Mechanisms of transcription regulation by full-length CREB



Jeremie Gaudez

The Queen's College

University of Oxford

A Thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2023

To my family, full of doctors

Acknowledgements

Firstly, I would like to thank my primary and secondary supervisors, Dr Sarah Shammas and Prof Jane Mellor, for their support, insightful comments, and valuable advice throughout this process. I am grateful to Sarah for her guidance and for her trust in allowing me to undertake this project, and thanks to whom I have greatly grown as a scientist. This process would have been much harder without advice from my Thesis committee, Dr Tim Nott, who generously suggested and donated the SUMO template, and Prof Christina Redfield. Both routinely took time to contribute to the project and made always wise and helpful suggestions. I am also indebted to the Shammas lab members both past and present for their presence throughout this project and stimulating, intelligent discussions. Especially, Chirs for helping shape the project at the beginning, and both Mikhail and Conor for their great advice and help in setting up experiments, ensuring continued success, and their steady presence whenever I needed them. They've given me great wisdom that I will carry forward with me, and hopefully transmit to the next generation of scientists. I give thanks to David Staunton for his experimental demonstrations and SEC-MALS, AUC and DSC contributions, and Dr. Gemma Harris for her Fidabio contribution. I want to dedicate a special mention to the neighbouring members of the Berks lab for being so welcoming, friendly, and helpful. I would also like to thank the Queen's College academic office for their support.

Of course, this would have been impossible without extraordinary friends and family, present through thick and thin. Thanks especially to the J's and honorary J members: Jack, Jacques, Jochem, Jai, Nic, Andrew, for making Oxford a better place; and to my friends dotted across Europe but there since the beginning, especially Lawrence. I can only hope I've been half as good to them as they've been to me. Heartfelt thanks to my family (and the family I've been welcomed into since), my parents, my brother, and my grandparents especially, who never stopped to support or believe in me once, giving me great strength. Last, but not least, heartfelt, loving gratitude to a certain special PhD student in The Other Place, Sara, whom I cannot wait to spend the rest of my life with.

Abstract

CREB (cAMP response element binding-protein) is a disordered essential mammalian transcription factor, activated by the cAMP signalling cascade. It is involved in a myriad of biological processes including memory and the circadian clock. Moreover, CREB is overexpressed in many cancers, such as leukaemia, sarcomas, melanoma. Thus, to understand both mammalian and cancer biology, it is important to characterise it well.

The C-terminal end of CREB contains a bZIP DNA-binding domain (DBD), and the N-terminal end of CREB contains three domains - two glutamine-rich domains surround its acidic activation domain KID (kinase inducible domain). Previous biophysical (and structural) studies have focused on understanding CREB's behaviour using isolated KID and bZIP domains, but few have used full-length CREB or looked at the influence of the other domains on its biophysical properties. Recent work has challenged simplistic modular architectural understandings of transcription factors, for example: there is increasing evidence that these disordered sequences are essential for efficient target search. This Thesis aims to directly address this issue through comparative studies with full-length CREB and its isolated domains.

In this Thesis I developed a high yield methodology for production of highly pure full-length CREB protein and used the material to conduct thorough biophysical characterisation of CREB and its binding to DNA *in-vitro*. I found CREB's non-DBDs lower DNA affinity around 5-fold compared to the isolated bZIP, but the homodimerization and target DNA affinity remain in the nanomolar range whereas affinity to non-target is in the micromolar range. Altogether, this makes the full-length CREB target search more efficient and less likely to be sequestered away by non-target DNA in a cellular context. Moreover, each of CREB's non-bZIP domains impedes binding to DNA to a similar extent. Finally, current literature suggests bZIP transcription factors dimerise following target binding through a monomer search pathway. However, I find a high CREB homodimer stability which, together with unquantifiable binding of CREB monomers to DNA, favours a dimeric pathway instead.

Declaration of own work

This work presented in my Thesis is my own, apart from when otherwise stated. I show some results from Richards, 1996, and some data acquired with CREB bZIP at pH 6.5 for comparison by Dr Conor Kelly and Dr Mikhail Kuravsky. Dr David Staunton carried out AUC, SEC-MALS and DSC experiments. Dr. Gemma Harris carried out Fidabio experiments. Mr. Chris Chung designed the initial HaloTag-CREB purification protocols. Dr Sarah Shammas derived equations used throughout this Thesis if not already present in the literature.

Table of Contents

List of Figures xii

List of Tables xvi

List of Abbreviations xvii

Chapter 1 Introduction 1

1.1. Transcription factors and their crucial role within the central dogma of molecular biology 1

1.2. Typical characteristics of transcription factors 4

1.3. Disorder in transcription factors 5

1.4. The target search process 7

1.5. Effects of disordered domains on target search 9

1.6. The CREB transcription factor 13

1.7. Overview of Thesis 19

Chapter 2 Methods 21

2.1. Cloning 21

2.1.1. Typical cloning procedure 21

2.1.2. Homemade competent cells 22

2.2. Protein and oligonucleotides preparation and purification 23

2.2.1. Full-length CREB and domain deletions purification protocol 23

2.2.2. bZIP-specific purification protocol 27

2.2.3. Protease purifications 30

2.2.4. Preparation of oligonucleotides 31

2.2.5. Maleimide labelling 32

2.2.6. Forced dimer generation 33

2.2.7. Western blot 34

2.3. Biophysical methods 34

2.3.1. Determination of extinction coefficients 35

2.3.2. Biophysical values relevant for the study of CREB interactions 36

2.3.3. Fluorescence anisotropy equilibrium titrations 37
2.3.4. Stopped-flow kinetics 40
2.3.5. SEC-MALS 46
2.3.6. Analytical ultracentrifugation (AUC) sedimentation velocity 47
2.3.7. Differential scanning calorimetry (DSC) 47
2.3.8. Isothermal titration calorimetry (ITC) 47
2.3.9. Circular dichroism (CD) and thermal unfolding 47
2.3.10. FIDA (flow-induced dispersion analysis) 49
2.3.11. Electrophoretic mobility shift assay 50
2.4. Simulations of CREB behaviour with large excess of non-target DNA 50

Chapter 3 Producing pure CREB samples suitable for biophysical studies 53
3.1. Inherited purification methods yield slightly contaminated CREB and could be improved 54
3.2. Pure CREB is obtained with the intermediate GB1 method but suffers from low yields and aggregation 56
3.2.1. DTT addition and mutation of each cysteine to serine to make pseudo wild-type CREB greatly increases purity 56
3.2.2. Testing of denaturants and other chromatography methods 59
3.2.3. The HaloTag acts as a solubility tag which GB1 but not GST can replace. 62
3.2.4. GB1-CREB can be heat purified 64
3.2.5. Polyethyleneimine does not improve purification 66
3.2.6. Moving the 6His tag to the N-terminus 66
3.2.7. The GB1 method still suffers from low yield and product aggregation 72
3.3. The SUMO method improves on yield and aggregation issues from the GB1 method . 75
3.3.1. GB1 to SUMO solubility tag switch and auto-induction of protein expression solves issues with the GB1 method 75
3.3.2. Extra chromatography methods tested with SUMO-CREB 77
3.4. Discussion 81

Chapter 4 Characterisation of CREB's biophysical properties 85
4.1. CREB secondary structure predictions 86
4.2. Investigation into CREB thermostability 89
4.2.1. CREB structure melts at high temperatures but recovers well 89
4.2.2. Finding out CREB's T_m 89
4.3. Investigation into CREB equilibrium binding properties to target and non-target DNA 92
4.3.1. CREB's affinity to target CRE probed by anisotropy titrations 92

4.3.2. Thermodynamics of CREB to CRE association 97
4.3.3. Exploration of non-target DNA affinity using fluorescence 97
4.3.4. Exploration of CREB binding to nCRE using isothermal titration calorimetry (ITC) 103
4.4. CREB's kinetics of binding to target DNA 106
4.4.1. CREB has fast, electrostatically directed binding to CREh 106
4.4.2. CREB has an apparent multiphase residence-time 110
4.4.3. Kinetics and equilibrium-derived K_D differ, likely due to DNA-induced dimerization at low CREB concentrations 116
4.5. Influence of pH on CREB DNA binding 116
4.6. Discussion 120

Chapter 5 Dissection of influence of non-DNA binding domains on DNA binding 125
5.1. Sequence analysis of the individual CREB domains 128
5.2. Design of CREB constructs 134
5.3. Investigation of bZIP DNA binding properties at pH 7.4 135
5.3.1. Binding equilibrium titrations show bZIP has similar binding affinity for CREh but increased affinity for nCRE compared to CREB 135
5.3.2. Further investigation of bZIP binding to target DNA in kinetics assays show a faster and longer binding of bZIP to target DNA than CREB 141
5.3.3. Monitoring association rate constants in the presence of non-target decoys shows CREB's non-DBD facilitate target-search 147
5.4. Investigation of the individual contributions of CREB's non-DNA binding domains to DNA binding 152
5.4.1. Equilibrium binding titrations show each domain impedes binding to non-target DNA, but not apparently to CREh 152
5.4.2. Kinetics experiment show each domain similarly impedes association to target DNA but each domain's influence on complex lifetimes varies 158
5.5. CREB isoform A has similar DNA binding properties to CREB isoform B 167
5.6. Discussion 167

Chapter 6 Investigation of CREB dimerization and the stoichiometry of target search 177
6.1. Characterisation of the homodimerization process of CREB and CREB bZIP 180
6.1.1. Native protein size determination methods suggest CREB is present as a dimer in solution 180
6.1.2. Thermal unfolding and dissociation give an estimate of homodimerization affinity and thermodynamics 183

6.1.3. Perturbation by dilution is a viable method to find CREB homodimer formation kinetics and affinity 185
6.1.4. Various equilibrium titration methods attempted to obtain $K_{D,homo}$ 190
6.1.5. Investigation of bZIP homodimerization at pH 7.4 using urea unfolding and refolding kinetics methods 192
6.2. Influence of DNA on the homodimer population 200
6.3. Producing forced dimer and forced monomer CREB 203
6.3.1. Rationale for forced dimer and monomer 203
6.3.2. CREB forced dimer purification 205
6.4. Forced dimer biophysical properties 206
6.4.1. Equilibrium binding to target CREh and non-target nCRE DNA 206
6.4.2. Association rate constant with CRE 212
6.4.3. Dissociation rate constant with CRE 214
6.4.4. Effect of non-target DNA on CREB forced dimer's association kinetics 218
6.5. Characterising monomer-DNA interactions 222
6.6. Equilibrium modelling of species concentrations 225
6.7. Discussion 228

Chapter 7 Perspectives and future directions 231
7.1. In depth biophysical characterisation shows CREB is a strong, fast, dimeric binder of DNA 231
7.2. CREB most likely follows the dimer pathway, in what appears to be an exception for bZIPs so far 234
7.3. Disordered and activation domains self-inhibit CREB binding to DNA, improving the efficiency of target-search 236

References 241
Appendix 261
Appendix 1: derivation of the homodimerization-corrected binding model for anisotropy titrations 262
Appendix 2: Derivation of the CREB homodimerization model to fit the labelled CREB dilution anisotropy curve. 265
Appendix 3: Derivation of the reversible refolding model to fit bZIP refolding kinetic traces. 267

List of Figures

Figure 1.1: Adapted central dogma of molecular biology. 1
Figure 1.2: Overview of transcription initiation regulation. 3
Figure 1.3: Overview of transcription factor domain and cofactor functions. 5
Figure 1.4: Overview of the facilitated diffusion process. 7
Figure 1.5: Reported and proposed influence of disordered domains on target search. 11
Figure 1.6: CREB regulation by cellular signals. 14
Figure 1.7: Notable CREB structures. 16
Figure 1.8: CREB domain architecture. 17
Figure 1.9: Representation of dimer and monomer pathways. 19

Figure 2.1: Simplified kinetic scheme of CREB binding target and non-target DNA. 37

Figure 3.1: Initial expression tests using previously derived method. 55
Figure 3.2: DTT addition and pseudo wild type purification. 57
Figure 3.3: Addition of denaturant and size-exclusion chromatography (SEC) improves purity. 58
Figure 3.4: Sequencing mass spectrometry CREB coverage. 61
Figure 3.5: Denaturants, ion-exchange and HaloLink beads early purification tests. 61
Figure 3.6: HaloTag likely acts as a solubility tag. 62
Figure 3.7: Excitation-emission spectra of purified CREB. 63
Figure 3.8: GST is not a suitable solubility tag for CREB. 63
Figure 3.9: Heat can be used as a first purification step. 65
Figure 3.10: Polyethyleneimine (PEI) and ammonium sulphate (AS) tests to precipitate CREB. 67
Figure 3.11: Degradation products are co-purified with 6His-HaloCREB. 68
Figure 3.12: FXa is not a suitable protease to remove the C-terminal 6His. 68
Figure 3.13: Carboxypeptidase A (CPA) cannot reliably cleave C-terminal 6His. 69
Figure 3.14: Purification to apparent homogeneity of bacterially expressed 6His-GB1-CREB. 70
Figure 3.15: Investigation of CREB aggregation. 73
Figure 3.16: TEV-CREB aggregates affect concentration determination, causing issues with experimental reproducibility. 74

Figure 3.17: Example of concentration determination using Bradford assays with albumin and TEV-CREB standards. 75
Figure 3.18: Yields and nucleic acid purity comparison of the different CREB purification methods. 76
Figure 3.19: SCooP thermostability prediction of SUMO homolog yeast SMT3. 77
Figure 3.20: CREB is purified to apparent homogeneity using the SUMO method. 78
Figure 3.21: Mass spectrometry shows pure CREB is obtained with the SUMO method. 79
Figure 3.22: CREB from the SUMO method can be further purified if needed. 80
Figure 3.23: Overview and comparison of the four methods used and developed in this Chapter.. 82

Figure 4.1: CREB CD spectra and structure predictions. 88
Figure 4.2: Disorder predictions for CREB. 87
Figure 4.3: AlphaFold predictions of CREB. 87
Figure 4.4: CREB CD spectra melt and structure predictions at different temperatures. 90
Figure 4.5: CD thermal melt analysis. 91
Figure 4.6: CD thermal melt recovery. 91
Figure 4.7: Differential scanning calorimetry (DSC) of CREB. 92
Figure 4.8: Anisotropy titrations of CREB binding to CRE labelled with Alexa Fluor™ 488. 94
Figure 4.9: CREB labelling with Alexa Fluor™ 594. 95
Figure 4.10: Reverse anisotropy titrations using 10nM CREB-Alexa Fluor™ 594 as a probe bound by unlabelled CRElin. 96
Figure 4.11: ITC titration with CREB bound by CRElin. 98
Figure 4.12: Anisotropy titrations of CREB binding to 5 nM nCRE-Alexa Fluor™ 488. 100
Figure 4.13: Two-sites models are better suited to fit CREB anisotropy titrations of binding to nCRE labelled with Alexa Fluor™ 488. 101
Figure 4.14: Reverse intensity nCRE titration using CREB labelled with Alexa Fluor™ 594 as a probe. 102
Figure 4.15: There is almost no difference between using the “Single” or “All” corrections for ITC. 104
Figure 4.16: CREB ITC titrations with nCRE and comparison of “FO” and “Single” corrections. 105
Figure 4.17: Association kinetics of CREB to CREh-Alexa Fluor™ 488. 107
Figure 4.18: Investigation of deadtime cutoff influence on association kinetics. 109
Figure 4.19: Dissociation kinetics of CREB from CREh follow a double-exponential decay model. 111
Figure 4.20: Starting CREB concentration has little influence on k_{off}. 112
Figure 4.21: DNA structure affects dissociation of CREB from DNA. 113

Figure 4.22: Control traces for stopped-flow dilutions. 114
Figure 4.23: Dissociation experiments with attempted correction for experimental artefacts. 115
Figure 4.24: Effect of pH on anisotropy titrations. 117
Figure 4.25: Effect of pH and C-terminal tag on CREB kinetics. 119

Figure 5.1: Bioinformatic analysis of CREB domains. 129
Figure 5.2: Sequence composition of CREB domains. 130
Figure 5.3: AlphaFold predictions of CREB self-interactions and sequence motifs/hydrophobicity of the Q domains. 132
Figure 5.4: CREB deletion mutants do not suffer from aggregation issues even when using the GB1 method. 134
Figure 5.5: Anisotropy titrations of bZIP binding to labelled targets CREh and CRElin. 136
Figure 5.6: ITC and fluorescence anisotropy binding titrations of bZIP to nCRE 138
Figure 5.7: Comparison of bZIP and CREB's equilibrium binding properties to target and non-target DNA. 139
Figure 5.8: Electrophoretic mobility shift assays using fluorescent DNA oligos. 140
Figure 5.9: bZIP associates to CRE twice as fast as full length CREB 142
Figure 5.10: k_{off} is independent of starting bZIP concentration. 143
Figure 5.11: Variability in rate constants and amplitudes extracted from double exponential decay fits of CREB dissociation. 144
Figure 5.12: Comparison of dissociation kinetics to target DNA for CREB and bZIP. 145
Figure 5.13: Simple control correction of dissociation traces for bZIP. 146
Figure 5.14: Single and double-exponential decay fitting of association to CREh in the presence of nCRE competitors. 148
Figure 5.15: CREB's association to CREh is impeded proportionally to competitor K_D. 150
Figure 5.16: CREB association rate constants to CREh in the presence of competitor. 151
Figure 5.17: Anisotropy titrations to CREh and K_D compared CREB mutants. 153
Figure 5.18: Anisotropy titrations to of CREB mutants to 5 nM nCRE-Alexa488 154
Figure 5.19: ITC raw heat and normalised heat titrations of CREB mutants with nCRE. 155
Figure 5.20: $K_{D,nCRE}$ estimated from ITC, anisotropy titrations and pooled replicates compared for each mutant. 156
Figure 5.21: Thermodynamic values and n-value extracted from ITC binding to nCRE titrations for CREB mutants. 157
Figure 5.22: Q1/Q2 deletion mutants association experiments to 5 nM CREh-Alexa488 159
Figure 5.23: GS-KID mutant association experiment to 5 nM CREh-Alexa488. 160
Figure 5.24: Comparison of CREB mutants' association kinetics. 161
Figure 5.25: Q1 and Q2 deletion mutants' dissociation kinetics 163

Figure 5.26: GS-KID mutant dissociation kinetics. 164
Figure 5.27: CREB mutants k_{off} comparison. 165
Figure 5.28: GS-KID control-corrected dissociation kinetics. 166
Figure 5.29: Equilibrium binding titrations of CREB-A to nCRE. 168
Figure 5.30: CREB-A association and dissociation kinetics with CREh. to nCRE. 169
Figure 5.31: CREB-A thermodynamics of binding to nCRE using ITC. 170
Figure 5.32: Comparison of CREB and CREB-A's binding kinetics to CREh. 170
Figure 5.33: Multiple sequence alignment of CREB, htCREB and human CREM family members. 174

Figure 6.1: Simplified kinetic scheme of CREB binding target and non-target DNA. 177
Figure 6.2: SEC-MALS of CREB and CREBΔLZ. 181
Figure 6.3: Analytical ultra-centrifugation (AUC) CREB run. DNA. 182
Figure 6.4: CD thermal melt of CREB to extract $K_{D,homo}$. 184
Figure 6.5: There is great variability in extracted parameters from individual CREB-Alexa594 dilution traces fits. 186
Figure 6.6: Example of pairwise CREB-Alexa594 dilution relaxation fits. 187
Figure 6.7: Variability of fitted parameters using pairwise fitting. 188
Figure 6.8: Discarding of the early signal change in fluorescently labelled CREB dilution relaxation traces has little effect on fitting. 189
Figure 6.9: Variability of fitted parameters from fluorescently labelled CREB dilution relaxation kinetics based on acquisition method. n fitting. 190
Figure 6.10: Anisotropy attempts with urea titration and CREB-Alexa594 dilution. 191
Figure 6.11: Fits of CREB binding CREh anisotropy titrations accounting for dimerization; K_D to CREh and $K_{D,homo}$ comparison by method. dilution. 193
Figure 6.12: bZIP unfolding in urea experiments. 195
Figure 6.13: bZIP refolding from 8 M urea experiments. 196
Figure 6.14: Comparison of bZIP's homodimerization parameters at pH 6.5 and 7.4 with CREB's. 197
Figure 6.15: Comparison of bZIP's K_Ds of dimerization and binding to CREh by method. 198
Figure 6.16: Individual FRET tests of CREB homodimerization and DNA binding. 201
Figure 6.17: FRET titration of CREB-Alexa488/594 mixture bound by nCRE. 202
Figure 6.18: CREB-GGC forced dimer generation. 204
Figure 6.19: CREB-GGC oxidation trials in the presence of various chemicals. 206
Figure 6.20: CREB-GGC oxidation trials in the presence of various chemicals. 207
Figure 6.21: Forced dimers anisotropy binding titrations to 1 nM CREh-Alexa488. 208
Figure 6.22: Forced dimers anisotropy binding titrations to 5 nM nCRE-Alexa488. 209

Figure 6.23: ITC titrations with addition of DNA to forced dimer CREB and bZIP. 210
Figure 6.24: Forced dimer binding to nCRE parameters comparison with non-forced dimers. 211
Figure 6.25: CREB-GGC and bZIP-GGC association kinetics..... 213
Figure 6.26: CREB-GGC investigation of time cutoff influence on association. 214
Figure 6.27: Forced dimer dissociation traces of CREB-GGC and bZIP-GGC from 5 nM CREh-Alexa488. 215
Figure 6.28: Forced dimer dissociation kinetics analysis of CREB-GGC and bZIP-GGC. 216
Figure 6.29: Rudimentary correction of DNA mixing artefacts for CREB-GGC dissociation..... 217
Figure 6.30: Rudimentary correction of DNA mixing artefacts for bZIP-GGC dissociation. 218
Figure 6.31: CREB-GGC competitor inhibited association to 5 nM CREh-Alexa488. 220
Figure 6.32: CREB-GGC target-search inhibition follows the simple sequestration model..... 221
Figure 6.33: CREBΔLZ forced monomer anisotropy binding titrations. 223
Figure 6.34: Other affinity methods suggest CREBΔLZ does not bind DNA. 224
Figure 6.35: Simulations of CREB behaviour with increasing non-target competitor concentration. 227

List of Tables

Table 2.1: Buffers used in experiments. 24
Table 2.2: List of oligonucleotides sequences and properties. 32
Table 2.3: Extinction coefficients of CREB mutants. 35

Table 3.1: List of non-native sequences used in purification. 54
Table 3.2: Sequencing mass spectrometry s results shows contaminants are CREB degradation products. 60
Table 3.3: Sequencing mass spectrometry shows CREB is obtained by the GB1 method. 71

Table 4.1: List of constructs used with their isoelectric points and net charge at given pH. 118

Table 5.1: Basic properties of CREB mutants. 135
Table 5.2: Anisotropy and kinetics K_D of CREB and mutants binding to CREh labelled with Alexa Fluor™ 488. 166
Table 5.3: Association and dissociation rate constants of CREB and mutants from kinetic studies using CREh labelled with Alexa Fluor™ 488. 171
Table 5.4: $K_{D,nCRE}$ of CREB and mutants through various methods. 171
Table 5.5: Thermodynamic parameters and n-value of CREB and mutants binding to nCRE via ITC. 171

Table 6.1: Cellular concentration, homodimerization K_D and dimeric fraction of example bZIPs. 179
Table 6.2: Experimental results of bZIP kinetic refolding. 198
Table 6.3: Experimental results of bZIP kinetic unfolding. 198
Table 6.4: Comparison of CREB's and bZIP's $K_{D,homo}$ from different methods. 199
Table 6.5: Comparison of CREB's and bZIP's homodimerization kinetics. 199
Table 6.6: Thermodynamics and n-value of binding to nCRE in ITC experiments using forced dimers. 212
Table 6.7: Monomerically-adjusted K_D to CREh and nCRE of forced dimer mutants. 219
Table 6.8: Monomerically-adjusted kinetics values of forced dimer mutants binding to CREh. 222
Table 6.9: Values set as constant for CREB equilibrium modelling. 225

List of Abbreviations

AA: amino acid

Alexa488: Alexa Fluor™ 488

Alexa594: Alexa Fluor™ 594

AD: activation domain

AS: ammonium sulphate

AUC: analytical ultracentrifugation

BR: basic region

bZIP: basic zipper

CBP: CREB binding protein

CD: circular dichroism

CPA: carboxypeptidase A

nCRE: non-target CRE

CREB: cAMP response element binding protein

CV: column volume(s)

DBD: DNA binding domain

D.I.T.: digital integration time

DNA: deoxyribonucleic acid

DSC: differential scanning calorimetry

ED: effector domain

EMSA: electromobility shift assay.

FIDA: flow-induced dispersion analysis

FPLC: fast protein liquid chromatography

FRET: Förster resonance energy transfer

FXa: Factor Xa

GdnHCl: guanidine hydrochloride

GS: glycine and serine linker

GST: glutathione S-transferase

HPLC: high performance liquid chromatography

ITC: isothermal titration calorimetry

IDP: intrinsically disordered protein

IDR: intrinsically disordered region

KID: kinase inducible domain

KIX: KID interacting domain

LCD: low complexity domain

LC-MS: liquid chromatography mass spectrometry

LZ: leucine zipper

MALS: multi-angle light scattering

MW: molecular weight

Ni-NTA: nickel-nitrilotriacetic acid.

nt: nucleotide

PAGE: polyacrylamide gel electrophoresis

PEI: polyethyleneimine

pI: isoelectric point

PKA: protein kinase A

PTM: post -translational modification

pWT: pseudo wild type

Q1: glutamine-rich domain 1

Q2: glutamine-rich domain 2

RD: repressor domain

RNA: ribonucleic acid

SDS: sodium dodecyl sulfate

SEC: size exclusion chromatography

SUMO: small ubiquitin-like modifier

TAD: transactivation domain

TEV: tobacco etch virus

TRE: TPA response element

ULP-1: Ubl-specific protease 1

UV-vis: ultraviolet-visible spectroscopy

WT: wild type

Chapter 1

Introduction

1.1. Transcription factors and their crucial role within the central dogma of molecular biology

The central dogma of molecular biology (Figure 1.1) (1) is a well-known triangle, composed of DNA, RNA and proteins. DNA is the cellular carrier of genetic information with an alphabet of four bases (A, T, G, C) and, in eukaryotic cells, segregated and protected within the nucleus. DNA can reach quite large sizes, more than 3 billion bases in the human genome which equates to 2 meters if all bases are aligned lengthwise one by one (2). Hence,

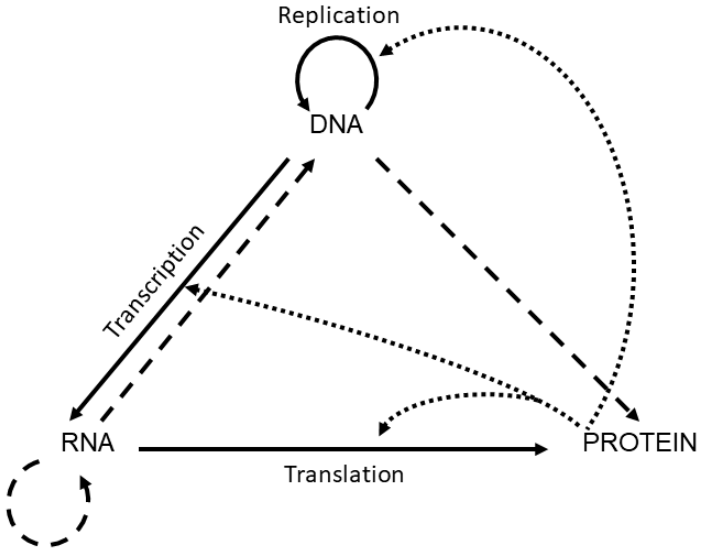


Figure 1.1: Adapted central dogma of molecular biology. Full arrows are known processes: transcription (DNA to RNA), translation (RNA to protein) and DNA replication. Dashed arrows are tentative processes. Dotted arrows show proteins are involved in all processes. Adapted from (1)

in eukaryotes, it is packed quite tightly by various proteins, mainly histone octamers and looping proteins, which also greatly affect DNA-related processes (reviewed in (3–6)).

Genes, a unique sequence of DNA bases, are read and replicated into messenger RNA (mRNA) by the transcription process during which RNA polymerase enzymes read and create complementary single-stranded mRNA molecules. These mRNAs are processed via splicing, 5' capping and 3' polyadenylation to be distinguished from other RNAs, then exported from the nucleus into the cell and read by the protein making machinery to make protein in the translation process (processing and its transcription/translation effects are reviewed in (7–10)). Hence, they act as intermediates between the DNA and the protein-making translation process, another barrier to limit DNA interactions and safeguard this al-important molecule.

Transcription factors are proteins that participate in transcription by modulating transcriptional activity at target genes, often acting downstream of signalling cascades (11). Specific transcription factors tend to bind DNA regulatory elements or target sites with sequences specific for (but not unique to) each factor. These elements either have "promoter" and/or "enhancer potential". "Promoter potential" elements are usually gene-specific and initiate transcription, whereas "enhancer potential" elements can greatly promote (but rarely initiate) transcription of many distal genes, up to 1 million bases away (reviewed in (12)). Transcription factors typically recruit other factors to affect a specific part of the transcription process. Most recruit coactivator proteins like Mediator (13) or corepressors, but some can directly interact with parts of the pre-initiation complex (RNA polymerase II and the general transcription factors essential for transcription initiation) (Figure 1.2) (14,15). Hence, transcription factors are uniquely important within the central dogma, representing all three facets at once: proteins that interact with DNA to help create

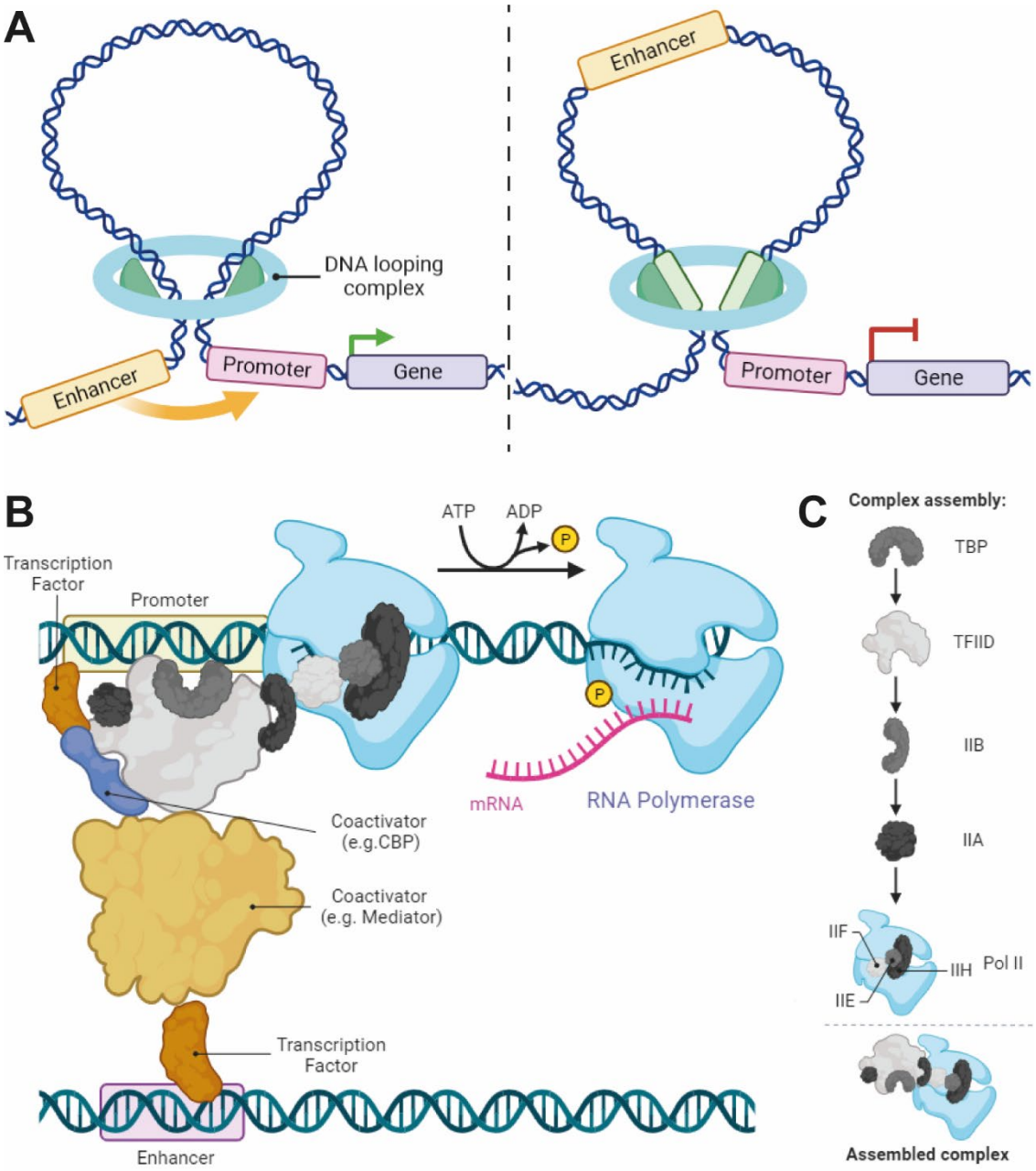


Figure 1.2: Overview of transcription initiation regulation. A. Example of how chromosomal structure can affect transcription, by bringing enhancer and promoter elements in proximity or separating them. Adapted from (6). **B.** Transcription initiation using promoter and enhancer elements. Transcription factors can bind promoters or enhancers and recruit the Pre-Initiation Complex (PIC) (grey factors and RNA polymerase II) themselves or via coactivators, like CBP and mediator. Phosphorylation via ATP hydrolysis then releases the polymerase to begin transcription. Adapted from (15). **C.** Sequential assembly of the PIC. Adapted from (15). Created with Biorender.com.

RNA. Studying them is crucial to understand the dogma and how our genetic information is read and used. Moreover, cancers typically arise from a deregulated transcription process and many transcription factors are overexpressed in cancers, so their study can have medical repercussions (reviewed in (16)).

1.2. Typical characteristics of transcription factors

Traditionally, transcription factor architecture is understood in terms of a modular format, with different domains being largely independent and performing different roles. Most, but not all, transcription factors are composed of at least a DNA binding domain (DBD), which recognises the aforementioned target sites, and an effector domain (ED) (14). The DBD adopts a typical fold and is used to separate transcription factors into families (14,17).

The effector domain is traditionally called an activation domain (AD) if the transcription factor activates transcription, or a repressor domain (RD) if it represses it. Effector domains often recruit cofactors (termed coactivators or corepressors respectively) which modulate transcription by recruiting other partners, modifying DNA or proteins and stimulate or inhibit association with the transcription machinery (Figure 1.3) (14,17). ADs are extensively more studied than RDs. Multiple screens in human and yeast cells have revealed ADs typically consist of acidic and hydrophobic amino acids and are depleted of basic amino acids (18–21). Despite these commonalities the actual sequence itself is quite malleable since 1% of sequences tested in screens usually activate transcription (for a random 20 amino acid sequence, that would mean 10^{25} hits for 10^{27} possibilities) (20). Activation domains are also, on average, 91 amino acids long but vary greatly in size: from 4 to 1248 amino acids (17). Effector domains are sometimes bound by ligands or modified by post-translational modifications that trigger activity e.g. phosphorylation by kinases that

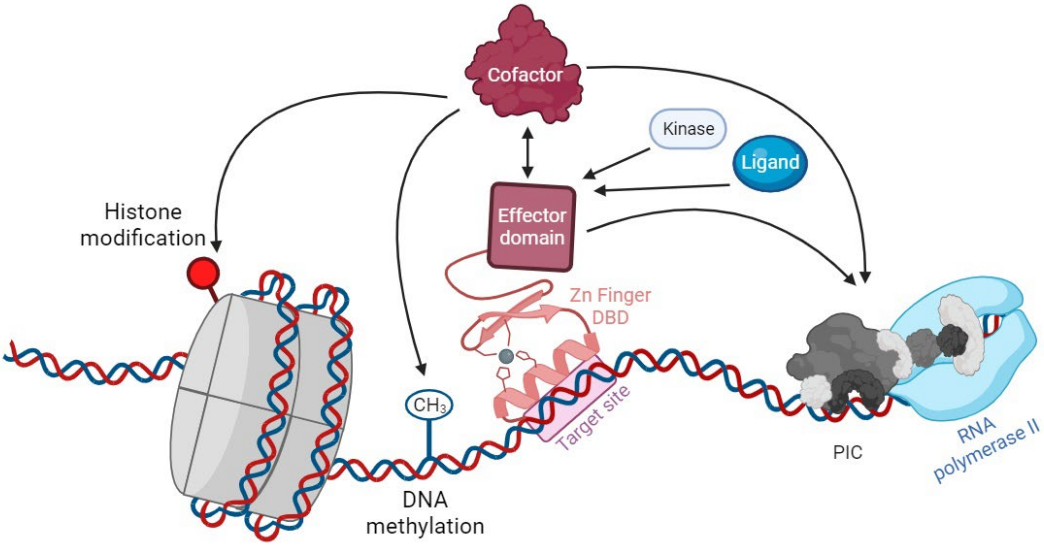


Figure 1.3: Overview of transcription factor domain and cofactor functions. Transcription factors DBD (DNA binding domain, here a Zn finger) bind target sites. Effector domains then affect the transcription process, often by recruiting cofactors, parts of the PIC (pre-initiation complex) or RNA polymerase II. Transcription factors can be activated via the effector domains, by ligands or kinases. Cofactors can have many activities, from histone or DNA modifying enzymes to promoting or inhibiting association with other partners. Adapted from (17).

are themselves activated during a signalling cascade (11). These events can create a binding surface for coactivators, like phosphorylation for the KID-KIX interaction (22–25) or alter electrostatic properties involved in autoinhibition (26). Furthermore, corepressors and coactivators, like CBP, can modify chromatin to prevent or promote binding by other factors (Figure 1.3).

1.3. Disorder in transcription factors

The greatest common characteristic in eukaryotic transcription factors is arguably that they are predominantly disordered proteins (27). Disordered proteins do not adopt a stable structure in a water-based environment, due to a high proportion of polar and charged residues which results on a lack of hydrophobic core to fold on (28,29). This does not mean they always adopt random coils. Short-lived conformational ensembles can be sampled by the protein, and disordered proteins often fold when binding to their partners in a process called folding upon binding (29,30). Folding upon binding is typically modelled by two

proposed mechanisms: induced-fit, when near proximity to the partner triggers folding to match the partner's surface; or conformational selection when multiple conformations are sampled by the disordered protein and those compatible for binding are selected by the partner's presence. A mixture of both is possible (31). Disorder can also be localised to particular regions within a protein with other regions being stably folded. An estimated 35% of human proteins contain disordered regions (32). Although, this is raised to 80% when considering transcription factors exclusively (27). This suggests disorder could be particularly important for transcription factor function.

Disordered content is linked intimately with sequence composition. Activation domains tend to be negatively charged, and DBDs are positively charged to bind negatively charged DNA (due to its phosphate groups) (17). Such a high charge distribution is likely to lead to disorder in a polar solvent like water.

However, there properties linked with protein disorder which can greatly benefit transcription factors. Firstly, both DBDs and effector domains can display typical folding upon binding behaviour when binding to their partner DNA and proteins (33). This apparent flexibility could expand a factor's interactome, direct it to different targets based on cellular signals, and modulate the strength of its response. Another consequence of this flexibility could be greater binding surface than in similarly folded proteins (34) or sampling area to find partners more efficiently, known as the fly-casting mechanism (35) which could improve signalling response speed. Disordered proteins also typically have quicker turnover due to shorter half-lives (36) and greater opportunities for post-translational modifications (PTM) to modify behaviour (reviewed in (37)). This contributes to fast adaptation to cellular conditions and limit unwanted sustained activity in a more efficient manner than by making and folding new proteins.

1.4. The target search process

One of the biggest outstanding questions regarding eukaryotic transcription factors, is how they reach such small target sites sequences in the vastness of eukaryotic genomes. Target sites are small (in eukaryotes, 5 to 14 bp, average of 10 bp, median of 8 (38)) and exist amongst more than 6 billion base pairs in the human diploid genome. The disordered nature of transcription factors may be key in the target search process.

Target search is mathematically found to be optimised through a facilitated diffusion model (39–44), where free 3D diffusion of the factor is interspersed with binding events. During such events, the transcription factors slide on the DNA (typically referred as 1D diffusion) and has a probability to stop at a binding site, with an increased probability with increased sequence similarity to the optimal recognition sequence. There are other discrete events during the sliding process, such as hopping, jumping to a nearby strand, immediate re-binding, or intersegmental transfer when two DNA regions are in such proximity when the transfer of one to the other does not interrupt the sliding process (Figure 1.4) (45). The

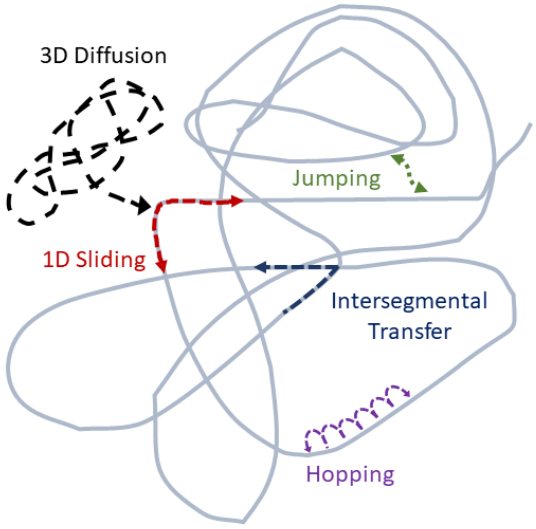


Figure 1.4: Overview of the facilitated diffusion process. Facilitated diffusion stipulates that target search is optimised by alternating 3D free diffusion with 1D sliding on DNA, where targets may be found and bound. Other events can happen, such as jumping to nearby strands, intersegmental transfer during sliding, or hopping on the DNA, for instance to avoid obstacles. Adapted from (45).

model has since been validated by fluorescence microscopy, mainly by using bacterial transcription factors (46–52). It was found, as the model suggests, that most binding events during the search are indeed to non-target DNA (53). However, prokaryotic genome sizes are much smaller and their DNA-binding sites are longer (around 10 to 24 bp, median and average of 15 bp (38)), so the chance of off-target binding events is reduced. This is more problematic in eukaryotes: assuming that some of the biophysical values extracted from these experiments hold for eukaryotic transcription factors (such as length and duration of the sliding window), they would take days to find their targets (46,54). Hence, there are mechanisms at play to direct factors to their target sites, even more so when considering that some factors still retain high affinity for incomplete target sequences.

Fluorescence experiments have unveiled some other ways which facilitate target search, such as the compartmentalisation of transcription factors to certain regions (55,56), or frequent re-binding events promoted by liquid-liquid phase separation (57). "Antenna" sequences with some similarity to binding sites (such as higher GC content than normal) could also trigger sliding events closer to target sites (58,59). Similarly, it was recently demonstrated that, if a transcription factor stops scanning and adopts a stably-bound conformation at this stage, it tends to stay bound for the same amount of time regardless of DNA sequence (a time specific to each transcription factor) (60). This means binding affinity of factors for a specific sequence is not determined by complex lifetime, but by the probability of binding. This would be proportional to the association frequency with a particular stretch of DNA. The probability of stopping is not always that high: for the *E. coli lac* repressor binding to the O1 sequence, it is 12% (51). Thus, repeated associations may be paramount for binding to occur. Lastly, the influence of chromatin should be kept in mind, as around 50% of the DNA is tightly packed in heterochromatin and inaccessible, although

this can be raised to as much as 90%. This could greatly reduce the binding pool available in the first instance (61,62).

1.5. Effects of disordered domains on target search

The folding upon binding particularity of disordered DNA binding domains presents a unique opportunity for target search. Theoretically, it allows the transcription factor to adopt multiple conformations while sliding on the DNA (low affinity conformations with a flat energy landscape that meets criteria for efficient search) but also adopt a "binding" high affinity conformation (43–45,63). In folded DBD's the opposing requirements for efficient search and recognition would instead be met by having multiple domains or through two separate stable structures that interchange. Indeed, the p53 transcription factor and tumour suppressor possesses an ordered DBD and an important C-terminal disordered region, which allows its oscillation between different sliding and binding modes (64,65). The disordered region is also required for intersegmental transfer (48) and fast sliding (65) the latter especially even in the absence of the DBD, showing the importance of the disordered region in the search process. The presence of DNA antennas or target sites could induce a structural switch gradually for disordered DBDs, resulting in a higher association probability during sliding. A similar process has been reported by simulations: these showed gradual folding of DBDs with increased proximity of these to DNA (66). There are suggestions that a similar structural switch exists in p53 (64,65) as well as Zn-finger transcription factors (63), but the existence of such a switch in most transcription factors remains to be proven.

Most of this hypothesis only considers the disordered DBDs. Conversely, what is very recently understood is that disordered activation domains have a major influence on target search – if not are solely responsible for some proteins (67,68). Indeed, sequencing experiments in yeasts have shown that deletions of disordered domains lead to delocalisation

of transcription factors, despite the DBD being intact (Figure 1.5A) (68). Moreover, the effect is dependent on the proportion of deleted disordered regions more than on a particular part of the disordered domain, and the disordered domain alone was better able to localise to expected target site than the DBD alone (Figure 1.5B). Follow-up studies have found that switching around such domains between transcription factors belonging to the hypoxia inducible factors family also exchanged some of their target search behaviour (69), and that transcription factor paralogs differed in their target genes not because of differences in their DBD, but in their activation domains (70), themselves predominantly disordered (17). Furthermore, adding disordered low-complexity domains (LCD) to the LacI repressor increased presence at reporters via multivalent interactions between other LCDs of similar composition (71), again showing how important these domains are for target search and DNA binding. They can also influence intersegmental transfer, for instance through the proposed "monkey bar" mechanism, where disordered N-terminal tails are paramount in finding, binding and switching to an alternative DNA strand while the DBD is bound to another in intersegmental transfer simulations (72). One notable exception to this behaviour is the largest eukaryotic transcription factor family, the zinc fingers. Although, this family is mainly composed of transcriptional repressors and have more structured non-DBD domains. They may behave differently as a result (73).

Hence, it is now almost undeniable that these disordered, activation domains influence target search massively, although some factors are more DBD dependent than AD dependent for their search (73). As mentioned previously, these domains are mainly negatively charged with some minor hydrophobic character (18–21). One model suggests this "surfactant" behaviour can be used in chromatin remodelling, dislodging DNA from chromatin and nucleosomes (21,74,75). There is evidence that DBD-dependent factors bind at nucleosome-free regions whereas AD-dependent ones localise to regions with "fuzzy"

chromatin (i.e., unstably located but present nucleosomes) favouring this model (Figure 1.5D) (73). Another model suggests these domains are used to "sample" protein complexes on the DNA, before committing to some regions, in a two-step process (Figure 1.5C) (67). The latter model also uses the basis that such disordered domains are involved in binding other proteins through multiple low-affinity interaction sites called "fuzzy" interactions (76). Indeed, it has been observed that disordered, low-complexity regions form highly dynamic and selective association hubs with other disordered domains, which could help locate

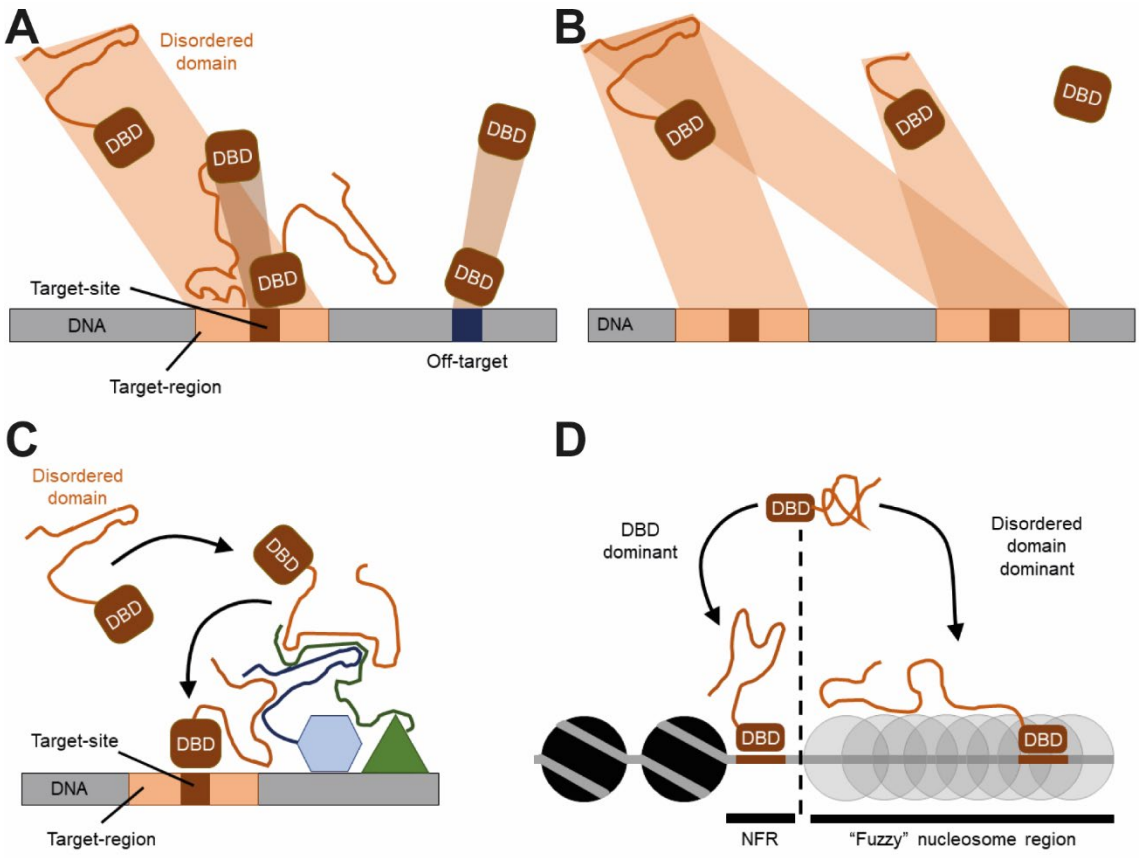


Figure 1.5: Reported and proposed influence of disordered domains on target search. A. Disordered domains help localise some factors to the correct regions before the DNA binding domain (DBD) binds its target. Deletion of such disordered regions led to off-target binding. **B.** This effect is dependent on the disordered domain's length, with decreasing length leading to reduced association to target sites. A. and B. adapted from (68). **C.** Two-step model where disordered domains help localise factors by first probing a disordered "protein cloud" before the DBD binds DNA. Adapted from (67). **D.** Transcription factors with DBD-dominated target search bind at NFR (nucleosome-free regions), whereas those with disordered domain-dominated search bind at "fuzzy" nucleosome regions with delocalised but still present histone octamers. Adapted from (73).

factors indirectly. The hubs also resulted in stabilised DNA binding, increased recruitment of RNA polymerase and increased transcription (71).

Another possible effect of disordered domains on target search is their ability to participate in liquid-liquid phase separation (77–82). Condensates, liquid-liquid phase separated droplets, or membrane-less organelles all refer to the observation in cells of highly localised concentration of certain proteins/nucleic acids observed via fluorescence microscopy. These condensates display liquid-like behaviour and exist in a different phase to their surroundings, like oil and water. Entry and exit into these droplets is regulated, but proteins can diffuse inside as if they were in water, and condensates can be dissolved with the application of the appropriate chemical, salt concentration, or by lowering protein concentration (83–85). The ability of disordered domains to participate in phase separation is well documented (79,83,84,86,87) and some transcription factors are known to participate in phase separation (57,81,82,88), although this appears to be somewhat dependent on partners (the Mediator coactivator complex is often involved (89), and nucleic acids can facilitate phase separation (90)). Additionally, not all transcription factors benefit from phase-separation (91). Regardless, the theory of phase-separated transcription factories or, at least, localised hubs is appealing (82,92), especially since these could facilitate re-binding by keeping factors and target sites in close proximity (57), and could similarly be sampled by factors during target search.

Lastly, there may be a much simpler use case for disordered activation domains in target search, as they tend to be negatively charged but disordered DBD tend to be positively charged (17). AD could then simply act as self-inhibitory domains for DBD via electrostatic interactions to limit off-target association and improve target search efficiency since, presumably, this effect would not work as well for target sequences. This was reported for the FoxM1 transcription factor, whose autoinhibition by activation domains is relieved by

post-translational modification (93), creating a binding surface for its coactivators simultaneously, showing the impressive interplay between all different facets, domains, processes and proteins involved in target search.

Nevertheless, there is great variability in individual transcription factors and their domains, as highlighted by the numerous transcription factor families (14). Effector domains can be grouped into multiple distinct clusters as well (17). Not all of the above findings and suggestions may apply to all of them. Hence, there is still a need to characterise important transcription factors individually, especially in some deregulation of their activity leads to disease.

1.6. The CREB transcription factor

CREB (cAMP response element binding-protein) is a transcription factor activated downstream of the cAMP (94–96) and other signalling cascades (97–104) (Figure 1.6A). CREB binds to the palindromic CRE DNA sequence: 5'-TGACGTCA-3' with very high affinity, in the low nM (94–96,105–107), although half-sites are also bound (5'-TGACG-3' or 5'-CGTCA-3') (95). CREB is member of the old bZIP transcription factor family that dates back to before the metazoan-fungi split, so-called due to their basic zipper DNA binding and dimerization domain, and also the second largest family of dimerising transcription factors (108). CREB is additionally an essential transcription factor, whose deletion leads to perinatal death in mice and developmental defects (109,110), but also plays a role in memory retention (111–116) and in circadian rhythms (117) by adjusting the circadian clock. CREB is also overexpressed in many cancer types, indicative of tumour initiation, is linked to worse prognosis, and responsible for tumour progression (118–121). CREB also has links with diseases such as major depressive disorder, schizophrenia,

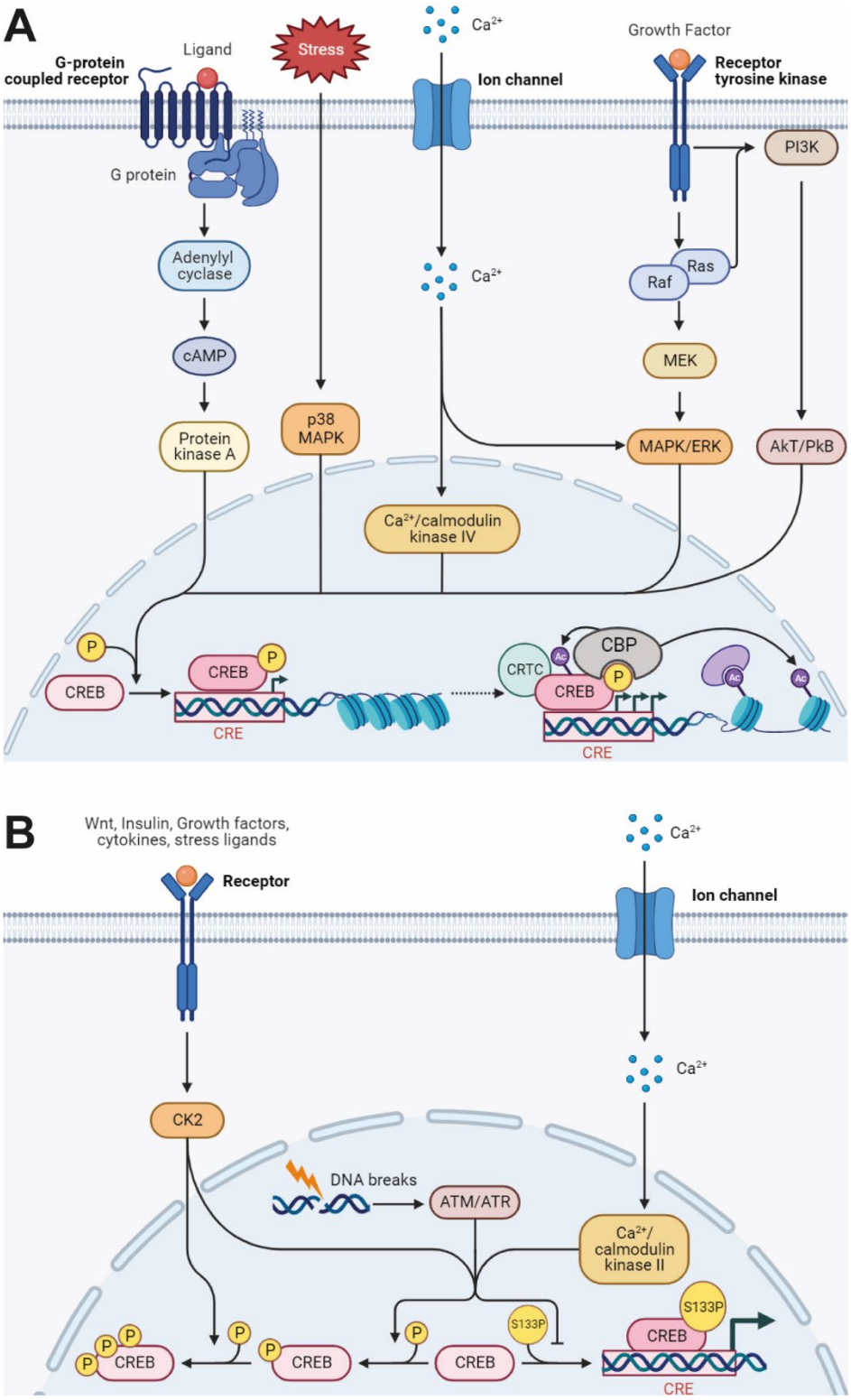


Figure 1.6: CREB regulation by cellular signals. A. Diagram of CREB activation by signalling cascades. Multiple pathways can trigger CREB phosphorylation ("P") at S133, mainly through cAMP (leftmost cascade), which induces activity and association with coactivators CRTC and CBP. CBP can also acetylate ("Ac") CREB and histones to open chromatin/recruit other factors. Adapted from (104). **B.** Diagram of CREB inhibition. CREB can be phosphorylated at other sites which inhibit S133 phosphorylation and prevent activation. Casein kinase 2 (CK2) can further increase inhibition via phosphorylation.

epilepsy and Alzheimer's (122–125). It is therefore an extremely important transcription factor to study to gain insights into the functional mechanisms of a predominant transcription factor family but also into general biology and cancer mechanisms.

The bZIP DBD is a bipartite domain split into the actual DNA binding region, subdomain is called the basic region due to its high number of basic (positively charged) residues. The other domain is the leucine zipper, a dimerization domain with interspersed hydrophobic (mainly leucine or L) residues that form an interaction interface (126–130). bZIPs are typically disordered in solution but form helical coiled-coil domain structures as dimers (131,132), with a further increase in helicity upon binding to DNA (107,133,134). An example of which is a structure of CREB bound to its DNA target (135) (Figure 1.7A). Most bZIPs (including CREB) have the bZIP domain at their C-terminus (17). As aforementioned, bZIPs are an old family, and many bZIPs can form heterodimers with other bZIPs, showcasing a wide range of interaction affinities: for instance, the Jun and Fos bZIPs have a high affinity for each other, but Jun can also bind to BATF with comparable affinity and Fos can also homodimerize with but with at least 20-fold reduced affinity (136). However, CREB is an exception somewhat, as it is one of few human bZIPs to exclusively form high-affinity homodimer complexes with almost no other bZIP binding partners (136).

CREB has three other domains N-terminal of the bZIP: two glutamine-rich domains, Q1 and Q2, and a kinase inducible domain (KID) which lies in between the longer Q domains (Figure 1.8). Together, they form the CREB transactivation domain (TAD). The KID domain is by far the most studied, as it contains the serine S133 residue whose phosphorylation leads to its activation. Various kinases acting downstream of signalling cascades can activate CREB (94,97,100,121,137,138), most notably protein kinase A (PKA) downstream of the cAMP cascade (94). Phosphorylation also acts as a binding interface for CREB's coactivator, CBP (CREB binding protein) (25). When bound, both the KID and

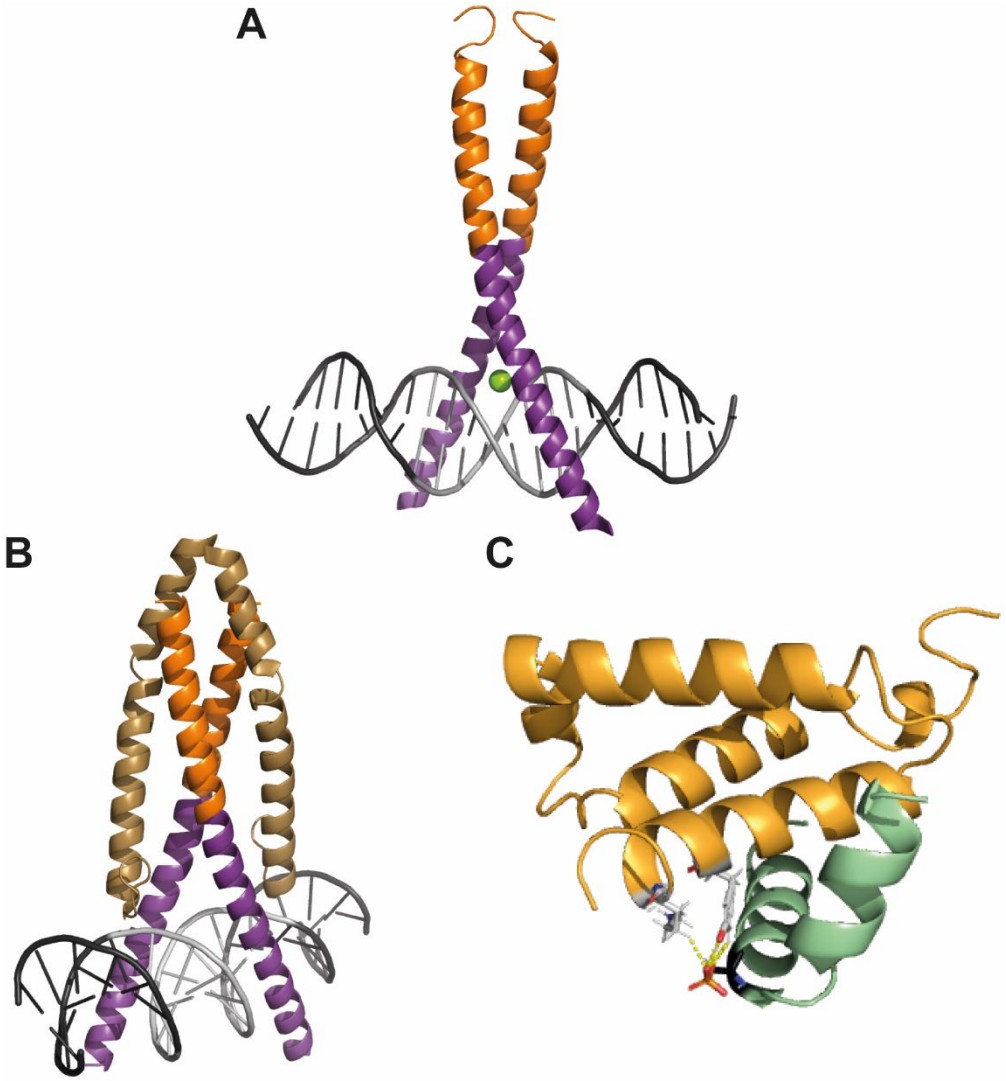


Figure 1.7: Notable CREB structures. A. CREB binding to DNA (black), with its CRE site highlighted (grey). Basic region is in purple, leucine zipper in orange. An Mg^{2+} ion (green) is involved in binding PDB: 1dh3. **B.** CRTC2 (gold) binding to CREB, with same colour scheme as A. CRTC2 binds to pre-formed CREB dimers by fitting in the groove the dimer forms and may interact with DNA too. PDB: 5zko. **C.** CBP's KIX (yellow) interacts with CREB's phosphorylated KID (green). S133P is involved in the reaction (bottom, black and licorice) by forming contacts with a tyrosine and lysine residue of KIX (grey and dashed yellow lines). PDB: 1kdx.

CBP's KIX (kinase inducible) domain form helices, another example of folding upon binding (23) (Figure 1.7C). Phosphorylation is crucial for this interaction because it increases the lifetime of the complex (22). CBP can then promote transcription by a combination of RNA polymerase recruitment (139,140) and acetylation of nearby histone octamer protein complexes (141). The latter can create a binding surface for transcriptional activators and disrupt the histone's ability to package DNA, making it more accessible for

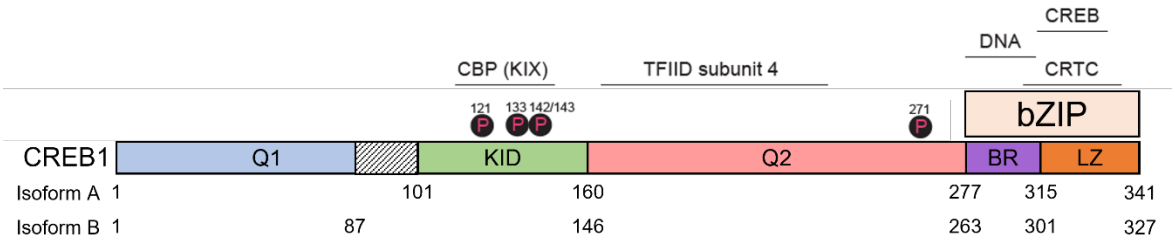


Figure 1.8: CREB domain architecture. CREB has two main isoforms, with isoform A being historically used (e.g. Serine 133 is with isoform A numbers) but isoform B is mainly used in this study as it currently is the standard Uniprot sequence. The only difference is a 14 amino acid deletion at the C-terminus of Q1 (shown as a hashed rectangle with horizontal black lines). Also shown are known phosphorylated residues (isoform A numbers) and interaction sites with partners.

transcription-inducing factors (142). The CRTC/TORC (CREB regulated transcription coactivator/transducers of regulated CREB activity) (143) is the other important class of CREB coactivators. CRTCs form helices upon binding DNA-bound CREB (144) at its bZIP, stabilising binding on target DNA, increasing transcription activity, and potentially favouring binding to certain promoters (145) (Figures 1.6A, 1.7C, 1.8).

CREB is extensively post-translationally modified, both by phosphorylation at other serine residues (Figure 1.8) (138) and by other modifications (121), but the roles of these modifications is still unclear, sometimes even contradictory (138). Some interesting examples are the Ca^{2+}/calmodulin kinases modulating CREB activity downstream of calcium induced signals (97,98) and ataxia-telangiectasia mutated / Rad3 related (ATM/ATR) which inactivate CREB during DNA damage through phosphorylation at other serine residues, not S133 (Figure 1.6B) (146). Casein kinase (CK) also phosphorylates CREB at multiple serine residues including four in the CK cassette just upstream of the KID, but the effect is inhibitory with its magnitude dependent on the number of phosphorylated residues in a graded fashion (Figure 1.6B) (26). CREB is also acetylated by CBP (147,148) at multiple lysine residues including K136 next to S133, although the effect appears situation dependent, with both increased and decreased effects on transcription and CBP association reported (121,147,148) (Figure 1.6A).

Similarly, the Q-rich domains have been largely neglected in both biophysical and transcriptional studies. There is evidence that these regions participate in transcription activation, since their deletion both decreased basal (149) and induced (150) transcription from target promoters. Moreover, Q2 appears critical for promoter occupancy and prolonged residence times (151). This is perhaps because both regions bind the TFIID (152), likely via its TAF4 subunit. Not only does inactivation of this subunit consequently reduce CREB-mediated transcriptional activation (153), but there are reports of Q-rich regions and TAF4 interactions (154). Yet very little is known about the individual impact of CREB's corresponding domains on target search, besides a very recent report that they appear to decrease affinity for DNA (155). This is especially damning considering how important non-DBD disordered domains appear to be in target search. Furthering our understanding of the glutamine-rich domains is paramount especially since a special isoform of CREB found in human testis (htCREB) lacks the Q2 domain (156), which makes it much more susceptible to CK phosphorylation-dependent inhibition (26).

Lastly, how CREB dimerization affects its target-search remains to be elucidated. Given that bZIPs are stably bound to DNA as a dimer, there are two pathways which they can take during the search process: the monomer pathway, through which dimerization occurs upon binding to the DNA; or the dimer pathway, when dimerization occurs beforehand (Figure 1.9). The consensus is that bZIPs, including CREB, follow the monomer pathway (157–161), especially since CREB is supposedly monomeric in solution (162). However, as previously stated, CREB is somewhat of an anomaly when considering bZIP interactions, and proteins used in the monomer vs dimer pathway are different, either forming heterodimers or have low homodimerization affinity. Indeed, preliminary studies from our lab show that the dimer pathway is preferred by isolated CREB bZIP domains (163).

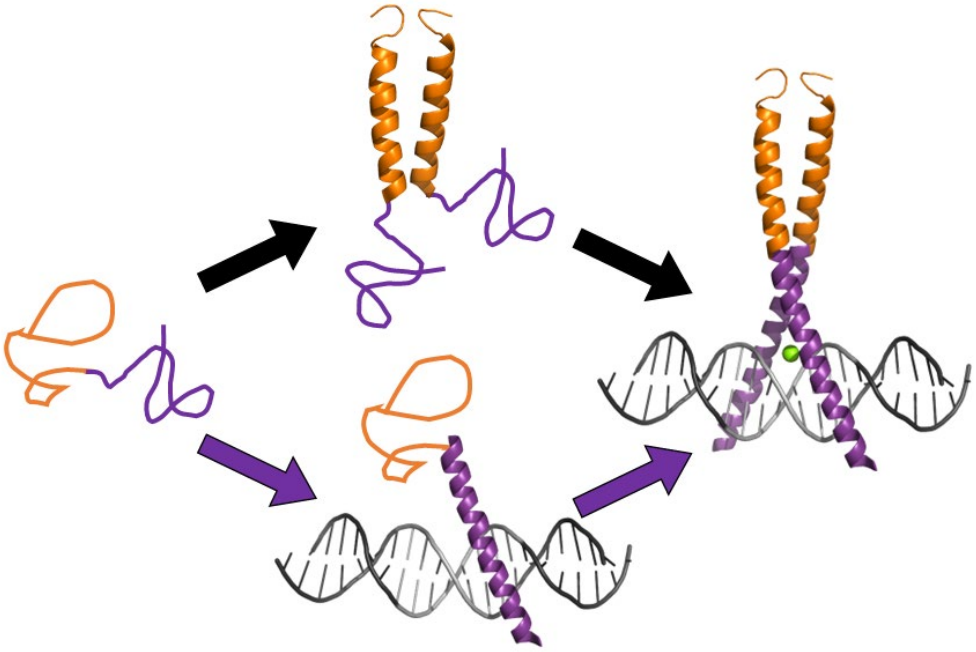


Figure 1.9: Representation of dimer and monomer pathways. CREB (here represented as a simplified bZIP, with other domains omitted) can bind its target as a monomer beforehand (bottom, orange arrows) or dimerise first (top, black arrows), called the monomer and dimer pathway respectively.

1.7. Overview of Thesis

The aim of this Thesis is to conduct a thorough characterisation of the full-length CREB transcription factor and of its individual domains' contributions to homodimerization and DNA binding. This will further our understanding of CREB function, the bZIP family and the role of disordered effector domains. I approached this through *in-vitro* biophysical methods, focusing on binding affinity and binding kinetics as described in Chapter 2.

These methods typically require large amounts of pure protein to ensure accuracy and repeatability of experiments. In Chapter 3, I show how I have achieved this by vastly improving available purification protocols to improve yield at no cost to end purity, obtaining high quality protein usable in biophysical assays.

In Chapter 4, I conduct a thorough biophysical characterisation of CREB, including secondary structure, thermostability, binding kinetics to target DNA, binding affinities to

target and non-target DNA, and thermodynamics of DNA binding. I propose that CREB forms beta-sheet secondary structures, unusual for a disordered protein. I then show through multiple methods that CREB has an almost 1000-fold increased affinity to target DNA compared to non-target DNA, alongside extremely high association kinetic rate constants.

Moreover, biophysical characterisation of the standalone bZIP domain in Chapter 5 reveals a new role for disordered effector domains in target search by limiting association to non-target DNA. Indeed, isolated bZIP has a 5-fold increased affinity to target and non-target DNA, and slightly increased homodimerization affinity. I show this greatly impedes bZIP's target search compared to CREB's and derive a model showing the influence of non-target DNA affinity on the search. The influence of each domain on limiting association to DNA is further tested in this Chapter which shows that no single CREB domain is fully responsible for the difference between CREB and bZIP behaviour. i.e. CREB's disordered non-bZIP domains combine to contribute inhibitory effects for DNA binding and thereby improve target-search efficiency. This represents another role of such disordered regions and activation domains for target search.

Lastly, I explore the stoichiometry of target search by CREB in Chapter 6, both by studying CREB homodimerization behaviour and through experiments with forced monomer and dimer mutants to inform mathematical simulations. I show that, unlike previously thought (162), CREB is dimeric in solution. However, forced monomers are virtually incapable of binding to DNA, demonstrating for the first time a bZIP that strongly favours the dimer pathway over the monomer pathway.

Chapter 2

Methods

2.1. Cloning

2.1.1. Typical cloning procedure

Plasmid backbones used were either the pRSET-A backbone with the GB-1 solubility tag (164) and ampicillin resistance, or pET backbone with kanamycin resistance and the SUMO solubility tag (165,166). If inserts over 40 bp were needed, they were ordered through Integrated DNA Technologies. Else, polymerase chain reaction (PCR) mutagenesis was carried out using primers ordered from Life Technologies. PCRs were done using the New England Biolabs (NEB) Q5 hot-start polymerase on an Applied Biosystems™ ProFlex™ thermocycler, following manufacturer's instructions, with products checked on a 1% agarose gel with 1/10000 SYBR Safe. If product plasmid ligation was desired, the product was then treated with the NEB KLD mix, following manufacturer's instructions. If performing Gibson assembly instead, PCR products were cleaned up using the Thermo Scientific GENEJet Gel Extraction and DNA Cleanup kit, before ligation using the NEBuiler® HiFi assembly mix, as per manufacturer's instructions. Both end product types were transformed into chemically competent NEB-5α *E. coli* strains, following manufacturer's instructions. Briefly: 5 µL product was diluted into 50 µL NEB-5α. After incubation on ice for 30 minutes, cells were heat-shocked using a 42 °C water bath for 45

seconds, before a 5-minute recovery period on ice, addition of 950 µL SOC medium and recovery incubation at 37 °C. 100 µL product was plated on 2xTY agar plates. If transformation is of a particularly low efficiency, cells were centrifuged using an Eppendorf Benchtop 5418 centrifuge at 3000 rpm for 5 minutes, 900 µL medium was removed, cells were resuspended in the 100 µL leftover medium and plated immediately.

50-100 µL transformed cells were plated on 2xTY agar plates with the desired antibiotic (final concentration ampicillin: 100 µg/mL, kanamycin: 50 µg/mL) and incubated overnight at 37 °C. Individual colonies were selected and resuspended in 2 mL (for pRSET-A) or 5 mL (pET) 2xTY with the appropriate antibiotic, incubated at 37 °C, shaking at 180 rpm in a New Brunswick Innova® 44 incubator shaker. Plasmids were extracted using QIAGEN's QIAprep Spin Miniprep kit, following manufacturer's instructions. Purified products concentrations were checked using a Nanopore 2000c and plasmids were checked by Sanger Sequencing through Source BioScience.

2.1.2. Homemade competent cells

Except for HiFi assembled plasmids where commercial cells were used, transformations involved homemade chemically competent cells were made from NEB-5α or Rosetta2 (DE3) cells. An aliquot (50-100 µL) was resuspended in a 5 mL 2xTY (with 34 µg/mL chloramphenicol if using Rosetta2). This pre-culture was incubated overnight at 37 °C, shaking 180 rpm. 100 mL 2xTY (with chloramphenicol if applicable) in a conical flask was inoculated the following day with 1 mL pre-culture and incubated at 37 °C until OD_{600} of 0.4-0.6 is reached. Flasks were placed on ice for 20 minutes before centrifugation for 10 minutes at 3000 rpm and 4 °C in a pre-chilled Beckman Allegra x12R centrifuge with a swinging bucket rotor. Supernatant was decanted, discarded and the pellet drained for 2 minutes by inverting the centrifuge tube, before resuspension in $1/5^{th}$ their original volume

in chilled, sterile 100 mM $CaCl_2$, 15% glycerol. After a 20-minute rest on ice, the centrifuge, decanting and draining steps were repeated, but this time cells were resuspended in 1/25th their original volume before aliquoting 50 µL in sterile, chilled microcentrifuge tubes. Aliquots were flash-frozen into liquid nitrogen and stored at -80 °C.

2.2. Protein and oligonucleotides preparation and purification

2.2.1. Full-length CREB and domain deletions purification protocol

2.2.1.a. Protein expression

This section describes the final full-length CREB purification procedure – earlier versions and differences are discussed in Chapter 3. Plasmids expressing CREB were cloned into pET vectors downstream of the SUMO solubility tag and transformed into Rosetta2 (DE3) *E. coli* strains following the heat shock procedure described in 2.1.2. 100 mL 2xTY with kanamycin (50 µg/mL), chloramphenicol (34 µg/mL) and 1% glucose pre-cultures in conical flasks were inoculated with either individual colonies or scraped 25% glycerol stocks and grown overnight in a shaking incubator at 37 °C, 180 rpm. 1 mL pre-culture was used to inoculate 1 L 2xTY-5052 autoinduction medium (with final concentrations: 0.5% v/v glycerol, 0.05% w/v glucose, 0.2% w/v lactose, typically added from a sterile 45x stock) supplied with kanamycin and chloramphenicol, in 2.5 L flasks. Cells were grown at 30 °C, 180 rpm until OD_{600} of 0.6, when temperature was lowered to 18 °C for overnight induction.

2.2.1.b. Heat purification

Cells were harvested by centrifugation at 5000 rpm (~6000 g) for 30 minutes at 4 °C in an Avanti J26XP centrifuge with the JLA 8.1000 rotor. Supernatant was discarded and pellet resuspended in 50 mL per 4 g pellet resuspension buffer (RB, see buffer Table 2.1)

Name	Components	pH
Biophysical buffer	10 mM HEPES, 150 mM NaCl, 10 mM $MgCl_2$, 0.05% Tween-20	7.4
Biophysical buffer pH 6.5	10 mM MES, 150 mM NaCl, 10 mM $MgCl_2$, 0.05% Tween-20	6.5
Resuspension buffer (RB)	100 mM Tris, 50 mM NaCl	8.5
Salt-wash buffer (SW)	100 mM Tris, 2 M NaCl	8.5
Imidazole elution buffer (IEB)	100 mM Tris, 50 mM NaCl, 1.5 M imidazole	8.5
Dialysis buffer	10 mM sodium phosphate, 1 mM EDTA, 2 mM DTT	8
Heparin wash buffer (HWB)	10 mM sodium phosphate, 2 mM DTT	8
Heparin urea buffer (HUB)	10 mM sodium phosphate, 2 mM DTT, 8 M urea	8
Heparin elution buffer (HEB)	10 mM sodium phosphate, 2 mM DTT, 2 M NaCl, 8 M urea	8
SEC buffer	10 mM sodium phosphate, 2 mM DTT, 1 M NaCl, 8 M urea	8
Tris ion-exchange buffer	20 mM Tris, 2 mM DTT	7.6
Tris ion-exchange elution buffer	20 mM Tris, 2 mM DTT, 2 M NaCl	7.6
HEPES wash buffer	10 mM HEPES	7
HEPES eletion buffer	10 mM HEPES, 2 M NaCl	7
Tris-TEV	20 mM Tris, 125 mM NaCl	7.5
ULP-1 lysis buffer	20 mM Tris, 300 mM NaCl, 20 mM imidazole, 1 mM TCEP	8.7
ULP-1 salt-wash buffer	ULP-1 lysis buffer with 2 M NaCl instead of 300 mM	8.7
ULP-1 elution buffer	ULP-1 lysis buffer with 400 mM imidazole instead of 20 mM	8.7
Labelling buffer	20 mM Tris, 500 mM NaCl	7.5
Annealing buffer	30 mM HEPES, 100 mM potassium acetate	7.5
Oligo wash buffer	50 mM Tris, 1 mM EDTA	7.2
Oligo elution buffer	50 mM Tris, 1 mM EDTA, 1.5 M NaCl	7.2
Oxidation buffer	100 mM Tris	9
2xTY	16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, autoclaved	N/A
2xTY Agar	2xTY + 20 g/L agar	N/A

Table 2.1: Buffers used in experiments. All buffers were filter sterilised through 0.22 µm pores, except 2xTY which was autoclaved.

containing 1 mM DTT and Pierce EDTA-free protease inhibitors. Cells were lysed by sonication while kept on ice, 150 mL at a time, with 10 minutes total on time in 10 seconds on/off cycles at 100% amplitude, using the Sonics Vibra-Cell™ VCX-130 and its associated 12 mm tip diameter probe. Lysate was heat-purified through a 75 °C, 30 minutes incubation in centrifuge tubes using a Thermo Fisher Heratherm™ oven before resting on ice for 10 minutes and centrifugation in the Avanti J26XP centrifuge with the JA 25.50 rotor, 21000 rpm (~53000 g), 45 minutes, 4 °C.

2.2.1.c. Ni-NTA affinity chromatography

The heat-soluble CREB fraction was filtered through 0.45 µm then 0.22 µm STARLAB filters, then loaded using a peristaltic pump at a 1-3 mL/min rate onto pre-washed (as per manufacturer's instructions, using DTT-free buffers) Cytiva 5 mL pre-packed HisTrap™ HP columns. Columns were washed with 40-column volumes (CV) RB, 60-CVss salt-wash (SW) buffer, and 5-CVs RB again, both with 1 mM DTT. CREB was eluted through gradient elution using an ÄKTApurifier UPC 100 with imidazole elution buffer (IEB) and RB, both with 1 mM DTT. Gradient was 0-30% for 6-CVs, then only IEB for 10-CVs. Fractions of interest were analysed on pre-cast NuPAGE™ 4 to 12% Bis-Tris gels, with run samples mixed with 4x NuPAGE™ LDS sample buffer, 5 mM Bond-Breaker™ TCEP, and boiled beforehand. Electrophoresis conditions were 200 V, 40 minutes, using NuPAGE™ MES SDS Runing Buffer, in a XCell *SureLock*™ gel electrophoresis system. Desired fractions were pooled, with additions of EDTA to 1 mM and 1/1000 dilution of ULP-1 stock (purified following Section 2.2.3.b), then diluted to 50 mL and dialysed against 2 L of dialysis buffer overnight at 4 °C, using a 10 kDa molecular weight cutoff SnakeSkin™ dialysis tubing. If working with TEV protease instead of ULP-1, TEV was added to 1-5 µM final concentration.

2.2.1.d. Heparin chromatography

Cytiva HiTrap™ heparin pre-packed columns were equilibrated with heparin wash buffer (HWB) and heparin urea elution buffer (HEB). Dialysed CREB was filtered using STARLAB sterile PES 0.22 μm filters and loaded at 1 mL/min using a peristaltic pump. Flowthrough was reloaded on the column to ensure binding. Column was washed with 10-CVs heparin urea buffer (HUB) before elution using a shallow 0-25% (to 500 mM NaCl) HEB gradient on the ÄKTApurifier UPC 100 for 15-CVs, followed by a 25-100% gradient for 5-CVs. Fractions of interest were analysed on pre-cast NuPAGE™ 4 to 12% Bis-Tris gels, as described previously, and by UV-vis spectrophotometry using a Varian Cary® 50 spectrophotometer to measure the A260/A280 ratio indicative of DNA contamination. Fractions contaminant-free protein-wise and with a A260/A280 ratio below 0.5 were pooled, diluted or concentrated to 50 μM (the latter using Amicon® Ultra-15 centrifugal filter units) and twice dialysed against 1 L biophysical buffer each time. Purified CREB was filtered through 0.22 μm pores, concentration recorded through UV-Vis spectroscopy (or, exceptionally, using the Pierce™ Bradford Protein Assay kit as per manufacturer's instructions, against a pre-determined CREB gradient). 0.5-1 mL aliquots were flash-frozen with liquid nitrogen and stored at -80 °C. If aggregation was observed spectrally, it was accounted for using equations derived in (167) which was, at a given wavelength λ in nm:

$$A_c = A_\lambda - c_1\lambda^{-4} - A_{340} + c_1 340^{-4}$$

$$c_1 = \frac{A_{340} - A_{360}}{340^{-4} - 360^{-4}}$$

Where:

A_x denotes uncorrected absorbance at wavelength x in nm.

A_c denotes corrected absorbance value.

2.2.1.e. Optional: anion-exchange chromatography

If contaminants were still present post heparin chromatography, anion-exchange chromatography was undertaken. Heparin eluate was dialysed in Tris ion-exchange buffer instead of dialysis in biophysical buffer, filtered through 0.22 μm pores and loaded at a 1 mL/min rate on a HiTrap™ Q HP pre-packed column using a peristaltic pump. Column was washed with 10-CVs Tris ion-exchange buffer before elution using a Tris ion exchange elution buffer gradient (same gradient as heparin chromatography) on the ÄKTApurifier UPC 100. Fractions are similarly analysed on pre-cast NuPAGE™ 4 to 12% Bis-Tris gels, with desired fractions dialysed against biophysical buffer, processed, and stored as described for the heparin method.

2.2.1.f. Optional: size-exclusion chromatography

If ion-exchange was not possible, size-exclusion chromatography was undertaken, using both Superdex™ 200 HiLoad™® 16/600 or Superdex™ 75 HiLoad™® 26/600 pre-packed columns. Columns were washed with 2-CVs MilliQ® ultrapure water and 3-CVs SEC buffer. CREB was buffer exchanged or concentrated (for 3 successive runs maximum) into SEC buffer, filtered through 0.22 μm pores and loaded on the column using a Cytiva 50 mL Superloop™. Fractions were again analysed on pre-cast NuPAGE™ 4 to 12% Bis-Tris gels, dialysed against biophysical buffer, processed, and stored as described in the heparin chromatography (Section 2.2.1.d).

2.2.2. bZIP-specific purification protocol

2.2.2.a. Protein expression

CREB's bZIP domain was cloned into pRSET vectors downstream of the GB1 solubility tag and transformed into BL21 (DE3) *E. coli* strains following the heat shock

procedure described in 2.1.2. 100 mL 2xTY with ampicillin (100 µg/mL), chloramphenicol (34 µg/mL) and 1% glucose pre-cultures in conical flasks were inoculated with individual colonies or 25% glycerol stocks. Cultures were incubated overnight in a New Brunswick Innova® 44 at 37 °C, 180 rpm. 10 mL pre-culture was used to inoculate 1 L 2xTY culture medium supplied with ampicillin and chloramphenicol, in 2.5 L flasks. Cells were grown at 37 °C, 180 rpm until OD_{600} of 0.6-0.8 is reached. Expression was induced by addition of IPTG to 1 mM for 4h, at 37 °C, 180 rpm. Cells were harvested by centrifugation at 5000 rpm (~6000 g) for 30 minutes at 4 °C in an Avanti J26XP centrifuge with the JLA 8.1000 rotor. Supernatant was discarded and pellets stored at -80 °C pending purification.

2.2.2.b. Ni-NTA affinity chromatography

No DTT was used at this stage. Cells were resuspended in 50 mL RB per 4 g pellet. Cells were lysed by sonication, 150 mL at a time for 5 minutes total on time while kept on ice, 15 seconds on/off cycles at 100% amplitude, using the Sonics Vibra-Cell™ VCX-130 and its associated 12 mm tip diameter probe. Lysates were centrifuged in an Avanti J26XP with the JA 25.50 rotor at 21000 rpm (~53000g) and 4 °C for 45 minutes.

The supernatant was filtered through 0.45 µm then 0.22 µm STARLAB filters, then loaded using a peristaltic pump at a 1-3 mL/min rate onto pre-washed Cytiva 5 mL pre-packed HisTrap™ HP columns. Columns were washed with 40 columns volumes RB, 60-CVs SW buffer, and 5-CVs RB again. bZIP was eluted using an ÄKTApurifier UPC 100 with an IEB gradient. Gradient was 0-30% for 6 columns volumes, then only IEB for 10-CVs. Fractions of interest were analysed on pre-cast NuPAGE™ 4 to 12% Bis-Tris gels, with run samples mixed with 4x NuPAGE™ LDS sample buffer, 5 mM Bond-Breaker™ TCEP, and boiled beforehand – same running conditions as CREB. Desired fractions were pooled, with additions of EDTA and DTT to 1 mM and TEV protease to 2.5 µM, then diluted

to 50 mL and dialysed against 2 L of RB with 1 mM EDTA and DTT overnight at 4 °C, using a 3.5 kDa molecular weight cutoff SnakeSkin™ dialysis tubing.

2.2.2.c. Cation-exchange affinity chromatography

Cytiva HiTrap™ SP-HP pre-packed columns were pre-equilibrated using MilliQ® water, HEPES wash and elution buffers, as per manufacturer's instructions. Dialysed bZIP was filtered using STARLAB sterile PES 0.22 μm filters and loaded at a 1 mL/min rate using a peristaltic pump. Column was washed with 10-CVs HEPES wash buffer before elution using a 0-100% HEPES elution buffer gradient over 20-CVs on the ÄKTApurifier UPC 100. If bZIP mutants had free thiols, 2 mM DTT was added to the buffers. Fractions of interest were analysed on pre-cast NuPAGE™ 4 to 12% Bis-Tris gels, as described above. Fractions contaminant-free fractions were pooled, optionally concentrated using Amicon® Ultra-15 centrifugal filter units and buffer exchanged into biophysical buffer. bZIP was then filtered through 0.22 μm pores, its concentration recorded through UV-Vis spectroscopy, aliquoted into 0.5-1 mL samples, flash-frozen with liquid nitrogen and stored at -80 °C.

2.2.2.d. Optional: reverse Ni-NTA affinity chromatography.

On some occasions, some contaminants were present, most likely TEV or GB1 from the purification procedure. Ion-exchange eluate was dialysed against 2 L RB and loaded onto a HisTrap™ HP pre-packed column using a peristaltic pump, saving the flowthrough. The column was washed with 5-CVs RB, saving the flowthrough again. bZIP tended to bind non-specifically through ionic interactions to the resin and was typically eluted upon addition of 5-CVs SW buffer. IEB was finally added, with the SW and IEB eluates saved. Eluates and flowthroughs were analysed on pre-cast NuPAGE™ 4 to 12% Bis-Tris gels, processing, analysing and storing the fraction with pure bZIP as per cation exchange chromatography.

2.2.3. Protease purifications

2.2.3.a. TEV protease

100 mL 2xTY pre-cultures supplied with chloramphenicol (34 µg/mL) and kanamycin (50 µg/mL) were inoculated with either chemically transformed BL21 colonies or 25% glycerol stocks from previous pre-cultures, and incubated overnight at 37 °C, 180 rpm. 10 mL pre-culture was used to inoculate each 1 L of 2xTY culture with chloramphenicol and kanamycin, incubated at 37 °C, 180 rpm until OD_{600} of 0.8-1.0 is reached. Expression was induced with 1 mM IPTG and carried for 3 h, 37 °C, 180 rpm. Cells were harvested as per the bZIP procedure and pellets stored at -80 °C until purified.

Pellets were resuspended in 35 mL Tris-TEV buffer per 2 g. Resuspended solutions were sonicated on ice using the apparatus described previously, for 5 minutes total on time, 15/45 seconds on/off time. Lysates were centrifuged in a J26XP centrifuge with the JA 25.50 rotor at 18000 rpm for 45 minutes at 8 °C. Supernatant was filtered using 0.22 µm syringe filters and loaded onto pre-packed, equilibrated HisTrap™ HP columns (usually 1 column per 3 L expression) with a peristaltic pump, for at least 1 h at room temperature. Columns were washed with 5-CVs Tris-TEV + 20 mM imidazole and eluted with 5-CVs Tris-TEV + 500 mM imidazole. Eluate was analysed by running on NuPAGE™ 4 to 12% Bis-Tris gels and, if of sufficient purity, 5 mM EDTA was added to the eluate followed by overnight dialysis (enough for a 1/100 dilution) in Tris-TEV. Concentration of dialysed product was measured with 280 nm absorbance using an extinction coefficient of 32400 $M^{-1}cm^{-1}$ from the ProtParam server (168). Purified TEV was then concentrated to roughly 100 µM using Amicon centrifugal filter units pre-washed with Tris-TEV. TEV resuspended in Tris-TEV + 2 mM EDTA, 2 mM DTT, 50% (v/v) glycerol, divided into 200 µL aliquots, flash-frozen in liquid nitrogen and kept at -20 °C.

2.2.3.b. ULP-1 Protease

Plasmids containing ULP-1 protease (for SUMO-tag cleavage) were transformed into chemically competent Rosetta2 (DE3) cells, plated on 2xTY agar with kanamycin (50 µg/mL) and chloramphenicol (34 µg/mL). 100 mL 2xTY pre-cultures, supplied with the same antibiotics, were inoculated with colonies or scraped 25% glycerol stocks and incubated overnight, 37 °C, 180 rpm. 4 L of 2xTY medium with antibiotics were inoculated with 10 mL pre-culture for each L culture, and grown at 37 °C, 180 rpm until OD_{600} of 0.5 was reached. Expression was induced with 1 mM IPTG for 4 h at the same conditions. Cells were harvested by centrifugation as described previously and stored at -80 °C.

Pellets were resuspended in ULP-1 lysis buffer, 50 mL per 4 g pellet, before lysis by sonication and centrifugation as per the bZIP protocol. All steps were then carried out at 4 °C to limit ULP-1 aggregation. Supernatant was filtered through 0.22 µm pores before loading on a pre-equilibrated HisTrap™ HP column, using a peristaltic pump at a 1 mL/min flow rate. Column was washed with ULP-1 salt wash buffer to remove nucleic acids for 1 h at 5 mL/min, before elution with 3-CVs ULP-1 Elution Buffer. Each step flowthrough was kept and analysed on NuPAGE™ 4 to 12% Bis-Tris gels to ensure purity. As soon as possible to limit precipitation, glycerol was added to the eluate at 10-20% (v/v) final concentration before flash freezing 50-100 µL aliquots in liquid nitrogen and storage at -80 °C. Protease from this protocol was found to completely cleave SUMO-CREB (auto-induced in 4 L 2xTY) post Ni-NTA affinity chromatography at 1/1000-1/10000 dilutions.

2.2.4. Preparation of oligonucleotides

Oligonucleotides were ordered desalted from Thermo Fisher or Integrated DNA Technologies. Oligonucleotide sequences are listed in Table 2.2, with CREh a self-annealing hairpin described in (144,169). DNA was resuspended to 1 mM in annealing buffer, equal

Name	Sequence (5'-3')	ε ($M^{-1}cm^{-1}$)*	GC (%)
CREh	CCTGACGTCAGCCCCCTGACGTCAGG	234800	69
CRElin f	CCTGACGTCATCCG	236296	64
CRElin r	CGGATGACGTCAGG		
nCRE f	GGCTAAAGCATTCT	225537	43
nCRE r	AGAATGCTTTAGCC		

Table 2.2: List of oligonucleotides sequences and properties. * indicates a double-stranded oligos value, calculated using Molbiotools's DNA calculator, available at https://molbiotools.com/dnacalculator.php.

molar amounts of forward and reverse primer were mixed and diluted to 100 µM. Hairpins were simply diluted to 100 µM. Oligos were annealed in an Applied Biosystems™ ProFlex™ thermocycler by denaturation for 5 minutes at 92 °C before slowly ramping down incubation temperature by 0.1 °C/s until 4 °C. Separation of unannealed or incomplete oligos was carried out by ion-exchange chromatography, with typical yields of 85%. Annealed products were diluted in oligo wash buffer before loading on a HiTrap Q HP pre-packed column, either with a syringe or a Superloop™. Column was washed with 5-CVs oligo wash buffer before elution through a 0-100% gradient of oligo elution buffer lasting 50-CVs at 2.5 mL/min on a FPLC which was pre-washed by overnight contact with 1 M NaOH at 0.1 mL/min. Fractions of interest were analysed on a 4% agarose gel with 1/25000 GelGreen® nucleic acid stain, with chosen fractions concentrated and buffer exchanged using Amicon centrifugal filter units before storage at -20 °C.

2.2.5. Maleimide labelling

A modified protocol from (155) was used. Briefly: protein was diluted to 10-20 µM and dialysed in 1 L labelling buffer with 1 mM TCEP for reduction of cysteines. TCEP was removed by buffer exchange into labelling buffer using Cytiva PD-10 gravity columns and protein was labelled overnight at 4 °C on a centrifuge tube roller with 4-fold molar excess

fluorophore, either Alexa Fluor™ 488 or 594 C5 maleimide. Reaction was quenched with 1 mM DTT. Dye, unlabelled and labelled proteins were dialysed post-labelling into Tris ion-exchange buffer and separated using anion-exchange chromatography, following the same method as described above for CREB, only gels were checked using a Syngene G:BOX XT-16 fluorescence imager before staining. Extent of labelling was also checked by LC-MS mass spectrometry on an Orbitrap Elite ETD.

2.2.6. Forced dimer generation

CREB-GGC or bZIP-GGC was buffer exchanged into oxidation buffer and incubated overnight at 37 °C, 180 rpm, ensuring oxygen could reach the solution to permanently link monomers together via disulfide bond formation, as shown in (131). Only CREB-GGC was then subjected to either denaturing anion-exchange chromatography or size-exclusion chromatography as described for CREB, only with 8 M urea in all buffers. The extent of forced dimerization was checked both by SDS-PAGE gels as described above and by LC-MS mass spectrometry.

Specific to the catalysis test, samples were reduced with 10 mM DTT overnight before buffer exchange into 100 mM Tris-HCl pH 8. Tested catalysts were:

- Simple heat oxidation control with no catalysts.
- 1 M DMSO.
- 5 mM Potassium ferricyanide ($K_3[Fe(CN)_6]$).
- 5 mM $CuCl_2$.
- 3 mM : 0.3 mM Glutathione reduced : oxidised.

Each reaction was run with either no denaturants or with 1.5 M GdnHCl to test if aggregation was affecting dimerization.

2.2.7. Western blot

Western blot were carried out to find contaminants. Filter paper and transfer membranes were soaked in NuPAGE™ transfer buffer supplied with methanol to 10% (v/v) and assembled in an Invitrogen XCell II™ Blot Module, as per manufacturer's instructions, before electrotransfer at 30 V for 1 h. Membrane was rinsed with MilliQ® water and blocked in 5% (w/v) milk for 1 h at room temperature, before overnight contact in primary antibody diluted 1:1000 in 5% milk. Membrane was washed 15 minutes with PBST three times before 1 h contact in 1:10000 secondary antibody diluted in 5% milk. Membrane was washed 15 minutes with PBST three times before imaging. Antibody combinations were: primary mouse Promega Anti-HaloTag® monoclonal antibody (G921A) and Cell Signalling Technology anti-mouse IgG, HRP-linked antibody (#7076); or, anti-CREB rabbit monoclonal antibody (#9197) and anti-rabbit IgG, HRP-linked antibody (#7074), both from Cell Signalling Technology.

2.3. Biophysical methods

Unless specified otherwise, all experiments were carried out in biophysical buffer and curve fitting was performed on Python3 using the lmfit package (170) or on GraphPad Prism version 10.0.0 for Windows, GraphPad Software, Boston, Massachusetts USA, www.graphpad.com, using the Levenberg-Marquardt least-squares algorithm.

2.3.1. Determination of extinction coefficients

Extinction coefficients were determined using the Gill, Von Hippel method (171). Briefly: proteins were diluted 1 in 4 with an 8 M > 99% purity GdnHCl solution to 6 M GdnHCl. Unfolded proteins' extinction coefficients were calculated from primary sequence since unfolded tyrosine, tryptophan and cysteine disulfide bonds extinction coefficients values are known. Concentrations were obtained from unfolded proteins absorbance spectra using Hellma® Quartz 1 cm pathlength cuvettes. Protein absorbance spectra from the same stock in biophysical buffer were taken and their extinction coefficients were back calculated since protein stock concentration is now known. Calculated coefficients are shown in Table 2.3.

Name	Number of absorbing residues	Denatured ε ($M^{-1}cm^{-1}$)	Folded ε ($M^{-1}cm^{-1}$)
CREB	7 Y	8960	9100 ± 100
CREB-A	7 Y	8960	9100 ± 100
CREBΔQ1	7 Y	8960	8200 ± 150
CREBΔQ2	3 Y	3840	3840 ± 100
CREB GS-KID	6 Y	7680	8110 ± 90
CREB-GGC	14 Y, 2 C	18160	19000 ± 370
CREB L318V	7 Y	8960	N/A (Bradford assay)
CREBΔLZ	6 Y	7680	8000 ± 250
bZIP	2 Y	2560	2630 ± 30
bZIP-GGC	4 Y, 2 C	5630	5450 ± 100

Table 2.3: Extinction coefficients of CREB mutants. "Disordered" values were calculated using Gill, von Hippel extinction coefficient values for unfolded amino acids absorbing at 280 nm. Averaged "Folded" values with standard deviation errors from 3 or more replicates were obtained following the Gill, von Hippel method. Folded values are very close to denatured values.

2.3.2. Biophysical values relevant for the study of CREB interactions

A standard two-state reaction scheme for CREB binding to DNA would be:

$$\mathrm{CREB} + \mathrm{DNA} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} \mathrm{CREB:DNA}$$

where k_{on} is the association rate constant and k_{off} is the dissociation rate constant. k_{on} is usually a second-order rate constant in $M^{-1}s^{-1}$ units and k_{off} a first-order rate constant in s^{-1} units.

At equilibrium, there is an equal amount of flux through both association and dissociation reactions, and concentrations of free and bound species will be constant. This means the following, as derived previously (172):

$$k_{\mathrm{on}}[\mathrm{CREB}]_{free}[\mathrm{DNA}]_{free} = k_{\mathrm{off}}[\mathrm{CREB:DNA}]$$

$$\frac{[\mathrm{CREB}]_{free}[\mathrm{DNA}]_{free}}{[\mathrm{CREB:DNA}]} = \frac{k_{\mathrm{off}}}{k_{\mathrm{on}}} = K_{\mathrm{D}}$$

Where K_{D} is the dissociation constant, in concentration units (M), a measure of binding affinity. Since complex concentration is the denominator and free species in the numerator, a lower number means more of the species are in complex and therefore higher affinity.

The above can be applied to CREB homodimerization too:

$$\mathrm{M} + \mathrm{M} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{D}$$

$$\frac{[\mathrm{M}]^2}{[\mathrm{D}]} = \frac{2k_2}{k_1} = K_{\mathrm{D,homo}}$$

where M is the monomer and D is the dimer species, k_1 and k_2 are the homodimerization association and dissociation rate constants respectively.

Hence, CREB's possible routes of binding target and non-target DNA can be represented by a simplified kinetic scheme (Figure 2.1):

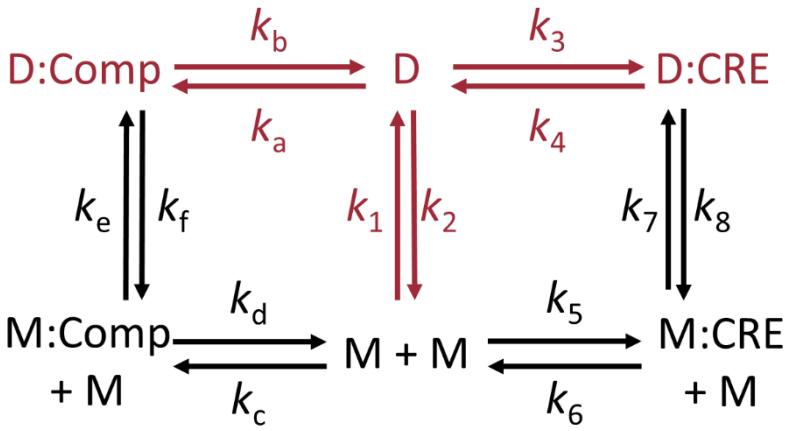


Figure 2.1: Simplified kinetic scheme of CREB binding target and non-target DNA. All isolated reaction schemes and associated kinetic rate constants of CREB binding competitor (Comp) and target (CRE) DNA. Black routes are preferred by monomer pathway, dark red routes by the dimer pathway.

2.3.3. Fluorescence anisotropy equilibrium titrations

Total labelled target DNA CREh (1 nM), non-target DNA nCRE (5 nM) (Table 2.2) or CREB concentrations were kept constant throughout the experiment by preparing one sample for each titrant concentration, with assays at 25 °C. For higher accuracy, dilutions were performed (and exact weights recorded) using an analytical balance, and samples were rested for at least 1 hour away from light before measurements. For each sample, between 3 and 10 measurements were taken and averaged. Fluorescence measurements were recorded on a Horiba FluoroMax-4 Spectrophotometer with a fluorescence polarization accessory, using excitation/emission wavelengths of 495/519 nm respectively for Alexa Fluor™ 488, or 590/618 nm excitation/emission respectively for Alexa Fluor™ 594, with FRET measurements between this fluorophore pair using excitation/emission wavelengths of 495 nm/618 nm. During FRET experiments, fluorescence intensity measurements were always recorded for the individual fluorophores as well. This enabled FRET efficiency calculations

using the following equation with the acceptor FRET signal intensity (F_a) and donor fluorescence signal intensity (F_d), both upon excitation of the donor only:

$$\text{FRET efficiency} = \frac{F_a}{F_d + F_a} \tag{2.1}$$

Samples were maintained at 25.0 °C using an external water bath probe. A 3 mm pathlength quartz cuvette was used (Hellma®). Anisotropy measurements were directly recorded using polarizing filters in VV, VH, HV and HH format, but can be back-calculated:

$$Anisotropy = \frac{I_{VV} - I_{VH}}{I_{VV} + 2I_{VH}} \tag{2.2}$$

With I being the intensity recorded using the different filters. V denotes vertically polarised light and H horizontally polarised light, with the first letter indicating the excitation direction and the last letter the emission direction. Anisotropy binding titrations were fitted to one of the following models:

Equation 2.3 for a single DNA-binding site, as per (173,174):

$$A = R_f + \Delta R\left(\frac{K + D + x - \sqrt{(K + D + x)^2 - 4Dx}}{2D}\right) \tag{2.3}$$

Equation 2.4 for two DNA-binding sites with different binding affinities:

$$A = R_f + \Delta R\left(\frac{K + D + x - \sqrt{(K + D + x)^2 - 4Dx}}{2D}\right) + \Delta R_2\left(\frac{K_2 + D + x - \sqrt{(K_2 + D + x)^2 - 4Dx}}{2D}\right) \tag{2.4}$$

Equation 2.5 for homodimerization corrected anisotropy (derived in Appendix 1):

$$A = R_f + \Delta R\left(\frac{[Dim] + D + K - \sqrt{([Dim] + D + K)^2 - 4D[Dim]}}{2D}\right) \tag{2.5}$$

With:

$$[Dim] = \frac{(4x + L) - \sqrt{(4x + L)^2 - 16\left(x^2 + L\frac{D(A - R_f)}{\Delta R}\right)}}{8}$$

Equation 2.6 for CREB dilution, derived in Appendix 2:

$$A = R_f + \Delta R\left(0.5\ + L\frac{1-\sqrt{1+\frac{8x}{L}}}{8x}\right) \tag{2.6}$$

Where:

A = anisotropy

R_f = the initial anisotropy value

ΔR_n = the difference between the saturated and initial anisotropy value

$K_n = K_D$ in nM, of DNA binding

L = homodimerization K_D.

D = the probe concentration (typically DNA), in nM

x = titrant concentration (typically CREB), in nM

Intensity, if not recorded directly, was calculated using the following equation (173):

$$I = I_{vv} + 2G \times I_{vh} \tag{2.7}$$

Where $G = \frac{I_{hv}}{I_{hh}}$ is the G factor, and I_{xx} are polarised intensities given by the fluorimeter.

If an intensity change was noticed between bound and unbound species, anisotropy and intensity were fit by a single-site model and anisotropy corrected following the formula (173,175) (Equation 2.8):

$$A_{\mathrm{adj}} = \frac{\frac{R - R_{\mathrm{f}}}{R_{\mathrm{b}} - R} \times \frac{Q_{\mathrm{f}}}{Q_{\mathrm{b}}} \times R_{\mathrm{b}} + R_{\mathrm{f}}}{1 + \frac{R - R_{\mathrm{f}}}{R_{\mathrm{b}} - R} \times \frac{Q_{\mathrm{f}}}{Q_{\mathrm{b}}}} \tag{2.8}$$

Where:

R = anisotropy measurement of the sample

Q = intensity measurement of the sample

R_f, Q_f = anisotropy and intensity of the free sample, respectively, obtained from fitting.

R_b, Q_b = anisotropy and intensity of the bound sample, obtained from fitting (e.g., $R_b = R_f + \Delta R$)

2.3.4. Stopped-flow kinetics.

The stopped-flow procedure was described by (173). Briefly, measurements were performed on an Applied Photophysics SX-20 Stopped Flow Spectrometer, monitoring fluorescence intensity, with 2 mm slit widths for both excitation and emission. Two independent syringes were filled with samples of interest which were quickly mixed in a reaction chamber. Mixing can be 1:1 volume ratio (3 bar pressure) or 1:10 (2.2 bar pressure). If acquiring for less than 2 seconds, a pressure-hold was applied to the syringes. A cutoff for deadtime was applied, often 1-3 ms if pressure-hold was on, A 100 ms cutoff was used if mixing 1:10 without pressure hold as artefacts in this time range were observed in this setup. All solutions were weighed during preparation on an analytical balance as these provided greater accuracy than by pipetting alone. The following long-pass filters were used:

- 515 nm filter at an excitation wavelength of 495 nm, used with Alexa Fluor™ 488.
- 610 nm filter at an excitation wavelength of 590 nm, used with Alexa Fluor™ 594.
- 305 nm filter at an excitation wavelength of 278 nm, used with tyrosine fluorescence.

The change in fluorescent signal was monitored over time on a linear or logarithmic scale, and observed rate constants were estimated by averaging traces and fitting to the appropriate function.

2.3.4.a. Association rate constants (k_{on})

One syringe was filled with CREB (often 200-800 nM), the other with 10 nM labelled CREh-Alexa488 and, if studying the influence of competitors, nCRE (1-20 μM). 50+ individual traces were averaged in 1:1 mixing setup and fit to one of the following functions:

$$y = ae^{-k_{\text{obs}}x} + c \text{ (single exponential decay)}$$

$$y = ae^{-k_{\text{obs}}x} + a_2e^{-k_{\text{obs},2}x} + c \text{ (double exponential decay)}$$

Where:

a_n, c_n = constants

y = fluorescence signal measured in V

x = time in seconds (s)

$k_{obs,n}$ = observed rate constant for the reaction, in s^{-1}

Appropriate timescales for traces were chosen to be approximately 10 half-lives (173) which is, for a single exponential decay fit:

$$\frac{10\ln(2)}{k_{\text{obs}}}$$

For a simple two-state binding process with 1:1 binding ratio, k_{obs} is linearly dependent upon protein concentration due to the pseudo first-order conditions maintained in the experiment, meaning at least 10-fold protein excess compared to DNA:

$$y = k_{\text{on}} \times [CREB] + k_{\text{off}} \quad (2.9)$$

although k_{off} is generally poorly estimated using this equation due to extrapolation.

2.3.4.b. Dissociation rate constants (k_{off})

Dissociation rate constants were determined using an out-competition method. One syringe contained a pre-mixed solution of labelled complex (10 nM CREh-Alexa488 with 40 nM protein), the other syringe contained unlabelled CRElin in large excess (2-20 μM). k_{obs} were obtained by fitting 30+ traces to single or double exponential decay functions (see association kinetics) and plotted against competitor concentrations. At high enough [CRElin], dissociation is essentially irreversible, and the observed rate constant is simply k_{off}, independent of concentration. Straight line fitting this plateau gives the k_{off}, with the equation:

$$y = k_{\mathrm{off}} \tag{2.10}$$

2.3.4.c. CREB homodimer dilution perturbation (k_2, $K_{D,homo}$)

Labelled CREB-Alexa488 or CREB-Alexa594, 200 nM or less was perturbed by an 11-fold dilution into biophysical buffer using a 1:10 setup. Relaxation to equilibrium was monitored through fluorescence intensity. 30+ traces were averaged and paired averaged traces from different CREB concentrations, from the same day, taken without changing instrument calibration, were simultaneously fit to the following equations derived both in (176) and again by the Shammas lab (Equation 2.11):

$$y = Ff_{in} + \Delta F \times \frac{ze^{-Aa}}{\frac{1}{N} - e^{-Aa}} \tag{2.11}$$

With:

$$A_0 = \frac{K_{\mathrm{D}}}{11} \times \sqrt{0.0625 + 5.5 \times \frac{P}{K_{\mathrm{D}}}} - 0.25$$

$$z = 2K_{\mathrm{D}} \times \sqrt{0.0625 + \frac{P}{2K_{\mathrm{D}}}}$$

$$N = \frac{\frac{K_D - 2z}{4} + A_0}{\frac{K_D + 2z}{4} + A_0}$$

$$Aa = k_2 \times x \times \sqrt{1 + \frac{8P}{K_D}}$$

Where:

K_D =homodimerization dissociation constant

P = protein concentration post dilution (nM)

y = fluorescence signal measured in V

x = time in seconds (s)

k_2 = dissociation rate constant of CREB homodimer (s^{-1})

ΔF = constant relating protein concentration to fluorescent signal strength.

Ff_{in} = fluorescence signal at the start of the experiment.

2.3.4.d. bZIP Chevron unfolding and refolding (k_2, k_1, $K_{D,homo}$)

bZIP stock solution in biophysical buffer alone (unfolding experiments) or with 8 M urea (refolding experiments) was diluted 11-fold using a 1:10 stopped-flow setup into solutions of increasing (unfolding) or decreasing (refolding) urea concentrations. A minimum of 30 traces were obtained for each dilution, monitoring intrinsic tyrosine fluorescence. Traces were fit using a model considering unfolding-refolding equilibrium, equivalent to the homodimerization reaction scheme in Section 2.3.2 assuming dimerization leads to folding. These were derived in Appendix 2, or as previously described (177). However, for unfolding into urea, a single exponential decay fit was found to be sufficient, likely because unfolding could not be reversed in high urea concentrations. Those unfolding-refolding models were, for unfolding into urea:

$$Y = \mathrm{Fc} \times P + \Delta\mathrm{F} \times \left(0.5 - \frac{e^{-k_2 x \sqrt{1+4\left(\frac{P}{K_\mathrm{D}}\right)}}}{2\left(\frac{1}{K_\mathrm{D}+z} e^{-k_2 x \sqrt{1+4\left(\frac{P}{K_\mathrm{D}}\right)}} - \frac{1}{K_\mathrm{D}-z}\right)} \right) \qquad (2.12)$$

(reversible unfolding, Equation 2.12 as per (177))

With:

$$z = K_\mathrm{D} \sqrt{1 + \frac{4P}{K_\mathrm{D}}}$$

Where:

x is time elapsed in seconds.

K_D is the unfolding dissociation constant.

k_2 is the unfolding rate constant.

P is the total protein concentration.

ΔF is a constant relating fluorescence change to molar protein concentration.

Fc is the initial fluorescence signal.

And for refolding into lower urea concentrations (derived in Appendix 3):

$$Y = \mathrm{Fc} + \Delta\mathrm{F} \left(\frac{\frac{K_\mathrm{D}-2z}{4} - \frac{K_\mathrm{D}+2z}{4} \times N e^{-k_1 z x}}{N e^{-k_1 z x} - 1} \right) \qquad (2.13)$$

(reversible refolding, Equation 2.13)

With:

$$z = \sqrt{\frac{{K_\mathrm{D}}^2}{4} + 2K_\mathrm{D}P}$$

$$N = \frac{\frac{K_\mathrm{D}-2z}{4} + A_0}{\frac{K_\mathrm{D}+2z}{4} + A_0}$$

Where:

x is time elapsed in seconds.

K_D is the unfolding dissociation constant.

k_1 is the folding association rate constant.

P is the total protein concentration.

ΔF is a constant relating fluorescence change to molar protein concentration.

Fc is the initial fluorescence signal.

A_0 is the initial concentration of unfolded protein, which is equal to P.

Rate constants and K_D without urea were extrapolated using the linear relationship between $\ln(k_1)$, $\ln(k_2)$ or $\ln(K_D)$ with urea concentration and finding the y-intercept. Linear regression gradients (here, noted "a_x") of $\ln(k_x)$ against urea concentration are used to give the m-values of unfolding (m_u) or folding (m_f) from $m_x = a_x/R$, with R being the gas constant.

2.3.4.e. Competitor inhibited association and the segregation model.

A CREB solution was rapidly mixed 1 to 1 with a solution of CREh labelled with Alexa Fluor™ 488 (5 nM final concentration) and different amounts of unlabelled nCRE competitor. Traces were fit with single exponential decay functions to extract observed rate constants k_{obs}, although double exponential function were more appropriate at very high competitor concentrations. At each competitor concentration, $k_{on,comp}$ was extracted by linear regression of k_{obs} against mutant concentration, as described in Section 2.3.4.a.

A segregation model for $k_{on,comp}$ behaviour with varying nCRE concentration behaviour was derived from the assumption that $k_{on,comp}$ would depend on how much CREB was segregated away by nCRE DNA, reaching rapid equilibrium (Equation 2.14):

$$k_{\mathrm{on,comp}} = k_{\mathrm{on,app}} \times \frac{[CREB]_{free}}{[CREB]_{total}} \qquad (2.14)$$

The dissociation constant for nCRE is equal to:

$$K_{\mathrm{D,nCRE}} = \frac{[CREB]_{free} \times [nCRE]_{free}}{[CREB{:}nCRE]} \tag{2.15}$$

Since $[CREB{:}nCRE] = [CREB]_{total} - [CREB]_{free}$, this can be rearranged to yield:

$$[CREB]_{total} - [CREB]_{free} = \frac{[CREB]_{free} \times [nCRE]_{free}}{K_{\mathrm{D,nCRE}}} \tag{2.16}$$

$$\frac{[CREB]_{total} - [CREB]_{free}}{[CREB]_{free}} = \frac{[nCRE]_{free}}{K_{\mathrm{D,nCRE}}} \tag{2.17}$$

$$\frac{[CREB]_{total}}{[CREB]_{free}} = \frac{[nCRE]_{free}}{K_{\mathrm{D,nCRE}}} + 1 \tag{2.18}$$

$$\frac{[CREB]_{total}}{[CREB]_{free}} = \frac{[nCRE]_{free} + K_{\mathrm{D,nCRE}}}{K_{\mathrm{D,nCRE}}} \tag{2.19}$$

Assuming nCRE concentration remains constant, Equation 2.19 can be plugged into Equation 2.14 to yield the equation to illustrate the segregation model (Equation 2.20):

$$k_{\mathrm{on,comp}} = k_{\mathrm{on,app}} \times \frac{K_{\mathrm{D,nCRE}}}{[nCRE] + K_{\mathrm{D,nCRE}}} \tag{2.20}$$

2.3.5. SEC-MALS

MALS experiments were performed during size-exclusion chromatography, loading protein at a maximum concentration of 40 µM on a Superdex200 Increase 10/300 GL column (GE Healthcare) equilibrated with biophysical buffer. Elution was monitored via online static light-scattering (DAWN HELEOS-8+, Wyatt Technology), differential refractive index Optilab T-rEX, Wyatt Technology) and UV (SPD-20A, Shimadzu) detectors. Data were analysed using the ASTRA package v6 (Wyatt Technology)

2.3.6. Analytical ultracentrifugation (AUC) sedimentation velocity

Sedimentation velocity experiments were performed with a Beckman An-60 Ti rotor in a Beckman Optima XL-1 analytical ultracentrifuge at 20 °C with a rotor speed of 40,000 rpm. 40 μM protein was prepared in biophysical buffer without Tween-20 and filtered through 0.22 μm pores. Radial absorbance scans were recorded using absorbance optics at 280 nm in continuous scan mode. Scans were fitted to a continuous size distribution using SEDFI (178), and buffer density and viscosity as well as sample partial specific volumes from amino acid sequences were calculated with SEDNTERP (179).

2.3.7. Differential scanning calorimetry (DSC)

20 μM protein in biophysical buffer was loaded on a Malvern (Malvern, UK) VP-PEAQ DSC instrument. Samples and buffer blanks were measured using a thermal ramp of 200 °C/hour between 10 °C and 110 °C at 55 psi pressure. Heat capacity vs temperature data were analysed by MicroCal PEAQ-DSC software.

2.3.8. isothermal titration calorimetry (ITC)

ITC titrations with at least 5 μM protein were performed on a Malvern Panalytical MicroCal PEAQ-ITC with 19 injections (0.4 μL for the first injection, 2 μL subsequently). Conditions were 25 °C with high feedback, 750 rpm stir speed, 60 seconds delay, 150 seconds injection spacing, 4 seconds injection duration. Each titration was control corrected by dilution of the same titrant solution into biophysical buffer. Titrations were fit to a single-site binding model using in-built equations from the MicroCal PEAQ-ITC analysis software.

2.3.9. Circular dichroism (CD) and thermal unfolding

CD measurements were taken on a Jasco J-815 equipped with a CDF-426S/15 Peltier, with hydrogen gas flowing through the instrument and at least 5 μM protein.

Temperature was controlled by a Jasco MCB-100 mini circulation bath. Scans were taken using a 0.2 mm or 1 mm pathlength high-grade quartz cuvette from Hellma®, typically with the following settings: 8 seconds D.I.T. (digital integration time), 5 seconds accumulations, 20 nm/min, 0.1 nm step size. CREB spectra values were subtracted using biophysical buffer blank spectra values and an 11-values moving window was applied to spectra to smoothen data and reduce noise (average of a central value and 5 values before and after).

Spectra were analysed for secondary structure content using the BeStSel webserver (180) which converted values from millidegrees to Δε ($M^{-1}cm^{-1}$) using:

$$\Delta\varepsilon = \frac{x}{3298 \times 10 \times L \times [\mathrm{CREB}] \times AAL} \tag{2.21}$$

Where:

x = CD values in millidegrees (mdeg)

L = cuvette pathlength in cm

[CREB] = CREB concentration in molar (M)

AAL = amino acid number (or length) in the protein

Thermal unfolding, or melt, was achieved by monitoring CD signal at 222 nm, as it is indicative of helicity, with an 0.1 °C/second temperature increase rate and 8 seconds D.I.T. A data-smoothing 5-values moving window was applied. Linear unfolded and folded baselines were manually extracted from the data and the CREB fraction unfolded calculated at each temperature with Equation 2.22:

$$F_u = \frac{\Delta\varepsilon - l_b}{u_b - l_b} \tag{2.22}$$

Where:

l_b is the folded baseline at the current temperature.

u_b is the unfolded baseline at the current temperature.

Fraction unfolded were fit to extract melting temperatures and unfolding K_D using a form of the Gibbs-Helmholtz equation adapted from (181,182) (Equation 2.23):

$$F_u = \frac{-K_D + \sqrt{K_D^2 + 8P \times K_D}}{4P} \quad (2.23)$$

With:

$$K_D = e^{-\frac{\Delta G}{RT}}$$

$$\Delta G = RT \times log(P) + \Delta H\left(1 - \frac{T}{T_m}\right) - \Delta C_p\left((T_m - T) + T \times log\left(\frac{T}{T_m}\right)\right)$$

Where:

T is the temperature of measurement in Kelvin (K).

T_m is the melting temperature in Kelvin (K).

K_D is the unfolding dissociation constant.

P is the total molar protein concentration.

R is the gas constant in J/(mol.K).

ΔG is the Gibbs free energy in J/mol.

ΔH is the enthalpy in J/mol.

ΔC_p is the heat capacity change of unfolding in J/K.

A plot of log(K_D) against1/T is linear from which the K_D at 25 °C is extracted.

2.3.10. FIDA (flow-induced dispersion analysis)

FIDA is a technique for measuring hydrodynamic radii of fluorescent species through fluorescent signals and laminar flow. Upon addition of titrants, the radius changes due to complex formation, and this allows the technique to find out affinities from the pM to mM range. Fluorescent CREBΔLZ, unlabelled CRElin and unlabelled nCRE were sent to the Research Complex at Harwell and loaded on a Fida 1 instrument by Fidabio. Labelled

protein was titrated with the oligonucleotides (each in separate experiments) and R_h measured with each new addition of titrant.

2.3.11. Electrophoretic mobility shift assay

Mixtures of CREB with labelled DNA with Alexa Fluor™ 488 were mixed with Novex™ Hi-Density TBE sample buffer and run on Invitrogen™ Novex™ 4-20% TBE gels in a XCell *SureLock*™ apparatus filled with 1x TBE running buffer. Constant 200 V current was run through the gel for 1 h at room temperature before imaging the gel via fluorescence in a Syngene G:BOX XT-16 imager.

2.4. Simulations of CREB behaviour with large excess of non-target DNA

` Simulations of CREB and bZIP behaviour were realised in Wolfram Mathematica 13.0.0 by simultaneously solving all the following equations, applicable to the kinetic scheme (Figure 2.1), substituting where applicable with the values obtained in this Thesis.

$$[\mathrm{CREB}]_{\mathrm{total}} = [\mathrm{M}] + [\mathrm{M{:}\,CRE}] + [\mathrm{M{:}\,nCRE}] + 2([\mathrm{D}] + [\mathrm{D{:}\,CRE}] + [\mathrm{D{:}\,nCRE}])$$

$$[\mathrm{nCRE}]_{\mathrm{total}} = [\mathrm{nCRE}]_{\mathrm{free}} + [\mathrm{M{:}\,nCRE}] + [\mathrm{D{:}\,nCRE}]$$

$$[\mathrm{CRE}]_{\mathrm{total}} = [\mathrm{CRE}]_{\mathrm{free}} + [\mathrm{M{:}\,CRE}] + [\mathrm{D{:}\,CRE}]$$

$$K_{\mathrm{D,homo}} = \frac{[\mathrm{M}]^2}{[\mathrm{D}]}$$

$$K_{\mathrm{D,nCRE,D}} = \frac{[\mathrm{D}][\mathrm{nCRE}]}{[\mathrm{D{:}\,nCRE}]}$$

$$K_{\mathrm{D,CREh,D}} = \frac{[\mathrm{D}][\mathrm{CRE}]}{[\mathrm{D{:}\,CRE}]}$$

$$K_{\mathrm{D,nCRE,M}} = \frac{[\mathrm{M}][\mathrm{nCRE}]}{[\mathrm{M{:}\,nCRE}]}$$

$$K_{\mathrm{D,CRE,M}} = \frac{[\mathrm{M}][\mathrm{CRE}]}{[\mathrm{M{:}\,CRE}]}$$

Where:

M is the free monomer CREB species.

D is the free dimeric CREB species.

CRE and nCRE are target and non-target DNA, where a column (":") denotes a complex.

Importantly, there are 2 more equilibrium equations that pertain to the kinetic scheme:

$$K_{\mathrm{D,homo,CREh}} = \frac{[\mathrm{M{:}\,CRE}][\mathrm{M}]}{[\mathrm{D{:}\,CRE}]}$$

$$K_{\mathrm{D,homo,CREh}} = \frac{[\mathrm{M{:}\,nCRE}][\mathrm{M}]}{[\mathrm{D{:}\,nCRE}]}$$

However, because of the presence of two thermodynamic squares, these do not need to be included in the analysis, as they are difficult to obtain.

CREB's cellular concentration ($[\mathrm{CREB}]_{\mathrm{total}}$) was kept constant 570 nM (183), although it is likely higher in the nucleus. There are around 5000 (184–186) CREB binding sites reported, although a low 900 sites value is also reported (187). Since a human cell nucleus volume is 500 fL (188), the concentration of CRE sites ($[\mathrm{CRE}]_{total}$) was kept constant at the 35 nM estimate calculated using:

$$[\mathrm{CRE}] = \frac{2 \times n_{\mathrm{CRE}}}{V_{\mathrm{nucleus}} \times N_A} \tag{2.24}$$

Where:

N_A is the Avogadro constant.

V_{nucleus} the nucleus volume in litres.

n_{CRE} the total number of CRE sites.

Since there are 3.088 billion DNA base pairs in a human genome (189), the amount of 14 bp nCRE sites (as per the oligo length) will be, assuming all of the DNA is accessible to CREB:

$$[\mathrm{nCRE}] = \frac{2 \times \frac{3.088 \times 10^9 - (8 * n_{\mathrm{CRE}})}{length_{\mathrm{nCRE}}}}{V_{\mathrm{nucleus}} \times N_A} = 1.5 \times 10^{-3}\mathrm{M} = 1.5\mathrm{mM} \qquad (2.25)$$

Although this increases to 2.6 mM if considering that the length of non-target binding sites should be 8 bp, as per CRE.

Chapter 3

Producing pure CREB samples suitable for biophysical studies

CREB, being a long, disordered transcription factor, is prone to many issues during purification. Long disordered proteins frequently interact non-specifically with many other proteins (29,190) increasing the level of contamination and are especially likely to aggregate at high concentration and be toxic to bacteria. Disordered proteins, particularly long ones, are also far more prone to digestion than folded proteins (191) so samples are often contaminated with partially truncated versions of the product. Furthermore, as a transcription factor, it is likely that DNA would contaminate the final product. For biophysical assays however, the final purified product needs to be of very high quality to enable accurate determination, and DNA must not be present as it could outcompete oligos used in downstream experiments. Both factors would lead to systematic errors in observed values for binding affinities and rate constants. Ensuring the protein I used was as contaminant-free as possible required exhaustive testing and purification procedures were continually refined throughout my doctoral work. In this Chapter, I describe how I devised a high-yield protocol yielding pure CREB.

A remark about terminology used in this Chapter

Name	*Function*	*NTD site non-native sequence*	*CTD site non-native sequence*
TEV	Protease	-SDNAIA	EDLYFQ-
FXa	Protease	-	IEGR-
ULP-1	Protease	-GAGMS	N/A
Pro	CPA stop site		P
GGC	Forced dimerization		GGC

Table 3.1: List of non-native sequences used in purification. For cleavage sites ("Protease"), "-" indicates the protease cleavage site, following by the linker space before the protein sequence starts. See Chapter 6 for more information on the GGC extension function.

During the protocol optimisation process, I tested many CREB clones with various non-native sequences at both N and C-termini. To improve understanding of what each recombinant protein contains, tags (amongst HaloTag, GB1, GST, SUMO, 6His for 6x poly-histidine) and such sequences will be referred to when listing the protein. Non-native sequences are given an identifier in Table 3.1. For instance, TEV-CREB-6His will have the TEV non-native sequence at the N-terminus and the 6His tag at the C-terminus. Additionally, CREB isoform B has 3, and isoform A 4 cysteines which were mutated to serine to make pseudo wild-type (as per (107)), but some constructs will have one non-mutated cysteine for labelling (indicated by the "-1C" suffix) or one in the forced-dimerization extension ("-GGC" suffix, Table 3.1).

3.1. Inherited purification methods yield slightly contaminated CREB and could be improved

I inherited the progress previously achieved by Ngai Lam Chung on the purification and characterisation of full-length CREB (192). He had tested the method by Shnitkind *et al.* (26) but found it inappropriate for his HaloCREB-WT-6His construct (Figure 3.1), and instead designed his own method (Figures 3.1, 3.23). Briefly: CREB was expressed in *E.*

coli, pellets were resuspended in buffer with protease inhibitors and sonicated. The lysate was subjected to an Ni-NTA affinity chromatography step with an overnight salt wash and then a heparin chromatography step.

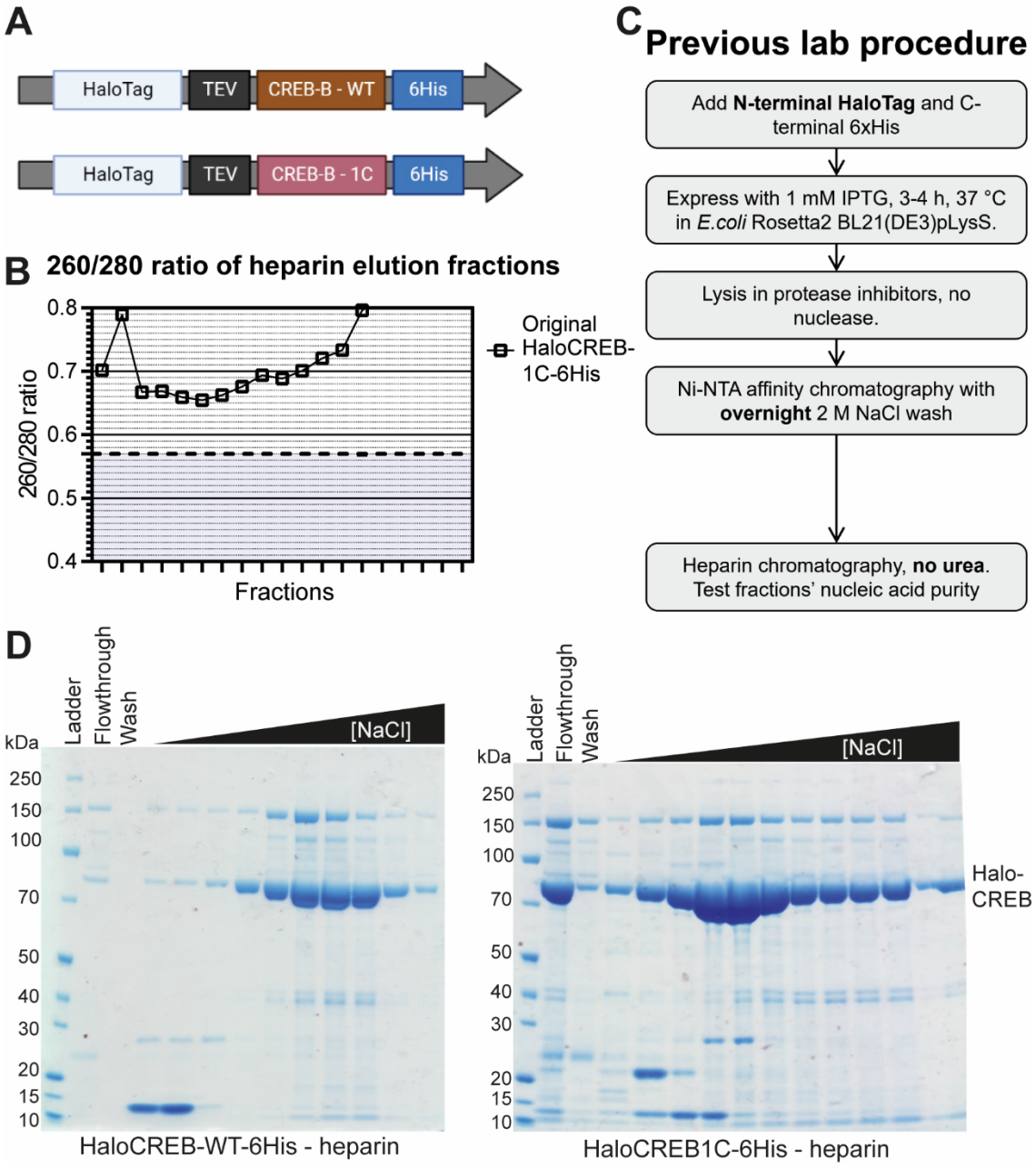


Figure 3.1: Initial expression tests using previously developed method. A. The two constructs used in the experiment, both are CREB isoform B, but the wild type (WT) has three cysteines and another only has one. **B.** A260/A280 of heparin chromatography fractions are all above the 0.57 "pure" ratio. **C.** Purification method overview. **D.** SDS-PAGE gels of heparin fractions show contamination for both constructs.

Unfortunately, I could not replicate Mr. Chung's success with this method - various HaloCREB constructs were both nucleic acid and protein contaminated to a substandard level for biophysical experiments (Figure 3.1). Although contamination was limited by careful fractionation, yields were greatly limited as a result. Hence, the purification method clearly needed further refinement to make CREB suitable for biophysical experiments.

3.2. Pure CREB is obtained with the intermediate GB1 method but suffers from low yields and aggregation

3.2.1. DTT addition and mutation of each cysteine to serine to make pseudo wild-type CREB greatly increases purity

Another published CREB purification procedure (193) (Figures 3.2, 3.23) is incompatible as it includes DNAses. Their inclusion risks contamination of the final product and the degradation of critical DNA in our nucleic acid-based assays. Yet, one notable feature of this procedure is usage of the reducing agent DTT. Furthermore, the replacement of cysteine by serine in CREB to create a pseudo-WT construct was shown to improve CREB behaviour at no apparent cost (107). However, such mutations seem to increase binding affinity to TRE (not true CRE) sites by 5-fold according to gel-shift assays (194). In contrast, the more sensitive fluorescence anisotropy technique indicates a less than two-fold change and a 5 nM upper estimate $K_{D,app}$ for the mutants binding to the canonical CRE sites (192).

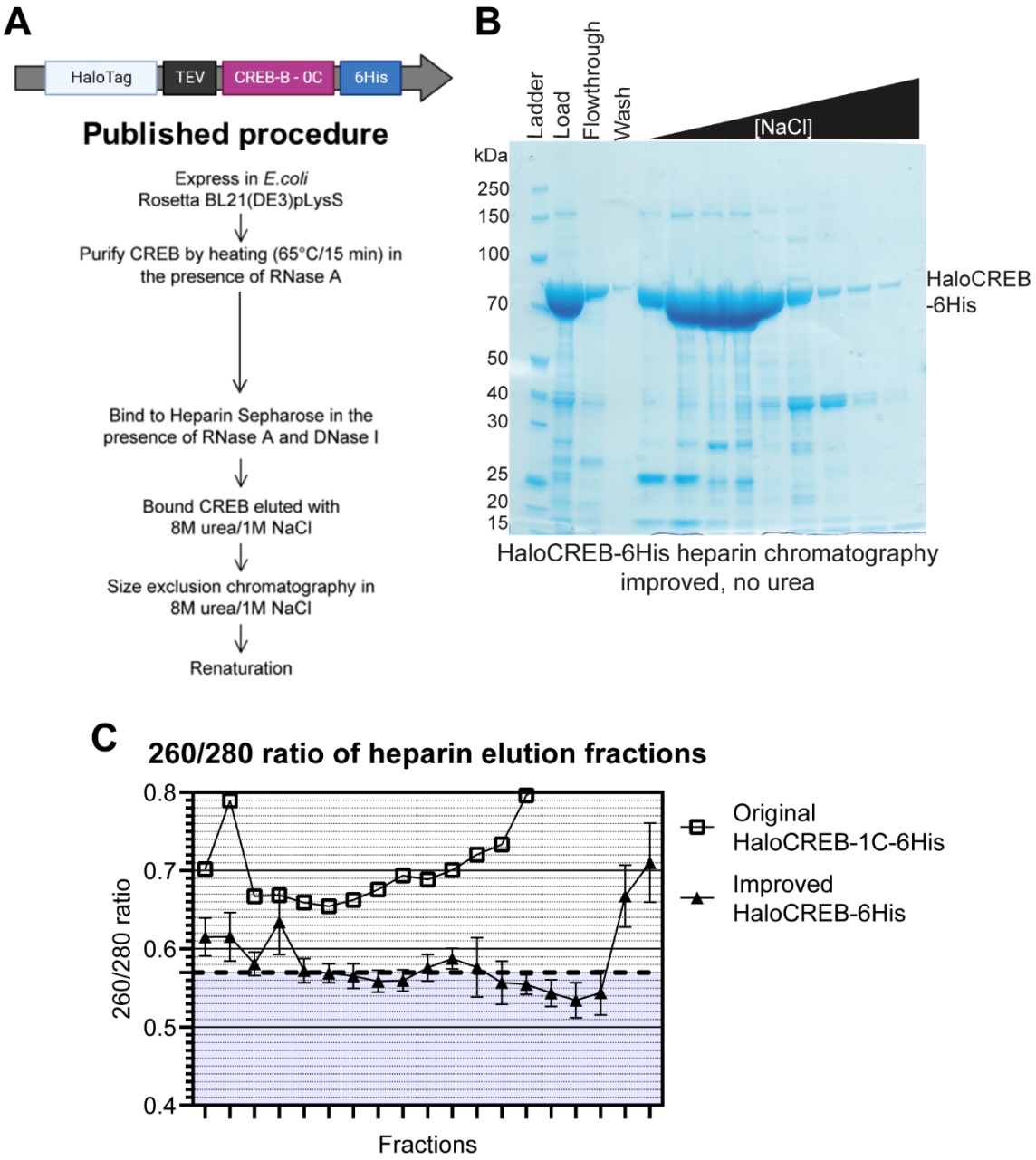


Figure 3.2: DTT addition and pseudo wild type purification. A. Overview of the published procedure in (193) and new CREB construct. **B.** A260/A280 ratios of heparin chromatography for the original and improved method. DTT and cysteine removal improve removal of nucleic acids. **C.** Example of heparin chromatography fractions on an SDS-PAGE gel. Lots of lower molecular weight contaminants remain.

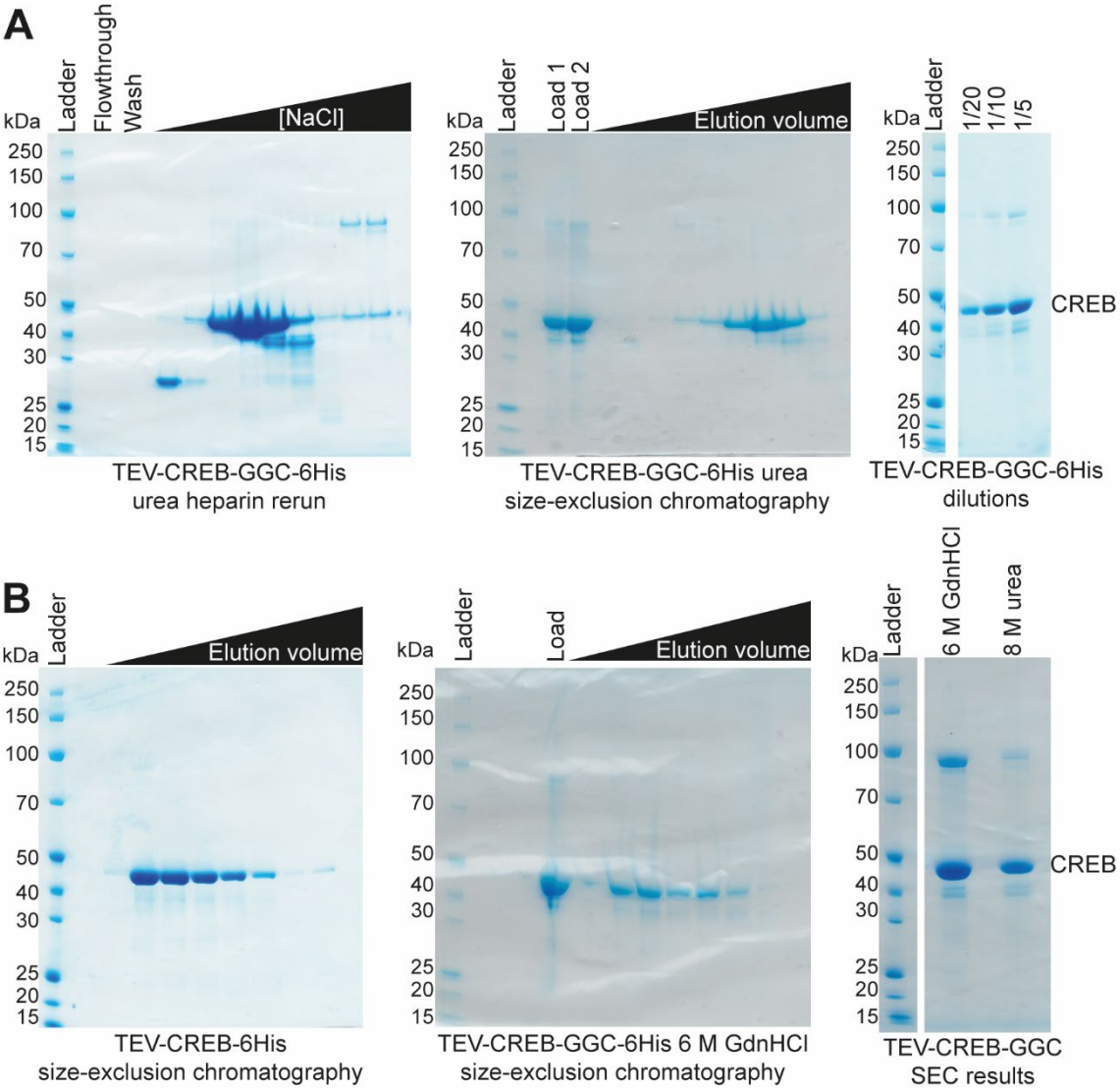


Figure 3.3: Addition of denaturant and size-exclusion chromatography (SEC) improves purity. A. Urea improves purity of heparin and size-exclusion chromatography fractions. Some minor contaminants remain, which make up roughly 1/20th of the final product. **B.** SEC is possible with and without denaturant but there is less protein in each fraction. Guanidine hydrochloride (GdnHcl) does not improve on urea and lower molecular weight contaminants are not resolved as well with increasing elution volume.

Using the published study as inspiration, I added DTT to all buffers and mutated all cysteine to serine. These changes had a positive impact on sample quality, lowering the A260/A280 ratios to quasi-pure levels and removing some protein contaminants (Figure 3.2). Unfortunately, more contaminants remained so further work was required.

3.2.2. Testing of denaturants and other chromatography methods

The published method also includes 8 M of urea denaturant during heparin chromatography and size-exclusion chromatography (SEC), therefore I tested whether 8 M urea could improve purity by dissolving CREB-contaminant complexes. As this is incompatible with the folded HaloTag, it is cleaved by TEV post Ni-NTA affinity chromatography. I found that adding 8 M urea to heparin chromatography and including an additional SEC step increased purity, but some stubborn contaminants remained (Figure 3.3). The addition of urea also improved A260/A280 ratios (Figure 3.18), likely because it dissolves aggregates that otherwise contribute to the 260 nm value. Only two minor lower molecular weight contaminants remained which sequencing mass spectrometry identified as CREB degradation products (Table 3.2) with intact C-terminal domains (Figure 3.4).

A previous DPhil student in the lab, Conor Kelly, had observed carbamylation of CREB bZIP by urea under some conditions, which might occur during size-exclusion chromatography as it takes several hours. I tested both 6 M guanidine hydrochloride (GdnHCl) and the absence of denaturants with worse results (Figure 3.3). SEC without denaturant is still possible and an option for denaturant incompatible tags but yields are reduced (not shown). I further trialled 6 M GdnHCl (to replace the 2 M NaCl salt wash) during Ni-NTA affinity chromatography, but there was no significant improvement in yield or purity (Figure 3.5). Given its additional expense and safety requirements this change was not implemented. I also tried anion-exchange chromatography to replace size-exclusion but didn’t improve yield or purity either (Figure 3.5).

To summarise, DTT, denaturants and size-exclusion each improved purity both nucleic-acid and protein-wise, with degradation products as the only minor contaminants.

					Contaminant #1				Contaminant #2			
Accession	# AAs	MW [kDa]	calc. pI	Description	Score A2	Coverage A2	# Peptides A2	# PSM A2	Score B2	Coverage B2	# Peptides B2	# PSM B2
Halo-CREB-6His	649	71.5	5.21	Halo-CREB-6His	269.57	20.03	14	73	383.20	18.03	14	102
CON_IPI00708398.1	607	69.2	6.18	SWISS-PROT:P02769 (Bos taurus) Bovine serum albumin precursor	100.04	29.82	15	32	86.86	32.95	16	28
CON_IPI00009866.6	458	49.6	4.96	SWISS-PROT:P13646-1 Tax_Id=9606 Gene_Symbol=KRT13 Isoform 1 of Keratin, type I cytoskeletal 13	5.47	5.02	2	2	74.87	38.21	12	25
CON_CASA1	199	23.0	4.94	SWISS-PROT:P02662 Alpha-S1-casein - Bos taurus (Bovine).	46.21	25.13	4	12	44.76	41.71	6	12
CON_IPI00290078.5	594	63.9	8.25	SWISS-PROT:P19013 Tax_Id=9606 Gene_Symbol=KRT4 keratin 4	2.76	1.85	1	1	65.77	30.30	13	21
CON_IPI00220327.3	644	66.0	8.12	SWISS-PROT:P04264 Tax_Id=9606 Gene_Symbol=KRT1 Keratin, type II cytoskeletal 1	37.00	14.13	8	10	37.87	18.94	9	12
CON_CASB	209	23.6	5.19	SWISS-PROT:P02666 Beta-casein - Bos taurus (Bovine).	26.69	32.54	4	8	23.55	32.54	4	7
CON_IPI00299145.9	564	60.0	8.00	SWISS-PROT:P48668 Tax_Id=9606 Gene_Symbol=KRT6C Keratin, type II cytoskeletal 6C	3.22	2.13	1	1	32.73	18.97	7	10
CON_IPI00009865.2	593	59.5	5.21	SWISS-PROT:P13645 Tax_Id=9606 Gene_Symbol=KRT10 Keratin, type I cytoskeletal 10	26.86	16.53	6	8	8.67	6.41	2	3
CON_IPI00021304.1	645	65.8	8.00	SWISS-PROT:P35908 Tax_Id=9606 Gene_Symbol=KRT2 Keratin, type II cytoskeletal 2 epidermal	26.77	20.00	7	7	8.50	5.58	3	3
CON_IPI00384444.5	472	51.6	5.16	SWISS-PROT:P02533 Tax_Id=9606 Gene_Symbol=KRT14 Keratin, type I cytoskeletal 14	6.28	6.57	2	2	19.19	6.57	2	7
CON_IPI00217963.3	473	51.2	5.05	SWISS-PROT:P08779 Tax_Id=9606 Gene_Symbol=KRT16 Keratin, type I cytoskeletal 16	2.64	2.54	1	1	24.87	6.55	2	8
CON_IPI00009867.3	590	62.3	7.74	SWISS-PROT:P13647 Tax_Id=9606 Gene_Symbol=KRT5 Keratin, type II cytoskeletal 5	6.24	4.58	2	2	14.53	7.29	3	5
CON_IPI00290857.2	629	64.5	6.48	SWISS-PROT:P12035 Tax_Id=9606 Gene_Symbol=KRT3 Keratin, type II cytoskeletal 3					23.97	14.79	5	6
CON_CASK	169	19.0	6.30	SWISS-PROT:P02668 Kappa-casein [Contains: Casoxin C; Casoxin 6; Casoxin A; Casoxin B; Casoplatelin	9.30	26.04	2	3	6.46	10.65	1	2
CON_IPI00698900.1	375	41.7	5.48	SWISS-PROT:P60712 (Bos taurus) Actin, cytoplasmic 1	3.06	4.80	1	1	9.99	9.87	2	3
CON_IPI00019359.3	623	62.1	5.30	SWISS-PROT:P35527 Tax_Id=9606 Gene_Symbol=KRT9 Keratin, type I cytoskeletal 9	16.73	12.68	4	4				
CON_CASA2	207	24.3	8.27	SWISS-PROT:P02663 Alpha-S2-casein [Contains: Casocidin-1 - Bos taurus (Bovine).	5.43	12.56	2	2	2.80	7.25	1	1
CON_TRY1_BOVIN	243	25.4	7.91	SWISS-PROT:P00760 Cationic trypsin precursor (EC 3.4.21.4) (Beta-trypsin) [Contains: Alpha-trypsin	8.56	8.23	1	2	4.56	8.23	1	1
CON_IPI00011694.1	247	26.5	6.51	SWISS-PROT:P07477 Tax_Id=9606 Gene_Symbol=PRSS1 Trypsin-1 precursor	3.65	8.10	1	1	4.35	8.10	1	1
CON_IPI00700865.3	451	50.1	5.06	ENSEMBL:ENSBTAP00000016242 (Bos taurus) similar to alpha-tubulin I isoform 1					5.70	7.76	2	2
CON_IPI00103481.3	511	55.8	6.89	SWISS-PROT:Q14CN4-1 Tax_Id=9606 Gene_Symbol=KRT72 Isoform 1 of Keratin, type II cytoskeletal 72					6.35	5.28	2	2
CON_GFP	249	28.1	6.52	SWISS-PROT:Q9U6Y5 Green fluorescent protein (GFP-Cter-HisTag)	3.90	6.83	1	1				
CON_IPI00713814.2	333	35.8	8.35	SWISS-PROT:P10096 (Bos taurus) Glyceraldehyde-3-phosphate dehydrogenase	2.56	4.50	1	1				
CON_IPI00710250.3	401	43.9	5.35	ENSEMBL:ENSBTAP00000024466 (Bos taurus) 44 kDa protein					2.87	3.99	1	1
CON_ALDOA_RABIT	360	38.8	8.09	SWISS-PROT:P00883 Rabbit fructose-bisphosphate aldolase A					2.96	3.89	1	1

Table 3.2: **Sequencing mass spectrometry s results shows contaminants are CREB degradation products.** HaloCREB-6His is the protein corresponding the most to the degradation products.

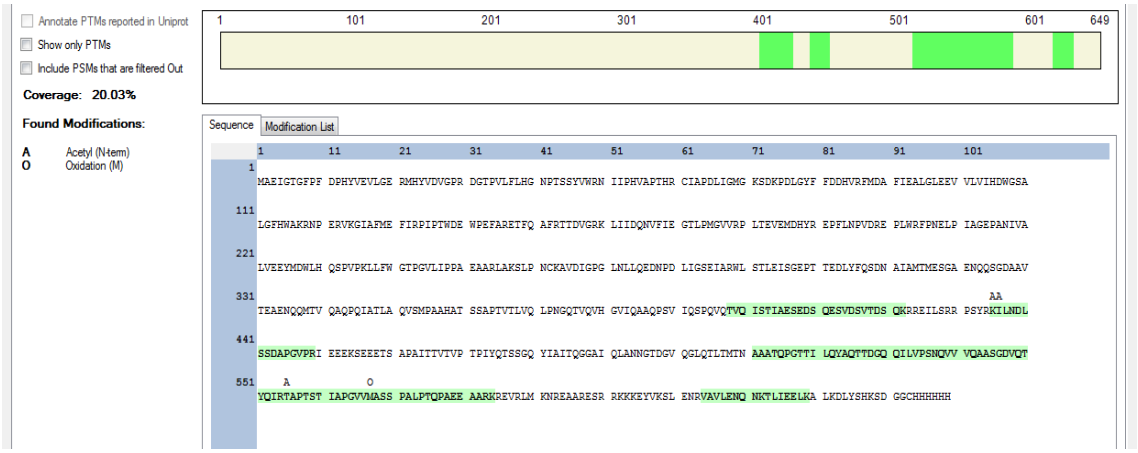


Figure 3.4: Sequencing mass spectrometry CREB coverage. Results coverage suggests CREB degradation products keep most of their C-termini.

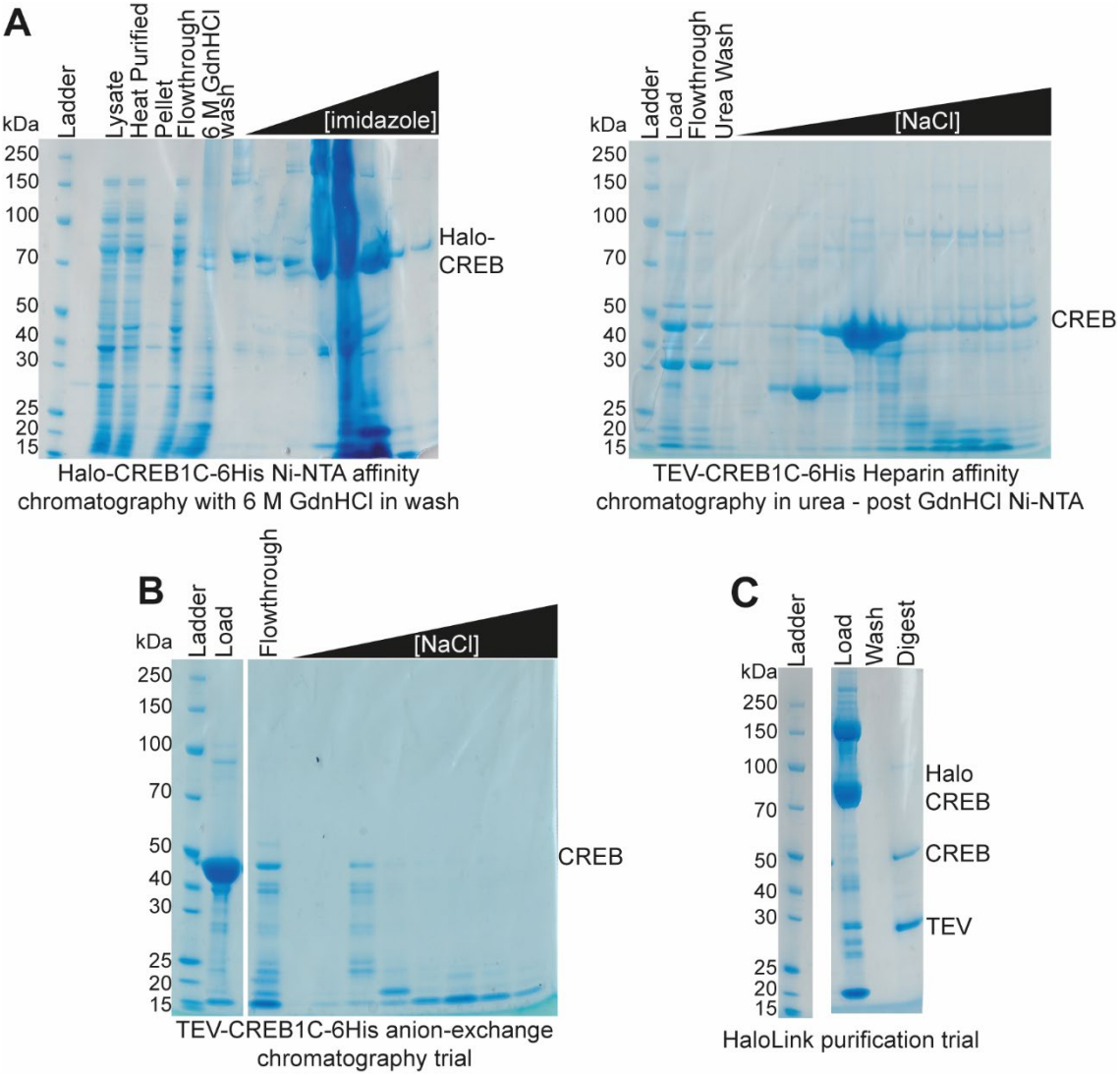


Figure 3.5: Denaturants, ion-exchange and HaloLink beads early purification tests. A. GdnHCl wash during Ni-NTA affinity chromatography does not improve purity. **B.** Anion-exchange chromatography doesn't remove contaminants if used after the "previous" purification protocol. **C.** HaloLink beads can isolate protein but are too low yield.

3.2.3. The HaloTag acts as a solubility tag which GB1 but not GST can replace

Initially intended as a dual affinity tag and labelling platform for cellular work done in the laboratory, the HaloTag is not needed for this project since cellular worked ceased. Worse, it may even hinder CREB activity, and targeting the HaloTag using HaloLink resin is too low yield to be suitable (Figure 3.5). Therefore, there did not appear to be a particular advantage to keeping it and I attempted to express CREB without it, but this reduced yields (Figure 3.6), suggesting that the HaloTag acts as a solubility tag. Instead, HaloTag was cleaved-off mid-purification with TEV protease, making TEV-CREB-6His (Table 3.1).

At this point, I tested if any contaminants that remained would affect concentration determination. I ran an excitation-emission scan to check for the noticeable presence of tryptophan since CREB only has the 5-fold less absorbent tyrosine, which all showed a single peak corresponding to tyrosine (Figure 3.7). Hence, if I had to work with contaminants, they would at least not perturb CREB concentration determination.

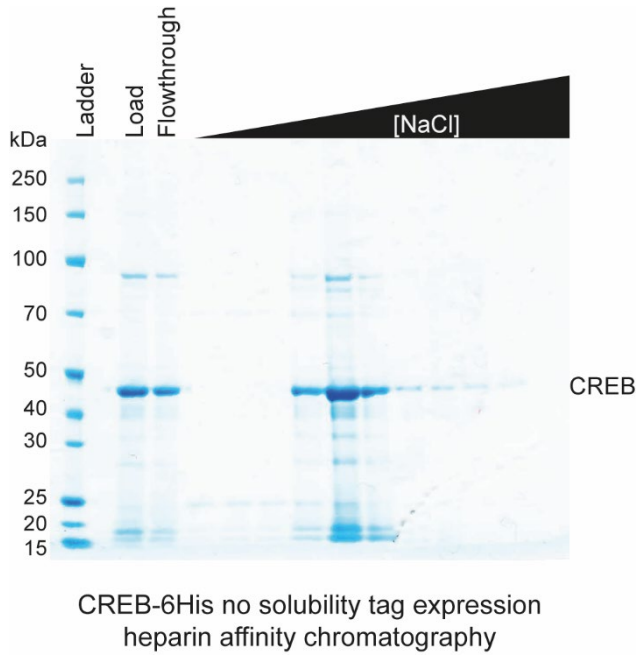


Figure 3.6: HaloTag likely acts as a solubility tag. Removal of the HaloTag greatly lowers soluble protein yields from bacterial expression, suggesting HaloTag acts as a solubility tag.

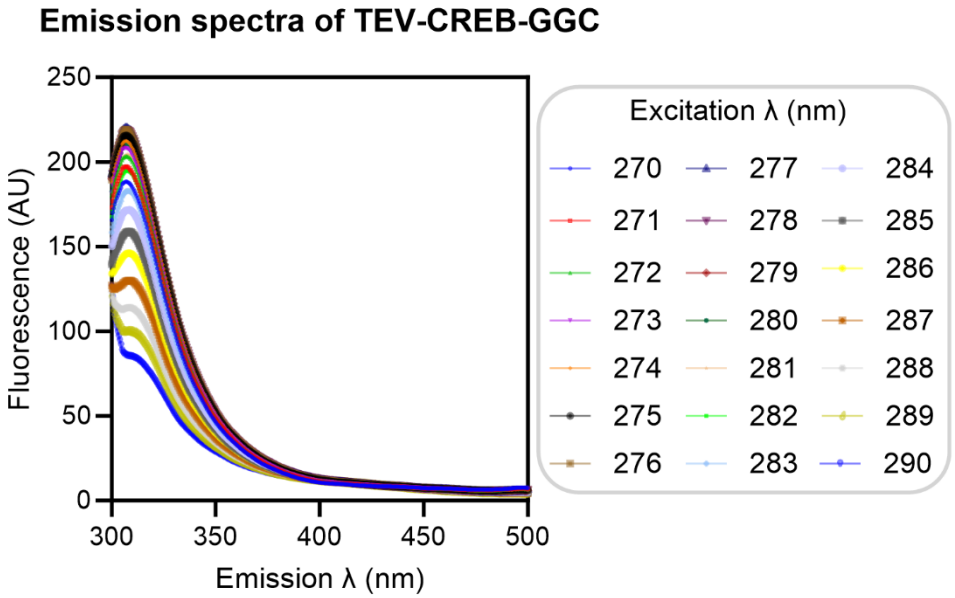


Figure 3.7: Excitation-emission spectra of purified CREB. A near 300 nm peak and a lack of a second peak shows only tyrosine is present at high enough concentration to affect absorbance reads. Contaminants do not have the higher absorbing tryptophan that could skew concentration reads.

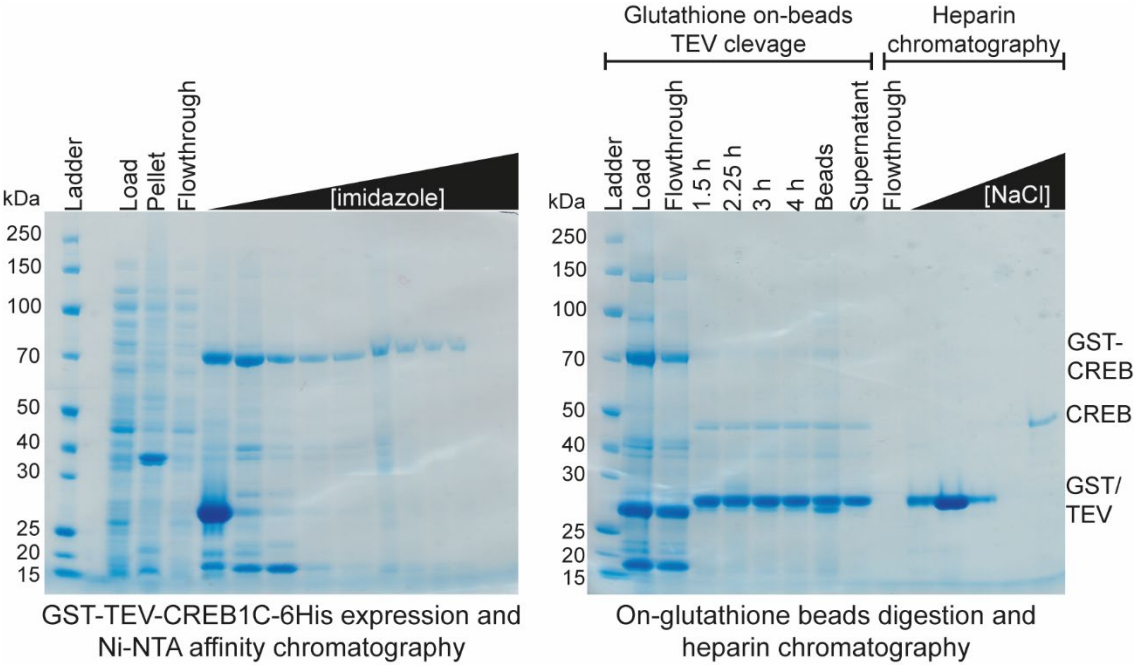


Figure 3.8: GST is not a suitable solubility tag for CREB. There are low yields for test GST-TEV-CREB1C-6His bacterial expression followed by Ni-NTA affinity chromatography (left), glutathione beads and heparin chromatography (right)

Replacement with an alternative tandem affinity-solubility tag, GST, was unsuccessful. Bacterial expression of GST-CREB rarely reached OD_{600} of induction (not shown) perhaps due to GST forming oligomers (195) with CREB, resulting in toxic aggregates. Plus, GST beads trial expression and purification were low yield (Figure 3.8), so this avenue appeared unproductive. However, the dual change from HaloTag to GB1 solubility tag and from a pET to a pRSET backbone was highly successful. It notably tripled the yield up to the heparin step and improved A260/A280 ratios (see later Figure 3.18), at no cost to purity (see later Figure 3.14), perhaps because the pRSET backbone has a higher copy number. In turn, the higher leaky expression affected bacterial growth during expression, although this was alleviated by the addition of 1% glucose to bacterial pre-cultures. Hence, although the HaloTag-pET combination can be used to express CREB, the GB1:pRSET combination has higher yields with no cost to purity.

3.2.4. GB1-CREB can be heat purified

Previous methods suggest heating bacterial lysates as an initial purification step to 65 °C (193). This appears to have little to no effect on the end product since CREB recovers well from heat treatments from circular dichroism data (107). This is compatible with GB1 as its T_m is around 80 °C (196) but not HaloTag due to its T_m of 50-60 °C (197). However, this means heat treatment can denature the HaloTag (Figure 3.9) and TEV protease (not shown), allowing removal by centrifugation. This does not improve purity of heparin chromatography eluates (Figure 3.9). Conversely, bacterial lysate pellet size increases if heat-treated at 65-80 °C for at least 30 minutes before centrifugation, likely due to protein aggregation. This is slightly perceptible on SDS-PAGE gels, as heat-treated pellets tend to have more protein than non-heat-treated ones (Figure 3.9). Therefore, the removal of the HaloTag allows heat treatment as a quick and useful first step to clean-up the lysate before immediate loading onto HisTrap columns for Ni-NTA affinity chromatography.

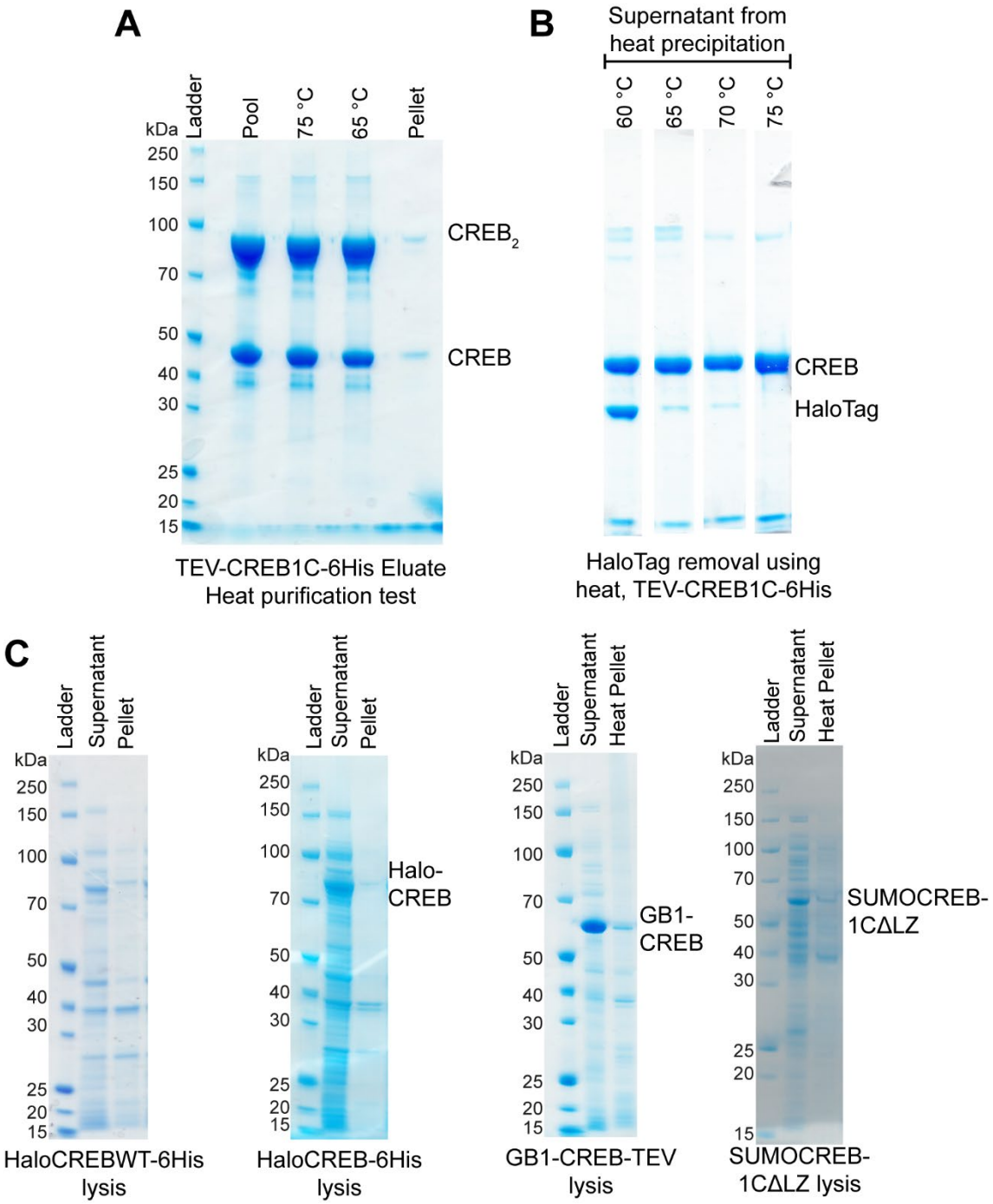


Figure 3.9: Heat can be used as a first purification step. A. Heat purification doesn't improve purity of samples having already undergone the original purification procedure. **B.** Heat can separate HaloTag from CREB above it's T_m. **C.** Differences between non-heat-treated lysis pellets (both left-hand side gels) and heat-treated ones (both right-hand side gels) show heat treatment causes a decrease in supernatant load and increase in pellet load by observing a change in the background blue dye shade.

3.2.5. Polyethyleneimine does not improve purification.

Above a certain concentration, polyethyleneimine (PEI) polymers can be used to selectively precipitate proteins and nucleic acids based on salt concentration (198). Only nucleic acids are precipitated by PEI in 1 M NaCl. Ammonium sulphate precipitation is recommended to subsequently remove PEI to prevent interference with downstream applications. Regrettably, tests proved inconclusive as there was never a clear combination to reliably precipitate CREB even at high PEI concentrations (0.5% (w/v) being already enough to precipitate most proteins (198)) (Figure 3.10). When PEI was used before Ni-NTA affinity chromatography, all protein eluted in the flowthrough (Figure 3.10). Lots of PEI and time-intensive centrifuge cycles are therefore needed with little upside, since GB1-expressed CREB tend to have less nucleic acid contamination already (Figure 3.18). Therefore, PEI steps were not incorporated.

3.2.6. Moving the 6His tag to the N-terminus.

C-terminal 6x His may affect CREB since it is close to the dimerization and DNA-binding domains, and it is likely to impede downstream processes (see more in Chapters 4, 6). Initially, I moved the tag to the N-terminus to be jointly cleaved with the HaloTag. However, this created too much contamination from degradation products (Figure 3.11). I therefore considered alternative purification approaches for cleavage.

C-terminal cleavage was then attempted with carboxypeptidase A (CPA) protease. CPA digests proteins at their C-terminus and is usually stopped by a proline (199). I purified CREB with a proline before the 6His tag (CREB-Pro-6His) using the “previous” method Figure 3.23). Digestion trials with carboxypeptidase A were unsuccessful and showed general degradation of protein, regardless of time or amount of enzyme used (Figure 3.12).

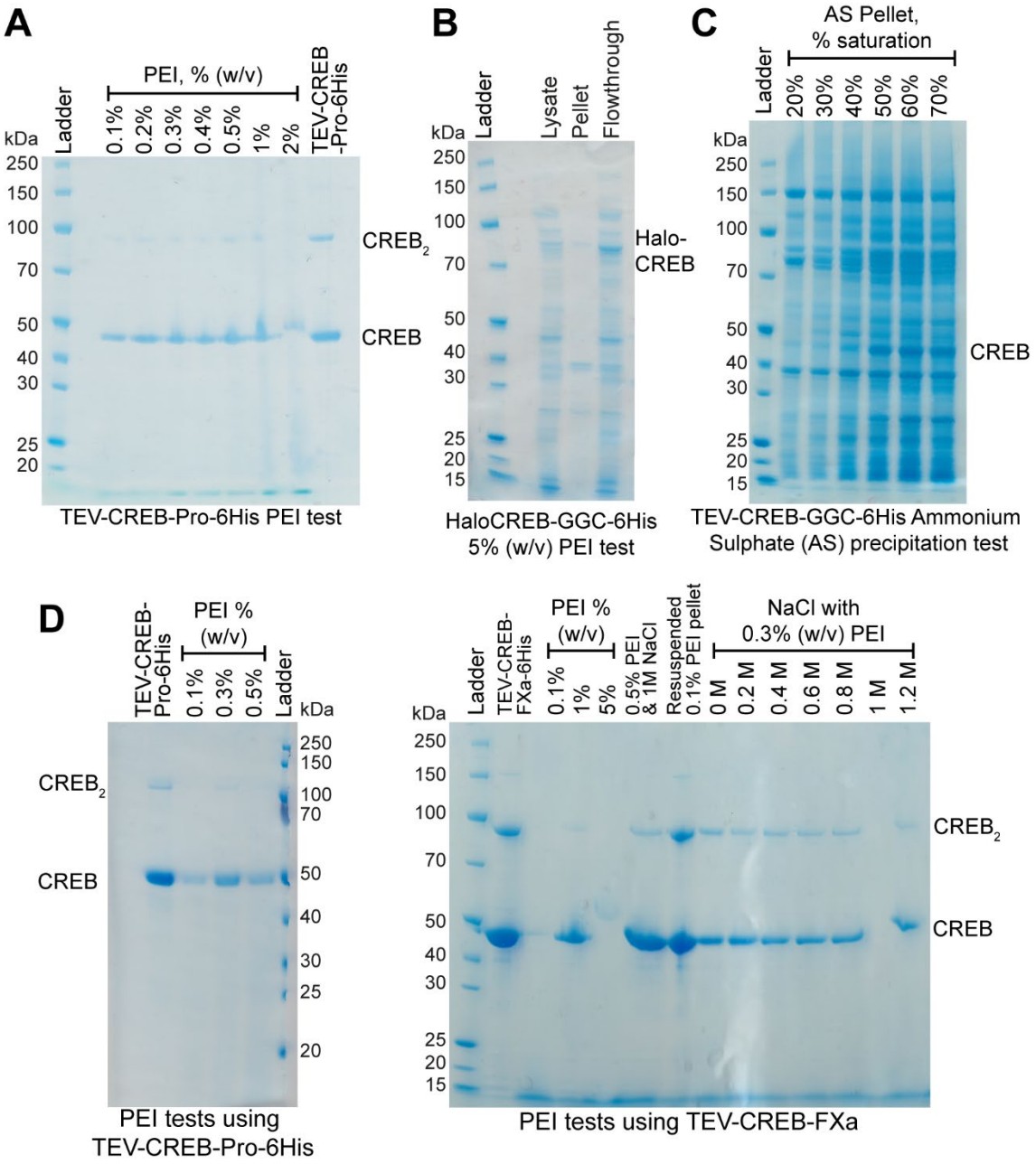


Figure 3.10: Polyethyleneimine (PEI) and ammonium sulphate (AS) tests to precipitate CREB. A. PEI slightly precipitates CREB at low concentrations and appears to fully precipitate it at 2%. **B.** PEI precipitation of nucleic acid at 1 M NaCl without AS treatment made CREB elute in the flowthrough. **C.** Effect of different AS saturation on CREB solubility, with resuspended pellets run on the gels. AS precipitates CREB at around 50% saturation. **D.** PEI can partially precipitate one CREB mutant around 0.1-0.5% w/v (left), and 0.1% for another mutant (right) as CREB is present in an 0.1% PEI pellet and not in the supernatant, although this may be unstable since 1% PEI appears to not fully precipitate TEV-CREB-FXa. 1 M NaCl keeps CREB in the soluble fraction at 0.5% PEI. However, 0.3% PEI looks unable to fully precipitate CREB, and salt has little effect up until 1-1.2 M. PEI is, overall, quite unpredictable for CREB.

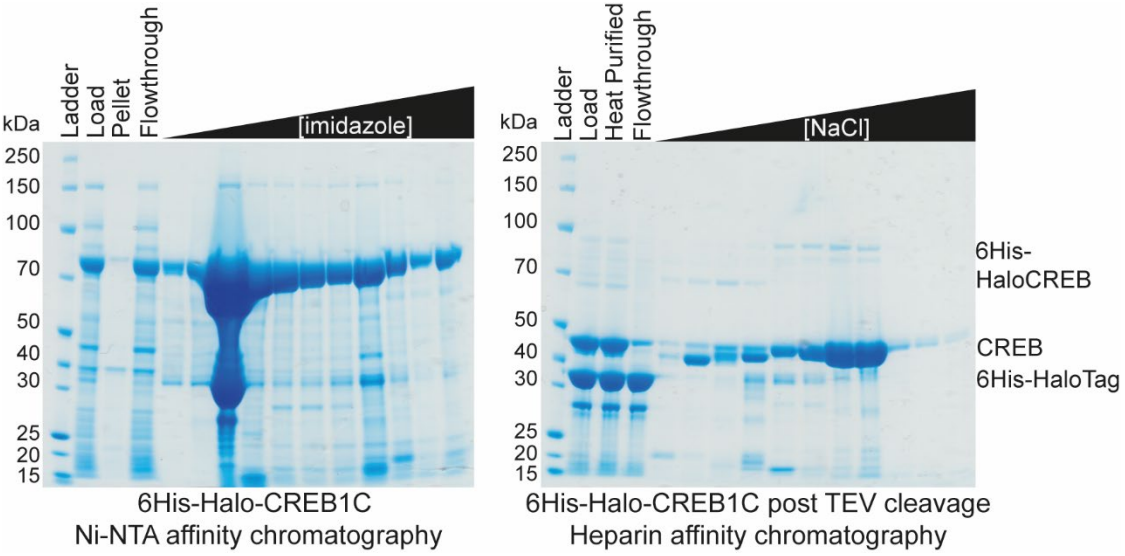


Figure 3.11: Degradation products are co-purified with 6His-HaloCREB. Both in Ni-NTA (left) and heparin (right) affinity chromatography, degradation products around CREB are co-purified.

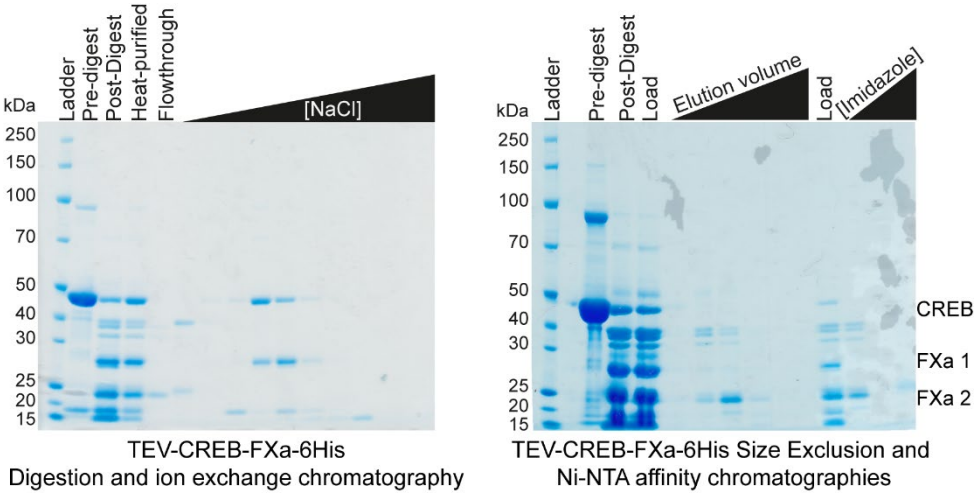


Figure 3.12: FXa is not a suitable protease to remove the C-terminal 6His. Regardless of chromatography technique used (left, ion exchange; right, SEC and Ni-NTA affinity), FXa cannot be separated from CREB, and overnight cleavage causes degradation, making it unsuitable for use.

An alternative is Factor Xa, a dimeric protease with a shorter 4 residue leftover if cleaving at the protein's C-terminus (Table 3.1). While overnight cleavage of CREB with a FXa cleavage site before the 6His tag, purified with the "previous" method (Figure 3.23), was successful, it was impossible to separate Factor Xa from CREB, regardless of the method used (Figure 3.13). Furthermore, some CREB appears degraded during the reaction.

Henceforth, I decided to use TEV since it is a reliable, already used and easily produced protease. I engineered an additional TEV cleavage site to simultaneously cleave

off the N-terminal solubility and C-terminal 6His tags to make "TEV-CREB-TEV" with leftover residues described in Table 3.1. Despite the 5-residue non-native sequence at TEV-CREB-TEV's C-terminus (one less than the 6His tag), that it wasn't a histidine repeat with a pK_a close to experimental pH is likely to lower potential interference in biophysical assays.

In order to compare the impact of the his-tag and TEV non-native extensions, I attempted to express CREB with a N-terminal 6His again, expecting a repeat of the earlier

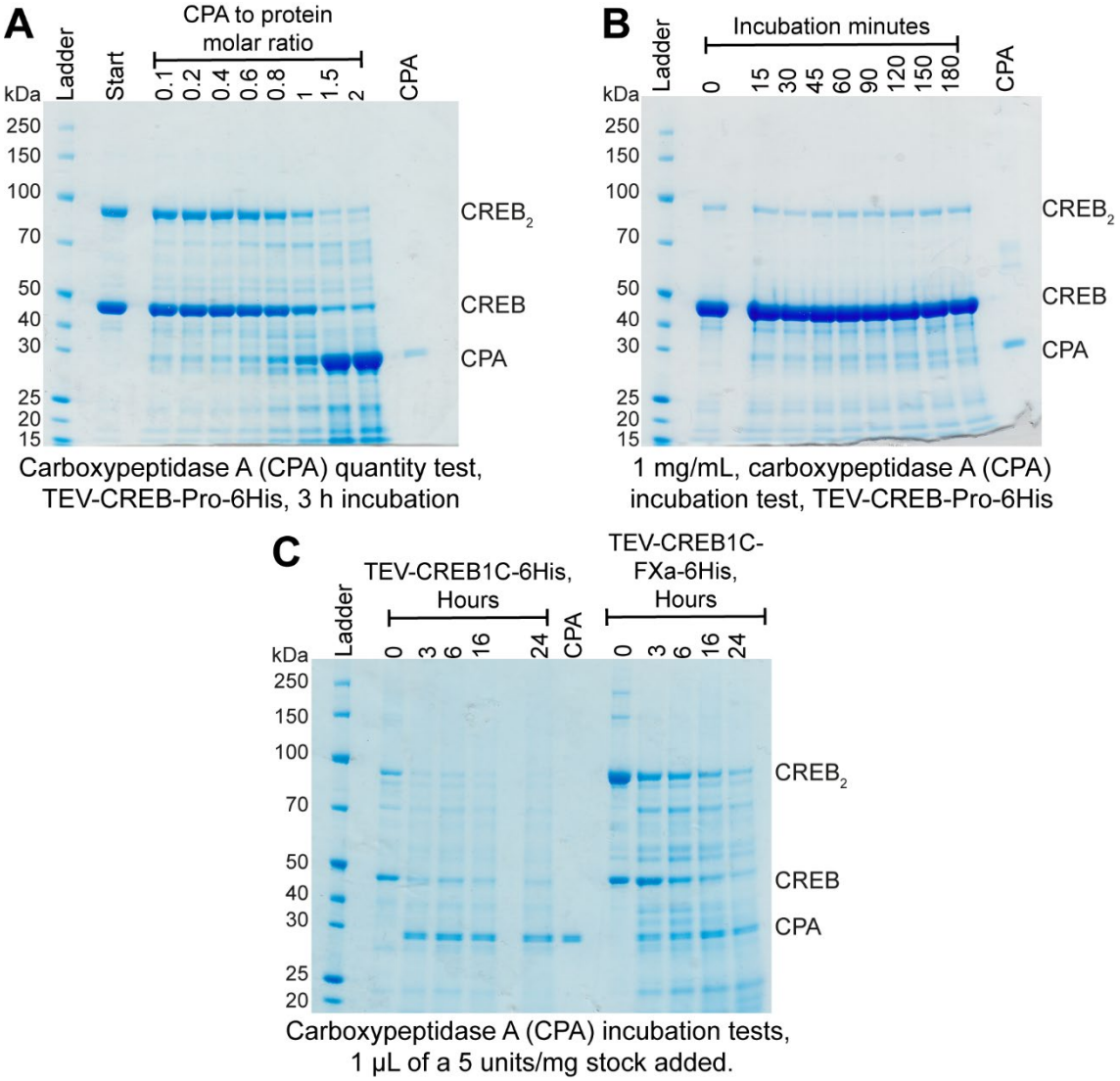


Figure 3.13: Carboxypeptidase A (CPA) cannot reliably cleave C-terminal 6His. A. Even at low CPA concentrations, non-specific CREB degradation is observed despite the presence of proline (pro). **B.** CPA quickly generates degraded CREB and low-molecular weight products, suggesting non-specific cleavage of the 6His tag. **C.** Similar consequences are observed with other CREB mutants. Therefore, CPA is ill-suited for C-terminal 6His removal.

test and thus a very low-yield eluate with lots of degradation product. Unexpectedly, with all of the above modifications to the protocol (DTT, urea, size-exclusion, GB1), the presence of degradation product was minimal, and CREB product was virtually contaminant-free (Figures 3.14, 3.18). It only has a TEV N-terminal non-native sequence, so it is referred to as "TEV-CREB". The pure product is indeed CREB as checked by sequencing mass-spectrometry (Table 3.3).

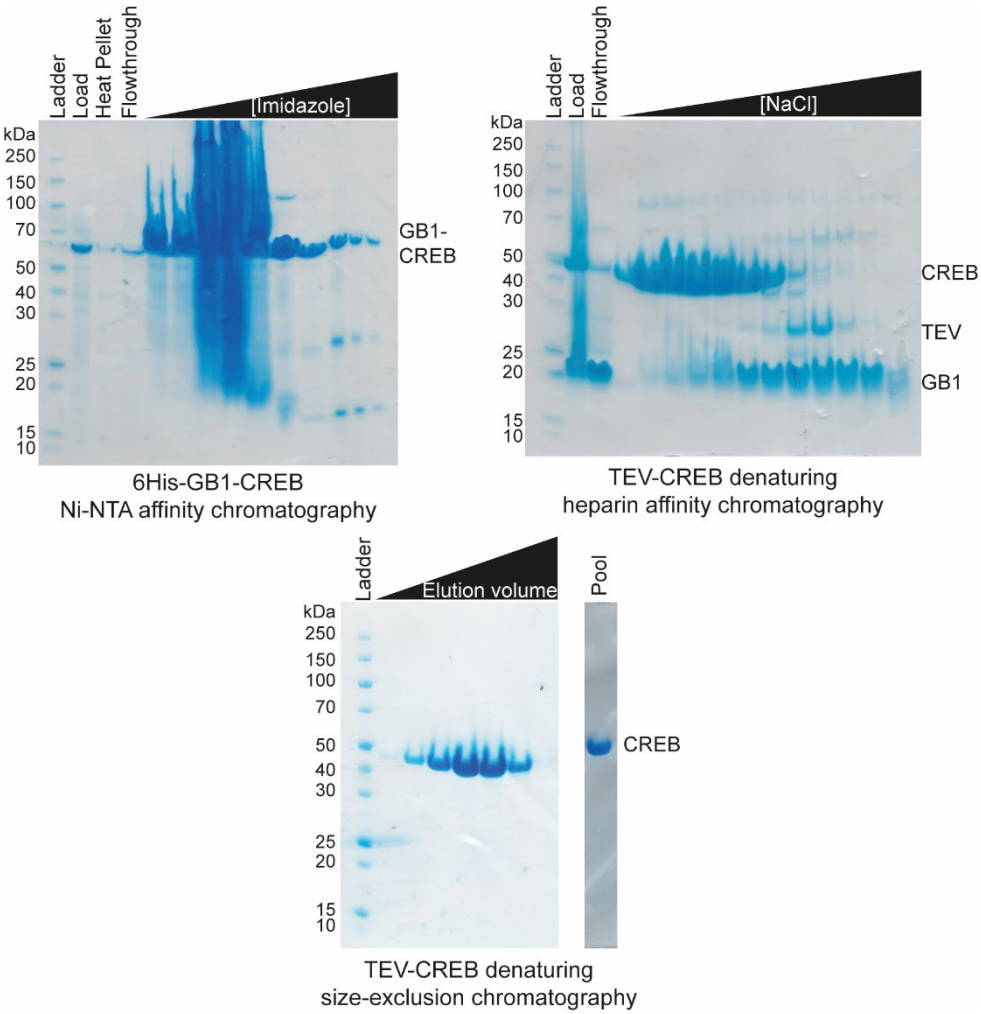


Figure 3.14: Purification to apparent homogeneity of bacterially expressed 6His-GB1-CREB. After heat purification, Ni-NTA affinity chromatography (top left), 6His-GB1 cleavage by TEV followed by denaturing heparin chromatography (top right) and denaturing SEC (bottom) with DTT present throughout, TEV-CREB (pWT) is free of protein contaminants.

Accession	TEV-CREB	CON_Chymotrypsin	CON_IPI00689304.2	A0A140NGU5	A0A140NEQ7
Description	CREB-B with TEV digestion overhang	SWISS-PROT:P00766 Chymotrypsinogen A - Bos taurus (Bovine).	TREMBL:Q2KIT0 (Bos taurus) Similar to collagen, type X, alpha 1	Acetyl-coenzyme A synthetase OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=acs PE=3 SV=1 - [A0A140NGU5_ECOBD]	Apolipoprotein N-acyltransferase OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=lnt PE=3 SV=1 - [A0A140NEQ7_ECOBD]
Score	166.87	3.23	0.00	0.00	0.00
Coverage	58.86	5.71	22.51	6.13	7.62
# proteins	1	1	1	1	1
# Unique Peptides	9	1	1	1	1
# Peptides	9	1	1	1	1
# PSMs	73	1	3	3	1
Area	9.733E9	7.384E7	4.347E5		9.062E8
# Aas	333	245	191	652	512
MW [kDa]	35.6	25.6	20.6	72.0	57.0
calc. pI	5.19	8.16	8.59	5.80	9.48

Table 3.3: Sequencing mass spectrometry shows CREB is obtained by the GB1 method. TEV-CREB (red column) is the only protein identified with a 50-fold higher score than the second hit, an enzyme used in the digestion process. CON = contaminant introduced by the procedure, PSM = peptide spectrum match.

3.2.7. The GB1 method still suffers from low yield and product aggregation.

In absolute terms, the yield of the GB1 method, although improved, is still half that of the published method (Figure 3.18). Size-exclusion chromatography causes reduced yields (Figure 3.18), perhaps because CREB interacts with the Superdex 200 resin or aggregates don't enter the column, so its replacement is desired. Moreover, TEV-CREB is prone to aggregation which, surprisingly, isn't the case for TEV-CREB-TEV (Figure 3.15).

Aggregation is especially troublesome since, due to Rayleigh scattering, it affects the A260/A280 purity assessment ratio and A280 used for concentration determinations. The latter caused problems with reproducibility in our biophysical experiments (Figure 3.16). The increase in A260/A280 isn't caused by nucleic acid, as overnight nucleic acid digestion with Benzonase doesn't affect the A260/A280 ratio (Figure 3.15).

It is possible to limit the presence of aggregates by filtering and CREB domain deletion mutants did precipitate less. Regardless, there were still cases where correction of Rayleigh scattering was required using published methods (167) or by working out the log-log behaviour of absorbance vs wavelength (λ). Since the effect is theoretically proportional to λ^{-4}, we can fit a log-log curve using a linear regression at wavelengths which CREB doesn't absorb, then using the fit to extrapolate the expected contribution of scattering at a desired wavelength (Figure 3.15). Using corrections does lowers the A260/A280 ratio to nucleic-acid free values (Figure 3.15), showing once-more that higher A260/A280 ratios are not due to nucleic acid, but to aggregation. The downside of this approach is that lots of corrections are needed to find a true value and it can "overcorrect", sometimes giving negative absorbance (Figure 3.15) and consequently underestimated extinction coefficients obtained by using the Gill, von Hippel method (171).

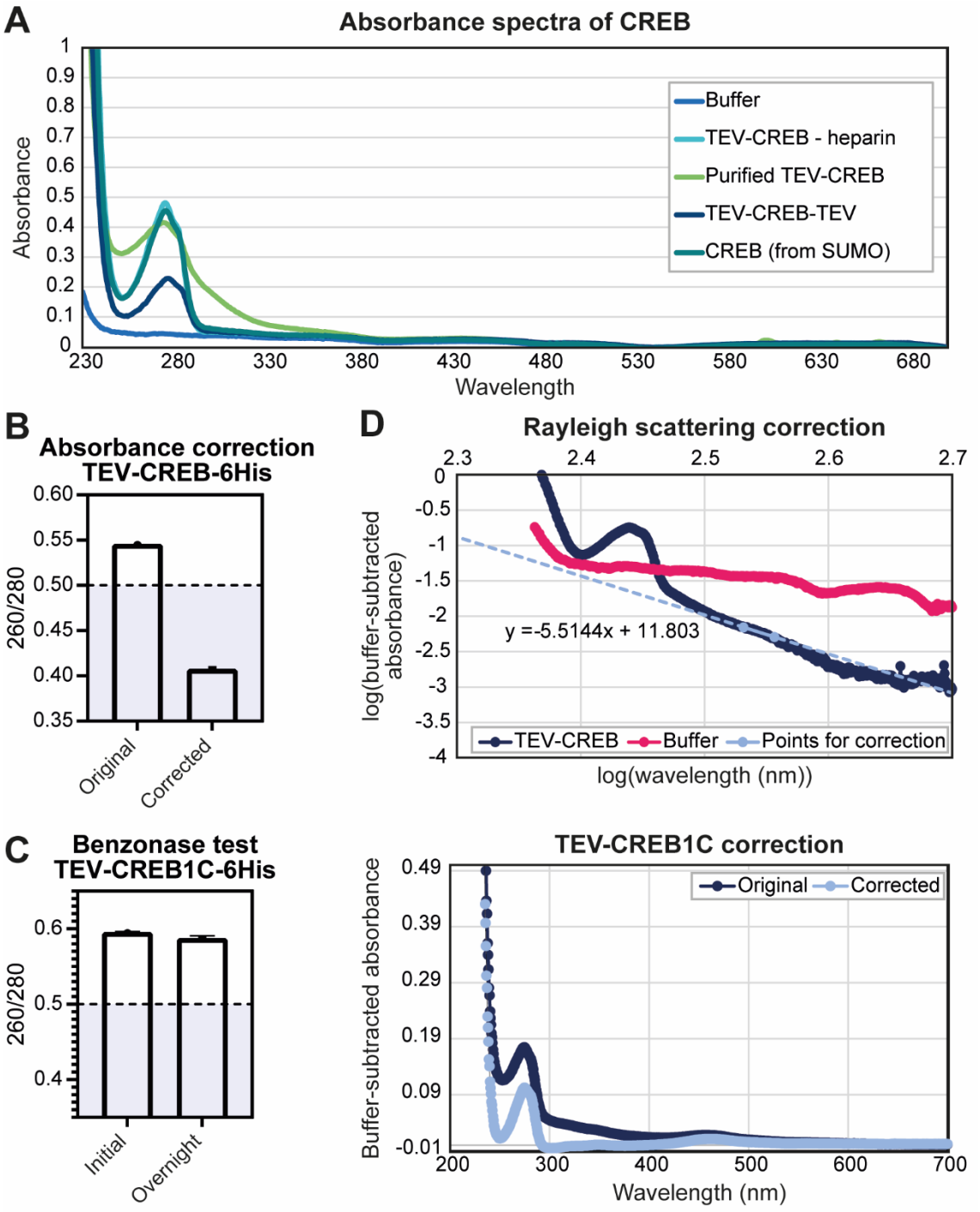


Figure 3.15: Investigation of CREB aggregation. A. Absorbance spectra of purified recombinant CREB. Only post-purification TEV-CREB shows signs of aggregation (rise at ~350 nm of the purified product but not of the denatured heparin eluate). **B.** A260/A280 values improve with regards to nucleic acid contamination after correcting for Rayleigh scattering using the method described in (167). **C.** Overnight incubation with Benzonase® nuclease doesn't improve A260/A280 ratio, suggesting the high A260/A280 ratio is caused by aggregation, not nucleic acid contamination. **D.** Top: example of log-log plot to show effect of Rayleigh scattering on a sample, with correction points used for correction by (167). Bottom: Rayleigh scattering correction can overcorrect and give some values in the negative.

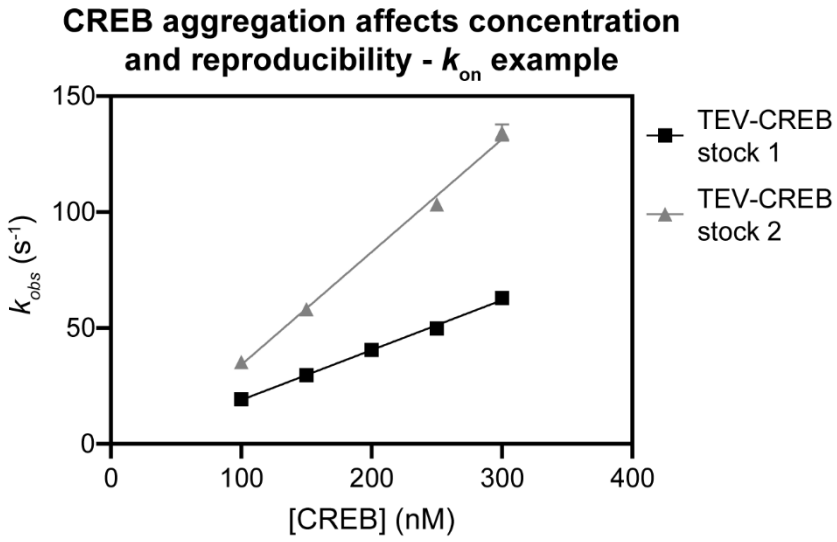


Figure 3.16: TEV-CREB aggregates affect concentration determination, causing issues with experimental reproducibility. Here, there are large differences in association rate constants to CREh, despite using the same TEV-CREB stock (see Chapter 4 for association kinetics analysis). 25 °C, pH 7.4.

An alternative is to use Bradford assays, comparing absorbance to a known CREB standard whose concentration was accurately determined by amino acid analysis off-site. While the high wavelengths and diluted samples used in Bradford assays limit the influence of Rayleigh scattering, the assay can be unreliable (200) and would require amino acid analysis for different mutants, especially since CREB's response to the Bradford reagent is different than albumin's (Figure 3.17).

In summary, the GB1 method greatly improved on the inherited procedure to make pure CREB, free of both nucleic acids (Figure 3.18) and protein contaminants (Figure 3.14), at a higher yield (Figure 3.18). I achieved this by changing plasmid backbone and solubility tag, moving the 6His tag, incorporating heat purification, size exclusion, and the presence of denaturants. Yet, this method still has two important issues, with overall low yields and product aggregation affecting concentration reads. Nevertheless, multiple methods can correct for the influence of aggregates and can be combined to give concentration estimates.

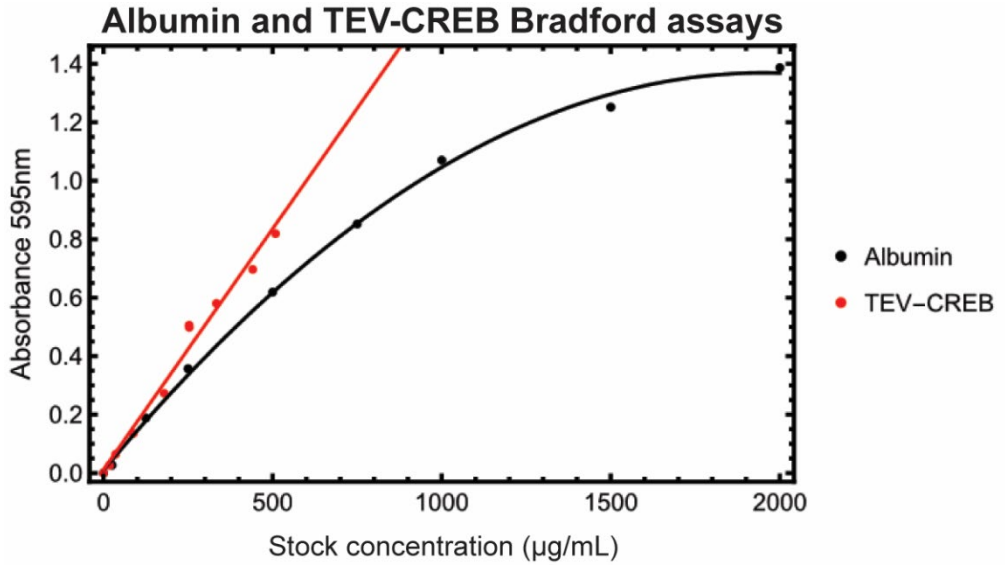


Figure 3.17: Example of concentration determination using Bradford assays with albumin and TEV-CREB standards. Albumin is fitted by a quadratic equation, whereas TEV-CREB by a linear one.

3.3. The SUMO method improves on yield and aggregation issues from the GB1 method

3.3.1. GB1 to SUMO solubility tag switch and auto-induction of protein expression solves issues with the GB1 method.

Although it is possible to correct somewhat for aggregation in terms of concentration and purity determinations, it is ideal to remove or avoid aggregates in the first place. Aggregation might reduce yields, and ultimately result in non-homogenous preparations for biophysical experiments. I trialled one further method for producing full-length CREB, replacing the GB1 tag with SUMO tag (here, *S. cerevisiae* SMT3 protein from the SUMO family), cleavable with homemade ULP-1 protease. I also switched IPTG-induced expression to overnight auto-induction upon suggestions from my peers who had found success with it. I supplied 2xTY media with glucose, glycerol and lactose concentrations as per the ZYM-5052 convention for autoinduction (201).

This method produced more bacteria per litre compared to IPTG induction. While increasing yields (Figure 3.18), it required more resuspension buffer be used and more thorough sonication of cells by splitting resuspended cells into smaller batches, otherwise causing clogged columns or filters when filtering lysate supernatant. As human SUMO-1's T_m is 65 °C (202) and the T_m predictor SCooP (203) suggest SMT3's is in the same range (Figure 3.19), it is possible to heat lysates up to this temperature in this method too, but I have used 80 °C treatment in my protocol without issues. The higher pH in resuspension buffers or possible thermo-recoverability of the tag may explain why SUMO-CREB doesn't precipitate when heat-treated above the tag's T_m. Some degradation products and standalone SUMO were formed but these are removed by the purification procedure (Figure 3.20).

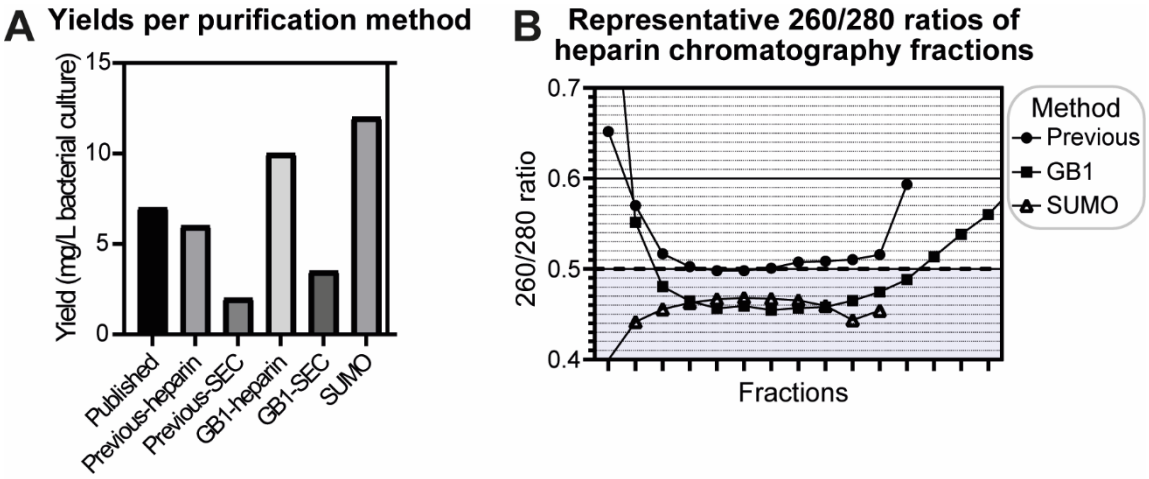


Figure 3.18: Yields and nucleic acid purity comparison of the different CREB purification methods. **A.** Yields from the published method were only reproducible if SEC was discarded, and the SUMO method has the greatest yield, near doubling that of the published method and five-fold more than the GB1 method post SEC. **B.** All methods have little to no nucleic acid contaminants post denaturing heparin chromatography, but GB1 and SUMO methods are purer than the previous method.

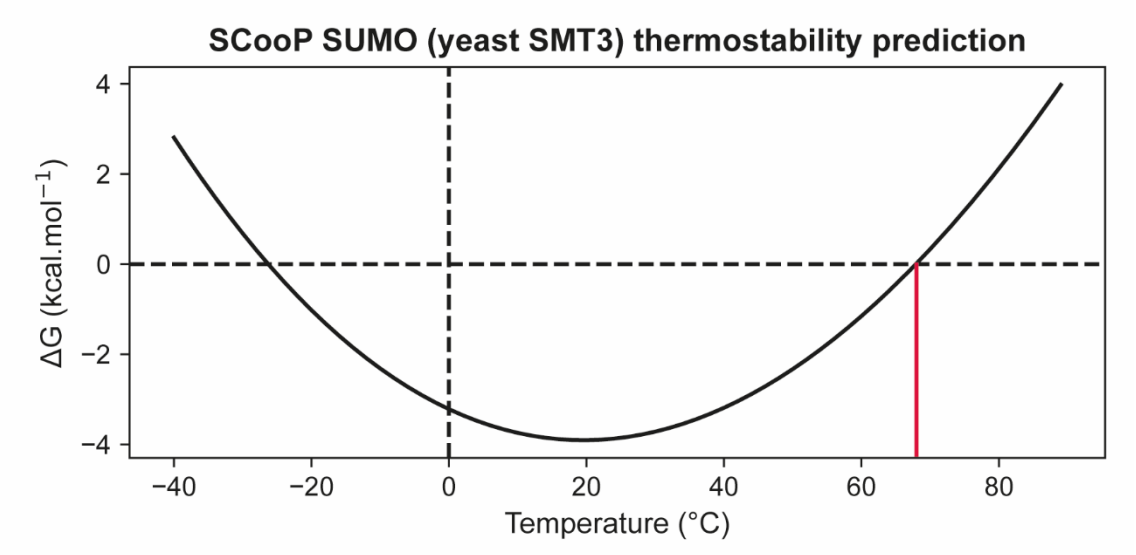


Figure 3.19: SCooP thermostability prediction of SUMO homolog yeast SMT3. SUMO *S. cerevisiae* homolog SMT3's T_m is predicted at 65-70 °C (red line) using the SCooP algorithm.

Two induction temperatures were tested, 37 °C and 18 °C, and the 18 °C product was less contaminated without reduced yield (Figure 3.20). Importantly, the product was virtually contaminant-free post heparin chromatography (Figure 3.20), so size-exclusion chromatography was eliminated. Combined with more bacteria grown per litre, yields quintupled as a result, despite SUMO-CREB being expressed from a lower copy-number pET backbone (Figure 3.18). Purity and lack of carbamylation were confirmed by LC-MS mass spectrometry (Figure 3.21). Importantly, CREB from this procedure doesn't aggregate as much as TEV-CREB (Figure 3.15), and therefore concentration estimates are much more reliable by using absorbance without much deviation due to Rayleigh scattering.

3.3.2. Extra chromatography methods tested with SUMO-CREB

Just like TEV-CREB from the GB-1 method, CREB denaturing size-exclusion chromatography is possible if needed (Figure 3.22). A difference from the previous methods is that native anion-exchange is now a viable separation method too (Figure 3.22), especially since the SUMO protease ULP-1, SUMO and CREB have different pIs (7.2, 4.2, 5.6

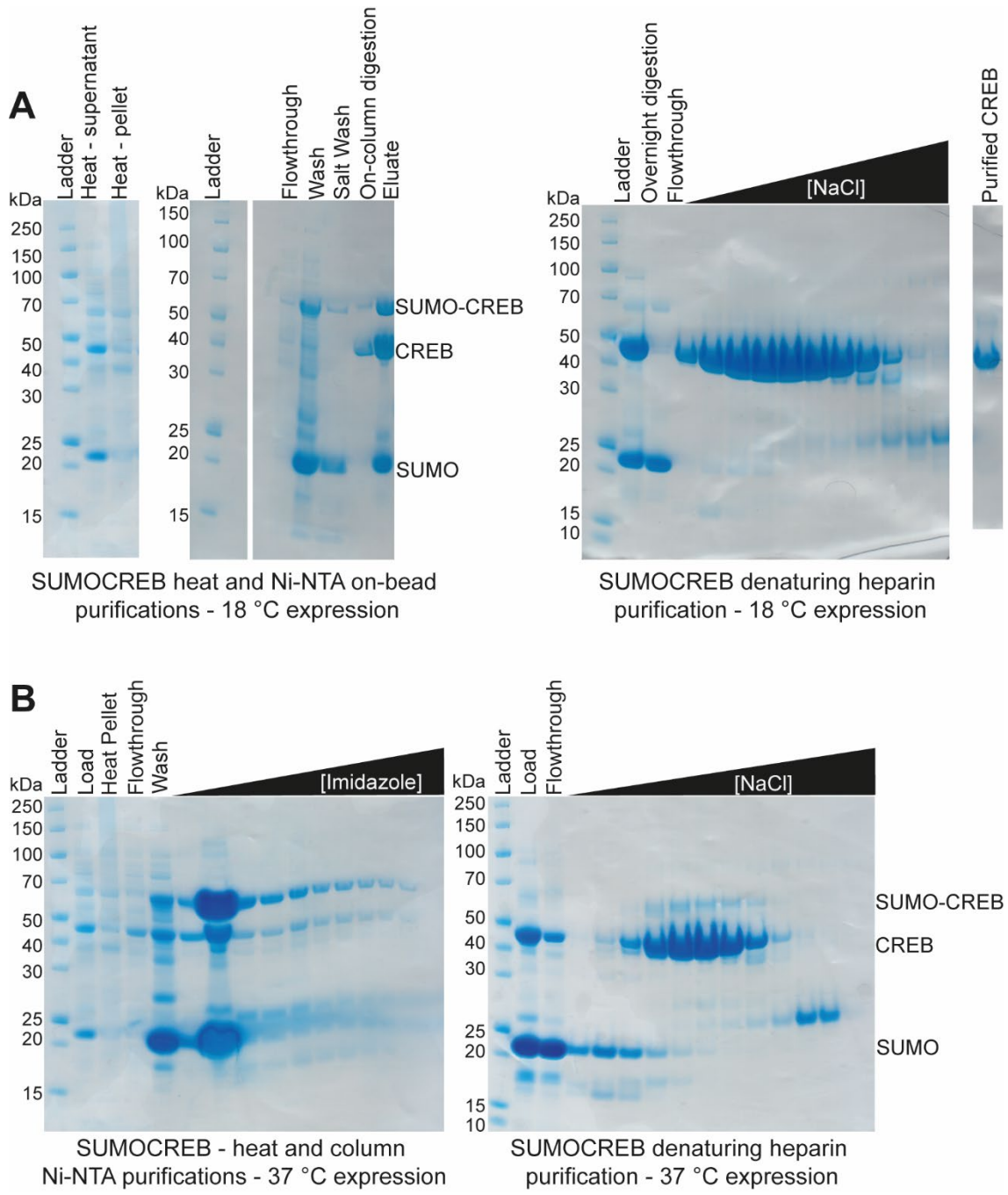


Figure 3.20: CREB is purified to apparent homogeneity using the SUMO method. A. On-bead Ni-NTA (left) and heparin (right) affinity chromatography using SUMO-CREB auto-induced overnight at 18 °C. On-beads digestion with ULP-1 is not as high yield compared to a typical elution ("eluate"), but capable of isolating CREB. With these conditions, pure CREB is obtained post the denaturing heparin step. **B.** Auto-induction of CREB at 37 °C is lower yield and more contaminated than at 18 °C.

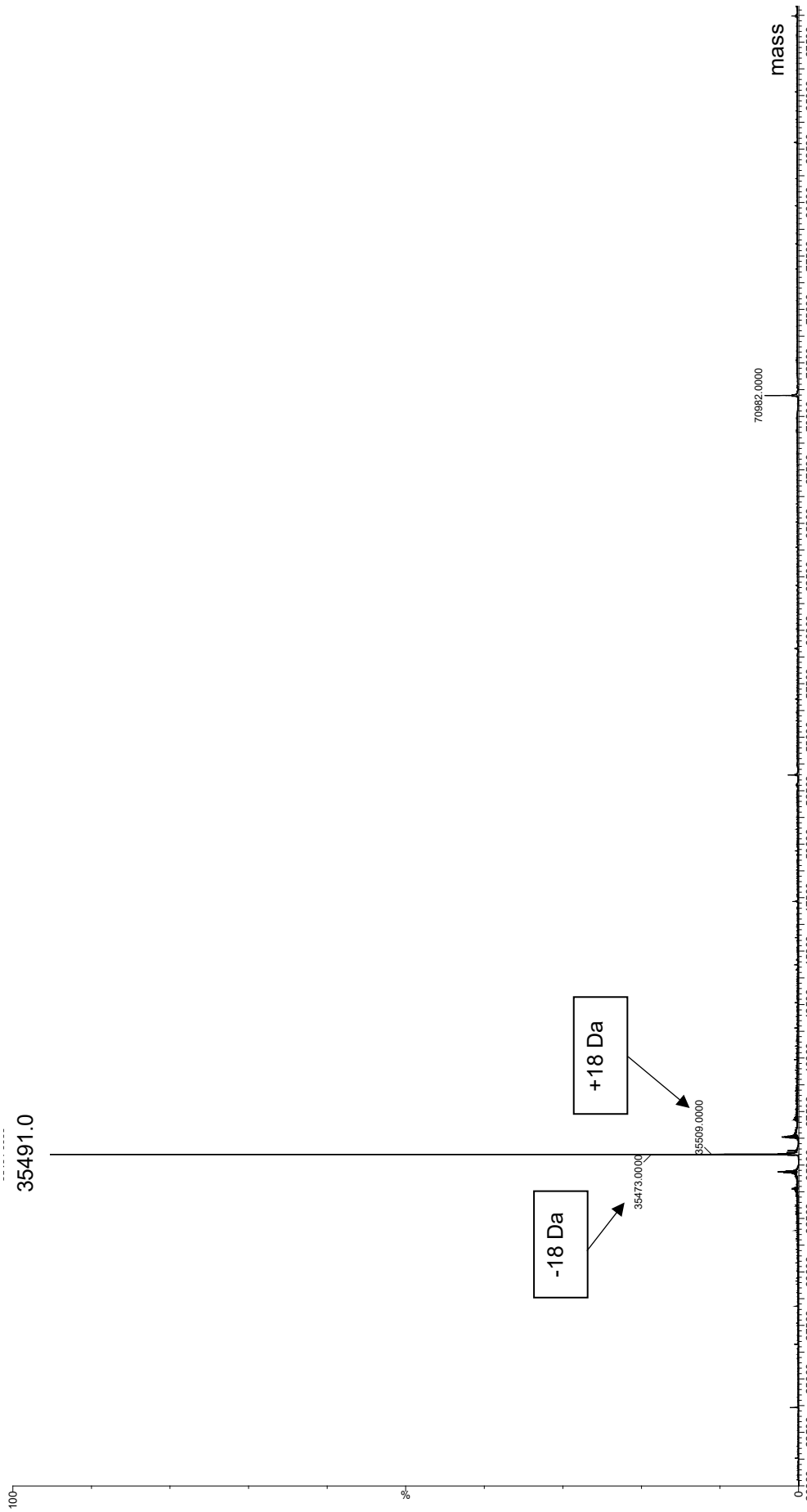


Figure 3.21: Mass spectrometry shows pure CREB is obtained with the SUMO method. LC-MS deconvoluted spectrum shows CREB from the SUMO method is contaminant-free, with 18Da size differences explained by water or $NH4^+$ adducts. No carbamylation (sometimes a consequence of incubation with urea) appears present.

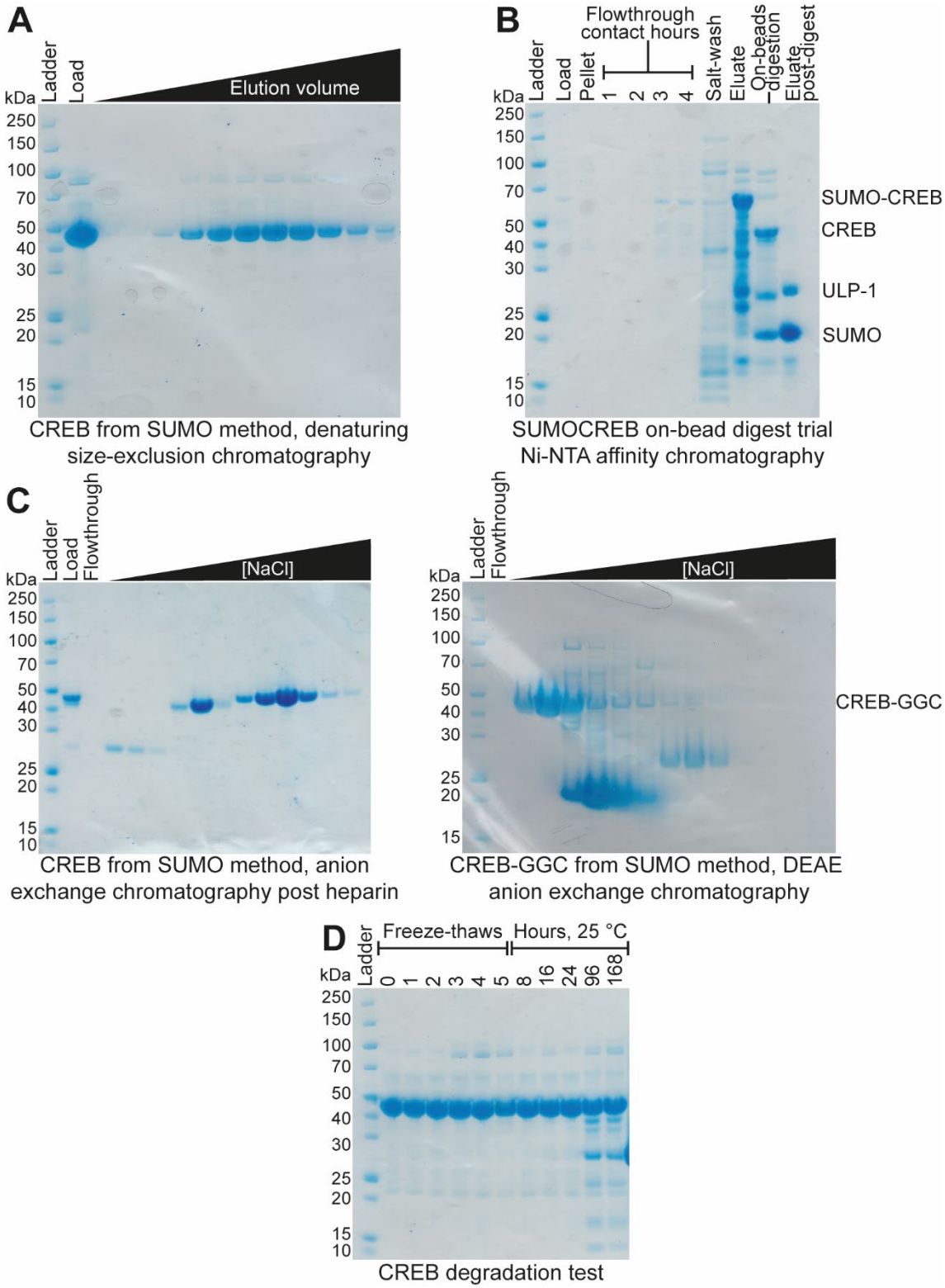


Figure 3.22: CREB from the SUMO method can be further purified if needed. A. SEC can further separate CREB from some minor contaminants. **B.** On-beads digest of 6His-SUMO-CREB using ULP-1 after some hours of contact-time is low yield, leaving a lot of uncleaved protein. **C.** Anion exchange can separate some contaminants post heparin (left) but isn't ideal pre-heparin, even if the different DEAE is used (right). **D.** Leaving CREB at room temperature for more than a day causes degradation. Freeze-thaws cycles don't degrade CREB much but may increase aggregation (thicker dimer band).

respectively) but GB1, HaloTag, TEV and TEV-CREB's pI are all around 5. Yet, this is better as a "refinement" step as it isn't as good as heparin if a Ni-NTA eluate is loaded (Figure 3.22). This is still useful for cysteine-containing CREB mutants destined for labelling, which are harder to purify. DEAE resin is not as powerful as HiTrap Q for ion exchange (Figure 3.22). Ni-NTA beads instead of columns can be used too, as this can circumvent the column and filter-clogging issue caused by increased bacterial production. On-bead ULP-1 cleavage of 6His-SUMO is, additionally, a potentially powerful step, ensuring CREB is almost exclusively the only protein released from the beads. Yet, this decreased yield with as much protein was left non-cleaved (Figures 3.20, 3.22), so it should only be considered if the current procedure doesn't remove contaminants for some recombinant protein preparations.

3.4. Discussion

Throughout this process, three methods were developed and compared to published ones (Figure 3.23). Throughout, I increased yield until I surpassed the published method and achieved similar purity levels without requiring nucleases. The importance of denaturant is obvious as 8 M urea is critical for a successful size-exclusion chromatography step, and all the methods that generate pure CREB also use denaturants.

The importance of a good solubility tag is also clear from this development. Lack of one clearly hampers the successful expression of soluble CREB; conversely, both SUMO and GB1 solubility tags are used to make pure CREB. It appears like SUMO is the superior tag, as SUMO-CREB expression limits aggregation compared to GB1-CREB. Plus, the requirement of glucose to limit basal expression in bacterial cultures further highlights the strain of full-length GB1-CREB expression on bacteria. This may be why auto-induction

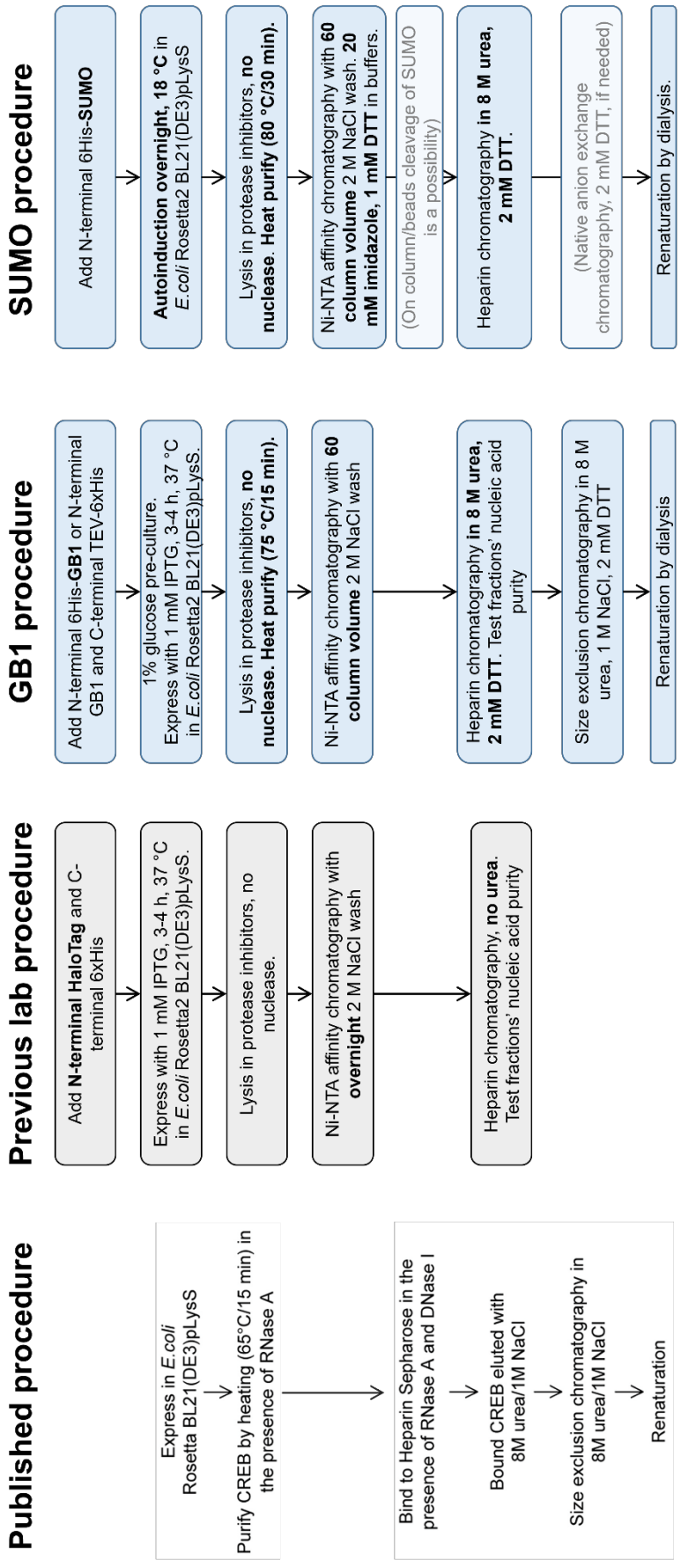


Figure 3.23: Overview and comparison of the four methods used and developed in this Chapter.

coupled with a pET vector is a good choice on many levels: glucose is in the expression medium, limiting basal expression until consumed, and pET is a lower copy-number vector, decreasing how much CREB is produced by basal expression, perhaps also indicating less CREB produced per cell upon induction and limiting aggregation. Yet, the greater number of cells offset the decrease in production to generate high yields. The different pI of both ULP-1 protease and SUMO to CREB also mean they are easily separated by ion-exchange chromatography if needed, thereby eliminating the need for yield-costly size-exclusion chromatography. Altogether, this makes the SUMO a powerful purification procedure and the go-to for this project. From here on, all pseudo wild-type CREB made from the SUMO procedure are referred to simply as "CREB" for convenience, with labelling-destined CREB with 1 cysteine is "CREB-1C"; and so-on.

However, such precautions highlight how fickle and cumbersome the current methods are. Cysteine presence favoured aggregation and often required some of the additional steps described in Section 3.4. Indeed, the absence of cysteines was one of the great initial improvements, perhaps showing the role of cysteine in binding contaminants via disulfide bonds, stabilising aggregates or DNA-bound CREB complexes. Moreover, changing a solubility tag for other useful tags (e.g. HaloTag) can have dire consequences on the final product. Since I use CREB exclusively *in-vitro*, such concerns can be discarded, but if one needs to use HaloCREB or wild-type CREB, it would be preferable to express it in another system (if needing purification) or expressing it directly in cells for cell-biology-style experiments, bypassing the need for purification entirely.

Chapter 4

Characterisation of CREB's biophysical properties

CREB is an essential transcription factor (204) acting downstream of the cAMP cascade (94) and is involved in a variety of biological processes, including regulation of the circadian clock (117), brain development and memory (112,113). Moreover, it is overexpressed in many varied cancer types (non-small cell lung cancer, clear cell sarcomas of soft tissue, leukaemia (118,204)) and neurological disorders (122–125). The study of this factor is thus essential both for both understanding human and cancer biology. Usually, such studies have focused on the isolated bZIP and KID domains, both of which are structurally resolved in a bound state (23,135,145). bZIP CD spectra also show folding upon binding (205), further highlighting the disordered nature of the domain. There have also been efforts to investigate CREB post-translational modifications. KID S133 phosphorylation by PKA causes CREB activation (94) and creates a high affinity binding interface to its co-activator CBP (22,24). Similarly, acetylation of KID lysine residues by CBP could increase CREB transactivation activity although it isn't quite clear how (147). Conversely, multisite phosphorylation of a nearby cassette promotes auto-inhibition between KID and bZIP (26).

Concerning full-length CREB itself, there have been CD spectra, a limited foray into CREB thermostability (107) and target binding affinity estimates (26,105,107,205). Until

very recently (155), there was only a single, inconclusive estimate for non-target affinity (107). Estimated homodimerization affinities differed wildly, although it was believed CREB is mostly monomeric from AUC data (162). The only available kinetic data are DNA residence times from single-molecule measurements in cells and *in-vitro* (206). Lastly, while there is some evidence that Q2 is important for CREB residency at promoters (151) and both Q1, Q2 recruit the general transcription machinery (152,154), the biophysical contributions of CREB's non-bZIP domains to DNA binding and target search as a whole is mostly unknown, although it is known they are important for basal expression (149).

In this Chapter, I've presented a detailed *in-vitro* biophysical characterisation of full-length CREB, including kinetics of its binding interactions. I present novel values, including kinetics, T_m and thermodynamics of DNA binding. I also estimate some previously examined affinities with greater precision. Lastly, I explore the influence of pH and C-terminal sequences on CREB parameters.

4.1. CREB secondary structure predictions

Purified CREB was scanned using circular dichroism to infer secondary structure content (Figure 4.1A). My spectra are very similar to published ones, which is another indication that good quality CREB is purified by the methods discussed in the previous Chapter (Figure 4.1B). Secondary structure content was estimated from a variety of online software, the most recent being BeStSel (180) (Figure 4.1D). The protein was expected to be largely disordered because disordered predictors (207–209) suggest full-length CREB will be almost entirely disordered, except at the bZIP (Figure 4.2). Surprisingly, CD analysis consistently estimated CREB be almost 20% beta-sheet (Figure 4.1D). The lowest wavelength reliably scanned in biophysical buffer (at the concentrations and pathlength

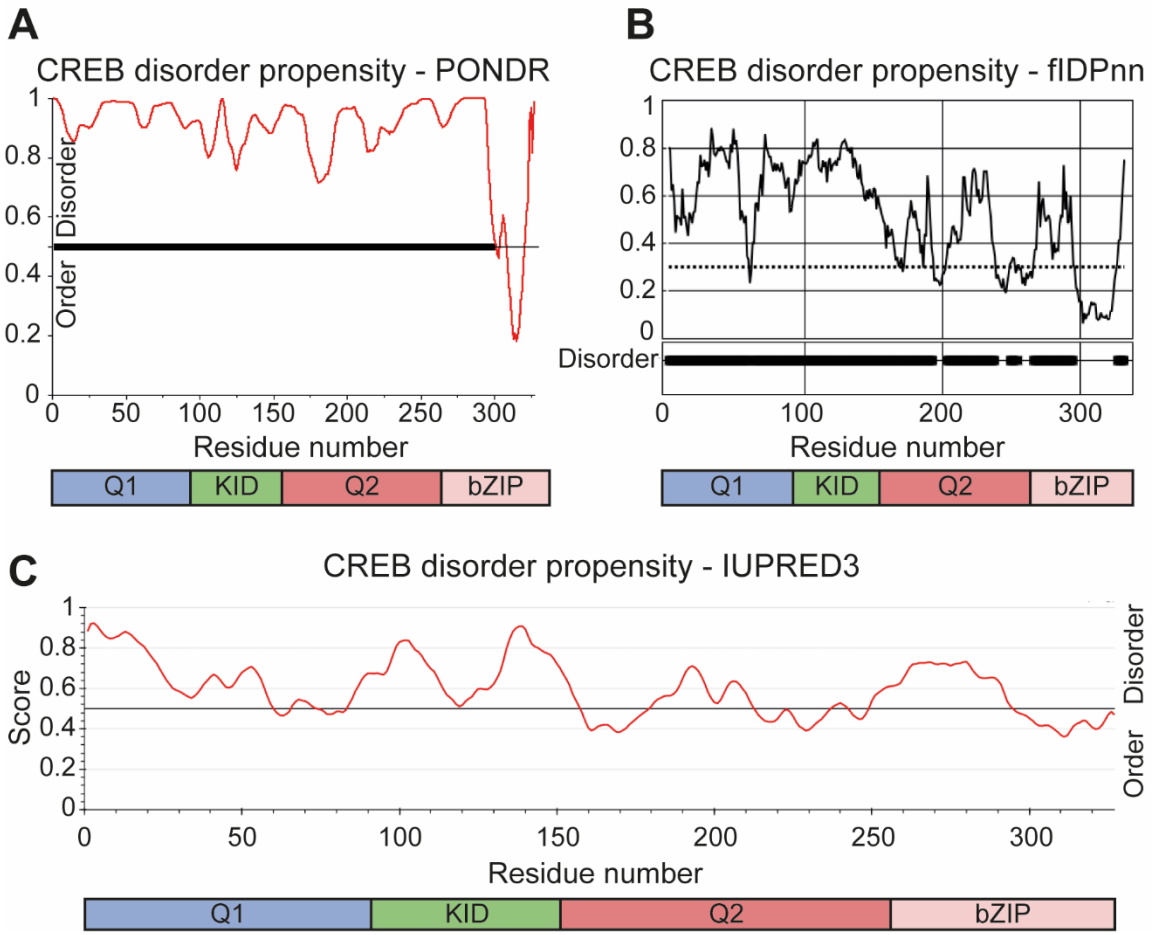


Figure 4.1: Disorder predictions for CREB. PONDR (**A**), fIDPnn (**B**), and IUPRED3 (**C**) servers concur that only the C-terminal region of the bZIP (corresponding to the zipper) have structure, although fIDPnn and IUPRED3 predict some structures in Q2 – where AlphaFold predicts some sheets (Figure 4.2).

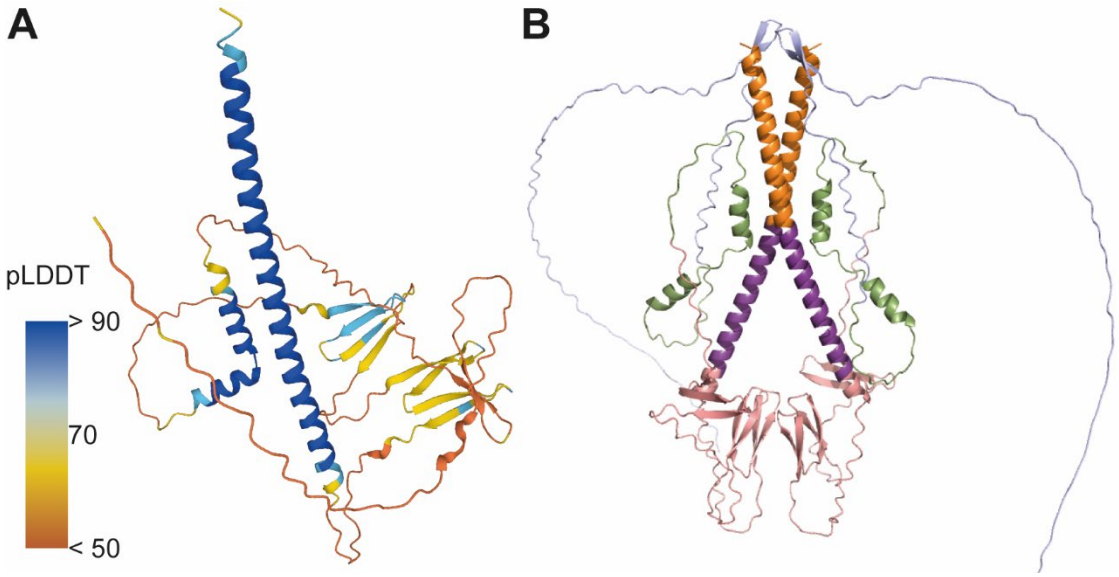


Figure 4.2: AlphaFold predictions of CREB. A. CREB monomer, with the confidence in pLDDT (per-residue local distance difference test), and **B.** CREB dimer, with colours corresponding to CREB domains: as per Figure 1.8. Both show a substantial, unexpected number of beta-sheets as well as the expected helical regions in bZIP and KID, or the many unstructured regions.

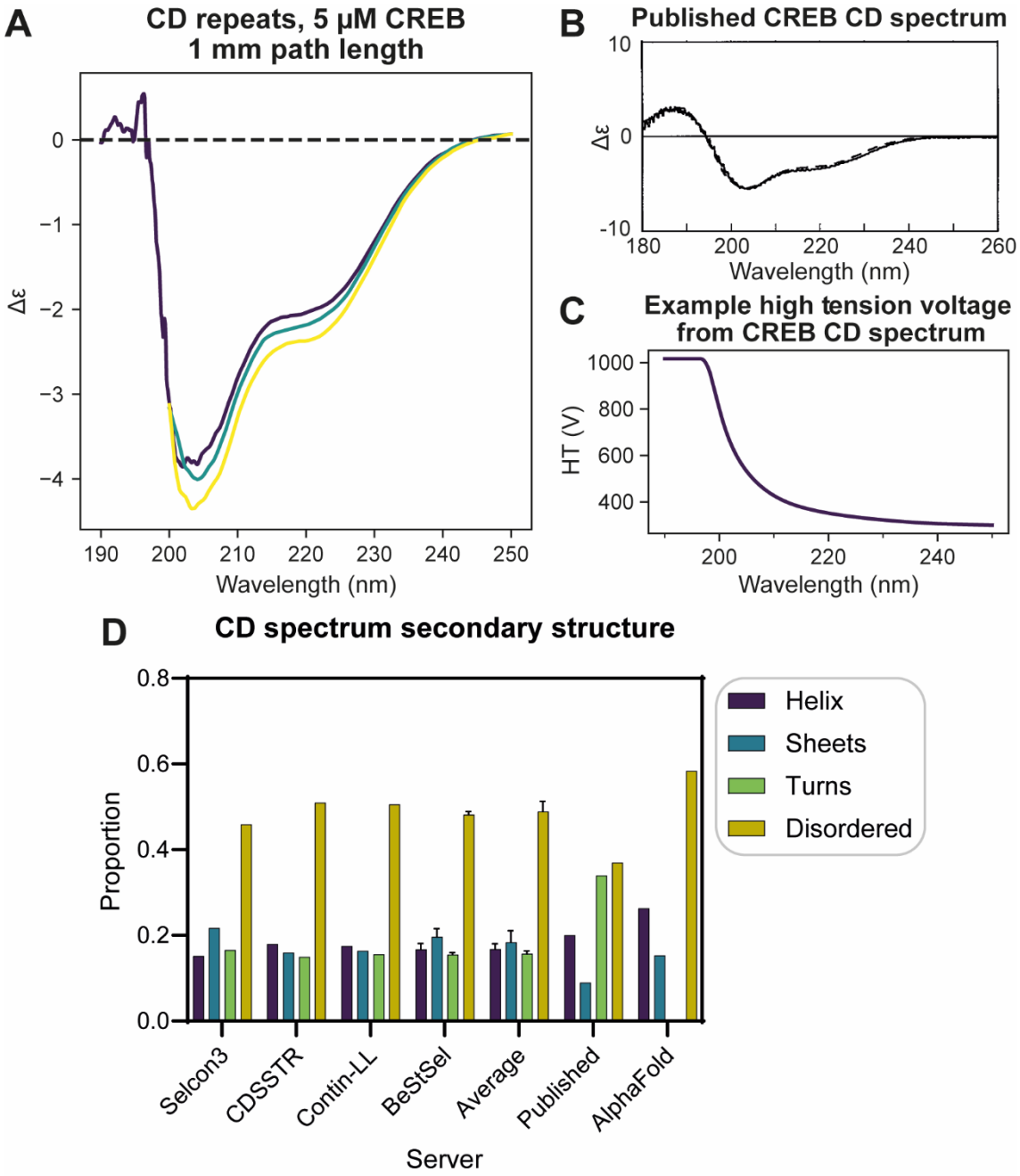


Figure 4.3: CREB CD spectra and structure predictions. A. CD spectra of CREB. General shape is reproducible. **B.** Spectrum of CREB as published by (108). The shape and raw values are like those obtained here. **C.** Example high tension voltage from the longest-range spectrum in (**A.**). Beyond 200 nm, the high voltage renders spectra unreliable. **D.** Secondary structure predictions by various servers from the CD spectra. CREB is predicted to be mostly disordered, but an almost equal proportion of sheets to helices is predicted. AlphaFold predictions are somewhat similar, with an in-between sheet content to the average predictions and publication, but with a higher helical content. Conditions: 25 °C, pH 7.4.

used) is 200 nm. The presence of salt causes high absorbances (and therefore great high-tension values), making CD signal below this wavelength unreliable so this may be an overinterpretation of the data (Figure 4.1C) However, it is notable that AlphaFold predictions (Figure 4.3) also have beta-sheets.

4.2. Investigation into CREB thermostability

4.2.1. CREB structure melts at high temperatures but recovers well.

Incubating CREB at increasing temperature gradually melts CREB structure. This is seen through both whole CD spectra (Figure 4.4A) and at 222 nm only, a wavelength chosen because its signal reflects helical content well (Figure 4.5A). Importantly and as shown previously (107) pre and post-melt spectra at the same temperature overlap well (Figure 4.6). This suggests that CREB has structure that melts but can recover it on cooldown in a reversible fashion, and without significant thermally induced aggregation. Correspondingly, the BeStSel secondary structure predictor algorithm suggests both sheets and helices melt with increasing temperature, with a compensating rise in disorder (Figure 4.4B). Interestingly, some new structure may be formed as helical content rises again after a certain temperature around the T_m (more in Section 4.2.ii), and β-turn percentage stays constant.

4.2.2. Finding out CREB's T_m

Looking at the different melted spectra of CREB, it seems like there starts to be some overlap from 55 °C, which suggests the T_m is in that range. The melt data from 222 nm can be used to get more detail and give the fraction folded and unfolded protein at each temperature by extracting the folded and unfolded baselines at low and high temperatures respectively (Figure 4.5A, Equation 2.22). This can be fit to Equation 2.23 that suggests the T_m is (53 ± 2) °C (Figure 4.5B).

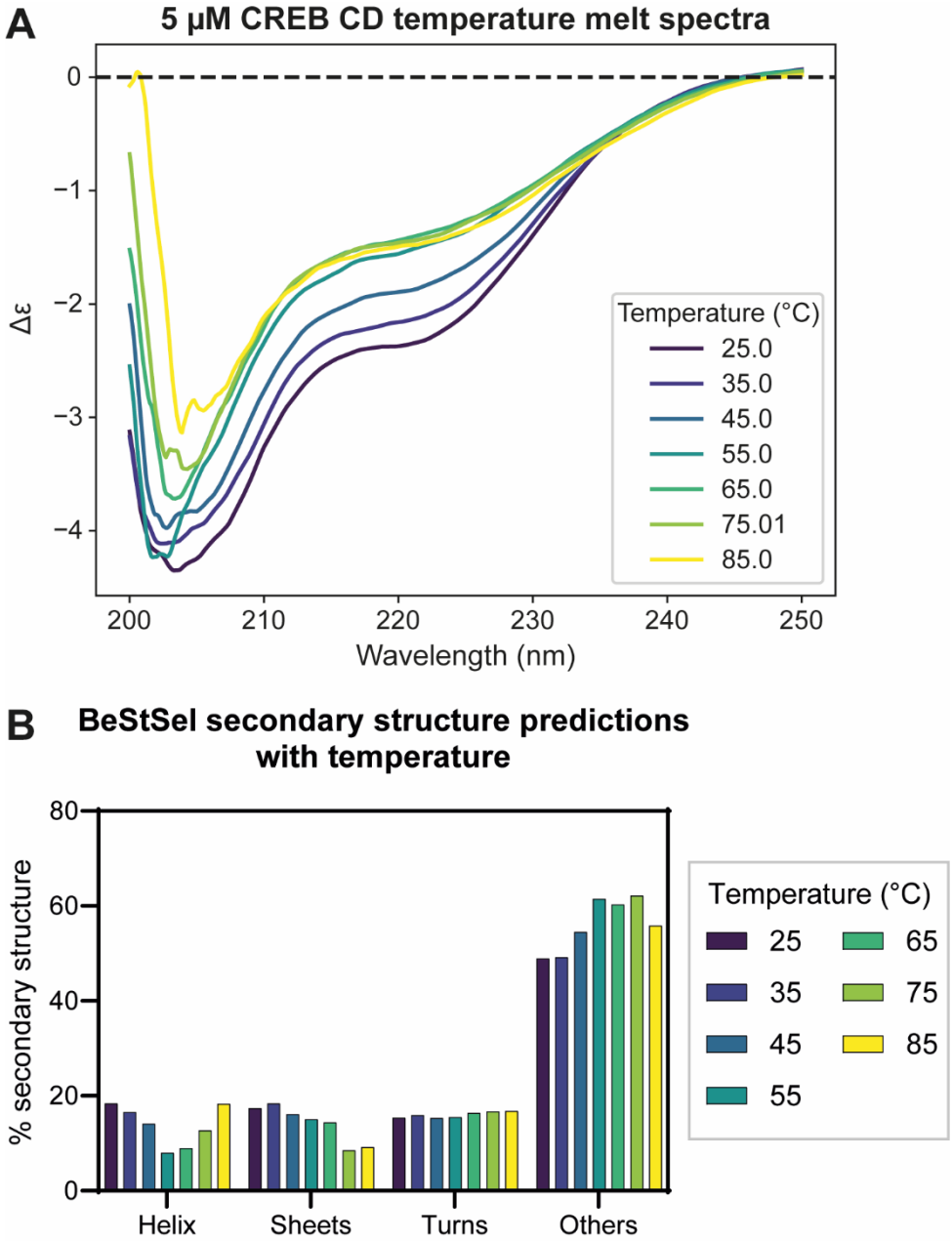


Figure 4.4: CREB CD spectra melt and structure predictions at different temperatures. A. CD spectra of CREB with increasing temperature at pH 7.4. Some structure is melted although, starting from 65 °C, there is no signal change before 210 nm. **B.** BeStSel secondary structure prediction from the spectra. Whilst the amount of turns stays constant, there is a consistent decrease in sheet content and a dip in helical content follow by a rise after 55 °C. Conversely, disordered ("others") content rises until 55-75 °C before falling again.

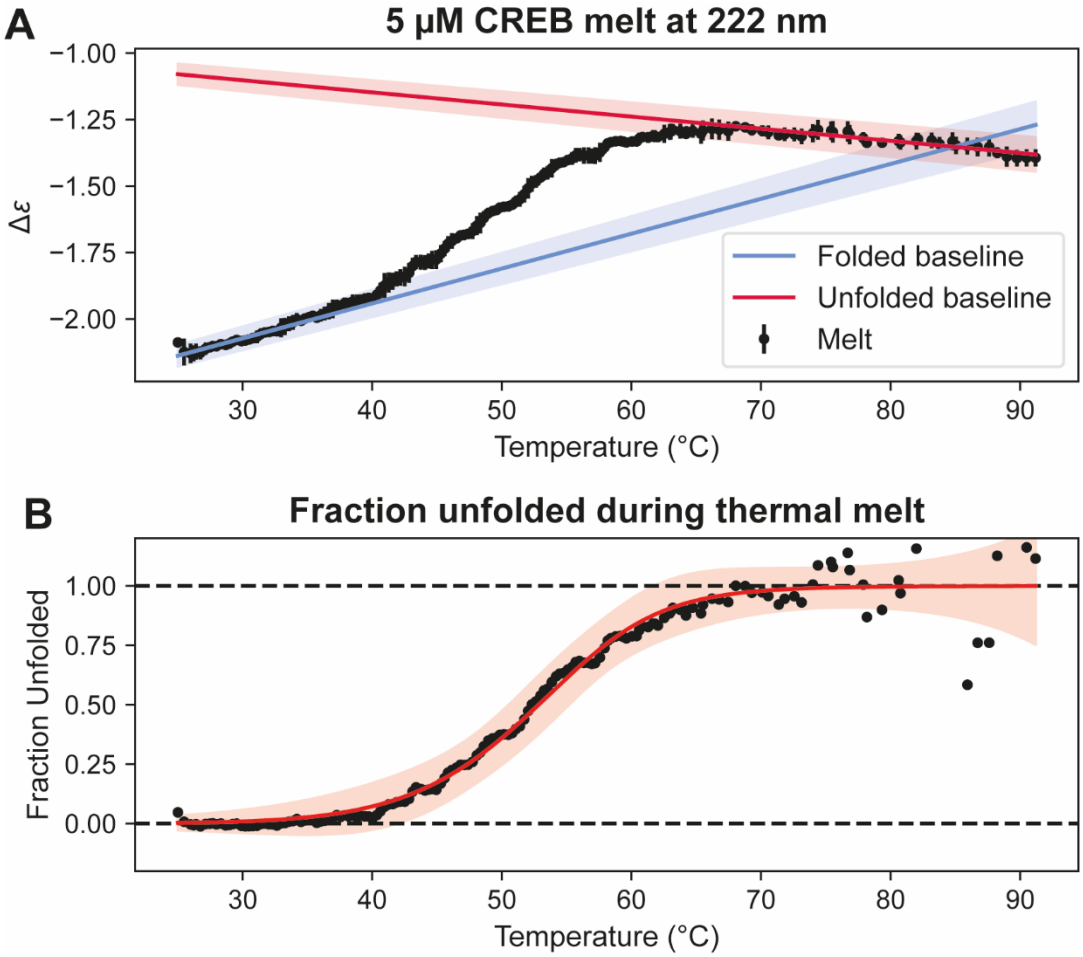


Figure 4.5: CD thermal melt analysis. A. CD melt at 222 nm, pH 7.4, with extrapolated unfolded and folded baselines. **B.** Fraction unfolded with fit to extract the T_m = (53 ± 2) °C.

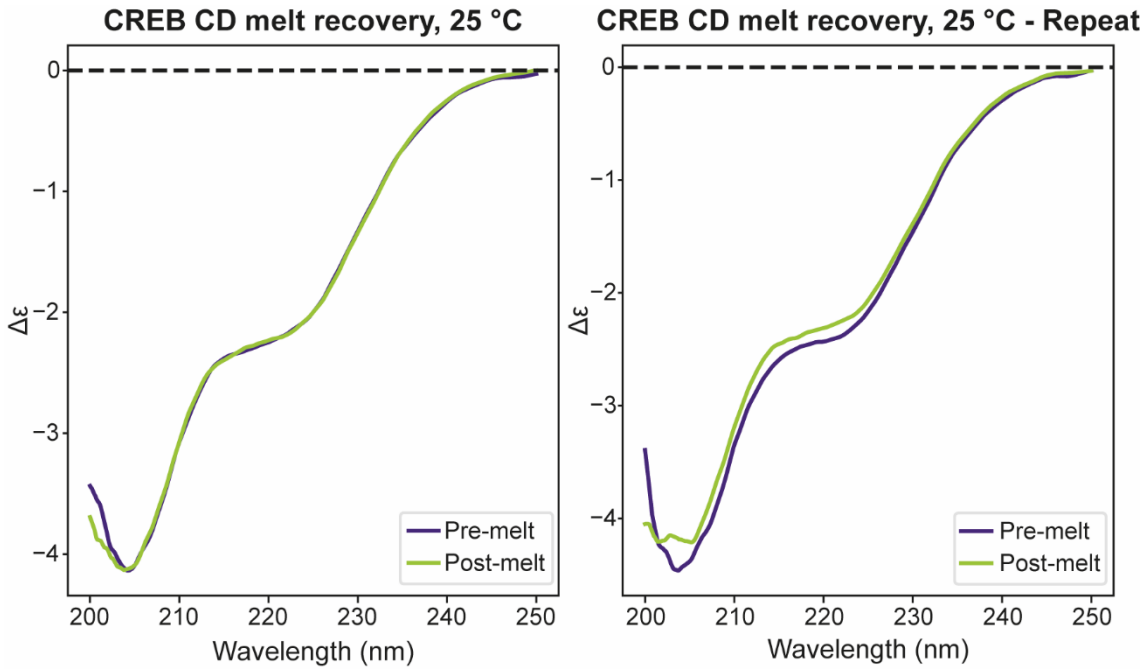


Figure 4.6: CD thermal melt recovery. Both pre and post-melt spectra at 25 °C, pH 7.4, 5µM CREB, overlap well, confirming previous claims that CREB recovers from heat treatment (107).

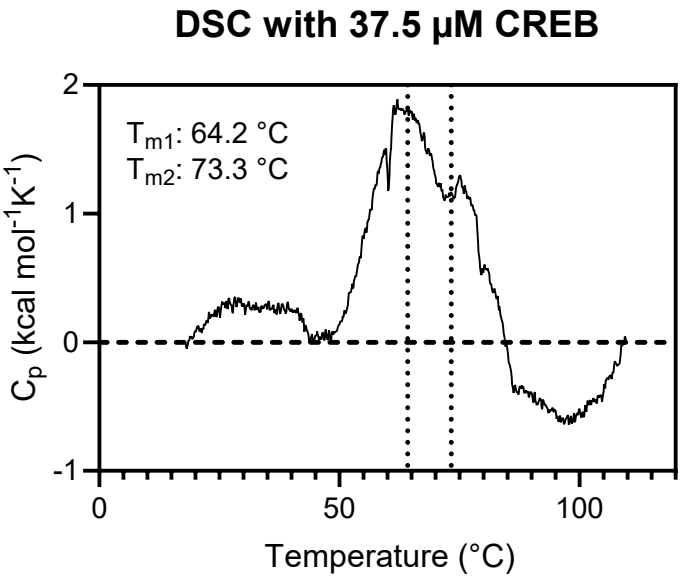


Figure 4.7: Differential scanning calorimetry (DSC) of CREB. Although the data are reportedly low quality, likely due to disordered nature of CREB, T_m are extracted around 64 °C and 73 °C (biophysical buffer pH 7.4). The discrepancy with the CD value is likely due to the difference in protein concentration.

A differential scanning calorimetry experiment was attempted to corroborate this number. Although the data were low quality (likely due to disordered nature of CREB), two higher T_m of 64 °C and 73 °C are extracted (Figure 4.7). The discrepancy could be explained by a reportedly low-quality DSC data along with 8 times more protein used in the DSC experiment compared to the CD melt experiment.

Certainly, CREB does have structure that can be melted with a T_m around 60 °C but isn't permanently affected by this loss and recovers its structure, suggesting that the use of heat in the purification process is justified.

4.3. Investigation into CREB equilibrium binding properties to target and non-target DNA

4.3.1. CREB's affinity to target CRE probed by anisotropy titrations.

Hairpin DNA containing a CRE site and labelled with Alexa Fluor™ 488 (CREh-Alexa488), was titrated using CREB and fit using a single site binding model (Equation 2.3).

Repeat titrations all showed CREB binding to DNA in the low nanomolar range (Figure 4.8A,C), as reported previously (26,105,107,205). This is binding specific to CRE and not to the hairpin DNA structure, as control titrations with linear CRE ("CRElin"), also Alexa Fluor™ 488-labelled, have similar affinity (Figure 4.8A,C), and shows again that CREB has very high affinity for CRE. Intensity change upon binding must be accounted for (Figure 4.8B), and applying corrections (Equation 2.8) (175,210) does increase the $K_{D,app}$ slightly (Figure 4.8C). Post corrections, $K_{D,app}$ = (5.4 ± 0.9) nM. As these titrations were taken at very low 1 nM probe concentration I ensured intrinsic buffer fluorescence intensity has virtually no effect on the $K_{D,app}$ (Figure 4.8C). Yet, CREB concentrations below the homodimerization K_D are used, which is around 10-30 nM ((136,155,163), more in Chapter 6). Consequently, the equilibrium curve probes both dimerization and CRE binding simultaneously. This is showcased by a small deviation from the single-site binding model, as shown by residuals around the transition region of 1-10 nM (Figure 4.8A).

Anisotropy titrations were also trialled using fluorescently labelled CREB as the probe and unlabelled CRElin as the titrant. CREBs with a single cysteine were purified and fluorescently labelled with Alexa Fluor™ 488 (CREB-Alexa488) or Alexa Fluor™ 594 (CREB-Alexa594) using thiol-maleimide following published procedures (155). TCEP was found to compete for labelling and was removed from the labelling mixture by PD-10 columns before addition of dye. Mass spectrometry and gel analysis suggests CREB is fully labelled (Figure 4.9). Titrations with CREB as a probe have intensity changes over larger concentration ranges and, interestingly, have better-looking intensity than anisotropy change (Figure 4.10). Regardless, the transitions don't fit as well to a single-site binding model and, although the extracted $K_{D,app}$ using intensity values are in the same range as the preceding results, they are also below the probe CREB-Alexa594 concentration, which means they are unreliable. Overall, using CREB as the probe is possible but less suitable than using DNA.

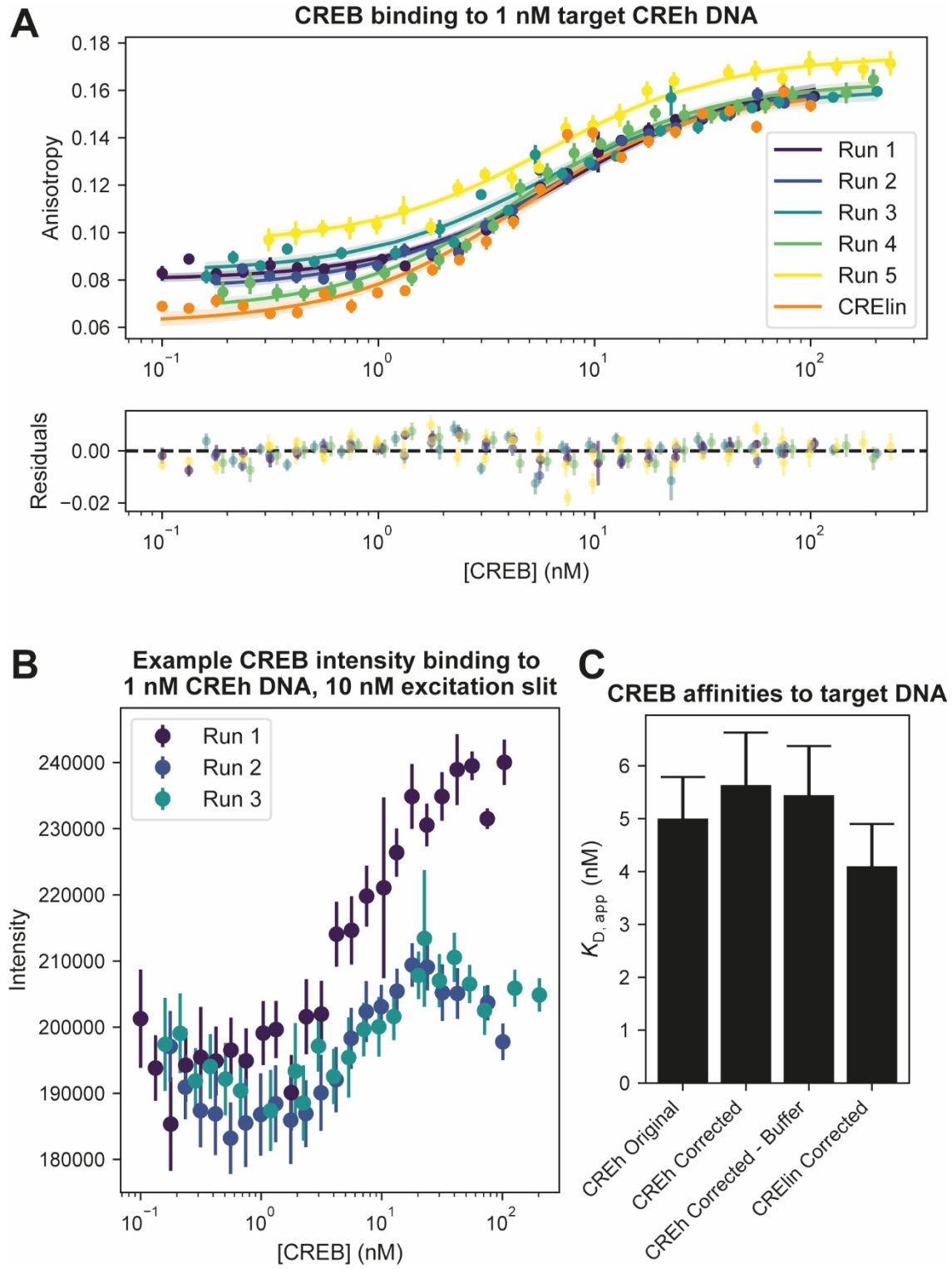


Figure 4.8: Anisotropy titrations of CREB binding to CRE labelled with Alexa Fluor™ 488. A. All titrations (pH 7.4, 25 °C) show reproducible behaviour in absolute anisotropy values and in concentrations where transitions happen. Both CRElin and CREh behave similarly. **B.** Binding to DNA causes a fluorescence intensity change, which must be accounted and corrected for when analysing anisotropy data. **C.** While all are within error, correcting for intensity changes and increases K_D slightly. Intrinsic buffer fluorescence has almost no influence. There appears to be slightly higher affinity for a linear CRElin oligo, but all are within error of each other.

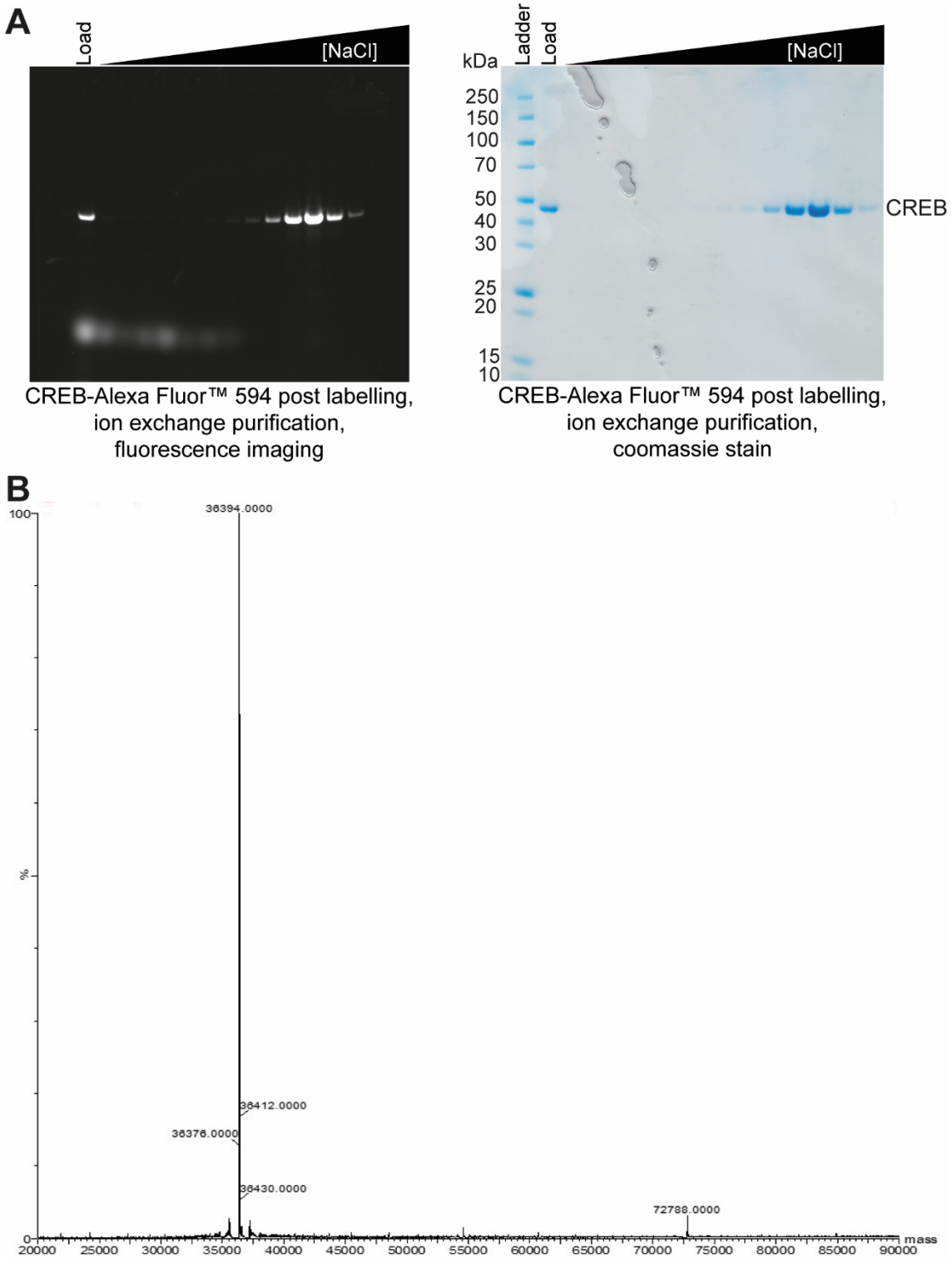


Figure 4.9: CREB labelling with Alexa Fluor™ 594. A. Post labelling ion-exchange chromatography shows only a single labelled species is eluted. **B.** Mass spectrometry shows one main species with a molecular weight +68 Da above expected labelled species. This is likely labelled CREB ionic such as: TFA, two Na or Acetonitrile+Na, all commonly found in buffers.

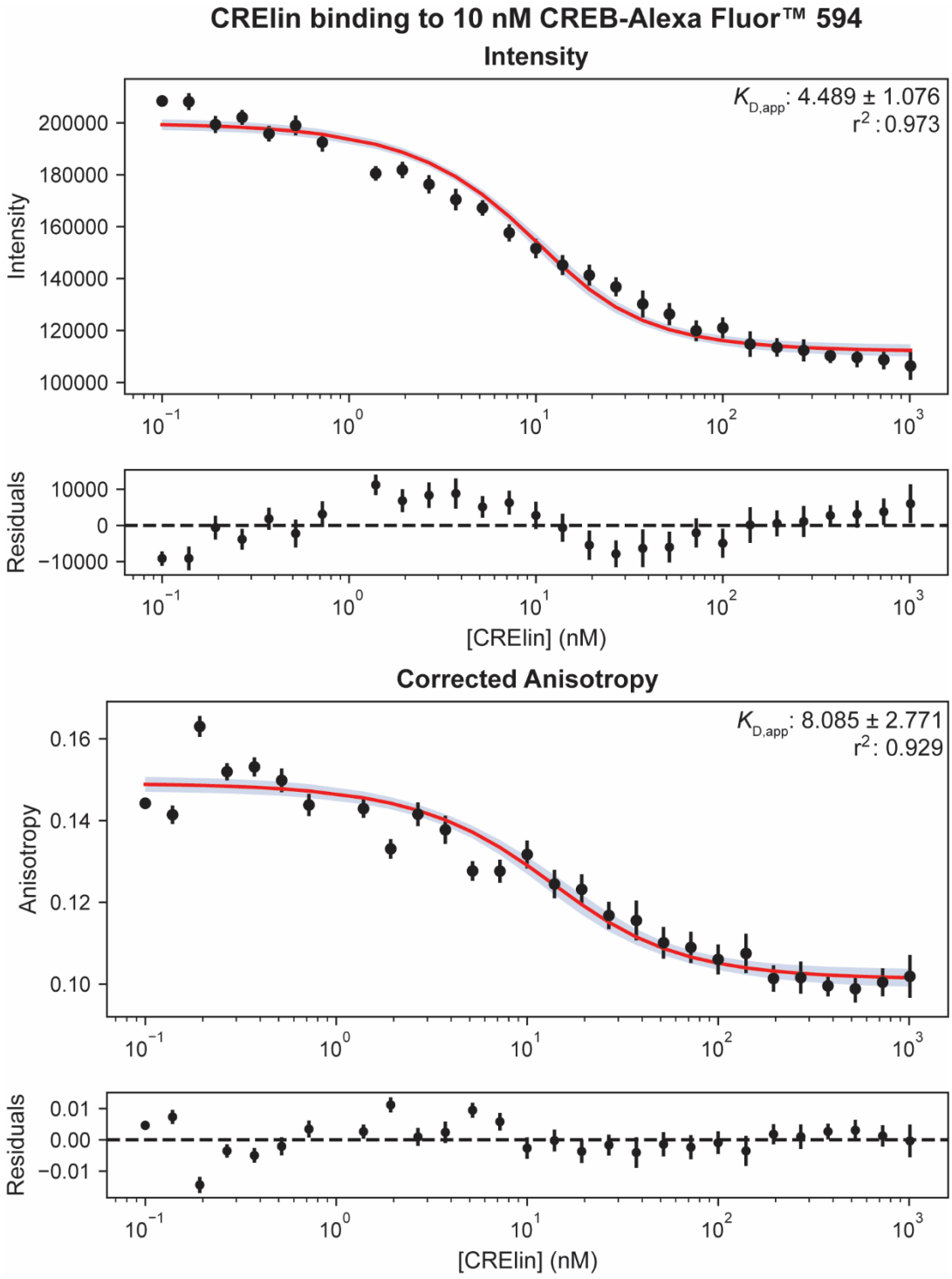


Figure 4.10: Reverse anisotropy titrations using 10 nM CREB-Alexa Fluor™ 594 as a probe bound by unlabelled CRElin. Using CREB as a probe gives a good intensity change that can be fit with a single-site binding model (top). It fits slightly better and gives a value closer to those from usual titrations than the corrected anisotropy from the same titration (bottom). However, both give a K_D below the concentration of the probe which makes the estimate reliable. Titration at pH 7.4, 25 °C.

4.3.2. Thermodynamics of CREB to CRE association.

Isothermal titration calorimetry (ITC) is a technique that uses the heat generated during a biomolecular reaction by adding titrant to extract a variety of parameters – notably K_D of interaction, but also stoichiometry and a variety of thermodynamic parameters. It is also label-free so potentially a more "native" way to study interactions. Whilst both are advantages compared to fluorescence anisotropy, ITC consumes more material. ITC's narrower 1 nM-100 μM working K_D range (211) can be greatly reduced by low reaction heat as well and require unconventional setups to expand the range as a result.

This is why ITC is not well-suited to find the K_D of binding to CRE. At least 5 μM CREB is required in the ITC reaction cell for an observable signal change if using CRElin as a titrant. Due to high affinity, the ITC c-value in these conditions is above 500, so these titrations cannot find K_D but can still investigate binding thermodynamics and stoichiometry. Perhaps surprisingly, ITC suggests that the reaction stoichiometry is n=0.3, or that around 3 CREB monomers bind to a CRElin oligo, when it is accepted that CREB binds DNA as a dimer (135) (Figure 4.11, 4.16C). Regarding the interaction's thermodynamics, ΔG and therefore ΔS are unreliable due to the titration's inadequacy for extraction of a K_D, but the binding enthalpy, ΔH is estimated around -40 kcal/mol (Figure 4.11).

4.3.3. Exploration of non-target DNA affinity using fluorescence

How CREB binds to non-target DNA has not been studied thoroughly. Considering how much more likely it is that transcription factors encounter random DNA rather than their target-site, and that the facilitated diffusion model predicts that factors will bind non-target DNA to find their target DNA, this seems like an oversight that needs to be explored. Indeed, studies of target search are increasing in prominence in the transcription factor community (18,46,48,50,53,57,67–69,73,74,212).

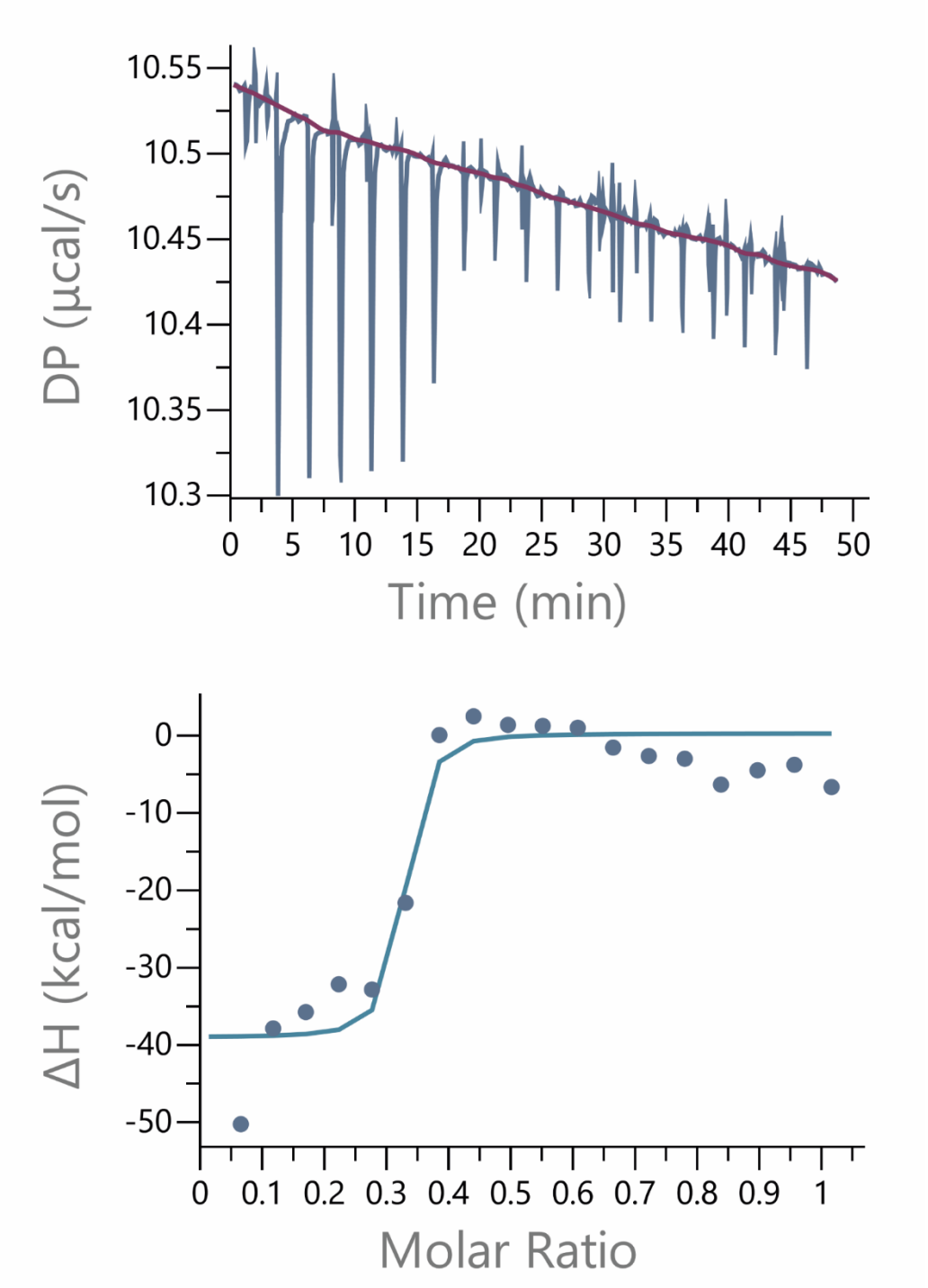


Figure 4.11: ITC titration with CREB bound by CRElin. Top: raw heat change. 5µM CREB is used (c-value above 500) and the change is very small. It is unlikely that good quality signal would be obtained by using less protein, and therefore this is a K_D that lies outside the ITC capability unless a competition assay is used. Bottom: normalised heat change, typical for a high c-value titration, where K_D is inaccessible but the "n-value", an approximation of the reaction's stoichiometry, is observed at the location of the signal change. n~0.3, substantially lower than the 0.5 expectation. ΔH of binding is around -40 kcal/mol according to the fit (the difference between the lower and upper plateau). Conditions were pH 7.4, 25 °C.

Linear 14 bp nCRE DNA (labelled with Alexa Fluor™ 488, "nCRE-Alexa488") was titrated with increasing concentrations of CREB and fluorescence anisotropy change monitored to get an estimate of CREB binding affinity to non-target DNA (Figure 4.12A). Because the affinity much reduced compared with target DNA, it is possible to use more labelled DNA probe (5 nM rather than 1 nM) and still maintain probe concentrations below the binding affinity as required in these types of studies. This limits noise and also the effect of buffer fluorescence on anisotropy calculations.

CREB binding to nCRE does not cause an intensity change, so there is no need to correct the measured anisotropy (Figure 4.12B). Since the signal change also occurs at much higher CREB concentrations, it is unlikely dimerization needs to be corrected for either. Unfortunately, no upper plateau from these titrations is obtained. even when attempted with higher [CREB] to 100 μM. Much higher concentrations CREB, can lead to aggregation (as disordered proteins are prone to) and, subsequently, unreliable fluorescence values, so titration curves with a clear plateau will not be possible from this method.

Titrations under 50 μM can be fitted using equations suitable for a single or two binding sites. However, the higher concentration titration is only fitted well with a two binding-sites model (Equation 2.4, Figure 4.13A). Moreover, changing the fit to a two-sites model also halves the $K_{D,nCRE}$ (Figure 4.13B), showing the fitting model does affect the extracted value. Therefore, fitting with a two-sites model seem preferred to consider the lack of upper plateau. Averaging the result of multiple titrations gives a nCRE $K_{D,app} = (1.4 \pm 0.2)$ μM, almost 1000x higher than target affinity.

In an attempt to access higher concentration ranges I also performed "reverse titrations" with CREB-Alexa594 as the probe and unlabelled linear nCRE as the titrant. DNA oligos can reach much higher concentrations without aggregating. Moreover, one

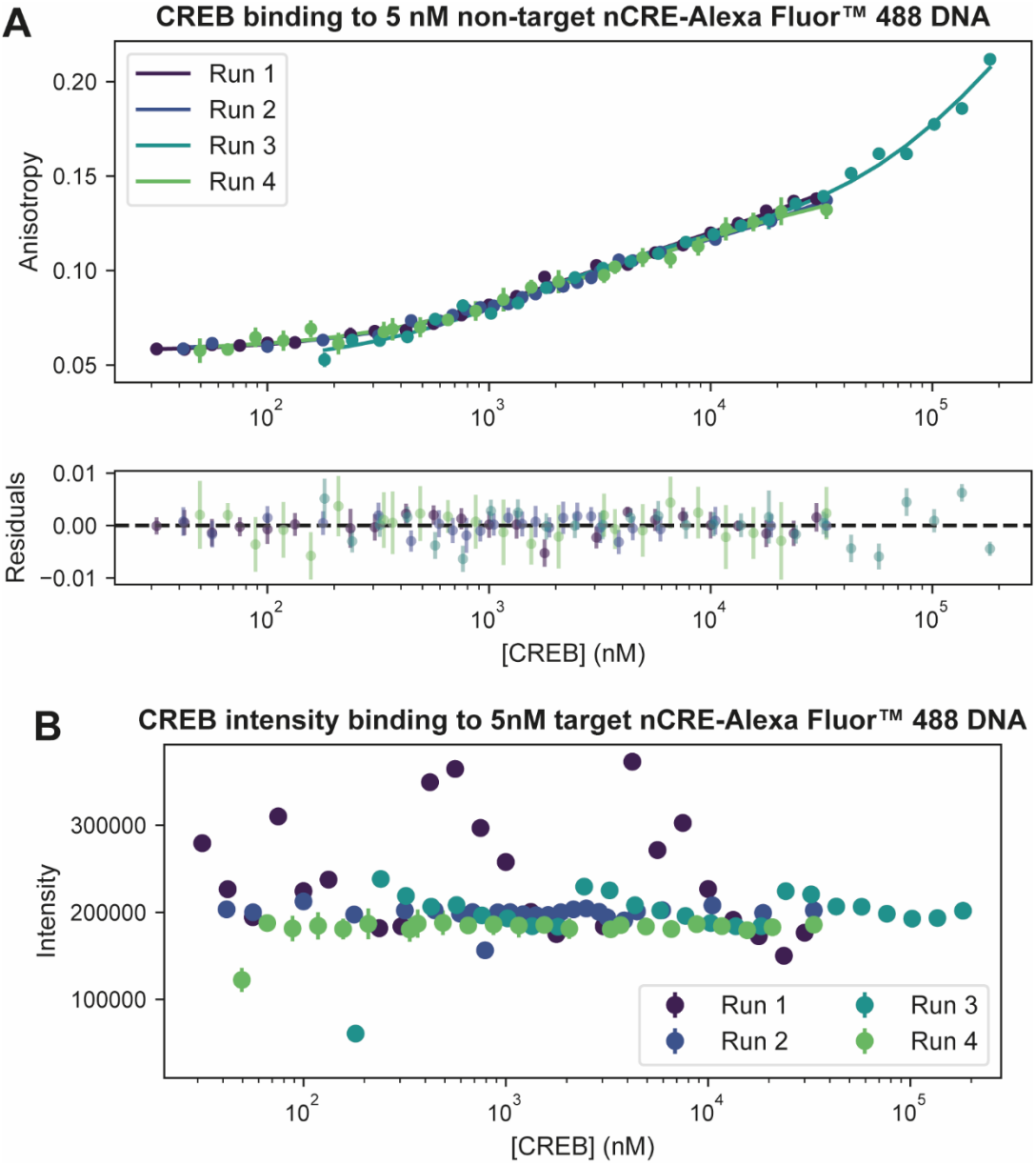


Figure 4.12: Anisotropy titrations of CREB binding to 5 nM nCRE-Alexa Fluor™ 488. A. Anisotropy titrations of CREB binding to nCRE show a transition around the µM range, but a desirable upper plateau is not observed. All traces overlap well and are fit by a two binding-sites model with different affinity for each site. **B.** Intensity values do not change with binding, making anisotropy corrections unnecessary. Titration conditions were pH 7.4, 25 °C.

interpretation of the second transition is that it represents secondary binding events. In this circumstance we can expect that the value for the first transition is similar between reverse titrations, but that for second transition may be different (or not present) in one direction. This is because loading multiple proteins onto a single DNA molecule is likely to be fundamentally different to loading multiple DNA molecules onto a single protein.

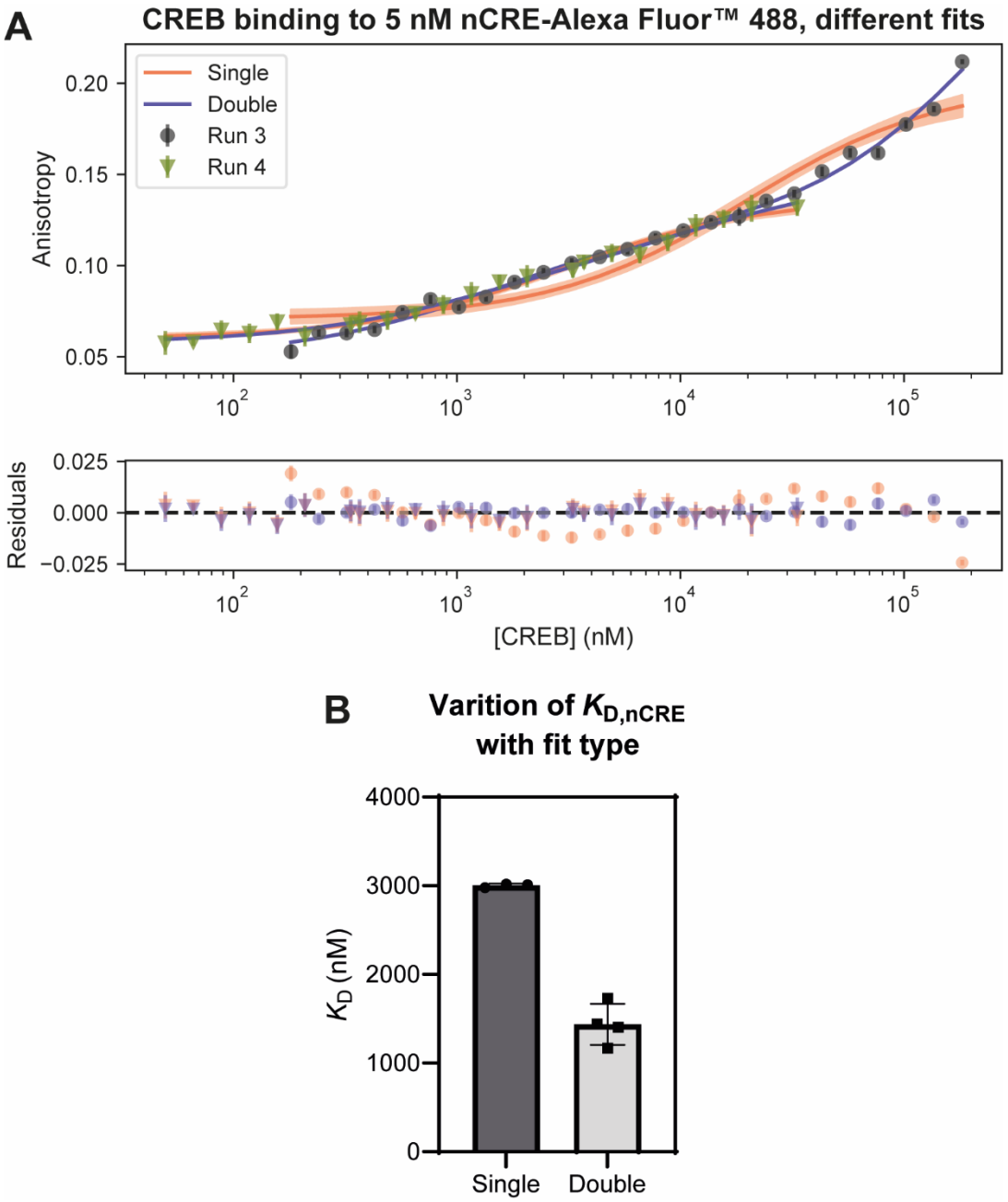


Figure 4.13: Two-sites models are better suited to fit CREB anisotropy titrations of binding to nCRE labelled with Alexa Fluor™ 488. A. Two-sites models (with different K_D for each site) fit nCRE titrations better, although the effect is less prevalent if titrations use lower CREB concentrations. **B.** $K_{D,nCRE}$ are two-fold higher if using a single-site compared to a two-sites ("Double") model (only the first K_D is considered). In both cases, extracted values are reproducible.

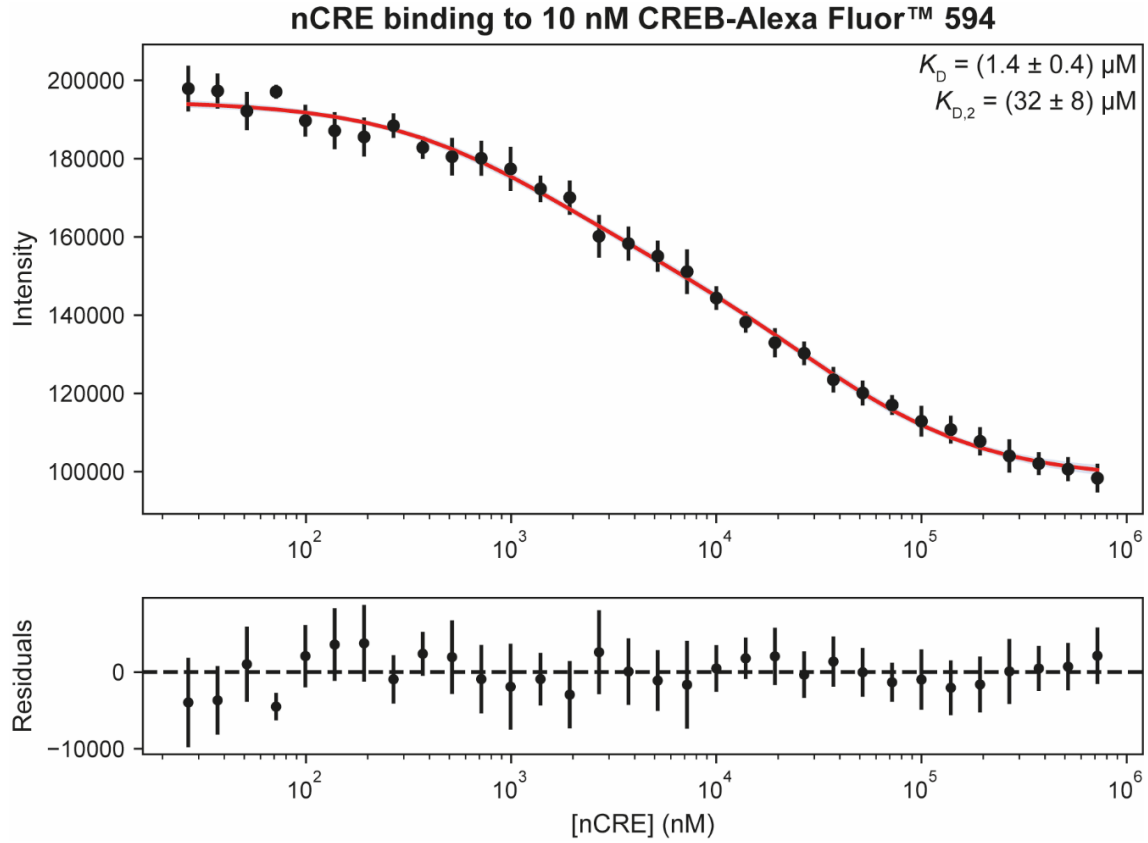


Figure 4.14: Reverse intensity nCRE titration using CREB labelled with Alexa Fluor™ 594 as a probe. Just like for CRE (Figure 4.10), labelled CREB is less suited than labelled DNA for titrations. Although a greater range of DNA concentrations can be used, a high concentration plateau is still missing despite approaching the millimolar range. Intensity changes as well as anisotropy and can be used to fit and extract $K_{D,nCRE}$. Here the titration that used the most DNA is shown and is the only one able to be fit only with a two-sites model, giving a K_D value similar to those obtained from titrations with labelled nCRE. Conditions were pH 7.4, 25 °C.

Like those attempted with target DNA, there is a good fluorescence intensity decrease with binding and therefore no need for fluorescence anisotropy measurements (Figure 4.14). Regardless, these titrations are less reliable and reproducible with transitions occurring over large concentration ranges. Even when nearing 1 mM nCRE, the upper plateau is not resolved fully. The titration with the highest DNA concentration is fit well by a two binding-sites model which gives a K_D = (1.3 ± 0.4) µM, but titrations which stop at a lesser [nCRE] give a K_D into the 3 µM range (not shown). Unfortunately, using CREB as a probe means using it at a concentration near the homodimerization K_D which may affect the equation needed to fit the titrations. Hence, using CREB as a probe is possible and gives

$K_{D,nCRE}$ in the μM range also, but values still do not reach an upper plateau and data are less reliable and reproducible than using DNA as a probe.

4.3.4. Exploration of CREB binding to nCRE using isothermal titration calorimetry (ITC)

The micromolar range of CREB binding to nCRE is ideal for ITC. I performed ITC titrations with solutions of typically 30-40 μM CREB (in the cell) being titrated with 190-240 μM nCRE. The use of such high concentrations ensures an ideal titration range ("c-value") for ITC and limits the influence of dimerization.

There are multiple corrections possible to choose from during ITC analysis: Fitted Offset (FO, simply subtracting the baseline calculated from datapoints during an experiment), Single (subtracting results with a titrant into buffer titration using the exact same titrant stock), and All (adding to Single, the effect of buffer into buffer and buffer into titrand solution). Upon testing the various types, "All" and "Single" have almost no difference (Figure 4.15), but FO and Single do differ slightly (Figure 4.16). This means that dilution of DNA into buffer contributes to the ITC signal change and should be accounted for, but there is no need to correct for other factors.

The ITC titrations were well-fit by single-site binding models, with a $K_{D,nCRE}$ = (1.4 ± 0.4) μM, very similar to the lowest (first) $K_{D,nCRE}$ extracted from the anisotropy titrations using a two-sites model. However, it appears ITC is likely a better tool to study this mechanism – extra data is gathered like stoichiometry and thermodynamics, no labels are required, and only a single transition is observed as an upper plateau is reached (Figure 4.16A). This is likely because the titrant:titrand ratio is closer to 1:1. Therefore, there isn't excess protein or DNA, and CREB dimers only binds to a single nCRE site in turn.

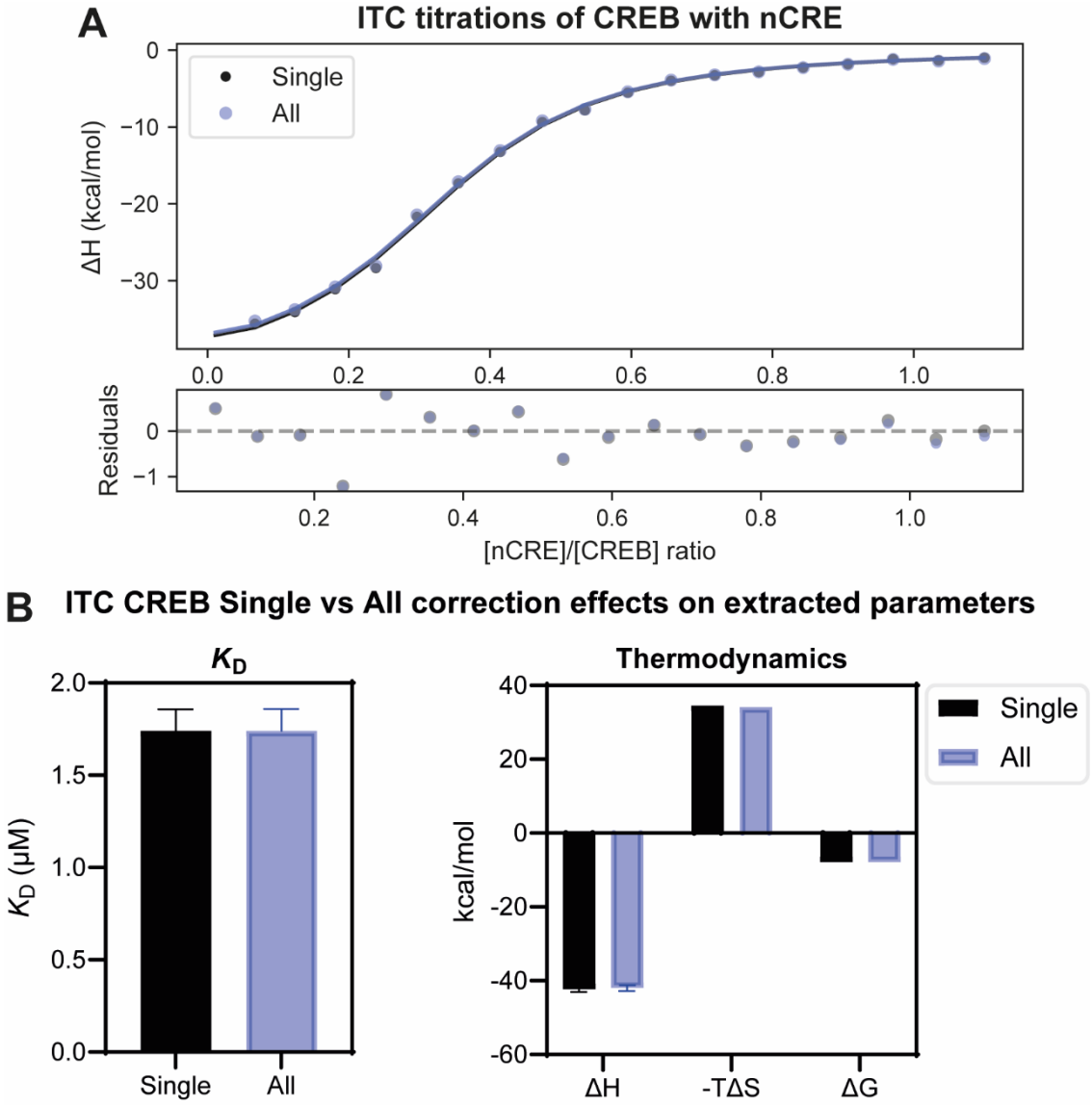


Figure 4.15: There is almost no difference between using the "Single" or "All" corrections for ITC. **A.** Normalised heat from ITC titrations using 38 μM CREB in the cell, titrated with 224 μM nCRE at pH 7.4, 25 °C. The "All" correction (considering heat change from buffer-buffer, buffer-nCRE, CREB-buffer dilutions) and "Single" correction (just the buffer-nCRE dilution) curves overlap well. **B.** Correspondingly, all extracted parameters (K_D and thermodynamics) are almost identical. Error bars are fitting errors.

More puzzlingly, (and like for CRElin earlier), the binding stoichiometry is roughly 0.35, lower than the expected 0.5 (Figure 4.16C). This could mean that CREB trimers, not dimers, bind to nCRE. Given the near 1:1 ratios in ITC experiments and homodimeric X-ray crystal structures (135), this seems unlikely. It is possible that CREB concentrations are 30-50% off. This also seems unlikely since both have been estimated using established

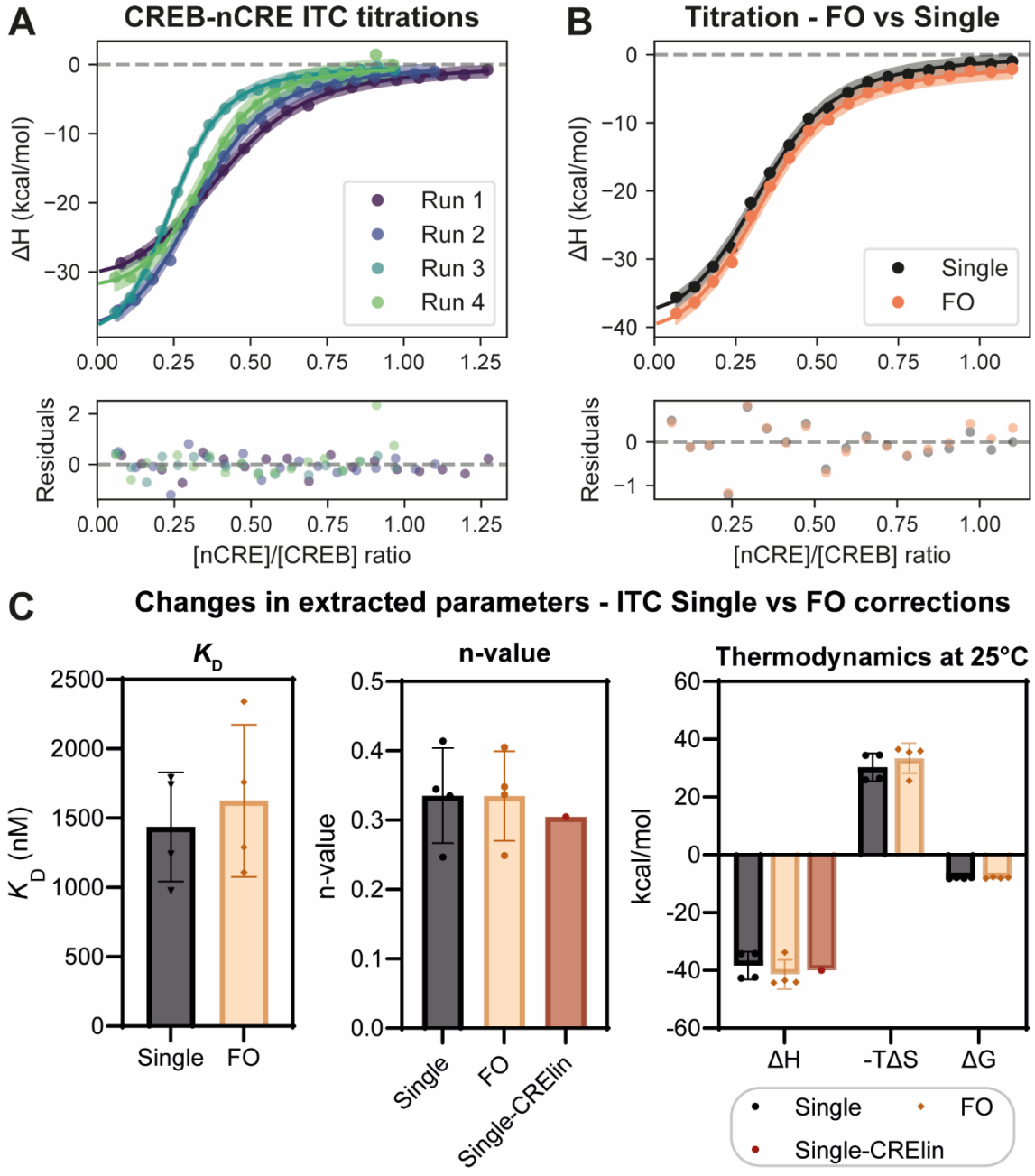


Figure 4.16: CREB ITC titrations with nCRE and comparison of "FO" and "Single" corrections. Conditions are pH 7.4, 25 °C throughout **A.** All ITC titrations of 30-40 µM CREB in the cell titrated with 180-240 µM nCRE, using normalised heat and the "Single" correction (dilution of nCRE in buffer). Titrations overlap and are well-fit by a single binding-site model. **B.** Representative ITC titrations with both "Single" and "FO" (fitted offset) corrections. Adding the "Single" correction noticeably changes the titration values. **C.** Although the variability between repeats is quite large, making them within error (standard deviation) between the correction types, the extracted parameters change value between FO and Single corrections – therefore the performing the buffer-nCRE run for the "Single" correction is recommended. Only the n-value is almost identical, which is also the case for CRE.

methodologies. Instead, tetramers (observed in Chapter 6, albeit in very small quantities) could contribute to dragging the stoichiometry down – binding DNA induces localised concentration increases that could stabilise higher-order oligomers.

Thermodynamic values for binding nCRE are: ΔG = (-8.0 ± 0.2) kcal/mol, ΔH = (-39 ± 5) kcal/mol, $-T\Delta S$ = (31 ± 5) kcal/mol, all at 25 °C (Figure 4.16C). Interestingly, the enthalpy of binding is almost the same as for CRElin (Figure 4.16C). This is surprisingly favourable binding for non-target and the enthalpic contribution suggests similar contacts are made between CREB and both target and non-target DNA. ΔG for CREh binding is lower than nCRE since it is directly linked to the K_D, which is lower for CRE than nCRE itself. As ΔH is similar for both, this suggests the entropy difference accounts for most of the thermodynamic difference instead. This may be somewhat surprising since NMR and CD studies have shown that CREB bZIP gains more additional helical structure on binding to target DNA than non-target DNA.

4.4. CREB's kinetics of binding to target DNA

4.4.1. CREB has fast, electrostatically directed binding to CREh.

The kinetics of CREB binding to CREh can be investigated simply through monitoring fluorescence intensity change using stopped-flow apparatus, since binding of the DNA causes a fluorescence intensity change (Figure 4.8B). Traces fit well to single exponential decay functions (Figure 4.17A) to give the observed k_{obs}. Assuming a two-state model: $k_{obs}=k_{app}+k_{off}$, with k_{off} being the dissociation rate constant. If a high enough CREB concentration is used, then it is considered constant throughout the experiment, and $k_{app}=k_{on}$[CREB], with k_{on} being the association rate constant. This threshold is 10-fold excess of CREB dimer to DNA. Hence, plotting k_{obs} against [CREB] should give a linear relationship with gradient k_{on} and y-intercept k_{off} which is observed (Figure 4.17B), although

this k_{off} extrapolation is often unreliable. As seen in the kinetic scheme in Figure 2.1, CREB kinetics are more complex due to the influence of dimerization, and this particular k_{on} is referred to as $k_{on,app}$ as it may encompass multiple processes.

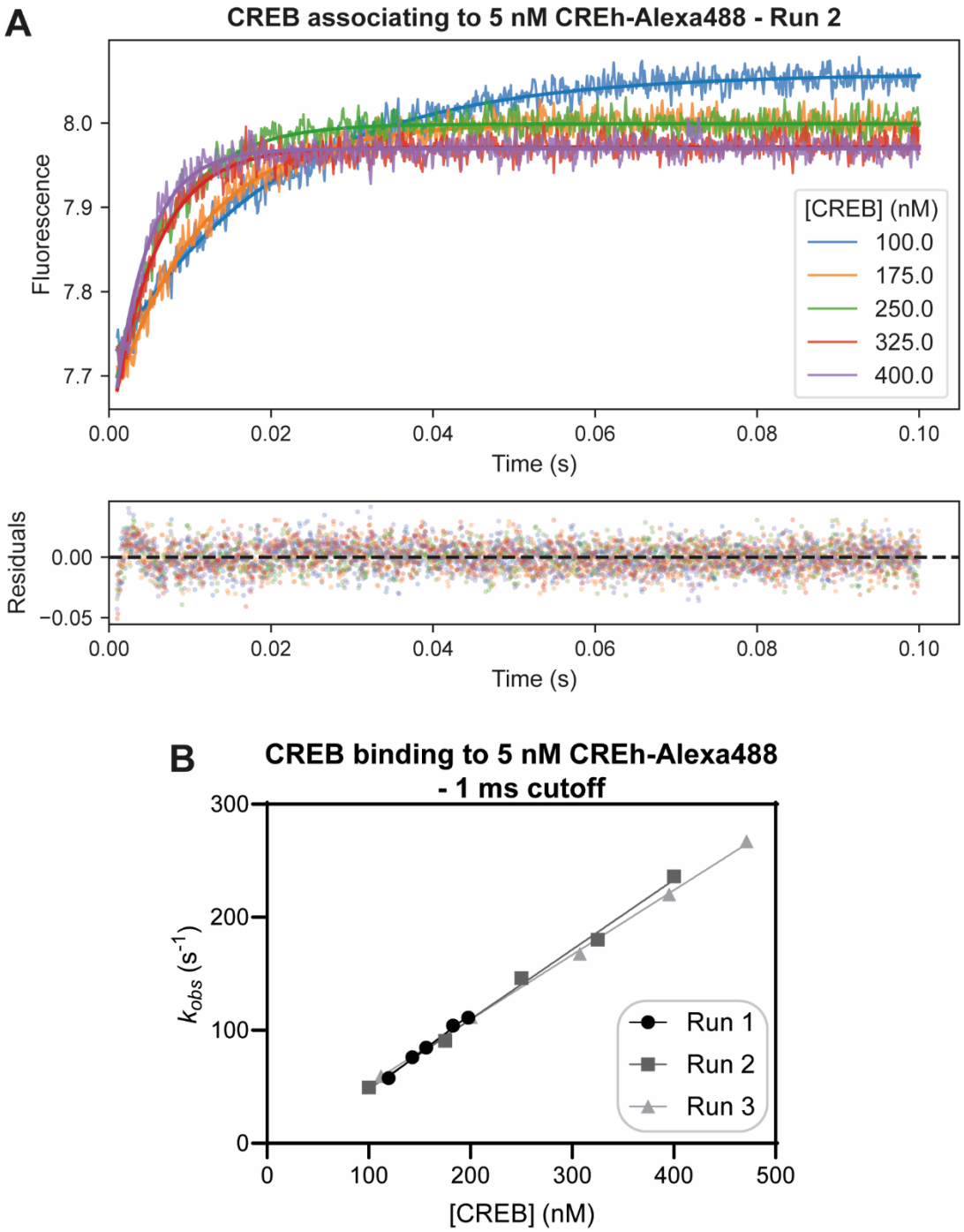


Figure 4.17: Association kinetics of CREB to CREh-Alexa Fluor™ 488. Conditions are 25 °C, pH 7.4. **A.** Example association kinetic traces of CREB to CREh. Single exponential decay equations fit the traces well, as shown by the residuals. **B.** k_{obs} extracted from the traces can be fit linearly which, according to a two-phase model, gives k_{on} from the gradient. This is repeatable across different runs and different concentration ranges. Individual k_{obs} fitting errors are small and not shown.

Deadtime and equilibration time for stopped-flow usually are in the 1-3 ms range. While this difference is not relevant for longer reactions, it becomes quite important for the very fast CREB association to DNA, since there are many datapoints in this range. Changing cutoff also changes k_{on} (Figure 4.18B), so an appropriate cutoff must be found. To investigate this, I analysed a dataset (the one with the highest concentration range) using various data cutoffs. Firstly, it is clear from the traces themselves that there is little signal change before 3 ms (Figure 4.18A). All traces are fit similarly well regardless of cutoff (not shown) and, while all cutoffs lead to $k_{on,app}$ in a similar range, around 0.6-0.7 $nM^{-1}s^{-1}$, $k_{on,app}$ is most reproducible with a cutoff of 2 ms, which also gives the highest $k_{on,app}$ gradient (Figure 4.18B).

k_{obs} are more linear with shorter cutoffs, with a gradual increase in y-intercept values, from negative towards 0 (Figure 4.18C,D). Under the two-state model, y-intercepts approximate k_{off} which must be positive. However, "looking for linearity" may be inappropriate if k_{obs} behaviour is, unexpectedly, not linear (suggested by 3.0 ms cutoff, Figure 4.18D). Hence, either the two-state model (and k_{obs} linear fit) is not strictly applicable, or lower cutoffs give a more "truthful" value. This y-intercept increase at low dead times is greater with CREB concentration increases (not shown), meaning more pre-dimerised CREB, so dimerization could provide a simple rationalisation for non-linear behaviour (see Chapter 6). Considering the amplitudes also suggests a 1 ms deadtime. For this dataset (and if ignoring the last trace, which has very small amplitude), the amplitude is only roughly constant with 1 ms cutoff. The experimental design never changes the amount of CREh available for binding, and therefore amplitude should be approximately constant.

These characteristics advise for a low cutoff due to fast binding of CREB to its target DNA, in spite of less reproducible values. At 1 ms cutoff, $k_{on,app} = (0.62 \pm 0.06)$ $nM^{-1}s^{-1}$ (from triplicate).

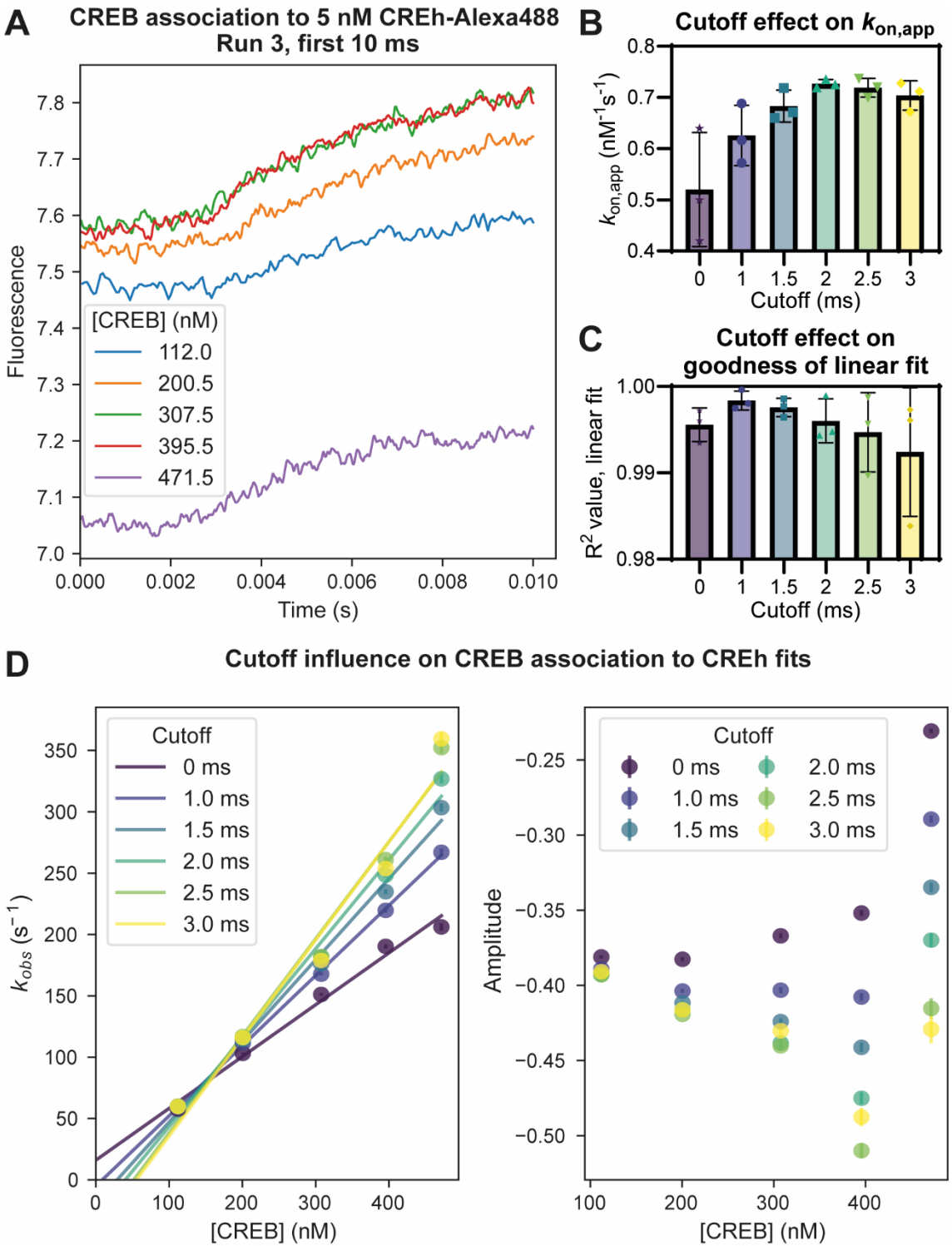


Figure 4.18: Investigation of deadtime cutoff influence on association kinetics. pH 7.4, 25 °C. **A.** Association traces only show signal change starting from around 3 ms. Notice the highest concentration trace has less amplitude. **B.** k_{on} is maximised and most reproducible with a 2 ms cutoff (standard deviation error). **C.** All cutoffs show a high goodness of linear k_{obs} fit (as per the R^2 value) but 1 ms cutoff is slightly better. **D.** k_{obs} and amplitudes from (A)'s traces, single exponential decay fits. Decreasing cutoff increases the y-intercept. At high cutoffs, the highest k_{obs} is either an outlier or suggests a linear relationship isn't suitable. Amplitudes are also affected by cutoff, with 1 ms showing the most consistency.

4.4.2. CREB has an apparent multiphase residence-time.

To investigate how long CREB stays bound to DNA, a pre-equilibrated CREB and CREh-Alexa488 solution was rapidly mixed with vast excess of unlabelled competitor CRElin (linear CRE). This ensures any dissociated CREB would not rebind to the fluorescent oligo, instead binding to the excess competitor. Fitting the resulting traces gives the dissociation rate constant k_{off}.

Single exponential decay fits do not fully explain all the variation in the traces, as shown by the residuals; double exponential decay functions prove to be a much better fit (Figure 4.19A,B). The two k_{off} obtained are (1.18 ± 0.06) s^{-1} and (0.32 ± 0.09) s^{-1} (Figure 4.21C). The faster phase tends to dominate, as shown by the amplitudes of the traces (Figure 4.19D). Similar rate constants were extracted by varying CREB rather than competitor concentrations (Figure 4.20A), but amplitudes were more variable, especially at higher CREB concentrations (Figure 4.20B). If fitted to a single exponential decay function, extracted rate constants are similar to the fastest one (Figure 4.19C).

Residence times is also the only DNA binding biophysical parameter that appears different between CREh and CRElin currently. k_{off} from CRElin-Alexa488 (using CREh competitors), at (0.7 ± 0.1) s^{-1} and (0.125 ± 0.003) s^{-1}, are lower than CREh values (Figure 4.21C,D). Importantly, the dominant k_{off} is now the lower one with regards to the amplitude (Figure 4.21A,B). This could show that nearby DNA secondary structure may destabilise CREB:DNA complexes or affect how it dissociates from DNA.

Explanations for this double exponential decay behaviour are either: (i) there are two residence time phases, (ii) I am resolving a reaction intermediate, or (iii) one rate constant may be caused by experimental artefacts. Explanation (i) is possible as previous experiments

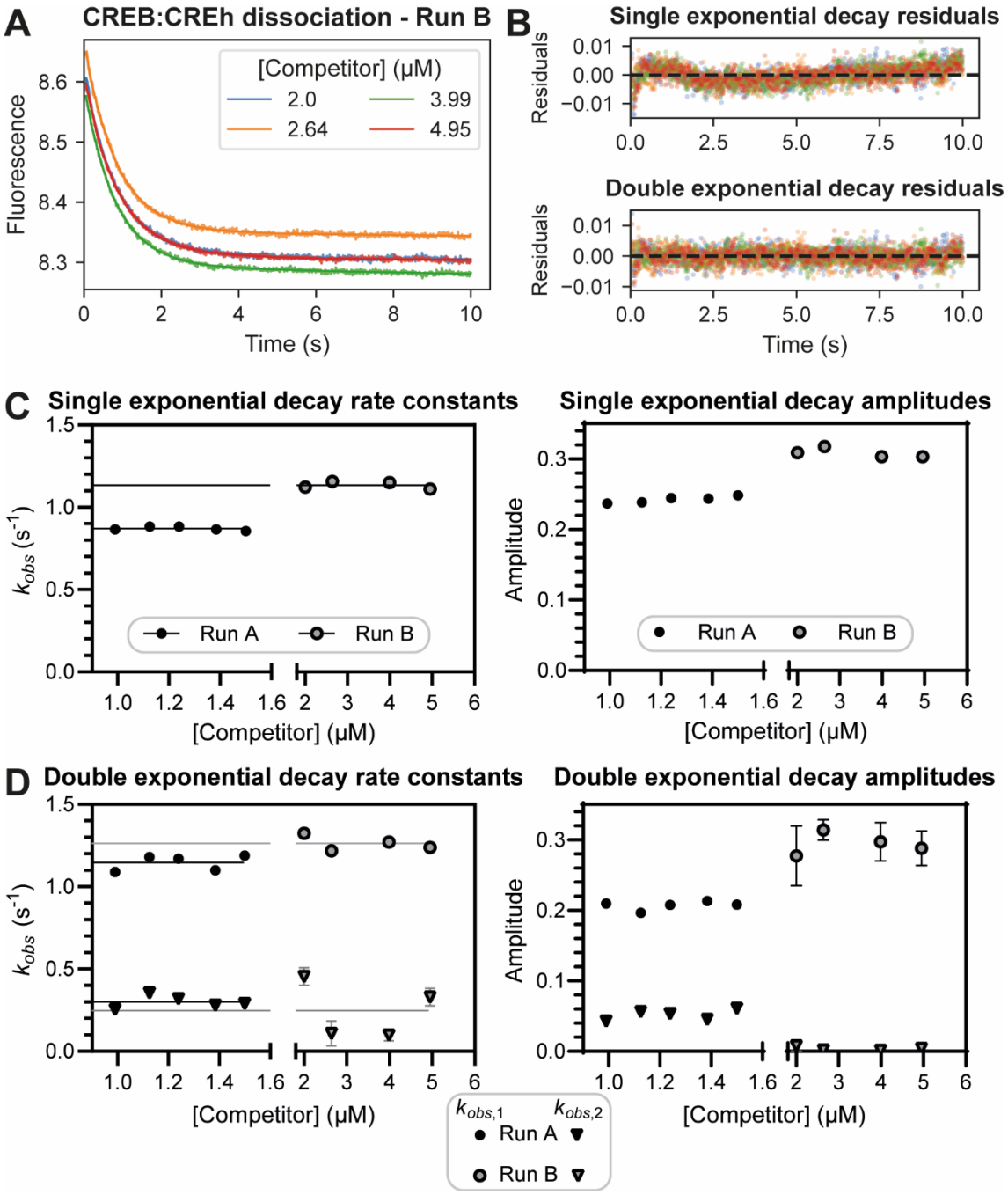


Figure 4.19: Dissociation kinetics of CREB from CREh follow a double-exponential decay model. pH 7.4, 25 °C. **A.** Example 20 nM CREB dissociation traces from 5 nM CREh-Alexa488 by CRElin competitor, fit by a double-exponential decay model. **B.** Residuals show variation only accounted for by double-exponential decay fits. **C.** Single exponential decay observed rate constants and amplitudes. k_{obs} are unvarying, as expected from an out-competition experiment with vast excess of competitor. Single exponential decay fits give k_{off} around the 1 s^{-1} range. **D.** Double exponential decay fitted rate constants and associated amplitudes. There is a fast and a slow phase, around 1.2 s^{-1} and 0.3s^{-1}. Amplitudes and rate constants are consistent between repeats, with the fast phase consistently having the higher amplitude. Note that the slow phase in the "B" experiment has more variation likely due to decreased amplitude, which may be due to the different experimental setup: "A" is a 1 to 1 dilution, "B" is a 1 in 11 dilution.

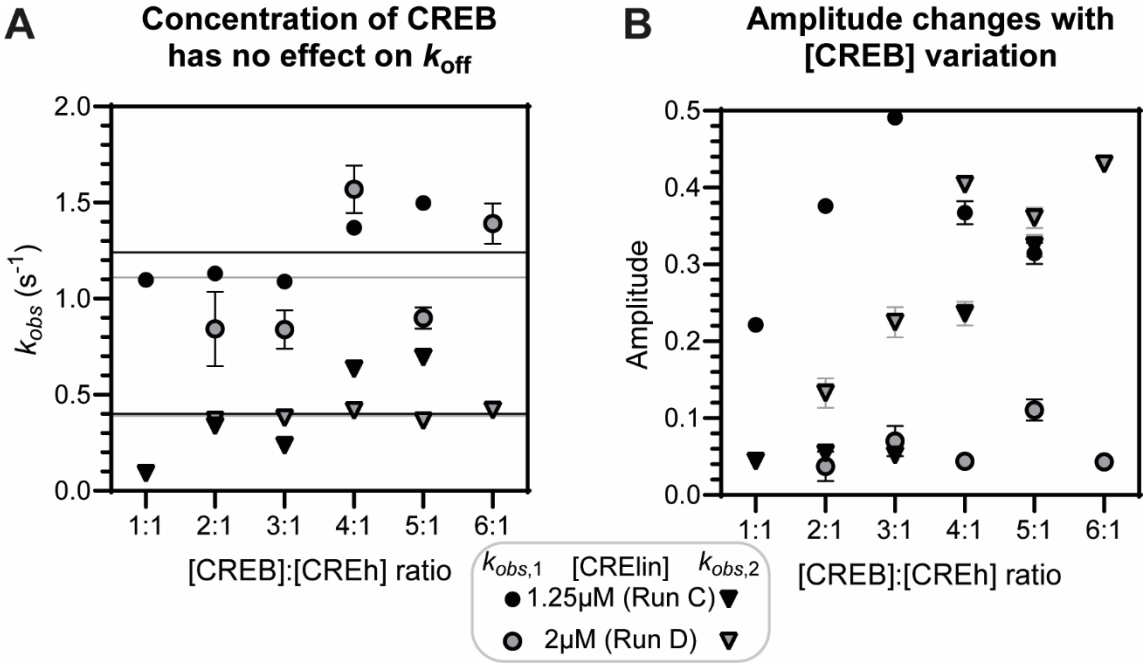


Figure 4.20: Starting CREB concentration has little influence on k_{off}. A. k_{obs} (from double-exponential decay fits of CREB dissociation from 5 nM CREh-Alexa488 using CRElin, pH 7.4, 25 °C) are consistent with those from Figure 4.19 at most [CREB], but there is more variability at higher [CREB]. **B.** Amplitude increases with increasing [CREB], likely because more DNA is initially bound. Interestingly, in run C, amplitudes appear to converge for each phase, whereas it is constant for the slow rate constant in run D.

using single molecule fluorescence microscopy have found that CREB has two residence phases – both *in-vitro* and in cells (206). However, one of the two rate constants is too high to be comparable here. Most illuminatingly, mixing of CREh-Alexa488 with CRElin (in the absence of protein) showed a signal change over timescales correlating with the lower rate constant (Figure 4.22C). Attempting to correct traces by subtracting raw traces with the control trace, then summing up all values with a constant to ensure positive fluorescence values, changes most traces to have apparent single exponential decay behaviour (Figure 4.23A). Extracted k_{off} are like the higher rate constant for CREh, although amplitudes are not constant and suggest a possible competitor-concentration dependence of the artefact (Figure 4.23C), possibly accounting for the correction's inability to capture the second phase fully in some experiments (Figure 4.23B). Hence, it appears explanation (iii) is correct: the double exponential decay behaviour of the dissociation process is caused by artefacts, and the rate constant with the highest amplitude is likely the "true" dissociation phase.

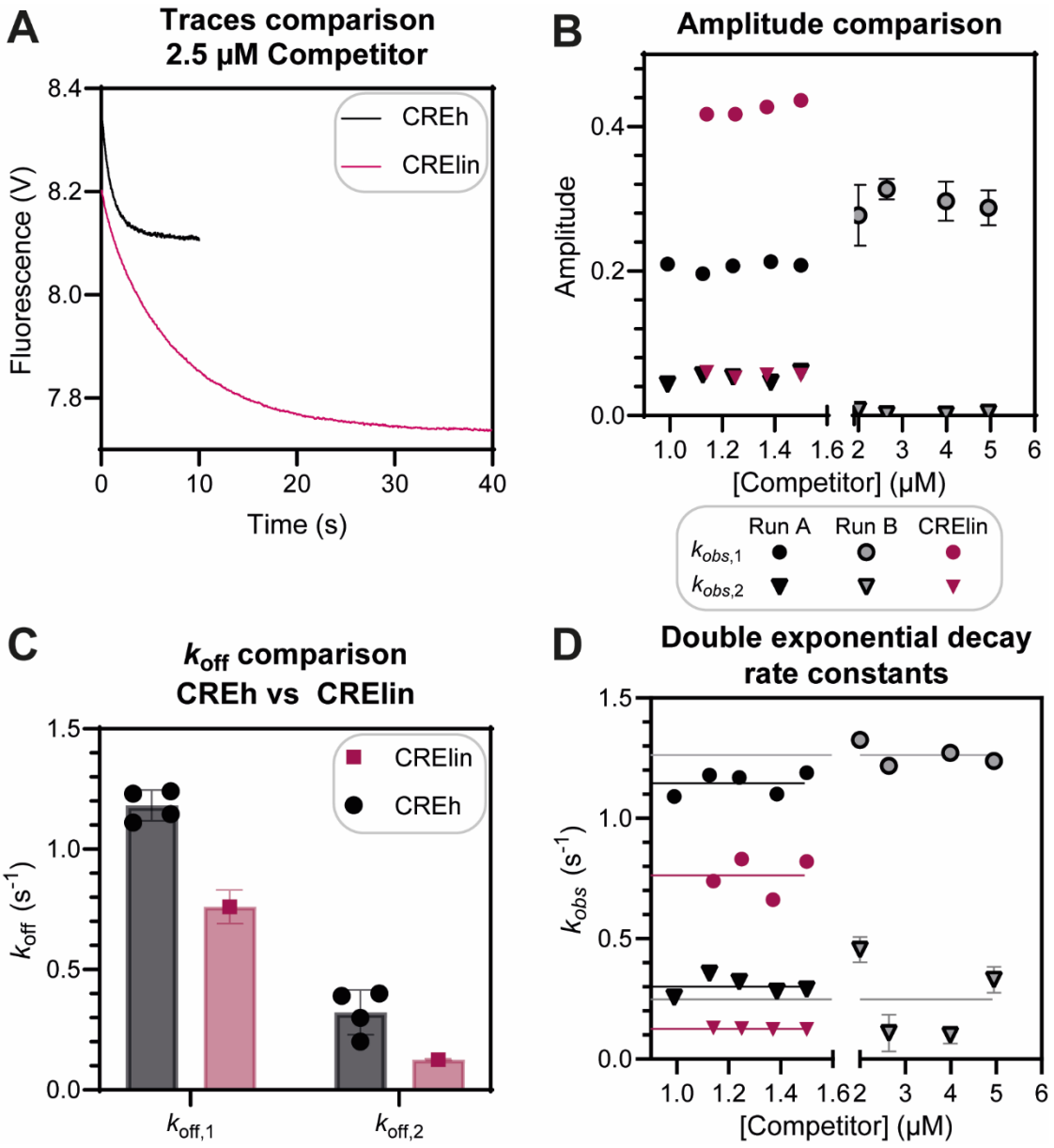


Figure 4.21: DNA structure affects dissociation of CREB from DNA. Conditions are pH 7.4, 25 °C throughout. **A.** Example of dissociation traces of 20nM CREB from 5 nM labelled CREh or CRElin, both oligos Alexa Fluor™ 488-labelled. CREB takes longer to dissociate from CRElin. **B.** Amplitude from the double-exponential decay fits. **C**. Both rate constants extracted using double exponential decay functions are smaller than those for CREh, but their importance is reversed – the slow phase is now more important and accounts for most of the amplitude of the trace. **D.** Individual rate constants extracted from the double exponential decay fits.

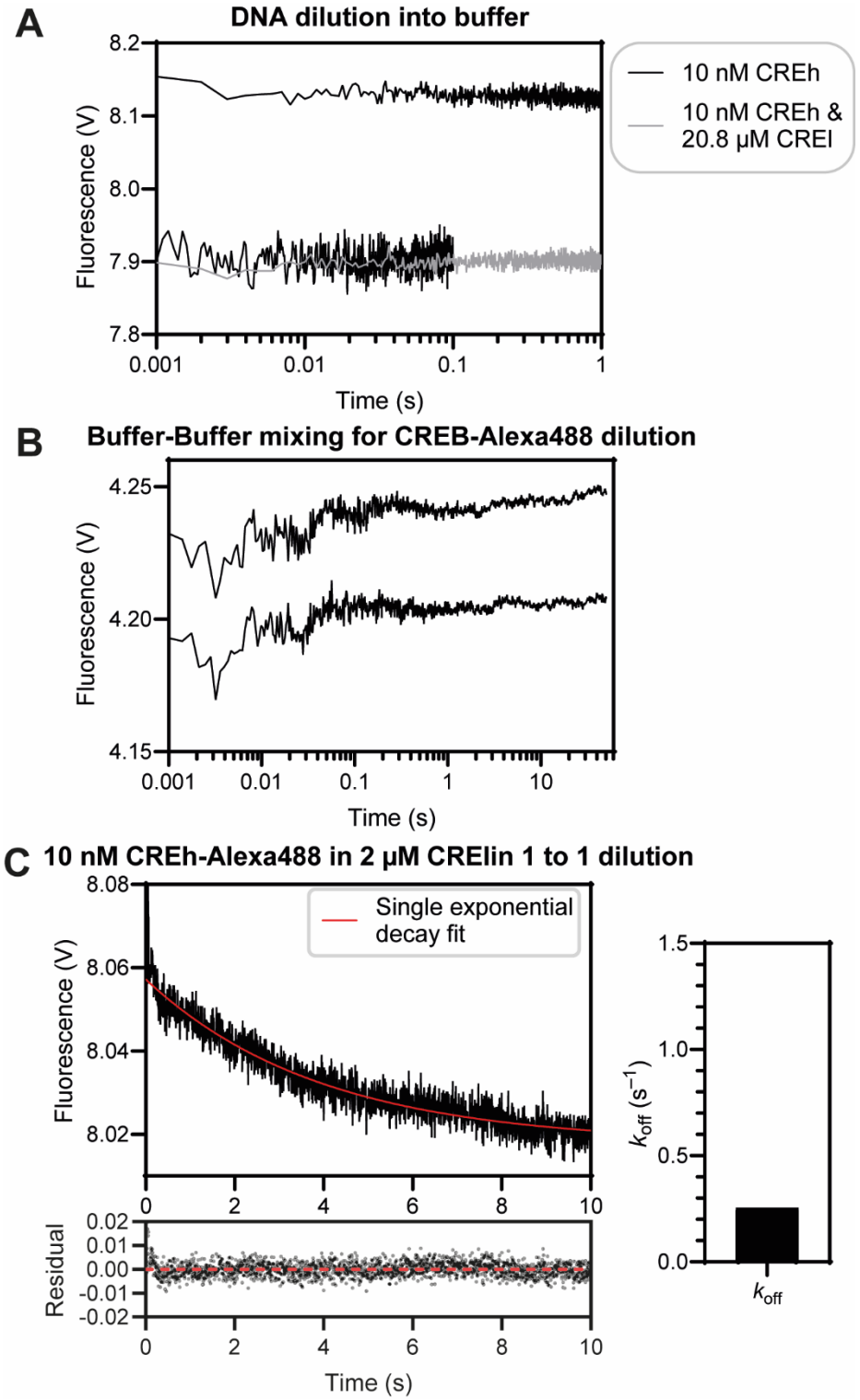


Figure 4.22: Control traces for stopped-flow dilutions. 25 °C, pH 7.4 throughout. **A.** Mixing labelled CREh-Alexa488, with or without unlabelled CRElin, into buffer has no signal change. **B.** Buffer-buffer mixing does shows minor change, likely due to some minute amounts of leftover protein in the system. The larger change around 0.01-0.1s from the 1 in 11 dilution setup also favours its removal in the analysis of traces in this setup. **C.** Mixing labelled CREh into unlabelled CRElin causes a signal change which, if fitted to a single exponential decay equation, extracts a low k_{off} comparable to one of those extracted from CREB dissociation experiments. This 2nd rate constant could therefore be an experimental artefact.

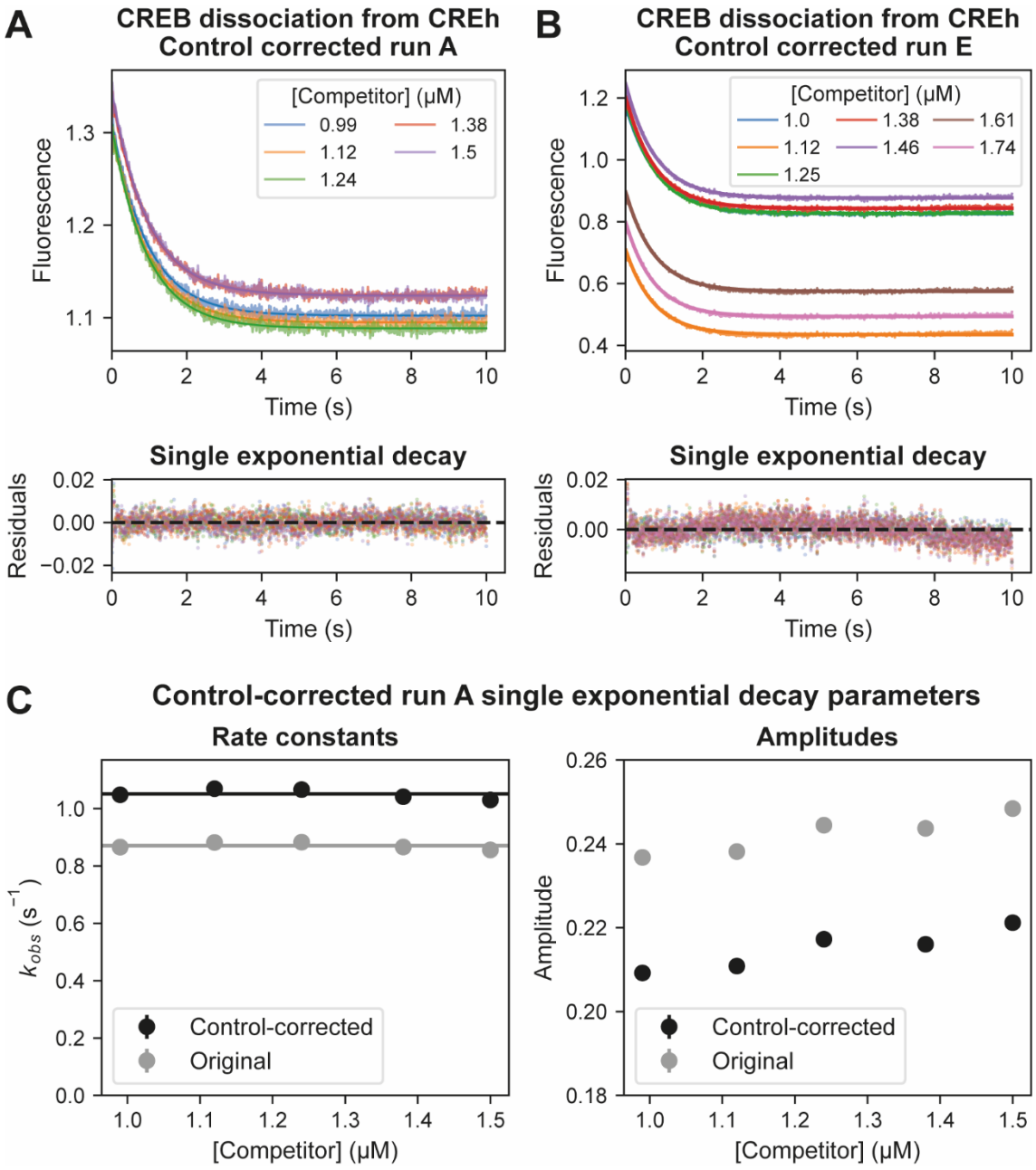


Figure 4.23: Dissociation experiments with attempted correction for experimental artefacts. A. Run A traces after subtraction with signal from the control CREh-Alexa488 mixed with CRElin alone trace are well-fitted by a single exponential decay function, as expected for a two-state model. **B.** Not all traces become well-fit by a single exponential decay function, as with this test run E (although run E's double exponential decay k_{off} are outliers and not included in previous analysis). **C.** Attempting to account for artefacts changes parameters, lowering amplitude and bringing single exponential decay rate constants closer to $k_{off,1}$ from earlier double exponential decay fits. However, amplitudes show dependence on competitor concentration, which suggest the need to account for artefacts at every competitor concentration.

Interestingly, a change in stopped-flow setup, from a 1 to 1 dilution to a 1 in 11 dilution (the CREB and CREh-Alexa488 mixture is diluted 11-fold), appears to limit the importance of the second rate constant (Figure 4.19B). One hypothesis is that the 1 to 1 setup creates an equally large perturbation upon mixing the two DNA mixtures, but that a 1 in 11 setup causes comparatively little change in one of the solutions – here, the CRElin one. This might limit the extent to which an equilibrated CRElin solution affects the signal. Unfortunately, this was attempted very late during this project, and there was no opportunity to test this further, but it could prove an easy solution to mitigate the artefact issue.

4.4.3. Kinetics and equilibrium-derived K_D differ, likely due to DNA-induced dimerization at low CREB concentrations.

Equilibrium dissociation constants K_D can be calculated using kinetics rate constants since, for a two-phase system like (theoretically) CREB binding to DNA, $K_D = k_{off}/k_{on}$. Doing this using the largest k_{off} gives kinetics $K_D = (1.9 \pm 0.2)$ nM for CREh, which is lower than the value obtained in equilibrium titrations. Yet, this does not constitute a discrepancy. As stated previously, dimerization likely plays a factor, as CREB concentrations below the homodimerization K_D are used (136,155,163). On the other hand, higher CREB concentrations are used in the kinetics experiment, so dimerization becomes less of a factor and CRE binding is probed specifically.

4.5. Influence of pH on CREB DNA binding

Prior to the establishment of the SUMO-CREB purification protocol (see Chapter 3), comparative binding experiments were performed at pH 6.5 (MES) and pH 7.4 (HEPES), as pH 6.5 was used to match experimental conditions in NMR studies. These data obtained with TEV-CREB-6His or TEV-CREB-TEV are worth considering briefly even though the

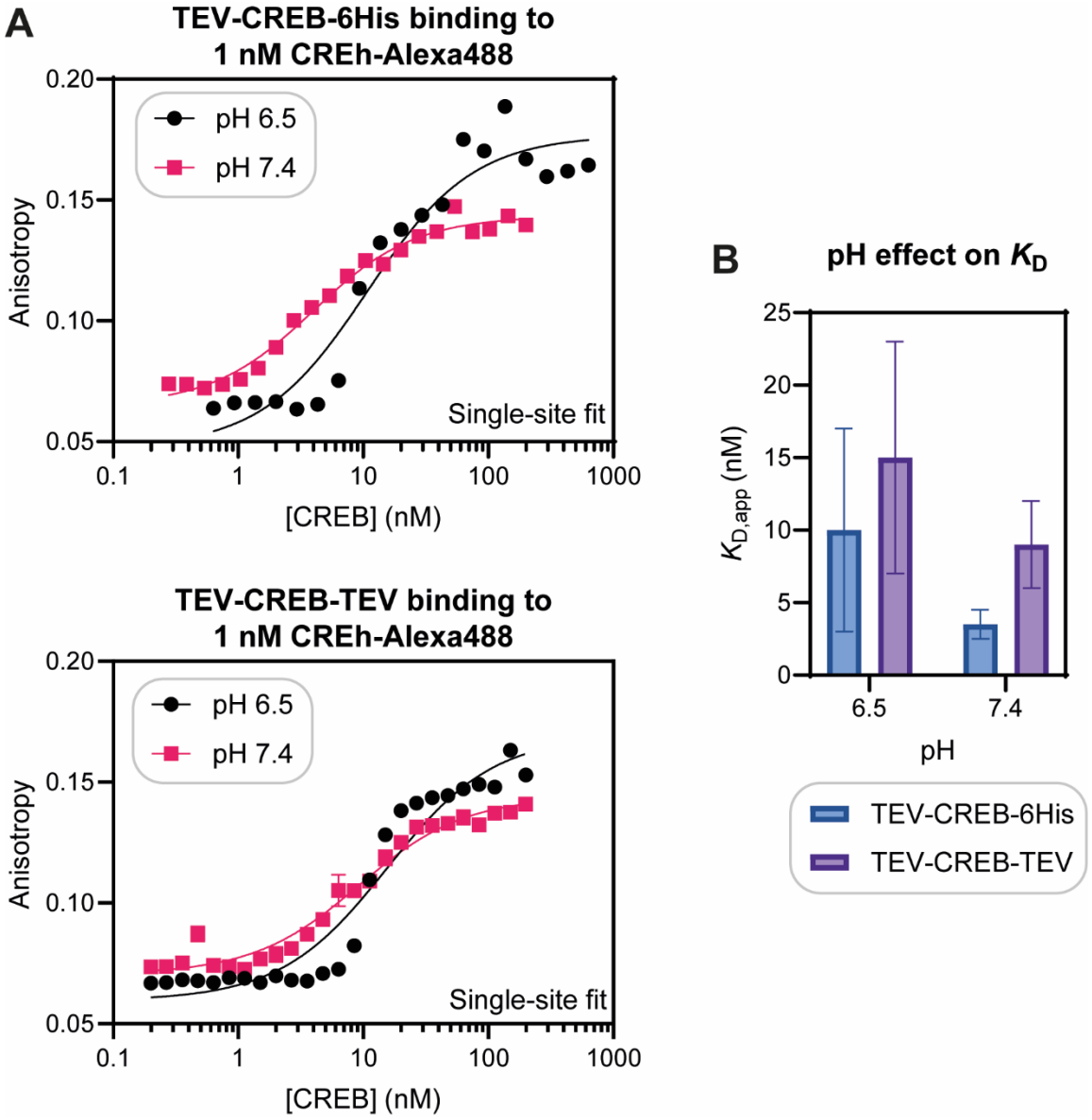


Figure 4.24: Effect of pH on anisotropy titrations. A. Titrations at pH 6.5 biophysical buffer have a very sharp shifted transition compared to pH 7.4 biophysical buffer titrations, all at 25 °C. **B.** In turn, the $K_{D,app}$ extracted is higher, but less reliable.

absolute biophysical values extracted from these are not as reliable (due to higher errors in concentrations, reduced purity and different tags or non-native sequences).

At pH 6.5 anisotropy titrations for CREh are much "sharper" in their transition than expected for a simple 2-state binding mechanism, and K_D extracted from them are unreliable (Figure 4.24). This could indicate some co-operativity in binding (e.g. multi-site binding with a Hill co-efficient). However, I have seen earlier the importance of dimerization in determining the forms of these curves – it may instead reflect a change in the

homodimerization affinity. k_{off} and k_{on} were respectively decreased and increased from pH 7.4 to 6.5, suggesting tighter binding overall (Figure 4.25). The pI of CREB is close to 5.5, so this pH change brings it much closer to its isoelectric point, perhaps explaining the drastic changes (Table 4.1). One hypothesis to explain this behaviour is that changing pH could in some way relieve CREB auto-inhibition of CRE binding. However, this effect may be less important in CREB than TEV-CREB-TEV or TEV-CREB-6His, as the tags and non-native sequences bring with them more charges and perhaps more reactivity to pH changes – despite minimal effect on pI (Table 4.1, calculated using ProtPI and ProtParam (213)).

A parallel conclusion from these short experiments is that changing C-terminal tags does seem to affect biophysical properties slightly, which suggests the absence of C-terminal tag in the study of CREB to be of paramount importance.

Protein	pI	pH	Charge
bZIP	9.99	6.5	7.23
		7.4	6.58
TEV-CREB-6His	5.63	6.5	-9.31
		7.4	-10.7
TEV-CREB-TEV	4.97	6.5	-9.31
		7.4	-10.7
TEV-CREB	5.1	6.5	-7.33
		7.4	-8.66
CREB with SUMO procedure overhang	5.19	6.5	-6.15
		7.4	-7.35
Pseudo WT	5.19	6.5	-6.3
		7.4	-7.62
Wild Type	5.19	6.5	-6.31
		7.4	-7.7

Table 4.1: List of constructs used with their isoelectric points and net charge at given pH. pI, was calculated using ProtParam and charges with ProtPI, using ExPASY-derived pK_a values.

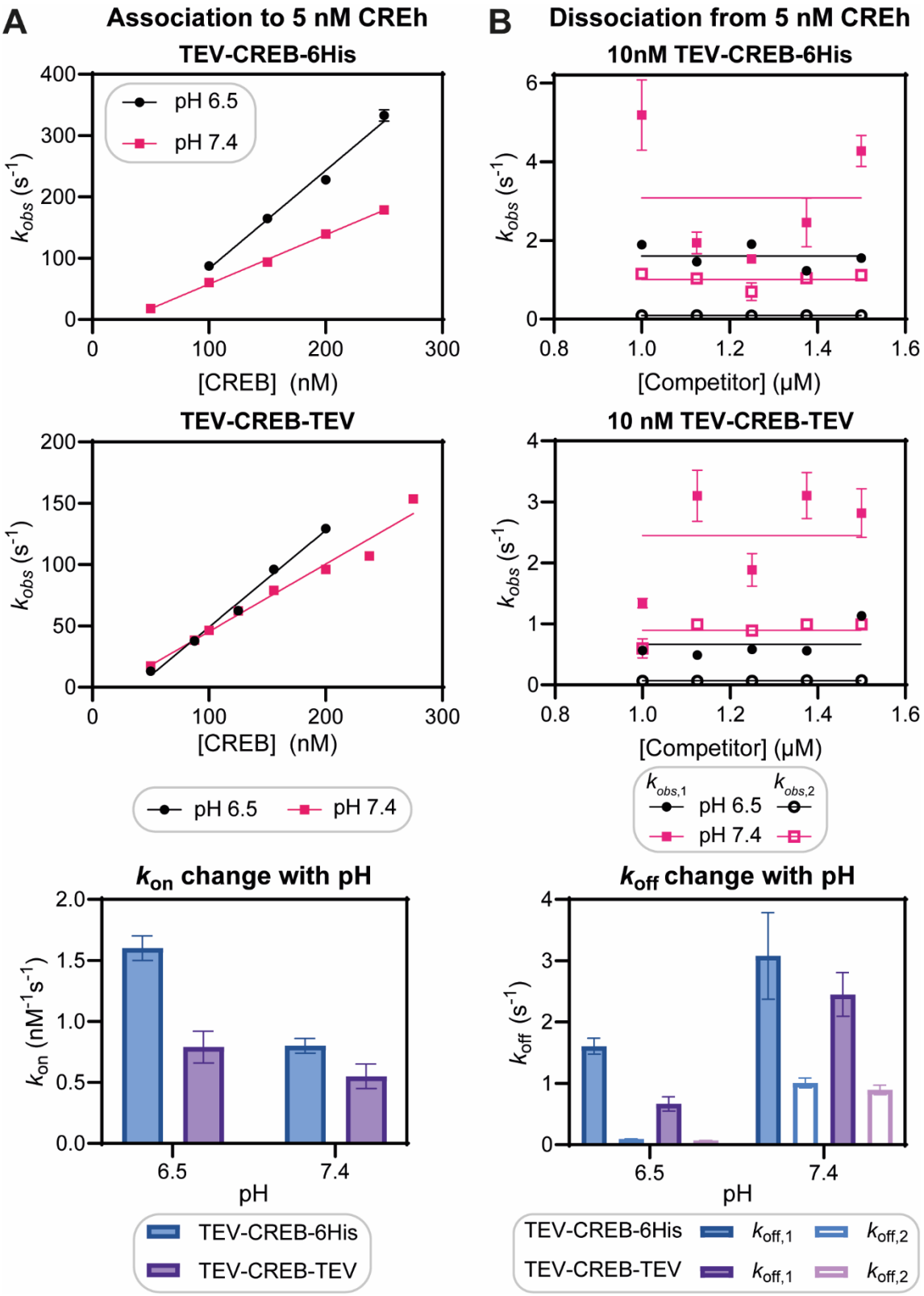


Figure 4.25: Effect of pH and C-terminal tag on CREB kinetics. All taken at 25 °C. **A.** Effect on association to CREh-Alexa488. k_{on} in biophysical buffer pH 6.5 is higher than in biophysical buffer pH 7.4, regardless of construct. It is also higher with a 6His than a TEV non-native extension. **B.** k_{off} from CREh-Alexa488 is similar, with tighter binding (decreased k_{off}) at pH 6.5 compared to 7.4, and for a CREB with 6His than a CREB with TEV extension.

4.6. Discussion

In this Chapter, I have presented a thorough biophysical characterisation of CREB and its DNA binding properties. Although the binding affinity to CRE DNA has been estimated previously (26,105,107,205), my kinetic approach uniquely accounts for dimerization and shows how it changes the $K_{D,app}$. Hence, K_D of CREB for its target is now estimated with greater confidence. A similarly low $K_{D,app}$ value was reported previously by fluorescence anisotropy (107) and, although similar to my kinetics K_D, was 2.5 times lower than my anisotropy-derived one. However, different buffer conditions, namely the presence of BSA protein competitors, along with altered detergent and salt concentrations, explain the discrepancy and could contribute to lowering this $K_{D,app}$ value. The CREB construct in (107) also had all of its natural cysteines which could affect the K_D too but, in this case, it would not have done so to the extent previously reported (105). Thus, the CREB cysteines have, at the most, a minor contribution to target DNA binding.

Furthermore, I have successfully recorded CREB association kinetics and (unlabelled) CREB dissociation kinetics. The association rate constant is extremely high, less than10-fold below the diffusion limit for k_{on} from the Smoluchowski limit for uncharged particles (214,215). Given typical requirements for orientation and activation energies this suggests strong electrostatic steering. This is not unheard of for both transcription factors and bZIPs, as the yeast GCN4 has a similar association rate constant (157).

Conversely, the apparent biphasic nature of CREB *in-vitro* dissociation is puzzling, as it does not conform to the two-state model assumption. CREB residency has been observed as biphasic from single-molecule fluorescence imaging experiments using linear DNA (206). Yet, one of the dissociation rate constants in this study is very high and could reflect interaction with "non-target" sequences or sliding, as the oligos used were around 20

nt, longer than those I used. The authors also observed two rate constants for a non-consensus containing DNA sequence, however the sequence appears to (inadvertently) contain a CREB half-site so this is difficult to interpret. Although my dissociation kinetics best fit double-exponential decay functions, it is likely one rate constant is due to experimental artefacts of DNA mixing, whilst the other reflects simple two-state model of dissociation. More accurate determination of the dissociation rate constant likely requires fully accounting for artefacts or changing setup entirely.

Interestingly, the other dissociation rate constant from this study was lower than the one I obtained by using labelled CREh, but was similar to the one from CRElin. This suggests the nearby DNA hairpin affects residency, although deconvolution from artefacts is needed to confirm this. This could be explored further too, for instance by moving the hairpin further from the binding site. This could give insight into how CREB dissociates from the DNA. It might slide off the single-stranded DNA of the hairpin for instance. Another interesting takeaway from this study is that CREB residency times are similar in cells and *in-vitro* (206), although CREB expression was unstimulated and triggers interaction with numerous coactivators (23,145,152,154), which could stabilise CREB at promoters and prolong residency times to levels observed in other studies (151).

Furthermore, I show that ITC is a viable method to study CREB:nCRE interaction. A CREB non-target K_D of a similar value has also been reported in a very recent publication (155). This is especially important, since how transcription factors bind to non-target DNA is not well known, even though non-target DNA interactions are likely far more prevalent than target ones. Although the binding affinity to non-target is almost 1000-fold higher than target, it is still relatively high and thermodynamically favoured – the enthalpic change is quite close to that for target suggesting similar interactions are formed upon binding. Due to

the vast excess of non-target DNA, chromatin structure is likely very important in limiting how much DNA is available to bind and for CREB to not get "lost".

Despite the viability of ITC data, the stoichiometry of binding it returns is puzzling, as an n-value of 0.5 is widely expected since CREB binds as a dimer. Yet, a value of 0.3 is consistently returned no matter if CRElin or nCRE is used. However, a similar behaviour was observed with the bZIP domain alone at pH 6.5 (163). Interestingly, reverse titrations with DNA in the cell and concentrated proteins in the syringe returned a n-value of 2, as expected (reversing the titration also reverses the stoichiometry). Hence, this discrepancy could be explained by the experimental setup. Dilution of DNA into an excess of CREB could somehow skew stoichiometry, perhaps by increasing local CREB concentration or allowing for more than one CREB molecule to be bound to DNA, whereas this would not occur by dilution of CREB into excess DNA, where there are plenty of partners to choose from. Another interpretation is that the n-value is a measure of not just stoichiometry but activity, and that a sizeable portion of CREB is "inactive" and binding-incapable under these experimental conditions. To control for potential concentration determination mistakes, repeats using oligonucleotides whose concentration is determined through other methods, like absorbance of a fluorescent tag, is a possibility. High c-value titrations, ideal for the investigation of n-value, are possible using CRElin oligos which, additionally, would inquire into the binding enthalpy further and verifying that both CRE and nCRE binding have similar binding enthalpy.

Another conclusion from this Chapter is that both CD and AlphaFold suggest CREB might form beta-sheets, which is unusual for a disordered protein. Confirming this, a recent NMR study has shown that parts of the CREB's TAD domains form small beta-hairpins (154). Interestingly, these are involved in the recruitment of the general transcription machinery, showing the importance of these structural motifs.

To finish, I also bring preliminary data which suggests both the pH and tags or short non-native sequences added to CREB might affect its biophysical properties, and this should be considered whenever working with this protein. For instance, NMR studies typically require a certain pH, and how this would affect CREB structure and interactions should be considered.

Chapter 5

Dissection of influence of non-DNA binding domains on DNA binding

The traditional architectural/modular understanding of transcription factors has understandably focused on the role of DNA-binding domains (DBDs) in DNA target search. However it has recently been shown that non-DNA binding disordered domains of transcription factors can guide them to binding sites in the genome, even without a DBD (68). Moreover, recent screens have unveiled the promiscuity of activation domains (AD) which suggests many sequences could activate or regulate binding, often within the factor itself, provided these have a balance of negatively charged, aromatic and hydrophobic residues (18,19). It is plausible that such negatively charged ADs could interact with the positively charged basic regions of DBDs and compete with DNA binding. Negative regions are known to accelerate association to DNA targets in the presence of decoys (212) so impeding binding to DNA in such a way could be beneficial and improve target-search efficiency. However, it is currently unclear how these disordered sequences contribute to the binding of DNA.

Besides its bZIP DBD, CREB contains a large, disordered transactivation domain (TAD) divided into three subdomains: the kinase inducible domain (KID) surrounded by two Q-rich domains, Q1 and Q2 (Figure 1.8). The TAD is known to increase basal CREB

expression, suggesting a role in DNA binding even in the absence of activators (149). KID is considered the more traditional and thereby more studied AD due to its S133 activating phosphorylation site. It has an expectedly high local composition of negative residues but it is unexpectedly depleted of aromatic residues, only containing a single tyrosine whereas ADs are typically enriched in aromatic residues, especially phenylalanine and tryptophan (18,19). Nevertheless, it does contain an aromatic binding motif that co-ordinates binding to the KIX domain of transcriptional coactivator CBP (23). The interaction between (p)KID and KIX is well studied structurally and mechanistically (22–24) and is arguably the most well-known example of an induced fit mechanism for coupled folding and binding. Furthermore, KID does interact with the bZIP DBD, although this is strongly mediated by post-translational modifications (26) and it is unclear if the KID can impact DNA binding on its own. It is possible that CBP binding to CREB could relieve KID-mediated inhibition and reinforce binding – or, reversely, that DNA binding could expose KID to the CBP.

The other two domains are glutamine (Q)-rich domains, often found in transcription factors (216). Q-rich domains work synergistically with more traditional AD and protein binding partners but are largely ineffective alone (216,217). These domains are involved in non-specific interaction, often with other Q-rich regions which can recruit the general transcription machinery (218) or aggregation which lead to disease (216,219–221). These are typical of multivalent fuzzy complexes interactions (76) which are relevant for the control of transcription (71,91). Indeed, both Q-rich domains can recruit the general transcription machinery through TFIID's own Q-rich regions, although Q1 less than Q2 (152,154), which may be why only Q2 deletion shortens CREB residence time at promoters (151). Other Q-rich domains have been shown to bind to CBP as well as phosphorylated KID (23,222–224) so these could facilitate CBP recruitment and, as Q-rich domains have been shown to participate in phase-separation (225), it is also possible that they coordinate

integration into phase-separated liquid-liquid condensates formed by CBP (226). This is especially relevant as phase-separation can both accelerate (57) or impede target search by preventing binding to chromatin (227), could impact the local concentrations of binding partners and can drive (81,82,88,228) or inhibit (91) transcription. The GCN4 bZIP can phase separate with Mediator (88), so CREB could behave similarly with CBP. Furthermore, both Q-rich domains have already been shown to impact CREB behaviour as they deregulate CREB transcription in *Drosophila* (229). Lastly, the general disordered nature and flexibility of these domains suggest they could help position the KID for self-inhibition or act as a flexible steric hindrance to dimerization or DNA binding themselves. Indeed, it was recently found that both Q1 and Q2 can bind to each other (154), perhaps indicating a role in activity modulation via intra- or intermolecular interactions with other CREB proteins.

In this Chapter, I will first bioinformatically describe CREB's domains and show that the Q-rich domains do not conform to typical activation domain architecture and are mostly disordered stretches of small uncharged amino acids. I then use my established binding assays to investigate the biophysical properties of the standalone CREB domains. Although the bZIP has been thoroughly explored by the laboratory at pH 6.5 to match conditions in NMR studies (163), I showed in the previous Chapter that pH can affect CREB and I recharacterized the bZIP at pH 7.4 to draw fair comparisons to full-length CREB. These show that isolated bZIP, unaffected by a pH change to 7.4, has increased binding affinities for both CRE and non-target DNA in comparison to CREB. Subsequently, I find that each of the TAD's subdomains contribute to impeding DNA binding. Interestingly, Q1 and Q2 have similar effects except on dissociation. Therefore, I suggest that the surrounding, and rather understudied, disordered domains of CREB can thereby play at least one functional role in DNA target search. Finally, CREB exists as separate isoforms within the

cell and I show that CREB isoform A, despite a small potentially helical insertion (154) differs little in its *in-vitro* binding behaviour to CREB isoform B.

5.1. Sequence analysis of the individual CREB domains

CREBs's non-bZIP domains are predicted to be disordered (Chapter 4). However, this may not be fully representative, as AlphaFold (230,231) and CD spectra suggest beta-sheets are present. Q2 seems to interact with the bZIP and could compete with DNA for binding as well. Moreover, the KID forms helices upon binding partners and, like many other disordered proteins (232), contains residual levels of helicity in their absence (22,23).

I used the predictor fIDPnn (209) to examine the broad function of CREB's domains. KID is predicted to be mostly a protein binding region, whereas the glutamine-rich domains are not predicted to have any function with high confidence but contain discrete regions predicted to bind DNA and RNA (Figure 5.1A). Q1 has a protein-binding region, and Q2 is identified as a linker region. Since Q-rich regions are expected to be involved in protein binding (216,219–221), this is expected for Q1 but not Q2, especially because the latter appears to bind general transcription machinery factors (152) like other Q-rich regions (218).

CIDER (233) is another webserver which calculates the fraction of positively and negatively charged residues of a protein to predict its structural behaviour. It also calculates κ, a measure of charge patterning with a value of 1 indicating complete charge separation. CIDER shows the glutamine-rich domains are mainly uncharged, whereas KID and the bZIP are almost half composed of charged residues, with twice as much negative and positive residues respectively (Figure 5.1C). Perhaps surprisingly, the κ for CREB, bZIP and KID is around 0.1-0.2 (Figure 5.1B). This suggests a relatively well mixed repartition of charges throughout the domain but with areas of highly concentrated, unmixed charges, which is clearly observed when looking at sequence histograms from PEPinfo (234) (Figure 5.2A).

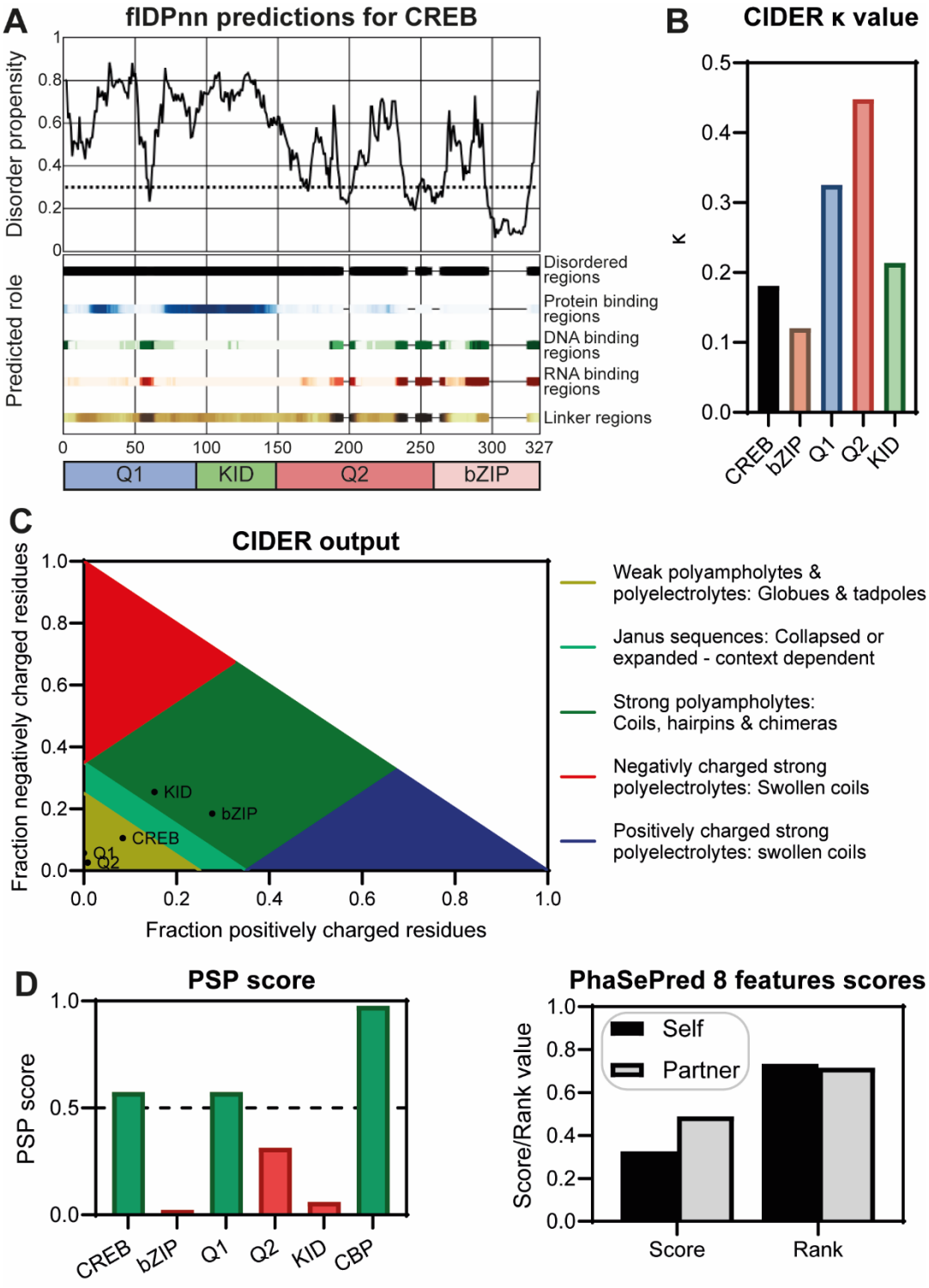


Figure 5.1: Bioinformatic analysis of CREB domains. A. fIDnn predicts CREB is mostly disordered. Non-bZIP domains are linker-like. KID is predicted to be protein-binding, Q domains to be nucleic-acid binding. **B.** CIDER-calculated Kappa values show CREB, bZIP and KID's charges are spread-out with pockets of packed, same-charge residues. **C.** CIDER predictions. Q domains have few charged residues, whilst KID and bZIP are strongly charged, making CREB's profile stand in-between but as a weak polyampholyte still. **D.** Phase-prediction algorithms suggest CREB could phase-separate but not with great confidence compared to CBP, and that partners could be important for this.

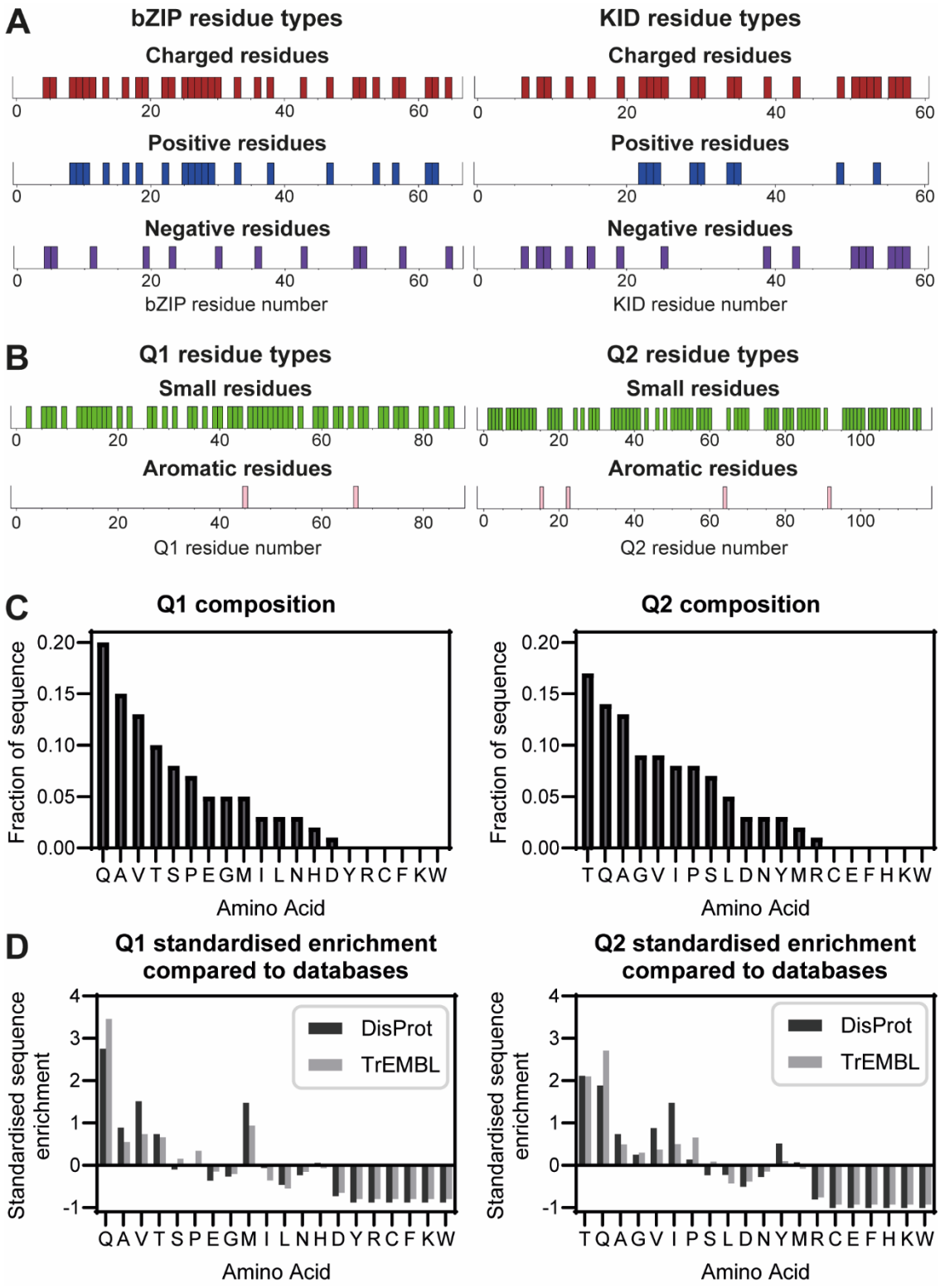


Figure 5.2: Sequence composition of CREB domains. A. As shown by figure 5.1, bZIP and KID are heavily charged, with overall spread-out charges but with regions of densely packed same-charge residues. **B.** Q-rich regions are mostly composed of small residues, with some aromatic residues which the other domains lack. **C.** Q, A, T, V are the main components of both Q-rich regions, but Q2 does not have Q as its most common amino acid. **D.** Q, V, A, T are also enriched in Q1 and Q2 compared to disordered (DisProt) and general (TrEMBL) proteins, with aromatic and charged residues depleted.

Conversely, the Q1 and Q2 domains are mostly of composed of small uncharged residues (Figures 5.1C, 5.2B), so their high κ value is not truly indicative since there are so few charged residues in the first place. Interestingly for Q-rich domains, neither Q1 nor Q2 have polyQ stretches (at least 4/7 Q residues) (Figure 5.3C) and Q2 only has Q as its second most common residue (Figure 5.2C). Q1 is 1/5th composed of Q (Figure 5.2C), and there are a series of QxxQ, QxQ and QQ motifs in both domains (Figure 5.3C). Additionally, polar and hydrophobic residues, mainly Q, A, T and V, are highly present in both domains and are also enriched compared to both disordered proteins from the DisProt database (235) and general proteins from the TrEMBL database (236), whereas aromatic and charged residues are depleted (Figure 5.2C,D). Proline, Isoleucine and tyrosine are slightly enriched in Q2 only, while methionine is slightly enriched in Q1 (Figure 5.2D). Despite serine being prevalent in both domains, it is not overrepresented compared to either database (Figure 5.2D). Although highly enriched in some amino acids, these regions do not pass the threshold to be considered low-complexity sequences by bioinformatic analysis (237).

As previously mentioned, phase-separation, in which Q-rich domains can participate, has the potential to regulate target-search. Both PhaSePred (238) and PSP (239) predictors do not suggest CREB does phase-separate with high confidence, especially when compared with phase-separating CBP (226) (Figure 5.1D). However, PhaSePred does score CREB higher for partner-dependent phase-separation and, since the CREB coactivator CBP does undergo phase-separation, CREB could integrate its granules. PSP also suggests that, if CREB does phase separate, Q1 likely plays a role (Figure 5.1D) which is quite unusual, as Q2 has more tyrosine (Figure 5.2C, 5.3C) which are very important for phase separation (227).

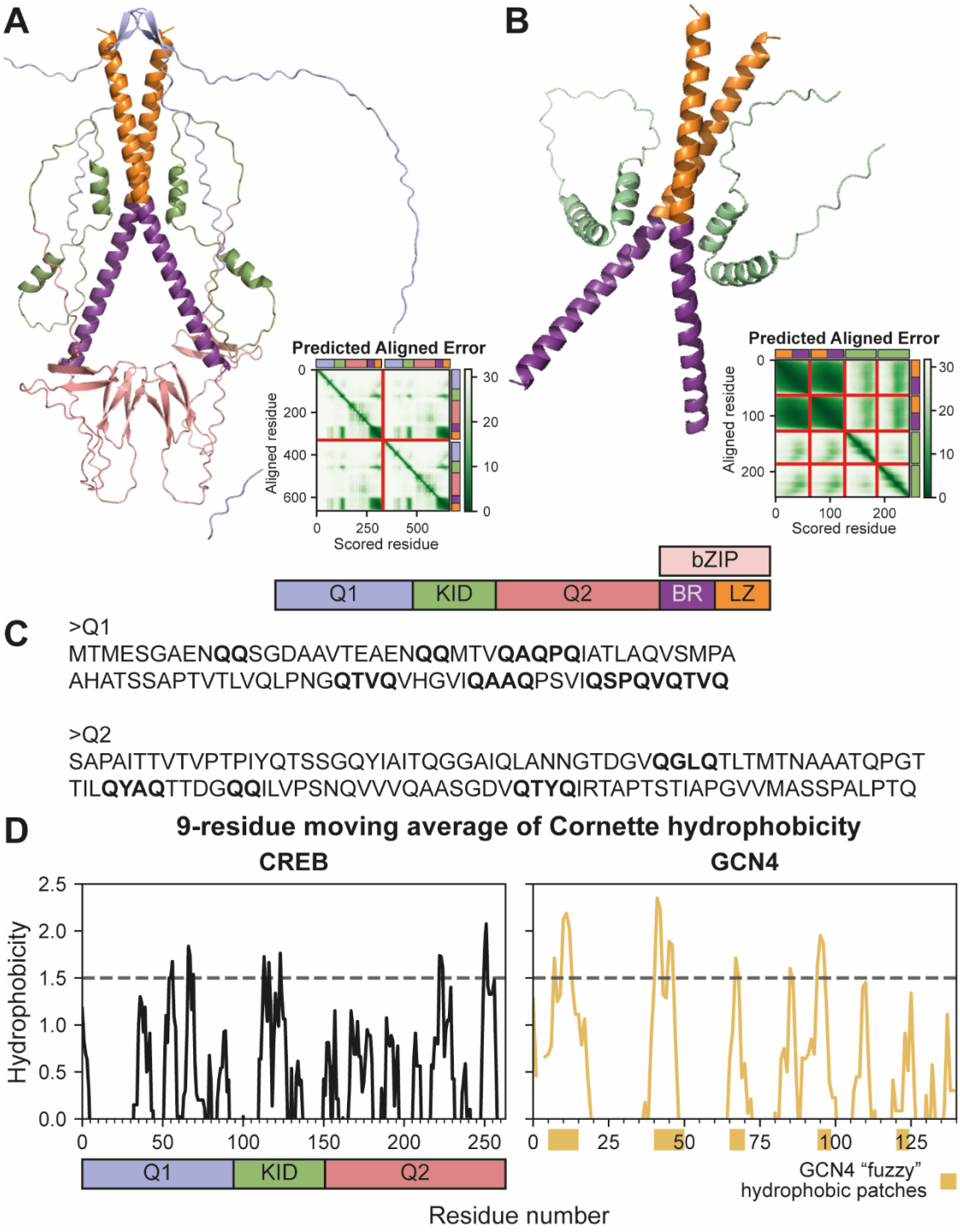


Figure 5.3: AlphaFold predictions of CREB self-interactions and sequence motifs/hydrophobicity of the Q domains. A. AlphaFold CREB dimer show Q2 β-sheets near the basic region, while KID interacts with the bZIP hinge, although with high predicted error. **B.** Prediction of bZIP and KID interaction alone shows a change in angle, with unchanged interacting regions, although with high predicted error for the KID. **C.** Q1 has more QQ, QxQ and QxxQ motifs than Q2 and neither have a polyQ patch (4+/7 consecutive Qs). **D.** 9-residue moving average of Cornette scale residue hydrophobicity for CREB (left) and Gcn4 (right). Hydrophobic Gcn4 patches involved in fuzzy interactions (as per (76, 243)) are highlighted, and tend to have hydrophobicity higher than 1.5, as per the dotted line. CREB's Q domains have similar regions of local hydrophobicity.

Predicting the structure of CREB dimers using AlphaFold and the KID:bZIP interaction suggest both Q2 and KID may interact with the bZIP (Figure 5.3A,B). However, the position of the KID in the prediction suggests that it does not impede DNA binding as it is positioned more towards the hinge between the leucine zipper and the basic region. Conversely, AlphaFold predicts Q2 to form structures (notably sheets) located near to identified DNA binding region (135,145), indicating a potential hindrance to DNA binding (Figure 5.3A). Although AlphaFold can predict folding upon binding structures for disordered proteins (240) (as seen here for the bZIP and the KID structure resembling KID bound by KIX), it appears this is a case where AlphaFold is not as reliable, as NMR studies have shown the KID can bind the bZIP but closer to the DNA binding interface rather than the hinge (26) and the Q-rich domains can bind each other, which is not predicted (154). However, this may also represent a heterogeneous ensemble, where both types of interactions are possible, and AlphaFold only solves some of them. Moreover, this same NMR study found that beta-hairpins are formed and important for binding TAF4 (154), confirming the presence of beta-sheets found by AlphaFold and CD in Chapter 4.

Lastly, the Q-rich domains appears to have regions of localised hydrophobicity, calculated using the well-performing (241) Cornette hydrophobicity scale (242), akin to Gcn4 activator domain regions that are involved in non-specific, additive “fuzzy” interactions (Figure 5.3D) (76,243). This would reinforce the role of Q-rich regions as non-specific binding partners potentially involved in phase-separation, both for CREB partner but also possibly for CREB itself.

Overall, while the KID is predicted to be protein binding and has negative patched capable of interacting with bZIP’s positive patches, structural prediction does not suggest it does. The role of the glutamine-rich domains is less clear from their sequence analysis due to their low charge, unremarkable composition, and absence of motifs, although these may

be important in phase-separation should it occur for CREB, and, structurally, Q2 is predicted to impede DNA binding.

5.2. Design of CREB constructs

To test the influence of KID domain on binding, it was replaced with a GS flexible linker of equal size. Theoretically, this tests the influence of the domain sequence and composition without significantly impacting size. This construct is referred from hereon as GS-KID. Unfortunately, cloning GS linkers of the length required to match the Q-rich domains is impossible in *E. coli*, so Q-rich domains were deleted instead. ΔQ mutants and the bZIP domain were purified following the GB1 procedure (or a previously derived bZIP-specific one, see methods). These therefore have the "TEV" N-terminal non-native sequence, unlike CREB and GS-KID, both from the SUMO method. However, the largest impact of this TEV extension was in apparently causing aggregation that complicated

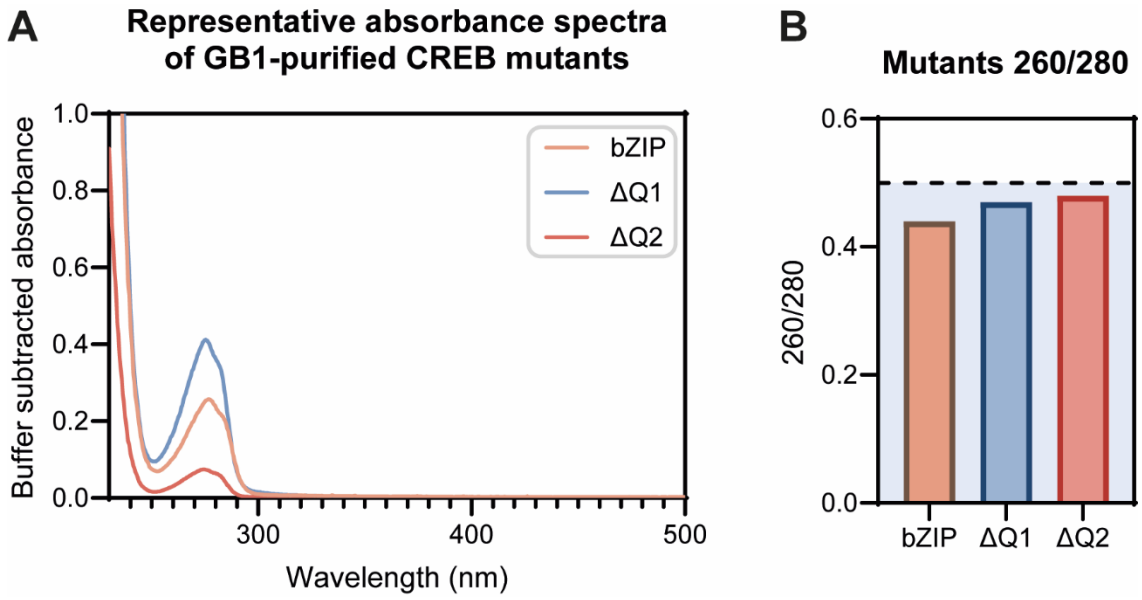


Figure 5.4: CREB deletion mutants do not suffer from aggregation issues even when using the GB1 method. A. Buffer subtracted absorbance spectra of CREB deletion mutants in biophysical buffer. There is no sign of aggregation at higher wavelengths. **B.** Uncorrected A260/A280 ratios of CREB deletion mutants are all under the acceptable purity threshold of 0.5, showing both a lack of nucleic acid and Rayleigh scattering contribution.

Protein	Length (with NTD linker)	pI	pH	Charge
bZIP	64	9.99	6.5	7.23
			7.4	6.58
CREB	332	5.19	6.5	-6.15
			7.4	-7.35
CREB-A	346	5.44	7.4	-5.36
ΔQ1	245	5.45	7.4	-3.73
ΔQ2	215	5.29	7.4	-6.65
GS-KID	332	6.37	7.4	-1.36

Table 5.1: Basic properties of CREB mutants. Length, pI and charge at different pH of CREB and deletion mutants. Charge was calculated using ProPI.

determination of CREB concentration. Rather interestingly, the bZIP and ΔQ constructs were not aggregation prone even with the TEV non-native sequence (Figure 5.4).

5.3. Investigation of bZIP DNA binding properties at pH 7.4

5.3.1. Binding equilibrium titrations show bZIP has similar binding affinity for CREh but increased affinity for nCRE compared to CREB.

I performed equilibrium binding studies with the isolated bZIP with the experimental methodology I developed in Chapter 4 for full-length CREB. I used the same labelled oligos, hairpin target CREh-Alexa488 and non-target nCRE-Alexa488, and unlabelled oligo nCRE for ITC, always using the same temperature and buffer. Experiments were performed on the same apparatus, and with matching experimental and analysis protocols.

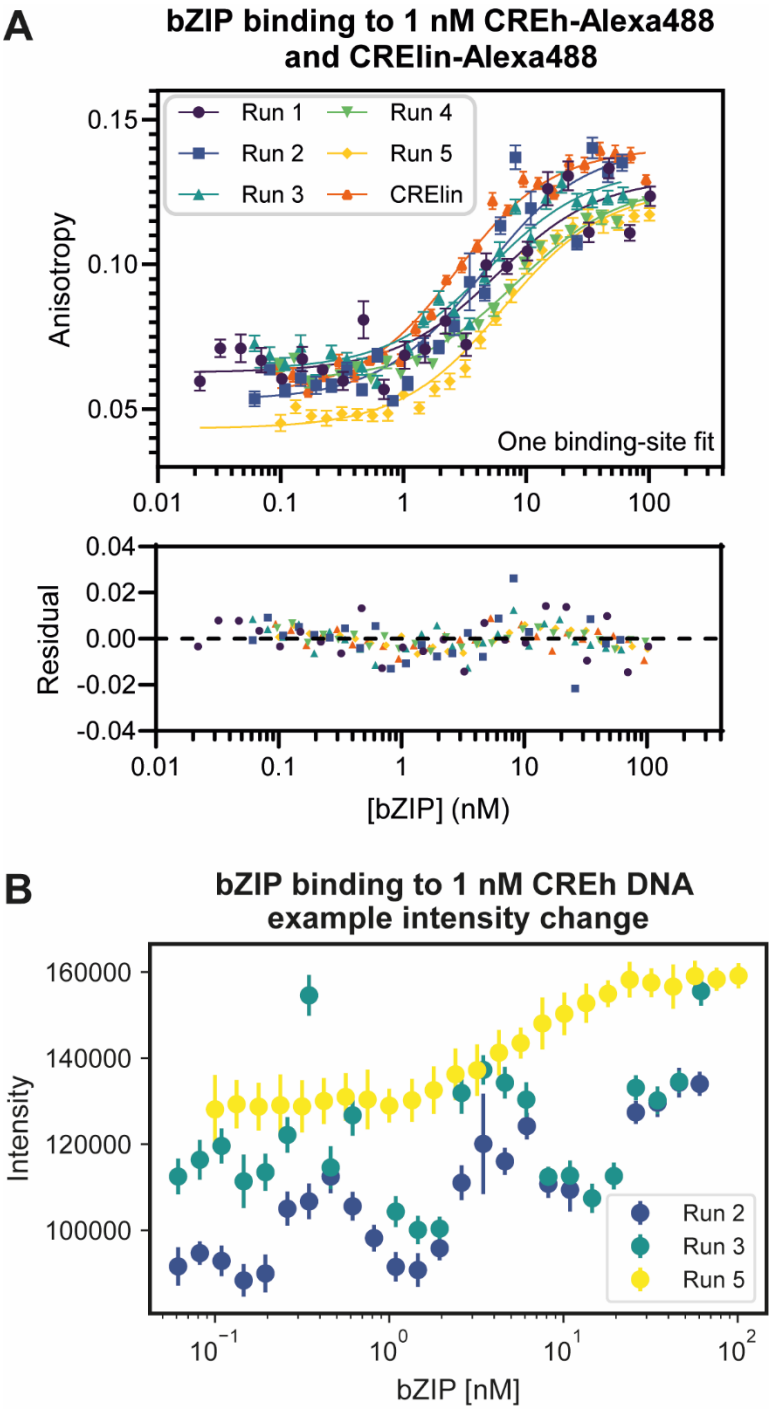


Figure 5.5: Anisotropy titrations of bZIP binding to labelled targets CREh and CRElin. A. Titrations of 1 nM target DNA by addition of bZIP at pH 7.4, 25 °C. All have a lower and upper plateau, with similar absolute values and transition range. There doesn't appear to be much difference between CRElin and CREh, although the CRElin transition happens slightly earlier and is slightly sharper. Residuals show deviation from the model around the titration range, presumably due to the influence of homodimerization **B.** Intensity values of bZIP binding CREh from selected runs. In all three, the data are not as usable as intensity values but there is an intensity value change between free and bound bZIP, requiring correction of anisotropy values.

CREB bZIP binding to labelled CREh caused a similar fluorescence intensity and anisotropy change as for CREB (Figure 5.5A), and anisotropy was corrected to account for the intensity change (Figure 5.5B). Fitting the corrected anisotropy titrations to a single binding-site model suggests a $K_{D,app}$ in the 5.5 ± 1.4 nM, similar to CREB (Figure 5.7E, Table 5.2). However (as described in the previous Chapter) this apparent equilibrium constant must include a significant contribution from dimerization since the homodimerization $K_{D,homo}$ is higher ((163) and Chapter 6). This is indicated by the slight deviation of residuals around the transition range (Figure 5.5A). Therefore, this number is likely unreliable, but it does show that both bZIP and CREB have high nanomolar affinity to CREh and a similar homodimerization range.

Just as for CREB, nCRE anisotropy titrations do not need to be intensity-corrected and are well-fit by a two binding-sites model with different affinities (as shown by the largest range titration, Figure 5.6A). However, the fitted binding affinity is 5-fold lower, with a $K_{D,nCRE}$ = (268 ± 41) nM for the first binding transition (Figure 5.7F, Table 5.4). The value is fully consistent with those from ITC titrations, with ITC $K_{D,nCRE}$ = (218 ± 82) nM obtained by adding concentrated nCRE into bZIP (Figure 5.6B, 5.7F, Table 5.4). Electromobility shift assay gels also show bZIP binding at much lower concentration to nCRE than CREB, whereas target binding is very similar (Figure 5.8). Thermodynamic values for bZIP binding to nCRE are ΔG = (-9.1 ± 0.3) kcal/mol, ΔH = (-30 ± 2) kcal/mol, -TΔS = (21 ± 2) kcal/mol at 25 °C, or ΔS = (70 ± 7) cal/mol/K (Table 5.5). The enthalpic contribution is lower than for CREB but the entropic penalty is decreased significantly for the bZIP, perhaps because it is much smaller (Figure 5.7D). Therefore, it appears that the CREB's non-DNA binding domains impede binding to non-target DNA around 4 to 5-fold.

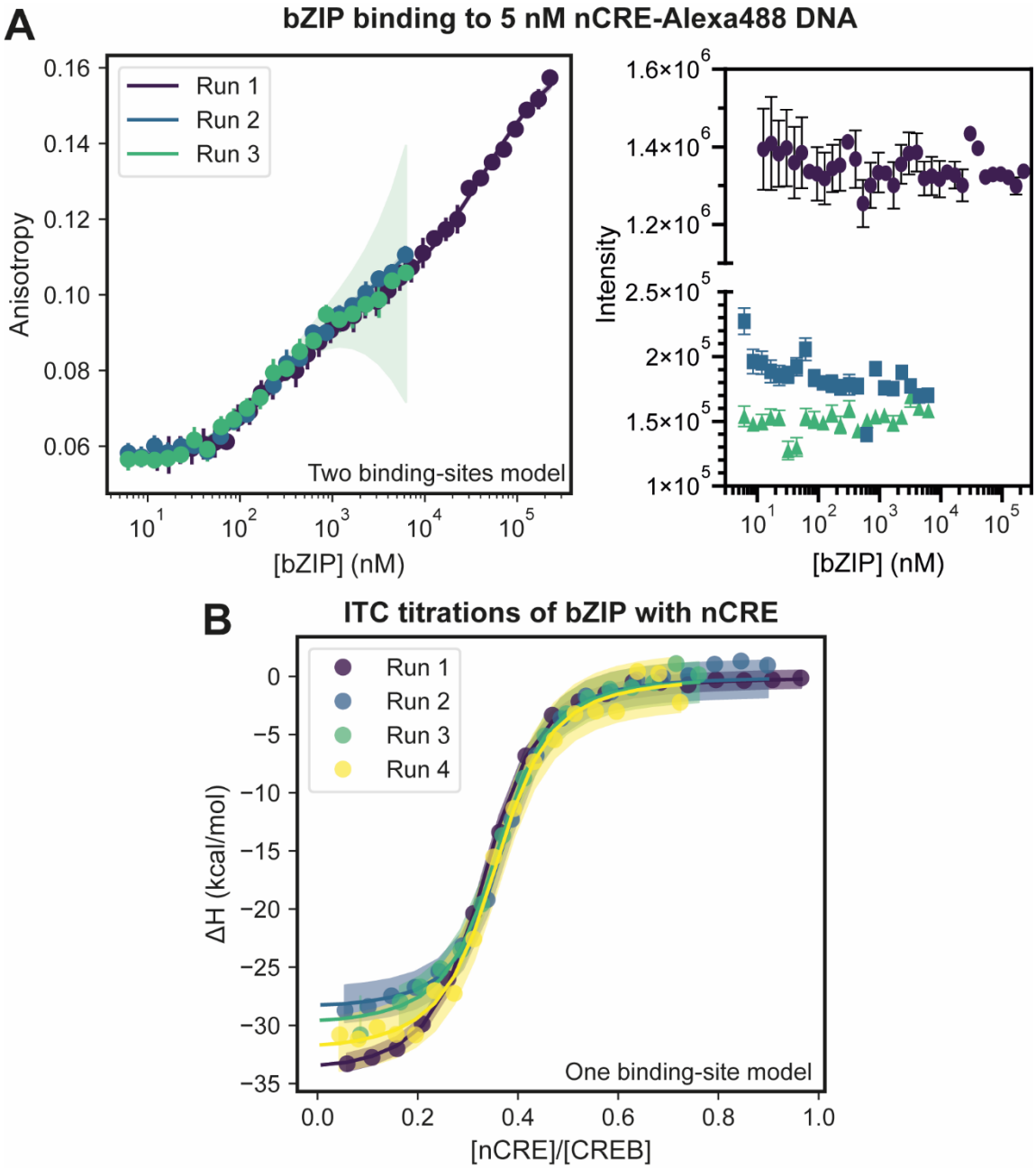


Figure 5.6: ITC and fluorescence anisotropy binding titrations of bZIP to nCRE A. Anisotropy titrations of bZIP binding to 5 nM nCRE-Alexa488 at pH 7.4, 25 °C, overlap well with each other and are fit by a two binding-sites model. Intensity values are constant, if decreasing slightly due to photobleaching, and anisotropy values thus do not need to be corrected. **B.** ITC titrations by addition of 150-200µM nCRE to 35-40µM bZIP (except Run D: 15µM [bZIP], 57µM [nCRE]) at pH 7.4, 25 °C. These overlap well with each other, especially around the transition area, and are well fit by a single-site binding model.

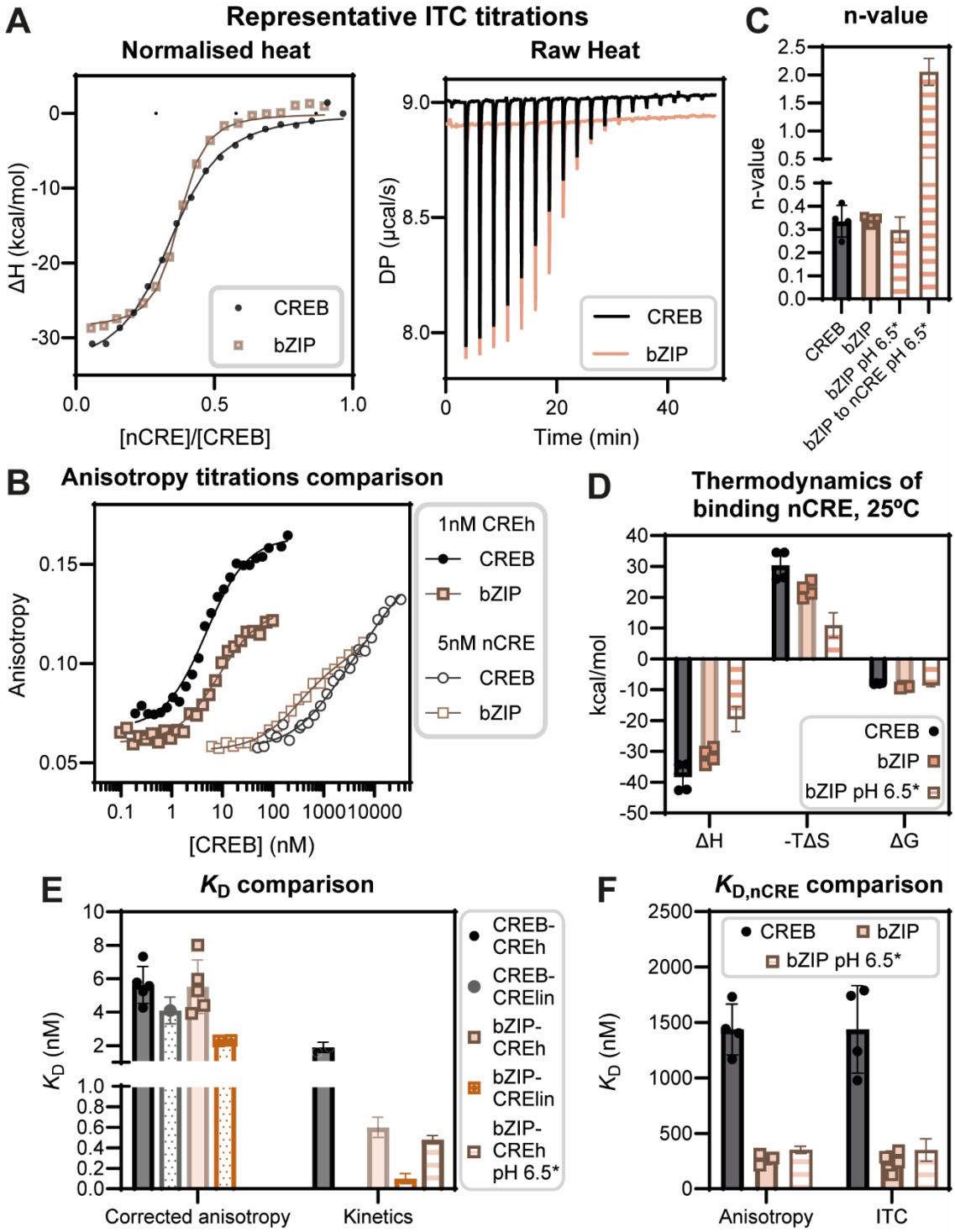


Figure 5.7: Comparison of bZIP and CREB's equilibrium binding properties to target and non-target DNA. Conditions are biophysical buffer pH 7.4, 25 °C unless stated otherwise. **A.** bZIP's nCRE ITC titrations have a steeper gradient, indicative of higher affinity. **B.** bZIP fluorescence anisotropy titrations have an earlier transition for nCRE than CREB. **C.** ITC n-values are below the expected 0.5. However, reverse titrations found the expected value. **D.** ΔG is slightly lower for bZIP than CREB. Both ΔH and ΔS are decreased for bZIP and at pH 6.5 compared to 7.4. **E.** Only more sensitive kinetic values show bZIP has increased affinity for CREh compared to CREB and for CRElin compared to CREh, unaffected between pH 6.5 and 7.4. **F.** bZIP has increased affinity for nCRE, unaffected between pH 6.5 and 7.4. *Data from Kuravsky and (163).

The ITC n-value did not change compared to CREB and was still around 0.35, when a value of 0.5 is expected (Figure 5.7C, Table 5.5). This was observed previously when nCRE was diluted into bZIP for ITC titrations (163) (Figure 5.7C). However, when doing a reverse titration (dilution of bZIP into nCRE), the n-value was 2, as expected (since the titration is reversed, the n-value is too) (Figure 5.7C). ITC n-values are indicative of stoichiometry or, more exactly, effective concentration ("activity") of the reagents. This change and deviation from expected value either means the concentrations used are slightly

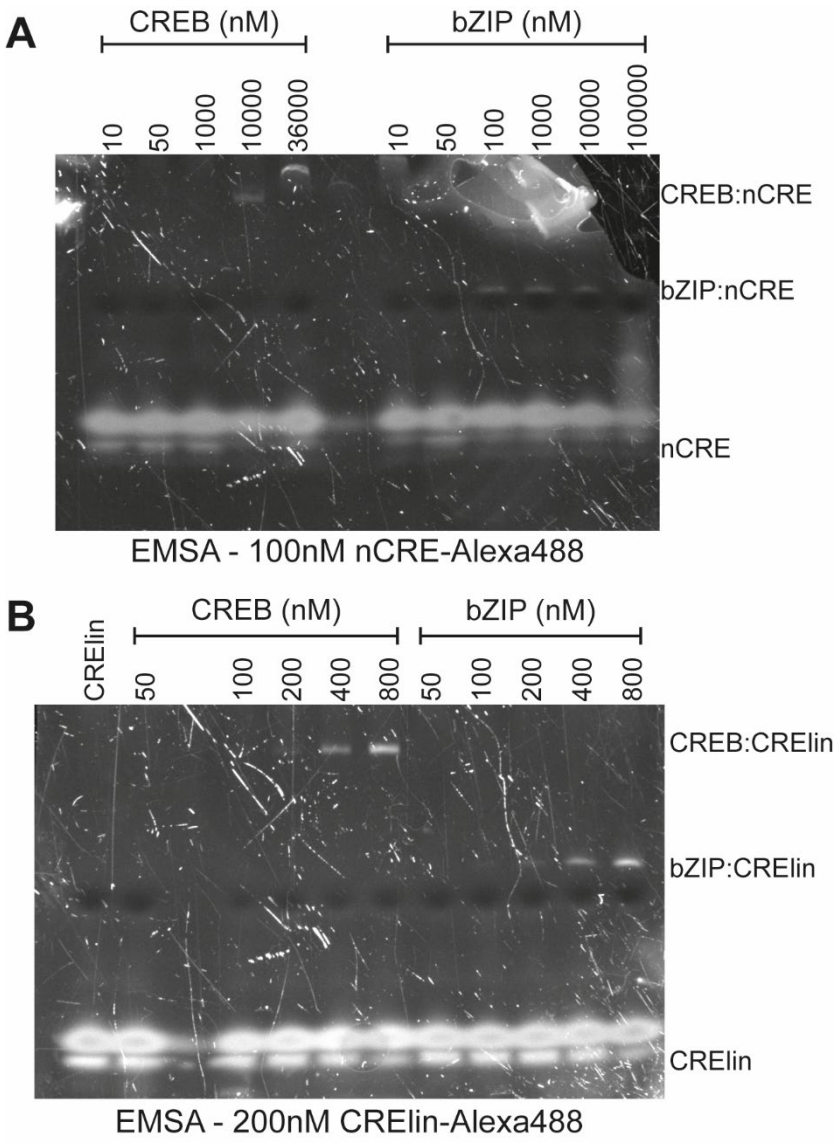


Figure 5.8: Electrophoretic mobility shift assays using fluorescent DNA oligos. pH 7.4, room temperature. **A.** bZIP:nCRE complex starts to appear around 100 nM bZIP whereas it starts appearing at 10 µM for CREB, showing greater affinity of bZIP for nCRE. **B.** Both CREB and bZIP start showing binding around 200 nM protein, showing that they have similar affinity for target DNA.

erroneous compared to those done when using a reverse titration, or effective concentration of CREB is altered under this experimental setup. This could be because the addition of DNA to a protein solution catalyses subsequent dimerization and binding, therefore artificially changing the n-value, whereas addition of a more concentrated and more dimerised CREB mixture to DNA stays dimerised and bound. Or, because there is less DNA than protein, more than one CREB dimer may attempt to bind DNA, skewing the n-value somewhat, whereas if the DNA pool is greater than the protein pool, this doesn't occur.

Lastly, there appears to be little difference in binding affinities measured for bZIP between pH 6.5 and 7.4, as most parameters measured are in the same range. Only thermodynamics appear to show pH dependence (Figure 5.7).

5.3.2. Further investigation of bZIP binding to target DNA in kinetics assays show a faster and longer binding of bZIP to target DNA than CREB.

As anisotropy titrations may not be indicative of true binding K_D, kinetics measurements can give more reliable values instead. I used stopped-flow to monitor association upon rapid mixing vast excess of bZIP with 5 nM labelled CREh, again with 1 ms cutoff and single exponential decay fits. bZIP associates roughly two-fold faster to target than CREB, with $k_{on} = (1.2 \pm 0.1)$ $nM^{-1}s^{-1}$ (Figure 5.9, Table 5.3). This also meant the limits of the apparatus (k_{obs} = 500 s^{-1}) were reached at a lower bZIP concentration than CREB, reducing the bZIP concentration working range. The k_{on} is virtually identical at pH 6.5 (163) and if labelled CRElin was used instead of CREh (Figure 5.9C,D). Once more I noticed that using higher concentrations decreased k_{on} gradients and increased y-intercepts (Figure 5.9A,C). A simple two-state model does not explain this behaviour or negative y-intercepts. At lower concentrations, a third state (free CREB monomer) is present in larger proportions, and this could offer an explanation.

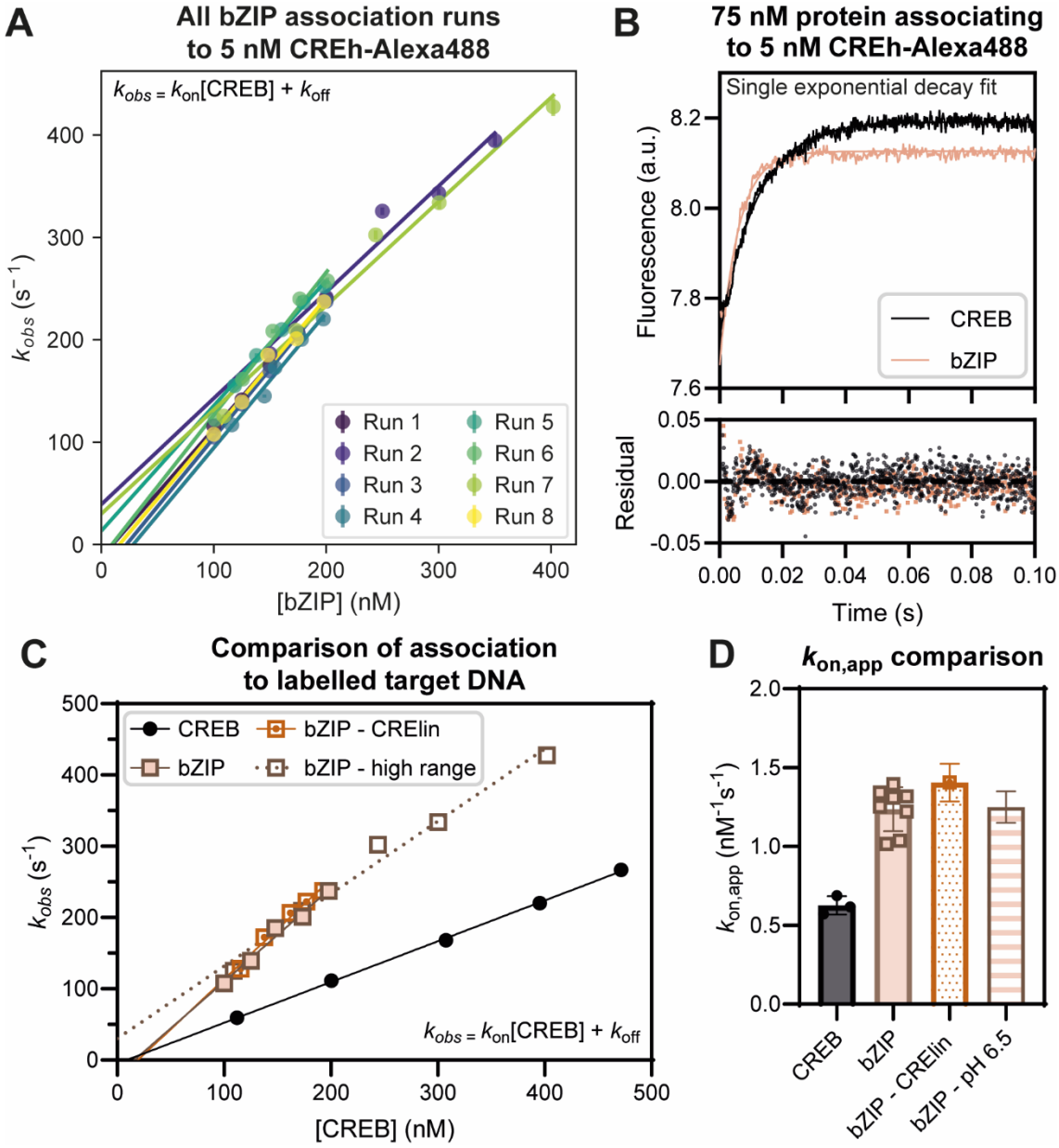


Figure 5.9: bZIP associates to CRE twice as fast as full length CREB. A. All association runs of bZIP binding to 5 nM CREh-Alexa488 show rate constants are reproducibly obtained at 25 °C, pH 7.4. Using higher concentrations tend to decrease gradients and increase y-intercepts. **B.** Raw trace examples of bZIP, which reaches the bound plateau faster than CREB. Residuals show traces are well fit by single exponential decay equations. **C.** bZIP observed association rate constants are similar whether binding to 5 nM CREh or CRElin and are greater than CREB's. Differences in gradients and y-intercepts from using greater bZIP concentrations are again observed here. **D.** Regardless of hairpin or pH, bZIP tends to associate twice as fast to target DNA than CREB. pH 6.5 values obtained in (163).

Dissociation rate constant was again measured by out-competition of bound 5 nM Alexa Fluor™ 488 labelled DNA with unlabelled target competitor. If CREh was labelled, CRElin was the competitor, and vice-versa. Traces were, unusually for a presumed two-state model, better fitted by double exponential decay than single exponential decay functions (Figure 5.12B). k_{off} was independent of starting bZIP concentration (Figure 5.10). Of the two k_{off}, one (typically with the highest amplitude) was consistent across experiments and the other fluctuated more (Figures 5.11, 5.12C). The high value, observed less often, was considered an outlier (Figure 5.12C). That one of the two rate constants was volatile shows perhaps that only one rate constant is the "true" one, and the other is an artefact of dilution (as discussed in Chapter 4). Only forced dimer (bZIP-GGC, see Chapter 6) k_{off} is available for the bZIP at pH 6.5, but it is very close to a non-forced dimer bZIP $k_{off,1}$ (Figure 5.12C).

Nevertheless, rate constants were lower than their CREB counterpart, showing that bZIP stays bound to target DNA for longer timescales (Figure 5.12C, Table 5.3). If monitoring bZIP dissociation from CRElin, both extracted k_{off} were the same (Figure 5.12C), but the amplitudes were reversed as with CREB (Figure 5.11C), showing that it is likely the smaller value is now the true k_{off} and that the hairpin disturbs binding to DNA.

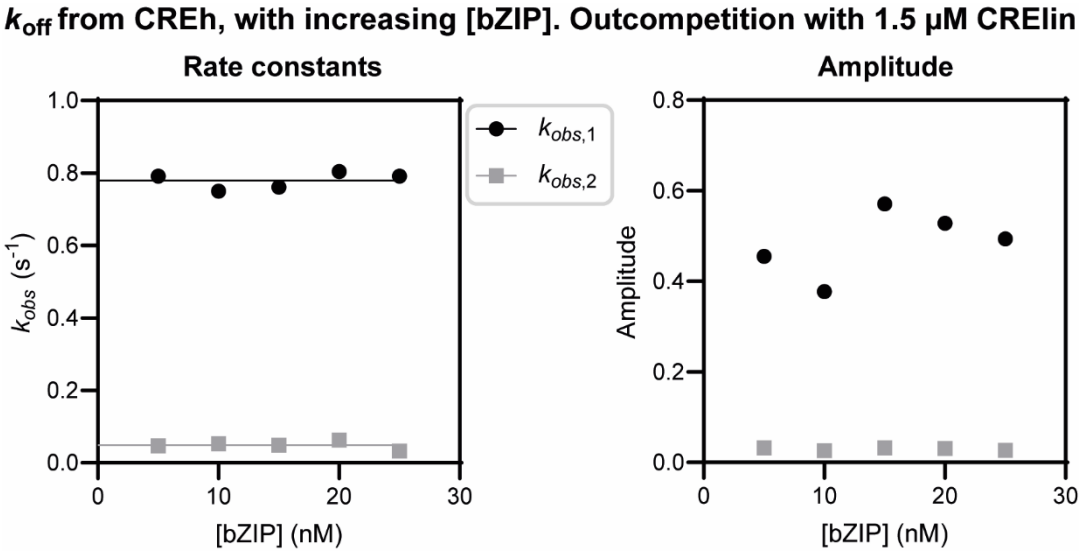


Figure 5.10: k_{off} is independent of starting bZIP concentration. Neither observed rate constant (left) nor amplitude (right) changes with increasing bZIP concentration in the cell when outcompeting from 5nM labelled CREh-Alexa488 with 1.5 µM CRElin at 25 °C, pH 7.4.

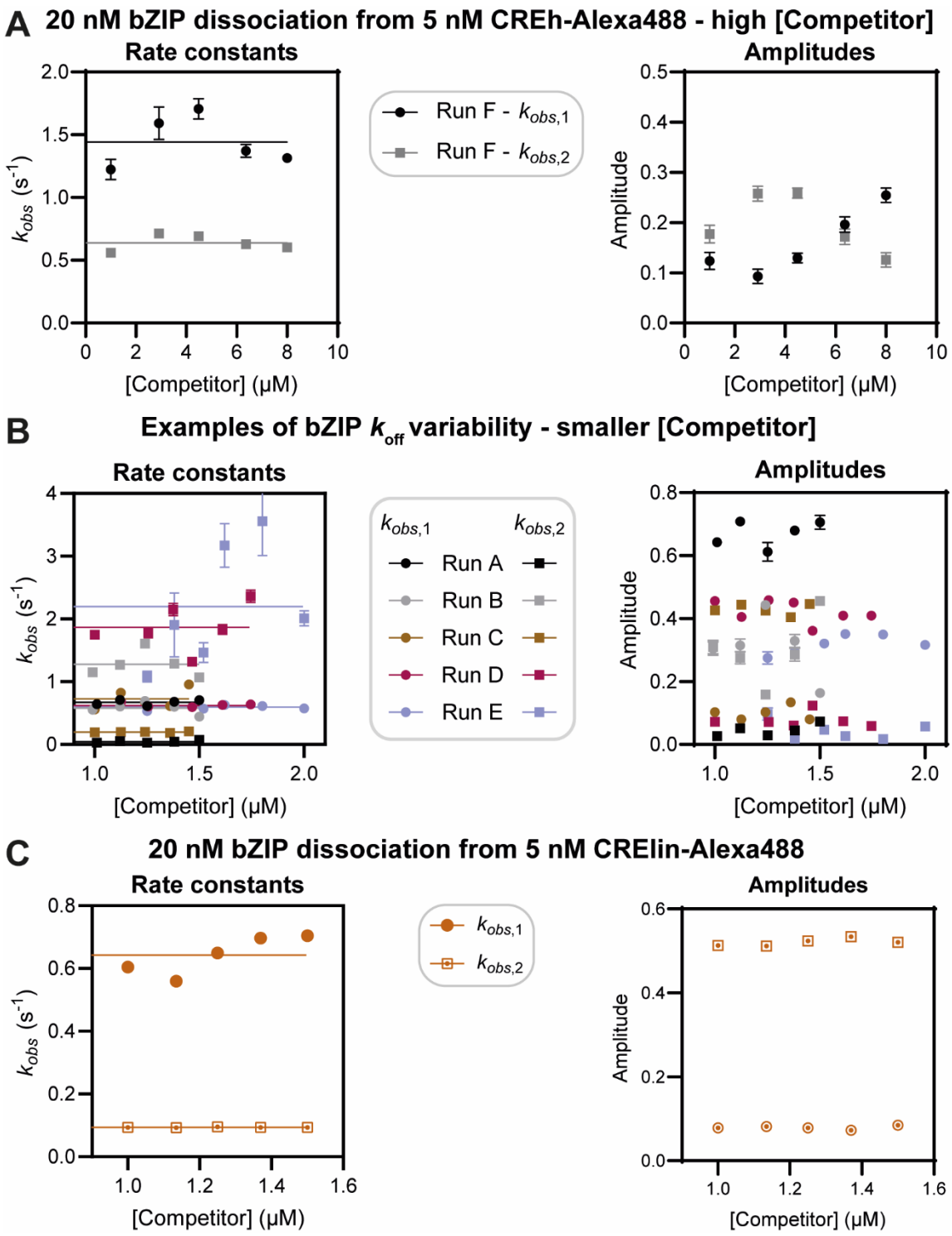


Figure 5.11: Variability in rate constants and amplitudes extracted from double exponential decay fits of CREB dissociation. Conditions: pH 7.4, 25 °C. **A.** Double exponential decay k_{obs} from dissociation experiments with high competitor concentrations are constant, but their respective amplitudes cross over at higher concentrations. **B.** k_{obs} from shorter range experiments, with 20-30 nM bZIP and 5 nM CREh-Alexa488. Only $k_{obs,1}$ is consistent across all runs and usually has the highest amplitude (except for runs B and C). The $k_{obs,2}$ rate constant itself is either very low (run C) or very high (run E). **C.** A double exponential decay fit of dissociation from CRElin obtains rate constants of similar values, but with reversed amplitudes.

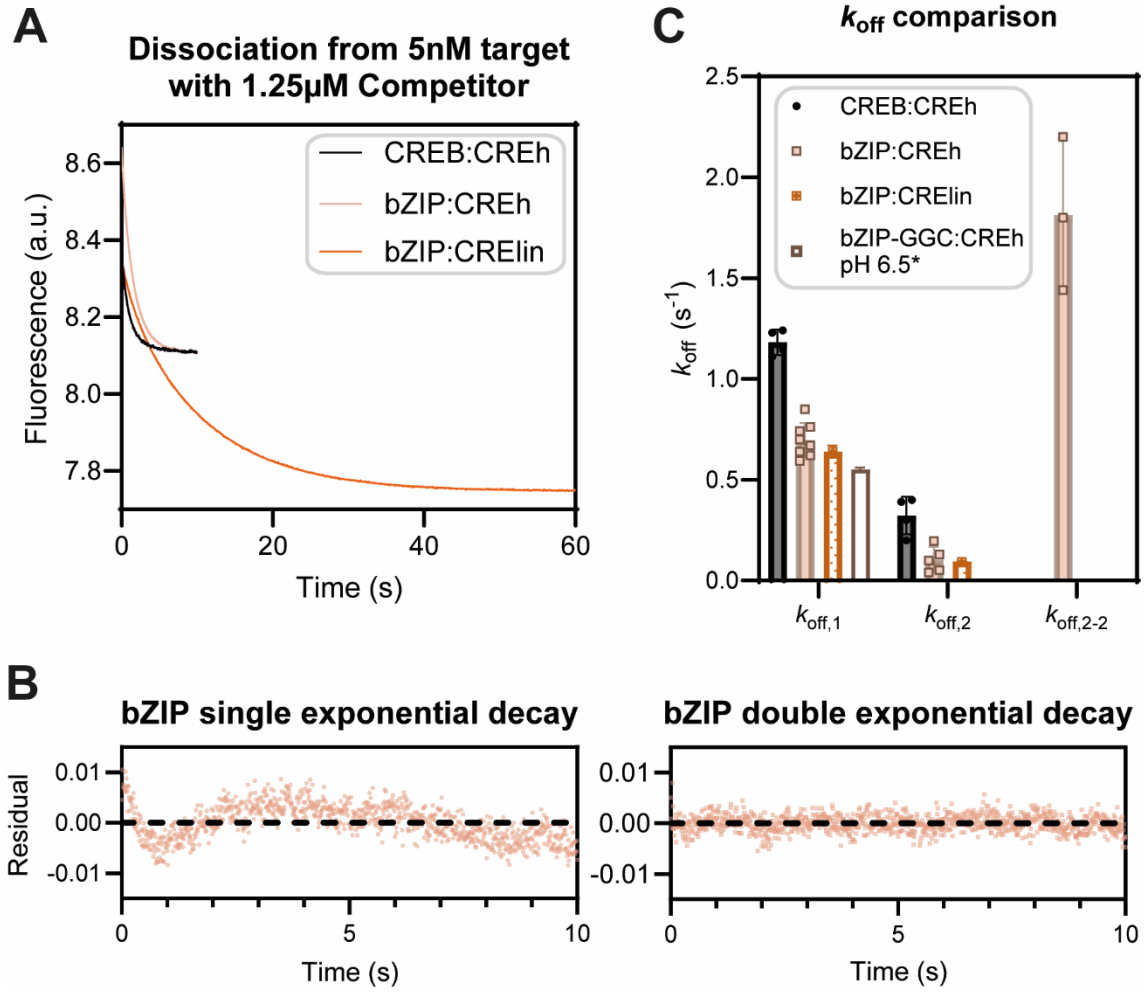


Figure 5.12: Comparison of dissociation kinetics to target DNA for CREB and bZIP. Unless otherwise stated, conditions are pH 7.4, 25 °C. **A.** Dissociation traces examples. CREB appears to reach plateau earlier, and its amplitude is lower than the bZIP equivalent trace, likely because more DNA is bound by bZIP beforehand. CRElin dissociation trace shows a much longer-lived complex and increased amplitude. **B.** Representative residuals from single (left) and double (right) exponential decay fits of bZIP dissociation traces. Double exponential decay captures more of the curve's behaviour. **C.** Comparison of extracted rate constants. Most bZIP traces have a $k_{off,1}$ around $0.7s^{-1}$ and a $k_{off,2}$ around $0.1s^{-1}$, although $k_{off,2}$ can be highly variable. This suggests $k_{off,1}$ is the "true" rate constant especially due to its higher amplitude (Figure 5.10) and the other rate constant is due to dilution artefacts. The two rate constants extracted from CRElin dissociation are in the same range. Accepted CREh dissociation k_{off} from a similar bZIP forced dimer mutant is in the same range as $k_{off,1}$ (*Kuravsky, unpublished).

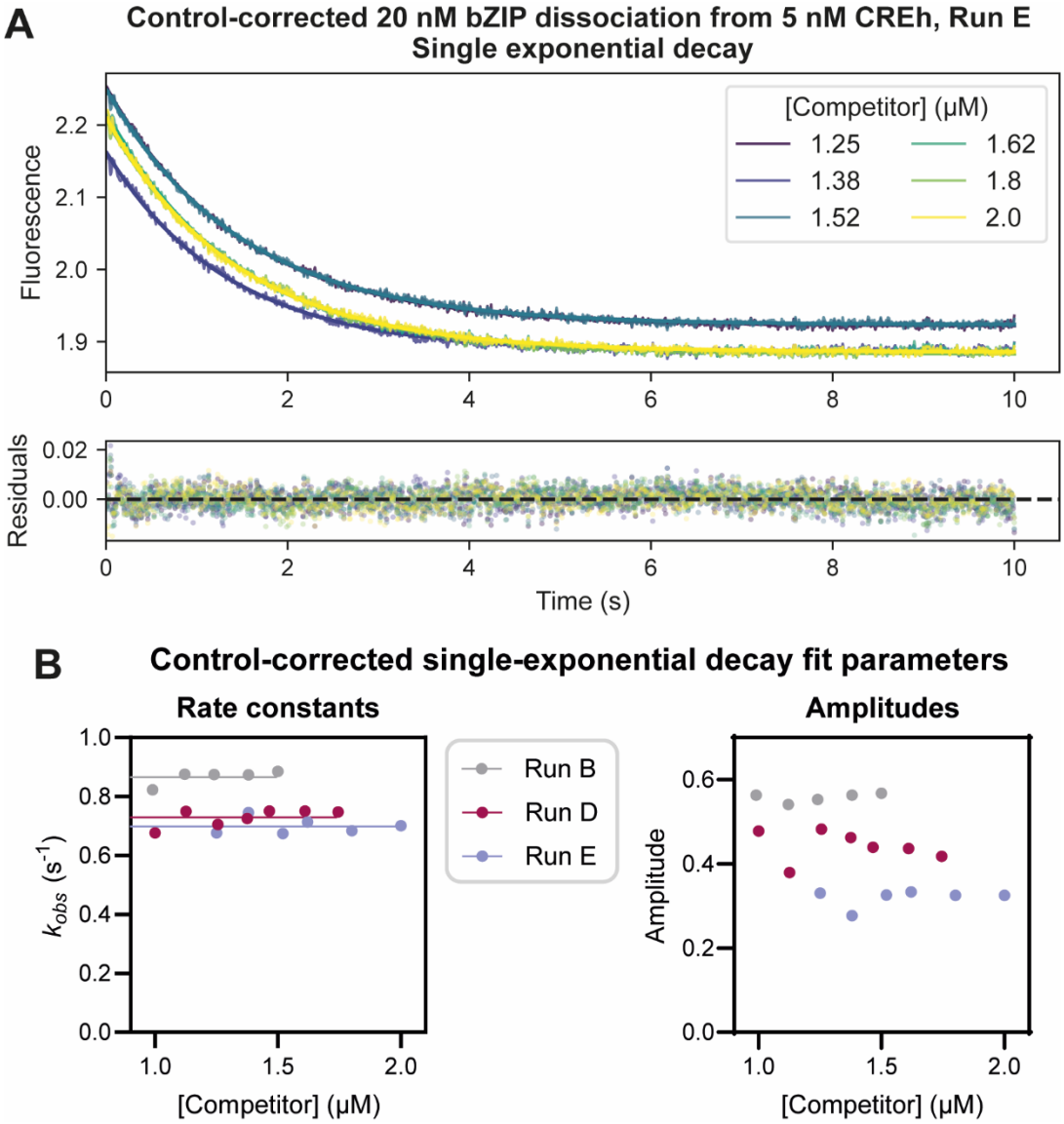


Figure 5.13: Simple control correction of dissociation traces for bZIP. A. Example of control-corrected dissociation traces, now with single exponential decay behaviour. **B.** Amplitudes (right) are mostly constant across traces from the same run and so are the rate constants (left), giving values around $k_{off,1}$ (figure 5.12).

If the correction for potential DNA artefacts was applied to a compatible experiment, subtracting the trace fluorescent signal using signal from a CREh into CRElin, single exponential decay functions were better fits, as observed previously for CREB (Figure 5.13A). The extracted k_{off} was similar to the higher amplitude double exponential decay $k_{off,1}$ (Figure 5.13B). Therefore, it is likely that the high amplitude rate constant ($k_{off,1}$ for CREh, $k_{off,2}$ for CRElin) is the "true" dissociation phase, and the other is an artefact.

Through this assumption, the kinetics K_D can be calculated, using $\frac{k_{off}}{k_{on}}$. The kinetics K_D shows that bZIP has, in fact, almost four-fold increased affinity for CREh compared to CREB (Figure 5.7E, Table 5.2). Hence, CREB's non-bZIP domains impede binding to target to a similar extent than the non-target, both through decreased association and increased dissociation (Table 5.3).

5.3.3. Monitoring association rate constants in the presence of non-target decoys shows CREB's non-DBD facilitate target-search.

Despite the non-bZIP domains limiting affinity to both target and non-target DNA, the affinity for target remains high for both bZIP and CREB. Therefore, a difference in non-target DNA binding affinity may have significant impacts on target search in the nuclear environment since concentrations of competing DNA are huge, especially since both bZIP's and CREB's nCRE affinity is still relatively high and common for important biological interactions (244). It is likely that many off-target binding events occur, and the influence of non-DBD domains could limit those and improve target-search efficiency.

To test this, I performed multiple k_{on} determination experiments in the presence of increasing nCRE concentrations to better mimic the target search situation. Both CREB and bZIP fit to association traces with single exponential decay functions at lower competitor concentration. Above 5 μM nCRE, a double exponential decay function appears more appropriate for CREB only (Figure 5.14A), as this change does not improve the bZIP traces fit greatly and one of the extracted rate constants is very close to the single exponential decay one (Figure 5.14B,C). This other rate constant could be due to a change of binding mechanism (perhaps the resolution of an intermediate by affecting its lifetime, or the appearance of intersegmental strand transfer). DNA mixing artefacts is possible but unlikely, as dilution of a DNA mixture in water did not yield a signal change (Figure 4.22).

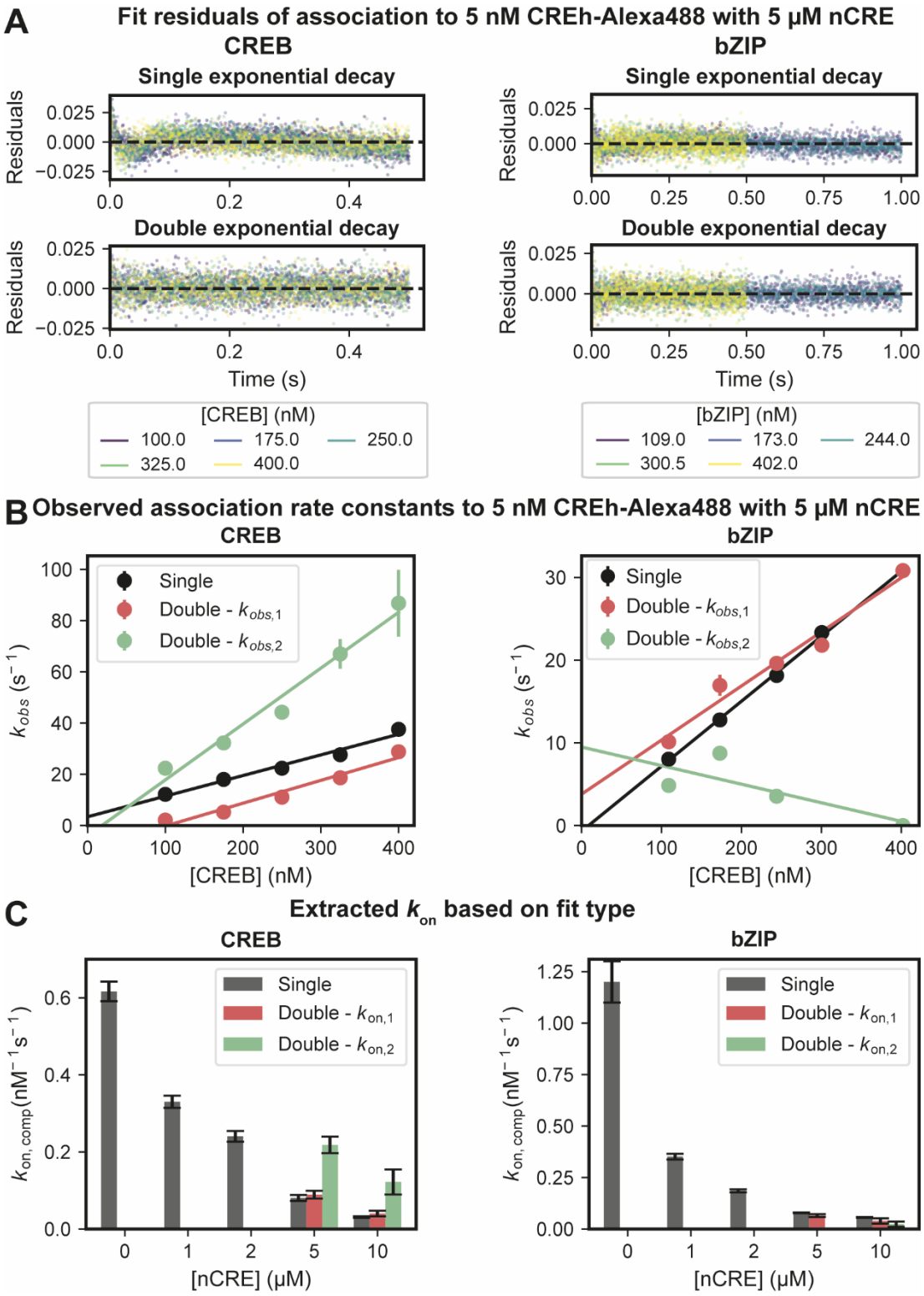


Figure 5.14: Single and double-exponential decay fitting of association to CREh in the presence of nCRE competitors. Conditions: pH 7.4, 25 °C. **A.** Above 5 µM, traces start to show double exponential decay character, more so for CREB (left) than bZIP (right). **B.** Extracted k_{obs} when binding 5 nM CREh with 5 µM competitor and **C.** association rate constants with competitor ($k_{on,comp}$) for single and double-exponential decay fits. The latter do not reliably show a second process occurs for bZIP and, for both, the $k_{on,1}$ is similar to the single exponential decay rate constant.

While bZIP has a higher CREh k_{on} than CREB, it was reduced more sharply by the addition of competitor (Figures 5.15A,B, 5.16A,B). I can explain this behaviour quantitatively with a simple model, calculating the free CREB population using the $K_{D,nCRE}$ (method's Equation 2.20) (Figure 5.15C). This suggests target search is greatly dependent on $K_{D,nCRE}$, which is represented by the linear relationship between $k_{on,app}$ to predicted $[CREB]_{free}$ population, supposing that [nCRE] is constant (Figure 5.15D) (although the 10-fold excess [nCRE] requirement for this assumption is only fully met above 5 μM). The model also obtained similar k_{on} and $K_{D,nCRE}$ to other methods if allowed to fit them (Figures 5.15E, 5.16E, Tables 5.3, 5.4). This model predicts the full-length protein would have a higher k_{on} at high competitor concentrations, 2.5-fold higher with extrapolation to 1 mM nCRE (Figure 5.15F). Although, both k_{on} are around 1000-fold reduced compared to the uninhibited rate constant, and it is likely that other processes such as intersegmental transfer (perhaps observed already with CREB) play a greater role than in our experimental setup.

Given the success of this "sequestration" model, I generated association "master curves" by adjusting the free CREB population for every datapoint using $K_{D,nCRE}$. This extracted rate constant values at lower CREB concentrations that are usually inaccessible due to the requirements of pseudo-first-order kinetics (Figure 5.16). The bZIP datapoints are well-modelled linearly, with a k_{on} virtually identical to the one obtained previously (Figure 5.16C, Table 5.3). However, CREB's appear to deviate from linearity at lower concentrations, levelling up and therefore adjusting for the impossible negative y-intercept, as discussed in Chapter 4 (Figure 5.16C). This behaviour is slightly accounted for if using the first of the double-exponential decay rate constants, delaying the point at which k_{obs} level off (Figure 5.16D). Therefore, something different seems to happen with CREB compared to bZIP at these lower concentrations. Regardless of single or double exponential decay fits, the master curve k_{on} is similar to those obtained previously (Figure 5.16E, Table 5.3).

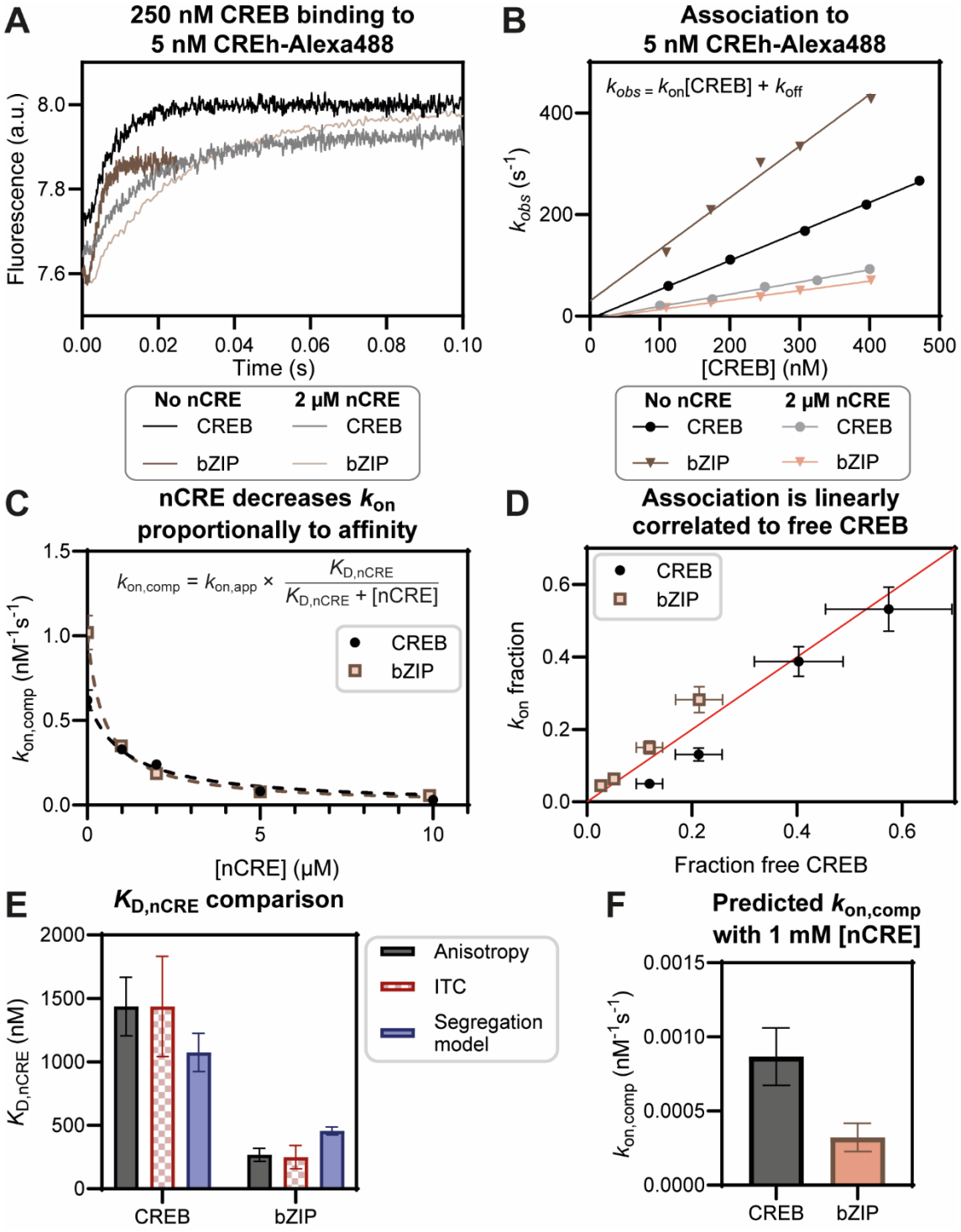


Figure 5.15: CREB's association to CREh is impeded proportionally to competitor K_D. Conditions: pH 7.4, 25 °C. **A.** Example association traces to labelled CREh. Addition of nCRE DNA slows association. **B.** Addition of nCRE competitor decreases bZIP's extracted single exponential decay rate constants to below CREB's. **C.** A model calculating the portion of free CREB (dependent on $K_{D,nCRE}$) fits $k_{on,comp}$ against competitor concentration well and predicts CREB will bind to CRE faster with increasing competitor concentration. **D.** The dependence on $K_{D,nCRE}$ is shown by the apparent linear dependence between fraction free CREB and $k_{on,comp}/k_{on,app}$. **E.** Freely fitting the model in (**C**) extracts a simikar $K_{D,nCRE}$ to the other methods. **F.** The simple model predicts CREB's $k_{on,comp}$ would be twice that of the bZIP at 1 mM nCRE.

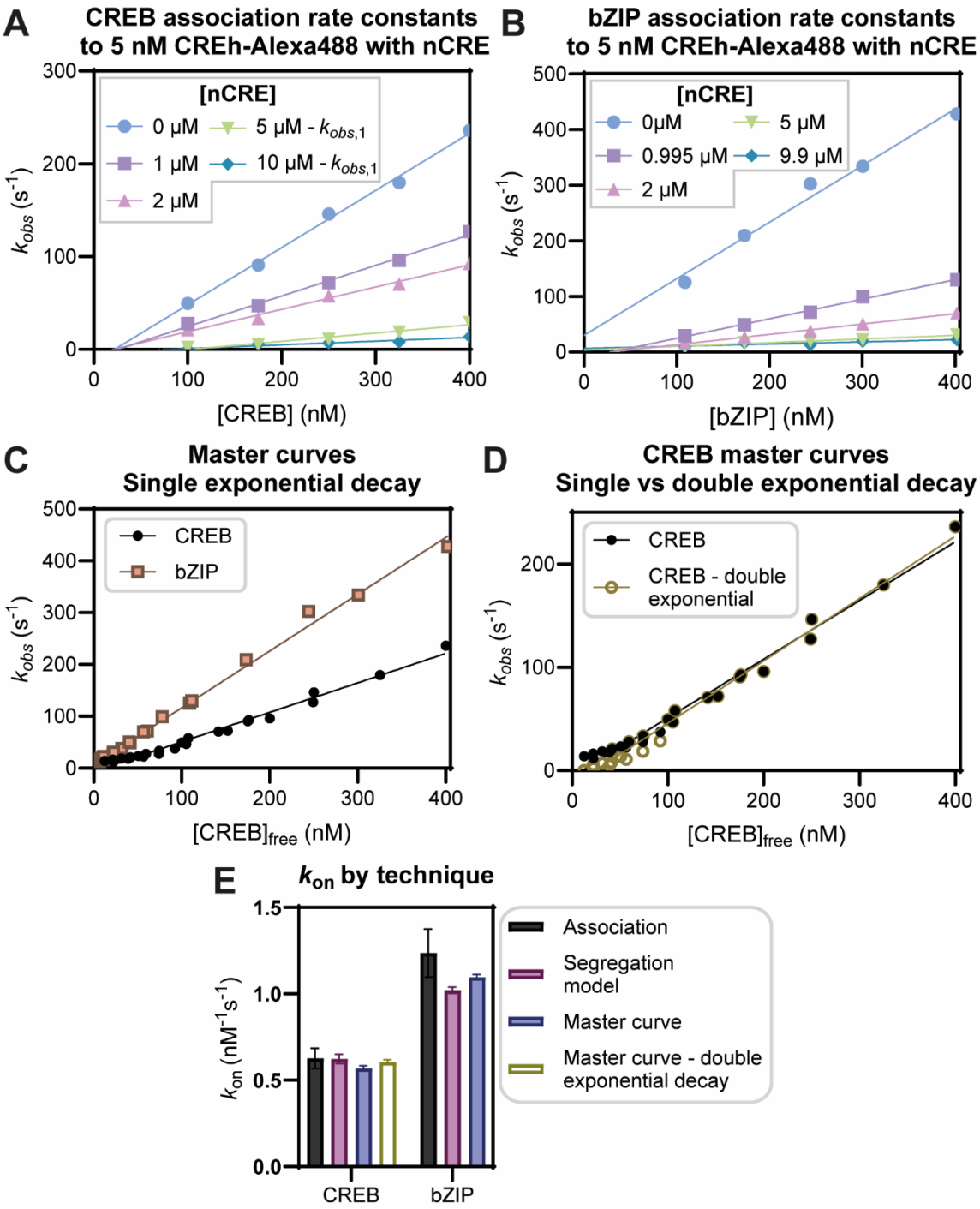


Figure 5.16: CREB association rate constants to CREh in the presence of competitor. Conditions: pH 7.4, 25 °C. **A.** CREB and **B.** bZIP k_{obs} to 5 nM labelled CREh with increasing nCRE competitor concentrations. Above 5 µM nCRE, the first of the double exponential decay k_{obs} is used for CREB. **C.** Assuming the model discussed in Figure 5.14 holds, the concentration of free CREB (rather than total CREB) can be used to create a master curve from all the rate constants obtained in the experiment. While bZIP is well-represented linearly, CREB deviates from it at lower concentration. **D.** Using double-exponential decay rate constants delays the point at which CREB k_{obs} level off. **E.** Master curves (regardless of single or double exponential decay fits), free fitting of the segregation model and traditional association setup all extract similar k_{on}.

5.4. Investigation of the individual contributions of CREB's non-DNA binding domains to DNA binding

Having established that the CREB's TAD plays a role in DNA binding, I then proceeded to investigate the individual contributions of the Q1, Q2 and KID domains, either through deletions (ΔQ1, ΔQ2) or by substitution of the KID domain with a GS linker.

5.4.1. Equilibrium binding titrations show each domain impedes binding to non-target DNA, but not apparently to CREh.

Broad characterisation of binding to Alexa Fluor™ 488-labelled CREh was performed using equilibrium titrations, correcting for intensity change as previously explained. Titrations were well-fit by a single-site binding model as for bZIP and CREB, with $K_{D,app}$ values in the nM range (Figure 5.17B, Table 5.2). As alluded to previously, this method is not very suitable for determining accurate K_D values in this range because it is smaller than the homodimerization K_D, as indicated by the small deviation in the residuals around the transition (Figure 5.17A). Meaning, the data might be indicative of DNA-induced dimerization more than CRE affinity outright.

Affinity to Alexa Fluor™ 488 nCRE DNA was tested through the established equilibrium anisotropy (Figure 5.18) and ITC titrations (Figure 5.19). However, there are not enough replicates to analyse each method individually for the mutants. Yet, for CREB and bZIP, which have the largest number of replicates, both techniques provided similar estimates, but both also showed some variability in their respective results (Figure 5.7F). I therefore pooled all the estimates for each mutant together to account for said variability. It is also worth noting that, for the mutants, anisotropy titrations had greater variability than ITC titrations, although more anisotropy titrations were carried out (Figure 5.20C).

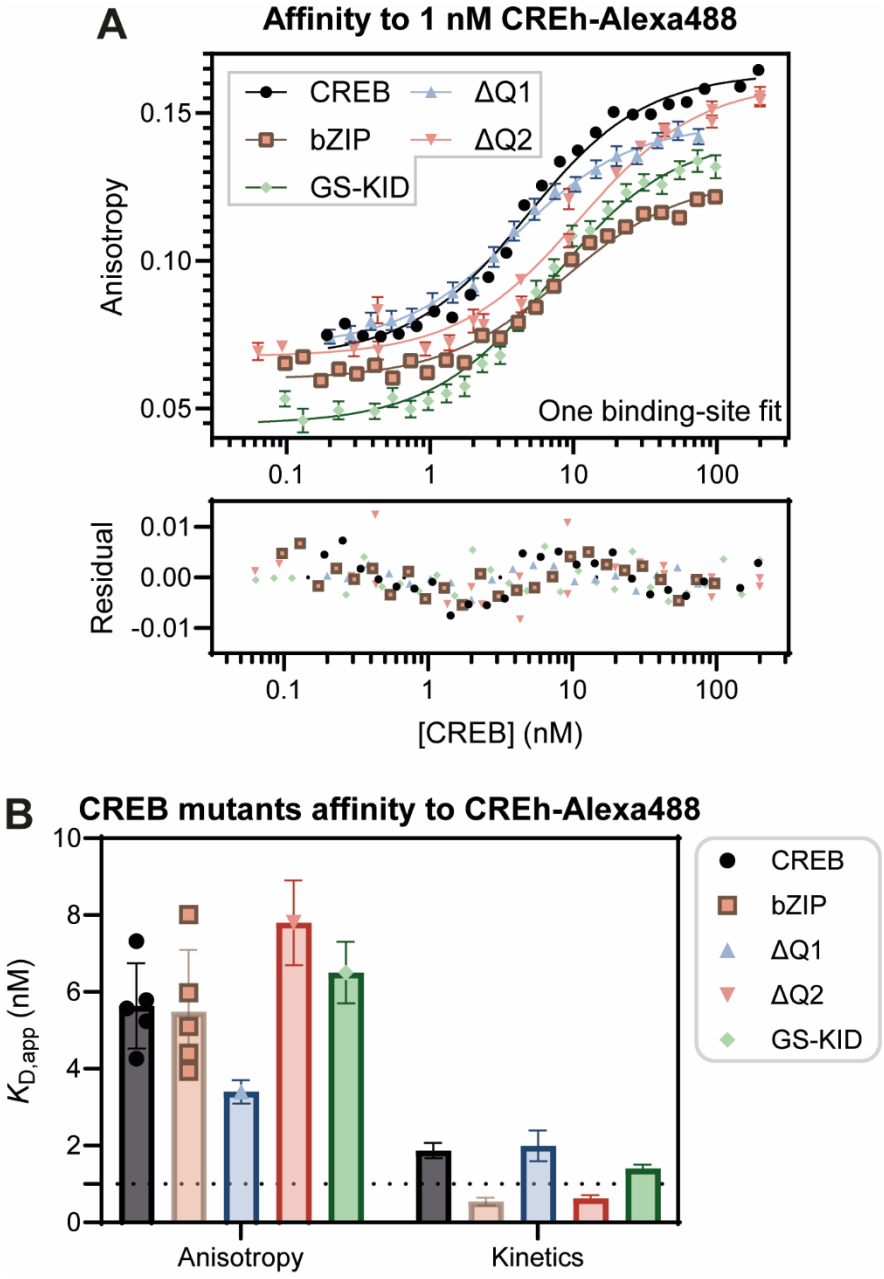


Figure 5.17: Anisotropy titrations to CREh and K_D compared CREB mutants. A. A single anisotropy titration binding to labelled 1 nM CREh-Alexa488 was performed for each mutant at pH 7.4, 25 °C. All show binding transitions in a similar range with a small deviation from the single binding-site model, consistent with DNA-induced dimerization at low CREB concentrations. **B.** Comparison of both anisotropy-derived and kinetics-derived K_D at 25 °C, pH 7.4. All values are in a similar, low nM range for anisotropy. Only Q1 deletion appears to increase affinity, whilst other mutants decrease it. Kinetics-derived K_D are all around 1 nM and show that Q2 and KID mutants have affinity to target in-between CREB's and bZIP's. CREB-A and Q1 mutants' K_D are within error of CREB's. Kinetics values were calculated using $K_D = k_{off,1}/k_{on}$ assuming the second rate constant is caused by DNA artefacts, with errors calculated following propagation of uncertainty. Anisotropy bar-chart values are averages of repeats with standard-deviation (if multiple datapoints) or fitting value and standard error from (**A**) (if a single datapoint).

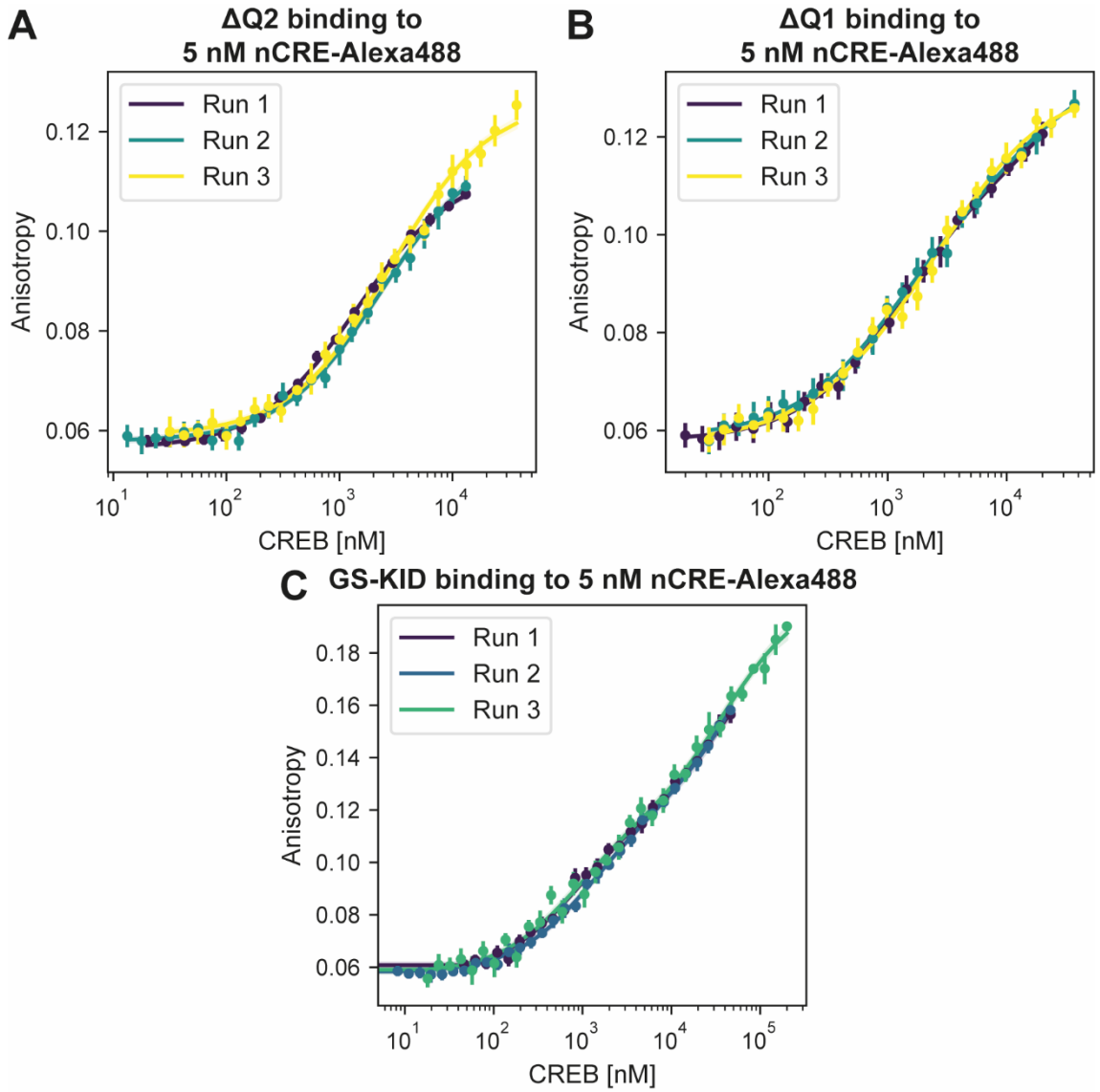


Figure 5.18: Anisotropy titrations to of CREB mutants to 5 nM nCRE-Alexa488. Titrations at 25 °C, pH 7.4 tended to overlap well for each **A.** ΔQ2, **B.** ΔQ1 or **C.** GS-KID if repeats were performed and are well-fit by a two binding-sites model. Just as for CREB, an upper plateau was not reached, but high concentration titrations were not performed to verify this with absolute certainty.

Using the pooled data, none of the mutants had similar affinity to that of the bZIP, so no single domain appears responsible for the difference (Figure 5.20D, Table 5.4). While ΔQ2 has a similar affinity to nCRE than CREB and both GS-KID and ΔQ1 have increased nCRE affinity, all three mutants' values are within error of each other and of CREB, as shown by the similarly in ITC and anisotropy titrations (Figure 5.20). Therefore, it is likely that all three domains contribute to limiting association with nCRE, but their effects are less evident with individual domain deletions. Furthermore, each of GS-KID and ΔQ1 only have

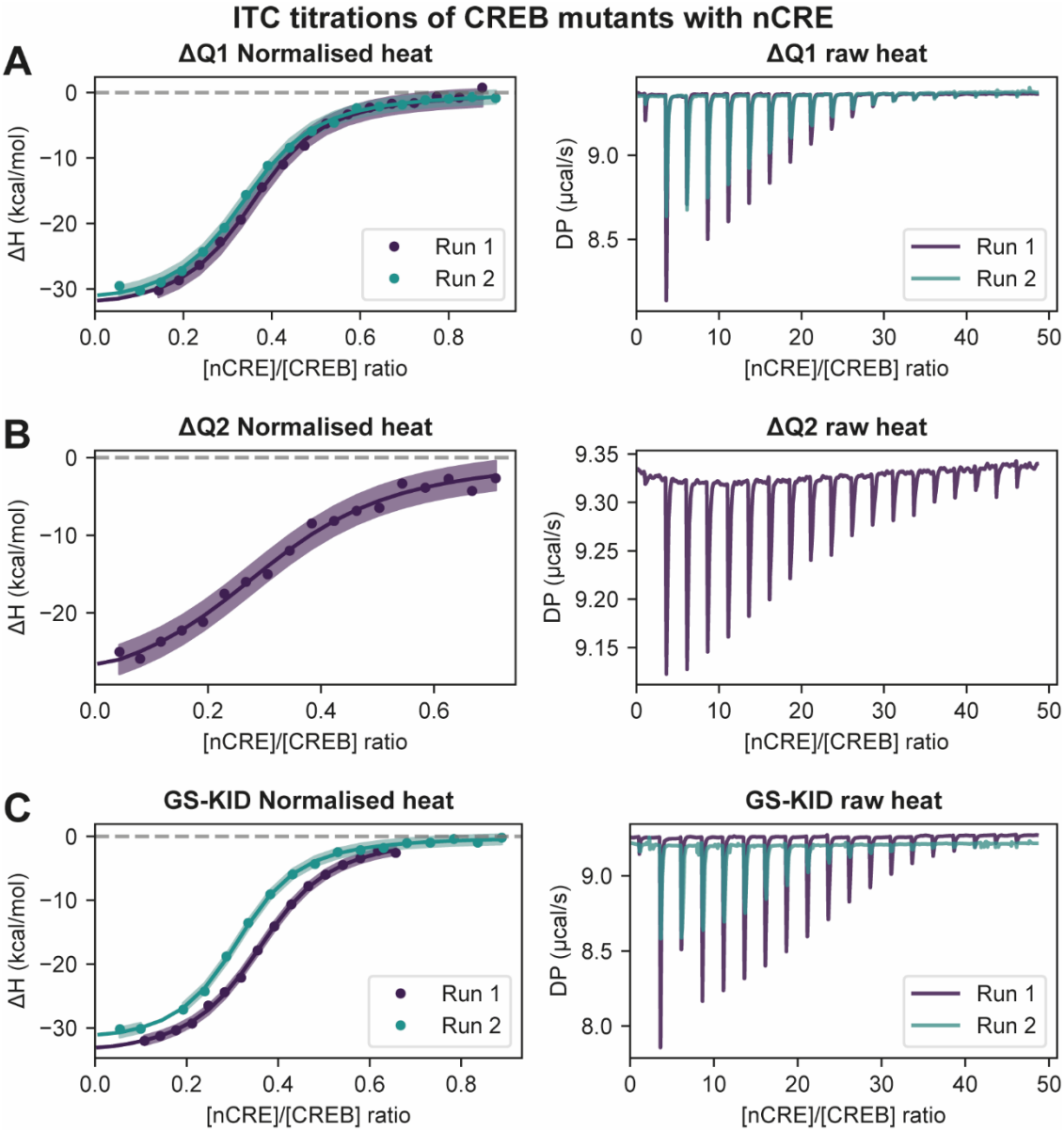


Figure 5.19: ITC raw heat and normalised heat titrations of CREB mutants with nCRE. A. ΔQ1, **B.** ΔQ2, and **C.** GS-KID ITC titrations at 25 °C, pH 7.4 tend to overlap well or at least, transition with a similar gradient and in a similar range. This is especially the case for **A.** ΔQ1. Two runs were performed for each mutant, except for **B.** ΔQ2 for which a single run was made.

a 2/3rd reduced $K_{D,nCRE}$ compared to CREB which combine for an only two-fold increased affinity, still far off the four to five-fold increased nCRE affinity of the bZIP (Figure 5.20D, Table 5.4). Hence, there likely is a synergistic effect from the deletion of multiple domains rather than the deletion of standalone domains.

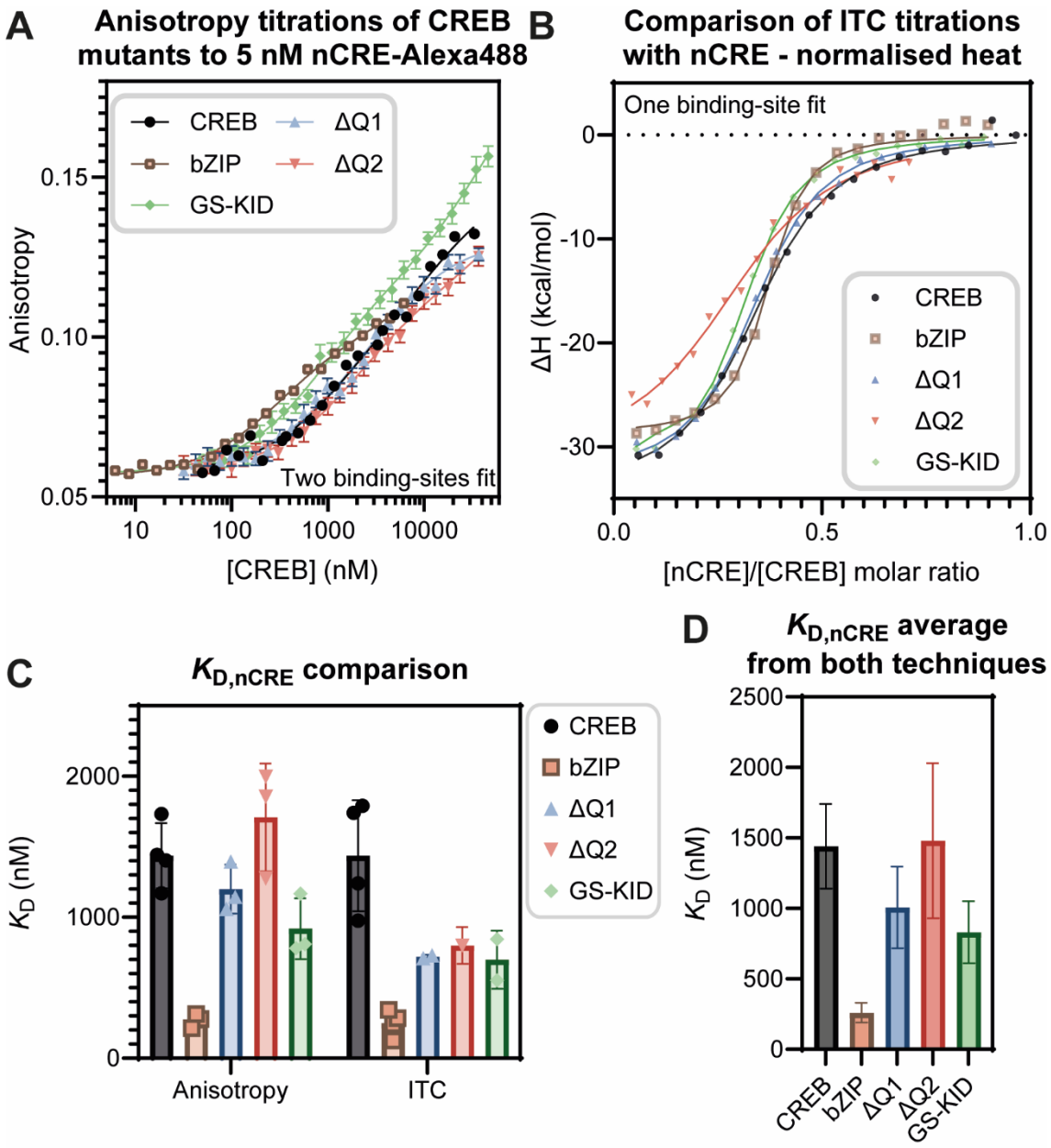


Figure 5.20: $K_{D,nCRE}$ estimated from ITC, anisotropy titrations and pooled replicates compared for each mutant. Conditions: 25 °C, pH 7.4. **A.** Representative anisotropy titrations to labelled nCRE. Only GS-KID visibly transitions in-between CREB and bZIP, while the other mutants follow CREB's behaviour. **B.** bZIP visibly transition more sharply than CREB in ITC titrations, with ΔQ1 and GS-KID mutants having a slightly sharper transition. ΔQ2's unique behaviour is due to lower concentration of protein (giving lower c-value). **C.** Anisotropy and ITC $K_{D,nCRE}$ compared. Anisotropy-derived values suggest only GS-KID (and possibly ΔQ1) increase binding to non-target. ITC titrations suggest all domain mutants increase affinity to non-target by a similar amount but lack the replicates to show this convincingly. **D.** Average of pooled replicates based on mutants in (**C**) suggest only Q1 and KID domains impede binding to nCRE to a similar extent. Interestingly, none show affinity comparable to the bZIP, suggesting an additive effect of all non-bZIP domains in impeding association to DNA. All values are average of replicates, with errors from standard deviation except for ΔQ2's single value and associated fitting standard error.

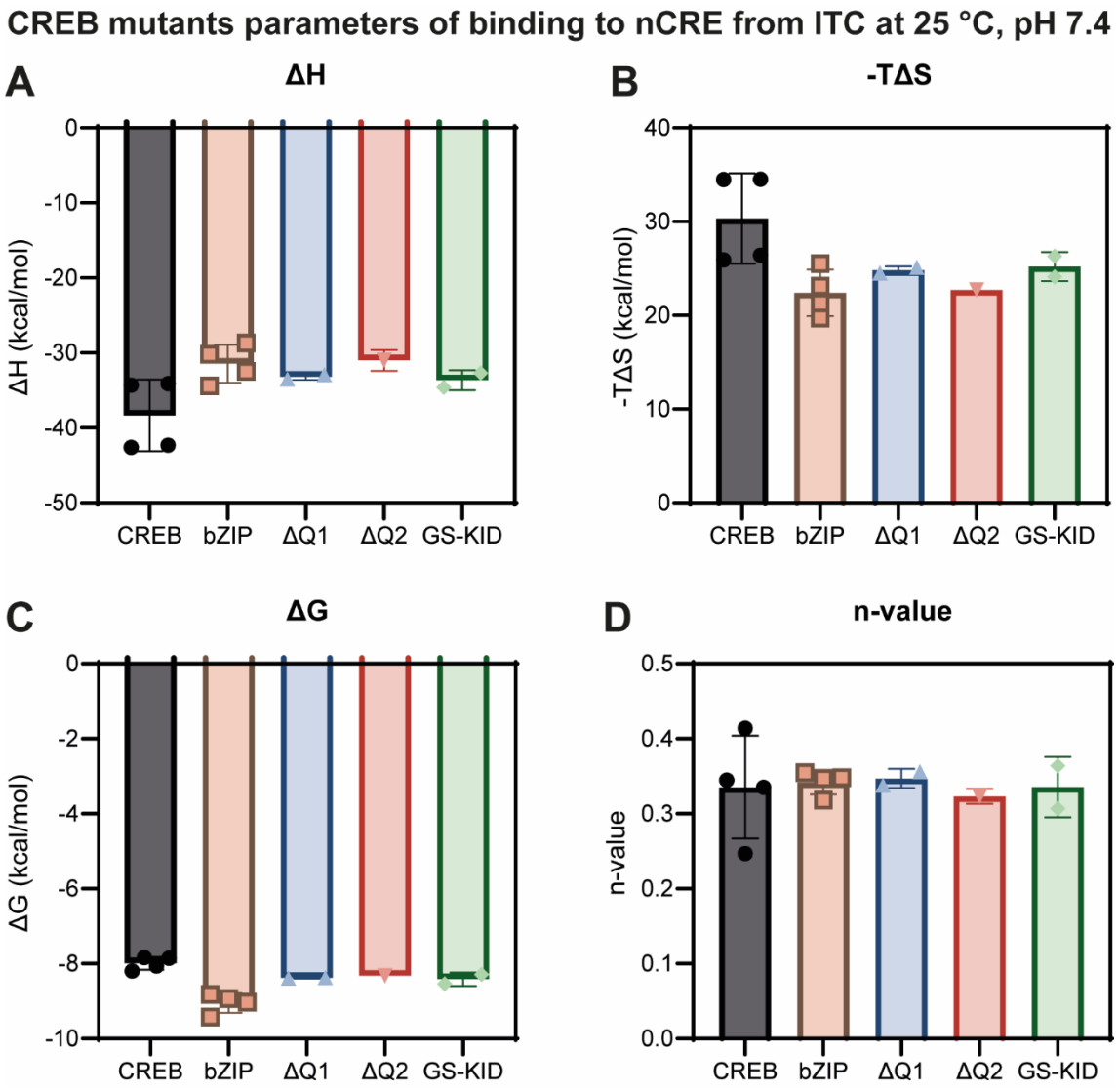


Figure 5.21: Thermodynamic values and n-value extracted from ITC binding to nCRE titrations for CREB mutants. Bar values are averages of shown datapoints with standard deviation or fitting value and error (if available) if a single datapoint is shown. **A.** ΔH and **B.** ΔS values of the domain mutants are closer to bZIP's than CREB but are all within error of each other. ΔQ1 and GS-KID values are almost the same and ΔQ2 values are closer to that of the bZIP but the low c-value used does not allow to make this observation with absolute certainty. **C.** ΔG behaviour mirrors that of the K_D and values are in-between CREB and bZIP. **D.** n-values are similar for all mutants.

Notably, most ΔQ2 titrations suffer from low protein concentrations used throughout, highlighted by lower range anisotropy titrations (Figure 5.18A) and low c-value shape of ITC titration (Figure 5.19B). Coupled with a low extinction coefficient, this could increase the inaccuracy of results. As it stands, there is no indication that this domain contributes to decreasing affinity for nCRE but, ideally, some repeat experiments with

higher ΔQ2 concentrations would validate this or, conversely, show that it does impede DNA binding.

If exploring the thermodynamics of binding to nCRE (only available with ITC), GS-KID's and ΔQ1's average values were almost identical and were in-between those of the bZIP's and CREB (Figure 5.21, Table 5.5). Deletion of Q2 appeared to have a greater effect with thermodynamic values closer to that of the bZIP, potentially explained by its closer proximity to the binding domain, but this is from a single experiment, furthermore with low c-value, giving it a unique behaviour (Figure 5.20B), and these observations are therefore preliminary at best. Furthermore, all thermodynamic values (besides ΔG) are within error of each other, so there may not be much thermodynamic variation for the mutants at all. All mutants also have the same n-value, around 0.35 (Figure 5.21C, Table 5.5). The implications of this low -n-value have already been explored previously, but it's replication across different mutants strongly disfavours that it is a result of errors in the determination of CREB concentration.

5.4.2. Kinetics experiment show each domain similarly impedes association to target DNA but each domain's influence on complex lifetimes varies.

Having established that none of the mutations had a drastic impact on CRE binding affinity I therefore moved on to characterise the interaction kinetically instead. As described previously, association kinetics were tested by rapidly mixing vast excess of mutants with 5 nM Alexa Fluor™ 488 labelled CREh. All traces were well fit by single exponential decay functions with extracted rate constants between that of CREB and bZIP (Figures 5.22, 5.23, 5.24, Table 5.3). The mutants' k_{on} are very similar in value (Figure 5.24C, Table 5.3). Since

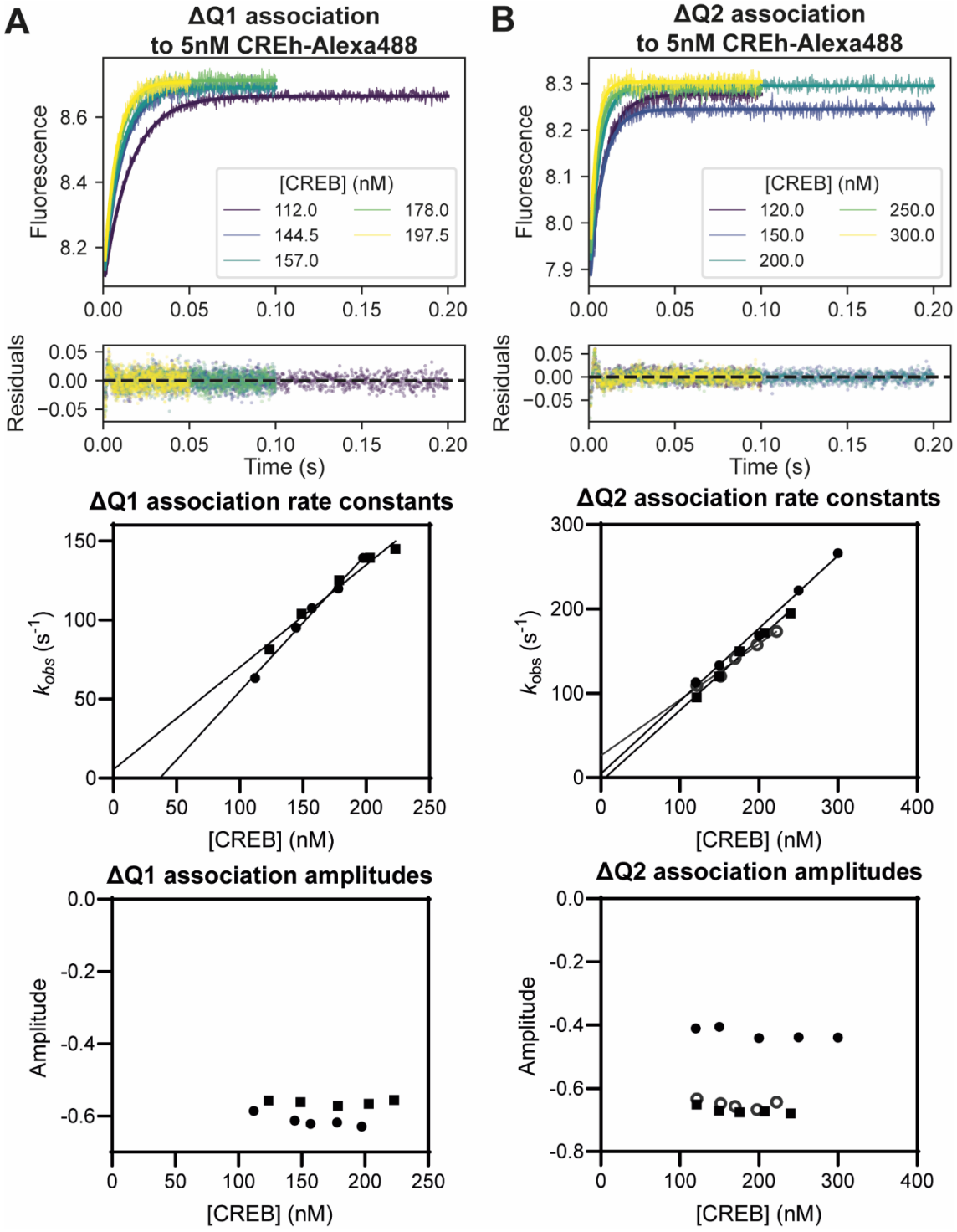


Figure 5.22: Q1/Q2 deletion mutants association experiments to 5 nM CREh-Alexa488. Conditions: pH 7.4, 25 °C. **A.** ΔQ1 and **B.** ΔQ2 mutants' association traces to labelled CREh are both fit well by single exponential decay functions with 1ms dead-time. ΔQ2 repeat association experiments overlap well, but there is some fluctuation with ΔQ1. Amplitudes are slightly dependent on concentration with a minor increase in amplitude with increasing concentration but are consistent across a single experiment.

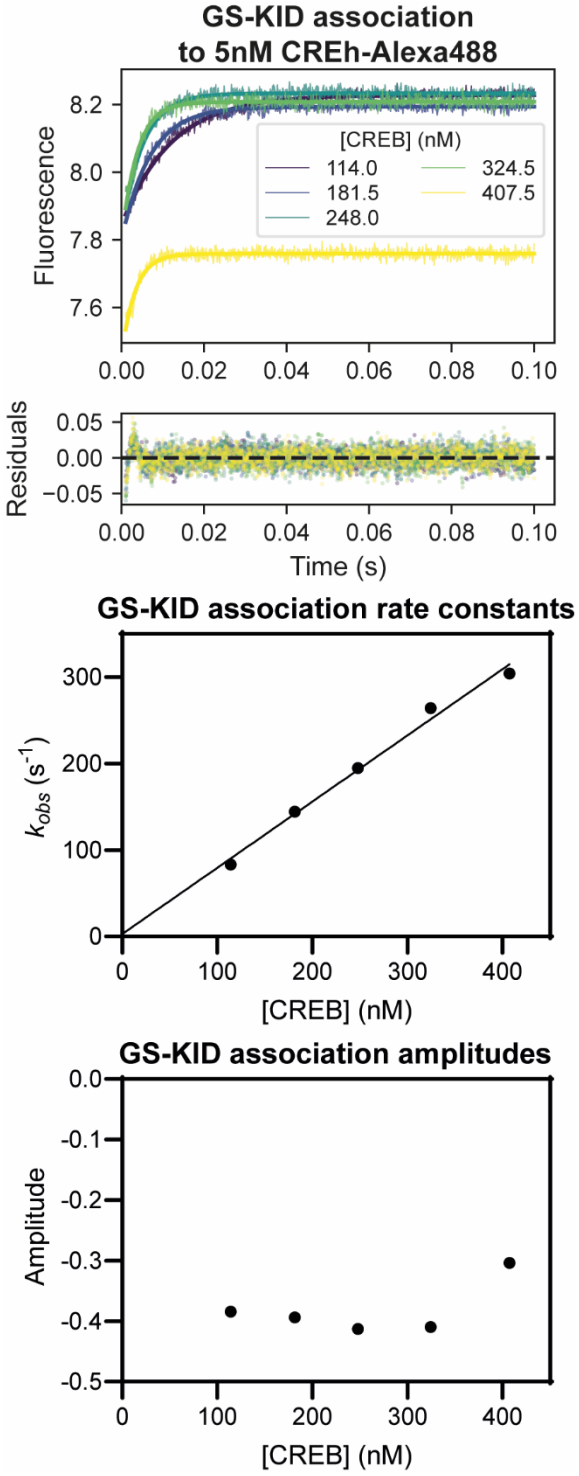


Figure 5.23: GS-KID mutant association experiment to 5 nM CREh-Alexa488. GS-KID mutant association traces to labelled CREh at pH 7.4, 25 °C are fit well by single exponential decay equations with 1ms dead-time and the resulting k_{obs} align well linearly, although there is only one association experiment. Just like for the Q regions mutants, amplitudes are almost consistent but slightly increase in absolute value with concentration. GS-KID's last amplitude datapoint is higher due to the last trace showing different behaviour.

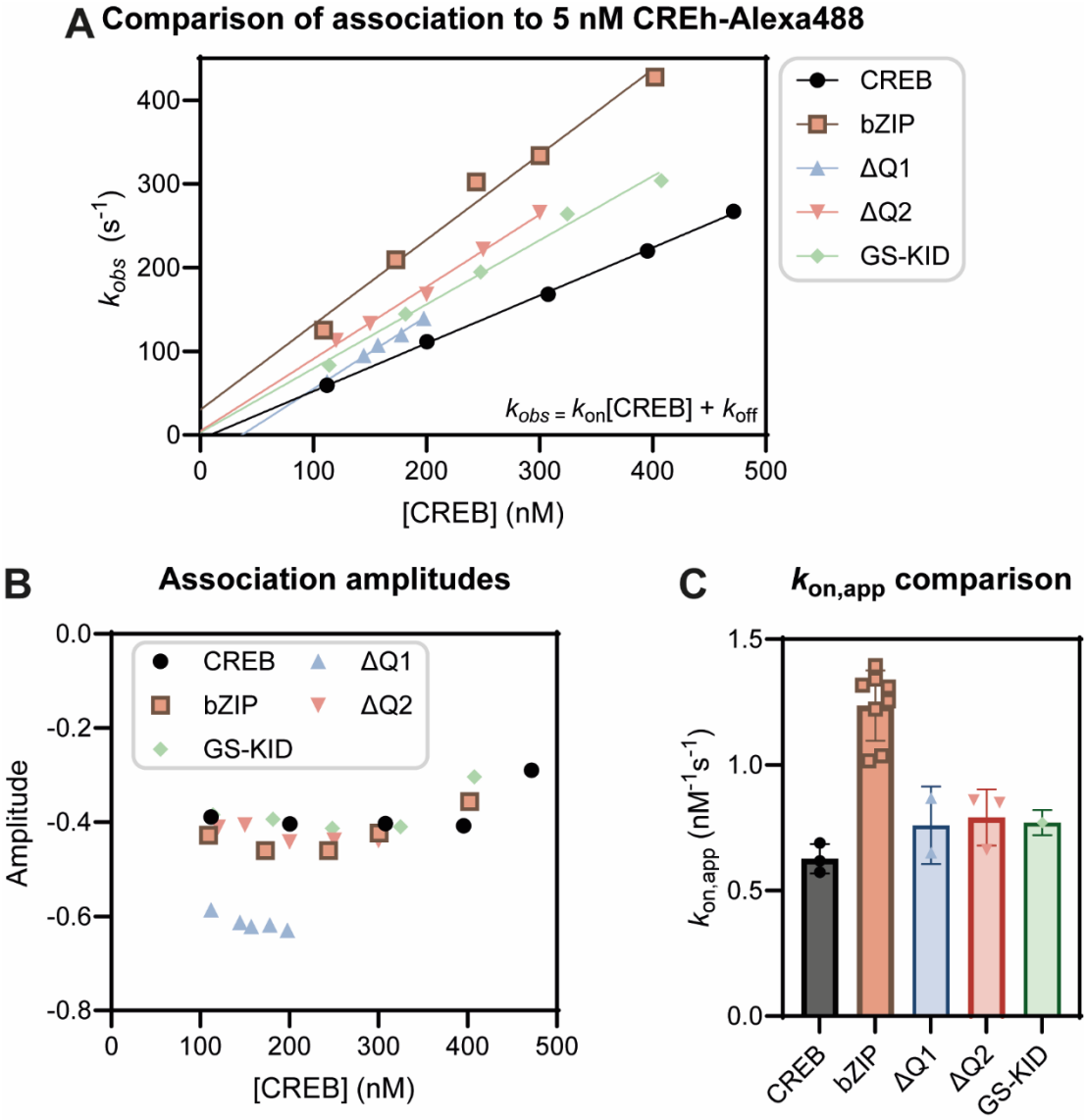


Figure 5.24: Comparison of CREB mutants' association kinetics. A. Association k_{obs} from a single run for each mutant to labelled CREh, with **B.** their associated amplitudes, which are consistently showing similar behaviour. **C.** k_{on} values to labelled CREh at pH 7.4, 25 °C compared for each mutant, with bar-charts showing the average value for each mutant with their standard deviation – or fitting error if a single value was available. All mutants have a similar k_{on} in-between CREB and bZIP, showing a likely cumulative effect in limiting CREB's association to CREh.

the Q-rich regions are similar in both composition (Figure 5.2), this could indicate a general Q-rich region dependent effect rather than a Q1 or Q2-specific one like GS-KID, especially since the charge disparity between both is quite large as ΔQ1 is -3 charges lower (Table 5.1).

I next compared dissociation kinetics. However, not enough replicates were obtained, henceforth these are preliminary results. As before, complex dissociation was

tested by dilution of a 4:1 CREB mutant to Alexa Fluor™ 488 labelled CREh mixture into a solution with vast excess of unlabelled target competitor. Again, traces are best fit by double-exponential decay functions rather than the typical single exponential decay ones (Figures 5.25, 5.26). As outlined both in previous Sections 4.5 and 5.3, a kinetic phase apparently exists even in the absence of CREB protein (labelled CREh mixing with unlabelled CRElin, Section 4.5, which is likely the cause for an expected single to double exponential decay shift. Generally, the rate constant with the highest amplitude was the "true" k_{off}, (often $k_{off,1}$) and the other appeared to be an artefact.

The ΔQ2 result is particularly troublesome for this assumption as amplitude is not consistently higher for one of the k_{off} and a single exponential decay fit, which is usually closer to the dominant rate constant, extracts a k_{off} that is in-between the double exponential decay ones (Figure 5.25). However, it is closer to the lower k_{off}, which is also more consistent across the experiment. This would indicate that the lower phase is the "true" phase. Both $k_{off,1}$ and $k_{off,2}$ are lower than CREB's $k_{off,1}$ (Figure 5.27, Table 5.3), which would highlight a role for the Q2 domain in destabilising CREB bound to target, but the magnitude of which depends on which is the artefact rate constant, which can only be obtained with repeats or an alternative approach.

Conversely, Q1 appears to stabilise the bound complex, as ΔQ1 deletion increases $k_{off,1}$ 1.5-fold (Figure 5.27, Table 5.3). GS-KID deletion appeared to have no effect on dissociation (Figure 5.27, Table 5.3), although there was much experimental variability for this mutant's dissociation experiments (Figure 5.26). In case the lower amplitude $k_{off,2}$ assumption is wrong, and it is relevant to dissociation, it is higher for those two mutants compared to CREB (Figure 5.27, Table 5.3). Overall, these preliminary results interestingly indicate a unique domain-specific effect on dissociation compared to association.

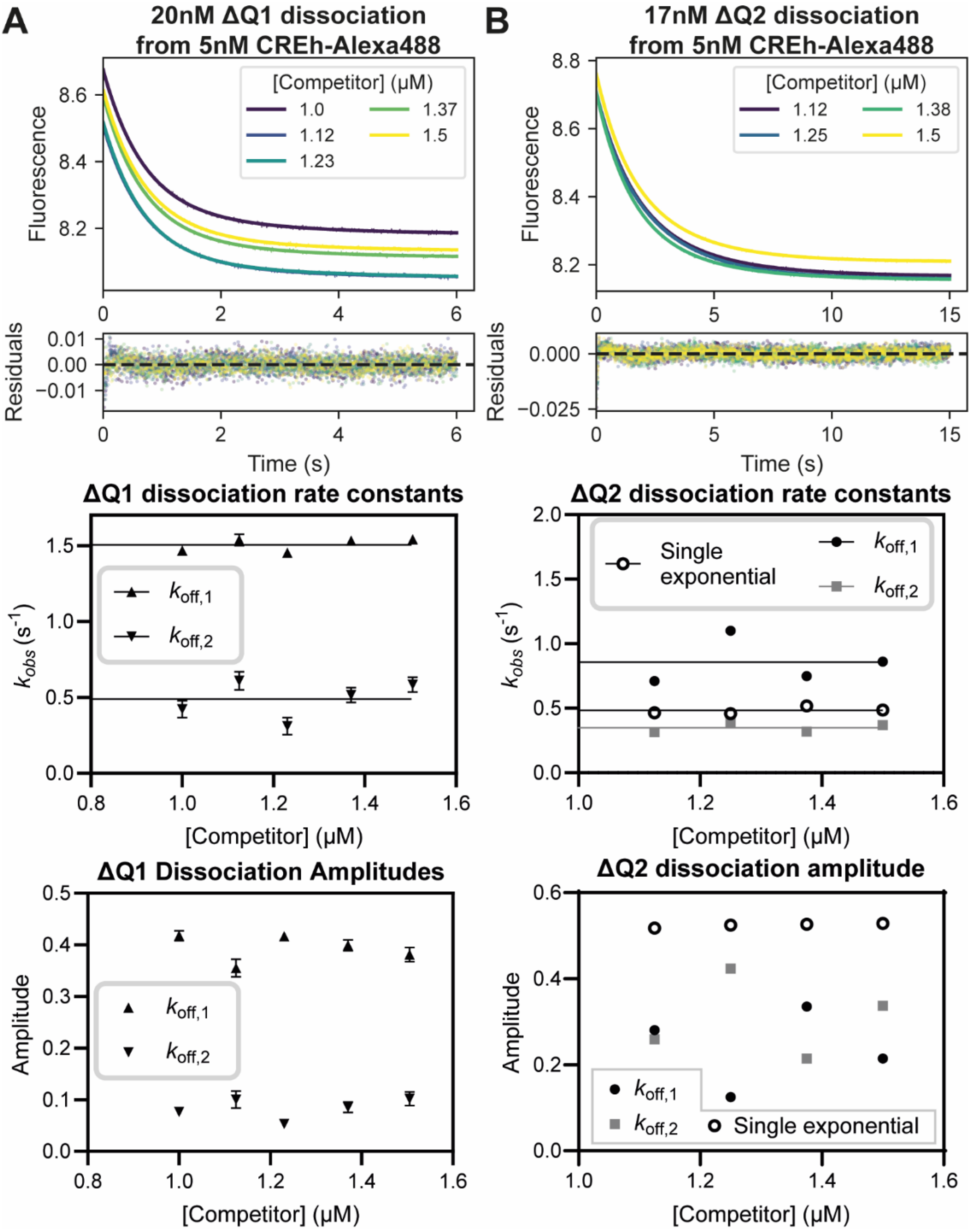


Figure 5.25: Q1 and Q2 deletion mutants' dissociation kinetics. A. Single 20nM ΔQ1 and **B.** 17nM ΔQ2 attempt to obtain dissociation kinetics from labelled 5 nM CREh-Alexa488 through CRElin out-competition, at pH 7.4, 25 °C. Traces fit well to double-exponential decay models, with two broadly constant k_{obs} although the upper one is more reproducible for ΔQ1 and the lower one for ΔQ2. Amplitudes are also consistent for ΔQ1 but sometimes switch around for ΔQ2, so the dominant rate constant is unknown. Single exponential decay fits from ΔQ2 are closer to $k_{off,2}$ which suggest it may be the dominant k_{off}. Single exponential decay values are used as a k_{off} approximation due to this uncertainty.

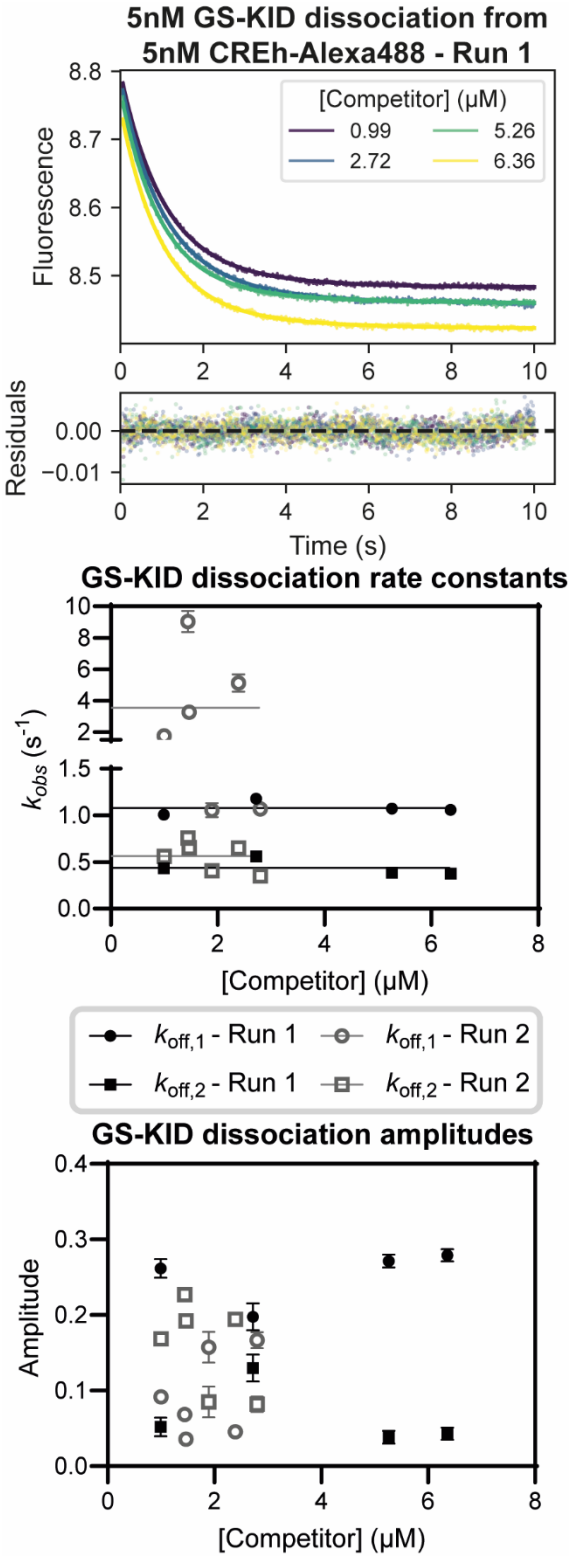


Figure 5.26: GS-KID mutant dissociation kinetics. GS-KID mutant (5 nM for run 1, 20 nM for run 2) dissociation from labelled 5 nM CREh-Alexa488 using out-competition with CRElin at pH 7.4, 25 °C fit to double exponential decay functions well but there is variability between runs both in amplitudes and rate constants, especially run 2 with interchanging amplitudes and $k_{off,1}$ variability. Only two datapoints from that repeat are consistent with the other experiment (both in k_{obs} and amplitudes).

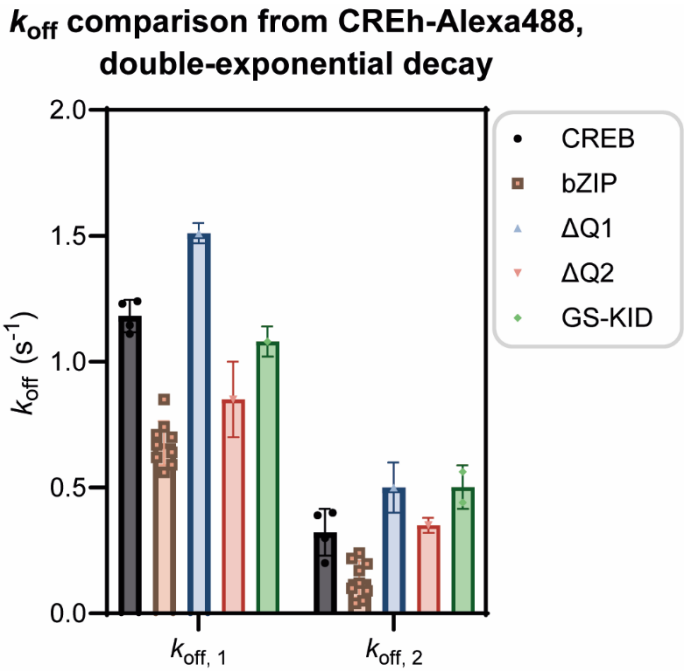


Figure 5.27: CREB mutants k_{off} comparison. Values are at pH 7.4, 25 °C. The $k_{off,2}$ are quite consistent between mutants, only greatly decreasing for bZIP. This may be because of DNA artefacts which show a phase in this range, although it is very consistent for ΔQ2 and single exponential decay k_{obs} for ΔQ2 are closer to its $k_{off,2}$, hence an argument for the opposite could be made for this mutant. Nevertheless, ΔQ mutants have great influence on dissociation, with ΔQ1 deletion decreasing residence time, but ΔQ2 deletion increasing it, almost to the extent of the bZIP. KID has minimal effect. Values are either averaged replicates with errors from standard deviation or fitting value with fitting error for a single datapoint.

Unfortunately, due to the different acquisition timescales, the correction by subtraction of a control labelled CREh into unlabelled CRElin can only be applied to the GS-KID traces. Application of this correction allowed traces to have single exponential decay character (Figure 5.28A) and extracted rate constants much more reproducible (Figure 5.28B). These are close to GS-KID's $k_{off,1}$ from double exponential decay fits, likely the "true" k_{off}, which itself is close to CREB's $k_{off,1}$ (Figure 5.27).

I estimated K_D to CREh using a kinetic approach, assuming $k_{off,1}$ is the "true" k_{off}, whereas the (largely invariant) lower k_{off} was artefactual. Due to the confusion with the ΔQ2 k_{off}, I used the single exponential decay fit instead. All values are around the 1 nM range showing strong target binding (Figure 5.17B). Expectedly, these are lower than the apparent anisotropy-derived K_D which is not corrected for dimerization effects. Only ΔQ2 and GS-

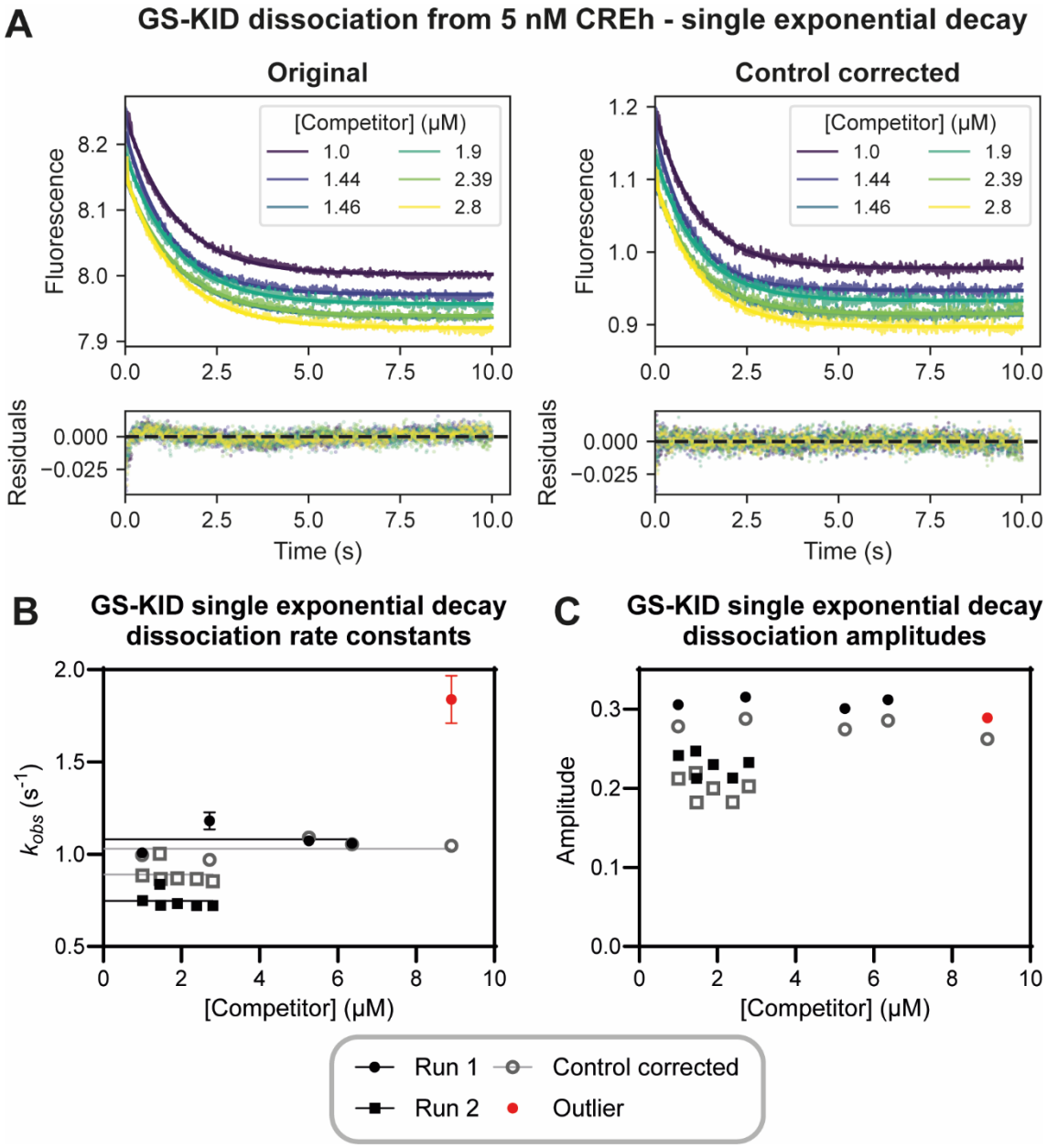


Figure 5.28: GS-KID control-corrected dissociation kinetics. A. Comparison of single exponential decay-fitted dissociation traces without and with rudimentary control correction (subtraction of signal obtained from a single labelled CREh-Alexa488 into unlabelled CRElin dilution, then addition by a constant to bring signal above 0). Rudimentary correction removes the second phase, conforming with two-state model expectations. **B.** Applying this correction makes rate constants (grey) more consistent over multiple runs than without (black) and removes outlier. **C.** Correction also lowers amplitudes.

K_D by method	CREB	bZIP	ΔQ1	ΔQ2	GS-KID	CREB-A
Anisotropy (nM)	5.6 ± 1.1	5.5 ± 1.6	3.4 ± 0.3*	7.8 ± 1.1*	6.5 ± 0.8*	N/A
Kinetics (nM)	1.9 ± 0.2	0.55 ± 0.1	2.0 ± 0.4	0.9 ± 0.2	1.4 ± 0.1	2.1 ± 0.1

Table 5.2: Anisotropy and kinetics K_D of CREB and mutants binding to CREh labelled with Alexa Fluor™ 488. Values are averages of at least three replicates with standard deviation at 25 °C, pH 7.4 († denotes only two replicates, * a single value with error from fitting).

KID appear to slightly increase affinity for CRE, while ΔQ1 deletion does not appear to change target affinity (Table 5.2).

5.5. CREB isoform A has similar DNA binding properties to CREB isoform B

Despite a minor insertion in CREB isoform B between Q1 and KID, capable of forming residual structures (154) and changing the overall charge of the protein (Table 5.1), there appears to be little difference between both isoforms with regards to their biophysical properties, regardless of which biophysical parameter was investigated. Indeed, nCRE binding titrations overlap well, whether with anisotropy or ITC binding titrations experiments (Figure 5.29), and both $K_{D,nCRE}$ (Figure 5.29C, Table 5.4) or thermodynamic parameters (Figure 5.31, Table 5.5) are similar as well. $K_{D,nCRE}$ from ITC titrations have a large variation, but are at the extreme edges of values obtained from CREB, and therefore likely represent the extremes of CREB-A as well (Figure 5.29). Kinetics values are almost equivalent too (Figure 5.30, 5.32, Table 5.3). Therefore, the obtained data does not show that this small domain insertion affects CREB biophysical properties.

5.6. Discussion

After extensive focus on transcription's factors activation signals and DBDs, transcription factor studies' focus is shifting to their transactivation domains' influence. Fascinatingly, a wide range of sequences can activate transcription, provided these have a mixture of negative charges and aromatic residues (18–21), and disordered transcription factor regions are essential in directing their transcription factor to their targets (48,68–71,73,212). One particularly striking example is that swapping the disordered domains of

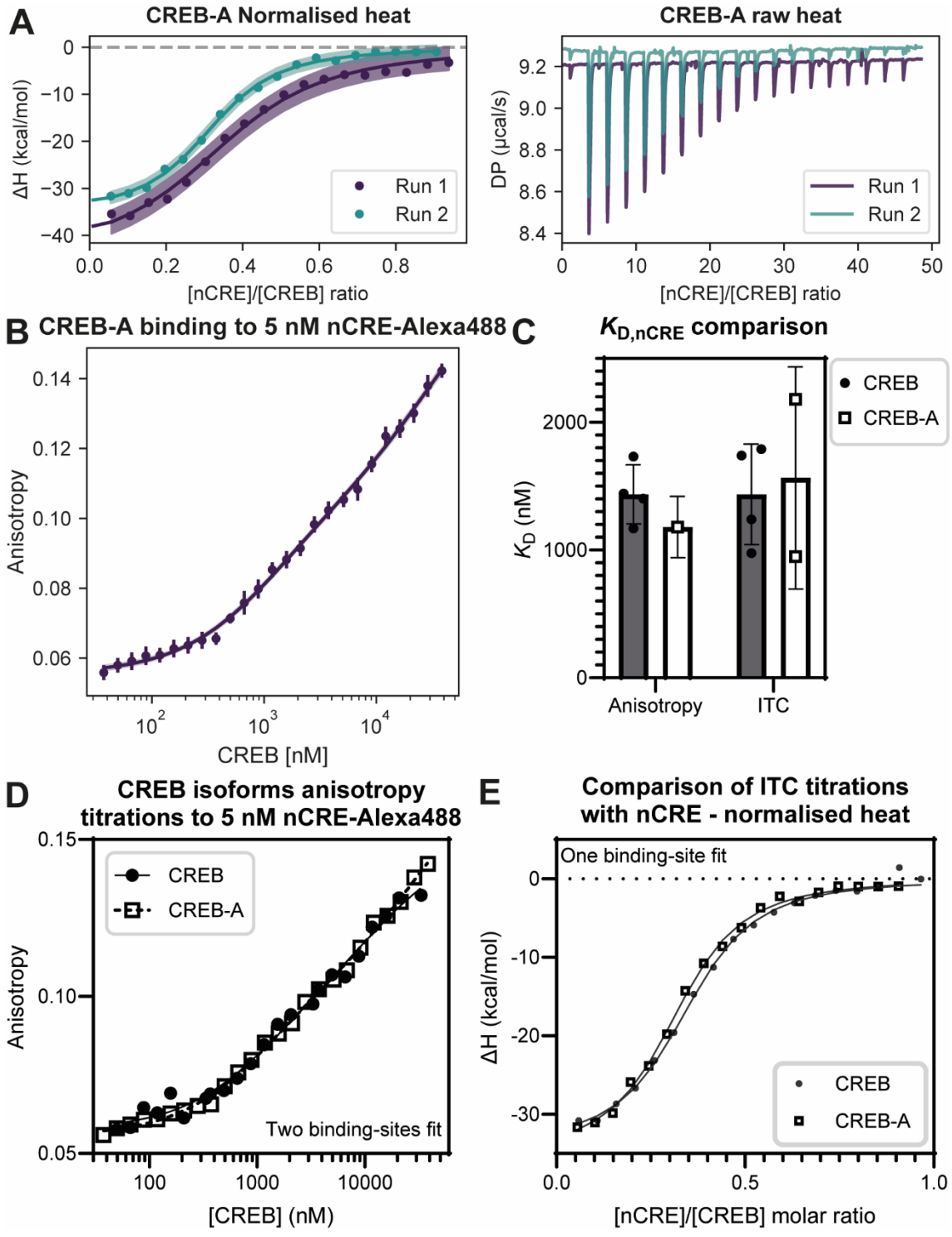


Figure 5.29: Equilibrium binding titrations of CREB-A to nCRE. Conditions: pH 7.4, 25 °C. **A.** ITC titrations of CREB-A bound by nCRE have some variability. **B.** Anisotropy binding titration to Alexa Fluor™ 488-labelled nCRE is well fit by a two binding-sites model. **C.** $K_{D,nCRE}$ is virtually identical between CREB and CREB-A. CREB-A ITC has more variability but likely represents the extreme deviation possible and is close to CREB's extremes as well. Values are averages of replicates with standard deviation errors, or fitting value with fitting error if a single datapoint is present. **D.** CREB isoforms' anisotropy titrations to labelled nCRE overlap almost perfectly. **E.** ITC titrations of nCRE into CREB are also similar.

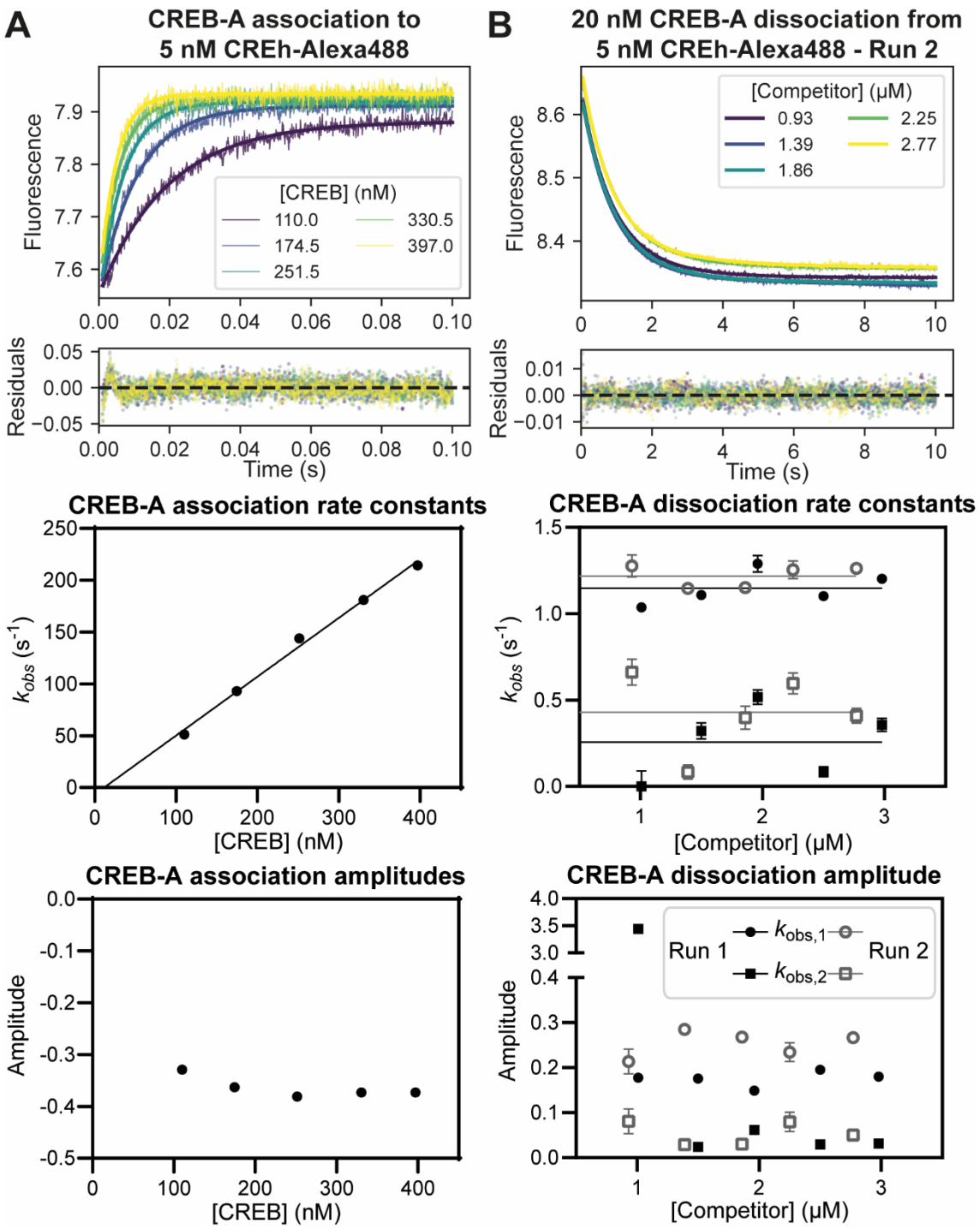


Figure 5.30: CREB-A association and dissociation kinetics with CREh. to nCRE. A. Association data and **B.** dissociation data (from CRElin out-competition) of CREB-A using Alexa Fluor™ 488-labelled CREh at pH 7.4, 25 °C. Association traces are fit by a single exponential decay function, whereas dissociation requires a double exponential decay function. However, only $k_{obs,1}$ which also has the highest amplitude is consistent across replicates and within an experiment. Amplitudes tend to be constant within an experiment for both association and dissociation.

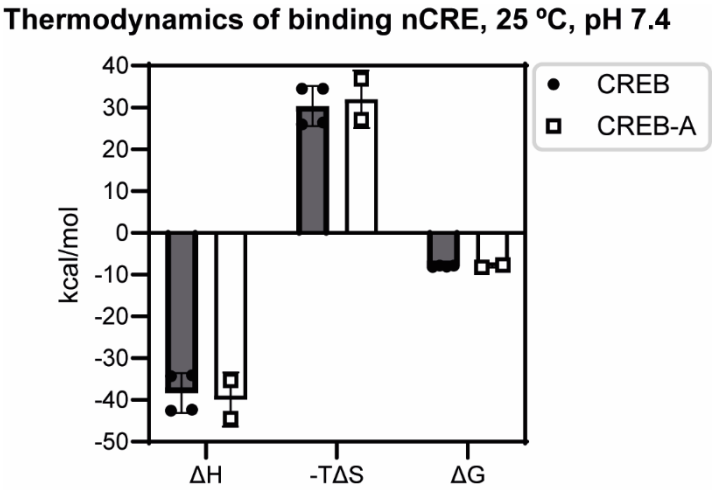


Figure 5.31: CREB-A thermodynamics of binding to nCRE using ITC. CREB-A thermodynamics of binding nCRE are like CREB's. Values are averaged replicates, with standard deviation errors.

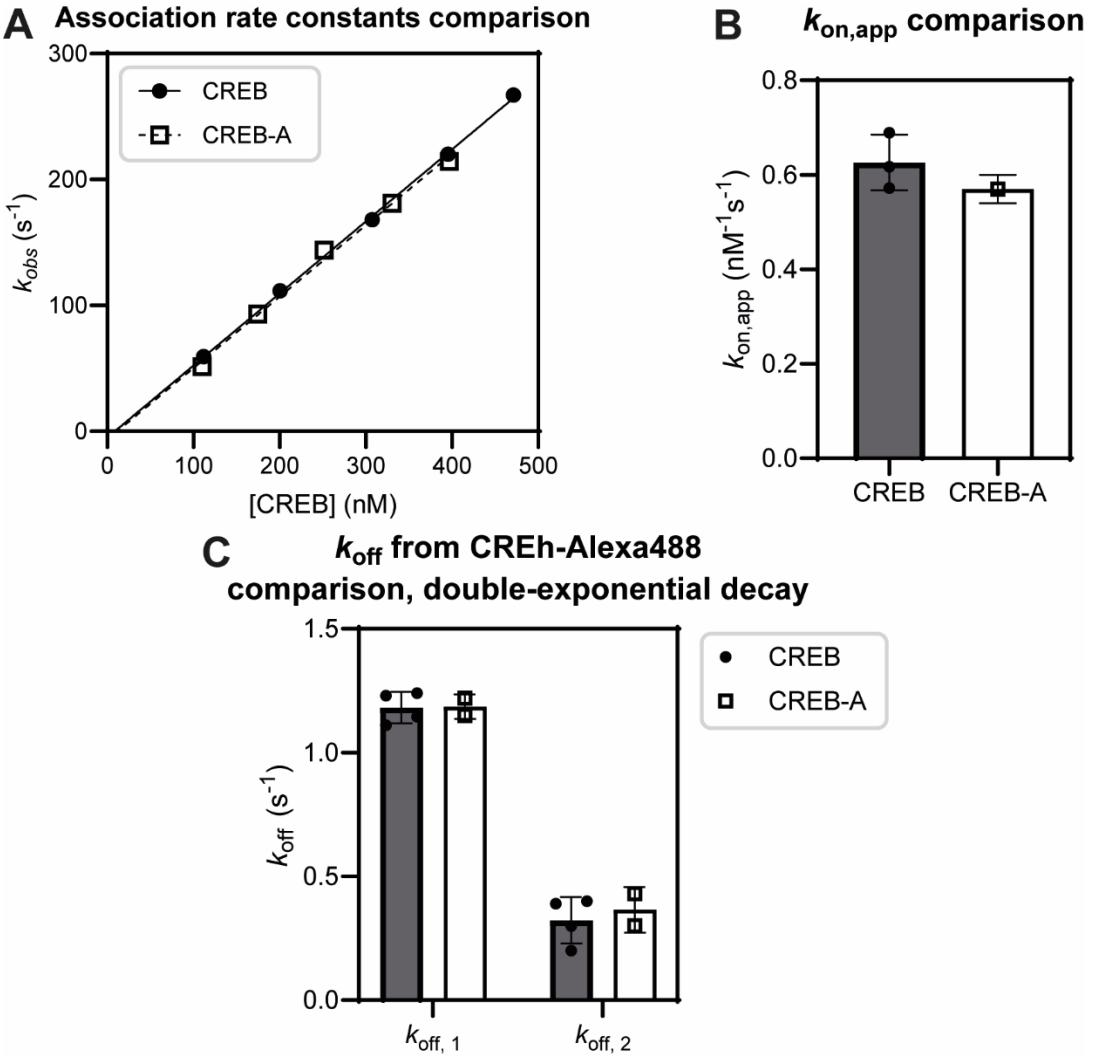


Figure 5.32: Comparison of CREB and CREB-A's binding kinetics to CREh. A. Association rate constants of CREB and CREB-A binding to CREh-Alexa488 overlap almost perfectly. **B.** k_{on} of binding to CREh are in similar ranges. **C.** Both $k_{off,1}$ and $k_{off,2}$ from CREh-Alexa488 are similar too. Values are averaged replicates with standard deviation error.

Mutant	CREB	bZIP	ΔQ1	ΔQ2	GS-KID	CREB-A
k_{on} ($nM^{-1}s^{-1}$)	0.63 ± 0.06	1.24 ± 0.14	0.76 ± 0.15†	0.79 ± 0.11	0.77 ± 0.05*	0.57 ± 0.03*
Master curve k_{on} ($nM^{-1}s^{-1}$)	0.57 ± 0.01 (SE)* 0.60 ± 0.01 (DE)*	1.10 ± 0.02*	-	-	-	-
Segregation model k_{on} ($nM^{-1}s^{-1}$)	0.62 ± 0.03*	1.02 ± 0.02*	-	-	-	-
$k_{off,1}$ (s^{-1})	1.18 ± 0.06	0.68 ± 0.09	1.51 ± 0.04*	0.85 ± 0.15*	1.08 ± 0.06*	1.19 ± 0.05†
$k_{off,2}$ (s^{-1})	0.32 ± 0.09	0.14 ± 0.07	0.5 ± 0.1*	0.35 ± 0.03*	0.50 ± 0.09†	0.37 ± 0.09†

Table 5.3: Association and dissociation rate constants of CREB and mutants from kinetic studies using CREh labelled with Alexa Fluor™ 488. SE: single exponential decay, DE: double exponential decay. Values are averages of at least three replicates with standard deviation at pH 7.4, 25 °C († denotes only two replicates, * a single value with error from fitting).

$K_{D,nCRE}$ by method	CREB	bZIP	ΔQ1	ΔQ2	GS-KID	CREB-A
Anisotropy (µM)	1.4 ± 0.2	0.27 ± 0.05	1.0 ± 0.4	1.7 ± 0.4	1.0 ± 0.3†	1.3 ± 0.2†
ITC (µM)	1.4 ± 0.4	0.25 ± 0.09	0.72 ± 0.01†	0.8 ± 0.1*	0.7 ± 0.2†	1.5 ± 0.9†
Pooled (µM)	1.4 ± 0.3	0.26 ± 0.07	0.9 ± 0.3	1.5 ± 0.55	0.84 ± 0.25	1.4 ± 0.65
Segregation model (µM)	1.07 ± 0.03*	0.49 ± 0.02*	-	-	-	-

Table 5.4: $K_{D,nCRE}$ of CREB and mutants through various methods. Anisotropy uses labelled nCRE with Alexa Fluor™ 488, ITC and the segregation model use label-free nCRE, all at pH 7.4, 25 °C. Pooled represents an average of all individual anisotropy and ITC experiments. Values are averages of at least three replicates with standard deviation († denotes only two replicates, * a single value with error from

ITC parameter at 25°C	CREB	bZIP	ΔQ1	ΔQ2	GS-KID	CREB-A
n value	0.34 ± 0.07	0.34 ± 0.02	0.35 ± 0.01†	0.32 ± 0.01*	0.34 ± 0.04†	0.35 ± 0.04†
ΔH (kcal/mol)	-38 ± 5	-31 ± 3	-33.2 ± 0.4†	-31 ± 1*	-34 ± 1†	-40 ± 7†
-TΔS (kcal/mol)	30 ± 5	22 ± 2	24.8 ± 0.4†	22.7*	25 ± 2†	32 ± 7†
ΔG (kcal/mol)	-8.0 ± 0.2	-9.1 ± 0.3	-8.38 ± 0.01†	-8.32*	-8.4 ± 0.2†	-8.0 ± 0.3†

Table 5.5: Thermodynamic parameters and n-value of CREB and mutants binding to nCRE via ITC. Values are averages of at least three replicates with standard deviation at pH 7.4, 25 °C († denotes only two replicates, * a single value with error from fitting).

HIF-2α and HIF-1α also affected how much each was bound to chromatin, despite not changing their DBD (69). Here, I show that CREB's disordered domains impede binding to DNA but that the affinity to target DNA remains high, which allows for a more efficient target-search process. It is likely other transcription factors' disordered domains have similar function, which would prevent off-target binding events and therefore partially explain why their deletion leads to transcription factors delocalisation (68). Interestingly, there appeared to be a gradual effect, with individual domain deletions never recreating the standalone bZIP domain behaviour. This would also help explain previous studies, where gradual deletions of disordered domains also gradually altered transcription factor behaviour (68). Of course, this is probably only one aspect of disordered domains' influence, and that they can bind coactivators, are activation platforms and possibly even displace nucleosome (21,73–75) probably act synergistically to help factors in their target search (67). It would be interesting to test how these partners influence dissociation from targets, and even how disordered domains affect other parts of the target search process, like sliding or intersegmental transfer.

Interestingly, not all domains seemed to affect the same facet of DNA binding. Q2 mainly affected kinetics, decreasing both dissociation and association to target DNA but not binding affinity to non-target DNA. Conversely, Q1 appears to stabilise binding slightly to target, but impede association and non-target binding. KID impedes all aspects of DNA binding apart from dissociation kinetics which appear unchanged. As KID is the smallest domain, that it has as big an influence as Q1 and was the only domain to consistently impedes DNA binding highlights its importance in this regard, even more-so since that it wasn't deleted but instead replaced with a disordered linker. Since KID has many negative charges, as expected for activation domain, this shows a potential role of activation domain sequences to limit DNA binding and improve target-search efficiency. This has been shown

previously but with artificial E repeats increasing association rate constants to target DNA in the presence of inhibitors (212), rather than bona-fide domains.

On the topic of association, two more pioneered methods are able to extract k_{on} with a similar value to the classical, straightforward mixing method. Although the latter is still recommended, as it is easier to implement and analyse, the master curve method can be useful if k_{on} values are needed at concentrations below the 10-fold excess of protein usually required. The segregation model, although an elegant and simple explanation for the protein's behaviour, relies on critical assumptions (simplified binding, rapid equilibrium reached for CREB binding to nCRE, constant nCRE concentrations) which might not always hold and will therefore not be as reliable as the other methods, but can always be used as a check for values obtained by the other methods.

What gives further credence to the hypothesis that it is KID's negative charges that compete with DNA for binding to the basic region of the bZIP, is that phosphorylation of a CK cassette nearby to the KID, which adds negatively charged groups, also impedes DNA binding (26). This also shows how the influence of KID can be modulated, as the cassette's inhibitory effect increases with the number of phosphorylated residues. It is unclear how phosphorylation of S133, which activates CREB, affects CRE binding. Supposedly, this should inhibit binding too. However, this phosphorylation leads to an expansion of KID, which may provide an easier binding interface for the CBP (154) which could prevent binding to the bZIP. On the other hand, binding to target DNA could prevent KID association with the bZIP and therefore make it available both for phosphorylation and binding by the CBP.

Although how KID impedes association to DNA appears rather straightforward, how the Q-rich domains impede binding to DNA appears less so. These are very important in a

cellular context, with both Q1 and Q2 notably binding the TFIID component of the transcription machinery (152,154). Plus, Q2 is almost entirely deleted in the htCREB isoform (found in human testes) (156) and parts of either Q1 or Q2 are deleted in the CREB-related CREM (or ICER) family (Figure 5.33). It appears that deletion of either could modulate the lifetime of bound protein-CRE complex, at the same time partly preventing recruitment of Q regions interactors. Another study has found, on the contrary, that Q2 deletion greatly decreases association to DNA and lifetime of complexes bound at promoters (151). However, these observations were made with less sensitive techniques (245) and selecting for long-lived complexes. The latter is likely a result of CREB being held at transcriptionally active promoters by the general transcription machinery and coactivators,

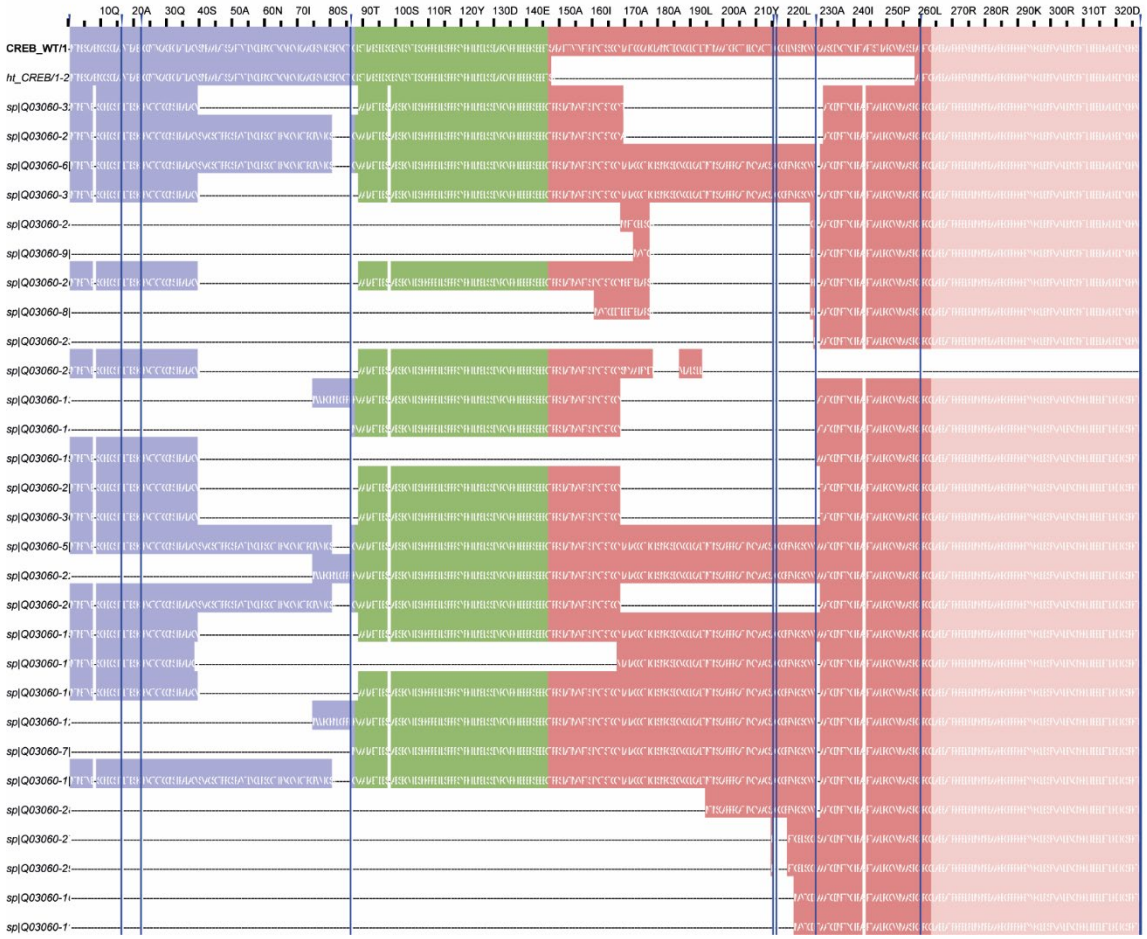


Figure 5.33: Multiple sequence alignment of CREB, htCREB and human CREM family members. htCREB (second line) is missing a large Q2 segment compared to CREB (top line), and many human CREM family members are missing parts of Q2 and Q1. The bZIP and the KID are more often kept. Q1 is shown in blue, KID in green, Q2 in red, bZIP in beige.

which require Q2 for recruitment, and therefore these findings are complementary rather than contradictory.

Both Q-rich regions have areas of local hydrophobicity comparable to those involved in GCN-4 fuzzy complexes (76) and are reported to interact with each other (154). Furthermore, their deletion limits the propensity of purified CREB to aggregate. Therefore, it is possible these regions help prevent DNA binding via non-specific binding either intra-molecularly, potentially forming structures that inhibit association to DNA, as predicted by AlphaFold, or inter-molecularly, causing CREB to preferentially interact with itself rather than DNA. Favouring this, low-complexity domains of SP1 similar in composition to CREB's Q-rich domains are known to cooperatively increase presence at promoters by forming self-interacting hubs, but exclusively with similarly composed regions, and can recruit RNA polymerase II as well (71). Furthermore, these Q regions can indeed form beta-hairpins (154), which are involved in TFIID recruitment, but that AlphaFold places near the bZIP DBD, so they could also prevent DNA binding. Still, more experiments are required to unravel how each Q-rich region affects binding to DNA and the target search process.

How elements of DNA secondary structure affect target-search is also misunderstood. Preliminary results from the previous Chapter showed that dissociation is faster from hairpin CRE than linear CRE. Although factors are reported to be able to bypass such obstacles (47), these might affect CREB and force it to dissociate. Here, preliminary results with the bZIP show that CREB's other domains do not affect this behaviour. Further experiments are needed to confirm this and to investigate exactly DNA secondary structure elements mechanistically affect the target search process.

Chapter 6

Investigation of CREB dimerization and the stoichiometry of target search

It is common knowledge that bZIP transcription factors bind to target DNA as a dimer, as shown in structural studies (135). CREB is stably bound as a homodimer (135,246), but bZIPs can both homodimerize and heterodimerize with varying ranges of affinity (136). The Jun-Fos dimer is a famous example of a bZIP heterodimer.

This dimerization behaviour is likely to matter in the bZIP target-search process, itself understood in terms of the facilitated diffusion model (39–44) (Figure 1.4), discussed at length in the introductory Chapter 1. This simple kinetic scheme for target search in Figure

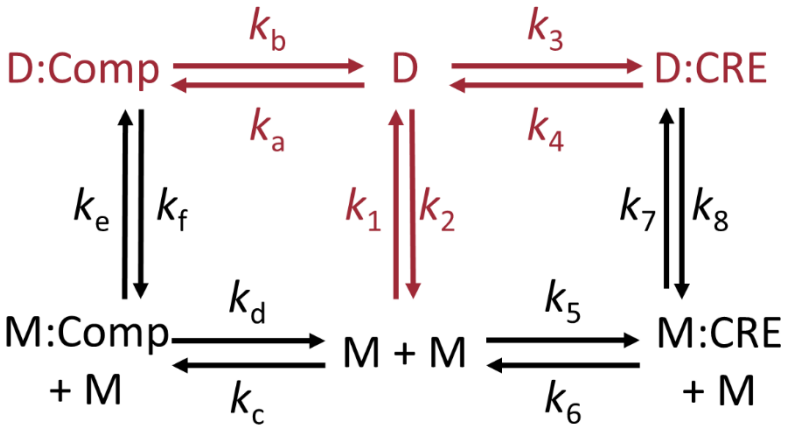


Figure 6.1: Simplified kinetic scheme of CREB binding target and non-target DNA. All isolated reaction schemes and associated kinetic rate constants of CREB binding competitor (Comp) and target (CRE) DNA. Black routes are preferred by monomer pathway, dark red routes by the dimer pathway.

2.1 (reused here as 6.1 for convenience) shows how bZIP could either find and bind its target as a monomer and subsequently dimerise on the target (monomer pathway) or dimerise before binding DNA altogether. This kinetic scheme simplification ignores some of the other facilitating processes of target search (such as intersegmental transfer, additional binding partners or phase separation) but is applicable in my isolated *in-vitro* setup using short oligos where said aspects are less likely to occur.

Multiple studies have suggested the monomer pathway is dominant for bZIPs, including the human Jun, Fos (159,160), ATF-2 (158) and the yeast Gcn4 (157,158) proteins. This could offer significant advantages, such as greater flexibility for the monomer to bind DNA, more probes to find target and then dimerise once one is found. In this scenario, dimerization would stabilise bound CREB. Yet, it is also possible that monomers are unable to form DNA-compatible bound coiled-coiled structures or enough to differentiate target sites from random DNA. These helices could propagate from dimer-induced helices, and dimerization could therefore act as a pre-requisite to DNA binding.

The concentration of these proteins in the cell, according to the OpenCell repository (183), is relatively low, and much lower than CREB (Table 6.1). A naïve expectation would be that this could favour the monomer pathway as there would be more monomer present. CREB should follow the monomer pathway too, since it was reported to be monomeric in solution (162). Yet, it was observed (163) and recently confirmed (155) that CREB has a low homodimerization K_D indicating high dimerization propensity. Although Gcn4's homodimerization K_D is 10-fold higher than CREB's (247), dimerization K_D for Jun:Fos heterodimers and ATF2 homodimers are comparable or higher to CREB's (136) (Table 6.1), so it appears high dimerization propensity is not a *sine qua non* condition for favouring the dimer pathway. Indeed, equilibrium concentrations of species do not give a complete picture

of mechanism – kinetic experiments are needed to understand fluxes through different pathways.

Preliminary evidence from the Shammas laboratory ((163) and Kuravsky, unpublished) with CREB bZIP domain suggests CREB does favour the dimer pathway, mainly because of:

1. low threshold for homodimerization,
2. similar bZIP and bZIP forced dimer behaviour, and
3. very low affinity of bZIP monomer for DNA, with a K_D above 100 μM and inability to distinguish target from non-target DNA well.

In this Chapter, I firstly explore CREB's and the bZIP standalone domain's dimerization behaviour at pH 7.4. As said in Chapter 5, the laboratory gathered bZIP data at pH 6.5 to match NMR conditions, but there is the possibility of pH-dependent behaviour as shown in Chapter 4. Therefore, data at pH 7.4 was taken to make fair comparisons. I then generated full-length CREB forced dimer and forced monomer mutants of CREB to probe the stoichiometry of target search. I show that forced dimers have near identical behaviour to full-length CREB, while forced monomers appear to not bind to DNA at all. DNA enhances dimerization but this does not imply that target search is performed as monomer.

Protein	Homodimerization K_D (nM)	Cellular concentration (nM)	Dimeric %	References
CREB	15	570	89	(136,155,183)
Jun / Fos	<10	150 / not detected	N/A	(136,183)
ATF2	10-50	73	56-77	(136,183)
GCN4	500	40	10	(247–249)

Table 6.1: Cellular concentration, homodimerization K_D and dimeric fraction of example bZIPs.

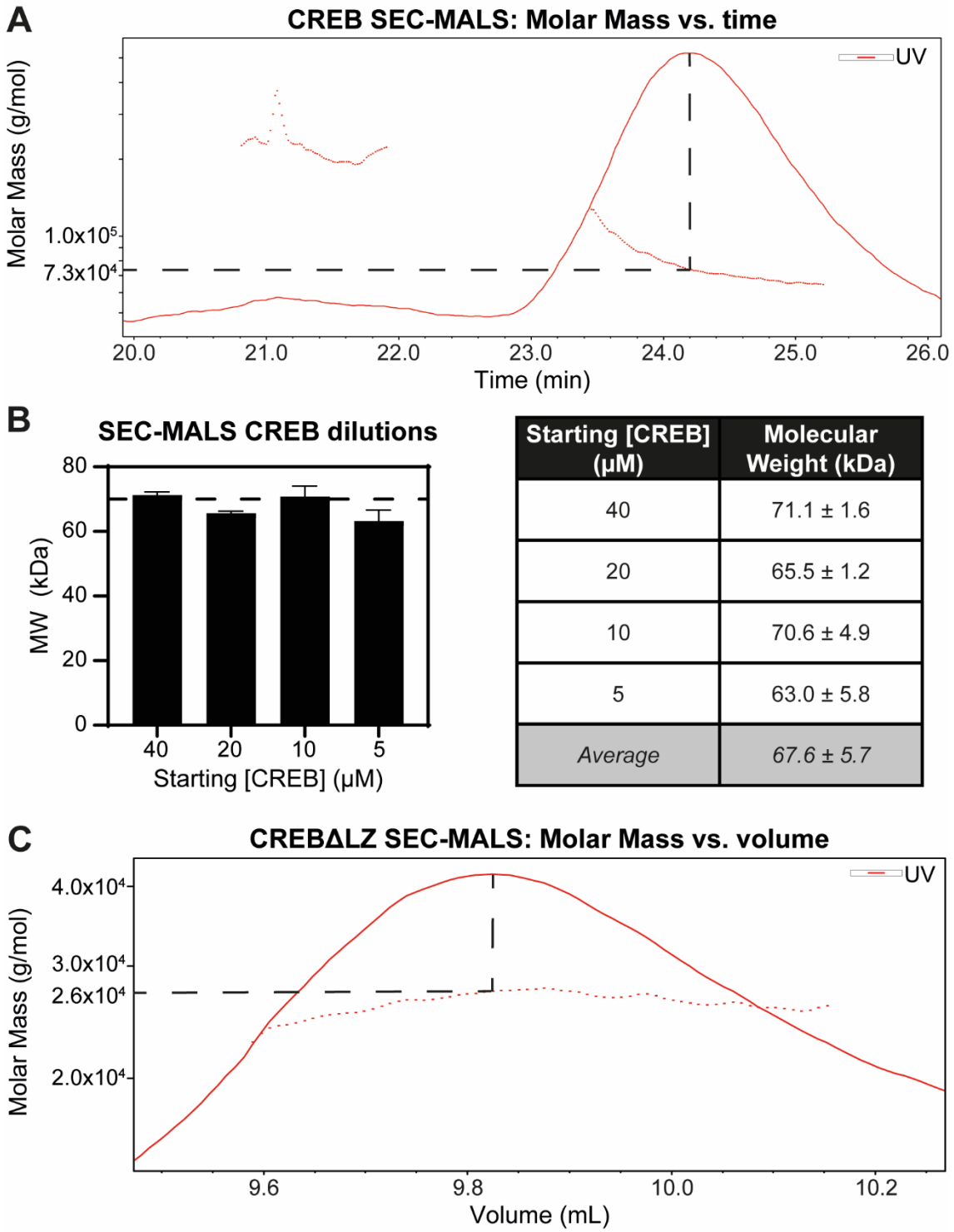
By extrapolation of the simple model to mimic CREB behaviour in the nucleus, I suggest that CREB favours the dimer pathway, the first bZIP factor reported to do so.

6.1. Characterisation of the homodimerization process of CREB and CREB bZIP

6.1.1. Native protein size determination methods suggest CREB is present as a dimer in solution.

Spectra from Size exclusion chromatography multi-angle light scattering (SEC-MALS), which can describe a protein's size, show a main peak at around 70 kDa and another, smaller peak at twice this size (Figure 6.2A). Serial dilutions of the sample did not change the size of the main peak (Figure 6.2B). Since pure CREB from the SUMO method has a molecular weight of 35491 Da, this signal fits well with a CREB dimer, and the weaker signal at high concentrations with a tetramer. As the lowest concentration achieved was 5 μM, it is safe to assume that the homodimerization K_D is in the hundreds of nM range or below, otherwise there would be some monomer signal. As a control, a forced monomer CREB (CREBΔLZ, see Section 6.2) also analysed by SEC-MALS and returned, as expected, a weight of 26.4 kDa (Figure 6.2C).

AUC data were more puzzling, with an apparent mixture of monomer and dimer signal at around 52 kDa (Figure 4.9). These findings match those reported before (162), where it was interpretated as a monomeric signal with some precipitation. However, this interpretation appears inconsistent with the robust SEC-MALS analysis. Analysis of a forced monomer or dimer version would be useful to understand this AUC signal.



Starting [CREB] (μM)	Molecular Weight (kDa)
40	71.1 ± 1.6
20	65.5 ± 1.2
10	70.6 ± 4.9
5	63.0 ± 5.8
Average	*67.6 ± 5.7*

Figure 6.2: SEC-MALS of CREB and CREBΔLZ. All runs are made in biophysical buffer, pH 7.4, at room temperature. **A.** SEC-MALS run of CREB, 40 μM, Superdex200. Two species are observed, one main at ~73 kDa and a minor at 150 kDa, consistent with dimer and tetramer respectively. **B.** Two-fold dilution SEC-MALS runs of CREB (starting at 40 μM) on Superdex200 all show a single species around the expected weight of the dimer. Below 5 μM, no signal is observed. **C.** CREBΔLZ (forced monomer) SEC-MALS run on Superdex75 shows a single species at 26 kDa, consistent with monomer.

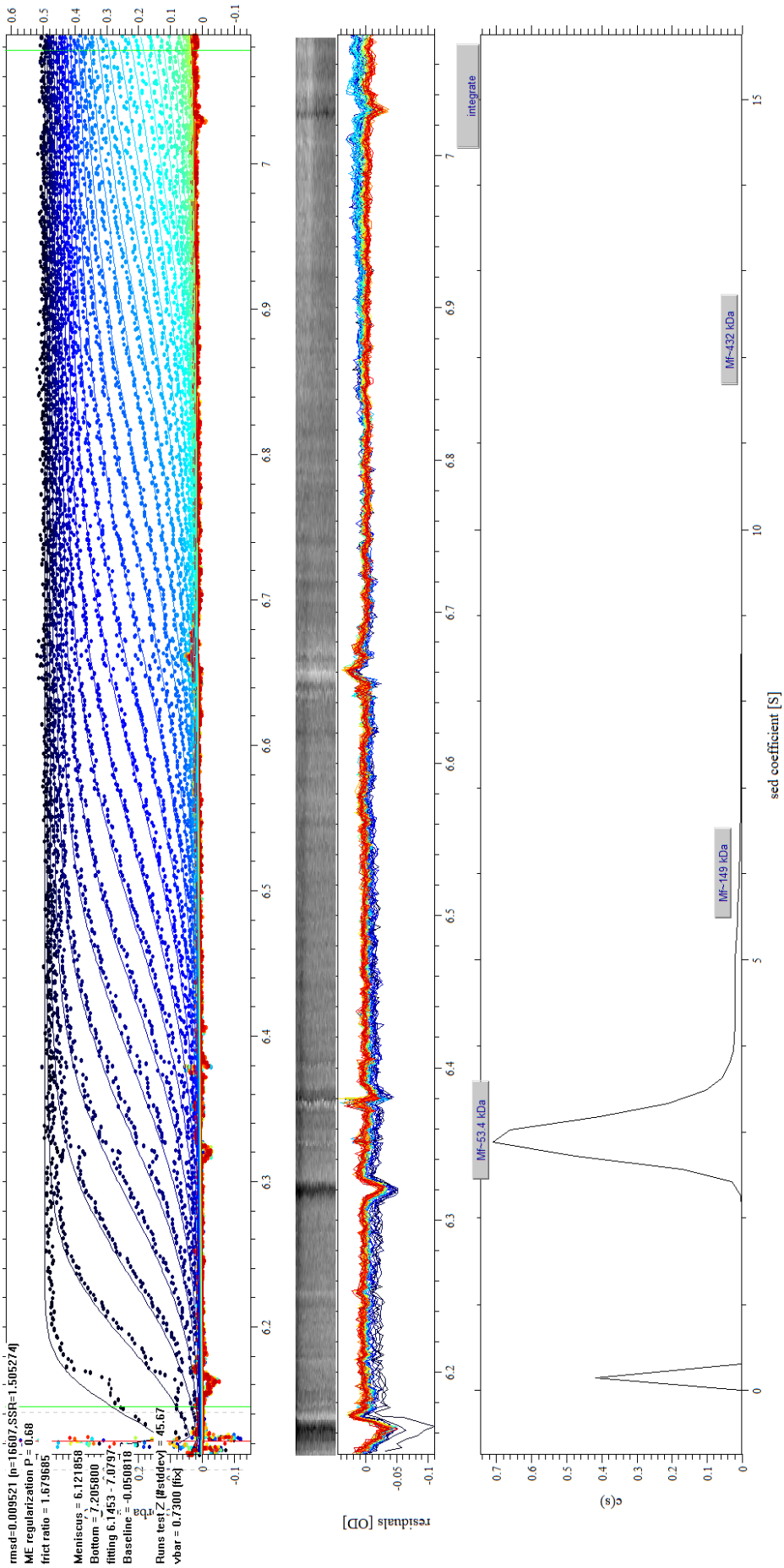


Figure 6.3: Analytical ultracentrifugation (AUC) CREB run. DNA. As previously reported one major species is observed (MW =53.4 kDa), confusingly in between the expected weights of dimer (70 kDa) and monomer (35 kDa). Run at pH 7.4.

6.1.2. Thermal unfolding and dissociation give an estimate of homodimerization affinity and thermodynamics.

CREB homodimers form coiled-coil helices (see Section 1) (131,135). If it is assumed that these secondary structure elements are concomitant with dimerization, then unfolding CREB will also inform on dimerization properties. Thermal unfolding was used, in conjunction with CD signal at 222 nm indicative of helicity, to calculate the CREB unfolded fraction (see Section 4.2), which, under this assumption, could be considered the monomeric fraction. This can be used to calculate the homodimerization affinity constant $K_{D,homo}$ at every temperature (see methods Equation 2.23) Since the natural logarithm of K_D for interactions are linearly dependent on temperature, as per the Gibbs free energy equations, the relationship can be extracted by plotting $\ln(K_{D,homo})$ against temperature.

As is seen from the plot, this relationship is only accurate between 40 °C and 70 °C where the transition is apparent. From extrapolation, $K_{D,homo}$ = (1 ± 6) nM at 25 °C (Figure 6.4C , Table 6.4), This provides a broad estimate rather than an accurate value, especially since the error on some of the values are so large. Nevertheless, the range is in the low nM, which shows high homodimerization affinity and that CREB is likely dimeric in solution, as per SEC-MALS data.

An advantage of this method is that thermodynamic values can be estimated too. ΔH° for homodimerization is (-274 ± 100) kJ.mol^{-1} and, at 25 °C and from the $K_{D,homo}$, ΔG° = (-50.4 ± 0.3) kJ.mol^{-1} and ΔS° = (-0.75 ± 0.33) kJ.K^{-1}.mol^{-1}. The specific heat change cannot be reliably estimated however, with ΔC_p = (3.5 kJ.kg^{-1}.K^{-1} ± 25) kJ.kg^{-1}.K^{-1} or a 700% uncertainty What can at least be extracted is that homodimerization is heavily favoured enthalpically and thermodynamically at biologically relevant temperatures.

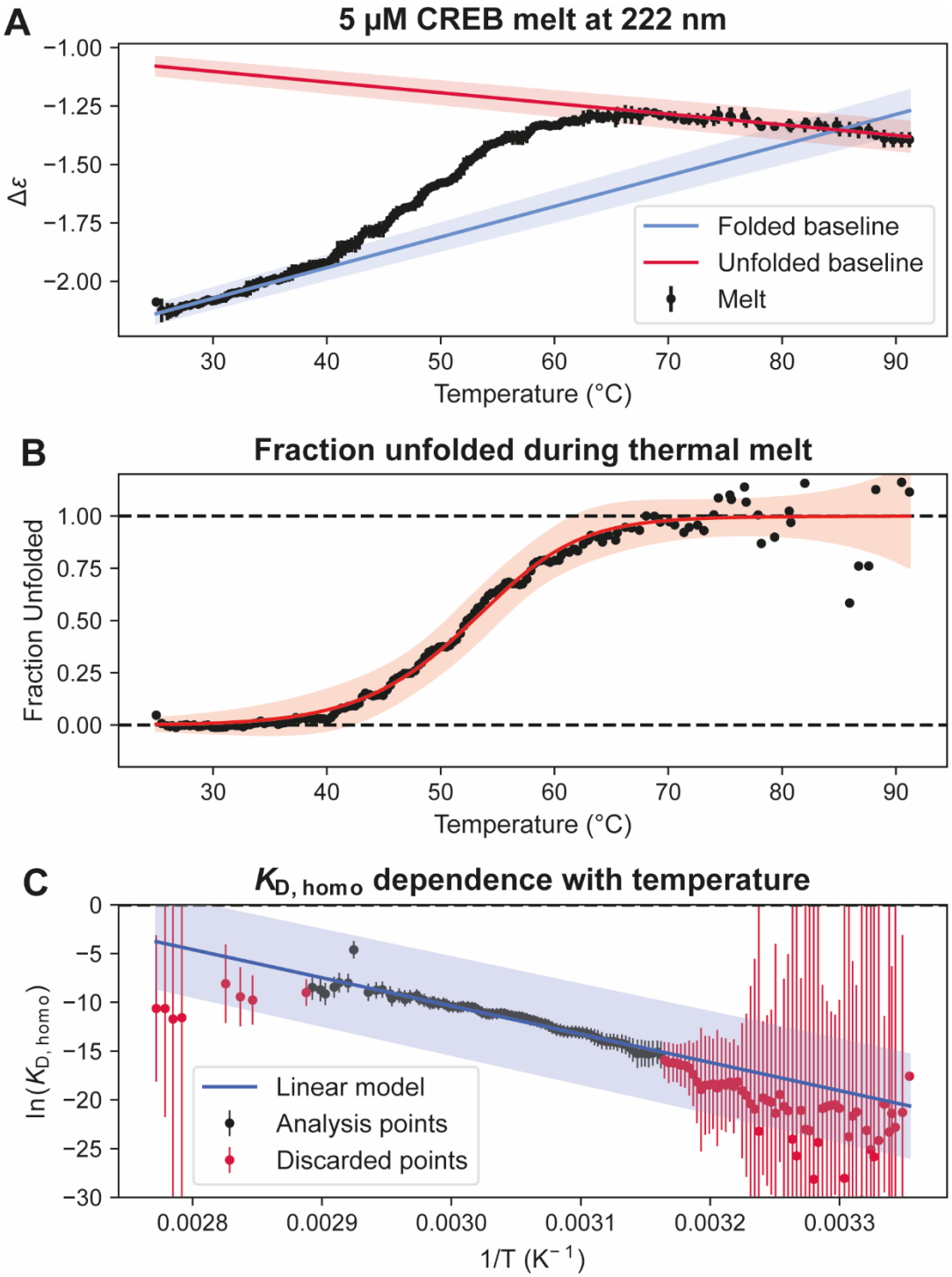


Figure 6.4: CD thermal melt of CREB to extract $K_{D,homo}$. A. CD melt at 222 nm, pH 7.4, with extrapolated unfolded and folded baselines. **B.** Fraction unfolded with fit to extract the T_m = (52.8 ± 0.5) °C. **C.** Predicted linear relationship between temperature and $K_{D,homo}$ to extrapolate value at 25 °C.

6.1.3. Perturbation by dilution is a viable method to find CREB homodimer formation kinetics and affinity.

It is possible to study homodimerization using relaxation kinetics after causing a perturbation by 11-fold dilution of a labelled CREB solution using stopped-flow apparatus and monitoring the relaxation to equilibrium. Labelled CREB was generated as described previously (Section 4.4.2). The relaxation traces were fit to derived equations (see (176) and methods' Equation 2.11) to find both the $K_{D,homo}$ and the dimerization dissociation rate constant, k_2 (see kinetic scheme, Figure 6.1). I found that fitting individual kinetic traces did not give reliable estimates for biophysical quantities (except k_2), as the values given had both large variability and errors (Figure 6.5). However, simultaneous curve fitting dilutions from multiple starting concentrations provided additional information, which allowed for reliable fits. Estimated values were robust to the number of curves fitted (even pairwise fitting), and regardless of the pair combinations, $K_{D,homo}$ and k_2 were always in a similar range: low tens of nanomolar and $0.1s^{-1}$, respectively (Figures 6.6, 6.7). It is important that these dilutions are taken on the same day, with identical instrument setup, as this allows for the fitting variable ΔF, a measure of fluorescent change adjusted for protein concentration, to be also shared between the curves. Otherwise, it must be individually fit to each trace, which can add avoidable uncertainty to the result.

The estimates were robust to the exact fluorophore and linear/logarithmic stopped-flow setup (Figure 6.9). There was some variability as some samples gave two-fold higher $K_{D,homo}$ values. Notably, there was a sharp signal change at the beginning of each trace below 0.1s, which are excluded from fits. These are likely artefacts due to fluorophore adjustments post dilution or extended lag time from the 1 in 11 stopped-flow dilution setup. Regardless, the inclusion of this change only slightly increased the output values (Figure 6.8).

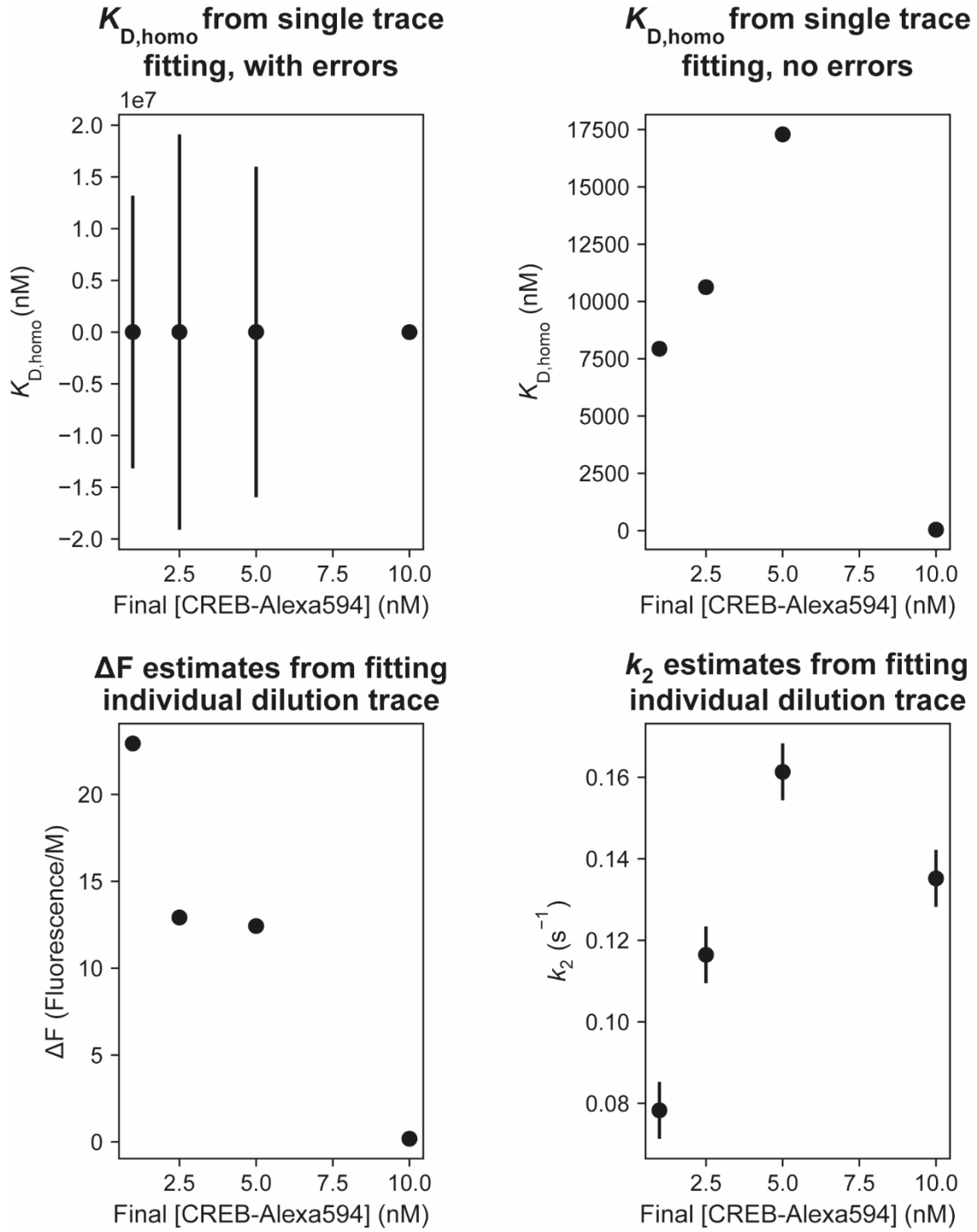


Figure 6.5: There is great variability in extracted parameters from individual CREB-Alexa594 dilution trace fits. All parameters show great variability from the fits (sometime extreme, if compared between the highest and lowest concentration trace) and, in the case of $K_{D,}$ there are very large errors. Traces were taken at pH 7.4, 25 °C.

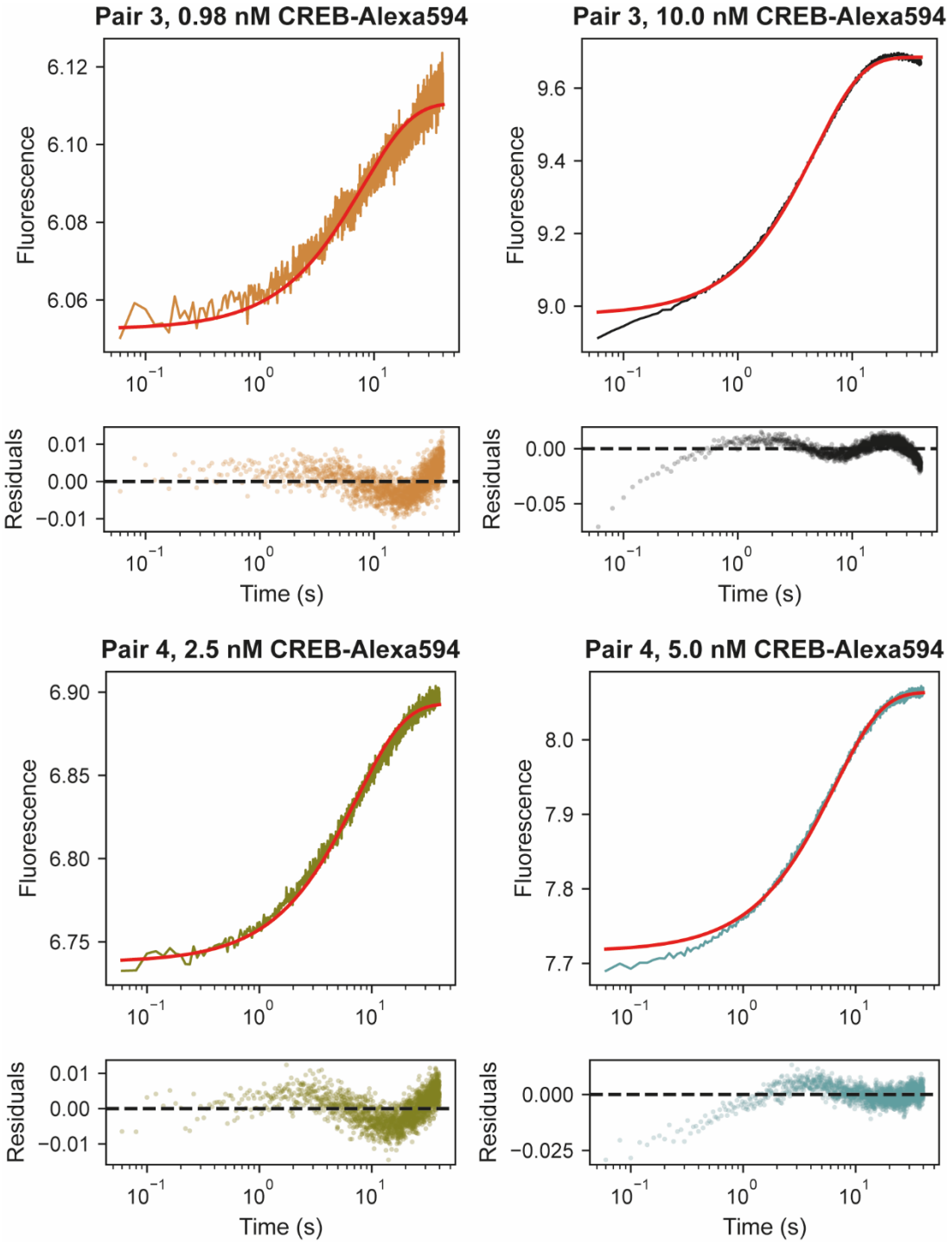


Figure 6.6: Example of pairwise CREB-Alexa594 dilution relaxation fits. Fits (in red) are calculated to give the best fit to each pair, here pairs 3 and 4 from run 1. Traces taken at pH 7.4, 25 °C.

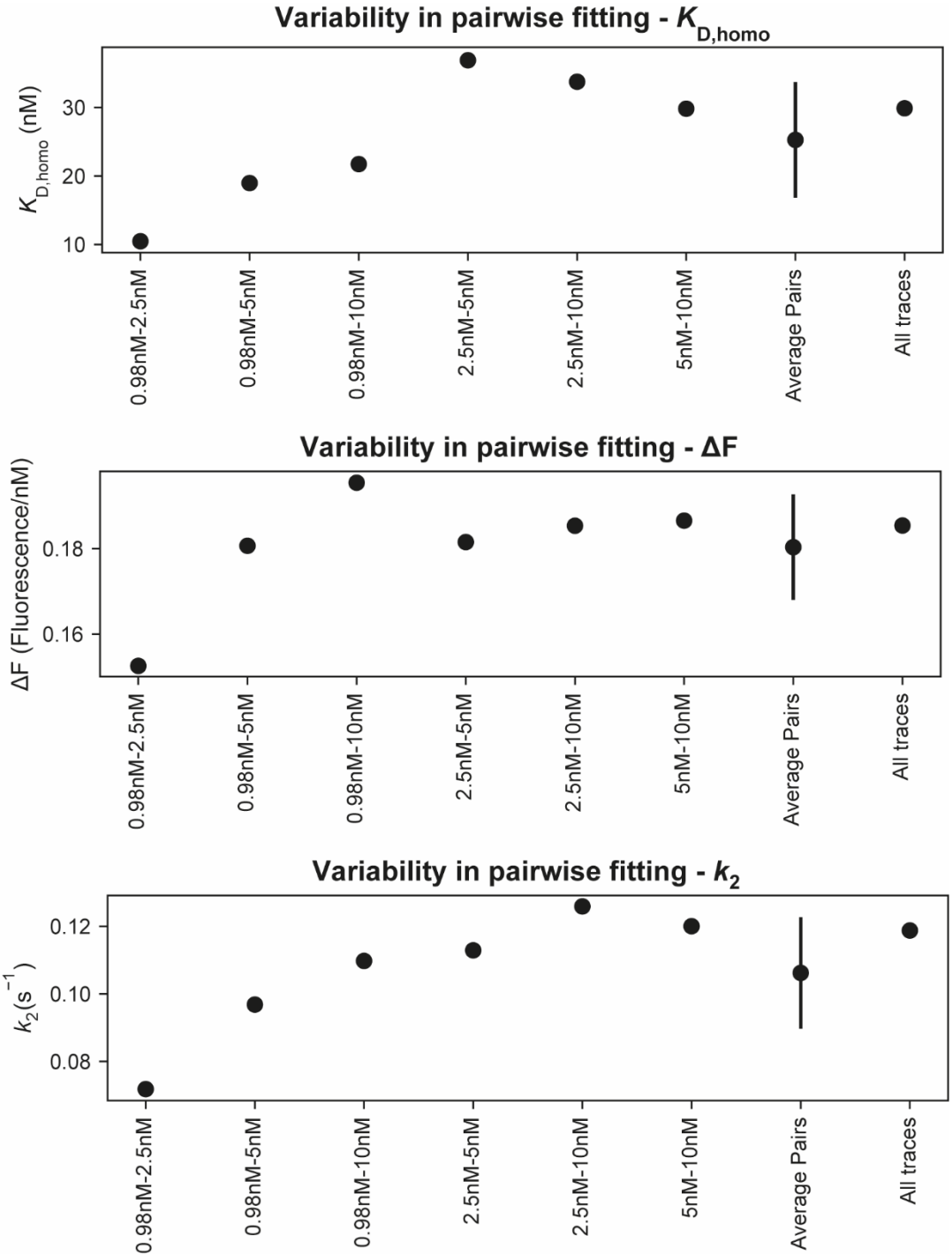


Figure 6.7: Variability of fitted parameters using pairwise fitting. Compared to individual fitting (Figure 6.5), all pairs give similar values, especially for the dimeric dissociation rate constant k_2. $K_{D,homo}$ is greatly improved, although it is still the most unreliable variable. Lastly, the average result of all pairs is always within error to the value obtained by simultaneous fitting of "all traces". Conditions: pH 7.4, 25 °C.

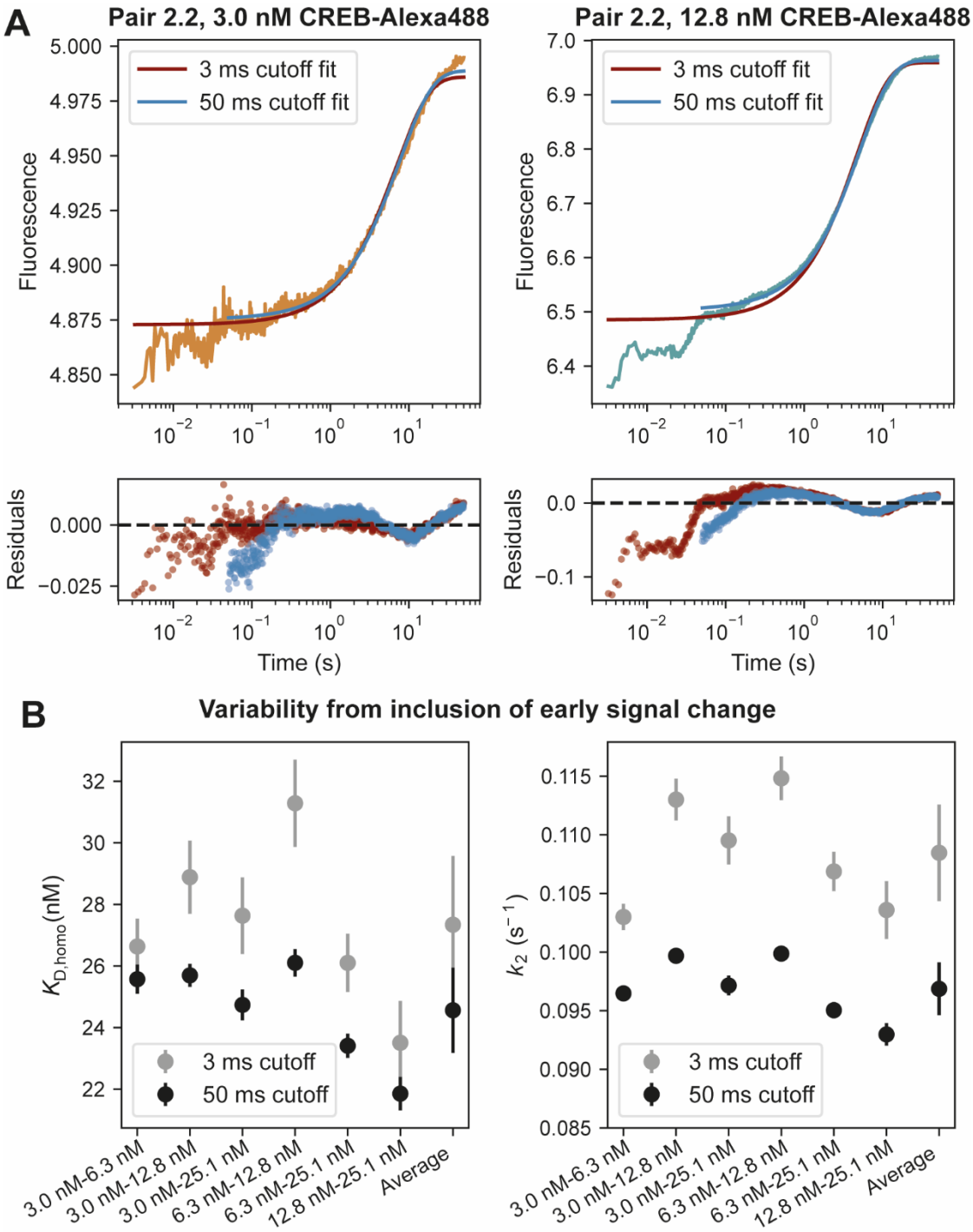


Figure 6.8: Discarding of the early signal change in fluorescently labelled CREB dilution relaxation traces has little effect on fitting. Conditions: pH 7.4, 25 °C. **A.** Example paired traces with a 3 ms cutoff, with both 3 ms and 50 ms cutoff fits. **B.** Results for all pairs in a single run. Including the change slightly increases the output values, but there is less than 10% different in the averaged values for all pairs.

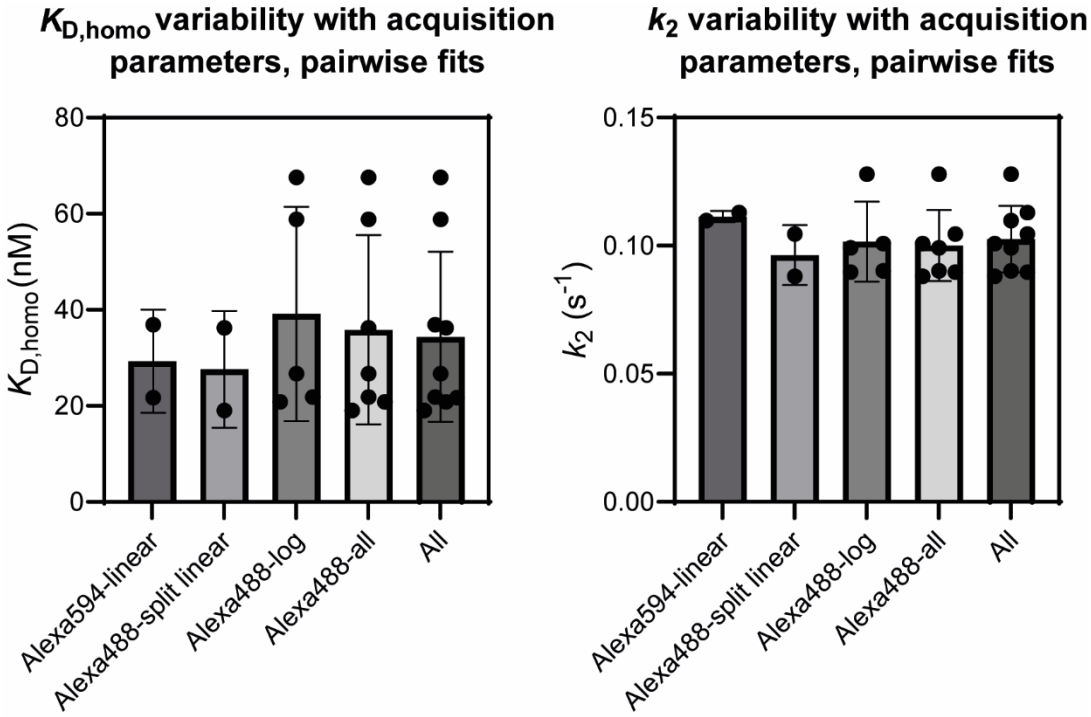


Figure 6.9: Variability of fitted parameters from fluorescently labelled CREB dilution relaxation kinetics based on acquisition method. n fitting. Overall, there is no effect on either k_2 or $K_{D,homo}$ if the acquisition method or fluorophore is different at pH 7.4, 25 °C. There are two datapoints that appear slightly higher but, regardless, most points are in the same range.

Using each trace only once (e.g. if 4 traces were taken on a day, using only 2 pairs out of all possible six) gives a k_2 = (0.10 ± 0.01 s^{-1}) (Table 6.5) and $K_{D,homo}$ = (35 ± 18) nM (Figure 6.9, Table 6.4). The homodimerization association rate constant k_1 can be calculated assuming a two-state model using $k_1 = \frac{2k_2}{K_{D,homo}}$ = (6 ± 3) $\mu M^{-1}s^{-1}$ (Figure 6.14, Table 6.5).

6.1.4. Various equilibrium titration methods attempted to obtain $K_{D,homo}$

Attempts to obtain $K_{D,homo}$ through fluorescence anisotropy titration with urea were unsuccessful since a substantive "folded baseline" was not available. When attempted with CREB bZIP a clear transition was observed (163). However, 2 µM bZIP was used instead of the 50 nM used for the (more costly) CREB labelled with Alexa Fluor™ 594, so it is possible the technique may be fruitful if attempted at higher protein concentrations.

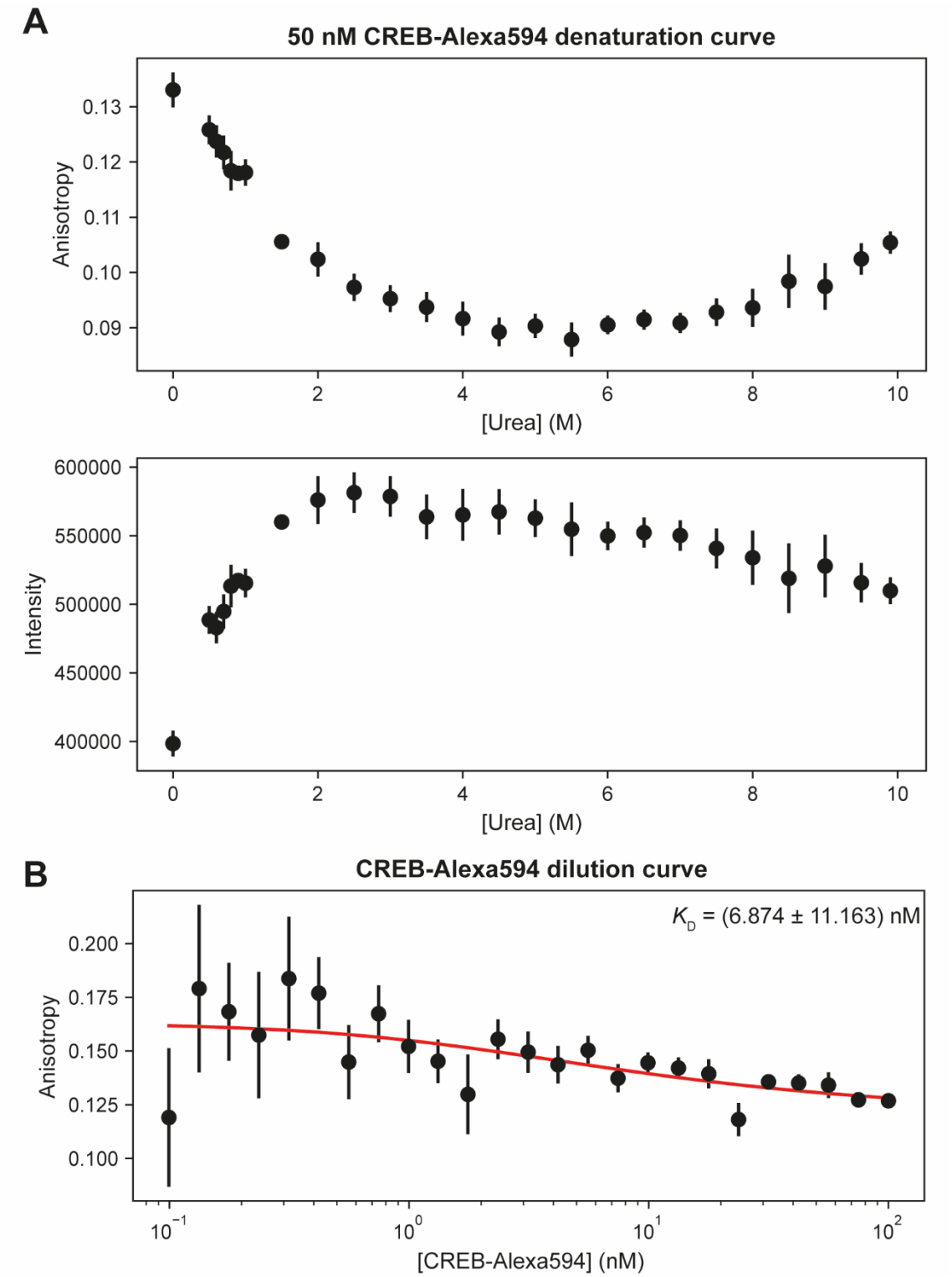


Figure 6.10: Anisotropy attempts with urea titration and CREB-Alexa594 dilution. A. urea titration of 50 nM CREB-Alexa594, both anisotropy (top) and intensity (bottom). Intensity suggests an abrupt signal change at very low urea concentration, perhaps due to the low concentration of CREB used, and makes this dataset unusable. **B.** CREB-Alexa594 anisotropy dilution curve. A lower plateau is not reached which limits the accuracy of this method, although the $K_{D,homo}$ extracted is in the low nM, in a similar range from the other methods. Titrations were carried out at pH 7.4, 25 °C.

Labelled CREB anisotropy upon dilution had a similar issue. (Figure 6.10B). This means the extracted K_D (via Equation 2.6), is likely underestimated and has 200% uncertainty. Nevertheless, it remains in a range near the value extracted from relaxation kinetics experiments (Figure 6.11, Table 6.4). It is likely that, if repeated over a larger concentration range, both the estimate's accuracy and precision would improve.

Lastly, fits of anisotropy titrations of binding to CREh are imperfect due to the K_D of binding being below the $K_{D,homo}$. It is possible to account for this with a more complex fit (Equation 2.5 in methods), but it requires that either K_D of binding to CRE or $K_{D,homo}$ be set (Figure 6.11A). The kinetics-derived K_D can be used for the former since higher CREB concentrations are used in the kinetics experiments, so CRE binding is probed more specifically. The relaxation kinetics $K_{D,homo}$ can be used for the latter. Fixing the K_D returns a very low homodimer $K_{D,homo}$ but has a high relative uncertainty (Figure 6.11C, Table 6.4). It may not be entirely trustworthy as doing the reverse (fixing $K_{D,homo}$) to the relaxation kinetics value returns a lower K_D than with kinetics (Figure 6.11B).

6.1.5. Investigation of bZIP homodimerization at pH 7.4 using urea unfolding and refolding kinetics methods

CREB bZIP, like full-length CREB, forms strong homodimers and their affinity are best assessed using kinetic and/or single molecule approaches. A successful strategy for determining bZIP homodimerization affinity and kinetics using stopped flow refolding kinetic studies has already been developed (163). Briefly, bZIP has a fluorescent tyrosine that becomes buried in the helix that forms upon dimerization, so native

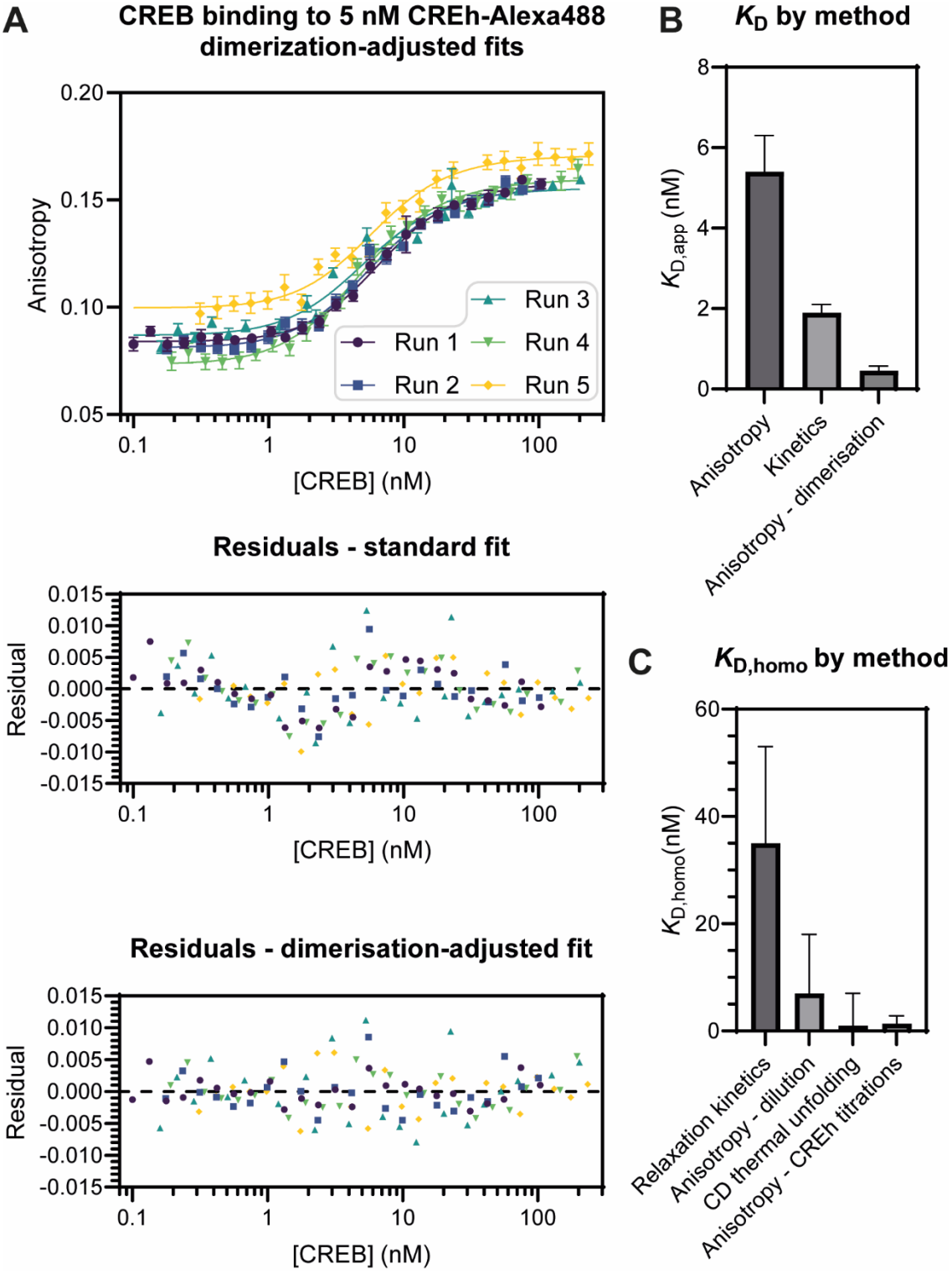


Figure 6.11: Fits of CREB binding CREh anisotropy titrations accounting for dimerization; K_D to CREh and $K_{D,homo}$ comparison by method. dilution. Conditions: pH 7.4, 25 °C. **A.** CREB binding to labelled CREh-Alexa488 anisotropy titrations, fit with a model accounting for DNA-induced dimerization (top). Residuals appear slightly better than with a non-dimerization fit. **B.** Kinetics K_D to CREh, assuming a two-state model, is nearer that of the better-fit titrations. **C.** $K_{D,homo}$ is variable based on method but is always in the low nanomolar. The relaxation method is 3-4-fold higher than estimates from other methods.

fluorescence can be used as a probe for dimerization. The dimer fraction can be modulated either through unfolding (dilution into urea) or refolding (diluting bZIP in 8 M urea into lower urea concentrations). Kinetic traces are fit to a single exponential decay function or Equation 2.13 (see methods) to extract estimates of k_u (unfolding) or unfolding K_D and k_f (refolding) respectively. Indeed, a single exponential decay fit was just as good as the more complex Equation 2.12 for unfolding traces (Figure 6.12B), likely because dilution into large urea concentrations prevent refolding. The logarithm of the extracted parameters is then plotted against [urea]) to enable extrapolation to the desired experimental conditions (0 M urea) – also called a chevron plot, which is what this experiment is referred to for convenience. These can then be compared to $K_{D,homo}$, k_2 and k_1 respectively, assuming folding is an approximation of dimerization.

Above 5 μM bZIP, the signal change is sufficient to accurately analyse the kinetic traces. This indicates that the observed rate constants reflect unfolding alone. In contrast, refolding traces are not well fit by simple hyperbolic equations that are appropriate for irreversible binding. This indicates the need for a more complex model that includes the unfolded-refolded equilibrium and provides an opportunity to extract K_D as well as the folding rate constant (Figure 6.13). Extracted parameters are well-fitted by linear models, with extrapolated values at 25 °C from triplicates in Tables 6.2 and 6.3. These values are like those obtained at pH 6.5 (163) (Figure 6.14, Tables 6.4. 6.5), showing pH change from 6.5 to 7.4 has little effect on bZIP homodimerization. Even the kinetic m-values (Section 2.3.4.d) are almost identical, with $m_{equil} = m_f + m_u = (1.2 \pm 0.2)$ kcal·mol^{-1}·M^{-1}.

Comparing these values to those obtained via perturbation for CREB in the previous section shows that the bZIP's $K_{D,homo}$ is 4-fold lower than CREB's. The unfolding rate constant k_u is 7-fold higher than the dimerization dissociation rate constant k_2, while the association rate constant k_f is within error of the dimerization association rate constant k_1

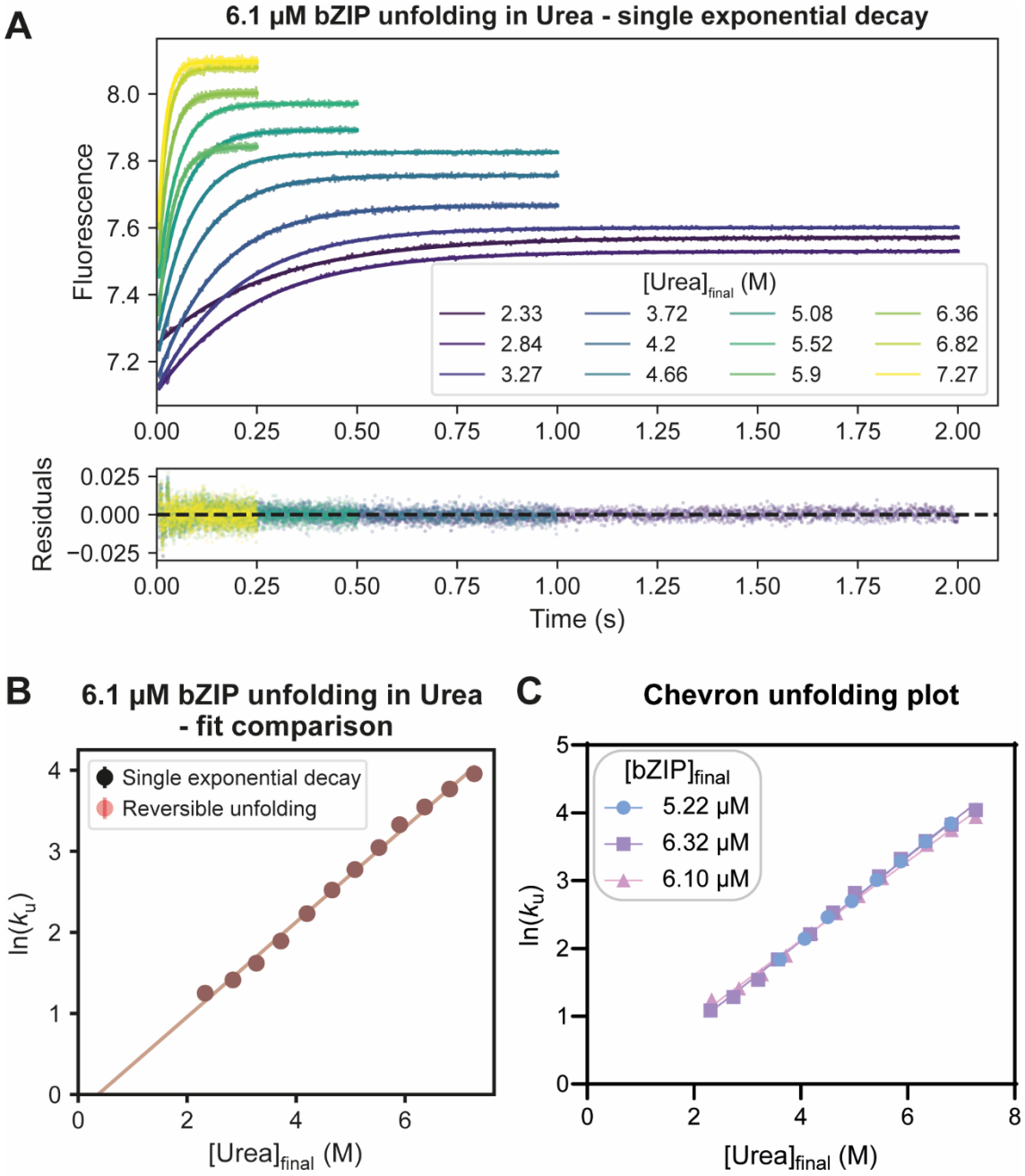


Figure 6.12: bZIP unfolding in urea experiments. Conditions: pH 7.4, 25 °C. **A.** Example intrinsic tyrosine fluorescence traces showing bZIP unfolding into different urea concentrations. Amplitudes gradually decrease and half-times increase with decreasing urea concentration. **B.** Traces fit just as well to single exponential decay equations than more complex models, likely because excess urea prevents refolding. **C.** Results of all three unfolding experiments. The value for each experiment is extracted by linear extrapolation.

(Figure 6.14, Tables 6.4, 6.5). This suggests that the CREB complex, despite having lower affinity for itself, is longer-lived. However, this is not an exact comparison due to the different methods used. Moreover, folding and unfolding can be cooperative in a way CREB's perturbation kinetics will not catch, and urea might trigger a more brutal unfolding,

therefore explaining the contrasting k_2 and k_u values. Yet, it still offers a starting point for comparisons, assuming folding is linked to dimerization.

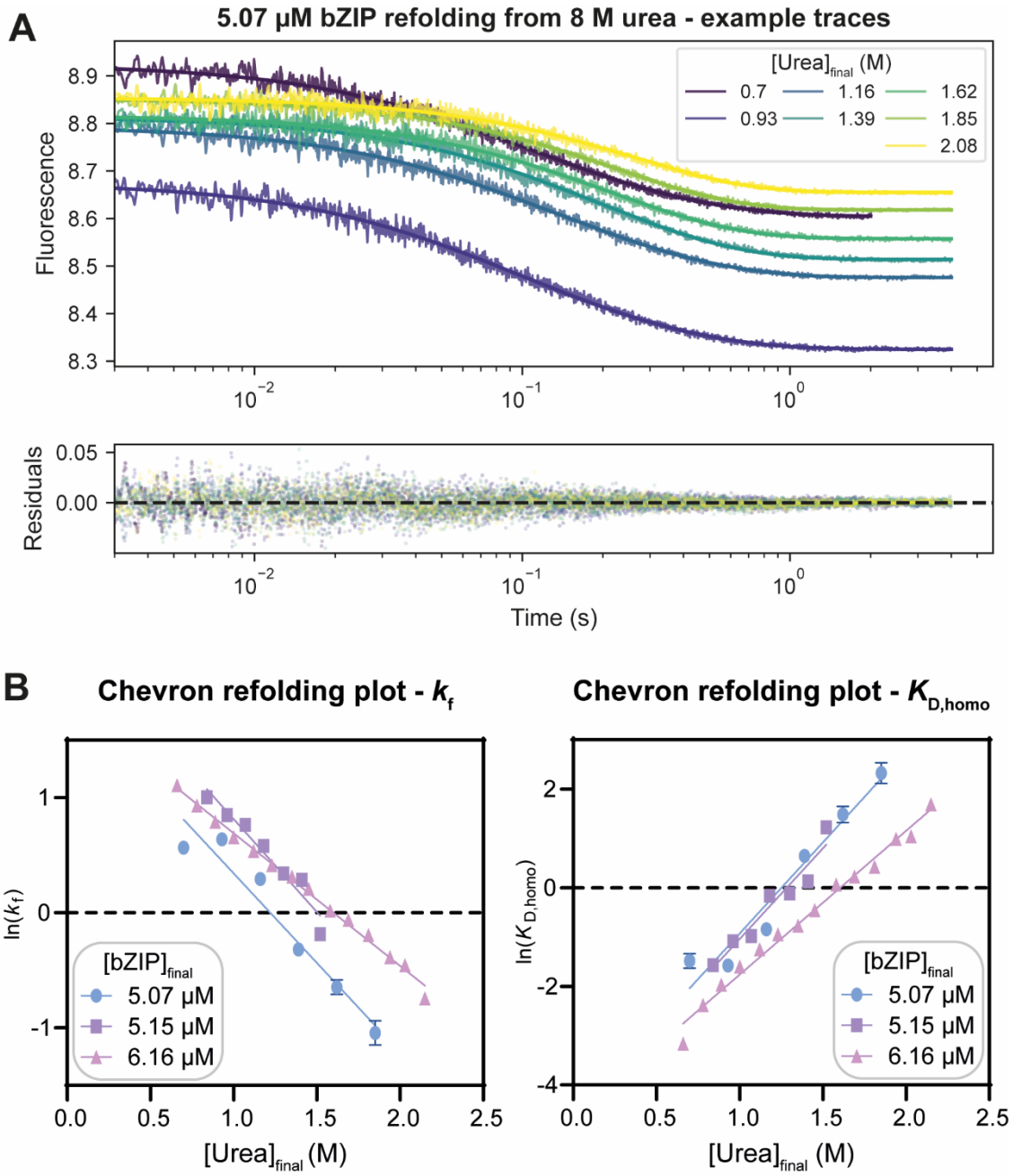


Figure 6.13: bZIP refolding from 8 M urea experiments. Conditions: pH 7.4, 25 °C. **A.** Example of intrinsic tyrosine fluorescence traces showing refolding bZIP traces, fit well by a model accounting for dynamic unfolding and refolding. Amplitudes gradually decrease and half-times increase sharply with increasing urea concentration, and $[urea]_{final}$ must remain very low **B.** The model can extract both K_D of unfolding (which is assumed to be near $K_{D,homo}$) and k_f, and the value for each experiment is extracted by linear extrapolation.

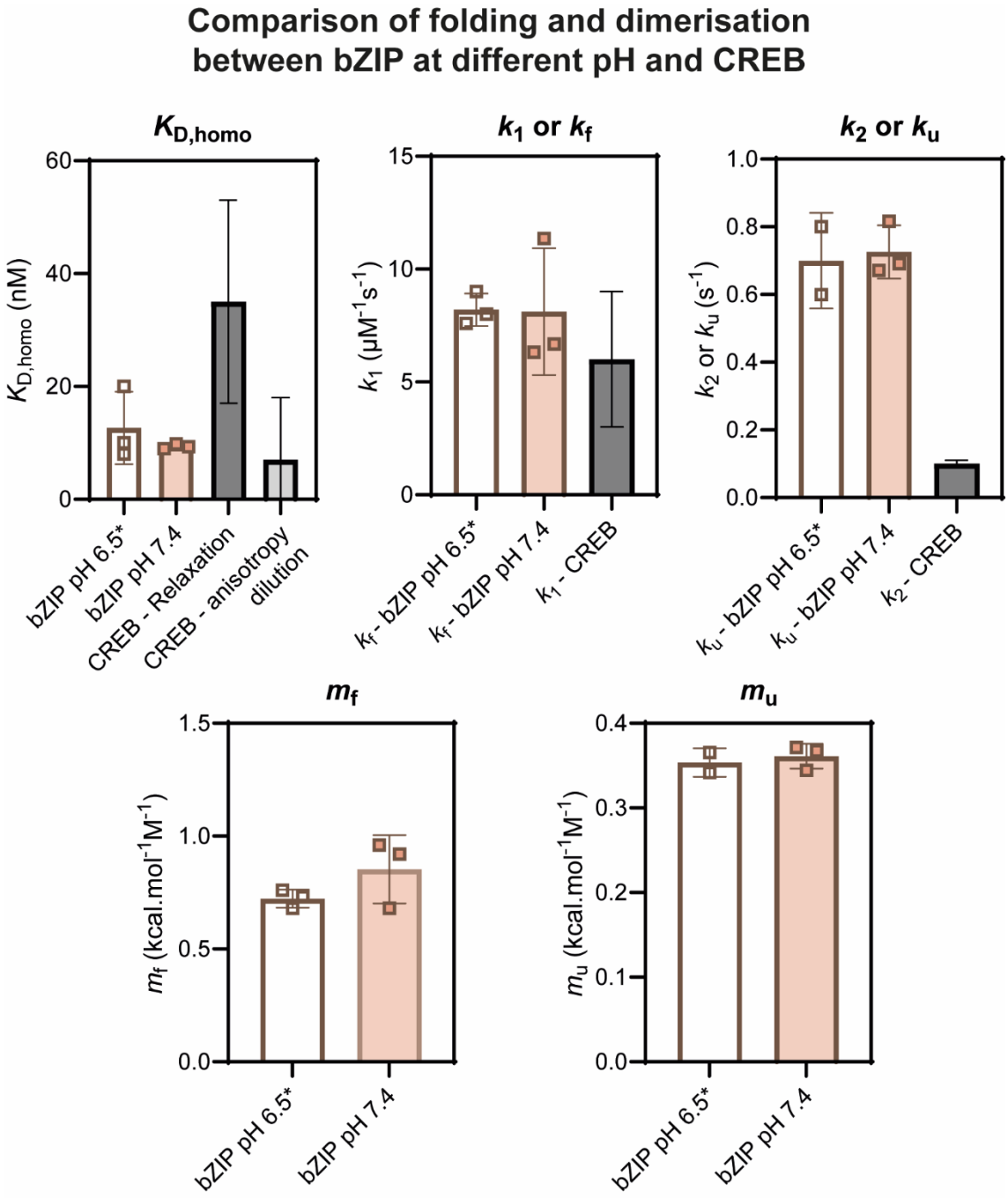


Figure 6.14: Comparison of bZIP's homodimerization parameters at pH 6.5 and 7.4 with CREB's. Clockwise: bZIP's $K_{D,homo}$ obtained via unfolding and refolding kinetics does not change between pH 6.5 and 7.4, and is close to CREB's value from some experiments (e.g. anisotropy dilution) but below that of the relaxation kinetics method. Association rate constant k_1 (or folding rate constant k_f) are within error of each other regardless of mutant or condition, but CREB's dissociation rate constant is much lower than bZIP's unfolding rate constant. m_u and m_f are very similar for bZIP between pH 6.5 and 7.4. All values obtained at 25 °C. * data from Dr. Kelly (163).

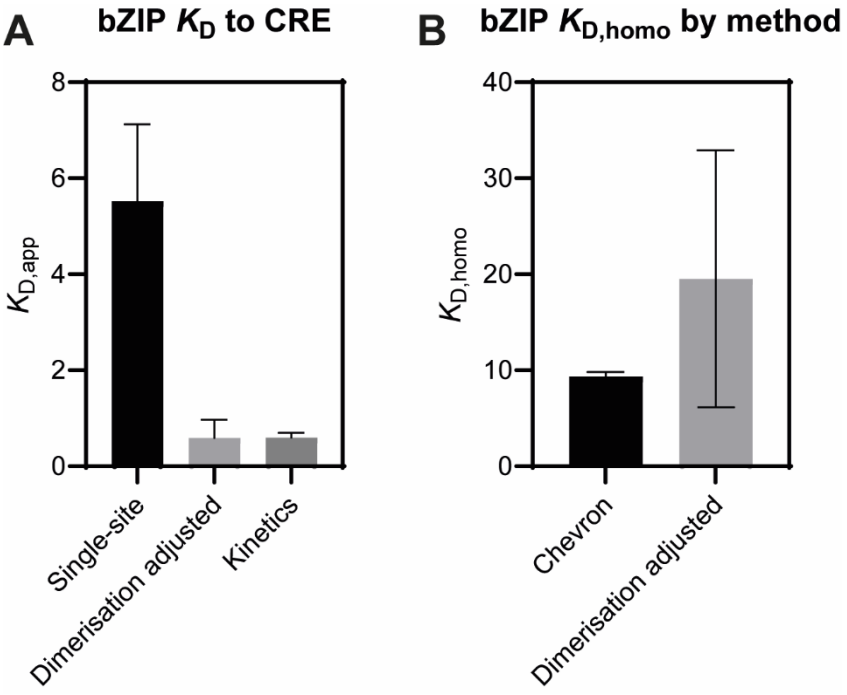


Figure 6.15: Comparison of bZIP's K_Ds of dimerization and binding to CREh by method. Conditions: 25 °C, pH 7.4. **A.** K_D of binding to CREh extracted by fitting anisotropy titrations accounting for dimerization is within error that of the kinetics' one. **B.** $K_{D,homo}$ extracted by fitting CREh anisotropy binding titrations with dimerization-adjusted models (fixing the K_D of binding to CREh) has large associated uncertainty and variability, but is within error to the value calculated from unfolding and refolding kinetics. This shows that, although monitoring folding and unfolding, the range it estimates is credible.

[bZIP] (μM)	$K_{D,homo}$ (nM)	k_f ($\mu M^{-1}s^{-1}$)	m_f ($kcal.mol^{-1}M^{-1}$)
5.07	9.8 ± 5.9	6.7 ± 1.7	0.92 ± 0.12
5.15	9.0 ± 4.9	11 ± 2	0.96 ± 0.1
6.16	9.3 ± 1.3	6.3 ± 0.2	0.68 ± 0.01
Average	**9.4 ± 0.4**	**8 ± 3**	**0.85 ± 0.15**

Table 6.2: Experimental results of bZIP kinetic refolding. 25 °C, pH 7.4. Mean with standard deviation.

[bZIP] (μM)	k_u (s^{-1})	m_u ($kcal.mol^{-1}M^{-1}$)
5.22	0.67 ± 0.03	0.371 ± 0.005
6.32	0.69 ± 0.04	0.368 ± 0.006
6.1	0.82 ± 0.05	0.344 ± 0.007
Average	**0.73 ± 0.08**	**0.36 ± 0.01**

Table 6.3: Experimental results of bZIP kinetic unfolding. 25 °C, pH 7.4. Mean with standard deviation.

Method to determine $K_{D,homo}$ in nM	CREB	bZIP pH 7.4	bZIP pH 6.5*
Folding kinetics	-	9.4 ± 0.4	13 ± 6
Complex fit of CREh titrations	1.4±1.5	20 ± 13	-
Relaxation kinetics	35 ± 18	-	-
CD thermal melt	1 ± 6†	-	2 ± 3
Anisotropy dilution	7 ± 11†	-	4 ± 1
FRET (from (155))	-	33 ± 13	-
Mass photometry (from (155))	15 ± 3	15 ± 3	-

Table 6.4: Comparison of CREB's and bZIP's $K_{D,homo}$ from different methods. Regardless of construct or condition, there was much variability in the $K_{D,homo}$ extracted, although it is almost always in the low nM range, around 10-50 nM. CD and anisotropy-based methods tend to extract lower values. bZIP's folding kinetics value is within error between pHs. Although there is a 4-fold difference between the chevron value for bZIP and relaxation kinetics for CREB, another study found similar number for both from mass photometry, which was half that of fluorescently labelled bZIP's value from FRET (155). * data from (163). Values are averaged triplicates or duplicates for † with standard deviation.

Kinetic parameter	CREB (relaxation kinetics)	bZIP pH 7.4 (folding kinetics)	bZIP pH 6.5 (folding kinetics)*
k_1 (k_f for urea) ($\mu M^{-1}s^{-1}$)	6 ± 3	8 ± 3	8.2 ± 0.7
k_2 (k_f for urea) (s^{-1})	0.1 ± 0.01	0.73 ± 0.08	0.7 ± 0.1

Table 6.5: Comparison of CREB's and bZIP's homodimerization kinetics. bZIP folding kinetics values are near identical at pH 6.5 (* from (163)) and 7.4. While the association rate constant k_1 is within error for all mutants and conditions, the dissociation rate constant k_2 is 7-fold smaller for CREB's, but this parameter could be skewed by the method used. All values are averaged triplicates with standard deviation at 25 °C.

As mentioned in the previous section 6.1.4, it is possible to use more complex models to account for dimerization during CREh binding, provided either bZIP's K_D to CREh or $K_{D,homo}$ is fixed. There is a high variability of $K_{D,homo}$ if fixing K_D using the kinetics value, but it is in the vicinity of the unfolding/refolding kinetics value (Figure 6.15). Similarly, K_D of binding to target is within error to the kinetics value when fixing $K_{D,homo}$ using the unfolding/refolding kinetics value (Figure 6.15), reinforcing this method's validity.

6.2. Influence of DNA on the homodimer population

In Chapter 4, labelled CREB was used as a probe to measure affinity to target and non-target DNA. In both instances, the absolute anisotropy signal decreased, which suggests a possible decrease in complex size and, therefore, that, adding DNA could be dissociating CREB homodimers into bound monomers. This would certainly favour the monomer pathway in the nucleus where there are huge concentrations of non-target DNA.

To test this, FRET experiments were performed, either with labelled CREB and labelled DNA to measure DNA binding (Figure 6.16A-B), or with two different labelled CREB populations to measure dimerization (Figure 6.16C-D). In all cases, the donor fluorophore was Alexa Fluor™ 488 and acceptor fluorophore Alexa Fluor™ 594.

There was always an increase in FRET signal upon binding to DNA or upon mixing CREB populations, suggesting that FRET can be used to study both interactions (Figure 6.16). When using two CREB populations, there is a chance that two CREB from the same population interact, and I cannot correct for this – but, by chance, there will be some dimer formation with one subunit from each population, as shown by FRET signal and efficiency increase when mixing labelled populations (Figure 6.16). Interestingly, the addition of unlabelled DNA, whether target or non-target, increased the FRET signal of a CREB dimer.

To explore this further, a FRET intensity titration was performed by adding unlabelled nCRE DNA to a 1:1 mixture of CREB-Alexa488 and CREB-Alexa594. The addition of DNA results in a continuously increasing FRET signal intensity and efficiency with corresponding decrease in donor and acceptor standalone emissions, suggesting that the intensity increase comes from increased levels of dimerization (Figure 6.17). Hence, it appears that DNA actually increases the dimeric fraction of CREB. Interestingly, there was

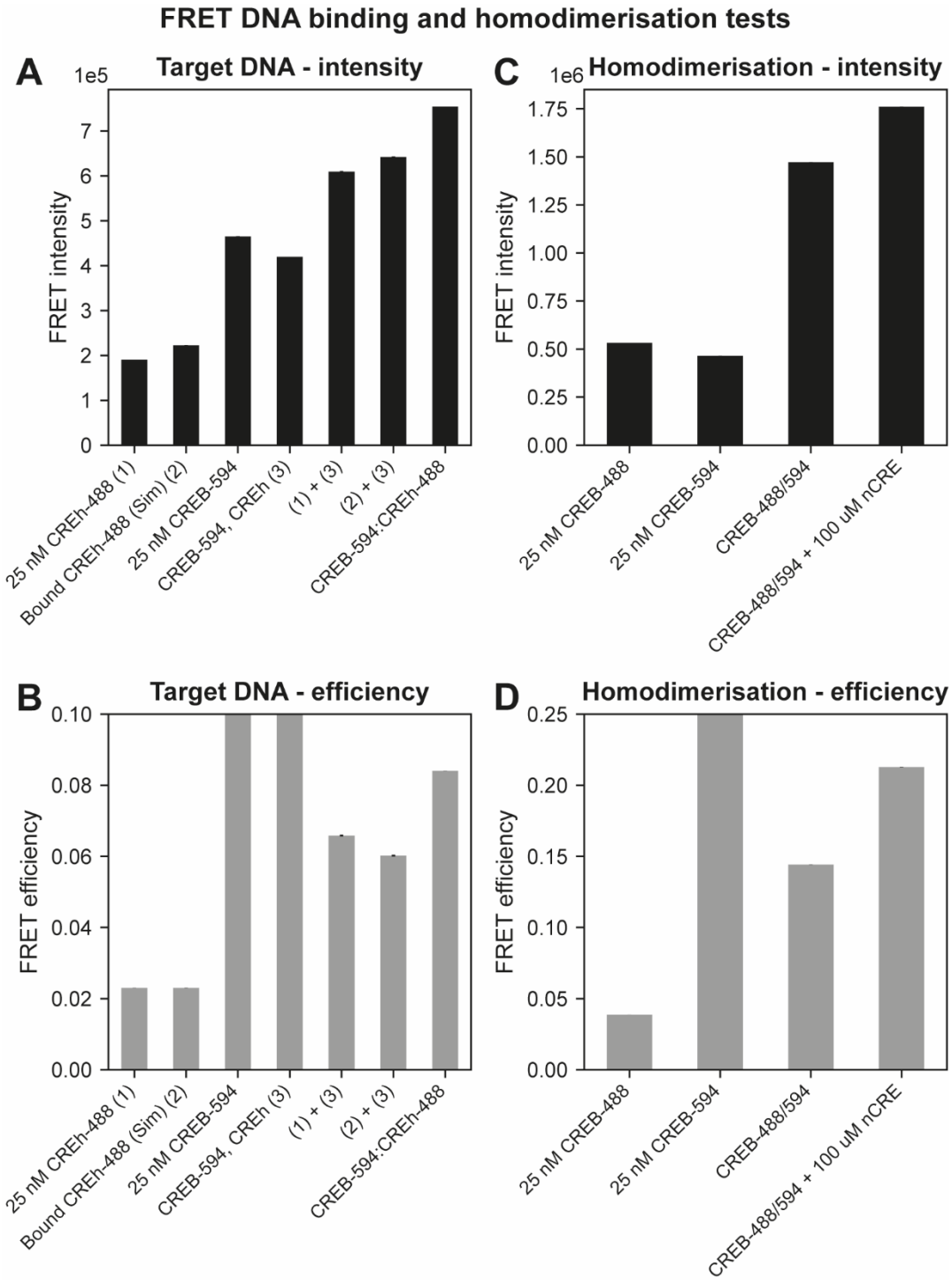


Figure 6.16: Individual FRET tests of CREB homodimerization and DNA binding. FRET pair is Alexa Fluor™ 488 and 594. Conditions: pH 7.4, 25 °C. **A.** FRET intensity and **B.** efficiency of CREB binding to DNA. (Sim) indicates simulated, by adjusting intensity signals according to expectations upon binding DNA. Both intensity and efficiency signals increase upon mixing labelled CREB and DNA more than expected. **C.** FRET intensity and **D.** efficiency of homodimerization tests. Mixing equal amounts of labelled CREB with donor 488 and acceptor 594 fluorophores leads to increased FRET intensity and efficiency signals. Efficiency is artificially high for CREB-A594 alone due to low fluorescence intensity emission of the Alexa Fluor™ 488 channel. Efficiency is calculated as per section 2.3.3 of the methods.

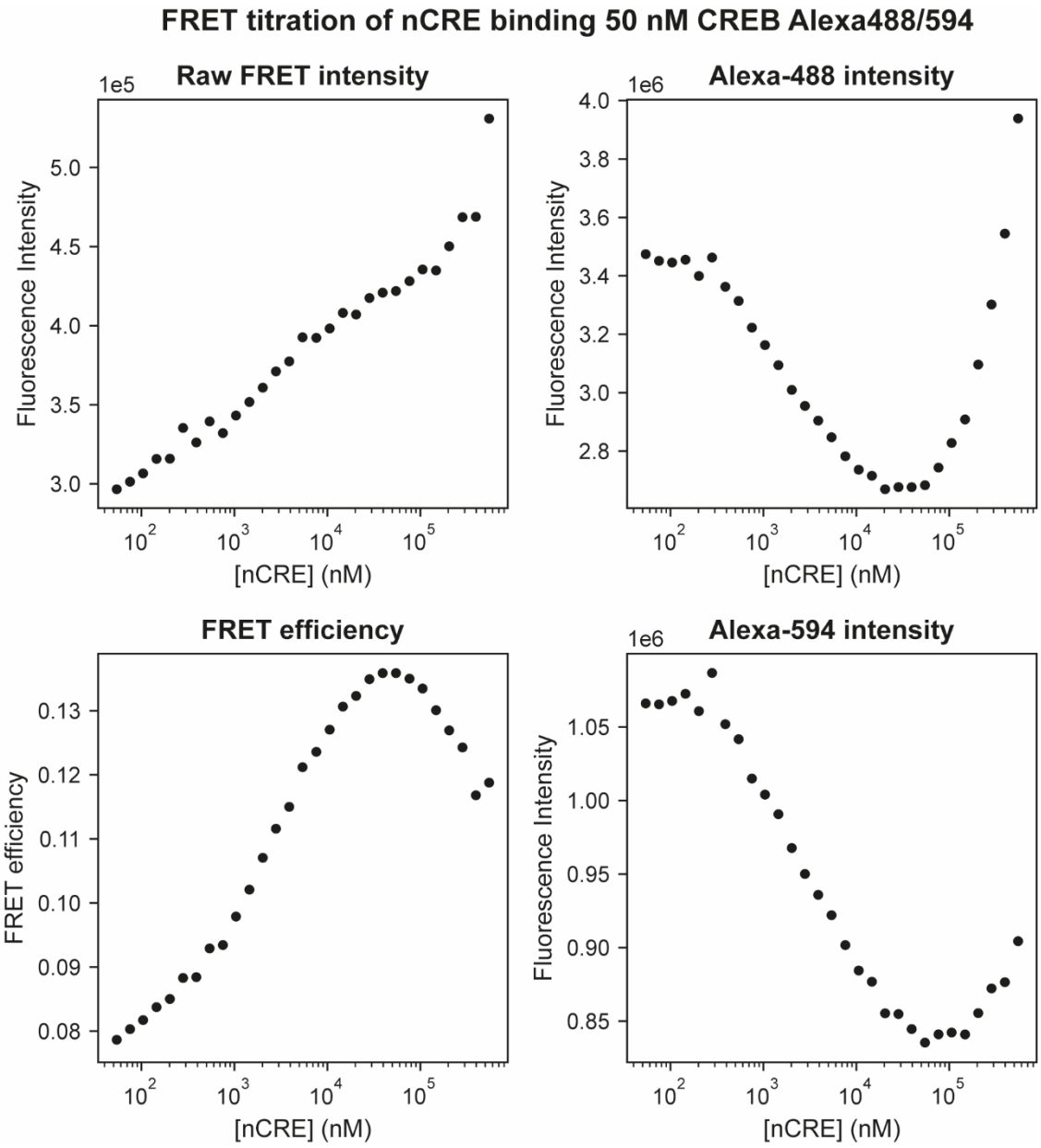


Figure 6.17: FRET titration of CREB-Alexa488/594 mixture bound by nCRE. All panels are from the same titration, using 50 nM CREB half labelled with Alexa Fluor™ 488 (donor) or 594 (acceptor) at 25 °C, pH 7.4. Panels are FRET intensity (top left), efficiency (bottom left) and fluorescence intensities from individual donor (top right) and acceptor (bottom right) channels. While individual channels intensities decrease, both FRET signals increase with DNA concentration. Around 100 µM nCRE, both intensities sharply increase instead, which suggest a signal unreliability threshold.

a sharp increase in fluorescence intensity beyond 100 μM DNA, which could show a viscosity threshold beyond which signal becomes unreliable.

6.3. Producing forced dimer and forced monomer CREB

Despite the above showing evidence for DNA-induced dimerization (which seems to favour the monomer pathway), there is still not enough evidence to say CREB follows it. In terms of a strict definition of the monomer pathway, where monomers meet once at the target, it is not clear if CREB dimerises on the target site or earlier on non-target DNA. More importantly because the measurements are equilibrium ones rather than kinetic ones they cannot inform about the order of the events. For example, the dimeric fraction could simply be increased by an increased dimer-half-life when bound to DNA. Hence, there is a need to completely isolate dimeric from monomeric contributions to target search, which is achieved with forced dimer and monomer mutants.

6.3.1. Rationale for forced dimer and monomer.

To facilitate purification, labelling, and for comparative purposes, purified CREB and mutants have no exposed cysteine residues (they are either mutated to serine or used in maleimide labelling). Hence, CREB dimers can be covalently linked by introducing GGC C-terminal extensions, forming disulfide bonds under oxidising conditions (Figure 6.18A)(131). Since this is the only cysteine available, there is no risk of forming intra-subunit bonds. This technique has been developed for bZIPs in the laboratory to great success ((163); Kuravsky, unpublished). Separately, the CREBΔLZ mutant, a complete deletion of the leucine zipper dimerization platform beyond residue V315 (or V301 for CREB isoform B), was used as a forced monomer.

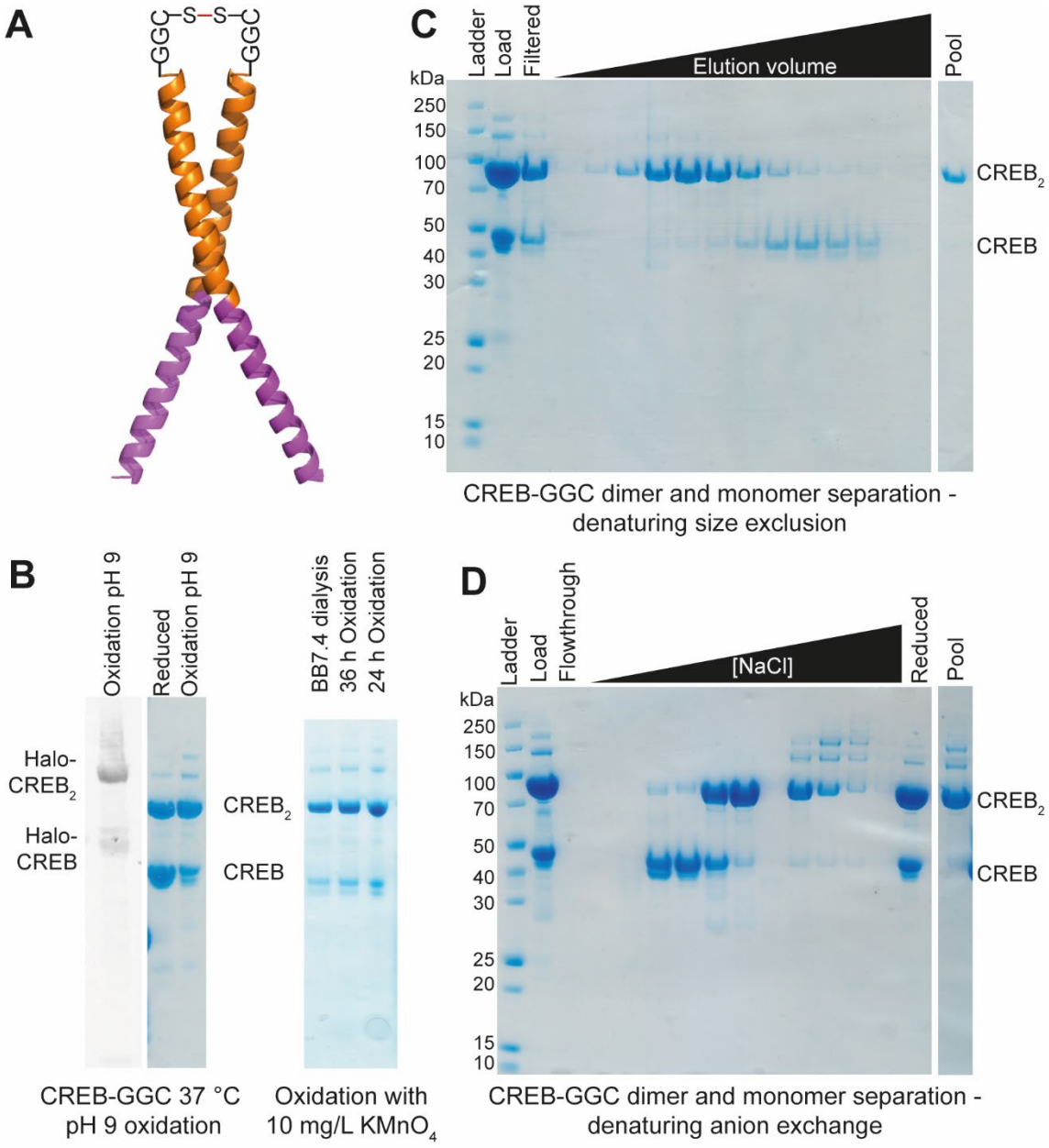


Figure 6.18: CREB-GGC forced dimer generation. A. Diagram of a CREB forced dimer, with a disulfide-bond (red) between a GGC C-terminal extension. **B.** There remains ~25% non-forced dimer post CREB-GGC oxidation at high pH and temperature, as shown in (192) with HaloCREB-6His (leftmost gel lane). Removal of the C-terminal tag appears to improve dimerization efficiency slightly. The presence of $KMnO_4$ oxidising agent helps (right gel) and the protein stays dimeric after removal of the agent by dialysis in biophysical buffer pH 7.4 (here BB7.4). **C.** Denaturing size-exclusion chromatography allows the separation of most dimeric from the monomeric species. **D.** Denaturing anion-exchange chromatography also separates both species, with a higher yield but with some minor monomeric contamination (< 5%).

6.3.2. CREB forced dimer purification.

The bZIP-GGC purification method was developed previously and simply consists of a standard bZIP purification, followed by concentration of the product and overnight 37 °C incubation in a 100 mM Tris pH 9 buffer to generate disulfide bonds. This was often sufficient to get a near complete dimeric species (163). However, when attempted with CREB (both in this work and in (192)), it was impossible, using the same oxidation method, to obtain almost complete forced dimerization, as about 25% of the protein remained monomeric (Figure 6.18B). Removing C-terminal tags slightly increases efficiency of the method (compared to (192), Figure 6.18B).

There are multiple chemicals reported to improve disulfide bond formation: glutathione (250), potassium ferricyanide (251), copper(II)(252) and DMSO (253). I performed various tests for CREB oxidation, incubating CREB with these chemicals with and without denaturants present (to prevent aggregation or folding cysteines away from each other). I pre-reduced with DTT to ensure availability of all cysteines. None of these chemicals improved oxidation significantly compared to simple high pH buffer and, if they did, often led to undesired multimerization (Figure 6.19). Instead, I opted to separate forced dimers from non-forced dimers by denaturing ion-exchange chromatography or denaturing size-exclusion chromatography, which is observed both on SDS-PAGE gels (Figure 6.18C-D). Unfortunately, the yield is very low, 25% compared to that achieved with pWT CREB. CREB-GGC is also more susceptible to precipitation (Figure 6.20), and some monomer can remain, although it usually makes up less than 5% of the final product (Figure 6.18D). Therefore, the purification method could be improved, potentially by crosslinking the cysteines or by purifying a single protein construct of two linked CREB. Nevertheless, the eluate is mostly pure forced dimer and suitable for biophysical characterisation.

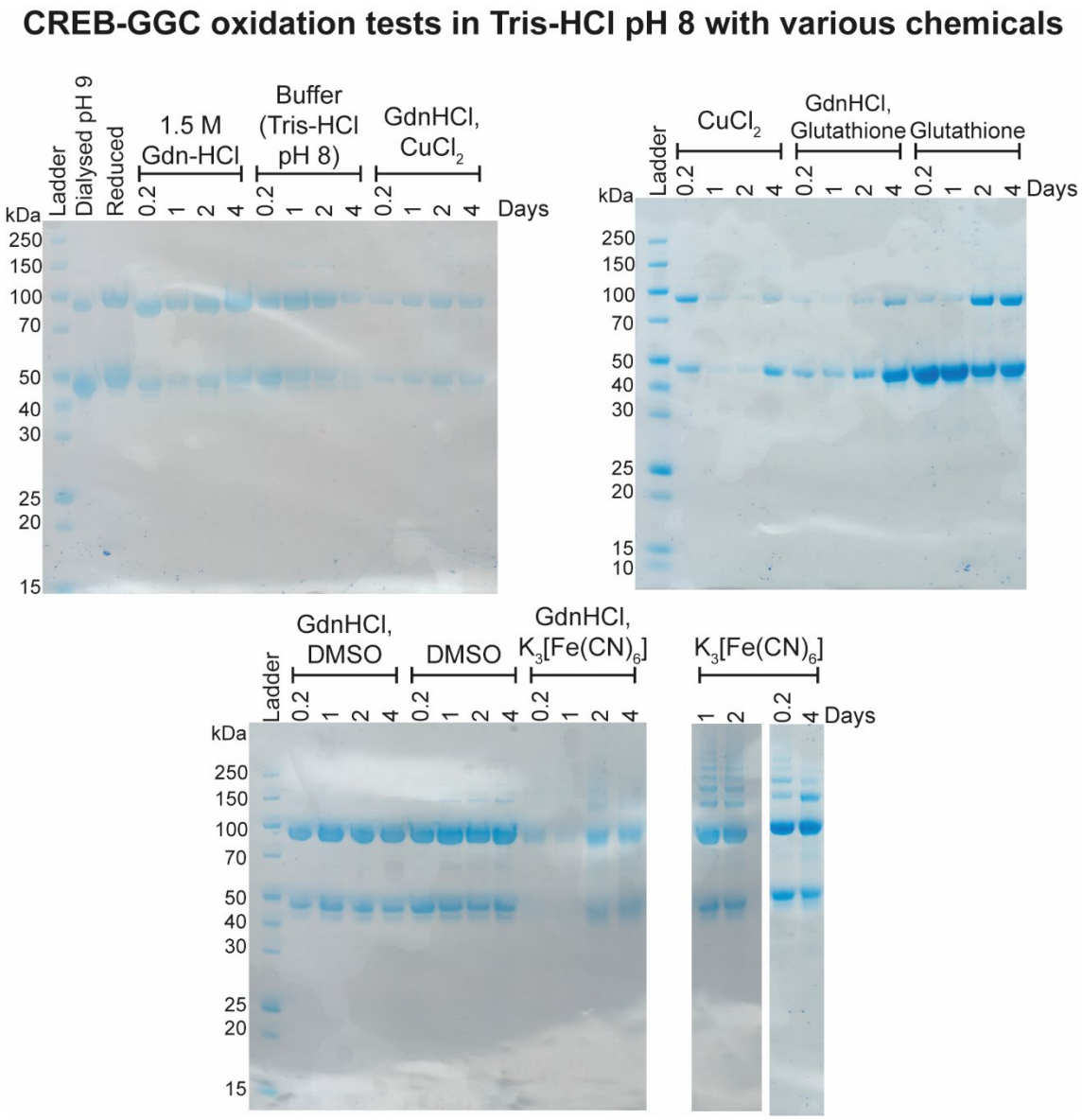


Figure 6.19: CREB-GGC oxidation trials in the presence of various chemicals. Incubation with various chemicals ($K_3[Fe(CN)_6]$, DMSO, GdnHCl, $CuCl_2$) known to catalyse disulfide bond formation do not improve forced dimer yields compared to simple incubation at high pH. Worse, $K_3[Fe(CN)_6]$ causes multimerization of the protein, when only dimeric species are desired.

6.4. Forced dimer biophysical properties

6.4.1. Equilibrium binding to target CREh and non-target nCRE DNA

As discussed in all previous chapters, there is a slight deviation in the 1 nM CREh-Alexa488 anisotropy titrations from the one binding-site model for all proteins. This is apparent when looking at residuals plots and is likely due to the presence of additional

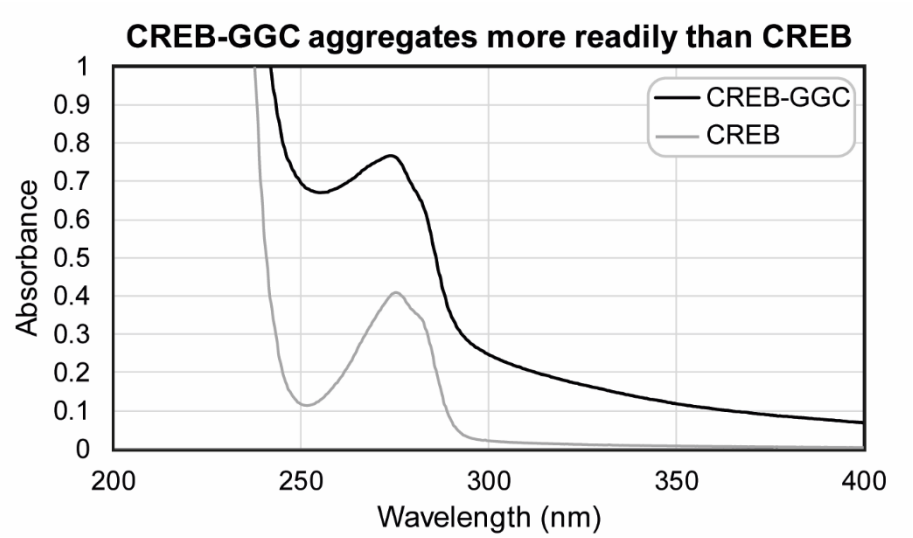


Figure 6.20: CREB-GGC oxidation trials in the presence of various chemicals. UV-vis buffer-subtracted absorbance spectrum of a CREB-GGC eluate showing an extreme case of aggregation, shown by Rayleigh scattering away from the absorbance baseline in the near-UV range, between 300-400 nm, which non-forced dimer CREB does not have.

species and processes. The transition regions in equilibrium curves are at such low CREB concentrations that CREB would be largely monomeric without the DNA present. Fitting to a more complex model to account for this behaviour needs either the homodimer or "true" CREh binding dissociation constant to be known (see Chapter 4).

Of course, forced dimers are already in the dimeric form. Thus, anisotropy titrations with 1 nM CREh-Alexa488 bound by CREB-GGC are well-fit by a single-site model, without deviation, and more likely to produce a "true" K_D (Figure 6.21A,C, Table 6.7). However, bZIP-GGC has too great an affinity for target DNA such that K_D cannot be obtained by anisotropy titration, as it displays typical saturation behaviour (Figure 6.21B).

Affinity to non-target was tested both with anisotropy titrations with 5 nM nCRE-Alexa488 (Figure 6.22). It was also estimated by ITC by addition of concentrated nCRE to CREB in the syringe and cell, respectively (Figure 6.23). Anisotropy titrations were well-fit by a two binding-sites model and ITC titrations by a single-site model. Titrations had some

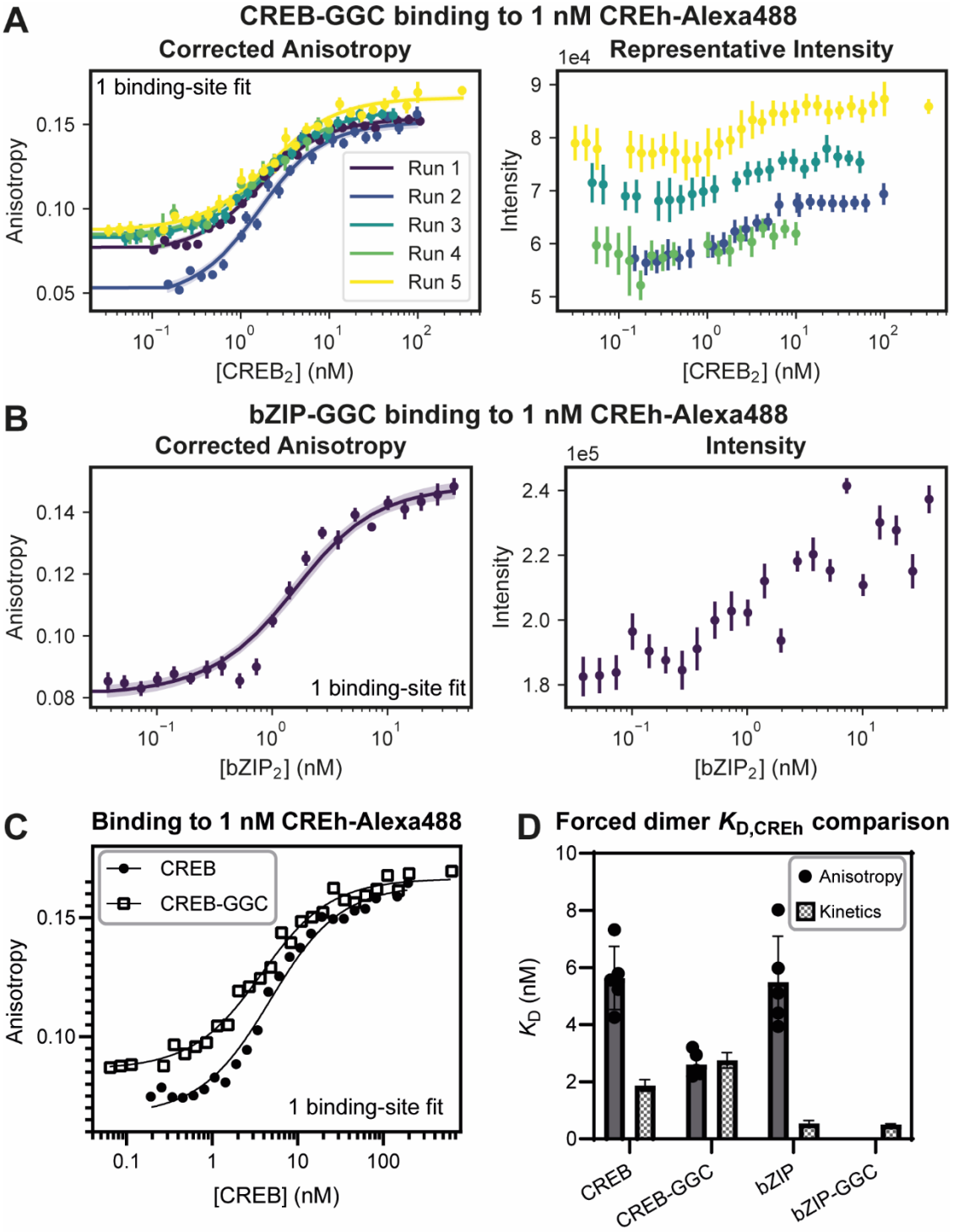


Figure 6.21: Forced dimers anisotropy binding titrations to 1 nM CREh-Alexa488 Conditions are pH 7.4, 25 °C. **A.** CREB-GGC binding to labelled CREh-Alexa488. Intensity (right) changes with binding, requiring anisotropy corrections. **B.** bZIP-GGC has greater than 1 nM affinity for CREh, as shown by the sudden "all or none" transition around 1 nM. K_D cannot be extracted using such titrations. **C.** CREB-GGC titrations transition in a similar range to CREB's but are better fit by a single binding-site model, likely because DNA-induced dimerization does not occur. **D.** Anisotropy and kinetics-derived K_D to CREh. Kinetics values are similar between pWT and forced dimers. CREB-GGC's anisotropy-derived K_D is close to the kinetics K_D as well, unlike CREB's, showing the influence of dimerization for DNA binding.

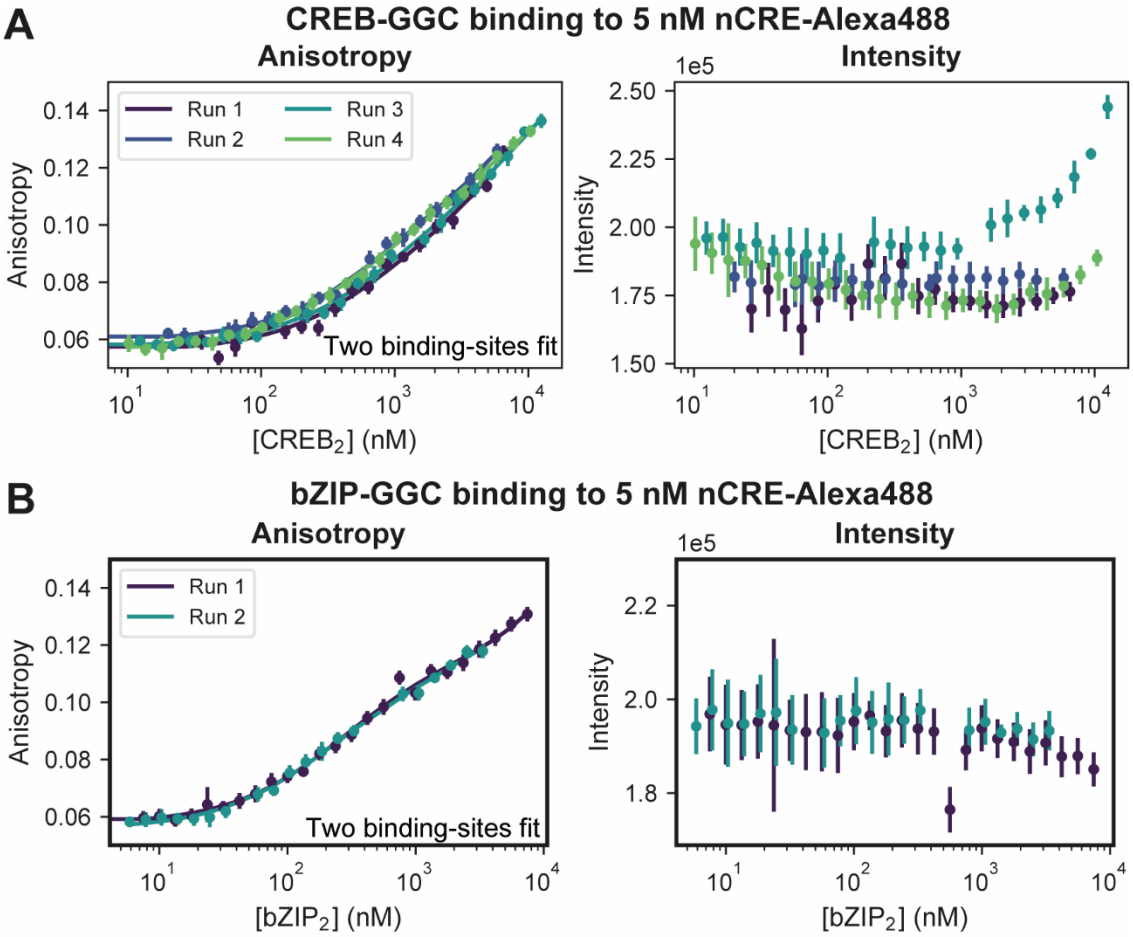


Figure 6.22: Forced dimers anisotropy binding titrations to 5 nM nCRE-Alexa488. Anisotropy binding titrations of **A.** CREB-GGC and **B.** bZIP-GGC to labelled nCRE, at pH 7.4 and 25 °C, overlap well with each other. Intensity tends to be constant except for a single CREB-GGC run where an upper plateau is not reached, and therefore anisotropy cannot be corrected. The apparent decrease in intensity for some titrations could be due to photobleaching.

variation, but generally gave a similar K_D in the vicinity of the non-forced dimer version (Figure 6.24D, Table 6.7). Of note, due to the low concentration of CREB-GGC used (both because of low yield and aggregation) it was not possible to estimate n-values reliably. Instead, a set n-value of 0.7 was used since it was close to double that of CREB titrations and similar to that of both high c-value CRElin into CREB ITC titration (Figure 6.23C), and nCRE into bZIP-GGC titrations (Figure 6.24C, Table 6.6). Knowing that non-forced dimer mutants have an n-value of 0.35, it is concordant that n-values were around 0.7 for forced dimers. Surprisingly, thermodynamic parameters for bZIP and bZIP-GGC are similar, but

CREB-GGC's are lower than CREB's and closer to the bZIP (Figure 6.24C, Table 6.6), but it could be a consequence of low c-value titrations.

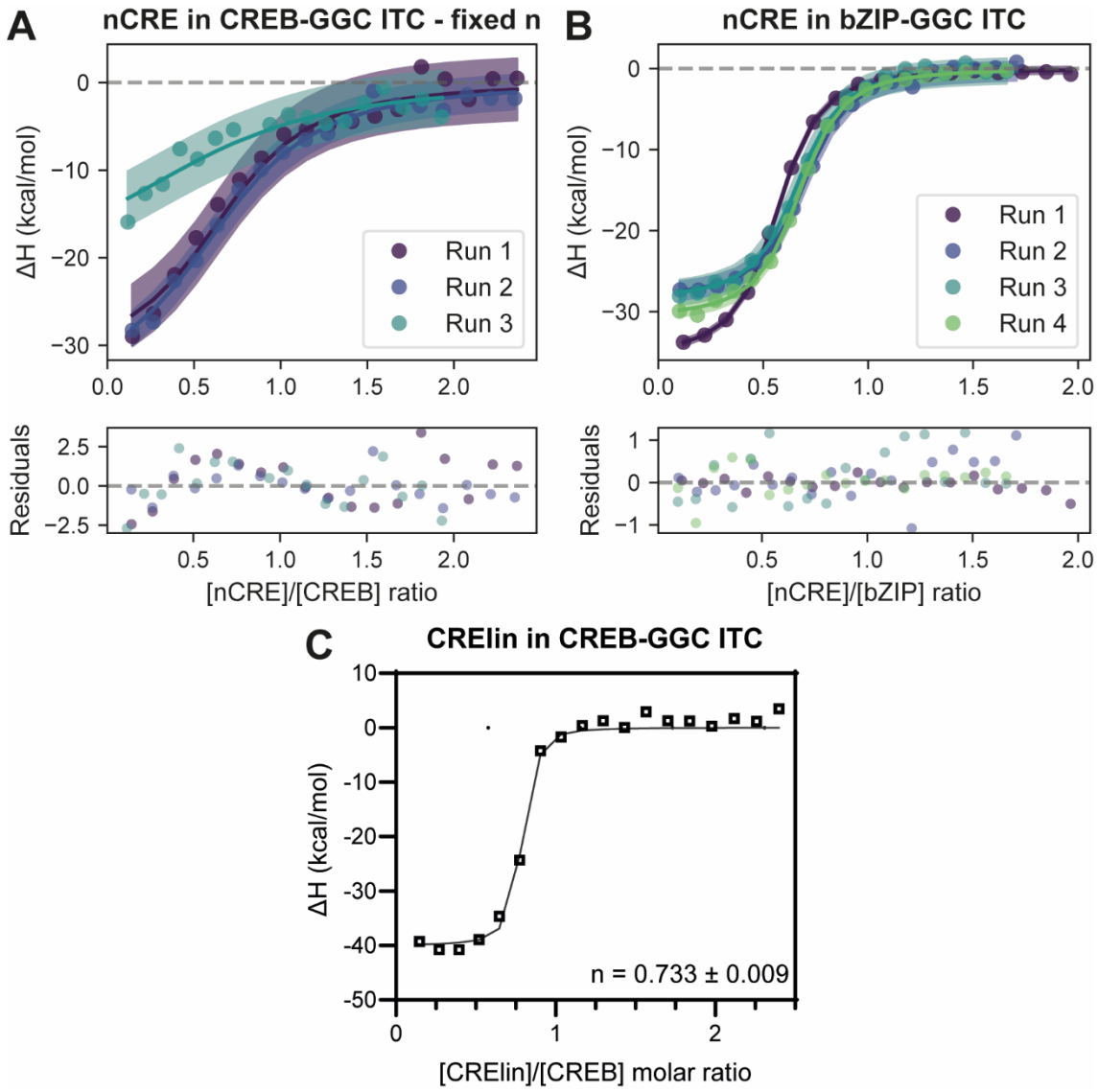


Figure 6.23: ITC titrations with addition of DNA to forced dimer CREB and bZIP. For all titrations, DNA was added to protein in the reaction cell at pH 7.4, 25 °C. **A.** CREB-GGC and **B.** bZIP-GGC titrations overlap very well with each other, although the difference in binding affinity is clear from the gradient differences, with bZIP-GGC's sharper gradient showing higher affinity. All CREB-GGC titrations (**A.**) have a fixed n-value at n=0.7 due to their low concentration in the ITC cell (an extreme case is run 3), which is required to extract $K_{D,nCRE}$ reliably. **C.** High-c titration of CRElin in CREB-GGC useful to extract reliable n-value, here 0.733, showing the n=0.7 estimate for CREB-GGC bound by nCRE is reasonable.

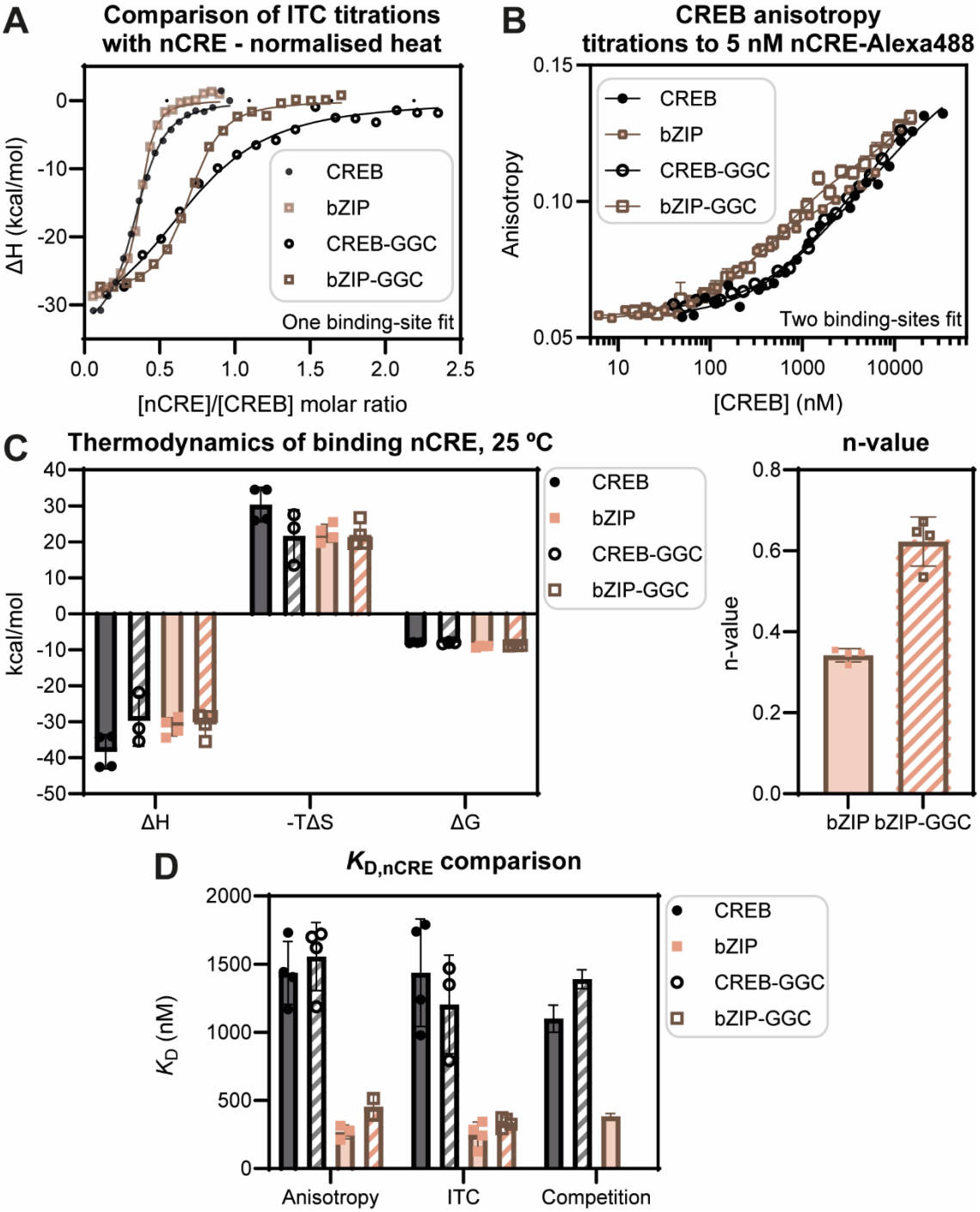


Figure 6.24: Forced dimer binding to nCRE parameters comparison with non-forced dimers. Conditions: 25 °C, pH 7.4. **A.** ITC titrations with nCRE to protein setup show the n-value difference between forced-dimers and pWT. **B.** Anisotropy titrations binding to labelled nCRE (with converted monomeric concentrations) show good overlap between forced dimers and pWT, although bZIP-GGC does not reach the plateau that bZIP does. **C.** Thermodynamics of nCRE binding by ITC. bZIP and bZIP-GGC are similar but CREB-GGC's enthalpy and entropy are reduced, although this could be due to lower concentrations used. N-value for forced dimers is roughly twice that of pWT. **D.** Forced dimers have comparable affinity to nCRE, although both bZIP-GGC's and CREB-GGC's antitropy and competition (through association to CREh with nCRE competitors) $K_{D,nCRE}$ are slightly higher.

ITC parameter at 25 °C	CREB-GGC	bZIP-GGC
ΔH (kcal/mol)	-30 ± 7	-31 ± 3
-TΔS (kcal/mol)	22 ± 7	22 ± 3
ΔG (kcal/mol)	-8.1 ±0.2	-8.84 ±0.07
n-value	Fixed at 0.7	0.62 ± 0.06

Table 6.6: Thermodynamics and n-value of binding to nCRE in ITC experiments using forced dimers. All values are averages of triplicates or more, with standard deviation errors, at pH 7.4, 25 °C.

6.4.2. Association rate constant with CRE

CREB-GGC and bZIP-GGC k_{on} was determined, as previously, by rapidly mixing various concentrations of protein with 5 nM CREh-Alexa488 in a stopped-flow apparatus. The difference is only on the concentrations – since forced dimers are covalently linked, their concentration is two-fold lower than the monomeric equivalent. Kinetics values must account for this before comparisons are made.

Traces were well-fit by single exponential decay functions (Figure 6.25A). The observed rate constants align well linearly with protein concentration and, if analysed with 1 ms cutoff, are similar in value to CREB's, at k_{on} = (1.08 ± 0.09) $nM^{-1}s^{-1}$, (or (0.54 ± 0.05) $nM^{-1}s^{-1}$ if converted to monomeric concentration) (Figure 6.25B-D, Table 6.8). Importantly, this value closely matches the one obtained for CREB at the highest concentration range (when y-intercepts are also positive). This value is also more robust to cutoff changes (Figure 6.26), which was not the case for CREB (see Chapter 4). bZIP-GGC's k_{on} is also very similar to bZIP's, albeit y-intercepts from bZIP-GGC's k_{on} gradients are closer to the origin or positive (Figure 6.25B-C, Table 6.8). Hence, they are more representative of the "true" k_{on} value for dimeric CREB. This analysis demonstrates that the variability in CREB's (or bZIP's) k_{on} and y-intercept values (see Chapters 4, 5), depending upon chosen concentration range, are due to changes in the dimer fraction. If additional processes (dimerization) are taking place then a simple single exponential decay function may not

completely accurately describe the kinetic traces for CREB, which could further explain the cutoff-dependence of extracted rate constants. It also likely indicates that 1 ms cutoff is the correct choice for association analysis, since the 1 ms CREB k_{on} is like CREB-GGC's.

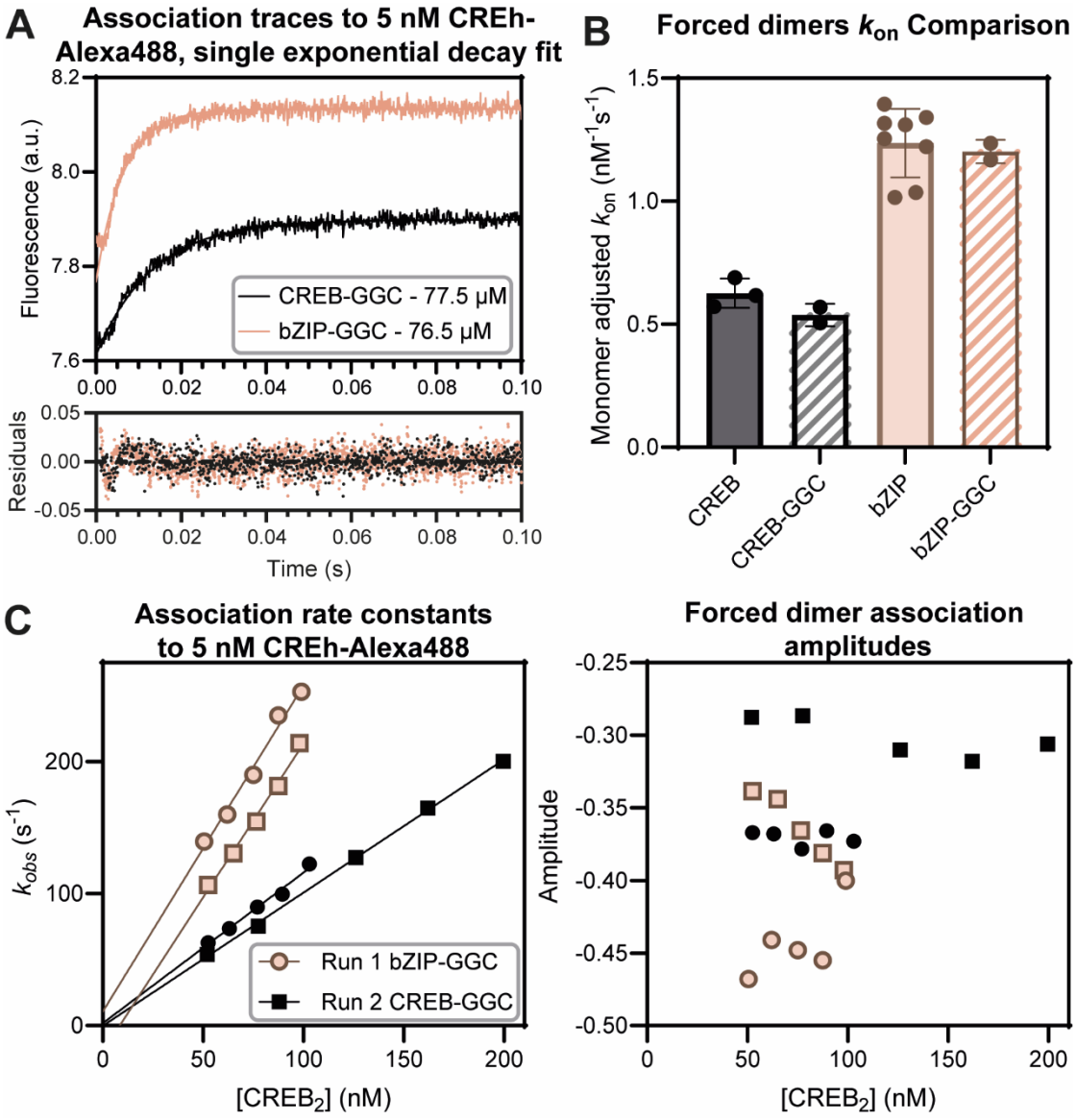


Figure 6.25: CREB-GGC and bZIP-GGC association kinetics. Conditions are pH 7.4, 25 °C. **A.** Comparison of CREB-GGC and bZIP-GGC association traces to labelled CREh. Both are well-fit by single exponential decay functions, as shown by the residuals, and bZIP reaches equilibrium saturation before CREB. **B.** Comparison of extracted k_{on}. Both forced dimer constructs are almost identical to their pWT counterparts. **C.** All runs with associated rate constants and amplitudes for forced dimers. Gradients are reproducible and, importantly, go close or through the origin, as expected for a two-state model (left). Amplitudes are largely independent of protein concentration (though not for bZIP-GGC, run 2).

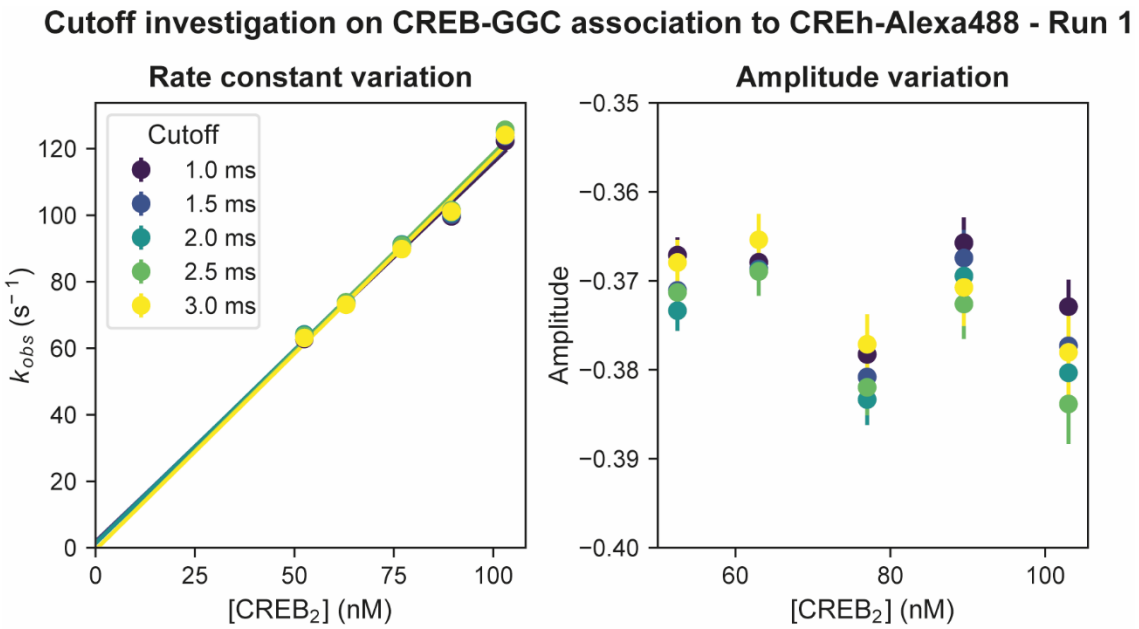


Figure 6.26: CREB-GGC investigation of time cutoff influence on association. CREB-GGC association rate constants (left) and amplitudes (right) to 5 nM Alexa Fluor™ 488-labelled CREh at pH 7.4, 25 °C are almost identical, regardless of chosen time cutoff.

6.4.3. Dissociation rate constant with CRE

Dissociation kinetics (again, tested by mixing a preformed 2:1 dimer:CREh-Alexa488 complex into large amounts of CRElin) are well-fit by a double exponential decay function for both CREB-GGC and bZIP-GGC, as previously seen with non-forced dimer CREB (Figure 6.27). This is consistent with the second phase not being dimerization-related but, as discussed in Chapter 4, an artefact from DNA into DNA dilution. Once again, these values are similar to pWT CREB, although slightly higher for CREB-GGC (Figure 6.28C, Table 6.8). The supposed artefact rate constant is usually very low in absolute value, and this was the case too for CREB-GGC, whose k_{off} values are similar to CREB's. However, there was more variability when using the bZIP, and the artefact rate constant was either very low or very high compared to the reproducible rate constant (see Chapter 5). bZIP-GGC showed similar behaviour, as $k_{obs,2}$ has the higher amplitude and is closer to bZIP's $k_{off,1}$ but is smaller than $k_{obs,1}$ which is much more variable within a run (Figure 6.28B-C). Interestingly, using large amounts of competitor DNA appeared to affect amplitudes and

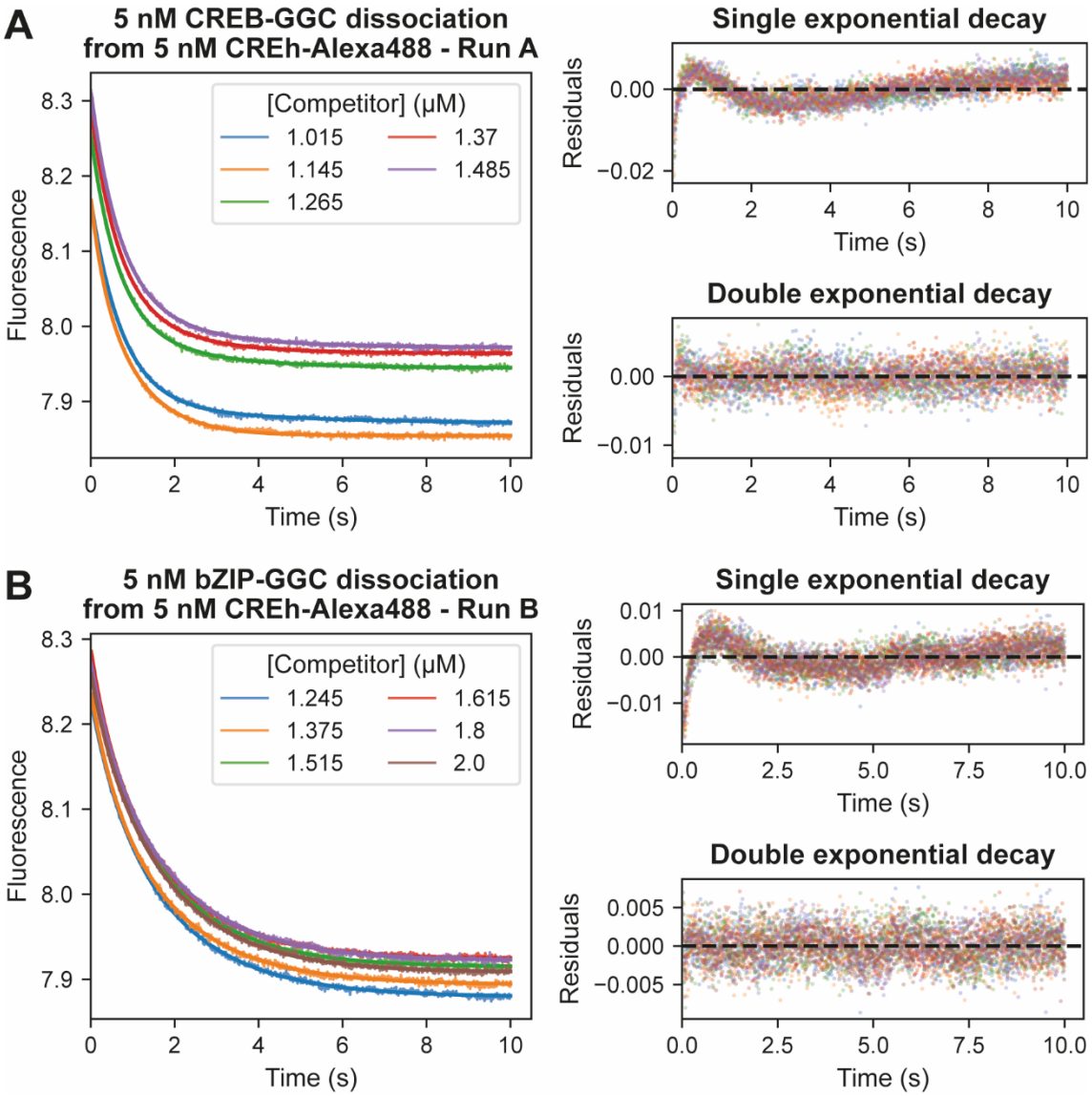


Figure 6.27: Forced dimer dissociation traces of CREB-GGC and bZIP-GGC from 5 nM CREh-Alexa488. Both CREB-GGC (**A.**) and bZIP-GGC (**B.**) traces at pH 7.4, 25 °C are better fit by double rather than single exponential decay equations. All traces also appear to have similar behaviour (rate constant and amplitude).

rate constants (6.28B), showing that the artefacts may affect reads in a concentration-dependent manner. The rudimentary "control correction" developed in Chapter 4 goes some way towards retrieving single exponential decay profiles (Figures 6.29, 6.30), such that the double exponential decay rate constant with the greatest amplitude is virtually identical to the single exponential decay one. The correction is not perfect which likely represents the somewhat unpredictable nature of the second phase.

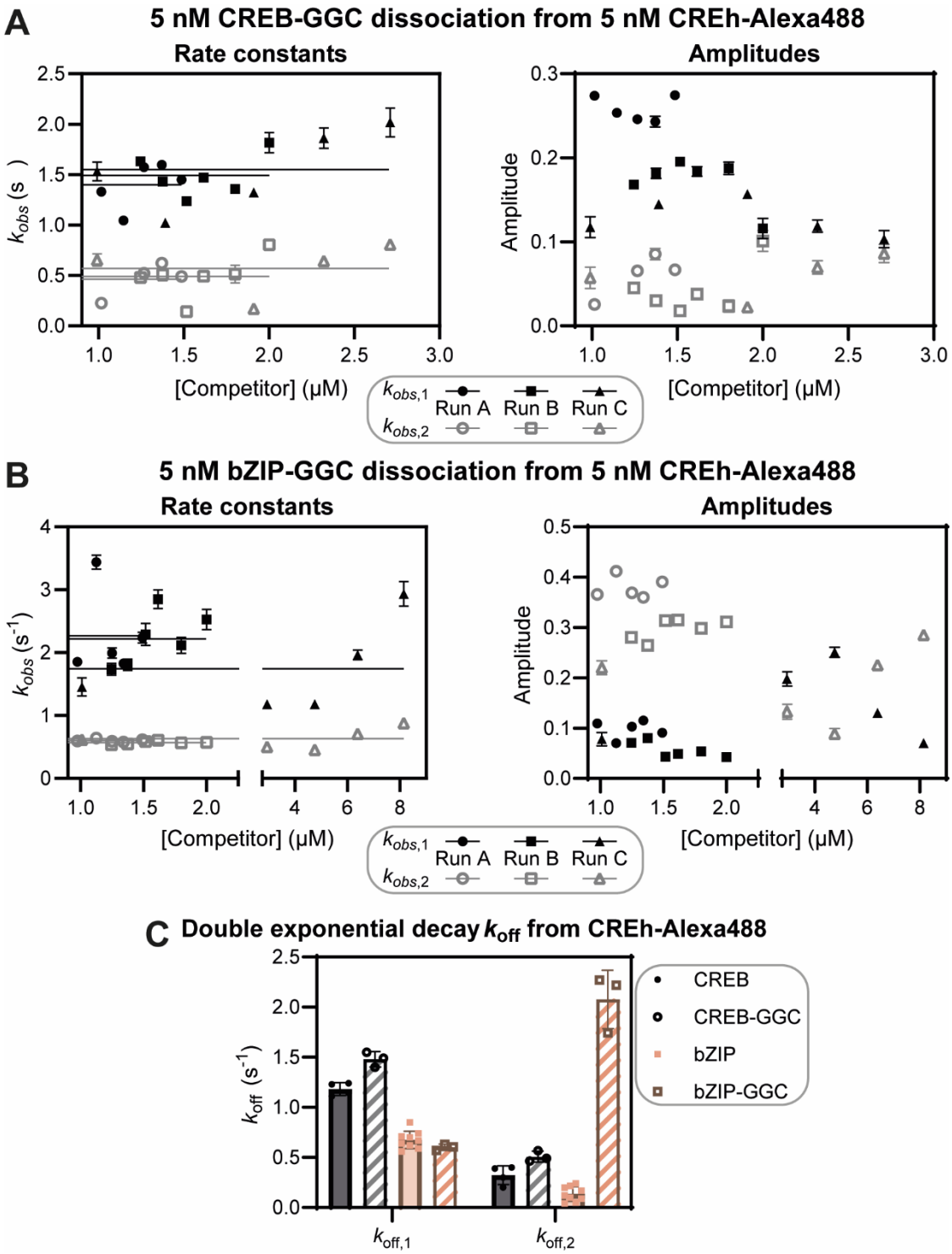


Figure 6.28: Forced dimer dissociation kinetics analysis of CREB-GGC and bZIP-GGC. A. CREB-GGC and **B.** bZIP-GGC dissociation rate constants and amplitudes from labelled CREh from double exponential decay fits at pH 7.4, 25 °C. Despite some single value heterogeneity, CREB-GGC run averages of all runs are consistent. $k_{obs,1}$ also has the highest amplitude. $k_{obs,2}$ from different bZIP-GGC runs overlap well and, unlike CREB-GGC, have the highest amplitude. The highest concentration run interestingly shows a swap in amplitude and apparent rise in k_{obs}. **C.** Comparison of rate constants, with $k_{off,\,1}$ extracted from k_{obs} with the highest amplitude. Both forced dimers are quite similar to their pWT counterparts, although CREB-GGC appears to stay bound less long, and $k_{off,\,2}$ is higher for both forced dimers.

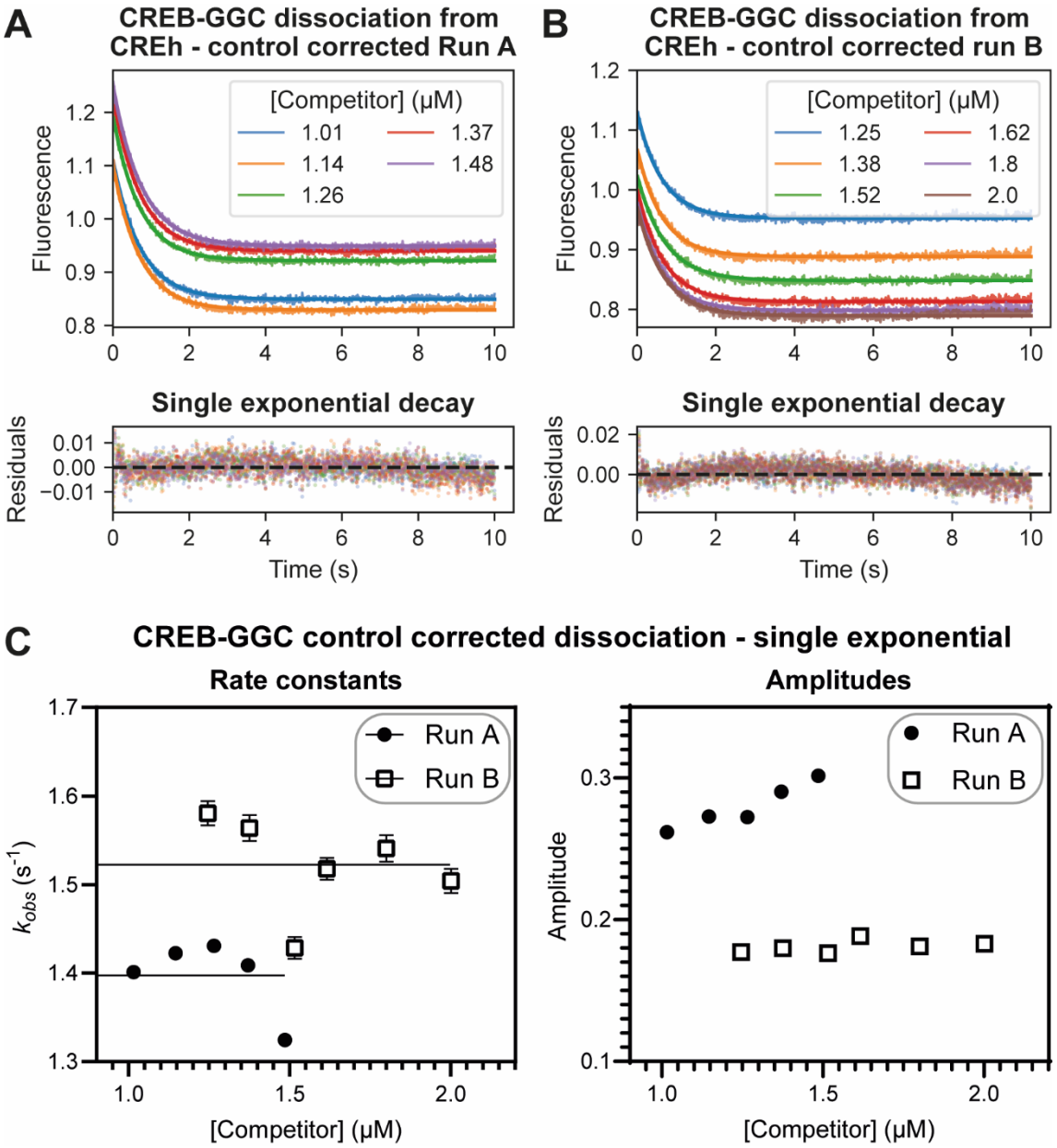


Figure 6.29: Rudimentary correction of DNA mixing artefacts for CREB-GGC dissociation. A. Example where the control correction works well, with residuals showing single exponential decay functions account for most of the behaviour and residuals are smaller than in the non-control corrected trace (Figure 6.27). **B.** Example where control correction captures some but not all the second phase, as highlighted by the residuals' greater deviation from the fit. **C.** Rate constants and amplitudes from control-corrected traces. k_{obs} are similar to CREB-GGC's $k_{off,1}$ showing $k_{off,2}$ is likely the result of artefacts. Amplitudes are relatively constant but slowly increase, showing the need for more robust correction.

Assuming the rate constant with the highest amplitude ($k_{off,1}$ in Figure 6.28C), it can be used to calculate the K_D of binding CREh via kinetics values (Figure 6.21D). It is very similar for bZIP and bZIP-GGC and, as mentioned in section 6.4.1, the anisotropy K_D and kinetics K_D are concordant for CREB-GGC as dimerization is no longer relevant for binding.

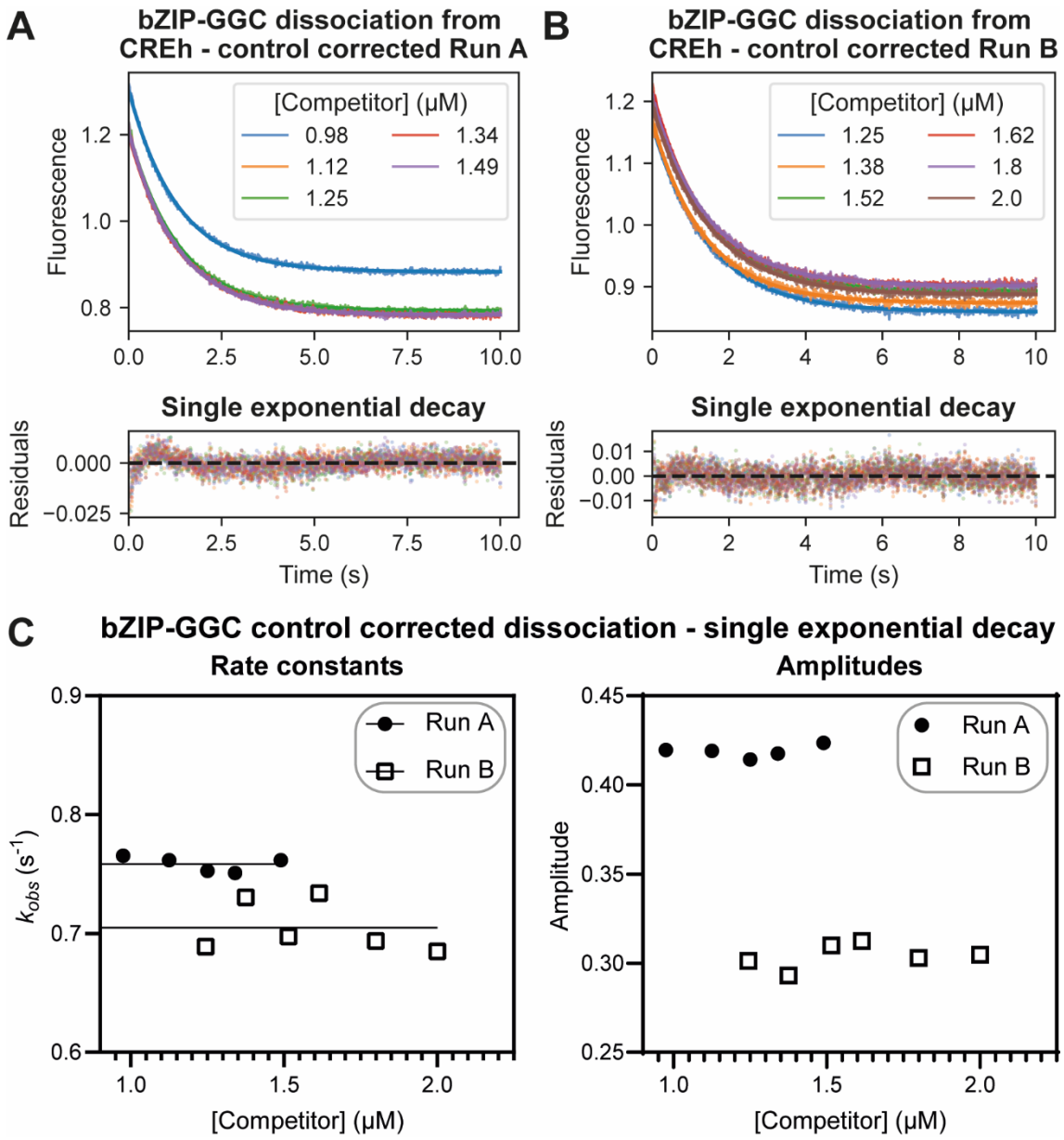


Figure 6.30: Rudimentary correction of DNA mixing artefacts for bZIP-GGC dissociation. A. and **B.** Examples where the control correction transforms behaviour to single exponential decay (compared to Figure 6.27), although both show some variation at the beginning of the trace. **C.** Rate constants and amplitudes from control-corrected traces. k_{obs} are like bZIP-GGC's lower $k_{off,\,1}$ showing the higher $k_{off,\,2}$ is likely the result of artefacts. Amplitudes are relatively constant, showing the efficacy of the correction.

CREB's kinetics K_D value is slightly lower than CREB-GGC, which suggests there may be an advantage to being monomeric, favouring binding to CREh.

6.4.4. Effect of non-target DNA on CREB forced dimer's association kinetics.

Similarly to CREB and bZIP, the addition of non-target oligos slows down association of CREB-GGC to CREh (Figure 6.31C). This is also fit by a simple-model

dependent on the affinity of CREB-GGC to DNA (Figure 6.32C). If allowed to be freely fit, the model extracts a comparable $K_{D,nCRE}$ to ITC and anisotropy and a similar k_{on} to stopped flow and the master curve (Figure 6.32D, Table 6.7, 6.8) obtained by calculating [CREB-GGC]$_{free}$ using $K_{D,nCRE}$ (Figure 6.32B). The link between the affected k_{on} and [CREB-GGC]$_{free}$ is seen by the linear dependence between the $k_{on,comp}/k_{on,app}$ and [CREB-GGC]$_{free}$/[CREB-GGC]$_{total}$, even more so than for CREB (Figure 6.32A).

High concentrations of nCRE appear to change CREB-GGC association traces from single to double exponential decay behaviour (as seen with CREB) (Figure 6.31A). Control traces of labelled CREh and unlabelled nCRE mixed with buffer alone do not have a signal change (see Chapter 4), suggesting the second phase is linked to the DNA binding process, however the effect is still very small as seen per the residuals. Two rate constants were reliably extracted at the highest nCRE concentration (double exponential decay fits for 5 μM as shown for CREB in Chapter 4, were unreliable here) (Figure 6.31B), but neither fits the model as well as the single exponential decay rate, suggesting it is still applicable here (Figure 6.32C). Interestingly, the amplitudes appear to converge with increasing protein concentrations, perhaps indicating the rising prevalence of a second process, such as intersegmental transfer (Figure 6.31B). Higher competitor amounts are needed to verify.

K_D parameter and associated method	CREB-GGC	bZIP-GGC
K_D (anisotropy) (nM)	2.6 ± 0.4	-
K_D (kinetics) (nM)	2.8 ± 0.3	0.50 ± 0.03
$K_{D,nCRE}$ (anisotropy) (nM)	1560 ± 250	450 ± 90†
$K_{D,nCRE}$ (ITC) (nM)	1200 ± 360	340 ± 40
$K_{D,nCRE}$ (segregation model) (nM)	1390 ± 70*	-

Table 6.7: Monomerically-adjusted K_D to CREh and nCRE of forced dimer mutants. All values (at pH 7.4, 25 °C) are averages of triplicates or more, with standard deviation errors, except for * values which are from a single experiment with fitting error, or † for two experiments.

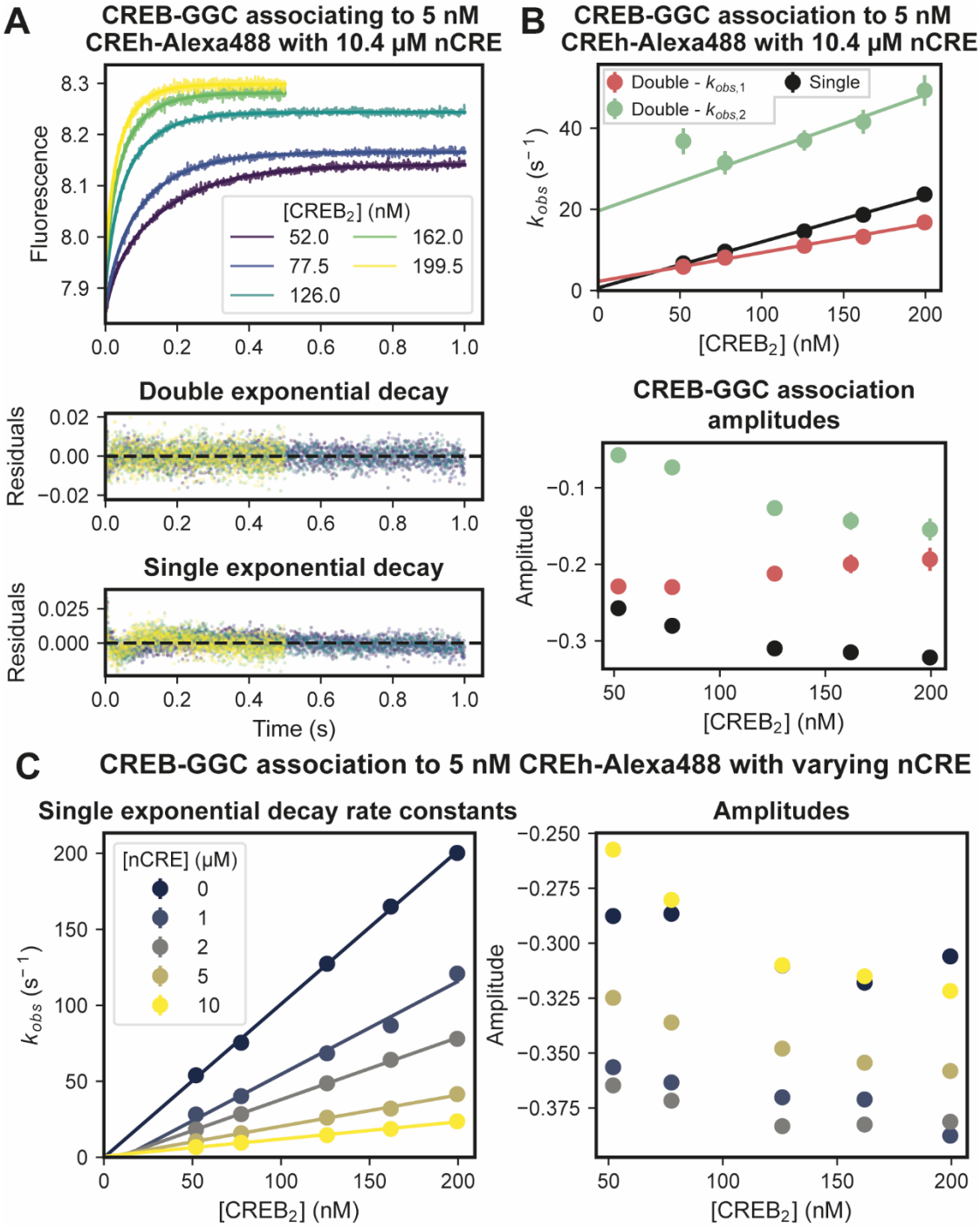


Figure 6.31: CREB-GGC competitor inhibited association to 5 nM CREh-Alexa488. Conditions: pH 7.4, 25 °C. **A.** Association traces to labelled CREh show some double-exponential decay behaviour at the highest nCRE competitor concentration, although single exponential decay still is a good fit. **B.** Extracted single and double exponential decay observed association rate constants. Interestingly, amplitudes seem to converge for double exponential decay, perhaps indicative of a second process (e.g. intersegmental transfer) becoming more important with increased protein and DNA concentrations. **C.** All single exponential decay observed rate constants, which decrease with increasing competitor concentration. Amplitudes appear to decrease but this effect is small, as seen from the y-axis scale. This decrease does become more prevalent with increasing competitor concentration, perhaps due to a second process.

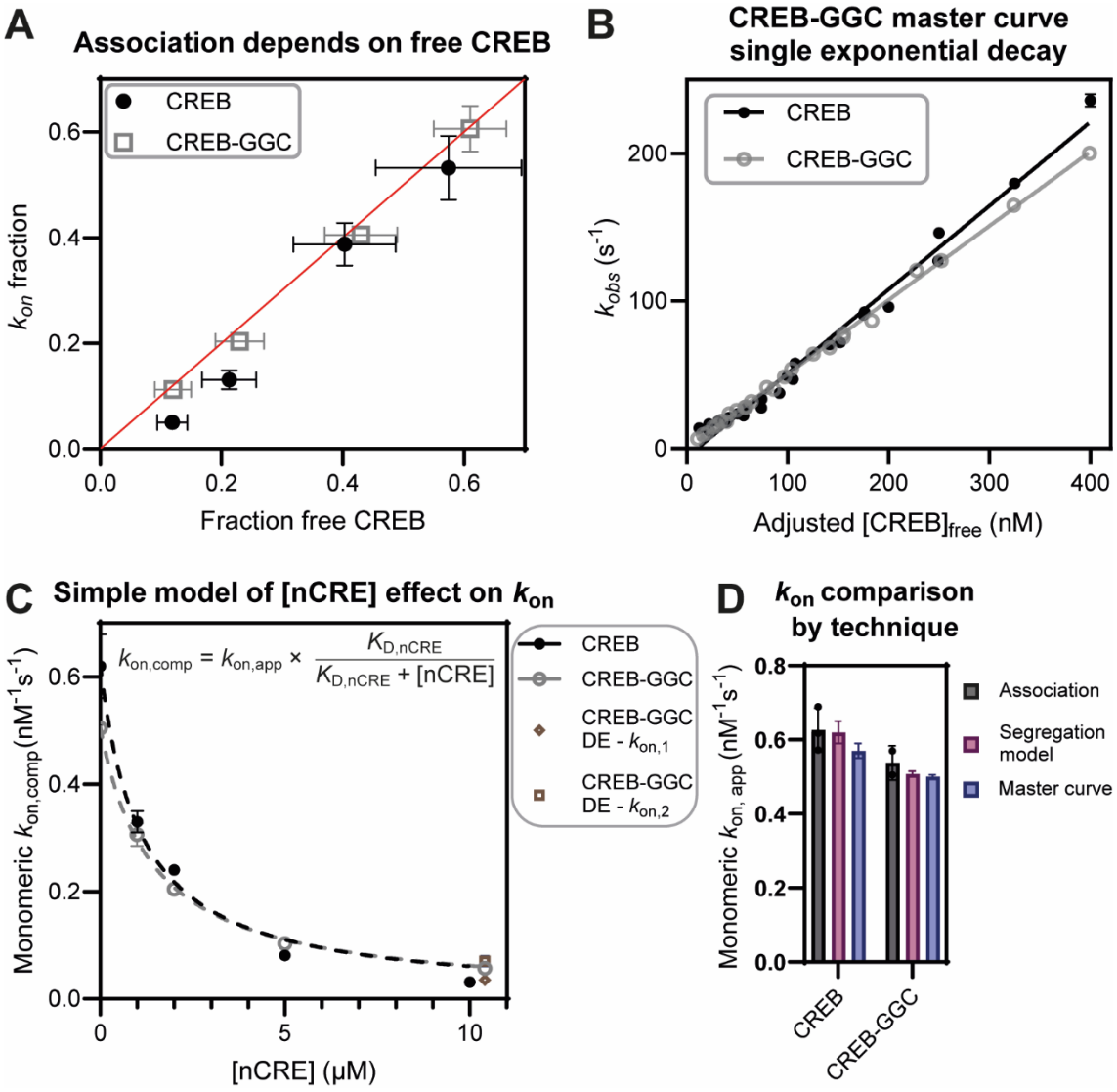


Figure 6.32: CREB-GGC target-search inhibition follows the simple sequestration model. Conditions: pH 7.4, 25 °C. **A.** CREB-GGC's fractional k_{on} (compared with no competitor) shows an almost 1:1 dependence with its expected free fraction, as per CREB **B.** CREB-GGC master curve association to CREh, using $[CREB]_{free}$ rather than total CREB concentration, calculated using $K_{D,nCRE}$ All rate constants align very well linearly, with a similar k_{on} to Figure 6.25, with a slightly decreased gradient compared to CREB. Of note, low $[CREB]_{free}$ concentrations do not deviate from linearity. **C.** Extracted $k_{on,app}$ are well fit by the simple exclusion model shown, dependent on $K_{D,nCRE}$. Double exponential decay rate constants do not conform to the model as well. **D.** k_{on} obtained by freely fitting the model and from the master curve conform to those obtained by association kinetics without competitors.

Kinetic parameter	CREB-GGC	bZIP-GGC
k_{on} (kinetics) ($nM^{-1}s^{-1}$)	0.54 ± 0.05†	1.20 ± 0.05†
Segregation model k_{on} ($nM^{-1}s^{-1}$)	0.507 ± 0.008*	-
Master curve k_{on} ($nM^{-1}s^{-1}$)	0.503 ± 0.005*	-
$k_{off,1}$ (s^{-1})	1.48 ± 0.08	0.60 ± 0.03
$k_{off,2}$ (s^{-1})	0.51 ± 0.05	2.1 ± 0.3

Table 6.8: Monomerically-adjusted kinetics values of forced dimer mutants binding to CREh. Values at pH 7.4, 25 °C to labelled CREh are averages of triplicates or more, with standard deviation errors, except for * values which are from a single experiment with fitting error, or † for two experiments.

6.5. Characterising monomer-DNA interactions

Deletion of the leucine zipper (CREBΔLZ mutation) ensures mutants are monomeric, as confirmed by SEC-MALS (Figure 6.2C), since they are missing their dimerization platform (163). Binding to both target and non-target was measured by fluorescence anisotropy using DNA labelled with Alexa Fluor™ 488. However, an upper plateau was never reached in anisotropy, and changes were only noticed at extremely high titrant concentrations (Figure 6.33A). Saturating concentrations could cause this change by increasing viscosity and, thereby, fluorescence rather than binding. Similar observations were made for the FRET titration (Figure 6.17) where, at high nCRE concentration, fluorescence intensities, previously decreasing, increase sharply past a certain threshold in the high μM range. Even should this reflect a binding event, it is impossible to extract K_D reliably from such titrations, with at best an estimate around 1 mM, and there is no preference for CREh over nCRE.

To try and account for this a reverse anisotropy titration was performed i.e. the protein was labelled (here CREBΔLZ-Alexa594) and titrated with unlabelled CRElin (Figure 6.33B). It was reasoned the smaller size of DNA would delay the saturation effect

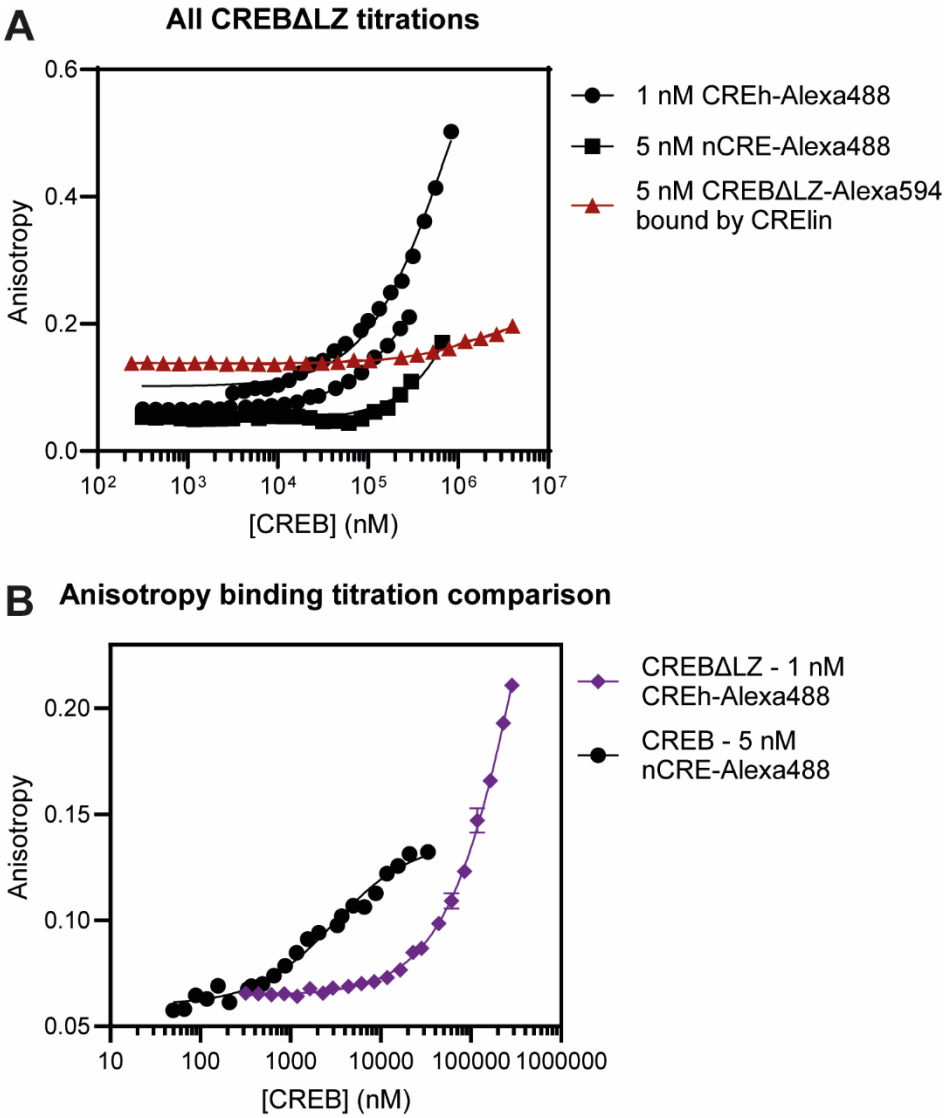


Figure 6.33: CREBΔLZ forced monomer anisotropy binding titrations. Conditions: pH 7.4, 25 °C. **A.** Whether using labeleld CREBΔLZ or DNA as a probe, an upper plateau is never reached, and K_D never reliably obtained. Anisotropy increases massively at high titrant concentrations, perhaps due to viscosity. Best possible affinity estimates are in the 500 µM - 1 mM range. Moreover, there does not appear to be great distinction between CREh and nCRE affinity for CREBΔLZ. **B.** Comparison of CREBΔLZ and CREB binding titrations show the much greater range at which CREBΔLZ curves start to transition compared to even nCRE binding titrations for CRE.

at higher concentrations. The signal increase did occur again albeit at higher titrant concentrations than previously. In support of the viscosity hypothesis, the anisotropy increases whereas, in previous reverse titrations, it decreased (Chapter 4). The results suggest an extremely low affinity of CREBΔLZ for CRE DNA, in the mM range or above.

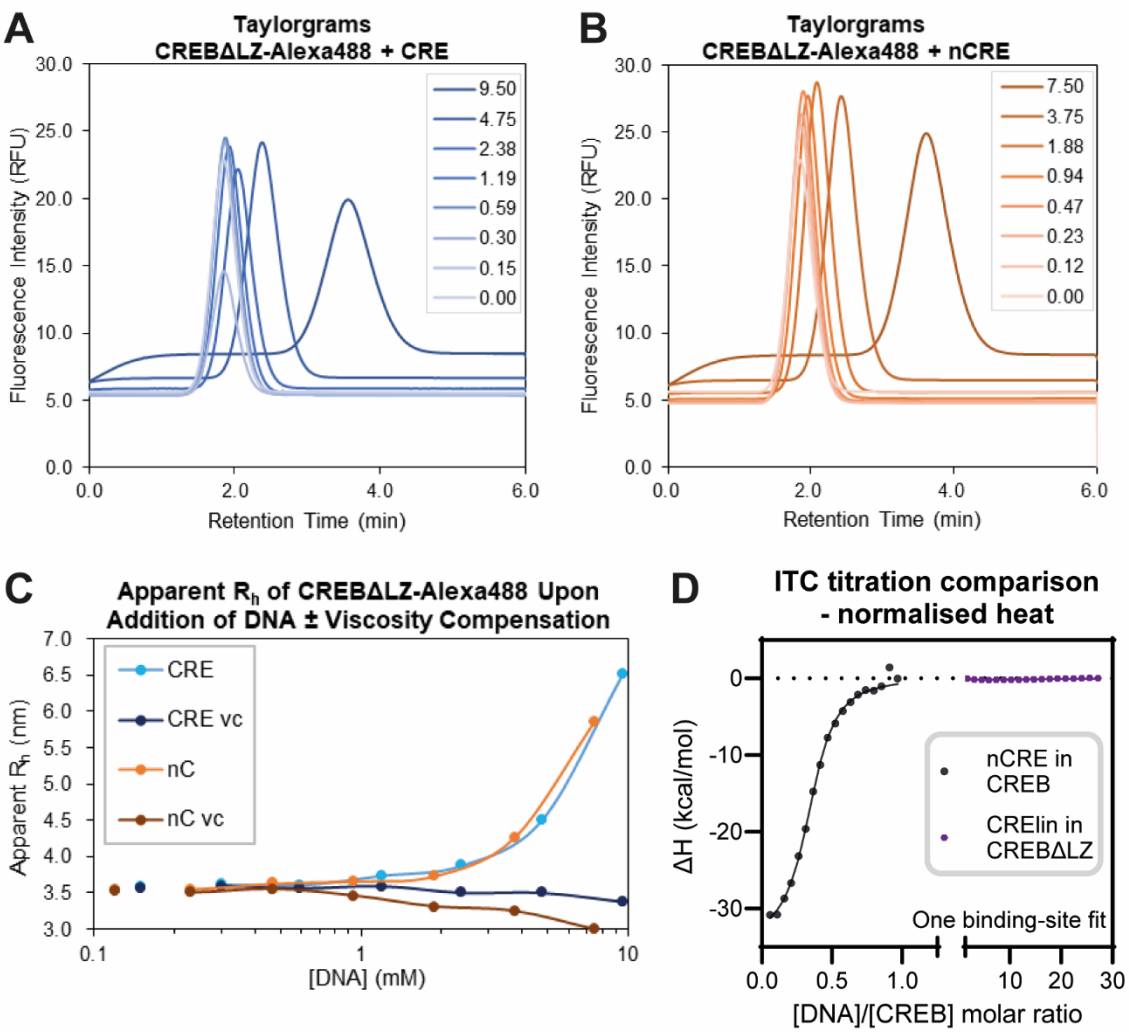


Figure 6.34: Other affinity methods suggest CREBΔLZ does not bind DNA. Conditions: pH 7.4, 25 °C. **A.** and **B.** Fidabio fluorescence reads of 1-5µM labelled CREB measuring hydrodynamic radius (R_h) show similar fluorescence intensity reads with increased DNA concentrations, changing only at very high concentrations. **C.** Titration of CREBΔLZ using Fidabio instruments, when correcting for viscosity ("vc"), show CREBΔLZ does not bind DNA. **D.** Low-c ITC titration of CREBΔLZ shows no thermodynamic change (especially compared to CREB), suggesting again that no binding occurs.

I also tried to characterise any interaction by ITC, but a low-c ITC titration (usually reliable for K_D but not n-value extraction) showed no signal change despite addition of up to 500 µM titrant (Figure 6.34D). This means that either the binding reaction does not cause a high enough heat change (unlikely considering previous binding reactions did), no binding occurs at the current concentrations, or the DNA is already saturated. Another method, looking at hydrodynamic radii changes, also shows no changes upon the addition of very high DNA concentrations, whether target or non-target (Figure 6.34A-C). Despite multiple

attempts to obtain affinities of CREB monomers to DNA, none of the methods give a reliable K_D estimate beyond the mM range and there is, in fact, no evidence to suggest monomers bind DNA at all.

6.6. Equilibrium modelling of species concentrations

Having obtained several K_D from the various species of CREB, both dimer and monomer it is possible to use simple equilibrium models to predict the equilibrium concentrations of various species in the kinetic scheme shown in Figure 6.1. This is done by solving a set of simultaneous equations (each pertaining to a different K_D). These equations are simple to write down and are described in the methods (Section 2.4). Of note, although there are 8 equations in total, there are only 6 independent ones due to the presence of two thermodynamic squares. It is convenient to select the two equations for which it is not possible to obtain a K_D estimate for exclusion from the process, which are those for CREB dimerization involving an already DNA bound monomer.

All that remains is to assign values to the binding affinities and total concentrations (Table 6.9). The most controversial K_D estimate is arguably for monomer-DNA interactions, for which an estimate from the CREBΔLZ-Alexa594 anisotropy titrations is used (1 mM).

Value	CREB	bZIP
$K_{D,homo}$	35 nM	9 nM
Monomer $K_{D,CRE}$	1000 µM	200 µM (Kuravsky)
Monomer $K_{D,nCRE}$	1000 µM	400 µM (Kuravsky)
Dimer $K_{D,CRE}$	1.4 nM	0.25 nM
Dimer $K_{D,nCRE}$	700 nM	175 nM
Total [CREB]	570 nM	
Total [CRE]	35 nM	

Table 6.9: Values set as constant for CREB equilibrium modelling.

6. Investigation of CREB dimerization and the stoichiometry of target search

In terms of concentrations there are roughly 5000 CREB binding sites (184–186), which means roughly 35 nM CRE sites (see methods). Meanwhile, cellular CREB concentration is in the OpenCell database (183) and can be set as a 570 nM constant, although this appears to be for whole-cells, and are likely higher in the smaller nucleus volume.

I calculated the concentration of different CREB and bZIP species at equilibrium for a range of competitor concentrations (Figure 6.35) up to 1 mM. There are expected to be ~1 mM nCRE sites equivalent in a human cell nucleus (see methods), though not all may be accessible due to the nature of chromatin. Therefore, considering the extrapolations to high competitor concentrations gives the best understanding of the situation within the cell nucleus. These show that CREB is mostly dimeric in the cell and that the DNA-bound monomer species is quasi non-existent. Moreover, it shows the non-bZIP domains of CREB contribute to the efficiency of target search, as more CREB is freely available to bind CRE at any given non-target DNA concentration e.g. 2-4 fold between 1 and 100 μM nCRE.

These simulations can also give an indirect glimpse into kinetic fluxes. In a steady-state situation the flux through monomer and dimer pathways can be understood as $k_7[\text{M.CRE}]_{eq}[\text{M}]_{eq}$ and $k_3[\text{D}]_{eq}[\text{CRE}]_{eq}$ respectively, k_3 being the dimeric k_{on} to target. Assuming that $k_7 \sim k_3$ this is equivalent to saying that $[\text{D}]_{eq}[\text{CRE}]_{eq}/[\text{M.CRE}]_{eq}[\text{M}]_{eq}$ complexes are formed through the dimer pathway for every one formed via the monomer pathway, and simulations show the numerator is greater than the denominator. This analysis therefore informs a quantitative understanding of why the dimer pathway is so favoured for CREB. It also implies, somewhat surprisingly, that the relative fluxes are independent of protein concentration and non-target DNA concentration.

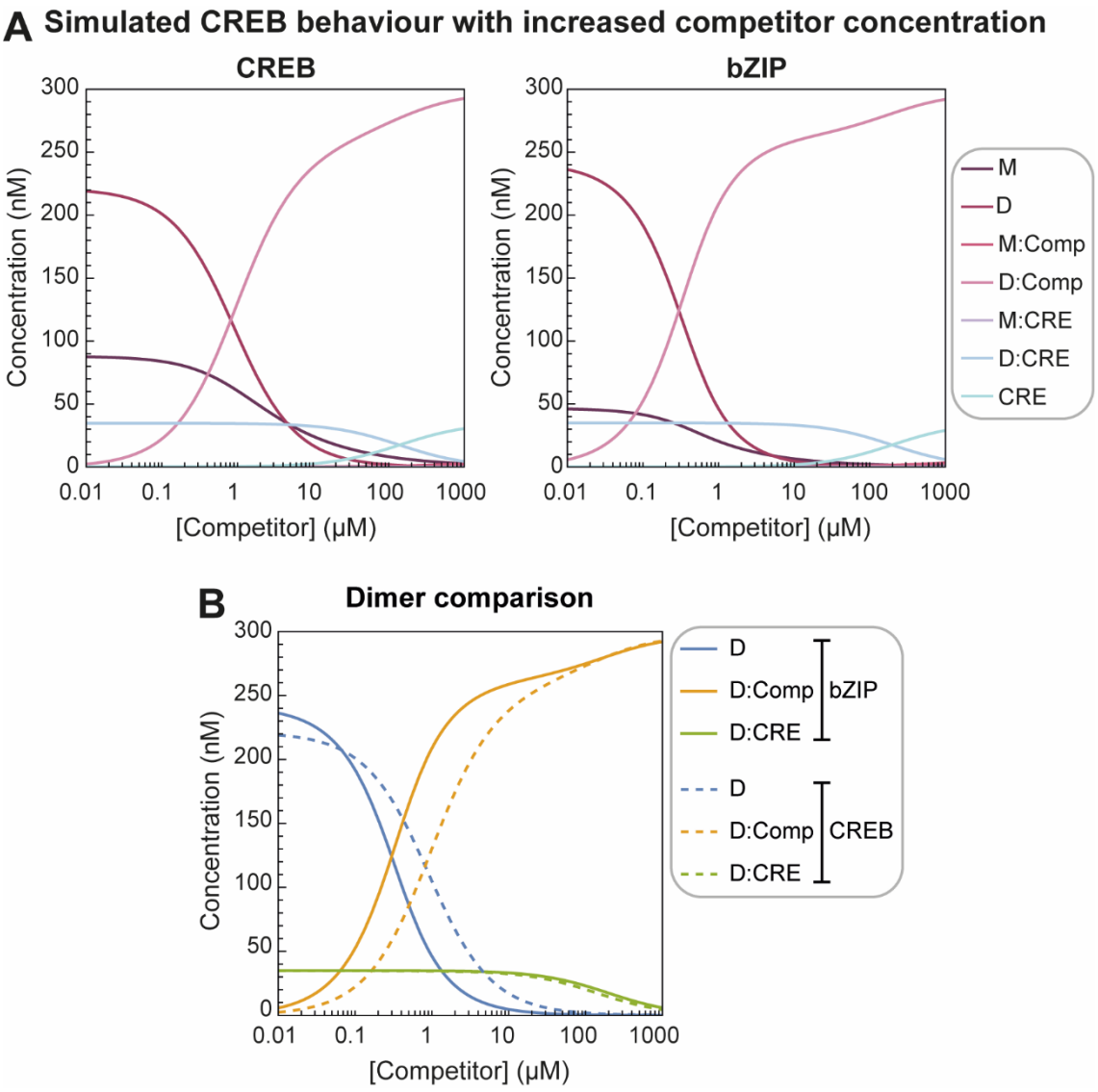


Figure 6.35: Simulations of CREB behaviour with increasing non-target competitor concentration. Simulations were achieved by simultaneously solving multiple K_D equations, keeping total [CREB] and [CRE] constant (Table 6.9). **A.** Both CREB and bZIP are mostly dimeric ("D"), with the monomer ("M") almost non-existent. CRE binding is eventually outcompeted by non-target DNA, showing the importance of other factors in ensuring CRE is bound in physiological conditions **B.** The influence of $K_{D,nCRE}$ on target search is apparent as there is more free CREB than bZIP at high competitor concentration. Although CRE is comparably bound, these are likely due to equilibrium values used for the simulations, and more free CREB is likely important in finding targets faster when initially exposed to DNA.

6.7. Discussion

In this Chapter, I have explored the contributions of both dimeric CREB and monomeric CREB to the target search process. Surprisingly, all evidence shows that the monomer pathway is inadequate: CREB has high homodimerization affinity, is mainly dimeric in solution and forced dimer mutants behave very similarly to pseudo wild-type while forced monomers are unable to bind DNA. Modelling shows a lack of monomer binding to DNA even at high competitor concentrations expected in the cell, and suggests that if CREB target search can be considered in terms of equilibrium properties the relative flux through the monomer pathway is tiny. This contradicts previously assertions that the monomer pathway is more adequate for bZIPs (157–160). However, these bZIPs have different properties altogether, that could favour the monomer pathway – for instance have a lower homodimer affinity like Gcn4 (247). Hence, the pathway preference may be more of a case-by-case basis.

Although CREB did behave mostly like a dimer, there were some interesting differences between CREB and CREB-GGC. Notably, CREB's kinetics were indicative of stronger binding to CREh than CREB-GGC, with slightly faster k_{on} and slightly lower ($k_{off,1}$). The k_{off} would ideally need repeating due to the presence of the artefacts, and these differences are rather small – within error for some methods – but it suggests there is some minor kinetic benefit to being monomeric.

This is the first recorded investigation into CREB homodimerization kinetics, and I have confirmed the K_D of homodimerization through a variety of methods and, uniquely for CREB, via stopped-flow apparatus. The value shows high homodimerization affinity, suggesting CREB is mainly present as a dimer in solution, contradicting a previous publication which suggested that CREB was monomeric in solution, therefore favouring the

monomer pathway (162). However, this conclusion was based on inconclusive AUC data which was interpreted to show CREB as a monomer in solution. This conclusion was recently confirmed by others (155). Interestingly, they obtained a 33 nM homodimerization K_D for labelled bZIP through FRET, very close to the relaxation kinetics value of 34 nM for fluorescently labelled CREB. The authors considered this value to be fairly unreliable due to absorption issues and mass photometry returned a lower K_D of 15 nM both for bZIP and CREB, close my chevron bZIP value. Because I used Tween-20 in my buffers, I suspect adsorption is limited. Therefore, this two-fold difference is likely due to the fluorophore or slightly altered buffer conditions. Both bZIP and CREB have comparable homodimerization K_D under similar conditions (Table 6.4), showing the CREB's TAD has little to no influence on the process. This is indirectly indicated by the anisotropy binding titrations to labelled CREh using bZIP and CREB being very similar (see Chapter 5), hinting at a comparable homodimerization K_D.

Furthermore, all methods across their study and my own give values in a very similar range both for bZIP and CREB – tens of nanomolar, showing limited to no influence of CREB's non-bZIP domains on dimerization (Table 6.4). Hopefully this goes some way to resolve the disagreement in the literature of this important value. As far as homodimerization affinities go, CREB is in the average for bZIPs, with some families (JUN, ATF6, CEBP) having tighter homodimerization affinity, and some (ATF3, NFE2) having weaker (136). Although these studies were done with bZIP domains only, CREB seems to have a relatively, but not remarkably, tight affinity for itself.

The increase in FRET signal with increasing DNA concentration could, at a glance, favour the monomer pathway. However, this does not necessarily follow. Either: monomers bind DNA with low affinity and DNA acts as a dimerization catalyst, on which CREB dimers can then find targets (a hybrid model). Or: the addition of DNA shifts the equilibrium

concentrations in favour of the dimer. Even if the monomer does not bind DNA, the free dimer population is depleted by DNA binding but repopulated from the free monomer population, increasing the proportion in dimer. This could be what the FRET titration shows. Although, the total CREB concentration in this experiment was low compared to the cell, 50 nM, and around that of CREB's $K_{D,homo}$. Hence, a large part of CREB would have been monomeric, and this effect, while undeniable, could be artificially overestimated as a result.

A similar increase in FRET signal is what led to the Fos:Jun monomer pathway conclusion (159,160) so it is worth re-examining the work in light of my findings. The modelling performed here for CREB shows it is perfectly possible to increase dimer populations with increased DNA concentration even when the dimer pathway operates almost exclusively. It is possible that other bZIPs also favour the dimer pathway.

The modelling may also indirectly show the importance of coactivators, antenna sequences and chromatin in target search. CREB, despite the influence of its non-bZIP domains, is largely sequestered by non-target DNA – to such an extent that search kinetics ought to be considerably reduced (near a 1000-fold, shown in section 5.3). Target search may therefore be greatly helped by: chromatin by limiting how much DNA is available in the first place; chromatin modifications, namely cytosine methylation, preventing CREB binding to DNA (184,254); antenna sequences (58,59) by increasing chances that a sliding event occurs near a target-site; and coactivators for either directing proteins to the correct area or creating long-lived binding events. Of course, there may be other processes that take place during target search (hopping, intersegmental transfer) which the modelling does not account for and could limit the influence of competitor DNA. This is already apparent at 10 μM competitor DNA, where traces begin to have two extractable phases, but more so for CREB than bZIP (section 5), potentially highlighting the role of the disordered transactivation domains in that regard, reminiscent of the monkey bar hypothesis (72).

Chapter 7

Perspectives and future directions

7.1. In depth biophysical characterisation shows CREB is a strong, fast, dimeric binder of DNA

Throughout the Thesis process, I have carried out multiple biophysical experiments to further understand the important protein CREB. In doing so, I have improved on past purification procedures to obtain high yields of CREB, almost twice as much as the reported dedicated method (193), at no additional time cost (Figures 3.18, 3.20, 3.23). I have also fluorescently labelled these products (Figure 4.9) and managed to circumvent issues related to precipitation (Figure 3.15) or forced dimerization (Figure 6.18-20) to fully explore CREB behaviour. Analyses of the protein have found heat denaturation to be a useful purification step since, although CREB is denatured by heat with a melting temperature of around 60 °C, it recovers from heat treatment extremely well (Figures 3.9, 4.6).

Overall, I have found that CREB is a high affinity binder for its target (Figures 4.8, 5.17, Table 5.2) and surprisingly medium affinity binder for non-target DNA (Figures 4.12-16, Table 5.4). This is one of the first few reported instance of both CREB's and a bZIP's affinity to non-target DNA (155). Anisotropy binding curves are not as reliable as kinetics measurements as the binding affinity is higher than the dimerization affinity and both

homodimerization and CRE binding are probed by this method (Figures 6.11, Table 6.4). However, kinetic measurements use more protein, which probes CRE binding more specifically (Figures 4.17-23, 5.17, Table 5.2). This has confirmed the very high affinity of CREB for its target DNA, around the nM range (107,155). There is also no evidence cysteine removal affects affinity for DNA to the extent that was previously reported (105).

CREB associates to DNA in an extremely fast manner, around 10-fold below the supposed diffusion limit (Figures 4.17, 5.24, Table 5.3) (214,215). This suggests electrostatically enhanced and directed binding, a known trait amongst transcription factors (as for the lac repressor (255) and Gcn4 (157,256), despite some uncertainty on the latter (257)). Relatedly, CREB's dissociation kinetics (Figures 4.19-23, Table 5.3), although showing some relatively long-lived complexes, are still lower compared to other bZIPs, where the residence time (inverse of the k_{off}) is longer (157,158,257). Another showcase of the importance and usefulness of kinetic studies is the apparent influence of the hairpin structure on DNA binding, as to it seems to affect dissociation exclusively (Figure 4.21). The published dissociation rate constant for CREB to linear CRE is also 5-fold lower than CREh's dominant k_{off} (206). CREB might therefore have an even higher affinity for linear CRE, which is not seen with the anisotropy titrations (Figure 4.8). The longer residence time could increase the window of opportunity for partners to bind and retain CREB at target promoters, especially since some longer CREB residence times have been observed (151).

That only dissociation is affected would indicate a mechanistic aspect of dissociation is at play, which could be explored further. Yet, an alternative method to obtain k_{off} should undeniably be sought. Expectations for dissociation traces is to be well-fit by a single exponential decay function. However, traces using CREh diluted into excess linear CRE show clear second exponential decay behaviour (Figure 4.19). Although such a behaviour was reported before (206), the values for the higher rate constant do not match. That could

be attributed to differences in the sequences used, as oligos involved did not have a full CRE site, or that the second rate constant was due to scanning or sliding factor, which is unlikely to be relevant in our case due to the smaller oligo size. The presence of DNA dilution artefacts from control traces (Figure 4.22) and switch to single exponential decay behaviour upon control correction (Figure 4.23) shows that this is likely a result of DNA into DNA dilution artefacts.

The behaviour of such artefacts should be accounted for. It is unknown how different concentration of unlabelled competitor affects the artefacts' extracted rate constants and amplitude. Once understood, it may be possible to correct the dissociation traces individually, allowing for a quick, cost-efficient verification of the currently obtained double exponential decay k_{off}. Interestingly, the 1 in 11 stopped-flow dilution setup appears to limit the influence of the artefacts, and this could prove an interesting work-around for this issue without a change in setup (Figure 4.19). However, this technique may simply be unsuitable to study dissociation kinetics, and alternatives may be required. These include retaining the stopped-flow but using relaxation kinetics (as used for homodimerization, Figures 6.5-9), monitoring fluorescence anisotropy instead of intensity, or switching technique entirely and using single-molecule fluorescence assays.

Preliminary results from our collaborators in the Kapanidis lab suggest it could be possible to study kinetics with their fluorescence microscopy-based setup. Although dissociation of CREB through fluorescence assays have been achieved already (206), applying them to our study would be useful to further study dissociation. As stated, potential ventures include further exploring the influence of the hairpin or CREB's individual domains on the dissociation process. Another possible use of our collaborator's setup is the study of nCRE binding kinetics and how domain deletions affect them. The CREh and dimerization aspects of the kinetic scheme were explored in this work, but the nCRE aspect

remains unexplored (figure 2.1, 6.1). The high K_D of binding to nCRE and lack of intensity signal change (figure 4.12) makes the kinetics of this interaction near-impossible to study using fluorescence intensity signal in a stopped-flow apparatus, hence the need for alternative methods. Completion of the scheme would allow for kinetic simulations, rather than equilibrium ones, more indicative of monomeric and dimeric fluxes. Lastly, there have been advanced kinetic studies of target search which used fluorescently labelled bacterial transcription factors (50,51,60). Therefore, ensuring that consistent k_{off} are obtained would allow an investigation into these aspects of CREB behaviour.

7.2. CREB most likely follows the dimer pathway, in what appears to be an exception for bZIPs so far

Previous experiments have all suggested a monomer pathway for bZIPs (157–160,162), meaning DNA-bound monomers would dimerise on the DNA. Yet, my data suggests CREB binds DNA following the dimer pathway, which indicates that dimerization occurs prior to binding DNA. Not only did CREB forced dimer constructs have similar biophysical properties to CREB (Figures 6.21-32, Tables 6.7-8); it was also impossible to accurately determine forced monomer K_D or to determine whether CREB monomers were binding DNA at all (Figures 6.33-34). This is similar to reports of dimerization impeded L318/325V (or L304/311V in isoform B) mutants and CREBΔLZ mutants' reduced ability to bind DNA (206,258). Still, greater dimerization levels were observed with increasing DNA concentration in FRET experiments (Figure 6.17). This can be explained by DNA-induced dimerization of monomers, apparently favouring the monomer pathway. However, a simple equilibrium explanation is also possible, where DNA depletes the dimer pool which is repopulated from the monomer pool. This would be favoured as simulations seldom show

monomer binding to DNA, and approximate predictions of monomer and dimer pathway fluxes also show the dimer pathway is greatly favoured (Figure 6.35).

Similarly, CREB has a strong homodimerization K_D, around 30nM (Figures 6.6-9, Table 6.4), but not extremely strong homodimer K_D for a bZIP (136). This makes it mainly dimeric in solution, as shown by SEC-MALS (Figure 6.2). Presence of a fluorophore looks to impede homodimerization, as another study found a similar K_D of homodimerization using labelled bZIP, but a 15 nM K_D when using unlabelled CREB (Table 6.4) (155). Nevertheless, this is an important milestone in addressing this important biophysical parameter on which there has been some controversy. Indeed, this represents a tens of nanomolar consensus for the homodimerization K_D reached using numerous methods (Table 6.4) (136,155). Notably, this refutes previous claims that CREB is monomeric (162), although there were made using trying to interpret confusing AUC results (Figure 6.3).

As previously stated, Gcn4, Jun, Fos and ATF2 all follow the monomer pathway, making CREB an exception amongst bZIPs thus far (157–160,162). Although, lower cellular concentrations (183) and higher homodimerization K_D (247) can explain why some bZIPs follow the monomer pathway when CREB follows the dimer pathway (Table 6.1). Nevertheless, an outstanding question is how widespread the monomer and dimer pathways are amongst bZIPs. This could be explored via binding assays using monomeric bZIP mutants. Tantalisingly, a few bZIPs (like the ATF family and CREB) have very similar target site sequences, making high-throughput assays a possibility (259).

It is also unclear if the cellular environment affects the pathway behaviour. This could be verified in cells using FRET using published procedures for homodimers (260), perhaps with a combination of endogenous and transfected protein with the FRET acceptor in excess. Single-particle tracking could be applied to compare the behaviour of monomers

(FRET acceptor alone) and dimers (FRET signal) during DNA target search. Should such a setup be made, it could be used to investigate if CREB incorporates liquid droplets formed by coactivators like CBP (226) by tagging both and checking for colocalization in droplets.

7.3. Disordered and activation domains self-inhibit CREB binding to DNA, improving the efficiency of target-search

There have been increasing indications that non-DNA binding domains (DBD) are involved in the DNA binding and target search process. For instance, aromatic and negatively charged residues are key elements of activation domains (18–21). Theoretically, these could compete with DNA for binding the negatively charged DNA binding domain (17), dislodge nucleosomes (74,75) or recognise other proteins at target sites (67). Correspondingly, multiple studies have observed that mutations of disordered non-DNA binding domains greatly impact the target search process of some (68–71) but not all transcription factors, as some are more DBD dependent (73). Multivalent interactions from these domains, typical of fuzzy complexes (76), are likely to be involved (69,91). These could even participate in phase separation which the Mediator complex and many associated factors are involved in, to increase transcription factor activity and facilitate target search (57,81,82,88,228,261). Disordered domains could also impact specific parts of the facilitated diffusion process like jumping to a nearby strand or intersegmental transfer, as proposed with the monkey bar mechanism (72), where the disordered domains would find and bind a nearby DNA strand from which an already DNA bound factor would switch to. Yet, much remains to be understood about the role of disordered domains in target search.

Here, I reinforce this hypothesis, by showing that CREB's non-bZIP domains decrease affinity to DNA around 4 to 5-fold (Figures 5.6-16), which has very recently been confirmed through different methods (155). However, I build on these findings to show that

all non-bZIP domains, both the KID activation domain and Q-rich IDRs, contribute to this self-inhibition, even in the absence of partners (Figure 5.17-28, Tables 5.2-4). It is likely that the negatively charged KID decreases affinity by auto-inhibitory binding to the bZIP, as previously reported by NMR (26). This gives credence to the hypothesis that most activation domains would have a similar function. Indeed they are mainly negatively charged like the KID (Figures 5.1-2) (18–21), and short negatively charged stretches are known to kinetically improve target search (212). Yet, it is unclear how the Q-rich domains decrease affinity. A possibility is that they form fuzzy-complexes, segregating CREB away from the DNA. Lastly, these domains do not affect expression induced by cAMP signalling, which is wholly dependent on the KID domain (149).

It was previously reported that gradual deletions of a transcription factor disordered domain also led to gradual delocalisation of its DNA binding profile (68). Interestingly, the findings that each of CREB's non-DBD domains impedes DNA binding can help explain this. Gradual deletion of these domains would concomitantly increase affinity to DNA in general, leading to more frequent off-target binding events. This also complements another cellular study which showed that CREB's non-bZIP domains greatly contributed to increasing basal expression (149), as these other domains would help localise it. The fact that CREB likely binds targets without activation by phosphorylation (and CBP binding), as shown by basal expression of reporters, and that CRTC coactivators only bind already DNA-bound CREB (145), suggests the influence of disordered domains is paramount for the target search process of CREB, whereas the influence of coactivators may be limited. Somewhat outside the scope of this study, the decrease of non-target affinity could also impact aspects of the facilitated diffusion process. Perhaps it facilitates dissociation during sliding, potentially increasing the number of intersegmental transfer or jumping events, giving more importance to the aforementioned monkey bar hypothesis (72).

This is interestingly reminiscent of findings on MITF, a bZIP transcription factor that can be acetylated to promote activity. Interestingly, it was shown that using an acetylated-mimic mutant of MITF resulted in a decreased affinity to DNA (262). It was reasoned that MITF was bound to non-target DNA and released by acetylation, increasing the available factor pool, activity and specific binding. Here, the difference between the bZIP mutant and full length CREB difference is similar, where bZIP binds to non-target DNA more readily than CREB. This highlights the importance of the TAD domains in limiting off-target association and making CREB more available for target binding, akin to the role of acetylation in MITF. Moreover, CREB behaviour could similarly be altered with its extensive post-translational modifications (103,121,147,148), as already shown by CREB phosphorylation decreasing affinity for DNA (26).

Of course, this does not invalidate previously formulated search pathways, which are likely still predominant for transcription factor localisation. Indeed, it was observed via the equilibrium binding simulations with increasing nCRE concentrations (figure 6.35), that, eventually, all CREB would be bound to non-target DNA, regardless of the presence of its TAD domains. Other factors are likely at play to prevent this from happening. Chromatin is likely important in this regard, as DNA packing limits how much DNA is available for binding in the first place. This would be crucial in a case where coactivators may not be as important, which may be relevant to CREB. The 2-step search process, which involves the transcription factor's disordered domains binding protein partners before the DNA (Figure 1.5) (67), can also explain many of the aforementioned results. Relatedly, coactivators could refine the search space too. Moreover, not all factors explore DNA in an unrestricted manner (56), sometimes being even segregated away from the DNA (55). Additionally, a bacterial transcription factor transfected in human cells, without such coactivator help, takes very

long to find its target, showing the importance of those partners (although this could be due to differences in scanning DNA between bacteria and eukaryotes) (46).

Kinetics assays could be modified to test the effect of some of these parameters, for instance, how the presence of coactivators would affect the kinetics of binding. As CTRC binds already DNA-bound CREB (145), perhaps it lengthens the lifetime of the complex. Similarly, CBP binding to phosphorylated KID via its KIX domain (23) could relieve its inhibitory effect on association and affect dissociation as well. Relatedly, phosphorylation is essential for CREB activity and it is known to impede CREB binding to DNA (26). Since these sites are located near the KID, this gives further indication that the KID negative charges contribute to self-inhibition. Post-translational modifications were largely ignored in this Thesis, but there remain questions as to their influence related to DNA binding. For example, activated CREB is phosphorylated at S133 (CREB-A numbering)(94,149) and the effect of S133 on target search, if any, is unclear. Moreover, CREB is modified in other way, most notably acetylation (147), and their influence on DNA binding parameters is unknown. The proposed role of disordered activation domains as nucleosome detergents, helping dislodge DNA from nucleosomes (73–75) could also be tested by using packed DNA instead of linear short oligonucleotides.

Nevertheless, there is now a clearly defined role for IDRs in limiting off-target binding events, increasing the efficiency of the facilitated diffusion target search process (Figures 5.14-16) (39–42). This also potentially gives greater role to “antenna” regions for target localisation (58,59). Antenna sequences usually have certain characteristics similar to that of a target site. As the CRE sequence (TGACGTCA) has a higher GC content (50%) than the rest of the genome (40%), antenna sequences could have a higher GC content to direct transcription factors. Hence, measuring affinity to random sequences with varying GC content could give insight into the CREB search pathway. A kinetic approach could also be

tested with CRE oligos flanked by various sequences, and the search time modulated with the influence of competitors to probe the DNA sliding and intersegmental parameters too, as attempted previously (49). The influence of disordered domains on decreasing affinity for DNA could be reduced for antenna regions, allowing for more efficient funnelling to target sites. Along with the potential features of disordered DBDs, such as the hypothetical ability to switch between fast, low-affinity "scanning" and slow, high-affinity "binding" conformations (43,63,263), perhaps already observed in p53 (64,65), this could greatly improve target search efficiency.

Lastly, both AlphaFold structural predictions and CD spectra analysis have suggested that CREB does form some β-sheets, possibly in the Q2 region (figures 4.1, 4.3). This is highly unusual for a disordered protein. Yet, it was recently confirmed via NMR studies that functionally relevant, residual β-sheets are formed by CREB's TAD, which help recruit transcriptional partners (154). Interestingly, AlphaFold suggests these structures form near the bZIP's DNA binding interface. These could impede binding to DNA, which binding to partners to would relieve, a testable hypothesis through mutagenesis or the addition of said partners in kinetics and equilibrium binding assays.

References

1. Crick F. Central Dogma of Molecular Biology. Nature. 1970 Aug;227(5258):561–3.

2. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Chromosomal DNA and its packaging in the chromatin fiber. In: Molecular Biology of the Cell 4th edition. Garland Science; 2002.

3. Hoencamp C, Rowland BD. Genome control by SMC complexes. Nat Rev Mol Cell Biol. 2023 Sep;24(9):633–50.

4. Luger K, Dechassa ML, Tremethick DJ. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nat Rev Mol Cell Biol. 2012 Jul;13(7):436–47.

5. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. Nat Rev Genet. 2019 Apr;20(4):207–20.

6. Kim S, Yu NK, Kaang BK. CTCF as a multifunctional protein in genome regulation and gene expression. Exp Mol Med. 2015 Jun 5;47(6):e166–e166.

7. Bentley DL. Coupling mRNA processing with transcription in time and space. Nat Rev Genet. 2014 Mar;15(3):163–75.

8. Sonenberg N, Hinnebusch AG. Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. Cell. 2009 Feb;136(4):731–45.

9. Moore MJ, Proudfoot NJ. Pre-mRNA Processing Reaches Back to Transcription and Ahead to Translation. Cell. 2009 Feb;136(4):688–700.

10. Licatalosi DD, Darnell RB. RNA processing and its regulation: global insights into biological networks. Nat Rev Genet. 2010 Jan;11(1):75–87.

11. Weidemüller P, Kholmatov M, Petsalaki E, Zaugg JB. Transcription factors: Bridge between cell signaling and gene regulation. Proteomics. 2021 Dec;21(23–24):2000034.

12. Andersson R, Sandelin A. Determinants of enhancer and promoter activities of regulatory elements. Nat Rev Genet. 2020 Feb;21(2):71–87.

13. Richter WF, Nayak S, Iwasa J, Taatjes DJ. The Mediator complex as a master regulator of transcription by RNA polymerase II. Nat Rev Mol Cell Biol. 2022 Nov;23(11):732–49.

14. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. Cell. 2018 Feb;172(4):650–65.

15. Cooper G M. Regulation of Transcription in Eukaryotes. In: The Cell: A Molecular Approach 2nd edition. Sunderland (MA): Sinauer Associates; 2000.

16. Bushweller JH. Targeting transcription factors in cancer — from undruggable to reality. Nat Rev Cancer. 2019 Nov;19(11):611–24.

17. Soto LF, Li Z, Santoso CS, Berenson A, Ho I, Shen VX, et al. Compendium of human transcription factor effector domains. Molecular Cell. 2022 Feb;82(3):514–26.

18. Staller MV, Ramirez E, Kotha SR, Holehouse AS, Pappu RV, Cohen BA. Directed mutational scanning reveals a balance between acidic and hydrophobic residues in strong human activation domains. Cell Syst. 2022 Apr 20;13(4):334-345.e5.

19. Erijman A, Kozlowski L, Sohrabi-Jahromi S, Fishburn J, Warfield L, Schreiber J, et al. A High-Throughput Screen for Transcription Activation Domains Reveals Their Sequence Features and Permits Prediction by Deep Learning. Molecular Cell. 2020 Jun;78(5):890-902.e6.

20. Ravarani CN, Erkina TY, De Baets G, Dudman DC, Erkine AM, Babu MM. High-throughput discovery of functional disordered regions: investigation of transactivation domains. Molecular Systems Biology. 2018 May;14(5):e8190.

21. Broyles BK, Gutierrez AT, Maris TP, Coil DA, Wagner TM, Wang X, et al. Activation of gene expression by detergent-like protein domains. iScience. 2021 Sep;24(9):103017.

22. Dahal L, Shammas SL, Clarke J. Phosphorylation of the IDP KID Modulates Affinity for KIX by Increasing the Lifetime of the Complex. Biophysical Journal. 2017 Dec;113(12):2706–12.

23. Radhakrishnan I, Pérez-Alvarado GC, Parker D, Dyson HJ, Montminy MR, Wright PE. Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of CREB: A Model for Activator:Coactivator Interactions. Cell. 1997 Dec;91(6):741–52.

24. Zor T, Mayr BM, Dyson HJ, Montminy MR, Wright PE. Roles of Phosphorylation and Helix Propensity in the Binding of the KIX Domain of CREB-binding Protein by Constitutive (c-Myb) and Inducible (CREB) Activators. Journal of Biological Chemistry. 2002 Nov;277(44):42241–8.

25. Chrivia JC, Kwok RPS, Lamb N, Hagiwara M, Montminy MR, Goodman RH. Phosphorylated CREB binds specifically to the nuclear protein CBP. Nature. 1993 Oct;365(6449):855–9.

26. Shnitkind S, Martinez-Yamout MA, Dyson HJ, Wright PE. Structural Basis for Graded Inhibition of CREB:DNA Interactions by Multisite Phosphorylation. Biochemistry. 2018 Dec 26;57(51):6964–72.

27. Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK. Intrinsic Disorder in Transcription Factors. Biochemistry. 2006 Jun 1;45(22):6873–88.

28. Uversky VN, Gillespie JR, Fink AL. Why are 'natively unfolded' proteins unstructured under physiologic conditions? Proteins. 2000 Nov 15;41(3):415–27.

29. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol. 2005 Mar;6(3):197–208.

30. Uversky VN. Natively unfolded proteins: a point where biology waits for physics. Protein Sci. 2002 Apr;11(4):739–56.

31. Shammas SL, Crabtree MD, Dahal L, Wicky BIM, Clarke J. Insights into Coupled Folding and Binding Mechanisms from Kinetic Studies. J Biol Chem. 2016 Mar 25;291(13):6689–95.

32. Oates ME, Romero P, Ishida T, Ghalwash M, Mizianty MJ, Xue B, et al. D^2P^2: database of disordered protein predictions. Nucleic Acids Res. 2013 Jan;41(Database issue):D508-516.

33. Smith NC, Kuravsky M, Shammas SL, Matthews JM. Binding and folding in transcriptional complexes. Curr Opin Struct Biol. 2021 Feb;66:156–62.

34. Gunasekaran K, Tsai CJ, Kumar S, Zanuy D, Nussinov R. Extended disordered proteins: targeting function with less scaffold. Trends in Biochemical Sciences. 2003 Feb;28(2):81–5.

35. Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. Proc Natl Acad Sci USA. 2000 Aug;97(16):8868–73.

36. van der Lee R, Lang B, Kruse K, Gsponer J, Sánchez de Groot N, Huynen MA, et al. Intrinsically disordered segments affect protein half-life in the cell and during evolution. Cell Rep. 2014 Sep 25;8(6):1832–44.

37. Darling AL, Uversky VN. Intrinsic Disorder and Posttranslational Modifications: The Darker Side of the Biological Dark Matter. Front Genet. 2018 May 4;9:158.

38. Stewart AJ, Hannenhalli S, Plotkin JB. Why transcription factor binding sites are ten nucleotides long. Genetics. 2012 Nov;192(3):973–85.

39. Berg OG, Winter RB, Von Hippel PH. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. Biochemistry. 1981 Nov 1;20(24):6929–48.

40. Von Hippel PH, Berg OG. Facilitated Target Location in Biological Systems. Journal of Biological Chemistry. 1989 Jan;264(2):675–8.

41. Winter RB, Berg OG, Von Hippel PH. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor-operator interaction: kinetic measurements and conclusions. Biochemistry. 1981 Nov 1;20(24):6961–77.

42. Winter RB, Von Hippel PH. Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The Escherichia coli lac repressor-operator interaction: equilibrium measurements. Biochemistry. 1981 Nov 1;20(24):6948–60.

43. Slutsky M, Mirny LA. Kinetics of Protein-DNA Interaction: Facilitated Target Location in Sequence-Dependent Potential. Biophysical Journal. 2004 Dec;87(6):4021–35.

44. Mirny L, Slutsky M, Wunderlich Z, Tafvizi A, Leith J, Kosmrlj A. How a protein searches for its site on DNA: the mechanism of facilitated diffusion. J Phys A: Math Theor. 2009 Oct 30;42(43):434013.

45. Kong M, Van Houten B. Rad4 recognition-at-a-distance: Physical basis of conformation-specific anomalous diffusion of DNA repair proteins. Progress in Biophysics and Molecular Biology. 2017 Aug;127:93–104.

46. Normanno D, Boudarène L, Dugast-Darzacq C, Chen J, Richter C, Proux F, et al. Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. Nat Commun. 2015 Jul 7;6(1):7357.

47. Cui TJ, Klein M, Hegge JW, Chandradoss SD, Van Der Oost J, Depken M, et al. Argonaute bypasses cellular obstacles without hindrance during target search. Nat Commun. 2019 Sep 26;10(1):4390.

48. Itoh Y, Murata A, Takahashi S, Kamagata K. Intrinsically disordered domain of tumor suppressor p53 facilitates target search by ultrafast transfer between different DNA strands. Nucleic Acids Research. 2018 Aug 21;46(14):7261–9.

49. Esadze A, Iwahara J. Stopped-Flow Fluorescence Kinetic Study of Protein Sliding and Intersegment Transfer in the Target DNA Search Process. Journal of Molecular Biology. 2014 Jan;426(1):230–44.

50. Marklund E, Van Oosten B, Mao G, Amselem E, Kipper K, Sabantsev A, et al. DNA surface exploration and operator bypassing during target search. Nature. 2020 Jul 30;583(7818):858–61.

51. Hammar P, Leroy P, Mahmutovic A, Marklund EG, Berg OG, Elf J. The *lac* Repressor Displays Facilitated Diffusion in Living Cells. Science. 2012 Jun 22;336(6088):1595–8.

52. Tafvizi A, Huang F, Leith JS, Fersht AR, Mirny LA, Van Oijen AM. Tumor Suppressor p53 Slides on DNA with Low Friction and High Stability. Biophysical Journal. 2008 Jul;95(1):L01–3.

53. Stracy M, Schweizer J, Sherratt DJ, Kapanidis AN, Uphoff S, Lesterlin C. Transient non-specific DNA binding dominates the target search of bacterial DNA-binding proteins. Molecular Cell. 2021 Apr;81(7):1499-1514.e6.

54. Jana T, Brodsky S, Barkai N. Speed–Specificity Trade-Offs in the Transcription Factors Search for Their Genomic Binding Sites. Trends in Genetics. 2021 May;37(5):421–32.

55. Yao J, Fetter RD, Hu P, Betzig E, Tjian R. Subnuclear segregation of genes and core promoter factors in myogenesis. Genes Dev. 2011 Mar 15;25(6):569–80.

56. Izeddin I, Récamier V, Bosanac L, Cissé II, Boudarene L, Dugast-Darzacq C, et al. Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. eLife. 2014 Jun 12;3:e02230.

57. Kent S, Brown K, Yang C hsun, Alsaihati N, Tian C, Wang H, et al. Phase-Separated Transcriptional Condensates Accelerate Target-Search Process Revealed by Live-Cell Single-Molecule Imaging. Cell Reports. 2020 Oct;33(2):108248.

58. Afek A, Schipper JL, Horton J, Gordân R, Lukatsky DB. Protein−DNA binding in the absence of specific base-pair recognition. Proc Natl Acad Sci USA. 2014 Dec 2;111(48):17140–5.

59. Castellanos M, Mothi N, Muñoz V. Eukaryotic transcription factors can track and control their target genes using DNA antennas. Nat Commun. 2020 Jan 28;11(1):540.

60. Marklund E, Mao G, Yuan J, Zikrin S, Abdurakhmanov E, Deindl S, et al. Sequence specificity in DNA binding is mainly governed by association. Science. 2022 Jan 28;375(6579):442–5.

61. Janssen A, Colmenares SU, Karpen GH. Heterochromatin: Guardian of the Genome. Annu Rev Cell Dev Biol. 2018 Oct 6;34(1):265–88.

62. International Human Genome Sequencing Consortium, Whitehead Institute for Biomedical Research, Center for Genome Research:, Lander ES, Linton LM, Birren B, Nusbaum C, et al. Initial sequencing and analysis of the human genome. Nature. 2001 Feb 15;409(6822):860–921.

63. Zandarashvili L, Esadze A, Vuzman D, Kemme CA, Levy Y, Iwahara J. Balancing between affinity and speed in target DNA search by zinc-finger proteins via modulation of dynamic conformational ensemble. Proc Natl Acad Sci U S A. 2015 Sep 15;112(37):E5142-5149.

64. Leith JS, Tafvizi A, Huang F, Uspal WE, Doyle PS, Fersht AR, et al. Sequence-dependent sliding kinetics of p53. Proc Natl Acad Sci USA. 2012 Oct 9;109(41):16552–7.

65. Murata A, Ito Y, Kashima R, Kanbayashi S, Nanatani K, Igarashi C, et al. One-Dimensional Sliding of p53 Along DNA Is Accelerated in the Presence of Ca2+ or Mg2+ at Millimolar Concentrations. Journal of Molecular Biology. 2015 Aug;427(16):2663–78.

66. Munshi S, Gopi S, Asampille G, Subramanian S, Campos LA, Atreya HS, et al. Tunable order–disorder continuum in protein–DNA interactions. Nucleic Acids Research. 2018 Sep 28;46(17):8700–9.

67. Staller MV. Transcription factors perform a 2-step search of the nucleus. Rine J, editor. Genetics. 2022 Sep 30;222(2):iyac111.

68. Brodsky S, Jana T, Mittelman K, Chapal M, Kumar DK, Carmi M, et al. Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. Molecular Cell. 2020 Aug;79(3):459-471.e4.

69. Chen Y, Cattoglio C, Dailey GM, Zhu Q, Tjian R, Darzacq X. Mechanisms governing target search and binding dynamics of hypoxia-inducible factors. eLife. 2022 Nov 2;11:e75064.

70. Gera T, Jonas F, More R, Barkai N. Evolution of binding preferences among whole-genome duplicated transcription factors. eLife. 2022 Apr 11;11:e73225.

71. Chong S, Dugast-Darzacq C, Liu Z, Dong P, Dailey GM, Cattoglio C, et al. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science. 2018 Jul 27;361(6400):eaar2555.

72. Vuzman D, Azia A, Levy Y. Searching DNA via a 'Monkey Bar' mechanism: the significance of disordered tails. J Mol Biol. 2010 Feb 26;396(3):674–84.

73. Kumar DK, Jonas F, Jana T, Brodsky S, Carmi M, Barkai N. Complementary strategies for directing in vivo transcription factor binding through DNA binding domains and intrinsically disordered regions. Molecular Cell. 2023 May;83(9):1462-1473.e5.

74. Erkine AM. 'Nonlinear' Biochemistry of Nucleosome Detergents. Trends in Biochemical Sciences. 2018 Dec;43(12):951–9.

75. Erkina TY, Erkine AM. Nucleosome distortion as a possible mechanism of transcription activation domain function. Epigenetics & Chromatin. 2016 Dec;9(1):40.

76. Tuttle LM, Pacheco D, Warfield L, Luo J, Ranish J, Hahn S, et al. Gcn4-Mediator Specificity Is Mediated by a Large and Dynamic Fuzzy Protein-Protein Complex. Cell Reports. 2018 Mar;22(12):3251–64.

77. Harmon TS, Holehouse AS, Rosen MK, Pappu RV. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. eLife. 2017 Nov 1;6:e30294.

78. Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, et al. Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. Cell. 2012 May;149(4):753–67.

79. Dignon GL, Zheng W, Kim YC, Mittal J. Temperature-Controlled Liquid–Liquid Phase Separation of Disordered Proteins. ACS Cent Sci. 2019 May 22;5(5):821–30.

80. Lu H, Liu R, Zhou Q. Balanced between order and disorder: a new phase in transcription elongation control and beyond. Transcription. 2019 May 27;10(3):157–63.

81. Zamudio AV, Dall'Agnese A, Henninger JE, Manteiga JC, Afeyan LK, Hannett NM, et al. Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes. Molecular Cell. 2019 Dec;76(5):753-766.e6.

82. Sabari BR, Dall'Agnese A, Boija A, Klein IA, Coffey EL, Shrinivas K, et al. Coactivator condensation at super-enhancers links phase separation and gene control. Science. 2018 Jul 27;361(6400):eaar3958.

83. Alberti S, Gladfelter A, Mittag T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. Cell. 2019 Jan 24;176(3):419–34.

84. Ruff KM, Roberts S, Chilkoti A, Pappu RV. Advances in Understanding Stimulus-Responsive Phase Behavior of Intrinsically Disordered Protein Polymers. Journal of Molecular Biology. 2018 Nov;430(23):4619–35.

85. Banani SF, Rice AM, Peeples WB, Lin Y, Jain S, Parker R, et al. Compositional Control of Phase-Separated Cellular Bodies. Cell. 2016 Jul 28;166(3):651–63.

86. Wang J, Choi JM, Holehouse AS, Lee HO, Zhang X, Jahnel M, et al. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. Cell. 2018 Jul 26;174(3):688-699.e16.

87. Nott TJ, Petsalaki E, Farber P, Jervis D, Fussner E, Plochowietz A, et al. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. Mol Cell. 2015 Mar 5;57(5):936–47.

88. Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio AV, et al. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell. 2018 Dec;175(7):1842-1855.e16.

89. Cho WK, Spille JH, Hecht M, Lee C, Li C, Grube V, et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science. 2018 Jul 27;361(6400):412–5.

90. Zhang H, Elbaum-Garfinkle S, Langdon EM, Taylor N, Occhipinti P, Bridges AA, et al. RNA Controls PolyQ Protein Phase Transitions. Molecular Cell. 2015 Oct;60(2):220–30.

91. Trojanowski J, Frank L, Rademacher A, Mücke N, Grigaitis P, Rippe K. Transcription activation is enhanced by multivalent interactions independent of phase separation. Molecular Cell. 2022 May;82(10):1878-1893.e10.

92. Rieder D, Trajanoski Z, McNally JG. Transcription factories. Front Genet. 2012;3:221.

93. Marceau AH, Brison CM, Nerli S, Arsenault HE, McShan AC, Chen E, et al. An order-to-disorder structural switch activates the FoxM1 transcription factor. eLife. 2019 May 28;8:e46131.

94. Gonzalez GA, Montminy MR. Cyclic AMP stimulates somatostatin gene transcription by phosphorylation of CREB at serine 133. Cell. 1989 Nov;59(4):675–80.

95. Sands WA, Palmer TM. Regulating gene transcription in response to cyclic AMP elevation. Cellular Signalling. 2008 Mar;20(3):460–6.

96. Montminy MR, Bilezikjian LM. Binding of a nuclear protein to the cyclic-AMP response element of the somatostatin gene. Nature. 1987 Jul;328(6126):175–8.

97. Enslen H, Sun P, Brickey D, Soderling SH, Klamo E, Soderling TR. Characterization of Ca2+/calmodulin-dependent protein kinase IV. Role in transcriptional regulation. J Biol Chem. 1994 Jun 3;269(22):15520–7.

98. Tokumitsu H, Brickey DA, Glod J, Hidaka H, Sikela J, Soderling TR. Activation mechanisms for Ca2+/calmodulin-dependent protein kinase IV. Identification of a brain CaM-kinase IV kinase. Journal of Biological Chemistry. 1994 Nov;269(46):28640–7.

99. Xing J, Ginty DD, Greenberg ME. Coupling of the RAS-MAPK pathway to gene activation by RSK2, a growth factor-regulated CREB kinase. Science. 1996 Aug 16;273(5277):959–63.

100. Du K, Montminy M. CREB Is a Regulatory Target for the Protein Kinase Akt/PKB. Journal of Biological Chemistry. 1998 Dec;273(49):32377–9.

101. Zhang L, Jope RS. Oxidative stress differentially modulates phosphorylation of ERK, p38 and CREB induced by NGF or EGF in PC12 cells☆. Neurobiology of Aging. 1999 May;20(3):271–8.

102. Di Giacomo V, Sancilio S, Caravatta L, Rana RA, Di Pietro R, Cataldi A. Regulation of CREB activation by p38 mitogen activated protein kinase during human primary erythroblast differentiation. Int J Immunopathol Pharmacol. 2009;22(3):679–88.

103. Steven A, Seliger B. Control of CREB expression in tumors: from molecular mechanisms and signal transduction pathways to therapeutic target. Oncotarget. 2016 Jun 7;7(23):35454–65.

104. Ahmed MB, Alghamdi AAA, Islam SU, Lee JS, Lee YS. cAMP Signaling in Cancer: A PKA-CREB and EPAC-Centric Approach. Cells. 2022 Jun 24;11(13):2020.

105. Williams JS, Dixon JE, Andrisani OM. Binding Constant Determination Studies Utilizing Recombinant ΔCREB Protein. DNA and Cell Biology. 1993 Mar;12(2):183–90.

106. Busch SJ, Sassone-Corsi P. Dimers, leucine zippers and DNA-binding domains. Trends Genet. 1990 Feb;6(2):36–40.

107. Richards JP, Bächinger HP, Goodman RH, Brennan RG. Analysis of the Structural Properties of cAMP-responsive Element-binding Protein (CREB) and Phosphorylated CREB. Journal of Biological Chemistry. 1996 Jun;271(23):13716–23.

108. Amoutzias GD, Veron AS, Weiner J, Robinson-Rechavi M, Bornberg-Bauer E, Oliver SG, et al. One billion years of bZIP transcription factor evolution: conservation and change in dimerization and DNA-binding site specificity. Mol Biol Evol. 2007 Mar;24(3):827–35.

109. Rudolph D, Tafuri A, Gass P, Hämmerling GJ, Arnold B, Schütz G. Impaired fetal T cell development and perinatal lethality in mice lacking the cAMP response element binding protein. Proc Natl Acad Sci U S A. 1998 Apr 14;95(8):4481–6.

110. Bleckmann SC, Blendy JA, Rudolph D, Monaghan AP, Schmid W, Schütz G. Activating transcription factor 1 and CREB are important for cell survival during early mouse development. Mol Cell Biol. 2002 Mar;22(6):1919–25.

111. Bourtchuladze R, Frenguelli B, Blendy J, Cioffi D, Schutz G, Silva AJ. Deficient long-term memory in mice with a targeted mutation of the cAMP-responsive element-binding protein. Cell. 1994 Oct;79(1):59–68.

112. Kandel ER. The molecular biology of memory: cAMP, PKA, CRE, CREB-1, CREB-2, and CPEB. Mol Brain. 2012;5(1):14.

113. Ortega-Martínez S. A new perspective on the role of the CREB family of transcription factors in memory consolidation via adult hippocampal neurogenesis. Front Mol Neurosci. 2015;8:46.

114. Sano Y, Shobe JL, Zhou M, Huang S, Shuman T, Cai DJ, et al. CREB Regulates Memory Allocation in the Insular Cortex. Current Biology. 2014 Dec;24(23):2833–7.

115. Sekeres MJ, Neve RL, Frankland PW, Josselyn SA. Dorsal hippocampal CREB is both necessary and sufficient for spatial memory. Learn Mem. 2010 Jun;17(6):280–3.

116. Zhou Y, Won J, Karlsson MG, Zhou M, Rogerson T, Balaji J, et al. CREB regulates excitability and the allocation of memory to subsets of neurons in the amygdala. Nat Neurosci. 2009 Nov;12(11):1438–43.

117. Jagannath A, Butler R, Godinho SIH, Couch Y, Brown LA, Vasudevan SR, et al. The CRTC1-SIK1 Pathway Regulates Entrainment of the Circadian Clock. Cell. 2013 Aug;154(5):1100–11.

118. Xiao X, Li BX, Mitton B, Ikeda A, Sakamoto KM. Targeting CREB for Cancer Therapy: Friend or Foe. CCDT. 2010 Jun 1;10(4):384–91.

119. Sapio L, Salzillo A, Ragone A, Illiano M, Spina A, Naviglio S. Targeting CREB in Cancer Therapy: A Key Candidate or One of Many? An Update. Cancers. 2020 Oct 28;12(11):3166.

120. Zhang H, Kong Q, Wang J, Jiang Y, Hua H. Complex roles of cAMP–PKA–CREB signaling in cancer. Exp Hematol Oncol. 2020 Dec;9(1):32.

121. Steven A, Friedrich M, Jank P, Heimer N, Budczies J, Denkert C, et al. What turns CREB on? And off? And why does it matter? Cell Mol Life Sci. 2020 Oct;77(20):4049–67.

122. Sakamoto K, Karelina K, Obrietan K. CREB: a multifaceted regulator of neuronal plasticity and protection. J Neurochem. 2011 Jan;116(1):1–9.

123. Wang JQ, Mao L. The ERK Pathway: Molecular Mechanisms and Treatment of Depression. Mol Neurobiol. 2019 Sep;56(9):6197–205.

124. Wang G, Zhu Z, Xu D, Sun L. Advances in Understanding CREB Signaling-Mediated Regulation of the Pathogenesis and Progression of Epilepsy. Clin Neurol Neurosurg. 2020 Sep;196:106018.

125. Amidfar M, de Oliveira J, Kucharska E, Budni J, Kim YK. The role of CREB and BDNF in neurobiology and treatment of Alzheimer's disease. Life Sci. 2020 Sep 15;257:118020.

126. Landschulz WH, Johnson PF, McKnight SL. The DNA Binding Domain of the Rat Liver Nuclear Protein C/EBP Is Bipartite. Science. 1989 Mar 31;243(4899):1681–8.

127. Metallo S. The role of a basic amino acid cluster in target site selection and non-specific binding of bZIP peptides to DNA. Nucleic Acids Research. 1997 Aug 1;25(15):2967–72.

128. Metallo SJ, Schepartz A. Distribution of labor among bZIP segments in the control of DNA affinity and specificity. Chemistry & Biology. 1994 Nov;1(3):143–51.

129. Niu X, Renshaw-Gegg L, Miller L, Guiltinan MJ. Bipartite determinants of DNA-binding specificity of plant basic leucine zipper proteins. Plant Mol Biol. 1999 Sep;41(1):1–13.

130. Llorca CM, Potschin M, Zentgraf U. bZIPs and WRKYs: two large transcription factor families executing two different functional strategies. Front Plant Sci. 2014;5:169.

131. O'Shea EK, Rutkowski R, Kim PS. Evidence that the leucine zipper is a coiled coil. Science. 1989 Jan 27;243(4890):538–42.

132. Ellenberger TE, Brandl CJ, Struhl K, Harrison SC. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. Cell. 1992 Dec 24;71(7):1223–37.

133. Santiago-Rivera ZI, Gorenstein DG, Williams JS, Andrisani OM. Bacterial expression and characterization of the CREB bZip module: Circular dichroism and 2D ^{1} H-NMR studies. Protein Science. 1993 Sep;2(9):1461–71.

134. Weiss MA, Ellenberger T, Wobbe CR, Lee JP, Harrison SC, Struhl K. Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. Nature. 1990 Oct 11;347(6293):575–8.

135. Schumacher MA, Goodman RH, Brennan RG. 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. 2000 Nov;275(45):35242–7.

136. Reinke AW, Baek J, Ashenberg O, Keating AE. Networks of bZIP Protein-Protein Interactions Diversified Over a Billion Years of Evolution. Science. 2013 May 10;340(6133):730–4.

137. Tan Y, Rouse J, Zhang A, Cariati S, Cohen P, Comb MJ. FGF and stress regulate CREB and ATF-1 via a pathway involving p38 MAP kinase and MAPKAP kinase-2. The EMBO Journal. 1996 Sep;15(17):4629–42.

138. Johannessen M, Moens U. Multisite phosphorylation of the cAMP response element-binding protein (CREB) by a diversity of protein kinases. Front Biosci. 2007 Jan 1;12:1814–32.

139. Boija A, Mahat DB, Zare A, Holmqvist PH, Philip P, Meyers DJ, et al. CBP Regulates Recruitment and Release of Promoter-Proximal RNA Polymerase II. Mol Cell. 2017 Nov 2;68(3):491-503.e5.

140. Narita T, Ito S, Higashijima Y, Chu WK, Neumann K, Walter J, et al. Enhancers are activated by p300/CBP activity-dependent PIC assembly, RNAPII recruitment, and pause release. Molecular Cell. 2021 May;81(10):2166-2182.e6.

141. Ogryzko VV, Schiltz RL, Russanova V, Howard BH, Nakatani Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell. 1996 Nov 29;87(5):953–9.

142. Clayton AL, Hazzalin CA, Mahadevan LC. Enhanced Histone Acetylation and Transcription: A Dynamic Perspective. Molecular Cell. 2006 Aug;23(3):289–96.

143. Conkright MD, Canettieri G, Screaton R, Guzman E, Miraglia L, Hogenesch JB, et al. TORCs: transducers of regulated CREB activity. Mol Cell. 2003 Aug;12(2):413–23.

144. Luo Q, Viste K, Urday-Zaa JC, Senthil Kumar G, Tsai WW, Talai A, et al. Mechanism of CREB recognition and coactivation by the CREB-regulated transcriptional coactivator CRTC2. Proc Natl Acad Sci U S A. 2012 Dec 18;109(51):20865–70.

145. Song Y, Zhai L, Valencia Swain J, Chen Y, Wang P, Chen L, et al. Structural Insights into the CRTC2–CREB Complex Assembly on CRE. Journal of Molecular Biology. 2018 Jun;430(13):1926–39.

146. Dodson GE, Tibbetts RS. DNA Replication Stress-induced Phosphorylation of Cyclic AMP Response Element-binding Protein Mediated by ATM. Journal of Biological Chemistry. 2006 Jan;281(3):1692–7.

147. Lu Q, Hutchins AE, Doyle CM, Lundblad JR, Kwok RPS. Acetylation of cAMP-responsive Element-binding Protein (CREB) by CREB-binding Protein Enhances CREB-dependent Transcription. Journal of Biological Chemistry. 2003 May;278(18):15727–34.

148. Paz JC, Park S, Phillips N, Matsumura S, Tsai WW, Kasper L, et al. Combinatorial regulation of a signal-dependent activator by phosphorylation and acetylation. Proc Natl Acad Sci USA. 2014 Dec 2;111(48):17116–21.

149. Quinn PG. Distinct activation domains within cAMP response element-binding protein (CREB) mediate basal and cAMP-stimulated transcription. Journal of Biological Chemistry. 1993 Aug;268(23):16999–7009.

150. Gonzalez GA, Menzel P, Leonard J, Fischer WH, Montminy MR. Characterization of Motifs Which Are Critical for Activity of the Cyclic AMP-Responsive Transcription Factor CREB. Molecular and Cellular Biology. 1991 Mar 1;11(3):1306–12.

151. Mayr BM, Guzman E, Montminy M. Glutamine Rich and Basic Region/Leucine Zipper (bZIP) Domains Stabilize cAMP-response Element-binding Protein (CREB) Binding to Chromatin. Journal of Biological Chemistry. 2005 Apr;280(15):15103–10.

152. Ferreri K, Gill G, Montminy M. The cAMP-regulated transcription factor CREB interacts with a component of the TFIID complex. Proc Natl Acad Sci USA. 1994 Feb 15;91(4):1210–3.

153. Mengus G, Fadloun A, Kobi D, Thibault C, Perletti L, Michel I, et al. TAF4 inactivation in embryonic fibroblasts activates TGFβ signalling and autocrine growth. EMBO J. 2005 Aug 3;24(15):2753–67.

154. Martinez-Yamout MA, Nasir I, Shnitkind S, Ellis JP, Berlow RB, Kroon G, et al. Glutamine-rich regions of the disordered CREB transactivation domain mediate dynamic intra- and intermolecular interactions. Proc Natl Acad Sci USA. 2023 Nov 21;120(47):e2313835120.

155. Bentley EP, Scholl D, Wright PE, Deniz AA. Coupling of binding and differential subdomain folding of the intrinsically disordered transcription factor CREB. FEBS Letters. 2023 Apr;597(7):917–32.

156. Huang X, Zhang J, Lu L, Yin L, Xu M, Wang Y, et al. Cloning and expression of a novel CREB mRNA splice variant in human testis. Reproduction. 2004 Dec;128(6):775–82.

157. Cranz S, Berger C, Baici A, Jelesarov I, Bosshard HR. Monomeric and Dimeric bZIP Transcription Factor GCN4 Bind at the Same Rate to Their Target DNA Site. Biochemistry. 2004 Jan 1;43(3):718–27.

158. Metallo SJ, Schepartz A. Certain bZIP peptides bind DMA sequentially as monomers and dimerize on the DMA. Nat Struct Mol Biol. 1997 Feb;4(2):115–7.

159. Kohler J. Effects of nucleic acids and polyanions on dimer formation and DNA binding by bZIP and bHLHZip transcription factors. Bioorganic & Medicinal Chemistry. 2001 Sep;9(9):2435–43.

160. Kohler JJ, Schepartz A. Kinetic Studies of Fos·Jun·DNA Complex Formation: DNA Binding Prior to Dimerization. Biochemistry. 2001 Jan 1;40(1):130–42.

161. Wang X, Cao W, Cao A, Lai L. Thermodynamic characterization of the folding coupled DNA binding by the monomeric transcription activator GCN4 peptide. Biophys J. 2003 Mar;84(3):1867–75.

162. Wu X, Spiro C, Owen WG, McMurray CT. cAMP Response Element-binding Protein Monomers Cooperatively Assemble to Form Dimers on DNA. Journal of Biological Chemistry. 1998 Aug;273(33):20820–7.

163. Kelly C. Target search mode of a disordered transcription factor [Internet]. 2021. Available from: https://ora.ox.ac.uk/objects/uuid:31ea252d-644a-431e-b1b5-048919175731/files/d9p290945r

164. Huth JR, Bewley CA, Jackson BM, Hinnebusch AG, Clore GM, Gronenborn AM. Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. Protein Sci. 1997 Nov;6(11):2359–64.

165. Bayer P, Arndt A, Metzger S, Mahajan R, Melchior F, Jaenicke R, et al. Structure determination of the small ubiquitin-related modifier SUMO-1. J Mol Biol. 1998 Jul 10;280(2):275–86.

166. Malakhov MP, Mattern MR, Malakhova OA, Drinker M, Weeks SD, Butt TR. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. J Struct Funct Genomics. 2004;5(1–2):75–86.

167. Porterfield JZ, Zlotnick A. A simple and general method for determining the protein and nucleic acid content of viruses by UV absorbance. Virology. 2010 Nov;407(2):281–8.

168. Walker JM, editor. The proteomics protocols handbook. Totowa, NJ: Humana Press; 2005. 988 p.

169. Craig JC, Schumacher MA, Mansoor SE, Farrens DL, Brennan RG, Goodman RH. Consensus and variant cAMP-regulated enhancers have distinct CREB-binding properties. J Biol Chem. 2001 Apr 13;276(15):11719–28.

170. Newville M, Otten R, Nelson A, Stensitzki T, Ingargiola A, Allan D, et al. LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python [Internet]. Zenodo; 2023. Available from: https://zenodo.org/record/598352

171. Gill SC, Von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. Analytical Biochemistry. 1989 Nov;182(2):319–26.

172. Pollard TD. A Guide to Simple and Informative Binding Assays. Kellogg D, editor. MBoC. 2010 Dec;21(23):4061–7.

173. Crabtree MD, Shammas SL. Chapter Fourteen - Stopped-Flow Kinetic Techniques for Studying Binding Reactions of Intrinsically Disordered Proteins. In: Rhoades EBTM in E, editor. Intrinsically Disordered Proteins. Academic Press; 2018. p. 423–57.

174. Jarmoskaite I, AlSadhan I, Vaidyanathan PP, Herschlag D. How to measure and evaluate binding affinities. Elife. 2020 Aug 6;9:e57264.

175. Dandliker WB, Hsu ML, Levin J, Ramanath Rao B. Equilibrium and kinetic inhibition assays based upon fluorescence polarization. In: Methods in Enzymology. Elsevier; 1981. p. 3–28.

176. Sóley Kristjánsdóttir S, Moody AE, Norton JR. Approach to equilibrium after dilution of a monomer/dimer mixture. Measurement of the rate constant for dissociation of trityl dimer by stopped-flow methods. Int J Chem Kinet. 1992 Oct;24(10):895–902.

177. Milla ME, Sauer RT. P22 Arc repressor: folding kinetics of a single-domain, dimeric protein. Biochemistry. 1994 Feb 8;33(5):1125–33.

178. Schuck P. Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. Biophysical Journal. 2000 Mar;78(3):1606–19.

179. Philo JS. SEDNTERP: a calculation and database utility to aid interpretation of analytical ultracentrifugation and light scattering data. Eur Biophys J. 2023 Jul;52(4–5):233–66.

180. Micsonai A, Bulyáki É, Kardos J. BeStSel: From Secondary Structure Analysis to Protein Fold Prediction by Circular Dichroism Spectroscopy. In: Chen YW, Yiu CPB, editors. Structural Genomics. New York, NY: Springer US; 2021. p. 175–89. (Methods in Molecular Biology; vol. 2199).

181. Jain P, Shah K, Sharma N, Kaur R, Singh J, Vinson C, et al. A-ZIP53, a dominant negative reveals the molecular mechanism of heterodimerization between bZIP53, bZIP10 and bZIP25 involved in Arabidopsis seed maturation. Sci Rep. 2017 Oct 30;7(1):14343.

182. Greenfield NJ. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. Nat Protoc. 2006;1(6):2527–35.

183. Cho NH, Cheveralls KC, Brunner AD, Kim K, Michaelis AC, Raghavan P, et al. OpenCell: Endogenous tagging for the cartography of human cellular organization. Science. 2022 Mar 11;375(6585):eabi6983.

184. Zhang X, Odom DT, Koo SH, Conkright MD, Canettieri G, Best J, et al. Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues. Proc Natl Acad Sci USA. 2005 Mar 22;102(12):4459–64.

185. Impey S, McCorkle SR, Cha-Molstad H, Dwyer JM, Yochum GS, Boss JM, et al. Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. Cell. 2004 Dec 29;119(7):1041–54.

186. Lesiak A, Pelz C, Ando H, Zhu M, Davare M, Lambert TJ, et al. A Genome-Wide Screen of CREB Occupancy Identifies the RhoA Inhibitors Par6C and Rnd3 as Regulators of BDNF-Induced Synaptogenesis. Hetman M, editor. PLoS ONE. 2013 Jun 6;8(6):e64658.

187. Tanis KQ, Duman RS, Newton SS. CREB Binding and Activity in Brain: Regional Specificity and Induction by Electroconvulsive Seizure. Biological Psychiatry. 2008 Apr;63(7):710–20.

188. Wu Y, Pegoraro AF, Weitz DA, Janmey P, Sun SX. The correlation between cell and nucleus size is explained by an eukaryotic cell growth model. Igoshin OA, editor. PLoS Comput Biol. 2022 Feb 18;18(2):e1009400.

189. Piovesan A, Pelleri MC, Antonaros F, Strippoli P, Caracausi M, Vitale L. On the length, weight and GC content of the human genome. BMC Res Notes. 2019 Dec;12(1):106.

190. Chakrabarti P, Chakravarty D. Intrinsically disordered proteins/regions and insight into their biomolecular interactions. Biophys Chem. 2022 Apr;283:106769.

191. Liu CW, Corboy MJ, DeMartino GN, Thomas PJ. Endoproteolytic activity of the proteasome. Science. 2003 Jan 17;299(5605):408–11.

192. Chung NL. Investigating potential allostery in the transcription factor CREB [Internet]. 2020. Available from: https://ora.ox.ac.uk/objects/uuid:c5923fdf-00cf-4f6b-a64c-76cfcf24d965/files/dz029p486b

193. Lopez DI, Mick JE, Nyborg JK. Purification of CREB to apparent homogeneity: Removal of truncation products and contaminating nucleic acid. Protein Expression and Purification. 2007 Oct;55(2):406–18.

194. Goren I, Tavor E, Goldblum A, Honigman A. Two cysteine residues in the DNA-binding domain of CREB control binding to CRE and CREB-mediated gene expression. Journal of Molecular Biology. 2001 Nov;313(4):695–709.

195. Fabrini R, De Luca A, Stella L, Mei G, Orioni B, Ciccone S, et al. Monomer−Dimer Equilibrium in Glutathione Transferases: A Critical Re-Examination. Biochemistry. 2009 Nov 3;48(43):10473–82.

196. Alexander P, Fahnestock S, Lee T, Orban J, Bryan P. Thermodynamic analysis of the folding of the streptococcal protein G IgG-binding domains B1 and B2: why small proteins tend to have high denaturation temperatures. Biochemistry. 1992 Apr;31(14):3597–603.

197. Frei MS, Tarnawski M, Roberti MJ, Koch B, Hiblot J, Johnsson K. Engineered HaloTag variants for fluorescence lifetime multiplexing. Nature Methods. 2022 Jan 1;19(1):65–70.

198. Burgess RR. Chapter 20 Protein Precipitation Techniques. In: Methods in Enzymology. Elsevier; 2009. p. 331–42.

199. Austin BP, Tözsér J, Bagossi P, Tropea JE, Waugh DS. The substrate specificity of Metarhizium anisopliae and Bos taurus carboxypeptidases A: Insights into their use as tools for the removal of affinity tags. Protein Expression and Purification. 2011 May;77(1):53–61.

200. Lu TS, Yiao SY, Lim K, Jensen RV, Hsiao LL. Interpretation of biological and mechanical variations between the Lowry versus Bradford method for protein quantification. N Am J Med Sci. 2010 Jul;2(7):325–8.

201. Studier FW. Protein production by auto-induction in high-density shaking cultures. Protein Expression and Purification. 2005 May;41(1):207–34.

202. Drobecq H, Boll E, Sénéchal M, Desmet R, Saliou JM, Lacapère JJ, et al. A Central Cysteine Residue Is Essential for the Thermal Stability and Function of SUMO-1 Protein and SUMO-1 Peptide–Protein Conjugates. Bioconjugate Chem. 2016 Jun 15;27(6):1540–6.

203. Pucci F, Kwasigroch JM, Rooman M. SCooP: an accurate and fast predictor of protein stability curves as a function of temperature. Bioinformatics. 2017 Nov 1;33(21):3415–22.

204. Steven A, Friedrich M, Jank P, Heimer N, Budczies J, Denkert C, et al. What turns CREB on? And off? And why does it matter? Cell Mol Life Sci. 2020 Oct;77(20):4049–67.

205. Santiago-Rivera ZI, Gorenstein DG, Williams JS, Andrisani OM. Bacterial expression and characterization of the CREB bZip module: Circular dichroism and 2D ^{1} H-NMR studies. Protein Sci. 1993 Sep;2(9):1461–71.

206. Sugo N, Morimatsu M, Arai Y, Kousoku Y, Ohkuni A, Nomura T, et al. Single-Molecule Imaging Reveals Dynamics of CREB Transcription Factor Bound to Its Target Sequence. Sci Rep. 2015 Jun 3;5(1):10662.

207. Garner E, Romero P, Dunker AK, Brown C, Obradovic Z. Predicting Binding Regions within Disordered Proteins. Genome Inform Ser Workshop Genome Inform. 1999;10:41–50.

208. Erdős G, Pajkos M, Dosztányi Z. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. Nucleic Acids Research. 2021 Jul 2;49(W1):W297–303.

209. Hu G, Katuwawala A, Wang K, Wu Z, Ghadermarzi S, Gao J, et al. flDPnn: Accurate intrinsic disorder prediction with putative propensities of disorder functions. Nat Commun. 2021 Jul 21;12(1):4438.

210. Crabtree MD, Shammas SL. Stopped-Flow Kinetic Techniques for Studying Binding Reactions of Intrinsically Disordered Proteins. In: Methods in Enzymology. Elsevier; 2018. p. 423–57.

211. Damian L. Isothermal Titration Calorimetry for Studying Protein–Ligand Interactions. In: Williams MA, Daviter T, editors. Protein-Ligand Interactions. Totowa, NJ: Humana Press; 2013. p. 103–18. (Methods in Molecular Biology; vol. 1008).

212. Wang X, Bigman LS, Greenblatt HM, Yu B, Levy Y, Iwahara J. Negatively charged, intrinsically disordered regions can accelerate target search by DNA-binding proteins. Nucleic Acids Research. 2023 Jun 9;51(10):4701–12.

213. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, editor. The Proteomics Protocols Handbook. Totowa, NJ: Humana Press; 2005. p. 571–607.

214. Smoluchowski MV. Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen. Zeitschrift für Physikalische Chemie. 1918 Nov 1;92U(1):129–68.

215. Schlosshauer M, Baker D. Realistic protein–protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness. Protein Science. 2004 Jun;13(6):1660–9.

216. Gemayel R, Chavali S, Pougach K, Legendre M, Zhu B, Boeynaems S, et al. Variable Glutamine-Rich Repeats Modulate Transcription Factor Activity. Mol Cell. 2015 Aug 20;59(4):615–27.

217. Escher D, Bodmer-Glavas M, Barberis A, Schaffner W. Conservation of glutamine-rich transactivation function between yeast and humans. Mol Cell Biol. 2000 Apr;20(8):2774–82.

218. Saluja D, Vassallo MF, Tanese N. Distinct subdomains of human TAFII130 are required for interactions with glutamine-rich transcriptional activators. Mol Cell Biol. 1998 Oct;18(10):5734–43.

219. Fiumara F, Fioriti L, Kandel ER, Hendrickson WA. Essential Role of Coiled Coils for Aggregation and Activity of Q/N-Rich Prions and PolyQ Proteins. Cell. 2010 Dec;143(7):1121–35.

220. Urbanek A, Popovic M, Morató A, Estaña A, Elena-Real CA, Mier P, et al. Flanking Regions Determine the Structure of the Poly-Glutamine in Huntingtin through Mechanisms Common among Glutamine-Rich Human Proteins. Structure. 2020 Jul;28(7):733-746.e5.

221. Shattuck JE, Waechter AC, Ross ED. The effects of glutamine/asparagine content on aggregation and heterologous prion induction by yeast prion-like domains. Prion. 2017 Jul 4;11(4):249–64.

222. Schaffar G, Breuer P, Boteva R, Behrends C, Tzvetkov N, Strippel N, et al. Cellular Toxicity of Polyglutamine Expansion Proteins. Molecular Cell. 2004 Jul;15(1):95–105.

223. Jung J, Bonini N. CREB-Binding Protein Modulates Repeat Instability in a *Drosophila* Model for PolyQ Disease. Science. 2007 Mar 30;315(5820):1857–9.

224. McCampbell A, Taylor JP, Taye AA, Robitschek J, Li M, Walcott J, et al. CREB-binding protein sequestration by expanded polyglutamine. Hum Mol Genet. 2000 Sep 1;9(14):2197–202.

225. Murthy AC, Dignon GL, Kan Y, Zerze GH, Parekh SH, Mittal J, et al. Molecular interactions underlying liquid-liquid phase separation of the FUS low-complexity domain. Nat Struct Mol Biol. 2019 Jul;26(7):637–48.

226. Ma L, Gao Z, Wu J, Zhong B, Xie Y, Huang W, et al. Co-condensation between transcription factor and coactivator p300 modulates transcriptional bursting kinetics. Molecular Cell. 2021 Apr;81(8):1682-1697.e7.

227. Li L, Liu H, Dong P, Li D, Legant WR, Grimm JB, et al. Real-time imaging of Huntingtin aggregates diverting target search and gene transcription. eLife. 2016 Aug 3;5:e17056.

228. Shrinivas K, Sabari BR, Coffey EL, Klein IA, Boija A, Zamudio AV, et al. Enhancer Features that Drive Formation of Transcriptional Condensates. Molecular Cell. 2019 Aug;75(3):549-561.e7.

229. Iijima-Ando K, Wu P, Drier EA, Iijima K, Yin JCP. cAMP-response element-binding protein and heat-shock protein 70 additively suppress polyglutamine-mediated toxicity in *Drosophila*. Proc Natl Acad Sci USA. 2005 Jul 19;102(29):10261–6.

230. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug 26;596(7873):583–9.

231. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Research. 2022 Jan 7;50(D1):D439–44.

232. Hammarström P, Carlsson U. Is the unfolded state the Rosetta Stone of the protein folding problem? Biochem Biophys Res Commun. 2000 Sep 24;276(2):393–8.

233. Holehouse AS, Das RK, Ahad JN, Richardson MOG, Pappu RV. CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. Biophysical Journal. 2017 Jan;112(1):16–21.

234. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Research. 2022 Jul 5;50(W1):W276–9.

235. Quaglia F, Mészáros B, Salladini E, Hatos A, Pancsa R, Chemes LB, et al. DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation. Nucleic Acids Res. 2022 Jan 7;50(D1):D480–7.

236. The UniProt Consortium, Bateman A, Martin MJ, Orchard S, Magrane M, Ahmad S, et al. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Research. 2023 Jan 6;51(D1):D523–31.

237. Lee B, Jaberi-Lashkari N, Calo E. A unified view of low complexity regions (LCRs) across species. Elife. 2022 Sep 13;11:e77058.

238. Chen Z, Hou C, Wang L, Yu C, Chen T, Shen B, et al. Screening membraneless organelle participants with machine-learning models that integrate multimodal features. Proc Natl Acad Sci USA. 2022 Jun 14;119(24):e2115369119.

239. Chu X, Sun T, Li Q, Xu Y, Zhang Z, Lai L, et al. Prediction of liquid–liquid phase separating proteins using machine learning. BMC Bioinformatics. 2022 Feb 15;23(1):72.

240. Alderson TR, Pritišanac I, Kolarić Đ, Moses AM, Forman-Kay JD. Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2. Proc Natl Acad Sci USA. 2023 Oct 31;120(44):e2304302120.

241. Simm S, Einloft J, Mirus O, Schleiff E. 50 years of amino acid hydrophobicity scales: revisiting the capacity for peptide classification. Biol Res. 2016 Dec;49(1):31.

242. Cornette JL, Cease KB, Margalit H, Spouge JL, Berzofsky JA, DeLisi C. Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. Journal of Molecular Biology. 1987 Jun;195(3):659–85.

243. Jackson BM, Drysdale CM, Natarajan K, Hinnebusch AG. Identification of Seven Hydrophobic Clusters in GCN4 Making Redundant Contributions to Transcriptional Activation. Molecular and Cellular Biology. 1996 Oct 1;16(10):5557–71.

244. Kastritis PL, Moal IH, Hwang H, Weng Z, Bates PA, Bonvin AMJJ, et al. A structure-based benchmark for protein-protein binding affinity. Protein Sci. 2011 Mar;20(3):482–91.

245. Ishikawa-Ankerhold HC, Ankerhold R, Drummen GPC. Advanced Fluorescence Microscopy Techniques—FRAP, FLIP, FLAP, FRET and FLIM. Molecules. 2012 Apr 2;17(4):4047–132.

246. Hollenbeck JJ, Oakley MG. GCN4 Binds with High Affinity to DNA Sequences Containing a Single Consensus Half-Site. Biochemistry. 2000 May 1;39(21):6380–9.

247. Wendt H, Baici A, Bosshard HR. Mechanism of assembly of a leucine zipper domain. J Am Chem Soc. 1994 Jul;116(15):6973–4.

248. Chan YHM, Marshall WF. Organelle size scaling of the budding yeast vacuole is tuned by membrane trafficking rates. Biophys J. 2014 May 6;106(9):1986–96.

249. Ho B, Baryshnikova A, Brown GW. Unification of Protein Abundance Datasets Yields a Quantitative Saccharomyces cerevisiae Proteome. Cell Syst. 2018 Feb 28;6(2):192-205.e3.

250. Hwang C, Sinskey AJ, Lodish HF. Oxidized redox state of glutathione in the endoplasmic reticulum. Science. 1992 Sep 11;257(5076):1496–502.

251. Rivier J, Kaiser R, Galyean R. Solid-phase synthesis of somatostatin and glucagon-selective analogs in gram quantities. Biopolymers. 1978 Aug;17(8):1927–38.

252. Smith RC, Reed VD, Hill WE. Oxidation Of Thiols By Copper(II). Phosphorus, Sulfur, and Silicon and the Related Elements. 1994 May 1;90(1–4):147–54.

253. Wallace TJ, Mahon JJ. Reactions of Thiols with Sulfoxides. II. Kinetics and Mechanistic Implications. J Am Chem Soc. 1964 Oct;86(19):4099–103.

254. Iguchi-Ariga SM, Schaffner W. CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. Genes Dev. 1989 May;3(5):612–9.

255. Riggs AD, Bourgeois S, Cohn M. The lac repressor-operator interaction. 3. Kinetic studies. J Mol Biol. 1970 Nov 14;53(3):401–17.

256. Berger C, Piubelli L, Haditsch U, Rudolf Bosshard H. Diffusion-controlled DNA recognition by an unfolded, monomeric bZIP transcription factor. FEBS Letters. 1998 Mar 20;425(1):14–8.

257. Okahata Y, Niikura K, Sugiura Y, Sawada M, Morii T. Kinetic Studies of Sequence-Specific Binding of GCN4-bZIP Peptides to DNA Strands Immobilized on a 27-MHz Quartz-Crystal Microbalance. Biochemistry. 1998 Apr 21;37(16):5666–72.

258. Dwarki VJ, Montminy M, Verma IM. Both the basic region and the 'leucine zipper' domain of the cyclic AMP response element binding (CREB) protein are essential for transcriptional activation. The EMBO Journal. 1990 Jan;9(1):225–32.

259. Lin YS, Green MR. Interaction of a common cellular transcription factor, ATF, with regulatory elements in both E1a- and cyclic AMP-inducible promoters. Proc Natl Acad Sci U S A. 1988 May;85(10):3396–400.

260. Basu S, Needham LM, Lando D, Taylor EJR, Wohlfahrt KJ, Shah D, et al. FRET-enhanced photostability allows improved single-molecule tracking of proteins and protein complexes in live mammalian cells. Nat Commun. 2018 Jun 28;9(1):2520.

261. Zhang Y, Brown K, Yu Y, Ibrahim Z, Zandian M, Xuan H, et al. Nuclear condensates of p300 formed though the structured catalytic core can act as a storage pool of p300 with reduced HAT activity. Nat Commun. 2021 Jul 29;12(1):4618.

262. Louphrasitthiphol P, Siddaway R, Loffreda A, Pogenberg V, Friedrichsen H, Schepsky A, et al. Tuning Transcription Factor Availability through Acetylation-Mediated Genomic Redistribution. Mol Cell. 2020 Aug 6;79(3):472-487.e10.

263. Gill ML, Byrd RA, Palmer, Iii AG. Dynamics of GCN4 facilitate DNA interaction: a model-free analysis of an intrinsically disordered region. Phys Chem Chem Phys. 2016;18(8):5839–49.

Appendix

Appendix 1: derivation of the homodimerization-corrected binding model for anisotropy titrations

Assuming the following reaction scheme:

$$M + M \rightleftharpoons Dim + \mathrm{DNA} \rightleftharpoons Dim{:}\,\mathrm{DNA}$$

Where:

M is a CREB monomer

Dim is a CREB dimer

*Dim:*DNA is a DNA bound by CREB dimer.

There are two dissociation constants involved in the above scheme:

$$\frac{[M]^2}{[Dim]_{free}} = L \ (1)$$

$$\frac{D_{free}[Dim]_{free}}{[Dim{:}D]} = K \ (2)$$

Where D is the concentration of DNA. Since $D - [Dim{:}D] = D_{free}$ and $[Dim]_{free} = [Dim]_{total} - [Dim{:}D]$, (2) can be rearranged to give:

$$K[Dim{:}D] = (D - [Dim{:}D])([Dim]_{total} - [Dim{:}D])$$

$$0 = [Dim{:}D]^2 - [Dim{:}D](D + K + [Dim]_{total}) + D[Dim]_{total}$$

The roots of the above quadratic equation are extracted to give:

$$[Dim{:}D] = \frac{(D+K+[Dim]_{total}) - \sqrt{(D+K+[Dim]_{total})^2 - 4[Dim]_{total}}}{2} \ (3)$$

Assuming there is an anisotropy signal change between upon DNA binding by CREB, with free DNA having anisotropy value R_f and bound DNA having anisotropy value R_b:

$$R_b - R_f = \Delta R$$

Assuming the change is proportional to bound species, anisotropy value at any concentration of labelled CREB (A) will be equal to:

$$A = R_f + \Delta R \frac{[Dim:D]}{D} \quad (4)$$

Plugging (3) into (4) yields:

$$A = R_f + \Delta R \frac{(D + K + [Dim]_{total}) - \sqrt{(D + K + [Dim]_{total})^2 - 4[Dim]_{total}}}{2D} \quad (5)$$

However, $[Dim]_{total}$ is still an unknown, but it can be approximated using (1). Knowing that the total monomeric concentration of CREB is equal to:

$$x = [M] + 2[Dim]_{total} \quad (6)$$

(1) can be rearranged to give:

$$(x - 2[Dim]_{total})^2 = L([Dim]_{total} - [Dim:D])$$

$$(L[Dim:D] + x^2) - [Dim]_{total}(4x + L) + 4[Dim]_{total}^2 = 0$$

This again a quadratic equation. Assuming we need to solve for $[Dim]_{total}$:

$$[Dim]_{total} = \frac{(4x+L) - \sqrt{(4x+L)^2 - 16(L[Dim:D] + x^2)}}{8} \quad (7)$$

However, this still leaves $[Dim:D]$ as an unsolvable variable. It is possible to use the anisotropy signal to obtain it. Indeed, (4) can be rearranged to yield:

$$\frac{(A-R_f)D}{\Delta R} = [Dim{:}\, D] \quad (8)$$

Plugging (8) into (7) yields:

$$[Dim]_{total}\frac{(4x+L) - \sqrt{(4x+L)^2 - 16\left(L\frac{(A-R_f)D}{\Delta R} + x^2\right)}}{8}$$

Which allows for (5) to be fitted against the known total CREB (*x*) and DNA (*D*) concentrations, but also against the anisotropy value (*A*) itself.

Appendix 2: Derivation of the CREB homodimerization model to fit the labelled CREB dilution anisotropy curve

Assuming there is an anisotropy signal change between monomeric and dimeric CREB, with free CREB having anisotropy value R_f and bound, dimeric CREB having anisotropy value R_b:

$$R_b - R_f = \Delta R$$

Assuming the change is proportional to bound dimeric species, anisotropy value at any concentration of labelled CREB will be equal to:

$$A = R_f + \Delta R \frac{[\mathrm{D}]}{x} \quad (1)$$

Where:

[D] is the dimeric CREB concentration.

x is the total CREB concentration.

The homodimerization equilibrium constant *K* is known:

$$\frac{[\mathrm{M}]^2}{[\mathrm{D}]} = K \quad (2)$$

The equation can be rearranged to include x and make a quadratic equation where [D] is the unknown:

$x = [\mathrm{M}] + 2[\mathrm{D}]$ (dimeric concentration is halved since it has two CREB monomers)

$$[\mathrm{M}] = \mathrm{x} - 2[\mathrm{D}] \quad (3)$$

Substituting (3) into (2):

$$(\mathrm{x} - 2[\mathrm{D}])^2 = [\mathrm{D}]K$$

$$4[\mathrm{D}]^2 - [\mathrm{D}](4x + K) + x^2 = 0$$

As a quadratic equation, this is solved for [D] using:

$$[\mathrm{D}] = \frac{(4x + K) - \sqrt{(4x + K)^2 - 4 \times 4x^2}}{2 \times 4}$$

Developing the root term gives:

$$[\mathrm{D}] = \frac{4x + K - \sqrt{K^2 + 8xK}}{8}$$

$$[\mathrm{D}] = 0.5x + K\frac{1 - \sqrt{1 + \frac{8x}{K}}}{8}$$

Substitution of [D] into (1) yields:

$$A = R_f + \Delta R\left(0.5 + K\frac{1 - \sqrt{1 + \frac{8x}{K}}}{8x}\right)$$

Appendix 3: Derivation of the reversible refolding model to fit bZIP refolding kinetic traces

$$\mathrm{M} + \mathrm{M} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{D}$$

Assuming the following homodimerization scheme:

Where k_1 and k_2 are the association and dissociation rate constants for the homodimerization scheme. The rate of change in monomer concentration is given by:

$$\frac{d[\mathrm{M}]}{dt} = -k_1[\mathrm{M}]^2 + 2k_2[\mathrm{D}]$$

Where [M] and [D] are the monomer and dimer concentrations respectively. Identifying the dissociation constant as

$$K_\mathrm{D} = 2\frac{k_2}{k_1} = \frac{[\mathrm{M}]^2}{[\mathrm{D}]}$$

And requiring mass balance:

$$[\mathrm{M}] + 2[\mathrm{D}] = P$$

Where P is the total protein concentration leads to:

$$-k_1 dt = \frac{d[\mathrm{M}]}{[\mathrm{M}]^2 + \frac{K_\mathrm{D}}{2}[\mathrm{M}] - \frac{K_\mathrm{D}}{2}P}$$

Finding the roots of the quadratic expression in the denominator allows this to be rewritten as:

$$-k_1 dt = \frac{d[\mathrm{M}]}{\left([\mathrm{M}] + \frac{K_\mathrm{D} + 2z}{4}\right)\left([\mathrm{M}] + \frac{K_\mathrm{D} - 2z}{4}\right)}$$

Where:

$$z = \sqrt{\frac{K_D^2}{4} + 2K_D P}$$

Which may then be integrated to obtain an expression for the concentration of monomeric protein with time, to provide equation X:

$$[\mathrm{M}] = \frac{\frac{K_D - 2z}{4} - \left(\frac{K_D + 2z}{4}\right) N e^{-ik_1 zt}}{N e^{-ik_1 zt} - 1}$$

where *N* is a constant of integration that may be written in terms of the initial monomer concentration ($[\mathrm{A}]_0$) by evaluation of Equation X at time zero to give equation Y:

$$N = \frac{\frac{K_D - 2z}{4} + [\mathrm{A}]_0}{\frac{K_D + 2z}{4} + [\mathrm{A}]_0}$$

Kinetic refolding traces were fit to:

$$F(t) = F_0 + \Delta F[\mathrm{M}]$$

Where [M] is given by Equation X, N is given by Equation Y, $z = \sqrt{\frac{K_D^2}{4} + 2K_D P}$, and due to the conditions of our refolding experiment starting in 8 M urea, $[\mathrm{A}]_0 = P$.