DEPARTMENT OF BIOCHEMISTRY
UNIVERSITY OF OXFORD

Investigating potential allostery in the transcription factor CREB

Ngai Lam Chung
Master of Science by Research in Biochemistry
St. John's College
Candidate Number: 1040402

Supervised by Dr. Sarah Shammas

Department of Biochemistry
3 South Parks Road, Oxford, OX1 3QU

Word Count: 30,000

Acknowledgements

I would like to take this opportunity to thank Dr. Sarah Shammas for her meticulous guidance, support and insight throughout this project. I am very grateful be given the opportunity to work on such an important project for the Shammas laboratory. Without her efforts and time, this project would not be possible. I would like to thank the other members of the Shammas lab: Dr. Mikhail Kuravsky, Conor Kelly, Jeremie Gaudez and Iris Kalenderoglou. They were extremely helpful in times of need and were great colleagues to work alongside with. I would also like to thank Professor Colin Kleanthous in providing his insight and help in the latter stages of the project.

I would also like to thank Professor Justin Benesch and his lab members, particularly Joshua Bishop and Zihao Wang. The data they collected for was an essential part of this project. I would also like to thank them for their incredible flexibility and willingness to help even at the shortest notice of time.

Finally, thank you to my family who has been there for me since day 1, in the most literal sense. Without their love and support, I would never be who I am today and done what I have achieved. Thank you.

Abstract

Intrinsically disordered proteins (IDPs) are incredibly prevalent in eukaryotic transcription factors as disorder in structure allows for great flexibility and multi-faceted pathways of signal transduction and protein regulation. The basic leucine zipper (bZIP) is one of the three eukaryotic disordered DNA-binding domains. One particular basic zipper, cAMP-response element binding protein (CREB), is predicted to be one of the most disordered proteins in the human genome. CREB is a major signal transducer of key signalling pathways including cAMP, Ca^{2+} etc., and regulate the transcription of over 4,000 genes. Ubiquitously expressed in all eukaryotic cells, CREB has been found to be essential in cell survival and memory formation. The misregulation of CREB *in vivo* has been attributed to various forms of cancer and hence must be tightly regulated by many sources.

Despite being one of the most well-studied transcription factors in terms of signal transduction, the mechanism of CREB DNA binding and search is yet to be elucidated. It is known that CREB must form a homodimer, bound to its specific DNA sequence - the cAMP response element (CRE) and activated by serine-133 phosphorylation, in order to promote downstream gene transcription by recruiting co-factors and members of the basal transcription complex. However, the exact order of events that occur leading up to cofactor recruitment remains unknown.

The Shammas lab, through biophysical and structural studies of the DNA-binding domain (bZIP), found that the basic zipper strictly follows a dimer pathway in binding to CRE DNA. This study aims to contextualise this finding by performing biophysical characterisation on a HaloTagged full length CREB construct. First, an expression and purification protocol was developed and optimised to purify HaloCREB free from nucleic acid and protein contamination. Fluorescent anisotropy, stopped-flow kinetics and other biophysical techniques were then used to estimate various key parameters in determining the DNA binding and search pathway in CREB.

This study found HaloCREB to exhibit similar binding kinetics as the bZIP construct. A fast association rate of 10^{10} M^{-1} s^{-1} combined with a dissociation constant of 0.1 s^{-1} yields a remarkable $K_{d,CRE,kinetics}$ of subnanomolar

scale. However, a 10-fold mismatch between the K_d from equilibrium and kinetics study suggests that, if the results are reproducible, a two-state reaction is insufficient to describe how HaloCREB binds to DNA. This infers the involvement of the monomer binding pathway.

In comparison with existing literature and the bZIP data collected by the Shammas laboratory, my finding suggests that HaloCREB does not interfere with CREB DNA-binding. This observation is consistent with other *in vivo* studies. Moreover, a 3-fold difference in homodimerisation K_d and a 5-fold difference in association and dissociation binding constants separates HaloCREB from the bZIP construct. While small, these changes reveal possible, yet to be understood allosteric effect of the rest of the CREB protein in modulating the DNA-binding kinetics of CREB.

Lastly, using an experimental set-up that is a better approximation to the cellular environment, I observed a slow-down in association rates when CRE DNA is in presence of 200-fold excess non-target DNA. This novel result highlights the need to consider the effect of non-target DNA in the biophysical characterisation of transcription factors. The estimation of biophysical parameters presented in this study will enable a model to be developed for a holistic understanding of the binding and searching mechanism of CREB.

Table of Contents

Abbreviations

BR	Basic region
bZIP	Basic zipper DNA binding domain
CBP	cAMP response element binding protein-binding protein
CRE	cAMP response element
CREB	cAMP response element binding protein
CREh	Hairpin CRE
CREM	cAMP response element modulator
CRElin	Linear CRE
FA	Fluorescence anisotropy
FD	Forced dimer
HaloCREB	HaloTag-CREB construct
His6	Hexahistidine tag
ID	Intrinsic disorder
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
ITC	Isothermal calorimetry
KID	Kinase inducible domain
KIX	Kinase inducible domain-interacting domain
LZ	Leucine zipper
miR/miRNA	microRNA
NLS	Nuclear localisation sequence
PCR	Polymerase chain reaction
pI	Isoelectric point
PTM	Post-translational modification
Q1	Glutamine rich region 1
Q2	Glutamine rich region 2
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SF	Stopped-flow
TEV	Tobacco Etch Virus

1. Introduction

1.1 Disorder but not Dysfunctional: The Case for Intrinsic Disorder

The *protein-structure paradigm*, as postulated by Fischer in 1894[1] describes a binary relationship between a protein's tertiary structure and its physiological function: a protein's structure determines its function and its physiological relevance is limited by the three-dimensional structure it can assume. The first structural determination of a protein — myoglobin[2], revealed the globular basis of protein structure that is familiar to us all. The explosive growth of protein structure publications paired with functional studies in the following decades further strengthened the undeniable link between a protein's structure with its functionality.

However recent bioinformatic analysis of the human proteome has revealed a sizeable number of apparent exceptions to this paradigm, ones that are overlooked by structural studies[3,4]. These protein and domains are referred to as intrinsically disordered (ID), meaning they have no fixed tertiary structure, and only fluctuating levels of partial secondary structure[5]. These domains are found to be enriched in charged, disordered-promoting residues such as proline (P), glutamic acid (E), lysine (K), aspartic acid (D) and serine (S), while depleted in hydrophobic and bulky order-promoting residues of tryptophan (W), tyrosine (Y), phenylalanine (F), isoleucine (I) and cysteine (C)[5,6] that are often found in the core of folded proteins. Most notably, prolines appear twice as often in intrinsically disordered proteins compared to globular proteins[7]. The cyclic nature of proline means it has no backbone hydrogen atom, preventing it from forming stabilizing hydrogen bonds, hence proline is rarely found in α-helices or β-sheets[8]. The bias in amino acid composition, particularly the increased prevalence of charged residues and the decrease in hydrophobic and aromatics residues, reduces the contribution of hydrophobic effect in the initiation and propagation of protein folding[9]. Additionally, the electrostatic repulsion from the enrichment of charged residues further promotes the formation of elongated, disordered structures[10]. It is important to note that the formation of transient secondary structure and folding-upon-binding are both a common occurrence within ID proteins (IDP) and domains (IDR)[11].

A common metric to define whether a protein or domain is ID is the Uversky plot[10] (Figure 1.1). The Uversky plot, or the charge-hydropathy plot, describes a relationship between the mean hydropathy and the mean net charge of a protein. The mean hydropathy represents the contribution of the hydrophobic force in protein folding, whereas the net charge corresponds to the contribution of charge repulsion in unfolding. A highly charged protein with relatively low hydropathy is considered disordered, whereas a protein with high hydropathy and low net charge is natively folded.

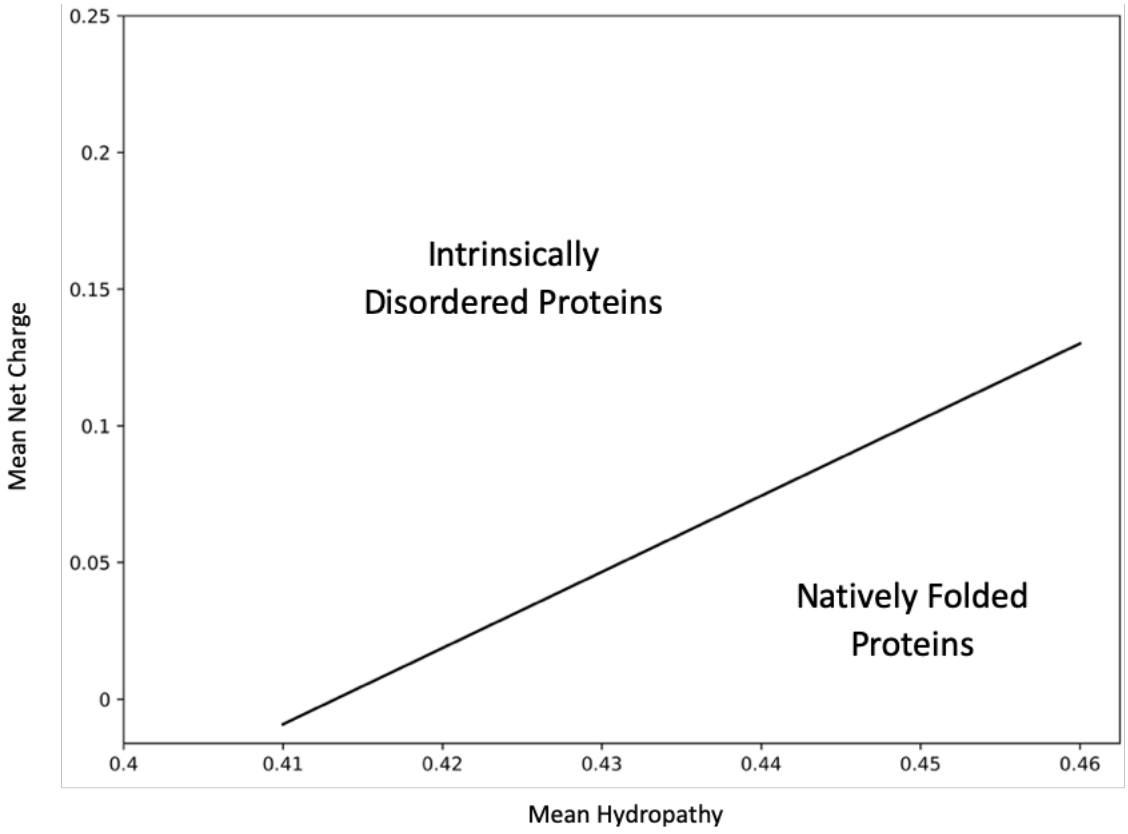


Figure 1.1 | The Uversky plot of predicted protein structure in solution. Mean hydropathy of a protein is calculated by Kyte and Doolittle approximation across 5 amino acids[12]. Mean net charge is the net charge of the protein at pH 7.0 divided by the number of residues. Black line represents the border between natively folded proteins and ID proteins. The equation of the border is <NC> = 2.785<H> - 1.151, where <NC> describes the mean net charge and <H> describes the mean hydropathy.

The lack of structure in ID proteins grant these proteins unique characteristics over the conventional globular proteins. Their ability to form only transient secondary structures is reflected in their characteristic energy landscape (Figure 1.2). The landscape of a globular protein shapes like a funnel, describing the directed folding in search the global energy minima — its native state[13]. On the other hand, the energy landscape of ID proteins are comparatively flat, containing countless local energy minima of similar energy levels[14]. Given the low energetic barrier between the local minima, the protein is able transverse the landscape and

transition between local minima at ease. The lack of a global energy minima entails the disordered protein or domain to sample the entirety of the conformational landscape. As such, ID proteins are heterogeneous in structural composition and exist as an ensemble in their native state. In turn, their ability to adopt various secondary structures establish these proteins as excellent candidates for protein interaction hubs, transcription factors and signal transducers[15].

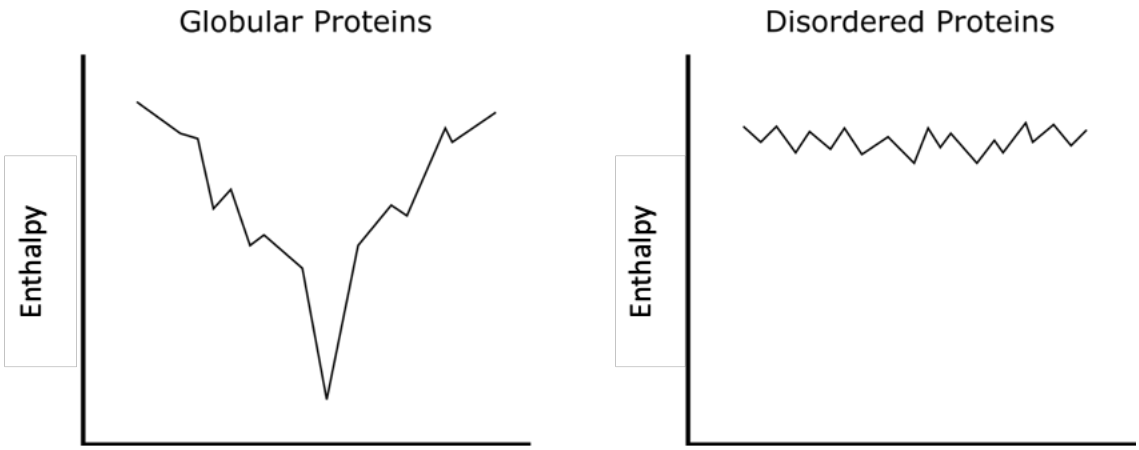


Figure 1.2 | Conformational landscape of globular and ID proteins. The funnel-like landscape describes the process of protein folding. The global energy minima represents the native state of the protein. The flat landscape of disordered proteins contain many local energy minima of similar energy level, representing the protein's ability to shift between locations within the conformational landscape and form transient secondary structures. The width of the funnel represents the entropy of the system.

1.2 Disorder in eukaryotes

Intrinsically disordered proteins frequently function in interaction hubs, and eukaryotic transcription factors (TF) are very often largely disordered[16]. So called "coupled folding-upon-binding" mechanisms[17], where disordered regions can fold into a desired conformation upon association with their binding partner are quite common in these signalling reactions.

This enables precise control over the reaction, essential in transient protein-protein or protein-nucleic acid interactions[18,19], where any mistake in protein function could be detrimental to the survival of the cell or multicellular organism.

In addition, molecular recognition is promoted within ID regions through the disproportionally large binding and interacting surface area to overcome steric restrictions[20]; faster rate of association by the reduction in dependence of orientation-specific interactions[21]; and the ability to fine tune the binding free energy to maximise the specificity of the interaction[15].

These characteristics are a fit for key biological processes which require a high degree of adaptability, rapidity and complexity gained through their protein interactions[22]. This phenomenon is reflected in the distribution of ID regions in the human proteome and varying protein functionality. A third of the human proteome contains significant regions of ID (≥30 residues)[23], and this number was predicted to be up to 65% in signalling proteins; 75% in cancer-associated proteins[24]; and 94% in TFs[19]. Conversely, proteins involved in enzymatic catalysis or transport are strongly associated with ordered proteins[15,25] (Table 1.1). Catalytic proteins, have a strong preference for a stable folded structure with $\Delta G \lesssim -1$ kcal/mol, consistent with the strong thermodynamic requirement in enzymes. Conversely, ID proteins whose major function is to bind other proteins, can tolerate much higher degrees of disorder without reducing functionality[15].

Additional complexity and promiscuity in protein-protein interaction and signal transduction can be introduced through post-translational modifications (PTMs). ID proteins and regions frequently contain motifs that stimulate modification[26], and are more solvent accessible by modifying enzymes[27]. These motifs are much more easily accommodated in disordered regions because they are accessible, and do not interrupt the evolved folded structure. PTMs can modulate rate(s) of formation and dissolution, and therefore binding affinities, of complexes involving folding and without folding[26,28]. ID proteins typically contain multiple types of PTM sites, such as: phosphorylation, acetylation, methylation, glycosylation and ubiquitination. Sequential or combinatorial modifications can generate ultra-sensitive switch-like behaviour[29,30] or a rheostat-like sensing mechanism[31] to further potentiate transcription and binding regulation.

The high degree of regulation of gene expression and signal transduction from ID proteins and regions reflect the need of a complex but efficient system to regulate the chromosomal nature of eukaryotic gene organization. In organisms with simpler gene systems, such as archaea and bacteria, the prevalence of ID proteins or regions is significantly lower. Only 2% of archaean and 4.2% bacterial proteins contain a long region of ID[23] compared to 33% in eukaryotes. Proteins in these classes of organisms are less involved in the cellular functions where protein disorder is commonly found in eukaryotes (table 1.i)[11]. As such, proteins of ID play an incredibly important role in eukaryotic gene regulation and transcription.

Cellular Processes	
Top correlation with predicted intrinsic disorder	Top correlation with predicted order
Differentiation	GMP biosynthesis
Transcription	Amino acid biosynthesis
Transcription regulation	Transport
Spermatogenesis	Electron transport
DNA condensation	Lipid A biosynthesis
Cell cycle	Aromatic hydrocarbon metabolism
mRNA processing	Glycolysis
mRNA splicing	Purine biosynthesis
Mitosis	Pyrimidine biosynthesis
Apoptosis	Carbohydrate metabolism

Table 1.i | Comparison between cellular processes that correlate strongest with predicted intrinsic disorder and order[25]. Proteins from SWISS-PROT were analysed through a disorder predictor and correlated with the associated SWISS-PROT keywords.

1.3 Modular eukaryotic transcription factors

Eukaryotic transcription factors (TFs) are often described in terms of a modular architecture and generally include at least one DNA-binding domain and transactivation domain, separated by flexible linker regions[32]. DNA-binding domains allow the TF to associate with the nucleic acid, while the transactivation domains interacts with other co-activators and cofactors, contributing to the transcriptional regulation activity of the protein[33]. The two modules working together yield a functional protein that can transduce cellular signals and affect patterns of gene expression.

Transactivation domains are 30-100 residues in length and are significantly disordered in solution. Main trans-activating motifs, such as acidic, glutamine-rich and proline-rich domains are characterized by their low sequence complexity with a predicted disorder content ranging from 73% to 94%[15]. These domains contain a wide range of protein interaction sites and PTM motifs and can therefore act as an interaction hub, facilitating multifaceted protein-protein and protein-nucleic interactions to create a complex gene regulation network.

1.4 Disordered DNA Binding domains

While most of the DNA-binding domains are ordered in nature, several are disordered in solution but fold upon binding to their target DNA, most notably the AT-hook, basic helix loop helix (bHLH) and the basic leucine zipper (bZIP)[15]. Ordered transcription factors search along DNA by following a helical path along the major groove.[34] However, the search mechanism of transcription factors with a disordered DNA binding domain is poorly understood[28].

1.4.1 AT-hooks

AT-hooks originate from the mammalian high mobility group (HMG-I/Y) chromosomal proteins and can bind to the minor groove of AT-rich DNA[35]. The short motif centers around a highly conserved pentapeptide of

proline-arginine-glycine-arginine-proline[36]. NMR simulated annealing[37] structures reveal that the AT-hook remains entirely disordered while bound to DNA (Figure 1.3). The arginine-glycine-arginine core inserts itself into the minor groove while nearby lysines form stabilising electrostatic interactions with the phosphate backbone[37,38]. AT-hooks bind to the minor groove, they have been found to act as accessory DNA-binding domains for proteins that bind the major groove and have been attributed to stabilising unconventional DNA structures such as four-way junctions[39].

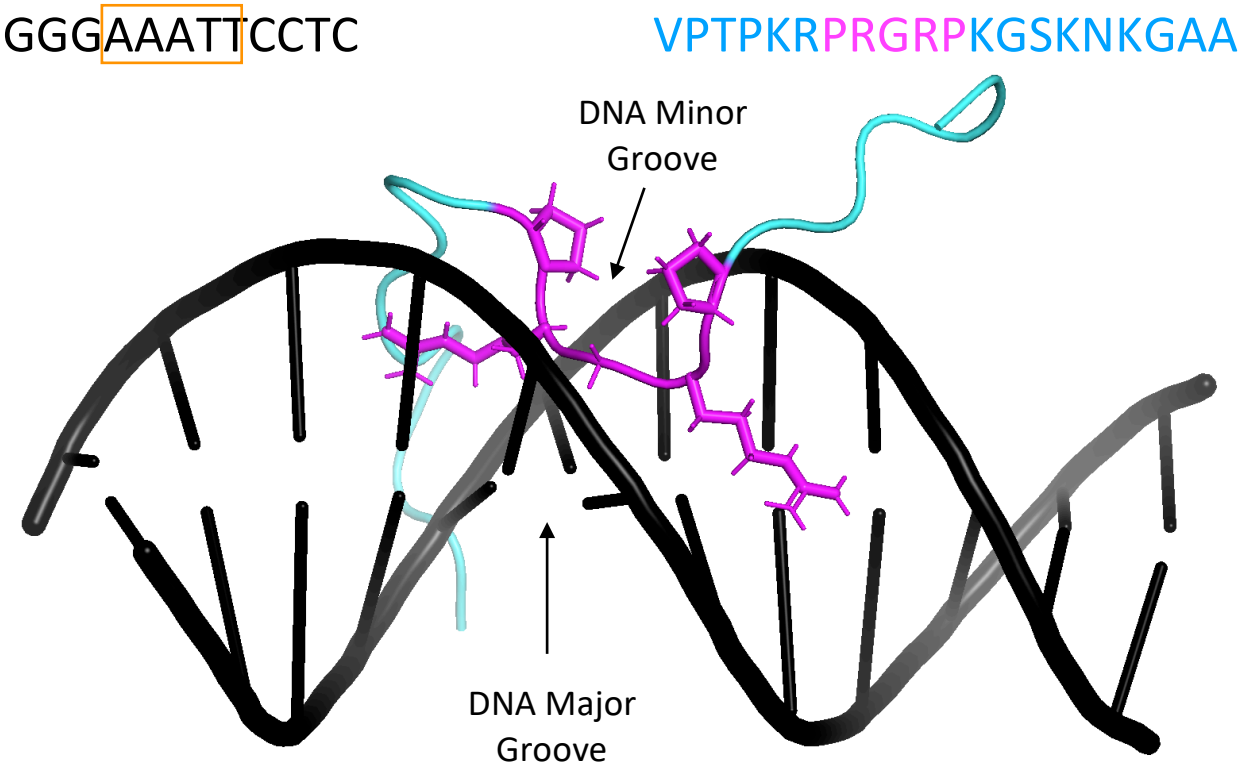


Figure 1.3 | Binding pose of AT-hook from HMG-I (residues 51-71) to the minor groove of an AT-rich 12-mer DNA sequence (PDB: 2EZD). Structure determined by NMR simulated annealing[37]. The 12-mer DNA is represented in black with the AT-rich region highlighted by orange. Truncated HMG-I in cyan. The five conserved residues that forms the AT-hook are highlighted in magenta. The arginine-glycine-arginine core inserts itself into the minor groove while nearby lysines form stabilising electrostatic interactions with the phosphate backbone[37,38].

1.4.2 Basic leucine zippers (bZIPs)

The basic zipper (bZIP) is the second most common DNA-binding motif[15] and can be found within proteomes of all eukaryotes[40]. A bZIP dimer is a coiled-coil chopsticks-like structure and consists of two bZIP domains each containing a basic region (BR) and a leucine zipper (LZ)(Figure 1.4)[41].

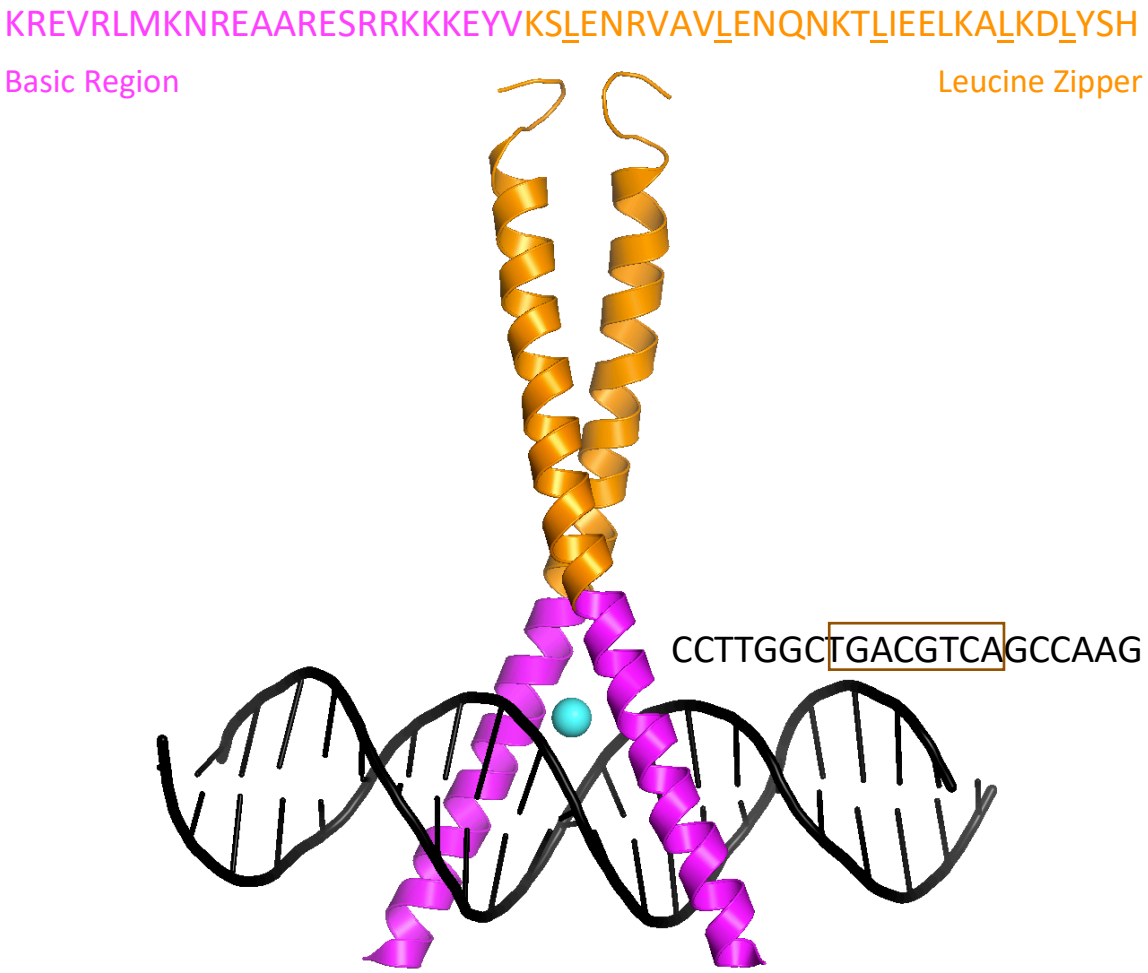


Figure 1.4 | Binding pose of bZIP dimer from CREB (residues 285-339) to a 21-mer DNA sequence containing the somatostatin CRE site (PDB: 1DH3). Structure determined by x-ray crystallography[42]. Magenta represents the basic region and orange represents the leucine zipper. Cyan represents the Mg^{2+} ion. DNA sequence in black and the somatostatin CRE is highlighted in brown. Coiled-coil formation upon dimerisation allow the basic regions to interact with the palindromic CRE sites

The Basic Region (BR)

The basic region is enriched in positively charged residues, and through the formation of stabilizing ionic interactions with the phosphate backbone serves as the DNA-binding domain[40]. It interacts with the major groove of DNA, folding into an α-helical structure[42]. In the absence of DNA, biophysical and structural studies have revealed that the disorder of the BR varies between species. GCN4 and C/EBPβ are mostly unfolded in

solution[43] and have been described with transient helical structures, whereas the BR of ATF-4 forms a stable helix in solution. In addition to its DNA-binding activities, the BR also facilitates specific interactions between cellular and viral proteins to further modulate gene expression[41]. The BR also contains a bipartite nuclear localisation sequence (NLS), separated by a 10-12 residue linker[44]. The NLS motif is one of the common motifs recognised by importin-α, facilitating the nuclear import of bZIPs to regulate gene transcription[45].

The DNA specificity of the varying bZIP proteins is modulated by the BR sequence but they preferentially recognise short palindromic sequences. The two most commonly recognised motifs within the bZIP family are AP-1 sites (TGACTCA) and cAMP response element (CRE) sites (TGACGTCA)[46]. The sequence is recognised through the highly conserved key recognition element of N-X_7-R/K. Substitutions amongst the five contacting amino acids dictates the exact recognition sequence[41]. The DNA specificity of the dimer is further modulated by the hinge region between the BR and the LZ[41], as well as the presence of ion between the basic regions[42].

The Leucine Zipper

The leucine zipper acts as the dimerisation domain. Organised in heptads corresponding to two turns in an α-helix, a leucine or other hydrophobic amino acids can be found in every three to four residues[40]. This organisation results in a hydrophobic face, which forms the dimer interface[41]. While both parts of the bZIP fold into an α-helix (upon dimerisation or DNA binding), bZIPs in monomeric form are highly disordered[15].

Dimerisation can greatly accentuate complexity of the protein's functionality since both heterodimers and homodimers may be formed. 53 bZIP sequences can be found in the human genome[47], giving the possibility to form 2809 unique dimers. This tendency of bZIP proteins to form heterodimers can lead to differential gene regulation patterns adding to the complexity of gene expression. Dimerisation in general can modulate the sensitivity of recognition (through signal noise dampening from stochastic fluctuation) and allow sharp responses to change in signal concentration[48].

While most bZIPs are known to form extensive heterodimer networks with one another[49], members of the cAMP-response element binding protein (CREB) family, OASIS family, PAR family and ATF2 family strongly favour homodimerisation within the family[47]. In fact, only 15% of the possible interactions between human bZIPs actually occur *in vivo*[50]. The specificity is determined by the sequence of heptads within the LZ and rules governing this are well described.

Each heptad corresponds to two full turns in the α-helix and the residues are labelled **a** - **g**. Residues a and d are typically hydrophobic, packed in a "hole and knob" pattern to create a hydrophobic core[51], contributing to the majority of the stabilising interactions within the leucine zipper[52]. A leucine (L) in position **d** is more thermodynamically favourable than hydrophobic alternatives of similar size (isoleucine and valine)[53], and are found in 84% of the **d** positions[47]. On the other hand, amino acids found in the **a** position are more variable as thermodynamic preference of leucine is not observed[54]. Furthermore, the interhelical **a**-**a**' interaction contributes to the dimerisation specificity. For example an asparagine (N) in position **a** prefers integration with another asparagine in **a**'[55], favouring the formation of a homodimer. Asparagine residues in the **a** position of the second heptad within homodimer-forming bZIPs are highly conserved[47]. In contrast, a lysine (K) or serine (S) residue in the **a** position prefers interaction with other amino acids, favouring the formation of heterodimers instead[55].

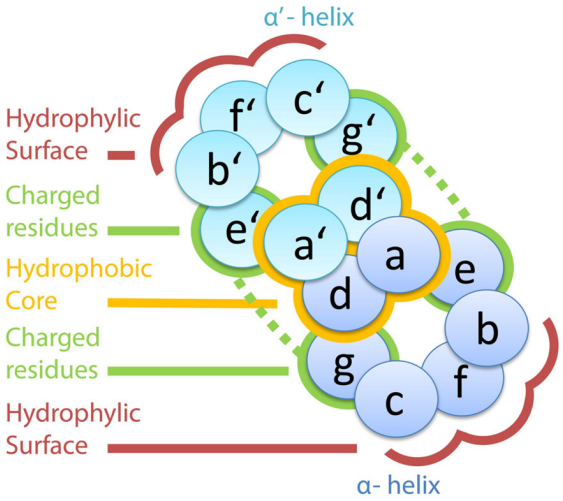


Figure 1.5 | Heptad organisation within leucine zipper dimers. Figure from Llorca et al.[39]

76% of residues **e** and **g** are one of the four charged residues: glutamine acid, aspartic acid, lysine and arginine[56]. The ionic interactions between **g** and **e**' not only contributes to the stability of the coiled coil[57], but also modulate dimerisation specificity. An aspartic acid (E) is most commonly found in the **g** position, paired with an oppositely charged lysine (K) or arginine (R) in the **e** position[47]. The E-K or E-R pair provides

stabilising coupling energy to maintain the dimer. Conversely, electrostatic repulsion between pairs of the same charge result in a destabilising coupling energy and are not favoured[55]. While optimisation of stabilising energy is preferred, studies have shown preventing repulsive interactions is more important in driving the dimerisation process than selecting for optimal coupling energy[47].

1.4.3 Basic helix-loop-helix (bHLHs)

The basic helix-loop-helix (bHLH) is another common disordered DNA binding domain[58]. Similar with bZIPs, transcription factors in the bHLH family dimerises for preferential binding to target DNA[59]. bHLH proteins are also widely involved in a diverse set of regulatory function. Likewise, notable bHLH proteins such as c-Myc are essential in survival and have been attributed to different types of cancer upon misregulation[60].

1.5 cAMP response element binding protein (CREB)

cAMP response element binding protein, CREB, is a 32kDa transcriptional activator in the bZIP family. CREB binds to the palindromic CRE sequence as a homodimer, and is activated by phosphorylation at serine 133 in its transactivation domain. Promoter-bound CREB can activate transcription of downstream genes by recruiting co-activators such as CREB-binding protein (CBP)[61], p300, CREB-regulated transcriptional coactivators (CRTCs)[62], and members of the basal transcription complex[63]. CREB is predicted to be essentially disordered along its entire length, and is one of the top ten most disordered proteins in the human proteome[15].

CREB has four well-identified domains (Figure 1.5). Two glutamine rich domains Q1 and Q2 flank the kinase induced domain (KID) – all of these are identified as activation domains[64]. The bZIP domain is in the C-terminus. 17 splice variants of CREB are currently documented on Ensembl database. The two main CREB variants: CREB_A and CREB_B, differ only by the presence of the α domain from residues 88-101 in the CREB_A isoform.

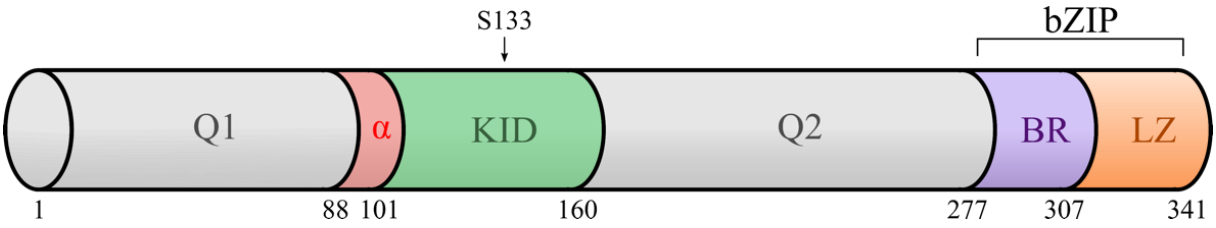


Figure 1.6 | Domain organisation of CREB_A (UniProt: P16220). Q1 comprises residues 1-88, followed by the α domain from residues 88-101. The α domain is only found in the A isoform of CREB. The kinase induced domain (KID) in green. Serine-133 can be phosphorylated, enhancing CREB's ability to activate downstream transcription. The second glutamine rich domain (Q2) follows KID from residues 160-277. Finally, the bZIP domain can be found in the C-terminus. The two subdomains: the basic region (BR) and leucine zipper (LZ) are labelled respectively in purple and orange.

1.5.1 Q1

Q1 is not required for transcriptional regulation[65] and is not found in all CREB splice isoforms. While the exact functionality of the Q1 domain remains unknown, it is termed a protein-protein interaction domain, and is known to interact with HDAC1[66] and a TATA-binding protein associated factor, $dTAF_{II}110/hTAF_{II}135$[67].

1.5.2 KID

The CREB KID acts as the transactivating domain of the protein. The most important function of the KID is the recruitment of the coactivator protein CREB Binding-Protein (CBP), and thereby assembly of the rest of the transcriptional machinery[68]. This recruitment occurs through its interaction with the folded KIX domain of CBP. Upon association with KIX, pKID folds into a kinked helical structure[69] (Figure 1.7). ϕ-value analyses revealed that the pKID has a largely unstructured transition state, with the majority of helical structure forming later in the process, ruling out a pure conformational selection mechanism[70,71].

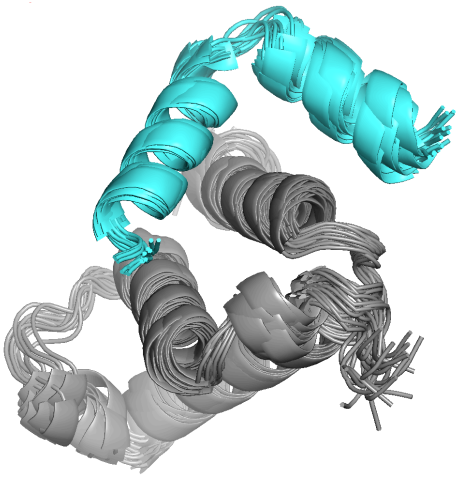


Figure 1.7 | Solution NMR ensemble structures of the pKID-KIX interaction between CREB and CBP (PDB:1KDX)[69]. All 17 structures from the PDB file are displayed. The KIX domain of CBP is in grey and the pKID domain of CREB is in cyan.

The KID-KIX interaction is enhanced by S133 phosphorylation. While KID can associate with KIX in its unphosphorylated state, the half-life of the complex is drastically reduced (~1 s for pKID vs. ~10 ms for KID) through modulation of the dissociation rate constant[72]. Along with the activating serine-133, the KID also contains other serine residues, when phosphorylated, modulate the activity of CREB. Residues S108, S111, S114, S117 form the "CK cassette" and can be phosphorylated by casein kinases I and II, yielding a rheostat-like autoinhibition response to CREB's DNA-binding ability through ionic interactions with the basic region[73,74]. S121 can be phosphorylated by ataxia-telangiectasia-mutated (ATM) kinase for a similar effect[75]. On the other hand, phosphorylation of S142 can inhibit CREB's function by disrupting tyrosine-650 in CBP, a key residue in facilitating the pKID-KIX (S133 phosphorylated KID-KID interacting domain) interaction[76]. Phosphorylation of S142 is also recently found to interfere with CREB homodimerisation and consequently reduce binding affinity to DNA[77].

The formation of a structural domain upon pKID-KIX association makes the interaction an attractive target for drug design against CREB-associated diseases. Traditionally, small molecule drugs mostly antagonise against enzymes or surface receptors by mimicking the natural substrate[6]. The structural integrity of the target proteins contains well-defined binding pockets, allowing for an effective structured-based drug design

pipeline[78]. This is generally not the case in IDPs. The lack of a consistent structure and defined substrate binding site thereby render traditional structured-based drug design inapplicable[6]. Furthermore, the disordered nature of ID proteins introduces significant difficulties in designing small molecules that are specific, efficient and potent[79]. However, the coupled folding upon binding of the pKID domain presents great similarities with that of an ordered protein for drug design: a well-defined interface with accessible binding surfaces. This strategy has already been employed to identify CREB-inhibiting small molecules[80,81], but they have not yet been proven effective in a clinical setting.

1.5.3 Q2

The Q2 domain of CREB is 117 residues in length. The similarities of the two glutamine rich regions are reflected in the nomenclature, with both regions having a relatively low sequence complexity and being enriched in glutamine residues. Q2 has been found to interact with multiple protein partners, most notably the TATA-binding protein associated factor $hTAF_{II}130/135$, a part of $TF_{II}D$[67]. This interaction is necessary for mediating the recruitment of the RNA polymerase II transcription complex. Mutations in Q2 have shown to abolish the interaction and inhibit CREB-mediated gene transcription[64].

The importance of Q2 is reflected in the activity of various CREM isoforms. CREM, cAMP response element modulator, is a bZIP transcription factor in the CREB family that also binds to CRE sequences[82]. CREM and CREB have 65% sequence identity and the same domain organisation (Figure 1.8a). Inhibiting CREM isoforms are much more common than inhibiting CREB isoforms[83]. Whilst all activating CREM isoforms contain KID, Q2 and bZIP domains, inhibiting isoforms have at least one of the three missing (Figure 1.8b). Given the high sequence identity between the two proteins, it is likely that the contribution of the domains is similar in CREB. Q2 has also been attributed a critical role in enhancing residence time and retaining promoter occupancy[84]. However, the exact mechanism of which is yet to be undetermined. Intriguingly, a splice isoform of CREB without Q2 is highly expressed in human testes (htCREB)[85]. The lack of Q2 suggests that this splice isoform is inhibitory and plays a role in spermatogenesis[85].

A)

```
                                                                                                                    Q1
CREB_A  MTMESGAENQQSGDAAVTEAENQQMTVQAQP---QIATLAQVSMPAAHATSSAPTVTLVQLPNGQTVQVHGVIQAAQPSVIQSPQVQTVQSSCKDLKRLFSGT
CREM_1  MTMET-VESQHDGSITASLTESKSAHVQTQTGQNSIPALAQVSVAGSGTRRGSPAVTLVQLPSGQTIHVQGVIQTPQPWVIQSSEIHTV--------------
        ****: .*.*:.*. :.: :*.:.  **:*.   .*.:*****:..: :  .:*:*******.***::*:****:.** ****.:::**

                KID                                                                                                 Q2
CREB_A  QISTIAESEDSQESVDSVTDSQKRREILSRRPSYRKILNDLSSDAPGVPRIEEEKSEEETSAPAITTVTVPTPIYQTSSGQYIAITQGGAIQLANNGTDGVQG
CREM_1  QVAAIAETDESAES-EGVIDSHKRREILSRRPSYRKILNELSSDVPGVPKIEEERSEEEGTPPSIATMAVPTSIYQTSTGQYIAIAQGGTIQISNPGSDGVQG
        *:::***:::* ** :.* **:**********:******:****.****.****.**** :.*:*:*::***.*****:******:***:**::* *:*****

                                                                                                                    BR
CREB_A  LQTLTMTNAAATQPGTTILQY-AQTTDG-QQILVPSNQVVVQ----------AASGDVQTYQIRTAPTSTIAPGVVMASSPA---LPTQPAEEAARKREVR
CREM_1  LQALTMTNSGAPPPGATIVQYAAQSADGTQQFFVPGSQVVVQDEETELAPSHMAAATGDMPTYQIR-APTAALPQGVVMAASPGSLHSPQQLAEEATRKRELR
        **:*****:.*. **:**:** **::** **::**..*****           **:**: ***** ***:::. *****:**.    * * ****:****:*

                                                                                                    LZ
CREB_A  LMKNREAARECRRKKKEYVKCLENRVAVLENQNKTLIEELKALKDLYCHKSD-
CREM_1  LMKNREAAKECRRRKKEYVKCLESRVAVLEVQNKKLIEELETLKDICSPKTDY
        ********.****.*********.****** ***.*****::***: . *:*
```

B)

CREM Isoform		Q1	Q1		KID	Q2		BR	LZ	Activity
	A	B	C	θ	EF	G	γ	H	Ia/Ib	
1		■	■		■	■	■	■	■	
2	■	■	■		■	■		■	■	
3	■	■	■		■	■		■	■	
5				■	■	■		■	■	
10				■	■	■		■	■	
13	■	■			■	■	■	■	■	
14	■	■			■	■		■	■	
19				■	■	■	■	■	■	
28	■	■			■	■		■	■	
6 (ICER_I)							■	■	■	
7 (ICER_Iγ)								■	■	
8 (ICER_II)							■	■	■	
9 (ICER_IIγ)								■	■	
11				■	■		■	■	■	
12				■	■		■	■	■	
15	■	■				■		■	■	
16	■	■					■	■	■	
17	■	■	■		■			■	■	
18	■	■	■		■			■	■	
20								■	■	
21								■	■	
23	■	■			■		■	■	■	
24								■	■	
27		■			■			■	■	
29	■	■			■			■	■	
4	■	■			■			■	■	
22		■			■					
25						■		■	■	
26							■	■	■	

Figure 1.8 | Comparison between CREM and CREB. A) Multiple sequence alignment between CREB_A (UniProt: P16220) and CREM_1 (UniProt: Q03060) using MUSCLE (EMBL). Activating serine-133 highlighted by black rectangle. Residues are shaded according to their corresponding domain: Q1 (grey), KID (yellow), Q2 (cyan), BR (pale), LZ (purple) B) Functional comparison between CREM splice isoforms. Sequence and isoform numbers obtained from UniProt. The presence of domain is determined by domain organisation according to UniProt and Borlikova & Endo[82]. Isoforms that are labelled green in the activity column are transcriptional activators, red for transcriptional repressors and white for undetermined activity. Function of domains were determined through multiple sequence alignment with CREB. Activity of isoforms were extracted from UniProt.

1.5.4 bZIP

The CREB bZIP domain consists of the highly conserved basic region and leucine zipper, spanning residues 277-341 in the C-terminus. As its namesake, CREB binds specifically to palindromic consensus cAMP response elements (CRE) sites in the genome, as well as closely related motifs[86]. A whole genome study by Zhang et al. revealed that out of the 740,390 half CRE sites and 10,447 full CRE sites in the genome, CREB regulates only 4,084 genes[87]. The large discrepancy between available sites and occupancy is likely due to binding restrictions imposed by the methylation state of the site[87,88].

The specificity of CREB to CRE DNA is modulated by the sequence in the basic region. R286, K292, R294, R298, K303 and K305 form contacts with the phosphate backbone, stabilising the interaction[42]. Two other particularly important residues are R301 and K304. Both residues are absolutely conserved amongst the CREB family and their substitution to another amino acid abolishes the ability of CREB to bind to DNA[89,90]. The R301 side chain interacts with the central CG bases of the CRE sequence (TGACGTCA), forming stabilising hydrogen bonds. While K304 does not form direct contacts with DNA, it is known to play a role in coordinating the Mg^{2+} ion[47].

As described earlier, the CREB BR strongly favours homodimerisation[47,55]. The characteristic asparagine in the **a** position of the second heptad can be found in all members of the CREB family. The thermodynamically favourable E-K and E-R contacts in the **e** and **g'** position can also be found in respectively the first and third heptad (Figure 1.9)[47]. Intersubunit contacts between Y307 and E312' further contribute to the stability of the homodimer and as Y307 is not found outside the CREB family, this interaction also modulates the homodimer specificity[42]. However, the preferential homodimerisation tendency of CREB bZIP lies largely in the length of its LZ. Most leucine zippers contain 5 heptads, whereas the CREB LZ only contains 4 - the two missing helical turns greatly disfavour heterodimerisation with other bZIP proteins[47].

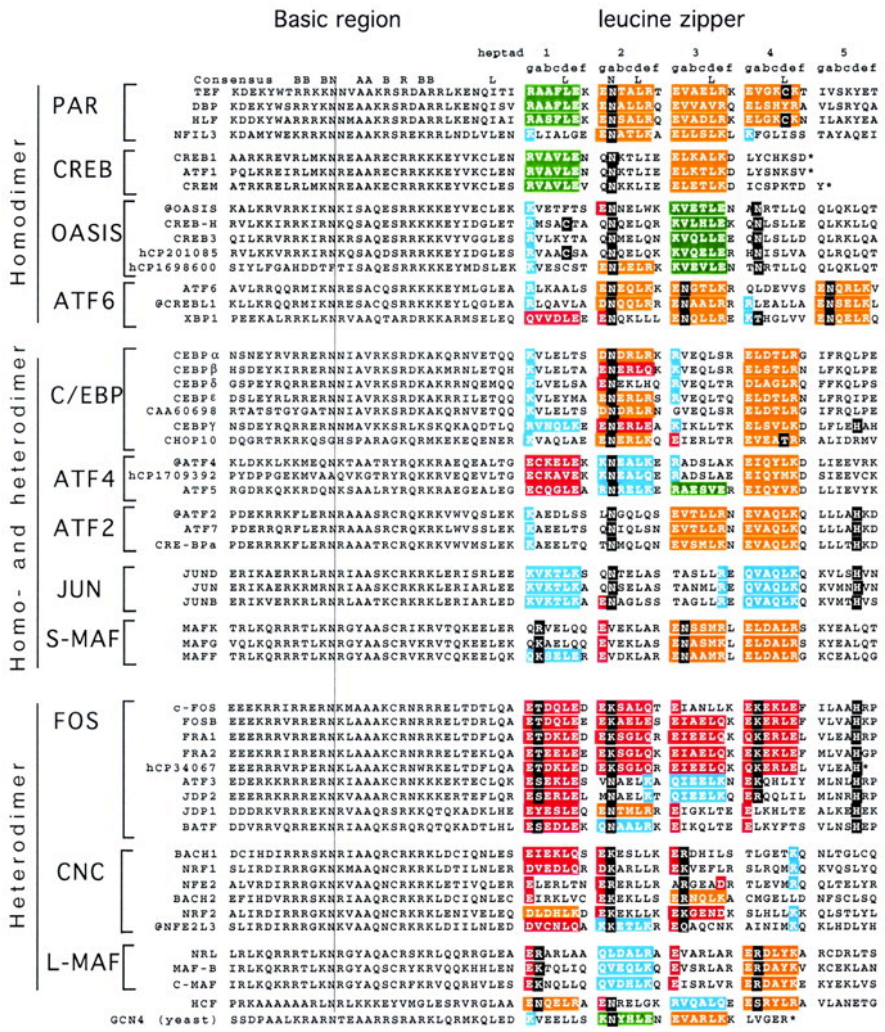


Figure 1.9 | Comparison of the bZIP domain between human bZIP transcription factors. Leucine zipper heptads are denoted by **gabcdef**. **gabcde** residues are coloured if a charged residue is found in both **g** and **e** positions. The four colours denote the four combinations of basic and acidic residues between the g and e positions. Figure from Vinson et al.[47]

The crystal structure of CREB reveals a hexahydrated Mg^{2+} ion sandwiched between the two basic region monomers (Figure 1.4)[42]. The role of the Mg^{2+} ion has been attributed to modulating CREB's DNA binding specificity[91,92], however with conflicting results in literature, it is unclear how the Mg^{2+} ion affects nonspecific DNA binding. One study found that it has no effect in binding to other sequences[91] whereas another study has shown that the presence of Mg^{2+} abolishes binding to nonspecific sequences[92]. The presence of 5 mM $MgCl_2$ was also found to decrease the autoinhibitory effect of phosphorylating serine residues in the CREB KID[74]. The role of magnesium in CREB activity is yet to be wholly elucidated.

It is important to note that while the bZIP domain does dictate DNA-binding and homodimerisation affinity, the rest of the protein can also play a role in mediating these interactions. This is known as the allosteric

effect[93]. The CK cassette of S108, S111, S114, S117 allosterically inhibits CREB when phosphorylated through ionic interactions with the basic region[74]. While the role of the transactivating domains of CREB in mediating protein-protein interaction is known, their possible allosteric effect in mediating bZIP DNA-binding and homodimerization through ionic interactions is to be determined.

1.6 CREB signalling pathways and physiological implications

CREB is involved in numerous cellular signalling pathways as one of the final effectors to modulate gene transcription. Most signalling pathways converge onto CREB by kinases which phosphorylate serine-133. These kinases include protein kinase A (PKA), protein kinase B (PKB), protein kinase C (PKC), protein kinase G (PKG), MEK/ERK, CaMKI/II/IV, Dyrk1A, LIM kinase I and MAPKAPK[64]. Upon signal transduction, the catalytic subunits are released from their regulatory subunit, allowing them to freely diffuse into the nucleus and phosphorylate CREB[68] (figure 1.10).

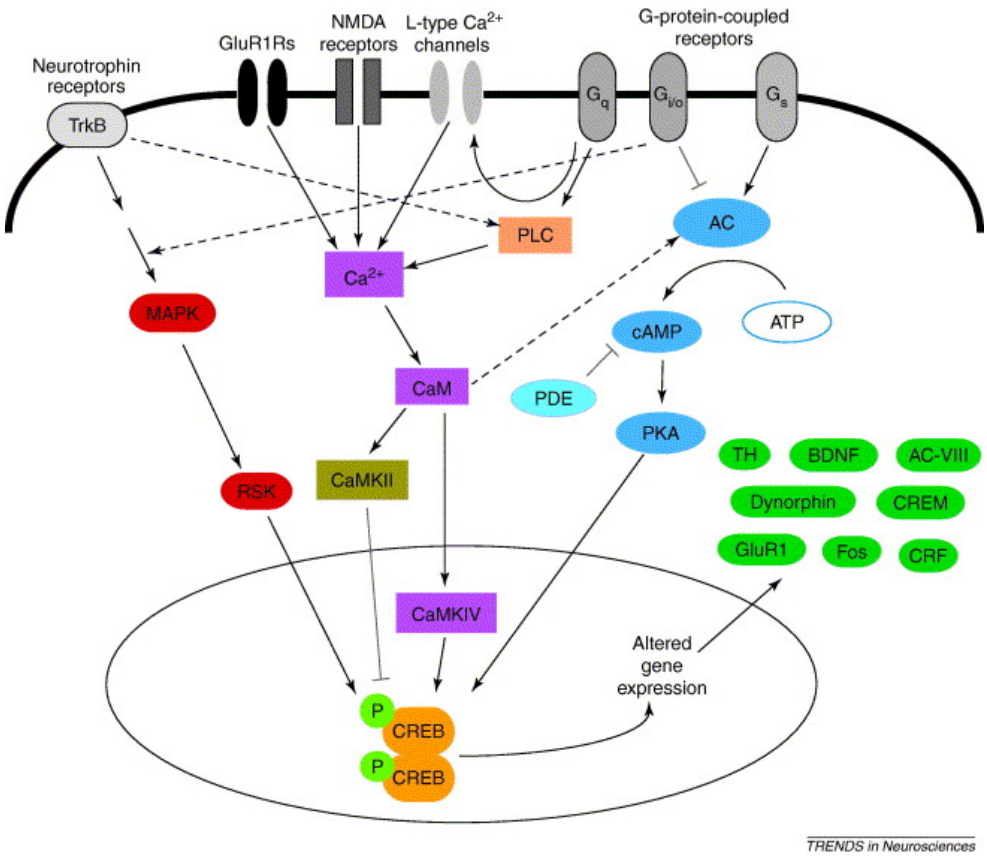


Figure 1.10 | Examples of signalling pathways that CREB participates in. Dashed lines represent simplified interactions between signalling pathways. Figure from Carlezon et al.[68]

Around 300 stimuli have already been shown to induce CREB phosphorylation[64] Combined with regulating over 4,000 genes[87] and an ubiquitous expression pattern[94], it is no surprise that CREB plays an incredibly important role in a wide range of cellular processes and is essential in survival[94]. Notably, CRE-sites are found enriched in genes involved with transcription factor activity, metabolism, cell cycle, secretory pathways, neurotransmitters, growth factors and maintaining neuronal development and plasticity[87,95]. As such, CREB has been identified as an important intermediate in mediating the coupling of neuronal activity to growth and cellular metabolism.

1.6.1 CREB and Memory

CREB plays a critical role in memory formation and maintaining neuronal synaptic plasticity. In a neuronal setting, CREB is essential as it acts as the converging point between cAMP and Ca^{2+} signalling cascades. These cascades can be activated by stress cytokine activity, growth factor/hormonal activity and phospholipid C activity[96]. cAMP and Ca^{2+} are particularly important in a synaptic junction as cAMP is one of the most commonly found secondary messenger, while Ca^{2+} acts as a prominent neurotransmitter between synapses[97]. Furthermore, CREB activity can be modulated by a wide range of receptors and neurotransmitters found in synaptic membranes[96].

Activation of CREB from synaptic activity results in the expression of downstream proteins, yet to be discovered, that strengthen the synaptic connection through the growth of new synapses between the two neurons[97]. This process, known as cellular consolidation, modulates the synaptic plasticity as it encodes the memory of the response triggered by the stimuli. This model explains how CREB can affect long-term memory as these new synaptic connections can last from hours to days and even years depending on the *de novo* gene expression[97].

CREB also affects cellular consolidation through regulating adult neurogenesis. Adult neurogenesis describes the formation of new neurons from neuronal stem cells and their incorporation into functional networks[95]. The formation of new network connections contributes to cellular consolidation. Phosphorylated CREB has

been found in immature neurons from important neurogenic niches[98], and CREB activation has been reported to result in postnatal neurogenesis[99]. Furthermore, studies have shown that knock out and mutated mice with inhibitory CREB are impaired in long term memory formation[100] and deficient in long-term potentiation[101].

CREB has also been shown to modulate the intrinsic plasticity of a synapse, affecting memory formation. Through affecting the properties of Na^+/K^+ ion channels[102], CREB can effectively altering the threshold of which the action potential is triggered, thus potentiating the depolarising and polarising currents. This response encodes the cellular memory to particular behavioural response, resulting in memory formation[97]. These evidences and models highlight the various ways that CREB can affect memory formation and accentuating its cellular importance.

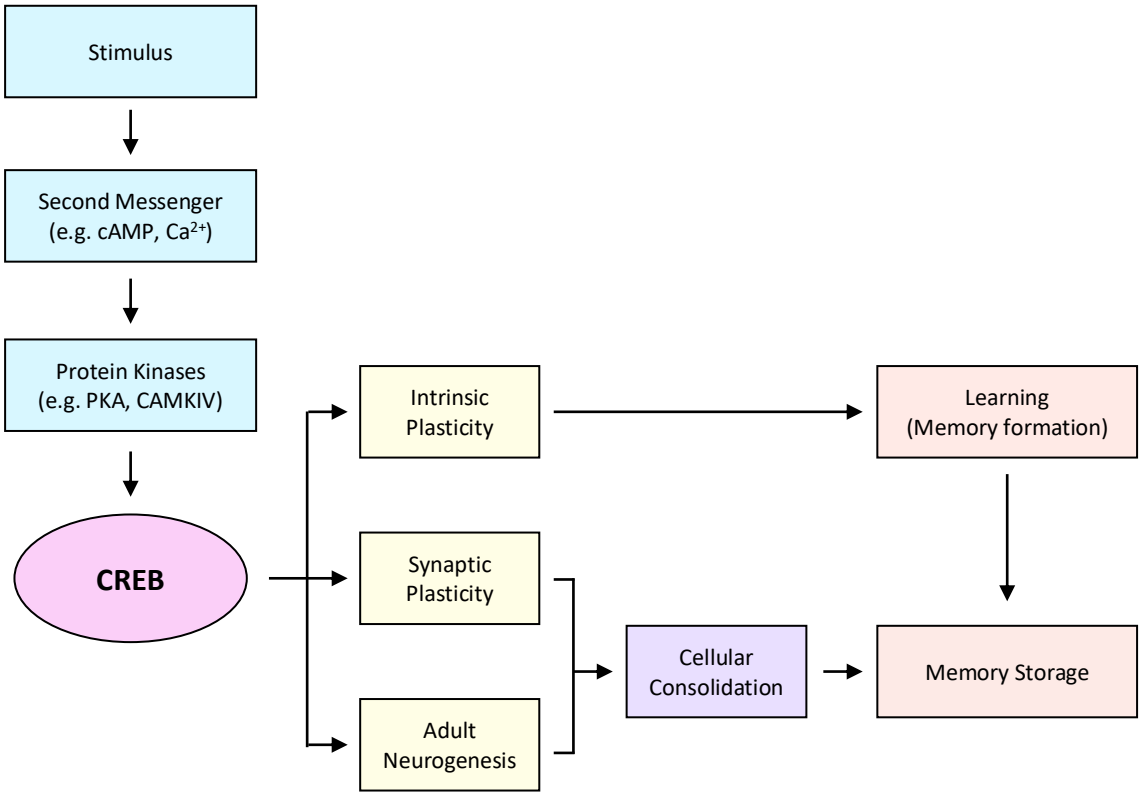


Figure 1.11 | Schematic of how CREB affects learning and memory. Upon neuronal stimulus, CREB is activated through signalling cascades. CREB then modulates the intrinsic and synaptic plasticity of the neuron, leading to cellular consolidation, memory formation and memory formation. Figure adapted from Benito & Barco[1].

1.6.2 CREB and Cancer

Frequent and persistent activation of CREB is sufficient to turn a normal cell into tumour cells[103]. From solid tumours such as non-small lung carcinoma[104], breast cancer[105] and melanoma[106], to hematopoietic cancer such as leukemia[1], a variety of cancer cells are found to overexpress CREB. The overexpression and prolonged activation of CREB were also attributed to enhanced cell proliferation, increased angiogenesis, increased metabolism, genome instability and rendering resistance to apoptosis[103,104,107]. In a clinical setting, CREB overexpression has been associated with a worse prognosis, reduced patient survival, increased tumour size, growth and greater chance of recurrences[103]. Interestingly, while high levels of CREB expression can be found corresponding to tumour cell lines, CREB mutations and deletions are rarely observed in tumours[108].

1.7 The CREB Regulon

The regulatory mechanisms of CREB can be extracellular or intracellular. Extracellular mechanisms involve signal transduction, where cytokines and growth factors can induce CREB activation, or changes in micro-environment, where CREB is activated through stress response or hypoxic environments[75,109]. Intracellular regulatory mechanisms include transcriptional regulation, post-translational modifications (PTM) or post transcriptional modifications[107]. The coupling of extracellular stimuli and intracellular responses form the CREB regulon. Considering the vast amount of effectors and downstream genes that is involved in CREB pathways, it is no surprise that CREB is a prominent oncoprotein when misregulated.

1.7.1 CREB regulation by Post-translational modifications (PTMs)

Phosphorylation

Phosphorylation is, amongst all the CREB PTMs, the most studied one. Phosphorylation at S133 activates CREB and allows it to recruit CBP and other transcription machinery. Inhibitory phosphorylation sites can be

found in proximal serines in the KID[74]. The effects of phosphorylation in the KID was discussed in section 1.4.2.

Ubiquitination and SUMOlyation

Ubiquitination mediates the proteasomal degradation pathway and is particularly important for proto-oncogenic transcription factors like CREB[107]. Prolonged serine-133 phosphorylation of CREB has been linked with oncogenic activity, shedding light on the importance to regulate the half-life of key transcription factors[103]. The ubiquitination of CREB was shown to be activated by hypoxia and the presence of reactive oxygen species[110–112]. While the exact ubiquitination site is unknown, CREB contains a proteasomal degradation sequence (DSVTDS, residues 115-121) that was shown to be targeted by the ubiquitin and proteasome pathway in a phosphorylation-dependent manner[113].

On the other hand, under prolonged hypoxic environment, CREB was found to be SUMOlyted in K285 and K304. The introduction of a SUMO (small ubiquitin-like modifier) stabilises CREB[112]. A K304R mutation to eliminate the SUMO site while maintaining the nuclear localisation sequence affected the sublocalization of CREB, implicating its role in CREB subcellular localisation regulation[64,110].

Acetylation

The transcriptional activity of CREB can also be tuned by acetylation. Acetylation by CBP/p300 has been correlated with an increased DNA-binding and transcriptional activity[114,115]. Acetylation of K136 was shown to be triggered by a low glucose and calorie restricted environment. This increases CREB activity, allowing glucogenic genes to be activated[116].

1.7.2 Post-transcriptional regulation of CREB

MicroRNAs (miR/miRNA) are small double-stranded RNA molecules that can induce gene silencing. They bind to the 3'-UTR (untranslated regions) of mRNA transcripts, targeting them for cleavage or translational

repression[117]. Interestingly, the 3'-UTR of the CREB mRNA is over 10 times longer than the average transcript (~9,000 nucleotides to ~800 nucleotides), enabling extensive regulation in this region[107]. 10 CREB-regulating miRNAs were found in tumour cell lines[107]. Overexpression of these miRNAs, such as miR-200b, miR-301 and miR-343 have been attributed to altered tumour growth properties through miRNA-mediated degradation of CREB[103]. These miRNA were also shown to reduce tumorigensis *in vivo*[118].

For example, CREB and miRNA-9 are highly expressed in glioma cells[108]. CREB plays a pro-proliferative and anti-migratory role in the tumour cell, whereas miR-9 plays the opposite by targeting CREB mRNA for degradation. This creates a negative feedback loop as migration and proliferation are mutually exclusive[108]. This loop mechanically represents the migration-proliferation dichotomy commonly found in glioma cells, showcasing the close relationship between CREB and regulating microRNAs. Given the anti-cancer nature of regulatory miRNAs, a better understanding of the relationship between CREB, microRNA and other forms of regulation could bring insight to novel therapeutics in tackling various forms of CREB-related cancer.

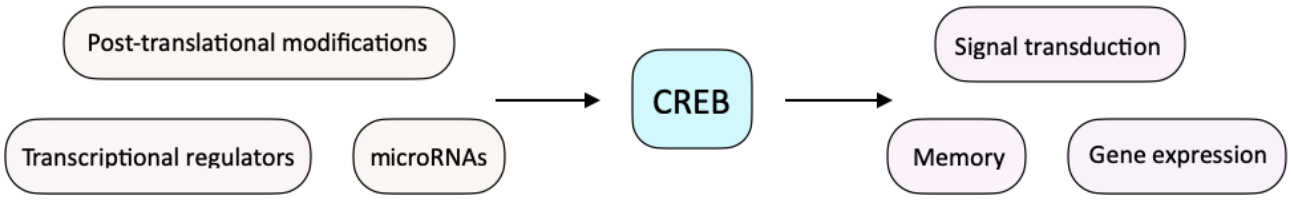


Figure 1.12 | Schematic of the CREB regulon. Post-translational modifications, transcriptional regulators and miRNAs regulate CREB's functionality, modulating signal transduction, gene expression and memory formation.

1.8 Methods for studying IDRs

The disordered nature of IDPs presents a great challenge in studying them at an atomic scale. Structural techniques such as x-ray crystallography and cryo-electron microscopy require structural homogeneity to obtain high resolution data[119]. NMR spectroscopy and molecular dynamics are the two structural techniques suitable for analysing disordered proteins or domains, but they are not without their limitations.

Transient secondary structures[120] and long range contacts[121] from NMR can give insight into dynamics of the disordered region, but ensemble averaged data alone is generally insufficient to assemble a high-resolution conformational ensemble[122]. Alternatively, molecular dynamics can be used to simulate protein dynamics and structure to a femtosecond time scale[123], but the trajectories generated varies greatly on the choice of forcefield and solvent model[124]. The dependence on a suitable forcefield and water model is exemplified in simulations of disordered region due to significant increase in protein-solvent interactions arising from the disordered nature of domains[125,126]. Furthermore, the required sampling time for examining binding reactions of disordered proteins represents a significant challenge.

Non-structural approaches to investigate the interactions of intrinsically disordered proteins include biophysical and kinetic assays. Techniques such as isothermal calorimetry (ITC), fluorescence anisotropy, and fluorescence stopped-flow kinetics (SF) are applied to estimate binding affinities for IDP interactions.

1.9 Protein Kinetic and Biophysical Studies

ITC measures the binding energy associated with a reaction by titrating a small amount of one component into a temperature-controlled cell containing the other reactant[127]. By extracting the area under the titration curve, the technique can directly probe the binding enthalpy of the interaction (ΔH) and the equilibrium association rate constant (K_a)[128]. However, the upper limit of ITC lies at around 10^8 to 10^9 M^{-1} for the binding constant (10^{-8} to 10^{-9} M for the dissociation constant). Beyond its upper limit, the ITC titrations lose their curvature and the data cannot be reliably fitted to an equation[127].

Fluorescence anisotropy (FA) is another biophysical method to probe the changes that occur in a system upon a binding event and relies on changes in the polarization of fluorescent light during the timescale of fluorescence emission[129]. The anisotropy of the sample is mostly affected by the size of the molecule the probe is attached to, but is also modulated by the shape of the probe. Importantly, when a protein is bound to the fluorescent probe, it will increase the molecular size of the probe and alter its shape and tumbling rate. In turn, the anisotropy will increase upon a binding event[130]. By the same principle, when the probe is

excited with a polarised light source, under a fast tumbling regime, the sample will emit a depolarised light. Upon a binding event, the tumbling is slowed and the light emitted is more polarised (Figure 1.13)[129]. By measuring the polarisation of emission in both vertical and horizontal manner, the anisotropy of the sample can be determined. The equilibrium dissociation constant (K_d) of the binding event can then be extracted by measuring the anisotropy with various concentrations of a binding partner and fitting the data to a curve[131].

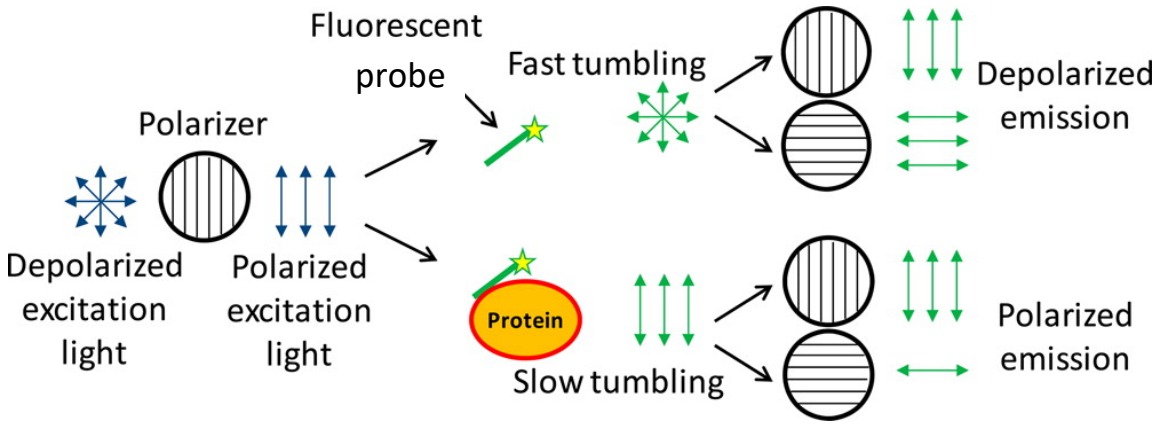


Figure 1.13 | Measuring fluorescent anisotropy of a sample by polarised excitation. A free fluorescent probe tumble fast in solution, and therefore depolarises the excitation light upon emission. When a protein is bound to the probe, the speed of the tumbling decreases. This causes inefficient depolarisation of the light source, and hence the emission is more polarised than that of the free probe. Figure from Cheow et al.[129]

The limit of which binding affinities can be determined by FA depends on the sensitivity of the machine and the brightness of the fluorophore. As the anisotropy is a ratio of intensity measurements, the signal-to-noise ratio decreases along with the probe concentration[131]. Therefore, a lower bound of the binding affinity is determined by the concentration of the fluorescent probe. The lowest probe concentration that can be measured is dependent on the sensitivity of the machine and the size of the change in anisotropy. There is always a possibility that fluorescently labelling alters the 'true' interaction between the two molecules, particularly for smaller ligands such as DNA and RNA, as the dyes would be comparatively bigger. Furthermore, fluorescent dyes are often charged[132], and will modulate the ionic interaction between the two reactants. Again, its effect is more prominent in interactions that rely heavily in ionic contacts, such as protein-DNA interactions.

FA and other 'mix and observe' biophysical methods typically have a dead time of 5-10 seconds between samples for experimental setup and efficient mixing which limits their capacity for monitoring the binding reaction itself[131]. Fluorescence stopped-flow (SF) kinetics involve rapid mixing of the two solutions and do not suffer from this limitation. The two solutions in the reaction cell are mixed rapidly, with the mixing event triggering the detector to start measuring the signal through means of a stop syringe (Figure 1.14)[133]. This method of rapid mixing has a dead time of 1-10ms and is suitable for recording up to around 500 seconds[130]. This makes SF an excellent technique to probe pre-equilibrium kinetics of reaction between these two timescales.

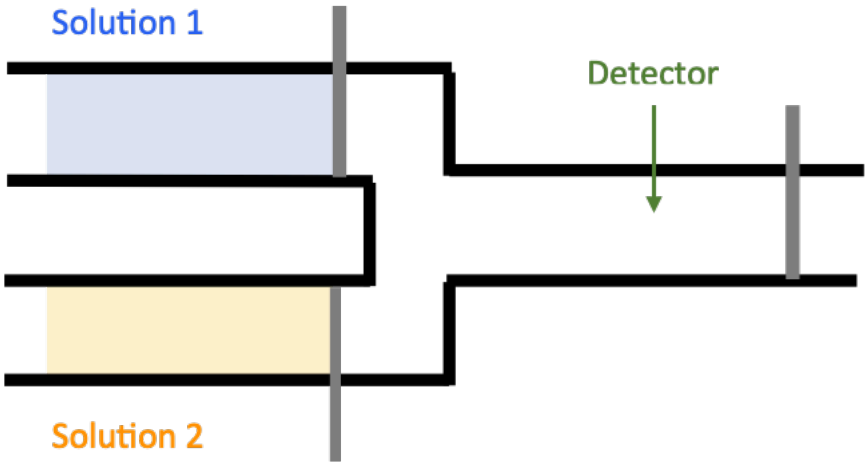


Figure 1.14 | Simplified schematic of a stopped-flow machine. Solutions 1 and 2 are kept in separate reaction cells. The valves for the two chambers are connected by a stop syringe. The two solutions are then mixed in the mixing chamber, of which the detector detects the signal change over a designated period of time.

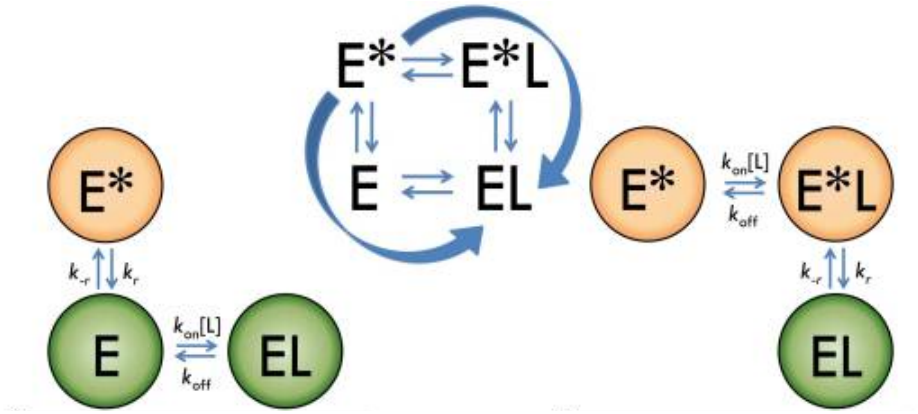
The stopped-flow detector can detect absorbance, but fluorescence is the most commonly used output. The fluorescence can originate from a fluorescent dye, or the intrinsic tryptophan/tyrosine fluorescence of a protein. Various kinetic parameters can be extracted depending on the experimental setup. For example, the association rate constant (k_{on}) and the dissociation rate constant (k_{off}) can be probed with a pseudo-first-order mixing experiment and a competition mixing experiment respectively[131]. If the reaction is known to be two-state, the K_d can then estimated by k_{off}/k_{on}[130].

These varying techniques can be used in complementary of one another to obtain a holistic understanding of an interaction. A ϕ-value analysis, based upon several conservative point mutants, can be performed to probe the transition state of the two-state reaction[134]. Through understanding the kinetic parameters of the protein-protein or protein-nucleic acid interaction, their interaction mechanism can be elucidated. The addition of structural studies via NMR or molecular dynamics, or even circular dichroism (CD) spectroscopy can then be employed to supplement the kinetic data, providing additional information on a residual level.

1.9 Mechanisms of Protein Interaction

The binding rates of interactions can vary greatly depending on macromolecular partners, the type of process involved, the specificity of the reaction and the interaction mechanism involved, particularly in ID proteins or regions that fold upon binding to its partner. Some IDPs such as bZIP proteins wholly transition into an ordered state, whereas some others retain their disordered nature, forming a fuzzy complex[28].

The two main models to describe protein interactions with another species are induced fit and conformational selection. The induced fit mechanism entails the protein to recognise its binding partner to initialise the interaction, upon which the protein will fold into a conformation that would be most energetically favourable for the interaction. Opposingly, conformational selection describes interactions that favour molecules that already adopt the binding conformation in their native state (Figure 1.15)[135].



Conformational Selection Induced Fit

Figure 1.15 | Comparison between the conformation selection and induced fit mechanism of recognition. E represents enzyme while L represents ligand. In both schemes, the equilibrium between E and E* represents the forms that the enzyme take in solution. Only E can interact with the ligand L. In the conformational selection scheme, only the protein in the E conformation in solution can bind the ligand. Alternatively, in the induced fit model, any conformation of E or E* can bind the ligand. Upon binding, the conformation of E* will change to fit the ligand in the EL complex. Figure from Vogt & Di Cera[135].

Of course, while some protein interactions strictly follow either binding mechanisms, many fall between the spectrum. Understanding the interaction mechanism and the rate for each reaction can allow more sophisticated models to be developed. In the case of IDPs, an accurate interaction network can be mapped between the various protein partners. With the knowledge of the binding pathway and the equilibrium between each state, protein engineering could modulate the interactions such that one pathway or state is preferred. This could deter proteins from being in a misfolded or disease-causing state, such as the Alzheimer's causing fibrillation of amyloid beta[136]. Similarly, antagonising small molecules can be designed against a certain pathway or stabilise certain transition states for therapeutic purposes.

Aims of the Project

Despite being one of the most well-studied bZIP in signal transduction pathways, the molecular mechanism by which CREB binds to DNA is yet to be uncovered. It is known that dimerisation, DNA binding and phosphorylation are essential in activating downstream gene transcription, but in which order do they occur? Does CREB bind specific DNA sequences as a monomer, or only as a dimer? Does CREB follow the monomer or the dimer pathway? There are many questions to be answered regarding the binding mode of CREB to itself and to DNA.

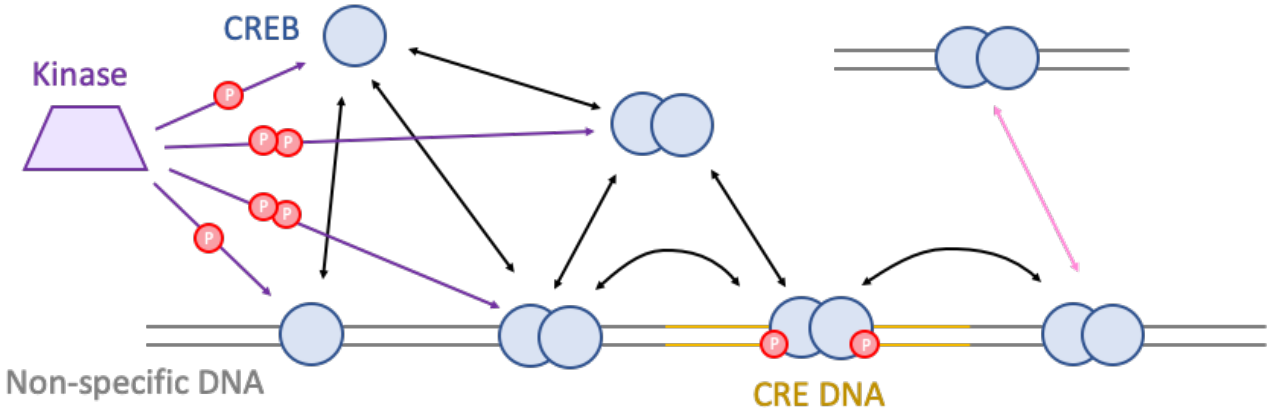


Figure 1.16 | Simplified schematic of the events during CREB binding to target CRE DNA. Light blue circle represents CREB monomers. Purple trapezoid represents the activating kinase, and red circles for phosphates. Grey represents nonspecific DNA and the yellow represents the CRE recognition sequence (TGACGTCA). Black arrows represent the possible pathways that CREB can take to reach the final activating state. Purple arrows represent possible phosphorylation pathways. The pink pathway shows intersegmental transfer from a section of nonspecific DNA to another. Currently, the rates of which these interactions occur is still yet to be determined.

The Shammas lab is currently investigating the DNA binding pathways of CREB using structural and biophysical analysis through a construct containing only the bZIP domain. Moreover, the lab sets to uncover the contribution of key residues and elucidate the role of disorder in DNA binding.

The aim of this project is to contextualise the findings on the bZIP domain by expanding the kinetic studies to the full length CREB. This will require the development of an expression and purification protocol to obtain

pure samples suitable for detailed biophysical studies, and adaptation of existing biophysical assays to suite the new construct.

The first objective of this study is to purify HaloCREB that is suitable for biophysical analysis. This includes construction of an expression plasmid and optimisation of the protocol to remove all nucleic acid contamination. Additionally, mutant versions of HaloCREB are generated and purified to shed light on its binding pathways. Biophysical assays are applied to characterise these various HaloCREB constructs.

The second objective is to characterize the purified full length CREB using biophysical techniques to allow comparison of parameters with the bZIP domain construct. Fluorescence anisotropy was used to measure the equilibrium dissociation constant (K_d) to CRE and nonCRE DNA; and stopped-flow spectroscopy to investigate the kinetics parameters for HaloCREB binding to CRE DNA (k_{on} and k_{off}). Lastly, the homodimerisation of HaloCREB was examined by fluorescent anisotropy, native mass spectrometry and mass photometry.

2. Materials and Methods

Unless specified otherwise, all experiments were conducted at room temperature. Unless specified otherwise, all chemicals were purchased from Sigma-Aldrich. All instances of the use of H_2O or water refers to the MilliQ water from Millipore Milli-Q system. All solutions were made up with the same MilliQ water above. All culture media and agar plates were prepared by the Department of Biochemistry Media Glassware Sterilization Lab.

2.1 Microbiology

Both strains used in the experiments are of *Escherichia coli* derivatives. The NEB 5α is a derivative of the DH5α strain while Rosetta2(DE3) strain is a derivative of the BL21 strain. All culture growths were done in a shaking incubator at 37°C and 180 rpm unless stated otherwise.

Strain	Genotype	Source
NEB 5α	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	NEB
Rosetta2(DE3)	F^- *ompT hsd*S_B(r_B^- m_B^-) *gal dcm* (DE3) pRARE2 (Cam^R)	Sigma-Aldrich

Table 2.i | List of bacterial strains used in this study.

2.2 Molecular biology

2.2.1 Plasmids and primers used in this study

Plasmid	Properties	Source
pET303/CT-His	Contains a C-terminal His6 tag before stop codon.	Thermofisher (#K630203)
pFN21A HaloTag CMV Flexi	Contains the HaloTag to be cloned	Promega (#G2821)
pGEM-CREB1	Full length human ORF of CREB1 (NCBI: NM_004379.2)	Sino Biological (#HG10074-G)
pET303-HT-CREB-His6	Expression construct for HaloCREB. T7 inducible (IPTG), Amp^R	This work; PL103
pET303-HT-CREB-His6 C286S C296S	Cysteine to serine mutations in position 286 and 296 of CREB. Single cysteine mutant (C323)	This work; PL145
pET303-HT-CREB-His6 C286S C296S C323S GGC_{ins}	Cysteine to serine mutations in position 286, 296 and 323 of CREB. GGC insert between CREB and His6.	This work; PL146
pET303-HT-CREB-His6 Xa_{ins}	Factor Xa cleavage site (IEGR) insert between CREB and His6	This work; PL147

Table 2.ii | List of plasmids used in this study.

Template	Primer	Primer Sequence	T_m (°C)
pET303/CT-His	pET303_F	5'-tttctgccatTTCTAGACCTCCTTCTTAAAGTTAAACAAAATTATTACTAGAGG-3'	68°C
	pET303_R	5'-caaatcagatCACCACCACCACCACCACT-3'	66°C
pFN21A HaloTag CMV Flexi	FL_HT_F	5'-aggtctagaaATGGCAGAAATCGGTACTGGC-3'	66°C
	FL_HT_R	5'-ccatggtCATGGCGATCGCGTTATCG-3'	66°C
pGEM-CREB1	FL_CREB_F	5'-cgcgatcgccATGACCATGGAATCTGGAGCC-3'	65°C
	FL_CREB_R	5'-ggtggtggtgATCTGATTTGTGGCAGTAAAGGTCC-3'	65°C

Table 2.iii | List of amplification primers used in this study. Upper case bases denote annealing bases with the template. Lower case bases denote non-annealing bases.

2.2.2 HiFi Plasmid Assembly

Gibson fragments of the inserts and vector with directional overlaps were amplified with the primers outlined in table 2.iii using the NEB Q5 HotStart High-Fidelity Polymerase kit (#M0493). The respective forward and reverse primers for each reaction were pre-mixed to a working concentration of 5 µM. The reactions were performed both with and without GC enhancer buffer as outlined in table 2.iv. The thermocycling conditions for can be found in table 2.v. Touchdown PCR was performed to amplify the pET303 backbone. PCR products were then analysed on a 0.5% agarose gel.

Reagents	Volume (µL)	
	A	B
5X Q5 Buffer	10	10
10mM dNTPs	1	1
5 µM Combined Primers	5	5
Template DNA	1	1
5X GC Enhancer Buffer	/	10
10X Q5 Polymerase	0.5	0.5
H_2O	32.5	22.5

Table 2.iv | Reagent list for amplification of Gibson fragments. Template DNA concentration is 1.69ng/µL for pET303 and 1ng/µL for pGEM-CREB1 and pFN21A HaloTag.

98°C	30 sec	
98°C T_m 72°C	10 sec 30 sec 2 min 45 sec	30 cycles
72°C	2 min	
10°C	Hold	

Table 2.v | Thermocycling condition for amplifying Gibson fragments to assemble HaloTag-CREB-His6. T_m for pGEM-CREB1 and pFN21A HaloTag are respectively 65°C and 66°C. In the touchdown PCR protocol for amplifying pET303, T_m was gradually reduced from 71°C to 62°C over the course of the first 10 cycles.

The PCR products in 0.5% agarose were cut out and purified using the Thermo Scientific GeneJET DNA cleanup micro kit (#K0832), following the protocol outlined in the kit. The concentration of DNA was then measured with the Thermo Scientific NanoDrop 2000.

The NEB HiFi DNA assembly master mix (#E2621) was used to assemble the plasmid. 0.1 pmol of each fragment and 0.05 pmol of the pET303 backbone were used to perform the reaction. 10µL of the master mix was combined with the corresponding amount of each fragment, and H_2O was added to the final volume of 20 µL. The reaction was then incubated at 50°C for 15 minutes before transforming into *E. coli* NEB 5α cells by the protocol outlined in section 2.2.4.

2.2.3 Generating competent NEB 5α and Rosetta2(DE3) cells

5mL overnight cultures of the both strains were grown with the respective commercial stock using 2xTY media. The Rosetta2(DE3) starter culture was supplemented with 34 µg/mL chloramphenicol. A 100 mL culture with the same media composition was inoculated with 1 mL of the overnight culture and grown to OD_{600} of 0.6 at 37°C. The cultures were then incubated on ice for at least 20 minutes. The cells were then harvested by centrifugation at 4,250 x g and 4°C for 10 minutes.

The supernatant was discarded and drained for 2 minutes by inverting the centrifugation tube. Cells were then resuspended in 1/5th of the original volume with [sterile 100 mM $CaCl_2$, 15% glycerol], and was incubated on ice for 20 minutes. The cells were spun down again at 4,250 x g and 4°C for 10 minutes, discarding and draining the supernatant in the same manner. The cells were then resuspended in 1/25th of the original volume with [sterile 100 mM $CaCl_2$, 15% glycerol]. The cells were then aliquoted in 50 µL aliquots before flash-freezing by liquid nitrogen and stored at -80°C for future use.

2.2.4 Transformation of PL103 into NEB 5α

An aliquot of chemically competent *E. coli* NEB 5α was defrosted on ice for 5 minutes before adding 5 µL of the assembly product from 2.2.2. The mixture was then incubated on ice for 30 minutes before a 30 seconds heat shock at 42°C in a water bath. The cells were then returned on ice for 5 minutes before plating on agar plates with 100 µg/mL ampicillin and incubated overnight at 37°C.

Colonies of transformed *E coli.* NEB 5α cells were then inoculated in a 2 mL 2xTY starter culture supplemented with 100µg/mL ampicillin. The cultures were grown at 37°C for 6 hours. The plasmid from each culture were isolated with the QIAprep Spin Miniprep Kit (#27104) following standard protocol and eluding in 20 µL H_2O. The plasmids were checked with 0.5% agarose gel electrophoresis before being sent off for DNA sequencing to Source Bioscience using the sequencing primers in table 2.vi for verification.

Template	Sequencing Primer	Sequencing Primer Sequence
All plasmids	T7_Forward	5'-TAATACGACTCACTATAGGG-3'
	T7_Reverse	5'-GCTAGTTATTGCTCAGCGG-3'

Table 2.vi | Sequencing primers for all sequencing results.

2.2.5 Site-Directed Mutagenesis of PL103

Site-directed mutagenesis was performed on PL103 using the primers outlined in table 2.vii below.

Template	Primer	Primer Sequence	Annealing T_m Q5 (°C)
pET303-HT-CREB-His6	C286S_C296S_F	5'-AATATGTGAAATCTTTAGAAAACAGAGTGGC-3'	65°C
	C286S_C296S_R	5'-CTTTCTTCTTTCTACGAGACTCTCGAG-3'	65°C
	Xa_F	5'-ggccgcCACCACCACCACCACCACTGAGATC-3'	68°C
	Xa_R	5'-ttcaatATCTGATTTGTGGCAGTAAAGGTCCTTAAGTGC-3'	66°C
pET303-HT-CREB-His6 C286S C296S	C323S_GGC$_{ins}$_F	5'-ctgcCACCACCACCACCACCACTGAGATC-3'	66°C
	C323S_GGC$_{ins}$_R	5'-ccgccATCTGATTTGTGGGAGTAAAGGTCCTTAAGTGC-3'	66°C

Table 2.vii | List of site directed mutagenesis primers used in this study. Upper case bases denote annealing bases. Lower case bases denote non-annealing bases. Bold and underlined denote mutation or insertion to be introduced.

Reagents	Volume (µL)
PCR Product	1
2X KLD Reaction Buffer	5
10X KLD Enzyme Mix	1
H_2O	3

Table 2.viii | List of reagents for circularising the PCR product.

The procedure was carried out using the NEB Q5 site-directed mutagenesis kit (#E05544S) following the manufacturer's instructions. The reagents are the same as outlined in table 2.iv, but with each respective pair of primers and 1 ng/µL of template DNA. The reactions were then performed cycling condition outlined in table 3.v but with their respective annealing temperatures. The PCR products were then circularised by assembling the reagents described in table 2.viii and incubated at room temperature for at least 5 minutes. 5 µL of the reaction mix was then used to transform into *E. coli* NEB 5α using the protocol outlined in section 2.2.4. The plasmids were subsequently isolated and sent for sequencing with the same primers in table 2.vi.

2.3 Protein Expression

2.3.1 SDS-PAGE Analysis

For each sample, 10 µL of SDS sample buffer [NuPAGE LDS sample buffer (#NP0007) supplemented with 5 mM Bond-Breaker TCEP (#77720)] was added to 30 µL of sample in an Eppendorf tube. The samples were then boiled on a heat block at 100°C for 5-10 minutes. 7 µL of a suitable protein ladder was loaded onto the first well of a NuPAGE 4-12% bis Tris gel (#NP0323). Subsequently, 10 µL of each sample was then loaded onto the following wells. The loading dock was then filled with either 1X MES (#NP0002) or 1X MOPS (#NP0001) running buffer. The gel was run at 200 mV for 45 minutes before staining with InsantBlue coomasie stain (#ISB1L). Incubate at room temperature for at least 30 minutes on a gel rocker before imaging.

2.3.2 HaloCREB Expression Test in Escherichia coli Rosetta2(DE3)

PL103 was first transformed into Rosetta2 cells and plated onto agar plates as per the same protocol described in 2.2.4. A 10mL starter culture of 2xTY supplemented with 100 µg/mL ampicillin and 34 µg/mL chloramphenicol was grown overnight using colonies from the transformation plate. 500 µL of the overnight culture (after incubation) was combined with 500 µL of sterile 50% glycerol solution and stored at -80°C as a glycerol stock.

Two 100 mL cultures were grown with the same media composition with a 1 in 10 inoculation from the starter culture. Upon reaching OD_{600} of 0.6, both cultures were induced by 1 mM isopropylthio-β-galactoside (IPTG). One culture was returned to the 37°C shaking incubator, and a sample was taken hourly until 6 hours upon induction. The second culture was placed on ice for 20 minutes and incubated at 18°C for 21 hours. A sample of the 18°C expression was taken at 6 hours mark. All samples were then sonicated for 15 seconds. The tubes were then centrifuged at 15,000 x g for 10 minutes. The supernatant was then collected for SDS-PAGE analysis.

2.3.3 Large Scale Expression of HaloCREB in E. coli Rosetta2(DE3)

A 100mL starter culture of 2xTY supplemented with 100 µg/mL ampicillin and 34 µg/mL chloramphenicol was grown overnight from the glycerol stock. A 1 in 10 inoculation of the starter culture in 1 L of the same expression media is grown to OD_{600} of 0.6 in a 37°C shaking incubator at 180 rpm. The cultures were then induced with 1 mM isopropylthio-β-galactoside (IPTG) and grown for at least 4 hours. Cells were harvested by centrifuging at 6,250 x g and 4°C for 30 minutes. Cell pellets were then stored at -80°C until further purification.

2.4 Protein Purification

2.4.1 Buffers used in this study

Resuspension buffer*	100 mM Tris-HCl (pH 8.5), 50 mM NaCl
Shnitkind HisTrap buffer	20 mM Tris-HCl (pH 7.5), 1 M NaCl
Shnitkind cation buffer	20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 2 mM DTT
SEC buffer	100 mM Tris-HCl (pH 8.5/11), 50 mM NaCl, 2 mM DTT
IEC buffer	20 mM bis-tris (pH 6.5/10), 2 mM DTT
Heparin buffer*	10 mM Na_2PO_4 (pH 7), 2 mM DTT
HaloLink buffer	50 mM HEPES (pH 7.5), 150 mM NaCl, 1 mM DTT
Carboxypeptidase A buffer	50 mM Tris-HCl (pH 7.5), 500 mM NaCl
Forced dimerisation buffer	100 mM Tris-HCl (pH 9.2)
Biophysical buffer*‡	10 mM MES, 150 mM NaCl, 10 mM $MgCl_2$, 0.05% TWEEN, 1 mM DTT

Table 2.ix | List of buffers used for protein purification in this study. *Buffers for the final purification protocol. DTT is only used when purifying the WT and C286S_C296S construct but must be avoided when purifying the forced dimer construct. ‡Must be made up volumetrically. DTT is only added before use.

2.4.2 Pre-Purification Preparation Steps

Pellets from the expression protocol was defrosted at room temperature if necessary, then resuspended in 25 mL of resuspension buffer [100 mM Tris-HCl (pH 8.5), 50 mM NaCl] and supplemented with Pierce EDTA-free protease inhibitor per pellet from 1 L of culture. Cells were lysed on ice for 15 minutes by sonication. The lysate was then spun down at 48,400 x g at 4°C for 45 minutes. The supernatant was then collected and filtered through a 0.45 µm and subsequently a 0.22 µm syringe filter (Millipore, #SLGP033RS & #SLGP033RS).

2.4.3 Temperature Test

500 µL of cell lysate were transferred to six Eppendorf tubes. Using a thermoshaker, five of the Eppendorf tubes were sequentially incubated 5 minutes at 55°C, 60°C, 65°C, 70°C and 75°C. The last sample was kept at room temperature as a control. The six tubes were then centrifuged for 15 minutes at 16,800 x g. Insoluble fractions were discarded, and the supernatant was diluted 10-fold prior to analysis with SDS-PAGE.

2.4.4 Shnitkind et al. Purification Protocol

The purification protocol from Shnitkind et al.[74] was tested for HaloCREB. The cell lysate was prepared as described in 2.4.2, but were resuspended in [20 mM Tris-HCl (pH 7.5), 1 M NaCl supplemented with Pierce EDTA-free protease inhibitor] instead.

A 5 mL HisTrap High Performance (#GE17-5248) column was equilibrated at 5 mL/min with 10 column volumes of buffer, followed by 10 column volumes of [buffer with 2M imidazole], and 10 more column volumes of buffer prior to loading. Cell lysate was loaded onto the column at 1.5 mL/min using a peristaltic pump. The column was then washed with 10 column volumes of buffer, followed by an overnight wash of [buffer with 2 M NaCl] at 1 mL/min to reduce nucleic acid contamination. The column was then eluted via gradient elution with [buffer with 2 M imidazole] using an AKTA FPLC. Fractions were checked via SDS-PAGE and ones that contain HaloCREB were collected. The eluate was diluted 3-fold and dialysed overnight against 2 L of [20 mM Tris (pH 7.5), 100 mM NaCl and 2 mM DTT]. The A_{260}/A_{280} of the eluate was measured.

A 5 mL HiTrap SP High Performance (#17-1152-01) column was equilibrated in the same manner as the Ni-affinity column, except with the dialysis buffer and [dialysis buffer with 2 M NaCl]. The column was loaded with the dialysis product in the same manner as the Ni-affinity column. The column was washed with 10 column volumes of dialysis buffer before eluted via gradient elution with 2 M NaCl using an AKTA FPLC. Fractions that correspond to a peak in the elution profile were checked via SDS-PAGE.

2.4.5 Ni-Affinity Chromatography

The purification of HaloCREB and its mutants with Ni-affinity column chromatography is as described above in 2.4.3, but instead with the use of resuspension buffer [100 mM Tris-HCl (pH 8.5), 50 mM NaCl] supplemented with the same components for the respective steps.

WT HaloCREB and C286S_C296S mutant were incubated with 10mM DTT at room temperature for at least 1 hour upon elution before dialysis into the next working buffer.

2.4.6 Size Exclusion Chromatography (SEC)

Size exclusion chromatography was tested with the Superdex 200 5/150 GL (#GE28-9065-61) analytical column. The purification protocol was performed with the AKTA FPLC at a flow rate of 0.3 mL/min. Prior to sample loading, the column was cleaned with 2 column volumes of 1 M NaOH and H_2O. The column was then equilibrated with 10 column volumes of [resuspension buffer with 2 mM DTT]. The eluate was not dialysed before loading onto the size exclusion column.

Using a 100 µL sample loop, 50 µL of the Ni-affinity purification eluate was loaded onto the column. The column was then eluted with 4 mL of [resuspension buffer with 2 mM DTT], collecting 100 µL fractions in Eppendorf tubes. Fractions corresponding to peaks in the absorbance chromatogram were collected and checked via SDS-PAGE. The same test was repeated using [resuspension buffer (pH 11)].

2.4.7 Anion Exchange Chromatography

A 5 mL HiTrap Q High Performance column (#GE29-0513-25) was used for this purification step. The reduced Ni-affinity column eluate was dialysed into 2 L of IEC buffer [20 mM bis-tris (pH 6.5), 2 mM DTT] overnight. The column was equilibrated, loaded and eluted the same method as was described for the cation exchange column in 2.4.4, but with the IEC buffer supplemented with the same components for the respective steps. The same test was repeated with a buffer of [20 mM bis-Tris (pH 10), 2 mM DTT]

2.4.8 Cation Exchange Chromatography

The same protocol that was used to purify HaloCREB as described above in 2.4.7 was used for cation exchange chromatography. The purification was performed with the eluate from anion exchange chromatography.

2.4.9 Heparin Column Chromatography

Upon reduction by 10 mM DTT incubation for 1 hour, Ni-affinity purification eluate was diluted 3-fold before dialysing into 2 L of heparin buffer [10 mM Na_2PO_4 (pH 7), 2 mM DTT] overnight. A 5 mL HiTrap Heparin High Performance column (#GE17-0407-01) was equilibrated at 3 mL/min with 10 column volumes of heparin buffer, 10 column volumes of [heparin buffer with 2 M NaCl] and 10 column volumes of heparin buffer prior to loading.

Dialysis product was loaded onto the column with a peristaltic pump at 1 mL/min. The column was washed with 10 column volumes of heparin buffer, then eluted via gradient elution with 2 M NaCl using an AKTA FPLC. Fractions were checked via SDS-PAGE, and the A_{260}/A_{280} ratio for each fraction was measured on the Varian Cary 50 UV-vis spectrophotometer with a Hellma quartz ultra micro absorption cuvette (#Z600199). Fractions that correspond to HaloCREB with an A_{280} of >0.1 and an A_{260}/A_{280} of <0.58 were collected.

2.4.10 HaloLink Column Purification

4 mL of the HaloLink resin slurry (#G1912) was transferred to a 15 mL falcon tube after thoroughly suspending the stock bottle. The resin was then centrifuged at 1,000 x g for 5 minutes at room temperature and the supernatant was discarded. 10 mL of the HaloLink buffer [50 mM HEPES (pH 7.5), 150 mM NaCl, 1 mM DTT] was then added to the resin, and the resin was thoroughly mixed. This process was then repeated twice to equilibrate the resin.

Upon removing the supernatant from the last equilibration, 5 mL of cell lysate was added the column and mixed thoroughly by inversion. The falcon tube was then placed in a tube rotor for 1 hour at room temperature for binding. The column was then centrifuged at 1,000 x g for 5 minutes. The supernatant was then collected as flowthrough.

To wash the column, 10 mL of the HaloLink buffer was added to the resin and mixed thoroughly. The column was then centrifuged at 1,000 x g for 5 minutes and the supernatant was discarded. This process was then repeated twice.

To elute the protein, 1 mL of a 3 µM TEV protease solution was added to the resin and mixed by pipetting several times. The column was then incubated at room temperature overnight on a tube rotor. The column was then centrifuged at 3,000 x g for 5 minutes and the supernatant was collected. A sample was taken after 1 hour of incubation for SDS-PAGE analysis.

1 mL of HaloLink buffer was then added to the column, mixed thoroughly, and centrifuged at 3,000 x g for 5 minutes. The supernatant was collected as wash. 1 mL of [HaloLink buffer with 2M NaCl] was then added to the column and collected in a similar fashion upon centrifugation. All samples were then analysed via SDS-PAGE.

2.4.11 TEV Cleavage and Anion Exchange Chromatography

Nucleic acid-free HaloCREB after heparin column was dialysed into 2 L of IEC buffer [20 mM bis-tris (pH 6.5), 2 mM DTT] overnight. TEV protease (#T4455) was added to the dialysis product at a molar ratio of 1:100. The mixture was incubated at 37°C for 1 hour.

The protocol for anion exchange chromatography follows that of described in 2.4.7. A_{260}/A_{280} was measured for fractions that correspond to CREB on SDS-PAGE. If the signal was too low, the fractions were first concentrated with a 10,000 MWCO centrifuge concentrator before remeasuring. Fractions with an A_{260}/A_{280} of <0.46 are collected.

2.4.12 TEV Cleavage and Cation Exchange Chromatography

Similarly, the same purity of CREB can be achieved by employing the methodology described by Shnitkind et al.[74] Nucleic acid-free HaloCREB from heparin elution was dialysed into 2 L of [20 mM Tris-HCl (pH 7.5), 1 mM DTT] overnight.

The protocol follows that of described in 2.4.4. The dialysis product was bound to the column at 1 mL/min and gradient eluted on the AKTA FPLC with 2 M NaCl. Similarly, fractions with CREB as seen in SDS-PAGE was measured spectrophotometrically. If the signal is too low, the fractions were first concentrated with a 10,000 MWCO centrifuge concentrator before remeasuring. Fractions with an A_{260}/A_{280} of <0.46 are collected.

2.4.13 Carboxypeptidase A Test

A small sample of the Ni-affinity purification eluate was dialysed into Carboxypeptidase A buffer [50 mM Tris-HCl (pH 7.5), 500 mM NaCl] for 2.5 hours. Protein concentration of the solution was estimated with the Beer-Lambert law spectrophotometrically. 50 units of carboxypeptidase A was added per 1 mg of HaloCREB and incubated at room temperature for 2 hours. A sample was taken from the mixture for each time point, and the reaction quenched with the addition of SDS sample buffer. The samples were then analysed by SDS-PAGE.

2.4.14 Forced dimerisation of HaloCREB-GGC$_{ins}$

Upon elution from heparin column, the eluate was concentrated to a volume of 1.25 mL using a Millipore 10,000 MWCO centrifuge concentrator. 1.25 mL of [100 mM Tris (pH 9.2)] was added to the eluate and transferred to a 15 mL falcon tube. The lid was taped on loosely and placed into the 37°C shaking incubator for an overnight incubation.

2.4.15 Desalting and Buffer Exchange

Nucleic acid-free HaloCREB or CREB was first concentrated using a 10,000 MWCO centrifuge concentrator to a volume of 2.5 mL. PD-10 desalting columns (#GE17-0851-01) equilibrated with 25 mL of biophysical buffer [10 mM MES (pH 6.5), 150 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, 0.05% TWEEN20] were then used to exchange the protein into the same buffer before storing at -20°C for future use. DTT was not included in the biophysical buffer when desalting the forced dimer HaloCREB.

2.5 Fluorescent Labelling

2.5.1 HaloTag TMR Labelling with HaloCREB

As a HaloTag functionality test, 1 µL of the 30 mM HaloTag TMR ligand stock (#G8251) was incubated with 30 µL of a concentrated fraction of HaloCREB for 1 hour in the dark at room temperature. The fraction was then analysed by SDS-PAGE and visualised with a transilluminator.

To label HaloCREB for biophysical measurements, 1.5-fold molar excess of HaloTag TMR ligand was added to the protein sample in biophysical buffer and incubated for 3 hours in the dark at room temperature, before incubating at 4°C overnight. A 10,000 MWCO centrifugal concentrator was used to concentrate to a volume of 2.5 mL. Unbound TMR ligand was removed by PD-10 desalting column, exchanging the protein into biophysical buffer.

2.5.2 Measuring HaloTag TMR Labelling Efficiency

The absorbance of a 1.5 mM solution of HaloTag-TMR at 548 nm and 280 nm were measured using the Varian Cary 500 UV-vis spectrophotometer. As the manufacturer does not provide molar extinction coefficient (ε) or correction factor (CF) for HaloTag TMR, the two values at 280 nm were calculated using the extinction coefficient at 548 nm for TMR[137] as an estimate with respectively Eq. 1 and 2.

$$\varepsilon_{280\text{nm}} = A_{280\text{nm}} \frac{\varepsilon_{548\text{nm}}}{A_{548\text{nm}}} \quad \text{[Eq.1]}$$

$$C\text{F}_{280\text{nm}} = \frac{\varepsilon_{280\text{nm}}}{\varepsilon_{548\text{nm}}} \quad \text{[Eq.2]}$$

The A_{548} and A_{280} of labelled protein was then measured and used to calculate the labelling efficiency with Eq.3. The extinction coefficient of HaloCREB was determined experimentally as outlined in 2.6.2.

$$\text{Labelling Efficiency} = \frac{A_{548\text{nm}} \times \varepsilon_{\text{HaloCREB}}}{(A_{548\text{nm}} - A_{548\text{nm}} \times \text{CF}_{280\text{nm}}) \times \varepsilon_{548\text{nm}}} \quad \text{[Eq.3]}$$

2.6 Biophysical Methods

2.6.1 Electrospray Ionisation Mass Spectrometry

The mass and purity of HaloCREB and CREB at various A_{260}/A_{280} were verified by electrospray ionisation mass spectrometry (ESI-MS) in cationic mode. The ESI-MS was carried out by Dr. David Staunton from the Department of Biochemistry Molecular Biophysics Suite at the University of Oxford.

2.6.2 Determining the extinction coefficient of HaloCREB

The Gill von Hippel method[138] was used to experimentally determine the extinction coefficient of HaloCREB. First, two buffers: native buffer [20 mM Na_2PO4 (pH 6.5)] and denaturing buffer [20 mM Na_2PO_4 (pH 6.5), 8 M guanidium chloride] were made up volumetrically. The denaturing buffer was made up with BioUltra grade or equivalent (≥99.5%) guanidium chloride to minimise background absorbance at 280 nm. A pure sample of HaloCREB was then buffer exchanged into the native buffer using a PD-10 desalting column.

Three solutions were then made up as accurately as possible in Eppendorf tubes. First, a mixture of 50 µL of HaloCREB with 150 µL of native buffer. Second, 50 µL of native buffer and 150 µL denaturing buffer. Third, 50 µL of HaloCREB and 150 µL of denaturing buffer. 150 µL of each sample was then loaded into a Hellma absorbance cuvette micro and its absorbance at 280 nm measured with the Varian Cary 500.

The Hellma cuvette was extensively washed between each sample to prevent contamination. The cuvette was washed with water (3x), 2% Hellmanex (#9-307-011-4-507) (3x), water (3x), 70% ethanol (3x) and lastly water (3x). The cuvette was then dried with pressurised nitrogen gas to expel any remaining droplets. The same washing protocol was applied when changing solutions in any quartz cuvette during all biophysical measurements.

The background absorbance of the two respective buffers were subtracted from the protein sample. The A_{280} of the $HaloCREB_{6M\ GdHCl}$ sample was then corrected for its density using Eq.4 below, where ρ indicates the density of each solution[139,140]: $\rho_{native\ buffer}$ = 1 g/cm^3, $\rho_{6M\ GdnHCl}$ = 1.1844 g/cm^3 and $\rho_{8M\ GdnHCl}$ = 1.4445 g/cm^3.

$$A_{\text{HaloCREB 6M GdHCl, corr}} = A_{\text{HaloCREB 6M GdHCl}} \times \frac{50\mu L \times \rho_{8M\ GdHCl} + 150\mu L \times \rho_{\text{native buffer}}}{\rho_{6M\ GdHCl}} \div 200\mu L \qquad \text{[Eq.4]}$$

The extinction coefficient of HaloCREB at denaturing condition ($\varepsilon_{HaloCREB,\ denaturing}$, in M^{-1} cm^{-1}) was then calculated with Eq.5, with *#Tyrosine* as 9 and *#Tryptophan* as 14. The corrected absorbances at 280 nm of the two HaloCREB conditions was then used in Eq.6 to calculate the extinction coefficient of HaloCREB at its native state.

$$\varepsilon_{\text{HaloCREB, denaturing}} = \#\text{T}rptophan \times 5690 + \#\text{Tyrosine} \times 1280 \qquad \text{[Eq.5]}$$

$$\varepsilon_{\text{HaloCREB, nat}} = \frac{A_{\text{HaloCREB, native}} \times \varepsilon_{\text{HaloCREB, denaturing}}}{A_{\text{HaloCREB 6M GdHCl, corr}}} \qquad \text{[Eq.6]}$$

2.6.3 Preparation of DNA oligos for biophysics

DNA oligonucleotides in table 2.x below were purchased from Invitrogen with a 1 µM scale of synthesis. The lyophilised primers were made up to a concentration of 1 µM according to the oligo nanomole yield from the manufacturer. Equal volumes of the corresponding primers (if annealing linear DNA) are mixed, diluted 10-fold with water and aliquoted into a 96-well PCR plate.

To anneal hairpin DNA (CREh), the mixture was heated to 95°C for 2 minutes in a thermocycler, upon which the DNA was immediately placed into ice and incubate for 30 minutes. Alternatively, to anneal linear DNA (nCRElin and CRElin), the mixture was heated to 95°C for 2 minutes and gradually cooled to 4°C over the course of 30 minutes.

The concentration of the DNA was then measured spectrophotometrically using the Beer-Lambert Law (Eq.7). If the DNA is tagged with AlexaFluor488, the measurement is made at 495 nm and calculated using the extinction coefficient of 73,000 M^{-1} cm^{-1}. Alternatively, the concentration of non-fluorescent DNA was measured by with the same equation but with the extinction coefficient of the oligonucleotide at 260 nm.

$$A = \varepsilon\, c\, l \qquad \text{[Eq.7]}$$

DNA Oligonucleotide		Sequence
AF488-CREh		5'-AF488-CCTGACGTCAGCCCCCTGACGTCAGG-3'
AF488-nCRElin	F	5'-AF488-GGCTAAAGCATTCT-3'
	R	5'-AGAATGCTTTAGCC-3'
CRElin	F	5'-CCTGACGTCATCCG-3'
	R	5'-CGGATGACGTCAGG-3'
nCRElin	F	5'-GGCTAAAGCATTCT-3'
	R	5'-AGAATGCTTTAGCC-3'

Table 2.x | Sequence of DNA oligos used for biophysical measurements. Half CRE sites are labelled in blue.

2.6.4 Equilibrium measurements for DNA-binding with fluorescence anisotropy

Aliquots of HaloCREB was defrosted from storage at -20°C and their concentration was remeasured using the experimentally determined extinction coefficient and the Beer-Lambert law (Eq.7). A 120 µL solution of HaloCREB and fluorescent DNA were mixed to the desired final concentration, diluting with biophysical buffer [10 mM MES (pH 6.5), 150 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, 0.05% TWEEN20] in 1.5 mL amber microcentrifuge tubes (#I1415-1007, Starlab). The same buffer without DTT is used when doing measurements with the forced dimer mutant.

Fluorescent DNA was then diluted to the same probe concentration with biophysical buffer in LoBind DNA Eppendorf tubes (#0030108051) wrapped with tin foil. 60 µL of the fluorescent DNA solution was aliquoted to each of the successive dilution concentration point. 60 µL of the first solution was pipetted into the second

tube and mixed thoroughly via vortexing. This was then repeated until the last concentration point. All solutions were then incubated at room temperature in the dark for at least 30 minutes for equilibration before measurements.

The measurements were performed on two machines: the PerkinElmer LS 55 fluorescence spectrometer and the Horiba FluoroMax-4 spectrofluorometer, both of which uses the Hellma fluorescence cuvette ultra micro cuvette for measurements in a volume of 60 µL. The excitation and emission wavelengths of machine were set at respectively 495 nm and 519 nm, respectively. The bandpass size is 8/8 nm for the PerkinElmer and 10/10 nm for the Horiba.

When using the PerkinElmer, a blank was first performed by measuring the I_{vv} and I_{vh} of the biophysical buffer with the gain set to high and an integration time of 1 second. The G-factor was then measured with a solution of a fluorescent DNA before taking anisotropy measurements, starting with the lowest protein concentration. The samples were allowed 30 seconds to equilibrate in the measurement cell to the appropriate temperature. The protocol for using the Horiba is largely similar, except without the need of measuring the G-factor.

The HaloCREB monomer concentration / 2 was then plotted against the fluorescence anisotropy for each sample. A single-state curve with dimerisation, or a two-state binding curve, was fitted using respectively Eq. 8 and Eq. 9 below.

In a single-state binding curve with dimerization (Eq.8), R_f = initial anisotropy, dR = change in anisotropy and K = equilibrium constant. D is constrained to the fluorescent DNA concentration and L is constrained to the $K_{d,homodimersation}$. Equation was derived by Dr. Mikhail Kuravsky from the Shammas laboratory.

$$y = R_f + \frac{dR}{D}\left(0.5x - \frac{D(y-R_f)}{dR\dfrac{D-\dfrac{D(y-R_f)}{dR}}{K}}\right) - 0.5\sqrt{\frac{D(y-R_f)}{dR\dfrac{D-\dfrac{D(y-R_f)}{dR}}{KL}}} \qquad \text{[Eq.8]}$$

In the two-state binding curve, R_f = initial anisotropy (Eq.9), dR_1 = change in anisotropy upon the first transition, K_1 = first equilibrium constant, dR_2 = change in anisotropy upon the second transition and K_2 = second equilibrium constant. D was constrained to the fluorescent DNA concentration. Equations were derived by the Shammas laboratory.

$$y = R_f + \frac{dR_1(K_1 + D + \frac{x}{2} - \sqrt{(K_1 + D + x)} \div 2D)}{dR_2(K_2 + D + \frac{x}{2} - \sqrt{(K_2 + D + x)} \div 2D)} \quad \text{[Eq.9]}$$

2.6.5 Competition fluorescence anisotropy for HaloCREB to unlabelled nCRE DNA

700 µL of solutions of 1) 100 nM AF488-nCRElin, 500 nM HaloCREB and 2) 100 nM AF488-nCRElin, 500 nM HaloCREB, 5 µM nCRElin were made up in amber Eppendorf tubes. The two solutions were then mixed according to table 2.x below.

Fluorescence anisotropy of the 11 samples was then measured using the PerkinElmer as described in 2.6.4. A curve was then fitted with Eq. 11 to yield IC_{50} of the reaction. Using the IC_{50} and $K_{d,nCRE}$ determined in section 2.6.4, $K_{i,unlabelled\ nCRE}$ can then be extracted using the method from Nikolovska-Coleskato et al.[141]

Sample	Solution 1 (µL)	Solution 2 (µL)
1	60	0
2	54	6
3	58	12
4	42	18
5	36	24
6	30	30
7	24	36
8	18	42
9	12	48
10	6	54
11	0	60

Table 2.xi | Schematic of assembling the samples to measure the equilibrium binding constant of HaloCREB to unlabelled nonCRE DNA. Solution 1 contains 100 nM AF488-nCRElin, 500 nM HaloCREB. Solution 2 contain 100 nM AF488-nCRElin, 500 nM HaloCREB, 5 µM nCRElin.

2.6.6 Association DNA-binding kinetics with fluorescence stopped-flow

The association rate constant (k_{on}) for HaloCREB with CREh was measured by rapidly mixing: 1) fluorescent CREh DNA solution; and 2) various concentrations of unlabelled HaloCREB. To ensure there is a pseudo-first order reaction, HaloCREB monomer was at least 20-fold in excess compared to the fluorescent DNA. The unlabelled HaloCREB was diluted to a range of concentration from 200 nM to 300 nM. For each protein concentration, 5 mL of the fluorescent DNA (10 nM AF488-CREh) and the unlabelled HaloCREB solution was needed. The solution were made up to the correct final volume by the addition of biophysical buffer using a balance and the weight of the solution as a measure of the total volume. Measurements of the forced dimer were done in biophysical buffer without DTT.

The measurement regime and washing protocol follow the protocol described by Crabtree & Shammas[131]. The measurements were done in an Applied Photophysics SX20 stopped flow. A cut-off filter of 515 nm was used. To ready the machine for measurement, the machine was first washed with biophysical buffer. Bubbles were expelled by repeatedly emptying and filling the drive syringes with buffer. The mixing chamber and the rest of the system was also washed by manually rotating the valves to expel the buffer through the chamber to the waste. Syringes with the respective solutions were then attached to the machine. Bubbles were expelled from the drive syringes in the same manner. The system was then washed with the solutions by manually emptying the mixing chamber three times.

Calibrating the emission wavelength to 495 nm, the detector voltage was adjusted using the AutoPMT function. An initial spectrum of three replicate shots was recorded for 60 seconds on a logarithmic scale without the pressure hold. The averaged kinetic trace was then fitted to a single exponential decay. The observed association binding constant k_{obs} can then be estimated as the exponent of the fit. The estimate was then used to determine the timescale for data collection using Eq. 10.

$$\frac{10 \ln 2}{k_{\mathrm{obs}}} \qquad \text{[Eq.10]}$$

To measure the concentration dependence of k_{obs}, 36 replicate shots were performed with the selected timescale on a linear scale for each protein concentration. The machine was washed in a similar manner when exchanging between solutions of different protein concentration. The protein was diluted to its working concentration by the addition of biophysical buffer using a balance to measure the weight of the solution. The averaged spectrum of each concentration was then fitted to a single exponential (Eq.11).

$$F_x = \Delta F^{(-k_{obs} \times x)} + F_1 \quad \text{[Eq.11]}$$

In which, F_x describes the fluorescence at HaloCREB dimer concentration X, ΔF is the change in fluorescence, k_{obs} is the observed association rate constant and F_1 is the end fluorescence. The k_{on} (in nM^{-1} s^{-1}) of HaloCREB to CREh can then be calculated by obtaining the slope of the linear fit describing the concentration dependence of k_{obs}.

The association rate constant (k_{on}) to CREh in presence of nCRE DNA was measured in a similar manner as described above, with the inclusion of 2 µM nCRElin DNA in the DNA solution.

2.6.7 Dissociation DNA-binding kinetics with competition fluorescence stopped-flow

The dissociation rate constant (k_{off}) was measured in a competition experiment by rapidly mixing: 1) 10 nM HaloCREB with 10 nM fluorescent CREh DNA (AF488-CREh); and 2) various concentrations of competitor DNA (CRElin) at minimum 300-fold excess of the CREh. The two solutions were prepared in the same manner as the k_{on} measurements.

The machine was also calibrated in the same way as described above. An initial spectrum was taken for 600 seconds instead to obtain the timescale for data collection. 36 replicate shots were performed per concentration of competitor DNA to obtain an average trace that can be fit with a single exponential decay to calculate the k_{app} (Eq.11). By plotting the concentration dependence of k_{app}, k_{off} (in s^{-1}) can then be extracted as the plateau of the single-phase exponential decay.

2.6.8 CREh binding kinetics from fluorescence stopped-flow

Dissociation constants (K_d, nM) estimated using kinetic parameters of HaloCREB to CREh were calculated by Eq. 12. k_{on} and k_{off} were obtained from stopped flow spectroscopy as described in above.

$$K_{\mathrm{d,CRE,kinetics}} = \frac{k_{\mathrm{off}}}{k_{\mathrm{on}}} \quad \text{[Eq.12]}$$

2.6.9 Homodimerisation equilibrium kinetics with fluorescent anisotropy

The homodimerisation equilibrium constant ($K_{d,homodimerisation}$) was measured with a fluorescently tagged WT HaloCREB. The HaloCREB was labelled with the HaloTag-TMR via the protocol outlined in 2.5.2. As the labelling efficiency is not 100%, the concentration of the HaloCREB was measured spectrophotometrically using the determined CF_{280} and ε_{280}.

A 25 µM solution labelled HaloCREB was serially diluted 2-fold with biophysical buffer in amber Eppendorf tubes. The solutions were then incubated in the dark at room temperature for 30 minutes to allow to reach equilibrium.

Measurements were performed on the PerkinElmer starting from the lowest HaloCREB concentration. The excitation and emission wavelengths were set to 555 nm and 585 nm respectively with a bandpass width of 8/8 nm. The background I_{vv} and I_{vh} was measured with biophysical buffer and G-factor was measured with the sample of 200 nM HaloCREB. Samples was first measured with a high detector gain. When one of the recorded intensities reached 990, the gain was set to medium to prevent saturation. The same was done from medium to low detector gain.

2.6.10 Homodimerisation equilibrium measurements with native MS and mass photometry

Sample preparation and native mass spectrometry (MS) was kindly performed by Zihao Wang from the Benesch Lab of the Oxford Department of Chemistry. WT HaloCREB was buffer exchanged into 1 M

ammonium acetate (pH 6.9) using a Bio-Spin size exclusion spin column and ionised by nanospray ionisation. The sample was performed with the Waters Synapt G2-Si using TOF-MS.

Mass photometry[142] and spectrum analysis of WT HaloCREB was kindly performed by Joshua Bishop from the Benesch Lab of the Oxford Department of Chemistry. The samples were collected in biophysical buffer without TWEEN [10 mM MES (pH 6.5), 150 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT], diluted with [137 mM NaCl, 2.7 mM KCl, 9.5 mM Phosphate (pH 7.4)]. Samples were analysed using the Refeyn One^{MP} and a spectrum was obtained, detailing the hits corresponding for the monomer and the dimer peak. The percentage of dimer in solution was then calculated. Given the K_d of a homodimer is defined by Eq. 13, the $K_{d,homodimer}$ can then be calculated using Eq.14.

$$K_d = \frac{[M]^2}{[D]} \quad \text{[Eq.13]}$$

$$K_d = [P]_0 \frac{1}{\frac{[D]}{[M]}(2\frac{[D]}{[M]} + 1)} \quad \text{[Eq.14]}$$

3. Results

3.1 Establishing a protocol for purifying DNA-free HaloCREB and CREB

3.1.1 Choice of expression construct: HaloTag-CREB-His6 in pET303

The base expression construct used in this study contains an N-terminal HaloTag and C-terminal polyhistidine tag. A TEV cleavage site separates the HaloTag from the N-terminal of CREB. The polyhistidine tag allows the protein to be purified by immobilised metal affinity chromatography and is widely used in protein purification.

HaloTag is a highly specific and flexible tag that can be both used for protein visualization and protein purification. HaloTag is a 33 kDa deactivated haloalkane dehydrogenase. A mutation in the catalytic histidine-272 to a phenylalanine enable the protein to trap its substrate through the formation of an irreversible covalent bond[143]. The haloalkane substrate can then be functionalized through various means. The HaloTag-TMR ligand contains a fluorophore attached to the reactive linker, whereas the HaloLink resin immobilize the substrate on a sepharose bead to act as resin for affinity chromatography. Introduction of HaloTag enables downstream *in vivo* studies by providing a straight-forward fluorescent labelling protocol.

Functionalizing proteins with tags, especially those of similar size with the original protein, can significantly alter its functionality and affinity. This is not the case with HaloCREB. Through chromatin immunoprecipitation (ChIP) assays, it has been shown that the addition of HaloTag does not affect CREB's ability to bind to DNA *in vivo*[144]. Another study also found that HaloTag does not affect or hinder CREB's ability to induce transcription[145]. The combination of both studies shows compelling evidence that despite being similar in size, the HaloTag does not alter CREB's functionality. The added flexibility makes the HaloCREB construct an attractive option to explore the mechanism of CREB and effect of its allostery on a molecular level. It is important to note that the linker region between HaloTag and CREB for both studies were not described. As such, it might differ from my construct, potentially leading to different results.

The choice of using the pET plasmid as the backbone is largely due to convenience brought by the Invitrogen Champion Vector Kit. The vectors from the kit contains either a N-terminal His6 tag or C-terminal His6 tag, allowing the polyhistidine tag to be inserted with ease. The polyhistidine tag was chosen to be attached to the C-terminus to prevent any truncated products from purifying during Ni-affinity column chromatography. Considering the large number of rare codons within the CREB ORF, truncation is possible even with the use of the Rosetta2 host strain. Therefore, the C-terminal polyhistidine tag is chosen to improve the purification efficiency.

3.1.2 Choice of expression strain: Rosetta2(DE3)

Rosetta2 (DE3) cells are chosen as the expression strain for this protocol due to the study by Lopez et al.[146] In which, Lopez and colleagues reported that the use of the Rosetta (DE3) strain significantly increases yield of CREB compared to the more commonly used BL21. Rosetta and Rosetta2 cells carry the pRARE and pRARE2 plasmid, supplying rare E. coli codons for eukaryotic protein expression. Specifically, the seven tRNAs that are supplemented by the pRARE2 plasmid correspond to the codons AUA, AGG, AGA, CUA, CCC, GGA and CGG.

Codon	Occurrence in HaloCREB	Frequency
AUA	I177	1
AGG	R111, R116, R117, R121, R136, R280	6
AGA	R271, R288, R300	3
CUA	L55, L276, L315	3
CCC	P12(HaloTag), P51, P59, P222, P244	5
GGA	G6, G13, G68, G133, G166, G174, G175, G216, G233, G251	10
CGG	/	0

Table 3.i | Occurrence of rare E. coli codons supplemented by the pRARE2 plasmid in Rosetta2 (DE3) cells. Unless specified, denoted residue number are of the CREB sequence.

As the CREB gene was purchased from an ORF library, codon optimization for E. coli could not be performed. Using the Rosetta2 strain for expression would greatly alleviate the pressure from codon bias. Furthermore, as the rare codons appear to be clustered within certain parts of the transcript (Table 3.i), the use of the

Rosetta2 cells would vastly reduce the likelihood of transcript termination due to ribosome stalling, leading to a larger expression yield.

3.1.3 Plasmid construction and expression test

A HaloTagged construct (HaloCREB) was used for the biophysical parameterisation done in this study to complement future cell culture work. HiFi assembly was performed to obtain a plasmid of HaloCREB in a pET303 backbone. The fragments (with overhangs) of the two inserts and pET303 backbone were amplified as described. As the initial pET303 amplification failed, the reaction was repeated with touchdown PCR and increasing the working concentration template DNA to 1.69 ng/µL. This generated a sufficient fragment for HiFi assembly. Upon assembly, the PCR product was immediately transformed into NEB 5α cells. Agarose gel electrophoresis analysis of the plasmids from each colony purified by QIAprep Miniprep is seen in Figure 3.1. Plasmids from all four colonies corresponds to the bands at expected size of 7 kbp. The sequences of the plasmids were confirmed with DNA sequencing. The resultant plasmid is hereby referred to as PL103. PL103 consists of a N-terminal HaloTag and a C-terminal His6 tag of CREB. The HaloTag and CREB are separated by a TEV cleavage site.

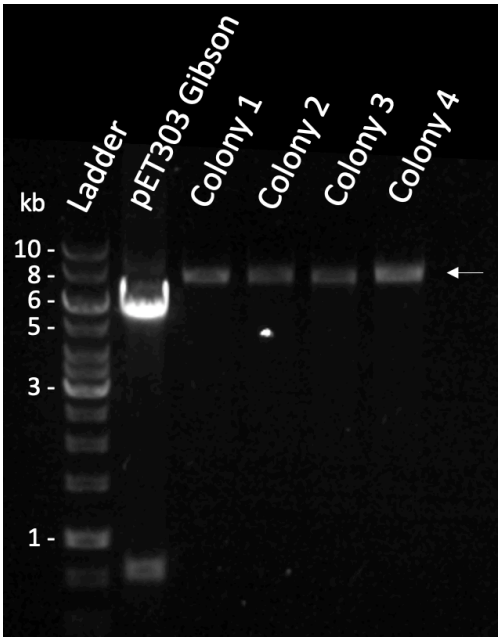


Figure 3.1 | Gel electrophoresis of the Gibson assembly. Lane 2 contains the pET303 Gibson fragment as positive control. Lane 3-6 contains plasmids from colonies harvested by the QIAGEN MiniPrep kit, all of which correspond to the expected size of 7 kbp of the PL103 plasmid.

PL103 was subsequently transformed into Rosetta2(DE3) cells for expression. The expression of HaloCREB was tested at 37°C over the course of 6 hours.

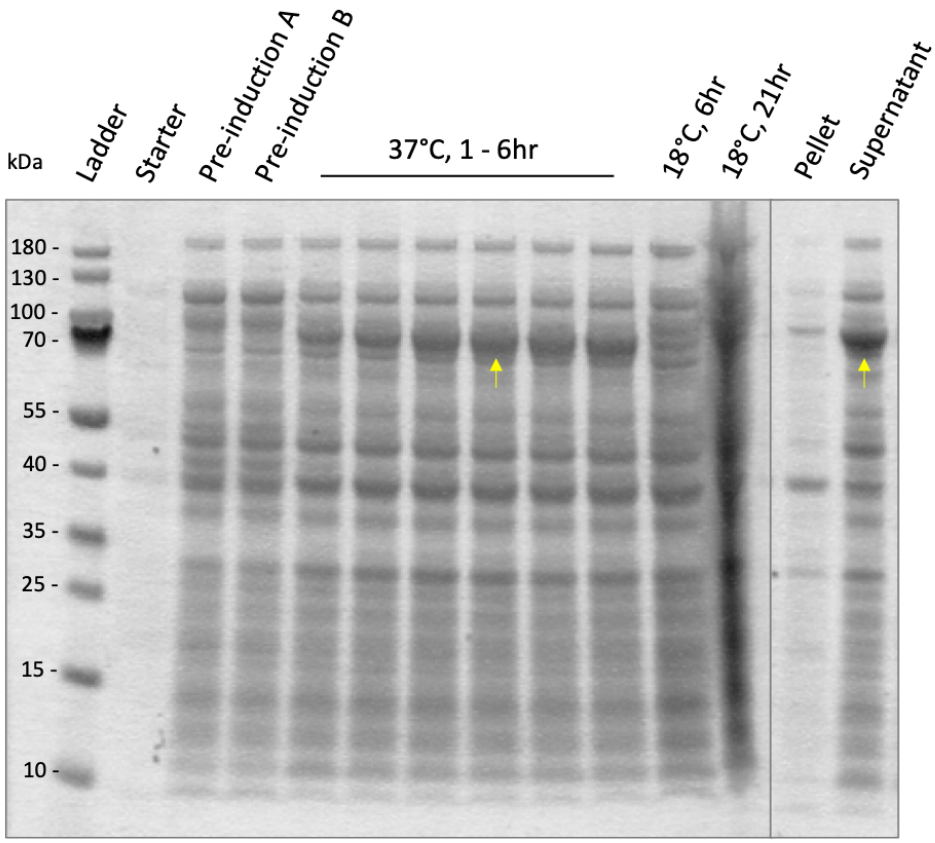


Figure 3.2 | Expression test of PL103 in Rosetta2(DE3) E. coli. Analysis performed on NuPAGE 4-12% bis-tris gel with MES running buffer. Both pre-induction samples originate from the same starter culture. Yellow arrow indicates HaloCREB, corresponding the expected molecular weight of 70 kDa. Lane 12 (18°C, 21hr) is inadequately lysed. Right gel shows the pellet and supernatant upon lysis via sonication and centrifugation. HaloCREB can be found in the supernatant but not the insoluble fraction.

Figure 3.2 shows that the expression of HaloCREB plateaus at 4 hours after induction at 37°C. Unfortunately, the effectiveness of 18°C overnight culture cannot be determined due to the insufficient lysis of the 21-hour sample prior to loading on the polyacrylamide gel. The right panel shows that HaloCREB was found in the soluble fraction upon cell lysis, which simplifies downstream protein purification.

3.1.4 Existing CREB Purification Protocols

Conventional CREB purification protocols begin with heat treating cell lysate to various temperatures[146–149]. Figure 3.3 shows that this is ineffective for HaloCREB, presumably due to the globular nature of the HaloTag. A minor decrease of background proteins (without affecting HaloCREB) can be seen for heat treatment at 55°C for 5 minutes, however incubation at higher temperatures significantly reduces the amount of soluble HaloCREB. While HaloCREB formed no visible aggregates at 55°C, whether the incubation will cause partial unfolding and disruption in functionality of HaloTag is unknown.

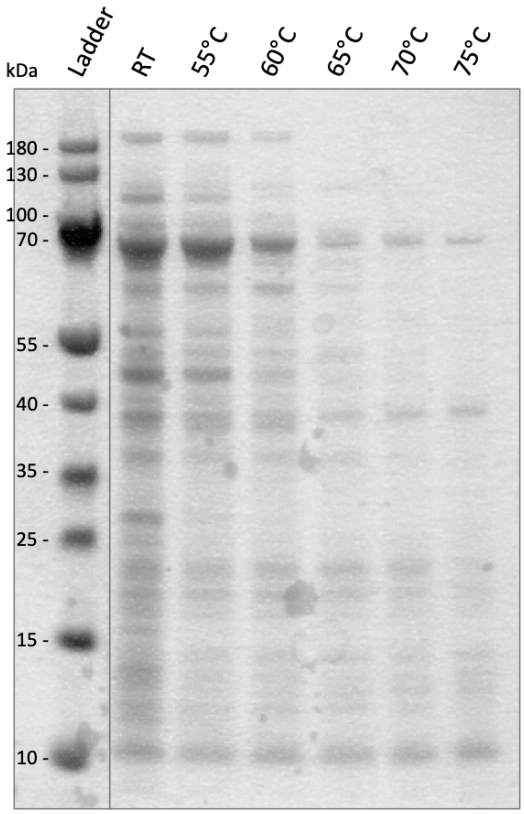


Figure 3.3 | Temperature denaturing test on HaloCREB. Cell lysate are heated to the respective temperature on a thermoshaker for 5 minutes. The samples were then centrifuged at 16,800 x g for 5 minutes, then the supernatant for each sample is collected for analysis. Room temperature cell lysate in lane 2 acts as the positive control. No significant reduction in HaloCREB was obsevered in the 55°C sample, but a reduction in HaloCREB can be seen starting from 60°C.

A recent NMR paper by Shnitkind et al. purified CREB by Ni-affinity column followed by cation exchange chromatography[74]. I followed this protocol for HaloCREB but the sample had remaining contaminants, rendering it unsuitable for biophysical studies (Figure 3.4). A gradient elution was used for both chromatography columns, and an extensive 2 M NaCl wash was performed on the Ni-affinity column prior to

elution. The Ni-affinity column was very effective in the initial purification of HaloCREB, but significant contaminants remain in the elution fractions (data not shown, refer to Figure 3.5 for Ni-affinity column).

Furthermore, the samples were not clear from nucleic-acid contamination. A quick but sensitive test was performed by measuring the A_{260}/A_{280} ratio. A ratio of 0.57 for tryptophan containing proteins indicate a nucleic-acid free same. The A_{260}/A_{280} of the Ni-affinity eluate was 0.61, indicating the use of 1 M NaCl and extensive 2 M NaCl wash was only slightly insufficient in removing nucleic acid contamination.

Dialysis of the eluate into the ion exchange buffer resulted in slight aggregation of HaloCREB, likely due the large change in salt concentration (from 1 M to 100 mM NaCl). The cation exchange column yielded two elution peaks: the first one being significantly sharper than the second. The peak towards the end of the broad peak (indicated by the purple arrow) corresponds with a peak in the conductivity curve (brown on chromatogram), suggesting that it is an artefact from the pump change during AKTA operation.

Unfortunately, as seen in the SDS-PAGE in figure 3.4c, none of the elution fraction contains a pure fraction of HaloCREB, and therefore the use of cation exchange column following Ni-affinity column is not sufficient for our purposes as it is.

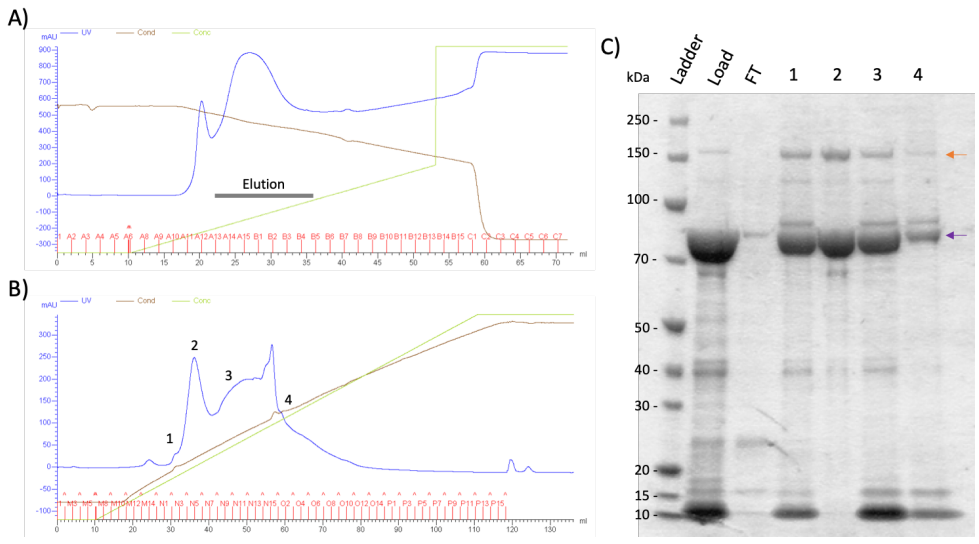


Figure 3.4 | Purification of HaloCREB by Ni-affinity chromatography and cation exchange chromatography developed by Shnitkind et al[73]. A) Elution profile of the Ni-affinity chromatography purification. The increase in absorbance over the elution period is due to the background absorbance of imidazole. SDS-PAGE for the individual fractions are not shown. Grey bar represents fractions that were collected as eluate, corresponding to the load fraction in panel C. B) Elution profile of the cation exchange chromatography purification. The peak between 3 and 4 is an artefact from pump change during AKTA operation. C) SDS-PAGE analysis of the cation exchange chromatography. Residual contamination can still be seen in all four fractions of the protein. Orange arrow indicates HaloCREB dimer and purple arrow indicates CREB.

It is worth noting that HaloCREB monomer and dimer are disordered proteins which are enriched in negatively charged residues, SDS treatment would make the protein more negatively charged than usual. As such, they would migrate less towards the cathode and correspond to a higher molecular weight on polyacrylamide gels.

3.1.5 Development of HaloCREB purification protocol

As existing protocols were insufficient, various new HaloCREB purification strategies were tested. Previously Ni-affinity column chromatography was shown to be an effective first step to rid of most cellular contaminant (Figure 3.5). The buffer composition was changed to match the protocol used to purify the bZIP domain construct in the Shammas lab – reducing NaCl concentration from 1 M to 500 mM helps reduce the formation of aggregates upon dialysis but may make it harder to remove nucleic acids bound to CREB. To compensate, the 2 M NaCl wash on the His-trap column was extended to be performed overnight to maximise the amount of nucleic acid removed during the wash.

Unfortunately, the Ni-affinity column alone (under both buffer conditions) was insufficient to remove all protein and nucleic acid contamination. The A_{260}/A_{280} ratio of the eluate was roughly 0.6, which again suggests a small amount of residual nucleic acid contamination. Prior to further purification steps, the eluate was incubated with 10 mM DTT to reduce the protein to a monomer state. All subsequent purification buffers contain at least 1 mM DTT to ensure the protein remains reduced.

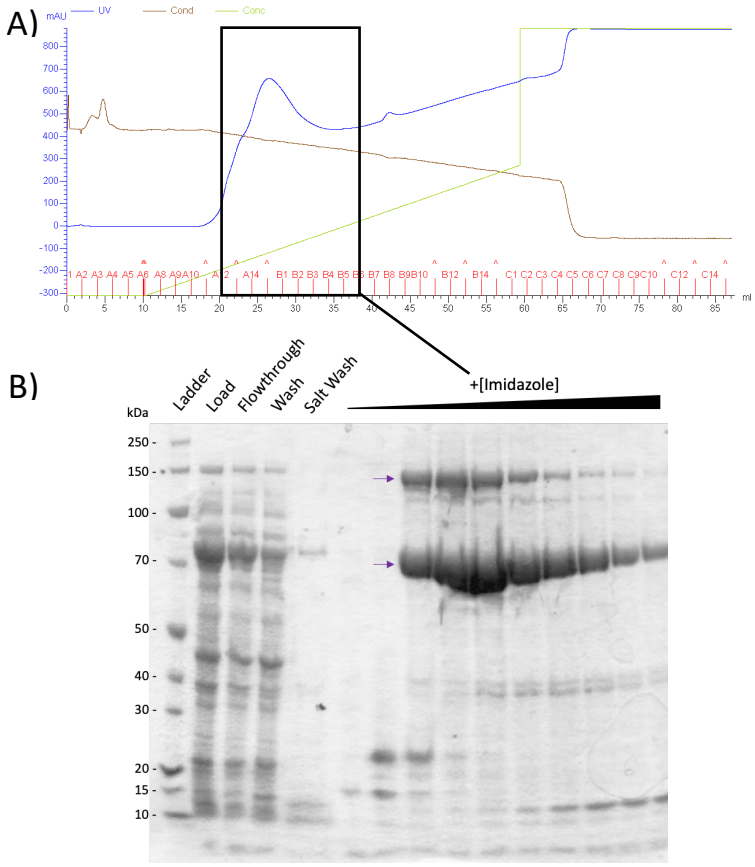


Figure 3.5 | Gradient elution of Ni-affinity column. A) Elution profile of the purification. Increase in absorbance over the course of elution due to background absorbance of imidazole. Fractions highlighted by black rectangle were analysed by SDS-PAGE. B) SDS-PAGE analysis of purification. Purple arrows indicate monomer and dimer HaloCREB, corresponding to the correct molecular weight (70 kDa and 140 kDa).

Next I tested size exclusion chromatography as a secondary step by loading the Ni-affinity eluate onto a Superdex 200 5/150 GL column (Figure 3.6). No protein was observed from the elution chromatogram or the SDS-PAGE gel of all fractions. However, a spike in the chromatogram during the routine 1 M NaOH wash later suggested that HaloCREB might be bound to the column nonspecifically. The purification was repeated with a buffer at pH 11 to mimic the high pH of the 1 M NaOH in an attempt to remove the nonspecific interaction, but was ultimately ineffective as well (data not shown).

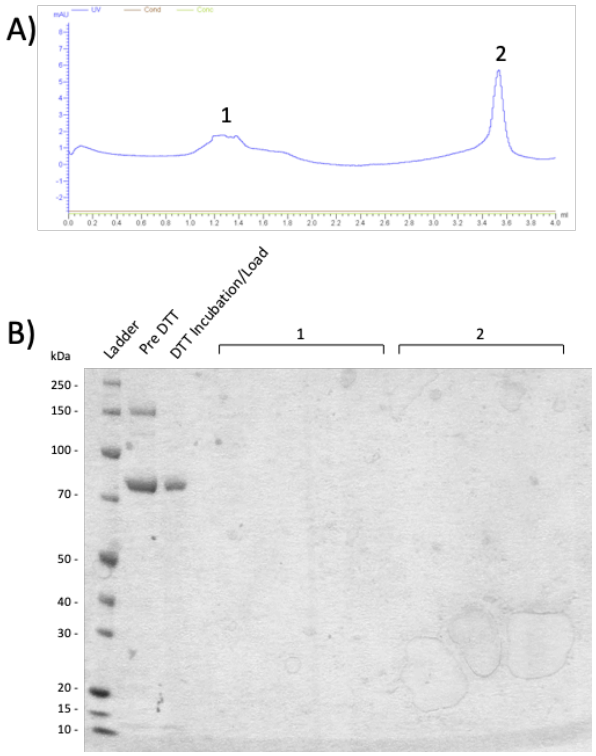


Figure 3.6 | Size Exclusion Chromatography with Ni-affinity eluate. The elution fraction was incubated with 10 mM DTT for 1 hour before loaded onto the column. A) Elution profile of the purification. Peak 2 likely corresponds to imidazole or DTT. B) SDS-PAGE of the first and second peak seen in the elution profile. No protein can be seen in any of the fractions.

Instead I examined the possibility of using ion exchange as a secondary purification step. Anion and cation exchange columns (Q and SP) were considered. As HaloCREB has a pI of 5.5, a buffer at pH 6.5 should ensure binding to an anion exchange column (Figure 3.7). This was confirmed by the elution profile and SDS-PAGE. However, various protein contamination remained amongst the elution fractions. This indicates that either: 1) the contaminants have a similar pI; or 2) the contaminants were associated with HaloCREB and therefore co-purifying.

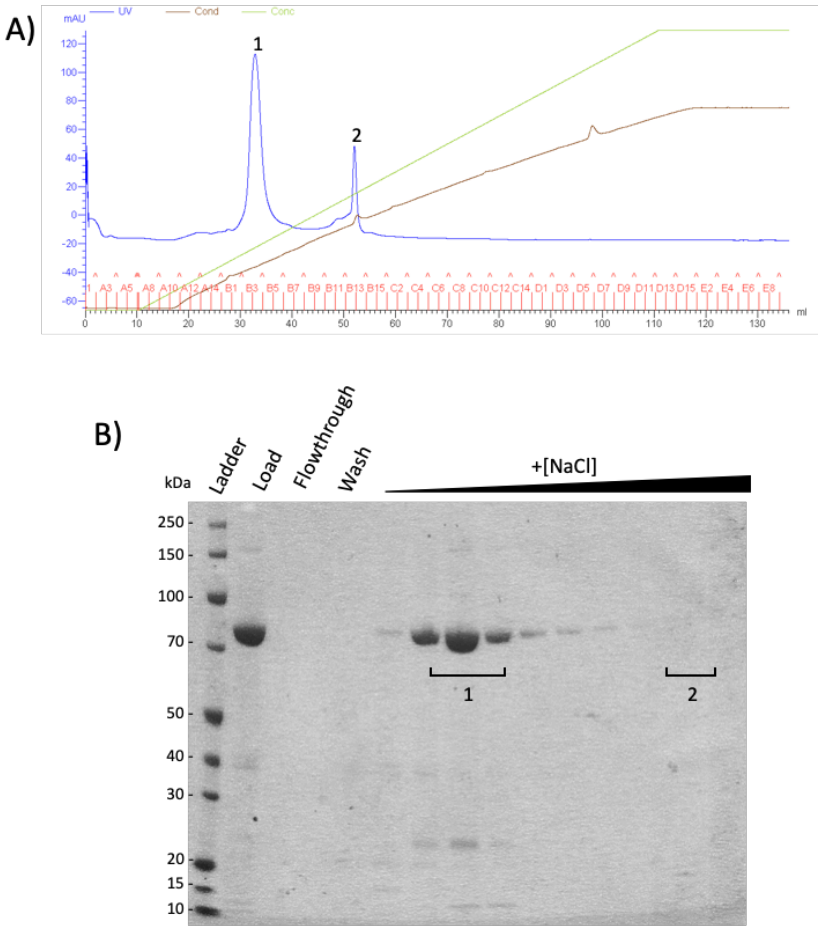


Figure 3.7 | Gradient elution of anion exchange chromatography after Ni-affinity purification. The purification was performed with 20mM bis-tris, 2mM DTT and at pH 6.5. A) Elution profile of the purification. B) SDS-PAGE analysis of the two elution peaks. Residual contamination can be seen around 12 kDa and 25 kDa of the first peak. The second peak is likely non-protein contamination. The purity of the fractions would not be sufficient for biophysics analysis.

Cation exchange column was then tested next. The purification was first performed with the same buffer as the anion column at pH 6.5. HaloCREB binds to the cation exchange column at pH 7.5 as discussed in section 3.1.2, and reducing the pH to 6.5 without crossing the pI yielded the same result

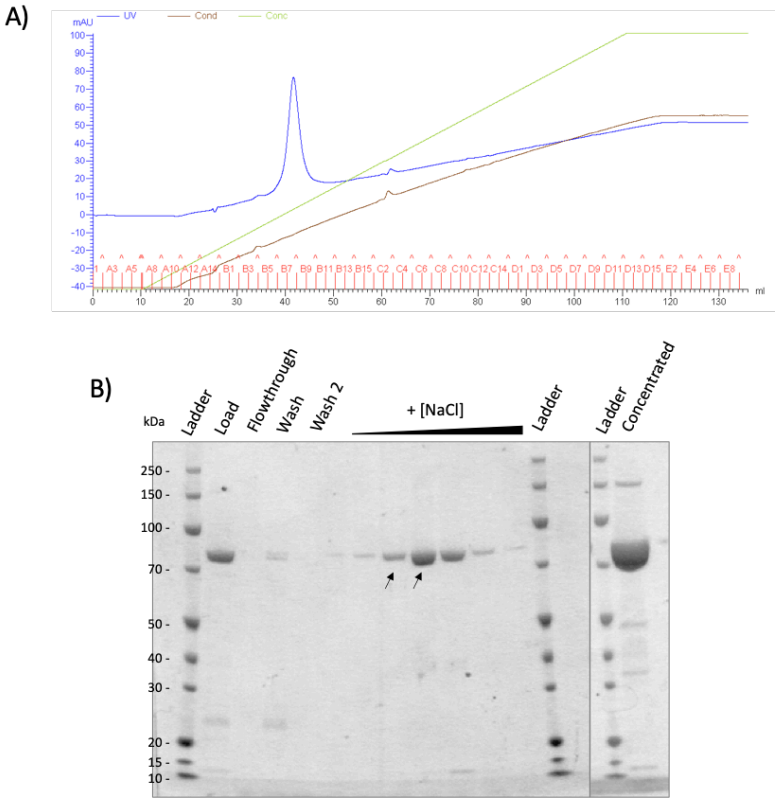


Figure 3.8 | Gradient elution of cation exchange chromatography after Ni-affinity and anion exchange purification. A) Elution profile of the purification. The increase in absorbance over the elution is due to the different levels of DTT oxidation between the two buffers. B) SDS-PAGE of the elution peak in panel A. Fractions indicated by the black arrow was then concentrated on a 10,000 MWCO centrifugal concentrator and shown on the right gel. A faint smear in the background and residual contamination can be seen in the concentrated fraction.

While the two elution fractions indicated by the black arrow seem pure on SDS-PAGE, some contamination with masses corresponding to 12 kDa, 35 kDa and 50 kDa are present. To identify the contaminants and the functionality of HaloTag after purification steps, the sample was incubated with HaloTag-TMR fluorescent ligand and analysed by fluorescent SDS-PAGE (Figure 3.9).

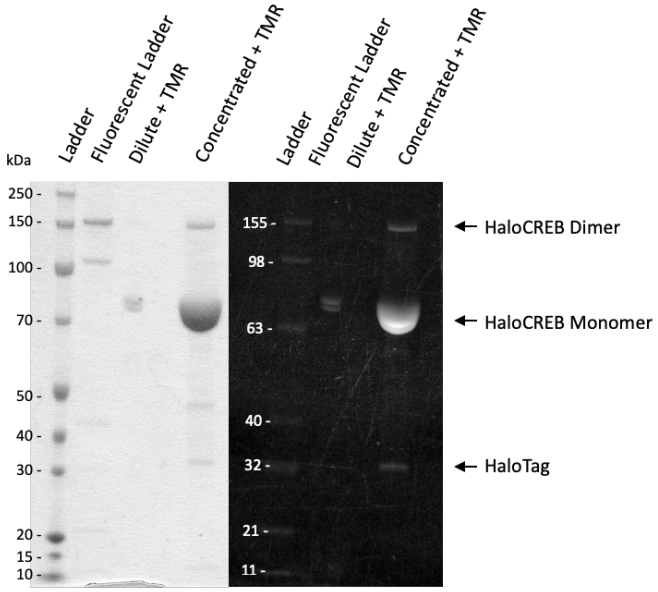


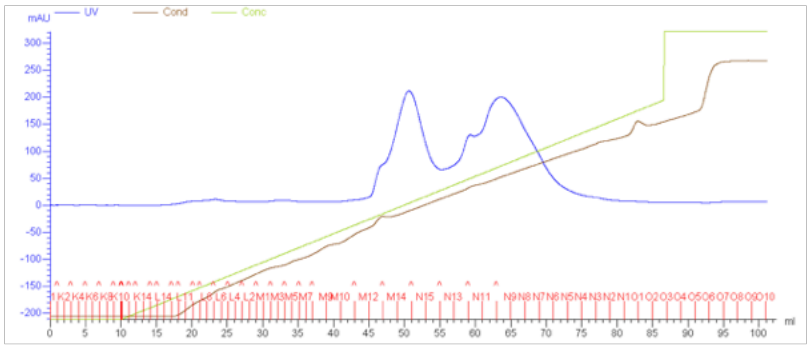
Figure 3.9 | HaloCREB fluorescent ligand test. HaloCREB was incubated with HaloTag TMR ligand for 1 hour and analysed by SDS-PAGE. The polyacrylamide gel was first imaged using an UV-transilluminator (right panel) before staining with coomassie (left panel).

Figure 3.9 first shows that the HaloTag remains functional after three steps of purification. Second, it confirms that the major band at 70kDa corresponds to HaloCREB monomer and the band at 150 kDa corresponds to HaloCREB dimer. Third, a visible fluorescent band at 32 kDa corresponds to HaloTag with a molecular weight of 33 kDa. This suggests that the contamination may be due to residual TEV protease in one of the purification columns, which may be solved by using a new column.

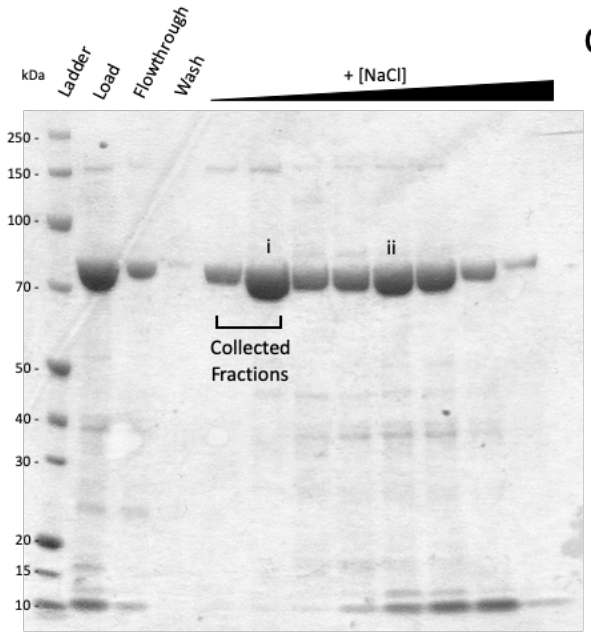
However, the protein sample had a A_{260}/A_{280} of 0.6, meaning it was not completely free of nucleic acid. Furthermore, the same purity was not achieved in a larger scale purification despite the use of a new column for each step (data not shown). Therefore, a better protocol to obtain a nucleic acid and contaminant-pure sample of HaloCREB has to be developed. I continued with further attempts to optimise the purification.

Heparin chromatography was tested next as the resin mimics the polyanionic structure of DNA so, it might displace any bound nucleic acid to from HaloCREB. An initial test was performed with the Ni-affinity chromatography eluate in a 2 hour reflow binding step to increase binding events and possibility to displace bound nucleic acids (Figure 3.10).

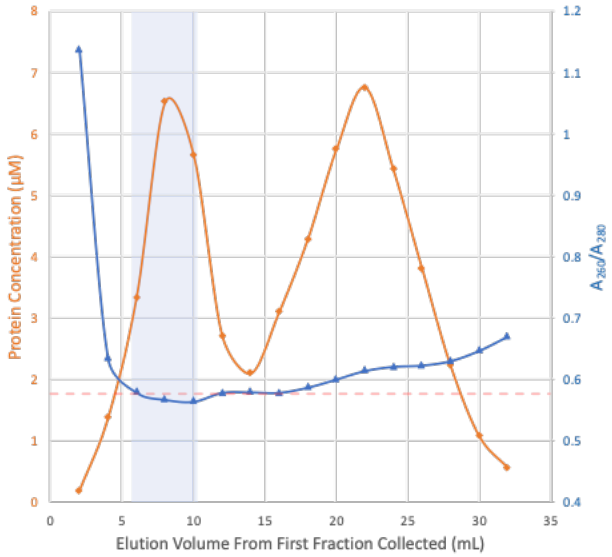
A)


mAU
UV
Cond
Conc
ml

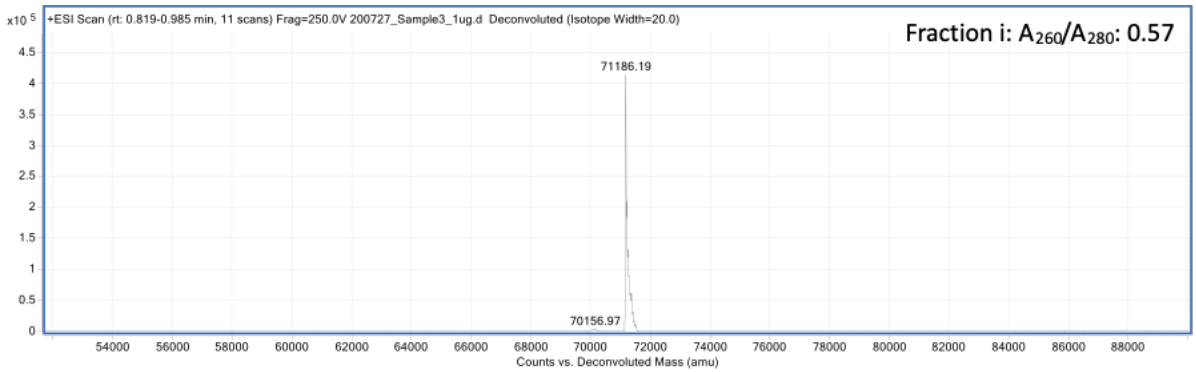
B)


kDa
Ladder
Load
Flowthrough
Wash
+ [NaCl]
250
150
100
70
50
40
30
20
15
10
i
ii
Collected Fractions

C)


Protein Concentration (µM)
A260/A280
Elution Volume From First Fraction Collected (mL)

D)


+ESI Scan (rt: 0.819-0.985 min, 11 scans) Frag=250.0V 200727_Sample3_1ug.d Deconvoluted (Isotope Width=20.0)
Fraction i: A260/A280: 0.57
71186.19
70156.97
Counts vs. Deconvoluted Mass (amu)

E)

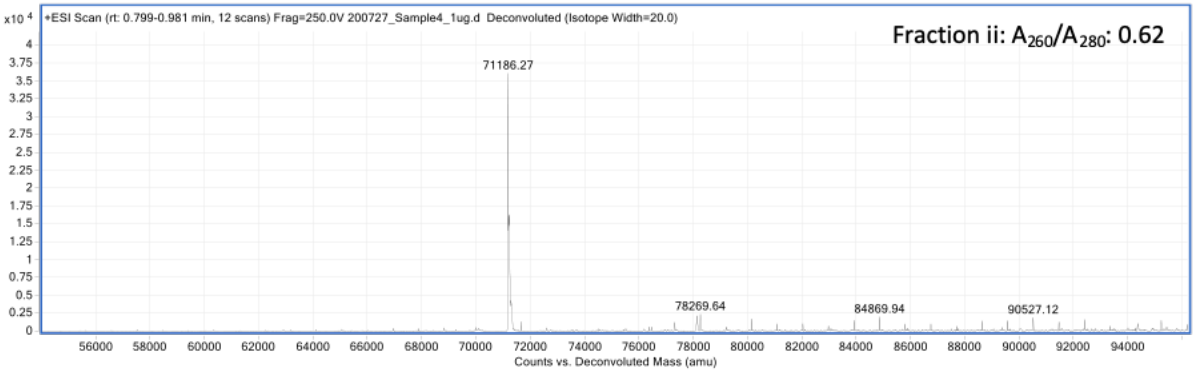

+ESI Scan (rt: 0.799-0.981 min, 12 scans) Frag=250.0V 200727_Sample4_1ug.d Deconvoluted (Isotope Width=20.0)
Fraction ii: A260/A280: 0.62
71186.27
78269.64
84869.94
90527.12
Counts vs. Deconvoluted Mass (amu)

Figure 3.10 | Gradient elution of heparin column chromatography following HisTrap purification. A) Elution profile of the purification. B) SDS-PAGE analysis of the elution fractions. Residual contamination bands can be seen in fractions of the second peak. Fractions i and ii were sent for mass spectrometry analysis. C) Graph of A_{260}/A_{280} and protein concentration of each fraction of the two peaks. The protein concentration is determined based on the A_{280} of the fraction and the molar extinction coefficient from ExPaSY. Red dashed line represents an A_{260}/A_{280} of 0.57, indicating the solution contains 100% protein and 0% nucleic acids. Blue shaded region indicate fractions collected for biophysical analysis. D) ESI-MS of fraction i which has a A_{260}/A_{280} of 0.57. E) ESI-MS of fraction ii which has a A_{260}/A_{280} of 0.62. Both are 128 Da off the expected mass of 71314, corresponding of the cleavage of the N-terminal methionine.

The absorbance elution profile showed two sharp peaks of equal height. SDS-PAGE analysis revealed that these both contain HaloCREB. When the fractions were analysed by the spectrophotometer, fractions in the first peak had a A_{260}/A_{280} of ≤ 0.57, whereas fractions in the second peak had A_{260}/A_{280} ranging from 0.59 to 0.67. A fraction from each elution peak was then sent for mass spectrometry analysis to confirm the purity. However, due to difference in protein ionisation efficiency, denaturing mass spectrometry is not necessarily the best to quantitate sample purity. A better method to quantify purity would be to perform liquid chromatography-mass spectrometry (LC-MS) and analyze its chromatogram.

Figure 3.10D and 3.10E shows that both proteins correspond to a mass 128Da less than the expected mass of 71314, indicating that the N-terminal methionine was cleaved sometime during expression or purification. No protein or other contaminants are now observable in the first peak, and therefore is of suitable purity for biophysical analysis. The case is not the same in the second sample with a A_{260}/A_{280} of 0.62. The spectrum contains contamination in higher molecular mass, regularly spaced by 943 Da. This protocol yields 1 mg of HaloCREB per 1 L of bacterial culture.

3.1.6 Development of CREB purification protocol

In order to establish any effects of the HaloTag upon CREB's properties, a construct without the HaloTag was generated. A HaloTag-affinity chromatography (HaloLink) was first used as this chromatography technique is highly specific and remove major contaminants from the purification when HaloTag is irreversibly bound to the resin upon elution. Following the manufacturer's instructions, I attempted to purify HaloCREB from cell lysate.

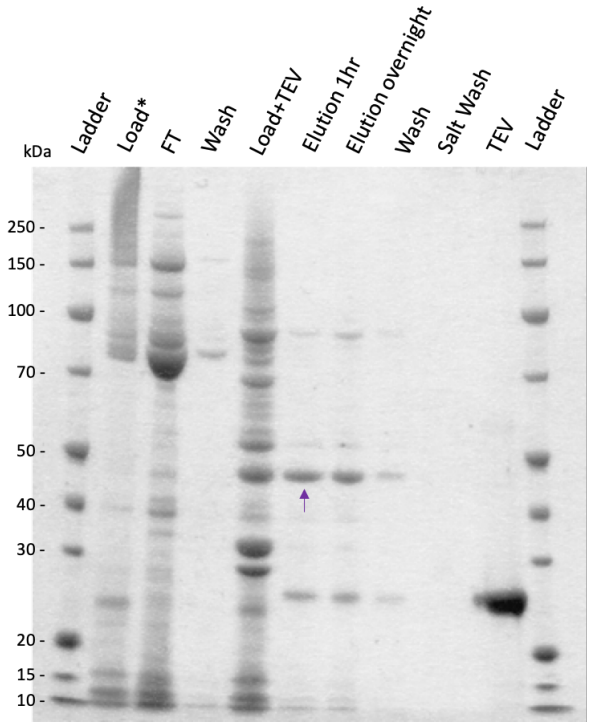


Figure 3.11 | HaloLink purification with cell lysate. Load sample was from another batch of purification. The lack of the CREB in the sample might be due to insufficient vortexing before collecting. Purple arrow corresponds to CREB (35 kDa), but migrating to a higher molecular weight in the same fashion as HaloCREB. The two elution fractions correspond to the amount of time the resin was incubated with Tobacco Etch Virus (TEV) protease.

Despite the improper treatment of the load fraction, a large band of HaloCREB was observed in the flowthrough. The low yield of the purification protocol was further reflected by the low amount of protein that was eluted, as the elution was only 1 mL in volume. A repeat was performed with an overnight binding step, but resulted the same yield. The yield would be further reduced as another step of purification was required to remove Tobacco Etch Virus (TEV) protease from the sample.

The initial test had disappointing yield that is insufficient for subsequent essential purification step. Furthermore, as HaloTag irreversibly binds to the resin, these resin could only be used once, and therefore are expensive to scale up. It did not appear viable as a long-term purification strategy, so I tried another strategy.

Another approach was to take relatively pure Halo-CREB and cleave the HaloTag with TEV protease, and then repurify by anion exchange chromatography (figure 3.12) to remove TEV protease and HaloTag.

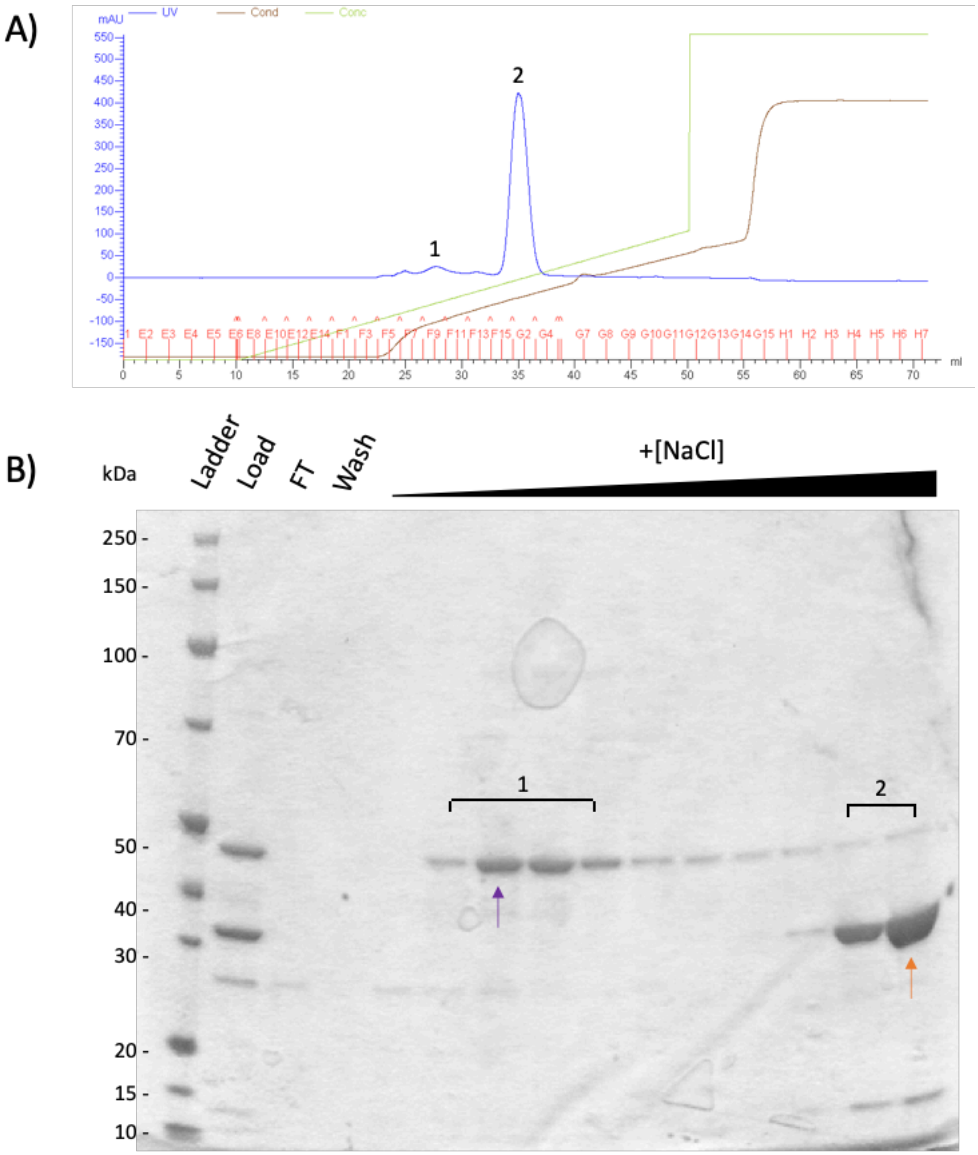


Figure 3.12 | Anion exchange chromatography following TEV cleavage of HaloCREB at pH 6.5. A) Elution profile of the purification, Two major peaks can be seen. B) SDS-PAGE of the purification. Fractions corresponding to the two peaks were labelled. Purple arrow indicates CREB and orange arrow indicates HaloTag.

The elution profile (figure 3.15A) showed two sets of peaks. The first set containing three peaks were of much lower absorbance compared to the single sharp second peak. SDS-PAGE analysis showed that the first peaks correspond to CREB and the second peak correspond largely to HaloTag. Although the absorbance is much higher in the HaloTag peak, the two are of similar concentration as there is a 6-fold difference in extinction coefficient (HaloTag: 59930 M^{-1} cm^{-1}, CREB: 10430 M^{-1} cm^{-1} estimated by ProtParam).

The CREB fractions showed minimal or no contamination by other proteins so the A_{260}/A_{280} of each fraction was measured. Since CREB does not contain tryptophan residues but only tyrosine, the expected value for a nucleic acid-free sample is around 0.45.

3.1.7 Cleavage of the polyhistidine tag

As the polyhistidine tag is at the C-terminus, where the basic region and leucine zipper can be found, I was concerned that the charged histidine residues might disrupt the interaction between HaloCREB and DNA. I tried to cleave the histidine tag using carboxypeptidase A test to see if the histidine tag can be cleaved off without non-specific cleavage of the protein (figure 3.13).

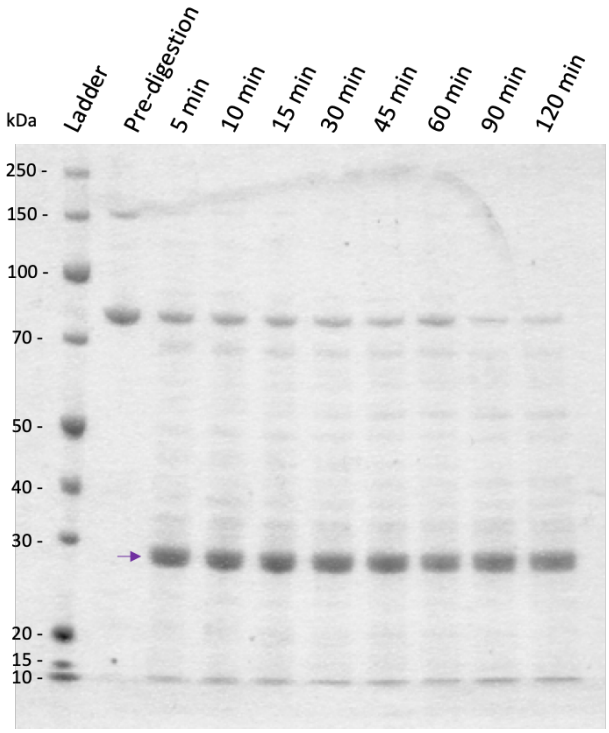


Figure 3.13 | Carboxypeptidase A digestion. Labels represent the amount of time the sample is incubated with carboxypeptidase A. Purple arrow indicates carboxypeptidase A. Significant amounts of nonspecific cleavage can be seen from 5 minutes.

Unfortunately, while the reaction seem to be improperly quenched by the addition of SDS sample buffer, significant non-specific cleavage can be seen in all samples. Even if the histidine tag can be cleaved off without cleavage in the rest of the protein, it would be very hard to purify it from the other cleaved contaminants. Hence, a site directed mutagenesis approach was taken instead to insert a cleavage site between CREB and the polyhistidine tag.

3.1.8 Introducing mutants for further analysis

Specific cleavage of the polyhistidine tag with carboxypeptidase was unsuccessful. Hence, a site directed mutagenesis approach was taken instead to insert a cleavage site between CREB and the polyhistidine tag. Other mutants of HaloCREB were also desired for experimental studies as follows:

1) Single cysteine mutant of CREB with C286S and C296S substitutions to enable maleimide labelling with a fluorophore for single-molecule imaging;

2) forced dimer (FD) of HaloCREB with C286S, C296S, C323S mutations and a GGC insert between CREB and the polyhistidine tag (HaloCREB-GGC_{ins}) to investigate the stoichiometry of the binding pathway;

3) insertion of a Factor Xa cleavage site (IEGR) between CREB and the polyhistidine tag.

Primers were designed with *E. coli* optimised codons for insertions to allow site directed mutagenesis. Sequencing of produced plasmids demonstrated these were successful.

3.1.9 Forced dimerisation of HaloCREB-GGC_{ins}

The HaloCREB-GGC_{ins} mutant only has one surface exposed cysteine and therefore can be oxidised in a high pH environment. The covalent disulphide bridge should be stable in solution as long as no reducing agent such as DTT or TCEP is present in the buffer. Figure 3.14 presents an SDS-PAGE gel of the reduced and non-reduced FD HaloCREB. TCEP was not added in the loading buffer for the FD HaloCREB, but the samples were boiled prior to loading. The gel was performed after several freeze-thaw cycles of the protein and therefore a small amount of degradation can be seen.

The presence of monomer HaloCREB in the oxidized lane showed that the pre-established protocol was ineffective in fully dimerizing HaloCREB. Furthermore, a smear in the higher molecular weight demonstrate that there might be nonspecific complex formation between the HaloTag and the bZIP.

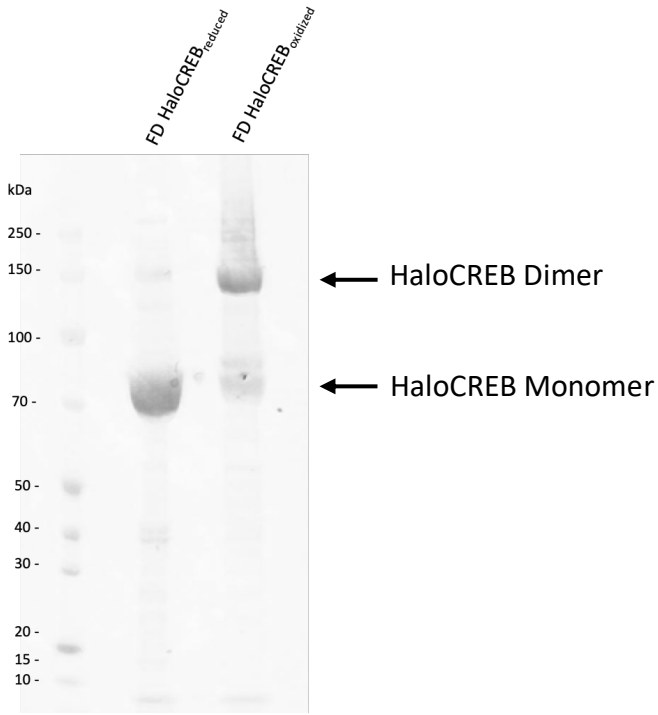


Figure 3.14 | Forced dimerization of HaloCREB-GGC$_{ins}$. Left lane shows a HaloCREB-GGC$_{ins}$ incubated with 10 mM DTT. A strong band at 70 kDa corresponds to the monomeric HaloCREB. Right lane is the forced dimerized HaloCREB-GGC$_{ins}$. A monomer band can still be seen and a smear can be seen in the higher molecular weight for HaloCREB.

3.2 Biophysical characterisation of HaloCREB

The interaction between CREB bZIP and DNA has already been characterised within the Shammas laboratory. Key parameters such as: the equilibrium and kinetic rate constants for binding to CRE DNA; binding affinity to nCRE DNA; and equilibrium and kinetic rate constants for homodimerisation were elucidated. The parameters suggested a tight binding of low nM affinity for both bZIP homodimerisation and CRE binding, generally supporting a dimer pathway for the DNA search.

As existing literature demonstrated the fusion of HaloTag does not affect the DNA binding and functionality of CREB[144,145,150], I set out to expand upon the findings of the Shammas laboratory through determining these parameters for HaloCREB. By comparing the data with the bZIP construct, I can investigate the impact of the remainder of the CREB sequence, outside the bZIP domain, on dimerization and DNA binding.

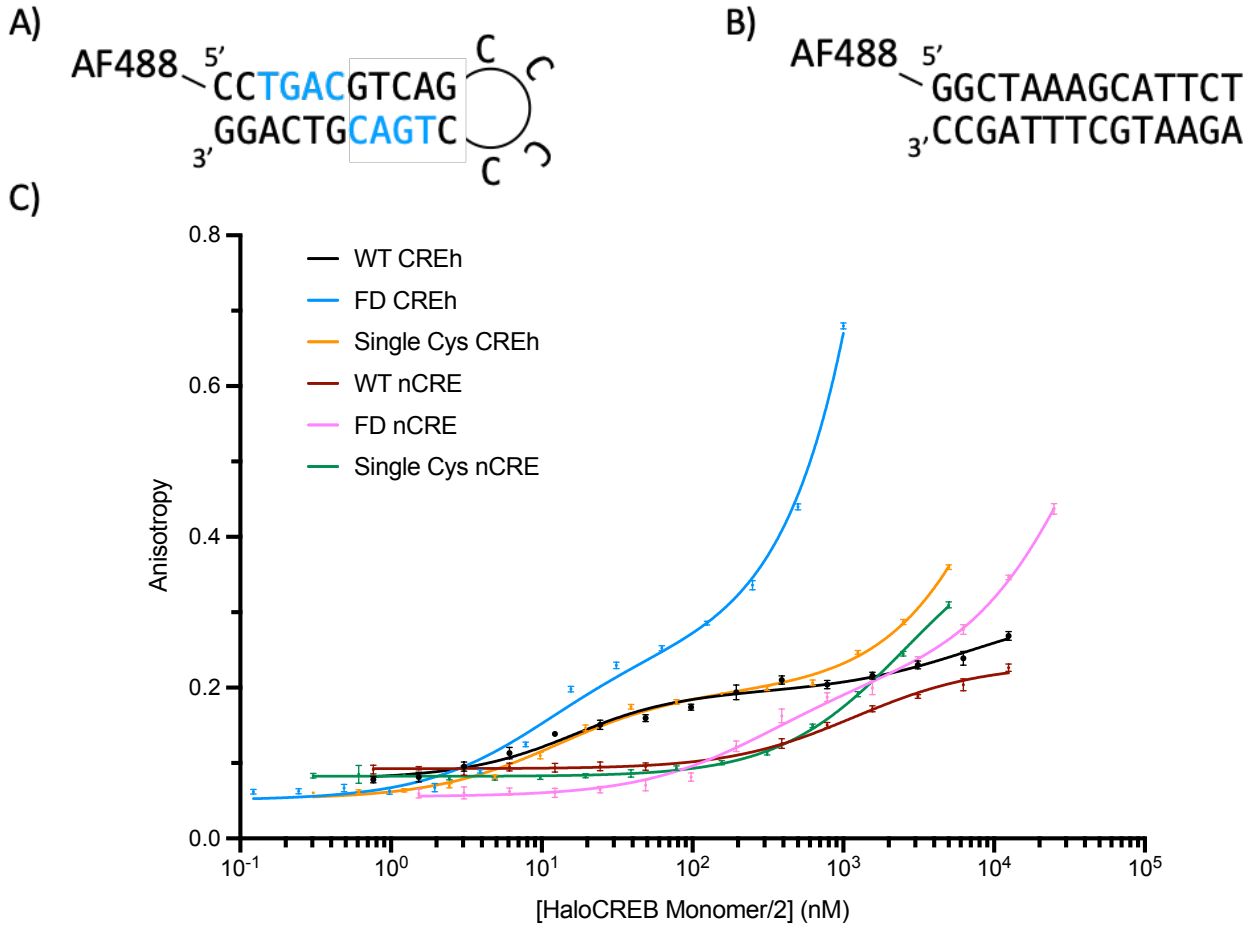
3.2.1 Experimental determination of HaloCREB extinction coefficient

To accurately define the protein concentration in biophysical assays, the extinction coefficient of HaloCREB was experimentally determined using the Gill von Hippel[138] method described in 2.6.2. Three repeats were done and the extinction coefficient was determined to be 71500 ± 700 M^{-1} cm^{-1}. The experimental extinction coefficient differs only 2% from that of predicted by ProtParam (70360 M^{-1} cm^{-1}).

3.2.2 HaloCREB binds specifically to CRE DNA

I first utilised a pre-established assay from the lab to determine whether my HaloCREB construct demonstrates the same specificity to CRE DNA. Using AlexaFluor®-488 labelled dsDNA oligonucleotides (figure 3.15a,b), I monitored binding of HaloCREB constructs to DNA using fluorescence anisotropy as a probe. Specific DNA binding was probed with AF488-hairpin CRE (CREh, fig 3.15A) while nonspecific DNA binding was probed with AF488-linear non CRE (nCRE, fig 3.15B). A linear blunt end oligonucleotide for nCRE titrations as it better matches the secondary structure of DNA *in vivo*, but hairpin DNA was used for CRE titrations to match the pre-established setup of the competitive stopped-flow experiment.

Anisotropy measurements for the oligonucleotides incubated in the presence of various HaloCREB concentrations were collected to form an equilibrium curve. Initial equilibrium curves were collected on an LS55 PerkinElmer instrument using 5 nM AF488-labelled DNA. However, as the first transition of the AF488-hairpin CRE (CREh) curves was reaching completion by around 5 nM, the assay was not suitable for determining the K_d for this reaction. Instead AlexaFluor®-488 labelled dsDNA at 1 nM was used in subsequent equilibrium curves, and measurements were performed on a more sensitive Horiba FluoroMax-4 fluorimeter.



	WT		FD		Single Cys	
	K_1	K_2	K_1	K_2	K_1	K_2
AF488-CREh (nM)	5 ± 4	3620	5 ± 4	3753727	3 ± 2	7544
AF488-nCRElin (nM)	1100 ± 300		140 ± 80	30210	2600 ± 400	

Figure 3.15 | HaloCREB equilibrium binding constant (K_d) determination to CREh and nCRElin by fluorescent anisotropy. a) Structure of the AF488-CREh DNA. Half CRE sites labelled in blue. b) Structure of the AF488-nCRElin DNA. c) Concentration-dependent change in anisotropy from DNA binding events. All CREh curves and FD nCRElin fitted with a two-state binding equation (Eq. 9) where WT and single cysteine mutant nCRElin curves are fitted with a one-state with dimerisation equation (Eq. 8) ($R^2 > 0.98$). Errors represent the standard deviation of the measurements.

Table 3.i | Equilibrium dissociation rate constants (K_d) extracted from figure 3.15c. K_d values were obtained by either fitted to a two-state binding reaction ($R^2 > 0.98$). All data are single repeat experiments. Errors were extracted from the upper and lower bound of the 95% confidence limit of the fit.

As expected, an increase in fluorescence anisotropy was observed as the concentration of HaloCREB increases for both AF488-hairpin CRE (CREh) and AF488-non CRE (nCRE) DNA, confirming HaloCREB's ability to bind to DNA and confirming the validity of my purification protocol.

The equilibrium curves show a clear distinction between binding affinity to nCRE and CREh amongst all three HaloCREB constructs, demonstrating HaloCREB's specificity to CRE DNA. A single-state transition was observed in WT and single cysteine mutant to nCRE, whereas a two state-transition can be seen for all three CREh titrations and the FD nCRE.

The single transition of nCRE for the WT, single cysteine mutant HaloCREB are at respectively 1100 ± 300 nM and 2600 ± 400 nM, corresponding to the K_d of the constructs to nCRE DNA. On the other hand, the two transitions for CREh titrations correspond to two binding events to CREh. The first transition ranges from 3 – 5 nM for all three constructs. The second transition does not reach a plateau so the value reported is greatly varied, from 3620 nM to 3753727 nM. Both FD HaloCREB titration curves increase to very high values and do not reach plateau, reflecting the possible aggregation of the sample seen in figure 3.14.

The three constructs that was characterised were: 1) wild-type (WT) HaloCREB, 2) forced dimer (FD) HaloCREB and 3) single cystine mutant HaloCREB. Measurements of the FD HaloCREB is essential as a control as it must bind DNA via the dimer pathway. The dimerisation of the single cysteine mutant is compromised as the two mutated cysteines are located in the bZIP domain. Characterisation of this mutant can also reveal insight to the importance of dimerisation in modulating HaloCREB's ability to bind specific and nonspecific DNA.

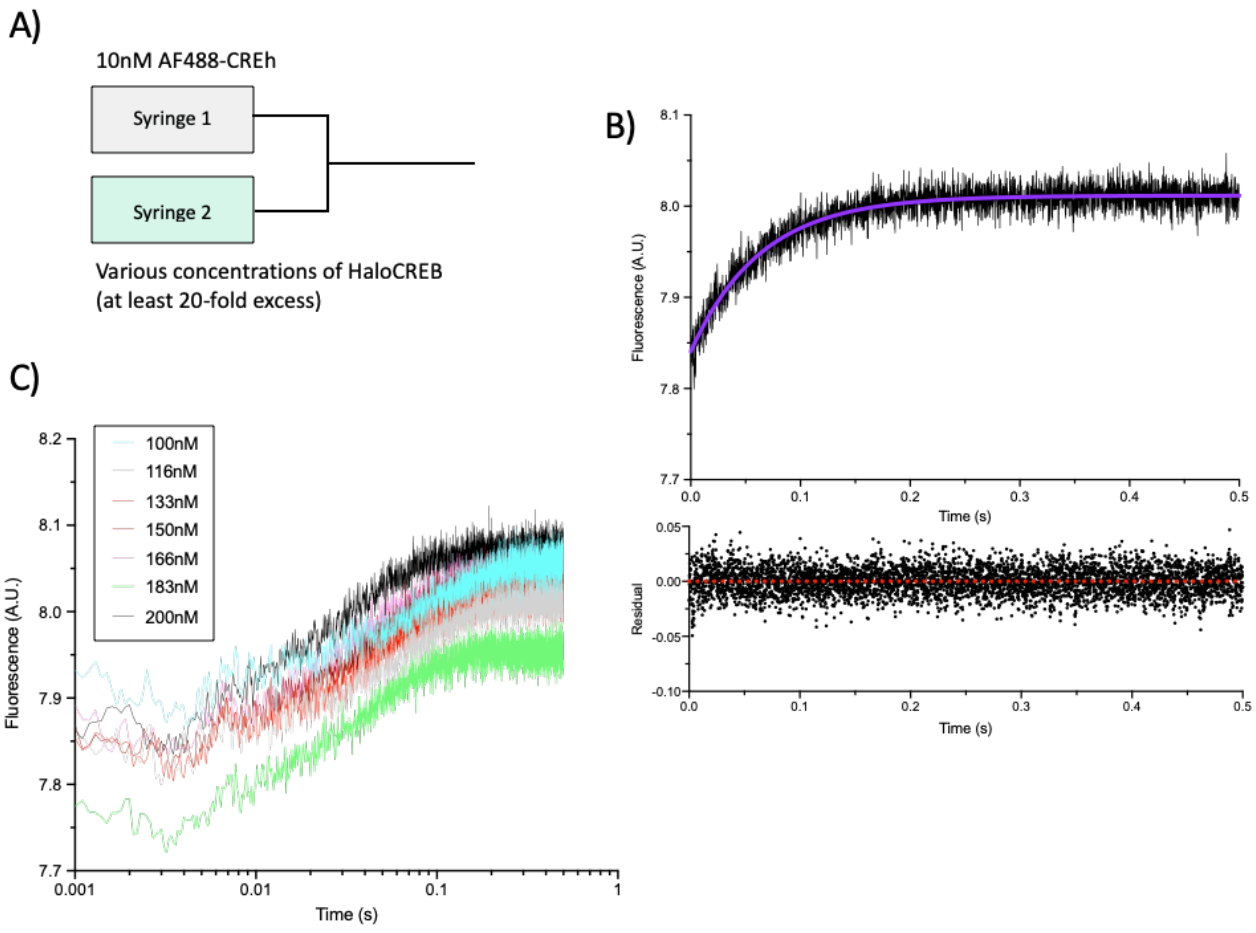
Measurements of the WT and single cysteine mutant were done in biophysical buffer [10 mM MES (pH 6.5), 150 mM NaCl, 10 mM MgCl2, 0.05% TWEEN] with 1 mM DTT, as the two constructs have respectively 3 and 1 free cysteines. On the other hand, measurements with the FD HaloCREB were done with no DTT as the two HaloCREB monomers are linked by a disulphide bridge. Inclusion of DTT in FD HaloCREB measurements would reduce the protein, rendering the results ineffective. As the two cysteines mutated in the single cysteine mutant are located in the bZIP, characterisation of the single cysteine mutant can reveal insight into the importance of dimerisation.

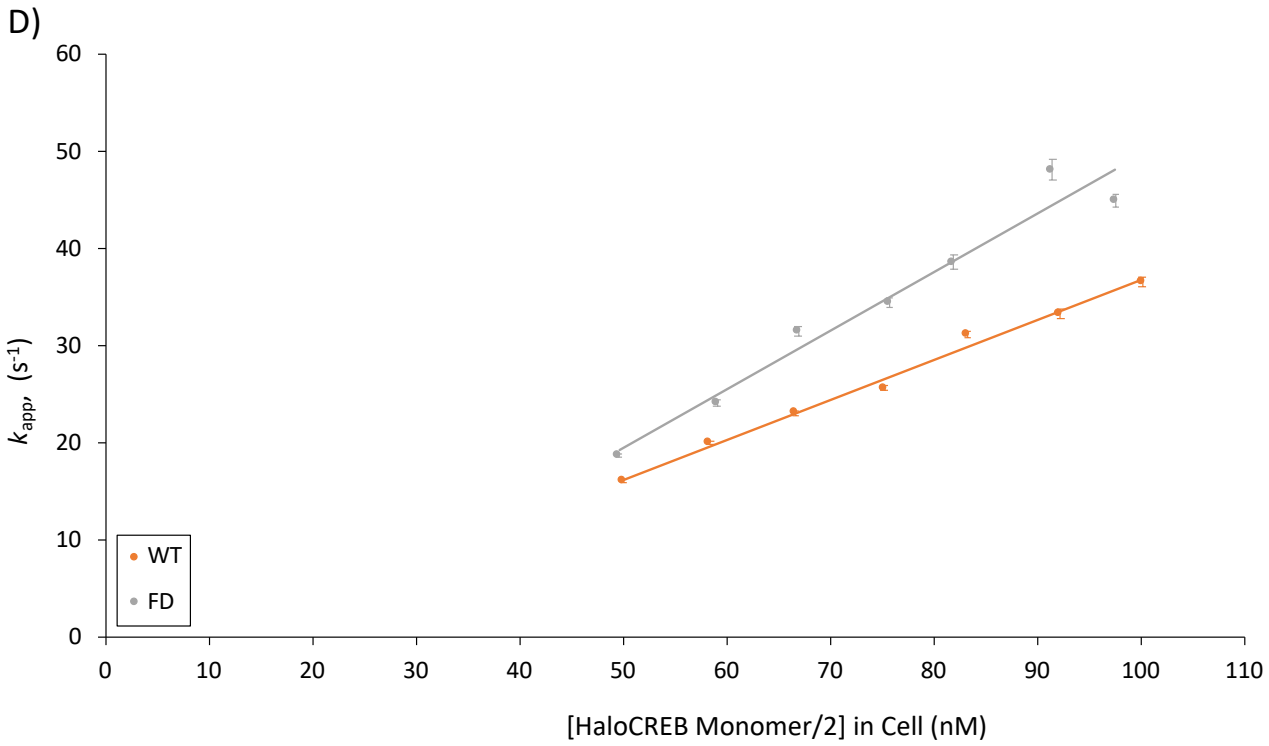
3.2.3 HaloCREB associates rapidly with CRE DNA *in vitro*

Next I performed stopped-flow fluorescent spectroscopy to determine the association rate of various mutants of HaloCREB to CRE DNA with pre-established lab protocols. Buffer blanks were performed to ensure there was no fluorescent change without a binding reaction. The three control shows that without mixing of the fluorescent CREh and HaloCREB, no change in fluorescence intensity was observed.

10 nM of AF488-hairpin CRE was rapidly mixed against various concentrations of WT and FD HaloCREB (figure 3.16). The concentration of HaloCREB in the reaction cell was kept at 20-fold excess over AF488-CREh in all reactions to maintain pseudo-first order condition for the dimer.

A single-phase exponential increase in fluorescence intensity was recorded for all of the experimental conditions. All kinetic traces were well fit with a single exponential decay function (Eq.10), and a single apparent rate constant extracted (k_{app}), which was linearly dependent on protein concentration ($R^2 > 0.95$). As pseudo-first-order conditions were maintained the slope of the linear fit can be interpreted as the association rate constant (k_{on}) of the construct to CRE DNA (table 3.ii).





	WT	FD
k_{on} (nM^{-1} s^{-1})	0.41 ± 0.02	0.60 ± 0.06
Y-intercept (s^{-1})	-4 ± 1	-10 ± 4

Figure 3.16 | Association kinetics of WT and FD HaloCREB with 5 nM of AF488-CREh. a) Schematic of the experimental setup. Solution of 10 nM of AF488-CREh was rapidly mixed with various concentrations of HaloCREB with a minimum of 20-fold excess. b) Example trace of the reaction with a collection time of 0.5 s. Purple curve was fitted to a single exponential ($R^2 > 0.93$). Residual graph shows the goodness of fit. The constant for the exponential was extracted as k_{app}. c) Kinetic traces of all WT HaloCREB measurements in a log scale. Legend shows the concentration of HaloCREB monomer in the mixing cell (2-fold less than the concentration of HaloCREB dimer in mixing cell). d) Concentration dependence of k_{app}. The data are fitted to a linear equation of y=mx+c ($R^2 > 0.95$). Error bars represent the fitting error to the exponential increase.

Table 3.ii | Parameters extracted from association stopped-flow. k_{on} and y-intercept for the two constructs are from the linear fit of k_{app} in figure 3.16d. Errors are from the fitting error to the linear regression.

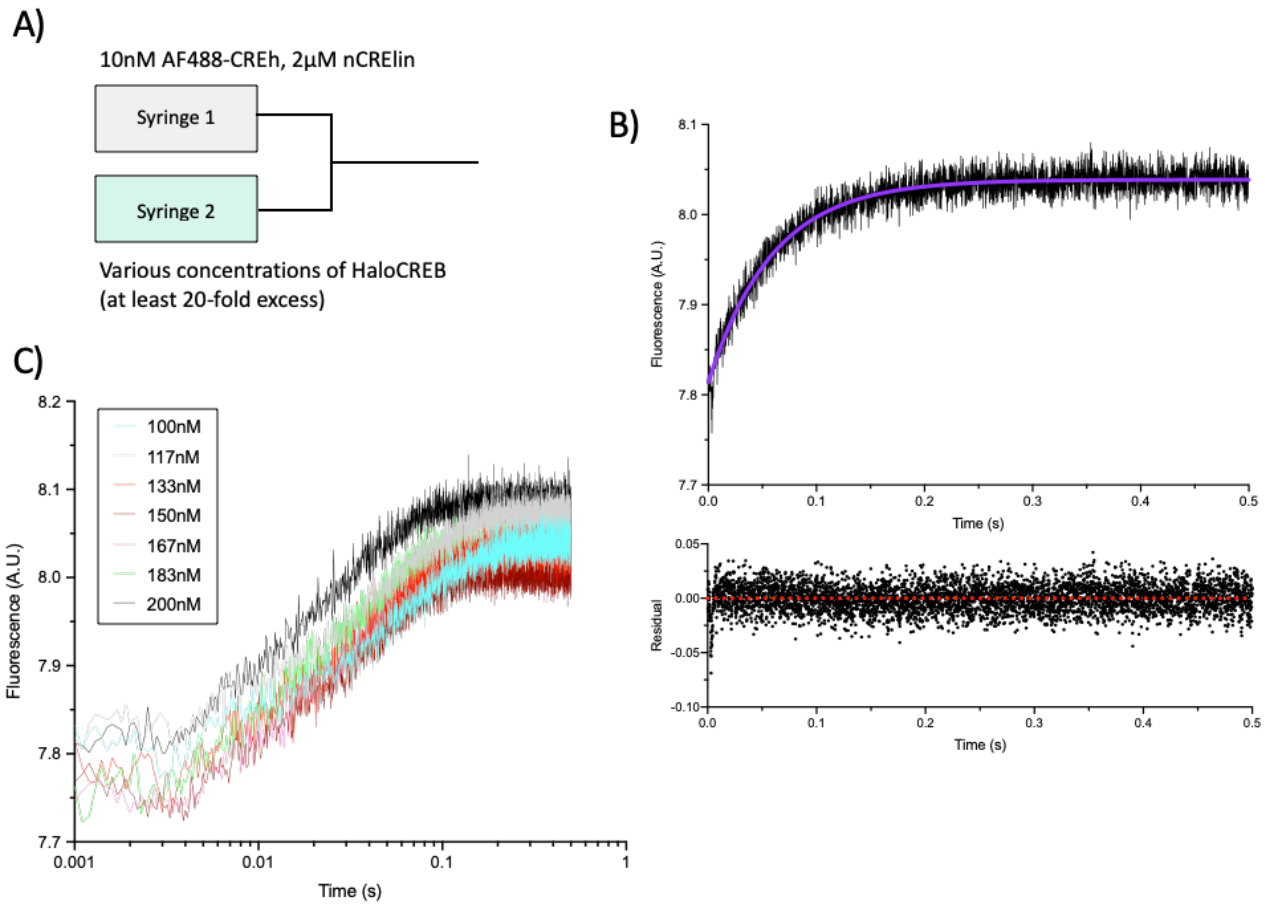
The association rate constant ($k_{on,CRE}$) for WT HaloCREB was determined to be 0.41 ± 0.02 nM^{-1} s^{-1}. Comparatively, forced dimer (FD) HaloCREB binds to CRE DNA 30% faster at 0.60 ± 0.06 nM^{-1} s^{-1}. The two constructs both rapidly bind to CRE DNA, beyond the commonly regarded threshold of 'diffusion-controlled' of 10^8 - 10^9 M^{-1} s^{-1}.

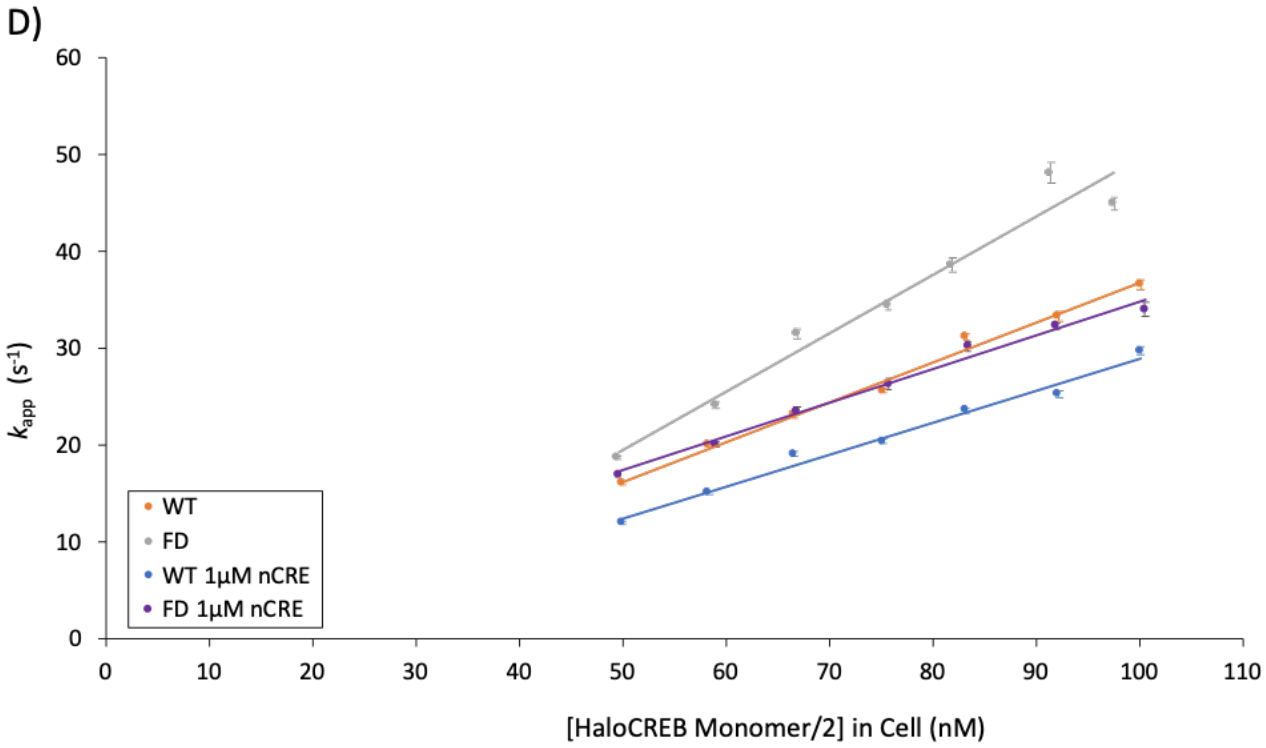
Y-intercepts of the linear fit indicate the dissociation of the of the complex from DNA. Both linear fits yielded a negative y-intercept value, respectively -4 ± 1 s^{-1} and -10 ± 4 s^{-1} for WT and FD HaloCREB so it was not possible to make a sensible estimate of the dissociation rate constant.

3.2.4 Competitor dependence of HaloCREB CRE DNA-binding

Next I investigated the impact of competitor DNA in the association rate constant to CRE DNA. As CRE sites *in vivo* are surrounded by a 30,000-fold excess of nCRE DNA[87,151], HaloCREB is expected to bind to nCRE sequences despite the 50 to 100-fold difference in affinity, resulting in a slowdown of association to CRE sites.

To measure the slowdown, the same stopped-flow experimental setup as determining the association rate constant was used, with the inclusion of 2 µM nCRElin DNA in the DNA syringe (figure 3.18). A single-phase exponential increase was also observed from the kinetic traces. The apparent rate constant (k_{app}) displayed the same linear dependence to WT and FD HaloCREB concentration as previously described. The associate rate constant (k_{on}) for the two HaloCREB constructs were therefore extracted from the slope of the linear fit.





	WT	FD	WT + 1µM	FD + 1µM
k_{on} (nM^{-1} s^{-1})	0.41 ± 0.02	0.60 ± 0.06	0.33 ± 0.02	0.34 ± 0.01
Y-intercept (s^{-1})	-4 ± 1	-10 ± 4	-4 ± 1	0 ± 1

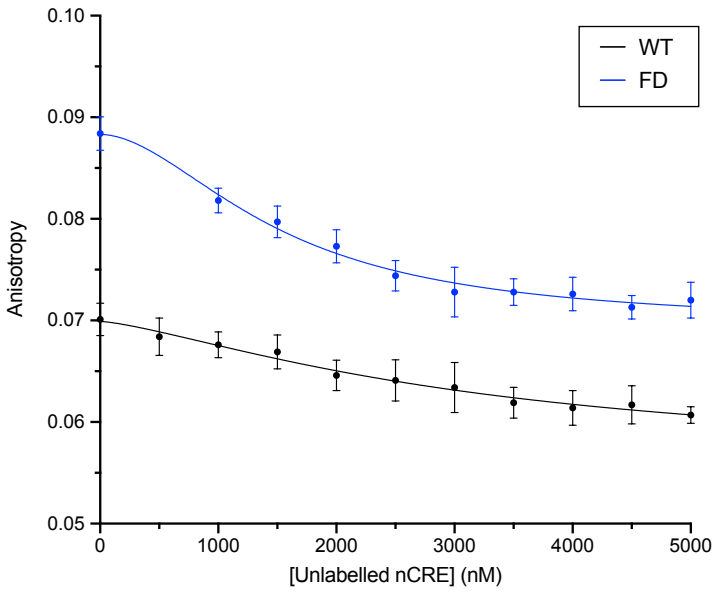
Figure 3.17 | Association kinetics of WT and FD HaloCREB with 5 nM of AF488-CREh in presence of 1 µM nCRElin. a) Schematic of the experimental setup. Solution of [10 nM of AF488-CREh, 2 µM nCRE] was rapidly mixed with various concentrations of HaloCREB with a minimum of 20-fold excess. b) Example trace of the reaction with a collection time of 0.5 s. Purple curve was fitted to a single exponential ($R^2 > 0.90$). Residual graph shows the goodness of fit. The constant for the exponential was extracted as k_{app}. c) Kinetic traces of all WT HaloCREB measurements in a log scale. Legend shows the concentration of HaloCREB monomer in the cell (2-fold less than the concentration of HaloCREB dimer in mixing cell). d) Concentration dependence of k_{app}. The data are fitted to a linear equation of y=mx+c ($R^2 > 0.95$). Error bars represent the fitting error to the exponential increase.

Table 3.iii | Parameters extracted from association stopped-flow in presence of non-specific DNA. k_{on} and y-intercept for the two constructs are from the linear fit of k_{app} in figure 3.16d. Errors are from the fitting error to the linear regression.

The association rate constant ($k_{on,CRE + 1µM nCRE}$) for WT HaloCREB in presence of 200-fold excess competitor DNA was calculated to be 0.33 ± 0.02 nM^{-1} s^{-1}, a 20% decrease from the rate constant without competitor DNA. The k_{on} for FD HaloCREB with nCRE was decreased by 43% to a similar value of 0.34 ± 0.01 nM^{-1} s^{-1}.

Interestingly, the y-intercept of the linear fit behaves differently between the two constructs. The linear fit for WT HaloCREB with nCRE DNA did not yield a change in y-intercept, whereas the fit for FD HaloCREB essentially intercepts the y-axis at the origin.

A slowdown in association rate constant indicate that both WT and FD HaloCREB bind to nCRE DNA with a given rate constant. Fluorescent anisotropy measurements in section 3.2.2 and table 3.i determined the equilibrium dissociation constant (K_d) to the AlexaFluor®-488 labelled nCRE DNA. A pre-established competition assay using fluorescence anisotropy was used to determine the equilibrium inhibition constant (K_i) of unlabelled nCRE DNA to the HaloTag-AF488-nCRElin complex.



	WT	FD
$K_{d,nCRE}$ (nM)	1100 ± 300	140 ± 80
$K_{i,nCRE}$ (nM)	2000 ± 1000	260 ± 90

Figure 3.18 and Table 3.iv | Determination of K_i unlabelled nCRE DNA to WT and FD HaloCREB. Fluorescent anisotropy curves were fitted according to Eq. 11 (WT R^2: 0.77, FD R^2: 0.91). $K_{i,nCRE}$ were then determined using the Kd,nCRE measured in figure 3.15 and the equations from Nikolovska-Coleska et al.[141] Errors bars are the standard deviation of the data points, whereas the errors for K_i are from fitting errors to the equation.

Fitting to the equations from Nikolovska-Coleska et al.[141] yielded a $K_{i,nCRE}$ of 2000 ± 1000 nM and 260 ± 90 nM for respectively WT and FD HaloCREB, both of which presents a 2-fold difference compared to the K_d to AF488-nCRE. The fit for WT HaloCREB is not particularly well as an anisotropy change of less than 0.1 is observed over the titration. A more pronounced change, from adjusting the saturation of HaloCREB-AF488nCRElin complex, would result in a larger signal change needed for a better estimation.

3.2.5 HaloCREB dissociates slowly from CRE DNA

I then moved on to investigate the dissociation rate constant (k_{off}) of WT and FD HaloCREB from CRE DNA. This was done by a pre-established stopped-flow competition mixing experiment. By rapidly mixing a solution of [10 nM AF488-CREh, 10 nM HaloCREB] with various concentration of CRElin DNA, the HaloCREB would be displaced from the AF488-CREh, resulting in a decrease in fluorescence signal (figure 3.19).

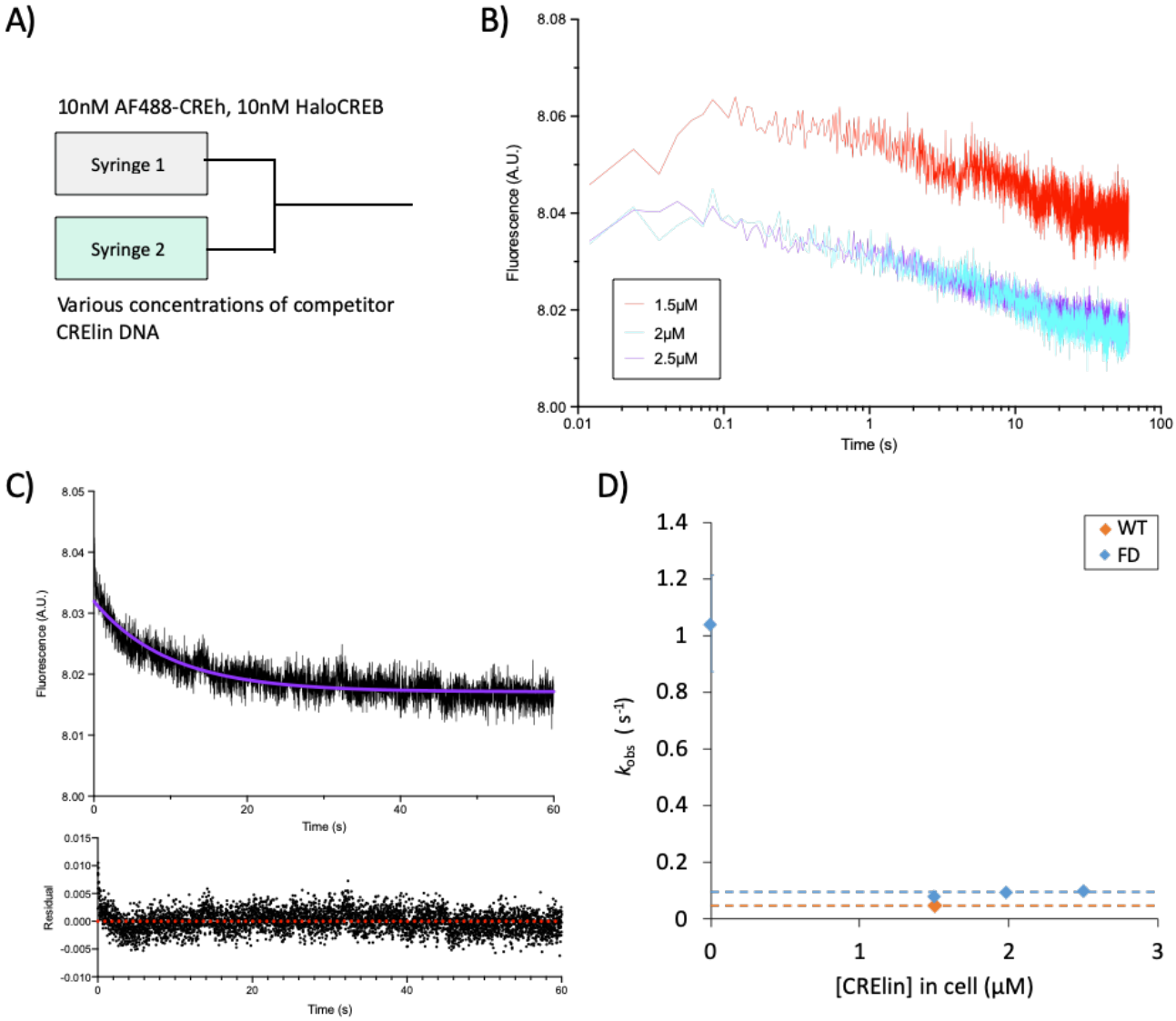


Figure 3.19 | Dissociation kinetics of WT and FD CREB by competition stopped-flow. a) Schematic of the experimental setup. [10 nM AF488-CREh, 10 nM HaloCREB] were mixed with various concentrations of unlabelled CRElin DNA. b) Kinetic traces of FD HaloCREB at the measured concentrations in a log scale. Legend shows the concentration of CRElin DNA in the mixing cell. c) Example trace with a collection time of 60 s. Purple curve was fitted to a single exponential decay ($R^2 > 0.78$). The constant for the exponential was extracted as k_{obs}. d) Concentration dependence of k_{obs}. An average of the data points was taken as the plateau. Error bars represent the fitting error to the exponential decay.

	WT	FD
k_{off} (s^{-1})	0.05 ± 0.001	0.094 ± 0.009

Table 3.v | Dissociation rate constant (k_{off}) for WT and FD HaloCREB inferred from figure 3.19. k_{off} for FD was estimated by the average of the individual data points. k_{obs} for WT at 1.5 μM CRElin was estimated to be the k_{off} for the construct. Error for FD is the error of the averaged samples, whereas error for the WT is the fitting error of the data point to the exponential decay.

A single-phase exponential decay was observed in the kinetic traces after a collection time of 60 seconds. The observed dissociation constant (k_{obs}) was be extracted from the exponential decay. An exponential relationship was expected between k_{obs} and HaloCREB concentration, in which the plateau of the decay represents the dissociation rate constant (k_{off}, s^{-1}). At low competitor concentrations, there would be incomplete dissociation of HaloCREB from the fluorescent CREh DNA and re-binding events would be observed. Hence, k_{on} contributes to the k_{obs} at low competitor concentration. The increase of competitor DNA eliminate re-binding events to the fluorescent probes upon dissociation, resulting in a scenario where k_{obs} = k_{off}.

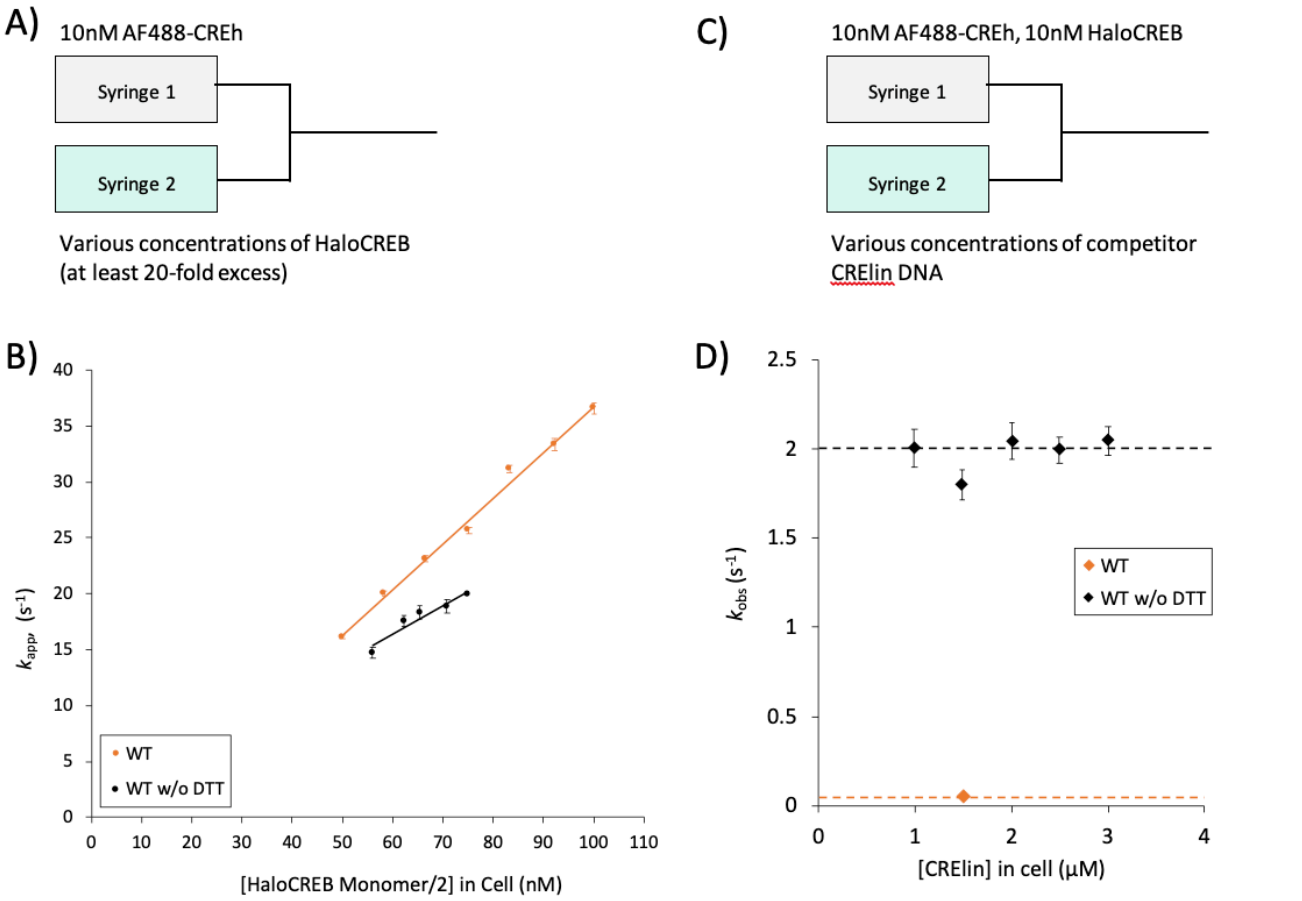
However, no obvious concentration dependence of k_{obs} was observed at the measured competitor DNA concentrations. The k_{off} values were therefore extracted with an average of the independent estimations. Furthermore, due to time and material constraint, only one data point was collected for WT HaloCREB. Since a plateau can be seen at the same CRElin concentration for FD HaloCREB, the k_{obs} is used as an estimate as the k_{off} for WT HaloCREB.

The dissociation rate constant (k_{off}) for WT HaloCREB was estimated at 0.050 ± 0.001 s^{-1}, twice as slow as the FD HaloCREB at 0.094 ± 0.009 s^{-1}. The two dissociation rate constants are both remarkably slow, suggesting that HaloCREB remains bound to CRE DNA for an extended period of time.

3.2.6 Non-reduced cysteine residues alter association and dissociation rate constants

As the WT HaloCREB contain 3 free cysteine residues in the bZIP domain, 1 mM DTT was kept in the working biophysical buffer to maintain a reducing environment and prevent unwanted oligomerisation by disulphide bridges. I have repeated the association and dissociation kinetic experiments with stopped-flow without DTT

in the working buffer to investigate the effects of non-reduced cysteine residues and observe any potential changes in the association and dissociation binding constants.



	WT	WT w/o DTT
k_{on} (nM^{-1} s^{-1})	0.41 ± 0.02	0.25 ± 0.04
Y-intercept (s^{-1})	-4 ± 1	1 ± 3
k_{off} (s^{-1})	0.050 ± 0.001	2.0 ± 0.1

Figure 3.20 | Association and dissociation of WT HaloCREB by stopped-flow in absence of DTT. A) Schematic of association stopped-flow experimental setup. B) HaloCREB concentration dependence of k_{app} from association kinetics. The data were fit with a linear fit ($R^2 > 0.9$), and the slope of the fits were inferred as k_{on}. C) Schematic of the dissociation competition stopped-flow experimental setup. D) Concentration dependence of k_{obs} from dissociation kinetics. An average of the individual measurements were taken as the estimate of the k_{off}. Error bars represent the fitting error to the exponential increase or decrease.

Table 3.vi | Comparison of kinetic parameters of WT HaloCREB with and without presence of DTT. Kinetic parameters for WT without DTT was calculated from the fits in figure 3.20b and 3.20d.

An exponential increase can be fit for the kinetic traces of the association experiments. A linear correlation was fitted to the concentration dependence of the k_{app} in the association experiments, and a k_{on} value of

0.25 ± 0.04 nM^{-1} s^{-1} was extracted. This value represents a 2-fold decrease compared to the measurements done with 1mM DTT. The difference can clearly be seen in figure 3.20c represented by the slope of the linear fits. A difference in the y-intercept was also observed. In a non-reducing environment, the y-intercept corresponds to 1 ± 3 s^{-1}. The increase of y-intercept to a positive value suggests that HaloCREB dissociates at a much quicker rate under non-reducing conditions.

This observation is complemented by the dissociation experiment. Similar to the previous experiments, an exponential decay was observed in the kinetic traces and there is no clear competitor DNA concentration dependence of k_{obs}. As such, the k_{off} was determined by an average of the individual measurements and was calculated to be 2.0 ± 0.1 s^{-1}. The values demonstrate that WT HaloCREB in a non-reducing condition dissociate from CRE DNA 40-fold quicker than that in a reducing condition.

3.2.7 CREh Binding Kinetics with Stopped-Flow

The association and dissociation rate constants (k_{on} and k_{off}) obtained from stopped-flow can be used to estimate an equilibrium dissociation constant ($K_{d,\ kinetics}$) of HaloCREB binding to CRE DNA. Assuming a two-state reaction, $K_{d,\ kinetics} = k_{off} / k_{on}$. The $K_{d,\ kinetics}$ for the various HaloCREB constructs can be seen in table 3.vii.

	WT	FD	WT 1μM nCRE	FD 1μM nCRE	WT w/o DTT
k_{on} (nM^{-1} s^{-1})	0.41 ± 0.02	0.60 ± 0.06	0.33 ± 0.02	0.34 ± 0.01	0.25 ± 0.04
k_{off} (s^{-1})	0.050 ± 0.001	0.094 ± 0.009	N.A.		2.0 ± 0.1
$K_{d,\ kinetics}$ (nM)	0.12 ± 0.01	0.17 ± 0.02			8 ± 0.05

Table 3.v | Biophysical parameters of WT and FD HaloCREB under various experimental conditions as determined by stopped-flow fluorescence spectroscopy. k_{on} of the constructs were determined by association kinetics, while k_{off} are determined by dissociation competition experiments. $K_{d,\ kinetics}$ was estimated by k_{off} / k_{on} assuming a two-state reaction. *The $K_{d,\ kinetics}$ corresponding to WT and FD with 1μM nCRE was calculated using the respective k_{off} of the construct.

3.2.8 Fluorescently Labelling HaloCREB

HaloCREB was fluorescently labelled with HaloTag-TMR ligand to measure the homodimerisation rate constant. Previous small-scale test (figure 3.9) demonstrated the fluorescent labelling protocol to be effective. As the manufacturer does not provide the extinction coefficient and correction factor at 280 nm (CF_{280}), measurements were made experimentally. Extinction coefficient of TMR at 548 nm from Grimm et

al.[137] was used to calculate the extinction coefficient and correction factor at 280 nm. The two experimental values were respectively 15255 M^{-1} cm^{-1} and 0.20.

The protein was then labelled with HaloTag-TMR by incubating 3 hours at room temperature and overnight at 4°C. The labelling efficiency was measured after 3 hours and after the overnight incubation. A 64% labelling efficiency was measured after both 3 hours and overnight incubation.

3.2.9 HaloCREB homodimerises with a 10^{-8} M K_d

I first tried to extract the equilibrium homodimerisation rate constant with fluorescent anisotropy by a serial dilution of TMR fluorescently tagged HaloCREB (HaloCREB-TMR) (figure 3.21). An exponential increase in anisotropy was expected as the homodimer is expected to tumble slower than the monomer HaloCREB, emitting a more polarised light upon excitation from a polarised light source.

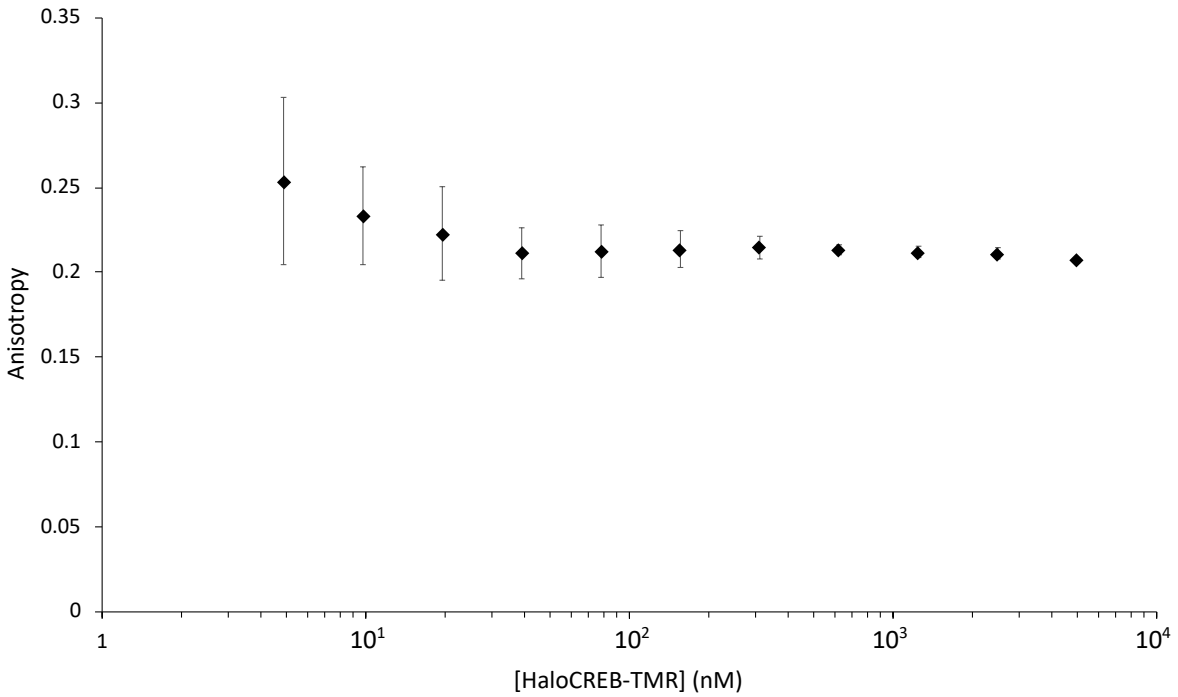


Figure 3.21 | Fluorescent anisotropy serial dilution with HaloCREB-TMR. Error bars from standard deviation of the data set. The large error bars are a result of low signal-to-noise ratio at low HaloCREB-TMR concentration.

No anisotropy change was observed in the serial dilution and therefore methodology was deemed ineffective. I then tried to use native mass spectrometry (N-MS) to determine the same equilibrium

homodimerisation constant. Unfortunately, native MS was also ineffective in determining the homodimerisation K_d as HaloCREB was found to aggregate in 1 M ammonium acetate and stuck to concentrating columns.

Mass photometry was used as an alternative method to investigate the homodimerisation tendencies of (figure 3.20). The aggregation of HaloCREB in 1 M ammonium acetate was later confirmed by mass photometry.

Mass photometry experiments[142] were performed at two separate protein concentrations. Two peaks were observed in both experiments. The first peak was observed at 58 ± 9 kDa and 61 ± 10 kDa, while the second peak was observed at 95 ± 12 kDa and 94 ± 22 kDa. The peaks were downshifted from the expected mass of 71 kDa and 142 kDa for monomer and dimer HaloCREB. The purity of the sample was previously confirmed with denaturing mass spectrometry.

Both monomer and dimer species were observed in both HaloCREB concentrations. There were very similar numbers of counts for both species, suggesting that the homodimer K_d is approximately of this range. The homodimerization K_d inferred from the mass proportion is in the range of tens of nM (10^{-8} M) in both cases (table 3.vi). Note that each dimer count corresponds to two CREB monomers, and thus the proportions by number and mass are different.

	22 nM WT HaloCREB	44 nM WT HaloCREB
Monomer counts	1094	984
Dimer counts	771 (1542)	885 (1770)
Total monomer count	2636	2754
% CREB in dimer	58%	64%
$K_{d,\ homodimerisation}$ (nM)	13 ± 1	17 ± 2

Table 3.vi | Calculation of WT HaloCREB $K_{d,\ homodimerisation}$ from mass photometry data in figure 3.20. Dimer count in parenthesis represent the number of monomers associated with the dimer count. $K_{d,\ homodimerisation}$ calculated from eq. 13 in section 2.6.10. Errors were propagated from the standard deviation of the peak counts in figure 3.20.

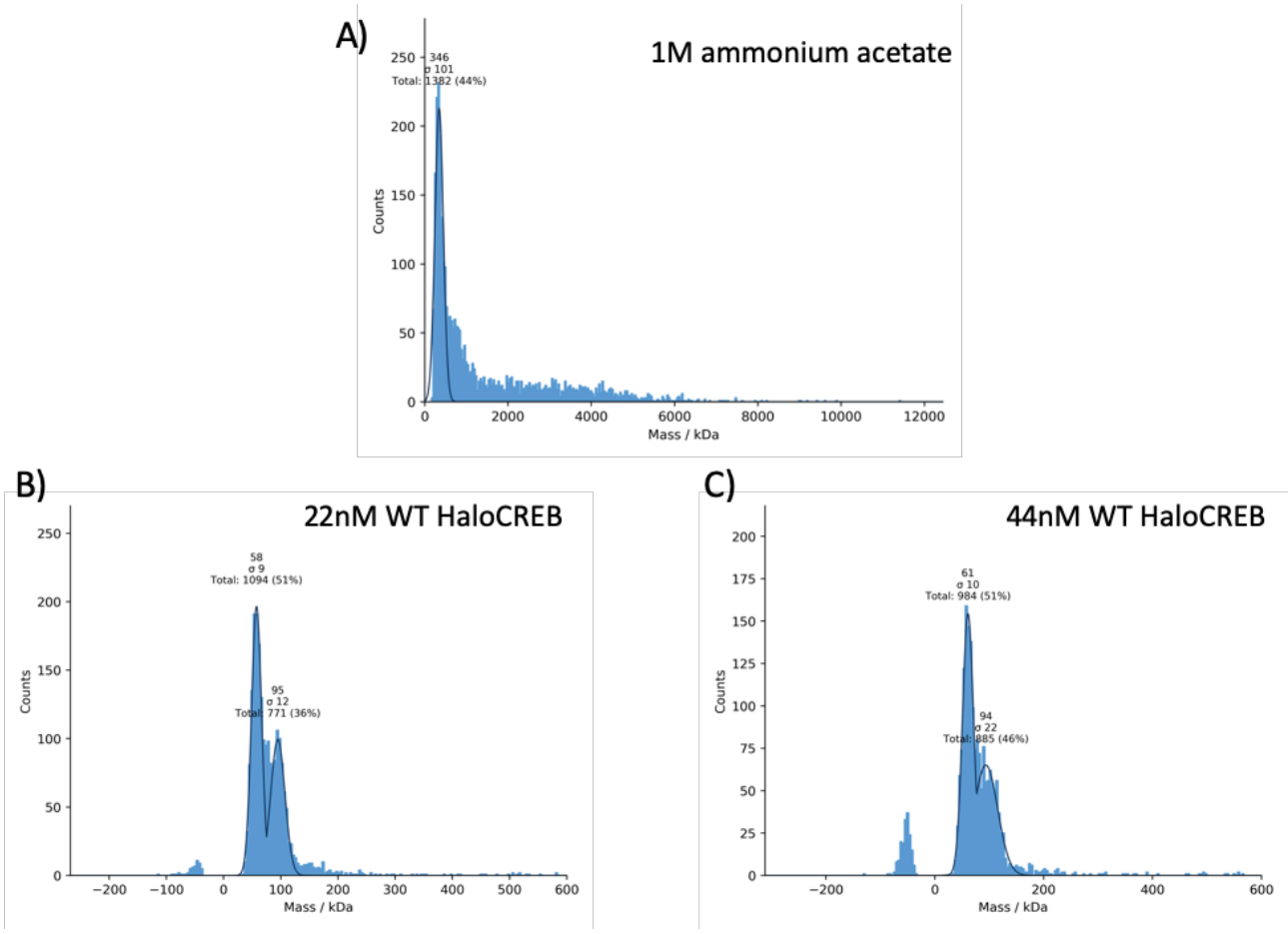


Figure 3.20 | Mass photometry of WT HaloCREB at various concentrations. a) WT HaloCREB in 1 M ammonium acetate. Significant aggregation can be seen in the sample. b) 22 nM WT HaloCREB. c) 44 nM WT HaloCREB. Counts of negative mass corresponds to unbinding events from the glass cover slip of the mass photometer.

4. Discussion

4.1 Purification of HaloCREB suitable for biophysical analysis

4.1.1 HaloCREB was successfully purified

This study establishes that a protocol of Ni-affinity chromatography followed by heparin chromatography was effective in the purification of HaloCREB from both protein and nucleic-acid contamination and to a suitable purity for biophysical analysis. The yield of 1 mg/L was reasonable. This is despite notorious difficulties in obtaining pure IDP samples such as nonspecific association with columns, and the method developed is novel.

Studies of HaloCREB in existing literature are either *in vivo*[144,150] or expressed with a cell-free expression system[145]. Biophysical methods, in particularly kinetics by stopped-flow, require a large amount of protein to perform measurements. A cell-free system would be too costly to scale up and therefore not applicable in this study. As such, a protein expression and purification methodology must be devised first to enable kinetic studies.

The purification of HaloCREB was expectedly tricky as it contains both a folded protein domain and an extended disordered domain. As such, common protocols to purify disordered proteins are ineffective for HaloCREB. On the other hand, methods to purify globular proteins are also reduced in effectiveness due to the extended disordered domain of CREB. In general, I avoided strategies that risked denaturing the HaloTag – without testing it is impossible to know whether this slowed progress unnecessarily, however I have generated pure samples without needing to denature the HaloTag.

4.1.2 Heat treatment is not an effective purification strategy for HaloCREB

Existing CREB purification protocols in literature implement a high temperature treatment varying between 65°C[146], 70°C[149], 75°C[148] and 100°C[147] (as described in section 2.4.2 and 3.1.2), exploiting the increased thermal stability of disordered proteins to precipitate globular proteins. However, treatment beyond 55°C

precipitated HaloCREB without significantly reducing background protein contamination. The potential of denaturing of HaloTag at 55°C deemed the heat treatment to be ineffective.

4.1.3 HaloCREB interacts with Sephadex columns

Size exclusion chromatography (SEC) using a Sephadex column was ineffective in purification as it appeared HaloCREB non-specifically interacts with the column and did not elute upon application of several column volumes worth of biophysical buffer. The ineffectiveness of size exclusion chromatography in purifying HaloCREB is particularly interesting. The disordered nature of IDPs gives them a larger hydrodynamic radii than that of globular proteins of the same size. As SEC columns separate protein by their hydrodynamic radii, the IDP would be expected to elute from the column faster than their corresponding mass, giving a better separation from contaminants of lower molecular mass[152]. The protein may have been nonspecifically interacting with the column beads. This was suggested as the protein was eluted from the column during the 1 M NaOH wash. A second attempt was done with a buffer with pH 11 to mimic the conditions of the wash, but the HaloCREB remained stuck to the column. SEC with a Superdex 200 HR 16/60 column under denaturing conditions (8 M urea) was used successfully in the CREB purification protocol from Lopez et al., though this construct did not have a HaloTag[146]. It is possible that SEC purification of HaloCREB might only be possible under denaturing conditions. However, as the HaloTag is a folded protein, purification under denaturing conditions would unfold the HaloTag. While experiments of HaloTag refolding has been detailed in existing literature[153], it is unknown to what degree this process affects the functionality of the HaloTag so I avoided the use of chaotropic agents and denaturing conditions with the aim of finding an effective alternative. If required in the future this option could be investigated.

4.1.4 Nickel-affinity is highly successful at removing protein contaminants

Fortunately, the Ni-affinity column was highly effective in isolating HaloCREB from the cell lysate. This step has been used in a recent purification strategy for CREB by Shnitkind et al.[74], and I found it to remove the vast majority of other components from Halo-CREB samples as well. Significant amounts of protein bound to

the column and was successfully eluted without aggregation or significant contamination. HaloCREB's affinity to metal-affinity columns is slightly low, as it only requires 0.25 M imidazole to fully elute all the bound HaloCREB. This might suggest that some HaloCREB may get removed during washing steps, however the reasonable yield achieved suggests this is not a major problem. This might be improved by increasing the length of the polyhistidine tag.

4.1.5 HaloCREB can be purified from other proteins using both anion and cation exchange chromatography

Intriguingly, HaloCREB bound to both anion and cation exchange columns at pH 6.5. With an isoelectric point of 5.5, HaloCREB is expected to be negatively charged at pH 6.5, and therefore binds to an anion exchange column. HaloCREB's affinity to bind to cation exchange columns can be attributed to the enrichment of positively charged residues amongst the protein, particularly in CREB where the residues are likely to be accessible by resin due to its disordered nature. The isoelectric point of a protein however does not describe the local charge distribution of a protein, especially in charge-enriched domains such as the two glutamine-rich domains and the basic region (of the bZIP) within CREB.

The initial test of the sequential application of the two ion exchange chromatography (IEC) columns resulted in an apparently pure fraction of HaloCREB by SDS-PAGE. Upon concentrating and overloading the SDS-PAGE gel, two additional bands can be seen at 33 kDa and 150 kDa. To identify the contaminant bands, the protein sample was incubated with HaloTag-TMR fluorescent ligand and visualised on an UV transilluminator. All three bands, including the 33 kDa and 150 kDa bands were fluorescent. This shows that at least some of the HaloTag is functional following the purification steps. The 150 kDa is likely a HaloCREB dimer band, and the 33kDa might be a HaloTag contaminant band. The presence of the cleaved HaloTag might be due to residual TEV in the laboratory communal IEC columns or FPLC system and can be fixed with rigorous system cleaning and the use of a new column.

Unfortunately, when the same protocol was repeated at a larger scale, the same purity was not achieved. Despite the sample being pure of protein contamination, the sample was not pure from nucleic acid as

reflected by the A_{260}/A_{280} ratio. The A_{260}/A_{280} ratio is a measurement of protein purity from nucleic acid. A ratio of 0.57 in tryptophan containing proteins are considered nucleic-acid pure. The few repeats of the same protocol at a larger scale yielded ratios of roughly 0.7. This was not ideal as any nucleic acid contamination will affect the results of downstream kinetic assays. An alternate purification method that can remove nucleic contamination was therefore devised.

4.1.6 Removal of nucleic acid contamination

The removal of nucleic acid contamination was the biggest hurdle in the optimisation of purification protocol for HaloCREB and CREB. The use of DNase and RNase is typically the first choice in removing nucleic acid contamination in the routine protein purifications, as such is the case in the study by Lopez et al.[146]. However, as I intend to investigate CREB-DNA interaction and DNA concentrations will be critical, trace amounts of DNase remaining in my protein sample might affect the assay results – such effects have been anecdotally reported for biophysical studies with other DNA-binding proteins. The current protocol involves two steps where DNase might be removed: purifications with his-trap and heparin columns. At least one of these is ineffective at removing DNase - DNase binds to heparin columns due to the DNA-like polyanion structure of the resin. Ni-affinity chromatography is then the only method amongst my protocol to remove DNase. The contaminating nucleic acid is likely to be bound to CREB and therefore at least partially protected from cleavage by DNase cleavage. With no guaranteed methods to remove the DNase, I decided against it as it is preferable to avoid use of nucleases in our lab environment.

Instead a 2 M NaCl wash was routinely done during the Ni-affinity chromatography purification. The wash was extended to overnight but it was still insufficient, yielding a eluate of A_{260}/A_{280} of 0.61. Although this figure indicates the vast majority of RNA and DNA has been removed (5% nucleic acid contamination gives a ratio of 1.06), I aimed for an entirely nucleic-acid free sample with a ratio of 0.57. Following this fairly robust purification with a round of heparin affinity chromatography was ultimately effective in removing nucleic acid contamination. As the polyanionic structure of the resin mimics DNA, it was able to displace the nucleic acid still bound to CREB. However, as a clear band of CREB remains in the flowthrough, it is likely that the

heparin resin could not displace the more tightly bound nucleic acid, e.g.. *E.coli* bacterial DNA containing the CRE sequence. This is demonstrated by the two elution peaks, first of which contains CREB without nucleic acids, but the second of which does.

ESI-MS revealed the first peak was pure of protein contamination and suitable for biophysical analysis. Curiously, contamination of regularly spaced higher molecular mass was seen in fractions of the second peak. Moreover, the regular spacing corresponds to 943Da each, with significantly increased peak size from the 7th iteration. A mass of 943Da is 20Da away from the average mass of a trinucleotide (307.9 Da), and perfectly corresponds to a trinucleotide of thymine (303.2 Da), adenine (312.2 Da) and guanine (328.2 Da)[154] i.e. TGA. Suggestively CRE sites have the consensus motif TGACGTCA, with half-site CRE having the sequence TGA(C). This is highly indicative of nucleic acid oligomers bound to CREB and not dissociated from the mass spectrometry process. This suggests that three nucleotides are the minimum required for CREB binding.

As electrospray ionisation is not a hard ionisation source, it does not cause fragmentation and has been used to observe protein-DNA noncovalent complexes[155–157]. Dissociation constants (K_d) in the range of 10^{-6} to 10^{-9} M have been measured using ESI-MS[158,159]. This falls into the range of the dissociation constants that were measured for HaloCREB and bZIP, and therefore the observation of HaloCREB-DNA complex would be in line with existing literature.

Therefore, this study establishes that a protocol of Ni-affinity chromatography followed by heparin chromatography was effective in the purification of HaloCREB from nucleic-acid contamination and to a purity for biophysical analysis.

4.1.7 Purification of CREB from HaloCREB

I have also been able to produce CREB from Halo-CREB. Nucleic-acid free HaloCREB (after heparin chromatography) was cleaved, and CREB isolated using ion exchange. Both ion exchange chromatography protocols from my first tests and from Shnitkind et al.[74] were effective at separating CREB from both HaloTag and TEV protease, and yielded the same purity. This means that CREB is also able to bind to both cation and

anion exchange columns within a range of pH of 1 without crossing its isoelectric point, consistent with my finding in HaloCREB.

Initially I had tried to purify using the HaloLink resin from Promega. The resin contains immobilised HaloTag ligand. HaloTag would irreversibly bind to the resin, and the bound protein could be released by cleaving the HaloTag. However, despite following the manufacturer's protocol, the yield of the purification was much lower than expected. As the resin is only single use, I inferred that the scale up of this protocol would be costly. On reflection it would have been better to try this strategy after an initial step of nickel-affinity chromatography, however since an effective and less costly solution was developed it did not seem worth further investigation.

4.1.8 Forced dimerisation of HaloCREB

The pre-established protocol of the forced dimerisation of CREB was ineffective in fully dimerising HaloCREB. An estimate from the SDS-gel picture indicate a monomer:dimer population of roughly 25% and 75% can be seen. However, the FD is known to be rather unstable and must be freshly oxidise the night before for experiments.

Unfortunately, although experiments were performed with freshly oxidised version of HaloCREB, a sample was not taken immediately for SDS-PAGE analysis. HaloCREB is already particularly prone to degradation, and such is reflected in the noisy background in both the reduced and oxidised version of HaloCREB. Significant degradation was observed upon multiple freeze-thaw cycles. Furthermore, the analysed sample was already previously forced dimerised, and might not prefer to do so again.

Regardless, while the SDS-PAGE presented might not accurately describe the purity of the sample used in the experiments, it is clear that this step must be further optimised to yield a complete dimerisation of HaloCREB. A higher initial HaloCREB concentration, incubation pH and longer incubation time are all methods to possibly obtain a larger shift in the monomer:dimer equilibrium.

4.2 Biophysical characterisation of HaloCREB *in vitro*

This study is the first to biophysically characterise the DNA-binding properties of HaloCREB *in vitro*. The parameters estimated include: 1) the equilibrium dissociation constants (K_d) to CRE and non-target nCRE DNA; 2) the association rate constant (k_{on}) to CRE DNA; 3) the dissociation rate constant from CRE DNA and; 4) the equilibrium homodimerisation constant ($K_{d,\ homodimerisation}$) of WT HaloCREB. These parameters were deemed essential in understanding the DNA-binding pathway of CREB through studying and modelling of the CREB bZIP domain by the Shammas laboratory. In vitro experiments performed in Shammas lab with CREB bZIP are well described by a dimer pathway, with dimerization occurring prior to CRE binding, which is different from that of current literature[160–162]. Through biophysical characterisation of HaloCREB, I wish to probe the impact of the rest of the CREB protein on these parameters, and on the DNA search pathway.

4.2.1 Experimental conditions

The characterisation with fluorescence anisotropy and stopped-flow kinetics were performed in the biophysical buffer composing of [10 mM MES (pH 6.5), 150 mM NaCl, 10 mM $MgCl_2$, 0.05% TWEEN20]. The buffer composition was chosen to match that conditions that the CREB bZIP was characterised in to enable direct comparison. This pH (6.5) differs from the intracellular and nuclear pH of 7.2[163]. While it is understood that fluctuations in pH, even at a minute scale can influence binding affinities and properties of proteins, a pH of 6.5 allow the biophysical characterisation to match the conditions of CREB bZIP NMR studies[164]. Significant peak dispersion was observed at pH 7.2, which lead to unresolved and unassignable peaks in ^{1}H and ^{13}C NMR spectra. Peaks became more observable at lower pH, and hence the biophysical assays were performed at the same lowered pH to match experimental conditions. Preliminary investigations showed the pH did not significantly alter binding affinity for the bZIP (data not shown). The expected charge of CREB (and HaloCREB) changes only moderately from -8 (-27) to -7 (-20) upon changing pH from 7.2 to 6.5.

The ionic strength of the buffer is a crucial choice as ionic interaction is a major contributing factor in HaloCREB-DNA interaction. The cell nucleus, particularly near genomic DNA regions, is known to have a NaCl concentration of 150mM, significantly higher than the rest of the cell[165,166]. Since CREB sublocalizes to the

nucleus and tends to be in close proximity to genomic DNA, a high ionic concentration was used to match *in vivo* conditions. We chose an equivalent NaCl concentration for the biophysical buffer, but it is worth noting that all charges contribute to ionic strength and so it is not entirely clear what the ionic strength actually is inside the nucleus where DNA is present at very high and non-homogenous concentrations.

The inclusion of 10 mM Mg^{2+} modulates CREB's DNA binding affinity.[74,91,92] While the crystal structures of CREB -DNA complex revealed the importance of Mg^{2+} in mediating CREB-DNA interaction[43], how it affects CREB DNA binding and specificity is of debate in current literature[74,91,92]. As the nuclear concentration of Mg^{2+} was determined to be 10 mM[167], the same concentration was used for my biophysical assays, however in the future it would be interesting to see how the various parameters estimated in the project depend upon the Mg^{2+} concentration.

Lastly the use of 0.05% TWEEN20 prevents nonspecific association of HaloCREB and DNA to eppendorf tubes or other containers. This is particularly important as the assays involve very low concentrations of both HaloCREB and DNA. Any nonspecific association will significantly alter the actual concentration and hence yielding inaccurate results.

4.2.2 Equilibrium homodimerisation constant ($K_{d,\ homodimerisation}$) of WT HaloCREB

The equilibrium homodimerisation constant ($K_{d,\ homodimerisation}$) of WT HaloCREB was estimated to be in the range of 10^{-8} M using mass photometry. The use of mass photometry to characterise WT HaloCREB is advantageous over native mass spectrometry as it can be performed in the native buffer. However, as the technique infers mass from the scattering properties of the protein, it is likely that an intrinsically disordered protein (IDP) might scatter light differently in solution, resulting in the observed downshifted mass of the monomeric and dimeric HaloCREB. The monomer:dimer ratios from two experiments yielded a similar estimated $K_{d,\ homodimerization}$ (13 ± 1 nM at 22 nM HaloCREB, 17 ± 2 nM at 44 nM HaloCREB).

However, PBS at pH 7.4 was used to dilute the protein as the biophysical buffer yielded unsatisfactory signal-to-noise ratio. This means the values that were obtained better approximates the cellular environment but

is not representative in the experimental conditions of my other biophysical assays. Furthermore, as mass photometry relies on nonspecific binding of protein to a glass cover slip, the use of detergents such as TWEEN would eliminate this interaction and hence must be avoided. This means that the protein concentrations might not be as accurate in lower ranges, introducing a source of error in the experiment.

The serial dilution with fluorescent intensity or anisotropy using HaloCREB-TMR was ineffective in estimating $K_{d,\ homodimerisation}$ as there was no observed change in anisotropy even over this concentration range. It is likely that the fluorophore is positioned too far away from the dimerisation domain as ~280 disordered residues separate the N-terminal HaloTag from the C-terminal bZIP domain. Even if HaloCREB was dimerised, the tumbling of the HaloTag may not be slowed down significantly if the long disordered linker already leads to significant opportunity for movement irrespective of the size of the complex it is in. This also suggests that further studies, particularly single molecule microscopy, might experience difficulties using HaloTag-TMR as a fluorophore to investigate or quantify N-terminally Halo-tagged CREB-DNA interactions. It may be better to position any label closer to the bZIP domain at the C-terminus of the protein.

The reported $K_{d,\ homodimerisation}$ amongst existing literature varies greatly, from the sub-nanomolar affinity reported by Lundblad et al.[149], to an estimation in the tens of µM range from Wu et al.[162] and Santiago-Rivera et al.[161] (Figure 4.1). My estimation of $K_{d,\ homodimerisation}$ (13-17 nM) is on the lower end of the spectrum, and is on a similar order of magnitude as the CREB bZIP (9 ± 3 nM) determined by the Shammas laboratory under identical buffer conditions but via a different experimental technique (urea refolding kinetics).

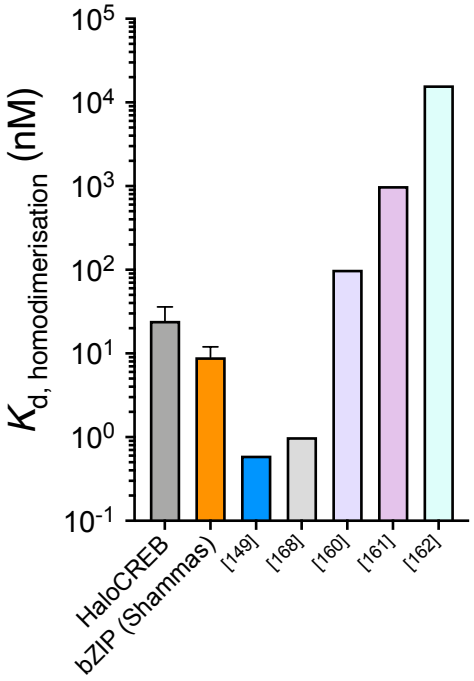


Figure 4.1 | Comparison of the equilibrium homodimerisation constant ($K_{d,\ homodimerisation}$) from various sources. Value for HaloCREB corresponds to the average of the two estimations presented in table 3.vi. K_d of bZIP collected by Conor Kelly (Shammas Lab, unpublished data). K_d of CREB rom varying existing literature[149,160–162,168].

Unlike GFP[169], there have been no reports of HaloTag oligomerisation despite its wide usage as a functional tag. It is therefore unlikely that HaloTag is contributing to the HaloCREB dimerisation affinity. The polyhistidine tag in the C-terminus might affect dimerisation but the histidines are mostly uncharged at pH 6.5, reducing the possibility of unfavourable charge repulsion leading to the destabilisation of the homodimer. Furthermore it is compelling that the bZIP and HaloCREB values are so similar.

As CREB has so many charged residues in its dimerisation domain, small changes in the pH of the buffer have the potential to cause a large change in homodimerisation affinity from specific charge-charge repulsions. Most of the values in figure 4.1 were obtained in buffers of pH 7.4-7.5, with the exception of [168] and the estimations from the Shammas lab, measured in pH of respectively 7.9 and 6.5. Examination of the HaloCREB sequence suggests that there is a change of net charge of 8 between pH 6.5 and 7.5 (Figure 4.2), due to ionisations of the 19 histidine residues.

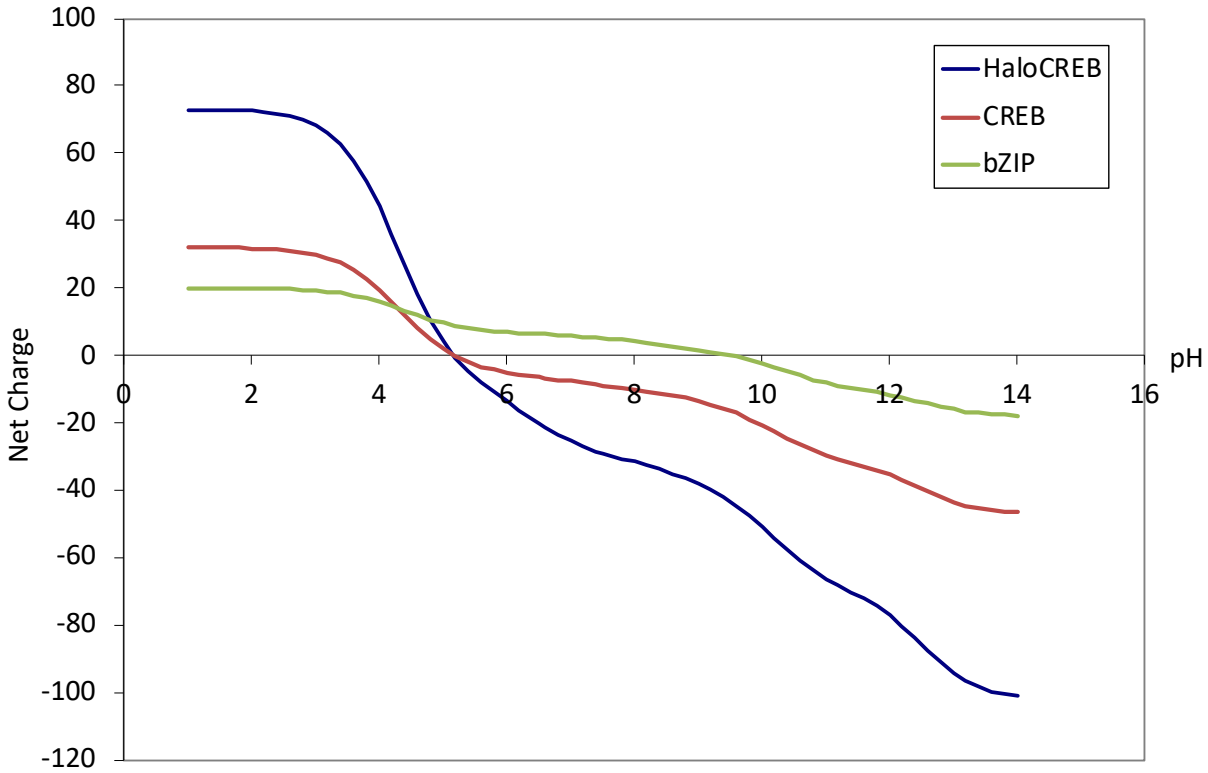


Figure 4.2 | Effect of pH in net charge of HaloCREB, CREB and bZIP. X-intercept of the curves represent the isoelectric point (pI) of the protein. Charge at different pH calculated using tool by Gale Rhodes of University of Southern Maine (from https://spdbv.vital-it.ch/TheMolecularLevel/Goodies/PeptChg.xls).

The isoelectric point of the bZIP domain is significantly higher than that of the CREB and HaloCREB construct (9.5 vs. 5.5 vs. 5.8). However, the HaloCREB is significantly more charged at both pH 6.5 and 7.5 due to the enrichment of histidine residues.

The various measurements also differ in the ionic strength of the buffer. Notably, the two 10^{-9} M range values were measured in buffers of lower ionic strength (50 mM NaCl or KCl)[149,168], whereas the values in the 10^{-6} to 10^{-8} M range were recorded in buffers of higher ionic strength much closer to physiological conditions (100-150 mM NaCl or KCl)[160–162]. DNA and CREB bZIP are both heavily and oppositely charged so salt concentration is expected to have a very significant effect on association rates through shielding of long-range electrostatic attraction. For example, association rates for barnase and barstar (highly oppositely charged surfaces) are accelerated around 4 orders of magnitude by electrostatic forces[170]. Nevertheless, the values in the literature have varied by more than two orders of magnitude even for those collected under “physiological conditions”.

It is also worth noting that nearly each measurement uses a different assay with the exception of circular dichroism for both [160] and [162].

As the basic region of the CREB monomers diverge from each other in the coiled-coil formation, even residues immediately before the basic region are not expected to be in close proximity with each other. Therefore, it is not likely the rest of the CREB protein will modulate the CREB homodimer dissociation rate constant. However, association rate constants might be altered by several features including different net charges and protein sizes, and competing (auto-inhibition) interactions between the basic region and other areas of the protein such as that reported recently by Shnitkind et al.[74] upon phosphorylation of the CK cassette. Nonetheless the similar order of magnitude between the HaloCREB and bZIP values reported here suggests that the rest of the CREB protein does not play a major role in homodimerisation affinity.

A more thorough investigation through additional mass photometry measurements at various concentrations and use of other techniques at the same buffer condition would provide a clearer prospect regarding the homodimerisation of (Halo)CREB.

4.2.3 Equilibrium dissociation constant (K_d) of HaloCREB to CRE DNA

The equilibrium dissociation constants (K_d) of HaloCREB versions to CRE DNA were estimated using equilibrium fluorescence anisotropy. A hairpin CRE DNA was used as the DNA probe as the pre-established protocol for the dissociation stopped-flow experiment required competition of CRElin DNA against the HaloCREB-CREh complex. Compared to $K_{d,\ CRE}$ values in existing literature, my obtained $K_{d,\ CREh,\ equil}$ for both WT and FD largely agree with the majority of the reports at the scale of 10^{-9} M (figure 4.3). This is also in agreement with existing *in vivo* studies of HaloCREB, where DNA-binding is not compromised from the fusion of HaloTag[144,145,150]. This also adds on to the findings of Wu et al. regarding the possible connotations of a polyhistidine tag[161], as a C-terminal His6 tag also does not appear to hinder CREB's DNA binding ability at pH 6.5.

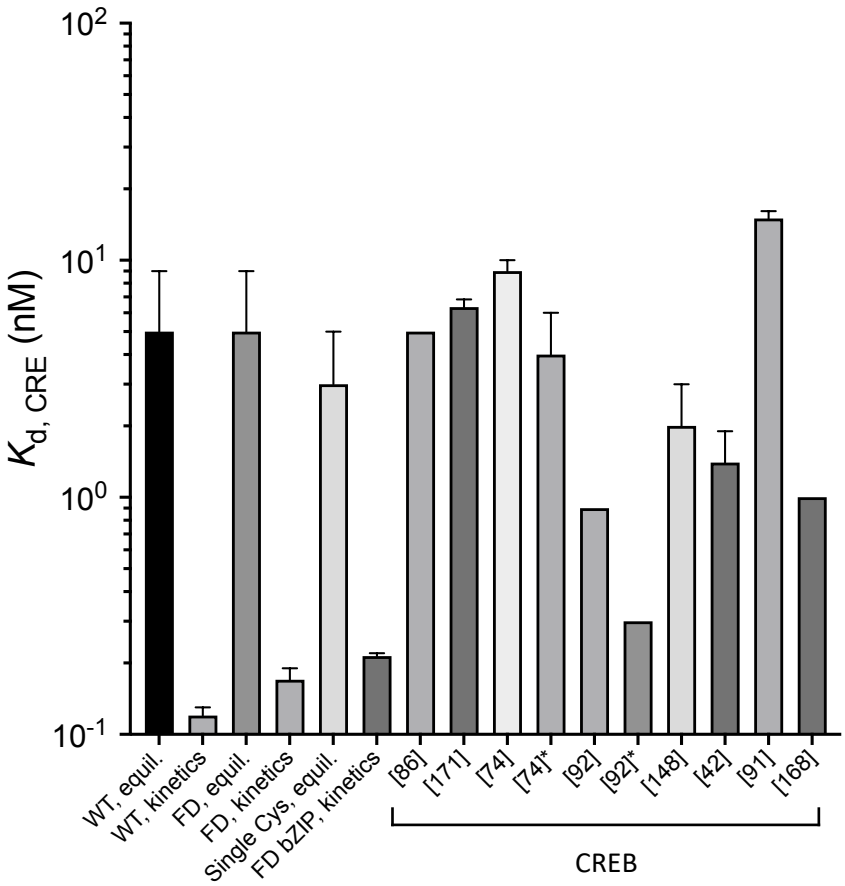


Figure 4.3 | Comparison of the equilibrium dissociating constant ($K_{d, CRE}$) of (Halo)CREB to CRE from various sources. Value for FD $bZIP_{kinetics}$ was measured by Dr. Mikhail Kuravskiy (Shammas lab, unpublished data). Values of CREB were compiled from various sources[42,74,86,91,92,148,168,171]. Asterisk (*) denote measurements were made at either 0 or < 0.7 mM $MgCl_2$

However, despite agreeing with the FD bZIP construct from the Shammas lab, the WT and FD $K_{d, CRE, kinetics}$ calculated from kinetic rate constants is 10-fold lower than the majority of literature estimates. The closest of which is a K_d of 0.3 nM reported by Moll et al., was recorded with no Mg^{2+} present[92]. Others with the exception of [168] and Shammas Lab was performed with a buffer of pH 7.5-7.6 with 50 mM NaCl or KCl, which used a buffer of pH 7.9 and 6.5 instead.

The $K_{d,CREh}$ estimation of WT HaloCREB from fluorescent anisotropy and stopped-flow are respectively 5 ± 4 nM and 0.12 ± 0.01 nM. This large mismatch might be attributed to the high concentration of labelled ligand in equilibrium studies (relative to the K_d). This has already been confirmed to cause inaccurate estimates in the same assay for CREB bZIP. However a repeat with a more sensitive instrument and lower DNA

concentration resulted in the same $K_{d,\ CREh,\ equil}$. The lower bound of the error for WT $K_{d,CREh}$ (from fitting, 95% confidence interval) does overlap with the probe concentration so this may still be the case.

The 20-fold difference between the $K_{d,\ CREh,\ equil.}$ from fluorescent anisotropy and $K_{d,\ CREh,\ kinetics}$ from stopped-flow might instead suggest the interaction is not two-state, i.e. HaloCREB dimer binding to CRE DNA. A model requiring both dimerization and DNA binding might be able to account for the difference. This is particularly the case since in equilibrium measurements the protein concentrations around the transition zone are 1-10 nM (where HaloCREB would be mostly monomeric), whereas the association stopped-flow experiments were performed at higher protein concentrations of 100-200 nM (where HaloCREB would be mostly dimer).

At the estimation of 15 nM for $K_{d,homodimerisation}$, 66% of HaloCREB would be monomeric at 5 nM, increasing up to 98% at 0.12 nM. This would be expected to make a significant difference to the estimated $K_{d,CRE}$ if unaccounted for in fitting. Surprisingly then there is no significant difference in the estimated $K_{d,\ CREh,\ equil.}$ and $K_{d,\ CREh,\ kinetics}$ for the FD construct (5 ± 4 nM and 0.17 ± 0.02 nM) compared with the WT construct. While the FD HaloCREB construct might not be entirely dimeric, a significant shift in the monomer:dimer equilibrium towards the dimer should result in a much lowered $K_{d,CRE}$ in the equilibrium data.

There are clearly some differences between WT and FD since the anisotropy curves reach different maxima, which suggests some reaction has taken place – furthermore SDS-PAGE gel of a similarly prepared sample showed around 75% dimer formation. The increased anisotropy for the DNA complex with FD indicates the molecule is tumbling slower than expected - perhaps multiple proteins bound e.g. nonspecific oligomerisation with the HaloTag cysteines. Ultimately since this result was unexpected, only a single repeat has been performed, and the percentage dimer was not estimated for that sample, the next step should be to repeat the equilibrium curve for FD.

However, if taken as accurate the K_d data with HaloCREB might support the notion that there are other populated intermediates in the pathway which have not been observed in the kinetic studies. For example, monomeric HaloCREB may associate significantly with CRE DNA, which was not the found to be the case for CREB bZIP (data from Shammas lab). Experiments utilising mutant versions of HaloCREB with mutations known to cause destabilisation of the leucine zipper[89] might address this question. Other methods more

suitable for obtaining equilibrium K_d estimates at subnanomolar concentrations e.g. FCS, single molecule fluorescence should also be attempted.

The obtained $K_{d,CRE,equil}$ for the single cysteine mutant of HaloCREB is on the same order of magnitude as the WT and FD HaloCREB. This finding is consistent with the existing literature where the cysteines are dispensable and mutations do not compromise the DNA-binding affinity of CREB[148,149].

4.2.4 Equilibrium dissociation constant (K_d) of HaloCREB to nCRE DNA

The K_d of WT, FD and single cysteine mutant HaloCREB to nCRE DNA were estimated using fluorescent anisotropy as 1100 ± 300 nM, 140 ± 80 nM and 2600 ± 400 nM (fitting errors, 95% confidence interval). As the affinity of CREB to DNA is modulated by the sequence of the oligonucleotide, direct comparison of $K_{d,\ nCRE}$ with existing literature of varying DNA probe is somewhat inconclusive. As CREB can bind anywhere along the nCRE oligonucleotide, changes in the probe nucleotide length can also modulate the $K_{d,nCRE}$. Existing literature report K_d values ranging from 10nM to 1000nM for nonconsensus CRE and nCRE sequences of various lengths.[86,91,92,171]

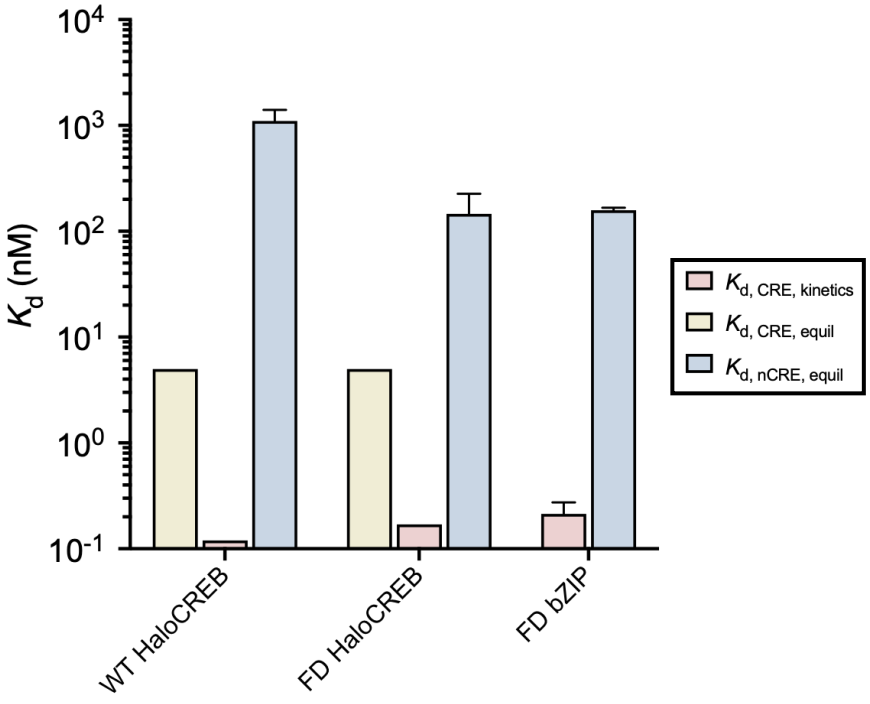


Figure 4.4 | Comparison of the equilibrium dissociating constant ($K_{d,\ nCRE}$) of HaloCREB and CREB bZIP to nCRE DNA. Values for FD bZIP was measured by Dr. Mikhail Kuravskiy (Shammas lab, unpublished data).

Interestingly, the $K_{d,nCRE,equil}$ for FD HaloCREB matches the value obtained for FD bZIP in the Shammas lab (FD HaloCREB $K_{d,nCRE,equil.}$: 140 ± 80 nM, FD bZIP $K_{d,nCRE,equil.}$: 158 ± 9 nM). The FD also displays two binding transitions with nCRE DNA as the bZIP does, while the WT and single cysteine mutant was fitted with a single-state curve. Along with the agreement in FD $K_{d,\ CRE,\ kinetics}$ between HaloCREB and bZIP, it is likely that the rest of the CREB protein does not allosterically interfere with the DNA-binding properties of dimeric CREB.

The WT (and single cysteine mutant) displayed an approximate 10-fold decrease in affinity towards nCRE compared to FD. The simplest explanation is a shift in the monomer:dimer equilibrium. The 10-fold difference would then suggest that monomeric HaloCREB is not binding tightly to nCRE DNA, and that a significant proportion of the protein is monomeric under the assay conditions. This is not expected given the previously determined $K_{d,homodimerization}$ estimate. Could the nCRE DNA itself be promoting the monomeric species? It is unlikely since the protein is present in significant excess over DNA (5 nM). Unfortunately, equilibrium data for WT bZIP to the CREh DNA construct are yet to be obtained from the Shammas lab, and therefore such comparison cannot be made.

While it cannot be concluded that HaloTag does not modulate specificity due to a lack of suitable comparison, it can be seen that the specificity of CREB is retained to an unknown extent upon fusion with HaloTag. This observation is expected as ~280 disordered amino acid residues separate the HaloTag from the HaloCREB bZIP domain and the HaloTag has been shown *in vivo* not to affect the CREB transcriptome.[144]

4.2.5 Association rate constants (k_{on}) of HaloCREB to CRE DNA

The association binding rate (k_{on}) of WT and FD HaloCREB to CREh DNA was estimated to be respectively 0.41 ± 0.02 $nM^{-1}\ s^{-1}$ and 0.60 ± 0.06 $nM^{-1}\ s^{-1}$. Negative y-intercepts were obtained for both linear fittings of WT and FD HaloCREB, and therefore was not used as an estimation for the dissociation constant. An association rate in the range of $10^{10}\ M^{-1}\ s^{-1}$ suggests the interaction is well beyond the threshold of 'diffusion-limited' and likely reflects a strong electrostatic rate enhancement. The fact that the WT associates with CREh 70% slower than the FD HaloCREB is highly suggestive given that the $K_{d,homodimerization}$ would indicate around 80% of the

protein will be dimeric under the assay conditions (100-200 nM HaloCREB). The positive correlation between k_{on} and the dimer population suggests the assay is monitoring association of the dimeric species alone. However an estimation of the $K_{d,homodimerisation}$ in biophysical buffer and a k_{on} with a uniformly dimerised FD will be essential in fully characterising the DNA binding pathway of CREB.

In comparison with the FD CREB bZIP, both WT and FD HaloCREB appear to associate with CRE DNA slower by order of magnitude (k_{on} bZIP: 2.57 ± 0.06 nM^{-1} s^{-1}) demonstrating that the rest of the protein contributes to the association kinetics. The most obvious difference is that a HaloCREB monomer is 10 times heavier in molecular weight and considerably larger in surface area compared to a bZIP monomer. This is likely to cause a reduced diffusion co-efficient, and also means that the "reactive" surface area of the protein is smaller. Importantly there are differences in net charges to be considered (-20 for HaloCREB and 6 for bZIP at pH 6.5) which are important given the large effect that electrostatic rate enhancement appears to be playing.

Regions of enriched negatively-charged residues can be found outside of the basic region that might cause auto-inhibitory behaviour. Unfortunately, the allosteric effect of the rest of the CREB protein in the k_{on} to CRE hairpin DNA cannot be fully discussed as only the k_{on} (and k_{off}) for WT bZIP is yet to be determined by the Shammas laboratory, however both bZIP constructs displayed almost identical k_{on} for a linear CRE construct (data not shown).

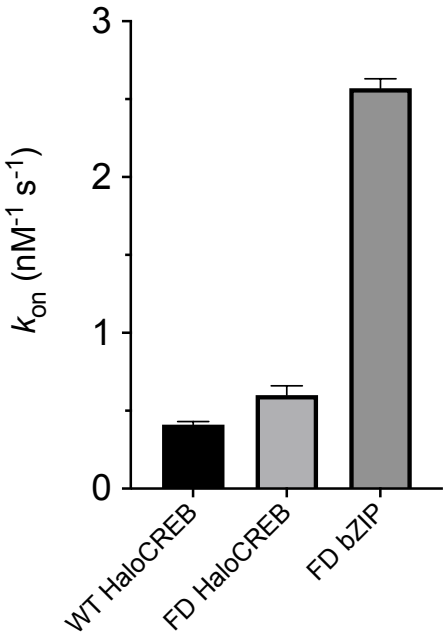


Figure 4.5 | Comparison of the association binding constant (k_{on}) of HaloCREB and CREB bZIP to CREh DNA. Value for FD bZIP was measured by Dr. Mikhail Kuravskiy (Shammas lab, unpublished data).

While this study is the first to characterise the association rate constant of HaloCREB, specific transcription factors had been identified to bind to DNA at similar rates. In particularly, the *lac* repressor binds to its target DNA with a k_{on} of 0.5 nM^{-1} s^{-1}.[172,173] Such a fast association rate is therefore not a precedent in scientific literature and is within reasonable accounts.

4.2.6 Dissociation rate constants (k_{off}) of HaloCREB from CRE DNA

In out-competition experiments with unlabelled DNA kinetic traces were well fit by single exponential decay functions. As the competition stopped-flow was performed at 10 nM HaloCREB, the soluble WT proteins should be roughly half monomeric and half dimeric. However, an appropriate fitting to a single exponential suggests that there is only one dissociation event being observed, and this is likely to be dimeric since 1) previous experiments have shown monomeric CREB to have a remarkably low binding affinity for DNA (~150 µM K_d for CREB BR) and; 2) WT and FD k_{off} are so similar.

Dissociation rate constants of WT and FD HaloCREB were estimated to be respectively 0.050 ± 0.001 s^{-1} and 0.094 ± 0.009 s^{-1}. These values suggest that WT and FD HaloCREB reside on the CREh probe for a remarkable period of time. The range of dissociation constants that was observed for HaloCREB is consistent with the FD bZIP under the same experimental conditions, suggesting that the HaloTag and surrounding CREB sequence do not significantly perturb CREB-DNA interaction.

Comparison of k_{off} between HaloCREB, bZIP and existing literature suggests that FD HaloCREB behave in a similar manner as the slow-dissociating sample in the study by Sugo et al.[145]. The two populations are separated by their observed residence time[145], respectively corresponding to CREB who are bound to nCRE and to CRE DNA. It is important to note that the DNA used in the single molecular study is a 19mer containing a half CRE site which might be expected to have a higher k_{off}. Furthermore, as the $k_{off, Sugo}$ was calculated based on residence time, diffusion of the HaloCREB along the molecule on and off the CRE site, and off the detection limit of TIRF microscopy might artificially increase the k_{off}. Additionally, the study also observed a stepwise dissociation of HaloCREB from the DNA probe, alluding to a two-step dissociation reaction.

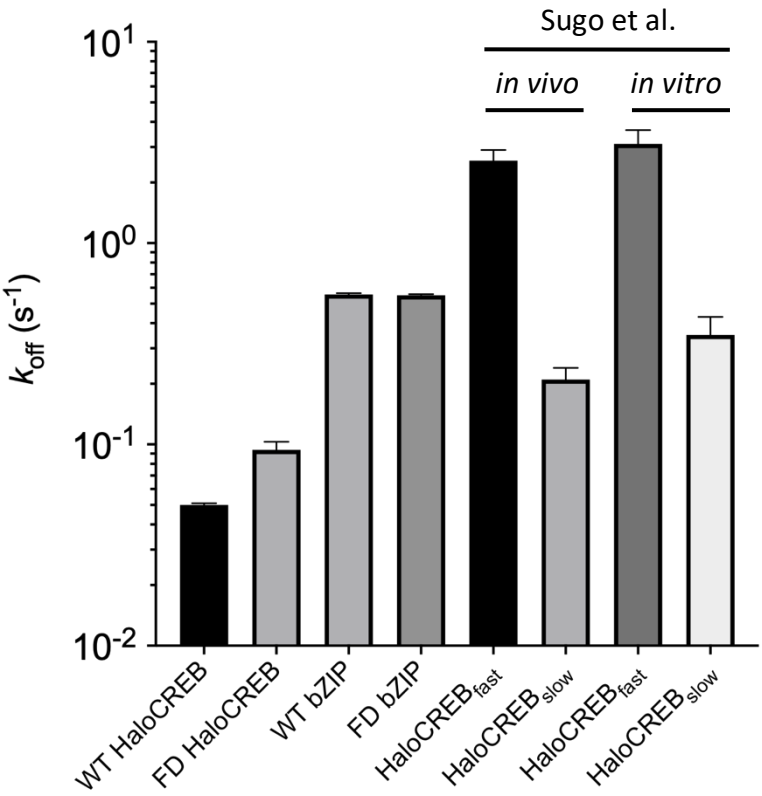


Figure 4.6 | Comparison of the dissociation binding constant (k_{off}) of HaloCREB and CREB bZIP to CRE DNA. Value for WT and FD bZIP was measured by Dr. Mikhail Kuravskiy (Shammas lab, unpublished data). The remaining pairs of dissociation rates are from Sugo et al.[145]. In vivo data collected by observing CREB sliding on immobilized DNA, in vitro data from single molecule tracking of live cells.

4.2.7 Competitor dependence of HaloCREB association with CRE DNA

The determined association rate constants and equilibrium binding conduits do not fully represent the rate of which CREB associates with genomic DNA *in vivo* as there are many more factors to consider in a physiological condition. One particular is the presence of nCRE DNA in massive excess. 10,447 full CRE sites were found in an extensive scan of the human genome[87]. This attributes a 0.003% saturation of CRE sites amongst the 3.2 x 10^9 base pairs of the human genome[151]. This highlights the need of understanding how massive excess of non-target DNA can affect the binding properties, particularly the association constant of CREB to CRE DNA.

Stopped-flow association experiments were performed with 5 nM of CREh and 1 µM of nCRE DNA to investigate any potential effects. Such mixing conditions represent a 0.5% saturation of CRE sites. It therefore is not a perfect representation of the theoretical physiological saturation but is meant as an initial test to

probe any changes in parameters. A reduction in the $k_{on,CRE}$ was observed in presence of 200-fold excess of nCRE DNA to 0.33 ± 0.02 nM^{-1} s^{-1} and 0.34 ± 0.01 nM^{-1} s^{-1} i.e. 80 and 57% of the values without competitor. This observation may be described if the nCRE DNA acts as a reservoir through transient interactions, decreasing the effective concentration of the transcription factor, and has been demonstrated in the lab mathematically for bZIP FD (data not shown). Recently Stracy et al. detailed the effect of nontarget DNA on the search mechanism of eleven representative DNA-binding proteins[174]. One of which is the *lac* operon repressor LacI, which was reported to have a similar k_{on} to its specific DNA compared to HaloCREB (0.5 nM^{-1} s^{-1})[172,173]. They found that the five chosen transcription factors spend from 56% to 93% of their search time non-specifically bound to DNA. The slowdown in association rate of CREB in presence of nCRE DNA suggests that the same might also be applicable for CREB.

However, comparison of the slowdown of HaloCREB and bZIP does not appear to be consistent. The k_{on} for FD bZIP in presence of 1 µM nCRE DNA is 0.68 ± 0.03 nM^{-1} s^{-1}, a 3.5-fold decrease from the unperturbed k_{on} which is the precise amount expected according to sequestration of bZIP by a pool of nCRE.

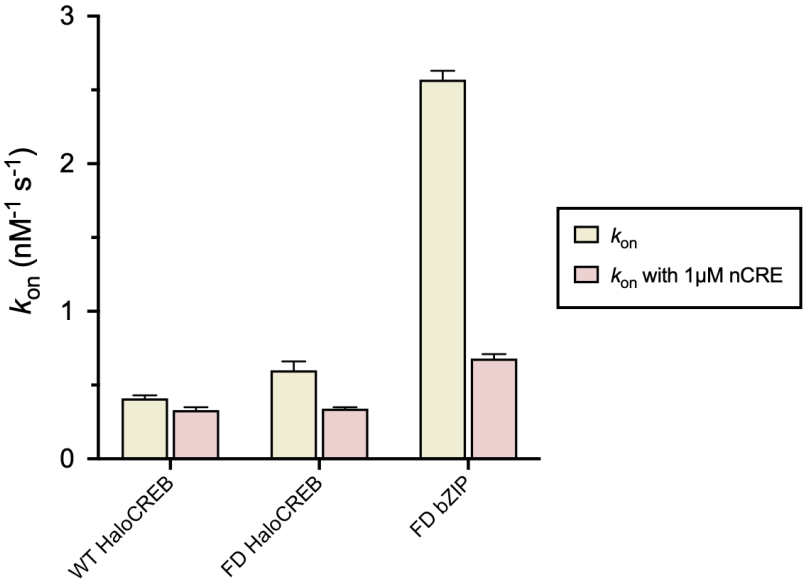


Figure 4.7 | Comparison of the slowdown in association binding constant (k_{on}) of HaloCREB and CREB bZIP to CRE DNA in presence of 1 µM nCRE DNA. Value for FD bZIP was measured by Dr. Mikhail Kuravskiy (Shammas lab, unpublished data).

Slowdown is attributed to nonspecific association with the non-target DNA, so the decrease in k_{on} is predicted by the $K_{d,nCRE}$. According to the $K_{d,nCRE}$ estimated for FD the rate expected is only 13% of that without competitor. Intriguingly if the $K_{d,nCRE}$ estimated for HaloCREB WT is used instead then the rate constant is expected to be around 54% of that without competitor (c.f. 57% observed). Given the discrepancy between $K_{d,nCRE}$ of WT HaloCREB and that for FD HaloCREB this figure is worth reconsidering in further experiments – however an order of magnitude error seems unlikely, and if it is accurate then it may provide very important implications for search mechanism. The appropriate $K_{d,nCRE}$ is actually for nCRE without a label so this was investigated as a possible explanation.

A competition fluorescence anisotropy experiment was used to determine the binding constant of HaloCREB to unlabelled nCRE DNA. This is needed to identify the contribution of the highly negatively charged fluorophore in the ionic interactions that mediate DNA binding. By using unlabelled nCRE DNA to outcompete CREB-nCRE DNA complex and observing the decrease in anisotropy, a $K_{i,\ unlabelled\ nCRE}$ can be obtained. It is noteworthy that as the K_i only approximates $K_{d,\ unlabelled\ nCRE}$. Furthermore, Nikolovska-Coleskaa et al. found that the method deviates at higher K_i values in the range of 100 μM[141]. Accurate estimates for $K_{d,\ unlabelled\ nCRE}$ are reliant on accurate $K_{d,nCRE}$.

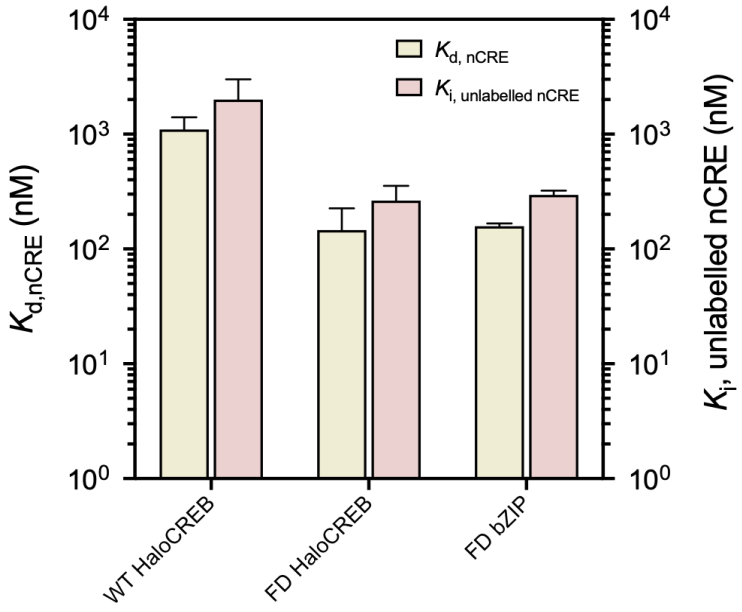


Figure 4.8 | Comparison between $K_{d,nCRE}$ and $K_{i,\ unlabelled\ nCRE}$ of various HaloCREB and bZIP constructs. Value for FD bZIP was measured by Dr. Mikhail Kuravskiy (Shammas lab, unpublished data).

A 2-fold increase can be seen between the $K_{d,nCRE}$ and $K_{i,unlabelled\ nCRE}$ of all three constructs suggesting a consistent effect from the fluorophore. It is important to note that the data for WT HaloCREB had a very small signal change, which could be improved by adjusting the initial concentration to better match the $K_{d,nCRE}$ such that there is a reasonable amount of protein bound to the labelled nCRE to be dissociated. Repeats and competition experiments at various concentration of nCRE must be done to further investigate the manner of which k_{on} is decreased.

4.2.8 Non-reduced cysteine residues decrease binding affinity

The WT CREB contains 3 cysteine residues in the bZIP domain. While existing literature and this study suggest that the cysteines are not essential for DNA binding they are important *in vivo* for binding affinity of accessory TORC proteins[148]. CREB is unlikely to be oxidised *in vivo* given the generally reducing environment of the cell. Association and dissociation kinetic constants of WT HaloCREB to CRE DNA were compared under reducing and non-reducing conditions to see if this is an important consideration for buffers.

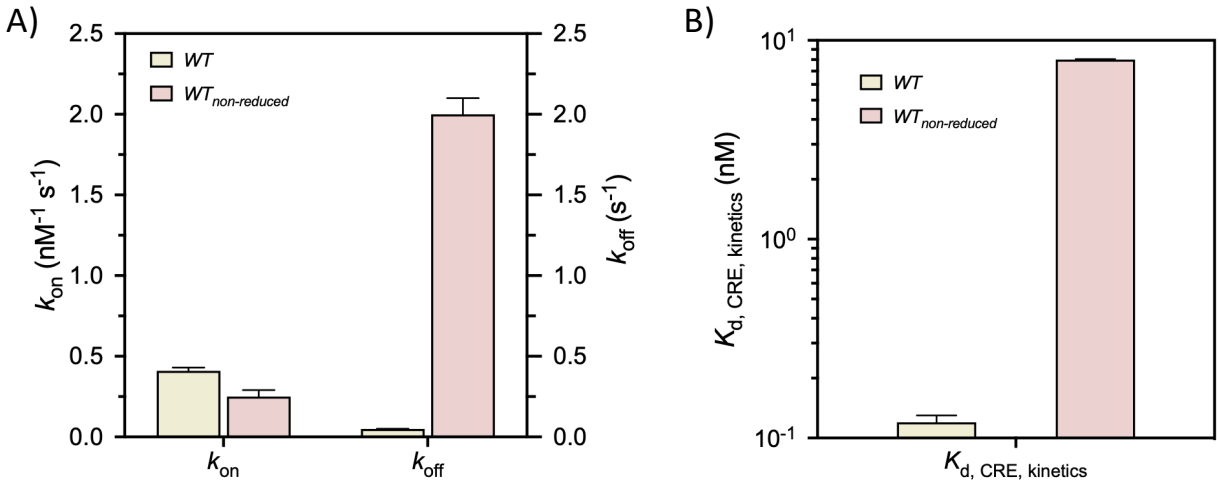


Figure 4.9 | Effect of non-reduced cysteines in the association and dissociation binding constants of WT HaloCREB to CRE DNA. Experiments with $WT_{non\text{-}reduced}$ was performed with WT HaloCREB with biophysical buffer in absence of DTT. A) Comparison of k_{on} and k_{off} between WT and $WT_{non\text{-}reduced}$. B) Comparison of $K_{d,\ CRE,\ kinetics}$ between WT and $WT_{non\text{-}reduced}$.

In a non-reducing environment, i.e. absence of DTT in biophysical buffer, WT HaloCREB associates with CRE DNA slower and dissociates at a rate increased by 8-fold. This disparity resulted in a $K_{d,CRE,kinetics}$ greater by an 2 orders of magnitude. The 100-fold reduction in affinity is likely a result of the formation of nonspecific

oligmers. Such oligomers by disulphide bridges in the bZIP may disrupt the orientation and alignment for dimerisation, forming domains that are not optimised for DNA binding. HaloTag also contains two cysteine residues. However, its crystal structure (PDB: 5UY1) indicated that only one is surface accessible, but with the side chain facing inwards. While it is not currently determined whether the HaloTag is involved, the presence of reducing agent in the biophysical buffer greatly enhances the DNA-binding ability of HaloCREB. In the cell redox state is controlled through cellular redox factors. AP-1, a member of the bZIP family was found be strictly maintain reduced by Redox-factor 1 (Ref-1) and the thioredoxin-1 system[175]. Similar systems might also be in place to regulate the redox state of CREB *in vivo*.

5. Conclusion

This study is the first to develop and optimise a complete expression and purification protocol for nucleic-acid free HaloCREB for biophysical studies. The purification was expectedly tricky as it was essentially the 'worst of both worlds' between disordered protein and globular proteins: unable to exploit the disorder nature of the protein while experiencing difficulties with particular robust purification techniques.

Regardless, this study found the subsequent use of Ni-affinity column chromatography and heparin affinity column to be sufficient in eliminating both protein and nucleic-acid interaction. The protein purity of the sample was confirmed by both SDS-PAGE and denaturing mass spectrometry, whereas the nucleic-acid purity was additionally confirmed by A_{260}/A_{280} reading of the sample. This expression and purification protocol yielded 1 mg of HaloCREB per 1 L of culture which is sufficient for biophysical studies.

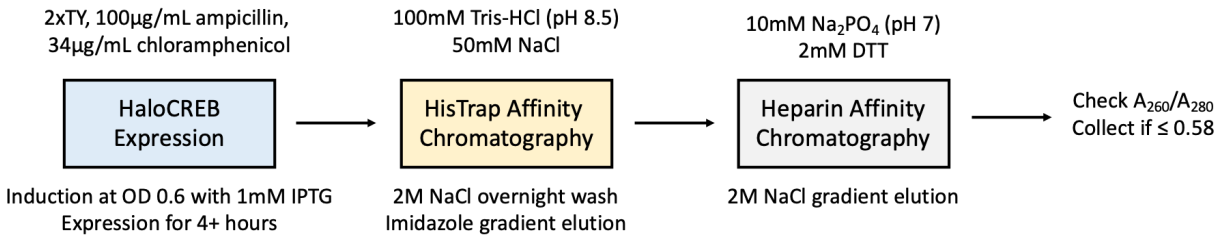


Figure 5.1 | Final expression and purification protocol to obtain a nucleic-acid free ample of HaloCREB suitable for biophysical analysis.

This study also provides preliminary estimates for key binding affinities and kinetic parameters of HaloCREB. HaloCREB was found to demonstrate specificity towards CRE DNA with a K_d of 5 ± 4 nM, agreeing with the affinities reported by the majority of the current literature. An estimation of the homodimerisation K_d (~15 nM) with mass photometry revealed the equilibrium to be only marginally perturbed by the context of the full CREB sequence compared with the bZIP alone. This value suggests that that CREB is mostly dimeric in most of the experiments I have performed. Simplistically this may also suggest the protein is mostly dimeric in the cell since CREB is present at about 0.4 μM, however there are many other contributing factors in this environment.

Utilizing stopped-flow kinetics, association and dissociation binding constants of HaloCREB was also estimated. HaloCREB was found to have an association rate constant to CRE DNA of 10^{10} M^{-1} s^{-1}, and a dissociation rate constant of around 0.1 s^{-1}. The fast association rate constant implies an electrostatically enhanced reaction between CREB and CRE DNA, while the slow dissociation rate constant suggests a long residence time consistent with those observed in the cellular environment.

Combined, these yield a $K_{d,CRE,kinetics}$ of subnanomolar scale, significantly different to that measured by equilibrium studies. If the results of equilibrium studies are reproducible then this suggests a simple two-state reaction is insufficient to describe how HaloCREB binds to CRE DNA, supporting the notion of the involvement of the monomer pathway.

In comparison with existing literature and the parameters for the bZIP construct collected by members of the Shammas laboratory, this study suggests that the fusion of a HaloTag in the N-terminus does not interfere with the DNA binding or recognition of CREB. The full-length construct has a homodimerisation constant 3-fold higher than bZIP, and association and dissociation rate constants for CRE DNA that are 5-lower. These are small changes. However, this indicates that the rest of the CREB protein exerts an allosteric effect in mediating the dimerisation and DNA interaction in a manner a yet to be understood.

The effect of surrounding excess non-target DNA on the association rates of CREB was also investigated. A slow-down in target search was observed for HaloCREB with the presence of 200-fold excess nCRE DNA. This reflect that despite the significant lower affinity to nCRE DNA in comparison with bZIP, massive excess may sequester dimer so that the observed association rate is reduced by nonspecific interactions. This sheds a light on how CREB DNA binding might differ from what was measured *in vitro* due to a multitude of physiological factors.

Through biophysical characterisation of HaloCREB, this study presents preliminary estimate for key parameters in determining the target search function of CREB. With repeat experiments and confirmation of biophysical and kinetic parameters by different experimental methods, a precise model can be developed to understand how CREB searches along DNA.

An exact understanding of the search model of CREB can reveal similarities that might hold true for other transcription factors, particularly ones that are also intrinsically disordered. Knowing the equilibrium between binding pathways can also aid in the design of therapeutics against CREB through identifying additional interactions that can be targeted against.

6. Future research opportunities

There are several unexplained findings in my studies, which may reflect new biology to be discovered. To be more specific – the $K_{d,CREh}$ obtained by equilibrium and kinetic methods do not match, and the $K_{d,nCRE}$ for FD and WT HaloCREB differ by an order of magnitude despite the soluble protein being largely dimeric under the conditions of the experiment. It is possible these findings indicate the involvement of a monomer search pathway, which has not been observed for bZIP.

To fully understand the DNA search mechanism of CREB, the experiments detailed in this study will have to be thoroughly repeated for a more accurate estimation. The biophysical parameters will also have to be

repeated by other experimental methods as a cross reference. Isothermal calorimetry can be used to determine $K_{d,\ nCRE,\ equil}$; analytical ultracentrifugation or stopped-flow with intrinsic tyrosine fluorescence can be used for a more careful determination of the $K_{d,\ homodimerisation}$. Selective cleavage of different parts of the CREB protein can also further understand the role it may play in mediating interactions.

Given the uncertainty now shed on the stoichiometry of the target search pathway a particularly useful experiment would involve determining the stoichiometry of CREB within the cell. This might be assessed using single-molecule tracking of labelled HaloCREB in the cell e.g. using a dual-labelling strategy.

Beyond alternative methods in obtaining the same biophysical parameters, the transition state of CREB binding to DNA can be probed through ϕ-value analysis. This analysis would be particularly insightful if a multi-phase binding model is confirmed. Understanding the transition state of CREB associating with DNA might reveal underlying energetic barriers of the interaction.

Observing CREB at a single molecular level via microscopy techniques can yield DNA sliding mechanics and further the understanding of CREB's DNA search. In particularly, the use of optical tweezers to precisely track the rate at which CREB slides along specific and nonspecific DNA would be of particular interest. Additionally, the experimental setup from Stacey et al. can be modified and used to determine the occupancy rate of CREB on non-specific DNA.

Finally, structural analysis with NMR spectroscopy and molecular simulation would also be particularly insightful as it can complement all the studies mentioned above. Understanding the dynamic tendencies of CREB in solution and how it interacts with various DNA sequences on a residual level would be a giant step in completing the understanding of the CREB DNA binding mode, and particularly how it compares with other ordered transcription factors.

7. References

(1) Fischer, E. Einfluss Der Configuration Auf Die Wirkung Der Enzyme. *Eur. J. Inorg. Chem.* **1894**, *27* (3), 2985–2993. https://doi.org/10.1002/cber.18940270364.

(2) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **1958**, *181* (4610), 662–666. https://doi.org/10.1038/181662a0.

(3) Garner, A.; Cannon, P.; Romero, P.; Obradovic, Z,; Dunker, A.K. Predicting Disordered Regions from Amino Acid Sequence: Common Themes Despite Differing Structural Characterization. *Genome Inform. Workshop Genome Inform.* **1998**, *9*, 201–213.

(4) Dunker, A. K.; Garner, E.; Guilliot, S.; Romero, P.; Albrecht, K.; Hart, J.; Obradovic, Z.; Kissinger, C.; Villafranca, J. E. Protein Disorder and the Evolution of Molecular Recognition: Theory, Predictions and Observations. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* **1998**, 473–484.

(5) Williams, R. M.; Obradovi, Z.; Mathura, V.; Braun, W.; Garner, E. C.; Young, J.; Takayama, S.; Brown, C. J.; Dunker, A. K. The Protein Non-Folding Problem: Amino Acid Determinants of Intrinsic Order and Disorder. *Pac. Symp. Biocomput. Symp. Biocomput.* **2001**, No. Journal Article, 89–100.

(6) Uversky, V. N. Intrinsically Disordered Proteins and Novel Strategies for Drug Discovery. *Expert Opin. Drug Discov.* **2012**, *7* (6), 475–488. https://doi.org/10.1517/17460441.2012.686489.

(7) Theillet, F. X.; Kalmar, L.; Tompa, P.; Han, K. H.; Selenko, P.; Dunker, A. K.; Daughdrill, G. W.; Uversky, V. N. The Alphabet of Intrinsic Disorder: I. Act like a Pro: On the Abundance and Roles of Proline Residues in Intrinsically Disordered Proteins. *Intrinsically Disord. Proteins* **2013**, *1* (1), e24360. https://doi.org/10.4161/idp.24360.

(8) Tompa, P. Intrinsically Unstructured Proteins. *Trends Biochem. Sci.* **2002**, *27* (10), 527–533.

(9) Dyson, H. J.; Wright, P. E.; Scheraga, H. A. The Role of Hydrophobic Interactions in Initiation and Propagation of Protein Folding. *Proc. Natl. Acad. Sci.* **2006**, *103* (35), 13057–13061. https://doi.org/10.1073/pnas.0605504103.

(10) Uversky, V. N. Natively Unfolded Proteins: A Point Where Biology Waits for Physics. *Protein Sci. Publ. Protein Soc.* **2002**, *11* (4), 739–756. https://doi.org/10.1110/ps.4210102.

(11) Wright, P. E.; Dyson, H. J. Intrinsically Unstructured Proteins: Re-Assessing the Protein Structure-Function Paradigm. *J. Mol. Biol.* **1999**, *293* (2), 321–331. https://doi.org/10.1006/jmbi.1999.3110.

(12) Kyte, J.; Doolittle, R. F. A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* **1982**, *157* (1), 105–132. https://doi.org/10.1016/0022-2836(82)90515-0.

(13) Onuchic, J. N.; Wolynes, P. G. Theory of Protein Folding. *Curr. Opin. Struct. Biol.* **2004**, *14* (1), 70–75. https://doi.org/10.1016/j.sbi.2004.01.009.

(14) Fisher, C. K.; Stultz, C. M. Constructing Ensembles for Intrinsically Disordered Proteins. *Curr. Opin. Struct. Biol.* **2011**, *21* (3), 426–431. https://doi.org/10.1016/j.sbi.2011.04.001.

(15) Liu, J.; Faeder, J. R.; Camacho, C. J. Toward a Quantitative Theory of Intrinsically Disordered Proteins and Their Function. *Proc. Natl. Acad. Sci.* **2009**, *106* (47), 19819–19823. https://doi.org/10.1073/pnas.0907710106.

(16) Dunker, A. K.; Cortese, M. S.; Romero, P.; Iakoucheva, L. M.; Uversky, V. N. Flexible Nets. *FEBS J.* **2005**, *272* (20), 5129–5148. https://doi.org/10.1111/j.1742-4658.2005.04948.x.

(17) Nadassy, K.; Wodak, S. J.; Janin, J. Structural Features of Protein−Nucleic Acid Recognition Sites †. *Biochemistry* **1999**, *38* (7), 1999–2017. https://doi.org/10.1021/bi982362d.

(18) Uversky, V. N.; Dunker, A. K. The Case for Intrinsically Disordered Proteins Playing Contributory Roles in Molecular Recognition without a Stable 3D Structure. *F1000 Biol. Rep.* **2013**, *5*. https://doi.org/10.3410/B5-1.

(19) Liu, J.; Perumal, N. B.; Oldfield, C. J.; Su, E. W.; Uversky, V. N.; Dunker, A. K. Intrinsic Disorder in Transcription Factors. *Biochemistry* **2006**, *45* (22), 6873–6888. https://doi.org/10.1021/bi0602718.

(20) Gunasekaran, K.; Tsai, C.-J.; Kumar, S.; Zanuy, D.; Nussinov, R. Extended Disordered Proteins: Targeting Function with Less Scaffold. *Trends Biochem. Sci.* **2003**, *28* (2), 81–85. https://doi.org/10.1016/S0968-0004(03)00003-3.

(21) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. Intrinsically Disordered Protein. *J. Mol. Graph. Model.* **2001**, *19* (1), 26–59. https://doi.org/10.1016/S1093-3263(00)00138-8.

(22) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. Showing Your ID: Intrinsic Disorder as an ID for Recognition, Regulation and Cell Signaling. *J. Mol. Recognit.* **2005**, *18* (5), 343–384. https://doi.org/10.1002/jmr.747.

(23) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J. Mol. Biol.* **2004**, *337* (3), 635–645. https://doi.org/10.1016/j.jmb.2004.02.002.

(24) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradovic, Z.; Dunker, A. K. Intrinsic Disorder in Cell-Signaling and Cancer-Associated Proteins. *J. Mol. Biol.* **2002**, *323* (3), 573–584.

(25) Xie, H.; Vucetic, S.; Iakoucheva, L. M.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N.; Obradovic, Z. Functional Anthology of Intrinsic Disorder. 1. Biological Processes and Functions of Proteins with Long Disordered Regions. *J. Proteome Res.* **2007**, *6* (5), 1882–1898. https://doi.org/10.1021/pr060392u.

(26) Xie, H.; Vucetic, S.; Iakoucheva, L. M.; Oldfield, C. J.; Dunker, A. K.; Obradovic, Z.; Uversky, V. N. Functional Anthology of Intrinsic Disorder. 3. Ligands, Post-Translational Modifications, and Diseases Associated with Intrinsically Disordered Proteins. *J. Proteome Res.* **2007**, *6* (5), 1917–1932. https://doi.org/10.1021/pr060394e.

(27) Csizmok, V.; Follis, A. V.; Kriwacki, R. W.; Forman-Kay, J. D. Dynamic Protein Interaction Networks and New Structural Paradigms in Signaling. *Chem. Rev.* **2016**, *116* (11), 6424–6462. https://doi.org/10.1021/acs.chemrev.5b00548.

(28) Shammas, S. L. Mechanistic Roles of Protein Disorder within Transcription. *Curr. Opin. Struct. Biol.* **2017**, *42*, 155–161. https://doi.org/10.1016/j.sbi.2017.02.003.

(29) Borg, M.; Mittag, T.; Pawson, T.; Tyers, M.; Forman-Kay, J. D.; Chan, H. S. Polyelectrostatic Interactions of Disordered Ligands Suggest a Physical Basis for Ultrasensitivity. *Proc. Natl. Acad. Sci.* **2007**, *104* (23), 9650–9655. https://doi.org/10.1073/pnas.0702580104.

(30) Ferrell, J. E. Tripping the Switch Fantastic: How a Protein Kinase Cascade Can Convert Graded Inputs into Switch-like Outputs. *Trends Biochem. Sci.* **1996**, *21* (12), 460–466. https://doi.org/10.1016/S0968-0004(96)20026-X.

(31) Lee, C. W.; Ferreon, J. C.; Ferreon, A. C. M.; Arai, M.; Wright, P. E. Graded Enhancement of P53 Binding to CREB-Binding Protein (CBP) by Multisite Phosphorylation. *Proc. Natl. Acad. Sci.* **2010**, *107* (45), 19290–19295. https://doi.org/10.1073/pnas.1013078107.

(32) Triezenberg, S. J. Structure and Function of Transcriptional Activation Domains. *Curr. Opin. Genet. Dev.* **1995**, *5* (2), 190–196. https://doi.org/10.1016/0959-437X(95)80007-7.

(33) Minezaki, Y.; Homma, K.; Kinjo, A. R.; Nishikawa, K. Human Transcription Factors Contain a High Fraction of Intrinsically Disordered Regions Essential for Transcriptional Regulation. *J. Mol. Biol.* **2006**, *359* (4), 1137–1149. https://doi.org/10.1016/j.jmb.2006.04.016.

(34) Blainey, P. C.; Luo, G.; Kou, S. C.; Mangel, W. F.; Verdine, G. L.; Bagchi, B.; Xie, X. S. Nonspecifically Bound Proteins Spin While Diffusing along DNA. *Nat. Struct. Mol. Biol.* **2009**, *16* (12), 1224–1229. https://doi.org/10.1038/nsmb.1716.

(35) Reeves, R.; Nissen, M. S. The A.T-DNA-Binding Domain of Mammalian High Mobility Group I Chromosomal Proteins. A Novel Peptide Motif for Recognizing DNA Structure. *J. Biol. Chem.* **1990**, *265* (15), 8573–8582.

(36) Aravind, L.; Landsman, D. AT-Hook Motifs Identified in a Wide Variety of DNA-Binding Proteins. *Nucleic Acids Res.* **1998**, *26* (19), 4413–4421.

(37) Huth, J. R.; Bewley, C. A.; Nissen, M. S.; Evans, J. N. S.; Reeves, R.; Gronenborn, A. M.; Clore, G. M. The Solution Structure of an HMG-I(Y)–DNA Complex Defines a New Architectural Minor Groove Binding Motif. *Nat. Struct. Biol.* **1997**, *4* (8), 657–665. https://doi.org/10.1038/nsb0897-657.

(38) Fonfría-Subirós, E.; Acosta-Reyes, F.; Saperas, N.; Pous, J.; Subirana, J. A.; Campos, J. L. Crystal Structure of a Complex of DNA with One AT-Hook of HMGA1. *PLoS ONE* **2012**, *7* (5). https://doi.org/10.1371/journal.pone.0037120.

(39) Zhang, W.; Wu, Q.; Pwee, K.-H.; Manjunatha Kini, R. Interaction of Wheat High-Mobility-Group Proteins with Four-Way-Junction DNA and Characterization of the Structure and Expression of HMGA Gene. *Arch. Biochem. Biophys.* **2003**, *409* (2), 357–366. https://doi.org/10.1016/S0003-9861(02)00630-6.

(40) Llorca, C. M.; Potschin, M.; Zentgraf, U. BZIPs and WRKYs: Two Large Transcription Factor Families Executing Two Different Functional Strategies. *Front. Plant Sci.* **2014**, *5*. https://doi.org/10.3389/fpls.2014.00169.

(41) Miller, M. The Importance of Being Flexible: The Case of Basic Region Leucine Zipper Transcriptional Regulators. *Curr. Protein Pept. Sci.* **2009**, *10* (3), 244–269.

(42) Schumacher, M. A.; Goodman, R. H.; Brennan, R. G. The Structure of a CREB BZIP·Somatostatin CRE Complex Reveals the Basis for Selective Dimerization and Divalent Cation-Enhanced DNA Binding. *J. Biol. Chem.* **2000**, *275* (45), 35242–35247. https://doi.org/10.1074/jbc.M007293200.

(43) O'Neil, K. T.; Shuman, J. D.; Ampe, C.; DeGrado, W. F. DNA-Induced Increase in the Alpha-Helical Content of C/EBP and GCN4. *Biochemistry* **1991**, *30* (37), 9030–9034. https://doi.org/10.1021/bi00101a017.

(44) Williams, S. C.; Angerer, N. D.; Johnson, P. F. C/EBP Proteins Contain Nuclear Localization Signals Imbedded in Their Basic Regions. *Gene Expr.* **1997**, *6* (6), 371–385.

(45) Weis, K. Importins and Exportins: How to Get in and out of the Nucleus. *Trends Biochem. Sci.* **1998**, *23* (5), 185–189. https://doi.org/10.1016/s0968-0004(98)01204-3.

(46) Fujii, Y.; Shimizu, T.; Toda, T.; Yanagida, M.; Hakoshima, T. Structural Basis for the Diversity of DNA Recognition by BZIP Transcription Factors. *Nat. Struct. Biol.* **2000**, *7* (10), 889–893. https://doi.org/10.1038/82822.

(47) Vinson, C.; Myakishev, M.; Acharya, A.; Mir, A. A.; Moll, J. R.; Bonovich, M. Classification of Human B-ZIP Proteins Based on Dimerization Properties. *Mol. Cell. Biol.* **2002**, *22* (18), 6321–6335. https://doi.org/10.1128/MCB.22.18.6321-6335.2002.

(48) Sassone-Corsi, P. Coupling Gene Expression to CAMP Signalling: Role of CREB and CREM. *Int. J. Biochem. Cell Biol.* **1998**, *30* (1), 27–38. https://doi.org/10.1016/S1357-2725(97)00093-9.

(49) Amoutzias, G. D.; Bornberg-Bauer, E.; Oliver, S. G.; Robertson, D. L. Reduction/Oxidation-Phosphorylation Control of DNA Binding in the BZIP Dimerization Network. *BMC Genomics* **2006**, *7* (1), 107. https://doi.org/10.1186/1471-2164-7-107.

(50) Newman, J. R. S.; Keating, A. E. Comprehensive Identification of Human BZIP Interactions with Coiled-Coil Arrays. *Science* **2003**, *300* (5628), 2097–2101. https://doi.org/10.1126/science.1084648.

(51) Crick, F. H. C.; IUCr. The packing of -helices: simple coiled-coils http://scripts.iucr.org/cgi-bin/paper?S0365110X53001964 (accessed Oct 10, 2020). https://doi.org/10.1107/S0365110X53001964.

(52) Thompson, K. S.; Vinson, C. R.; Freire, E. Thermodynamic Characterization of the Structural Stability of the Coiled-Coil Region of the BZIP Transcription Factor GCN4. *Biochemistry* **1993**, *32* (21), 5491–5496. https://doi.org/10.1021/bi00072a001.

(53) Moitra, J.; Szilák, L.; Krylov, D.; Vinson, C. Leucine Is the Most Stabilizing Aliphatic Amino Acid in the d Position of a Dimeric Leucine Zipper Coiled Coil. *Biochemistry* **1997**, *36* (41), 12567–12573. https://doi.org/10.1021/bi971424h.

(54) Wagschal, K.; Tripet, B.; Mant, C.; Hodges, R. S.; Lavigne, P. The Role of Position a in Determining the Stability and Oligomerization State of α-Helical Coiled Coils: 20 Amino Acid Stability Coefficients in the Hydrophobic Core of Proteins. *Protein Sci.* **1999**, *8* (11), 2312–2329. https://doi.org/10.1110/ps.8.11.2312.

(55) Acharya, A.; Ruvinov, S. B.; Gal, J.; Moll, J. R.; Vinson, C. A Heterodimerizing Leucine Zipper Coiled Coil System for Examining the Specificity of a Position Interactions: Amino Acids I, V, L, N, A, and K. *Biochemistry* **2002**, *41* (48), 14122–14131. https://doi.org/10.1021/bi020486r.

(56) O'Shea, E. K.; Rutkowski, R.; Kim, P. S. Mechanism of Specificity in the Fos-Jun Oncoprotein Heterodimer. *Cell* **1992**, *68* (4), 699–708. https://doi.org/10.1016/0092-8674(92)90145-3.

(57) Vinson, C. R.; Hai, T.; Boyd, S. M. Dimerization Specificity of the Leucine Zipper-Containing BZIP Motif on DNA Binding: Prediction and Rational Design. *Genes Dev.* **1993**, *7* (6), 1047–1058. https://doi.org/10.1101/gad.7.6.1047.

(58) Tarczewska, A.; Greb-Markiewicz, B. The Significance of the Intrinsically Disordered Regions for the Functions of the BHLH Transcription Factors. *Int. J. Mol. Sci.* **2019**, *20* (21). https://doi.org/10.3390/ijms20215306.

(59) Jones, S. An Overview of the Basic Helix-Loop-Helix Proteins. *Genome Biol.* **2004**, *5* (6), 226. https://doi.org/10.1186/gb-2004-5-6-226.

(60) Conacci-Sorrell, M.; McFerrin, L.; Eisenman, R. N. An Overview of MYC and Its Interactome. *Cold Spring Harb. Perspect. Med.* **2014**, *4* (1), a014357. https://doi.org/10.1101/cshperspect.a014357.

(61) Chrivia, J. C.; Kwok, R. P. S.; Lamb, N.; Hagiwara, M.; Montminy, M. R.; Goodman, R. H. Phosphorylated CREB Binds Specifically to the Nuclear Protein CBP. *Nature* **1993**, *365* (6449), 855–859. https://doi.org/10.1038/365855a0.

(62) Luo, Q.; Viste, K.; Urday-Zaa, J. C.; Senthil Kumar, G.; Tsai, W.-W.; Talai, A.; Mayo, K. E.; Montminy, M.; Radhakrishnan, I. Mechanism of CREB Recognition and Coactivation by the CREB-Regulated Transcriptional Coactivator CRTC2. *Proc. Natl. Acad. Sci.* **2012**, *109* (51), 20865–20870. https://doi.org/10.1073/pnas.1219028109.

(63) Ferreri, K.; Gill, G.; Montminy, M. The CAMP-Regulated Transcription Factor CREB Interacts with a Component of the TFIID Complex. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91* (4), 1210–1213.

(64) Johannessen, M.; Delghandi, M. P.; Moens, U. What Turns CREB On? *Cell. Signal.* **2004**, *16* (11), 1211–1227. https://doi.org/10.1016/j.cellsig.2004.05.001.

(65) Quinn, P. G. Distinct Activation Domains within CAMP Response Element-Binding Protein (CREB) Mediate Basal and CAMP-Stimulated Transcription. *J. Biol. Chem.* **1993**, *268* (23), 16999–17009.

(66) Canettieri, G.; Morantte, I.; Guzmán, E.; Asahara, H.; Herzig, S.; Anderson, S. D.; Yates, J. R.; Montminy, M. Attenuation of a Phosphorylation-Dependent Activator by an HDAC–PP1 Complex. *Nat. Struct. Mol. Biol.* **2003**, *10* (3), 175–181. https://doi.org/10.1038/nsb895.

(67) Felinski, E. A.; Quinn, P. G. The CREB Constitutive Activation Domain Interacts with TATA-Binding Protein-Associated Factor 110 (TAF110) through Specific Hydrophobic Residues in One of the Three Subdomains Required for Both Activation and TAF110 Binding. *J. Biol. Chem.* **1999**, *274* (17), 11672–11678. https://doi.org/10.1074/jbc.274.17.11672.

(68) Carlezon, W. A.; Duman, R. S.; Nestler, E. J. The Many Faces of CREB. *Trends Neurosci.* **2005**, *28* (8), 436–445. https://doi.org/10.1016/j.tins.2005.06.005.

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

(70) Dahal, L.; Kwan, T. O. C.; Shammas, S. L.; Clarke, J. PKID Binds to KIX via an Unstructured Transition State with Nonnative Interactions. *Biophys. J.* **2017**, *113* (12), 2713–2722. https://doi.org/10.1016/j.bpj.2017.10.016.

(71) Turjanski, A. G.; Gutkind, J. S.; Best, R. B.; Hummer, G. Binding-Induced Folding of a Natively Unstructured Transcription Factor. *PLoS Comput. Biol.* **2008**, *4* (4). https://doi.org/10.1371/journal.pcbi.1000060.

(72) Dahal, L.; Shammas, S. L.; Clarke, J. Phosphorylation of the IDP KID Modulates Affinity for KIX by Increasing the Lifetime of the Complex. *Biophys. J.* **2017**, *113* (12), 2706–2712. https://doi.org/10.1016/j.bpj.2017.10.015.

(73) Shanware, N. P.; Trinh, A. T.; Williams, L. M.; Tibbetts, R. S. Coregulated Ataxia Telangiectasia-Mutated and Casein Kinase Sites Modulate CAMP-Response Element-Binding Protein-Coactivator Interactions in Response to DNA Damage. *J. Biol. Chem.* **2007**, *282* (9), 6283–6291. https://doi.org/10.1074/jbc.M610674200.

(74) Shnitkind, S.; Martinez-Yamout, M. A.; Dyson, H. J.; Wright, P. E. Structural Basis for Graded Inhibition of CREB:DNA Interactions by Multisite Phosphorylation. *Biochemistry* **2018**, *57* (51), 6964–6972. https://doi.org/10.1021/acs.biochem.8b01092.

(75) Kim, S. H.; Trinh, A. T.; Larsen, M. C.; Mastrocola, A. S.; Jefcoate, C. R.; Bushel, P. R.; Tibbetts, R. S. Tunable Regulation of CREB DNA Binding Activity Couples Genotoxic Stress Response and Metabolism. *Nucleic Acids Res.* **2016**, *44* (20), 9667–9680. https://doi.org/10.1093/nar/gkw643.

(76) Wu, X.; McMurray, C. T. Calmodulin Kinase II Attenuation of Gene Transcription by Preventing CAMP Response Element-Binding Protein (CREB) Dimerization and Binding of the CREB-Binding Protein. *J. Biol. Chem.* **2001**, *276* (3), 1735–1741. https://doi.org/10.1074/jbc.M006727200.

(77) Parker, D.; Ferreri, K.; Nakajima, T.; LaMorte, V. J.; Evans, R.; Koerber, S. C.; Hoeger, C.; Montminy, M. R. Phosphorylation of CREB at Ser-133 Induces Complex Formation with CREB-Binding Protein via a Direct Mechanism. *Mol. Cell. Biol.* **1996**, *16* (2), 694–703.

(78) Hopkins, A. L.; Groom, C. R. The Druggable Genome. *Nat. Rev. Drug Discov.* **2002**, *1* (9), 727–730. https://doi.org/10.1038/nrd892.

(79) Dunker, A. K.; Uversky, V. N. Drugs for 'Protein Clouds': Targeting Intrinsically Disordered Transcription Factors. *Curr. Opin. Pharmacol.* **2010**, *10* (6), 782–788. https://doi.org/10.1016/j.coph.2010.09.005.

(80) Li, B. X.; Xiao, X. Discovery of a Small-Molecule Inhibitor of the KIX–KID Interaction. *ChemBioChem* **2009**, *10* (17), 2721–2724. https://doi.org/10.1002/cbic.200900552.

(81) Li, B. X.; Gardner, R.; Xue, C.; Qian, D. Z.; Xie, F.; Thomas, G.; Kazmierczak, S. C.; Habecker, B. A.; Xiao, X. Systemic Inhibition of CREB Is Well-Tolerated in Vivo. *Sci. Rep.* **2016**, *6* (1), 34513. https://doi.org/10.1038/srep34513.

(82) Borlikova, G.; Endo, S. Inducible CAMP Early Repressor (ICER) and Brain Functions. *Mol. Neurobiol.* **2009**, *40* (1), 73–86. https://doi.org/10.1007/s12035-009-8072-1.

(83) De Cesare, D.; Fimia, G. M.; Sassone-Corsi, P. Signaling Routes to CREM and CREB: Plasticity in Transcriptional Activation. *Trends Biochem. Sci.* **1999**, *24* (7), 281–285. https://doi.org/10.1016/S0968-0004(99)01414-0.

(84) Mayr, B. M.; Guzman, E.; Montminy, M. Glutamine Rich and Basic Region/Leucine Zipper (BZIP) Domains Stabilize CAMP-Response Element-Binding Protein (CREB) Binding to Chromatin. *J. Biol. Chem.* **2005**, *280* (15), 15103–15110. https://doi.org/10.1074/jbc.M414144200.

(85) Huang, X.; Zhang, J.; Lu, L.; Yin, L.; Xu, M.; Wang, Y.; Zhou, Z.; Sha, J. Cloning and Expression of a Novel CREB MRNA Splice Variant in Human Testis. *Reproduction* **2004**, *128* (6), 775–782. https://doi.org/10.1530/rep.1.00036.

(86) Williams, J. S.; Andrisani, O. M. The Hepatitis B Virus X Protein Targets the Basic Region-Leucine Zipper Domain of CREB. *Proc. Natl. Acad. Sci.* **1995**, *92* (9), 3819–3823. https://doi.org/10.1073/pnas.92.9.3819.

(87) Zhang, X.; Odom, D. T.; Koo, S.-H.; Conkright, M. D.; Canettieri, G.; Best, J.; Chen, H.; Jenner, R.; Herbolsheimer, E.; Jacobsen, E.; Kadam, S.; Ecker, J. R.; Emerson, B.; Hogenesch, J. B.; Unterman, T.; Young, R. A.; Montminy, M. Genome-Wide Analysis of CAMP-Response Element Binding Protein Occupancy, Phosphorylation, and Target Gene Activation in Human Tissues. *Proc. Natl. Acad. Sci.* **2005**, *102* (12), 4459–4464. https://doi.org/10.1073/pnas.0501076102.

(88) Iguchi-Ariga, S. M.; Schaffner, W. CpG Methylation of the CAMP-Responsive Enhancer/Promoter Sequence TGACGTCA Abolishes Specific Factor Binding as Well as Transcriptional Activation. *Genes Dev.* **1989**, *3* (5), 612–619. https://doi.org/10.1101/gad.3.5.612.

(89) Dwarki, V. J.; Montminy, M.; Verma, I. M. Both the Basic Region and the 'Leucine Zipper' Domain of the Cyclic AMP Response Element Binding (CREB) Protein Are Essential for Transcriptional Activation. *EMBO J.* **1990**, *9* (1), 225–232. https://doi.org/10.1002/j.1460-2075.1990.tb08099.x.

(90) Walton, K. M.; Rehfuss, R. P.; Chrivia, J. C.; Lochner, J. E.; Goodman, R. H. A Dominant Repressor of Cyclic Adenosine 3',5'-Monophosphate (CAMP)-Regulated Enhancer-Binding Protein Activity Inhibits the CAMP-Mediated Induction of the Somatostatin Promoter in Vivo. *Mol. Endocrinol.* **1992**, *6* (4), 647–655. https://doi.org/10.1210/mend.6.4.1350057.

(91) Craig, J. C.; Schumacher, M. A.; Mansoor, S. E.; Farrens, D. L.; Brennan, R. G.; Goodman, R. H. Consensus and Variant CAMP-Regulated Enhancers Have Distinct CREB-Binding Properties. *J. Biol. Chem.* **2001**, *276* (15), 11719–11728. https://doi.org/10.1074/jbc.M010263200.

(92) Moll, J. R.; Acharya, A.; Gal, J.; Mir, A. A.; Vinson, C. Magnesium Is Required for Specific DNA Binding of the CREB B-ZIP Domain. *Nucleic Acids Res.* **2002**, *30* (5), 1240–1246.

(93) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. The Ensemble Nature of Allostery. *Nature* **2014**, *508* (7496), 331–339. https://doi.org/10.1038/nature13001.

(94) Groussin, L.; Massias, J. F.; Bertagna, X.; Bertherat, J. Loss of Expression of the Ubiquitous Transcription Factor CAMP Response Element-Binding Protein (CREB) and Compensatory Overexpression of the Activator CREMτ in the Human Adrenocortical Cancer Cell Line H295R. *J. Clin. Endocrinol. Metab.* **2000**, *85* (1), 345–354. https://doi.org/10.1210/jcem.85.1.6307.

(95) Sakamoto, K.; Karelina, K.; Obrietan, K. CREB: A Multifaceted Regulator of Neuronal Plasticity and Protection. *J. Neurochem.* **2011**, *116* (1), 1–9. https://doi.org/10.1111/j.1471-4159.2010.07080.x.

(96) Ortega-Martínez, S. A New Perspective on the Role of the CREB Family of Transcription Factors in Memory Consolidation via Adult Hippocampal Neurogenesis. *Front. Mol. Neurosci.* **2015**, *8*. https://doi.org/10.3389/fnmol.2015.00046.

(97) Benito, E.; Barco, A. CREB's Control of Intrinsic and Synaptic Plasticity: Implications for CREB-Dependent Memory Models. *Trends Neurosci.* **2010**, *33* (5), 230–240. https://doi.org/10.1016/j.tins.2010.02.001.

(98) Nakagawa, S.; Kim, J.-E.; Lee, R.; Malberg, J. E.; Chen, J.; Steffen, C.; Zhang, Y.-J.; Nestler, E. J.; Duman, R. S. Regulation of Neurogenesis in Adult Mouse Hippocampus by CAMP and the CAMP Response Element-Binding Protein. *J. Neurosci.* **2002**, *22* (9), 3673–3682. https://doi.org/10.1523/JNEUROSCI.22-09-03673.2002.

(99) Young, D.; Lawlor, P. A.; Leone, P.; Dragunow, M.; During, M. J. Environmental Enrichment Inhibits Spontaneous Apoptosis, Prevents Seizures and Is Neuroprotective. *Nat. Med.* **1999**, *5* (4), 448–453. https://doi.org/10.1038/7449.

(100) Vogt, M. A.; Inta, D.; Luoni, A.; Elkin, H.; Pfeiffer, N.; Riva, M. A.; Gass, P. Inducible Forebrain-Specific Ablation of the Transcription Factor Creb during Adulthood Induces Anxiety but No Spatial/Contextual Learning Deficits. *Front. Behav. Neurosci.* **2014**, *8*. https://doi.org/10.3389/fnbeh.2014.00407.

(101) Han, M.-H.; Bolaños, C. A.; Green, T. A.; Olson, V. G.; Neve, R. L.; Liu, R.-J.; Aghajanian, G. K.; Nestler, E. J. Role of CAMP Response Element-Binding Protein in the Rat Locus Ceruleus: Regulation of Neuronal Activity and Opiate Withdrawal Behaviors. *J. Neurosci.* **2006**, *26* (17), 4624–4629. https://doi.org/10.1523/JNEUROSCI.4701-05.2006.

(102) Dong, Y.; Green, T.; Saal, D.; Marie, H.; Neve, R.; Nestler, E. J.; Malenka, R. C. CREB Modulates Excitability of Nucleus Accumbens Neurons. *Nat. Neurosci.* **2006**, *9* (4), 475–477. https://doi.org/10.1038/nn1661.

(103) Steven, A.; Seliger, B. Control of CREB Expression in Tumors: From Molecular Mechanisms and Signal Transduction Pathways to Therapeutic Target. *Oncotarget* **2016**, *7* (23), 35454–35465. https://doi.org/10.18632/oncotarget.7721.

(104) Abramovitch, R.; Tavor, E.; Jacob-Hirsch, J.; Zeira, E.; Amariglio, N.; Pappo, O.; Rechavi, G.; Galun, E.; Honigman, A. A Pivotal Role of Cyclic AMP-Responsive Element Binding Protein in Tumor Progression. *Cancer Res.* **2004**, *64* (4), 1338–1346. https://doi.org/10.1158/0008-5472.can-03-2089.

(105) Chhabra, A.; Fernando, H.; Watkins, G.; Mansel, R. E.; Jiang, W. G. Expression of Transcription Factor CREB1 in Human Breast Cancer and Its Correlation with Prognosis. *Oncol. Rep.* **2007**, *18* (4), 953–958. https://doi.org/10.3892/or.18.4.953.

(106) Melnikova, V. O.; Dobroff, A. S.; Zigler, M.; Villares, G. J.; Braeuer, R. R.; Wang, H.; Huang, L.; Bar-Eli, M. CREB Inhibits AP-2alpha Expression to Regulate the Malignant Phenotype of Melanoma. *PloS One* **2010**, *5* (8), e12452. https://doi.org/10.1371/journal.pone.0012452.

(107) Steven, A.; Friedrich, M.; Jank, P.; Heimer, N.; Budczies, J.; Denkert, C.; Seliger, B. What Turns CREB on? And off? And Why Does It Matter? *Cell. Mol. Life Sci.* **2020**. https://doi.org/10.1007/s00018-020-03525-8.

(108) Tan, X.; Wang, S.; Yang, B.; Zhu, L.; Yin, B.; Chao, T.; Zhao, J.; Yuan, J.; Qiang, B.; Peng, X. The CREB-MiR-9 Negative Feedback Minicircuitry Coordinates the Migration and Proliferation of Glioma Cells. *PLOS ONE* **2012**, *7* (11), e49570. https://doi.org/10.1371/journal.pone.0049570.

(109) Meyuhas, R.; Pikarsky, E.; Tavor, E.; Klar, A.; Abramovitch, R.; Hochman, J.; Lago, T. G.; Honigman, A. A Key Role for Cyclic AMP-Responsive Element Binding Protein in Hypoxia-Mediated Activation of the Angiogenesis Factor CCN1 (CYR61) in Tumor Cells. *Mol. Cancer Res.* **2008**, *6* (9), 1397–1409. https://doi.org/10.1158/1541-7786.MCR-07-2086.

(110) Comerford, K. M.; Leonard, M. O.; Karhausen, J.; Carey, R.; Colgan, S. P.; Taylor, C. T. Small Ubiquitin-Related Modifier-1 Modification Mediates Resolution of CREB-Dependent Responses to Hypoxia. *Proc. Natl. Acad. Sci.* **2003**, *100* (3), 986–991. https://doi.org/10.1073/pnas.0337412100.

(111) Özgen, N.; Guo, J.; Gertsberg, Z.; Danilo, P.; Rosen, M. R.; Steinberg, S. F. Reactive Oxygen Species Decrease CAMP Response Element Binding Protein Expression in Cardiomyocytes via a Protein Kinase D1-Dependent Mechanism That Does Not Require Ser133 Phosphorylation. *Mol. Pharmacol.* **2009**, *76* (4), 896–902. https://doi.org/10.1124/mol.109.056473.

(112) Steven, A.; Leisz, S.; Sychra, K.; Hiebl, B.; Wickenhauser, C.; Mougiakakos, D.; Kiessling, R.; Denkert, C.; Seliger, B. Hypoxia-Mediated Alterations and Their Role in the HER-2/Neuregulated CREB Status and Localization. *Oncotarget* **2016**, *7* (32), 52061–52084. https://doi.org/10.18632/oncotarget.10474.

(113) Taylor, C. T.; Furuta, G. T.; Synnestvedt, K.; Colgan, S. P. Phosphorylation-Dependent Targeting of CAMP Response Element Binding Protein to the Ubiquitin/Proteasome Pathway in Hypoxia. *Proc. Natl. Acad. Sci.* **2000**, *97* (22), 12091–12096. https://doi.org/10.1073/pnas.220211797.

(114) Lu, Q.; Hutchins, A. E.; Doyle, C. M.; Lundblad, J. R.; Kwok, R. P. S. Acetylation of CAMP-Responsive Element-Binding Protein (CREB) by CREB-Binding Protein Enhances CREB-Dependent Transcription. *J. Biol. Chem.* **2003**, *278* (18), 15727–15734. https://doi.org/10.1074/jbc.M300546200.

(115) Barral, S.; Reitz, C.; Small, S. A.; Mayeux, R. Genetic Variants in a "cAMP Element Binding Protein" (CREB)-Dependent Histone Acetylation Pathway Influence Memory Performance in Cognitively Healthy Elderly Individuals. *Neurobiol. Aging* **2014**, *35* (12), 2881.e7-2881.e10. https://doi.org/10.1016/j.neurobiolaging.2014.06.024.

(116) Fusco, S.; Ripoli, C.; Podda, M. V.; Ranieri, S. C.; Leone, L.; Toietta, G.; McBurney, M. W.; Schütz, G.; Riccio, A.; Grassi, C.; Galeotti, T.; Pani, G. A Role for Neuronal CAMP Responsive-Element Binding (CREB)-1 in Brain Responses to Calorie Restriction. *Proc. Natl. Acad. Sci.* **2012**, *109* (2), 621–626. https://doi.org/10.1073/pnas.1109237109.

(117) Bartel, D. P. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **2004**, *116* (2), 281–297. https://doi.org/10.1016/S0092-8674(04)00045-5.

(118) Jq, Z.; Qh, Y.; Yq, K.; Y, M.; Lb, Y.; Hd, H.; Jm, C.; T, Y.; Ey, L.; L, L.; Kx, F.; K, Z.; X, X.; Jw, G. Prognostic Value of Coexistence of Abnormal Expression of Micro-RNA-200b and Cyclic Adenosine Monophosphate-Responsive Element-Binding Protein 1 in Human Astrocytoma. *Hum. Pathol.* **2014**, *45* (10), 2154–2161. https://doi.org/10.1016/j.humpath.2014.01.025.

(119) Mittag, T.; Forman-Kay, J. D. Atomic-Level Characterization of Disordered Protein Ensembles. *Curr. Opin. Struct. Biol.* **2007**, *17* (1), 3–14.

(120) Kosol, S.; Contreras-Martos, S.; Cedeno, C.; Tompa, P. Structural Characterization of Intrinsically Disordered Proteins by NMR Spectroscopy. *Mol. Basel Switz.* **2013**, *18* (9), 10802–10828. https://doi.org/10.3390/molecules180910802.

(121) Eliezer, D. Biophysical Characterization of Intrinsically Disordered Proteins. *Curr. Opin. Struct. Biol.* **2009**, *19* (1), 23–30. https://doi.org/10.1016/j.sbi.2008.12.004.

(122) Rauscher, S.; Pomes, R. Molecular Simulations of Protein Disorder. *Biochem. Cell Biol. Biochim. Biol. Cell.* **2010**, *88* (2), 269–290. https://doi.org/10.1139/o09-169.

(123) Karplus, M.; Kuriyan, J. Molecular Dynamics and Protein Function. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (19), 6679–6685.

(124) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem.* **2015**, *119* (16), 5113–5123. https://doi.org/10.1021/jp508971m.

(125) Anandakrishnan, R.; Drozdetski, A.; Walker, R. C.; Onufriev, A. V. Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations. *Biophys. J.* **2015**, *108* (5), 1153–1164. https://doi.org/10.1016/j.bpj.2014.12.047.

(126) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a Molecular Dynamics Force Field for Both Folded and Disordered Protein States. *Proc. Natl. Acad. Sci.* **2018**, *115* (21), E4758–E4766. https://doi.org/10.1073/pnas.1800690115.
(127) Leavitt, S.; Freire, E. Direct Measurement of Protein Binding Energetics by Isothermal Titration Calorimetry. *Curr. Opin. Struct. Biol.* **2001**, *11* (5), 560–566. https://doi.org/10.1016/S0959-440X(00)00248-7.
(128) Gibbs, E. B.; Showalter, S. A. Quantitative Biophysical Characterization of Intrinsically Disordered Proteins. *Biochemistry* **2015**, *54* (6), 1314–1326. https://doi.org/10.1021/bi501460a.
(129) Cheow, L. F.; Viswanathan, R.; Chin, C.-S.; Jennifer, N.; Jones, R. C.; Guccione, E.; Quake, S. R.; Burkholder, W. F. Multiplexed Analysis of Protein–Ligand Interactions by Fluorescence Anisotropy in a Microfluidic Platform. *Anal. Chem.* **2014**, *86* (19), 9901–9908. https://doi.org/10.1021/ac502605f.
(130) Shammas, S. L.; Crabtree, M. D.; Dahal, L.; Wicky, B. I. M.; Clarke, J. Insights into Coupled Folding and Binding Mechanisms from Kinetic Studies. *J. Biol. Chem.* **2016**, *291* (13), 6689–6695. https://doi.org/10.1074/jbc.R115.692715.
(131) Crabtree, M. D.; Shammas, S. L. Chapter Fourteen - Stopped-Flow Kinetic Techniques for Studying Binding Reactions of Intrinsically Disordered Proteins. In *Methods in Enzymology*; Rhoades, E., Ed.; Intrinsically Disordered Proteins; Academic Press, 2018; Vol. 611, pp 423–457. https://doi.org/10.1016/bs.mie.2018.09.026.
(132) Zanetti-Domingues, L. C.; Tynan, C. J.; Rolfe, D. J.; Clarke, D. T.; Martin-Fernandez, M. Hydrophobic Fluorescent Probes Introduce Artifacts into Single Molecule Tracking Experiments Due to Non-Specific Binding. *PLoS ONE* **2013**, *8* (9). https://doi.org/10.1371/journal.pone.0074200.
(133) Jia, Y.; Kumar, A.; Patel, S. S. Equilibrium and Stopped-Flow Kinetic Studies of Interaction between T7 RNA Polymerase and Its Promoters Measured by Protein and 2-Aminopurine Fluorescence Changes. *J. Biol. Chem.* **1996**, *271* (48), 30451–30458. https://doi.org/10.1074/jbc.271.48.30451.
(134) Fersht, A. R.; Sato, S. Φ-Value Analysis and the Nature of Protein-Folding Transition States. *Proc. Natl. Acad. Sci.* **2004**, *101* (21), 7976–7981. https://doi.org/10.1073/pnas.0402684101.
(135) Vogt, A. D.; Di Cera, E. Conformational Selection or Induced-Fit? A Critical Appraisal of the Kinetic Mechanism. *Biochemistry* **2012**, *51* (30), 5894–5902. https://doi.org/10.1021/bi3006913.
(136) Grasso, G.; Leanza, L.; Morbiducci, U.; Danani, A.; Deriu, M. A. Aminoacid Substitutions in the Glycine Zipper Affect the Conformational Stability of Amyloid Beta Fibrils. *J. Biomol. Struct. Dyn.* **2020**, *38* (13), 3908–3915. https://doi.org/10.1080/07391102.2019.1671224.
(137) Grimm, J. B.; English, B. P.; Chen, J.; Slaughter, J. P.; Zhang, Z.; Revyakin, A.; Patel, R.; Macklin, J. J.; Normanno, D.; Singer, R. H.; Lionnet, T.; Lavis, L. D. A General Method to Improve Fluorophores for Live-Cell and Single-Molecule Microscopy. *Nat. Methods* **2015**, *12* (3), 244–250. https://doi.org/10.1038/nmeth.3256.
(138) Gill, S. C.; von Hippel, P. H. Calculation of Protein Extinction Coefficients from Amino Acid Sequence Data. *Anal. Biochem.* **1989**, *182* (2), 319–326. https://doi.org/10.1016/0003-2697(89)90602-7.
(139) Sosnick Group. Denaturant Concentration Calculator http://sosnick.uchicago.edu/gdmcl.html.
(140) Kawahara, K.; Tanford, C. Viscosity and Density of Aqueous Solutions of Urea and Guanidine Hydrochloride. *J. Biol. Chem.* **1966**, *241* (13), 3228–3232.

(141) Nikolovska-Coleska, Z.; Wang, R.; Fang, X.; Pan, H.; Tomita, Y.; Li, P.; Roller, P. P.; Krajewski, K.; Saito, N. G.; Stuckey, J. A.; Wang, S. Development and Optimization of a Binding Assay for the XIAP BIR3 Domain Using Fluorescence Polarization. *Anal. Biochem.* **2004**, *332* (2), 261–273. https://doi.org/10.1016/j.ab.2004.05.055.

(142) Young, G.; Hundt, N.; Cole, D.; Fineberg, A.; Andrecka, J.; Tyler, A.; Olerinyova, A.; Ansari, A.; Marklund, E. G.; Collier, M. P.; Chandler, S. A.; Tkachenko, O.; Allen, J.; Crispin, M.; Billington, N.; Takagi, Y.; Sellers, J. R.; Eichmann, C.; Selenko, P.; Frey, L.; Riek, R.; Galpin, M. R.; Struwe, W. B.; Benesch, J. L. P.; Kukura, P. Quantitative Mass Imaging of Single Biological Macromolecules. *Science* **2018**, *360* (6387), 423–427. https://doi.org/10.1126/science.aar5839.

(143) Los, G. V.; Encell, L. P.; McDougall, M. G.; Hartzell, D. D.; Karassina, N.; Zimprich, C.; Wood, M. G.; Learish, R.; Ohana, R. F.; Urh, M.; Simpson, D.; Mendez, J.; Zimmerman, K.; Otto, P.; Vidugiris, G.; Zhu, J.; Darzins, A.; Klaubert, D. H.; Bulleit, R. F.; Wood, K. V. HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* **2008**, *3* (6), 373–382. https://doi.org/10.1021/cb800025k.

(144) Hartzell, D. D.; Trinklein, N. D.; Mendez, J.; Murphy, N.; Aldred, S. F.; Wood, K.; Urh, M. A Functional Analysis of the CREB Signaling Pathway Using HaloCHIP-Chip and High Throughput Reporter Assays. *BMC Genomics* **2009**, *10* (1), 497. https://doi.org/10.1186/1471-2164-10-497.

(145) Sugo, N.; Morimatsu, M.; Arai, Y.; Kousoku, Y.; Ohkuni, A.; Nomura, T.; Yanagida, T.; Yamamoto, N. Single-Molecule Imaging Reveals Dynamics of CREB Transcription Factor Bound to Its Target Sequence. *Sci. Rep.* **2015**, *5* (1), 1–9. https://doi.org/10.1038/srep10662.

(146) Lopez, D. I.; Mick, J. E.; Nyborg, J. K. Purification of CREB to Apparent Homogeneity: Removal of Truncation Products and Contaminating Nucleic Acid. *Protein Expr. Purif.* **2007**, *55* (2), 406–418. https://doi.org/10.1016/j.pep.2007.06.011.

(147) Hoeffler, J. P.; Lustbader, J. W.; Chen, C.-Y. Identification of Multiple Nuclear Factors That Interact with Cyclic Adenosine 3',5'-Monophosphate Response Element-Binding Protein and Activating Transcription Factor-2 by Protein-Protein Interactions. *Mol. Endocrinol.* **1991**, *5* (2), 256–266. https://doi.org/10.1210/mend-5-2-256.

(148) Richards, J. P.; Bächinger, H. P.; Goodman, R. H.; Brennan, R. G. Analysis of the Structural Properties of CAMP-Responsive Element-Binding Protein (CREB) and Phosphorylated CREB. *J. Biol. Chem.* **1996**, *271* (23), 13716–13723. https://doi.org/10.1074/jbc.271.23.13716.

(149) Lundblad, J. R.; Kwok, R. P. S.; Laurance, M. E.; Huang, M. S.; Richards, J. P.; Brennan, R. G.; Goodman, R. H. The Human T-Cell Leukemia Virus-1 Transcriptional Activator Tax Enhances CAMP-Responsive Element-Binding Protein (CREB) Binding Activity through Interactions with the DNA Minor Groove. *J. Biol. Chem.* **1998**, *273* (30), 19251–19259. https://doi.org/10.1074/jbc.273.30.19251.

(150) Kitagawa, H.; Sugo, N.; Yamamoto, N. Live-Cell Single-Molecule Imaging with Optogenetics Reveals Dynamics of a Neuronal Activity-Dependent Transcription Factor. In *Single Molecule Microscopy in Neurobiology*; Yamamoto, N., Okada, Y., Eds.; Neuromethods; Springer US: New York, NY, 2020; pp 59–79. https://doi.org/10.1007/978-1-0716-0532-5_4.

(151) McPherson, J. D.; Marra, M.; Hillier, L.; Waterston, R. H.; Chinwalla, A.; Wallis, J.; Sekhon, M.; Wylie, K.; Mardis, E. R.; Wilson, R. K.; Fulton, R.; Kucaba, T. A.; Wagner-McPherson, C.; Barbazuk, W. B.; Gregory, S. G.; Humphray, S. J.; French, L.; Evans, R. S.; Bethel, G.; Whittaker, A.; Holden, J. L.; McCann, O. T.; Dunham, A.; Soderlund, C.; Scott, C. E.; Bentley, D. R.; Schuler, G.; Chen, H.-C.; Jang, W.; Green, E. D.; Idol, J. R.; Maduro, V. V. B.; Montgomery, K. T.; Lee, E.; Miller, A.; Emerling, S.; Kucherlapati, R.; Gibbs, R.; Scherer, S.; Gorrell, J. H.; Sodergren, E.; Clerc-Blankenburg, K.; Tabor, P.; Naylor, S.; Garcia, D.; de Jong, P. J.; Catanese, J. J.; Nowak, N.; Osoegawa, K.; Qin, S.; Rowen, L.; Madan, A.; Dors, M.; Hood,

L.; Trask, B.; Friedman, C.; Massa, H.; Cheung, V. G.; Kirsch, I. R.; Reid, T.; Yonescu, R.; Weissenbach, J.; Bruls, T.; Heilig, R.; Branscomb, E.; Olsen, A.; Doggett, N.; Cheng, J.-F.; Hawkins, T.; Myers, R. M.; Shang, J.; Ramirez, L.; Schmutz, J.; Velasquez, O.; Dixon, K.; Stone, N. E.; Cox, D. R.; Haussler, D.; Kent, W. J.; Furey, T.; Rogic, S.; Kennedy, S.; Jones, S.; Rosenthal, A.; Wen, G.; Schilhabel, M.; Gloeckner, G.; Nyakatura, G.; Siebert, R.; Schlegelberger, B.; Korenberg, J.; Chen, X.-N.; Fujiyama, A.; Hattori, M.; Toyoda, A.; Yada, T.; Park, H.-S.; Sakaki, Y.; Shimizu, N.; Asakawa, S.; Kawasaki, K.; Sasaki, T.; Shintani, A.; Shimizu, A.; Shibuya, K.; Kudoh, J.; Minoshima, S.; Ramser, J.; Seranski, P.; Hoff, C.; Poustka, A.; Reinhardt, R.; Lehrach, H.; The International Human Genome Mapping Consortium; Washington University School of Medicine, G. S. C.; Wellcome Trust Genome Campus:; National Center for Biotechnology Information:; National Human Genome Research Insititute:; Albert Einstein College of Medicine:; Baylor College of Medicine, H. G. S. C.; Roswell Park Cancer Institute:; Multimegabase Sequencing Center:; Fred Hutchinson Cancer Research Institute:; The Children's Hospital of Philadelphia:; Genoscope:; US DOE Joint Genome Institute:; Stanford Human Genome Center and Department of Genetics:; University of California, S. C.; British Columbia Cancer Research Centre:; Department of Genome Analysis, I. of M. B.; Departments of Human Genetics and Pediatrics, U. of C.; RIKEN Genomic Sciences Center:; Department of Molecular Biology, K. U. S. of M.; Max-Planck-Institute for Molecular Genetics: A Physical Map of the Human Genome. *Nature* **2001**, *409* (6822), 934–941. https://doi.org/10.1038/35057157.

(152) Graether, S. P. Troubleshooting Guide to Expressing Intrinsically Disordered Proteins for Use in NMR Experiments. *Front. Mol. Biosci.* **2019**, *5*. https://doi.org/10.3389/fmolb.2018.00118.

(153) Samelson, A. J.; Bolin, E.; Costello, S. M.; Sharma, A. K.; O'Brien, E. P.; Marqusee, S. Kinetic and Structural Comparison of a Protein's Cotranslational Folding and Refolding Pathways. *Sci. Adv.* **2018**, *4* (5), eaas9098. https://doi.org/10.1126/sciadv.aas9098.

(154) Burnham, B.; Nass, S.; Kong, E.; Mattingly, M.; Woodcock, D.; Song, A.; Wadsworth, S.; Cheng, S. H.; Scaria, A.; O'Riordan, C. R. Analytical Ultracentrifugation as an Approach to Characterize Recombinant Adeno-Associated Viral Vectors. *Hum. Gene Ther. Methods* **2015**, *26* (6), 228–242. https://doi.org/10.1089/hgtb.2015.048.

(155) Schultz, J. C.; Hack, C. A.; Benner, W. H. Mass Determination of Megadalton-DNA Electrospray Ions Using Charge Detection Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **1998**, *9* (4), 305–313. https://doi.org/10.1016/S1044-0305(97)00290-0.

(156) Cheng, X.; Morin, P. E.; Harms, A. C.; Bruce, J. E.; Ben-David, Y.; Smith, R. D. Mass Spectrometric Characterization of Sequence-Specific Complexes of DNA and Transcription Factor PU.1 DNA Binding Domain. *Anal. Biochem.* **1996**, *239* (1), 35–40. https://doi.org/10.1006/abio.1996.0287.

(157) Hofstadler, S. A.; Griffey, R. H. Analysis of Noncovalent Complexes of DNA and RNA by Mass Spectrometry. *Chem. Rev.* **2001**, *101* (2), 377–390. https://doi.org/10.1021/cr990105o.

(158) Hanson, C. L.; Robinson, C. V. Protein-Nucleic Acid Interactions and the Expanding Role of Mass Spectrometry. *J. Biol. Chem.* **2004**, *279* (24), 24907–24910. https://doi.org/10.1074/jbc.R300037200.

(159) Kapur, A.; Beck, J. L.; Brown, S. E.; Dixon, N. E.; Sheil, M. M. Use of Electrospray Ionization Mass Spectrometry to Study Binding Interactions between a Replication Terminator Protein and DNA. *Protein Sci.* **2002**, *11* (1), 147–157. https://doi.org/10.1110/ps.27702.

(160) Ahn, S.; Olive, M.; Aggarwal, S.; Krylov, D.; Ginty, D. D.; Vinson, C. A Dominant-Negative Inhibitor of CREB Reveals That It Is a General Mediator of Stimulus-Dependent Transcription of c-Fos. *Mol. Cell. Biol.* **1998**, *18* (2), 967–977. https://doi.org/10.1128/MCB.18.2.967.

(161) Santiago-Rivera, Z. I.; Gorenstein, D. G.; Williams, J. S.; Andrisani, O. M. Bacterial expression and characterization of the CREB bZip module: Circular dichroism and 2D 1H-NMR studies. *Protein Sci.* **1993**, *2* (9), 1461–1471. https://doi.org/10.1002/pro.5560020910.
(162) Wu, X.; Spiro, C.; Owen, W. G.; McMurray, C. T. CAMP Response Element-Binding Protein Monomers Cooperatively Assemble to Form Dimers on DNA. *J. Biol. Chem.* **1998**, *273* (33), 20820–20827. https://doi.org/10.1074/jbc.273.33.20820.
(163) Casey, J. R.; Grinstein, S.; Orlowski, J. Sensors and Regulators of Intracellular PH. *Nat. Rev. Mol. Cell Biol.* **2010**, *11* (1), 50–61. https://doi.org/10.1038/nrm2820.
(164) Kelly, C.; Kuravsky, M.; Redfield, C.; Shammas, S. L. The Dynamic Search Mode of a Disordered Transcription Factor. *Biophys. J.* **2020**, *118* (3), 6a. https://doi.org/10.1016/j.bpj.2019.11.224.
(165) Almagor, M.; Cole, R. D. In physiological salt conditions the core proteins of the nucleosomes in large chromatin fragments denature at 73 degrees C and the DNA unstacks at 85 degrees C - PubMed https://pubmed.ncbi.nlm.nih.gov/2703503/ (accessed Oct 6, 2020).
(166) Terry, C. A.; Fernández, M.-J.; Gude, L.; Lorente, A.; Grant, K. B. Physiologically Relevant Concentrations of NaCl and KCl Increase DNA Photocleavage by an N-Substituted 9-Aminomethylanthracene Dye. *Biochemistry* **2011**, *50* (47), 10375–10389. https://doi.org/10.1021/bi200972c.
(167) Gotoh, H.; Kajikawa, M.; Kato, H.; Suto, K. Intracellular Mg2+ Surge Follows Ca2+ Increase during Depolarization in Cultured Neurons. *Brain Res.* **1999**, *828* (1–2), 163–168. https://doi.org/10.1016/s0006-8993(99)01298-6.
(168) Yun, Y.; Dumoulin, M.; Habener, J. F. DNA-Binding and Dimerization Domains of Adenosine 3', 5'- Cyclic Monophosphate-Responsive Protein CREB Reside in the Carboxyl-Terminal 66 Amino Acids. *Mol. Endocrinol.* **1990**, *4* (6), 931–939. https://doi.org/10.1210/mend-4-6-931.
(169) Costantini, L. M.; Fossati, M.; Francolini, M.; Snapp, E. L. Assessing the Tendency of Fluorescent Proteins to Oligomerize under Physiologic Conditions. *Traffic Cph. Den.* **2012**, *13* (5), 643–649. https://doi.org/10.1111/j.1600-0854.2012.01336.x.
(170) Schreiber, G.; Fersht, A. R. Energetics of Protein-Protein Interactions: Analysis Ofthe Barnase-Barstar Interface by Single Mutations and Double Mutant Cycles. *J. Mol. Biol.* **1995**, *248* (2), 478–486. https://doi.org/10.1016/S0022-2836(95)80064-6.
(171) Yin, M. J.; Gaynor, R. B. Complex Formation between CREB and Tax Enhances the Binding Affinity of CREB for the Human T-Cell Leukemia Virus Type 1 21-Base-Pair Repeats. *Mol. Cell. Biol.* **1996**, *16* (6), 3156–3168. https://doi.org/10.1128/MCB.16.6.3156.
(172) Hippel, P. H. von; Berg, O. G. Facilitated Target Location in Biological Systems. *J. Biol. Chem.* **1989**, *264* (2), 675–678.
(173) Riggs, A. D.; Bourgeois, S.; Cohn, M. The Lac Represser-Operator Interaction: III. Kinetic Studies. *J. Mol. Biol.* **1970**, *53* (3), 401–417. https://doi.org/10.1016/0022-2836(70)90074-4.
(174) Stracy, M.; Schweizer, J.; Sherratt, D. J.; Kapanidis, A. N.; Uphoff, S.; Lesterlin, C. *Transient Non-Specific DNA Binding Dominates the Target Search of Bacterial DNA-Binding Proteins*; preprint; Cell Biology, 2020. https://doi.org/10.1101/2020.08.13.249771.
(175) Go, Y.-M.; Jones, D. P. Redox Compartmentalization in Eukaryotic Cells. *Biochim. Biophys. Acta* **2008**, *1780* (11), 1273–1290. https://doi.org/10.1016/j.bbagen.2008.01.011.