

Target search mode of a disordered transcription factor



Conor Francis Kelly
Balliol College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Hilary 2021

Abstract

Protein disorder is prevalent in proteins which are involved in DNA binding, such as transcription factors. Transcription factors commonly contain disorder within protein-protein interaction interfaces, however a small number of families contain disorder in their DNA binding domain such as the basic leucine zipper proteins (bZIP). bZIP proteins like cAMP response element binding protein (CREB) have a wide range of roles in human health, however the role of disorder in how bZIP proteins find their cognate DNA site within the nucleus is still unknown. In addition, significant controversy exists about whether bZIP proteins search for the target as a preformed dimer or by sequential association of monomers. Using a range of biophysical techniques including stopped flow kinetics and nuclear magnetic resonance spectroscopy with constructs to mimic the monomer or dimer species, this question has been examined. The CREB homodimer affinity was found to be high, at the upper limit of what had been previously seen for bZIPs. CREB is able to recognise the cognate DNA CRE motif specifically as a preformed dimer and undergoes coupled folding upon binding. In the search mode, however, the basic region remains dynamic with more disordered character. Compared to the dimeric construct, a monomeric CREB model exhibited a lack of specificity for the cognate site and several orders of magnitude reduction in affinity. The affinity for DNA by the monomer appears to be dependent on the propagation of helical content into the DNA-binding basic region, which directly contacts the DNA, from the leucine zipper in the dimer. CREB bZIP also searches for its target DNA using a combination of intersegmental transfer and sliding, with the residual helical content in the basic region controlling these rates. CREB therefore appears to perform effective target search preferentially through the pre-formed dimer pathway, independent of the presence of excess competitor DNA which is consistent with the homodimer stability, but can also proceed through the monomer pathway albeit with a reduced affinity for the target. This work therefore extends our understanding towards describing the target search mode and mechanism by a disordered transcription factor.

Acknowledgements

Firstly I would like to thank Dr Sarah Shammass, Dr Alice Bowen and Prof Christina Redfield for their amazing support and guidance throughout this process. With Sarah, being able to share the experience of building a new lab group and for the trust placed in me throughout the years, I am extremely grateful. The backbone of this project has been the NMR, which would really not have been possible without Christina's passion and enthusiasm for the project. Spending many hours discussing the data and having fun in the process was really enlightening for me. And finally for Alice, who where even though the project did not work out, provided a great input in the early stages of the project and provided great personal support throughout. I have learned a significant amount across the past four years or so and admire you all.

Leaving the Shammass Lab as its longest standing member is a strange feeling, seeing the lab grow to where it is today. A big thanks must go to the current and previous members of the lab for the help and discussions during my time here: Mikhail, Iris, Chris, Jeremie, Marina, Matt, Wilson and Lasse. Members of the Department of Biochemistry that have provided support are Dr David Staunton, Profs Jason Schnell, Jane Mellor and Tim Nott for their helpful advice and discussions or experimental support. In the very early days of the project I was heavily supported by members of Prof Ben Berk's, Dr Tim Nott and Prof Penny Handford's group for which I am very grateful for. I would like to thank the members of Balliol College who provided academic support: Prof's Sebastian Shimeld and Tom Melham as well as Rev Bruce Kinsey.

This whole process would not have been possible without the amazing support and guidance from my parents - I am forever grateful for this and hope that I do them proud! My family and friends back home have supported and pushed me on those tough days. On those long frustrating days, my friend Claire provided an ear to complain to and without, Oxford would have been a different place. Thank you.

Declaration of own work

The work presented in this Thesis is my own, aside from a few experiments. Dr Mikhail Kuravsky performed the protein/DNA equilibrium and kinetic (k_{on} and k_{off}) assays for CREB 277-339 (pWT and FD) as well as the CREB 277-315/DNA binding assays. Dr Kuravsky also designed the V288I/M291R construct. Mr Wilson Bridge performed the CREB 285-315/DNA binding assays, under my supervision. Mr Lasse Staby designed 5 of the 10 probe DNA constructs for use in the target search kinetics. Initial NMR data collection and processing was performed with help from Prof Christina Redfield (CREB 285-339 apo and CRE DNA complexes). Dr Sarah Shammass designed the binding equations used throughout this thesis.

Contents

List of Figures	viii
List of Abbreviations	xiv
1 Introduction	1
1.1 The Central Dogma of Biology	2
1.2 Genomic regulation at the level of transcription	2
1.2.1 DNA packaging within the nucleus	2
1.2.2 Transcription factors regulate gene expression	4
1.3 Disordered Proteins	10
1.4 Protein disorder in transcription factors	13
1.5 bZIP family is disordered and the pathway of association to target DNA is controversial	15
1.5.1 cAMP Response Element Binding Protein	18
1.6 Pathway for target association by bZIPs can occur via two routes .	23
1.6.1 Dimerisation competent bZIP constructs remain specific for cognate DNA sequences	23
1.6.2 Analysis of monomeric constructs unveils detail for the monomer pathway	25
1.6.3 Forced dimer bZIP constructs to investigate the pre-formed dimer pathway	27
1.7 Target search process of transcription factors	27
1.7.1 Target search mode of bZIP proteins	30
1.8 Aims of the thesis	30
2 Materials and Methods	32
2.1 Buffers and Reagents	33
2.2 Chemically competent <i>Escherichia Coli</i>	35
2.3 Cloning and mutagenesis	36
2.4 DNA used for biophysics experiments	40
2.5 Elongation of probe DNA for sliding search rate kinetics	40
2.6 Expression of His ₆ -GB1-CREB bZIP, incorporating isotope enrich- ment for NMR	46

2.7	Purification of CREB bZIP	47
2.8	Protein concentration determination and buffer exchange	48
2.9	Nuclear Magnetic Resonance of CREB with DNA	50
2.9.1	NMR data acquisition parameters	50
2.9.2	Analysis of DNA titrations using chemical shift perturbation	54
2.9.3	Assignment of CREB complexes and determination of secondary structure and dynamics	54
2.10	Urea denaturation at equilibrium	55
2.11	Thermal denaturation at equilibrium	55
2.12	Labelling of CREB with extrinsic fluorophores	56
2.13	Fluorescence measurements of CREB and CREB/DNA complexes	56
2.13.1	Homodimer affinity with labelled CREB bZIP	57
2.13.2	CREB equilibrium affinity with fluorescently labelled DNA	57
2.13.3	Isothermal titration calorimetry (ITC)	58
2.14	Circular Dichroism (CD) for coupled folding upon binding	58
2.15	Kinetics	59
2.15.1	Folding and unfolding of bZIP dimer	59
2.15.2	k_{on} and k_{off} of protein to DNA and target search model	60
2.16	Bioinformatics	61
2.16.1	Contact map analysis and structural alignments	61
2.16.2	Promoter flanking analysis	62
3	Characterisation of the dimeric state of the CREB bZIP domain	69
3.1	Introduction	69
3.2	Results	72
3.2.1	Kinetics identified a reliable and low K_d for the homodimer	72
3.2.2	Dilution of CREB bZIP using labelled protein constructs	77
3.2.3	Structural characterisation of the dimer	87
3.3	Discussion	109
3.3.1	Homodimer affinity estimate was consistent between a variety of techniques	109
3.3.2	DNA binding region remains dynamic in the apo-state of the bZIP complex	112
4	Investigation of the dimer pathway of CREB association to the target site	115
4.1	Introduction	115
4.2	Results	117
4.2.1	Design of a "forced" dimer construct	117
4.2.2	Equilibrium specificity of Dimer to DNA	118

4.2.3	Kinetics of CREB associating to target containing DNA. . .	124
4.2.4	Structural changes of bZIP interacting with DNA	131
4.2.5	Varied helical content in basic region of CREB 285-339 in complex with different DNA sequences	132
4.2.6	More helical bZIP results in a non-target conformational change that is similar to that bound to target DNA.	141
4.3	Discussion	144
5	Investigations into a monomer only model reveals a lack of target DNA specificity	152
5.1	Introduction	152
5.2	Results	154
5.2.1	Design for the monomer construct	154
5.2.2	Binding of monomer model to DNA	155
5.2.3	Structural characterisation of the monomer	156
5.2.4	non-CRE DNA and monomeric CREB	163
5.2.5	Monomer model with target DNA	171
5.2.6	Decoupling of the leucine zipper from the basic region	178
5.3	Discussion	183
6	Productive target search of CREB is enhanced by disorder	188
6.1	Introduction	188
6.2	Results	191
6.2.1	Development of kinetic strategy for determining target search parameters	191
6.2.2	Characterisation of target search of CREB bZIP	201
6.2.3	Computational and structural examination of the effective sliding length of CREB bZIP	209
6.3	Discussion	214
7	Discussion and future work	223
7.1	Summary of main results	223
7.2	Future work	228
7.3	COVID and Mitigating Circumstances	230
Appendices		
References		300

List of Figures

1.1	DNA is packaged in supermolecular structures <i>in vivo</i> within the nucleus.	3
1.2	Transcription is regulated by a large complex of general and specific transcription factors.	5
1.3	Transcription factors can directly and indirectly read a DNA sequence.	7
1.4	Three major structural families of transcription factors.	9
1.5	α -synuclein is a well characterised model IDP.	11
1.6	Four families of transcription factors contain disorder in the DNA binding domain.	15
1.7	Basic region is disordered and leucine zipper is a coiled coil.	17
1.8	bZIP domain of CREB is located at the C-terminus and contains one of the predicted ordered regions.	19
1.9	Multiple signalling pathways converge on CREB for regulation of transcriptional response.	21
1.10	Simplified model of the possible association pathways of a bZIP protein (CREB) to DNA.	24
1.11	GCN4 was shown to have high affinity for two cognate DNA sequences.	24
1.12	Target search of proteins to their cognate DNA site occurs via multiple macro- and microdissociation events, coupled with sliding, hopping and intersegmental transfer.	28
3.1	Simplified model of the possible association pathways of CREB bZIP to DNA.	70
3.2	Cooperative folding model and energy diagram for CREB bZIP folding	73
3.3	Folding of CREB bZIP can be estimated using stopped flow kinetics.	74
3.4	Rates of folding/unfolding and K_d values were extracted from kinetics.	75
3.5	Designing the location of site directed cysteine mutagenesis in CREB bZIP pWT.	79
3.6	Labelling trials of CREB S337C.	80
3.7	Successful labelling of CREB bZIP S337C with CPM.	82
3.8	Homodimer K_d of labelled CREB bZIP shows a tight interaction.	84
3.9	Homodimer K_d estimates for CREB bZIP pWT.	85

3.10	Mg ²⁺ dependence on homodimer conformation.	85
3.11	High protein concentration dependence on homodimer stability . . .	86
3.12	Ensemble structural description of CREB bZIP pWT by CD.	87
3.13	Optimisation of expression for isotopically labelled CREB.	89
3.14	CREB bZIP pWT assignments at pH 7.4	91
3.15	CREB bZIP pWT assignments using NOESY.	92
3.16	pH titration of CREB bZIP using NMR.	93
3.17	Results from pH titration by NMR.	94
3.18	CREB assignments at two pH values.	96
3.19	Triple resonance assignments of CREB bZIP.	97
3.20	Sequential side chain proton assignments of CREB bZIP.	97
3.21	¹³ C-HSQC showed lack of dispersion in BR peaks.	99
3.22	Helical propensity prediction shows two helical regions.	101
3.23	Secondary chemical shifts from assignments under two pH values. .	102
3.24	Secondary structure predictions and information on dynamics of CREB bZIP pWT.	103
3.25	Comparison of SSP between two bZIP proteins.	105
3.26	Extension of N-terminus increases predicted helical propensity in basic region.	106
3.27	Refolding kinetics of CREB bZIP 277-339	107
3.28	Assigned spectra of CREB bZIP 277-339.	108
3.29	Structural and dynamic properties altered by an N-terminal extension.	109
4.1	Simplified model of the possible association pathways of CREB bZIP to DNA, highlighting the dimer pathway.	116
4.2	Design of the forced dimer CREB bZIP construct.	118
4.3	Equilibrium binding curves for CREB 285-339 pWT demonstrate high specificity of binding to CRE.	119
4.4	CREB bZIP with CRE is stable across varying ionic strengths. . . .	120
4.5	Specificity for target DNA was increased with a combination of forced dimerisation and extension of the N-terminus.	121
4.6	A larger enthalpy change for target DNA association was seen for CREB bZIP.	122
4.7	Thermodynamics of CREB interacting with DNA shows a preference for target containing sequences.	123
4.8	Comparison of the K_d derived from two equilibrium state experimen- tal techniques.	124
4.9	k_{on} of CREB 285-339 FD to target containing DNA can be fitted to a single exponential.	126

4.10	Association rates to target DNA sites are faster for the pWT than for the FD.	127
4.11	k_{obs} could be determined from a single exponential fit to stopped flow dissociation experiments.	129
4.12	k_{off} was markedly different between pWT and FD CREB 285-339 constructs.	129
4.13	Comparison of K_d estimates of CREB 285-339 to CRE-DNA between those obtained from equilibrium and kinetics.	131
4.14	Secondary structure of CREB increases upon addition of DNA. . .	132
4.15	Peak dispersion seen of CREB bZIP in association with CRE-DNA.	133
4.16	Increased dispersion in ^1H dimension of CREB 285-339 upon addition of CRE-DNA.	134
4.17	Carbon chemical shifts of CREB 285-339 in complex with CRE DNA did not vary with temperature.	135
4.18	CREB 285-339 is in fast exchange with the non-CRE DNA.	136
4.19	CREB 285-339 with non-CRE DNA assignments at 25°C.	137
4.20	CREB 285-339 is more helical and less dynamic when in complex with DNA.	138
4.21	Arginine C δ region provides detail on the level of specific interaction.	140
4.22	CREB 277-339 was assigned in complex with CRE DNA at 35°C. .	142
4.23	CREB 277-339 was assigned in complex with non-CRE DNA at 35°C.	143
4.24	The N-terminal extended construct still retains a disordered and dynamic basic region in complex with non-target DNA.	144
4.25	Location of the assigned arginine C δ peaks in CREB 285-339 with CRE DNA complex.	150
5.1	Simplified model of the possible association pathways of CREB bZIP to DNA, highlighting the monomer pathway.	153
5.2	Monomeric CREB design.	154
5.3	A weak, non-specific interaction was seen with monomeric CREB to target DNA.	155
5.4	CREB 285-315 assignments show disorder.	156
5.5	Side chain carbon chemical shifts with CC(CO)NH on CREB 285-315	158
5.6	NMR temperature titration of CREB 285-315.	159
5.7	Assigned ^{15}N -HSQC of CREB 277-315.	160
5.8	The monomer model was dynamic and disordered.	162
5.9	Non-CRE DNA titration with CREB 285-315.	164
5.10	CREB 285-315 with non-CRE DNA assignments.	165
5.11	non-CRE DNA binding curve with CREB 285-315 by NMR.	166
5.12	Assignment of CREB 277-315 with non-CRE DNA.	167

5.13	CREB 285-315 remains disordered associated to non-CRE DNA. . .	169
5.14	CREB 277-315 remains disordered associated to non-CRE.	170
5.15	Addition of CRE DNA to the monomer model reveals different exchange regime.	171
5.16	CREB 285-315 is in slow/intermediate exchange with CRE-DNA. .	172
5.17	Limited chemical shift perturbation with CREB 285-315 during a CRE DNA titration due to a reduction in peak intensity.	173
5.18	Binding affinity of monomeric CREB model to CRE and non-CRE DNA with NMR.	174
5.19	CREB 285-315 in CRE-DNA complex backbone amide peaks reappear at higher temperature.	174
5.20	Minimal chemical shift perturbation of CREB 285-315 in complex with CRE-DNA at different temperatures.	175
5.21	No major change in secondary structure or dynamics of monomeric CREB when associated to DNA.	176
5.22	Specific interactions were observed by CREB 277-315 with CRE-DNA over non-CRE DNA.	178
5.23	Basic region residues in a decoupled bZIP are as disordered as a monomer.	179
5.24	Carbon chemical shifts show basic region in decoupled bZIP are disordered.	180
5.25	Lack of coupled folding upon binding and a weak interaction of decoupled bZIP to DNA.	182
5.26	Accuracy of Agadir prediction is closer for the basic region only constructs.	183
6.1	Illustration of the stopped flow set-up and experimental parameters that can be extracted.	192
6.2	Probe DNA purity.	194
6.3	Experimental setup for the stopped flow experiments to determine the target search parameters for CREB bZIP.	195
6.4	Stopped flow of CREB mixed with probe and competitor DNA was challenging to fit to a single or double exponential fit.	196
6.5	Significant slowing of CREB target search was seen in the presence of competitor DNA.	197
6.6	k_{app} was increased for CREB 285-339 FD associating to various lengths of CRE-containing probe DNA.	199
6.7	Ionic strength increases the rate of 1D diffusion target search in the presence of competitor DNA.	200

6.8	A linear dependence on the k_{app} was seen for CREB 285-339 pWT and FD during stopped flow with probe DNA in the presence of competitor DNA.	201
6.9	Varying the competitor DNA showed that the intersegmental transfer is a predominant mode of translocation for CREB.	203
6.10	Varying the competitor DNA showed that the intersegmental transfer and sliding model falters for CREB at extreme competitor DNA concentrations.	204
6.11	Marked increase in secondary structure propensity and reduction in dynamics of the V288L/M291R mutation of CREB bZIP.	206
6.12	Disorder increases the rate of sliding of the disordered DNA binding domain of CREB, but not the effective sliding length.	208
6.13	GC content flanking various <i>H. sapiens</i> transcription factor DNA target sequences is higher than the genome average.	210
6.14	Assigned CREB 277-339 with non-CRE p1 DNA at 35°C.	213
6.15	Predicted secondary structure of CREB 277-339 in complex with non-CRE p1 DNA is intermediate between apo and non-CRE DNA.	214
7.1	Model of CREB bZIP association to DNA (target or non-target sequences) via dimer or monomer pathways.	224
S1	Labelling trials of CREB S337C.	235
S2	Labelling efficiency of CREB bZIP S337C with AF594 was estimated with ESI-MS.	237
S3	Chemical denaturation of CREB bZIP pWT.	238
S4	Thermal unfolding of CREB bZIP pWT unveils a tight homodimer interaction.	239
S5	Thermal melt extraction of baselines.	240
S6	Van't Hoff plot from thermal melt data.	241
S7	ESI-mass spec analysis of the CREB bZIP 285-339 FD	248
S8	k_{off} determination of CREB from non-CRE DNA required temperature dependence measurements.	249
S9	Anisotropy of free AlexaFluor488 changes upon addition of CREB 277-339 FD.	250
S10	Estimate of k_{off} at 25°C from linear extrapolation of $\ln(k_{\text{obs}})$ against $1/T$	250
S11	k_{on} of CREB 277-339 FD to non-CRE by stopped flow anisotropy required temperature dependence.	251
S12	Estimate of k_{on} at 25°C from linear extrapolation of $\ln(k_{\text{obs}})$ against $1/T$	255
S13	CREB 285-339 in complex with CRE-DNA was assigned at 35°C.	256

S14	CREB 285-339 with non-CRE DNA assignments at 35°C.	260
S15	Carbon chemical shifts of CREB 285-339 with non-CRE DNA assignments at 25°C and 35°C.	264
S16	CREB 277-339 was assigned in complex with CRE-DNA at 25°C. .	265
S17	CREB 277-339 is in slow/intermediate exchange with CRE-DNA. .	269
S18	CREB 277-339 was assigned in complex with CRE-DNA at 25°C. .	270
S19	Arginine C δ chemical shifts show increased dispersion in CREB 277-339 with non-CRE DNA.	273
S20	Limited carbon chemical shift changes with non-CRE DNA addition to CREB 285-315.	285
S21	Stability of the decoupled construct was similar to that of the pWT.	287
S22	Varying temperature did not change exchange regime for CREB 277-315 with CRE-DNA.	288
S23	Stopped flow of CREB mixed with probe and competitor DNA was challenging to fit to a single or double exponential fit.	289
S24	Assigned ^{15}N -HSQC of CREB 277-339 V288L/M291R.	290
S25	DNA GC content of flanking sequences of promoter sequences. . . .	293
S26	Chemical shifts of basic region in CREB 277-339 in complex with non-CRE p1 DNA appear intermediate between apo and non-CRE DNA.	294
S27	No chemical shift perturbation of CREB 277-339 in complex with non-CRE DNA was observed upon dilution.	297
S28	Small difference in dispersion of the arginine C δ between random and symmetrical non-target DNA.	298

List of Abbreviations

RB	Resuspension Buffer
FPLC	Fast Protein Liquid Chromatography
TEV	Tobacco Etch Virus
CV	Column Volume
RT	Room Temperature
SWB	Salt Wash Buffer
IEL	Ion Exchange Load
CD	Circular Dichroism
ITC	Isothermal Titration Calorimetry
AF	AlexaFluor
ESI-MS	Electrospray Ionisation Mass Spectrometry
CPM	7-Diethylamino-3-(4'-Maleimidylphenyl)-4-Methylcoumarin
MW	Molecular Weight
pWT	pseudo Wild Type
SSP	Secondary Structure Propensity
CREB	cAMP response element binding protein
CRE	cAMP response element
CBP	CREB-binding protein
cAMP	cyclic adenosine monophosphate
bZIP	basic leucine zipper
bHLH	basic Helix-Loop-Helix
HetNOE	Heteronuclear nuclear Overhauser effect
FD	Forced Dimer
TAD	Transactivation Domain
DBD	DNA binding domain

EMSA		Electrophoretic Mobility Shift Assay
VK		Veksler and Kolomeisky
IDP		Intrinsically Disordered Protein
IDR		Intrinsically Disordered Region
NDR		Nucleosomal Depleted Region
CSP		Chemical shift perturbation
PKA		Protein Kinase A
TORC		Transducer of the Regulation of CREB

1

Introduction

Contents

1.1	The Central Dogma of Biology	2
1.2	Genomic regulation at the level of transcription	2
1.2.1	DNA packaging within the nucleus	2
1.2.2	Transcription factors regulate gene expression	4
1.3	Disordered Proteins	10
1.4	Protein disorder in transcription factors	13
1.5	bZIP family is disordered and the pathway of association to target DNA is controversial	15
1.5.1	cAMP Response Element Binding Protein	18
1.6	Pathway for target association by bZIPs can occur via two routes	23
1.6.1	Dimerisation competent bZIP constructs remain specific for cognate DNA sequences	23
1.6.2	Analysis of monomeric constructs unveils detail for the monomer pathway	25
1.6.3	Forced dimer bZIP constructs to investigate the pre-formed dimer pathway	27
1.7	Target search process of transcription factors	27
1.7.1	Target search mode of bZIP proteins	30
1.8	Aims of the thesis	30

1.1 The Central Dogma of Biology

The central dogma of biology was first described by Francis Crick in 1957. The general scheme that was proposed was that the information stored in DNA, RNA and protein was transferred in a well defined mode: $\text{DNA} \rightarrow \text{RNA} \rightarrow \text{Protein}$ ¹. Whilst this scheme does still hold true, it was a significant oversimplification of what is currently known today and Crick does indeed postulate that other "special" transfers of information may occur.

Since that time, the various levels of regulation have been investigated and well characterised. The process of the transfer of the information encoded into the DNA sequence to RNA is called transcription, with the subsequent process of transfer from RNA to protein called translation. At each stage, there are significant levels of regulation. The levels of RNA and their lifetimes contribute to the levels of proteins².

In comparison to transcription, which will be discussed more in detail later on, the regulation of translation is a very fast process². This is down to the use of pre-assembled pre-initiation complexes which associate to the 5' end of mRNA (reviewed by Sonenberg *et al*³). A level of regulation for translation can occur from the perspective of RNA, with miRNA also shown to affect the rate of translation³, showing that the Crick model is more accurately drawn with feedback loops at each stage.

1.2 Genomic regulation at the level of transcription

1.2.1 DNA packaging within the nucleus

Within the nucleus, DNA is packaged into macromolecular structures called chromatin. The now-famous arrangement of chromatin, that of beads on a string (reviewed in Baldi *et al*⁴), is a repeating polymer unit called nucleosomes (Figure 1.1). The nucleosomes are an octamer of two copies of four histone proteins (H2A, H2B, H3 and H4), with 146 bp of DNA wrapped around this⁵. The DNA associated to

nucleosomes is inaccessible to transcription factors and other DNA binding proteins, and the process of making DNA accessible is regulated by a variety of mechanisms.

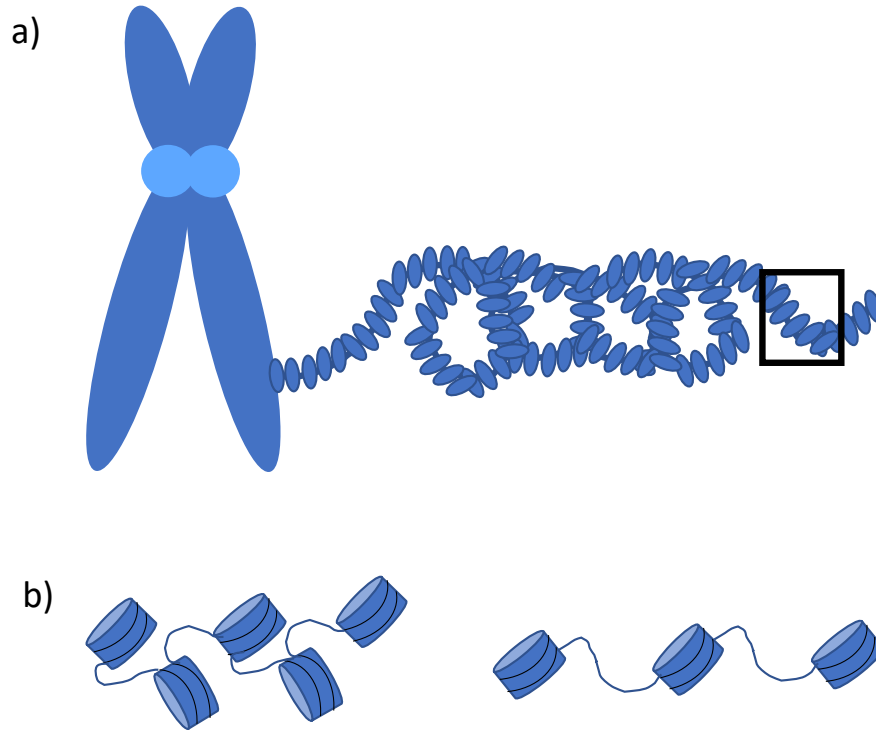


Figure 1.1: DNA is packaged in supramolecular structures *in vivo* within the nucleus. a) DNA is contained within chromosomes, themselves being built from highly packaged chromatin. The chromatin is a beads on a string polymer built up from nucleosomes where the DNA is wrapped around histone octamers. The square indicates a closer examination shown in panel b). b) Closer view of the nucleosome structure in two versions of chromatin. Left: heterochromatin is where the nucleosome octamers are compacted together, with access to the DNA closed⁶. Right: open form of chromatin, with the DNA able to act as linkers between nucleosome units.

Histones are heavily modified post-translationally, with these having a marked impact on their function of regulating DNA accessibility. A large study identified multiple sites where histone residues were modified by either methylation, ubiquitination, acetylation or phosphorylation and that the modifications were tissues specific⁷. This complexity therefore enables the tight regulation of genes. However, the DNA itself can be modified for a further level of regulation, with an elegant FRET study showing that methylated CpG dinucleotides led to compaction of the histone complex⁸.

These modifications on histones and/or DNA thereby created a theoretical framework for two modes of chromatin: open and closed (Figure 1.1b). However, recent evidence points towards topological domains of chromatin, consistent with the closed mode of chromatin⁶. Between the boundary of these domains are where the so-called "housekeeping" genes were found to be located⁶. The position of the nucleosome on the DNA is dependent on the methylation status of H3K4, and is highly dynamic between double and triple methylation of H3K4 surrounding the enhancer region of a transcription start site⁹.

The role of phase separation, which describes the formation of proteinaceous and membraneless organelles¹⁰, within epigenetics is now starting to be more well understood (reviewed in Hnisz *et al*¹¹). Due to high local concentrations of DNA, DNA-binding proteins and their co-factors, there has been evidence of phase separation occurring¹². For instance, HP1 α was shown to have a significantly reduced diffusional coefficient that was consistent with diffusional coefficients surrounding the hetero-euchromatin border. Further single molecule work *in vivo* with the glucocorticoid receptor revealed multiple phases, two of which were low mobility states¹³. One of the low mobility states was found to be bound to chromatin. This mode of regulation appears consistent for many other transcription factors as a way of regulating the transcriptional output^{14,15}.

1.2.2 Transcription factors regulate gene expression

The regulation of gene expression appears predominantly controlled by a large number of proteins called transcription factors (reviewed in Rohs *et al*¹⁶). Transcription factors can be defined as two generic families: General and Specific. General transcription factors are those that are involved in the pre-initiation complex of transcription¹⁷ for all transcribed sequences. They were discovered as accessory proteins to RNA polymerase II¹⁸, and their primary function is to be recruited to upstream promoter regions of genes (Figure 1.2) as a pre-initiation complex. This complex typically consists of Transcription Factor's (TF) II A, B, D, E-H, with

specific DNA recognition imparted by TATA-box binding protein for promoters containing the TATA box DNA sequence¹⁹.

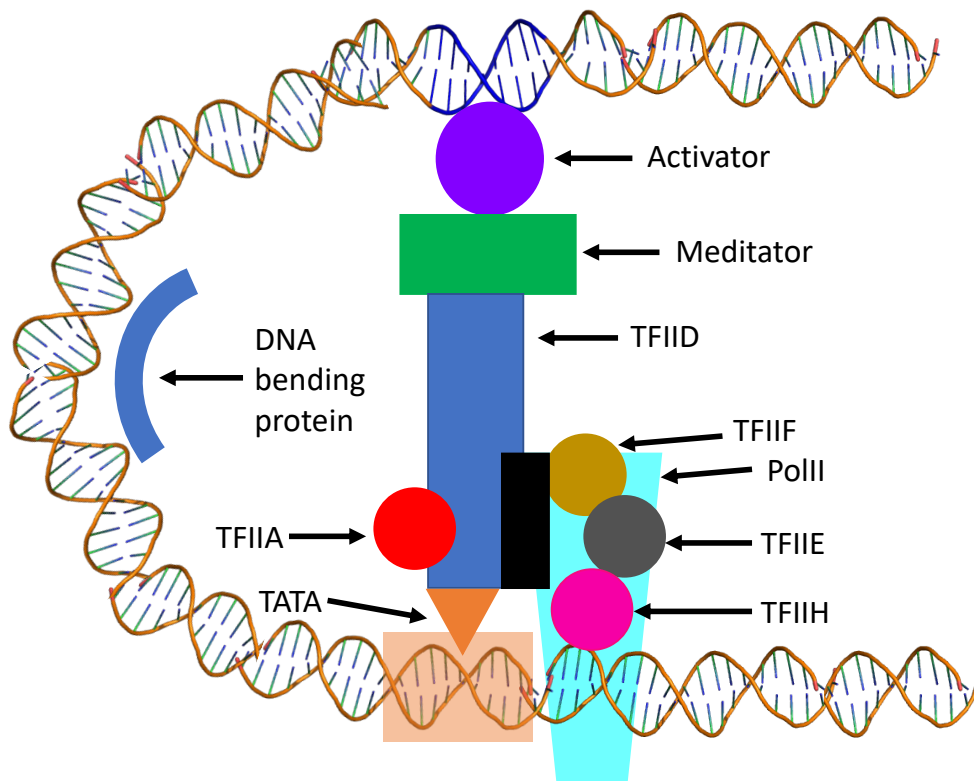


Figure 1.2: Transcription is regulated by a large complex of general and specific transcription factors. The specific promoter site (in blue) is associated with a specific transcription factor (like an enhancer or repressor) to regulate levels of gene expression in *cis* to the transcription start site. Typically a mediator protein provides a scaffold for the activator to aid in recruitment of general transcription factors which become assembled in the following order: TATA box binding protein $\rightarrow$ TFIID $\rightarrow$ IIB $\rightarrow$ IIA $\rightarrow$ PolII with IIE/F/H. This example for a TATA box associated gene expression¹⁶.

Specific transcription factors on the other hand typically associate to the DNA directly themselves, in a sequence specific manner, and will be referred to as just transcription factors from this point. All transcription factor domain architecture can be summarised into at least two functionally similar domains: a trans-activation domain (TAD) and a DNA binding domain (DBD). Discussed in further detail below, the trans-activation domains are responsible for associating to accessory

proteins and general transcription factors, and can be regulated at the level of post-translational modifications. The DNA binding domains are used for interacting with the DNA itself and are discussed throughout the thesis.

Transcription factors have long been known to directly and indirectly read the DNA sequence. Seeman *et al* proposed that a set of hydrogen bond donor and acceptors were utilised for proteins to associate to DNA directly²⁰. Amino acid side chains of Asn and Gln contributed to the hydrogen bonding network to adenine in the major groove, whereas in the minor groove these side chains interact with guanine. Many crystal structures are now available of transcription factors in complex with DNA, mostly always specifically bound to the target containing DNA^{19,21–24}, indicating the difficulties in characterising a non-target DNA/protein complex.

The crystal structures show that a large range of interactions occurs between the transcription factor and the DNA. These include hydrogen bonding between the DNA and protein side chain plus backbone atoms, electrostatic interactions between positively charged residues to the phosphate backbone and also hydrophobic interactions. A combination of these contribute to specific and non-specific interactions. Little is known about the particular DNA sequence dependence against the protein sequence dependence however. It is widely accepted that a combination of DNA shape and sequence enable increased specificity of protein-DNA interaction. A study showed a distinct preference for hydrophobic interactions to A-DNA structures and for polar interactions to B-DNA²⁵, with the A-DNA preference being due to the way A-DNA-like sugars expose larger hydrophobic surfaces.

Examples of this joint DNA structure and sequence dependence is that in the TATA box binding protein¹⁹ (Figure 1.3a). The TATA box binding protein associates to the minor groove but the interaction perturbs the conformation of the DNA significantly, enlarging the minor groove to expose the DNA to make preferential hydrophobic interactions. Proteins which recognise an AT rich sequence are most frequently found to associate to the minor groove. In contrast, the zinc-finger binding protein zif268 associates to a GC rich sequence, which exhibits B-DNA conformation when the structure was solved²², with no perturbation of the

DNA structure (Figure 1.3b). An overlay of the crystal structures of GCN4 and CREB in complex with CRE DNA^{21,26} show that for GCN4, the B-DNA is slightly kinked compared to the cAMP response element binding protein (CREB) structure, highlighting the influence of an indirect readout of DNA shape in conjunction with the direct DNA sequence (Figure 1.3c). In contrast, AP-1 DNA does not show a significant change in DNA conformation when associated to GCN4²³.

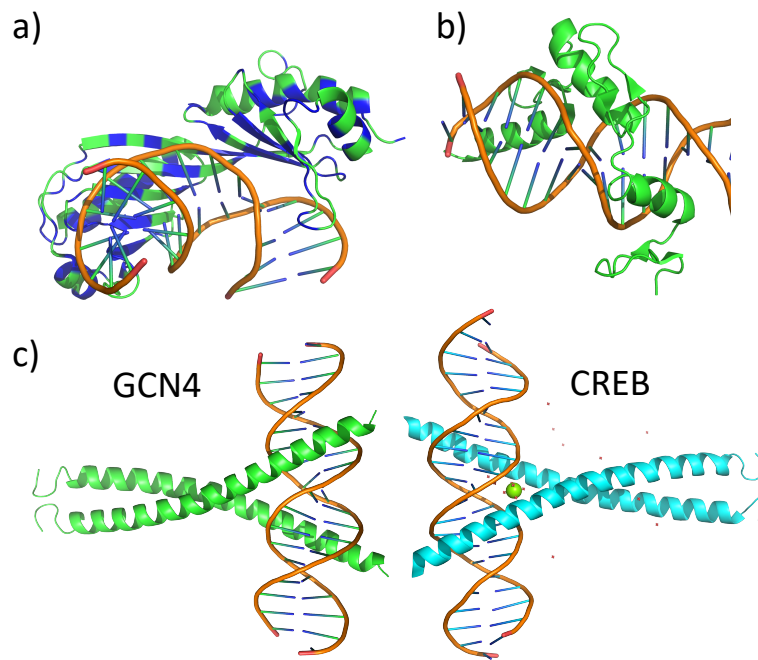


Figure 1.3: Transcription factors can directly and indirectly read a DNA sequence. a) Crystal structure of TATA box binding protein associated to 5'-TATAAAAG-3'¹⁹. Residues in blue are hydrophobic residues (L,I,M,V,G,F,A). A significant number of these are in the interface between protein and DNA. PDB:1vtl. b) Crystal structure of Zif268 (Egr-1) in complex with 5'-GTGGCGTGGGCGGCGTGGGCGT-3'²². PDB: 1p47. c) CRE-DNA in complex with GCN4 (left) and CREB (right) from PDB 1DGC and 1DH3 respectively. CRE DNA is kinked due to the complex with GCN4, whereas CRE in complex with CREB is not and remains in B-DNA conformation.

Transcription is also regulated at temporal and spatial levels. For instance, a study of the Forkhead promoter sites in different tissues and during different times of development in *Drosophila nidogen* showed a different expression pattern of the gene *nidogen*²⁷. Mutation of each site upstream to a GFP reporter led to increased expression from the *nidogen* enhancer region in the heart, whereas in somatic

myoblasts mutating one or more Fkh sites reduced expression²⁷. Complementary to this, Homeobox domain containing transcription factors can bind to different *cis*-regulatory modules in different tissues for the same gene²⁸, which increases the number of promoter sites for a particular transcription factor. Various co-factors may be the reason for the highly regulated spatial and temporal regulation of transcription factors, and these will be discussed in the context of CREB below.

Various families of transcription factors have now been found using DNase footprinting initially²⁹ and more recently by DNA affinity purification mass spectrometry or stable isotope labelling of amino acids in cell culture³⁰, due to many family members associating to a similar or identical DNA sequence. An exception to this general DNA target sequence rule are zinc finger family member proteins³¹. Transcription factors can be split into three main structural families, based on the structure the DNA-binding domain has in complex with the DNA: α -helix, β -sheet and mixed α/β . The α -helix family consists of helix-turn-helix, helix-loop-helix and basic leucine zipper (bZIP) proteins, with an example shown in Figure 1.4 of the helix-turn-helix containing protein, lac repressor³². The lac repressor contacts both the major groove in a specific manner, with the interface between two symmetric headpiece domains interacting with the minor groove (Figure 1.4a). Similarly the bZIPs, which will be discussed within this thesis, use a single α -helix per monomer to associate to the major groove only³³.

The β -sheet family of transcription factors however typically involve a hydrophobic protein-DNA interface, seen with the TATA-box binding protein and discussed above¹⁹ (Figure 1.3). This is not the only type of interaction though, with AML1 actually interacting with the DNA using two loop sections at the extreme C-terminus called "tail" and "wing" elements that interact non-specifically between the major and minor groove, clamping the DNA in position³⁴ (Figure 1.4b). The tail domain makes most of the specific contact.

The major example of the mixed α/β family are the zinc finger transcription factors. This is the only transcription factor family where the target DNA sequence is vastly different between the members³⁶, as most other families share a similar

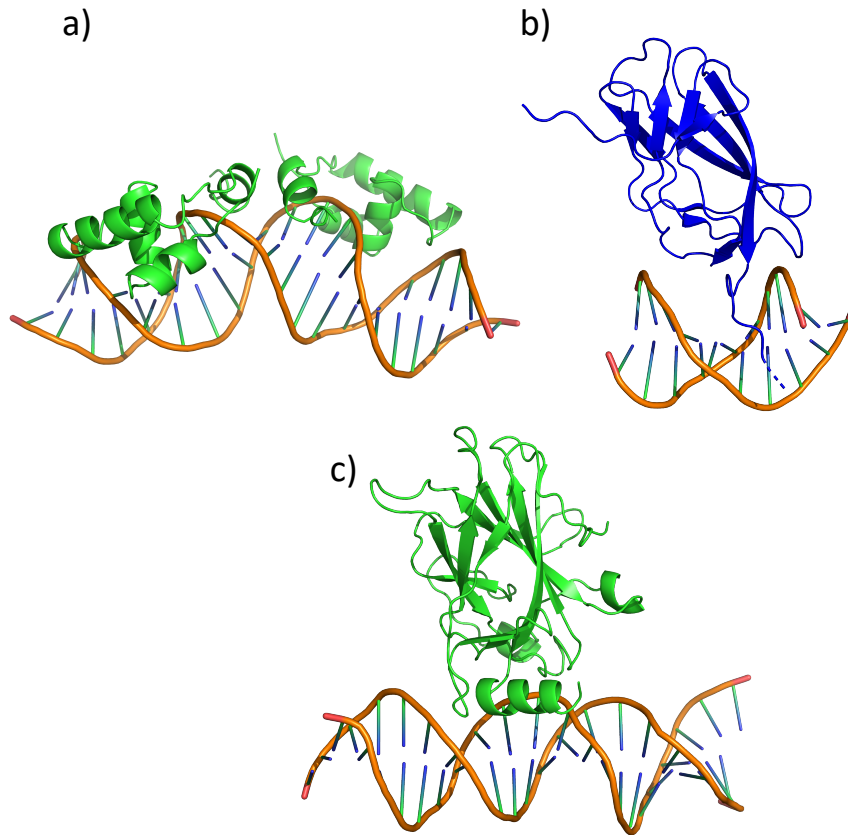


Figure 1.4: Three major structural families of transcription factors. a) α -helical family have α -helices as their DNA binding domain. Example shown is the lac repressor headpiece domain³² in complex with symmetric operator site DNA (PDB: 1CJG). b) β -sheet family transcription factor member AML1 (blue) associated to DNA (PDB: 1H9D)³⁴. c) p53 is an example of a mixed α/β structure family member³⁵ (PDB: 1TSR).

cognate motif. Positions within the DNA are well defined for this class, with position -1, 3 and 6 contacting the major strand, with position 2 contacting the reverse strand. With the evidence above, it was therefore assumed that transcription factors could associate in a specific manner to a single cognate site. p53 is another example of a mixed α/β family member, and has been well characterised to due its role as a tumour suppressor³⁷ (Figure 1.4c).

Interestingly, the theory of one transcription factor family associating to one DNA motif was challenged by Badis *et al* using protein microarrays³⁸. The study used 104 known and predicted mouse transcription factor DNA binding domains

and identified that, in half of the proteins, a secondary DNA binding site was seen. Analysis of these revealed binding motifs like variable spacer length for Jun (CRE to AP-1 sequences TGAC**G**TCA where **G** is the spacer) or position interdependence like for estrogen related receptor α (CAAGGTCA or AGGGGTCA, but not CAGGGTCA or CGGGGTCA)³⁸.

1.3 Disordered Proteins

Over the past 20 years, intrinsically disordered proteins (IDP) have become increasingly recognised as a major class of proteins^{39–41}. Initially thought of as disagreeing with the general theory that a protein requires structure to be functional, disordered proteins are defined as existing in a conformational ensemble with little to no defined regular secondary or tertiary structure (Figure 1.5). A large scale bioinformatics screen identified the propensity of intrinsically disordered regions (IDR) of differing lengths (minimum of 30 residues) and revealed a large population of IDRs in eukaryotes (up to 41%)⁴².

Sequence characteristics of IDR/IDPs were characterised in a subsequent study, with the sequence found to be enriched in R, K, E, P and S⁴⁵. Expanding on this, disorder can be predicted due to low sequence complexity that is enriched in charged residues and containing the secondary structure breakers P and G (reviewed by Dyson and Wright⁴⁰). In contrast, Romero *et al* showed a lower propensity of C, W, Y, I and V residues, all of which, except C, are bulky and hydrophobic. Cysteine residues in folded proteins are often involved in inter/intramolecular disulphide bond formation, but in disordered proteins they have actually been shown to coordinate intermolecular interactions as well⁴⁵. *In vivo*, typically intracellular proteins that contain cysteines would remain in a reduced state.

IDRs/IDPs were found to have a high prevalence in molecular recognition elements; these involve a molecule with a strong and specific binding to the partner molecule. Molecular recognition elements with a propensity for α -helix content are involved in nascent interactions with protein partners through a coupled folding upon binding interaction⁴⁶. This importance of α -molecular recognition elements was

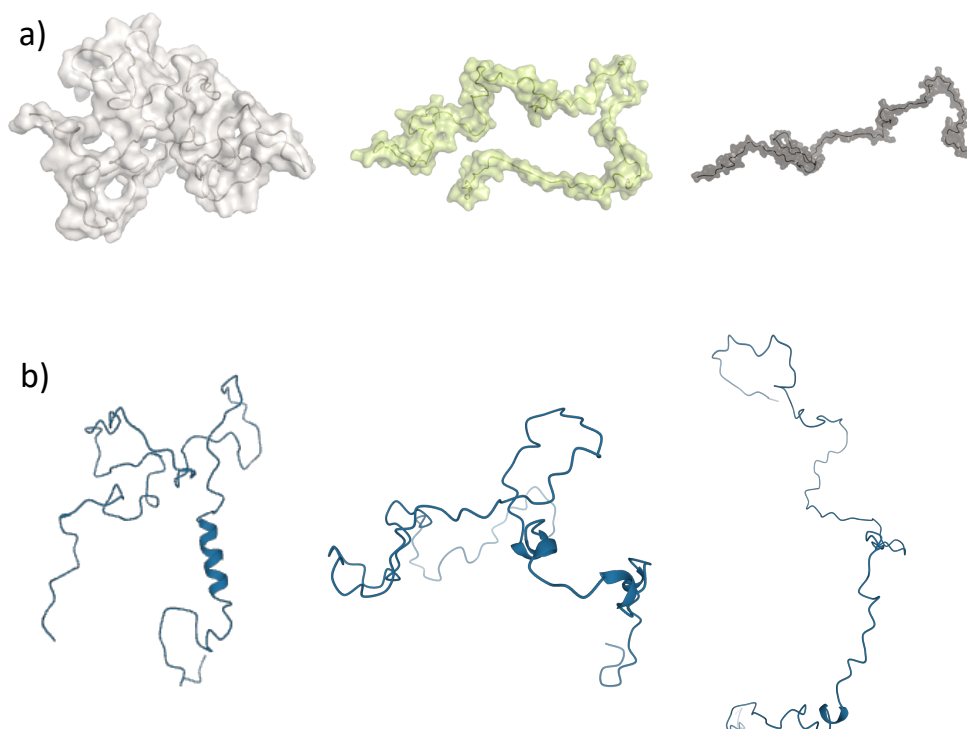


Figure 1.5: α -synuclein is a well characterised model IDP. a) Structures of the conformational ensemble of α -synuclein, solved by paramagnetic relaxation enhancement NMR spectroscopy (PED00024)⁴³. b) Structure of N(TAIL) domain from measles virus nucleoprotein solved by various NMR methods (PED00020)⁴⁴. Structures were taken from Protein Ensemble Database (PED).

shown because, when lacking the element, the nucleoprotein of measles virus N(TAIL) domain truncation loses the ability to associate to phosphoprotein. The solution state structural ensemble of the N(TAIL) also shows a region between residues 497-509⁴⁴ with a higher helical propensity, however the domain was still highly dynamic.

The function of disorder in proteins appears to have been selected for to operate separately on affinity and specificity³⁹. Remaining disordered would aid in the high affinity but low specificity interactions⁴⁷, whereas a disorder to order transition leads to low affinity but high specificity interactions^{48,49}. In eukaryotes, the amount of disorder in proteins was found to be almost three times higher than that in archaea, indicating a major functional role in complex multicellular organisms^{42,50}.

A large bioinformatic screen for IDR/IDP sequences in *S. cerevisiae* revealed that IDR's were over-represented in proteins involved in transcription, DNA binding

and signalling cascades⁵⁰. In fact, these proteins contained longer disordered regions than the average population due to being large multi-domain proteins that contains more linker regions and higher probability of incorporating a disordered region⁵⁰. It was found that insertion/deletion mutations are particularly enriched in IDRs⁵¹. Furthermore, whole genome duplication results in proteins that are dose sensitive, with many of these being disordered proteins (reviewed by Ahrens *et al*⁵²). A study of p53 evolution shows a lack of the disorder in the DNA binding domain in neopterans⁵³. In vertebrates, whole genome duplication events have led to p63 and p73, except in most avian genomes.

Disordered proteins have been shown to be hubs within protein interaction networks⁵⁴ and are enriched in short linear motifs that are evolutionary conserved⁵⁵. Therefore, the disordered hub protein can engage in multivalent interactions with a wide range of protein targets. Tight regulatory control of translation of disordered proteins that enables faster regulation of populations of IDPs⁵⁶ was shown with higher turnover of mRNA encoding for disordered proteins.

Studies into whether disordered proteins confer a kinetic advantage for protein-protein interactions have revealed that, with cMyb binding to KIX, that the k_{on} was one of the fastest protein association rates recorded⁵⁷. This was conferred on cMyb due partly to an electrostatic rate enhancement, as the kinetics were relatively ionic strength independent, and the role of disorder in this interaction was believed to lower the activation barrier from encounter to bound complex. Investigations of other known binding partners of KIX that are also disordered showed a similar coupled folding upon binding mechanism, with the residual charge of the partner peptide playing a role in the rate of association⁵⁸.

Residual helical propensity within IDRs is known to play a role in association kinetics of disordered proteins. The activator for thyroid hormone and retinoid receptor showed a linear dependence on residual helicity against k_{on} and k_{off} ⁵⁹, supporting the idea that partial secondary structure is a molecular recognition element. A similar finding was observed using the IDP PUMA associating to MCL1, however in this instance k_{on} enhancement is independent of increased helical content

but helps to stabilise the complex by modulating k_{off} ⁶⁰. The helicity was shown to be modulated by conserved helix-flanking proline residues in p53 and MLL, and the increased affinity was due to a reduction in the k_{off} ⁶¹.

Post-translational modifications are utilised to regulate the function of disordered proteins. Sic1 was found to be phosphorylated at eight N-terminal serines. pSic1 was more extended than Sic1, and remain disordered in the complex with Cdc4, which was to enable Cdc4 to associate to specific phosphorylated sites that are not charge neutral due to the adjacent protein sequence⁶².

IDR's have been found to be linked to some proteins that have been shown to undergo phase separation (discussed initially by Mittag and Forman-Kay⁶³). The disordered N-terminal domain of Ddx4 was found to be necessary for incorporation into membraneless organelles *in vivo* and *in vitro*, and these responded to temperature and isotonic shock¹⁰. The process was driven by electrostatics, due to destabilisation of the droplets at high ionic strength, and is responsible for sequestering ssDNA or RNA from the cellular milieu. Phase separation of nephrin intracellular domain occurred due to homotypic interactions and the amino acid composition was found to be more important than the primary sequence⁶⁴. This lack of sequence dependent characteristic was used to design synthetic peptides that were found to undergo phase separation⁶⁵.

1.4 Protein disorder in transcription factors

As mentioned above, IDPs/IDR were found to be highly populated in protein groups related to DNA binding and transcription⁵⁰. Typically the disordered regions can be found in Transactivation Domains (TADs) or in loops immediately adjacent to the DBD. An example of this is the disordered linker region between the Core and Tet domains of p53, and this was shown to control the 1D diffusion of p53 along non-target DNA⁶⁶. Furthermore, the residual secondary structure content of the N-terminal TAD in p53 was found to control the affinity of the interaction with Mdm2, with more helical mutants of p53 TAD showing lower DNA damage in *in*

vivo assays⁶⁷. A significant quantity of work has been done with CREB and its TAD domain, and that is discussed in Section 1.5.1.

In contrast to TADs, disorder in the DBD is relatively rare and exists in four families (Figure 1.6). The family of AT-hooks contain a protein consensus sequence (PRGRP), flanked by other positively charged residues⁶⁸ (Figure 1.6a). The proline residue was found to exist in the *trans* isomer only, with the glycine enabling a turn-like conformation⁶⁹. The arginine side chains within the consensus sequence, rather than neutralising the phosphates in the DNA backbone, point inwards and make hydrophobic contacts with adenine bases⁷⁰. This consensus sequence enables AT-hook domains to associate preferentially to the minor groove of DNA⁶⁸, but was shown to associate more frequently to non-regular DNA structures such as supercoiled plasmids or four-way junctions⁷¹.

Another transcription factor family with a disordered DBD is the helix-loop-helix containing proteins (Figure 1.6b). This family contains a range of disorder within their DBD, ranked as Myb > Homeobox > Tryptophan clusters > Forked/winged helix⁷². The second helix in helix-turn-helix proteins functions as the recognition helix and associates with the major groove of DNA⁷³. This domain can form constitutive hetero- or homodimers via the leucine zipper domain, and the transition from being disordered to ordered occurs during the formation of the dimer⁷⁴. A study revealed significant helical structure in the monomeric, molten-globule like state of Myc, in particular in the leucine zipper region which interacts with the partner protein Max⁷⁵.

Zinc finger proteins, due to their requirement for multiple finger domains to facilitate DNA binding, can be classified as disordered due to flexible linkers between the fingers (Figure 1.6c). The DBD itself is structured, with Zn^{2+} coordinated by two Cys and two His residues⁷⁶. A well defined α -helix is packed against two β strands. The linker appears to play a major determinant role in the affinity of the Zn-finger and DNA complex, with a conserved linker sequence (TGEKP), in particular, increasing the helical content of each zinc finger in complex with DNA⁷⁷, due to the helix capping P.

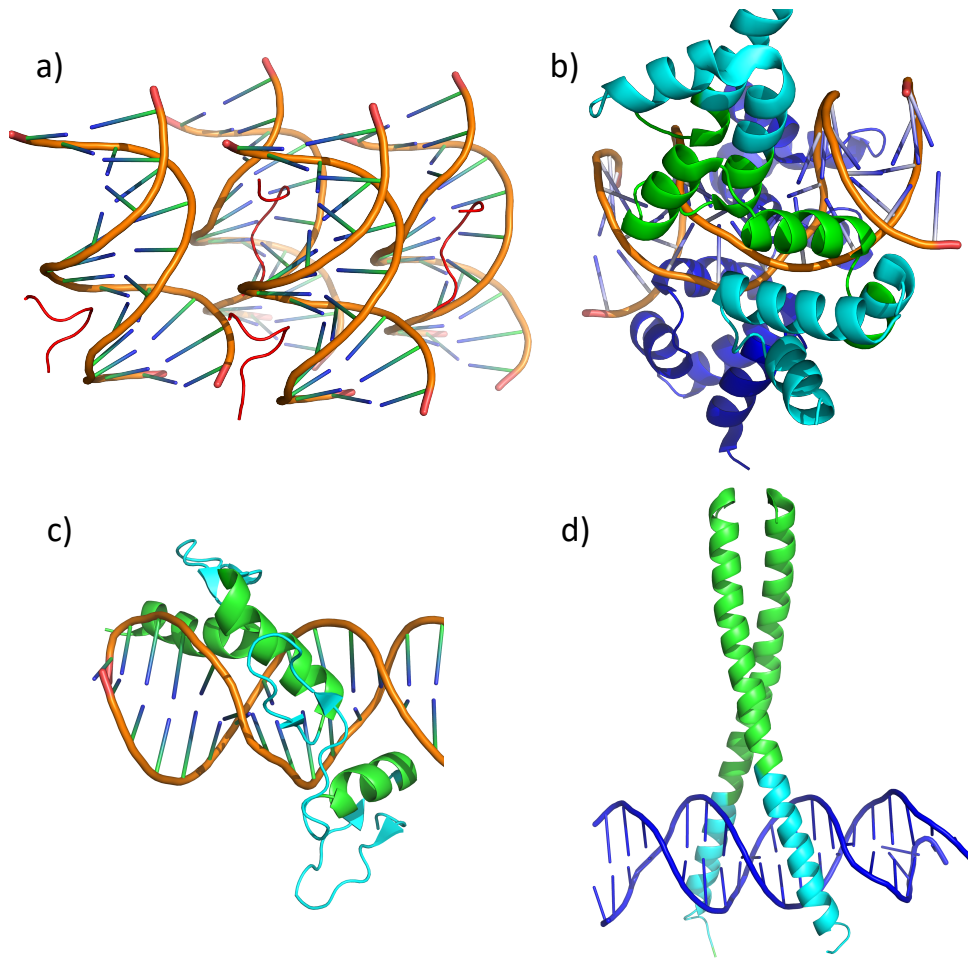


Figure 1.6: Four families of transcription factors contain disorder in the DNA binding domain. a) Crystal structure of an AT hook from HMGA1 in complex with DNA. Note - crystal packing renders different conformation for the AT hook peptide. PDB 3UXW. b) Structure of a helix-turn-helix containing protein KorB. PDB 1R71. c) Crystal structure of the zinc finger transcription factor Zif268 in complex with DNA. PDB 1P47. d) Crystal structure of the homodimer bZIP domain of GCN4 in complex with DNA. PDB: 1YSA. In all structures, predicted disorder domains are highlighted in cyan, aside from in a) where for contrast protein is shown in red only.

1.5 bZIP family is disordered and the pathway of association to target DNA is controversial

A final example of disorder within DNA binding domains are that of the basic leucine zipper domains (bZIP, Figure 1.6d). This domain can be split into regions: the N-terminal basic region and the C-terminal leucine zipper. The basic region protein sequence is predicted to be disordered and is enriched in positively charged residues

(Arg and Lys, Figure 1.7a), which are used to interact directly with the DNA^{78–81}. The leucine zipper functions as the homo/heterodimer interaction interface and is not involved in DNA binding⁸². Furthermore, the leucine zipper in the dimer complex is in a coiled-coil conformation, with a hydrophobic interface⁸³ between the monomers. The sequence characteristics can be summarised in a helical wheel (Figure 1.7b-c), with the hydrophobic leucine residues occupying position d on the wheel and e and g contributing to intermolecular ionic interactions as charged residues are frequented at these positions⁸⁴.

The leucine zipper interaction has been shown to enable homo and heterodimers, such as between Fos/Jun for the latter⁸⁵ and GCN4 for the former²³. However, the vast majority of the dimer interactions that bZIPs could form are preferentially homodimeric, from a large scale study of eukaryotic bZIPs^{86,87}. Various properties have been shown to contribute to this specificity of coiled-coil interaction, with interhelical salt bridges seen in crystal structures as well as in engineered protein inhibitors^{21,88}. Further investigations with Fos-Jun identified that 8 residues in the *a* and *d* heptad positions were required for specificity using thermal melts of various bZIP chimera's⁸³. That being said, the stability of bZIP dimers has been controversial, with K_d ranging from μM to pM ^{89,90}.

The most characterised bZIP domain is GCN4. Initial investigations utilising the leucine zipper region only⁹¹, identified a completely helical structure by CD at 25°C. The formation of the dimer using the leucine zipper region only in work by Wendt *et al* also identified an association model with at least one conformational rearrangement step, as two rate constants were observed upon dilution of the protein, with the first rate being concentration independent⁹². Destabilised GCN4 leucine zipper domains had homodimer K_d values in the low μM , and exhibited fewer phases in the folding process than stabilised versions⁹³, revealing another example of the conformational rearrangement.

However, when looking at the GCN4 bZIP domain (which contains the basic region), a single phase was seen corresponding to the two-state process of unfolded monomer directly converting to folded dimeric complex⁹¹. NMR has also been

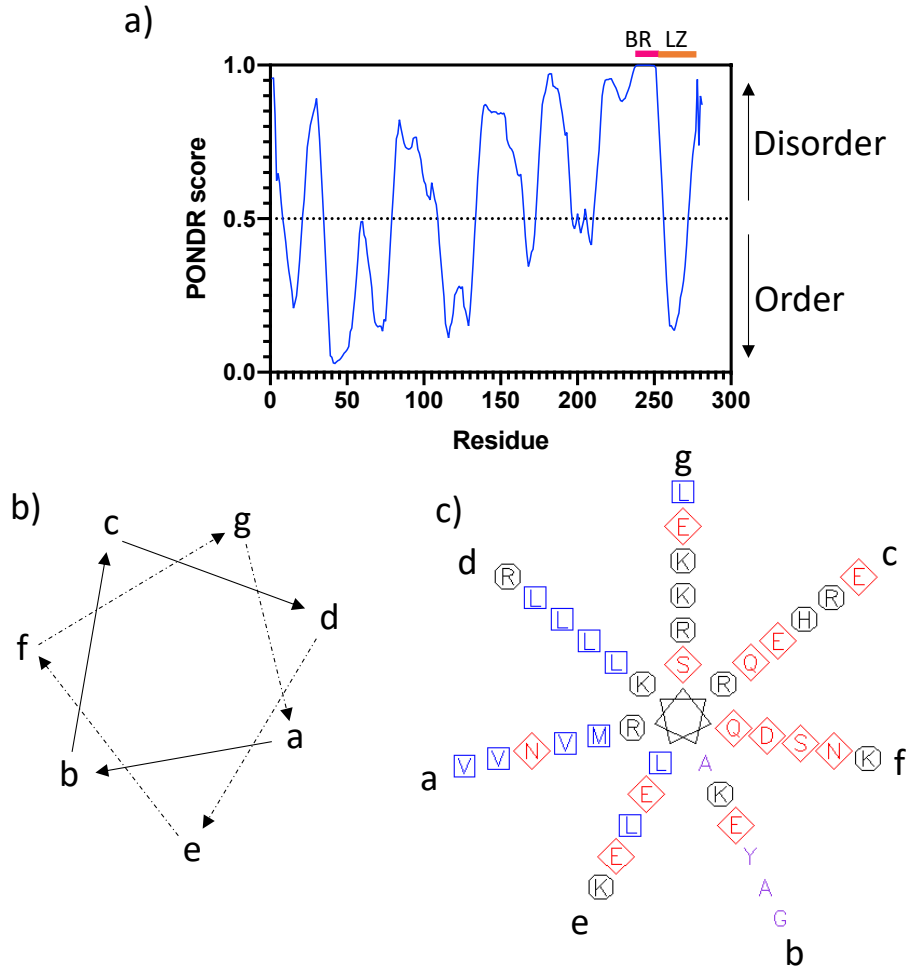


Figure 1.7: Basic region is predicted to be disordered and leucine zipper sequence can be summarised as a helical wheel. a) Disorder prediction of GCN4 protein sequence (Uniprot ID: P03069) using PONDR. BR = basic region, LZ = leucine zipper. b) Generic diagram showing a helical wheel for a coiled-coil domain. Positions **a** and **d** are typically hydrophobic, **e** and **g** are predominantly charged with **b**, **c** and **f** characterised as solvent exposed. c) Helical wheel plot for GCN4 (residues 242-281, Uniprot: A6ZQT9-1) from EMBOSS pepwheel.

performed on GCN4 bZIP previously in the apo-state, which showed the basic region populating dynamic transient structures, but with helical propensity⁹⁴ which decreases at higher temperatures (max of 310 K). Two-state unfolding of GCN4 bZIP was also identified by time-domain ¹³C NMR spectra at several temperatures⁹⁵.

1.5.1 cAMP Response Element Binding Protein

An example of a member of the bZIP transcription factor family, the cAMP Response Element Binding (CREB) protein is 341 residues long and contains four distinct domains. The domain architecture of CREB consists of two poly-glutamine (polyQ) domains that flank a Kinase Inducible Domain (KID). The DNA binding and dimerisation bZIP domain is located at the extreme C-terminus (Figure 1.8a) and is responsible for associating to cAMP Response Elements (CRE) to regulate expression of target genes. CREB has been shown to associate to full CRE sites (5'-TGACGTCA-3') and half-sites (5'-TGAC-3')⁹⁶. Interestingly, the association to full and half-CRE sites have different affinities, and the populations of CREB occupancy to either site being phosphorylation-dependent⁹⁷. The function of the Q₁ domain is currently unknown, but the predicted structure of this domain is disordered except the C-terminus transition to the KID domain (Figure 1.8b), where a marked increase in predicted order is observed.

The Q₂ domain has been coined the constitutively active domain, due to the phosphorylation independent transcriptional activity that was reported⁹⁸. The Q₂ domain could associate to the TATA-box binding protein cofactors *in vivo* in yeast, in a hydrophobic interaction. Concomitantly, an investigation *in vivo* in liver extracts was completed which showed the same domain interacts with RNA polymerase II, independently of CREB phosphorylation⁹⁹. On the other hand, the KID domain has been studied intensively as a prototypical model system for disordered proteins/regions. In the absence of a partner protein, the KID domain is predicted to be unstructured and this was defined experimentally by several groups^{48,102}. The KID domain contains several residues that have been shown to be phosphorylated, in particular by Protein Kinase A (PKA) and casein kinases, amongst others^(103,104). Phosphorylation of S133 in the KID domain by PKA has so far been the only post-translational modification to lead to enhanced transcriptional activity, which was found to be the interaction of CREB pKID to CREB-binding protein (CBP)¹⁰³. This was confirmed in an *in vivo* set of experiments by Shaywitz

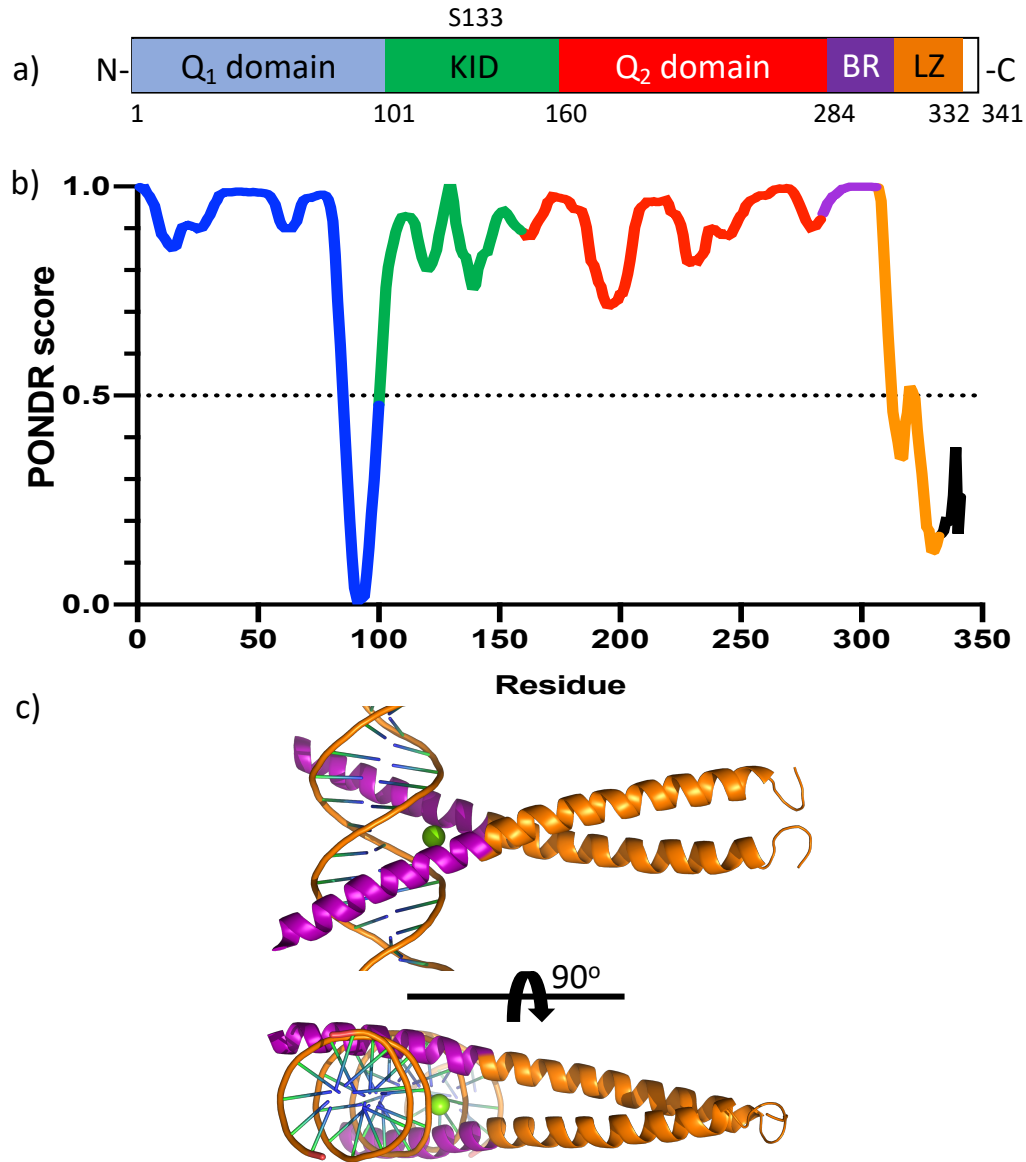


Figure 1.8: bZIP domain of CREB is located at the C-terminus and contains one of the predicted ordered regions. a) Domain architecture of CREB1a sequence from *H. sapiens*. S133 is the residue that is phosphorylated during the signalling pathway events. Leucine zipper = LZ, basic region = BR, Q1/Q2 domains are polyQ domains, Kinase Inducible Domain = KID. b) Disorder prediction of CREB1a using the PONDR webserver^{100,101}. CREB1a sequence is Uniprot: p16220. c) Crystal structure of the bZIP DNA binding domain of CREB bZIP (residues 285-339). LZ is orange, BR is purple. Mg²⁺ is shown as a green sphere. PDB=1DH3.

et al by *E. coli* and HEK293 cells to show the increased transcriptional output¹⁰⁵, but that this was predominantly mediated by hydrophobic interactions.

Other phosphorylation sites in the KID domain, particularly S142, can be targetted by Calmodulin dependent protein kinases (CaMKII) in addition to S133¹⁰⁶. In contrast, phosphorylation at S142 does not lead to activation of gene expression in the presence pS133, and is thought to block the binding of CBP to S133 due to the close proximity of pS142¹⁰⁴. The phosphorylation state of CREB does not impact the binding affinity to CRE-DNA *in vitro*¹⁰⁷, with both CREB and pCREB associating to CRE with a K_d of ~ 1 nM.

The bZIP domain contains the basic region and the leucine zipper, which in CREB's case, preferentially forms dimeric interactions with the CREB/CREM/ATF subfamily^{108–110}. There is a considerable range of estimates for the stability of the homodimer interaction of CREB bZIP. CREBs homodimer stability has been inferred indirectly from DNA binding studies^{90,111}, indicating that the protein was dimeric when bound to the target CRE site. However, limited and contrasting evidence has been reported for the oligomerisation state of CREB in the absence of DNA. For instance, Wu *et al* identified that full length CREB was monomeric in the apo-state⁹⁰, whereas another study with a similar length of construct identified a K_d of 0.6 nM⁸⁹.

The CREB homodimer has been shown to associate to Transducer of the regulation of CREB (TORC) to form a tetrameric complex¹¹², with TORC wrapping around the solvent exposed face of the leucine zipper, in an phosphorylation independent manner. TORC was shown to bind to the bZIP domain alone and could result in an increase of activation in gene expression¹¹². Therefore, the regulation of CREB occurs across a large region of the protein, involving specific protein-protein interactions as well as post-translational modifications. Some of the signalling pathways CREB is involved in are summarised in Figure 1.9.

CREB in particular is unique within the bZIP family, having been found to have a divalent cation dependence to its DNA binding specificity^{113,114}. This divalent cation, Mg^{2+} , significantly reduced CREB bZIP association to non-target DNA sequences, whilst reducing the binding affinity for target DNA sites by ~ 3 -fold¹¹³. Overall, this resulted in a large increase in the specificity of CREB for the target DNA (note: specificity being the difference between target and non-target

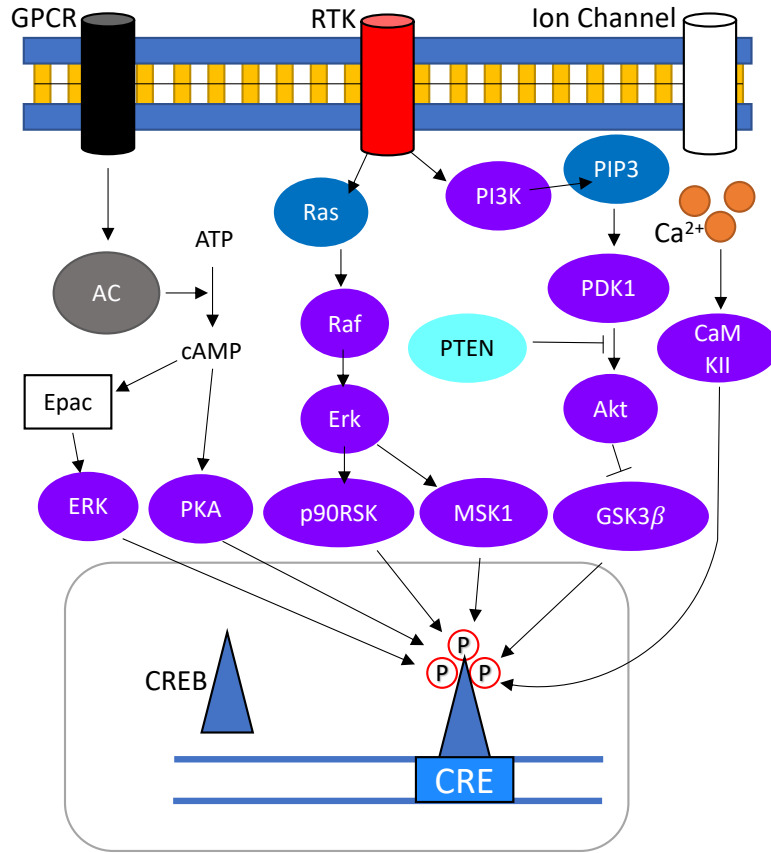


Figure 1.9: Multiple signalling pathways converge on CREB for regulation of transcriptional response. Signalling from G-protein coupled receptors (GPCR), receptor tyrosine kinases (RTK) or ion channels activates downstream kinases. PKA for instance is dependent on cAMP levels, which leads to phosphorylation of S133 on CREB. Multiple pathways are activated by RTK signalling, which lead to phosphorylation of CREB at multiple sites. Diagram was adapted from Wang *et al*¹⁰⁴.

binding affinities). CREB stability on target DNA has been shown to be enhanced upon phosphorylation of S133 by Protein Kinase A¹⁰³, due to the recruitment of CREB Binding Protein (CBP).

The crystal structure of CREB bZIP homodimer in complex with the somatostatin CRE DNA was solved²¹ and identified which residues contacted the DNA phosphate backbone (R286, K292, R294, R298, K303, K305) and the DNA bases directly (N293, A296, A297, S300). The role of K305 was believed to aid in CREB associating to the CRE-DNA without distorting the DNA structure, unlike GCN4²¹. Also, typical NMR experiments are usually performed in a non-biologically relevant buffer, such as with a lower ionic strength or non biological pH, so the scope for

characterising the structure of CREB bZIP in complex with target and non-target DNA by solution NMR is high.

Functionally, CREB is responsible for the regulation of over 4000 genes, through a combination of clusters of full and half site CRE's⁹⁶. Zhang *et al* also investigated the role of methylation at CRE motifs, finding that CREB associated with unmethylated CRE motifs and did not associate to methylated sites, *in vivo*. This agrees with earlier *in vitro* and *in vivo* work by Iguchi-Ariga and colleagues, whereby the CpG dinucleotide in the centre of the CRE motif (5'-TGAC**CG**TCA-3') was examined in a methylated and unmethylated state¹¹⁵. They found that the CREB was unable to associate to the methylated CRE and that this also led to no expression output. Transcriptional output of CREB activated by cAMP targets different sets of target genes in different cell lines. Equivalently, those cAMP inducible genes may be dependent on the recruitment of CBP via the phosphorylated S133⁹⁶.

One regulatory partner that can associate with CREB and CBP is p53³⁷. Phosphorylated CREB (pCREB) at S133 was shown to enhance the association of KIX to p53, with pCREB and CREB both associating with p53 alone. Giebler and colleagues therefore state that CREB can function as a scaffold for the association of p53 to KIX³⁷. This interaction of CREB with p53 is altered in the presence of Huntington's disease protein¹¹⁶, indicating the important mediary function that CREB plays.

CREB has been found to play a major role in several other disease states as well, due in part to its role as a terminal protein in many signalling cascades (reviewed in¹⁰⁴). Many neurological disorders are implicated with CREB signalling pathways, with Alzheimer's Disease amongst others (review in¹¹⁷) being implicated in an over- or down-regulation of CREB induced gene expression. Furthermore, Huntington's Disease was modelled to show that a loss-of-function CREB model reconstituted the disease symptoms, whereas a gain-of-function model reduced these¹¹⁸. CREB was shown to be down-regulated in mouse models of vascular disease, with oxidised low-density-lipoproteins decreasing levels of CREB in a dose dependent manner¹¹⁹. CREB controls hepatic liver metabolism which was

dependent on PPAR- γ ¹²⁰. Rather than a dysregulation through the trans-activation domains, CREB has been found to be hijacked by human T-cell leukaemia virus-1 transcriptional activator protein (Tax-1), using CREB to associate to Tx response elements (TxRE's, 5'-TGACG-3')⁸⁹.

Therefore, being able to understand the mechanisms by which CREB searches for its target DNA motif is an important question. With the wide functionality of CREB and its very large implications for human health, understanding these molecular mechanisms could help inform design of inhibitors to CREB associating to DNA to aid in those disease processes whereby CREB is over-regulated for instance.

1.6 Pathway for target association by bZIPs can occur via two routes

With the dimerisation capability of bZIPs, this results in two possible pathways that bZIP protein can use to associate to their cognate DNA site: 1) monomer pathway or 2) pre-formed dimer pathway (summarised in Figure 1.10). Controversy exists as to whether bZIPs prefer the monomer or dimer pathway, due to range of dimer stability that has been reported and also the binding affinity to target DNA sites in both the monomeric and dimeric models.

1.6.1 Dimerisation competent bZIP constructs remain specific for cognate DNA sequences

The specificity of bZIP interactions with their cognate DNA site has been investigated using a wide variety of experimental techniques, from Electrophoretic Mobility Shift Assay (EMSA) to fluorescence anisotropy with labelled DNA/protein and Isothermal Titration Calorimetry (ITC), including differing buffer conditions and protein construct lengths^{78–82,90,97,108,113,121–127}. What remains consistent throughout is the specificity of the interaction with the cognate DNA site over a non-target DNA sequence, typically with an order of magnitude lower K_d when associated to target DNA. This can occur whether associated to full or half sites¹²⁸ (Figure

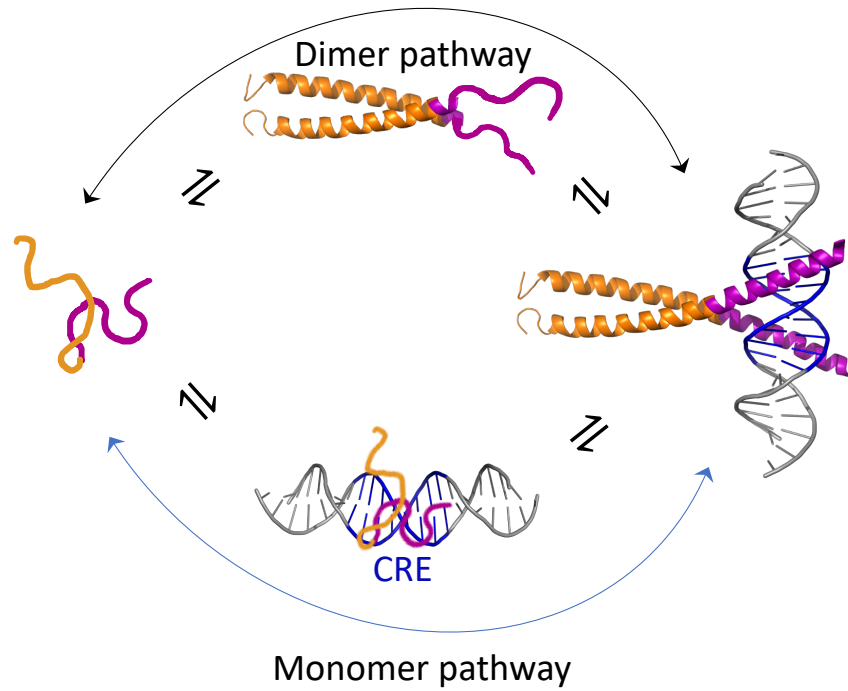


Figure 1.10: Simplified model of the possible association pathways of a bZIP protein (CREB) to DNA. Blue DNA is target CRE motif. Structure used was PDB: 1DH3.

1.11). Usually bZIPs only associate to one target DNA sequence, however GCN4 can recognise two target sequences specifically over non-target DNA¹²⁸.

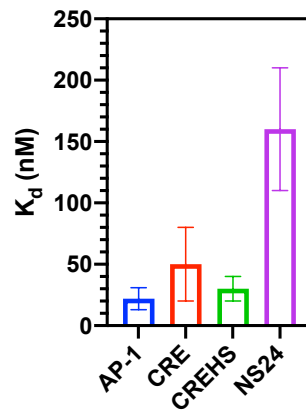


Figure 1.11: GCN4 was shown to have high affinity for two cognate DNA sequences¹²⁸. Electrophoretic mobility shift assay (EMSA) determined binding affinities of GCN4 in complex with different linear DNA sequences. CREHS = CRE Half Site, NS24 = non-specific sequence of 24bp length. Errors are SEM.

The specificity for the target DNA has been largely determined through CD experiments reporting on the secondary structure change of the bZIP domain^{79,85,123,127–131}.

In all cases, the non-target DNA sequence used induces a small increase in helicity of the protein signal when compared to the interaction with target DNA. Thermal melts of the protein in complex with the non-target DNA sequences for CREB bZIP domain showed a melting temperature 13°C lower than the protein in complex with the target, CRE-DNA, indicating a change in stability. In complex with target DNA though, bZIPs undergo coupled folding upon binding, forming a complete α -helix to form the specific contacts in the major groove^{21,23,26,85,128}.

The kinetics of bZIP dimers associating to target DNA has also been investigated *in vitro*, using both EMSA and stopped-flow techniques^{80,81,126,132–134}. In the cases where stopped flow was utilised for GCN4 (bZIP and BR)⁸⁰, the k_{on} was estimated to be on the order of $10^8 \text{ M}^{-1} \text{ s}^{-1}$. Berger *et al* also identified a $k_{\text{on}} > 10^8$ for GCN4 bZIP. However, when another bZIP containing protein, ATF-2, was used by a different group, this showed a significantly slower k_{on} ($1.9 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$)¹³⁴. The k_{off} appeared to be the determining factor in the K_d , with this varying more greatly between protein constructs, indicating that the k_{on} is driven initially by electrostatics. With regards to CREB bZIP, the kinetics of the formation of the DNA complexes have not been investigated previously.

1.6.2 Analysis of monomeric constructs unveils detail for the monomer pathway

The monomer pathway has equally been well characterised with various monomeric models of bZIPs, typically using only the basic region^{78,80,81,121,125,129,135}. One study used a mutant CREB bZIP with a number of the homodimer interface leucines mutated to valine, which abrogated the formation of the homodimer¹³⁵.

As mentioned above, the disagreement about whether the pathway of bZIP proteins going through a monomer or dimer pathway to bind to target DNA can be found when looking at the binding affinity of the monomeric constructs (i.e. basic regions only in most cases) to their respective target DNA sequences. Although a significant majority of them were completed using EMSA or DNase protection assays, a qualitative account of specificities can be estimated. For instance, GCN4

basic region has been studied by several groups, with a wide range of binding affinities reported from lack of observable binding (via EMSA)⁷⁹ to a 400 nM affinity interaction with CRE DNA⁸¹.

Further evidence in support for the monomer pathway was seen by Busch *et al*, who showed that two basic clusters within the basic region are responsible for the specificity that peptides based on Fos/Jun and CREB exhibited when binding to CRE⁷⁸. Various GCN4 constructs (bZIP, basic region and forced dimer versions of those) associating to CRE DNA was measured by stopped flow and showed that the rate was independent of the construct used, even with different dimeric population states at the start of the reaction⁸¹. Studies on bZIP containing proteins CCAATT-enhancer-binding-protein and GCN4 basic regions showed that, due to the homodimer K_d for each protein being high ($> 100 \mu\text{M}$ range) that for the observed k_{on} to target DNA (by EMSA and stopped flow, respectively) the system was using the monomer pathway^{80,133}. This observation was also reported for ATF-2¹³⁴.

A focus of equilibrium based methods does reveal a global lack of specificity of basic region/monomeric bZIP constructs. For instance, a direct comparison between GCN4 bZIP and basic region binding to CRE revealed a 200x greater specificity for the bZIP compared to basic region⁸⁰, with another study showing a ~ 1000 fold difference between the affinities of GCN4 bZIP and basic region⁸¹. Investigations into two target DNA sequences that GCN4 is known to associate specifically with (CRE and AP-1) revealed a lowest K_d of 50 μM by ITC, with the association driven by the enthalpic contribution¹²⁵. Use of a mutated leucine region by Dwarki *et al*, whereby the leucines involved in CREB's homodimer interface were mutated to valine, revealed that a double mutant (L318/325V) completely abrogated target site binding¹³⁵.

A lack of structural information exists for a native protein sequence monomer-DNA complex however. However, a model basic region peptide was disordered in the absence of DNA, and addition of double stranded DNA induced helical conformational change¹³⁰. Single stranded DNA did not induce a conformational change, whereas increasing the ionic strength the absence of DNA did increase the

helical content¹³⁰. Therefore, the contrast of evidence for target site recognition by the monomeric constructs of bZIPs is striking.

1.6.3 Forced dimer bZIP constructs to investigate the pre-formed dimer pathway

Direct investigations into describing a pre-formed dimer pathway have been utilised with "forced" dimer versions of bZIP constructs^{79,136}. A forced dimer version of the GCN4 bZIP was used, with a C-terminal GGC linker, which revealed⁷⁹ in the bZIP length of construct no change in binding affinity to target DNA was seen compared to WT. However, with the basic region only version of GCN4, a disulphide linked "forced" dimer did associate to target DNA sequences⁷⁹ with a coupled folding upon binding interaction. Therefore, a potential role of the dimer, with increased stability when associated to the DNA, is to kinetically trap the protein at sites in the DNA.

1.7 Target search process of transcription factors

Transcription factors are required to associate to target DNA sequence site that, under *in vivo* conditions, occur at a very low amount compared to those of a non-target site (in *E. coli* lac operator site is 17 bp amongst a genome size of 4.6 Mbp for example¹³⁷). Despite this, the rate constant was $\sim 10^{10} \text{ M}^{-1} \text{ s}^{-1}$ for lac repressor to the operator site¹³⁸.

Work by Riggs *et al* began the development of several theories of how DNA binding proteins were able to associate to their targets, the majority of which were summarised by Berg *et al* in 1981¹³⁹. As a non-target sequence can be defined as being $m \pm 1$ out of frame from the reading frame of the target DNA sequence (where m is the position of the first bp of the target sequence), then for target search to occur effectively, the non-target binding must confer a kinetic advantage. Berg *et al* therefore defined this being a two (or more) step process for protein to associate to the target¹³⁹.

Two types of dissociation events have been determined to occur during a protein target search for its DNA target sites: macroscopic and microscopic dissociation.

Macroscopic dissociation would be essentially a full dissociation of the protein from the DNA, followed by a random association to another distinct site (Figure 1.12). On the other hand, microscopic dissociation releases the protein from the DNA but to a limited extent, whereby re-association occurs near to the original site. A combination of macro- and microscopic dissociation thus occurs to find the target¹³⁹.

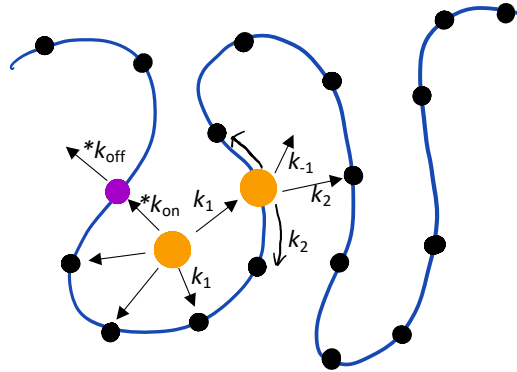


Figure 1.12: Target search of proteins to their cognate DNA site occurs via multiple macro- and microdissociation events, coupled with sliding, hopping and intersegmental transfer. Diagram illustrating the various mechanisms that a DNA binding protein could use for productive target search. Rates for the various events are illustrated where possible, assuming a two-step kinetic equation. * indicates a k_{on} and k_{off} to target sites. Black spots are non-target sequences, purple spot is target sequence, orange spot is protein.

In addition to these mechanisms, sliding, intersegmental transfer and DNA looping have been proposed as mechanisms for efficient target search^{139,140}. Historically, the rate of sliding was the simplest to extract due to experimental limitations (Table 6.1 contains a review of the subset of experimentally determined values). Sliding is completed in one dimension, following the dimensionality of the DNA molecule and will occur either until the protein reaches the target site or dissociates from the non-target DNA after an effective sliding length (λ). In more recent years, single molecule approaches have been able to determine the sliding rate *in vivo* and *in vitro*, with LacI being very well characterised^{137,141–143}. LacI is known to perform the 1D diffusion or sliding along DNA at a very fast but variable rate (2.3×10^{-12} to $1.3 \times 10^{-9} \text{ cm}^2 \text{ s}^{-1}$)¹⁴⁴ *in vivo*. For eukaryotic DNA binding proteins, the tumor suppressor p53 has been the most well characterised, with several studies

identifying the association rate to the target sequence and sliding along non-target DNA that is significantly slower than that of the bacterial LacI^{145,146}.

Conversely, the significance of intersegmental transfer has only recently been able to be characterised. Intersegmental transfer is defined as being the translocation of a DNA binding protein between two distinct sites, via an intermediate where the protein is associated to both sites at the same time (often found via a dimeric protein or a DBD with two DNA binding domains). The DBD of the glucocorticoid receptor was shown to have increased rates of dissociation from DNA upon higher DNA concentration, which would aid to increase the rate of association of the target¹⁴⁷. Structural and kinetic characterisation of the zinc finger protein Egr-1 identified a highly dynamic zinc finger 1 domain in the non-target DNA complex¹⁴⁸. This may be linked to the dominance of the intersegmental transfer pathway over the sliding pathway for Egr-1 target search¹⁴⁹. NMR on Oct-1 showed using residues within the two HOU domains that Pou_s exchanged faster between two DNA sequences that differ by 2 bp only¹⁵⁰, but that the rate of interchange between the two states did not fully account for intersegmental transfer and involved a contribution to hopping as well. DNA looping has only recently been proposed, and may itself be indistinguishable from intersegmental transfer¹⁴⁰.

Conformational switches have been experimentally determined to show a search and recognition mode during target search. For instance, the zinc finger protein Egr-1 showed that the zinc finger 1 domain showed limited chemical shift perturbation changes, by NMR, from the apo to non-target DNA states, confirmed by residual dipolar couplings¹⁴⁸. Zandarashvili *et al* then expanded this investigation further by performing ion-pair mutagenesis between the zinc finger - zinc finger and zinc finger-DNA interfaces, which pushed the conformational equilibrium towards either the recognition or search modes¹⁵¹. A dimeric lac headpiece in complex with non-target DNA was shown to undergo a "tilt" of $\sim 25^\circ$ compared to the target bound complex¹⁵², with many of the same residues involved in the target complex which confer the specificity now contacting the DNA phosphate backbone in the non-target complex.

The underlying DNA sequence appears to play a marked role in the rates of target search for a DNA binding protein. Using the Engrailed transcription factor from *Drosophila melanogaster*, Castellanos *et al* identified a relatively small K_d change when the binding site diverged quite considerably from the target sequence¹⁵³. Modelling the binding free energy landscapes for flanking DNA sequences or the target, a rough energy landscape profile was observed (similar to that for protein folding energy landscapes)¹⁵³, indicating that flanking regions have local energy minima that may aid in kinetically trapping the protein to regions near to the target sites. The symmetry of the DNA sequence flanking the proteins DNA target site was also shown to enhance the binding occupancy of transcription factors surrounding target DNA¹⁵⁴, with a high-throughput screen of 10,000's DNA sequences. All of this provides evidence for an antenna effect, whereby association of the DNA binding protein to a site in local proximity to the target will aid in the search process.

1.7.1 Target search mode of bZIP proteins

In the case of bZIP proteins, little to no evidence has been recorded as to how these proteins perform target search. A large number of those proteins that have been studied contain structured DNA binding domains, often joined by flexible linkers to the other domains. However, disordered DNA binding domain containing proteins have not been investigated as thoroughly. For instance, a disordered DNA binding domain may have significantly more than a single search mode conformation, which was identified with CytR¹⁵⁵. The DNA sequence itself provided a significant regulatory role in structure formation within CytR by charge screening. The limited structural evidence in the literature would suggest that a similar mechanism would occur, due to the smaller increase in helical content being seen by CD when bZIP proteins are in complex with non-target DNA compared to target DNA^{127,128,131}.

1.8 Aims of the thesis

The role of protein disorder in transcription covers a wide range of functions. However, the role of disorder within DNA binding domains in particular remains

poorly characterised. Using CREB as my model system, I will have to characterise both the dimer and monomer pathways, in order to determine which is preferred. This will be completed using a range of biophysical and structural techniques such as equilibrium affinity assays, kinetics assays using stopped flow, circular dichroism and NMR spectroscopy. Therefore, my main aims for the thesis are as follows:

1. Determining the homodimer affinity of CREB bZIP in order to start to build a model of how CREB associates to DNA. The structural propensity of the basic region will also be examined to begin to describe on both a structural and kinetic level. This is studied in Chapter 3.
2. Validating the specificity of CREB bZIP for CRE DNA and the dynamics of the basic region during the target search state when associated to non-target DNA. This is covered in Chapter 4.
3. Focusing on a monomeric CREB bZIP construct to determine if the basic region is sufficient for target DNA recognition and what the mechanism could be? This is examined in Chapter 5.
4. Does altering the helical propensity in the bZIP domain influence the target search kinetics and what role does the DNA sequence itself have to enable CREB to perform productive target search? This will be focused on in Chapter 6.

2

Materials and Methods

Contents

2.1	Buffers and Reagents	33
2.2	Chemically competent <i>Escherichia Coli</i>	35
2.3	Cloning and mutagenesis	36
2.4	DNA used for biophysics experiments	40
2.5	Elongation of probe DNA for sliding search rate kinetics	40
2.6	Expression of His₆-GB1-CREB bZIP, incorporating isotope enrichment for NMR	46
2.7	Purification of CREB bZIP	47
2.8	Protein concentration determination and buffer exchange	48
2.9	Nuclear Magnetic Resonance of CREB with DNA	50
2.9.1	NMR data acquisition parameters	50
2.9.2	Analysis of DNA titrations using chemical shift perturbation	54
2.9.3	Assignment of CREB complexes and determination of secondary structure and dynamics	54
2.10	Urea denaturation at equilibrium	55
2.11	Thermal denaturation at equilibrium	55
2.12	Labelling of CREB with extrinsic fluorophores	56
2.13	Fluorescence measurements of CREB and CREB/DNA complexes	56
2.13.1	Homodimer affinity with labelled CREB bZIP	57
2.13.2	CREB equilibrium affinity with fluorescently labelled DNA	57
2.13.3	Isothermal titration calorimetry (ITC)	58
2.14	Circular Dichroism (CD) for coupled folding upon binding	58
2.15	Kinetics	59
2.15.1	Folding and unfolding of bZIP dimer	59
2.15.2	k_{on} and k_{off} of protein to DNA and target search model	60

2.16 Bioinformatics	61
2.16.1 Contact map analysis and structural alignments	61
2.16.2 Promoter flanking analysis	62

2.1 Buffers and Reagents

Details and components of buffers used throughout this thesis are in Table 2.1. All buffers were weighed using a weighing or analytical balance. Buffer pH values were adjusted by addition of strong acid or base and verified using a pH meter calibrated using standard pH calibration solutions. For biophysical and NMR buffers, these were made up to the correct volume using volumetric flasks, before filtering. All buffers were filtered through a 0.22 μm membrane filter (cellulose acetate/nitrate, WhatmanTM) using a bottle top vacuum filtration unit. When tween-20 was used, this was added following filtering.

Table 2.1: Details on the components of buffers used throughout this work.

Name	Components	pH	Ionic Strength (I) (M)
NMR buffer	10 mM Tris-d ₁₁ , 150 mM NaCl, 10 mM MgCl ₂ , 1 mM NaN ₃ , 5% D ₂ O	6.5	0.189
MES Biophysical Buffer	10 mM MES, 150 mM NaCl, 10 mM MgCl ₂ , 0.05% tween-20	6.5	0.185
MES Biophysical Buffer - ITC	10 mM MES, 150 mM NaCl, 10 mM MgCl ₂	6.5	0.185
MES Refolding Buffer	10 mM MES, 150 mM NaCl, 10 mM MgCl ₂ , 8 M Urea	6.5	0.185
Tris Biophysical Buffer	10 mM Tris, 150 mM NaCl, 10 mM MgCl ₂ , 0.05% tween-20	6.5	0.185
Tris Biophysical Buffer - ITC	10 mM Tris, 150 mM NaCl, 10 mM MgCl ₂	6.5	0.185
Resuspension Buffer (RB)	100 mM Tris, 50 mM NaCl	8.5	0.150
Salt wash buffer (SWB)	100 mM Tris, 2 M NaCl	8.5	2.100
Imidazole Elution Buffer (IEB)	100 mM Tris, 50 mM NaCl, 2 M imidazole	8.5	0.150
Ion Exchange Load (IEL)	10 mM HEPES	7.0	0.010
Ion Exchange Elute (IEE)	10 mM HEPES, 2 M NaCl	7.0	2.010
Labelling Buffer	100 mM NaPO ₄	6.7	0.189
Disulphide oxidation Buffer	100 mM Tris	9.5	0.100
2x TY*	Tryptone (16g/L), yeast extract (10g/L), NaCl (5g/L)	7.0	n/a
2x TY plates*	As 2x TY, with 20g/L agar and autoclaved before pouring	7.4	n/a
M9 Minimal media (10x)*	Na ₂ HPO ₄ (61.31 g/L), KH ₂ PO ₄ (30 g/L) NaCl (5 g/L)	n/a	n/a

2.2 Chemically competent *Escherichia Coli*

Chemically competent *E. coli* cells were used for DNA amplification and also for expression of recombinant proteins. Strains of *E. coli* used for DNA amplification were XL-1 Blue (genotype: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacIqZΔM15 Tn10* (Tet^r)], Agilent Technologies) and DH5α (genotype: F- Φ80*lacZΔM15 Δ(lacZYA-argF)* U169 *recA1 endA1 hsdR17*(r_k⁻, m_k⁺) *phoA supE44 thi-1 gyrA96 relA1 λ*⁻, ThermoFisher). Further strains used for protein expression were C41 (genotype: F⁻ *ompT gal hsdS_B*(r_B⁻ m_B⁻(DE3), Lucigen) and BL21 (Gold) pLysS (genotype: B F⁻ *ompT hsdS*(r_B⁻ m_B⁻) *dcm*⁺ *Tet*^r *gal λ*(DE3) *endA Hte* [pLysS Cam^r], Agilent Technologies).

Stocks of chemically competent *E. coli* were amplified using a standard CaCl₂ protocol. In short, 100 mM CaCl₂ with 15% glycerol was chilled at 4°C before use. *E. coli* cells were plated onto 2x TY agarose plates (cloning strains without antibiotics and expression strains containing 34 μg/mL chloramphenicol only) and incubated overnight at 37°C. Single colonies were used for a starter culture in 2x TY, with the same antibiotic components as above. A 1:100 ratio of confluent starter culture to 100 mL of 2x TY with antibiotics as mentioned above, was incubated until the cell density reached OD₆₀₀ of 0.4-0.6. The flask of cells was placed on ice for at least 20 mins, before centrifugation at 3000 rpm for 10 mins at 4°C. Supernatant was decanted and pellet was resuspended in 1/5th of the initial volume in chilled 100 mM CaCl₂ with 15% glycerol solution, and incubated on ice for 20 mins. The suspension was centrifuged once again at 3000 rpm for 10 mins at 4°C, supernatant was removed and pellet was resuspended with 100 mM CaCl₂ with 15% glycerol in 1/25th of the original volume before aliquoting and flash freezing in liquid N₂.

For transformation, ~100 ng of vector was added to the cell suspension on ice, incubated for at least 5 mins before heat shock at 42°C for 45 s (DH5α used 30 s incubation). Treated cells were restored on ice for 2 mins, then 200 μL of 2x TY media was added and incubated at 37°C for 40 mins before plating on 2x TY agarose plates containing relevant antibiotics. Plates were incubated at 37°C overnight and were checked for colonies the following day.

2.3 Cloning and mutagenesis

pRSETA vector containing His₆-GB1-CREB bZIP (residues 285-339, legacy plasmid from Jane Clarke’s group, University of Cambridge) was used as the template plasmid for PCR mutagenesis. This vector contained a thrombin cleavage site (LVPRGS) between His₆-GB1 and a tobacco etch virus (TEV) cleavage site (ENLYFQG) between GB1-CREB bZIP. Polymerase chain reaction (PCR) was utilised in order to perform site directed mutagenesis, with *pfu* (Promega) and Q5 polymerases respectively. The protocol used for each polymerase is shown in Table 2.2 and Table 2.3, respectively.

Mutagenesis of the pRSETA-His₆-GB1-CREB bZIP was performed using the sets of primers shown in Table 2.4. All primers for mutagenesis were obtained from Sigma or Thermo Fisher as desalted. The primers used with *pfu* polymerase were designed with the rule of having a G/C at the 5’ end, and the mutation site in the centre of primer, where possible.

Table 2.2: *Pfu* polymerase PCR conditions

Component	Final Concentration
<i>Pfu</i> polymerase 10X buffer	1X
dNTP mix (10 mM each)	200 μ M each
forward primer	50 ng/ μ L
reverse primer	50 ng/ μ L
DNA template	50 ng
<i>Pfu</i> polymerase	1.25 U

Table 2.3: Q5 polymerase PCR conditions

Component	Final Concentration
Q5 reaction buffer 5X	1X
dNTP mix (10 mM each)	200 μ M each
forward primer	0.5 μ M
reverse primer	0.5 μ M
DNA template	1 ng
Q5 polymerase	0.02 U/ μ L

Conditions for the *pfu* polymerase and Q5 polymerase are shown in Table 2.5 and 2.6, respectively.

When using *pfu* polymerase, DpnI restriction enzyme (NEB) digestion is required, following manufacturer's protocol. DpnI digestion occurred o/n at 37°C. When using Q5 polymerase, the KLD Enzyme Mix (NEB) reaction was used on the PCR reaction product, following manufacturer's protocols. In both cases, the products of these reactions were added to chemically competent *E. coli*, as referred to in Section 2.2, for transformation.

Single colonies were selected for growth in 5 mL of 2x TY containing ampicillin (100 µg/mL) overnight at 37°C with 180 rpm of shaking. The following day, the cells were pelleted at 13,000 rpm and supernatant was drained. The cell pellets were used for isolation of the plasmid by Miniprep (QIAgen), following the manufacturer's protocol. Isolated plasmid concentration was determined by absorbance at 260 nm and purity by A_{260}/A_{280} . Plasmids were sequenced by Sanger Sequencing with the T7 promoter primer (Source Bioscience and Genscript).

Table 2.4: Primers used for PCR mutagenesis

Primer Name	Sequence (5'-3')	T _m (°C)	Polymerase used
CREB bZIP M291C forward	GTGAAGTGGCTCTGTGCAAAAATCGTGAGGCG	68	<i>pfu</i>
CREB bZIP M291C reverse	CGCCTCACGATTTTTCACACAGACGCACTTCAC		
CREB bZIP I326C forward	CCAAAATAAAACCCTGTGCGAAGAACTGAAG	62	<i>pfu</i>
CREB bZIP I326C reverse	CTTCAGTTCTTTCGCACAGGGTTTATTATTGG		
CREB bZIP S310C forward	GAAAGATATGTGAAATGCCCTGGAAAATCGTGTG	62	<i>pfu</i>
CREB bZIP S310C reverse	CACACGATTTTCCAGGCATTTTCACATACTCTTTC		
CREB bZIP A331C forward	GAAGAACTGAAGTGCCCTGAAAGACCTGTATAG	62	<i>pfu</i>
CREB bZIP A331C reverse	CTATACAGGTCTTTTCAGGCACCTTCAGTTCTTC		
CREB bZIP L318V forward	GTGGAGAACCAAAATAAAACCGTG	66	Q5
CREB bZIP L318V reverse	CTGGAAAATCGTGTGGCGGTT		
CREB bZIP L325V forward	CCAAAATAAAACCGTGATTTGAAGAACTGAAGCGC	55	<i>pfu</i>
CREB bZIP L325V reverse	GCGCCTTCAGTTCTTCAATCAGGGTTTATTTTGG		
CREB bZIP S337C forward	AAGCGCTGAAAGACCTGTAT T GCCATAAGTGATAAG	72	Q5
CREB bZIP S337C reverse	CTTATCACTTATGGCAATACAGGTCTTTTCAGCGCCTT		
CREB bZIP insG307G308G309G310 forward	GGCGGCGGCGGTATGTGAAAAGCCCTGGAAAATCGTG	69	Q5
CREB bZIP insG307G308G309G310 reverse	CTCTTTCTTTTTCGACGGCTCTC		
CREB bZIP insGGC C-term forward	TTGTTGATAAGAAATTCGAAGCTTG	62	Q5
CREB bZIP insGGC C-term reverse	CCACCCTTATGGCTATACAGGTC		

Table 2.5: PCR conditions when using pfu polymerase.

Stage	Temperature (°C)	Time (min)	No. of Cycles
Initial Denaturation	95	2	1
Denaturation	95	1	
Annealing	42-65	0.5	35
Extension	72	3.5	
Final Extension	72	5	1
Hold	4	Indefinite	1

Table 2.6: PCR conditions when using Q5 polymerase.

Stage	Temperature (°C)	Time (min)	No. of Cycles
Initial Denaturation	98	0.5	1
Denaturation	98	0.2	
Annealing	50-72	0.5	25
Extension	72	1.75	
Final Extension	72	2	1
Hold	4	Indefinite	1

2.4 DNA used for biophysics experiments

DNA was acquired as desalted from Life Tech and Sigma-Aldrich. Desalted oligos were resuspended in ddH₂O and absorbance of ssDNA was determined at 260 nm. Equimolar quantities of forward and reverse, if required, was added and underwent annealing in a PCR thermocycler from 95°C to 10°C in 0.2°C/s intervals. DNA was then buffer exchanged or diluted into the required buffer before use and absorbance measured for concentration determination.

Table 2.7: Details of DNA oligos used for equilibrium, kinetic and NMR experiments. * for dsDNA only. f and r are forward and reverse. * Supplied by Dr M. Kuravskiy

Name	Sequence (5'-3')	ϵ (M ⁻¹ cm ⁻¹)*	GC content (%)
AF488 CRE 14bp (SF) f	CCTGACGTCATCCG	236296	64.3
AF488 CRE 14bp (SF) r	CGGATGACGTCAGG	236296	64.3
CRE hairpin*	CCTGACGTCAGCCCCCTGACGTCAGG	n/a	n/a
nonCRE hairpin*	TCGGGCGGGGCGAGTGCCCCGCCCCGA	n/a	n/a
AF488 nonCRE 14bp f	GGCTAAAGCATTCT	225537	42.9
AF488 nonCRE 14bp r	AGAATGCTTTAGCC	225537	42.9
CRE 14 bp (NMR)	AGATGACGTCATCT	231705	42.9
nonCRE 14 bp f	GGCTAAAGCATTCT	225537	42.9
nonCRE 14 bp r	AGAATGCTTTAGCC	225537	42.9
nonCRE p1	GGCTAAATTTAGCC	225704	45.9

2.5 Elongation of probe DNA for sliding search rate kinetics

Various lengths of probe DNA were required to be synthesised in order to reproduce the experimental design that was used by Esadze and colleagues¹⁴⁹. The Vent polymerase conditions are shown in Table 2.8. The labelled probe DNA was purchased containing a 5'-AlexaFluor (AF) 488 dye (Life Technologies), with the various primers described in Table 2.9. Vent DNA polymerase was used with pHSG299 as the template, with the primers designed to add the somatostatin CRE motif at the 5' end via the forward primer, and then to extend the various lengths with the reverse primers. All probe DNA constructs were designed not to incorporate full or half somatostatin CRE sites. Components of the PCR reaction followed the manufacturer's protocol, except the template DNA concentration was 0.1 ng.

Table 2.8: PCR conditions for probe DNA elongation using Vent polymerase.

Stage	Temperature (°C)	Time (min)	No. of Cycles
Initial Denaturation	98	0.5	1
Denaturation	98	0.17	
Annealing	63	0.5	25
Extension	72	0.25	
Final Extension	72	10	1
Hold	4	Indefinite	1

Products of PCR elongation of probe DNA were verified using agarose gel electrophoresis. 4% agarose gels in 1x TBE were set, with 1 μ L per 50 mL gel of GelGreenTM dye, and run for at least 35 mins at 80 V. The agarose gel was visualised using a UV transilluminator. Labelled probe DNA constructs were purified using the QIAquick Nucleotide Removal Kit (QIAGEN), following the manufacturer's protocol. Due to the high volume of Buffer PNI used throughout the protocol, this was made in the lab using 40% (v/v) 5 M guanidium chloride and 60% (v/v) isopropanol. Purity of DNA constructs was verified by agarose gel electrophoresis, using the same conditions as above. DNA concentration was determined by absorbance at 280 nm, using extinction coefficient values calculated from the expected DNA sequence (IDT MolBiolTools) and the extinction coefficient of AF488 at 280 nm = 21300 M⁻¹ cm⁻¹, detailed in Table 2.10.

Table 2.9: Primers used for extension of probe DNA using pHSG299 as the template.

Name	Sequence (5'-3')	Tm °C
AF488 31 bp forward	CCTGACGTCATCCGGTAACTATCGTCTAGAG	n/a
31 bp reverse	CTCAAGACGATAGTTACCGGATGATGACGTCAGG	61
48 bp reverse	CTTACCGGGTTGGACTCAAGACG	59
56 bp reverse	CGTCTTGAGTCCAACCCGGTAAGACACGACT	61
68 bp reverse	CCAGTGGCGATAAGTCGTGTCTTAC	58
79 bp reverse	AGTGGCTGCTGCCAGTGGCG	63
93 bp reverse	CTAATCCTGTTACCAGTGGCTGCTG	58
106 bp reverse	ATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGC	61
123 bp reverse	TCTGTAGCACCGCCTACATACCTC	59
137 bp reverse	ACCACTTCAAGAACTCTGTAGCACCGCCTAC	61
153 bp reverse	TAGCCGTAGTTAGGACACCACTACAAG	58
175 bp reverse	ACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACC	61
203 bp reverse	AGGTAACCTGGCTTCAGCAGAGC	59

Table 2.10: Probe and competitor DNA sequences and properties used in the sliding rate kinetics experiments.

Name	Sequence (5'-3')	ϵ (mol ⁻¹ cm ⁻¹)	GC content (%)
31 bp	CCTGACGTCATCATCCGGTAACTATCGTCTAGAG	317100	52
48 bp	CCTGACGTCATCATCCGGTAACTATCGTCTTGAG TCCAAACCCGGTAAG	815760	56
56 bp	CCTGACGTCATCATCCGGTAACTATCGTCTTGAG TCCAAACCCGGTAAGACACGACT	949585	55
68 bp	CCTGACGTCATCATCCGGTAACTATCGTCTTGAG TCCAAACCCGGTAAGACACGACTTATCGC- CACTGG	1146668	56
79 bp	CCTGACGTCATCATCCGGTAACTATCGTCTTGAG TCCAAACCCGGTAAGACACGACTTATCGCCAC TGGCAGCAGCCACT	1326628	57
93 bp	CCTGACGTCATCATCCGGTAACTATCGTCTTGAG TCCAAACCCGGTAAGACACGACTTATCGCCAC TGGCAGCAGCCACTGGTAACAGGATTAG	1553817	55
106 bp	CCTGACGTCATCATCCGGTAACTATCGTCTTGAG TCCAAACCCGGTAAGACACGACTTATCGCCAC TGGCAGCAGCCACTGGTAACAGGATTAGC AGAGCGAGGTAT	1768950	55

123 bp	CCTGACGTCATCATCCGGTAACTATCGTCTTGAG TCCAACCCCGGTAAGACACGACTTATCGCCAC TGGCAGCAGCCACTGGTAACAGGATTAGC AGAGCGAGTATGTAGGCGGTGCTACAGA	2050154	55
137 bp	CCTGACGTCATCATCCGGTAACTATCGTCTTGAG TCCAACCCCGGTAAGACACGACTTATCGCCAC TGGCAGCAGCCACTGGTAACAGGATTAGC AGAGCGAGGTATGTAGGCGGTGCTACAGA GTTCTTGAAGTGGT	2275986	54
153 bp	CCTGACGTCATCATCCGGTAACTATCGTCTTGAG TCCAACCCCGGTAAGACACGACTTATCGCCAC TGGCAGCAGCCACTGGTAACAGGATTAGC AGAGCGAGGTATGTAGGCGGTGCTACAGA GTTCTTGAAGTGGTGCCCTAACTACGGCTA	2540142	54
175 bp	CCTGACGTCATCATCCGGTAACTATCGTCTTGAG TCCAACCCCGGTAAGACACGACTTATCGCCAC TGGCAGCAGCCACTGGTAACAGGATTAGCA GAGCGAGGTATGTAGGCGGTGCTACAGA GTTCTTGAAGTGGTGCCCTAACTACGGCT ACACTAGAGAACAGTATTGGT	2895013	52

203 bp	CCGCTGCGCCTTATCCGGTAACTATCGTCTTGAG TCCAAACCCGGTAAGACACGACTTATCGCCAC TGGCAGCAGCCACTGGTAACAGGATTAGCA GAGCGAGGTATGTAGCGGCTGCTACAGAG TTCTTGAAAGTGGTGGCCTAACACGGCTA CACTAGAAAGAACAGTATTTGGTATCTGCGC TCTGCTGAAGCCAGTTACCT	3351147	52
Competitor DNA 31 bp	GGCTAAAGCATTCTGCCACTTATTGGAAACACCTCGG	335893.8	50

2.6 Expression of His₆-GB1-CREB bZIP, incorporating isotope enrichment for NMR

pRSETA plasmid vector containing the insert for expression of His₆-GB1-CREB constructs was transformed to *E. coli* expression strains, as detailed in Section 2.2. BL21(Gold) pLysS strain was used as the expression strain, unless otherwise noted (the protocol was identical for all expression strains used in this work). Multiple colonies were taken from a 2x TY agarose plate containing ampicillin (100 µg/mL, unless otherwise stated) and chloramphenicol (34 µg/mL, unless otherwise stated) to inoculate a starter culture of 2x TY containing ampicillin, chloramphenicol and 0.2% glucose, which was incubated at 37°C overnight with 180 rpm of shaking.

The following morning, a 1:20 ratio of confluent starter culture was used to inoculate several litres of 2x TY with ampicillin, chloramphenicol and 0.2% glucose, in 1L UltraYield flasks (Thomson Instrument Company). Cell suspensions were incubated at 37°C with 180 rpm of shaking, before being induced with IPTG (1 mM) once OD₆₀₀ reached 0.6-1.0. For regular protein expression, cells were induced for 4 hours at 37°C with 180 rpm shaking. Cells were harvested by centrifugation at 5000 x g at 4°C for 15 mins before storage at -80°C until further use.

For expression with isotope labelling, a modification of the protocol mentioned by Marley and colleagues was used¹⁵⁶. Upon reaching OD₆₀₀ of 0.6-0.8, the cells were centrifuged using a JLA 8.1000 rotor at 5000 x g for 15 mins. Supernatant was removed and autoclaved 1x M9 minimal media (see Table 2.1) was used to rinse the cell pellet. The pellet was drained and 1x M9 minimal media containing FeCl₃, CaCl₂ and MgSO₄ (1 mM), with 4 g/L glucose and/or 1 g/L ammonium chloride as additives. When incorporating the isotopes, ¹³C₆-glucose (Cambridge Isotopes) and ¹⁵N-ammonium chloride (Cambridge Isotopes) were used. The pellet was resuspended in the Minimal Media buffer and incubated in UltraYield flasks for 45 mins at 37°C with 180 rpm shaking, before induction of protein expression with IPTG (1 mM) for 4 hours at 37°C with 180 rpm shaking. Cells were harvested and stored under the same conditions as regular expression.

2.7 Purification of CREB bZIP

Cell pellets stored at -80°C from Section 2.6 were thawed at RT and resuspended in ~ 50 mL of resuspension buffer (RB) for 3 L worth of bacterial pellet. For isotopically labelled proteins, PierceTM Protease Inhibitor tablets were added according to manufacturer's instructions. The cell suspension was sonicated with a rate of 15 s on, 15 s off, for a total 'on' time of 9 mins. The solution was then centrifuged in SS34 tubes with a JA-20 rotor at 20,000 rpm for 45 mins at 4°C . The fraction that the protein was located in was determined by SDS-PAGE, using NOVEXTM NuPAGETM 4-12% Bis-Tris gels, running at 200 V for 35 mins in MES running buffer (NuPAGE, ThermoFisher). All protein constructs were found to be in the supernatant in all cases.

The supernatant was filtered initially with $0.45\text{ }\mu\text{m}$ then $0.22\text{ }\mu\text{m}$ filters (Millex) using a syringe, before loading using an ÄKTA Fast Protein Liquid Chromatography (FPLC) with a RB pre-equilibrated HisTrap HP column (charged with Ni^{2+}) at 2 mL/min. All buffers are detailed in Table 2.1. The protein load step was followed by a RB wash at 5 mL/min for 10 CV, then a wash with salt wash buffer (SWB) for 100 mL per 1 L of initial cell pellet that was resuspended. A further RB wash for 10 CV at 5 mL/min was done before a gradient of 0-100% of Imidazole elution buffer against RB across 20 CV at 3 mL/min. Absorbance at 280 nm was measured during the ÄKTA FPLC method and 2 mL fractions were collected across the elutions. Fractions of interest were analysed by SDS-PAGE, with fractions containing protein of interest pooled.

His₆-GB1 was cleaved from CREB using TEV protease (synthesised in house). The His₆-GB1-CREB was incubated with 3 μM TEV protease during dialysis to 2L of RB + 1 mM DTT + 5 mM EDTA using a membrane with 3000 MWCO (Snakeskin), and was incubated o/n at RT. At this stage, either a second HisTrap step and/or an ion exchange step would be performed using ÄKTA FPLC. In the case of the second HisTrap step, protein would be loaded as described above. The loaded column is washed with 10 CV of RB at 5 mL/min. A gradient of 0-100% of salt wash buffer against RB across 20 CV at 3 mL/min was performed to elute the

cleaved CREB bZIP. For the ion exchange step, a pre-equilibrated SP HP cation exchange column (GE Healthcare) in ion exchange load (IEL) buffer was loaded with the dialysate at 3-5 mL/min. A 10 CV wash with IEL at 5 mL/min was performed, followed by a 0-100% gradient of ion exchange elute buffer at 3 mL/min. As with the HisTrap elutions, fractions where collected were absorbance peaks were seen on the Äkta system and analysed by SDS-PAGE.

2.8 Protein concentration determination and buffer exchange

Protein purity and concentration was determined by UV/Vis absorbance scans (Cary Spectrometer) using a QS Quartz cuvette (Hellma Analytics). Protein extinction coefficients were determined by the Gill and Von Hippel method¹⁵⁷, whereas proteins containing cystines incorporated the extinction coefficient determined by Pace *et al*¹⁵⁸. Protein constructs used throughout this work, with their various properties, can be seen in Table 2.11. The A_{260}/A_{280} ratio was used to determine DNA contamination, with a value of 0.42 used for pure proteins due to a lack of tryptophan residues in the protein influencing the ratio. Dilutions of proteins, when required, were weighed before measuring absorbance for accurate dilution ratios.

Proteins were concentrated, if required, using centrifugal filters with a 3000 Da MWCO (Millipore). For isothermal titration calorimetry (ITC), proteins were buffer exchanged using mini dialysis tubes into the ITC buffer overnight at RT. For all other experiments, mini dialysis tubes or PD10 buffer exchange columns were utilised, following the manufacturers instructions. To store proteins in their respective buffers, aliquots were made and flash frozen in liquid N₂ and stored at -80°C until further use.

Table 2.11: Recombinantly expressed protein constructs used in this work and their respective properties. * = C300/310/337S, ** = assuming dimer and cysteine is only species.

Construct	Residues	Name	Mutations	No of residues	MW - monomer	MW - dimer	$\epsilon(\text{M}^{-1} \text{cm}^{-1})$
CREB	285-339	CREB 285-339 pWT	C300/310/337S*	56	6725.83	13451.66	2560
	285-339	CREB 285-339 S337C	S337C	56	6741.89	13481.78	5240**
	285-339	CREB 285-339 FD	* GGC-C'	59	6943.07	13884.14	5240**
	285-315	CREB 285-315	*	31	3945.61	n/a	1280
	285-339 + 4	CREB 285-339ins4G306	* + ins4GE306	60	6954.03	13908.06	2560
	277-339	CREB 277-339 pWT	*	63	7578.73	15157.46	2560
	277-339	CREB 277-339 FD	* GGC-C'	67	7795.97	15589.94	5240**
	277-315	CREB 277-315	*	38	4798.51	n/a	1280
His ₆ -GB1-CREB	285-339	CREB 285-339 pWT	C300/310/337S*	143	16473.45	n/a	15930
	285-339	CREB 285-339 FD	* GGC-C'	146	16708.71	n/a	15930
	285-315	CREB 285-315	*	118	13711.25	n/a	14440
	285-339 + 4	CREB 285-339ins4G306	* + ins4GE306	147	16719.67	n/a	15930
	277-339	CREB 277-339 pWT	*	150	17344.37	n/a	15930
	277-339	CREB 277-339 FD	* GGC-C'	154	17561.61	n/a	15930
	277-315	CREB 277-315	*	125	14564.15	n/a	14440
	n/a	n/a	n/a	87	9765.64	n/a	12950

2.9 Nuclear Magnetic Resonance of CREB with DNA

2.9.1 NMR data acquisition parameters

All nuclear magnetic resonance (NMR) experiments were performed in NMR buffer (see Table 2.1), unless otherwise stated, containing 5% D₂O. Protein was buffer exchanged and concentration determined as described in Section 2.8, with concentrations of protein ranging from 0.02-1 mM and DNA ranging from 0.02-2.4 mM. pH was determined and adjusted if necessary between each experiment. The list of NMR experiments and their acquisition parameters is shown in Tables 2.12 to 2.14. NMR experiments were performed in D₂O-matched Shigemi tubes (5 mm diameter, Sigma) or in shaped NMR tubes (Bruker #Z106898). Minimum sample volumes used for each NMR tube were 280 and 320 μ L, respectively. Sodium trimethylsilylpropanesulfonate (DSS) was added as the chemical shift reference compound.

NMR spectrometers that were used throughout this work were located at Department of Biochemistry, University of Oxford, ranging from 600-950 MHz. All spectrometers were equipped with 5 mm TCI Cryoprobes for improved sensitivity, running with Bruker Avance III HD consoles. Experiments were acquired using Bruker Topspin (750 MHz and 600 M Hz = v3.2, 950 MHz = v3.6). BEST experiments were performed using the pulse sequences designed by Schanda *et al*¹⁵⁹. Data was processed using NMRPipe¹⁶⁰, with NUS data processing using iterative threshold sampling to reconstruct the data. NMR data was analysed using CcpNMR Analysis (v2.4 and 2.5)¹⁶¹.

Table 2.12: Acquisition Parameters for Triple Resonance NMR Experiments used in Backbone Assignment.

Experiment	Spectrometer	¹ H Pre-frequency (MHz)	Sweep Width (Hz)			No. of Complex Points			Reference Chemical Shift (ppm)			Temp (°C)	NUS (%)
			¹ H	¹⁵ N	¹³ C/[¹⁵ H]	¹ H	¹⁵ N	¹³ C/[¹ H]	¹ H	¹⁵ N/[¹ H]	¹³ C/[¹ H]		
HNCA	Bruker 600	600.133	6024.096	1703.578	3168.568	512	32	48	4.771	118.08	61.325	25	25
CBCA(CO)NH	Bruker 600	600.133	6024.096	1703.578	9057.971	512	32	60	4.772	117.262	42.734	25	25
HBHA(CO)NH	Bruker 600	600.133	6024.096	1703.578	[5617.976]	512	32	85	4.772	117.95	[4.772]	25	25
HNHA	Bruker 600	600.133	6024.133	1703.578	[5405.405]	512	128	20	4.771	121.586	[4.771]	25	0
HNCO	Bruker 600	600.133	6024.096	1703.578	2267.574	512	32	40	4.771	118.077	176.205	25	25
HN(CA)CO	Bruker 600	600.133	6024.096	1703.578	2267.574	512	32	96	4.771	118.073	176.201	25	25
CC(CO)NH	Bruker 600	600.133	6024.096	1703.578	10526.316	512	32	128	4.772	117.262	42.734	25	25
H(CCO)NH	Bruker 600	600.133	6024.096	1703.578	[5347.594]	512	32	128	4.772	117.95	[4.772]	25	25
CBCA(CO)NH	Bruker 750	749.914	7936.508	2132.196	11312.217	512	32	60	4.676	117.966	42.623	35	25
HBHA(CO)NH	Bruker 750	749.914	7936.508	2132.196	[6896.552]	512	32	85	4.676	117.966	4.676	35	25
HNCO	Bruker 750	749.914	7936.508	2132.196	2832.861	512	32	40	4.676	117.966	176.094	35	25
HN(CA)CO	Bruker 750	749.914	7936.508	2132.196	2832.861	512	32	40	4.676	117.966	176.094	35	25
BEST-HNCA	Bruker 750	749.914	7936.508	2132.196	3958.828	512	32	36	4.676	117.966	61.121	35	25

Table 2.13: Acquisition Parameters for ^{15}N edited NMR experiments.

Experiment	Spectrometer	^1H Frequency (MHz)	Sweep Width (Hz)			No. of Complex Points			Reference Chemical Shift (ppm)			Mixing times (ms)	Temp ($^{\circ}\text{C}$)	NUS (%)
			^1H	^{15}N	^1H	^1H	^{15}N	^1H	^1H	^{15}N	^1H			
^{15}N HSQC	Bruker 600	600.133	6024.069	1703.578	n/a	n/a	128	n/a	4.771	118.087	n/a	n/a	25	0
^{15}N BTROSY	Bruker 600	600.133	6024.069	1703.578	n/a	n/a	100	n/a	4.771	118.087	n/a	n/a	25	0
^{15}N HetNOE	Bruker 600	600.133	6024.069	1582.278	n/a	n/a	100	n/a	4.771	118.087	n/a	n/a	25	0
^{15}N TOCSY	Bruker 600	600.133	6024.096	1490.313	5434.783	512	32	128	4.772	117.827	4.772	10	25	25
^{15}N NOESY	Bruker 600	600.133	6024.096	1490.313	5434.783	512	32	128	4.772	117.827	4.772	75	25	25
^{15}N HSQC	Bruker 750	749.914	7936.508	1672.241	n/a	n/a	128	n/a	4.771	119.065	n/a	n/a	25	0
^{15}N HSQC	Bruker 950	949.664	10000	1695.418	n/a	n/a	128	n/a	4.771	118.045	n/a	n/a	25	0
^{15}N BTROSY	Bruker 750	749.914	7936.508	1672.241	n/a	n/a	128	n/a	4.724	119.017	n/a	n/a	30	0
^{15}N BTROSY	Bruker 750	749.914	7936.508	1672.241	n/a	n/a	128	n/a	4.676	118.966	n/a	n/a	35	0
^{15}N HetNOE	Bruker 600	600.133	6024.096	1582.278	n/a	n/a	100	n/a	4.771	118.08	n/a	n/a	25	0
^{15}N Trosy NOE	Bruker 600	600.133	6024.096	1582.278	n/a	n/a	100	n/a	4.771	118.074	n/a	n/a	25	0

Table 2.14: Acquisition parameters for the ^{13}C edited NMR experiments.

Experiment	Spectrometer	^1H Frequency (MHz)	Sweep Width (Hz)			No. of Complex Points			Reference Chemical Shift (ppm)		Temp ($^{\circ}\text{C}$)	NUS (%)
			^1H	^{13}C	^1H	^1H	^{13}C	^1H	^{13}C	^1H		
^{13}C HSQC	Bruker 600	600.133	6024.069	9900.99	n/a	1024	128	n/a	4.772	42.23	25	0
^{13}C HSQC	Bruker 750	749.914	7936.508	11764.706	n/a	1024	128	n/a	4.771	41.723	25	0
^{13}C HSQC	Bruker 950	949.664	9615.385	14492.754	n/a	1024	128	n/a	4.771	42.703	25	0
^{13}C HSQC	Bruker 750	749.914	7936.508	11764.706	n/a	1024	128	n/a	4.724	41.674	30	0
^{13}C HSQC	Bruker 750	749.914	7936.508	11764.706	n/a	1024	128	n/a	4.676	41.623	35	0
HCCO-TOCSY	Bruker 600	600.133	6024.096	5347.594	9900.99	512	128	48	4.772	42.143	25	25
HCCO-TOCSY	Bruker 600	600.133	6024.096	5347.594	9900.99	512	128	48	4.676	42.143	35	25

2.9.2 Analysis of DNA titrations using chemical shift perturbation

Where DNA titrations were performed at 35°C, a straight line was drawn from the apo to saturated state and assignments were made manually. At each titration point, the chemical shift perturbation (CSP, in ppm) could be calculated using equation 2.1¹⁶².

$$CSP = \sqrt{(\Delta^1H^N)^2 + (0.14(\Delta^{15}N))^2} \quad (2.1)$$

If 6 or more titration points were performed, an estimate of the binding affinity (K_d) could be determined using equation 2.2:

$$Y = Rf + \frac{dR(K_d + ligand + X - \sqrt{(K_d + ligand + X)^2 - 4 \bullet ligand \bullet X})}{2 \bullet ligand} \quad (2.2)$$

where Rf is maximum CSP (or signal), dR is change in CSP (or signal), X is [DNA] with ligand being protein with fixed concentration and Y being one of fluorescence anisotropy/intensity or CSP.

2.9.3 Assignment of CREB complexes and determination of secondary structure and dynamics

Assignment of CREB bZIP/BR complexes were completed using the sequential resonance assignment strategy (reviewed by Prof. C. Redfield¹⁶³), using various triple resonance experiments. All apo and BR constructs experiments were performed at 25°C (aside from CREB 285-315 with CRE), with the remainder performed at 35°C. Temperature titrations in steps of 5°C were performed to track assignments between 35°C and 25°C.

Chemical shifts at the assigned temperatures were used to determine the secondary structure using TALOS+¹⁶⁴ and Secondary Structure Propensity (SSP)¹⁶⁵, using default parameters. As assignments were referenced with DSS, no re-referencing was used in the programs. As several chemical shifts could not be assigned, various combinations of chemical shifts were used in the SSP predictions.

2.10 Urea denaturation at equilibrium

Estimation of binding affinity using urea denaturation was performed with CREB 285-339 with 10, 50 and 250 μ M. Tris biophysical buffer was used at this stage. Equimolar CREB solutions in Tris biophysical buffer with and without 8 M urea were mixed in various ratios for a range of concentrations between 0 M and 8 M urea. Each CREB concentration was performed at 25°C. Mixed solutions were incubated for 1 hour before analysis. Horiba FluoroMax-4 Spectrofluorometer was used for analysis with excitation wavelength at 278 nm and emission at 308 nm, slit width 5 nm, equilibration time 0.5 min. All conditions were performed in triplicate. Data was fitted using Graphpad Prism (v8) with the following equations provided by Dr. S. Shammash.

$$y = (a_f + b_f[D])(1 - Fu) + (a_u + b_u[D]) \bullet Fu \quad (2.3)$$

where a_f , b_f , a_u and b_u are used to define the folded and unfolded baselines, $[D]$ is denaturant concentration and Fu is fraction unfolded. Fu is defined by equation 2.4:

$$Fu = \frac{-K_d + \sqrt{K_d^2 + 8K_d[P]_T}}{4[P]_T} \quad (2.4)$$

where $[P]_T$ is total protein concentration and K_d is defined in equation 2.5:

$$K_d = \exp \frac{RT \ln[P]_T - m([D]_{50\%} - [D])}{(RT)} \quad (2.5)$$

where $[D]_{50\%}$ is denaturant concentration for 50% unfolded protein.

2.11 Thermal denaturation at equilibrium

Thermal denaturation of CREB 285-339 (pWT and FD) was performed using Circular Dichroism (Jasco J-800) and DSC (Malvern VP Capillary DSC) in Tris biophysical buffer. DSC was performed with and without 0.5% tween-20. DSC was performed with ramp rate of 60°C/hr, but limited signal change was observed at a range of protein concentrations. CD data was collected with ramp rate of 2°C/min, with intervals of 0.5°C being stored. Data was analysed with the following

equations provided by Dr. S. Shammas, starting by manually extracting the folded and unfolded baselines using linear regression. Fraction unfolded (Fu) was then calculated using equation 2.6:

$$Fu = (\theta_x - l_b)/(u_b - l_b) \quad (2.6)$$

where θ_x is MRE at each temperature, l_b being the folded baseline equation from the linear regression and u_b the unfolded baseline. The K_d at each temperature was then evaluated using equation 2.7.

$$K_d = (2 \times Fu^2 \times [P])/(1 - Fu) \quad (2.7)$$

Following this, a plot of $\log(K_d)$ against temperature was used with linear regression to extract the K_d at 25°C.

2.12 Labelling of CREB with extrinsic fluorophores

CREB 285-339 S337C (50 μ M) was buffer exchanged into 100 mM sodium phosphate, pH 6.7 and incubated with 5-fold excess of TCEP for 30 mins at 50°C. 200 μ M CPM (7-Diethylamino-3-(4'-Maleimidylphenyl)-4-Methylcoumarin) was added and incubated for 2 hours at 50°C. Protein/Dye mixture was diluted 4-fold into IEC start buffer and loaded to an SP column, with the same method used as for the protein purification in Section 2.7. Mass spectrometry (ESI-MS) was used to determine efficiency of labelling to be 100%.

2.13 Fluorescence measurements of CREB and CREB/DNA complexes

All equilibrium fluorescence intensity and anisotropy measurements were performed using Horiba FluoroMax-4 Spectrofluorometer. Temperature was controlled by water bath at 25°C. Cuvette was Hellma 105-250-15-40. Anisotropy value was defined as equation 2.8:

$$r = \frac{I_{vv} - GI_{vH}}{I_{vv} + 2GI_{vH}} \quad (2.8)$$

with G being defined as $I_{\text{VH}}/I_{\text{HH}}$. VH and VV are defined as the Vertical and Horizontal positions of the two polarisers within the fluorimeter and the fluorescence intensity observed after passing through each. Where intensity was measured, it was either measured directly (i.e. without the anisotropy mode) or indirectly with equation 2.9:

$$I = I_{\text{VV}} + 2I_{\text{VH}} \quad (2.9)$$

2.13.1 Homodimer affinity with labelled CREB bZIP

Fluorescence intensity and anisotropy of CPM-labelled CREB 285-339 pWT S337C was measured using excitation at 384 nm and emission at 470 nm, with 5 nm slit widths and 0.5 s integration time. Serial dilution of labelled CREB was fitted with equation 2.10.

$$Y = F + dF \times (1 - 0.25 \times (-(K_d/x) + \sqrt{((K_d/x)^2) + (8 \times (-K_d/x))})) \quad (2.10)$$

F is signal of free DNA, and dF is signal change.

A competition experiment whereby 5 nM labelled CREB bZIP was mixed with varying concentrations of pWT was fitted with equation 2.11 from Daugherty *et al*¹⁶⁶.

$$F = F_{\text{max}}(C/(C_{50} + C)) \quad (2.11)$$

where F_{max} is the maximum fluorescence, C is concentration of monomeric protein and C_{50} being the concentration of protein at which 50% of the labelled protein is bound to pWT.

2.13.2 CREB equilibrium affinity with fluorescently labelled DNA

AlexaFluor488 labelled DNA constructs were purchased desalted from Sigma. 5 nM labelled DNA was measured with varying concentrations of CREB protein (pWT and FD concentrations assumed as dimer) using the serial dilution strategy. CRE-containing DNA sequences with pWT and FD bZIPs were measured by intensity, with all other constructs and non-CRE containing sequences using anisotropy. Data was fitted to the one-site binding model (equation 2.2).

2.13.3 Isothermal titration calorimetry (ITC)

ITC was performed using PEAQ ITC (Malvern). CREB and DNA constructs were buffer exchanged to Tris biophysical buffer without tween-20 using mini-dialysis tubes inserted into a float, which were placed into a 1 L beaker with parafilm enclosing the beaker. The dialysis took place overnight at RT with stirring of the solution. Concentrations were determined prior to the experiment and necessary dilutions were made using the dialysis buffer and weighed.

Experiments were performed by having protein or DNA in cell, with DNA or protein in the syringe. Controls of buffer to buffer, protein to buffer, DNA to buffer and vice-versa were performed and used to subtract the necessary controls on the PEAQ-ITC Analysis Software (Malvern). Data was fitted using this software to one-site binding, with component concentrations fixed during the fitting. The system was cleaned following each use following manufacturer's protocol.

2.14 Circular Dichroism (CD) for coupled folding upon binding

Equimolar protein (monomer) and DNA were mixed together in Tris biophysical buffer with tween-20. Spectra were acquired at 25°C, with 4 accumulations per sample, from 240-200 nm in 0.5 nm intervals. Buffer signal was subtracted from the samples containing protein/DNA. Three independent measurements were acquired per construct. Mean residual ellipticity (MRE) was calculated using the conversion in equation 2.12, where θ is CD signal, n in number of residues and l is pathlength (cm). Signal from the DNA CD spectra was subtracted from the protein signal in the combined plot, otherwise a difference plot was shown. Cuvettes used were Hellma 110-1-40, 106-0.2-40 and 1042002B-10-40 with pathlengths of 1, 0.2 and 10 mm respectively.

$$MRE = \frac{\theta \times 0.1}{n \times l \times [Protein]} \quad (2.12)$$

2.15 Kinetics

All kinetics experiments were performed with stopped flow apparatus (SX-20 and SX-18, Applied Photophysics). Deadtimes for experiments were 1 or 2.5 ms, depending on the experiment and data preceding those times were removed from further analysis. For homodimer K_d estimates, 1:10 mixing was employed otherwise 1:1 mixing was performed.

2.15.1 Folding and unfolding of bZIP dimer

CREB 285-339 and 277-339 constructs underwent unfolding and refolding experiments using urea for homodimer K_d determination. 250 μ L and 2.5 mL drive syringes were utilised with the protein (in either MES biophysical buffer for unfolding or MES biophysical buffer plus urea for refolding) in the 250 μ L syringe and MES biophysical buffer with varied concentrations of urea in the 2.5 mL syringe. All buffer components and solutions were weighed before mixing. Stopped flow experiments took place at 25°C, with the excitation wavelength at 278 nm, slit widths 2 nm for both excitation and emission. A 305 nm cut-off filter was used for intrinsic tyrosine fluorescence measurements.

The rate equation for this interaction would be (equation 2.13).



To determine the k_{on} and k_{off} for folding of the bZIP homodimer, equations generated by Dr. S. Shammass were used that assumed that the rate of folding and unfolding reactions were significant to one another, and were fitted using Profit software (QuantumSoft). The final script is inserted into the appendix, as well as the unfolding optimised equation. Solving for the change in A,

$$\frac{d[A]}{dt} = -k_1[A]^2 + k_2P - k_2[A] \quad (2.14)$$

as $[C] = (P - [A])/2$, where P is the total protein concentration in the solution. Solving for dt and writing in terms of K (where K is K_d),

$$-k_1 dt = \frac{d[A]}{[A]^2 + K[A] - KP} \quad (2.15)$$

where the solving the roots of the quadratic in the denominator of equation 2.15 was

$$\frac{-K \pm \sqrt{K^2 + 4KP}}{2} \quad (2.16)$$

Z can then be defined as $\frac{\sqrt{K^2 + 4KP}}{2}$ and equation 2.15 then becomes

$$-k_1 dt = \frac{d[A]}{([A] + \frac{K-Z}{2})([A] + \frac{K+Z}{2})} \quad (2.17)$$

Solving using partial fractions and equalling the coefficients would lead to

$$-k_1 dt = \frac{d[A]}{Z([A] + \frac{K-Z}{2})} - \frac{d[A]}{Z([A] + \frac{K+Z}{2})} \quad (2.18)$$

Integration of equation 2.18 would result in,

$$N \bullet e^{-k_1 t Z} = \frac{[A] + \frac{K-2Z}{4}}{[A] + \frac{K+2Z}{4}} \quad (2.19)$$

This solution enables the extraction of k_1 and the K_d by directly fitting the stopped flow data. A plot of $\ln(k_1)$ or $\ln(K_d)$ against [Urea] leads to the rate of refolding and K_d under ideal conditions. The unfolding arm was also fitted to equations that assumed the unfolding rate was the dominant rate, defined by Milla and Sauer¹⁶⁷.

2.15.2 k_{on} and k_{off} of protein to DNA and target search model

1:1 mixing of CREB either alone or in the presence of the competitor DNA was carried out with AlexaFluor labelled dsDNA by stopped flow fluorescence. Solutions were weighed before use for accurate concentration determination. For measurements performed by intensity, [DNA] was fixed for k_{on} at 2.5 nM (in the cell). For k_{off} , the [DNA] was 5 nM. For measurements performed by anisotropy, a g-factor was measured to be 0.73. Timescales for the experiments were optimised per condition, and data was collected on the log scale for 2000 points collected for each run. For

k_{on} and k_{off} experiments, data was averaged then fitted to a single (equation 2.20) or double exponentials (equation 2.21) to extract the k_{app} .

$$Y = Ae^{-kt} + C \quad (2.20)$$

$$Y = Ae^{-k_1t} + Be^{-k_2t} + C \quad (2.21)$$

In equations 2.20 and 2.21, A and B are the amplitude, t is time and C is the Y-intercept. A plot of the k_{app} against the independent variable and fitting by linear regression enables the extraction of k_{on} from the gradient, whereas k_{off} was determined using equation 2.20. Attempts were made to fit the k_{off} data using an equation taking into account reassociation to the labelled DNA (equation 2.22)⁵⁸. Data was analysed using Prism (v8, Graphpad).

$$Y = k_{\text{off}} + (k_{\text{on}} \times \frac{1}{1 + \frac{x}{K_d}}) \quad (2.22)$$

As for the target search, a similar stopped flow set up was used as mentioned above. A single exponential was utilised for extraction of k_{app} , and a plot of k_{app} against the independent variable was fitted with linear regression when protein was on the X-axis. For DNA probe lengths and competitor DNA concentration dependence, matlab scripts were provided from Prof J. Iwahara and are available from Esadze *et al*¹⁶⁸. Scripts were used unmodified, with number of orientations for non-specific sites (ϕ) set to 1 for symmetric dimers. K_d values for the non-specific DNA parameter were obtained from the fluorescence equilibrium measurements.

2.16 Bioinformatics

2.16.1 Contact map analysis and structural alignments

Crystal structures used in this thesis were obtained from the Protein Data Bank (RCSB) and were visualised using PyMol Molecular Graphics Suite. Contact map and interface analysis was performed using PDBePISA (https://www.ebi.ac.uk/msd-srv/prot_int/pistart.html), using the default settings. Structural alignments were made using Pymol, with alignments made on the DNA molecules only.

2.16.2 Promoter flanking analysis

Searching for the DNA promoter sequences utilised published annotated genome datasets (*H. sapiens*: hg16 and GCF_000001405.39_GRCh38.p13_genomic; *S. cerevisiae* S288C - R64-2-1-20150113; *E. coli* K12 - R64-2-1-20150113). DNA promoter sequences were identified and the region flanking each site was extracted using a python script:

```
# Python script to find the position of the target sequence
(can be changed) and extract a 3000 base flanking
region either side
# Designed by Conor Kelly 21/7/20. Tested by Conor Kelly
21/7/20.
# Extract DNA sequence data of 3000 bp flanking residues
print ("Reminder - change DNA sequence to extract sequences within the code file and the number of bases the target is ")
import sys
import pandas as pd
import numpy as np
import regex as re
# import text file containing sequences
with open(sys.argv[1], 'r') as file:
    dna_lower = file.read().replace('\r\n', '')
    dna = dna_lower.upper()
# open list objects
    CRE_reading_Start = []
    CRE_antisense_Start = []
    run_start = []
    run_start_2 = []
# Start position of each DNA motif that wants to be
extracted (needs adjusting in code)
    runs = re.finditer(r"TGACGTCA", dna)

for match in runs:
    run_start = match.start()
    CRE_reading_Start.append(run_start)

# Make a dataframe using the matched start positions
df_CRE_start = pd.DataFrame(CRE_reading_Start, columns=['pos_start'])
```

```

# Region of DNA that will be stored and append to the
  previous dataframe.

df_CRE_end = df_CRE_start + 8
df_CRE_sense = pd.concat([df_CRE_start, df_CRE_end.rename(
    columns={'pos_start': 'pos_end'})], axis=1)
df_CRE_sense['Strand'] = '+'

df_final = df_CRE_sense

df_final['5prime_extreme'] = df_final['pos_start'] - 3000
df_final['5prime_adj'] = df_final['pos_start'] - 1
df_final['3prime_extreme'] = df_final['pos_end'] + 3000
df_final['3prime_adj'] = df_final['pos_end'] + 1

# Using the start and end of each region that we want to
  extract from above, code to do this and end up with a
  list of DNA sequences
# CRE sense orientation and 3 prime flanking residues
list_DNA_sequences = []

for row in df_final.itertuples():
    start_string = row[4]
    end_string = row[6]
    splicestring = dna[int(start_string):(end_string)]
    list_DNA_sequences.append(splicestring)

# Append the list of DNA sequences to the corresponding row
  of the dataframe and export this file with all of the
  information extracted.
# print(len(list_DNA_sequences))
df_DNA_sequences = pd.DataFrame(list_DNA_sequences, columns
    =['DNA_sequences'])
df_final['DNA_sequences'] = df_DNA_sequences['DNA_sequences']
df_final.to_csv(sys.argv[2] + '_extractedDNAsequences.csv',
    index=True, header=True)

```

This script could be run for each chromosome individually and then appended to a separate file containing all of the data for each chromosome. This next step details are for when data was filtered based on known CREB target sites. The bedtools software package was then used on advice of Dr N. Gray (Centre for Computational Biology, University of Oxford). Bedtools *closest* was used with

the known gene target positions for a certain transcription factor, in the case of CREB this was obtained from Zhang *et al*⁹⁶. This showed how close to the known targets that my extracted positions were. Filtering for the gene name from the dataset provided by the authors (without taking into account gene synonyms) yielded a dataset of known CREB targets and the immediate flanking regions of just those targets using the following script:

```
# Python script to remove all Bedtools matches between the
entire chromosome and the sites.
# Script designed by Conor Kelly 14/7/20. tested by Conor
kelly 14/7/20.
import sys
import pandas as pd

# Load the data files and add headers.
df_load = pd.read_csv(sys.argv[1], sep=" ", header=
    None)
# print(df_load)
df_load.columns = ["Chromosome_2", "MatchStart", "MatchEnd"
    , "Chromosome", "Start", "End", "Gene_Name", "N/A1", "
    Plus/Minus", "Source", "Chrom_Region", "N/A3", "ID", "
    DistanceAway_bp"]
print(df_load)

# Make a dataframe containing only the Chrom_Region column
selection "gene"
df_gene = df_load.loc[df_load['Chrom_Region'] == 'gene']
# print(df_gene)

df_gene["Gene_Name"] = df_gene["Gene_Name"].str.replace("
    gene-", "")
# print(df_gene)
df_gene.to_csv(sys.argv[2]+'__genenearby.csv', index=False,
    header=True)
```

From this point, the data analysis is consistent whether the data has been filtered by known positions or not. With the datasets with the flanking DNA sequences around targets, the GC content of regions of increasing lengths can be measured with descriptive statistics using the following script:

```
# Python script to generate a table of GC content from
various flanking DNA sequences. This script can be
repeated several times.
```

```

# Script takes in arguments: 1) file with the extracted
regions, 2) the flanking region number you want to use
and 3)name for the beginning of the summary file of GC
content for each length (usually the number of bp)
# Designed by Conor Kelly 21/7/20. Tested by Conor Kelly
21/7/20.

import sys
import pandas as pd
import numpy as np
import regex as re
import matplotlib
matplotlib.use('TkAgg')
import matplotlib.pyplot as plt
print("Reminder—uncomment sections regarding post_
      bedtools analysis!!")
# Load in the csv file that has extracted a string of 6008
residues, with CRE motif in the middle
df_first = pd.read_csv(sys.argv[1])

# Drop other columns that aren't required for calculation
of GC content.
# This section requires commenting depending on whether the
file is from new or from bedtools analysed data.
df_DNAseqonly = df_first.drop(['Unnamed: 0', 'pos_start', '
      pos_end', '5prime_extreme', '5prime_adj', '3
      prime_extreme', '3prime_adj'], axis=1)
# df_DNAseqonly = df_first.drop(['Unnamed: 0', '+ Start', '+
End', '5prime_extreme', '5prime_adj', '3prime_extreme',
'3prime_adj'], axis=1)

# As 3000 bases flanking the specific site was extracted,
the total length is 6008
# Flanking DNA range to calculate the GC content from.
flanking_range = int(sys.argv[2])

flank5primeregion = 3000 - flanking_range
flank3primeregion = 3008 + flanking_range

df_DNAseqonly['Flank_5prime_Sequence'] = df_DNAseqonly.
      DNA_sequences.str[int(flank5primeregion):3000]
df_DNAseqonly['Flank_3prime_Sequence'] = df_DNAseqonly.
      DNA_sequences.str[3008:int(flank3primeregion)]

df_DNAseqonly.to_csv(sys.argv[3]+'bp_flankingDNAsequences.

```

```

    csv', index=True, header=True)

# Load in df_DNAseqonly.csv file if this is required, just
    uncomment the next section and treat the first argument
    as the file you load in
# df_DNAseqonly = pd.read_csv(sys.argv[1])

df_5prime_flank = df_DNAseqonly.drop(['DNA_sequences', '
    Flank_3prime_Sequence'], axis=1)
df_split_data = df_5prime_flank.Flank_5prime_Sequence.str.
    split('(\w)', expand=True)
df_split_data2 = df_split_data.ix[:, 1::2]

df_3prime_flank = df_DNAseqonly.drop(['DNA_sequences', '
    Flank_5prime_Sequence'], axis=1)
df_split_data3 = df_5prime_flank.Flank_5prime_Sequence.str.
    split('(\w)', expand=True)
df_split_data4 = df_split_data3.ix[:, 1::2]

df_split = df_split_data2.append(df_split_data4)
print(df_split)
# df_split looks to be the dataframe with all splits in the
    data - will use this as the template for analysis.
# Convert the dataframe iteratively to a list
list_dna_data = df_split.values.tolist()

# Calculate GC content of each list item and create new
    list of results
length_list_dna_data = len(list_dna_data)

list_df_frac = []

for i in range(length_list_dna_data):
    GC_count = list_dna_data[i].count("C") + list_dna_data[i].
        count("G")
    GC_frac = float(GC_count)/len(list_dna_data[i])
    GC_frac_final = 100*GC_frac
    list_df_frac.append(GC_frac_final)

df_GC_frac = pd.DataFrame(list_df_frac, columns=[str(sys.
    argv[2])+'_bp'])

# Comment the next line out after the first run.
# df_first.to_csv('GC_dataforhistogram.csv', index=True,
    header=True)

```

```
summary_df_GC_frac = df_GC_frac[[str(sys.argv[2])+'_bp']].
    describe()
summary_df_GC_frac.to_csv(sys.argv[3]+'_GCcontent_summary.
    csv', index=True, header=True)

# Append the current flanking length to the df_GC_frac that
I ahve previously exported
# Un comment this section for all other lengths
df_append = pd.read_csv('GC_dataforhistogram.csv')
df_final = df_append.join(df_GC_frac)
df_final.to_csv('GC_dataforhistogram.csv', header=True)
```

Several runs of this script with different lengths produces a table that can be plotted as a histogram using the following script:

```
# Python script to use the GC_dataforhistogram.csv file and
plot the data as a histogram in a 1 by 8 layout.
# Designed by Conor Kelly on 7/5/20. Test 7/5/20.
```

```
import sys
import pandas as pd
import numpy as np
import regex as re
import matplotlib
matplotlib.use('TkAgg')
import matplotlib.pyplot as plt

# Load in the GC_dataforhistogram.csv file generated by the
GC_contenttableonly.py script. Headerlist creates the
name of the title of each file.
# Note: attach 00's to the beginning of title names, for
how many significant figures are present in the largest
flanking region looked at.
df_append = pd.read_csv('GC_dataforhistogram.csv')
# df_append = df_append_2.drop('Unnamed: 0', axis=1)
headerlist = list(df_append)

# Make a histogram of the data and save this
fig, ax = plt.subplots(figsize=(5,20))
df_append.hist(column=headerlist, sharex=True, sharey=False,
    , layout=(9,1), bins = 'auto', ax=ax, grid=False)
# plt.tight_layout()
# plt.subplots_adjust(bottom=0.15)
# plt.subplots_adjust(right=0.2)
plt.autoscale()
```



```
plt.xlabel( 'GC_content_(%) ' )  
plt.ylabel( 'Frequency ' )  
  
plt.savefig( sys.argv[1]+ 'test_GC_content_histogram.png' )
```

If a cumulative frequency distribution is required as plotted in this thesis, then a python module called LogoMaker¹⁶⁹ was utilised which created a table of values of A, T, G and C for all positions along multiple sequences. The GC content was calculated from this data and then plotted in Prism 8.

3

Characterisation of the dimeric state of the CREB bZIP domain

Contents

3.1	Introduction	69
3.2	Results	72
3.2.1	Kinetics identified a reliable and low K_d for the homodimer	72
3.2.2	Dilution of CREB bZIP using labelled protein constructs	77
3.2.3	Structural characterisation of the dimer	87
3.3	Discussion	109
3.3.1	Homodimer affinity estimate was consistent between a variety of techniques	109
3.3.2	DNA binding region remains dynamic in the apo-state of the bZIP complex	112

3.1 Introduction

A significant amount of previous work has gone into characterising bZIP domains and their oligomeric states (see Table 3.1). However the evidence that is available for CREB is limited as well as controversial, making the relative importance of the monomer and dimer states of CREB difficult to assess (Figure 3.1). In this chapter, CREB bZIP will be characterised with a similar range of methods previously applied to GCN4 bZIP, including kinetics and NMR, in order to determine if the

characteristics determining bZIP folding and dimerisation that were shown for GCN4 are general for bZIP containing proteins.

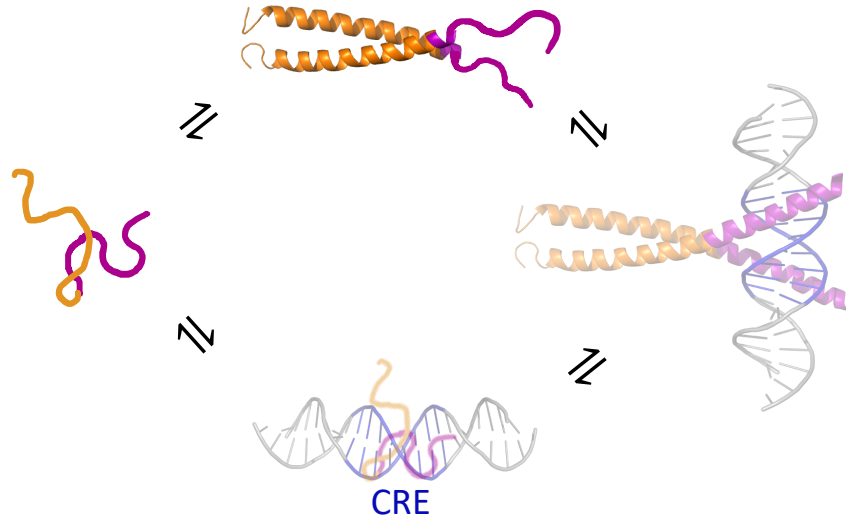


Figure 3.1: Simplified model of the possible association pathways of CREB bZIP to DNA. The focus of this chapter is the monomer-dimer equilibrium of CREB bZIP. The transparent parts of the model indicate the regions that are not focused on in this Chapter. Structure used was PDB: 1DH3.

Table 3.1: Dimer stability for various homo/hetero-dimeric proteins, determined by the dimer K_d determined *in vitro*. * - heterodimer, ** - thiol labelled protein, *** - bHLH protein

Protein	[Protein] (μ M)	K_d (μ M)	Buffer	Temp. ($^{\circ}$ C)	Method	Construct Length (# of aa)	Ref
His6-GABP-beta	2-13	1.8 ± 0.27	50 mM phosphate, pH 7.4, 100 mM KCl, 10% glycerol, 1 mM DTT	20	AUC	n/a LZ-like c-terminus	[170]
C/EBP GCN4	200	n/a (<25% dimeric) n/a	1.4 mM phosphate, pH 7.4, 2.7 mM KCl, 40 mM NaCl, 5 mM DTT	25	CD - dilution	54	[171]
GCN4		0.48 ± 0.23	100 mM sodium phosphate pH 7.0	25	Fluorescence	249-281	[92]
Max*** ATF2	50-1 30-3	0.6 3	1.4 mM phosphate, pH 7.4, 2.7 mM KCl, 137 mM NaCl, 1 mM DTT	25	CD - dilution	22-113 350-505	[126]
Fos/Jun**	n/a	0.054 ± 0.012 0.106 ± 0.04	1x PBS	25	FRET Polarisation	139-200/263-324	[126]
MyoD***	Varied	9.4 ± 2.2	20 mM ammonium acetate, pH 7, 100 mM KCl, 1 mM DTT, 0.1 EDTA	25	CD - dilution	n/a	[172]
E47***	15	0.02 ± 0.01	20 mM HEPES, 100 mM KCl, pH 7.5, 10 mM DTT		CD - urea denaturation	Full length	[90]
CREB	0.5-16	n/a (>16 μ M)			SEC		
CREB	5.00E-04	0.0006	25 mM Tris, pH 7.5, 1 mM EDTA, 100 mM NaCl, 1 mM DTT, 5% glycerol, 50 μ g/mL BSA	20	Fluorescence intensity (quenching)	Full length	[89]
Fos/Jun**	n/a	0.05	1x PBS, pH 7.4, 1% BSA, 0.1% Tween-20	25	Protein Array	143-217/272-327	[86]
ATF-2	n/a	3				359-428	

3.2 Results

3.2.1 Kinetics identified a reliable and low K_d for the homodimer

The characterisation of the homodimer stability of CREB bZIP was crucial to start to build a model for how CREB performs effective target search. With the large controversy over a monomer or dimer pathway, based on the homodimer stability of CREB and other bZIP domains varying over several orders of magnitude, getting information on the dimerisation reaction was required. Several methods were discussed in order to answer the question of how stable is the bZIP homodimer complex in the apo state. A kinetics based approach was chosen to determine the rates of folding/unfolding of the homodimer complex, as these rates will be dependent on the affinity of the homodimer (Figure 3.2). This utilised CREB bZIP pWT (residues 285-339), where the native cysteine residues were substituted by serines. The A260/280 ratio was used as a marker for the protein purity, with the ratio of 0.42 required based on the A260/280 ratio for tyrosine (there are no tryptophan residues present), as DNA contamination could markedly affect the affinity of the homodimer complex.

Stopped flow fluorescence was used, rapidly mixing folded or urea-unfolded CREB bZIP with various concentrations of urea, and observing the fluorescence intensity change of the intrinsic tyrosine fluorescence, as seen in Figure 3.3. There are two tyrosine residues in CREB bZIP, one located at the extreme C-terminus, Y336, in the disordered region from the crystal structure and another at position Y307. Kinetic traces showed a suitable intensity change across the various urea concentrations was observed. The data were fit to the equation that takes into account the reversibility of the folding reaction, which enables the extraction of the K_d . Figure 3.3a shows a range of urea concentrations that were mixed with CREB bZIP, with the amplitude highest at the lower urea concentrations and lowest towards 3 M urea (concentration within the stopped flow mixing cell). For refolding and unfolding arms of the experiment, the equations defined by Dr S.

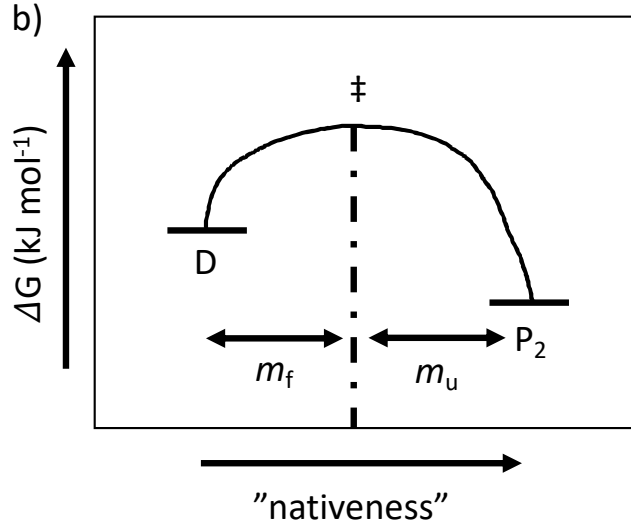


Figure 3.2: Cooperative folding model and energy of folding diagrams to describe the process by which CREB bZIP pWT can fold in the homodimer complex. a) Equation describing the folding/unfolding equilibrium of CREB bZIP pWT. P_2 is folded dimer protein complex, with D being denatured protein. P^* indicates a potential transition state during the folding/unfolding process. b) Gibbs free energy vs "nativeness" plot of the homodimer complex (P_2) for CREB bZIP. ‡ signifies the transition state. m describes the m -value for the folding and unfolding arms. As stoichiometry is changing, energy should strictly not be on the same scale as nativeness, however this is a widely used.

Shammas fitted the data remarkably well, as shown by the residuals for each fit (Figure 3.3b and c respectively).

The use of the reversible refolding equation allowed the extraction of both the folding rate as well as the homodimer K_d , at various concentrations of urea (Figure 3.4). A linear dependence of the natural logarithm of k_f and K_d vs urea was seen when the high urea concentration were removed, enabling the extrapolation of the rate and K_d at 0 M urea, as well as the m_f -value.

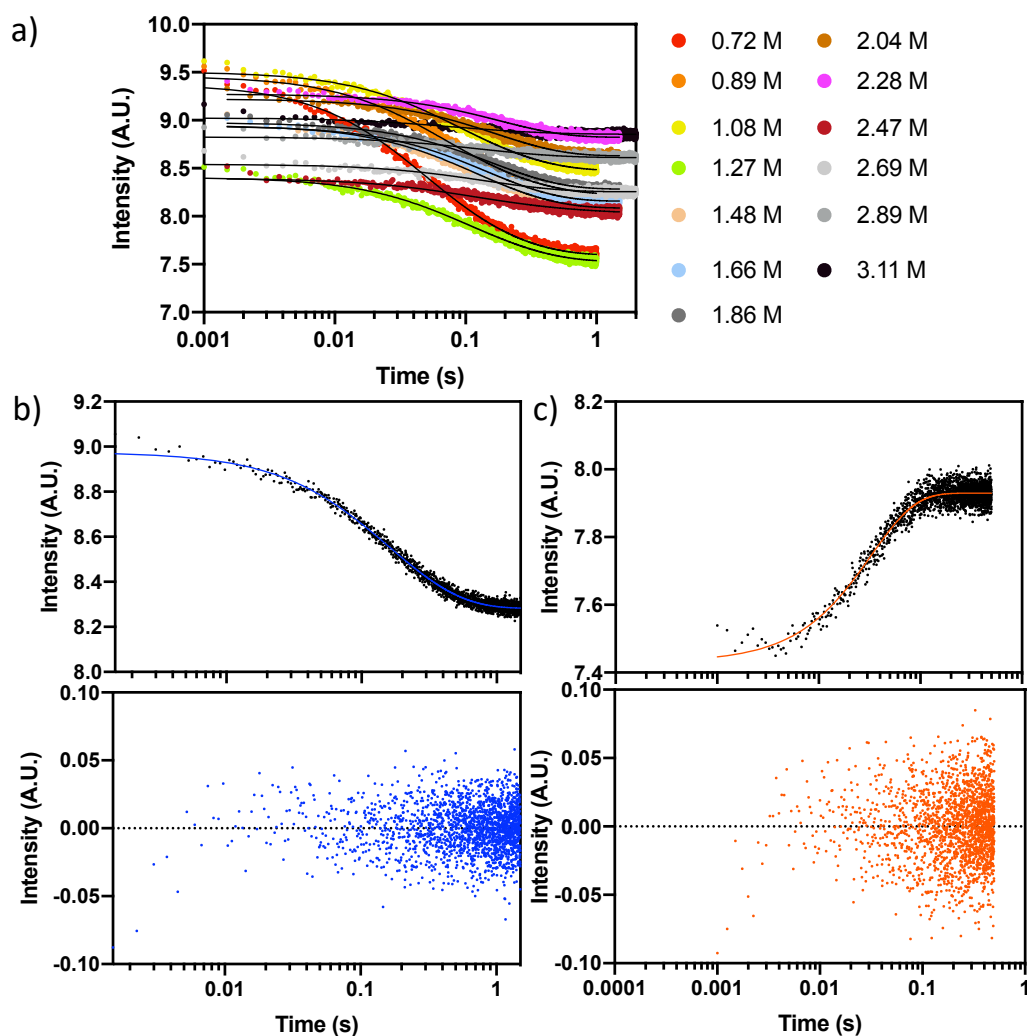


Figure 3.3: Folding of CREB bZIP can be estimated using stopped flow kinetics. a) Several concentrations of urea were used and fit to a reversible equation (section 2.15) in order to extract the folding rate and the K_d . b) Top - Representative stopped flow fluorescence intensity trace observed at 278 nm of 55 μ M CREB bZIP in 8 M urea mixed 1:11 with 1.2 M urea., fit to a reversible equation with the residuals shown below. c) Similar to b), but for the unfolding experiment where 55 μ M CREB bZIP was mixed with 6 M urea in a 1:11 ratio. Errors bars are shown as SEM.

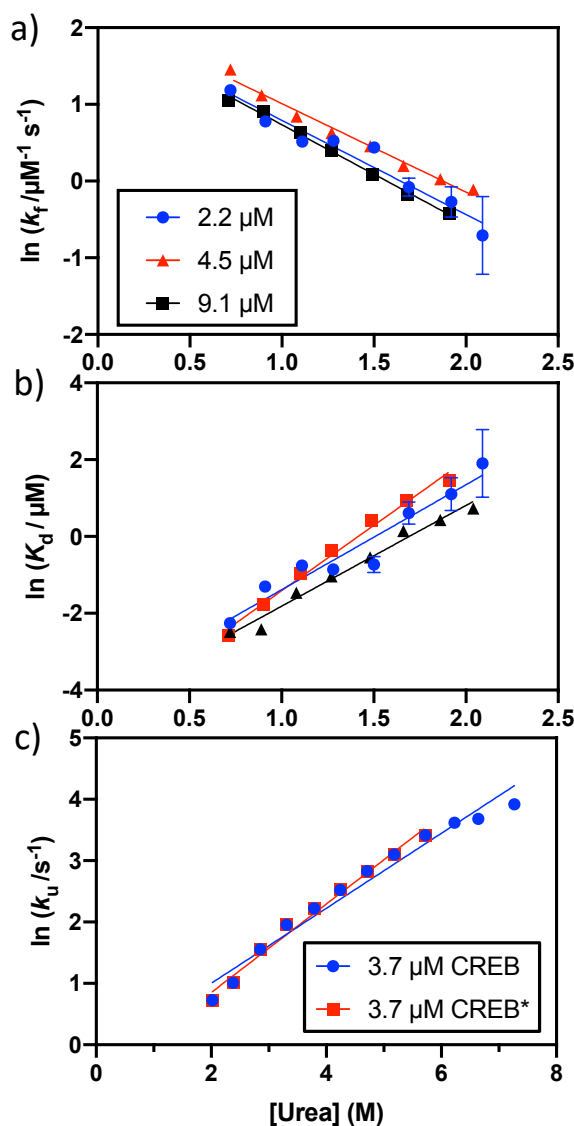


Figure 3.4: Extracted parameters from Figure 3.3 reveals k_f , k_u and K_d . a) Folding rate of CREB bZIP pWT was determined at three protein concentrations across a range of urea concentrations. Linear extrapolation of the fit was used to extract the $k_{f, \text{H}_2\text{O}}$ and m -value, at the Y-intercept and gradient respectively. b) Similar to a), with the $K_{d, \text{H}_2\text{O}}$ determined from the Y-intercept. c) Similar to a), showing the unfolding rate at a single CREB bZIP pWT concentration. * denotes the three high urea concentration points removed for a more suitable linear dependent fit. k_u determined from fitting unfolding data as described in section 2.15.1. Error bars are shown as SEM.

The $k_f^{\text{OM urea}}$ obtained from the refolding kinetics at the three protein concentrations were consistent within error, indicating no significant protein concentration dependence on the estimates, in support of the two-state model (Table 3.2). A similar observation with the m_f and K_d estimates was also seen. Therefore, a weighted average of these parameters can be used, resulting in a K_d of 9 ± 3 nM, k_f of $7.74 \pm 0.08 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ and m_f of $-0.7431 \pm 0.0003 \text{ kcal mol}^{-1} \text{ M}^{-1}$.

Table 3.2: Fitted parameters from the refolding kinetic analysis of CREB bZIP pWT shown in Figure 3.4. Errors are Standard Error (SE).

[CREB bZIP] (μM)	K_d (nM)	k_f ($\mu\text{M}^{-1} \text{ s}^{-1}$)	m_f ($\text{kcal mol}^{-1} \text{ M}^{-1}$)
2.2	20 ± 7	8 ± 1	-0.73 ± 0.07
4.5	10 ± 3	9 ± 1	-0.76 ± 0.02
9.1	8 ± 2	7.6 ± 0.3	-0.68 ± 0.04

For the unfolding rate, an irreversible unfolding fit (i.e. the unfolding rate under these high [Urea] conditions is significantly greater than the folding rate) was used from 5.72 M urea and greater, with a reversible fit for the remainder. The natural logarithm of the unfolding rate vs urea was plotted and a slight non-linear dependence along all the urea concentrations measured was observed. Removal of the highest three urea concentration points restored the linear dependence and both versions were fit with linear regression for comparison (Table 3.3).

Table 3.3: Fitted parameters from the unfolding kinetics of CREB bZIP pWT, shown in Figure 3.4. * denotes fitted parameters extracted from the linear fit where the three highest urea concentrations points were removed. Errors are SE.

[CREB bZIP] (μM)	K_d (nM)	k_u (s^{-1})	m_u ($\text{kcal mol}^{-1} \text{ M}^{-1}$)
3.7	n/a	0.8 ± 0.6	1.43 ± 0.07
3.7*	n/a	0.6 ± 0.1	1.53 ± 0.07

The error in the extracted K_d parameter from the reversible unfolding fits rendered it unreliable. The gradient of the linear fit to the natural logarithm of folding/unfolding rates vs urea yields the m -value, which is itself related to the solvent accessible surface area. The result of $m_u - m_f$ gives m_{equil} , which for CREB bZIP pWT was $2.27 \pm 0.07 \text{ kcal mol}^{-1} \text{ M}^{-1}$. The results from the kinetics yielded

a high level of confidence in the reliability of the estimate for the homodimer K_d , when measured from the perspective of the refolding experiment. Due to the controversy over the CREB homodimer stability, described in the section 3.1, one method was not sufficient evidence to progress with. Several alternative methods were attempted, with the urea equilibrium and CD thermal melts of CREB shown in the Appendix (Appendix Figures S3, S4 and S5 and Appendix Tables S2, S3 and S4). However, the homodimer K_d values determined from the urea denaturation at equilibrium experiments ranged from 9 pM to 500 nM, depending on whether m -value was fixed and/or protein concentration. Similarly, CD thermal melts revealed a range of homodimer K_d 's being estimated ranged from 13 pM to 15 nM, indicating the variability of extracted values from both of these techniques. However, the range within which these homodimer K_d values are in is consistent with those obtained in kinetics. The m -values are consistently lower in equilibrium based approaches than those by kinetics.

3.2.2 Dilution of CREB bZIP using labelled protein constructs

Method of labelling CREB bZIP

Due to the controversy about the bZIP homodimer stability, a further method was required to validate the kinetics approach that was used in section 3.2.1. Labelling of proteins with a fluorescent probe has commonly been used for characterising protein complexes, and has previously been used with a GCN4 system by Daugherty *et al*¹⁶⁶. With this system potentially being used in complex with DNA, the N-terminal labelling strategy used by Daugherty and colleagues would not be suitable for potential downstream applications. The extreme N-termini of a protein containing an amino group, and a labelling strategy involving amino groups would be likely to also label the N ϵ of lysine side chains; this was not considered further due to the enrichment of lysine residues in the basic region of CREB (Figure 3.5a).

As an alternative, thiol-reactive groups were chosen due to their increased specificity. As an added benefit, three positions within the CREB bZIP pWT

construct, S300/310/337, are cysteine in the native sequence (Uniprot ID: p16220). Using the crystal structure of CREB bZIP in complex with specific (CRE-containing) DNA²¹, a contact map between the protein chains was calculated and from this, the buried surface area can be estimated as shown in Figure 3.5b. Several potential candidates for site directed mutagenesis had little to no buried surface area, however the native cysteine locations of S300/310/337 were of particular interest. S300 forms part of the interface between the protein and DNA strand and was excluded from further choice of mutation sites, whereas S310 could have been a possible candidate. However, S337 is located near the C-terminus of the protein construct, in a loop section identified from the crystal structure, and thus was identified as the preferred option. From this point the construct will be denoted as CREB bZIP S337C.

To label the protein constructs, maleimide-linked fluorophores were utilised. This reaction produces a covalent link between the maleimide group and any free thiol group in the protein (Figure 3.6a), with a wide variety of chemical groups that can be attached to the maleimide. Labelling strategies involving free thiols and maleimides are well characterised, being available as commercial kits.

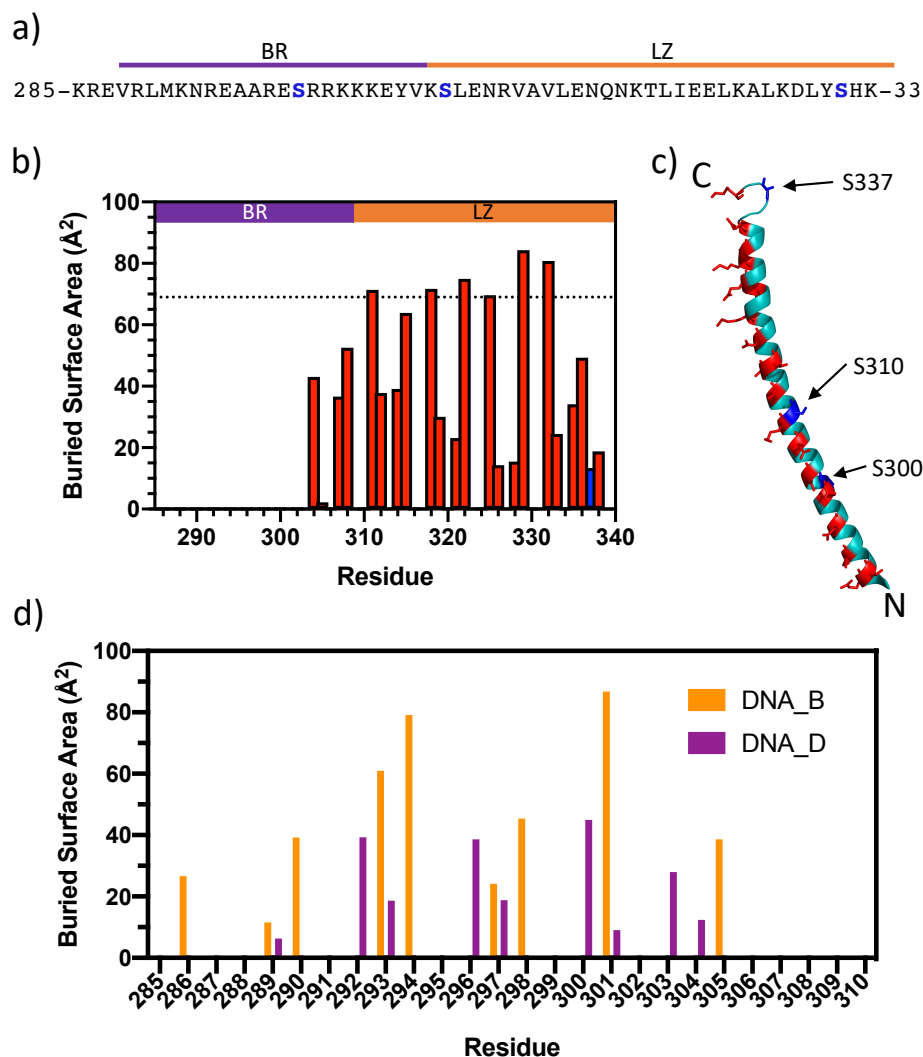


Figure 3.5: Designing the location of site directed cysteine mutagenesis in CREB bZIP pWT. a) Protein sequence of the CREB bZIP pWT construct. Residues in blue are the three serine residues which are Cys in the native sequence. b) Buried surface area was determined using contact map analysis on the CREB bZIP pWT crystal structure (PDB:1DH3). Data shown is of chain C residues in contact with chain A (both protein chains). Dashed line denotes a cut-off at $69\text{\AA}$ for the leucines involved in the zipper interaction. c) Location of all the residues that are not in a contact interface, are shown as red sticks. The residues that are in a contact interface are cyan. The three Ser residues positions are shown as blue sticks. d) Buried surface area of CREB bZIP chain A in complex with both of the DNA chains from PDB:1DH3.

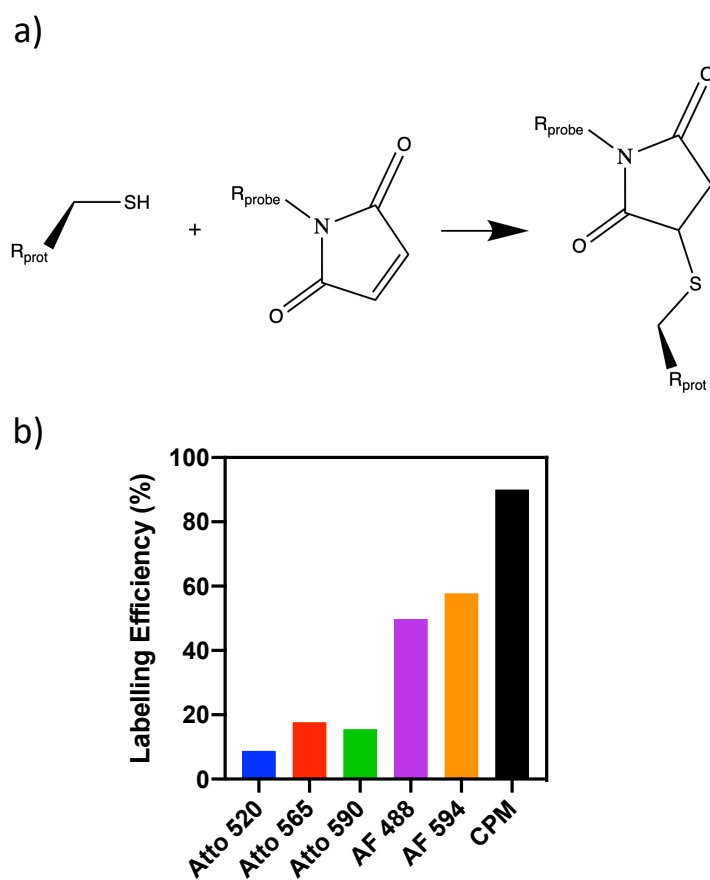


Figure 3.6: Multiple labelling conditions were trialled for CREB bZIP S337C with various fluorophores, with some success. a) Chemical reaction diagram for maleimide reaction with a free thiol group found in cysteine residues. b) Labelling efficiency was determined by UV/vis absorbance scans for each of the dyes.

Multiple labelling strategies were attempted, including reproducing the commercial kit conditions, which are summarised in Appendix Table S1 and Appendix Figure S1. In short, the use of various fluorophore dyes (AlexaFluor and Atto) with different buffer conditions (chaotropic agents, ionic strength, DNA and temperature), resulted in varied but unsatisfactory labelling efficiencies. A separate control protein, c-Myb, also underwent labelling as a control due to the difficulties faced. Notably, the use of urea at a high labelling temperature resulted in "rooster" type tails of a protein band during SDS-PAGE analysis and extra peaks in the ESI-MS trace with a difference of 42.5 Da (or multiples of), indicating that carbamylation is occurring. The labelling efficiency was at a maximum of ~60% (Figure 3.6b).

However, a combination of low pH (pH 6.7), high temperature (50°C) and 7-Diethylamino-3-(4'-Maleimidylphenyl)-4-Methylcoumarin (CPM) resulted in the efficient labelling of CREB bZIP S337C, with a efficiency >90% (Figure 3.7 and 3.6b). The use of CPM was not an ideal choice for a fluorophore when compared to modern AlexaFluor dyes, due to its poor photostability. However, CPM does not contain a charged fluorescent moiety, which may provide an advantage for any potential downstream DNA binding assays. Therefore, the opportunity then existed for the characterisation of the labelled protein construct, to validate the kinetics results.

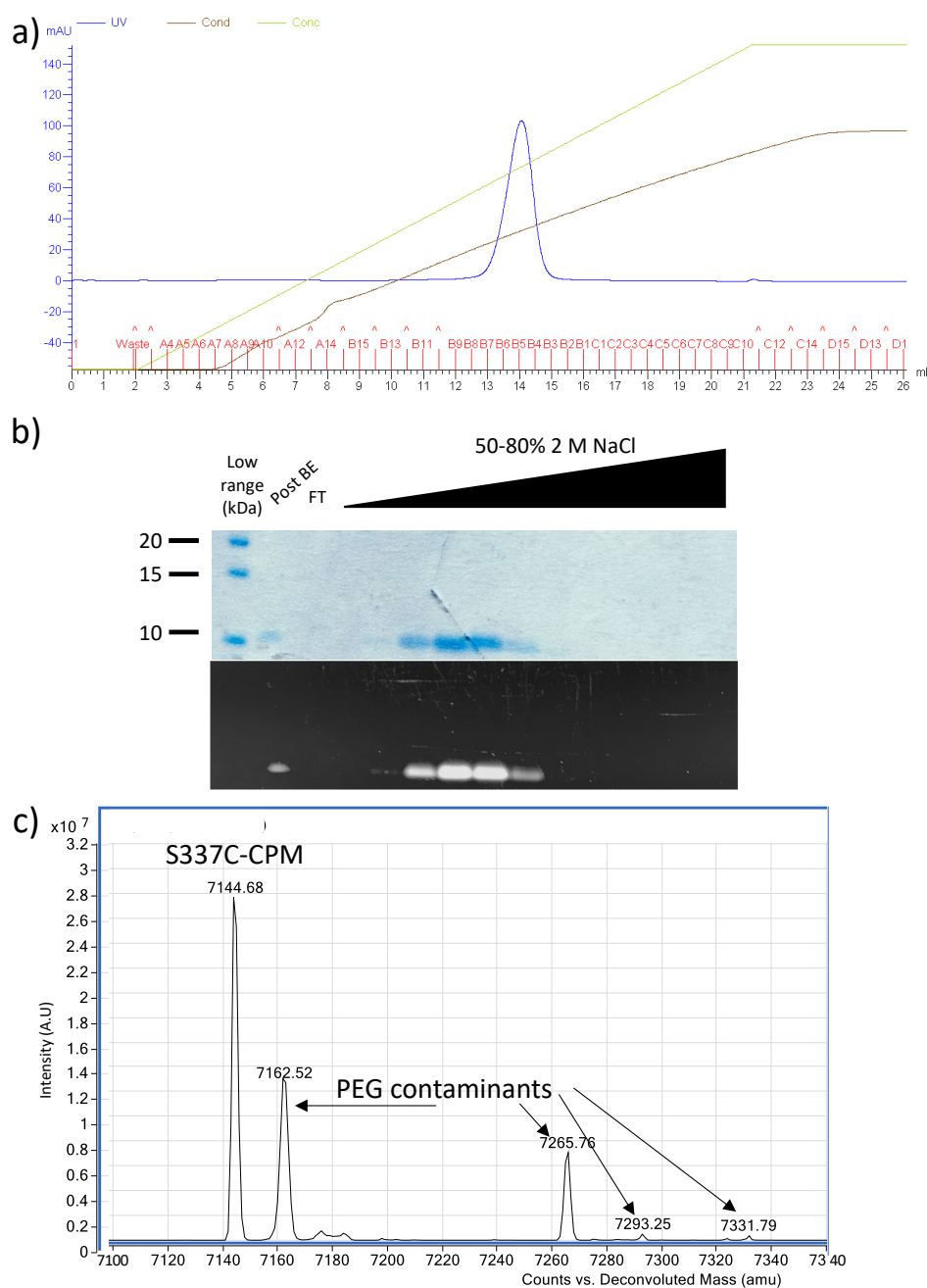


Figure 3.7: Successful labelling of CREB bZIP S337C with CPM. a) Cation exchange chromatogram showing absorbance (blue line) at 280 nm of CREB bZIP S337C-CPM protein along a gradient of 2 M NaCl (green line). Brown line indicates conductivity of the eluent. b) SDS-PAGE of the fractions where high absorbance was seen in a), stained by coomassie on the top and a UV plate reader image on the bottom. c) ESI-MS of the purified construct in b). Polyethyleneglycol (PEG contaminants) contaminants due to mass spectrometry procedure by provider.

Equilibrium analysis of labelled CREB bZIP

Following the successful labelling strategy of CREB bZIP at position C337 in section 3.2.2, the ability to validate the homodimer stability estimate from the

folding kinetics in section 3.2.1 was available. Serial dilutions were performed on the labelled protein into biophysical buffer with the fluorescence anisotropy used to report on the state of the complex, shown in Figure 3.8.

The estimate of the K_d from the labelled protein serial dilution strategy was 4 ± 1 nM, which is similar to the kinetics estimate in section 3.2.1 of 9 ± 3 nM within the same order of magnitude. The stability of the homodimer did not vary significantly across a wide range of ionic strengths (Figure 3.8c-d). Due to the low protein/fluorophore concentrations used in this approach, the fluorescence intensity that was observable was perturbed by the background fluorescence of buffer alone, seen in Figure 3.8b. It was clear at this stage that the S337C mutation and addition of CPM did not perturb the homodimer to a great extent. To provide further evidence for this, a competition assay was used with CPM labelled S337C construct and the pWT CREB bZIP, where an IC_{50} of 1.09 ± 0.04 nM and a K_i of 0.193 nM was obtained when using the Cheng-Prusoff equation¹⁷³. This then confirms that the homodimer complex is very stable, with a high affinity for the homodimer interaction.

Various other equilibrium approaches were attempted, and are summarised in Figure 3.9 and in the Appendix (pages 238 to 241), which also showed a consistently high affinity for the homodimer interaction. From the CD thermal melts, the thermodynamics of the interaction could start to be understood (Appendix Figure S6) using the extracted T_m . The y-intercept from Figure S6 shows the $\Delta S = 500 \pm 100$ J K⁻¹ and the slope suggests $\Delta H = 100 \pm 40$ kJ mol⁻¹. From the Gibbs Free Energy relationship at 25°C, $\Delta G = -49$ kJ mol⁻¹, indicating that it was favourable for formation of the homodimer.

NMR experiments also helped to validate the stability of the homodimer complex was independent of divalent cation presence or protein concentration ≥ 20 μ M (Figures 3.10 and 3.11 respectively), due to no chemical shift perturbation seen in the ¹⁵N-HSQC. Assignments for these constructs were not performed at this time and data was analysed qualitatively. Details on the assignments can be seen in the following section (section 3.2.3).

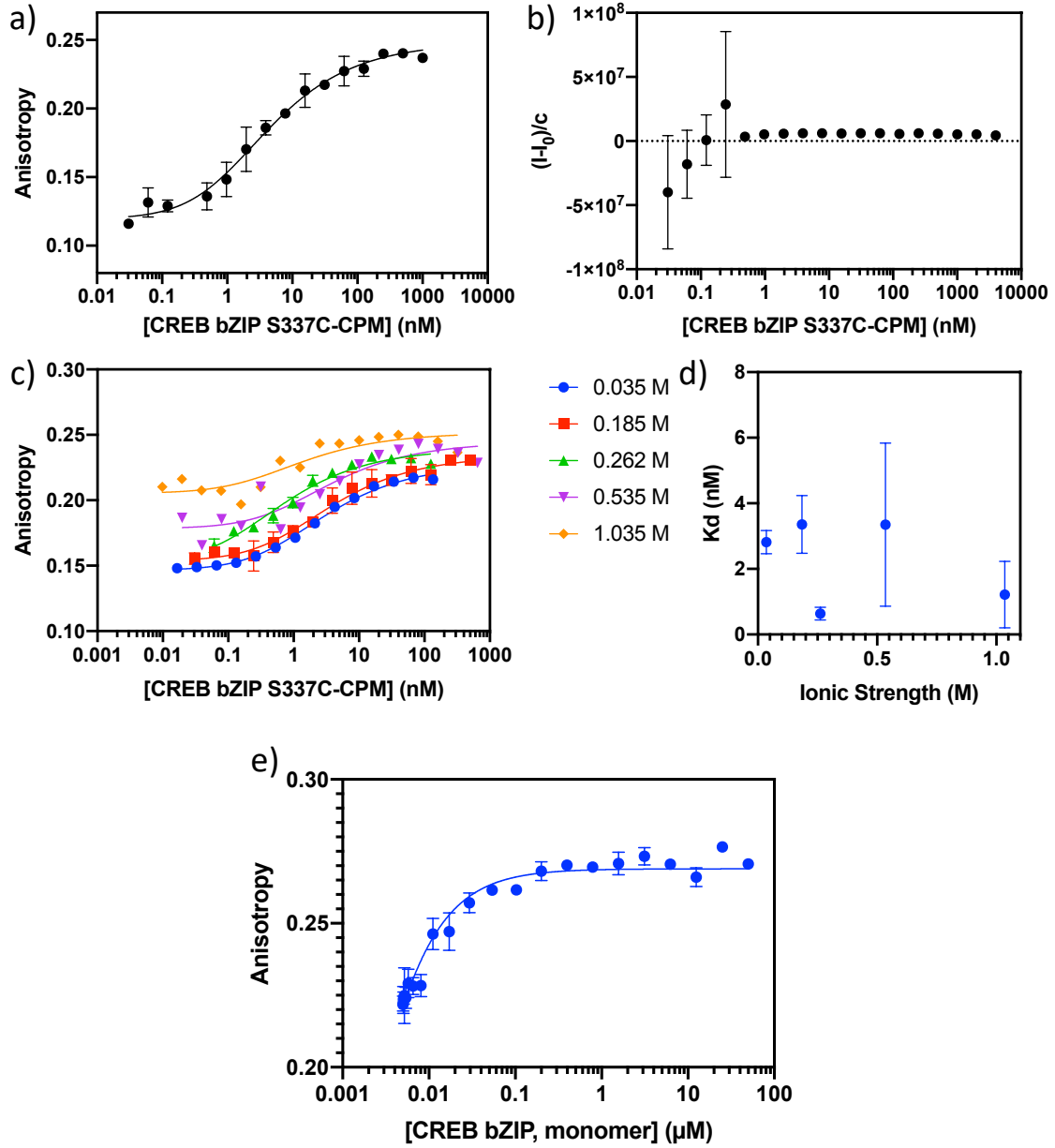


Figure 3.8: Homodimer K_d of labelled CREB bZIP shows a tight interaction. a) Fluorescence anisotropy of CPM labelled CREB bZIP S337C serially diluted into buffer. b) Normalised intensity of CREB bZIP S337C-CPM. Errors are SEM, $n=3$. c) and d) Ionic strength dependence of CREB bZIP S337C-CPM serial dilutions. Serial dilutions were fitted with Equation 2.10. e) Labelled CREB bZIP was competed with pWT and data was fit with Equation 2.11.

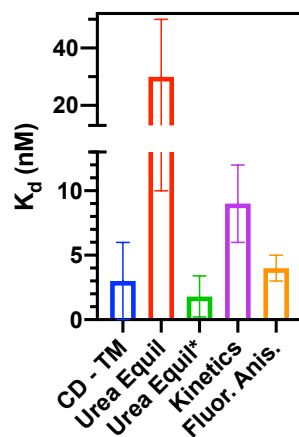


Figure 3.9: Homodimer K_d estimates for CREB bZIP pWT from multiple experimental techniques. Errors are SE. * contains m -value determined experimentally from kinetic re/unfolding experiments fixed in fitting procedure. Fluor. Anis. = fluorescence anisotropy from the labelled protein serial dilution with buffer.

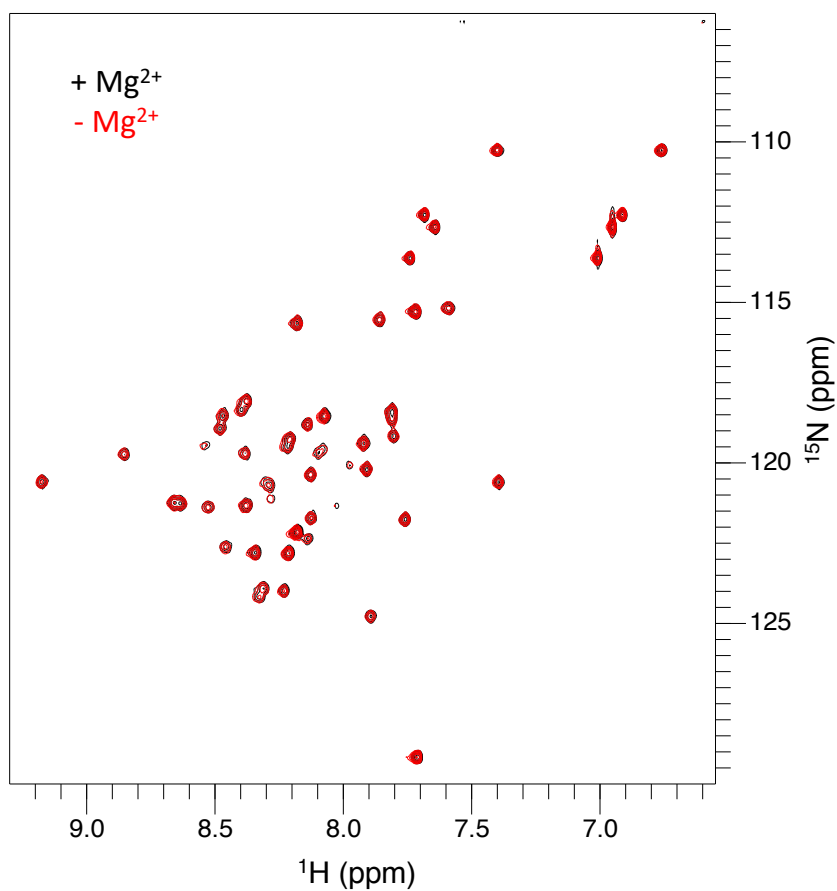


Figure 3.10: Mg^{2+} dependence on homodimer conformation using NMR. ^{15}N -HSQC of CREB bZIP (164 μM assuming dimer) collected at pH 7.4 at 25°C, in the presence or absence of 10 mM $MgCl_2$.

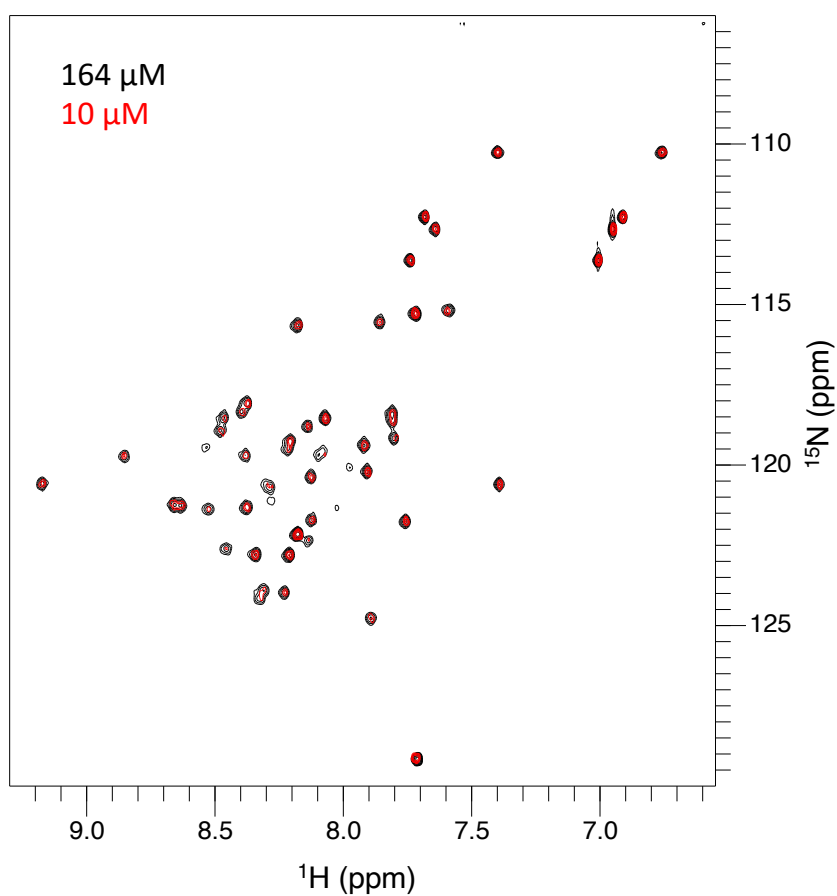


Figure 3.11: Protein concentration dependence on homodimer stability. ^{15}N -HSQC of CREB bZIP collected at pH 7.4 at 25°C, at high and low [CREB bZIP], assuming dimer concentration.

3.2.3 Structural characterisation of the dimer

Partially folded bZIP domain observed with CD

Following the characterisation of the dimeric state of CREB bZIP pWT in Sections 3.2.1 and 3.2.2, the next step was to characterise the structure of the dimeric state, as this appears to be the native state for the DNA binding domain of CREB. CD spectra showed a Mean Residual Ellipticity (MRE) at 222 nm of $-17100 \theta \text{ cm}^{-1} \text{ M}^{-1}$ (Figure 3.12a). Assuming a theoretical MRE of $-30100 \theta \text{ cm}^{-1} \text{ M}^{-1}$ for a 100% helical protein at 25°C ¹⁷⁴, then CREB bZIP pWT contained 56% helical content, which corresponds well to the number of residues in the leucine zipper structure (62%). However, it was unknown at this stage what structure the basic region populated in the apo-state, so NMR was chosen to characterise the protein structure.

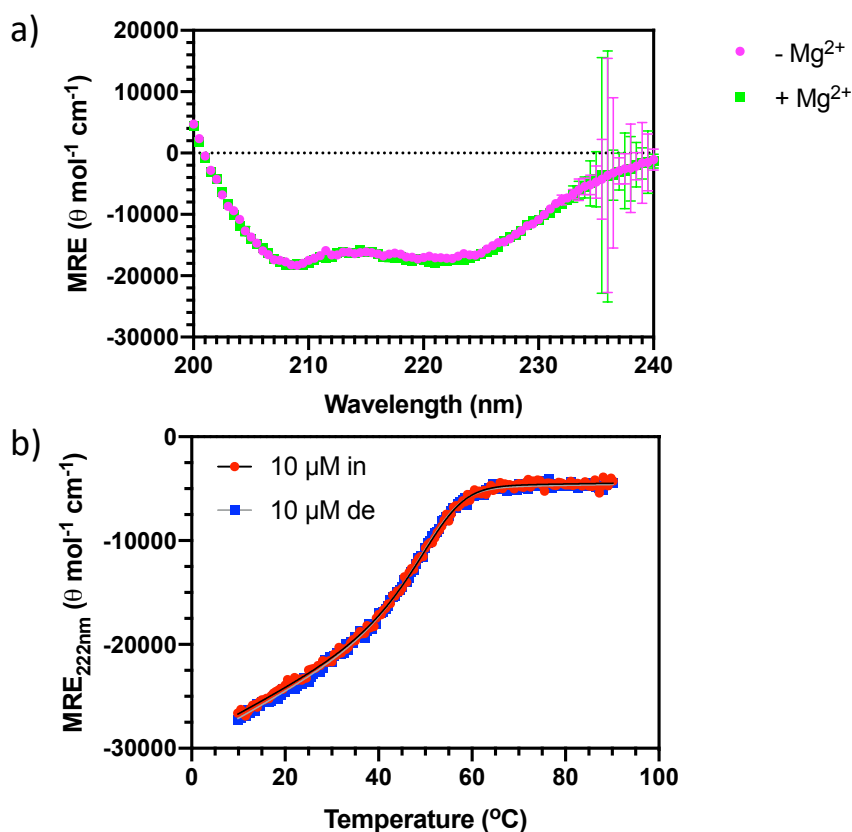


Figure 3.12: Ensemble structural description of CREB bZIP pWT by CD. a) CD spectra of CREB bZIP pWT (285-339) at 25°C . Errors are from independent repeats and are SEM. b) Thermal melt of CREB bZIP by CD. Unfolding is reversible as shown by the agreement between MRE upon heating (in) and cooling (de).

Optimisation of the expression of isotopically labelled constructs

The structure of the CREB bZIP domain was shown using CD to be partially helical, which was thought to correspond to the leucine zipper region. To confirm this, a residue specific approach was required, with NMR ideally placed to provide this analysis of the dimeric complex. Assignments of CREB bZIP domain have previously not been obtained, therefore isotopically labelled protein was required to perform the sequential assignment strategies that are commonly employed on proteins and summarised by C. Redfield¹⁶³.

Initially, the expression of the CREB bZIP pWT construct needed to be optimised for isotope labelling, as a significant amount of leaky expression was seen when using *E. coli* strain C41(DE3, Figure 3.13a), seen with the band at 20 kDa which corresponds to the His₆-GB1-CREB bZIP. Although the construct is estimated to be 16 kDa, the high disorder content coupled with the high net positive charge results in a SDS-PAGE migration that is higher than expected. BL21 Gold (DE3) pLysS strain was then used, as this was designed to express toxic proteins by limiting the level of leaky expression.

The pLysS plasmid encodes T7 lysozyme which inhibits the T7 RNA polymerase. The level of leaky expression was reduced to a significant extent, so this strain was used for isotope expression. The level of expressed construct after induction did not vary significantly between 3-6 hours, therefore 4 hours was used as the preferred incubation time. A test to see whether moving from small (100 mL) to large (1 L) cultures resulted in minimal differences in the levels of the expressed construct after the incubation period (Figure 3.13b), with 0.2% glucose being added in the starter and expressed culture reducing the level of leaky expression even further. Therefore an isotope expression test could be pursued. An initial trial where a confluent starter culture was used to inoculate 1x M9 media with ¹⁴N-NH₄Cl and ¹²C-glucose resulted in no growth of bacteria, therefore the isotope enrichment strategy developed by Marley *et al*¹⁵⁶ was attempted.

Isotope enrichment uses rich media grown bacteria and concentrates them at a desired density in minimal media containing the isotopes, therefore being highly

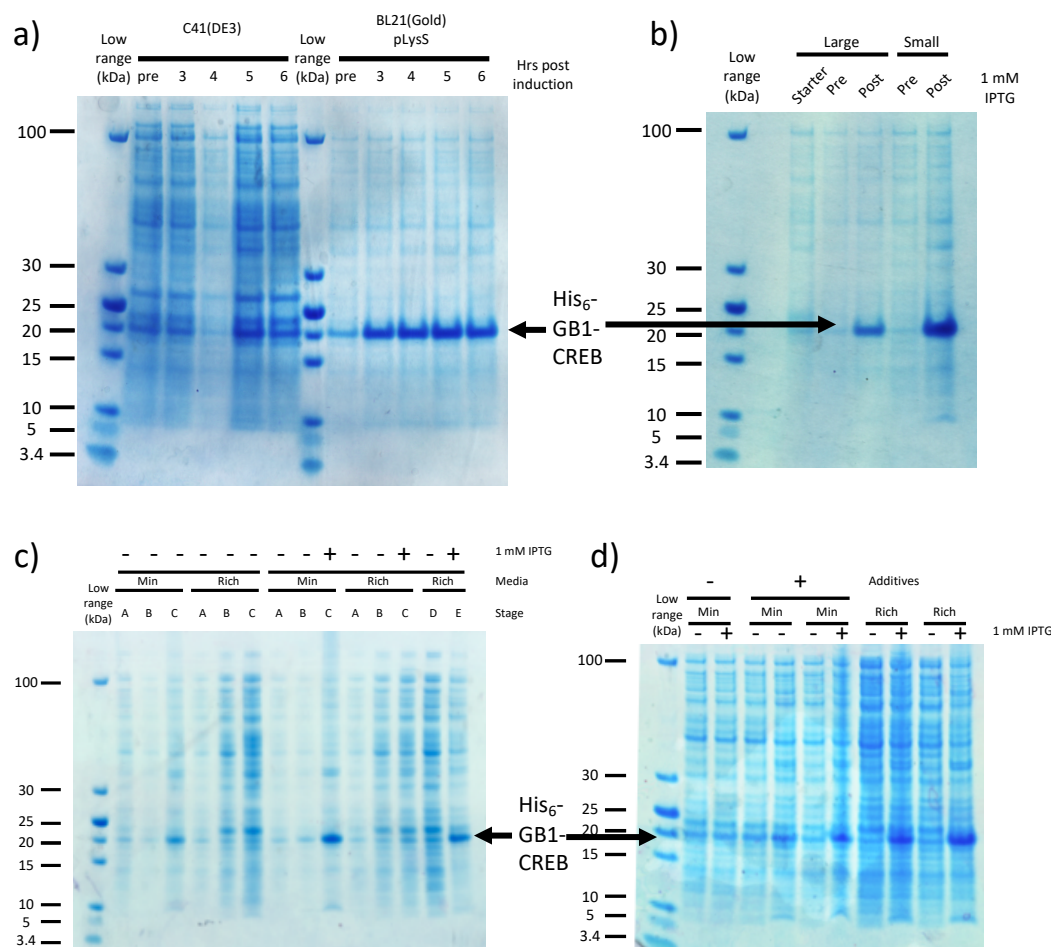


Figure 3.13: Leaky expression of expressed CREB bZIP construct required optimisation to increase yields for NMR. a) Two *E. coli* strains were transformed with the pRSETA plasmid containing His₆-GB1-CREB and expressed in rich media on a small scale. Samples were taken before and at several points following induction. b) Expression in BL21(Gold)pLysS to examine the change from small to large scale expression, coupled with the inclusion of 0.2% glucose in the media. c) Isotope labelling was trialled with ¹⁴N-NH₄Cl and ¹²C-glucose, following the isotope enrichment protocol¹⁵⁶. Rich media with 2xTY, min(imal) media was 1x M9 salts. Stage A was cells at OD₆₀₀ 0.6, B was cells resuspended in minimal media before induction, C was post induction of cells in B. D and E were pre and post IPTG incubation for a regular expression respectively. d) Similar to c), but containing further additives to increase expression yield. a-d) All samples analysed by non-reducing SDS-PAGE.

efficient and removes the bacterial grow up stage in minimal media (Figure 3.13c). The estimated level of expression, from the size of the band at 20 kDa, was comparable between the minimal media with IPTG to the regular expression in rich media (stage D and E). Inclusion of additives (FeCl₂, CaCl₂ and MgSO₄ (1 mM)) improved the yield of isotope enrichment over when no additives were present,

hence these were included in the final protocol. The yield of ^{13}C and ^{15}N labelled CREB bZIP pWT was 2.2 mg/L, and provided protein which could be used for assignment and characterisation by NMR.

Initial assignment of CREB bZIP pWT at neutral pH

In the previous section, CREB bZIP pWT was able to be isotopically labelled with ^{13}C and/or ^{15}N to a suitable yield. With this isotopically labelled protein, experiments were performed in NMR buffer (see Table 2.1) at pH 7.4. Initial assessment showed that only 33 peaks were visible under these conditions, via ^{15}N -HSQC (Figure 3.14). Peaks were assigned using ^{15}N -edited NOESY-HSQC and TOCSY-HSQC, as the leucine zipper peaks were expected to be in a helical conformation. The assignment followed the well established protocol of linking consecutive residues by the $\text{NH}_{i\pm 1}$ in the NOESY (summarised in the chapter written by C. Redfield¹⁶³, shown in Figure 3.15).

The peaks that were assigned corresponded to the residues belonging to the leucine zipper region, by virtue of the sidechain proton resonances from the ^{15}N -edited TOCSY. This enabled me to find the directionality of the linked resonances as certain residues such as valine and isoleucine have more up-field shifted H_α resonances. With the leucine zipper residues assigned at pH 7.4, further optimisation was required of the experimental conditions in order to observe the residues corresponding to the basic region.

Examination of conditions for NMR analysis of CREB bZIP pWT

The experimental conditions required optimisation in order to observe peaks for all residues. Initially starting with NMR buffer (see Table 2.1) at pH 7.4, ^{15}N -HSQC's were collected at various pH's and the number of peaks was counted and compared to the number of peaks expected for the construct, which was 55 (excluding the N-terminal glycine, Figure 3.16).

At the highest pH analysed in this series of experiments, 37 peaks were observed, indicating that a significant number of peaks were not detected (Figure 3.16). pH is well known to determine the intrinsic rate of backbone amide proton exchange

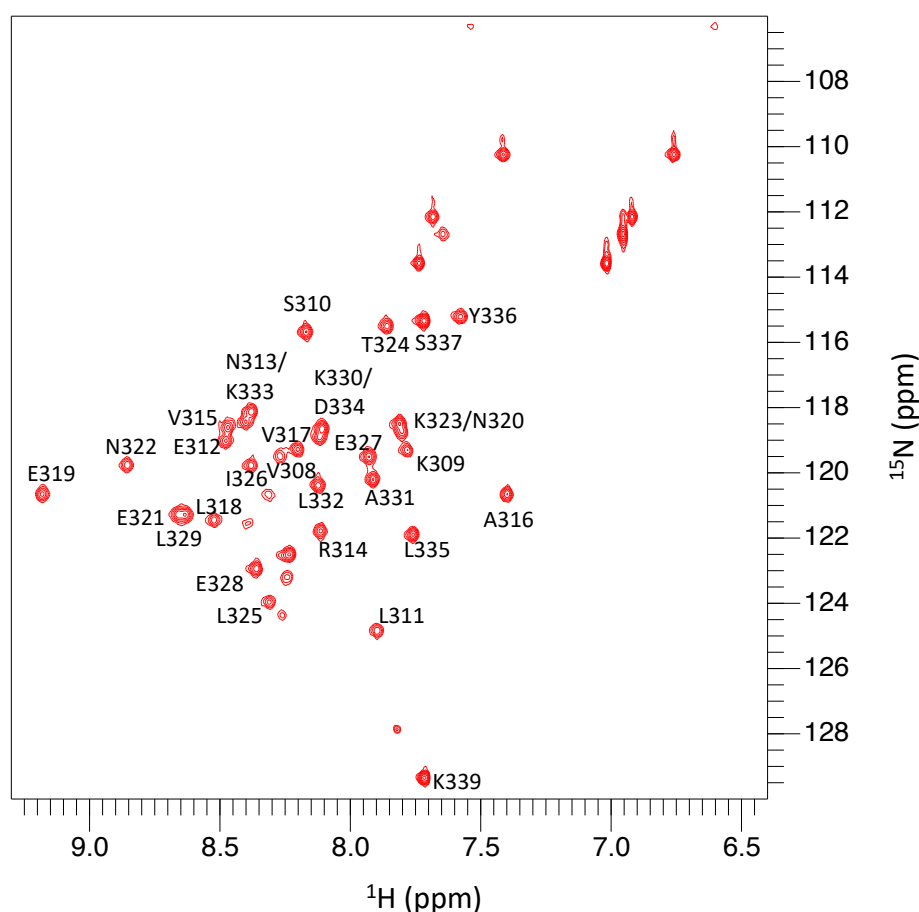


Figure 3.14: CREB bZIP pWT assignments at pH 7.4. Assigned ^{15}N -HSQC of CREB bZIP pWT collected at 25°C with NMR buffer at pH 7.4.

with solvent, and can be estimated using the SPHERE webserver¹⁷⁵. The backbone amide proton exchange rate could be reduced by a significant factor (~ 10 -fold) upon reduction of the pH to 6.5 (Figure 3.17a). Experimentally, the effect of the change in pH was validated by performing ^{15}N -HSQC's at those pH values, resulting in 54 peaks appearing (peaks count not including the side chain amide resonances, identified through the chemical shifts and their paired relationship).

Peak heights can be obtained from the assigned spectra (see Section 3.2.3 for assignments), which show an increase in the peak height at pH 6.5 for residues in the basic region (residues 285-306, Figure 3.17b). Only the residues located close to His338 are perturbed by the change in pH, due to the change in protonation state of the histidine side chain group ($\text{pK}_a \sim 6$, Figure 3.17c). Therefore, pH

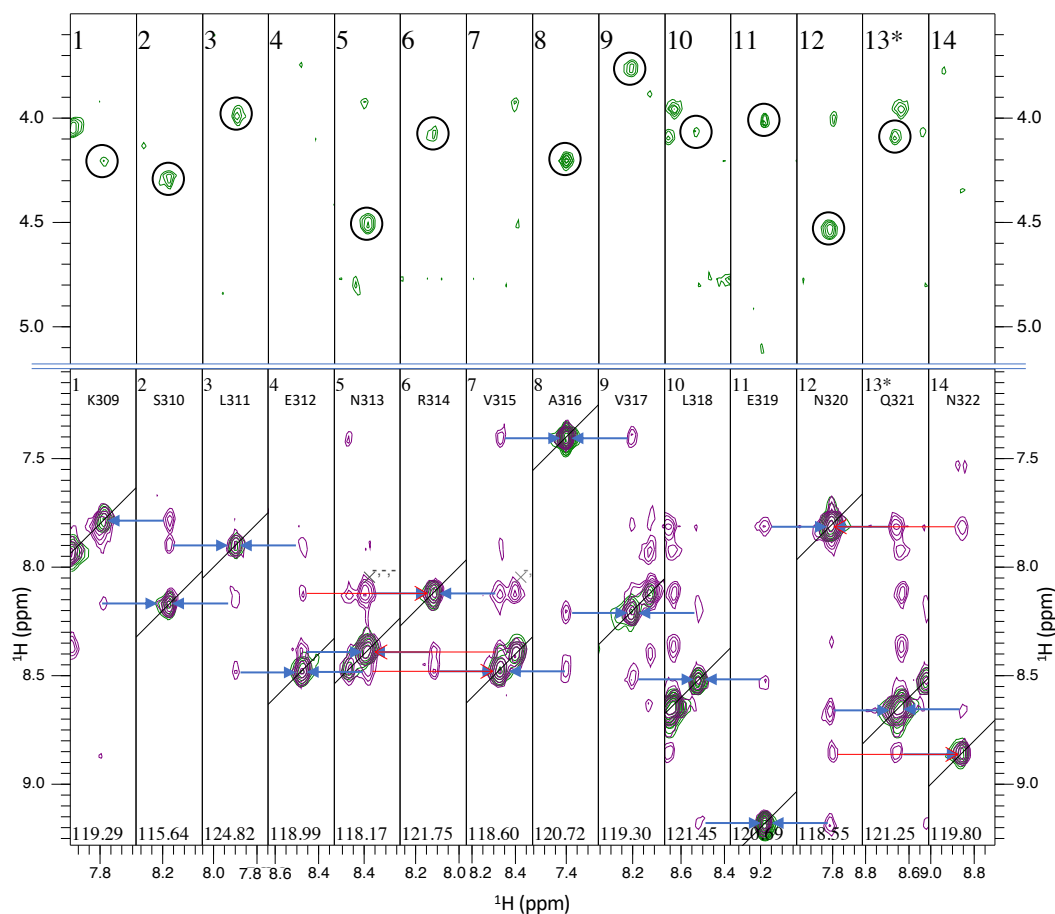


Figure 3.15: Assignment of peaks attributed to the leucine zipper was possible using ^{15}N -edited NOESY amide backbone resonance linking. Linked $\text{NH}_{i\pm 1}$ are shown using blue arrows, $\text{NH}_{i\pm 2}$ with red arrows. Residues 309-322 are shown as an example. Green = ^{15}N -edited TOCSY, purple = ^{15}N -edited NOESY. Circles indicate $\text{H}\alpha$ assigned for that residue using the TOCSY only.

6.5 was chosen as the condition under which the peaks that corresponded to the basic region had the greatest intensity, with the next step to assign the apo-state of the bZIP dimer complex.

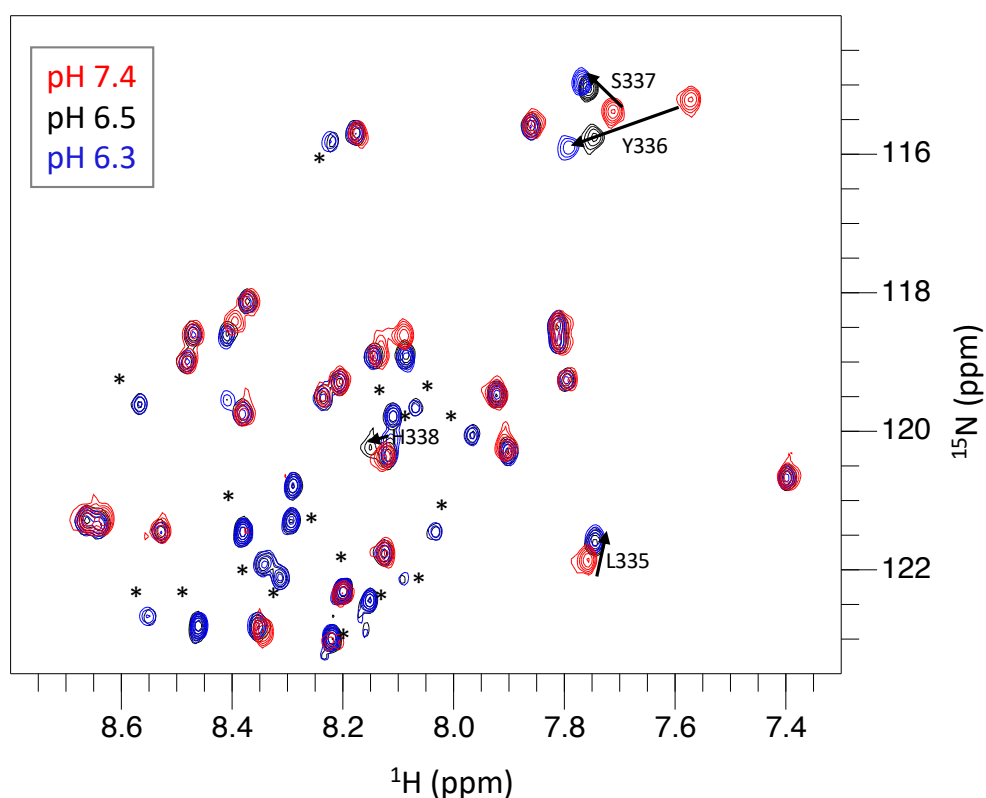


Figure 3.16: pH titration of ^{15}N -CREB bZIP was undertaken to visualise all backbone amide resonances. ^1H - ^{15}N -HSQC spectra were acquired at several pH values, all at 25°C . Arrows and labels indicate peaks that were sensitive to changes in pH. * indicates peaks that were not present at pH 7.4, which appeared in the lower pH conditions.

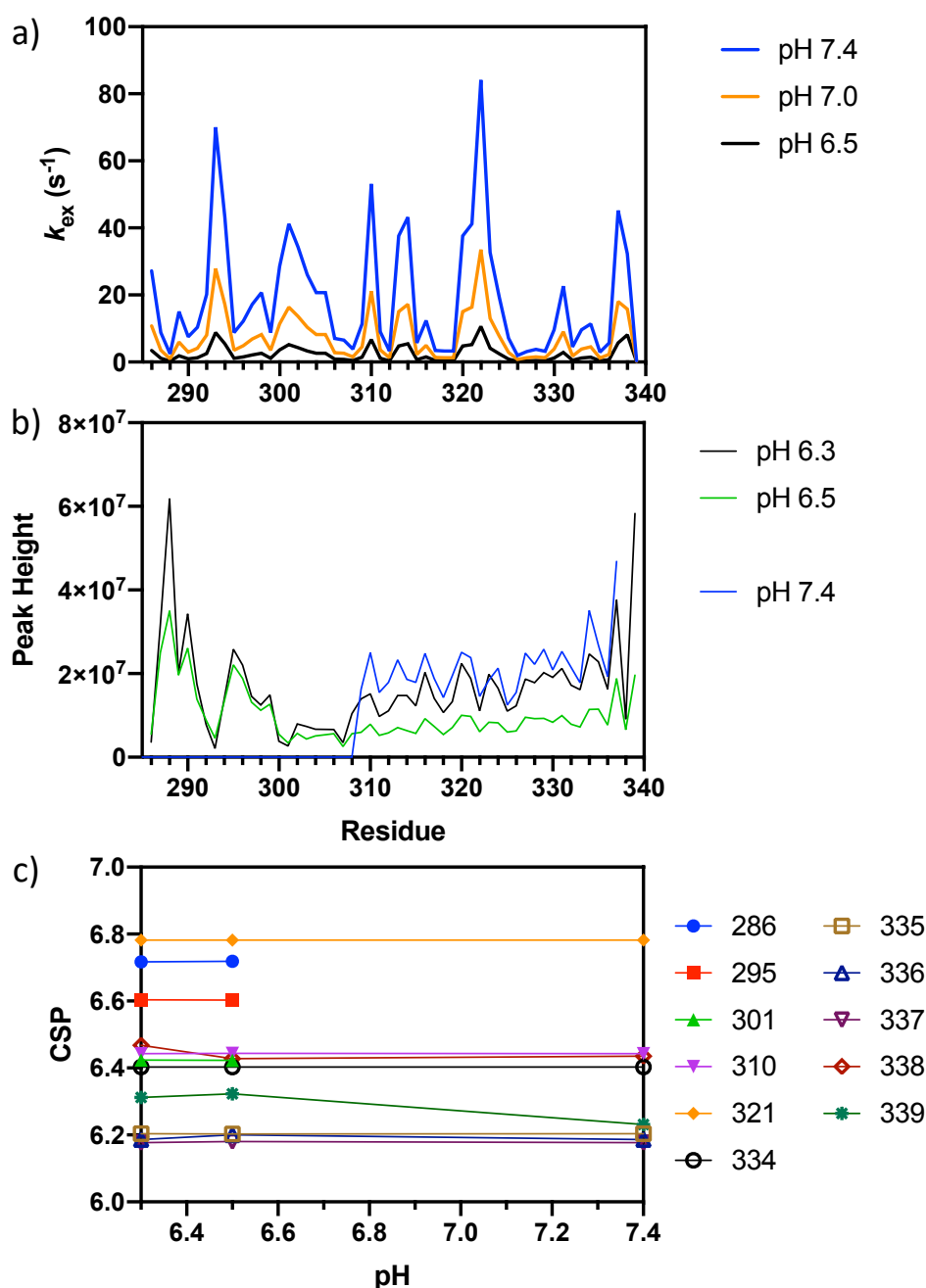


Figure 3.17: Basic region residues are resolved at low pH and leucine zipper residues are unaffected. a) Exchange rate of amide proton with solvent was estimated using the SPHERE webserver at several pH values to identify the highest pH at which the backbone proton exchange rates were low enough to observe peaks. b) Peak heights as determined from CCPNMR v2, using assignments at pH 6.5 initially to titrate back to other pH values, where available. c) Chemical shift perturbation calculated using the equation described by Williamson¹⁶².

Complete assignment of the apo state of CREB bZIP pWT

Following the assignment of CREB bZIP pWT under sub-optimal conditions and the subsequent optimisation of the conditions for analysis, the complete assignment of the basic region could be obtained. The change in pH did not alter the position of the majority of amide backbone resonances in the ^{15}N -HSQC (aside from those residues located near to H338, Figure 3.18) and these have been previously shown to be in a dynamic disordered region at the extreme C-terminus²¹. At this point, doubly labelled ^{15}N and ^{13}C CREB bZIP pWT was used, with triple resonance experiments to assign the peaks corresponding to the basic region. Triple resonance experiments were utilised at this stage as it was expected that the ^{15}N -edited NOESY may not work as successfully to link the $\text{H}^{\text{N}}\text{-H}^{\text{N}}$ with a disordered region of the protein.

For assigning triple resonance experiments, a minimum of five experiments were performed: HNCA, CBCA(CO)NH, HN(CA)CO, HNCO and HBHA(CO)NH. These experiments enabled me to link residues i , $i-1$ via the $\text{C}\alpha$ when using HNCA and CBCA(CO)NH, or via the CO when using the HNCO and HN(CA)CO. An example of sequential assignment using triple resonance experiments is shown in Figure 3.19.

The sequential assignment using the $\text{C}\alpha$ and CO also enabled the validation the assignments of the leucine zipper using the ^{15}N -edited experiments, unequivocally determining the orientation of the linked resonances. When using the CBCA(CO)NH, this yielded the $\text{C}\beta$ and $\text{C}\alpha$ chemical shifts for the preceding residue, therefore another atom's chemical shift was available for determining the residue's identity and enough information to estimate the secondary structure propensity. Further triple resonance data to obtain the $\text{H}\alpha$ and $\text{H}\beta$ chemical shifts were more difficult to obtain, particularly for the leucine zipper region at pH 6.5 (Figure 3.20).

The use of HCCH-TOCSY did not aid in improvement of side chain proton assignments due to the lack of chemical shift dispersion in the ^{13}C -HSQC $\text{H}\alpha$ and $\text{H}\beta$ region (Figure 3.21). The idea here was to use the known $\text{C}\alpha$ and $\text{C}\beta$ chemical shifts from the CBCA(CO)NH and HNCA experiments, to aid in the assignment of the corresponding protons. However, the majority of the backbone and $\text{H}\beta/\text{C}\beta$ atoms was able to be assigned, with the proportion of atoms assigned

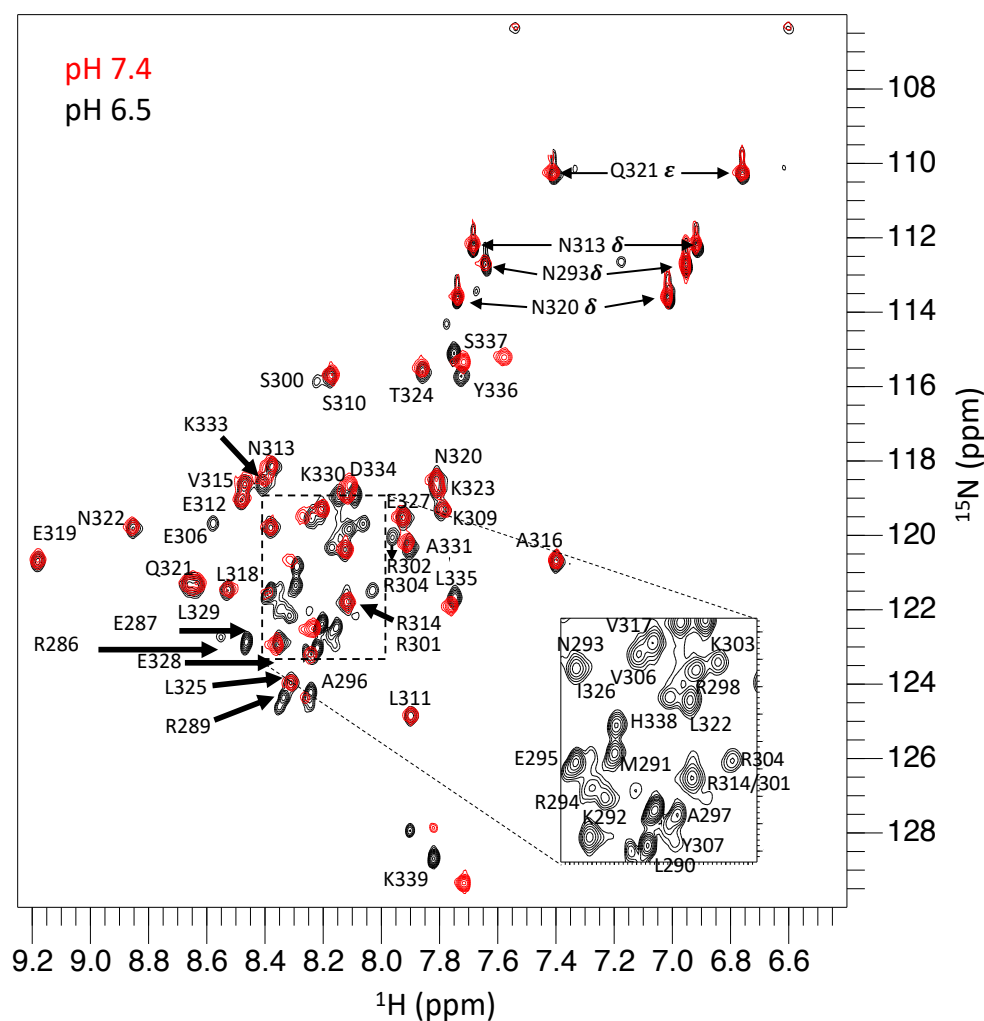


Figure 3.18: CREB bZIP pWT was assigned at two pH values to determine the effect of the pH change. Assignments shown in the ^{15}N -HSQC's are for pH 6.5, with the pH 7.4 residues that are different shown in Figure 3.17.

$\geq 90\%$, but extended side chain information was unfortunately limited (Table 3.4 and Figure S6). Therefore, with the high quality of assignments at pH 6.5, this information could be used to determine the secondary structure characteristics of the apo complex for dimeric CREB bZIP.

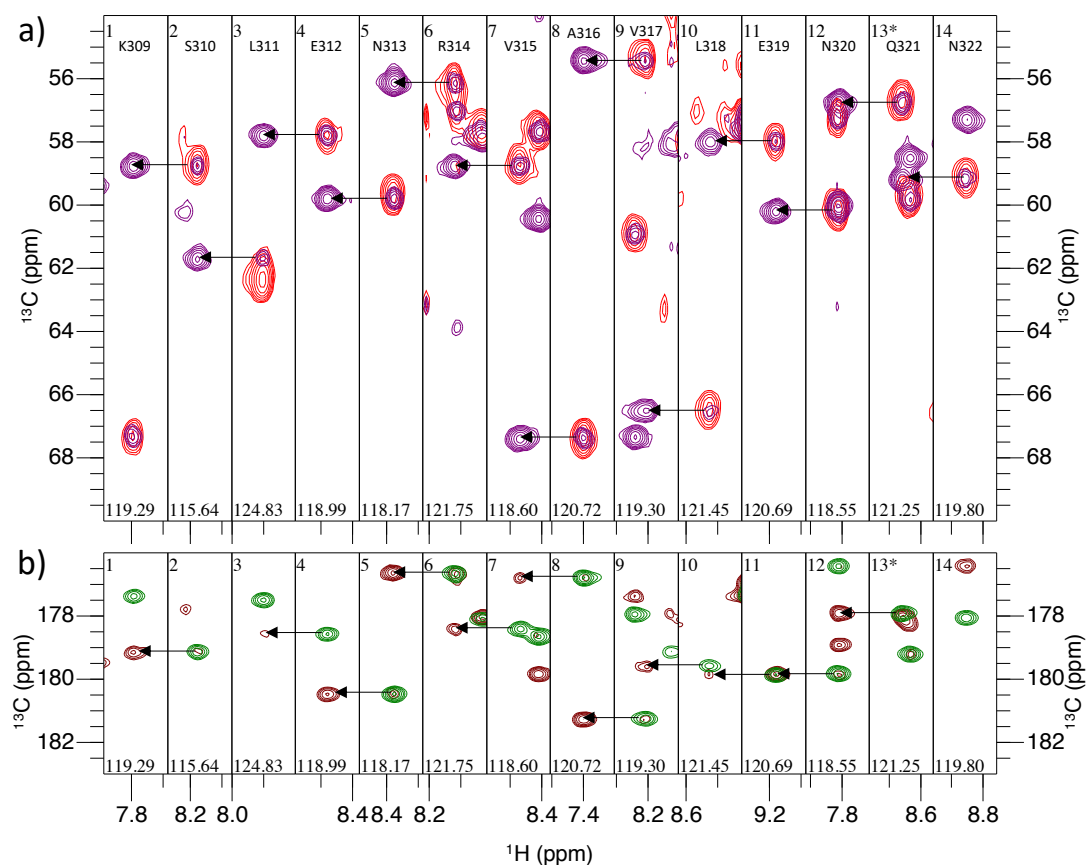


Figure 3.19: Various triple resonance experiments were used to assign the basic region residues in CREB bZIP constructs. a) Representative CBCA(CO)NH and HNCA spectra overlays from the same region in Figure 3.15. Purple = HNCA, Red = CBCA(CO)NH. b) CO region of spectrum shown in a), indicating the use of the HNCO and HN(CA)CO for sequential assignment. Green = HNCO, Brown = HN(CA)CO. Arrows indicate direction of connection from i to $i-1$.

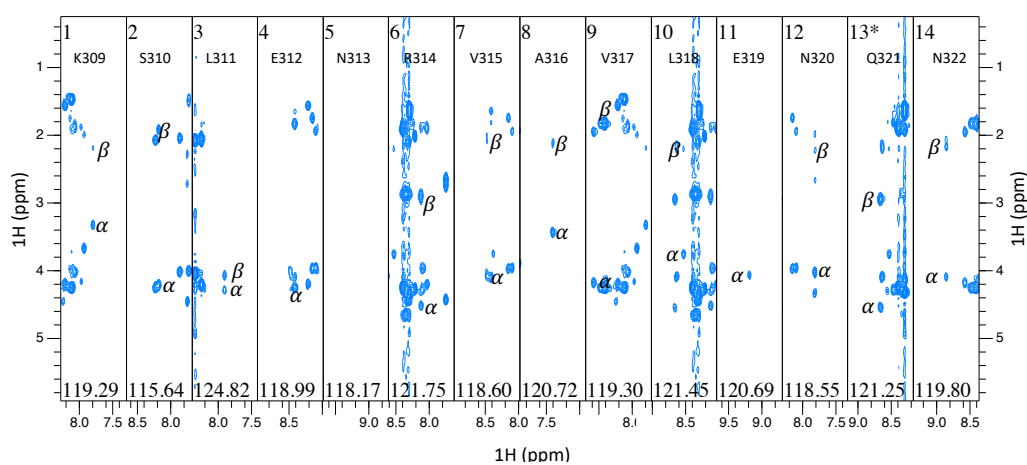


Figure 3.20: HBHA(CO)NH was performed on CREB bZIP pWT to obtain limited side chain proton assignments. α and β assignments are for $i-1$ to the assigned residue at the top of each strip.

Table 3.4: Assignment report of CREB bZIP pWT (285-339) under two different conditions. The construct feature 56 residues, including the N-terminal residue. Obtained from ccpnmr v2.4.

Category	% Assigned	
	pH 7.4	pH 6.5
All Residues	53.6	98.2
Type C	51.8	98.2
Type CA	53.6	98.2
Type CB	49.1	98.2
Type CD	0.0	18.2
Type CE	0.0	0.0
Type CG	0.0	17.0
Type CZ	0.0	0.0
Type H	52.7	98.2
Type HA	52.6	96.5
Type HB	45.0	92.0
Type HD	0.0	21.7
Type HE	0.0	6.1
Type HG	0.0	11.0
Type N	51.8	96.4
Type ND	0.0	60.0
Type NE	0.0	11.1
Type NH	0.0	0.0
Type NZ	0.0	0.0

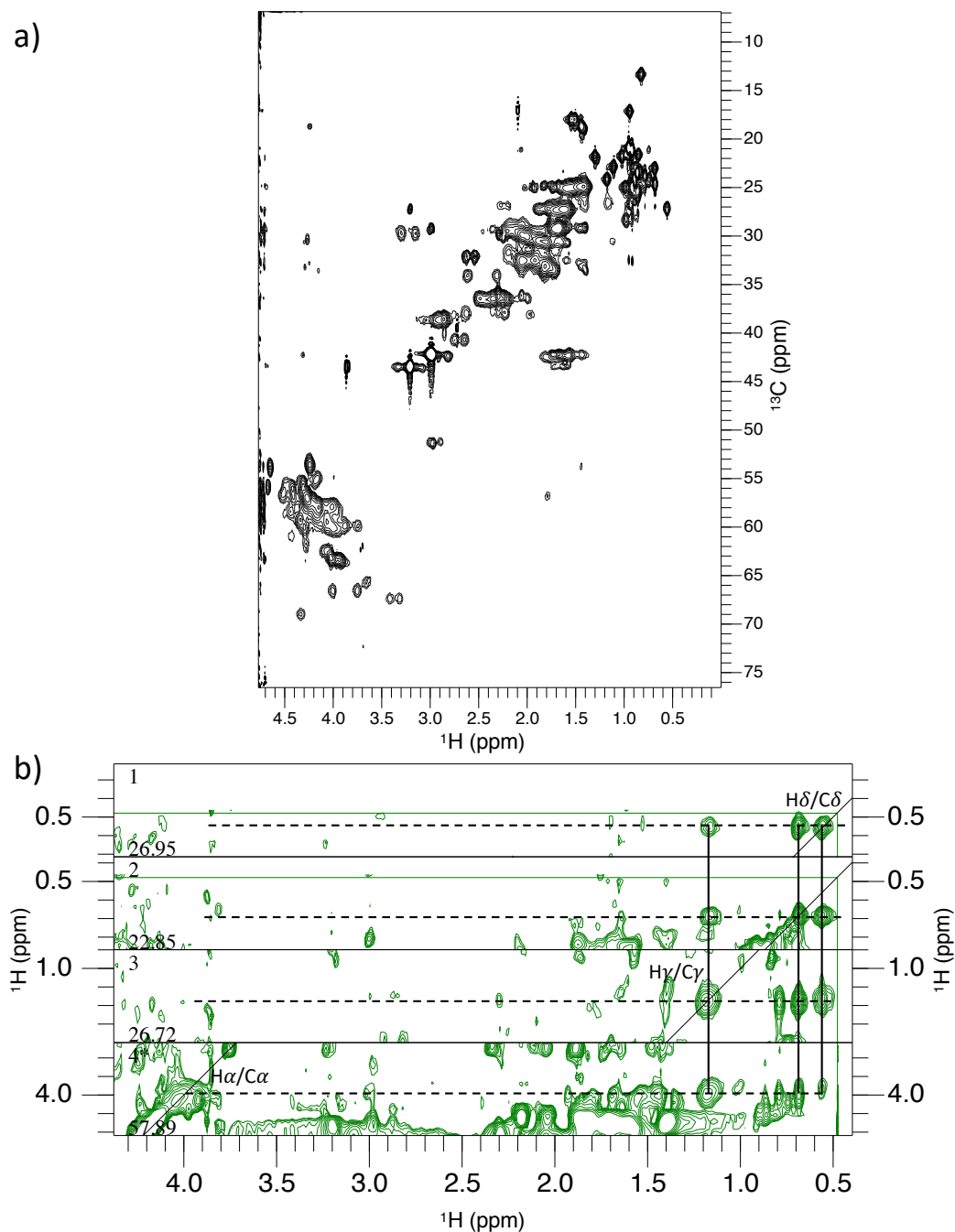


Figure 3.21: Lack of chemical shift dispersion in the $C\alpha$ and $C\beta$ regions of ^{13}C -HSQC limited to scope of sidechain assignment with HCCCH-TOCSY. a) ^{13}C -HSQC of $^{15}\text{N}/^{13}\text{C}$ CREB bZIP pWT. b) Spectra obtained from HCCCH-TOCSY of CREB bZIP pWT. Lines indicate linked connections within the side chain of L311. Collected at pH 6.5, 298 K.

Estimation of secondary structure from NMR

Following on from the successful assignment of CREB bZIP in the previous section, the secondary structure and dynamics of the protein was able to be estimated. As seen in the CD data earlier in this chapter (Figure 3.12), there was evidence to suggest the CREB bZIP domain had a significant helical character in the apo-state of the dimeric complex. Whilst this is most likely due to the leucine zipper, CD cannot yield the residue specific information to categorically show this.

The secondary structure propensity of protein sequences can be estimated using bioinformatic software. Agadir was chosen to estimate the level of helical content of protein constructs¹⁷⁶, with one drawback being that it does not estimate oligomer formation. Several bZIPs were compared against one another using CREB as the model (Figure 3.22). This identified two regions that were populated by increased helical content, one towards the N-terminus and the other towards the C-terminus, in most of the proteins that were studied. The same proteins, with the exception of Cys3, underwent structural alignment using published crystal structures^{21,23,85,177} (Figure 3.22b). Helical wheel analysis of CREB bZIP was performed with HeliQuest¹⁷⁸, which showed the amphipathic α -helix of the leucine zipper region facing one side of the wheel (Figure 3.22c) and a random distribution of residues in the basic region itself. Several methods exist for interpreting chemical shift data in terms of secondary structure, the simplest being the difference between the experimentally observed and expected random coil chemical shifts. When this was done for the $C\alpha$ and $C\beta$ atoms, there was evidence of the peaks associated with the leucine zipper being in a helical conformation from residue 308 onwards, having a large positive difference when looking at the $C\alpha$ shifts, with a negative difference for the $C\beta$ shifts (Figure 3.23). Residues with a positive $\Delta C\beta$ (311, 318 and 325) are all leucine and are involved in the leucine zipper interactions located in the dimer interface (Figure 3.23b and c).

The idea of using chemical shifts to estimate the secondary structure of proteins has been extended by the development of various programs. These use pre-compiled libraries of random coil chemical shifts and expect shifts in helices and β -sheets,

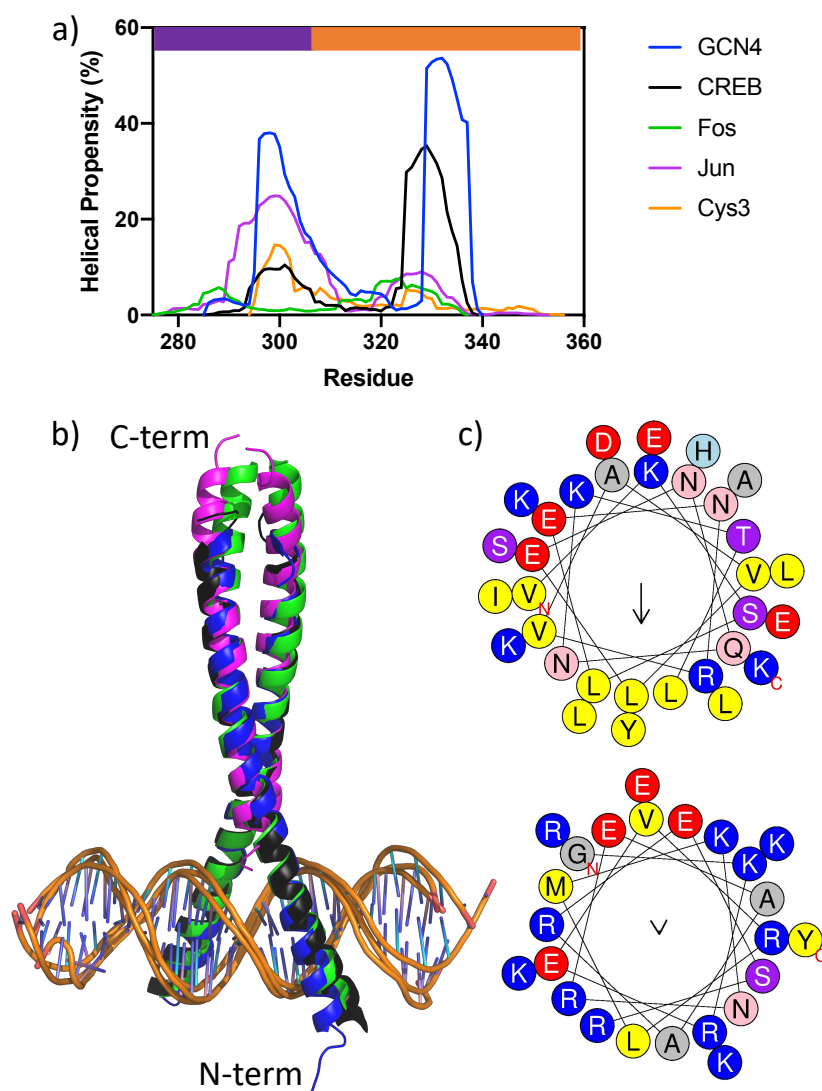


Figure 3.22: Helical propensity prediction of several bZIP protein sequence shows two helical regions. a) Multiple bZIP containing proteins were analysed using Agadir¹⁷⁶. Sequences were aligned based on structures shown in b) and numbered based on the CREB1a sequence (Uniprot: p16220). Purple = basic region, orange = leucine zipper. b) Structural alignment of various bZIP domains from different proteins based on the protein chains, performed in PyMOL. Colour of protein chains matches that in a). PDB codes for each construct are CREB = 1DH3, GCN4 = 1YSA, Fos = 1FOS, Jun = 1JUN*. * indicates NMR assignment. c) Helical wheel analysis of CREB bZIP domain (285-339). Top) Residues 307-339. Bottom) Residues 285-306. Analysis performed with HeliQuest¹⁷⁸. Red = negatively charged residues, blue = positively charged residues, yellow = hydrophobic residues.

to estimate the secondary structure propensity of a protein of interest. One of these programs, called TALOS+, was developed by the Bax group and is designed to estimate protein backbone torsion angles, but in doing so can provide a

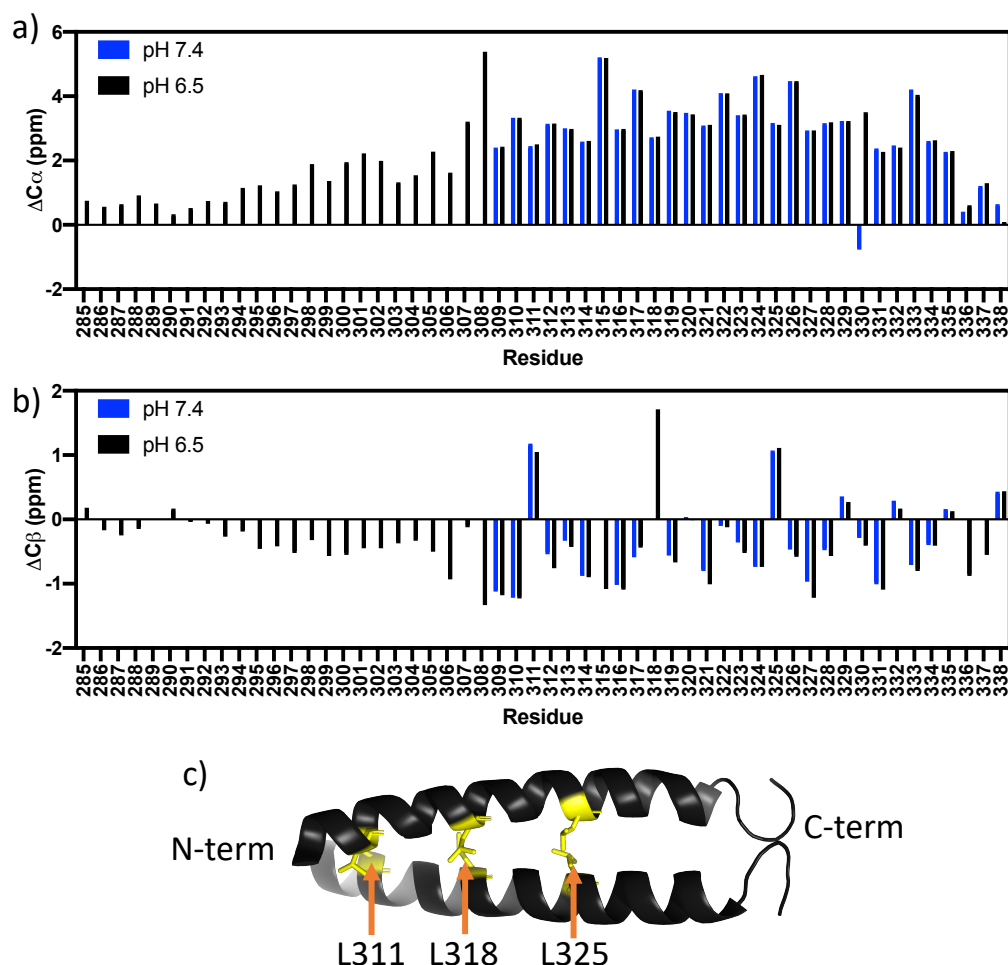


Figure 3.23: Secondary chemical shifts of CREB bZIP from assignments at two different pH values. a) $C\alpha$ secondary chemical shifts. b) $C\beta$ secondary chemical shifts. Random coil chemical shifts were obtained from ncIDP¹⁷⁹. c) Leucine residues that have the positive $\Delta C\beta$ are indicated in yellow on the crystal structure of CREB bZIP (PDB: 1DH3).

estimate of the probability that a residue populates a particular secondary structure (Figure 3.24a)¹⁶⁴. An alternate method developed by the Forman-Kay laboratory was designed specifically for IDPs, called Secondary Structure Propensity¹⁶⁵ (SS, Figure 3.24b).

Comparing the three methods overall, the leucine zipper is fully formed, and helical content decreases steadily from around residues 306-310 towards the N-terminus. This agrees well with the BLAST bioinformatic result of the CREB1a sequence identifying the leucine zipper starting at residue 307, with a decrease in α -helical content extending towards the N-terminus. TALOS+ reveals that for

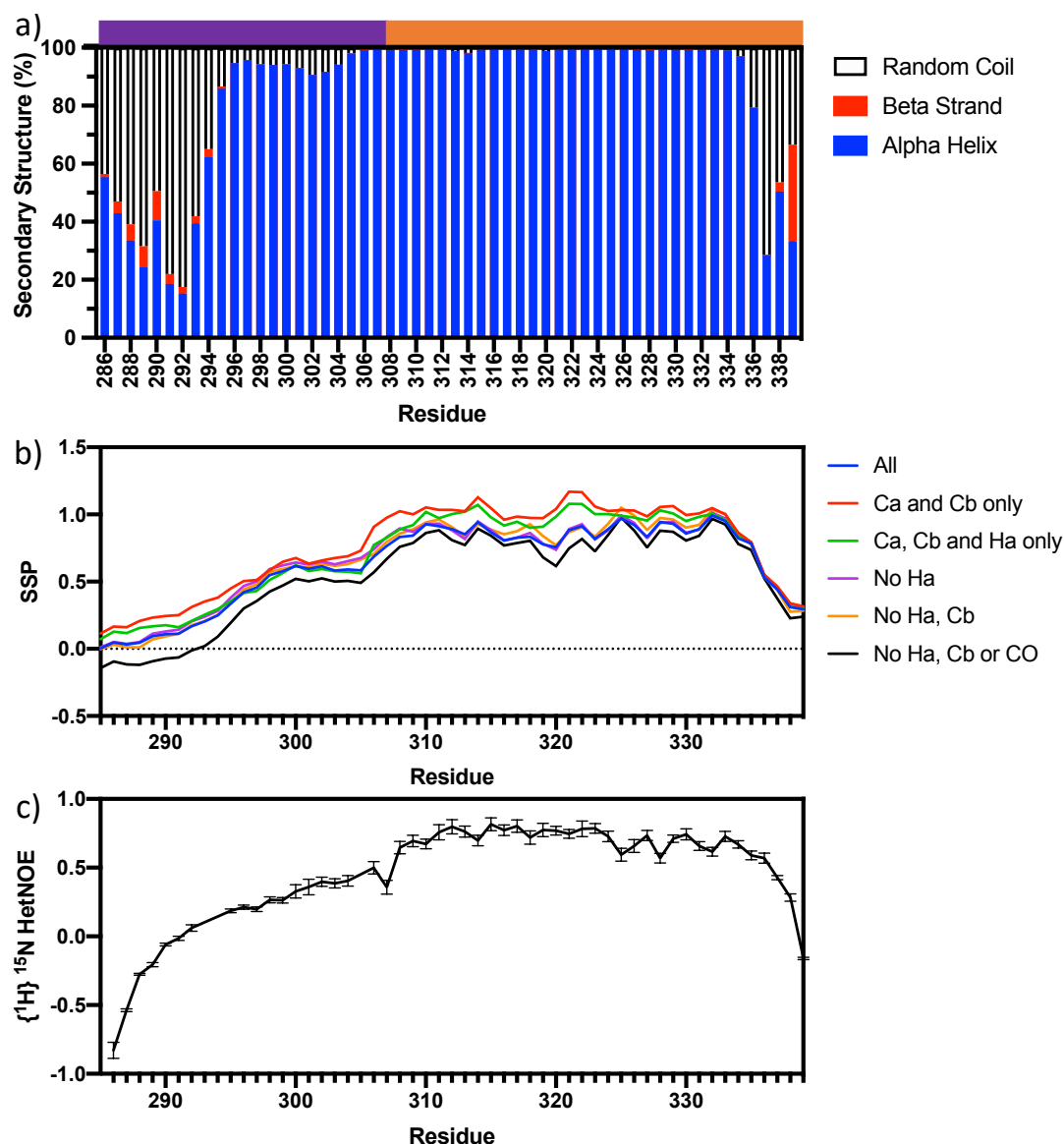


Figure 3.24: Information on CREB bZIP secondary structure propensity can be determined from assigned chemical shifts. a) Secondary structure populations can be determined on a residue specific level using TALOS¹⁶⁴. b) Secondary Structure Propensity (SSP) was calculated from CREB bZIP chemical shifts¹⁶⁵, with variations of atoms used to determine the effect on the result. A score of 1 = α -helix, 0 = random coil, -1 = β -sheet. c) $\{^1\text{H}\}$ - ^{15}N Heteronuclear NOE on CREB bZIP, collected at 600 MHz, provides information about backbone dynamics. Errors are SD. a-c were analysed using data collected at 25°C. Purple = basic region, orange = leucine zipper.

residues ~ 295 -308, there is a higher prediction of helix when compared to SSP. This could be due to TALOS overpredicting secondary structure, with a large positive change in $\text{C}\alpha$ being interpreted as α -helix. Interestingly, the SSP results seem dependent on the type of atoms and combinations thereof used as chemical shift

inputs, with a maximum difference identified as ~ 0.5 for residue Q321. The same may be observed for TALOS+, but not directly investigated in this thesis. The basic region residues do not appear to be as affected by the removal or change in combinations of atoms, however using a limited amount of information without $H\alpha$, $C\beta$ or CO does give the lowest overall SSP value per residue. Due to difficulties encountered later in this thesis with obtaining $H\alpha$ chemical shifts, $C\alpha$ and $C\beta$ this particular combination is used for the remainder of the thesis.

Dynamic information can also be obtained by performing a $\{^1H\}$ - ^{15}N Heteronuclear NOE experiment, which yields information on the flexibility of a particular residue, on the ps/ns timescale, within the overall tumbling of a protein construct/complex (Figure 3.24c). This proved insightful by agreeing with the SSP analysis, particularly for the basic region residues being more dynamic than those in the leucine zipper. This indicates that the basic region, in the absence of DNA, is in a dynamic state as helical propagation from the leucine zipper tails off around residue 300-306.

Secondary Structure comparison between GCN4 and CREB

Following on from the calculation of the secondary structure of CREB bZIP pWT, this data was used to compare previously assigned bZIP containing constructs. GCN4 was the only one with assignments publicly available⁹⁴. Analysing the chemical shifts using the SSP program revealed an almost identical periodicity in the helical propensity of the leucine zipper region (Figure 3.25), coming from the use of 1H N and N^H shifts being included in the analysis. The periodic minima corresponded to the residues identified in the contact map analysis for CREB bZIP as being not in a structural interface between the protein chains (Figure 3.5b). As for the residues corresponding to the basic region, GCN4's contains a higher helical propensity than CREB.

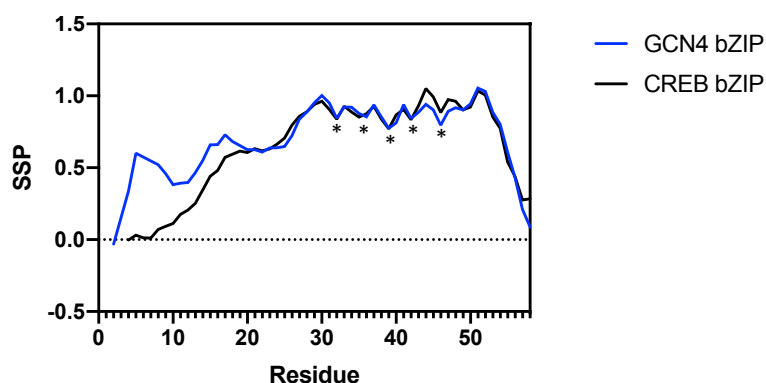


Figure 3.25: Comparison of secondary structure propensity between two bZIP domains. GCN4 bZIP (residues 226-281) data was obtained from Bracken *et al*⁹⁴. Chemical shifts used in analysis were ^1HN , ^{15}N , $^{13}\text{C}\alpha$ and ^{13}CO for both constructs.

N-terminal extension of CREB bZIP alters the helical propensity in the basic region

Throughout this Chapter, the experiments have focused on characterising the CREB bZIP pWT construct spanning residues 285-339, of the CREB1a protein sequence. This was the construct used previously for the X-ray structure of the complex with CRE DNA²¹. A comparison between the full length sequence of CREB and that of the bZIP domain residues 285-339 showed a significant difference in helical propensity (Figure 3.26).

The helical propensity could be increased by addition of several residues at the N-terminus of the Schumacher *et al* length of construct, where the predicted helicity of the basic region matches that of the full length protein when looking at 275-339. The initial work was performed by Dr M. Kuravskiy, and the 277-339 construct was chosen at this time and this particular construct was therefore used. The longer bZIP construct will be defined as CREB bZIP (277-339), with the shorter being CREB bZIP (285-339).

To validate if changing the length of CREB bZIP at the N-terminus affected the homodimer binding affinity, kinetic refolding experiments were performed on the 277-339 construct (Figure 3.27). The protocol established for the 285-339 construct in section 3.2.1 was used here and showed a similar homodimer K_d within error for both bZIP construct lengths, indicating the increase in length of the N-terminus

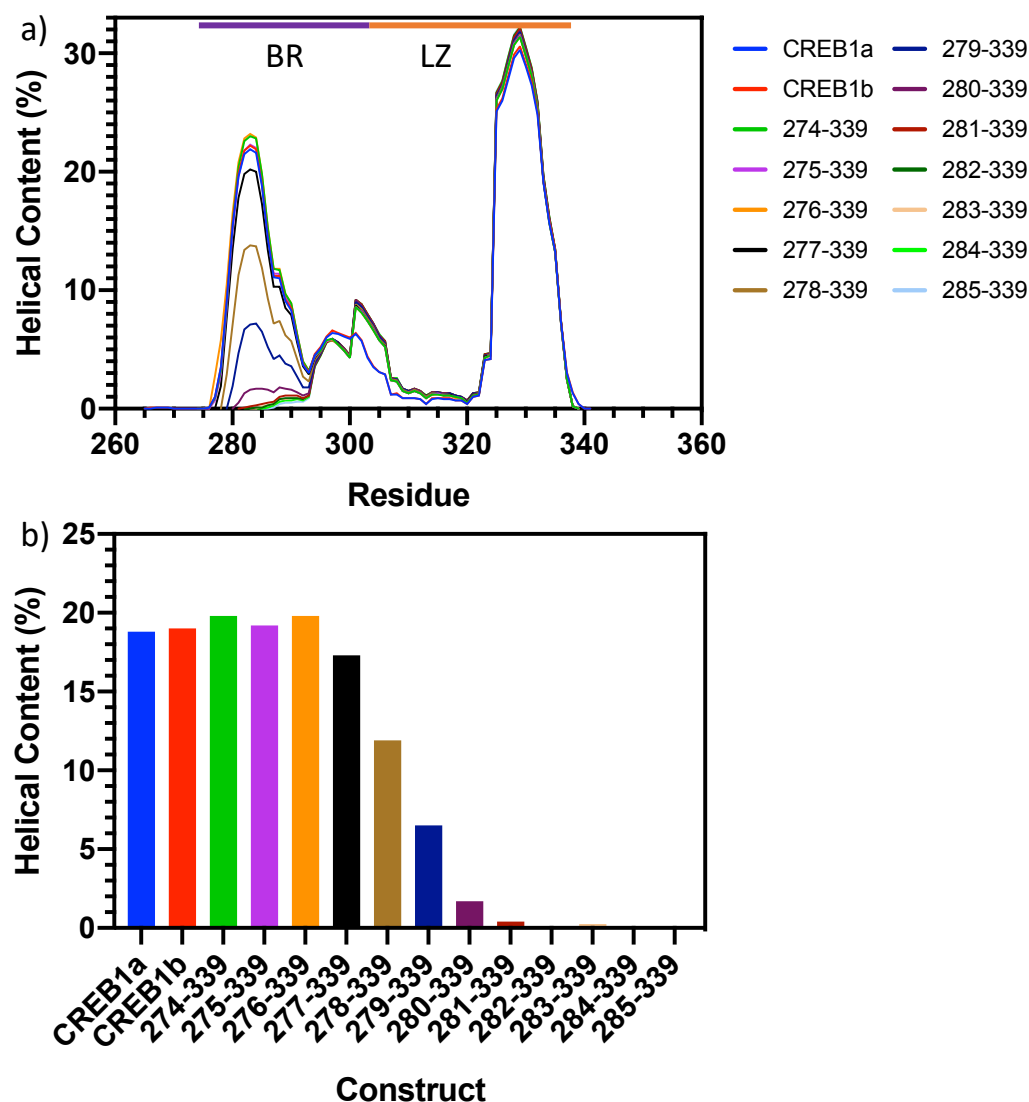


Figure 3.26: Extension of N-terminus increases predicted helical propensity in basic region. a) Agadir was used to calculate the helical propensity of the residues that corresponded to the bZIP domain. BR = basic region, LZ = leucine zipper. b) Comparison of the estimated helical content of residue 285 from all the construct shown in a). CREB1a/1b were calculated using the full sequence, with only the C-terminal residues shown. Native Cys remained in the sequence during the calculations.

has not affected the dimer stability, nor the rate of dimer formation. In order to determine the structural effect experimentally of the N-terminal extension on the CREB bZIP domain, NMR was repeated with the longer construct. As the helical content in the basic region may influence the association to the DNA, the residue specific information that NMR yielded for the 285-339 construct would be applicable here. The CREB bZIP 277-339 was assigned using the sequential

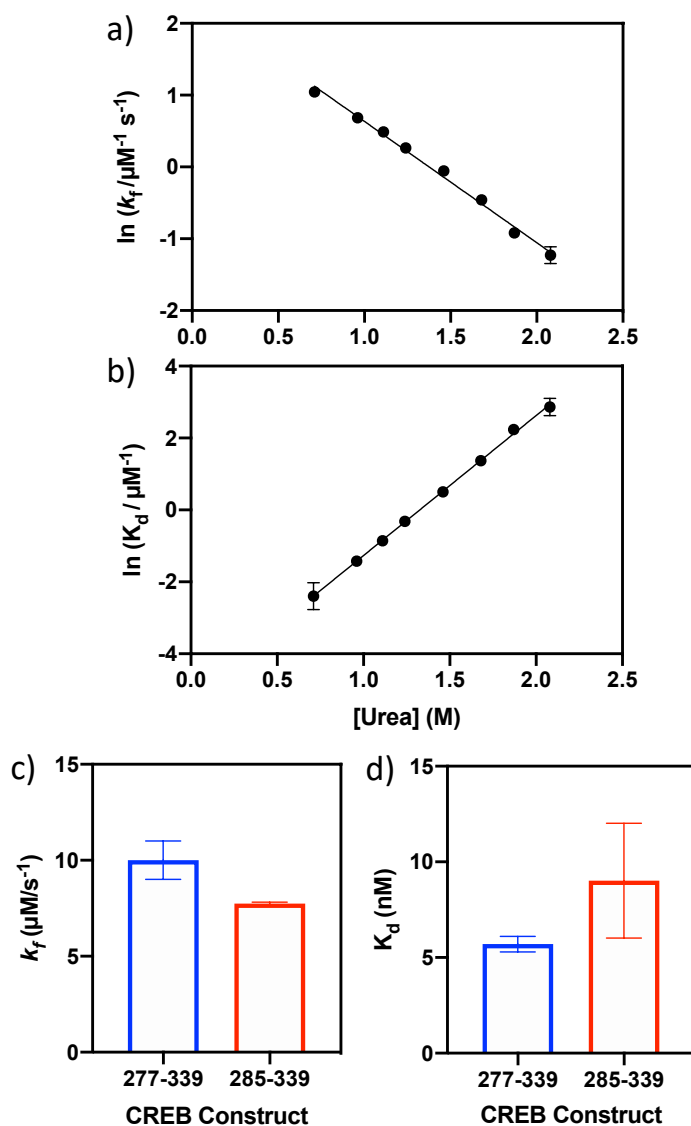


Figure 3.27: Dimer stability is not dependent on extension of basic region length. a) Natural logarithm of the folding rate (k_f) against urea concentration of unfolded CREB bZIP 277-339 at 9 μM . b) Similar to a) but with natural logarithm of K_d against urea concentration. c) Comparison of folding rate of both CREB bZIP construct lengths used. d) Similar to c), but with K_d . Errors are SE and are shown in all plots.

assignment strategy (see Figures 3.19, 3.20 and 3.28). From the assignment of CREB bZIP 277-339 pWT, the same characteristics were observed as for the 285-339 pWT construct, with the majority of the peaks assigned to the leucine zipper being dispersed in contrast to the basic region peaks. Due to the increased number of peaks in the 277-339 construct however, the region between 8-8.5 ppm in the ^1H dimension appears more crowded. Some assignments of the protein backbone

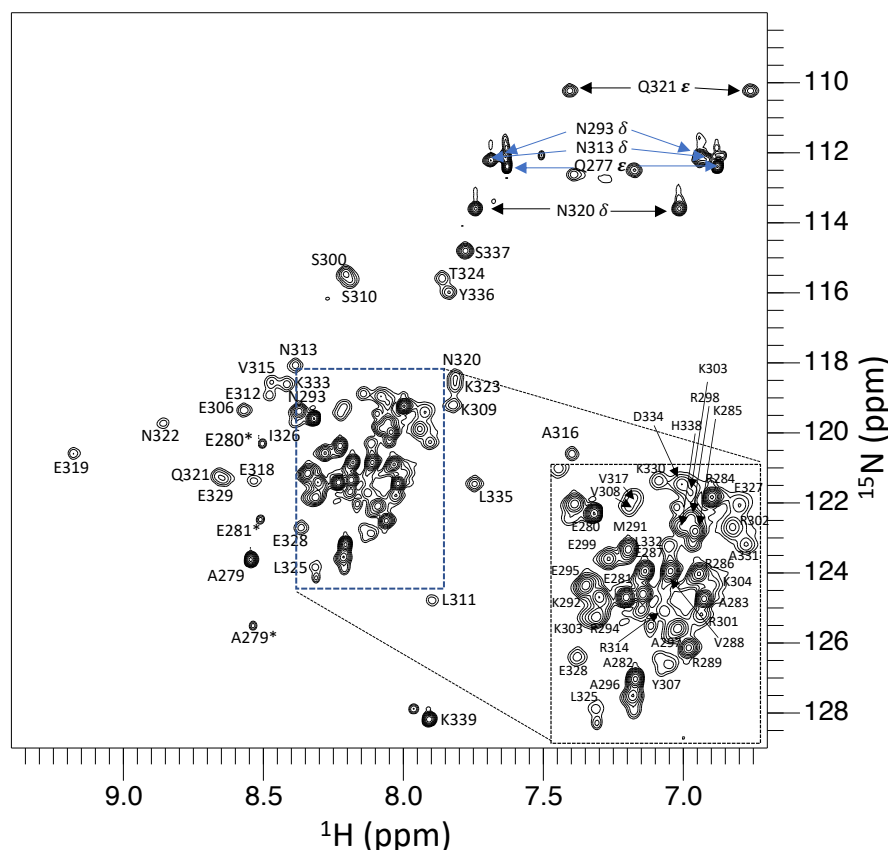


Figure 3.28: Assigned spectra of CREB bZIP 277-339. ^{15}N -HSQC of CREB bZIP 277-339, collected at 25°C in NMR buffer at pH 6.5. Inset is a larger version of the highlighted region in the spectra. * indicates assignment of the residue due to the *cis*-Pro at position 278.

were not as well characterised, for instance the $\text{H}\alpha$ and $\text{C}\beta$ (see Appendix for quality report (Appendix Table S7) and chemical shift table (Appendix Table S8). Also identified were three less intense peaks attributed to the *cis*-proline isomer of residues 279-281. Residues towards the C-terminus of residue 281 were not affected by the *cis-trans* isomerisation of the proline residue (P278).

Dynamic and structural information was obtained for CREB bZIP 277-339 pWT, following assignment. The SSP calculation was carried out with the chemical shifts, which unveiled more helical content in the entire span of the basic region when compared to the 285-339 construct (Figure 3.29a). This was corroborated with the Heteronuclear NOE, which also shows a less mobile basic region compared to the 285-339 construct (Figure 3.29b). Therefore, the increase in N-terminal residues in the 277-339 construct has significantly altered the helical content in the basic region.

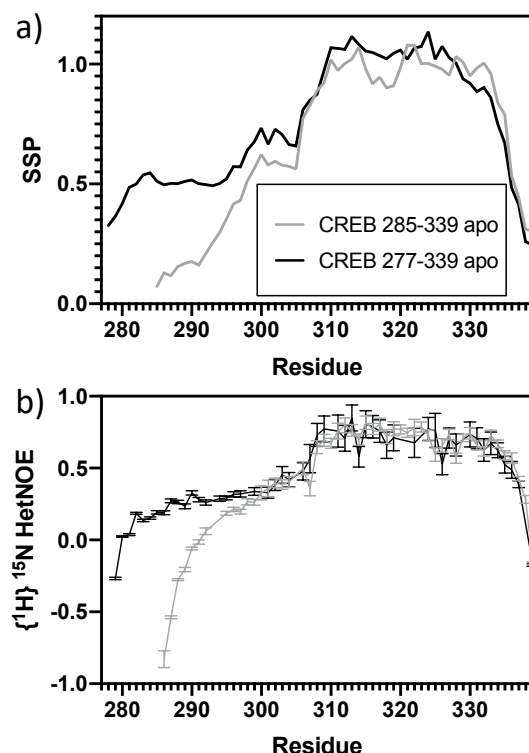


Figure 3.29: Structural and dynamic properties altered by an N-terminal extension. a) SSP prediction of CREB bZIP 277-339 vs 285-339 using $\text{C}\alpha$, $\text{C}\beta$ and $\text{H}\alpha$ chemical shifts. b) (^1H) - ^{15}N Heteronuclear NOE of CREB bZIP 277-339 pWT, compared with 285-339. Data for both constructs were collected at 25°C at 600 MHz.

3.3 Discussion

3.3.1 Homodimer affinity estimate was consistent between a variety of techniques

With any discussion surrounding the bZIP association pathway controversy, the homodimer binding affinity for each protein is crucial to understanding which of these pathways would be preferred. Therefore, a significant amount of effort was placed into answering this question. The approach of focusing on the apo-state of the homodimer complex only was deliberate, but this was with the acknowledgement that these may not be representative of *in vivo* conditions, which involve the full-length protein and many binding partners.

CREB bZIP shows a similar sigmoidal thermal unfolding curve to other bZIP proteins (GCN4^{91,180}, Trp repressor¹⁸¹) and also previously published data on CREB

bZIP (residues 274-341)¹¹³, consistent with the analysis that the folding/unfolding mechanism for bZIP/bHLH proteins is two-state. Interestingly, this is in direct contrast to the thermal unfolding of a longer CREB bZIP construct (259-327)¹²⁷ (equivalent to 273-341 in CREB 1a used in this thesis), which showed a biphasic melting curve. An engineered coiled-coil with a mutation at position *a* of the heptad repeat did show an unfolded intermediate oligomer¹⁸². The CREB bZIP domain used by Moll and colleagues¹¹³ contained a reduced thiol, with the T_m remaining consistent at 49°C in the presence or absence of Mg^{2+} , suggesting that the Mg^{2+} does not influence the dimerisation mechanism. This was confirmed with the NMR of the bZIP domain in the presence or absence of Mg^{2+} not showing any chemical shift perturbation, indicating that Mg^{2+} does not interact with the protein in the apo-state.

Interestingly, the estimated K_d for CREB bZIP was not consistent to previous estimates of 20 μM ¹²⁷, with values presented in this thesis being consistently several orders of magnitude lower (low nM to high pM). Whether the conversion of the data by Santiago-Riveria and colleagues to MRE would unveil a lack of relationship in peptide concentration to ellipticity at 222 nm remains to be seen. The initial slope in the pre-transition, seen in the CREB bZIP melting curve, was also present in GCN4 bZIP¹⁸⁰ and then was noted as a pre-transition due to unfolding in the basic region. Whilst the use of CD data to obtain the homodimer K_d worked, the fitting of the folded and unfolded baselines may have contributed to the unreliability during the extraction of the K_d value, and is notably not the preferred method for estimating the K_d of complexes.

When chemical denaturation of the protein constructs was used, a similar sigmoidal transition to the thermal melts was observed, with no biphasic behaviour. This is consistent with the unfolding process being two-state. Interestingly, Zhu *et al* showed that a GCN4 leucine zipper mutant showed biphasic behaviour with a thermal melt, but a single phase in a urea denaturation¹⁸². The highest concentration of CREB bZIP used did show a large change in the transition curve when compared

to 50 μM , possibly involving intermediates that are resolved at higher protein concentration, however the values were not statistically significant so this is unclear.

Investigations into the folding kinetics of the homodimer complex has been carried out previously for GCN4^{92,183,184}, which unveiled a K_d of 480 nM. In particular, Wendt and colleagues determined that assembly of the coiled coil progresses through at least one kinetically observable step of conformational rearrangement⁹². The folding rate constant of the GCN4 construct is the same order of magnitude as that seen with CREB bZIP pWT, but was 4-fold lower^{183,184}, as well as a model LZ being 2-fold lower⁹³, indicating that the kinetic approach can be vindicated for its reliability. Also, the model for folding of CREB bZIP can be described as two-state, consistent with the previous investigations of other bZIP proteins mentioned above. The rate of folding for CREB is also approaching the diffusion limit for association of proteins, with work by Sosnick *et al* potentially revealing that this is ionic strength independent as their chaotropic agent was guanidium hydrochloride¹⁸⁴, however this remains to be validated. This is supported via $^{13}\text{C}\alpha$ -NMR performed on the GCN4 leucine zipper, which used line-shape analysis at various temperatures, and showed that two conformational states needed to be considered⁹⁵. Interestingly, the unfolding arm of the experiment used with CREB did not yield a reliable K_d estimate. Signal to noise was an issue for this experiment, therefore a higher number of repeats would aid in increasing the signal to noise. Also, the range of urea concentrations that was examined in the unfolding arm was large; a narrower range nearer the $D_{50\%}$ of urea coupled to increased repeats of at each urea concentration would provide a benefit in being able to extract the K_d and unfolding rate of CREB bZIP.

The stability of the homodimer interaction of CREB is high, and this appears to be dependent on intermolecular leucine-leucine hydrophobic interaction. For instance, the half life of a model leucine zipper was shown to be ~ 30 mins, and independent of peptide concentration, obtained by measuring fluorescent quenching⁹³. Mutating one of two leucine residues resulted in a reduction in the half-time, and also presented the evidence for a concentration-independent

unimolecular reorganisation preceded by a concentration-dependent bimolecular association of two monomer chains. However, in each of these mutations there was a reduction in the helical content, as determined by relative fluorescence⁹³. The homodimer complex of CREB remains stable even under high ionic strength conditions, consistent with the hydrophobic interaction being the driving force.

3.3.2 DNA binding region remains dynamic in the apo-state of the bZIP complex

A key result from the work on CREB bZIP domain presented in this thesis was that of the difference in predicted helical content when determined by CD compared to NMR. Conversion to MRE showed that ~60% of the bZIP domain, which in total spanned 285-339, was predicted to be helical. This is consistent with other experiments on CREB and other bZIP containing proteins^{91,92,133,180}. The helical region then agrees with it corresponding to the leucine zipper residues, indicating that the basic region residues are in a disordered state.

In order to confirm that this was the case, NMR was performed on the bZIP domain to obtain information at the residue specific level. Various approaches for estimating the secondary structure content were used, and are in agreement with each other. The leucine zipper residues (residues: 307-339) are in an α -helical conformation, consistent with the evidence from the crystal structure of the residues being in a coiled-coil²¹. A periodicity can be seen from the SSP prediction which appears absent from the $C\alpha$ and $C\beta$ secondary chemical shifts, with the local minima in the SSP corresponding to the residues not in a structural interface. Investigations into this phenomena have been done using residual dipolar couplings on cGK1 α ¹⁸⁵, which revealed under typical experimental conditions, the resolution of the Quality factor varied with the exact heptad periodicity of the coiled-coil.

Interestingly, the estimation of secondary structure using the experimentally determined NMR results differed by a great extent to that obtained by the prediction software Agadir¹⁷⁶. One of the major discrepancies was the helical content in the leucine zipper region, with the NMR results (and CD at lower protein concentrations)

indicating complete α -helical content. Agadir does not predict α -helical content in this region and thus underestimates the amount of helix present in the complex, due to the program not taking into account stabilisation via oligomer complex formation. This presents an issue as the α -helix propagates from the leucine zipper to the basic region in both CREB and GCN4 bZIP domains, potentially playing a major role in function.

As for the basic region, from the SSP results there is a defined decrease in α -helical content from residue 306 extending towards the N-terminus, which is consistent with previous work performed suggesting that the structure of the basic region alters in the presence of DNA^{113,180}. NMR characterisation of the GCN4 bZIP domain had been performed previously⁹⁴ and using the chemical shift information from this the structure of the basic region shows a similar behaviour to that of CREB. A distinct step can be seen from the SSP results of CREB, which correspond to the region of the intramolecular salt bridge between R302 to E299/306, which has been previously identified from the crystal structure²¹. However, this is not seen in the crystal structure of GCN4 bZIP domain²³, where the step can also be identified using SSP. This indicates that the underlying protein sequence similarity for the basic region in bZIPs is controlling the helical content of these ~ 5 -7 residues.

Interestingly, the choice of where to start the N-terminus of bZIP constructs could play a crucial role in determining how they function. Whilst this is investigated in subsequent chapters, the 277-339 construct contained significantly more helical content than the 285-339 construct. With a large range of bZIP domain lengths being reported in the literature (Table 3.1), the evidence shown here for CREB confirms that the N-terminus has a role in controlling the helical content of the basic region¹⁸⁶. The helical content in the 277-339 may be stabilised by the presence of P278, with prolines known to cap regions containing α -helix⁶¹.

An investigation into the quality of the prediction of secondary structure from chemical shifts showed a wide variety of detail, when focusing on the methods themselves, or population of secondary structure at the residual level. For instance, the TALOS+ program was developed to estimate the ϕ and ψ angles of polypeptides

with a 74% success rate¹⁸⁷, with a bias for accurate prediction of structured regions over disordered ones. In contrast, SSP was developed specifically to identify small differences in secondary structure content between α -synuclein and γ -synuclein respectively¹⁶⁵, in order to investigate the role of structural stability in their amyloid forming regions. Therefore, the increased detail for disordered regions shows that the CREB bZIP basic region exhibits a significant amount of disorder at the extreme N-terminus. One interesting point is the variations of use of the various atoms for the SSP prediction, and the variation that this yields on the SSP prediction. Marsh *et al* identified a 0.2 unit difference in the amyloid-forming region of α -synuclein and γ -synuclein, which utilised $C\alpha$ and $C\beta$ chemical shifts only, which with CREB experimental data yielded results greater than a score of 1. The decision for the most suitable parameters to use was difficult, with $C\alpha$, $C\beta$ and $H\alpha$ being recommended by the authors of the program for disordered proteins. However, due to issues with obtaining $H\alpha$ chemical shifts from the HBHACONH experiments in the subsequent Chapters, only $C\alpha$ and $C\beta$ will be used as these provide a reliable estimate of secondary structure content. Therefore, the choice of residue and also the proportion of atoms assigned could yield a significant difference in the SSP output.

4

Investigation of the dimer pathway of CREB association to the target site

Contents

4.1	Introduction	115
4.2	Results	117
4.2.1	Design of a "forced" dimer construct	117
4.2.2	Equilibrium specificity of Dimer to DNA	118
4.2.3	Kinetics of CREB associating to target containing DNA.	124
4.2.4	Structural changes of bZIP interacting with DNA	131
4.2.5	Varied helical content in basic region of CREB 285-339 in complex with different DNA sequences	132
4.2.6	More helical bZIP results in a non-target conformational change that is similar to that bound to target DNA.	141
4.3	Discussion	144

4.1 Introduction

The interaction of CREB bZIP domain with target and non-target DNA will be the focus of this Chapter. In particular, the dimer pathway will be examined in detail (Figure 4.1) using an approach of maintaining the buffer conditions across all experimental techniques, as this has been varied across previous work (Table 4.1). In particular, the non-target DNA interaction was of interest due to the increased

likelihood of this interaction taking place *in vivo*, due to the larger concentration of non-target sites in the genome. Using a range of equilibrium, kinetic and structural methods, in combination with a forced and unforced dimer, a model could start to be developed to help explain why CREB appears to associate to the target DNA site via the dimer pathway.

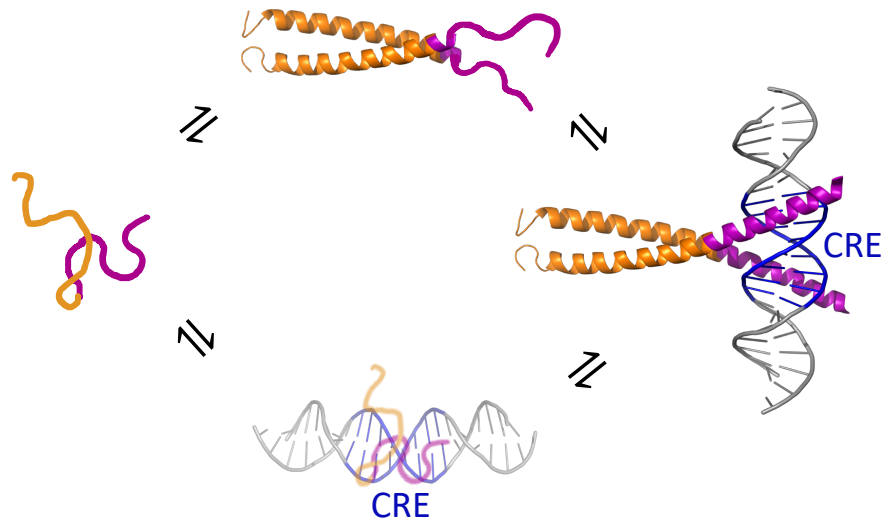


Figure 4.1: Simplified model of the dimer association pathway of CREB bZIP to DNA in the context of both of the possible pathways. The transparent parts of the model indicate the regions that are not focused on in this Chapter. Structure used was PDB: 1DH3.

Table 4.1: Binding affinities for various bZIP transcription factors to target and non-target DNA. Errors are SEM.

Construct	DNA	K_d (nM)	Reference
GCN4	AP-1	22 ± 9	[128]
	CRE	50 ± 30	
	CREHS	30 ± 10	
	NS24	160 ± 50	
Jun	AP-1	250 ± 10	[124]
Fos/Jun	TRE	150 ± 10	[123]
	CRE	210 ± 20	
GCN4 (bZIP)	CRE (full)	2	[80]
Fos/Jun	AP-1	12	[126]
GCN4 bZIP	CRE	2.2 ± 2.9	[81]
CREB bZIP	CRE	50	[127]
CREB bZIP	CRE + Mg	1.8	[108]
	CRE - Mg	> 20	
CREB FL	CRE	4 ± 4	[114]

4.2 Results

4.2.1 Design of a "forced" dimer CREB bZIP construct

In order to see which pathway CREB used for the target association, a comparison would need to be made between a pseudo Wild Type (pWT) and forced dimer (FD) construct. The FD construct is a mutant of the CREB bZIP that is unable to dissociate due to a covalent link. One of the initial plans for designing this construct was to use the CREB bZIP S337C construct used for fluorescent labelling (see section 3.2.2 and Figure 3.5 and 4.2a). However the location was deemed too close to the structured region of the leucine zipper determined by the crystal structure²¹ and NMR (Figure 3.24), and could destabilise the leucine zipper due to torsional stress.

Previous attempts with GCN4 bZIP used a C-terminal linker^{80,129,136}. This approach was adopted for CREB bZIP, by adding a GGC amino acid sequence to the C-terminus of the CREB pWT 285-339 construct. The protocol to ensure complete disulphide formation involved an overnight incubation of the protein in a buffer at pH 9.2, as this pH value is above the pKa for the thiol group of Cys

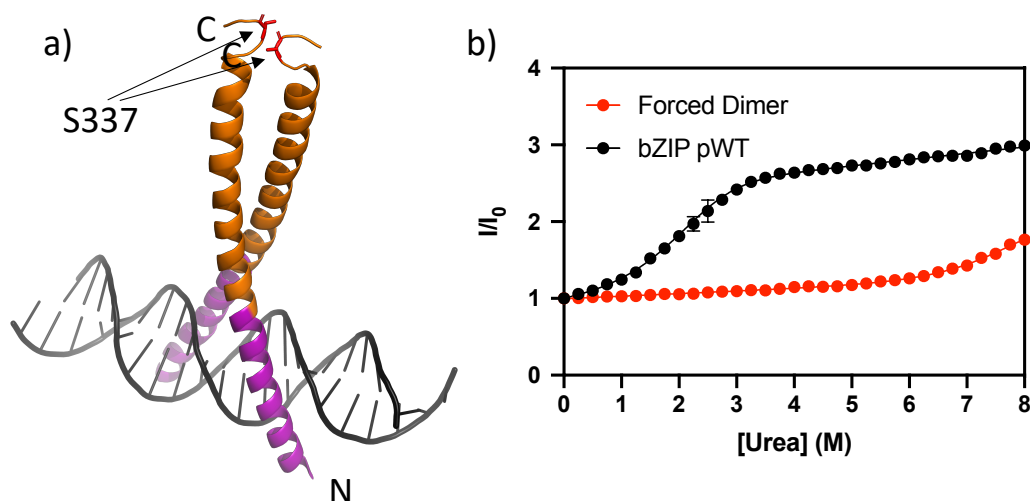


Figure 4.2: Design of the forced dimer CREB bZIP construct. a) Location of the S337 residue in the CREB bZIP crystal structure. PDB = 1DH3. S337C shown as red sticks. b) Equilibrium chemical denaturation of CREB bZIP 285-339 pWT and FD constructs using urea. Normalised Intensity of intrinsic tyrosine fluorescence was measured with 10 μ M protein. Errors are SE, $n=3$.

residue ($pK_a \sim 8.5$), and the deprotonation of the thiol group drives the formation of the disulphide bond¹⁸⁸ (Figure 4.2a). This behaviour was confirmed by ESI-mass spectrometry of CREB 285-339 GGC, which showed only the presence of disulphide linked CREB 285-339 FD (expected MW = 13884.14 Da, Appendix Figure S7). The resulting disulphide linked CREB 285-339 dimer will be referred to as the CREB 285-339 Forced Dimer (FD). The stability of CREB 285-339 FD was examined by a chemical denaturation curve and no unfolding event was observed, indicating the forced dimer construct remained dimeric after an overnight incubation (Figure 4.2b).

4.2.2 CREB bZIP specifically recognises the target CRE consensus site

From the perspective of CREB 285-339, several methods have been used in this thesis to determine the binding affinity of the protein to different DNA sequences. Fluorescence anisotropy was used as probe of protein binding to fluorescently labelled DNA, and performing measurements at various protein concentrations resulted in an equilibrium binding curve. The DNA used here was CRE-DNA, containing the consensus full CRE motif, and non-CRE DNA, which was a scrambled version

of the CRE-DNA sequence maintaining the same GC content. A fit of the curve could then be made to estimate the K_d of CREB associating to CRE and non-CRE DNA, which showed an order a magnitude tighter interaction for CRE compared to non-CRE DNA (Figure 4.3). Note here due to the one site binding equation used, the K_d for the CRE complex is $K_{d,app}$ due to the homodimer K_d being within this range. An intensity change was observed for the CRE DNA sample only ($\sim 25\%$), with a negligible difference in K_d to that observed by anisotropy. In comparison, the intensity change with the non-CRE DNA was minimal within error and not used for fitting.

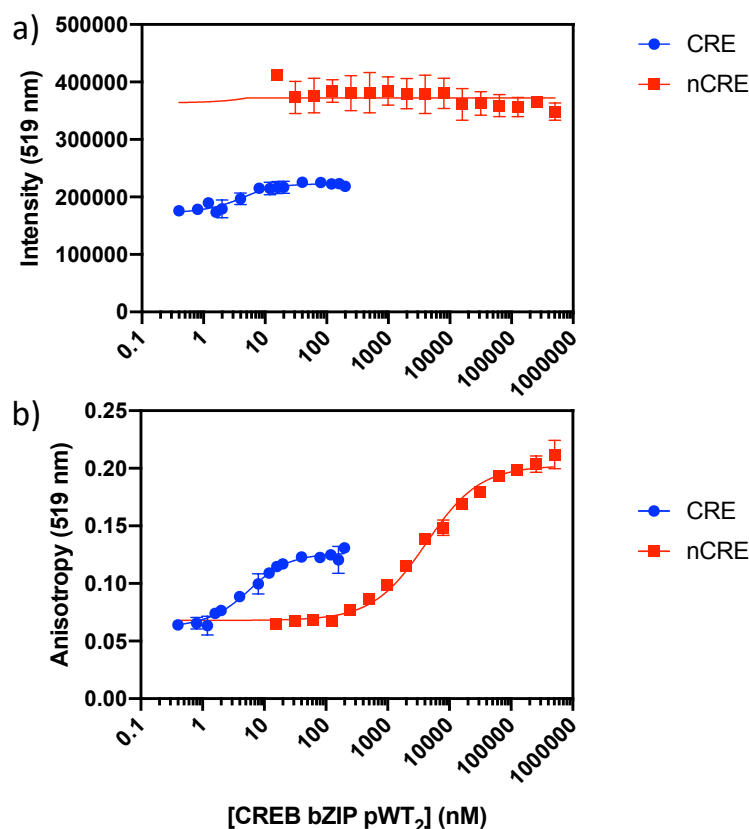


Figure 4.3: Equilibrium binding curves for CREB 285-339 pWT demonstrate high specificity of binding to CRE. a) Fluorescence intensity of 5 nM AF-488 labelled DNA (CRE or n(on)-CRE) titrated against varying concentrations of CREB 285-339 pWT. b) Similar to a), but using fluorescence anisotropy. $n=3$, errors are SD. Performed at 25°C.

An ionic strength dependence on the K_d of CREB 285-339 pWT to DNA was performed in order to interrogate the stability of the protein/DNA complexes

(Figure 4.4). A linear dependence was seen on a log-log plot of K_d vs ionic strength for both DNA constructs, but interestingly the gradient of these opposed each other, with the CRE DNA $K_{d,app}$ becoming lower at increasing ionic strength. As for the non-CRE DNA, the K_d increased at higher ionic strengths. A point to note is the K_d at 185 mM ionic strength is the biophysical buffer ionic strength used throughout this thesis unless otherwise noted.

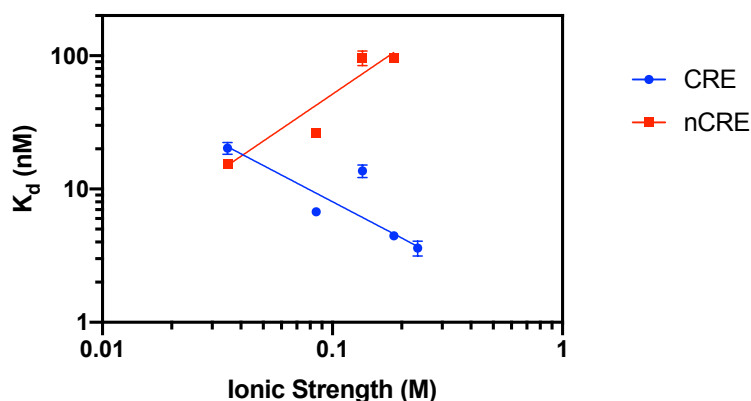


Figure 4.4: CREB bZIP with CRE is stable across varying ionic strengths. Log-log plot of K_d or $K_{d,app}$ against ionic strength for CREB bZIP associating to CRE and non-CRE DNA at 25°C. Errors are SEM.

A similar experimental design for the forced dimer CREB 285-339 construct was used and showed an order of magnitude tighter interaction with CRE DNA, but the same affinity for the non-CRE DNA (Figure 4.5). Interestingly, the use of the CREB 277-339 forced dimer construct with the hairpin CRE DNA showed the highest affinity interaction. CREB 277-339 (FD) also has a similar K_d for non-CRE DNA as the pWT and FD versions of CREB 285-339. However, a comparison between the hairpin and linear DNA sequences used with CREB 277-339 should be taken with caution due to their different designs and potential for unknown secondary structure (Figure 4.5).

To confirm the specificity that CREB bZIP has shown for target containing DNA sequences, Isothermal Titration Calorimetry (ITC) was used. This technique is a label free analysis of the enthalpy change upon an interaction taking place, the analysis of which can yield information on the binding affinity and stoichiometry of the interaction. ITC was performed with CREB bZIP pWT 285-339 with two variants

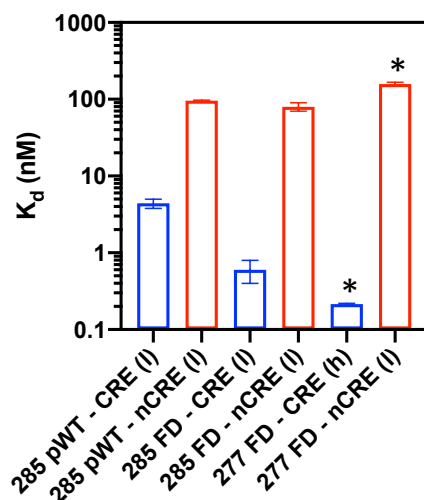


Figure 4.5: Specificity for target DNA was increased with a combination of forced dimerisation and extension of the N-terminus. K_d or $K_{d,app}$ of the interaction between CREB bZIP constructs starting from residue 285 or 277-339 and DNA. Sequences of CRE DNA used in these experiments were the NMR-CRE. Data was taken from a titration of protein to AF488-labelled DNA. * indicates data collected by Dr M. Kuravskiy. Errors are SEM. h = hairpin, l = linear.

of target DNA (NMR CRE: 5'-AGATGACGTCATCT-3', Stopped flow (SF) CRE: 5'-CCTGACGTCATCCG-3' and the non-CRE DNA (5'-GGCTAAAGCATTCT-3'), Figure 4.6). The specificity that CREB 285-339 had for the target containing, CRE DNA was reproduced here with ITC, with a K_d over a magnitude lower than for non-CRE DNA (Figure 4.7a). The specificity was not due to a change in the number of sites for the interaction, as the approach of CREB being titrated to DNA indicated one binding site per DNA molecule regardless of sequence (similar to the fluorescence equilibrium binding curves). In contrast, the approach of titrating DNA to protein showed two binding sites per protein, which could be down to the protein being dimeric (in this experiment the minimum concentration of protein would suggest it remaining dimeric throughout) and therefore contains two binding sites per one DNA molecule (Figure 4.7a).

The specificity of the interaction is thermodynamically driven, with the ΔG being lower for the CRE DNA interactions than for the non-CRE DNA (Figure 4.7b). Interestingly the enthalpy of the interaction seems dependent on the approach of the experiments (i.e. protein to DNA or vice versa), with CREB being added

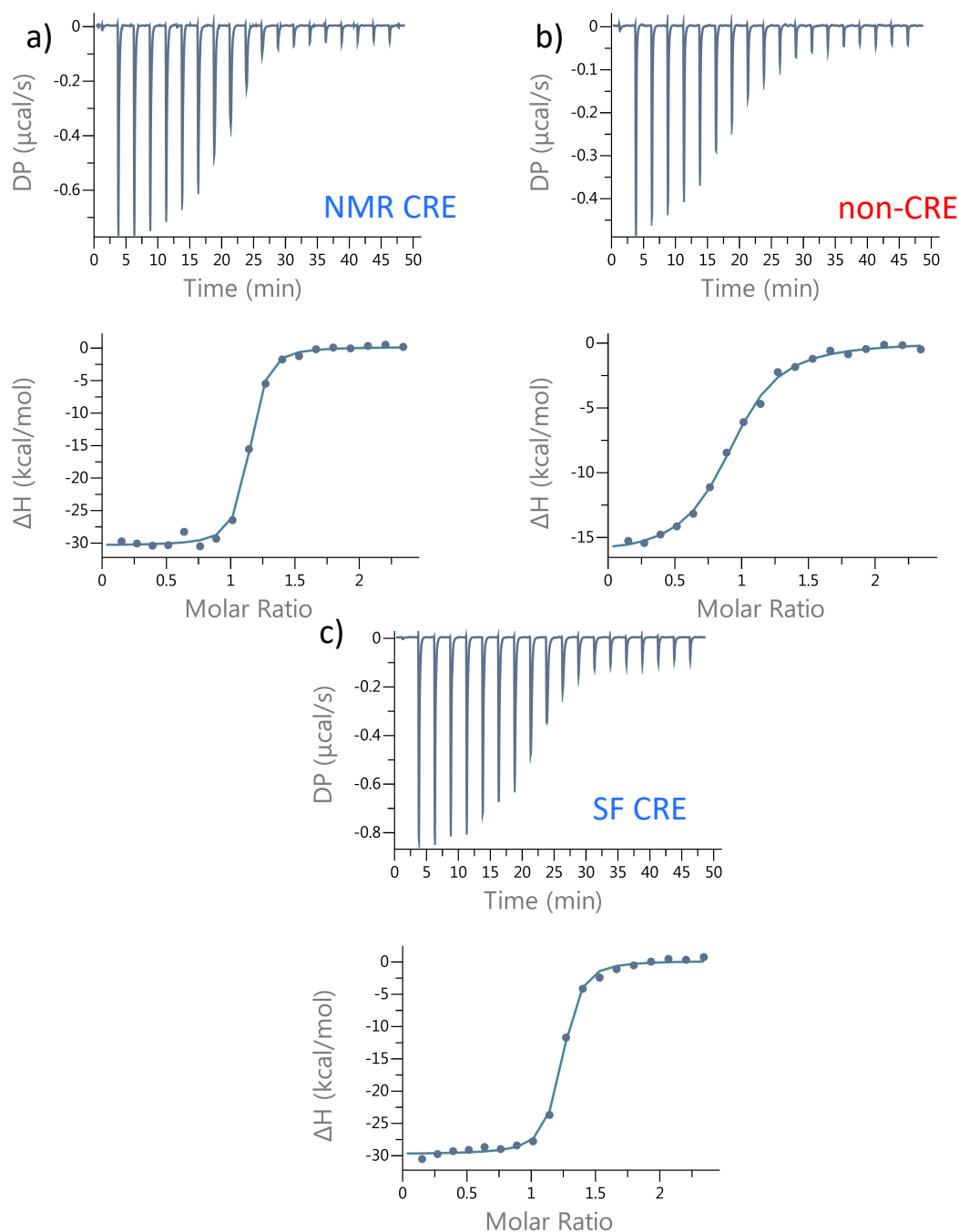


Figure 4.6: A larger enthalpy change for target DNA association was seen for CREB bZIP. a) ITC was performed with CREB 285-339 pWT. Top: representative power differential (DP) plot between sample and reference cells. Bottom: Representative enthalpy change upon addition of DNA to protein plot. b) and c) The same as a) for addition of non-CRE and SF-CRE DNA respectively.

to CRE DNA resulting in a enthalpy almost twice as low as the CRE DNA to CREB approach, but with a significantly larger error. A similar trend was

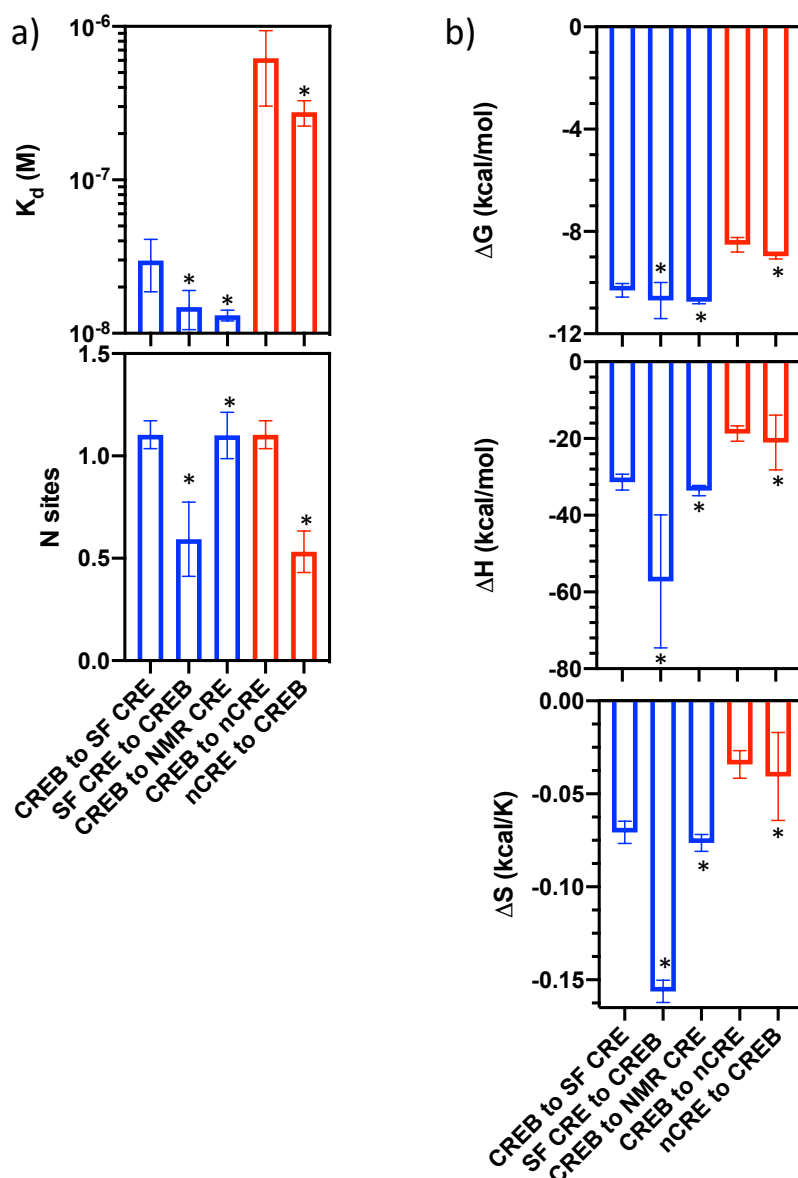


Figure 4.7: Thermodynamics of CREB interacting with DNA shows a preference for target containing sequences. a) ITC was performed with CREB bZIP titrating into DNA and vice-versa. Top: K_d values for the various complexes that underwent analysis by ITC. Bottom: number of sites that was estimated from the fitting procedure. b) Thermodynamic properties that were able to be fitted from the binding isotherms in Figure 4.6. Top: Gibbs free energy of the interaction of CREB interacting with different DNA sequences. Middle: same as the top but focusing on enthalpy change. Bottom: Same as top and middle looking at the change in entropy. Errors are SD, $n=3$ or 2^* .

observed for the entropy of the interaction with the CRE DNA samples being significantly different within error.

A comparison of the K_d determined from fluorescence anisotropy/intensity

to ITC showed that the affinity was higher during the fluorescence anisotropy experiments than during the ITC experiments (Figure 4.8). The same ionic strength was used as well as the same pH, with the significant difference being the fluorescently labelled DNA used in the fluorescence anisotropy/intensity titration. An examination of CREB 277-339's propensity to associate to AlexaFluor488 was examined at two concentrations, one without protein and one with a high protein concentration (Appendix Figure S9), and an anisotropy change was observed.

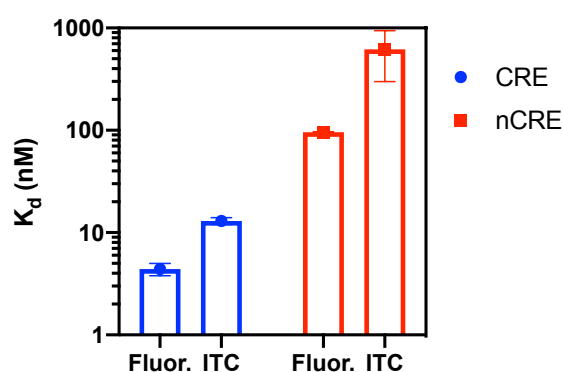


Figure 4.8: Comparison of the K_d of CREB 285-339 pWT associating to CRE and non-CRE DNA, derived from two equilibrium state experimental techniques at 25°C. Errors are SEM.

4.2.3 Kinetics of the interaction of CREB with target containing DNA.

Following the kinetics of the interaction between CREB bZIP constructs with target or non-target DNA sequences would reveal whether the difference in the K_d for the different complexes are the result of the association or dissociation rate. As the $K_{d,app}$ of CREB to CRE DNA could be followed by a fluorescent intensity change of the fluorescently labelled DNA, this was used for the stopped flow fluorescence experiments.

For the determination of the association rate (k_{on}), AF488 labelled CRE DNA (sequence name: Stopped-Flow CRE (SFCRE)) was rapidly mixed with varying concentrations of CREB 285-339 pWT and FD constructs under pseudo first-order conditions, to compare if being completely dimeric influenced the association rate.

Each of the protein concentration points were able to fit to a single exponential (Figure 4.9), indicating that a two-state process was observed for both the pWT and FD constructs (Figure 4.10a). A plot of k_{obs} against protein concentration showed linear dependence for both FD and pWT constructs, however the straight line did not pass through the origin. Forcing the fit through the origin did not affect the results (Table 4.2). Therefore, the k_{on} for CREB 285-339 could be determined and showed a very fast association reaction, with an advantage for the pWT construct over the FD.

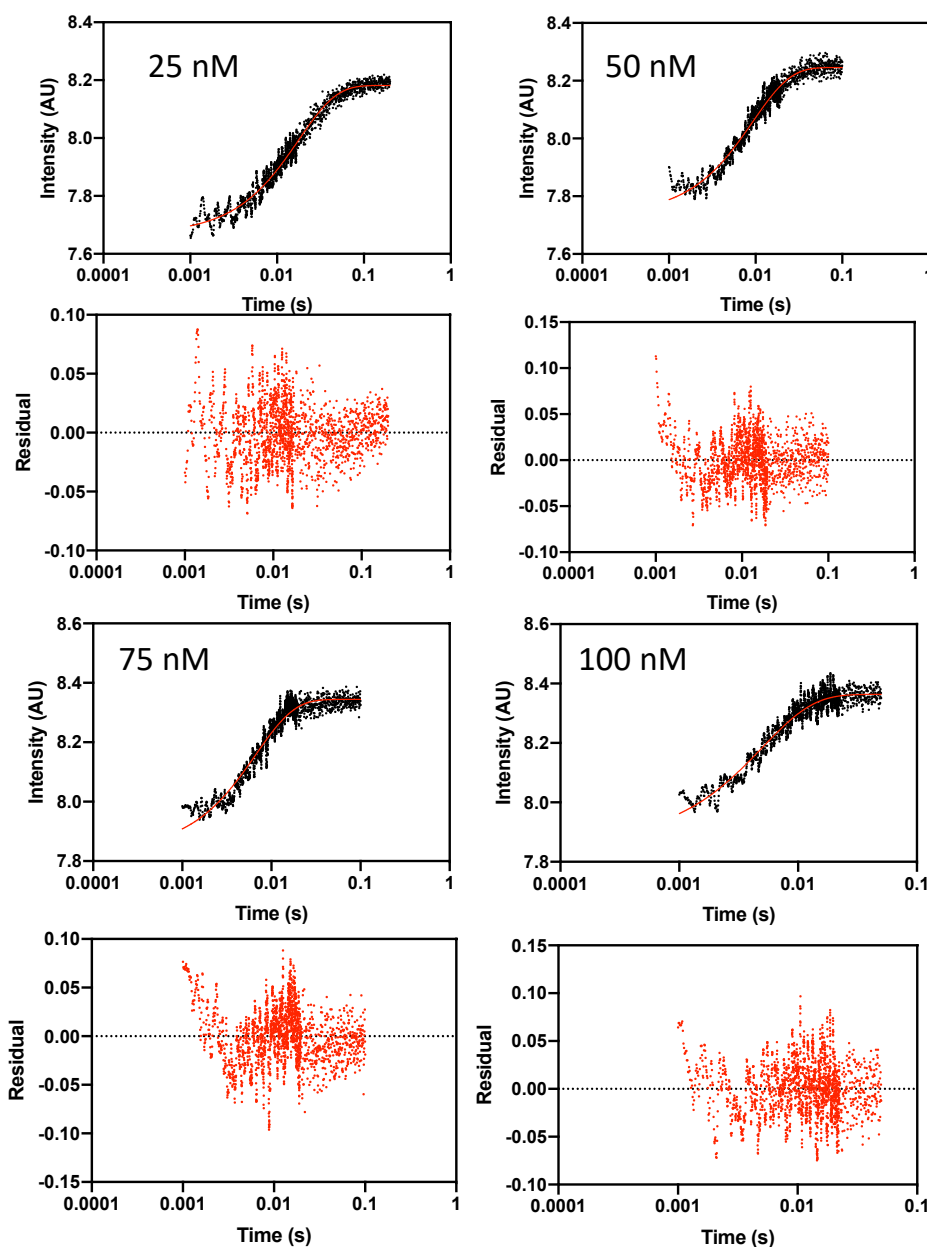


Figure 4.9: k_{on} of CREB 285-339 FD to target containing DNA can be fitted to a single exponential. Fluorescence intensity changes of AF488-CRE DNA being rapidly mixed by stopped flow to CREB 285-339 FD. Over 40 traces were averaged in each case and fit to a single exponential equation, with the residuals of each fit shown below.

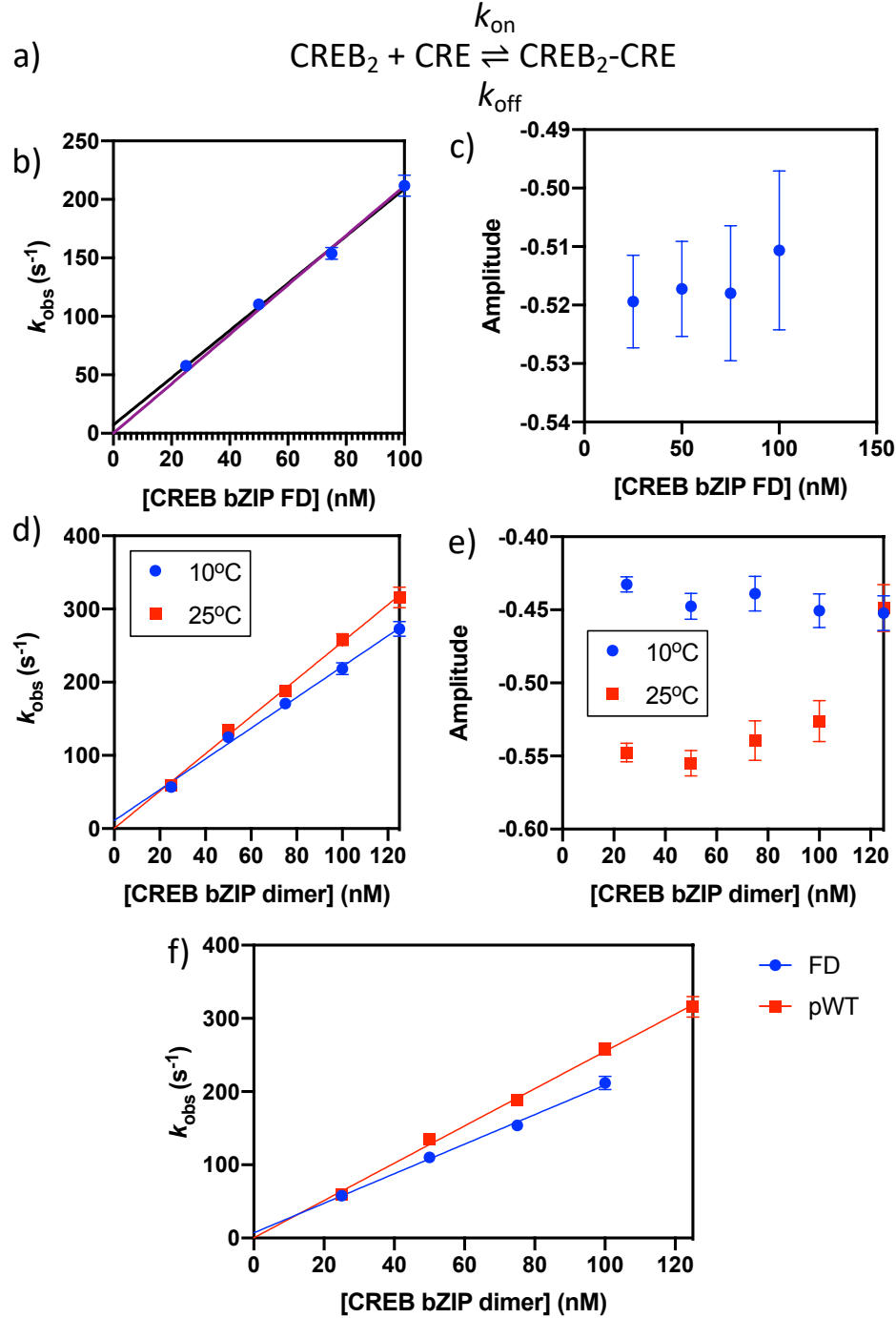


Figure 4.10: Association rates to target DNA sites are faster for the pWT than for the FD. a) Kinetic scheme for a two-state interaction between CREB 285-339 constructs and CRE DNA (SFCRE). b) k_{obs} of CREB 285-339 FD associating with SFCRE DNA via stopped flow fluorescence of labelled DNA. Black line of best fit was not forced through the origin purple was forced through. c) Amplitudes of each point in a). d) Similar to a), but with CREB 285-339 pWT performed at 10 and 25°C. e) Amplitudes extracted from the data in d). f) Comparison of the CREB 285-339 pWT and FD shown in b) and d). Rates and amplitudes extracted from single exponential fits, errors are SEM.

Table 4.2: Fitted parameters from the plot shown in Figure 4.10 showing the influence of forcing through the origin.

CREB 285-339	$k_{\text{on}}(\text{nM}^{-1} \text{ s}^{-1})$	Y-intercept
FD	2.0 ± 0.1	7.1 ± 6.8
	2.11 ± 0.04	0
pWT	2.5 ± 0.1	0.4 ± 6.2
	2.55 ± 0.03	0

The dissociation rate constant (k_{off}) was estimated using stopped flow, by rapidly mixing 40 nM CREB (assuming dimer concentration) saturated AF488-SF-CRE DNA (5 nM) with unlabelled SF-CRE DNA at varying concentrations, and observing the intensity change with time. The averaged fluorescence intensity data was fitted to a single exponential decay function to extract the observed rate constant (k_{obs} , Figure 4.11), due to the lack of improvement a double exponential fit yielded. The amplitude was the opposite sign to the k_{on} measurements. To extract the k_{off} , data was fit to an equation assuming reversibility during the dissociation reaction (Equation 2.22). The k_{off} for the CREB 285-339 pWT from SFCRE DNA was shown to be ~ 1.5 times higher than the FD, consistent with the forced dimer construct having a tighter K_d (Table 4.3).

Table 4.3: k_{off} parameters of CREB 285-339 constructs from SFCRE DNA. Data fitted with equation 2.22 Error is SEM.

CREB 285-339	$k_{\text{off}} (\text{s}^{-1})^*$
FD	0.22 ± 0.05
pWT	0.33 ± 0.08

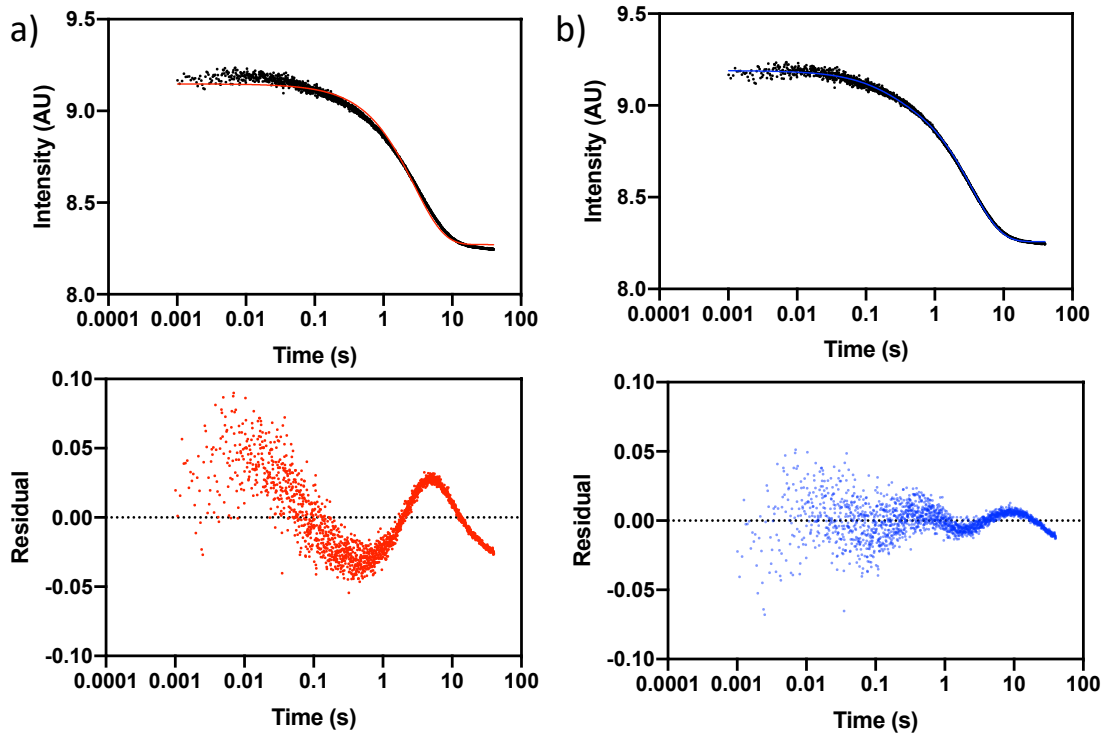


Figure 4.11: k_{obs} could be determined from a single exponential fit to stopped flow dissociation experiments. a) Top - Fluorescent intensity change of CREB 285-339 FD in complex with AF488-SFCRE DNA mixed with unlabelled SFCRE DNA via stopped flow. Data was averaged from over 20 kinetics traces and fit to single exponential. Bottom - residuals of the fit from the top. b) Same as a) but fit with a double exponential equation.

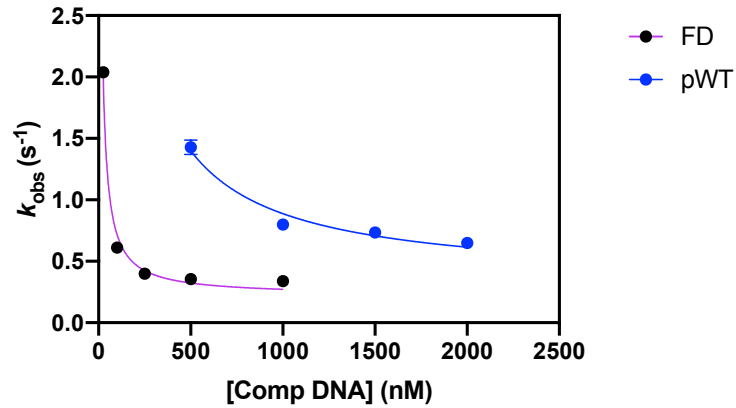


Figure 4.12: k_{off} was markedly different between pWT and FD CREB 285-339 constructs. The dissociation of CREB 285-339 pWT and FD (40 nM dimer) constructs from AF488-SFCRE DNA (10 nM) using non-labelled SFCRE DNA was measured using stopped flow fluorescence intensity. Data was fitted to equation 2.22. Errors are SEM.

CREB 277-339 FD construct also underwent association and dissociation kinetics, with k_{on} being $2.57 \pm 0.06 \text{ nM}^{-1} \text{ s}^{-1}$ and k_{off} being $0.55 \pm 0.008 \text{ s}^{-1}$ (collected by Dr M. Kuravskiy). Therefore, the values obtained for the two constructs have similar values for the association and dissociation rates. Using the k_{on} and k_{off} values determined from stopped flow kinetics to evaluate the K_d of CREB 285-339 to associate to CRE-DNA, K_d of pWT to CRE was 0.44 nM and FD to CRE was 0.16 nM.

Stopped flow anisotropy was utilised to determine the kinetics of the CREB 277-339 FD interaction with non-CRE DNA, due to the lack of an fluorescence intensity change upon complex formation as mentioned above. To determine the k_{off} of CREB from non-CRE DNA, 50 nM CREB 277-339 FD in complex with 50 nM AF488-nonCRE DNA was rapidly mixed with non-labelled CRE DNA and anisotropy change observed over time (Appendix Figure S8). Initially attempted at 25°C, no change in amplitude was observed so a temperature dependence was performed. At low temperatures, the amplitude remained consistent and the rate increased, however at higher temperatures the amplitude started to decrease with a corresponding decrease in the k_{obs} . A plot of $\ln(k_{\text{obs}})$ against $1/T$ with the lower temperature points only was able to be fitted by linear regression, with an extrapolated k_{off} at 25°C to be $600 \pm 50 \text{ s}^{-1}$ (Appendix S10). To obtain the k_{on} , a similar approach had to be utilised (Appendix Figure S11 and S12) to that for the k_{off} . Oddly, the amplitude change for the k_{on} of CREB 277-339 FD to non-CRE DNA (hpn) was the same sign as the k_{off} experiment, with the amplitude expected to become more positive rather than negative. The k_{on} from this experiment could thus be obtained using the similar method for the k_{off} above, but yielded unexpected results ($k_{\text{on}} = 8000 \text{ M}^{-1} \text{ s}^{-1}$).

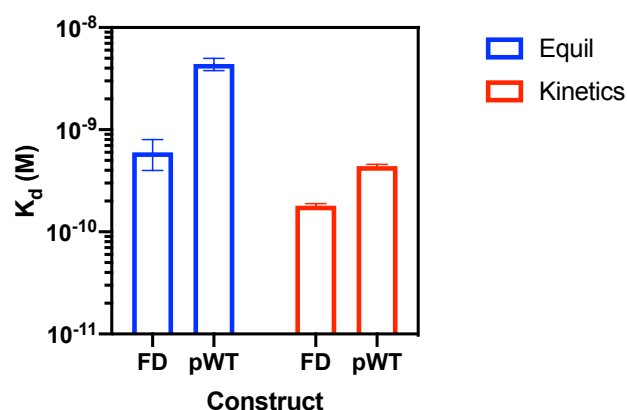


Figure 4.13: Comparison of K_d estimates of CREB 285-339 to CRE-DNA between those obtained from equilibrium and kinetics. Errors are SEM.

4.2.4 CREB bZIP undergoes coupled folding upon binding to DNA

With the specificity for target DNA being confirmed for CREB bZIP (section 4.2.2), the next stage was to determine whether a structural change in the protein during the interaction was influencing the affinity for the different DNA sequences. From section 3.2.3, it was shown that the basic region of the bZIP domain was disordered in the absence of DNA. Therefore, by adding in DNA it is expected to change the structure of the basic region, and this was evaluated initially using Circular Dichroism (CD, Figure 4.14). The CD spectra showed that, in complex with CRE-DNA, there is a significant increase of helical content, compared to the non-CRE DNA complex which contained a smaller increase in helical content. What the data does not show is where within the protein structure to the partial increase in helical content that can be seen in the CREB with non-CRE DNA complex.

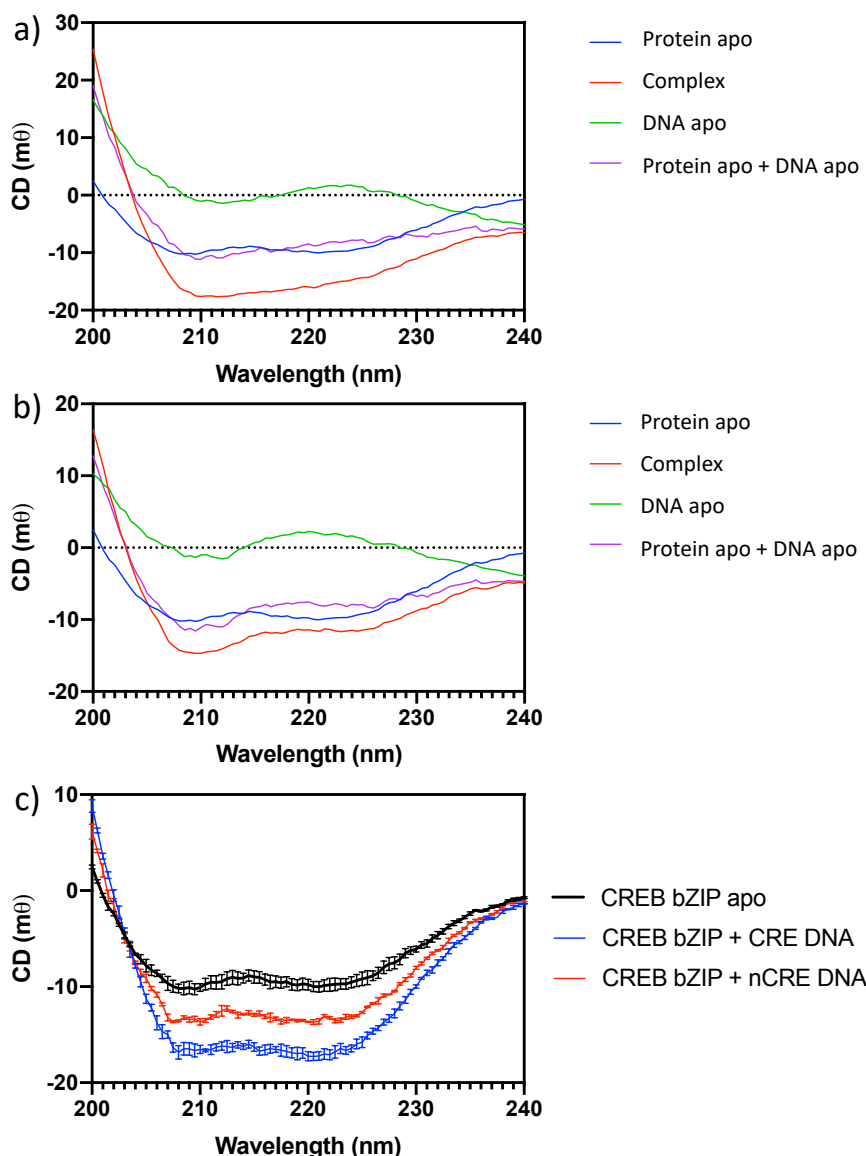


Figure 4.14: Secondary structure of CREB increases upon addition of DNA. a) and b) Difference plot of CREB bZIP in complex with CRE- and non-CRE DNA, respectively, collected at 25°C. c) CD spectra of the DNA signal subtracted from the protein signal. Errors are SD, n=3.

4.2.5 Varied helical contents in basic region of CREB 285-339 in complex with different DNA sequences

Following the results from the CD experiments showing an increase in helicity of CREB bZIP when associating to DNA (section 4.2.4), further structural characterisation of the complex was undertaken using NMR. This was in order to determine the conformational change, in solution, on the residue specific level which

NMR can provide. Therefore, assignments of the bZIP domain (CREB 285-339) were undertaken in a saturating concentration of the target containing CRE DNA (Figure 4.15). The resonances that correspond to the basic region (residues 285-307) were more dispersed than in the apo-state (see Figure 4.16), indicating a significant conformational change in these residues. Assignment of the CREB

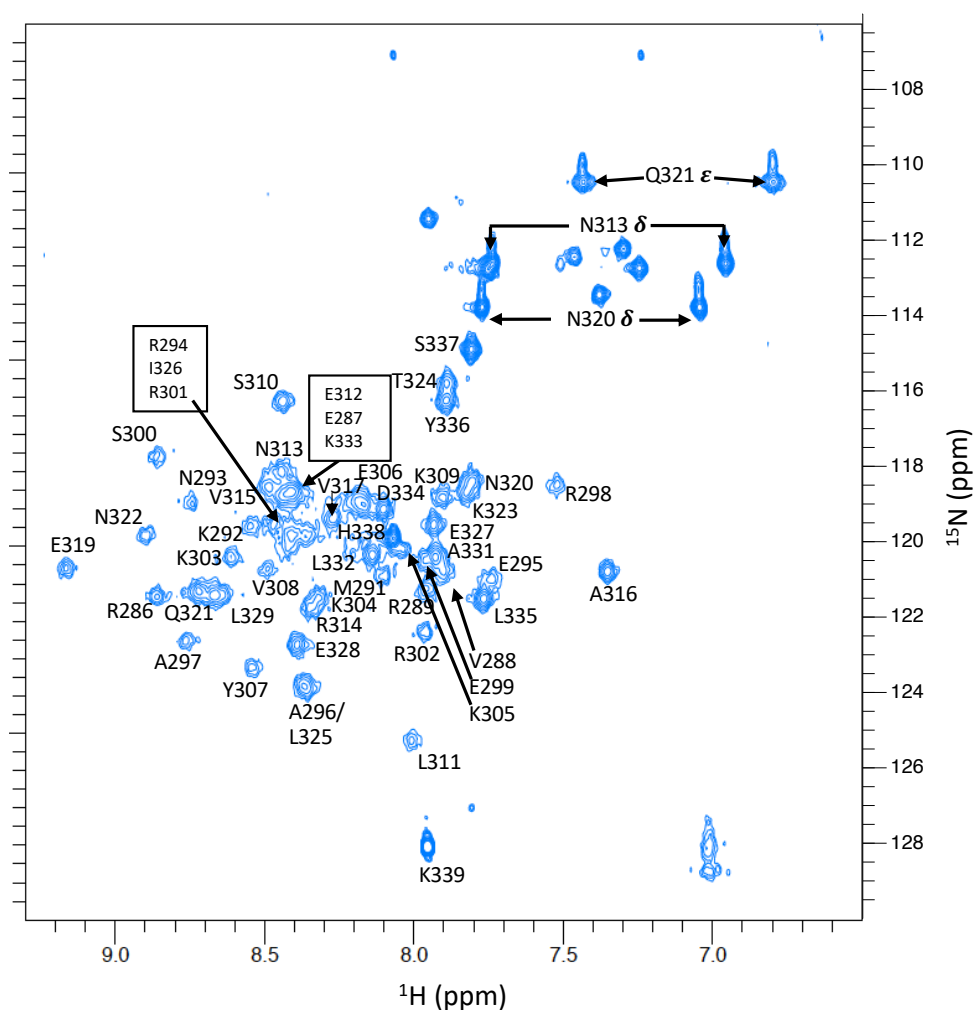


Figure 4.15: Peak dispersion seen of CREB bZIP in association with CRE-DNA. ^{15}N -HSQC of 1 mM CREB bZIP 285-339 (monomer concentration) in complex with a saturating concentration of CRE-DNA (1 mM), collected at 25°C.

285-339 with CRE DNA complex were performed with triple resonance experiments as previously described in Section 3.2.3. Initial experiments were performed at 25°C, but low signal from triple resonance experiments (despite the use of a BEST-HNCA

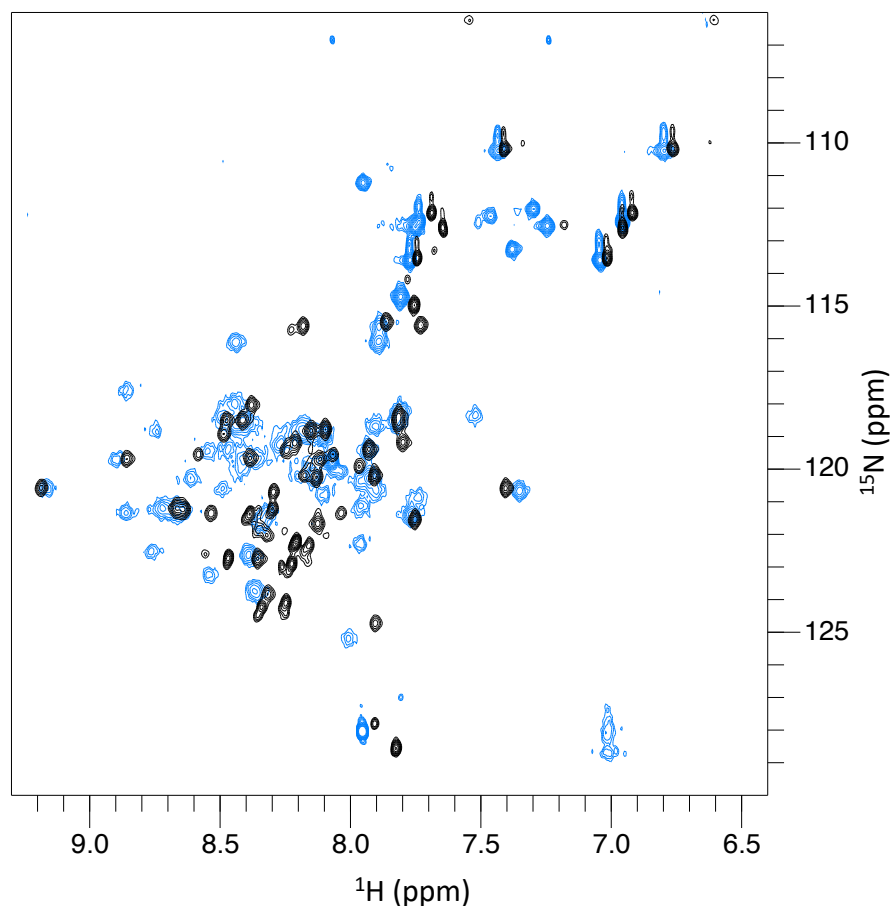


Figure 4.16: Increased dispersion in ^1H dimension of CREB 285-339 upon addition of CRE DNA. Overlay of ^{15}N -HSQCs of CREB 285-339 in the apo (black) and in a saturating complex with CRE DNA (blue). Data collected at 25°C.

pulse sequence) hampered the ability to link resonances for complete assignment. Following a discussion with Prof's C. Redfield and J. Schnell, moving towards a non-uniform sampling method of data acquisition coupled to a increase in temperature to 35°C, improved the quality of the signal sufficiently to assign the protein spectrum. Increasing the temperature was initially concerning due to the potential of changing the population of free and bound states that could influence the secondary structure, but no secondary structure changes were observed in a ^{13}C -HSQC (Figure 4.17). The chemical shift table and assignment report can be found in the appendix (Appendix Tables S9 to S11), with the assigned ^{15}N -HSQC at 35°C shown in appendix Figure S13. With the complex of CREB 285-339 and CRE DNA, the appearance of peaks possibly corresponding to those of the Arg N ϵ were seen at ~ 7.0 ppm on the ^1H

axis and ~ 128 ppm on the ^{15}N , which were folded in this spectrum.

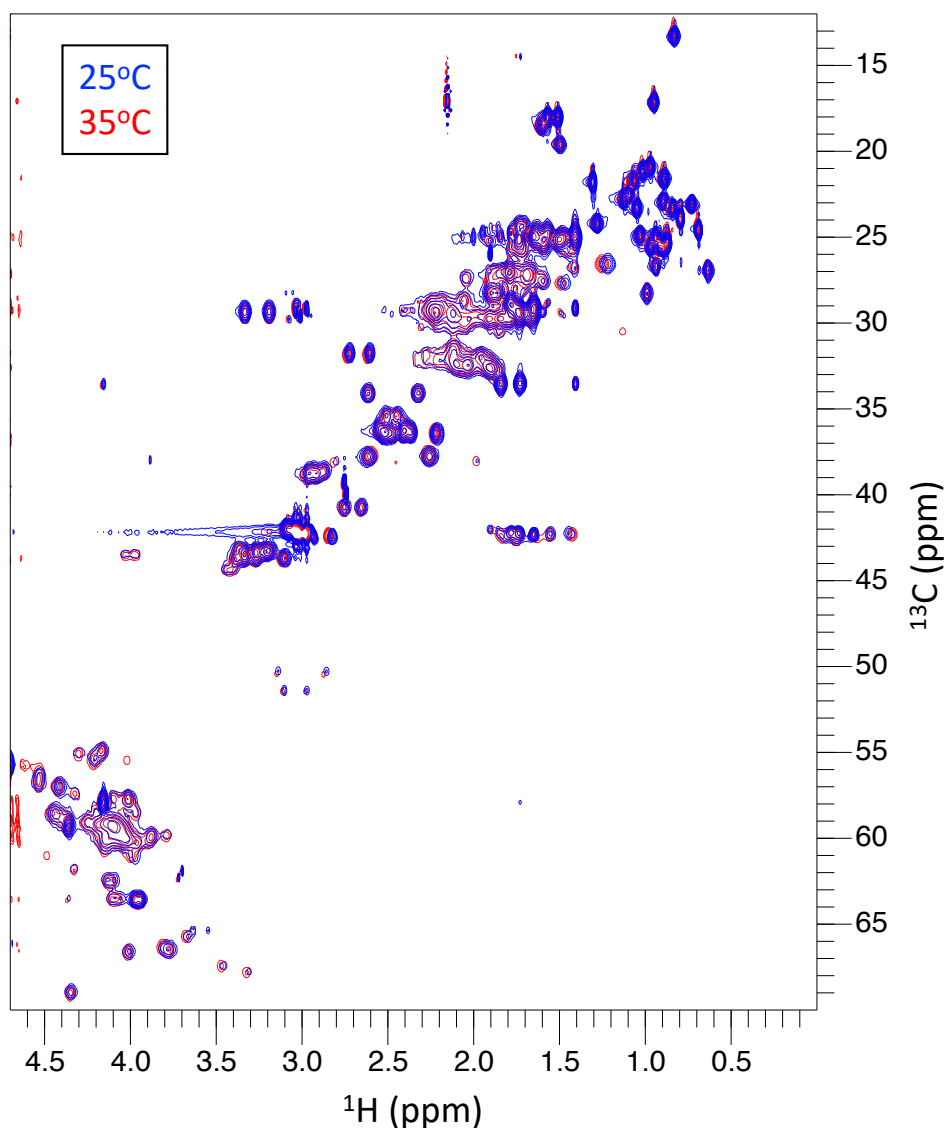


Figure 4.17: Carbon chemical shifts of CREB 285-339 in complex with CRE DNA did not vary with temperature. ^{13}C -HSQC of CREB 285-339 with CRE DNA, performed at two temperatures (25°C and 35°C). The spectra were referenced using DSS.

In comparison to the assignment of CREB 285-339 in complex with CRE DNA (and also the monomer model in section 5.2.4), assignment of peaks of CREB 285-339 in complex with non-target DNA sequence was more challenging. One of the first steps was to try a non-CRE DNA titration into isotopically labelled protein. Even though the affinity for this interaction was known to be high, the titration revealed the complex was in fast exchange (Figure 4.18).

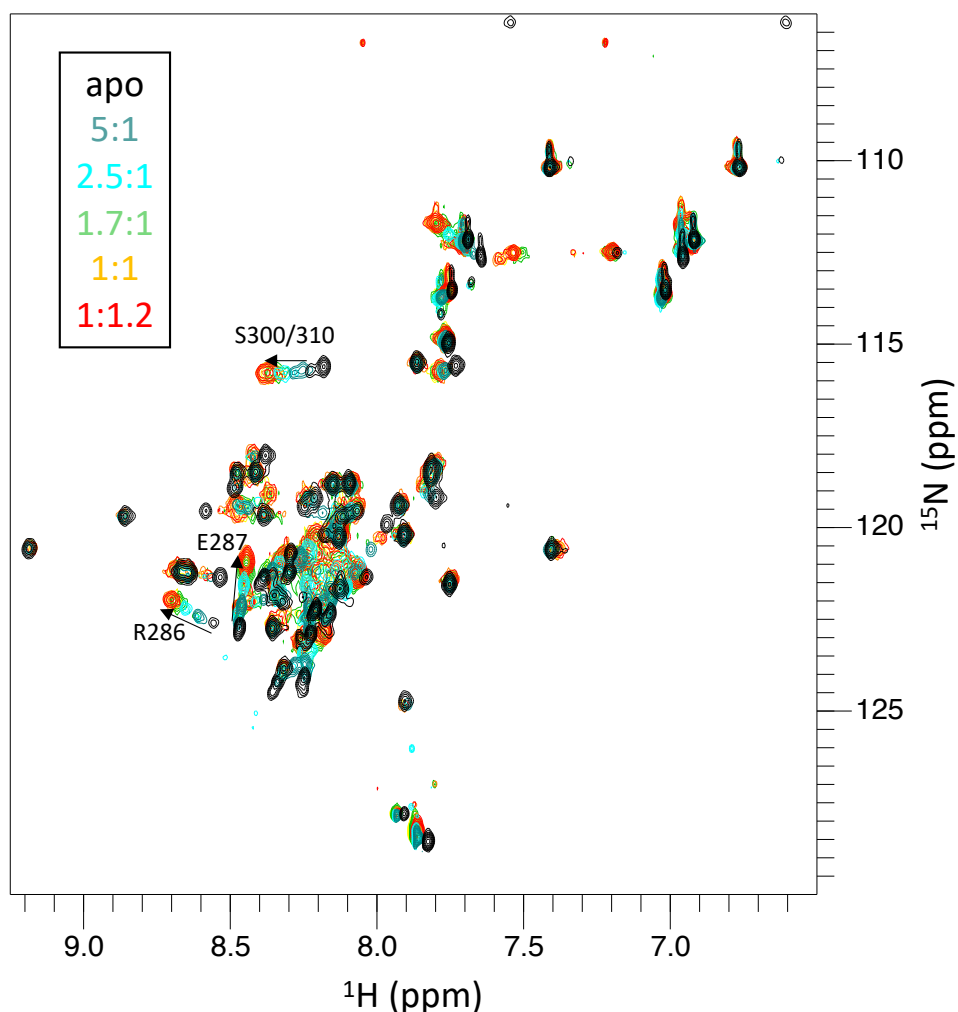


Figure 4.18: CREB 285-339 is in fast exchange with the non-CRE DNA. ^{15}N -HSQC of CREB 285-339 (1 mM of monomer) in complex with varying concentrations of non-CRE DNA. All data collected at 25°C at 750 MHz.

However, following all of the peaks to their saturated state was not possible. This was due to a heavily overlapped region between the ^1H region of 8.0-8.5 ppm in the ^{15}N -HSQC. Triple resonance experiments were attempted at 25°C initially, following which all of the peaks in the ^{15}N -HSQC were assigned by sequential assignments, coupled with the initial titration points to follow the general peak trajectories (Figure 4.19). However, a significant number of $\text{H}\alpha$, $\text{H}\beta$, $\text{C}\alpha$ and $\text{C}\beta$ peaks were missing (see appendix Table S12 and S13).

As going to the higher temperature increased the signal of the triple resonance with the protein and CRE DNA complex, this was used for the triple resonance

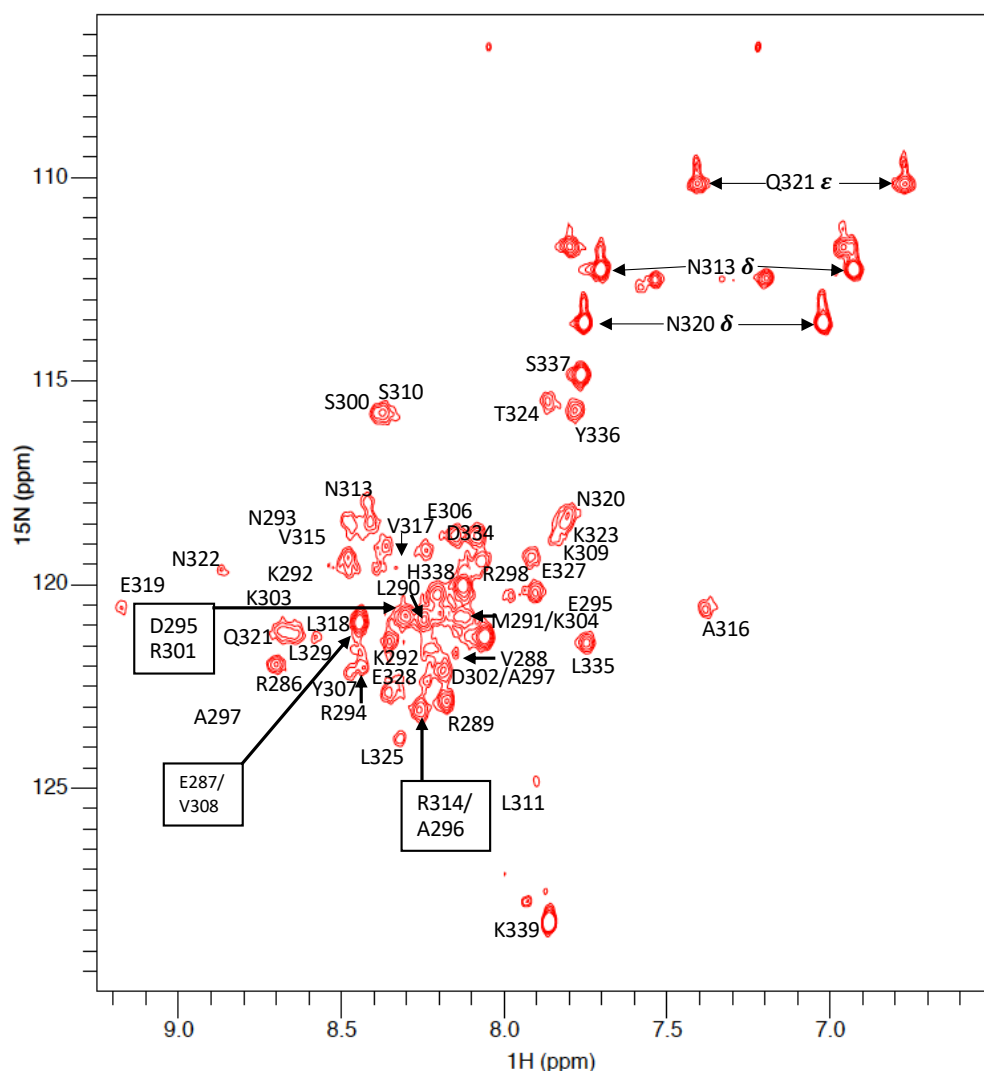


Figure 4.19: CREB 285-339 was fully assigned with non-CRE DNA at 25°C. ^{15}N -HSQC of CREB 285-339 in complex with non-CRE DNA. Assignments were performed at 25°C, from triple resonance experiments also performed at these conditions.

experiments of CREB 285-339 with non-CRE DNA. All of the residues were able to be assigned at this temperature, aside from R302, but there was a marked improvement in the number of $\text{H}\alpha$, $\text{H}\beta$, $\text{C}\alpha$ and $\text{C}\beta$ assigned (Appendix Tables S14 and S15). Similarly, there were only a few observed differences when an overlay of the two ^{13}C -HSQC spectra, with a few $\text{C}\alpha$ peaks having shifted (Appendix Figure S15). The difference in the $\text{C}\alpha$ chemical shifts between 25 and 35°C revealed that between residues 295 and 298, the difference was at most 0.186 ppm, with residues

N-terminal of this region having a larger difference (maximum difference of 0.885 ppm for residue 287, average was 0.54 ppm).

Secondary structure was calculated using the measured chemical shifts for each of the complexes, with the non-CRE DNA analysis being performed at two temperatures, using SSP¹⁶⁵ (Figure 4.20a-b). CREB 285-339 in complex with CRE

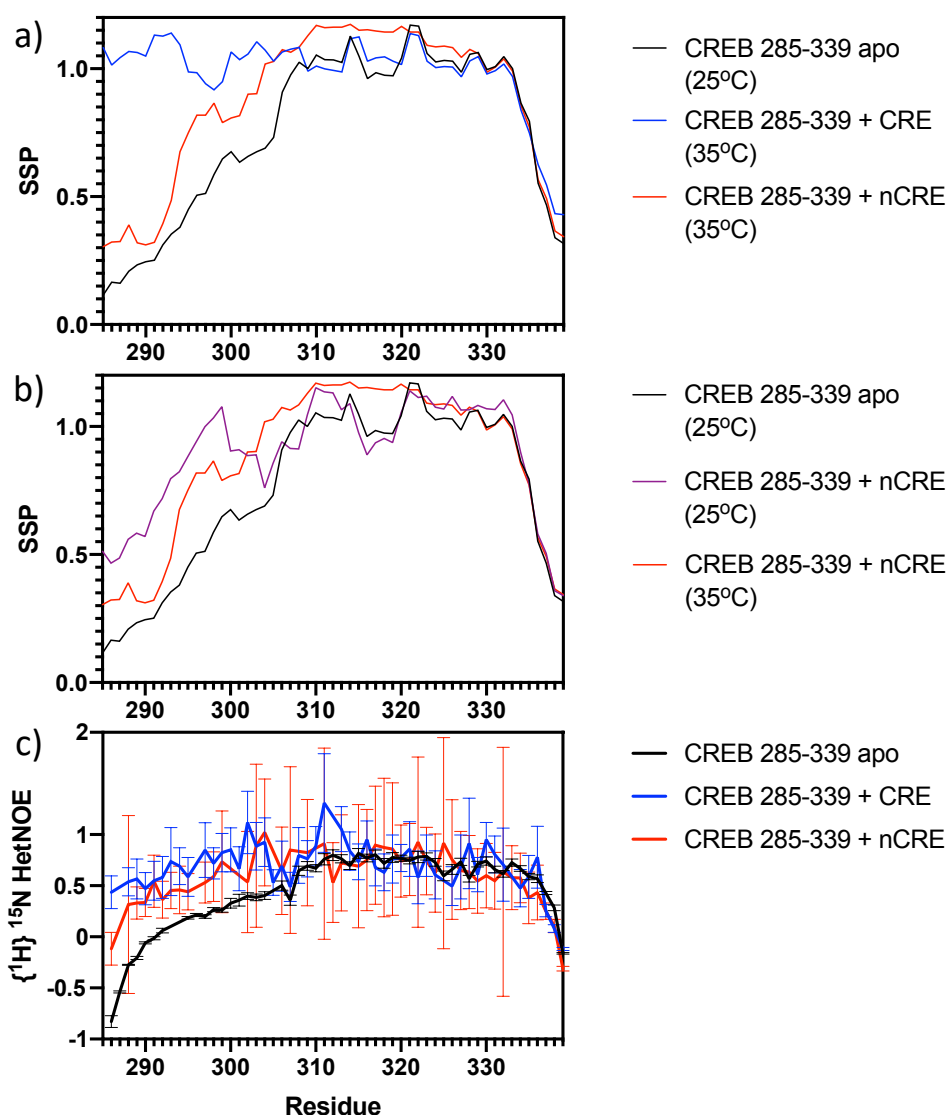


Figure 4.20: CREB 285-339 is more helical and less dynamic when in complex with DNA. a) SSP of CREB 285-339 in the apo, CRE and non-CRE DNA complexes, with the DNA complexes using 35°C chemical shift assignments of C α and C β only. b) Similar to a), but with assignments of the non-CRE DNA complex at the two temperatures under which backbone chemical shifts were assigned. c) $\{^1\text{H}\}$ - ^{15}N Heteronuclear NOE of CREB 285-339 in the apo, CRE- and non-CRE DNA complexes performed at 25°C at 600 MHz.

DNA showed near 100% helical character, aside from a slight decrease to $\sim 75\%$ helical propensity in the region of residues 296-300 and the extreme C-terminus (Figure 4.20a). Contrast this to the structural propensity of the non-CRE DNA complex, which showed lower helical propensity than the CRE DNA complex, but more than the apo-state. This intermediate behaviour was consistent with an SSP calculation using chemical shifts assigned at the two temperatures (Figure 4.20b), with the SSP on average being lower at 35°C than 25°C across residues 285-304 only. The secondary structure of the leucine zipper remains consistently helical whether bound to DNA or not.

The disordered character of the non-CRE DNA complex was confirmed when a $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE was performed on the CREB 285-339/DNA complexes. In the non-CRE DNA complex, the heteronuclear NOE experiment for the residues in the basic region indicated that they are dynamic in this state, but not as dynamic as the apo protein. The errors for these experiments were large, particularly for the leucine zipper residues (307-339).

To begin to determine what the mechanism for the specificity that CREB has for the CRE DNA sequence, the arginine C δ region was focused upon (Figure 4.21). In the apo-state, the Arg C δ peaks are overlapping one another, indicating that the side chain of the arginine residues are disordered and sampling the same environment. In the non-CRE DNA complex, the peaks become more dispersed, however there was still a lack of resolution for individual peaks. In the CRE DNA complex however, there is a large change with the Arg C δ peaks becoming more dispersed and resolved. This helped to assign some of the Arg C δ peaks to their respective residues using the HCCH-TOCSY experiment.

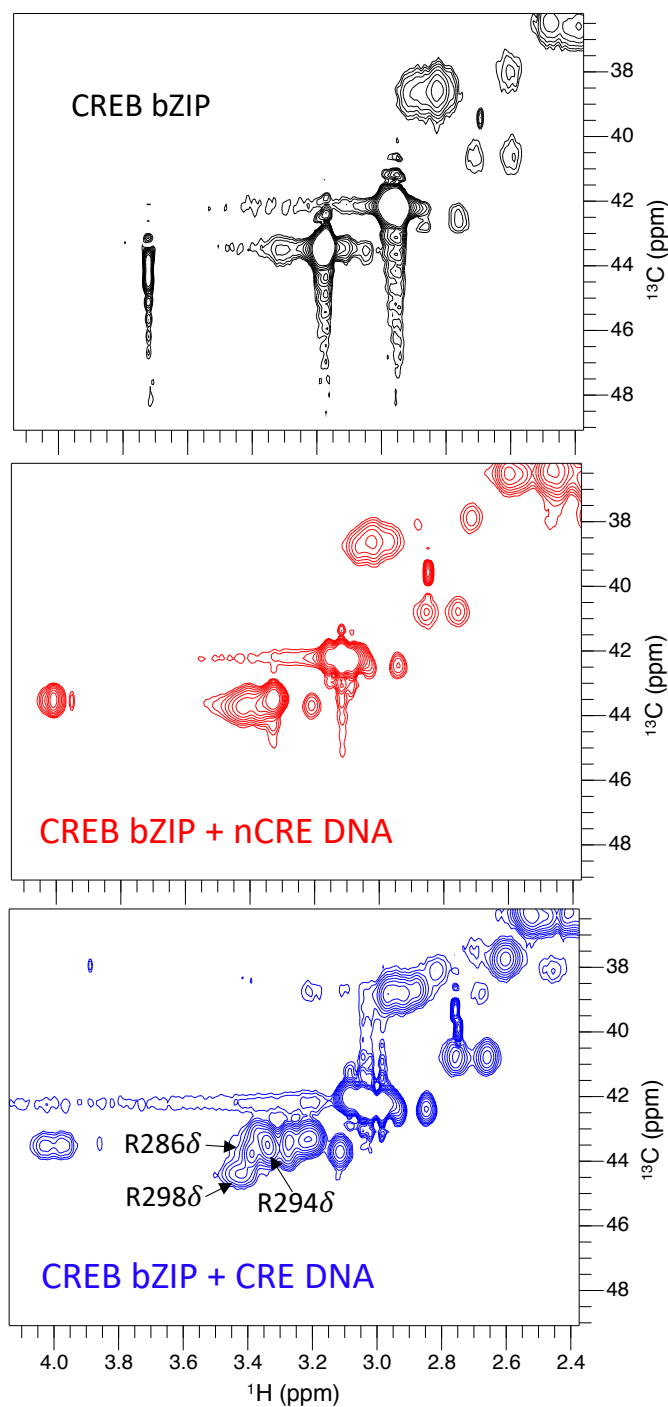


Figure 4.21: Arginine C δ region provides detail on the level of specific interaction. Arginine C δ region from the ^{13}C -HSQC of CREB 285-339 in the apo, CRE DNA and non-CRE DNA complexes shown in the top, middle and bottom panels respectively. Spectra were collected 25°C.

4.2.6 More helical CREB bZIP bound to non-target DNA has a conformational change that is similar to the target DNA complex.

In section 3.2.3 it was shown that the helical content of the basic region in the CREB bZIP domain can be significantly increased by using the longer 277-339 construct. As CREB 285-339 was completed first, the experimental conditions for the CREB/DNA complexes had already been optimised. Therefore, the various steps will not be repeated here.

In all of the CREB 277-339/DNA complexes, the assignments were completed at 35°C and SSP predictions were based on data collected at this temperature. The $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE experiments were performed at 25°C and the ^{15}N -BTROSY or HSQC were used to track peaks between 25-35°C (data not shown). The assignments for the CREB 277-339 with CRE DNA at 35°C can be seen in Figure 4.22, with the assignments at 25°C shown in the appendix Figure S16 (with the chemical shift tables for 35°C, 25°C and assignment report of 35°C in Appendix Tables S16, S17 and S18 respectively).

A titration of CREB 277-339 with CRE DNA was performed, as this was not undertaken for the 285-339 construct. The results showed that peaks do appear in slow/intermediate exchange, as broadening occurs for a significant number of residues. However, as this occurred upon the first addition of DNA and peaks shifted enough to lose their identity, limited information could be gained (Appendix Figure S17). As with the CREB 285-339 construct, the addition of CRE DNA to CREB 277-339 creates an improved dispersion of the peaks, corresponding to those assigned to the basic region. Extra peaks due to the *cis/trans* isomerism were not significantly perturbed by the addition of DNA. Alignment of CREB 285-339 with CRE DNA spectra to the CREB 277-339 with CRE-DNA reveals the vast majority of residues in a similar peak position in the ^{15}N -BTROSY/HSQC spectra (data not shown).

Like the CREB 285-339 with non-CRE DNA, when CREB 277-339 was examined in complex with this DNA sequence a titration of DNA was used, and the peaks were found to be in fast exchange (titration data not shown). A small chemical

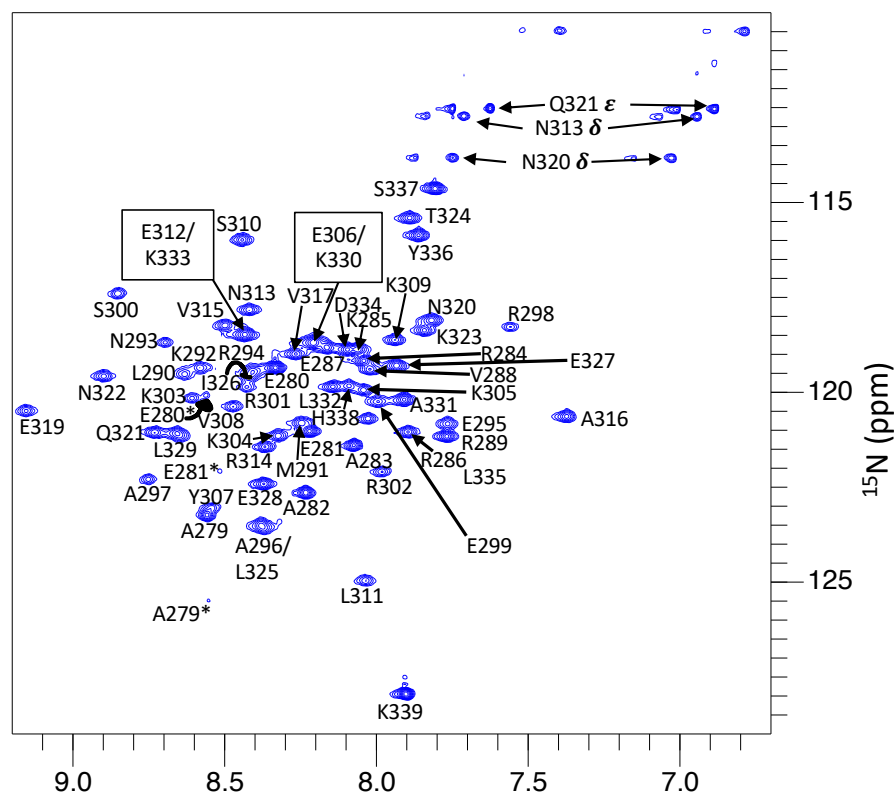


Figure 4.22: CREB 277-339 was assigned in complex with CRE DNA at 35°C. ^{15}N -BTOSY of CREB 277-339 with CRE DNA acquired at 35°C.

shift perturbation was seen once again, and triple resonance assignments of the complex were completed at 35°C in order to assign the resonances sequentially. The assigned spectrum at 35°C can be seen in Figure 4.23. The 25°C assigned spectrum (Appendix Figure S18), as well as the chemical shift tables at 35°C (Appendix Table S19) and 25°C (Appendix Table S21) and the assignment report (Appendix Table S20) can be found in the Appendix.

The SSP results with $C\alpha$ and $C\beta$ were used identify a CRE DNA bound complex that is helical in the basic region residues (277-307), with only the extreme N-terminus decreasing to a SSP score of ~ 0.55 (Figure 4.24a). The non-CRE DNA complex, using the same set of chemical shifts, appears to contain almost as much helical content as the CRE DNA sample. There is a significant difference in the helical propensity of the non-CRE DNA complex only, with this becoming intermediate between the apo and CRE DNA bound complexes. A focus on the

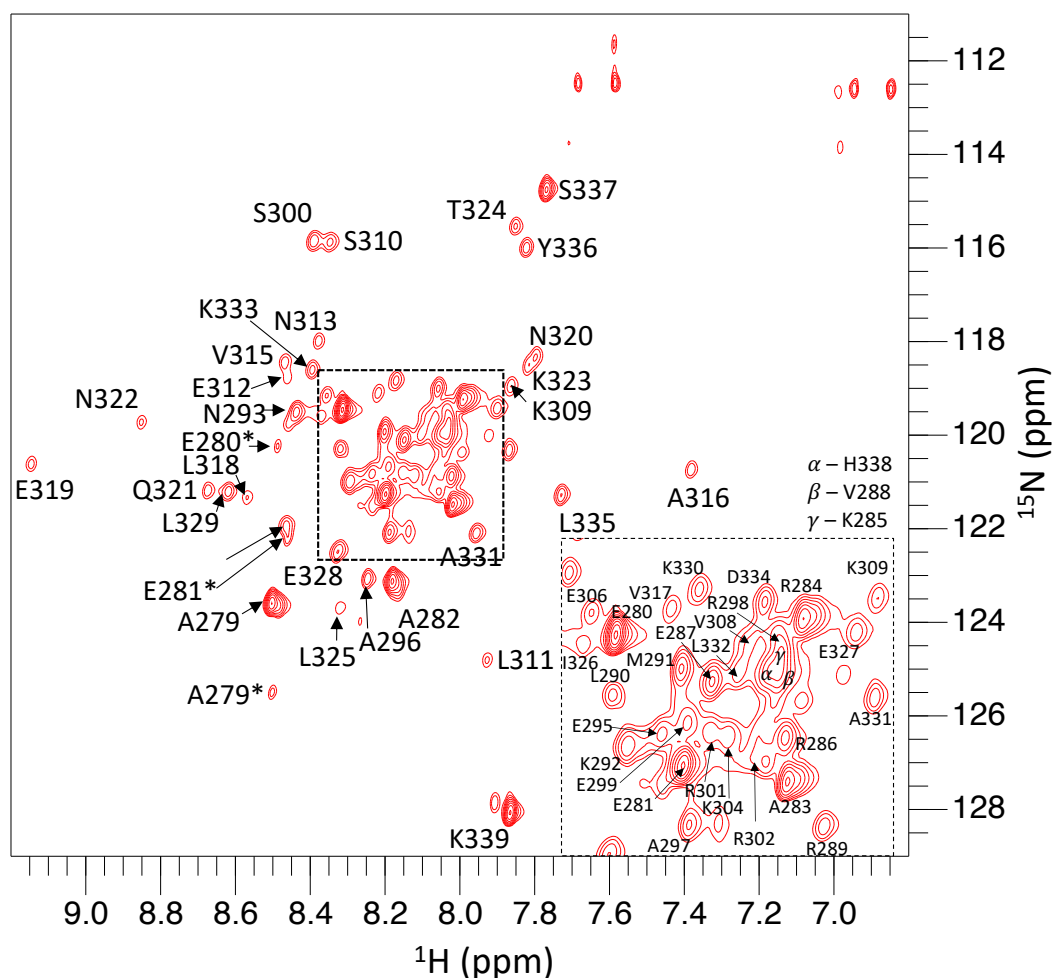


Figure 4.23: CREB 277-339 was assigned in complex with non-CRE DNA at 35°C. ^{15}N -BTOSY of CREB 277-339 with non-CRE DNA acquired at 35°C.

arginine C δ region of a ^{13}C -HSQC reveals significantly more dispersion for the non-CRE DNA complex compared the apo complex (Appendix Figure S19).

The dynamics of CREB 277-339 in each complex was measured by $\{^1\text{H}\}$ - ^{15}N Heteronuclear NOE (Figure 4.24b). This identified that the CRE DNA bound protein was the most ordered of the three complexes measured, with the apo-state containing order in the leucine zipper and disorder in the basic region. In the CREB 277-339 with non-CRE DNA complex, the basic region appears slightly more dynamic than the CRE DNA bound protein, with the difference not being statistically significant.

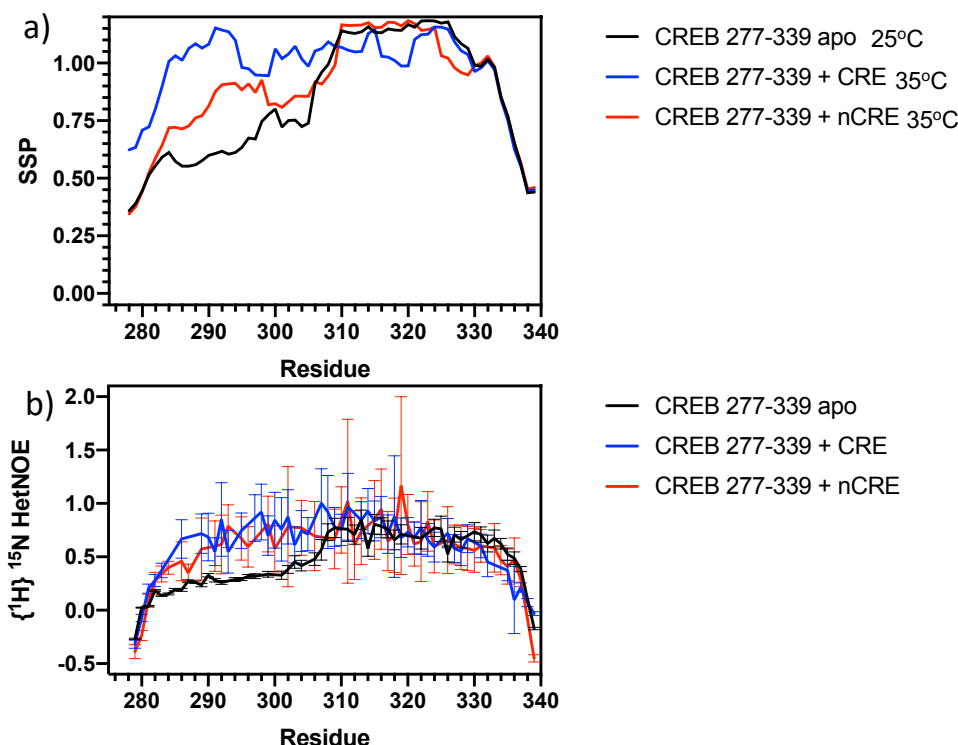


Figure 4.24: The N-terminal extended construct still retains a disordered and dynamic basic region in complex with non-target DNA. a) SSP of CREB 277-339 in complex with CRE and non-CRE DNA. $C\alpha$, $C\beta$ and $H\alpha$ were used in the calculation. With DNA, chemical shifts at 35°C were used, apo at 25°C. b) Same as in a), but with $C\alpha$ and $C\beta$ chemical shifts only. c) $\{^1\text{H}\}$ - ^{15}N Heteronuclear NOE of CREB 277-339 in the apo, CRE- and non-CRE bound states. All were performed at 25°C at 600 MHz.

4.3 Discussion

One of the key functions of a transcription factor is to be able to recognise its cognate target DNA site in a specific manner, amongst the several order of magnitude greater number of non-target sites that are present in the genome. It has long been known that transcription factors are able to associate to a non-target site¹⁸⁹, and it was postulated that this acts as a level of regulation for the activation of target genes. CREB was able to recognise its target site, CRE (somatostatin CRE: 5'-TGACGTCA-3') specifically over a non-CRE containing DNA sequence at a biophysically relevant ionic strength ($I = 0.185\text{ M}$) by a factor of ~ 10 . This specificity was consistent across a CREB 285-339 pWT, 285-339 forced dimer and also a 277-339 forced dimer construct, with work previously done for GCN4¹²⁸.

The data acquired by Hollenbeck *et al* revealed a similar affinity of GCN4 for full CRE and CRE half-site¹²⁸. EMSA, which was used by Hollenbeck *et al* for GCN4 experiments are not at complete equilibrium, as the populations of bound and free DNA may change as the complex migrates through the gel, and modern methods for K_d determination are available now which alleviate this issue. The CRE-half site was not investigated in this study.

In the context of the shorter bZIP pWT protein sequence, the apparent K_d to the CRE DNA was an order of magnitude higher for the dimerisation-competent pWT rather than the forced dimer version. Thus a large increase in stability of the homodimer in the forced dimer increases the apparent binding specificity for the target DNA. The result was not seen for GCN4 as the binding affinity of a native and forced dimer construct, as determined by fluorescence intensity change of labelled DNA, was consistent within error⁸⁰. The ionic strength of the buffer used by Cranz *et al* was only 68 mM, nearly three times less than the buffer used in this work and it should increase the protein-DNA binding affinity. The order of magnitude difference in K_d for CRE DNA between pWT and forced dimer was not replicated with non-CRE DNA, with this difference not down to the number of binding sites when looking at the ITC data. In the concentration range of this binding transition, the protein is expected to be almost fully dimeric. Interestingly, use of a labelled ATF-2 bZIP construct and titrating in DNA showed an increase in anisotropy until a point, where a sharp decrease in anisotropy was observed¹⁹⁰, indicating a possible mechanism whereby the dimer dissociates whilst bound to DNA if excess DNA is present. This potentially has implications in the target search pathway for CREB but was not investigated in this work.

When looking at the thermodynamics of the CREB-DNA complexes by ITC, the large enthalpy change observed for both the CRE and non-CRE DNA constructs provided the major contributor to the free energy of each complex. Whilst not examined with CREB, ITC with the Fos/Jun heterodimer in complex with both TRE (5'-TGA(C/G)TCA-3') and CRE DNA sequences showed that the enthalpy and entropy compensated after changes of temperature to maintain the free energy

of the complex¹²³. Truncating the leucine zipper one heptad repeat at a time for the Jun bZIP was also shown to successively compromise the binding energetics to target TRE DNA, to the point that a monomeric only Jun construct showed no heat upon injection¹²⁴. Also, the enthalpy change was similar for the Jun homodimer as it was for CREB homodimer at ~ 30 kcal mol⁻¹, indicating that the enthalpy change of bZIPs associating to their target DNA may be relatively consistent¹²⁴. Ionic strengths were in the physiologically relevant range for both experiments (CREB $I = 0.185$ M, Jun $I = 0.252$ M). A difference in binding affinity was reported for the same constructs using fluorescence anisotropy/intensity and the unlabelled constructs in ITC, with roughly a order of magnitude tighter for the interaction of CREB 285-339 to AlexaFluor488-CRE by fluorescence when compared to unlabelled CRE DNA with ITC. The change in fluorescence intensity of free AlexaFluor488 when titrated with CREB bZIP indicates that a tighter interaction with Alexa labelled DNA is possible, due to the Alexa dye being negatively charged. However, neither technique is ideal for determining K_d values in this range, and systematic errors may be significant. The effect of the homodimer K_d being within the range of the DNA binding affinity also meant that, for the fluorimeter measurements in particular, the $K_{d, \text{app}}$ was only able to be determined. Fitting the data to a model assuming the full monomer/dimer pathway directions was not possible due to poor signal to noise, therefore an assumption was taken into account for the protein to remain dimeric throughout the concentration range.

Limited kinetic experiments have been performed with bZIP domains associating to target DNA, with all of them based upon target or half-target sites only. GCN4, Fos/Jun and ATF-2 all showed k_{on} values that are very high (i.e. $\geq 10^8$ M⁻¹ s⁻¹)^{80,81,134}, which are two orders of magnitude higher than the k_{on} for protein-protein interactions assuming no electrostatic rate enhancement ($k_{\text{on}} \sim 10^5$ - 10^6 M⁻¹ s⁻¹) but close to the Debye-Smoluchowski limit of 10^9 M⁻¹ s⁻¹ which describes the rate limit for perfectly reactive spheres¹³⁷. The k_{on} values for CREB were shown to be 2×10^9 M⁻¹ s⁻¹, and are of a similar order of magnitude to the previous work on bZIPs. The LacI system has been intensively studied and showed a k_{on} that was on

order of $10^9 \text{ M}^{-1} \text{ s}^{-1}$ or greater¹³⁷. CREB was found to associate to target containing CRE DNA with a second order rate constant determined to be beyond the diffusion limit, similar to LacI. Therefore, the electrostatics aid in facilitating the diffusion of CREB to CRE DNA, which is not surprising given the relative net charges of the two reactants of -28 and +7.5 for DNA and CREB 277-339 respectively (CREB full length has a charge of +8.5) Similar k_{on} values were observed for two different CREB bZIP constructs, which have varying affinities for target site DNA.

The difference in binding affinity for CREB bZIP to target-containing DNA sequences is therefore a result of the dissociation rate constant, which varied ~ 1.5 fold more than the k_{on} . In the LacI system, the dissociation rate becomes higher at higher ionic strengths, indicating the role of a predominantly electrostatic interaction becoming increasingly perturbed by the addition of charged ions¹³⁸. The k_{off} *in vitro* for CREB 285-339 pWT from CRE DNA by stopped flow was identical within error to that obtained for the long CRE residence time that Sugo *et al* determined by single molecule fluorescence of a Halo-tagged full length CREB *in vivo*. However, the k_{off} experiments were not ideal as the concentration of competitor was not high enough to reliably extract the off rate itself. hence the use of the single exponential for fitting the data which is a oversimplification of the model. Increasing the concentration of competitor DNA would be required. However, the k_{off} values determined here can be used as an upper limit on the off rate. Therefore, the experimental set-up *in vitro* can be a suitable representation of the *in vivo* conditions, even though it is a very simple system compared to that in the cellular environment (further experiments in Chapter 6 that try to replicate those conditions better).

The basic region of CREB underwent coupled folding upon binding upon association to DNA, with CRE DNA forming complete α -helix and non-CRE showing a partial increase. This consistent behaviour has been measured with CD using GCN4, Fos/Jun, and CREB previously^{79,123,127,128,131}, with non-target DNA complexes resulting in an intermediate helical propensity between apo and target bound complexes. This sequence specific coupled folding upon binding has also been seen in p53¹⁹¹. Mutations within the L1 loop of p53 that altered the

population levels of various conformational switch structures led to increased off rates, as determined by fluorescence anisotropy, suggesting that the disordered loop region regulates k_{off} from the DNA.

One major limitation in the work presented here is the lack of kinetics of reliable CREB association/dissociation to non-target containing DNA. The temperature dependence that was attempted did not predict the theoretical k_{off} of $\sim 100 \text{ s}^{-1}$ at 25°C , and was 6-fold higher. This was due to a number of reasons, the k_{off} is high and is difficult to measure reliably, and a low amplitude due to the requirement of fluorescence anisotropy to measure the change in fluorescence. Previous work with bZIPs did not determine the k_{on} or k_{off} with non-target DNA sequences *in vitro*, possibly due to the lack of a fluorescence intensity change that was seen for CREB in both equilibrium and kinetics experiments. Labelling the bZIP within the basic region/N-terminus would increase the likelihood that a fluorescent intensity change would be observed upon DNA binding. Stopped flow anisotropy was attempted as discussed in the Results, however the signal to noise was significantly lower. One way around this would be to use either increased concentration of fluorescently labelled DNA or using a 2-4 fold increase in the number of mixing events to try to improve the signal, as these combined could lead to the observation of the association or dissociation reaction.

One of the possible explanations for the inherent difficulty in obtaining the kinetic rates of CREB with non-target DNA was answered by NMR. With CREB 285/277-339 constructs, the complex with CRE DNA was in slow/intermediate exchange on the μs -ms timescales. This is consistent with the high affinity, coupled folding upon binding interaction seen here and elsewhere^{90,97,108,113,127}, as the exchange regime seen in NMR is related to the k_{on} plus k_{off} ¹⁶². The non-CRE DNA titration with CREB bZIP by NMR was initially expected to reveal a similar slow/intermediate exchange regime, by virtue of the relatively high affinity of the non-CRE DNA interaction, however a fast exchange regime was observed. This observation has been reported by Loth *et al* with the lac repressor headpiece¹⁹².

A difficulty with the analysis of CREB with non-CRE DNA was the low chemical shift dispersion and a significant overlap of peaks in the ^{15}N -HSQC/BTOSY experiments during the titration. This then did not allow CSP or line-widths to be measured. These results have been used to good effect with other systems however. Zandarashvili *et al* looked at a zinc finger transcription factor associated to non-target DNA, with the chemical shift perturbation of the zinc finger 1 region showing intermediate chemical shift perturbations between the apo- and target-DNA bound states¹⁴⁸. This non-target binding mode remained dynamic over the specific complex. NMR analysis of the HoxD9 homeodomain also show a highly mobile non-target DNA interaction using paramagnetic relaxation enhancement, which revealed significant line broadening at the extreme N-terminus and at the N-terminus of helix 2, indicating that these two regions are highly mobile in the non-specific interaction¹⁹³. Also, the lac repressor headpiece was investigated with multiple non-target DNA sequences which identified a sequence dependence between each when looking at complexes via ^{15}N -HSQC line-width and chemical shift perturbation analysis¹⁹².

Little to no evidence exists for a structural description of the search mode for a disordered transcription factor. The results for CREB suggest that the basic region remains partially dynamic (not completely helical) upon interaction with non-target DNA. For both CREB 285-339 and 277-339 constructs, the complex with CRE DNA resulted in complete helical formation of the basic region, consistent with that seen in the crystal structure²¹. The Arg C δ region showed a large chemical shift dispersion in the CRE DNA complex, indicative of the stable interactions of the arginine sidechains to the DNA. Some of these particular peaks were assigned in the 285-339 construct to be R286, R298 and R294, which are also the residues in the crystal structure that interact with the phosphate backbone as a non-specific interaction (Figure 4.25), however the other arginine residues were too difficult to assign with any confidence. R289 and R314 H δ /C δ were not assigned in the CRE DNA complex, which is interesting as these are not in an interaction interface and therefore the

dynamics of the side chain may be high enough to not be able to do assignment of these. These Arg resonances overlaid with the 277-339 construct with CRE DNA.

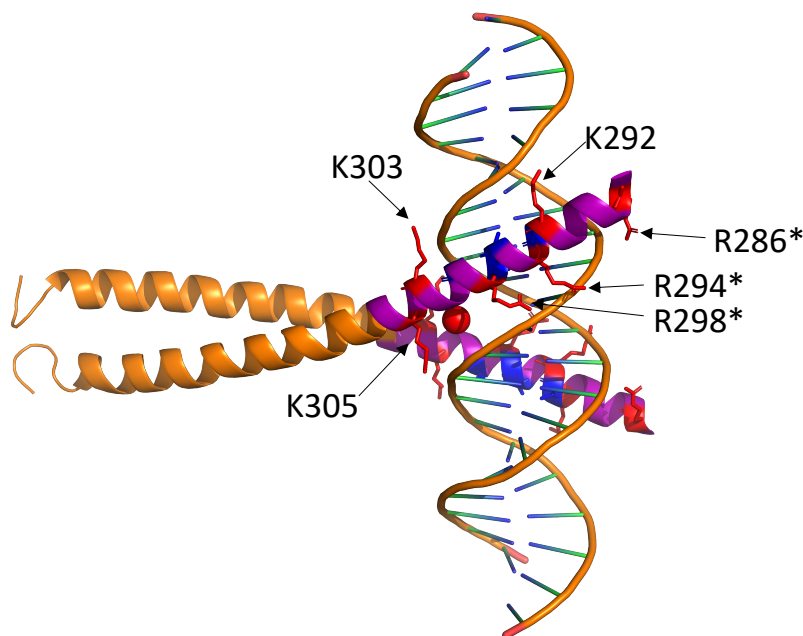


Figure 4.25: Location of the assigned arginine $C\delta$ peaks in CREB 285-339 with CRE DNA complex. Crystal structure of CREB 285-339 in complex with CRE DNA²¹ with the residues creating the non-specific contacts highlighted and annotated. * indicates assigned $C\delta$ shifts from the NMR experiments. The leucine zipper is orange, basic region is purple, specific contacts on protein are blue and non-specific are red. Mg^{2+} is shown as a red sphere. PDB = 1DH3.

Whilst the amide backbone resonance provides useful information in the specificity of protein/DNA complexes, it is sensitive to changes in buffer composition, temperature and pH as well as secondary structure. For instance, the 285-339 vs 277-339 with non-target DNA SSP revealed that, although the H^N and N^H chemical shifts were quite similar, the secondary structure as determined from backbone carbon and proton chemical shifts revealed quite different behaviour. The heteronuclear NOE showed that the dynamics of the non-specific complex complemented that of the SSP prediction quite remarkably well, with a lower heteronuclear NOE ratio for the non-CRE DNA complexes compared to the CRE DNA, albeit a smaller effect for the 277-339 construct. This behaviour indicates that the non-CRE DNA/285-339 complex is more dynamic in the ps-ns timescales, and that a more disordered structure of the basic region in this construct is the result. This can be compared

to a study by Zandarashvili *et al*, were mutations of the zinc finger protein Egr-1 resulted in a push of the conformational ensemble towards a "target search" mode, which exhibited a lower R_1 relaxation value than in a recognition mode¹⁵¹. The longitudinal relaxation constant is equivalent to the ^{15}N heteronuclear NOE used for CREB in terms of the timescale that is being investigated.

In agreement with this, a focus on the Arg C δ region of the ^{13}C -HSQC revealed that a minimal chemical shift dispersion for CREB 285-339 with non-CRE DNA compared to apo, whereas the CREB 277-339 with non-CRE DNA showed a large chemical shift perturbation in this region, more similar to that in the CRE DNA complex. It stands to reason that the more helical the bZIP is in the apo-state, the increased likelihood that the non-target DNA interaction is more stable as energetically there is less conformational change to occur. This effect has been used as a potential design area for bZIP inhibitors, whereby GCN4 basic region was covalently stapled to force helical content across one or two turns of an α -helix¹⁹⁴ showed increased affinity to target DNA over WT. A tunable version which uses unnatural amino acids which chelate to metal ions which act to stabilise i to $i + 4$ or $i + 7$ ¹⁹⁵ showed a similar effect.

5

Investigations into a monomer only model reveals a lack of target DNA specificity

Contents

5.1	Introduction	152
5.2	Results	154
5.2.1	Design for the monomer construct	154
5.2.2	Binding of monomer model to DNA	155
5.2.3	Structural characterisation of the monomer	156
5.2.4	non-CRE DNA and monomeric CREB	163
5.2.5	Monomer model with target DNA	171
5.2.6	Decoupling of the leucine zipper from the basic region	178
5.3	Discussion	183

5.1 Introduction

With the controversy over the binding specificity of monomeric models of bZIP proteins and an inherent lack of information for the structural propensity of the protein/DNA complexes, further investigation was required. A lot of work has been previously completed for monomeric bZIP models, with a summary of the binding affinities to various DNA sequences shown in Table 5.1. However, CREB is not well understood in this regard as to whether the dimer or monomer pathway

is preferred (Figure 5.1). Hence, a monomeric model needs to be designed and analysed using the same approaches as used for the bZIP protein in Chapter 4. This will enable the determination of which pathway CREB uses in order to associate to target containing DNA.

Table 5.1: Binding affinities for various monomeric bZIP transcription factors to target DNA. Errors are SEM. * indicates a forced dimer version of basic region construct.

Construct	DNA	K_d (nM)	Reference
Fos/Jun CREB	CRE	Selectively over SP-1	[78]
CREB FL	CRE	No binding Monomer or dimer protects site during DNase	[135]
v-Jun	AP-1		[121]
GCN4*	AP-1	Binding EMSA No binding EMSA	[79]
GCN4	AP-1 ATF-CREB	50000 ± 2500 70000 ± 2000	[125]
GCN4	CRE	406 ± 32	[80]
GCN4	CRE	1700 ± 100	[81]

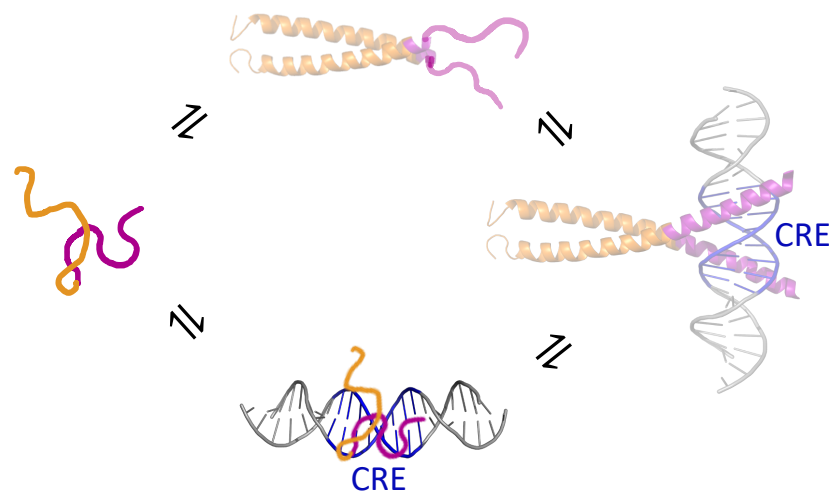


Figure 5.1: Simplified model of the monomer association pathway of CREB bZIP to DNA in the context of both of the possible pathways. The transparent parts of the model indicate the regions that are not focused on in this Chapter. Structure used was PDB: 1DH3.

5.2 Results

5.2.1 Design for the monomer model construct

When considering both CREB constructs (N-terminus starting at 277 and 285), the helical propensity of the basic region residues is maintained when the construct length is terminated at V315. The removal of residues sequentially from this position towards the N-terminus is predicted to result in less overall helicity (Figure 5.2). V315 is several residues beyond the estimated interface between the basic and leucine zipper regions at E306/Y307 by BLAST-p. Therefore, the monomer models that will be used in the remainder of this chapter are 285-315 and 277-315.

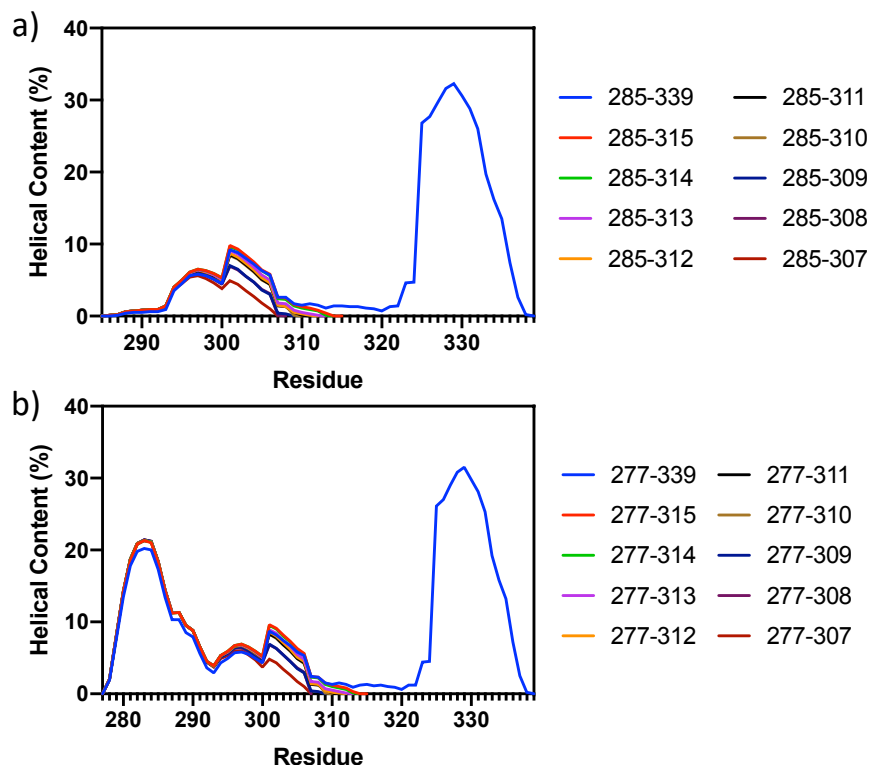


Figure 5.2: Design of monomeric CREB using Agadir for helical propensity predictions. a) Helical propensity of C-terminal truncations of CREB bZIP residues 285-339. b) Similar to a), for the CREB 277-339 construct. Constructs were analysed on the Agadir webserver¹⁷⁶.

5.2.2 Basic region only cannot discriminate between target and non-target DNA sequences

As the dimeric form of CREB bZIP domain appears to be able to discriminate between target and non-target containing DNA sequences, the first step was to determine whether this specificity was carried over to the monomer model protein. Fluorescence anisotropy of AlexaFluor488-labelled DNA titrated with CREB 285-315 and 277-315 was utilised, as no change in fluorescence intensity of the labelled DNA was observed upon complex formation (Figure 5.3). A fit to a one-site binding event showed a very weak affinity for the monomeric model proteins to DNA, at least three orders of magnitude lower to the dimeric constructs. A lack of significant specificity was seen for CREB 285-315 for CRE over the non-CRE DNA sequences, however the CREB 277-315 construct showed a even weaker affinity in general, but with an two-fold increase in specificity for CRE DNA.

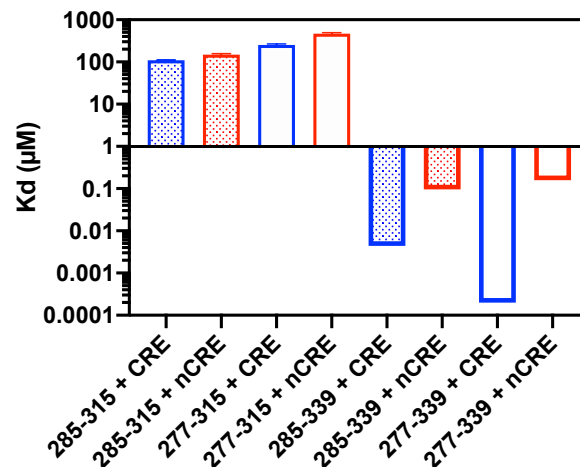


Figure 5.3: A weak, non-specific interaction was seen with monomeric CREB to target DNA. K_d of monomeric CREB (285-315 or 277-315) when titrated against fluorescently labelled CRE- and non-CRE DNA (both NMR versions). Errors are SEM. 277/285-339 FD data was presented in Figure 4.5. Data collected by Wilson Bridge (285-315) and Dr M. Kuravskiy (277-315).

5.2.3 Structural characterisation of the apo-state of the monomer

In section 5.2.2, the binding affinity for the basic region only construct for both CRE-containing and non-CRE DNA was shown to be quite low, indicating a weak interaction. To understand why this was the case, a structural study of the basic region in the absence and presence of DNA was performed.

CREB 285-315 was expressed with ^{13}C and ^{15}N isotopes and assigned using the sequential assignment strategy with triple resonance experiments, as previously described in section 3.2.3. The ^{15}N -HSQC identified the correct number of peaks, which were assigned to all of the residues in the CREB 285-315 construct, aside from the extreme N-terminal Gly and Lys residues (Figure 5.4 and Table 5.2). A lack of dispersion of the peaks is seen in the ^{15}N -HSQC of the CREB 285-315 construct, with all of the peaks in Figure 5.4 falling between 8.0-8.5 ppm on the ^1H axis.

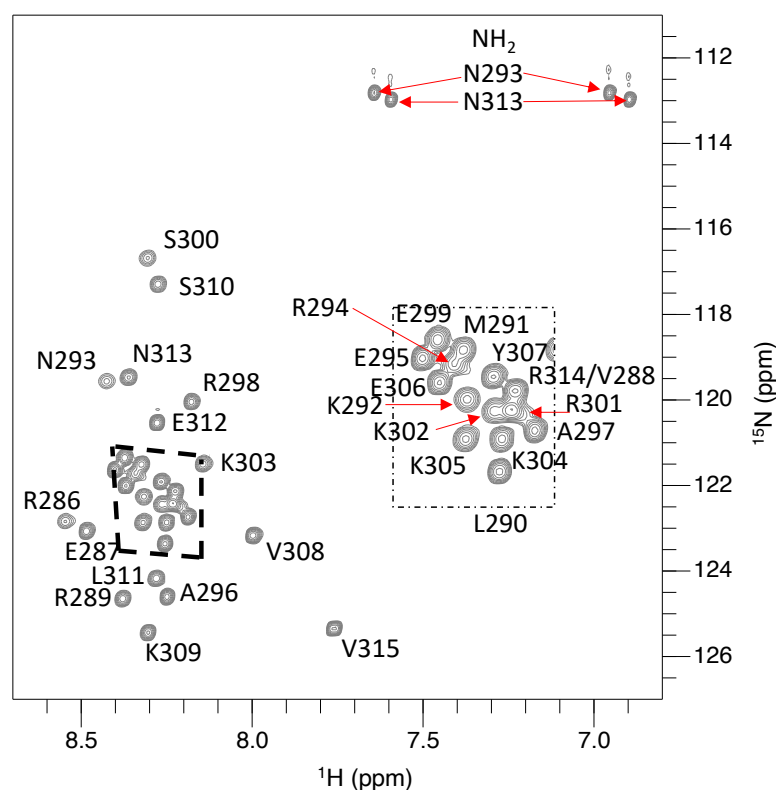


Figure 5.4: Assignment of CREB 285-315 shows disordered characteristics. ^{15}N -HSQC of CREB 285-315 with annotated peak assignments at 25°C. Inset - larger scale of the crowded region for ease of seeing peak labels.

Table 5.2: Assignment report for CREB 285-315 at 25°C.

Category	% Assigned
All Residues	96.875
Type C	96.875
Type CA	96.875
Type CB	96.7741935
Type CD	73.9130435
Type CE	37.5
Type CG	90
Type CZ	0
Type H	96.7741935
Type HA	93.9393939
Type HB	98.245614
Type HD	97.1428571
Type HE	61.9047619
Type HG	95.6521739
Type N	93.75
Type ND	0
Type NE	0
Type NH	0
Type NZ	0

The process of assignment for this particular construct was significantly easier than the bZIP, due in part to the smaller monomeric protein yielding better signal to noise. An example of this can be seen in Figure 5.5, where CC(CO)NH and H(CCCO)NH experiments were utilised as part of the triple resonance experiment set to assign more of the side chain resonances of each $i-1$ residue where possible (Appendix Table S22). These experiments provided more conclusive evidence for the identity of a particular resonance in the ^{15}N -HSQC.

For one of the subsequent CREB 285-315/DNA complexes (Section 5.2.5), a higher temperature was used for assignment. A temperature titration was used to track the chemical shift changes between 9°C and 35°C (Figure 5.6). A linear dependence on the temperature was observed between the minimum and maximum temperatures, which was used for tracking the assigned resonances from 25°C to 35°C. Chemical shifts assigned at 35°C can be seen in the Appendix Table S23.

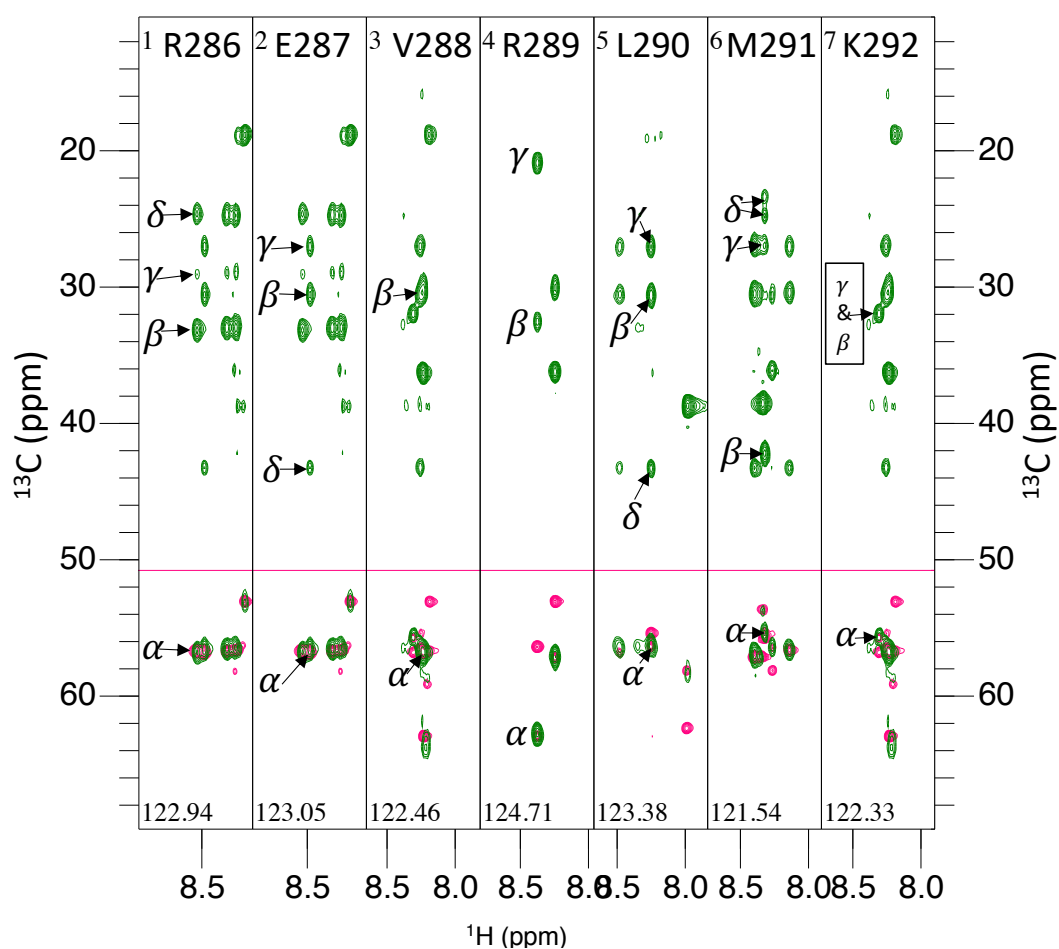


Figure 5.5: Assignment of side chain carbon chemical shifts with CC(CO)NH on CREB 285-315. Representative strip plot showing the assignments between 286-292 of residue $i-1$ in each corresponding strip. Green = CC(CO)NH, pink = HNCA.

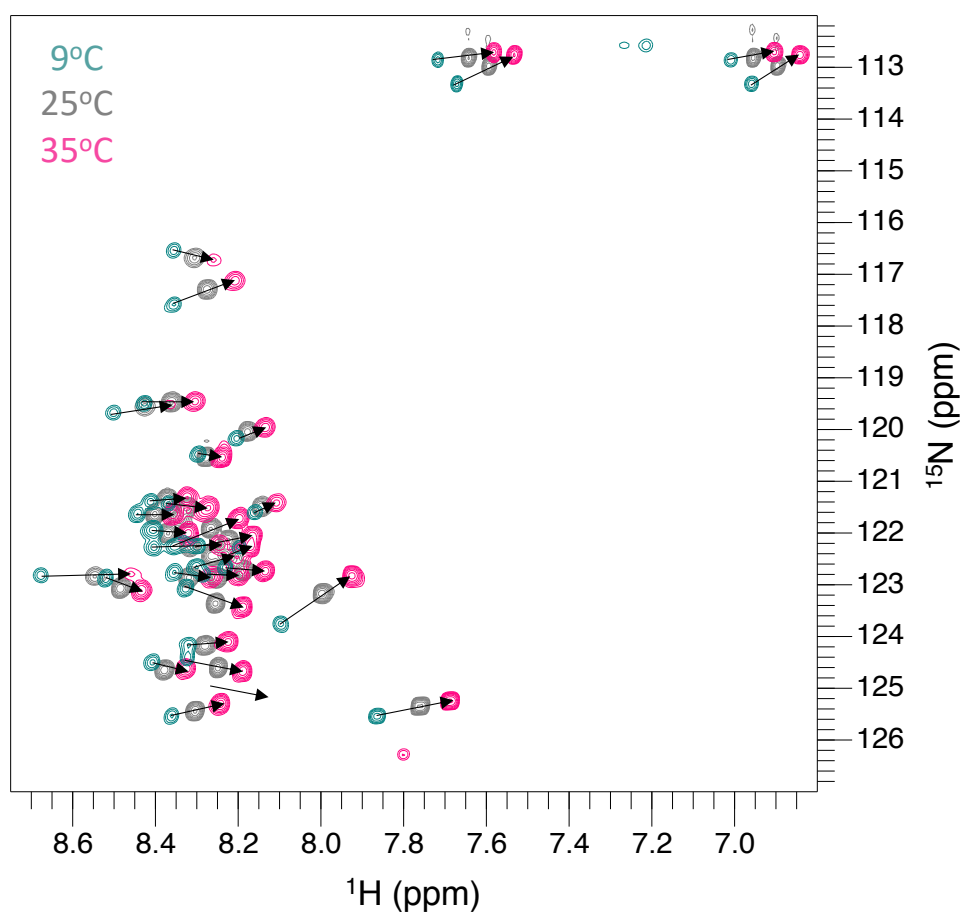


Figure 5.6: Temperature titration of CREB 285-315 by NMR for assignments at the higher temperature. ^{15}N -HSQC's were performed at 9, 25 and 35°C on CREB 285-315. Peak shifts due to temperature are notated with an arrow.

The N-terminal extended version of the monomeric model, CREB 277-315, was also assigned following an identical strategy to CREB 285-315, with the exception of a missing assignment for R286. Similarly to CREB 277-339, three extra peaks appeared in the ^{15}N -HSQC, corresponding to the *cis*-proline isomer of residues 279-281 (Figure 5.7, Appendix Tables S24 and S25). While expected to contain more helicity in the apo-state than the 285-315 construct, the lack of dispersion in the CREB 277-315 spectra indicates that the protein remains disordered and possibly only a small increase in helical propensity is seen. As the apo state of the monomer models have been assigned, the secondary structure propensity and dynamics of these proteins could be investigated.

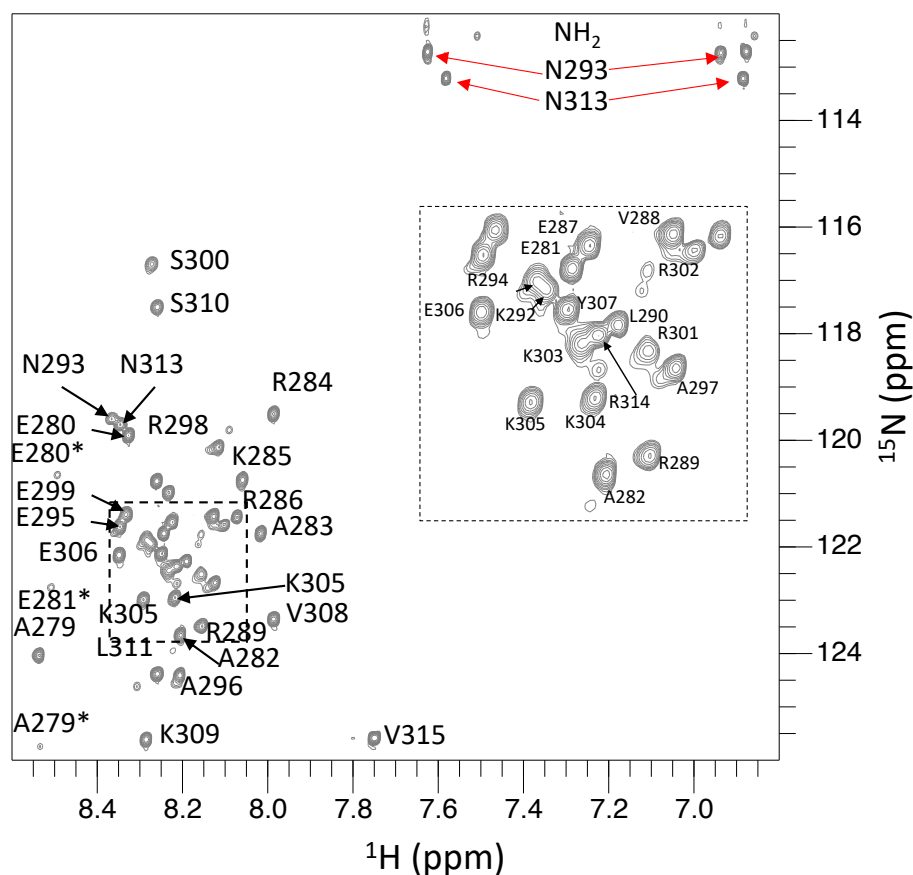


Figure 5.7: Assigned ^{15}N -HSQC of CREB 277-315. ^{15}N -HSQC was collected at 25°C and assignments were made at this temperature. * indicates peaks assignments due to the close proximity of the *cis*-proline (P278).

The ^1H - ^{15}N Heteronuclear NOE experiment was performed on the monomeric CREB model proteins. For the shorter length CREB 285-315, this was performed at two temperatures (25°C and 35°C, Figure 5.8a). Increasing the temperature resulted in a more dynamic peptide backbone overall, but the general trend of the extreme termini being more dynamic than the centre of the protein sequence remained, indicating that the central region of the CREB 285-315 exhibits slightly more restricted dynamics.

Comparing the CREB 277-315 to CREB 285-315 constructs, the ^1H - ^{15}N Heteronuclear NOE reveals that, similarly to the N-terminal extension in the context of the bZIP (see Section 3.2.3), that CREB 277-315 is less disordered than the 285-315 construct (Figure 5.8b). In particular, the region of CREB that is known to interact with target containing DNA, R286-K305, is significantly less disordered in the N-terminally extended construct.

Secondary structure propensity was calculated for the monomeric constructs (Figure 5.8c). The SSP agreed with the decrease in dynamics of the CREB 277-315 by revealing that this protein had a higher SSP score than the CREB 285-315 in general, indicating a higher amount of helical content, with this also being consistent with the CREB 277-339/285-339 comparison. What effect the DNA had on the structure of the monomer was thus the next step and is discussed in the following section.

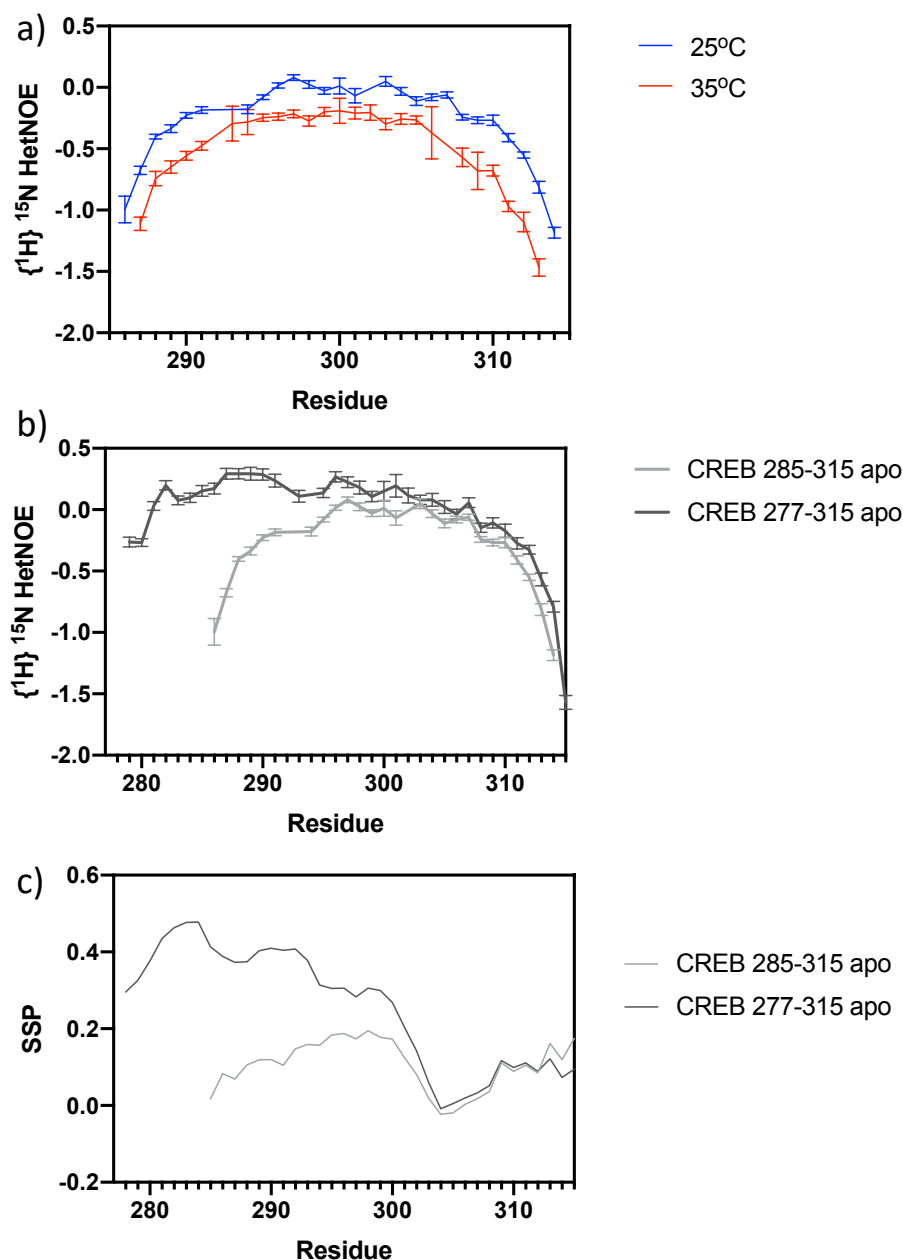


Figure 5.8: The monomer model CREB was dynamic and disordered. a) ${}^1\text{H}$ - ${}^{15}\text{N}$ Heteronuclear NOE of CREB 285-315 performed at 25°C and 35°C. b) ${}^1\text{H}$ - ${}^{15}\text{N}$ Heteronuclear NOE of CREB 285-315 and CREB 277-315 performed at 25°C. ${}^1\text{H}$ - ${}^{15}\text{N}$ Heteronuclear NOE were performed at 600 MHz. Errors are SD. c) SSP of CREB 285-315 and CREB 277-315 of data collected at 25°C. $\text{C}\alpha$ and $\text{C}\beta$ chemical shifts were used for the SSP calculation.

5.2.4 Monomeric CREB remains disordered when interacting with non-CRE containing DNA

Following the assignment of a monomeric model protein for CREB in Section 5.2.3, the next question that was addressed was whether the monomer undergoes coupled folding upon binding to the target containing CRE DNA or non-CRE DNA. The lack of specificity for either DNA sequences in Section 5.2.2 would suggest that the mode of interaction of the monomer for DNA would be similar. Therefore, NMR experiments were performed with monomer and DNA complexes to identify if this would be the case.

Addition of non-CRE DNA (sequence of forward: 5'-GGCTAAAGCATTCT-3') to CREB 285-315 was tracked with ^{15}N -HSQC's at several DNA concentrations until the point where further additions of DNA did not result in any perturbation of chemical shifts (Figure 5.9 and Appendix Table S26). This indicated that the protein (324 μM) was saturated after the addition of 1.8 mM non-CRE DNA. All of the peaks could be tracked at each DNA concentration, as the system was in a fast exchange regime. Also, the chemical shift perturbation by the addition of DNA was limited, with all the resonances in the ^1H axis remaining within 8.0-8.5 ppm. This indicated that the final conformation of the protein may remain disordered when associated to non-CRE DNA. Assignment of the saturated non-CRE complex used information from the DNA titration (Figure 5.9) and from triple resonance experiments, with the assigned spectra shown in Figure 5.10.

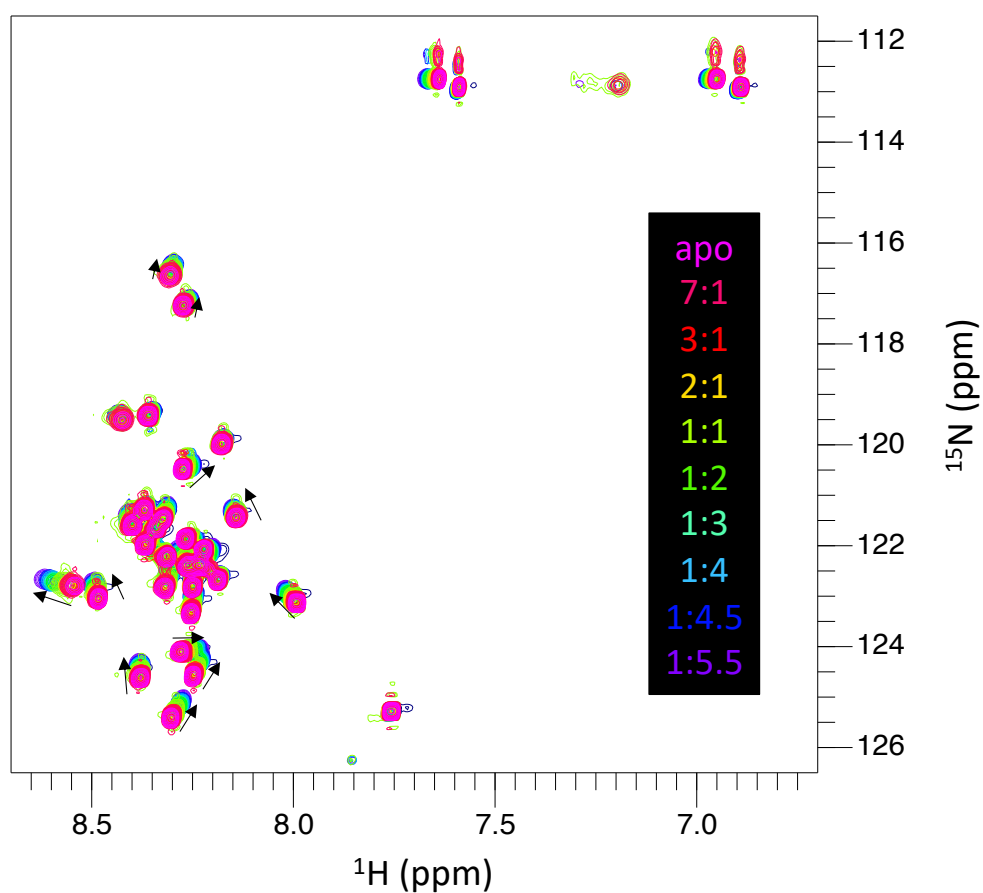


Figure 5.9: NMR assignments could be tracked from the apo to saturated complex using a non-CRE DNA titration with CREB 285-315. ^{15}N -HSQC titration of CREB 285-315 (324 μM) with non-CRE DNA (5'-GGCTAAAGCATTCT-3') at varying concentrations. Spectra are displayed as a rainbow from low [DNA] = pink to high [DNA] = purple. Arrows indicate general trajectories for a selection of peaks.

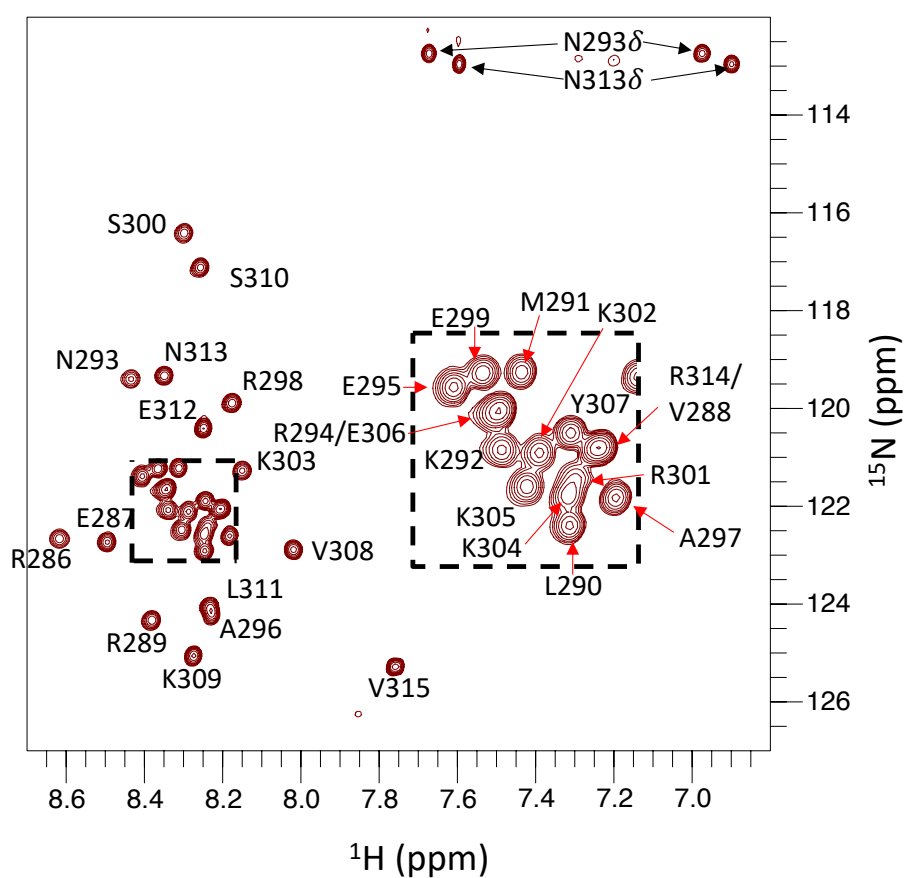


Figure 5.10: CREB 285-315 with non-CRE DNA was assigned using a combination of DNA titration and triple resonance experiments. ^{15}N -HSQC of CREB 285-315 with saturating concentration of non-CRE DNA at 25°C.

As a large number of titration points were collected, a DNA-binding curve was formed by plotting the chemical shift perturbation against [DNA], with the chemical shift perturbation in equation 2.1¹⁶² (see Appendix Table S26 for chemical shifts, Figure 5.11). The data were fitted to a one-site equilibrium binding event, with a parameter that accounted for the change in protein concentration with each dilution step and can be seen in Figure 5.11. A weighted average of each residues fitted K_d was used, excluding residues 285, 288, 294, 301, 311, 314 and 315, as these residues either were not able to be tracked during the titration or the equation did not converge during fitting. The K_d value of CREB 285-315 with non-CRE DNA was thus calculated to be $260 \pm 10 \mu\text{M}$, two-fold higher than the K_d determined using fluorescence anisotropy.

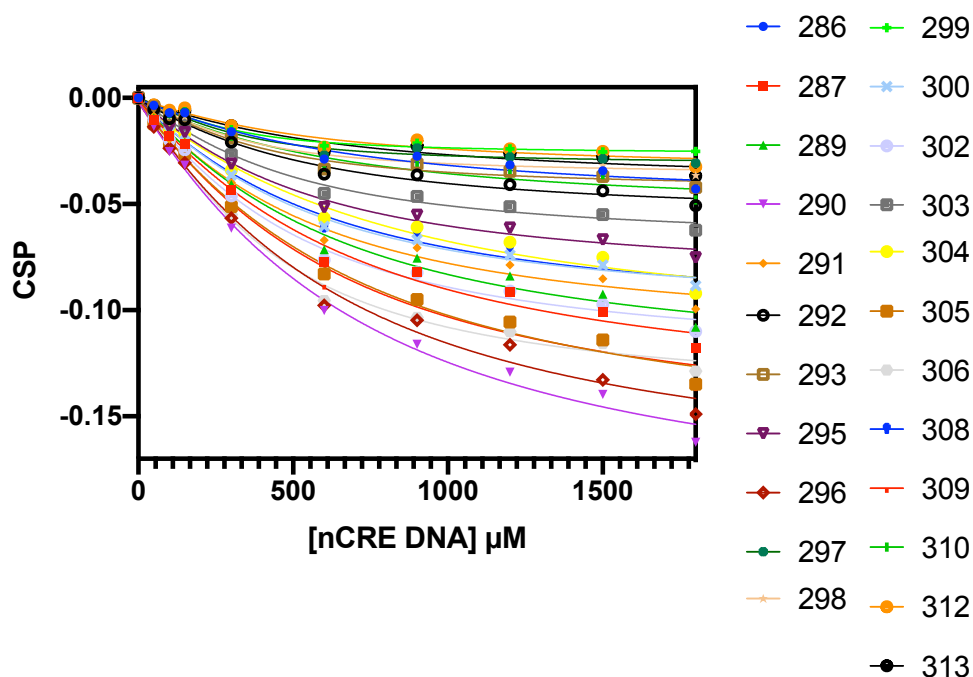


Figure 5.11: Equilibrium binding of non-CRE DNA with CREB 285-315 reveals a weak interaction. ^{15}N -HSQC of CREB 285-315 titrated against varying concentrations of non-CRE DNA. Data were collected at 25°C at 600 MHz and fit with equation 2.2.

A similar set of experiments was performed for the N-terminal extended version of the monomeric CREB model (residues 277-315) with the non-CRE DNA. A lower number of titration points were performed, therefore no DNA binding curves could be plotted for this construct. The CREB 277-315 construct was assigned

as shown in Figure 5.12 (see Appendix Tables S29 and S30 for chemical shifts and assignment report, respectively). As with CREB 285-315, there was minimal chemical shift perturbation upon addition of the non-CRE DNA, indicating that the protein remains disordered.

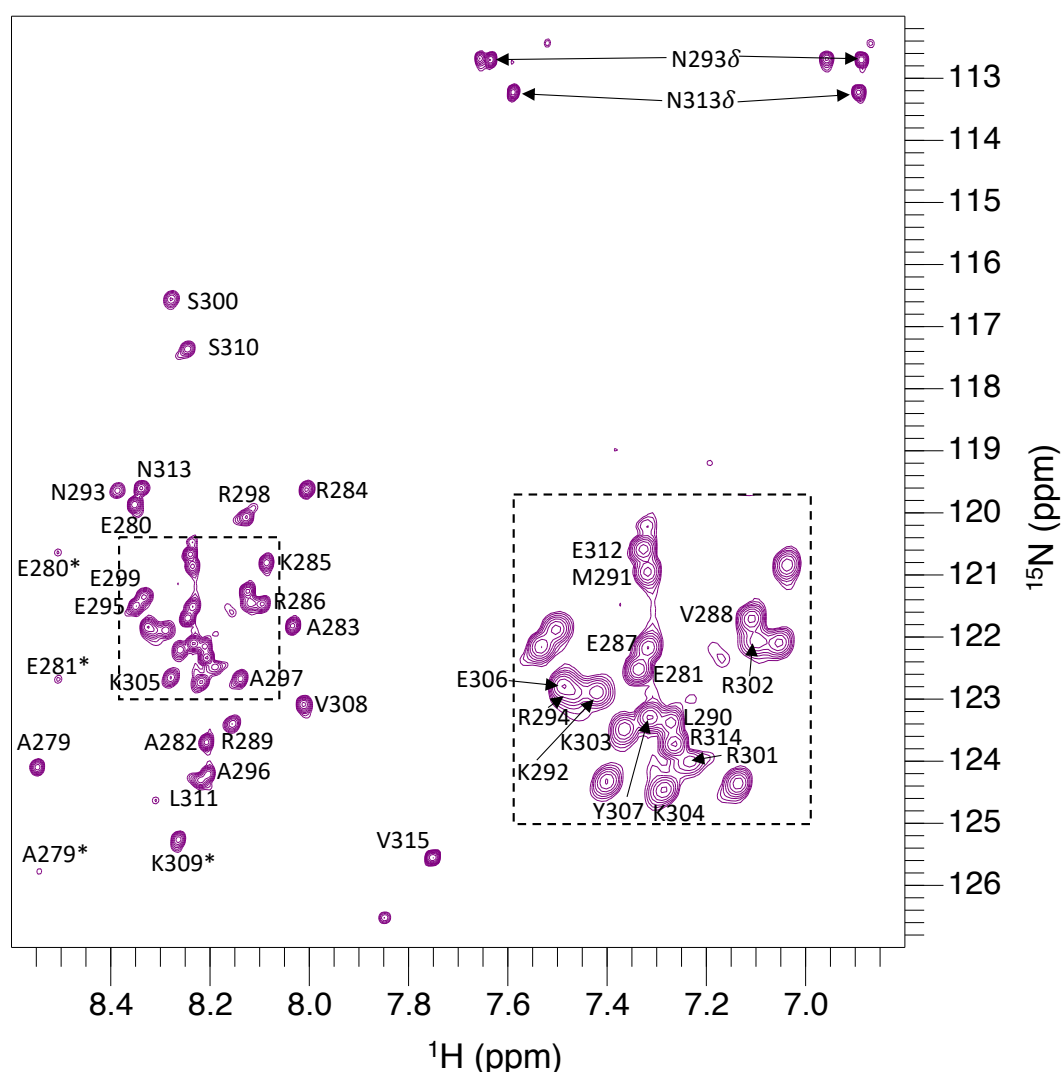


Figure 5.12: CREB 277-315 with non-CRE DNA showed limited chemical dispersion. ^{15}N -HSQC of CREB 277-315 in complex with a saturating concentration of non-CRE DNA. Assignments were collected at 25°C. * indicates peaks in close proximity to the *cis*-proline.

To confirm this, SSP calculations were performed using information from the triple resonance experiments for both constructs (Figure 5.13a). $\text{H}\alpha$ chemical shifts were not collected due to limited dispersion in the ^{13}C -HSQC (Appendix S20).

SSP revealed a small increase in helical content from residue 285-308, with the C-terminus being more similar to the apo-state. However, the SSP ratio does indicate that CREB 285-315 in the protein/non-CRE DNA complex is disordered. The ^1H - ^{15}N heteronuclear NOE revealed that the formation of non-CRE complex does not greatly increase the amount of order in the protein, indicating that the protein is still dynamic when bound to the DNA. Following this, the next step was to determine if the same structural behaviour is seen with the target, CRE-containing DNA. The same analysis was completed for the CREB 277-315 construct with non-CRE DNA (Figure 5.14). This revealed that 277-315 in complex with non-CRE was as dynamic as the apo state in general, with decreased dynamics at residues 292-293 and from 297 towards the C-terminus which is characterised by the highly similar secondary structure content by SSP.

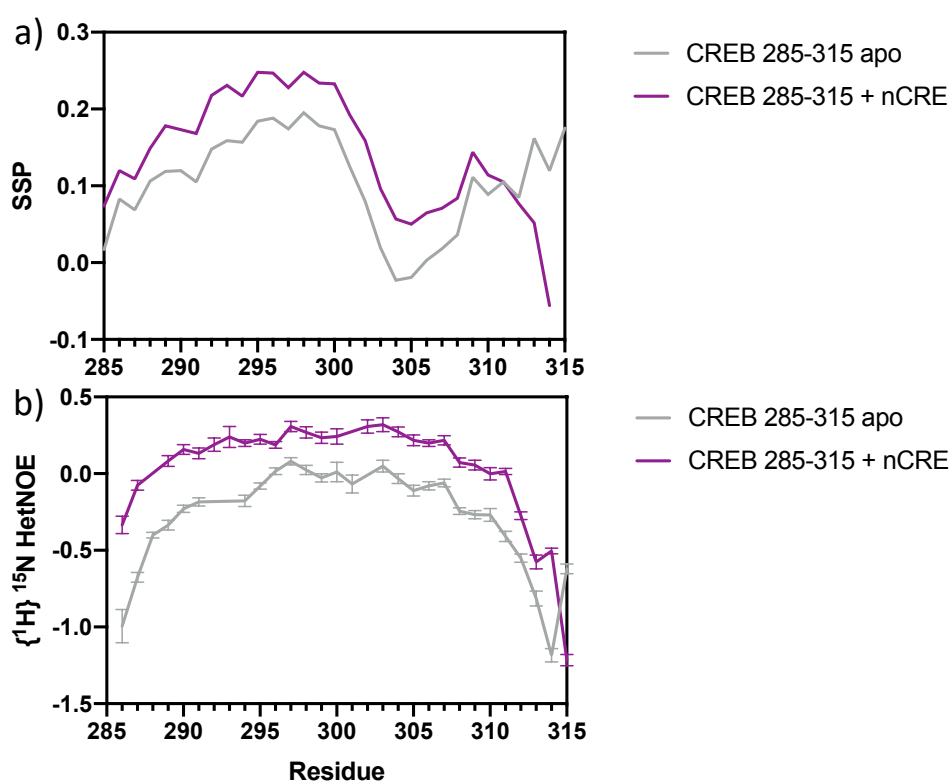


Figure 5.13: CREB 285-315 remains disordered and dynamic when associated to non-CRE DNA. a) SSP plot of CREB 285-315 in the apo state and non-CRE DNA saturated complex. $C\alpha$ and $C\beta$ were used as the atoms in the prediction. b) ^1H - ^{15}N heteronuclear NOE of CREB 285-315 in the apo and non-CRE bound states, collected at 25°C at 600 MHz.

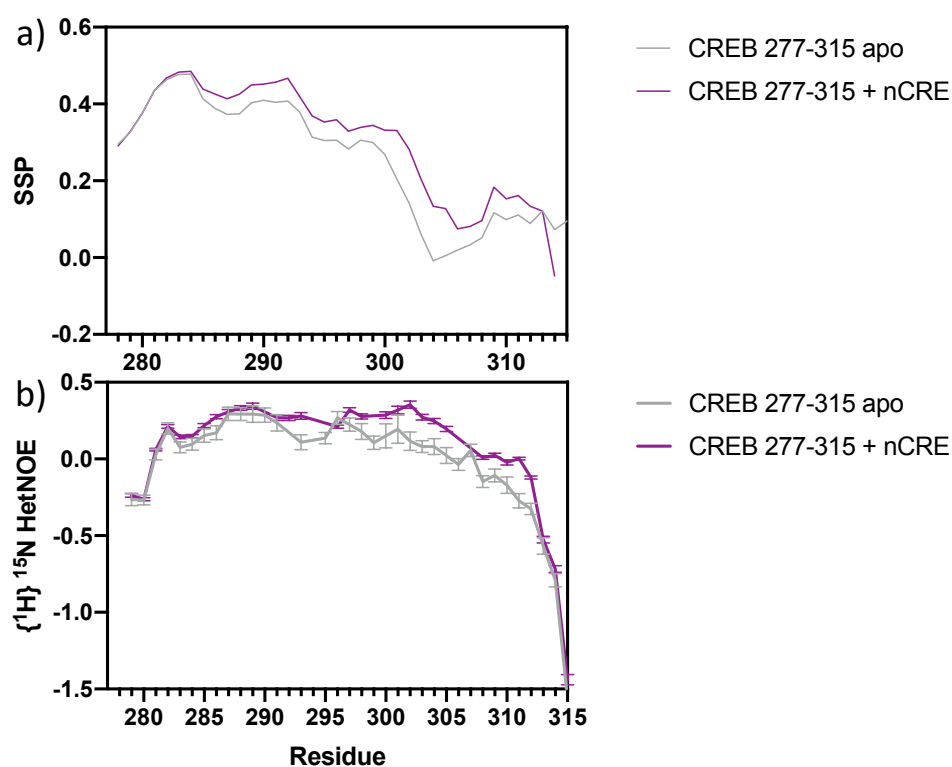


Figure 5.14: CREB 277-315 remains disordered and dynamic when associated to non-CRE. a) SSP plot of CREB 277-315 in the apo state and non-CRE DNA saturated complex. $\text{C}\alpha$ and $\text{C}\beta$ were used as the atoms in the prediction. b) ^1H - ^{15}N heteronuclear NOE of CREB 285-315 in the apo and non-CRE bound states, collected at 25°C at 600 MHz. Errors are SD.

5.2.5 Monomer model with target DNA is in an intermediate/slow exchange regime

In previous sections of this chapter, monomeric CREB has shown reduced affinity for DNA, identified in Section 5.2.2, and a lack of substantial conformational change upon binding to non-CRE DNA in Section 5.2.4. Therefore, it was expected that a similar behaviour would occur with the target site containing CRE DNA. Similar to the non-CRE assignments, a DNA titration was used to aid in the assignment from the apo to the saturated CRE-DNA bound state (Figure 5.15). However, a significant number of the peaks in the ^{15}N -HSQC were in slow or intermediate exchange and disappeared during the titration (Figure 5.16).

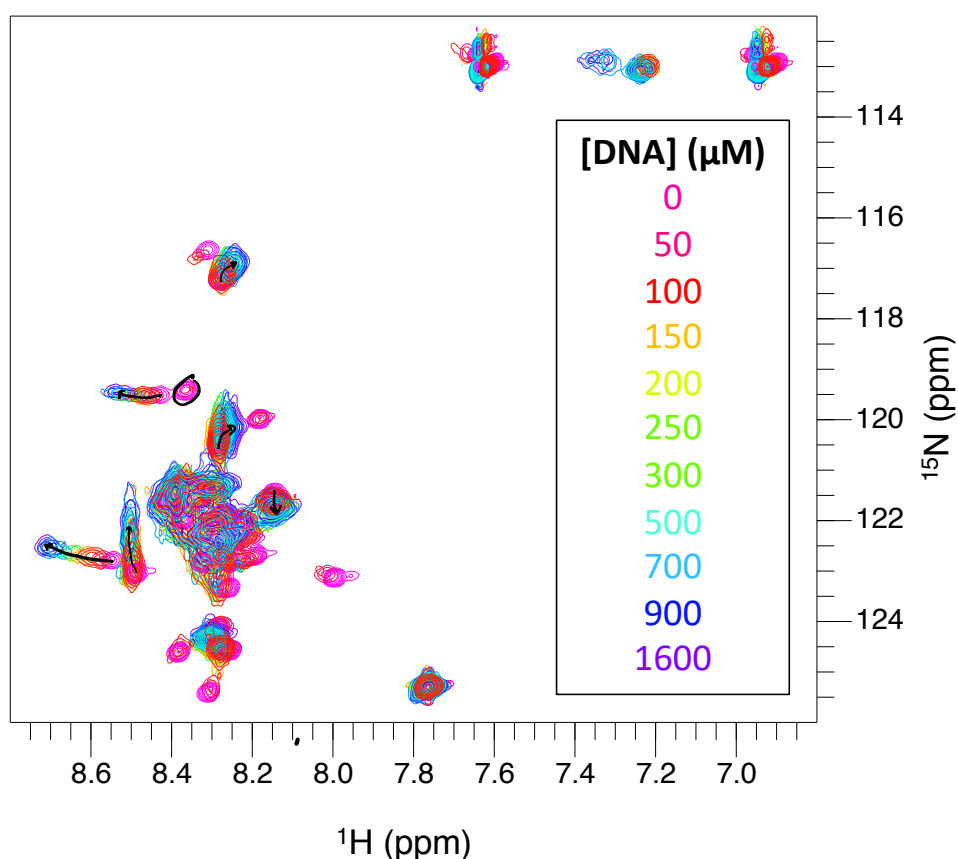


Figure 5.15: Addition of CRE DNA to the CREB 285-315 reveals different exchange regime. ^{15}N -HSQC's were performed with $^{13}\text{C}/^{15}\text{N}$ -CREB 285-315 (342 μM) with increasing CRE-DNA concentration. ^{15}N -HSQC's were performed at 25°C. Arrows indicate direction of the chemical shift perturbation for selected peaks. Circles indicate resonances that were in fast exchange/visible but were not perturbed by the addition of DNA.

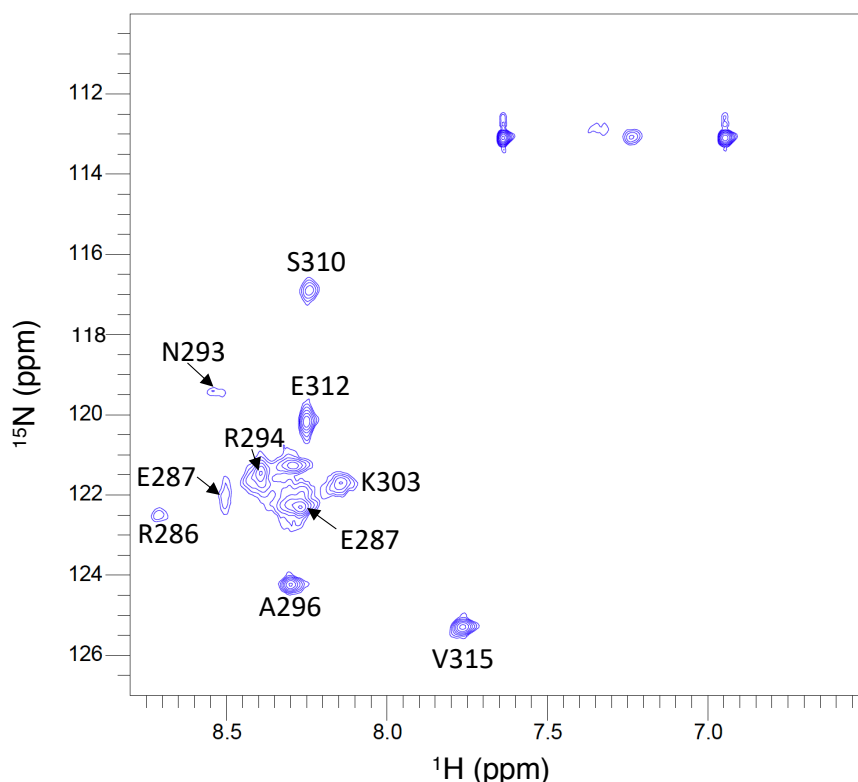


Figure 5.16: CREB 285-315 is in slow/intermediate exchange with CRE-DNA. Assigned ^{15}N -HSQC of $^{13}\text{C}/^{15}\text{N}$ -CREB 285-315 with CRE DNA, collected at 25°C.

Of the peaks that remained and could be followed during the titration, some appeared to move via a non-linear trajectory, in contrast to the CREB 285-315 with non-CRE DNA titration. Nevertheless, the chemical shift perturbation and peaks intensities were measured and are shown in Figure 5.17. Only six residues were used in the weighted average analysis however (R286, E287, V288, M291, S310, E312), which resulted in an estimate of the K_d for CREB 285-315 with CRE DNA of $130 \pm 20 \mu\text{M}$. However, the peak intensity of all residues decreased which was consistent with the visual analysis of the ^{15}N -HSQC mentioned above.

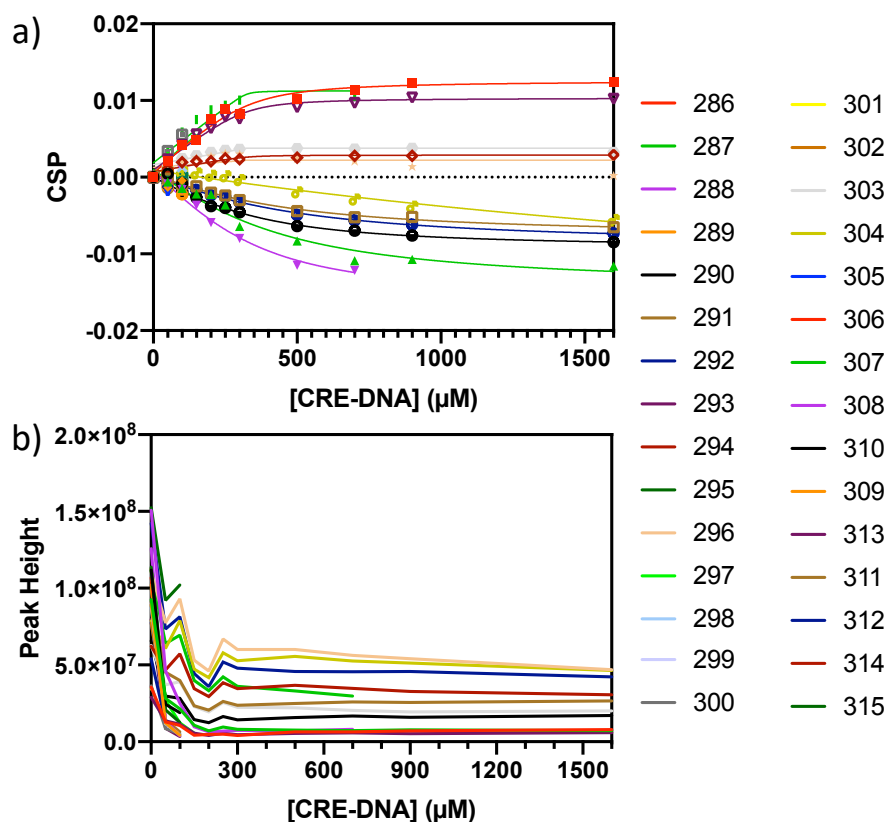


Figure 5.17: Limited chemical shift perturbation with CREB 285-315 during a CRE DNA titration due to a reduction in peak intensity. a) CSP of CREB 285-315 ^1H - ^{15}N from ^{15}N -HSQC's with varied CRE DNA concentration. Data were fitted to a single site binding equation 2.2. b) Peak height from the same data shown in a). Data analysed using CCPN v2.6. Data collected at 25°C at 600 MHz.

As the DNA binding affinity can be estimated from the peak shifts, seen with the non-CRE DNA above in Section 5.2.4, this was used for the CRE DNA as well (Figure 5.17a and 5.18). The result shows that, for each residue in Figure 5.18, the K_d for CRE DNA is lower than the non-CRE, however in some cases it is not statistically significant.

In order to observe any conformational changes of CREB 285-315 with CRE DNA, a decision was taken to increase the temperature to try to observe the remainder of the peaks. Increasing the temperature to 35°C gave the correct number of peaks and CREB 285-315 with CRE DNA was assigned using triple resonance experiments only under these conditions (Figure 5.19). A comparison of the ^{13}C -HSQC at the two temperatures showed minimal movement of the peaks from 25°C to 35°C (Figure

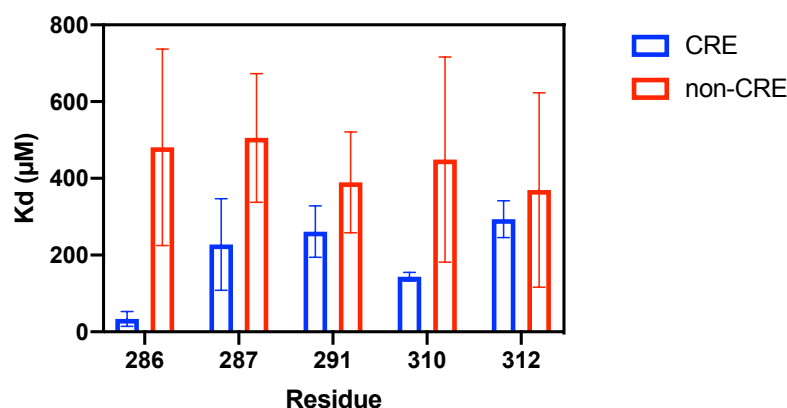


Figure 5.18: Binding affinity of monomeric CREB model to CRE and non-CRE DNA with NMR. A comparison of the extracted K_d from the data presented in Figure 5.17a and 5.11, between CRE and non-CRE DNA binding to CREB 285-315. Error is SEM.

5.20), aside from the methyl region. At 35°C, the peaks in the ^{15}N -HSQC appeared

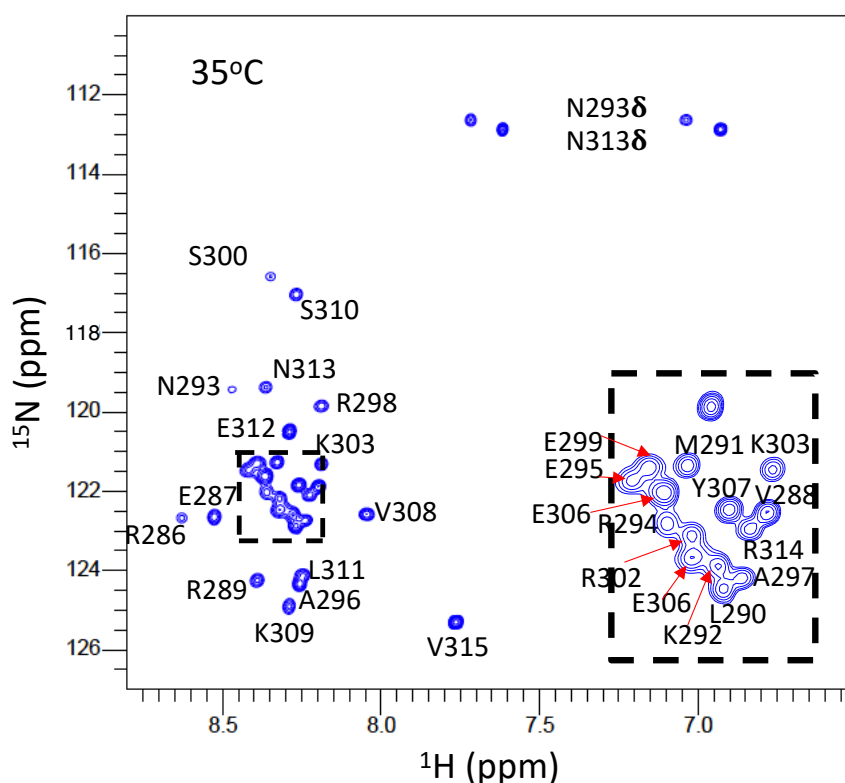


Figure 5.19: CREB 285-315 in CRE-DNA complex backbone amide peaks reappear at higher temperature. ^{15}N -HSQC of ^{13}C - ^{15}N CREB 285-315 with CRE-DNA. Data was collected at 35°C.

with the ^1H axis range of 8.0-8.5 ppm, indicative of the protein being disordered.

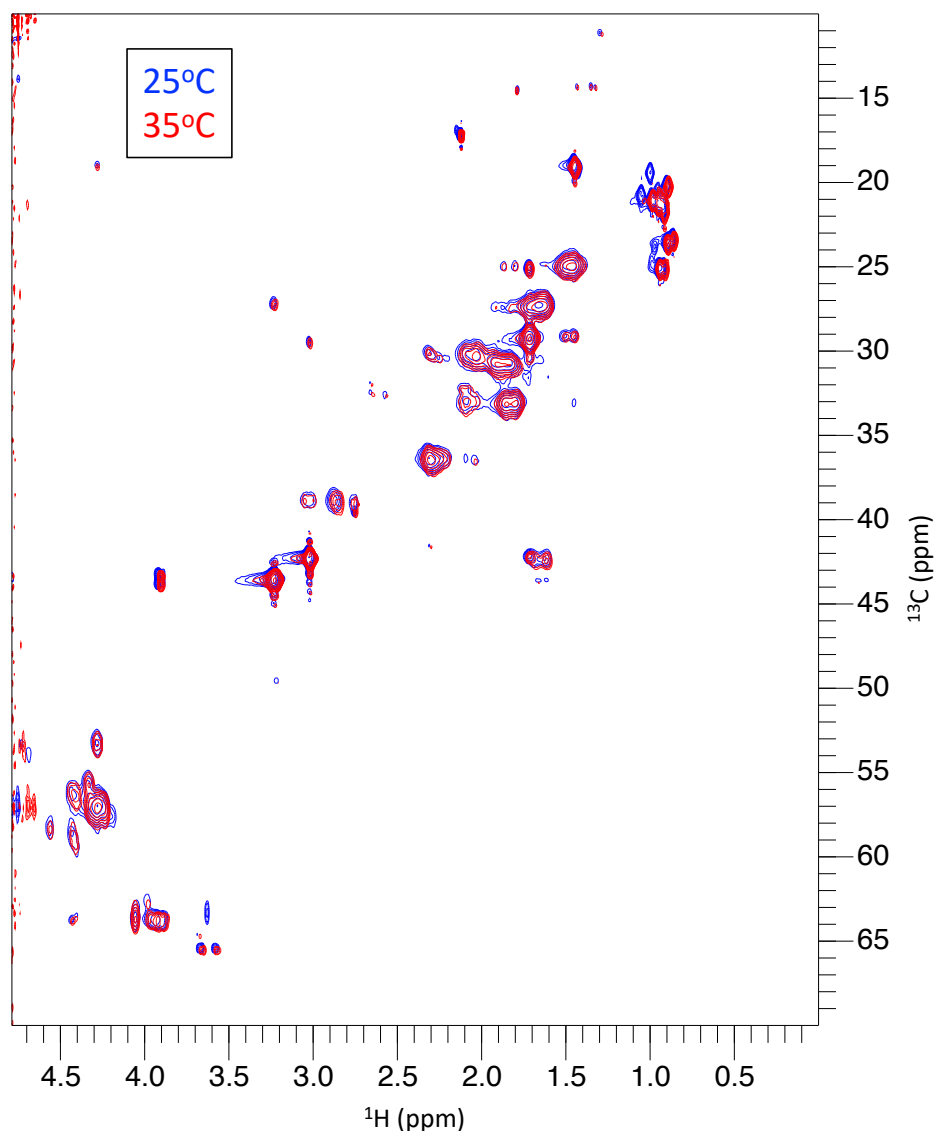


Figure 5.20: Minimal chemical shift perturbation of CREB 285-315 in complex with CRE-DNA at different temperatures. ^{13}C -HSQC of CREB 285-315 in complex with saturating concentration of CRE-DNA, collected at 25°C and 35°C.

As the triple resonance and dynamic based experiments were now completed for the CREB 285-315 complexes, investigations into the what effect the DNA has on the secondary could be determined at the residue specific level. The chemical shift table and assignment report for this complex are shown in Appendix Tables S31 and S32.

A calculation of the secondary structure propensity revealed that, under the conditions that each CREB 285-315 state was analysed, there is no major conformational change upon association to DNA (Figure 5.21). This is consistent

with the lack of chemical shift dispersion that was noted in the ^{15}N -HSQC. This is reflected when looking at the dynamics with the heteronuclear NOE, which identified that the DNA-bound state of CREB 285-315 was very dynamic, but the apo state appeared more dynamic to a degree.

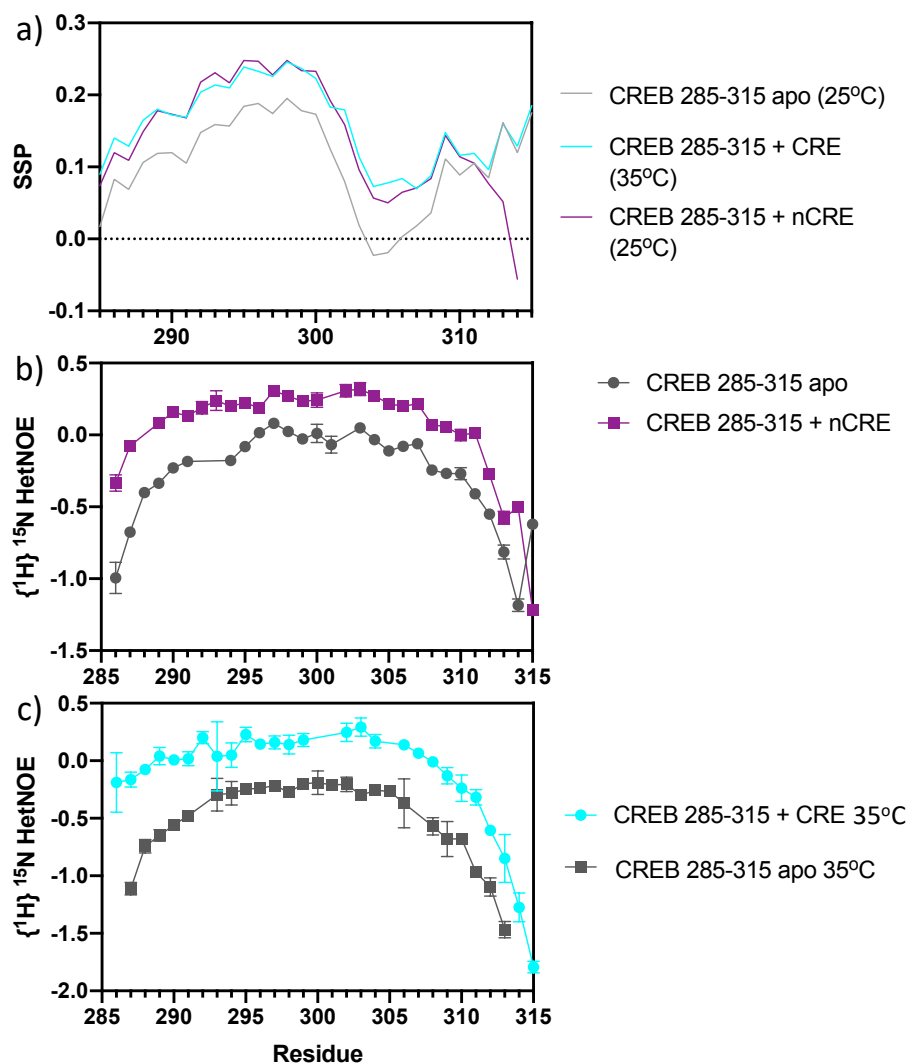


Figure 5.21: No major change in secondary structure or dynamics of monomeric CREB when associated to DNA. a) SSP of CREB 285-315 in the apo-state and in complex with CRE- and non-CRE DNA. Estimation used $C\alpha$ and $C\beta$ chemical shifts. b) and c) ^{15}N -Heteronuclear NOE of CREB 285-315 in complex with non-CRE (collected at 25°C) and with CRE-DNA (collected at 35°C), respectively. Heteronuclear NOE data were collected at 600 MHz.

The more helical version of the monomer, CREB 277-315, was then used to see if a more helical construct may be able to undergo coupled folding upon binding to the

CRE target DNA sequence. The ^{15}N -HSQC revealed that, similarly to the 285-315 construct, addition of CRE DNA resulted in a slow/intermediate exchange regime at 25°C (Appendix Figure S22). Interestingly, increasing the temperature to 35°C did not push the system into fast exchange, nor did it result in the appearance of the missing peaks. The ^{13}C -HSQC at the two temperatures did reveal a more specific mode of interaction for CREB 277-315 with CRE DNA over the nonCRE-DNA (Figure 5.22), in particular the Arg C δ region being more dispersed with CRE-DNA. Hence, full assignment of this construct was not possible and therefore analysis could not be extended further. What was evident throughout the results presented in this chapter was that the monomer model of CREB appears to have very little or no specificity for the CRE target DNA sequence. In support of this, there was no conformational change when the protein associates to the CRE DNA sequence, however NMR showed a change in exchange regime between the CRE DNA and a DNA sequence not containing the target site, suggesting a difference in on and off rates. In Chapter 4, there is a stark contrast in the way that CREB bZIP can associate to DNA sequences and specifically recognise the target site. With the only difference being the presence of the leucine zipper region of CREB bZIP, it was assumed at this stage that one of the possible mechanisms controlling CREB specificity was helical propagation from the leucine zipper into the basic region.

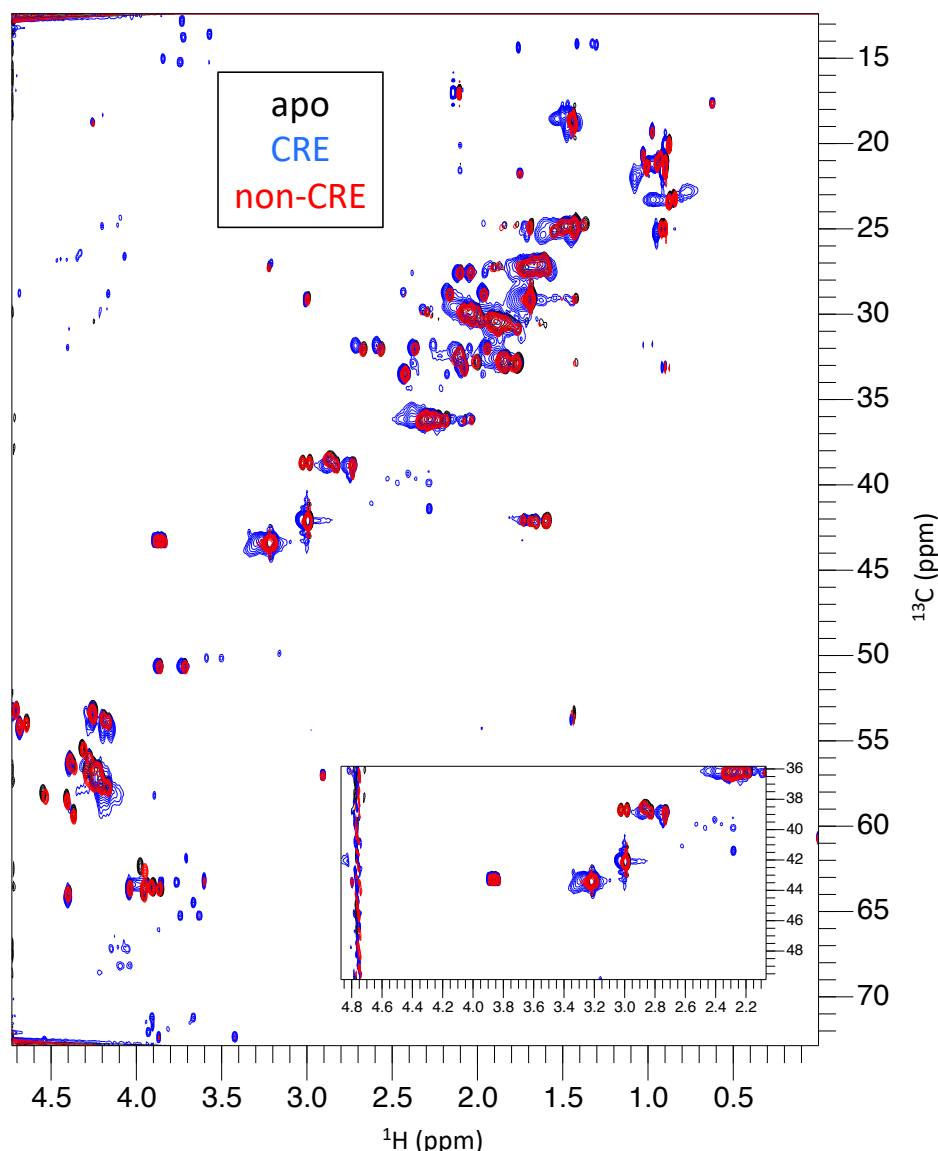


Figure 5.22: Specific interactions were observed by CREB 277-315 with CRE-DNA over non-CRE DNA. ^{13}C -HSQC of CREB 277-315 (300 μM) in complex with CRE-DNA (3200 μM) and non-CRE DNA (2400 μM). Data were collected at 25°C.

5.2.6 Decoupled leucine zipper construct shows a dependency on homodimer formation for target DNA specificity

In the previous sections of this chapter, a monomeric CREB construct has shown low target site specificity (Sections 5.2.2, 5.2.4 and 5.2.5) and limited to no coupled folding upon binding to any DNA sequence (Sections 5.2.4 and 5.2.5). When looking at the same complexes in the context of the bZIP sequence, the opposite behaviour

is seen. To investigate the role of the leucine zipper, a CREB bZIP construct was designed which would abrogate any structural propagation into the basic region. The resulting construct, CREB 285-339 4G_{ins307}, had four glycine residues inserted at the interface between the basic region and the leucine zipper in order to maintain the residue positions in the event of helical formation. A confirmation of reduced helicity in the basic region compared to the apo-state of CREB bZIP pWT was done using natural abundance NMR (Figure 5.23 and 5.24). What was clear immediately

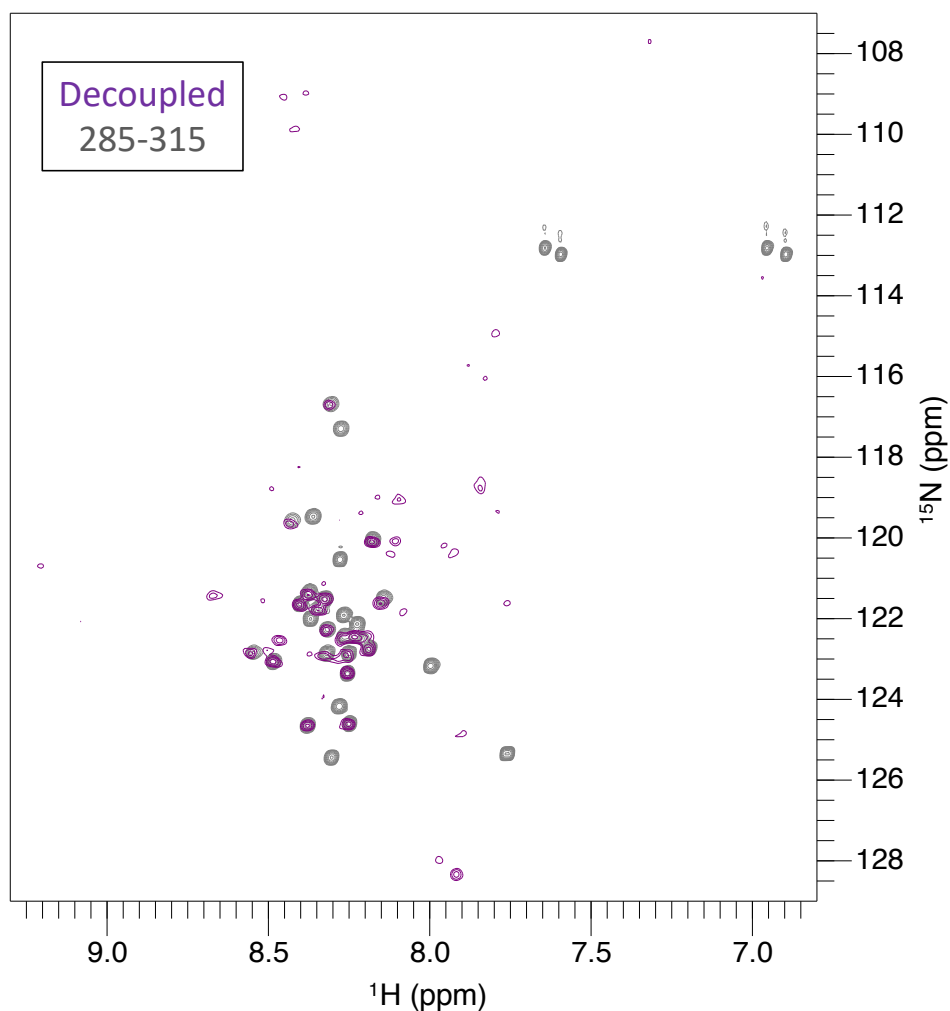


Figure 5.23: Basic region residues in a decoupled bZIP are as disordered as a monomer. Natural abundance ^{15}N -BTROSY of CREB 285-339 4G_{ins307} at 25°C. Overlaid with a ^{15}N -HSQC of CREB 285-315, assignments can be seen in Figure 5.4.

from the natural abundance ^{15}N -BTROSY experiments was that several residues in CREB 285-339 4G_{ins307} gave peaks of high intensity, and these were located

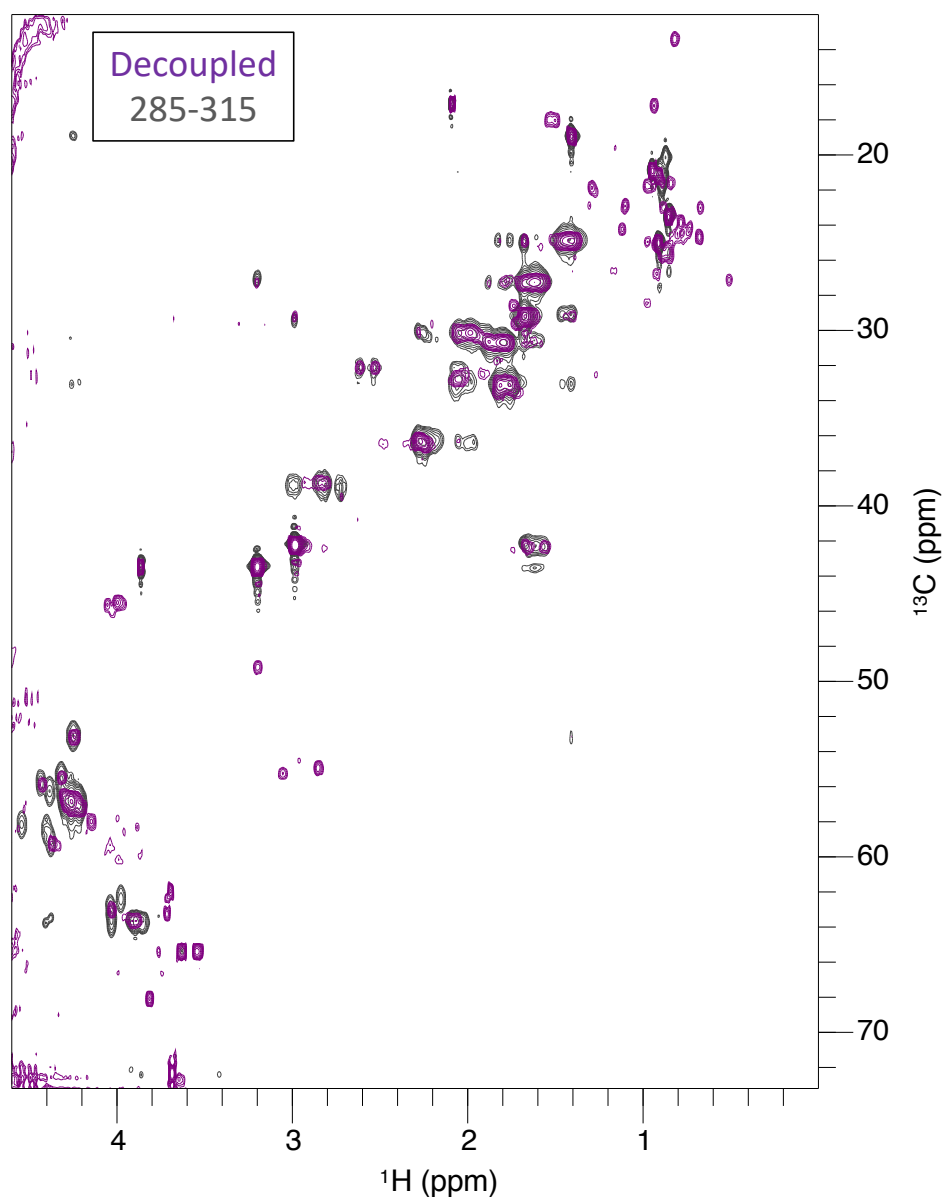


Figure 5.24: Carbon chemical shifts show basic region in decoupled bZIP are disordered. Natural abundance ^{13}C -HSQC of CREB 285-339 $4\text{G}_{\text{ins307}}$ at 25°C , overlaid with a ^{13}C -HSQC of CREB 285-315.

within the disordered region of the proton axis between 8.0-8.5 ppm. Conversely, the well dispersed peaks that correspond to some of the leucine zipper residues were not visible in the decoupled construct. Overlaying the spectra of the decoupled construct with that of CREB 285-315 identified that those high intensity peaks overlaid exactly to the residues in CREB 285-315. This suggested that decoupling the helical propensity of the leucine zipper from the basic region by the insertion of

the four glycine residues renders the the basic region disordered to a similar extent to the monomeric CREB 285-315. The ^{13}C -HSQC confirms that the carbon chemical shifts, which directly report on secondary structure, also indicate the basic region residues overlap significantly between the decoupled construct and CREB 285-315.

Further examination of the decoupled protein construct involved interaction with DNA. Circular dichroism of the decoupled construct alone revealed a significantly reduced helical signal than CREB bZIP pWT alone (Figure 5.25a). The addition of DNA led to small increase of helical content, but this is still much reduced compared to the CREB bZIP pWT apo. Using fluorescence anisotropy, the binding affinity of the decoupled protein to CRE and non-CRE DNA revealed that there is a weak interaction with both types of DNA, but that both are one order of magnitude tighter than the CREB 285-315 construct (Figure 5.25b). However, the small difference between the target CRE DNA and non-CRE DNA K_d with the decoupled construct indicates that the helical propagation that the leucine zipper would have provided aids in the specificity of the interaction of CREB bZIP for target over the non-target DNA. This was not down to a difference in the stability of the homodimer, which was investigated via urea denaturation at equilibrium and resulted in a D_{50} of 3.8 ± 0.1 M whereas the pWT was at 3.6 ± 0.1 M (Appendix Figure S21).

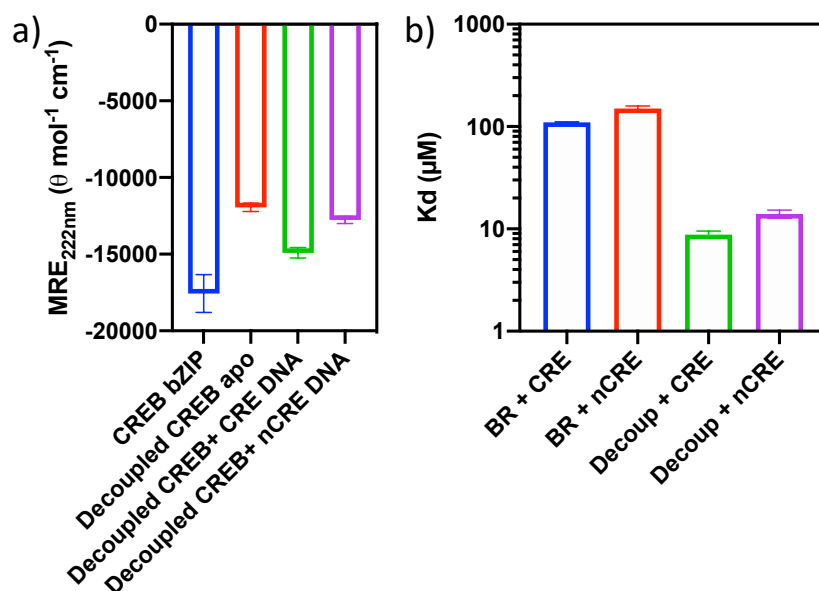


Figure 5.25: Lack of coupled folding upon binding and a weak interaction of decoupled bZIP to DNA. a) Circular dichroism of 10 μM CREB bZIP 285-339 4G_{ins307} apo or with equimolar concentrations of CRE- and non-CRE DNA. CD was performed at 25°C. b) K_d estimates determined from single site binding fits of fluorescence anisotropy data of decoupled CREB bZIP with CRE- and non-CRE DNA. Data obtained at 25°C. Errors are SE, from three independent replicates.

5.3 Discussion

The monomer model of CREB's DNA binding domain has been investigated by biophysical techniques to define its role in the pathway of target association, and thus start to understand which of the two pathways are preferred by bZIPs. The design of the monomeric construct, by simply removing the leucine zipper region of the bZIP domain was a certain method to eliminate the formation of the dimer and has been used with a variety of bZIP containing proteins^{78–81,121,125}, so this approach was executed for CREB. A comparison between the SSP calculation and Agadir predictions^{165,176} of the apo-state of the basic region showed a greater reliability in the predictions compared to the full bZIP (Figure 5.26). This is indicative of Agadir being more reliable for monomeric proteins, rather than those that have the propensity to be dimeric, and therefore was a useful tool in describing the influence of the number of residues in the basic region on the helical propensity.

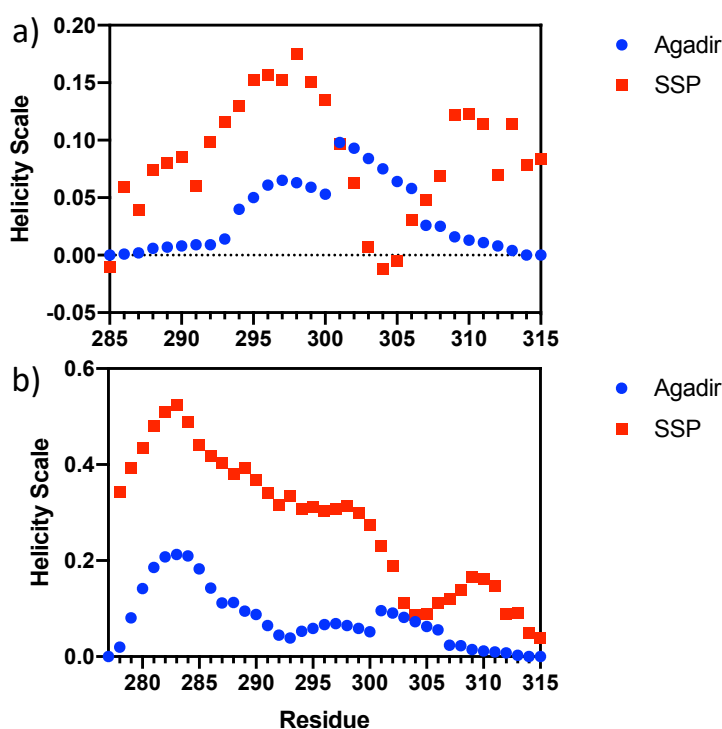


Figure 5.26: Accuracy of Agadir prediction is closer for the basic region only constructs. a) Comparison of the Agadir and SSP scores for CREB 285-315. b) Same as a) with the CREB 277-315 construct. SSP chemical shifts used were $C\alpha$, $C\beta$ and $H\alpha$. Agadir score was divided by 100 to give the ratio in the same scale as the SSP.

A characteristic drop in helical propensity was observed around residues 300-305 in both constructs (sequence: SRRKKK), which is in direct contrast to the prediction by Agadir and independent of the helicity at the N-terminus of that region. Interestingly, the bZIP apo analysis in section 3.2.3 showed a plateau of the helicity within this same region, which was attributed to the intra-helical electrostatic interaction between E299/R302/E306²¹. The general helical content of the monomeric constructs are significantly lower than that found in the bZIP constructs, which was confirmed by analysis of the apo-state of the CREB 277/285-315 constructs using NMR. Both protein constructs were significantly disordered, lacking the helical content seen in the corresponding bZIP domains. The basic region only constructs were also very dynamic, with no ordered regions across the protein length.

This also leads into a possible source of experimental differences between previous investigations into the pathway of bZIP associating to target DNA. As it was shown for CREB, the length of the construct in both the bZIP and more importantly the basic region construct is important for the helical propensity in the basic region. For instance, work by Busch *et al* used a CREB basic region consisting of residues 285-309⁷⁸, which is a few residues shorter at the C-terminus than the CREB basic region used here, and they reported a selectivity of CREB 285-309 for CRE DNA via EMSA. This discrepancy in construct length is also seen with GCN4, with constructs spanning 222-253, 222-249, 226-252, and 222-256 being utilised^{79-81,125,129}. These differences could have a marked impact on the stability of protein/DNA complexes and should be treated with caution.

One of the implications of the lack of structure seen in the basic region only constructs is that thermodynamically it would be more expensive to associate to DNA, in particular a target containing DNA sequence. Interrogation of this via fluorescence anisotropy yielded a K_d for the CREB 285-315 with CRE and non-CRE DNA that showed a slight specificity for CRE-DNA, but this specificity was significantly lower than that seen for the corresponding bZIP domain. Interestingly, the more helical CREB 277-315 construct showed a weaker affinity for both types

of DNA, but this was more specific with a two-fold difference observed between CRE and non-CRE DNA. The specificity of other basic region proteins has been investigated, with the majority in agreement that the monomeric version can associate to target containing DNA (footprinting assay¹²¹, fluorescence titration⁸¹, ITC¹²⁵), however with a K_d in the μM range. The ITC experiment showed that the association of GCN4 basic region to the two target containing DNA sequences was enthalpically driven, and also incurred an increased helical content¹²⁵. However, with the estimate of the K_d for CREB basic region constructs, ITC was not feasible to complete for this work as the K_d is outside the working range for analysis. A direct comparison between GCN4 bZIP to basic region by Berger *et al* showed that the bZIP showed increased affinity for CRE on the order to 1000-fold⁸¹. However, one issue was that the binding affinity to non-target DNA was not characterised in these experiments, only indirectly as with the footprinting assay. One of the potential issues the use of the basic region only construct is the removal of a significant portion of the leucine zipper. However, the results presented here are consistent previous work that mutated several leucine residues in the zipper region, which was also shown to abrogate the specific binding of CREB to CRE DNA¹³⁵.

Due to the difficulties in labelling the bZIP domain in Chapter 3 and the estimated K_d of the protein/DNA complex, stopped flow kinetics was not performed. This was unfortunate as previous stopped flow experiments with GCN4 basic region showed a k_{on} that was at the diffusion limit^{80,81}, similar to many other transcription factors such as Cro and TatR repressors, and the TATA binding protein amongst others (reviewed in¹⁹⁶). However, the LacI were found to associate with rates beyond the diffusion limit ($\geq 10^{10}$)¹⁹⁷. This perhaps isn't surprising, due to high net positive charge found in the basic regions due to the enrichment of arginines and lysines in the sequence aiding in the association kinetics. Limited information exists about the k_{off} of basic region proteins from DNA, however one study with GCN4 basic region did reveal a k_{off} of 83 ± 20 (s^{-1}), twice as fast as the bZIP⁸⁰. However the k_{on} for the basic region was also found to be twice as fast as the bZIP, thereby resulting in a similar K_d . The authors implied that due to the homodimer

K_d , that association of two monomers to target DNA was faster than going through the dimer pathway. Therefore, the result may be down to use of a 68 mM ionic strength buffer for the stopped flow experiments to examine the K_d , which would markedly increase the affinity of a basic region to the DNA⁸⁰.

That being said, NMR did reveal that on the μ s-ms timescales, the mode of interaction for CREB 285/277-315 to CRE-DNA is different compared to the non-CRE DNA. The broadening of the peaks observed for the CRE-DNA complex at 25°C possibly suggests one of a few different mechanisms at play: 1) Two sites are available for CREB basic region constructs due to the full CRE-DNA being used, resulting in two sites with different K_d 's due to steric hindrance or co-operativity and broadening due to the different specificities incurring different structural propensities or 2) single site binding of CREB basic region but in a three state mechanism with an unfolded intermediate that is resolvable under the timescales of the ^{15}N -HSQC. Further experiments would need to be performed to eliminate these various mechanisms, such as the use of half site CRE-DNA to answer 1). The curved nature of the titration paths of peaks for residues that were not broadened during the titration appears to agree with a three-state complex with a conformational re-arrangement¹⁶². One of the possibilities is that, due to the very high K_d as determined by equilibrium, the complex is not saturated. However, in the case of both protein constructs, the complex formation percentage would have been >95%. The ^{13}C -HSQC data showed a limited chemical shift perturbation with the 285-315 construct upon associating to both DNA sequences. In direct contrast, the 277-315 sequence did show a specific association to the CRE-DNA, as shown by the chemical shift perturbation in the Arg C δ region. Therefore, the increased helical propensity in the 277-315 construct plays a role in the specificity of the DNA interaction by CREB, with the N-terminal segments of bZIPs having been shown to be important for this structural function¹⁸⁶.

Increasing the temperature of the CREB 285-315 protein with CRE-DNA resulted in a push of the system into fast exchange, similar to what was seen with the non-CRE DNA complex. However, a discrepancy in the results of the

277-315 complex exists. Whilst the non-CRE DNA complexes for CREB 285-315 and 277-315 both showed limited secondary structure change and a small decrease in the dynamics of the protein (most likely due to the increased viscosity due to the presence of mM quantities of DNA), the CRE-DNA complex with CREB 277-315 did not move to the fast exchange regime upon increased temperature. As the affinity of this complex is believed to be lower than the CREB 285-315 with CRE-DNA, this is a surprising result, as it would be expected a weaker affinity interaction would be more easily pushed towards the fast exchange regime whereas this is not case.

A key piece of evidence that does support dimer pathway was that of the lack of reconstitution of the specificity that CREB bZIP has for target DNA when the leucine zipper was decoupled from the basic region. The basic region in the decoupled construct was shown to be as disordered as the basic region only construct by NMR, in fact the peaks associated to the basic region were very sharp and well resolved whereas the signal to noise made the peaks corresponding to the leucine zipper broad due to data acquisition being performed under natural abundance. There was a limited difference in the binding affinity for both CRE and non-CRE DNA and limited coupled folding upon binding due to the interaction. This showed that the helical propensity that was propagated from the leucine zipper into the basic region was important for the function of the protein. Also, it showed that the leucine zipper had to be folded, which it is only capable of doing so when interacting as a dimer. A full comparison and the implications for the pathway preference for CREB will be examined in the Discussion Chapter.

6

Productive target search of CREB is enhanced by disorder

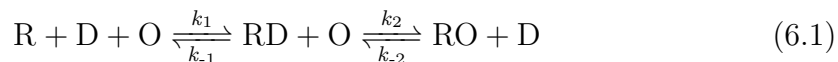
Contents

6.1	Introduction	188
6.2	Results	191
6.2.1	Development of kinetic strategy for determining target search parameters	191
6.2.2	Characterisation of target search of CREB bZIP	201
6.2.3	Computational and structural examination of the effective sliding length of CREB bZIP	209
6.3	Discussion	214

6.1 Introduction

As discussed in Section 1.7, the target search process for transcription factors requires the specific association to a target DNA site in the presence of an excess of non-target sites. The initial theory proposed by Berg *et al* suggested that the first step would be the binding to an extended target, followed by another step which tethers the protein (R) to the macromolecule containing the target site (D), which reduces the dimensionality of the second step helping to increase the overall

rate of association to the target (O), shown in equation 6.1¹³⁹.



A simpler two-state model with a first order rate of association fails based on evidence from the Debye-Smoluchowski equation. The equation predicts the maximum rate for the association of two uncharged species by random diffusion to be $\sim 10^7$ to $\sim 10^8 \text{ M}^{-1} \text{ s}^{-1}$. This is several orders of magnitude lower than the k_a of LacI to Lac operator ($\sim 10^{10} \text{ M}^{-1} \text{ s}^{-1}$)¹³⁸, therefore detailed descriptions of facilitated diffusion events were proposed (discussed in section 1.7). Several proteins have now been studied and are summarised in Table 6.1.

A controversy arose surrounding the speed-selectivity paradox. In short, the experimentally observed fast search of proteins to the target DNA (LacI¹³⁸, TetR¹⁹⁸, p53¹⁴⁶ to summarise a few) requires smooth binding potentials, whereas the protein recognising the target DNA sequence itself presents a large energy barrier which would slow down the target search (summary by Veksler and Kolomeisky¹⁹⁹). Estimated standard deviations of the energy distribution, σ , for the smooth binding potentials were $< 1\text{-}2 \text{ k}_B\text{T}$ during target search whereas the stability of the protein-target DNA complex increases σ to $> 5 \text{ k}_B\text{T}$. To account for this energy difference, different modes were defined for DNA binding proteins, typically referred to as search and recognition modes in the simplest case²⁰⁰. The low binding potentials are thus the search mode, with the high binding potentials being the recognition mode, with a switch occurring between the two states.

Table 6.1: Available target search parameters obtained in the literature, highlighting the use of different models to obtain the sliding rate in particular. λ is the effective sliding length. Errors are SEM.

Protein	Target	Competitor DNA	Method	k_a ($M^{-1} s^{-1}$)	λ (bp)	Sliding rate ($\times 10^5$)	Model	Reference
LacI	Primary and secondary operator sequences	10.2 kbp with O in the centre	Optical tweezers			4.1 ± 2.3 $nm^2 s^{-1}$	Continuum	[141]
LacI	Operator		FRET and FCS		39 ± 9	0.027 ± 0.001 $\mu m^2 s^{-1}$	Continuum	[142]
LacI	Operator	in vivo	Single molecule (YFP tagged LacI)	0.73 ± 0.02 min^{-1}	45 ± 10			[143]
LacI	Various Non-target sequences		NMR	4×10^{12}		4×10^{12} $cm^2 s^{-1}$	Continuum	[192]
LacI	operator - various lengths of sequence		Filter assay	$4.2-70 \times 10^8$				[201]
LacI	operator	lambda plac5	Filter assay	$< 10^9$			Continuum	[137]
TetR	<i>tetO</i>	in vivo	Single molecule tagged protein Atto647N	$(9.2 \pm 0.2) \times 10^4$	250-750	~ 8 $\mu m^2 s^{-1}$	Continuum	[198]
Egr-1	Egr-1 target	Random (see ¹⁴⁹)	Stopped flow and NMR	$\sim 0.5 \times 10^7$	50 ± 2	~ 2 $bp^2 s^{-1}$	Discrete	[151]
p53	lambda phage		TIRF with microfluidic assay of fixed DNA			10 ± 15 $bp^2 s^{-1}$	Discrete	[145]
p53	48.5 kb dsDNA	n/a	TIRF with microfluidic assay of fixed DNA			1.62 ± 0.17 $nm^2 s^{-1}$	Continuum	[146]
p53	48.5 kb dsDNA	n/a	TIRF with microfluidic assay of fixed DNA			14 ± 1.5 $bp^2 s^{-1}$	Discrete	[146]

However, investigations by Veksler and Kolomeisky (VK) revealed that the paradox was an artefact from the model that was being utilised¹⁹⁹. For instance, in the Berg, Winter and von Hippel model the target search along DNA was defined as a continuum, with the DNA being modelled as a negatively charged rod¹³⁹. The continuum model predicts that at a limit of small sliding lengths, the search rates would go to zero which is not experimentally observed. Modelling the process as discrete sites removes this particular limit, and also negates the requirements for conformational rearrangements which had been used to explain this phenomenon.

As discussed in section 1.5.1, there is limited information on what search mechanism CREB uses to find the target site within the genome. Therefore, the aims for this chapter were to investigate the mechanism that CREB bZIP predominantly uses to perform effective target search. With the research group having prior knowledge of stopped flow kinetics, the approach utilised by Esadze *et al* using stopped flow kinetics will be adapted for use with CREB¹⁴⁹. By altering the residual helical content of the basic region in the context of the bZIP, the role of disorder will begin to be elucidated during the search process also. The flanking sequence characteristics of the DNA will be briefly examined to provide an *in vivo* context to the work described in this chapter.

6.2 Results

6.2.1 Development of kinetic strategy for determining target search parameters

To determine the 1D and 3D diffusional coefficient for CREB bZIP, various methods were considered including single molecule microscopy. However, the method by Esadze *et al*¹⁴⁹ was chosen to be applied to CREB bZIP primarily due to ease of access and previous experience of the group with stopped flow apparatus. This approach utilises two experiments, each of which can extract the 1D and 3D diffusional coefficients for sliding and intersegmental transfer respectively (Figure 6.1). The theoretical framework for the use of the apparent pseudo first order rate constant (k_{app}) can be viewed in Esadze *et al*¹⁴⁹. In short, the conditions

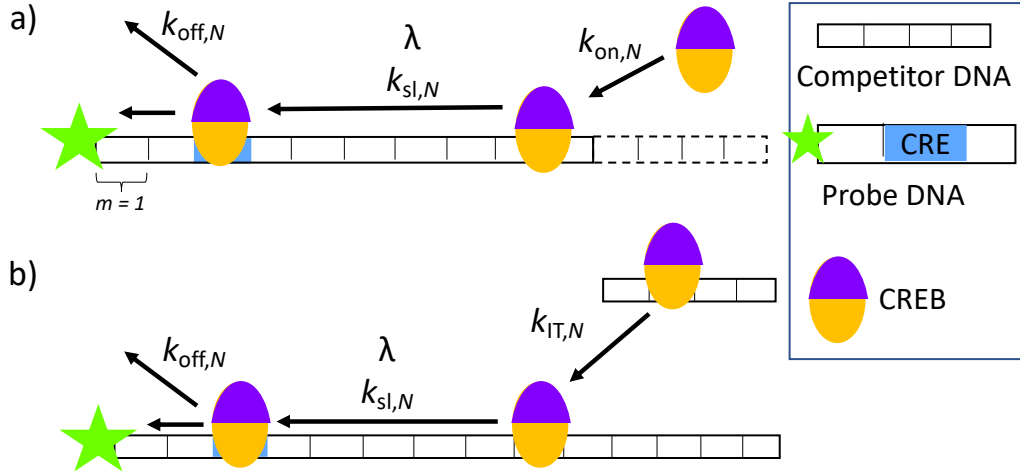


Figure 6.1: Illustration of the stopped flow set-up and experimental parameters that can be extracted. a) 1D diffusion coefficient ($k_{sl,N}$) can be extracted by measuring the k_{app} at various lengths of probe DNA under the conditions set out in a). Hashed lines indicate increasing length of the probe DNA, which contains the target site in blue. c) Similar to b), but with a fixed probe DNA length and varying concentration of competitor DNA the intersegmental transfer rate on non-target DNA can be extracted. Diagrams were adapted from Esadze *et al*¹⁴⁹.

under which the experiments are performed ($D_{tot} \ll P_{tot} \ll C_{tot}$, where D_{tot} is probe DNA concentration, P_{tot} is total protein concentration and C_{tot} is competitor DNA concentration) in the presence of a large excess of competitor DNA means that only one protein will be associated to any probe DNA molecule at a time. This removes any interference with the measurement of sliding by reducing the possibility of multiple proteins on the DNA at the same time. Furthermore, it avoids introducing significant proportions of non-target DNA as the probe lengths gets longer than the sliding length. This also has the advantage of providing a more realistic picture of proteins undergoing DNA target search. k_{app} can be extracted and by varying either the probe DNA length or the competitor DNA concentration, the sliding rate and the intersegmental transfer respectively can be analytically determined. To complete this, the probe DNA needed to be synthesised and experimental conditions optimised.

Synthesis of various lengths of probe DNA

The approach that Esadze *et al* used to extract the 1D diffusional coefficient involved a fluorescently labelled probe DNA, which included the transcription factor target site in close proximity to the fluorophore. The number of sites, m , exists as a sliding window of 8 bp, starting at 1 where the 5' end contains the fluorophore. Work in the lab by L. Staby had identified that for CREB 285-339, a fluorescently labelled oligonucleotide where the target site started at $m = 3$ sites away (i.e. 2 bp) from the fluorophore, yielded the highest amplitude change upon protein association. The probe DNA of various lengths were produced using the pHSG299 plasmid as a template, making sure no CRE full or half sites were present 3' to the expected CRE site (see Table 2.10 for details).

PCR was employed to extend the lengths of the 31 bp oligo to a maximum length of 155 bp. Longer DNA oligo's were synthesised but issues arose with analysis of that data, with large errors being obtained for k_{app} using single exponential equations for the fitting. Above ~ 100 bp there was a potential for bending of the DNA, which could cause a complication of the analysis by making intersegmental transfer possible. Following a purification step with a nucleotide removal kit, agarose gel electrophoresis was used to ascertain the purity of the DNA sample (Figure 6.2). A clear ladder of DNA lengths can be seen without smearing, indicating that the PCR extension and subsequent purification is a suitable method for obtaining the labelled DNA of various lengths. The yield varied between the DNA lengths, with the smaller lengths having larger yields (4 mg compared to 10 μg for example).

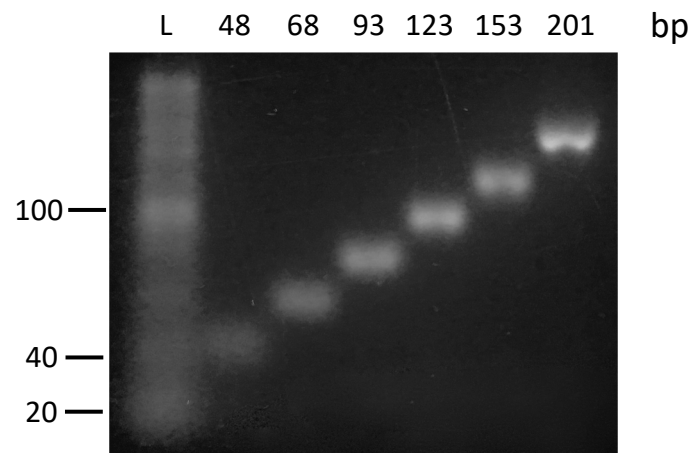


Figure 6.2: Probe DNA purity. Forward primer contains AlexaFluor488 at 5' end. False colour agarose gel electrophoresis of the purified labelled DNA constructs analysed on a UV/vis reader. L = DNA ladder, in bp.

Optimisation of conditions for analysis of the pseudo first-order apparent rate constant

The initial setup for the stopped flow experiments followed the protocols designed by Esadze *et al*¹⁴⁹. All protein concentrations from this point are referred to as concentration of the dimer, unless otherwise noted. A solution of protein was rapidly mixed with a DNA solution containing the labelled probe DNA and competitor DNA as shown in Figure 6.3a. Examples of the resulting fluorescent traces are shown in Figure 6.3b, indicating that a sufficient amplitude change across all of the DNA lengths is present and can be used for determination of the target search parameters.

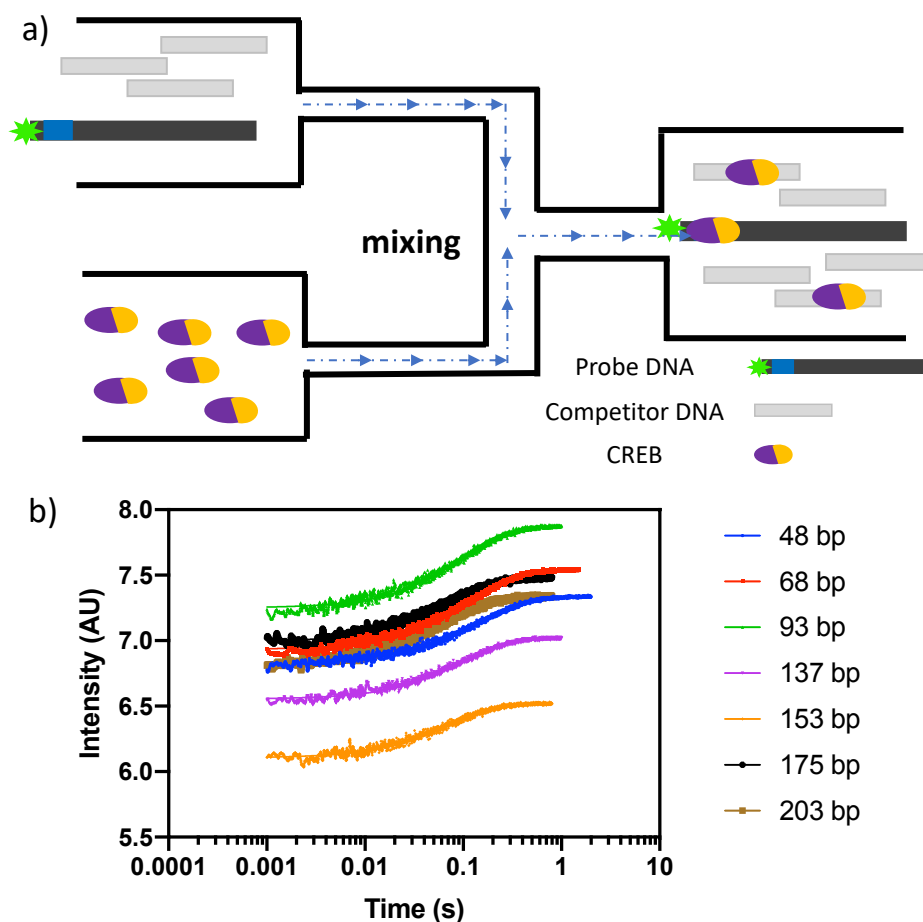


Figure 6.3: Experimental setup for the stopped flow experiments to determine the target search parameters for CREB bZIP. a) Illustration of the mixing strategy for the stopped flow experiments. DNA constructs were in one syringe and the protein in the other, which are mixed 1:1 in the sample cell. b) Representative averaged stopped flow fluorescence traces of the mixing of 100 nM CREB 285-339 FD (assuming dimer concentration) with 5 nM AF488 probe DNA and 4 μ M competitor DNA.

An average of multiple (e.g. 15-40) stopped flow kinetic traces was fitted to single and double exponential equations (Figure 6.4 and Appendix Figure S23). This may represent a failure of the model, or the assumptions made in deriving the equations in Esadze *et al* are an oversimplification¹⁴⁹. Residual traces for single exponential fits showed systematic deviation from zero, however they were not much improved by double (or triple) exponential fits (Figure 6.4 and Appendix Figure S23).

An examination of the various methods of mixing (1-protein to probe DNA only, 2-protein to probe & competitor DNA, 3-protein & competitor DNA to probe DNA) revealed a significant decrease of the k_{app} when competitor DNA

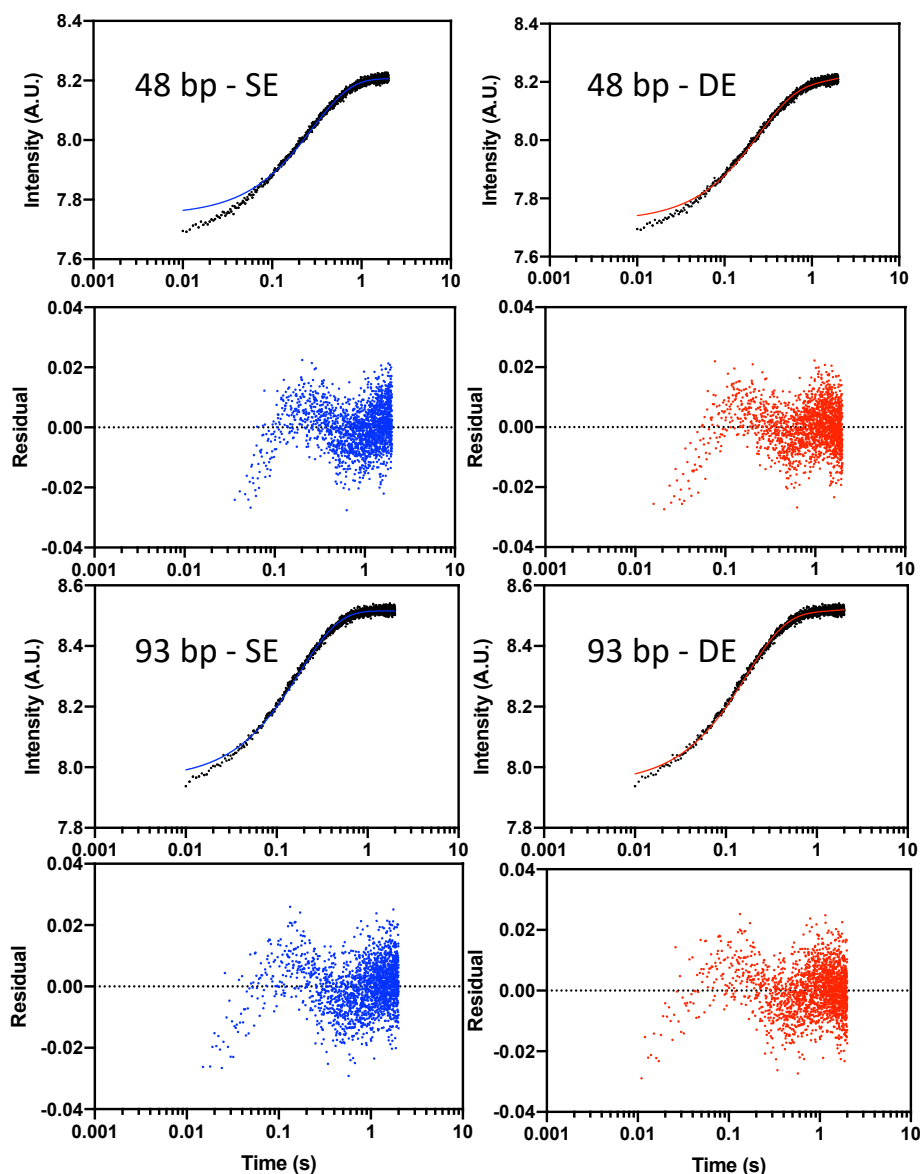


Figure 6.4: Stopped flow of CREB mixed with probe and competitor DNA was challenging to fit to a single or double exponential fit. Examples of single (SE) and double exponential (DE) fits for a selection of DNA probe lengths with the residuals shown below for each.

was present (Figure 6.5). This decrease in k_{app} was independent of whether the competitor was pre-mixed with probe DNA or protein. This work was completed using CREB 277-339 as the main construct.

Comparing CREB 277-339 pWT and FD, there was no difference in k_{app} or amplitude between the two protein constructs with each of the conditions that were analysed. This gave the confidence that the potential of the pWT construct to

associate as a monomer was either not occurring, due to the same k_{app} or amplitude as FD, or that the monomer association occurs at the same rate as a forced dimer.

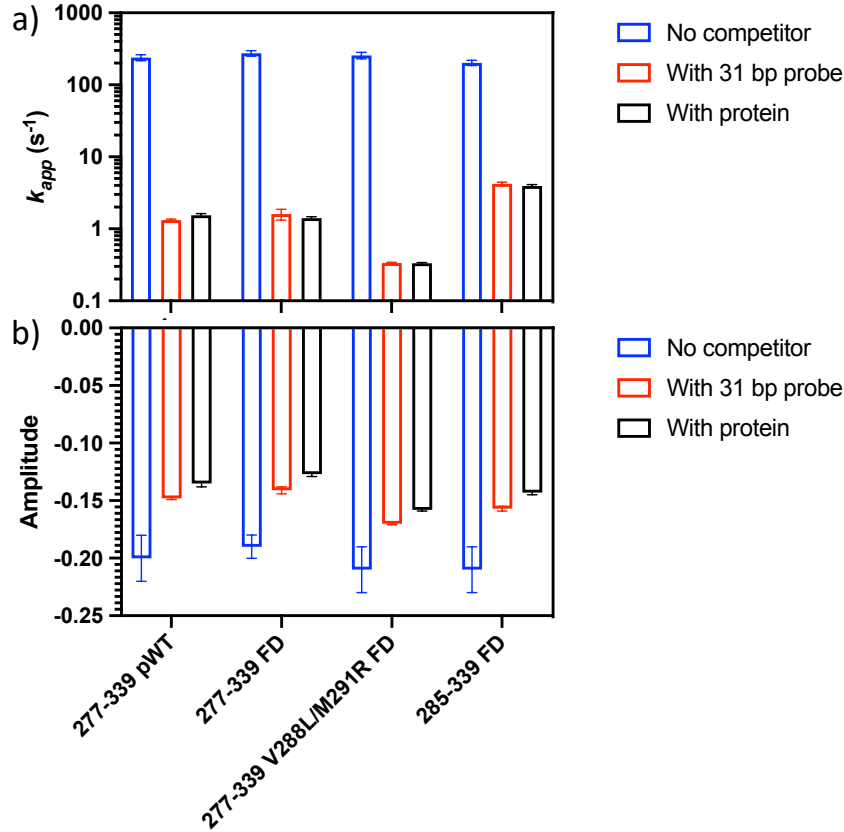


Figure 6.5: Significant slow-down in CREB target search was seen in the presence of competitor DNA. a) k_{app} of CREB constructs (50 nM) mixed under various conditions ([Probe DNA] = 2.5 nM, [Competitor DNA] = 2 μ M) was determined via stopped flow kinetics. Legend shows where the competitor DNA was before mixing. b) Same as a), showing the amplitude instead of k_{app} . All errors are SEM from a single exponential fits of the stopped flow data, n ≥ 16.

The CREB 285-339 FD associated to target at the same rate as the 277-339 FD without competitor DNA, however when competitor DNA was present the rate was ~ 4 fold faster. This is in contrast to the more helical CREB 277-339 V288L/M291R FD (more details on this construct are provided in section 6.2.2), which saw a reduction in the k_{app} to the CREB 277-339 FD by ~ 5 -fold (Figure 6.5a). The contrast between the more disordered 285-339 FD and the CREB 277-339 V288L/M291R FD (compared to 277-339 FD) was striking.

Effect of ionic strength on rate constants

Extracting k_{app} from these fits was possible and one of the first applications of this was to examine the effect of ionic strength. At a higher ionic strength (0.262 M) the k_{app} increased (k_{app} being the rate constant extracted from single exponential fits from the kinetic traces, Figure 6.6a and b). Amplitude change was independent of ionic strength and DNA probe lengths. Here, the single exponential rate constant is dominant, as seen high similarity to the fast phase rate constant in the double exponential fits. The k_{app} was dependent on the number of sites of the DNA being present on the probe DNA (Figure 6.6c).

k_{app} is related to the apparent second order rate constant k_a , by taking into account the protein concentration. Fitting k_a against variable probe DNA length assuming intersegmental transfer is occurring is shown in Figure 6.7, using scripts kindly provided by Prof. J. Iwahara. The parameters from the fit were the effective sliding length (λ) and the 1D diffusional coefficient (also known as the sliding rate). At the lower ionic strength (0.185 M), the sliding rate was ~ 50 times lower than the rate at $I = 0.262$ M. The effective sliding length, which is the average length of DNA that the protein will slide along investigating each discrete non-target site before dissociating, was ~ 13 bp longer at $I = 0.185$ M compared to $I = 0.262$ M.

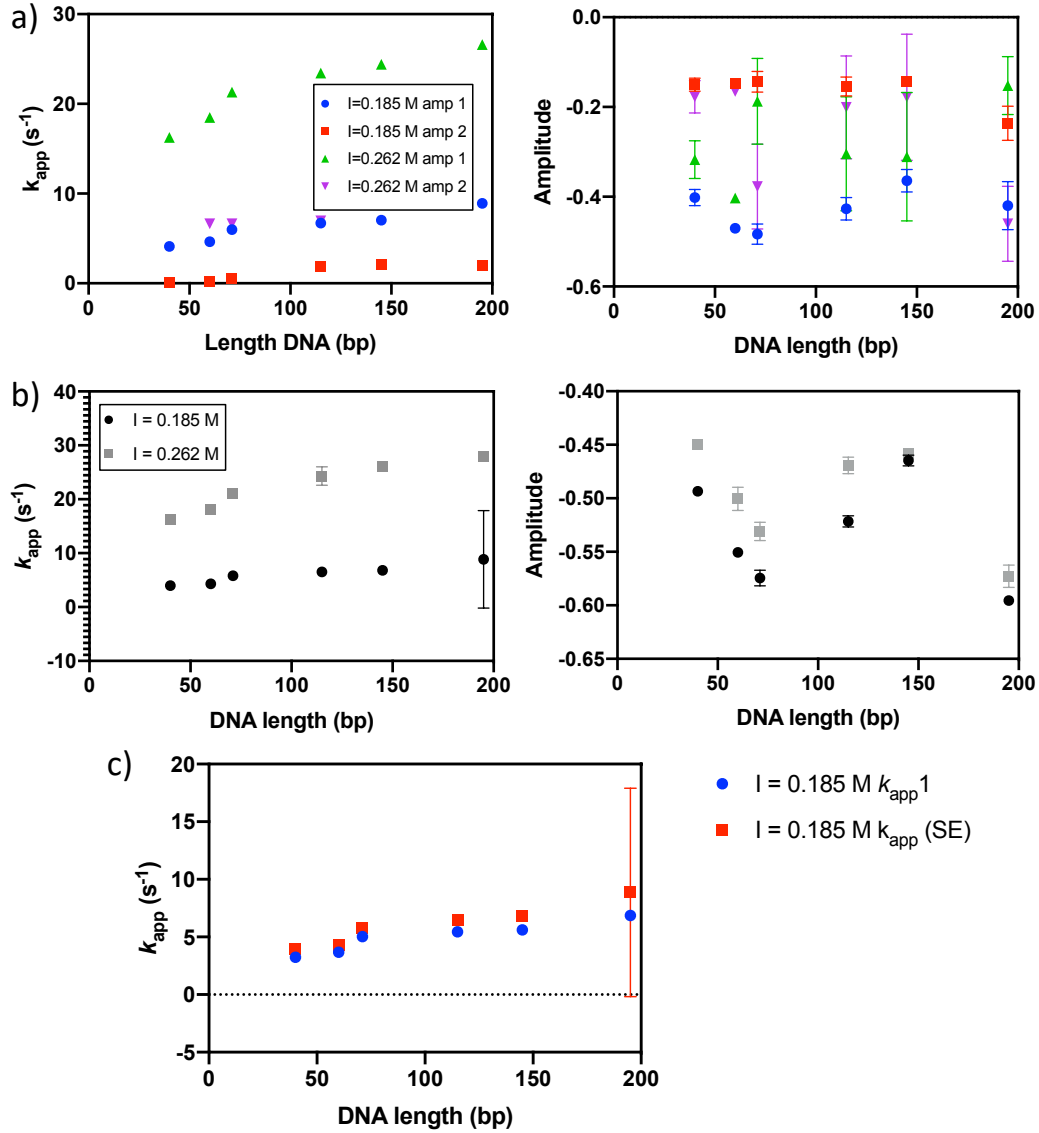


Figure 6.6: k_{app} was increased for CREB 285-339 FD associating to various lengths of CRE-containing probe DNA. a) Left - k_{app} of a double exponential fit to the data from 100 nM CREB 285-339 FD was mixed with 5 nM probe DNA and 4 μ M competitor DNA. Right - the amplitudes of the data in the left panel. b) Single exponential fit of the data in a) (left and right corresponding to panel a). c) Comparison between the largest amplitude determined from the double exponential fit and the single exponential fit for the $I = 0.185$ M condition. Error bars are SEM, $n \geq 59$.

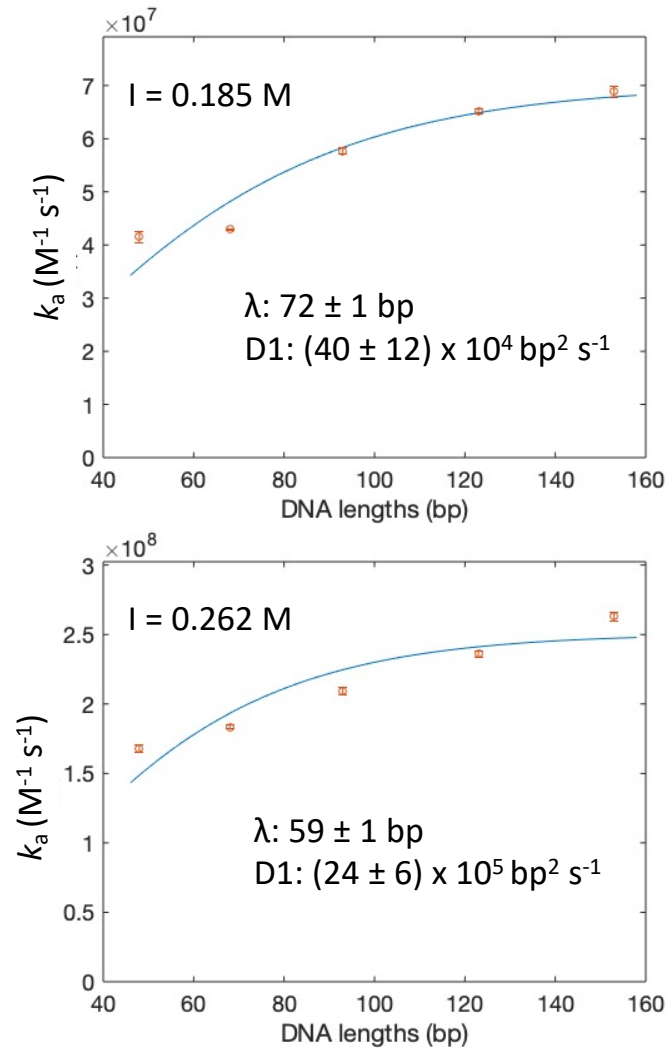


Figure 6.7: Ionic strength increases the rate of 1D diffusion target search in the presence of competitor DNA. k_a against probe DNA lengths of 100 nM CREB 285-339 FD mixed with 4 μM competitor DNA and 5 nM probe DNA of various lengths. Top - stopped flow experiment performed at $I = 0.185 \text{ M}$, bottom - performed at $I = 0.262 \text{ M}$. λ is the effective sliding length, $D1$ is the 1D dimensional coefficient/sliding rate. Errors are SEM.

6.2.2 Characterisation of target search of CREB bZIP

Analytical form of k_{app} is consistent with a linear dependence on protein concentration

One of the experiments that Esadze *et al* used to verify their analytical forms of k_{app} was to measure the k_{app} against varied protein concentration. Under these conditions, the probe DNA length and competitor DNA concentrations were kept constant but in all cases, the pseudo first-order conditions were maintained throughout. Both CREB 285-339 pWT and FD were used, with the k_{app} extracted from single exponential fits of stopped flow traces. This showed a linear dependence of the k_{app} against [CREB] (Figure 6.8a), which corresponds to an apparent second order rate constant, k_a , to target DNA.

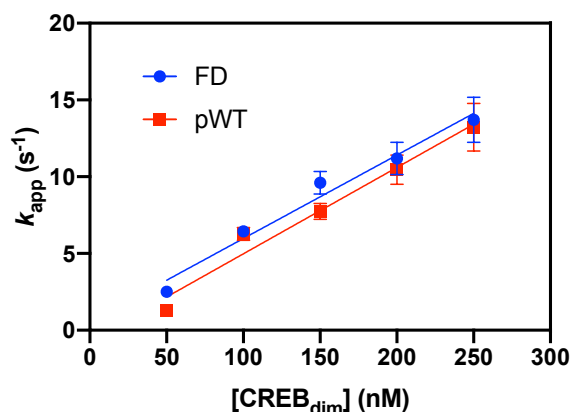


Figure 6.8: A linear dependence on the k_{app} on protein concentration was seen for CREB 285-339 pWT and FD during stopped flow with probe DNA in the presence of competitor DNA. 5 nM 48 bp probe DNA and 4 μM competitor DNA. Data was fitted to a single exponential to extract k_{app} and a linear regression was performed without forcing through the origin. Errors are SEM, $n = \geq 12$.

k_a was found to be the same for the forced dimer construct, at $(54 \pm 6) \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ and the pWT at $(56 \pm 6) \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$. The Y-intercepts were the same within error, with both going through the origin within error, consistent with a low k_{off} . Thus, in the presence of excess competitor DNA, the pWT construct is not slower to associate to the target due to the possibility of forming monomeric species. The linear dependence was similar to the work on Egr-1¹⁴⁹. Subsequent experiments

were performed at a single protein concentration, with k_{on} inferred from k_{app} . This allowed investigation of other parameters within a reasonable timeframe.

Intersegmental pathway followed by sliding is a major translocation pathway for CREB

With the stopped flow experiments of CREB bZIP 285-339 satisfying the analytical forms of the k_{app} equation, similar to Esadze *et al*, the next stage was to determine the rate of intersegmental transfer in the model. To accomplish this, the competitor DNA concentration was varied to observe its effect on k_{a} (Figure 6.9) with CREB 285-339 pWT and FD as the proteins of interest.

The script provided by Prof J. Iwahara takes into account two models whereby intersegmental transfer does or does not occur. In Figure 6.9, datasets for each construct appear to be modelled best by the equation taking intersegmental transfer into account. Interestingly, CREB 285-339 pWT follows this pathway with a similar rate to the FD variants (Table 6.2). However, when looking at the competitor DNA concentrations across a larger range (i.e. with lower and higher competitor DNA concentrations than those in Figure 6.9), both models of target search taking into account the presence or absence of intersegmental transfer fails to reliably fit the data (Figure 6.10).

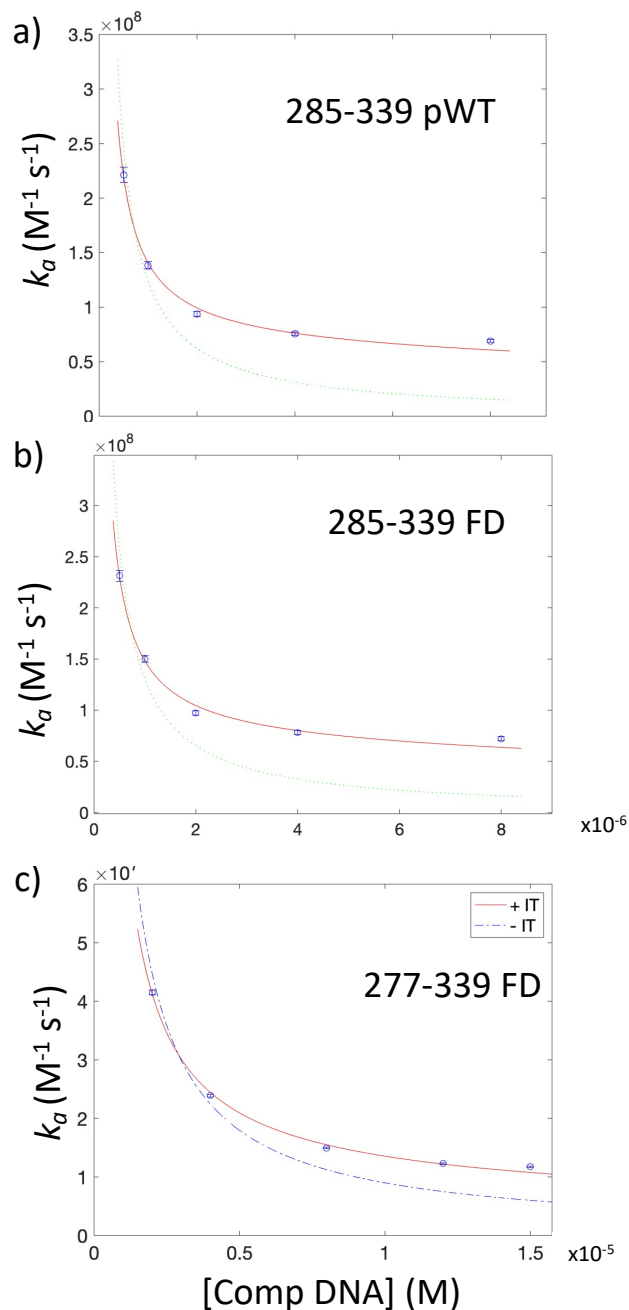


Figure 6.9: Varying the competitor DNA showed that the intersegmental transfer (IT) is a predominant mode of translocation for CREB. a) Rate of intersegmental transfer of CREB 285-339 FD was determined via stopped flow of 100 nM dimeric CREB, mixed with 5 nM 48 bp probe DNA and varied competitor DNA concentration. Data was fitted using J. Iwahara scripts¹⁶⁸ where k_a was determined from single exponential fits of the stopped flow data. b) Similar to a), but with CREB 285-330 pWT used. c) Similar to a-b), but with CREB 277-339 FD as the protein. Errors are SEM, $n \geq 15$.

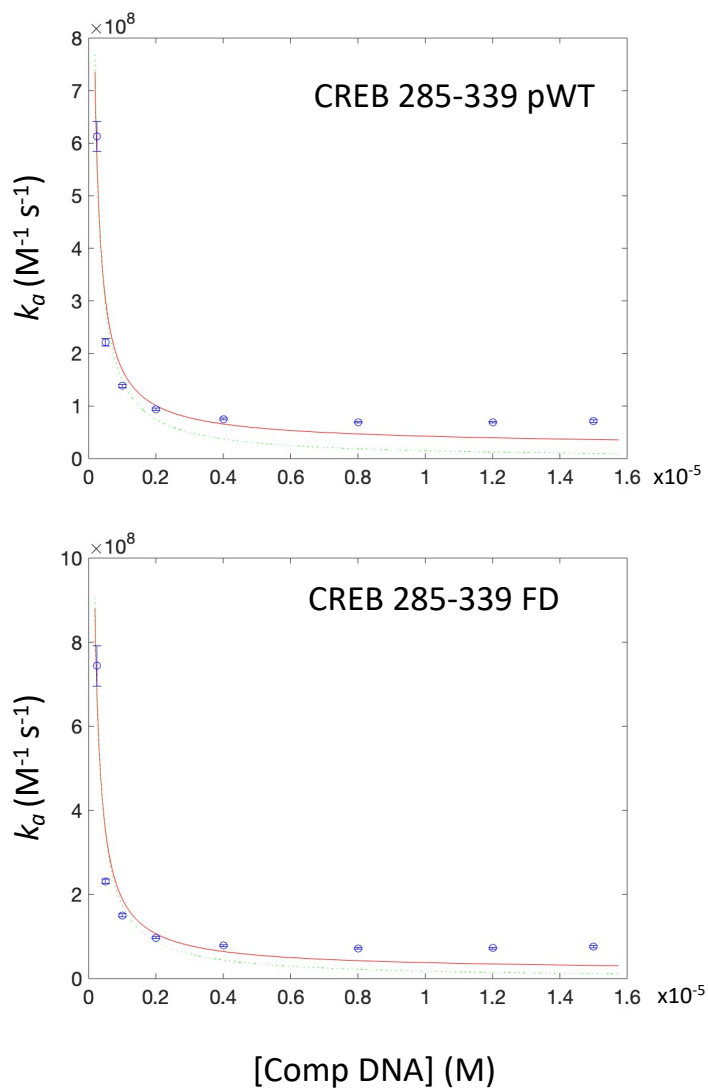


Figure 6.10: Varying the competitor DNA showed that the intersegmental transfer (IT) and sliding model filters for CREB at extreme concentrations of competitor DNA. a) Rate of intersegmental transfer of CREB 285-339 pWT was determined via stopped flow of 100 nM dimeric CREB, mixed with 5 nM 48 bp probe DNA and varied competitor DNA concentration. Data was fit using J. Iwahara scripts (Appendix) where k_a was determined from single exponential fits of the stopped flow data. b) Similar to a), but with CREB 285-330 pWT used. Red line with intersegmental transfer, blue line without intersegmental transfer. Errors are SEM.

Table 6.2: Kinetics parameters identified from the stopped flow experiments with CREB bZIP constructs to probe DNA with varied concentrations of competitor DNA from Figure 6.10. Errors are SEM. * $k_{\text{on},N}$ can be estimated from k_{off} and $K_{d,N}$, with K_d of CREB to non-target DNA being 250 nM for 285-339 constructs and 176 nM for 277-339.

Protein	$k_{\text{it},N}$ ($\text{M}^{-1} \text{s}^{-1}$)	$k_{\text{off},N}$ (s^{-1})	$k_{\text{on},N}$ ($\text{M}^{-1} \text{s}^{-1}$ *)
285-339 pWT	$(17 \pm 2) \times 10^5$	42 ± 3	$2.8 \pm 0.2 \times 10^8$
285-339 FD	$(18 \pm 2) \times 10^5$	45 ± 3	$3.0 \pm 0.2 \times 10^8$
277-339 FD	$(37 \pm 4) \times 10^4$	45 ± 3	$1.8 \pm 0.1 \times 10^8$

The $k_{\text{off},N}$ was determined to be similar between the protein constructs used when varying the competitor DNA concentrations during mixing. From the known K_d from section 4.2.2, the $k_{\text{on},N}$ can be estimated which is also fairly similar for each construct.

A V288L/M291R mutant of CREB bZIP significantly increases the residual helical propensity

To investigate the role of disorder in the sliding rate of the CREB bZIP, several constructs were used. Two of these have been discussed extensively, these being CREB 285-339 and 277-339. One of the other constructs used, CREB 277-339 V288L/M291R, was designed by Dr. M. Kuravskiy from DNA sequence alignments of expressed sequence tags during a bioinformatic screen to intentionally increase helical propensity. CREB 277-339 V288L/M291R underwent NMR analysis in order to determine the secondary structure on the residue specific level, using the same suite of triple resonance experiments used for the apo-state proteins throughout this thesis.

The assigned ^{15}N -HSQC of CREB 277-339 V288L/M291R is shown in Appendix Figure S24, with all residues assigned aside from K303. The side chain amide groups of Gln and Asn were not assigned due to low signal to noise in the triple resonance experiments. The chemical shift table and assignment report are in Appendix Tables S34 and S35 respectively. SSP was utilised to determine the secondary structure at the residue specific level (Figure 6.11a) which identified a marked increase in helicity within the basic region residues 277-296 over CREB 277-339 pWT. The differences from residue 296 onwards are not significant, following an near identical trend. The ^1H - ^{15}N Heteronuclear NOE agrees with the SSP prediction,

showing that alongside the increased helicity there is a decrease in the dynamics of those residues in the apo state. The leucine zipper peaks overlay between the pWT and variant, indicating that the differences in the basic region are not due to a change in structural propagation from the leucine zipper.

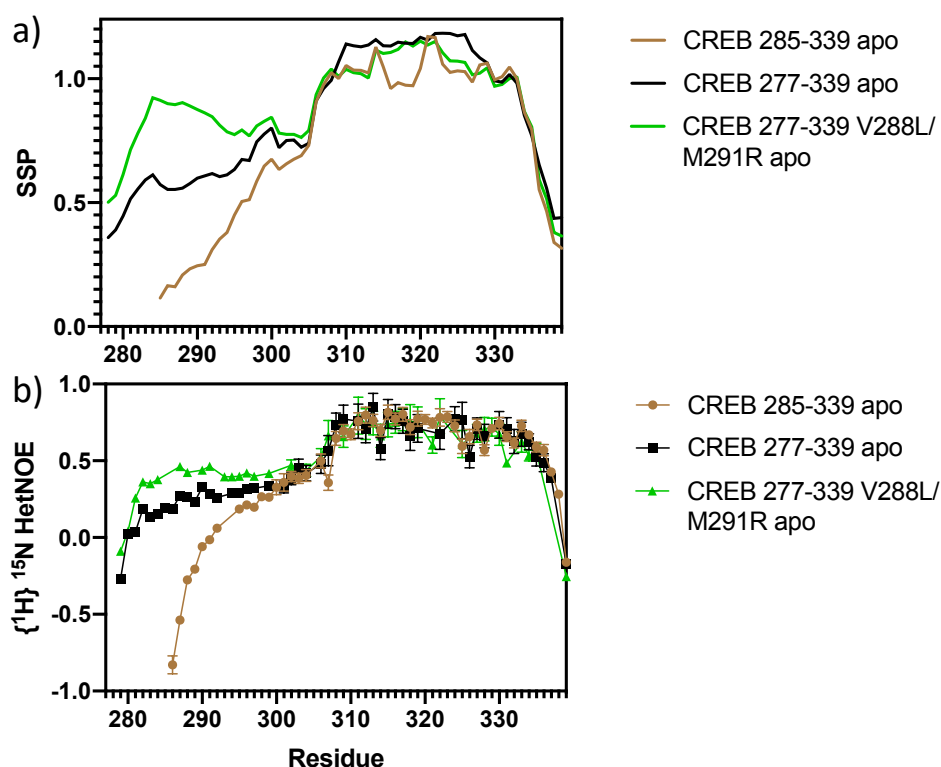


Figure 6.11: Marked increase in secondary structure propensity and reduction in dynamics of the V288L/M291R mutation of CREB bZIP. a) SSP prediction of CREB 277-339 V288L/M291R plotted alongside the two other apo-state bZIP proteins that have been calculated. $C\alpha$ and $C\beta$ only were used in this plot. b) $\{^1\text{H}\} \text{ } ^{15}\text{N}$ Heteronuclear NOE of CREB 277-339 V288L/M291R plotted alongside the two other apo-state proteins used throughout this thesis. Errors are SD. All data were collected at 25°C and with a 600 MHz spectrometer.

Kinetics of 1D diffusion of CREB bZIP is dependent on the residual helical structure in the basic region

1D diffusion, or sliding, is a prevalent mechanism for DNA binding proteins to find their target site on the DNA. To determine the role of protein disorder in the mechanism of sliding, four constructs based on the bZIP domain were utilised.

The constructs used were CREB 285-339 FD, 277-339 pWT and FD as well as CREB 277-339 V288L/M291R FD.

To investigate the role of the residual helicity in each CREB construct, k_{app} for various lengths of the probe DNA was analysed by stopped flow fluorescence, using the apparent pseudo-first order conditions set out in previous sections of this Chapter. The kinetic traces were fitted to a single exponential to extract the k_{app} , which is shown in Figure 6.12.

With all constructs, the plot of k_{app} against DNA probe length was able to be fitted with the Veksler and Kolomeisky derived model by J. Iwahara¹⁴⁹. However, the fit for the CREB 277-339 construct variants was slightly poorer than the 285-339 construct. The $k_{\text{sl,N}}$ (or D1) and λ (effective sliding length) can be extracted for all of the constructs (Figure 6.12b and c), Table 6.3). A comparison between the CREB 277-339 pWT and FD constructs shows a very slightly faster search time for the FD construct, however within error this is not significant. There is a marked reduction in the search time for the V288L/M291R construct, which is the most helical construct in the apo-state. In contrast, the 285-339 FD construct has an order of magnitude faster sliding rate than the 277-339 FD construct.

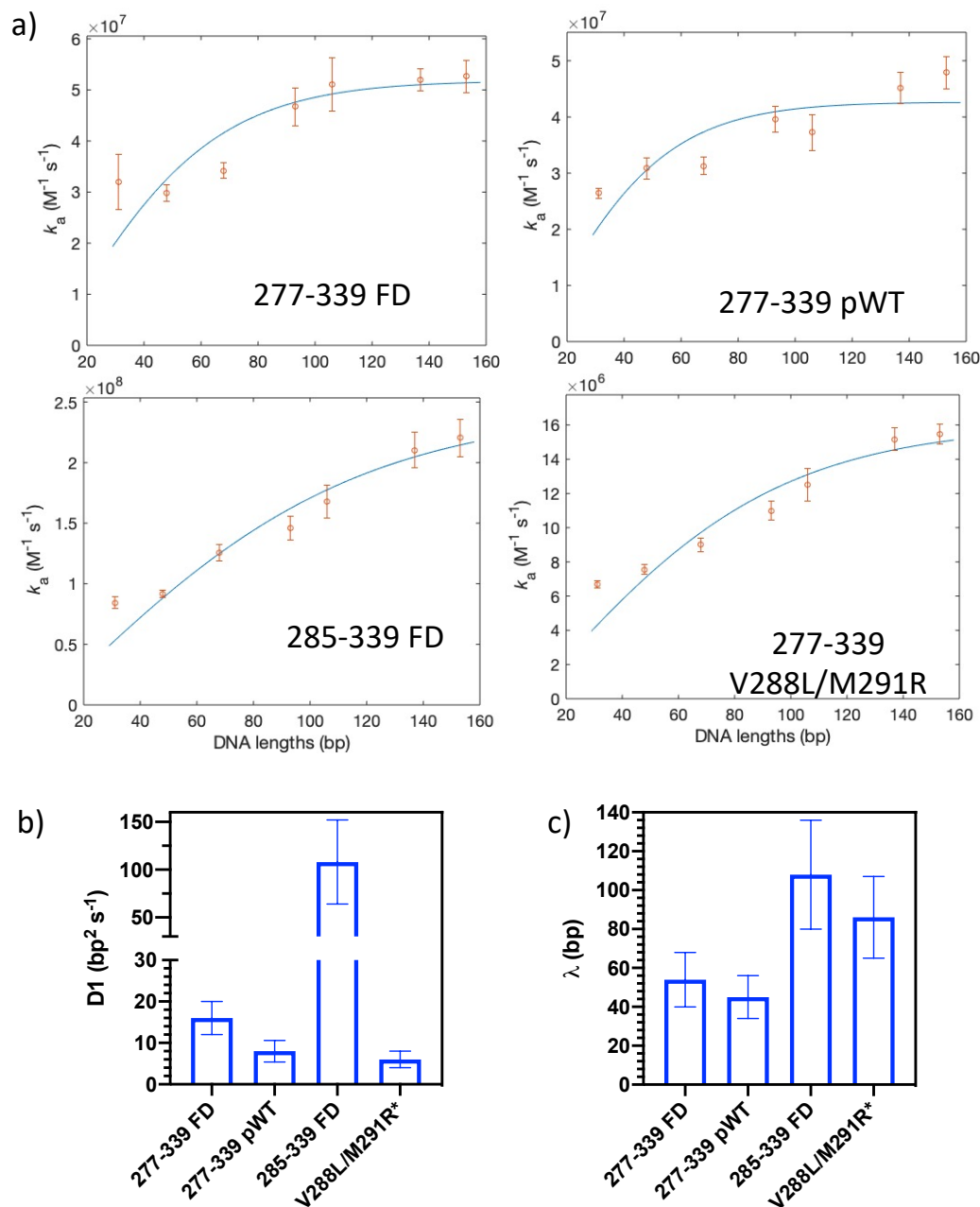


Figure 6.12: Disorder increases the rate of sliding of the disordered DNA binding domain of CREB, but not the effective sliding length. a) Plots of k_{app} vs length of probe DNA of 50 nM dimeric CREB mixed with 2.5 nM probe DNA and 2 μ M competitor DNA. FD is forced dimer, pWT is pseudo wild type. Data was fitted with the equations from Prof J. Iwahara¹⁴⁹. b) 1D diffusion coefficient is dependent on CREB's residual helicity, extracted from the fit in a). c) Similar to b), but plotting the effective sliding length. All errors are SEM, $n \geq 14$.

Despite this, the sliding length is very similar between constructs. For instance, the λ for 285-339 FD is statistically significantly higher than 277-339 pWT/FD, where it is not with V288L/M291R mutant construct. λ for the 285-339 FD construct

Table 6.3: Extracted parameters for the sliding rate (D1) and effective sliding length (λ) from stopped flow experiments of CREB bZIP constructs against various lengths of probe DNA. Errors are SEM.

Protein	λ (bp)	D1 ($\times 10^4$, $\text{bp}^2 \text{ s}^{-1}$)	K_d (non-specific DNA, nM)
CREB 277-339 FD	54 ± 11	11.6 ± 4.0	176
CREB 277-339 pWT	45 ± 14	7.6 ± 2.6	153
CREB 285-339 FD	108 ± 28	109.2 ± 44.6	189
CREB 277-339 FD V288L/M291R	86 ± 21	5.6 ± 2.0	20

was twice as high that those for the 277-339 pWT and FD constructs. Large errors are reported with the extraction of the λ parameter however (Table 6.3).

6.2.3 Computational and structural examination of the effective sliding length of CREB bZIP

Following the kinetics experiments which identified a 50 bp effective sliding length for the bZIP construct which was designed to reflect most accurately the full-length WT CREB, CREB 277-339, the properties of the DNA sequence began to be investigated. *In vivo* CREB must find its target sequence within 3×10^9 bp, therefore was there a defined characteristic within the flanking regions of the CRE target sequence which enabled effective target search?

To investigate this, every occurrence of the somatostatin CRE-motif (5'-TGACGTCA-3') was found in a human genome dataset. A region around the occurrence of each CRE sequence was extracted and used to determine the average GC content at each position flanking the target site (Figure 6.13). The number of targets identified can be seen in Table 6.4. For CRE, 19394 sites were extracted from the human genome dataset. When filtered for those identified as being close to a known CREB

target gene using data from Zhang *et al*⁹⁶, this reduced to 272 sites. This was less than those found by Zhang *et al*, due to gene synonyms not being taken into account.

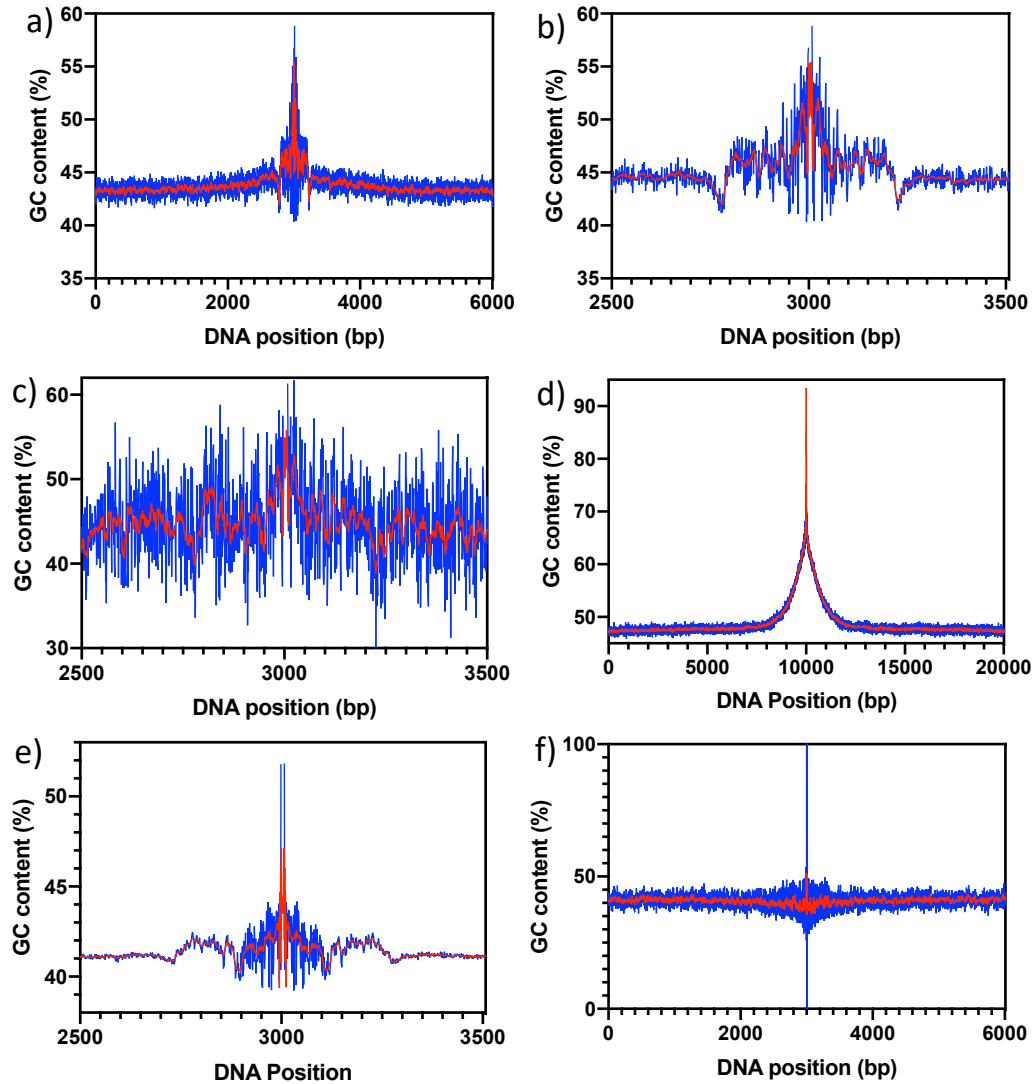


Figure 6.13: GC content flanking various *H. sapiens* transcription factor DNA target sequences is higher than the genome average. Flanking sequences of 3000 bp were extracted from transcription factor DNA target sequences (CRE: 5'-TGACGTCA-3', AP-1: 5'-TGA(G/C)TCA-3', Egr-1: 5'-GCGTGGGCG-3' and random control: 5'-CGTATACG-3') from human genome dataset hg16. Panels a) - c) show searches for CRE, with a-b) being from whole genome wide search for the target and c) being data filtered as being close to known full CRE CREB target sites⁹⁶. d-f) Data from whole genome search of Egr-1, AP-1 and random control respectively. Blue is raw data, red is a sliding window average of the raw data with $n=10$.

Table 6.4: Sequences of the target genes which were used for the characterisation of the flanking sequences of various transcription factors.

Target Sequence (5'-3')	Name	Transcription Factor	Genome Dataset	Number of sites found (total)
TGACGTCA	somatostatin CRE	CREB	Homo sapiens - hg16	19394
TGACGTCA	somatostatin CRE	GCN4	Yeast S288C - R64-2-1-20150113	234
GCGTGGGCG	Egr-1 target	Egr-1/Zif268	Homo sapiens - hg16	26998
TGA(C/G)TCA	AP-1	Fos/Jun, GCN4	Homo sapiens - hg16	46248
TGA(C/G)TCA	AP-1	GCN4	Yeast S288C - R64-2-1-20150113	1976
AATTGTGAGCGGATAACAATT	Lac operator	LacI	E. coli - K12	2
AGCGTTTC	Control 1	n/a	Homo sapiens - hg16	38246
AGCGTTTC	Control 1	n/a	Yeast S288C - R64-2-1-20150113	222
CGTATACG	Control 2	n/a	Homo sapiens - hg16	1171
CGTATACG	Control 2	n/a	Yeast S288C - R64-2-1-20150113	204
GGCTAAATTAGCC	Control 3	p1-DNA (from lab)	Homo sapiens - hg16	6

Focusing on the whole genome data, for each of the three target sites that were focused on (CRE, AP-1 and Egr-1 target DNA sequences), there was a common increase of GC content flanking the target site (starting at 3000 bp in each case). With CRE flanking sequences, there is marked increase in GC content from ± 50 bp away from the centre of the target sequence, regardless of having the datasets from the whole genome or filtered to known CREB target sites. The effect is more pronounced with the whole genome dataset. The highly symmetric behaviour reveals a small decrease in GC content beyond the 50 bp increase in GC content also. The DNA target sequence for Egr-1 is very high in GC content, but remarkably searching the whole genome for target sequences shows a 400 bp range over which the GC content increases towards the midpoint. In contrast, pseudo-symmetric sequence containing 50% GC content does not show this corresponding increase in GC content. Investigations in other organisms (lac operator in *E. coli* and AP-1 in *S. cerevisiae*) did not show a similar behaviour (Appendix Figure S25).

As a higher than genome average of GC content exists ± 50 bp of the CREB target, this led to the question of why does this exist and is there a functional role of this sequence characteristic that aids CREB in the target search process. Also, at this time discussions with Prof. C. Redfield around the design of the non-CRE DNA that was used for the NMR (and biophysics) experiments (5'-)GGCTAAAGCATTCT-3') may lead to the difficulties in assigning these protein-DNA complexes due to broadening effects. To alleviate these, Prof C. Redfield suggested designing a pseudo-symmetric version of a non-target sequence. The basis for the design of these pseudo-symmetric sequences was the original non-CRE DNA and resulted in the following sequences taking the first and last 7 bp of the non-CRE sequence: non-CRE p1 = 5'-GGCTAAATTTAGCC-3' and non-CRE p2 = 5'-AGAATGCGCATTCT-3'.

Similar to that described in section 3.2.3 for assignment of CREB using triple resonance methods, the complex of CREB 277-339 with non-CRE p1 DNA was examined by NMR. A lower protein concentration was used (monomer = 300 μ M compared to a minimum of 800 μ M for those in Chapter 4.1) with a saturating concentration of non-CRE p1 DNA as determined via multiple titrations (method

shown previously). Assignments of the CREB 277-339 with non-CRE p1 DNA are shown in Figure 6.14. A comparison between p1 and original non-CRE DNA

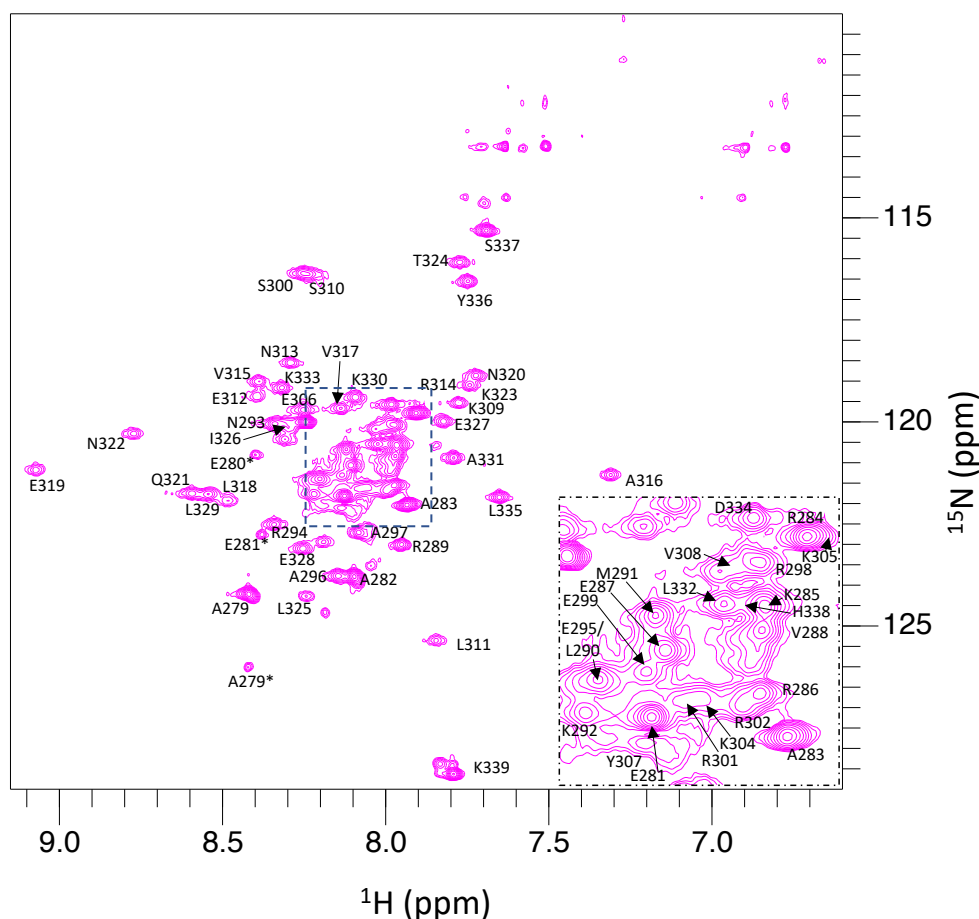


Figure 6.14: Assigned CREB 277-339 with non-CRE p1 DNA at 35°C. ^{15}N -BTROSY of CREB 277-339 in complex with saturating concentration of non-CRE p1 DNA. Data was collected at 35°C.

sequences of the H^{N} and ^{15}N peak positions revealed that in the non-CRE p1 titration, CREB peaks migrated on a similar trajectory as the peaks in the non-CRE DNA titration, but saturation occurred at a position intermediate between apo and non-CRE (Appendix Figure S26). The chemical shift table and assignment report are shown in Appendix (Appendix Table S36 and S37). A check was made using the ^{15}N -HSQC of the CREB 277-339 with non-CRE complex by diluting that sample to an equivalent protein concentration to see if the peaks shifted due to concentration dependent effects; this was not observed (Appendix Figure S27).

The SSP of the non-CRE p1 DNA complex with CREB 277-339 was challenging to determine due to a high number of missing $C\beta$ assignments. However, the region from 278-299 contained both the $C\alpha$ and $C\beta$ chemical shifts. The SSP in this region showed an intermediate behaviour of the SSP in-between apo and non-CRE DNA, indicating that on a different non-target DNA sequence, the helical content upon association to the DNA sequence can vary (Figure 6.15).

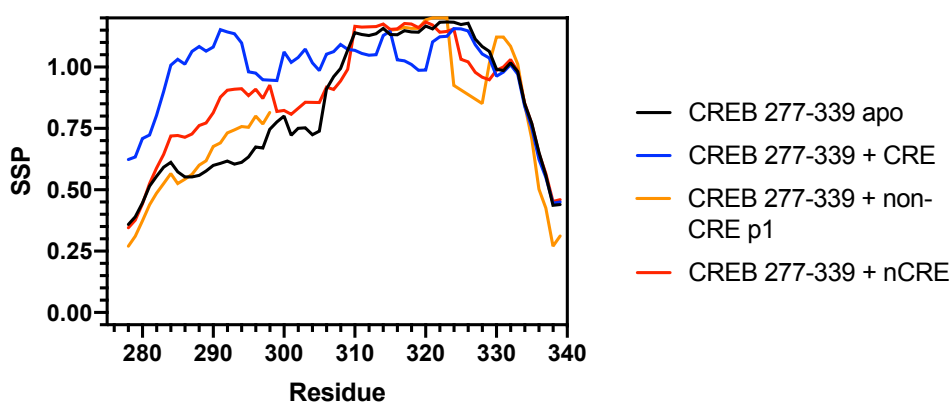


Figure 6.15: Prediction of secondary structure of CREB 277-339 in complex with non-CRE p1 DNA is intermediate between apo and non-CRE DNA. SSP plot of CREB 277-339 using only $C\alpha$ and $C\beta$ of the apo, non-CRE, non-CRE p1 and CRE DNA. The data with DNA was collected at 35°C, apo at 25°C. non-CRE p1 DNA data was missing between 299-309 and is not shown.

6.3 Discussion

The target search process for a bZIP containing protein been investigated to determine if sliding or intersegmental transfer is a potential mechanism to enable the location the target DNA site. Using the discrete state model developed by Veksler and Kolomeisky¹⁹⁹, the stopped flow kinetics setup that was followed in this thesis utilised this model over a continuum model. The continuum model is a root cause for the speed-selectivity paradox, that dictates protein-DNA interactions are homogeneous when the protein translocates along the DNA, leading to a slower target search time as the sliding rate is reduced. The effect on the rate of association thus suggests that the facilitated diffusion explanation for the fast target search

times fails^{139,199}. The continuum model does satisfy a target search model when the effective sliding length (λ) is significantly greater than the target size.

In most cases, experimentally derived λ were ~ 40 bp for p53^{143,192}, regardless if they were derived *in vivo* or *in vitro*, with remarkably different DNA lengths present in each case (maximum 22 bp used by Loth *et al*). Therefore, the discussion surrounding the speed-stability paradox may be invisible experimentally, which was noted as 3D diffusion can still occur even when the 1D diffusion approaches the lower limit (of 0)²⁰⁰. Early work for determining the target search process, particularly with LacI, used filter binding assays of protein/protein-DNA complexes with ³¹P-DNA. This technique does not take into account free DNA, and investigations into this revealed that multiple operons would produce artefacts with this type of assay²⁰².

The discrete model was the most suitable to use for the approach with CREB bZIP as a model protein, as the protein appears to recognise discrete sites. Association of CREB bZIP to different non-target sequences (a random and a symmetrical non-target sequence) revealed different helical contents in the basic region using NMR. The ¹³C-HSQC revealed a slight difference in the range of arginine C δ dispersion between the two DNA sequences (Appendix Figure S28). NMR with a range of different non-target DNA sequences in complex with LacI showed a small range of K_d values, but the CSP showed large differences based on the underlying DNA sequence¹⁹². Ideally, this range of DNA constructs would be examined with CREB to show if the change in helical content is significant or not, but the evidence from CREB supports the use of a discrete DNA binding model for DNA target search.

Various advantages and disadvantages of the stopped flow approach used in this thesis over single molecule microscopy were initially discussed. With single molecule techniques, much longer lengths of DNA strands can be investigated, and combined with optical tweezers the shape (and hence conformation) of the DNA can be modulated based on the force applied¹⁴¹. Investigating the effects of a single molecule allows investigations on a sub-ms timescale, with NMR experiments revealing interesting characteristics of how proteins perform the target search on the μ s-ms timescales. Three binding modes of LacI were seen, corresponding to

target bound with high affinity, target bound with low affinity and sliding on the DNA, based on the velocity and position of the protein. Based on the position, the sequence characteristics of that particular region will be known and can be examined, which is important as this can show the free energy landscape based on the non-target DNA sequence¹⁵⁴. However, intersegmental transfer would be difficult to visualise with this experimental set-up, as a second DNA strand would be freely diffusing in solution. Here, a reduction in the force applied could lead to DNA looping, which has been theorised as a novel mechanism for target search but as of yet not determined experimentally¹⁴⁰, and itself may be indistinguishable from intersegmental transfer. *In vivo* analysis can also be performed using single molecule approaches. Single molecule imaging has been performed for CREB but did not distinguish freely-diffusing and sliding species²⁰³.

Stopped flow apparatus on the other hand allows the use of pre-mixed or un-mixed components at a very high throughput in comparison to single molecule microscopy. With this approach, the sliding or intersegmental arms of the target search can be investigated by varying the length of a probe DNA or the competitor DNA concentration respectively. By maintaining pseudo-first order conditions during mixing, the k_{app} could be used to examine the target search process of proteins to DNA containing the target at any site along the probe DNA^{149,199}. One limitation of this, particularly with CREB bZIP, was that the fluorescent probe needed to be close to the target DNA site, which meant that the probe was towards the 5' end of the DNA construct in all cases. The fluorescence could not be reported from the perspective of the protein, due to the inherent difficulties faced with labelling at a C-terminal position of the bZIP protein, particularly with a photostable dye (section 3.2.2). The effect of the competitor DNA being present (and easily varied), makes the stopped-flow set-up more biologically relevant than an *in vitro* single molecule approach, with at least a 10-fold excess of competitor being present.

With regards to CREB, the rate of intersegmental transfer appears dependent on the helical content present in the basic region, rather than the ability of the protein to sample the monomeric and dimeric pathways, by a factor of ~ 4 -fold.

The intersegmental transfer rate was a similar order of magnitude for the pWT compared to the FD (CREB 285-339 only), indicating that there is no observable kinetic advantage for the protein to search as a monomer within these experimental conditions. Interestingly, the rate of intersegmental transfer for Egr-1, determined by Esadze *et al*¹⁴⁹ to be $1 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$, is of an similar order of magnitude to CREB 285-339 pWT ($17 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$). The intersegmental transfer rate is reported to be highly dependent on the ionic strength¹⁹⁹, with Esadze *et al* performing the analysis with a much lower ionic strength than used for CREB (90 mM for Egr-1 compared to 185 mM for CREB), which may account for the difference that is observed. An issue arose with the extreme DNA concentration values that were used with CREB when attempting to fit the data using the equations set out by Esadze *et al*. The low DNA concentrations can be answered by not following the pseudo first-order conditions set out for this type of experiments, however fitting of the data still did not occur with the high DNA concentrations, indicating that the Esadze *et al* experimental set-up perhaps fails at these concentrations of DNA. Therefore, by removing these extreme DNA concentrations, using inspiration from the range of DNA concentrations used by Esadze *et al* the model appeared to fit successfully.

The rate of sliding for CREB was also an order of magnitude lower than that observed for Egr-1¹⁴⁹. With Egr-1, the pathway that was proposed showed a dominance of intersegmental transfer as the major translocation mechanism, followed by sliding of the protein to the cognate target site. Given the results obtained for CREB, it would appear that a mechanism by which CREB searches for the target initially by intersegmental transfer followed by a sliding translocation step is also the case. Similarly, LacI showed limited DNA length dependency during association to an operator sequence²⁰², indicating a possible enhancement via intersegmental search. Therefore there appears to be a precedent for several transcription factors from different organisms using the intersegmental transfer process as the main translocation mechanism during target search.

The rate of sliding appears highly dependent on the residual helical content in the basic region, with the V288L/M291R mutant construct in CREB 277-339

showing a ~ 2 -fold decrease in the calculated rate compared to the CREB 277-339 FD. Conversely, the shorter, less helical, bZIP construct (CREB 285-339) revealed a significant increase in the sliding rate by an order of magnitude. The binding affinities, and therefore likely k_{off} , for the non-CRE DNA were similar between 277-339 and 285-339 constructs, suggesting that the increased disorder content in the basic region significantly increases the rate of target search via this mechanism, with double the effective sliding length.

These results suggests that reducing the helical structure in the basic region aids in the sampling of the underlying DNA sequence during the search mode, perhaps by lowering the energy barriers during the process¹⁵⁴. It therefore stands to reason that, for disordered DNA binding domains, increasing the amount of disorder content in the domain may in some cases lead to more effective target search of the cognate site without majorly affecting the stability of the protein-target site interaction. A balance may be required *in vivo* for efficient transcription. This could have implications in the design of synthetic transcription factors or inhibitors^{194,204}. To investigate this further, a range of mutants with varying disorder contents would need to be investigated. Previous work has shown that a range of proteins translocate via sliding by following the helical pitch of the DNA²⁰⁵ using single molecule microscopy, therefore it would be reasonable to assume a similar occurrence here. Of course, the modification of the charge on the various protein constructs (+7.5 with 277-339, +8.5 with 277-339 V288L/M291R, + 8.4 with 285-339) could increase the rate in intersegmental transfer but might also expect a decreased sliding rate for 285-339 (which is not observed). Further investigations into whether the change in the charge of the constructs or the helical content (or a combination of the two), are contributing to the changes in target search rates that are seen are required.

As the helical content affects the rate of sliding of the protein along the DNA, this led to a question of what would happen when the DNA sequence itself changes? The large number of non-target sites that CREB would sample during the target search mechanism may influence the structure of the basic region in the bZIP domain, and therefore could influence the sliding kinetics, in particular. This had

been shown previously for several bZIP proteins by CD, in particular GCN4 was shown to have increased helicity when associating to DNA (AP1 > CREHS > NS24)¹²⁸. Santiago-Riveria showed a similar dependence with CRE > AP-1 > lac operator¹²⁷ with regards to the helicity. However, what the work with GCN4 and CREB lacked was information on the residue specific level.

The evidence from NMR with CREB presented here showed that the basic region residues have varied helical contents dependent on the DNA construct that it is in complex with. As discussed above, previous evidence has shown a DNA association landscape that proteins sample¹⁵⁴. The arginine side chains, shown to make specific (to the bases) and non-specific (to the phosphate backbone) contacts to the DNA, proved a valuable tool for elucidating the nature of the interactions within the complex. From the crystal structure of CREB with CRE-DNA²¹, arginine side chains that make non-specific interactions in that complex were R286, R294, R298 and R301. The arginine C δ region in the ¹³C-HSQC showed a subtle difference in the chemical shift dispersion between the symmetric and non-symmetric non-target DNA that was used, but the two non-target DNA sequences led to a significant difference in secondary structure propensity.

One of the possible theories for this behaviour of the residues in the basic region having varied dynamics and disorder content was investigated with an bioinformatics screen of known CREB target sites. ChIP-seq analysis by Zhang *et al* revealed all known CREB promoter regions⁹⁶. Using this information, the CRE target sites had their flanking sequences extracted to determine the average GC content at each position. Within 50 bp of the centre of the CRE site, a significant increase of GC content was seen compared to the regions outside of this region. A large number of CREB targets are known to be regulated by multiple full and half CRE sites⁹⁶, so this may account for the increased GC content, as well as CpG islands^{206,207} which aid in regulation of CREB-induced target gene expression. This behaviour appears to occur with and without filtering for known CREB target sites in the genome as well.

The increased GC content may correspond to nucleosomal depleted regions (NDR) that are known to be proximal or overlapping with core promoter regions.

For instance, the human NDR sequence TTAGGG was found to act as a NDR region *in vivo*²⁰⁸, with a yeast system containing these human NDR regions showing a 20-fold increase in gene expression from the CYC1 core promoter regulating lacZ²⁰⁹. The functional role of NDRs was identified as enacting a high degree of reliability of target gene expression during the cell cycle, whereas promoters without NDRs showed no change in peak expression levels²¹⁰. Therefore, the sequence characteristics identified for the CRE promoter region may be reflective of an NDR region, which would provide reliable control for CREB to target these particular sites.

Comparing to GCN4 sites (AP-1) that were detected in *S. cerevisiae*, there appears to be no GC content dependence in the region flanking each site. In contrast, the same site was used for detecting Fos/Jun (GCN4 homolog) interactions in the human genome and a similar GC content behaviour was observed to CRE, indicating that in higher complexity organisms, this characteristic is important for gene targetting or regulation of expression (see below). Interestingly, the same method used for the Egr-1 target site showed a markedly different behaviour, with a much longer region of increased GC content than that seen for the bZIP proteins. What this signifies is unknown. Further investigation would involve detailed designing of the probe DNA flanking region and observing the effect by stopped flow.

The 50 bp sequence either side of the CRE sites also agrees to match with the experimentally defined effective sliding length (λ) of ~ 50 bp. This would provide a kinetic advantage for CREB to perform effective target search, by possibly creating an antenna for CREB to use to be actively directed towards the target site itself²¹¹. Further studies are required to investigate this characteristic, for instance to perform the sliding experiments with a different DNA sequence to the 3' end of the target site. Presumably, the increased GC content may stabilise the non-target interaction by providing pseudo CRE sites, slowing down the dissociation of the protein in regions near the target, or increasing local CREB concentrations. Analysis of various DNA constructs containing varied GC content by NMR to observe the difference, if any, between them with the helical content in the basic region would be of interest. The errors on the extraction of λ for CREB were large using the

stopped-flow method, possibly due to the poor fitting of the data using the model proposed by Esadze *et al.* Further repeats of the experiment would be required to reduce the experimental error, or an alternative strategy utilised such as the single molecule microscopy described above.

Limited evidence of bZIP target search has been reported before within the context of the full length protein. One single molecule approach using N-terminally Halo-tagged CREB revealed short and long residency times on DNA²⁰³. The short residency times were believed to be reflective of CREB in the target search process, with an experimentally derived off rate from non-target DNA of 2.74 s⁻¹. This rate of dissociation was 10-fold lower than that determined by stopped flow for the bZIP domain, indicating that *in vivo* and with full length CREB, there is potentially a significant difference.

One explanation for this discrepancy in the rate of dissociation could be the accessory proteins that are known to associate with CREB *in vivo*. For instance, the CREB-regulated transcriptional co-activator (CRTC) 2 has been shown to associate to the leucine zipper region of the bZIP domain of CREB²¹² using SEC-MALS in a DNA-dependent manner in a 2:2:1 ratio. This was thought to increase the stability of the CREB bZIP homodimer-CRE complex. In addition to this, another enhancer protein associates to the KID domain of CREB, CEBP, in a S133 phosphorylation dependent manner¹⁰³. CEBP leads to recruitment of the transcriptional machinery as it is sufficient for activation of gene expression²¹³. The formation of the transcription initiation complex would aid in stabilising CREB at the particular target location, thereby reducing the dissociation rate.

The role of the disorder in transcription factors was recently shown to guide the protein towards the target site on the DNA²¹⁴. By deleting the DBD and disordered N-terminus separately of zinc-finger protein Msn2, Brodsky *et al* showed that, rather intriguingly, the DBD was not required for localisation to the target. Instead, certain lengths of the disordered N-terminus showed preferential binding to TATA box containing sequences and to the lengths of promoter regions (protein truncations resulted in favouring short promoter regions for example) using

Chromatin-Endonuclease sequencing²¹⁴. Brodsky *et al* also focused on the yeast bZIP protein Yap-1, which showed an almost identical behaviour to the Msn2 protein for Yap-1 targets. However, the evidence does not fully show if the number of co-activator proteins are responsible for the change in promoter occupancy. This may reveal a role for the significantly disordered Q-rich domains in CREB and by increasing the range of the target site, also therefore impacting the dissociation rate from non-target DNA, and also increasing the number of targets that CREB could associate to in the genome.

7

Discussion and future work

7.1 Summary of main results

The role of a disordered DBD in associating to and searching for the target DNA site has been characterised using a variety of biophysical techniques. The results presented in this thesis using CREB bZIP as the model have added a significant amount of understanding into these processes, being able to partially model both the monomer and dimer pathways using kinetics and equilibrium based techniques, and structurally characterise all but one of the states (Figure 7.1)

NMR provided the necessary level of detail in order to characterise the apo-state of CREB bZIP constructs that were examined in this work. Crucially, NMR revealed that, in the apo-state, the basic region residues were disordered with only partial helical content, and highly dynamic even in the context of the dimeric bZIP domain. The monomeric basic region alone was found to be even more disordered and more dynamic than this region in the bZIP constructs, but still containing small helical character. Focusing briefly on the CREB constructs starting at residue 285, the helical content was shown to be dependent on the propagation of the helical content from the leucine zipper in a preformed dimer; this propagation could be disrupted by introduction of glycine residues between the zipper and basic regions. Remarkably, the basic region residues in the decoupled construct were in an identical structural

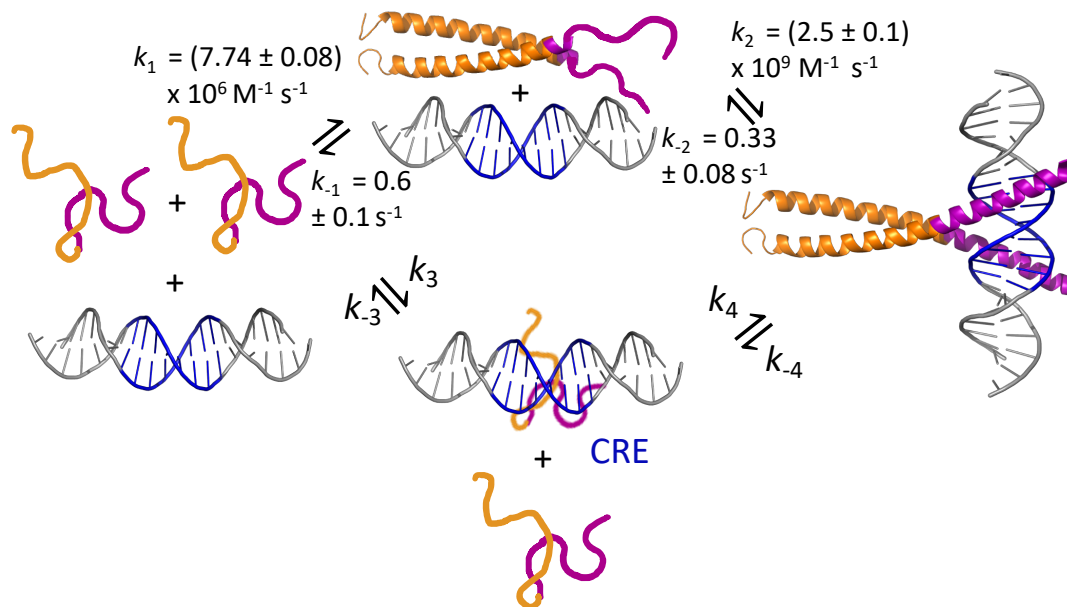


Figure 7.1: Model of CREB bZIP association to DNA (target or non-target sequences) via dimer (k_1 and k_2) or monomer (k_3 and k_4) pathways. Values for various parameters can be seen in Table 7.1.

and dynamical environment as the basic region only construct. Therefore, this defines a regulatory role for the formation of the homodimer, as the binding affinity of the decoupled construct to DNA was reduced compared to the WT.

The N-terminal residues preceding residue 285 also modulate the helical content within the basic region quite dramatically, with a construct designed between residues 277-339 being predicted to contain the helical content of the full length CREB protein. The 277-339 construct contained higher helical content in the basic region than 285-339, which shows the importance of the construct length of bZIPs when using a domain as the representative protein model for analysis¹⁸⁶. This behaviour of the N-terminal extension was confirmed in the basic region only constructs, with 277-315 showing markedly increased helical propensity over 285-315, highlighting that helix propagates from both the N- and C-terminal regions around the basic region controlling the helical propensity. A wide range of lengths of bZIP proteins have been used in previous work, potentially making any direct comparisons difficult (dimer details in Table 3.1).

Importantly, this work has been able to describe the structural propensity of CREB bZIP in complex with target and non-target DNA sequences. The 277-339 and 285-339 constructs both underwent coupled folding upon binding to CRE DNA, resembling the published crystal structure of CREB 285-339 with CRE DNA²¹ showing α -helix in the basic region. Unfortunately, structural descriptions of the monomeric CREB 285-315 and 277-315 with target DNA could not be determined fully, which is discussed in more detail below in the context of the equilibrium binding affinity. CREB 277-315 construct appears to be able to make limited specific interactions with the full CRE sequence as seen by NMR, indicating the potential for an ensemble of structures in solution composed of folded CREB bound to one CRE site, unfolded CREB bound to one CRE site and the same situations but with two CREB molecules associated to the full CRE. Focusing on the non-target DNA complex, the basic region residues in both (dimeric) bZIP and (monomeric) basic region only constructs undergo coupled folding upon binding intermediate to the apo and CRE DNA complexes. Also, the basic region residues remain dynamic when in complex with non-target DNA. Fascinating insights into the target search mechanism were obtained with the use of alternative non-target DNA sequences, highlighting that the helical content is dependent on the underlying DNA sequence, indicating that there is no single "non-specific" DNA sequence and that all possible permutations of DNA sequence are likely to produce a different helical content in the basic region.

The structural description works hand in hand with a description of the target search model, a summary of which can be seen in Figure 7.1 and Table 7.1. The kinetic diagram reveals a full experimental characterisation of the dimer pathway, with kinetic and experimental data observed. Only partial characterisation of the monomer pathway could be observed however, with no kinetic information. The reason behind the limited kinetic information for the monomer was due to the observed K_d of the monomer/DNA complex being high, which can infer that the association and dissociation rate constants also being high. Unfortunately

timescales for the kinetics experiments would not be suitable for analysing the monomer/DNA complex.

Table 7.1: Summary of the kinetic and equilibrium parameters obtained throughout this work for CREB 285-339 in complex with CRE-DNA. Parameters can be seen in Figure 7.1. nd is not determined. Errors are SEM.

Parameter	Value	Units
k_1	$(7.74 \pm 0.08) \times 10^6$	$\text{M}^{-1} \text{s}^{-1}$
k_{-1}	0.6 ± 0.1	s^{-1}
K_{d1}	9 ± 3	nM
k_2	$(2.5 \pm 0.1) \times 10^9$	$\text{M}^{-1} \text{s}^{-1}$
k_{-2}	0.33 ± 0.08	s^{-1}
K_{d2}	4.4 ± 0.6	nM
k_3	nd	
k_{-3}	nd	
K_{d3}	130 ± 20	μM
k_4	nd	
k_{-4}	nd	

That being said, the equilibrium affinity to DNA for the dimer and monomer models could be determined, taking into account the CREB homodimer K_d that was estimated to be 9 ± 3 nM. Remarkably, the affinity of bZIP dimer to the target DNA was 10000-fold tighter than the basic region monomer, with the apparent dimer affinity essentially independent of whether it was a forced dimer or not. This high affinity corresponded to a slow/intermediate exchange regime by NMR for the dimer in complex with target DNA. Surprisingly, the basic region monomer was also in slow/intermediate exchange with CRE DNA as observed by NMR despite the experimentally determined K_d being 10000-fold weaker. This may indicate that a two-state model fails for this complex and that potential intermediates exist, adding complication to the kinetic "square" in Figure 7.1 and reported elsewhere for other bZIP containing proteins^{80,134,171}.

When looking at the bZIP containing protein ATF2, adding in an 25-fold excess of calf thymus DNA resulted in the dissociation of the ATF-2 homodimer to monomer-DNA complexes¹⁹⁰; this study used FRET to validate the intermolecular distance on the protein. This would suggest that the homodimer stability is

dependent on the presence of DNA. This phenomenon may be related to the DNA dependent-rate enhancement in dimer formation of Fos-Jun¹²⁶ and the increased stability upon thermal unfolding of GCN4 with CRE-DNA compared to GCN4 alone¹⁸⁰ when DNA concentration is equal to or lower than the protein concentration. Therefore, increasing DNA concentration stabilises the homodimer, until a point at high DNA: protein ratios whereby the monomer-DNA complex is preferred. As *in vivo* conditions contain a significant excess of DNA (possibly greater than 1000-fold, notwithstanding that occluded by nucleosomes), it is possible that the monomer-DNA complex may be the most populated state. *In vivo* single molecule experiments with halo-tagged CREB and mutants that destabilised the homodimer interaction did show two residency times, a short residency time which was theorised by the authors to be CREB associated to non-target sites and a long residency time which was CREB associated specifically to CRE²⁰³. It could therefore also be possible that the short residency times are monomeric CREB associated to CRE-sites, with the short residency times appearing relatively equal for CREB WT and CREB L318/325V^{135,203}.

Whilst undergoing the search process, the basic region of both the bZIP and monomeric model proteins remains dynamic, however the bZIPs show an increase in helical content. In accordance with this, looking at two different non-target DNA sequences actually revealed a difference in the secondary structure within the basic region, indicating that CREB samples discrete non-equivalent DNA sites during the target search process. The intersegmental transfer rate has been able to be experimentally determined for CREB homodimer, and appears to be dependent on the residual helical content in the basic region. There was no evidence to suggest a significant proportion of the protein has become monomeric under these conditions (160-fold excess of DNA) since the target rate was equivalent in the two cases. For 1D diffusion along DNA, the rate of pWT and FD constructs were similar, meaning that we have access to the sliding rate of the dimeric species only. Therefore, it is still currently unclear as to whether CREB has a preference for target DNA association and search through the monomer or dimer pathway, the evidence presented here

would suggest that both pathways are utilised. This would provide an answer to the monomer/dimer pathway controversy, with bZIPs in general possibly utilising both pathways rather than being restricted to one.

All of the work in this thesis was performed *in vitro*. *In vivo* conditions are different and even variable depending on the result of certain signalling pathways¹⁰⁴. These could change the flux of populations through the monomer-dimer pathways, with phosphorylation of CREB leading to association with CBP and assembly of transcriptional machinery^{103,104} and mTORC associating to the homodimer only²¹². DNA methylation may play a significant role in CREB target search, with cell extracts loaded with methylated and unmethylated CRE-DNA resulting in no band-shift with methylated DNA¹¹⁵.

Overall, this work has furthered the understanding of the target search process by a disordered transcription factor. The apo state of the bZIP domain of CREB has been characterised in depth, describing the homodimer affinity which is a key parameter to be able to model the process. The results presented here provide detailed information on the dynamics of the bZIP domain whilst undergoing target search, with the specificity being controlled by helical content in the basic region and propagation from the leucine zipper. This provides context to various bioinformatic and theoretical work regarding how DNA binding proteins are able to recognise the target site within the large excess of non-target DNA.

7.2 Future work

Several areas can be identified where further work would aid in answering some of the outstanding questions posed by work in this thesis. Firstly, investigation of CRE half sites was alluded to above. Many of the genes that CREB regulation is involved in contain CRE clusters, containing full and half-site CRE motifs⁹⁶. To fully investigate the target search mode of CREB, this will need to be investigated experimentally using the techniques described throughout this thesis. An interesting experiment would be repeating the stopped flow target search kinetics but with half-site CRE's within the probe and/or competitor DNA sequences, to investigate

how the rates are affected by the half-sites being present. Also, NMR with the monomeric model CREB might be less complex with just one binding site per DNA rather than the two used throughout this work.

Subsequent to this, analysing the CREB/DNA complex from the perspective of the DNA would be of high importance. Initial experiments into the feasibility of this had been performed and quite promisingly the downfield shifted proton resonances that correspond to the thymine and guanine NH groups from the bases were able to be assigned in the apo state, and these DNA resonances shift when DNA is added. Therefore, further investigations using multiple DNA sequences could start to answer the question of what makes a DNA sequence specific? Similarly, investigations of the dynamics of the arginine side chains could provide detailed information into the complex between CREB and non-target DNA; this could be done using the pulse sequences developed by Fleming Hansen's group²¹⁵ to determine the rotational dynamics of the guanidino group. Complementary to this, use of methylated DNA throughout the experiments would be useful to investigate, and would provide further *in vivo* significance to the work.

An inherent drawback of this work was the difficulty in fluorescently labelling the CREB bZIP construct efficiently. Further work would need to be completed to fully optimise the conditions for labelling with photostable dyes, such as AlexaFluor's. This would enable equilibrium and kinetic experiments to be done from the perspective of the protein, which is currently not possible. This would reveal if the two-state association model is correct, as swapping the fluorophore between two protein partners can reveal increased complexity in protein/protein interactions via the presence of intermediates if the state is populated at the timescales observed for kinetics and/or equilibrium. The bioinformatics presented is at an early and basic stage and could go down two routes: 1) determine the significance of the regions found by comparing the data to ChIP seq information for chromatin occupancy. 2) Determine common sequence motifs and use these in the *in vitro* experiments already performed, similar to the half-site CRE DNA and methylated CRE DNA plan above. This will help to determine what the role, if any, in the flanking regions

of promoters have in the context of CREB target search. Equivalently, the NMR was at the limit of signal to noise to assign each of the complexes and was time consuming. To improve this, ^2H labelling of the protein could significantly improve the signal-to-noise ratio of the protein/DNA complexes and lead to improved efficiency of NMR data collection.

7.3 COVID and Mitigating Circumstances

Due to the COVID-19 pandemic, various sections of work could not be completed and these will be outlined here.

- k_{off} and k_{on} experiments with non-CRE DNA (linear) using stopped flow fluorescence anisotropy. This experiment would have been valuable to further describe the exchange regime identified from NMR, as well as describing the target search process from the non-target DNA perspective. Stopped flow experiments were stopped in February 2020.
- K_d determination of the p1, p2 and non-CRE DNA sequences at 35°C with CREB 277/285-339 using fluorescence equilibrium and/or ITC. The K_d may vary with temperature, therefore it is of importance to determine the fraction bound for analysis of the NMR experiments. CREB constructs with non-CRE DNA titrations were performed at 35°C until no further chemical shift perturbation was observed, therefore saturation of the complex was deemed to have occurred. The NMR triple resonance experiments with the CREB bZIP constructs in complex with DNA were assigned at 35°C, with the chemical shifts for secondary structure plotted from that dataset. The assignments of the apo state of protein was completed at 25°C. Due to the lack of dispersion in the ^{13}C -HSQC in particular at the $\text{C}\alpha/\text{H}\alpha$ and $\text{C}\beta/\text{H}\beta$ region of the spectrum, temperature dependence titrations could not be used for tracking of assignments. This had been planned for completion in March/April 2020 as assignment of spectra had priority.

- Similar to the point above, assignment of CREB bZIP bound to p1 and p2 non-target DNA sequences by NMR at the regular NMR concentrations used for the other complexes ($[\text{protein}] > 800 \mu\text{M}$). The p1 DNA with CREB 277-339 used in this thesis was performed at 300 μM protein, and although a control experiment with non-CRE DNA was performed at the same concentration and showed no effects, the signal to noise of the p1 complex was lower and therefore data quality was poor. Protein and DNA was being prepared the day of notice of Departmental shutdown.
- Intersegmental transfer rate experiments with CREB 277-339 (pWT) by stopped flow. As CREB 285-339 pWT and FD and 277-339 FD had been performed, a difference in the intersegmental transfer rate was observed between the length of constructs. This could be inferred to be due to the change in helical content so consistency between the two lengths of constructs would be preferable. Following this, a range of mutants with different helical contents could be examined to further investigate the role of the helical content in the basic region.

Appendices

Kinetics refolding equation for homodimer designed by Dr S. Shammass for use in Profit (QuantomSoft). K is K_d , k_1 is the k_{on} of homodimer formation, F_c and DF are the maximal and change in fluorescence intensity upon mixing respectively and P_{tot} and A_0 are fixed and are the Protein concentration in the reaction mixture and the starting concentration of monomer respectively.

```
function SarahReversibleHomodimerFit(k1, Ptot, A0, K, Fc, DF);

defaults
K:= 1, active;
k1:= 1.2, active;
Fc:=4, active;
DF:=2, active;
Ptot:= 5.469, inactive;
A0:= 5.469, inactive;

var
z, N, Aa:real;

begin
z:= sqrt(((K^2)/4)+(2*K*Ptot));
N:= (((K-2*z)/4)+A0)/(((K+2*z)/4)+A0);

Aa:= (((K-2*z)/4)-((K+2*z)/4)*N*exp(-k1*z*x))/(N*exp(-k1*z*x)-1));

Y:= Fc+DF*Aa;

end;
```

Similarly, for unfolding of the homodimer the following Profit script was generated.

```

function ReversibleHomodimerFit_Unfold(Ptot, K, Fc, DF, k2);

defaults
k2:= 40, active;
Fc:=4, active;
DF:=2, active;
Ptot:= 5.131, inactive;
K:= 0.02, active;
var
z, N, Aa:real;

begin
z:= K*sqrt(1+((4*Ptot)/K));
Aa:= 0.5*(1-exp(-k2*sqrt(1+4*(Ptot/K))*x)/((1/(K+z))*exp(-k2*sqrt
(1+4*(Ptot/K))*x)-(1/(K-z)))));
N:= (K-z)/(K+z);

Y:= Fc*Ptot + DF*Aa;
end;

```

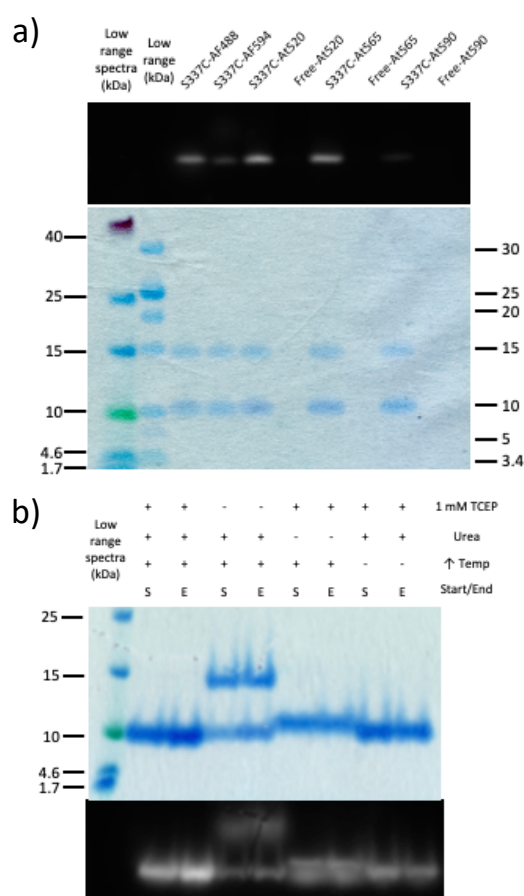


Figure S1: Multiple labelling conditions were trialled for CREB bZIP S337C with AF594 and other fluorophores, with limited success. a) Several fluorophore maleimide linked compounds were used for labelling and analysed by SDS-PAGE, using fluorescence intensity and coomassie stained scans of the same gel. b) Various optimisation of the labelling conditions were attempted, with the result being analysed similar to that in a).

Table S1: Conditions used for labelling the CREB bZIP S337C construct with fluorophores.

Date	Conditions	Atto 520	Atto 565	Atto 590	Alexa 488	Alexa 594	CPM	Label with Dimer
180923	RT 4 h + TCEP)	8.8%	17.7%	15.5%	47.8%	57.8%		No
180926	5M Urea. 1 mM TCEP. 50C. 4h	5.3%	14.7%	14.8%	49.8%	56.4%		No
181003	5M Urea. 1 mM TCEP. 50C. 4h				82.0%			No
181026	5M Urea. 1 mM TCEP. 50C. 4h							Yes
181030	0.9M salt. 5M Urea. 1 mM TCEP. 50C. 4h				50.0%			No
181031	0.9M salt. 5M Urea. 1 mM TCEP. 50C. 4h	poor	poor	poor				No
181106	0.9M salt. 5M Urea. 1 mM TCEP. 50C. 4h				<i>Carbamylation</i>			No
181109 - cMyb	100 mM sodium phosphate, pH 7.4, 0.05% tween-20	Yes	Yes	Yes	Yes	Yes		No
181112	10 mM sodium phosphate, pH 7.4. With DNA, Gdn-HCl, Tween-20 respectively.					Yes		No
181114	100 mM Tris, pH 7.4, 5 M Gdn-HCl				4% (after IEC)			No
181121	100 mM Tris pH 7, 50 deg C					60%		No
	100 mM Tris pH 7, 5 M urea 50 deg C					>90%		No
	100 mM Tris pH 7, 5 M Gdn-Hcl 50 deg C					<10%		No
190312	100 mM Tris, pH 7.4, 250 mM NaCl, 5x dye:protein						>90%	No

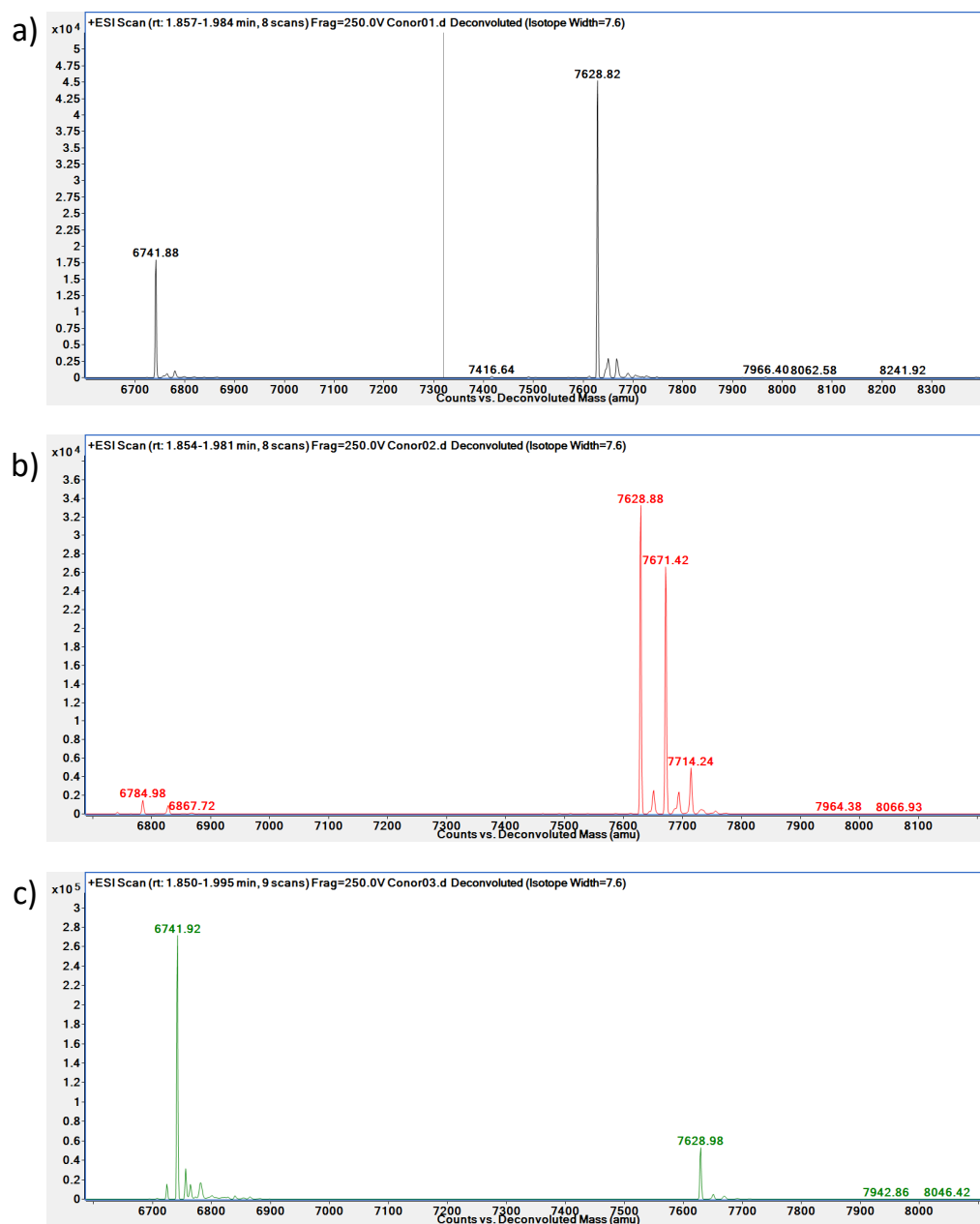


Figure S2: Labelling efficiency of CREB bZIP S337C was estimated with ESI-MS. a) -c) Details on buffer can be seen in Appendix Table S1 dated 181121. a) No denaturant. b) With urea. c) With guanidinium hydrochloride. Mass is Da.

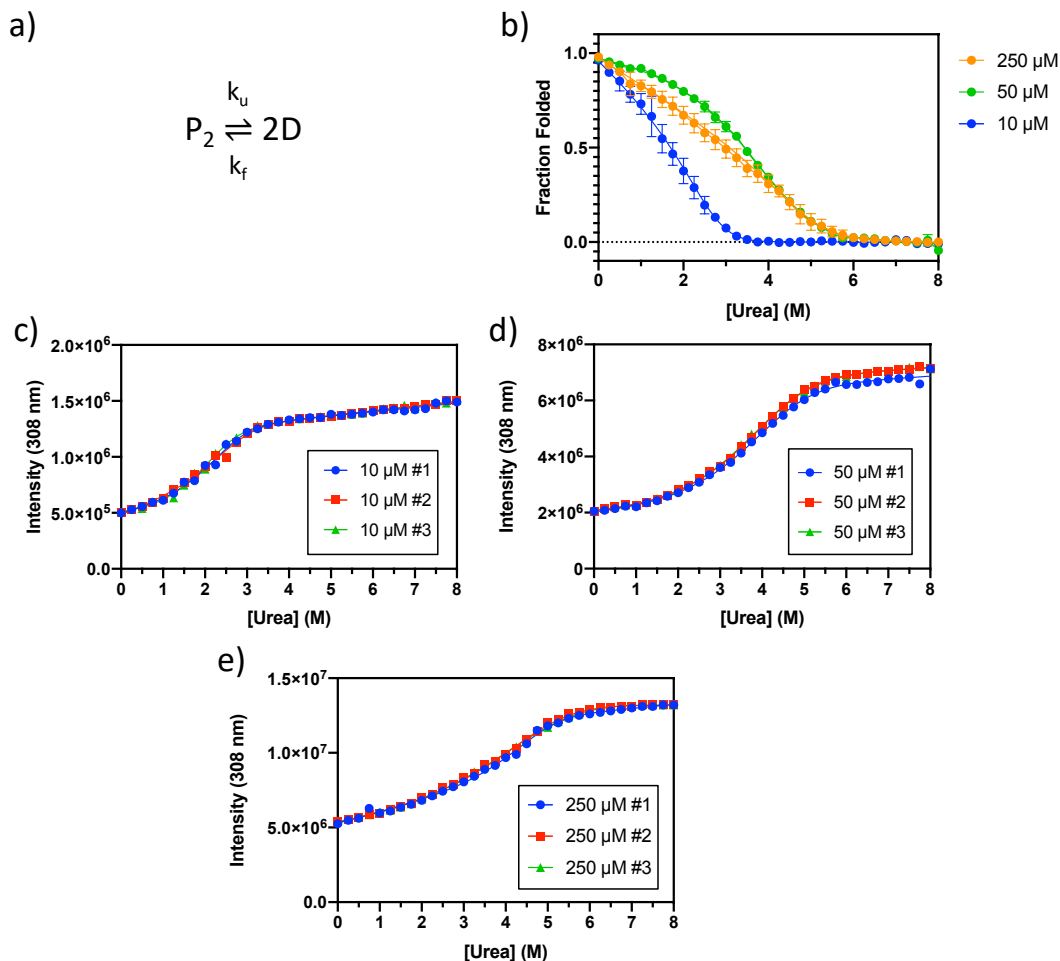


Figure S3: Chemical denaturation of CREB bZIP pWT unveiled a possible route for determining homodimer affinity. a) Two-state folding model for CREB bZIP was assumed. P = folded dimer, D = denatured monomer. b) Fraction folded of three concentrations of CREB bZIP at different urea concentrations, obtained from the data in c)-e). $n=3$. c)-e) Intrinsic tyrosine fluorescence was measured at several urea concentrations to see a fluorescence intensity change. b)-e) were fit with equations in section 2.10.

Table S2: Fitted parameters from the urea denaturation of CREB bZIP pWT at equilibrium. Errors are SD, from $n=3$.

[CREB bZIP pWT] (μM)	K_d (nM)	$[D]_{50\%}$ (M)	m -value ($\text{kcal mol}^{-1} \text{M}^{-1}$)
10	30 ± 20	2.1 ± 0.2	1.1 ± 0.1
50	90 ± 50	3.6 ± 0.1	0.9 ± 0.1
250	500 ± 700	4.1 ± 0.2	0.8 ± 0.2

Table S3: Fitted parameters from the urea denaturation of CREB bZIP pWT at equilibrium, with m -value fixed from Section 3.4. Errors are SD, from $n=3$.

[CREB bZIP pWT] (μM)	K_d (pM)	$[D]_{50\%}$ (M)	m -value ($\text{kcal mol}^{-1} \text{M}^{-1}$)
10	1800 ± 1600	2.4 ± 0.3	2.25
50	13 ± 3	4.0 ± 0.1	
250	9 ± 2	4.5 ± 0.1	

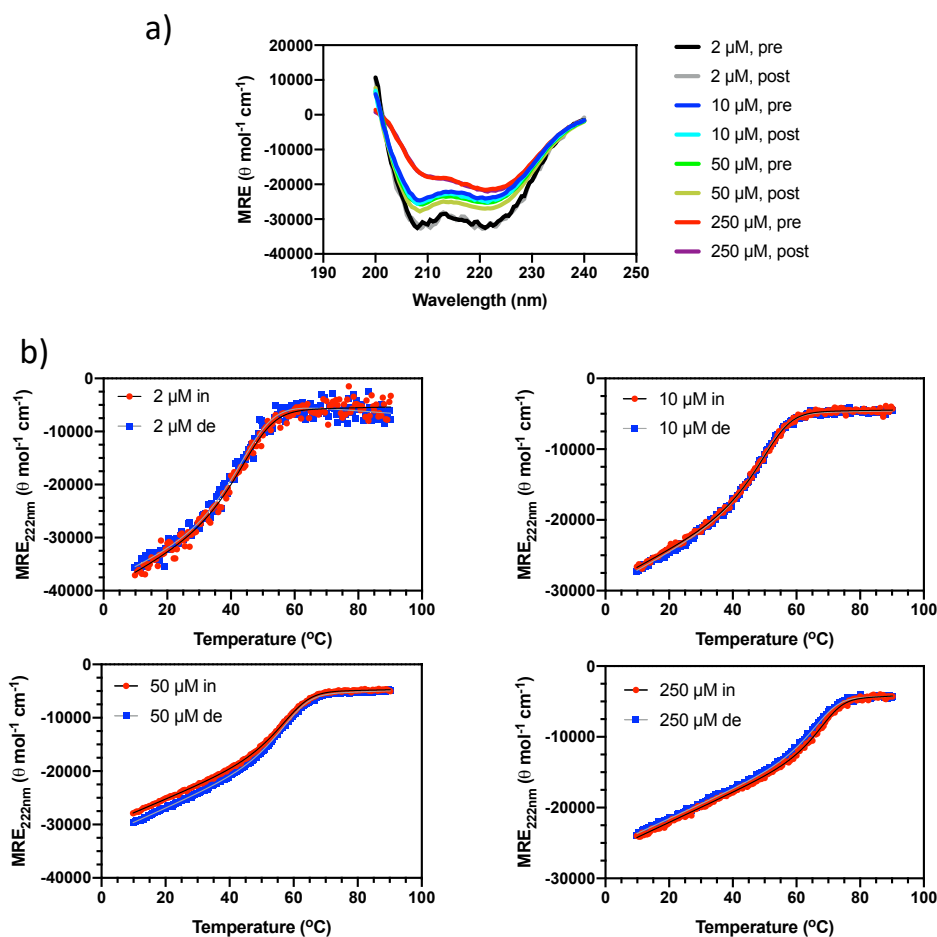


Figure S4: Thermal unfolding of CREB bZIP pWT unveils a tight homodimer interaction. a) Pre- and post melt CD spectra of CREB bZIP pWT at various concentrations at 25°C, in order to view the reversibility of the thermal melt. b) Thermal melts of CREB bZIP at various concentrations of protein (noted as monomeric protein concentration) across a 10-90°C range, acquiring the CD signal at 222 nm. In and De are increase and decrease curves respectively. Data was fit with the method outlined in section 2.11.

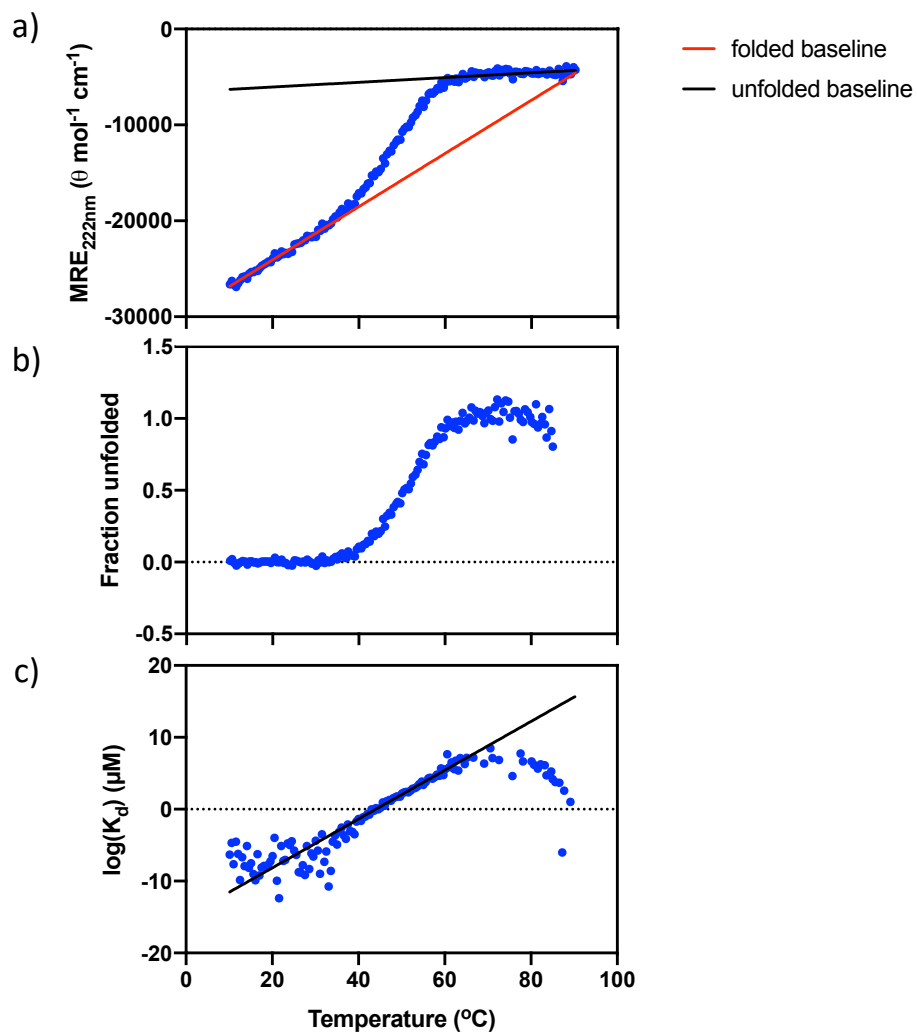


Figure S5: Baselines for the CREB bZIP thermal melts were manually extracted to determine K_d . a) A representative example of a 10 μM CREB bZIP thermal melt, with the folded and unfolded baseline fits shown that were manually extracted. b) Fraction unfolded plot was generated from the folded and unfolded baseline fits. c) The $\log(K_d)$ vs Temperature was determined by equation 2.7 and linear regression was fit to obtain K_d at a certain temperature.

Table S4: Thermal denaturation of CREB bZIP resulted in low estimates for homodimer affinity. K_d determined at 25°C. * indicates a independent repeat. Errors are standard errors.

Experiment Trial	[CREB bZIP pWT] (μ M)	Temp. Change	K_d (nM)	T_m ($^{\circ}$ C)	Cell type (mm)
A	10	Increase	1.5 ± 1.8	49.9 ± 0.3	1
		Decrease	3 ± 60	49.6 ± 0.3	
	50	Increase	2 ± 1	57.7 ± 0.1	1
		Decrease	0.7 ± 0.2	57.4 ± 0.1	
	250	Increase	0.013 ± 0.012	68.4 ± 0.3	0.2
		Decrease	0.2 ± 0.1	66.7 ± 0.2	
B	2	Increase	4 ± 20	44 ± 1	2
		Decrease	0.007 ± 770	43 ± 1	
	50	Increase	15 ± 6	59.3 ± 0.1	1
C	10	Increase	2.4 ± 1.3	50.2 ± 0.2	1
		Decrease	1.1 ± 0.8	49.6 ± 0.3	
D	10	Increase	3 ± 3	48.8 ± 0.6	1
		Decrease	1.5 ± 2.6	49.1 ± 0.6	
		Increase*	1.1 ± 4.2	49.3 ± 0.6	
		Decrease*	2.8 ± 4.3	47.0 ± 0.7	
	50	Increase	0.10 ± 0.03	56.9 ± 0.1	
		Decrease	0.09 ± 0.03	56.7 ± 0.1	
		Increase*	0.09 ± 0.04	56.8 ± 0.1	
		Decrease*	0.08 ± 0.08	56.5 ± 0.1	

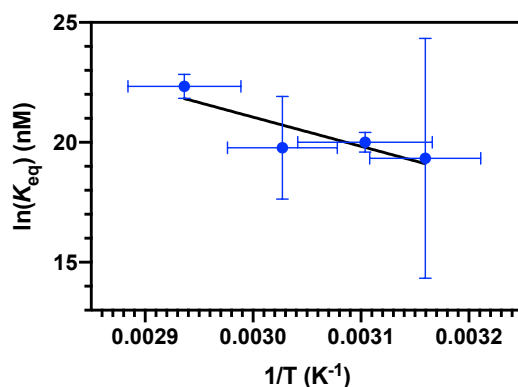


Figure S6: Van't Hoff plot using averaged values for each concentration extracted from CREB bZIP pWT thermal denaturation curves, using data from Table S4. Errors are SEM.

Table S5: Assigned chemical shifts for CREB bZIP pWT (285-339) at pH 7.4.

Residue No.	Name	H ^N	N	H α	H β	C	C α	C β
309	Lys	7.79	119.29	4.21	1.9	179.13	58.74	31.74
310	Ser	8.17	115.64	4.29	4.03	177.49	61.69	62.39
311	Leu	7.9	124.83	3.99	-	178.58	57.79	43.3
312	Glu	8.48	118.99	3.76	-	180.48	59.81	29.53
313	Asn	8.39	118.17	4.51	2.88	176.65	56.14	38.39
314	Arg	8.12	121.75	4.08	2.01	178.4	58.76	29.65
315	Val	8.47	118.6	3.42	1.87	176.8	67.4	37.02
316	Ala	7.4	120.72	4.2	1.55	181.25	55.43	17.98
317	Val	8.21	119.3	3.76	2.21	179.59	66.53	31.9
318	Leu	8.52	121.45	4.07	-	179.86	57.98	-
319	Glu	9.19	120.69	4.02	1.98	179.84	60.22	29.51
320	Asn	-	118.55	4.53	2.94	177.91	56.78	38.7
321	Gln	8.66	121.25	4.09	2.06, 2.18	178.08	59.16	28.49
322	Asn	8.86	119.8	4.01	2.64	176.41	57.31	38.66
323	Lys	7.81	118.63	4	2.04, 2.15	178.93	59.99	32.44
324	Thr	7.86	115.52	4.01	4.34	176.99	66.56	68.9
325	Leu	8.31	123.95	4.15	1.89	178.67	58.44	43.3
326	Ile	8.38	119.79	3.67	-	179.21	65.73	38.01
327	Glu	7.93	119.49	4.05	2.22, 2.12	179.49	59.43	29.21
328	Glu	8.36	122.94	4.08	2.19, 2.14	179.2	59.8	29.57
329	Leu	8.64	121.29	3.96	1.75	178.28	58.52	42.48
330	Lys	8.22	118.9	3.88	1.92	178.69	55.54	32.51
331	Ala	7.91	120.21	4.17	1.51	180.86	54.99	17.94
332	Leu	8.13	120.35	4.08	1.83	178.62	57.75	42.4
333	Lys	8.4	118.44	3.93	1.92, 1.97	179.78	60.61	32.15
334	Asp	8.11	118.64	4.42	2.75, 2.64	178.67	56.95	40.53
335	Leu	7.76	121.89	3.97	1.47	178.84	57.67	42.21
336	Tyr	7.59	115.21	4.45	2.67, 2.65	175.64	58.1	-
337	Ser	7.72	115.34	4.26	-	-	59.07	-
338	His	-	-	4.6	3.19	174.4	56.62	30.03
339	Lys	7.76	129.35	-	-	-	-	-

Table S6: Assigned chemical shifts for CREB bZIP pWT (285-339) at pH 6.5.

Residue No.	Name	H	N	ND	NE	HA	HB	HG	HD	HE	C	CA	CB	CG	CD
285	Lys	-	-	-	-	4.3	1.81, 1.81	-	-	-	176.91	56.93	33.11	-	-
286	Arg	8.55	122.56	-	-	4.27	1.81, 1.81	-	-	-	176.36	56.77	30.48	-	-
287	Glu	8.47	122.77	-	-	4.27	2.01, 2.01	-	-	-	176.81	57.11	29.98	-	-
288	Val	8.21	122.33	-	-	4.03	2.08	-	-	-	176.44	63.18	32.45	-	-
289	Arg	8.34	124.22	-	-	4.3	1.82, 1.82	-	-	-	176.42	56.64	30.49	-	-
290	Leu	8.22	122.9	-	-	4.35	1.60, 1.67	-	0.86, 0.92	-	177.71	55.56	42.14	-	23.50, 25.03
291	Met	8.3	121.21	-	-	4.43	2.11, 2.10	-	-	-	176.77	56.05	32.54	-	-
292	Lys	8.32	122.02	-	-	4.25	1.83	-	-	-	177	57.11	32.7	-	-
293	Asn	8.42	119.46	112.51	-	4.65	2.87, 2.86	-	7.64, 6.96	-	176.09	53.89	38.46	176.67	-
294	Arg	8.35	121.9	-	-	4.25	1.89, 1.89	-	-	-	177.35	57.46	30.36	-	-
295	Glu	8.39	121.46	-	-	4.2	2.07, 2.07	-	-	-	177.34	57.72	29.75	-	-
296	Ala	8.25	124.07	-	-	4.24	1.45	-	-	-	179.14	53.62	18.58	-	-
297	Ala	8.19	122.44	-	-	4.25	1.46	-	-	-	179.09	53.76	18.49	-	-
298	Arg	8.11	119.78	-	-	4.13	1.92, 1.68	-	-	-	177.91	58.07	30.3	-	-
299	Glu	8.3	120.7	-	-	4.25	2.07	-	-	-	177.78	57.89	29.7	-	-
300	Ser	8.22	115.61	-	-	4.3	3.97	-	-	-	176.27	60.21	63.14	-	-
301	Arg	8.09	122.03	-	-	4.16	1.88, 1.88	-	-	-	177.62	58.2	30.13	-	-
302	Arg	7.96	119.9	-	-	4.02	1.87, 1.87	-	-	-	177.44	57.97	30.23	-	-
303	Lys	8.07	119.5	-	-	4.2	1.89, 1.89	-	-	-	178.12	57.62	32.54	-	-
304	Lys	8.03	121.4	-	-	4.26	2.00, 2.00	-	-	-	178	58	32.52	-	-
305	Lys	8.35	121.95	-	-	4.18	1.95, 1.95	-	-	-	178.59	58.82	32.38	-	-
306	Glu	8.58	119.49	-	-	4.25	1.83, 2.15	-	-	-	178.22	58.32	29.21	-	-
307	Tyr	8.16	122.85	-	-	4.45	-	1.04	-	-	177.95	60.93	38.55	-	-
308	Val	8.25	119.41	-	-	3.32	2.19	-	-	-	177.38	67.34	31.39	-	-
309	Lys	7.8	119.22	-	-	4.22	1.91	-	-	-	179.15	58.77	31.68	-	-
310	Ser	8.18	115.56	-	-	4.29	4.07	-	-	-	177.5	61.69	62.38	-	-
311	Leu	7.9	124.7	-	-	4.01	-	1.17	0.56, 0.69	-	178.57	57.84	43.17	26.65	26.91, 23.02
312	Glu	8.48	118.94	-	-	3.75	1.93, 1.93	-	-	-	180.47	59.82	29.31	-	-
313	Asn	8.38	118.07	112.07	-	4.52	2.9	-	7.68, 6.92	-	176.64	56.12	38.29	175.92	-
314	Arg	8.12	121.6	-	-	4.08	2.09, 2.00	-	-	-	178.42	58.78	29.63	-	-
315	Val	8.47	118.53	-	-	3.42	2.12	-	-	-	176.79	67.38	31.57	-	-
316	Ala	7.4	120.66	-	-	4.21	1.56	-	-	-	181.26	55.44	17.91	-	-
317	Val	8.21	119.23	-	-	3.76	2.2	-	-	-	179.59	66.51	32.05	-	-
318	Leu	8.53	121.39	-	-	4.07	-	-	-	-	179.85	58	43.84	-	-
319	Glu	9.18	120.64	-	-	4.02	2.22, 1.97	-	-	-	179.82	60.18	29.4	-	-
320	Asn	7.81	118.43	113.45	-	4.54	2.94	-	7.74, 7.02	-	177.92	56.74	38.66	176.1	-
321	Gln	8.66	121.17	-	110.05	4.09	2.08, 2.17	-	-	6.76, 7.41	178.06	59.18	28.28	33.85	179.54
322	Asn	8.86	119.73	-	-	4.33	3.33	-	-	-	176.43	57.3	38.64	-	-
323	Lys	7.81	118.56	-	-	4.01	2.04	-	-	-	178.93	60.01	32.28	-	-

Table S6 continued from previous page

324	Thr	7.86	115.43	-	-	4.01	4.33	1.29, 1.29	-	-	176.97	66.6	68.9	21.91	-
325	Leu	8.32	123.82	-	-	4.13	1.83	-	-	-	178.68	58.38	43.34	-	-
326	Ile	8.38	119.64	-	-	3.66	1.97	0.94, 1.83, 1.83	0.83	-	179.16	65.73	37.9	30.63, 17.14	13.36
327	Glu	7.93	119.39	-	-	4.06	2	-	-	-	179.51	59.43	28.96	-	-
328	Glu	8.36	122.68	-	-	4.08	2.16	-	-	-	179.23	59.83	29.48	-	-
329	Leu	8.64	121.19	-	-	3.97	1.75	-	-	-	178.29	58.52	42.39	-	-
330	Lys	8.15	118.87	-	-	3.88	1.93	-	-	-	178.75	59.81	32.39	-	-
331	Ala	7.91	120.24	-	-	4.2	1.5	-	-	-	180.82	54.88	17.86	-	-
332	Leu	8.13	120.24	-	-	4.1	1.64	-	-	-	178.64	57.68	42.28	-	-
333	Lys	8.41	118.5	-	-	3.96	1.94	-	-	-	179.83	60.44	32.06	-	-
334	Asp	8.09	118.82	-	-	4.42	2.74, 2.65	-	-	-	178.56	56.97	40.52	-	-
335	Leu	7.75	121.57	-	-	4	1.46, 1.48	1.18	0.69, 0.79	-	178.99	57.7	42.18	27.15	24.62, 23.83
336	Tyr	7.73	115.54	-	-	4.45	2.71, 2.28	-	-	-	176.06	58.3	37.81	-	-
337	Ser	7.75	114.94	-	-	4.31	3.96	-	-	-	174.2	59.16	63.27	-	-
338	His	8.17	120.22	-	-	4.66	3.12, 3.26	-	-	-	174.07	56.05	30.04	-	-
339	Lys	7.78	128.82	-	-	4.13	-	-	-	-	181.3	57.81	-	-	-

Table S7: Assignment report for CREB bZIP 277-339 pWT. Generated from ccpnmr analysis (v2.4).

Category	% Assigned
All Residues	98.4375
Type C	95.3125
Type CA	96.875
Type CB	80.95238
Type CD	22.44898
Type CE	7.14286
Type CG	41.37931
Type CZ	0
Type H	98.3871
Type HA	69.23077
Type HB	50.44248
Type HD	9.375
Type HE	11.11111
Type HG	2.40964
Type N	95.3125
Type ND	60
Type NE	18.18182
Type NH	0
Type NZ	0

Table S8: Chemical shift table of assignments from CREB bZIP 277-339 pWT at pH 6.5.

Residue No.	Name	H	N	ND	NE	NH	HA	HB	HG	HD	HE	C	CA	CB	CG	CD	CE
277	Gln	-	-	-	112.39	-	-	1.96,2.16	2.43,2.43	-	7.63,6.87	-	-	28.75	33.47	-	-
278	Pro	-	-	-	-	-	4.4	1.93,2.36	-	-	-	177.58	64.11	31.95	27.49	50.71	-
279	Ala	8.54	123.68	-	-	-	-	-	-	-	-	179.11	53.63	18.78	-	-	-
280	Glu	8.31	119.52	-	-	-	4.2	2.06,2.06	-	-	-	177.76	57.66	29.77	36.26	-	-
281	Glu	8.23	121.42	-	-	-	4.17	2.05,2.05	-	-	-	177.76	57.87	29.75	35.94	-	-
282	Ala	8.2	123.16	-	-	-	4.16	1.44	-	-	-	179.34	54.08	18.48	-	-	-
283	Ala	8.01	121.47	-	-	-	4.17	1.47	-	-	-	179.2	54.09	18.49	-	-	-
284	Arg	7.99	119.21	-	-	-	4.2	1.90,1.90	-	-	-	177.74	57.73	30.23	27.11	43.29	-
285	Lys	8.04	119.96	-	-	-	4.14	1.87,1.87	-	-	-	178.06	58.06	32.56	25.29	29.13	-
286	Arg	8.03	120.89	-	-	-	4.17	1.89,1.89	-	-	-	177.14	57.87	30.42	27.49	43.33	-
287	Glu	8.18	120.8	-	-	-	4.17	2.09,2.09	-	-	-	177.87	57.93	29.77	36.19	-	-
288	Val	8.11	120.8	-	-	-	3.91	2.1	-	-	-	177.24	64.53	32.1	21.32,21.32	-	-
289	Arg	8.06	122.49	-	-	-	4.17	1.89,1.89	-	-	-	177.35	57.76	30.24	27.25	43.39	-
290	Leu	8.18	121.34	-	-	-	4.24	1.76,1.60	-	-	-	178.56	56.44	41.91	26.61	23.43,24.95	-
291	Met	8.22	120.33	-	-	-	4.31	2.15,2.15	-	-	-	177.6	57.01	32.34	32.07	-	-
292	Lys	8.3	121.42	-	-	-	4.2	1.87,1.87	-	-	-	177.96	57.79	32.36	24.89	29.13	42.16
293	Asn	8.37	119.38	112.11	-	-	4.62	2.89,2.89	-	6.94,7.64	-	176.76	54.48	38.2	176.45	-	-
294	Arg	8.31	121.85	-	-	-	4.23	1.93,1.93	-	-	-	178.04	58.15	30.41	27.37	43.51	-
295	Glu	8.34	121.14	-	-	-	4.17	2.03,2.03	-	-	-	177.83	58.15	29.78	36.11	-	-
296	Ala	8.2	123.44	-	-	-	4.24	1.47	-	-	-	179.53	54.04	18.37	-	-	-
297	Ala	8.09	122.06	-	-	-	4.24	1.46	-	-	-	179.5	54.04	18.34	-	-	-
298	Arg	8.05	119.68	-	-	-	4.1	1.93,1.93	-	-	-	178.28	58.51	30.25	27.28	43.49	-
299	Glu	8.27	120.6	-	-	-	4.24	2.09,2.09	-	-	-	177.98	58.1	29.53	35.94	-	-
300	Ser	8.2	115.45	-	-	-	-	-	-	-	-	176.55	60.51	62.98	-	-	-
301	Arg	8.03	121.78	-	-	-	4.13	1.87,1.87	-	-	-	177.83	58.51	30.19	-	-	-
302	Arg	7.94	119.87	-	-	-	3.98	-	-	-	-	177.66	58.41	30.23	27.34	-	-
303	Lys	8.05	119.16	-	-	-	4.14	1.86,1.86	-	-	-	178.32	57.8	32.73	-	-	-
304	Lys	7.99	121.13	-	-	-	-	-	-	-	-	178.08	57.96	30.38	-	-	-
305	Lys	8.32	121.86	-	-	-	4.17	1.96,1.96	-	-	-	178.73	58.94	32.33	24.91	-	-
306	Glu	8.56	119.34	-	-	-	4.21	-	-	-	-	178.28	58.42	-	-	-	-
307	Tyr	8.13	122.79	-	-	-	4.23	-	-	-	-	178.01	60.95	38.59	-	-	-
308	Val	8.22	119.43	-	-	-	-	-	-	-	-	177.42	67.35	-	-	-	-
309	Lys	7.82	119.2	-	-	-	-	-	-	-	-	179.18	58.78	31.75	-	-	-
310	Ser	8.19	115.59	-	-	-	4.38	-	-	-	-	177.53	61.69	-	-	-	-
311	Leu	7.89	124.72	-	-	-	4.09	-	-	-	-	178.61	57.79	-	-	-	-
312	Glu	8.48	118.92	-	-	-	3.85	-	-	-	-	180.51	59.77	-	-	-	-
313	Asn	8.38	118.05	112.21	-	-	4.3	-	-	6.92,7.69	-	176.66	56.1	38.27	-	-	-
314	Arg	8.12	121.69	-	-	-	-	-	-	-	-	178.46	58.72	29.71	-	-	-
315	Val	8.47	118.52	-	-	-	-	-	-	-	-	-	67.38	-	-	-	-

Table S8 continued from previous page

316	Ala	7.39	120.61	-	-	-	4.3	1.65	-	-	181.28	55.39	17.9	-	-	-
317	Val	8.21	119.27	-	-	-	-	-	-	-	179.61	66.47	-	-	-	-
318	Leu	8.53	121.37	-	-	-	-	-	-	-	179.89	57.99	-	-	-	-
319	Glu	9.17	120.61	-	-	-	4.1	-	-	-	179.86	60.18	29.51	-	-	-
320	Asn	7.81	118.46	113.6	-	-	-	-	7.74,7.01	-	177.96	56.73	38.64	-	-	-
321	Gln	8.66	121.19	-	110.23	-	4.17	-	-	-	178.06	59.09	28.35	33.99	179.58	-
322	Asn	8.85	119.74	-	-	-	-	-	-	-	176.45	57.29	38.63	-	-	-
323	Lys	7.81	118.53	-	-	-	4.09	-	-	-	178.95	59.98	32.31	-	-	-
324	Thr	7.86	115.51	-	-	-	-	-	-	-	176.95	66.57	-	-	-	-
325	Leu	8.31	123.81	-	-	-	-	-	-	-	178.72	58.35	-	-	-	-
326	Ile	8.38	119.69	-	-	-	-	-	-	-	179.2	65.7	-	-	-	-
327	Glu	7.92	119.4	-	-	-	4.39	1.93,1.93	-	-	179.53	59.4	29.05	36.38	-	-
328	Glu	8.36	122.71	-	-	-	3.87	-	-	-	179.26	59.73	29.31	36.46	-	-
329	Leu	8.64	121.24	-	-	-	-	-	-	-	178.35	58.5	42.38	-	-	-
330	Lys	8.14	118.88	-	-	-	3.98	2.01,2.01	-	-	178.83	59.76	32.51	-	-	-
331	Ala	7.9	120.26	-	-	-	4.26	1.6	-	-	180.8	54.86	17.94	-	-	-
332	Leu	8.11	120.25	-	-	-	4.2	-	-	-	178.69	57.61	42.32	-	-	-
333	Lys	8.41	118.56	-	-	-	4.06	-	-	-	179.89	60.31	32.22	-	-	-
334	Asp	8.08	118.94	-	-	-	-	-	-	-	178.49	56.92	40.54	-	-	-
335	Leu	7.74	121.49	-	-	-	4.1	-	-	-	179.13	57.63	42.13	-	-	-
336	Tyr	7.83	115.93	-	-	-	4.54	2.85,2.85	-	-	176.35	58.41	37.93	-	-	-
337	Ser	7.77	114.79	-	-	-	4.44	4.06,4.06	-	-	174.25	59.2	63.55	-	-	-
338	His	8.07	119.82	-	-	-	4.79	3.28,3.42	-	-	173.91	55.72	29.45	-	-	-
339	Lys	7.9	128.15	-	-	-	-	-	-	-	181.41	57.85	-	-	-	-

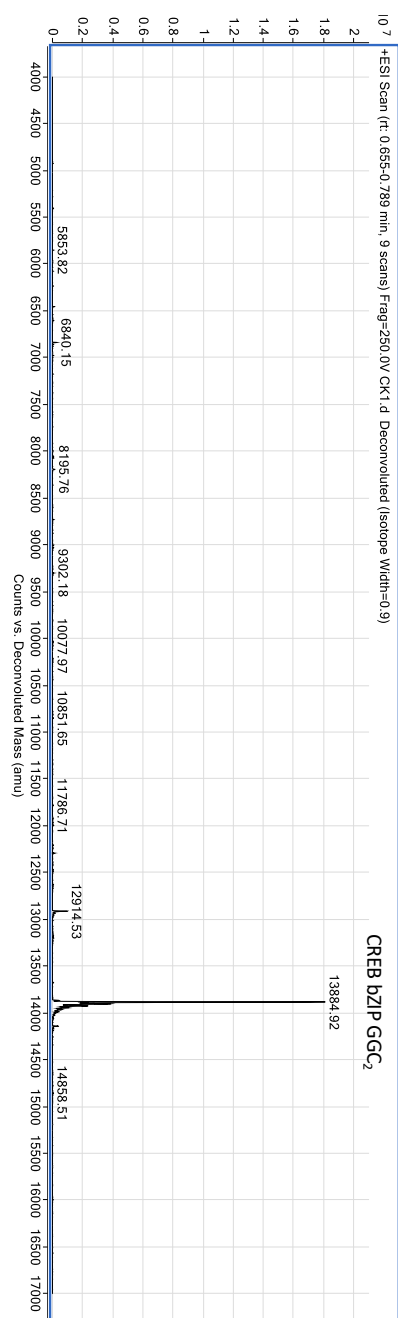


Figure S7: ESI-mass spec analysis of the CREB bZIP 285-339 FD construct following the disulphide bond formation step.

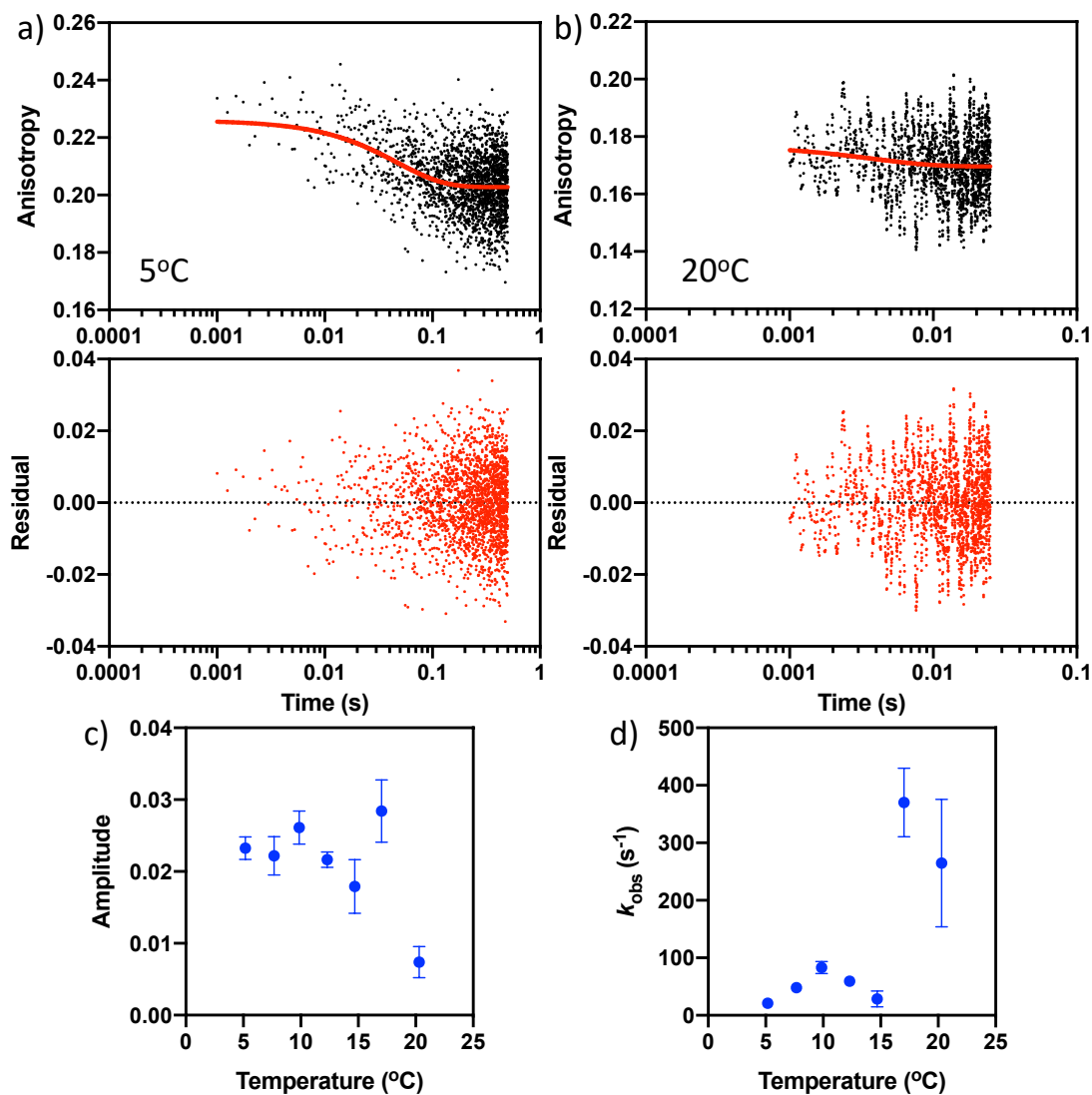


Figure S8: k_{off} determination of CREB from non-CRE DNA required temperature dependence measurements. a) Stopped flow anisotropy trace of 50 nM CREB 277-339 FD in complex with 50 nM AF488-non-CRE DNA (hpn) mixed with 300 nM CRE-DNA (hpn), collected at 5°C, with residuals from a single exponential decay function on bottom. b) Similar to a), but performed at 20°C. c) Amplitude from the single exponential decay function fits at the varying temperature. d) Extracted k_{obs} from the temperature dependence on the k_{off} data. Errors are SEM, minimum n=15.

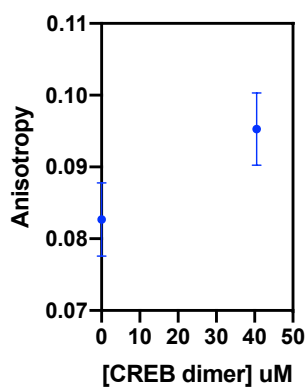


Figure S9: Anisotropy of free AlexaFluor488 changes upon addition of CREB 277-339 FD. Anisotropy of 5 nM AF488 was measured at two concentrations of CREB 277-339 FD. Errors are SEM.

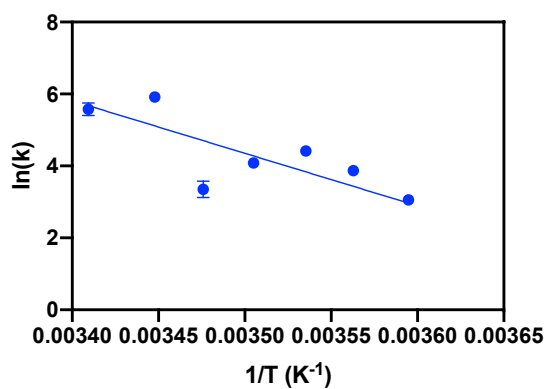


Figure S10: Estimate of k_{off} at 25°C from linear extrapolation of $\ln(k_{\text{obs}})$ against $1/T$. Errors are SEM from data in Appendix Figure S8.

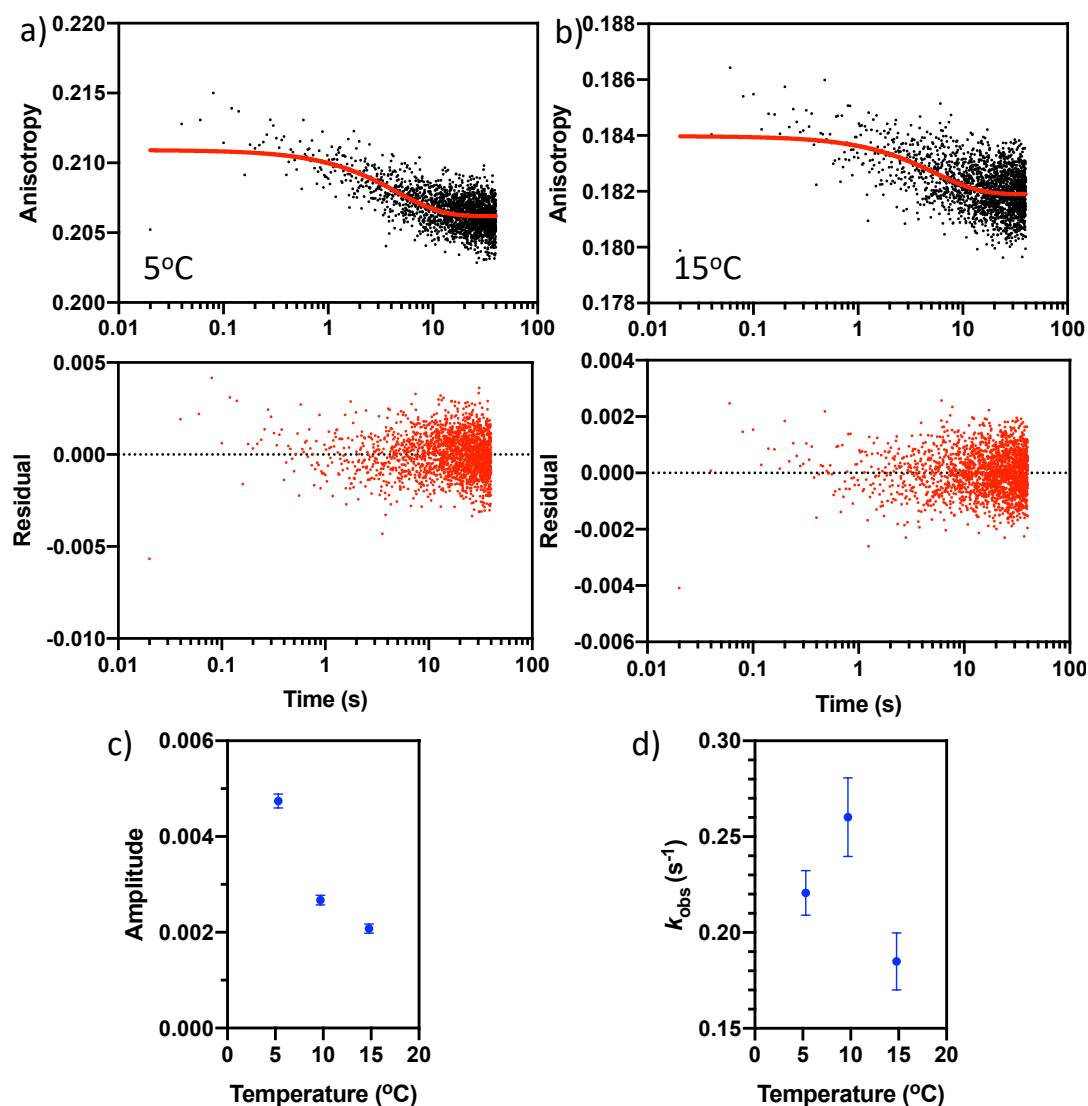


Figure S11: k_{on} of CREB 277-339 FD to non-CRE by stopped flow anisotropy required temperature dependence. a) 100 nM CREB 277-339 FD was rapidly mixed to 100 nM AF488-non-CRE DNA (hpn) and analysed by stopped flow anisotropy at 5°C. b) Similar to a), but collected at 15°C. Data in a) and b) were fit to a single exponential decay function and the residuals are shown underneath. c) Amplitude from the single exponential decay function fits at the varying temperature. d) Extracted k_{obs} from the temperature dependence on the k_{on} data. Errors are SEM, minimum n=13.

Table S9: Chemical shift table for CREB 285-339 in complex with CRE-DNA, with H^N and N^H collected at 25°C and the side chain chemical shifts collected at 35°C.

Residue	Name	H	N	ND	NE	HA	HB	HG	HD	HE	C	CA	CB	CG	CD
285	Lys	-	-	-	-	-	-	-	-	-	178.56	59.3	32.75	-	-
286	Arg	8.85	121.37	-	-	3.78	2.16	1.79	3.49	-	177.5	59.28	29.92	27.22	43.78
287	Glu	8.42	118.54	-	-	3.97	1.84,1.88	-	-	-	179.24	61.1	28.73	-	-
288	Val	7.9	120.62	-	-	3.85	2.13	1.06,1.14	-	-	177.66	66.37	31.89	22.49,21.17	-
289	Arg	7.93	121.1	-	-	-	2.01	-	-	-	179.14	59.46	29.78	-	-
290	Leu	8.47	119.42	-	-	4.34	1.92,2.00	-	1.09,1.01	-	174.24	58.32	42.02	-	23.32,25.53
291	Met	8.08	120.8	-	-	-	-	-	-	-	178.43	59.03	32.29	-	-
292	Lys	8.54	119.5	-	-	-	-	-	-	-	180.33	59.25	31.84	-	-
293	Asn	8.74	118.82	-	-	4.11	-	-	-	-	176.93	57.27	39.5	-	-
294	Arg	8.34	119.72	-	-	-	-	1.83	3.44	-	179.8	59.57	30.08	27.13	43.49
295	Glu	7.72	120.89	-	-	4.17	2.2	-	-	-	178.82	59.17	29.22	-	-
296	Ala	8.36	123.68	-	-	4.33	1.65	-	-	-	180.75	55.04	18.58	-	-
297	Ala	8.73	122.51	-	-	-	-	-	-	-	179.69	55.47	19.56	-	-
298	Arg	7.5	118.4	-	-	4.34	2.35	1.98	3.52	-	179.78	60.2	30.2	27.21	44.39
299	Glu	7.95	120.39	-	-	-	-	-	-	-	178.67	58.68	28.99	-	-
300	Ser	8.84	117.6	-	-	4.58	4.07	-	-	-	178.24	62.66	63.53	-	-
301	Arg	8.4	120.01	-	-	-	-	2.14	-	-	179.74	60.6	-	27.43	-
302	Arg	7.94	122.3	-	-	-	-	-	-	-	178.93	59.99	29.81	-	-
303	Lys	8.61	120.26	-	-	4.03	2.03	-	1.53	-	179.99	60.03	32.6	-	27.76
304	Lys	8.3	121.4	-	-	3.93	1.49	-	-	-	178.56	58.78	31.41	-	-
305	Lys	8.02	120.06	-	-	4.12	2.09	-	-	-	179.94	60.09	-	-	-
306	Glu	8.18	118.69	-	-	-	2.21	-	-	-	179.12	59.14	29.31	-	-
307	Tyr	8.54	123.25	-	-	4.29	-	-	-	-	178.11	60.9	38.26	-	-
308	Val	8.49	120.62	-	-	3.97	2.17	1.11,1.34	-	-	177.8	67.79	31.52	21.72,24.20	-
309	Lys	7.88	118.71	-	-	4.21	1.94	-	-	-	179.39	58.98	31.86	-	-
310	Ser	8.41	116.11	-	-	4.18	3.50,3.68	-	-	-	177.45	61.74	62.49	-	-
311	Leu	7.98	125.18	-	-	4.12	1.46,1.66	1.35	0.83,0.74	-	178.7	57.59,57.81	43.5	26.59	23.14,26.80
312	Glu	8.4	118.55	-	-	-	-	-	-	-	180.47	59.75	29.5	-	-
313	Asn	8.42	118	112.43	-	4.01	-	-	-	-	176.72	56.25	38.71	174.75	-
314	Arg	8.33	121.65	-	-	-	2.15	-	-	-	178.56	58.93	29.93	-	-
315	Val	8.46	118.43	-	-	3.57	-	0.93	-	-	176.8	67.39	31.6	21.62	-
316	Ala	7.32	120.66	-	-	4.23	1.61	-	-	-	181.29	55.42	18.03	-	-
317	Val	8.26	119.24	-	-	3.81	2.15	1.16,1.07	-	-	176.8	66.47	29.67	22.74,20.97,24.93	-
318	Leu	8.63	121.31	-	-	3.93	1.9	-	0.93,1.03	-	179.91	58.05	43.99	-	22.93,28.16
319	Glu	9.14	120.6	-	-	4.01	-	-	-	-	179.82	60.22	29.43	-	-
320	Asn	7.78	118.28	113.59	-	4.31	-	-	7.77,7.04	-	177.95	56.8	38.82	173.87	-
321	Gln	8.7	121.18	-	110.24	-	-	-	-	6.83,7.43	178.08	59.24	28.58	-	179.6
322	Asn	8.87	119.73	-	-	4	-	-	-	-	176.51	57.31	38.73	-	-

Table S9 continued from previous page

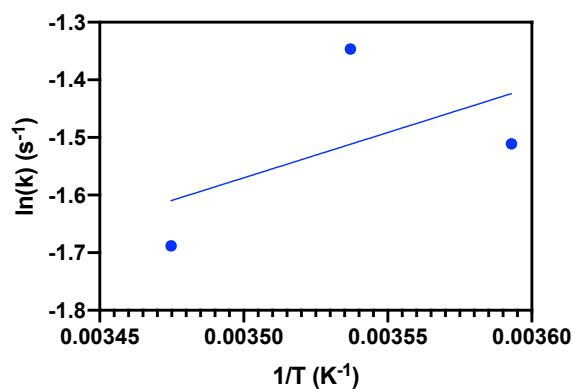
323	Lys	7.8	118.49	-	-	4.54	2.03,1.98	-	-	-	178.99	60.02	32.36	-	-
324	Thr	7.88	115.64	-	-	4.04	4.38	1.41	-	-	176.93	66.6	69	-	-
325	Leu	8.35	123.73	-	-	4.1	-	-	1.04,0.94	-	178.75	58.58	43.29	23.34,26.72	-
326	Ile	8.39	119.72	-	-	3.97	2.01	0.98	0.86	-	179.19	65.66	38.04	13.3	17.18
327	Glu	7.9	119.4	-	-	4.02	2.17	-	-	-	179.54	59.35	29.19	-	-
328	Glu	8.37	122.66	-	-	4.06	2.2	-	-	-	179.62	59.71	29.54	-	-
329	Leu	8.64	121.28	-	-	4.07	1.76	-	1.00,1.04	-	178.43	58.47	42.46	25.06,25.65	-
330	Lys	8.17	118.83	-	-	-	2.13	-	-	-	178.82	59.82	32.45	-	-
331	Ala	7.92	120.36	-	-	4.23	1.56	-	-	-	180.76	54.9	18.06	-	-
332	Leu	8.13	120.21	-	-	4.13	1.91,1.78	-	1.02,0.97	-	178.69	57.66	42.33	25.51,25.10	-
333	Lys	8.41	118.63	-	-	-	-	-	-	-	179.82	60.32	32.21	-	-
334	Asp	8.08	118.96	-	-	4.4	2.71,2.63	-	-	-	178.46	56.96	40.73	-	-
335	Leu	7.76	121.42	-	-	3.99	1.41,1.52	-	0.73,0.84	-	179.04	57.65	42.27	23.99,24.50	-
336	Tyr	7.88	116.07	-	-	4.43	2.78,2.43	-	-	-	-	58.45	38.09	-	-
337	Ser	7.8	114.75	-	-	4.19	3.93	-	-	-	174.25	59.22	63.63	-	-
338	His	8.05	119.76	-	-	4.69	3.35,3.21	-	-	-	173.78	55.69	29.44	-	-
339	Lys	7.92	128.07	-	-	-	-	-	-	-	181.66	57.89	-	-	-

Table S10: Chemical shift table for CREB 285-339 in complex with saturating concentration of CRE-DNA, with H^N and N^H collected at 35°C. For side chain carbon and hydrogen assignments, see Appendix Table S9.

Residue	Name	H	N
285	Lys	-	-
286	Arg	8.785	121.33
287	Glu	8.355	118.5
288	Val	7.835	120.58
289	Arg	7.865	121.06
290	Leu	8.405	119.38
291	Met	8.015	120.76
292	Lys	8.475	119.46
293	Asn	8.675	118.78
294	Arg	8.275	119.68
295	Glu	7.655	120.85
296	Ala	8.295	123.64
297	Ala	8.665	122.47
298	Arg	7.435	118.36
299	Glu	7.885	120.35
300	Ser	8.775	117.56
301	Arg	8.335	119.97
302	Arg	7.875	122.26
303	Lys	8.545	120.22
304	Lys	8.235	121.36
305	Lys	7.955	120.02
306	Glu	8.115	118.65
307	Tyr	8.475	123.21
308	Val	8.425	120.58
309	Lys	7.815	118.67
310	Ser	8.345	116.07
311	Leu	7.915	125.14
312	Glu	8.335	118.51
313	Asn	8.355	117.96
314	Arg	8.265	121.61
315	Val	8.395	118.39
316	Ala	7.255	120.62
317	Val	8.195	119.2
318	Leu	8.565	121.27
319	Glu	9.075	120.56
320	Asn	7.715	118.24
321	Gln	8.635	121.14
322	Asn	8.805	119.69
323	Lys	7.735	118.45
324	Thr	7.815	115.6
325	Leu	8.285	123.69
326	Ile	8.325	119.68
327	Glu	7.835	119.36
328	Glu	8.305	122.62
329	Leu	8.575	121.24
330	Lys	8.105	118.79
331	Ala	7.855	120.32
332	Leu	8.065	120.17
333	Lys	8.345	118.59
334	Asp	8.015	118.92
335	Leu	7.695	121.38
336	Tyr	7.815	116.03
337	Ser	7.735	114.71
338	His	7.985	119.72
339	Lys	7.855	128.03

Table S11: Assignment report of CREB 285-339 in complex with CRE-DNA.

Category	% Assigned
All Residues	98.21429
Type C	96.42857
Type CA	98.21429
Type CB	94.54545
Type CD	45.45455
Type CG	32.07547
Type H	98.18182
Type HA	68.42105
Type HB	62
Type HD	45
Type HE	6.06061
Type HG	26.0274
Type N	96.42857
Type ND	40
Type NE	11.11111

**Figure S12:** Estimate of k_{on} at 25°C from linear extrapolation of $\ln(k_{\text{obs}})$ against $1/T$. Errors are SEM from data in Appendix Figure S11

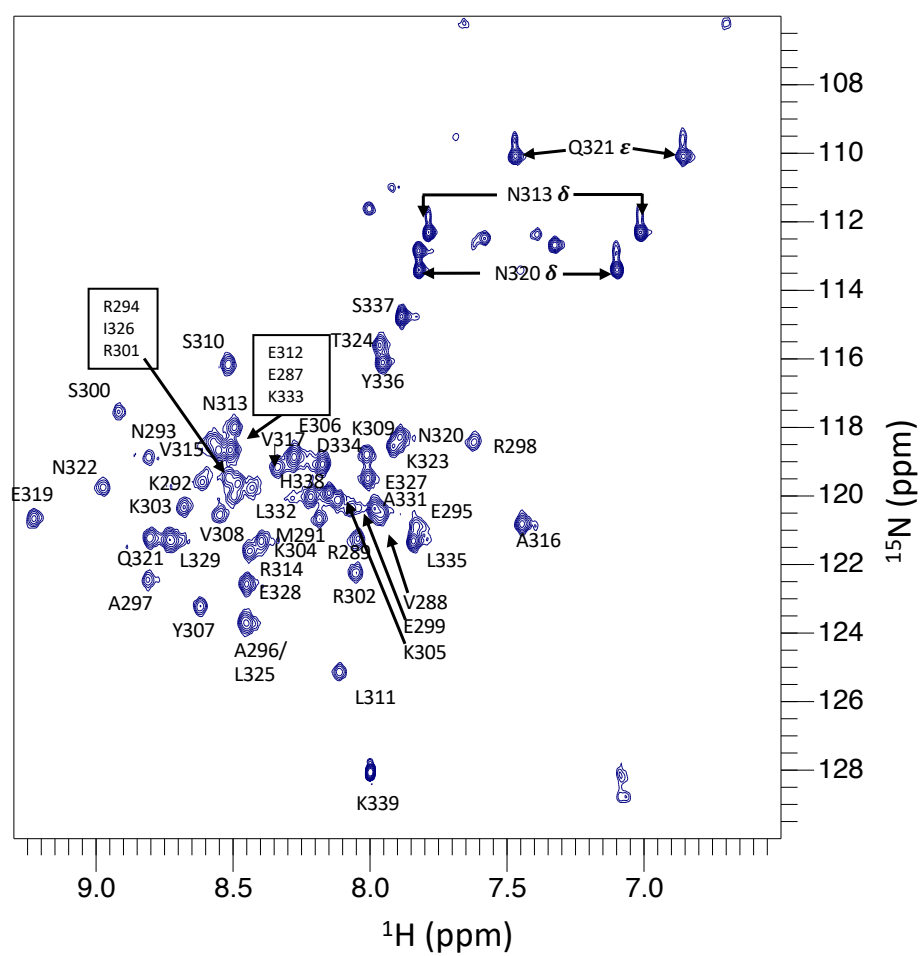


Figure S13: CREB 285-339 in complex with CRE-DNA was assigned at 35°C. Annotated ^{15}N -HSQC of CREB 285-339 in complex with CRE-DNA collected at 35°C.

Table S12: Chemical shift table for CREB 285-339 in complex with saturating concentration of non-CRE DNA collected at 25°C.

Residue	Name	H	N	ND	NE	NZ	HA	HB	HG	HD	HE	C	CA	CB	CG	CD
285	Lys	-	-	-	-	-	-	-	-	-	-	177.64	57.93	33	-	-
286	Arg	8.7	121.95	-	-	-	-	-	-	-	-	176.85	57.82	30.33	-	-
287	Glu	8.44	120.93	-	-	-	-	-	-	-	-	177.03	58.76	29.99	-	-
288	Val	8.05	121.28	-	-	-	3.92	2.09	0.96,1.01	-	-	177.1	63.48	32.16	21.59,19.37	-
289	Arg	8.17	122.8	-	-	-	-	-	-	-	-	177.3	57.75	30.22	-	-
290	Leu	8.25	120.81	-	-	-	-	-	-	-	-	177.35	56.8	41.98	-	-
291	Met	8.2	120.25	-	-	-	-	-	-	-	-	177.6	57.42	32.5	-	-
292	Lys	8.35	121.4	-	-	-	-	-	-	-	-	178.44	58.19	32.57	-	-
293	Asn	8.48	119.33	111.67	-	-	-	-	-	7.79,6.93	-	177.26	54.84	-	-	-
294	Arg	8.43	122.1	-	-	-	-	-	-	-	-	178.75	58.73	30.34	-	-
295	Glu	8.31	120.76	-	-	-	-	-	-	-	-	178.47	58.88	29.24	-	-
296	Ala	8.26	123.08	-	-	-	4.27	1.37	-	-	-	180.27	54.61	18.39	-	-
297	Ala	8.18	122.17	-	-	-	-	-	-	-	-	180.02	54.61	18.31	-	-
298	Arg	8.07	119.43	-	-	-	-	-	-	-	-	178.77	59.37	-	-	-
299	Glu	7.98	120.27	-	-	-	-	-	-	-	-	178.22	59.57	-	-	-
300	Ser	8.39	115.81	-	-	-	-	-	-	-	-	174.24	61.42	-	-	-
301	Arg	8.31	120.32	-	-	-	-	-	-	-	-	-	58.95	-	-	-
302	Arg	8.18	122.32	-	-	-	-	-	-	-	-	178.64	54.78	-	-	-
303	Lys	8.46	119.58	-	-	-	-	-	-	-	-	179.73	59.39	-	-	-
304	Lys	8.11	120.82	-	-	-	-	-	-	-	-	-	59.29	-	-	-
305	Lys	7.94	120.26	-	-	-	-	-	-	-	-	179.52	59.76	-	-	-
306	Glu	8.37	119.08	-	-	-	-	-	-	-	-	176.53	58.87	30.39	-	-
307	Tyr	8.47	122.15	-	-	-	-	-	-	-	-	176.84	57.59	-	-	-
308	Val	8.44	120.9	-	-	-	-	-	-	-	-	177.55	67.38	30.08	-	-
309	Lys	7.84	118.82	-	-	-	-	-	-	-	-	179.29	58.86	-	-	-
310	Ser	8.37	115.81	-	-	-	4.3	4.09	-	-	-	177.52	61.42	-	-	-
311	Leu	7.9	124.84	-	-	-	-	-	1.19	0.58,0.70	-	178.6	57.79	-	26.55	27.10,23.00
312	Glu	8.45	118.58	-	-	-	-	-	-	-	-	180.53	59.76	-	-	-
313	Asn	8.42	117.99	112.26	-	-	-	-	-	6.92,7.69	-	176.92	56.15	-	175.97	-
314	Arg	8.26	123.1	-	-	-	-	-	-	-	-	178.58	58.82	-	-	-
315	Val	8.48	118.48	-	-	-	3.43	2.11	1.00,0.86	-	-	176.8	67.38	31.95	24.95,21.63	-
316	Ala	7.38	120.59	-	-	-	4.19	1.55	-	-	-	181.31	55.37	17.66	-	-
317	Val	8.24	119.21	-	-	-	3.77	2.2	1.11,0.96	-	-	179.6	66.5	33.52	22.79,21.10	-
318	Leu	8.57	121.36	-	-	-	4.06	1.77	-	0.89,0.99	-	179.86	58.2	43.83	-	28.33,22.94
319	Glu	9.17	120.57	-	-	-	-	-	-	-	-	179.84	60.25	29.04	-	-
320	Asn	7.8	118.36	113.58	-	-	-	-	-	7.02,7.74	-	177.9	56.75	38.74	176.12	-
321	Gln	8.67	121.21	-	110.16	-	-	-	-	-	7.40,6.76	178.1	59.22	-	-	179.55
322	Asn	8.87	119.69	-	-	-	-	-	-	-	-	176.44	57.32	-	-	-
323	Lys	7.81	118.51	-	-	-	-	-	-	-	-	179	60.05	32.69	-	-

Table S12 continued from previous page

324	Thr	7.86	115.52	-	-	-	4.01	4.34	-	-	-	176.73	66.56	68.95	-	-	-
325	Leu	8.31	123.8	-	-	-	-	-	-	-	-	178.72	58.4	-	-	26.72,23.45	-
326	Ile	8.39	119.62	-	-	-	3.66	1.98	0.94,1.12,1.12	0.84,0.93	-	179.16	65.68	38.04	26.54,17.12	13.35	-
327	Glu	7.91	119.33	-	-	-	-	-	-	-	-	179.54	59.4	-	-	-	-
328	Glu	8.35	122.68	-	-	-	-	-	-	-	-	179.25	59.82	-	-	-	-
329	Leu	8.64	121.24	-	-	-	-	-	-	-	-	178.34	58.64	42.47	-	-	-
330	Lys	8.14	118.81	-	-	-	-	-	-	-	-	178.77	59.81	32.08	-	-	-
331	Ala	7.9	120.17	-	-	-	4.16	1.51	-	-	-	179.72	54.86	17.95	-	-	-
332	Leu	8.16	120.38	-	-	-	4.18	1.56	-	0.83,0.87	-	178.66	57.66	-	-	25.28,24.90	-
333	Lys	8.41	118.46	-	-	-	-	-	-	-	-	179.84	60.43	-	-	-	-
334	Asp	8.08	118.82	-	-	-	-	-	-	-	-	178.53	56.97	40.72	-	-	-
335	Leu	7.75	121.44	-	-	-	3.99	1.42,1.52	1.11	0.68,0.78	-	179.03	57.72	41.98	26.53	24.73,23.73	-
336	Tyr	7.79	115.78	-	-	-	-	-	-	-	-	176.14	58.33	37.91	-	-	-
337	Ser	7.76	114.86	-	-	-	-	-	-	-	-	174.23	59.23	63.35	-	-	-
338	His	8.12	120.03	-	-	-	-	-	-	-	-	-	56.01	29.95	-	-	-
339	Lys	7.86	128.32	-	-	-	-	-	-	-	-	-	57.82	-	-	-	-

Table S13: Assignment report of CREB 285-339 in complex with non-CRE DNA at 25°C.

Category	% Assigned
All Residues	98.21429
Type C	91.07143
Type CA	98.21429
Type CB	56.36364
Type CD	27.27273
Type CE	0
Type CG	22.64151
Type CZ	0
Type H	98.18182
Type HA	21.05263
Type HB	16
Type HD	28.33333
Type HE	6.06061
Type HG	15.06849
Type N	96.42857
Type ND	60
Type NE	11.11111
Type NH	0
Type NZ	0

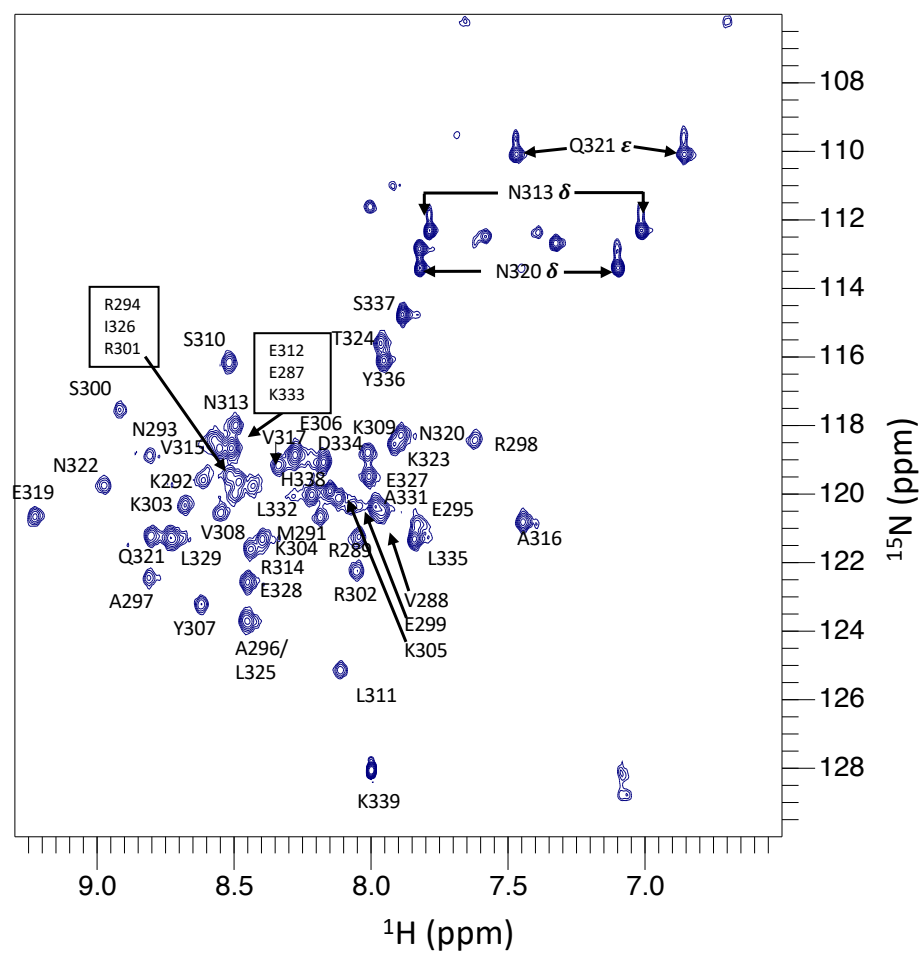


Figure S14: CREB 285-339 was fully assigned with non-CRE DNA at 35°C. ^{15}N -HSQC of CREB 285-339 in complex with non-CRE DNA.

Table S14: Chemical shift table for CREB 285-339 in complex with saturating concentration of non-CRE DNA collected at 35°C.

Residue	Name	H	N	ND	NE	HA	HB	HG	HD	HE	C	CA	CB	CG	CD
285	Lys	-	-	-	-	4.26	1.80,1.80	-	-	-	177.12	57.3	33.08	-	-
286	Arg	8.59	122.25	-	-	4.22	1.80,1.80	-	-	-	176.55	57.16	30.39	-	-
287	Glu	8.44	121.77	-	-	4.21	2.01,2.01	-	-	-	177.19	57.87	29.83	-	-
288	Val	8.06	121.38	-	-	3.96	2.06	-	-	-	176.69	63.71	32.38	-	-
289	Arg	8.25	123.52	-	-	4.25	1.86,1.86	-	-	-	176.76	57.12	30.38	-	-
290	Leu	8.18	121.82	-	-	4.28	1.59,1.69	-	-	-	178.15	56.13	42.05	-	-
291	Met	8.23	120.6	-	-	4.35	2.11,2.11	-	-	-	177.17	56.77	38.3	-	-
292	Lys	8.33	121.76	-	-	4.22	1.84,1.84	-	-	-	177.64	57.64	32.58	-	-
293	Asn	8.42	119.25	112.1	-	4.65	2.88,2.88	-	7.71,6.95	-	176.66	54.33	38.36	-	-
294	Arg	8.37	121.9	-	-	4.25	1.95,1.95	-	-	-	178.15	58.25	30.33	-	-
295	Glu	8.33	120.8	-	-	4.12	2.08,2.08	-	-	-	178.12	58.69	29.53	-	-
296	Ala	8.2	123.19	-	-	4.23	1.48	-	-	-	179.88	54.43	18.41	-	-
297	Ala	8.13	122.17	-	-	4.25	1.49	-	-	-	179.76	54.45	18.42	-	-
298	Arg	8.06	119.31	-	-	-	-	-	-	-	178.76	59.05	30.3	-	-
299	Glu	8.21	120.57	-	-	4.28	2.10,2.10	-	-	-	178.16	58.42	29.5	-	-
300	Ser	8.33	115.8	-	-	4.25	4.03,4.03	-	-	-	176.96	61.08	63.05	-	-
301	Arg	8.13	121.27	-	-	4.2	2.01,2.01	-	-	-	178.69	59.24	32.14	-	-
302	Arg	-	-	-	-	-	-	-	-	-	178.64	59.26	30.58	-	-
303	Lys	8.06	120.97	-	-	-	-	-	-	-	179.38	59.14	-	-	-
304	Lys	8.1	120.95	-	-	-	-	-	-	-	178.64	59.15	32.14	-	-
305	Lys	8.05	120.32	-	-	4.15	2.01,2.01	-	-	-	179.33	59.49	32.39	-	-
306	Glu	8.33	119.08	-	-	4.11	-	-	-	-	178.84	58.89	29.41	-	-
307	Tyr	8.26	122.26	-	-	3.93	-	-	-	-	178.01	60.96	38.3	-	-
308	Val	8.11	119.58	-	-	4.61	-	-	-	-	177.56	67.51	29.59	-	-
309	Lys	7.85	118.93	-	-	4.18	1.92,1.92	-	-	-	179.34	58.94	31.73	-	-
310	Ser	8.3	115.8	-	-	-	-	-	-	-	177.53	61.76	62.44	-	-
311	Leu	7.91	124.79	-	-	-	-	-	-	-	178.6	57.93	-	-	-
312	Glu	8.45	118.74	-	-	3.78	-	-	-	-	180.47	59.84	-	-	-
313	Asn	8.36	117.91	112.13	-	4.25	1.85,1.85	-	7.65,6.89	-	176.74	56.26	38.47	-	-
314	Arg	8.2	121.38	-	-	4.03	2.01,2.01	-	-	-	178.58	58.97	29.66	-	-
315	Val	8.45	118.37	-	-	-	-	-	-	-	176.8	67.41	-	-	-
316	Ala	7.37	120.76	-	-	4.18	1.55	-	-	-	181.28	55.45	17.89	-	-
317	Val	8.2	119.04	-	-	3.74	2.18	-	-	-	179.55	66.53	-	-	-
318	Leu	8.55	121.35	-	-	4.04	-	-	-	-	179.9	58.1	-	-	-
319	Glu	9.14	120.55	-	-	4.02	-	-	-	-	179.78	60.23	29.58	-	-
320	Asn	7.79	118.28	113.31	-	-	2.89,2.89	-	7.70,6.97	-	178.01	56.79	38.68	-	-
321	Gln	8.66	121.19	-	109.98	4.04	-	-	-	7.34,6.73	178.08	59.22	28.31	34.09	179.58
322	Asn	8.84	119.69	-	-	-	-	-	-	-	176.49	57.38	38.7	-	-
323	Lys	7.8	118.44	-	-	3.97	1.98,1.98	-	-	-	178.98	60.04	32.37	-	-

Table S14 continued from previous page

324	Thr	7.84	115.48	-	-	4	4	-	-	-	176.92	66.61	69.07	-	-
325	Leu	8.3	123.76	-	-	4.25	1.59,1.59	-	-	-	176.57	58.5	42.23	-	-
326	Ile	8.36	119.59	-	-	3.67	-	-	-	-	179.04	65.65	-	-	-
327	Glu	7.89	119.42	-	-	4.02	2.20,2.20	-	-	-	179.57	59.44	29.14	-	-
328	Glu	8.31	122.53	-	-	4.04	2.10,2.15	-	-	-	179.18	59.84	29.57	-	-
329	Leu	8.61	121.21	-	-	3.93	1.73,1.73	-	-	-	178.42	58.48	42.48	-	-
330	Lys	8.16	118.73	-	-	3.86	1.88,1.88	-	-	-	178.82	59.82	32.36	-	-
331	Ala	7.86	120.27	-	-	3.93	1.49	-	-	-	179.74	54.91	17.9	-	-
332	Leu	8.09	120	-	-	4.06	1.57,1.57	-	-	-	178.74	57.69	42.34	-	-
333	Lys	8.38	118.51	-	-	3.93	1.89,1.89	-	-	-	179.77	60.35	32.1	-	-
334	Asp	8.05	118.92	-	-	4.4	2.72,2.62	-	-	-	178.47	56.91	40.62	-	-
335	Leu	7.72	121.31	-	-	3.98	1.51,1.40	-	-	-	178.98	57.68	42.15	-	-
336	Tyr	7.77	115.77	-	-	4.43	2.74,2.35	-	-	-	176.27	58.37	37.98	-	-
337	Ser	7.75	114.73	-	-	4.32	3.93,3.93	-	-	-	174.25	59.21	63.53	-	-
338	His	8.08	120.05	-	-	4.65	3.13,3.27	-	-	-	173.97	55.96	29.69	-	-
339	Lys	7.82	128.18	-	-	-	-	-	-	-	181.35	57.89	-	-	-

Table S15: Assignment report of CREB 285-339 in complex with non-CRE DNA at 35°C.

Category	% Assigned
All Residues	98.21429
Type C	98.21429
Type CA	98.21429
Type CB	85.45455
Type CD	2.27273
Type CE	0
Type CG	1.88679
Type CZ	0
Type H	96.36364
Type HA	78.94737
Type HB	69
Type HD	10
Type HE	6.06061
Type HG	0
Type N	94.64286
Type ND	60
Type NE	11.11111
Type NH	0
Type NZ	0

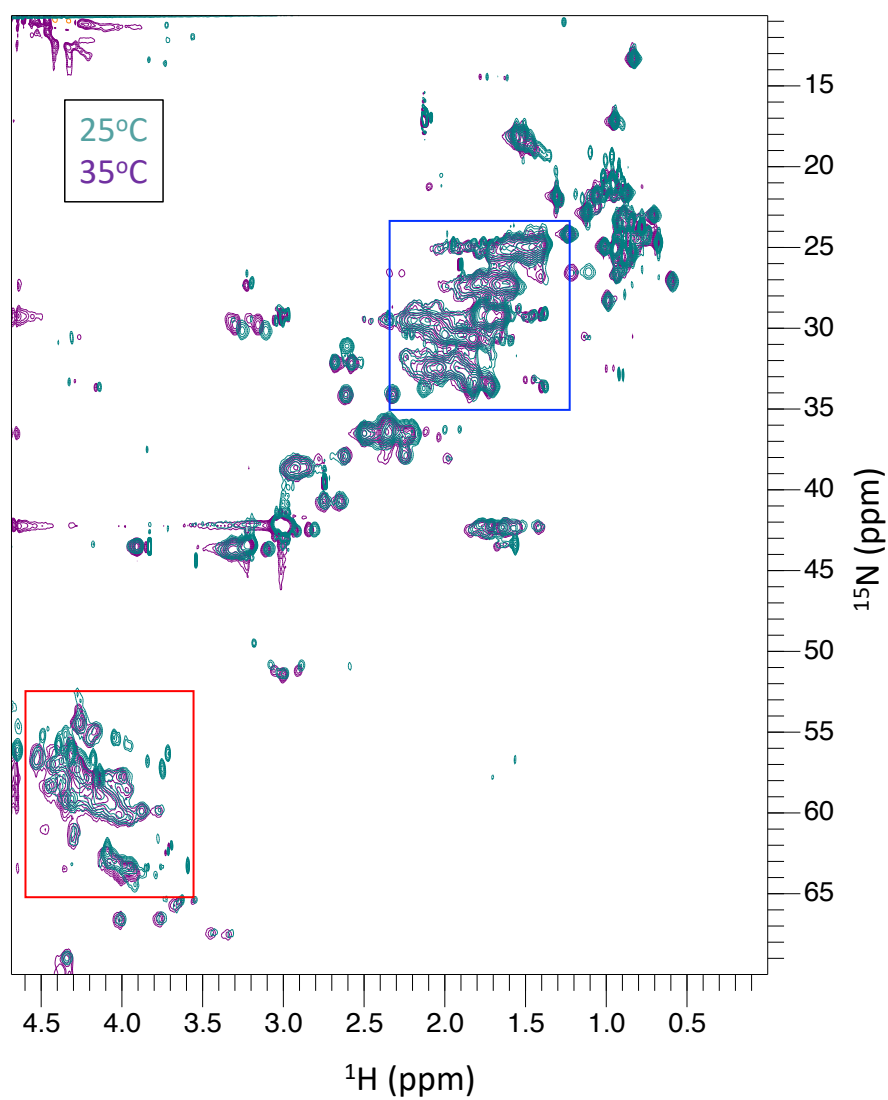


Figure S15: ^{13}C -HSQC of CREB 285-339 with non-CRE DNA performed at two temperatures (25°C and 35°C). Red and blue boxes indicate the $\text{H}\alpha/\text{C}\alpha$ and $\text{H}\beta/\text{C}\beta$ regions respectively. Spectra were referenced using DSS as a reference.

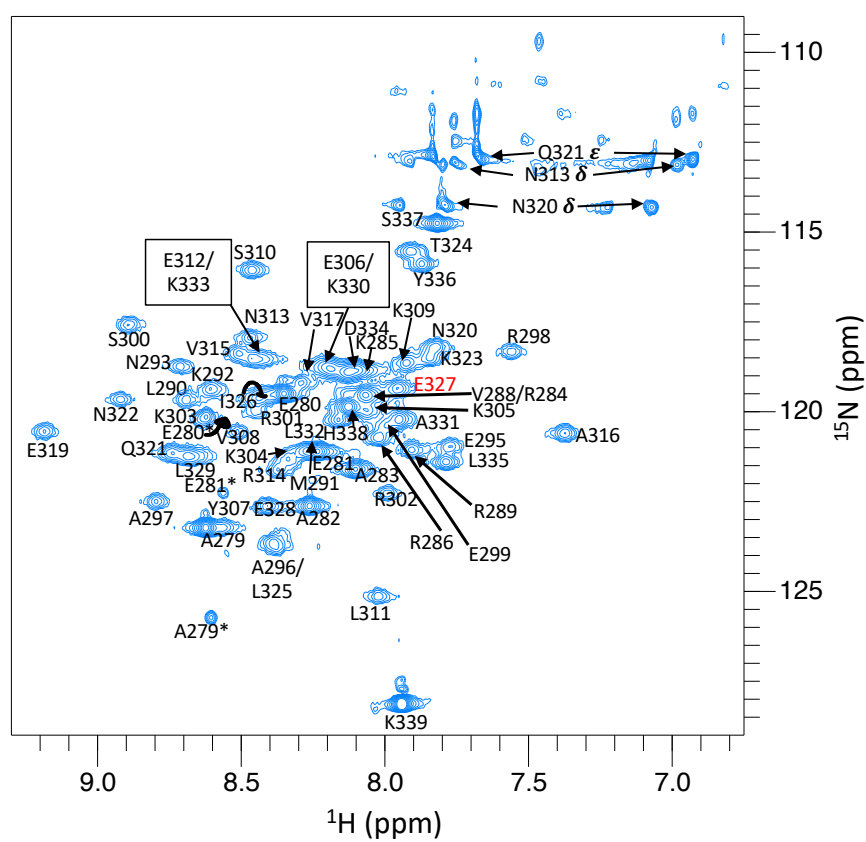


Figure S16: CREB 277-339 was assigned in complex with CRE-DNA at 25°C. ^{15}N -BTOSY of CREB 277-339 with CRE-DNA acquired at 25°C.

Table S16: Chemical shift table of CREB 277-339 in complex with CRE-DNA at 35°C. * indicates residues located close to the *cis* isomer of proline.

Residue	Name	H	N	HA	HB	C	CA	CB
278	Pro	-	-	4.41	1.96,2.38	177.48	64.1	32.01
278	Pro*	-	-	4.75	2.40,2.20	176.05	62.86	34.4
279	Ala	8.5	123.64	4.06	1.45	178.92	53.72	18.9
279	Ala*	8.5	125.44	4.32	1.46	178.09	53.3	19.18
280	Glu	8.32	119.47	4.02	2.08,2.08	177.67	57.81	29.9
280	Glu*	8.49	120.19	-	-	176.97	57.33	-
281	Glu	8.21	121.18	3.98	2.04,2.04	177.72	58.01	29.89
281	Glu*	8.46	122.23	-	-	-	-	-
282	Ala	8.18	123.19	3.97	1.46	179.42	54.31	18.6
283	Ala	8.02	121.48	4	1.49	179.35	54.32	18.57
284	Arg	7.99	119.27	4.03	1.94,1.94	177.76	58.01	30.15
285	Lys	8.03	119.78	3.91	1.88,1.88	178.35	58.52	32.65
286	Arg	8.03	120.93	3.98	1.94,1.94	177.12	58.45	30.16
287	Glu	8.15	120.14	3.76	-	178.23	58.66	29.83
288	Val	8.04	120.12	3.5	-	177.42	65.28	32.06
289	Arg	7.96	122.1	4.08	-	177.79	58.46	30.26
290	Leu	8.32	120.3	4.05	-	179.29	57.35	41.91
291	Met	8.21	119.97	4.05	-	177.79	58.02	32.5
292	Lys	8.3	120.99	3.82	-	178.88	58.59	32.19
293	Asn	8.44	119.59	4.05	1.97,2.06	177	55.2	38.18
294	Arg	8.47	121.92	3.91	-	178.92	59.14	30.42
295	Glu	8.24	120.84	3.77	-	178.47	59.05	29.66
296	Ala	8.25	123.1	4.08	2.07	180.29	54.77	18.4
297	Ala	8.19	122.07	3.81	-	179.93	54.7	18.63
298	Arg	8.04	119.53	-	-	179.12	59.51	30.4
299	Glu	8.19	120.67	3.93	-	178.31	58.61	29.62
300	Ser	8.39	115.81	-	-	177.23	61.44	62.97
301	Arg	8.16	120.82	-	-	177.6	59.75	32.66
302	Arg	8.07	121.19	-	-	-	-	-
303	Lys	-	-	-	-	179.63	59.37	32.61
304	Lys	8.12	120.9	-	-	-	59.37	-
305	Lys	-	-	3.86	-	179.48	59.57	32.46
306	Glu	8.36	119.19	-	-	-	59	-
307	Tyr	-	-	3.85	-	177.91	58.18	38.39
308	Val	8.09	119.65	-	-	177.56	67.54	-
309	Lys	7.87	118.94	3.88	-	179.29	58.98	31.9
310	Ser	8.35	115.88	-	-	177.48	61.8	62.5
311	Leu	7.93	124.76	-	-	178.55	57.89	37.72

Table S16 continued from previous page

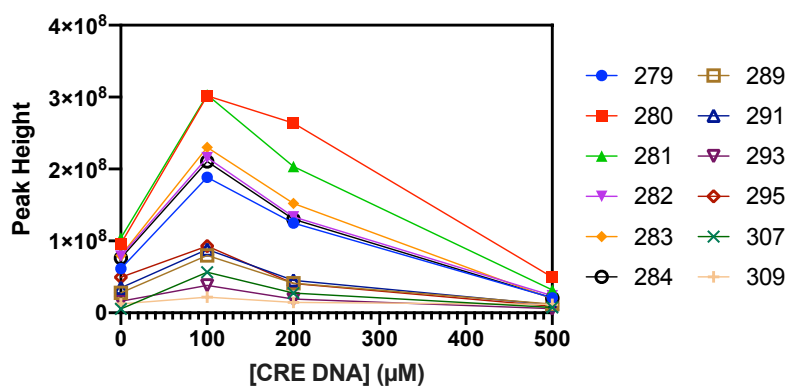
312	Glu	8.46	118.8	-	-	180.47	59.9	-
313	Asn	8.38	117.97	-	-	-	56.26	-
314	Arg	-	-	-	-	178.5	59.04	-
315	Val	8.47	118.43	-	-	176.78	67.42	-
316	Ala	7.38	120.77	-	-	177.64	55.47	18.02
317	Val	8.22	119.18	-	-	179.57	66.56	-
318	Leu	8.57	121.37	-	-	179.87	58.13	-
319	Glu	9.15	120.57	-	-	179.74	60.18	-
320	Asn	7.8	118.33	3.67	-	177.93	56.84	38.82
321	Glu	8.68	121.2	-	-	178.05	59.23	28.52
322	Asn	8.85	119.7	-	-	176.46	57.39	-
323	Lys	7.81	118.46	-	-	178.98	60.09	32.28
324	Thr	7.85	115.56	4.02	4.35	176.91	66.58	69.02
325	Leu	8.32	123.7	3.88	-	178.84	58.46	-
326	Ile	8.37	119.6	-	-	179.02	65.64	-
327	Glu	7.9	119.45	4.1	1.83,1.83	179.56	59.42	30.85
328	Glu	8.33	122.48	3.47	-	179.13	59.79	29.56
329	Leu	8.62	121.19	3.57	-	178.4	58.53	42.44
330	Lys	8.17	118.84	3.47	-	178.77	59.87	32.41
331	Ala	7.87	120.41	3.55	-	180.74	54.87	17.98
332	Leu	8.11	119.96	-	-	178.74	57.69	42.27
333	Lys	8.4	118.65	3.96	-	179.69	60.29	32.22
334	Asp	8.06	119.03	4.44	-	178.41	56.98	40.67
335	Leu	7.73	121.37	4.02	-	178.98	57.69	42.24
336	Tyr	7.82	115.92	4.27	2.81,2.81	176.3	58.45	38.09
337	Ser	7.77	114.86	4.16	3.96,3.96	174.16	59.23	63.47
338	His	8.06	119.98	4.5	3.12,2.98	173.77	55.85	29.62
339	Lys	7.87	128.05	-	-	181.25	57.93	-

Table S17: Chemical shifts of CREB 277-339 with CRE-DNA at 25°C.

Residue	Name	H	N
279	Ala	8.61	123.68
279	Ala*	8.59	125.96
280	Glu	8.36	119.8
280	Glu*	8.6	120.43
281	Glu	8.23	121.53
281	Glu*	8.54	122.64
282	Ala	8.25	123.16
283	Ala	8.09	121.88
284	Arg	8.07	119.79
286	Arg	8.02	120.93
287	Glu	8.1	119.74
289	Arg	7.89	121.85
291	Met	8.25	119.99
292	Lys	8.33	120.95
293	Asn	8.49	119.77
294	Arg	8.62	122.39
295	Glu	-	-
296	Ala	8.3	123.15
297	Ala	8.26	122.4
298	Arg	8	119.68
299	Glu	8.19	120.9
300	Ser	8.47	116.07
302	Arg	8.09	121.08
304	Lys	8.15	120.98
306	Glu	8.41	119.42
308	Val	8.07	119.39
309	Lys	7.89	119.13
310	Ser	8.43	116.16
311	Leu	7.95	125.19
312	Glu	8.48	119.04
313	Asn	8.47	118.3
315	Val	8.51	118.74
316	Ala	7.41	120.88
317	Val	8.27	119.48
318	Leu	8.62	121.62
319	Glu	9.2	120.85
320	Asn	7.82	118.58
321	Gln	8.71	121.46
322	Asn	8.9	119.98
323	Lys	7.85	118.85
324	Thr	7.9	115.83
325	Leu	8.35	124.04
326	Ile	8.42	119.9
327	Glu	7.95	119.67
328	Glu	8.39	122.95
329	Leu	8.67	121.55
330	Lys	8.18	119.12
331	Ala	7.94	120.53
333	Lys	8.44	118.85
334	Asp	8.11	119.17
335	Leu	7.77	121.71
336	Tyr	7.85	116.17
337	Ser	7.8	115.08
339	Lys	7.92	128.46

Table S18: Assignment report for CREB 277-339 in complex with CRE-DNA at 35°C.

Category	% Assigned
All Residues	94.28571
Type C	87.14286
Type CA	91.42857
Type CB	72.05882
Type H	90.90909
Type HA	55.55556
Type HB	24.59016
Type N	85.71429

**Figure S17:** CREB 277-339 is in slow/intermediate exchange with CRE-DNA. CRE-DNA titration was performed with CREB 277-339 at 35°C. Data was collected via ^{15}N -BTROSY at various DNA concentrations, with 500 μM saturating the complex. Peak heights were obtained from ccpnmr v2.4.

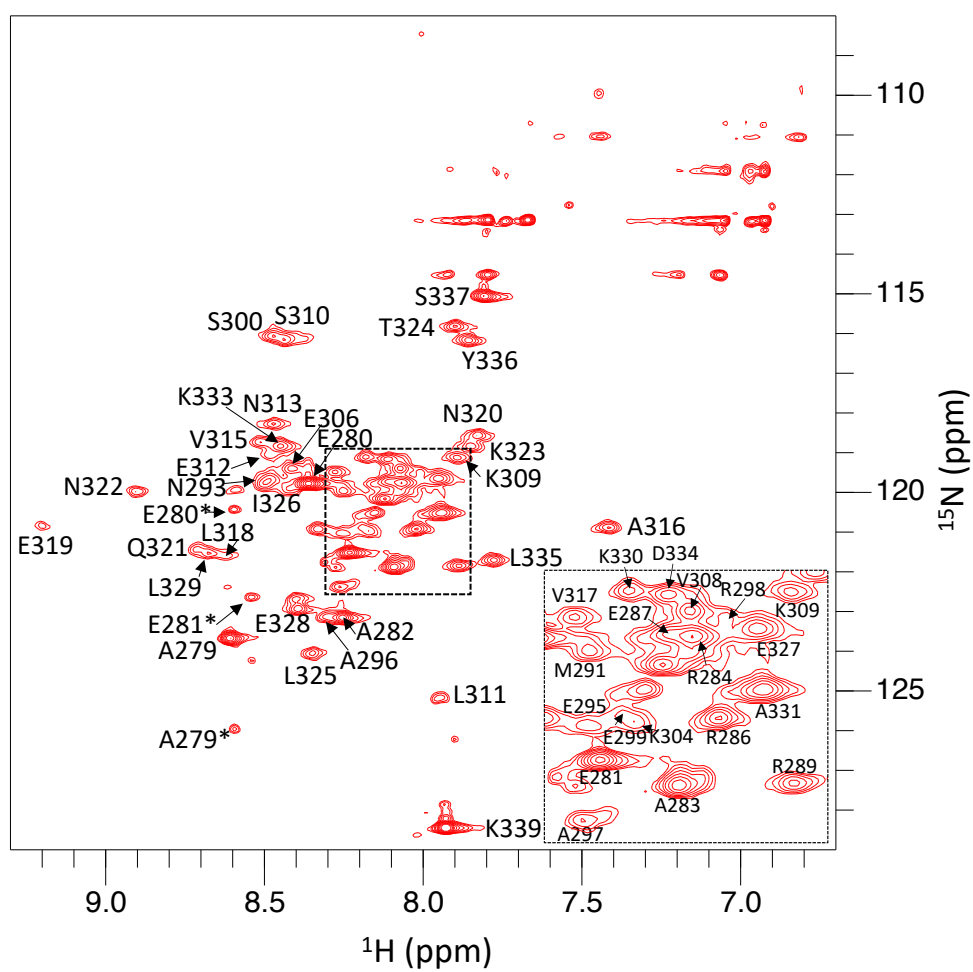


Figure S18: CREB 277-339 was assigned in complex with CRE-DNA at 25°C. ^{15}N -BTRSY of CREB 277-339 with CRE-DNA acquired at 25°C.

Table S19: Chemical shift table of CREB 277-339 in complex with non-CRE DNA at 35°C. * indicates residues located close to the *cis* isomer of proline.

Residue	Name	H	N	HA	HB	C	CA	CB
278	Pro	-	-	4.41	1.96,2.38	177.48	64.1	32.01
278	Pro*	-	-	4.75	2.40,2.20	176.05	62.86	34.4
279	Ala	8.5	123.64	4.06	1.45	178.92	53.72	18.9
279	Ala*	8.5	125.44	4.32	1.46	178.09	53.3	19.18
280	Glu	8.32	119.47	4.02	2.08,2.08	177.67	57.81	29.9
280	Glu*	8.49	120.19	-	-	176.97	57.33	-
281	Glu	8.21	121.18	3.98	2.04,2.04	177.72	58.01	29.89
281	Glu*	8.46	122.23	-	-	-	-	-
282	Ala	8.18	123.19	3.97	1.46	179.42	54.31	18.6
283	Ala	8.02	121.48	4	1.49	179.35	54.32	18.57
284	Arg	7.99	119.27	4.03	1.94,1.94	177.76	58.01	30.15
285	Lys	8.03	119.78	3.91	1.88,1.88	178.35	58.52	32.65
286	Arg	8.03	120.93	3.98	1.94,1.94	177.12	58.45	30.16
287	Glu	8.15	120.14	3.76	-	178.23	58.66	29.83
288	Val	8.04	120.12	3.5	-	177.42	65.28	32.06
289	Arg	7.96	122.1	4.08	-	177.79	58.46	30.26
290	Leu	8.32	120.3	4.05	-	179.29	57.35	41.91
291	Met	8.21	119.97	4.05	-	177.79	58.02	32.5
292	Lys	8.3	120.99	3.82	-	178.88	58.59	32.19
293	Asn	8.44	119.59	4.05	1.97,2.06	177	55.2	38.18
294	Arg	8.47	121.92	3.91	-	178.92	59.14	30.42
295	Glu	8.24	120.84	3.77	-	178.47	59.05	29.66
296	Ala	8.25	123.1	4.08	2.07	180.29	54.77	18.4
297	Ala	8.19	122.07	3.81	-	179.93	54.7	18.63
298	Arg	8.04	119.53	-	-	179.12	59.51	30.4
299	Glu	8.19	120.67	3.93	-	178.31	58.61	29.62
300	Ser	8.39	115.81	-	-	177.23	61.44	62.97
301	Arg	8.16	120.82	-	-	177.6	59.75	32.66
302	Arg	8.07	121.19	-	-	-	-	-
303	Lys	-	-	-	-	179.63	59.37	32.61
304	Lys	8.12	120.9	-	-	-	59.37	-
305	Lys	-	-	3.86	-	179.48	59.57	32.46
306	Glu	8.36	119.19	-	-	-	59	-
307	Tyr	-	-	3.85	-	177.91	58.18	38.39
308	Val	8.09	119.65	-	-	177.56	67.54	-
309	Lys	7.87	118.94	3.88	-	179.29	58.98	31.9
310	Ser	8.35	115.88	-	-	177.48	61.8	62.5
311	Leu	7.93	124.76	-	-	178.55	57.89	37.72

Table S19 continued from previous page

312	Glu	8.46	118.8	-	-	180.47	59.9	-
313	Asn	8.38	117.97	-	-	-	56.26	-
314	Arg	-	-	-	-	178.5	59.04	-
315	Val	8.47	118.43	-	-	176.78	67.42	-
316	Ala	7.38	120.77	-	-	177.64	55.47	18.02
317	Val	8.22	119.18	-	-	179.57	66.56	-
318	Leu	8.57	121.37	-	-	179.87	58.13	-
319	Glu	9.15	120.57	-	-	179.74	60.18	-
320	Asn	7.8	118.33	3.67	-	177.93	56.84	38.82
321	Glu	8.68	121.2	-	-	178.05	59.23	28.52
322	Asn	8.85	119.7	-	-	176.46	57.39	-
323	Lys	7.81	118.46	-	-	178.98	60.09	32.28
324	Thr	7.85	115.56	4.02	4.35	176.91	66.58	69.02
325	Leu	8.32	123.7	3.88	-	178.84	58.46	-
326	Ile	8.37	119.6	-	-	179.02	65.64	-
327	Glu	7.9	119.45	4.1	1.83,1.83	179.56	59.42	30.85
328	Glu	8.33	122.48	3.47	-	179.13	59.79	29.56
329	Leu	8.62	121.19	3.57	-	178.4	58.53	42.44
330	Lys	8.17	118.84	3.47	-	178.77	59.87	32.41
331	Ala	7.87	120.41	3.55	-	180.74	54.87	17.98
332	Leu	8.11	119.96	-	-	178.74	57.69	42.27
333	Lys	8.4	118.65	3.96	-	179.69	60.29	32.22
334	Asp	8.06	119.03	4.44	-	178.41	56.98	40.67
335	Leu	7.73	121.37	4.02	-	178.98	57.69	42.24
336	Tyr	7.82	115.92	4.27	2.81,2.81	176.3	58.45	38.09
337	Ser	7.77	114.86	4.16	3.96,3.96	174.16	59.23	63.47
338	His	8.06	119.98	4.5	3.12,2.98	173.77	55.85	29.62
339	Lys	7.87	128.05	-	-	181.25	57.93	-

Table S20: Assignment report for CREB 277-339 in complex with non-CRE DNA. Data acquired from the 35°C triple resonance experiments.

Category	% Assigned
All Residues	94.28571
Type C	87.14286
Type CA	91.42857
Type CB	72.05882
Type H	90.90909
Type HA	55.55556
Type HB	24.59016
Type N	85.71429

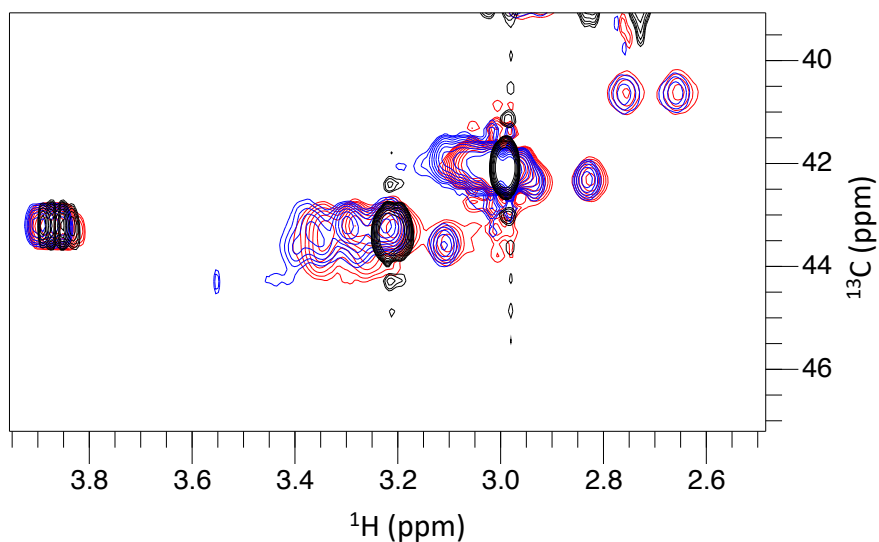


Figure S19: Arginine C δ chemical shifts show increased dispersion in CREB 277-339 with non-CRE DNA. ^{13}C -HSQC of CREB 277-339 in apo-state, in complex with CRE-DNA or non-CRE DNA which is shown as black, blue or red spectra respectively. Data were collected at 25°C at 950 MHz.

Table S21: Chemical shifts of CREB 277-339 with non-CRE DNA at 25°C.

Residue	Name	H	N
277	Gln	-	-
278	Pro	-	-
279	Ala	8.62	123.22
279	Ala*	8.6	125.71
280	Glu	8.35	119.48
281	Glu	8.24	121.08
281	Glu*	8.56	122.24
282	Ala	8.26	122.61
283	Ala	8.1	121.59
284	Arg	8.07	119.52
284	Arg	7.15	85.24
285	Lys	8.1	118.85
286	Arg	8.02	120.7
288	Val	8.07	119.52
289	Arg	7.91	121.03
290	Leu	8.69	119.65
291	Met	8.27	121.08
292	Lys	8.61	119.36
293	Asn	8.71	118.72
295	Glu	7.78	120.94
296	Ala	8.38	123.66
297	Ala	8.8	122.47
298	Arg	7.56	118.3
299	Glu	7.99	120.37
300	Ser	8.89	117.57
301	Arg	8.46	119.95
302	Arg	7.99	122.26
303	Lys	8.63	120.17
304	Lys	8.34	121.31
305	Lys	8.06	119.93
306	Glu	8.19	118.77
307	Tyr	8.56	123.19
308	Val	8.51	120.56
309	Lys	7.92	118.62
310	Ser	8.46	116.04
311	Leu	8.02	125.12
312	Glu	8.46	118.53
313	Asn	8.47	117.93
314	Arg	8.37	121.59
315	Val	8.51	118.34
316	Ala	7.37	120.58
317	Val	8.29	119.18
318	Leu	8.64	121.23
319	Glu	9.19	120.51
320	Asn	7.83	118.19
321	Gln	8.74	121.13
322	Asn	8.92	119.63
323	Lys	7.84	118.46
324	Thr	7.91	115.54
325	Leu	8.38	123.66
326	Ile	8.44	119.61
327	Glu	7.95	119.32
328	Glu	8.41	122.59
329	Leu	8.68	121.23
330	Lys	8.19	118.77
331	Ala	7.95	120.2
332	Leu	8.16	120.19
333	Lys	8.46	118.53
334	Asp	8.12	118.88
335	Leu	7.79	121.35
336	Tyr	7.87	115.86
337	Ser	7.82	114.7
338	His	8.11	119.8
339	Lys	7.95	128.08

Table S22: Chemical shift table for CREB 285-315 apo at 25°C.

Residue No.	Name	H	N	HA	HB	HG	HD	HE	C	CA	CB	CG	CD	CE
285	Lys	-	-	4.3	1.80,1.80	1.44,1.44	1.69,1.69	3.00,3.00	176.77	56.87	33.21	24.64	29.07	-
286	Arg	8.54	122.94	4.29	1.80,1.80	1.67,1.59	3.21,3.21	-	176.26	56.63	30.63	27.04	43.25	-
287	Glu	8.48	123.05	4.29	1.98,1.98	2.26,2.26	-	-	176.59	56.8	30.18	36.34	-	-
288	Val	8.23	122.46	4.04	2.07	0.95,0.95	-	-	176.3	62.94	32.56	20.87,20.87	-	-
289	Arg	8.38	124.71	4.32	1.81,1.81	1.62,1.62	3.20,3.20	-	176.22	56.38	30.67	27.05	43.4	-
290	Leu	8.25	123.38	4.33	1.66,1.59	1.57	0.92,0.86	-	177.48	55.32	42.26	-	23.40,24.75	-
291	Met	8.32	121.54	4.44	2.06,2.06	2.63,2.55	-	2.07	176.53	55.74	32.9	31.97	-	-
292	Lys	8.31	122.33	4.26	1.81,1.81	1.42,1.45	1.68,1.68	2.99,2.99	176.64	56.82	32.85	24.66	28.88	-
293	Asn	8.42	116.63	4.65	2.84,2.84	-	6.96,7.64	-	175.64	53.63	38.61	177.03	-	-
294	Arg	8.34	121.66	4.26	1.87,1.88	1.63,1.67	3.22,3.22	-	176.86	56.98	30.52	26.96	43.26	-
295	Glu	8.4	121.63	4.22	2.06,2.00	2.29,2.29	-	-	176.81	57.18	29.91	36.33	-	-
296	Ala	8.25	124.69	4.25	1.41	-	-	-	178.22	53.08	18.89	-	-	-
297	Ala	8.19	122.69	4.25	1.42	-	-	-	178.39	53.08	18.85	-	-	-
298	Arg	8.18	119.96	4.23	1.86,1.86	1.63,1.71	3.21,3.21	-	177.11	57.03	30.61	26.94	43.43	-
299	Glu	8.37	121.29	4.27	2.04,2.04	2.28,2.28	-	-	177.03	57.14	30.01	36.07	-	-
300	Ser	8.3	116.72	4.39	3.91,3.91	-	-	-	175.1	59.09	63.66	-	-	-
301	Arg	8.22	122.51	4.29	1.84,1.79	1.63,1.63	3.20,3.20	-	176.52	56.72	30.76	27	43.24	-
302	Arg	8.26	122.44	4.31	1.89,1.81	1.67,1.64	3.21,3.21	-	176.64	56.6	30.49	27.06	43.26	-
303	Lys	8.14	121.5	4.26	1.85,1.77	1.47,1.41	1.68,1.68	2.99,2.99	176.71	56.56	33.01	24.76	28.93	42.2
304	Lys	8.25	122.88	4.26	1.82,1.80	1.46,1.41	1.69,1.69	2.99,2.99	176.61	56.66	33.05	24.69	28.97	-
305	Lys	8.32	122.89	4.22	1.74,1.74	1.38,1.40	1.69,1.69	2.98,2.98	176.43	56.74	32.84	24.66	28.91	42.12
306	Glu	8.37	122.05	4.26	1.88,1.96	2.19,2.19	-	-	176.05	56.54	30.49	36.09	-	-
307	Tyr	8.27	121.96	4.55	3.00,3.00	-	-	-	175.62	58.13	38.72	-	-	-
308	Val	7.99	123.2	3.98	2	0.92,0.92	-	-	175.82	62.37	32.85	20.95,20.95	-	-
309	Lys	8.3	125.51	4.22	1.79,1.79	1.45,1.45	1.77,1.77	3.00,3.00	176.65	56.78	32.88	24.67	28.94	41.95
310	Ser	8.27	117.28	4.41	3.89,3.89	-	-	-	174.67	58.44	63.73	-	-	-
311	Leu	8.28	124.21	4.32	1.65,1.65	1.62	0.91,0.85	-	177.51	55.53	42.15	26.83	23.37,24.76	-
312	Glu	8.28	120.49	4.23	2.03,1.92	2.25,2.25	-	-	176.11	56.77	30.2	36.12	-	-
313	Asn	8.36	116.45	4.7	2.82,2.73	-	7.60,6.89	-	174.71	53.22	38.83	176.81	-	-
314	Arg	8.22	122.1	4.39	1.77,1.88	1.62,1.62	3.20,3.20	-	175.34	56.24	30.84	26.91	43.38	-
315	Val	7.76	125.28	4.03	-	-	-	-	181.03	63.64	-	-	-	-

Table S23: H^N and N chemical shifts from the CREB 285-315 temperature titration.

Residue	9°C		25°C		35°C	
	H	N	H	N	H	N
285	-	-	-	-	-	-
286	8.68	122.82	8.54	122.94	8.49	122.8
287	8.52	122.88	8.48	123.05	8.47	123.12
288	8.3	122.66	8.23	122.46	8.2	122.25
289	8.41	124.5	8.38	124.71	8.36	124.63
290	8.33	123.06	8.25	123.38	8.22	123.45
291	8.37	121.44	8.32	121.54	8.31	121.53
292	8.4	122.28	8.31	122.33	8.28	122.23
293	8.5	119.68	8.42	116.63	8.4	119.53
294	-	-	8.34	121.66	-	-
295	8.44	121.63	8.4	121.63	8.39	121.65
296	8.32	124.4	8.25	124.69	8.22	124.68
297	8.23	122.67	8.19	122.69	8.17	122.74
298	8.2	120.17	8.18	119.96	8.17	119.98
299	8.41	121.4	8.37	121.29	8.36	121.32
300	8.36	116.54	8.3	116.72	8.29	116.73
301	-	-	8.22	122.51		
302	8.3	122.25	8.26	122.44	8.24	122.48
303	8.16	121.6	8.14	121.5	8.14	121.43
304	8.29	122.77	8.25	122.88	8.23	122.83
305	8.35	122.76	8.32	122.89	8.3	122.86
306	8.41	121.95	8.37	122.05	8.35	122.02
307	8.35	122.25	8.27	121.96	8.23	121.73
308	8.1	123.76	7.99	123.2	7.96	122.84
309	8.36	125.53	8.3	125.51	8.28	125.32
310	8.36	117.58	8.27	117.28	8.24	117.13
311	8.32	124.17	8.28	124.21	8.26	124.11
312	8.3	120.48	8.28	120.49	8.27	120.54
313	8.43	119.49	8.36	116.45	8.34	119.46
314	8.34	122.23	8.22	122.1	8.21	122.08
315	7.86	125.54	7.76	125.28	7.72	125.25

Table S24: Chemical shift table of CREB 277-315 assignments. * indicates the *cis* isomer of those residues surrounding P278.

Residue No.	Name	H	N	ND	HA	HB	HG	HD	C	CA	CB	CG	CD
278	Pro	-	-	-	4.3	2.26,2.26	-	3.72,3.86	177.5	63.97	31.91	27.54	50.61
279	Ala	8.53	124.05	-	4.18	1.33	-	-	178.92	53.43	18.79	-	-
279*	Ala*	8.53	125.75	-	-	-	-	-	-	-	-	-	-
280	Glu	8.33	119.87	-	4.11	1.95,1.95	2.29,2.29	-	177.63	57.54	29.78	36.31	-
280*	Glu*	8.49	120.65	-	-	-	-	-	-	-	-	-	-
281	Glu	8.24	121.74	-	4.07	1.92,1.92	2.27,2.27	-	177.59	57.68	29.76	36.06	-
281*	Glu*	8.51	122.76	-	-	-	-	-	-	-	-	-	-
282	Ala	8.21	123.59	-	4.1	1.34	-	-	179.1	53.87	18.52	-	-
283	Ala	8.02	121.71	-	4.12	1.35	-	-	179	53.74	18.54	-	-
284	Arg	7.99	119.51	-	4.12	1.78,1.78	1.69,1.69	3.22,3.22	177.5	57.37	30.3	27.17	43.28
285	Lys	8.06	120.65	-	4.07	1.74,1.74	1.55,1.43	-	177.63	57.63	32.62	25.13	29.09
286	Arg	-	-	-	4.21	1.86,1.86	1.70,1.60	3.23,3.23	-	-	-	-	-
287	Glu	8.23	121.47	-	4.11	1.86,1.95	2.28,2.28	-	177.48	57.5	29.92	36.17	-
288	Val	8.13	121.42	-	3.86	2.03	0.98	-	176.95	63.89	32.31	21.18,21.19	-
289	Arg	8.15	123.52	-	4.14	1.65,1.75	1.61,1.67	-	176.93	57.18	30.4	27.2	43.31
290	Leu	8.19	122.24	-	4.28	1.53,1.39	-	-	178.09	55.97	42.06	26.95	24.92,23.40
291	Met	8.24	121.13	-	4.38	2.11,2.11	2.56,2.67	-	177.07	56.4	32.02	-	-
292	Lys	8.27	121.94	-	4.24	1.84,1.84	1.49,1.43	-	177.24	57.25	32.58	24.8	29.01
293	Asn	8.37	119.58	112.71	4.6	2.86,2.86	-	6.88,6.94	177.04	54	38.46	-	-
294	Arg	8.28	121.87	-	4.26	1.81,1.81	1.64,1.72	-	177.32	57.3	30.44	27.21	43.37
295	Glu	8.35	121.61	-	4.07	1.96,1.96	2.30,2.30	-	177.06	57.37	29.82	36.24	-
296	Ala	8.21	124.38	-	4.2	1.37	-	-	178.51	53.24	18.72	-	-
297	Ala	8.13	122.75	-	4.2	1.37	-	-	178.6	53.2	18.69	-	-
298	Arg	8.12	120.14	-	4.21	1.81,1.79	1.63,1.72	-	177.32	57.26	30.46	27.2	43.36
299	Glu	8.34	121.45	-	4.19	1.91,1.91	2.30,2.30	-	177.18	57.22	29.85	36.06	-
300	Ser	8.28	116.68	-	4.32	3.87,3.87	-	-	175.23	59.22	63.47	-	-
301	Arg	8.16	122.52	-	4.23	1.90,1.83	-	-	176.7	56.78	30.38	27.24	43.32
302	Arg	8.11	121.53	-	4.19	1.87,1.87	1.69,1.61	-	176.61	56.62	30.6	27.16	43.3
303	Lys	8.23	122.44	-	4.21	1.84,1.84	1.43,1.43	-	176.77	56.58	32.86	24.86	29.04
304	Lys	8.22	122.96	-	4.16	1.74,1.67	1.43,1.43	-	176.63	56.62	32.92	24.79	29.12
305	Lys	8.29	123.02	-	4.15	1.65,1.65	1.39,1.39	1.75,1.75	176.47	56.68	32.76	24.7	-
306	Glu	8.35	122.17	-	4.19	1.91,1.81	2.20,2.10	-	176.1	56.48	30.26	36.1	-
307	Tyr	8.25	122.09	-	4.48	2.97,2.95	-	-	175.67	58.1	38.67	-	-
308	Val	7.99	123.31	-	3.93	1.95	0.91,0.91	-	175.85	62.35	32.79	20.88,20.80	-
309	Lys	8.29	125.62	-	4.16	1.58,1.73	1.45,1.45	-	176.67	56.74	32.79	24.73	29.09
310	Ser	8.26	117.47	-	4.36	3.82,3.81	4.74	-	174.66	58.39	63.65	-	-
311	Leu	8.26	124.39	-	4.32	1.63,1.63	-	-	177.52	55.44	42.11	26.89	23.26,24.89
312	Glu	8.26	120.69	-	4.22	2.03,1.92	2.25,2.25	-	176.11	56.69	30.09	36.16	-
313	Asn	8.35	119.68	113.21	4.71	2.73,2.83	-	6.88,7.58	174.73	53.17	38.77	177.01	-

Table S24 continued from previous page

314	Arg	8.21	122.38	-	4.39	1.86,1.77	1.60,1.60	3.19,3.19	175.35	56.17	30.76	26.94	43.31
315	Val	7.75	125.53	-	4.04	2.08	0.89,0.88	-	181.03	63.62	33.12	20.03	-

Table S25: Assignment report for CREB 277-315 at 25°C.

Category	% Assigned
All Residues	95
Type C	95
Type CA	95
Type CB	97.4359
Type CD	64.28571
Type CE	0
Type CG	88.57143
Type CZ	0
Type H	97.36842
Type HA	92.68293
Type HB	97.14286
Type HD	35.89744
Type HE	0
Type HG	85.71429
Type N	92.5
Type ND	100
Type NE	0
Type NH	0
Type NZ	0

Table S26: Chemical shifts from the assigned peaks during the non-CRE DNA titration with CREB 285-315.

Residue (μM)	0		50		100		150		300		600		900		1200		1500		1800	
	¹ H (ppm)	¹⁵ N (ppm)	¹ H (ppm)	¹⁵ N (ppm)	¹ H (ppm)	¹⁵ N (ppm)	¹ H (ppm)	¹⁵ N (ppm)	¹ H (ppm)	¹⁵ N (ppm)	¹ H (ppm)	¹⁵ N (ppm)	¹ H (ppm)	¹⁵ N (ppm)	¹ H (ppm)	¹⁵ N (ppm)	¹ H (ppm)	¹⁵ N (ppm)	¹ H (ppm)	¹⁵ N (ppm)
286	8.54253	122.79777	8.54746	122.78725	8.55193	122.77669	8.55816	122.77512	8.56691	122.74877	8.58148	122.71056	8.59371	122.71079	8.59789	122.69873	8.6051	122.68968	8.6172	122.6635
287	8.48281	123.05465	8.48364	123.02667	8.48474	123.00559	8.48848	122.99411	8.48629	122.93747	8.49067	122.84426	8.4922	122.83103	8.49094	122.80565	8.49311	122.77978	8.49478	122.7336
288	8.23442	122.39863	8.23004	122.36919	8.2275	122.34733	8.22806	122.33376	8.22149	122.30763										
289	8.3783	124.62204	8.37782	124.59675	8.37808	124.57559	8.38129	124.56327	8.37731	124.51244	8.37923	124.42925	8.38014	124.41384	8.37835	124.39593	8.37972	124.37251	8.38083	124.33044
290	8.25388	123.34935	8.25237	123.31352	8.25152	123.28179	8.25442	123.26383	8.24892	123.1852	8.24748	123.08111	8.24748	123.03829	8.24501	123.00326	8.24632	122.97493	8.24564	122.91433
291	8.32394	121.48725	8.32251	121.4646	8.32086	121.43563	8.32244	121.43069	8.31729	121.38126	8.31602	121.30871	8.315	121.29912	8.33219	122.27815	8.31303	121.26032	8.31177	121.22233
292	8.31303	122.22176	8.31422	122.2068	8.31591	122.19471	8.32036	122.19242	8.32065	122.16372	8.32698	122.12165	8.33229	122.1194	8.33219	122.10701	8.33578	122.09841	8.33913	122.0787
293	8.42389	119.51542	8.42441	119.50382	8.42666	119.49102	8.42777	119.48799	8.42627	119.46379	8.42803	119.42404	8.43116	119.42969	8.43002	119.42071	8.4326	119.41341	8.43373	119.3901
294	8.33802	121.62299	8.33922	121.63346	8.34098	121.6392	8.34493	121.64899	8.34009	121.50969	8.34944	121.73776	8.34799	121.71089	8.34528	121.69935	8.34612	121.684	8.34388	121.64956
295	8.39828	121.59387	8.39886	121.57503	8.39934	121.55662	8.40285	121.54009	8.40009	121.43426	8.40219	121.45384	8.40438	121.44358	8.40273	121.42788	8.40484	121.41234	8.40538	121.3895
296	8.24881	124.58498	8.2458	124.54855	8.24446	124.52136	8.24615	124.50261	8.23824	124.43426	8.23344	124.32474	8.23237	124.30618	8.22921	124.2758	8.2307	124.23087	8.23007	124.18758
297	8.18729	122.6885	8.1858	122.67705	8.1854	122.66759	8.18839	122.66842	8.18294	122.64387	8.18303	122.61508	8.18394	122.62416	8.18036	122.61362	8.18191	122.61074	8.18182	122.60548
298	8.17783	119.98889	8.17649	119.97718	8.17591	119.96614	8.17877	119.9652	8.17435	119.94167	8.17402	119.90851	8.17544	119.91525	8.17342	119.90708	8.17537	119.90374	8.17556	119.89101
299	8.36948	121.30065	8.36865	121.28952	8.36806	121.28047	8.37065	121.28259	8.36538	121.2613	8.3651	121.24083	8.3657	121.24387	8.3637	121.23744	8.36547	121.23315	8.36532	121.23358
300	8.3062	116.65152	8.30417	1.17E+02	8.30343	116.60957	8.30397	116.60498	8.30072	116.55466	8.2989	116.49016	8.29913	116.47338	8.29637	116.45454	8.29765	116.44154	8.29831	116.41491
301	8.21051	122.48915																		
302	8.26039	122.41396	8.26242	122.38657	8.264	122.35947	8.26879	122.34609	8.26863	122.28763	8.27451	122.20506	8.27916	122.18771	8.27884	122.16432	8.2827	122.14666	8.28723	122.11084
303	8.14205	121.44536	8.1414	121.42748	8.14195	121.41386	8.14306	121.41008	8.1421	121.37332	8.14361	121.32377	8.14656	121.31919	8.14522	121.30649	8.14781	121.29587	8.1501	121.27546
304	8.25011	122.82743	8.24853	122.80616	8.24843	122.78804	8.25151	122.78164	8.24673	122.73682	8.24694	122.67581	8.24758	122.66409	8.24576	122.64536	8.24702	122.62578	8.2472	122.5799
305	8.31874	122.84216	8.31682	122.80985	8.31586	122.78446	8.3183	122.77107	8.31127	122.70588	8.30899	122.62124	8.30822	122.58873	8.30553	122.561	8.30665	122.53801	8.30445	122.48224
306	8.36713	121.99073	8.36488	121.95968	8.36294	121.93302	8.36436	121.91731	8.35531	121.84547	8.34944	121.73776	8.34799	121.71089	8.34528	121.6935	8.34612	121.684	8.34388	121.64956
307	8.26771	121.86521	8.26491	121.8653	8.26335	121.86511	8.26498	121.87253	8.25739	121.8748	8.25203	121.86773	8.25038	121.80167	8.24734	121.80265	8.2474	121.80481	8.2439	121.80218
308	7.9913	123.13823	7.99335	123.11438	7.99534	123.09418	7.99898	123.08559	8.0003	123.03931	8.00899	122.96912	8.01178	122.96168	8.01214	122.94224	8.01539	122.92688	8.01883	122.88482
309	8.30295	125.41666	8.30027	125.38689	8.29822	125.3608	8.29976	125.34763	8.29052	125.28394	8.28391	125.18392	8.28243	125.16818	8.27875	125.1404	8.27872	125.11846	8.2737	125.05755
310	8.27353	117.24421	8.27128	117.23037	8.2703	117.219	8.27263	117.21987	8.26581	117.19642	8.26286	117.15881	8.26239	117.16513	8.25933	117.15154	8.26008	117.14596	8.25653	117.11879
311	8.28079	124.10373	8.27641	124.09852	8.27555	124.09424	8.27389	124.09562	8.26208	124.08214	8.24976	124.06874	8.24697	124.07817	8.24175	124.07999	8.23925	124.09559	8.23031	124.08508
312	8.27553	120.48943	8.27298	120.48125	8.27149	120.47492	8.27318	120.4772	8.26418	120.45709	8.25954	120.43106	8.25692	120.4407	8.25302	120.43044	8.25316	120.42694	8.2485	120.40911
313	8.35941	119.42868	8.35813	119.41959	8.35773	119.41075	8.36006	119.4117	8.35452	119.39448	8.35295	119.36167	8.35281	119.36918	8.35023	119.35996	8.35131	119.35541	8.34878	119.33191
314	8.22297	122.07393	8.22142	122.0613	8.22063	122.06813	8.22293	122.07959	8.21825	122.06446	8.21298	122.11353	8.21238	122.11353	8.20891	122.1011	8.20891	122.08949	8.20587	122.05827
315	7.75841	125.28815	7.75764	125.28376	7.75866	125.28198	7.75906	125.287	7.75813	125.28247	7.75814	125.28921	7.75814	125.28921	7.75824	125.28746	7.75788	125.28706	7.75842	125.28256

Table S27: Chemical shift table for CREB 285-315 in complex with non-CRE DNA. Data collected at 25°C.

Residue No.	Name	H	N	ND	HA	HD	C	CA	CB	CG
285	Lys	-	-	-	-	-	176.81	57.02	33.27	-
286	Arg	8.62	122.67	-	-	-	176.28	56.77	30.69	-
287	Glu	8.49	122.74	-	-	-	174.64	57.17	30.12	-
288	Val	8.2	122.06	-	-	-	176.33	63.15	32.58	-
289	Arg	8.38	124.31	-	-	-	176.27	56.66	30.7	-
290	Leu	8.25	122.91	-	-	-	177.52	55.56	42.24	-
291	Met	8.31	121.21	-	-	-	176.55	55.98	32.74	-
292	Lys	8.34	122.07	-	-	-	176.71	57	32.83	-
293	Asn	8.43	119.4	112.79	-	6.97,7.67	175.71	53.71	38.66	176.71
294	Arg	8.34	121.65	-	-	-	176.99	57.24	30.56	-
295	Glu	8.4	121.39	-	-	-	176.91	57.46	29.92	-
296	Ala	8.23	124.18	-	-	-	178.34	53.25	18.9	-
297	Ala	8.18	122.6	-	-	-	178.48	53.25	18.82	-
298	Arg	8.17	119.88	-	-	-	177.22	57.29	30.68	-
299	Glu	8.36	121.23	-	-	-	177.05	57.26	30.01	-
300	Ser	8.3	116.41	-	-	-	175.25	59.29	63.59	-
301	Arg	8.24	122.38	-	-	-	176.64	56.92	30.68	-
302	Arg	8.29	122.1	-	-	-	176.72	57	30.54	-
303	Lys	8.15	121.31	-	-	-	176.86	56.91	33.02	-
304	Lys	8.24	122.55	-	-	-	176.71	56.94	33.04	-
305	Lys	8.3	122.47	-	-	-	176.54	56.96	32.86	-
306	Glu	8.34	121.65	-	-	-	176.18	56.69	30.42	-
307	Tyr	8.24	121.89	-	-	-	175.73	58.34	38.72	-
308	Val	8.02	122.86	-	3.95	-	175.94	62.65	32.81	-
309	Lys	8.27	125.06	-	-	-	176.69	56.9	32.87	-
310	Ser	8.26	117.12	-	-	-	174.64	58.6	63.77	-
311	Leu	8.23	124.11	-	-	-	177.48	55.56	42.21	-
312	Glu	8.25	120.41	-	-	-	176.05	56.8	30.24	-
313	Asn	8.35	119.34	112.94	-	7.60,6.90	176.65	53.24	38.86	176.95
314	Arg	8.2	122.06	-	-	-	175.27	56.28	30.9	-
315	Val	7.77	125.3	-	-	-	-	-	-	-

Table S28: Assignment report for CREB 285-315 with nCRE DNA.

Category	% Assigned
All Residues	96.875
Type C	93.75
Type CA	93.75
Type CB	96.77
Type CD	0
Type CE	0
Type CG	6.67
Type CZ	0
Type H	96.77
Type HA	3.03
Type HB	0
Type HD	11.43
Type HE	0
Type HG	0
Type N	93.75
Type ND	100
Type NE	0
Type NH	0
Type NZ	0

Table S29: Chemical shift table for CREB 277-315 with non-CRE DNA.

Residue No.	Name	H	N	ND	HA	HB	HG	HD	C	CA	CB	CG	CD	CE
278	Pro	-	-	-	4.4	-	-	3.86,3.71	177.44	63.92	31.92	27.52	50.63	-
279	Ala	8.55	124.11	-	4.25	1.42	-	-	178.92	53.41	18.8	-	-	-
280	Glu	8.35	119.87	-	4.18	2.04,2.04	2.27,2.27	-	177.63	57.55	29.76	36.27	-	-
281	Glu	8.24	121.69	-	4.15	1.98,1.98	2.27,2.27	-	177.56	57.69	29.71	36.06	-	-
282	Ala	8.21	123.66	-	4.14	1.42	-	-	179.05	53.84	18.52	-	-	-
283	Ala	8.03	121.79	-	4.18	1.45	-	-	178.97	53.7	18.54	-	-	-
284	Arg	8.01	119.63	-	4.21	1.86,1.86	1.69,1.69	-	177.46	57.35	30.24	27.17	43.28	-
285	Lys	8.09	120.8	-	4.15	1.83,1.83	-	-	177.7	57.68	32.6	25.12	29.09	-
286	Arg	8.1	121.48	-	4.18	1.85,1.85	-	-	176.87	57.47	30.4	27.36	43.34	-
287	Glu	8.24	121.5	-	4.16	2.01,2.01	2.25,2.25	-	177.52	57.57	29.81	36.15	-	-
288	Val	8.12	121.25	-	3.95	2.08	0.92,0.98	-	176.93	63.92	32.21	21.24	-	-
289	Arg	8.15	123.39	-	4.21	1.85,1.85	-	-	177	57.26	30.37	27.18	43.32	-
290	Leu	8.21	122.13	-	4.25	1.62,1.62	-	-	178.17	56.09	41.93	26.94	23.45	-
291	Met	8.23	120.87	-	4.34	2.11,2.11	-	-	177.12	56.44	32	32	-	-
292	Lys	8.29	121.92	-	4.21	1.83,1.83	-	-	177.35	57.3	32.55	24.81	29	-
293	Asn	8.39	119.64	112.7	4.64	2.86,2.86	-	6.89,7.63	176.19	-	38.41	-	-	-
294	Arg	8.31	121.89	-	4.22	1.88,1.88	-	-	177.3	57.47	30.42	27.21	43.41	-
295	Glu	8.35	121.48	-	4.18	2.02,2.02	2.29,2.29	-	177.23	57.55	29.64	36.22	-	-
296	Ala	8.2	124.18	-	4.23	1.42	-	-	178.68	53.32	18.73	-	-	-
297	Ala	8.14	122.69	-	4.25	1.42	-	-	178.76	53.28	18.71	-	-	-
298	Arg	8.13	120.07	-	4.18	1.86,1.86	-	-	177.5	57.38	30.45	27.19	-	-
299	Glu	8.34	121.38	-	4.25	-	-	-	177.27	57.29	29.81	36.04	-	-
300	Ser	8.28	116.57	-	4.36	3.92,3.93	-	-	175.44	59.37	63.45	-	-	-
301	Arg	8.19	122.45	-	4.27	-	-	-	176.9	57.02	30.37	27.31	43.34	-
302	Arg	8.11	121.43	-	4.21	1.85,1.85	-	-	176.85	56.88	30.5	27.16	43.29	-
303	Lys	8.26	122.18	-	4.23	1.83,1.78	-	-	176.98	56.78	30.29	-	-	-
304	Lys	8.22	122.8	-	4.25	1.84,1.78	-	-	176.84	56.85	32.85	24.84	-	-
305	Lys	8.28	122.68	-	4.19	1.75,1.75	-	-	176.69	56.86	32.76	24.76	-	-
306	Glu	8.32	121.91	-	4.22	1.88,1.88	2.23,2.23	-	176.28	56.64	30.29	36.14	-	-
307	Tyr	8.23	122.19	-	4.52	2.98,2.98	-	-	175.85	58.21	38.65	-	-	-
308	Val	8.01	123.11	-	3.95	1.99	0.89,0.89	-	176.02	62.58	32.64	20.95	-	-
309	Lys	8.26	125.24	-	4.2	1.80,1.75	1.44,1.44	-	176.82	56.83	32.76	24.75	29.08	42.08
310	Ser	8.25	117.36	-	4.4	3.87,3.87	-	-	174.74	58.47	63.62	-	-	-
311	Leu	8.22	124.31	-	4.28	1.65,1.59	-	-	177.59	56.04	42.11	26.8	24.06	-
312	Glu	8.24	120.71	-	-	-	-	-	-	56.7	30.26	36.16	-	-
312	Glu	-	-	-	4.2	1.91,1.91	2.10,2.10	-	176.13	-	-	-	-	-
313	Asn	8.34	119.6	113.22	4.71	2.72,2.81	-	6.89,7.59	174.74	53.23	38.87	-	-	-
314	Arg	8.21	122.3	-	4.37	1.74,1.86	1.59,1.59	3.18,3.18	175.37	56.17	30.73	26.99	43.31	-
315	Val	7.75	125.62	-	4.02	2.07	0.88,0.88	-	181.04	-	-	-	-	-

Table S30: Assignment Report of CREB 277-315 with non-CRE DNA.

Category	% Assigned
All Residues	95
Type C	95
Type CA	90
Type CB	94.87179
Type CD	53.57143
Type CE	12.5
Type CG	80
Type CZ	0
Type H	97.36842
Type HA	92.68293
Type HB	88.57143
Type HD	20.51282
Type HE	0
Type HG	42.85714
Type N	92.5
Type ND	100
Type NE	0
Type NH	0
Type NZ	0

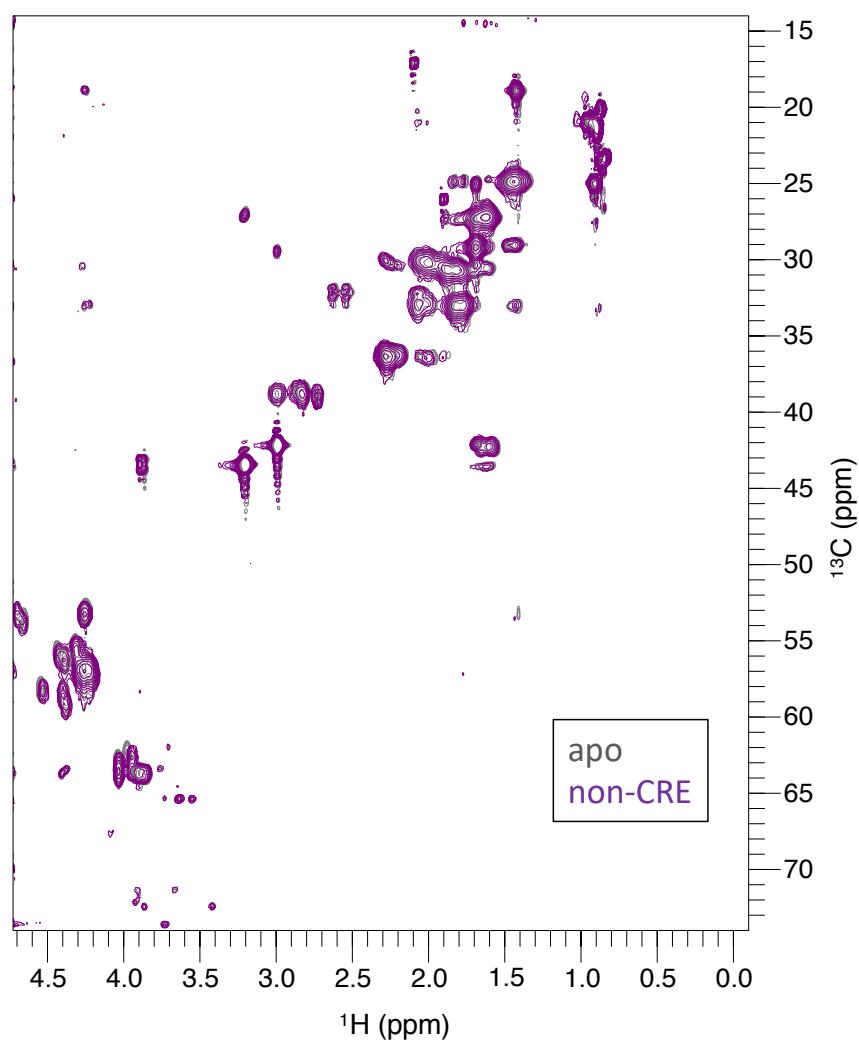


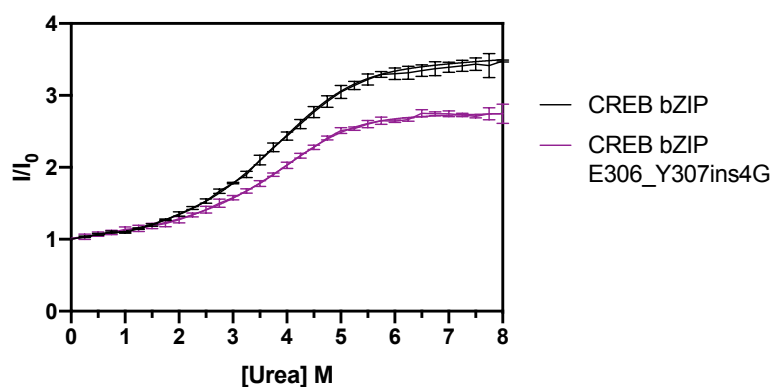
Figure S20: Limited carbon chemical shift changes with non-CRE DNA addition to CREB 285-315. ^{13}C -HSQC of CREB 285-315 in the apo and non-CRE bound states, collected at 25°C at 600 MHz.

Table S31: Chemical shift table of CREB 285-315 in complex with CRE-DNA, collected at 35°C.

Residue No.	Name	H	N	HA	HB	HG	HD	C	CA	CB	CG	CD
285	Lys	-	-	4.43	1.92,1.92	1.56,1.56	-	176.86	57.05	33.25	24.79	-
286	Arg	8.63	122.66	4.4	1.92,1.92	1.70,1.78	3.32,3.32	176.34	56.89	30.69	27.14	43.43
287	Glu	8.53	122.65	4.38	2.11,2.11	2.37,2.37	-	176.76	57.26	30.17	36.62	-
288	Val	8.2	121.89	4.16	2.19	1.06,1.06	-	176.38	63.27	32.56	21.05,21.05	-
289	Arg	8.39	124.35	4.44	1.92,1.92	1.73,1.73	3.32,3.32	176.38	56.76	30.71	27.26	43.44
290	Leu	8.27	122.88	4.44	1.70,1.78	1.7	1.03,1.03	177.61	55.68	42.39	27.06	24.50,24.50
291	Met	8.33	121.28	4.4	1.94,1.94	-	-	176.91	56.05	33.01	33.01	-
292	Lys	8.28	122.58	4.37	1.92,1.92	1.55,1.55	-	176.79	57.01	32.88	24.9	29.04
293	Asn	-	112.63	-	-	-	7.04	-	-	-	-	-
293	Asn	8.47	117.19	4.54	2.96,2.96	-	7.72	175.7	53.92	38.69	-	-
294	Arg	8.36	122.03	4.37	2.01,1.94	1.77,1.77	3.33,3.33	176.98	57.16	30.62	27.44	43.41
295	Glu	8.42	121.51	4	2.18,2.10	2.40,2.40	-	176.94	57.52	30.05	36.42	-
296	Ala	8.25	124.34	4.38	1.53	-	-	178.33	53.33	18.98	-	-
297	Ala	8.24	122.74	4.37	1.54	-	-	178.46	53.36	18.95	-	-
298	Arg	8.19	119.91	4.34	2.00,1.95	1.75,1.82	3.34,3.34	177.22	57.39	30.66	27.31	43.58
299	Glu	8.39	121.31	4.4	2.18,2.12	2.40,2.40	-	177.1	57.39	30.07	-	-
300	Ser	8.35	116.54	4.37	4.03,4.03	-	-	175.34	59.45	63.78	-	-
301	Arg	-	-	4.37	1.92,1.99	1.75,1.75	3.31,3.31	176.68	57.03	30.73	-	-
302	Arg	8.32	122.21	4.41	1.93,2.01	1.75,1.82	3.33,3.33	176.79	56.99	30.66	27.35	43.49
303	Lys	8.19	121.33	4.37	1.90,1.96	1.59,1.54	1.80,1.80	176.69	57.01	33.03	24.77	28.54
304	Lys	8.32	122.48	-	-	-	-	-	57.04	-	-	-
305	Lys	-	-	-	-	-	-	176.59	56.89	32.88	24.76	29.04
306	Glu	8.37	121.66	4.36	2.01,2.09	2.33,2.30	-	176.28	56.89	30.43	36.3	-
307	Tyr	8.26	121.83	4.66	3.11,3.11	-	-	175.84	58.39	38.75	-	-
308	Val	8.04	122.6	4.08	2.13	1.01,1.01	-	176.05	62.82	33.02	21.04,21.04	-
309	Lys	8.29	124.93	4.35	1.89,1.92	1.56,1.56	-	176.77	57	32.93	-	-
310	Ser	8.27	117.04	4.33	4.01,4.01	-	-	174.72	58.71	63.83	-	-
311	Leu	8.25	124.17	4.42	1.73,1.73	1.73	1.01,1.01	177.54	55.7	42.3	27.11	23.70,24.77
312	Glu	8.29	120.51	4.34	2.14,2.04	2.34,2.34	-	176.16	56.99	30.32	36.28	-
313	Asn	8.36	116.11	4.81	2.94,2.84	-	7.62	174.72	53.4	38.91	-	-
313	Asn	-	112.87	-	-	-	6.93	-	-	-	-	-
314	Arg	8.23	122.11	4.5	1.99,1.88	1.74,1.74	3.30,3.30	175.3	56.39	30.94	27.03	43.46
315	Val	7.76	125.32	-	-	-	-	-	63.82	-	-	-

Table S32: Assignment report for CREB 285-315 with CRE-DNA at 35°C.

Category	% Assigned
All Residues	100
Type C	93.54839
Type CA	100
Type CB	93.54839
Type CD	56.52174
Type CE	0
Type CG	70
Type CZ	0
Type H	90.32258
Type HA	90.32258
Type HB	91.22807
Type HD	68.57143
Type HE	0
Type HG	82.6087
Type N	90.32258
Type ND	0
Type NE	0
Type NH	0
Type NZ	0

**Figure S21:** Stability of the decoupled construct was similar to that of the pWT. Urea denaturation curve of CREB 285-339 4G_{ins}306-307 in comparison to pWT. 50 μ M of both proteins were examined by fluorescence intensity at 308 nm using intrinsic tyrosine fluorescence. Errors are SEM from triplicate of 3 independent measurements. Data was fit using equations in section 2.10.

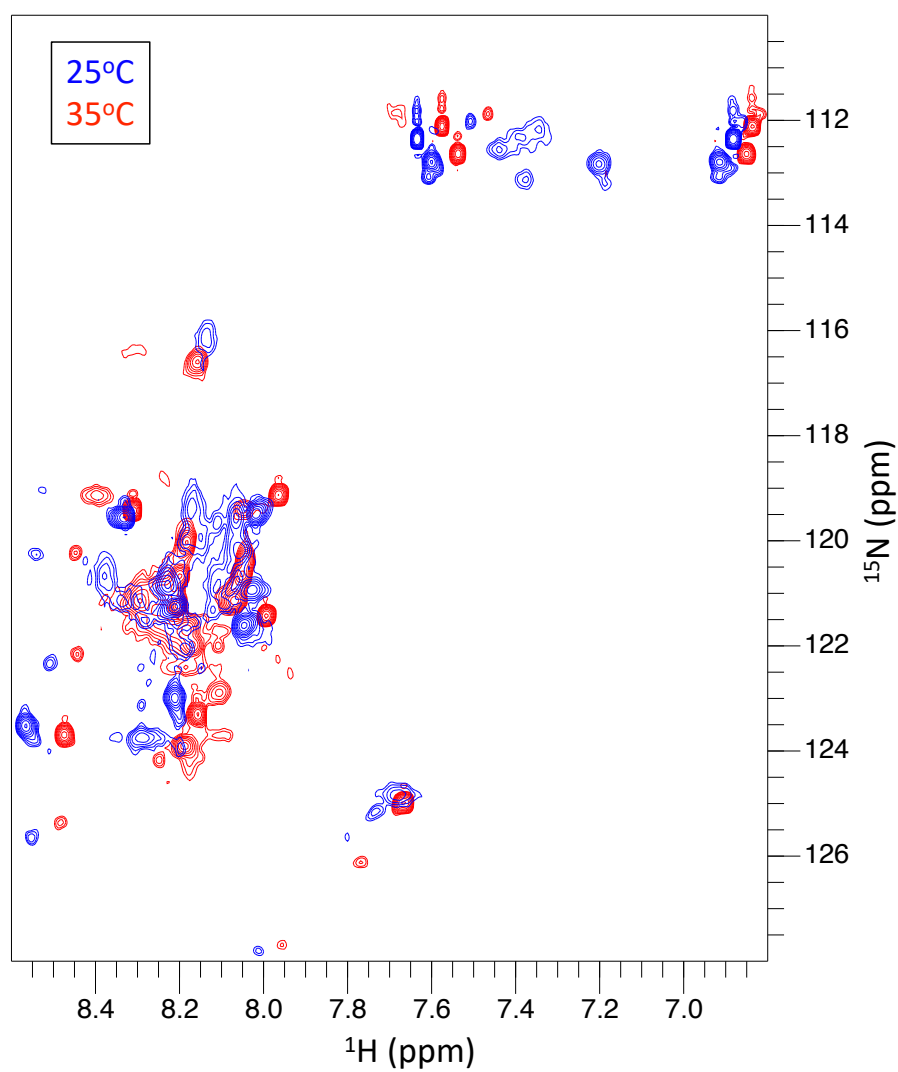


Figure S22: Varying temperature did not change exchange regime for CREB 277-315 with CRE-DNA. ^{15}N -HSQC of CREB 277-315 with CRE DNA, collected at two temperatures at 600 MHz.

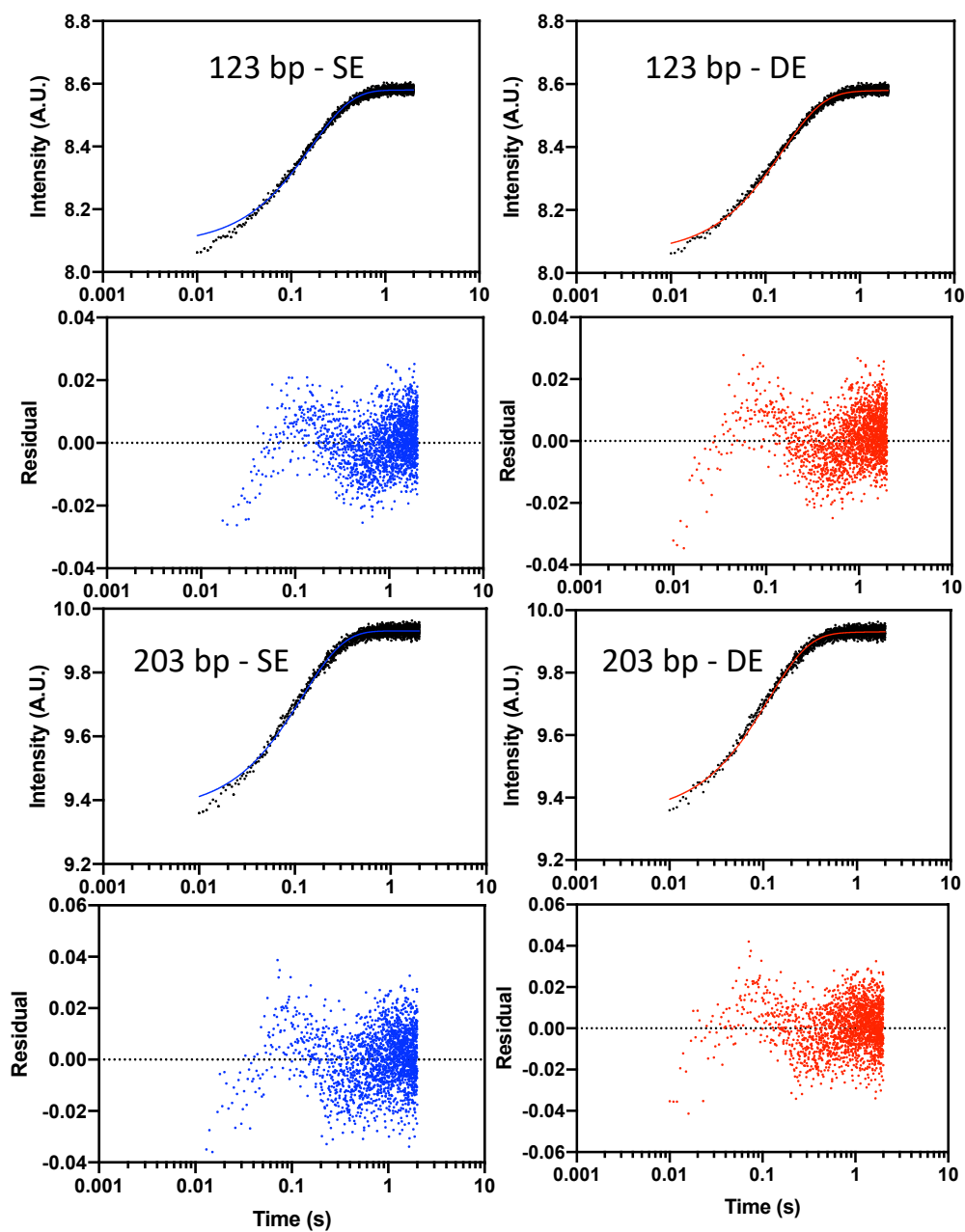


Figure S23: Stopped flow of CREB mixed with probe and competitor DNA was challenging to fit to a single or double exponential fit. Examples of single (SE) and double exponential (DE) fits for a selection of DNA probe lengths with the residuals shown below for each.

Table S33: Kinetics parameters identified from the stopped flow experiments with CREB bZIP constructs to probe DNA with varied concentrations of competitor DNA, where a large range of competitor concentrations were utilised. Errors are SEM. * $k_{on,N}$ can be estimated from k_{off} and $K_{d,N}$, with K_d of CREB to non-target DNA being 250 nM and

Protein	$k_{it,N}$ (s^{-1})	$k_{off,N}$ (s^{-1})	$k_{on,N}$ ($M^{-1} s^{-1}$ *)
CREB 285-339 pWT	$4.4 \pm 3.2 \times 10^5$	37 ± 4	$15 \pm 2 \times 10^7$
CREB 285-339 FD	$3.2 \pm 4.5 \times 10^5$	45 ± 6	$18 \pm 2 \times 10^7$
CREB 277-339 FD	$3.8 \pm 4.1 \times 10^4$	25 ± 2	$10 \pm 0.7 \times 10^7$

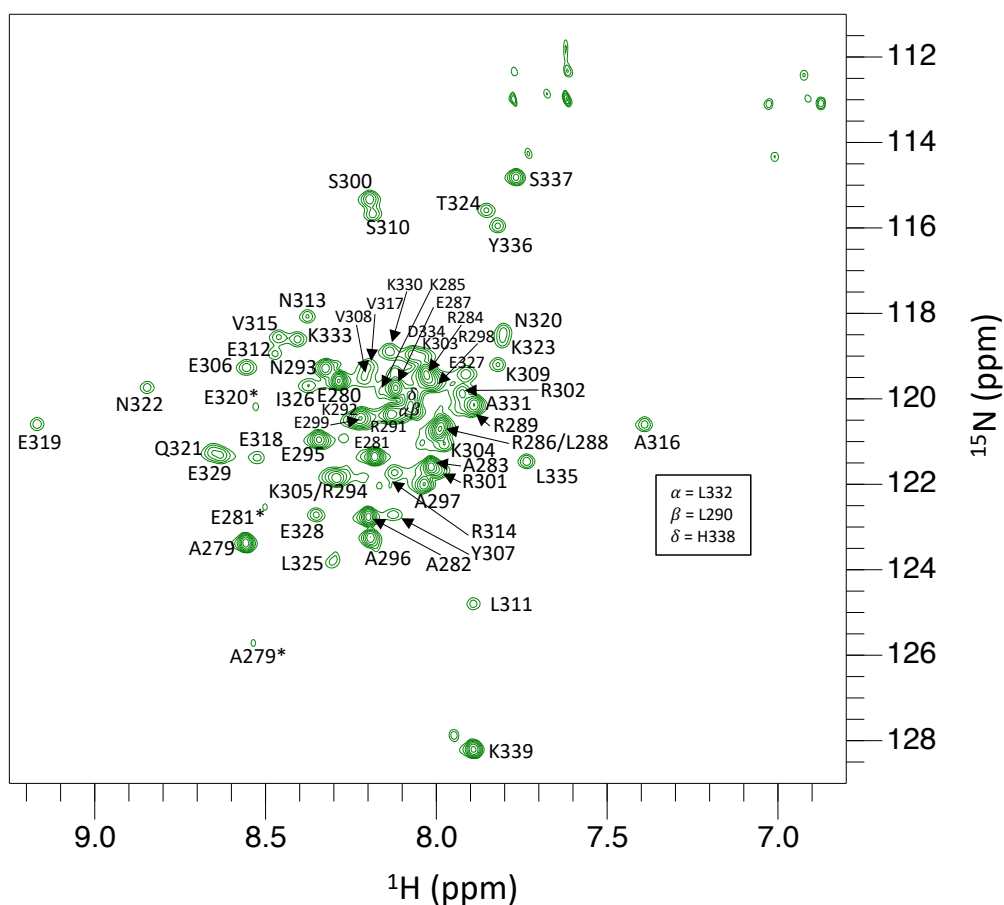


Figure S24: Assigned ^{15}N -HSQC of CREB 277-339 V288L/M291R. ^{15}N -HSQC of CREB 277-339 V288L/M291R was assigned at 25°C. Residue K303 is not assigned and is therefore missing.

Table S34: Chemical shift table of assignments of CREB 277-339 V288L/M291R collected at 25°C.

Residue	Name	H	N	HA	HB	C	CA	CB
278	Pro	-	-	4.37	2.35,1.92	177.86	64.43	31.95
279	Ala	8.56	123.34	4.22	1.42	179.42	54.01	18.73
280	Glu	8.29	119.5	4.16	2.07,2.07	178.12	58.09	29.75
281	Glu	8.19	121.34	4.12	2.01,2.01	178.08	58.29	29.74
282	Ala	8.2	122.79	4.09	1.44	180.01	54.61	18.26
283	Ala	8.02	121.57	4.11	1.48	179.73	54.69	18.26
284	Arg	8.03	119.39	4.08	1.89,1.89	178.7	58.7	30.21
285	Lys	8.16	119.8	4.01	1.86,1.86	178.71	58.76	32.49
286	Arg	7.99	120.69	4.07	1.9	177.87	58.99	30.18
287	Glu	8.12	119.78	4.06	2.08,2.08	178.62	58.81	29.42
288	Leu	7.99	120.68	4.1	1.63,1.76	178.74	57.47	41.93
289	Arg	7.89	120.24	4.08	1.92,1.92	178.07	58.53	30.15
290	Leu	8.08	120.21	4.15	1.79,1.57	179.15	57.18	41.7
291	Arg	8.13	120.33	4.05	1.93,1.93	178.11	58.79	30.23
292	Lys	8.22	120.26	4.13	1.88,1.88	178.63	58.33	32.27
293	Asn	8.33	119.32	4.58	2.88,2.88	177.32	55.03	38.18
294	Arg	8.29	121.9	4.18	1.92,1.92	178.53	58.65	30.43
295	Glu	8.35	121.01	4.09	2.07,2.07	178.18	58.5	29.52
296	Ala	8.19	123.24	4.21	1.48	179.78	54.25	18.31
297	Ala	8.04	121.99	4.2	1.46	179.65	54.18	18.33
298	Arg	8.01	119.65	4.07	1.92,1.92	178.35	58.77	30.23
299	Glu	8.23	120.5	4.2	2.07,2.07	178.11	58.27	29.54
300	Ser	8.2	115.37	4.23	3.95,3.95	176.67	60.68	63.01
301	Arg	8	121.62	4.09	1.86,1.86	177.91	58.64	30.22
302	Arg	7.92	119.86	3.94	1.84,1.84	177.72	58.46	-
303	Lys	-	-	-	-	-	58.01	32.51
303	Lys	8.05	119.01	4.13	1.86,1.86	178.38	-	-
304	Lys	7.98	121.02	4.29	1.94,1.94	178.13	58.13	32.51
305	Lys	8.31	121.88	4.13	1.93,1.93	178.8	59.04	32.34
306	Glu	8.56	119.31	4.12	2.01,2.01	178.33	58.43	29.34
307	Tyr	8.13	122.72	4.19	-	177.98	60.97	-

Table S34 continued from previous page

308	Val	8.22	119.43	3.29	-	177.41	67.34	31.47
309	Lys	7.82	119.21	4.15	1.88,1.88	179.15	58.82	31.82
310	Ser	8.18	115.7	4.05	3.20,3.20	177.35	61.74	62.44
311	Leu	7.89	124.79	3.94	-	178.59	57.81	43.36
312	Glu	8.47	118.93	3.71	-	180.48	59.78	29.34
313	Asn	8.38	118.06	4.16	2.82,2.82	176.69	56.16	38.41
314	Arg	8.12	121.69	4.05	-	178.43	58.75	29.75
315	Val	8.46	118.53	3.39	-	176.79	67.35	31.66
316	Ala	7.39	120.62	4.17	1.52	181.27	55.46	18
317	Val	8.2	119.31	3.72	-	179.58	66.52	-
318	Leu	8.53	120.44	3.82	-	179.88	58.04	-
319	Glu	9.17	120.61	3.98	-	179.84	60.18	29.52
320	Asn	7.81	118.45	4.49	2.89,2.89	177.94	56.77	38.79
321	Gln	8.66	121.21	4.04	-	178.06	59.15	28.66
322	Asn	8.85	119.74	3.98	-	176.45	57.28	38.66
323	Lys	7.81	118.52	3.97	1.99,1.99	178.94	60.03	32.33
324	Thr	7.86	115.56	4.17	3.98	177.45	66.59	68.8
325	Leu	8.31	123.78	-	-	179.61	58.41	42.15
326	Ile	8.38	119.69	3.63	1.92	179.2	65.71	38.11
327	Glu	7.92	119.42	4.01	2.18,2.07	179.55	59.41	29.24
328	Glu	8.35	122.73	4.05	2.07,2.07	179.26	59.78	29.53
329	Leu	8.64	121.3	3.94	1.71,1.71	178.3	58.52	42.41
330	Lys	8.14	118.88	3.85	1.89,1.89	178.77	59.77	32.52
331	Ala	7.89	120.15	4.12	1.47	178.07	54.89	18
332	Leu	8.11	120.29	4.07	1.62,1.79	178.67	57.62	42.38
333	Lys	8.41	118.58	3.93	1.86,1.86	179.87	60.39	32.26
334	Asp	8.07	118.95	4.38	2.62,2.71	178.5	56.97	40.72
335	Leu	7.74	121.47	3.97	1.42,1.42	179.09	57.67	-
336	Tyr	7.81	115.9	4.41	2.73,2.73	176.33	58.37	37.95
337	Ser	7.76	114.85	4.31	3.93,3.93	174.23	59.23	63.32
338	His	8.08	119.93	4.66	3.14,3.28	173.87	55.92	29.74
339	Lys	7.88	128.21	-	-	181.42	57.87	-

Table S35: Assignment report of CREB 277-339 V288L/M291R collected at 25°C.

Category	% Assigned
All Residues	96.875
Type C	96.875
Type CA	96.875
Type CB	88.88889
Type CG	1.75439
Type H	98.3871
Type HA	92.30769
Type HB	78.07018
Type HE	5.55556
Type N	95.3125
Type NE	8.33333

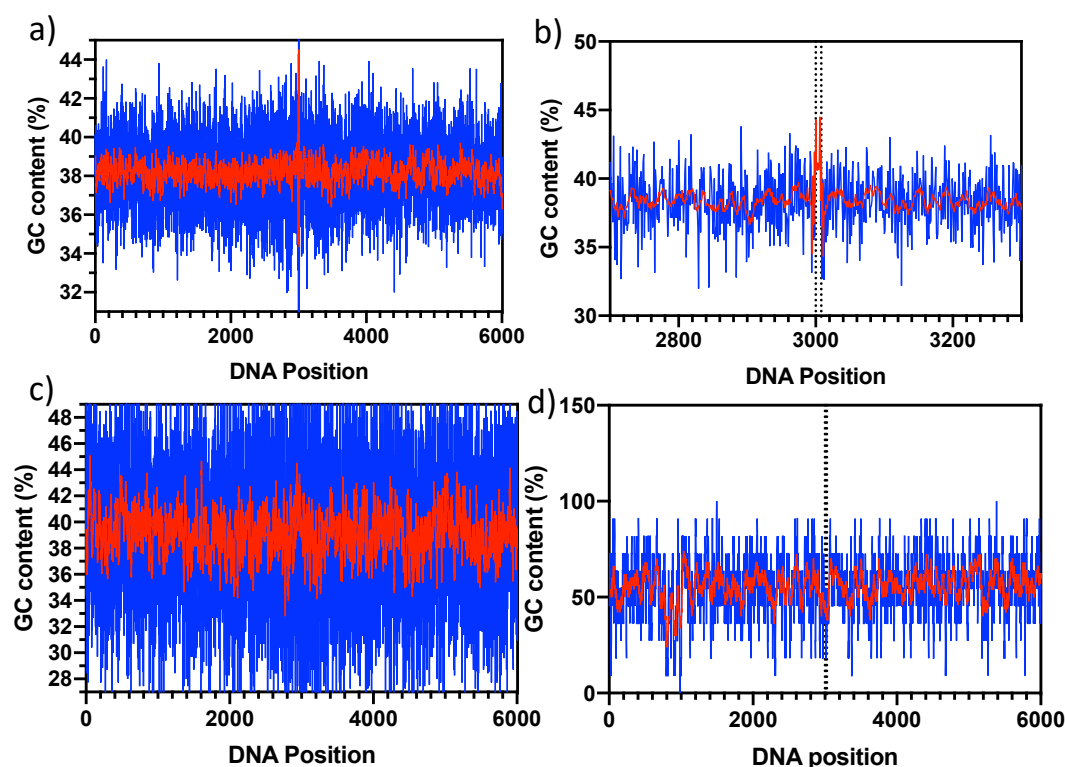


Figure S25: DNA GC content of flanking sequences of promoter sequences. Flanking sequences of 3000 bp were extracted from transcription factor DNA target sequences (CRE: 5'-TGACGTCA-3', AP-1: 5'-TGA(G/C)TCA-3', LacO: 5'-AATTGTGAGCGGATAACAATT-3', random control: 5'-CGTATACG-3') from *S. cerevisiae* S288C and *E. coli* K12. a-b) Yeast promoters AP-1 and CRE respectively. c) Random sequence in yeast. d) LacO in *E. coli*.

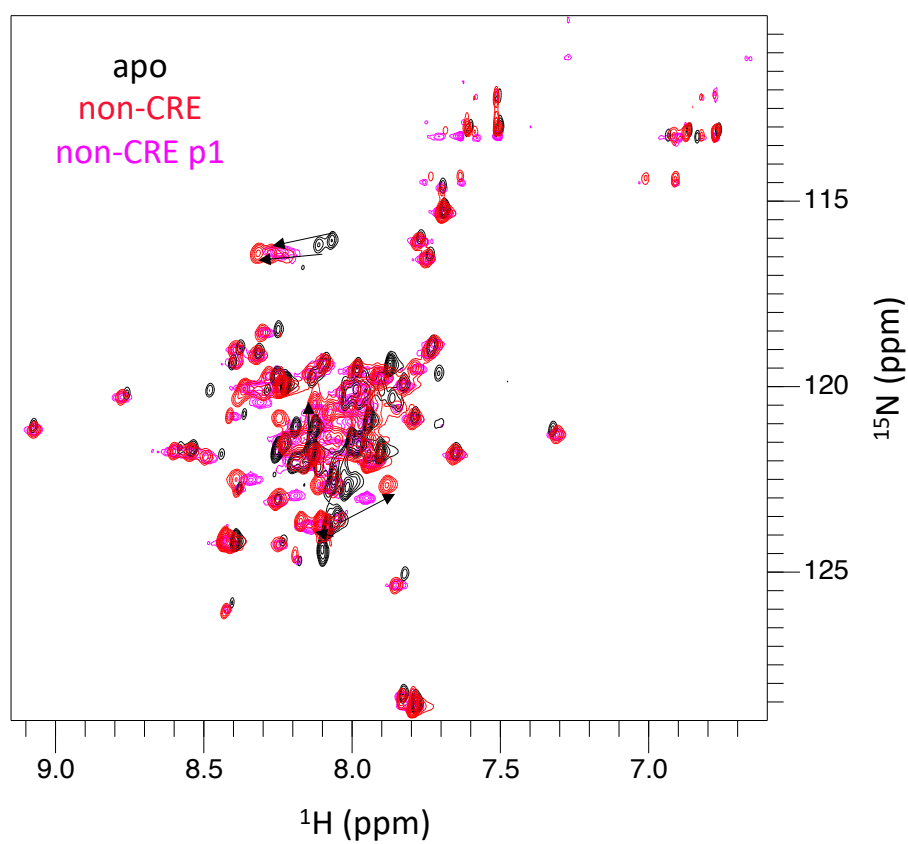


Figure S26: Chemical shifts of basic region in CREB 277-339 in complex with non-CRE p1 DNA appear intermediate between apo and non-CRE DNA. ^{15}N -BTROSY of CREB 277-339 apo or in complex with saturating concentrations of non-CRE DNA or non-CRE p1 DNA. All data collected at 35°C. Arrows indicate trajectories of a selection of peaks.

Table S36: Chemical shift table of CREB 277-339 with non-CRE p1 DNA at 35°C.

Residue	Name	H	N	HA	HB	C	CA	CB
278	Pro	-	-	4.41	1.96,2.39	177.44	63.91	31.94
279	Ala	8.42	124.21	4.27	1.45	178.9	53.41	18.91
280	Glu	8.25	119.92	4.22	2.07,2.07	177.65	57.63	29.88
281	Glu	8.13	121.73	4.19	2.04,2.04	177.63	57.79	29.82
282	Ala	8.1	123.77	4.18	1.46	179.23	54.02	18.55
283	Ala	7.93	122.02	4.22	1.49	179.16	53.89	18.56
284	Arg	7.91	119.81	4.22	1.92,2.11	177.58	57.66	30.41
285	Lys	7.97	120.63	4.15	1.89,1.89	178.07	58.08	32.64
286	Arg	7.98	121.53	4.19	-	176.98	57.98	30.35
287	Glu	8.11	121.09	-	-	177.75	58.17	30.12
288	Val	7.97	120.76	-	-	177.33	64.68	-
289	Arg	7.95	123	4.22	2.07,2.07	177.66	57.97	30.36
290	Leu	8.2	121.46	4.25	-	178.92	56.73	41.81
291	Met	8.12	120.72	4.22	-	177.74	57.41	32.56
292	Lys	8.22	121.82	-	-	178.31	58	32.29
293	Asn	8.34	120.04	-	-	177.1	54.8	38.12
294	Arg	8.34	122.47	-	-	178.52	58.64	30.36
295	Glu	8.21	121.41	4.18	-	178.27	58.63	29.6
296	Ala	8.15	123.74	-	-	180.03	54.39	18.42
297	Ala	8.08	122.72	4.25	-	177.63	54.34	18.6
298	Arg	7.97	120.08	4.19	-	178.81	59.06	-
299	Glu	8.13	121.3	-	-	178.16	58.25	-
300	Ser	8.25	116.31	-	-	177.08	61.08	-
301	Arg	8.08	121.63	-	-	178.66	57.71	32.58
302	Arg	8	121.67	-	-	176.87	59.16	-
303	Lys	-	-	-	-	179.42	-	-
304	Lys	8.06	121.59	-	-	-	59.16	-
306	Glu	8.25	119.68	-	-	-	58.79	-
307	Tyr	8.13	122.12	-	-	177.99	-	-
308	Val	8.03	120.19	-	-	-	67.38	-
309	Lys	7.78	119.53	-	-	179.3	-	-

Table S36 continued from previous page

310	Ser	8.24	116.36	-	-	-	-	-
311	Leu	7.85	125.36	-	-	-	-	-
312	Glu	8.39	119.37	-	-	-	-	-
313	Asn	8.29	118.55	-	-	-	-	-
314	Arg	-	-	-	-	178.53	-	-
315	Val	8.39	118.99	-	-	176.77	67.32	-
316	Ala	7.31	121.24	-	-	181.24	55.35	-
317	Val	8.14	119.67	-	-	-	66.42	-
318	Leu	8.48	121.9	-	-	179.85	57.95	-
319	Glu	9.07	121.14	-	-	179.76	60.16	-
320	Asn	7.73	118.87	-	-	177.94	56.69	38.7
321	Gln	8.59	121.75	-	-	178.08	59.12	-
322	Asn	8.78	120.25	-	-	176.52	57.26	-
323	Lys	7.74	119.08	-	-	179	59.92	-
324	Thr	7.77	116.11	-	-	176.91	66.44	-
325	Leu	8.25	124.23	-	-	178.7	-	-
326	Ile	8.3	120.13	-	-	179.05	59.01	-
327	Glu	7.83	119.99	4.31	-	179.57	59.34	30.94
328	Glu	8.25	123.1	-	-	179.24	59.53	-
329	Leu	8.54	121.75	-	-	178.39	58.44	-
330	Lys	8.09	119.34	-	-	178.79	59.71	-
331	Ala	7.79	120.88	-	-	180.79	54.8	17.94
332	Leu	8.03	120.56	-	-	178.71	57.56	-
333	Lys	8.32	119.15	-	-	179.77	60.25	-
334	Asp	7.99	119.61	-	-	178.42	56.85	40.61
335	Leu	7.65	121.85	-	-	178.99	57.59	42.12
336	Tyr	7.74	116.56	-	-	-	58.32	-
337	Ser	7.69	115.31	4.37	3.97,3.97	174.23	59.41	63.69
338	His	7.99	120.63	4.69	3.32,3.17	173.88	55.9	29.9
339	Lys	7.77	128.61	-	-	181.32	57.91	-

Table S37: Assignment report for the assignment of CREB 277-339 in complex with non-CRE p1 DNA at 35°C.

Category	% Assigned
All Residues	92.85714
Type C	88.57143
Type CA	88.57143
Type CB	48.52941
Type H	95.45455
Type HA	27.77778
Type HB	15.57377
Type N	90

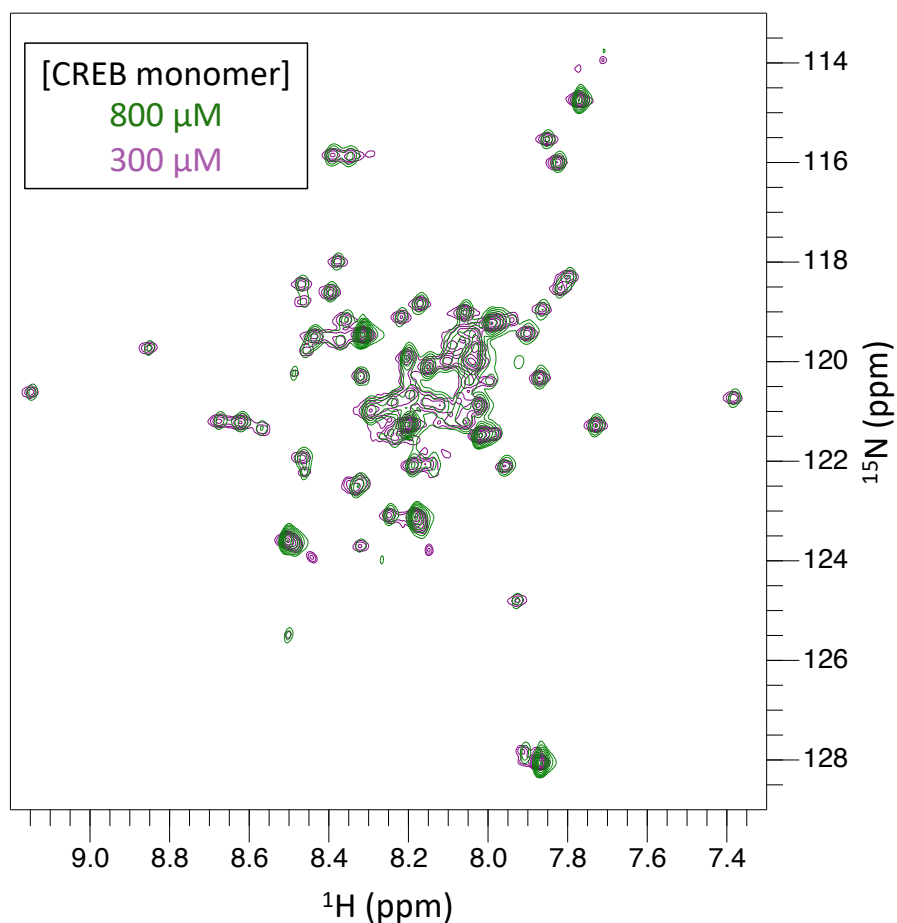


Figure S27: No chemical shift perturbation of CREB 277-339 in complex with non-CRE DNA was observed upon dilution. ^{15}N -BTOSY of CREB 277-339 in complex with a saturating concentration of non-CRE DNA. A dilution of the 800 μM sample was made to 300 μM in order to account for any potential differences with the non-CRE p1 DNA datasets.

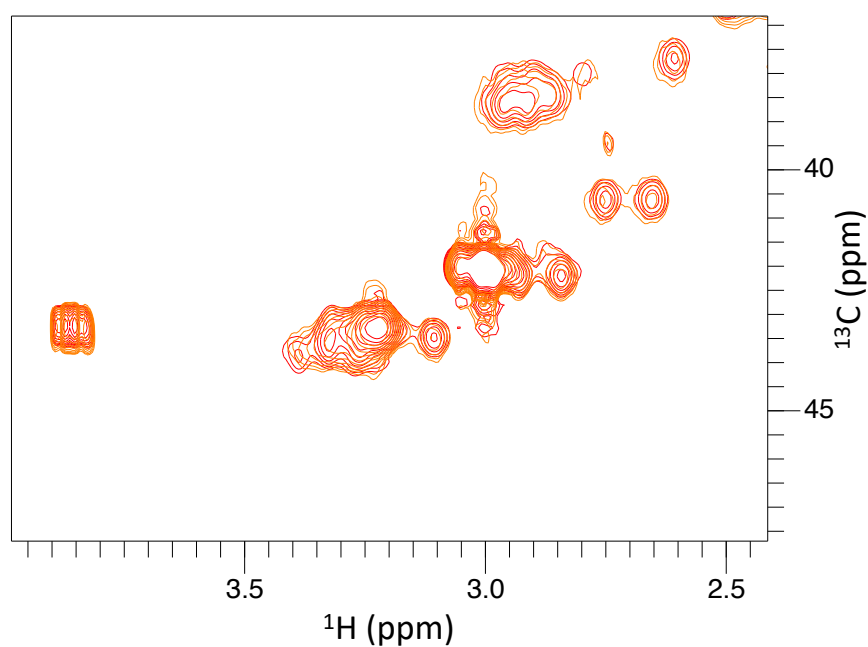


Figure S28: Small difference in dispersion of the arginine C δ between random and symmetrical non-target DNA. ^{13}C -HSQC of CREB 277-339 in complex with random (Red) and pseudo-symmetric (orange) non-target containing DNA sequences. Protein concentration was 800 μM and 300 μM respectively. Data collected at 35°C at 750 MHz.

I

References

- [1] Francis Crick. “Central dogma of molecular biology”. In: *Nature* 227.5258 (Aug. 1970), pp. 561–563.
- [2] Thomas Esquerré et al. “Genome-wide investigation of mRNA lifetime determinants in *Escherichia coli* cells cultured at different growth rates”. In: *BMC Genomics* 16.1 (Dec. 2015), p. 275.
- [3] Nahum Sonenberg and Alan G. Hinnebusch. “Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets”. In: *Cell* 136.4 (Feb. 2009), pp. 731–745.
- [4] Sandro Baldi, Philipp Korber, and Peter B. Becker. “Beads on a string—nucleosome array arrangements and folding of the chromatin fiber”. In: *Nature Structural and Molecular Biology* 27.2 (Feb. 2020), pp. 109–118.
- [5] Karolin Luger et al. “Crystal structure of the nucleosome core particle at 2.8 Å resolution”. In: *Nature* 389.6648 (Sept. 1997), pp. 251–260.
- [6] Jesse R. Dixon et al. “Topological domains in mammalian genomes identified by analysis of chromatin interactions”. In: *Nature* 485.7398 (May 2012), pp. 376–380.
- [7] Jacek R. Wiśniewski et al. “Mass spectrometric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue”. In: *Molecular and Cellular Proteomics* 6.1 (Jan. 2007), pp. 72–87.
- [8] John S Choy et al. “DNA methylation increases nucleosome compaction and rigidity”. In: *Journal of the American Chemical Society* 132.6 (2010), pp. 1782–1783.
- [9] Housheng Hansen He et al. “Nucleosome dynamics define transcriptional enhancers”. In: *Nature Genetics* 42.4 (2010), pp. 343–347.
- [10] Timothy J. Nott et al. “Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles”. In: *Molecular Cell* 57.5 (Mar. 2015), pp. 936–947.
- [11] Denes Hnisz et al. “A Phase Separation Model for Transcriptional Control”. In: *Cell* 169.1 (2017), pp. 13–23.
- [12] Amy R. Strom et al. “Phase separation drives heterochromatin domain formation”. In: *Nature* 547.7662 (July 2017), pp. 241–245.
- [13] David A. Garcia et al. “An intrinsically disordered region-mediated confinement state contributes to the dynamics and function of transcription factors”. In: *Molecular Cell* (Feb. 2021).

- [14] Yi Lu et al. “Phase separation of TAZ compartmentalizes the transcription machinery to promote gene expression”. In: *Nature Cell Biology* 22.4 (Apr. 2020), pp. 453–464.
- [15] Ann Boija et al. “Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains”. In: *Cell* 175.7 (Dec. 2018), 1842–1855.e16.
- [16] Remo Rohs et al. “Origins of specificity in protein-DNA recognition”. In: *Annual Review of Biochemistry* 79 (2010), pp. 233–269.
- [17] Richard N. Freiman. “Specific variants of general transcription factors regulate germ cell development in diverse organisms”. In: *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1789.3 (Mar. 2009), pp. 161–166.
- [18] D. Reinberg and R. G. Roeder. “Factors involved in specific transcription by mammalian RNA polymerase II. Purification and functional analysis of initiation factors IIB and IIE”. In: *Journal of Biological Chemistry* 262.7 (Mar. 1987), pp. 3310–3321.
- [19] Joseph L Kim, Dimitar B Nikolov, and Stephen K Burley. “Co-crystal structure of TBP recognizing the minor groove of a TATA element”. In: *Nature* 365.6446 (Oct. 1993), pp. 520–527.
- [20] N C Seeman, J M Rosenberg, and A Rich. “Sequence specific recognition of double helical nucleic acids by proteins”. In: *Proceedings of the National Academy of Sciences of the United States of America* 73.3 (Mar. 1976), pp. 804–808.
- [21] Maria A. Schumacher, Richard H. Goodman, and Richard G. Brennan. “The Structure of a CREB bZIP.Somatostatin CRE Complex Reveals the Basis for Selective Dimerization and Divalent Cation-enhanced DNA Binding”. In: *Journal of Biological Chemistry* 275.45 (Nov. 2000), pp. 35242–35247.
- [22] Ezra Peisach and Carl O. Pabo. “Constraints for zinc finger linker design as inferred from x-ray crystal structure of tandem Zif268-DNA complexes”. In: *Journal of Molecular Biology* 330.1 (June 2003), pp. 1–7.
- [23] Thomas E Ellenberger et al. “The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α Helices: Crystal structure of the protein-DNA complex”. In: *Cell* 71.7 (Dec. 1992), pp. 1223–1237.
- [24] Mitchell Lewis et al. “Crystal structure of the lactose operon repressor and its complexes with DNA and inducer”. In: *Science* 271.5253 (Mar. 1996), pp. 1247–1254.
- [25] Michael Y. Tolstorukov, Robert L. Jernigan, and Victor B. Zhurkin. “Protein-DNA Hydrophobic Recognition in the Minor Groove is Facilitated by Sugar Switching”. eng. In: *Journal of Molecular Biology* 337.1 (Mar. 2004), pp. 65–76.
- [26] Peter König and Timothy J. Richmond. “The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility”. In: *Journal of Molecular Biology* 233.1 (Sept. 1993), pp. 139–154.
- [27] Xianmin Zhu et al. “Differential regulation of mesodermal gene expression by Drosophila cell type-specific Forkhead transcription factors”. In: *Development* 139.8 (Apr. 2012), pp. 1457–1466.

- [28] Jonathan Enriquez et al. “Multi-step control of muscle diversity by Hox proteins in the *Drosophila* embryo”. In: *Development* 137.3 (Feb. 2010), pp. 457–466.
- [29] Peter F. Johnson and S. L. McKnight. “Eukaryotic transcriptional regulatory proteins”. In: *Annual Review of Biochemistry* 58 (1989), pp. 799–839.
- [30] Gerhard Mittler, Falk Butter, and Matthias Mann. “A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements”. In: *Genome Research* 19.2 (Feb. 2009), pp. 284–293.
- [31] Samuel A Lambert et al. “The Human Transcription Factors”. In: *Cell* 172.4 (2018), pp. 650–665.
- [32] Christian A.E.M. Spronk et al. “The solution structure of Lac repressor headpiece 62 complexed to a symmetrical lac operator”. In: *Structure* 7.12 (Dec. 1999), pp. 1483–1492.
- [33] Tom Ellenberger. “Getting a grip on DNA recognition: structures of the basic region leucine zipper, and the basic region helix-loop-helix DNA-binding domains”. In: *Current Opinion in Structural Biology* 4.1 (Jan. 1994), pp. 12–21.
- [34] Jerónimo Bravo et al. “The leukemia-associated AML1 (Runx1)-CBF β complex functions as a DNA-induced molecular clamp”. In: *Nature Structural Biology* 8.4 (Apr. 2001), pp. 371–378.
- [35] Yunje Cho et al. “Crystal structure of a p53 tumor suppressor-DNA complex: Understanding tumorigenic mutations”. In: *Science* 265.5170 (July 1994), pp. 346–355.
- [36] Scot A. Wolfe, Lena Nekludova, and Carl O. Pabo. “DNA recognition by Cys2His2 zinc finger proteins”. In: *Annual Review of Biophysics and Biomolecular Structure* 29.1 (June 2000), pp. 183–212.
- [37] Holli A Giebler, Isabelle Lemasson, and Jennifer K Nyborg. “p53 Recruitment of CREB Binding Protein Mediated through Phosphorylated CREB: a Novel Pathway of Tumor Suppressor Regulation”. In: *Molecular and Cellular Biology* 20.13 (July 2000), pp. 4849–4858.
- [38] Gwenael Badis et al. “Diversity and complexity in DNA recognition by transcription factors”. In: *Science* 324.5935 (2009), pp. 1720–1723.
- [39] A K Dunker et al. “Protein disorder and the evolution of molecular recognition: theory, predictions and observations.” eng. In: *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* (1998), pp. 473–84.
- [40] Peter E Wright and H Jane Dyson. “Intrinsically disordered proteins in cellular signalling and regulation”. In: *Nature Reviews Molecular Cell Biology* 16.1 (Jan. 2015), pp. 18–29.
- [41] H Jane Dyson and Peter E Wright. “Intrinsically unstructured proteins and their functions”. In: *Nature Reviews Molecular Cell Biology* 6.3 (Mar. 2005), pp. 197–208.
- [42] A. K. Dunker et al. “Intrinsic protein disorder in complete genomes.” In: *Genome informatics. Workshop on Genome Informatics* 11 (2000), pp. 161–171.
- [43] Jane R. Allison et al. “Determination of the free energy landscape of α -synuclein using spin label nuclear magnetic resonance measurements”. In: *Journal of the American Chemical Society* 131.51 (Dec. 2009), pp. 18314–18326.

- [44] Valéry Ozenne et al. “Mapping the Potential Energy Landscape of Intrinsically Disordered Proteins at Amino Acid Resolution”. In: *Journal of the American Chemical Society* 134.36 (Sept. 2012), pp. 15138–15148.
- [45] Pedro Romero et al. “Sequence Complexity of Disordered Protein”. In: *Proteins: Structure, Function and Genetics* 42.1 (Jan. 2001), pp. 38–48.
- [46] Christopher J Oldfield et al. “Coupled folding and binding with α -helix-forming molecular recognition elements”. In: *Biochemistry* 44.37 (2005), pp. 12454–12470.
- [47] Rashmi Sharma et al. “Fuzzy complexes: Specific binding without complete folding”. In: *FEBS Letters* 589 (Sept. 2015), pp. 2533–2542.
- [48] Liza Dahal, Sarah L. Shammas, and Jane Clarke. “Phosphorylation of the IDP KID Modulates Affinity for KIX by Increasing the Lifetime of the Complex”. In: *Biophysical Journal* 113.12 (2017), pp. 2706–2712.
- [49] S L Shammas et al. “Slow, reversible, coupled folding and binding of the spectrin tetramerization domain”. In: *Biophysical Journal* 103.10 (Nov. 2012), pp. 2203–2214.
- [50] J.J. Ward et al. “Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life”. In: *Journal of Molecular Biology* 337.3 (Mar. 2004), pp. 635–645.
- [51] Sara Light et al. “Long indels are disordered: A study of disorder and indels in homologous eukaryotic proteins”. In: *Biochimica et Biophysica Acta - Proteins and Proteomics* 1834.5 (May 2013), pp. 890–897.
- [52] Joseph B. Ahrens, Janelle Nunez-Castilla, and Jessica Siltberg-Liberles. “Evolution of intrinsic disorder in eukaryotic proteins”. In: *Cellular and Molecular Life Sciences* 74.17 (2017), pp. 3163–3174.
- [53] Helena G. Dos Santos, Janelle Nunez-Castilla, and Jessica Siltberg-Liberles. “Functional diversification after gene duplication: Paralog specific regions of structural disorder and phosphorylation in p53, p63, and p73”. In: *PLoS ONE* 11.3 (Mar. 2016). Ed. by Klaus Roemer, e0151961.
- [54] Zsuzsanna Dosztányi et al. “Disorder and sequence repeats in hub proteins and their implications for network evolution”. In: *Journal of Proteome Research* 5.11 (Nov. 2006), pp. 2985–2995.
- [55] Alex N. Nguyen Ba et al. “Proteome-wide discovery of evolutionary conserved sequences in disordered regions”. In: *Science Signaling* 5.215 (Mar. 2012).
- [56] Jörg Gsponer et al. “Tight regulation of unstructured proteins: From transcript synthesis to protein degradation”. In: *Science* 322.5906 (Nov. 2008), pp. 1365–1368.
- [57] Sarah L Shammas, Alexandra J Travis, and Jane Clarke. “Remarkably fast coupled folding and binding of the intrinsically disordered transactivation domain of cMyb to CBP KIX”. In: *Journal of Physical Chemistry B* 117.42 (Oct. 2013), pp. 13346–13356.
- [58] Sarah L. Shammas, Alexandra J. Travis, and Jane Clarke. “Allostery within a transcription coactivator is predominantly mediated through dissociation rate constants”. In: *Proceedings of the National Academy of Sciences of the United States of America* 111.33 (Aug. 2014), pp. 12055–12060.

- [59] Vytautas Iešmantavičius et al. “Helical propensity in an intrinsically disordered protein accelerates ligand binding”. In: *Angewandte Chemie - International Edition* 53.6 (Feb. 2014), pp. 1548–1551.
- [60] Joseph M Rogers, Chi T Wong, and Jane Clarke. “Coupled folding and binding of the disordered protein PUMA does not require particular residual structure”. In: *Journal of the American Chemical Society* 136.14 (Apr. 2014), pp. 5197–5200.
- [61] Michael D. Crabtree et al. “Conserved Helix-Flanking Prolines Modulate Intrinsically Disordered Protein:Target Affinity by Altering the Lifetime of the Bound Complex”. In: *Biochemistry* 56.18 (2017), pp. 2379–2384.
- [62] Tanja Mittag et al. “Structure/Function Implications in a Dynamic Complex of the Intrinsically Disordered Sic1 with the Cdc4 Subunit of an SCF Ubiquitin Ligase”. In: *Structure* 18.4 (Mar. 2010), pp. 494–506.
- [63] Julie D. Forman-Kay and Tanja Mittag. “From sequence and forces to structure, function, and evolution of intrinsically disordered proteins”. In: *Structure* 21.9 (Sept. 2013), pp. 1492–1499.
- [64] Chi W. Pak et al. “Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein”. In: *Molecular Cell* 63.1 (July 2016), pp. 72–85.
- [65] Michael Dzuricky et al. “De novo engineering of intracellular condensates using artificial disordered proteins”. In: *Nature Chemistry* 12.9 (Sept. 2020), pp. 814–825.
- [66] Dwiky Rendra Graha Subekti et al. “The Disordered Linker in p53 Participates in Nonspecific Binding to and One-Dimensional Sliding along DNA Revealed by Single-Molecule Fluorescence Measurements”. In: *Biochemistry* 56.32 (Aug. 2017), pp. 4134–4144.
- [67] Wade Borchers et al. “Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells”. In: *Nature chemical biology* 10.12 (Dec. 2014), pp. 1000–1002.
- [68] R Reeves and M S Nissen. “The A · T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure”. In: *Journal of Biological Chemistry* 265.15 (May 1990), pp. 8573–8582.
- [69] JEREMY N.S. EVANS et al. “¹H and ¹³C NMR assignments and molecular modelling of a minor groove DNA-binding peptide from the HMG-I protein”. In: *International Journal of Peptide and Protein Research* 45.6 (Jan. 1995), pp. 554–560.
- [70] Jeffrey R. Huth et al. “The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif”. In: *Nature Structural Biology* 4.8 (1997), pp. 657–665.
- [71] David A. Hill, Marisa L. Pedulla, and Raymond Reeves. “Directional binding of HMG-I(Y) on four-way junction DNA and the molecular basis for competitive binding with HMG-1 and histone H1”. In: *Nucleic Acids Research* 27.10 (Jan. 1999), pp. 2135–2144.

- [72] Jiangang Liu et al. “Intrinsic disorder in transcription factors”. In: *Biochemistry* 45.22 (June 2006), pp. 6873–6888.
- [73] R. G. Brennan and B. W. Matthews. “The helix-turn-helix DNA binding motif”. In: *Journal of Biological Chemistry* 264.4 (Feb. 1989), pp. 1903–1906.
- [74] Elizabeth M. Blackwood and Robert N. Eisenman. “Max: A helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc”. In: *Science* 251.4998 (Mar. 1991), pp. 1211–1217.
- [75] Stanislava Panova et al. “Mapping Hidden Residual Structure within the Myc bHLH-LZ Domain Using Chemical Denaturant Titration”. In: *Structure* 27.10 (Oct. 2019), 1537–1546.e4.
- [76] Min S. Lee et al. “Three-dimensional solution structure of a single zinc finger DNA-binding domain”. In: *Science* 245.4918 (Aug. 1989), pp. 635–637.
- [77] John H. Laity, H. Jane Dyson, and Peter E. Wright. “DNA-induced α -helix capping in conserved linker sequences is a determinant of binding affinity in Cys2-His2 zinc fingers”. In: *Journal of Molecular Biology* 295.4 (Jan. 2000), pp. 719–727.
- [78] S. J. Busch and P. Sassone-Corsi. “Fos, Jun and CREB basic-domain peptides have intrinsic DNA-binding activity enhanced by a novel stabilizing factor”. In: *Oncogene* 5.10 (1990), pp. 1549–1556.
- [79] Robert V Talanian et al. *Minimum Length of a Sequence-Specific DNA Binding Peptide1-of GCN4 involved in sequence-specific DNA contacts and provide a uniquely simple model for studying protein-DNA interactions in detail*. Tech. rep. 1992, pp. 6871–6875.
- [80] Susanne Cranz et al. “Monomeric and Dimeric bZIP Transcription Factor GCN4 Bind at the Same Rate to Their Target DNA Site †”. In: *Biochemistry* 43.3 (Jan. 2004), pp. 718–727.
- [81] Christine Berger et al. “Diffusion-controlled DNA recognition by an unfolded, monomeric bZIP transcription factor”. In: *FEBS Letters* 425.1 (Mar. 1998), pp. 14–18.
- [82] Peter Agre, Peter F Johnson, and Steven L McKnight. “Cognate DNA binding specificity retained after leucine zipper exchange between GCN4 and C/EBP”. In: *Science* 246.4932 (Nov. 1989), pp. 922–926.
- [83] Erin K. O’Shea, Rheba Rutkowski, and Peter S. Kim. “Mechanism of specificity in the Fos-Jun oncoprotein heterodimer”. In: *Cell* 68.4 (Feb. 1992), pp. 699–708.
- [84] Charles R. Vinson, Paul B. Sigler, and Steven L. McKnight. “Scissors-grip model for DNA recognition by a family of leucine zipper proteins”. In: *Science* 246.4932 (Nov. 1989), pp. 911–916.
- [85] J. N. Mark Glover and Stephen C. Harrison. “Crystal structure of the heterodimeric bZIP transcription factor c-Fos–c-Jun bound to DNA”. In: *Nature* 373.6511 (Jan. 1995), pp. 257–261.
- [86] John R. S. Newman and Amy E. Keating. “Comprehensive identification of human bZIP interactions with coiled-coil arrays.” In: *Science (New York, N.Y.)* 300.5628 (June 2003), pp. 2097–2101.

- [87] Carles M Llorca, Maren Potschin, and Ulrike Zentgraf. “bZIPs and WRKYs: two large transcription factor families executing two different functional strategies.” In: *Frontiers in plant science* 5.April (2014), p. 169.
- [88] C R Vinson, T Hai, and S M Boyd. “Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: Prediction and rational design”. In: *Genes and Development* 7.6 (June 1993), pp. 1047–1058.
- [89] James R Lundblad et al. “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”. In: *Journal of Biological Chemistry* 273.30 (July 1998), pp. 19251–19259.
- [90] Xiling Wu et al. “cAMP response element-binding protein monomers cooperatively assemble to form dimers on DNA”. In: *Journal of Biological Chemistry* 273.33 (Aug. 1998), pp. 20820–20827.
- [91] Kelly S. Thompson, Ernesto Freire, and Charles R. Vinson. “Thermodynamic Characterization of the Structural Stability of the Coiled-Coil Region of the bZIP Transcription Factor GCN4”. In: *Biochemistry* 32.21 (1993), pp. 5491–5496.
- [92] Hans Wendt, Antonio Baici, and Hans Rudolf Bosshard. “Mechanism of Assembly of a Leucine Zipper Domain”. In: *Journal of the American Chemical Society* 116.15 (July 1994), pp. 6973–6974.
- [93] Hans Wendt et al. “Kinetics of Folding of Leucine Zipper Domains”. In: *Biochemistry* 34.12 (Mar. 1995), pp. 4097–4107.
- [94] Clay Bracken et al. “Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: implications for the entropy of association with DNA”. In: *Journal of Molecular Biology* 285.5 (Feb. 1999), pp. 2133–2146.
- [95] Marilyn Emerson Holtzer et al. “Temperature dependence of the folding and unfolding kinetics of the GCN4 leucine zipper via $^{13}\text{C}\alpha$ -NMR”. In: *Biophysical Journal* 80.2 (Feb. 2001), pp. 939–951.
- [96] Xinmin Zhang et al. *Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues*. Tech. rep. 12. 2005, pp. 4459–4464.
- [97] Mark Nichols et al. “Phosphorylation of CREB affects its binding to high and low affinity sites: implications for cAMP induced gene transcription.” In: *The EMBO journal* 11.9 (1992), pp. 3337–3346.
- [98] Edward A. Felinski and Patrick G. Quinn. “The coactivator dTAFII110/hTAFII135 is sufficient to recruit a polymerase complex and activate basal transcription mediated by CREB”. In: *Proceedings of the National Academy of Sciences of the United States of America* 98.23 (Nov. 2001), pp. 13078–13083.
- [99] Edward A Felinski et al. “Recruitment of an RNA Polymerase II Complex Is Mediated by the Constitutive Activation Domain in CREB, Independently of CREB Phosphorylation”. In: *Molecular and Cellular Biology* 21.4 (Feb. 2001), pp. 1001–1010.
- [100] Garner et al. “Predicting Binding Regions within Disordered Proteins.” eng. In: *Genome informatics. Workshop on Genome Informatics* 10 (1999), pp. 41–50.

- [101] Li et al. “Predicting Protein Disorder for N-, C-, and Internal Regions.” eng. In: *Genome informatics. Workshop on Genome Informatics* 10 (1999), pp. 30–40.
- [102] Kenji Sugase, H. Jane Dyson, and Peter E. Wright. “Mechanism of coupled folding and binding of an intrinsically disordered protein”. In: *Nature* 447.7147 (June 2007), pp. 1021–1025.
- [103] John C. Chrivia et al. “Phosphorylated CREB binds specifically to the nuclear protein CBP”. In: *Nature* 365.6449 (Oct. 1993), pp. 855–859.
- [104] Haitao Wang et al. “cAMP Response Element-Binding Protein (CREB): A Possible Signaling Molecule Link in the Pathophysiology of Schizophrenia”. In: *Frontiers in Molecular Neuroscience* 11.August (Aug. 2018), pp. 1–14.
- [105] Adam J Shaywitz et al. “Magnitude of the CREB-Dependent Transcriptional Response Is Determined by the Strength of the Interaction between the Kinase-Inducible Domain of CREB and the KIX Domain of CREB-Binding Protein”. In: *Molecular and Cellular Biology* 20.24 (2000), pp. 9409–9422.
- [106] P. Sun et al. “Differential activation of CREB by Ca²⁺/calmodulin-dependent protein kinases type II and type IV involves phosphorylation of a site that negatively regulates activity”. In: *Genes and Development* 8.21 (Nov. 1994), pp. 2527–2539.
- [107] M Hagiwara et al. “Coupling of hormonal stimulation and transcription via the cyclic AMP-responsive factor CREB is rate limited by nuclear entry of protein kinase A.” In: *Molecular and Cellular Biology* 13.8 (Aug. 1993), pp. 4852–4859.
- [108] Johanna C. Craig et al. “Consensus and Variant cAMP-regulated Enhancers Have Distinct CREB-binding Properties”. In: *Journal of Biological Chemistry* 276.15 (2001), pp. 11719–11728.
- [109] B. M. Laoide et al. “The functional versatility of CREM is determined by its modular structure”. In: *EMBO Journal* 12.3 (Mar. 1993), pp. 1179–1191.
- [110] Marc M. Loriaux, Richard G. Brennan, and Richard H. Goodman. “Modulatory function of CREB · CREM α heterodimers depends upon CREM α phosphorylation”. In: *Journal of Biological Chemistry* 269.46 (Nov. 1994), pp. 28839–28843.
- [111] Yungdae Yun, Michele Dumoulin, and Joel F Habener. “Dna-binding and dimerization domains of adenosine 3,5- cyclic monophosphate-responsive protein creb reside in the carboxyl-terminal 66 amino acids”. In: *Molecular Endocrinology* 4.6 (1990), pp. 931–939.
- [112] Michael D Conkright et al. “TORCs: Transducers of regulated CREB activity”. In: *Molecular Cell* 12.2 (2003), pp. 413–423.
- [113] Jonathan R Moll et al. “Magnesium is required for specific DNA binding of the CREB B-ZIP domain”. In: *Nucleic Acids Research* 30.5 (Mar. 2002), pp. 1240–1246.
- [114] Sergey Shnitkind et al. “Structural Basis for Graded Inhibition of CREB:DNA Interactions by Multisite Phosphorylation”. In: *Biochemistry* 57.51 (2018), pp. 6964–6972.

- [115] S. M. Iguchi-Arigo and W. Schaffner. "CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation." In: *Genes & development* 3.5 (May 1989), pp. 612–619.
- [116] Joan S. Steffan et al. "The Huntington's disease protein interacts with p53 and CREB-binding protein and represses transcription". In: *Proceedings of the National Academy of Sciences of the United States of America* 97.12 (June 2000), pp. 6763–6768.
- [117] Carlos A. Saura and Jorge Valero. "The role of CREB signaling in Alzheimer's disease and other cognitive disorders". In: *Reviews in the Neurosciences* 22.2 (2011), pp. 153–169.
- [118] Yun Sik Choi et al. "CREB is a key regulator of striatal vulnerability in chemical and genetic models of Huntington's disease". In: *Neurobiology of Disease* 36.2 (Nov. 2009), pp. 259–268.
- [119] Irene E. Schauer et al. "CREB downregulation in vascular disease: A common response to cardiovascular risk". In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 30.4 (Apr. 2010), pp. 733–741.
- [120] Stephan Herzig et al. "CREB controls hepatic lipid metabolism through nuclear hormone receptor PPAR-gamma." In: *Nature* 426.6963 (Nov. 2003), pp. 190–193.
- [121] Changmoon Park, Judy L. Campbell, and William A. Goddard. "Can the monomer of the leucine zipper proteins recognize the dimer binding site without dimerization?" In: *Journal of the American Chemical Society* 118.18 (1996), pp. 4235–4239.
- [122] Qunhui Li and George A. Marzluf. "Determination of the *Neurospora crassa* CYS 3 sulfur regulatory protein consensus DNA binding site: Amino acid substitutions in the CYS3 bZIP domain that alter DNA-binding specificity". In: *Current Genetics* 30.4 (Sept. 1996), pp. 298–304.
- [123] Kenneth L. Seldeen et al. "Coupling of folding and DNA-binding in the bZIP domains of Jun-Fos heterodimeric transcription factor". In: *Archives of Biochemistry and Biophysics* 473.1 (May 2008), pp. 48–60.
- [124] Kenneth L. Seldeen et al. "Dissecting the role of leucine zippers in the binding of bZIP domains of Jun transcription factor to DNA". In: *Biochemical and Biophysical Research Communications* 394.4 (Apr. 2010), pp. 1030–1035.
- [125] Xu Wang et al. "Thermodynamic Characterization of the Folding Coupled DNA Binding by the Monomeric Transcription Activator GCN4 Peptide". In: *Biophysical Journal* 84.3 (Mar. 2003), pp. 1867–1875.
- [126] J. J. Kohler and A. Schepartz. "Kinetic studies of Fos · Jun · DNA complex formation: DNA binding prior to dimerization". In: *Biochemistry* 40.1 (2001), pp. 130–142.
- [127] Z I Santiago-Rivera et al. "Bacterial expression and characterization of the CREB bZip module: circular dichroism and 2D 1H-NMR studies". In: *Protein Sci* 2.9 (1993), pp. 1461–1471.

- [128] Jessica J. Hollenbeck and Martha G. Oakley. “GCN4 binds with high affinity to DNA sequences containing a single consensus half-site”. In: *Biochemistry* 39.21 (May 2000), pp. 6380–6389.
- [129] Robert V. Talanian, C. James Mcknight, and Peter S. Kim. *Sequence-specific DNA binding by a short peptide dimer*. Tech. rep. 4970. 1990, pp. 769–771.
- [130] Neil P Johnson et al. “Double-stranded DNA templates can induce alpha-helical conformation in peptides containing lysine and alanine: functional implications for leucine zipper and helix-loop-helix transcription factors.” In: *Proceedings of the National Academy of Sciences* 91.11 (May 1994), pp. 4840–4844.
- [131] M A Weiss et al. “Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA.” In: *Nature* 347.6293 (1990), pp. 575–8.
- [132] Dionisios Rentzeperis, Thorlakur Jonsson, and Robert T. Sauer. “Acceleration of the refolding of Arc repressor by nucleic acids and other polyanions”. In: *Nature Structural Biology* 6.6 (June 1999), pp. 569–573.
- [133] Steven J. Metallo and Alanna Schepartz. “Certain bZIP peptides bind DNA sequentially as monomers and dimerize on the DNA”. In: *Nature* 4.2 (1997), pp. 115–117.
- [134] Jennifer J. Kohler et al. “DNA specificity enhanced by sequential binding of protein monomers”. In: *Proceedings of the National Academy of Sciences of the United States of America* 96.21 (Oct. 1999), pp. 11735–11739.
- [135] V J Dwarki, M Montminy, and I M Verma. “Both the basic region and the ‘leucine zipper’ domain of the cyclic AMP response element binding (CREB) protein are essential for transcriptional activation.” In: *The EMBO journal* 9.1 (Jan. 1990), pp. 225–32.
- [136] Erin K. O’Shea, Rheba Rutkowski, and Peter S. Kim. “Evidence that the leucine zipper is a coiled coil”. In: *Science* 243.4890 (Jan. 1989), pp. 538–542.
- [137] Robert B. Winter, Otto G. Berg, and Peter H. von Hippel. “Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 3. The Escherichia coli lac Repressor-Operator Interaction: Kinetic Measurements and Conclusions”. In: *Biochemistry* 20.24 (1981), pp. 6961–6977.
- [138] Arthur D. Riggs, Suzanne Bourgeois, and Melvin Cohn. “The lac repressor-operator interaction. III. Kinetic studies”. In: *Journal of Molecular Biology* 53.3 (Nov. 1970), pp. 401–417.
- [139] Otto G. Berg, Robert B. Winter, and Peter H. von Hippel. “Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 1. Models and Theory”. In: *Biochemistry* 20.24 (Nov. 1981), pp. 6929–6948.
- [140] Jaehoh Shin and Anatoly B. Kolomeisky. “Facilitation of DNA loop formation by protein-DNA non-specific interactions”. In: *Soft Matter* 15.26 (July 2019), pp. 5255–5263. arXiv: 1901.10123.
- [141] Alessia Tempestini et al. “Sliding of a single lac repressor protein along DNA is tuned by DNA sequence and molecular switching”. In: *Nucleic Acids Research* 46.10 (June 2018), pp. 5001–5011.
- [142] E Marklund et al. “Mechanism of DNA surface exploration and operator bypassing during target search”. In: *bioRxiv* (2020).

- [143] Petter Hammar et al. “The lac repressor displays facilitated diffusion in living cells”. In: *Science* 336.6088 (June 2012), pp. 1595–1598.
- [144] Y. M. Wang, Robert H. Austin, and Edward C. Cox. “Single molecule measurements of repressor protein 1D diffusion on DNA”. In: *Physical Review Letters* 97.4 (July 2006), p. 048302.
- [145] Jason S. Leith et al. “Sequence-dependent sliding kinetics of p53”. In: *Proceedings of the National Academy of Sciences of the United States of America* 109.41 (2012), pp. 16552–16557.
- [146] Anahita Tafvizi et al. “A single-molecule characterization of p53 search on DNA.” In: *Proceedings of the National Academy of Sciences of the United States of America* 108.2 (2011), pp. 563–568.
- [147] Benjamin A. Lieberman and Steven K. Nordeen. “DNA intersegment transfer, how steroid receptors search for a target site”. In: *Journal of Biological Chemistry* 272.2 (1997), pp. 1061–1068.
- [148] Levani Zandarashvili et al. “Asymmetrical roles of zinc fingers in dynamic DNA-scanning process by the inducible transcription factor Egr-1”. In: *Proceedings of the National Academy of Sciences of the United States of America* 109.26 (June 2012), E1724–E1732.
- [149] Alexandre Esadze and Junji Iwahara. “Stopped-Flow Fluorescence Kinetic Study of Protein Sliding and Intersegment Transfer in the Target DNA Search Process”. In: *Journal of Molecular Biology* 426.1 (Jan. 2014), pp. 230–244.
- [150] Michaelleen Doucleff and G Marius Clore. “Global jumping and domain-specific intersegment transfer between DNA cognate sites of the multidomain transcription factor Oct-1”. In: *Proceedings of the National Academy of Sciences of the United States of America* 105.37 (2008), pp. 13871–13876.
- [151] Levani Zandarashvili et al. “Balancing between affinity and speed in target DNA search by zinc-finger proteins via modulation of dynamic conformational ensemble.” In: *Proceedings of the National Academy of Sciences of the United States of America* 112.37 (2015), E5142–5149.
- [152] Charalampos G. Kalodimos et al. “Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes”. In: *Science* 305.5682 (July 2004), pp. 386–389.
- [153] Milagros Castellanos, Nivin Mothi, and Victor Muñoz. “Eukaryotic transcription factors can track and control their target genes using DNA antennas”. In: *Nature Communications* 11.1 (Dec. 2020), p. 540.
- [154] Ariel Afek et al. “Protein-DNA binding in the absence of specific base-pair recognition”. In: *Proceedings of the National Academy of Sciences of the United States of America* 111.48 (2014), pp. 17140–17145.
- [155] Sneha Munshi et al. “Tunable order–disorder continuum in protein–DNA interactions”. In: *Nucleic Acids Research* 46.17 (Aug. 2018), pp. 8700–8709.
- [156] J Marley, M Lu, and C Bracken. “A method for efficient isotopic labeling of recombinant proteins.” In: *J. Biomol. NMR* 20.1 (2001), pp. 71–5. arXiv: 0005074v1 [arXiv:astro-ph].

- [157] Stanley C Gill and Peter H von Hippel. “Calculation of protein extinction coefficients from amino acid sequence data”. In: *Analytical Biochemistry* 182.2 (Nov. 1989), pp. 319–326.
- [158] C. Nick Pace et al. “How to measure and predict the molar absorption coefficient of a protein”. In: *Protein Science* 4.11 (Nov. 1995), pp. 2411–2423.
- [159] Paul Schanda, Hélène Van Melckebeke, and Bernhard Brutscher. “Speeding up three-dimensional protein NMR experiments to a few minutes”. In: *Journal of the American Chemical Society* 128.28 (2006), pp. 9042–9043.
- [160] Frank Delaglio et al. “NMRPipe: A multidimensional spectral processing system based on UNIX pipes”. In: *Journal of Biomolecular NMR* 6.3 (Nov. 1995), pp. 277–293.
- [161] Wim F. Vranken et al. “The CCPN data model for NMR spectroscopy: Development of a software pipeline”. In: *Proteins: Structure, Function and Genetics* 59.4 (Apr. 2005), pp. 687–696.
- [162] Mike P. Williamson. “Using chemical shift perturbation to characterise ligand binding”. In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 73 (Aug. 2013), pp. 1–16.
- [163] Lawrence Berliner. *Protein NMR: Modern techniques and biomedical applications*. 2015, pp. 1–185. arXiv: [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- [164] Yang Shen et al. “TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts”. In: *Journal of Biomolecular NMR* 44.4 (2009), pp. 213–223.
- [165] Joseph A. Marsh et al. “Sensitivity of secondary structure propensities to sequence differences between α - and γ -synuclein: Implications for fibrillation”. In: *Protein Science* 15.12 (Dec. 2006), pp. 2795–2804.
- [166] David L. Daugherty and Samuel H. Gellman. “A fluorescence assay for leucine zipper dimerization: Avoiding unintended consequences of fluorophore attachment”. In: *Journal of the American Chemical Society* 121.18 (May 1999), pp. 4325–4333.
- [167] Marcos E Milla and Robert T Sauer. “P22 Arc Repressor: Folding Kinetics of a Single-Domain, Dimeric Protein”. In: *Biochemistry* 33.5 (Feb. 1994), pp. 1125–1133.
- [168] Alexandre Esadze et al. “Positive and negative impacts of nonspecific sites during target location by a sequence-specific DNA-binding protein: Origin of the optimal search at physiological ionic strength”. In: *Nucleic Acids Research* 42.11 (June 2014), pp. 7039–7046.
- [169] Ammar Tareen and Justin B. Kinney. “Logomaker: Beautiful sequence logos in Python”. In: *Bioinformatics* 36.7 (2020), pp. 2272–2274.
- [170] Yurii Chinenov, Michael Henzl, and Mark E Martin. “The α and β subunits of the GA-binding protein form a stable heterodimer in solution. Revised model of heterotetrameric complex assembly”. In: *Journal of Biological Chemistry* 275.11 (Mar. 2000), pp. 7749–7756.

- [171] Steven J Metallo and Alanna Schepartz. “Certain bZIP peptides bind DMA sequentially as monomers and dimerize on the DMA”. In: *Nature Structural Biology* 4.2 (Feb. 1997), pp. 115–117.
- [172] Hans Wendt, Richard M Thomas, and Tom Ellenberger. “DNA-mediated folding and assembly of MyoD-E47 heterodimers”. In: *Journal of Biological Chemistry* 273.10 (Mar. 1998), pp. 5735–5743.
- [173] Cheng Yung-Chi and William H. Prusoff. “Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction”. In: *Biochemical Pharmacology* 22.23 (Dec. 1973), pp. 3099–3108.
- [174] Robert W. Woody. “Circular Dichroism of Peptides”. In: Academic Press, Jan. 1985, pp. 15–114.
- [175] Yu-Zhu Zhang. “Protein and peptide structure and interactions studied by hydrogen exchanger and NMR”. PhD thesis. University of Pennsylvania, 1995, pp. 1–203.
- [176] Victor Muñoz and Luis Serrano. “Elucidating the folding problem of helical peptides using empirical parameters”. In: *Nature Structural Biology* 1.6 (June 1994), pp. 399–409.
- [177] F. Keith Junius et al. “High resolution NMR solution structure of the leucine zipper domain of the c-Jun homodimer”. In: *Journal of Biological Chemistry* 271.23 (1996), pp. 13663–13667.
- [178] Romain Gautier et al. “HELIQUEST: A web server to screen sequences with specific α -helical properties”. In: *Bioinformatics* 24.18 (Sept. 2008), pp. 2101–2102.
- [179] Kamil Tamiola, Burçin Acar, and Frans A.A. Mulder. “Sequence-specific random coil chemical shifts of intrinsically disordered proteins”. In: *Journal of the American Chemical Society* 132.51 (Dec. 2010), pp. 18000–18003.
- [180] M a Weiss. “Thermal unfolding studies of a leucine zipper domain and its specific DNA complex: implications for scissor’s grip recognition.” In: *Biochemistry* 29.35 (1990), pp. 8020–4.
- [181] Lisa M. Gloss and C. Robert Matthews. “Urea and thermal equilibrium denaturation studies on the dimerization domain of Escherichia coli Trp repressor”. In: *Biochemistry* 36.19 (1997), pp. 5612–5623.
- [182] Hai Zhu et al. “An engineered leucine zipper a position mutant with an unusual three-state unfolding pathway”. In: *Protein Science* 10.1 (Jan. 2001), pp. 24–33.
- [183] Liam B. Moran et al. “Transition state heterogeneity in GCN4 coiled coil folding studied by using multisite mutations and crosslinking”. In: *Proceedings of the National Academy of Sciences of the United States of America* 96.19 (Sept. 1999), pp. 10699–10704.
- [184] Tobin R. Sosnick et al. “The role of helix formation in the folding of a fully α -helical coiled coil”. In: *Proteins: Structure, Function and Genetics* 24.4 (Apr. 1996), pp. 427–432.

- [185] Jason R Schnell et al. “Rapid and accurate structure determination of coiled-coil domains using NMR dipolar couplings: Application to cGMP-dependent protein kinase I α ”. In: *Protein Science* 14.9 (Sept. 2005), pp. 2421–2428.
- [186] Rahul K. Das, Scott L. Crick, and Rohit V. Pappu. “N-terminal segments modulate the alpha-helical propensities of the intrinsically disordered basic regions of bZIP proteins”. In: *Journal of Molecular Biology* 416.2 (Feb. 2012), pp. 287–299.
- [187] Yang Shen and Ad Bax. “Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks”. In: *Journal of Biomolecular NMR* 56.3 (July 2013), pp. 227–241. arXiv: [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- [188] Philip J. Robinson and Neil J. Bulleid. “Mechanisms of Disulfide Bond Formation in Nascent Polypeptides Entering the Secretory Pathway”. In: 9.9 (Aug. 2020).
- [189] P. H. von Hippel et al. “Non-specific DNA binding of genome regulating proteins as a biological control mechanism: I. The lac operon: equilibrium aspects.” In: *Proceedings of the National Academy of Sciences of the United States of America* 71.12 (1974), pp. 4808–4812.
- [190] Jennifer J. Kohler and Alanna Schepartz. “Effects of nucleic acids and polyanions on dimer formation and DNA binding by bZIP and bHLHZip transcription factors”. In: *Bioorganic and Medicinal Chemistry* 9.9 (Sept. 2001), pp. 2435–2443.
- [191] Tom J Petty et al. “An induced fit mechanism regulates p53 DNA binding kinetics to confer sequence specificity”. In: *EMBO Journal* 30.11 (2011), pp. 2167–2176.
- [192] Karine Loth et al. “Sliding and target location of DNA-binding proteins: An NMR view of the lac repressor system”. In: *Journal of Biomolecular NMR* 56.1 (May 2013), pp. 41–49.
- [193] Junji Iwahara, Markus Zweckstetter, and G. Marius Clore. “NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA”. In: *Proceedings of the National Academy of Sciences of the United States of America* 103.41 (Oct. 2006), pp. 15062–15067.
- [194] Abhishek Iyer et al. “Stapling monomeric GCN4 peptides allows for DNA binding and enhanced cellular uptake”. In: *Organic and Biomolecular Chemistry* 13.13 (Mar. 2015), pp. 3856–3862.
- [195] Sarah J. Smith et al. “Tunable helicity, stability and DNA-binding properties of short peptides with hybrid metal coordination motifs”. In: *Chemical Science* 7.8 (July 2016), pp. 5453–5461.
- [196] Stephen E Halford and John F Marko. “How do site-specific DNA-binding proteins find their targets?” In: *Nucleic Acids Research* 32.10 (2004), pp. 3040–3052.
- [197] Otto G Berg and Clas Blomberg. “Association kinetics with coupled diffusional flows. Special application to the lac repressor-operator system”. In: *Biophysical Chemistry* 4.4 (July 1976), pp. 367–381.
- [198] Davide Normanno et al. “Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher”. In: *Nature Communications* 6.1 (Nov. 2015), p. 7357.

- [199] Alex Veksler and Anatoly B. Kolomeisky. “Speed-selectivity paradox in the protein search for targets on DNA: Is it real or not?” In: *Journal of Physical Chemistry B* 117.42 (2013), pp. 12695–12701.
- [200] Leonid Mirny et al. “How a protein searches for its site on DNA: The mechanism of facilitated diffusion”. In: *Journal of Physics A: Mathematical and Theoretical* 42.43 (2009), pp. 1–23.
- [201] Anastasia M. Khoury et al. “Lac repressor-operator interaction: DNA length dependence”. In: *BBA - Gene Structure and Expression* 1087.1 (Sept. 1990), pp. 55–60.
- [202] Reimund Fickert and Benno Müller-Hill. “How Lac repressor finds lac operator in vitro”. In: *Journal of Molecular Biology* 226.1 (1992), pp. 59–68.
- [203] Noriyuki Sugo et al. “Single-Molecule Imaging Reveals Dynamics of CREB Transcription Factor Bound to Its Target Sequence”. In: *Scientific Reports* 5. April (2015), p. 10662.
- [204] Tatsuya Morisaki et al. “Rapid transcriptional activity in vivo and slow DNA binding in vitro by an artificial multi-zinc finger protein”. In: *Biochemistry* 47.38 (Sept. 2008), pp. 10171–10177.
- [205] Paul C Blainey et al. “Nonspecifically bound proteins spin while diffusing along DNA”. In: *Nature Structural and Molecular Biology* 16.12 (Dec. 2009), pp. 1224–1229.
- [206] Katsue Sunahori, Yuang-Taung Juang, and George C. Tsokos. “Methylation Status of CpG Islands Flanking a cAMP Response Element Motif on the Protein Phosphatase 2A α Promoter Determines CREB Binding and Activity”. In: *The Journal of Immunology* 182.3 (2009), pp. 1500–1508.
- [207] Catherine Bui et al. “cAMP response element-binding (CREB) recruitment following a specific CpG demethylation leads to the elevated expression of the matrix metalloproteinase 13 in human articular chondrocytes and osteoarthritis”. In: *The FASEB Journal* 26.7 (2012), pp. 3000–3011.
- [208] Yuichi Ichikawa et al. “Telomeric repeats act as nucleosome-disfavouring sequences in vivo”. In: *Nucleic Acids Research* 42.3 (Feb. 2014), pp. 1541–1552.
- [209] Yuichi Ichikawa et al. “Sequence-directed nucleosome-depletion is sufficient to activate transcription from a yeast core promoter in vivo”. In: *Biochemical and Biophysical Research Communications* 476.2 (2016), pp. 57–62.
- [210] Lu Bai et al. “Nucleosome-Depleted Regions in Cell-Cycle-Regulated Promoters Ensure Reliable Gene Expression in Every Cell Cycle”. In: *Developmental Cell* 18.4 (2010), pp. 544–555.
- [211] Tao Hu, A Yu Grosberg, and B I Shklovskii. “How proteins search for their specific sites on DNA: The role of DNA conformation”. In: *Biophysical Journal* 90.8 (2006), pp. 2731–2744.
- [212] Qianyi Luo et al. “Mechanism of CREB recognition and coactivation by the CREB-regulated transcriptional coactivator CRTC2”. In: *Proceedings of the National Academy of Sciences of the United States of America* 109.51 (2012), pp. 20865–20870.

- [213] Jean-Rene Cardinaux et al. “Recruitment of CREB Binding Protein Is Sufficient for CREB-Mediated Gene Activation”. In: *Molecular and Cellular Biology* 20.5 (Mar. 2000), pp. 1546–1552.
- [214] Sagie Brodsky et al. “Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity”. In: *Molecular Cell* 79.3 (Aug. 2020), 459–471.e4.
- [215] Karola Gerecht, Angelo Miguel Figueiredo, and D. Flemming Hansen. “Determining rotational dynamics of the guanidino group of arginine side chains in proteins by carbon-detected NMR”. In: *Chemical Communications* 53.72 (2017), pp. 10062–10065.