

DEPARTMENT OF BIOCHEMISTRY
UNIVERSITY OF OXFORD



Pathways of target search for bZIP proteins

MSc by Research in Biochemistry

Iris Elpida Kalenderoglou

Linacre College

Supervisors; Dr. Sarah Shammas, Prof. Christina Redfield

Department of Biochemistry

University of Oxford. OX1 3QU

Abstract

The General Control Non-depressible protein 4 (GCN4) is a transcription factor found in *Saccharomyces cerevisiae* and is involved in the regulation of amino acid biosynthesis in response to stress response. GCN4 is member of the bZIP transcription factors superfamily and is an intrinsically disordered protein that recognises and binds to specific DNA sites through its Basic Region (BR), located in the basic leucine zipper domain (bZIP). GCN4 is found as a monomer in solution and is highly disordered but becomes fully helical upon dimerization on its target site. To this day, GCN4's target recognition and search pathway, via a monomer or a dimer, has not been established. Previous work has suggested that GCN4 prefers a monomer pathway, however, recent experimental results for one of its mammalian homologs, the CREB bZIP, pointed towards a dimer pathway. In this work, fluorescence-based equilibrium and kinetics methods were employed to investigate the DNA-binding kinetics of three different GCN4 constructs; the monomer BR-only peptide, the wild-type bZIP and the forced-dimerized bZIP, in complex with two of its target sites AP-1 and CRE, and a non-target DNA (nsDNA). Interestingly, the monomer bound relatively weakly to the DNA sites (K_d range ~30-150 μM) and possesses only a 5-fold specificity for the target. In contrast, full-length bZIP constructs bind tightly to target and non-target sites and are more specific (~2 nM and ~110 nM, respectively). The dissociation rate constants between the bZIP and the forced-dimer were similar (10 s^{-1} for CRE and 2 s^{-1} for AP-1) meaning that the protein appears to leave target sites as a dimer. Association rate constants were the same within error for the two versions, and high, indicating rate enhancement by electrostatic interactions. Further work is required to unambiguously identify the preferred binding pathway. In a further investigation the dynamics of the wild-type GCN4 bZIP in its free or DNA-bound state (with CRE or a non-target DNA) were explored through NMR spectroscopy. In all cases, the Secondary Structure Propensity (SSP) analysis of GCN4 backbone chemical shifts in combination with heteronuclear NOE results demonstrated a rigid coiled-coil leucine zipper. The basic region showed similar helicity with the leucine zipper when bound to the CRE site and was less structured and more dynamic in the apo- and the nsDNA-bound states. Therefore, although the full-length bZIP constructs have relatively high affinity for nsDNA (K_d in the nM range), GCN4 bZIP in complex with nsDNA showed a similar dynamic behaviour to its apo-state and significantly less α-helix content than in the creDNA-bound state. In the intrinsically disordered protein field this would be described as a fuzzy complex.

Acknowledgements

Studying at Oxford University has been an invaluable experience for me, and this project would not have been possible without the support and guidance that I received from many people.

Firstly, I would like to acknowledge Dr Sarah Shammas for giving me the chance to carry out my research project in her group and allowing me to gain important experience in the field of protein biochemistry. I am also grateful that she gave me the freedom to develop my own ideas and for supporting me in their realisation.

My sincere thanks go to Prof. Christina Redfield, who taught me the value of research and offered me the opportunity to work in a high-quality research environment. I also appreciate the training she offered in the NMR technique, and the help and valuable advice provided during my studies.

I would also like to thank Conor Kelly and Dr Mikhail Kuravsky for their training in laboratory methods and their support and advice throughout my thesis.

Many thanks go to Prof. Antonios Kolocouris who has always supported my dreams to become a researcher and his invaluable advice and help during my research journey.

My gratitude extends to Oxford University, Department of Biochemistry for helping me to develop my background skills in Biochemistry, Biophysics and Structural Biology and to Linacre College for providing a supporting environment to me.

A big thank you to my two wonderful sisters, Nikolette and Jacqueline for their unwavering support and for always believing in me. They have been always there for me, as a voice of reason I've needed to hear on several occasions (and undoubtedly will in the future).

Special thanks go to Tonia, my long-lasting friend, thank you for being there even during the lock down period, in spite of the distance.

I would also like to say a heartfelt thanks to Dimitris, who has been by my side throughout my studies, living every single minute of it, and without whom, I would not have had the courage to embark on this journey in the first place.

Finally, I reserve my deepest expression of gratitude to my mum and dad, who always encouraged me to pursue my dreams and never doubted my abilities.

It is for my family that I have completed this thesis.

Abbreviations

Amp	Ampicillin
AP-1	Activator-Protein-1
BR	Basic Region
bZIP	Basic Region Leucine Zipper
CD	Circular Dichroism
Chl	Chloramphenicol
CRE	cAMP Response Element
CREB	cAMP Response Element Binding protein
DNA	Deoxyribonucleic Acid
GCN4	General Control Nonderepressible protein 4
hetNOE	Heteronuclear Nuclear Overhauser Effect
IDP	Intrinsically Disordered Protein
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K_d	Dissociation constant
k_{off}	Dissociation rate
k_{on}	Association rate
LZ	Leucine Zipper
NMR	Nuclear Magnetic Resonance
nsDNA	non-specific DNA
OD	Optical Density
ON	overnight
^{R}	Resistance
rpm	revolutions per minute
SD	Standard Deviation
SDM	Site-Directed Mutagenesis
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SSP	Secondary Structure Propensity

Temp	Temperature
TF	Transcription Factor
wt	wild type

Contents

Abstract 1

Acknowledgements 2

Abbreviations 4

Contents 6

List of Figures 10

List of Tables 13

Chapter 1. Introduction 14

1.1 The Intrinsically Disordered Proteins 14

1.2 Transcription regulation 16

1.3 The class of bZIP Transcription Factors 17

1.4 Target search by specific transcription factors 18

1.5 Fluorescence anisotropy 19

1.6 GCN4; An intrinsically disordered transcription factor 19

Aim of this work 23

Chapter 2. Expression and purification of GCN4 constructs 24

2.1 Introduction 24

2.2 Material and Methods 25

Microbiology 25

2.2.1 Plasmids, bacterial strains and cell culture conditions 25

Molecular Biology 27

2.2.2 Gibson Assembly 27

2.2.3 Polymerase Chain Reaction (PCR) 28

2.2.4 Agarose gel nucleic acid electrophoresis 30

2.2.5 Production of $CaCl_2$ competent cells 30

2.2.6 Transformation of competent cells 31

Biochemical Methods 31

2.2.7 Polyacrylamide gel electrophoresis (PAGE) 31

2.2.8 Measurement of protein and DNA concentrations 32
2.2.9 Expression of GCN4 protein constructs 33
2.2.10 Purification of GCN4 protein constructs 33
2.2.11 Purification of His6-GB1-GCN4 proteins 34
2.2.12 Purification of His6-GB1-cleaved GCN4 proteins 34
2.2.13 Protein sample preparation 35
AGADIR analysis: Choice of GCN4 bZIP protein sequence for biophysical experiments 35
2.3 Results 37
2.3.1 Design of the GCN4 constructs 37
2.3.2 Cloning, expression and purification of GCN4 constructs 40
2.4 Discussion 49
2.4.1 Developing suitable models for monomeric and dimeric GCN4 49
2.4.2 Cloning, expression and purification of GCN4 protein constructs 50
2.5 Conclusion 51
Chapter 3. Biophysical characterisation of GCN4 binding to DNA 52
3.1 Introduction 52
3.2 Materials and Methods 55
3.2.1 DNA sample preparation 55
3.2.2 Fluorescence Anisotropy equilibrium binding experiments 55
3.2.3 Association and dissociation kinetics by Stopped-flow 56
3.2.4 Binding interaction equations 57
3.3 Results 60
3.3.1 Equilibrium binding constant (K_d) of GCN4 bZIP and FD 60
3.3.2 Binding association kinetics of GCN4 with its targets 68
3.3.3 Dissociation of GCN4 from its target (k_{off}) 71
3.3.4 Kinetics of binding reaction for GCN4 with linear CRE DNA 74
3.3.5 DNA binding by monomeric GCN4 basic region 78

3.4 Discussion 81
3.4.1 Wild-type bZIP and forced-dimer GCN4 bind specifically to target DNA with nM affinity 81
3.4.2 Monomeric GCN4 binds less tightly to target DNA than dimeric GCN4 82
3.4.3 GCN4 binding rate constants are similar for dimeric and dimerization-competent versions 84
3.4.4 Binding of dimeric GCN4 to AP-1 and CRE are thermodynamically very similar 85
3.4.5 Choice of DNA constructs 86
3.5 Conclusions and further work 88
Chapter 4. Structural characterisation of the wild-type GCN4 bZIP dimer 90
4.1 Introduction 90
4.2 Material and Methods 92
4.2.1 ^{15}N and ^{13}C-labelled protein expression and expression trials 92
4.2.2 Purification of labelled GCN4-bZIP 94
4.2.3 Nuclear magnetic resonance experiments 94
4.2.4 Data processing and analysis 94
4.3 Results 97
4.3.1 ^{15}N and ^{13}C-labelled protein expression trials 97
4.3.2 NMR experiments of the apo-GCN4 bZIP dimer 98
4.3.3 NMR experiments of the GCN4 bZIP dimer in complex with non-specific DNA (nsDNA) and target CRE DNA (creDNA) sites 106
4.3.4 Comparative NMR data analysis of the three systems; apo-GCN4 bZIP, GCN4 bZIP-nsDNA and GCN4 bZIP-creDNA 119
4.4 Discussion 123
4.4.1 Expression and purification of ^{15}N and ^{13}C-labelled GCN4 bZIP 123
4.4.2 GCN4-creDNA is the most structured complex followed by the apo-state and the GCN4-nsDNA complex 123
4.4.3 nsDNA and creDNA show different exchange regime 126

4.4.4 How does CREB bZIP compare to GCN4 bZIP?.....126
4.5 Conclusions and further work.....130
Appendices.....133
References.....141

List of Figures

Figure 1.1. Sumo-1 soluble protein; an example of protein structure with regions of various secondary structure. 14
Figure 1.2. An illustration of a dimerized bZIP bound to DNA forming a fully structured complex. 17
Figure 1.3. The full-length GCN4 protein sequence. 20
Figure 1.4. GCN4 structure in complex with dsDNA (PDB ID: 1YSA)....22
Figure 2.1. Residual prediction of helical propensity as a function of GCN4 full length sequence analysed by the AGADIR algorithm. 36
Figure 2.2. Helical content calculation using the AGADIR algorithm. 38
Figure 2.3. Illustration of the three designed protein constructs. 39
Figure 2.4. Agarose nucleic acid electrophoresis gels validating DNA fragments. 41
Figure 2.5. Representative 4 to 12% Bis-Tris protein gel of GCN4 FD early stages of purification process. 43
Figure 2.6. Representative 4 to 12% Bis-Tris protein gels from GCN4 FD IMAC purification. 44
Figure 2.7. Representative chromatography traces of GCN4 FD IMAC purification. 45
Figure 2.8. Representative GCN4 bZIP IMAC purification SDS-PAGE gels and chromatography traces. 46
Figure 2.9. Representative 4 to 12% Bis-Tris protein gels from GCN4 BR IMAC purification. 47
Figure 2.10. Representative chromatography traces of GCN4 BR IMAC purification 48
Figure 3.1. A model example of the association and dissociation reactions between a bZIP and its target DNA or its non-target DNA (competitor). 53
Figure 3.2. Pilot experiments: Fluorescence anisotropy (a) and intensity (b) equilibrium data of GCN4 bZIP binding to linCRE and NS. 61
Figure 3.3. Pilot experiments: Fluorescence anisotropy (a)(c) and intensity (b)(d) equilibrium data of GCN4 FD (a)(b) or bZIP (c)(d) binding to linCRE. 62
Figure 3.4. Illustration of the linear and hairpin version of CRE site....63
Figure 3.5. Fluorescence anisotropy equilibrium data of GCN4 bZIP and FD binding. 64
Figure 3.6. Fluorescence intensity equilibrium data of GCN4 bZIP and FD binding. 65
Figure 3.7. Dissociation constants determined for the full-length bZIP constructs from various methods. 66

Figure 3.8. Observed kinetic rate constants of GCN4 bZIP or FD binding to hCRE or hAP-1. 69
Figure 3.9. Representative traces for association experiments using fluorescence stopped-flow. 70
Figure 3.10. Association rate constants determined from fluorescence-based kinetics. ...71
Figure 3.11. Representative traces for dissociation experiments using fluorescence intensity stopped-flow. 72
Figure 3.12. k_{off} determination: Fluorescence intensity kinetic data for the dissociation of the complexes GCN4 bZIP or FD binding to hCRE or hAP-1 using the stopped-flow machine. 73
Figure 3.13. Dissociation rate constants determined from fluorescence-based kinetics. ..74
Figure 3.14. Fluorescence anisotropy kinetic data of GCN4 FD binding using stopped-flow. 75
Figure 3.15. Representative traces for association experiments for GCN4 FD binding to linCRE using fluorescence anisotropy stopped-flow....76
Figure 3.16. Representative traces for dissociation experiments for GCN4 FD binding to linCRE using fluorescence anisotropy stopped-flow....77
Figure 3.17. Pilot experiments: Fluorescence anisotropy (a) and intensity (b) equilibrium data of GCN4 BR monomer binding to linCRE and NS. 78
Figure 3.18. Fluorescence anisotropy equilibrium data of GCN4 BR monomer binding. 79
Figure 3.19. Equilibrium dissociation constants determined for the monomer BR. 80
Figure 3.20. Equilibrium dissociation constants determined for GCN4 binding to NS DNA. 80
Figure 4.1. Expression trials for production of ^{15}N and ^{13}C-labelled GCN4-bZIP.....97
Figure 4.2. (a) Assigned ^{1}H-^{15}N HSQC spectrum of GCN4 bZIP (at pH 6.5 and 25 °C). (b) The crowded region around Q248 is shown enlarged. 100
Figure 4.3. Triple resonance experiments to assign GCN4 bZIP residues. 101
Figure 4.4. HBHA(CO)NH triple resonance spectra showing Hα and Hβ hydrogens of consecutive residues. 102
Figure 4.5. HNCO triple resonance spectra showing backbone carbonyl carbons of consecutive residues. 102
Figure 4.6. Secondary structure propensity and dynamics for apo-GCN4 bZIP. 104
Figure 4.7. 950 MHz ^{1}H-^{15}N BTROSY spectra of the GCN4 bZIP in the presence of varying concentrations of nsDNA....107

Figure 4.8. 950 MHz ^{1}H-^{15}N BTROSY spectra of the GCN4 bZIP in the presence of varying concentrations of creDNA 108
Figure 4.9. 950 MHz ^{1}H-^{15}N BTROSY spectra of the GCN4 bZIP in complex with nsDNA collected over a range of temperatures: 25 °C (red), 30 °C (yellow) and 35 °C (green).110
Figure 4.10. 950 MHz ^{1}H-^{15}N BTROSY spectra of the GCN4 bZIP in complex 375 μM creDNA collected over a range of temperatures: 25 °C (red), 30 °C (yellow) and 35 °C (blue).....111
Figure 4.11. (a) Assigned ^{1}H-^{15}N BTROSY spectra of nsDNA-bound GCN4 bZIP (at pH 6.5 and 35 °C). (b) The crowded region around R232 is shown enlarged. 113
Figure 4.12. (a) Assigned ^{1}H-^{15}N BTROSY spectra of creDNA-bound GCN4 bZIP (at pH 6.5 and 35 °C). (b) The crowded region around Y265 is shown enlarged. 114
Figure 4.13. Secondary structure propensity and dynamics for GCN4-nsDNA. 116
Figure 4.14. Secondary structure propensity and dynamics for GCN4-creDNA. 117
Figure 4.15. Secondary structure propensity (SSP) of GCN4 bZIP. 119
Figure 4.16. {^{1}H}-^{15}N Heteronuclear NOE experiment for the apo-state (orange), nsDNA-GCN4 (black) and creDNA-GCN4 (blue). 120
Figure 4.17.Combined ^{1}H - ^{15}N weighted shift differences between the apo-GCN4 and GCN4 DNA systems.....121
Figure 4.18. GCN4 backbone Cα atoms shift change upon addition of nsDNA (black) or creDNA (blue). 122
Figure 4.19. Secondary structure propensity (SSP) of CREB bZIP. 127
Figure 4.20. Secondary structure propensity (SSP) of the GCN4 and CREB bZIPs in the apo-state or in complex with nsDNA or creDNA.....128

List of Tables

Table 2.1. Amino acid sequence, molecular weight (MW) and extinction coefficient (EC) at 280 nm for monomer protein constructs. 25
Table 2.2. Bacterial strains and plasmids used. 26
Table 2.3. Antibiotics and chemical compounds used for cell cultures in molecular biology and protein expression. 27
Table 2.4. Plasmid DNA templates and primers used for the PCR reactions 29
Table 2.5. PCR programme for PCR reactions using the Q5® High-Fidelity 29
Table 2.6. DNA sequence and extinction coefficient (EC) at 260 nm for unlabelled DNA and 496 nm for AF488-labelled DNA. 32
Table 3.1. Dissociation constants (K_d) for target (hCRE and hAP-1) and non-target DNA (NS) in complex with GCN4 bZIP, FD or BR from fluorescence anisotropy experiments. 66
Table 3.2. Dissociation constants (K_{d2}) for target (hCRE and hAP-1) and non-target DNA (NS) in complex with GCN4 bZIP, FD or BR. 67
Table 3.3. Dissociation constants (K_d) for target (hCRE and hAP-1) in complex with GCN4 bZIP or FD from fluorescence intensity equilibrium experiments. 67
Table 3.4. Association and dissociation rate constants for the GCN4 bZIP and forced dimer obtained by fluorescence anisotropy or intensity measurements 74
Table 3.5. Association and dissociation rate constants for the GCN4 FD-linCRE complex obtained by fluorescence anisotropy measurements. 77
Table 4.1. Protein and DNA sequences used in the NMR experiments. 92
Table 4.2. Acquisition parameters for the triple resonance NMR experiments used in backbone assignment. 95
Table 4.3. Acquisition parameters for the 2D ^{1}H-^{15}N NMR experiments. 96
Table 4.4. Detailed interaction information of GCN4 bZIP binding to CRE site. 125
Table 4.5. Saccharomyces cerevisiae GCN4 bZIP and Homo sapiens CREB bZIP sequences used in NMR studies 127

Chapter 1. Introduction

1.1 Intrinsically Disordered Proteins

Intrinsically Disordered Proteins or Regions (IDPs/IDRs) are proteins that do not have a well-defined tertiary structure, or persistent secondary structure. They are involved in several diseases. For instance, Alzheimer's and Parkinson's diseases are associated with protein misfolding which leads to amyloid aggregation (Araki et al., 2019; Ashraf et al., 2014). However, there are also many biological processes in which IDPs play a vital role. More interestingly, it is estimated that one third of eukaryotic proteins possess a disordered region (Ward et al., 2004) and at least 10% of those are fully disordered (Obradovic et al., 2003). Furthermore, it has been shown that there are IDPs of a semi-disordered state with 50% of possibility switching to structured conformation or a disordered state (Zhang et al., 2013). One example of a semi-disordered protein is SUMO-1 shown in Figure 1.1.

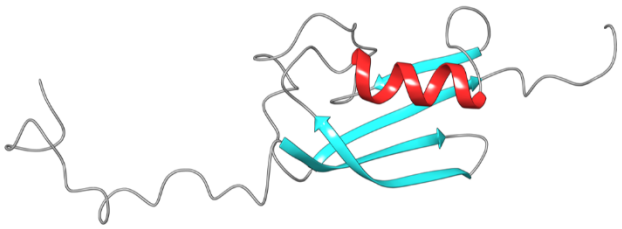


Figure 1.1. Sumo-1 soluble protein; an example of protein structure with regions of various secondary structure.
The PDB ID 15AR model 1 was prepared in Maestro protein preparation wizard. α-helix is shown in red; β-sheet is shown in cyan and unstructured regions are shown in light grey.

These findings suggest protein disorder contributes to biological procedures in a unique way that structured proteins do not. Having a large ensemble of available 3D conformations may be advantageous to IDPs as it gives them the capability of interacting with many substrates – either by adopting different conformations as they bind, or by

allowing different regions to be exposed to the solvent. Indeed, disordered proteins are significantly over-represented in signalling proteins and are found as hubs in signalling networks (Haynes et al., 2006). They may have fluctuating levels of residual secondary structure. IDPs have a broad conformational space and flat energy landscapes (Chong & Ham, 2019). Some families of disordered proteins have the ability to fold into a structured state upon binding to their macromolecular partners. Such is the case of many transcription factors (TFs), where the final protein-DNA complex is ordered and fully functional (i.e. bZIP family of TF).

Level of intrinsic disorder determined by amino-acid sequence

Intrinsic protein disorder, and more accurately, the ability of proteins to fold or not to fold is encoded it their amino-acid sequence (Dunker et al., 2001). In general, IDPs are characterised by the combination of a low content of hydrophobicity (bulky hydrophobic amino acids) and a high net charge (polar and charged amino acids) (Uversky et al., 2000). This conclusion was derived from the comparative analysis of sequences of natively folded and natively unfolded proteins. The aforementioned low mean hydrophobicity leads to a low driving force for globular protein folding, which typically involves formation of a hydrophobic core protected from solvent. Instead, the charged and polar residues lead to good interactions with water and high solvation. These features are common between the IDPs and the intrinsically disordered regions of natively folded proteins. More specifically, after analysis of the amino acid composition and sequence of IDPs it has been suggested that the disorder-promoting residues are Ala, Arg, Gly, Gln, Ser, Glu, Lys, and Pro whereas the least frequent amino acids within a disordered region are Ile, Leu, Val, Trp, Tyr, Phe, Cys, and Asn (Dunker et al., 2001; Radivojac et al., 2007; Romero et al., 2001; Williams et al., 2001). The remaining amino acids (His, Met, Thr and Asp) remain ambivalent since they can be found in both structured and disordered regions in equal measures.

The next step from separating the disorder-defining residues was to specify an order-disorder ranking of the amino acids. An analysis by Vacic et al. (2007), ranked amino acids by their propensity to form disordered regions: Trp, Phe, Tyr, Ile, Met, Leu, Val, Asn, Cys, Thr, Ala, Gly, Arg, Asp, His, Gln, Lys, Ser, Glu, Pro, where order promoting to disorder promoting residues are shown (Vacic et al., 2007). Such studies, have given rise to computational tools and more specifically algorithms that are able to calculate a protein's

disorder profile, depending on the composition and sequence of the order-disorder ranked amino acids, at residual level.

1.2 Transcription regulation

Organisms can adapt to different environmental conditions by producing responses to new physiological triggers by their immediate environment. These responses are time and dose-regulated and include the production of new chemical compounds and proteins/enzymes as a reply to the environmental changes. In cell molecular biology and genetics, transcriptional regulation is the means and process by which DNA information is transferred and converted into RNA. Thus, organisms can modulate DNA transcription to further control protein expression that is crucial for survival.

Physiological growth, development and adaptation through cellular responses triggered by environmental stress, are tightly controlled by transcriptional networks. Their role is to regulate gene expression by increasing or repressing the amount of expression of proteins in response to environmental triggers. These networks are highly complex, contain many regulatory biomolecules and encode the essential information for development and survival. Transcription factors (TFs) play a vital role in these regulatory networks. These proteins control the expression of genes, by turning them on and off, making sure the concentration, timing and location of essential proteins are right. Once they find their DNA target, they bind and attract further transcription machinery such as the DNA polymerase to commence the transcription process. Transcription factors are present both in prokaryotes and eukaryotes, but eukaryotic TFs are on average twice as long as their prokaryotic counterparts and contain extended disordered regions. More specifically, at least 82% of eukaryotic TFs have intrinsically disordered regions (Liu et al., 2006) whereas up to 50% of the entire human TF sequence is intrinsically disordered (Minezaki et al., 2006). Protein sequences and their consensus regulatory elements' sequences demonstrate a high level of conservation within species (Iwama & Gojobori, 2004).

1.3 The class of bZIP Transcription Factors

The family of the bZIP (basic-region leucine zipper) transcription factors is exclusively present in eukaryotes and in a limited number of viruses e.g. the Zta bZIP found in the Epstein-Barr virus (Chang et al., 1990). Interestingly, bZIP TF proteins are the largest and most conserved family of TFs found in eukaryotes (Amoutzias et al., 2007). The bZIP family of TFs are specific TFs, meaning that they bind to their cognate DNA sites to promote transcription. They are involved in the transcription of genes by forming homo- or hetero-dimers (Newman & Keating, 2003). The transcriptional networks formed by the bZIP TFs play vital roles in many cellular processes such as proliferation, maintenance and differentiation (Miller, 2009). Likewise, dysfunctional bZIP TFs can result in pathophysiological conditions, such as cancer and autoimmune diseases (Davudian et al., 2016; Igarashi et al., 2017; Kaser et al., 2008; Zhao et al., 2013).

The proteins of this family have at least one bZIP domain in their sequence. The name bZIP comes from the two regions, the basic region and the leucine zipper, which constitute the bZIP domain. The N-terminal basic region is composed of highly positively charged amino residues, and the following C-terminal leucine zipper is characterised by leucine residues that repeat along the helical turn, forming an amphipathic helix (Schumacher et al., 2000). The whole bZIP structure - when folded – forms an uninterrupted α-helix.

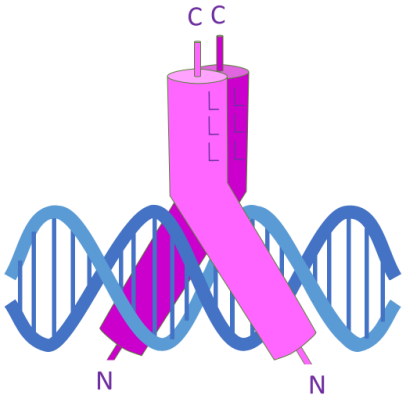


Figure 1.2. An illustration of a dimerized bZIP bound to DNA forming a fully structured complex.
The bZIP is shown in pink and the DNA in blue. L=leucine, N=N-terminus, C=C terminus.

bZIP domains dimerize with other bZIPs by the formation of a coiled-coil. Many bZIP TFs form homodimers to carry out their role in transcription regulation, however, most heterodimerize with various bZIP TFs leading to different functionalities. Newman and Keating (2003) suggest that only 15% of all possible interactions among different bZIPs take place however, with specificity being determined by the leucine-zipper sequence.

As mentioned previously, the basic region consists of positively charged amino acids which interact with the negatively charged DNA backbone to form a final bZIP-DNA complex. Crystal structures of bZIPs (Ellenberger et al., 1992) with their target-DNA site show that the basic region is fully ordered and makes base-specific contacts with the target DNA. However, when the bZIP is in solution and in the absence of target DNA sites, the DNA-binding domain (BR) is disordered. Depending on their DNA-binding specificities, bZIP proteins have been arranged into several families, each of which recognizes different DNA motifs (Hakoshima et al., 2000; Vinson et al., 2002). The crystal structures of bZIPs bound to their cognate DNA have revealed functional variability of the conserved residues in DNA recognition. For this reason, a universal code connecting the BR sequence with the cognate DNA sequence has not been established.

1.4 Target search by specific transcription factors

The cell nucleus contains an extraordinary concentration of DNA of around 10 mg/ml and for specific transcription factors that recognise specific DNA sequences, the target sites constitute a minute fraction of the cellular DNA. As a result, transcription factors have to first scan the genome so they can find their target sites and then bind there to help turn specific genes “on” or “off”. TFs do not simply undergo 3D diffusion and bind immediately to their targets. Rather they also bind to non-cognate DNA sequences through non-specific and DNA-sequence-independent interactions and then intramolecularly translocate to their target DNA sequence for transcription initiation. One of the suggested mechanisms is that the protein is “sliding”. In this case, the TF remains in contact with the non-specific DNA and “slides” through one-dimensional (1-D) diffusion for a long enough period to find its target DNA site (Berg et al., 1981). The second proposed mechanism is the protein “hopping” where the TF “hops” from one site to another by dissociating from one site and re-associating with a site closer to the target site. In this scheme, there is a three-

dimensional (3D) movement in space (Berg et al., 1981). The previously described schemes involve protein translocating between adjacent sites located on the same DNA molecule. In a third scenario, a TF may dissociate from one DNA site and bind to a different one, far away in the DNA sequence, due to simultaneous protein binding to the two sites (Bellomy & Record, 1990; Vologodskii & Cozzarelli, 1996). However, this mechanism has been proposed for DNA-binding proteins that have at least two binding sites such as the Lac repressor (Solomon, 2019).

1.5 Fluorescence anisotropy

Fluorescence intensity and anisotropy offer sensitive tools for investigating and measuring the binding of fluorescent ligands and are particularly useful probes in the case of kinetic studies. Fluorescence anisotropy gives a measure of how the orientation of the fluorophore changes in space, between the excitation and the emission processes. Attaching a fluorophore to a ligand therefore provides the opportunity for anisotropy to be used as a probe - when a ligand binds to its receptor, the overall formed complex tumbles slower than the ligand when it is free in solution. If the fluorophore is excited with polarized light the emitted light will maintain polarisation depending on how fast the molecule rotates in the solution. If it is rotating slowly, then the emitted light will retain its polarisation while if the molecule is rotating fast, then the emitted light will lose its polarisation.

1.6 GCN4; An intrinsically disordered transcription factor

The General Control Nonderepressible protein 4 (GCN4) is a transcription factor (TF) found in yeast and controls the stress response by regulating amino acid biosynthesis (Natarajan et al., 2001). Under conditions of amino acid starvation (Arndt & Fink, 1986; Hill et al., 1986), GCN4 cellular levels are increased, resulting in the enhanced transcription of the coregulated genes (Arndt & Fink, 1986; Hinnebusch, 1984; Roussou et al., 1988). In addition to the de-repression of genes involved in amino acid biosynthetic pathways, GCN4 is increasingly understood to be involved in cell metabolism. For instance, GCN4 is involved in purine and organelle biosynthesis, autophagy, glycogen homeostasis, and

multiple stress responses (Natarajan et al., 2001), and therefore has a broad role in cell metabolism.

```
        10         20         30         40         50
MSEYQPSLFA LNPMGFSPLD GSKSTNENVS ASTSTAKPMV GQLIFDKFIK
        60         70         80         90        100
TEEDPIIKQD TPSNLDFDFA LPQTATAPDA KTVLPIPELD DAVVESFFSS
       110        120        130        140        150
STDSTPMFEY ENLEDNSKEW TSLFDNDIPV TTDDVSLADK AIESTEEVSL
       160        170        180        190        200
VPSNLEVSTT SFLPTPVLED AKLTQTRKVK KPNSVVKKSH HVGKDDESRL
       210        220        230        240        250
DHLGVVAYNR KQRSIPLSPI VPESSDPAAL KRARNTEAAR RSRARKLQRM
       260        270        280
KQLEDKVEEL LSKNYHLENE VARLKKLVGE R
```

Figure 1.3. The full-length GCN4 protein sequence.
The C-terminal bZIP domain is coloured in red (residues 225-281).

GCN4 has a mammalian homolog, activating transcription factor-4 (ATF4). Like ATF4, GCN4 is recognised to participate in more complex cellular functions regarding longevity and response to cellular stress (Mittal et al., 2017). GCN4 serving both amino acid deprivation by increasing transcription, and environmental stress by reducing protein expression/suppressing translation (Mittal et al., 2017), is counter intuitive at first. However, GCN4 is really a master *regulator* (Natarajan et al., 2001), dependent on upstream environmental messenger concentrations and therefore allows early control of protein biosynthesis even in instances where biosynthesis suppression is required. This functional promiscuity dictates a very modular role for GCN4 and related transcription factors in cell metabolism control. A similar case has been observed for ATF4 protecting against neuronal death in Parkinson's (Sun et al., 2013), while increasing protein synthesis and cell death in endoplasmic reticulum (ER) stress (Han et al., 2013).

As with other bZIP domains GCN4-bZIP is intrinsically disordered but its dimerization module folds into a coiled coil upon homodimerization (LZ) (Saudek et al., 1991). The final GCN4-target-DNA complex consists of a dimerized fully structured assembly. (Bracken et al., 1999; Ellenberger et al., 1992; M. L. Gill et al., 2016; Robustelli et al., 2013; Weiss et al., 1990) GCN4 recognises and binds to AP-1 and CRE DNA sites in the upstream 5'-regulatory regions of these genes (Hope & Struhl, 1986; Roussou et al.,

1988; Sellers et al., 1990). Among other targets, GCN4 binds the *his3* promoter in yeast (Hill et al., 1986) but it has general preference for the 9-base-pair (bp) dyad-symmetric DNA sequence, ATGA(C/G)TCAT (Sellers et al., 1990). This sequence is a general consensus of the GCN4 naturally occurring targets and therefore is not a strict requirement for binding. For example, the respective *his3* recognition sequence is 5'-ATGACTC**T**T-3' (Hill et al., 1986) (variation base shown in bold). GCN4 binds its targets as a dimer and therefore two monomers share the same binding site. However, the central 7-bp sequence proven to be most important for dimer binding, 5'-TGA(C/G)TCAT-3', is asymmetric due to the central G-C pair. This means there is overlap in the key half-site recognition sequence of a GCN4 monomer (Sellers et al., 1990). Given the complementarity in the DNA strands, these observations mean that in order to bind to the consensus sequence, a GCN4 monomer must bind to ATGAG while the other to ATGAC. The latter has been found optimal for monomer binding (Sellers et al., 1990) explaining why variations in ATGAG are better tolerated than the ATGAC half (Oliphant et al., 1989). Additionally, insertions in the recognition sequence do not necessarily abrogate GCN4 binding. For example, the sequence 5'-ATGA**CG**TCAT-3' (CREB recognition sequence) allows for GCN4 binding (Sellers et al., 1990) while providing with the two optimal half-sites, ATGAC. It seems though that a 9-bp as opposed to a 10-bp sequence is preferred (Sellers et al., 1990) despite optimal half-site availability for both GCN4 monomers.

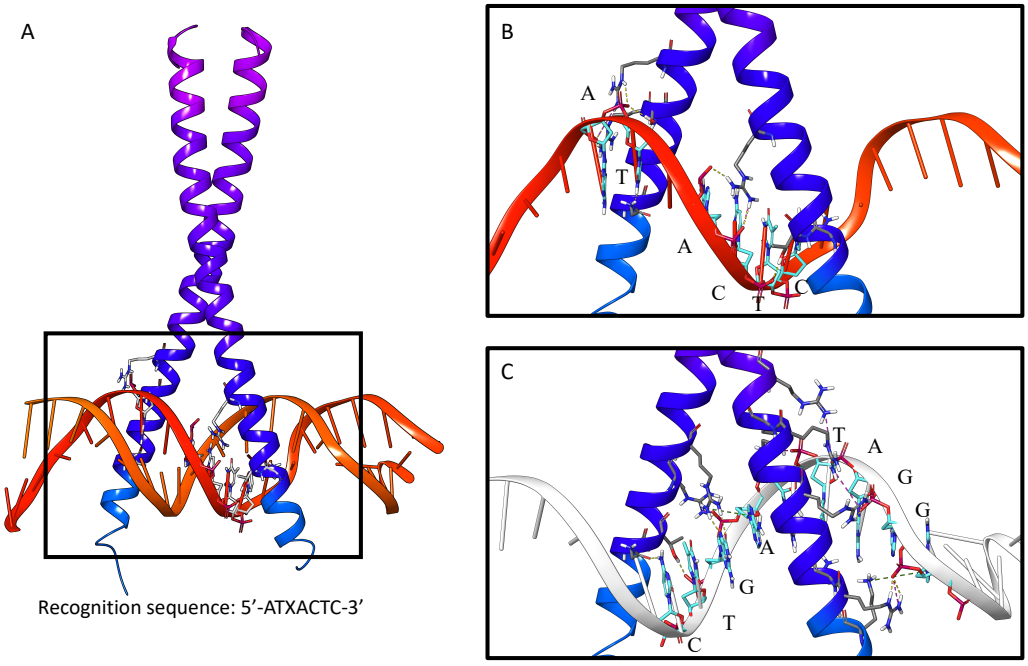


Figure 1.4. GCN4 structure in complex with dsDNA (PDB ID: 1YSA).
Panel (A) displays a cartoon representation of the protein (blue ribbon) – DNA (red helix) binding complex. Only residues interacting with the sense strand are shown along with the respective deoxyribonucleotides, based on the X-ray crystal structure. (B) Close-up of the interactions in (A) with protein-DNA H-bonds as yellow connecting lines. Interacting deoxyribonucleotides are labelled explicitly. (C) Close-up showing the protein interactions with the complementary DNA strand. Interacting deoxyribonucleotides are labelled explicitly. Figure made with Schrodinger's v2020.3 PyMol molecular graphics engine.

Aim of this work

Dr. Sarah Shammas' group focuses on a mammalian CRE-binding bZIP protein; CREB. Unpublished equilibrium binding and kinetic results for experiments that use target DNA in the presence of competing non-specific DNA, indicate a dimeric mode of DNA searching due to monomeric CREB's inability to bind CRE sequences more tightly than non-CRE sequences. However, a monomer search pathway has previously been suggested for the bZIP GCN4. Structurally, NMR data show that the dimeric search state of CREB (bound to non-target DNA) is less helical than the CREB-CRE complex and more helical than the CREB apo-state. My project has two main questions to address:

1. How does GCN4 behave in similar pathway studies to those performed by Shammas lab for CREB bZIP? If there is a difference, from where does it originate?
2. What does GCN4 bZIP dimer look like during target search? Is it partially structured in a similar manner to CREB bZIP dimer?

Chapter 2. Expression and purification of GCN4 constructs

2.1 Introduction

Equilibrium and kinetic binding studies require high-quality preparations of protein constructs. Previous experiments in the lab with the bZIP CREB had identified the importance of creating monomeric, dimeric, and dimerization-competent versions of the bZIP domain for determining the target search pathway.

My first task was therefore to create pure versions of these constructs, free from contamination with other proteins, and critically, nucleic acids.

2.2 Material and Methods

Microbiology

2.2.1 Plasmids, bacterial strains and cell culture conditions

All bacterial strains and plasmids for cloning and protein expression are included in Table 2.2. Protein constructs designed for this project are shown in Table 2.1. For molecular biology methods and for protein expression, liquid cultures were prepared as following; strains were grown at 37 °C in autoclaved 2xTY which contains 16 g/L tryptone, 10 g/L yeast extract and 5 g/L NaCl. In each case, medium was supplemented with the required antibiotics, depending on resistance of bacterial strains and plasmids used (See antibiotics and chemical compounds used in Table 2.3. 2xTY is a nutrient-rich microbial growth medium, suitable for cultivation of *E. coli* strains for cloning and production of recombinant proteins.

Table 2.1. Amino acid sequence, molecular weight (MW) and extinction coefficient (EC) at 280 nm for monomer protein constructs.
The glycine residue highlighted in blue is a legacy amino acid left from TEV cleavage. The tryptophan at the end of the BR sequence is present to assist protein concentration determination. EC were used as determined by Gil and von Hippel method (S. C. Gill & von Hippel, 1989).

GCN4 construct	Amino acid sequence	MW (kDa)	EC_{280} (M^{-1} cm^{-1})
BR G(S225-E259)W	GSDPAALKRARNTEAARRSRARKL QRMKQLEDKVEEW	4.3	5690
bZIP G(S225-R281)	GSDPAALKRARNTEAARRSRARKL QRMKQLEDKVEELLSKNYHLENEV ARLKKLVGER	6.7	1280
Forced Dimer (FD) G(S225-R291)GGC	GSDPAALKRARNTEAARRSRARKL QRMKQLEDKVEELLSKNYHLENEV ARLKKLVGERGGC	6.9	1400

Table 2.2. Bacterial strains and plasmids used.

Strain	Genotype	Source
NEB® 5-alpha Competent E. coli	fhuA2 Δ(argF-lacZ) U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17 No antibiotic resistance	New England Biolabs
BL21 (DE3)	F– ompT hsdSB (rB–, mB–) gal dcm (DE3) Chl^R	ThermoFischer Scientific

Plasmid	Properties	Source
1. pcDNA3	Mammalian expression vector with the CMV promoter. Includes GCN4 partial mRNA. Amp^R.	Addgene
2. Champion™ pET303	Expression vector used for mutagenesis of the subsequent vectors. Amp^R	Invitrogen
3. pRSET-A-His6-GB1-GCN4 wtbZIP	Expression vector for His6-GB1-GCN4 wtbZIP, residues 226-281, (IPTG inducible). Amp^R	Shammas' lab, Invitrogen (plasmid number 1)
4. pET303-His6-GB1-GCN4 wtbZIP	Expression vector for His6-GB1-GCN4 wtbZIP, T7 promoter (IPTG inducible). Used with BL21 E. coli expression strain. Final resistance: $AmpChl^R$	This work
5. pET303-His6-GB1-GCN4 bZIPGGC	Expression vector for His6-GB1-GCN4 bZIPGGC, T7 promoter (IPTG inducible). Used with BL21 E. coli expression strain. Final resistance: $AmpChl^R$	This work
6. pET303-His6-GB1-GCN4 BR	Expression vector for His6-GB1-GCN4 BR, T7 promoter (IPTG inducible). Used with BL21 E. coli expression strain. Final resistance: $AmpChl^R$	This work

Table 2.3. Antibiotics and chemical compounds used for cell cultures in molecular biology and protein expression.

Compound	Stock solution	Working concentration
Ampicillin (Amp)	100 mg/ml	100 μg/ml
Chloramphenicol (Chl)	50 mg/ml	34 μg/ml
IPTG	1 M	1 mM

Molecular Biology

2.2.2 Gibson Assembly

Gibson Assembly (GA) is a molecular cloning method which offers the possibility of joining sequential DNA fragments into one. In order to create the Champion™ pET303-His6-GB1-GCN4 wtbZIP vector, three fragments were needed:

1. pET303
2. His6-GB1
3. GCN4 wtbZIP (residues 225-281)

GA utilises three enzymes for joining fragments with complementary ends: T5 exonuclease, Phusion DNA polymerase and Taq DNA ligase – in a single mix. GA was performed following manufacturer's instructions. Gibson Assembly® Master Mix kit purchased from New England Biolabs was used. DNA templates with sharing terminal overlaps were derived from PCR reactions (see paragraph 2.2.3) using the primers listed in Table 2.4 and were assembled in a 20 μL reaction mix. Following PCR amplification, the PCR products were checked using agarose gel electrophoresis (see protocol in paragraph 2.2.4).

Next, the PCR products were purified using the GeneJET PCR Purification Kit from Thermo Scientific™. The final products were cloned into a final vector using one-step isothermal assembly reaction (Gibson et al., 2009). The molar ratio of pET303:GCN4bZIP:His6-GB1 was 1:2:2. The GA was set on ice where all of the reagents

were added. The samples were incubated in a thermocycle at 50 °C for 15 min and stored in -20 °C until DH5α transformation.

2.2.3 Polymerase Chain Reaction (PCR)

Five total PCR reactions were performed: three for the amplification of the three DNA fragments required for Gibson Assembly (mentioned in paragraph 2.2.2) to produce the His6-GB1-GCN4bZIP construct in the pET303 vector, and two for SDM to create the BR and the FD-containing plasmids by using the GA product as template. For every PCR reaction, ~100 pg of plasmid template DNA was used. A complete list of the templates, primers, melting temperatures and experimental temperatures used are shown in Table 2.4. The experimental conditions for each PCR reaction are demonstrated in Table 2.5. For the calculation of the optimal annealing temperatures for each primer set, the NEB Tm calculator (https://tmcalculator.neb.com) was used which is recommended for use in combination with the Q5® High-Fidelity DNA Polymerase kit. Primers were purchased from Life Technologies Ltd.

PCR products were verified by agarose gel electrophoresis (See paragraph 2.2.4). The final DNA products were linear and thus, for the FD and BR plasmid constructs the KLD (kinase-ligase-DpnI) reaction was performed, prior to DH5α transformation. In this case the KLD Enzyme Mix was used (New England Biolabs). KLD reaction utilizes three enzymes (kinase-ligase-DpnI) for rapid circularisation of the PCR product. The KLD reaction was set at room temperature and in each sample 1 μL of PCR product, 5 μL of KLD Reaction Buffer, 1 μL of KLD Enzyme Mix and 3 μL of deionised nuclease-free water were added. The final sample was incubated for 5 min at room temperature. Next, 5 μL of the final sample were used for DH5α transformation.

Final plasmid sequences were confirmed by DNA sequencing (Source Bioscience Sequencing).

Table 2.4. Plasmid DNA templates and primers used for the PCR reactions.
The template numbers correspond to the numbering shown in Table 2.2. (Tm = melting temperature of primer, Ta = annealing temperature used in PCR experiments)

Final vector	Template	Primers (forward and reverse, 5' to 3' end)	Tm (°C)	Ta (°C)
His6-GB1 fragment for GA	3	5'-ctttaagaaggaggtctagaATGCGGGGTTCTCATCATC-3' 5'-caggatcactCCCTTGAAAATATAGGTTCTCTC-3'	74 69	61
GCN4 wtbZIP fragment	3	5'-ttttcaagggAGTGATCCTGCTGCTCTAAAAC-3' 5'-ccggatctcaTTAGCGTTCGCCAACTAATTTC-3'	72 72	64
pET303 fragment for GA	2	5'-cgaacgctaaTGAGATCCGGCTGCTAAC-3' 5'-TCTAGACCTCCTTCTTAAAGTTAAAC-3'	73 61	62
pET303-His6-GB1-GCN4 BR	5	5'-taaTTGCTTTCGAAAAATTATCAC-3' 5'-ccaTTCTTCAACCTTGTCTTC-3'	57 61	58
pET303-His6-GB1-GCN4 bZIPGGC	4	5'-ctgcTAATGAGATCCGGCTGCT-3' 5'- ccgccGCGTTCGCCAACTAATTTC-3'	69 74	70

Table 2.5. PCR programme for PCR reactions using the Q5® High-Fidelity DNA Polymerase

Step	Cycles	Temp (°C)	Time
Initial denaturation	1	98	30 sec
Denaturation	30	98	10 sec
Annealing		Ta	30 sec
Extension		72	30 sec/kb
Final extension	1	72	2 min
Hold	-	4	-

2.2.4 Agarose gel nucleic acid electrophoresis

DNA samples were verified by agarose gel nucleic acid electrophoresis. For the 0.5% agarose gel preparation, 0.25 g of agarose powder (Merck) and 50 mL of TAE (tris base, acetic acid, EDTA) was added into a conical flask to make 0.5% agarose gel. The mixture was microwaved until it became homogenous and was cooled at RT for 5 min. 3 µL of SYBR Safe DNA Gel Stain (Invitrogen) was added into the mixture which was lightly shaken and poured into an agarose gel tank to be solidified.

5 µL of each DNA sample was mixed with 1 µL Orange DNA Loading Dye (ThermoFischer Scientific) and 5 µL of the final sample was loaded into the agarose gel. 5 µL of the GeneRuler 1 kb DNA Ladder was used and the agarose gel run at 120 V until the dye line was approximately 60-70% of the way down the gel. DNA fragments were visualized under the UV light.

2.2.5 Production of $CaCl_2$ competent cells

All solutions, tubes and centrifuges were kept on ice during the procedure and conditions were kept sterile. A 5 ml over-night culture of *E. coli* was set up and the appropriate antibiotics were added according to

Table 2.3 (Amp for BL21 and no antibiotics for DH5α). 1.0 mL of over-night culture was used to inoculate 100ml 2xTY. The appropriate antibiotics were added.

Cultures were grown at 37 °C in an orbital incubator until the OD600 of the cells reached 0.4-0.6. Flasks were placed on ice for 20 min. Cells were centrifuged at 3000 rpm for 10 min at 4 °C in a pre-cooled centrifuge. The supernatant from the cell pellet was discarded and the pellet was drained for 2 minutes. Cells were resuspended in 20 mL of cooled 100 mM $CaCl_2$, and 15% v/v glycerol solution (filtered for sterilisation). Mixture was kept on ice for 20 min. Cells were centrifuged again at 3000 rpm for 10 min at 4 °C in a pre-cooled centrifuge. The supernatant from the cell pellet was discarded and the pellet was drained for 2 minutes. Cells were resuspended in 4 mL of cooled 100 mM $CaCl_2$, and 15% v/v glycerol solution (filtered for sterilisation) and were dispensed into sterile Eppendorf tubes into 50 µL aliquots. Aliquots were snap-frozen immediately into liquid N_2 and stored at -80 °C.

2.2.6 Transformation of competent cells

The transformation of DH5α and BL21 competent cells was carried out as follows. Two aliquots of cells were thawed on ice for 15-30 min. Each aliquot was inoculated with 2-5 µL of plasmid sample and was kept on ice for 30 min. Cells were heat shocked by placing them in a 42 °C water bath for 45 sec. Cells were cooled for 3 min on ice and were immediately spread on a cell culture plate. Cells were grown over-night at 37 °C. The next day, cell plates were sealed with parafilm tape and kept at 4 °C.

Biochemical Methods

2.2.7 Polyacrylamide gel electrophoresis (PAGE)

Protein samples were verified using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). 5 µL of NuPAGE™ LDS Sample Buffer (Invitrogen) and 1 µL of 2 M DTT in cases where reducing of cysteine linked peptides was needed were added to 15 µL of protein sample. The final mixture was denatured at 95 °C for 2 min. Protein samples were spun down for 1 min and 10 µL of each sample were loaded into a NuPAGE™ 4 to 12%, Bis-Tris protein gel (Invitrogen). The PageRuler™ Unstained Low Range Protein Ladder was used (Thermo Scientific). Protein SDS PAGE gel was subjected

at 200 V for 34 min and was stained using the InstantBlue® Coomassie Protein Stain (Thermo Scientific) for at least 2 hours.

2.2.8 Measurement of protein and DNA concentrations

Protein and DNA concentrations were determined using the Beer-Lambert law:

$$A = \varepsilon c l \quad (1)$$

Where A = absorbance at 280 nm (for proteins), 260 nm (for nucleic acids) and 495 nm for AF-488-labelled DNA,

ε = molar extinction coefficient (M^{-1} cm^{-1}),

c = concentration (M),

l = pathlength (1 cm),

using a spectrophotometer. For proteins, EC are shown in Table 2.1 and for DNA samples, EC are shown in Table 2.6.

Table 2.6. DNA sequence and extinction coefficient (EC) at 260 nm for unlabelled DNA and 496 nm for AF488-labelled DNA.
Target sites recognised by GCN4 are shown in bold. EC for non-labelled oligos were calculated using the http://www.molbiotools.com calculator.

DNA construct	Sequence	EC (M^{-1} cm^{-1})
linCRE^{AF488}	5'-CC**TGACGTCA**TCCG-3' 3'-GG**ACTGCAGT**AGGC-5'	73000
hCRE ^{AF488}	5'- CC**TGACGTCA**GCC… ...CCC**TGACGTCA**GG-3'	73000
hAP-1 ^{AF488}	5'- CC**TGACTCA**GCC… ...CCC**TGAGTCA**GG-3'	73000
NS ^{AF488}	5'-GGCTAAAGCATTCT-3' 3'-CCGATTTCGTAAGA-5'	73000

linCRE (competitor)	5'-CC**TGACGTCA**TCCG-3' 3'-GG**ACTGCAGT**AGGC-5'	236296

2.2.9 Expression of GCN4 protein constructs

For the starter cultures 100 mL of 2xTY was inoculated with 4-5 colonies from the GCN4 BL21 cell culture plates, 100 μg/mL ampicillin, 34 μg/mL chloramphenicol and glucose (0.2% w/v final concentration). They were incubated overnight at 37 °C and 180 rpm shaking.

In order to produce GCN4 recombinant proteins, 3 L of 2xTY media was inoculated with confluent starter culture with a ratio 50:1. The media also contained 100 μg/mL ampicillin, 34 μg/mL chloramphenicol and 0.2% glucose. The bacterial cultures were incubated at 37 °C with 180 rpm shaking until OD_{600} reached 0.6-0.8. Next, cells were induced with 1 mM IPTG and incubated for 4 hours. Cells were harvested by centrifugation using JLA8.1000 rotor for 30 mins at 5000 x g at 4 °C. Cell pellet was collected and stored at -80 °C for protein purification.

2.2.10 Purification of GCN4 protein constructs

For the purification of His6-GB1-GCN4 constructs the following buffers were prepared and filtered through cellulose nitrate (CN) filters:

1. 1 M Tris-HCl stock, pH 8.5
2. 2 M NaCl stock
3. Resuspension buffer (RB): 100 mM Tris-HCl, 50 mM NaCl
4. Guanidine wash buffer (GW): 4 M guanidine HCl, 100 mM Tris-HCl, pH 8.5
5. Imidazole elution buffer: 100 mM Tris-HCl, 50 mM NaCl, 2 M imidazole
6. Salt wash buffer (SW): 2 M NaCl, 100 mM Tris-HCl, pH 8.5

Imidazole containing buffers were stored at 4 °C away from light.

2.2.11 Purification of His6-GB1-GCN4 proteins

Cells were thawed at room temperature and resuspended in 50-100 mL of RB. 2 M DTT was added to a final concentration of 1 mM. Next, cells were sonicated at 100% output with 15 sec pulse and 15 sec rest for a total duration of 15 min using Sonics Vibra-Cell VCX130, cv18 probe and 13 mm tip. Cells were centrifuged at 20000 rpm using the JA-20 rotor at 4 ºC for 45 min.

Cell pellet was discarded, and cell lysate was filtered through a 0.45 µm filter and then twice through 0.22 µm filters. Filters used for this step were hydrophilic polytetrafluoroethylene (PTFE) Millex® syringe filters suitable for filtration of biological samples.

Purification of His6-GB1-GCN4 proteins was performed using the AKTA FPLC system. The filtered lysate containing His6-GB1 was manually loaded (rate 3 mL/min) onto a pre-equilibrated with RB 5 ml HisTrap HP nickel column (GE Healthcare). The column was washed with 100 mL of RB (rate 5 mL/min) to remove any unbound material. Next, GW concentration was gradually increased (0-100% gradient over 100 mL), and the column was washed with 1-1.6 L of GW to remove protein bound DNA contamination until UV absorbance at 280 nm plateaued at minimum absorbance of GW. RB concentration was gradually increased to 100% and the column was washed with 200 mL of RB (rate 5 mL/min) to equilibrate with RB. Bound His6-GB1-GCN4 was eluted at 1 ml/min using a linear gradient of 0 to 2 M imidazole (EB). Protein elution was monitored by UV absorbance at 280 nm. Eluted fractions were selected and pooled after verification of protein aliquots running on SDS-PAGE gels to locate fractions of protein.

Protein fractions were pooled and in the final protein sample 5 µM TEV, 1 mM EDTA and 1 mM DTT were added to cleave the His-Tag. The final mixture was dialysed against 2 L of 100 mM Tris-HCl supplied with 1 mM EDTA and 1 mM DTT, overnight at RT using the SnakeSkin™ Dialysis Tubing, 3.5K MWCO (Thermo Scientific).

2.2.12 Purification of His6-GB1-cleaved GCN4 proteins

His-tag cleaved GCN4 protein solution was diluted x3 with 100 mM Tris-HCl and was loaded onto a pre-equilibrated with 100 mM Tris-HCl (pH 8.5) 5 ml HisTrap HP Nickel

column to separate cleaved GCN4 from His6-GB1 performing reversed phase HisTrap chromatography. The cleaved-GCN4 protein binds to HisTrap columns non-specifically, and thus can be eluted using a high concentration of salt buffers. Bound GCN4 was eluted at 1 ml/min using a linear gradient of 0 to 2 M NaCl (SW). His6-GB1 was eluted with 25 mL of EB. Protein elution was monitored by UV absorbance at 280 nm. The fractions were analysed by SDS-PAGE and those containing protein were pooled.

2.2.13 Protein sample preparation

Protein samples were concentrated using the Vivaspin® 2 Centrifugal concentrators. PD-10 columns were pre-equilibrated with 100 mM sodium phosphate, ionic strength 0.185 M controlled by NaCl, and pH 6.5. Proteins samples were buffer exchanged to biophysical buffer (100 mM sodium phosphate, ionic strength 0.185 M controlled by NaCl, 0.05% v/v Tween-20 and pH 6.5) using PD-10 columns. Protein samples were further concentrated using the Vivaspin® 2 Centrifugal concentrators and concentrations were verified measuring UV absorbance at 280 nm. Protein aliquots were snap-frozen immediately into liquid N_2 and stored at -80 °C.

AGADIR analysis: Choice of GCN4 bZIP protein sequence for biophysical experiments

The most widely-used helix content predictor is AGADIR, which predicts the conformational behavior of monomer helical peptides in solution (Muñoz & Serrano, 1994). This algorithm characterizes a peptide's helicity content at a residual content by considering intramolecular interactions taking place in an α-helix formation and it is based on the helix/coil transition theory. Figure 2.1 shows the AGADIR analysis prediction of the GCN4 full length sequence at pH 6.5, ionic strength 0.185M and 25 °C temperature.

Although, the majority of GCN4 is suggested to have <2% helicity, the bZIP domain shows two high helicity regions; the BR and the LZ, connected by a lower helicity linker.

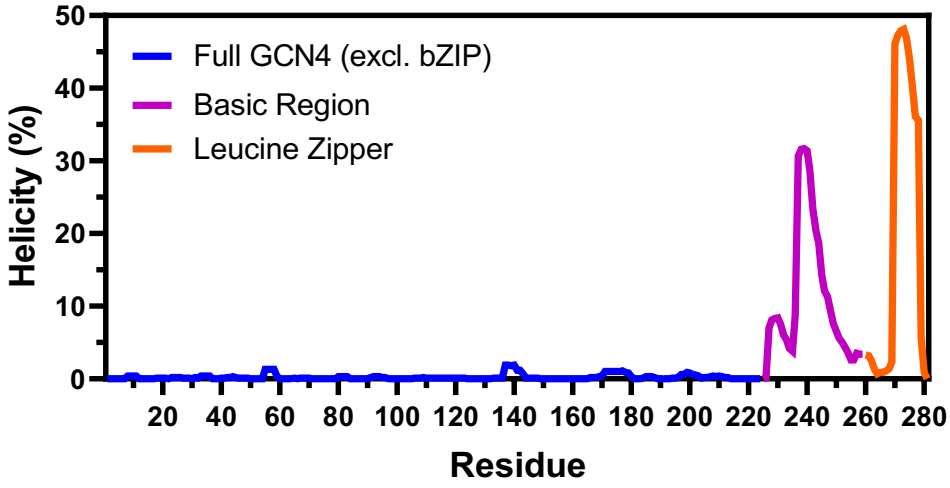


Figure 2.1. Residual prediction of helical propensity as a function of GCN4 full length sequence analysed by the AGADIR algorithm.
Sequence identifier code: P03069-1. From right to left: the LZ domain is coloured in orange, the BR in pink and the GCN4 sequence excluding bZIP domain is coloured in blue.

For the determination of the GCN4 BR-only construct used in the experiments, I performed AGADIR analysis of the *S. cerevisiae* GCN4 bZIP domain. In this analysis, I used 25 °C, ionic strength 0.185 M and pH 6.5 to match my experimental conditions. Initially, I analysed one small region which I219-R243, including the previously characterised conserved DNA-binding domain (res. 231-243) (Hakoshima et al., 2000). Next, I proceeded by performing analysis using longer sequences by adding one residue at a time.

2.3 Results

2.3.1 Design of the GCN4 constructs

For the experiments, I designed GCN4 constructs to investigate the stoichiometry of the search process (See Figure 2.3 and Table 2.1 for information regarding their sequence, MW and EC). I performed molecular biology (Site Directed Mutagenesis) and cloning (Gibson Assembly) to generate these constructs as described in the Methods.

All proteins were expressed with an N-terminal addition of six consecutive histidines, allowing purification of the protein due to his-tag's high affinity to immobilized metal chelate affinity chromatography (IMAC) columns. His-tag competes for binding to the metal-charged resin and the protein can be eluted from an IMAC column. In all of the proteins, the folded GB1 domain was also added, to increase protein solubility and half-life, and thus increase purification yield. Finally, a TEV cleavage site (Glu-Asn-Leu-Tyr-Phe-Gln-Gly) was included between the His6-GB1 domain and GCN4 in order to cleave His6-GB1 domain from the functional GCN4 protein using TEV protease.

The sequence of the obtained dimerization-competent GCN4 construct was based upon that for the crystal structure of GCN4, but with additional amino acids at the N-terminus. The aim was that the residual helical structure predicted in the basic region would not be altered significantly from that of the full-length GCN4 protein. AGADIR analysis of the GCN4 BR domain was performed to obtain the predicted helical propensity of various BR lengths. The BR in full-length GCN4 is expected to be highly helical, with a peak helicity of ~68 % located from Ala238 to Ala239. The peak BR helicity is completely unchanged for peptides starting at S225 and earlier, with incrementally lower helicity as more residues are removed from the N-terminus. Therefore, S225 was chosen as the starting residue. This starting residue was used as a basis for the other constructs.

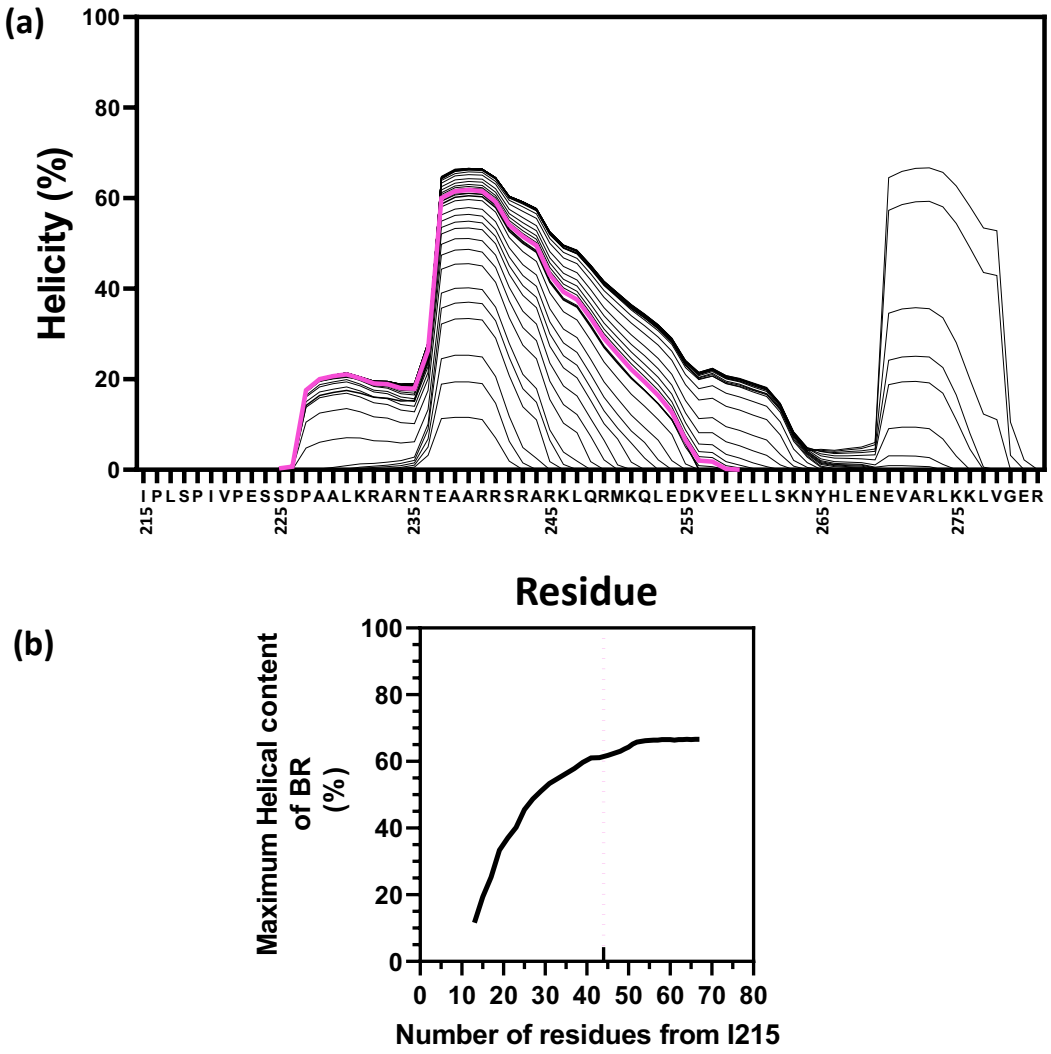


Figure 2.2. Helical content calculation using the AGADIR algorithm.
(a) Helical content as a function of GCN4 bZIP truncated sequences and (b) maximum helical content of BR as a function of number of residues from I215. The AGADIR output for BR-only peptide (S225-E259) used in this study is coloured in pink.

To investigate the behaviour of monomeric GCN4, I designed one GCN4 construct removing most of the LZ, hereafter referred to as BR to reflect the fact that it is largely the sequence identified as the basic region. This construct lacks the native leucine heptads in the leucine zipper and cannot dimerize through a coiled coil formation. The end of the basic region and start of the zipper is a difficult point to define. I investigated sequences with various endpoints using AGADIR and found that increasing the number of residues increased helical content within the basic region marginally. I decided upon endpoint of

E259, which contains 7 residues (LEDKVEE) of the Uniprot defined zipper, since the helicity in the DNA binding region almost matches the peak obtained in the full sequence, however most of the zipper is not present and therefore the construct should not dimerize (Figure 2.2). This also corresponds roughly with the point of termination of the CREB BR that is in use in the Shammas lab. In order to allow protein concentration calculation using the EC determined by S. C. Gill & von Hippel, (1989) at least one tryptophan or tyrosine is needed to be present in the GCN4 sequence. There is a native tyrosine in the LZ region, but none in the BR construct. Thus, one tryptophan was added to the C-terminus of the BR to allow protein concentration determination.

To investigate the behaviour of dimeric GCN4 I added three amino acids in the carboxyl end of the bZIP sequence: glycine, glycine and cysteine, resulting in the FD construct (Figure 2.3). In this way, using oxidative conditions, FD could be cysteine linked through disulphide bond formation. This approach has been used previously in the literature (Berger et al., 1998; Cranz et al., 2004; Park et al., 1996) for GCN4-bZIP and also in our lab with CREB-bZIP.

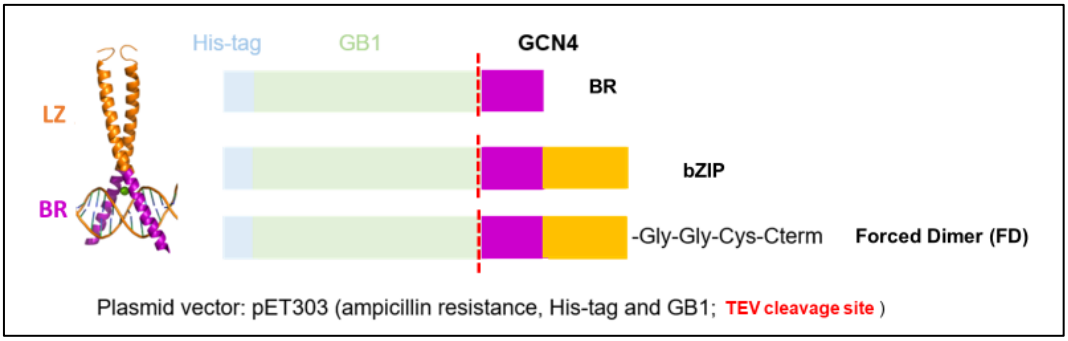


Figure 2.3. Illustration of the three designed protein constructs.
The six-histidines-tag is shown in light blue, the GB1 domain is shown in light green, BR is shown in pink and the LZ in orange. The TEV cleavage site is demonstrated as a dashed red line.

2.3.2 Cloning, expression and purification of GCN4 constructs

Cloning of GCN4 constructs

The plasmid vector for the expression of GCN4 bZIP was produced using the protocol described in paragraphs 2.2.2 and 2.2.3. All agarose gel nucleic acid electrophoresis experiments were performed following the protocol in paragraph 2.2.4. At first, three linear DNA fragments with complementary flanking regions were generated: 1) pET303 (~5300 bp), 2) His6-GB1 (~260 bp) and 3) GCN4 bZIP (~170 bp). These fragments were created and amplified through the SDM method and PCR amplification. The PCR products were verified through agarose gel nucleic acid electrophoresis. The agarose gels are shown in Figure 2.4(b). The bZIP fragment gel band is located at the 75 bp region, the His6GB1 fragment is located at the 75-200 bp region, slightly higher than the bZIP since it has higher MW, and the pET303 fragment band is seen at 5000 bp region.

Next, Gibson Assembly reaction was performed from which the PCR products were cloned into circular vectors. The plasmids were then used to transform DH5α cells. Vectors from successfully amplified colonies were isolated and sent for sequencing.

The pET303-His6-GB1-GCN4bZIP plasmid was used as a template in the SDM reactions to further create the FD and BR expression plasmids. These fragments were created and amplified through the SDM method and PCR amplification. The PCR products were verified through agarose gel nucleic acid electrophoresis. The agarose gels are shown in Figure 2.4(c). The pET303-His6-GB1-GCN4BR and the pET303-His6-GB1-GCN4FD (which are ~5600 bp each) plasmids gel bands are both located between the 5000 and 7000 bp region. The FD Lane also shows one additional faint band at MW<75 bp. This band corresponds to primers forming primer-dimers. However, this issue did not impede the KLD reaction used for circularization of linear DNA PCR products. This was evident by the fact that DH5α cells were able to form bacterial colonies when transformed with the KLD reaction products and grown on antibiotic-containing agar plates. The plasmids sequence was verified with Sanger sequencing and confirmed the correct vector sequences for pET303-His6-GB1-GCN4BR and pET303-His6-GB1-GCN4FD.

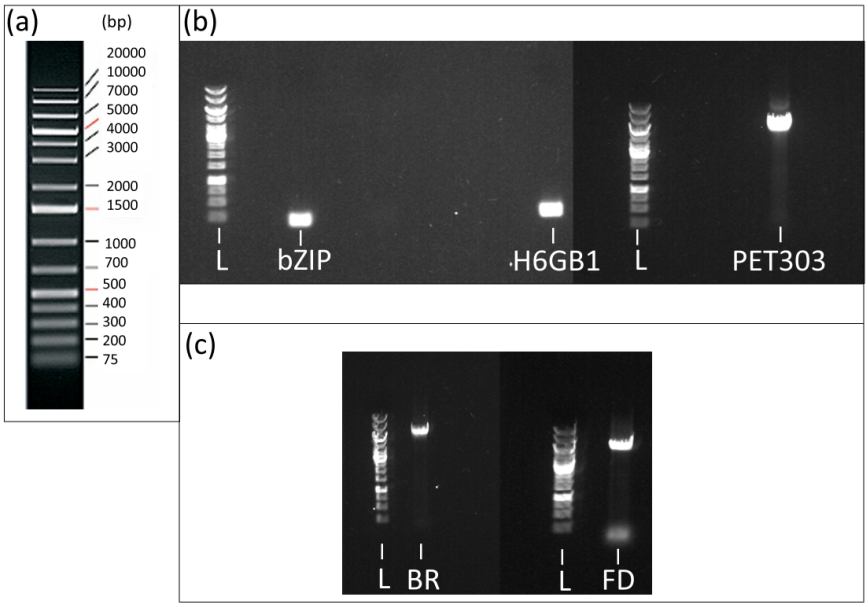


Figure 2.4. Agarose nucleic acid electrophoresis gels validating DNA fragments.
(a) GeneRuler 1 kb Plus DNA ladder indicating the MW that correspond to each band (b) L: ladder, bZIP = GCN4 bZIP PCR product to use in Gibson Assembly reaction, H6GB1 = His6GB1 PCR product to use in Gibson Assembly reaction, pET303 = pET303 PCR product to use in Gibson Assembly reaction. (c) Site-directed mutagenesis PCR products (uncircularised plasmids) to use in KLD reaction. L = ladder, BR = pET303-His6-GB1-GCN4BR linearized plasmid, FD = pET303-His6-GB1-GCN4FD linearized plasmid**.**

GCN4 constructs expression and purification

GCN4 bZIP, BR and FD were expressed using, initially, the protocol for the expression CREB bZIP developed by the Shammas group. In order to test whether the particular protocol is also suitable for GCN4, I performed a trial expression of GCN4 FD. I ran a sample from each stage on an SDS-PAGE gel (Figure 2.5 and Figure 2.6). In Figure 2.6, lane 1 demonstrates that there is no leaky expression, and that protein expression is successfully induced after addition of IPTG (lane 2) with a satisfactory yield. In Figure 2.5, protein samples from resuspended cells to the supernatant from sonicated and centrifuged cells is shown. The protein samples were collected in the following order: 2 and 3, 4, 5, 1. The gel indicates that GCN4 monomers and dimers are not lost during the protein purification stage where the cell lysate is prepared for AKTA purification. In addition to this, lane 1 (sonicated-cell pellet) compared to lane 5 (sonicated-cell supernatant) demonstrates that GCN4 (soluble protein) is not deposited in inclusion bodies in the cell

pellet where insoluble proteins are usually found and thus, the supernatant was used for further GCN4 purification.

Figure 2.6(a and b) demonstrates the samples collected from an IMAC His6GB1-GCN4 FD purification using FPLC. Absorbance traces obtained during this purification are shown in Figure 2.7. Lanes 4 to 7 show no significant protein elution. At these stages, the column was subjected to buffer washes in order to clean the column-bound protein from cell contaminants. Lanes (a):8-14 and (b):1-4 demonstrate protein aliquots from an imidazole gradient elution. His6GB1-FD monomer band is located at ~20 kDa and His6GB1-FD dimer is located at ~35 kDa. Lanes 8 and 9 contain samples from the first and second sharp peaks of the chromatography trace (Figure 2.7(a)), where there are a lot of cell contaminants and no His6GB1-FD. His6GB1-FD elutes at the third flattened peak as observed in lanes (a)10-14 and (b) 1-4. Thus, the aliquots B2-C15 were pooled for subsequent TEV cleavage.

Figure 2.6(c) shows the samples collected from the reverse IMAC GCN4 FD purification. Lanes 1 and 2 show the protein before and after TEV cleavage. As observed in lane 1, uncleaved GCN4 FD is located at 15 kDa and 35 kDa. In lane 2, cleaved GCN4 FD monomer is shown at 10 kDa and dimer at 15 kDa. TEV protease whose MW has been determined equal to ~25 kDa, is also present at little higher MW of 25 kDa on the gel. Lanes 4 to 12 correspond to aliquots obtained from the 2 M NaCl gradient elution. In all samples, GCN4 FD is present, however, lanes 4 and 5 include some faint bands of contamination. Thus, these aliquots were discarded and the B8-E8 aliquots were pooled for further pure GCN4 FD dimerization. The respective chromatography trace is demonstrated in Figure 2.7(b).

Pure GCN4 FD was further subjected to dimerization using oxidative conditions. The first trial was performed buffer-exchanging the GCN4 FD sample in 10 mM Tris pH 9.5 buffer and diluting 10 times. After overnight incubation at room temperature, the pre-oxidization and post-oxidization samples were analysed on an SDS-PAGE (Figure 2.6(d)). Unfortunately, the “post-oxidization” sample showed unsuccessful dimerization but also a band at 25 kDa corresponding to contamination. To produce pure dimerized GCN4 FD protein, I buffer-exchanged the protein sample with RB and repeated reverse IMAC purification. Next, I buffer-exchanged the pooled pure GCN4 FD aliquots with 10 mM Tris pH 10.2 buffer and diluted 10 times. Post-oxidization sample is shown is lane 5 where the

majority of protein is sufficiently shown to be dimerized. His6GB1 protein construct is 15 kDa and therefore it is highly expected to run at the same position as the GCN4 FD dimer. In order to confirm whether GCN4 FD is pure from His6GB1, 1 mM of DTT (redox reagent) was added and run on gel. Lane 4 shows that upon addition of 1 mM DTT all GCN4 FD is reduced and no HisGB1 protein is present.

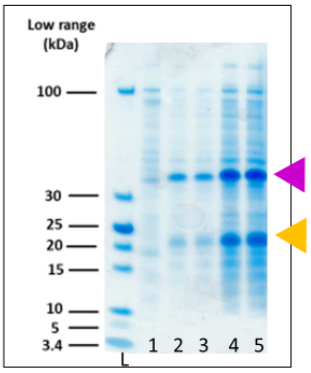


Figure 2.5. Representative 4 to 12% Bis-Tris protein gel of GCN4 FD early stages of purification process.
One protein sample from each purification stage is shown: 1=pellet collected from cell sonication, 2 and 3=post-thawing cell resuspension, 4=post-sonication, 5=post-centrifugation supernatant. Pink arrow shows GCN4 FD dimer; Orange arrow shows GCN4 FD monomer.

DNA contamination in final GCN4 FD and bZIP samples

The OD 260:280 ratio is a measure of sample purity. Nucleic acid contamination in a protein sample should be kept to a minimum, as it can interfere with the activity of nucleic acid-binding proteins like GCN4 transcription factor, and in biophysical experiments would create errors in the calculated DNA concentrations. Nucleic acid contamination does not show up on a protein gel, which is why measuring the ratio is essential for an accurate picture of purity. Due to the differential absorbance spectra of DNA and proteins contamination is typically assessed using the ratio of absorbance at 260 nm and 280 nm. High values of 1-2 indicate the presence of nucleic acids and lower values of around 0.6 (or below) indicate the presence of pure proteins.

Washes during purification in high salt buffers help minimize the presence of nucleic acid. The initial CREB bZIP purification protocol included a 2 M NaCl wash to get rid of DNA

contamination. However, when final 260 nm and 280 nm UV-Vis absorbance of pure protein samples was measured, a final ~0.96 ratio of 260 nm:280 nm was calculated. Thus, a more successful step to avoid DNA-contamination was needed. For this purpose, I incorporated a 4 M guanidine wash during IMAC purification of his-tagged construct instead. High concentrations of guanidine hydrochloride denature GCN4 and substantially decreases its ability to bind to DNA. Also, the high ionic strength of 4 M guanidine hydrochloride decreases GCN4 electrostatic interactions with the negative charged DNA. With the incorporation of this step, pure protein was isolated with a 260:280 ratio ~0.55.

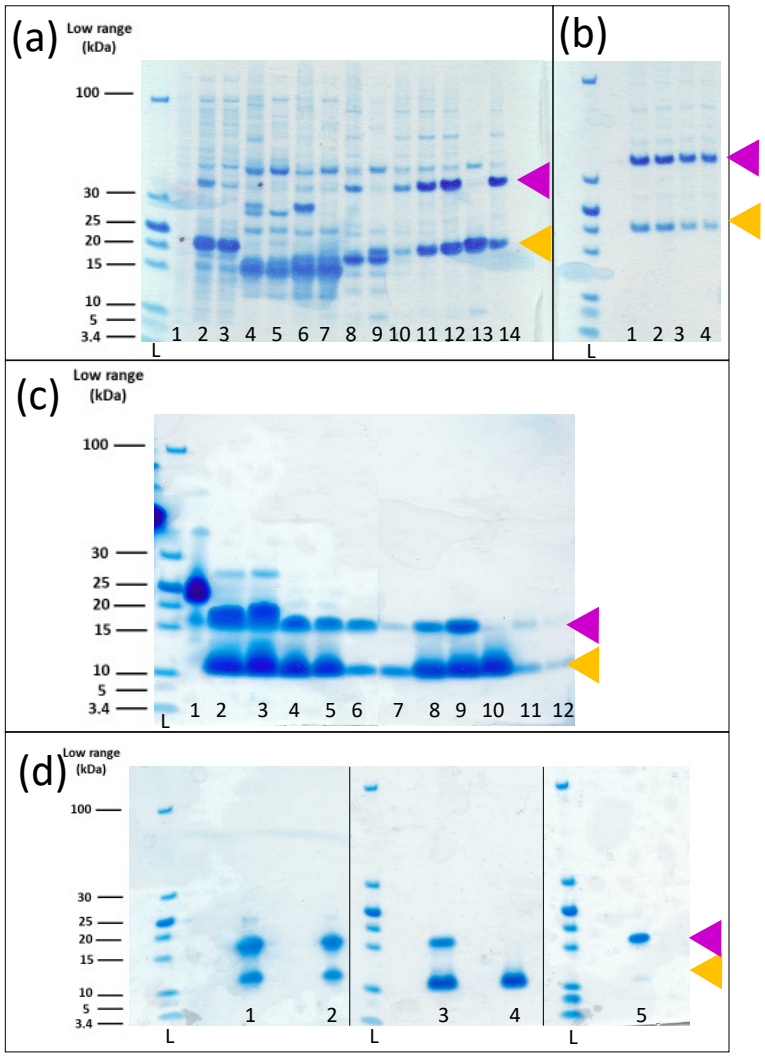


Figure 2.6. Representative 4 to 12% Bis-Tris protein gels from GCN4 FD IMAC purification.

(a) and (b): IMAC purification of His6-GB1-GCN4-FD where monomer GCN4 is present at ~20 kDa range (shown with a yellow arrow) and dimer GCN4 is present at ~30 kDa

(shown with a pink arrow) and (c) reverse IMAC purification of GCN4-FD where monomer GCN4 is present at ~12 kDa range (shown with a yellow arrow) and dimer GCN4 is present at ~15 kDa (shown with a pink arrow). Lanes 1 to 14 are showing (a) L=Ladder, 1= pre-IPTG-induced cell lysate, 2 and 3=post-IPTG-induced cell lysate, 4 and 5=cell lysate loading flow-through, 6=guanidine wash, 7= RB wash, 8=A13, 9=A15, 10=B2, 11=B1, 12=B6, 13=B10, 14=C1, (b) L=Ladder, 1= C3, 2=C5, 3=C7, 4=C9. (c) L=Ladder, 1=pre-TEV, 2 and 3=post-TEV, 4=B4, 5=B5, 6=B8, 7=B13, 8=B15, 9=C2, 10=C8, 11=D9, 12=E2. (d) 4 to 12% Bis-Tris protein gel showing oxidisation trials of FD dimerization: L=Ladder, 1=pre-oxidization sample, 2=post-incubated sample with 10 mM Tris pH 9.5, 3=pre-oxidization sample, 4=pre-oxidization sample with 1 mM DTT addition, 5=post-incubated sample with 10 mM Tris pH 10.2

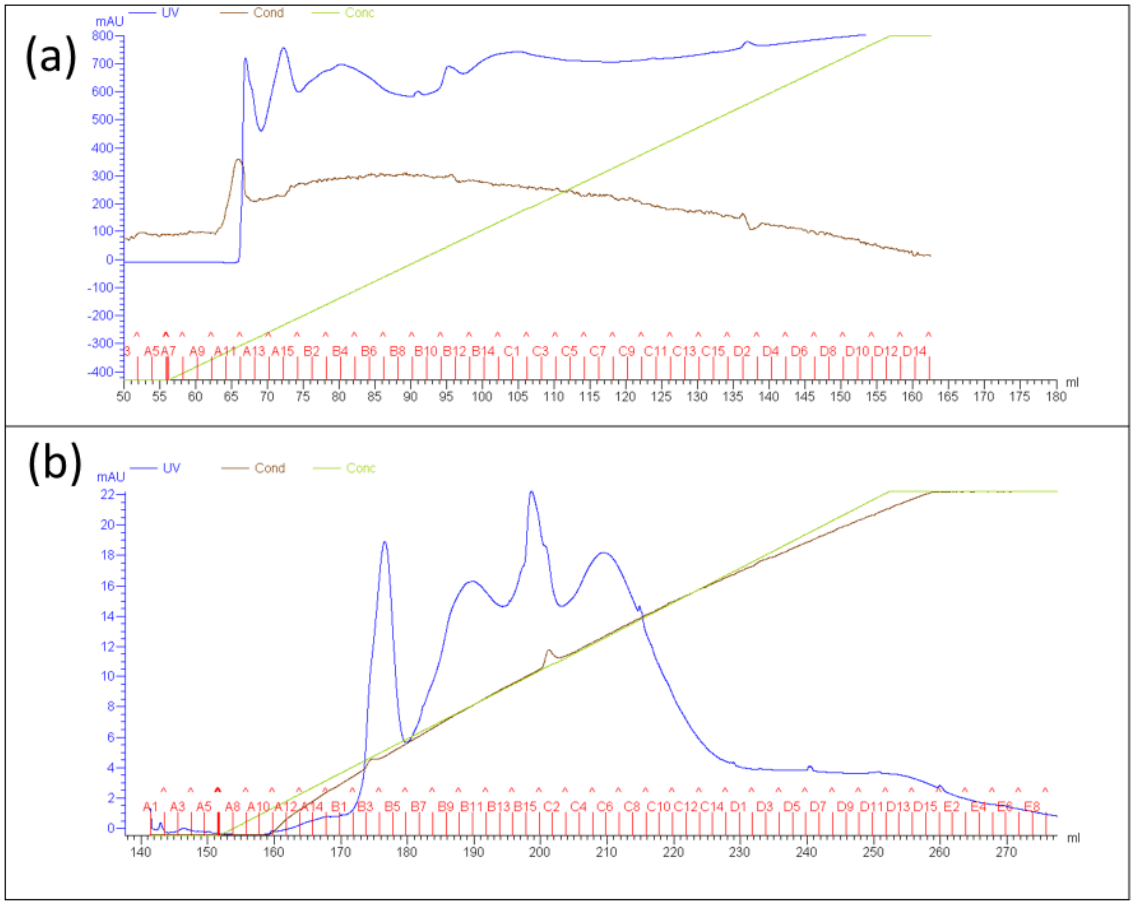


Figure 2.7. Representative chromatography traces of GCN4 FD IMAC purification. UV absorbance profile at 280 nm of the eluents from (a) His6-GB1-GCN4-FD and (b) GCN4-FD from IMAC purification, as a function of volume eluted. Black line (Cond) represents eluents' conductance fluctuation, and green (Conc) line represents percentage of concentration of elution buffer used at the particular volume.

The same protocol was used for the expression and purification of GCN4 bZIP. Similar gels and chromatography traces were obtained (Figure 2.8) and the expression protocol worked successfully for protein isolation.

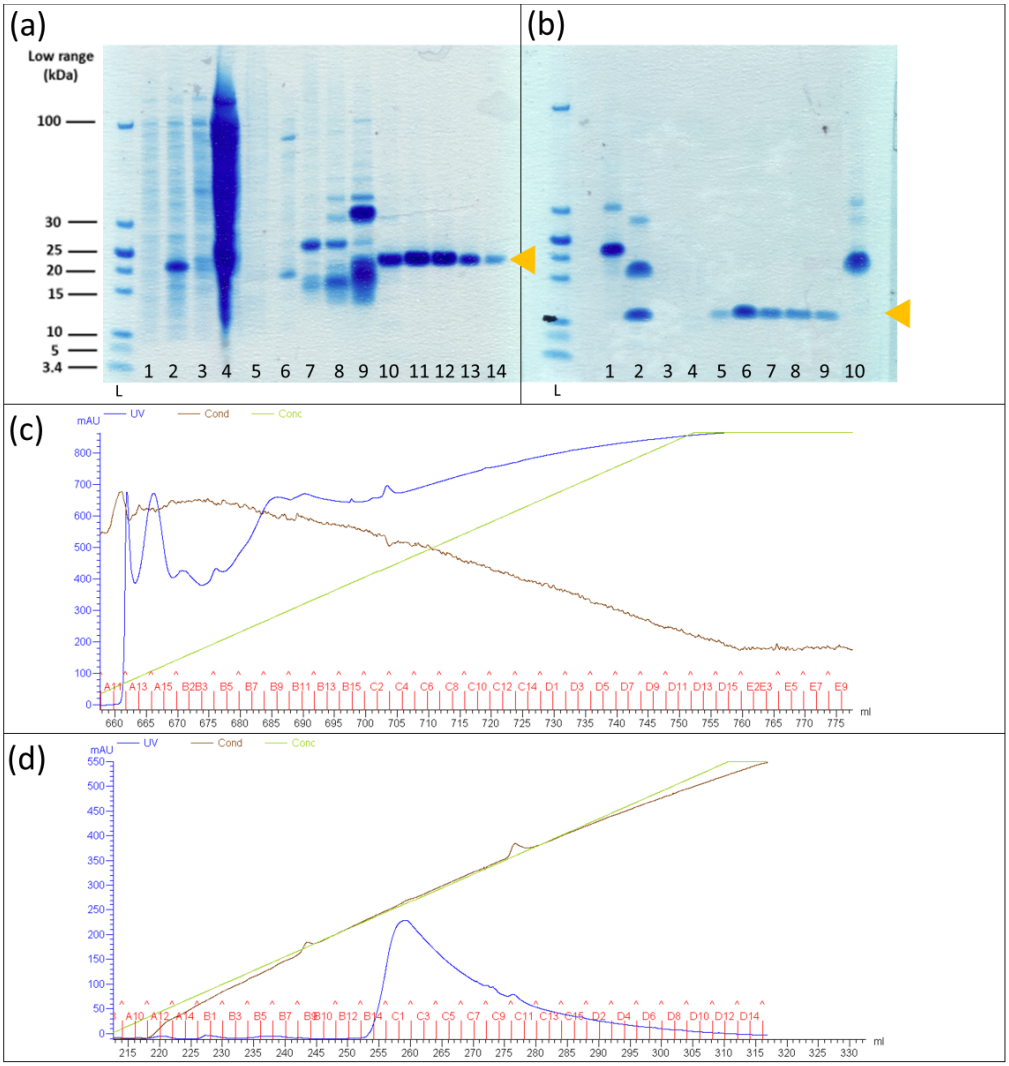


Figure 2.8. Representative GCN4 bZIP IMAC purification SDS-PAGE gels and chromatography traces.

(a) and (b): 4 to 12% Bis-Tris protein gel showing (a) IMAC purification of His6-GB1-GCN4bZIP where GCN4 is present at ~23 kDa range (shown with a yellow arrow) and (b) reverse IMAC purification of GCN4bZIP where GCN4 is present at ~12 kDa range (shown with a yellow arrow). Lanes 1 to 14 are showing (a) L=Ladder, 1=pre-IPTG-induced cell lysate, 2=post-IPTG-induced cell lysate, 3=cell lysate loading flow-through, 4=guanidine wash, 5=RB wash, 6=A11, 7=A12, 8=A13, 9=A15, 10=B8, 11=B11, 12=B15, 13=C14, 14=C8, (b) L=Ladder, 1=pre-TEV, 2=post-TEV, 3=post-TEV sample flow-through, 4=B1, 5=B13, 6=C2, 7=C5, 8=C9, 9=D2, 10=post-purification 2 M imidazole wash of column. (c) and (d): Chromatograms showing UV absorbance at 280 nm of the eluents from (c)

His6-GB1-GCN4bZIP and (d) GCN4bZIP from IMAC purification, as a function of volume eluted. Black line (Cond) represents eluents' conductance fluctuation and green line (Conc) represents percentage of concentration of elution buffer used at the particular volume.

Expression and purification of GCN4 BR-only peptide

The same protocol was also used for the expression and isolation of GCN4 BR construct based on IMAC purification technique.

As previously mentioned, the 2 M NaCl wash initially used for the DNA contamination in full-length bZIP protein samples was inadequate and was substituted with a 4 M guanidine wash. However, this deemed adequate in the case of the BR-only because the final 260:280 ratio was calculated equal to ~0.59 (which was considered satisfactory).

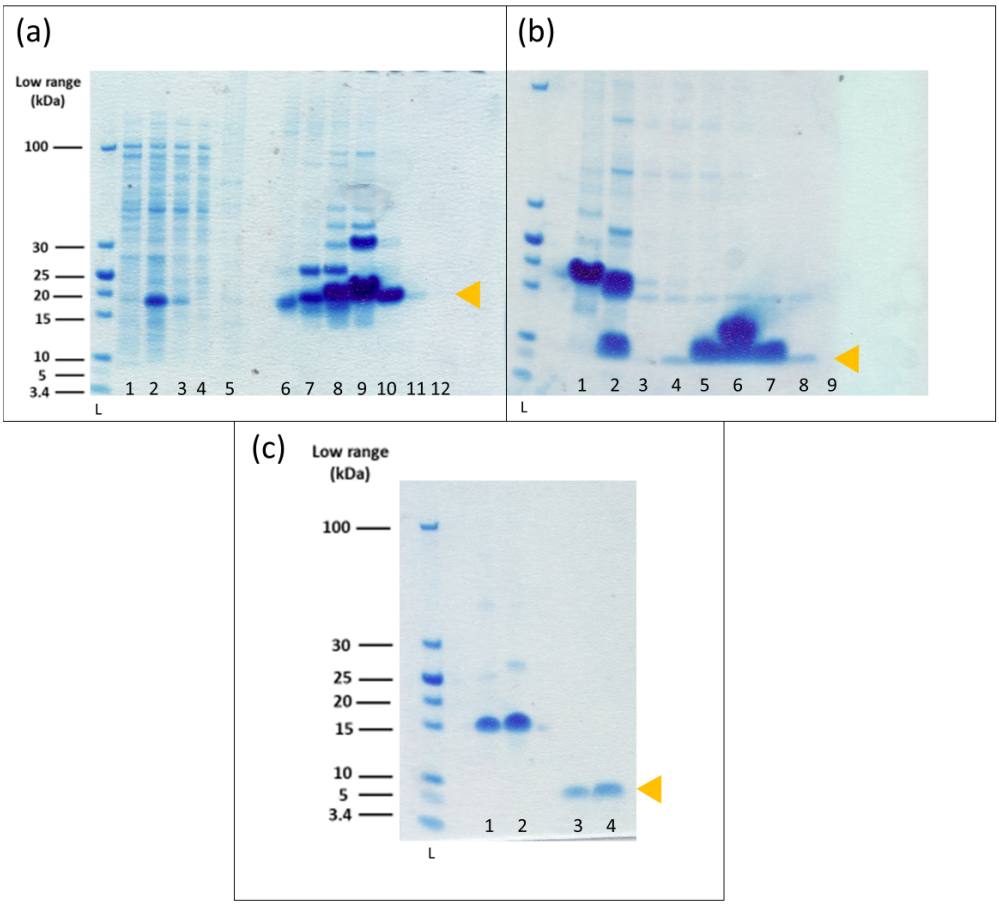


Figure 2.9. Representative 4 to 12% Bis-Tris protein gels from GCN4 BR IMAC purification.
(a) IMAC purification of His6-GB1-GCN4-BR where GCN4 is present at ~18 kDa range (shown with a yellow arrow), (b) reverse IMAC purification of GCN4-BR where GCN4 is

present at ~6 kDa range (shown with a yellow arrow) and (c) second reverse IMAC purification of GCN4-BR where GCN4 is present at ~6 kDa range (shown with a yellow arrow). Lanes 1 to 14 are showing (a) L=Ladder, 1= pre-IPTG-induced cell lysate, 2=post-IPTG-induced cell lysate, 3=cell lysate loading flow-through, 4= RB wash, 5=guanidine wash, 6=A12, 7=B1, 8=B2 , 9=B5, 10=B8, 11=B13, 12=C1, (b) L=Ladder, 1=pre-TEV, 2=post-TEV, 3=B4, 4=B6, 5=B10, 6=B11, 7=B13, 8=B14, 9=B15, (c) L=Ladder, 1 and 2= 2 M imidazole wash, 3 and 4=SW-eluted GCN4 BR.

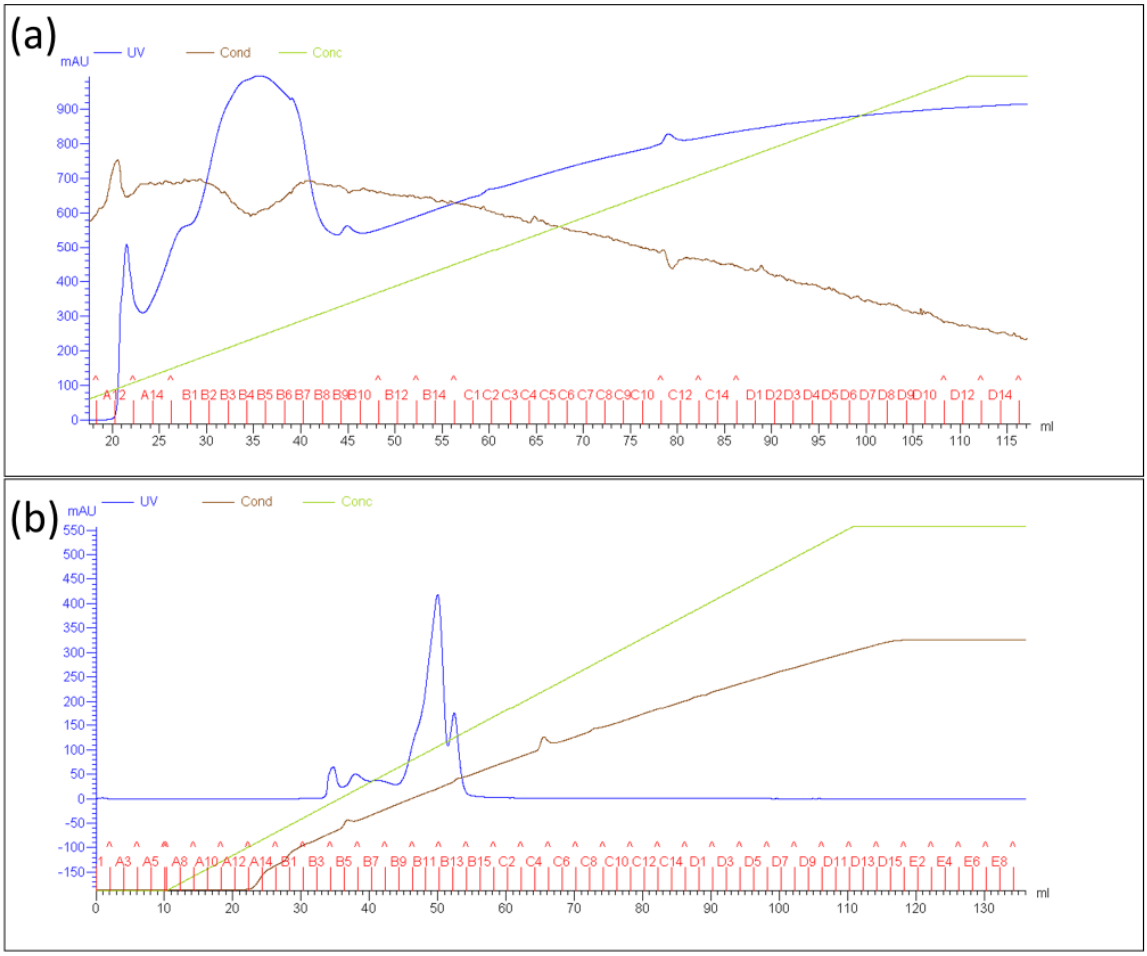


Figure 2.10. Representative chromatography traces of GCN4 BR IMAC purification
UV absorbance at 280 nm of the eluents from (a) His6-GB1-GCN4-BR and (b) GCN4-BR from IMAC purification, as a function of volume eluted. Black line (Cond) represents eluents' conductance fluctuation and green line (Conc) represents percentage of concentration of elution buffer used at the particular volume.

2.4 Discussion

2.4.1 Developing suitable models for monomeric and dimeric GCN4

I designed three different GCN4 constructs: two with a full-length bZIP, and one with only the BR domain. My approach is not unprecedented. bZIP domains have frequently been used as models of dimeric GCN4, and a C-terminal GGC extension to force dimerization via oxidation, as I have used here, has been utilised in at least two GCN4 studies (Cranz et al., 2004; Moran et al., 1999; Wakankar & Borchardt, 2006). Furthermore, the basic region alone is typically used as a model for monomeric GCN4, as I have also done here. In both cases the vast majority of the GCN4 sequence is removed, and of course this might affect the behaviour of the system – for example other regions of the protein might interact with the bZIP itself via an auto-inhibition scheme for binding. However, the approach does have an advantage in making the results less specific to GCN4 itself, and the protein is amenable to purification with large yields. Another possibility is that neighbouring regions of the sequence might impact the residual helical propensity within the GCN4 bZIP. Since my project aims to characterise the helicity, particularly in the basic region, I was careful to consider this possibility when deciding the residue numbers to use in my constructs. The full-length bZIP constructs I made are based on S225- R281. R281 is the last amino acid in GCN4 sequence. This sequence has a matching predicted helicity to the full-length GCN4 protein by AGADIR and includes the N-capping residues DPAAL (res. 226-230) that Hollenbeck, McClain, and Oakley (2009) showed play an important role in helix stabilization and DNA binding affinity. Figure 2.1 shows that the GCN4 bZIP domain has significantly higher helical content than the rest of the sequence with helical propensity dropping to lower than 1% in amino acids before S225 in the protein sequence. Thus, no propagation of helical structure from residues outside my chosen construct is expected. The basic region construct starts at this same residue but ends at E259. Although AGADIR analysis showed a small increase in helicity might be achieved for longer versions, the peak helicity was almost identical to that predicted for full-length monomeric GCN4 and terminating the sequence at this point reduced the likelihood of any unwanted dimerization.

2.4.2 Cloning, expression and purification of GCN4 protein constructs

In this work, I have successfully used molecular biology and cloning techniques (Gibson Assembly and SDM) to generate GCN4 bZIP, FD and BR expression vectors. I adapted a previously developed protein expression and purification protocol for the high quality and quantity production of GCN4.

I confirmed identity and purity by collecting samples along the process and performing gel electrophoresis methods, nucleotide sequencing, measuring 260:280 nm ratios etc. These steps allowed me to monitor the process of cloning, expression and purification of the GCN4 constructs and identify, as early as possible in the process, complications that would decrease protein sample's quality and quantity. The expression and purification of BR was very straight forward, with suitable protein being obtained with no required changes to the protocol. Initial samples of bZIPs suffered from significant nucleic acid contamination however, presumably because it binds more tightly to DNA than the monomeric BR peptide. It is important to note that biophysical studies of TFs and especially kinetics experiments using the stopped-flow or NMR spectroscopy, require DNA-free samples of high purity and concentration. I found that inclusion of a 4 M guanidine HCl wash step, which is expected to destabilise the bZIP dimer and also the putative electrostatically favourable complexes with DNA, significantly decreases the DNA contamination so the purity is suitable for biophysics experiments. This avoided the fairly common practice of introducing nucleases into the lab, which is preferable because small quantities of these enzymes in samples might interfere with accurate DNA concentrations needed for biophysical studies.

2.5 Conclusion

I have developed a protocol to produce monomeric, dimerization-competent, and completely dimerized versions of GCN4 protein in high purity and yield.

Chapter 3. Biophysical characterisation of GCN4 binding to DNA

3.1 Introduction

The final bZIP transcription factor complex with DNA consists of two protein monomers binding together at a single binding site. Although studies have characterised both apo- and bound-forms, there is more to be explored regarding the search/binding process. In particular, little is known about the process by which disordered DNA binding domains find their target sites, and whether this differs to the search process of folded DNA binding domains. For bZIP domains there is also a question of search stoichiometry - whether a bZIP searches and binds to the target site as monomer or as a dimer. Of course, the target recognition process, being a complicated biological process may even involve a combined phenomenon.

A simple example model to describe the process from a single monomer bZIP leading to a full bZIP dimer-DNA complex formation, is shown in Figure 3.1. where k_i and k_{-i} represent fundamental forward and reverse rate constants of the *i* reaction, respectively. This model is suitable for bZIPs that can only homodimerize. Including heterodimerizing bZIPs into this model will lead to a more complicated scheme. Descriptions of the binding pathway have traditionally ignored the presence of non-target DNA, however. In typical terminology reaction 3 followed by reaction 4 describes the monomer pathway. In this case, a bZIP monomer recognises the cognate site, binds so that the second bZIP monomer interacts and forms the ternary complex. On the other hand, there is the dimer pathway which is represented by reaction 7 followed by reaction 5. There may be a kinetic preference for one particular pathway or a combined one, and whether the protein binds to the target as a monomer or dimer may also depend on the concentration of non-target DNA present. Studying the kinetics and determining the thermodynamic constants of each reaction, by designing *in vitro* experiments that are suitable to model each reaction, may suggest the search pathway of a particular bZIP.

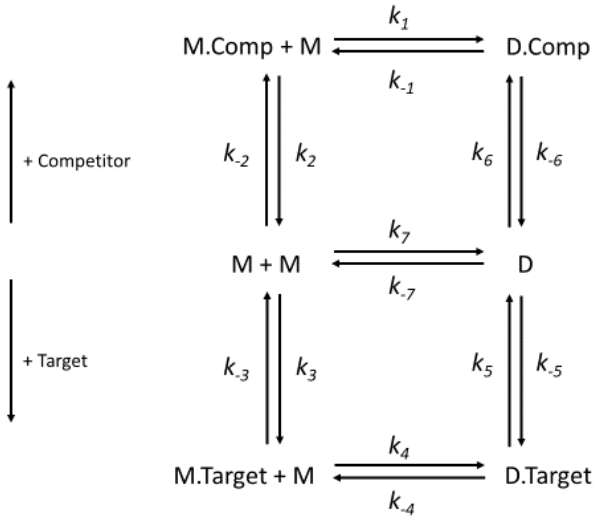


Figure 3.1. A model example of the association and dissociation reactions between a bZIP and its target DNA or its non-target DNA (competitor).
Target= bZIP target DNA sites, M= bZIP monomer, D=bZIP homodimer, M.Comp=bZIP monomer in complex with competitor non-target DNA, M.Target=bZIP monomer in complex with target DNA, D.Comp= bZIP homodimer in complex with competitor, D.Target=bZIP homodimer in complex with target DNA. k_i and k_{-i} are the rate constants for each indicated process.

GCN4's biological action is not only determined by its affinity to specific DNA sequences and specificity but also by the association and dissociation rate constants (i.e. k_{on} and k_{off}), and by its interaction with non-specific DNA. These constants, and not only the equilibrium constant, K_d, determine the success of the protein-DNA interactions in molecular signalling. Previous studies (Berger et al., 1998; Cranz et al., 2004) have addressed whether GCN4 is monomeric or dimeric upon DNA searching, suggesting the monomer pathway as advantageous for GCN4. For example, Cranz et al. (2004) used stopped-flow kinetics studies with fluorescently labelled DNA constructs and found that GCN4 bZIP and BR both bound to targets with similar rate constants. They suggest that binding as a monomer is likely favoured in typical *in vitro* experiments since the protein does not need to "wait" to dimerize free in solution first. However, they also discuss how this may differ significantly within a cellular context, notably because of non-target DNA.

Previous results with monomeric versions of CREB in Shammas lab showed that it displayed < 2-fold binding specificity for the CRE target. It is thought that this feature is

a major reason that it follows a dimeric search pathway in the presence of competitor DNA. In contrast previous work with GCN4 has suggested that binding of monomer is specific for the target. Firstly, substantial helix is acquired upon binding to target DNA, whereas minimal changes are observed with non-target DNA. Secondly in ITC experiments binding to target DNA was well characterised, but no heat changes were observed upon mixing with non-target DNA (X. Wang et al., 2003). How specific is GCN4 monomer for its targets?

In this Chapter, my aim is to thermodynamically characterise GCN4 binding to AP-1 and CRE sites by using different versions of the GCN4 bZIP domain and provide essential information regarding the kinetics of GCN4 in complex with target or non-target DNA sequences.

3.2 Materials and Methods

3.2.1 DNA sample preparation

For the preparation of linear DNA samples, the appropriate oligonucleotides were purchased from Life Technologies. Oligonucleotides were diluted with nuclease-free water to 1 mM. Forward and reverse oligonucleotides were added at a ratio of 1:1 and were heated up to 95 °C for 1 hour and then slowly cooled down using a thermocycler. Next, DNA samples were immediately kept on ice.

For the preparation of hairpin DNA samples, one self-annealing oligonucleotide for each DNA construct was purchased from Life Technologies. Primer was diluted with nuclease-free water and was subjected to 2 min heating at 95 °C using a thermocycler. Next, DNA samples were immediately kept on ice.

DNA concentrations were calculated by measuring UV absorbance at 260 nm for non-labelled DNA or 495 nm for AF488-labelled DNA. When buffer exchange was required, DNA samples were dialysed against dialysis buffer overnight at RT using the Slide-A-Lyzer MINI Dialysis Devices.

3.2.2 Fluorescence Anisotropy equilibrium binding experiments

For the equilibrium binding fluorescence anisotropy experiments, which were performed in order to obtain the K_d values, fluorescently labelled DNA samples were used. Each linear double-stranded DNA was produced by annealing one 5'-fluorescently-labelled primer in combination with one complementary unlabelled primer. For the fluorescently labelled hairpin DNA versions, one 5'-labelled primer was used. The oligonucleotides used in this study are shown in Table 2.6 and the annealing protocol is described in Paragraph 3.2.1.

For the K_d determination, protein samples were prepared in 100 mM sodium phosphate, ionic strength 0.185 M controlled by NaCl, 0.05% v/v Tween-20 and pH 6.5. A series of samples were used with increasing protein concentration. For each protein-DNA system, I performed a pilot experiment using protein samples ranging from 0.3 nM to 100

μM to determine the range of K_d constant. Each equilibrated protein sample contained 5 nM labelled DNA whose fluorescence anisotropy was measured using a fluorimeter. After determining the K_d range and to obtain the K_d value, I optimized the protein and the DNA concentrations according to the pilot experiments. I collected at least 10 anisotropy measurements for each equilibrated protein-DNA sample using a square 3 mm path length quartz.

All equilibrium binding experiments were performed at 25 °C and the temperature was monitored during the data collection. The Alexa Fluor 488 (AF488) is a fluorescent dye that absorbs at 495 nm and emits at 519 nm. Thus, 495 nm and 519 nm were used for excitation and emission wavelengths, respectively. The band passes were set at 10 nm. AF488's EC_{495} of 73000 cm^{-1} M^{-1} was used to calculate labelled DNA concentration from OD_{495}.

All equilibrium binding experiments were performed using the spectrofluorometer (PerkinElmer LS55) based in the Shammas group and the spectrofluorometer (Horiba FluoroMax-4) located in the biophysical suite at the Department of Biochemistry, University of Oxford.

3.2.3 Association and dissociation kinetics by stopped-flow

I used a stopped-flow mixing machine to obtain the k_{on} and k_{off} rate constants. The stopped-flow technique involves rapid mixing of two sample solutions in a small reaction cell and the fluorescence is recorded. For the k_{on} estimation, five different concentrations of protein samples were prepared, which were mixed with a constant concentration of DNA, using the stopped-flow. For the k_{off} rates, three or five saturating concentrations of unlabelled competitor was mixed with steady concentration of a mixture containing protein and fluorescent DNA. The unlabelled linCRE was used as a competitor in all the experiments.

The experiments were performed at 25 °C and the same biophysical buffer as the equilibrium binding assays was used. 495 nm and 519 nm were used for excitation and emission wavelengths, respectively. The excitation slit was 0.5 mm. In all cases, 1:1 ratio of fluorophore (DNA or DNA with protein) with protein or competitor was mixed. For the fluorescence intensity measurements, at least 40 traces were collected while for the fluorescence anisotropy, at least 90 traces were collected.

For association pilot experiments, samples were measured for 15 seconds, recording 2000 points and 2 traces to determine the optimal acquisition time. Samples were excited at 495 nm, and emission measured using a 515 nm cut-off filter. For the final experiments, I used 10 nM labelled hCRE/hAP-1 in biophysical buffer and this was rapidly mixed at a ratio 1:1 with varying concentrations of GCN4 dimer (50 nM, 100 nM, 150 nM 200 nM and 250 nM). 40 traces were collected for each of them. The fluorescence kinetic traces were fitted from 2 ms to a single-exponential rate equation (Equation 8).

For the dissociation pilot experiments, samples were measured for 30 seconds, recording 2000 points and 2 traces to determine the optimal acquisition time. Samples were excited at 495 nm, and emission measured using a 519 nm cut-off filter. A 2 ml syringe concentration of 10 nM labelled hCRE or hAP-1 in biophysical buffer premixed with ~50 nM of GCN4 bZIP or FD dimer was mixed with 1 μM of unlabelled competitor. In all of the cases, the linCRE was used as a competitor.

For the final experiments, five different concentrations of competitor were used; 1 μM, 2 μM, 3 μM, 4 μM and 5 μM, and 40 traces were collected for each of them. The temperature used was 25 °C. The fluorescence kinetic traces were fitted from 2 ms to a single-exponential rate equation (Equation 8).

The fluorescence intensity kinetic data were obtained using the stopped-flow machine (SX20 by Applied Photophysics) based in the Shammas group. In the case where fluorescence anisotropy measurements were collected, the stopped-flow machine (SX20 by Applied Photophysics) located in the biophysical suite at the Department of Biochemistry, University of Oxford, was used.

3.2.4 Binding interaction equations

A basic binding interaction can be described by the following equilibrium reaction:

$$P + L \rightleftharpoons PL \quad (2)$$

Where P and L are the two interacting species (in our case protein and DNA respectively) in solution, which for the bound product PL. The equilibrium association constant defines its reaction quotient at chemical equilibrium and is described by the following equation:

$$K_a = \frac{[PL]}{[P][L]} \tag{3}$$

Where K_α is the association constant with unit (M^{-1}) and [P], [L] and [PL] are the concentrations of protein, DNA and protein-DNA complex with units (M). The time-dependent rate equations for the formation and the decay of product [PL] are:

$$\frac{d[PL]}{dt} = k_{\text{on}}\,[P][L] \tag{4}$$

$$\frac{d[PL]}{dt} = -\,k_{\text{off}}\,[PL] \tag{5}$$

with the forward reaction rate constant k_{on} (on-rate) and the reverse reaction rate constant k_{off} with units ($M^{-1}\ s^{-1}$) and (s^{-1}), respectively. In equilibrium, the sum of all-time derivatives is equal to zero, which leads to the equation:

$$K_a = \frac{1}{K_d} = \frac{k_{\text{on}}}{k_{\text{off}}} = \frac{[PL]}{[P][L]} \tag{6}$$

Where K_d is the dissociation constant with units M.

The equations used for the data fitting regarding the equilibrium and kinetics are described in the Methods in Enzymology section, by M. Crabtree and S. Shammas. (Crabtree & Shammas, 2018)

Equilibrium binding curve fitting equations for one-site or two-sites reactions

$$S = S_f + \sum_i \Delta S_i \left(\frac{K_{di}+[L]+[P]-\sqrt{((K_{di}+[L]+[P])^2-4[L][P])}}{2[L]}\right) \tag{7}$$

Where i=1 for one-site binding reaction and i=2 for two-sites binding reaction. S is the anisotropy/intensity signal, S_f is the final signal of the DNA, ΔS is the difference in signal change between free and bound DNA, and [L] and [P] are the total concentration of reactants L and P, respectively.

Exponential decay functions for pseudo-first reaction conditions

$$[S] = S_{final} + \sum_i \Delta S_i e^{-k_{obs(i)}\, t} \tag{8}$$

Where S is the anisotropy/intensity signal, S_f is the final signal of the DNA, ΔS is the difference in signal change between free and bound DNA, and [L] and [P] are the total concentration of reactants L and P, respectively.

Exponential decay functions for equimolar or near-equimolar reaction conditions

$$F = F_0 + \Delta F \frac{(b-z)(1-e^{(zk_{on}t)})}{2\left(\left(\frac{b-z}{b+z}\right)e^{(zk_{on}t)}-1\right)} \tag{9}$$

where

$$b = -(K_d + (1+x)[L]) \tag{9a}$$

and

$$z = \sqrt{K_d^2 + 2(1-x)K_d[L] + (x^2 - 2x + 1)[L]^2} \tag{9b}$$

and

$$x = \frac{[P]}{[L]} \tag{9c}$$

Where F is the anisotropy/intensity signal, F_0 is the signal of the bound DNA, ΔF is the difference in signal change between free and bound, k_{on} is the association rate constant, K_d is the equilibrium dissociation constant, and [L] and [P] are the total concentration of reactants L (DNA) and P (protein), respectively.

3.3 Results

3.3.1 Equilibrium binding constant (K_d) of GCN4 bZIP and FD

The stability of GCN4 bZIP or FD complexes were determined using an equilibrium fluorescence-based method (3.2.2 and 3.2.3). Briefly, a low concentration (1-50 nM) of A488-DNA oligo was incubated with varying concentrations of the GCN4 protein, and the fluorescence intensity/anisotropy determined for each sample. The fluorescence parameters are plotted as a function of the protein concentration to generate an equilibrium curve, which can be fitted to extract an equilibrium dissociation constant (K_d) for the process.

I performed pilot experiments (Figure 3.2) with GCN4 bZIP in incubated with linCRE or NS to obtain a rough profile of GCN4's behaviour with these DNA constructs and determine a protein concentration range to use in more accurate studies.

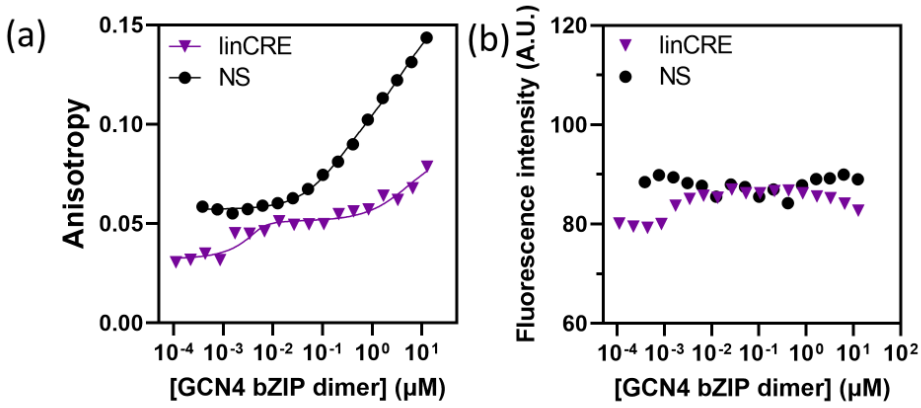


Figure 3.2. Pilot experiments: Fluorescence anisotropy (a) and intensity (b) equilibrium data of GCN4 bZIP binding to linCRE and NS.
Coloured in magenta: addition of 5 nM linCRE^{AF488} into different concentrations of GCN4 bZIP dimer ranging from 0.1 nM to 10 μM. Coloured in black: addition of 5 nM NS^{AF488} into different concentrations of GCN4 bZIP dimer ranging from 0.5 nM to 10 μM. All experiments were performed at 25 °C. Data were fitted to two-site binding equation where the lower and the higher K_d values were determined as following: 4.7 ± 2.3 nM and 25 ± 8 μM for linCRE and 100 ± 30 nM and 150 ± 45 μM for NS. (Errors represent the standard deviations (SD) of the best-fit parameters.)

The equilibrium curve for binding of GCN4 bZIP to linCRE (purple triangles) showed two transitions. The fluorescence intensity change for both appears very small, however the fluorescence anisotropy showed increases over similar concentration scales suggesting binding had occurred, and this technique is less prone to signal/noise issues due to differences in probe concentration (which can occur in such experiments due to probe molecules sticking to surfaces such as tips and Eppendorf tubes, as well as natural pipetting variation). The data were fit to Equation 7, which allows for two binding events. The lower K_d estimate showed an approximate K_d of bZIP binding to linCRE of 4.7 ± 2.3 nM. This is around the value of the probe concentration, which can lead to difficulties in obtaining accurate K_d values, so in subsequent experiments with DNA containing target sites, I decreased the probe concentration as much as possible whilst maintaining reasonable signal (to 1 nM). The higher K_d estimate is in the hundreds of nM but since a plateau is not reached this value is very approximate. Only one clear transition was observed for binding of GCN4 bZIP to NS, and the K_d of bZIP binding to NS was estimated around 100 ± 30 nM.

Next, I attempted to determine the K_d for fluorescently labelled linCRE more accurately (Figure 3.3). Fluorescence anisotropy again showed two transitions where

anisotropy was increased, presumably representing binding events, and FD version appeared to bind more tightly than the bZIP. However, there was not a reliably detectable change in the fluorescence intensity for the first transition as observed in Figure 3.3(b). Although binding could be confirmed using fluorescence anisotropy the general requirement for higher probe concentrations with this technique posed difficulties in accurately measuring the very low K_d values. Lack of a significant fluorescence intensity change upon binding would also lead to significant difficulties obtaining kinetic rate constants using stopped-flow spectroscopy. Therefore, I investigated alternative DNA constructs. I was able to obtain a fluorescence intensity change (and more reliable K_d estimates) using AlexaFluor®488-labeleld DNA hairpin versions of the target sites which were already available in the lab (Figure 3.4).

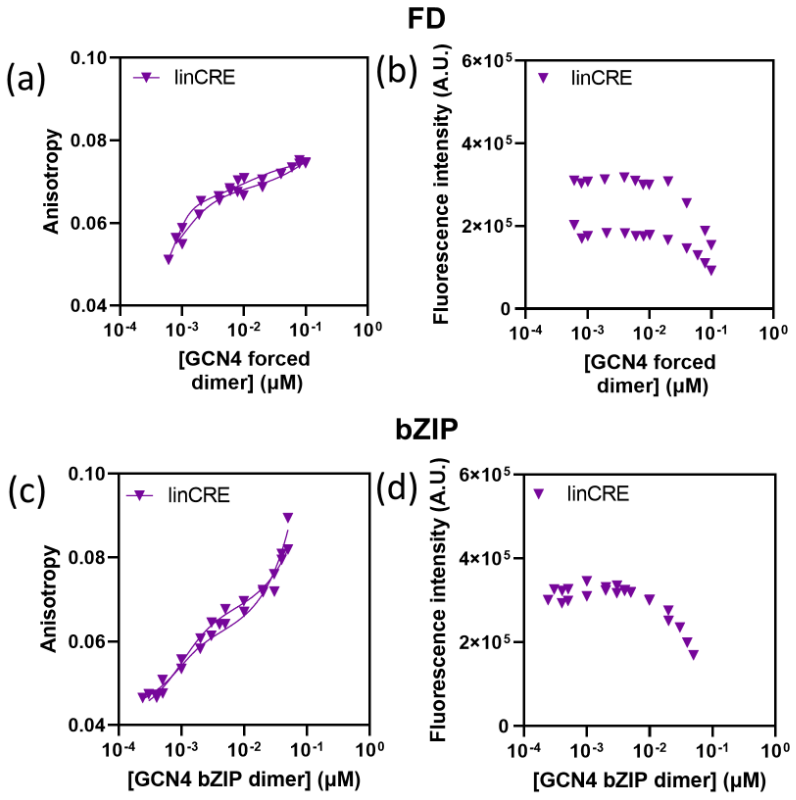


Figure 3.3. Pilot experiments: Fluorescence anisotropy (a)(c) and intensity (b)(d) equilibrium data of GCN4 FD (a)(b) or bZIP (c)(d) binding to linCRE.
Addition of 1 nM linCRE^{AF488} into different concentrations of GCN4 dimer ranging from 0.1 nM to 100 nM. Experiments performed in duplicate. Data were fitted with Equation 7 where the lower and the higher dissociation constants were determined as following: 0.5 ± 0.4 nM and 1.7 ± 0.5 μM for the bZIP and 0.4 ± 0.3 nM and 200 ± 35 nM for the FD protein. (Errors represent the standard deviations (SD) of the best-fit parameters.)

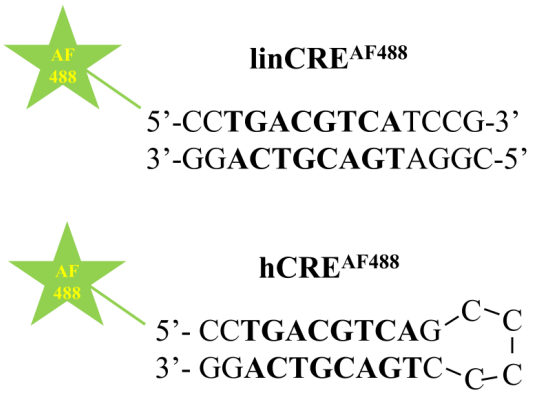


Figure 3.4. Illustration of the linear and hairpin version of CRE site.
The CRE target sequence is shown in bold. The AlexaFluor-488 fluorescent probe (AF488) is demonstrated as green asterisks attached to the 5' of the oligonucleotide.

Fluorescence anisotropy equilibrium measurements were made with GCN4 bZIP and FD to confirm their interaction with labelled hCRE, hAP-1 and NS (Figure 3.5). In both cases the fluorescence anisotropy increased with protein concentration, and I was able to fit the equilibrium curves using Equation 7 to extract K_d estimates (Table 3.1). I have focussed on the lower K_d (higher affinity) binding event because these are likely to represent binding to the specific site. GCN4 bZIP and FD bind to hCRE and hAP-1 with similar affinities at the nM range, with the FD apparently having 2-fold higher affinity to the target sites compared to the wt bZIP.

I examined the specificity of the binding by performing similar binding experiments using a DNA construct without either of the previously described target sites (NS). Again GCN4-bZIP and GCN4-FD bind to the non-target DNA (NS) with similar affinities, in this case K_d is hundreds of nM, with bZIP dimer showing a slightly higher affinity to NS compared to the FD. Thus, an approximate 50-fold difference in K_d was observed between the target sites and the non-target site. Interestingly the final anisotropy is also much higher than with the target, suggesting that more than one protein may bind.

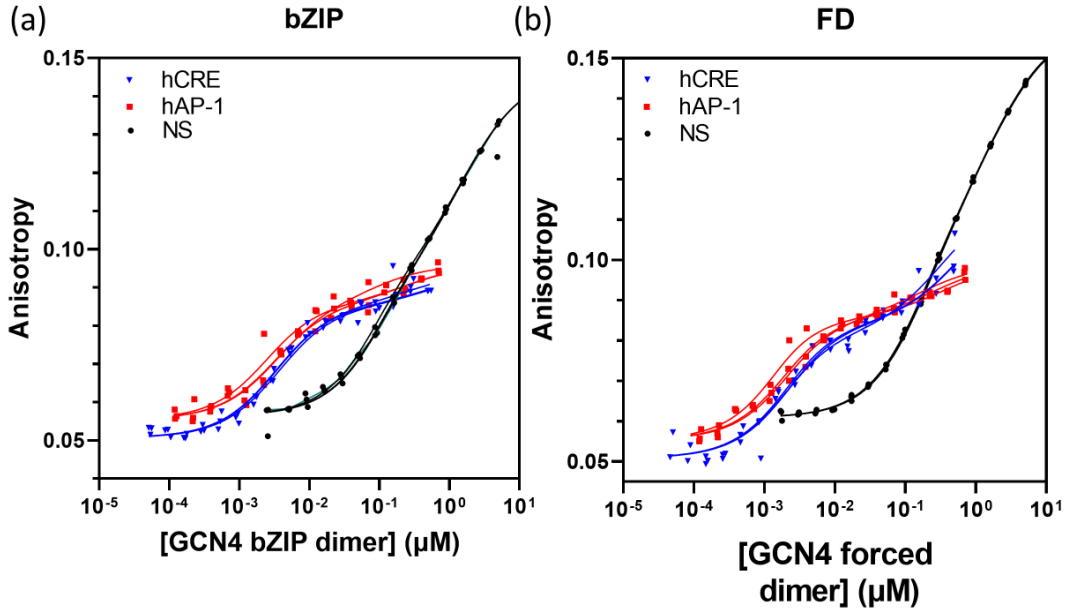


Figure 3.5. Fluorescence anisotropy equilibrium data of GCN4 bZIP and FD binding. (a) Coloured in red: addition of 1 nM hCRE^{AF488} into different concentrations of GCN4 bZIP dimer ranging from 0.1 nM to 500 nM. Coloured in blue: addition of 1 nM hAP-1^{AF488} into different concentrations of GCN4 bZIP dimer ranging from 0.1 nM to 500 nM. Coloured in black: addition of 5 nM NS^{AF488} into different concentrations of GCN4 bZIP dimer ranging from 5 nM to 5 μM. (b) Coloured in red: addition of 1 nM hCRE^{AF488} into different concentrations of GCN4 FD dimer ranging from 0.1 nM to 500 nM. Coloured in blue: addition of 1 nM hAP-1^{AF488} into different concentrations of GCN4 FD dimer ranging from 0.1 nM to 500 nM. Coloured in black: addition of 5 nM NS^{AF488} into different concentrations of GCN4 FD dimer ranging from 5 nM to 5 μM. All experiments were performed in triplicate at 25 °C.

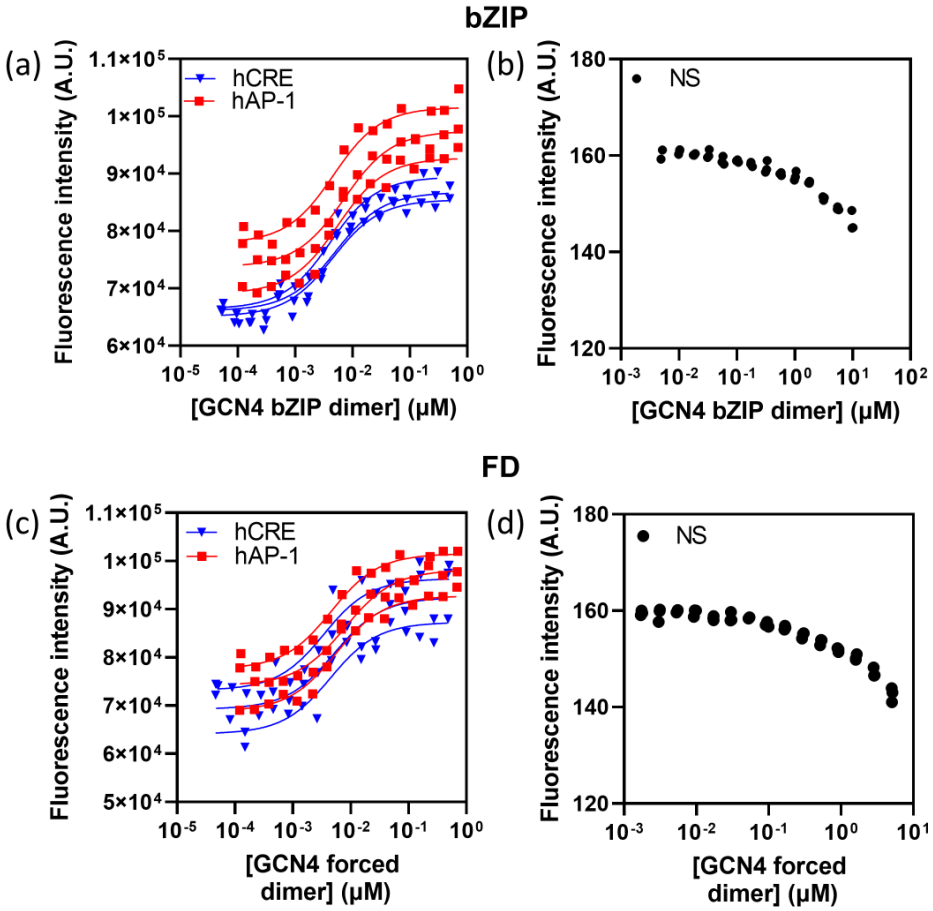


Figure 3.6. Fluorescence intensity equilibrium data of GCN4 bZIP and FD binding. For the GCN4 bZIP: (a) Coloured in red: addition of 1 nM hCRE^{AF488} into different concentrations of GCN4 bZIP dimer ranging from 0.1 nM to 500 nM. Coloured in blue: addition of 1 nM hAP-1^{AF488} into different concentrations of GCN4 bZIP dimer ranging from 0.1 nM to 500 nM. (b) Coloured in black: addition of 5 nM NS^{AF488} into different concentrations of GCN4 bZIP dimer ranging from 5 nM to 5 μM.
For the GCN4 FD: (c) Coloured in red: addition of 1 nM hCRE^{AF488} into different concentrations of GCN4 FD dimer ranging from 0.1 nM to 500 nM. Coloured in blue: addition of 1 nM hAP-1^{AF488} into different concentrations of GCN4 FD dimer ranging from 0.1 nM to 500 nM. Coloured in black: addition of 5 nM NS^{AF488} into different concentrations of GCN4 FD dimer ranging from 5 nM to 5 μM. All experiments were performed in triplicate at 25 °C. K_d constants that could be extracted from these equilibrium curves are shown in Table 3.3.

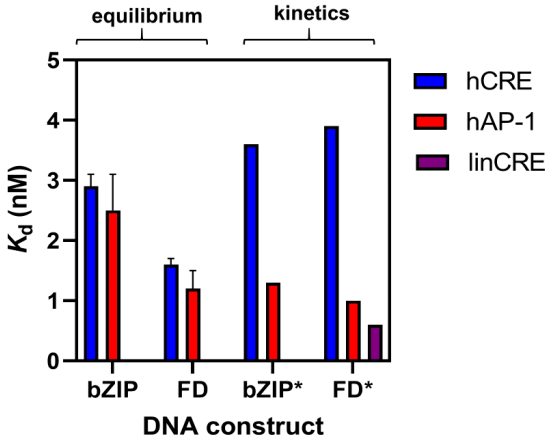


Figure 3.7. Dissociation constants determined for the full-length bZIP constructs from various methods.
All constants were obtained by equilibrium fluorescence anisotropy experiments except from the ones marked with an asterisk where: hCRE and hAP-1 related constants were obtained from fluorescence intensity kinetics and linCRE from fluorescence anisotropy kinetics. Error bars represent SD from multiple measurements.

Table 3.1. Dissociation constants (K_d) for target (hCRE and hAP-1) and non-target DNA (NS) in complex with GCN4 bZIP, FD or BR from fluorescence anisotropy experiments.
The thermodynamic parameters were obtained at 25 °C. (N.D. = Not Determined). The error for the average values represents the standard deviation (SD) calculated from the three replicates of each experiment. Data fitted using either a single-site[†] or a two-site[‡] binding fitting equation. Here the first transition K_d (K_{d1}) is reported. (Equation 8). (*) Calculated from kinetics experiments.

Protein Construct	**hCRE^{AF488}**	**linCRE ^{AF488}**	**hAP-1 ^{AF488}**	**NS ^{AF488}**
GCN4 BR	86 ± 1 μM[†]	32 ± 9 μM[†]	100 ± 1 μM[†]	174 ± 7 μM[†]
GCN4 bZIP	2.9 ± 0.2 nM[‡]	N.D.	2.5 ± 0.6 nM[‡]	87 ± 10 nM[‡]
GCN4 FD	1.6 ± 0.1 nM[‡]	0.6 ± 0.1 nM*	1.2 ± 0.3 nM[‡]	127 ± 11 nM[‡]

Table 3.2. Dissociation constants (K_{d2}) for target (hCRE and hAP-1) and non-target DNA (NS) in complex with GCN4 bZIP, FD or BR.

The thermodynamic parameters were obtained at 25 °C. The error for the average values represents the standard deviation (SD) calculated from the three replicates of each experiment. Data fitted using either a single-site[†] or a two-site[‡] binding fitting equation. Here the second transition K_d (K_{d2}) is reported. (Equation 8)

Protein Construct	hCRE^{AF488}	linCRE ^{AF488}	hAP-1 ^{AF488}	NS ^{AF488}
GCN4 BR	-	-	-	-
GCN4 bZIP	460 ± 140 nM[‡]	-	430 ± 120 nM[‡]	1.70 ± 0.01 mM[‡]
GCN4 FD	510 ± 150 nM[‡]	-	560 ± 160 nM[‡]	1.6 ± 0.1 mM[‡]

Table 3.3. Dissociation constants (K_d) for target (hCRE and hAP-1) in complex with GCN4 bZIP or FD from fluorescence intensity equilibrium experiments.

The thermodynamic parameters were obtained at 25 °C. Errors represent standard deviation (SD) from multiple replicates. (N.D.=Not Determined)

Protein Construct	hCRE^{AF488}	hAP-1 ^{AF488}
GCN4 bZIP	3.4 ± 0.3 nM	5.2 ± 1.2 nM
GCN4 FD	3.9 ± 0.8 nM	5.7 ± 0.9 nM

3.3.2 Binding association kinetics of GCN4 with its targets

Having already established that fluorescence intensity of A488-DNA could be used as a probe of GCN4 binding in equilibrium studies, I proceeded to use fluorescence stopped-flow techniques to assess the kinetic rate constants for the interactions. I performed mixing experiments to monitor the progress of association process with time. Solutions of 10 nM AlexaFluor-488®-hCRE or hAP-1 in biophysical buffer were mixed rapidly in a 1:1 ratio with ~50 nM of GCN4 dimer. I performed pilot experiments to determine the approximate timescale and intensity change observed for the process and found these to be suitable for the stopped-flow technique (data not shown), before embarking upon stopped-flow fluorimetry experiments (Materials and Methods paragraph 3.2.3),

The kinetic traces were well fit by a monophasic exponential decay in fluorescence intensity (Figure 3.9). The extracted apparent rate constants (k_{app}) appeared to depend linearly on the protein concentration (Figure 3.8). The gradient of a straight line fit in such cases can be used to extract an association rate constant (k_{on}). Remarkably k_{on} was almost identical for both GCN4 bZIP and GCN4 FD with both hAP-1 (2.3 ± 0.1 nM^{-1} s^{-1} and 2.4 ± 0.1 nM^{-1} s^{-1}, respectively) and hCRE DNA (2.2 ± 0.2 nM^{-1} s^{-1} and 2.7 ± 0.1 nM^{-1} s^{-1}, respectively), being 2-3 nM^{-1} s^{-1} in all cases.

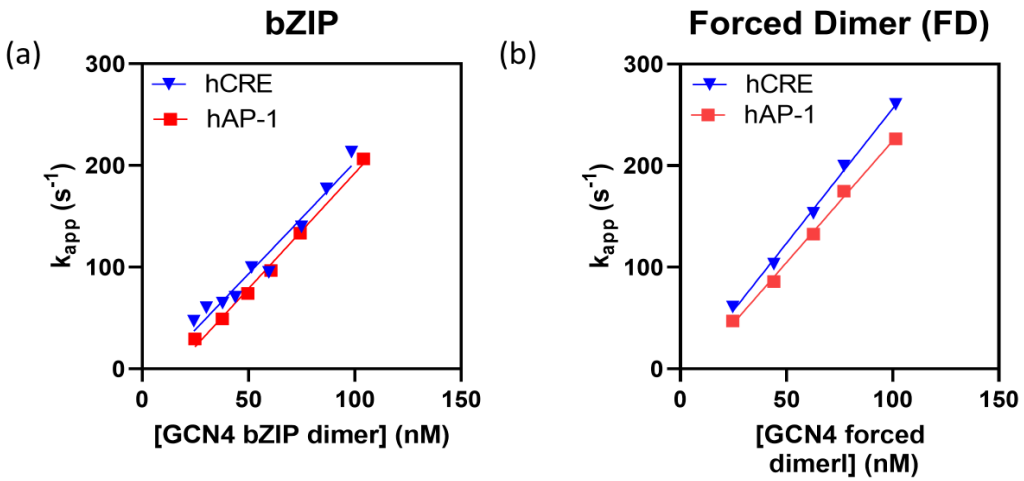


Figure 3.8. Observed kinetic rate constants of GCN4 bZIP or FD binding to hCRE or hAP-1.

(a) GCN4 bZIP and (b) GCN4 FD in complex with hCRE (blue) or hAP-1 (red). All traces were fitted to a single exponential decay to extract k_{app}. Lines represent straight line of best fit, with gradient indicating k_{on}.

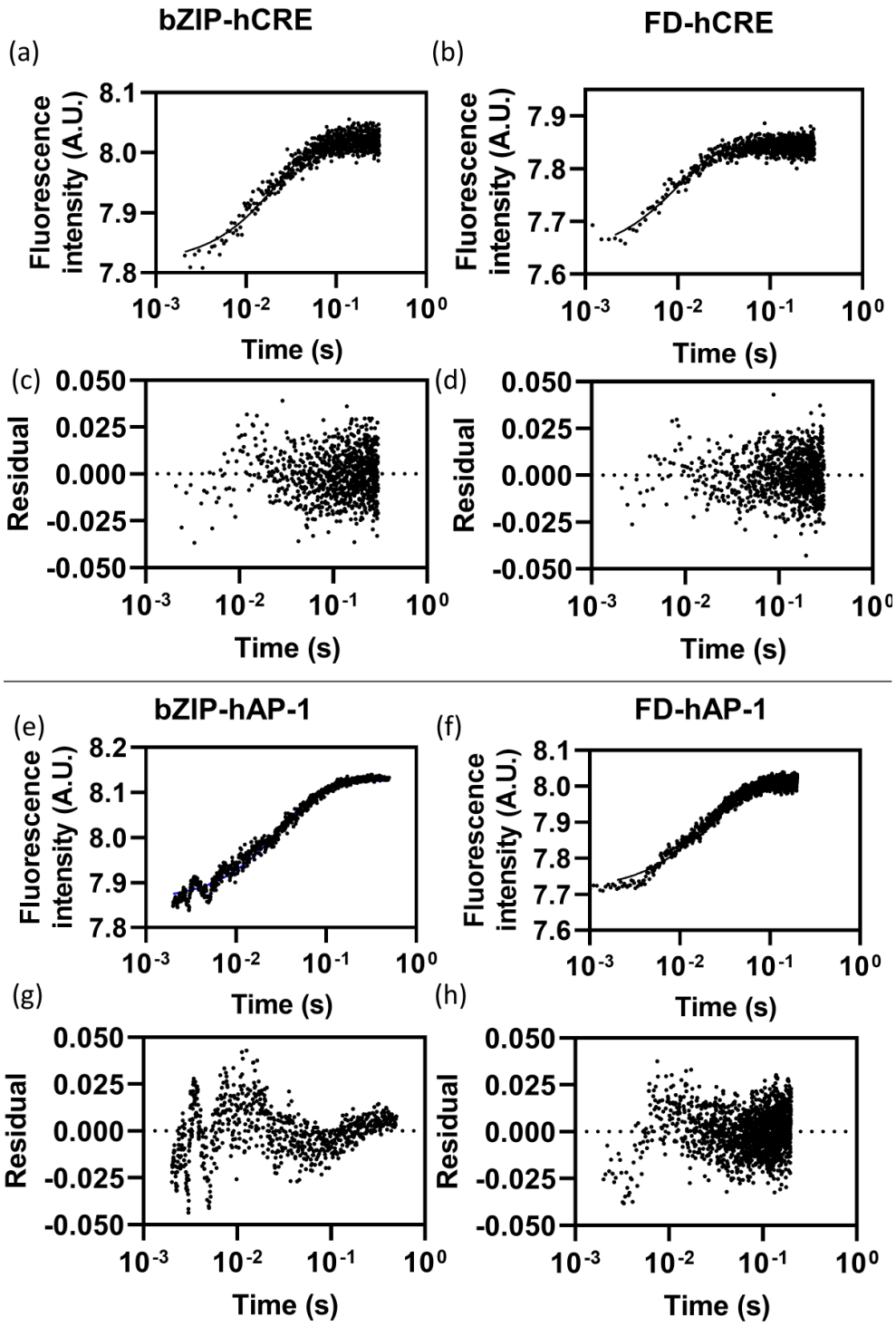


Figure 3.9. Representative traces for association experiments using fluorescence stopped-flow.

(a) and (b): Intensity trace from the reaction between 10 nM hCRE and 50 nM GCN4 dimer protein with ration 1:1, (c) and (d): Residual plot from the fitting of hCRE and GCN4 dimer protein dataset from (a) and (b) respectively. (e) and (f): Intensity trace from the reaction

between 10 nM hAP-1 and 50 nM GCN4 dimer protein with ration 1:1, (g) and (h): Residual plot from the fitting of hAP-1 and GCN4 dimer protein dataset from (e) and (f) respectively.

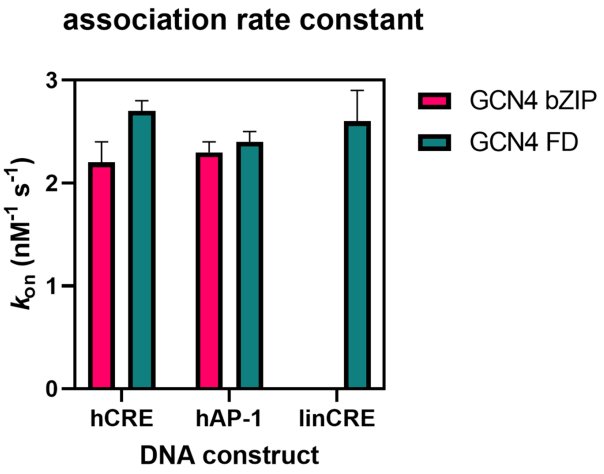


Figure 3.10. Association rate constants determined from fluorescence-based kinetics. hCRE and hAP-1 were investigated using fluorescence intensity stopped-flow whereas linCRE used fluorescence anisotropy stopped-flow. Error bars represent SD of best-fit parameters.

3.3.3 Dissociation of GCN4 from its target (k_{off})

Traditional competition experiments were used to assess the dissociation rate constant (k_{off}) for the interaction. Preformed complexes of GCN4-bZIP with A488-labelled target DNA (10 nM A488-hAP1/h-CRE and 50 nM protein) were rapidly mixed with large excesses of unlabelled linCRE DNA (1 to 5 μM) with a ratio 1:1. As a result of the GCN4-competitor complex formation, fluorescent DNA was freed from GCN4 and thus the dissociation process was observed as a decrease in fluorescence intensity. In the presence of a large enough excess of unlabelled competitor the observed apparent rate constant (k_{app}) in such experiments is equal to the dissociation rate constant (k_{off}), and independent of the competitor concentration (Figure 3.12 and Figure 3.13).

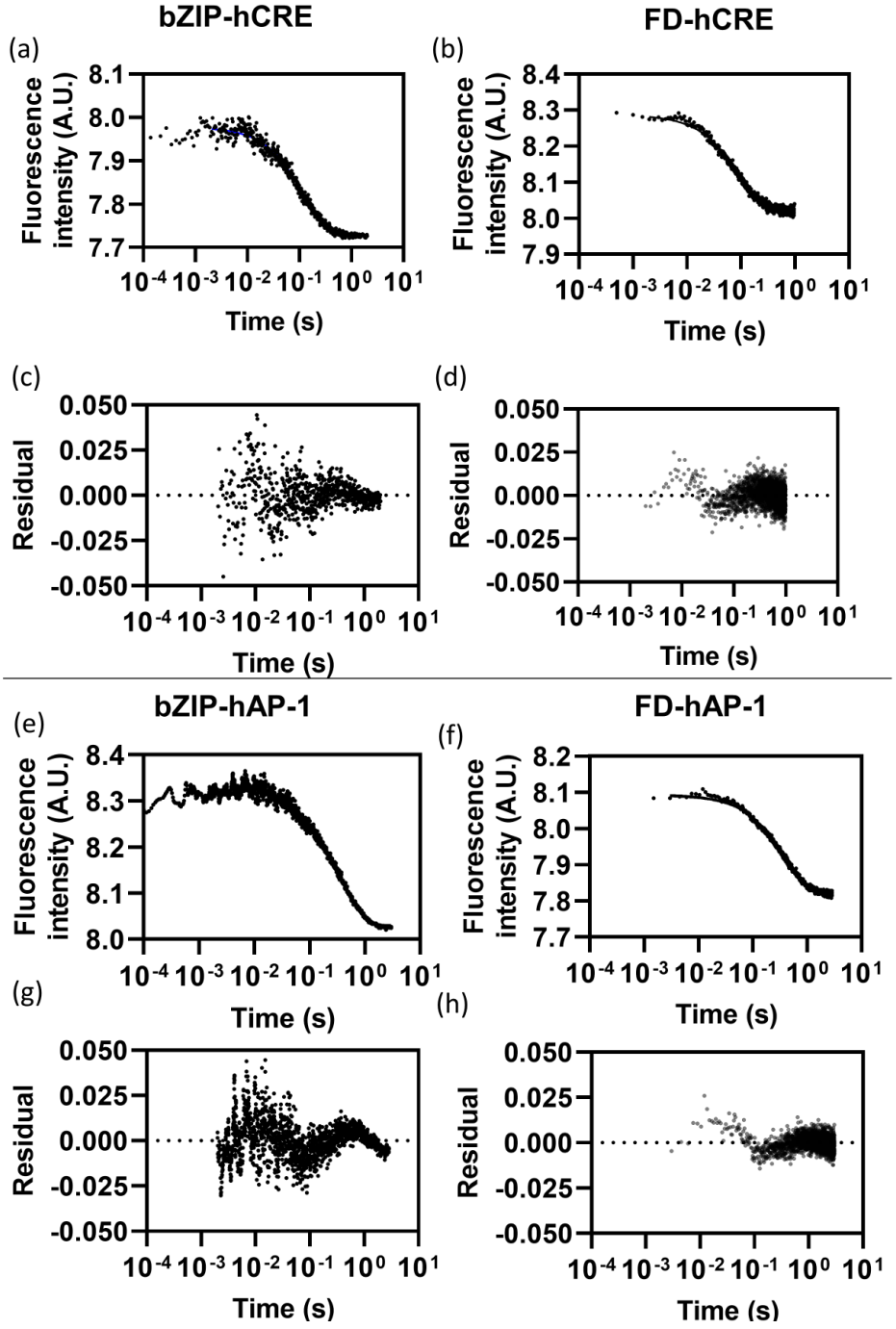


Figure 3.11. Representative traces for dissociation experiments using fluorescence intensity stopped-flow.

(a) and (b): Intensity trace from the reaction between 10 nM hCRE - 50 nM GCN4 dimer protein and 1 μM competitor with ration 1:1, (c) and (d): Residual plot from the fitting of hCRE and GCN4 dimer protein dataset from (a) and (b) respectively. (e) and (f): Intensity trace from the reaction between 10 nM hAP-1 - 50 nM GCN4 dimer protein and 1 μM competitor with ration 1:1, (g) and (h): Residual plot from the fitting of hAP-1 and GCN4 dimer protein dataset from (e) and (f) respectively.

The kinetic traces showed a monophasic exponential decay in fluorescence intensity (Figure 3.11). The observed rate constants were independent of the competitor concentration, indicating that the dissociation reaction was essentially irreversible under my conditions. The k_{off} was therefore obtained by averaging for the five competitor concentrations. Values for k_{off} depended upon the DNA construct – being higher for hCRE (8.1 ± 0.3 s^{-1} for bZIP and 10.4 ± 0.2 s^{-1} for FD) than for hAP1 (2.9 ± 0.1 s^{-1} for bZIP and 2.5 ± 0.1 s^{-1} for FD) but were very similar for bZIP and FD. For a simple 2-state reaction the equilibrium dissociation constant K_d can be calculated using the ratio of the dissociation and association rate constants, Equation 6. Comparison of the K_d calculated from kinetics to that previously obtained by equilibrium methods shows that the values observed are similar, but not necessarily the same within error.

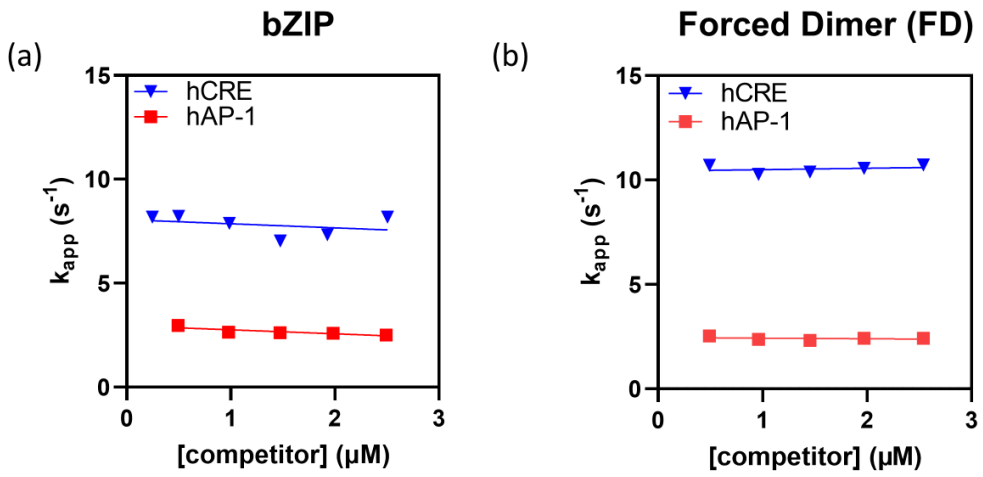


Figure 3.12. k_{off} determination: Fluorescence intensity kinetic data for the dissociation of the complexes GCN4 bZIP or FD binding to hCRE or hAP-1 using the stopped-flow machine.

k_{off} measurements for the (a) GCN4 bZIP and (b) GCN4 FD in complex with hCRE (coloured in blue) or hAP-1 (coloured in red). All traces were fitted to a single exponential decay. All experiments were performed at 25 °C.

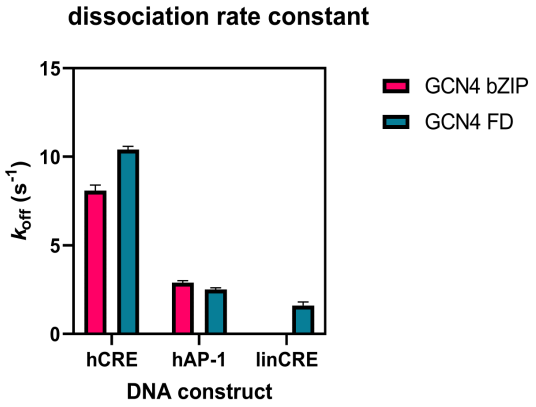


Figure 3.13. Dissociation rate constants determined from fluorescence-based kinetics. hCRE and hAP-1 measurements were determined using fluorescence intensity, whereas linCRE were determined using fluorescence anisotropy as a probe. Error bars represent SD of best-fit parameters.

Table 3.4. Association and dissociation rate constants for the GCN4 bZIP and forced dimer obtained by fluorescence anisotropy or intensity measurements.
The thermodynamic parameters were obtained at 25 °C. Errors represent the standard deviations (SD) of the best-fit parameters. (N.D.=Not Determined)

	GCN4 bZIP		***GCN4 FD***	
	k_{on} (nM^{-1} s^{-1})	k_{off} (s^{-1})	k_{on} (nM^{-1} s^{-1})	k_{off} (s^{-1})
hCRE	2.2 ± 0.2	8.1 ± 0.3	2.7 ± 0.1	10.4 ± 0.2
hAP-1	2.3 ± 0.1	2.9 ± 0.1	2.4 ± 0.1	2.5 ± 0.1
linCRE	N.D.	N.D.	2.6 ± 0.3	1.6 ± 0.2

3.3.4 Kinetics of binding reaction for GCN4 with linear CRE DNA

The previous fluorescence anisotropy equilibrium binding experiments did not allow the determination of the K_d of GCN4 bZIP binding to linCRE (See paragraph 3.3). Thus, I needed to find a different way to accurately calculate this thermodynamic parameter to compare it with hCRE results. For this purpose, I employed fluorescence anisotropy-based kinetics. A similar protocol to the one described in paragraph 3.3.2 was followed.

Pilot experiments showed that in order to monitor the process a probe concentration of at least 50 nM of fluorescently labelled-linCRE was required for

anisotropy stopped-flow experiments. For dissociation experiments, a preformed complex of 50 nM of labelled linCRE and 50 nM of bZIP dimer was rapidly diluted with a high concentration of competitor (2 μM, 4 μM and 6 μM unlabelled linCRE). The kinetic traces were well fit with a single exponential decay function (equation 8, Figure 3.16) and since the derived k_{app} were independent of competitor concentration, they were averaged to calculate k_{off} (Figure 3.14).

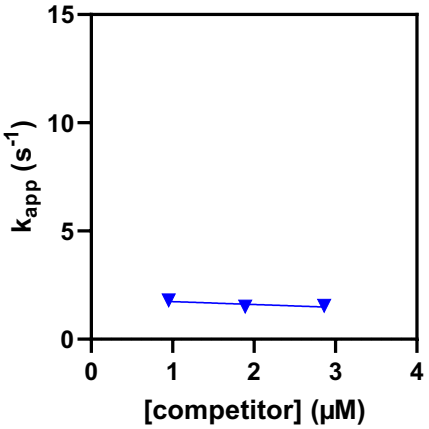


Figure 3.14. Fluorescence anisotropy kinetic data of GCN4 FD binding using stopped-flow.
k_{off} measurements for the GCN4 bZIP in complex with hCRE (coloured in blue) using unlabelled linCRE as a competitor. All traces were fitted to a single exponential decay and the derived k_{app} values were averaged to obtain final $\boldsymbol{k_{off}}$. All experiments were performed at 25 °C.

In the previous intensity-based association experiments, pseudo-first order reaction conditions were maintained since the concentration of protein was at least 10-fold higher than the labelled-DNA. However, in this case to maintain pseudo-first order reaction conditions required more than 1 mM of GCN4 bZIP dimer, which would mean the reaction was too fast to observe. Thus, for the determination of k_{on} I was forced to use near-equimolar conditions. More precisely, 100 nM of linCRE were mixed with 100 nM-500 nM bZIP dimer, in a ratio 1:1. The traces were fit to the Equation 9, which is suitable for equimolar and near-equimolar conditions, and k_{off} was fixed to the value obtained by the k_{off} experiments described below. Each experiment gave an independent estimate of k_{on} and the weighted average of 2.6 ± 0.3 nM^{-1} s^{-1} was the same within error as the k_{on} rate constant for hCRE and hAP-1. Finally, I was able to calculate the dissociation constant (K_d)

describing GCN4 FD binding to linCRE, using Equation 6. The final K_d value (0.6 ± 0.1 nM) shows a 2-fold/5-fold higher affinity compared to the equilibrium/kinetic estimates for the hCRE construct. All constants measured in this paragraph are shown in Table 3.5.

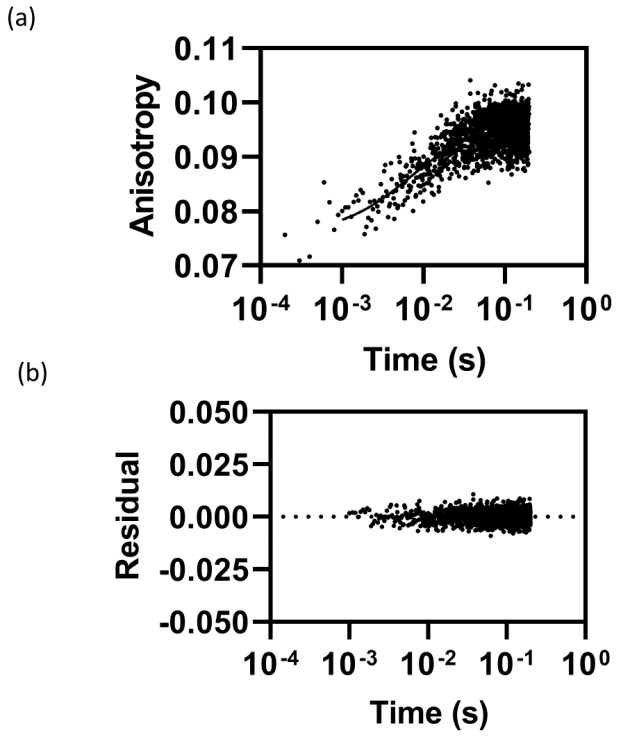


Figure 3.15. Representative traces for association experiments for GCN4 FD binding to linCRE using fluorescence anisotropy stopped-flow.
(a) Anisotropy trace from the reaction between 100 nM hCRE and 100 nM GCN4 FD (mixed with ratio 1:1) and (b) associated residual plot.

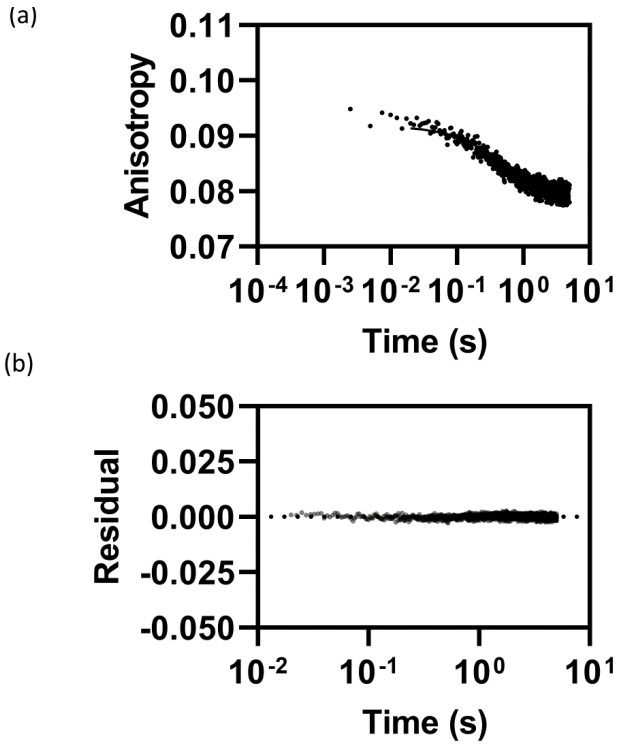


Figure 3.16. Representative traces for dissociation experiments for GCN4 FD binding to linCRE using fluorescence anisotropy stopped-flow.
(a) Anisotropy trace from the reaction between 100 nM hCRE, 100 nM GCN4 FD dimer and 2 μM competitor (mixed with ratio 1:1) and (b) associated residual plot.

Table 3.5. Association and dissociation rate constants for the GCN4 FD-linCRE complex obtained by fluorescence anisotropy measurements.
The thermodynamic parameters were obtained at 25 °C. Errors represent standard deviation (SD) from multiple replicates. (N.D.=Not Determined)

DNA construct	**GCN4 FD**		
	k_{on} (nM^{-1} s^{-1})	k_{off} (s^{-1})	K_d (k_{off} /k_{on}) (nM)
linCRE	2.6 ± 0.3	1.6 ± 0.2	0.6 ± 0.1

3.3.5 DNA binding by monomeric GCN4 basic region

The stability of the monomeric BR-DNA complex was investigated using the same protocol as described in the previous paragraph (3.3) for the GCN4 bZIP and FD constructs. I performed pilot experiments (Figure 3.17) with GCN4 BR in reaction with AlexaFluor®488-labelled linCRE or NS to determine a K_d range. For this purpose, 50 nM of A488-DNA oligo was incubated with varying concentrations of the BR protein (from 0.5 nM to 100 µM). Increases in both fluorescence and anisotropy were detected at the highest protein concentrations, however the concentrations used were not high enough to observe a plateau (Figure 3.17). In addition to this, there was only a very small change in the fluorescence intensity of AlexaFluor®488-NS DNA upon binding (Figure 3.17(b), black points). Initial K_d estimates showed for linCRE and NS around 40 µM and 100 µM, respectively, which are both orders of magnitude higher than those for the dimeric GCN4 constructs.

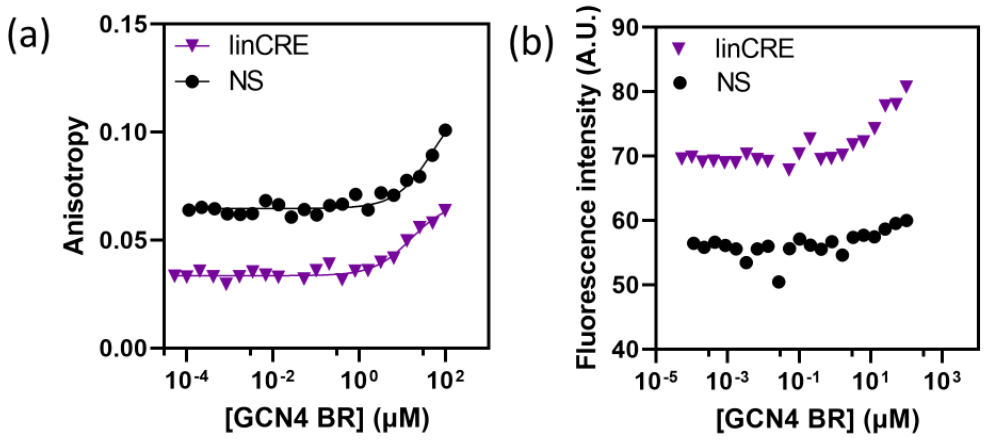


Figure 3.17. Pilot experiments: Fluorescence anisotropy (a) and intensity (b) equilibrium data of GCN4 BR monomer binding to linCRE and NS.
Coloured in magenta: addition of 5 nM linCRE^{AF488} into different concentrations of GCN4 BR monomer ranging from 0.1 nM to 100 µM. Coloured in black: addition of 5 nM NS^{AF488} into different concentrations of GCN4 BR monomer ranging from 0.5 nM to 100 µM. All experiments were performed at 25 °C.

I attempted to obtain more accurate K_d estimates for AlexaFluor®488-labelled hCRE, hAP-1 and NS by preparing even higher concentrations of BR samples (Figure 3.18). Above around 1.5 mM I observed aggregation in my protein solutions so this created a practical limit for my equilibrium curve. I prepared samples with BR concentration up to 1.2 mM. In order to obtain an approximate dissociation constant (K_d), the obtained binding data were fitted into a single-site binding equation (Table 3.1). Binding of BR to hAP1 and hCRE DNA is only around 2-fold tighter than that to non-specific (NS) DNA. In order to rule out a difference caused by the topology of the DNA construct I also performed an equilibrium curve with linCRE (purple triangles). Approximately 5-fold tighter binding is observed for linCRE compared with linear NS DNA of identical length.

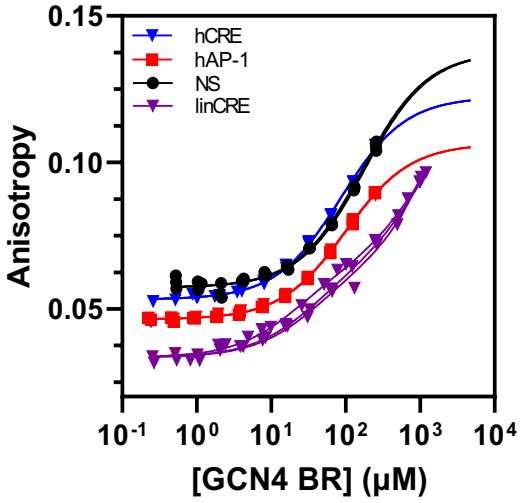


Figure 3.18. Fluorescence anisotropy equilibrium data of GCN4 BR monomer binding.
Coloured in red: addition of 50 nM hCRE^{AF488} into different concentrations of GCN4 bZIP dimer ranging from 0.1 μM to 0.8 mM. Coloured in blue: addition of 50 nM hAP-1^{AF488} into different concentrations of GCN4 BR monomer ranging from 0.1 μM to 1.2 mM. Coloured in black: addition of 50 nM NS^{AF488} into different concentrations of GCN4 BR monomer ranging from 0.1 μM to 0.8 mM. Coloured in magenta: addition of 50 nM linCRE^{AF488} into different concentrations of GCN4 BR monomer ranging from 0.1 μM to 1.2 mM. All experiments were performed in triplicate at 25 °C.

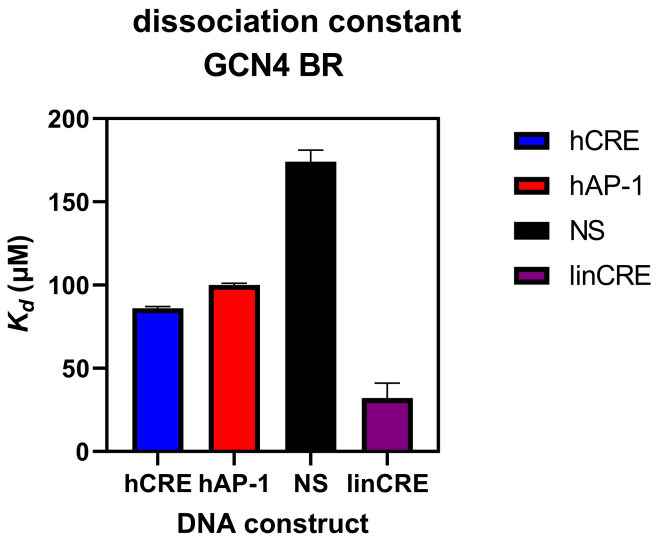


Figure 3.19. Equilibrium dissociation constants determined for the monomer BR.
Values were used from Table 3.1. Error bars represent SD from multiple replicates.

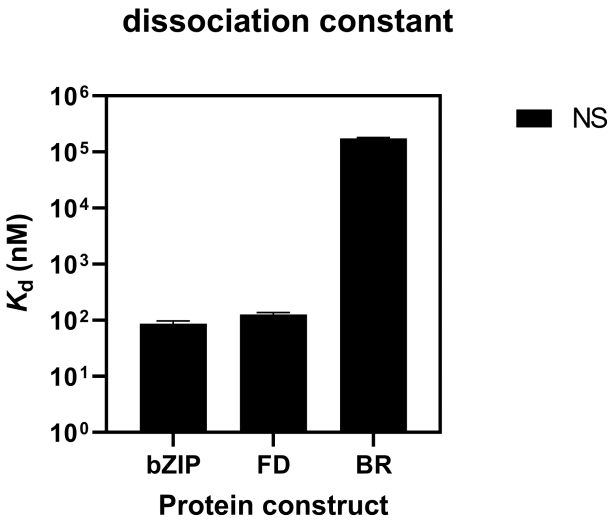


Figure 3.20. Equilibrium dissociation constants determined for GCN4 binding to NS DNA.
Values were used from Table 3.1. Error bars represent SD from multiple replicates.

3.4 Discussion

3.4.1 Wild-type bZIP and forced-dimer GCN4 bind specifically to target DNA with nM affinity

In this study, I performed equilibrium binding fluorescence anisotropy experiments which showed that the bZIP version of GCN4 (wild-type bZIP and FD) showed significantly higher affinities to the target sites compared to the BR monomer version. The two bZIP-length constructs have very similar dissociation constants for the target sites, in the low nM range. The comparison between these two constructs shows a 2-fold higher affinity for the FD than the bZIP protein, which may not be considered significant. If we suppose that GCN4 bZIP physiologically binds as monomer i.e. the monomer pathway is preferred, then the forced-dimerized version of GCN4 bZIP could show a lower affinity to the target sites. This was not observed. Instead, the forced dimer binds slightly more tightly, which could reflect the requirement for dimerization either before or after DNA binding. Nevertheless, the results demonstrated that the forced-dimerization of the GCN4 bZIP did not impede its binding to the target site (in the absence of non-target DNA). It is difficult to conclude the pathway from my result as it is consistent with both.

Specificity refers to the ability of GCN4 to recognise and bind to one DNA site in preference to other DNA sites. In order to introduce a definition of specificity in this study, I needed one DNA construct that did not include a similar nucleotide sequence to the ones in the widely known cognate sites. For this purpose, I used the 5'-GGCTAAAGCATTCT-3' sequence, that does not include either the full or the half-target sites. The full-length dimers showed a ~55-fold specificity to AP-1 and CRE than NS, the dissociation constant corresponding to the dimer-NS complex being calculated in hundreds of nM range (~100 nM). This is a surprisingly small difference, given the vast excess of non-specific DNA in the context of the nucleus.

A previous study by Dragan et al. (2004), used fluorescence anisotropy equilibrium assays and ITC to measure force-dimerized GCN4 bZIP affinity to AP-1 and the CRE in 30 mM sodium phosphate buffer (pH 7.4) containing 100 mM NaCl (ionic strength 160

mM). The AP-1 and the CRE complex showed K_d values of 1.5 nM and 6.4 nM, respectively. These values are in close agreement with my results considering that experimental conditions were not identical since in this work, the ionic strength was 185 mM and the pH was 6.5. The thermodynamic binding of GCN4 bZIP has also been investigated in older studies with different methods such as electrophoretic mobility shift assays, where GCN4 bZIP binds to target sites with a K_d approximately 20-50 nM (Hollenbeck & Oakley, 2000) and ~20 nM (Weiss et al., 1990). These values are 10-fold higher than the most recent results described in the present work and by (Anatoly I. Dragan et al., 2004). The accurate determination of very low binding constants is a problem since many of the methods used for studies of protein–DNA interactions (e.g. calorimetry, CD) are not sensitive enough, while gel-shift assays are not always suitable for quantitative thermodynamic analysis. The fluorescence anisotropy titration has proved efficient since it can give well-resolved and accurate binding curves using the very low concentration of reagents (e.g. <10 nM of fluorescently labelled DNA) which are required for the determination of high affinity binding constants. However, in my work I was still working very close to the limit for this technique, with K_d values of around 1 nM. In the future a shift to a technique with lower fluorophore concentrations, such as fluorescence correlation spectroscopy (FCS) could be considered.

3.4.2 Monomeric GCN4 binds less tightly to target DNA than dimeric GCN4

The BR monomer construct lacks the leucine zipper domain, making GCN4 unable to dimerize, however the positively charged residues consisting of the DNA binding site which also restrict dimerization (due to high electrostatic repulsive forces) are maintained, so the protein can interact with target DNA. Yet, the equilibrium binding results for the BR-target-DNA showed significantly higher dissociation constants (around 80-100 μM), compared with the full-bZIP-length versions. This is significantly higher than the likely protein concentration within the nucleus, suggesting that monomer-target complexes will be essentially unpopulated. In contrast Cranz et al. (2004) found GCN4 BR and CRE interact with a ~ 0.4 μM K_d. It is important to note that different techniques, experimental conditions and biophysical buffers have a significant impact on the thermodynamic results. Cranz et al. (2004) use less biologically relevant conditions - a lower temperature (6.5 °C)

and ionic strength (68 mM), which might both reasonably be expected to lead to significantly tighter binding. They use Tris HCl (pH 7.8) in combination with $MgCl_2$ and not phosphate buffer similar to this study. The Mg^{2+}, or other divalent ions, might also affect complex stability, although no divalent ions are present in any of the GCN4 crystal structures in the PDB.

Previous works have shown that forced-dimerized BR lacking the leucine-zipper binds successfully to DNA (K_d~58 nM) (Cranz et al., 2004) so it may be that dimerization, rather than the leucine-zipper per se, is required for tight binding. This fact is in agreement with previous studies where cross-linked BR peptides were capable to bind DNA with high affinities (Cao et al., 2000; Hollenbeck et al., 2009; Talanian et al., 1990). In these studies, higher affinities were demonstrated for the binding of a cross-linked BR compared to the wild-type monomer BR peptide.

One major similarity between the results of this study and the results found by the Shammas group for CREB, is the K_d for DNA binding of the monomeric BR. In the case of CREB, the monomeric BR peptide showed a K_d approximately 200 μM for target-DNA binding, which is significantly higher than the K_d value of the full-length bZIP of ~0.2 nM, and instead similar to that for a control DNA sequence without the target site. The binding affinity found for the BR-target-DNA complex here is only approximately 2-fold higher than the equivalent for CREB BR monomer, and again it is fairly similar to that for non-specific DNA, displaying 5-fold tighter binding to target CRE over non-specific DNA of an equivalent length. It binds only weakly - in the tens-hundreds μM range - despite the high affinity of the full-length bZIP version. If binding to target and non-target DNA by monomer is relatively similar in strength, then this reduces strongly the chance of a monomer being already bound at the target during the limited time that another investigates the target and makes the monomer search pathway more unlikely. In the case of CREB bZIP it is this lack of specificity, combined with a low homodimerization K_d, which ultimately drives the dimer pathway *in vitro*. The same situation may apply here.

3.4.3 GCN4 binding rate constants are similar for dimeric and dimerization-competent versions

The association rate constants obtained in this study, with values of 2-3 nM^{-1} s^{-1} , are high when compared to the typical values for diffusion-limited protein-protein interactions of 10^{-5}-10^{-6} M^{-1} s^{-1}. This high k_{on} value is reasonable if we consider the strong electrostatic interactions contributing to GCN4-DNA association. Thus, given that the electrostatic interactions are important for the complex formation, an increase in ionic strength should reduce their contribution to binding free energy, and thus slow binding. Indeed, previous works have shown that increasing concentration of NaCl reduces GCN4 bZIP affinity to DNA (Anatoly I. Dragan et al., 2004).

The k_{on} obtained were very similar between the bZIP and the FD proteins. This is consistent with bZIP-binding to DNA as a dimer if the dimerization-competent version is largely dimeric under the experimental conditions i.e. $K_{d,homodimerization}$ is below 100 nM. It is not inconsistent with binding as monomer but would seem rather coincidental because of the lower affinity of monomer for the target. Under the stopped-flow concentrations most of the target would not have monomer bound, and K_d of ~ 100 μM would indicate that it also does not remain bound for long. It would therefore require that the k_{on} for monomer to the target, multiplied by the probability of another monomer being bound already, happened to match k_{on} for dimer. Furthermore, as previously described the k_{on} for the dimer is already extremely high at > 10^9 M^{-1} s^{-1} and a value orders of magnitude higher than this may be unlikely. Therefore, my data rather suggest that the K_d for homodimerization is below 100 nM. Previous CD experiments performed with a full bZIP under relatively similar conditions to those here have proposed that GCN4 bZIP homodimerizes with a K_d approximately 1.5 μM (Anatoly I. Dragan et al., 2004), but the authors themselves note a relatively large variation in estimated values within the literature.

k_{off} for GCN4 bZIP and FD are also similar, however it is slightly higher for the forced dimer version for both targets. A previous study made fusion versions of GCN4 and ATF2 proteins and demonstrated an apparent relationship between zipper stability and the rate of dissociation from target (Metallo & Schepartz, 1997). This would suggest that the protein comes off the target as monomer, and the authors argued using microscopic reversibility that bZIP will also arrive at the target as monomer. However, the changes they made to the proteins were much larger than those I have made. In my case I have a forced

dimer with an extremely high dimer stability, and I observe no difference in k_{off}. My result suggests that the protein leaves the target as a dimer.

Finally, I note that K_d for the binding of GCN4 bZIP and FD to the target sites was investigated using two methods; by obtaining equilibrium curves measuring fluorescence anisotropy and by calculating the k_{on} and k_{off} by performing fluorescence intensity stopped-flow kinetics. Both methods gave similar results (Figure 3.7), in support of a 2-state mechanism.

3.4.4 Binding of dimeric GCN4 to AP-1 and CRE are thermodynamically very similar

X-ray structures have been obtained for the GCN4 bZIP complex with DNA target sites, and the main DNA cognate sequences used were the AP-1 and the CRE. These two sites differ from each other by only one nucleotide in their sequence; CRE has an additional guanidine-cytosine base pair in its double-stranded sequence compared to the 9-base-pair AP-1. CRE DNA gets slightly bent by this interaction, whereas AP-1 DNA does not (König & Richmond, 1993). In this study I used both target sites to examine whether GCN4 behaves differently upon introduction of the additional GC bp.

The fluorescence anisotropy equilibrium binding titrations showed similar K_d values for CRE and AP-1, independently from the protein construct used. There is a slightly higher difference that can be noticed if the K_d constants derived from kinetics - where the K_d is ~4-fold lower for the hCRE construct. The full-length bZIP constructs associate with either CRE or AP-1 with the same rate (~2.4 nM^{-1} s^{-1}) and therefore, the difference in their dissociation binding constant comes from the different k_{off}. The k_{off} is inversely related with the lifetime of the protein-DNA complex. This lifetime could be of critical importance if the complex is involved in further various extra-molecular interactions and in the case of a transcription factor, downstream interactions, with other transcriptional machinery, are very common. This means that if the mean life span of the bZIP-DNA complex is too short, many of the necessary protein-protein interactions with other protein may be abortive because GCN4 would have had insufficient time to be activated. Therefore, the difference in k_{off} could have functional biochemical consequences in the cell.

Furthermore, the dissociation constants for the monomer BR in complex with hCRE or hAP-1 were determined. In both cases, I measured a significantly higher K_d (~100 μM) fairly similar to the non-specific DNA construct. These results suggest that the monomer BR-DNA complex is formed via non-specific interactions, probably involving the binding site's amino acids and the negatively charged DNA phosphate backbone.

3.4.5 Choice of DNA constructs

In this study, I used fluorescently labelled DNA constructs which allowed me to extract equilibrium and kinetic rate constants for binding. My pilot experiments started by using linear DNA versions which gave signal change by anisotropy. However, they did not give a fluorescence intensity change. The general requirement for higher label concentrations with an anisotropy-based technique poses difficulties in accurately measuring the very low K_d values. Lack of a significant fluorescence intensity change upon binding would also lead to significant difficulties obtaining kinetic rate constants using stopped-flow spectroscopy. Therefore, the target DNA binding of GCN4 was investigated using hairpin versions of the DNA.

In order to compare the GCN4 full-length bZIP behaviour in complex with the hairpin or the linear CRE site, I performed fluorescence anisotropy kinetics with the labelled linCRE using the bZIP construct. The obtained K_d for the linCRE was relatively similar the hCRE showing a 4-fold increase in binding affinity. This slight difference is due to the dissimilar k_{off} values obtained by the two methods since the k_{on} values were the same within error. Clearly therefore the surrounding sequence can impact the interaction somewhat, however it is less obvious which is "correct", and both may be reasonable models for the specific interaction. An apparent 3-fold difference using the two DNA versions was also observed in the monomer BR results where the K_d was calculated equal to ~30 μM for linCRE and ~90 μM for hCRE, which is interesting given the apparent differences in their binding modes.

The slight difference in fluorescence intensity might suggest some difference in how the protein is bound to hairpin and linear DNA. It could be possible that the fluorophore on the DNA is altered slightly in orientation compared with the protein in the two cases, due to the presence of the adjacent DNA loop. Previous studies have shown that

binding of GCN4 bZIP to its target sites induce DNA bending toward the minor groove (Antony I Dragan et al., 2004) and thus, DNA flexibility may play an important role to GCN4 bZIP association. However, the above proposal is an implication rather than a conclusion and further explanation of this difference requires more investigation whether this includes an *in vitro* or a computational approach (e.g. MD simulations of protein-hairpin complex).

A further insight into this phenomenon could be offered performing additional experiments using more versions of DNA. For example, the anisotropy kinetics of a linear AP-1 are missing as well as the equilibrium binding data from a NS DNA hairpin version. The length of double-stranded DNA between the site and the hairpin loop could be extended.

Furthermore, I used the AlexaFluor-488 fluorescent probe as a means to record the fluorescence intensity/anisotropy behaviour of the studied protein-DNA complex. One reason behind this decision was to replicate the system set-up that was used by the Shammas' group for CREB investigation. In addition to this, AlexaFluor®488 is photostable and allows for the sensitive detection and recording of low-abundance biological complexes, even over long-timescales of dissociation experiments. One reasonable alternative would be to test different fluorescent dyes to obtain a fluorescence intensity change whilst still using linear DNA constructs. Another possible way would be to fluorescently label the bZIP monomers instead of the DNA itself. This approach was used by Kohler & Schepartz, (2001), to study the kinetics of the Fos-Jun bZIP by performing forster resonance energy transfer (FRET) experiments. In this technique, each monomer was labelled with either rhodamine or fluorescein (a commonly used FRET pair) and the fluorescence intensity change upon dimerization was recorded. These could provide complementary information for pathway studies because they provide an opportunity to inspect the situation from the protein perspective.

3.5 Conclusions and further work

The mechanism of target search by GCN4 has yet to be determined. In this study, I created a series of GCN4 bZIP proteins: wild-type bZIP, forced-dimerized bZIP (FD) and monomer basic region only peptide (BR) to provide an insight into the process. I used fluorescence-based equilibrium and kinetic experiments, which enabled me to confirm that the aforementioned protein constructs bind to DNA with binding affinities ranging from low nM to higher μM, and to characterise their kinetics with GCN4's previously reported target sites; AP-1 and CRE.

It is not yet possible to make a definite statement regarding the preferred pathway, but on balance a dimer search may be more likely, rather than the monomer pathway previously suggested. The monomer is only 5-fold specific for the target, and the protein appears to leave target sites (and therefore presumably also non-target sites) as a dimer. However, the tight-binding of dimer with non-specific DNA would disfavour this pathway. It may be that there is a balance in the cell, and different bZIPs have differential fluxes through the pathways depending upon the values of these critical parameters. More experiments need to be done in order to determine whether wild-type GCN4 bZIP is a monomer or dimer, in the bZIP concentrations and conditions used in this work.

To clarify the complexity of the target search process and the contribution of the non-specific DNA, more investigation needs to be done. For example, various non-target DNA sequences of different lengths could be designed. Also, even though the monomer BR can be a good example of an incapable-to-dimerize version, it does not precisely simulate the monomer version of bZIP. A more realistic way to accomplish this, would be to create a bZIP mutant with a destabilised leucine zipper. This can be achieved by mutating one leucine residue to a helix-inducing amino acid (e.g. alanine) from the leucine zipper region. This will result in a bZIP protein with a higher homodimerization K_d. Such experiments have commenced in the Shammas lab. In this work I was not able to calculate folding and unfolding rate constants for homodimerization due to limited time, however this will be established in the lab soon. Once these figures become available it will be possible to model the flux through both pathways and establish whether the 5-fold specificity is sufficient to support the monomer pathway in the case of GCN4 in the presence of non-target DNA *in vitro*. It may be possible that it is, since the dimer pathway

is itself necessarily slowed by quite tight interactions with non-target DNA i.e. dimer is sequestered and not free to bind target.

Finally, my conclusions are based on *in vitro* studies using short DNA sequences and truncated versions of GCN4. The assembly of the GCN4-DNA complex *in vivo* is a more complicated process. Given that the cell nucleus has an enormous amount of genomic DNA packed in chromatin, GCN4 must not only search for its target DNA but also compete with other protein binding to DNA (e.g. transcription factors). Previous studies have also biophysically characterised the GCN4 binding to target sites in the presence of calf thymus DNA (Kohler & Schepartz, 2001). Calf thymus DNA contains a distribution of DNA sequences, some of which are likely to be similar to the AP-1 site, whilst most are completely nonspecific. In conclusion, there are many variables contributing to the target recognition process of GCN4 and a more detailed insight is desirable. Ultimately these questions may be best answered by determining the stoichiometry of the bZIP transcription factors as they perform their search within living cells using advanced single-molecule fluorescence tracking techniques.

Chapter 4. Structural characterisation of the wild-type GCN4 bZIP dimer

4.1 Introduction

The structurally heterogenous nature of IDPs is a big challenge in structural biology. Structured conformations (e.g. in a bound DNA-state) can on occasion be captured by X-ray crystallography (Ellenberger et al., 1992). However, disordered states cannot be crystallized. Even if an X-ray disordered structure was hypothetically possible, this would not represent the structural ensemble IDPs sample in solution. Instead, nuclear magnetic resonance (NMR) spectroscopy proves to be a useful tool for the investigation of IDP conformational dynamics and offers the unique potential to link structural propensities to biological functionality (Dunker et al., 2002). Furthermore, solution-state NMR provides key dynamic information regarding the soluble biomolecular complex in its native state (i.e. in solution). Backbone chemical shifts can be analysed to determine the secondary structure propensity as a function of the protein amino acid sequence.

The yeast GCN4 bZIP is an intrinsically disordered transcription factor (TF), and a "model" bZIP system; the AP-1 and CRE sites are two of its high affinity DNA binding sequences. bZIP domains consist of a basic region that contacts the DNA, and a leucine zipper (LZ). GCN4 homodimerizes through its leucine zipper forming a coiled-coil and the final protein-DNA structure is fully helical, with helices contacting the DNA through the major groove (Bracken et al., 1999; Ellenberger et al., 1992; M. L. Gill et al., 2016; Robustelli et al., 2013; Weiss et al., 1990). The structure of GCN4 bZIP in complex with its target site (CRE) has been determined by X-ray crystallography (Keller et al., 1995). In the absence of substrate GCN4 bZIP DNA-binding domain is disordered with residual helicity and the LZ forms a stable coiled coil, as demonstrated by CD (Bracken et al., 1999; Weiss et al., 1990) and NMR spectroscopy (Bracken et al., 1999; Lajmi et al., 2000). In the presence of non-target DNA, CD experiments have demonstrated a rigid structured LZ but also varying results regarding the BR domain, from helicity levels being similar to the apo-state (Weiss et al., 1990), to a 100% structured complex (Hollenbeck & Oakley, 2000). However, there have been no NMR studies involving GCN4 bZIP complexes with DNA.

The cell nucleus contains an extraordinary concentration of competing non-specific DNA of 10 mg/mL (Sprague et al., 2006), yet there has been only a limited structural assessment of the impact of binding "non-specific" sequences on GCN4 structure. In contrast to AP-1 binding, CD thermal melts show that the BR acquires limited additional helical structure upon binding non-specific DNA, and that unfolding is not co-operative (X. Wang et al., 2003). In this work, a wt GCN4 bZIP construct was studied in the apo-state, bound to CRE DNA and to non-specific (NS) DNA. In each case, relaxation data and backbone chemical shifts were obtained to calculate residue-specific structural and dynamic parameters (secondary structure propensity (SSP) (Marsh et al., 2006) and heteronuclear nuclear overhauser effect (hetNOE), respectively).

4.2 Material and Methods

4.2.1 ^{15}N and ^{13}C-labelled protein expression and expression trials

^{15}N and ^{13}C-labelled wt GCN4 bZIP was expressed using the protocol for the expression of double-labelled CREB bZIP developed by the Shammas group. The protein and DNA sequences used, and their molar extinction coefficients (EC) are shown in Table 4.1.

Table 4.1. Protein and DNA sequences used in the NMR experiments.

Construct	Sequence	EC (M^{-1} cm^{-1})
GCN4 bZIP residues 224-281	GSDPAALKRARNTEAARRS RARKLQRMKQLEDKVEELLSKN YHLENEVARLKKLVGER	1280
Double stranded non-specific DNA (nsDNA)	5'-GGCTAAATTTAGCC-3'	225704
Double stranded CRE DNA (creDNA)	5'-AGA**TGACGTCA**AGA -3'	235026

For the starter cultures 100 mL of 2xTY was inoculated with 4-5 colonies from the GCN4 bZIP BL21 cell culture plates, 100 μg/mL ampicillin, 34 μg/mL chloramphenicol and glucose (0.2% w/v final concentration). They were incubated overnight at 37 °C and 180 rpm shaking.

In order to produce double-labelled GCN4 bZIP, 3 L of 2xTY media was inoculated with confluent starter culture with a ratio 50:1. 100 μg/mL ampicillin, 34 μg/mL chloramphenicol and 0.2 % glucose was used. The bacterial cultures were incubated at 37 °C and 180 rpm until OD_{600} reached 0.6-0.8. Next, cells were centrifuged using JLA 8.1000 rotor at 5000 x g for 15 mins at 4 °C (step missed in controls 1 and non-labelled). The pellet was washed with 100 mL of 1x M9 minimal media. The pellet was resuspended in 1/4 initial volume with 1x M9 minimal. The following additives were added (step excluded in controls 1, non-labelled and 2):

1. 4 g/L glucose (^{13}C uniformly labelled) *
2. 3 g/L ammonium chloride (^{15}N labelled) *
3. 1 mM $MgSO_4$
4. 1 mM $CaCl_2$
5. 1 mM $FeCl_3$
6. 100 μg/mL ampicillin
7. 34 μg/mL chloramphenicol

*In the expression trials, non-labelled additives were used.

Cells were incubated for 1 hour at 37 °C with 180 rpm shaking, before being induced with 1 mM IPTG (IPTG addition missed in control 1 and 3) and incubated as previously for 4 hours. Cells were harvested using JLA8.1000 rotor for 30 mins at 5000 x g at 4 °C. Cell pellet was taken and frozen at -80 °C until purification.

4.2.2 Purification of labelled GCN4-bZIP

For the purification process, the protocol for the non-labelled GCN4 bZIP was followed (For protocol please see paragraph 2.2.10). Subsequently, protein samples were buffer exchanged into NMR buffer (100 mM sodium phosphate, ionic strength 0.185 M controlled by NaCl, 1 mM NaN_3, protease inhibitors and pH 6.5), concentrated to a final monomer concentration of 1.5 mM, snap-frozen using liquid N_2 and stored at -80 °C.

4.2.3 Nuclear magnetic resonance experiments

All nuclear magnetic resonance (NMR) experiments were performed in NMR buffer, containing 5% D_2O. For the apo-GCN4 experiments, the initial monomer concentration was 1 mM. For the protein-DNA experiments, initial monomer concentration of protein was 0.3 mM and double-stranded DNA samples varied from 50-375 μM. pH was determined and adjusted, if necessary, prior to measurement. The list of NMR experiments and their acquisition parameters is shown in Table 4.2 and Table 4.3. NMR experiments were performed in Bruker shaped NMR tubes which tolerate high ionic strength. The minimum sample volume was ~320 μL.

All NMR experiments were performed using the NMR spectrometers in the NMR facility at the Department of Biochemistry, University of Oxford. NMR spectrometers ranged from 600-950 MHz and were equipped with 5 mm TCI Cryoprobes, for improved sensitivity and with Bruker Avance III HD consoles. Experiments were performed using Bruker Topspin (750 MHz and 600 MHz = v3.2, 950 MHz = v3.6).

4.2.4 Data processing and analysis

All data were processed in NMRPipe (Delaglio et al., 1995). The peak assignment was performed using CCPNmr Analysis (Vranken et al., 2005).

Table 4.2. Acquisition parameters for the triple resonance NMR experiments used in backbone assignment.

Experiment (BT-BestTrosy)	Spectrometer	^{1}H Frequency (MHz)	Sweep Width (Hz)			Number of Complex Points		
			^{1}H	^{15}N	^{13}C/[^{1}H]	^{1}H	^{15}N	^{13}C/[^{1}H]
BT-HNCA	Bruker 600	600.133	6024.096	1398.601	3937.008	512	29	48
BT-HNCO	Bruker 600	600.133	6024.096	1398.601	2272.727	512	29	32
BT-HN(CO)CACB	Bruker 600	600.133	6024.096	1398.601	10309.278	512	29	64
HBHACONH	Bruker 600	600.133	6024.096	1522.070	5617.978	512	32	85
BT-HNCA	Bruker 950	949.664	9615.385	2222.222	7163.324	512	32	45
BT-HNCO	Bruker 950	949.665	9615.385	2222.222	3584.229	512	30	32
CBCACONH	Bruker 750	749.914	7575.758	1748.252	11312.217	512	26	50
HBHACONH	Bruker 750	749.914	7575.758	1748.252	6896.552	512	26	85

Experiment	Reference Chemical Shift (ppm)			Temp (°C)	Nus (%)
	^{1}H	^{15}N	^{13}C/[^{1}H]		
BT-HNCA	4.771	119.570	55.726	25	25
BT-HNCO	4.772	119.571	176.199	25	25
BT-NCOCACB	4.772	119.571	42.728	25	25
HBHACONH	4.772	119.074	4.772	25	25
BT-HNCA	4.676	118.947	55.603	35	25
BT-HNCO	4.676	118.939	176.068	35	25
CBCACONH	4.676	118.963	42.620	35	25
HBHACONH	4.676	118.963	2.377	35	25

Table 4.3. Acquisition parameters for the 2D ^{1}H-^{15}N NMR experiments.

Experiment	Spectrometer	^{1}H Frequency (MHz)	Sweep Width (Hz)		Number of Complex Pints	
			^{1}H	^{15}N	^{1}H	^{15}N
^{1}H-^{15}N HSQC	Bruker 600	600.133	6024.096	1398.601	1024	128
^{1}H-^{15}N BestTROSY	Bruker 600	600.133	6666.667	1398.601	1024	128
^{1}H-^{15}N BestTROSY	Bruker 750	749.914	7575.758	1748.252	1024	128
^{1}H-^{15}N BestTROSY	Bruker 950	949.665	9615.385	2222.222	1024	128
^{1}H-^{15}N HetNOE	Bruker 600	600.133	6024.096	1522.070	1024	96

Experiment	Reference Chemical Shift (ppm)		Temp (°C)
	^{1}H	^{15}N	
^{1}H-^{15}N HSQC	4.771	119.570	25
^{1}H-^{15}N BestTROSY	4.772	119.571	25
^{1}H-^{15}N BestTROSY	4.676	119.463	35
^{1}H-^{15}N BestTROSY	4.676	118.947	35
^{1}H-^{15}N HetNOE	4.772	119.074	25

4.3 Results

4.3.1 ^{15}N and ^{13}C-labelled protein expression trials

^{15}N and ^{13}C-labelled wt GCN4 bZIP was expressed using the protocol for the expression of double-labelled CREB bZIP developed by the and Shammas group. In order to test whether the particular protocol is also suitable for GCN4, I performed a series of scaled-down expression trials and ran a sample from crude-cell lysate for each, on an SDS-PAGE gel (Figure 4.1).

A relatively faint band was observed at the expected position for His-GB1-bZIP (20 kDa), matching the migration of an unlabelled sample produced according to my standard protocol, indicating that protein was successfully expressed. Although over-expression is not as successful as with the standard method employed, a similar reduction in yield was observed for production of labelled CREB-bZIP (Shammas group, unpublished data), and this expression level should be sufficient for the purposes of the NMR experiments (Figure 4.1).

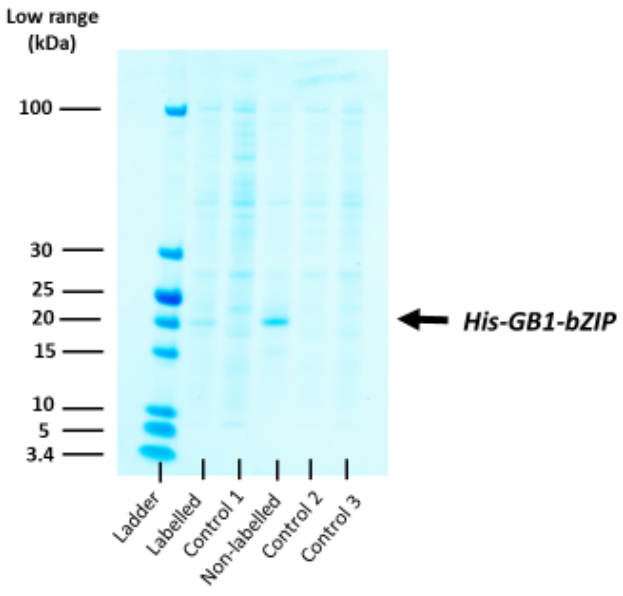


Figure 4.1. Expression trials for production of ^{15}N and ^{13}C-labelled GCN4-bZIP. SDS-PAGE gel electrophoresis of protein samples taken from each expression trial. Lanes left to right: 1st lane; PageRuler™ Unstained Low Range Protein Ladder, 2nd lane; purified ^{15}N and ^{13}C-labelled protein based on final protocol, 3rd lane; control without

centrifugation, additives, and IPTG addition, 4th lane; non-labelled protein from standard expression protocol (See Materials and Methods for non-labelled GCN4 bZIP purification Chapter 2.), 5th lane; control without additives, 6th lane; control without addition of IPTG. The expected migration position for His-GB1-bZIP is marked with an arrow.

4.3.2 NMR experiments of the apo-GCN4 bZIP dimer.

Peak assignment

The scaled-up expression of double labelled ^{15}N and ^{13}C-labelled produced sufficient GCN4 bZIP for NMR experiments. The protein was buffer exchanged into NMR buffer (See Materials and Methods) for NMR analysis. ^{1}H-^{15}N HSQC and ^{1}H-^{15}N BestTROSY (BTROSY) were collected to assess whether conditions used in biophysical studies are optimal for peak assignment. Conditions of 25 °C, pH 6.5 and 0.185 M ionic strength proved to be appropriate in order to assign 56 out of 58 protein residues (G224 and S225, at the N-terminus, could not be assigned in the ^{1}H-^{15}N HSQC) (Figure 4.2).

The GCN4 bZIP samples were prepared as described in the previous section and were used for the double and triple resonance experiments which were performed at 25 °C and pH 6.5. For the peak assignment of the apo-GCN4 bZIP, four protein backbone experiments were collected at 600 MHz: BT-HNCA, BT-HN(CO)CACB, BT-HNCO and HBHA(CO)NH. The assignment followed the well-established protocol of linking consecutive residues by the backbone amide ($i\pm1$) (described in the chapter written by C. Redfield (Redfield, 2015), shown in Figure 4.3).

For the peak assignment, I visualized 3D experiments in the form of plane strips (Figure 4.3, Figure 4.4 and Figure 4.5). Specifically, each strip visualizes one plane from the 3D spectra with a constant value of the ^{15}N dimension. As regards to the 3D experiments, the HN(CO)CACB correlates the backbone amide group (NH) of the *i* residue with the Cα and Cβ of the preceding residue (*i*-1). Once an NH group of a characteristic residue was spotted, then the identification of the amino acid types located in the particular string was made possible. For example, valine has downfield ppm values for the Cα compared to other residues and therefore, served as a good starting point for subsequent alignments. The HNCA and the HNCO spectra correlate the *i* NH group to the *i* and *i*-1 Cα

carbon and the *i* NH group to the *i*-1 CO carbon respectively. Furthermore, the HBHA(CO)NH spectrum links the *i* NH group to the *i*-1 Hα and Hβ.

These experiments allowed me to identify NH of adjacent residues by connecting the Cα of residue *i* (in the HNCA) with the Cα of residue *i*-1 in the HN(CO)CACB (connection illustrated in Figure 4.3b with red arrows). The Cβ chemical shifts provide information about amino acid type. An example of sequential assignment using triple resonance experiments is demonstrated in Figure 4.3. For the assignment of the α and β hydrogens, the HBHA(CO)NH spectrum was collected and was used in combination with the ^{1}H-^{15}N HSQC (Figure 4.4). The HBHA(CO)NH did not have as high signal to noise ratio (S/N) as required for some of the Hα and Hβ assignments. The HNCO spectrum was used to assign CO carbonyl carbon resonances (Figure 4.5).

A complete list of the chemical shifts is shown in the Appendices (Tables I-III) with a quality report (Table IV), summarising the percentage of assignments completed, for each of the constructs studied. The backbone information obtained from the above experiments at pH 6.5 and 25 °C were of high quality and suitable to be used for secondary structure prediction calculations.

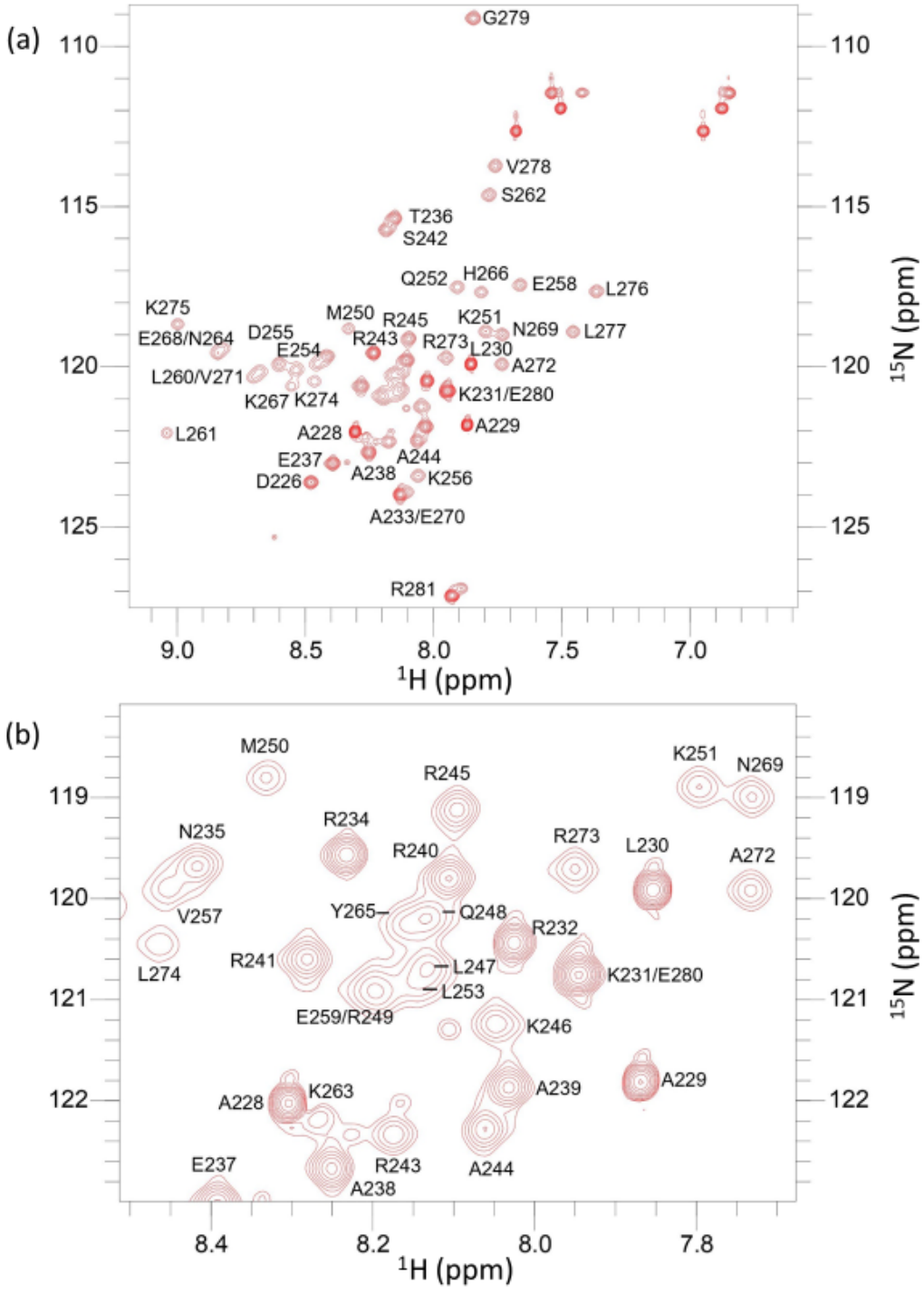


Figure 4.2. (a) Assigned ^{1}H-^{15}N HSQC spectrum of GCN4 bZIP (at pH 6.5 and 25 °C). (b) The crowded region around Q248 is shown enlarged.

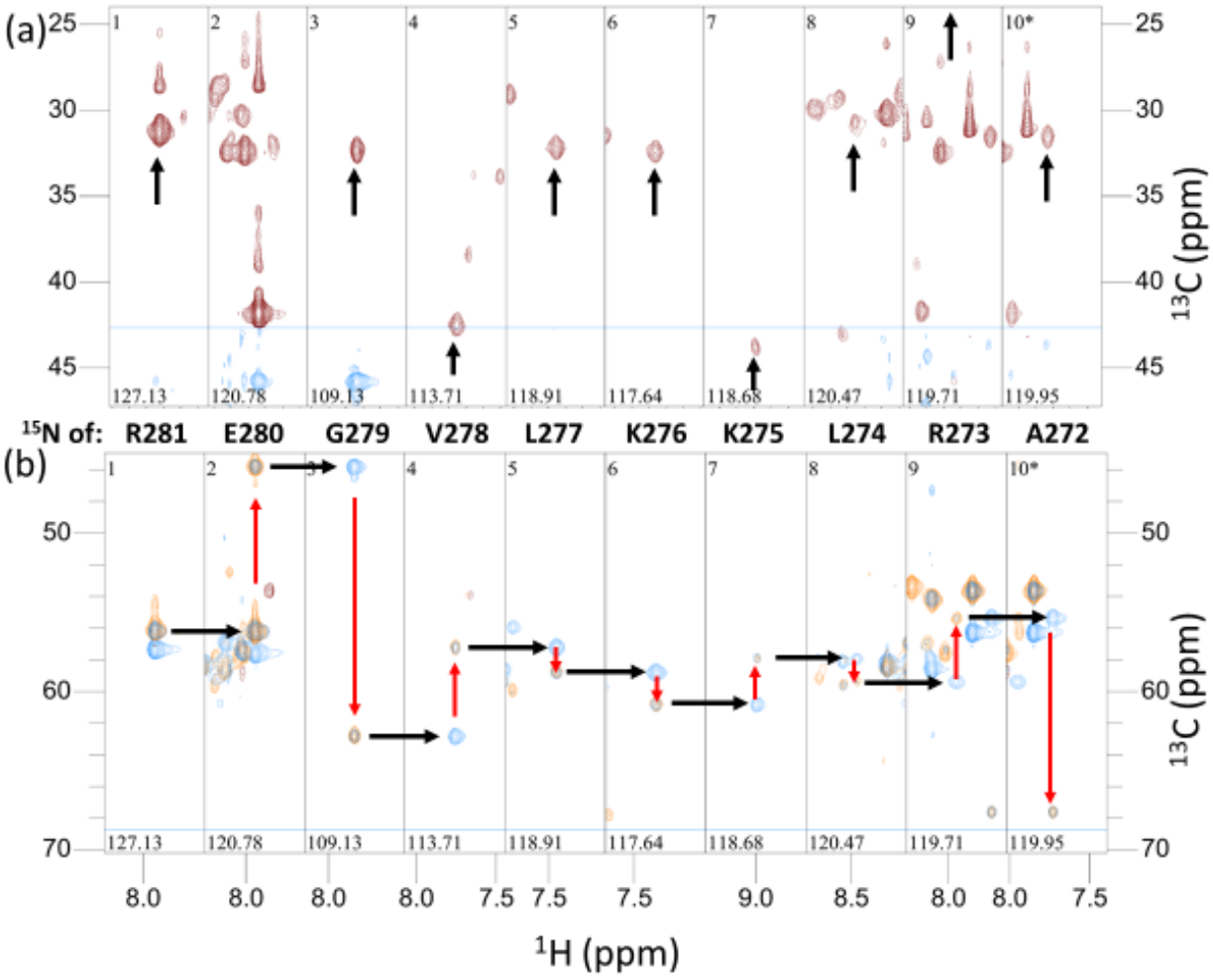


Figure 4.3. Triple resonance experiments to assign GCN4 bZIP residues.
(a) Representative HN(CO)CACB spectra (peaks coloured brown). Cβ carbon of the previous residue (i-1) is being shown with a black arrow. (b) Representative overlaid HNCA (blue peaks) and HN(CO)CACB spectra (orange peaks) for the Cα carbons. Arrows indicate direction of connection from i to i-1 residue.

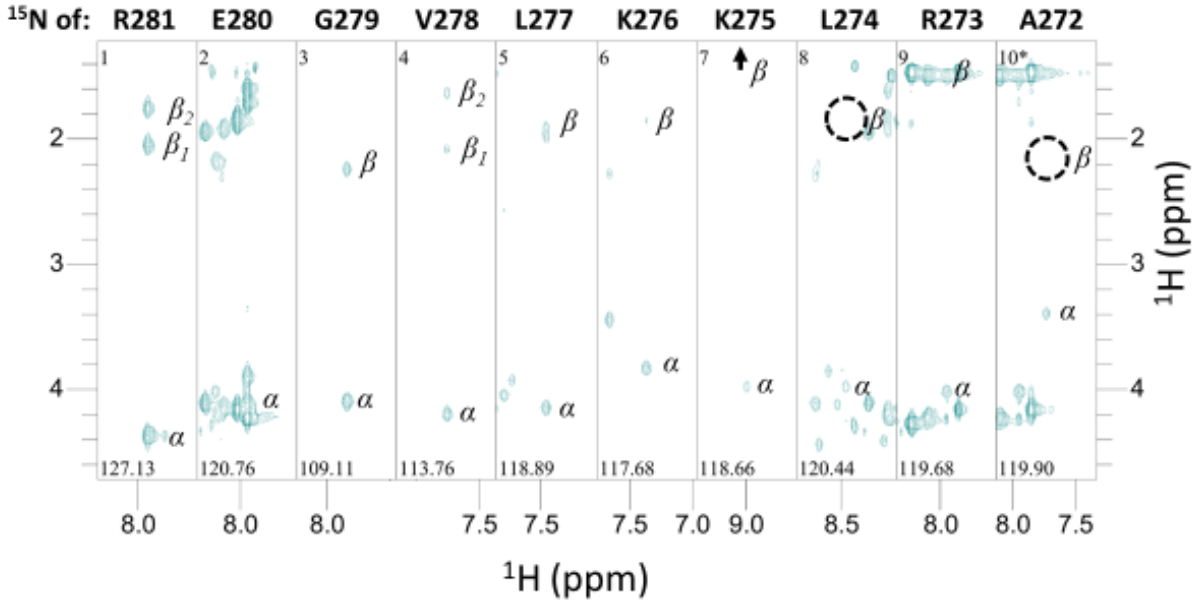


Figure 4.4. HBHA(CO)NH triple resonance spectra showing Hα and Hβ hydrogens of consecutive residues.

Dashed black circles and black arrow show location of peaks not shown in the spectra due either to their weak intensity at the contour level chosen or their upfield chemical shift, respectively.

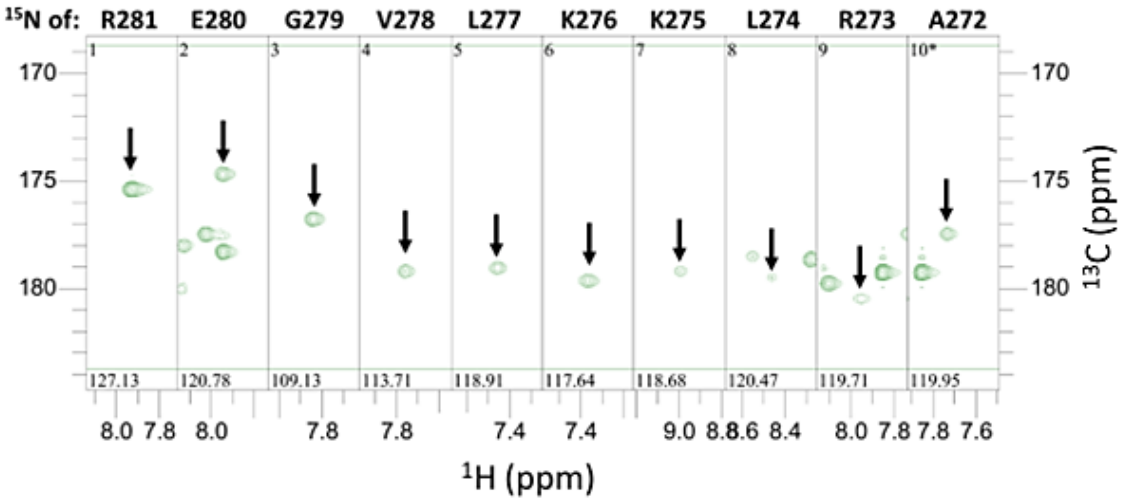


Figure 4.5. HNCO triple resonance spectra showing backbone carbonyl carbons of consecutive residues.

Black arrows indicate the CO of the preceding residue.

Secondary structure and dynamic characterisation of the apo-GCN4 bZIP

The derived chemical shifts from the assigned apo-GCN4 bZIP were used for the characterisation of secondary structure within the peptide. SSP (Marsh et al., 2006) is an algorithm that combines chemical shift information from different nuclei and assigns each individual residue to a different preference of canonical secondary structure. Each amino acid is allocated with a different value of SSP – the more helical residues are assigned with higher values of SSP (with full helix being approximately 1 and a random coil being 0). Only $^{13}C\alpha$, $^{13}C\beta$ and $^{1}H\alpha$ chemical shifts have been suggested as very useful for disordered proteins and to determine α-helical content. Other chemical shifts (^{1}H and ^{15}N) are suggested to be avoided for IDPs in particular (Marsh et al., 2006; Y. Wang, 2002). Different combinations of types of atoms were used in order to assess the effect of each atom type on the SSP output (Figure 4.6a) in which the three combinations showed similar SSP levels and pattern of secondary structure propensity.

In all cases, the LZ (residues 250-281) demonstrated higher helical propensity than the BR (residues 224-249) (Figure 4.6a). Moving towards the N-terminus, SSP values drop below 0.8 from R249 and below 0.5 from T236. The residues at the C-terminus of the bZIP also show lower helical propensity (K276-R281). The SSP between these regions indicates stable helical structure in the LZ. In the BR, the helical propensity drops to a value of ~0.33 at A233-N235 and then increases back to ~0.5 again at L230. This low-helicity region corresponds to the positively charged residues constituting the DNA-binding domain.

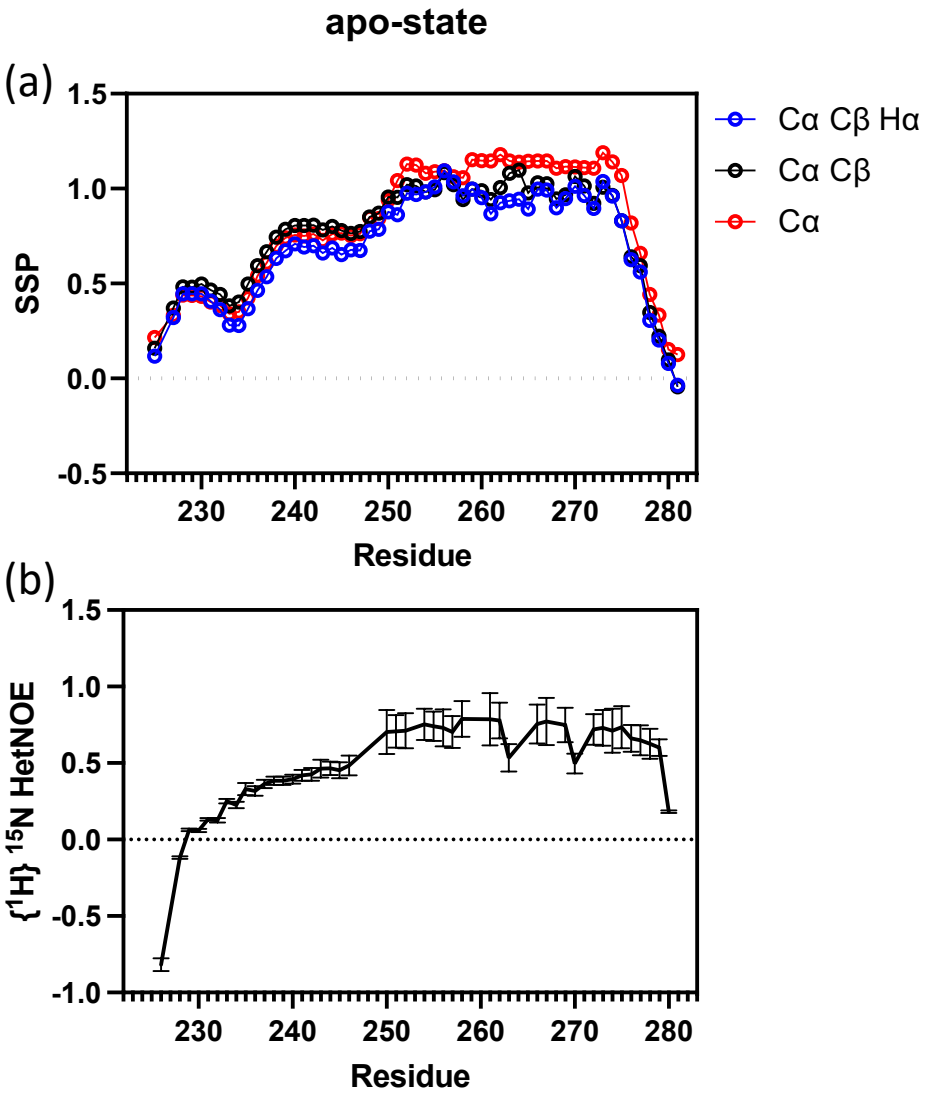


Figure 4.6. Secondary structure propensity and dynamics for apo-GCN4 bZIP. Calculated using: (a) Secondary structure propensity (Marsh et al., 2006) (SSP) algorithm using the assigned chemical shift information obtained from the backbone triple resonance experiments. Different combinations of atoms were used to assess the possible effect on predicted residual helicity. (b) $\{^{1}H\}$-^{15}N heteronuclear NOE experiment collected at 25 °C and pH 6.5 using the 600 MHz spectrometer. The error bars are standard deviations (SD). The BR and the LZ domains consist of residues 225-251 and 253-281, respectively, as determined by the uniprot.org database.

Backbone dynamic information was also obtained from the $\{^1H\}$-^{15}N heteronuclear NOE experiment (Figure 4.6b). The $\{^1H\}$-^{15}N heteronuclear NOE gives details about backbone flexibility in a residue-specific manner, on the ps/ns timescale (faster than the overall molecular tumbling). The heteronuclear NOE for each residue was extracted and plotted as a function of protein sequence including the SD of each value. A high value of the heteronuclear NOE (0.65-0.8) indicates a relatively rigid backbone while a lower value (< 0.65) indicates a more flexible backbone. The results parallel the SSP results, demonstrating high values of the heteronuclear NOE in the LZ and a gradual decrease in the BR region, moving from K246 to the N-terminus. Thus, the LZ which forms a coiled coil, shows low flexibility, while the BR, not participating in dimerization, is highly flexible with respect to the LZ i.e. in this case helicity and flexibility appear anti-correlated. However, even though the SSP analysis demonstrates apparent low helicity at the positively charged residues and increased helicity at the N-terminal tail, the $\{^1H\}$-^{15}N hetNOE displays a gradual and consistent increase in flexibility, with respect to the LZ core structure, in the whole BR towards the N-terminus. In conclusion, the DNA-binding residues of GCN4 bZIP (positively charged amino acids) are highly disordered in the absence of DNA.

The results for the apo-state of GCN4 bZIP will now be compared with data for GCN4 bZIP bound to DNA.

4.3.3 NMR experiments of the GCN4 bZIP dimer in complex with non-specific DNA (nsDNA) and target CRE DNA (creDNA) sites

In a further investigation, GCN4 bZIP in complex with nsDNA and creDNA was studied to understand the changes in helicity and flexibility observed in interaction with non-specific and specific DNA sequences. At the present time, there have been no NMR studies reported investigating GCN4 bZIP behaviour in complex with DNA.

DNA titrations to fully saturate GCN4 bZIP with DNA and temperature increase from 25 °C to 35 °C

Initially the same conditions as for the apo-GCN4 bZIP, were used for the NMR experiments. An initial ^{1}H-^{15}N BTROSY was collected for the apo-state at 950MHz. The nsDNA and creDNA constructs that were used in the experiments are GCN4 bZIP binders with binding affinities in the hundreds of nM and lower, as demonstrated in the equilibrium and kinetic binding assays described in the previous chapters. Thus, the chemical shifts of the BR, which is the DNA-binding domain, were expected to change upon addition of DNA. For the nsDNA titrations, double-stranded DNA concentrations of 0 μM, 50 μM, 100 μM, 150 μM, 200 μM, 250 μM, 350 were used with a GCN4 bZIP concentration of ~0.3 mM. As can be seen in Figure 4.7, the GCN4 bZIP peaks gradually move in the spectra as the concentration of nsDNA is increased; this allows the peaks to be tracked between apo-GCN4 bZIP and nsDNA-bound bZIP. For creDNA, addition of 75 μM DNA led to significant peak broadening and disappearance of many peaks; these observations were indicative of slow and/or intermediate exchange. In these circumstances a detailed titration does not allow the peaks to be tracked. For this reason, further experiments with 300 μM and finally 375 μM of creDNA were collected; no differences were observed between these two spectra indicating that the GCN4 bZIP was fully saturated with creDNA (Figure 4.8). Because the peaks in the creDNA spectra cannot be tracked directly from the apo-GCN4 bZIP spectrum, assignment of the creDNA-bound state of GCN4 bZIP is more difficult.

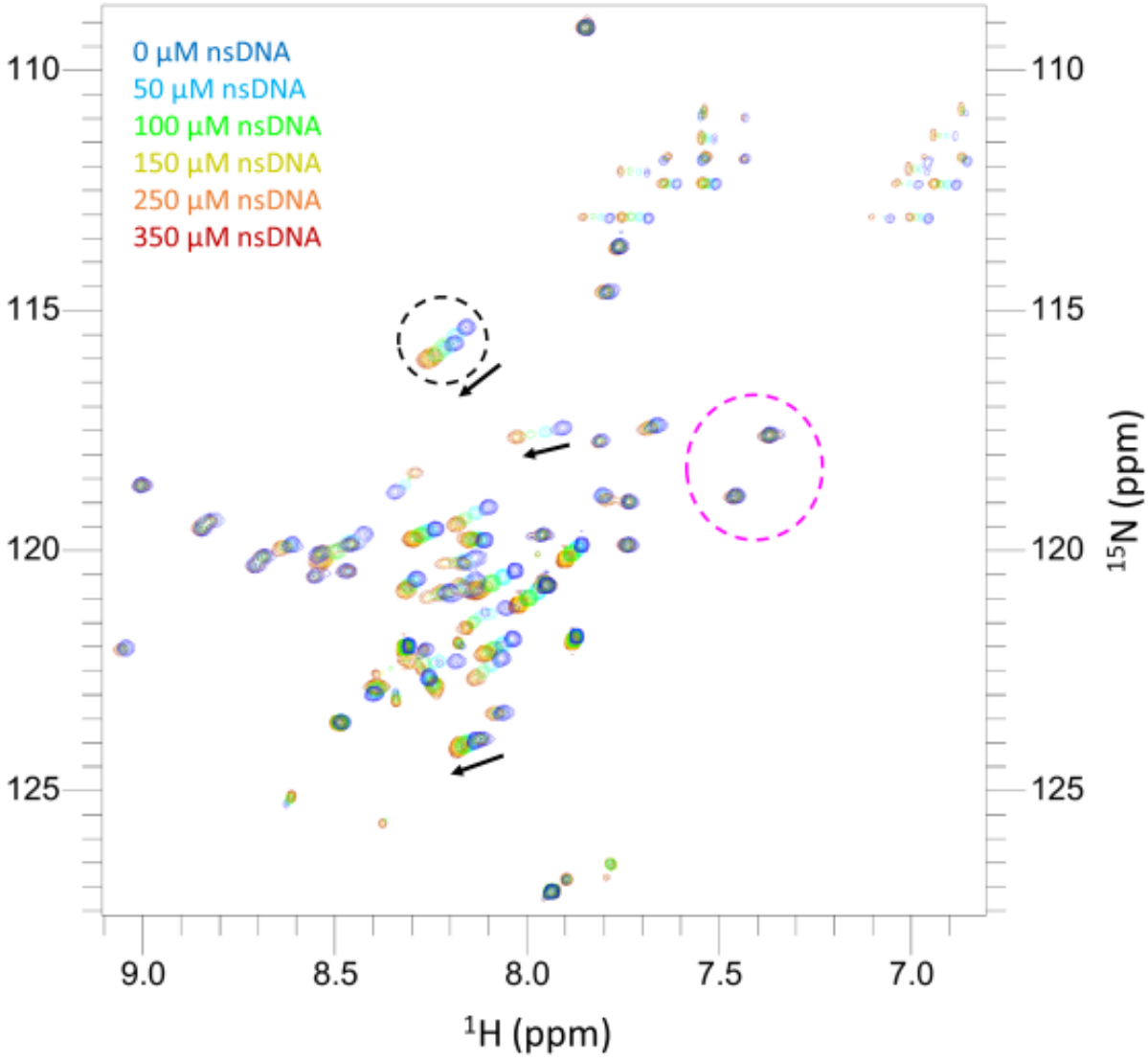


Figure 4.7. 950 MHz ^{1}H-^{15}N BTROSY spectra of the GCN4 bZIP in the presence of varying concentrations of nsDNA

Concentrations of nsDNA: 0 µM nsDNA (blue), 50 µM nsDNA (cyan), 100 µM nsDNA (green), 150 µM nsDNA (yellow), 250 µM nsDNA (orange), 350 µM nsDNA (maroon). Black arrows show directionality of peak shifts after addition of increasing concentrations of nsDNA. Dashed black circle shows the location of basic region residues T236 and S242 which are perturbed by the addition of nsDNA. Dashed pink circle shows the peaks of L276 and L277 which are located in the LZ.

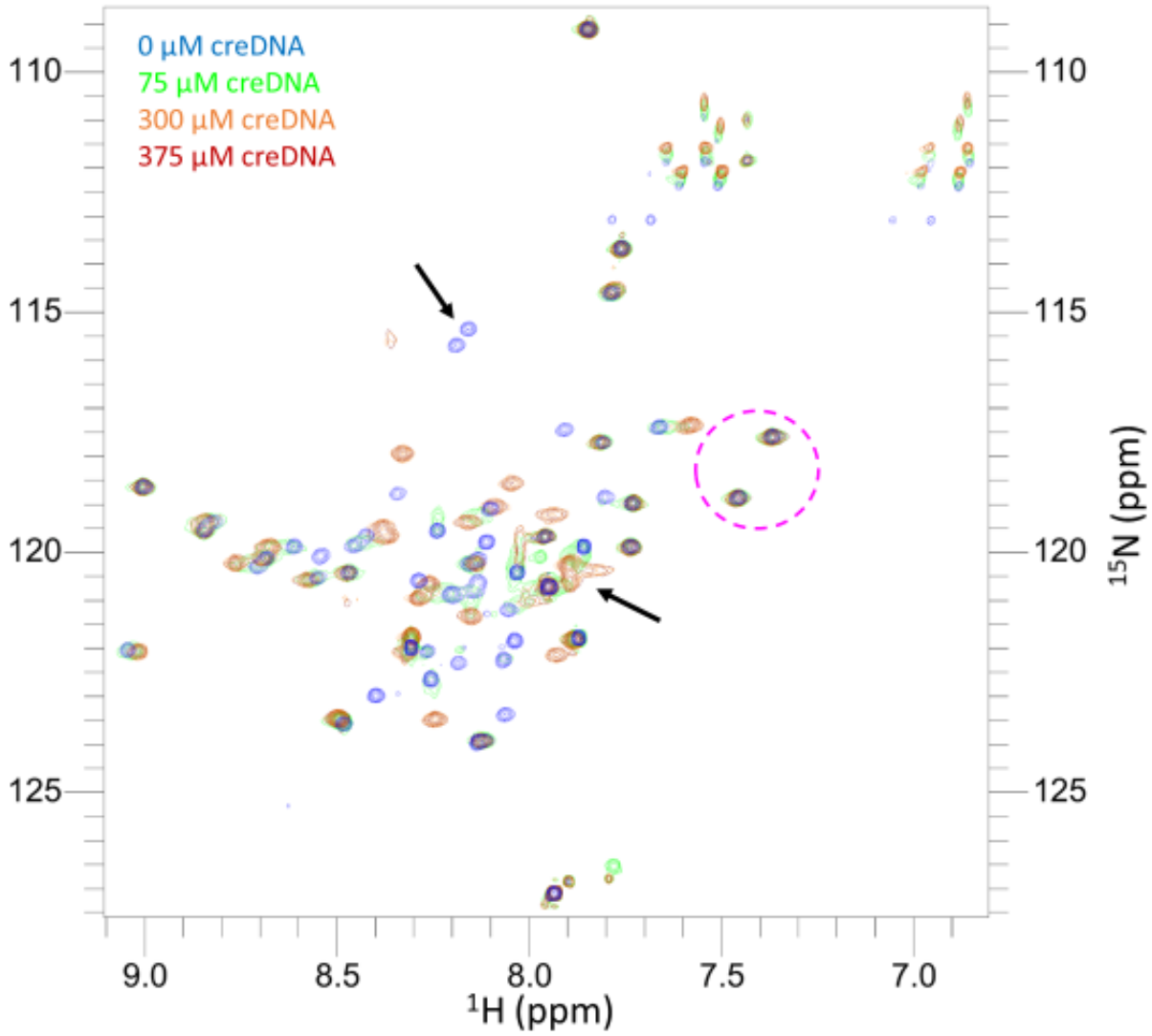


Figure 4.8. 950 MHz ^{1}H-^{15}N BTROSY spectra of the GCN4 bZIP in the presence of varying concentrations of creDNA

Concentrations of creDNA: 0 µM creDNA (blue), 75 µM creDNA (green), 300 µM creDNA (orange), 375 µM creDNA (maroon). Black arrow shows missing or broadened peaks of unsaturated GCN4 bZIP in the T236/S242 region (~8.2 ppm, ~115 ppm) and in the BR (~8.0 ppm, ~120 ppm). Dashed pink circle show the maintained peaks of L276 and L277 which are located in the LZ.

For the creDNA, the fully saturated GCN4-DNA complex showed broadened peaks, especially in the ^{1}H 8.2 ppm and ^{15}N 121 ppm region where most of the BR residue peaks are found. Previous NMR experiments from the Shammas' group using CREB-DNA complexes, required higher temperatures (up to 35 °C) to obtain good signal-to-noise in 3D backbone assignment spectra. To obtain sharper peaks that would enable me to assign more BR residues, I chose to perform GCN4-DNA NMR experiments at 35 °C. For this purpose, I performed ^{1}H-^{15}N BTROSY experiments at 25 °C, 30 °C and 35 °C for both fully saturated GCN4-DNA systems (Figure 4.9 and Figure 4.10) to acquire suitable information for comparison of GCN4 bZIP nsDNA and creDNA. To confirm that the apoGCN4 bZIP behaves in a similar way at 35 °C compared to 25 °C, 3D experiments at 35 °C were also collected (SSP comparison is shown in Appendices, Figure I).

For the peak assignment, I was able to easily spot LZ residues peaks in both the nsDNA and creDNA spectra since they did not deviate from their initial location in the apo-GCN4 bZIP spectra. I could easily follow peaks from the 0 μM DNA spectra to the full-bound nsDNA spectrum at 25 °C due to the fast exchange conditions and could also follow the peaks from the 25 °C to the 35 °C spectra. 3D experiments were used for the nsDNA-bound sample to confirm these assignments and to measure the ^{1}H and ^{13}C chemical shifts needed for the SSP calculations. For the creDNA-bound state of GCN4 bZIP, the DNA-binding amino acids, located in the BR, could not be followed in the titration due to slow/intermediate exchange conditions and these could only be assigned with the 3D triple-resonance backbone experiments. Thus, the same 3D experiments as collected for the apo-state, were collected for the nsDNA and creDNA-bound samples. On the basis of the assignments obtained from these spectra, the SSP and heteronuclear NOE analysis were then carried out.

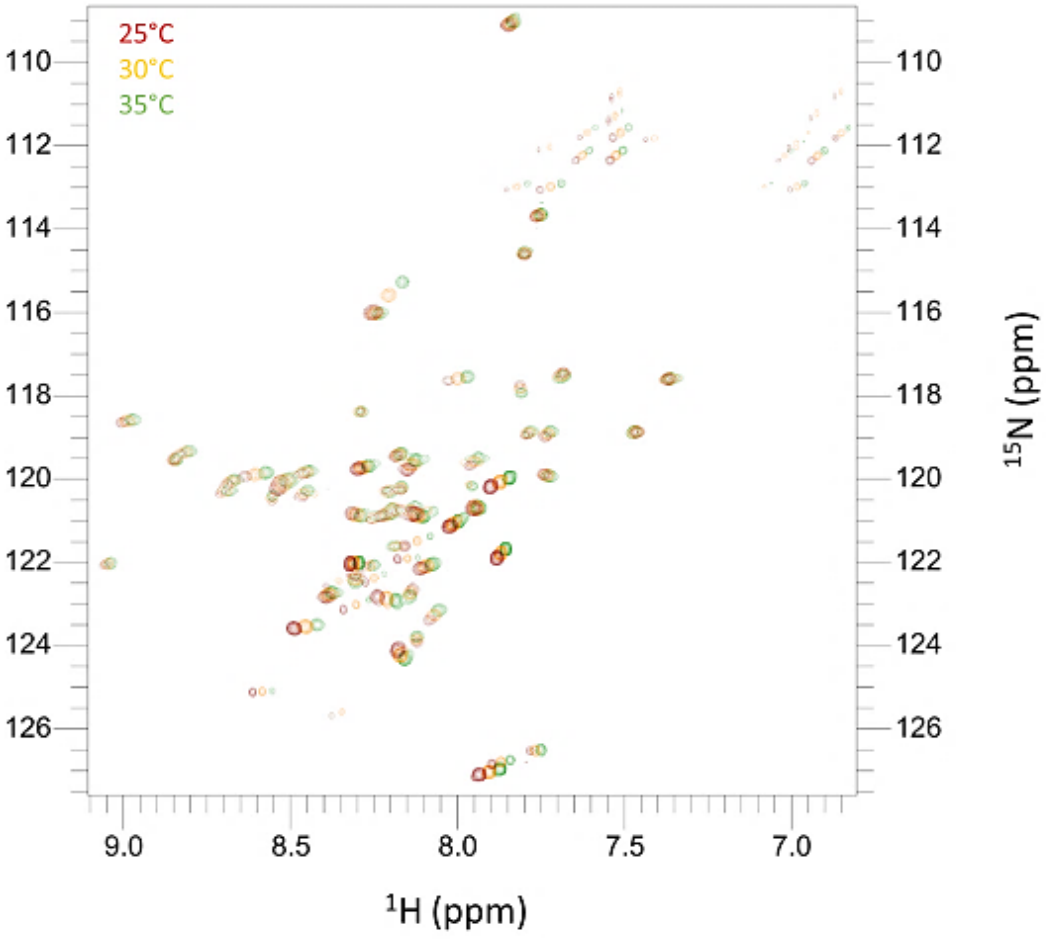


Figure 4.9. 950 MHz ^{1}H-^{15}N BTROSY spectra of the GCN4 bZIP in complex with nsDNA collected over a range of temperatures: 25 °C (red), 30 °C (yellow) and 35 °C (green)

These experiments were performed by using the fully-saturated GCN4 bZIP sample (350 μM nsDNA) from the DNA titrations.

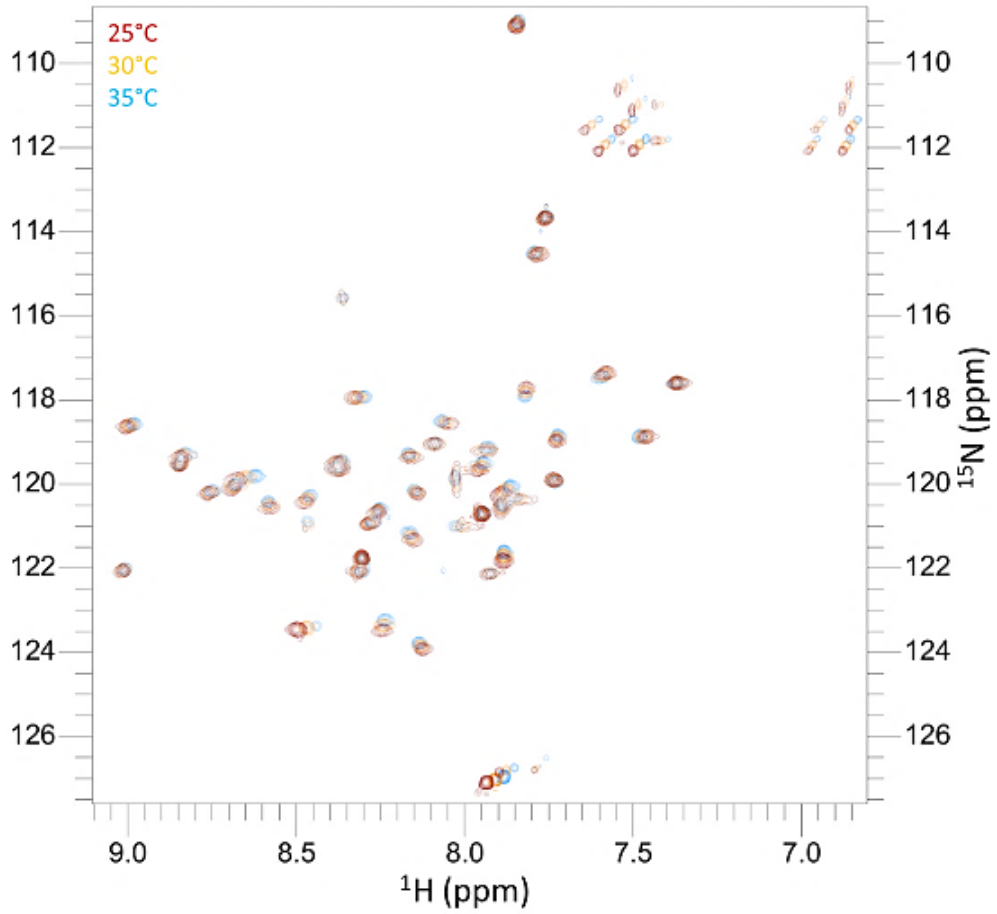


Figure 4.10. 950 MHz ^{1}H-^{15}N BTROSY spectra of the GCN4 bZIP in complex 375 µM creDNA collected over a range of temperatures: 25 °C (red), 30 °C (yellow) and 35 °C (blue)

These experiments were performed by using the fully-saturated GCN4 bZIP sample (375 µM creDNA) from the DNA titrations.

Triple resonance backbone experiments of the GCN4 bZIP-DNA complexes for peak assignment.

For the peak assignment of the GCN4 bZIP DNA-bound state, four protein backbone experiments were collected: BT-HNCA, CACB(CO)NH, BT-HNCO and HBHA(CO)NH. Next, I followed the same protocol used for the assignment of the apo-GCN4 state, where I linked consecutive residues by the backbone amide ($i\pm1$) and visualized 3D experiments in the form of strips (described in Paragraph 4.3.2). The assigned spectra are shown in Figure 4.11 for the nsDNA and in Figure 4.12 for the creDNA-bound bZIP. A complete list of the chemical shifts is shown in the Appendices (Tables I, II and III) with a quality report of the assigned spectra (Table IV).

For the nsDNA system, the backbone information obtained from the above experiments at pH 6.5 and 35 °C allowed me to determine all the ^{1}H and ^{15}N chemical shifts for all residues. However, in the case of the creDNA systems, only the LZ and the N-terminal capping residues of the BR could be assigned. Many peaks for the creDNA-binding residues were absent from the BTROSY spectrum collected at 950 MHz. A BTROSY spectrum collected at the lower frequency of 750 MHz showed slightly less broadening for some of the BR residues which is consistent with a chemical exchange process for this region of GCN4 bZIP in the presence of creDNA (Figure 4.12). The CBCA(CO)NH experiment was collected for the creDNA sample at 750 MHz to take advantage of the decreased line broadening. In both nsDNA and creDNA, less information was collected for the Cβ, Hα and Hβ atoms, especially for the DNA-binding amino acids, due to broader peaks in the DNA complexes. Nevertheless, the data were considered of high enough quality and suitable to be used for secondary structure prediction calculations and for further comparison with the apo-state.

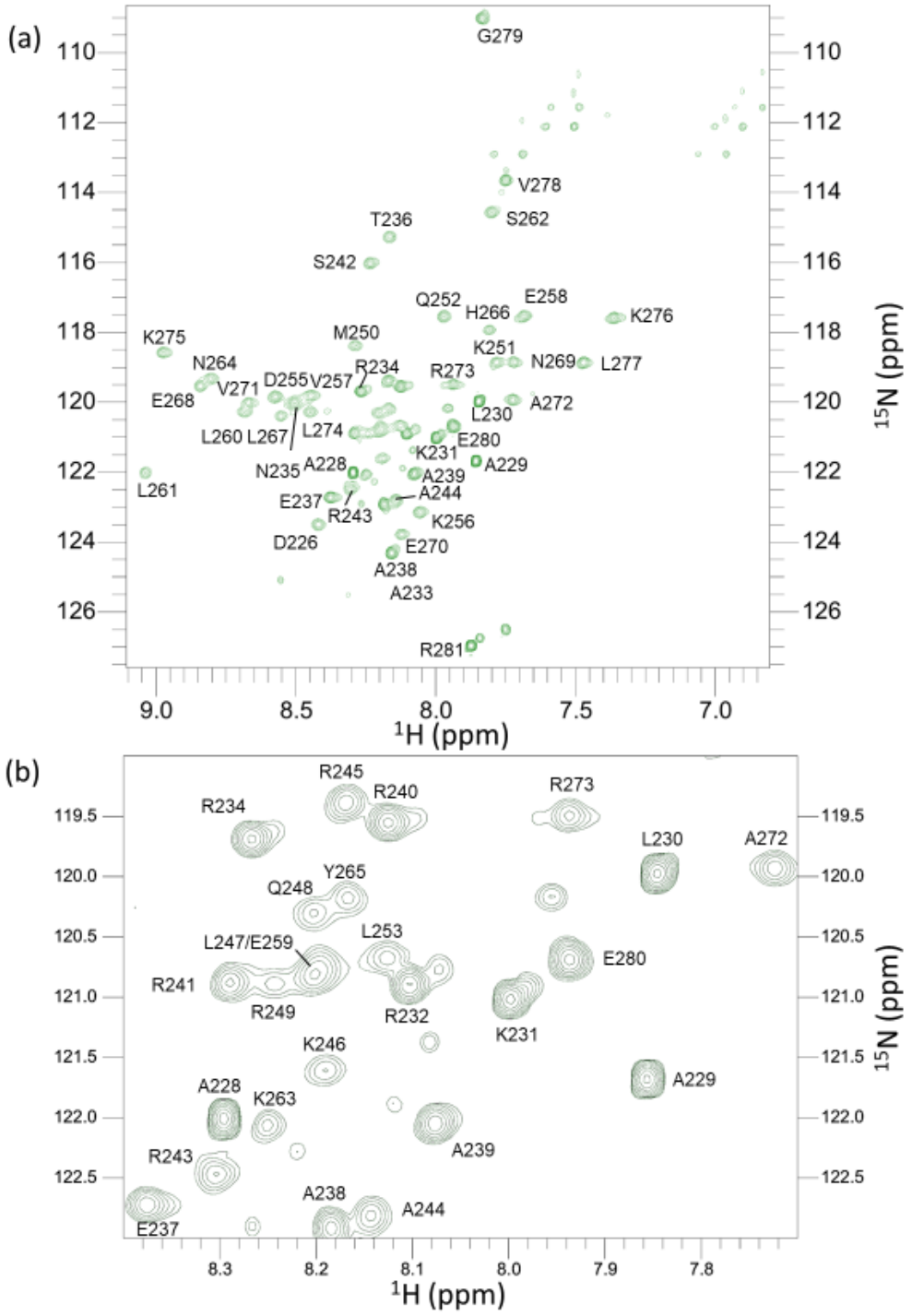


Figure 4.11. (a) Assigned ^{1}H-^{15}N BTROSY spectra of nsDNA-bound GCN4 bZIP (at pH 6.5 and 35 °C). (b) The crowded region around R232 is shown enlarged.

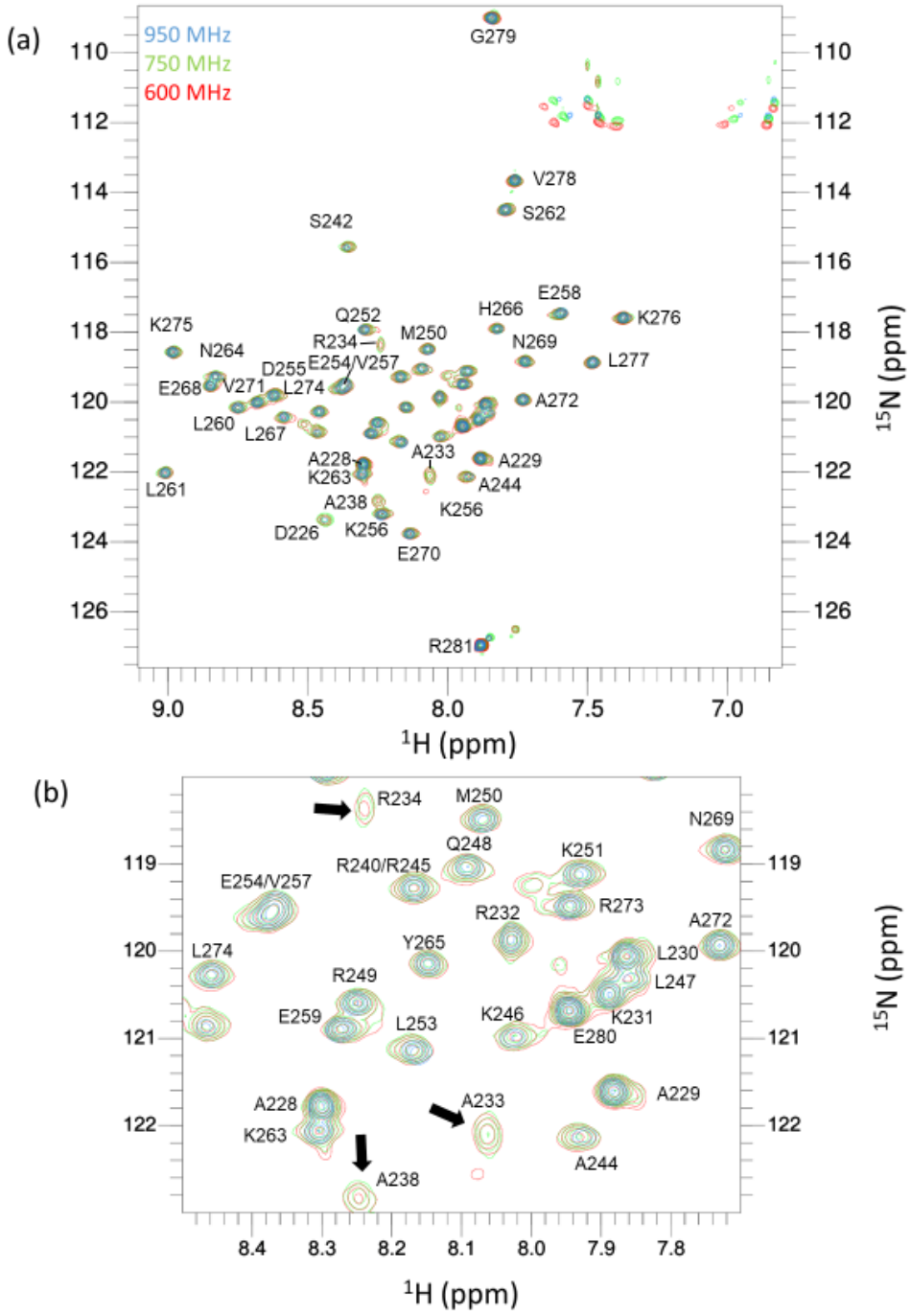


Figure 4.12. (a) Assigned ^{1}H-^{15}N BTROSY spectra of creDNA-bound GCN4 bZIP (at pH 6.5 and 35 °C). (b) The crowded region around Y265 is shown enlarged.
Spectra were collected at different frequencies (600 MHz coloured in red, 750 MHz coloured in green and 950 MHz coloured in blue). Characteristic examples of amino acids whose peaks were only visible in the lowest frequency (600 MHz) are R234, A238 and A233 shown with black arrows.

Secondary structure and dynamic characterisation of the nsDNA- and the creDNA-bound GCN4 bZIP

As previously done for apo-GCN4, SSP calculations based on the 3D NMR spectra described above were calculated for the bound protein. This enables direct the comparison of secondary structure regions based on the presence or absence of DNA binders. As with apo-GCN4 SSP, different nuclei were combined to assess the impact of each atom on the SSP profile.

In the case of the nsDNA, high helicity is observed in the LZ region (250-281), using all the combinations of atoms in the SSP algorithm (Figure 4.13). However, the SSP values decrease from the residue R249 towards to the N-terminus until residue A238. At residue A238, there is an apparent drop in helicity, which reaches near 0.2-0.3 SSP values at residue 233. In the beginning of the N-terminus, the N-capping amino acids (res. S225-R232) demonstrate an increase in helicity, reaching SSP values up to 0.5 and then drop again.

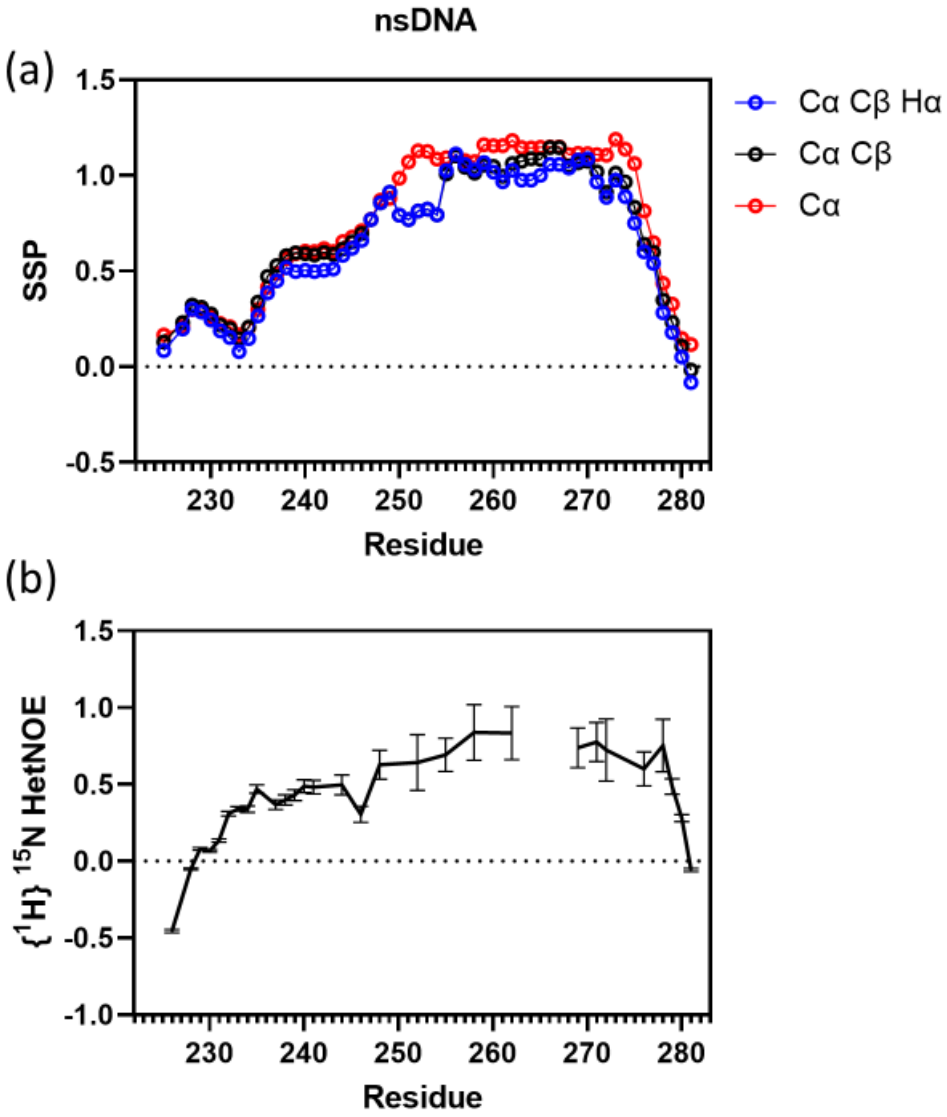


Figure 4.13. Secondary structure propensity and dynamics for GCN4-nsDNA. Calculated using: (a) Secondary structure propensity (Marsh et al., 2006) (SSP) algorithm using the assigned chemical shift information obtained from the backbone triple resonance experiments. Different combinations of atoms were used to assess the possible effect on residual helicity (b) $\{^{1}H\}$-^{15}N heteronuclear NOE experiment collected at 25 °C and pH 6.5 using the 600 MHz spectrometer. The error bars are standard deviations (SD). The BR and the LZ domains consist of residues 225-251 and 253-281, respectively, as determined by the uniprot.org database.

In the case of the creDNA-bound GCN4, the same pattern is observed across the entire sequence using any variation of atom types (Cα, Cα-Cβ or Cα-Cβ-Hα) in the SSP input. It is interesting that amino acids located in the BR show SSP values as high as the LZ domain (~1.1-1.2) but the SSP values decrease before residue R232 in the sequence. The missing assignments for many of the DNA-interacting residues in the BR in the presence of creDNA makes a detailed analysis of secondary structure in this region difficult.

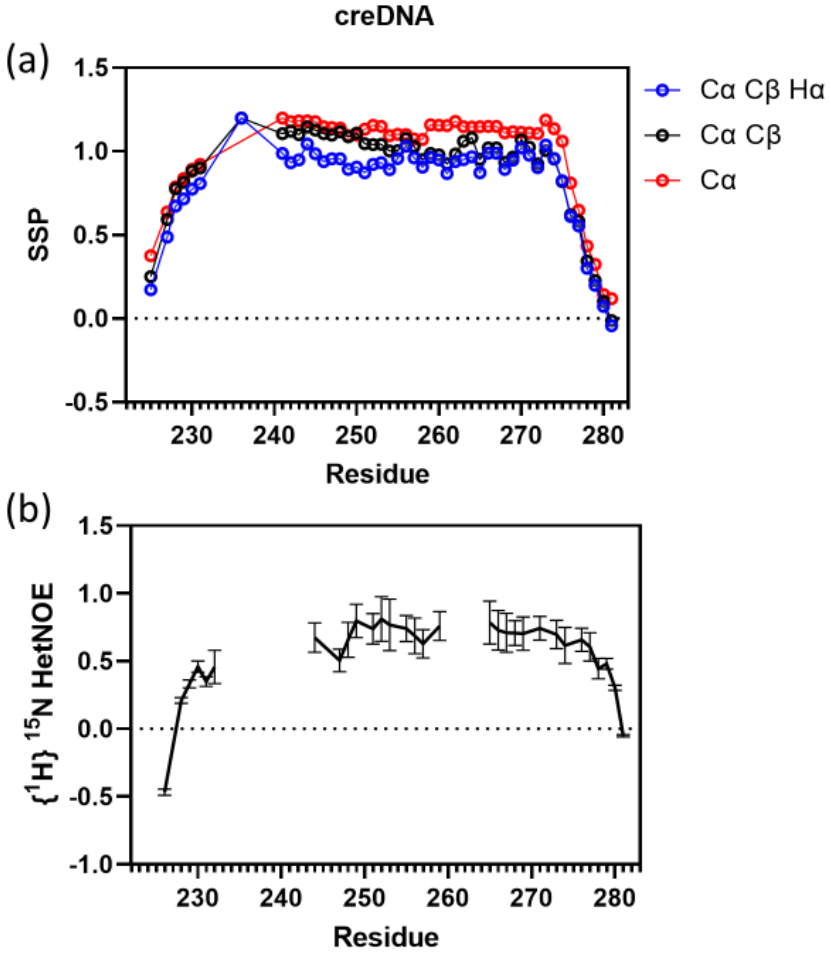


Figure 4.14. Secondary structure propensity and dynamics for GCN4-creDNA. Calculated using: (a) Secondary structure propensity (Marsh et al. 2006) (SSP) algorithm using the assigned chemical shifts information obtained from the backbone triple resonance experiments. Different combinations of atoms were used to assess the possible effect on residual helicity. Data points are connected with a line if information for no more than 5 residues are missing. (b) $\{^{1}H\}$-^{15}N heteronuclear NOE experiment collected at 25 °C and pH 6.5 using the 600 MHz spectrometer. The error bars are standard deviations (SD). The

BR and the LZ domains consist of residues 225-251 and 253-281, respectively, as determined by the uniprot.org database.

Protein backbone dynamics were also determined from $\{^{1}H\}$-^{15}N heteronuclear NOE experiments at 25 °C (Figure 4.13(b), Figure 4.14(b)). However, due to the higher MW of the final complexes which resulted in reduced signal-to-noise (compared to the apo-state) and the fact that not all residues were assigned in creDNA ^{1}H-^{15}N spectra, cut-off values were set. This decision was in order to carefully select pairs of peaks in the saturated and unsaturated spectra. More specifically, mean values with SD>0.2 and mean hetNOE values > 0.8, which are not theoretically possible, were discarded. Overlapping peaks were also excluded.

For the nsDNA-GCN4 complex, the heteronuclear NOE in combination with the SSP plot demonstrates a rigid LZ domain with hetNOE values reaching 0.8. The values corresponding to amino acids located at the BR gradually decrease as we move towards the N-terminus. For the creDNA-GCN4 complex, the heteronuclear NOE data show a similar pattern in the LZ region, as expected, since the two complexes are structurally similar with close MWs. However, the values for most residues of the BR could not be obtained due to extensive peak broadening for these residues. Thus, a more detailed comparison of the hetNOE data is further discussed in Paragraph 4.4.

4.3.4 Comparative NMR data analysis of the three systems; apo-GCN4 bZIP, GCN4 bZIP-nsDNA and GCN4 bZIP-creDNA

The NMR data obtained for the three systems offer a good basis for the comparison of the apo-GCN4 bZIP to GCN4 bZIP bound to DNA. Each system possesses different physical characteristics (e.g. MW, protein-DNA dissociation constant etc). As a result, different amounts of data were available for each sample. For example, a significant number of Hα peaks could not be assigned in the creDNA whereas in the apo-GCN4 protein most of the Hα were assigned. For a fair comparison of the three systems, the NMR results need to be more uniformly processed and merged into single plots.

1. **Secondary structure propensity (SSP) calculations.**

The SSP output for the studied complexes is shown in Figure 4.15, where the combination of ^{13}Cα, ^{13}Cβ and ^{1}Hα was used as an input.

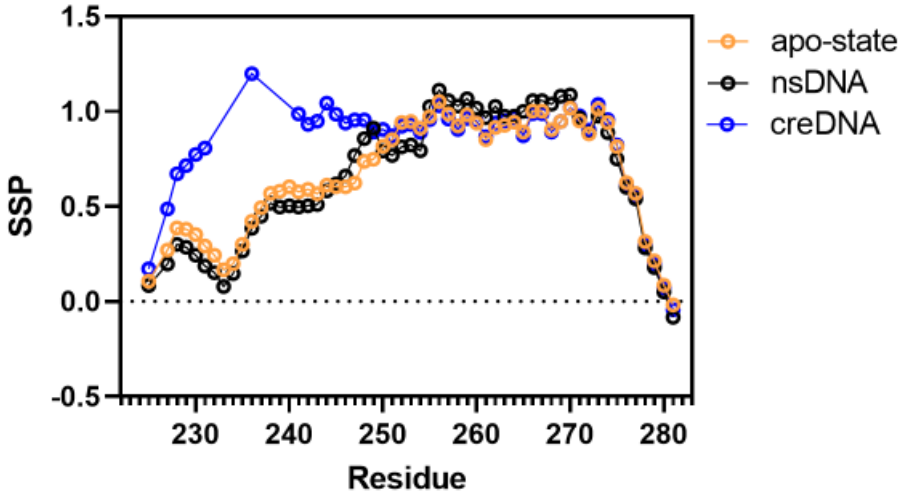


Figure 4.15. Secondary structure propensity (SSP) of GCN4 bZIP.
Calculated using: secondary structure propensity (Marsh et al. 2006) (SSP) algorithm using the assigned chemical shifts information obtained from the ^{13}Cα, ^{13}Cβ and ^{1}Hα backbone triple resonance experiments. The BR and the LZ domains consist of residues 225-251 and 253-281, respectively, as determined by the uniprot.org database.

2. {^{1}H}-^{15}N Heteronuclear NOE plots - merged.

Here, the heteronuclear NOE plots of the apo-GCN4, GCN4-nsDNA and GCN4-creDNA are merged (Figure 4.16).

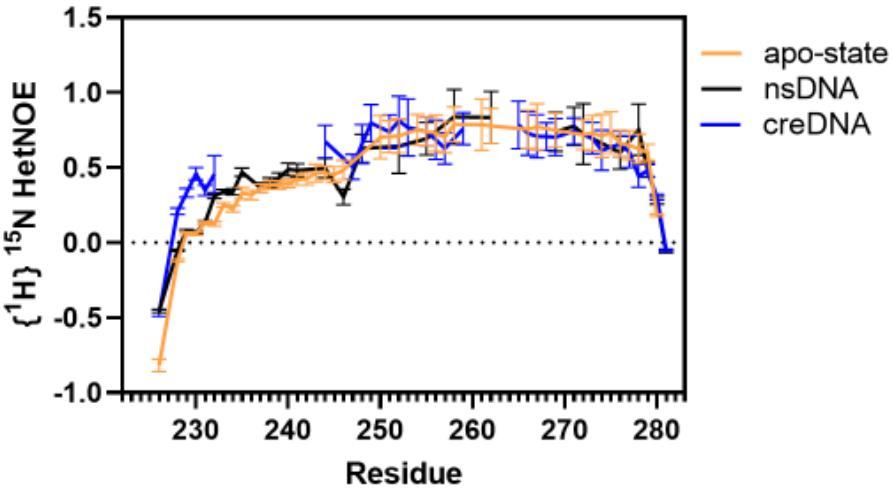


Figure 4.16. {^{1}H}-^{15}N Heteronuclear NOE experiment for the apo-state (orange), nsDNA-GCN4 (black) and creDNA-GCN4 (blue).
Collected at 25 °C and pH 6.5 using the 600 MHz spectrometer. The error bars are standard deviations (SD). The BR and the LZ domains consist of residues 225-251 and 253-281, respectively, as determined by the uniprot.org database.

3. Change in the Chemical Shifts of DNA-GCN4 complexes compared to the apo-GCN4 state.

A further investigation to assess the impact caused by the addition of DNA to GCN4 bZIP and the differences between the two DNA sequences, involves the direct comparison of the chemical shifts between two states: the free and the fully-bound GCN4 state. For this purpose, the chemical shift changes can be calculated for the ^{1}H and ^{15}N atoms upon addition of nsDNA or creDNA (Figure 4.17). For both of the two systems, the most significant changes are noted in the BR whereas the LZ region shows only small chemical shift changes. Further description of these results is provided in the Discussion section.

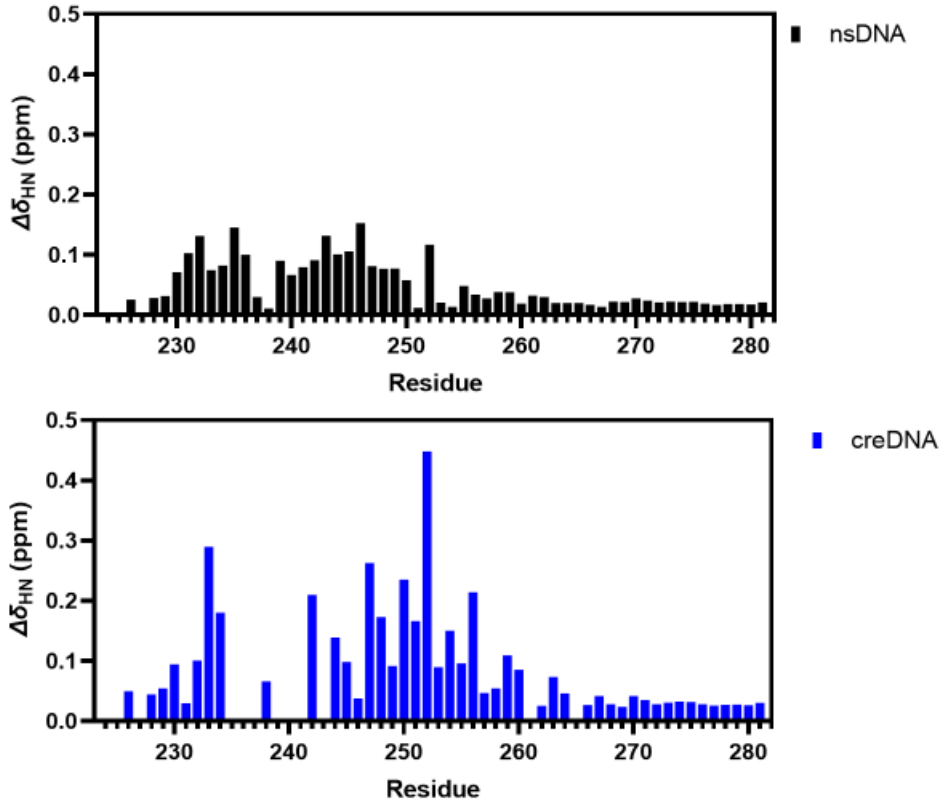


Figure 4.17.Combined ^{1}H - ^{15}N weighted shift differences between the apo-GCN4 and GCN4 DNA systems.
Black; shift changes between apo-GCN4 and nsDNA-GCN4, blue; shift changes between apo-GCN4 and creDNA-GCN4. Calculated using CCPNmr where a scale factor of 1.0 for ^{1}H and 0.14 for ^{15}N were used. The peaks for residues 224-225, 227 in the nsDNA sample and residues 224-225, 227, 235-237, 239-241, 243 in the creDNA sample have not been assigned and therefore the shift difference for these is missing. The BR and the LZ domains consist of residues 225-251 and 253-281, respectively, as determined by the uniprot.org database.

The Cα, Cβ and Hα nuclei give the most information regarding the secondary structure propensity. Among these atoms, Cα has been suggested to be a good predictor of α-helix (Y. Wang, 2002). Thus, I calculated the Cα chemical shift difference between the apo-state and the DNA-bound state, to directly investigate whether Cα moved upfield or downfield upon DNA binding (Figure 4.18). When the Cα chemical shift moves downfield, it means that there is an increase in helical structure content and vice versa. In Figure 4.18, it is clear for both of the systems that the LZ residues do not show a significant change whereas the residues located in the BR appear to show the most important changes. Overall,

creDNA appears to show a downfield movement which means that the complex becomes more helical. However, in the case of nsDNA, there is an overall upfield chemical shift change for Cα, especially in the BR, meaning that GCN4 backbone becomes more unstructured upon binding to nsDNA. Many Cα carbons could not be assigned in the BR domain of bZIP-creDNA. Nevertheless, looking at the Cα changes, the overall BR becomes more helical in the presence of target DNA and more random coil upon addition of non-specific DNA.

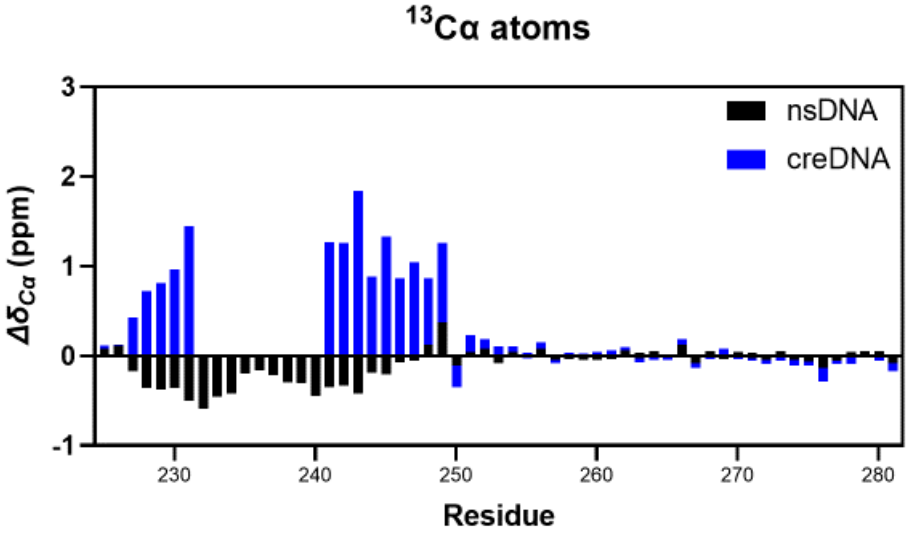


Figure 4.18. GCN4 backbone Cα atoms shift change upon addition of nsDNA (black) or creDNA (blue).
The chemical shift difference plotted on the y-axis comes from the subtraction of Cα (GCN4-DNA) atoms from Cα (apo-GCN4) atoms. A positive value indicates a downfield shift and an increase in helical structure. The peaks for residues 232 to 240 in the creDNA sample have not been assigned and therefore the shift difference for these is missing. The BR and the LZ domains consist of residues 225-251 and 253-281, respectively, as determined by the uniprot.org database.

4.4 Discussion

4.4.1 Expression and purification of ^{15}N and ^{13}C-labelled GCN4 bZIP

For the structural characterization of GCN4 bZIP with NMR, the ^{15}N and ^{13}C-labelling of the protein construct was required. Following the GCN4 protein expression method described in Paragraph 4.3.1 and isolating isotopically labelled GCN4 using the purification protocol for non-labelled GCN4, led to satisfactory yield and purity. Although the electrophoresis gels, which included protein samples from the expression trials, demonstrated a fainter band for the labelling pipeline compared to the expression of non-labelled protein, the amount of pure GCN4 isolated was high enough to be used in NMR experiments of the apo-state, nsDNA-bound and creDNA-bound states.

4.4.2 GCN4-creDNA is the most structured complex followed by the apo-state and the GCN4-nsDNA complex

The apo-state of the GCN4 bZIP has been analysed before with NMR spectroscopy (Bracken et al., 1999). However, Bracken et al. (1999) used a low pH value of 4.7 for the NMR experiments, which deviates from physiological conditions. In the present study a pH of 6.5, which is closer to physiological pH, has been used. In the previous study only HN, ^{15}N, Cα and CO shifts were reported. Here, the Cβ and Hα have also been assigned allowing a more accurate description of the secondary structure propensity. The chemical shifts reported by Bracken et al. (1999) are qualitatively similar to the chemical shifts determined in this study.

The final SSP plots for the three systems, apo-GCN4, nsDNA-GCN4 and creDNA-GCN4, obtained using the NMR data collected in the 3D backbone experiments are shown in Figure 4.15. From a first glance, the SSP values between these three systems are well conserved for the LZ region, from residue D255 to R281. The amino acids in the LZ region

reach an SSP value close to ~1 which is indicative of structure formation. This is in agreement with the X-ray structures of GCN4 both the CRE and AP1 DNA site (e.g. PDB ID 1YSA or 2DGC) which show that the LZ forms a coiled coil.

However, from E234 towards the N-terminus, a different secondary structure profile is observed between the systems. Overall, creDNA shows significantly higher helicity values compared to the other two systems. Although, many Cα carbons could not be assigned in the BR domain of bZIP-creDNA, a comparison of the Cα data clearly show a more helical BR (Figure 4.18). A further proof for this conclusion is provided by comparing the S225-N235 region where SSP values for all systems have been obtained. This region (mostly referred to as N-capping residues) demonstrates enhanced helicity, reaching SSP values of 0.8 – values considered to indicate high helical content. It is, thus, reasonable to postulate that the entire BR bound to creDNA is more helical compared to the nsDNA and apo-state systems.

Previous CD studies of GCN4 bZIP bound to non-target DNA sequences show different overall helicity for the entire bZIP. For example, (Weiss et al., 1990), found that the λ O_L1 operator site, which is not known as a GCN4 binding site, induces an insignificant amount of helicity compared to the apo-state of bZIP. On the other hand, Hollenbeck and Oakley (2000) showed that GCN4 bZIP bound to a non-target DNA sequence is 100% helical. Interestingly, the current study shows that the BR is more disordered when it binds to nsDNA than when it binds to creDNA and also somewhat less helical compared to the apo-state. To further investigate if this decrease in helicity is real, I collected 3D backbone data of the apo-state at 35 °C (Figure Appendices). Even though a higher temperature does induce more disorder in the BR, the apo-GCN4 BR is still a bit more structured than the nsDNA-bound BR. A further proof of these results is the comparison of the raw Cα atom data (Figure 4.18). An increase in Cα chemical shifts (downfield) is indicative of structure formation whereas a decrease in Cα chemical shifts (upfield) indicates higher disorder. The creDNA-GCN4 demonstrates higher Cα values in the S225-R249 region meaning a significant increase in helicity, whereas nsDNA-GCN4 shows a decrease in chemical shift values in the region P227-L247 meaning lower helicity.

Regarding the creDNA-GCN4 bZIP complex, much backbone chemical shift information is missing at the BR domain L230-R243. In a further investigation, I investigated the protein-DNA binding interactions involved in the X-ray structure of GCN4-CRE (PDB ID 2DGC) (Table 4.4). I found that the DNA-interacting residues are K231-R243 matching to the residues that could not be assigned in the creDNA-GCN4 system. These residues, which are mostly assigned in the apo- and nsDNA-bound state show a significantly lower helicity in the SSP results for these states. The absence of peaks for L230-R243 in the creDNA-bound state suggests that the observed peak broadening results directly from the interaction with the DNA. In the case of AP-1 site (GCN4-AP-1 complex PDB ID: 1YSA), the majority of interactions are similar. The most prevalent differences are the following; GCN4 T236 residue interacts through hydrophobic forces with AP-1, K231forms a hydrogen bond with the DNA base side-chain, R240 instead of water bridge, forms a hydrogen bond with AP-1 and, R234 and R245 interact with the phosphate backbone through salt bridges (Ellenberger et al., 1992).

Table 4.4. Detailed interaction information of GCN4 bZIP binding to CRE site.
Information was obtained by analysing GCN4 complex with PDB ID: 2DGC and using the protein-ligand interaction profiler (PLIP) (Salentin et al., 2015).

Residue	Distance (Å) (H-A and D-A for hydrogen bonds)	Ligand group (DNA)	Type of interaction
T236	3.92	Base side-chain	Hydrophobic
A239	3.67	Base side-chain	Hydrophobic
N235	3.30,3.94	Base side-chain	Hydrogen bond
N235	2.05,3.00	Base side-chain	Hydrogen bond
T236	2.69,3.30	Phosphate backbone	Hydrogen bond
R240	2.75	Phosphate backbone	Water-bridge
K231	5.47	Phosphate backbone	Salt bridge
R232	4.03	Phosphate backbone	Salt bridge
R240	4.44	Phosphate backbone	Salt bridge
R240	4.89	Phosphate backbone	Salt bridge
R243	4.76	Phosphate backbone	Salt bridge

The backbone dynamics were also obtained for the three systems to determine the backbone flexibility in a residue-specific manner. The ^{1}H-^{15}N heteronuclear NOE experiment for the nsDNA-GCN4 and creDNA-GCN4 are missing data for significantly more residues compared to the apo-state, especially in the BR domain. However, a fair comparison can be made for the N-capping residues (S225-R232). In the creDNA data, A228 starts with a value of 0.2 and increases to 0.5 at residue R232. In the case of apo-GCN4 and nsDNA-bound GCN4, a more flexible N-terminal tail is observed since A228 starts with values of -0.1 and -0.05, respectively, increasing up to 0.1 and 0.3, respectively, at residue R232. The LZ region demonstrates the same level of rigidity with residues showing values around 0.7-0.8 in all of the three systems.

4.4.3 nsDNA and creDNA show different exchange regime

In the case of the nsDNA titration, the GCN4 peaks are observed to move in the spectrum as the nsDNA concentration is increased; this is known as fast exchange on the NMR time scale. For the creDNA titration, many peaks broadened and disappeared in the 75 μM spectrum and the peaks that remained did not move in the spectrum. With 300 μM creDNA a new set of peaks appeared in the spectrum; this behaviour is known as intermediate to slow exchange on the NMR time scale. These results appear consistent with the previous equilibrium and kinetic data where I determined creDNA binding to GCN4 bZIP to be 30-fold tighter than that for nsDNA site, with the difference presumed to originate from a higher k_{off} for nsDNA (recently demonstrated in the lab for CREB bZIP).

4.4.4 How does CREB bZIP compare to GCN4 bZIP?

Similar NMR experiments have been employed by Conor Kelly (DPhil candidate in Shammas' group) to investigate the structure of CREB bZIP free or bound on nsDNA or creDNA. The same DNA constructs presented in this study, were used, as well as the same pH, ionic strength and temperature. The CREB bZIP construct used is of similar MW to the GCN4 bZIP used in this study. Both constructs are illustrated in Table 4.5.

Table 4.5. Saccharomyces cerevisiae GCN4 bZIP and Homo sapiens CREB bZIP sequences used in NMR studies.
Sequences were aligned according to the LZ leucines (highlighted in yellow). Identity and similarity between the full-length protein sequences were calculated equal to 6.73% and 18.65%, respectively (Stothard, 2000). For GCN4, the BR and the LZ domains consist of residues 225-251 and 253-281, respectively, as determined by the uniprot.org database. For CREB, the BR and the LZ domains consist of residues 270-295 and 297-318, respectively, as determined by the uniprot.org database.

bZIP protein	Residue sequence
GCN4 bZIP (res. 225-281)	SDPAALKRARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVGER
CREB bZIP (res. 277-339)	QPAEEAARKREVRLMKNREAARESRRKKKEYVKSLENRVAVLENQNKTLIEELKALKDLYSHK

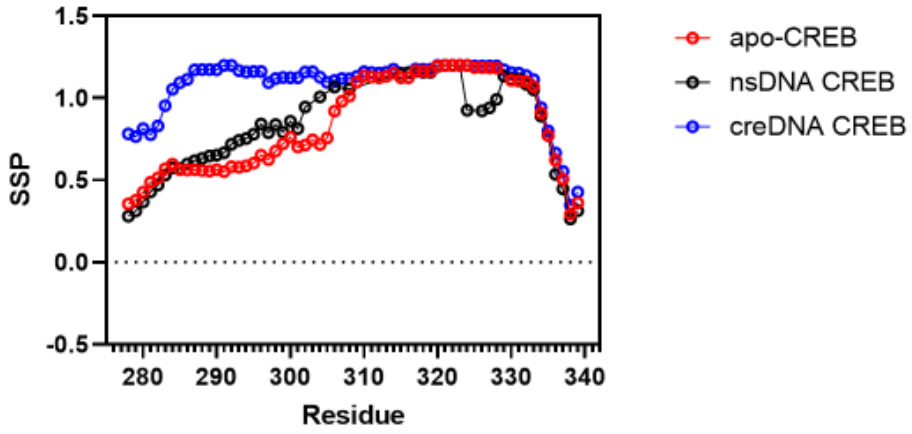


Figure 4.19. Secondary structure propensity (SSP) of CREB bZIP.
Calculated using: Secondary structure propensity (Marsh et al. 2006) (SSP) algorithm using the assigned $^{13}C\alpha$ chemical shifts (data provided by Conor Kelly). The BR and the LZ domains consist of residues 270-295 and 297-318, respectively, as determined by the uniprot.org database.

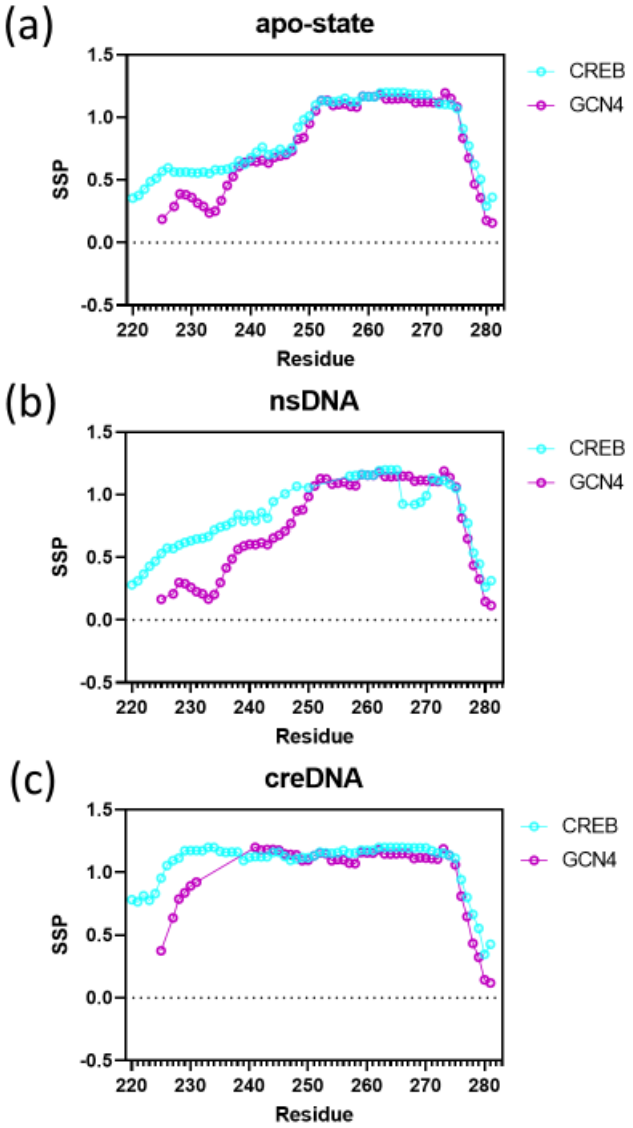


Figure 4.20. Secondary structure propensity (SSP) of the GCN4 and CREB bZIPs in the apo-state or in complex with nsDNA or creDNA.

CREB is shown in light blue and GCN4 is shown in magenta in (a) apo-state and in complex with: (b) nsDNA (c) For these calculations $^{13}C\alpha$ atoms alone were considered due to the lack of $^{13}C\beta$ and $^{1}H\alpha$ assignments for many residues in CREB. For GCN4, the BR and the LZ domains consist of residues 225-251 and 253-281, respectively, as determined by the uniprot.org database. For CREB, the BR and the LZ domains consist of residues 270-295 and 297-318, respectively, as determined by the uniprot.org database.

The SSP profile of CREB and GCN4 (Figure 4.19 and Figure 4.20) demonstrates a very conserved pattern in the LZ region with high helicity, in apo and DNA-bound states. It also shows that CREB bZIP has many similarities with GCN4 bZIP. CREB bound to creDNA gains a high level of helicity in the BR domain just like GCN4. The apo-state of the two different bZIPs have a similar pattern of SSP values, showing high helicity in the LZ then decreasing gradually towards the N-capping residues. The most significant difference between CREB and GCN4 is their behaviour in complex with nsDNA. As observed earlier in this chapter, GCN4 DNA binding residues are more disordered in complex with nsDNA than in the apo-state. By contrast, CREB DNA-binding residues (BR domain residues 285-310) show higher levels of helicity in the presence of nsDNA than in the apo state (Figure 4.19).

4.5 Conclusions and further work

In conclusion, the dynamic information (heteronuclear NOE) in combination with the secondary structure propensity profiles show a rigid and structured LZ in all of the systems. In the presence of creDNA, the BR shows significant structure in agreement with the previous X-ray structures and CD data. Apo-GCN4 has a less helical and more flexible BR, with lower amounts of helicity, and the addition of the nsDNA does not induce helical structure in this region.

Better conditions to acquire more NMR chemical shifts and assign all GCN4 residues, particularly in the presence of the creDNA, could be developed in future. For example, signal-to-noise can be improved if higher concentrations of protein are used. This may allow the assignment of more Hα resonances which are important for secondary structure characterisation. In addition to this, many Cβ atoms could not be assigned in the DNA-bound states, due to higher MW of the final complexes leading to poor signal-to-noise in 3D spectra. Deuteration, in addition to $^{13}C/^{15}N$ labelling, might give better quality HN(CO)CACB datasets allowing assignment of more Cβ peaks. Decreasing the overall MW of complexes, using truncated GCN4 bZIP constructs or shorter nucleotide oligos, may not be a suitable solution given that the GCN4 bZIP already has the minimum length required (See paragraph 2.3.1) and shortening of the DNA may affect the binding affinity or interaction mode. It is interesting that different secondary structure is observed between the GCN4 and CREB in the presence of the nsDNA. In order to confirm that these differences are real and are not induced due to different experimental conditions, further experiments could be performed using the Tris buffer conditions from CREB experiments.

An important investigation, to address the GCN4 bZIP behaviour bound to DNA, is to include different sequences of DNA. In this study, a palindromic non-target DNA sequence was used (nsDNA). The pool of different non-target sequences in the yeast genome though, is very large and involve non-palindromic sequences. Thus, it would be very interesting to test various non-target sequences. GCN4 has been also shown to bind to the AP-1 site and as demonstrated in the previous chapter, GCN4-AP-1 formation shows different kinetics results. As a result, NMR studies of AP-1 in complex with GCN4 would

be very interesting and provide detailed insight into GCN4 binding to a high affinity specific DNA site that differs from CRE by only one nucleotide. It is also possible, that the complex of GCN4 bZIP with the AP-1 DNA may not show the very extensive broadening observed for the creDNA and might allow a more detailed analysis of the chemical shift changes in the BR.

The NMR experiments in this study, utilised un-physiologically high protein concentrations which lead to GCN4 dimerization. Although these data highlight the behaviour of wild-type GCN4 bZIP in the apo or DNA bound state, a further investigation involving the BR-only peptide in the absence or presence of DNA needs to be made. In the case of CREB, the BR-only construct has been designed and investigated with NMR by Conor Kelly and the SSP results show that BR is highly disordered in the presence of either nsDNA or creDNA. Future work involving LZ-destabilised GCN4 bZIP mutants or the BR-only peptide would answer some important questions about how the GCN4 protein might interact with DNA through a monomer pathway.

There are more techniques that can offer important information for the GCN4 bZIP binding to DNA. For example, CD (circular dichroism) experiments allow rapid secondary structure characterisation demanding significantly lower protein concentration than NMR. This would allow characterisation of the interaction with a larger range of DNA sequences than is feasible using NMR. However, CD experiments cannot provide secondary structure details at a residual-specific level. In addition to this, ITC (isothermal titration calorimetry) is universally accepted as a reliable tool to assess the thermodynamic parameters of inter-molecular interactions without the need for labelling. A single titration experiment can allow the determination of the change of enthalpy (ΔH), the binding constant (K_a), and the stoichiometry (n) of the reaction and furthermore, the Gibbs free energy (ΔG) and the change in entropy (ΔS) can be calculated. This measurement offers detailed information for protein-DNA binding thermodynamics.

Word count=28175

Appendices

Table I Chemical shift table of assigned residues.

Residue number	Amino acid	H (ppm)	N (ppm)	Hα (ppm)	Hβ (ppm)	CO (ppm)	Cα (ppm)	Cβ (ppm)
225	Ser	-	-	4.51	3.84,3.84	173.66	58.03	63.85
226	Asp	8.48	123.59	4.92	2.65,2.80	-	51.95	41.28
227	Pro	-	-	4.34	1.98,2.38	177.99	64.29	32.07
228	Ala	8.31	122.01	4.22	-	179.02	53.71	18.46
229	Ala	7.87	121.81	4.16	-	179.24	53.66	18.61
230	Leu	7.86	119.92	4.22	1.71,1.60	178.30	56.29	41.88
231	Lys	7.95	120.69	4.16	1.85,1.85	177.48	57.66	32.57
232	Arg	8.03	120.43	4.23	1.86,1.86	177.24	57.41	30.49
233	Ala	8.13	123.97	4.28	-	178.69	53.48	18.61
234	Arg	8.23	119.56	4.25	1.87,1.87	177.18	57.18	30.45
235	Asn	8.42	119.69	4.75	2.90,2.90	176.48	53.93	38.40
236	Thr	8.16	115.39	4.39	-	175.68	63.51	69.40
237	Glu	8.39	123.00	4.18	2.08,2.08	177.95	58.41	29.33
238	Ala	8.25	122.65	4.14	-	179.82	54.36	18.33
239	Ala	8.03	121.87	4.24	-	179.74	54.18	18.38
240	Arg	8.11	119.80	4.10	1.95,1.95	178.66	58.53	30.23
241	Arg	8.28	120.59	4.22	1.91,1.91	178.25	58.20	30.31
242	Ser	8.18	115.68	4.31	3.99,3.99	176.44	60.36	63.00
243	Arg	8.18	122.31	4.17	1.93,1.93	177.61	58.39	30.19
244	Ala	8.06	122.25	4.21	-	179.70	54.19	18.37
245	Arg	8.10	119.11	4.20	1.95,1.95	178.29	58.11	30.24
246	Lys	8.05	121.22	4.16	1.93,1.93	178.02	58.61	32.48
247	Leu	8.13	120.69	4.22	1.82,1.61	179.10	56.93	41.76

248	Gln	8.13	120.19	4.01	2.17,2.17	177.70	57.85	28.69
249	Arg	8.20	120.90	4.22	1.97,1.97	177.88	58.28	29.87
250	Met	8.33	118.80	4.04	2.57,2.57	177.46	59.38	32.31
251	Lys	7.80	118.88	4.18	1.94,1.94	178.12	58.62	32.01
252	Gln	7.91	117.50	-	2.30,2.30	180.04	58.95	28.50
253	Leu	8.14	120.75	4.11	1.81,1.38	178.42	57.98	43.08
254	Glu	8.54	120.09	3.85	2.06,2.34	180.21	60.16	29.48
255	Asp	8.60	119.90	4.42	2.85,2.70	178.92	57.12	39.71
256	Lys	8.06	123.39	-	2.02,2.16	178.52	58.04	31.10
257	Val	8.45	119.89	3.44	-	177.28	67.81	31.57
258	Glu	7.66	117.43	4.01	2.17,2.17	179.60	59.74	29.03
259	Glu	8.20	120.90	4.11	2.28,2.28	179.65	59.25	30.01
260	Leu	8.70	120.25	4.04	2.11,2.11	179.61	57.83	43.22
261	Leu	9.04	122.06	4.07	1.97,1.49	179.97	58.14	41.94
262	Ser	7.78	114.63	4.34	4.10,4.10	177.79	61.56	62.59
263	Lys	8.27	122.12	4.23	1.91,1.91	178.79	59.59	32.72
264	Asn	8.83	119.44	4.40	3.30,2.76	176.69	57.18	38.87
265	Tyr	8.15	120.22	4.38	3.20,3.27	178.15	60.76	37.83
266	His	7.81	117.67	4.37	3.37,3.37	178.47	59.67	29.34
267	Leu	8.55	120.57	3.97	-	178.87	58.13	43.49
268	Glu	8.84	119.54	3.92	1.99,2.17	180.40	60.12	29.07
269	Asn	7.73	118.98	4.45	2.76,2.89	177.96	55.93	37.74
270	Glu	8.11	123.93	4.44	2.17,1.97	178.37	58.59	29.95
271	Val	8.68	120.17	3.39	-	177.46	67.59	31.59
272	Ala	7.73	119.91	4.02	-	180.48	55.39	17.84
273	Arg	7.95	119.70	3.98	1.89,1.89	179.48	59.40	30.68
274	Leu	8.46	120.44	3.98	1.32,1.32	179.20	57.98	43.82
275	Lys	9.00	118.66	3.83	1.85,1.85	179.66	60.81	32.48

276	Lys	7.36	117.63	4.14	1.99,1.90	179.07	58.90	32.23
277	Leu	7.46	118.89	4.19	1.64,2.08	179.16	57.22	42.55
278	Val	7.76	113.73	4.09	2.24	176.75	62.87	32.37
279	Gly	7.84	109.11	-	3.90,3.90	174.68	45.78	-
280	Glu	7.95	120.74	4.36	1.75,2.05	175.39	56.22	31.21
281	Arg	7.93	127.11	-	-	-	57.41	-

Table II Chemical shift table of assigned residues for the GCN4 bZIP-nsDNA complex.

Residue number	Amino acid	H (ppm)	N (ppm)	Hα (ppm)	Hβ (ppm)	CO (ppm)	Cα (ppm)	Cβ (ppm)
225	Ser	-	-	4.52	3.84,3.84	173.65	58.1	63.91
226	Asp	8.42	123.52	4.92	2.78,2.62	-	52.05	41.34
227	Pro	-	-	4.35	2.36,1.98	177.69	64.13	32
228	Ala	8.29	122.06	4.23	1.43	178.57	53.38	18.63
229	Ala	7.86	121.71	4.2	1.45	178.69	53.22	18.82
230	Leu	7.84	119.95	4.25	1.60,1.69	177.9	55.91	42.05
231	Lys	8	121.01	4.22	1.84,1.84	177.06	57.16	32.69
232	Arg	8.1	120.94	4.28	1.86,1.86	176.61	56.8	30.72
233	Ala	8.16	124.29	4.31	1.46	178.27	53.02	19.01
234	Arg	8.26	119.65	4.3	1.87,1.87	176.82	56.76	30.66
235	Asn	8.5	120	4.77	2.90,2.93	176.2	53.73	38.46
236	Thr	8.16	115.25	4.25	4.36	175.56	63.34	69.47
237	Glu	8.37	122.79	4.19	2.08,2.08	177.63	58.21	29.56
238	Ala	8.18	122.89	4.16	1.47	179.25	54.06	18.5
239	Ala	8.08	122.05	4.24	1.47	179.24	53.86	18.6
240	Arg	8.13	119.56	4.17	1.96,1.96	178.17	58.08	30.39
241	Arg	8.29	120.9	4.26	1.93,1.93	177.75	57.8	30.47
242	Ser	8.23	115.99	-	3.98,4.00	176.12	60.01	63.38

243	Arg	8.3	122.45	-	-	177.3	58.01	30.36
244	Ala	8.14	122.81	4.25	1.49	179.27	53.96	18.6
245	Arg	8.17	119.43	-	-	178.1	57.91	30.37
246	Lys	8.19	121.65	-	-	177.87	58.54	32.68
247	Leu	8.19	120.72	-	-	179.03	56.88	41.65
248	Gln	8.2	120.28	-	-	177.86	57.97	28.52
249	Arg	8.24	120.91	-	-	178.06	58.66	-
250	Met	8.29	118.43	-	-	177.49	59.25	-
251	Lys	7.78	118.85	-	-	178.24	58.66	31.8
252	Gln	7.97	117.51	-	-	179.95	59.03	-
253	Leu	8.13	120.65	-	-	178.42	57.95	43.18
254	Glu	8.5	119.97	-	-	180.1	60.21	29.4
255	Asp	8.57	119.89	-	-	178.92	57.12	39.74
256	Lys	8.05	123.14	-	-	178.5	58.16	30.72
257	Val	8.45	119.87	3.45	-	177.25	67.75	31.39
258	Glu	7.69	117.6	-	-	179.54	59.76	29.22
259	Glu	8.2	120.88	-	-	179.59	59.24	29.96
260	Leu	8.68	120.28	-	-	179.57	57.84	-
261	Leu	9.04	122.06	-	1.49,1.97	179.9	58.17	41.96
262	Ser	7.8	114.52	4.35	-	177.59	61.61	-
263	Lys	8.25	122.07	-	-	178.71	59.56	32.7
264	Asn	8.81	119.3	-	-	176.74	57.18	38.97
265	Tyr	8.17	120.22	4.38	3.19,3.27	178.14	60.73	37.79
266	His	7.81	118	-	3.33,3.37	178.54	59.77	29.65
267	Leu	8.55	120.34	-	-	178.85	58.05	-
268	Glu	8.84	119.56	-	-	180.31	60.12	29.13
269	Asn	7.72	118.86	-	-	177.89	55.95	37.96
270	Glu	8.12	123.83	-	-	178.35	58.56	29.93

271	Val	8.67	119.99	3.41	-	177.42	67.58	-
272	Ala	7.72	119.94	4.01	1.5	180.42	55.34	17.91
273	Arg	7.94	119.52	3.84	-	179.37	59.37	30.6
274	Leu	8.44	120.28	-	-	179.19	57.93	43.78
275	Lys	8.97	118.53	3.84	-	179.58	60.71	32.32
276	Lys	7.36	117.66	4.13	1.93,1.93	178.95	58.74	32.2
277	Leu	7.47	118.87	4.2	2.05	179.1	57.16	42.52
278	Val	7.75	113.72	4.1	2.24	176.69	62.81	32.33
279	Gly	7.83	109.06	3.89,4.11	-	174.62	45.77	-
280	Glu	7.94	120.67	4.37	2.05,1.76	175.33	56.22	31.18
281	Arg	7.87	127.02	-	-	-	57.33	-

Table III Chemical shift table of assigned residues for the GCN4 bZIP-creDNA complex.

Residue number	Amino acid	H (ppm)	N (ppm)	Hα (ppm)	Hβ (ppm)	CO (ppm)	Cα (ppm)	Cβ (ppm)
225	Ser			4.54	3.84,3.85	173.45	58.06	63.98
226	Asp	8.44	123.34	4.96	-	-	51.97	41.34
227	Pro	-	-	4.34	2.43,2.03	178.59	64.72	32.02
228	Ala	8.30	121.79	4.23	1.50	179.61	54.43	18.34
229	Ala	7.88	121.62	4.14	1.54	180.33	54.47	18.47
230	Leu	7.86	120.06	4.23	1.77,1.67	178.69	57.26	41.7
231	Lys	7.89	120.53	-	-	178.72	59.11	-
232	Arg	8.03	119.91	-	-	-	-	-
233	Ala	8.06	122.05	-	-	-	-	-
234	Arg	8.24	118.28	-	-	-	-	-

235	Asn	-	-	-	-	-	-	-
236	Thr	-	-	-	4.43	-	-	68.61
237	Glu	-	-	-	-	-	-	-
238	Ala	8.25	122.78	-	-	-	-	-
239	Ala	-	-	-	-	-	-	-
240	Arg	-	-	-	-	-	-	-
241	Arg	-	-	4.19	-	178.89	59.47	30.60
242	Ser	8.36	115.57	-	-	-	61.62	-
243	Arg	-	-	-	-	179.21	60.23	-
244	Ala	7.92	122.13	4.29	1.60	180.56	55.07	17.91
245	Arg	8.17	119.26	-	-	179.23	59.44	30.44
246	Lys	8.03	121.01	-	-	178.81	59.48	32.22
247	Leu	7.86	120.33	4.23	1.78	179.67	57.97	41.72
248	Gln	8.09	119.06	4.13	2.24	178.33	58.60	-
249	Arg	8.25	120.63	4.17	2.04	178.43	59.16	29.86
250	Met	8.07	118.50	4.15	-	177.67	59.14	32.26
251	Lys	7.93	119.15	4.19	-	178.48	58.82	31.84
252	Gln	8.29	117.94	4.18	2.21,2,35	179.89	59.06	28.33
253	Leu	8.17	121.14	4.16	-	178.48	58.09	42.90
254	Glu	8.37	119.53	3.90	2.12	180.17	60.23	29.56
255	Asp	8.62	119.85	4.45	-	178.91	57.09	39.98
256	Lys	8.23	123.20	4.25	-	178.49	58.11	31.30
257	Val	8.38	119.61	3.47	-	177.23	67.77	31.46
258	Glu	7.59	117.44	4.03	2.19	179.57	59.72	29.28
259	Glu	8.27	120.92	4.11	2.30	179.61	59.27	29.99
260	Leu	8.75	120.22	4.04	-	179.60	57.81	43.13
261	Leu	9.01	122.05	4.08	1.97	179.92	58.19	42.00
262	Ser	7.79	114.49	4.11	-	177.57	61.60	62.77

263	Lys	8.30	122.08	4.24	-	178.70	59.55	32.75
264	Asn	8.83	119.26	4.42	-	176.74	57.20	38.96
265	Tyr	8.15	120.20	4.38	3.30,3.19	178.15	60.78	37.76
266	His	7.82	117.93	4.37	3.34	178.53	59.74	29.69
267	Leu	8.59	120.39	3.98	-	178.86	58.08	43.62
268	Glu	8.85	119.54	3.92	-	180.32	60.09	29.18
269	Asn	7.72	118.83	4.44	2.76,2.88	177.89	55.99	37.83
270	Glu	8.13	123.81	4.43	2.15,1.99	178.36	58.58	29.93
271	Val	8.68	120.00	3.41	2.12	177.42	67.57	31.45
272	Ala	7.73	119.93	4.02	1.50	180.42	55.34	17.92
273	Arg	7.95	119.51	3.97	-	179.36	59.41	30.58
274	Leu	8.46	120.28	3.98	-	179.19	57.91	43.86
275	Lys	8.98	118.54	3.83	1.86	179.58	60.76	32.44
276	Lys	7.37	117.66	4.14	1.96	178.95	58.76	32.20
277	Leu	7.48	118.87	4.20	1.63,2.06	179.10	57.18	42.56
278	Val	7.76	113.71	4.09	2.25	176.69	62.80	32.34
279	Gly	7.84	109.05	3.90,4.11	-	174.63	45.78	-
280	Glu	7.95	120.67	4.37	1.76,2.05	175.33	56.23	31.19
281	Arg	7.88	127.01	-	-	-	57.32	-

Table IV Assignment report of GCN4 bZIP (224-281) The construct feature 58 residues, including the N-terminal Gly residue. Obtained from CCPNmr v2.4.2.

Category	% Assigned		
	apo-state	GCN4-nsDNA	GCN4-creDNA
All Residues	98.3	98.3	89.7
Type C	94.8	54.4	77.6
Type CA	98.3	98.3	84.5
Type CB	98.2	76.8	76.8
Type H	98.2	35.0	83.9
Type HA	91.7	51.7	71.7
Type HB	98.0	44.6	48.5
Type N	94.8	94.8	81.0

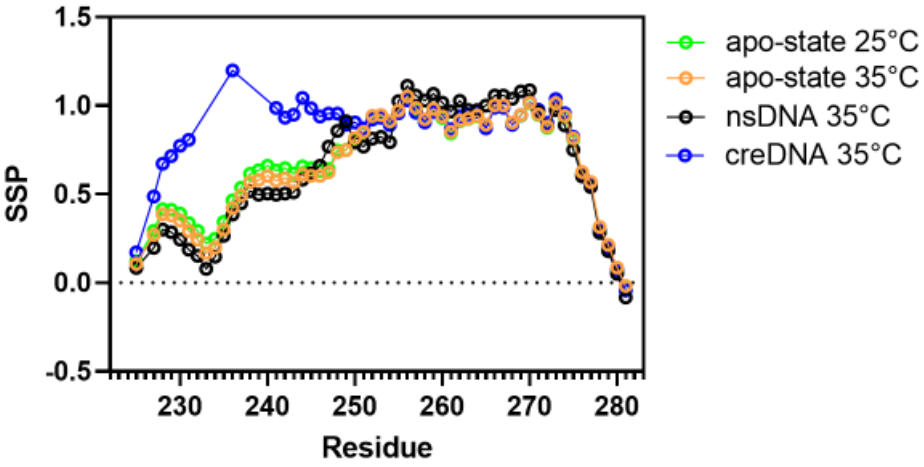


Figure I. Secondary structure propensity of GCN4.
Calculated using: Secondary Structure Propensity (Marsh et al. 2006) (SSP) algorithm using the assigned chemical shifts information obtained from the $^{13}C\alpha$, $^{13}C\beta$ and $^{1}H\alpha$ backbone triple resonance experiments.

References

Amoutzias, G. D., Veron, A. S., Weiner, J., Robinson-Rechavi, M., Bornberg-Bauer, E., Oliver, S. G., & Robertson, D. L. (2007). One billion years of bZIP transcription factor evolution: Conservation and change in dimerization and DNA-binding site specificity. *Molecular Biology and Evolution*, *24*(3), 827–835. https://doi.org/10.1093/molbev/msl211

Araki, K., Yagi, N., Aoyama, K., Choong, C. J., Hayakawa, H., Fujimura, H., Nagai, Y., Goto, Y., & Mochizuki, H. (2019). Parkinson's disease is a type of amyloidosis featuring accumulation of amyloid fibrils of α-synuclein. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(36), 17963–17969. https://doi.org/10.1073/pnas.1906124116

Arndt, K., & Fink, G. R. (1986). GCN4 protein, a positive transcription factor in yeast, binds general control promoters at all 5' TGACTC 3' sequences. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(22), 8516–8520. https://doi.org/10.1073/pnas.83.22.8516

Ashraf, G., Greig, N., Khan, T., Hassan, I., Tabrez, S., Shakil, S., Sheikh, I., Zaidi, S., Akram, M., Jabir, N., Firoz, C., Naeem, A., Alhazza, I., Damanhouri, G., & Kamal, M. (2014). Protein Misfolding and Aggregation in Alzheimer's Disease and Type 2 Diabetes Mellitus. *CNS & Neurological Disorders - Drug Targets*, *13*(7), 1280–1293. https://doi.org/10.2174/18715273136661409170955 14

Bellomy, G. R., & Record, M. T. (1990). Stable DNA Loops in Vivo and in Vitro: Roles in Gene Regulation at a Distance and in Biophysical Characterization of DNA. *Progress in Nucleic Acid Research and Molecular Biology*, *39*(C), 81–128. https://doi.org/10.1016/S0079-6603(08)60624-8

Berg, O. G., Winter, R. B., & von Hippel, P. H. (1981). Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 1. Models and Theory. *Biochemistry*, *20*(24), 6929–6948. https://doi.org/10.1021/bi00527a028

Berger, C., Piubelli, L., Haditsch, U., & Rudolf Bosshard, H. (1998). Diffusion-controlled DNA recognition by an unfolded, monomeric bZIP transcription factor. *FEBS Letters*, *425*(1), 14–18. https://doi.org/10.1016/S0014-5793(98)00156-2

Bracken, C., Carr, P. A., Cavanagh, J., & Palmer, A. G. (1999). Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: Implications for the entropy of association with DNA. *Journal of Molecular Biology*, *285*(5), 2133–2146. https://doi.org/10.1006/jmbi.1998.2429

Cao, W., Liu, L., Klein, D. E., Wei, L., & Lai, L. (2000). The monomer of the yeast transcriptional activator GCN4 recognizes its dimer binding DNA target sites without dimerization. *Thermochimica Acta*, *360*(1), 47–56. https://doi.org/10.1016/S0040-6031(00)00538-4

Chang, Y. N., Dong, D. L., Hayward, G. S., & Hayward, S. D. (1990). The Epstein-Barr virus Zta transactivator: A member of the bZIP family with unique DNA-binding specificity and a dimerization domain that lacks the characteristic heptad leucine zipper motif. *Journal of Virology*, *64*(7), 3358–3369. https://doi.org/10.1128/jvi.64.7.3358-3369.1990

Chong, S.-H., & Ham, S. (2019). Folding Free Energy Landscape of Ordered and Intrinsically Disordered Proteins. *Scientific Reports*, *9*(1), 14927. https://doi.org/10.1038/s41598-019-50825-6

Crabtree, M. D., & Shammas, S. L. (2018). Stopped-Flow Kinetic Techniques for Studying Binding Reactions of Intrinsically Disordered Proteins. In *Methods in Enzymology*

(Vol. 611, pp. 423–457). Academic Press Inc. https://doi.org/10.1016/bs.mie.2018.09.026

Cranz, S., Berger, C., Baici, A., Jelesarov, I., & Bosshard, H. R. (2004). Monomeric and Dimeric bZIP Transcription Factor GCN4 Bind at the Same Rate to Their Target DNA Site. *Biochemistry*, *43*(3), 718–727. https://doi.org/10.1021/bi0355793

Davudian, S., Mansoori, B., Shajari, N., Mohammadi, A., & Baradaran, B. (2016). BACH1, the master regulator gene: A novel candidate target for cancer therapy. *Gene*, *588*(1), 30–37. https://doi.org/10.1016/j.gene.2016.04.040

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., & Bax, A. (1995). NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, *6*(3), 277–293. https://doi.org/10.1007/BF00197809

Dragan, Anatoly I., Frank, L., Liu, Y., Makeyeva, E. N., Crane-Robinson, C., & Privalov, P. L. (2004). Thermodynamic signature of GCN4-bZIP binding to DNA indicates the role of water in discriminating between the AP-1 and ATF/CREB sites. *Journal of Molecular Biology*, *343*(4), 865–878. https://doi.org/10.1016/j.jmb.2004.08.101

Dragan, Antony I, Liu, Y., Mekeyeva, E. N., & Privalov, P. L. (2004). DNA-binding domain of GCN4 induces bending of both the ATF-CREB and AP-1 binding sites of DNA. *Nucleic Acids Research*, *32*(17), 5192–5197. https://doi.org/10.1093/nar/gkh854

Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., & Obradović, Z. (2002). Intrinsic disorder and protein function. *Biochemistry*, *41*(21), 6573–6582. https://doi.org/10.1021/bi012159+

Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C. H., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu,

W., Garner, E. C., & Obradovic, Z. (2001). Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling*, *19*(1), 26–59. https://doi.org/10.1016/S1093-3263(00)00138-8

Ellenberger, T. E., Brandl, C. J., Struhl, K., & Harrison, S. C. (1992). The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α Helices: Crystal structure of the protein-DNA complex. *Cell*, *71*(7), 1223–1237. https://doi.org/10.1016/S0092-8674(05)80070-4

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345. https://doi.org/10.1038/nmeth.1318

Gill, M. L., Byrd, R. A., & Palmer, A. G. (2016). Dynamics of GCN4 facilitate DNA interaction: A model-free analysis of an intrinsically disordered region. *Physical Chemistry Chemical Physics*, *18*(8), 5839–5849. https://doi.org/10.1039/c5cp06197k

Gill, S. C., & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry*, *182*(2), 319–326. https://doi.org/10.1016/0003-2697(89)90602-7

Hakoshima, T., Fujii, Y., Shimizu, T., Toda, T., Yanagida, M., & Hakoshima, T. (2000). Structural basis for the diversity of DNA recognition by bZIP transcription factors. *Nature Structural Biology*, *7*(10), 889–893. https://doi.org/10.1038/82822

Han, J., Back, S. H., Hur, J., Lin, Y. H., Gildersleeve, R., Shan, J., Yuan, C. L., Krokowski, D., Wang, S., Hatzoglou, M., Kilberg, M. S., Sartor, M. A., & Kaufman, R. J. (2013). ER-stress-induced transcriptional regulation increases protein synthesis leading to cell death. *Nature Cell Biology*, *15*(5), 481–490. https://doi.org/10.1038/ncb2738

Haynes, C., Oldfield, C. J., Ji, F., Klitgord, N., Cusick, M. E., Radivojac, P., Uversky, V. N., Vidal, M., & Iakoucheva, L. M. (2006). Intrinsic Disorder Is a Common Feature of Hub Proteins from Four Eukaryotic Interactomes. *PLoS Computational Biology*, *2*(8). https://doi.org/10.1371/journal.pcbi.0020100

Hill, D., Hope, I., Macke, J., & Struhl, K. (1986). Saturation mutagenesis of the yeast his3 regulatory site: Requirements for transcriptional induction and for binding by GCN4 activator protein. *Science*, *234*(4775), 451–457. https://doi.org/10.1126/science.3532321

Hinnebusch, A. G. (1984). Evidence for translational regulation of the activator of general amino acid control in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *81*(20 I), 6442–6446. https://doi.org/10.1073/pnas.81.20.6442

Hollenbeck, J. J., McClain, D. L., & Oakley, M. G. (2009). The role of helix stabilizing residues in GCN4 basic region folding and DNA binding. *Protein Science*, *11*(11), 2740–2747. https://doi.org/10.1110/ps.0211102

Hollenbeck, J. J., & Oakley, M. G. (2000). GCN4 binds with high affinity to DNA sequences containing a single consensus half-site. *Biochemistry*, *39*(21), 6380–6389. https://doi.org/10.1021/bi992705n

Hope, I. A., & Struhl, K. (1986). Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of Yeast. *Cell*, *46*(6), 885–894. https://doi.org/10.1016/0092-8674(86)90070-X

Igarashi, K., Kurosaki, T., & Roychoudhuri, R. (2017). BACH transcription factors in innate and adaptive immunity. *Nature Reviews Immunology*, *17*(7), 437–450. https://doi.org/10.1038/nri.2017.26

Iwama, H., & Gojobori, T. (2004). Highly conserved upstream sequences for transcription factor genes and implications for the regulatory network. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(49), 17156–17161. https://doi.org/10.1073/pnas.0407670101

Kaser, A., Lee, A. H., Franke, A., Glickman, J. N., Zeissig, S., Tilg, H., Nieuwenhuis, E. E. S., Higgins, D. E., Schreiber, S., Glimcher, L. H., & Blumberg, R. S. (2008). XBP1 Links ER Stress to Intestinal Inflammation and Confers Genetic Risk for Human Inflammatory Bowel Disease. *Cell*, *134*(5), 743–756. https://doi.org/10.1016/j.cell.2008.07.021

Keller, W., König, P., & Richmond, T. J. (1995). Crystal structure of a bZIP/DNA complex at 2.2 Å: Determinants of DNA specific recognition. *Journal of Molecular Biology*, *254*(4), 657–667. https://doi.org/10.1006/jmbi.1995.0645

Kohler, J. J., & Schepartz, A. (2001). Kinetic studies of Fos·Jun·DNA complex formation: DNA binding prior to dimerization. *Biochemistry*, *40*(1), 130–142. https://doi.org/10.1021/bi001881p

König, P., & Richmond, T. J. (1993). The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. *Journal of Molecular Biology*, *233*(1), 139–154. https://doi.org/10.1006/jmbi.1993.1490

Lajmi, A. R., Lovrencic, M. E., Wallace, T. R., Thomlinson, R. R., & Shin, J. A. (2000). Minimalist, alanine-based, helical protein dimers bind to specific DNA sites. *Journal of the American Chemical Society*, *122*(23), 5638–5639. https://doi.org/10.1021/ja993025a

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

Marsh, J. A., Singh, V. K., Jia, Z., & Forman-Kay, J. D. (2006). Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein Science*, *15*(12), 2795–2804. https://doi.org/10.1110/ps.062465306

Metallo, S. J., & Schepartz, A. (1997). Certain bZIP peptides bind DNA sequentially as monomers and dimerize on the DNA. *Nature Structural Biology*, *4*(2), 115–117. https://doi.org/10.1038/nsb0297-115

Miller, M. (2009). The Importance of Being Flexible: The Case of Basic Region Leucine Zipper Transcriptional Regulators. In *Current protein & peptide science* (Vol. 10, Issue 3, pp. 244–244). NIH Public Access. http://www.disprot.org

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

Mittal, N., Guimaraes, J. C., Gross, T., Schmidt, A., Vina-Vilaseca, A., Nedialkova, D. D., Aeschimann, F., Leidel, S. A., Spang, A., & Zavolan, M. (2017). The Gcn4 transcription factor reduces protein synthesis capacity and extends yeast lifespan. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-00539-y

Moran, L. B., Schneider, J. P., Kentsis, A., Reddy, G. A., & Sosnick, T. R. (1999). Transition state heterogeneity in GCN4 coiled coil folding studied by using multisite mutations and crosslinking. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(19), 10699–10704. https://doi.org/10.1073/pnas.96.19.10699

Muñoz, V., & Serrano, L. (1994). Elucidating the folding problem of helical peptides using empirical parameters. *Nature Structural Biology*, *1*(6), 399–409. https://doi.org/10.1038/nsb0694-399

Natarajan, K., Meyer, M. R., Jackson, B. M., Slade, D., Roberts, C., Hinnebusch, A. G., & Marton, M. J. (2001). Transcriptional Profiling Shows that Gcn4p Is a Master Regulator of Gene Expression during Amino Acid Starvation in Yeast. *Molecular and Cellular Biology*, *21*(13), 4347–4368. https://doi.org/10.1128/mcb.21.13.4347-4368.2001

Newman, J. R. S., & Keating, A. E. (2003). Comprehensive identification of human bZIP interactions with coiled-coil arrays. *Science*, *300*(5628), 2097–2101. https://doi.org/10.1126/science.1084648

Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C. J., & Dunker, A. K. (2003). Predicting Intrinsic Disorder From Amino Acid Sequence. *Proteins: Structure, Function and Genetics*, *53*, 566–572. https://doi.org/10.1002/prot.10532

Oliphant, A. R., Brandl, C. J., & Struhl, K. (1989). Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: Analysis of yeast GCN4 protein. *Molecular and Cellular Biology*, *9*(7), 2944–2949. https://doi.org/10.1128/mcb.9.7.2944

Park, C., Campbell, J. L., & Goddard, W. A. (1996). Can the monomer of the leucine zipper proteins recognize the dimer binding site without dimerization? *Journal of the American Chemical Society*, *118*(18), 4235–4239. https://doi.org/10.1021/ja950653t

Radivojac, P., Iakoucheva, L. M., Oldfield, C. J., Obradovic, Z., Uversky, V. N., & Dunker, A. K. (2007). Intrinsic disorder and functional proteomics. *Biophysical Journal*, *92*(5), 1439–1456. https://doi.org/10.1529/biophysj.106.094045

Redfield, C. (2015). Assignment of protein NMR spectra using heteronuclear NMR-A tutorial. In *Protein NMR: Modern Techniques and Biomedical Applications* (pp. 1–42). Springer International Publishing. https://doi.org/10.1007/978-1-4899-7621-5_1

Robustelli, P., Trbovic, N., Friesner, R. A., & Palmer, A. G. (2013). Conformational dynamics of the partially disordered yeast transcription factor GCN4. *Journal of Chemical Theory and Computation*, *9*(11), 5190–5200. https://doi.org/10.1021/ct400654r

Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., & Dunker, A. K. (2001). Sequence complexity of disordered protein. *Proteins: Structure, Function and Genetics*, *42*(1), 38–48. https://doi.org/10.1002/1097-0134(20010101)42:1<38::AID-PROT50>3.0.CO;2-3

Roussou, I., Thireos, G., & Hauge, B. M. (1988). Transcriptional-translational regulatory circuit in Saccharomyces cerevisiae which involves the GCN4 transcriptional activator and the GCN2 protein kinase. *Molecular and Cellular Biology*, *8*(5), 2132–2139. https://doi.org/10.1128/mcb.8.5.2132

Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F., & Schroeder, M. (2015). PLIP: Fully automated protein-ligand interaction profiler. *Nucleic Acids Research*, *43*(W1), W443-447. https://doi.org/10.1093/nar/gkv315

Saudek, V., Pastore, A., Morelli, M. A. C., Frank, R., Gausepohl, H., & Gibson, T. (1991). The solution structure of a leucine-zipper motif peptide. *Protein Engineering, Design and Selection*, *4*(5), 519–529. https://doi.org/10.1093/protein/4.5.519

Schumacher, M. A., Goodman, R. H., & Brennan, R. G. (2000). The structure of a CREB bZIP·somatostatin CRE complex reveals the basis for selective dimerization and

divalent cation-enhanced DNA binding. *Journal of Biological Chemistry*, *275*(45), 35242–35247. https://doi.org/10.1074/jbc.M007293200

Sellers, J. W., Vincent, A. C., & Struhl, K. (1990). Mutations that define the optimal half-site for binding yeast GCN4 activator protein and identify an ATF/CREB-like repressor that recognizes similar DNA sites. *Molecular and Cellular Biology*, *10*(10), 5077–5086. https://doi.org/10.1128/mcb.10.10.5077

Solomon, G. S. (2019). Two are better than one. *Nature Physics*, *15*(9), 882–883. https://doi.org/10.1038/s41567-019-0566-9

Sprague, B. L., Müller, F., Pego, R. L., Bungay, P. M., Stavreva, D. A., & McNally, J. G. (2006). Analysis of binding at a single spatially localized cluster of binding sites by fluorescence recovery after photobleaching. *Biophysical Journal*, *91*(4), 1169–1191. https://doi.org/10.1529/biophysj.105.073676

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques*, *28*(6), 1102, 1104. https://doi.org/10.2144/00286ir01

Sun, X., Liu, J., Crary, J. F., Malagelada, C., Sulzer, D., Greene, L. A., & Levy, O. A. (2013). ATF4 protects against neuronal death in cellular Parkinson's disease models by maintaining levels of parkin. *Journal of Neuroscience*, *33*(6), 2398–2407. https://doi.org/10.1523/JNEUROSCI.2292-12.2013

Talanian, R. V., Mcknight, C. J., & Kim, P. S. (1990). Sequence-specific DNA binding by a short peptide dimer. *Science*, *249*(4970), 769–771. https://doi.org/10.1126/science.2389142

Uversky, V. N., Gillespie, J. R., & Fink, A. L. (2000). Why are 'natively unfolded' proteins unstructured under physiologic conditions? *Proteins: Structure, Function and*

Genetics, *41*(3), 415–427. https://doi.org/10.1002/1097-0134(20001115)41:3<415::AID-PROT130>3.0.CO;2-7

Vacic, V., Uversky, V. N., Dunker, A. K., & Lonardi, S. (2007). Composition Profiler: A tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics*, *8*(1), 211–211. https://doi.org/10.1186/1471-2105-8-211

Vinson, C., Myakishev, M., Acharya, A., Mir, A. A., Moll, J. R., & Bonovich, M. (2002). Classification of Human B-ZIP Proteins Based on Dimerization Properties. *Molecular and Cellular Biology*, *22*(18), 6321–6335. https://doi.org/10.1128/mcb.22.18.6321-6335.2002

Vologodskii, A., & Cozzarelli, N. R. (1996). Effect of supercoiling on the juxtaposition and relative orientation of DNA sites. *Biophysical Journal*, *70*(6), 2548–2556. https://doi.org/10.1016/S0006-3495(96)79826-0

Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., & Laue, E. D. (2005). The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function and Genetics*, *59*(4), 687–696. https://doi.org/10.1002/prot.20449

Wakankar, A. A., & Borchardt, R. T. (2006). Formulation considerations for proteins susceptible to asparagine deamidation and aspartate isomerization. *Journal of Pharmaceutical Sciences*, *95*(11), 2321–2336. https://doi.org/10.1002/jps.20740

Wang, X., Cao, W., Cao, A., & Lai, L. (2003). Thermodynamic characterization of the folding coupled DNA binding by the monomeric transcription activator GCN4 peptide. *Biophysical Journal*, *84*(3), 1867–1875. https://doi.org/10.1016/S0006-3495(03)74994-7

Wang, Y. (2002). Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Science*, *11*(4), 852–861. https://doi.org/10.1110/ps.3180102

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

Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C., & Struhl, K. (1990). Folding transition in the DMA-binding domain of GCN4 on specific binding to DNA. *Nature*, *347*(6293), 575–578. https://doi.org/10.1038/347575a0

Williams, R. M., Obradovi, Z., Mathura, V., Braun, W., Garner, E. C., Young, J., Takayama, S., Brown, C. J., & Dunker, A. K. (2001). The protein non-folding problem: Amino acid determinants of intrinsic order and disorder. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 89–100. https://doi.org/10.1142/9789814447362_0010

Zhang, T., Faraggi, E., Li, Z., & Zhou, Y. (2013). Intrinsically Semi-disordered State and Its Role in Induced Folding and Protein Aggregation. *Cell Biochemistry and Biophysics*, *67*(3), 1193–1205. https://doi.org/10.1007/s12013-013-9638-0

Zhao, M., Liu, Q., Liang, G., Wang, L., Luo, S., Tang, Q., Zhao, H., Su, Y., Yung, S., Chan, T. M., & Lu, Q. (2013). E4BP4 overexpression: A protective mechanism in CD4+ T cells from SLE patients. *Journal of Autoimmunity*, *41*, 152–160. https://doi.org/10.1016/j.jaut.2013.01.004