Biophysical and structural characterisation of Protein-Protein interactions, HIF-1α/p300 and eIF4E/eIF4G to inform inhibitor rational design.

Hannah Frances Kyle

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds School of Chemistry

Astbury Centre for Structural and Molecular Biology

1.May.2015

Intellectual Property and Publication Statements

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 3 is based on the work contributed to jointly authored papers:

Small-Molecule Proteomimetic Inhibitors of the HIF-1α–p300 Protein–Protein Interaction.
George M. Burslem, Hannah F. Kyle, Dr. Alexander L. Breeze, Dr. Thomas A. Edwards, Prof. Adam Nelson, Dr. Stuart L. Warriner and Prof. Andrew J. Wilson

Exploration of the HIF-1α/p300 interface using peptide and Adhiron phage display technologies.
Hannah. F. Kyle, Kate. F. Wickson, Jonathon. Stott, George. M. Burslem, Alexander. L. Breeze, Christian. Tiede, Darren. C. Tomlinson, Stuart. L. Warriner, Adam. Nelson, Andrew. J. Wilson and Thomas. A. Edwards.

Work performed by Hannah F Kyle –

- Cloning, expression and purification of all proteins.
- Biophysical assays (FA, ITC and SPR) of direct binding and of inhibitors
- Phage display experiments (peptide and Adhrion)
- HSQC experiments
- Crystallisation

Work performed by George M Burslem –

- Peptide synthesis (where not purchased)
- Compound synthesis (where not purchased)
- Modelling and docking

Work performed by Jonathan Stott

- Computational analysis of peptide phage display NGS

Work performed by Christian Tiede

- Assistance with Adhiron phage display experiment

Kate Wickson, Darren Tomlinson, Alex Breeze, Stuart Warriner, Adam Nelson, Andy Wilson and Thomas Edwards supervised the project and assisted with manuscript preparation. Thomas Edwards also assisted with structural determination from crystallisation experiments.

Acknowledgments

Firstly I would like to thank George Burslem, for every peptide, every mass spec and generally all things Chemistry but mainly for having the same unwavering dedication for OUR project as I do. I hope when you look back on your thesis you feel, every late night in the Jif centre, every compound not in solution and every non-binding curve was worth it and you can look back with pride in what we have achieved in this 3.5 years. We didn't make a bad team!

I have had the honour of working under 4 great supervisors at Leeds.

- Thomas Edwards, Ed, despite a "tendency towards the bellicose" we rubbed along quite nicely. For every salt crystal and every spec of dust I though (prematurely) was a crystal I am truly sorry. And for every late night looking at said salt crystals I am truly grateful. I would like to take this opportunity to state on record that I acknowledge that you are ~~always~~ generally correct. It has been exceptional.
- Andy Wilson, I can only hope to have as much passion for the science I will do throughout my career as you do. If I do then I know I can't be doing half bad.
- Adam Nelson for keeping complete order, believe me the meetings you weren't there – utter chaos!
- Stuart Warriner, you may have one the sharpest chemistry (and maths and physics and pretty much everything) brains I have come across but really I would just like to thank-you for your starring role in Emmerdale, you were a star.

Thank-you must also go to AstraZeneca and the EPSRC for funding the project and, especially Kate and Alex for giving us the freedom to follow the project as we wanted with lots of help and advise for us to do so.

And finally to my friends and family thank-you for being a constant reminder that there is a world outside of the lab and that precipitated protein is not and never will be the end of the world.

Abstract

Protein-protein interactions (PPIs) are an important class of therapeutic target, however due to their large interaction interface they are considered difficult to inhibit. The two PPIs of interest in this thesis are the HIF-1α (hypoxia inducible factor-1α)/p300 and eIF4E (eukaryotic initiation factor 4E)/eIF4G interactions, both of which have been shown to be involved in many different cancers and are hypothesised to be good targets for targeted therapy. A rational design approach is favoured for PPIs, however for this to be possible a detailed understanding of the binding interface is required.

Biophysical assessment of the HIF-1α CTAD/p300 CH1 interface has revealed a key binding "hot-spot" of p300 where the helix three region of HIF-1α binds to p300, this area has subsequently been targeted using oligoamide α-helix mimetics to disrupt the interaction. This binding site was found by two approaches, used to probe the HIF-1α binding surface on p300, first, by analysis of the binding of shorter HIF-1α peptide fragments; and second, by phage display experiments. The HIF-1α fragment study demonstrated that HIF-1α helix 3 region binds to p300 with a higher affinity than any of the short (<20 amino acids) peptide regions of HIF-1α and a peptide containing helices 2 and 3 binds with a higher affinity than a peptide containing helices 1 and 2, this importance of the helix 3 binding site was validated using mutagenesis. The phage display experiment found 12mer peptides that bind to p300 with a higher affinity than any short peptide region of HIF-1α. Structural techniques and mutagenesis were the used to verify that this binding site was similar to that of HIF-1α helix 3. The rationally designed mimetics of HIF-1α helix 3 were able to disrupt the interaction with low micromolar affinity. A second phage display experiment found Adhirons which bound with low nanomolar affinity and were able to disrupt the interaction with low micromolar affinity, docking of the crystal structure of this Adhiron indicated that it may be acting at a different site to the HIF-1α helix 3 region.

This is a proof of principle that a detailed understanding of an interaction, using both biophysical and structural techniques, can directly lead to the development of inhibitors of challenging and therapeutically relevant PPIs.

Table of contents

ACKNOWLEDGMENTS IV
ABSTRACT V
LIST OF FIGURES IX
LIST OF TABLES XII
ABBREVIATIONS XIII
AMINO ACID CODES XVI

CHAPTER 1 : INTRODUCTION 1
1.1 PROTEIN-PROTEIN INTERACTIONS GENERAL OBSERVATIONS 1
1.1.1 Hot spots 2
1.1.2 Previous misconceptions with targeting PPIs 3
1.2 DRUG DESIGN APPROACHES 8
1.2.1 Alternative to HTS 8
1.2.2 Improvements in structure determination methodology 8
1.2.3 Rational drug design strategies 13
1.2.3.1 Peptide mimetics 13
1.2.3.2 Non-peptidic scaffold mimics 14
1.3 CHARACTERISATION OF BINDING INTERFACE 17
1.3.1 Ligand discovery by phage display 17
1.3.2 Biophysical analysis 20
1.4 THE TARGETS OF INTEREST 27
1.4.1 Targeted therapy 27
1.4.2 HIF-1α/p300 30
1.4.3 eIF4E/eIF4G 32

CHAPTER 2 : MATERIALS AND METHODS 34
2.1 MATERIALS 34
2.1.1 Bacterial Strains and Vectors 34
2.1.3 Media 35
2.1.4 General reagents 36
2.2 METHODS 38
2.2.1 Recombinant DNA techniques 38
2.2.1.1 Touchdown PCR 38
2.2.2 Agarose Gel Electrophoresis 39
2.2.3 Restriction endonuclease digests 39
2.2.4 Ligations 40
2.2.5 Competent DH5α E. coli cell preparation 40
2.2.6 Transformations 40
2.2.7 Colony PCR 41
2.2.8 p300 mutant cloning 41
2.2.9 Overnight cultures 42
2.2.9 Plasmid DNA extraction and glycerol stock preparation 42
2.2.10 Sanger DNA sequencing 42
2.3 Protein expression and purification 43
2.3.1 General techniques 43
2.3.1.1 Protein Expression Trials 43
2.3.1.2 OD_{600} measurement 43
2.3.1.3 Nickel affinity chromatography 43
2.3.1.4 Glutathione affinity chromatography 44
2.3.1.5 Size exclusion chromatography 44
2.3.1.6 Protein concentration 45
2.3.1.7 Protease expression 45
2.3.2 Specific Protein expression 46
2.3.2.1 p300 and HIF-1α/p300 fusion construct protein expression 46
2.3.2.2 ^{15}N-p300 protein expression 46
2.3.2.3 Biotinylation of p300 47

2.3.2.4 eIF4E protein expression 47
2.3.2.5 Adhirons 49
2.3.3 Protein characterization 49
2.3.3.1 SDS PAGE gel electrophoresis 49
2.3.3.2 Mass spectrometry 50
2.3.3.3 Circular Dichroism (CD) Spectroscopy 50
2.3.4 Biophysical Assays 51
2.3.4.1 Fluorescence Anisotropy 51
2.3.4.1.1 Direct binding 51
2.3.4.2 Competition 51
2.3.4.2 Isothermal Titration Calorimetry 52
2.3.4.3 SPR 52
2.3.4.4 BLitz 53
2.4 Phage Display Experiments 53
2.4.1 Peptide phage display 53
2.4.1.1 Enrichment ELISA 53
2.4.1.2 Peptide Phage Display Sequencing 54
2.4.2 Adhiron phage display 54
2.5 Structural analysis 55
2.5.1 Crystallography 55
2.5.1.1 General techniques 55
2.5.1.1.1 Factorials 55
2.5.1.1.2 Optimisation 56
2.5.1.1.3 Cryo cooling 56
Data collection 56
2.5.1.2 Apo eIF4E crystallography 56
2.5.1.3 Adhiron crystallography 56
2.5.2 NMR 57
2.5.3 SAXS 57
2.5.4 Docking 57

CHAPTER 3 : HIF-1A/P300 58
3.1 INTRODUCTION TO HIF-1A/P300 58
3.1.1 Hypoxic response 58
3.1.2 Inhibitors of the hypoxic response via HIF-1 62
2.1.2.1 HIF-1α mRNA expression 62
2.1.2.2 HIF-1α protein translation 63
2.1.2.3 HIF-1α degradation pathway 63
2.1.2.4 HIF-1 binding to DNA 64
2.1.2.5 HIF-1α transcriptional activity 65
3.1.3 Structural information of HIF-1α/p300 66
3.2 HIF-1A/P300 AIMS 76
3.3 HIF-1A/P300 CLONING AND EXPRESSION 77
3.3.1 p300 CH1 cloning 77
3.3.2 Expression Trials 79
3.3.3 p300 labeling 87
3.3.4 HIF-1α peptide 91
3.4 HIF-1A/P300 CHARCTERISATION 91
3.4.1 Biophysical assessment of HIF-1α$_{786-826}$/p300 interaction 92
3.4.2 Exploration of the binding interface using HIF-1α peptide fragments 98
3.4.2.1 Small molecule HIF-1α/p300 inhibitors 105
3.4.3 Exploration of the binding interface by phage display 109
3.4.3.1 Peptide phage display 109
3.4.3.1.1 Peptide phage display clone selection and sequencing 111
3.4.3.1.2 Phage display derived peptide analysis 115
3.4.3.2 Adhiron phage display experiment 118
3.4.3.3 Summary of phage display. 128
3.4.4 Structural characterization by crystallography 129
3.4.4.1 SAXS analysis 135
3.4.4.1.1 Sample Quality 135
3.4.4.1.2 Size and shape comparison of p300, co-purified p300/HIF-1α and HIF-1α/p300 fusion construct 139
3.4.4.2 HIF-1α/p300 fusion construct crystallisation 143

3.5 HIF-1A/P300 CHAPTER SUMMARY AND FURTHER WORK 146

CHAPTER 4 : EIF4E/EIF4G CHAPTER 150

4.1 INTRODUCTION 150

4.1.1 Eukaryotic translation 150

4.1.2 Inhibition of cap dependent eukaryotic translation 152

4.1.2.1 Phosphorylation state of 4E-BP 152

4.1.2.2. eIF4E/mRNA Cap interaction 154

4.1.2.3 eIF4F complex 155

4.1.3 Structural characterization of eIF4E/eIF4G 158

4.2 EIF4E/EIF4G AIMS 159

4.3 PROTEIN PRODUCTION 159

4.4 EIF4E BINDING STUDIES 162

4.5 EIF4E/EIF4G INHIBITOR DEVELOPMENT 166

4.6 PHAGE DISPLAY 170

4.7 eIF4E Crystallography 175

4.8 eIF4E/eIF4G Summary and Further Work 179

CHAPTER 5 : OVERALL CONCLUSION AND DISCUSSION 180

REFERENCES 186

List of Figures

Figure 1 *- The difference in binding sites and inhibition of enzymes and PPIs. A) An interaction between two proteins, illustrating an inhibitor displacing one of the binding partners. B) An enzyme substrate complex, showing an inhibitor fitting in to the enzyme cleft using the lock and key principle. 2*
Figure 2 *- Proteomimetic scaffolds[96] a) Core structue of a 3,2',2'' – terphenyl scaffold developed by the Hamilton group; the first true α helix mimetic. b) Wilson group 3-O Alkylated oligobenzamide scaffold. 16*
Figure 3 *- A schematic showing the general protocol of a phage display experiment. 18*
Figure 4 *– Adhiron non-antibody binding scaffold, variable loops are highlighted in blue. PBD 4N6T[122]. 20*
Figure 5 *– Schematic demonstrating the fluorescence anisotropy assay, both in direct binding and competition mode 21*
Figure 6 *- An illustration of the theory behind microscale thermophoresis, figure adapted from the NanoTemper user manual 24*
Figure 7 *- UV CD spectra associated with various types of secondary structure. The two most common secondary structures are labelled. Solid line = α-helix; dotted line = β sheet. Figure adapted from[134] 25*
Figure 8 *– Schematic showing the specific hallmarks of cancer, adapted from[144]. 28*
Figure 9 *- HIF-1α/p300, cartoon representation illustrating the importance of HIF-1α for transcription in hypoxic environments. 31*
Figure 10 *– Schematic illustrating the importance of the eIF4e/eIF4g binding event in translation. A) complex formed for cap dependent translation to occur. B) complex formed which limits cap dependent translation. 33*
Figure 11 - *Cartoon representation of the domain structure of HIF-1α 59*
Figure 12 *- p300 and CBP domain representation and comparison. A) Cartoon representation if the domain structure of both p300 and CBP. B) Comparison of the amino acid sequence of the CH3 domain of p300 and CBP 61*
Figure 13 *– Structure of Topotecan (chem draw image drawn from pubchem, pubchem ID – 60700) and EZN-2208 (chem draw image drawn from pubchem, pubchem ID – 59443782) 63*
Figure 14 *– Structure of Geldanalycin (chem draw image drawn from pubchem, pubchem ID – 5288382) 64*
Figure 15 *– Structure of Enhinomycin (chem draw image drawn from pubchem, pubchem ID – 23724556) 65*
Figure 16 *– Structure of Bortezomib (chem draw image drawn from pubchem, pubchem ID – 387447) and Chetomin (chem draw image drawn from pubchem, pubchem ID – 73583) 66*
Figure 17 *- Pymol illustration of the interaction between p300 (green) and HIF-1α (blue). PBD ID:1L8C[182] 67*
Figure 18 *-Pymol image illustrating the key binding residues (cyan) of HIF-1α helix 2 (blue) binding to p300 (green). PDB:1L8C[182]. 68*
Figure 19 *- Pymol illustration of the key binding residues (cyan) of HIF-1α (blue) helix 3 binding to p300 (green). PDB:1L8C[182]. 68*
Figure 20 *- Peptide lengths tested for binding affinity. Image adapted from[220] 69*
Figure 21 *- OHM helix mimetics. A) molecular modelling suggests the scaffold projects the side chains in the same special orientation as a helix. B) Four OHM compounds were designed and synthesised. Image adapted from[225]. 74*
Figure 22 *– Illustration of the the additional C terminal residues (cyan) for the two p300 constructs (Pymol, cartoon). PDB ID: 1L8C[182] 78*
Figure 23 *- Cloning of the $p300_{330-420}$ construct, in to the pGex-6p-2 plasmid, image made in Snapgene 79*
Figure 24 *- SDS PAGE analysis of expression trials of $p300_{330-420}$ in expression strains E. coli BL21 (DE3) pLysS GOLD/ROSETTA 2 (R2)/STAR. I =insoluble fraction S=soluble fraction A) GST-p300 (37 kDa). B) SUMO-p300 (24 kDa). C) GFP-p300 (41 kDa). D) Broad range molecular weight marker (2-212 kDa). 80*
Figure 25 *- Effect of zinc and reducing agents on p300 expression, solubility and fold. A) pymol image of p300 coordinating 3 zinc molecules (pale blue) , each zinc is coordinated by 3 cysteines and a histamine. PDB:1L8C[182] B) The effect of zinc concentration on protein structure by CD (top) and*

NMR (bottom)figure adapted from[180]. C) Expression trails of GST-p300 (37 kDa) in the presence and absence of DTT and EDTA. *82*
***Figure 26** - p300$_{330-430}$ purification A) SDS PSGE gel of the purification of GST-p300 by affinity chromatography (GST column). B) SDS PAGE gel illustrating the cleavage of GST-p300 by PreScission protease. C) Size exclusion chromatography spectra showing the separation of GST (26.8 kDa – 154 mL) from p300 (10.8 kDa – 210 mL). Insert SDS PAGE gel showing pure p300 after size exclusion chromatography.* *83*
***Figure 27** - CD analysis of p300$_{330-420}$. A) p300 (18.5 μM) spectra in the presence and absence of EDTA. B) Thermal melt of p300, minima 208 plotted.* *84*
***Figure 28** - Illustration of the potential dimerization of p300 (323-423). The p300 peak elutes at 185 mL (opposed to 210 mL, the expected elution volume for a monomer) .SDS PAGE gels prior to and post size exclusion chromatography show a band at approximately 23 kDa.* *85*
***Figure 29** - CD analysis of p300$_{323-423}$ 0.2 mg/mL in both its monomeric and potential dimeric form.* *86*
***Figure 30 -** Depsispeptide ligation trials. A) and B) Analysis of depsipeptide and sortase pre-reaction. A) SDS PAGE coomassie stained. B) SDS PAGE under UV showing the fluorescence of the depsipepide. C) and D) Analyssi of reaction at the 5 hour time point. C) SDS PAGE coomassie stained showing a band at 10.8 kDa (p300) and 27 kDa (sortase). D) SCS PAGE under UV, the both the p300 and sortase bands fluoresce, indicating a protein-peptide complex*……*88*
***Figure 31** - Biotinylation of p300. A) Biotin-depsipeptide. B) SDS PAGE gel shows a slight increase in size in biotin-p300 from wildtype p300.* *89*
***Figure 32** - ^{15}Np300 purification. A) SDS PAGE gel post affinity chromatography showing a band at 37 kDa which is most likely to be GST-^{15}Np300. B) SDS PAGE analysis of the cut GST-^{15}Np300, both GST (27 kDa) and ^{15}Np300 (11 kDa) band can be seen. C) SDS PAGE analysis of the separation of GST (27 kDa) and ^{15}Np300 (11 kDa).* *90*
***Figure 33 -** The HIF -1α$_{785-826}$ construct used for the direct binding*……*92*
***Figure 34** - Fluorescence anisotropy binding analysis of the HIF-1α/p300 interaction expressed as % bound. The plate was incubated for 30 minutes at room temperature.* *93*
***Figure 35** – Time course of plate incubation for fluorescence anisotropy analysis. Plate incubated between 30 minutes and 18 hours for the HIF-1α/p300 interaction.* *94*
***Figure 36** – Effect of temperature on fluorescence anisotropy analysis, plates incubated between 25 °C and 40 °C for the HIF-1α/p300 interaction. A) Data represented as anisotropy. B) Data represented as fraction of HIF-1α bound.* *95*
***Figure 37** - ITC data for the HIF-1α/p300 interaction at 25 °C.* *96*
***Figure 38** - Schematic showing the peptide HIF-1**α** fragments purchased to investigate binding hot spot regions on HIF-1α* *98*
***Figure 39 -** Fluorescence anisotropy investigation of native HIF-1α fragments binding. a) Binding of the fluorescein labeled helical regions of HIF-1α C-TAD, FITC-HIF-1α$_{794-804}$ (violet) FITC-HIF-1α$_{816-826}$ (red) to p300 CH1 compared to FITC-HIF-1α$_{786-826}$ (black). b) Assessment of the co-operativity the helical regions of HIF-1α. FITC-HIF-1α$_{781-816}$ (green), FITC-HIF- 1α$_{816-826}$ (red) and FITC-HIF-1α$_{816-826}$ in the presence of unlabeled HIF-1α$_{794-804}$ (brown). Note the overlap of the red and grey data sets. c) Schematic of the peptide fragments using in this experiment*……*100*
***Figure 40** - Fluorescence anisotropy competition assay to test the disruption of the HIF-1α C-TAD/p300 CH1 complex by HIF-1α CTAD fragments. HIF-1α$_{782-826}$ (black) HIF-1α$_{794-826}$ (red) and HIF- 1α$_{782-805}$ (violet).* *101*
***Figure 41** – CD spectrum showing the α-helical secondary structure of the p300 mutants. a)H20A. b) L47M. c) I71M.* *102*
***Figure 42** - Investigation of the HIF-1α helix 3 binding pocket on p300. a) Schematic highlighting the three residues of the helix p300 binding pocket of p300 which were mutated. b) Fluorescence anisotropy comparison of the binding of FITC-HIF-1α$_{786-826}$ to wild-type p300 (black) and mutant p300 I71M (red).* *104*
***Figure 43** – Design and testing of p300-oligoamide 1 compound. A) The oligoamide presents three key side chains, Leu$_{818,}$ Leu$_{822,}$ Val$_{825.}$ Helix 3 of HIF-1α in blue, key side chains highlighted in magenta. Pymol cartoon. B) Fluorescence anisotropy competition assay measuring the disruption of FITC-HIF-1α/p300 interaction.* *106*
***Figure 44 -** N-linked oligoamide alpha helix mimetic*……*108*
***Figure 45** - Enrichment ELISA for each of the conditions of the peptide phage display experiment.* *110*
***Figure 46** – Increased representation of the 12mer peptides from the unpanned library to round1 and to round 3 as shown by next generation sequencing.* *115*

Figure 47 *- Analysis of the binding site of PDDP1. a) p300 ^{15}N-^{1}H HSQC experiment, p300 (black) upon titration of PDDP1 at increasing concentrations, 300 μM (blue), 545 μM (green) and 750 μM (red). Insert shows a concentration-dependent shift in one peak and no effect on another peak upon titration of peptide. b) Bar chart showing the size of the shift of all the assigned peaks (BMRB-6268) of p300 after titration of 750 μM PDDP1. c) Mapping of the largest shifts (red) in or around the helix 3 binding site of p300 (green), the helix 3 of HIF-1α is shown in blue with the rest of the peptide shown in teal. The two mutations which cause a statistically significant decrease in affinity (L47M and I71M) are highlighted in magenta. d) Fluorescence anisotropy direct binding measurement of FITC-PDDP1 binding to mutant p300 L71M, the K_d of the mutant is significantly higher then that of wild-type ($p<0.05$).* *117*
Figure 48 *– ELISA results for the selection of clones to be sent for sequencing* *119*
Figure 49 *– Expression and purification of Adhrions (Ad 24 as an example), showing size exclusion chromatography trace and SDS PAGE of Ni column washes and elute and minor and major peaks from size exclusion chromatography.* *121*
Figure 50 *– CD analysis of Ad34. A) spectra run at 20 °C showing a classical β sheet trace. B) thermal melt* *122*
Figure 51 *– SPR sensogram of Ad34 binding to biotin-p300* *123*
Figure 52 *– Inhibition of the HIF-1α/p300 interaction by Adhirons. A) Ad24. B) Ad34 and C) Ad41* *124*
Figure 53 *- Dimerisation of Ad34, one molecule per assymetric unit, hydrogen bonding illustrated by dotted lines* *126*
Figure 54 *- Adhiron structure and potential binding model to p300. a) Crystal structure of Ad34, variable loops shown in cyan. b) Adhiron Ad34 (magenta) docked to the NMR structure of p300 (green) using HADDOCK; key binding residues are shown in cyan. c) NMR structure of HIF-1α (blue) bound to p300 (green) in the same orientation as the docked Adhiron (shown in b). d) Overlay of the docked Ad34 (magenta) and HIF-1α (blue) bound p300 (green), highlighting Ad34 may not dock in a helix binding pocket.* *127*
Figure 55 *- The solubility phage diagram* *130*
Figure 56 *– Fused p300/HIF-1α. HIF-1α (blue), p300 (green) and TEV linker region (red). The amino acid sequence is shown, colour coded as above (ribbons Pymol).* *132*
Figure 57 *– CD analysis of the HIF-1α/p300 fusion construct.* *134*
Figure 58 - *SAXS intensity profiles (logarithmic) for 7 concentrations of each protein sample. A) p300. B) HIF-1α/p300 fusion construct. C) co-purified p300/HIF-1α*……..*136*
Figure 59 - *Linearity of Guinier plots with manual selection of Guinier region. a) Guinier plot of p300. b) Guinier plot of HIF-1α/p300 fusion construct. c) Guinier plot of co-purified p300/HIF-1α*……*137*
Figure 60 *- Normalised pair distribution function P(R) calculated automatically by GNOM . A**)** p300.B**)** HIF-1α/p300 fusion construct C**)** co-purified p300/HIF-1α[292].* *138*
Figure 61 *- Comparison of p300, co-purified p300/HIF-1α and HIF-1α/p300 fusion construct (concatomer). A) Comparison of the normalised pair distribution functions P(R) calculated automatically with AutoGNOM[292]. B) Kratky plots support a globular shape.* *140*
Figure 62 *- Fit of the theoretical scattering profile for the rigid body model (blue line) with the experimental p300 SAXS data[293]. A**)** model generated from p300/HIF-1α complex NMR[182]. B**)** model generated from p300 NMR data (HIF-1α data removed from complex model)* *141*
Figure 63 *- Fit of the theoretical scattering profile for the rigid body model (blue line) with the experimental co-purified SAXS data[293]. A**)** model generated from HIF-α/p300 complex NMR[182]. B**)** model generated from p300 NMR data (HIF-1α data removed from complex model)* *141*
Figure 64 *- Fit of the theoretical scattering profile for the rigid body model (blue line) with the experimental co-purified SAXS data[293]. A) model generated from HIF-1α/p300 complex NMR[182]. B**)** model generated from p300 NMR data (HIF-1α data removed from complex model). C) Average molecular envelope for the HIF-1α/p300 fusion construct [294] overlaid the NMR structure of p300/HIF-1α[182].* *142*
Figure 65 *– Trypsin digest analysis. A)SDS PAGE analysis of HIF-1α/p300 fusion construct cut with 250 nM trypsin over a time course. 0=before addition of trypsin, 1= 5 minutes, 2=10 minutes, 3=30 minutes, 4=1 hour, 5 =2 hours and 6=4.5 hours. B) Full fused HIF-1α/p300 with R_{785} highlighted (magenta), where trypsin potentially cuts (pymol, cartoon). C) Potential product of the after fused HIF-1α/p300 was treated with trypsin. (pymol cartoon).* *144*
Figure 66 *– Structures of helix mimetic scaffolds. A) O-linked oligoamide. B) N-linked oligoamide. C) Hybrid scaffold replacing the central monomer with a natural amino acid.* *147*

Figure 67 *- Control of cap dependant translation. A) Formation of the eIF4F complex of eIF4E, eIF4A and eIF4G, allows the binding of the cap of mRNA which subsequently recruits the ribosome. B) 4E-BP bindings at the same site as eIF4G, blocking eIF4G binding which stops the formation of the eIF4F complex.* *151*
Figure 68 *– Structure of Rapamycin (chem draw image drawn from pubchem, pubchem ID: 5284616)* *153*
Figure 69 *– Structures of Temsirolimus (pubchem ID:6918289), Deforolimus (pubcehm ID:11520894) and Everolimus (pubchem ID:6442177)* *154*
Figure 70 *– Structure of guanine-derived inhibitor (magenta) of the mRNA Cap/eIF4E interaction bound in Cap binding sire of eIF4E (cyan).* *155*
Figure 71 *– Structures of 4EIR-Cat (MDL number:MFCD01931282) and 4EGI-1 (CAS number:315706-13-9)* *156*
Figure 72 *- Crystal structure of eIF4G (red) binding eIF4E (cyan). Key binding residues of eIF4G are highlighted in green.* *158*
Figure 73 *- SDS PAGE analysis of GST-eIF4E as part of the initial expression trials. MW = molecular weight marker, I = insoluble fraction. S = soluble fraction UI = unindiced.* *159*
Figure 74 *- Size exclusion chromatography trace for SUMO-eIF4E, the major peak was collected and the fractions run on an SDS PAGE gel (inset).* *160*
Figure 75 *- CD analysis of the eIF4G peptide, both FITC labelled and unlabelled and in phosphate buffer and TFE.* *161*
Figure 76 *- Fluorescence anisotropy assay measuring the binding of SUMO-eIF4E (black) and eIF4E (red) binding to FITC-eIF4G.* *162*
Figure 77 *–Testing the robustness of the fluorescence anisotropy assay. A) testing the change in affinity of the eIF4E/eIF4G interaction over a 5 hour period, 30 minutes (black) and 5 hours (red). B) the change in affinity of the eIF4E/eIF4G interaction over a temperature range, 21 °C (black), 25 °C (red), 30 °C (purple) and 35 °C (blue).* *163*
Figure 78 *– Structures of Cap and Cap analogue, which have been tested in the fluorescence* *164*
Figure 79 *- Fluorescence anisotropy analysis of the effect of Cap (blue) and Cap analogue (red) on the binding of eIF4E (black) to eIF4G* *165*
Figure 80 *- Fluorescence anisotropy competition assay testing the inhibitor Capability of unlabelled eIF4G peptide* *166*
Figure 81 *– Fluorescence anisotropy competition assay testing the inhibitor 4E1RCat* *167*
Figure 82 *- Fluorescence anisotropy competition assay testing the alpha helix mimetics Oligoamide 1 and Oligamide 6* *168*
Figure 83 *- Fluorescence anisotropy competition assay testing the designed eIF4G mimetic (shown in the insert)* *169*
Figure 84 *- Enrichment ELISA for each condition of the eIF4E phage display experiment, showing an enrichment from the unpanned library to the third panning round* *181*
Figure 85 *- Fluorescence anisotropy assay testing the binding of the peptides (eIF PDDP1, black and eIF PDDP 2 red) generated by phage display to eIF4E.* *173*
Figure 86 *- Crystal structure of apo-eIF4e. Each of the 4 molecules in the asymmetric unit is coloured separately with the eIF4G binding site coloured orange. The squares highlight the non-solvent accessible binding site.* *177*

List of tables

Table 1 *- Table summarising resent example of PPI modulators in the clinic* *6*
Table 2 *– The names and sequences of all the primers used in this study (all 5'-3')* *35*
Table 3 *– Concentrations of antibiotics used in this study, all diluted in* ddH_2O *36*
Table 4 *– Buffers commonly used in this study* *37*
Table 5 *– Protocol used for touchdown PCR* *39*
Table 6 *– Protocol used for single colony PCR* *41*
Table 7 *– Method used for size exclusion chromatography* *45*
Table 8 *– Composition of the media used to express* ^{15}*N-p300* *47*
Table 9 *– Autoinduction overnight media* *48*

Table 10 *– Autoinduction expression media* *48*
Table 11 *– Composition of a 15 % SDS PAGE gel* *50*
Table 12 *– Binding affinities of fragments of HIF-1α*[220] *69*
Table 13 *- Summary of biophysical and in vitro data for peptides designed to target HIF-1α/p300 interactions*[217] *71*
Table 14 *– Binding affinities of OHM ligands*[225] *75*
Table 15 *– Summary of difference in expression strains used in expression trials*[231, 232] *79*
Table 16 *- Reaction conditions to investigate the ligation of a depsipeptide to the N terminal glycine of p300* *87*
Table 17 *- Summary of the ITC data for the interaction of HIF-α/p300 interaction measured at 25°C* *97*
Table 18 *- Affinity of p300 helix 3 binding pocket mutant proteins binding to FITC-HIF-1α*$_{786\text{-}826}$ *as measured by fluorescence anisotropy* *103*
Table 19 *– Structures and* IC_{50} *values for the oligoamide compound library, table adapted from*[244]*.* *107*
Table 20 *– Top 5 peptide phage display sequences for each of the peptide phage display condition* *114*
Table 21 *- Binding study of p300 helix 3 binding pocket mutant proteins binding to FITC-HIF-1α*$_{786\text{-}826}$ *as measured by fluorescence anisotropy* *116*
Table 22 *– Loop sequences of binding Adhirons* *120*
Table 23 *- Phage display derived Adhiron binding data,* K_d *measured by BLITZ and* IC_{50} *measured by fluorescence anisotropy competition assay.* *124*
Table 24 *- Crystallographic data for the adhiron Ad34* *125*
Table 25 *– SAXS derived size parameters for p300, co-purified HIF-1α/p300 and HIF-1α/p300 fusion construct determined by Primus* *135*
Table 26 *– Peptide sequences, affinity and helicity of eIF4G peptide (including mutation D5S) and stapled variants. The symbol (*) signifies positions that are linked by the hydrocarbon staple.* *157*
Table 27 *– The most frequent 5 clones from the unpanned libraries, round 1 and round 3 and the frequency they appear in the pool* *172*
Table 28 *- Crystallographic data and refinement statistics for apo_eIF4E* *176*
Table 29 *– Summary of commonly used expression systems*[345] *184*

Abbreviations

4E-BP	4E-Binding Protein
ΔG	Change in Gibbs free energy
ΔH	Change in Enthalpy
ΔS	Change in Entropy
Ad34	Adhiron 34
APS	Ammonium persulphate
ARNT	Aryl hydrocarbon receptor nuclear translocator protein
ATP	Adenosine triphosphate
BAK	Bcl-2 homologous antagonist killer
Bcl-x_L	B-cell lymphoma-extra large
bHLH	Basic helix-loop-helix

BMRB	Biologiclal magnetic resonanace bank
BSA	Bovine serum albumin
C-term	C-terminus
CD	Circular dichroism
CMV	Cytomegalovirus
cDNA	Complementary DNA
C-TAD	COOH-terminal transactivation domain
DSC	Differential scanning calorimetry
Dmax	maximum particle dimension
DMSO	Dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNAse	deoxyribonuclease
dNTP	Deoxynucleotide
DTT	Dithiothreitol
E.coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
eIF	Eukaryotic initiation factor
ELISA	Enzyme-linked immunosorbent assay
EM	Electron microscopy
FA	Fluorescence anisotropy
FDA	Food and Drug Administration
FBDD	Fragment based drug discovery
FIH	Factor inhibiting HIF
FITC	Fluorescein isothiocyanate
FP	Fluorescence polarisation
FRET	Förster resonance energy transfer
GFP	Green fluorescent protein
GST	glutathione S-transferase
HADDOCK	High Ambiguity Driven protein-protein docking
HBS	Hydrogen bond surrogate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIF	Hypoxia inducible factor
HLH	Helix-loop-helix

HPH	HIF-1α prolyl hyroxylase
HRE	Hypoxic response elements
Hsp90	Heat shock protein 90
HSQC	Heteronuclear Single Quantum Coherence
HTS	High throughput screen
IC_{50}	Inhibitory concentration (50%)
IR	Infra red
IPTG	Isopropyl β-D-1-thiogalactopyransodie
K_a	Association constant
K_d	Dissociation constant
LC-MS	Liquid chromatography–mass spectrometry
MAD	Multi-wavelength anomalous diffraction
Mdm2	Mouse double minute 2 homolog
MRE	Mean residue ellipticity
mRNA	Messenger Ribonucleic acid
MRC	Medical Research Council
MST	Microscale thermophoresis
MW	Molecular weight
N	Stoichiometry
N-TAD	N-terminal transactivation domain
N-term	N-terminal
NEB	New England Biolabs
NGS	Next generation sequencing
NMR	Nuclear magnetic resonance
OD	Optical density
ODDD	Oxygen Dependent Degradation Domain
OHM	Oxopiperazine helix mimetic
PAS	PER-ARNT-SIM
PBD	Protein data bank
PCR	Polymerase chain reaction
PDDP	Phage display derived peptide
PEG	Polyethylene glycol
PER	Period circadian protein

Pfu	Plaque forming units
PHD	Prolyl hydroxylase domain
PLP	Periodate-lysine-paraformaldehyde
PPI	Protein-protein interaction
P(R)	normalised pair distribution function
pVHL	Von Hippel–Lindau tumour suppressor
Rg	radius of gyration
RNA	Ribonucleic acid
RT-PCR	Real time – polymerase chain reaction
SAD	Single-wavelength anomalous diffraction
SAXS	Small angle X-ray scattering
SDS PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SIM	Single-minded protein
SPR	Surface plasmon resonance
SUMO	Small Ubiquitin-like Modifier
TBS	Tris buffer saline
TBS-T	Tris buffer saline - tween
TMB	Tetramethylbenzidine
TFE	trifluoroethanol
TRIS	Tris(Hydroxymethyl)aminomethane
UV	Ultra violet
VEGF	Vascular endothelial growth factor
VEGFR	Vascular endothelial growth factor receptor
WT	Wild type
x-fel	X-ray free-electron lasers
YFP	Yellow fluorescent protein

Amino Acid codes

Alanine	Ala	A
Cysteine	Cys	C
Aspartate	Asp	D
Glutamate	Glu	E

Phenylalanine	Phe	F
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Lysine	Lys	K
Leucine	Leu	L
Methionine	Met	M
Asparagine	Asn	N
Proline	Pro	P
Glutamine	Gln	Q
Arginine	Arg	R
Serine	Ser	S
Threonine	Thr	T
Valine	Val	V
Tryptophan	Trp	W
Tyrosine	Try	Y

Chapter 1 : Introduction

Protein–protein interactions (PPIs) control many physiological processes, from intercellular signaling to programmed cell death in both healthy and diseased states, and therefore represent a large and important class of targets for development of molecular therapies[1, 2]. However, despite the clear importance of PPI modulation, progress in the development of PPI modulators has been hampered by the historic view that PPIs are "undruggable". This view has been shown to be unfounded, as success has been achieved, with several PPI modulators currently in clinical development[1, 3-6] (Table 1[1]). However, despite such progress, PPI modulation remains a challenge of drug discovery. The challenge arises for numerous reasons, most prominently the general lack of small-molecule starting points for drug design and the large size of the PPI interface[2].

This thesis will address the challenges involved in inhibiting α-helix mediated PPIs, using HIF-1α/p300 and eIF4E/eIF4G as model systems. Also demonstrate that structural and biophysical information about the PPI can be utilized in the drug design process, principally for rational and structure based drug design.

1.1 Protein-Protein Interactions General Observations

Traditionally, drug discovery has focused on enzymes as drug targets, as enzymes tend to have well defined pocket-like binding sites, amenable to competitive inhibition by small molecule ligands[7, 8]. In contrast, the interacting surfaces involved in PPIs are larger (1500-3000 $Å^2$) and often lack a clearly defined pocket, making potent and selective inhibition more challenging[9, 10] (Figure 1), as a conventional small molecule[11] could not cover the whole interacting interface. Furthermore, transition-state analogues are available to inform enzyme inhibitor design[7], allowing for a lock and key principle of design. Such analogues are not available as starting points for PPI inhibitor development.

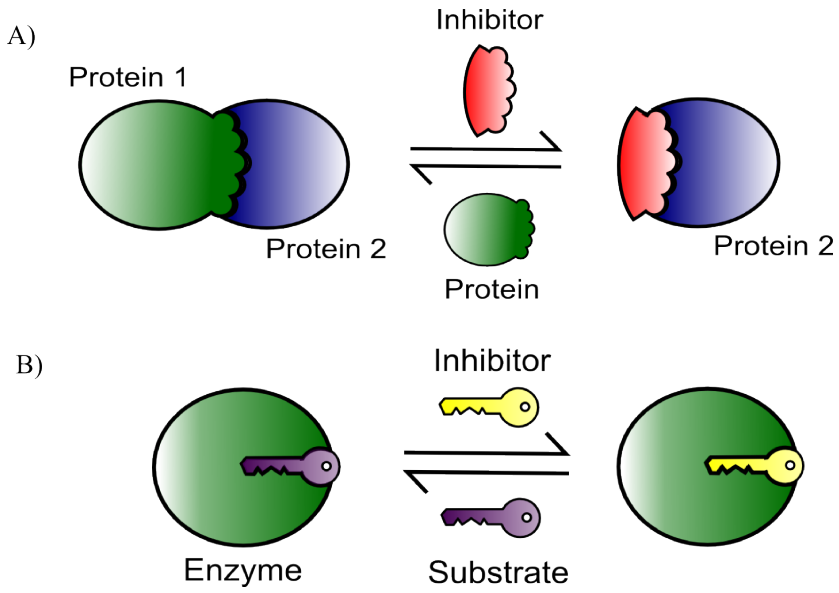


Figure 1 - The difference in binding sites and inhibition of enzymes and PPIs. A) An interaction between two proteins, illustrating an inhibitor displacing one of the binding partners. B) An enzyme substrate complex, showing an inhibitor fitting in to the enzyme cleft using the lock and key principle.

1.1.1 Hot spots

Although the binding interface of a PPI is large, Clackson and Wells demonstrated that a small subset of resides involved, contribute most of the free energy of binding, they termed such residues "hot-spots"[12]. Mutational studies indicate that many protein–protein interfaces contain compact, centralized regions of residues that are crucial for the affinity of the interaction[2, 12-14]. These hotspots, are generally found on both protein partners and constitute less than half of the contact surface of the proteins involved in the interaction[5]. Therefore if we consider the binding interface of only the hot-spots and not the full interface, then the challenge of designing a small molecule to cover the interface becomes less daunting. Proteins involved in PPIs can be 'promiscuous', binding to several targets using the same hotspot region[5, 15]. Structural studies show that these contact surfaces are adaptable, so one protein can interact with a range of structurally diverse partners; presenting the same residues in different structural contexts[5, 15-18]. Structural changes in the hot-spot are observed through side-chain rotations and in the protein backbone; usually in loops

and hinge regions between protein domains[2]. It has been observed that despite having different structures and functions, many PPIs have similar properties at their binding interface; hot-spots tend to be hydrophobic and contain a protruding component[19]. Consequently, although the whole PPI interface is large, there is often a concentrated region to focus small molecule design upon.

A detailed analysis of binding hot-spots has been carried out[20]. Examination of the PDB showed that multi-protein complexes constitute 15% of the databank, of these multi-protein complexes 62% feature an α-helix at the interface, highlighting the role of α-helices in PPIs. However, this does not necessarily mean the whole helix is critical for binding. Computational alanine scanning mutagenesis (using Rosetta) was performed to evaluate the energetic contribution of each residue within a helix to complex formation[21]. After normalization for natural abundance, it was found that aromatic residues and arginine, as well as leucine (the most abundant residue in proteins), are overrepresented as key binding residues at helical interfaces in comparison to polar residues. Hydrophobic and aromatic residues represent the majority of hot spot residues; however, polar and charged residues also contribute significantly[22]. This analysis supports the common perception that PPIs are generally hydrophobic but feature key salt bridges and other polar interactions that influence the binding energy[23].

1.1.2 Previous misconceptions with targeting PPIs

The limited success in targeting PPIs with small molecules has lead to many misconceptions as to the nature of PPIs and the reasoning for the limited number of modulators. The more common misconceptions are outlined below.

Historically PPIs were reported to have large, flat and rigid interfaces, therefore not presenting cavities for small molecules to bind. However, most contact surfaces show some adaptability and flexibility in terms of small loop perturbations and side chain motion[5]. In addition, most interacting surfaces do present cavities available for small molecule binding, although sometimes these cavities are not clear in the structure of either protein when free in solution or bound to one another. For this

reason it should not be assumed that the best binding site for a small molecule is always observed in static structures of either the free protein target or even the protein–protein complex. Molecular dynamic simulations can be used to search for such cavities, for example Bcl-x_L appears to have a flat surface in its static apo state, however simulations reveal small transient cavities; although these cavities are only observed for nanoseconds they still provide potential sites for small molecule modulation[24].

Conventional high-throughput screening (HTS) methods have been deemed unsuitable for PPIs[5]. However, the vast majority of PPI drug development programs still start with empirical screening, either fragment screening or traditional HTS involving small-molecule libraries[5], and this approach has yielded some success. For example Nutlin, an inhibitor of the interaction between Mdm2 and tumour suppressor p53, was discovered through a high throughput screen[25]. The Mdm2/p53 interaction is an attractive therapeutic target in oncology, as p53 tumour-suppressor activity can be stimulated to eradicate tumour cells. Mdm2 activity is dysregulated in many cancers, exerting its oncogenic effects predominantly by inhibiting p53, therefore inhibiting its tumour suppressor activity. It is hoped that inhibiting the Mdm2/p53 interaction with synthetic molecules will lead to p53-mediated cell-cycle arrest or apoptosis in p53-positive stressed cell[26, 27]. One reason why HTS may appear more successful for more traditional drug targets (e.g. enzymes) than for PPIs, is due to the relative ease of enzyme assays, which have a more measureable end point. In addition, current libraries are informed by historical medicinal-chemistry efforts within pharmaceutical companies. These 'chemical phenotypes' (chemotypes) are dominated by past drug-discovery research which have focused on more traditional drug targets such as G-protein-coupled receptors and enzymes. As PPIs are a new target and have different types of interface, a different type of chemotype will most likely be required. Indeed analysis comparing high affinity PPI modulators to compounds targeting more traditional targets using a compound similarity ensemble approach showed very little similarity[28]. It is therefore likely that the current HTS libraries are not suitable for PPIs, rather than the screening approach itself being the problem. Moreover, it should not be assumed that there will be a few 'privileged' scaffolds that will unlock this entire target class, as was the case for protein kinases and G-protein-coupled receptors. Except for close homologues, each

protein–protein interface is different, so the chemotypes of their inhibitors are likely to be more diverse in chemical space[5]. As fragment libraries allow for a greater search of chemical space, it is possible that fragment libraries will yield more success than traditional HTS libraries[28, 29].

Small molecules identified to be PPI inhibitors have been criticised as they were deemed too large to be drugs. It is widely believed that for good oral absorption (and bioavailability), drugs must be less than 500 Da[11, 30]. However, this principle, which is derived from a small subset of known agents, has notable exceptions relating to PPIs. For example ABT-737 (a Bcl-2 inhibitor) is 813 Da and has a reasonable (70%) bioavailability in rodents[31], and a derivative of ABT-737 of similar size is currently in clinical trials. Bcl-2 inhibitors are an anti-cancer strategy as Bcl-2 is a member of a family of anti-apoptotic proteins responsible for dysregulation of apoptosis and prevention of death in cancer cells[32]. Further examples of agents, which have made clinical development despite not lying within Lipinski rule of 5[11], can be seen in Table 1. Still, in terms of pharmacokinetic properties, toxicity and solubility it is usually the case that the smaller the drug candidate the better the properties, however there must always be a trade-off between compound binding affinity and such properties[5].

Table 1 -Table summarising resent example of PPI modulators in the clinic

Target PPI	Inhibitor	Molecular weight (Da)	clogP
Mdm2/p53[33]	RG7112	728	10
β-catenin–/CREB- binding protein[34]	ICG-001	549	6.1
IAP family/SMAC[35]	TL32711	807	2.8
Tubulin/α-tubulin-β[36, 37]	Vinflunine ditartrate/PM391	817	5.4
Tubulin/α-tubulin-β [38, 39]	Eribulin mesylate/E7389	730	1.2
Tubulin/α-tubulin-β[40]	Ixabepilone	507	3.1
IAP/family SMAC[41]	LCL161	501	3.6
BCL-2/BAX and BAK BH3 domain[42]	ABT-263	975	12.4
mTOR/FKBP12[43]	Temsirolimus	1030	7.5
BCL-2/family-BH3 domain; BCL/2-beclin 1; BCL-x_L/beclin 1[44, 45]	(-)-gossypol	519	6.1
Tubulin/α-tubulin-β[46]	Cabazitaxel	836	5.4

All the interactions outlined in Table 1 are implemented in cancer, however their applicability as therapeutic targets differ. The relevance and implication of inhibition of the BCL-2 family proteins and the mDM2 interaction have been outlined previously outlined. The β-catenin–/CREB- binding protein interaction is relevant predominantly in colon cancers, as β-catenin is mutationally activated is 85 % of colon cancers. The β-catenin pathway is implicated in cancers as it regulates expression of a range of genes involved in promoting cell proliferation and differentiation via the recruitment of transcription factors [34]. IAP, inhibitor of apoptosis, are a family of proteins involved in the regulation of the cell's decision to live or die in response to stresses. IAPs exert their effects via interactions with SMAC, second mitochondria-derived activator of caspases, a mitochondrial protein which promotes cytochrome-c dependent activation. Therefore disruption of this interaction may lead to apoptosis and programmed cell death[47]. Tubulins are components of microtubules, which make up the cytoskeleton; they are important in

a number of cellular processes and are critical for cell division, therefore disruption of the system is an attractive cancer target[48]. Mammalian target of rapamycin (mTOR) has emerged as a critical effector in cell-signaling pathways commonly deregulated in human cancers[49]. mTOR has many binding partners, one of which is FKBP12. FKBP12 belongs to a family of proteins that have propyl isomerase activity, therefore is involved in protein folding and is a regulator of the cell cycle[50]. Consequently dysregulation of mTOR may lead to dysregulation of protein folding and the cell cycle.

Thus far in this thesis the concept of protein-protein interactions has been thought of as two proteins with a defined 3-dimentional globular structure binding together. However, many proteins are inherently unfolding, meaning they do not adopt a defined three-dimensional structure; either entirely or in parts, when alone in solution[51, 52]. Indeed more than a third of the eukaryotic proteins have been shown to contain intrinsically disordered regions of over 30 residues in length[53]. There are multiple advantages for a protein to be inherently unfolded; firstly the unfolded nature of the protein will allow for a larger surface area for interactions than that of globular proteins. Secondly, the flexible nature of inherently unfolded proteins allows a conformational flexibility and exposure of short linear peptide regions and interaction-prone structural motifs, this allows the protein to interact with many other proteins, rendering many inherently unfolded proteins promiscuous. The structural and biochemical properties of disordered regions make such proteins ideal for specific recognition of interaction partners, therefore proteins that participate in the binding and regulatory/signalling functions are enriched for intrinsically discorded regions[51-55].

Many therapeutically relevant PPIs consist of one of the binding partners being an inherently disordered protein; with many undergoing a disorder to order transition upon binding[56]. This process means the there are often weak interactions per unit of surface area because the binding energy is spent to organize the disordered partner; the weakness of the interaction may facilitate tighter binding of a small drug molecule to the structured partner as compared with the natural but disordered protein ligand[56]. Knowledge of the disorder to order process and the final ordered structure may aid drug discovery.

1.2 Drug Design approaches

1.2.1 Alternative to HTS

As previously discussed current HTS libraries may not be amenable to PPI targets. This could be rectified by re-designing and building new compound HTS libraries; however this would be costly and time consuming. An alternative to the HTS approach is to use rational/structure based drug design. This process is still quite young and requires adaptation and revisions for each project, however it is becoming an integral part of the drug discovery process[57].

For rational based drug design high resolution structural information about the target (ideally both apo and in complex) should be available. For drug discovery, structural determination is primarily conducted through X-ray crystallography, however other structural techniques (such as NMR) could be useful during the process. Computational methods are also becoming widely used in drug discovery[58, 59]. There have been vast improvements in computational approaches, however there are still limitations; current state-of-the-art docking programs correctly dock ~70–80% of ligands when tested on large sets of protein–ligand complexes[60, 61]. Therefore despite the considerable progress achieved in structure-based virtual ligand screening and computer-aided drug design, the accurate prediction of binding affinities remains an elusive holy grail[57, 62, 63]. Consequently computational methods are often powerfully used in conjunction with both structural and biophysical methods.

1.2.2 Improvements in structure determination methodology

The increase in rational/structural based drug design, can at least in part be attributed to the improvement in structure-determination methodology[59]. The production and isolation of pure protein samples has become more routine leading to the increased production of higher quality protein for crystallisation (as well as for NMR and other structural methods)[64, 65].

X-ray crystallography accounts for ~90 % of all structures deposited in the Protein

Data Bank (PDB). It uses the diffraction of X-rays by molecules within a crystal to calculate an electron density map, into which an atomic model of the molecule can be built. There are a number of difficulties with X-ray crystallography, some of which are practical (e.g. crystal growth and manipulation) and others are fundamental (e.g. the phase problem), however the ability to achieve resolutions that are theoretically only limited by X-ray wavelength (and crystal quality) is the reason it is currently the preferred choice for high resolution structure determination.

Typically, crystal structures determined with data below 2-2.5 Å are acceptable for drug design purposes since they have a high data to parameter ratio, and the placement of atoms in the electron density map is unambiguous (hence why NMR is not ideal for structure determination for drug discovery)[58]. The R factor and R_{free} reported for a model are measures of the correlation between the model and experimental data. The R_{free} should be below 28% (ideally below 25%), and the R factor below 25% in order for the structure to be used for drug design[58]. A further advantage of crystallography over other structural determination techniques is that ordered water molecules can be visible in the experimental data[57]. Knowledge of the water content in binding sites maybe critical during lead optimization as water displacement may contribute significantly to the binding affinity.

There have been significant improvements in the crystallisation process, both in the attainment of crystals and in solving structures. The availability of commercial crystallisation screens and the use of liquid-handling robots that can dispense such screens have decreased the time spent in obtaining the first crystal hit[66-68]. Liquid-handling robots can dispense protein drops volumes of 25–100 nL of protein[69]; this has allowed poorly expressing proteins which were once unworkable for crystallisation, to be used in screens. X-ray data collection has been simplified by the introduction of crystal mounting and alignment robots and rapid detectors[70], allowing for unattended data collection and has generally allowed for more data to be collected within a given time. The development of hot beam lines has also decreased the time required per data set. All of the above has streamlined the crystallisation and data collection process. In addition to the improvements in crystallisation tools, there have been vast improvements in solving the experimental data[71]. Application of synchrotron radiation to single wavelength anomalous diffraction (SAD) and multi-

wavelength anomalous diffraction (MAD) techniques has revolutionized phase determination[72], commonly using substituted selenomethionines, but also other heavy atoms such as argon and even simple halides have been used[73]. The determination of phases has been accelerated by powerful Patterson methods and direct methods for locating the anomalous scatterers[71, 74]. New approaches have provided fully automated phasing[71, 75]. Molecular replacement has become the technique of choice for solving crystal data if suitable structural homologues are available in the PDB. Approximately 70% of structures in the PDB were solved by molecular replacement (although this number is constantly increasing). A general rule states that for molecular replacement to be successful there must be 50% sequence similarity between the two proteins.

NMR is another structure determination method which may be useful in rational/structure based drug design. Although not ideal for structure determination (due to a lower resolution than crystallography) it does have uses in terms of binding site determination and stoichiometry when a crystal structure is not available and protein dynamics. NMR is based upon the fact that certain NMR-active magnetic nuclei (predominantly ^{1}H, ^{15}N and ^{13}C) absorb electromagnetic radiation at the resonant frequency determined by the isotope, exciting these nuclei to a high-energy state. When these nuclei relax back to their equilibrium state they release radiation, the frequency of which is characteristic of their chemical environment and allows the mapping of interatomic distances (through bonds and through space). This results in an ensemble of possible structures and typically the quality of the 20 most probable structures is assessed by their root mean square deviation (rmsd). NMR has the advantage that proteins are observed in solution and flexible regions that may not be resolved by X-ray crystallography can be seen in multiple conformations, providing information about the dynamics of protein folding, protein binding interactions and stoichiometry. NMR also requires less protein at lower purity, and samples can be reused for several experiments[76]. However there are several disadvantages to using NMR, the most important of which are the size limitation (current technology means NMR for 3D structure solution is restricted to proteins <30 kDa in the majority of cases), and the low resolution. The major obstacle in the determination of large proteins (>30 kDa) and complexes by NMR spectroscopy is the increased difficulty of extracting useful structural information as the size of the system to be studied

increases. As resonance overlap becomes a significant problem with the increased size of macromolecules, unambiguous sequence specific assignment of both the backbone and side-chains becomes difficult. This, combined with the disappearance of peaks due to relaxation processes leads to a reduction in the number of distance constraints typically used for structure calculation. Orientation constraints derived from residual dipolar couplings help to overcome the size limitations by providing long-range information on the relative orientation of distant parts of the structure[77-80]. However, they do not provide translational information and by themselves cannot be used to generate an unambiguous solution[80]. NMR structures are not assigned a resolution in the same way X-ray crystal structures are, but a good NMR structure is considered to equate to a 3- 3.5 Å X-ray crystal structure i.e. inadequate resolution for structure-guided drug design. Larger proteins must also be labelled with ^{15}N or ^{13}C, which can result in altered expression and purification requirements, and they must be stable at room temperature in buffers that are typically low salt and below physiological pH (5-6), to slow exchange of amides with the solvent.

Small-angle X-ray scattering (SAXS). SAXS is a low resolution structure characterization method in solution, used to assess the overall shape (envelope) and oligomeric state of proteins and protein complexes[80]. As the technique only generates low-resolution data, high-level structural conclusions cannot be drawn. However SAXS may be useful for the analysis of flexible systems including multi-domain interactions (which could apply to PPIs)[81]. SAXS will also give data regarding the purity and homogeneity of the sample and the overall shape and size of the molecules present in the sample. Such details, purity, homogeneity and flexibility, can be used when assessing how likely a protein is to crystalize, therefore SAXS can be a useful tool when deciding on a construct for crystal trials (for example does removing a flexible loop region improve crystallisation properties?).

SAXS data can in no way replace X-ray crystallography data but instead can be a powerful tool when used in combination, as it can be used as a tool for crystal model validation. The use of SAXS to investigate the solution properties of crystal structures was pioneered in the 1970s and early 1980s, with the development of methods to predict theoretical scattering from crystal structures[80]. SAXS can be used as a tool for molecular replacement using the program FSEARCH[82], which uses the

shape envelopes generated by SAXS for the determination of low resolution phases. Following the correct positioning of the molecular envelope in the unit cell the low resolution phases can be extended to crystallographic resolution. This is a promising method which has already been used successfully for several proteins[80], and continues to be developed[83]. In the same way that SAXS will not replace crystallography data, it will not replace NMR, although it can be complementary. As previously mentioned there is a size limitation to NMR; SAXS data can provide information on the global shape of the macromolecule that can be introduced into the NMR structure calculation and reduce this ambiguity. The inclusion of potentials for the refinement of NMR structures against SAXS data have been introduced into several popular structure calculation packages, and have been shown to significantly improve the accuracy of calculated structures[84, 85].

There are other structure determination methods that are rapidly improving and increasing in popularity; namely cryo-electron microscopy (EM), neutron diffraction and X-ray free-electron lasers (X-FEL). EM was classed as a low resolution technique however the development of new detector hardware has led to much higher resolutions[86]. This is evident from a series of new structures at near-atomic resolution[87-92]. As a general rule EM has been limited to molecules >300 kDa[93], however this limit is decreasing, as last year human γ-secretase complex (170 kDa) was solved at a resolution of 4.5 Å[94]. One major advantage of EM over X-ray crystallography is that the molecule of interest does not have to be crystallized, therefore flexible proteins and larger complexes can be visualized by this method. Despite the vast advances in EM over recent years, the power of EM still lies in its complementarity with X-ray crystallography. Neutron diffraction is a technique analogous to X-ray crystallography, using neutrons diffracted from protons within a crystal instead of electrons. This allows hydrogens to be visualised which offers information about protonation states and hydrogen bonding. This technique is still limited by the lack of neutron sources and challenging experimental requirements (large crystals are required and data collection can take 1-3 days compared to minutes with synchrotron X-rays). X-FEL is a technique still very much in its infancy, however it is deemed the future of X-ray crystallography. As the X-ray pulses in this technique are coherent, the phase problem would be overcome and the increased brightness compared to synchrotron radiation is thought to limit radiation

damage, therefore overcoming two considerable problems with current X-ray methodologies[95].

1.2.3 Rational drug design strategies

One rational approach to PPI inhibitor design uses the secondary structure at the interface as a starting point for inhibitor design. In terms of PPIs there has been much focus and consequently success in targeting α-helix mediated PPIs as they have been recognized as more tractable from a chemical perspective[96]. The α-helix is the most abundant secondary structure motif and it stands to reason therefore that a significant population of PPIs should involve α-helices[20, 97]. The α-helix has 3.4 residues per turn, a rise of 0.54 nm per turn and is defined by backbone dihedral angles θ and ψ of −60° and −45°. The spacing of side chains is such that they are situated above one another every 3–4 residues; thus the α-helix can be considered to have three distinct faces[96-98]. In many cases of α-helix mediated PPIs, there are predominantly hydrophobic residues at the *i*, *i* + 3(4), *i* + 7 positions, which make key contacts with a reasonably deep α-helix binding cleft[96]. Should an α-helix be found at a PPI interface, a mimetic that reproduces the key recognition features of the helix (such as the key side chains) might act as a competitive inhibitor[96].

There are two leading mimetic inhibitor design strategies: peptide mimetics that replicate the local topography of the structural motif[99], and non-peptidic scaffolds that mimic the surface formed by non-sequential hot-spot residues (*i*, *i* + 3(4), *i* + 7) of the α-helix[100, 101].

1.2.3.1 Peptide mimetics

Peptides are highly conserved and have a complex set of functions which a small molecule may not be able to mimic fully[102]. However, as peptides have poor transport properties and are sensitive to proteolytic degradation they are considered therapeutically undesirable[103]. In addition, synthetic peptides are less organized and may lack secondary structure in solution compared to when part of their parent

protein, potentially increasing the entropic cost of binding[104]. Several stabilisation approaches are being explored to replicate the local topography of an α-helical motif found at a protein interface while also addressing the problems of proteolytic stability and other pharmacokinetic factors. Strategies for helix stabilisation include salt bridges, metal chelates and covalent cyclisation methods such as disulphide and lactam bridges[105, 106]. There are several considerations required before constraining peptides, firstly although it seems intuitive that helical content should be maximized to decrease the entropic penalty of binding, this risks constraining a peptide in an inactive confirmation. Consequently a less well-defined conformation may be considered adequate with a linker introduced merely to reduce the number of degrees of freedom of the peptide in the unbound state. Consideration should also be given to which amino acids to crosslink and their relative positions, where in the sequence they should be and the length and flexibility of the linkers[96].

1.2.3.2 Non-peptidic scaffold mimics

Proteomimetics match the topography of the original α-helix by mimicking the spatial orientation of its key recognition residues, rather than recapitulating the helical conformation[96, 107]. Within an α-helix there is a hydrogen bond network which stabilizes the helical conformation and orients the key binding residues along the solvent-exposed surface. It is possible to simplify this pharmacophore to a rod-shaped object presenting side chains in a spatial orientation comparable to that of a native α-helix. Such molecules are likely to be more synthetically tractable and confer more drug like properties. This pharmacophore may allow the development of a template scaffold amenable to many different PPIs simply by matching side chains to the helical target of interest[96].

The Hamilton group identified the first true α-helix mimetic[108]; a series of trisubstituted 3,2′,2″-terphenyl derivatives in which the aryl core assumes a staggered conformation projecting the *ortho* substituents to mimic the position of the *i*, *i* + 3(4) and *i* + 7 residues of the helix. This series of compounds, and derivatives have demonstrated inhibitory capabilities (IC_{50} in the nanomolar range) for numerous PPIs including calmodulin/phosphodiesterase, Bcl-x_L/Bak and gp41 assembly interactions[108-110]. Further studies have shown the ability of these ligands to disrupt

interactions in cells using the BH3-mediated interaction with Bcl-x_L in HEK293 (human embryonic kidney) cells as a model[111, 112]. As these ligands have a template scaffold it is important to establish selectivity, in this case selective inhibition of the Bcl-x_L/Bak interaction over the p53/hDM2 PPI was indeed achieved by a subtle exchange of a methyl-1-naphthyl for methyl-2-naphthyl side chain in the terphenyl sequence[111].

Since this initial success considerable efforts have been invested in development of scaffolds which have a more versatile synthesis amenable to library generation and greater drug like properties[96, 113-115]. Although structural rigidity is important in reducing the entropic cost, it may also be beneficial to have some flexibility to maximise the interactions with the target through induced fit[96]. Oligobenzamide mimics, developed by the Wilson group, appear to allow some flexibility[98]. Further structural studies by the Hamilton group on the oligobenzamide mimics suggested that backbone curvature can be controlled to match that of target α-helices by using combinations of pyridine and benzene rings in the scaffold[116]. The Wilson group have demonstrated the inhibition properties of 3-*O*-alkylated aromatic oligoamides demonstrating that this scaffold could generate compounds with comparable affinity to the native peptide against the p53/hDM2 interaction[98]. The synthesis of this scaffold is amenable to efficient library generation using solid-phase synthesis[117].

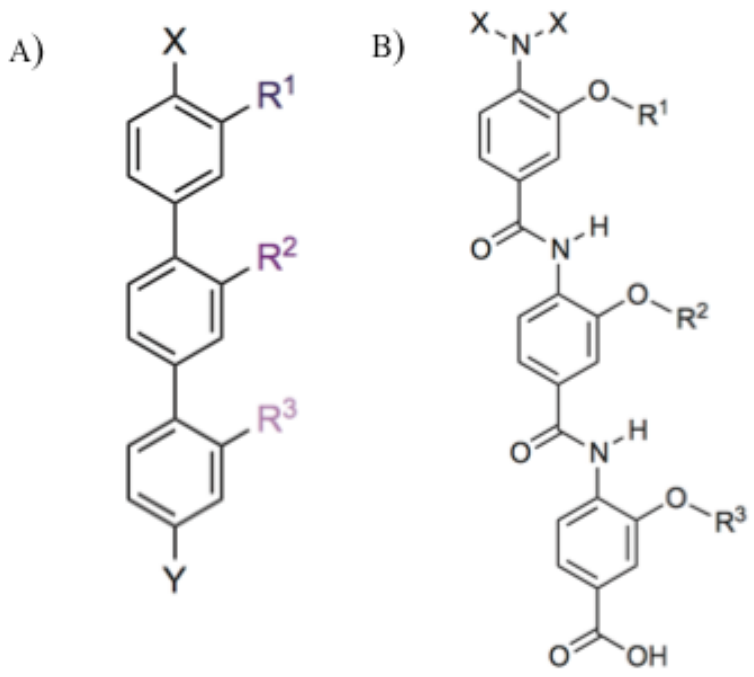


Figure 2 - Proteomimetic scaffolds[96] a) Core structue of a 3,2',2'' – terphenyl scaffold developed by the Hamilton group; the first true α helix mimetic. b) Wilson group 3-O Alkylated oligobenzamide scaffold**.**

1.3 Characterisation of binding interface

Both computational analysis and rational design require an intimate knowledge of the protein structure to identify binding hot-spots, ideally this structure will be a crystal structure. However, in some cases the crystal structure is not available, in such cases other methods must be utilised to understand the binding interface. These methods include biophysical assays, mutagenesis and phage display; as well as other structural methodologies.

1.3.1 Ligand discovery by phage display

Phage display is a powerful combinatorial method, using technology based upon the ability of phage particles to display various structures (including peptides and larger protein scaffolds) on the bacteriophage coat proteins, by insertion of the encoding nucleotides into the phage DNA. This allows the development of large and diverse libraries ($>10^{11}$) using simple molecular biology methods[118], which can be amplified via passage through a bacterial host (such as *E. coli*), making phage library generation relatively cheap and quick. The phage display experiment consists of: isolation of the protein target, incubation with the phage library, washing away of non-binding phage, elution and amplification of the phage that bind. This process (panning) is repeated to obtain high affinity binders to the target protein (Figure 3)[119]. Binders will then be sequenced, establishing a link between genotype and phenotype[120].

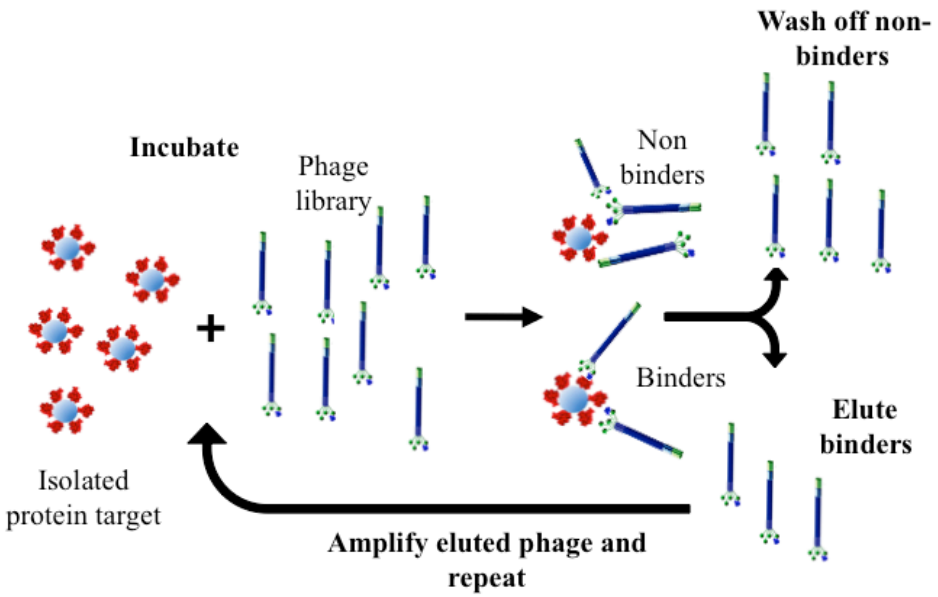


Figure 3 - A schematic showing the general protocol of a phage display experiment.

The high affinity binders can be assessed for binding affinity to the target protein and the location of high affinity binding peptides may highlight a PPI binding hot-spot. This peptide can then be taken forward to be used in computational screening or rational design. As phage do not operate under the same evolutionary pressures as natural binding partners, it is reasonable to believe that epitopes distinct from those found on the natural ligand will be discovered; it is also possible that transient binding sites or sites which are induced by the binding of the phage-derived ligand will be discovered. As PPI interfaces are large this further exploration of the interface may yield binding sites which otherwise would have remained unexplored. Phage display may also identify localised binding epitopes that are readily transferred to potent small molecules scaffolds. The pharmacophore of the binders may also shed light upon the properties of both the binding interface and any potential small molecules[119].

Phage display has advanced to not only display short peptides but also more complex binders such as non-antibody binding proteins. In terms of binding proteins antibodies are the most commonly used[121]. However, as there are limitations to using antibodies *in vitro*, efforts have been made to engineer non-antibody/artificial binding proteins which mimic the molecular recognition properties of antibodies but

with improved properties[122]. Non-antibody proteins are in general small (200 amino acids), monomeric, stable and easily expressed in *E. coli*. Most do not contain cysteines, enabling the introduction of a cysteine for site-specific coupling of biotin, fluorescent labels, or polyethylene glycol to enhance their utility or stability. These characteristics make non-antibody proteins powerful tools capable of replacing antibodies in a range of applications including research[123], diagnostics[124], *in vivo* drug discovery studies[125] and even as a novel class of therapeutics[126] including multivalent and/or multi-specific protein therapeutics[127]. Phage display methods will be a powerful tool for selection of high affinity binding scaffolds.

The Adhiron[122] (commercially known as Affimer) is a non-antibody scaffold based upon a consensus sequence of plant-derived photocystatins which are small (ca.100 a.a.) protein inhibitors of cysteine proteases[128]. This consensus protein displays good protease inhibitor activity as well as the properties of non-antibody binding proteins (small, monomeric, high solubility, high stability, lack of disulphide bonds and glycosylation sites), therefore it is a good scaffold for peptide presentation. The inhibitory sequences with Gln, Val, Val, Ala, Gly and Pro, Trp, Glu loops of the consensus photocystatin were replaced with nine randomized amino acid positions in each loop (Figure 4). The peptide loop regions are suitable for molecular recognition and are expected to adapt to form molecular contacts with a wide range of targets allowing interactions with protein pockets, protein surfaces, peptides and small molecules. The Adhiron scaffold yields high-level purification of soluble protein (typically 10–100 mg/L) from *E. coli* and displays high thermostability with a melting temperature of 101 °C[122]. There has been some success in the generation of high affinity binding Adhirons[122] and the generation of Adhirons which inhibit PPIs using phage display technologies (unpublished data – D. Tomlinson, University of Leeds).

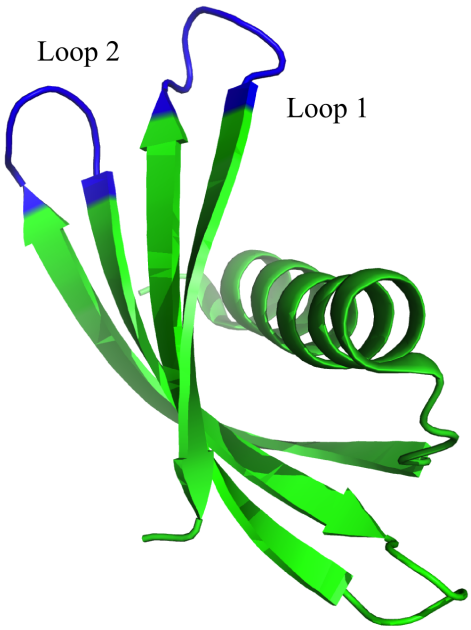


Figure 4 – Adhiron non-antibody binding scaffold, variable loops are highlighted in blue. PBD 4N6T[122].

1.3.2 Biophysical analysis

Biophysical analysis of the native PPIs as well as ligand binding (either in the form of phage display generated peptide/non-antibody binding proteins or small molecules) is an important part of the drug discovery process. Many biophysical methods will give an overall affinity of the two interacting molecules but different methods will give additional information (e.g. thermodynamic and kinetic information) and it is therefore important to consider the data parameters required when deciding which biophysical method to use. Below the biophysical techniques used during this project are described, and briefly those used within the drug discovery process.

Fluorescence anisotropy assays are used to measure rotational diffusion of a molecule (tumbling rate). This rotational diffusion is primarily affected by the molecule size, the larger the molecule the lower the rotational diffusion. This rotational diffusion can be measured by the ratio of polarized light within the total light intensity. When polarized light is applied to a group of randomly oriented fluorophores, most of the excited molecules will be those oriented within a particular

range of angles to the applied polarisation. If they have low rotational diffusion (larger molecules) the fluoresced light will remain polarized (high anisotropy). When the fluorophores can freely change their orientation (high rotational diffusion of small molecules) before re-emitting the photons, the degree of polarisation of the emitted light will be reduced (low anisotropy). Therefore this phenomenon can be used to measure the potency of binding events. One of the protein partners (generally a smaller peptide) is fluorescently labeled, upon binding to a larger protein the rotational diffusion is reduced, increasing the fluorescence anisotropy. Plotting the fluorescence anisotropy against protein concentration of the unlabeled protein partner allows the calculation of K_d. This assay can also be adapted to measure the affinity of inhibitors. Starting with a pre-bound complex the rotational diffusion of the fluorophore will be low; therefore the fluorescence anisotropy will be high. When an inhibitor disrupts this complex, the rotational diffusion of the fluorophore will increase, resulting in a decrease in fluorescence anisotropy (Figure 5).

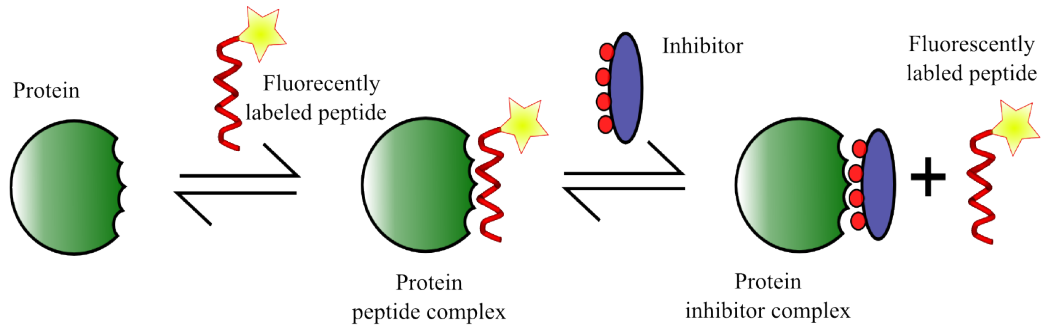


Figure 5 – Schematic demonstrating the fluorescence anisotropy assay, both in direct binding and competition mode

This technique is relatively high throughput and amenable to robotics, making it an appropriate tool for primary screening, both in direct binding and competition mode. It is limited as one of the proteins (ideally the smaller binding partner) is attached to a fluorophore. There is a chance that the fluorophore itself may bind the protein, however this is easily controlled for. A more likely situation is that the fluorophore inhibits binding. However there are many different fluorophores which can be used, and they can be attached at different positions on the molecule. Therefore it is likely that the limitation of fluorophore inhibiting binding can be overcome.

Isothermal titration calorimetry (ITC) gives thermodynamic data about an interaction by directly measuring the heat that is released or absorbed when a complex is formed (protein–protein or protein–ligand complex formation)[1]. An isothermal titration calorimeter consists of two cells, a sample cell and a reference cell. A syringe is used to titrate one binding partner into the sample cell containing the other binding partner. This will cause a heat change in the sample cell. The power required to maintain isothermal conditions between the sample and the reference cell is measured by a calorimeter. A plot of power ($\mu cals^{-1}$) against time (min) shows a series of peaks of heat flow (power). The area under the peaks gives the enthalpy (ΔH) of binding[129, 130]. A binding isotherm obtained from the plot of heat of formation against the molar ratio of ligand to biomolecule (gives information about the enthalpy (ΔH), the association constant (K_a) and the stoichiometry (n) of binding, which can be used to calculate changes in entropy (ΔS) and Gibbs Free Energy (ΔG)[130, 131].

ITC is a label free method, therefore there is no concern that a label, such as a fluorophore, will inhibit or indeed promote binding. ITC is also a method where neither binding partner is immobilized; therefore there is no concern about the orientation of one of the binding partners being restrictive for binding. In addition, it does not have limitations associated with clarity of the solution, molecular weight, temperature or pH. A limitation of this technique is that it is relatively low throughput and is therefore used as a secondary screening method to eliminate false positive hits after primary screening in the drug design process.

Surface plasmon resonance (SPR) is a technique that measures the interaction of molecules through changes in the refractive index that occur when ligands that are bound to a surface interact with another molecule in solution[1]. SPR occurs when light is reflected off thin metal films. A fraction of the light energy at a specific angle (SPR angle) can interact with delocalised electrons within the metal film (plasmon) resulting in a reduction in the reflected light intensity. If binding occurs to the immobilized target the local refractive index changes, leading to a change in SPR angle, which can be monitored in real-time by detecting changes in the intensity of the reflected light. The rates of the change in reflected light intensity can be analysed to give association and dissociation rate constants. The ratio of these values gives the

apparent equilibrium constant (affinity). The size of the change in SPR signal is directly proportional to the mass being immobilized and can thus be interpreted crudely in terms of the stoichiometry of the interaction.

SPR directly measures the on and off rates of the interaction and this may be important for lead optimization. A limitation of SPR is that one of the binding partners (ideally the smaller partner) must be immobilized on to the chip. This adds concern that the orientation of this molecule may hinder binding, for example if the binding site is occluded. However there are many different ways a molecule can be captured on a chip (amide coupling, streptavidin-biotin or His-nickel) so different approaches can be tested to optimize binding.

Microscale thermophoresis (MST) is a relatively new biophysical technique that enables the analysis of molecular interactions in solution at the microlitre scale. MST measures the movement of molecules along temperature gradients, an effect termed thermophoresis. Molecules will move away from localized heat; this movement is affected by size, charge and entropy of the molecules; therefore the thermophoresis of a molecule will be different when within a complex. The thermophoresis of fluorescently labeled molecules is measured by monitoring the fluorescence distribution inside a capillary. In solution the fluorescence distribution is homogeneous. When a microscopic temperature gradient is established by an IR-Laser, a new fluorescence distribution arises (Figure 6)[132]. MST is time and cost effective, with low sample consumption and fast measurement of samples. MST can measure a broad range of affinities and sizes, therefore is useful for the testing of small molecules. However like fluorescence anisotropy, one component must be fluorescently labeled, so again there is a concern that the fluorophore may impact on binding.

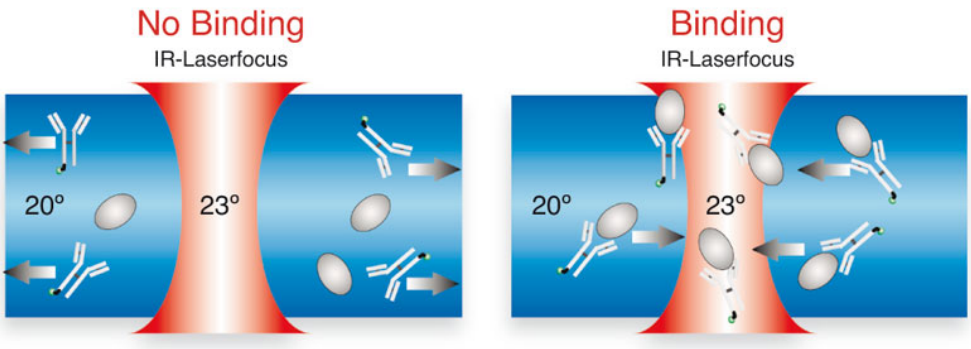


Figure 6 - An illustration of the theory behind microscale thermophoresis, figure adapted from the NanoTemper user manual

Förster Resonance Energy Transfer (FRET) is also used to measure molecular interactions. FRET is the radiationless transmission of energy from a donor molecule to an acceptor molecule. The donor initially absorbs the energy and the energy is subsequently transferred to the acceptor when in close proximity (typically 10-100 Å). This resonance interaction occurs over greater than interatomic distances, without conversion to thermal energy, and without any molecular collision. For FRET to occur the excitation spectrum of the acceptor must overlap the fluorescence emission spectrum of the donor[133]. One of the major advantages of FRET is that it can be conducted *in cellulo* (as well as *in vitro*), unlike the other biophysical techniques described. FRET can also be used as a molecular ruler as the FRET efficiency (Q) is proportional to 1/distance to the power 6 (r^6) between the two fluorophores.

As can be seen, each of the biophysical techniques, ITC, SPR, fluorescence anisotropy, MST and FRET, have different strengths and weaknesses and therefore are appropriate to use at different parts of the drug discovery process, both for PPI characterisation and testing of small molecules. Each of the biophysical techniques give an indication of binding affinity, however there are other techniques which give other information, for example circular dichroism (CD), differential scanning calorimetry (DSC) and mass spectrometry.

The main application of CD is measuring the secondary structure content of proteins, CD can also be used to give an indication of protein stability, both in different buffers and thermal stability by thermal melt. CD functions through the interaction of circularly polarized light with chiral molecules. Circular polarised light has a

"handedness"; it is either rotating left or right. Bonds around a chiral centre absorb left and right polarised light differently; it is this difference that is measured by the spectrapolarimeter (CD spectrometer). The spectrapolarimeter light source switches between left and right polarized light and the samples absorbance of each type of light is compared; the difference between the two signals is known as the ellipticity (θ). In practise the protein concentration and the path length of the cuvette are used to give mean residue ellipticity (MRE).

Each secondary structure has a characteristic CD signature, these spectra are shown in Figure 7[134]. The α-helix spectra has two minima 208 nm and 222nm; these can be followed to assess protein stability. As the protein unfolds, for example due to heat, these minima will increase towards 0; in stability studies the change in MRE is measured.

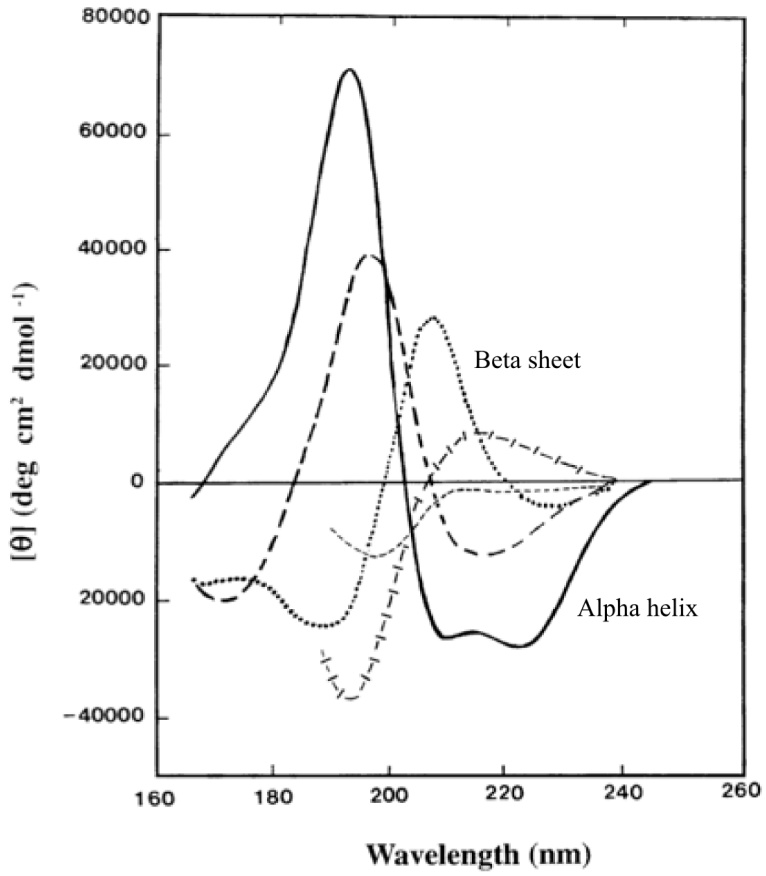


Figure 7 - UV CD spectra associated with various types of secondary structure. The two most common secondary structures are labelled. Solid line = α-helix; dotted line = β sheet. Figure adapted from[134]

Other techniques such as DSC and fluorescent reporter dye assays can give information about the thermal stability of a protein, although neither give information about secondary structure. DSC can be particularly useful if refolding information is required. Fluorescent reporter dye assays are high throughput and therefore can be useful for the characterisation of larger libraries. Mass spectrometry has been recognised as having a potential for fast, sensitive and high throughout analysis of weak covalent complexes[135]. As well as providing exact mass and sequence specific information, mass spectrometry has been reported to provide dissociation constants of protein–ligand complexes from titration experiments[136]; information on conformational changes upon binding (from ion mobility measurements and collision cross section calculations)[137]; location of the binding site (from epitope mapping approaches)[138]; binding stoichiometry and binding specificity (from competition experiments)[139]; isotopic information such as oxidation states of binding metal ligands[140]; relative quantification or subunit exchange that can occur within multicomponent- protein complexes[141].

Each of the techniques described have been used within drug discovery projects, either to characterise the binding event of the native interactions or the ligand/target interaction. As each of the techniques is able to measure different parameters consideration should be given to the techniques capabilities before deciding which techniques to use.

1.4 The targets of interest

1.4.1 Targeted therapy

The two PPIs investigated in this thesis are the HIF-1α/p300 and the eIF4E/eIF4G interactions; both targets are implicated in cancer development. Cancer is a major concern to public health and although the mortality rate is decreasing, cancer remains one of the biggest problems of the modern medical era[142]. As well as only seeing modest reductions in mortality rate, many current cancer treatments such as chemotherapy and radiotherapy are associated with an abundance of adverse effects. The adverse effects occur because such treatments target the obligate rapid proliferation of cancer cells[143]. Therefore such treatments are targeting all targeting all rapidly dividing cells, not just cancer cells. Patients can expect to suffer fatigue, hair loss and nausea as well as many other debilitating symptoms[143]. It is for these reasons that the development of novel anti-cancer agents is of paramount importance. To achieve this, new targets must be identified and novel methods of intervention developed. These new targets should not target all rapidly dividing cells but target cancer cells more specifically. Targeted therapies are substances that block the growth and spread of cancer by interfering with one of the specific hallmarks of cancer highlighted in Figure 8, it is suggested that most, if not all cancers have acquired the same set of functional capabilities during their development, albeit through different mechanistic strategies[144].

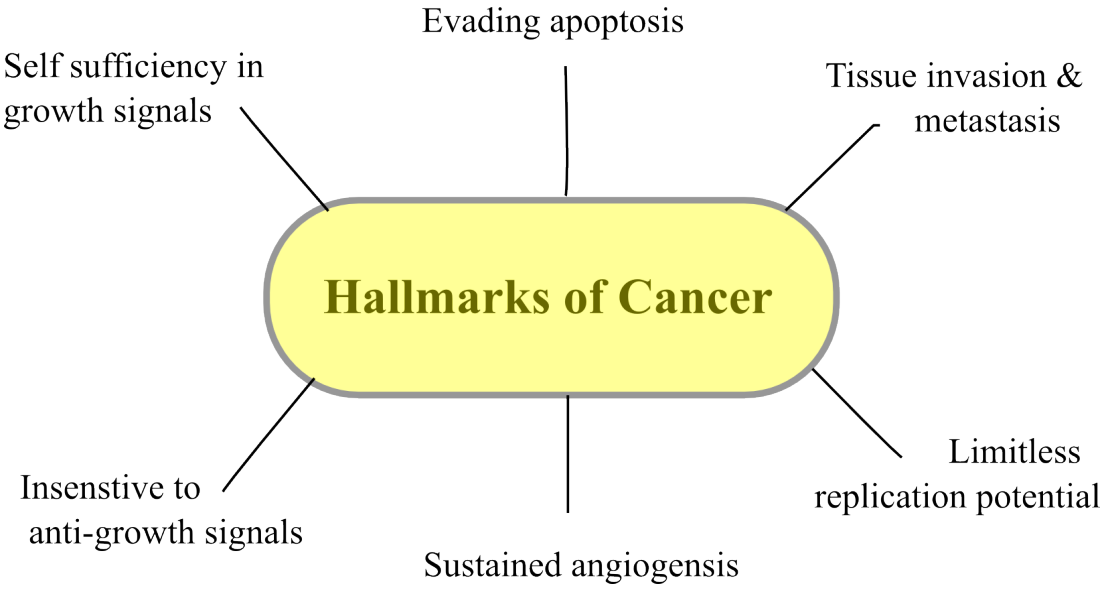


Figure 8 – Schematic showing the specific hallmarks of cancer, adapted from[144].

One approach for target selection is to compare the amounts of individual proteins in cancer cells with the amounts found in normal healthy cells. Proteins more abundant in cancer cells than in normal healthy cells would be potential targets, especially if they are known to be involved in cell growth or survival. Another approach to identify potential targets is to determine whether cancer cells produce mutant proteins that are involved in cancer progression. Once a potential target has been identified, the next step is to develop a therapy that affects the target in a way that interferes with its ability to promote cancer cell growth or survival. For example, a targeted therapy could reduce the activity of the target or prevent it from binding to a receptor. Targeted therapies do have some limitations, one is they are more prone to resistance, which mainly occurs in two ways; the target itself changes through mutation so that the targeted therapy no longer interacts well with it, and/or the tumor finds a new pathway to achieve tumor growth that does not depend on the target. Resistance can be overcome, or at least slowed; by using the therapies in combination with either another targeted therapy or chemotherapy.

Targeted therapy is based on the basic principle of preferentially killing cancer cells, without having any significant toxic effect on normal cells. This can be achieved by targeting specific proteins mutations and interactions, as previously described, and

additionally as a targeted delivery approach. Many of the adverse effects of current cancer therapy can be attributed to ineffective delivery of specific molecules, this leads to ineffective therapeutic drug concentration reaching the tumor site, adverse effects caused by nonspecific tissue distribution of anticancer agents, and acquired resistance of the cancer cell upon chemotherapy that triggers cross-resistance to a wide range of different drugs. A targeted delivery approach would mean that the active agent was spatio-temporally delivered to the correct location at the appropriate time[145]. Targeting strategies have attempted to take advantage of low extracellular pH often found in tumours, elevated enzymes in tumour tissues, the hypoxic environment inside the tumour core and tumour-specific antigens expressed on tumor cell surfaces[145]. Recent advances in targeted delivery have focused on prodrugs and carriers[146]. The prodrug concept is an alternative approach to design less reactive and less cytotoxic form of anticancer drugs[147, 148]. Pharmacologically inert drug molecule derivatives that can be converted to their active form *in vivo*, either enzymatically or non-enzymatically, to exert a therapeutic effect are used[146]. Traditional produgs temporarily alter the physicochemical properties of drugs to improve drug pharmacokinetics, prolong action, reduce adverse effects and increased selectivity. To become a targeted prodrug the prodrug is conjugated to a carrier such as sugars, growth factors, vitamins, antibodies, peptides and synthetic polymers that can transport the drugs to tumours and subsequently release it outside or inside the tumour cells, therefore the drug is inactive until at the tumour site. This targeting can be achieved in different ways. One such way is to develop a carrier such as monoclonal antibodies[149, 150] that binds to antigens or receptors expressed on tumor cell surface. If used in combination with a targeted therapy, targeted delivery will be a very powerful tool for limiting adverse effects and intensifying the effect of the anti-cancer agent.

Inhibitors against HIF-1α/p300 and the eIF4E/eIF4G would be a targeted therapy. HIF-1α/p300 inhibitors could be a targeted therapy for solid tumours such as breast cancers as such tumours have a limited blood supply[151]. eIF4E has been shown to be up-regulated in multiple cancer types, including malignancies of the prostate, breast, stomach, colon, lung, skin, and the hematopoietic system[152], eIF4E/eIF4G inhibitors could therefore be a targeted therapy against these cancers.

The role of each of these interactions in cancer development will be briefly described here however each interaction will be discussed in more detail in the respective chapters.

1.4.2 HIF-1α/p300

Hypoxia inducible transcription factor (HIF) is a transcription factor and is regarded as a potential protein for targeted cancer therapy due its role within cells in hypoxic environments[153]. As the multiplication of cancer cells is rapid, tumours quickly outgrow their blood supply. This results in a hypoxic environment, which is characteristic of cancerous cells. The hypoxic environment may therefore be exploitable in cancer therapy[153]. There are three isoforms of HIF; HIF-1, HIF-2 and HIF-3[143]. Although the exact role of each isoform is not fully established, it is HIF-1 that is described as the messenger to activate transcriptional responses to hypoxia and therefore regarded as a potential cancer target[143]. HIF-1 is a promiscuous heterodimeric transcription factor; composed of an α subunit and a β subunit[154]. The link between HIF-1 and cancer was established by immunohistochemical analysis of human cancer biopsies[155]. The level of HIF-1α has been shown to be increased in cancer cells relative to normal tissue, in the majority of cancers[155]. Furthermore, experimental data has complemented clinical data, showing that in the absence of HIF-1α there is decreased tumour growth, vascularization and metastasis[156]; whereas, the opposite is true when HIF-1α is over expressed. This demonstrates a causal relationship between HIF-1α and cancer progression[157]. Numerous studies have been conducted giving evidence that HIF-1α activation promotes oncogenesis or cancer progression. Clinical data has also linked high levels of HIF-1α to a poorer prognosis and increased mortality rate[158]. Therefore there is an extensive bank of data demonstrating the relationship between HIF-1α and cancer.

The β subunit is constitutively expressed in the nucleus whereas the stability, subcellular localization and transcriptional potency of the α subunit is affected by oxygen concentration[153]. HIF-1α is continuously expressed at a low level in the cell, however in normoxic conditions HIF-1α is rapidly degraded, most prominently via a Von Hippel–Lindau tumor suppressor (pVHL)-mediated ubiquitin-proteasome pathway (Figure 9)[143]. This process is initiated by the hydroxylation of two proline

residues by HIF-prolyl hydroxylase (HPH). HPH requires oxygen to complete this hydroxylation and oxygen availability is the rate-limiting step in this reaction[159]. Upon hydroxylation, pVHL is able to bind to HIF-1α, which subsequently allows the recruitment of the E3 ligase complex. This allows the attachment of ubiquitin via the lysine residues of HIF-1α, thereby marking HIF-1α for degradation. Due to the rapid nature of this pathway HIF-1α has a half life of less than 5 minutes in normoxic conditions; resulting in no detectable protein in normoxic cells[160]. However, in hypoxic environments oxygen availability is limited, therefore hydroxylation of the prolines is inhibited. Consequently HIF-1α is not rapidly degraded and the stability and transcriptional activity of HIF-1α increases within the cell. This allows HIF-1 to bind to hypoxia response elements[143].

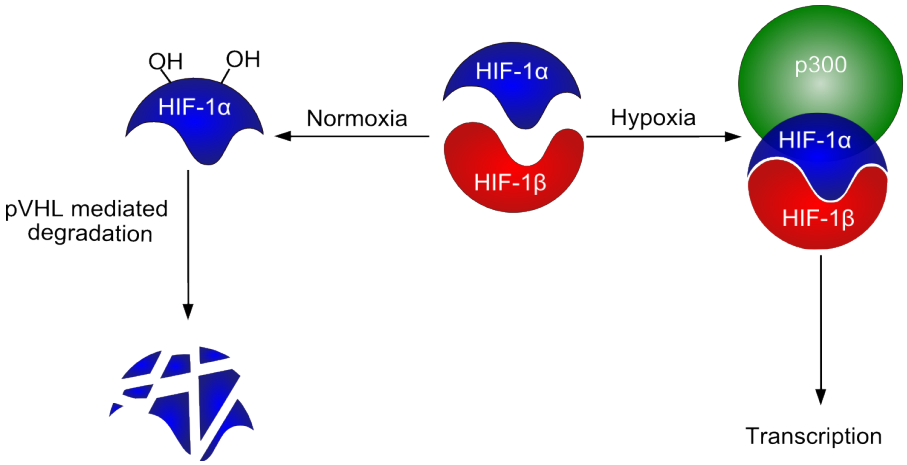


Figure 9 - HIF-1α/p300, cartoon representation illustrating the importance of HIF-1α for transcription in hypoxic environments.

The hypoxia response elements have many roles in normal and cancer biology. These include angiogenesis[161], stem cell maintenance[162], metabolic reprogramming[163], autocrine growth factor signaling[164], metastasis[165] and providing a mechanism of resistance to radiation and chemotherapy[166]. It is therefore unsurprising that there are also many molecular mechanisms to inhibit the activity of HIF-1α. These mechanisms include decreasing mRNA levels, decreasing protein synthesis, increasing degradation, inhibiting heterodimerization, inhibiting the HIF-1α/DNA interaction and decreasing the transcriptional activity of HIF-1[166]. A reduction of transcriptional activity could be achieved by targeting the interaction between HIF-1α and its co-activator p300[166]. It is this interaction which is of interest

for this project. p300 and the paralogous CREB binding protein (CBP) bind to many transcription factors to control expression of their targets genes. The structure and function of both HIF-1α and p300 will be discussed in more detail in 3.1 Introduction to HIF-1α/p300.

1.4.3 eIF4E/eIF4G

Another way in which cancer cells may be targeted preferentially over healthy cells is to target translation initiation factors (which translate oncogenes), one such initiation factor is Eukaryotic translation Initiation Factor 4E (eIF4E). The control of translation of mRNA is crucial for gene expression and is one of the key ways in which protein levels may be selectively regulated[167]. Translation requires an initiation step where the cap at the 5' end of an mRNA binds to eIF4E, which recruits the ribosome[168]. This initiation is mediated by many eukaryotic initiation factors (eIFs) and is regulated via changes in their phosphorylation states and by endogenous protein[167]. An important step in this initiation is the formation of eukaryotic initiation factor 4F, which is made up of three subunits; eIF4E, the Cap binding protein, eIF4A the RNA helicase, and eIF4G the multidomain adaptor. eIF4E is described as the master switch that controls eukaryotic translation[167]. eIF4E is responsible for the delivery of the 5' end of the mRNA to the RNA helicase eIF4A[167]. eIF4E is the rate limiting protein within the complex, as eIF4E activity is tightly regulated via an interaction between either eIF4G or endogenous 4E-binding protein (4E-BP) (Figure 10). The interaction with 4E-BP negatively regulates the formation of the eIF4F complex, as the availability of eIF4E to complete the complex relies on release from 4E-BP[169]. Therefore the binding of eIF4E is competitive between eIF4G and 4E-BP[168]. When hypophosphorylated 4E-BP binds eIF4E with high affinity, preventing the eIF4E/eIF4G interaction. Hyperphosphorylation of 4E-BP causes dissociation of the 4E-BP/eIF4E complex, allowing eIF4E to bind to eIF4G, resulting in the formation of eIF4F complex. mRNA's with unstructured UTR are translated preferentially to mRNAs with structured UTR's; many mRNA's of proteins associated with cancer have structured UTR's. Therefore an increase in eIF4F formation enhances the translation of a subset

of mRNAs involved in critical cellular processes implicated in oncogenesis such as cell proliferation, evasion of apoptosis, angiogenesis, and metastasis. Targeting the transcription of such mRNA's by either dose or targeted delivery of an eIF4F complex inhibitor may decrease the transcription of oncogenes without the adverse effects of inhibiting all cap depended transcription. The role of eIF4E in translation is discussed further in 4.1 Introduction to eIF4e/eIF4g.

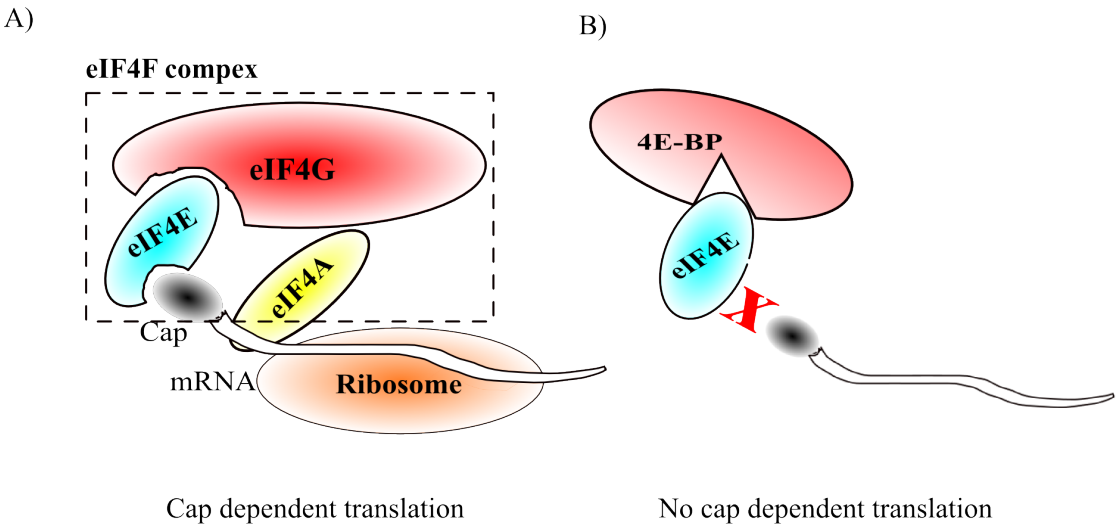


Figure 10 – Schematic illustrating the importance of the eIF4e/eIF4g binding event in translation. A) complex formed for cap dependent translation to occur. B) complex formed which limits cap dependent translation.

Chapter 2 : Materials and Methods

2.1 Materials

2.1.1 Bacterial Strains and Vectors

Competent DH5α *Escherichia coli (E. coli)* cells (Invitrogen, genotype: *F⁻ endAI hsdRJ7 (r^-k, Mk^+) supE44 thi-I λ⁻ recAl gyrA96 relAI deoR Δ(lacZYA-argF)-U169 Φ80dlacZΔM15)*[170] were used for propagation of ligand insert-vector construct DNA and plasmid amplification.

The *E. coli* expression strains BL21 (DE3) pLysS (*F⁻ ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3) pLysS(cm^R)*) Star, Gold and Rosetta 2 (Statagene) encoding T7 RNA polymerase and chloramphenicol resistance and were used for over-expression of protein from plasmid DNA. *E. coli* ER2738 cells (NEB, F′*$proA^+B^+$ $lacI^q$ Δ(lacZ)M15 zzf::Tn*10(Tet^R)*/ fhuA2 glnV Δ(lac-proAB) thi-1 Δ(hsdS-mcrB)5*) were used for the amplification of phage.

Vectors used throughout this study were: pGEX-6p-2 (GE Healthcare), with ampicillin resistance; pET-28b-Prescission and pET-28a (Novagen), with kanamycin resistance. The fusion tag SUMO was added to pET-28a and GFP to pET-28b-Prescission, by H. Jenkins (University of Leeds). All plasmid amplification and protein expression strains were stored as glycerol stocks at -80 °C.

2.1.2 cDNA and primers

The cDNA templates for proteins p300 and eIF4E were purchased from Source Bioscience and the HIF-1α cDNA template was purchased from Origene. All primers (

Table 2) were purchased from Sigma Aldrich.

Table 2 – The names and sequences of all the primers used in this study (all 5'-3')

Primer Name	**Sequence**
p300 (330-430) forward	TATTGGATCCGCTGATCCAGAGAAGCG
P300 (340-430) reverse	ATTCTCGAGTTAAGCATTTTTGAGGGGGAG
p300 (323-423) forward	TATTGGATCCATGGGTTCTGGAGCAC
P300 (323-423) reverse	AATA CTCGAG TTA CTTATCACCAGCATTTTTG
eIF4E forward	TATTGGATCCAAACATCCCCTACAGAACAGAT G
eIF4E reverse	TATTCTCGAGTTAAACACAAACCTATTTTTAGT GGTGG
Full HIF-1α forward	TATTGGTACCTCTGATTTAGCATGTAGACTGCT G
Full HIF-1α reverse	GCTCACCATACCGTAACCACCGTTAACTTGAT CCAAAGCTCTGAG
HIF-1α helix 2 forward	GAAGTGAACGCGGCGGCGGGTGGTTACGGTA TGGTGAGC
HIF-1α helix 2 reverse	TATTGGTACCCAGCTGACCAGCTATGATTGCG AAGTGAACGCGGCG
HIF-1α helix 3 forward	CGCGCGCTGGATCAGGTGAACGCGGCGGGTG GTTACGGTATGGTGAGC
HIF-1α helix 3 reverse	TATTGGTACCGAAGAACTGCTGCGCGCGCTGG ATCAG
HIF-1α helix 2+3 forward	TATTGGTACCCAGCTGACCAGTTATGATTGTG
HIF-1α helix 2+3 reverse	GCTCACCATACCGTAACCACCGTTAACTTGAT CCAAAGCTCTGAG
eIF4G forward	TTTCTGCTGGGCTTTCAGGGTGGTTACGGTATG GTGAGC
eIF4G reverse	TATTGGTACCCGCTATGATCGCGAATTTCTGCT GGGCTTTCAG
YFP reverse	TATTGGATCCTTACTTGTACAGCTCTC
TEV-nHIF (HIF-1α/p300 fusion construct)	CTGTACTTCCAGGGTGGATCTGATTTAGCATG TAGACTGC
P300c-TEV (HIF-1α/p300 fusion construct)	CTGGAAGTACAGGTTTTCACCAGCATTTTGAG GGGGA
HIF-reverse (HIF-1α/p300 fusion construct)	TATTCTCGAGTTAGTTAATTGATCCAAAGCTC AG
p300 mutant H20A forward	TTTGCATGCTGCTAAGTGCCAGCGCCG
p300 mutant H20A reverse	AGGAGAACAAGCTGCTGC
p300 mutant L20M forward	GAAGAATGTCATGAACCACATGACACAC
p300 mutant L20M reverse	ATTGTGCGACAGTGGGGA
p300 mutant I71M forward	TCGACAAATCATGTCACACTGGAAG
p300 mutant I71M reverse	GAAGATGCACAGTGTGCC

2.1.3 Media

All cultures were grown in 2x YT media supplemented with the appropriate antibiotic. The antibiotics were added from stock solutions, (see Table 3 for stock solutions and working concentrations). Stock solutions of antibiotics were stored at -20 °C. 2xYT Media was prepared with 16 g tryptone soya broth, 10 g yeast extract and 5 g sodium chloride per litre of dd H_2O which was then autoclaved.

Table 3 – Concentrations of antibiotics used in this study, all diluted in ddH_2O

Antibiotic	Stock Concentration (mg/mL)	Working Concentration (µg/mL)
Kanamycin	50	50
Ampicillin	100	100
Chloramphenicol	34	34
Carbenicillin	100	100

2.1.4 General reagents

The reagents zinc sulphate, magnesium chloride, nickel sulphate, ethanol, sodium phosphate, iron chloride, cobalt chloride, nickel chloride, potassium phosphate, xylene cyanol, ammonium chloride, lactose, Tetramethylethylenediamine (TEMED), coomassie brilliant blue R-250, PIPES, chloramphenicol and aspartate were all purchased from Sigma-Aldrich. Sodium chloride, bromophenol blue, glycerol, calcium chloride, glucose, boric acid, methanol and acetic acid were all purchased from Fisher Scientific. Tryptone soya broth and yeast extract were purchased from Oxoid. Reagents Ficoll 400 and glutathione were purchased from Alfa Aesar. Agar, imidazole, manganite chloride and agarose were all purchased from Acros Organics. Copper chloride was purchased from Riedel-de-Haen (Seelze). Tris(hydroxymethyl)aminomethane (Tris), Ethylenediaminetetraacetic acid (EDTA), sodium dodecyl sulfate (SDS), Isopropyl β-D-1-thiogalactopyranoside (IPTG), kanamycin and ampicillin were all purchased from Melford. Sodium sulphate and magnesium sulphate were purchased from Glycon. Triton x100 and potassium chloride were purchased from Fluka.

Table 4 – Buffers commonly used in this study

Experiment	**Buffer**
p300 CD	40 mM sodium phosphate pH 7.5 100 mM sodium chloride 1 mM DTT 5 % glycerol
eIF4E CD	40 mM sodium phosphate pH 7.5 200 mM sodium chloride 1 mM DTT 5 % glycerol
Adhiron CD	40 mM sodium phosphate pH 7.5 100 mM sodium chloride 1 mM DTT 5 % glycerol
p300 fluorescence anisotropy	40 mM sodium phosphate pH 7.5 100 mM sodium chloride 1 mM DTT 5 % glycerol 0.1 % triton x100
eIF4E fluorescence anisotropy	40 mM sodium phosphate pH 7.5 200 mM sodium chloride 1 mM DTT 5 % glycerol 0.1 % triton x100
p300 phage display peptide fluorescence anisotropy	40 mM sodium phosphate pH 7.5 100 mM sodium chloride 1 mM DTT 5 % glycerol
p300 ITC	40 mM sodium phosphate pH 7.5 100 mM sodium chloride 5 % glycerol
Adhrion SPR	8 mM disodium hydrogen phosphate 1.5 mM potassium dihydrogen phosphate 3 mM potassium chloride 137 mM sodium chloride, pH 7.4,
NMR	10 mM Tris-HCl pH 6.9 50 mM sodium chloride 2 mM DTT 2% glycerol
eIF4e crystallography	20 mM Hepes pH 7.5 100 mM sodium chloride 1 mM DTT

	5 % glycerol
Initial crystal screens (p300 and HIF-1α/p300 fusion construct)	20 mM Tris pH 7.5 100 mM sodium chloride 1 mM DTT 5 % glycerol
Adhiron crystallography	20 mM Hepes pH 7.5 100 mM sodium chloride 1 mM DTT 5 % glycerol
Lysis buffer	20 mM TRIS pH 7.9 500 mM sodium chloride
TBS-T	50 mM Tris pH 7.6 150 mM NaCl 0.05% Tween 20
TBE running buffer	45 mM TRIS-borate pH 8.3 1 mM EDTA pH 6.8
SDS loading buffer	100 mM TRIS 4% (w/v) SDS 20 % (v/v) glycerol 200 mM DTT 0.2% (w/v) bromophenol blue
SDS running buffer	25 mM Tris pH 6.8 0.1% (w/v) SDS 190 mM glycine
DNA loading buffer	0.43% bromophenol blue 0.42% xylene cyanol FF 25% Ficoll 400
Transformation buffer	10 mM PIPES 15 mM $CaCl_2$, 250 mM KCl. 55 mM $MnCl_2$
p300 biotinylation	50 mM HEPES pH 7.5 150 mM sodium phosphate 5 mM calcium chloride

2.2 Methods

2.2.1 Recombinant DNA techniques

2.2.1.1 Touchdown PCR

All Touchdown PCR experiments were conducted in 0.2 mL thin-walled dome capped tubes (Alpha Laboratories). The total reaction mixture was 100 µL containing 10µL 10x *Pfu* Ultra II polymerase buffer (Stratagene), 10 µL 10x dNTP stock

solution (2.5 mM of each dNTP, invitrogen), 10 µL 10 µM forward primer, 10 µL 10 µM reverse primer, 2 µL PCR DNA template, 1 µL *Pfu* Ultra II DNA polymerase (Stratagene) and 57 µL water. PCR was conducted in a G-storm thermocycler (GRI Labcare), The PCR protocol is outlined in Table 5.

Table 5 – Protocol used for touchdown PCR

Step number	Number of cycles	Temperature (°C)	Time
1	1	96	2
2	20	95	0.5
		60*	0.5
		72	1 min/kb
3	20	95	0.5
		64	0.5
		72	1 min/kb
4	1	72	4

* decrease by 0.5 °C each cycle

PCR samples were purified by agarose gel electrophoresis and extracted using the QIAquick gel extraction kit (QIAGEN), following the manufacturers recommended protocol.

2.2.2 Agarose Gel Electrophoresis

1.4% (w/v) agarose gels were prepared with 0.5x TBE buffer containing 1x RedSafe Nucleic Acid Staining Solution (Invitrogen). Prior to loading 10x DNA loading buffer was added to the samples. Gels were run at 95 V in 0.5x TBE buffer until bands were sufficiently separated. 2-log DNA ladder (New England Biolabs) was included alongside the samples.

2.2.3 Restriction endonuclease digests

For diagnostic digests, a total reaction mixture of 10 µL was used and for large scale digests a total reaction mixture of 100 µL was used. All digests were conducted in

1.5 mL Eppendorf tubes. The large scale digests consisted of 50 µL DNA product, 4 µL of each appropriate restriction enzyme (20,000 units New England Biolabs), 10 µL Buffer 4 (New England Biolabs), 10 µL 10x BSA (New England Biolabs) and water (22 µL). The mixture was incubated at 37 °C overnight and then the DNA was purified by agarose gel electrophoresis and gel extracted according to the QIAquick Gel Extraction Kit Protocol (QIAGEN). Diagnostic digests were carried out using 8 µL DNA product, 0.5 µL of each appropriate restriction enzyme (New England Biolabs), 10 µL Buffer 4 (New England Biolabs), 10 µL 10x BSA (New England Biolabs). The mixture was incubated at 37 °C for one hour and then the DNA was purified by agarose gel electrophoresis and gel extracted using the QIAquick Gel Extraction Kit Protocol (QIAGEN).

2.2.4 Ligations

Ligations were conducted in 1.5 mL Eppendorf tubes and incubated at room temperate for 30 minutes. Each reaction contained 1 µL digested insert, 6 µL purified digested plasmid, 1 µL T4 DNA ligase (New England Biolabs) and 2 µL 5x ligase buffer (New England Biolabs).

2.2.5 Competent DH5α E. coli cell preparation

200 mL 2xYT was inoculated with DH5α *E. coli* cells from glycerol stock and incubated at 37 °C, with aeration at 100-110 rpm, for approximately 6 hrs. The temperature was decreased to 18 °C, with shaking at 100-110 rpm until OD_{600} = 0.6-0.7. The culture was then incubated on ice for 10 min, followed by centrifugation at 13,000 × g for 10 min at 4 °C. The pellet was resuspended in a total of 80 mL ice cold transformation buffer and incubated for 10 min on ice. The centrifugation was repeated. The pellet was resuspended in 18.6 mL ice cold transformation buffer and 1.4 mL Dimethyl sulfoxide (DMSO) were slowly added with continual stirring. The culture was incubated on ice for a further 10 min, aliquoted and frozen in liquid nitrogen and stored at −80 °C.

2.2.6 Transformations

100 µL of competent *E. coli* DH5α or BL21 expression strains were transformed with 10 µL of ligated product and incubated on ice for 30 min. For kanamycin and chloramphenicol-resistant transformations 400 µL 2xYT was then added to the transformation mixture and incubated at 37 °C for one hour with aeration. The cells were then centrifuged at 3500 × g for one minute, 400 µL of supernatant was then removed and the pellet resuspended. Transformed cells were incubated at 37 °C overnight on agar plates (31 g/L 2xYT media, 15 g agar) which contained appropriate antibiotic.

2.2.7 Colony PCR

Single Colonies from the ligation plate were picked and transferred to a 2xYT agar plate supplemented with appropriate antibiotic and incubated at 37 °C overnight to create a master plate. PCR templates of each picked colony were made by resuspending the remainder of each colony in separate 1.5 mL Eppendorf tubes containing 30 µL of water and then boiling them for five minutes. The PCR was carried out in 0.2 mL thin-walled dome-capped PCR tubes (Alpha Laboratories) in the G storm-thermocycler (GRI Labcare). Each tube contained a total reaction mixture of 30 µL consisting of 3 µL 10x thermopol buffer (New England Biolabs), 3 µL 10x dNTP stock solution (2.5 mM each dNTP), 3 µL 10 µM vector primer, 3 µL 10 µM insert primer, 3 µL PCR template, 0.5 µL *Taq* polymerase (New England Biolabs) and 14.5 µL water was used. The PCR protocol used is outlined in Table 4. The resulting PCR products were run on an agarose gel and those which showed a single band of the correct size were presumed to be positive. The presumed positive colonies were picked from the master plate and overnight cultures were grown from which plasmids can be extracted. Presumed positive colonies were confirmed by sequencing (cogenics, UK).

Table 6 – Protocol used for single colony PCR

Step Number	Number of cycle	Temperature (°C)	Time (min)
1	1	96	2
2	25	95	0.5
		54	0.5
		72	1/kb

3	1	72	4

2.2.8 p300 mutant cloning

p300 mutants, H20A, L47M and I17M were cloned by Quickchange from the GST-p300 construct. Quickchange was performed as per the Q5 site directed mutagenesis kit protocol. Primers used are outlined in

Table 2.

2.2.9 Overnight cultures

Cultures were initiated by inoculation of 2xYT with appropriate antibiotic from either a single colony from a transformation plate or by a scrape from a glycerol stock. Starter cultures of 10 mL were grown in 50 mL falcon tubes and incubated at 37 °C with aeration overnight.

2.2.9 Plasmid DNA extraction and glycerol stock preparation

Plasmid DNA was extracted from 10 mL of overnight culture using QIAprep spin miniprep kit (QIAGEN) as per the manufacturers handbook. For large scale plasmid DNA extraction 3 mL overnight culture was used to inoculate 300 mL 2xYT, which were incubated at 37 °C overnight. The plasmid DNA was extracted using QIAprep spin maxiprep kit (QIAGEN) as per the manufacturers handbook. The extracted DNA samples were stored at -20 °C. Each new plasmid obtained from positive colonies was transformed into both DH5α and BL21 *E. coli* strains and glycerol stocks of the resulting cells were produced by the addition of 0.5 mL sterile 60 % glycerol to 1.2 mL of overnight culture before storing them at -80 °C.

2.2.10 Sanger DNA sequencing

Each new plasmid made was sequenced prior to expression and solubility trials to test for possible mutations introduced during the ampification PCR step. 1 mL of 2xYT agar with appropriate antibiotic was set in 2 mL sterile screw cap tubes. Stabs were made by pipetting of 5 µL overnight culture on to the agar surface and then stabbing a hole into the agar with the pipette tip. The bacterial stabs were sent to Cogenics (Essex) to be sequenced.

2.3 Protein expression and purification

2.3.1 General techniques

2.3.1.1 Protein Expression Trials

For small-scale over-expression trials, 1 mL of overnight culture was used to inoculate 10 mL 2xYT (supplemented with appropriate antibiotic) in a 50 mL falcon tube. Cultures were incubated at 37 °C with aeration until an OD_{600} 0.6-0.8 was reached. Protein expression was induced by addition of sterile IPTG to the culture at a final concentration of 0.1-0.4 mM (outlined in appropriate results section for protein expression trials and 2.3.2 Specific Protein Expression) and the culture was then incubated at 18 °C with aeration overnight. Cells were harvested by centrifugation at 3500 × g at 4 °C for 20 minutes. Cell pellets were resuspended in Lysis buffer. 5 U DNase I and 2 mM $MgCl_2$ were added to the suspensions prior to lysing the cells. Cells were lysed by 5 cycles of freeze/thaw. The lysate was

centrifuged at 13000 × g for 25 minutes at 4 °C. Both the pellet and supernatant were analysed by SDS-PAGE.

2.3.1.2 OD_{600} measurement

The Optical Density (OD) of cell cultures was read with a spectrophotometer (Helious β) set to 600 nm by using 1 mL of culture in a plastic cuvette (brand) after the spectrophotometer was zeroed against a blank sample (1 mL of un-inoculated media).

2.3.1.3 Nickel affinity chromatography

For purification of protein with a His tag from cultures ≤ 1 L a 1 mL Ni^{2+} sepharose HisTrap HP column (GE Healthcare) was used and for purification of protein from cultures > 1 L a 5 mL Ni^{2+} sepharose HisTrap HP column (GE Healthcare) was used. Columns were attached to a peristaltic pump at room temperature. The maximum flow rate used was 1 mL/min for the 1 mL column and 5 mL min for the 5 mL column. Columns were washed with 5 column volumes of water, charged with 0.1 M $NiSO_4$ and equilibrated with 5 column volumes of Lysis buffer before loading the filtered supernatant. The filtered supernatant was passed through the column and the flow through was collected. The column was then washed with 5 column volumes of Lysis buffer. Elution was carried out by increasing concentration of imidazole in Lysis buffer (50 mM-200 mM). All wash fractions were collected and analysed by SDS-PAGE. Columns were stored in 20% ethanol.

2.3.1.4 Glutathione affinity chromatography

Proteins fused to a glutathione S-transferase (GST) tag were purified with approximately 10 mL Glutathione Superflow Resin (Generon) packed in a free-flow gravity column, attached to a peristaltic pump at room temperature. The column was washed with 5 column volumes of water and equilibrated with 5 column volumes of Lysis buffer. The lysate was then added to the column and mixed with the beads. The mixture was left for 10 min to allow binding. The mixture was then allowed to flow through the column and the flow through was collected. The column was washed with 5 column volumes of Lysis Buffer, then with 5 column volumes of Lysis buffer supplemented with 1 M sodium chloride, followed with a further wash with 5

column volumes of Lysis buffer. The protein was eluted with 20 mM glutathione pH 7.0 (made up with Lysis buffer). All wash fractions were collected and analysed by SDS-PAGE. Columns were stored in 20% ethanol.

2.3.1.5 Size exclusion chromatography

Size exclusion chromatography was performed on a Superdex 75 (26/60) column (GE healthcare) attached to an Akta prime system at 4 °C. The absorbance of the sample was monitored at 280 nm throughout. The column, which is stored in 20% ethanol, was washed with water and equilibrated with appropriate buffer. 5 mL of protein sample (after affinity chromatography) was loaded onto the column using a 5 mL injection loop. The method for size exclusion chromatography is summarized in Table 7. Upon completion of the method, the column was washed with water and stored in 20% ethanol. Each fraction from within the peak of the UV trace was analysed by SDS-PAGE.

Table 7 – Method used for size exclusion chromatography

Breakpoint	**%B**	**Rate (ml/min)**	**Fraction size**	**Buffer valve**	**Injection volume**	**Auto zero**	**Event mark**
0	0	2	0	1	Load	No	No
10	0	2	0	1	Inject	Yes	Yes
20	0	2	0	1	Load	No	Yes
110	0	2	3	1	Load	No	Yes
320	0	2	0	1	Load	No	No

2.3.1.6 Protein concentration

Protein was concentrated using an Amicon Ultra centrifugal filter (Millipore), by centrifugation at 2770x g until the desired concentration was reached. The concentration of protein was routinely measured using a protein assay reagent dye (Bio-Rad). 200 µL of dye was added to 800 µL of water and then either 20 µL of buffer (blank) or protein sample was added to this mix. The solutions were placed in

a plastic cuvette (Brand) and the absorbance measured at 595 nm in a spectrophotometer (Helious β). Firstly the spectrophotometer was zeroed against the blank then the absorbance of the sample was measured. In the assay an absorbance of 1 signifies a concentration of 1 mg/mL. The assay is accurate for readings between 0.2-0.8, if outside this range the volume of protein sample added should be adjusted accordingly.

2.3.1.7 Protease expression

Both SUMO and PreScission protease were routinely produced in house (glycerol stock donated by Thomas Edwards, University of Leeds). 1 L of 2xYT was inoculated with 10 mL of each overnight culture. Cells were grown 37 °C until an OD_{600} 0.6-0.8 was reached, induced with 0.5 mM IPTG and incubated for 18 hours. Cells were harvested by centrifugation, lysed by sonication; soluble protein was separated from the insoluble fraction by centrifugation. SUMO-protease was purified by nickel affinity chromatography and PreScission-protease was purified by glutathione affinity chromatography.

2.3.2 Specific Protein expression

2.3.2.1 p300 and HIF-1α/p300 fusion construct protein expression

10 mL BL21 Gold cells containing GST-p300/GST-HIF-1α/p300 fusion construct overnight culture were used to inoculate 1 L of 2xYT (supplemented with ampicillin) in a 2 L baffled flask. Cells were grown to OD_{600} = 0.6-0.8, induced with 0.1 mM IPTG supplemented with 50 μM zinc sulphate and incubated for 18 hours at 18 °C. Cells were harvested by 20 minute centrifugation (3500 × g at 4 °C). Cell pellets were resuspended in Lysis buffer, 5 U DNase I and 2 mM $MgCl_2$ were added to the suspensions prior to lysing the cells. Cells were disrupted by sonication (soniprep 150, 200 mm probe) on ice in a glass beaker, 3 cycles of sonication, 20 seconds on 40 seconds off. Then the soluble protein was separated from insoluble protein by centrifugation (13000 × g for 25 minutes at 4 °C). The supernatant was decanted and filtered through a 0.45 μm Amicon syringe filter. GST-p300/GST-HIF-1α/p300

fusion construct was purified by glutathione affinity chromatography. The GST was cleaved by PreScission protease (1 mg of protease per 20 mg of GST-p300, incubated for 16 hours at 4°C) and separated from p300 and HIF-1α/p300 fusion construct by size exclusion chromatography.

p300 mutants (H20A, L47M and I71M) were all purified in the same way as wild-type p300.

2.3.2.2 ^{15}N-p300 protein expression

^{15}N-GST-p300 was purified after autoinduction with a glucose to lactose ratio 1:4; the full media recipe is outlined in Table 8. The cells were grown at 37 °C for four hours and then 20 °C for 40 hours. Cells were harvested and the protein purified in the same way as native p300.

Table 8 – Composition of the media used to express ^{15}N-p300

100 mL	**10x M9 medium** 60 g Na_2HPO_4 /L 30 g KH_2PO_4 /L 5 g NaCl /L 5 g $N^{15}H_4Cl$
10 mL	**100x Trace elements** 5 g EDTA /L 0.83 g $FeCl_3.6H_2O$ /L 84 mg $ZnCl_2$ /L 13 mg $CuCl_2.2H_2O$ /L 10 mg $CoCl_2.6H_2O$ /L 10 mg H_3BO_3 /L 1.6 mg $MnCl_2.6H_2O$ /L
20 mL	**50x 5052** 25 % glycerol (w/v) 2.5 % glucose (w/v) 10 % lactose (w/v)
1 mL	1 M $MgSO_4$
0.3 mL	1 M $CaCl_2$
1 mL	Biotin (1 mg/mL)

1 mL	Thiamin (1 mg/mL)
1 mL	1 M ampicillin

2.3.2.3 Biotinylation of p300

Biotin-depsipeptide coupling to the N-terminal glycine of p300 was catalysed by sortase A[171]. The reaction consisted of 100 μM p300, 300 μM biotin-depsipeptide and 20 μM sortase (both produced in house by Dan Williamson, University of Leeds), the reaction was incubated at 37 °C with agitation for 3.5 hours. The biotin-p300 was separated from the unreacted biotin-depsipeptide by dialysis and from the sortase A by nickel affinity chromatography (the sortase A has a His tag). The biotinylation was confirmed by western blot and mass spectrometry.

2.3.2.4 eIF4E protein expression

Overnight cultures were grown in minimal media (Table 9) and 400 μL of this overnight culture was used to inoculate 400 mL auto-induction media (

Table 10). The culture was grown for 4 hours at 37 °C with aeration, then reduced to 20 °C with aeration until no more cell growth was apparent (no more increase in OD_{600}).

Table 9 – Autoinduction overnight media

Solution	Composition	Volume (µL)
H_2O	H_2O	4775
$MgSO_4$	1 M $MgSO_4$	10
1000x trace elements	50 mM $FeCl_3$ 20 mM $CaCl_2$ 10 mM $MnCl_2$ 2 mM $CoCl_2$ 2 mM $CuCl_2$ 2 mM $NiCl_2$ 2 mM Na_2MoO_4 2 mM H_3BO_3	1
40 % glucose	40% glucose (w/v)	62.5
25% aspartate	25% aspartate (w/v)	50
50x M	1.25 M Na_2HPO_4 1.25 M KH_2PO_4 2.5 M NH_4Cl 0.25 M Na_2SO_4	100

Table 10 – Autoinduction expression media

Solution	Composition	Volume (mL)
ZY media	5 g/L yeast extract 10 g/L tryptone	383
$MgSO_4$	1 M $MgSO_4$	0.8
1000x trace elements	See Table 5	1
50x 5052	25% glycerol (w/v) 2.5% glucose (w/v) 10% lactose (w/v)	8
50x M	See Table 9	8

Cells were harvested by centrifugation (20 minutes 3500 × g at 4°C) and resuspended in 15 mL Lysis buffer per 400 mL culture. 5 U DNase I and 2 mM $MgCl_2$ were added to the suspensions prior to lysing the cells. Cells were disrupted by sonication, 3 cycles of sonication, 20 seconds on 40 seconds off. Then the soluble protein was separated from insoluble protein by centrifugation (13000 × g for 25 minutes at 4°C). The supernatant was decanted and filtered through a 0.45 µm Amicon syringe filter. SUMO-eIF4E was purified by nickel affinity chromatography

and then by size exclusion chromatography. The SUMO was then cleaved by SUMO protease and separated by a second a Ni^{2+} column after dialysis. eIF4E was then further purified by a second size exclusion chromatography.

2.3.2.5 Adhirons

Adhirons were in pET-11 and expressedin Rosetta 2 cells. 10 mL overnight culture was used to inoculate 1 L of 2xYT in a 2 L baffled flask, cells were were grown to OD_{600} = 0.6-0.8 and induced with 0.1 mM IPTG and incubated for 18 hours at 18 °C. Cells were harvested by 20 minute centrifugation (3500 × g at 4°C). Cell pellets were resuspended in Lysis bufer. 5 U DNase I and 2 mM $MgCl_2$ were added to the suspensions prior to lysing the cells. Cells were disrupted by sonication, 3 cycles of sonication, 20 seconds on 40 seconds off. Then the soluble protein was separated from insoluble protein by centrifugation (13000 × g for 25 minutes at 4°C). The supernatant was decanted and filtered through a 0.45 μm Amicon syringe filter. The adhirons were purified by nickel affinity chromatography and then by size exclusion chromatography.

2.3.3 Protein characterization

2.3.3.1 SDS PAGE gel electrophoresis

Sodium dodecyl sulphate polyacrylamide (SDS-PAGE) gel electrophoresis was performed based on the method described by Lamemmli[172]. 15 % gels were prepared as outlined in

Table 11. Samples were mixed with loading buffer in a 1:1 ratio and boiled for 10 minutes and were electrophoreses through SDS PAGE gels in SDS running buffer at 35mA. The broad range molecular marker (NEB 2-212 kDa) was included in each gel. Proteins were visualized using coomassie blue stain (45% (v/v) methanol, 10% (v/v) acetic acid, 0.25% (w/v) coomassie brilliant blue R-250) and destained in 30% (v/v) methanol, 10% (v/v) acetic acid.

Table 11 – Composition of a 15 % SDS PAGE gel

Solution	15% Resolving gel	Stacking gel
H_2O	3.3 mL	2.1 mL
30% bis-acrylamide (37.5:1)	4 mL	0.5 mL
1.5 m Tris (pH 8.8)	2.5 mL	-
1 M Tris (pH 6.8)	-	0.38 mL
10% SDS	100 µL	30 µL
10% APS	100 µL	30 µL
TEMED	10 µL	10 µL

2.3.3.2 Mass spectrometry

To confirm the mass of each protein 1 mg/mL of protein was analysed by high mass liquid chromatography mass spectrometry (LC-MS). The column was reverse phase; the method involved gradient elution of water in acetonitrile with 0.1% formic acid. Analysis was conducted on a Waters-Micromass ZMD spectrometer using electrospray ionization.

2.3.3.3 Circular Dichroism (CD) Spectroscopy

Spectra were recorded on a chirascan circular dichroism spectropolarimeter (Applied Photophysics), at 20 °C, using 1 mm cells and a scan speed of 5 nm/min. The spectra were averaged over 3 repeats with a buffer baseline subtracted. Protein concentrations of approximately 0.2 mg/mL were used (although the exact concentration was used to allow determination of MME). The helical content of each peptide was determined from the mean residue ellipticity at 222 nm, [θ] (deg cm^2 $dmol^{-1}$).

2.3.4 Biophysical Assays

2.3.4.1 Fluorescence Anisotropy

2.3.4.1.1 Direct binding

p300 or eIF4E protein was serially diluted in buffer (different buffers are used for different protein targets, outlined in Table 4) labeled peptide (40 nM) (FITC-HIF-1α or FITC-eIF4G) was added, and the plates were then incubated for 30 minutes at room temperature. Each experiment was run in triplicate and the fluorescence anisotropy measured using a EnVision 2103 MultiLabel plate reader (Perkin Elmer) with excitation at 480 nm and emission at 535 nm (5 nm bandwidths). In parallel, a control experiment was performed in which no labeled peptide was added and the volume made up with additional buffer, this blank was subracted from the raw data each of the three repeats. The intensity was calculated for each point using Equation 1 and used to calculated anisotropy using Equation 2. From a plot of anisotropy against protein concentration the minimum and maximum anisotropies were obtained using a logistic sigmoidal fit in OriginPro 8.6. This allowed the conversion to fraction bound (Equation 3). The data were then fitted using Equation 4 in OriginPro 8.6 to determine the dissociation constant, K_d.

$I = (2PG) + S$ **(Equation 1)**

$R = (S - PG) / I$ **(Equation 2)**

$Lb = (R - R_{min}) / ((\lambda(R_{max} – R)) + R – R_{min})$ **(Equation 3)**

$y=((Kd +x+[FL])-\sqrt{((Kd +x+[FL])2 -4x[FL]))/2}$ **(Equation 4)**

R= anisotropy, I= total intensity, P=perpendicular intensity, S= parallel intensity, G = an instrument factor set to 1, Lb = fraction ligand bound, λ = Ibound/Iunbound = 1, [FL] = concentration of fluorescent peptide, K_d = dissociation constant, y = Lb multiplied by [FL], x = protein concentration

2.3.4.2 Competition

Unlabeled peptide (HIF-1α/eIF4G) or compound was serially diluted across a 384 well plate in buffer (Table 4) and 40 nM labeled peptide and protein were added sequentially (0.1 μM p300 or 1 μM eIF4E). The plates were then incubated for 30 minutes at room temperature. Each experiment was run in triplicate and the fluorescence anisotropy measured using an EnVision 2103 MultiLabel plate reader (Perkin Elmer) with excitation at 480 nm and emission at 535 nm (5 nm bandwidths). A control experiment was performed in which no labelled peptide was added and the volume made up with additional buffer, this blank was subtracted from the raw data each of the three repeats. Intensity and anisotropy were calculated as above using Eq. 1 and Eq. 2 respectively. Plots of anisotropy against unlabeled peptide were fitted to a logistic sigmoidal dose response model to determine an IC_{50}.

2.3.4.2 Isothermal Titration Calorimetry

ITC experiments were conducted at 25 °C in 40 mM sodium phosphate, 100 mM sodium chloride and 5% glycerol. 10 μM p300 was present in the cell and 100 μM HIF-1α in the syringe. One injection of 10 μL for 20 seconds of HIF-1α was made every 300 seconds for 30 injections. Buffer was also titrated in to protein to give a baseline integral. This baseline was then subtracted to give a baseline-subtracted thermogram. This thermogram was then modeled by one-site interaction ITC data analysis, performed with Origin 5.0 software (MicroCal Inc).

2.3.4.3 SPR

1 pM biotin-p300 was isolated on to one flow cell of a streptavidin chip (120 response units), at a flow rate of 5 uL/min, while the other flow cell was left unfunctionalised. 5 concentrations of Ad34 (100 nM-1 μM) were tested. Each concentration was flowed over both the functionalised and the unfunctionalised flow cells at 40 μL/min and the on- and off-rates were calculated using the Biacore software. The on- and off-rates were used to calculate the K_d.

$d[AB]/-dt = k_a[A][B] \bullet d[AB]/dt = k_d[AB]$ **(Equation 5)**

$K_d = k_d/k_a$ **(Equation 6)**

2.3.4.4 BLitz

The BLitz^{TM} (ForteBio) dip and read Ni biosensors were used to estimate the affinity of p300 (10 μM) binding to immobilized Adhirons (1 μM) and a global fit was used to calculate the binding affinities using the advanced kinetic software.

2.4 Phage Display Experiments

2.4.1 Peptide phage display

The biotin-p300/biotin-eIF4E (10 pM) was mixed with each phage library (10 μL of $1x10^{13}$ plaque forming units (pfu) /mL), incubated at room temperature for one hour, then isolated on to a streptavidin plate via the biotin tag (ten minute incubation), additional biotin (0.1 mM final concentration) was added to each well to block the unbound streptavidin (5 minute incubation at room temperature). The wells were then emptied and washed 3 times with high and low salt TBS-T (0.5 M NaCl and 1.5 M NaCl respectively). The phage were then eluted (500 μM HIF-1α/500 μM eIF4G). To re- amplify the phage, ER2783 cells were grown to mid- log phase (OD_{600} = 0.6-0.8) and the elutes were added (one elute condition per culture). The cultures were then grown at 37 °C with vigorous shaking (300 rpm) for 4.5 hours, although the first 10 minutes of this growth period the shaking was set to 80 rpm. To isolate phage the cultures were then pelleted and the supernatant was mixed with a chilled 20% PEG 8000/2.5 M NaCl solution and the mixture was centrifuged at 13,000 rpm for 20 minutes (at 20 °C). The pellet was resusspended in 1 ml TBS, spun down at 13,000 rpm, then the chilled 20 % PEG 8000/2.5 M NaCl solution was again added to the supernatant and incubated on ice for 30 minutes. The mixture was again centrifuged at 13,000 rpm for 5 minutes and the pellet resuspended in TBS. 10^{11} pfu (as calculated by a phage titer as per the NEB phage titer protocol) were added to start the next round of panning, in total three panning rounds were completed.

2.4.1.1 Enrichment ELISA

Streptavidin plates coated in biotin-p300/biotin-eIF4E (5 μg/mL) were incubated at room temperature for 30 minutes. The plates were washed with TBS-T to remove

excess target. 1 x 10^{10} pfu of re-amplified phage from each panning elute were added and incubated for one hour at room temperature, (and also added to an uncoated well as a control). The plates were washed with TBS-T and anti M13-HRP antibody was added and incubated for one hour, washed with TBS-T, developed with 100 µl/well TMB. Plate was read at 370 nm.

2.4.1.2 Peptide Phage Display Sequencing

DNA was isolated from pan elute 1 and 3 and from the unpanned library and PCR was used to amplify the DNA. Blunt end repair of the resultant ds-DNA was conducted using Illumina Paired-end DNA sample preparation kit. The primers with the unique bar codes (Illumina adaptors) were ligated to each fragment and the fragments with the adaptors were amplified by a second round of PCR. This resulted in 12 different pools which each had a unique bar code; the unpanned 7mer and 12mer libraries were also sequenced to assess the presence of any propagation related clones at the start of the experiment. Sequencing was performed by the Centre for Genomic Research, University of Liverpool. The Matlab analysis was performed (Jonathan Stott, AstraZeneca) using scripts based on published code[173] and modified for correctness and the specific sequences used. Modifications used were:

- Cope with smaller files by not discarding partial blocks
- Allow for variable length adaptor sequences, and unequal adaptor sequences
- Allow for variable file tags (i.e. not just Illumina)
- Allow for variable length peptide libraries (tested with 7-mer and 12-mer)

Scripts were run over all 'pure' 7-mer and 12-mer libraries. Locus specific sequences rather than full primers were used and the quality control cut off was A (quality score of 32, the original scripts used a quality control of 15[173].

2.4.2 Adhiron phage display

p300 was expressed and biotinylated as described above. Biotin-p300 was added and incubated on pre-blocked streptavidin plate, the plate was then washed using a KingFisher robotic platform (ThermoFisher) and 10^{12} cfu of the pre-panned phage

library was added and incubated for 2.5 h with shaking. Wells were washed ten times and eluted with 100 µL 0.2 M glycine (pH 2.2) for ten minutes neutralized with 15 µL 1 M Tris-HCl (pH 9.1), further eluted with triethylamine 100 mM for 6 min, and neutralised with 1 M Tris–HCl (pH 7). Eluted phage were used to infect ER2738 cells for 1 h at 37 °C and 90 rpm then plated onto LB agar plates supplemented carbenicillin and grown overnight. All colonies were scraped into 5 mL of 2xYT with carbenicillin and 1 x 10^9 M13K07 helper phage were added. After an overnight incubation phage were precipitated with 4 % PEG 8000, 0.3 M NaCl and resuspended in 1 ml of 10 mM Tris, pH 8.0, 1 mM EDTA (TE buffer). 2 µL phage suspension was used for the second round panning using streptavidin magnetic beads as opposed to streptavidin plates (Invitrogen); otherwise the second pan was conducted in the same way as the first pan. The third pan was conducted using neutravidin high binding capacity plates (Pierce). During the fourth and final pan 50 µM of HIF-1α peptide was added as competitor before elution. After final pan colonies were picked, an ELISA was conducted to select positive clones (in the same way as the enrichment ELISA) which were sent for Sanger sequencing

2.5 Structural analysis

2.5.1 Crystallography

2.5.1.1 General techniques

2.5.1.1.1 Factorials

Initial factorial screens were set up routinely using the commercially available screens:

Hampton Research: Crystal Screen I and II, Index, Salt RX

Emerald Biosystems: Wizard I, II, III and IV

Molecular Dimensions: Morpheus, Midas, Pact

Screens were either set up manually or using a Douglas Instruments Oryx 6 plate loader/NT8 robotics system. Typically drops of 0.2 µL of protein in a 1:1 drop ratio with mother liquor were used in 96 well MRC 2-drop sitting drop plates (Molecular

Dimensions). Wells were filled with 80 µL mother liquor and sealed with Viewseal pressure adhesive clear seals and incubated at 4 °C, 18 °C and 25 °C.

2.5.1.1.2 Optimisation

Initial hits from the factorial screens were optimised in 24 well hanging drop plates, typically with 1 µL protein and 1 µL mother liquor (although this ratio was varied). Initial optimisations centred on the variation of precipitant and pH.

2.5.1.1.3 Cryo cooling

Crystals were picked using a nylon loop mounded on a cryo pin (Hampton Research) and submerged in to 1 µL of mother liquor + appropriate cryo protectant. Crystals were then immediately immersed in liquid nitrogen for storage.

Data collection

All data was collected at Diamond light source at various beamlines.

2.5.1.2 Apo_eIF4E crystallography

eIF4E was purified into buffer 20 mM Hepes pH 7.5, 100 mM sodium chloride, 1 mM DTT and 5 % glycerol. Crystals grew overnight in the condition 17.5% PEG 4000, 0.2 M ammonium citrate, 3 % trehalose at 25 °C. Crystals were cryo-protected with 20 % glycerol. Crystals diffracted to 2.4 Å at diamond light source. The structure was solved by molecular replacement using PDB:2W97[174], using Phaser[175] and the structure was then refined using Phenix[176] and Refmac[177]. Structure determination was primarily performed by Thomas Edward (University of Leeds).

2.5.1.3 Adhiron crystallography

Ad34 was concentrated to 10 mg/mL and crystalized in the condition 0.8 M di-sodium succinate pH 7. The crystals grew overnight at 18 °C. They were picked and cryo protected with 20 % glycerol. The structure was solved using molecular replacement using PDB:4N6U[122]. Initially Ad34 was subjected to molecular replacement with no loops present using the program Balbes[178], and Buccaneer[179] was used to rebuild the loops. The structure was then refined using Phenix[176] and Refmac[177]. Structure determination was primarily performed by Thomas Edward (University of Leeds).

2.5.2 NMR

The NMR experiments were all conducted in 10 mM Tris-HCl pH 6.9, 50 mM NaCl, 2 mM DTT and 2% glycerol. PDDP1 peptide (300-750 μM) was titrated into 230 μM ^{15}N-p300. Data were collected on a 600 MHz Agilent NMR system at 25 °C and. The size of the shift for each peak was measured and those with the largest shift were mapped on to p300 (assignments from BMRB: 6268[180]). Analysis was conducted using CCPN Analysis software.

2.5.3 SAXS

p300, co-purified HIF-1α/p300 (co-purificatoin was completed by incubation of HIF-1α and p300 in a 1.5:1 molar ratio for 30 minutes on ice and purification by size exclusion chromatography) and HIF-1α/p300 fusion construct were purified in to the buffer 20 mM Tris pH 7.5, 100 mM sodium chloride, 1 mM DTT and 5 % glycerol, various concentrations were tested (outlined in the HIF-1α/p300 results section). SAXS experiments were conduced at Stanford Synchrotron Radiation Lightsource. Data was analysed using ATSAS program suite.

2.5.4 Docking

Docking was performing by George Burlsem (University of Leeds) using HADDOCK (High Ambiguity Driven protein−protein Docking)[181]. Using the NMR structure of p300 (PDN:1L8C[182]) and the crystal structure of Ad34. The restraint used was that the binding was mediated through the loop regions.

Chapter 3 : HIF-1α/p300

3.1 Introduction to HIF-1α/p300

3.1.1 Hypoxic response

Hypoxia is an almost universal hallmark of solid tumours; the ability to adapt to hypoxic environments is crucial to their growth and survival[155, 183]. In the main, this adaption is mediated by transcriptional activation of genes that facilitate short-term adaptive mechanisms (e.g. increased vascular permeability, vasodilatation, glucose transport, switch to anaerobic metabolism), as well as long-term adaptive mechanisms (e.g. angiogenesis)[184-187]. This coordinated homeostatic response is mediated in large part through the activation of the transcription factor hypoxia-inducible factor 1 (HIF-1). HIF-1 is responsible for the activation/transcription of >100 genes which are crucial for the cells adaptation to hypoxia, including oncogenes and the inactivation of tumor suppressor genes. Tumour hypoxia and overexpression of HIF-1 have been associated with resistance to some therapies, increased risk of invasion and metastasis, and poorer outcome in malignancies[188]. HIF-1 activity in tumors is dependent upon availability of the HIF-1α subunit, the levels of which increase under hypoxic conditions (as discussed in Chapter 1 : Introduction). The near universality of hypoxia in human tumors and the centrality of the HIF-1 pathway in adapting to the hypoxic environment suggest that inhibition of the HIF-1 pathway may have therapeutic utility as an antitumor strategy. Inhibition of the function of HIF-1 in tumors may reduce angiogenesis and thereby contribute directly to tumour cell death[189].

HIF-1 is a heterodimer consisting of two subunits; an oxygen-sensitive HIF-1α and a constitutively expressed HIF-1β, both subunits are members of the basic helix–loop–helix (bHLH) proteins of the PER-ARNT-single-minded protein (SIM) (PAS) family of transcription factors[190]. The regulation region of HIF-1α is dependent on the Oxygen Dependent Degradation Domain (ODDD) and two transactivation domains: the N-terminal transactivation domain or N-TAD and the C-terminal transactivation domain or C-TAD[191]; located towards the C-terminus. The C-TAD is involved in modulating the transcriptional activation of HIF-1α under hypoxic conditions, in contrast to the N-TAD, which is involved in the stabilisation of HIF-1α. The N

terminal region of HIF-1α has a basic helix-loop-helix (HLH) domain and enables binding of HIF-1α to the hypoxia response elements (HRE), located in the enhancer and promoter regions of various genes[143]. The domain structure of HIF-1α is illustrated in Figure 11.

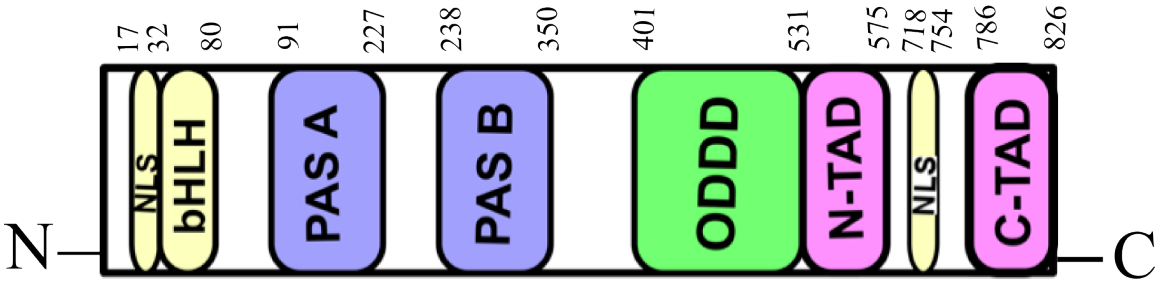


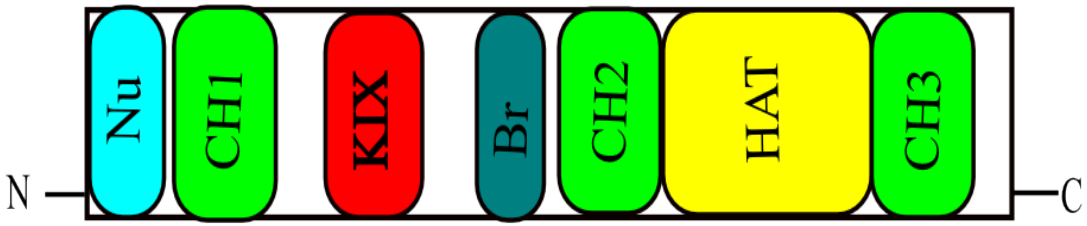
Figure 11 - Cartoon representation of the domain structure of HIF-1α

The stability and transcriptional activity of HIF-1α increases under hypoxic conditions. Regarding stability: HIF-1α is continuously expressed at a low level in the cell, however in normoxic conditions HIF-1α is rapidly degraded, most prominently via a Von Hippel–Lindau tumor suppressor (pVHL)-mediated ubiquitin-proteasome pathway[143]. This process is initiated by the hydroxylation of two proline residues within the ODDD by the enzyme prolyl hydroxylase-domain protein (PHD)[159]. Upon hydroxylation, pVHL is able to bind to HIF-1α, which subsequently allows the recruitment of the E3 ligase complex. This allows the attachment of ubiquitin via the lysine residues of HIF-1α, thereby marking HIF-1α for degradation. Due to the rapid nature of this pathway HIF-1α has a half life of less than 5 minutes in normoxic conditions; resulting in no detectable protein in normoxic cells[160]. However, in hypoxic environments oxygen availability is limited, therefore hydroxylation of the prolines is inhibited, resulting in an increased level of HIF-1α. Although the most prominent, the pVHL pathway is not the only pathway controlling levels of HIF-1α. Another mechanism is the recruitment of the murine double minute 2 (Mdm2) ubiquitin-protein ligase to HIF-1α by binding of the tumor suppressor p53, which results in a decrease in HIF-1α levels by promoting Mdm2-mediated ubiquitination and proteasomal degradation of HIF-1α[192]. In addition, Hsp90 interacts directly with HIF-1α and has been suggested to promote a conformational change in HIF-1α, which leads to inhibition of the dimerization with HIF-1β[193].

Regarding transcriptional activity: cells transduce decreased O_2 concentration into increased HIF-1 activity via an O_2-dependent post-translational modification. In hypoxic conditions the C-TAD is able to interact with transcriptional co-activators, such as p300 (CH1 domain). This interaction is unable to occur in normoxia due to the oxygen-dependent hydroxylation of Asn_{803}, located within the C-TAD. Hydroxylation of Asn_{803} is mediated by an asparaginyl hydroxylase, known as factor inhibiting HIF-1 (FIH-1), which prevents HIF-1α from interacting with the transcriptional co-activators such as the CH1 domain p300[194]. As a result, in hypoxic conditions, there is a decreased rate of HIF-1α degradation, so more HIF-1α is available, as well as an increase in transcriptional activity[195].

The co-activator protein p300 is paralogous to CBP, they are thought to control gene expression by relaxation of the chromatin structure at the gene promoter via intrinsic histone acetyltransferase activity, they also recruit basal transcriptional machinery including RNA polymerase to the promoter[196]. The multidomain proteins p300 and CBP are very similar in structure[197], they consist of key domains including; the nuclear interaction domain (Nu), the CREB and MYB interaction domain (KIX), cysteine/histadine regions (CH/TAZ), a histone acetyltransferase domain (HAT) and a bromodomain (Br)[198]. The structure if p300/CBP is summerised in Figure 12. The CH1 domain of both proteins interacts with the CTAD of HIF-1α, this domain only differs by four amino acids between the two proteins[180, 197].

A



B

	330	
p300	ADPEKRKLIQQQLVLLLHAHKCQRREQANGEVRQCNLPHCRTMKNVLNHMTHCQSGKSCQVAHCASSRQIISHWKNCTRHDCPVCPLKNAGDK	423
CBP	ADPEKRKLIQQQLVLLLHAHKCQRREQANGEVRACSLPHCRTMKNVLNHMTHCQAGKACQVAHCASSRQIISHWKNCTRHDCPVCPLKNAGDK	439
	246	

Figure 12 - p300 and CBP domain representation and comparison. A) Cartoon representation if the domain structure of both p300 and CBP. B) Comparison of the amino acid sequence of the CH3 domain of p300 and CBP.

As is highlighted above there are many different pathways for HIF-1 and increased activity in hypoxic environments. HIF-1 is also invol hypoxia signaling pathways. This is turn means there are many diffe targets for small molecule intervention.

3.1.2 Inhibitors of the hypoxic response via HIF-1

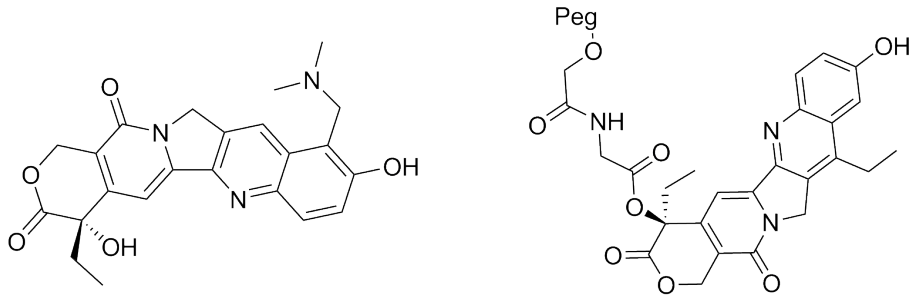
There has been much focus on inhibiting elements of the hypoxic respo inhibitors can be broadly classified by their mechanism of action. A common denominator of most, if not all, HIF-1 inhibitors identified thu of specificity, indicative of hitting multiple targets and pathways; HII cannot be easily separated from other activities exerted by these ag means that mechanism of action can be difficult to decipher, this is co the fact that many of the known inhibitors were discovered throug screening, which offers little information regarding the mechanism of classifications of HIF-1 inhibitors are (1) HIF-1α mRNA expression, protein translation, (3) HIF-1α protein degradation, (4) HIF-1α DNA bi HIF-1α transcriptional activity[199].

2.1.2.1 HIF-1α mRNA expression

It has been suggested that, under hypoxic conditions, levels of HIF-1α be a limiting factor affecting the rate of HIF-1α protein translation EZN-2698 is an RNA antagonist, which is composed of a thi oligonucleotide; a technology that specifically binds and inhibits the HIF-1α mRNA. It has shown potent (IC_{50} = 1-5 nM) and selective inhib 1α mRNA and protein expression in both normoxia and hypoxia. demonstrated a dose-dependent and highly potent down regulation o HIF-1α and VEGF in the liver. Tumor reduction was found in nude mi with DU-145 human prostate cancer cells treated with EZN-2968[201]. T that inhibition of HIF-1α mRNA has potential as a target for cancer ther

2.1.2.2 HIF-1α protein translation

Several agents have been described that may affect the rate of HIF-1α protein synthesis[199]. One such agent is Topotecan (Figure 13), an FDA approved drug currently used as a second line therapy for patients with small cell lung or ovarian cancer[199]. Topotecan works by inhibiting topoisomerase I by inducing the formation of stable Top1-DNA cleavage complexes, which in the presence of DNA replication generates double strand DNA breaks and cytotoxicity, limiting HIF-1α translation. Recently, it has been shown that administration of daily Topotecan in combination with the anti-VEGF antibody Bevacizumab exerts synergistic antitumour activity in xenograft models, providing a rationale for clinical development of this combination strategy[202]; which, as discussed previously, should limit resistance. Other topoisomerase 1 inhibitors have been developed, one such inhibitor is EZN-2208 (Figure 13). EZN-2208 has better pharmacokinetic properties and a longer half-life than Topotecan, making it more suitable for chronic suppression of the HIF-1 pathway[203]. Other agents and targets that affect HIF-1α protein translation include; Digoxin, a cardiac glycoside, which inhibits the translation of HIF-1α by an mTOR-independent mechanism, and PX-478, an agent that potentially inhibits HIF-1α translation through multiple mechanisms, although non have been confirmed.



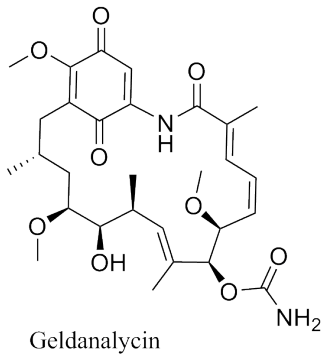
Topotecan EZN-2208

Figure 13 – Structure of Topotecan (chem draw image drawn from pubchem, pubchem ID – 60700) and EZN-2208 (chem draw image drawn from pubchem, pubchem ID – 59443782)

2.1.2.3 HIF-1α degradation pathway

Heat shock protein 90 (Hsp90) is a molecular chaperone that controls the folding and regulates the function of many proteins, including receptor tyrosine kinases, serine/threonine kinases, transcription factors and activated oncoproteins[204]. Disruption of Hsp90 function has been shown to promote HIF-1α degradation via a

novel, oxygen-independent E3 ubiquitin ligase and diminishes HIF-1α transcriptional activity[205]. In the presence of Hsp90 inhibitors HIF-1α undergoes VHL-independent proteasomal degradation[205]; HIF-1α heterodimers may also not acquire the proper conformation and therefore fail to recruit cofactors important for HIF-1-mediated transcriptional activity[206]. The first Hsp90 inhibitor was a natural product, Galdanamycin (Figure 14), which exerted its inhibitor activities by competing with the ATP binding site[205].



Geldanalycin

Figure 14 – Structure of Geldanalycin (chem draw image drawn from pubchem, pubchem ID – 5288382)

2.1.2.4 HIF-1 binding to DNA

Inhibition of HIF-1 DNA binding to the hypoxia responsive element (HRE); a step required for transcription induction, is also a potential mechanism by which small molecules may inhibit HIF-1 activity[199, 207]. Proof of principle was established using a cyclic peptide, Echinomycin (Figure 15), which was known to bind DNA in a sequence-specific fashion. It was shown that Echinomycin inhibits the DNA/HIF-1 interaction more potently than DNA/AP-1 or DNA/NF-$_K$B, binding, providing evidence of selective inhibition based on recognition of DNA sequences[207]. Polyamides, which have a similar mechanism, have also been produced[208].

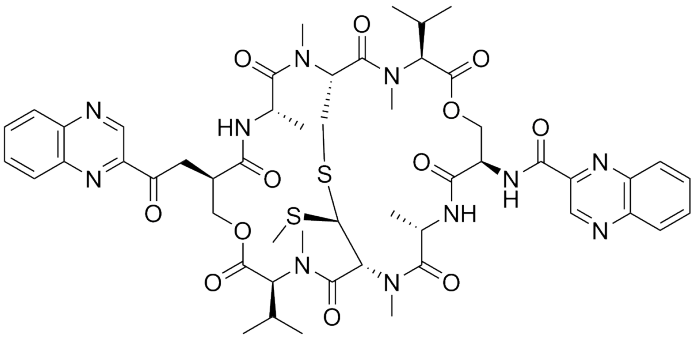


Figure 15 – Structure of Enhinomycin (chem draw image drawn from pubchem, pubchem ID –23724556)

2.1.2.5 HIF-1α transcriptional activity

Dimerization of HIF-1α and HIF-1β is critical for both transcriptional activity and DNA binding and therefore has been described as an optimal point of interception[209]. The Tavasolli group used a genetically encoded HTS platform[210, 211] for the identification of cyclic peptides that are able to disrupt the dimerization. The cyclic peptide *cyclo*-CLLFVY was identified from a plasmid encoded library of 3.2 million cyclic peptides. The compound was tested *in vitro* and in cells and was shown to disrupt HIF-1 dimerization by binding the PAS-B domain of HIF-1α[209].

Inhibition of the proteasome leads to an accumulation of HIF-1α[212], although, this HIF-1α that accumulates is transcriptionally inactive[213]. A proteasome inhibitor, Bortezomib (Figure 16), has been FDA approved. As well as inhibiting the proteasome at low nanomolar concentrations Bortezomib was also able to limit the HIF-1α/p300 interaction, by improving the binding of FIH to HIF-1α[214]. FIH is a dioxygenase that hydroxylates Asn_{803} in the C-terminal transactivation domain of HIF-1α; this prevents the recruitment of the co-activator p300. There have been efforts to directly disrupt the HIF-1α/p300 interaction, thus far with limited success. It was reported that an agent Chetomin (Figure 16) inhibited HIF-1α by disrupting the HIF-1α/p300 interaction[189], however work by the Schofield group has demonstrated that Chetomin works by ejecting structurally important zinc from p300, therefore destroying one of the binding partners not directly disrupting the interaction[215]. As p300 has other roles within the cell this is not a viable therapeutic. Therefore more work is needed to develop inhibitors of the HIF-1α/p300 interaction.

As discussed in the introduction current high throughput screens are not often amenable to PPIs, therefore focus may shift towards rational/structure based design methods. To do this, a detailed understanding of the interaction, both structurally and biophysically will be required.

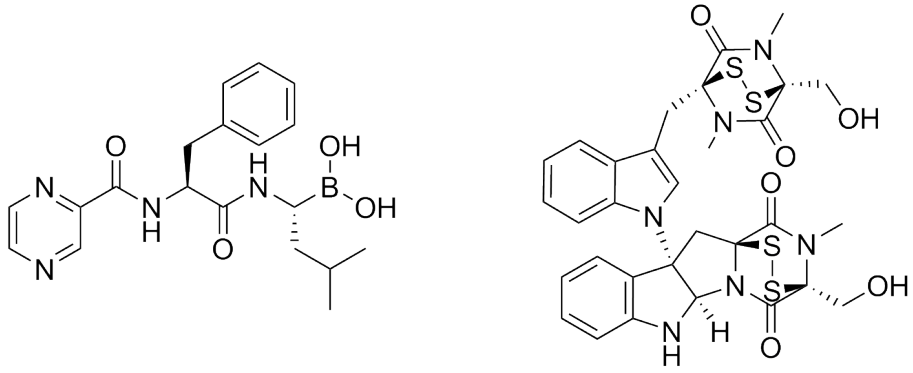


Figure 16 – Structure of Bortezomib (chem draw image drawn from pubchem, pubchem ID – 387447) and Chetomin (chem draw image drawn from pubchem, pubchem ID – 73583)

3.1.3 Structural information of HIF-1α/p300

The starting point for structure based and rational drug design of an inhibitor is to analyse the structural and biophysical data of the interaction. The interaction between the CH1 domain of p300 and the C-TAD of HIF-1α Figure 17[182] was solved by using multidimensional NMR methods (PDB:1L8C[182], 1L3E[197]).

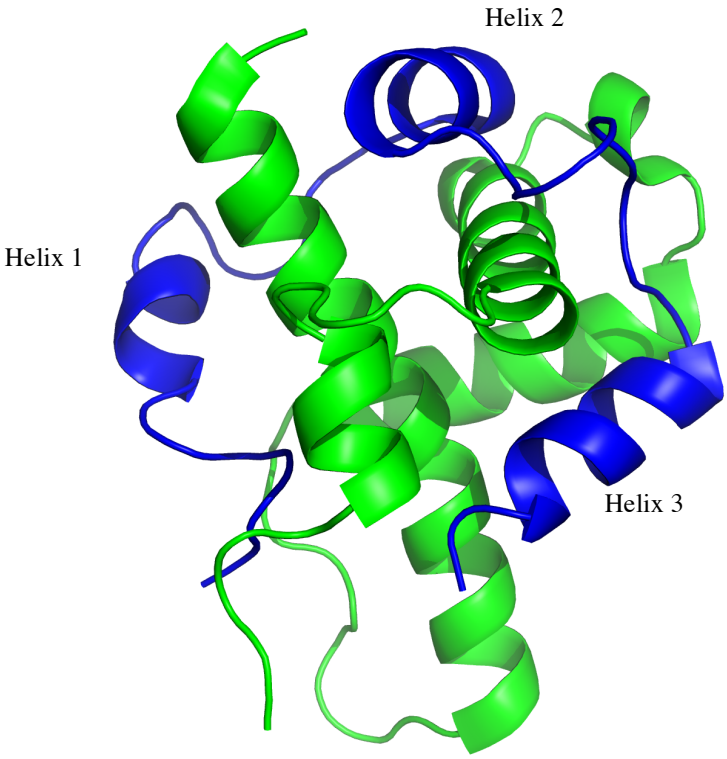


Figure 17 - Pymol illustration of the interaction between p300 (green) and HIF-1α (blue). PBD ID:1L8C[182]

Previous studies have highlighted key regions of HIF-1α (helix 2 and helix 3) to interacting with the CH1 domain of p300 by hydrophobic or polar interactions[197]. Further mutational studies such as alanine scanning have indicated key binding residues of HIF-1α[216]. Residues Cys_{800} and Asn_{803}[217-219] located in helix 2 have been highlighted as key binding residues in cellular screens. However the two residues are not hydrophobic but polar; which is unusual for PPIs and the binding events from these residue are not clear from the NMR structure[182, 197] (Figure 18).

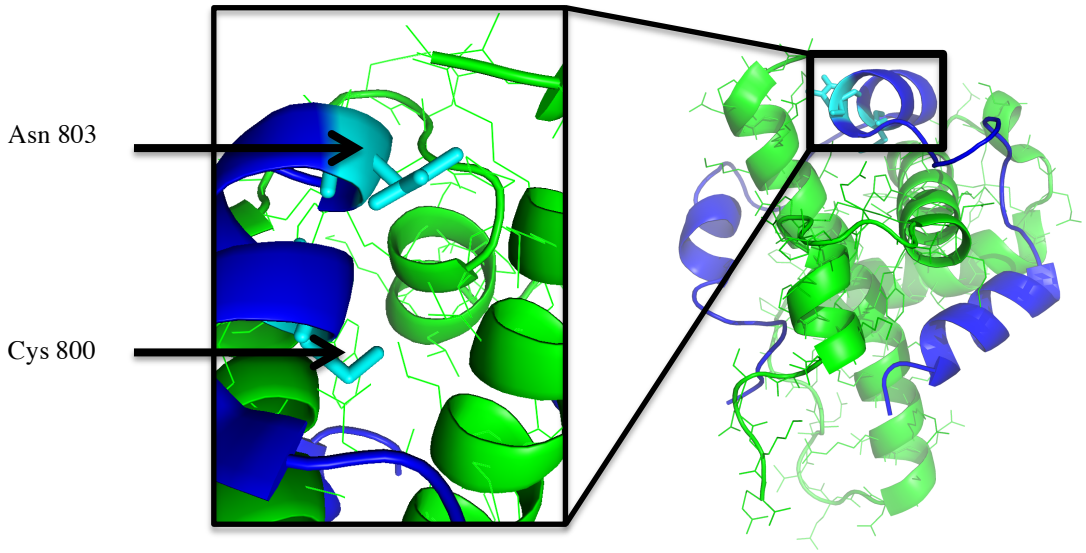


Figure 18 -Pymol image illustrating the key binding residues (cyan) of HIF-1α helix 2 (blue) binding to p300 (green). PDB:1L8C[182].

Helix 3 contains 3 key binding residues, Leu_{818}, Leu_{822}[216] and Val_{825}[197]; three hydrophobic residues, so this is a hydrophobically driven interaction. The binding of helix 3 is illustrated in Figure 19.

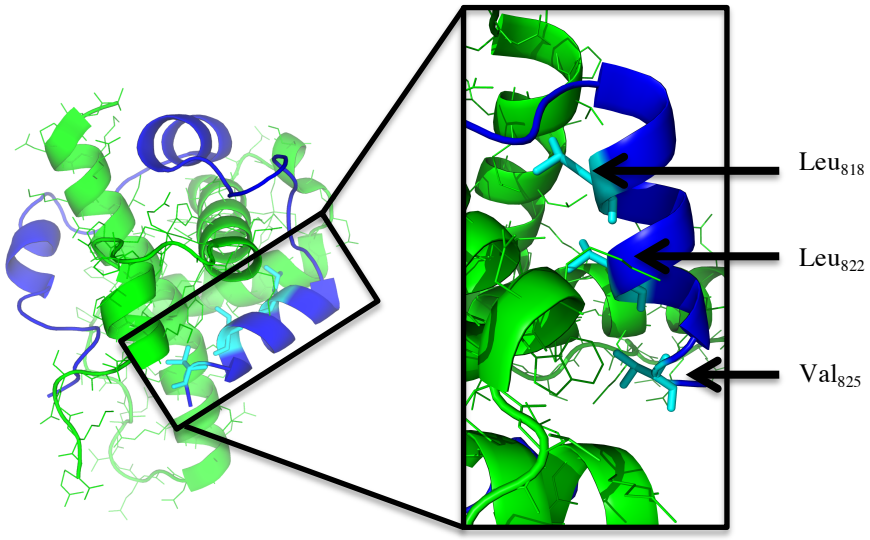


Figure 19 - Pymol illustration of the key binding residues (cyan) of HIF-1α (blue) helix 3 binding to p300 (green). PDB:1L8C[182].

Previous studies do not reveal a distinct site for small molecule modulation; therefore

further analysis is would be useful locate a binding hot-spot to target with a small molecule. A fluorescence polarisation (FP) experiment was conducted[220] to compare interactions of various peptide lengths derived from HIF-1α C-TAD with p300; the HIF-1α peptide lengths tested are illustrated in Figure 20[220]. The peptides were N-terminally labelled with fluorescein (FITC).

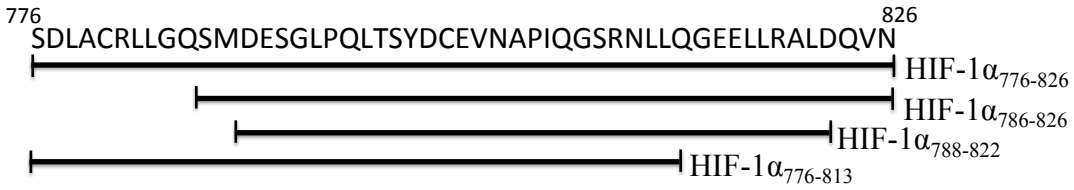


Figure 20 - Peptide lengths tested for binding affinity. Image adapted from[220]

The binding affinities for the peptides are outlined in Table 12[220]. From this experiment the authors concluded that the C-terminus of HIF-1α C-TAD is important for binding, this correlates with previous studies[197, 216, 218, 219]. However the data does not narrow down the interaction area to one of the individual helices and the interaction area is still relatively large.

Table 12 – Binding affinities of fragments of HIF-1α[220]

Peptide	K_d
FITC-HIF-1α 776–826	189 ± 45 nM
FITC-HIF-1α 786–826	166 ± 66 nM
FITC- HIF-1α 788–822	> 3000 nM
FITC-HIF- 1α 776–813	> 3000 nM

In 2010 work by the Arora group attempted to decrease the size of the interaction interface by focusing on one of the two key helices. An ITC binding experiment between the CH1 domain of p300 and C-TAD HIF-1α $_{799}$DCEVNA$_{804}$ (peptide

AcTAADCEYNAR), which corresponds to the helix 2 region; encompassing the key binding residues Cys_{800} and Asn_{803} was conducted.

They found that this short peptide region had a binding affinity to p300 of 825 nM. They recognised that synthetic mimics of these domains could potentially inhibit HIF-1α/p300 interaction and as a result down-regulate the expression of genes encoding vascular endothelial growth factor (VEGF) and its receptor VEGFR-2, which are involved in the induction of new blood vessels (angiogenesis) in solid tumors. As discussed in the introduction short peptides typically require stabilizing moieties to retain their folded conformation once excised away from the protein environment. To stabilize this helical region they utilized the hydrogen bond surrogate (HBS) approach. They made three hydrogen bond surrogates (HBS-3 was a negative control) and used CD to show all 3 hydrogen bond surrogates had a helical component, and a higher % of helicity than the unconstrained peptide. The potential for the HBSs to down-regulate the HIF-1α induced transcription of *VEGF* gene in HeLa cells under hypoxic conditions was assessed by real time quantitative RT-PCR, HBS-2 showed a good level of % transcriptional inhibition (45% comparable to Chetomin) (positive control) 50%). HBS-2 had a binding affinity for p300 by ITC of 420 nM. All of this data is summarized in Table 13.

Table 13 - Summary of biophysical and *in vitro* data for peptides designed to target HIF-1α/p300 interactions[217].

Ligand	Sequence	% helicity	Kd (nM)	Transcription inhibition (%)
HBS-1	XTAADCEYNA	40	950 ± 90	0 ± 3
HBS-2	XTAADCEYNAR	53	420 ± 35	45 ± 8
HBS-3	XTAADREYNAR	51	>>2200	2 ± 7
Peptide	AcTAADCEYNAR	15	825 ± 50	8 ± 3
Chetomin	-	-	120 ± 25	50 ± 5

Further analysis has indicated that HBS-2 does not disrupt the interaction by denaturation of p300 in the same way as has been shown with Chetomin; HBS-2 does not alter the CD spectrum of p300, indicating that the structure of p300 is not altered dramatically (i.e. unfolded) in the presence of HBS-2. Furthermore a cell viability assay demonstrated that HBS-2 does not display significant cytotoxicity[217].

This work would indicate helix 2 is a promising target; however further work by this group has focused on the helix 3 region, resulting in doubt as to which helix is best to target. A computational alanine scan was conducted which suggested that Leu_{819} was not a significant contributor to binding affinity as was initially suggested[216]; instead Leu_{822}, Asp_{823} and Gln_{824} were identified as key binding residues[221]. In a similar way to the piece of work described previously, HBS peptides were prepared, although this time mimicking the helix 3 region (Leu_{822}, Asp_{823} and Gln_{824}), HBS-1 was the molecule under investigation (expected to have the highest potency), HBS-2 had a point mutation of a key binding residue (Leu_{822}-Ala_{822}), so would be expected to bind with a lower affinity, acting as a negative control; a peptide (an unconstrained analogue of HBS-1), was also made allowing for the evaluation of the effect of helix stabilisation. The three agents described were all compared to the native helix 3 peptide region.

Again the constrained peptides showed a characteristic α-helical circular dichroism spectroscopic signature in aqueous buffer compared to the unconstrained peptide. The affinity of each of these agents for p300 was measured by tryptophan fluorescence spectroscopy. As there is one tryptophan in p300, located in the helix 3 binding pocket (Trp_{403}), it is a sensitive probe for interrogating binding of helix 3 mimetics. Using this fluorescence method, it was calculated that HBS 1 bound to p300 with a K_d = 690 ± 25 nM, this, as expected, was a higher affinity than HBS-2 (negative control with point mutation) and the unconstrained analogue of HBS-1, peptide 3; as a comparison HIF-1α C-$TAD_{786–826}$ binds p300-CH1 with a K_d = 38 ± 0.14 nM. In addition to the Trp_{403} fluorescence change experiment, evidence for binding in the helix 3 binding pocket has come from HSQC NMR experiments, with prominent shifts occurring from residues around the helix 3 binding site, including Trp_{403}. In addition it was shown that HBS-1 is able to disrupt the HIF-1α/p300 (K_i = 1.2 μM) using a fluorescence polarisation competition assay, whilst neither HBS-2 or peptide 3 caused reproducible inhibition of the complex[221].

Based on the confirmed ability of HBS-1 to bind p300 and disrupt the p300/HIF-1α interaction, the potential of HBS-1 to down-regulate hypoxia-inducible promoter activity *in cellulo* using a luciferase-based reporter gene system was investigated. A construct containing five tandem repeats of the HRE consensus sequence found in the VEGF promoter (TACGTGGG), cloned upstream of the human CMV minimal promoter was used to drive expression of firefly luciferase[222]. HBS-1 at a concentration of 50 μM reduced luciferase expression by 25%. At the same concentrations, specificity control HBS-2 and unconstrained peptide 3 were found to be less effective. The luciferase reporter assays suggest that treatment with HBS 1 results in a statistically significant down-regulation of HIF-1α inducible transcription in this cell line. It was confirmed by western blot that the down-regulation was not as a result of decreased expression of HIF-1α. The ability of HBS-1 to inhibit hypoxia-induced transcription of target genes (VEGFA, SLC2A1/GLUT-1, and LOX) was evaluated using real-time quantitative RT-PCR (qRT-PCR) assays. HBS-1 reduced expression levels of VEGF by 50% at 10 μM and greater than 60% at 50 μM; showing marked dose dependence. Expression of SLC2A1 (GLUT1) gene, one of the markers of glycolysis, showed dose dependent inhibition (50-60%) and LOX, the hypoxia-inducible gene that has been shown to promote metastasis, was also

decreased in a dose dependent manner (55-70%). An ELISA showed that HBS-1 also down-regulated VEGF protein levels in HeLa cells in a dose-dependent manner. As with HBS peptide targeted against the helix 2 binding site, this HBS peptide is not cytotoxic. The investigators went further using a mouse xenograft tumour model to assess the *in vivo* efficacy of HBS-1. HBS-1 was retained in plasma at much higher concentrations compared with the unconstrained peptide, suggesting that the internally constrained structure of HBS-1 impacts favourably on serum stability. This observation is supported by the fact that proteases largely bind and cleave peptides in extended conformations[223]; the plasma stability of HBS-1 is also consistent with the published stability of hydrocarbon-bridged helices[224]. Throughout the course of the treatment and at the experiment endpoint, mice treated with HBS 1 had smaller tumours with median tumour volume reduction of 53% compared with the mice from the control group; they also showed no distress or significant weight loss.

Another piece of work by this group again targeted the helix 3 domain, this time using an oxopiperazine helix mimetic (OHMs) scaffold[225]. OHMs are assembled from naturally occurring amino acids with the nitrogen atoms of neighbouring backbone amides constrained with ethylene bridges providing a non-peptidic chiral scaffold that displays protein-like functionality as the bridges confine the side chain groups in orientations that mimic α-helices. Molecular modelling suggests that the low-energy conformation of the oxopiperazine scaffold presents side chain functionality to mimic the arrangement of the *i*, *i*+4, and the *i*+6/*i*+7 residues on α-helices (Figure 21A)[226]. A further advantage of OHMs is their chiral backbone; chiral scaffolds are expected to interact with molecular binding pockets with higher specificity. Three of the key binding residues, Leu_{818}, Leu_{822}, and Gln_{824}, can be mimicked by oxopiperazine. Four analogues were designed and synthesised (Figure 21B), OHM-1 contains projections representing all three key residues from HIF-1α: R1 as Leu_{818}, R2 as Leu_{822}, and R4 as Gln_{824}. The R3 position of the oxopiperazine scaffold is not predicted to make contacts with the target protein; an alanine residue was inserted at this position. OHMs 2 and 3 are single mutants of OHM-1 with R4 and R2 positions, respectively, substituted with alanine residues. Based on computational analysis and the relative contributions of Leu_{822} and Gln_{824}, OHM-2 would be expected to bind p300 with a much higher affinity than OHM-3, as Gln_{824} in the R4 is a weak contributor to binding where as Leu_{822} in the R2 position is

predicted to make a larger contribution to binding affinity, therefore the substitution to alanine in OHM-2 would be expected to have a larger effect than in OHM-3.

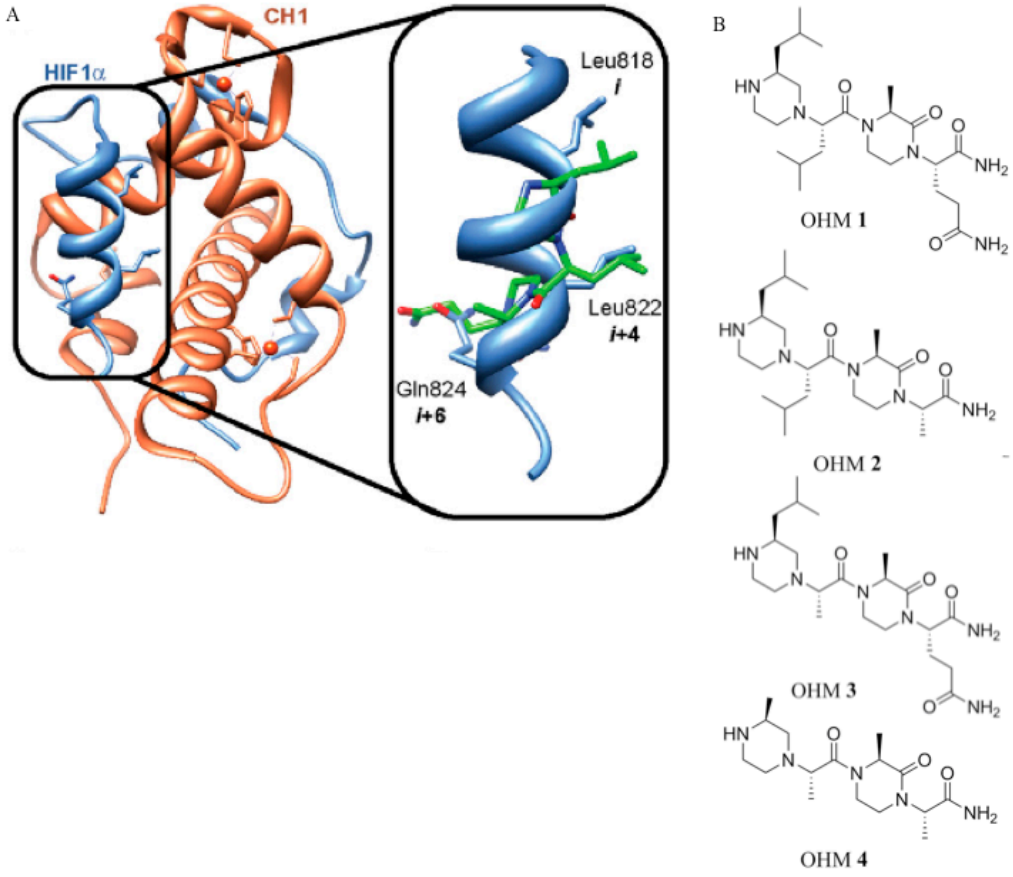


Figure 21 - OHM helix mimetics. A) molecular modelling suggests the scaffold projects the side chains in the same special orientation as a helix. B) Four OHM compounds were designed and synthesised. Image adapted from[225].

Intrinsic tryptophan fluorescence spectroscopy was again used to measure binding affinity to p300 CH1 domain; the results are summarised in Table 14. OHM-1, which represented all three key side chains, had the highest affinity of the OHMs. OHM-2, which contains the two critical leucine residues but lacks Gln_{824}, binds with a slightly reduced affinity for OHM-1, confirming the computational prediction that Gln_{824} is a weak contributor to binding. The negative controls OHMs 3 and 4 displayed very weak affinities for p300.

Table 14 – Binding affinities of OHM ligands[225]

Ligand	K_d
HIF1α C-TAD$_{786–826}$	38 nM
OHM-1	530 nM
OHM-2	620 nM
OHM-3	>>10 μM
OHM-4	>>10 μM

The results signify that the designed scaffolds are able to target the protein of interest in a predetermined manner. NMR was used to further characterise the binding; concentration-dependent shifts of several residues were observed upon addition of OHM-1. The NMR results, along with the fluorescence binding experiments, provide strong evidence that the rationally designed topographical mimics of helix 3 are binding in the predicted binding sites of their intended targets. The luciferase-based reporter assay again showed a dose-dependent reduction in the promoter activity. The ability of OHMs to inhibit transcription of three selected HIF target genes, VEGFA, LOX, and GLUT1, was assessed using real-time quantitative RT-PCR (qRT-PCR) assays in A549 cells. OHMs 1 and 2 at 10 μM down-regulate the mRNA expression levels of the critical angiogenesis regulator vascular endothelial growth factor (VEGFA) by 80% and 90%, respectively. In comparison, control compounds had no effect on VEGFA mRNA levels at these concentrations. Similar levels of decrease were observed for lysyl oxidase (LOX) and glucose transporter 1 (GLUT1) expression. OHM-1 caused changes in 32 transcripts by at least fourfold ($P \leq 0.005$) and 597 transcripts by at least twofold ($P \leq 0.005$), of these transcripts 11 are validated as HIF-1α target genes[227].

The ability of OHM-1 to reduce the tumour growth rate in a mouse xenograft models was assessed; the treated group had a smaller median tumour volume (103 mm3) compared with the control group (186 mm^3). This indicated that OMH-1, a mimetic of the helix 3 region of HIF-1α, is a potential cancer therapeutic.

It can be seen that both helix 2 and helix 3 have been mimicked and have successfully inhibited the HIF-1α/p300 interaction. Therefore further work is required to resolve which of the helical sites is the more prominent binding hot-spot, indeed if either of these regions is a hot-spot of the interface, as there is a potential that a hot-spot is located at an unexplored region of the interaction interface. There is also a potential that there are no binding hot-spots of this interaction and the full length HIF-1α C-TAD is required for high affinity binding. If this is the case it does not mean that the interaction is "undruggable", just structure-based rational design will be more difficult as there is no clear starting point.

3.2 HIF-1α/p300 Aims

The aim of this part of the project was to further characterise the HIF-1α/p300 interaction using a combination of biophysical and structural techniques. It was hoped that an increased understanding of the binding interface and the location of binding hot-spot will aid rational drug design. Two approaches were envisioned to probe the HIF-1α binding surface on p300: first, by analysis of the binding of shorter HIF-1α peptide fragments; and second, by phage display experiments. Binding analysis of fragments of the native peptide allow identification of the highest affinity region of the HIF-1α peptide, whereas phage display permits the unbiased exploration of the p300 CH1 protein surface to discover high affinity binders. The location and binding mode of phage display derived ligands provides new information on suitable chemotypes for orthosteric small-molecule inhibitor development.

3.3 HIF-1α/p300 cloning and expression

As discussed in the introduction structural and rational drug design methods may be a way to yield inhibitors of protein-protein interactions (such as HIF-1α/p300). For such design strategies to be successful an intimate knowledge of the binding interaction and protein binding partners, both structurally and biophysically must be gained. Before binding and structural studies can be conducted, the protein partners must first be cloned and expressed/synthesised. In this section the cloning and expression of p300 and HIF-1α proteins is described.

3.3.1 p300 CH1 cloning

Two different p300 CH1 lengths were cloned, $p300_{330-420}$ and $p300_{323-423}$. The two lengths were chosen because firstly, 330-420 fits with the structured region of p300 as shown in the NMR structure[182]; as a crystal structure of p300 was an aim of the project having a highly ordered protein structure will be an advantage. However it has been reported that for maintenance of full HIF-1α activity p300 resides 323-423 are required[197]. Also previous work characterising the HIF-1α/p300 interaction by the Arora group used this longer construct[217, 221, 225]. The additional C-terminal residues are illustrated in Figure 22; there is no structure available which shows the additional N-terminal residues. The entire unstructured region was not removed (cut to residue 417) as it is likely that some C-terminal residues will be required for the stability of the C-terminal helix.

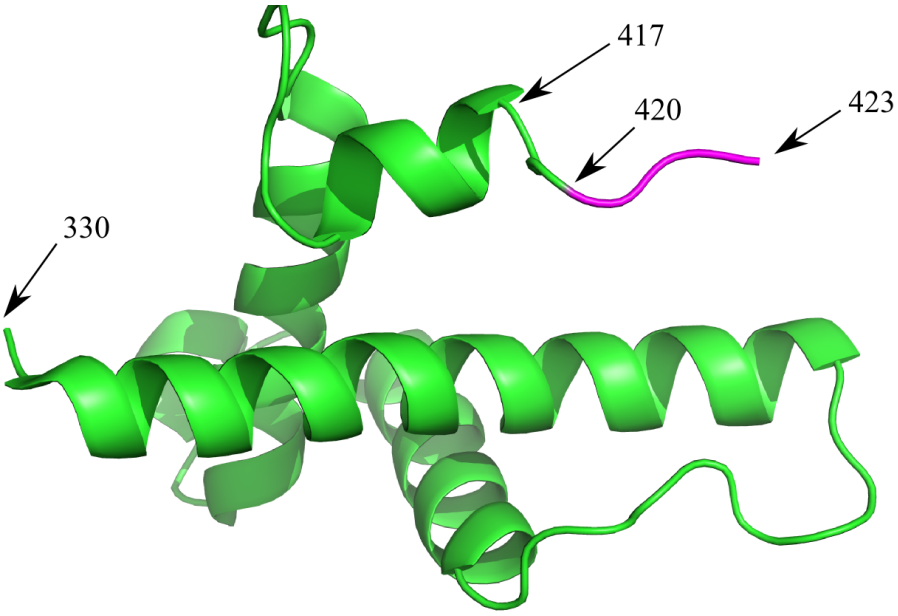


Figure 22 – Illustration of the the additional C terminal residues (cyan) for the two p300 constructs (Pymol, cartoon). PDB ID: 1L8C[182].

The CH1 domain of p300 ($p300_{330-420}$ $p300_{323-423}$) was produced by PCR during PCR the restriction sites BamHI (N-term) and Xho1 (C-term) were added. The primers used are out lined in the methods section (

Table 2). The PCR products were ligated in to three plasmids:

- pET-SUMO-28a, the SUMO tag aids solubility and purification[228];
- pGEX-6P-2, the GST tag again aids solubility and purification[229]
- pET-GFP-28b-PreScission, the GFP tag can be used for FRET, this plasmid also contained a HIS tag for purification[230].

Positive constructs were identified by colony PCR and diagnostic digest and subsequently confirmed by sequencing. The plasmid (pGex-6p-2 as an example) with the inserted p300 DNA is illustrated in Figure 23.

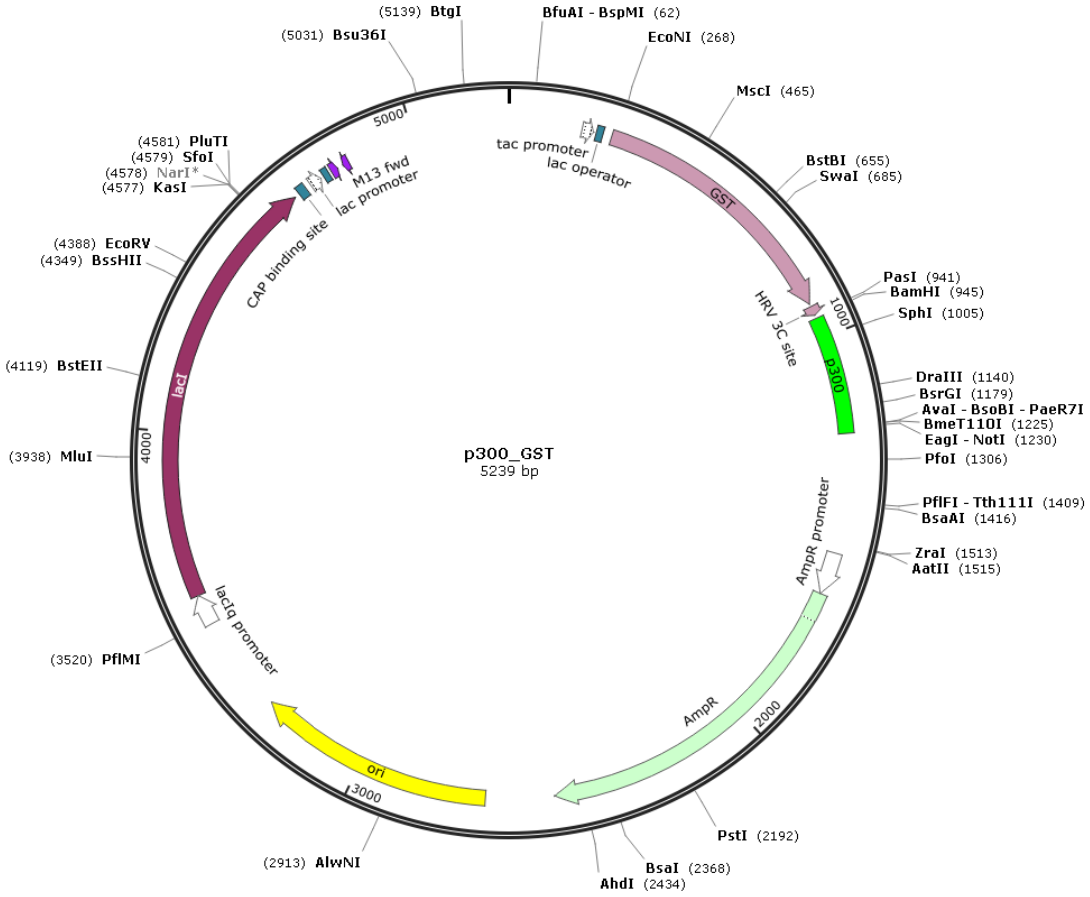


Figure 23 - Cloning of the $p300_{330\text{-}420}$ construct, in to the pGex-6p-2 plasmid, image made in Snapgene

3.3.2 Expression Trials

E. coli BL21 (DE3) pLysS expression strains Gold, Star and Rosetta 2, were transformed with each construct, from which expression trials were conducted. The different strains contained mutations from BL21 (DE3) pLysS[231, 232], outlined in Table 15.

Table 15 – Summary of difference in expression strains used in expression trials[231, 232]

Strain	Mutation
BL21 (DE3) pLysS star	mutated Rnase E
BL21 (DE3) pLysS gold	lack the Lon protease and deficient in OmpT protein
BL21 (DE3) pLysS rosetta 2	rare tRNA included

Initially expression trials were conducted using the $p300_{330-420}$ construct. The expression strains were tested first, small scale (10 mL) cultures were produced, initial induction was with 0.4 mM IPTG incubated at 18°C for 18 hours. For GST-$p300_{330-420}$ (37 kDa) each of the expression strains yielded soluble protein, with the greatest expression ratio of insoluble to soluble protein from Gold cells. SUMO-$p300_{330-420}$ (24 kDa) expressed well, however the protein was predominantly insoluble under these expression conditions. Again GFP-$p300_{330-420}$ (42 kDa) expressed well under these expression conditions, particularly in Star cells however the protein was largely insoluble. The SDS PAGE gels from these expression trials are summarized in Figure 24.

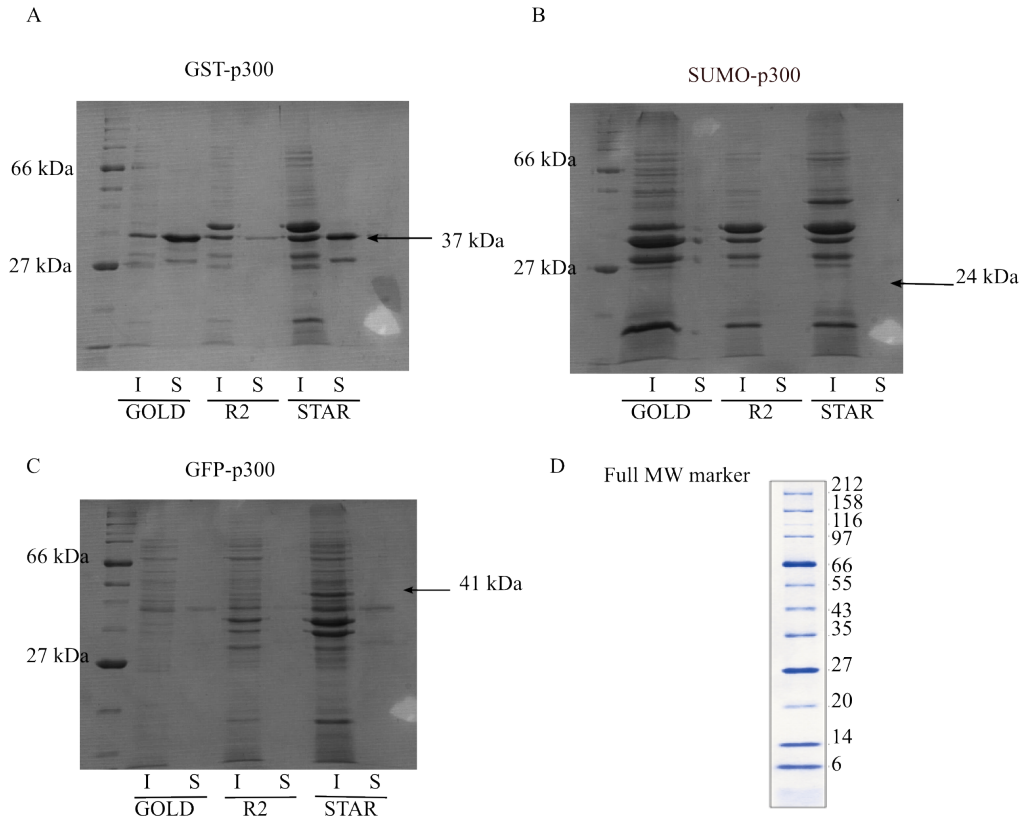


Figure 24 - SDS PAGE analysis of expression trials of $p300_{330-420}$ in expression strains *E. coli* BL21 (DE3) pLysS GOLD/ROSETTA 2 (R2)/STAR. I =insoluble fraction S=soluble fraction A) GST-p300 (37 kDa). B) SUMO-p300 (24 kDa). C) GFP-p300 (41 kDa). D) Broad range molecular weight marker (2-212 kDa**).**

The expression trials indicated that the most promising construct was GST-$p300_{330-420}$ expressed in *E. coli* BL21 (DE3) pLysS Gold cells; therefore this construct was

taken forward for further optimization. Although it should be noted that the other constructs may still be useful for specific experiments (for example GFP-p300 for fluorescence studies such as FRET).

p300 is a zinc binding protein (Figure 25A)[197], many zinc binding proteins require additional zinc either during induction or purification, and it has been shown, particularly for p300, that the concentration of zinc present has a profound influence on the fold of the protein[180]. De Guzman *et al.,* demonstrated by CD that with no additional zinc, p300 is only partially folded (Figure 25B top), however they also demonstrated by NMR when too much zinc was present p300 was unfolded (Figure 25B bottom)[180]. As the zinc concentration affects the fold of the p300 it would be expected that for suboptimal zinc concentrations there would be a higher percentage of protein in the insoluble fraction. IPTG concentration can also affect the expression levels of soluble protein. Therefore expression trials investigating the concentration of zinc (no zinc, 50 μM zinc sulphate at induction or 100 μM at lysis) and IPTG concentration (0.1 mM, 0.2 mM and 0.4 mM) were conducted. These trials showed the highest protein concentration in the soluble fraction to be under the conditions 50 μM zinc sulphate at induction and 0.1 mM IPTG. This condition was taken forward and the presence of EDTA (1 mM) and DTT (1 mM) was investigated. More protein remained soluble in the absence of EDTA and DTT (Figure 25C).

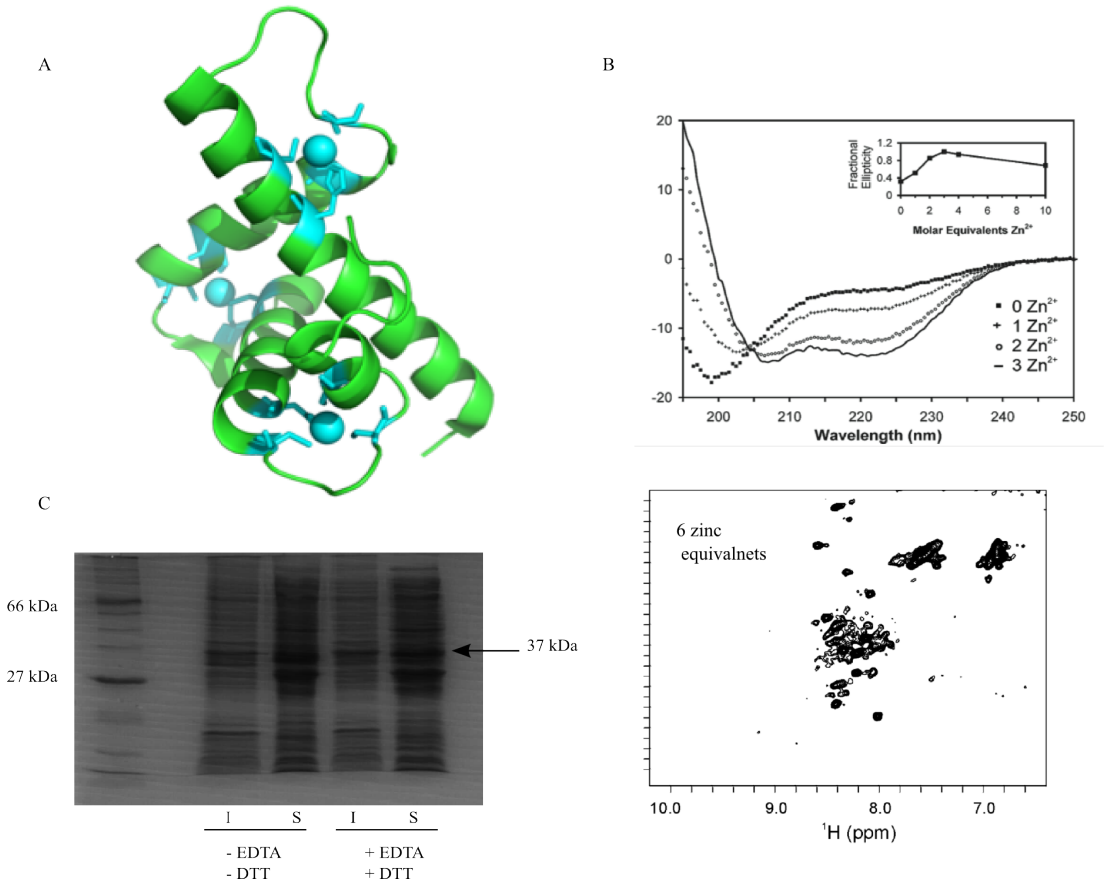


Figure 25 - Effect of zinc and reducing agents on p300 expression, solubility and fold. A) pymol image of p300 coordinating 3 zinc molecules (pale blue) , each zinc is coordinated by 3 cysteines and a histamine. PDB:1L8C[182] B) The effect of zinc concentration on protein structure by CD (top) and NMR (bottom)figure adapted from[180]. C) Expression trails of GST-p300 (37 kDa) in the presence and absence of DTT and EDTA.

From this series of expression trials the expression condition; induction by 0.1 mM IPTG, 50 µM zinc sulphate at induction at 18°C for 18 hours was taken forward to a larger scale induction, the full purification protocol is outlined in Chapter 2 : Materials and Methods. A 2 L culture under the outlined conditions yielded 40 mg of protein post affinity chromatography purification (GST column). The vast majority of this protein was GST-p300 however there was some smaller contaminating proteins present (as seen on the SDS page gel). Most likely, these proteins were GST and small degradation products (Figure 26A). This protein was then further purified by size exclusion chromatography. GST-p300 eluted from the S75 column at 140 mL, which from the calibration of column indicated a protein of the correct size. The size was confirmed by SDS PAGE and mass spectrometry. Unfused GST was still present (as seen on the SDS PAGE gel), however this was not a problem when p300

was cleaved from GST using PreScission protease (1 mg of protease per 20 mg of GST-p300, incubated for 16 hours at 4°C) as the size difference was large enough (27 kDa GST to 10.8 kDa p300) for good separation off the S75 gel filtration column (Figure 26B and C). 10 mg of pure $p300_{330-420}$ was obtained and confirmed by SDS PAGE (insert Figure 26C) and mass spectrometry.

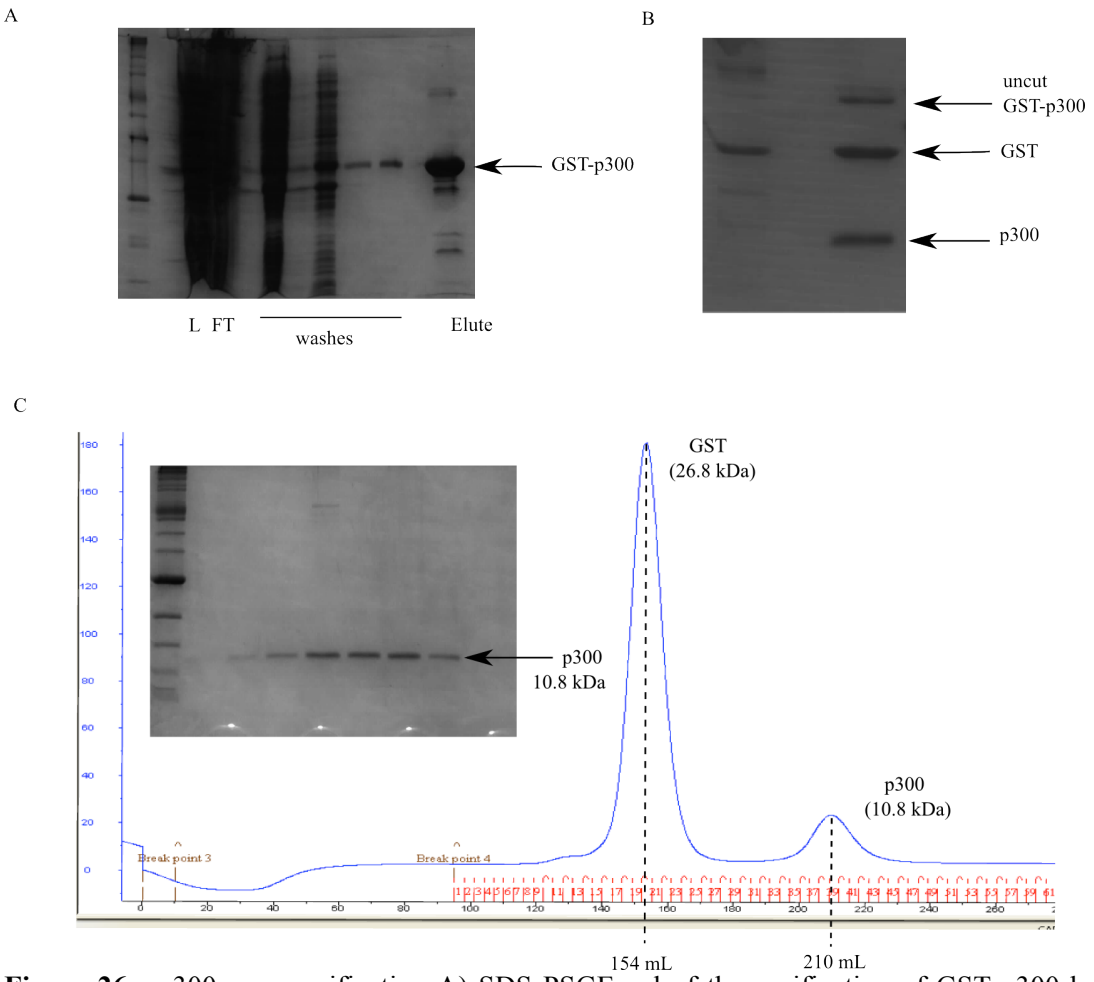


Figure 26 - $p300_{330-430}$ purification A) SDS PSGE gel of the purification of GST-p300 by affinity chromatography (GST column). B) SDS PAGE gel illustrating the cleavage of GST-p300 by PreScission protease. C) Size exclusion chromatography spectra showing the separation of GST (26.8 kDa – 154 mL) from p300 (10.8 kDa – 210 mL). Insert SDS PAGE gel showing pure p300 after size exclusion chromatography.

The fold and stability of $p300_{330-420}$ was assessed by CD. A classical spectra indicative of an α-helical secondary structure was obtained[134] (Figure 27A). The secondary structure could be disrupted by the addition of 1M EDTA, again confirming the structural importance of zinc (Figure 27A). A thermal melt demonstrated the stability if this construct. The protein was not fully denatured up to 90 °C, indicating that $p300_{330-420}$ is a highly thermostable protein (Figure 27B).

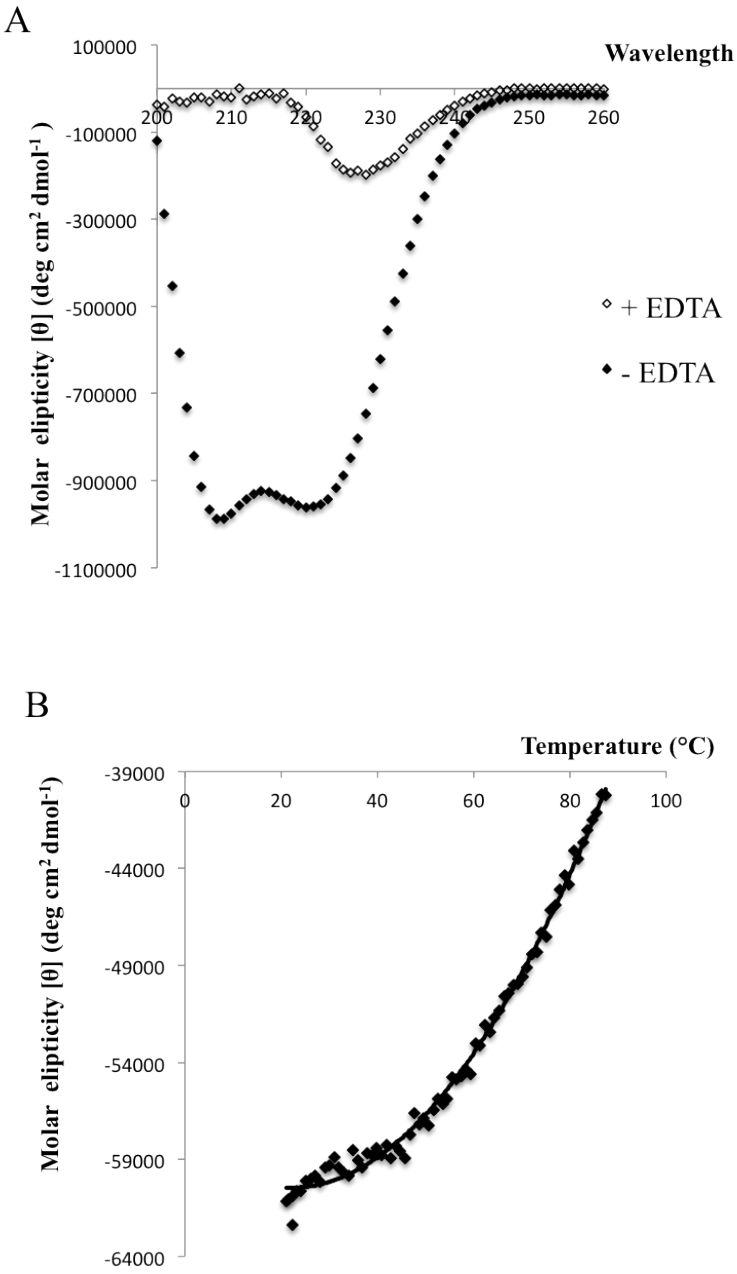


Figure 27 - CD analysis of $p300_{330-420}$. A) p300 (18.5 μM) spectra in the presence and absence of EDTA. B) Thermal melt of p300, minima 208 plotted.

The second p300 construct $p300_{323-423}$ was also cloned in to pET-SUMO-28a, pGEX-6P-2 and pET-GFP-28b-PreScission; *E. coli* BL21 (DE3) pLysS expression strains Gold, Star and Rosetta 2 (Table 15) were transformed with the constructs. $p300_{323-423}$ expressed well with a GST tag in Gold cells, GST-p300 also expressed well in Rosetta 2 and Star cells, however it was decided that GST-p300 in Gold cells would

be the construct taken forward to remain constant with the $p300_{330-420}$ construct. An identical expression and purification protocol to $p300_{330-420}$ was used. A 4 L growth of GST-$p300_{323-423}$ yielded 350 mg of protein after affinity chromatography and 200 mg post size exclusion chromatography. GST-$p300_{323-423}$ was cut by PreScission protease and $p300_{323-423}$ was separated from GST by size exclusion chromatography $p300_{323-423}$ was also pure and stable. However, the $p300_{323-423}$ expression and purification was not consistently reproducible, further use of this protein revealed a tendency for aggregation, as shown by gel filtration with the band at 23 kDa. It is believed $p300_{323-423}$ may dimerise as there was a band on the SDS PAGE gel slightly below the 27 kDa band for GST and the 185 mL elution volume corresponds to a 22.8 kDa protein (a dimer of the 11.4 kDa monomer p300 protein, which should be eluted at 210 mL) (Figure 28). However it is unusual for a dimer to be seen on the denaturing SDS PAGE gel, therefore this may be a contamination product as opposed to a dimer.

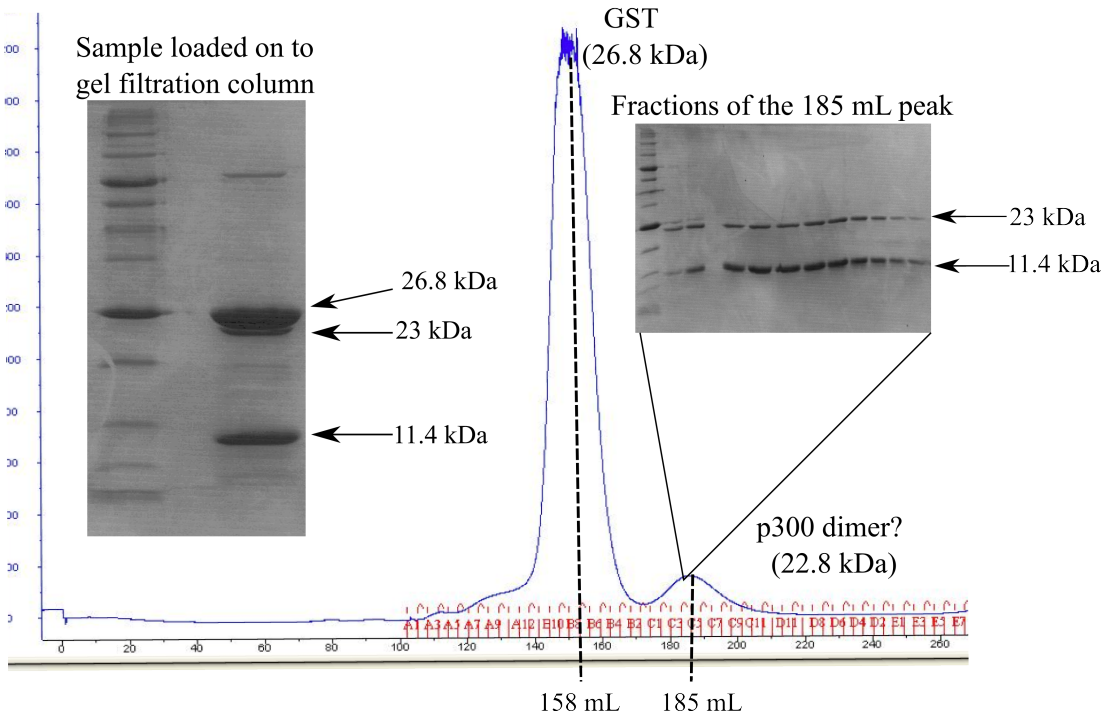


Figure 28 - Illustration of the potential dimerization of p300 (323-423). The p300 peak elutes at 185 mL (opposed to 210 mL, the expected elution volume for a monomer) .SDS PAGE gels prior to and post size exclusion chromatography show a band at approximately 23 kDa.

The fold and stability of the $p300_{323-423}$ monomer and the product eluted at 185 mL (potential dimer) were compared by CD. There was a difference in the spectra of the potential dimer and the monomer (Figure 29). The monomer has a classical α-helical secondary structure signature, however the dimer does not; this could be because there is more than one species present in the sample. The minima of the potential dimer spectra is closer to 200 nm, which is more indicative of an irregular structure[134]. Further analysis e.g. using NMR, may reveal if the structure is dimeric (for example is there a miss coordination of zinc). In addition alternative expression systems (such as arctic express cells or autoinduction) or simply re-optimisation of the expression and purification of the $p300_{323-423}$ construct could be conducted to reproducibly obtain pure monomeric protein.

As the $p300_{330-420}$ construct was pure and monomeric and obtained in good yield, this protein was taken forward for binding and structural studies. From this point forward $p300_{330-420}$ will be referred to as p300.

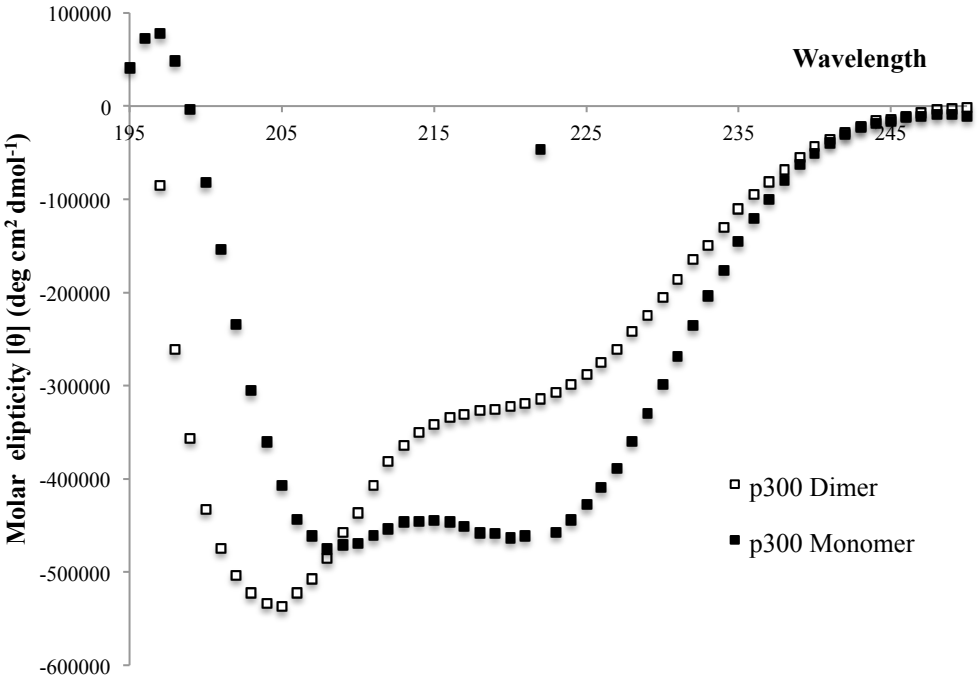


Figure 29 - CD analysis of $p300_{323-423}$ 0.2 mg/mL in both its monomeric and potential dimeric form.

3.3.3 p300 labeling

For some biophysical and structural assays the p300 protein needed to be labelled, for example to immobolise p300 for SPR and to ^{15}N labelled for NMR.

For protein isolation there are many methods for protein labelling including; derivatisation of cysteine residues[233], labelling lysine or N-terminal amino groups with activated esters, periodate or PLP-mediated oxidation of the N-terminus for oxime ligation[234] and native-chemical ligation[235]. Each of these methods has its limitations. The ability to control the location of the label has distinct advantages over uncontrolled labelling, therefore the method used was chemo-selective N-terminal labelling using depsipeptide substrates for the transpeptidase sortase A[171]. This method required an N-terminal glycine, although it must be in a sterically unhindered position. Fortuitously an N-terminal glycine was left after GST-p300 was cleaved with preScission protease.

Firstly it was assessed if the N-terminal glycine of p300 was sterically accessible and therefore amenable to this labelling method. A depsispeptide with a UV active group was synthesised and purified and the sortase A expressed by George Burslem and Dan Williamson (University of Leeds). Four reactions were set up in buffer 50 mM Hepes pH 7.5, 150 mM sodium phosphate and 5 mM calcium chloride at 37 °C and the reactions were sampled every hour for a 5 hour period to investigate optimal labelling. The reaction conditions are summarised in Table 16.

Table 16 - Reaction conditions to investigate the ligation of a depsipeptide to the N terminal glycine of p300

	Reaction 1	Reaction 2	Reaction 3	Reaction 4
Sortase (μM)	10	10	20	20
Depsipeptide (μM)	150	300	150	300
p300 (μM)	100	100	100	100

The reactions were sampled hourly and analysed by SDS PAGE. The fluorescent depsipeptide and sortase A can be seen in Figure 30A and B. After incubation, the SDS PAGE gels showed a band at around 11 kDa which is around the size for p300

and just above 27 kDa which is sortase. Both of these bands fluoresce when illuminated with UV (Figure 30), indicating that they are protein (p300 and sortase respectively) depsipeptide complex. An example gel is shown in Figure 30C and D (this gel was run at the 5 hour time point).

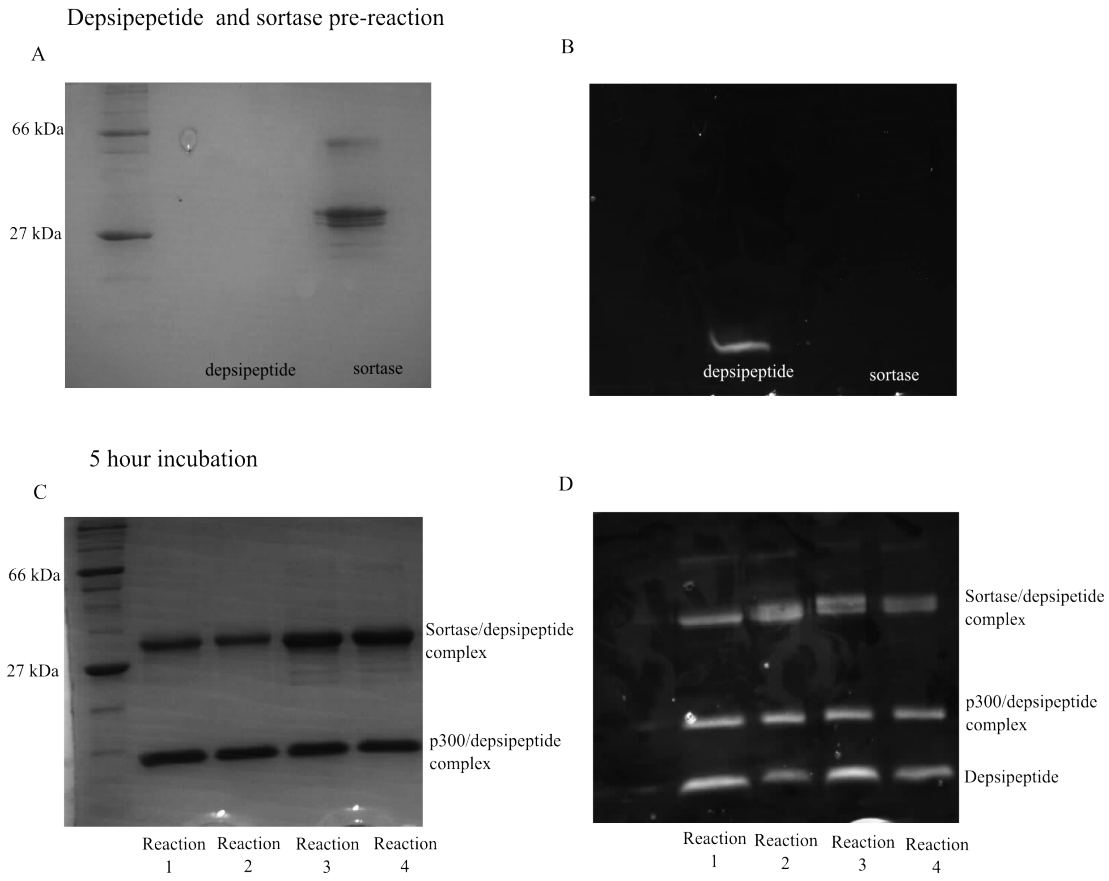


Figure 30 – Depsispeptide ligation trials. A) and B) Analysis of depsipeptide and sortase pre-reaction. A) SDS PAGE coomassie stained. B) SDS PAGE under UV showing the fluorescence of the depsipepide. C) and D) Analyssi of reaction at the 5 hour time point. C) SDS PAGE coomassie stained showing a band at 10.8 kDa (p300) and 27 kDa (sortase). D) SCS PAGE under UV, the both the p300 and sortase bands fluoresce, indicating a protein-peptide complex.

This trial indicated that the N terminal glycine was sterically accessible, as can be seen by the UV fluorescence of the p300, indicating that the depsipeptide was attached (Figure 30). Therefore a large scale reaction was conducted using a biotin-depispeptide (synthesised by George Burslem and Dan Williamson) (Figure 31A). From the trial reactions, the reaction 4 condition 100 μM p300, 300 μM biotin-

depsipeptide and 20 μM sortase was used and the reaction was incubated at 37 °C with agitation for 3.5 hours. The biotin-p300 was separated from the unreacted biotin-depsipeptide by dialysis and from the sortase A by affinity chromatography. As the sortase had a His tag a nickel column was used. The biotin-p300 was eluted in the 50 mM imidazole elute. SDS PAGE analysis indicated that the biotinylation was successful, with a slight increase in the band height of biotin-p300 compared with wild type p300 (Figure 31B), biotinylation was subsequently confirmed by immunoblot and mass spectrometry.

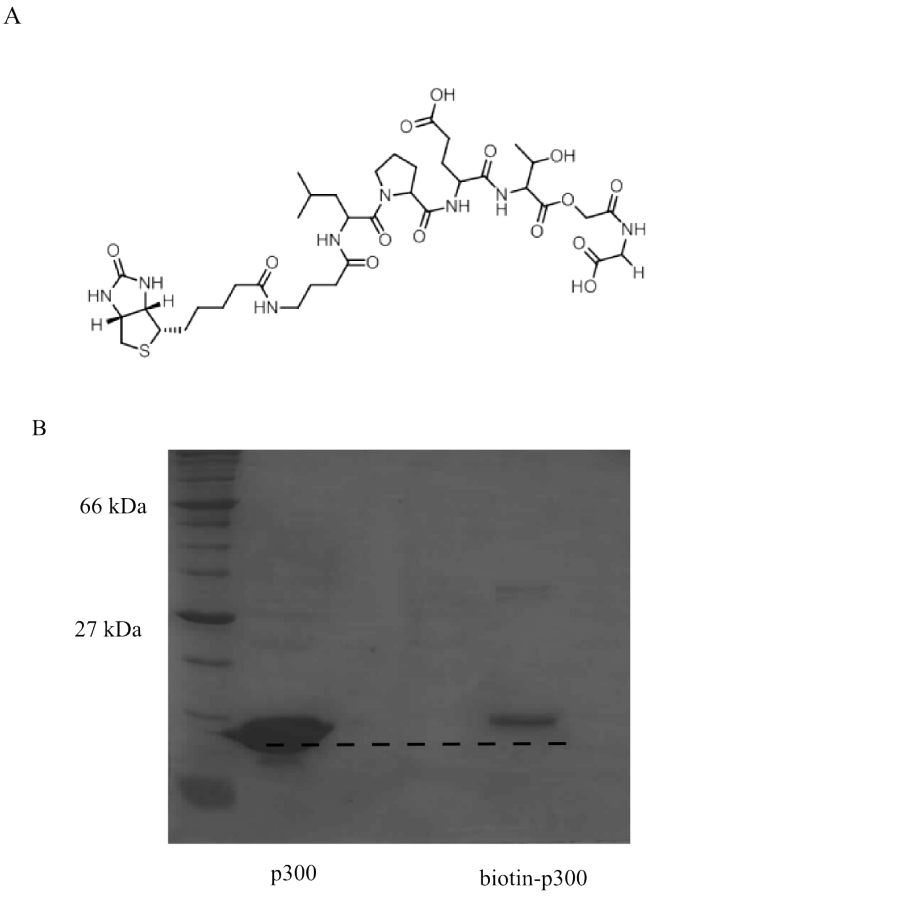


Figure 31 - Biotinylation of p300. A) Biotin-depsipeptide. B) SDS PAGE gel shows a slight increase in size in biotin-p300 from wildtype p300.

For NMR studies p300 was labelled with ^{15}N, as p300/CBP had already been assigned. ^{13}C labelling was not required[182]. Expression of ^{15}N-p300 by IPTG yielded potentially misfolded protein in low yield so autoinduction was attempted. An adaptation of the ^{15}N media (outlined in Chapter 2; Materials and Methods, 2.3.2.2

^{15}N-p300 protein expression) was used; the 20% (w/v) glucose generally used was replaced with the 50x 5052 from the Studier autoinduction protocol[236] (25% glycerol, 2.5% glucose and 10% lactose). Cells were grown at 37 °C for 4 hours (an OD_{600} of 0.22), temperature was then decreased to 20 °C and cells were incubated for 40 hours. The cells were harvested and purified as described in Chapter 2 ; Materials and Methods. From 2 L 50 mg of protein was eluted from the GST column, approximately 2/3 of this was GST-^{15}Np300 with the contaminating protein being GST (Figure 32A). The GST- ^{15}Np300 fusion was cut with preScission protease (Figure 32B) and separated by size exclusion chromatography (Figure 32C), the labelling was confirmed by mass spectrometry.

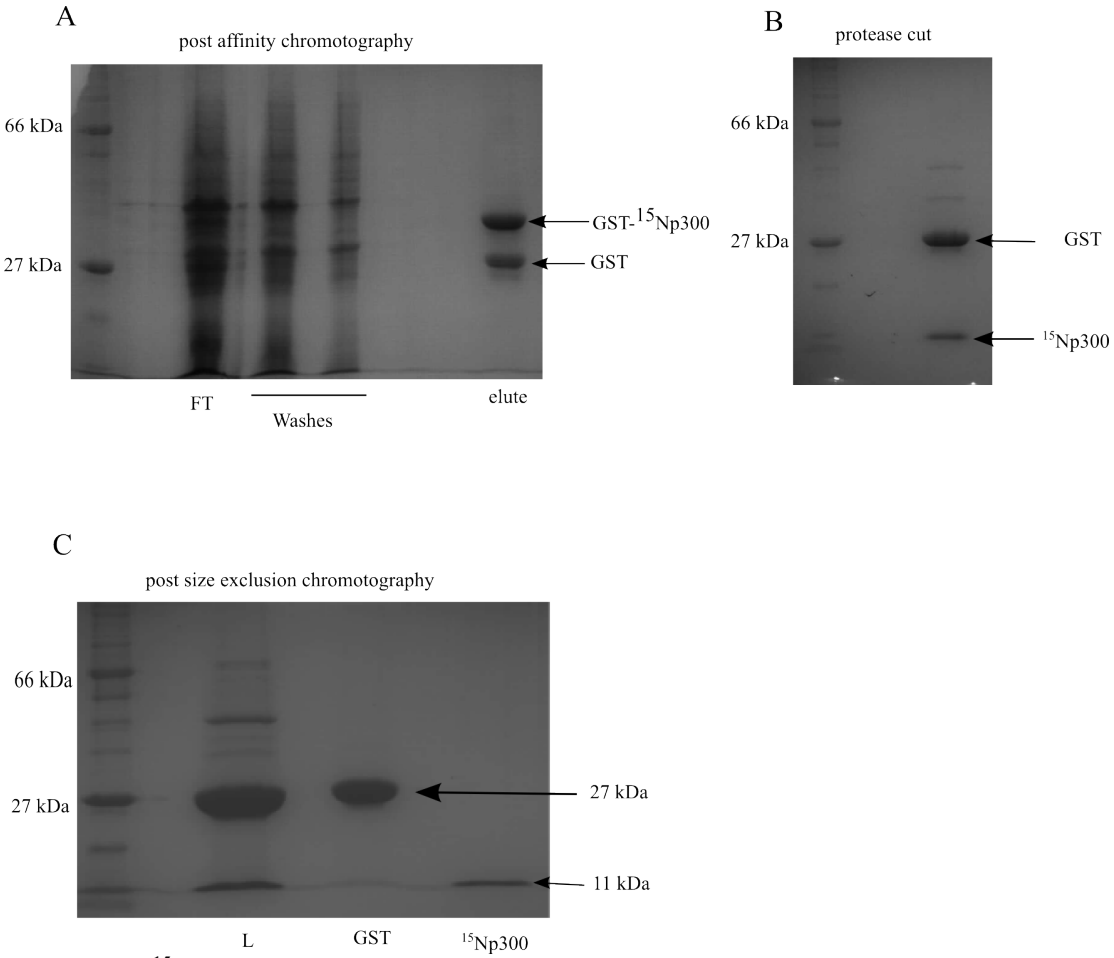


Figure 32 - ^{15}Np300 purification. A) SDS PAGE gel post affinity chromatography showing a band at 37 kDa which is most likely to be GST-^{15}Np300. B) SDS PAGE analysis of the cut GST-^{15}Np300, both GST (27 kDa) and ^{15}Np300 (11 kDa) band can be seen. C) SDS PAGE analysis of the separation of GST (27 kDa) and ^{15}Np300 (11 kDa).

3.3.4 HIF-1α peptide

For much of the biophysical and structural characterisation of the interaction peptides were synthesised by George Burslem (University of Leeds) using Fmoc protected solid phase peptide synthesis (FITC-HIF-1$\alpha_{794-804}$ and FITC-HIF-$\alpha_{816-826}$) or purchased from Proteogenix (HIF-1$\alpha_{786-826}$, HIF-$\alpha_{782-826}$, HIF-$\alpha_{782-793}$, HIF-$\alpha_{782-789}$, HIF-$\alpha_{790-804}$, HIF-$\alpha_{782-804}$, HIF-$\alpha_{794-826}$, HIF-$\alpha_{805-826}$, HIF-$\alpha_{794-815}$, HIF-$\alpha_{805-815}$, FITC-HIF-1$\alpha_{786-826}$ and FITC-HIF-$\alpha_{794-826}$). Both unlabelled and N terminally fluorescein (FITC) labelled peptides were used. However, for *in cellulo* studies, such as cellular FRET a synthesised peptide cannot be used, for this reason HIF-1α peptides of various lengths were cloned with a YFP tag. Yellow fluorescent protein (YFP) was selected as a partner for the GFP-p300 construct. However such clones were never used in the project therefore protein expression was not optimised.

Pure folded p300 protein (as seen by SDS PAGE, mass spectrometry and CD) was produced and was therefore taken forward to characterise the binding to HIF-1α.

3.4 HIF-1α/p300 charcterisation

Knowledge of the native HIF-1α/p300 interaction, including a detailed understanding of how the two proteins interact structurally and biophysically will aid rational drug design. In this section, the work completed to probe the interaction for the discovery of hot-spots will be discussed. Two different approaches were used; analysis of the binding of shorter HIF-1α peptide fragments and phage display experiments. Binding analysis of fragments of the native peptide allow identification of the highest affinity region of the HIF-1α peptide, whereas phage display permits the unbiased exploration of the p300 CH1 protein surface to discover high affinity binders. Two phage display technologies were used[237]: a peptide phage library (NEB[238-240]) and an Adhiron[122] (commercially known as Affimer[241, 242]) phage library. The location and binding mode of phage display derived ligands provides new information on suitable chemotypes for orthosteric small-molecule inhibitor development.

3.4.1 Biophysical assessment of HIF-1α$_{786-826}$/p300 interaction

p300 was cloned and expressed, and CD indicated that p300 is folded. To assess the activity of the p300 protein, binding experiments were conducted; these binding studies included fluorescence anisotropy and ITC. For the initial binding experiments a 41 residue HIF-1α construct was used (C-TAD HIF-1α$_{786-826}$). This construct consisted of the N terminal region which has previously been highlighted as key for binding[220] and both helix 2 and helix 3 which have been highlighted as containing key binding residues[216]. This peptide was purchased from Porteogenix, France. The peptide is shown in Figure 33.

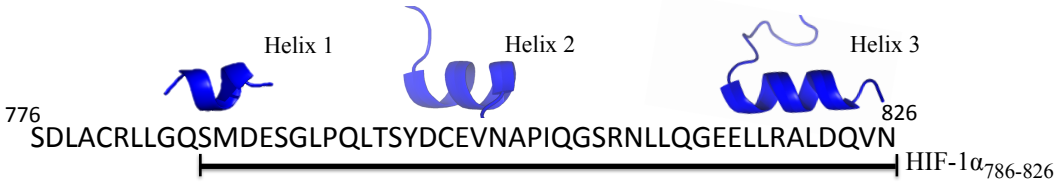


Figure 33 – The HIF -1α$_{785-826}$ construct used for the direct binding.

The direct binding of HIF-1α$_{786-826}$ to p300 was initially tested using fluorescence anisotropy. p300 was serially diluted into labelled FITC-HIF-1α$_{786-826}$ (40 nM) and incubated for 30 minutes at room temperature . The K_d for this interaction was measured as 16.11 ± 0.06 nM. The K_d reported in the literature for HIF-1α$_{776–826}$ binding to p300 is 189 ± 45 nM[220]. The difference in the K_d can be attributed to construct and assay variation. A control fluorescence anisotropy experiment was conducted where GST was serially diluted into labelled FITC-HIF-1α$_{786-826}$ (40 nM) and incubated for 30 minutes at room temperature, no change in anisotropy was measured up to 10 μM GST, indicating no binding of FITC-HIF-1α$_{786-826}$ to GST.

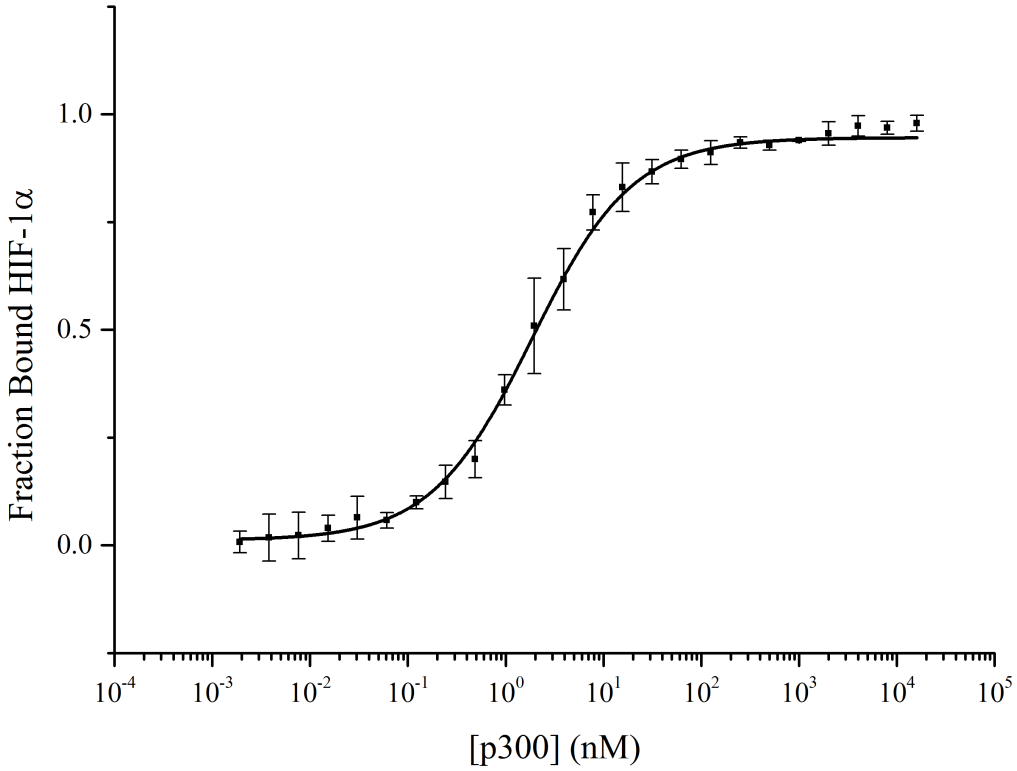


Figure 34 - Fluorescence anisotropy binding analysis of the HIF-1α/p300 interaction expressed as % bound. The plate was incubated for 30 minutes at room temperature.

A time course experiment was conducted to investigate the importance of incubation time (Figure 35). The experiment was conducted in the same way as above, however the plate was read after 30 minutes, after 5 hours and after 18 hours. As can be seen from Figure 35 there is little difference in binding over the time period tested (K_d 30 minutes 693.5 ± 95 nM, 5 hours 197.4 ± 48.6 nM and 16 hours 151.8 ± 10.3 nM).

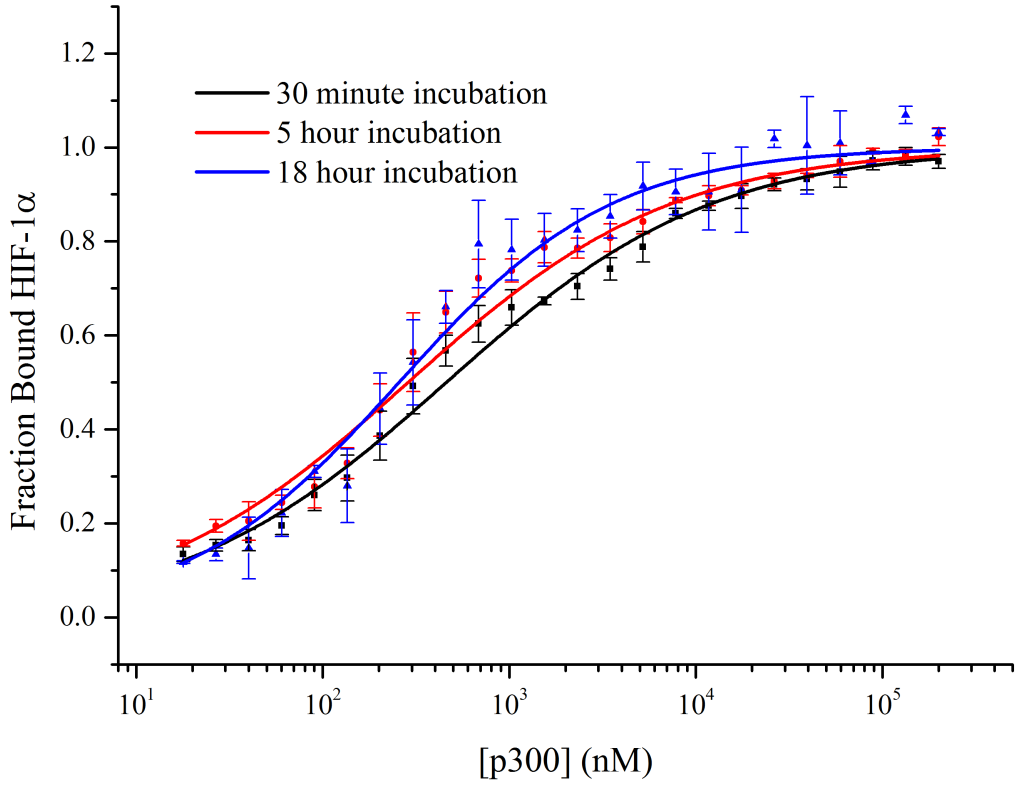


Figure 35 – Time course of plate incubation for fluorescence anisotropy analysis. Plate incubated between 30 minutes and 18 hours for the HIF-1α/p300 interaction.

The effect of incubation temperature, between 25°C and 40 °C, on binding was also investigated (Figure 36). The experiment was conducted in the same way as previously described, however rather than at room temperature the incubation was conducted at a controlled temperature. It can be seen from Figure 36A that the temperature had an effect on the overall anisotropy, which can be attributed to the amount of kinetic energy in the system, however temperature appears to have minimal effect on affinity (K_d, s: 25 °C = 384.1 ± 131 nM, 30 °C = 290.9 ± 97.5 nM, 35 °C = 401.2 ± 89.5 nM and 40 °C = 291.0 ± 106.9 nM).

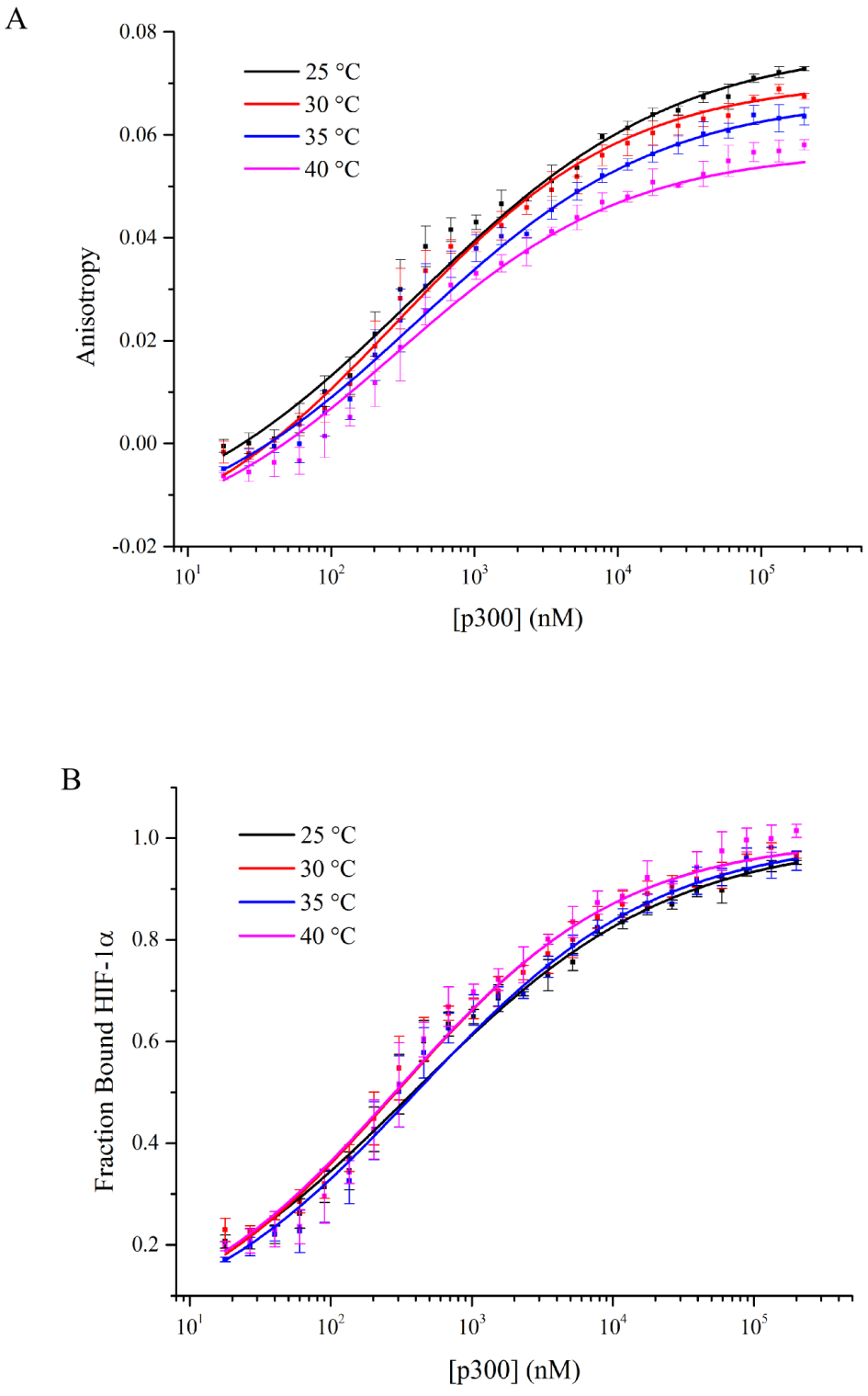


Figure 36 – Effect of temperature on fluorescence anisotropy analysis, plates incubated between 25 °C and 40 °C for the HIF-1α/p300 interaction. A) Data represented as anisotropy. B) Data represented as fraction of HIF-1α bound.

A lack of sensitivity to incubation time and incubation temperature within the specified range is important for a high throughput assay. The more sensitive an assay is the less amenable to robotics it will be.

An orthogonal assay, ITC, was used to confirm the direct binding of HIF-1$\alpha_{786\text{-}826}$. This assay confirms to what extent, if any, the fluorescein impacts upon binding. ITC also gives thermodynamic information regarding the entropic and enthalpic contributions. At 25 °C a K_d of 45 nM was measured; this fits well with the K_d measured by fluorescence anisotropy (at room temperature) and the published affinity. The stoichiometry (N) is 0.61, although slightly low this is likely indicative of 1:1 binding, N is likely to be slightly low if the concentration of active protein does not match the calculated total protein present or there are errors within calculated concentrations. The entropy (ΔS) was measured to be -52.46 $Jmol^{-1}Deg^{-1}$ and the enthalpy (ΔH) was measure to --57.32 $kJmol^{-1}$. These data, are summarised in Figure 37 and Table 17.

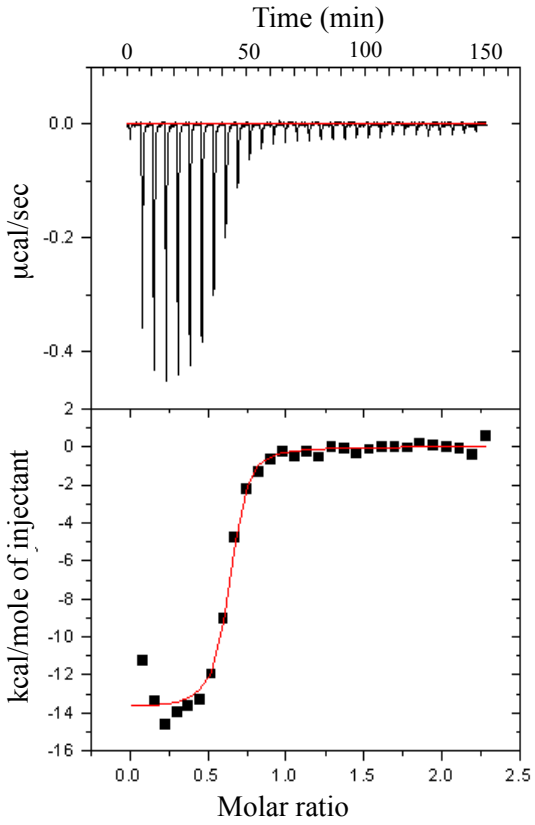


Figure 37 - ITC data for the HIF-1α/p300 interaction at 25 °C.

Table 17 - Summary of the ITC data for the interaction of HIF-α/p300 interaction measured at 25°C

N	0.61
ΔH ($kJmol^{-1}$)	-57.32
ΔS ($Jmol^{-1}Deg^{-1}$)	-52.46
K_d (nM)	45

The K_d of an interaction is dictated by the Gibbs free energy of binding (ΔG) (Equation 7), which results from the ultimate balance of desolvation and attractive forces, the more negative ΔG the more favorable the interaction. ΔG is made up of two components, enthalpy (ΔH) and entropy (ΔS) (Equation 8). The energies of all bonds that are made and broken during the process contribute to the enthalpy change, whereas all ordering and/or disordering processes in the protein, peptide/compound and solvent contribute to the entropy change[243]. The negative enthalpy for the HIF-$1\alpha_{786-826}$/p300 interaction indicates that the interaction is enthalpically driven, on a simplistic level this means that the binding occurs via specific molecular interactions, such as hydrogen bonds, as opposed to a entropically driven interactions which are more likely to be driven by hydrophobic interactions and solvent ordering effects[243].

$Ka = e^{-\Delta G/RT}$ **(Equation 7)**

$\Delta G = \Delta H - T\Delta S$ **(Equation 8)**

Both of the biophysical assays used, fluorescence anisotropy and ITC, measured the affinity of the HIF-1α/p300 interaction to be low nanomolar affinity. The ITC has confirmed the interaction to be enthalpically driven. To use rational based drug design the interaction interface had to be explored further, as the experiments conducted to measure overall binding affinity offer no information regarding the location of binding hot-spots. This information was ascertained in two ways; firstly

by the binding analyses of shorter fragments of the HIF-1α peptide constructs and secondly by phage display experiments.

3.4.2 Exploration of the binding interface using HIF-1α peptide fragments

HIF-1$\alpha_{786-826}$ which was measured to have a binding affinity of 16.11 ± 0.06 nM by fluorescence anisotropy, was broken up into potential key binding areas and peptides of such regions were either synthesised by George Burslem (University of Leeds) (FITC-HIF-1$\alpha_{794-804}$ and FITC-HIF-$\alpha_{816-826}$) or purchased from Proteogenix (HIF-1$\alpha_{786-826}$, HIF-$\alpha_{782-826}$, HIF-$\alpha_{782-793}$, HIF-$\alpha_{782-789}$, HIF-$\alpha_{790-804}$, HIF-$\alpha_{782-804}$, HIF-$\alpha_{794-826}$, HIF-$\alpha_{805-826}$, HIF-$\alpha_{794-815}$, HIF-$\alpha_{805-815}$, FITC-HIF-1$\alpha_{786-826}$ and FITC-HIF-$\alpha_{794-826}$). The length of the peptides are summarised in Figure 38. Each of these peptides was assessed for affinity to p300 to ascertain which regions are more important for binding.

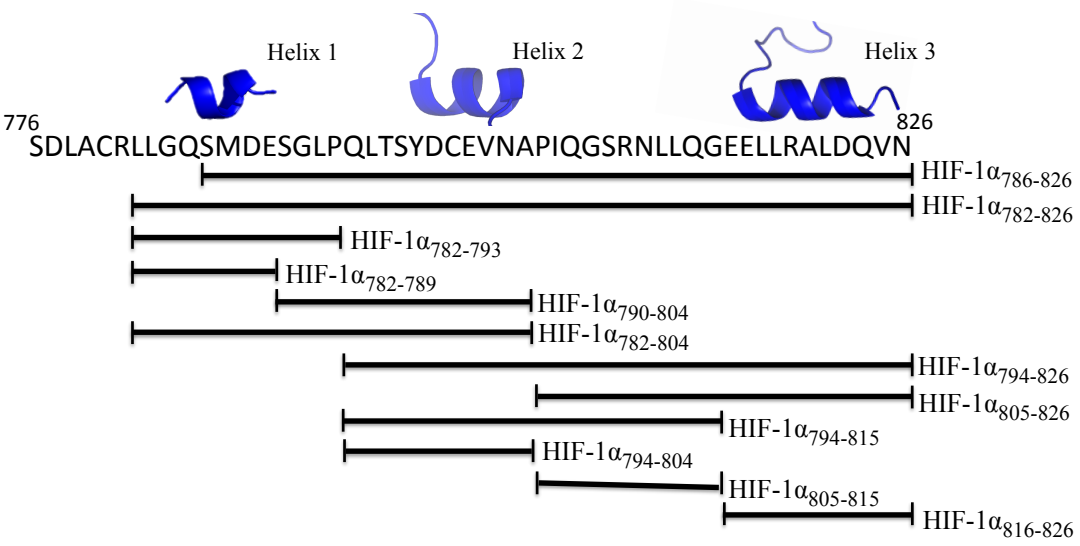


Figure 38 - Schematic showing the peptide HIF-1α fragments purchased to investigate binding hot spot regions on HIF-1α

α-helices have been highlighted as major binding hot-spots of PPIs[20, 96, 97], so initial assessment centered on identification of the highest affinity helix. The affinity of helices 2 and 3 was investigated by fluorescence anisotropy using FITC-HIF-1$\alpha_{794-804}$ and FITC-HIF-1$\alpha_{816-826}$ (**Figure 39**a). Both peptides showed weak binding to p300

(compared to 16 ± 0.06 nM for HIF-1$\alpha_{786-826}$): FITC-HIF-1$\alpha_{816-826}$ K_d approximately 200 μM and FITC-HIF-1$\alpha_{794-804}$ K_d > 1 mM. Notably, FITC-HIF-1$\alpha_{816-826}$ had a higher affinity than FITC-HIF-1$\alpha_{794-804,}$ indicating that helix 3 has a higher affinity than helix 2; therefore helix 3 may be important for high affinity binding of the native complex. To examine further the binding energy contributions of the individual helices to the interaction and the effect of linking, peptide FITC-HIF-1$\alpha_{794-826,}$ comprising helices 2 and 3 with the intervening linker region, was also tested in the fluorescence anisotropy assay. A K_d of 6.74 ± 0.54 μM was measured, showing that linking helices 2 and 3 together results in an increase in affinity for p300 compared with either individual helix. Linking can increase binding affinity either through enhanced local concentration of each binding entity (avidity or chelate effect), or through co-operative interaction between them mediated through allosteric conformational change in the binding site. To distinguish between chelate and allosteric co-operativity, the fluorescence anisotropy assay was used to test the binding of FITC-HIF-1$\alpha_{816-826}$ in the presence of HIF-1$\alpha_{794-804}$ (unlabeled). This demonstrated that the presence of HIF-1$\alpha_{794-804}$ had no effect on the affinity of FITC-HIF-1$\alpha_{816-826,}$ suggesting either an avidity (chelate) effect or a direct contribution from the linker, rather than allosteric co-operativity between helix 3 and helix 2 (**Figure 39**b). The avidity effect is modest, however, as the affinity measured for FITC-HIF-1$\alpha_{794-826}$ implies less than additive contributions from the individual binding energies of helices 2 and 3. This is consistent with the linker sequence having a considerable degree of flexibility when not bound to p300.

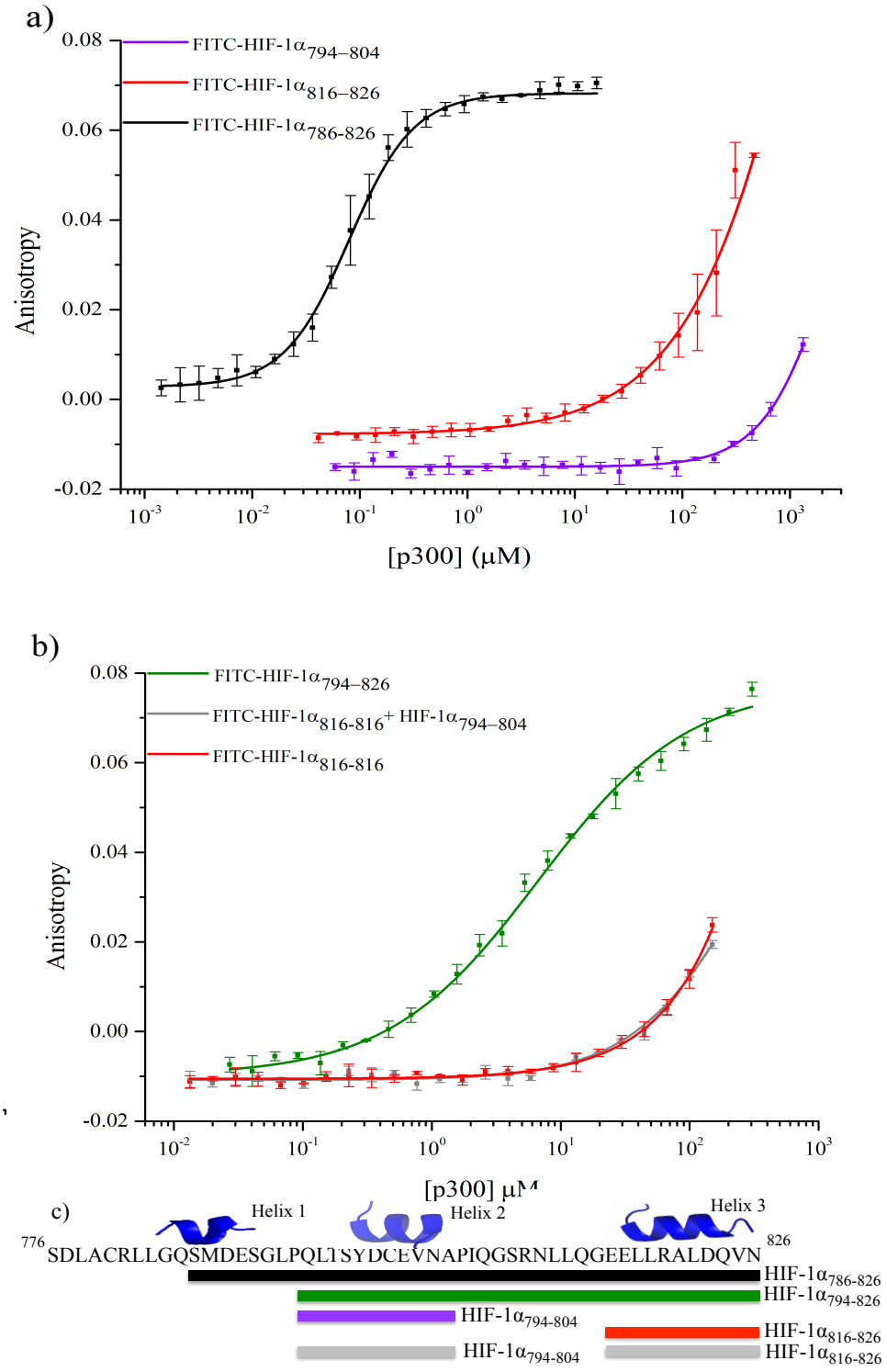


Figure 39 - Fluorescence anisotropy investigation of native HIF-1α fragments binding. a) Binding of the fluorescein labeled helical regions of HIF-1α C-TAD, FITC-HIF-1$\alpha_{794-804}$ (violet) FITC-HIF-1$\alpha_{816-826}$ (red) to p300 CH1 compared to FITC-HIF-1$\alpha_{786-826}$ (black). b) Assessment of the co-operativity the helical regions of HIF-1α. FITC-HIF-1$\alpha_{781-816}$ (green), FITC-HIF- 1$\alpha_{816-826}$ (red) and FITC-HIF-1$\alpha_{816-826}$ in the presence of unlabeled HIF-1$\alpha_{794-804}$ (brown). Note the overlap of the red and grey data sets. c) Schematic of the peptide fragments using in this experiment.

All of the peptides illustrated in Figure 38 were tested in a fluorescence anisotropy competition assay; however only peptides that contained more than one helix were able to disrupt the HIF-1α/p300 interaction (Figure 40). HIF-1$\alpha_{782\text{-}826}$ inhibited the interaction with the greatest potency (IC_{50} = 0.59 ± 0.05 μM). HIF-1$\alpha_{794\text{-}826}$ showed a greater capability to disrupt the complex than HIF-1$\alpha_{782\text{-}805}$ demonstrating the importance of the C-terminus of HIF-1α and reinforcing the dominance of helix 3.

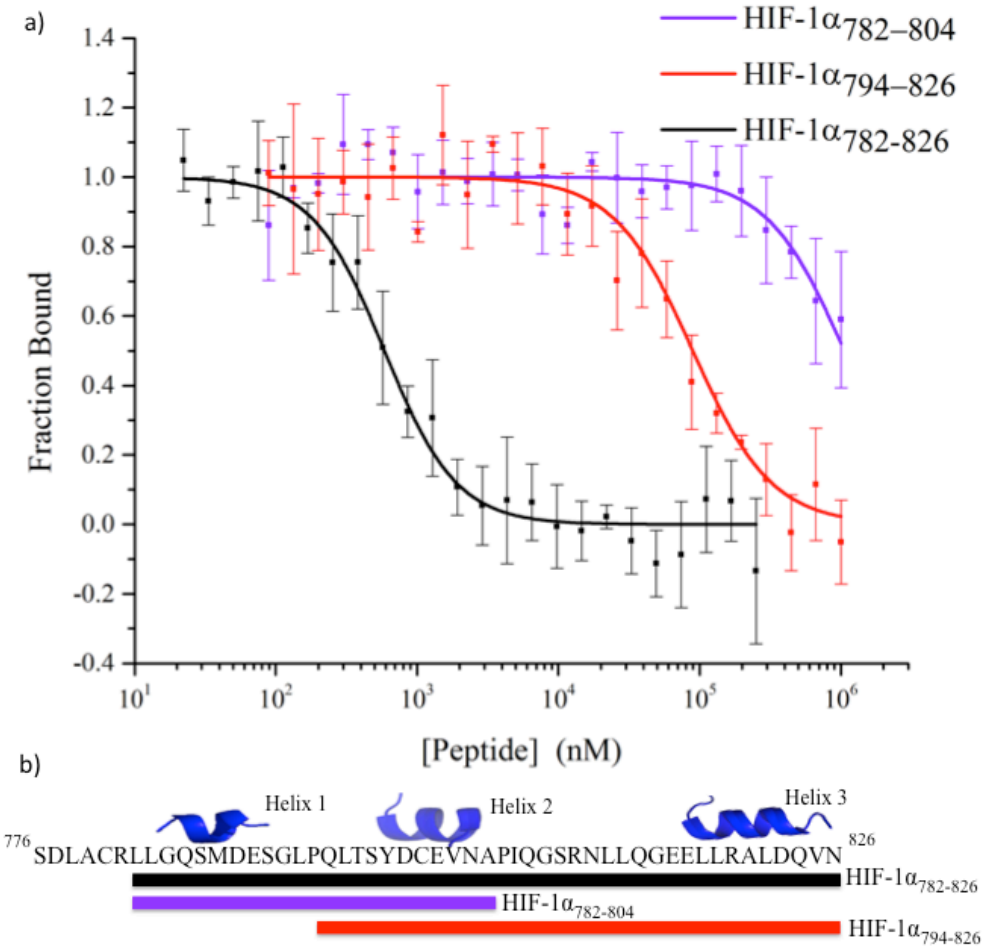


Figure 40 - Fluorescence anisotropy competition assay to test the disruption of the HIF-1α C-TAD/p300 CH1 complex by HIF-1α CTAD fragments. HIF-1$\alpha_{782\text{-}826}$ (black) HIF-1$\alpha_{794\text{-}826}$ (red) and HIF- 1$\alpha_{782\text{-}805}$ (violet).

To further investigate the importance of helix 3, mutants of p300 (H20A, L47M, I71M, Figure 42) in the HIF-1α helix 3 binding pocket (as shown in the NMR

structure[182, 197] Figure 19) were generated and their fold and stability confirmed by CD.

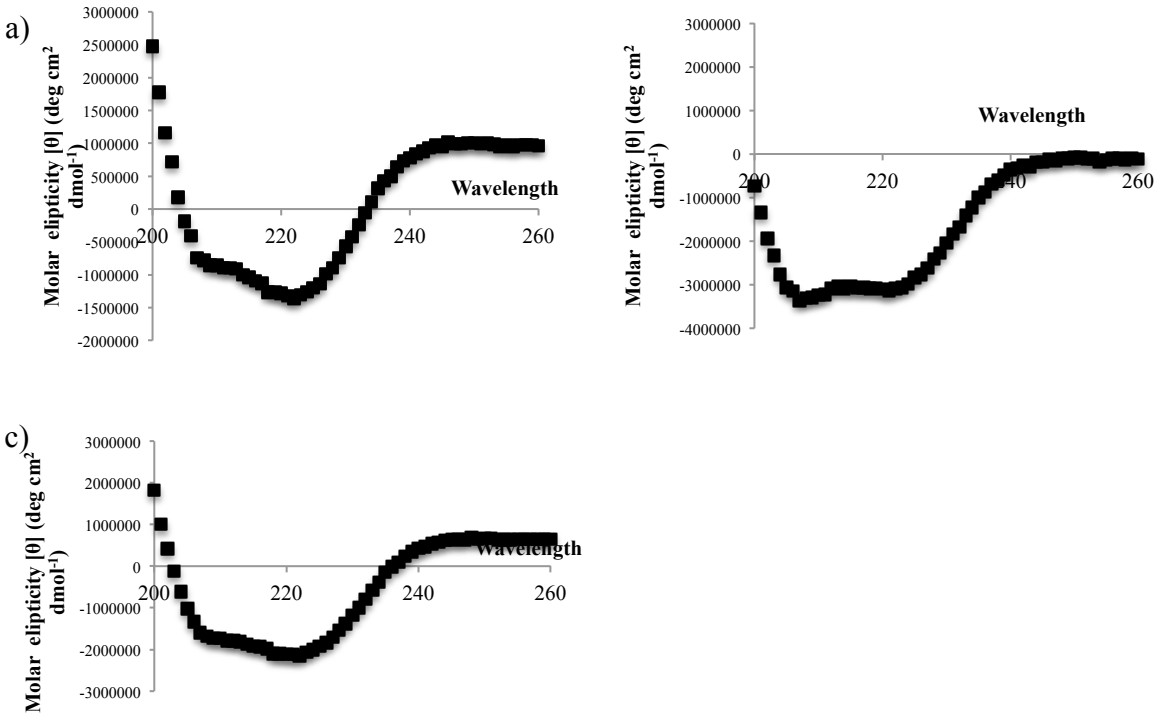


Figure 41 – CD spectrum showing the α-helical secondary structure of the p300 mutants. a)H20A. b) L47M. c) I71M

H20A was selected, as the NMR structure shows this residue makes a contact to HIF-1α, and mutation to alanine would remove this contact. Both L47M and I71M were selected as we hypothesised they would introduce steric clashes. Each of the three mutants caused a reduction in the binding affinity of to FITC-HIF-1$\alpha_{786\text{-}826}$, with the greatest effect observed for I71M, which caused a 18-fold reduction in binding affinity (**Table 18**) confirming that helix 3 has a critical role in driving p300 binding.

Table 18 - Affinity of p300 helix 3 binding pocket mutant proteins binding to FITC-HIF-1$\alpha_{786-826}$ as measured by fluorescence anisotropy

p300	K_d (nM)
Wild type	16.11 ± 0.06
H20A	25 ± 0.35
L47M	29.55 ± 5.58
I71M	285 ± 5.06

Taken together, these data validated helix 3 as the most important of the helical regions of HIF1-α in terms of affinity. This information was taken forward to design small molecules mimicking this helical region.

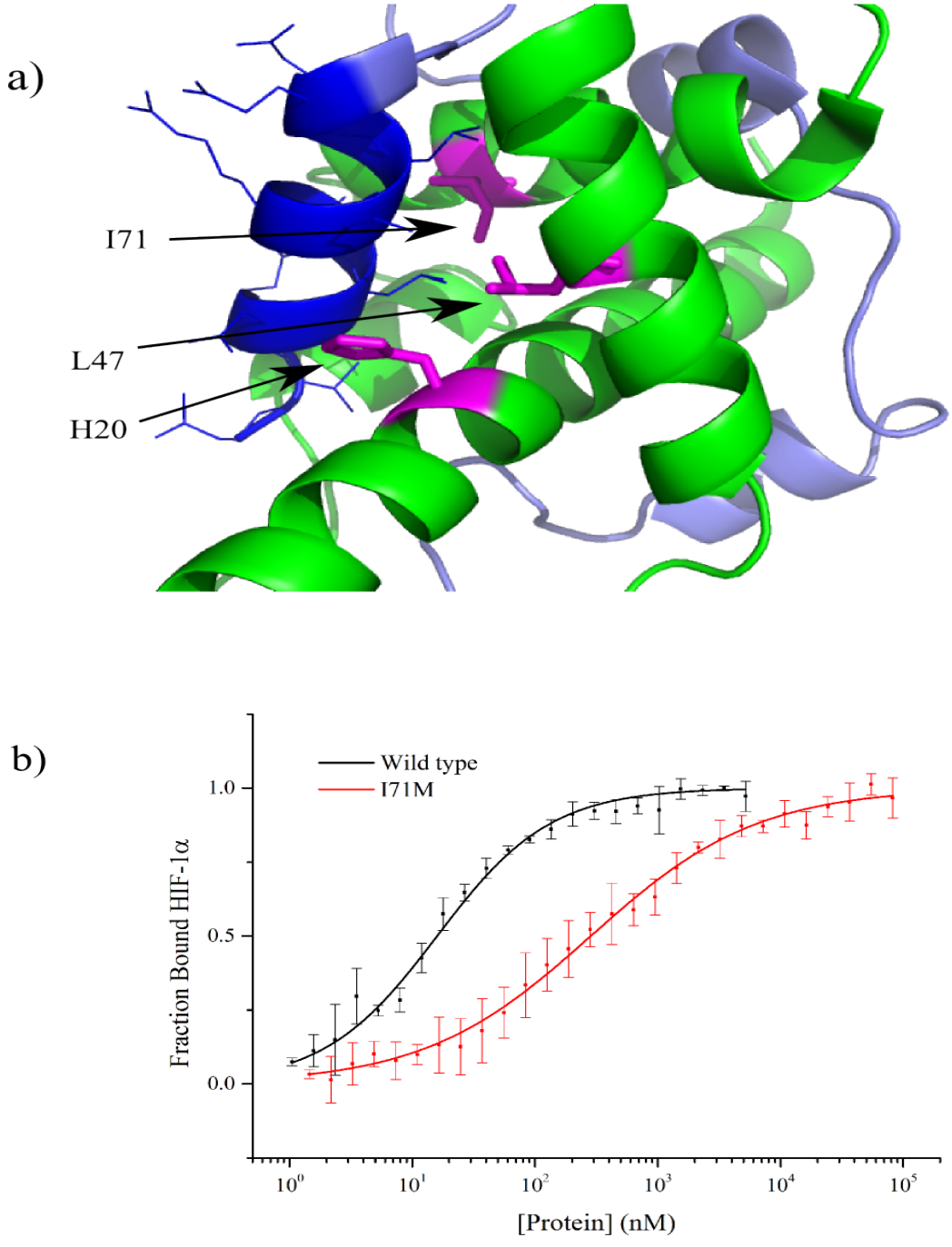


Figure 42 - Investigation of the HIF-1α helix 3 binding pocket on p300. a) Schematic highlighting the three residues of the helix p300 binding pocket of p300 which were mutated. b) Fluorescence anisotropy comparison of the binding of FITC-HIF-1$\alpha_{786\text{-}826}$ to wild-type p300 (black) and mutant p300 I71M (red).

3.4.2.1 Small molecule HIF-1α/p300 inhibitors

As previously discussed the exploration and characterisation of binding a given interface is required to aid rational drug design. Helix 3 has been identified as a key a contributor of HIF-1α binding to p300 via native fragment binding studies. This information can be used in different ways, e.g. by docking of large computational libraries to the helix 3 binding pocket and/or by the design of secondary structure mimics of helix 3. Docking studies identified numerous compounds, which were then purchased and screened. None of the compounds tested disrupted the HIF-1α/p300 interaction with meaningful potency (IC_{50} < 50 μM). Greater success was achieved using α-helix mimetics; this piece of work was published in ChemBioChem in 2014[244].

Oligoamide helix mimetics [98, 245] were designed that mimic the key functionalities and spatial orientation of Helix 3. Helix 3 presents hydrophobic residues (Leu_{818}, Leu_{822}, Val_{825}) along one face of the helix at the *i*, *i* + (3)4 and *I* + 7 positions[182]. A mimetic (p300-oligoamide **1**) was prepared (by George Burslem, University of Leeds) that recapitulates the Leu_{818}, Leu_{822}, Val_{825} residues of Helix 3 (Figure 43A). This compound was tested in the fluorescence anisotropy competition assay and disrupted the HIF-1α/p300 interaction with an IC_{50} = 9.19 ± 0.9 μM (
Figure 43B). This is significantly better than the sequence (EELLRALDVN) corresponding to helix 3 that the oligoamide is mimicking (IC_{50} approximately 200 μM).

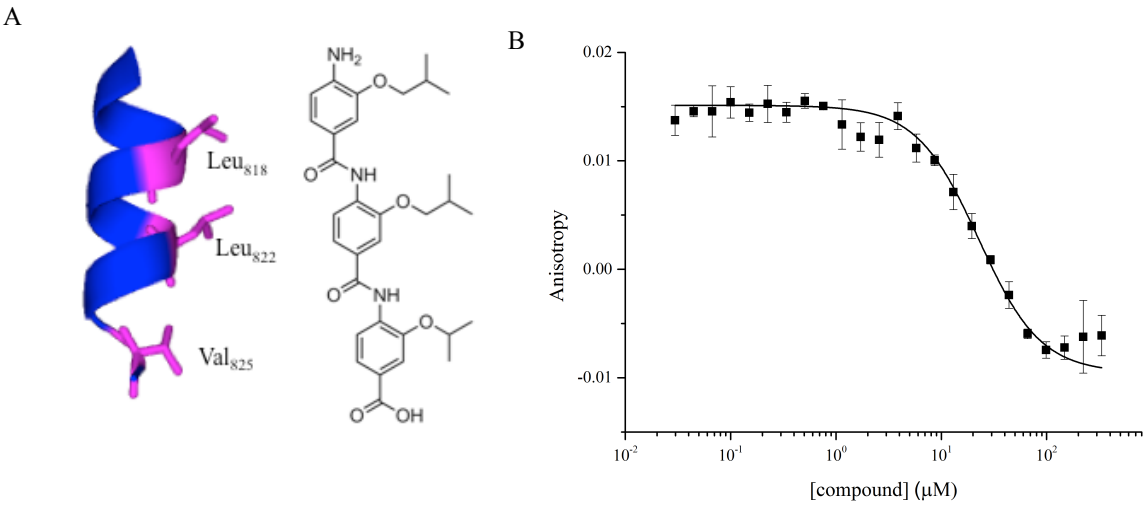


Figure 43 – Design and testing of p300-oligoamide **1** compound. A) The oligoamide presents three key side chains, $Leu_{818,}$ $Leu_{822,}$ $Val_{825.}$ Helix 3 of HIF-1α in blue, key side chains highlighted in magenta. Pymol cartoon. B) Fluorescence anisotropy competition assay measuring the disruption of FITC-HIF-1α/p300 interaction.

A small library of helix mimetics with alternative side chains to probe the relative importance of the side chain size and polarity was then prepared (George Burslem). The compounds and their IC_{50} values are summarised in **Table 19**. The most potent compound was the compound with the matched side chains to native helix 3 (p300-oligoamide **1** IC_{50} = 9.19 μM). Incorporation of larger aromatic side chains (as in p300-oligoamide **6**) in any position had a detrimental effect on inhibition. In addition the structure activity relationship studies suggest that the R1 side chain makes important contacts within the binding cleft through solvophobic effects. This is evidenced by the reduced affinity of p300-oligoamide **3** (because a methyl group was shown to diminish inhibition of the interaction) and p300-oligoamide **10** (because the introduction of polar functionality into a hydrophobic binding site is disfavoured).

Table 19 – Structures and IC_{50} values for the oligoamide compound library, table adapted from[244].

Native helix 3	Ac-GTEELLRALDQVNAAG-NH			Inactive
p300-oligoamide **1**	*i*Bu	*i*Bu	*i*Pr	9.2 ± 0.9
p300-oligoamide **2**	*i*Pr	*i*Bu	*i*Bu	24 ± 1.6
p300-oligoamide **3**	Me	*i*Pr	*i*Bu	216 ± 16
p300-oligoamide **4**	*i*Bu	*i*Bu	*i*Bu	9.8 ± 1.3
p300-oligoamide **5**	*i*Bu	*i*Pr	*i*Bu	13 ± 1.5
p300-oligoamide **6**	benzyl	benzyl	benzyl	56 ± 6.0
p300-oligoamide **7**	*i*Pr	*i*Pr	*i*Pr	39 ± 4.0
p300-oligoamide **8**	*i*Bu	*i*Pr	*i*Pr	17 ± 0.7
p300-oligoamide **9**	benzyl	*i*Pr	*i*Pr	20 ± 0.8
p300-oligoamide **10**	2-hydroxyethyl	*i*Pr	*i*Pr	416 ± 64

The most potent compound (oligoamide **1**, with the matched side chains to helix 3) was also tested against another PPI, eIF4E/eIF4G, to give an indication of selectivity. The compound had a 100-fold selectivity for disruption of HIF-1α/p300 over eIF4E/eIF4G (IC_{50} > 1mM). An N-alkylated oligoamide (Figure 44) as opposed to the O-alkylated scaffold was also synthesized[117, 246] (Valeria Azzarito, University of Leeds) and tested in the assay, however this mimetic did not disrupt the HIF-1α/p300. The magnitude of the decrease in affinity is somewhat surprising and implies that some affinity of oligoamide **1** may be coming form the scaffold itself and not just the side chains.

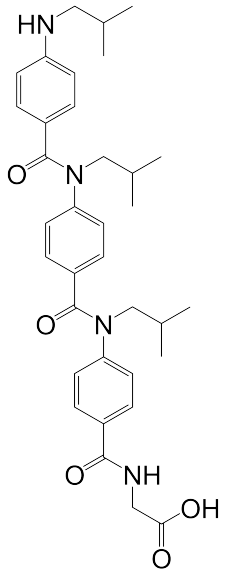


Figure 44 - N-linked oligoamide alpha helix mimetic

Therefore increased understanding of the binding interaction between HIF-1α/p300 directly allowed for the development of molecules that disrupt the interaction. The study of the binding interface highlighted the helix 3 binding region to be a key area for binding, and as a result the mimic of this region yielded low μM inhibitors of the interaction; the first biophysically characterised inhibitors of this interaction[244].

3.4.3 Exploration of the binding interface by phage display

The absence of a dominant fragment of the HIF-1α C-TAD, as seen by the native fragment studies, encouraged us to perform phage display studies to identify short but high potency peptide sequences that may serve as better starting points for small-molecule elaboration.

One of the limitations of using native HIF-1α is that the explored interface is confined to the native HIF-1α sequence. The individual helices and short peptide regions have a low affinity for p300; however it is possible that non-native short peptides are able to bind to p300 with higher affinity than the native peptides. Such peptides can be generated by unbiased phage display. High affinity binders will have multiple uses. Firstly a phage display derived high affinity peptide binder may bind to p300 at a different site to the HIF-1α peptide (which be used to generate potential allosteric inhibitors), this site would have remained unexplored if only the native sequence binding was investigated. Secondly the phage display derived higher affinity peptides may bind in the same hot-spot region identified by the native binding study (in the case of the HIF-1α/p300 interaction, the helix 3 binding pocket); however, as the phage display derived peptides maybe higher affinity modelling their pharmacophore in a small molecule or helix mimetic may yield a higher affinity ligand than modelling the pharamacophore of the native peptide. Phage display may also provide tools that can be used both *in vitro* and *in vivo*.

3.4.3.1 Peptide phage display

Phage display technologies can be used to present many different forms of binder; in this study 7mer and 12mer peptides libraries (NEB[238-240]) were tested. Both libraries were used; the 7mer library should have full coverage (all amino acids represented in all positions[238-240]), however due to the length of the 7mers the generation of high affinity binding peptides was anticipated to be less likely. The 12mer library does not have full coverage, but the longer peptide may yield higher affinity peptide binders. As well as the two phage libraries, two different buffer conditions were used in the wash step: a high sodium chloride (1.5 M) and low sodium chloride (0.5 M) buffer (TBS+0.1% tween) were used to eliminate any nonspecific electrostatic contributions. Elution using native HIF-1$\alpha_{786-826}$ (500 μM) was performed to ensure

elution of specific binders. The experiment was also performed with a different protein (eIF4E); this parallel experiment acted as a selectivity control.

p300 was selectively *N*-terminally labelled with biotin using a biotin-depsipeptide and sortase A (synthesised and expressed by George Burslem and Dan Williamson, University of Leeds)[171] to enable immobilisation on to streptavidin plates (optimisation for this method is outlined in the p300 labelling section (3.3.3)). Three panning rounds were completed. The phage pool was enriched from the unpanned library to round three, although in some cases this enrichment was not sequential (**Figure 45**). The eluents for the unpanned libraries, round 1 and round 3 were sent for next generation sequencing using the Illumina platform[173].

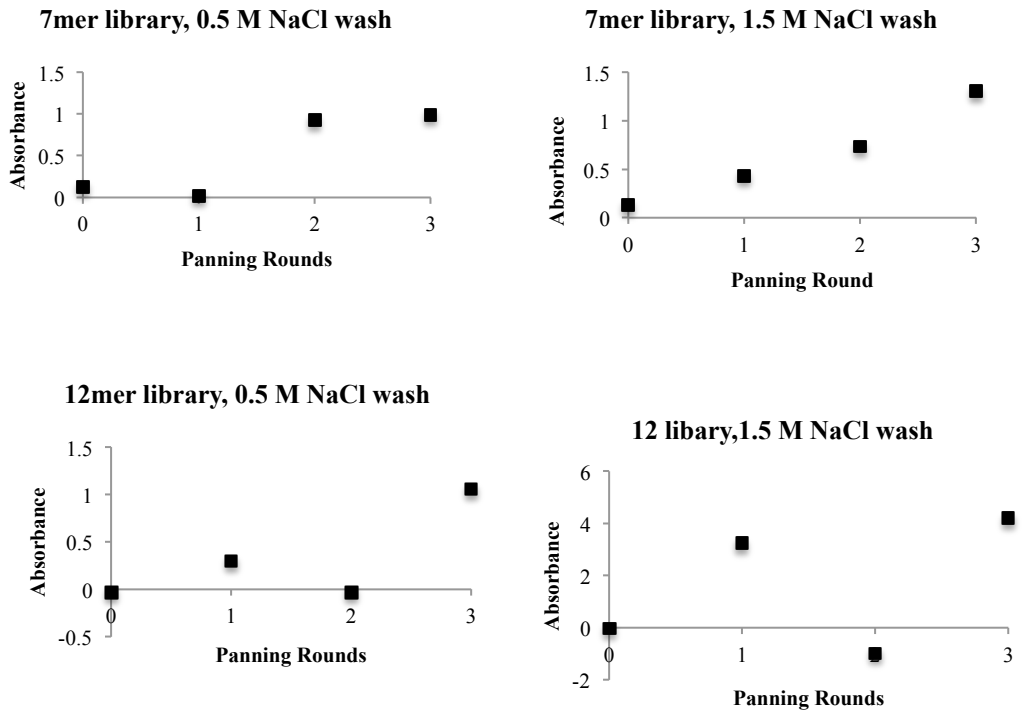


Figure 45 - Enrichment ELISA for each of the conditions of the peptide phage display experiment.

3.4.3.1.1 Peptide phage display clone selection and sequencing

Next generation sequencing was used to limit false positives. False positives have been reported for many phage display experiments. There are two main reasons for the occurrence of false positives; firstly the binding of the phage to other components that are present such as contaminants in the target sample, the solid phase (plastic plates, beads), the capturing reagents (streptavidin, protein A/G, biotin, secondary antibody) or substances used for blocking the solid surface (bovine serum albumin, milk); such binders are classified as selection-related target-unrelated peptides. The amount of selection-related target-unrelated peptides can be minimized by using different or alternate platforms (beads or plates) and substances (blocking or capturing reagents) through the panning rounds. The two wash conditions used during the experiment may also reduce the amount of selection-related target-unrelated peptides. Eluting the phage with an excess of the native binding partner (HIF-1α) will ensure that only the phage specifically bound to the protein will be eluted; however the limitation to this approach is that any phage which are not competitive with the native binding partner or any phage of significantly higher affinity than the native peptide will not be eluted[247]. To circumvent this limitation after the phage were eluted the wells were washed with 0.2 M glycine pH2.2 (which would elute any phage of higher affinity) and these washes were stored a -80 °C, ensuring a potential to sequence those phage of higher affinity then was able to be eluted with the native peptide if deemed necessary after the peptides were sequenced and tested for affinity to p300.

Another class of target-unrelated peptides is propagation-related clones[248, 249]. In this case selection is driven by a faster propagation rate of some phage clones (known as "growth advantage"). Recovery of such clones is independent of their affinity to the target protein (p300) because the advantage in replication enables them to prevail in the phage pool. During the amplification step after each round of panning such phage clones achieve a higher titer. Within a few rounds of panning, these clones may dominate although they are unrelated to the target or any other component of the screening system. This faster propagation of certain phage clones may result from a mutation in the phage genome, which influences the ability of the virus to infect host bacteria or accelerates the process of phage particle assembly. It may also be an

intrinsic property of the displayed peptide itself, without any causative mutation[247, 248]. Certain phage clones replicate slower because displayed sequences impede the phage assembly process to a greater extent. It is impossible to display some peptide sequences on the surface of phage, because they are not compatible with the phage replication process. On the other hand, some sequences allow a relatively faster propagation rate, and such clones are therefore more likely to be isolated. In M13-based libraries (such as the NEB libraries), proline is an example of an amino acid residue that is over represented, whereas cysteine is under represented[250]. This phenomenon is known as sequence bias. Certain peptides which exhibit this growth advantage are frequently reported, for example the 7mer peptide, HAIYPRH, has been selected numerous times against a plethora of different targets. It has been shown that phage clones displaying HAIYPRH amplify at a dramatically faster rate than other clones due to a G→A mutation in the Shine-Dalgarno sequence of p2, a protein involved in the replication process[249]. This mutation increases expression of p2, which results in faster replication of the phage genome and speeds up the entire phage generation process[247, 249, 251].

The propagation related clones are more difficult to control and minimize. Sanger sequencing of clones requires the isolation of DNA from individual phage clones, this is labour intensive and is therefore rarely used to analyse more than 100 clones, which may mean only propagation related clones are sequenced. However with the advance of next generation sequencing (NGS) it is now possible to sequence thousands of inserts in parallel[251]. Illumina platform is currently the leader in the NGS industry and most library preparation protocols are compatible with the Illumina system. Illumina deep-sequencing technology analyses a library of blunt-ended double stranded DNA (dsDNA) fragments and generates up to 10^9 base pair (bp) reads in a single run. Illumina offers the highest throughput of all platforms and the lowest per-base cost[252].

To prepare clones for sequencing, DNA was isolated from pan 1 and 3 elutes and from the unpanned library; PCR was used to amplify the DNA. Blunt end repair of the resultant ds-DNA was conducted using Illumina Paired-end DNA sample preparation kit. The primers with the unique bar codes (Illumina adaptors) were ligated to each fragment and the fragments with the adaptors were amplified by a

second round PCR. This resulted in 12 different pools which each had a unique bar code; the unpanned 7mer and 12mer libraries were also sequenced to assess the presence of any propagation related clones at the start of the experiment. The matlab analysis was performed by Jonathon Stott, AstraZeneca; the scripts were adapted from published scripts[173] and used to separate the sequences into the individual experimental conditions and rank the sequence in order of frequency[173]. The top 5 clones from each of the conditions and their frequency in the pool is summarised in Table 20.

Table 20 – Top 5 peptide phage display sequences for each of the peptide phage display condition

Condition	**Unpanned sequences and frequency**		**Round 1 sequences and frequency**		**Round 2 sequences and frequency**	
7mer library	WSLSELH	1148	DAIPTSV	55	ETALIAA	129
0.5 M wash	TTQVLEA	849	WSLSELH	50	DHAGLQV	126
condition	IDRTQFM	668	QLYREFN	47	TTQVLEA	110
	GTGSQAS	644	TTQVLEA	35	GTGSQAS	107
	SQNFVRE	492	KMISATE	28	NEAPRHA	98
7mer library	WSLSELH	1148	WSLSELH	592	ATNLFKS	35179
1.5 M wash	TTQVLEA	849	TTQVLEA	302	WDPRVNV	934
condition	IDRTQFM	668	IDRTQFM	258	LPVRLDW	821
	GTGSQAS	644	AGPWKSS	217	KVWDTRY	791
	SQNFVRE	492	VQYKPMK	206	KVWEIAR	712
12mer library	GLHTSATNLYLH	3109	SQVYKVAYDWQH	39	VHWDFRQWWQPS	1079
0.5 M wash	EGTSSWRYWLSP	2427	GLHTSATNLYLH	23	SGVYKVAYDWQH	287
condition	ASISNGPLTGYR	1099	SALKGLFPADHH	22	DPVGLGGWWAKV	156
	WPEFDILWAHPQ	304	MIQTNWDKLGLV	19	GTGLVTLPRLTV	92
	AVHLRLDHLSVL	149	SQDIRTWNGTRS	15	DWSSWVYRDPQT	84
12mer library	GLHTSATNLYLH	3109	VHWDFRQWWQPS	150	VHWDFRQWWQPS	2617
1.5 M wash	EGTSSWRYWLSP	2427	SGVYKVAYDWQH	37	SGVYKVAYDWQH	1922
condition	ASISNGPLTGYR	1099	GLHTSATNLYLH	32	GLHTSATNLYLH	1183
	WPEFDILWAHPQ	304	DPVGLGGWWAKV	15	AHHHTFKRLWAH	827
	AVHLRLDHLSVL	149	TENVSAELARSY	15	KLWSLPTSTIDL	624

Analysis of the sequences was conduced (using pLogo, a motif visualisation tool where residue heights are scaled relative to their statistical significance[253].) to assess if there was any strong consensus in the amino acid present. No strong consensus was seen. Three peptides were selected for synthesis (Proteogenix, France): VHWDFRQWWQPS, (phage display derived peptide 1; PDDP1) SGVYKVAYDWQH (PDDP2) and ATNLFKS (PDDP3). The next generation sequencing showed an increase in enrichment of these peptides through the panning rounds (Figure 46).

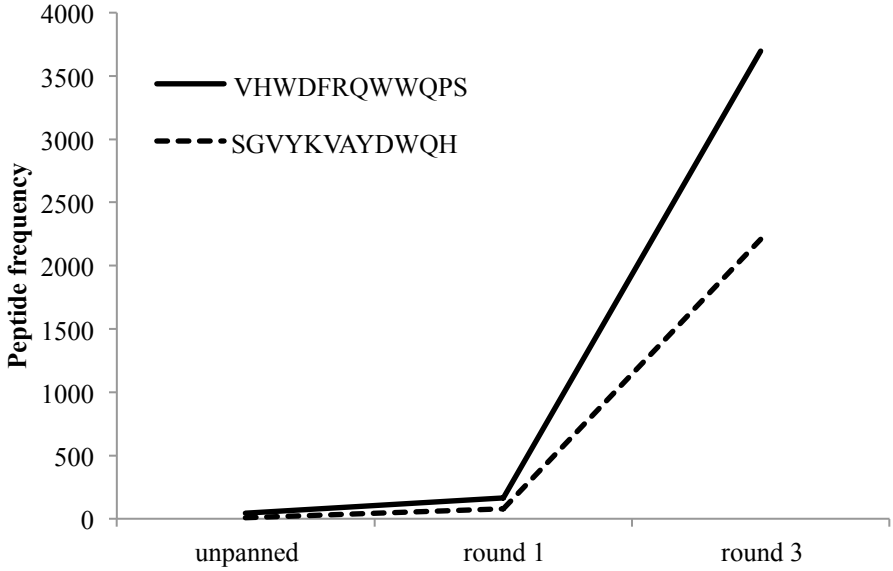


Figure 46 – Increased representation of the 12mer peptides from the unpanned library to round1 and to round 3 as shown by next generation sequencing.

3.4.3.1.2 Phage display derived peptide analysis

Initially the FITC-labelled peptides were tested for binding to p300 using a fluorescence anisotropy assay. The highest affinity peptide was PDDP1 with an affinity of 20.67 ± 3.17 μM (**Figure 47**). Although this is a relatively low affinity for phage display-derived peptides, where low nM and pM binders are often generated[251, 254], the phage display-derived peptide had a higher affinity for p300 than any of the HIF-1α C-TAD peptide fragments of similar length discussed previously and was within the same range as HIF-1$\alpha_{794\text{-}826,}$ the peptide containing both helix 2 and 3 (33 residues compared to 12 residues of PDDP1). PDDP1 showed

selectivity for p300 over eIF4e. PDDP1 is predicted be 35% helical in isolation (as assessed by Agadir[255]), indicating that a helical bound conformation is plausible.

To locate the binding site of the PDDP1 on p300 ^{15}N-^{1}H HSQC experiments were performed; PDDP1 was titrated (300 μM-750 μM) into a solution of ^{15}N-labelled p300 (230 μM). Peak shifts were observed confirming binding (**Figure 47**a). The size of the shift for each peak (**Figure 47**b) was measured and those with the largest shift were mapped on to p300 (assignments from Biological Magnetic Resonance Bank (BMRB): 6268[180]). The location of the shifts indicates that PDDP1 may bind towards the top of the helix 3 binding pocket; the red area highlighted in **Figure 47**c highlights a potential groove for interaction. In addition a shift was observed in the Trp indole peak; a change in the Trp environment has previously been used as an indicator of ligand binding to the helix 3 binding pocket[225].

To further corroborate the binding of the phage display derived peptide in the helix 3 binding site, the binding of the FITC-PDDP1 to both the wild type p300 and p300 with mutations in the helix 3 binding pocket (H20A, L47M, I71M) was investigated using fluorescence anisotropy. The mutant L47M and I71M bound to PDDP1 with a statistically significant reduced affinity ($P<0.05$) (**Table 21**). The reduction in binding affinity complements the NMR data, as the decrease in binding affinity of the mutant indicates that PDDP1 may bind towards the top of the helix 3 binding pocket. The location of L47M and I71M is highlighted in **Figure 47**c (magenta). This additionally highlights the HIF-1α helix 3 binding pocket as a key binding area.

Table 21 - Binding study of p300 helix 3 binding pocket mutant proteins binding to FITC-HIF-1$\alpha_{786-826}$ as measured by fluorescence anisotropy

p300	K_d (μM)
Wild type	20.67 ± 3.17
H20A	57.71 ± 7.87*
L47M	57.76 ± 7.82*
I71M	36.73 ± 11.0

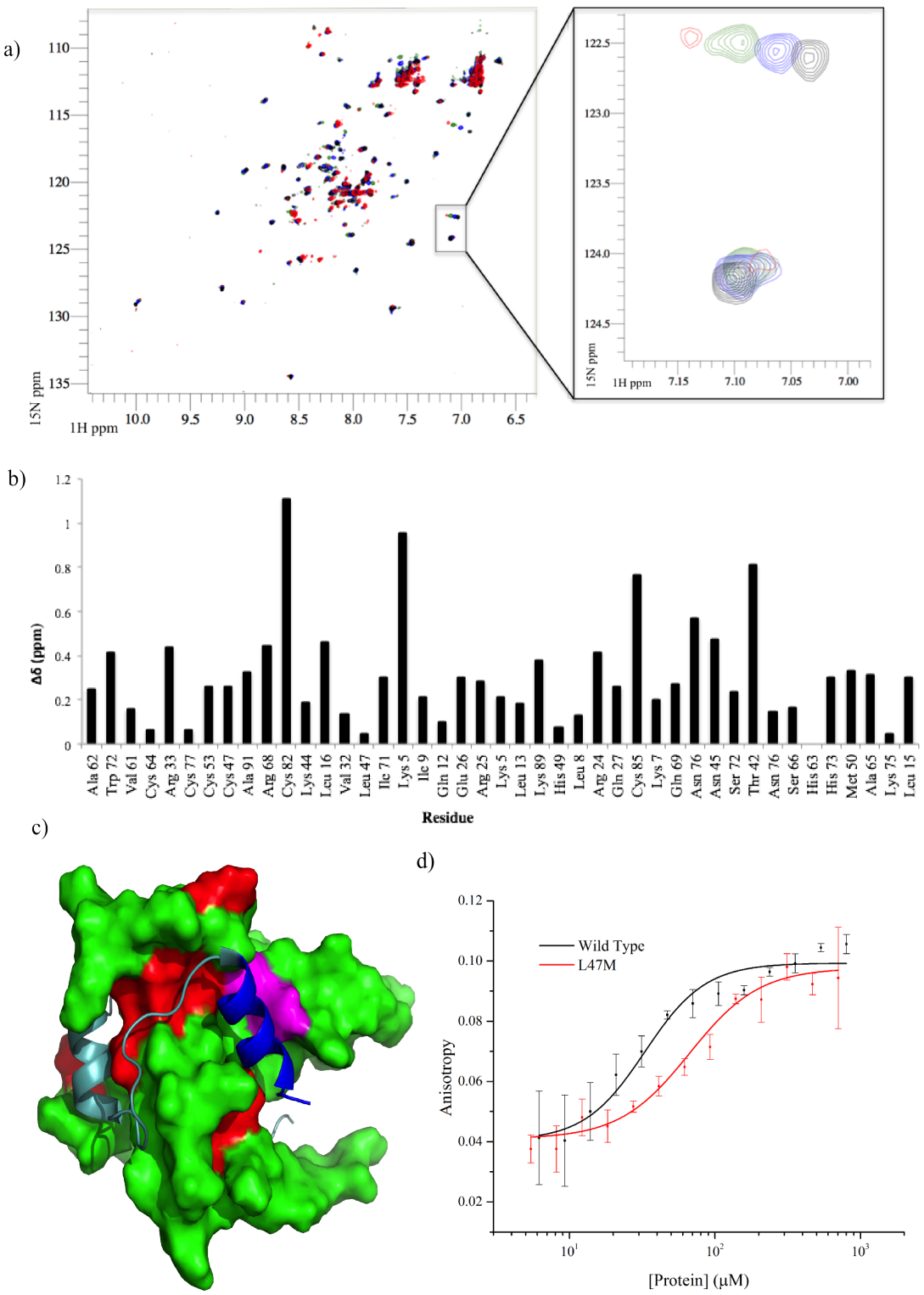


Figure 47 - Analysis of the binding site of PDDP1. a) p300 ^{15}N-^{1}H HSQC experiment, p300 (black) upon titration of PDDP1 at increasing concentrations, 300 μM (blue), 545 μM (green) and 750 μM (red). Insert shows a concentration-dependent shift in one peak and no effect on another peak upon titration of peptide. b) Bar chart showing the size of the shift of all the assigned peaks (BMRB-6268) of p300 after titration of 750 μM PDDP1. c) Mapping of the largest shifts (red) in or around the helix 3 binding site of p300 (green), the helix 3 of HIF-1α is shown in blue with the rest of the peptide shown in teal. The two mutations which cause a statistically significant decrease in affinity (L47M and I71M) are highlighted in magenta. d) Fluorescence anisotropy direct binding measurement of FITC-PDDP1 binding to mutant p300 L71M, the K_d of the mutant is significantly higher then that of wild-type ($p<0.05$).

3.4.3.2 Adhiron phage display experiment

The second phage display experiment used non-antibody binding proteins presented on the surface of the phage as opposed to short peptides; such scaffolds are designed to constrain and present variable peptide sequences for protein recognition[256]. Molecules that bind to proteins with high specificity and affinity are important as therapeutic and diagnostic agents in medicine, and as tools in basic and applied research. The majority of such binders are in the class of monoclonal antibodies, however there are limitations to use of antibodies including high molecular weight, limited tissue penetration, instability, high production cost and potential immunogenicity[257]. Therefore alternatives to antibodies which have favourable biochemical properties and lower production costs are being generated, proteins with a variety of functions have been tested as backbones for affinity molecules by converting them to a library of protein binders following randomisation of amino acids at the binding surface. Thus far 20 different scaffolds have been developed[242, 258-284]; the scaffold used in this study was the Adhiron scaffold (commercially named Affimer[241,242]). Adhirons are engineered non-antibody binding proteins which mimic the molecular recognition properties of antibodies but with improved properties (small, monomeric, thermostable, soluble and easy to express in *E. coli* and mammalian cells giving high protein yield)[122]. The Adhiron library has two randomised loops of 9 residues which are suitable for molecular recognition and are expected to adapt to form appropriate molecular contacts with a wide range of targets including protein pockets, protein surfaces, peptides and small molecules[122].

Four panning rounds were carried out, and an ELISA was used to select those clones to be sequenced (**Figure 48**). Nine clones were sent for sequencing (5, 11, 12, 24, 34, 36, 37, 41 and 43; see **Figure 48**) with four independent sequences present (Table 22). Three Adhirons were taken forward: Ad24, Ad34 and Ad41 (note that Ad41 is represented 6 times in the pool of 9).

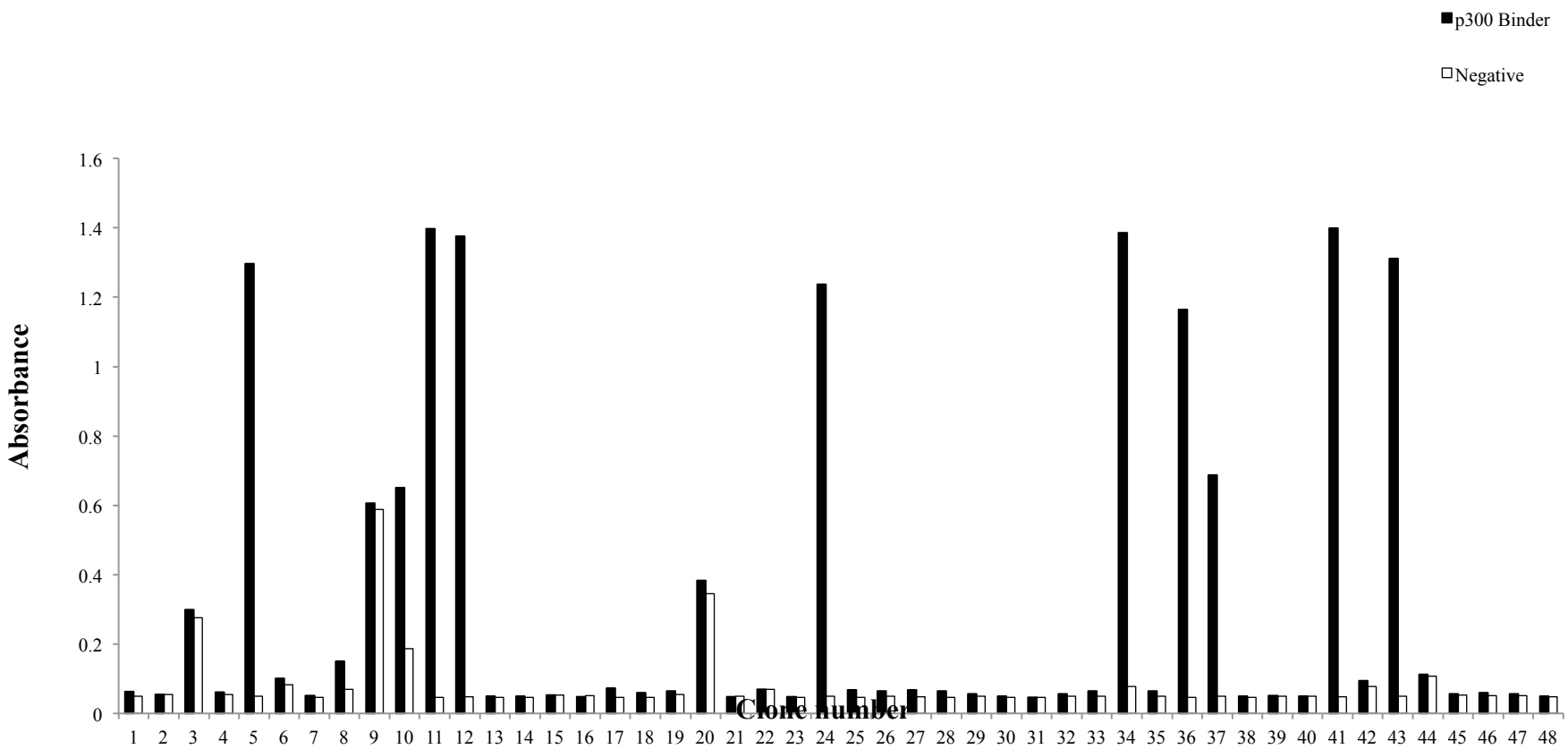


Figure 48 – ELISA results for the selection of clones to be sent for sequencing.

Table 22 – Loop sequences of binding Adhirons

	Loop 1	**Loop 2**
Ad41	AMHPTKNMD	DWGWIDEAY
Ad11	AMHPTKNMD	DWGWIDEAY
Ad12	AMHPTKNMD	DWGWIDEAY
Ad36	AMHPTKNMD	DWGWIDEAY
Ad5	AMHPTKNMD	DWGWIDEAY
Ad37	AMHPTKNMD	DWGWIDEAY
Ad43	PRISGDWEY	HGLYWLPKI
Ad24	PPDLSYYLF	MKSFPHAND
Ad34	ANLYLSRPI	KHIMYYPKT

Each of the three adhirons expressed in a high yield, although gel filtration did show a tendency for the adhirons to form higher oligomeric states (example Figure 49). It is difficult to confidently state that the major peak in each case is monomeric Adhiron as each of the major peaks elutes at a slightly different volume, this may indicate that the adhirons are interacting in someway with the column. As p300 is a hydrophobic protein hydrophobic molecules will be attracted to it, therefore the phage display may have selected for greasy Adhirons more prone to oligomerisation. The major peak in each case was taken forward for further studies.

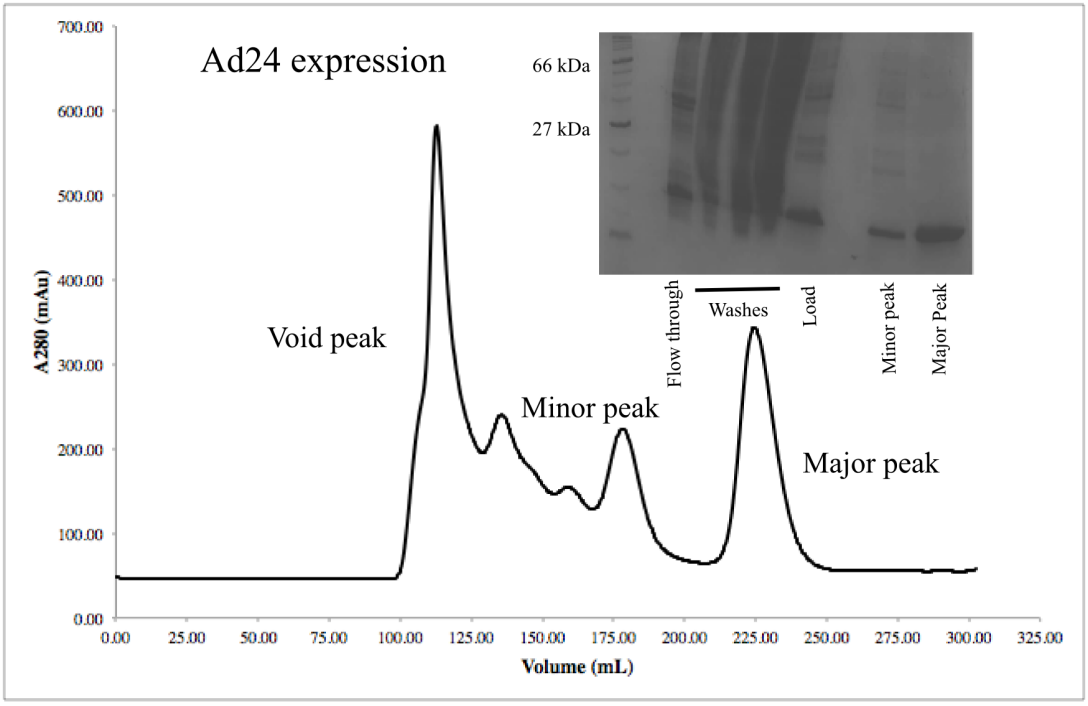


Figure 49 – Expression and purification of Adhrions (Ad 24 as an example), showing size exclusion chromatography trace and SDS PAGE of Ni column washes and elute and minor and major peaks from size exclusion chromatography.

To assess that each adhiron was folded and thermostable the major peak of each Adhrion was assessed by circular dichrosim. Each of the adhirons showed a classical β sheet spectrum[134], with a minima at 220 nM, as would be expected from the published Adhiron crystal structure[122]. The thermal melt data shows that all of the adhiron structures are thermostable.

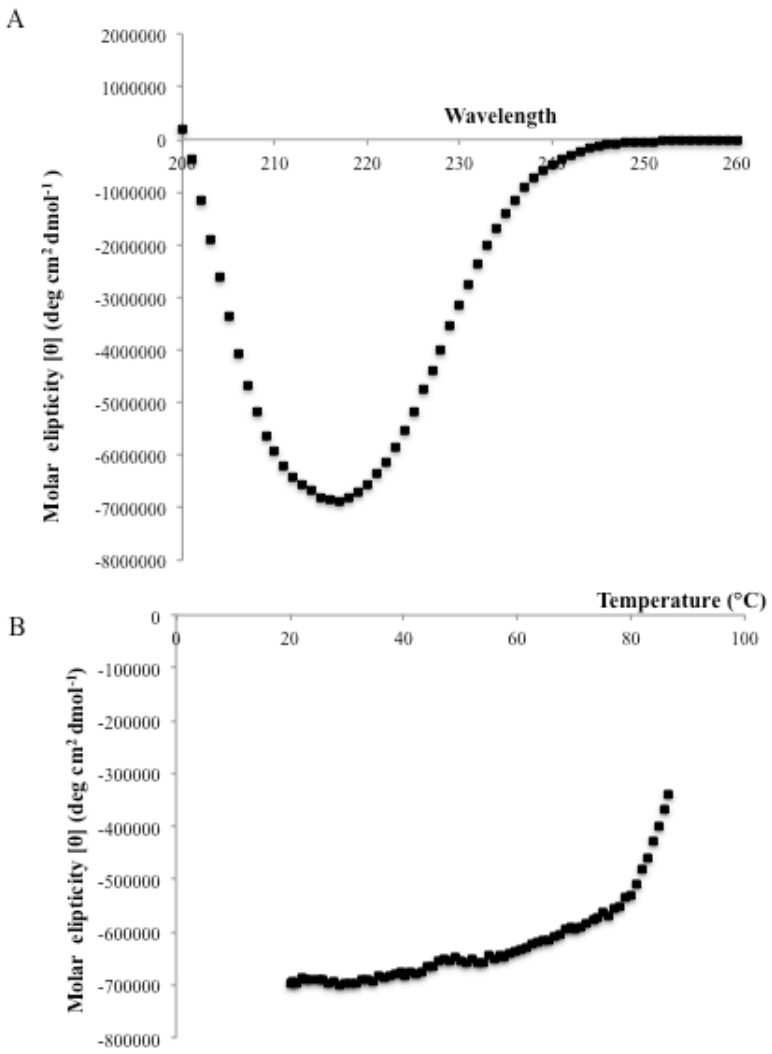


Figure 50 – CD analysis of Ad34. A) spectra run at 20 °C showing a classical β sheet trace. B) thermal melt

The affinity of the Adhirons for p300 was initially tested using BLitz™ (ForteBio), BLitz™ is a dip and read system which enables the real time (kinetic) quantification of molecular interactions in solution. The highest affinity Adhiron was Ad34 with an estimated K_d of 89 nM (Table 23). Therefore this Adhiron was taken forward to confirm binding by SPR, where an affinity of 157 nM (Chi^2 0.114) was measured for Ad34 (Figure 51), calculated from the on and off rates of the interaction.

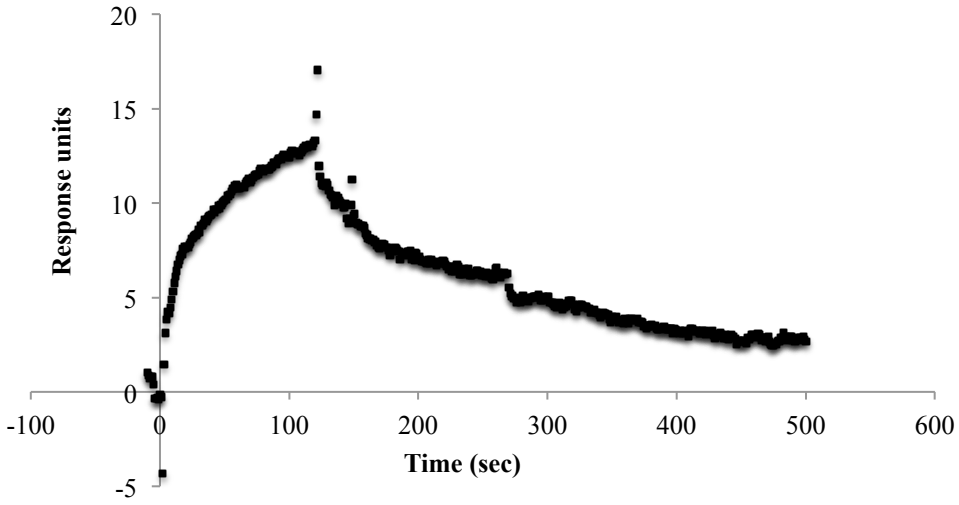


Figure 51 – SPR sensogram of Ad34 binding to biotin-p300

The affinity of the phage display-derived Adhiron, Ad34, for p300 is higher than the phage display-derived peptides and the shorter native peptide sequences and is in the same range as the HIF-1$\alpha_{786-826}$, which has an affinity of 16.11± 0.06 nM. However, Ad34 might be anticipated to bind at a more localised site than HIF-1$\alpha_{786-826.}$ Each of the Adhirons was assessed for their ability to inhibit the HIF-1α/p300 interaction using the fluorescence anisotropy competition assay. All 3 Adhirons had an IC_{50} of 1-5 μM (Table 23, Figure 52). This is significantly better than the fragments of native HIF-1α, or the phage peptides described previously (which were unable to disrupt the interaction). However, the binding site of the Adhiron on p300 is not yet defined, and unfortunately excessive peak broadening in HSQC spectra on addition of Adhirons precludes interpretation of the resulting NMR shifts. These Adhiron sequences do not appear in a selection of Adhirons raised against a variety of other PPI targets, suggesting selectivity.

Table 23 - Phage display derived Adhiron binding data, K_d measured by BLITZ and IC_{50} measured by fluorescence anisotropy competition assay.

Adhiron	K_d (nM)	IC_{50} (µM)
Ad41	105	1.98 ± 0.32
Ad34	89	4.78 ± 2.12
Ad24	140	2.96 ± 0.46

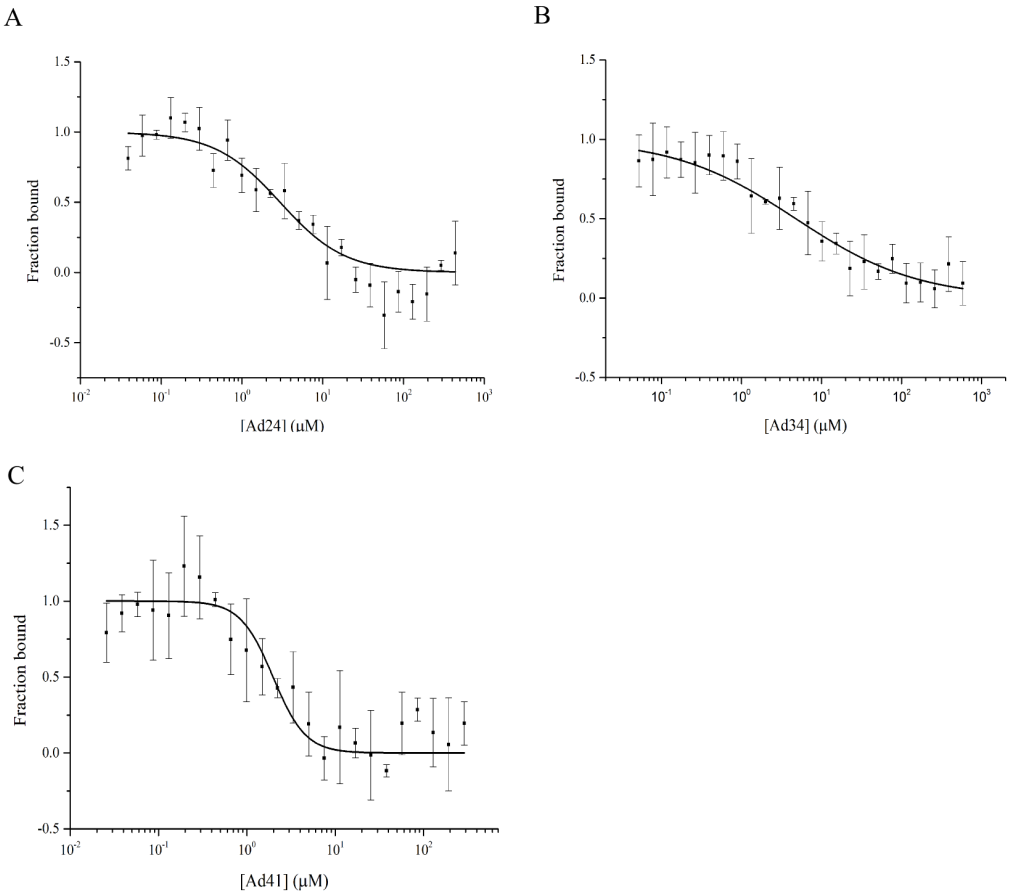


Figure 52 – Inhibition of the HIF-1α/p300 interaction by Adhirons. A) Ad24. B) Ad34 and C) Ad41

To further characterise Ad34 it was crystallised from 0.8 M di-sodium succinate pH 7; crystals were then supplemented with 20% glycerol for cryo-cooling and X-ray diffraction data collection. Data were solved to 2.8 Å. Crystallographic data is summarised

Table 24.

Table 24 - Crystallographic data for the adhiron Ad34

Data set	**Adhiron Ad34**
Wavelength (Å)	0.91741
Space group	I 21 21 21
Cell parameters (Å,º)	a = 69.29 b = 72.64 c = 107.93 α= 90 b = 90 γ = 90
Total reflections	87,748
Unique reflections	7,134
Resolution shells(Å)	
Total (High)	36.73 – 2.73 (2.80 – 2.73)
Rmerge (%) *	0.148 (1.644)
Rpim (%) ***	0.045 (0.477)
Completeness (%)	99.8 (100)
Multiplicity	12.6 (13.8)
I/s(I)	12.3 (1.7)
V_M ($Å^3$/Da)	3.67
Mol. per AU	2
Reflections working set	6,761
Free R-value set (No. of reflections)	373
R_{cryst} (%) ***	0.2388
R_{free} (%) **	0.2781
RMSD bond lengths (Å)	0.0106
RMSD bond angles (º)	1.725
No. of atoms used in refinement	
Non-hydrogen atoms	1,574
Water molecules	48
Mean B value ($Å^2$)	
Total	70.371
Water molecules	82.341
Ramachandran plot statistics (%)	
Preferred region	82.97
Allowed region	11.54
Outliers	5.49

There were two adhiron molecules in the asymmetric unit. Each of these adhirons forms a dimer to an adhiron in the adjacent asymmetric unit. This dimerization appears to be mediated predominantly via one loop binding to a β sheet of the adjacent adhiron by hydrogen bonding. This dimerization could be a crystallography

artefact, however the gel filtration showed a potential for oligermisation, therefore there is a strong potential that Ad34 will be able to form dimers when free in solution and in crystals. Also, the size of the contacts is larger than would be expected for crystal contacts and is more in line with the size expected for dimerisation. The likelihood of Ad34 being a dimer could be further explored using analytical ultracentrifugation and native SDS PAGE. The scaffold could be further optimised to prevent the dimerization. The insertion of charged residues in to the β sheet near the loop regions would potentially repel the adhirons from each other, therefore limiting dimerization.

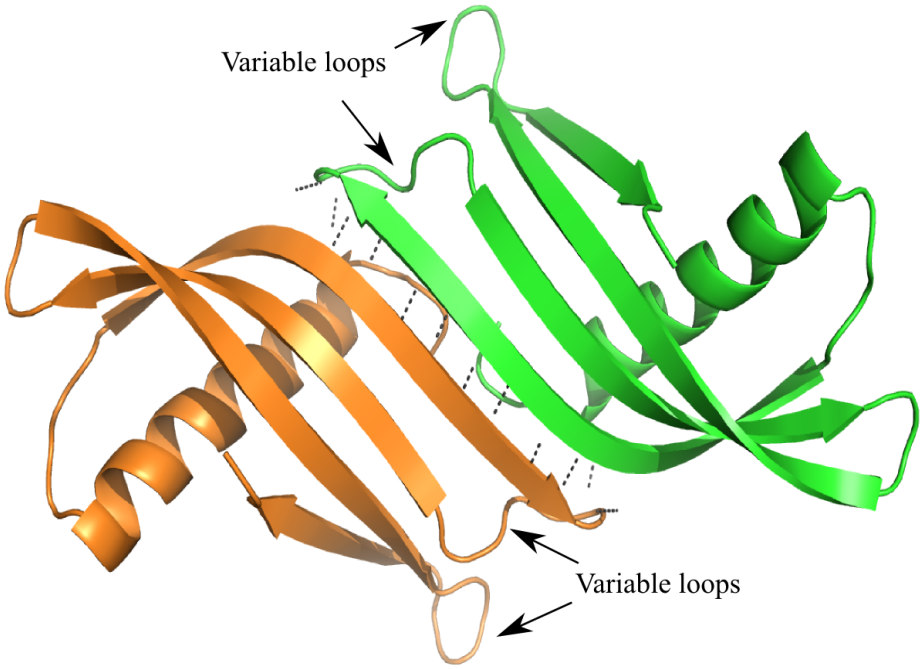


Figure 53 - Dimerisation of Ad34, one molecule per assymetric unit, hydrogen bonding illustrated by dotted lines

The binding mode of the Adhiron was then investigated by *in silico* docking. Ad34 was docked (by George Burlsem, University of Leeds) against the NMR structure of p300 (PDB 1L8C)[182] using HADDOCK (High Ambiguity Driven protein−protein Docking)[181]. The docking results for the highest scored and most frequently predicted complex suggested that both variable loops are involved in binding, with both of the loops buried in hydrophobic areas of p300 (Figure 54). Neither loop docked in the HIF-1α helix 3 region, suggesting that Ad34 may bind to a different

region from the highest affinity phage display derived peptide (PDDP1) and the highest affinity native helix (HIF-1$\alpha_{816-826}$) (Figure 54). The modelling suggests the two Adhiron loops bind in different crevices on p300; these pocket-like structures are in mutual proximity to each over. Given the proximity of the two sites identified, fragment-based approaches may be appropriate to link small molecules that target each site, although further experimental work is required to validate the binding site. The ability of the Adhirons to disrupt the interaction means that the pharmacophore of the adhiron could be modelled in a small molecule compound for the development of small molecule inhibitors. As it is suspected that the Adhiron binding, and therefore inhibitor potency, is mediated predominantly through the variable loop regions it is those loop regions that should be modelled.

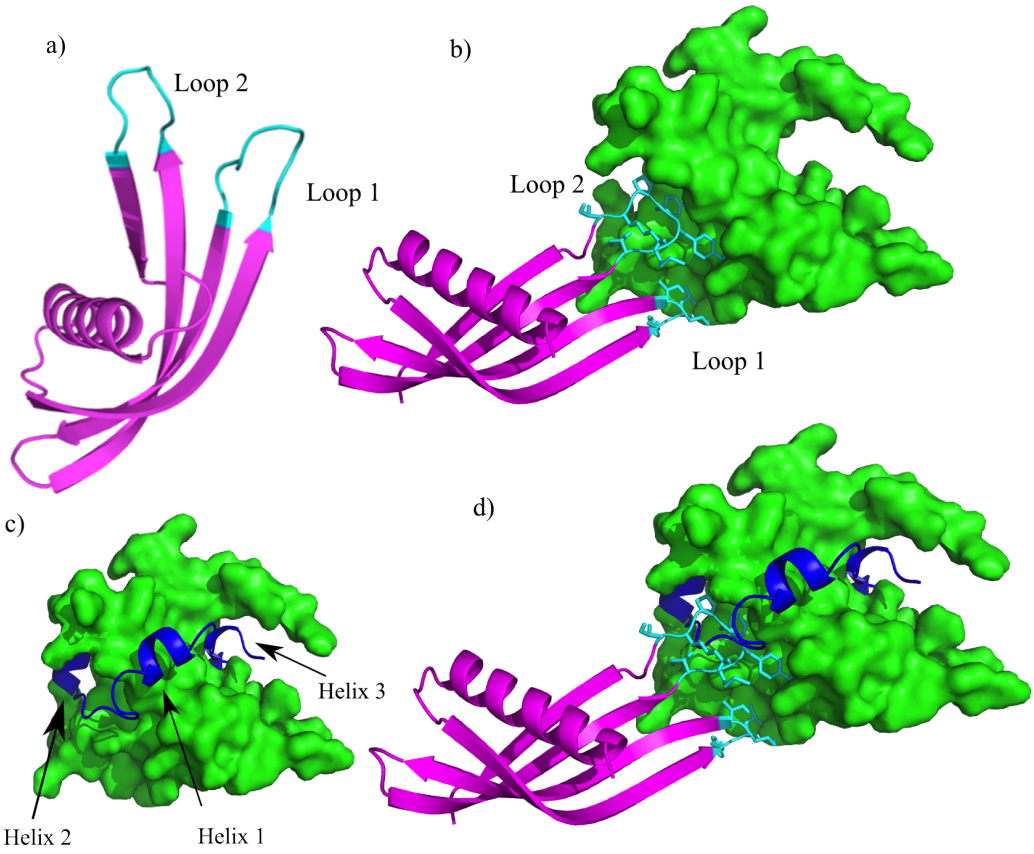


Figure 54 - Adhiron structure and potential binding model to p300. a) Crystal structure of Ad34, variable loops shown in cyan. b) Adhiron Ad34 (magenta) docked to the NMR structure of p300 (green) using HADDOCK; key binding residues are shown in cyan**.** c) NMR structure of HIF-1α (blue) bound to p300 (green) in the same orientation as the docked Adhiron (shown in b). d) Overlay of the docked Ad34 (magenta) and HIF-1α (blue) bound p300 (green), highlighting Ad34 may not dock in a helix binding pocket.

3.4.3.3 Summary of phage display.

This study has shown, like the native binding study, that it is difficult to obtain high affinity binding of small peptides to p300. The Adhirons were measured as the highest affinity binders, although the interaction may be mediated through 2 sites. That being said each loop is only 9 amino acids therefore the overall binding is mediated through a relatively small area, compared to the native HIF-1α. The Adhirons have a higher affinity for p300 than the native peptides containing individual helices, therefore the pharmacophore properties of the loops may be a better model for small molecule design than the native sequence. Certainly if a small molecule is able to span the two sites, in which the adhiron loops bind, potentially by fragment bridging or merging, an exciting new strategy to disrupt the HIF-1α/p300 PPI could be employed. However much further validation of the docking model is required before any such work could be carried out.

The peptide phage display experiment has again suggested the region of p300 where the helix 3 of HIF-1α binds as key for binding. This is supportive of the conclusion drawn from the native sequence analysis. Therefore from all the studies conducted the single highest affinity hot-spot is in the region of HIF-1α helix 3 binding. This also supports the conclusions drawn by Cho *et at*., that the C terminus of HIF-1α is important for binding[220], although the area highlighted has been decreased (in terms of size), aiding drug design.

3.4.4 Structural characterization by crystallography

A high resolution crystal structure of p300 or the HIF-1α/p300 complex would further characterise the binding interface as a higher resolution image will give more information regarding which residues interact. For crystallisation the protein must be a pure, homogenous sample, this was obtained for p300 as outlined in the p300 cloning and expression section (3.3.2.). The initial crystallisation trials for p300 were conducted at 10 mg/mL in the buffer 10 mM Tris pH 7.5, 150 mM sodium chloride and 5 % glycerol. Using commercially available sparse matrix screens, which were designed to sample a wide range of precipitants, salts, buffers, pH, polymers and organic molecules, in a high- throughput 96-well format, and were biased towards previously successful crystallisation conditions for macromolecules. The screens were set up in sitting-drops with various protein:mother liquor ratios across all conditions and at a range of temperatures. No crystals grew and very little protein precipitation was seen, indicating that a condition needed to be changed. To grow protein crystals, an undersaturated solution of protein and precipitant must be concentrated to supersaturation, i.e. beyond the concentration where the protein is soluble in the defined mixture of other solutes (the solubility line), but not allowed to reach the point of precipitation (disordered aggregation). The concentration of precipitant changes the position of the solubility line at a given protein concentration. Figure 55 shows a solubility phase diagram, illustrating the relationship between protein and precipitant concentrations – which varies depending on the specific protein and precipitant. Spontaneous nucleation (ordered aggregation) of crystals occurs in the labile zone, which reduces the protein concentration and moves the phase back into the metastable zone, where sustained crystal growth can occur. Crystals will continue to grow in the metastable zone until either the protein is depleted, or until impurities are incorporated into the lattice at the surfaces of the growing crystal[285]. The most common method used, and the method used in this case, for crystal growth is vapour-diffusion: The evaporation of water from a protein:crystallant drop as it approaches equilibrium with a reservoir of crystallant (in a closed system), concentrates both protein and crystallant to the point where nucleation and crystal growth can occur. Crystallant solutions (or 'mother liquor') typically contain combinations of precipitants (organic polyalcohols e.g. polyethylene glycol (PEG)), salts (e.g. ammonium sulphate or sodium chloride),

additives (e.g. detergents, metal ions, reducing agents etc.) and buffers. Protein:crystallant drops are either sat on platforms over a reservoir (sitting drop) or suspended below a coverslip over a reservoir (hanging drop) and sealed. As concentrations of free protein and precipitant increase, the phase moves from undersaturated to supersaturated. The supersaturated labile zone, where spontaneous nucleation occurs, is followed by the metastable zone, which supports sustained crystal growth. The precipitation zone results in disordered aggregation and no crystal growth[285].

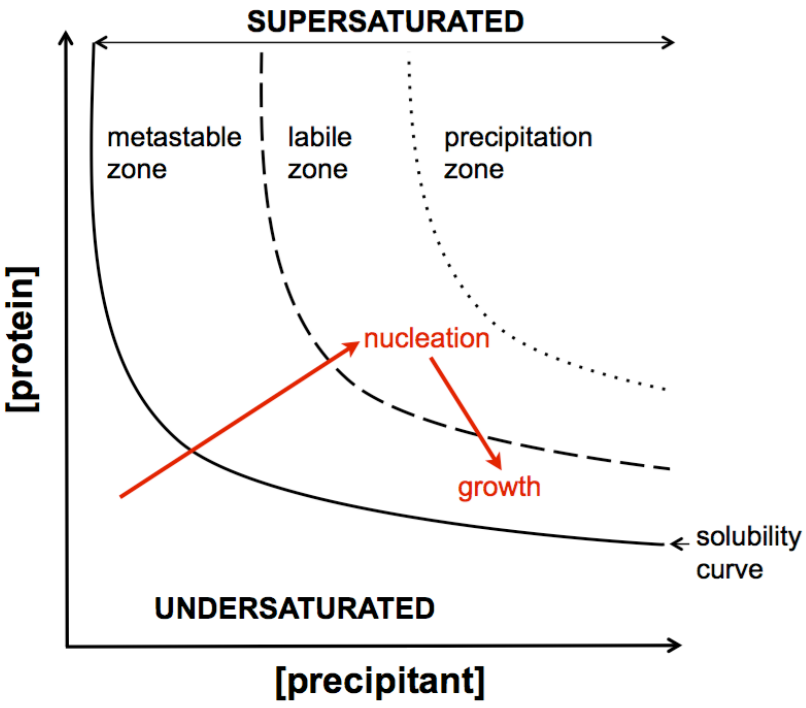


Figure 55 - The solubility phage diagram

As no crystals grew and very little protein precipitation occurred p300 was deemed to be in an undersatured state. To push p300 towards the labile zone the concentration of protein was increased. The sparse matrix screens were repeated at numerous concentrations up to 50 mg/mL, however even at this high concentration of protein very little protein precipitation was seen and no protein crystal grew. Therefore different buffers were investigated to try and alter the stability of the protein. Such buffers included ammonium acetate, Hepes and Tris at various pH,

ranging from pH 6 –pH 9. The sparse matrix screens were repeated at each of these buffers, still very little protein precipitation occurred.

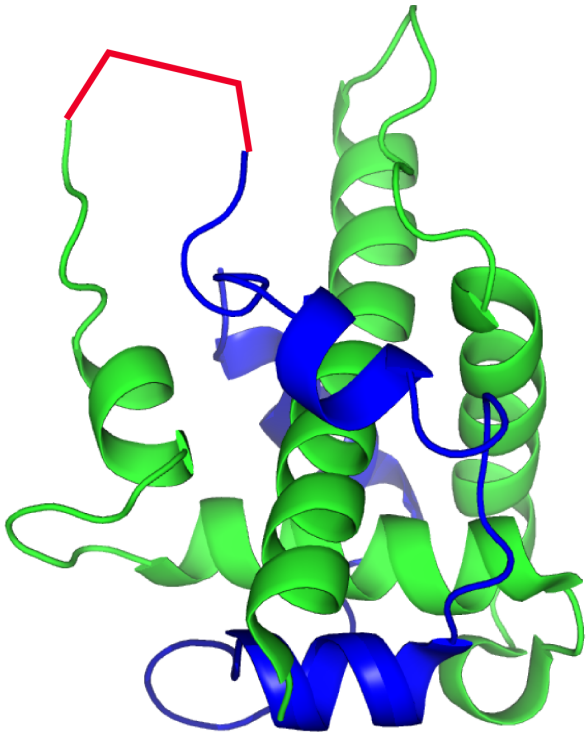
Another attempt to try and change protein stability was to conduct the sparse matrix screens under oil[286, 287]. The use of oil is adventitious as all components are directly combined into a single, supersaturated protein solution, which is then left undisturbed as oil prevents evaporation of the sample[287], therefore the equilibration process is slowed down by the insertion of an oil barrier over the reservoir[288, 289]. A further advantage to setting up crystal trials under oil is that the oil protects the airborne sample from airborne contamination. In order for nucleation to occur one must work with very clean solutions; when working under oil, the samples are never exposed to air therefore heterogeneous nucleation is controlled[290, 291]. In addition it is generally believed that external disturbances such as vibration can cause excess nucleation and lead to the formation of smaller crystals or to crystal imperfections[289]. Trials under oil are preserved from physical shock since the nuclei and the forming crystals are buoyed and cushioned by the viscous oil making trials less susceptible to vibration[290].

1536 cocktails were tested under oil (at the Hautpman-Woodward medical research institute, University of Buffalo). The stability of the protein appeared to change with a higher percentage of drops displaying protein precipitation. However no drops displaying protein crystals were observed. One of the many possible reasons that p300 was recalcitrant to forming crystals is the high flexibility of the protein (which will be discussed further in the SAXS section (3.4.4.1).

The complex HIF-1α/p300 may be less flexible than p300, therefore crystallisation trials were also conducted for the HIF-1α/p300 complex. HIF-1α was added to p300 at a molar ratio of 1.5:1, to ensure all p300 was bound, and the mixture was incubated for 30 minutes before co-purification by size exclusion chromatography. This co-purification should remove all unbound HIF-1α from the sample, leaving a homogeneous HIF-1α/p300 complex. A decrease in elution volume (from p300) signified complex formation. Again commercially available sparse matrix screens were run to find initial crystal hits. In much the same way as with p300 very little protein precipitation occurred and no protein crystals grew at the initial concentration

of 10 mg/mL. Again numerous concentrations, buffer conditions and temperatures were tested with little success. For crystallisation to occur the protein must be a homogenous sample, one of the limitations of co-purification is that the resolution of the gel filtration column allows for the potential of unbound protein to be present. To limit this possibility a HIF-1α/p300 fusion construct was produced.

A protein which consisted of p300 at the N terminus and HIF-1α at the C terminus, joined with a linker that includes a TEV protease site was produced by PCR. Multiple rounds of PCR were required, firstly to attach the N terminus of the linker region to the C terminus of p300 and the C terminus of the linker region to the N terminus of HIF-1α, and then to link these two fragments together (Figure 56). This fused p300/HIF-1α protein was then ligated into pGex-6p-2 plasmid and BL-21 PLysD DE3 Gold expression strain was transformed with the plasmid. The HIF-1α/p300 fusion construct expressed well under the same conditions as p300.

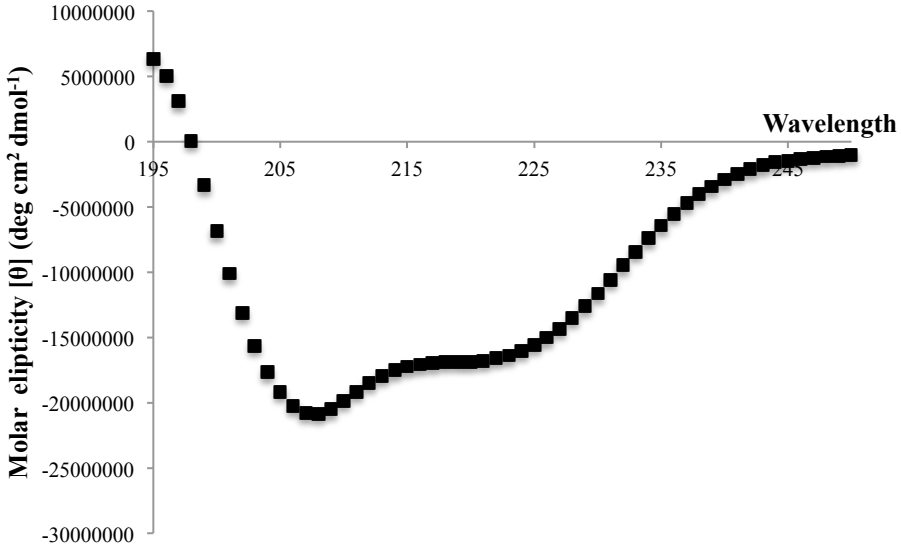


GPLGSADPEKRKLIQQQLVLLLHAHKCQRREQANGEVRQCNLPHCRTMKNVLNHM
THCQSGKSCQVAHCASSRQIISHWKNCTRHDCPVCLPLKNAGENLYFQGGSDLACR
LLGQSMDESGLPQLTSYDCEVNAPIQGSRNLLQGEELLRALDQVN

Figure 56 – Fused p300/HIF-1α. HIF-1α (blue), p300 (green) and TEV linker region (red). The amino acid sequence is shown, colour coded as above (ribbons Pymol).

As the linker region may have a detrimental effect on the fold and stability of the protein CD was used to check the integrity of the protein (Figure 57). The spectrum indicated that the HIF-1α/p300 fusion construct had an α-helical secondary structure[134]. The thermal melt did not give a sigmoidal curve, this may indicate that there is more than one species present in the sample or non-cooperative infolding. To investigate this further and to assess the flexibility and homogeneity of the HIF-1α/p300 fusion construct, as well as p300 and co-purified HIF-1α/p300, small angle X-ray scattering was used. The SAXS experiment should indicate which construct makes the most promising crystallisation target.

A



B

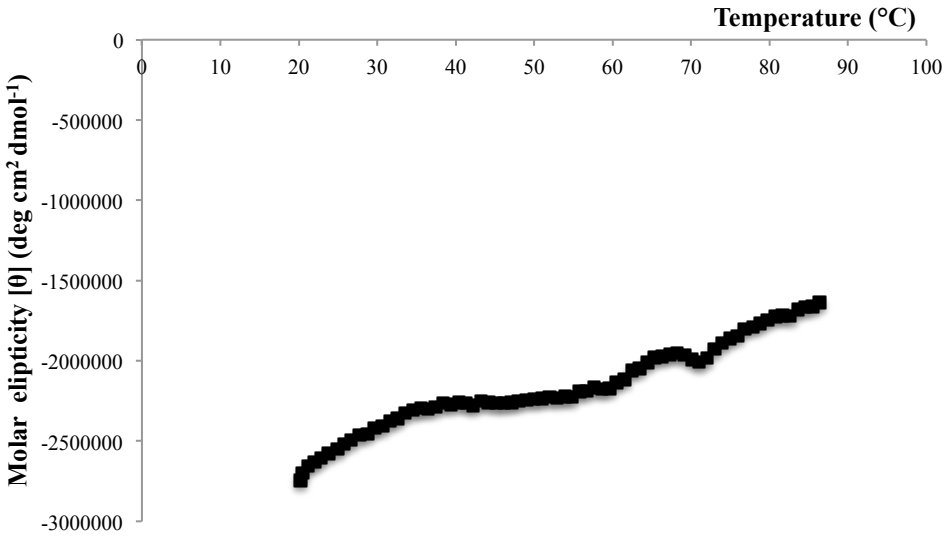


Figure 57 – CD analysis of the HIF-1α/p300 fusion construct.

3.4.4.1 SAXS analysis

3.4.4.1.1 Sample Quality

Before studying the conformations of p300, co-purified HIF-1α/p300 and HIF-1α/p300 fusion construct in solution the quality of the sample was assessed by collecting SAXS data for each protein over a concentration series (Table 25 and Figure 58). A persistent increase or decrease in the radius of gyration (Rg) or maximum particle dimension (Dmax) as the concentration of the sample increased would indicate protein aggregation or repulsion respectively. As Rg or Dmax remain consistent, within error, throughout the concentration range tested, the protein samples were deemed to be monodisperse and monomeric in solution. Data was analysed using Primus.

Table 25 – SAXS derived size parameters for p300, co-purified HIF-1α/p300 and HIF-1α/p300 fusion construct determined by Primus

Protein	Concentration (μM)	Rg	Dmax
P300			
	108	15.8 ± 1.7	51.8
	167	16.2 ± 0.9	55.6
	217	16.6 ± 1.4	55.6
	271	16.9 ± 1.0	55.5
	334	17.7 ± 1.3	56.7
	381	16.9 ± 2.4	56.6
	434	16.9 ± 0.8	58.0
HIF-1α/p300 fusion construct			
	115	17.8 ± 0.6	58.2
	177	17.6 ± 1.4	60.4
	230	17.6 ± 1.0	61.5
	288	18.9 ± 1.0	61.9
	354	17.8 ± 1.1	62.3
	404	17.6 ± 1.6	59.1
	460	17.9 ± 1.4	61.3
Co-purified HIF-1α/p300			
	95	16.8 ± 2.0	56.2
	146	17.2 ± 0.6	56.5
	190	17.2 ± 1.2	54.9
	238	16.6 ± 1.2	54.4
	292	16.8 ± 1.5	55.0
	333	16.7 ± 1.3	55.9
	380	16.7 ± 1.0	54.7

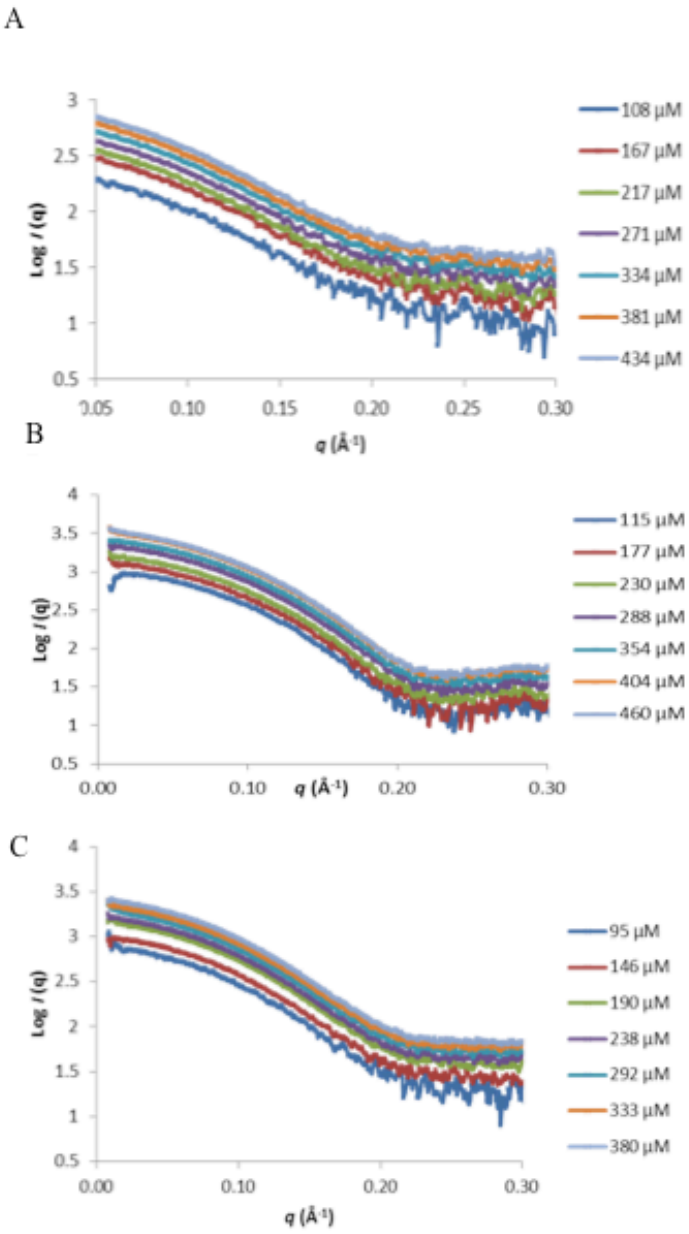


Figure 58 - SAXS intensity profiles (logarithmic) for 7 concentrations of each protein sample. A) p300. B) HIF-1α/p300 fusion construct. C) co-purified p300/HIF-1α.

The lack of aggregation and repulsion of each protein can be further confirmed by consistency in their Guinier plots across the concentration range. A positive inflection within the Guinier region (qXRg<1.3) indicates aggregation of the sample whereas a negative infection within the Guinier region indicated repulsion. As the Guinier region remains linear at the highest concentration of each protein; p300

(**Figure 59**A), HIF-1α/p300 fusion construct (**Figure 59**B) and co-purified HIF-1α/p300 (**Figure 59**C) the samples can be assumed to be resistant to aggregation and repulsion at the highest concentration tested.

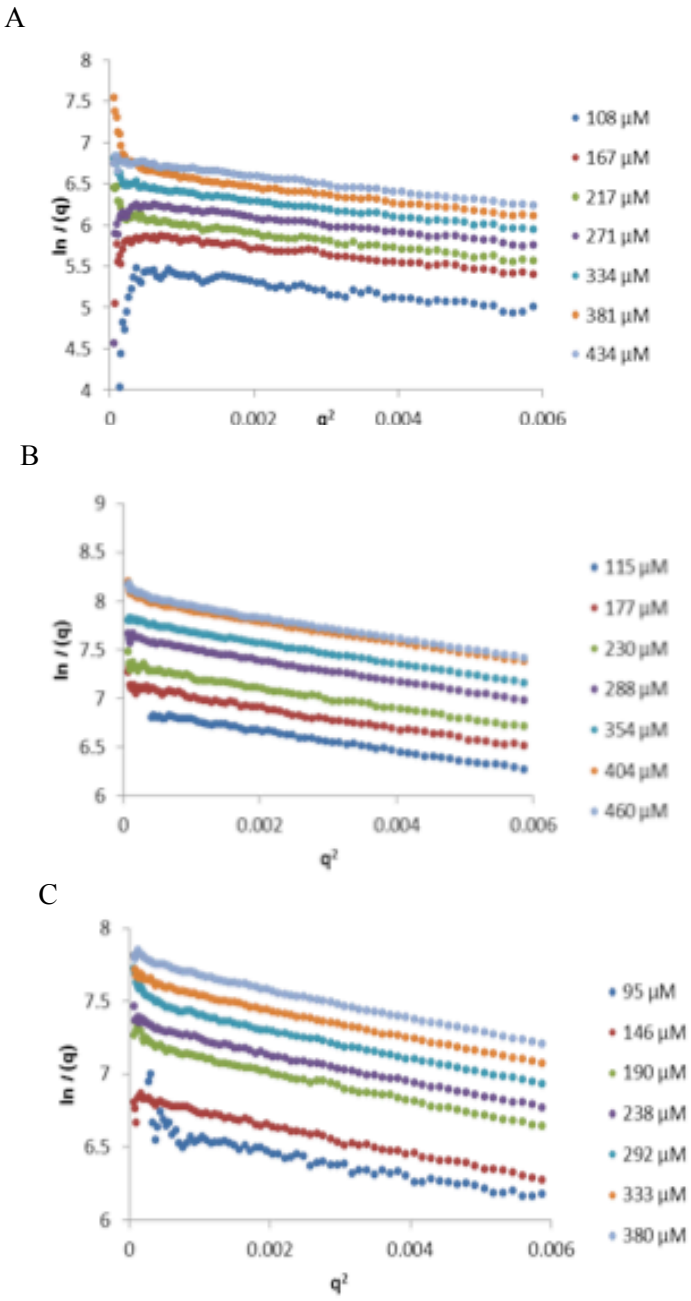


Figure 59 - Linearity of Guinier plots with manual selection of Guinier region. **a)** Guinier plot of p300. **b)** Guinier plot of HIF-1α/p300 fusion construct. **c)** Guinier plot of co-purified p300/HIF-1α.

The pair wise distribution function (P(R)) (**Figure 60**), which reflects inter-atomic difference distributions[292], was consistent across the concentration range for each protein sample. This, taken together with the gel filtration elution profile, the Guinier plots and Rg and Dmax values indicates a monodisperse and monomeric solution of each protein sample; p300, HIF-1α/p300 fusion construct and co-purified HIF-1α /p300.

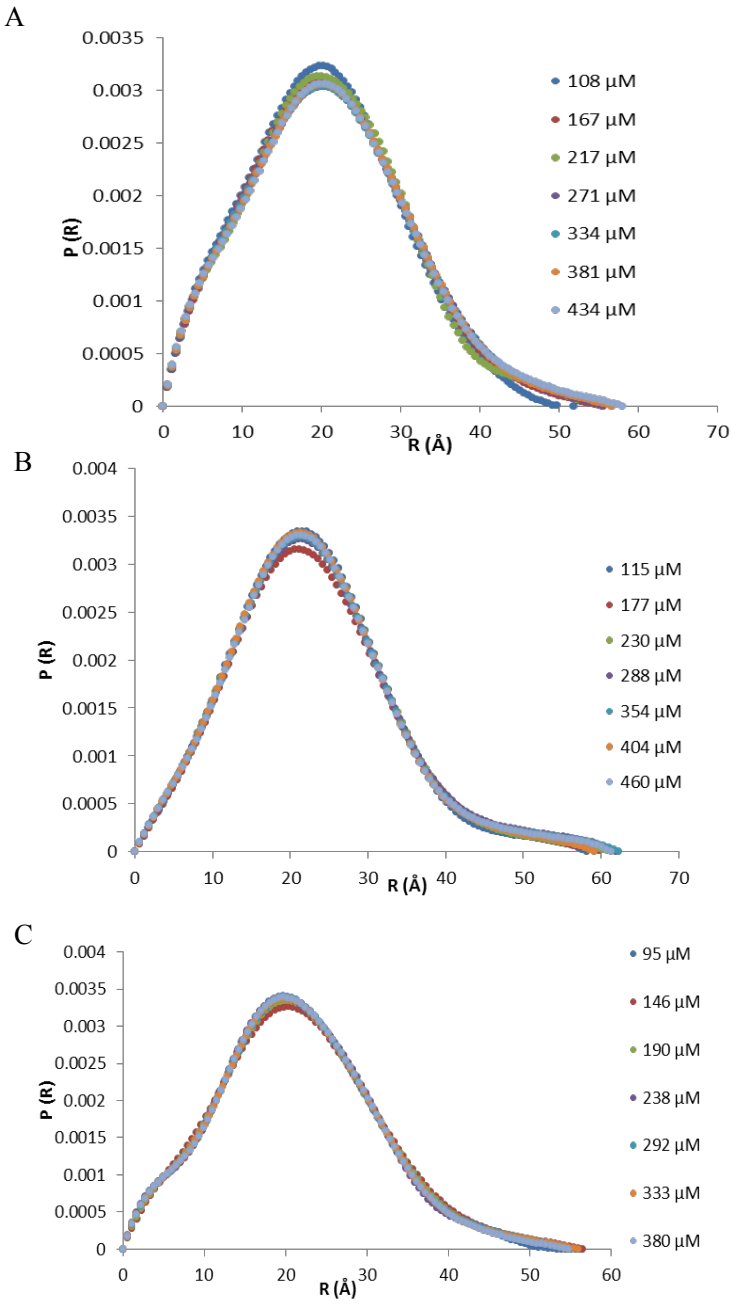


Figure 60 - Normalised pair distribution function P(R) calculated automatically by GNOM . A**)** p300.B**)** HIF-1α/p300 fusion construct C**)** co-purified p300/HIF-1α[292].

3.4.4.1.2 Size and shape comparison of p300, co-purified p300/HIF-1α and HIF-1α/p300 fusion construct

As there was no suggestion of aggregation or repulsion in the samples and no conformational change over concentration range was observed, indicated by consistency in Rg and Dmax (Table 25); further analysis was conducted using the data from the most concentrated sample of each protein due to the greater signal to noise ratio.

The radius of gyration is largest for the HIF-1α/p300 fusion construct indicating that the HIF-1α/p300 fusion construct is the least compact of the three samples whereas p300 has the smallest radius of gyration (Table 25) indicative of a more compact structure. The Dmax supports this, with the HIF-1α/p300 fusion construct having the largest Dmax and p300 having the smallest. This difference in size and shape is illustrated by the slight differences in the shape of the pair distribution curves (**Figure 61**A). These results align with what would be predicted for the size and shape of the samples from the known NMR structure (1L8C)[182], molecular weights and gel filtration elution profile.

Kratky plots describe the conformational heterogeneity of a protein, a relatively rigid protein will have a folded profile of a bell shape curve which returns to 0, whereas an unfolded protein with extreme conformational heterogeneity will not have a peak but will plateau at high q values, and flexible proteins with partial conformational heterogeneity will lie somewhere in between. Each of the Kratky plots show the characteristic globular peak of a folded protein for each sample (**Figure 61**B). However, differences between the p300, HIF-1α/p300 fusion construct and co-purified HIF-1α/p300 peaks illustrate a difference in conformational heterogeneity between all three of the samples. This difference was expected between p300 and the complexes (fusion and co-purified). The difference in Kratky plots between the HIF-1α/p300 fusion construct and co-purified p300/HIF-1α revealed that under the conditions tested there is a difference in conformational heterogeneity of the complex dependent on whether the complex is co-purified or fused. This difference in conformational heterogeneity is supported by the difference in Rg and Dmax. To investigate this difference further the SAXS data for each protein can be compared to theoretical SAXS data generated from NMR of HIF-1α/p300 complex (1L8C)[182].

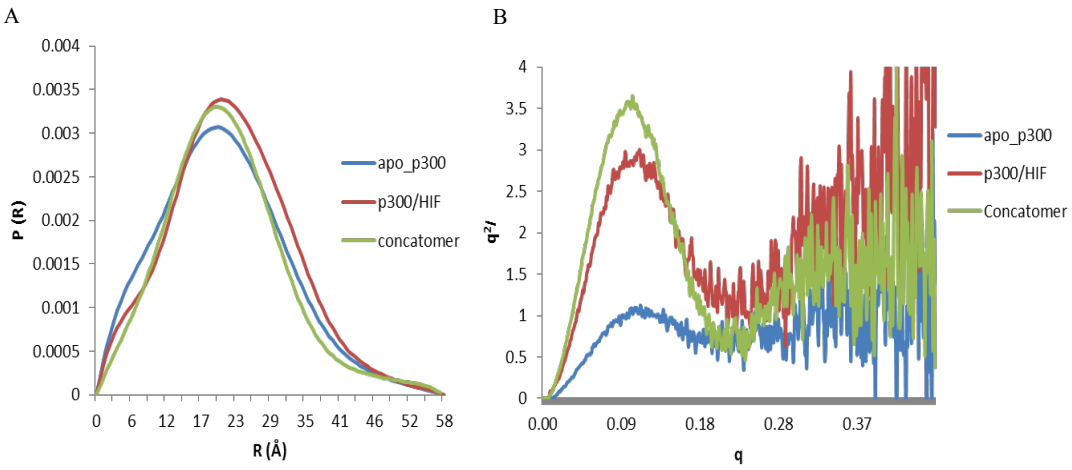


Figure 61 - Comparison of p300, co-purified p300/HIF-1α and HIF-1α/p300 fusion construct (concatomer). A) Comparison of the normalised pair distribution functions P(R) calculated automatically with AutoGNOM[292]. B) Kratky plots support a globular shape.

3.4.4.1.3 Comparison with NMR structure

Theoretical SAXS data can be generated from the known NMR structure (1L8C)[182] of the HIF-α/p300 complex (the model)[293]. The p300 experimental SAXS data did not match well to this model (**Figure 62**a). However when a model was generated containing only p300 theoretical SAXS data from the NMR structure of the HIF-1α/p300 complex (HIF-1α removed) the fit to p300 was improved **(Figure 62**b). The fit to this model would not be expected to match exactly as it is likely that p300 would exhibit some conformational changes upon binding, which can not be accounted for in this model.

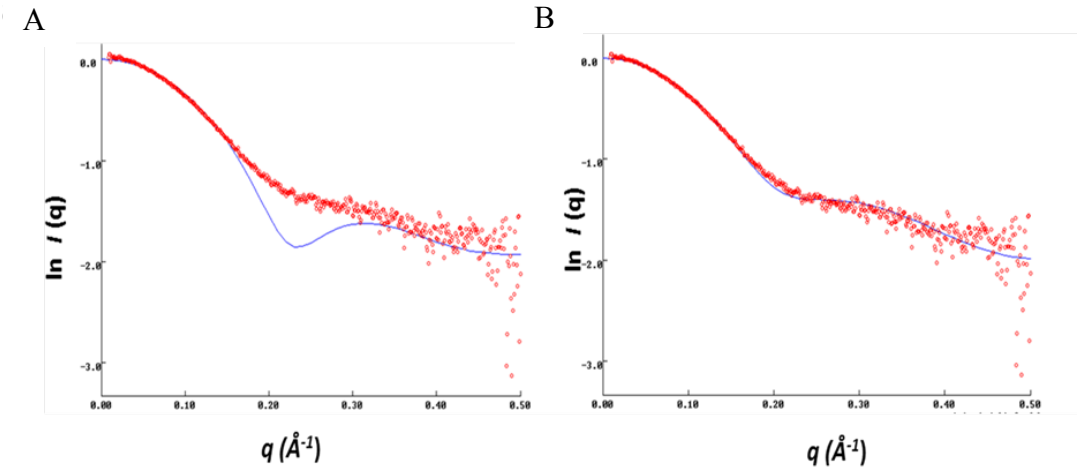


Figure 62 - Fit of the theoretical scattering profile for the rigid body model (blue line) with the experimental p300 SAXS data[293]. A**)** model generated from p300/HIF-1α complex NMR[182]. B**)** model generated from p300 NMR data (HIF-1α data removed from complex model)

The co-purified p300/HIF-1α data did not fit to the HIF-1α/p300 complex model (**Figure 63**a) or the p300 model (**Figure 63**b). This indicates that under the conditions tested the p300 and HIF-1α does not appear to remain completely as a complex. As the data does not fit well to either model it is suspected that the sample was a mixture of apo proteins and complex; further analysis is required to assess how much of the solution is of complex and how much is of apo protein.

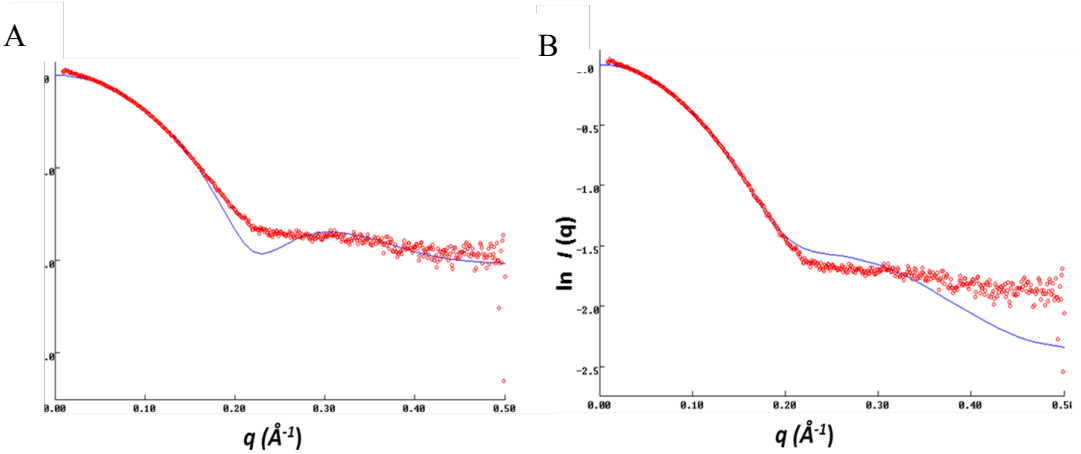


Figure 63 - Fit of the theoretical scattering profile for the rigid body model (blue line) with the experimental co-purified SAXS data[293]. A**)** model generated from HIF-α/p300 complex NMR[182]. B**)** model generated from p300 NMR data (HIF-1α data removed from complex model)

The HIF-1/p300 fusion construct α fit well to the HIF-1α/p300 model and poorly to the apo_300 model (**Figure 64**A/B). This indicated the HIF-1/p300 fusion construct is

folded and forms a complex in a similar way to the p300 and HIF-1α complex in the solution NMR structure[182]. This can be seen visually by generating an envelope from the HIF-1/p300 fusion construct and comparing it to the NMR structure of HIF-1α/p300 (**Figure 64**C). The envelope was generated using the P(R) function in DAMMIF[294], 10 individual *ab initio* molecular envelopes were reconstructed and averaged.

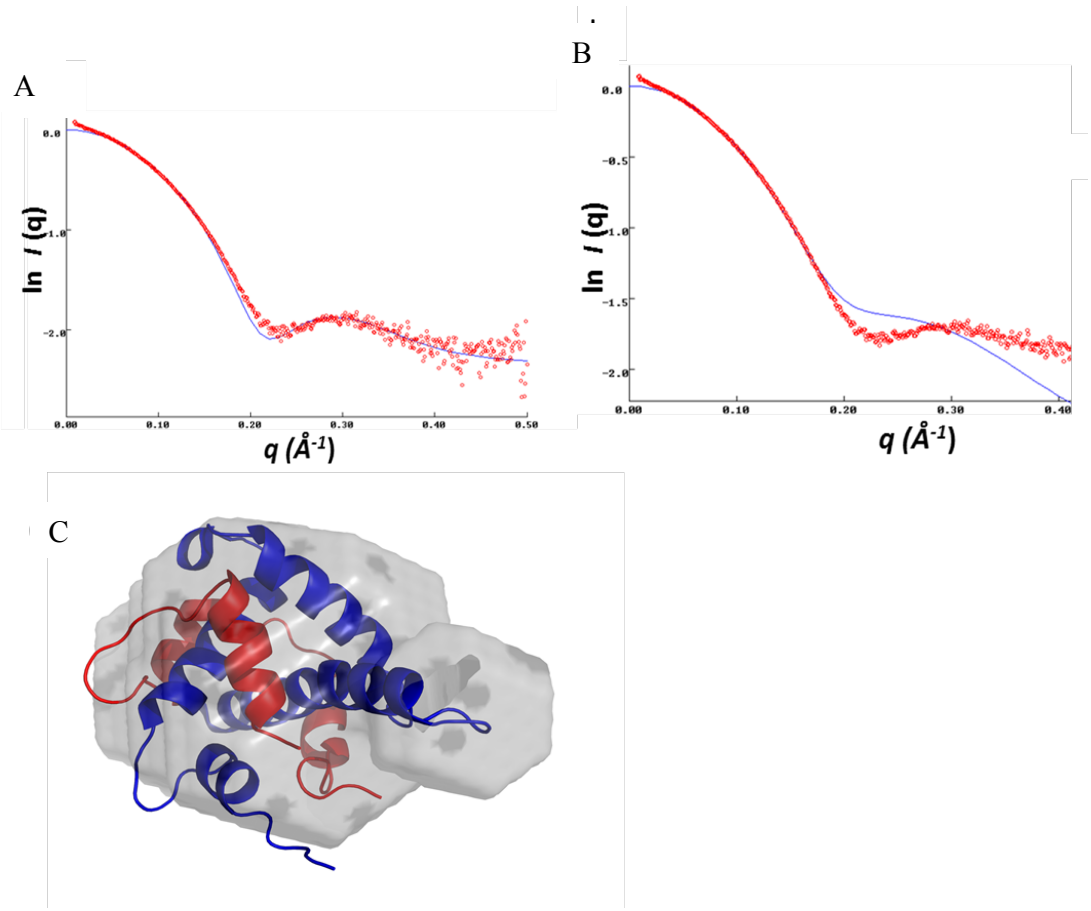


Figure 64 - Fit of the theoretical scattering profile for the rigid body model (blue line) with the experimental co-purified SAXS data[293]. A) model generated from HIF-1α/p300 complex NMR[182]. B**)** model generated from p300 NMR data (HIF-1α data removed from complex model). C) Average molecular envelope for the HIF-1α/p300 fusion construct [294] overlaid the NMR structure of p300/HIF-1α[182].

This SAXS experiment indicates that the HIF-1α/p300 fusion construct protein is a valid crystallisation target. The HIF-1α/p300 fusion construct has been shown to be resistant to oligomerisation (under the conditions tested), as have p300 and co-purified HIF-1α/p300, however when fused the p300/HIF-1α interaction has a

greater homogeneity than when co-purified. As the complex will give additional information regarding binding and will therefore inform drug design; the fused form may be deemed a more attractive crystallisation target than both the apo p300 and co-purified complex.

3.4.4.2 HIF-1α/p300 fusion construct crystallisation

Crystallisation trials of HIF-1α/p300 fusion construct were conducted, again using commercially available sparse matrix screens. Unfortunately no proteins crystals grew in these factorial screens. The factorials were also performed under oil with no success. To try and change the stability of the protein, trypsin was added to cut any flexible regions that may be limiting crystallisation, this technique has been used to aid crystallisation with some success[295, 296]. Initially an experiment was performed to assess the behaviour of the HIF-1α/p300 fusion construct in the presence of trypsin. A 10 μL stock of trypsin was prepared in the buffer 10 mM Tris pH 7.5, 1 mM $CaCl_2$ and was added to 10 mg/mL of HIF-1α/p300 fusion construct (10 mM Tris pH 7.5, 100 mM Nacl, 1 mM DTT and glycerol) to a final concentration of 1 nM, 250 nM and 500 nM. Each reaction was incubated at room temperature with samples at time points: 5 minutes, 10 minutes, 30 minutes one hour, two hours and four and a half hours.

SAS PAGE analysis showed that the HIF-1α/p300 fusion construct was cut to a stable fragment (in terms of no further significant degradation) over the time course tested. This product remained the minority protein after the addition of 1 nM trypsin, however for the 250 nM and 500 nM reaction this was the majority protein after 1 hour (Figure 65). The product was subject to mass spectrometry analysis and a mass of 12.48 kDa was measured for the product. Trypsin cuts at lysines and arginines preferentially, when looking at the structure there is an arginine present near the N terminus of HIF-1α; p300, the linker region and the remaining N terminus of HIF-α leaves a molecular weight of 12482 Da, therefore this is a likely product (Figure 65B and C).

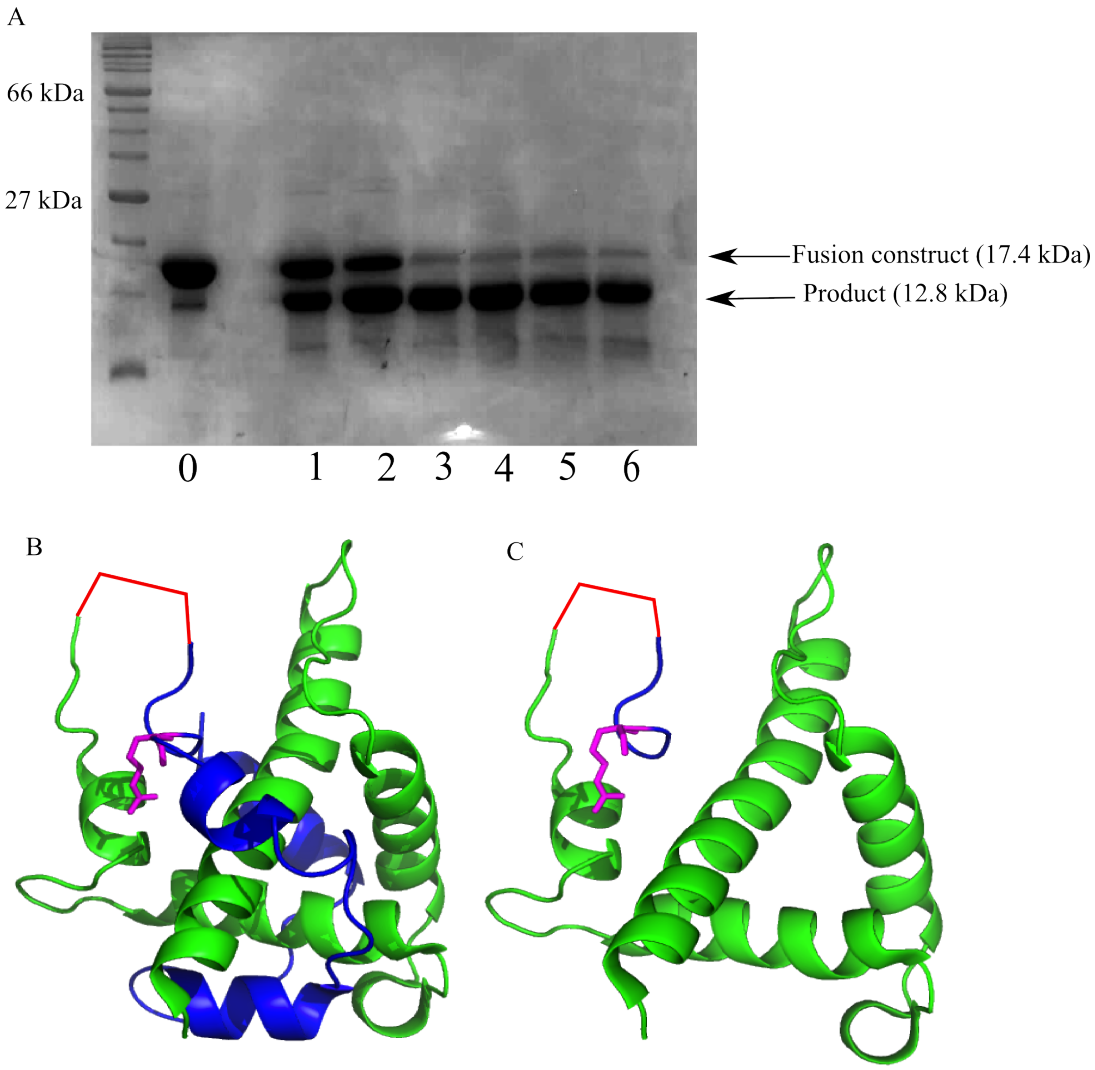


Figure 65 – Trypsin digest analysis. A)SDS PAGE analysis of HIF-1α/p300 fusion construct cut with 250 nM trypsin over a time course. 0=before addition of trypsin, 1= 5 minutes, 2=10 minutes, 3=30 minutes, 4=1 hour, 5 =2 hours and 6=4.5 hours. B) Full fused HIF-1α/p300 with R_{785} highlighted (magenta), where trypsin potentially cuts (pymol, cartoon). C) Potential product of the after fused HIF-1α/p300 was treated with trypsin. (pymol cartoon).

Crystal trials were set up using Morpheus screen (one of the sparse matrix screens). The HIF-1α/p300 fusion construct was incubated with 2 nM and 250 nM trypsin for 45 minutes (as well as a sample which was not incubated with trypsin) before the plates were set up. The plate was more promising, with a slightly larger percentage of drops displaying protein precipitation than when no trypsin was present. Further work could consist of trying different trypsin concentrations and incubations times as well as trying the full complement of commercially available screens. It may also be worth testing the use of trypsin in a different way, trypsin can be added to the mother liquor as opposed to protein sample.

Unfortunately despite a great number of factorial screens conducted from many protein concentrations, buffer conditions and temperatures none of the three constructs (p300, co-purified HIF-1α/p300 and the HIF-1α/p300 fusion construct) yielded protein crystals. In total for the three constructs over 30,000 crystallisation conditions were tested. Further work can be conducted to trial further conditions and constructs, particularly the HIF-1α/p300 fusion construct which was highlighted as the better crystal target by SAXS. There are almost boundless possibilities for crystallisation conditions with very little predictability as to which condition a protein will crystallise in.

3.5 HIF-1α/p300 chapter summary and further work

As has been previously discussed HIF-1α/p300 is a protein-protein interaction which is up-regulated in many solid tumours. The challenge of disruption of the HIF-1α/p300 interaction come as HIF-1α wraps around p300, with multiple potential sites for small molecule modulation, however which of these sites to target is unclear (Figure 17). Work conducted on this target involved biophysical and structural characterisation for the elucidation of the binding interface to try and locate binding hotspots, which could subsequently be used in the rational drug design progress. This was done by two approaches; firstly by analysis of the binding of shorter HIF-1α peptide fragments, this piece of work highlighted HIF-1α helix 3 as a key binding region. The second approach used phage display for unbiased exploration of the p300 surface. Two phage display technologies were used; a peptide phage library and an Adhiron (Affimer) phage display library.

The peptide phage display library experiment yielded a peptide which bound with a higher affinity than any of the short native fragments, NMR and mutations in p300 indicated that this peptide was binding in or around the helix 3 binding pocket. Oligoamide α-helix mimetics were designed against the helix 3 binding pocket, which disrupted the interaction with an $IC_{50} = 9$ μM. The Adhiron phage display experiment generated Adhirons around 100 nM, which disrupt the interaction $IC_{50} =$ 1-5 μM. The crystal structure of the Adhiron docked to the NMR structure of p300 suggested that the interaction was mediated through the two variable loops. Therefore potentially two sites are required for potent inhibition of the interaction.

This study demonstrated that HIF-1α/p300 is a high affinity interaction, which is difficult to inhibit; no short peptide region was able to disrupt the interaction. However, despite such challenges, inhibitors that do disrupt the interaction were designed. These compounds have been designed by mimicking the helix 3 region of HIF-1α using the Wilson group oligoamide scaffold[96, 98]. This work could be furthered by testing of these compounds in a cell-based assay, to give information about cytotoxicity and cellular potency. Other oligoamides, which have been

designed against different targets[7, 96, 98, 246, 297-299], have been shown to be active in cells[299].

Structure affinity relationship studies have been conducted for the oligoamide compounds[244], however further improvements to the compounds could be made to improve their properties. Attempts have been make to improve the solubility of scaffold[246, 297, 300], as this is a limiting factor for some biophysical (ITC) and structural (NMR and crystallography) techniques. An N-alkylated oligoamide as opposed to the O-alkylated scaffold Figure 66 used to disrupt the HIF-1α/p300 have been synthesized[117, 246], as have compounds which replace the central monomer which a natural amino acid[300] Figure 66. Both of these adaptations have reported to improve the solubility of the scaffold. Such adaptations have shown to be tolerated on other PPI targets[246], but suitably functionalized scaffolds did not inhibit the HIF-1α/p300 interaction.

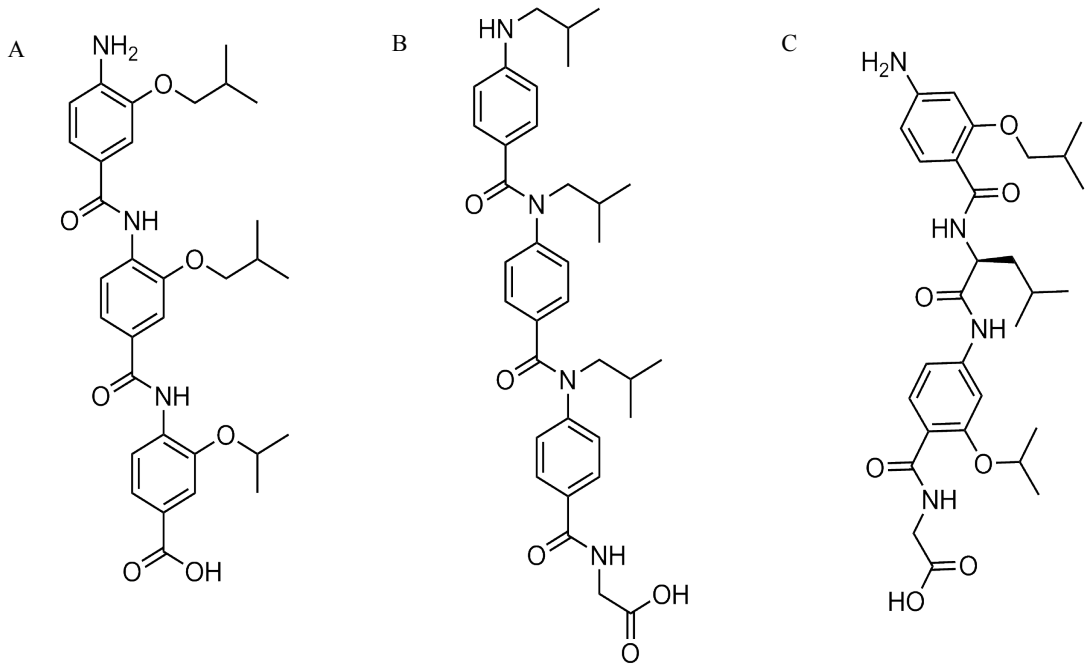


Figure 66 – Structures of helix mimetic scaffolds. A) O-linked oligoamide. B) N-linked oligoamide. C) Hybrid scaffold replacing the central monomer with a natural amino acid.

The phage display derived peptide binds to p300 with a higher affinity than any fragment of the native HIF-1α peptide of the same length, further studies to assess the binding confirmation of this peptide will allow for this peptide to be mimicked. As the phage peptide binds with a higher affinity than helix 3 it would be hoped that a mimetic would also be more potent. As the peptide may not be α-helical other small molecules may be able to mimic the peptide using a different scaffold. The ideal way to assess the binding conformation would be X-ray crystallography. However with a 20 μM binding affinity co-crystallisation might be difficult, and as p300 has proven difficult to crystallise the task may be more challenging. The NMR structure of the complex could be solved using multidimensional NMR studies, this would allow the complex to be visualised not just the binding region highlighted as in this study.

The Adhiron phage display experiment generated nanomolar affinity binders of p300, a higher affinity than both the fragments of native HIF-1α peptide and the phage display derived peptide. Unlike the fragments and the phage display derived peptide the Adhirons were able to disrupt the HIF-1α/p300 interaction. There are many possible reasons as to why the Adhirons perform better, two of which are; the constraints on the peptide loops decrease the entropic cost of binding, and more than one binding site is required to disrupt the interaction, which is achieved by the two loops. If the binding is mediated through the two variable loops; a larger interaction area is being reached. As was previously discussed further validation of the binding mode is required to verify the binding mode of the Adhirons. However, if the Adhiron is binding in the predicted way there may be a potential to exploit fragment based drug design to target the HIF-1α/p300 interaction.

The FBDD approach could be used in different ways to exploit the two sites identified by the two loops of the adhiron. If one fragment is identified which binds in one of the sites then this fragment could be optimized by fragment growth, with the fragment growing out to reach the second site. For this to be successful the sites need to be in relative close proximity and the initial fragment must act as an "anchor" and not change its binding mode during its growth[301]. Another method of exploiting the two sites is by fragment linking, which is possible if two fragments, one binding at each site, are identified. Such fragments could be linked directly

(although the two sites are likely to be too distant) or by a suitable linker. This may be challenging, as the linker must allow both fragments to maintain their original binding modes while combined in the new hybrid ligand[301].

Having structural data benefits the FBDD process greatly[302]. A soakable crystal system would not only allow for the discovery of binders but also give information regarding the binding mode of the fragment. Allowing for informed decisions regarding optimisation strategy. Therefore the lack of a crystallisable p300 construct may be a limiting factor. However the recent success of fragment based approaches and the indication that inhibition of HIF-1α/p300 is more likely through the targeting of multiple sites FBDD is an exciting prospect for inhibitor development of the HIF-1α/p300 interaction.

The Adhirons have been reported to be monomeric and stable[122], however the Adhirons which bind to p300 have demonstrated a tendency towards aggregation. The gel filtration trace displayed oligomerisation peaks and the crystals structure indicated dimerization. Analysis of the dimer revealed hydrogen bonding from the loop region to the β sheet of the second adhiron (Figure 53). This tendency for dimerisation (and even the formation of higher order oligomers) may mean that the concentration of active protein is lower than that measured, therefore the Adhiron is more active than reported. Polar and charged residues could be introduced to the scaffold, at the base of the loop regions, to obstruct the dimerisation.

As can be seen there is much more work that can be completed to further the project. However the characterisation of the interaction thus far has allowed for the development of small molecules, which disrupt the interaction. Therefore the structural and biophysical characterisation of the target has directly informed rational drug design

Chapter 4 : eIF4E/eIF4G Chapter

4.1 Introduction

As discussed briefly in the introduction, there was a second PPI target in this project eIF4E/eIF4G, which is also a potential cancer therapeutic target. This chapter will discuss the relevance and importance of this target and the biophysical and structural characterisation of the interaction with a view to informing inhibitor design.

4.1.1 Eukaryotic translation

The control of mRNA translation is important for cell growth, proliferation and differentiation[152], and is one way in which protein levels may be selectively regulated[167]. Translation requires an initiation step where the Cap (m^7GpppN) at the 5' end of an mRNA binds to a ribosome. This initiation is mediated by many eukaryotic initiation factors (eIFs)[167]; a key initiation factor is eIF4E. eIF4E is the Cap binding protein involved in directing the ribosome to Capped mRNA[303]. The Cap is located at the 5' end of all cellular eukaryotic mRNAs. eIF4E binds the Cap when in complex with eIF4A (the RNA helicase) and eIF4G (the multidomain adaptor), this complex is known as eIF4F (

Figure 67A). eIF4E is described as the master switch that controls eukaryotic translation, as it is rate limiting for eIF4F complex formation. The interaction of eIF4E with eIF4G to form eIF4F is tightly and negatively regulated by eIF4E-binding proteins (4E-BPs), a family of inhibitory proteins that sequester eIF4E by occupying the same binding site as eIF4G[304, 305], therefore the availability of eIF4E to bind to eIF4G relies on release from 4E-BP[169] (

Figure 67B).

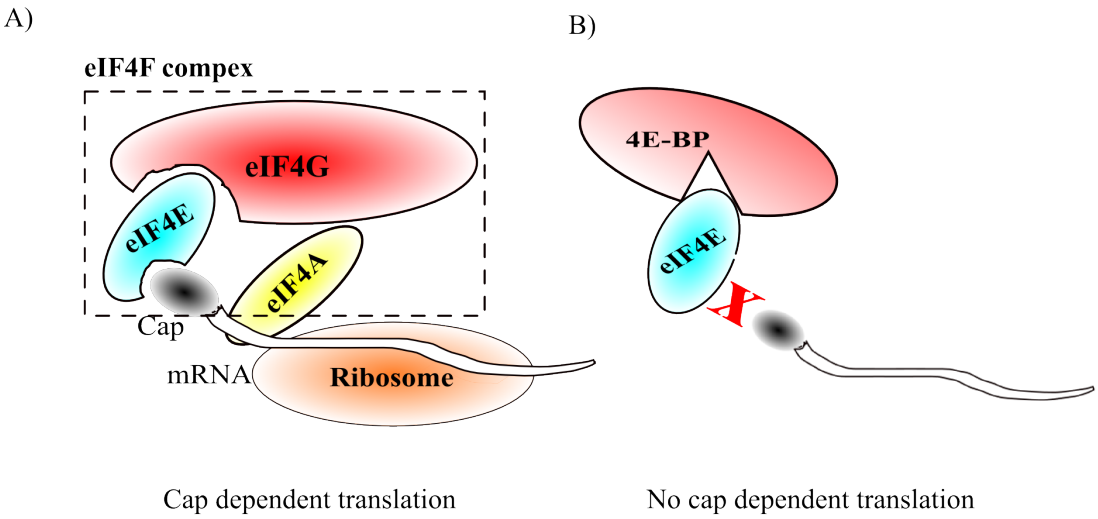


Figure 67 - Control of cap dependant translation. A) Formation of the eIF4F complex of eIF4E, eIF4A and eIF4G, allows the binding of the cap of mRNA which subsequently recruits the ribosome. B) 4E-BP bindings at the same site as eIF4G, blocking eIF4G binding which stops the formation of the eIF4F complex.

The eIF4E/4E-BP binding event is controlled by phosphorylation, hypophosphorylated 4E-BPs bind strongly to eIF4E (therefore less eIF4E is available for binding to eIF4G), resulting in little or no formation of the eIF4F complex, whereas the hyperphosphorylated forms bind weakly[152, 306], allowing eIF4G to compete for the binding of eIF4E and realise eIF4F assembly. Phosphorylation is mediated by many extracellular stimuli including growth factors, hormones, mitogens, amino acids, cytokines and G-protein-coupled receptor agonists[152, 306].

To become translated, mRNAs must interact with eIF4E within fully assembled eIF4F complex; those mRNAs with a short, unstructured, 5'UTR interact more readily than mRNA with highly structured 5'UTRs which are therefore translated less efficiently[152]. Consequently overexpression of eIF4E, which occurs in approximately 30 % of all cancers, results in increased translation of mRNAs containing highly structured 5'UTRs, which include mRNAs encoding cancer-related proteins[304] that control cell proliferation and viability[307] and growth-promoting gene products such as cyclin D1, c-myc and VEGF[152, 308-310]. It is this increased translation of growth-promoting gene products and cancer related proteins that may be a trigger for cancer development. *eIF4E* is a recognized oncogene as it is up-regulated in up to 30 % of human tumors, including breast, prostate and lung cancers as well as in many

leukemias and lymphomas[303]. The connection between overexpression of eIF4E and tumour development has been validated in cell culture and in animal models as the up-regulation led to oncogenic transformation (in cells) and tumor formation (in the animal models)[303].

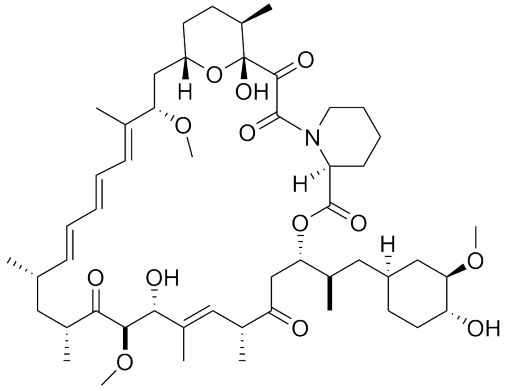
4.1.2 Inhibition of cap dependent eukaryotic translation

There are multiple points on the Cap dependent translation pathway that are potential targets for modulation to decrease the up-regulated translation of cancer-related proteins and growth promoting gene products. Areas that offer the opportunity for modulation include; the release of eIF4E from 4E-BP which increases the availability of eIF4E to form the eIF4F complex, the formation of the eIF4F complex and the binding of the mRNA Cap to eIF4E.

4.1.2.1 Phosphorylation state of 4E-BP

The binding of eIF4E to 4E-BP is controlled via the phosphorylation state of 4E-BP. When hypophosphorylated (little or no phosphorylation) 4E-BP binds to eIF4E with a greater affinity than when hyperphosphorylated. This affect is due to the fold of 4E-BP. Recent work has shown that in the hyperphosphylated state 4E-BP forms a highly structured β sheet fold. There is a key binding motif YXXXXLΦ (where X is any amino acid and Φ is hydrophobic) of 4E-BP (also present in eIF4G) which is required to be α-helical for binding, therefore when hyperphosorylated and in the β sheet fold no binding occurs[169]. When hypophosphylated 4E-BP is intrinsically disordered, the YXXXXLΦ motif is able to become α-helical and therefore 4E-BP is able to bind to eIF4E[311]. As 4E-BP and eIF4G compete for the same binding site of eIF4E[312], the higher the affinity of the interaction between 4E-BP and eIF4E the less eIF4E is available to interact with eIF4G; therefore the formation of the eIF4F complex is limited and Cap dependent translation is decreased. Controlling the phosphorylation of 4E-BP is therefore a potential way of controlling the formation of the eIF4F complex. If the phosphorylation of 4E-BP were inhibited then more eIF4E would be bound to 4E-BP (and therefore less in the eIF4F complex). However the phosphorylation of 4E-BP is complex and specific inhibition of phosphorylation is extremely challenging.

The natural product Rapamycin[313] (Figure 68) and its analogues, have been shown to inhibit 4E-BP phosphorylation by blockade of the mechanistic target of Rapamycin (mTOR) pathway[314, 315]. mTOR is a serine/threonine kinase, part of the phosphatidylinositol 3-kinase-related kinase family, and is regulated through the PI3K and Akt/PKB pathway[316].



Rapamycin

Figure 68 – Structure of Rapamycin (chem draw image drawn from pubchem, pubchem ID: 5284616).

The pharmacokinetic properties of Rapamycin are poor due to its lipophilicity[317]; therefore three analogues (Figure 69) that differ by their formulation and bioavailability but not by the mechanism of action have been developed. Temsirolimus (Wyeth/Pfizer), which has been FDA approved, Deforolimus or Ridaforolimus (Merck/Ariad) and Everolimus (Novartis)[315]. This is one example where the phosphorylation of 4E-BP has been inhibited resulting in an anti-cancer agent. It is likely that there are many more opportunities for modulation in the mTOR and indeed other pathways to regulate the phosphorylation of 4E-BP, however the network of multiple pathways is complex therefore the likelihood of generating a specific ligand maybe problematic.

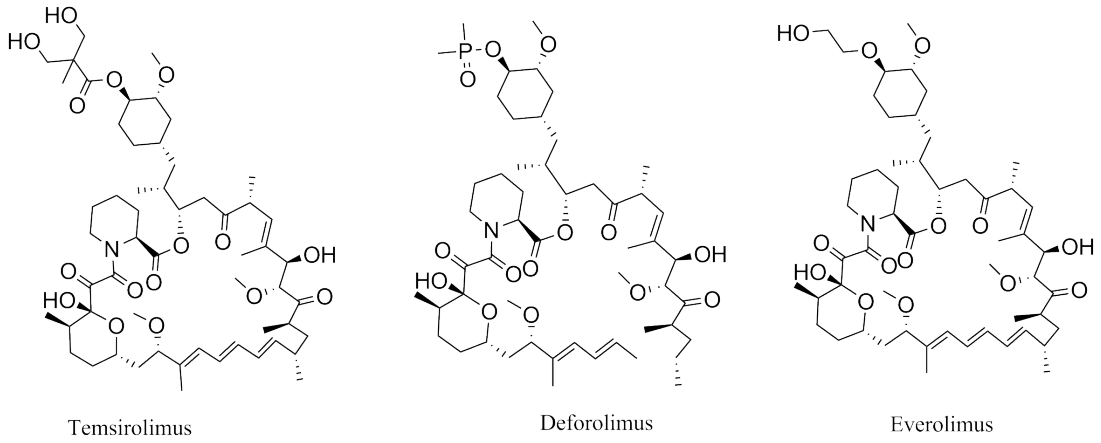


Figure 69 – Structures of Temsirolimus (pubchem ID:6918289), Deforolimus (pubcehm ID:11520894) and Everolimus (pubchem ID:6442177)

4.1.2.2. eIF4E/mRNA Cap interaction

Inhibition of the binding of the mRNA Cap to eIF4E is another potential point for inhibition of Cap dependent translation. For eIF4E to exert its role as an initiation factor it must bind to the methyl-7 guanosine (m7G) Cap structure on the 5' end of mRNAs[318]. A guanosine ribonucleoside analogue, 1-β-D-ribofuranosyl-1,2,4-triazole-3-carboxamide (Ribavirin), a drug which has been used for the treatment of infections including Lassa fever virus, respiratory syncytial virus, hepatitis C virus (HCV), and severe acute respiratory syndrome coronavirus[319], has been shown to also disrupt the eIF4E/Cap interaction[320]. Initially it was shown that Ribavirin was misincorporated into mRNA by viral RNA-dependent RNA polymerases[321, 322] due to its chemical similarity to guanosine, leading to lethal mutagenesis of polio and HCV genomes. More recently direct binding of Ribavirin to eIF4E (in the m7G mRNA Cap binding site) has been observed both *in vitro* and in cells; again most likely due to its similarity to guanosine[320]. Therefore Ribavirin inhibits eIF4E from promoting mRNA translation of eIF4E-sensitive transcripts, many of which are involved in cancer development[320].

Another inhibitor of the mRNA Cap/eIF4E interaction is a guanine-derived inhibitor (K_d (for eIF4E) = 95 nM, IC_{50} = 2.5 μM), discovered using structure guided drug design[323]. A crystal structure revealed that this compound bound by the same binding mode as Cap (Figure 70). This inhibitor was also shown to be selective, with no

notable off target effects up to 10 μM, however it does compromise cell permeability.

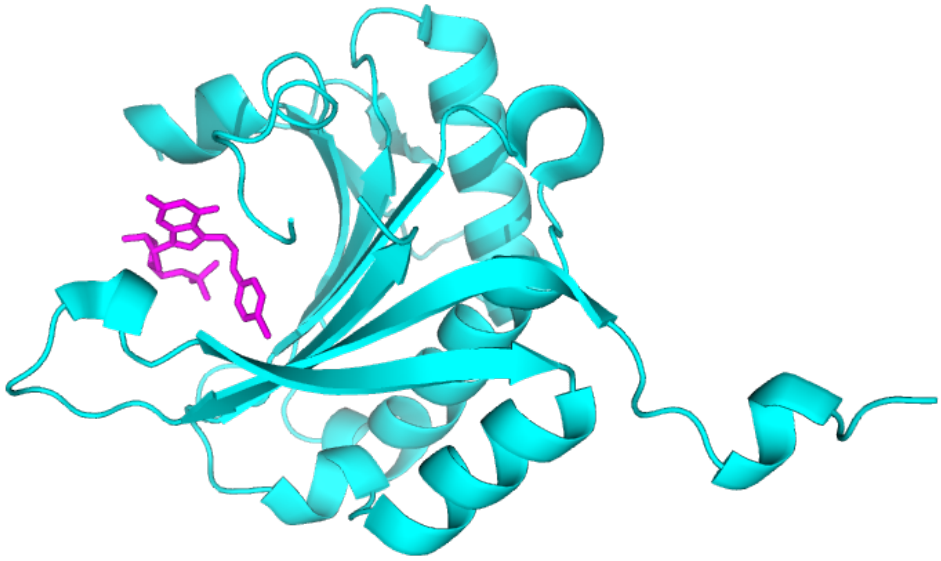


Figure 70 – Structure of guanine-derived inhibitor (magenta) of the mRNA Cap/eIF4E interaction bound in Cap binding sire of eIF4E (cyan).

4.1.2.3 eIF4F complex

The disruption of the eIF4F complex is another way to decrease cap dependent translation of cancer-related proteins and growth promoting gene products. Disruption of the eIF4E/eIF4G interaction would subsequently disrupt the eIF4F complex. There has been some success disrupting this interaction using small molecules. A high throughput approach was used to identify the small molecule 4EGI-1[324, 325] which disrupted the interaction with an IC_{50} = 16 ± 6 μM, this compound also inhibited cap-dependent translation *in vitro* in Renilla luciferase reporter assay system and similar affects were seen in mammalian cancer cell lines[326]. Structural studies have shown that this inhibitor acts through an allosteric mechanism. 4EGI-1 bindings to eIF4E at a site distinct from the eIF4G binding site, binding in a hydrophobic pocket which causes localized conformational changes, which causes an α-helical rearrangement, which is turn appears to stop the binding of eIF4G[327]. Another small molecule, 4E1R-Cat, was identified by library screening, this molecule inhibited eIF4E/eIF4G with an IC_{50} of 4 μM. Crucially 4E1R-Cat is

capable of reversing tumor chemoresistance in a genetically engineered Eμ−Myc lymphoma mouse model[328].

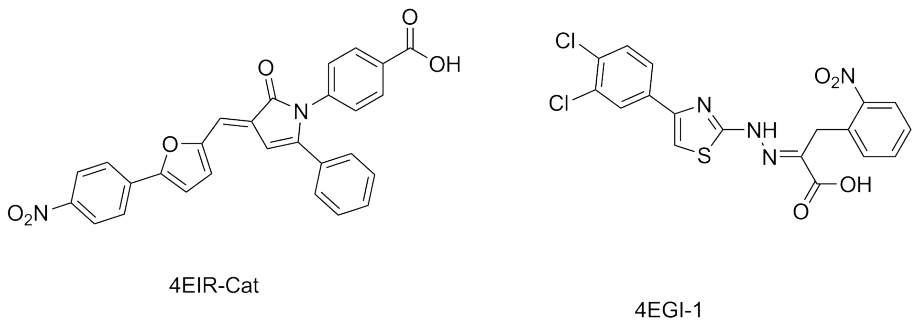


Figure 71 – Structures of 4EIR-Cat (MDL number:MFCD01931282) and 4EGI-1 (CAS number:315706-13-9)

Another approach used successfully to disrupt the eIF4E/eIF4G interaction is a stapled peptide approach. To staple the peptide, a hydrocarbon linker between turns of an α-helix was introduced; in theory this should stabilize the secondary structure[329]. This stabilisation should allow higher affinity binding due to the reduction of the entropic cost of binding. Proteolytic stability should also be improved, increasing the *in vivo* half life[330]. However, the introduction of a staple does decrease the conformational freedom of the peptide, therefore the location of the staple is important. A stapled variant of the eIF4G peptide has previously been produced[168]. The location of the staple as well as mutations in the native binding sequence were explored when designing the stapled version of the eIF4G peptide. Initially mutations (made outside the key binding motif YXXXXLΦ) were made to stabilize the helix[168], the mutations consisted of the inclusion of 1-aminocyclopentane carboxylic acid; this non-natural amino acid was chosen as it is thought to force the peptide in a conformation more similar to that observed in the peptide-protein complex and thus reduces the entropic cost of binding[168]. The second mutation, an α-carbon methylation of the peptide backbone was chosen as it was thought it would rigidify the peptide further. The location of each of the mutations was decided by sequential mutagenesis studies. The mutations were reported to cause

a 15 fold increase in affinity[168]. Phage display was then used to further improve the potency of the peptide, from this experiment the peptide ^{1}SLHYSRDQLVAL^{12} was selected. This peptide maintained the characteristic binding motif (Y_{624}, L_{629} and the hydrophobic position at 630) seen in the eIF4G peptide, although the rest of the peptide is different (^{1}KKRYDREFLLGF^{12}). This phage display derived peptide was insoluble and highly intractable therefore the N-terminal SLH sequence was changed to KKR, to match the native eIF4G peptide sequence. This modified peptide had an improved solubility and a 6-fold higher affinity for eIF4E than eIF4G. To investigate where this additional affinity was coming from, peptides were synthesized which introduced each residue from the modified peptide in to eIF4G individually. This study highlighted two key substitutions that improved the affinity of the eIF4G peptide, a D to S substitution at position 5 (4.6 fold improvement) and a G to A substitution at position 11 (2.3 fold improvement)[331]. This peptide was then constrained to further improve potency. Table 26 demonstrates that the location of the staple has a large effect on affinity (sTIP-01 compared to sTIP-02) The helicity of the stapled peptides were shown to be greater than that of the unstapled variant[332], and the stapled peptide which with the greatest helicity also had the greatest affinity. Competition fluorescence polarisation experiments showed the ability of stapled peptides to disrupt the eIF4E/eIF4G interaction, with the highest affinity analogue having an apparent K_d of 3 nM[332].

Table 26 – Peptide sequences, affinity and helicity of eIF4G peptide (including mutation D5S) and stapled variants. The symbol (*) signifies positions that are linked by the hydrocarbon staple.

Name	Sequence	K_d SPR/FP (nM)	Helicity (%)
eIF4G^{D5S}	KKRYSREFLLGF	99.9 ± 6.2 /195.2 ± 12.1	0
sTIP-01	KKRYSR*FLL*F	NA /558.0 ± 59.5	24
sTIP-02	KKRYSRE*LLG*	109.6 ± 4.6 /146.7 ± 1.7	18
sTIP-03	KKRYSRE*LL*R	3.4 ± 0.3 /4.4 ± 0.6	45

4.1.3 Structural characterization of eIF4E/eIF4G

The more structural and biophysical information available the more informed the rational design process. The crystal structure of eIF4G bound to eIF4E has previously been solved (PDB: 2W97)[174] (Figure 72). Mutagenesis studies have highlighted key binding residues of eIF4G (Y624, L629, Leu630, F628 and F632). Although not in the helix Y624 makes a significant contribution to binding by making multiple van der Waals contacts with eIF4E involving His37 and Val69[168].

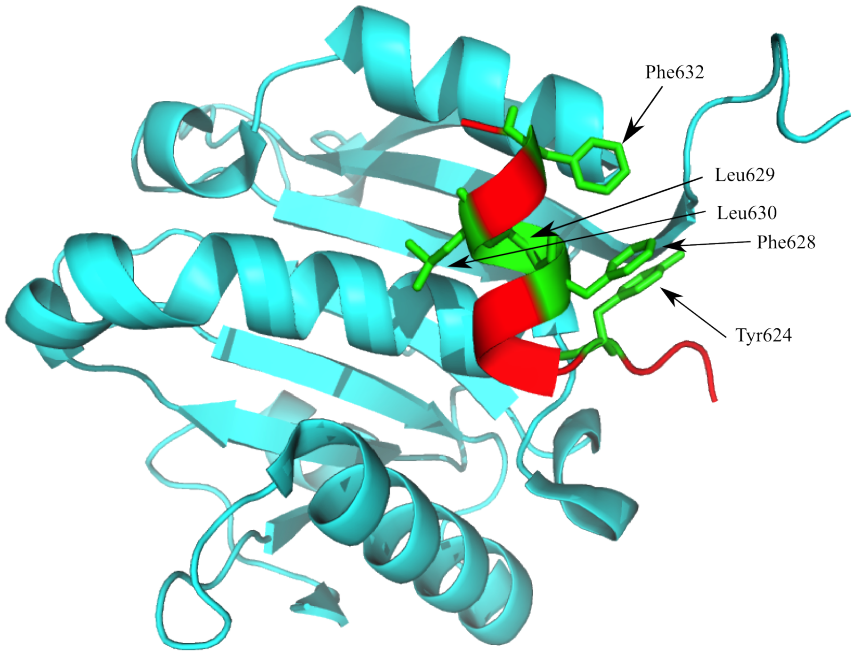


Figure 72: Crystal structure of eIF4G (red) binding eIF4E (cyan). Key binding residues of eIF4G are highlighted in green.

Complications in designing compounds may arise as not all the key binding residues are on the same face as the helix. Traditional α-helix mimetics have focused on designing scaffolds which mimic one face of the helix, although it should certainly be possible to design a small molecule scaffold which mimics side chain projections from more than one helical face. This chapter will discuss the work conducted to characterise the eIF4E/eIF4G interaction to inform rational drug design.

4.2 eIF4E/eIF4G Aims

The aims of this part of the project were to structurally and biophysically characterise the eIF4E/eIF4G interaction for the rational design of small molecule modulators able to disrupt the interaction.

4.3 Protein production

As with the work conducted for the distribution of the HIF-1α/p300 interaction the first step was to clone and express the protein. PCR was used to clone eIF4E into pET-SUMO, pET-GFP-28b-PreScission; and pGEX-6p-2. Expression trials for each of these constructs transformed into expression strains *E.coli* BL21 (DE3) pLysS Gold, Star and Rosetta 2 were conducted. Initial trials of small-scale cultures (10 mL) were produced and all were initially induced with 0.4 mM IPTG at 18 °C for 18 hours. SUMO-eIF4E (35 kDa), GST-eIF4E (48 kDa) and GFP-eIF4E (53 kDa) were all overexpressed, however all the protein appeared to be largely insoluble Figure 73. Further trials were conducted, investigating IPTG concentration, incubation temperature and incubation time, however eIF4E remained largely insoluble.

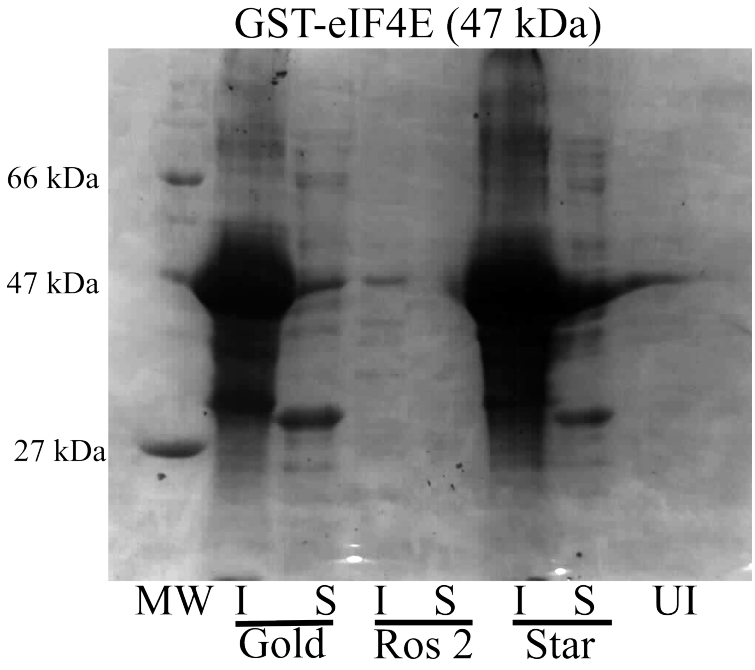


Figure 73 - SDS PAGE analysis of GST-eIF4E as part of the initial expression trials. MW = molecular weight marker, I = insoluble fraction. S = soluble fraction UI = uninduced

Autoinduction was used to try to obtain soluble protein. Autoinduction is generally considered to be a more sensitive approach to protein expression than IPTG induction[236], therefore it was hoped that a larger percentage of protein would remain soluble. The greatest expression by autoinduction was observed for the SUMO-eIF4E construct expressed from Star cells with a yield of approximately 33 mg SUMO-eIF4E from 400 mL growth culture. SUMO-eIF4E was purified by size exclusion chromatography and the major peak was collected (Figure 74), both SDS PAGE and mass spectrometry confirmed the size of the expressed protein (within the major peak) to be that of SUMO-eIF4E. The protein was confirmed to be folded by CD.

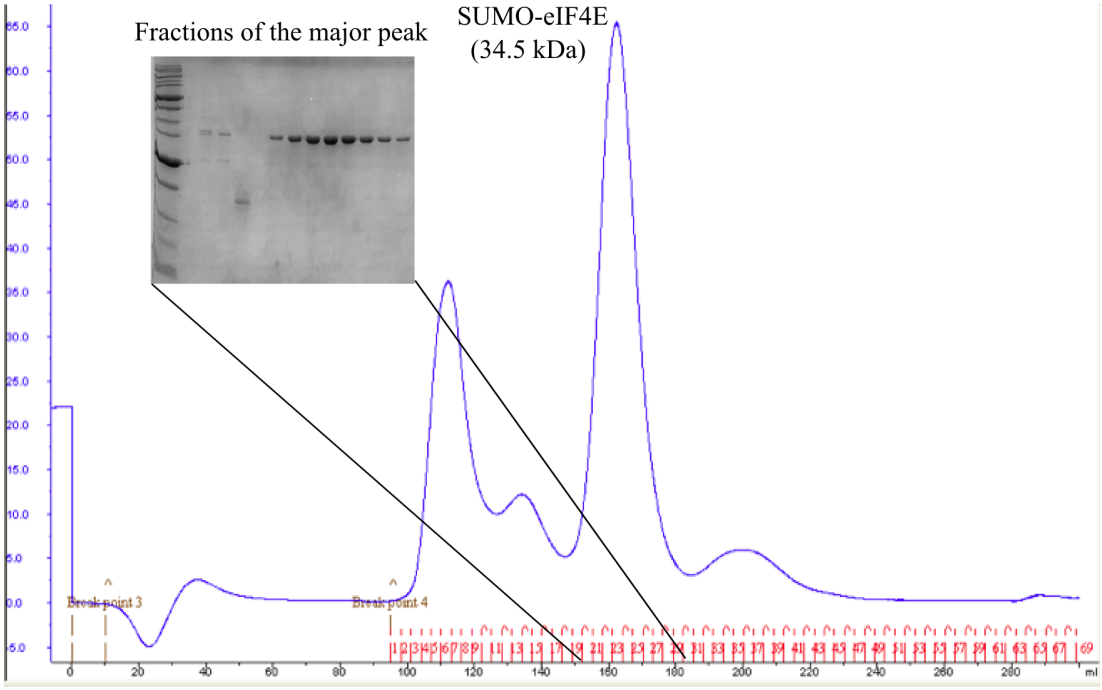


Figure 74 - Size exclusion chromatography trace for SUMO-eIF4E, the major peak was collected and the fractions run on an SDS PAGE gel (inset).

The SUMO was cleaved from eIF4E using SUMO protease by incubation for approximately 18 hours. Imidazole was removed by dialysis. The mixture was then applied to a second Ni^{2+} column, the eIF4E was eluted in the flow through, initial wash (0 % imidazole) and 50 mM imidazole wash. eIF4E was further purified by size exclusion chromatography.

The EIF4G peptides (RYDREFLLGFQ) both with and without a fluorescein tag were synthesised by solid phase synthesis by George Burslem (University of Leeds). The helicity of the eIF4G peptides was investigated by CD (Figure 75). In phosphate buffer (40 mM sodium phosphate pH 7.5, 200 mM sodium phosphate, 1 mM DTT and 5 % glycerol) both peptides adopted an irregular structure, however when solubilized in trifluoroethanol (TFE) both peptides appeared to contain an α helical component[134], although this spectra is not a classical α helical spectra.

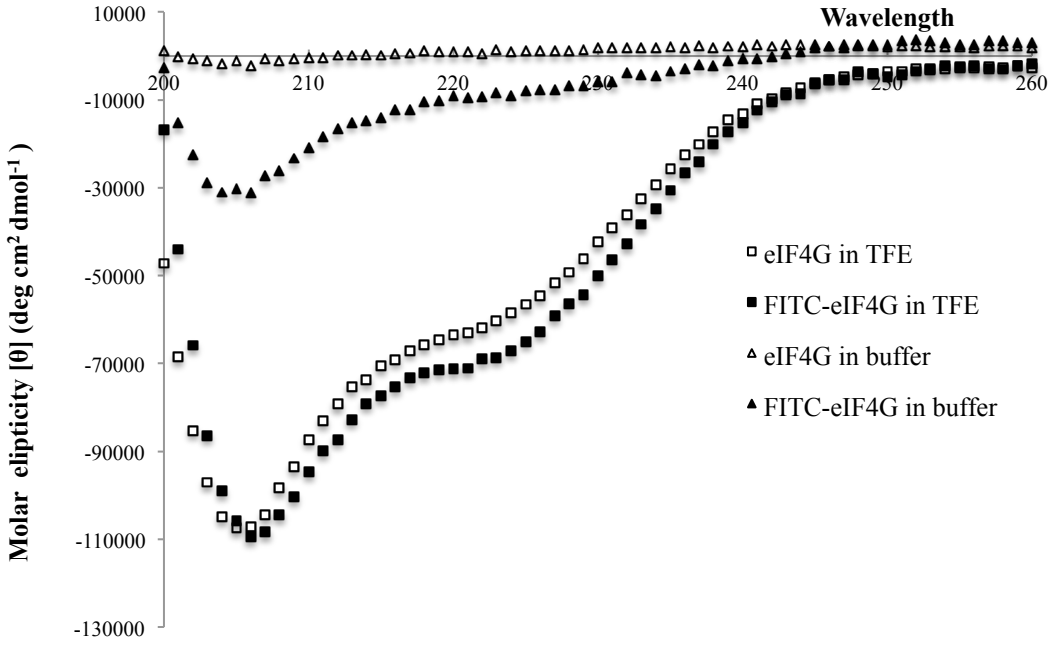


Figure 75 - CD analysis of the eIF4G peptide, both FITC labelled and unlabelled and in phosphate buffer and TFE.

4.4 eIF4E binding studies

The first binding study conducted was a fluorescence anisotropy direct binding assay. SUMO-eIF4E was titrated against 40 nM FITC-eIF4G, in buffer (40 mM sodium phosphate pH7.5, 200 mM sodium phosphate, 1 mm DTT and 5 % glycerol). A K_d of 1480 ± 730 nM was measured. To ensure that eIF4G was binding to eIF4E and not to SUMO the fluorescence anisotropy assay was conducted for eIF4E, a K_d = 710 ± 110 nM was measured (Figure 76). The control of a SUMO titration was also conducted to ensure that the binding was specific. No binding of FITC-eIF4G to SUMO was seen within the concentration range measured. Larger proteins should give greater changes in anisotropy; hence SUMO-eIF4E was used for further fluorescence anisotropy binding and competition assays (for a better signal to noise ratio).

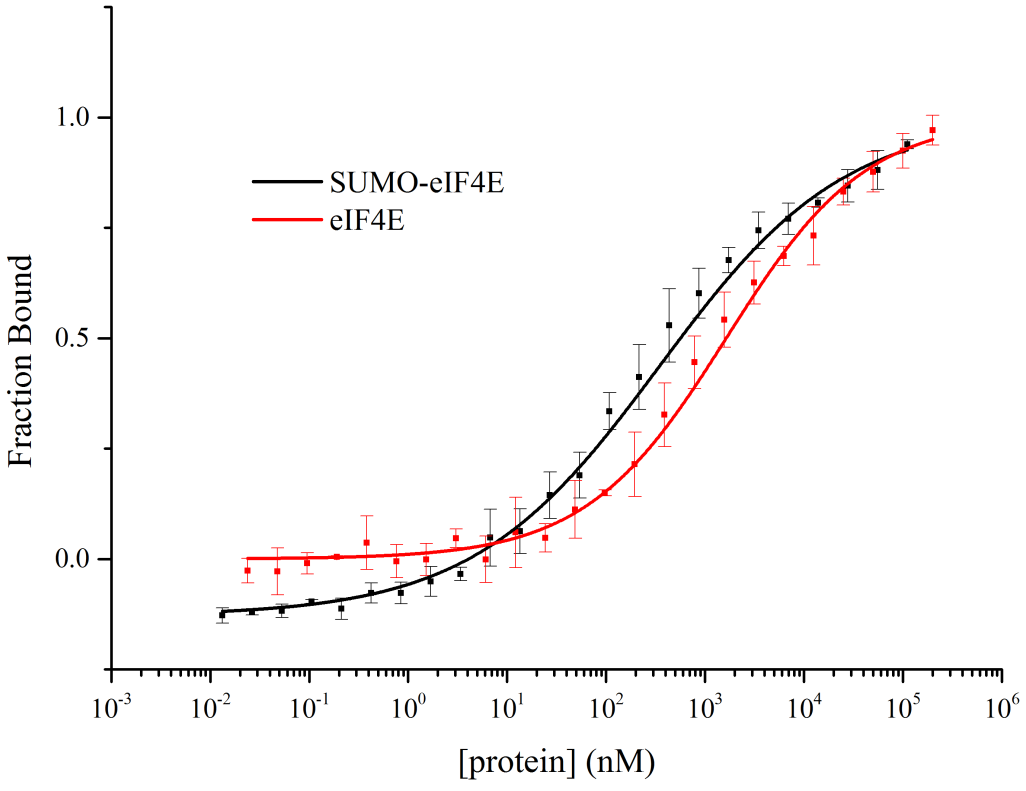


Figure 76 - Fluorescence anisotropy assay measuring the binding of SUMO-eIF4E (black) and eIF4E (red) binding to FITC-eIF4G.

The robustness of the assay was assessed by checking the variation at different time points (

Figure 77A) and temperatures (

Figure 77B). No change in affinity measured was observed between a 30 minute incubation and a 5 hour incubation (K_d after 30 minutes 1848 ± 252 nM and after 5 hours = 1754 ± 98 nM) (

Figure 77A). The next time tested was 18 hours, the data collected at this time point was poor. A change in anisotropy was seen in a temperature dependent manor, this change is most likely due to an increase of kinetic energy in the system at higher temperatures, resulting in a faster overall tumbling rate of both the unbound and bound fluorophores. Despite the change in anisotropy no temperature dependent affect on affinity was observed within the temperature range tested (K_d 21 °C = 1800 ± 300 nM, K_d 25 °C = 4100 K_d 30 °C = 3700 ± 400 and K_d 35 °C) (

Figure 77B).

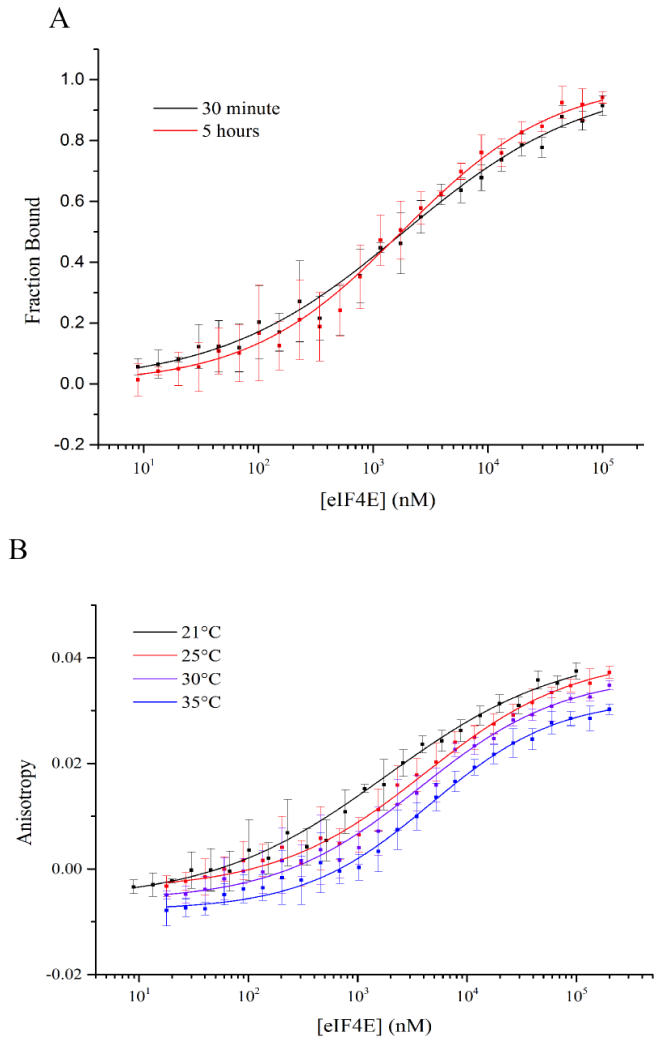


Figure 77 –Testing the robustness of the fluorescence anisotropy assay. A) testing the change in affinity of the eIF4E/eIF4G interaction over a 5 hour period, 30 minutes (black) and 5 hours (red). B) the change in affinity of the eIF4E/eIF4G interaction over a temperature range, 21 °C (black), 25 °C (red), 30 °C (purple) and 35 °C (blue).

It has been shown that the presence of eIF4G increases the affinity of Cap for eIF4E (approximately a 7-fold increase), although Cap does not bind to eIF4G itself[333]. A study to assess if the binding of Cap affects the affinity of eIF4G for eIF4E had previously been conducted. SPR was used to measure the K_d of the interaction in the presence and absence of Cap (m7GpppG). The results showed that eIF4G binds eIF4E and eIF4E·m7GpppG with the same affinity, although the authors recognised that this is contradictory to some previous studies[334, 335]. Due to this contradiction the effect of both Cap and Cap analogue (Figure 78) was tested in the fluorescence anisotropy assay (Figure 78). In this assay the presence of Cap and Cap analogue had no effect on the affinity of the eIF4E/eIF4G interaction (K_d with no additive =7740 ± 1400 nM, K_d + Cap analogue = 4810 ± 52 nM and K_d + Cap analogue = 7780 ± 1300 nM, due to the poorer quality data these data were fitted with a logistic fit rather than the K_d model fit). Therefore Cap or Cap analogue was not added to any further experiments.

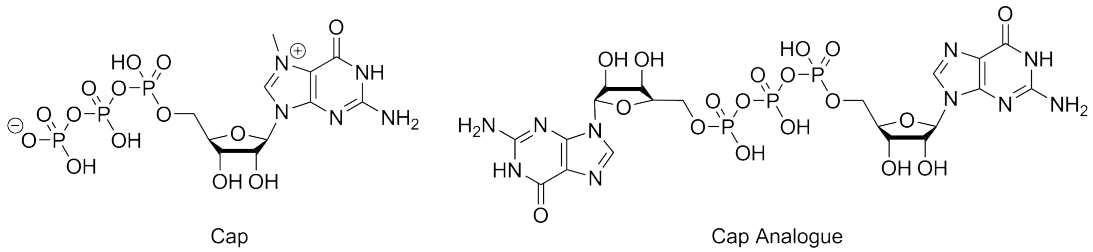


Figure 78 – Structures of Cap and Cap analogue, which have been tested in the fluorescence anisotropy assay.

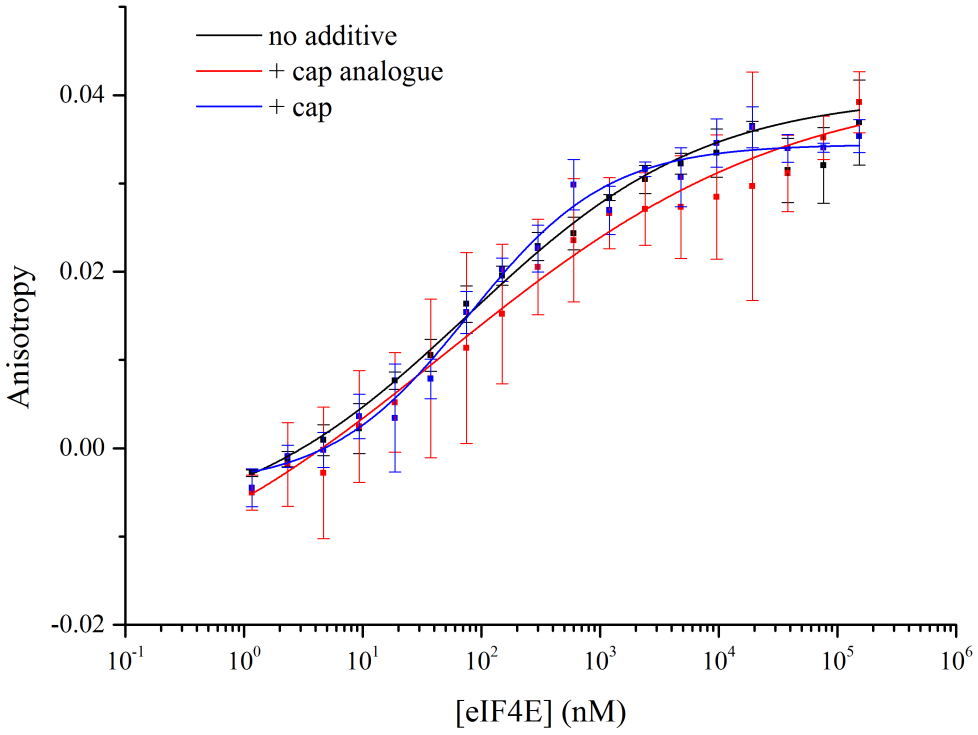


Figure 79 – Fluorescence anisotropy analysis of the effect of Cap (blue) and Cap analogue (red) on the binding of eIF4E (black) to eIF4G.

4.5 eIF4E/eIF4G inhibitor development

A fluorescence anisotropy competition assay was developed as a screen for potential inhibitors. To develop this assay the initial ligand used for disruption of the eIF4E/eIF4G interaction was unlabelled eIF4G. This ligand was able to disrupt the eIF4E/FITC-eIF4G interaction (1 μM eIF4E and 40 nM FITC-eIF4G) with an IC_{50} = 4878 ± 1103 nM (Figure 80).

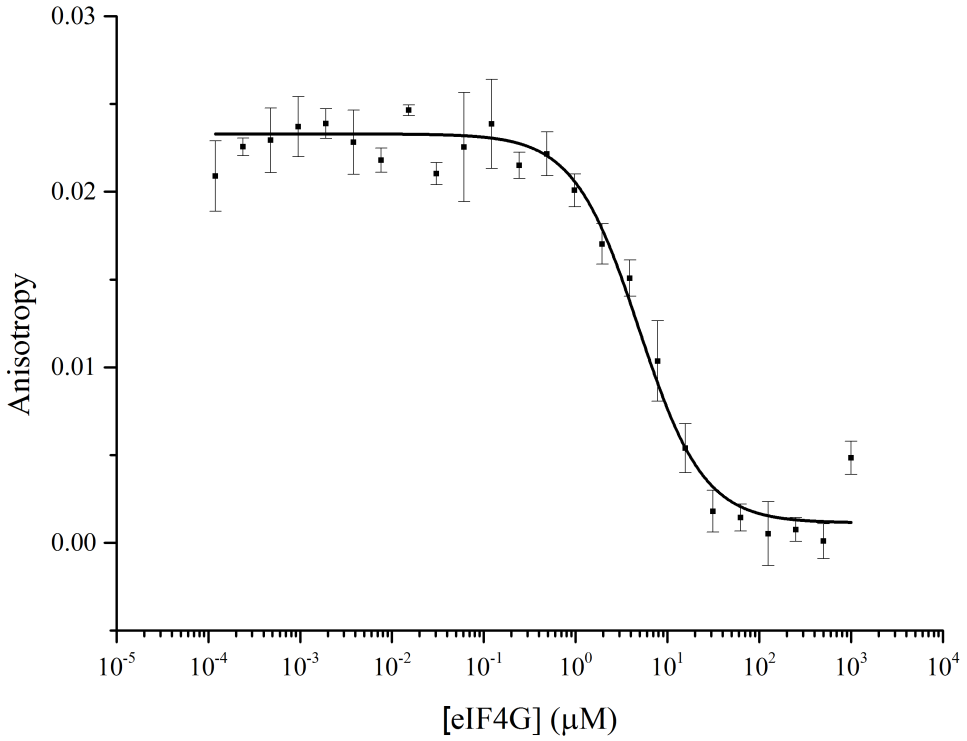


Figure 80 - Fluorescence anisotropy competition assay testing the inhibitor Capability of unlabelled eIF4G peptide

Initially, known inhibitors were selected and tested. The inhibitor 4E1RCat was measured to have an IC_{50} = 7.75 ± 0.41 μM, (**Figure 81**) this is similar to the published IC_{50} (≈ 4 μM[328]). The published inhibitor 4EGI-1 was also tested, however the inherent fluorescence of the compound disrupted the fluorescence anisotropy assay, therefore no IC_{50} was measured.

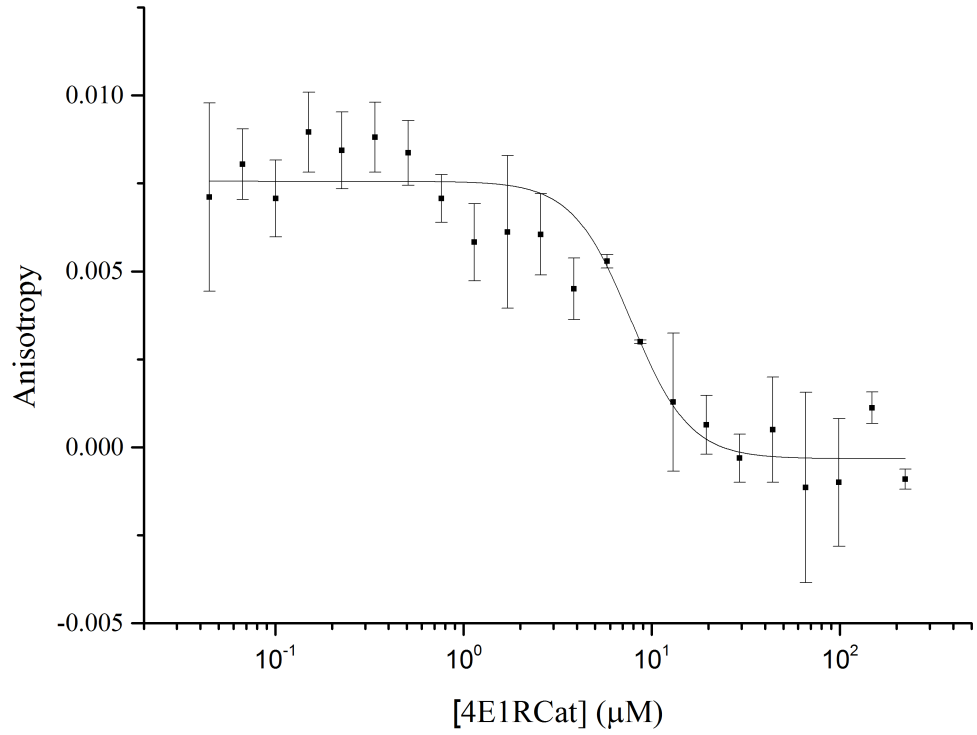


Figure 81 – Fluorescence anisotropy competition assay testing the inhibitor 4E1RCat

A rational design approach was employed to design small molecules to disrupt the eIF4E/eIF4G interaction. Initially the O-linked oligoamide helix mimetic scaffold[96] was used. Traditionally, as with the HIF-1α mimetics, this scaffold has been used to project 3 key side chains in the same spatial orientation as would be projected by the *i, i*+4 and *i*+7 side chains of an α-helix. A potential limitation of using this scaffold to disrupt this interaction is that the key side chains (Tyr_{624}, Phe_{628}, $Leu_{629,}$ Leu_{630} and Phe_{623}; highlighted as key binding residues by mutagenesis studies[168]) of eIF4G are not presented on one face of the helix. Two oligoamide scaffolds were tested (which had been previously tested against the HIF-1α/p300 interaction; synthesised by George Burselm, University of Leeds), one which included three phenylalanine side chains (oligoamide **6**) and the second which was designed to mimic HIF-1α, mimicking two leucine and a valine side chain (oligoamide **1**). Neither of these ligands disrupted interaction with an $IC_{50} < 300$ μM (Figure 81).

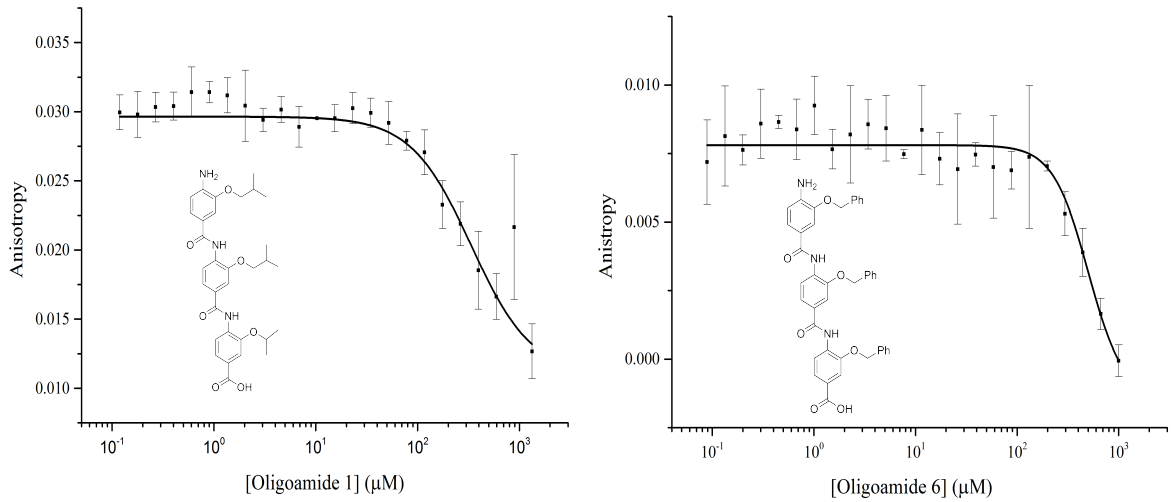


Figure 82 - Fluorescence anisotropy competition assay testing the alpha helix mimetics Oligoamide 1 and Oligamide 6.

As neither of the oligoamides disrupted the interaction with notable affinity a different scaffold was developed. It was hoped that this scaffold would project side chains Tyr, Phe and Leu, in a similar spatial orientation (from multiple helix faces) as the eIF4G helix. However this compound (eIF 1) did not disrupt the interaction with an IC_{50}<1mM (Figure 83), therefore inhibitor development for this target was not continued.

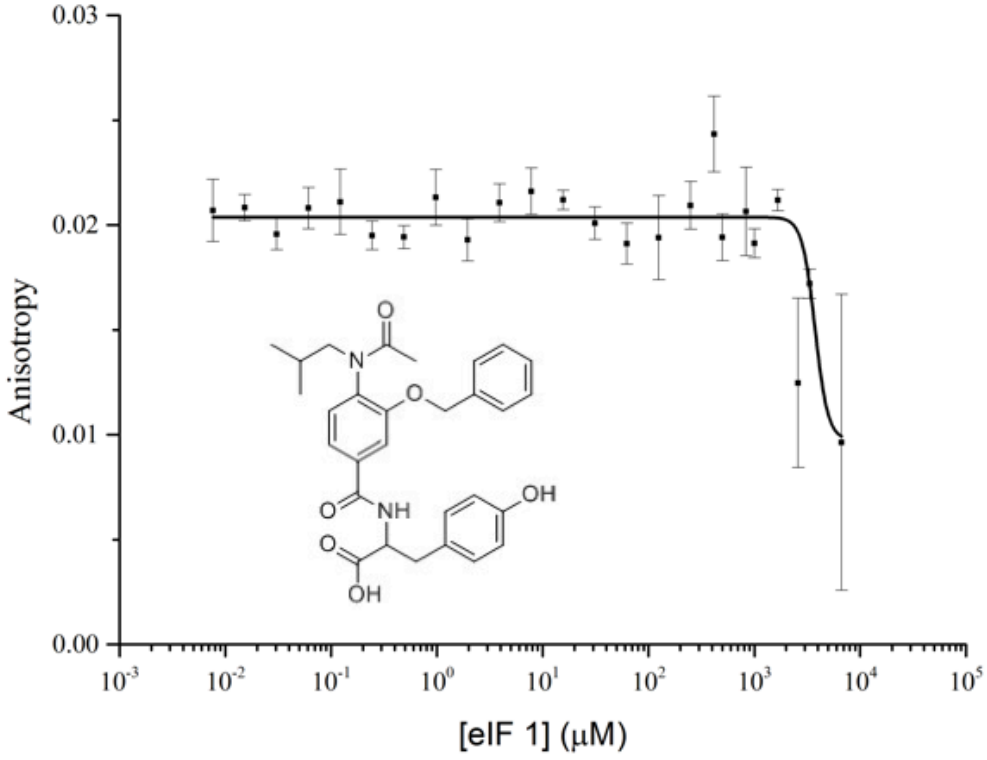


Figure 83 - Fluorescence anisotropy competition assay testing the designed eIF4G mimetic (shown in the insert)

4.6 Phage display

As with the HIF-1α/p300 study a peptide phage display experiment was performed to try and generate higher affinity peptides for eIF4E. As no inhibitors have been rationally designed, the discovery of a different pharmacophore would be a new starting point for the design process. It has been demonstrated, by solving the crystal structures of both 4E-BP (1EJ4) and eIF4G (1EJH) bound eIF4E that the two peptide binding regions are molecular mimics of each[312]. Both 4E-BP and eIF4G have the motif YXXXXLΦ, which becomes α-helical in the presence of eIF4E and is required for binding[312]. Phage display may identify a different binding motif or may show a preference for one of the central residues (X in the motif).

Initially eIF4E was biotinylated using the commercially available Sulfo-NHS-LC-biotin (Thermo Scientific), the activity was shown to be comparable to unlabelled eIF4G in the fluorescence anisotropy competition assay. The phage display libraries selected for use were the NEB 7mer and 12mer libraries. As with the p300 phage display experiment two buffer conditions were used (high and low salt) and the elution strategy used was competition with eIF4G peptide (500 μM). Three panning rounds were completed, the unpanned libraries, the first and third panning round clones were then sent for sequencing using the Illumina platform (sequencing results are summarised in (Table 27). An ELISA was conducted to assess the enrichment of the phage display libraries through the panning rounds. This ELISA measured the percent of binders in the eluted pool of each panning round. A general trend of enrichment was seen, however sequential enrichment through each round was not observed (Figure 84).

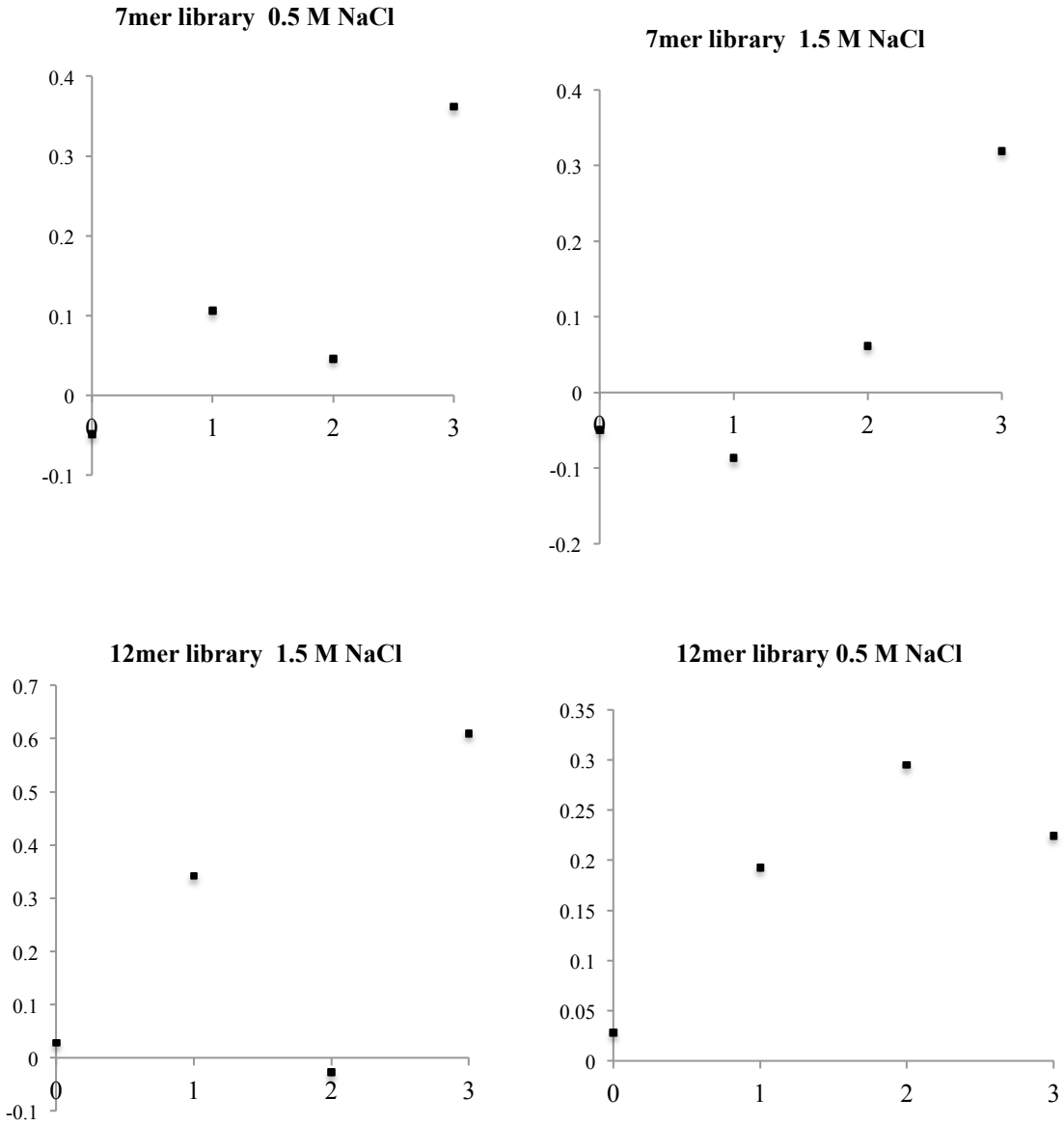


Figure 84 – Enrichment ELISA for each condition of the eIF4E phage display experiment, showing an enrichment from the unpanned library to the third panning round.

Table 27 – The most frequent 5 clones from the unpanned libraries, round 1 and round 3 and the frequency they appear in the pool

Condition	Unpanned sequence and frequency		Round 1 sequence and frequency		Round 3 sequence and frequency	
7 mer, standard salt	WSLSELH	1448	ATNLFKS	144	ATNLFKS	107307
	TTQVLEA	849	GTPLWMK	18	SLLGQTP	887
	IDRTQFM	667	MLADMHA	4	MNQASHY	810
	GTGSQAS	644	NPHVLRF	3	GLYPVSG	542
	SQNFVRE	492	IASTEKH	2	SAAWNKS	508
12 mer, standard salt,	GLHTSATNLYLH	3109	DWSSWVYRDPQT	7763	AADKVRPESWRM	894
	EGTSSWRYWLSP	2427	GLHTSATNLYLH	1898	HSKAFPVLYPLR	577
	ASISNGPLTGYR	1099	VHWDFRQWWQPS	1643	SFSPLTHMPPAG	404
	WPEFDILWAHPQ	340	SGVYKVAYDWQH	1264	SSDFLWNYRLLG	392
	AVHLRLDHLSVL	149	ASSAYLKSMDPA	1047	GNMAKARNGALY	333
7 mer, high salt	WSLSELH	1448	IDRTQFM	647	ATSSRLL	2376
	TTQVLEA	849	GLGFVGH	271	PLPVLLV	1619
	IDRTQFM	667	QGLILSR	263	FAQMRFV	937
	GTGSQAS	644	WSLSELH	257	HHVPGTV	839
	SQNFVRE	492	GNGMHSG	234	MVQWSVR	674
12 mer, high salt,	GLHTSATNLYLH	3109	GLPAVERIFFMA	11	EGTSSWRYWLSP	3059
	EGTSSWRYWLSP	2427	MDATTTLQRQLL	11	SPYSHVTMNSAG	54
	ASISNGPLTGYR	1099	AHEQVGYASQRA	8	VPGKVSFLSLPF	32
	WPEFDILWAHPQ	340	AAHSTLRQALAM	1	YVPGQSVIPDIR	32
	AVHLRLDHLSVL	149	AAPMIITLTSFG	1	GLHTSATNLYLH	28

The sequencing result did not clearly show one peptide which had been systematically enriched through the rounds of panning and no one peptide appeared in the elutes of multiple conditions. This indicated that the enrichment was poor. However two 12mer sequences were selected ^{1}AADKVRPESWRM^{12} (eIF PDDP1) and ^{1}HSKAFPVLYPLR^{12} (eIF PDDP2). Neither of these peptides showed the conserved motif of eIF4G and 4E-BP, although some comparisons can be made. The C terminal residue M^{12} of eIF PDDP1 is hydrophobic (in the same way as the hyrdrophobic residue 630 of eIF4G and 4E-BP. eIF PDDP2 contains a Tyr (position 9) and a Leu (in position 11), which is similar to the biding motif, however this peptide has a hydrophilic C terminus. These peptides were purchased with an N terminal fluorescein and tested for binding affinity using the fluorescence anisotropy assay (Figure 85). Neither of the phage display peptides bound with an affinity higher than 500 μM, which is lower affinity than the native eIF4G peptide, therefore neither peptide was taken forward for further study.

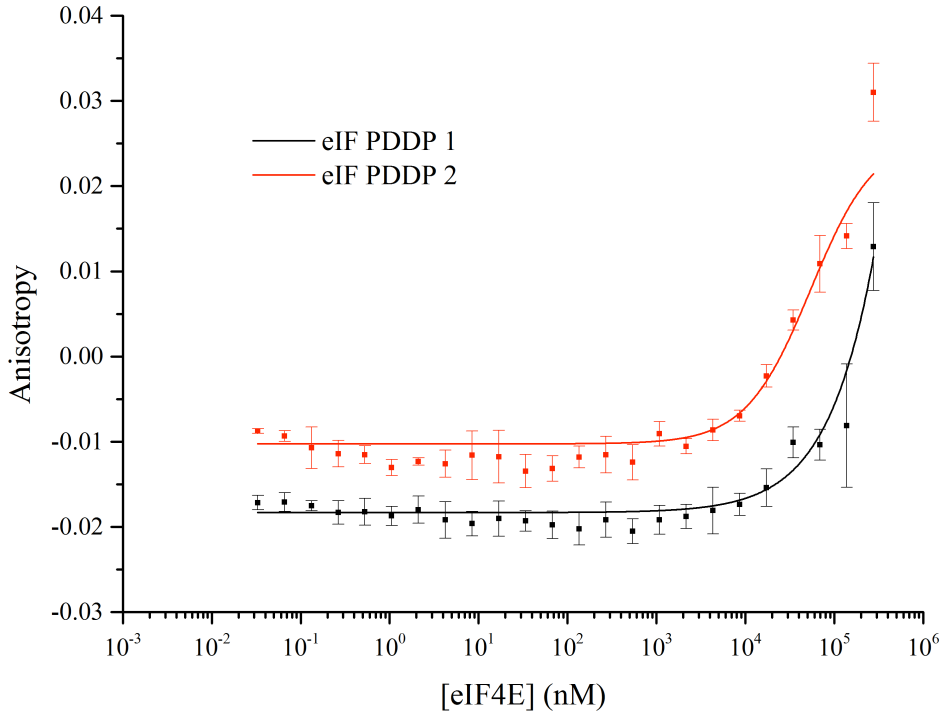


Figure 85 - Fluorescence anisotropy assay testing the binding of the peptides (eIF PDDP1, black and eIF PDDP 2 red) generated by phage display to eIF4E.

The low affinity of the peptide may be due to the high entropic cost of binding, the helical prediction software, Agadir[255], predicted each peptide to be only 2 % helical in isolation. The ability of the peptides to form a helix may therefore be poor. CD studies, as were conducted for the native eIF4G peptide, would confirm if the peptides were able to form helices when induced. Structural studies would confirm if the peptides bind in a helical confirmation.

4.7 eIF4E Crystallography

There have been several crystal structures published of eIF4E, both apo (PDB:3TF2[336], 2W97[174]) and in various complexes; bound to eIF4G (PDB:2W79[174]), 4E-BP (PDB:4UED[337], 3U7X[336], 3HX1[338], 1EJ4[312] and 1WKW[339]) and inhibitors (PDB:4TPW[327], 4BEA(stapled peptide)[332] and 4DT6[323]). However, none of these structures show eIF4E to be in a soakable form or indeed to bound to an eIF4G orthosteric inhibitor. To further the development of inhibitors a soakable crystal form that could be utilised for obtaining small molecule inhibitor bound structures would be immensely useful.

A pure and homogeneous sample of eIF4E was produced and screened at 8-10 mg/mL for initial crystallisation conditions using commercially available sparse matrix screens. The screens were set up in sitting-drops with various protein:mother liquor ratios across all conditions and at a range of temperatures. In parallel the literature condition, 20 % PEG 3350, 0.2 M diammonium citrate, 16 °C, from which apo_eIF4e crystals grew[336] was set up. Crystals grew in this condition and in one condition from the factorial screens, 20 % PEG 4000, 0.1 M sodium citrate pH 5.5 and 10 % 2-propanol (v/v). From this initial factorial hit and literature condition, optimization screens varying protein concentration, percent PEG 4000, pH, percent isopropanol and ratio of protein to mother liquor were set up. Microcrystals appeared in many of these conditions; the most promising condition was 17.5% PEG 4000, 0.2 M ammonium citrate, 25 °C; therefore an additive screen was conducted around this condition to improve crystal quality and size. A commercially available Additive Screen (Hampton Research) was used in 96-well sitting-drop format. This is a screen consisting of small molecules known to affect crystallisation by altering protein solubility, manipulating protein-solvent interactions, or perturbing the structure that water molecules form around the target protein. Several additives appeared to visually improve crystal quality, including sodium fluoride, gly-gly-gly, spermadine, trehalose and ethyl acetate. Multiple crystals were picked from each additive condition. Crystals were cooled supplemented with 20 % PEG 400 as cryo protectant.

Many crystals diffracted, three additive conditions (sodium fluoride, gly-gly-gly and spermidine) diffracted to 2.7Å, the highest resolution data came from the condition 17.5% PEG 4000, 0.2 M ammonium citrate, 25 °C with the additive trehalose and diffracted to 2.45 Å, the crystal structure was solved by molecular replacement Table 28.

Table 28 - Crystallographic data and refinement statistics for apo_eIF4E

Data set	**eIF4E-apo**
Wavelength (Å)	0.97949
Space group	$P2_1$
Cell parameters (Å,º)	a = 39.263 b = 110.333 c = 81.118 α= 90.00 β = 101.77 γ = 90.00
Total reflections	65,872 (9,483)
Unique reflections	22,265 (3,253)
Resolution shells(Å)	
Low	39.79-8.03
High	2.68-2.54
R_{merge} (%)	9.1 (41.3)
R_{free} (%)	27.89
R_{pim} (%)	6.9 (33.3)
R_{cryst} (%)	24.88
Completeness (%)	99.3 (99.7)
Redundancy	3.0 (2.9)
I/s(I)	6.9 (2.3)
V_M (Å^3/Da)	1.71
Mol. per AU	4
Reflections working set	21152
Free R-value set (No. of reflections)	5.0% (1113)
RMSD bond lengths (Å)	0.0048
RMSD bond angles (º)	0.916
No. of atoms used in refinement	
Non-hydrogen atoms	5,788
Water molecules	0
Mean B value (Å^2)	
Total	49.752
Water molecules	N/A
Ramachandran plot statistics (%)	
Preferred region	96.3
Allowed region	3.4
Outliers	0.3

It was initially hoped that the eIF4E would form a soakable crystal system, as a soakable crystal form maybe used for the hit discovery and optimisation.

However the eIF4G binding site appears to be enclosed within the crystal form and is not exposed to solvent channels (Figure 86). Therefore the apo-eIF4e crystals are unlikely to be a soakable system. Consequently, co-crystals structures of eIF4E with inhibitor bound will need to be made via co-purification of eIF4e and inhibitor, unless an alternative crystal condition is established which generates a soakable crystal form.

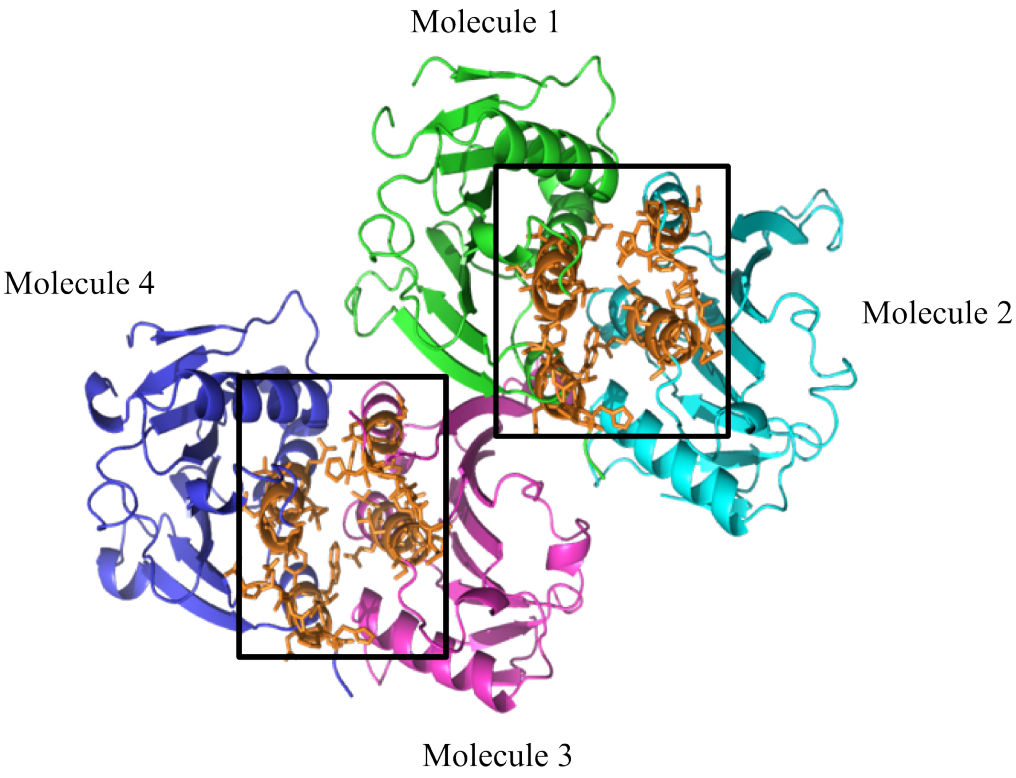


Figure 86 - Crystal structure of apo-eIF4e. Each of the 4 molecules in the asymmetric unit is coloured separately with the eIF4G binding site coloured orange. The squares highlight the non-solvent accessible binding site.

Attempts have been made to co-purify eIF4E and the inhibitor 4E1RCat, however the solubility limit of 4E1RCat has restricted the co-purification as the complex is not stable to gel filtration. Analogous compounds, which were hoped to be more soluble, were purchased and co-crystallisation was attempted, however all crystals that grew were in the apo form. None of these crystals differed in form to those structures previously solved therefore no further optimisations were made. Unfortunately as none of the ligands rationally designed in this study disrupted the interaction with a sufficient potency, co-crystal trials with these ligands were not possible. eIF4E appears to crystallise readily in many different conditions, as seen

from the number of crystal structures in the literature and the number of crystals obtained during this study, therefore a soakable crystal form may still be achievable. Many more crystals may be required to be screened; with the advances in high throughput crystallography this should be feasible.

4.8 eIF4E/eIF4G Summary and Further Work

In this project eIF4E was cloned, expressed and biophysically characterised for eIF4G binding. The disruption of the eIF4E/eIF4G interaction is a therapeutically relevant anticancer target, as the prevention of the binding of eIF4E to eIF4G would limit the formation of the eIF4F complex, resulting in a decrease in the cap dependent translation of proteins associated with cancer development and progression. This project has cloned and expressed the eIF4E protein and biophysically assessed its ability to form a complex with eIF4G in the presence and absence of mRNA cap. A robust assay has been established for the testing of potential inhibitors.

As the first generation ligands and phage display peptides did not bind to eIF4E or disrupt the eIF4e/eIF4G with notable affinity, much of the emphasis of the overall project focused on the HIF-1α/p300. However the robust biophysical assays for the testing of potential inhibitors has left a strong platform to build upon. Further work may consist of the development of a helix mimetic scaffold able to mimic more than one helical face. The Hamilton group have recently described a variation of the teryphlyl scaffold that mimics an additional residue in the i + 5 position (therefore projecting side chains off multiple faces) by incorporating an indane ring in the central position[340], although no inhibition of PPIs has been described as yet. The peptides generated by the phage display experiment may be a good starting point for the generation high affinity peptide binders, previous phage display experiments for eIF4E yielded peptides which required much optimization before the nM ligand was produced[168, 331]. Mutagenesis or peptide constraints could be used to improve this peptide, however evidence that the peptide is binding in a helical confirmation (as with eIF4G) must be obtained first. Phage display could also be used to develop a new pharmacophore, the peptide phage display did not yield high affinity binding peptides, however other phage display experiments using scaffolds such as the Adhiron scaffold[122] may be more successful.

Chapter 5 : Overall conclusion and discussion

This project has focussed on increasing our understanding of two therapeutically relevant protein-protein interactions, HIF-1α/p300 and eIF4E/eIF4G, which have been highlighted as a targeted anticancer therapy. Using both biophysical and structural techniques with the overall aim of informing rational inhibitor design.

Some of the difficulties in targeting these interactions can be attributed to the fact that they are protein-protein interactions. PPIs, in contrast too more traditional drug targets such as enzymes, have larger interacting surfaces which lack clearly defined pockets. This makes potent and selective inhibition more challenging. Another reason that enzymes have been targeted more successfully is the availability of transition-state analogues, which can be used as a starting point for inhibitor design. Theses difficulties have limited the success of PPI inhibitor development[9, 10, 341, 342].

It was hypothesised that increasing our understanding of an interaction, both biophysically and structurally, we can increase the success rate of PPI inhibitor development by designing small molecules more rationally rather than using a high throughput library screening method[343, 344]. During this project characterising the interaction successfully informed inhibitor design, particularly for the HIF-1α/p300 interaction. The difficulty with disrupting the HIF-1α/p300 interaction is that due to the size of the interaction a conventional small molecule[11] will not be able to cover the whole interacting area, as NMR studies[182, 197] have shown HIF-1α to have three helical regions, which wrap around p300 (Figure 17). Therefore a detailed analysis to locate binding hot-spots would be an integral part of the drug design process. Previous analysis had highlighted the C-terminus of HIF-1α to be important for binding[220], including helix 2 and helix 3. Residues of both of these helices have been shown by mutational studies to be key for binding (Cys_{800} and Asn_{803} (helix 2) and Leu_{818}, Leu_{822} and Val_{825} (helix 3))[216, 218, 219].

In this study two approaches to probe the HIF-1α binding surface on p300 were used to refine our understanding of the most productive regions to target using designed small-molecules: first, by analysis of the binding of shorter HIF-1α peptide fragments; and second, by phage display experiments. Binding analysis of fragments

of the native peptide allows identification of the highest affinity region of the HIF-1α peptide, whereas phage display permits the unbiased exploration of the p300 CH1 protein surface to discover high affinity binders. Two phage display technologies were used[237]: a peptide phage library (NEB[238-240]) and an Adhiron[122] phage library.

The binding analysis of fragments of the native HIF-1α peptide revealed HIF-1α helix 3 to have the highest affinity of any short region (< 15 residues) of the native HIF-1α sequence. This was supported by mutational studies; mutations in the HIF-1α helix 3 binding pocket of p300 reduced the binding affinity of the full length HIF-1α. The peptide phage display experiment also highlighted helix 3 to be a key binding area. A higher affinity peptide was generated and NMR studies indicated that it bound in or around the helix 3 pocket[182, 197]. This was again supported by mutation studies. The information that HIF-1α helix 3 is a high affinity binding area was used to inform inhibitor design. An α helix mimetic which projected the three key side chains of HIF-1α helix 3 was designed and synthesised[244], and disrupted the interaction in with an IC_{50} in the low micromolar range[244]. Therefore the understanding of the location of a binding hot-spot the of the HIF-1α/p300 interaction has lead directly to the design of a ligand which is able to disrupt the interaction.

The Adhiron phage display experiment yielded adhirons which bound to p300 with an affinity of approximately 100 nM and were able to disrupt the interaction with an IC_{50} = 1-5 μM. This is significantly better than fragments of native HIF-1α, or the phage peptides (which were unable to disrupt the interaction). After solving the crystal structure of the Adhiron, docking suggested that the Adhiron bound in a different location to helix 3 (Figure 54). The binding site of the Adhiron needs to be validated by structural studies as the location and binding mode of the adhirons will provide new information on suitable chemotypes for orthosteric small-molecule inhibitor development. As the Adhirons themselves inhibit the HIF-1α/p300 interaction with a low micromolar IC_{50}; mimicking the pharmacophore of the loops in a small molecule or using fragment based approaches has the potential to generate novel high affinity inhibitors.

One potential limitation is that throughout the studies the recombinant protein used was expressed using a bacterial protein expression system (*E. coli*). This expression system is used extensively[345], generally without problems with many proteins can now be produced routinely in secreted form with yields in the gram/litre scale[346] this system was chosen for these studies due to the low cost and potential for high protein yields. However, as a prokaryotic expression system, *E. coli* cannot perform post-translational modifications, which can be required for protein folding and protein functionality[345, 347, 348]. *E. coli* is not the only bacterial expression system, the soil bacterium *B. subtilis* can be used to produce recombinant protein and has several advantages, it does not produce lipopolysaccharide, a common byproduct of *E. coli* that can occasionally result in some degenerative disorders in humans and animals. Also, the bacterium can secrete some well-processed proteins directly into medium, facilitating further purification steps[349]. Furthermore, this cell type can grow to very high densities in relatively simple and cheap medium[345]. However problems remain which include degradation by proteases and instability of plasmids.

Eukaryotic expression systems such as yeast, filamentous fungus, insect/baculovirus systems and mammalian cells will overcome the limitation of the lack of post-translational modifications. Each of these systems has its own advantages and drawbacks. Yeast, like *E. coli*, has the advantage of culture simplicity, rapid growth, and low cost, but in addition the enriched endomembrane system of yeasts allows some intracellularly synthesized proteins to be secreted into the extracellular environment. As a unicellular eukaryote, yeast can potentially produce soluble, correctly folded recombinant proteins that have undergone all the post-translational modifications that are essential for their functions[345, 350]. Yeast is considered easier to culture and manipulate genetically than other eukaryotic expression systems and can be grown to higher cell densities[345]. Filamentous fungus has been used for protein expression, however the production of heterologous proteins is often limited[345]. One of the major advantages of baculovirus–insect expression systems is the eukaryotic protein processing capabilities of the host. Consequently, these systems are widely considered to be excellent tools for recombinant glycoprotein production. Although, the heterologous gene cannot be expressed continuously as the host cells infected by the nuclear polyhedrosis virus will eventually die. Also every round of synthesis of the protein of interest requires the infection of new insect cells,

therefore this system is limited in terms of high –throughput method. Finally mammalian cell expression systems have the distinct advantage that they promote signal synthesis, process, and can secrete and glycosylate proteins, particularly eukaryotic proteins. However this system has a relatively high cost and more complex due to the greater potential for contamination. A summary of the different expression systems is in Table 29.

Table 29 – Summary of commonly used expression systems[345]

System	Cell growth	Medium cost	Expression level	Post-translational modifications					
				N-linked glycosylation	O-linked glycosylation	Phosphorylation	Acetylation	Acylation	Gamma-carboxylation
E. coli	Rapid	Low	High	No	No	No	No	No	No
Yeast	Rapid	Low	Low-High	High mannose	Yes	Yes	Yes	Yes	No
Insect	Slow	High	Low-High	Simple	Yes	Yes	Yes	Yes	Yes
Mammalian	Slow	High	Low-Moderate	Yes	Yes	Yes	Yes	Yes	Yes

Of course there is much additional work that would allow this project to to design a useable therapeutic. However the combination of a detailed analysis of the interaction between HIF-1α and p300, along with two phage display techniques, has provided detailed information that could inform design of highly potent compounds to disrupt an extremely challenging pharmaceutically important protein-protein interaction. This work has prove the principle that biophysical and structural studies can be used in to characterise complex protein-protein interactions, and the information from such studies can be used to inform inhibitors which can be optimized as drugs for therapeutically important diseases

References

1. T. L. Nero, C. J. Morton, J. K. Holien, J. Wielens and M. W. Parker, *Nat Rev Cancer*, 2014, **14**, 248-262.
2. M. R. Arkin and J. A. Wells, *Nat Rev Drug Discov*, 2004, **3**, 301-317.
3. N. M. Baker and C. J. Der, *Nature*, 2013, **497**, 577-578.
4. M. Raj, B. N. Bullock and P. S. Arora, *Bioorg Med Chem*, 2013, **21**, 4051-4057.
5. J. A. Wells and C. L. McClendon, *Nature*, 2007, **450**, 1001-1009.
6. A. L. Garner and K. D. Janda, *Curr Top Med Chem*, 2011, **11**, 258-280.
7. A. J. Wilson, *Chem Soc Rev*, 2009, **38**, 3289-3300.
8. R. E. Babine and S. L. Bender, *Chem Rev*, 1997, **97**, 1359-1472.
9. S. Jones and J. M. Thornton, *Proc Natl Acad Sci U S A*, 1996, **93**, 13-20.
10. L. Lo Conte, C. Chothia and J. Janin, *J Mol Biol*, 1999, **285**, 2177-2198.
11. C. A. Lipinski, *J Pharmacol Toxicol Methods*, 2000, **44**, 235-249.
12. T. Clackson and J. A. Wells, *Science*, 1995, **267**, 383-386.
13. A. A. Bogan and K. S. Thorn, *J Mol Biol*, 1998, **280**, 1-9.
14. Z. Hu, B. Ma, H. Wolfson and R. Nussinov, *Proteins*, 2000, **39**, 331-342.
15. W. L. DeLano, M. H. Ultsch, A. M. de Vos and J. A. Wells, *Science*, 2000, **287**, 1279-1283.
16. I. Luque and E. Freire, *Proteins*, 2000, **Suppl 4**, 63-71.
17. B. Ma, M. Shatsky, H. J. Wolfson and R. Nussinov, *Protein Sci*, 2002, **11**, 184-197.
18. S. Atwell, M. Ultsch, A. M. De Vos and J. A. Wells, *Science*, 1997, **278**, 1125-1128.
19. S. Jones and J. M. Thornton, *Prog Biophys Mol Biol*, 1995, **63**, 31-65.
20. B. N. Bullock, A. L. Jochim and P. S. Arora, *J Am Chem Soc*, 2011, **133**, 14220-14223.
21. T. Kortemme and D. Baker, *Proc Natl Acad Sci U S A*, 2002, **99**, 14116-14121.
22. F. B. Sheinerman, R. Norel and B. Honig, *Curr Opin Struct Biol*, 2000, **10**, 153-159.
23. O. Keskin, A. Gursoy, B. Ma and R. Nussinov, *Chem Rev*, 2008, **108**, 1225-1244.
24. S. P. Brown and P. J. Hajduk, *ChemMedChem*, 2006, **1**, 70-72.
25. L. T. Vassilev, B. T. Vu, B. Graves, D. Carvajal, F. Podlaski, Z. Filipovic, N. Kong, U. Kammlott, C. Lukacs, C. Klein, N. Fotouhi and E. A. Liu, *Science*, 2004, **303**, 844-848.
26. P. Chene, *Nat Rev Cancer*, 2003, **3**, 102-109.
27. M. Wade, Y. C. Li and G. M. Wahl, *Nat Rev Cancer*, 2013, **13**, 83-96.
28. M. J. Keiser, B. L. Roth, B. N. Armbruster, P. Ernsberger, J. J. Irwin and B. K. Shoichet, *Nat Biotechnol*, 2007, **25**, 197-206.
29. R. A. Carr, M. Congreve, C. W. Murray and D. C. Rees, *Drug Discov Today*, 2005, **10**, 987-992.
30. C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv Drug Deliv Rev*, 2001, **46**, 3-26.

31. T. Oltersdorf, S. W. Elmore, A. R. Shoemaker, R. C. Armstrong, D. J. Augeri, B. A. Belli, M. Bruncko, T. L. Deckwerth, J. Dinges, P. J. Hajduk, M. K. Joseph, S. Kitada, S. J. Korsmeyer, A. R. Kunzer, A. Letai, C. Li, M. J. Mitten, D. G. Nettesheim, S. Ng, P. M. Nimmer, J. M. O'Connor, A. Oleksijew, A. M. Petros, J. C. Reed, W. Shen, S. K. Tahir, C. B. Thompson, K. J. Tomaselli, B. Wang, M. D. Wendt, H. Zhang, S. W. Fesik and S. H. Rosenberg, *Nature*, 2005, **435**, 677-681.
32. J. C. Reed, *J Clin Oncol*, 1999, **17**, 2941-2953.
33. B. Vu, P. Wovkulich, G. Pizzolato, A. Lovey, Q. Ding, N. Jiang, J. J. Liu, C. Zhao, K. Glenn, Y. Wen, C. Tovar, K. Packman, L. Vassilev and B. Graves, *ACS Med Chem Lett*, 2013, **4**, 466-469.
34. K. H. Emami, C. Nguyen, H. Ma, D. H. Kim, K. W. Jeong, M. Eguchi, R. T. Moon, J. L. Teo, H. Y. Kim, S. H. Moon, J. R. Ha and M. Kahn, *Proc Natl Acad Sci U S A*, 2004, **101**, 12682-12687.
35. S. Fulda and D. Vucic, *Nat Rev Drug Discov*, 2012, **11**, 109-124.
36. J. Bellmunt, C. Theodore, T. Demkov, B. Komyakov, L. Sengelov, G. Daugaard, A. Caty, J. Carles, A. Jagiello-Gruszfeld, O. Karyakin, F. M. Delgado, P. Hurteloup, E. Winquist, N. Morsli, Y. Salhi, S. Culine and H. von der Maase, *J Clin Oncol*, 2009, **27**, 4454-4461.
37. V. K. Ngan, K. Bellman, D. Panda, B. T. Hill, M. A. Jordan and L. Wilson, *Cancer Res*, 2000, **60**, 5045-5051.
38. M. J. Towle, K. A. Salvato, J. Budrow, B. F. Wels, G. Kuznetsov, K. K. Aalfs, S. Welsh, W. Zheng, B. M. Seletsky, M. H. Palme, G. J. Habgood, L. A. Singer, L. V. Dipietro, Y. Wang, J. J. Chen, D. A. Quincy, A. Davis, K. Yoshimatsu, Y. Kishi, M. J. Yu and B. A. Littlefield, *Cancer Res*, 2001, **61**, 1013-1021.
39. P. G. Morris, *Anticancer Drugs*, 2010, **21**, 885-889.
40. F. Y. Lee, R. Borzilleri, C. R. Fairchild, S. H. Kim, B. H. Long, C. Reventos-Suarez, G. D. Vite, W. C. Rose and R. A. Kramer, *Clin Cancer Res*, 2001, **7**, 1429-1437.
41. E. Weisberg, A. Ray, R. Barrett, E. Nelson, A. L. Christie, D. Porter, C. Straub, L. Zawel, J. F. Daley, S. Lazo-Kallanian, R. Stone, I. Galinsky, D. Frank, A. L. Kung and J. D. Griffin, *Leukemia*, 2010, **24**, 2100-2109.
42. L. Gandhi, D. R. Camidge, M. Ribeiro de Oliveira, P. Bonomi, D. Gandara, D. Khaira, C. L. Hann, E. M. McKeegan, E. Litvinovich, P. M. Hemken, C. Dive, S. H. Enschede, C. Nolan, Y. L. Chiu, T. Busman, H. Xiong, A. P. Krivoshik, R. Humerickhouse, G. I. Shapiro and C. M. Rudin, *J Clin Oncol*, 2011, **29**, 909-916.
43. M. E. Gore, M. L. Harrison and A. Montes, *Expert Rev Anticancer Ther*, 2007, **7**, 57-71.
44. R. S. Heist, J. Fain, B. Chinnasami, W. Khan, J. R. Molina, L. V. Sequist, J. S. Temel, P. Fidias, V. Brainerd, L. Leopold and T. J. Lynch, *J Thorac Oncol*, 2010, **5**, 1637-1643.
45. G. Liu, W. K. Kelly, G. Wilding, L. Leopold, K. Brill and B. Somer, *Clin Cancer Res*, 2009, **15**, 3172-3176.
46. A. C. Mita, L. J. Denis, E. K. Rowinsky, J. S. Debono, A. D. Goetz, L. Ochoa, B. Forouzesh, M. Beeram, A. Patnaik, K. Molpus, D. Semiond, M. Besenval and A. W. Tolcher, *Clin Cancer Res*, 2009, **15**, 723-730.
47. A. M. Hunter, E. C. LaCasse and R. G. Korneluk, *Apoptosis*, 2007, **12**, 1543-1568.

48. E. Pasquier and M. Kavallaris, *IUBMB Life*, 2008, **60**, 165-170.
49. J. B. Easton and P. J. Houghton, *Oncogene*, 2006, **25**, 6436-6446.
50. B. Aghdasi, K. Ye, A. Resnick, A. Huang, H. C. Ha, X. Guo, T. M. Dawson, V. L. Dawson and S. H. Snyder, *Proc Natl Acad Sci U S A*, 2001, **98**, 2425-2430.
51. M. M. Babu, R. van der Lee, N. S. de Groot and J. Gsponer, *Curr Opin Struct Biol*, 2011, **21**, 432-440.
52. V. N. Uversky and A. K. Dunker, *Biochim Biophys Acta*, 2010, **1804**, 1231-1264.
53. J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton and D. T. Jones, *J Mol Biol*, 2004, **337**, 635-645.
54. F. Diella, N. Haslam, C. Chica, A. Budd, S. Michael, N. P. Brown, G. Trave and T. J. Gibson, *Front Biosci*, 2008, **13**, 6580-6603.
55. J. Gsponer and M. M. Babu, *Prog Biophys Mol Biol*, 2009, **99**, 94-103.
56. Y. Cheng, T. LeGall, C. J. Oldfield, J. P. Mueller, Y. Y. Van, P. Romero, M. S. Cortese, V. N. Uversky and A. K. Dunker, *Trends Biotechnol*, 2006, **24**, 435-442.
57. V. Lounnas, T. Ritschel, J. Kelder, R. McGuire, R. P. Bywater and N. Foloppe, *Comput Struct Biotechnol J*, 2013, **5**, e201302011.
58. A. C. Anderson, *Chem Biol*, 2003, **10**, 787-797.
59. M. Congreve, C. W. Murray and T. L. Blundell, *Drug Discov Today*, 2005, **10**, 895-907.
60. J. W. Nissink, C. Murray, M. Hartshorn, M. L. Verdonk, J. C. Cole and R. Taylor, *Proteins*, 2002, **49**, 457-471.
61. R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *J Med Chem*, 2004, **47**, 1739-1749.
62. R. D. Malmstrom and S. J. Watowich, *J Chem Inf Model*, 2011, **51**, 1648-1655.
63. N. Singh and A. Warshel, *Proteins*, 2010, **78**, 1705-1723.
64. T. Lanio, A. Jeltsch and A. Pingoud, *Methods Mol Biol*, 2003, **205**, 199-203.
65. C. Scheich, V. Sievert and K. Bussow, *BMC Biotechnol*, 2003, **3**, 12.
66. R. Hui and A. Edwards, *J Struct Biol*, 2003, **142**, 154-161.
67. N. E. Chayen and E. Saridakis, *Acta Crystallogr D Biol Crystallogr*, 2002, **58**, 921-927.
68. L. Stewart, R. Clark and C. Behnke, *Drug Discov Today*, 2002, **7**, 187-196.
69. E. R. Bodenstaff, F. J. Hoedemaeker, M. E. Kuil, H. P. de Vrind and J. P. Abrahams, *Acta Crystallogr D Biol Crystallogr*, 2002, **58**, 1901-1906.
70. S. W. Muchmore, J. Olson, R. Jones, J. Pan, M. Blum, J. Greer, S. M. Merrick, P. Magdalinos and V. L. Nienaber, *Structure*, 2000, **8**, R243-246.
71. T. L. Blundell, H. Jhoti and C. Abell, *Nat Rev Drug Discov*, 2002, **1**, 45-54.
72. W. A. Hendrickson and C. M. Ogata, *Macromolecular Crystallography, Pt A*, 1997, **276**, 494-523.
73. Z. Dauter, M. Li and A. Wlodawer, *Acta Crystallogr D Biol Crystallogr*, 2001, **57**, 239-249.
74. C. M. Weeks and R. Miller, *Acta Crystallogr D Biol Crystallogr*, 1999, **55**, 492-500.
75. T. C. Terwilliger and J. Berendzen, *Acta Crystallogr D Biol Crystallogr*, 1999, **55**, 849-861.
76. M. Egli, *Curr Protoc Nucleic Acid Chem*, 2010, **Chapter 7**, Unit 7 13.

77. A. Bax, G. Kontaxis and N. Tjandra, *Methods Enzymol*, 2001, **339**, 127-174.
78. N. Tjandra, J. G. Omichinski, A. M. Gronenborn, G. M. Clore and A. Bax, *Nat Struct Biol*, 1997, **4**, 732-738.
79. J. R. Tolman, J. M. Flanagan, M. A. Kennedy and J. H. Prestegard, *Proc Natl Acad Sci U S A*, 1995, **92**, 9279-9283.
80. H. D. Mertens and D. I. Svergun, *J Struct Biol*, 2010, **172**, 128-141.
81. P. Bernado, L. Blanchard, P. Timmins, D. Marion, R. W. Ruigrok and M. Blackledge, *Proc Natl Acad Sci U S A*, 2005, **102**, 17002-17007.
82. Q. Hao, *Acta Crystallogr D Biol Crystallogr*, 2006, **62**, 909-914.
83. X. Hong and Q. Hao, *J Appl Crystallogr*, 2009, **42**, 259-264.
84. A. Grishaev, J. Wu, J. Trewhella and A. Bax, *J Am Chem Soc*, 2005, **127**, 16621-16628.
85. A. Grishaev, V. Tugarinov, L. E. Kay, J. Trewhella and A. Bax, *J Biomol NMR*, 2008, **40**, 95-106.
86. W. Kuhlbrandt, *Science*, 2014, **343**, 1443-1444.
87. W. Wong, X. C. Bai, A. Brown, I. S. Fernandez, E. Hanssen, M. Condron, Y. H. Tan, J. Baum and S. H. Scheres, *Elife*, 2014, **3**.
88. A. Amunts, A. Brown, X. C. Bai, J. L. Llacer, T. Hussain, P. Emsley, F. Long, G. Murshudov, S. H. Scheres and V. Ramakrishnan, *Science*, 2014, **343**, 1485-1489.
89. R. M. Voorhees, I. S. Fernandez, S. H. Scheres and R. S. Hegde, *Cell*, 2014, **157**, 1632-1643.
90. M. Liao, E. Cao, D. Julius and Y. Cheng, *Nature*, 2013, **504**, 107-112.
91. E. Cao, M. Liao, Y. Cheng and D. Julius, *Nature*, 2013, **504**, 113-118.
92. M. Allegretti, D. J. Mills, G. McMullan, W. Kuhlbrandt and J. Vonck, *Elife*, 2014, **3**, e01963.
93. W. Kuhlbrandt, *Elife*, 2014, **3**, e03678.
94. P. Lu, X. C. Bai, D. Ma, T. Xie, C. Yan, L. Sun, G. Yang, Y. Zhao, R. Zhou, S. H. Scheres and Y. Shi, *Nature*, 2014, **512**, 166-170.
95. I. Schlichting and J. Miao, *Curr Opin Struct Biol*, 2012, **22**, 613-626.
96. V. Azzarito, K. Long, N. S. Murphy and A. J. Wilson, *Nat Chem*, 2013, **5**, 161-173.
97. T. A. Edwards and A. J. Wilson, *Amino Acids*, 2011, **41**, 743-754.
98. J. P. Plante, T. Burnley, B. Malkova, M. E. Webb, S. L. Warriner, T. A. Edwards and A. J. Wilson, *Chem Commun (Camb)*, 2009, 5091-5093.
99. L. K. Henchey, A. L. Jochim and P. S. Arora, *Curr Opin Chem Biol*, 2008, **12**, 692-697.
100. J. M. Davis, L. K. Tsou and A. D. Hamilton, *Chem Soc Rev*, 2007, **36**, 326-334.
101. C. G. Cummings and A. D. Hamilton, *Curr Opin Chem Biol*, 2010, **14**, 341-346.
102. B. Leader, Q. J. Baca and D. E. Golan, *Nat Rev Drug Discov*, 2008, **7**, 21-39.
103. S. A. Marshall, G. A. Lazar, A. J. Chirino and J. R. Desjarlais, *Drug Discov Today*, 2003, **8**, 212-221.
104. J. M. Scholtz and R. L. Baldwin, *Annu Rev Biophys Biomol Struct*, 1992, **21**, 95-118.
105. M. J. I. Andrews and A. B. Tabor, *Tetrahedron*, 1999, **55**, 11711-11743.
106. J. Garner and M. M. Harding, *Org Biomol Chem*, 2007, **5**, 3577-3585.
107. A. S. Ripka and D. H. Rich, *Curr Opin Chem Biol*, 1998, **2**, 441-452.

108. B. P. Orner, J. T. Ernst and A. D. Hamilton, *J Am Chem Soc*, 2001, **123**, 5382-5383.
109. J. T. Ernst, O. Kutzki, A. K. Debnath, S. Jiang, H. Lu and A. D. Hamilton, *Angew Chem Int Ed Engl*, 2002, **41**, 278-281.
110. O. Kutzki, H. S. Park, J. T. Ernst, B. P. Orner, H. Yin and A. D. Hamilton, *J Am Chem Soc*, 2002, **124**, 11838-11839.
111. H. Yin, G. I. Lee, K. A. Sedey, O. Kutzki, H. S. Park, B. P. Orner, J. T. Ernst, H. G. Wang, S. M. Sebti and A. D. Hamilton, *J Am Chem Soc*, 2005, **127**, 10191-10196.
112. A. Kazi, J. Sun, K. Doi, S. S. Sung, Y. Takahashi, H. Yin, J. M. Rodriguez, J. Becerril, N. Berndt, A. D. Hamilton, H. G. Wang and S. M. Sebti, *J Biol Chem*, 2011, **286**, 9382-9392.
113. H. Yin, G. I. Lee, K. A. Sedey, J. M. Rodriguez, H. G. Wang, S. M. Sebti and A. D. Hamilton, *J Am Chem Soc*, 2005, **127**, 5463-5468.
114. J. M. Rodriguez, L. Nevola, N. T. Ross, G. I. Lee and A. D. Hamilton, *Chembiochem*, 2009, **10**, 829-833.
115. S. M. Biros, L. Moisan, E. Mann, A. Carella, D. Zhai, J. C. Reed and J. Rebek, Jr., *Bioorg Med Chem Lett*, 2007, **17**, 4641-4645.
116. I. Saraogi, C. D. Incarvito and A. D. Hamilton, *Angew Chem Int Ed Engl*, 2008, **47**, 9691-9694.
117. K. Long, T. A. Edwards and A. J. Wilson, *Bioorganic & Medicinal Chemistry*, 2013, **21**, 4034-4040.
118. S. S. Sidhu, H. B. Lowman, B. C. Cunningham and J. A. Wells, *Methods Enzymol*, 2000, **328**, 333-363.
119. S. S. Sidhu, W. J. Fairbrother and K. Deshayes, *Chembiochem*, 2003, **4**, 14-25.
120. G. P. Smith, *Science*, 1985, **228**, 1315-1317.
121. J. M. Reichert, *MAbs*, 2010, **2**, 84-100.
122. C. Tiede, A. A. Tang, S. E. Deacon, U. Mandal, J. E. Nettleship, R. L. Owen, S. E. George, D. J. Harrison, R. J. Owens, D. C. Tomlinson and M. J. McPherson, *Protein Eng Des Sel*, 2014, **27**, 145-155.
123. J. Wojcik, O. Hantschel, F. Grebien, I. Kaupe, K. L. Bennett, J. Barkinge, R. B. Jones, A. Koide, G. Superti-Furga and S. Koide, *Nat Struct Mol Biol*, 2010, **17**, 519-527.
124. J. P. Theurillat, B. Dreier, G. Nagy-Davidescu, B. Seifert, S. Behnke, U. Zurrer-Hardi, F. Ingold, A. Pluckthun and H. Moch, *Mod Pathol*, 2010, **23**, 1289-1297.
125. F. Grebien, O. Hantschel, J. Wojcik, I. Kaupe, B. Kovacic, A. M. Wyrzucki, G. D. Gish, S. Cerny-Reiterer, A. Koide, H. Beug, T. Pawson, P. Valent, S. Koide and G. Superti-Furga, *Cell*, 2011, **147**, 306-319.
126. T. Wurch, A. Pierre and S. Depil, *Trends Biotechnol*, 2012, **30**, 575-582.
127. P. J. Carter, *Exp Cell Res*, 2011, **317**, 1261-1269.
128. H. Kondo, K. Abe, Y. Emori and S. Arai, *FEBS Lett*, 1991, **278**, 87-90.
129. R. Perozzo, G. Folkers and L. Scapozza, *J Recept Signal Transduct Res*, 2004, **24**, 1-52.
130. G. A. Holdgate and W. H. Ward, *Drug Discov Today*, 2005, **10**, 1543-1550.
131. S. Nunez, J. Venhorst and C. G. Kruse, *Drug Discov Today*, 2012, **17**, 10-22.
132. S. Duhr and D. Braun, *Proc Natl Acad Sci U S A*, 2006, **103**, 19678-19682.
133. R. D. Mitra, C. M. Silva and D. C. Youvan, *Gene*, 1996, **173**, 13-17.
134. S. M. Kelly and N. C. Price, *Biochim Biophys Acta*, 1997, **1338**, 161-185.

135. J. A. Loo, *Mass Spectrom Rev*, 1997, **16**, 1-23.
136. S. Zhang, C. K. Van Pelt and D. B. Wilson, *Anal Chem*, 2003, **75**, 3010-3018.
137. Y. Berezovskaya, C. T. Armstrong, A. L. Boyle, M. Porrini, D. N. Woolfson and P. E. Barran, *Chem Commun (Camb)*, 2011, **47**, 412-414.
138. C. Moller, R. R. Sprenger, S. Sturup and P. Hojrup, *Anal Bioanal Chem*, 2011, **401**, 1619-1629.
139. D. A. Annis, N. Nazef, C. C. Chuang, M. P. Scott and H. M. Nash, *J Am Chem Soc*, 2004, **126**, 15495-15503.
140. M. J. Keith-Roach, *Anal Chim Acta*, 2010, **678**, 140-148.
141. F. Sobott, J. L. Benesch, E. Vierling and C. V. Robinson, *J Biol Chem*, 2002, **277**, 38921-38929.
142. R. Siegel, D. Naishadham and A. Jemal, *CA Cancer J Clin*, 2012, **62**, 10-29.
143. I. K. Nordgren and A. Tavassoli, *Chem Soc Rev*, 2011, **40**, 4307-4317.
144. D. Hanahan and R. A. Weinberg, *Cell*, 2000, **100**, 57-70.
145. D. Peer, J. M. Karp, S. Hong, O. C. Farokhzad, R. Margalit and R. Langer, *Nat Nanotechnol*, 2007, **2**, 751-760.
146. Y. Singh, M. Palombo and P. J. Sinko, *Curr Med Chem*, 2008, **15**, 1802-1826.
147. J. Rautio, H. Kumpulainen, T. Heimbach, R. Oliyai, D. Oh, T. Jarvinen and J. Savolainen, *Nat Rev Drug Discov*, 2008, **7**, 255-270.
148. A. Albert, *Nature*, 1958, **182**, 421-422.
149. P. Carter, *Nat Rev Cancer*, 2001, **1**, 118-129.
150. D. Schrama, R. A. Reisfeld and J. C. Becker, *Nat Rev Drug Discov*, 2006, **5**, 147-159.
151. E. M. De Francesco, R. Lappano, M. F. Santolla, S. Marsico, A. Caruso and M. Maggiolini, *Breast Cancer Res*, 2013, **15**, R64.
152. Y. Mamane, E. Petroulakis, L. Rong, K. Yoshida, L. W. Ler and N. Sonenberg, *Oncogene*, 2004, **23**, 3172-3179.
153. G. L. Semenza, *Nat Rev Cancer*, 2003, **3**, 721-732.
154. G. L. Semenza, *N Engl J Med*, 2011, **365**, 537-547.
155. H. Zhong, A. M. De Marzo, E. Laughner, M. Lim, D. A. Hilton, D. Zagzag, P. Buechler, W. B. Isaacs, G. L. Semenza and J. W. Simons, *Cancer Res*, 1999, **59**, 5830-5835.
156. M. A. Selak, S. M. Armour, E. D. MacKenzie, H. Boulahbel, D. G. Watson, K. D. Mansfield, Y. Pan, M. C. Simon, C. B. Thompson and E. Gottlieb, *Cancer Cell*, 2005, **7**, 77-85.
157. G. L. Semenza, *Crit Rev Biochem Mol Biol*, 2000, **35**, 71-103.
158. G. L. Semenza, *Oncogene*, 2010, **29**, 625-634.
159. D. J. Manalo, A. Rowan, T. Lavoie, L. Natarajan, B. D. Kelly, S. Q. Ye, J. G. Garcia and G. L. Semenza, *Blood*, 2005, **105**, 659-669.
160. G. Powis and L. Kirkpatrick, *Mol Cancer Ther*, 2004, **3**, 647-654.
161. D. Liao and R. S. Johnson, *Cancer Metastasis Rev*, 2007, **26**, 281-290.
162. Y. Wang, Y. Liu, S. N. Malek, P. Zheng and Y. Liu, *Cell Stem Cell*, 2011, **8**, 399-411.
163. H. Zhang, P. Gao, R. Fukuda, G. Kumar, B. Krishnamachary, K. I. Zeller, C. V. Dang and G. L. Semenza, *Cancer Cell*, 2007, **11**, 407-420.
164. A. Franovic, L. Gunaratnam, K. Smith, I. Robert, D. Patten and S. Lee, *Proc Natl Acad Sci U S A*, 2007, **104**, 13092-13097.

165. J. T. Erler, K. L. Bennewith, M. Nicolau, N. Dornhofer, C. Kong, Q. T. Le, J. T. Chi, S. S. Jeffrey and A. J. Giaccia, *Nature*, 2006, **440**, 1222-1226.
166. G. L. Semenza, *Trends Pharmacol Sci*, 2012, **33**, 207-214.
167. Y. Umenaga, K. S. Paku, Y. In, T. Ishida and K. Tomoo, *Biochem Biophys Res Commun*, 2011, **414**, 462-467.
168. C. J. Brown, J. J. Lim, T. Leonard, H. C. Lim, C. S. Chia, C. S. Verma and D. P. Lane, *J Mol Biol*, 2011, **405**, 736-753.
169. S. Mader, H. Lee, A. Pause and N. Sonenberg, *Mol Cell Biol*, 1995, **15**, 4990-4997.
170. S. G. Grant, J. Jessee, F. R. Bloom and D. Hanahan, *Proc Natl Acad Sci U S A*, 1990, **87**, 4645-4649.
171. D. J. Williamson, M. A. Fascione, M. E. Webb and W. B. Turnbull, *Angew Chem Int Ed Engl*, 2012, **51**, 9377-9380.
172. U. K. Laemmli, *Nature*, 1970, **227**, 680-685.
173. W. L. Matochko, K. Chu, B. Jin, S. W. Lee, G. M. Whitesides and R. Derda, *Methods*, 2012, **58**, 47-55.
174. C. J. Brown, C. S. Verma, M. D. Walkinshaw and D. P. Lane, *Cell Cycle*, 2009, **8**, 1905-1911.
175. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J Appl Crystallogr*, 2007, **40**, 658-674.
176. P. D. Adams, R. W. Grosse-Kunstleve, L. W. Hung, T. R. Ioerger, A. J. McCoy, N. W. Moriarty, R. J. Read, J. C. Sacchettini, N. K. Sauter and T. C. Terwilliger, *Acta Crystallogr D Biol Crystallogr*, 2002, **58**, 1948-1954.
177. G. N. Murshudov, A. A. Vagin and E. J. Dodson, *Acta Crystallogr D Biol Crystallogr*, 1997, **53**, 240-255.
178. F. Long, A. A. Vagin, P. Young and G. N. Murshudov, *Acta Crystallogr D Biol Crystallogr*, 2008, **64**, 125-132.
179. K. Cowtan, *Acta Crystallogr D Biol Crystallogr*, 2006, **62**, 1002-1011.
180. R. N. De Guzman, J. M. Wojciak, M. A. Martinez-Yamout, H. J. Dyson and P. E. Wright, *Biochemistry*, 2005, **44**, 490-497.
181. C. Dominguez, R. Boelens and A. M. Bonvin, *J Am Chem Soc*, 2003, **125**, 1731-1737.
182. S. A. Dames, M. Martinez-Yamout, R. N. De Guzman, H. J. Dyson and P. E. Wright, *Proc Natl Acad Sci U S A*, 2002, **99**, 5271-5276.
183. P. Vaupel, O. Thews, D. K. Kelleher and M. Hoeckel, *Adv Exp Med Biol*, 1998, **454**, 591-602.
184. A. L. Harris, *Nat Rev Cancer*, 2002, **2**, 38-47.
185. L. E. Huang and H. F. Bunn, *J Biol Chem*, 2003, **278**, 19575-19578.
186. C. W. Pugh and P. J. Ratcliffe, *Nat Med*, 2003, **9**, 677-684.
187. G. Semenza, *Biochem Pharmacol*, 2002, **64**, 993-998.
188. M. Hockel and P. Vaupel, *Semin Oncol*, 2001, **28**, 36-41.
189. A. L. Kung, S. D. Zabludoff, D. S. France, S. J. Freedman, E. A. Tanner, A. Vieira, S. Cornell-Kennon, J. Lee, B. Wang, J. Wang, K. Memmert, H. U. Naegeli, F. Petersen, M. J. Eck, K. W. Bair, A. W. Wood and D. M. Livingston, *Cancer Cell*, 2004, **6**, 33-43.
190. B. H. Jiang, E. Rue, G. L. Wang, R. Roe and G. L. Semenza, *J Biol Chem*, 1996, **271**, 17771-17778.
191. H. Li, H. P. Ko and J. P. Whitlock, *J Biol Chem*, 1996, **271**, 21262-21267.

192. R. Ravi, B. Mookerjee, Z. M. Bhujwalla, C. H. Sutter, D. Artemov, Q. Zeng, L. E. Dillehay, A. Madan, G. L. Semenza and A. Bedi, *Genes Dev*, 2000, **14**, 34-44.
193. K. Gradin, J. McGuire, R. H. Wenger, I. Kvietikova, M. L. fhitelaw, R. Toftgard, L. Tora, M. Gassmann and L. Poellinger, *Mol Cell Biol*, 1996, **16**, 5221-5231.
194. P. C. Mahon, K. Hirota and G. L. Semenza, *Genes Dev*, 2001, **15**, 2675-2686.
195. Y. Xia, H. K. Choi and K. Lee, *Eur J Med Chem*, 2012, **49**, 24-40.
196. R. H. Goodman and S. Smolik, *Genes Dev*, 2000, **14**, 1553-1577.
197. S. J. Freedman, Z. Y. Sun, F. Poy, A. L. Kung, D. M. Livingston, G. Wagner and M. J. Eck, *Proc Natl Acad Sci U S A*, 2002, **99**, 5367-5372.
198. N. Vo and R. H. Goodman, *J Biol Chem*, 2001, **276**, 13505-13508.
199. B. Onnis, A. Rapisarda and G. Melillo, *J Cell Mol Med*, 2009, **13**, 2780-2786.
200. R. M. Young, S. J. Wang, J. D. Gordan, X. Ji, S. A. Liebhaber and M. C. Simon, *J Biol Chem*, 2008, **283**, 16309-16319.
201. L. M. Greenberger, I. D. Horak, D. Filpula, P. Sapra, M. Westergaard, H. F. Frydenlund, C. Albaek, H. Schroder and H. Orum, *Mol Cancer Ther*, 2008, **7**, 3598-3608.
202. A. Rapisarda, M. Hollingshead, B. Uranchimeg, C. A. Bonomi, S. D. Borgel, J. P. Carter, B. Gehrs, M. Raffeld, R. J. Kinders, R. Parchment, M. R. Anver, R. H. Shoemaker and G. Melillo, *Mol Cancer Ther*, 2009, **8**, 1867-1877.
203. P. Sapra, H. Zhao, M. Mehlig, J. Malaby, P. Kraft, C. Longley, L. M. Greenberger and I. D. Horak, *Clin Cancer Res*, 2008, **14**, 1888-1896.
204. L. Neckers, *J Biosci*, 2007, **32**, 517-530.
205. J. S. Isaacs, Y. J. Jung, E. G. Mimnaugh, A. Martinez, F. Cuttitta and L. M. Neckers, *J Biol Chem*, 2002, **277**, 29936-29944.
206. E. Hur, H. H. Kim, S. M. Choi, J. H. Kim, S. Yim, H. J. Kwon, Y. Choi, D. K. Kim, M. O. Lee and H. Park, *Mol Pharmacol*, 2002, **62**, 975-982.
207. D. Kong, E. J. Park, A. G. Stephen, M. Calvani, J. H. Cardellina, A. Monks, R. J. Fisher, R. H. Shoemaker and G. Melillo, *Cancer Res*, 2005, **65**, 9047-9055.
208. N. G. Nickols, C. S. Jacobs, M. E. Farkas and P. B. Dervan, *ACS Chem Biol*, 2007, **2**, 561-571.
209. E. Miranda, I. K. Nordgren, A. L. Male, C. E. Lawrence, F. Hoakwie, F. Cuda, W. Court, K. R. Fox, P. A. Townsend, G. K. Packham, S. A. Eccles and A. Tavassoli, *J Am Chem Soc*, 2013, **135**, 10418-10425.
210. A. R. Horswill, S. N. Savinov and S. J. Benkovic, *Proc Natl Acad Sci U S A*, 2004, **101**, 15591-15596.
211. A. Tavassoli and S. J. Benkovic, *Angew Chem Int Ed Engl*, 2005, **44**, 2760-2763.
212. S. Salceda and J. Caro, *J Biol Chem*, 1997, **272**, 22642-22647.
213. S. Kaluz, M. Kaluzova and E. J. Stanbridge, *Mol Cell Biol*, 2006, **26**, 5895-5907.
214. D. H. Shin, Y. S. Chun, D. S. Lee, L. E. Huang and J. W. Park, *Blood*, 2008, **111**, 3131-3136.
215. K. M. Cook, S. T. Hilton, J. Mecinovic, W. B. Motherwell, W. D. Figg and C. J. Schofield, *J Biol Chem*, 2009, **284**, 26831-26838.
216. J. L. Ruas, L. Poellinger and T. Pereira, *J Biol Chem*, 2002, **277**, 38723-38730.

217. L. K. Henchey, S. Kushal, R. Dubey, R. N. Chapman, B. Z. Olenyuk and P. S. Arora, *J Am Chem Soc*, 2010, **132**, 941-943.
218. J. Gu, J. Milligan and L. E. Huang, *J Biol Chem*, 2001, **276**, 3550-3554.
219. D. Lando, D. J. Peet, D. A. Whelan, J. J. Gorman and M. L. Whitelaw, *Science*, 2002, **295**, 858-861.
220. H. Cho, D. R. Ahn, H. Park and E. G. Yang, *FEBS Lett*, 2007, **581**, 1542-1548.
221. S. Kushal, B. B. Lao, L. K. Henchey, R. Dubey, H. Mesallati, N. J. Traaseth, B. Z. Olenyuk and P. S. Arora, *Proc Natl Acad Sci U S A*, 2013, **110**, 15602-15607.
222. T. Shibata, A. J. Giaccia and J. M. Brown, *Gene Ther*, 2000, **7**, 493-498.
223. J. D. Tyndall, T. Nall and D. P. Fairlie, *Chem Rev*, 2005, **105**, 973-999.
224. R. E. Moellering, M. Cornejo, T. N. Davis, C. Del Bianco, J. C. Aster, S. C. Blacklow, A. L. Kung, D. G. Gilliland, G. L. Verdine and J. E. Bradner, *Nature*, 2009, **462**, 182-188.
225. B. B. Lao, I. Grishagin, H. Mesallati, T. F. Brewer, B. Z. Olenyuk and P. S. Arora, *Proc Natl Acad Sci U S A*, 2014, **111**, 7531-7536.
226. P. Tosovska and P. S. Arora, *Org Lett*, 2010, **12**, 1588-1591.
227. Y. Benita, H. Kikuchi, A. D. Smith, M. Q. Zhang, D. C. Chung and R. J. Xavier, *Nucleic Acids Res*, 2009, **37**, 4587-4602.
228. J. G. Marblestone, S. C. Edavettal, Y. Lim, P. Lim, X. Zuo and T. R. Butt, *Protein Sci*, 2006, **15**, 182-189.
229. D. B. Smith and K. S. Johnson, *Gene*, 1988, **67**, 31-40.
230. R. Y. Tsien, *Annu Rev Biochem*, 1998, **67**, 509-544.
231. F. W. Studier and B. A. Moffatt, *J Mol Biol*, 1986, **189**, 113-130.
232. P. Davanloo, A. H. Rosenberg, J. J. Dunn and F. W. Studier, *Proc Natl Acad Sci U S A*, 1984, **81**, 2035-2039.
233. J. M. Chalker and B. G. Davis, *Curr Opin Chem Biol*, 2010, **14**, 781-789.
234. J. Chen, W. Zeng, R. Offord and K. Rose, *Bioconjug Chem*, 2003, **14**, 614-618.
235. P. E. Dawson, T. W. Muir, I. Clark-Lewis and S. B. Kent, *Science*, 1994, **266**, 776-779.
236. F. W. Studier, *Protein Expr Purif*, 2005, **41**, 207-234.
237. S. S. Sidhu, *Curr Opin Biotechnol*, 2000, **11**, 610-616.
238. J. K. Scott and G. P. Smith, *Science*, 1990, **249**, 386-390.
239. S. E. Cwirla, E. A. Peters, R. W. Barrett and W. J. Dower, *Proc Natl Acad Sci U S A*, 1990, **87**, 6378-6382.
240. J. J. Devlin, L. C. Panganiban and P. E. Devlin, *Science*, 1990, **249**, 404-406.
241. T. Hoffmann, L. K. Stadler, M. Busby, Q. Song, A. T. Buxton, S. D. Wagner, J. J. Davis and P. Ko Ferrigno, *Protein Eng Des Sel*, 2010, **23**, 403-413.
242. L. K. Stadler, T. Hoffmann, D. C. Tomlinson, Q. Song, T. Lee, M. Busby, Y. Nyathi, E. Gendra, C. Tiede, K. Flanagan, S. J. Cockell, A. Wipat, C. Harwood, S. D. Wagner, M. A. Knowles, J. J. Davis, N. Keegan and P. K. Ferrigno, *Protein Eng Des Sel*, 2011, **24**, 751-763.
243. Y. Kawasaki and E. Freire, *Drug Discov Today*, 2011, **16**, 985-990.
244. G. M. Burslem, H. F. Kyle, A. L. Breeze, T. A. Edwards, A. Nelson, S. L. Warriner and A. J. Wilson, *Chembiochem*, 2014, **15**, 1083-1087.
245. J. Plante, F. Campbell, B. Malkova, C. Kilner, S. L. Warriner and A. J. Wilson, *Org Biomol Chem*, 2008, **6**, 138-146.

246. A. Barnard, J. A. Miles, G. M. Burslem, A. M. Barker and A. J. Wilson, *Organic & Biomolecular Chemistry*, 2015, **13**, 258-264.
247. M. Vodnik, U. Zager, B. Strukelj and M. Lunder, *Molecules*, 2011, **16**, 790-817.
248. W. D. Thomas, M. Golomb and G. P. Smith, *Anal Biochem*, 2010, **407**, 237-240.
249. L. A. Brammer, B. Bolduc, J. L. Kass, K. M. Felice, C. J. Noren and M. F. Hall, *Anal Biochem*, 2008, **373**, 88-98.
250. L. R. Krumpe, A. J. Atkinson, G. W. Smythers, A. Kandel, K. M. Schumacher, J. B. McMahon, L. Makowski and T. Mori, *Proteomics*, 2006, **6**, 4210-4222.
251. P. A. t Hoen, S. M. Jirka, B. R. Ten Broeke, E. A. Schultes, B. Aguilera, K. H. Pang, H. Heemskerk, A. Aartsma-Rus, G. J. van Ommen and J. T. den Dunnen, *Anal Biochem*, 2012, **421**, 622-631.
252. L. Liu, Y. Li, S. Li, N. Hu, Y. He, R. Pong, D. Lin, L. Lu and M. Law, *J Biomed Biotechnol*, 2012, **2012**, 251364.
253. J. P. O'Shea, M. F. Chou, S. A. Quader, J. K. Ryan, G. M. Church and D. Schwartz, *Nat Methods*, 2013, **10**, 1211-1212.
254. P. Molek, B. Strukelj and T. Bratkovic, *Molecules*, 2011, **16**, 857-887.
255. V. Munoz and L. Serrano, *Nat Struct Biol*, 1994, **1**, 399-409.
256. M. Gebauer and A. Skerra, *Curr Opin Chem Biol*, 2009, **13**, 245-255.
257. P. Chames, M. Van Regenmortel, E. Weiss and D. Baty, *Br J Pharmacol*, 2009, **157**, 220-233.
258. S. E. Hufton, N. van Neer, T. van den Beuken, J. Desmet, E. Sablon and H. R. Hoogenboom, *FEBS Lett*, 2000, **475**, 225-231.
259. S. J. McConnell and R. H. Hoess, *J Mol Biol*, 1995, **250**, 460-470.
260. B. Heyd, F. Pecorari, B. Collinet, E. Adjadj, M. Desmadril and P. Minard, *Biochemistry*, 2003, **42**, 5674-5683.
261. K. Nord, E. Gunneriusson, J. Ringdahl, S. Stahl, M. Uhlen and P. A. Nygren, *Nat Biotechnol*, 1997, **15**, 772-777.
262. E. Bianchi, A. Folgori, A. Wallace, M. Nicotra, S. Acali, A. Phalipon, G. Barbato, R. Bazzo, R. Cortese, F. Felici and et al., *J Mol Biol*, 1995, **247**, 154-160.
263. P. A. Dalby, R. H. Hoess and W. F. DeGrado, *Protein Sci*, 2000, **9**, 2366-2376.
264. C. Souriau, L. Chiche, R. Irving and P. Hudson, *Biochemistry*, 2005, **44**, 7143-7155.
265. K. Skrlec, B. Strukelj and A. Berlec, *Trends Biotechnol*, 2015, **33**, 408-418.
266. J. Nilvebrant, M. Astrand, J. Lofblom and S. Hober, *Cell Mol Life Sci*, 2013, **70**, 3973-3985.
267. V. Ramamurthy, S. R. Krystek, Jr., A. Bush, A. Wei, S. L. Emanuel, R. Das Gupta, A. Janjua, L. Cheng, M. Murdock, B. Abramczyk, D. Cohen, Z. Lin, P. Morin, J. H. Davis, M. Dabritz, D. C. McLaughlin, K. A. Russo, G. Chao, M. C. Wright, V. A. Jenny, L. J. Engle, E. Furfine and S. Sheriff, *Structure*, 2012, **20**, 259-269.
268. G. Varadamsetty, D. Tremmel, S. Hansen, F. Parmeggiani and A. Pluckthun, *J Mol Biol*, 2012, **424**, 68-87.
269. J. Lofblom, J. Feldwisch, V. Tolmachev, J. Carlsson, S. Stahl and F. Y. Frejd, *FEBS Lett*, 2010, **584**, 2670-2680.

270. H. Ebersbach, E. Fiedler, T. Scheuermann, M. Fiedler, M. T. Stubbs, C. Reimann, G. Proetzel, R. Rudolph and U. Fiedler, *J Mol Biol*, 2007, **372**, 172-185.
271. A. Hoffmann, M. Kovermann, H. Lilie, M. Fiedler, J. Balbach, R. Rudolph and S. Pfeifer, *PLoS One*, 2012, **7**, e31298.
272. M. Krehenbrink, M. Chami, I. Guilvout, P. M. Alzari, F. Pecorari and A. P. Pugsley, *J Mol Biol*, 2008, **383**, 1058-1068.
273. J. Desmet, K. Verstraete, Y. Bloch, E. Lorent, Y. Wen, B. Devreese, K. Vandenbroucke, S. Loverix, T. Hettmann, S. Deroo, K. Somers, P. Henderikx, I. Lasters and S. N. Savvides, *Nat Commun*, 2014, **5**, 5237.
274. M. Gebauer, A. Schiefner, G. Matschiner and A. Skerra, *J Mol Biol*, 2013, **425**, 780-802.
275. P. Byla, M. H. Andersen, T. L. Holtet, H. Jacobsen, M. Munch, H. H. Gad, H. C. Thogersen and R. Hartmann, *J Biol Chem*, 2010, **285**, 12096-12100.
276. J. Silverman, Q. Liu, A. Bakker, W. To, A. Duguay, B. M. Alba, R. Smith, A. Rivas, P. Li, H. Le, E. Whitehorn, K. W. Moore, C. Swimmer, V. Perlroth, M. Vogt, J. Kolkman and W. P. Stemmer, *Nat Biotechnol*, 2005, **23**, 1556-1561.
277. M. D. Diem, L. Hyun, F. Yi, R. Hippensteel, E. Kuhar, C. Lowenstein, E. J. Swift, K. T. O'Neil and S. A. Jacobs, *Protein Eng Des Sel*, 2014, **27**, 419-429.
278. A. Schweizer, H. Roschitzki-Voser, P. Amstutz, C. Briand, M. Gulotti-Georgieva, E. Prenosil, H. K. Binz, G. Capitani, A. Baici, A. Pluckthun and M. G. Grutter, *Structure*, 2007, **15**, 625-636.
279. H. K. Binz, P. Amstutz, A. Kohl, M. T. Stumpp, C. Briand, P. Forrer, M. G. Grutter and A. Pluckthun, *Nat Biotechnol*, 2004, **22**, 575-582.
280. M. A. Seeger, R. Zbinden, A. Flutsch, P. G. Gutte, S. Engeler, H. Roschitzki-Voser and M. G. Grutter, *Protein Sci*, 2013, **22**, 1239-1257.
281. M. Silacci, N. Baenziger-Tobler, W. Lembke, W. Zha, S. Batey, J. Bertschinger and D. Grabulovski, *J Biol Chem*, 2014, **289**, 14392-14398.
282. M. S. Dennis, A. Herzka and R. A. Lazarus, *J Biol Chem*, 1995, **270**, 25411-25417.
283. J. D. Steemson, M. Baake, J. Rakonjac, V. L. Arcus and M. T. Liddament, *PLoS One*, 2014, **9**, e86050.
284. S. C. Lee, K. Park, J. Han, J. J. Lee, H. J. Kim, S. Hong, W. Heu, Y. J. Kim, J. S. Ha, S. G. Lee, H. K. Cheong, Y. H. Jeon, D. Kim and H. S. Kim, *Proc Natl Acad Sci U S A*, 2012, **109**, 3299-3304.
285. B. Rupp, *Biomolecular Crystallography*, Taylor and Francis Group, 2010.
286. E. H. Snell, A. Cassetta, J. R. Helliwell, T. J. Boggon, N. E. Chayen, E. Weckert, K. Holzer, K. Schroer, E. J. Gordon and P. F. Zagalsky, *Acta Crystallogr D Biol Crystallogr*, 1997, **53**, 231-239.
287. N. E. Chayen, P. D. S. Stewart, D. L. Maeder and D. M. Blow, *Journal of Applied Crystallography*, 1990, **23**, 297-302.
288. N. E. Chayen, *Journal of Applied Crystallography*, 1997, **30**, 198-202.
289. N. E. Chayen, *Journal of Crystal Growth*, 1999, **196**, 434-441.
290. N. E. Chayen, *Structure*, 1997, **5**, 1269-1274.
291. N. E. Chayen, *Acta Crystallogr D Biol Crystallogr*, 1998, **54**, 8-15.
292. M. V. Petoukhov, P. V. Konarev, A. G. Kikhney and D. I. Svergun, *Journal of Applied Crystallography*, 2007, **40**, S223-S228.

293. D. Svergun, C. Barberato and M. H. J. Koch, *Journal of Applied Crystallography*, 1995, **28**, 768-773.
294. D. Franke and D. I. Svergun, *Journal of Applied Crystallography*, 2009, **42**, 342-346.
295. A. Dong, X. Xu, A. M. Edwards, G. Midwest Center for Structural, C. Structural Genomics, C. Chang, M. Chruszcz, M. Cuff, M. Cymborowski, R. Di Leo, O. Egorova, E. Evdokimova, E. Filippova, J. Gu, J. Guthrie, A. Ignatchenko, A. Joachimiak, N. Klostermann, Y. Kim, Y. Korniyenko, W. Minor, Q. Que, A. Savchenko, T. Skarina, K. Tan, A. Yakunin, A. Yee, V. Yim, R. Zhang, H. Zheng, M. Akutsu, C. Arrowsmith, G. V. Avvakumov, A. Bochkarev, L. G. Dahlgren, S. Dhe-Paganon, S. Dimov, L. Dombrovski, P. Finerty, Jr., S. Flodin, A. Flores, S. Graslund, M. Hammerstrom, M. D. Herman, B. S. Hong, R. Hui, I. Johansson, Y. Liu, M. Nilsson, L. Nedyalkova, P. Nordlund, T. Nyman, J. Min, H. Ouyang, H. W. Park, C. Qi, W. Rabeh, L. Shen, Y. Shen, D. Sukumard, W. Tempel, Y. Tong, L. Tresagues, M. Vedadi, J. R. Walker, J. Weigelt, M. Welin, H. Wu, T. Xiao, H. Zeng and H. Zhu, *Nat Methods*, 2007, **4**, 1019-1021.
296. A. Wernimont and A. Edwards, *PLoS One*, 2009, **4**, e5094.
297. P. Prabhakaran, A. Barnard, N. S. Murphy, C. A. Kilner, T. A. Edwards and A. J. Wilson, *European Journal of Organic Chemistry*, 2013, 3504-3512.
298. F. Campbell, J. P. Plante, T. A. Edwards, S. L. Warriner and A. J. Wilson, *Org Biomol Chem*, 2010, **8**, 2344-2351.
299. A. Barnard, K. Long, H. L. Martin, J. A. Miles, T. A. Edwards, D. C. Tomlinson, A. Macdonald and A. J. Wilson, *Angew Chem Int Ed Engl*, 2015, **54**, 2960-2965.
300. V. Azzarito, J. A. Miles, J. Fisher, T. A. Edwwards, S. L. Warriner and A. J. Wilson, *Chemical Science*, 2015.
301. M. Congreve, G. Chessari, D. Tisi and A. J. Woodhead, *J Med Chem*, 2008, **51**, 3661-3680.
302. C. W. Murray and T. L. Blundell, *Curr Opin Struct Biol*, 2010, **20**, 497-507.
303. M. Carroll and K. L. Borden, *J Interferon Cytokine Res*, 2013, **33**, 227-238.
304. Y. Jia, V. Polunovsky, P. B. Bitterman and C. R. Wagner, *Med Res Rev*, 2012, **32**, 786-814.
305. L. Rong, M. Livingstone, R. Sukarieh, E. Petroulakis, A. C. Gingras, K. Crosby, B. Smith, R. D. Polakiewicz, J. Pelletier, M. A. Ferraiuolo and N. Sonenberg, *RNA*, 2008, **14**, 1318-1327.
306. A. C. Gingras, B. Raught and N. Sonenberg, *Annu Rev Biochem*, 1999, **68**, 913-963.
307. O. Larsson, S. Li, O. A. Issaenko, S. Avdulov, M. Peterson, K. Smith, P. B. Bitterman and V. A. Polunovsky, *Cancer Res*, 2007, **67**, 6814-6824.
308. A. De Benedetti and A. L. Harris, *Int J Biochem Cell Biol*, 1999, **31**, 59-72.
309. S. G. Zimmer, A. DeBenedetti and J. R. Graff, *Anticancer Res*, 2000, **20**, 1343-1351.
310. J. R. Graff and S. G. Zimmer, *Clin Exp Metastasis*, 2003, **20**, 265-273.
311. A. Bah, R. M. Vernon, Z. Siddiqui, M. Krzeminski, R. Muhandiram, C. Zhao, N. Sonenberg, L. E. Kay and J. D. Forman-Kay, *Nature*, 2015, **519**, 106-109.
312. J. Marcotrigiano, A. C. Gingras, N. Sonenberg and S. K. Burley, *Mol Cell*, 1999, **3**, 707-716.
313. S. N. Sehgal, H. Baker and C. Vezina, *J Antibiot (Tokyo)*, 1975, **28**, 727-732.

314. Y. Mamane, E. Petroulakis, O. LeBacquer and N. Sonenberg, *Oncogene*, 2006, **25**, 6416-6422.
315. B. Seto, *Clin Transl Med*, 2012, **1**, 29.
316. S. Wullschleger, R. Loewith and M. N. Hall, *Cell*, 2006, **124**, 471-484.
317. M. Aoki, E. Blazek and P. K. Vogt, *Proc Natl Acad Sci U S A*, 2001, **98**, 136-141.
318. W. Filipowicz, Y. Furuichi, J. M. Sierra, S. Muthukrishnan, A. J. Shatkin and S. Ochoa, *Proc Natl Acad Sci U S A*, 1976, **73**, 1559-1563.
319. R. C. Tam, J. Y. Lau and Z. Hong, *Antivir Chem Chemother*, 2001, **12**, 261-272.
320. A. Kentsis, I. Topisirovic, B. Culjkovic, L. Shao and K. L. Borden, *Proc Natl Acad Sci U S A*, 2004, **101**, 18105-18110.
321. S. Crotty, D. Maag, J. J. Arnold, W. Zhong, J. Y. Lau, Z. Hong, R. Andino and C. E. Cameron, *Nat Med*, 2000, **6**, 1375-1379.
322. S. Crotty, C. E. Cameron and R. Andino, *Proc Natl Acad Sci U S A*, 2001, **98**, 6895-6900.
323. X. Chen, D. J. Kopecky, J. Mihalic, S. Jeffries, X. Min, J. Heath, J. Deignan, S. Lai, Z. Fu, C. Guimaraes, S. Shen, S. Li, S. Johnstone, S. Thibault, H. Xu, M. Cardozo, W. Shen, N. Walker, F. Kayser and Z. Wang, *J Med Chem*, 2012, **55**, 3837-3851.
324. P. Mahalingam, K. Takrouri, T. Chen, R. Sahoo, E. Papadopoulos, L. Chen, G. Wagner, B. H. Aktas, J. A. Halperin and M. Chorev, *J Med Chem*, 2014, **57**, 5094-5111.
325. L. Chen, B. H. Aktas, Y. Wang, X. He, R. Sahoo, N. Zhang, S. Denoyelle, E. Kabha, H. Yang, R. Y. Freedman, J. G. Supko, M. Chorev, G. Wagner and J. A. Halperin, *Oncotarget*, 2012, **3**, 869-881.
326. N. J. Moerke, H. Aktas, H. Chen, S. Cantel, M. Y. Reibarkh, A. Fahmy, J. D. Gross, A. Degterev, J. Yuan, M. Chorev, J. A. Halperin and G. Wagner, *Cell*, 2007, **128**, 257-267.
327. E. Papadopoulos, S. Jenni, E. Kabha, K. J. Takrouri, T. Yi, N. Salvi, R. E. Luna, E. Gavathiotis, P. Mahalingam, H. Arthanari, R. Rodriguez-Mias, R. Yefidoff-Freedman, B. H. Aktas, M. Chorev, J. A. Halperin and G. Wagner, *Proc Natl Acad Sci U S A*, 2014, **111**, E3187-3195.
328. R. Cencic, D. R. Hall, F. Robert, Y. Du, J. Min, L. Li, M. Qui, I. Lewis, S. Kurtkaya, R. Dingledine, H. Fu, D. Kozakov, S. Vajda and J. Pelletier, *Proc Natl Acad Sci U S A*, 2011, **108**, 1046-1051.
329. Y. W. Kim, T. N. Grossmann and G. L. Verdine, *Nat Protoc*, 2011, **6**, 761-771.
330. G. L. Verdine and G. J. Hilinski, *Methods Enzymol*, 2012, **503**, 3-33.
331. W. Zhou, S. T. Quah, C. S. Verma, Y. Liu, D. P. Lane and C. J. Brown, *PLoS One*, 2012, **7**, e47235.
332. D. Lama, S. T. Quah, C. S. Verma, R. Lakshminarayanan, R. W. Beuerman, D. P. Lane and C. J. Brown, *Sci Rep*, 2013, **3**, 3451.
333. A. Haghighat and N. Sonenberg, *J Biol Chem*, 1997, **272**, 21677-21680.
334. J. D. Gross, N. J. Moerke, T. von der Haar, A. A. Lugovskoy, A. B. Sachs, J. E. McCarthy and G. Wagner, *Cell*, 2003, **115**, 739-750.
335. T. von der Haar, Y. Oku, M. Ptushkina, N. Moerke, G. Wagner, J. D. Gross and J. E. McCarthy, *J Mol Biol*, 2006, **356**, 982-992.
336. N. Siddiqui, W. Tempel, L. Nedyalkova, L. Volpon, A. K. Wernimont, M. J. Osborne, H. W. Park and K. L. Borden, *J Mol Biol*, 2012, **415**, 781-792.

337. D. Peter, C. Igreja, R. Weber, L. Wohlbold, C. Weiler, L. Ebertsch, O. Weichenrieder and E. Izaurralde, *Mol Cell*, 2015, **57**, 1074-1087.
338. W. Liu, R. Zhao, C. McFarland, J. Kieft, A. Niedzwiecka, M. Jankowska-Anyszka, J. Stepinski, E. Darzynkiewicz, D. N. Jones and R. E. Davis, *J Biol Chem*, 2009, **284**, 31336-31349.
339. K. Tomoo, Y. Matsushita, H. Fujisaki, F. Abiko, X. Shen, T. Taniguchi, H. Miyagawa, K. Kitamura, K. Miura and T. Ishida, *Biochim Biophys Acta*, 2005, **1753**, 191-208.
340. I. C. Kim and A. D. Hamilton, *Org Lett*, 2006, **8**, 1751-1754.
341. S. Surade and T. L. Blundell, *Chem Biol*, 2012, **19**, 42-50.
342. A. D. Thompson, A. Dugan, J. E. Gestwicki and A. K. Mapp, *ACS Chem Biol*, 2012, **7**, 1311-1320.
343. S. Mandal, M. Moudgil and S. K. Mandal, *Eur J Pharmacol*, 2009, **625**, 90-100.
344. T. Mavromoustakos, S. Durdagi, C. Koukoulitsa, M. Simcic, M. G. Papadopoulos, M. Hodoscek and S. G. Grdadolnik, *Curr Med Chem*, 2011, **18**, 2517-2530.
345. J. Yin, G. Li, X. Ren and G. Herrler, *J Biotechnol*, 2007, **127**, 335-347.
346. G. Georgiou and L. Segatori, *Curr Opin Biotechnol*, 2005, **16**, 538-545.
347. E. Jung and K. L. Williams, *Biotechnol Appl Biochem*, 1997, **25 (Pt 1)**, 3-8.
348. M. H. Linskens, P. D. Grootenhuis, M. Blaauw, B. Huisman-de Winkel, A. Van Ravestein, P. J. Van Haastert and J. C. Heikoop, *FASEB J*, 1999, **13**, 639-645.
349. E. M. Mauriello, H. Duc le, R. Isticato, G. Cangiano, H. A. Hong, M. De Felice, E. Ricca and S. M. Cutting, *Vaccine*, 2004, **22**, 1177-1187.
350. R. Daly and M. T. Hearn, *J Mol Recognit*, 2005, **18**, 119-138.