesis with the plasmid containing the coding sequence for the TAZ2-IDR4 C4A CBP fragment used as the template. Therefore, all mutants also contain the C4A stability mutation of the TAZ2-IDR4 C4A fragment.

Name	Region	Mutations	Description
TAZ2	CBP 1763-1855	Wild type	Shorter WT contains no IDR4.
TAZ2 C4A	CBP 1763-1855	C1775A, C1783A, C1826A & C1827A	C4A mutation to non-zinc binding cysteines for increases stability and no IDR4
TAZ2-IDR4	CBP 1763-1876	Wild type	Longer WT contains initial helix forming region of IDR4
TAZ2-IDR4 C4A	CBP 1763-1876	C1775A, C1783A, C1826A & C1827A	C4A mutation to non-zinc binding cysteines for increases stability includes initial helix forming region of IDR4.
I1776T (TAZ2 mutant)	CBP 1763-1876	I1776T	Substitution of hydrophobic TAZ2 core residue for a polar residue.
C1819F	CBP 1763-1876	C1819F	Substitution of Cysteine residue that binds second Zn^{2+} ion.
C1819F	CBP 1763-1876	C1819F	Substitution of Histidine residue that binds third Zn^{2+} ion.

C1838Y	CBP 1763-1876	C1838Y	Substitution of Cysteine residue that binds third Zn^{2+} ion.
R1868W (IDR4 mutant)	CBP 1763-1876	R1868W	Substitution of a positively charged residue for a hydrophobic, bulky, aromatic residue.

Chapter 2: Materials & Methods.

2.1 General Methods

2.1.1 Autoclaving

Sterilisation of any equipment and media throughout this experimental work was carried out through autoclaving at 121ºC and $2x10^5$ Pa for 20 minutes.

2.1.2 Water

Ultrapure water was sourced from a NANOpure™ MilliQ system (Thermo Fisher Scientific).

2.1.3 Filter Sterilisation

Sterilisation of any liquids which are incompatible with autoclaving, were passed through either a 0.2 μm ReliaPrep ™ syringe filter (Ahlstrom) (<15 mL) or for larger volumes, a Millipore Express® PLUS 0.22 μm PES membrane mounted on an autoclaved glass bottle.

2.1.4 Sonication

In order to resolubilise cell pellets, to remove any potentially bound sample off vessel walls, or for mass spectrometry sample preparation a Fisherbrand™ Ultrasonic cleaning bath (Thermo Fisher Scientific) was used.

2.2 Microbiology

2.2.1 Bacterial Strains

Bacterial strains utilised throughout this work included *Escherichia coli (E. coli)* BL21 (DE3) for recombinant protein expression, and *E. coli* XL1 Blue or Top^{10} for plasmid maintenance.

2.2.2 Antibiotics

Ampicillin (Sigma) stock was made up to a concentration of 100 mg/mL using sterile MilliQ H_2O. 500 μL aliquots were stored at -20ºC in for later use. Aliquots

were thawed once, and any remaining was discarded and not refrozen for later use.

2.2.3 Media

Luria-Bertani (LB) (Sigma) agar plates were made up with MilliQ H_2O, LB broth at 25 g/L and bacteriological agar powder No. 2 (Neogen) at 1.5% (w/v). The LB agar liquid was sterilised by autoclave upon preparation and left to cool to 50ºC in a water bath. Upon cooling ampicillin was added at 100 μg/mL, and plates were able to be poured as 20 mL aliquots under a flame. Any air bubbles were removed using a Bunsen burner. Once plates were set, they were then stored inversely until use at 4ºC. 1 hour prior to use, the plates were transferred to a 37ºC incubator, still inverted, in order to remove condensation and pre-warm.

For liquid culture, LB (Sigma) was made up at 25 g/L with MilliQ H_2O and sterilised by autoclave. Ampicillin was then added at 100 μg/mL to desired cultures upon cooling to room temperature, just prior to inoculation.

2.2.4 Culture Optical Density

Cell culture optical density (OD) measurements were carried out using a SmartSpec™ Plus UV-visible Spectrophotometer (BioRad). Measurements were taken at a wavelength of 600nm (OD_{600}), with fresh, sterile media used as a blanking solution.

2.2.5 Preparation of Chemically Competent *E. coli*

A sterile toothpick swab of the desired *E. coli* strain, stored at -80ºC was inoculated into 5 mL of sterile SOB media (0.5% yeast extract, 2% tryptone, 25 mM KCl, 10 mM $MgSO_4$, and 10 mM $MgCl_2$), and left overnight to incubate at 37ºC on a shaking tray at 200 RPM. 50 mL of fresh, sterile SOB media was then inoculated with an aliquot of the starter culture for an initial OD_{600} of 0.05. The 50 mL culture was then incubated at 37ºC while shaking at 200 RPM and 37ºC for 2.5-3.5 hours or until and OD_{600} of 0.5 was reached.

The cells were then centrifuged in 50 mL Nunc™ tubes (Thermo Fisher Scientific) at 2,600 x g for 10 minutes at 4ºC. The supernatant was removed, and the cell pellet was resuspended in 16 mL of TB1 (75 mM $RbCl_2$, 30 mM potassium acetate, 10 mM $CaCl_2 2H_2O$, 15% Glycerol, 80mM $MnCl_2$, pH 5.8 with Acetic Acid) and incubated on ice for 15 minutes. The cells were again pelleted and resuspended in 16mL of fresh TB1. The cells were again pelleted and this time the supernatant was removed, and the cells were resuspended in 4mL of TB2 (7.5 mM $RbCl_2$, 10mM MOPS pH 6.8, 75 mM $CaCl_2 2H_2O$ and 15% glycerol). The cells were then snap frozen in liquid nitrogen and stored at -80ºC in 50 uL aliquots in sterile microcentrifuge tubes. Once aliquots were thawed for use, any remaining was discarded of appropriately, no repeated freeze thawing occurred.

2.2.6 Transformation & Starter Culture

Non-Expressing Host

2 μL of the 50 ng/μL plasmid stock was taken and mixed with 50 μL of chemically competent XL-1 Blue *E. coli* cells which were thawed on ice. 15 minutes after the addition of plasmid DNA to the cells, the mixtures were placed in 42ºC water bath for 42 seconds for heat shock treatment. The cells were then placed on ice for 15 minutes for recovery, prior to being plated on LB amp agar plates. All 52 μL of the cell solution was plated and spread under a flame or in a biosafety hood using a glass spreader sterilised with ethanol and flaming. The plates were left for up to 5 minutes to absorb the liquid before being placed inversely in the 37ºC incubator overnight. After incubation, sterile toothpicks were used to select singular colonies from the plates and they were inoculated into 5 mL of LB amp liquid broth. The liquid culture was then stored on a shaking tray at 200 RPM overnight in a 37ºC incubator.

Expression Host

For recombinant protein expression, the purified plasmid DNA was able to be transformed into chemically competent BL21 (DE3) *E. coli* cells. 50 μL of chemically competent BL21 (DE3) *E. coli* cells were first thawed on ice. 100 ng

of plasmid was then added and left on ice for 5 mins. The cells then under-went heat shock treatment at 42ºC for 42 seconds. The cells were then placed back on ice to recover for 5 minutes. The cells are then plated on LB amp agar and placed inverted into the 37ºC incubator overnight. Singular colonies were selected using sterile toothpick and inoculated into 5 mL of sterile LB amp and again incubated shaking at 200 RPM and 37ºC overnight, or until an OD_{600} of 0.5 was achieved.

2.2.7 Plasmid Purification

Plasmid DNA was purified from transformed XL-1 Blue *E. coli* liquid culture using the Zyppy Plasmid Miniprep Kit™ (Zymo Research), according to the manufacturer's instructions.

2.2.8 Culture Controls

As a positive control, untransformed cells were plated on non-amp containing LB agar. Cells were heat shock treated and recovered using the same method as transformed cells were, however 2 μL of H_2O was added in place of plasmid. Growth was expected on these plates. If no growth was seen, fresh competent cells were used in a new transformation.

As a negative control, non-transformed cells were plated on LB amp agar to ensure the cells were not ampicillin resistant in the absence of the plasmid, and that the ampicillin was active. Cells were heat shocked and recovered as transformed cells were, however 2 μL of H_2O was added in place of plasmid. If growth was seen on the negative control, this indicates contamination or expired/denatured antibiotics. In the occurrence of such, a fresh transformation was carried out with a new batch of competent cells and/or fresh ampicillin stocks were produced.

Negative controls were also used throughout the starter culture phase of growth. A liquid culture swabbed from the negative control agar plate was included with

ampicillin. If growth was seen at this stage, a fresh transformation was carried out and fresh LB was used.

2.3 DNA Methods

2.3.1 pET 15b Plasmid Stock Preparation

pET 15b plasmids containing the sequence encoding four variants of the human TAZ2 and IDR4 CBP fragments (table 1) were ordered from GenScript Biotech. 4 μg of the plasmid DNA was sent in lyophilised form. The 4 μg of DNA was dissolved in 80 μL TE buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.5), producing a stock with a concentration of 50 ng/μL which was stored at -20ºC for later use.

Table 2 Summary of WT CBP Fragments.

Name	Contents	Mutations
TAZ2	CBP 1763-1855	Wild type
TAZ2 C4A	CBP 1763-1855	C1775A, C1783A, C1826A & C1827A
TAZ2-IDR4	CBP 1763-1876	Wild type
TAZ2-IDR4 C4A	CBP 1763-1876	C1775A, C1783A, C1826A & C1827A

2.3.2 DNA Quantification

In order to determine DNA quantity and quality, a DeNovix DS-11 UV spectrophotometer (DeNovix) was used. 2 μL of sample buffer is used as a blank before loading 2 μL of sample. Measurements were taken at 260 nm and 280 nm.

2.3.3 DNA Agarose Gel Electrophoresis

TAE Gel

Tris-Acetate-Ethylenediaminetetraacetic acid (EDTA) (TAE) gels were utilised in order to visualise DNA after processes such as plasmid purification and plasmid amplification by colony PCR in order to analyse DNA yield and purity. 1% (w/v) TAE agarose gels were prepared using molecular grade agarose powder (Biolime) and 1x TAE buffer (50x TAE contained 2 M Tris HCl, 5 mM ETDA, 57.1 mL glacial acetic acid, pH 8 in 1 L H_2O). The solution was microwaved on medium-high setting for 3 minutes, stirring at 1 min intervals until boiling and all the agarose powder was dissolved. The SubCell GT Agarose gel electrophoresis system (BioRad) was used with 1x TAE as running buffer. Samples were loaded into the gel wells with 10x loading dye (0.2 g bromophenol blue and 6 mL 50% (v/v) glycerol in 4 mL pure H_2O). After loading, gels were run at 100 V for 1 hour.

TBE Gel

2% (w/v) Tris-Borate-Ethylenediaminetetraacetic acid (TBE) agarose gels were utilised for DNA visualisation in electrophoretic mobility shift assays (EMSAs). TBE gels were prepared using molecular grade agarose powder (Biolime) and 0.5x TBE buffer (10x TBE contained 108g Tris, 55g boric acid, 40 mL 0.5 M Na_2EDTA (pH 8.0) in 1 L H_2O). The solution was microwaved on medium-high setting for 3 minutes, stirring at 1 min intervals until boiling and all the agarose powder was dissolved. The SubCell GT Agarose gel electrophoresis system (BioRad) was used with 0.5x TBE running buffer. 30 minutes prior to gel loading, gels and running buffer were stored at 4º. Samples were loaded into gel wells with 10% (v/v) glycerol. After loading, the gels were run at 100 V for 1 hour in the 4ºC cold room.

Ethidium Bromide Staining

After electrophoresis TAE and TBE gels were stained with ethidium bromide (EtBr) solution (0.5 μg/mL) for 20 minutes, before being placed within MilliQ H_2O to destain for 10 minutes. Images of the gels where then captured under UV light using an UV Trans-illuminator Gel Doc system (BioRad).

2.3.4 Polymerase Chain Reaction

Polymerase chain reaction (PCR) was carried out in order to confirm the presence of the CBP fragment-containing plasmid within transformed BL21 (DE3) *E. coli,* this is also referred to as colony PCR. Toothpicks used to pick plates for liquid culture inoculation were dipped into sterile 2.5mL tubes containing PCR grade water, and then this water was boiled for 5 minutes for cell lysis to occur. 1 μL was then added to a thin wall 0.2 mL PCR tube along with 24 μL of PCR master mix (table 3).

Table 3 PCR Master Mix

Component	**Volume**	**Final Concentration**
10 mM dNTP	0.5 μL	200 μM
10 μM Forward Primer	1 μL	0.4 μM
10 μM Reverse Primer	1 μL	0.4 μM
Q5 HotStart Polymerase (New England BioLabs)	0.625 μL	0.625 units per 25 μL reaction
10x HotStart Buffer (New England BioLabs)	2.5 μL	1x
H_2O	18.9 μL	-
DNA template (cell lysate)	1 μL	<40 ng/μL
Final Volume	25 μL	

The samples were then placed within the thermocycler and the following reaction cycles were carried out (table 4).

Table 4 Colony PCR Thermocycle Protocol

Step	Temperature	Time	Cycles
Initial Denaturation	95ºC	30 seconds	-
Denaturation	95ºC	30 seconds	Repeat x30 cycles
Annealing	55ºC	60 seconds	
Extension	68ºC	60 seconds per Kb	
Final Extension	68ºC	5 minutes	-

2.3.5 Site Directed Ligase Independent Mutagenesis

In order to produce desired mutants for experimental assays, site directed ligase independent mutagenesis (SLIM) was carried out using the wild type (WT) CBP fragment 4 containing pET 15b plasmid as a template. The methodology, developed by (Chiu et al., 2008) is outlined in figure 9. Primer sets for the expression of 5 different mutant proteins (table 5) were designed. For each mutant a forward tail, a forward standard, a reverse tail and a reverse standard primer were required (Primers and resulting mutant cDNA sequences can be found in appendix). Lyophilised primers from Integrated DNA Technologies (IDT) were resuspended in nuclease free water to a final concentration of 100 μM. 10 μL aliquots of the primers were diluted 10-fold for a working concentration of 10 μM to be used in the SLIM amplification reaction (table 6 & 7).

Table 5 CBP mutants produced using SLIM

Name	Contents	Mutation
Mutant 1	CBP 1763-1876	I1776T
Mutant 2	CBP 1763-1876	C1819P
Mutant 3	CBP 1763-1876	H1829D
Mutant 4	CBP 1763-1876	C1838Y
Mutant 5	CBP 1763-1876	R1868W

Table 6 SLIM PCR Master Mix

Component	Volume	Final Concentration
5x Phusion buffer (Thermo Fisher Scientific)	4 μL	1x
Phusion Polymerase (Thermo Fisher Scientific)	0.2 μL	1 U/50 μL of reaction
10 mM dNTPs	0.4 μL	200 μM
5 ng/μL Template DNA	1 μL	0.25 ng/μL
25 mM $MgSO_4$	1.6 μL	2 mM
Nuclease free water	10.8 μL	-
Primers	2 μL (1 μL of each F & R)	1 μM (0.5 μM of each)
Final Volume	20 μL	-

Table 7 SLIM PCR Amplification Protocol (figure 9, step 1).

Step	Temperature	Time	Cycles
Hot start	95ºC	30 seconds	1x
Denaturation	95ºC	30 seconds	Repeat 25 cycles
Annealing	45-68ºC	30 seconds	
Extension	68ºC	30 seconds	
Final extension	68ºC	30 seconds	1x

Upon completion of the SLIM amplification, 1 μL of D*pn*I (10 U) and 5 μL of D buffer (20 mM $MgCl_2$, 20 mM Tris-Cl pH 8.0 and 5 mM dithiothreitol (DTT)) were added to each completed reaction in order to digest the template DNA within solution. This digestion was left to incubate at 37ºC for 1 hour. 10 μL of the two PCR products from each primer pair (figure 9, Fs & Rt with Ft & Rs) were then mixed together with 10 μL of H buffer (750 mM NaCl, 125 mM Tris pH 9.0, and

100 mM EDTA pH 8.0) and 20 μL of water. The mixtures were placed back within the thermocycler for SLIM hybridisation (table 8).

Table 8 SLIM PCR Hybridisation Protocol (figure 9, step 2).

Step	Temperature	Time	Cycles
Initial Denaturation	99ºC	3 minutes	
Denaturation	65ºC	5 minutes	Repeat x3 cycles
Reannealing	30ºC	40 minutes	

After hybridisation was complete, 10 μL of heteroduplex product was transformed into 50 μL of chemically competent XL1 blue *E. coli* via the standard protocol, plated on LB agar and grown overnight at 37ºC. Colonies were picked the next morning and grown in 5 mL of liquid LB amp, overnight at 37ºC. The following morning, cells were pelleted at 13,000 x g at 4ºC and lysed for plasmid purification. Purified plasmid DNA was then able to be sent for sequencing in order to confirm the presence of the desired mutation.

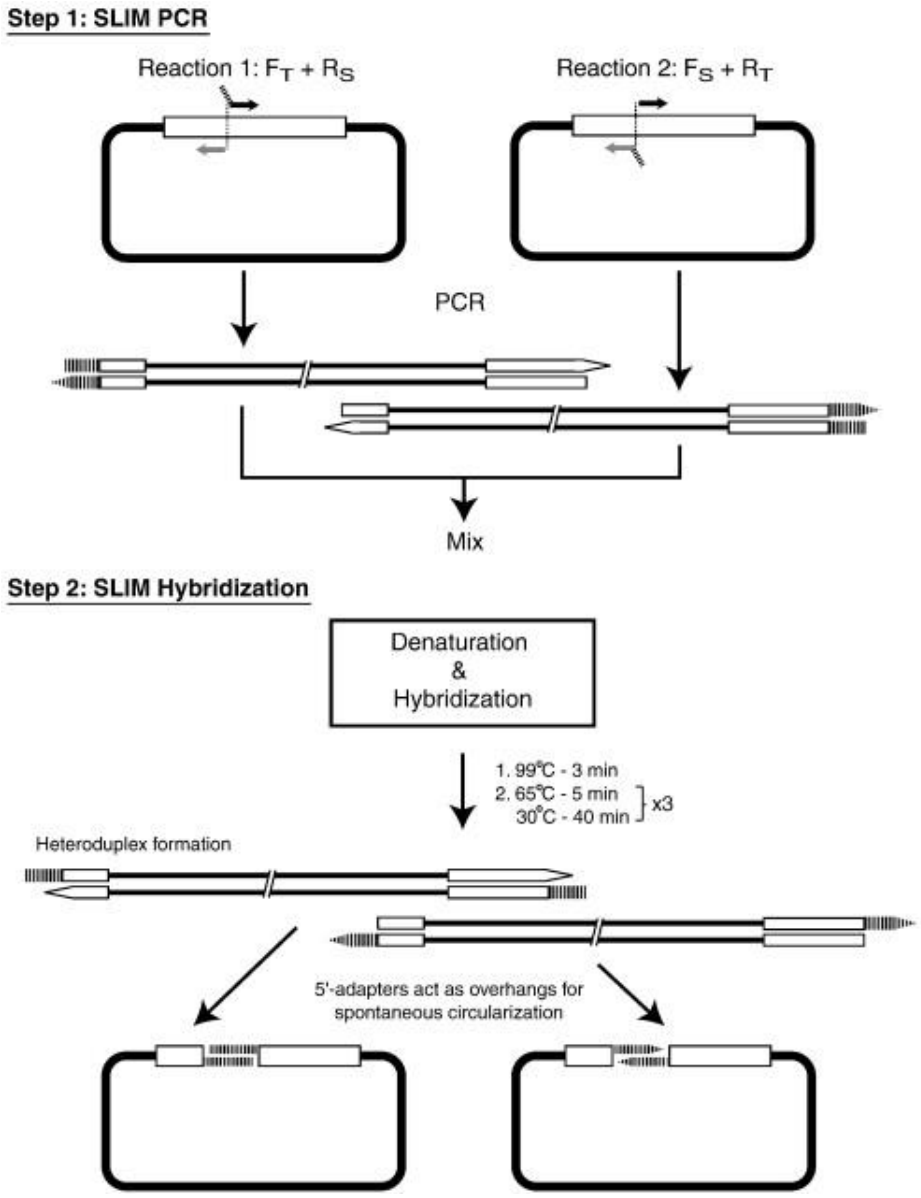

Figure 8 SLIM methodology overview. Step 1 involves amplifying template DNA with primers which have mismatch overhangs (Ft & Rt) that contains the desired substitution. (Ft) forward tail, (Rt) Reverse tail contain overhang sequence with the desired mutation, while (Fs) forward standard and (Rs) reverse standard are shorter and are identical to template DNA. The products of step one is double stranded DNA fragments which have overhanging ends of single stranded DNA that contain the desired mutation. Step two involves denaturing PCR products of step one and reannealing single strands with compatible ends to self-ligate into a circular plasmid, containing the mutation of interest. Image reproduced with permission (Chiu et al., 2008).

2.3.6 DNA Sequencing

All DNA sequencing was carried out at the Massey Genome Service Centre. 500 ng of each DNA sample to be sequenced was placed within sterile 0.2 mL PCR tubes along with 4 pM of T7 forward primer. Samples then had BigDye™ Sequencing Ready Reaction mix added by the Massey Genome Service centre prior to sequencing.

2.4 Protein Methods

2.4.1 Expression Culture

Small Scale Recombinant Protein Expression

Each expression began with a fresh transformation of competent BL21 (DE3) *E. coli*. 50 μL of freshly transformed cells were plated onto LB amp agar and grown inverted in a 37ºC incubator overnight. The next day, singular colonies were swabbed with sterile toothpicks and placed into a 3 mL LB amp starter culture, again at 37 ºC overnight on a shaker tray at 200 RPM. Cell density after overnight incubation was measured, and appropriate volumes of starter culture were transferred into 30 mL of fresh LB amp culture in order to obtain an initial OD_{600} of 0.05. The 30 mL cultures are then incubated at 37ºC for 3-4 hours on a shaking tray at 200 RPM, or until an OD_{600} of 0.5-0.6 was achieved.

Upon reaching desired optical cell density, protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to 1 mM concentration from 1 M IPTG stock. IPTG incubation time and cell growth temperatures were variable for small scale trial expression trials (chapter 3). In addition to IPTG, some samples also received $ZnCl_2$ at a 20 μM concentration during Zn^{2+} trials (chapter 3). $ZnCl_2$ was freshly prepared at 100 mM and added at the same time as IPTG. After the desired incubation period was complete, cells were then spun down for 15 minutes at 6,000 x g at 4ºC, washed with 1x phosphate buffered saline (0.8 g NaCl, 0.02 g KCl, 0.144 g Na_2HPO_4, 0.024 g KH_2PO_4, in 100 mL H_2O pH 7.4) (PBS) and frozen at -20ºC.

Medium Scale Recombinant Protein Expression

Medium scale expression trials also utilised sterile toothpick colony transfer into 5 mL starter cultures (>25 mL baffled flask used) and then after overnight incubation at 37ºC, transferred (as above in 2.4.1) into 300 mL expression cultures (>1 L baffled flask used). IPTG was added to a 1 mM final concentration from 1 M IPTG stocks, and again incubation time and temperature varied (chapter 3). Some samples also received $ZnCl_2$ at the time of IPTG induction at a concentration of 20 μM from a fresh 100 mM stock. After the desired incubation period was complete, cells were then spun down for 15 minutes at 6,000 x g at 4ºC, washed with 1x PBS and frozen at -20ºC.

Large Scale Recombinant Protein Expression

Large scale protein expression also utilised toothpick colony transfer into a 15 mL starter culture. 1 L expression cultures (>5 L baffled flasks used) were then inoculated after overnight incubation of the starter cultures, with an initial cell density of OD_{600} 0.05. The 1 L expression culture was then incubated at 37 ºC on a shaking tray for 3-4 hours or until an OD_{600} of 0.5 was achieved. The cells were then induced with 1 mM IPTG from 100 mM stock and $ZnCl_2$ was added from 100 mM stock to a final concentration of 20 μM. After the addition of IPTG, the cells were left to incubate at 20 ºC overnight on a shaking tray. After the desired incubation period was complete, cells were then spun down for 15 minutes at 6,000 x g at 4ºC, washed with 1x PBS and frozen at -20ºC.

2.4.2 Cell Lysis

French Press

Cell lysis for protein purification was carried out using a French press. Cells were harvested from expression culture by centrifugation at 6,000 x g at 4ºC for 15 minutes. Cell pellets were then resuspended in up to 15 mL of lysis buffer (20 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 10% glycerol, 2 mM tris(2-carboxyethyl) phosphine (TCEP), and 1x Pierce™ EDTA free protease inhibitor (Thermo Fisher Scientific). The resuspended cells were passed through the

French press cell (Aminco Instruments) twice with a hydraulic press (Wabash) applying a pressure of 5,000 kPa.

Sonication

After French press, samples were sonicated using a 1 mm tip in the Virtis Virsonic 600 Sonicator (SP Scientific) at 30% amplitude for 45 seconds while on ice. Samples were cooled on ice for 2 minutes before repeating the sonication again in order to shear DNA in solution. The cell lysate was then cleared by centrifugation at 23,000 x g for 30 minutes at 4ºC in order to pellet cellular debris. Finally, the cell lysis supernatant was passed through a sterile 0.8 μM filter before purification.

2.4.3 Protein Purification

Small Scale IMAC

Protein purification was carried out using immobilised metal affinity chromatography (IMAC). Bench top columns were made with 2 mL of iminodiacetic acid sepharose (Sigma Aldrich), charged with an excess of $NiCl_2$, until flow through ran green. The column was washed of any unbound $NiCl_2$ with 10 mL (5 column volumes) of H_2O before being equilibrated using 10 mL of equilibration buffer (300 mM NaCl, 20 mM NaH_2PO_4, 20 mM imidazole and 2 mM TCEP, pH 7.5). Lysed cell supernatant was then diluted 2-fold with equilibration buffer, then run through the column. The column was then washed with 5 mL of equilibration buffer, followed by 5 mL of wash buffer 1 (300 mM NaCl, 20 mM NaH_2PO_4, 50 mM imidazole and 2 mM TCEP, pH 7.5) followed by 5 mL of wash buffer 2 (300 mM NaCl, 20 mM NaH_2PO_4, 100 mM imidazole and 2 mM TCEP, pH 7.4). Finally, the protein was eluted using 300uL of elution buffer (300 mM NaCl, 20 mM NaH_2PO_4, 250 mM imidazole and 2 mM TCEP, pH 7.5). The fractions collected after elution were stored at 4ºC until analysed by SDS PAGE. Columns were reused by recharging them by chelating Ni^{2+} with 1 mL of 100 mM EDTA, washing with 1 mL H_2O and then recharging with $NiCl_2$.

Large Scale IMAC

Protein purification of medium and large-scale expression experiments utilised an NGC medium pressure chromatography system (BioRad). Equilibration buffer (300 mM NaCl, 20 mM NaH_2PO_4, 20 mM imidazole and 2 mM TCEP, pH 7.5) was used to equilibrate a prepacked HiTrap IMAC HP (2x 1 mL) column (Cytiva Life Sciences) prior to sample loading by washing with 5 column volumes. Samples were then loaded to the column at a flow rate of 0.2 mL/min. 20 column volumes of equilibration buffer then runs through the column at a rate of 1 mL/min once the sample was all bound in order to wash away unbound protein. Elution buffer (300 mM NaCl, 20 mM NaH_2PO_4, 500 mM imidazole and 2 mM TCEP, pH 7.4) was then added gradually to the column initially at a rate of 20% (80% equilibration buffer v/v), for 15 column volumes. Then 15 column volumes of 25% elution buffer, 75% equilibration buffer was used to wash the column. Finally, after the wash steps, elution buffer was gradually increased over a gradient until 100% elution buffer. Throughout, all flow through from the column after sample loading was collected as 1-2 mL fractions. Fractions were kept at 4ºC until analysis by SDS PAGE was complete.

2.4.4 SDS PAGE

Sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS PAGE) was carried out in order for protein analysis of a range of samples, from crude cell lysates to chromatography fractions. The resolving portion of the gels were typically 12-15% for optimal visualisation of the CBP fragments which are 13 kDa (table 6). A 4% stacking gel was placed on top of the resolving gel once set (table 7). Gels are left at room temperature for at least 1 hour to set. Gels were stored at 4ºC for up to two weeks if not used upon preparation.

Gels were run within Mini Protean system (BioRad) apparatus, with 500 mL of 1x running buffer (table 8). 20 μL of each sample is mixed with 5 μL 5x SDS page loading buffer (250 mM Tris-HCl pH 6.8, 10% SDS, 30% glycerol, 10 mM DTT and 0.05% bromophenol blue) before boiling for 60 seconds and then loading. Initially the gels were run at 100 V until the dye front passed the stacking gel,

then the voltage was increased to 200 V until the dye front reached the bottom of the gel.

Table 9 Polyacrylamide Resolving Gel

Component	**12%**	**15%**
H_2O	4.5 mL	3.75 mL
1.5 M Tris-HCl, pH 8.8	2.5 mL	2.5 mL
10% (w/v) SDS	0.1 mL	0.1 mL
40% Acrylamide	3 mL	3.75 mL
Mix and degas solution for 15 min		
10% Ammonium persulfate (w/v)	0.1 mL	0.1 mL
TEMED	5 μL	5 μL
Total volume	10 mL	10 mL

Table 10 Polyacrylamide Stacking Gel

Component	**4%**
H_2O	5.3 mL
0.5 M Tris-HCl, pH 6.8	2.5 mL
10% (w/v) SDS	0.1 mL
40% Acrylamide	1 mL
Mix and degas solution for 15 min	
10% Ammonium persulfate (w/v)	0.1 mL
TEMED	10 μL
Total volume	10 mL

Table 11 SDS PAGE Running Buffer

Component	Amount
Tris	15 g
Glycine	72 g
SDS	5 g
H_2O	Up to 1 L

Colloidal Coomassie Blue Staining

Once the SDS PAGE has been run, the gels were first placed in a staining tray submerged in fixative solution (40% MeOH, 10% Acetic acid, 50% H_2O) for 15 minutes on a shaker. The fixative is then decanted off, and the gels were resubmerged in freshly diluted Colloidal Coomassie stain (80% stock (table 11), 20% methanol) and placed back on the shaking tray and left overnight. The stain was then decanted off and the gels were washed with H_2O until the water discarded ran clear.

Table 12 Colloidal Coomassie Stock

Component	Amount
Coomassie Brilliant Blue G-250	0.25 g
85% Phosphoric Acid	3 mL
Ammonium Sulphate	25 g
Total volume	250 mL

2.4.5 Buffer Exchange

Chromatography fractions containing pure CBP fragment protein (analysed by SDS PAGE) were pooled and then underwent buffer exchange. Buffer exchange was carried out overnight using dialysis tubing. Pooled fractions were poured into

pre-soaked (300 mM NaCl, 20 mM NaH_2PO_4 and 2 mM TCEP, pH 7.5) dialysis tubing and then this was placed within at least a 10-fold volume of no imidazole IMAC buffer (300 mM NaCl, 20 mM NaH_2PO_4, and 2 mM TCEP, pH 7.5). A stirring flea was used while the sample was left at 4ºC overnight to dialyse. Any precipitate was pelleted by centrifuging the sample at 23,000 x g for 30 minutes at 4ºC.

2.4.6 Concentrating Samples

5 kDa molecular weight cut off Vivaspin® concentrators were utilised to concentrate any samples in order to be brought to appropriate concentrations for assays. Concentrators were centrifuged at 3,000 x g and 4ºC for 15-30-minute intervals until desired concentrations were achieved.

2.4.7 Protein Quantification

BCA Assay

Protein quantification was carried out using the Pierce™ Bicinchonic acid (BCA) assay kit (Thermo Fisher Scientific) and MultiSkan Go plate reader at 562 nm. The assay was carried out per the manufacturer's instructions. This method was utilised in instances such as determining protein concentrations before experimental assays (great accuracy required). Triplicates of each sample were used in these assays.

Bradford Assay

Bradford assays were utilised often to quickly quantify protein concentration of samples. 20 μL of each sample was mixed in a 96 well plate with 180 μL of 1x Bradford reagent (table 12) in triplicate. Protein standards were prepared using bovine serum albumin (BSA). 20 μL of each BSA standard and 180 μL of Bradford reagent were mixed in a 96 well plate in order to produce a protein concentration standard curve to compare protein samples with unknown concentrations against. MultiSkan Go plate reader was used to measure absorbance of samples at 595 nm.

Table 13 10x Bradford Reagent Stock

Component	Amount
Coomassie Brilliant Blue G250	100 mg
85% Phosphoric acid	100 mL

2.4.8 In Gel Digestion for Mass Spectrometry

In gel digestion was carried out on samples in preparation for mass spectrometry analysis in order to confirm the suspected bands within the SDS PAGE were indeed the CBP protein fragments. All of the following steps for the in-gel digestion protocol were carried out in a HEPA filtered laminar flow hood, and all reagents were prepared using LC-MS grade H_2O.

Destaining

After the gels were destained and imaged, the protein bands of interest were excised out of the gel using a scalpel and cut into small cubes before being placed within LoBind microcentrifuge tubes (Eppendorf). The gel cubes were then destained with 300 µL of freshly prepared 50 mM Ammonium Bicarbonate (ABC), 50% MeOH at 45ºC, for 2 hours. The destain liquid was then pipetted off, and 300 µL of 80% Acetonitrile (ACN) was added to dehydrate the gel cubes for 1 minute. The liquid was then pipetted off and the samples were placed in the SpeedVac™ centrifugal evaporator (Thermofisher) for 10 minutes to dry.

Reduction

To the dried gel pieces, 30 µL of fresh reducing solution (10 mM DTT and 50 mM ABC) was added and left to incubate for 1 hour at 37ºC, with periodic mixing. The supernatant was then removed, and the gel pieces were washed with 100 µL of 1x ABC for 5 minutes. The ABC wash was then removed before the addition of 100µL of 80% ACN for 1 minute. The ACN was then discarded and the gel pieces were placed within the SpeedVac™ (Thermo Fisher Scientific) again for 10 minutes.

Alkylation

30 μL of freshly prepared alkylation solution (20 mM iodoacetamide and 50 mM ABC) was added, and the samples were left to incubate in the dark for 30 minutes at room temperature. The supernatant was then removed, and the gel pieces were washed with 100 μL ABC for 5 minutes followed by 80% ACN for 1 minute. This wash and dehydration was then repeated once more before all liquid was removed from the gel pieces. The samples were then dried in the Speedvac again for 10 minutes.

Digestion

The gel pieces were swollen in 30 μL of digestion solution (20 ng/uL Trypsin and 50 mM ABC) while on ice for 10 minutes. The samples were then place in a 37°C incubator overnight.

Extraction

The samples where then sonicated for 2 minutes. The supernatant after sonication, was added to a fresh LoBind tube (Eppendorf). 50 μL of 5% formic acid, 50% ACN and 45% H_2O (v/v/v) was added to the gel pieces and they were sonicated for 2 minutes. The supernatant was collected and pooled with the previous supernatant. To the gel pieces, 50 μL of 0.1% formic acid, 80% ACN, 19.9% H_2O was added before 2 minutes of sonication. The supernatant was then taken and pooled again. Finally, the pooled supernatants were added to the Speedvac for 30-45 minutes until the volume was reduced to 30 μL. The samples at this point were frozen and stored at -80°C until sent to the Massey University Mass Spectrometry facility. Peptide analysis was carried out using a Q Exactive™ Focus Hybrid Quadrupole Orbitrap™ with a Nano Flex ion source (Thermo Fisher Scientific). Results were analysed using Proteome Discoverer™ Software (Thermo Fisher Scientific).

Chapter 3: Expression and Purification of TAZ2/IDR4 CBP Fragments.

3. CBP Fragment Expression & Purification Optimisation

Recombinant protein expression and purification protocols are all different and tailored to provide favourable conditions for each individual protein. Different proteins with different amino acid compositions have different requirements to be expressed in sufficient quantities and then to remain stable and soluble. Protein fragments are often used in experiments as a way of determining isolated activity or functions of a domain. It is for this reason, protein fragments are often expressed and purified in place of whole proteins.

Expressing protein fragments, as opposed to entire proteins would seem at first a more manageable task. However, removing surrounding domains can often have consequences. It may be that surrounding domains form critical interactions with the region trying to be expressed as a fragment. An additional challenge is the foreign prokaryotic environment and cellular machinery often utilised for recombinant protein expression. An isolated protein fragment, expressed by prokaryotic machinery into an unnatural cellular environment may be unexpressed, unstable, unable to fold, prone to aggregation, or all of the above (Zhang et al., 2004).

In attempts to prevent low rates of expression or production of insoluble protein, different expression and purification conditions can be trialled in order to determine a protocol which provides preferable conditions for the protein. There are many variables involved which have the potential to impact protein expression and purification, all of which fall under one of six main factors; expression host, vector, target DNA sequence, culture parameters, co-expression of aid proteins and purification conditions. Multiple of these variables were trialled in this study with the aim of establishing optimal conditions for the expression and purification of substantial quantities of the protein CBP fragments.

3.1 Features of an Expression System

There are some conditions which were predictably suitable for the expression and purification of the CBP protein fragments based upon previously published protocols of similar proteins. Utilising previous research is a time effective way of determining some aspects of the expression protocol. However, a number of variables were tested in an attempt to optimise yields of recombinant CBP fragments for analysis.

3.1.1 Host Organism

The most commonly utilised host organism for expression of recombinant protein, is *Escherichia coli* (*E. coli*) (Hemamalini et al., 2020). Bacterial expression systems are often used for recombinant protein expression due to being very cost effective, quick to culture and easy to genetically modify (Rosano & Ceccarelli, 2014; Yin et al., 2007). Many *E.* coli strains have been specifically developed in order to accommodate housing an expression vector and producing a foreign protein. The *E.* coli strain of choice for CBP fragment expression in this study was *E.* coli BL21 (DE3). The BL21 (DE3) strain contains a few specific genetic modifications in order to enhance its performance as a recombinant protein expression host.

Firstly, *E.coli* BL21 (DE3) contain a T7 RNA polymerase (T7RNAP) gene under the control of the leaky lacUV5 operator which has IPTG inducible expression (Du et al., 2021). Upon the addition of IPTG to cell culture, T7RNAPs are expressed and transcribe genes with T7 promoters at a rate eight times faster than other *E. coli* polymerases (Du et al., 2021). This makes it an ideal expression system for high levels of recombinant protein, when paired with a T7 site containing vector. Another key feature of *E. coli* BL21 DE3 is the natural deficiency of the *Lon* and *OmpT* genes. Both *Lon* and *OmpT* genes encode proteases which are known to degrade T7RNAPs and recombinant proteins, respectively (Gottesman, 1990).

E. coli BL21 (DE3) are a great host to begin protein expression trials with. They often provide an ideal environment for substantial expression of recombinant proteins which possess minimal rare codons and special requirements for folding. For this reason, this strain was used for the expression of the CBP fragments throughout this study.

3.1.2 Plasmid & Coding Sequence

The plasmid containing the coding sequence for the CBP fragment needs to be compatible with the expression host of choice. Additional components within the plasmid which aid the expression and stability of the recombinant protein fragment are equally as important as the genetic components in the expression host. The plasmid utilised to express the CBP fragments for this study was pET15b (figure 9).

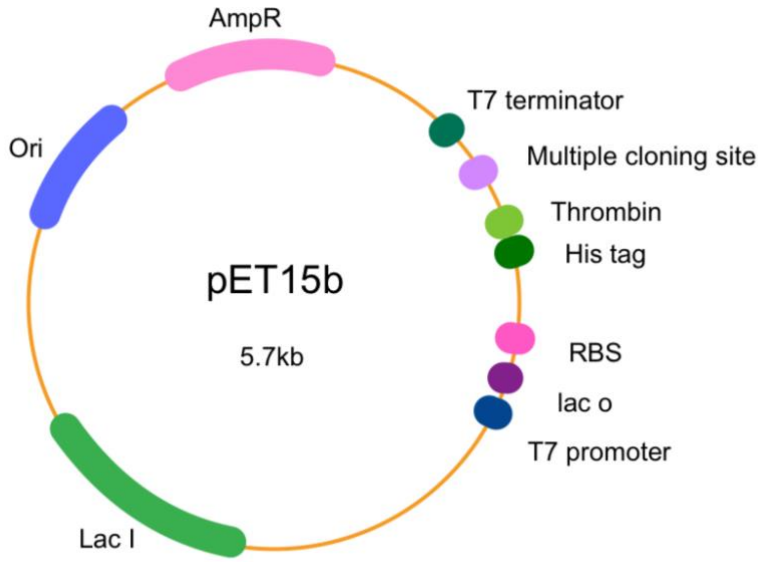

Figure 9 pET15b plasmid used in this study with CBP sequence inserted in multiple cloning site (MCS, lilac) to express CBP protein fragment (above). Within the plasmid is an ampicillin resistance gene (pink), origin of replication site (purple) and Lac I gene (teal). Flanking the MSC (lilac) upstream is the T7 promoter (navy), the Lac operator gene (purple), ribosome binding site (violet), the His tag sequence (olive) and the thrombin cleavage site sequence (lime). Following the MCS is the T7 terminator (green). Image produced using Goodnotes.

The first key feature of the pET15b plasmid, is the T7 promotor and termination sites which flank the multiple cloning site which houses the *CBP* insert (figure 9, navy and green). This means expression of the CBP fragment insert is carried out predominantly by T7RNAP which is provided by the BL21 (DE3) host. The T7RNAP operator lacUV5, is leaky and expression is near ubiquitous, but greater expression is inducible upon addition the of IPTG. This means upon IPTG induction T7RNAP is expressed, which consequently results in CBP fragment expression (Rosano & Ceccarelli, 2014). Another key feature of the plasmid is the sequence for a polyhistidine tag (histag) flanking the insert site (figure 9, olive). This means the CBP protein fragment will be expressed with six histidine residues at the N terminus, typically used for IMAC purification. Between the sequence for the six-residue long histag and the CBP coding region, is a thrombin cleavage site for removal of the tag after purification (figure 9, lime). Another key feature of the pET15b plasmid is the ampicillin resistance gene (figure 9, pink). This allows for selection of transformants within cell culture by addition of ampicillin to growth media.

The coding sequence of the *CBP* gene was also optimised for expression in a prokaryotic system. Rare or low usage codons were replaced with common *E. coli* codons, to ensure that translation was not limiting the quantities of CBP fragment produced.

3.1.3 Reducing Agent

A common cause of protein aggregation and insolubility in proteins with a large quantity of cysteine residues, is due to the formation of disulfide bonds which don't exist under standard conditions (Duong-Ly & Gabelli, 2014). The CBP fragments used during this study are composed of up to 115 residues in total, with some containing up to 13 cysteine residues (depending on mutations, see table 1). Prior to cell lysis the protein fragments are exposed to the reducing environment of the bacterial cytoplasm (Duong-Ly & Gabelli, 2014). However, upon cell lysis the environment becomes oxidizing. Two cysteines within close proximity under these conditions can be oxidised and form a disulfide bridge. The formation of non-native disulfide bridges between Zn^{2+} binding cysteine residues

may disrupt folding or cause aggregation of the protein fragments. It is common to add a reducing agent to buffers used from cell lysis onwards, in order to prevent these undesirable disulfide bonds from forming (Duong-Ly & Gabelli, 2014). Tris(2-carboxyethyl) phosphine (TCEP) was the reducing agent used throughout this study in buffers at a concentration of 2 mM in order to prevent unnatural disulfide formation. The concentration was limited to 2 mM to ensure protein purification or experimental techniques were not impaired.

3.1.4 Protease Inhibitor

An additional component which was not optimised by trials, was the addition of a protease inhibitor to lysis buffer. Another common problem in recombinant protein expression and purification is degradation of the recombinant protein by bacterial host proteases (Duong-Ly & Gabelli, 2014). Protease inhibitor cocktails prevent most protease activity. The addition of Pierce^{TM} EDTA free protease inhibitor (Thermo Fisher Scientific) to cell lysis buffer was another variable which was added without trial.

3.1.5 C4A Mutants

Throughout the optimisation trials, four different variants of CBP fragments were utilised (Table 14). The four fragments were all different by just a few residues. The first fragment is referred to as the TAZ2 fragment, and contains human CBP residues 1763-1855, encasing just the TAZ2 domain. The second fragment, termed TAZ2 C4A contained also just the TAZ2 domain, with the substitution of four non-zinc binding cysteine residues for four alanine residues (C1775A, C1783A, C1826A & C1827A). The third fragment, termed the TAZ2-IDR4 fragment is slightly longer (CBP 1763-1876), and contains the TAZ2 domain and the N-terminal region of the IDR4 (CBP 1852-1875), which has been shown to form a transient helical structure in research by Piai et al. (2016). This region of IDR4 (CBP 1852-1875) also contains nearly all of the IDR4 Menke Hennekam disease associated mutations. The final WT fragment referred to as the TAZ2-IDR4 C4A fragment and is the same length as the TAZ2-IDR4 fragment (CBP

1763-1876), with the same C4A substitution as the TAZ2 C4A fragment (C1775A, C1783A, C1826A & C1827A).

Table 14 Overview of CBP fragments used throughout this study with descriptions of region and mutations present.

Name	Region	Mutations	Description
TAZ2 Fragment	CBP 1763-1855	Wild type	TAZ2, contains no IDR4.
TAZ2 C4A Fragment	CBP 1763-1855	C1775A, C1783A, C1826A & C1827A	TAZ2 domain and no IDR4. C4A mutation to four cysteines (non-zinc binding Cys) for increased stability.
TAZ2-IDR4 Fragment	CBP 1763-1876	Wild type	Longer WT contains N-terminal helix forming region of IDR4
TAZ2-IDR4 C4A Fragment	CBP 1763-1876	C1775A, C1783A, C1826A & C1827A	Contains TAZ2 domain and N-terminal of IDR4. C4A mutation to four cysteines (non-zinc binding Cys) for increased stability.

This set of C4A mutations was first discussed by De Guzman et al. (2000), when they established the cysteine residues which were essential for folding of the TAZ2 domain by NMR studies. Upon substitution of four cysteine residues which were not involved in zinc binding and folding with four alanine residues, the addition of Zn^{2+} resulted in folding identical to the wildtype. The C4A mutants in this study were included as it was predicted based on these previous studies that these protein fragments may produce greater soluble yields of recombinant protein.

Expression and purification of all four CBP fragments appeared to be similar under all conditions throughout expression and purification trials. It was not until initial circular dichroism experiments were carried out (chapter 4.1) that it was established only the C4A fragments were folding. At this point the decision was made to utilise the C4A fragments as “wildtypes” relative to other mutant sequences throughout the remainder of the experimental work. Despite the circular dichroism results of the TAZ2-IDR4 fragment indicating minimal folding, sufficient soluble quantities were able to be produced to be included in some experimental work for ‘true’ WT comparison.

3.2 Expression Optimisation Trials

3.2.1 Incubation Trials

The first variable investigated by small scale trial was different incubation conditions for protein expression. This includes both incubation length and temperature, which is technically two variables however this is not without reason. An increase in incubation temperature after induction causes increased rates of protein expression (Vera et al., 2007). Over expression of recombinant proteins often results in aggregation and formation of inclusion bodies (Sahdev et al., 2008; San-Miguel et al., 2013; Vera et al., 2007; Zhang et al., 2004). Proteins are required to be soluble for structural and functional analysis, therefore conditions which bring about protein aggregation are avoided. It is for this reason that an increased incubation temperature is often paired with a shorter incubation period in order to prevent excess inclusion body formation.

Two different post IPTG induction incubation conditions were trialled; 37ºC for 3 hours and 20ºC for ~14 hours (overnight). Having a 3-hour incubation period saves time and is therefore preferred, however this trial was conducted to determine if the increased incubation temperature reduced the content of soluble protein. Duplicates cultures for all four protein fragments (table 14) were transformed, plated and grown in small scale starter and expression cultures (as per methods, chapter 2). It was upon IPTG induction that the duplicates were incubated in two different conditions. Cell lysis by sonication and small-scale purification was carried out, and an SDS PAGE was run to analyse results.

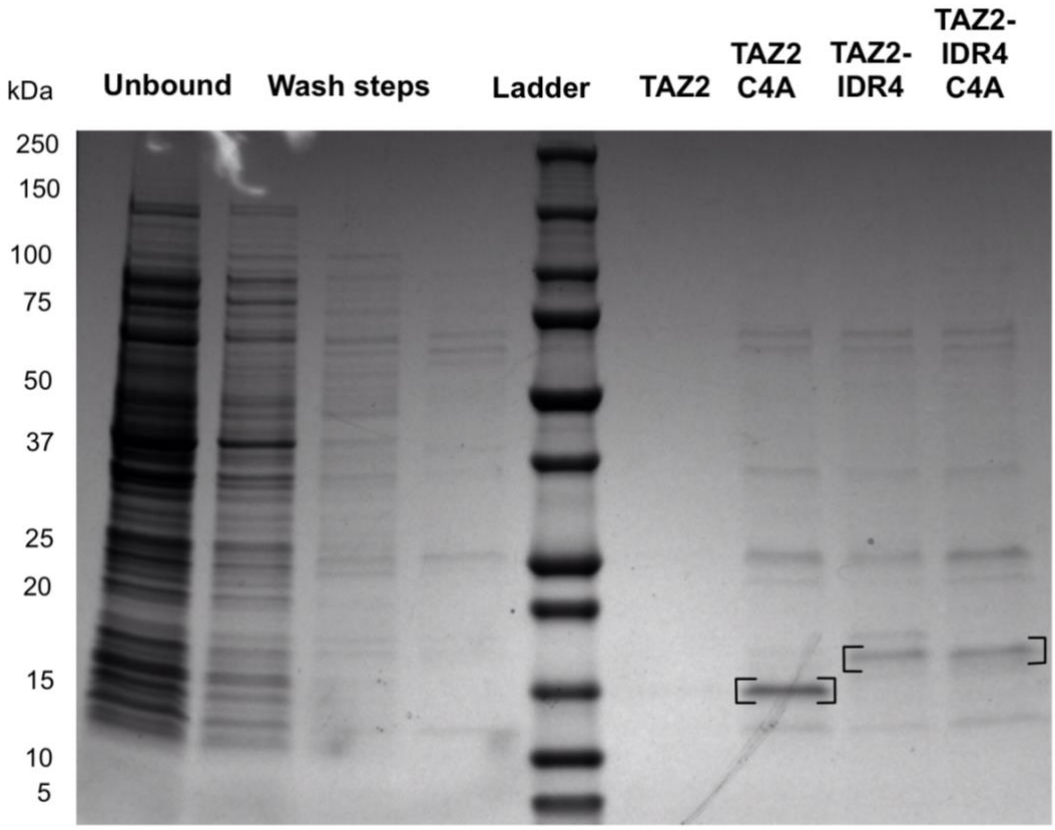

Figure 10 37ºC 3-hour incubation post IPTG induction. CBP fragment bands are encased by black brackets (~15 kDa). “Unbound” is the collected fraction after initial loading of the supernatant of lysed cell culture for the TAZ2 C4A fragment. “Wash steps” are the fractions collected during addition of wash buffers for the purification of the TAZ2 C4A fragment. “TAZ2”-“TAZ2-IDR4 C4A” are the elution fractions collected of each different CBP fragment (table 1). Expression of the TAZ2 fragment failed, likely due to human error at some point throughout protocol.

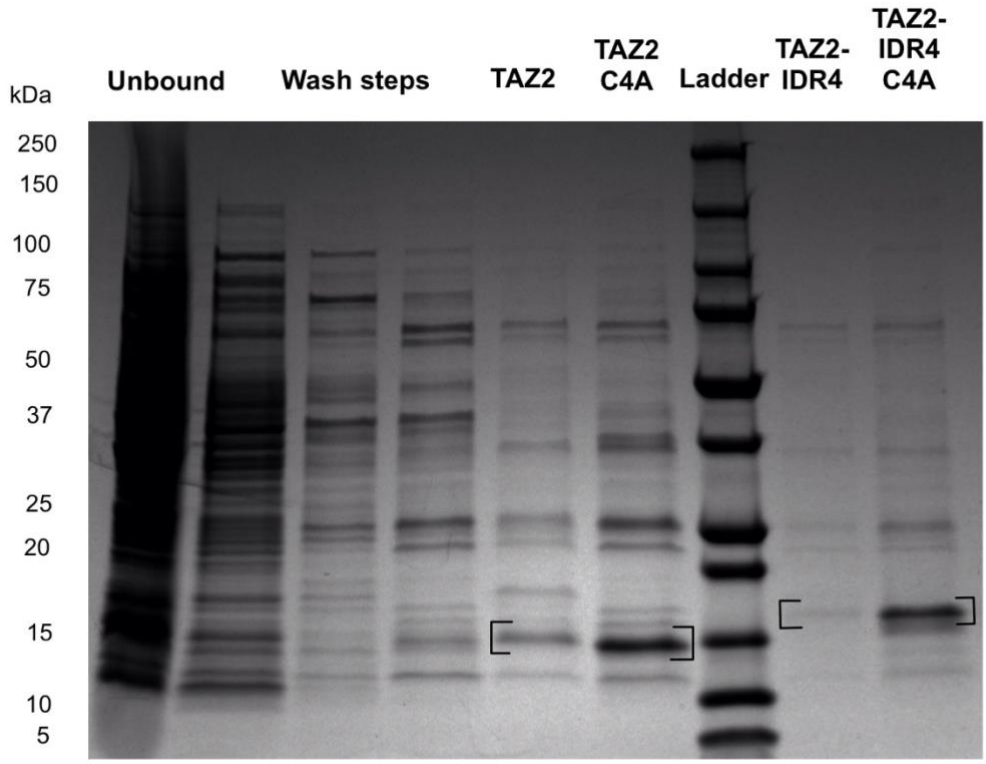

Figure 11 20ºC 14-hour incubation post IPTG induction. CBP fragment bands are encased by black brackets (~15 kDa). “Unbound” is the collected fraction after initial loading of the supernatant of lysed cell culture for the TAZ2 C4A fragment. “Wash steps” are the fractions collected during addition of wash buffers for the purification of the TAZ2 C4A fragment. “TAZ2”-“TAZ2-IDR4 C4A” are the elution fractions collected of each different CBP fragment (table 1). Expression of the TAZ2-IDR4 fragment failed, likely due to human error at some point throughout protocol.

SDS PAGE results (figures 10 & 11) alone are difficult to use to make a clear inference due to different sample concentrations loaded onto each gel. Therefore, GelAnalyzer software was used in order to compare the intensity of the CBP fragment bands relative to other bands consistently present within the elution fraction lanes (table 15). These values were used to produce a fold increase of CBP fragment within each lane compared to that of other proteins from the same culture. The two fragments which were not visible by SDS PAGE (The TAZ2

fragment 37ºC 3-hour incubation and the TAZ2-IDR4 fragment 20ºC 14-hour incubation**)** were not included within the calculations. It is unlikely due to the variable which was being trialled that no fragment is present, therefore they were excluded from the analysis.

Table 15 SDS PAGE band intensity values. Intensity is the intensity of CBP fragment band. Standard intensity is the average intensity of other proteins consistently present within all sample lanes (10, 15, 23, 25, 37 kDa) within the same lane. TAZ2 fragment (37ºC 3-hour) and TAZ2-IDR4 fragment (20ºC 14-hour) displayed very little protein upon SDS analysis and were therefore excluded from calculations.

Sample	37ºC 3-hour Intensity	37ºC 3-hour Standard Intensity	37ºC 3-hour Fold Increase	20ºC 14-hour Intensity	20ºC 14-hour Standard Intensity	20ºC 14-hour Fold Increase
TAZ2	-	-	-	74104	41710	**1.8**
TAZ2 C4A	55610	30236	**1.8**	144596	65229	**2.2**
TAZ2-IDR4	34182	26302	**1.3**	-	-	-
TAZ2-IDR4 C4A	31938	34237	**1.3**	79486	43755	**1.8**
Average fold increase			**1.5**			**1.9**

The SDS PAGE band intensity results from this trial confirm that although not a huge difference is visible between the different expression conditions, the CBP protein fragments are better expressed under cooler, slower growth conditions. Compared to the standard *E. coli* proteins (unknown, 10, 15, 23, 25 and 37 kDa) chosen to represent the other proteins within the samples, on average across the cultures a 1.91-fold increase of soluble CBP fragment expression was seen under the 20ºC growth conditions. This is compared to the 37ºC incubation producing a reduced 1.46-fold increase in soluble CBP fragment expression. Throughout the remainder of the study, cell culture was grown at 20ºC after the addition of IPTG for protein expression.

3.2.2 $ZnCl_2$ Trials

The second variable that was investigated by small scale expression trial was the addition of Zn^{2+} to the cell culture at induction in an attempt to aid in the folding of the CBP protein fragments. The TAZ2 domain of the CBP protein fragment folds into four alpha helices upon binding three Zn^{2+} ions (De Guzman et al., 2000). In hopes of increasing TAZ2 domain folding success and solubility of the expressed fragments, a final concentration of 20 µM $ZnCl_2$ was added to half of the duplicate cell cultures at the time of induction. The low concentration of zinc utilised in this trial is supported by the work carried out by Yao et al. (2005). They displayed that low zinc concentrations added to cell culture media increased *E. coli* growth, while higher concentrations had an inhibitory effect on growth. The work conducted by Sheahan et al. (2020) also used small concentrations (20 µM) of Zn^{2+} to supplement folding of the TAZ2 fragments used in their studies (CBP TAZ2 1752-1873).

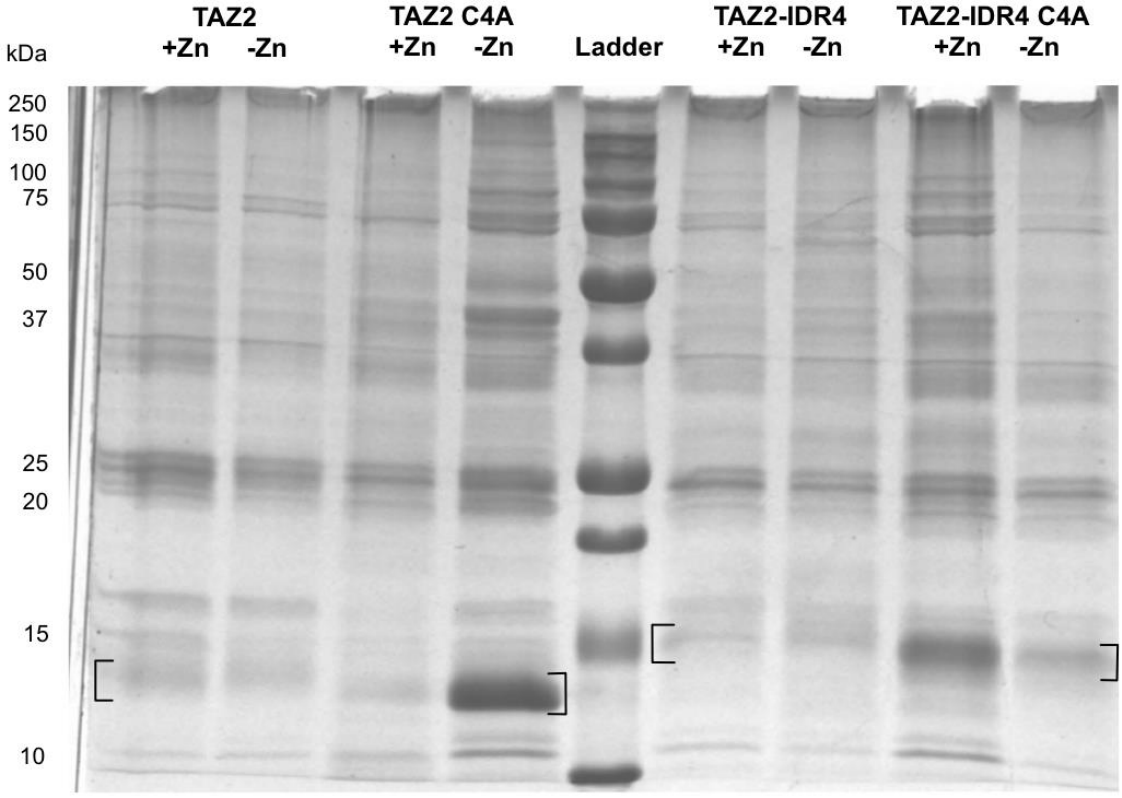

Figure 12 SDS PAGE Analysis Zn^{2+} Trial. "TAZ2" – "TAZ2-IDR4 C4A" are elution fractions from purified cell culture of CBP fragments (table 1). "+Zn" had $ZnCl_2$ added at IPTG induction and "-Zn" had no addition of $ZnCl_2$. Black brackets border CBP fragment bands. CBP fragments appear to have different migration, however the dye front has not migrated evenly, this is most noticeable at the bottom right side of the gel.

The results during $ZnCl_2$ trials were very inconsistent, and unable to confirm if the addition of Zn^{2+} was beneficial for solubility/protein stability (figure 12). There was no clear, consistent, significant difference between samples. This trial was repeated, with the same inconsistent and varied results. These trials did however confirm that the addition of zinc did not hinder the *E. coli* growth or CBP fragment expression or solubility. Upon upscaling the expression system, 20 µM of $ZnCl_2$ was added to cell culture at induction and the results displayed that very

substantial quantities of soluble CBP fragment was produced (figure 13 & 14). Therefore, it remained a part of the protocol throughout this study.

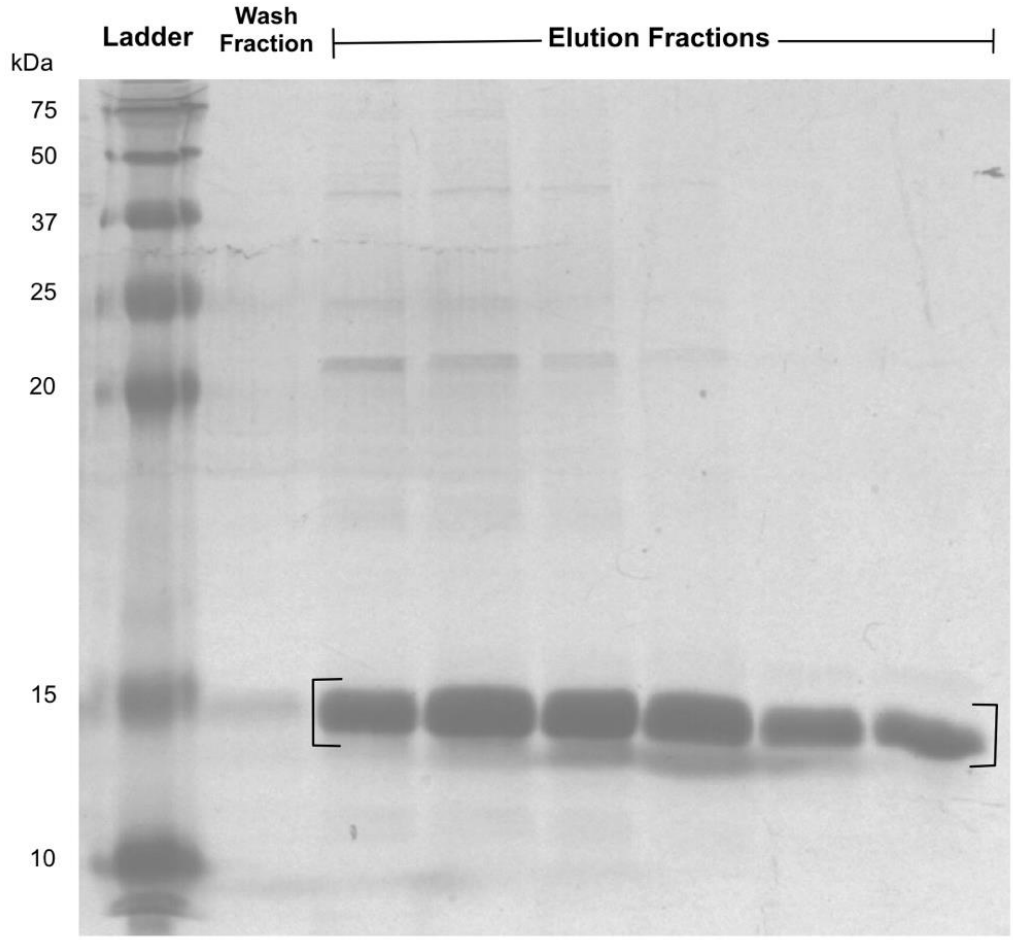

Figure 13 CBP TAZ2-IDR4 fragment Zn^{2+} elution fractions produced in an upscale trial with addition of 0.2 µM of $ZnCl_2$. Brackets encase CBP fragment (15 kDa).

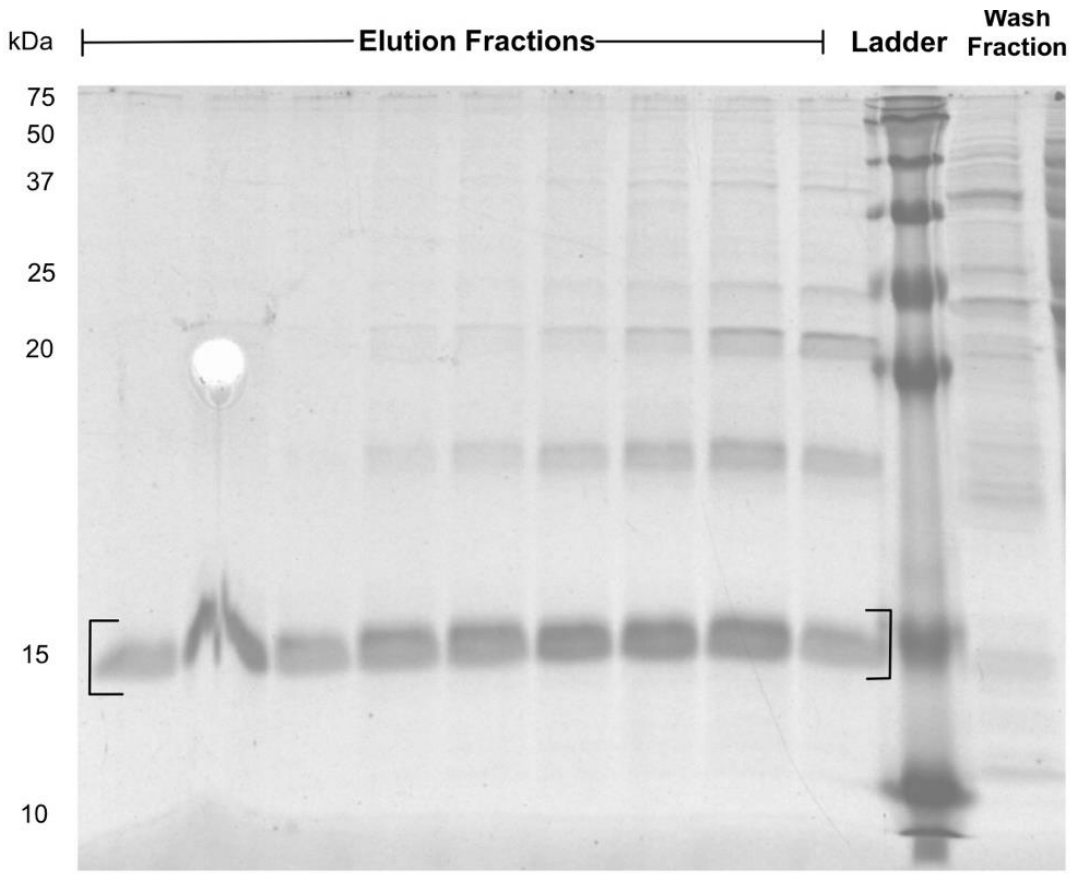

Figure 14 CBP TAZ2-IDR4 C4A fragment Zn^{2+} elution fractions produced in an upscale trial with addition of 0.2 µM of $ZnCl_2$. Black brackets encase CBP fragment (15 kDa).

3.2.3 Glycerol Stock Trials

Glycerol stocks of transformed cells can be stored and thawed at a later date in order to reduce the culture and expression procedure by a day. During an initial expression trial, an aliquot of starter culture was used to produce glycerol stocks, and then stored at -80ºC. At a later date, a sterile toothpick swab was then taken from the stock and inoculated into a fresh 3 mL starter culture broth. This was then transferred into an expression culture and induced with standard incubation periods. The cells were then lysed, only to discover by SDS PAGE there was no visible expression of the CBP fragments. This meant a fresh transformation and plating was required for every round of protein expression, as expression was not achieved to a great enough standard from glycerol stocks.

3.2.4 Purification Buffer Trial

Optimising the purification protocol for the CBP fragments involved trialling different buffers within an immobilised metal affinity chromatography (IMAC) system. All of the CBP fragments included in the study contained a histag at the N-terminus. Histags are commonly added to recombinant proteins in order to pair with an IMAC purification system. When in an environment where the histidine side chains of the tag are deprotonated, they are able to interact with immobilised Ni^{2+} ions which line the column that the sample passes through. Other proteins and contaminants within the cell lysate either don't bind to the column at all or are washed off by gradually changing the buffer flowing through the column. Changes to the buffer which can reduce the affinity between proteins and the column include; increasing the salt concentration, changing the pH or by introducing a competitive binding molecule. Eventually the change in pH, increase in salt content, or increase in concentration of the competitive binder in the buffer running through the column also elutes the target protein off the column.

The most common method for the elution of his-tagged proteins off an IMAC column is with the use of imidazole. Imidazole competes for the binding sites in the IMAC column, pushing off the his-tagged protein at high enough concentrations. It is typically used at low concentrations within the equilibration buffer in order to prevent any contaminant *E. coli* proteins binding weakly to the column. This helps to increase the purity of the final elution fractions. There can be adjustments made to the buffer system and column used, in order to get the most concentrated and pure elution fractions possible. For the purification of the CBP fragments in this study, these adjustments were determined by optimisation trials.

The initial buffer used for cell lysis and column equilibration needs to be lower in salt content and at a high enough pH where the histidine side chains of the tag are deprotonated. As the sample is passed through the column, the deprotonated side chains interact with the nickel which has a positive charge, holding the target protein within the column while others are washed away. Once sample binding

within the column has occurred, the buffer passing through the column changes either incrementally or over a gradient. The new buffer will have either an increased salt concentration, an increased imidazole concentration or a decreased pH. This will cause contaminants which are weakly bound to the column to be washed away by either being out competed for binding spots on the column or by becoming protonated and no longer attracted to the Ni^{2+} column.

In order to then elute the target protein, the buffer passing through the column is altered again. The first elution option is to decrease the pH (< 6.0) of the elution buffer in order to cause the histidine side chains of histag to become protonated, which will then reduce the attraction between the target protein and the Ni^{2+} column. This method will elute the protein from the column, however it may disrupt the folding of the target protein and cause precipitation. The second IMAC elution method is by increasing the imidazole concentration within the elution buffer to out compete the histag binding to the column.

Two methods were trialled with two different IMAC buffer systems. The first buffer system has a higher final imidazole concentration and is therefore referred to as the “high imidazole buffer system”. The second buffer system trialled uses greater salt concentrations during wash steps to remove contaminants from the column and contains less imidazole. This buffer system is referred to as the “high salt buffer system”. The high salt buffer system is used by Sheahan et al. (2020) in purifying different TAZ2 fragments.

In order to trial the two different buffer systems, one large cell culture was carried out with cells containing the CBP fragment 4 coding sequence. Cell lysis was carried out and the cleared lysate was divided equally and purified using the two different IMAC buffering systems. Fractions were then analysed by SDS PAGE (figure 16).

High Imidazole System

Equilibration buffer (300 mM NaCl, 20 mM NaH_2PO_4, 20 mM imidazole and 2 mM TCEP, pH 7.5) was used to equilibrate the prepacked 2x 1 mL HiTrap IMAC HP

column (Cytiva Life Sciences) prior to sample loading by washing with 5 column volumes. Samples were then loaded to the column at a flow rate of 0.2 mL/min. 20 column volumes of equilibration buffer then runs through the column at a rate of 1 mL/min once the sample was all bound in order to wash away unbound protein. Elution buffer (300 mM NaCl, 20 mM NaH_2PO_4, 500 mM imidazole and 2 mM TCEP, pH 7.4) was then added gradually to the column initially at a rate of 20% (80% equilibration buffer), for 15 column volumes. Then 15 column volumes of 25% elution buffer and 75% equilibration buffer was used to wash the column. Finally, after the wash steps, the elution buffer % is gradually increased over a gradient until reaching 100% elution buffer. Throughout the purification flow through from the column after sample loading was collected as 1-2 mL fractions. Fractions were kept at 4ºC until analysis by SDS PAGE was complete.

High Salt System

Low salt buffer (300 mM NaCl, 50 mM NaH_2PO_4, 20 mM imidazole and 2 mM TCEP, pH 8) was used to equilibrate the prepacked 2x 1mL HiTrap IMAC HP column (Cytiva Life Sciences) prior to sample loading by washing with 5 column volumes. Samples were then loaded to the column at a flow rate of 0.2 mL/min. 20 column volumes of equilibration buffer was run through the column at a rate of 1 mL/min once the sample was all bound in order to wash away unbound protein. High salt buffer (1 M NaCl, 50 mM NaH_2PO_4, 20 mM imidazole and 2 mM TCEP, pH 8) was then run through the column for 15 column volumes. Finally, 15 column volumes of the Sheahan et al. (2020) elution buffer (300 mM NaCl, 50 mM NaH_2PO_4, 250 mM imidazole and 2 mM TCEP, pH 8) was then used to elute off the CBP fragment.

SDS PAGE analysis showed that both systems produced fractions with satisfactory amounts of the CBP fragment, however the high imidazole buffer produced fractions with greater purity (lanes 2-5, figure 16). The high imidazole buffer system was therefore used throughout the remainder of the study (chapter 2).

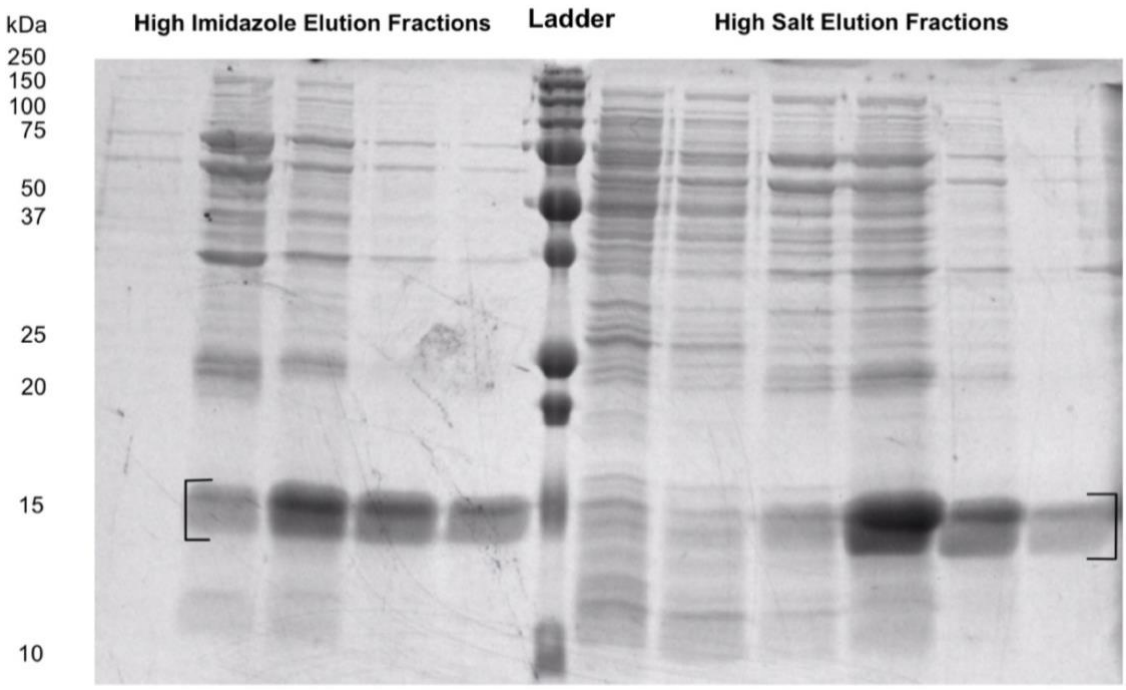

Figure 15 IMAC Buffer Trial comparing (left) high imidazole buffer system and (right) high salt Sheahan et al. 2020 buffer system for purifying the CBP TAZ2-IDR4 C4A fragment (table 1).

3.2.5 Sample Concentration & Buffer Exchange

Sample concentration and buffer exchange of the purified protein solutions proved to be a difficult step to optimise. Initially, sample concentration and buffer exchange of pooled, pure CBP fragment fractions was attempted using Vivaspin® concentrators with a 5 kDa molecular cut off. The concentrators were centrifuged at 3,000 x g and 4ºC in 30-minute intervals until the samples were reduced to approximately 20% of the original volume. The concentrators were then topped up with desired buffer (BC1000 buffer 100 mM KCl, 20 mM HEPES, 2mM TCEP, pH 7.5 or assay buffer 10 mM KCl, 20 mM HEPES, 2mM TCEP, pH 7.5) for buffer exchange by centrifugation. During this process at multiple stages, throughout various attempts, the protein became insoluble and precipitated. Protein concentration was monitored throughout (chapter 2.4.5-2.4.6) and never

surpassed 1.5 mg/mL. This indicates that the CBP fragment may be soluble at high concentrations, however it is not stable within a lower salt environment.

Owing to these difficulties it was decided that sample concentration would instead only be carried out if necessary, just prior to experimental assay. Buffer exchange would instead be carried out with a moderate salt buffer (20 mM NaCl, 300 mM NaH_2PO_4 and 2 mM TCEP, pH 7.5) instead of a low salt buffer in order to try keep the protein in solution. In addition to a moderate salt buffer, the use of dialysis tubing for buffer exchange was also trialled in case membrane associated aggregation was occurring as a result of the centrifugal concentration method. Overnight dialysis at 4ºC with pooled, pure CBP fragment fractions was carried out. This still produced a small amount of precipitate, however the passive, slower process produced greater yields of soluble CBP fragment. The precipitate was centrifuged at 23,000 x g into a pellet and the supernatant was kept at 4ºC until use. A trial was also conducted to determine if this pellet would resolubilise. The pellet was resuspended in moderate salt buffer and left overnight at 4ºC. The sample failed to resolubilise and remained as a visible precipitate.

After expression and purification optimisation was carried out, the quantities obtained for all CBP WT fragments were satisfactory (on average 2.5 mg of each CBP fragment per large scale expression culture). The experimental techniques utilised throughout this work required pure CBP fragment of moderate concentration. This was achievable with the protocols presented. Further optimisation is required for experimental techniques which require more concentrated sample or low salt conditions.

Upon optimisation of the expression and purification procedure of the CBP WT and C4A WT fragments, site directed ligase independent mutagenesis was carried out using the TAZ2-IDR4 C4A plasmid as a template to produce specific mutants (table 1). The mutants were assumed to behave similarly enough to the WT and C4A WT CBP fragments, that expression and purification of these fragments could be successfully carried out using the same protocol.

Chapter 4: Structural Analysis of TAZ2/IDR4 CBP Fragments.

4.1 Circular Dichroism

Structural and stability analysis of the CBP fragments were carried out in order to initially determine if the recombinant protein fragments were natively folded. Further studies were then carried out to determine of there were any differences in the structure and stability of the fragments which contained the Menke Hennekam disease associated mutations. The information provided by these structural experiments are essential to support and reinforce discussions made about the recombinant fragment activity during DNA binding experiments (chapter 5).

Circular dichroism (CD) spectroscopy is able to provide information about protein structure by the differential absorption of circular polarised ultraviolet (UV) light which passes through a cuvette holding the purified protein sample (Greenfield, 2006). The circularly polarised light is absorbed by the chromophores of the protein molecule; such as the peptide backbone bonds (far UV) or aromatic side chains (near UV) (Kelly & Price, 2000). Circularly polarised light is chiral and exists in two forms, left-handed and right-handed. When the circular polarised light passes through the sample chamber, absorption of the chiral light forms will be different due to the asymmetric nature of protein molecules (Greenfield, 2006). The data is collected as the difference in absorbance ($\Delta\varepsilon = \varepsilon_l - \varepsilon_r$) of the left ($\varepsilon_l$) and right ($\varepsilon_r$) handed light. This is presented commonly as degrees ellipticity ($\Delta\varepsilon * 32.98$). Where change in absorbance ($\Delta\varepsilon$) is the difference between the absorption of right handed (ε_r) and left-handed light (ε_l), which is multiplied by a standard factor of 32.98 (Greenfield, 2006). This can be converted to molar ellipticity which considers the molar mass, concentration and length of the light path through the sample (deg*cm^2/dmol^{-1}). This is plotted against wavelength and produces a spectrum characteristic for the fold of the protein.

The wavelength range between 180-260 nm (far UV) provides information about the conformation of the peptide backbone. Different secondary structures provide characteristic spectra within this range (figure 16). Alpha helical secondary structure spectra contain characteristic dips at 208 nm and 220 nm, and a peak

at 193 nm (figure 16, black trace). Beta pleated sheets produce spectra with a dip at 218 nm and a peak at 195 nm (figure 16, red trace). Disordered proteins produce far UV spectra with a dip at 195 nm and overall low signal intensity (figure 16, green trace).

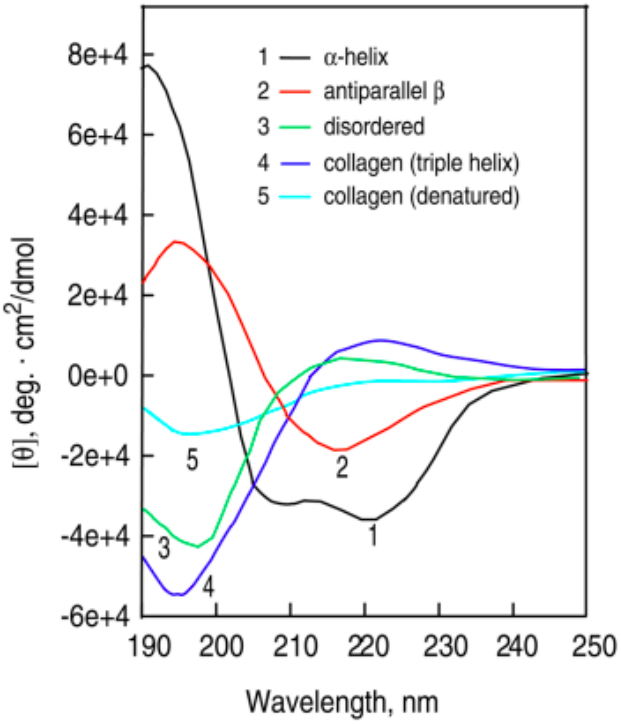

Figure 16 Characteristic CD spectra for different protein secondary structures. Alpha helix (black), beta sheet (red), disordered (green) collagen triple helix (blue) and denatured collagen (cyan). Figure reproduced with permission (Greenfield, 2006).

CD absorbance between 260-350 nm (near UV) provides information about the tertiary fold of the protein. Light within this wavelength range is absorbed by any disulphide bonds and the aromatic amino acids of the protein (Kelly & Price, 2000). Tryptophan residues produce a peak at 290 nm, tyrosine residues produce peaks at 275 nm and 282 nm, phenylalanine residues produce peaks at 255 nm and 270 nm and disulphide bonds produce a peak at 260 nm. Each protein will have a unique near UV spectra, reflective of the different tertiary folding and

aromatic composition (Kelly & Price, 2000). This means near UV spectra are unable to provide information about an unknown protein structure or specific residues. However, it is useful for determining if mutations cause any differences in structure, folding and stability.

CD spectra of either the near or far UV can be utilised in order to investigate protein stability. For stability studies CD spectra are collected over increasing temperatures or upon addition of increasing concentrations of denaturing agents. Protein unfolding upon exposure to these conditions is measurable by a change in signal intensity. This is useful for investigating the stability of different mutants.

4.1.1 CD Methodology

Structural and stability analysis of the CBP protein fragments was carried out using a Chirascan CD Spectrophotometer (Applied Photophysics). Samples (0.6 mg/mL) were loaded into a 0.1 mm or 10 mm precision quartz cell (Hellma Analytics) and placed within the temperature-controlled sample compartment which was purged with O_2 free N_2 for a minimum of 20 minutes before use. Sample buffer also had a spectrum measured to be used as a baseline (20 mM NaCl, 300 mM NaH_2PO_4 and 2 mM TCEP, pH 7.5). Structural analysis readings were taken at 0.5 nm increments with 0.5 seconds per point and 1 nm bandwidth. Stability, or thermal unfolding analysis also used 0.5 nm and 10 ºC increments, 1 nm bandwidth and each temperature was held for 60 seconds before taking measurements for 0.5 seconds. Processing of the raw data was done using the Chirascan software (Applied Photophysics) and included averaging the 8 reads of each sample, subtraction of the baseline, smoothing and conversion of units from millidegrees to molar ellipticity to account for sample concentration, molecular weight and cell path length.

4.1.2 CD Secondary Structure Analysis of CBP Fragments

The first investigation carried out was the comparison of secondary structure between the wild type fragments and MHS associated mutant fragments at room temperature (figure 17).

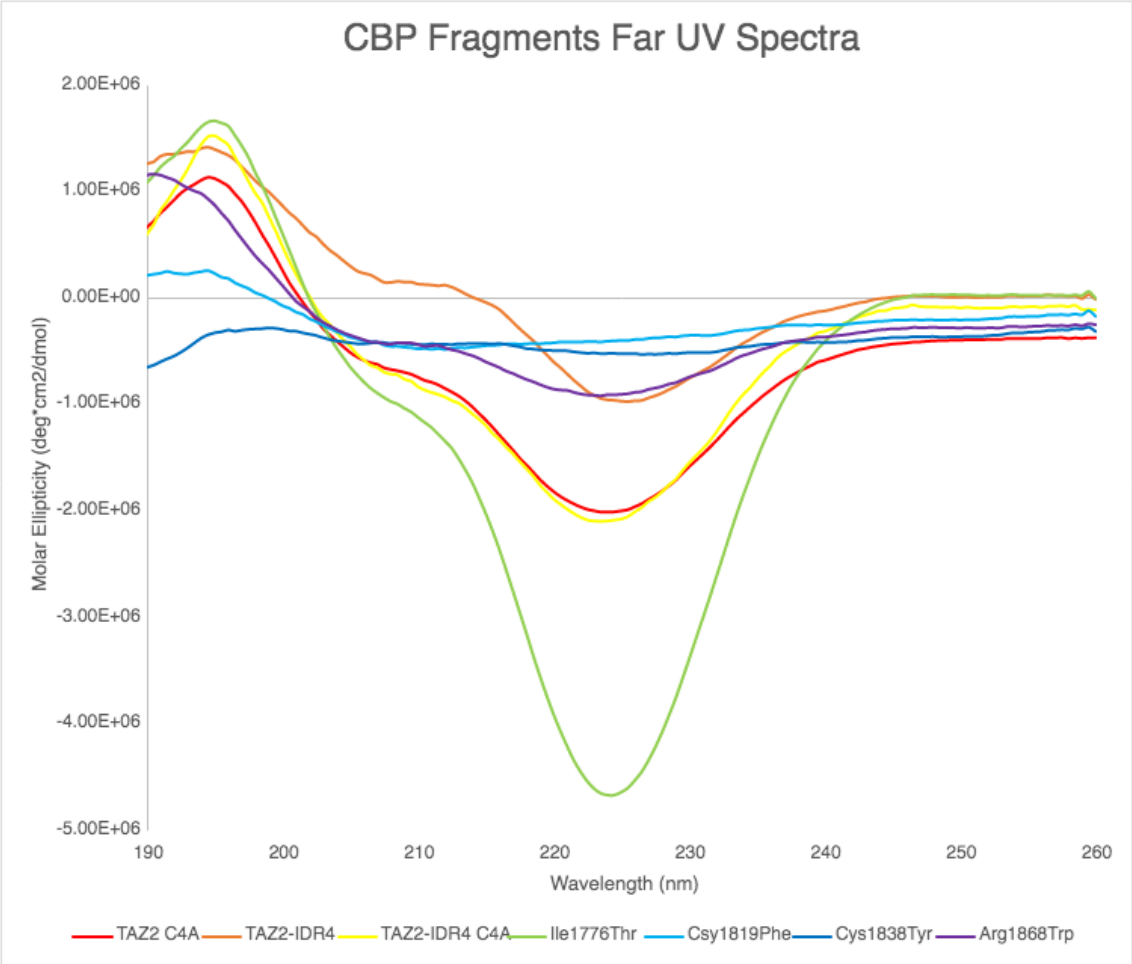

Figure 17 Circular Dichroism Far UV Spectrum of CBP fragments at room temperature. TAZ2 C4A (red), and TAZ2-IDR4 C4A (yellow) are C4A CBP fragments, TAZ2 (orange) is true CBP WT fragment. Ile1776Thr (green) contains an Ile1776Thr substitution in the TAZ2 domain. Cys1819Phe (blue) and Cys1838Tyr (navy) have zinc binding cysteine residue substitutions within the TAZ2 domain. Arg1868Trp (purple) has Arg1868Trp substitution within the IDR4. All mutant fragments additionally contain the C4A substitutions (see table 1).

Wild Type CBP Fragment Far UV Analysis

Wild type CBP fragments TAZ2 C4A and TAZ2-IDR4 C4A, contain the C4A substitution (table 1) and are not true wild types, wild type fragment TAZ2-IDR4 is the true CBP fragment wild type. The far UV spectra indicates that the TAZ2

C4A fragment (figure 17, red) and the TAZ2-IDR4 C4A (figure 17, yellow) appear to have almost identical secondary folding, while the TAZ2-IDR4 fragment (figure 17, orange) in comparison produces very weak signal intensity, indicating a lack of any folding or structure within the protein fragment. The solubility of a protein fragment can sometimes be an indicator towards folding success of the protein. When proteins fail to fold into their native state and instead aggregate together it is often visible as precipitate. It was difficult to produce and purify sufficient quantities of soluble TAZ2 and TAZ2-IDR4 fragments (true WT fragments), compared to the TAZ2 C4A and TAZ2-IDR4 C4A fragments. It was therefore assumed that there was a lack of correct folding in the TA2 and TAZ2-IDR4 CBP fragments which resulted in aggregation.

Enough soluble TAZ2-IDR4 fragment was produced to include in the structural analysis in order to investigate this assumption. The TAZ2-IDR4 fragment spectra displayed a lack in secondary structure, with weak signal intensity throughout. The TAZ2 C4A and TAZ2-IDR4 C4A fragments produce spectra which are characteristic of alpha helical secondary folding, with minima at ~208 nm and 220 nm, and a peak at ~193 nm. This is the folding which is expected of these fragments, as they include the entire TAZ2 domain which has been previously shown to form structures with four alpha helices upon Zn^{2+} binding (De Guzman et al., 2000).

There are only minimal differences visible between the TAZ2 C4A fragment and TAZ2-IDR4 C4A fragment spectra, despite the TAZ2-IDR4 C4A fragment containing an extra 21 residues. These 21 residues compose the N-terminal region of the IDR4 domain. These additional residues have been shown to form transient helical structures by NMR spectroscopy (Piai et al., 2016). However negligible differences between the CD spectra of the TAZ2 C4A fragment and the TAZ2-IDR4 C4A fragment suggests that under the experimental conditions provided, no helical structures were present in the IDR4 region of the TAZ2-IDR4 C4A fragment. If helical structures were forming within the additional region of the TAZ2-IDR4 C4A fragment, an increased minimal signal intensity would have

been visible at the characteristic alpha helical regions of the spectra (~208, 220 and 193 nm) compared to that of spectra produced by the TAZ2 C4A fragment.

CBP Menke Hennekam Mutant Fragments Far UV Analysis

Expression and purification of the CBP TAZ2 mutant (Ile1776Thr) fragment and the CBP IDR4 mutant (Arg1868Trp) fragment produced significant amounts of soluble protein. In contrast, a majority of the CBP Cys1819Phe mutant and CBP Cys1838Tyr mutant fragments (which contained Zn^{2+} binding cysteine substitutions) was insoluble at cell lysis and present within the cell pellet fraction.

The CBP Ile1776Thr mutant contained a substitution of a hydrophobic residue (Ile) within $\alpha 1$ of the TAZ2 domain with a polar residue (Thr). At a tertiary structure level, it was anticipated that this may reduce the packing of the helices which form the hydrophobic core of the TAZ2 domain with the introduction of the hydrophilic residue. However, change to the secondary structure with the introduction of this substitution was unexpected. The far UV spectra for this fragment displays a significantly increased signal intensity at the points characteristic for alpha helical secondary folding (~208, 220 and 193 nm). This may be a result of the polar residue introduced favouring high salt environment the fragment was exposed to, and less prone to aggregation and non-native folding.

The Cys1819Phe and Cys1838Tyr fragments contain substitution mutations for cysteine zinc binding residues to bulky aromatics. It was anticipated that these substitutions would prevent the TAZ2 region of the protein fragment from folding into the native helical structures which form upon Zn^{2+} binding. The weak signal intensity seen in the far UV CD spectra for these fragments agrees, and suggests minimal secondary folding is present in these fragments.

The Arg1868Trp IDR4 mutant fragment contains the substitution of a positive arginine residue at position 1868 for a hydrophobic aromatic tryptophan residue. It was anticipated that this substitution might have minimal impact on the secondary structure of the protein fragment as there is a lack of folding typically

seen in the IDR4. However, the far UV CD spectra of this mutant displayed a decreased signal intensity compared to the TAZ2-IDR4 C4A WT fragment. This suggests that the IDR4 mutation somehow hinders the ability of the TAZ2 domain within the fragment to fold. In contrast to the Ile1776Thr mutant, the non-conservative introduction of the hydrophobic residue may be enough to lower the solubility of the CBP fragment within the experimental buffer and cause aggregation and decreased native folding.

4.1.3 CD Tertiary Structure Analysis of CBP Fragments

Tertiary structure spectra were unable to be obtained for the CBP fragments. A variety of factors contributed to the failure to produce any quality near UV spectra for the CBP fragments. Tertiary structure is able to be investigated with near UV CD by the absorption of circularly polarised light by aromatics and disulfide bonds of the protein sample. Of all 115 residues which compose the CBP fragments, there is only one aromatic tyrosine residue in the WT fragments and one additional aromatic in three of the mutant fragments (phenylalanine, tyrosine and tryptophan). The TAZ2 domain does not contain any disulfide bonds, and the presence of the reducing agent (TCEP) also prevents the formation of any non-native disulfide bonds. The lack of aromatic residues and disulphides within the CBP protein fragments means that a greater sample concentration is likely required to produce a signal within the near UV spectra. Both 0.1 mm and 10 mm quartz cells (Hellma Analytics) were trialled, with protein concentrations of up to 3 mg/mL with no signal produced. Further optimisation of the purification process may be required in order to produce protein samples of greater concentration, in a lower salt buffer to produce signal within the near UV for tertiary structure analysis to be conducted.

4.1.4 CD Stability Analysis of CBP Fragments

Stability analysis of the CBP fragments was carried out utilising far UV CD. Spectra of the protein fragments were taken over increasing temperatures, in order to compare the thermal unfolding process of each fragment, and if the substitution of different residues associated with Menke Hennekam syndrome

alters the stability of the fragment. Spectra were obtained at 10ºC increments beginning at 20ºC (where the fragments are expected to be predominantly natively folded), through to 80ºC (where the fragments are expected to be completely unfolded).

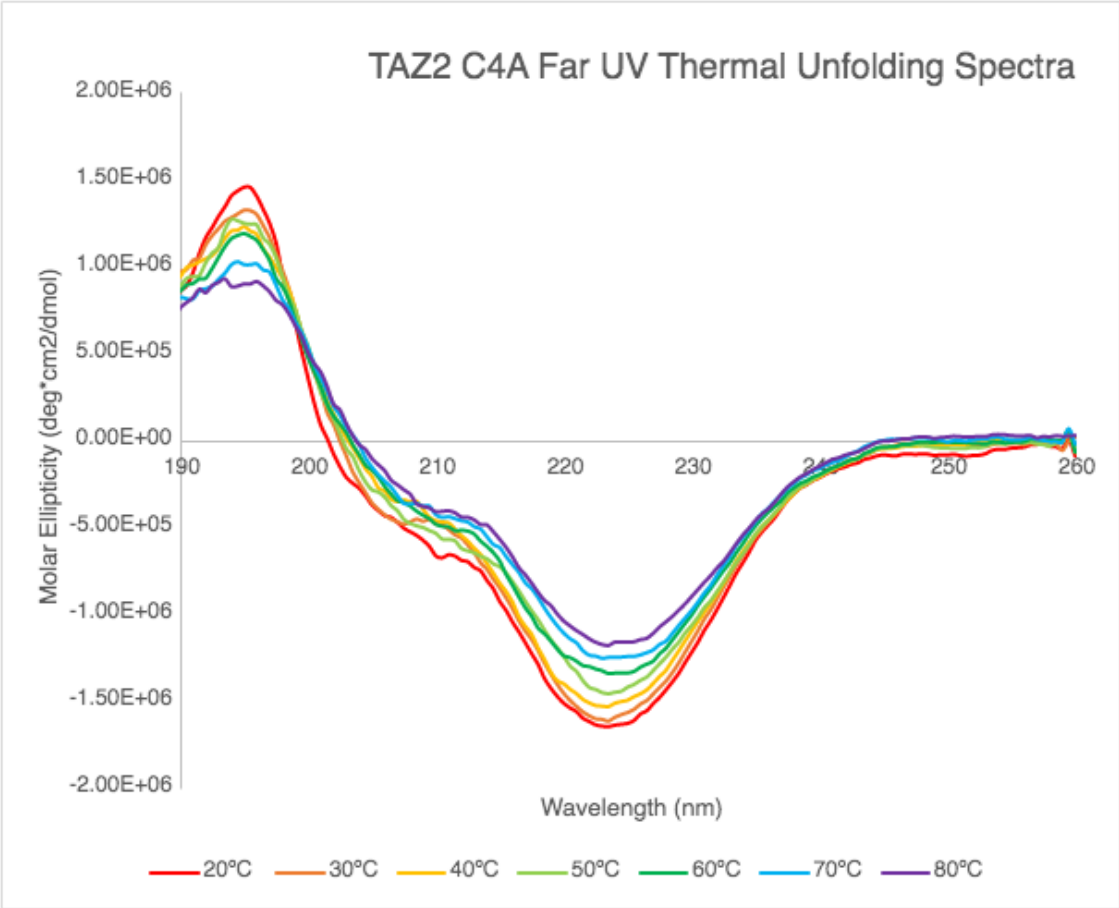

Figure 18 TAZ2 C4A CBP fragment far UV spectra obtained between 20ºC (red) and 80ºC (purple).

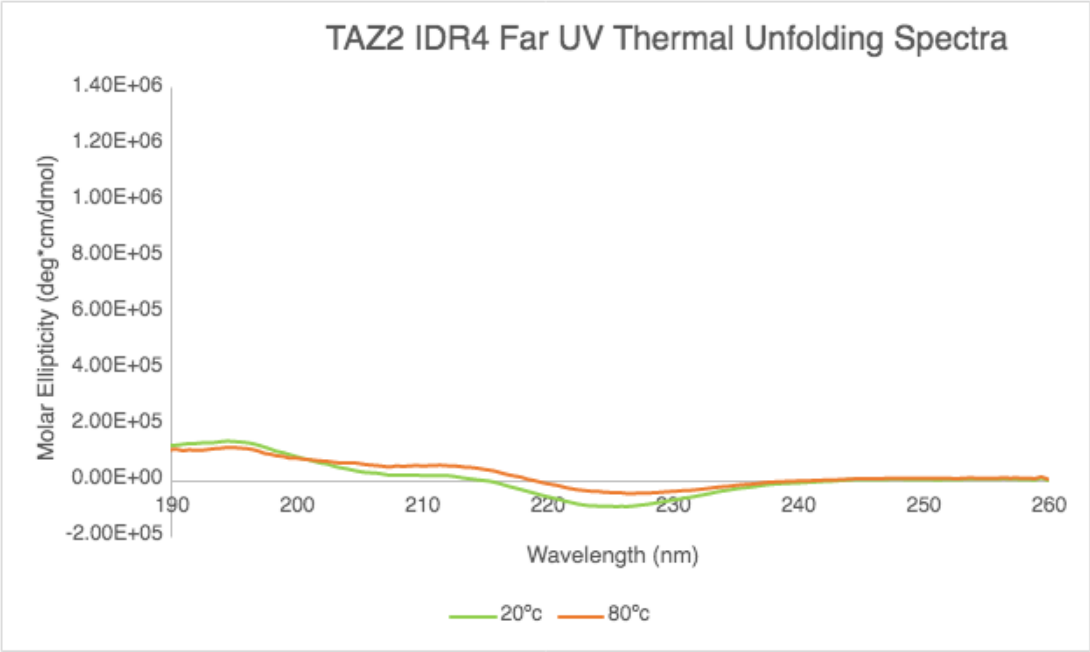

Figure 19 TAZ2-IDR4 CBP fragment far UV spectra obtained at 20ºC (green) and 80ºC (orange). Room temperature analysis displayed a lack of secondary structure for the TAZ2-IDR4 fragment, therefore the unfolding spectra was expected to be minimal as there was no folding present initially to denature with heat.

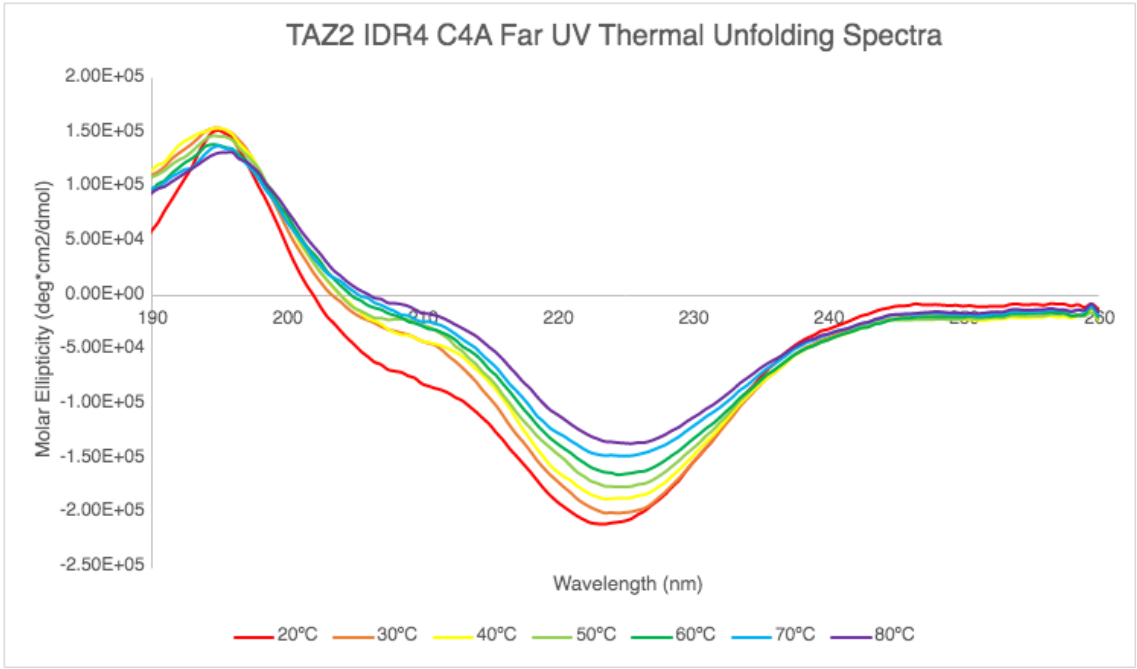

Figure 20 CBP fragment far UV spectra obtained between 20ºC (red) and 80ºC (purple).

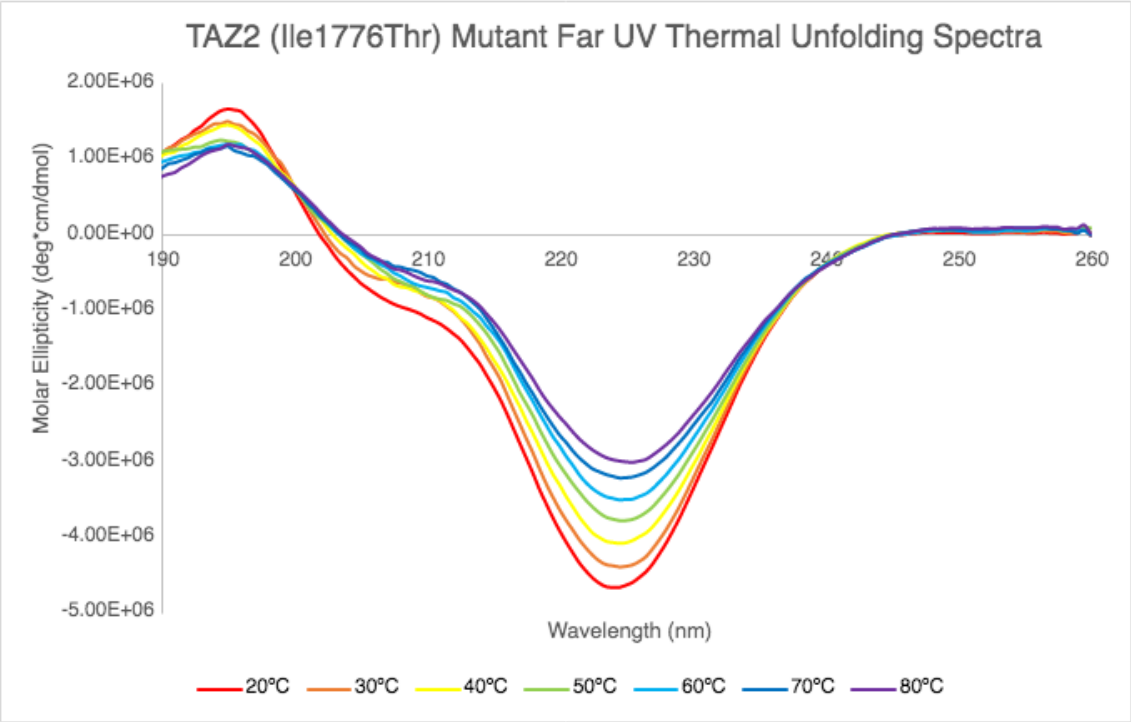

Figure 21 TAZ2 mutant CBP fragment far UV spectra obtained between 20ºC (red) and 80ºC (purple).

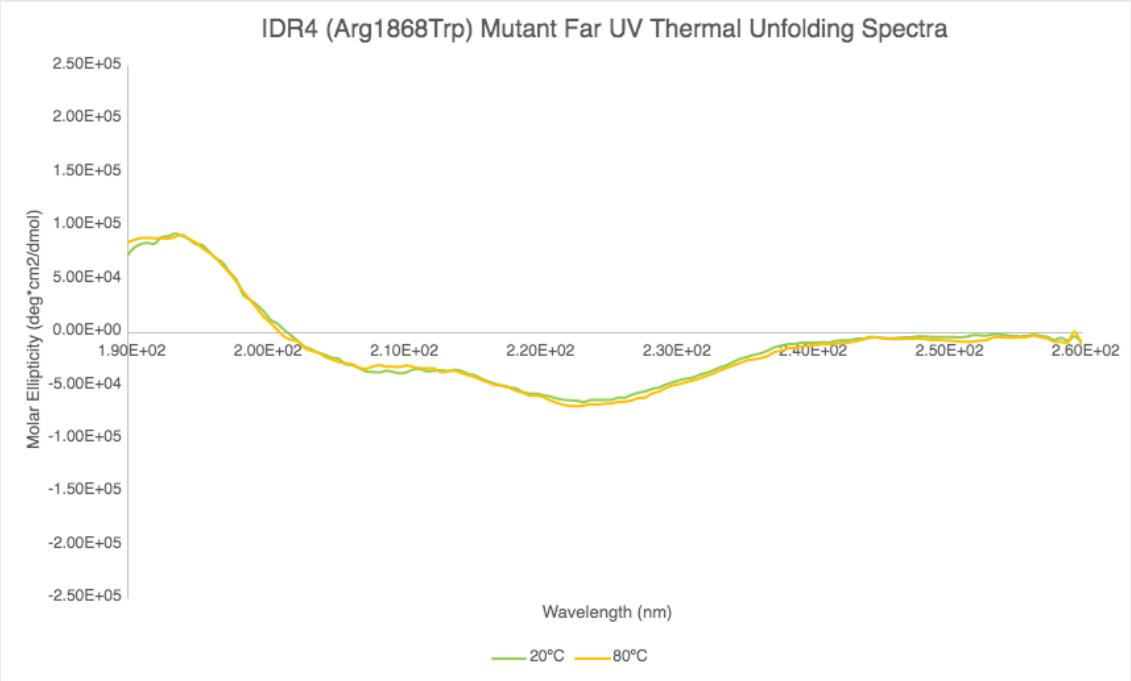

Figure 22 IDR4 mutant CBP fragment far UV spectra obtained at 20ºC (green) and 80ºC (orange). Room temperature analysis displayed a lack of secondary structure for IDR4 mutant, therefore the unfolding spectra was expected to be minimal as there was no folding present initially to denature with heat.

Utilising the thermal unfolding CD spectra, an estimate of the Tm (ºC) values were produced for each of the CBP protein fragments (Table 16). The Tm (ºC) values produced display only a minimal difference in the stability between the three fragments which achieved folding in the conditions provided. The Tm (ºC) value is the temperature at which 50% of the protein fragment within solution is unfolded. This value is able to be calculated using the unfolding spectra and provides a value which can be used to compare the stability of the different fragments.

Table 16 Tm ºC values of CBP Fragments which produced unfolding spectra in the far UV.

CBP Fragment	Tm (ºC)
TAZ2 C4A (CBP 1763-1855, C4A)	52ºC
TAZ2-IDR4 C4A (CBP 1763-1876, C4A)	50ºC
TAZ2 Mutant (CBP 1763-1876, C4A & I1776T)	49ºC

The difference in Tm (ºC) values seen between the three fragments which produced unfolding spectra is small. The TAZ2-IDR4 and IDR4 (Arg1868Trp) mutant fragment Tm (ºC) values were not calculated, as there was minimal folding present in these fragments to begin with. The TAZ2 C4A fragment appears to be slightly more stable than the TAZ2-IDR4 C4A fragment and the TAZ2 Ile1776Thr mutant, producing the greatest Tm value of 52ºC. The TAZ2 C4A fragment and TAZ2-IDR4 C4A fragment differ by 21 residues which compose the IDR4 domain at the C terminus of the TAZ2-IDR4 C4A fragment. Initial far UV room temperature spectra displayed it was unlikely that the IDR4 region contained within the TAZ2-IDR4 C4A fragment contained any folding. The

IDR4 portion of the TAZ2-IDR4 C4A fragment beginning the thermal unfolding experiments as unfolded would lower the Tm (ºC) value, which is likely why we see a 2ºC difference between the Tm values of the TAZ2 C4A fragment and the TAZ2-IDR4 C4A fragment. The TAZ2 (Ile1776Thr) mutant produced the smallest Tm value, with 50% of the protein unfolded at 49ºC. The TAZ2 Ile1776Thr mutant had a polar residue introduced in place of a hydrophobic residue. It would be expected that this would result in a lowered Tm as the polar residue is less resistant to unfolding as it favours interaction with the polar environment of the sample buffer.

Overall minimal impact on the CBP fragment stability is brought about by the introduction of the MHS associated mutation. Fragment behaviour under the experimental conditions provided, suggest that the molecular basis of MHS does not stem from a large difference of domain stability, impairing the ability of the region to function, keeping in mind that this analysis looked at CBP fragments outside of the context of full length IDR4 and CBP.

Chapter 5: DNA Binding Analysis of TAZ2/IDR4 CBP Fragments.

5. TAZ2 DNA Binding Analysis

In studies by Song et al. (2002) and Sheahan et al. (2020) it was shown that the CBP TAZ2 domain and N terminal region of the IDR4 possess DNA binding activity. Song et al. (2002) showed that the CBP TAZ2 domain and the N-terminal region of the IDR4 (CBP 1810-1860) had DNA binding activity. Further experiments carried out by Sheahan et al. (2020) on the TAZ2 domain displayed that DNA binding was sequence independent and occurred in both the presence and absence of histone/nucleosomal structure. This group also showed that other CBP fragments excluding the TAZ2 domain displayed off target histone acetylation by the HAT domain. Further investigations on the DNA binding activity of the TAZ2 domain are required in order to understand more about interaction.

A range of different CBP fragments were utilised throughout the experiments presented in this work to order to investigate the change in DNA binding activity when different mutations to the TAZ2 and IDR4 domains are present.

Three different DNA probes were used in this study to investigate the DNA binding activity of the CBP fragments. The probes were selected taking into consideration the existing knowledge of zinc finger DNA binding. Other zinc finger protein domains are known to form interactions with DNA by straddling the major groove of the double helix, making trinucleotide sequence specific interactions (Isalan et al., 1998).

Two of the probes used were randomly generated sequences (Random sequence generation website details are provided in the appendix) with differing guanine and cytosine (GC) content of 29 base pairs (60% GC) and 30 base pairs (50% GC) (full sequences provided in appendix). These probes were used to investigate sequence independent DNA binding of the CBP fragments. The third probe called gTEL45, (45 bases) is a telomeric, non-canonical DNA sequence known to form parallel and anti-parallel G quadruplex (G4) structures. G4 DNA structures are commonly found at promoter and telomeric regions and have been shown to be involved in gene regulation (Spiegel et al., 2020). As G4 DNA lacks

the major and minor grooves of canonical DNA, the gTEL45probe was used to investigate DNA binding by the CBP fragments in the absence of a major groove.

5.1 Electrophoretic Mobility Shift Assays (EMSA)

In order to investigate the presence of any interactions which may take place between the CBP protein fragments and DNA, two methodologies were used. The first of the two methods used was electrophoretic mobility shift assays (EMSAs). EMSAs are able to provide information on DNA-protein complexes because such complexes migrate differently on both agarose and acrylamide gels compared to their unbound versions due to changes in charge, shape and size upon DNA-protein complex formation. Both agarose and polyacrylamide gel electrophoresis systems can be utilised as long as the conditions are native to ensure interactions are not disrupted throughout the electrophoresis process.

Interactions between protein and DNA strong enough to result in the formation of a DNA protein complex, will result in a shifted band on the gel compared to where the unbound counterparts migrate to. Protein binding causes the DNA to migrate at a far slower rate due to the increased size of the complex. If correct molar ratios of protein and DNA are included in the reaction, the results after staining will display two bands in each lane where DNA protein interactions occur. The band further down an agarose gel, closer towards the positive electrode, represents the unbound DNA. The increased migration of this band is due to small size of the DNA probe (in comparison to the protein DNA complex), and the net negative charge of the DNA probe. Any second band visible on the gel, with reduced migration is the DNA probe which is interacting with the protein in solution. Not only is the size of the DNA-protein complex larger than that of the lone DNA probe but the complex also likely carries a net neutral or positive charge. The principle of electrophoresis is dependent on the components moving through the gel towards a positive electrode. A net positive charge, at a neutral pH, is a common feature amongst DNA binding proteins, as it is positively charged residues which often form interactions with negatively charged DNA molecules. Hence, migration of the complex is also typically reduced due to the

lessened negative charge of the complex as a result of the charge carried by the protein.

5.1.1 EMSA Methodology

The 50% GC and 60% GC DNA probes were selected using a random sequence generator site and ordered as single stranded lyophilised oligos from IDT. Upon arrival, the oligonucleotides were resuspended to a final concentration of 200 μM in annealing buffer (10 mM Tris pH 8, 50 mM NaCl, 1 mM EDTA). 50 μL of each complimentary strand was then mixed in a thin-walled PCR tube and placed in a 95 ºC tube warmer for 2 minutes, and then the warmer was switched off while the tubes remained in the metal compartments to cool slowly to room temperature over an hour. The tubes were then centrifuged to remove condensation from the tube lid and stored at -20ºC until use. The gTEL45 probe was provided by the Dr Tracy Hale, Chromatin Research Laboratory (Massey University).

The DNA probes were serially diluted to a concentration of 50 ηM for use in each assay. Protein samples were also diluted to a range of differing concentrations. Protein and DNA probes were incubated on ice for 30 minutes in a final reaction volume of 50 μL of EMSA buffer (20 mM HEPES pH 8, 150 mM KCl and 0.5 mM DTT). After incubation, 2 μL of 50% (v/v) glycerol was added to each tube and mixed. The reactions were then blind loaded (without any loading dye product which may interfere with interactions) into a pre-chilled 2% (w/v) TBE agarose gel (2 g agarose and 100 mL 1x TBE buffer (table 17)) in a 4ºC room. Pre-chilled 0.5x TBE running buffer (100 mL 1x TBE (table 17) and 100 mL H_2O) was used, and the gel was run at 70 v for a total of 20 minutes in 4ºC. The gels were placed within ethidium bromide solution within a fume hood at room temperature to stain for 20 minutes. The gel was then transferred into water to destain for 10 minutes, before being imaged for analysis.

Table 17 1x TBE Buffer used in constructing TBE gels and as running buffer for all EMSAs.

Component	Volume/Weight
Tris	10.8 g
Boric acid	5.5 g
0.5 M Na_2EDTA (pH 8)	4 mL
H_2O	Up to 1 L

5.1.2 EMSA Results

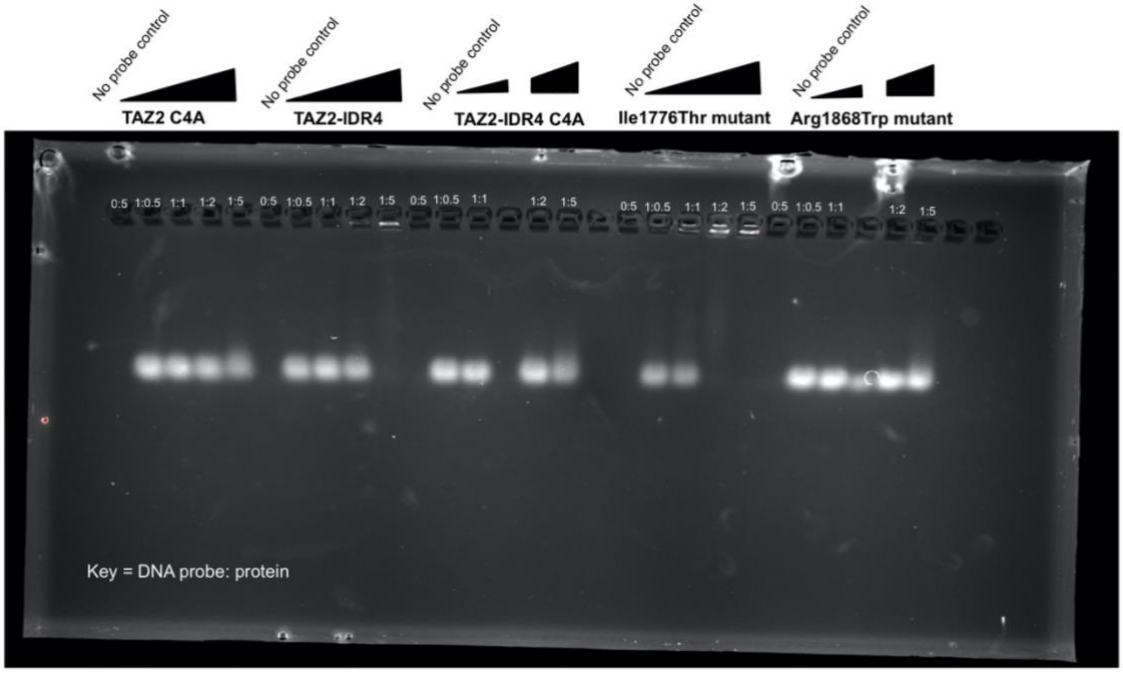

Figure 23 EMSA carried out with 50% GC probe and CBP fragments. Ratios refer to probe: CBP protein fragment molar ratio. From left to right; TAZ2 C4A, TAZ2-IDR4, TAZ2-IDR4 C4A, TAZ2 mutant (I1776T) and IDR4 mutant (A1868W). In order to prevent any disruption to interactions between DNA and protein interactions, loading dye was omitted. In two instances loading was interrupted with a blank lane (TAZ2-IDR4 C4A and IDR4 mutant). There is an unintentional gap between the loading of both TAZ2-IDR4 C4A 1:1 and 1:2. While loading IDR4 mutant 1:2 sample, the well was pierced slightly

during an attempt at loading, therefore the sample was loaded again, this is visible as a faint band in the middle of the IDR4 mutant samples.

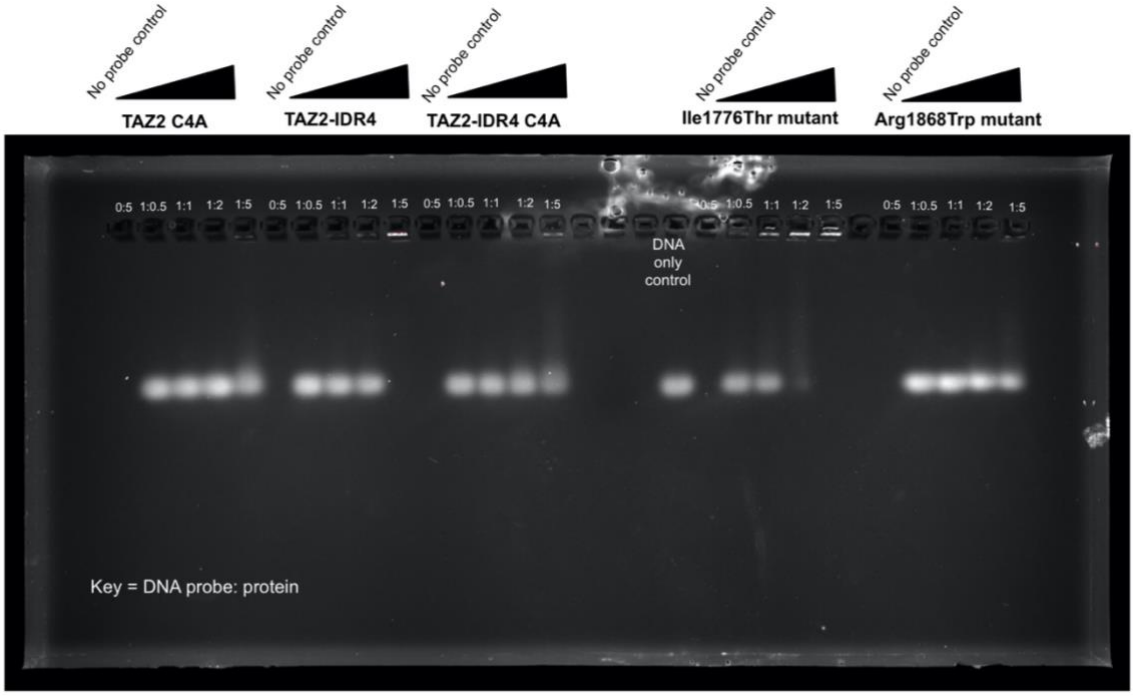

Figure 24 EMSA carried out with 60% GC Probe and CBP fragments. Ratios refer to probe: CBP protein fragment molar ratio. From left to right, with increasing quantities of protein; TAZ2 C4A, TAZ2-IDR4, TAZ2-IDR4 C4A, DNA only control, TAZ2 mutant (I1776T) and IDR4 mutant (A1868W).

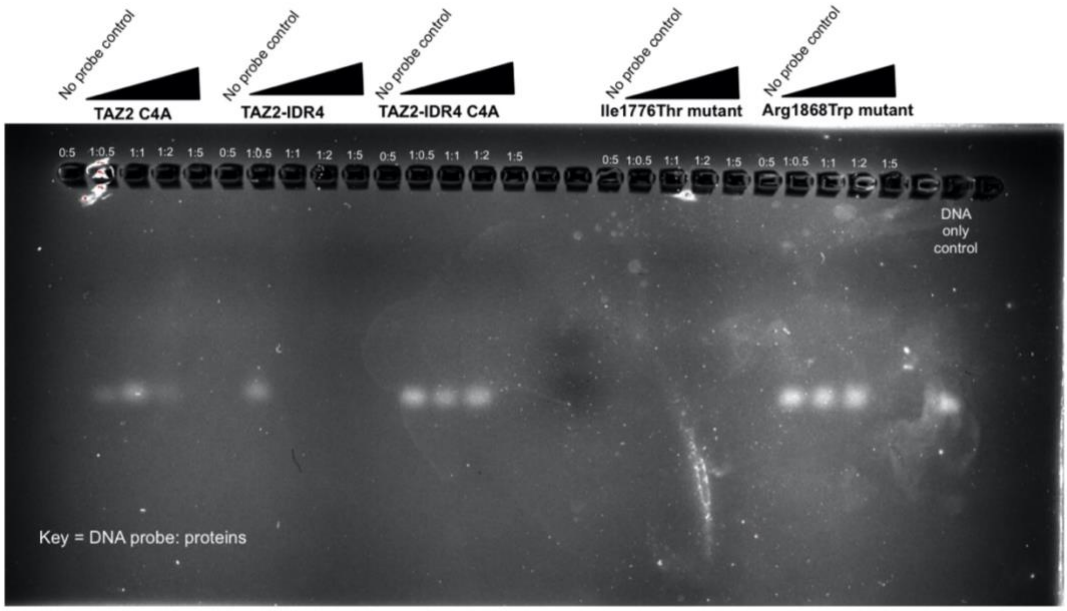

Figure 25 EMSA carried out with the gTEL45 probe and CBP fragments. Ratios refer to probe: CBP protein fragment molar ratio. From left to right, with increasing quantities of protein; TAZ2 C4A, TAZ2-IDR4, TAZ2-IDR4 C4A, TAZ2 mutant (I1776T) and IDR4 mutant (A1868W) and DNA only control.

All EMSA results produced during this study (figures 23-25) display one dominant band per lane, located midway on the agarose gel. This is representative of the DNA which did not interact with the CBP protein fragment. Typically, in an EMSA showing the presence of a DNA-protein interaction, a second band representative of the DNA protein complex is also visible on the agarose with reduced migration. The isoelectric point (pI) of the CBP fragments, is estimated to be 10.6 by ProtParam (details regarding Protparam website can be found in the appendix). This high pI value means the protein fragments possess a net positive charge in conditions with a pH lower that 10.6, including the environment provided by the EMSA buffer. This is a common attribute of DNA binding proteins, as a positive charge at physiological pH is ideal for forming interactions with DNA molecules which carry a net negative charge. This means the CBP fragments will have a positive charge within the experimental conditions also, likely not neutralised by interaction with the comparatively small DNA molecule. A potential limitation of

an EMSA run under these conditions is that the positive charge of the large DNA protein complex which may form between DNA and the CBP fragments, will not migrate throughout the agarose gel towards the positive electrode at the end of the gel. For this reason, DNA-protein complex formed in this experiment is not visible on the gel, but instead as slight fluorescence within the loading wells and/or as a lower mobility smear, accompanied by a decrease in the intensity of the free DNA probe band signal with increased protein concentration.

Table 18 CBP Fragment EMSA result summary discussing the DNA:protein ratio at which shifts are visible and to what extent.

	TAZ2 C4A	TAZ2-IDR4	TAZ2-IDR4 C4A	TAZ2 I1776T Mutant	IDR4 R1868W Mutant
50% GC probe	Mild shift at 1:5	Mild shift at 1:2, complete shift at 1:5,	Mild shift at 1:2, moderate shift 1:5	Moderate shift 1:0.5 and 1:1, complete shift 1:2	Mild shift 1:5
60% GC probe	Mild shift at 1:5	Mild shift at 1:2, complete shift at 1:5,	Moderate shift at 1:5	Moderate shift 1:0.5 and 1:1, complete shift at 1:2	Mild shift 1:5
gTEL45 probe	Mild shift at 1:1, complete shift at 1:2	Mild shift 1:0.5, complete shift at 1:1	Mild shift 1:1 and 1:2, complete shift at 1:5	Complete shift at 1:0.5	Complete shift 1:5

WT CBP Fragment EMSA Results

The gel-based nature of the EMSA method means that it is difficult to make definitive conclusions from these qualitative results. The TAZ2 C4A and TAZ2-IDR4 C4A fragments throughout all probe binding experiments produced very similar results. The common result seen across all CBP protein fragments was that at higher ratios of protein to probe, there was a shift in the probe DNA band. Throughout assays with all three probes, the TAZ2-IDR4 C4A fragment indicated slighter greater binding to the DNA probes than the TAZ2 C4A fragment, producing slightly more significant shifts at each probe to protein ratio. This opposes the suggestion that the IDR4 helical forming region (present within TAZ2-IDR4 C4A but absent within TAZ2 C4A), may act as an inhibitor of the TAZ2 DNA binding by competing for the DNA binding site. The opposite is in fact suggested, that DNA binding is enhanced in the presence of the IDR4 in agreements with the findings presented by Song et al. (2002).

The TAZ2-IDR4 fragment, despite displaying a lack of any folding during structural analysis (chapter 4), produced the greatest DNA binding capabilities of all three WT fragments with all three probes during the EMSA experiments. The TAZ2-IDR4 fragment does not contain the C4A stability enhancing substitution. It was thought that the C4A substitution would increase the solubility of the CBP protein fragments for experimental work without impairing or altering native structure. However, these experimental results suggest these four cysteine residues (non-zinc binding) which are substituted for alanine residues in the TAZ2 C4A and TAZ2-IDR4 C4A fragments, may actually be involved in forming interactions with DNA binding partners. By EMSA results alone, it is not possible to rule out that the disappearance of the free DNA probe band is due to binding aggregated protein at the higher protein concentrations.

Mutant CBP Fragment EMSA Results

Throughout EMSA trials, the TAZ2 mutant (I1776T) displayed increased DNA binding activity compared to all other CBP protein fragments. The increase in DNA binding of this mutant is not understood. The polar threonine residue

substituted, potentially could form any non-covalent interactions with DNA or may be able to provide further stabilising interactions.

The IDR4 mutant (R1868W), contains the substitution of a positively charged arginine residue for a bulky, hydrophobic tryptophan residue. Compared to other CBP fragments, this mutant displayed the weakest DNA binding, with only mild shifts seen at a 1:5 DNA probe to protein ratio. Removal of this positive residue from the IDR4, reduced DNA binding activity of the CBP fragment. This suggests DNA binding may also involve the positive arginine residues spread out along the IDR4, or that the presence of the tryptophan is disruptive. Substitution of these residues may result in decrease DNA binding capabilities of the protein, as displayed in these experimental EMSAs.

Probe Analysis of EMSA Results

The EMSA which utilised the 50% GC and 60% GC probes displayed very similar results, with CBP binding the 50% GC probe with a marginally apparent greater affinity. The EMSAs utilising the gTEL45 probe however, display significantly increased binding with all CBP fragments. The gTEL45 probe does not contain typical double stranded helical structure which zinc fingers are known to utilise in forming interactions with DNA. This does not appear to hinder the DNA binding of the CBP fragments, instead affinity is increased. However, the reduced intensity of unbound DNA on the gTEL45 gel, could also be the result of gTEL45 DNA interacting with a weaker intensity with the ethidium bromide dye. The latter suggestion is supported by the dull intensity of the DNA probe only control lane of the gTEL45 gel.

5.2 Bio-layer Interferometry (BLItz)

The Bio-layer interferometry (BLItz) technique is another method for investigating protein-DNA interactions which overcomes some of the limitations of the EMSA method. BLItz measures the reflection or interference of white light to detect the increased mass caused by binding of a molecule, DNA in this instance, to a protein of interest which is immobilised on a glass tip. The his-tag at the N-

terminus of the CBP protein fragments which is used for purification, is also utilised for attachment during BLItz experiments. The BLItz system offers a range of glass tips with different binding agents. Nickel-NTA coated glass tips were used throughout this study. The his-tag of the protein fragment binds to the nickel tip, and when mounted upon the BLItz sensor, any change in mass interacting with the tip is able to be detected through a change in the relative intensity of reflected white light. This includes DNA binding to the protein which is interacting with the nickel tip.

The BLItz system is extremely sensitive and provides quantitative values for both ligand association and dissociation. The K_D value or dissociation constant (M) for a ligand-protein interaction is the ligand concentration (M) at which half of the protein binding sites are occupied by ligand, which in this case is DNA. The smaller the K_D value, the less ligand required to occupy half of the binding site and therefore the greater the affinity between the two counterparts. In order for an accurate K_D value to be obtained, assays are carried out over a range of ligand concentrations with sufficient replicates. Provided the timeframe for this project, and the difficulties producing great quantities of CBP protein fragments, duplicates were carried out for each CBP fragment with each DNA probe at a set ligand concentration of 2 μM. This means the BLItz data obtained in this study is preliminary data and provides insight into what we might be able to expect from the WT and mutant DNA binding behaviour.

5.2.1 BLItz Procedure

The advanced kinetics programme was used for all BLItz experiments on the BLItz kinetics software. The Ni-NTA tips were submerged in 250 μL of moderate salt buffer (300 mM NaCl, 30 mM NaH_2PO_4, 2 mM TCEP pH 7.5) for 10 minutes prior to use. A Ni-NTA tip was then mounted onto the biosensor and submerged into an opaque tube containing 250 μL of moderate salt buffer again for 30 seconds while a baseline reading was established. The protein fragment solution was then added to the Ni-NTA tip by submerging the Ni-NTA within a 4 μL droplet of 60 μM protein solution with shaking for 120 seconds. Excess protein was then

able to be washed away from the tip by submersion back into moderate salt buffer within an opaque tube for 30 seconds with shaking. Association between the protein and DNA is then able to be measured by submerging the tip within a 4 μL drop of 2 μM DNA solution for 120 seconds with shaking. Dissociation is then able to be measured by submersion of the tip within 250 μL of moderate salt buffer in an opaque tube for 120 seconds with shaking. Tips were not stripped, recharged and reused. A new tip was used for every trial. Each CBP protein fragment was trialled a minimum of six times, with two trials for each probe.

Data is presented as a change in wavelength Δλ (nm) over time. Raw data obtained is processed by subtracting a baseline reading, where no DNA was introduced. The change in wavelength Δλ (nm) at which DNA association begins was also subtracted from all values, so that DNA association begins at 0 nm.

5.2.2 CBP Fragment DNA Binding BLItz Results

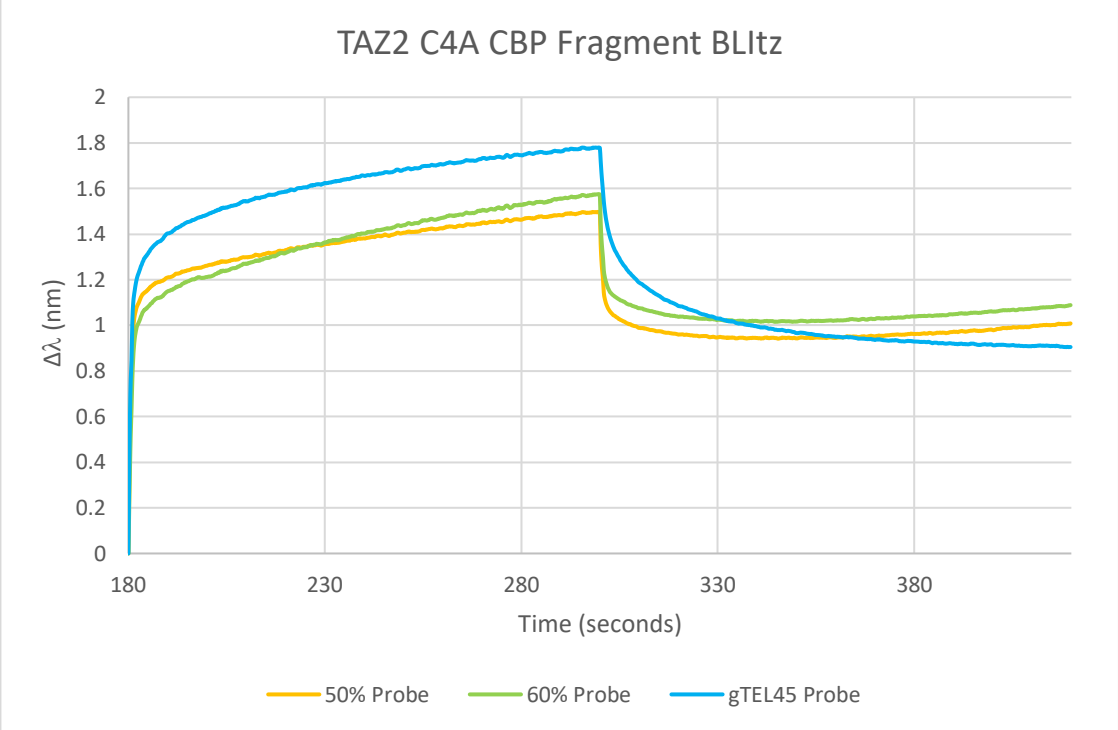

Figure 26 TAZ2 C4A CBP Fragment BLItz. DNA association takes place between 180 and 300 seconds, dissociation takes place between 300 and 420 seconds.

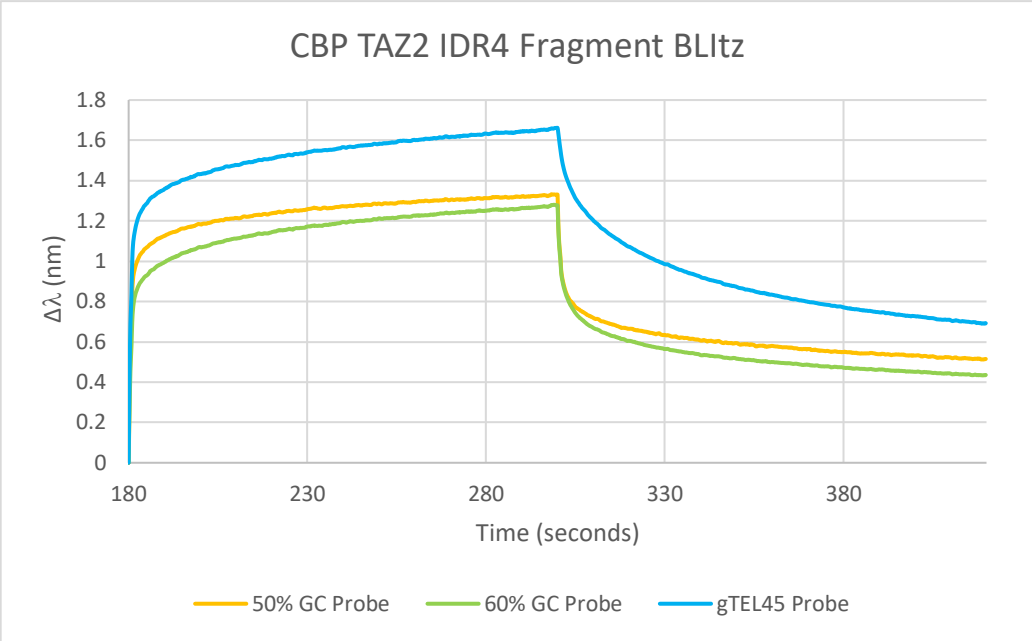

Figure 27 TAZ2-IDR4 CBP Fragment BLItz. DNA association takes place between 180 and 300 seconds, dissociation takes place between 300 and 420 seconds.

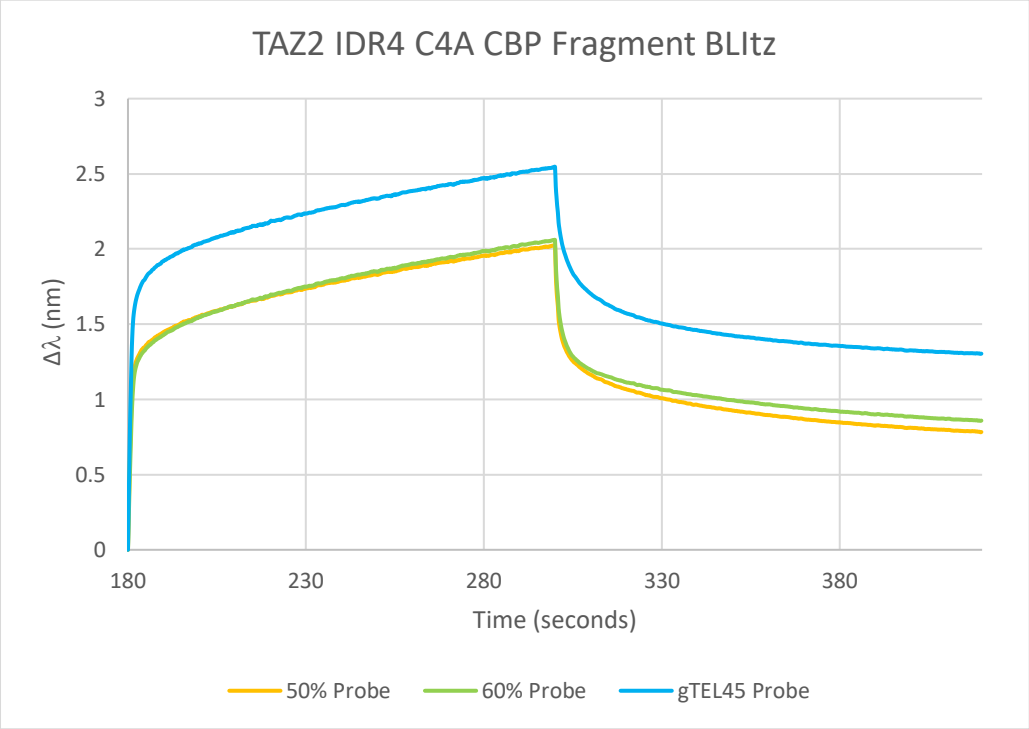

Figure 28 TAZ2-IDR4 C4A CBP Fragment BLItz. DNA association takes place between 180 and 300 seconds, dissociation takes place between 300 and 420 seconds.

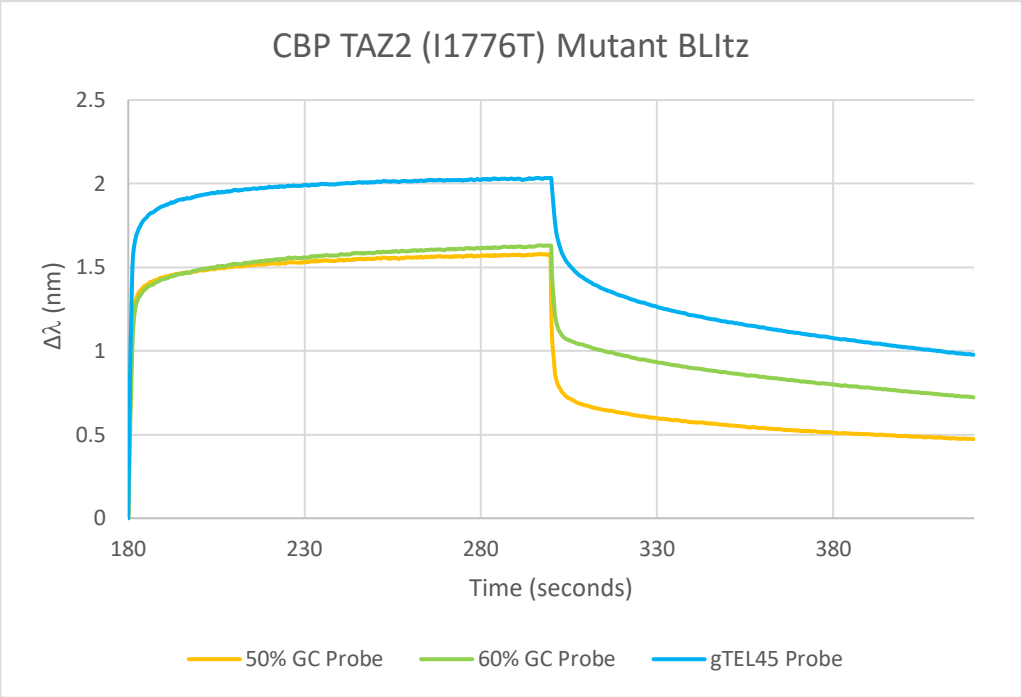

Figure 29 TAZ2 (Ile1776Thr) mutant CBP Fragment BLItz. DNA association takes place between 180 and 300 seconds, dissociation takes place between 300 and 420 seconds.

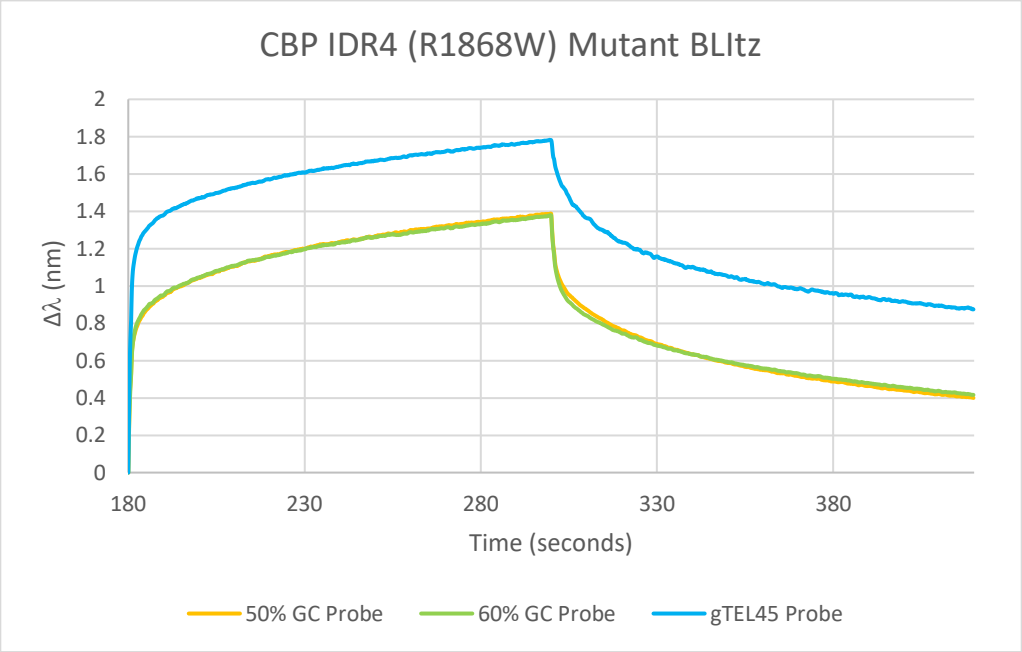

Figure 30 IDR4 mutant (Arg1868Trp) CBP Fragment BLItz. DNA association takes place between 180 and 300 seconds, dissociation takes place between 300 and 420 seconds.

Table 19 Dissociation (K_D) values (μM) produced by BLItz assays investigating DNA binding of CBP fragments. Data values presented are mean averaged K_D values of duplicates.

	TAZ2 C4A	TAZ2-IDR4	TAZ2-IDR4 C4A	TAZ2 Mutant	IDR4 Mutant
50% GC probe	0.66	0.56	0.98	0.20	1.48
60% GC probe	0.70	0.90	0.92	0.34	1.38
gTEL45 probe	0.65	0.30	0.34	0.25	0.35

CBP WT BLItz Results

All three WT and C4A WT CBP fragments displayed DNA binding activity, producing similar K_D values (figure 26-28, table 19). The TAZ2-IDR4 fragment provided the smallest K_D value which is also consistent with the results produced by the EMSAs. The K_D values provided by the WT fragments are within such a small range, which suggests the DNA binding activity of the WT fragments is all very similar. Further replicates would need to be carried out to determine consistent differences, if any, between these fragments DNA binding activity.

CBP Mutant BLItz Results

Both CBP mutants included in this study displayed DNA binding activity (figures 29 & 30), producing similar K_D values not only to each other, but also to the WT and C4A WT CBP fragments (table 19). The CBP fragment which produced the smallest K_D values, was the TAZ2 mutant (I1776T). This mutant also displayed dominant DNA binding activity of all CBP fragments during EMSAs. The introduced polar threonine may form further stabilising interactions with the charged DNA molecule, resulting in a smaller K_D value than other CBP fragments included in this study. The IDR4 mutant (R1868W) provided the greatest K_D value

of all CBP fragments. This result is consistent with the EMSAs and confirms the importance of the positively charged residue within the IDR4 for DNA binding.

Chapter 6: Discussion and Future Directions.

6.1 Discussion of Experimental Results

Compiling the information provided by the experiments presented, a few conclusions can be drawn about the in vitro behaviour of both WT and MHS associated mutant CBP protein fragments.

6.1.1 Expression and Purification of CBP Fragments

The expression and purification system used for the production of the CBP fragments was only able to produce sufficient quantities for the structural and functional activity assays presented in this thesis. The protocol was favoured by the C4A WT CBP fragments which were used for the optimisation process presented in chapter 3. The I1776T and R1868W CBP mutants were able to be expressed and purified well enough with this protocol that they were able to be included in structural and activity assays.

The other half of the MHS associated mutant fragments, which contained cysteine residue substitutions (Zn^{2+} binding residues) were unable to be produced in sufficient quantities to be taken past structural analysis. However, despite this hindering the ability for DNA binding analysis to be carried out, it does provide information in itself. The improved solubility displayed by the WT and C4A WT fragments compared to the Zn^{2+} binding cysteine MHS mutants, suggests a lack of native folding in the mutants, resulting in aggregation and the formation of inclusion bodies seen. The protocol requires further optimisation in order to produce larger quantities of soluble fragment for further, more in depth analysis with a greater number replicated experiments.

The key features of the expression and purification protocol which were essential for the success seen in WT, C4A WT, I1776T and R1868W CBP fragments include but is not limited to the following; the use of the reducing agent, TCEP, within all buffer systems post bacterial cell lysis in order to prevent the formation of any undesirable disulfide bridges. Secondly, the use of ‘slow and low’ cell culture conditions, growing the BL21 (DE3) *E. coli* over a three-day period, with the low incubation temperature of 20ºC upon induction of recombinant protein

expression. Additionally, the supplementation of $ZnCl_2$ to cell culture in order to provide the CBP fragment with cofactor required for folding. Finally, the use of dialysis for buffer exchange was also crucial as the CBP fragments would precipitate with use of centrifugal concentrators. This is likely a result of the protein being unstable within solution at higher concentrations or due to membrane association by the CBP fragments, causing aggregation.

6.1.2 Structure and Stability Discussion

Structural investigation methods attempted in this study included circular dichroism of both the near and far UV to investigate secondary and tertiary folding of the CBP fragments. Stability was investigated by circular dichroism also in the far UV while sample temperature was increased in order for denaturation profiles to be recorded. Near UV circular dichroism was attempted, but unable to provide any viable results. This is likely due to a lack of aromatic residues within the CBP fragments (>3), a lack of folding by the WT and MHS associated CBP fragments, and the high salt content of the buffer system which interferes with the near UV wavelength range.

In order to be able to draw conclusions about changes to structure, stability or activity caused by the Menke Hennekam associated mutations, the wild type CBP fragments of the study used for comparison ideally need to present native folding. The WT CBP fragment TAZ2-IDR4 (CBP 1763-1786) included in this study produced a weak far UV circular dichroism signal, suggesting a lack of ordered folding or native structure. Previous research carried out on the TAZ2 domain suggests expression and purification of the domain as a fragment is often hindered by the formation of non-native disulfide bonds as a result of the large quantity of cysteine residues within the domain (De Guzman et al., 2000). When the fragments are exposed to oxidising conditions, such as upon bacterial cell lysis, the environment favours the formation of salt bridges between cysteine residues residing within close proximity (Duong-Ly & Gabelli, 2014).

In order to prevent the formation of undesirable disulphide bridges which disrupt native folding, two steps were taken. Firstly, the addition of the reducing agent

TCEP to all buffer systems. Secondly, C4A versions of the CBP fragments were included in this study (table 1). The C4A fragments, contain the substitution of four non-zinc binding cysteine residues replaced by four alanine residues. The C4A fragments were first introduced by De Guzman et al. (2000), but have since been used in numerous TAZ2 fragment studies (Hanqiao Feng et al., 2009; Jenkins et al., 2009; Miller Jenkins et al., 2015; Miller et al., 2009), as they possess native folding and prove to be a more successful fragment for recombinant expression and purification due to their increased solubility. The C4A mutants included in this study are used as a wild type comparison for the CBP mutants, as the true CBP wild type, TAZ2-IDR4, failed to fold.

The TAZ2 C4A and TAZ2-IDR4 C4A fragments produced near identical far UV circular dichroism spectra. The spectrum produced indicated that the secondary structure of both of the fragments was predominantly alpha helical. This is expected and suggests a degree of native folding within the C4A fragments, since the TAZ2 domain native structure contains four alpha helices connected by loops (figure 5) (De Guzman et al., 2000). The minimal difference present between the far UV signals produced by the TAZ2 C4A fragment and the TAZ2-IDR4 C4A fragment, suggests there is no large difference in folding due to the presence of the IDR4 portion of the longer fragment. The IDR4 has been shown in previous research by Piai et al. (2016) to fold into transient helical structures, one of which forms with residues CBP 1855-1876, which is included within the TAZ2-IDR4 C4A fragment. In addition to the NMR studies by Piai et al. (2016) AlphaFold structure prediction software (Jumper & Hassabis, 2022) also estimates residues 1855-1876 to form helical folding continuous with the α4 of the TAZ2 domain (figure 26, yellow-red). If the CBP TAZ2-IDR4 C4A fragment in this experiment had such folding, the far UV CD signal produced would have been of a slightly greater intensity than that of the TAZ2 C4A fragment. It may be that the full IDR4 may need to be present, the conditions may need to be adjusted, or an external factor may be required to stimulate the formation of this predicted structure.

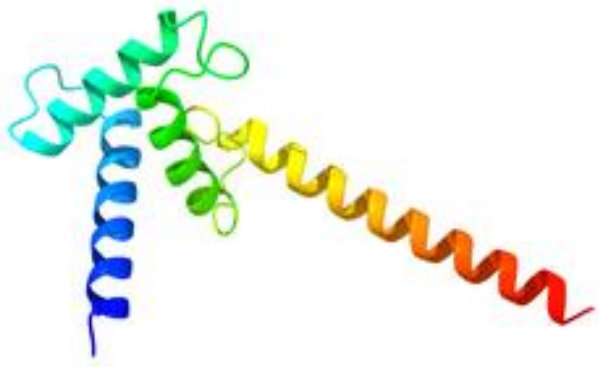

Figure 31 AlphaFold prediction software prediction of the TAZ2-IDR4 C4A CBP fragment (with high confidence). TAZ2 helices α1 (blue), α2 (cyan), α3 (green), α4 (yellow), IDR4 1855-1876 (red).

Although the Cys1819Phe and Cys1838Tyr Menke-Hennekam mutants proved to be difficult to obtain sufficient quantities of, enough of these MHS mutants were produced to be included in far UV CD analysis. The weak signal far UV spectra produced by these mutants, alongside the difficulty which was encountered in attempts to produce these recombinant protein fragments is consistent with a lack of native folding in these mutants. Cysteine residues 1819 and 1838 are two of twelve zinc binding residues within the TAZ2 domain. Of all currently recorded Menke Hennekam syndrome patients, a large portion contain substitution or deletion mutations of zinc binding cysteine or histidine residues (Menke et al., 2018; Menke et al., 2016; Sima et al., 2022). The behaviour of these fragments throughout this study suggests that the removal of just one zinc binding residue, results in the TAZ2 domain being unable for form any native folding. A lack of native folding may well hinder the functionality of the domain which in turn may give rise to the Menke Hennekam phenotype seen in the individuals carrying these mutations. Further optimisation of the expression and purification protocol for the zinc binding mutant fragments needs to be carried out, in order to obtain enough protein fragment to carry out DNA binding activity assays.

The IDR4 (Arg1868Trp) Menke Hennekam associated mutant fragment produced a far UV CD spectrum with a very weak signal suggesting a lack of native folding

within the fragment. Given this mutation resides within the IDR4 region of the fragment, the lack of TAZ2 folding was not predicted. Provided a degree of native folding is seen in the TAZ2 C4A fragment, which lacks the IDR4 all together suggests that the WT IDR4 is not required for folding of the TAZ2 domain. Further investigations are required to better understand and explain this behaviour.

The TAZ2 mutant (Ile1776Thr) included in the trial was included to investigate activity upon disruption to the hydrophobic core region, without making a change to a zinc binding residue. The impact which the introduction of a polar residue in place of a hydrophobic residue would have on the secondary structure was difficult to predict as protein secondary structure formation is the result of a complex interplay between neighbouring residues (Pirovano & Heringa, 2010). It was anticipated that there would potentially be reduced helical packing of the hydrophobic core, which is of course tertiary structure, and only detectable by near UV CD analysis. The Ile1776Thr TAZ2 mutant fragment displayed a significant increase in CD far UV spectra intensity compared to WT and C4A WT fragments. The far UV CD results suggest an increase in helical content of the I1776T TAZ2 mutant fragment which is likely a result of the threonine residue behaving differently to the isoleucine in the WT fragment, altering the secondary folding of the fragment.

Thermal unfolding analyses of the CBP fragments which displayed a degree of folding at room temperature was carried out using far UV CD, while increasing the sample compartment temperature in order to denature the sample. All initially folded fragments TAZ2 C4A, TAZ2-IDR4 C4A and TAZ2 mutant (I1776T) displayed similar unfolding patterns. The Tm (ºC) values calculated for these three CBP fragments displayed minimal difference, suggesting the TAZ2 I1776T mutant and both the WT and C4A WT CBP fragments are equally stable. The I1776T mutation does not disrupt the stability of the fragment.

6.1.3 CBP Fragment DNA Binding

The DNA binding assays carried out in this work were done so with the intention of establishing if the Menke Hennekam associated CBP mutant fragments had any DNA binding activity, and if so, did the activity differ significantly to the WT CBP fragments.

The two methods used to investigate CBP DNA binding activity were EMSAs and BLItz assays. All results provided across both assays were consistent, displaying similar DNA binding by all CBP fragments (K_D 0.20 μM – 1.48 μM). All fragments bound all three different DNA probes with a similar affinity, including the G-quadruplex DNA probe. It can be concluded that the MHS associated CBP R1868W mutant does possess DNA binding activity. The TAZ2 hydrophobic binding core mutant (I1776T) also displayed DNA binding activity which resembles the WT CBP fragments. Further speculations can be made about the small differences in DNA binding seen across the different CBP fragments. However further studies with increased replicates would be required to carry out statistical analysis to confirm whether any consistent differences exist for DNA binding by each fragment.

The Menke Hennekam associated IDR4 mutant included in this study (Arg1868Trp) displayed DNA binding activity in both the EMSA and BLItz assays, however it was consistently weaker than the binding by all other fragments included in this study. The previous research by Song et al. (2002) confirming that residues within the IDR4 also make interactions with DNA binding partners, means a weaker affinity should have been anticipated upon the substitution of a polar residue from this area. The substitution of the polar arginine residue, which may form direct interactions with DNA binding partners due to its charge, has been replaced by a hydrophobic tryptophan which likely has less attraction for the negatively charged DNA.

Taking into consideration the structural data obtained, it was shown that the R1868W mutation impairs the folding of the neighbouring TAZ2 domain. It may

be suggested that the hindered DNA binding activity of the IDR4 mutant could be a result of impaired domain folding. However, the TAZ2-IDR4 WT CBP fragment, which displayed unhindered DNA binding activity also lacked any folding of the TAZ2 domain. There are multiple arginine residues within the N-terminal helical forming region of the IDR4. Substitution of these arginine residues is one of the most commonly recorded mutations associated with Menke-Hennekam syndrome (Menke et al., 2018; Menke et al., 2016; Sima et al., 2022). Further investigations are required into the TAZ2 folding of the IDR4 arginine mutants and into IDR4 DNA binding.

The TAZ2 I1776T mutant, is not a MHS associated mutation. This mutant was included in experimental work in order to investigate the role of the TAZ2 hydrophobic binding core, without disrupting any zinc binding residues directly. The introduction of a polar residue to the region which forms the hydrophobic core of the TAZ2 domain, resulted in an increase in DNA binding activity. This suggests DNA binding interactions likely occur within this region, and the positive arginine and lysine residues which flank this region are probable key players in forming direct interactions with DNA binding partners. This hypothesis could be tested by mutagenesis of these arginine and lysine residues.

The results produced by the DNA binding experiments are consistent with the proposed GoF mechanism behind the MHS phenotype, which opposes the basis of the ‘clinically opposite’ presenting Rubenstein Taybi syndrome. Changes in DNA binding activity in the IDR4 MHS associated mutant, results in unregulated or less specific acetylation by the CBP HAT domain. This model suggests an increase in CBP HAT activity, results from the impaired ability of the TAZ2 domain to select and interact with DNA binding targets. This hyperactive HAT activity is the opposite to that which would be seen in Rubenstein Taybi individuals, where HAT activity is reduced due to direct mutations to the HAT catalytic core reducing acetylation.

6.2 Future Directions

6.2.1 Expression and Purification Protocol

Firstly, further optimisation of the expression and purification protocol for the TAZ2-IDR4 fragments needs to be conducted, especially for the zinc binding residue MHS associated mutant fragments. Trials utilising different *E. coli* strains and compatible plasmids may be required. Different *E. coli* which have been developed to contain the coding sequence for chaperone proteins which aid in the folding of eukaryotic recombinant protein fragments may produce increased quantities of soluble CBP fragment. Different concentrations of Zn^{2+} within the cell culture broth may be another variable to undergo further investigations, in addition to supplementing Zn^{2+} at different stages of fragment production. Another method to be trialled is the deliberate denaturation by guanidium hydrochloride as done by De Guzman et al. (2000) for the expression TAZ2 fragments (CBP 1764-1850). Upon bacterial cell lysis, lysate pellet containing inclusion bodies is dissolved in 6M guanidium hydrochloride before purification. Purified fragment is then reconstituted in buffer containing 200 μM $ZnCl_2$ for folding. Structural studies carried out by De Guzman et al. (2000) displayed that the TAZ2 and C4A TAZ2 fragments were folded by this method.

Additional optimisation needs to be carried out on the buffer exchange and concentration process for the purified protein fragments as this hinders the ability for a variety of experimental techniques to be utilised in investigating the structure and function of these domains. Many experimental techniques require greater protein concentrations, or protein concentration gradients which are limited if the protein fragment is limited to a concentration of 0.6 mg/mL. Experimental techniques are also often impaired by high salt content. Additionally, the high salt environment of the buffer utilised throughout these trials does not resemble physiological salt concentrations, which is not ideal for extrapolating or making claims about the behaviour of the protein within native conditions.

6.2.2 Structural and Stability Studies

Further structural studies focusing on the structures of the IDR4 need to be carried out. The conditions which bring about the formation of the IDR4 helical structures, suggested by Alphafold (figure 31) and displayed in the research by Piai et al. (2016), need to be defined in order to better understand the role of the IDR4. Additionally, further work needs to be carried out to investigate the relationship between the IDR4 and TAZ2 with respect to folding of the TAZ2 domain. In the presence of the IDR4 Arg1868Trp MHS associated mutation, folding of the TAZ2 domain was impaired for reasons which are not yet understood. This structural work is likely required to be conducted via a more extensive method than circular dichroism - NMR or crystallography studies will likely be required.

The protein stability work throughout this thesis was also limited by the quantities of soluble protein expressed and the inability to successfully optimise a differential scanning fluorometry (DSF) protocol. This was attempted however results were non-typical likely due to a lack of folding and hidden hydrophobic side chains. DSF stability analysis will allow for investigations into stability of the CBP fragments upon DNA binding as DNA does not interfere with DSF signal as it does for circular dichroism.

6.2.3 DNA Binding Activity Studies

In order to confirm the DNA binding behaviour of the fragments included in this work, further replicate measurements need to be conducted over varying protein concentrations. This will allow statistical analysis on the K_D values to determine if any differences between the CBP fragments and/or DNA probes are statistically significant. Additionally, upon optimisation of a successful expression and purification protocol for the MHS associated zinc binding residue mutants, DNA binding analysis of these mutants needs to be carried out. Further investigation of the conformational state of the TAZ2-IDR4 (Arg1868Trp) fragment when bound to DNA would be of interest to determine if the protein is bound to DNA in an unfolded state or weather the association induces or stabilises protein folding.

6.2.4 Other CBP Functional Analysis

Ultimately experimental work is required on larger CBP fragments including the TAZ2-IDR4 region with the HAT catalytic and other core domains, which would then allow for the analysis of any potential changes in HAT activity. This will allow for a greater understanding of how changes in the structure and function of the TAZ2 and IDR4 domains directly impacts the activity of the HAT domain. This will also require an optimised expression and purification system, likely involving insect cell culture as expression of such a large fragment is outside the scope for a bacterial cell culture system.

6.3 Final Summary

This study has shown that MHS associated CBP mutant fragments (containing full TAZ2 domain and partial IDR4) have a hindered ability to fold into native structure. Mutations which reside outside of the structured TAZ2 domain, within the IDR4 domain, also have been shown to impact the ability of the TAZ2 portion of the protein fragment to fold. This study has shown that MHS associated CBP mutant fragments possess DNA binding activity, with a similar affinity for DNA as WT CBP fragments. The experiments showed that CBP WT and MHS associated mutant fragments will bind G-quadruplex DNA, with a similar affinity as canonical double stranded DNA. In relation to the basis of MHS, these experimental results suggest that TAZ2 DNA binding affinity for random probes is not altered significantly in the presence of MHS associated mutations. Further DNA binding activity experiments are required to confirm if there is a statistically significant decrease seen in DNA binding by MHS associated mutants, such as that potentially displayed by the IDR4 mutant (R1868W) in these preliminary experiments. This agrees with the gain of function phenotype present in MHS patients, which results from a hyperactive CBP HAT domain due to a lack of regulation by the TAZ2/IDR4 domains in target DNA binding.

Bibliography

Banka, S., Sayer, R., Breen, C., Barton, S., Pavaine, J., Sheppard, S. E., . . . Clayton-Smith, J. (2019). Genotype–phenotype specificity in Menke–Hennekam syndrome caused by missense variants in exon 30 or 31 of CREBBP. *American Journal of Medical Genetics Part A, 179*(6), 1058-1062. doi:https://doi.org/10.1002/ajmg.a.61131

Bannister, A. J., & Kouzarides, T. (1996). The CBP co-activator is a histone acetyltransferase. *Nature, 384*(6610), 641-643.

Biel, M., Wascholowski, V., & Giannis, A. (2005). Epigenetics—an epicenter of gene regulation: histones and histone-modifying enzymes. *Angewandte Chemie International Edition, 44*(21), 3186-3216.

Boland, M. J., Nazor, K. L., & Loring, J. F. (2014). Epigenetic regulation of pluripotency and differentiation. *Circulation research, 115*(2), 311-324.

Cardinaux, J.-R., Notis, J. C., Zhang, Q., Vo, N., Craig, J. C., Fass, D. M., . . . Goodman, R. H. (2000). Recruitment of CREB binding protein is sufficient for CREB-mediated gene activation. *Molecular and cellular biology, 20*(5), 1546-1552.

Chawla, S., Hardingham, G. E., Quinn, D. R., & Bading, H. (1998). CBP: a signal-regulated transcriptional coactivator controlled by nuclear calcium and CaM kinase IV. *Science, 281*(5382), 1505-1509.

Chiu, J., Tillett, D., Dawes, I. W., & March, P. E. (2008). Site-directed, Ligase-Independent Mutagenesis (SLIM) for highly efficient mutagenesis of plasmids greater than 8kb. *Journal of microbiological methods, 73*(2), 195-198.

Contreras-Martos, S., Piai, A., Kosol, S., Varadi, M., Bekesi, A., Lebrun, P., . . . Tompa, P. (2017). Linking functions: an additional role for an intrinsically disordered linker

domain in the transcriptional coactivator CBP. *Scientific Reports, 7*(1), 4676. doi:10.1038/s41598-017-04611-x

Cunliffe, V. T. (2003). Memory by modification: the influence of chromatin structure on gene expression during vertebrate development. *Gene, 305*(2), 141-150. doi:https://doi.org/10.1016/S0378-1119(03)00386-X

De Guzman, R. N., Liu, H. Y., Martinez-Yamout, M., Dyson, H. J., & Wright, P. E. (2000). Solution structure of the TAZ2 (CH3) domain of the transcriptional adaptor protein CBP11Edited by M. F. Summers. *Journal of Molecular Biology, 303*(2), 243-253. doi:https://doi.org/10.1006/jmbi.2000.4141

Du, F., Liu, Y.-Q., Xu, Y.-S., Li, Z.-J., Wang, Y.-Z., Zhang, Z.-X., & Sun, X.-M. (2021). Regulating the T7 RNA polymerase expression in E. coli BL21 (DE3) to provide more host options for recombinant protein production. *Microbial Cell Factories, 20*(1), 189. doi:10.1186/s12934-021-01680-6

Duong-Ly, K. C., & Gabelli, S. B. (2014). In J. Lorsch (Ed.), *Methods in Enzymology* (Vol. 541, pp. 209-229): Academic Press.

Dyson, H. J., & Wright, P. E. (2016). Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-binding Protein (CBP) and p300 *. *Journal of Biological Chemistry, 291*(13), 6714-6722. doi:10.1074/jbc.R115.692020

Felsenfeld, G. (2014). A brief history of epigenetics. *Cold Spring Harbor perspectives in biology, 6*(1), a018200.

Feng, H., Jenkins, L. M., Durell, S. R., Hayashi, R., Mazur, S. J., Cherry, S., . . . Bai, Y. (2009). Structural basis for p300 Taz2-p53 TAD1 binding and modulation by phosphorylation. *Structure, 17*(2), 202-210. doi:10.1016/j.str.2008.12.009

Feng, H., Jenkins, L. M. M., Durell, S. R., Hayashi, R., Mazur, S. J., Cherry, S., . . . Appella, E. (2009). Structural basis for p300 Taz2-p53 TAD1 binding and modulation by phosphorylation. *Structure, 17*(2), 202-210.

Ferreon, J. C., Martinez-Yamout, M. A., Dyson, H. J., & Wright, P. E. (2009). Structural basis for subversion of cellular control mechanisms by the adenoviral E1A oncoprotein. *Proceedings of the National Academy of Sciences, 106*(32), 13260-13265.

Gelato, K. A., & Fischle, W. (2008). Role of histone modifications in defining chromatin structure and function.

Goldman, P. S., Tran, V. K., & Goodman, R. H. (1997). The multifunctional role of the co-activator CBP in transcriptional regulation. *Recent progress in hormone research, 52*, 103-119; discussion 119.

Gottesman, S. (1990). [11] Minimizing proteolysis in Escherichia coli: genetic solutions. In *Methods in enzymology* (Vol. 185, pp. 119-129): Elsevier.

Greenfield, N. J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols, 1*(6), 2876-2890.

Hemamalini, N., Ezhilmathi, S., & Mercy, A. A. (2020). Recombinant protein expression optimization in Escherichia coli: a review. *Indian Journal of Animal Research, 54*(6), 653-660.

Hennekam, R. C. M. (2006). Rubinstein–Taybi syndrome. *European Journal of Human Genetics, 14*(9), 981-985. doi:10.1038/sj.ejhg.5201594

Igo-Kemenes, T., Hörz, W., & Zachau, H. G. (1982). Chromatin. *Annual review of biochemistry, 51*(1), 89-121.

Imaizumi, K., & Kuroki, Y. (1991). Rubinstein-Taybi syndrome with de novo reciprocal translocation t (2; 16)(p13. 3; p13. 3). *American journal of medical genetics, 38*(4), 636-639.

Isalan, M., Klug, A., & Choo, Y. (1998). Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. *Biochemistry, 37*(35), 12026-12033.

Izumi, K. (2016). Disorders of Transcriptional Regulation: An Emerging Category of Multiple Malformation Syndromes. *Molecular Syndromology, 7*(5), 262-273. doi:10.1159/000448747

Janknecht, R., & Hunter, T. (1996). A growing coactivator network. *Nature, 383*(6595), 22-23.

Jenkins, L. M. M., Yamaguchi, H., Hayashi, R., Cherry, S., Tropea, J. E., Miller, M., . . . Mazur, S. J. (2009). Two Distinct Motifs within the p53 Transactivation Domain Bind to the Taz2 Domain of p300 and Are Differentially Affected by Phosphorylation. *Biochemistry, 48*(6), 1244-1255. doi:10.1021/bi801716h

Jumper, J., & Hassabis, D. (2022). Protein structure predictions to atomic accuracy with AlphaFold. *Nature methods, 19*(1), 11-12.

Kelly, S. M., & Price, N. C. (2000). The use of circular dichroism in the investigation of protein structure and function. *Current protein and peptide science, 1*(4), 349-384.

Klug, A. (1999). Zinc finger peptides for the regulation of gene expression. *Journal of Molecular Biology, 293*(2), 215-218.

Kwok, R. P., Lundblad, J. R., Chrivia, J. C., Richards, J. P., Bächinger, H. P., Brennan, R. G., . . . Goodman, R. H. (1994). Nuclear protein CBP is a coactivator for the transcription factor CREB. *Nature, 370*(6486), 223-226.

Laity, J. H., Lee, B. M., & Wright, P. E. (2001). Zinc finger proteins: new insights into structural and functional diversity. *Current Opinion in Structural Biology, 11*(1), 39-46. doi:https://doi.org/10.1016/S0959-440X(00)00167-6

Latchman, D. (2007). *Gene regulation*: Taylor & Francis.

Li, G., & Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. *Current Opinion in Genetics & Development, 21*(2), 175-186. doi:https://doi.org/10.1016/j.gde.2011.01.022

Menke, L. A., study, T. D., Gardeitchik, T., Hammond, P., Heimdal, K. R., Houge, G., . . . Hennekam, R. C. (2018). Further delineation of an entity caused by CREBBP and EP300 mutations but not resembling Rubinstein–Taybi syndrome. *American Journal of Medical Genetics Part A, 176*(4), 862-876. doi:https://doi.org/10.1002/ajmg.a.38626

Menke, L. A., van Belzen, M. J., Alders, M., Cristofoli, F., Study, T. D., Ehmke, N., . . . Hennekam, R. C. M. (2016). CREBBP mutations in individuals without Rubinstein–Taybi syndrome phenotype. *American Journal of Medical Genetics Part A, 170*(10), 2681-2693. doi:https://doi.org/10.1002/ajmg.a.37800

Miller Jenkins, L. M., Feng, H., Durell, S. R., Tagad, H. D., Mazur, S. J., Tropea, J. E., . . . Appella, E. (2015). Characterization of the p300 Taz2–p53 TAD2 complex and comparison with the p300 Taz2–p53 TAD1 complex. *Biochemistry, 54*(11), 2001-2010.

Miller, M., Dauter, Z., Cherry, S., Tropea, J. E., & Wlodawer, A. (2009). Structure of the Taz2 domain of p300: insights into ligand binding. *Acta Crystallographica Section D: Biological Crystallography, 65*(Pt 12), 1301.

Minezaki, Y., Homma, K., Kinjo, A. R., & Nishikawa, K. (2006). Human Transcription Factors Contain a High Fraction of Intrinsically Disordered Regions Essential for Transcriptional Regulation. *Journal of Molecular Biology, 359*(4), 1137-1149. doi:https://doi.org/10.1016/j.jmb.2006.04.016

O'Donnell, M., Langston, L., & Stillman, B. (2013). Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harbor perspectives in biology, 5*(7), a010108.

Ogryzko, V. V., Schiltz, R. L., Russanova, V., Howard, B. H., & Nakatani, Y. (1996). The Transcriptional Coactivators p300 and CBP Are Histone Acetyltransferases. *Cell, 87*(5), 953-959. doi:https://doi.org/10.1016/S0092-8674(00)82001-2

Petrif, F., Giles, R. H., Dauwerse, H. G., Saris, J. J., Hennekam, R. C. M., Masuno, M., . . . Breuning, M. H. (1995). Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. *Nature, 376*(6538), 348-351. doi:10.1038/376348a0

Piai, A., Calçada, Eduardo O., Tarenzi, T., Grande, Alessandro d., Varadi, M., Tompa, P., . . . Pierattelli, R. (2016). Just a Flexible Linker? The Structural and Dynamic Properties of CBP-ID4 Revealed by NMR Spectroscopy. *Biophysical Journal, 110*(2), 372-381. doi:https://doi.org/10.1016/j.bpj.2015.11.3516

Pirovano, W., & Heringa, J. (2010). Protein secondary structure prediction. *Data Mining Techniques for the Life Sciences*, 327-348.

Radhakrishnan, I., Pérez-Alvarado, G. C., Parker, D., Dyson, H. J., Montminy, M. R., & Wright, P. E. (1997). Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of CREB: A Model for Activator:Coactivator Interactions. *Cell, 91*(6), 741-752. doi:https://doi.org/10.1016/S0092-8674(00)80463-8

Ramos, Y. F., Hestand, M. S., Verlaan, M., Krabbendam, E., Ariyurek, Y., van Galen, M., . . . Zantema, A. (2010). Genome-wide assessment of differential roles for p300 and CBP in transcription regulation. *Nucleic acids research, 38*(16), 5396-5408.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature, 447*(7143), 425-432. doi:10.1038/nature05918

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology, 5*. doi:10.3389/fmicb.2014.00172

Rubinstein, J. H., & Taybi, H. (1963). Broad thumbs and toes and facial abnormalities: a possible mental retardation syndrome. *American journal of diseases of children, 105*(6), 588-608.

Sahdev, S., Khattar, S. K., & Saini, K. S. (2008). Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Molecular and cellular biochemistry, 307*, 249-264.

San-Miguel, T., Pérez-Bermúdez, P., & Gavidia, I. (2013). Production of soluble eukaryotic recombinant proteins in E. coli is favoured in early log-phase cultures induced at low temperature. *SpringerPlus, 2*(1), 89. doi:10.1186/2193-1801-2-89

Seera, S., & Nagarajaram, H. A. (2022). Effect of disease causing missense mutations on intrinsically disordered regions in proteins. *Protein and Peptide Letters, 29*(3), 254-267.

Sheahan, T. W., Major, V., Webb, K. M., Bryan, E., & Voigt, P. (2020). The TAZ2 domain of CBP/p300 directs acetylation towards H3K27 within chromatin. *bioRxiv*.

Sima, A., Smădeanu, R. E., Simionescu, A. A., Nedelea, F., Vlad, A. M., & Becheanu, C. (2022). Menke-Hennekam Syndrome: A Literature Review and a New Case Report. *Children (Basel), 9*(5). doi:10.3390/children9050759

Smith, J. L., Freebern, W. J., Collins, I., De Siervi, A., Montano, I., Haggerty, C. M., . . . Petersen, D. W. (2004). Kinetic profiles of p300 occupancy in vivo predict common features of promoter structure and coactivator recruitment. *Proceedings of the National Academy of Sciences, 101*(32), 11554-11559.

Sobulo, O. M., Borrow, J., Tomek, R., Reshmi, S., Harden, A., Schlegelberger, B., . . . Zeleznik-Le, N. J. (1997). MLL is fused to CBP, a histone acetyltransferase, in

therapy-related acute myeloid leukemia with a t(11;16)(q23;p13.3). *Proceedings of the National Academy of Sciences, 94*(16), 8732-8737. doi:doi:10.1073/pnas.94.16.8732

Song, C.-Z., Keller, K., Chen, Y., Murata, K., & Stamatoyannopoulos, G. (2002). Transcription coactivator CBP has direct DNA binding activity and stimulates transcription factor DNA binding through small domains. *Biochemical and biophysical research communications, 296*(1), 118-124.

Spector, D. L. (2003). The dynamics of chromosome organization and gene regulation. *Annual review of biochemistry, 72*(1), 573-608.

Spiegel, J., Adhikari, S., & Balasubramanian, S. (2020). The Structure and Function of DNA G-Quadruplexes. *Trends in Chemistry, 2*(2), 123-136. doi:10.1016/j.trechm.2019.07.002

Strålfors, A., & Ekwall, K. (2006). Heterochromatin and euchromatin—organization, boundaries, and gene regulation. *Reviews in Cell Biology and Molecular Medicine*.

Tanaka, Y., Naruse, I., Maekawa, T., Masuya, H., Shiroishi, T., & Ishii, S. (1997). Abnormal skeletal patterning in embryos lacking a single Cbp allele: a partial similarity with Rubinstein–Taybi syndrome. *Proceedings of the National Academy of Sciences, 94*(19), 10215-10220.

Theunissen, T. W., & Jaenisch, R. (2017). Mechanisms of gene regulation in human embryos and pluripotent stem cells. *Development, 144*(24), 4496-4509. doi:10.1242/dev.157404

van Genderen, M. M., Kinds, G. F., Riemslag, F. C. C., & Hennekam, R. C. M. (2000). Ocular features in Rubinstein-Taybi syndrome: investigation of 24 patients and review of the literature. *British Journal of Ophthalmology, 84*(10), 1177-1184. doi:10.1136/bjo.84.10.1177

Vera, A., González-Montalbán, N., Arís, A., & Villaverde, A. (2007). The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnology and bioengineering, 96*(6), 1101-1106.

Wojciak, J. M., Martinez-Yamout, M. A., Dyson, H. J., & Wright, P. E. (2009). Structural basis for recruitment of CBP/p300 coactivators by STAT1 and STAT2 transactivation domains. *Embo j, 28*(7), 948-958. doi:10.1038/emboj.2009.30

Wolfe, S. A., Nekludova, L., & Pabo, C. O. (2000). DNA recognition by Cys2His2 zinc finger proteins. *Annual review of biophysics and biomolecular structure, 29*(1), 183-212.

Woodcock, C. L., & Dimitrov, S. (2001). Higher-order structure of chromatin and chromosomes. *Current Opinion in Genetics & Development, 11*(2), 130-135. doi:https://doi.org/10.1016/S0959-437X(00)00169-6

Yao, J., Liu, Y., Liang, H. G., Zhang, C., Zhu, J. Z., Qin, X., . . . Yu, Z. N. (2005). The effect of zinc(II) on the growth of E . coli studied by microcalorimetry. *Journal of Thermal Analysis and Calorimetry, 79*(1), 39-43. doi:10.1007/s10973-004-0559-4

Yin, J., Li, G., Ren, X., & Herrler, G. (2007). Select what you need: a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *Journal of biotechnology, 127*(3), 335-347.

Zhang, Y.-B., Howitt, J., McCorkle, S., Lawrence, P., Springer, K., & Freimuth, P. (2004). Protein aggregation during overexpression limited by peptide extensions with large net negative charge. *Protein Expression and Purification, 36*(2), 207-216.

Appendix

SLIM Primers

Mutant 1:

Ile1776Thr Forward Standard Sequence (5' to 3'):

TTGGTTCATGCGGCTCAATGT

Ile1776Thr Forward Tail Sequence (5' to 3'):

ATACAACGTGCAACCCAAAGCTTGGTTCATGCGGCTCAATGT

Ile1776Thr Reverse Standard (reversed) Sequence (5' to 3'):

AGATACCCTTCTACTTTCTTG

Ile1776Thr Reverse Tail (reversed) Sequence (5' to 3'):

GCTTTGGGTTGCACGTTGTATAGATACCCTTCTACTTTCTTG

Mutant 2:

Cys1819Phe Forward Standard Sequence (5' to 3'):

TTGATCGCCCTGGCAGCGTAC

Cys1819Phe Forward Tail Sequence (5' to 3'):

GGTTGTCCAGTGTTCAAGCAGTTGATCGCCCTGGCAGCGTAC

Cys1819Phe Reverse Standard (reversed) Sequence (5' to 3'):

ACCATTGGTTTTGCGCTTGCA

Cys1819Phe Reverse Tail (reversed) Sequence (5' to 3'):

CTGCTTGAACACTGGACAACCACCATTGGTTTTGCGCTTGCA

Mutant 4

Cys1838Tyr Forward Standard Sequence (5' to 3'):

CCGTTCTGCCTGAACATCAAG

Cys1838Tyr Forward Tail Sequence (5' to 3'):

CAAGAGAACAAGTACCCGGTGCCGTTCTGCCTGAACATCAAG

Cys1838Tyr Reverse Standard (reversed) Sequence (5' to 3'):

GCAGTGTTTCGCATGGTACGC

Cys1838Tyr Reverse Tail (reversed) Sequence (5' to 3'):

CACCGGGTACTTGTTCTCTTGGCAGTGTTTCGCATGGTACGC

Mutant 5
Arg1868Trp Forward Standard Sequence (5' to 3'):
GCTACCATGAACACCCGTAAC
Arg1868Trp Forward Tail Sequence (5' to 3'):
CAGCTGATGAGATGGCGTATGGCTACCATGAACACCCGTAAC
Arg1868Trp Reverse Standard (reversed) Sequence (5' to 3'):
CGCCTGCTGCAGGCGGTGCTG
Arg1868Trp Reverse Tail (reversed) Sequence (5' to 3'):
CATACGCCATCTCATCAGCTGCGCCTGCTGCAGGCGGTGCTG

CBP Protein Sequences

N terminal tag: HHHHHHSSGLVPRG
Molecular weight: 1594.72

Mutant 1 Ile1776thr:
MSPQESRRVSIQRATQSLVHAAQCRNANCSLPSCQKMKRVVQHTKGCKRKT
NGGCPVCKQLIALAAYHAKHCQENKCPVPFCLNIKHKLRQQQIQHRLQQAQM
RRRMATMNTRN
Molecular weight: 13218.57

Mutant 2 Cys1819Phe:
MSPQESRRVSIQRAIQSLVHAAQCRNANCSLPSCQKMKRVVQHTKGCKRKT
NGGCPVFKQLIALAAYHAKHCQENKCPVPFCLNIKHKLRQQQIQHRLQQAQL
MRRRMATMNTRN
Molecular weight: 13274.66

Mutant 4 CYS1838TYR:
MSPQESRRVSIQRAIQSLVHAAQCRNANCSLPSCQKMKRVVQHTKGCKRKT

NGGCPVCKQLIALAAYHAKHCQENKYPVPFCLNIKHKLRQQQIQHRLQQAQL
MRRRMATMNTRN
Molecular weight: 13290.66

Mutant 5 Arg1868Trp:
MSPQESRRVSIQRAIQSLVHAAQCRNANCSLPSCQKMKRVVQHTKGCKRKT
NGGCPVCKQLIALAAYHAKHCQENKCPVPFCLNIKHKLRQQQIQHRLQQAQL
MRWRMATMNTRN
Molecular weight: 13260.65

CBP 1763-1855, WT TAZ2
MSPQESRRVSIQRCIQSLVHACQCRNANCSLPSCQKMKRVVQHTKGCKRKT
NGGCPVCKQLIALCCYHAKHCQENKCPVPFCLNIKHKLRQQQ
Molecular weight: 10622.63

CBP 1763-1855, TAZ2 C4A
MSPQESRRVSIQRAIQSLVHAAQCRNANCSLPSCQKMKRVVQHTKGCKRKT
NGGCPVCKQLIALAAYHAKHCQENKCPVPFCLNIKHKLRQQQ
Molecular weight: 10494.39

CBP 1763-1876, WT TAZ2-IDR4
MSPQESRRVSIQRCIQSLVHACQCRNANCSLPSCQKMKRVVQHTKGCKRKT
NGGCPVCKQLIALCCYHAKHCQENKCPVPFCLNIKHKLRQQQIQHRLQQAQL
MRRRMATMNTRN
Molecular weight: 13358.86

CBP 1763-1876, TAZ2-IDR4 C4A
MSPQESRRVSIQRAIQSLVHAAQCRNANCSLPSCQKMKRVVQHTKGCKRKT
NGGCPVCKQLIALAAYHAKHCQENKCPVPFCLNIKHKLRQQQIQHRLQQAQL
MRRRMATMNTRN
Molecular weight: 13230.62

EMSA and BLItz DNA Probes

50% Probe

5’ GGA GGC AAA CGC AGA ACA ATG GTT ACT ATT 3’

3’ CCT CCG TTT GCG TCT TGT TAC CAA TGA TAA 5’

60% Probe

5’ GTA GGC GGT GCT ACA CGG TTC CTG AAG TG 3’

3’ CAT CCG CCA CGA TGT GCC AAG GAC TTC AC 5’

gTEL45 G Quadruplex Probe

GGG TTA GGG TTA GGG TTA GGG TTA GGG TTA GGG TTA GGG TTA GGG

Web Resources

DNA Sequence generator: https://faculty.ucr.edu/~mmaduro/random.htm

GelAnalzer software: https://www.freeonlinegelanalyzer.com/

ProtParam: https://web.expasy.org/protparam/