

A single molecule approach to investigate how AP1  
transcriptional regulators find their target sites on DNA

Nicola Annette Don

A thesis submitted for the degree of Doctor of Philosophy  
in Cell and Molecular Biology

School of Biological Sciences

University of Essex

September 2015

## Abstract

Transcriptional regulator protein family members Activator Protein-1 (AP1) bind to their target site TGAC/GTCA during the normal cell cycle. Their over-expression is linked to the initiation of cancer. Regulating cFos and cJun interactions with AP1 binding sites is a potential cancer therapy strategy. How the proteins find their target sites and whether non-specific DNA binding occurs will be investigated. The Protein Fragment Complementation Assay (PCA) derived inhibitor FosW is also capable of interfering with its target cJun. To study these proteins, DNA tightropes were created where single strands of  $\lambda$ , pUC19, pUCap1 and target-free  $\lambda$  (TF $\lambda$ ) DNA were suspended above the surface of a glass coverslip on 5  $\mu$ m high pedestals. Oblique Angle Fluorescence (OAF) microscopy was used to image Quantum dot (Qdot) conjugated proteins *in vitro*.

The protein combinations cFos:cFos, cJun:cJun, cFos:cJun, FosW and FosW+cJun (Mason et al. 2006, Worrall and Mason 2011) were studied interacting with the different DNA substrates and within the AP1 family.  $71 \pm 3.1\%$  cJun:cJun,  $53 \pm 6.1\%$  cFos:cJun heterodimers diffused 3-Dimensionally and 1-Dimensionally along  $\lambda$  DNA, indicating this is a crucial part of their search mechanism. Surprisingly, cFos is capable of dimerising, a previously unseen observation.  $45 \pm 3.7\%$  of these cFos:cFos homodimers also diffused 3-Dimensionally and 1-Dimensionally. Diffusion decreased when the proteins interacted with pUC19 and pUCap1 and cJun only showed  $3 \pm 1.5\%$  movement on TF $\lambda$ , an unexpected observation.

The interaction between FosW and cJun:cJun indicated clear interference with cJun dimerization.  $55 \pm 11.0\%$  FosW and  $39 \pm 11.0\%$  FosW+cJun diffused 3-Dimensionally and 1-Dimensionally. This was observed to occur directly on DNA and clarifies the mechanism of competitive inhibition and partner exchange in the AP1 family. This insight may significantly impact our understanding on how these proteins regulate transcription and help define new mechanisms of inhibition.

## **Acknowledgments**

First of all, I would like to thank Dr. Neil Kad for allowing me to work on such a fascinating subject in his lab. I would also like to thank him for his expertise and help with the completion of this thesis. In addition, I would like to thank Dr. Jody Mason for allowing me to work with cFos and cJun (Granted I did end up nicknaming the proteins Candy Floss (cFos) and CaJun Chicken (cJun) to make the whole project seem less scary, but nevertheless, I did really enjoy working on this study!).

Next, I would like to thank my loving parents Mark and Heather and my husband Matthew for their support during my PhD journey. It has been a tough ride for me at times, but I couldn't have reached the end without you, for which I am eternally grateful for. I would not have achieved so much if it wasn't for you.

A special thanks to Craig 'aka' "Chug", "Hariman" Harish, Amna and Alessio from my lab, Victoria and Richard from Jody's lab and Sunitha from the plant lab for not only help, but fun times as well. I will have many happy memories working alongside all of you. Last but not least, Tati, the co-member of Team NicTat. We both formed this team back in 2012 when it was 11pm one day in the post-grad room. Since then, we formed a strong friendship and have supported each other through thick and thin. A very dear and special friend. Team NicTat GO GO GO!!! Outside of the lab, I would like to thank my good friends Steph and Ruth who I met through the mature Students Society MSX, Orestis and Sophia, Emma the administrator for lending an ear and also everyone at Rowans Care Ltd.

I feel very privileged to have had the chance to work with cFos and cJun. Back in the day when I was working as a General Staff Nurse on a cancer ward, I encountered many patients who had either walked out of the door with a new lease of life or lost their battle with cancer. After experiencing these happy and tragic stories first hand, I knew that I wanted to contribute towards cancer research, hoping that one day I could help make a difference. Throughout this PhD I was playing detective, just like my favourite Hercule Poirot and Miss Marple (thank you Agatha Christie!) and also the amazing Comoran Strike (thank you J.K Rowling!). Candy Floss and CaJun chicken, your mischievous endeavours will not go unfound forever...

## **Contents**

Abstract	ii
Acknowledgements	iii
List of tables	viii
List of figures	ix
List of abbreviations	xii

## **Chapter 1: Introduction**

1.1 What is cancer?	1
1.2 The role of cFos and cJun in cancer	1
1.3 The structure of cFos and cJun	11
1.4 The antagonistic peptide inhibitor FosW	16
1.5 Biochemical and biophysical methods for probing protein-DNA interactions	19
1.6 Single molecule techniques	20
1.7 Diffusion mechanisms and AP1 binding theories	28
1.8 The “waiting room model”	31
1.9 Aims of the thesis	34

## **Chapter 2: Imaging cFos, cJun, cFos:cJun and FosW interactions using single molecule techniques**

2.1 Introduction	36
2.2 Materials and methods	40
2.2.1 Flow cell construction	40
2.2.2 PEGylation of silica coverslips	42
2.2.3 PEGylation of silica coverslips <i>in situ</i>	42
2.2.4 Silica microspheres coated with poly-L-lysine	43
2.2.5 Buffers	44
2.2.5.1 ABC buffer	44

2.2.6 Bacteriophage Lambda methylated and non-methylated DNA tightropes	44
2.2.6.1 Bacteriophage Lambda DNA	44
2.2.6.2 Lambda DNA with additional variant AP1 target sites	47
2.2.7 Creating tightropes	49
2.2.8 Biotinylated proteins	51
2.2.9 Labelling the biotinylated peptide	55
2.2.10 Quantum dots	55
2.2.11 Imaging using Oblique Angle Fluorescence (OAF) microscopy	56
2.2.12 Data analysis	58
2.2.13 Statistical tests	58
2.2.14 Kymograph analysis	58
2.2.15 Random walker simulations with specific and non-specific pausing	61
2.2.16 Diffusion constant and $\alpha$ value	65
2.2.17 MSD analysis	65
2.2.18 The Binding Site Position Finder	74
2.2.18.1 Binding Site Position Finder for $\lambda$ DNA	74
2.2.18.2 The adapted Binding Site Position Finder incorporating variant AP1 target sites	76
2.2.19 Data for the Binding Site Position Finder	80
2.3 Results	83
2.3.1 Protein interaction with $\lambda$ DNA	83
2.3.2 Protein colocalisation	88
2.3.3 Protein binding positions along $\lambda$ DNA	91
2.4 Discussion	101
2.4.1 cFos, cJun and cFos:cJun heterodimers interacting with $\lambda$ DNA	101

2.4.2 Protein colocalisation	102
2.4.3 Interaction between FosW and cJun	103
2.4.4 Movement along DNA	104
2.4.5 Pausing along DNA	105
2.4.6 Conclusions	106
<b>Chapter 3: cFos, cJun and cFos:cJun interaction with different DNA substrates</b>	
3.1: Introduction	107
3.2 Materials and methods	110
3.2.1 Digestion of pUC19	110
3.2.2 AP1 oligonucleotide design	112
3.2.3 Making pUCap1	115
3.2.4 Sequencing of pUCap1	117
3.2.5 Making chemically competent cells	119
3.2.6 Transformation of chemically competent <i>E. coli</i> cells	119
3.2.7 Preparing a bacterial culture	120
3.2.8 Preparing a 1% (w/v) agarose gel	120
3.2.9 Inactivating T4 DNA Ligase using KCl	121
3.2.10 PEG <sub>6000</sub> concentration for optimal DNA ligation	123
3.2.11 A time course to create pUC19 tightropes that are comparable with the length of $\lambda$ DNA	125
3.2.12 Creating pUC19 concatemers	127
3.2.13 Creating pUC19 tightropes	127
3.2.14 Buffers	128
3.2.14.1 High salt buffer containing ATP	128
3.2.14.2 High salt buffer containing YOYO-1	128

3.2.15 Phenol:Chloroform extraction and DNA precipitation	128
3.2.16 Adaptation of Phenol:Chloroform extraction and DNA precipitation	133
3.2.17 Identifying target free $\lambda$ (TF $\lambda$ ) DNA	133
3.2.18 Double digestion of pUC19 for $\lambda$ pUC construction	136
3.2.19 Constructing $\lambda$ pUC	136
3.2.20 PEG <sub>6000</sub> concentration for optimal TF $\lambda$ DNA ligation	140
3.2.21 A time course to create TF $\lambda$ tightropes that are comparable with the length of $\lambda$ DNA	147
3.2.22 Adapted TF $\lambda$ DNA ligation method	149
3.2.23 The Binding site Finder for pUC19	149
3.2.24 The Binding site Finder for pUCap1	152
3.3 Results	154
3.3.1 Restriction mapping $\lambda$ pUC	154
3.3.2 Protein interaction with pUC19 DNA	156
3.3.3 Protein colocalisation	162
3.3.4 Protein binding positions along Phenol:Chloroform extracted pUC19 and pUCap1 DNA	164
3.4 Discussion	175
3.4.1 Protein movement along pUC19, pUCap1 and TF $\lambda$	175
3.4.2 Protein colocalisation	177
3.4.3 Movement along DNA	177
3.4.4 Pausing along DNA	178
3.4.5 Conclusion	179
<b>Chapter 4: Final conclusions</b>	<b>181</b>
4.1 Major conclusions from this thesis	184

4.2 Future work	186
References	188

## List of tables

### Chapter 2

1 Kymograph raw data	68
2 Order of binding sites along $\lambda$ DNA	78
3 Percentage of moving proteins along $\lambda$ DNA	86
4 Percentage of protein colocalisation	90
5 Percentage of pausing within each AP1 and variant AP1 target site by a simulated protein	93
6 Percentage of pausing within each AP1 and variant AP1 target site by a cFos	95
7 Percentage of pausing within each AP1 and variant AP1 target site by a cJun	97
8 Percentage of pausing within each AP1 and variant AP1 target site by a cFos:cJun heterodimers	98

### Chapter 3

9 Percentage of moving proteins along different DNA substrates	159
10 Percentage of protein colocalisation	163
11 Percentage of pausing within each AP1 and variant AP1 target site by a cFos:cJun heterodimers on $\lambda$ DNA	167
12 Simulated protein on Phenol:Chloroform extracted pUC19	169
13 Simulated protein on Phenol:Chloroform extracted pUCap1	169
14 Percentage of moving proteins on Phenol:Chloroform extracted pUC19	171
15 Percentage of moving proteins on Phenol:Chloroform extracted pUCap1	172

## List of figures

### Chapter 1

1 Images showing the effects of tumorigenesis	5
2 Cartoon diagram of MAP Kinase pathway	7
3 Cartoon diagram overview of transcription initiation and re-initiation pathway for RNA polymerase II	10
4 Diagram of a cFos:cJun heterodimer	12
5 Helical wheel and sequence alignment diagrams	14
6 Cartoon diagram of a Protein Fragment Complementation Assay	18
7 Cartoon diagram showing TIRFM	24
8 Cartoon diagram showing OAF microscopy	26
9 Cartoon diagram of the “waiting room model”	33

### Chapter 2

10 Flow cell construction diagram	41
11 Bacteriophage Lambda methylated and non-methylated DNA linear map	46
12 Lambda DNA with additional variant AP1 target sites linear map	48
13 DNA tightrope examples	50
14 cFos and cJun chromatograms	52
15 Absorbance graph	54
16 Cartoon diagram showing OAF microscopy setup	57
17 DNA and Kymograph analysis	60
18 Random walker simulation flow diagram	62
19 Simulated dimer taking a random walk with specific pausing	63
20 Simulated dimer taking a random walk with non-specific pausing	64
21 Examples of Gaussian distribution plots	67

22 Automated Gaussian fit raw data	70
23 Mean-squared displacement plot	71
24 Example Mean-squared displacement plots	72
25 The Binding Site Position Finder with AP1 target sites	75
26 The Binding Site Position Finder with AP1 and variant AP1 target sites	77
27 Data for the Binding Site Position Finder	81
28 Graphs showing diffusion constants against $\alpha$	87
29 Example of cFos:cJun colocalisation	89
30 Graph showing a simulated protein binding to $\lambda$ DNA	94
31 Graph showing protein correlation within $\lambda$ DNA binding sites	99
32 Graph showing correlation between the proteins and simulated protein	100
<b>Chapter 3</b>	
33 pUC19 plasmid map	111
34 AP1 oligonucleotide design	114
35 Making pUCap1 diagram	116
36 Sequencing of pUCap1	118
37 Inactivating T4 DNA Ligase using KCl gel	122
38 PEG <sub>6000</sub> concentration for optimal DNA ligation gel	124
39 A time course to create pUC19 tightropes that are comparable with the length of $\lambda$ DNA gel	126
40 DNA before and after Phenol:Chloroform extraction and DNA precipitation gel	131
41 Removing T4 DNA Ligase from pUC19 gel	132
42 Identifying target free $\lambda$ (TF $\lambda$ ) DNA diagram	135
43 Constructing $\lambda$ pUC gel	137
44 Constructing $\lambda$ pUC diagram	139

45 PEG <sub>6000</sub> concentration for optimal TF $\lambda$ DNA ligation gel	146
46 A time course to create TF $\lambda$ tightropes that are comparable with the length of $\lambda$ DNA gel	148
47 The Binding site Finder for pUC19	151
48 The Binding site Finder for pUCap1	153
49 Deconstructing $\lambda$ pUC gel	155
50 Graphs showing diffusion constants against $\alpha$ on Phenol:Chloroform extracted pUC19	160
51 Graphs showing diffusion constants against $\alpha$ on Phenol:Chloroform extracted pUCap1	161
52 Graph showing cFos:cJun heterodimers pausing on $\lambda$ DNA	168
53 Graphs of a computerised protein pausing on pUC19 and pUCap1	170
54 Graph showing visual correlation between the proteins	173
55 Graph showing visual correlation between the proteins and the simulated protein	174

## Abbreviations

Amp <sup>1</sup>	Ampicillin antibiotic
AP1	Activator Protein-1
ATP	Adenosine triphosphate
BSA	Bovine Serum Albumin
BPB	Bromophenol Blue DNA loading dye
d/d	Double digest
DPSS	Diode Pumped Solid State
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FastAP	Fast Alkaline Phosphatase
$\lambda$	<i>Bacteriophage</i> Lambda DNA
Ligase	T4 DNA Ligase
MCS	Multiple Cloning Site
MSD	Mean Squared Displacement
OAF	Oblique Angle Fluorescence microscopy
PCE	Phenol:Chloroform extracted
PCR	Polymerase Chain Reaction
PEG <sub>6000</sub>	Polyethylene Glycol <sub>6000</sub>
Qdot	Quantum dot
s/d	Single digest
SD	Standard deviation
TF $\lambda$	Target-free Lambda DNA
TIRF	Total Internal Reflection Fluorescence
TPA	phorbol 12-O-tetradecanoate-13-acetate
TRE	TPA-response element
TRIS	Tris(hydroxymethyl)aminomethane

v/v	Volume to volume
WT	Wild type
w/v	Weight per volume

## **Chapter 1: Introduction**

Cancer researchers are constantly trying to understand the intricate mechanisms used by cancerous cells to ensure their survival, and to find specific treatments to combat cancer and save the lives of those who suffer. This chapter looks at how single molecule techniques enable the direct visualisation of the AP1 transcription factors cFos and cJun as they interact with DNA tightropes.

### **1.1: What is cancer?**

Cancer, also referred to as malignant tumours and malignant neoplasms, is a generic term used to describe a collection of diseases that are caused by abnormal cellular growth that can occur within any part of the body. To date, over 100 different types of cancer have been found and it is crucial to understand the links between these cancer types (Kandoth et al. 2013). All types of cancerous cells are heterogeneous, meaning they are diverse in character and content. In most instances there are various mechanisms in place to promote their survival, such as increased cellular proliferation and the suppression of apoptosis due to mutations within the signalling cascade. This leads to the expansion and invasion of cancer within the body (Evan and Vousden 2001).

Throughout the last seventy years, researchers have been trying to provide a precise rationale for the pathogenic mechanisms causing cancer (Klausner 2002). This poses a significant challenge for cancer research as it is important to understand each molecular step within the different processes that influence cancer initiation and its survival. Understanding the molecular fingerprint of different types of tumours can help with the development of target specific therapies (Evan and Vousden 2001, Cairns et al. 2011).

### **1.2: The role of cFos and cJun in cancer**

Within the mammalian genome, the activation of a vast array of genes requires a combination of interactions between multiple transcription factors within specific promoter and enhancer regions that are located along the DNA (Hager et al. 2009, Champe et al. 2005, Lemon and Tjian 2000).

The transcription factors Fos and Jun are present within mammalian eukaryotic cells (Eferl and Wagner 2003). As proto-oncogenes, Fos and Jun encode functional proteins that prevent apoptosis, promote cell division and inhibit cellular differentiation, all of which are necessary for growth and development. However, if mutations occur within the proto-oncogenes, oncogenes develop; stimulating normal cells to become cancerous due to an increase in cell division, reduction in cellular differentiation and the prevention of apoptosis (Chial 2008).

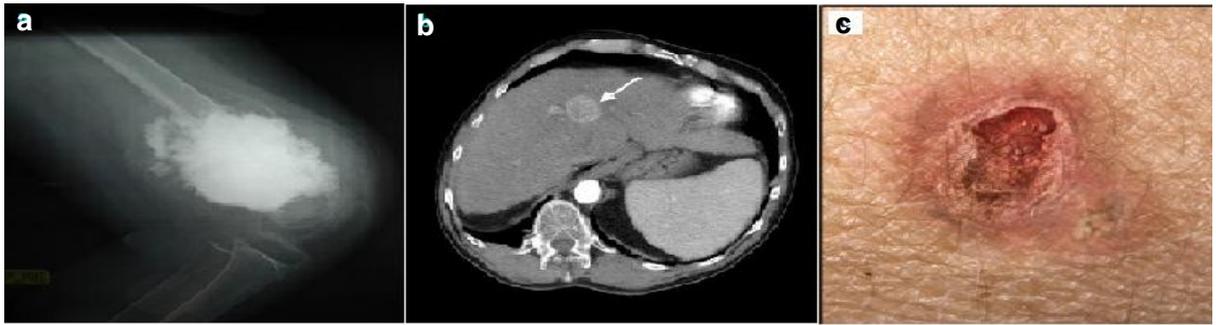
Interest in these particular transcription factors primarily arose due to the discovery of the viral oncoproteins v-Fos and v-Jun that were identified within the chicken retroviruses avian sarcoma virus ASV17 and Finkel-Biskis-Jenkins osteosarcoma, respectively (Latchman 1995, van Dam and Castellazzi 2001). Fos and Jun were identified as their cellular counterparts (Shaulian and Karin 2001). As members of an AP1 family that are a class of DNA binding proteins (Vinson et al. 1993), Fos and Jun were originally identified in the 1980's (Bohmann et al. 1987). This family contains an array of proto-oncogenic transcription factors consisting of 9 homologues. These are Fos (cFos, FosB, Fra-1, Fra-2) and Jun (cJun, JunB, JunD) (Rajaram and Kerppola 1997, Hai and Curran 1991, Angel and Karin 1991). Even though there is a high level of sequence homology, each AP1 functional protein has different transcriptional properties and biological effects that can modulate normal, healthy physiological functions (Nakabeppu and Nathans 1989, Leaner et al. 2003). These include cellular differentiation, proliferation and apoptosis (Jochum et al. 2001, Shaulian and Karin 2001, Santarpia et al. 2012). Under "normal circumstances," cFos has an important role in coordinating the function and differentiation of osteoclasts and osteoblasts, the two main bone tissue cell types. This occurs when the osteoclastogenic ligand RANKL, a Tumor Necrosis Factor (TNF) family member necessary for osteoclast differentiation becomes tightly regulated. This occurs when cFos specifically targets and induces RANKL genes within precursor osteoclast cells that are expressed via interferon  $\alpha/\beta$  signalling (Wagner 2002). cJun can also promote the proliferation of cells via the cell cycle and differentiation in cell lineages such as myloid, epithelial, hepatic and epidermal cells (Leppä and Bohmann 1999, Das et al. 2011, Zhang et al. 2006).

However, when cells within these cell lineages become cancerous, they present six distinct hallmarks (Hanahan and Weinberg 2011, Frémin and Meloche 2010). Firstly, maintaining proliferative signalling; this occurs when cancerous cells disable their senescence or apoptotic circuitry due to adapting to high levels of oncogenic signalling, resulting in increased cell proliferation and tumour growth. To sustain proliferation and growth, the oncogenic signalling between cancerous cells evoke maximum mitogenic stimulation while at the same time avoiding antiproliferative defences incurred by the cell cycle. Such hyperproliferation incurs DNA damage (Collado and Serrano 2010, Hanahan and Weinberg 2011). Secondly, growth suppressors avoidance; the TP53 protein that is responsible for triggering apoptosis downstream in the cell cycle in response to extracellular signalling such as stress becomes inactivated in cancerous cells (Hanahan and Weinberg 2011). Thirdly, invasion and metastasis activation; this occurs due to the suppression of TP53's tumour suppressing function, enabling cancerous cells to avoid the apoptotic circuitry within the cell cycle using anti-apoptotic mechanisms. Signalling within the cancerous cell population may promote the production of malignancies followed by metastasis where part of the tumour disseminates and travels to other organs around the body via blood vessels (Muller et al. 2011, Hanahan and Weinberg 2011, Muller et al. 2009). Fourthly, enabling immortal replication; typically there are two barriers imposed by the cell cycle which limit the number of cells that are able to grow and divide via the cell cycle. The first barrier is senescence, resulting in cells remaining viable but unable to proliferate. The second barrier is termed crisis, resulting in cell death. However there are occasions when cells overcome these two barriers, emerging from the crisis stage while displaying unlimited replication, resulting in cells becoming immortal and forming tumours (Hanahan and Weinberg 2011, Shay and Wright 2010).

Fifthly, causing angiogenesis; for tumours to grow like regular tissue, they also require oxygen, nutrition and be able to remove metabolic waste and carbon dioxide. To achieve this and assist with growth, the tumour induces new blood vessels by angiogenesis to develop from existing vessels via angiogenic regulator signalling proteins that bind to surface receptors located on vascular endothelial cells. Without angiogenesis, tumours could not grow or metastasise to

other organs around the body (Carmeliet and Jain 2000, Tonini et al. 2003, Duffy et al. 2008). Sixthly, cell death resistance; apoptosis is a natural barrier imposed by the cell cycle in order to avoid cancer progression. This occurs when apoptosis is regulated by effector components both upstream and downstream of the cell cycle. Balancing pro-apoptotic triggering proteins such as Bax and Bak and anti-apoptotic Bcl-2 family members helps control apoptosis. When this balance becomes disrupted, Bax and Bak promote the release of cytochrome c, a pro-apoptotic protein, from a cells mitochondrial outer membrane that activates proteolytic cascade that is part of the apoptotic program. In addition, oncoproteins such as Myc can also trigger apoptosis via hypoactive, oncogenic signalling (Zhang et al. 2013, Hanahan and Weinberg 2011).

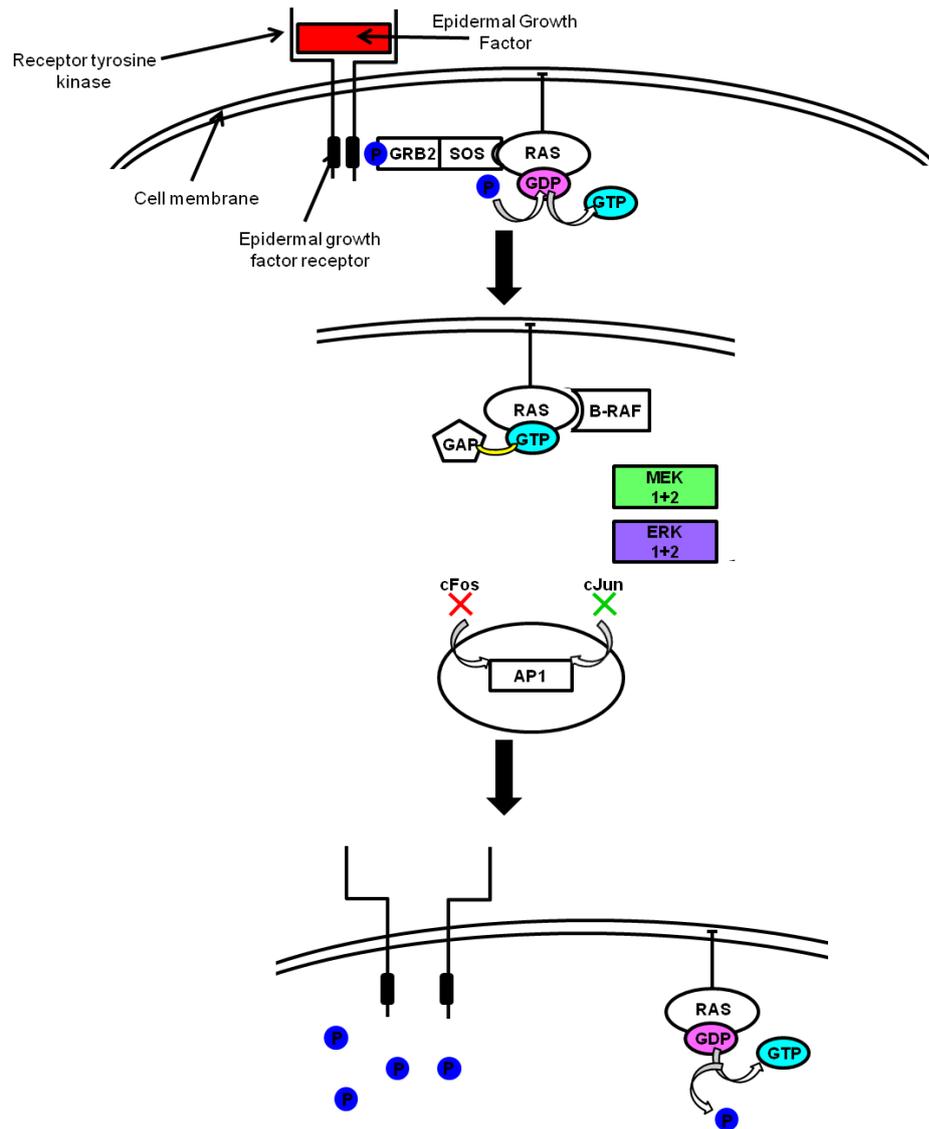
The AP1 family members cFos and cJun are known to implement tumorigenesis due to their over-expression and how this occurs is still under investigation (Eckhoff et al. 2013, Leaner et al. 2003, Hamacher et al. 2013). During tumorigenesis, cFos mediates the manifestation of malignant osteosarcomas and connective tissue neoplasms due to the transformation of chondroblasts and osteoclasts, causing tumorigenesis within these two different cell lines (Eferl and Wagner 2003, Karin et al. 1997, Wang et al. 1995). cJun mediates hepatocellular carcinoma which is the proliferation of hepatocytes within the liver (Das et al. 2011, Fausto 1998). Non-melanoma type skin cancer such as squamous cell carcinoma, a malignancy of the epidermal keratinocytes is also mediated by cJun (Zhang et al. 2006). Examples of these different cancer types are shown in Figure 1.



**Figure 1:** Images showing the effects of tumorigenesis initiated by over-expressed cFos and cJun. (a) X-ray of a knee showing osteosarcoma (McFadden 2002) that is caused by the over-expression of cFos. (b) Computerised tomography scan showing hepatocellular carcinoma. This is indicated by the arrow (Hinshaw. J. L. et al. 2012). Hepatocellular carcinoma can be caused by over-expressed cJun. (c) Squamous cell carcinoma (Healthwise 2008) can also be caused by the over-expression of cJun.

The proliferation and differentiation of healthy and tumorigenic cells occurs using different signalling pathways. In the case of cFos and cJun, the cell cycle becomes initiated through the mitogen-activated protein (MAP) kinase cellular signalling pathway (Figure 2) (Shaulian and Karin 2001, Brown 2007). The Ras protein, a member of the GTPase family, is a regulator of cellular growth within this signalling cascade (Downward 2003). Ras functions as a molecular switch, transmitting signals downstream to effector molecules, resulting in gene transcription. However, mutated Ras has also been reported in some types of cancer, whereby the signalling cascade is permanently on due to mutated Ras (Santarpia et al. 2012).

The transcription factors cFos and cJun become activated downstream from Ras at the closing stages of the MAP kinase pathway and regulate gene expression by controlling the rate of transcription (Whitmarsh 2007). This occurs by cFos and cJun binding to specific AP1 target sites (Angel et al. 1988, Brown 2007, Hager et al. 2009, Simonson et al. 1992, Nakabeppu and Nathans 1989), also known as a TPA (12-O-Tetradecanoylphorbol-13-acetate) response element (TRE), and this is characterised by the seven base pair, palindromic recognition sequence  $TGA \frac{C}{G} TCA$  (Nakabeppu and Nathans 1989, Simonson et al. 1992, Angel and Karin 1991). This is a potent tumour promoter and regulator of biological responses that include cellular proliferation and apoptosis (Shyy et al. 1995, Eferl and Wagner 2003, Angel and Karin 1991). Homologous to Fos and Jun is the yeast transcription factor GCN4 (Nakabeppu et al. 1988, Ellenberger et al. 1992). Unlike cFos and cJun, GCN4 can activate transcription in *Saccharomyces cerevisiae* yeast cells upon binding to its target sequence: TGACTC (Oliviero et al. 1992, Arndt and Fink 1986).

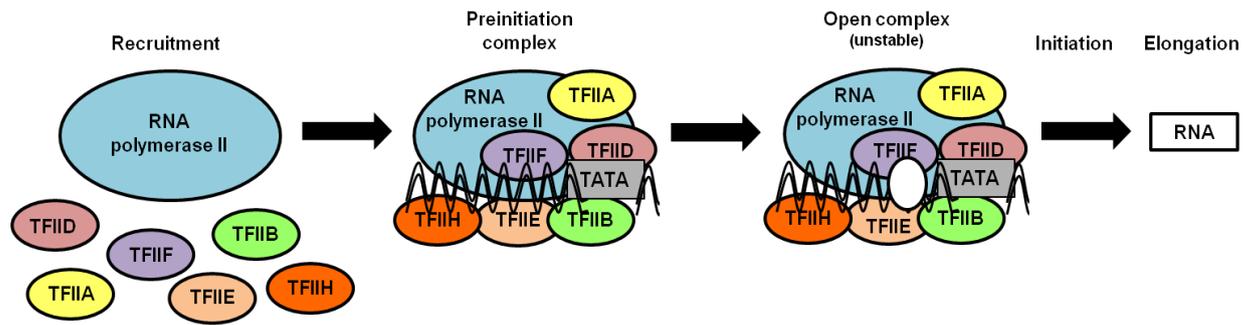


**Figure 2:** A cartoon diagram outlining the Map Kinase Pathway (adapted from (Frémin and Meloche 2010)). When the Epidermal Growth Factor (EGF) binds to the extracellular portion of the ligand receptor tyrosine kinase, causes the receptors two subunits to dimerise. The intracellular portions of the receptors catalyse self phosphorylation, enabling the binding of GRB2 (Growth Factor Receptor-bound protein-2). SOS (Son of Sevenless) can then bind onto GRB2 and the membrane bound protein Ras which is inactivated due to its GDP state. Ras activates upon catalysing GDP to GTP and can bind to proto-oncogenic effector protein, B-Raf, which in turn phosphorylates the kinases Mek1/2 which phosphorylate the kinases ERK 1/2. These kinases activate Fos and Jun that bind to their AP1 target site, enabling them to exert their biological function, such as cell proliferation. To inactivate GTP bound Ras, GAP (GTPase-Activating protein) assists with GTP hydrolysis to GDP, therefore turning off the MAP Kinase signalling pathway (Frémin and Meloche 2010, Weinberg 2007 ).

cFos and cJun's AP1 target sites are located within promoter regions along DNA that are adjacent to the genes they regulate (Chinenov and Kerppola 2001). Chromatin is the form in which eukaryotic DNA is packed into cells. The nucleosome is the basic unit of chromatin and is composed of two copies of four globular core histones: H3, H4, H2A and H2B that are wrapped with 147 bp of DNA (Kouzarides 2007, Shogren-Knaak et al. 2006). For transcription to occur, the DNA and its sequence-specific sites must become accessible to transcription factors, however, when chromatin is in a non-modified form, it represses transcription. To activate transcription, the chromatin must become available to the transcription factors (Narlikar et al. 2002). Histone modifying processes such as; acetylation which decreases interactions between internucleosomes and alters electrostatic charge and methylation which recruits proteins that require access to DNA without influencing electrostatic charge. This will cause the N-terminal histone tails to extend outwards from the nucleosomes core, enabling the DNA to become more accessible to transcription factors. In addition, modifications to histones can also alter the structure of the nucleosome and this contributes towards increasing DNA accessibility (Sims et al. 2003, Berger 2007).

The start of the transcription cascade occurs at the focus core promoter and commences when RNA is synthesised from a DNA template using RNA polymerase II machinery. However, for RNA polymerase II to recognise this promoter, it requires basal transcription factors such as; TFIID (Transcription Factor for RNA polymerase II D) which recognises focus core promoters, binding to the TATA promoter element deforming promoter DNA and recruiting TFIIA and TFIIB. TFIIA helps to stabilise TFIID DNA-binding, blocks transcription inhibitors and promotes positive and negative gene regulation. TFIIB assists with recruiting RNA polymerase II to the focused core promoter while helping to fix transcriptions starting site (Juven-Gershon and Kadonaga 2010). When TFIID, TFIIA and TFIIB form a stable complex, this phosphorylates RNA polymerase II and TFIIF which is part of the RNA polymerase II initiation complex. In parallel, an activated complex forms between TFIIIE which binds the promoter near to the site of transcription and TFIIH which assists with keeping the DNA unwound and open. These basal

transcription factors enable the promoter to be clear, enabling RNA synthesis to take place (Juven-Gershon and Kadonaga 2010, Hahn 2004). This process has been shown in Figure 3.

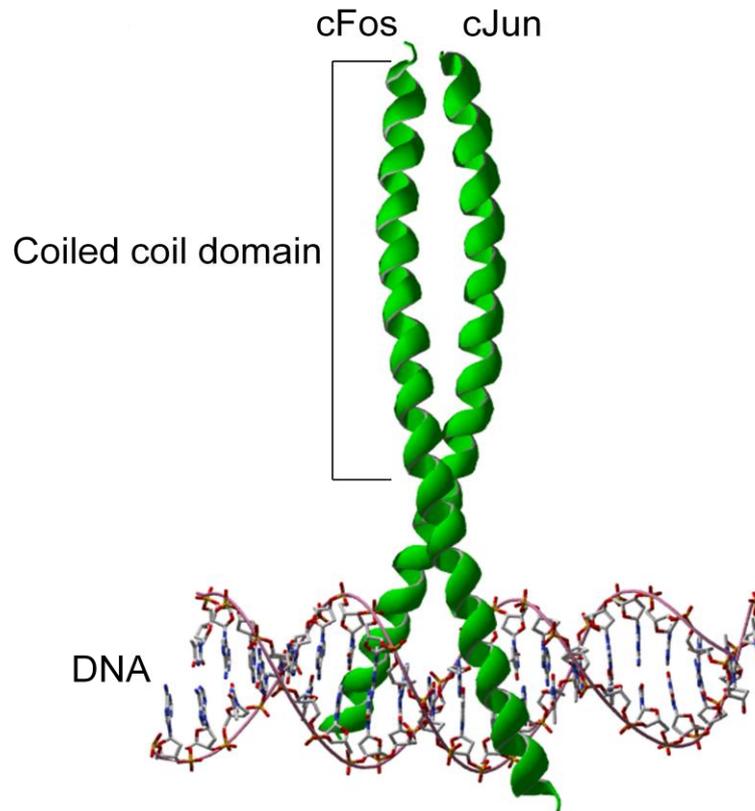


**Figure 3:** A cartoon diagram showing an overview of the transcription initiation and re-initiation pathway for RNA polymerase II (adapted from (Hahn 2004)).

These two transcription factors have a significant role in controlling cellular responses to a variety of extracellular stimuli, such as bacterial/viral infections, extracellular stress signals and various cytokines and growth factors (Zenz et al. 2008). However, when cFos and cJun become over-expressed, pleiotropic changes to the cells occur, known as malignancies, when the closely regulated stages of the cell cycle are no longer able to produce normal, healthy cells (Chow 2010). Cancerous cells are produced instead via the cell cycle. This occurs when the p53 gene responsible for tumour suppression becomes antagonised downstream, resulting in tumour cells being protected against apoptosis (Zhang et al. 2007, Weinberg 2007, Evan and Vousden 2001). It has also been reported that when cJun initiates the proliferation and differentiation of cells within the MAP kinase pathway, the MAP kinase p38 $\alpha$ , an important regulator of both cancer progression and cellular development, becomes activated due to auto-phosphorylation events in response to cellular stress (Bradham and McClay 2006).

### **1.3: The structure of cFos and cJun**

cFos and cJun are composed of amphipathic  $\alpha$ -helical coiled coils (Krylov et al. 1994). Located at one end of each coiled coil is the N-terminus. This contains a transactivation domain which enables chromatin to decondense downstream causing a transcription initiation complex to be established (Landschulz et al. 1988, Moras and Gronemeyer 1998). The basic region contains a DNA binding recognition sequence and a leucine zipper region that brings the two  $\alpha$ -helical coiled coils together. These are located at the C-terminus at the other end of the coiled coil (Vinson et al. 1993, Mason et al. 2006). This structure is shown in Figure 4.



**Figure 4:** A diagram showing a cFos:cJun heterodimer that has bound to DNA (Glover & Harrison, 1995 (PDB access number 1FOS)).

As cFos and cJun dimerise via the leucine zipper interface of each  $\alpha$ -helical coiled coil (Kohn et al. 1997, Viadiu and Aggarwal 2000, Ellenberger et al. 1992), they form a left-handed, superhelical conformation (Burkhard et al. 2000). This unique characteristic occurs due to the continuous contact of the hydrophobic amino acid residues present within each  $\alpha$ -helical leucine zipper region (Abate et al. 1991), forming tightly bound homodimeric or heterodimeric complexes due to their juxtaposed basic regions (Chen et al. 1998). The most common coiled coil structure is a left-handed supercoil that contains two right-handed  $\alpha$ -helix that have wrapped around each other (Lupas and Gruber 2005, Mason and Arndt 2004, Nakabeppu et al. 1988, Landschulz et al. 1988, Vinson et al. 1993). This occurs when side chains within the cores adopt their designated positions turn by turn, resulting in 3.6 residues per turn being reduced to 3.5 residues, the equivalent of 7 residues. This is known as a heptad repeat in a supercoiled structure (Lupas and Gruber 2005). This heptad repeats at least twice along each  $\alpha$ -helix (Landschulz et al. 1988) and has been described using the nomenclature '**abcdefg**', as shown in Figure 5.



During dimerisation, the two  $\alpha$ -helix oligomerise through the hydrophobic residues that are in positions **a** and **d** within the hydrophobic core (as seen in Figure 5). Predominantly, evolutionary conserved leucine is favoured within position **d** and these are aligned along the dimerisation interface within cFos and cJun (van Dam and Castellazzi 2001, Mechta-Grigoriou et al. 2001, Wagner 2001). As the core is slightly solvent exposed, having hydrophilic asparagine, a conserved residue in position **a** helps maintain the structures stability (Woolfson and Alber 1995, Arndt et al. 2000). This is in keeping with Crick's (1953) conventional 'knobs in holes' pattern along the dimerisation interface, ensuring that these hydrophobic residues remain unexposed to the external solvent (Crick 1953). Therefore, hydrophobic interactions occur when the residue in position **a** on one  $\alpha$  helix directly interacts with position **a** on the opposite  $\alpha$  helix. This interaction also occurs between residues in position **d** which have a major role in oligomerisation (Woolfson and Alber 1995).

Originally, Landschulz et al (1988) proposed that cFos and cJun can both homodimerise and heterodimerise. It is now thought that cFos can only form a stable heterodimer with cJun, whereas cJun can also homodimerise (Hai and Curran 1991, Ito et al. 2001). Upon dimerisation, cFos and cJun form a transcriptionally active, classic AP1 complex (Leaner et al. 2003) that can bind onto their AP1 target sites located on the major grooves of the DNA backbone (O'Neil et al. 1990). For DNA binding to occur, the proteins DNA binding domain, enriched with arginine and lysine amino residues (Olive et al. 1997, Talanian et al. 1990), enables specific base pair sequences along the double stranded DNA phosphate backbone to be recognised, such as the proteins specific target site sequence. These specific interactions are supported and stabilised by electrostatic interactions that occur between the negatively charged DNA phosphate backbone and the amino residues that surround the proteins target binding site (Von Hippel 2004).

However, DNA binding specificity can become complicated due to the occurrence of non-specific binding events taking place between the protein and non-specific areas of the DNA. Works performed using the *lac* repressor have helped with understanding how DNA binding proteins use different mechanisms when binding non-specifically to DNA (Von Hippel 2004,

Riggs et al. 1970). These non-specific DNA interactions assist with the protein locating sequence specific target sites (Kalodimos et al. 2004). The more non-specific binding that occurs on DNA, the less free *lac* repressor, or other DNA binding proteins, remain in solution. The binding of proteins to non-specific regions of DNA is weaker than target site binding when under physiological salt concentrations. To stabilise the structural integrity of the proteins when non-specifically bound to the DNA, a net positive charge is created around the binding site only which orientates the protein while facilitating its recognition of the target site (Von Hippel 2004). Therefore, 1-Dimensional diffusion such as sliding and hopping will enable the protein to locate its specific target site (Von Hippel 2004).

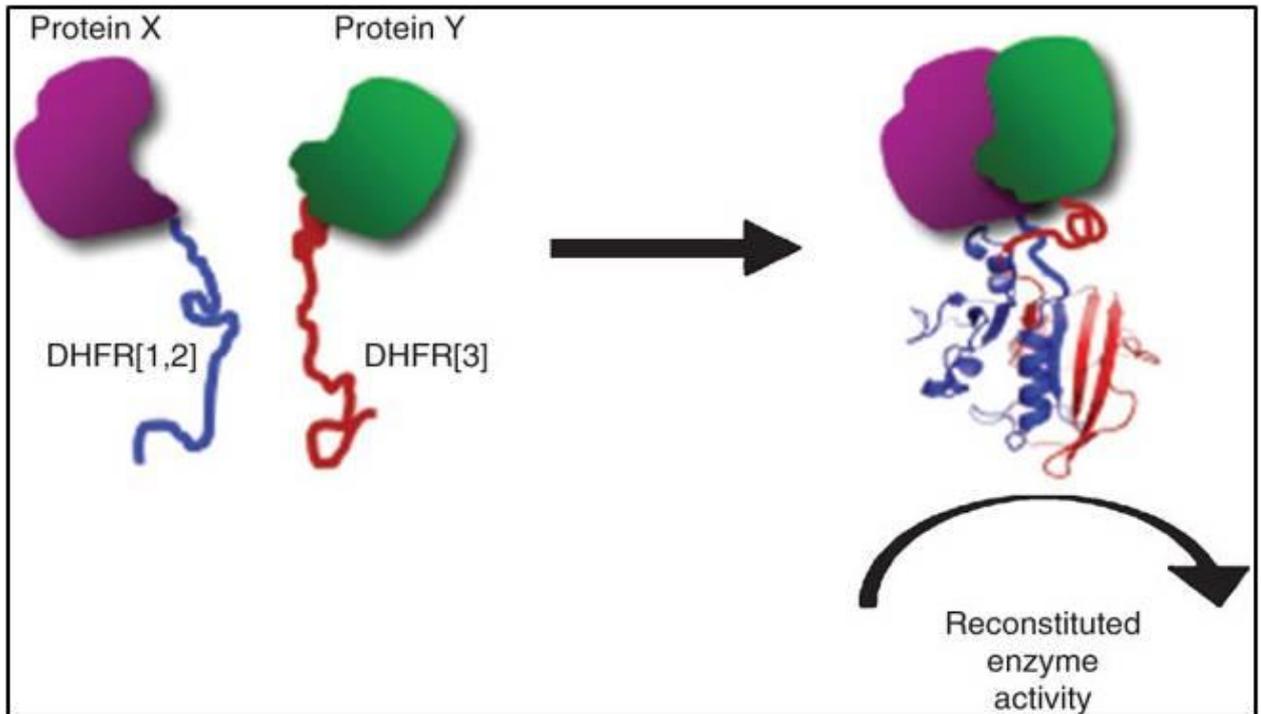
Glover and Harrison (1995) used X-ray crystallography to identify the actual structure of a bound cFos:cJun heterodimer to its AP1 target site, forming a protein-DNA complex. They showed that the DNA binding domain of cFos:cJun heterodimers can bind to the AP1 target site in any orientation and this has been supported by the work of others (Chen et al. 1995, Leonard et al. 1997, Glover and Harrison 1995).

Due to cFos' inability to form stable homodimers, it is believed that core residues within hydrophobic position **a**, such as hydrophobic isoleucine, hydrophilic threonine and basic lysine prevent cFos homodimerisation (Oliviero et al. 1992, Schuermann et al. 1989, Woolfson and Alber 1995). It is also thought that cFos:cJun heterodimers have a more stable core region than cJun homodimers (Halazonetis et al. 1988, Mason et al. 2006), however, the strength of this interaction within the cFos and cJun leucine zipper region still remains unknown (Newman and Keating 2003).

#### **1.4: The antagonistic peptide inhibitor FosW**

To overcome the tumorigenic effects of cJun over-expression, Fos Winner (FosW), an *in vitro* designed antagonistic peptide inhibitor, disrupts cJun homodimers with high specificity and affinity (Mason 2009). FosW is a member of a peptide library that contains many variations of cFos where residues within the dimerisation region were substituted to improve its stability. cFos was chosen due to its instability and has more scope for improvement when compared with cJun (Mason et al. 2006). This was selected using protein-fragment complementation

(PCA) assay (Remy et al. 2007), as shown in Figure 6, allowing *in vitro* protein-protein interactions to be determined. This assay is a biochemical process that enables the assembly and disassembly of protein complexes via protein-protein interaction. To achieve this, two proteins of interest, known as protein X and Y, are fused to a reporter protein, such as the enzyme dihydrofolate reductase (DHFR) that is composed of two complementary fragments. If an interaction between proteins X and Y occurs, this brings the two parts of the reporter protein together which then proceed to fold into its native structure, restoring its activity (Remy et al. 2007). Therefore, in the case of cJun+FosW, the interaction between the coiled coils of FosW and cJun to form a heterodimeric complex mediates the full activation of DHFR enzyme when FosW+cJun form a stable heterodimeric complex with high specificity, prior to production inside *Escherichia coli* (*E. coli*) (Hagemann et al. 2008, Hager et al. 2009).



**Figure 6:** A cartoon showing the basic principle used during Protein Fragment Complementation Assay (PCA) (Remy et al. 2007). The enzyme dihydrofolate reductase is separated into two complementary parts which are each fused to a protein of interest, that is protein X (blue) and Y (red). When an interaction occurs between the two proteins, this brings the DHFR complementary parts together, activating the enzyme.

### **1.5: Biochemical and biophysical methods for probing protein-DNA interactions**

Before the study of individual proteins interacting with DNA using single molecule techniques, more traditional biochemical and biophysical methods were used to determine protein-nucleic acid interactions. Electromobility Shift Assay (EMSA) is one classic method used to detect protein-nucleic interactions (Hellman and Fried 2007, Orchard and May 1993). The nucleic acid used varies in size from short oligonucleotides to thousands of base pairs long and is radioisotope-labelled with  $^{32}\text{P}$ . This results in a highly sensitive assay being performed with either protein-DNA concentrations at  $<0.1$  nM or sample volumes of  $\leq 20$   $\mu\text{L}$ . If protein-nucleic acid complexes form, these will migrate more slowly through the gel matrix showing a mobility shift unlike free nucleic acid (Hellman and Fried 2007).

Although EMSA provides quantitative results, this method also has further limitations, for example; the size of the protein interacting with the DNA determines how much mobility shift will be ascertained; proteins that rapidly dissociate from the DNA are not detected and proteins that slowly dissociate from the DNA results in their overall binding density being underestimated (Hellman and Fried 2007). Primarily, EMSA is a technique that formed the backbone of numerous protein-DNA interaction studies (Orchard and May 1993), but this methodology is potentially flawed by its inherent averaging and the forces experienced by the complexes as they travel through the gel matrix. When performed using cFos and cJun, EMSAs were only able to provide a bound protein-DNA ratio. Therefore it was concluded that cFos could not associate with DNA containing AP1 but cJun could form DNA-bound homodimers and heterodimers (Turner 1989).

Filter binding assays are also used to determine protein-DNA interactions (Rio 2012). The basic principle uses a nitrocellulose filter to quantify the affinities between protein and radioactive labelled  $^{32}\text{P}$ -DNA. As the nitrocellulose paper is negatively charged, any unbound DNA will not adhere to the paper, however the paper will immobilise protein-DNA complexes due to the proteins having a positive net charge. The quantification of protein-DNA complexes immobilised on the paper can be determined, once the paper has been dried, using a scintillation counter (Shindo et al. 1993).

Another method used to determine protein-DNA interactions is isothermal titration calorimetry (Freyer and Lewis 2008, Keller et al. 2012). This physical technique is used to determine the thermodynamic parameters of molecular interactions taking place in solution between large macromolecules like protein and DNA. The heat generated from each reaction can be determined by an aliquot containing the protein that is injected into a sample cell containing the DNA. The more protein that is injected into the sample chamber, the more saturated the DNA becomes and the heat generated starts decreasing (Keller et al. 2012). Like the fore mentioned methods, ITC is able to quantify the amount of protein-DNA interactions that occurred within solution. Positive attribute to this method when compared with the filter binding assay and the EMSA is that there are no molecular weight restrictions as the reactions all take place in solution and  $K_d$  values are obtained without any radioactive labels (Milev 2013).

From the data obtained from the fore mentioned techniques, dissociation constants ( $K_d$ ) can be determined from the point at which half of the DNA appears bound with protein and the other half appears unbound (Heffler and Walters 2012). Alternatively, a more precise  $K_d$  can be determined by using ImageJ to quantify the fluorescent signal in each DNA band and plotting a graph showing the fraction of bound DNA protein against the concentration of  $^{32}\text{P}$  labelled DNA. The use of additional software can obtain the  $K_d$  from non-linear regression (Heffler and Walters 2012, Shindo et al. 1993). All of these methods are able to provide a bound protein-DNA ratio, however they are unable to provide information about each individual protein that binds to DNA.

### **1.6: Single molecule techniques**

The first single molecule study was performed using patch clamp experiments on ion channels in frog muscle fibres (Neher and Sakmann 1976). This work initiated a new wave of experimentalism that offered a new view of biological mechanism not shrouded by the cloak of averaging as single studies allow individual molecules to be studied. Now single molecule techniques enable the direct visualisation of individual proteins and the search mechanisms they use to locate their target sites as they move along DNA in real time (Hilario and Kowalczykowski 2010, Kad et al. 2010). This is important as it allows each protein's movement and interaction with each other and with the DNA to be observed over long periods of time.

For proteins to be visualised, Quantum dots (Qdots) have become increasingly popular due to their high resistance to photobleaching and brightness (Wang et al. 2008). Their properties make them favourable over more traditional fluorescent labels, such as GFP (green fluorescent protein) that provide a much shorter photostable time (Bruchez et al. 1998, Leutwyler et al. 1996). Another advantage of using Qdots is they have a broad excitation spectrum and multiple coloured Qdots can be imaged simultaneously to enable many systems to be studied within a single assay. However, like all experimental techniques, Qdots also have their limitations. When under continuous excitation by the laser beam the Qdots 'blink'. The rationale behind this property is still not fully understood (Medintz et al. 2005).

To visualise protein-DNA interactions, the DNA must be elongated since in solution it normally exists as a polymer globule. One technique uses magnetic tweezers where a DNA molecule can be manipulated and elongated by tethering one end to a superparamagnetic Dynabead and the other to a surface or another bead. This bead is made from crosslinked polystyrene that contains  $\text{Fe}_2\text{O}_3$  and  $\text{Fe}_2\text{O}_4$  and displays magnetic properties only when in the presence of a magnetic field (Zlatanova and Leuba 2003, Haber and Wirtz 2000). Using molecular tweezers is a particularly useful single molecule technique as it enables a single protein to bind and interact with a single piece of DNA. One disadvantage however is this method would be very time consuming while producing minimal data.

The elongation and linearisation of DNA can also be achieved using DNA curtains. This technique involves one end of the DNA being anchored to a lipid bilayer that covers the surface of a flowcell (Granéli et al. 2006). The DNA remains linearised using a hydrodynamic force that orientates the curtains in the same direction as the buffer flow. Microscale diffusion barriers are etched into the flowcell to assist with keeping the DNA molecules aligned with respect to each other. This single molecule technique is useful as it allows the simultaneous imaging of approximately 100 bound proteins per DNA molecule, resulting in several hundred (or even thousands) of aligned DNA molecules being imaged within a single field-of-view (Song et al. 2016, Fazio et al. 2008). Another positive attribute is that data can be acquired from thousands of individual molecules rapidly. However, the DNA curtains technique also has limitations. The

accumulation of DNA molecules along the edge of the microscale diffusion barriers occurs, compromising the quality of the visual surface and leads to the scattering of light during imaging. Some of the DNA molecules may be unevenly aligned and the field-of-view may also be inefficiently covered with DNA molecules and bound proteins (Fazio et al. 2008).

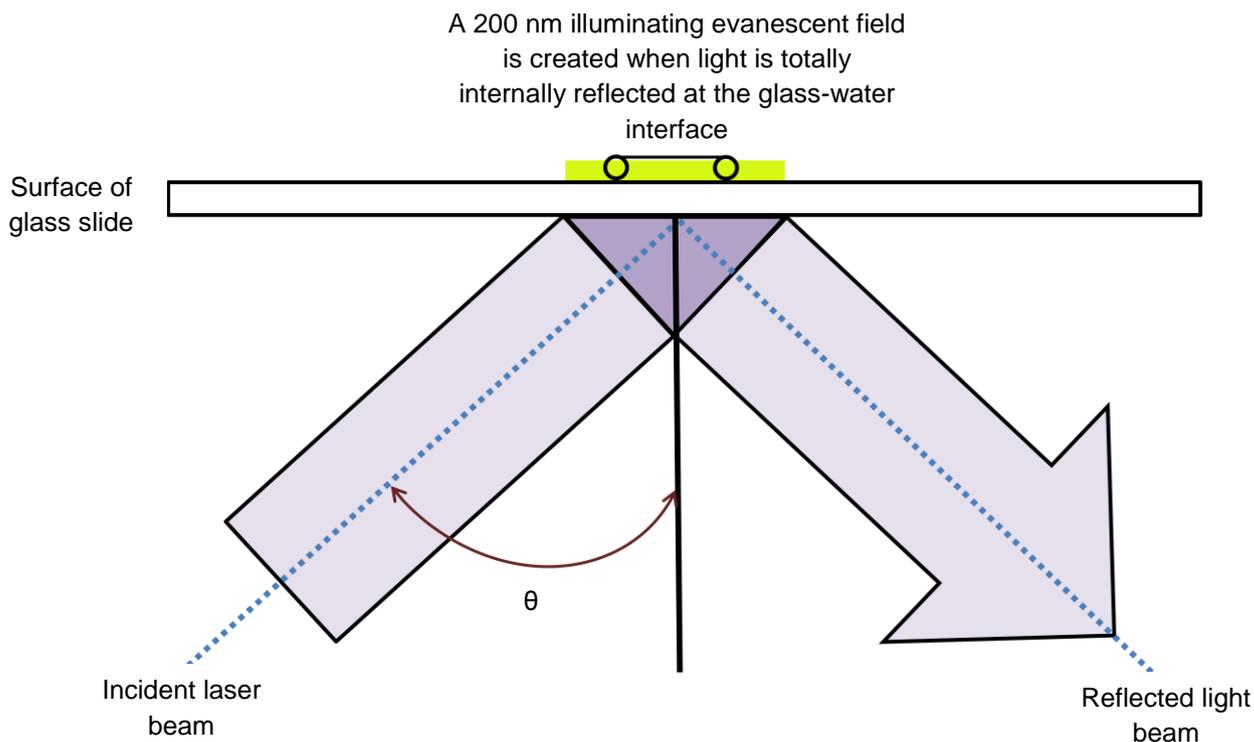
Another technique used to elongate and linearise DNA is using a DNA tightrope approach. DNA is suspended 5 µm above the surface of a slide on silica bead pedestals coated in Poly-L-lysine. This technique allows the direct visualisation of individual fluorescently labelled Quantum dot (Qdot) conjugated proteins in real time and is useful for observing protein-protein and protein-DNA interactions as there is no interference from artefacts adhered to the slides surface that can hinder protein movement (Hughes et al. 2013, Kad et al. 2010). The single molecule technique used in this study, the DNA tightrope approach, has provided an invaluable insight into how cFos, cJun and the biological peptide inhibitor FosW interact with each other and search the DNA for their targets. DNA tightropes have been used extensively with proteins from the Nuclear Excision Repair Pathway (Kad et al. 2010). cFos and cJun have never been studied using single molecule techniques before.

Total Internal Reflection Fluorescence microscopy (TIRFM), as shown in Figure 7, offers a means to detect single molecules with improved signal to noise (Joo et al. 2008, Mattheyses et al. 2010). This imaging technique is useful when visualising material, such as cells, that are localised to the surface of a slide (Fish 2009).

TIRFM uses an optical phenomenon where the light beam is refracted at a glass surface of a high refractive index meeting a solution of a lower refractive index. The refractive index is determined by how much an excitation light beam is deflected. In this case, the refractive index of the glass and water interface is determined by measuring the deflection of a light beam as it enters and leaves the glass and water interface at an angle, known as the angle of incidence (Fish 2009). This has been stated in Snell's law whereby a line is drawn perpendicular to the mediums interface and the angle of incidence,  $\theta_1$  is determined by the angle of which the light beam exits the medium,  $\theta_2$ . This is represented by the formula (Smith et al. 2004):

$$\sin(\theta_1) = n\sin(\theta_2)$$

It is the difference between the glass and water interface that determines the angle of incidence. In TIRFM, as the angle of incidence is raised so is that of the emergent refracted beam until it no longer propagates into the solution. This is known as the critical angle (Fish 2009) and at this point the laser beam is reflected back towards the source (Ishijima and Yanagida 2001, Pierce et al. 1997). A consequence of this geometry is that an evanescent field is produced approximately 50-200 nm above the reflecting interface in the solution, however the intensity of this evanescent wave decreases exponentially the further the wave is from the glass (Mattheyses et al. 2010, Walter et al. 2008). This evanescent wave is determined by the incident illumination angle (Fish 2009). Primarily, TIRFM works by selectively illuminating the fluorescently labelled proteins that remain close to the interface surface within the evanescent field (Ishijima and Yanagida 2001). Fluorescently labelled proteins outside of this evanescent field are not excited by the light beam. According to Granéli et al (2006), this results in a significant reduction in background fluorescence signals (Granéli et al. 2006). TIRFM is therefore a suitable visualisation tool for analysing biological matter such as cellular membranes that are on the surface of the slide. This is due to high resolution images being generated at the flow cells surface with minimal background fluorescence (Mattheyses et al. 2010).

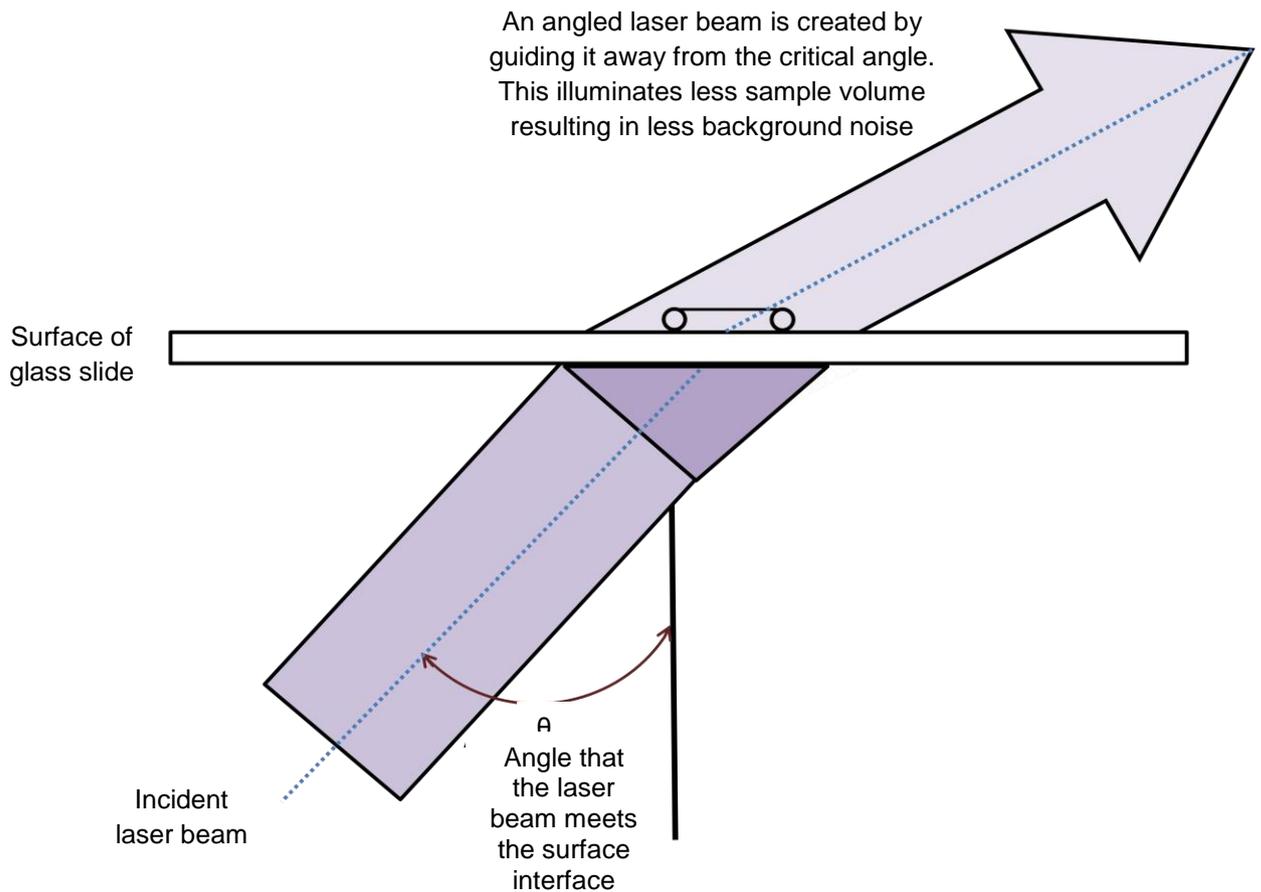


**Figure 7:** TIRFM imaging. TIRFM uses an evanescent wave that is generated when a beam of light, at a critical angle of  $63^\circ$ , is totally internally reflected at the glass-water interface. It is when this light beam encounters the interface between the glass and the slide that it becomes refracted by  $90^\circ$  in a direction parallel to the interface. This results in the illumination of fluorophores that are adjacent to the glass-water interface and within a 200 nm area above the surface. OAF microscopy, however, has its collimated light beam angled away from the critical angle. This results in the illumination of a thin section and the reduction in background noise.

When using TIRFM to visualise cellular material for example, fluorescently labelled molecules at a nanomolar range between 10-100 nM can be resolved at a sufficiently high signal/background ratio, ideally, one fluorescently labelled molecule per diffraction limit, the maximum angle at which the laser beam can be at before being totally internally reflected at the critical angle (van Oijen 2011, Geertsema et al. 2015). This nanomolar limit consequently decreases the amount of biomolecular interactions, resulting in dissociation constants ( $K_d$ ) being within the nanomolar range or lower (Geertsema et al. 2015).

TIRFM is a good technique to use when visualising cellular structures such as; cytoskeletons which are located near to the plasma membrane; the formation of endocytotic vesicles which assist with the recruitment of cytosolic proteins to the plasma membrane; secretory carriers can be visualised near the cell membrane and the visualisation of different types of intracellular signalling (Mattheyses et al. 2010).

Oblique Angle Fluorescence (OAF) microscopy however is an alternative illumination strategy whereby a lens is used to defocus and steer the beam off the edge of the back aperture unlike in TIRFM as shown in Figure 8. This results in an inclined collimated beam of light that is less than the critical angle. OAF can image deeper in the experimental chamber unlike TIRFM (Kad et al. 2010). Unlike TIRFM, using OAF to image at a surface to visualise individual proteins may be problematic since a protein's movement along DNA can be significantly hindered by interactions that occur on the surface, affecting their kinetic and dynamic behaviour. Raising the DNA platform, in the form of tightropes, away from the surface will reduce background fluorescence and the proteins can move without surface-induced artefacts. Any fluorescence within the focal plane does not occur due to fluorescently labelled proteins that may be attached to the surface, but from proteins bound to the DNA tightropes (Kad et al. 2010).



**Figure 8:** Our custom built OAF microscope uses a 20 mW 480 nm, 1.66 mm DPSS laser (JSDU) for excitation that expanded 17.5x and focussed onto the back focal plane (bfp) of a high NA objective lens (Olympus 1.45 NA 100x) using a 250 mm lens. By translating the laser across the bfp, its angle of incidence could be manually adjusted to achieve OAF. This optimisation was achieved by adjusting the angle of incidence until fluorescence could still be seen at the height of the beads. Any further adjustment would then begin to send the laser beam below the height of the tightropes, resulting in this imaging technique failing to work. All emitted light was passed through an Optosplit III to separate the image into three different wavelengths. The dichroic mirrors used for this were D565/20m, D605/20m and D655/20m (Chroma technology). Finally, images were collected using a DU897 EMCCD camera (Andor, Belfast UK) (Figure 5).

Variable time recordings were taken of proteins bound to DNA tightropes. Usually recordings ranged from 2-10 minutes with an exposure time of 0.1 seconds and a kinetic cycle time of 1.00011 seconds. This meant that an image was taken every 1.00011 seconds throughout the duration of the recording.

Another type of single molecule technique used to visualise fluorescently labelled material that is located deeper in the experimental chamber is epifluorescence microscopy. This was the first conventional, fluorescence imaging technique used to visualise individual fluorophores (Funatsu et al. 1995). During imaging using epifluorescence microscopy, the laser beam is passed directly through the sample and the slide, resulting in background noise from fluorophores deeper in solution. Therefore, this type of microscopy is more suited for imaging protein dynamics that are beyond the localised area of a slide surface (Funatsu et al. 1995).

Although not classed as a single molecule technique, chromatin immunoprecipitation coupled with high throughput sequencing (ChIP-Seq) is a powerful technique that can identify DNA binding sites for transcription factors and other proteins on a genome-wide scale (Håndstad et al. 2011). The transcription factors or proteins of interest are covalently cross-linked to the DNA. Small fragments of DNA containing cross-linked proteins suitable for sequencing are sheared by either sonication or enzymic digestion (Kharchenko et al. 2008). Prior to DNA purification, an antibody that is specific to the cross-linked proteins reverses the cross-linking. ChIP-Seq is useful as it can find differences between where transcription factors and other proteins mostly occupy within different cells (Håndstad et al. 2011, Raha et al. 2010, Kharchenko et al. 2008).

Raha *et al* (2009) studied the close association of the transcription factors cFos and cJun with RNA polymerase III (Pol III) using ChIP-Seq in human K562 erythroleukemia cells. Using ChIP-Seq, a map was generated to show the distribution of cFos and cJun within the K562 cells. The results showed that there were 23,369 and 35,746 binding sites detected for cFos and cJun respectively. They also found that 86% and 76% of cFos and cJun AP1 target sites were located 100 bp upstream from the Pol II and Pol III genes. Binding in close proximity to Pol II and Pol III suggests that cFos and cJun play a greater role in eukaryotic gene expression and both polymerases work together to assist with the coordination of gene expression (Raha et al. 2010).

In contrast to the study performed by Raha *et al* (2009), Seldeen *et al* (2009) used isothermal titration calorimetry to determine the thermodynamics behind cFos:cJun heterodimers binding to

synthetic dsDNA oligonucleotides that contained the AP1 target site; TGA<sup>C</sup><sub>G</sub>TCA and all single nucleotide variations of this sequence. In addition, it was shown that cFos:cJun heterodimers also bound to specific variations of the AP1 target site based on their lowest  $K_d$  values. The variant sites containing the single nucleotide substitutions (highlighted and underlined) were; TGAGGTCA, TGAITCA, TGACACA and TGACGCA. However, the study also found that cFos:cJun heterodimers could also bind to DNA containing the variant target sequences; CGACTCA, TGACTCG, TGITCTCA, TGACACA, TGCCCTCA and TGACGCA, despite having a higher  $K_d$ . Mutating the AP1 target sequence does not hinder cFos:cJun heterodimers from binding to these variant target sites (Seldeen et al. 2009).

### **1.7: Diffusion mechanisms and AP1 binding theories**

cFos and cJun can exert their biological function upon binding to their AP1 target sequence TGAC/GTCA as previously discussed in this chapter. However, how these transcription factors locate this site is yet to be understood. Winter et al (1981) tried to answer the question: 'How at a transcriptional level does gene expression become regulated by proteins once their target binding site has been located?' (Winter et al. 1981). Although the AP1 proteins were discovered in the 1980s, the way cFos and cJun locate their target sites still remains unclear (Lu et al. 2005). In recent years, researchers have come up with theories to try to answer this question (Gorski et al. 2006) by performing single molecule studies on protein-DNA interactions (Hughes et al. 2013, Kad et al. 2010). This enabled a greater insight into how DNA binding proteins locate their final destination.

How a protein can locate its specific target amongst a huge excess of non-specific DNA regions remains unclear (Viadiu and Aggarwal 2000). In 1970, Riggs et al studied how the LacI repressor located its target 'operator' site. They found that the LacI repressor travelled one thousand times faster than was originally predicted and used 3-Dimensional facilitated diffusion and random molecular collisions in order to locate its target site (von Hippel and Berg 1989, Riggs et al. 1970). To understand how this is possible, numerous models have been proposed (Gorman and Greene 2008).

Researchers such as Kampmann (2004) suggest that a mode of 3-Dimensional diffusion occurs when a DNA-binding protein diffuses through a 3-Dimensional space before binding to a non-specific part of the DNA sequence (Kampmann 2004, Von Hippel 1994, Berg et al. 1981). This implies that there are mechanisms *in situ* to prevent the protein from unnecessary target searching along a stretch of DNA that is target free (Halford and Marko 2004). This mode of diffusion was also confirmed by Wang et al (2006) on LacI-DNA binding and showed that once bound, the protein will diffuse for a period of time before dissociating from the DNA (Wang et al. 2006, Riggs et al. 1970).

Following on from the Riggs et al (1970) study, Gorski et al (2006) also suggested that proteins, such as transcription factors, can rapidly traverse throughout a nucleus within a few seconds whilst utilising a random 3-Dimensional scanning mechanism before rebinding to a specific or non-specific site (Berg et al. 1981, Hager et al. 2009, Riggs et al. 1970, Gorski et al. 2006). It was also suggested that once bound to DNA, the protein will also search for its specific site via sliding and hopping mechanisms within a localised region (Gorman et al. 2010).

In 2010, Gorman et al proposed that once a DNA binding protein bound onto the DNA non-specifically, there must be a mechanism used to facilitate its mode of diffusion while locating its target site (Elf et al. 2007, Gorman et al. 2010, Hughes et al. 2013). Upon binding, the proteins use 1-Dimensional diffusion mechanisms to scan the DNA for their target sites (Wang et al. 2006, Li et al. 2009). It has been proposed that there are different classifications of 1-Dimensional diffusion. These include hopping and sliding (Gorman and Greene 2008, Kampmann 2005, Li et al. 2009). A hopping mechanism has been defined as a series of microscopic association and dissociation events. The protein will re-associate itself with the DNA at the same site where it dissociated, or it will re-associate at a site near by (Kampmann 2004). Proteins employing a 1-Dimensional sliding mechanism follow the helical path defined by the DNA phosphate backbone or grooves (Kampmann 2005, Gorman and Greene 2008). This type of diffusion results in the protein taking a random walk, meaning it will continuously move in any direction along the DNA without dissociating (Gorman and Greene 2008, Berg 1993). Other modes of diffusion include jumping where a protein also uses a series of association and

dissociation events and this occurs over a greater distance. When a protein jumps, it will rebind at a site distal to its current location (Gorman and Greene 2008). Direct intersegment transfer is another mode of diffusion. This refers to DNA bound proteins transferring from one non-specific DNA site to another without dissociating and occurs when the protein is simultaneously bound to both sites (Halford and Marko 2004). It is thought for this type of diffusion to occur that a high concentration of proteins must be present within the vicinity or they can access two sites by folding the DNA (Kampmann 2005, von Hippel and Berg 1989). Gorski et al (2006) suggests when a protein locates and makes contact with its specific binding site, it becomes immobilised for a short time period due to having a high affinity, exerting its biological function (Gorski et al. 2006). The time duration necessary for the exertion of this biological action was unknown (Chiu et al. 1988). However, Elf et al (2007) found that the *lac* repressor bound to its target site with a residence time of <5 milliseconds (Elf et al. 2007).

Berg et al (1981) originally proposed that proteins using 1-Dimensional sliding mechanism while diffusing must form an electrostatic encounter with a neighbouring site before breaking contact with its current position. This enables the net DNA counter ions to be equal (Berg et al. 1981). It is believed that proteins target their desired location using both 1-Dimensional and 3-Dimensional diffusing mechanisms (Li et al. 2009, Mirny et al. 2009a). The final theorised mode of diffusion is direct intersegment transfer. This involves the bound protein moving from one site to another using an intermediate, looped piece of DNA (Gorman and Greene 2008).

A limitation that occurs when studying DNA binding proteins at a single molecule level is that the DNA used does not resemble chromosomal DNA that is present within all eukaryotic cells (Gorman et al. 2010, Brown 2007, Gorman and Greene 2008). Therefore, when proteins commence a 1-Dimensional random walk along the DNA backbone, they are not exposed to the same environment that *in vivo* proteins experience, such as overcoming obstacles that are in their path. This poses the question whether the proposed 1-Dimensional diffusion mechanism is relevant under such physiological diversity? (Hager et al. 2009, Kampmann 2005, Li et al. 2009). *In vivo*, DNA is tightly packed into the nucleus of eukaryotic cells in the form of chromatin. This structure is far more complex than the elongated, linear DNA used *in vitro* as

the DNA organisation in chromatin remains out of reach for the proteins (Bénichou et al. 2011). It is during transcription, as described in section 1.2, that enables the DNA to be more accessible (Berger 2007).

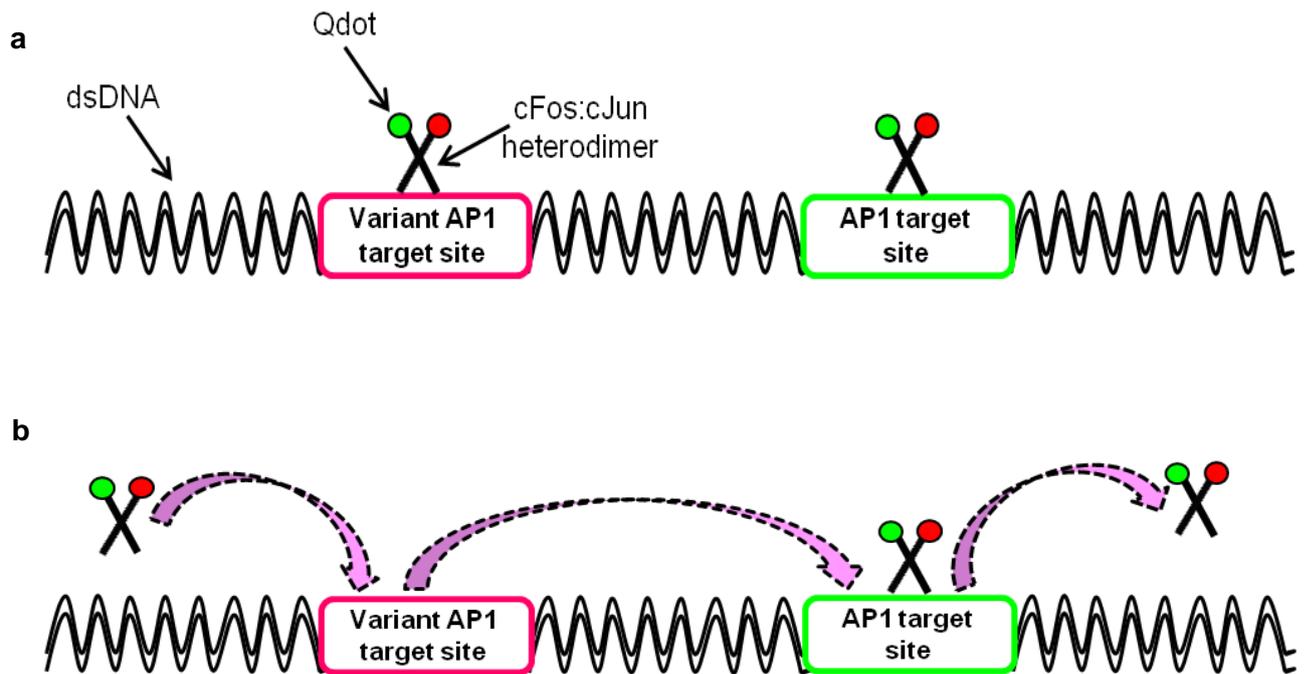
Elf et al (2007) studied how the *lac* repressor located its target sites within living *Escherichia coli* in real time using wide-field fluorescence microscopy to understand how transcription factors locate its *lac* operator sites and mediate transcription. They found that the *lac* repressor could use 3-Dimensional diffusion to translocate through areas of cytoplasm and search sections of DNA using 1-Dimensional diffusion. However ~90% of the time, the *lac* repressor was non-specifically binding and diffusing along the DNA (Elf et al. 2007). This study is useful as it gives an insight as to how transcription factors may be diffusing along chromosomal DNA *in vivo* based on the findings generated by the *lac* repressor and how gene expression may be regulated. The way in which the *lac* repressor associates and dissociates from DNA while using 1-Dimensional and 3-Dimensional diffusion assists with the understanding as to how DNA binding proteins, such as transcription factors, translocate along DNA in search of their target sites.

### **1.8: The “waiting room model”**

This work proposes a “waiting room model” as shown in Figure 9. This model can be defined as target areas of DNA that the transcription factors have a high affinity for in addition to their AP1 target sites, termed variant AP1 target sites. This has described in section 1.6. The transcription factors cFos, cJun and cFos:cJun heterodimers may intermittently bind and pause within these variant target sites until an AP1 target site becomes available, enabling them the ability to exert their biological function at this classic site.

Alternatively, it has been suggested that cFos:cJun heterodimers may be regulating specific genes upon binding to these variant sites (Seldeen et al. 2009). Currently, this remains unknown, however, the research presented within this thesis could lead to the knowledge that these variant AP1 target sites may be important in cellular pathways that controls the transcription of other genes in addition to the ones already controlled by AP1.

One way to prove that this model is feasible would be to perform an EMSA (as described in section 1.5). Oligonucleotides containing each variant AP1 target site could determine the ratio of proteins that bound to site. An alternative method to prove the model would be to use a ChIP-Seq assay (as described in section 1.6). This would enable all of cFos, cJun and cFos:cJun heterodimer AP1 and variant AP1 target sites to be determined on a genomic scale.



**Figure 9:** A cartoon diagram illustrating the “waiting room model.” (a) A transcription factor, in this example a cFos:cJun heterodimer, remains in the variant AP1 target site while the neighbouring AP1 target site is occupied by another transcription factor. (b) Once the AP1 target becomes available, the transcription factor that was waiting in the neighbouring variant AP1 target site is now able to occupy the AP1 target site and exert its biological function. Another transcription factor comes and binds to the now vacant variant AP1 target site.

## 1.9: Aims of this thesis

This study will investigate how the AP1 family members cFos, cJun and cFos:cJun, and the antagonistic peptide inhibitor FosW, interact with DNA tightropes while searching for their AP1 target sites. These protein-DNA interactions will be imaged using OAF microscopy throughout this investigation. The existence of homodimeric and heterodimeric species will also be investigated. These observations are important as they may provide insight into how these transcription factors can initiate cancer when in excess by binding to their AP1 target sites. In addition, studying FosW will provide a firsthand insight into how this inhibitor interacts with cJun on DNA. This insight will become the beginning of a long road towards the development of a cancer treatment when cJun becomes over-expressed.

The aim of Chapter 2 is to use  $\lambda$  DNA tightropes to observe for homodimerisation and heterodimerisation occurring between the transcription factors cFos, cJun, cFos:cJun heterodimers and FosW with cJun. Each transcription factors mode of diffusion, diffusion constant and  $\alpha$ -value will be determined through mean square displacement (MSD) analysis performed on each protein trace presented as a kymograph. To identify where these transcription factors were binding and pausing along  $\lambda$  DNA, the position of all AP1 and variant AP1 targets will be identified Gaussian distribution plots and a Binding Site Position Finder. This chapter shows that cFos is capable of homodimerising, binding to DNA and moving in search of its AP1 target sites like cJun and cFos:cJun heterodimers, and all of the complexes pause within variant AP1 target sites in addition to AP1. Partner swapping within the AP1 family does occur and FosW colocalises with cJun.

In chapter 3, the aim was to study how cFos, cJun and cFos:cJun heterodimers interact with different DNA substrates such as AP1 target free pUC19, a modified form of pUC19, pUCap1, where a single AP1 target site has been introduced, and Target Free  $\lambda$  (TF $\lambda$ ) DNA where no AP1 or variant AP1 target sites are present. Each proteins mode of diffusion, diffusion constant,  $\alpha$ -value and pause location were also determined as described above. This chapter shows that cFos, cJun and cFos:cJun are capable of binding and moving along these DNA substrates in addition to  $\lambda$  DNA. Binding to variant AP1 target sites in addition to AP1 does occur. Minimal

movement of cJun was also observed when bound to TF $\lambda$ . Chapter 4 summarises the major conclusions from this study and goes on to outline a number of possible future experiments. To summarise, in this thesis, cFos can homodimerise, bind and diffuse along  $\lambda$ , pUC19 and pUCap1 DNA substrates. cFos, cJun and cFos:cJun heterodimers can pause at sites other than AP1 and the antagonistic peptide inhibitor FosW is capable of interfering with cJun homodimers.

## **Chapter 2: Imaging cFos, cJun, cFos:cJun and FosW interactions using single molecule techniques**

This chapter aims to study how cFos, cJun, cFos:cJun heterodimers and FosW+/- cJun interact with  $\lambda$  DNA using OAF microscopy. From the proteins kymographic trace, the mode of diffusion, diffusion constant and  $\alpha$ -value will be determined using MSD analysis. Any partner swapping within the AP1 family to form homodimers or heterodimers will also be identified based on the combination of protein dimers.

Another aim in this chapter is to determine where the proteins are binding and pausing along the DNA and whether they are pausing directly within their AP1 target sites only or whether pausing occurs within the variant AP1 target sites as well. This will be determined using Gaussian distribution plots and the Binding Site Position Finder.

### **2.1: Introduction**

The transcription factors cFos and cJun have been identified as modulating physiological functions which include the initiation of cellular differentiation, proliferation and apoptosis. This happens through the regulation of gene expression at the final stages of the MAP kinase cellular signalling cascade (Jochum et al. 2001, Brown 2007, Shaulian and Karin 2001). When these transcription factors bind their AP1 target sequence TGACTCA (Eferl and Wagner 2003, Nakabeppu and Nathans 1989, Angel and Karin 1991), they are able to initiate certain biological functions that include cellular differentiation, proliferation and apoptosis (Campbell 2005, Shaulian and Karin 2001, Jochum et al. 2001). However, when the p53 gene becomes antagonised downstream, any tumour cells become protected against apoptosis due to the disruption of the cell cycle (Weinberg 2007, Latchman 1995) and cancer develops as a result of over-expressed cFos and cJun (Zhang et al. 2007): as described in Chapter 1. cFos causes osteosarcomas and cJun initiates hepatocellular carcinoma and squamous cell carcinoma (Wang et al. 1995, Zhang et al. 2006, Leaner et al. 2003). The mechanism behind tumour suppression is currently unknown (Wagner 2001). However, while this is essential information, it is important to quantify the amount of homodimeric and heterodimeric complexes and

understand how these transcription factors locate these AP1 target sites that are situated along the DNA backbone (Jochum et al. 2001, Woolfson and Alber 1995).

As mentioned in Chapter 1, cJun can homodimerise and heterodimerise with cFos while cFos can only heterodimerise with cJun (Hai and Curran 1991, Ito et al. 2001). As a brief structural overview, both native cFos and cJun are  $\alpha$ -helical coiled coils that are characterised by a repeating pattern of 7 amino acids that are hydrophobic and hydrophilic forming a heptad repeat (Woolfson and Alber 1995, Landschulz et al. 1988). These  $\alpha$ -helical coiled coils have a transactivation domain at the N terminus that is situated at one end of the protein. A basic region located above the transactivation domain contains a DNA binding region that is enriched with arginine and lysine enabling electrostatic interactions with the DNA. Near to the C-terminus at the other end of the protein is a leucine zipper region. Dimerisation occurs within the leucine zipper via the conserved leucine heptad repeat on the dimerisation interface (Woolfson and Alber 1995). It was therefore believed that a cFos:cJun heterodimer forms a more stable complex than a cJun homodimer, although the strength of this interaction currently remains unknown (Mason et al. 2006, Nakabeppu et al. 1988). Due to the continuous interaction between hydrophobic residues inside of the leucine zipper region, it is thought that tightly bound homodimers and heterodimers form in a left-handed, supercoiled conformation (Burkhard et al. 2000) as described in Chapter 1. In this chapter, when cFos was pre-mixed with cJun, red labelled cFos, green labelled cJun and dual labelled cFos:cJun complexes were observed bound to  $\lambda$  DNA tightropes. This is direct evidence that partner swapping within the AP1 family does occur. When cFos was labelled with 2 colours, as described in section 2.2.9, this confirmed that cFos is capable of homodimerising and binding to DNA. However, how this homodimerisation occurs is currently unknown.

When cJun causes tumorigenesis due to over-expression, the antagonistic peptide inhibitor Fos Winner (FosW) was designed to disrupt cJun homodimerisation with high specificity, rendering cJun inactive (Mason 2009, Worrall and Mason 2011). As a member of a peptide library containing variations of native cFos, FosW contains residues within its dimerisation region that enable the formation of a stable heterodimeric complex with its target cJun. This interaction is

more stable than a cJun homodimer (Mason et al. 2006). The FosW peptide was selected using protein-fragment complementation assays (PCA) as described in section 1.4. One advantage to this potential *in vitro* derived inhibitor is that it is highly specific for its target cJun, especially when cJun becomes implicated in cancer (Mason et al. 2007b).

Once bound to the DNA as either a homodimeric or heterodimeric complex, single molecule techniques have enabled cFos, cJun, cFos:cJun and the peptide inhibitor FosW to be tracked as they bind to the DNA using 3-Dimensional facilitated diffusion to locate part of the DNA for which they have a high affinity for. Once bound, 1-Dimensional diffusion was seen to be used to locate their target sites as described in Chapter 1. This suggests that the proteins have evolved to prevent excessive DNA searching. However, while travelling along the DNA, the transcription factors were observed non-specifically pausing. This suggests that there are other areas along the DNA that they had a high affinity for in addition to AP1. In this study, where the proteins are non-specifically binding in addition to their AP1 target sites was investigated to determine whether there were any sites in addition to AP1.

To understand where cFos:cJun heterodimers were binding on the DNA in addition to their known AP1 target sites, Seldeen et al (2009) performed a study where they substituted single nucleotides throughout the entire AP1 target site sequence. For example, TGACTCA became AGACTCA, CGACTCA and GGACTCA (Seldeen et al. 2009). All of these variant sites, known here as variant AP1 target sites, were integrated into this study. Therefore, in addition to the 8 known AP1 target sites along  $\lambda$  DNA, a total of 95 variant AP1 target sites were also located along  $\lambda$ .

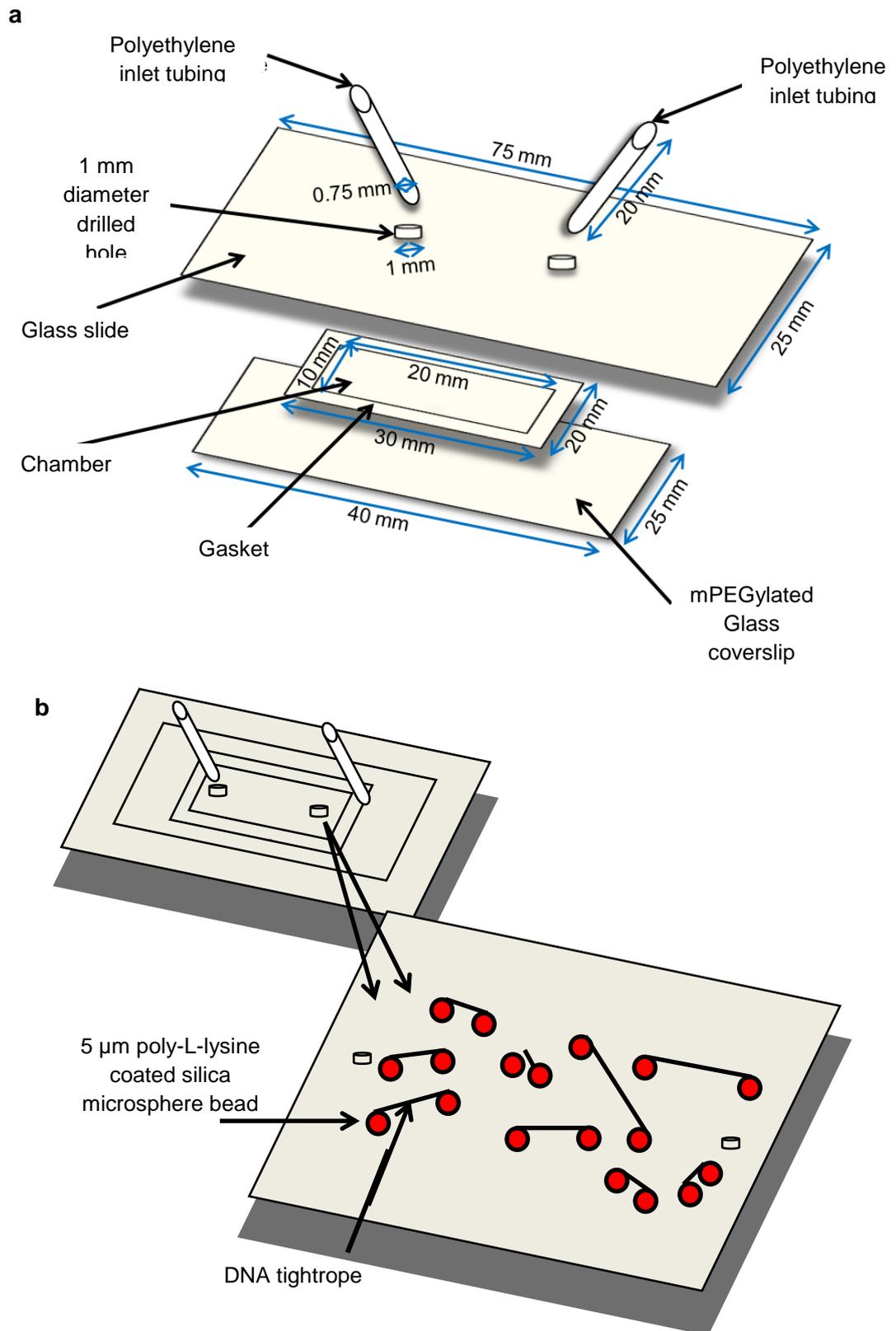
The methods in this chapter were used to try and understand how cFos and cJun transcription factors and the inhibitor FosW interacted with  $\lambda$  DNA tightropes, forming protein-DNA complexes. DNA tightropes enable the proteins to freely associate, dissociate and diffuse along the DNA without any obstructions from artefacts adhered to the surface of the slide. By performing an MSD analysis on the data obtained from each diffusing proteins kymograph, the mode of diffusion, diffusion constant and  $\alpha$ -value for each protein could be determined. As the

proteins diffused along the DNA, they also paused. The Binding Site Position Finder is an analytical method used to identify whether the proteins paused in their AP1 target sites only, or whether they were also pausing in variant AP1 target sites. This rationale is to try and understand whether non-specific binding within these variant AP1 target sites is a mechanism used by the proteins to control transcription or whether the AP1 transcription factors may be exerting their biological function within these variant sites in addition to AP1.

## Chapter 2.2: Materials and methods

### 2.2.1: Flow cell construction

To set up a DNA tightrope assay, a custom built flow cell was designed (Kad, 2010) as shown in Figure 10. This design enables microsphere beads, DNA, buffers and proteins to be introduced into the chamber. Applying methods 2.2.7, 2.2.8 and 2.2.9, DNA tightropes were constructed and DNA-protein interactions can be visualised. Firstly, two 1 mm holes were drilled 12 mm apart on a 75 mm x 25 mm glass slide (Thermo Scientific) using a Dremel electric hand drill with a diamond coated drill tip. The slide was cleaned and silanised as described in method 2.2.3. Next, polyethylene tubes (0.75 mm inner diameter and 1.22 mm outer diameter, (Smiths Medical) 20 mm long were placed in each hole and expanded using a heat gun. This expanded the tubing's end enabling a tight seal. These were fixed to each hole with UV glue (Norland optical adhesive 320-400 nm curing, Thorlabs) that cured under a 365 nm UV lamp (Model UVGL-58 365 nm. UVP) for approximately 30 minutes. Excess tubing and glue on the underside of the slide were removed using a scalpel, ensuring a smooth surface. A double sided adhesive gasket made from transparent polyester coated with a pressure sensitive modified acrylic adhesive, 180 µm deep, was applied to the underside of the glass side, ensuring that each hole was within the gasket. A watertight seal was created by affixing a PEGylated silica coverslip to the underside of the slide. As described in methods 2.2.2 and 2.2.3, these coverslips were pre-treated with ABT (1xABC buffer (50 mM KCl, 50 mM Tris-HCl (pH 7.5), 5 mM MgCl<sub>2</sub>), MilliQ ultrapure H<sub>2</sub>O (resistivity at 18.2 MΩ.cm), 1mg/mL bovine serum albumin (BSA), 0.001% (v/v) Tween<sub>20</sub>) buffer to reduce interactions occurring with the beads and proteins (Lei and MacDonald 2008). After construction, the flow cell was left blocking overnight in ABT buffer. It is the BSA component that reduces non-specific hydrophobic binding to the flow cell surfaces (Kad et al. 2010).



**Figure 10:** (a) A cartoon diagram showing the different components of a flow cell. Affixing a gasket and coverslip to the slide creates a watertight chamber. The tubing enabled buffers, DNA and proteins to be introduced into the chamber. The flow cell was then attached to an adapted microcentrifuge tube (contains 2 vertical holes enabling the addition of solutions) and a pump to create DNA tightropes. (b) A cartoon of a flow cell chamber. As DNA is pump in, the DNA unravels upon sticking to the silica beads (red), forming multiple DNA tightropes (black) of varying lengths.

### **2.2.2: PEGylation of silica coverslips**

Coverslips were PEGylated to prevent proteins from interacting with the surface. During imaging, this minimised any background interference.

Firstly, to remove any contaminants from the glass coverslips (Agar Scientific 25 mm x 40 mm), were sonicated for 30 minutes in 100% ethanol using a sonication bath. The Ethanol was then replaced with 1 M KOH and sonicated for a further 30 minutes. These 2 wash steps were repeated to ensure the coverslips were clean. Before being sonicated in acetone, the coverslips were dried using Nitrogen gas. The clean coverslips were gently agitated in a 3-aminopropyltriethoxysilane (Sigma Aldrich) solution for 2 minutes. The reaction was quenched with water prior to drying in an oven for 30 minutes at 100 °C. 100 mg/ml methoxypolyethylene glycol 5,000 propionic acid N-succinimidyl ester (mPEG) (Sigma Aldrich) was mixed with 100 mM NaHCO<sub>3</sub> at pH 8.2. 10 µL of mPEG mixture was pipetted onto the newly silanised coverslips. 2 coverslips were gently sandwiched together. This ensured that the mPEG mixture was evenly distributed over the 2 surfaces. The coverslips were left for 3 hours at room temperature (23 °C) before being carefully separated, washed with water and dried.

The coverslips were then ready to be attached to a flow cell. If they weren't used immediately, they were stored in a humidior at room temperature for up to a maximum of 1 week. After a week, it was generally found that the high fluorescence was of poor quality and images showed degradation due to a high noise to signal ratio caused by the surface of the slide becoming excessively covered with artefacts.

### **2.2.3: PEGylation of silica coverslips *in situ***

Several adaptations of the above PEGylation protocol were tested throughout the duration of this study. One modified method was found that significantly reduced the background signal.

The coverslips (Agar 25 mm x 40 mm) were sonicated in an ultrasonic bath for 30 minutes in 100% ethanol. The ethanol was replaced with 1 M KOH prior to sonication again for 30 minutes. These 2 wash steps were repeated to ensure that the coverslips were clean. The coverslips were thoroughly rinsed the slides with 100% acetone to remove residual KOH. The coverslips were sonicated in an ultrasonic bath for 10 minutes in 100% acetone. After sonication, the

slides were submerged and gently agitated in 2% (v/v) aminopropyltriethoxysilane in acetone for 2 minutes. The coverslips were thoroughly washed with water and dried using Nitrogen gas. To ensure the coverslips were completely dry, they were incubated at 110°C for >2 hours.

The coverslips were then ready to be attached to the flow cells, forming a watertight seal using a gasket that was die-cut to size using computer software and a die-cutting machine. The newly constructed flow cell and attached tubing were incubated at 4°C overnight with 25 mg/ml mPEG *in situ*. The flow cell was flushed through with water and blocked overnight with ABT buffer from inlet to outlet tube. These were incubated at 4°C until use. This method was more economical than the previous PEGylation method as it enabled more flow cells to be generated within a week and improved the signal to noise ratio.

#### **2.2.4: Silica microspheres coated with poly-L-lysine**

Silica microsphere (beads) (PolySciencesLtd. 2011) are commercially available, are stable and have non-specific interactions with proteins and DNA (Lauer and Nolan 2002). 5 µm beads were used. The beads were incubated at 4 °C in water and would settle to the bottom of the container when not in use. To resuspend the beads, they were vortexed for 1 minute. 100 µL of beads were immediately added to 500 µL of water. The beads were pelleted using centrifugation for 2 minutes at 13,300 xg. The supernatant was removed and the beads were resuspended in 400 µL of 350 µg/mL poly-L-lysine hydrobromide (stock: 5 mg/mL added to 372 µL MilliQ H<sub>2</sub>O. Stored at -20 °C) (Sigma Aldrich). Poly-L-lysine enables the beads to adhere to the flow cell surface as well as the DNA to the beads. The beads were stored at 4°C for >1 hour prior to use. Before introducing the poly-L-lysine coated beads into the flow cells chamber, they were resuspended by vortexing. 10 µL of resuspension was added to 500 µL of water and centrifuged for 2 minutes at 13,300 xg. The supernatant was removed, replaced with another 500 µL of water and recentrifuged for 2 minutes at 13,300 xg. This ensured that the beads were clean and free from contaminants. The pelleted beads were resuspended using a sonicator (cycle 8, amplitude 80%, up50H ultrasonic processor) 4 times using 1 second bursts and immediately added to the flow cell to randomly adhere on the chamber surface.

Before DNA was introduced, the flow cell was flushed with 100  $\mu$ L of water. This was to ensure that there were no bead clumps within the inlet and outlet tubes that could cause blockages. The density of the beads within the chamber was verified using brightfield microscopy (Nikon Diaphot).

There is one disadvantage to using Poly-L-lysine coated silica microspheres. The beads fluoresce due to YOYO-1 intercalated DNA also adhering to their sticky surface, they appear as fluorescent white objects either side of a tightrope. This meant that any Qdot conjugated proteins that bound at the ends of a tightrope could not be analysed accurately. However, one advantage is that these beads do offer a quick, yet reliable, DNA binding tool.

## **2.2.5: Buffers**

### **2.2.5.1: ABC buffer**

When introducing DNA tightropes into the chamber, a buffer containing 50 mM KCl, 50 mM Tris-HCl (pH 7.5), 5 mM MgCl<sub>2</sub> was used. This shall be referred to as 1x ABC buffer. When YOYO-1 was introduced into the chamber to label the DNA, this was present in 50 mM KCl, 20 mM MgCl<sub>2</sub>, 100 mM Tris-HCl (pH 7.5), H<sub>2</sub>O and 100 mM DTT. This will be referred to as low salt buffer. This buffer was modified by increasing the KCl molarity to 150 mM, referred to as high salt buffer, mimicking intracellular conditions in vivo (Khatter et al. 2014). 100 mM DTT was an essential component to this buffer as it reduced the Qdots blinking by encapsulating the Qdots, resulting in increased photostability during long exposure to fluorescence illumination (Antelman et al. 2009). DTT also limited the amount of YOYO-1 photo-bleaching.

## **2.2.6: Bacteriophage Lambda methylated and non-methylated DNA tightropes**

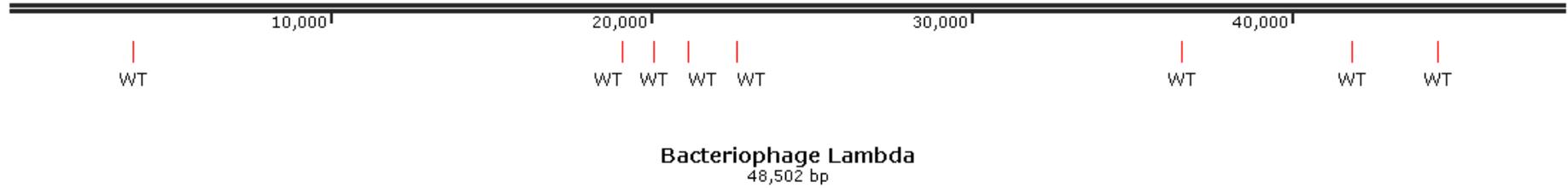
### **2.2.6.1: Bacteriophage Lambda DNA**

Bacteriophage Lambda DNA ( $\lambda$  DNA. New England BioLabs), as shown in Figure 11, is linear, double stranded, 48,502 bp in length and methylated.  $\lambda$  DNA contains the methylation sites Dam, Dcm and EcoKI (all methylated nucleotides have been underlined). The Dam gene encodes methylase to transfer bulky methyl groups to the adenine residues N6 position at the recognition sequence GATC. The Dcm gene encodes methylase to methylate cytosine residues C5 position at the recognition sequences CCAGG and CCTGG (Ringquist and Smith 1992). The

EcoKI gene methylates adenine at position N6 in the recognition sequences AACGTGC and GCACGTT (Powell et al. 1998).

Bacteriophage Lambda DNA ( $\lambda$  DNA, New England BioLabs) is linear, double stranded, 48,502 bp in length and contains an overhang of 12 bases either end enabling self annealing to take place between neighbouring strands.  $\lambda$  DNA is methylated from the *E. coli* host strains methylase enzymes, however which Dam and Dcm sites were methylated is unknown. Non-methylated  $\lambda$  DNA contains Dam sites that are only 50% methylated (Palmer and Marinus 1994). It was found that a DNA concentration of 500 ng/ $\mu$ L provided a sufficient amount of reliable tightropes (Kad et al. 2010, Hughes et al. 2013). If necessary, these pre-formed tightropes could be stored overnight at room temperature.

To reduce the number of freeze-thaw cycles during the storage of  $\lambda$  DNA, 50  $\mu$ L aliquots were prepared and stored at -20 °C in microcentrifuge tubes. If an aliquot was used for experiments, this was stored at 4°C. Over time it was found that the length of  $\lambda$  DNA increased due to self annealing and this was observed during imaging. This resulted in longer tightropes being formed. More care was taken as long DNA is more prone to shearing. To help prevent this from occurring, the DNA was never re-frozen and wider pipette tips were used. These were made by cutting the narrow end off of a 1000  $\mu$ L pipette tip.

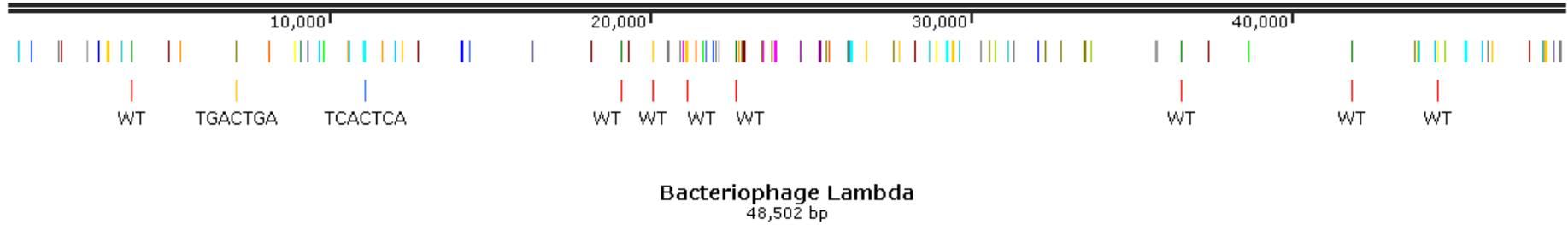


**Figure 11:** A linear map representing a single piece of double stranded  $\lambda$  DNA (made using SnapGene) that is 48,502 bp in length. AP1 'Wild type' (WT) target sites and their locations are indicated in red. There are 4 AP1 target sites that are within close proximity to each other. These are located between positions 19,119 bp – 22,688 bp.

### 2.2.6.2: Lambda DNA with additional variant AP1 target sites

As a protein travels along the DNA, it occasionally paused. Therefore, to try and understand where, the  $\lambda$  DNA sequence was reviewed to see whether there were any other alternative binding sites.

One study by Seldeen et al (2009) made single nucleotide substitutions (underlined) to each nucleotide base within the AP1 target site sequence. The following sequences were located along a single piece of  $\lambda$  DNA (as shown in Figure 12): AGACTCA, CGACTCA, GGACTCA, TAACTCA, TCACTCA, TTACTCA, TGCCTCA, TGGCTCA, TGICTCA, TGAATCA, TGAGTCA (palindrome sequence to AP1 target site), TGAITCA, TGACACA, TGACCCA, TGACGCA, TGACTAA, TGACTGA, TGACTIA, TGACTCC, TGACTCG and TGACTCI. In this study, these nucleotide substitutions shall be referred to as variant AP1 target sites. There were 95 variant AP1 target sites in total.



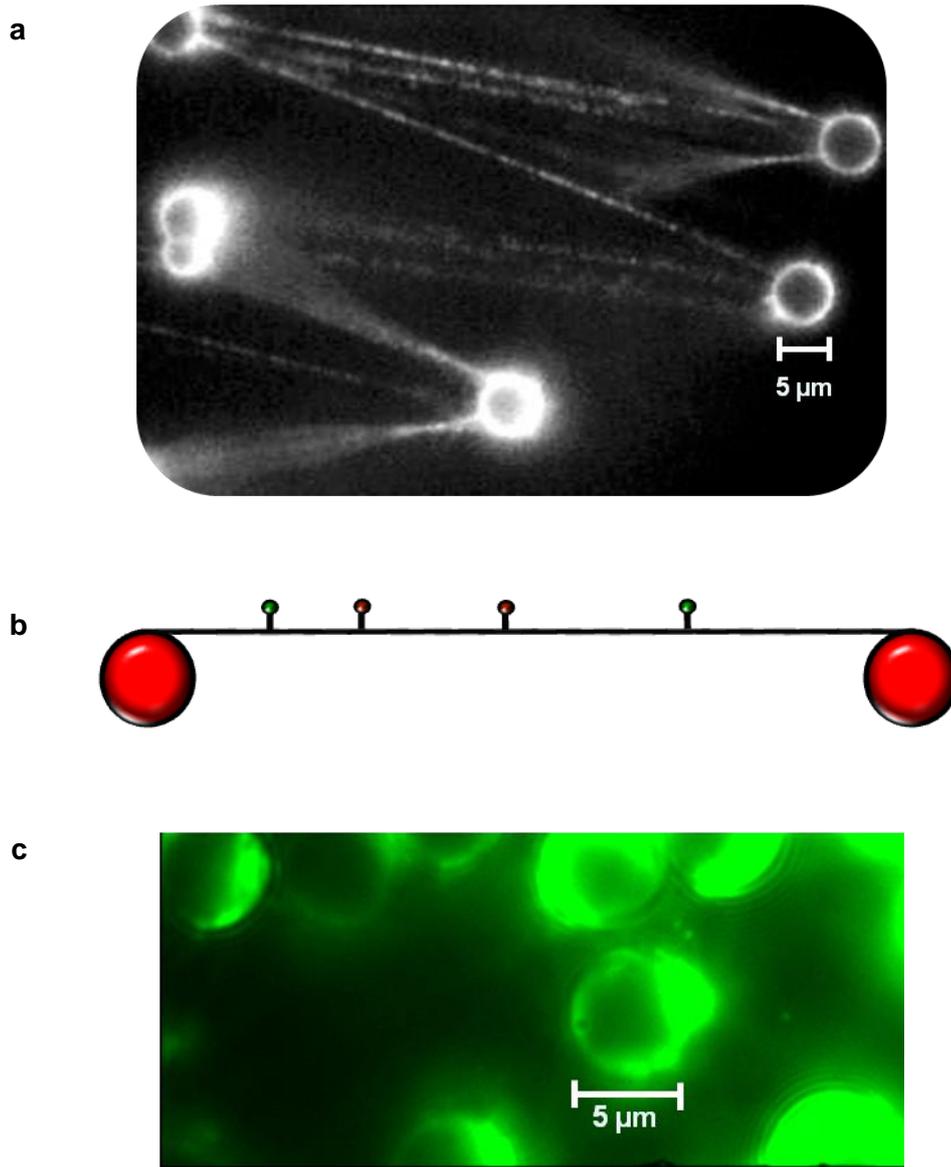
**Figure 12:** A diagram representing a single piece of  $\lambda$  DNA (created using SnapGene). In addition to the 8 known wild type (WT) AP1 target sites, there are also 95 variant AP1 target sites. Each target sites sequence has been labelled with the same colour. This diagram shows that  $\lambda$  DNA is overcrowded with WT and variant AP1 target sites.

### 2.2.7: Creating tigtropes

Beads were introduced into a flow cell chamber via the inlet tube. Prior to the addition of DNA, the flow cell outlet tube was connected to a pump (AL-1000 programmable syringe pump. World precision instruments) and a 10 mL plastic syringe (BP Plastipak) containing 500  $\mu\text{L}$  of 1x ABC buffer. Attached to the flow cell inlet tube was an adapted microcentrifuge tube. 1x ABC buffer was flowed through the flow cell and into the microcentrifuge tube. 500  $\mu\text{L}$  of 1x ABC buffer was withdrawn back through the flow cell towards the pump. This ensured that any contaminants within the flow cell were removed.

500 ng of DNA (1  $\mu\text{L}$  DNA in 99  $\mu\text{L}$  1x ABC buffer) was introduced into the chamber. Using the pumps bi-directional flow setting at 300  $\mu\text{L}$  per minute and switching every 100  $\mu\text{L}$ , tigtropes were formed by one end of the DNA attaching to a 'sticky' poly-L-lysine coated bead, unravelling and attaching to a neighbouring bead. As each bead is 5  $\mu\text{m}$  in diameter, this meant that tigtropes were suspended 5  $\mu\text{m}$  above the slides surface.

Before imaging, the tigtropes were labelled with YOYO-1 (1-20  $\mu\text{L}$  YOYO-1 in water making a 100  $\mu\text{L}$  final volume. YOYO-1 stock at 1  $\mu\text{M}$ ) as shown in Figure 13 image a.



**Figure 13:** Here are some examples of DNA tightropes and Qdot conjugated proteins that have bound onto the DNA. (a) An image of several DNA tightropes that have been labelled with YOYO-1 dye. It is important to identify tightropes before proceeding to the addition of Qdot labelled proteins. (b) A cartoon representing proteins that are conjugated with red and green Qdots and bound to a DNA tightrope. (c) An image of green Qdot labelled cJun proteins that have bound onto a single  $\lambda$  DNA tightrope. Labelled proteins (in high salt buffer containing 150 mM KCl to mimicking intracellular conditions *in vivo*) are added to the chamber after the YOYO-1 dye has been washed out with low salt buffer. Equilibrating the chamber using low salt ensures that the final KCl concentration within the chamber remains at 150 mM when imaging the proteins.

### 2.2.8: Biotinylated proteins

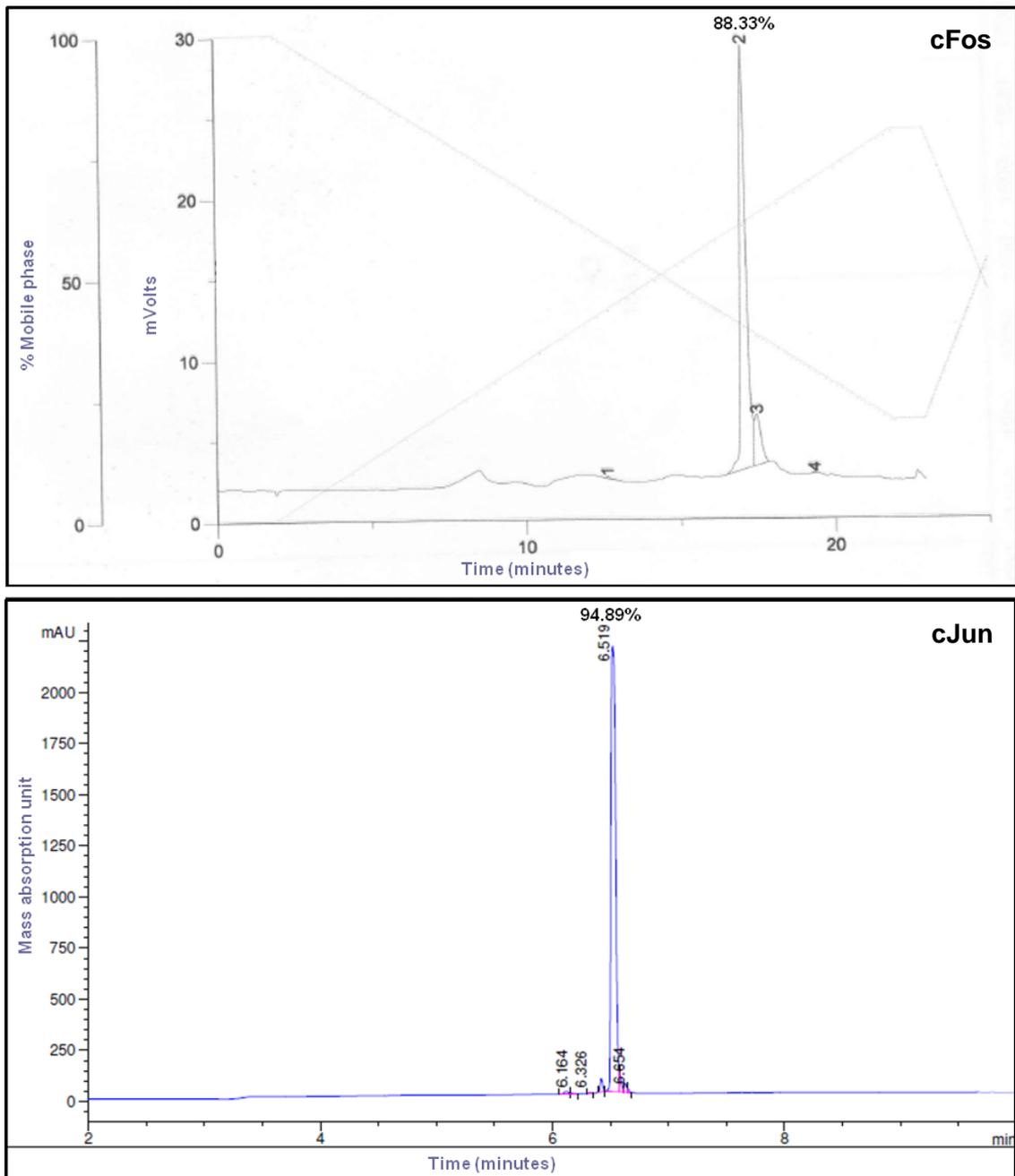
cFos and cJun transcription factors are basic zipper (bZIP) proteins and members of the AP1 family (as mentioned in Chapter 1). They do not have a transactivation domain but do have an additional Biotin tag that is situated at their C-terminus (Figure 4). This enables Qdots to attach through conjugation. The biotinylation tags were away from the proteins dimerisation and DNA binding domains and did not affect the proteins from colocalising, binding to DNA and diffusing as these were observed during imaging.

These are the only differences between the substitute peptides and their native forms. cFos contains 59 residues and is 7252 Da and cJun contains 60 residues and is 7247 Da. Their configurations and sequences can be seen in Figure 5.

Basic-FosW-Biotin (FosW) however is an antagonistic peptide inhibitor that contains 61 residues and is 7643 Da. This was designed to interfere with cJun homodimeric complexes (Mason et al. 2006). FosW was based on the structure of cFos, although its coiled coil region contains different residues surrounding the leucine zipper (LZIP) region and traditionally FosW does not contain a DNA binding region or a Biotin tag (Mason et al. 2006). However, for the purpose of this study, this peptide inhibitor has a DNA binding region that is an exact replica of cJun. This enabled the inhibitor to bind to the DNA and diffuse using 1-Dimensional sliding. Without this DNA binding domain, it would not be possible to visualise FosW colocalising with cJun due to a loss of signal. A Biotin tag at the C terminus also meant a Qdot could be conjugated.

The proteins cFos, cJun and the inhibitor FosW were ordered from the company PeptideSynthetics and were kindly supplied for this study by Dr. Jody Mason.

PeptideSynthetics performed Reverse-Phase High Performance Liquid Chromatography (RP-HPLC) to purify the proteins. cFos has >88% purity and cJun is >94% pure. Figure 14 shows the representative RP-HPLC data for cFos and cJun.



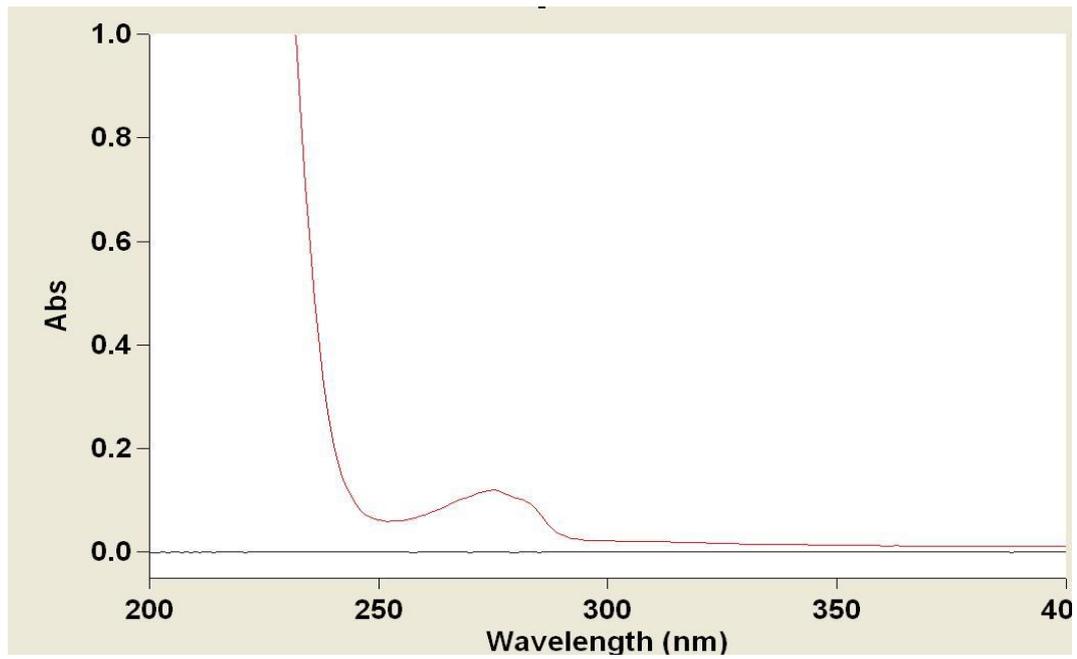
**Figure 14:** Chromatograms representing the purity of cFos and cJun using RP-HPLC. Performed by PeptideSynthetics. cFos is >88% pure and cJun has >94% purity.

A tyrosine residue is present within the LZIP region of cFos, cJun FosW. This has an extinction coefficient of  $1209 \text{ m}^{-1}\text{cm}^{-1}$  and is absorbent at A280 nm wavelength (shown by the example in Figure 15) and is the only residue in the sequence that absorbs at A280 nm. By using the Beer-lambert Law:

$$\mathbf{A = \epsilon cl}$$

**A** is the sample absorbance using spectrophotometry,  **$\epsilon$**  represents the molar extinction coefficient. This is the tyrosine (Y) extinction coefficient ( $1209 \text{ Lmol}^{-1}\text{cm}^{-1}$ ). **l** is the path length of the sample and **c** represents the concentration of the compound in solution ( $\text{mol/L}^{-1}$ ). This formula enabled the peptides molarity to be calculated.

To avoid freeze-thaw cycles that could potentially degrade the peptides, 1  $\mu\text{L}$  aliquots 0.5  $\mu\text{M}$  were stored in sterile microcentrifuge tubes at  $-80 \text{ }^\circ\text{C}$ .



**Figure 15:** An absorbance graph of cJun that was taken using a UV-Visible spectrophotometer (Cary 50). Like cFos and cJun, FosW has a single tyrosine (Y) residue that is present within the peptides LZIP region. The graph clearly shows that light is being absorbed by the Y residue at a wavelength of 280 nm. Similar graphs were observed for cFos and cJun.

### **2.2.9: Labelling the biotinylated peptide**

Prior to labelling the biotinylated peptide with Streptavidin conjugated Qdots (Invitrogen), a single 1  $\mu\text{L}$  aliquot of peptide at a concentration of 0.5  $\mu\text{M}$  was thawed on ice for 30 minutes. The peptide was diluted to a concentration of 100 nM upon the addition of high salt buffer and Qdots. This resulted in a final Qdot concentration of 400 nM. This 1:4 peptide to Qdot ratio increased the probability of each protein being conjugated to a single Qdot. The mixture was left to incubate for 15 minutes on ice. Prior to imaging, an aliquot of labelled peptide was diluted to 0.4- 2.5 nM in high salt buffer, making a final volume of 100  $\mu\text{L}$ . The concentration of labelled peptide imaged was dependent on the amount of protein-DNA binding observed in the previous experiment.

### **2.2.10: Quantum dots**

Proteins bound to DNA were visualised using Quantum dots (Qdots. Invitrogen) during this study.

Qdots are inorganic nanocrystals that are roughly protein-sized atomic clusters. These are composed of a mixture of cadmium and selenium which are semiconductor materials. An additional semiconductor shell made from zinc sulphide improves the optical properties of the material (Chan and Nie 1998). Qdots, unlike organic fluorophore labels, can enable multicolour assays to be visualised simultaneously due to the narrow emission spectrum. This is due to the direct relationship between the size of each Qdot and their emission wavelength (Alivisatos 1996).

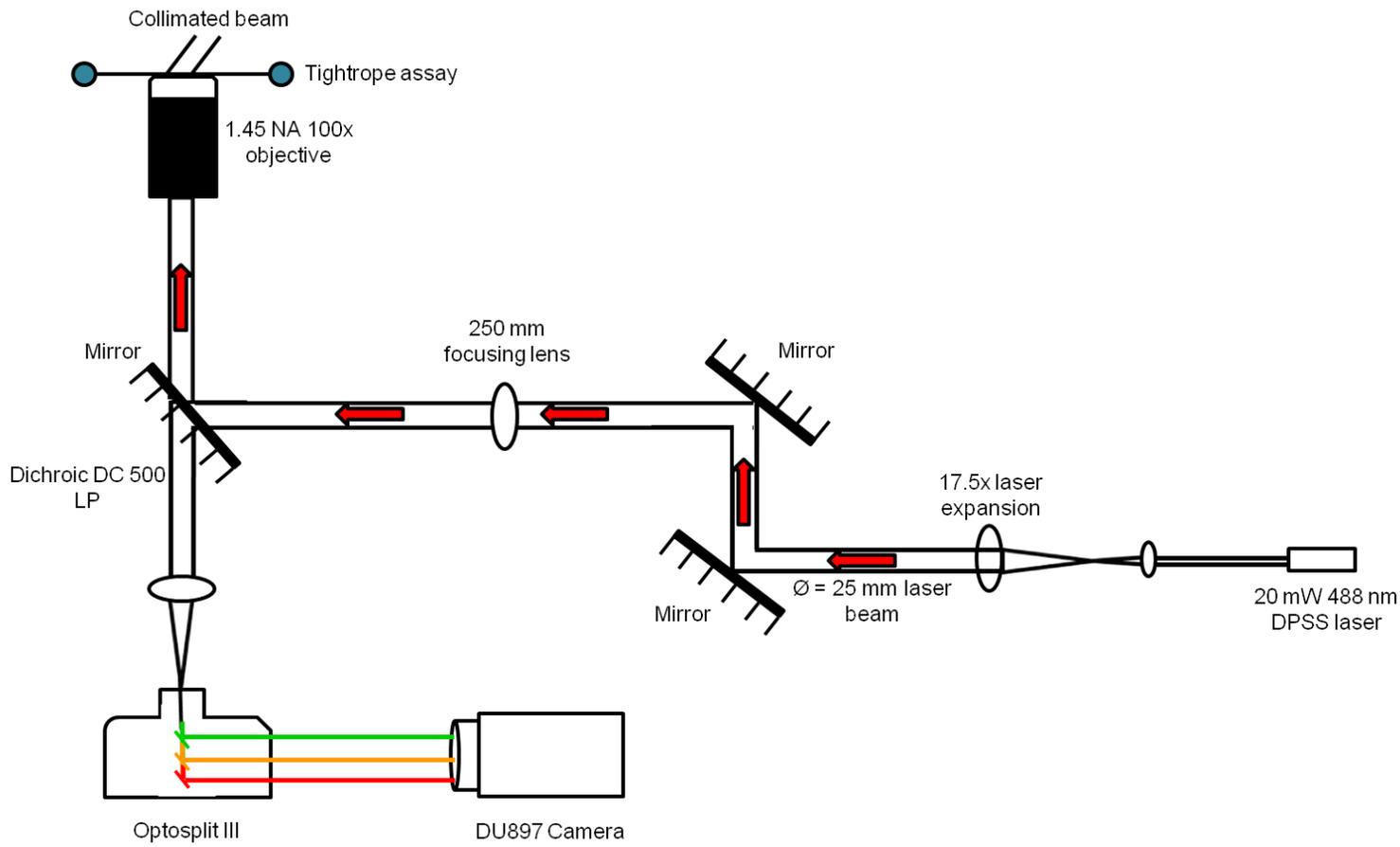
Unlike more traditional organic labels such as the Green Fluorescent Protein (GFP), Qdots are less susceptible to chemical and metabolic degradation, have a greater photostability threshold, enhanced brightness and a narrower emission spectrum (Jaiswal et al. 2002). It is these enhancements that enable individual protein-Qdot conjugates to be tracked over longer periods of time and their fast dynamics to be monitored during single molecule studies (Wang et al. 2008).

### **2.2.11: Imaging using Oblique Angle Fluorescence (OAF) microscopy**

Oblique Angle Fluorescence (OAF) microscopy, a variation of TIRFM was used to image all protein assays during this study. A diagram is shown in Figure 16 illustrating OAF setup.

TIRFM uses an evanescent wave to selectively illuminate and excite fluorophores within a 200 nm area that is directly adjacent to the glass-water interface. An evanescent wave occurs when the incident light beam becomes totally internally reflected at the glass-water interface. This occurs when the light beam encounters the interface at a critical angle of  $63^\circ$ . This refraction occurs due to the high refractive index of the glass medium and the low refractive index of the water/aqueous solution medium. This results in either a portion of the light becoming confined to the solution side of the interface or all of the light being dispensed in the higher index medium (Reichert and Truskey 1990).

OAF microscopy, like TIRFM, uses a collimated beam of light. It is the difference between the 2 mediums refractive indexes and the incident angle that determines whether the light beam becomes refracted as it enters the lower refractive index medium or reflected at the interface (Figure 10). Unlike TIRFM, the OAF light beam is angled away from the critical angle. This means a thin section becomes illuminated and background noise becomes reduced. This strategy also enables tightropes to be visualised at least  $5\ \mu\text{m}$  above the surface. Imaging away from the slides surface will reduce the amount of noise caused by excess fluorescence being emitted from surface artefacts upon excitation by the laser, such as adhered Qdots. This illumination strategy also increases the signal by only illuminating Qdot-bound proteins that are within the lasers path. This will also help to reduce background noise being emitted from nearby fluorophores as they will not be excited by the beam.



**Figure 16:** A diagram showing the optical setup of OAF microscopy. Firstly, a 1.66 mm DPSS laser passes through a 10 mm focusing lens before expanding 17.5 times as it passes through a second 250 mm focusing lens. This determines the size of OAFs illumination spot. On the back of the objective, the laser beam is focused down onto a diffraction limiting spot. The laser beam then passes through the flow cell before travelling back through the system and separated into three 565, 605, 655 nm beams. This happens when the beam hits the first dichroic and passing through 3 separate dichroic filters, located within the Optosplit III system. The images were collected using the DU897 EMCCD camera.

### **2.2.12: Data analysis**

This section will describe how the time lapse recordings taken of Qdot conjugated proteins interacting with DNA using OAF microscopy were analysed. Firstly, kymographs show the direct interaction of the Qdot bound proteins as they interact with the DNA, frame by frame throughout the entire recording. This data meant that the number of motile proteins could be quantified and what proportion of proteins visited their AP1 or variant AP1 target sites. Based on the location of the Qdot within the kymograph, the proteins motion could be ascertained and calculated by using an automated Gaussian fitter program. The error for all observed diffusing proteins, including their  $\alpha$  values and diffusion constants ( $\mu\text{m}^2\text{s}^{-1}$ ), have been recorded using the standard error of the mean (SEM).

### **2.2.13: Statistical tests**

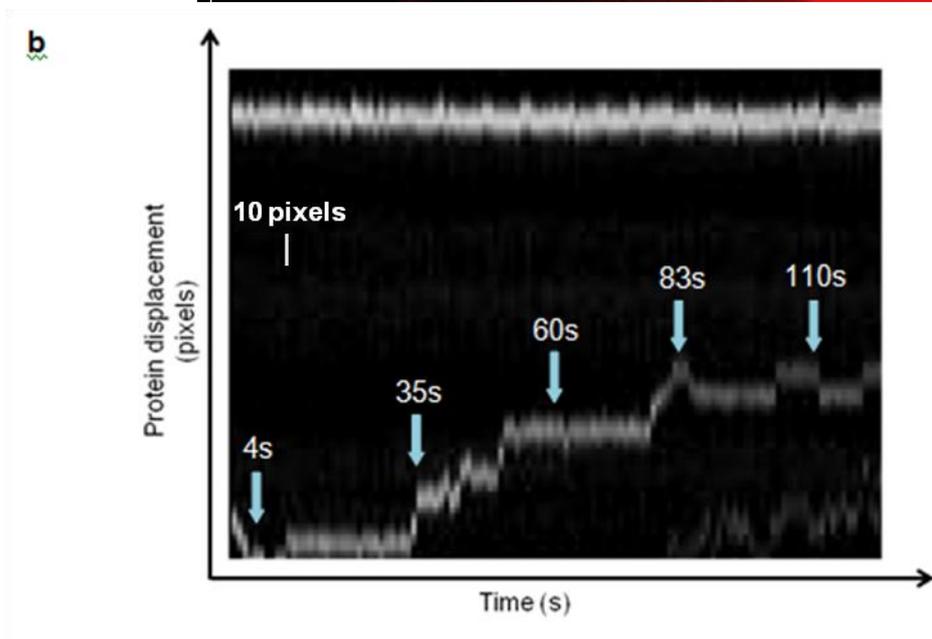
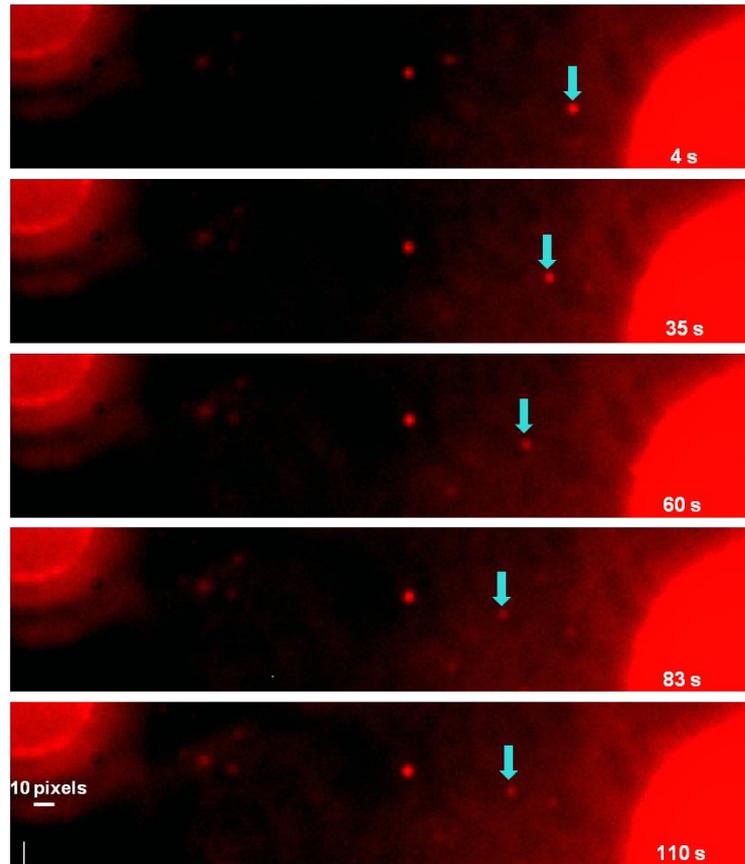
A t-test was used to prove the statistical significance of results when comparing two proteins that bound to DNA and were observed to be diffusing.

### **2.2.14: Kymograph analysis**

A kymograph, produced using ImageJ (NIH, USA), is a frame by frame report of the interaction between a Qdot conjugated protein with its DNA substrate. Kymographs can provide information about how each protein interacts with the DNA and how it changes its position during translocation.

Before creating a kymograph, a tightrope was chosen that was not overcrowded to avoid any interference from adjacent proteins during analysis. Prior to kymograph construction, the tightrope of interest was rotated (using ImageJ tools) so that it was horizontal. A region of interest (ROI) was then drawn around the tightrope. It is within this ROI of the movie that each horizontal line, a single pixel in height, was scanned in single pixel increments down the y-axis, forming a series of time slices. These were vertically stacked according to their original position and the stacked data was rotated by  $90^\circ$  to the right. This simply aided data analysis, meaning that the protein displacement was along the y-axis and time was along the x-axis. Using such a kymograph permitted every pixel per time frame to be visualised throughout the recording. If any Qdot conjugated proteins appeared extremely bright, these were not accounted for as this

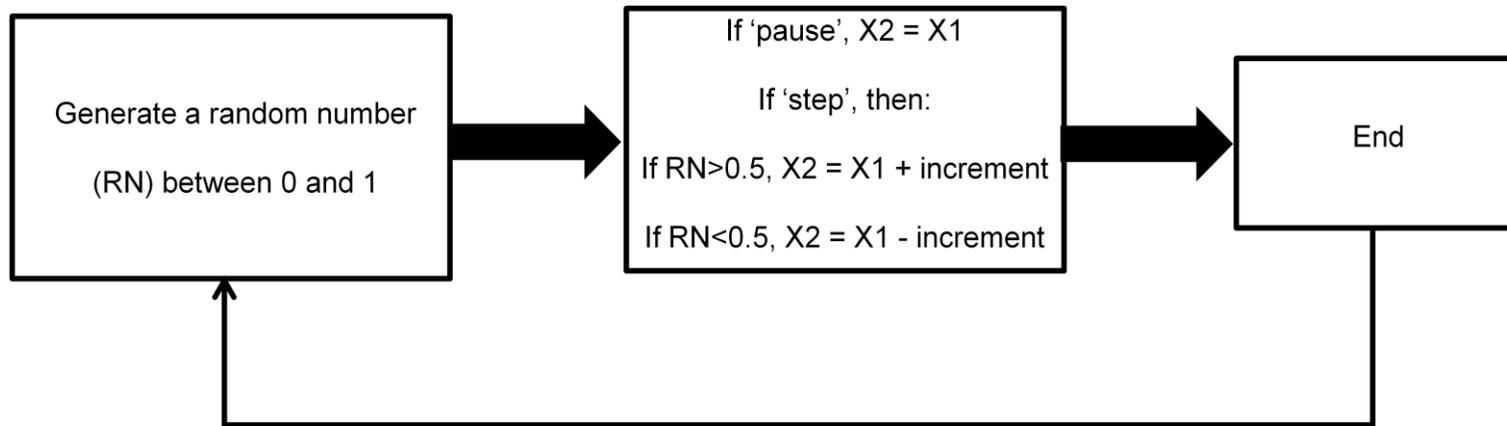
indicated that multiple Qdots may have bound. If a protein was observed to move 2 pixels or more from its previous position per time frame, this protein was considered to have moved (Figure 17).



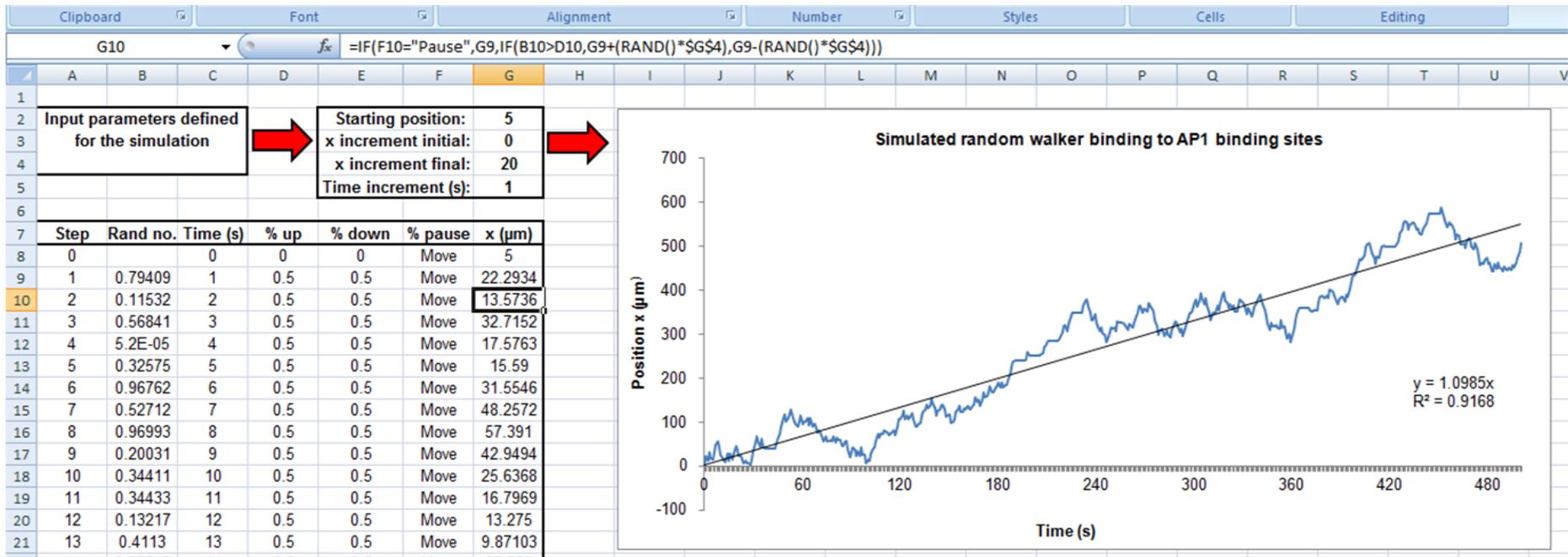
**Figure 17:** (a) Shows 2 cFos molecules that have bound to a DNA tightrope during a 120 second movie. The arrow above one of the more motile protein-DNA complexes shows the position of the protein at various time points in seconds. The orientation of the tightrope has been rotated to a vertical position using imageJ software to assist with the analysis of the kymograph. (b) Shows the same tightrope containing the 2 cFos molecules but displayed in the form of a kymograph using ImageJ. The most motile protein (as indicated by the arrows in (a) at various time intervals) shows its position along the DNA. The X axis represents time (s), that is 1 second per frame, and the Y axis shows the proteins position along the DNA. 10 pixels is 0.819  $\mu\text{m}$  (in height), equivalent to 2410 bp (represented by the scale bar).

### **2.2.15: Random walker simulations with specific and non-specific pausing**

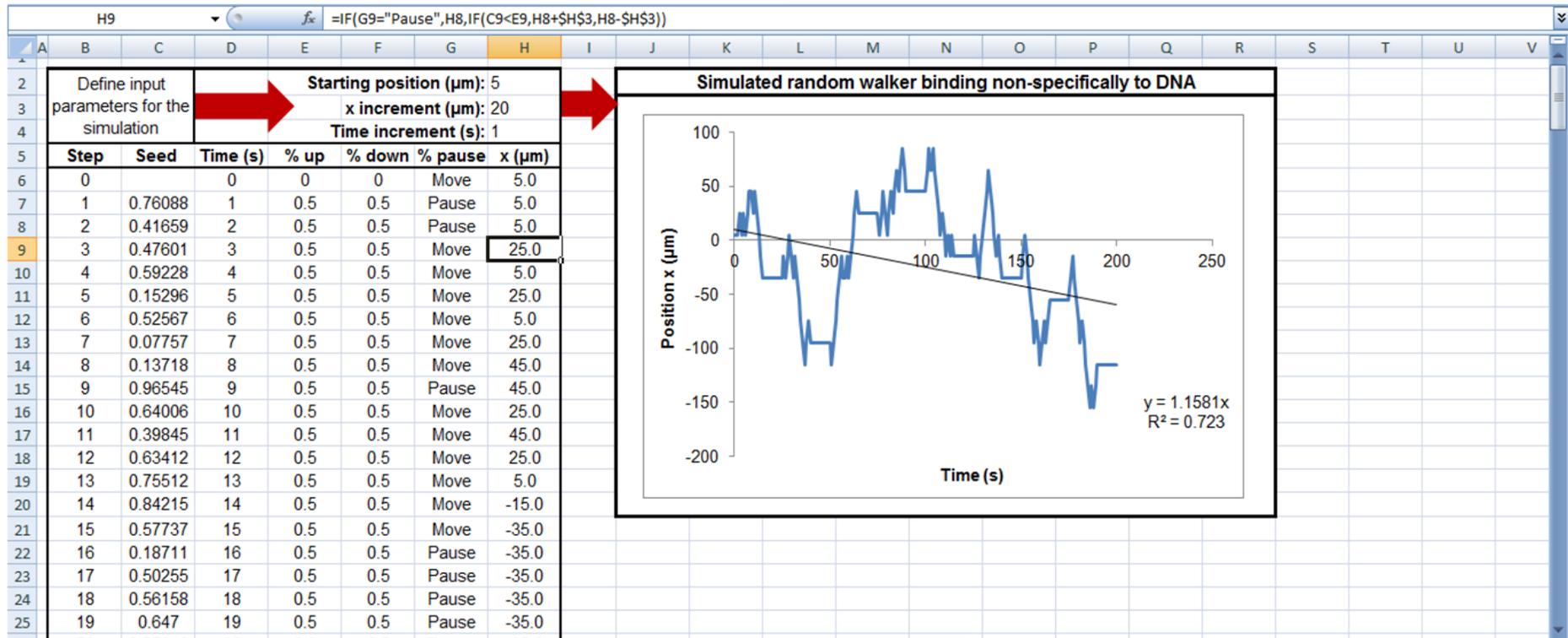
A model was designed to simulate a dimer taking a random walk. This has been described in Figure 18. The model assumes a 50 % probability of a dimer taking a step forwards or backwards during its translocation along  $\lambda$  DNA, mimicking *in vivo* DNA-bound protein movement. Pauses can be manually introduced by the user at any time. Random numbers between 0 and 1 were generated. Within this simulation, if a random number is above 0.5, a forward step will be taken. A backward step occurs with numbers less than 0.5. Depending on the random number, the dimer can either move forwards (new position = previous position + increment) or backwards (new position = previous position - increment). The increment ( $\mu\text{m}$ ) is user defined. Figure 18 shows the simulated dimer taking a random walk and specifically pausing along the DNA. Unlike in Figure 19, Figure 20 shows a simulated dimer taking a random walk along a DNA tightrope while pausing for varying lengths of time, mimicking real data traces.



**Figure 18:** A flow diagram of a dimer taking a random walk. X2 is equal to the dimers next position and X1 represents the proteins previous position. The increment is defined as the step size ( $\mu\text{m}$ ) taken by the dimer per time frame and is user defined.



**Figure 19:** A print screened image showing a model of a simulated dimer taking a random walk as described in method 2.2.15. In this example, the simulated dimer was formed using over 480 data points. It can be seen translocating along a single  $\lambda$  DNA tightrope, taking a step ( $\mu\text{M}$ ) forwards or backwards every second. The dimer also specifically pause at each of the 8 AP1 target sites, remaining static for 5 seconds for the purpose of this demonstrative model, before translocating again in search of the next target site. This simulation serves as a positive control as it depicts a dimers movement along the DNA 'in an ideal world', meaning, the dimer visits each target site without returning to the proceeding site. The model is an exact representation of  $\lambda$  DNA and has been scaled down 100 times.



**Figure 20:** A print screen image showing a model of a simulated dimer taking a random walk as described in method 2.2.15. Unlike the model shown in Figure 19, the user decides when and where the simulated dimer can move and pause along the  $\lambda$  DNA tightrope and for how long. Pausing of varying time durations was implemented to mimic actual data traces.

### 2.2.16: Diffusion constant and $\alpha$ value

Kymographs showing moving proteins were further analysed to identify their diffusion constant ( $\mu\text{m}^2\text{s}^{-1}$ ) and mechanism of motion, as determined by their  $\alpha$  value. An  $\alpha$  value of less than 1 suggests subdiffusion, 1 suggests unbiased diffusion and 2 indicates directed motion (Saxton 1997). This was defined by the slope of a log-log axis after plotting the MSD (as described in section 2.2.16). Firstly, using ImageJ application tools, a box was drawn around a single protein kymograph close to its trace. This helped prevent background fluorescence from being analysed.

An un-biased, customised and automated Gaussian distribution macro was installed into ImageJ. This generated a Gaussian distribution graph per time frame based on the point spread function of each Qdot fluorophore, providing positional accuracy that exceeded the diffraction limit. The peak of the Gaussian distribution determined the spatial accuracy of the system. This was determined to be 8.7 nm using Equation 1 (Hughes et al. 2013, Thompson et al. 2002):

$$\sigma_i = \sqrt{\frac{s^2 - a^2/12}{N} + \frac{4s^3b^2\sqrt{\pi}}{aN^2}}$$

**Equation 1:**

The standard deviation of the Gaussian distribution is represented by  $s$ ,  $a$  represents a pixels length at 82 nm and  $b$  represents the standard deviation of the background noise.  $N$  represents the number of photons that were collected. Brownian motion, background noise and camera noise have been incorporated into the value of uncertainty (Thompson et al. 2002). Throughout this study, the Andor camera used (as described in 2.2.11) has 160000 electrons per pixel count. This is an average conversion ratio. In the Andor camera, the gain is expressed in electrons per count. In this case, this means that the camera produces 1 count for every 160000 electrons.

### 2.2.17: MSD analysis

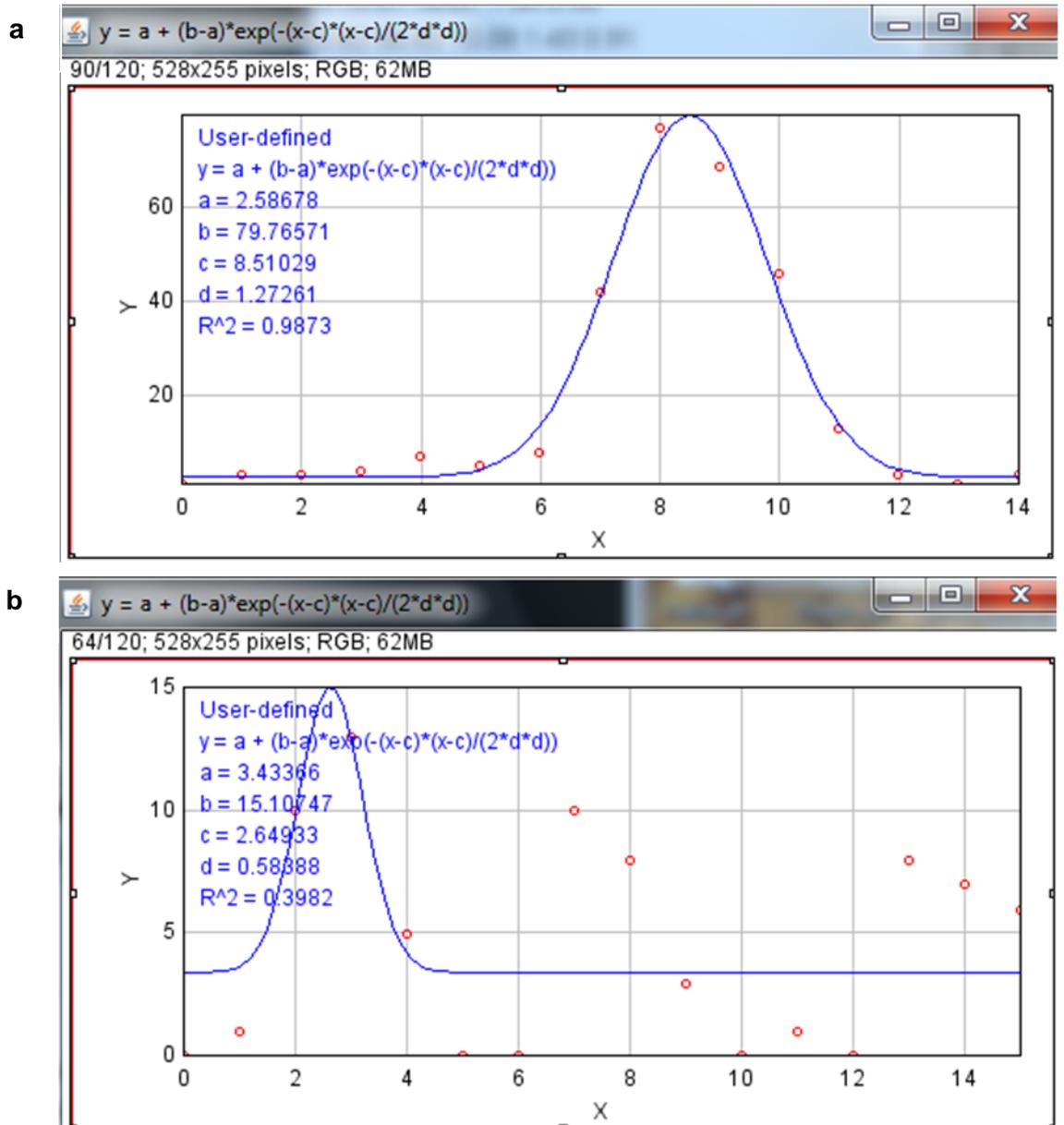
To understand how each protein moves along DNA, its mean-squared displacement (MSD) was determined as shown in Figure 23. To determine the proteins  $\alpha$  value, the MSD was logged per time frame. This can be seen in Figure 24. This methodology utilised the raw data that was

generated from the Gaussian distribution as described in Table 1 and Figures 22 and 23. Non-linearities occurred when the MSD was fitted to a straight line. This typically occurred after the first 10% of the frames, which translates into the first 12 seconds and 12 time steps, assuming the imaging occurred over 120 seconds. The standard deviation, showing the distribution of data, and  $R^2$  values, representing the Gaussian distribution fit, were used together to filter out any Gaussian distributions that did not fit within these parameters. This filtering method excluded data points were marked by an **X**. An example of this can be seen in Table 1.

To establish the standard deviation and  $R^2$  parameters to be used in this dual filter, 50 Gaussian distribution data were analysed and their means taken. It was concluded that any  $R^2$  values  $<0.84$  and any standard deviation values  $>2.32$  should be excluded as these data corresponded to a bad Gaussian distribution fit, as shown by the Gaussian distribution plots in Figure 21. If any values were on these dual filters threshold, they were analysed and introduced into the data set. The corresponding mean values to the Gaussian distributions standard deviations and  $R^2$  values that fitted within the parameters of the dual filter were converted from pixels into micrometers. Equation 2 was used to determine the proteins MSD:

**Equation 2:**

$$MSD(n\Delta t) = \frac{1}{N-1} \sum_{i=1}^{N-n} [(x_{i+n} - x_i)^2 + (y_{i+n} - y_i)^2]$$



**Figure 21:** Printscreen images of Gaussian distribution plots for a cJun protein. (a) A Gaussian distribution plot with a good fit. This can be seen by the  $R^2$  value 0.9. (b) This Gaussian distribution plot has an  $R^2$  value 0.3, illustrating a bad fit.

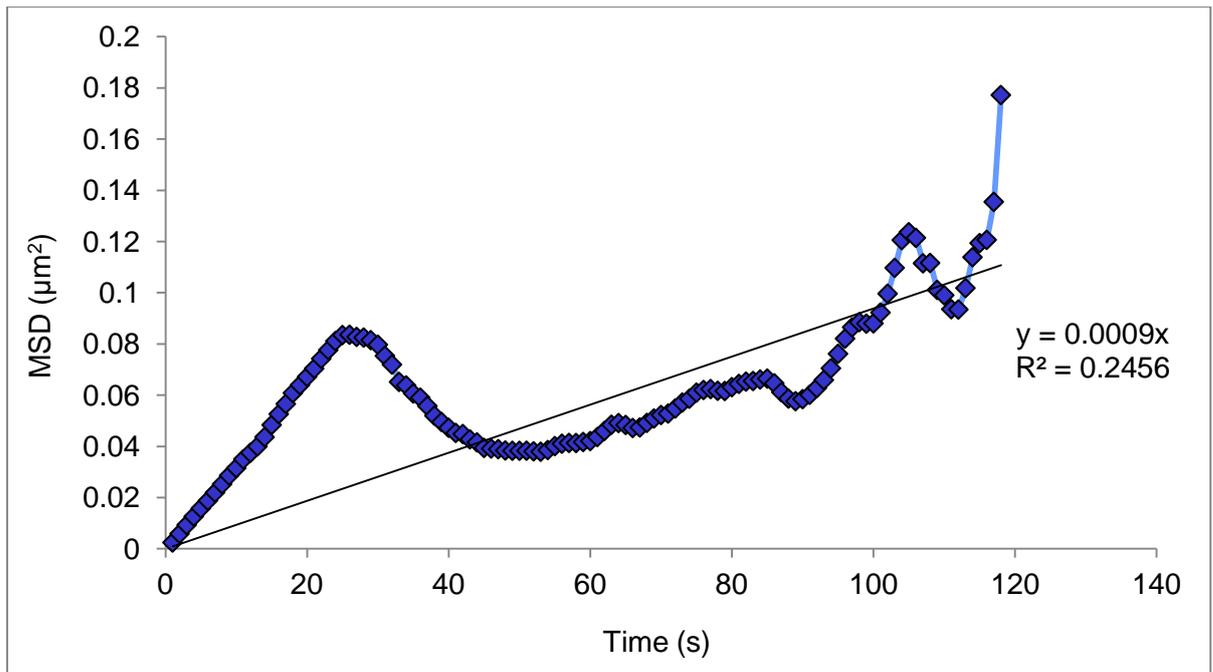
<b>Cycle Time (s)</b>	<b>Mean Position (pixels)</b>	<b>Mean Position (<math>\mu\text{m}</math>)</b>
1.00011	7.61	0.62
2.00022	7.93	0.65
3.00033	8.02	0.66
4.00044	8.14	0.67
5.00055	8.62	0.71
6.00066	9.31	0.76
7.00077	9.31	0.76
8.00088	8.91	0.73
9.00099	8.51	0.70
10.0011	8.08	0.66
11.00121	7.76	0.64
12.00132	8.08	0.66
13.00143	8.65	0.71
14.00154	8.86	0.73
15.00165	8.71	0.71
16.00176	8.31	0.68
17.00187	8.33	0.68
18.00198	8.62	0.71
19.00209	8.77	0.72
20.0022	8.57	0.70
21.00231	8.12	0.67
22.00242	7.82	0.64
23.00253	8.06	0.66
24.00264	8.42	0.69
25.00275	8.64	0.71

<b>Cycle Time (s)</b>	<b>Mean Position (pixels)</b>	<b>Mean Position (<math>\mu\text{m}</math>)</b>
93.01023	6.73	0.55
94.01034	7.58	0.62
95.01045	7.89	0.65
96.01056	8.01	0.66
97.01067	7.92	0.65
98.01078	8.8	0.72
99.01089	9.17	0.75
100.011	9.13	0.75
101.01111	9.24	0.76
102.01122	9.64	0.79
103.01133	9.64	0.79
104.01144	9.34	0.76
105.01155	9.72	0.80
106.01166	12.23	1.00
107.01177	12.97	1.06
108.01188	14.13	1.16
109.01199	12.64	1.04
110.0121	13.48	1.10
111.01221	13.08	
112.01232	12.71	1.04
113.01243	12.35	1.01
114.01254	11.58	0.95
115.01265	11.86	0.97
116.01276	12.87	1.05
117.01287	12.76	1.05

**Table 1:** An example of raw data obtained from a kymograph trace of cFos moving along  $\lambda$  DNA at 2 different time intervals over 120 seconds. The first column shows a cycle time of 1.00011 seconds. This is the time duration between each image captured by Andor Solis (as described in method 2.2.11). The second column shows the raw data that was produced by the automated Gaussian fit program. This shows the proteins displacement (pixels) per time frame. The third column shows the proteins displacement in  $\mu\text{m}$ . This unit conversion occurred by multiplying the data in pixels by 0.0819  $\mu\text{m}$ , the length of a single pixel as determined by a graticule. Any absent data from the second and third columns meant that these values did not fall within the data isolation parameters as described in method 2.2.16.

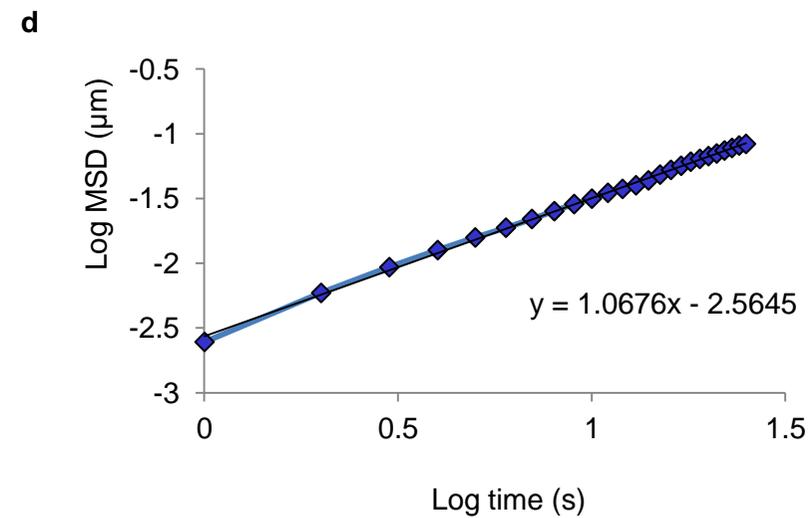
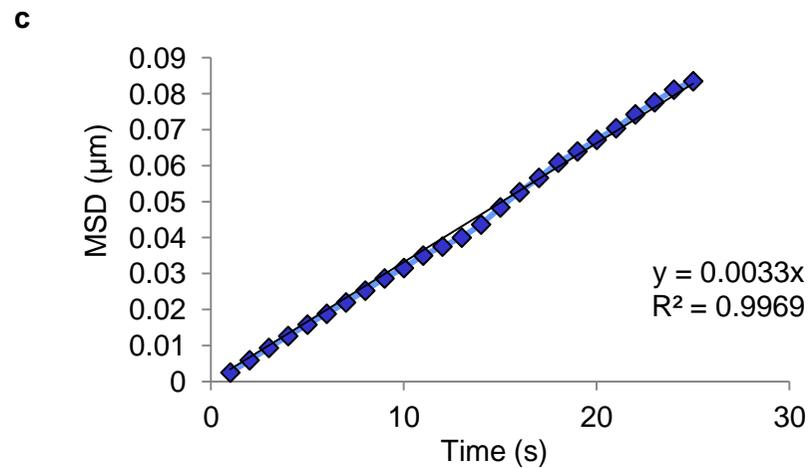
3																						
4	x1-x2	x1-x3	x1-x4	x1-x5	x1-x6	x1-x7	x1-x8	x1-x9	x1-x10	x1-x11	x1-x12	x1-x13	x1-x14	x1-x15	x1-x16	x1-x17	x1-x18	x1-x19	x1-x20	x1-x21	x1-x22	
5	1E-05	3E-05	0.0015	0.0015	0.0002	0.002	0.0628	0.0969	0.165	0.0808	0.1246		0.0841	0.0678	0.039	0.0485	0.0918	0.0864	0.0485	0.046	0.102	
6	8E-05	0.0017	0.0017	0.0003	0.0023	0.0645	0.0989	0.1677	0.0826	0.1269		0.086	0.0695	0.0403	0.05	0.0938	0.0884	0.05	0.0475	0.1041		
7	0.0011	0.0011	7E-05	0.0015	0.06	0.0933	0.1604	0.0775	0.1206		0.0808	0.0649	0.0367	0.046	0.0884	0.0831	0.046	0.0436	0.0984			
8	0	0.0006	4E-05	0.045	0.0744	0.1352	0.0604	0.0989		0.0632	0.0493	0.0252	0.0331	0.07	0.0653	0.0331	0.031	0.0789				
9	0.0006	4E-05	0.045	0.0744	0.1352	0.0604	0.0989		0.0632	0.0493	0.0252	0.0331	0.07	0.0653	0.0331	0.031	0.0789					
10	0.001	0.056	0.0884	0.1539	0.073	0.115		0.0762	0.0608	0.0337	0.0426	0.0836	0.0785	0.0426	0.0403	0.0933						
11	0.0423	0.0708	0.1305	0.0572	0.0948		0.06	0.0464	0.0232	0.0307	0.0666	0.062	0.0307	0.0287	0.0753							
12	0.0037	0.0242	0.0011	0.0105		0.0015	1E-04	0.0028	0.0009	0.0027	0.0019	0.0009	0.0013	0.0047								
13	0.009	0.0007	0.0017		0.0005	0.0026	0.013	0.0083	7E-05	0.0003	0.0083	0.0093	7E-05									
14	0.0149	0.0028		0.0135	0.0213	0.0436	0.0346	0.0106	0.0126	0.0346	0.0367	0.0075										
15	0.0047		3E-05	0.0006	0.0075	0.0041	0.0004	1E-04	0.0041	0.0048	0.0012											
16		0.004	0.0086	0.0242	0.0176	0.0025	0.0035	0.0176	0.0192	0.0011												
17																						
18	0.0009	0.0086	0.0048	0.0002	2E-05	0.0048	0.0057	0.0009														
19	0.004	0.0016	0.0018	0.0011	0.0016	0.0021	0.0035															
20	0.0005	0.0112	0.0093	0.0005	0.0003	0.0149																
21	0.0068	0.0054	0	3E-05	0.0098																	
22	8E-05	0.0068	0.0078	0.0003																		
23	0.0054	0.0063	0.0006																			
24	3E-05	0.0098																				
25	0.011																					
26																						

**Figure 22:** A section of cFos on  $\lambda$  DNA raw data generated from the automated Gaussian fit program that utilised the MSD equation as calculated in Equation 2. The first column shows the proteins displacement by subtracting each number with the number directly below, i.e., 1 to 2, 2 to 3 etc. The second column subtracts every third number, i.e., 1 to 3, 2 to 4 etc. The increments between the displacement numbers increases by 1 as you move towards the right of the table. This methodology covers all displacement configurations. The MSD was calculated by taking the average of each column.



**Figure 23:** A mean-squared displacement plot of a single cFos molecule (as shown in Table 1) that travelled along  $\lambda$  DNA. The protein was imaged every 1 second for 120 seconds. After the first 20% or 25 frames of the data trace, non-linearity can be seen. This is reflected by the  $R^2$  value that determines the quality of the fit to the data. Imaging typically has a time duration of 120 seconds to 600 seconds with a cycle time of between 0.2 to 1 frame per second.

a		b	
Cycle Time (s)	Displacement ( $\mu\text{m}^2$ )	Log time (s)	Log displacement ( $\mu\text{m}$ )
1.00011	0.00246733	4.77698E-05	-2.607772828
2.00022	0.005896399	0.301077765	-2.229413153
3.00033	0.009324392	0.477169024	-2.030379458
4.00044	0.012627559	0.602107761	-1.898680585
5.00055	0.01579894	0.699017774	-1.801372052
6.00066	0.018807474	0.77819902	-1.725669534
7.00077	0.021963504	0.84514581	-1.658298377
8.00088	0.025237621	0.903137757	-1.597951589
9.00099	0.028640824	0.954290279	-1.543014492
10.0011	0.031533462	1.00004777	-1.501228349
11.00121	0.034988248	1.041440455	-1.456077803
12.00132	0.037509349	1.079229016	-1.425860479
13.00143	0.040003242	1.113991122	-1.397904813
14.00154	0.043633753	1.146175805	-1.360177434
15.00165	0.048385484	1.176139029	-1.315284909
16.00176	0.052617981	1.204167752	-1.278865822
17.00187	0.056597932	1.230496691	-1.247199434
18.00198	0.06085024	1.255320275	-1.215737704
19.00209	0.063957767	1.278801371	-1.19410671
20.0022	0.067186496	1.301077765	-1.172718008
21.00231	0.070397327	1.322267064	-1.15244383
22.00242	0.074252986	1.342470451	-1.129286078
23.00253	0.077629717	1.361775606	-1.109972
24.00264	0.08112874	1.380259011	-1.09082527
25.00275	0.083473118	1.397987778	-1.078453364



**Figure 24:** An example MSD of cFos that travelled along  $\lambda$  DNA. This was calculated using Equation 2. (a) Shows the average displacement ( $\mu\text{m}$ ) per time frame. (b) Shows log displacement ( $\mu\text{m}$ ) per log time frame. Both panels a and b in this example show a total of 25 time steps, equating to the first 21% of the data set (the full data set is shown in Figure 21). The data was quantified based on how well the data fitted to the straight line. This was achieved by removing data from the bottom of the MSD upwards until an  $R^2$  value of 0.7 and above was achieved as this provided the best quality fit. Although the data set is smaller than the original data, linear data is required to perform an MSD. (c) A plot showing the proteins displacement ( $\mu\text{m}$ ) against time (s) using the data from panel a. The slope of the linear line determines the diffusion constant. (d) A plot showing log displacement ( $\mu\text{m}$ ) against log time (s) using the data from panel b.

## **2.2.18: The Binding Site Position Finder**

### **2.2.18.1: Binding Site Position Finder for $\lambda$ DNA**

The Binding Site Position Finder is an analytical method designed to calculate whether the measured distance between two pause sites corresponded with their AP1 target sites. Three pieces of virtual  $\lambda$  DNA were placed in tandem. This approach ensured that all binding distances were quantified.

First, the starting and end positions for each of the eight 7 bp AP1 target sites were identified and input into the program. An error margin was added to either end of the target site. To calculate this error, cJun was used as this protein is the most mobile out of all of the complexes. The error margin was calculated based on cJun pause locations as the protein travelled along  $\lambda$ , PCE pUC19 (Chapter 3) and PCE pUCap1 (Chapter 3) DNA substrates. These pause locations were identified by their Gaussian distribution graphs mean values (in bp) that were obtained from each peak. The difference between the mean values corresponded to 2 pause sites (as described in Figure 27). For each protein analysed, the standard deviations for these mean values was calculated. The average standard deviation was then determined for cJun on each DNA substrate. The mean average that determined the error margin was 80 bp.

Next, all possible target site configurations could be calculated, for example, the distance between sites 1 to 2, sites 1 to 3, sites 1 to 4 and so forth. If the distance corresponds to 2 AP1 target sites, a "Yes" will flag up within the grid situated to the right of the program. By clicking on the "Yes", the precise binding sites can be determined.

To activate the program, the user simply types the distance into the box and presses enter. Part of this Binding Site Position Finder can be seen in Figure 25.

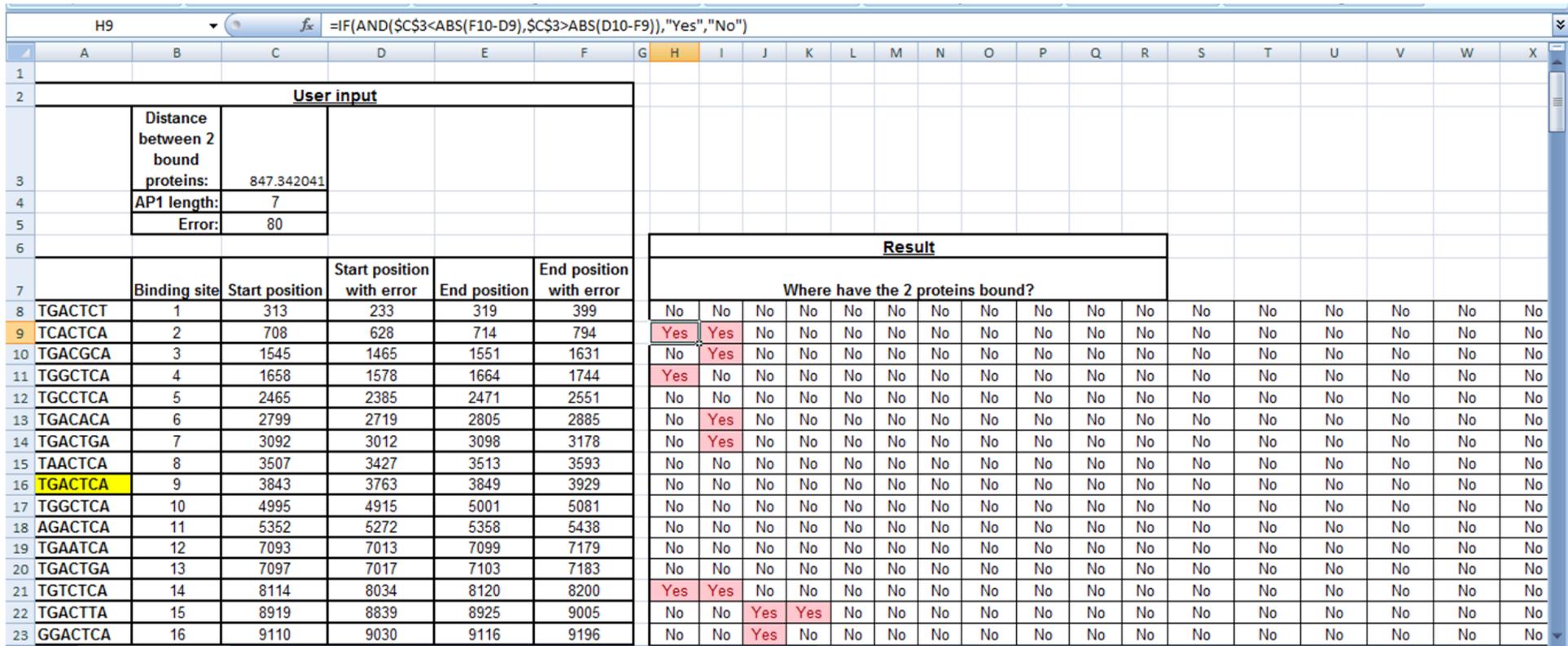
H14      fx =IF(AND(\$C\$3<E15-C14,\$C\$3>C15-E14),"Yes","No")

User input			Result																			
Distance between 2 bound proteins:	2688		Where has the protein bound?																			
AP1 length:	7																					
Binding site	Start Position	End Position																				
1	3843	3849	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
2	19119	19125	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
3	20096	20102	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
4	21153	21159	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
5	22682	22688	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
6	36560	36566	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
7	41893	41899	Yes	No																		
8	44581	44587	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
9	52345	52351	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
10	67621	67627	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
11	68598	68604	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
12	69695	69701	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
13	71184	71190	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
14	85062	85068	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
15	90395	90401	Yes	No																		

**Figure 25:** Showing part of the Binding Site Position Finder analysis output. Part of strand λ DNA 2 (highlighted green) and the whole of strand 1 (highlighted purple) can be seen in this screenshot image. This method used 3 strands (the third DNA strand cannot be seen in this image) of λ DNA which have been placed in tandem, covering 24 AP1 target sites in total. This analytical method was designed to calculate whether the measured distance between two pause sites corresponded to the protein landing within its AP1 target sites.

### **2.2.18.2: The adapted Binding Site Position Finder incorporating variant AP1 target sites**

Initially, the Binding Site Position Finder, as described in method 2.2.18.1, was used to identify whether the measured distance between two pause sites corresponded with their AP1 target sites. An unexpected percentage of translocating proteins paused directly within the 8 AP1 target sites. This contradicted studies showing that these AP1 transcription factors can only bind to their AP1 target site (Angel et al. 1988, Nakabeppu and Nathans 1989). The Binding Site Position Finder was therefore adapted to accommodate the single nucleotide substitutions within the AP1 target site sequence based on work performed by Seldeen *et al* (2009) described in sections 2.3.3 and 2.2.6.2. This adapted can only bind to their AP1 target site (Angel et al. 1988, Nakabeppu and Nathans 1989). The Binding Site Position Finder now shows 103 sites being present in total along a single piece of  $\lambda$  DNA. The order of these binding sites can be seen in Table 2. This analytical model functions in the same way as the previous Binding Site Position Finder.



**Figure 26:** The adapted Binding Site Position Finder. This program functions in the same way as the Binding Site Position Finder that is mentioned in Figure 25. It incorporates 8 AP1 target sites and 95 variant AP1 target sites. This results in 103 sites in total that span the length of a single piece of  $\lambda$  DNA. Only 1 piece of  $\lambda$  DNA has been used in this adapted program and all distances entered were quantified. This meant that a single piece of DNA was sufficient for this analytical technique. Each AP1 target site has been highlighted in yellow. Like the previous Binding Site Position Finder, the user simply types the known distance in the red outlined box and presses enter.

<b>Binding site</b>	<b>Site number</b>	<b>Binding site</b>	<b>Site number</b>
TGACTCT	1	TGAATCA	53
TCACTCA	2	TGACTAA	54
TGACGCA	3	TTACTCA	55
TGGCTCA	4	TCACTCA	56
TGCCTCA	5	TTACTCA	57
TGACACA	6	TGAATCA	58
TGACTGA	7	TGTCTCA	59
TAACTCA	8	GGACTCA	60
<b>TGACTCA</b>	9	TGACCCA	61
TGGCTCA	10	TGACTCG	62
AGACTCA	11	TGACTGA	63
TGAATCA	12	TGAATCA	64
TGACTGA	13	TGACTGA	65
TGTCTCA	14	TGGCTCA	66
TGACTTA	15	TAACTCA	67
GGACTCA	16	TGACTTA	68
TGACCCA	17	TGACTCG	69
TGACTCT	18	TGACTGA	70
TGACTCC	19	TAACTCA	71
AGACTCA	20	TGACGCA	72
TGACTCT	21	TGAATCA	73
TGACTCG	22	CGACTCA	74
TCACTCA	23	TAACTCA	75
AGACTCA	24	TGACGCA	76
TGACTCT	25	TGACACA	77
TGACTGA	26	TGAATCA	78
TGGCTCA	27	TGAATCA	79
TGACACA	28	TGAATCA	80
TCACTCA	29	CGACTCA	81
TGACCCA	30	TGCCTCA	82
TGGCTCA	31	<b>TGACTCA</b>	83
<b>TGACTCA</b>	32	TGGCTCA	84
TGGCTCA	33	TGACTCC	85
TGACTGA	34	<b>TGACTCA</b>	86
<b>TGACTCA</b>	35	TGAATCA	87
TGACGCA	36	CGACTCA	88
TGACGCA	37	TGACTCT	89
TGACTAA	38	TGACTCT	90
TGACTGA	39	TGACTTA	91
<b>TGACTCA</b>	40	<b>TGACTCA</b>	92
TGTCTCA	41	CGACTCA	93
TGACTCC	42	TGACTCG	94
TCACTCA	43	TGACTCT	95
TCACTCA	44	TGACGCA	96
TGACGCA	45	TGACTGA	97
TGCCTCA	46	TGGCTCA	98
<b>TGACTCA</b>	47	TAACTCA	99
AGACTCA	48	TGCCTCA	100
TGAATCA	49	TGACTGA	101
TGGCTCA	50	TGACCCA	102
TGAATCA	51	TGACGCA	103
TGACTAA	52		

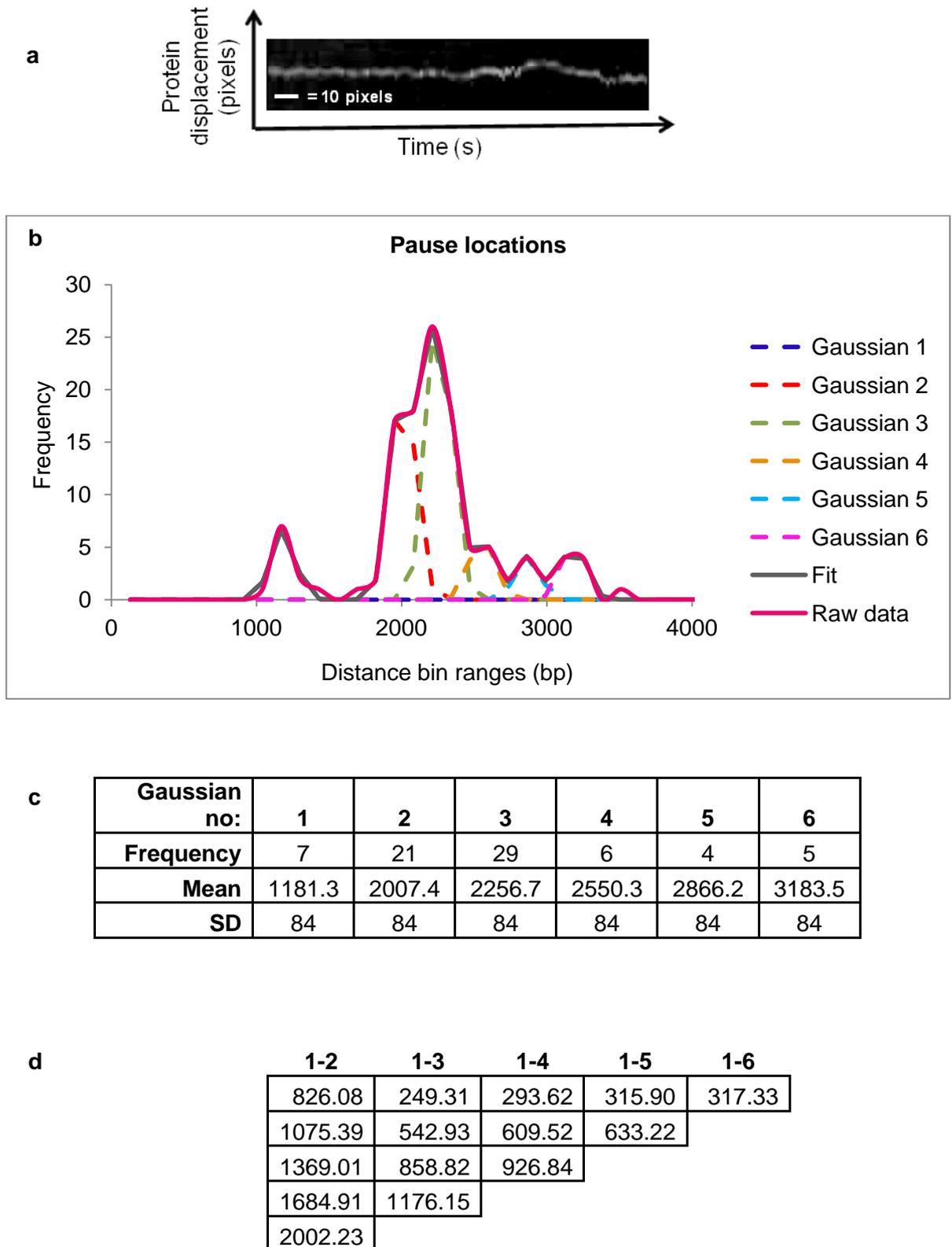
**Table 2:** A table showing the order of each binding site as they appear along a single piece of  $\lambda$  DNA. There are a total of 103 target sites. The wild type AP1 target sites have been highlighted in yellow.

### **2.2.19: Data for the Binding Site Position Finder**

cFos, cJun and cFos:cJun heterodimers travel along  $\lambda$  DNA in search of their AP1 target sites. As each single protein translocates, it can pause. Each pause time can vary in duration and some can be more obvious than others. Figure 15 shows examples of pauses that can be easily seen and those that cannot.

To identify pauses as a protein diffuses along the DNA, the mean (pixels) value from each Gaussian distribution was first converted into base pairs. The mean values (base pairs) were then binned. The bin ranges were 5 times greater than the 8.7 nm (equivalent of 26 bp) spatial resolution of the Andor camera used during imaging. This meant that large, clear peaks could be seen in the histogram. A Gaussian distribution of each peak was calculated based on the peaks amplitude, the mean and standard deviation. The sum of the Gaussian distributions were calculated and Microsoft Solver program was used to fit the graph.

The solved mean values for each Gaussian fit determined the position of each pause relative to each other. The difference between each pause site was quantified. This represented the distance in base pairs between pauses. These numbers were then entered into the adapted Binding Site Position Finder (as described in method 2.2.18.2) to identify whether the proteins paused specifically in their AP1 target sites or within the variant AP1 target sites. Examples of this have been shown in Figure 27.



**Figure 27:** As a protein travels along  $\lambda$  DNA, it pauses. This figure describes how each pause site was identified and whether the distances corresponded to the protein pausing inside AP1 or variant AP1 target sites. (a) An example kymograph of cFos travelling along  $\lambda$  DNA over a time frame of 120 seconds. 10 pixels is  $0.819 \mu\text{m}$ , the equivalent of 2410 bp (represented by the scale bar). Some pauses can be clearly

seen, whereas others are harder to identify. (b) The overlaid Gaussian distribution plots obtained from the above kymograph. The mean position of each solved Gaussian distribution plot are 1181 bp, 2007 bp, 2256 bp, 2550 bp, 2866 bp and 3183 bp. These distances determined the proteins pause positions relative to each other. (c) This table shows the number of times a site was visited as shown by the frequency. The mean position of each Gaussian distribution corresponds to a pause site and the standard deviation (SD) for the first plot has been identified. This number has been fixed across all Gaussian distributions in order to fit the data using Microsoft Solver. (d) Using the method outlined in Figure 22, the distances between each Gaussian distributions mean (as shown in Figure 27c) have been calculated. For example, the distance between peak 1 and 2, peak 1 and 3 etc. This method meant that all of the distances between each pause site could be quantified.

## 2.3: Results

To understand how cFos, cJun, cFos:cJun, FosW and FosW+cJun interact with  $\lambda$  DNA, the number of proteins bound to DNA were quantified and studied. The number of DNA tightropes bound complexes were quantified for the following: cFos ( $n = 330$ ), cJun ( $n = 174$ ) and cFos:cJun ( $n = 241$ ). Similarly, when determining the number of bound complexes for FosW and FosW+cJun, 109 and 67 tightropes were observed respectively. Following this, further studies were performed to gain an insight into how the proteins interacted with the DNA and the formation of homodimeric and heterodimeric complexes were identified using dual colour Qdot labelling.

### 2.3.1: Protein interaction with $\lambda$ DNA.

A Biotin tag was conjugated to the C-terminus of cFos, cJun and FosW. However, for the purpose of this study, a Biotin tag and a cJun DNA binding domain was a necessity as FosW could not be visualised without conjugated Streptavidin Qdots (as described in method 2.2.9). Qdot conjugated proteins could be visualised during imaging using OAF microscopy.

Typically, cFos and FosW were labelled with red 655 nm emitting Qdots and cJun was labelled with green 525 nm emitting Qdots (Invitrogen) using a 1:4 peptide to Qdot ratio (as described in method 2.2.9). cFos, cJun and FosW were labelled with the afore-mentioned Qdots colours. cFos:cJun and FosW+cJun complexes were labelled using 2 colours. Dual colour experiments provided a host of red, green and dual coloured complexes. Dual colour labelling identified true heterodimeric complex formation. This colour labelling strategy was adopted throughout the entire duration of this study, ensuring that all proteins were fluorescently tagged prior to being loaded onto the DNA by using a 4 Qdots: 1 protein ratio as described in method 2.2.9. All studies were performed in a high salt buffer, mimicking intracellular conditions *in vivo*.

Many Qdot-protein complexes were observed bound to the DNA. The longer the DNA tightrope, the more complexes bound. It was also found that very long tightropes tended to “sway”, resulting in the bound proteins not being analysed as they would present identical moving patterns that could be seen in the kymographs. Tightropes that appeared highly decorated with proteins were also avoided because of overcrowding. Therefore, tightropes with a minimum of 2

proteins bound were analysed. Proteins that were seen next to a bead were also not analysed. This was due to high fluorescence being emitted from the beads, obscuring protein movement from analysis.

The first study performed aimed to determine the number of Qdot labelled cFos, cJun, cFos:cJun, FosW and FosW+cJun complexes that were observed bound to  $\lambda$  DNA, forming protein-DNA complexes, and to calculate the percentage of molecules that moved. Surprisingly, cFos was observed binding to the DNA. This is an unexpected phenomenon as many studies reported that cFos is unable to homodimerise or bind to DNA (Mason et al. 2006, Nakabeppu and Nathans 1989). 518 cFos, 266 cJun and 206 FosW molecules were observed using their respective single colours. To quantify cFos:cJun heterodimers, equal amounts of red labelled cFos and green labelled cJun were mixed (as described in 2.2.9). From this mixture 326 molecules were analysed. 139 were red cFos complexes, 82 were green cJun complexes and 105 were dual colour cFos:cJun heterodimers. Similarly, FosW+cJun dual coloured complexes were also identified. From this dual coloured mixture, 112 molecules were analysed. 60 molecules were red FosW complexes, 24 molecules were green cJun and 28 molecules were dual labelled FosW+cJun heterodimeric complexes.

In order to identify the number of bound molecules that moved along the DNA, kymographs were produced. These provided a fast and effective way of enabling each individual protein displacement to be tracked per time frame. Kymographs also enabled the diffusion constant ( $\mu\text{m}^2\text{s}^{-1}$ ) and alpha value to be determined for each sliding protein. The data from this part of the study showed that 45% cFos, 71% cJun (t-test p-value = 0.03) and 55% of FosW diffused along  $\lambda$  DNA. 39% of cFos:cJun heterodimers (t-test p-value = 0.06) and 39% of FosW:cJun (t-test p-value = 0.7) dual labelled complexes were also observed to diffuse along  $\lambda$  DNA (shown in Table 3). This data shows that cJun is the most mobile and the antagonistic peptide inhibitor FosW is the second most mobile complex. Surprisingly, cFos is the third most mobile complex and is capable of moving along DNA in addition to simply binding to it. cFos:cJun heterodimers and FosW+cJun are the least mobile.

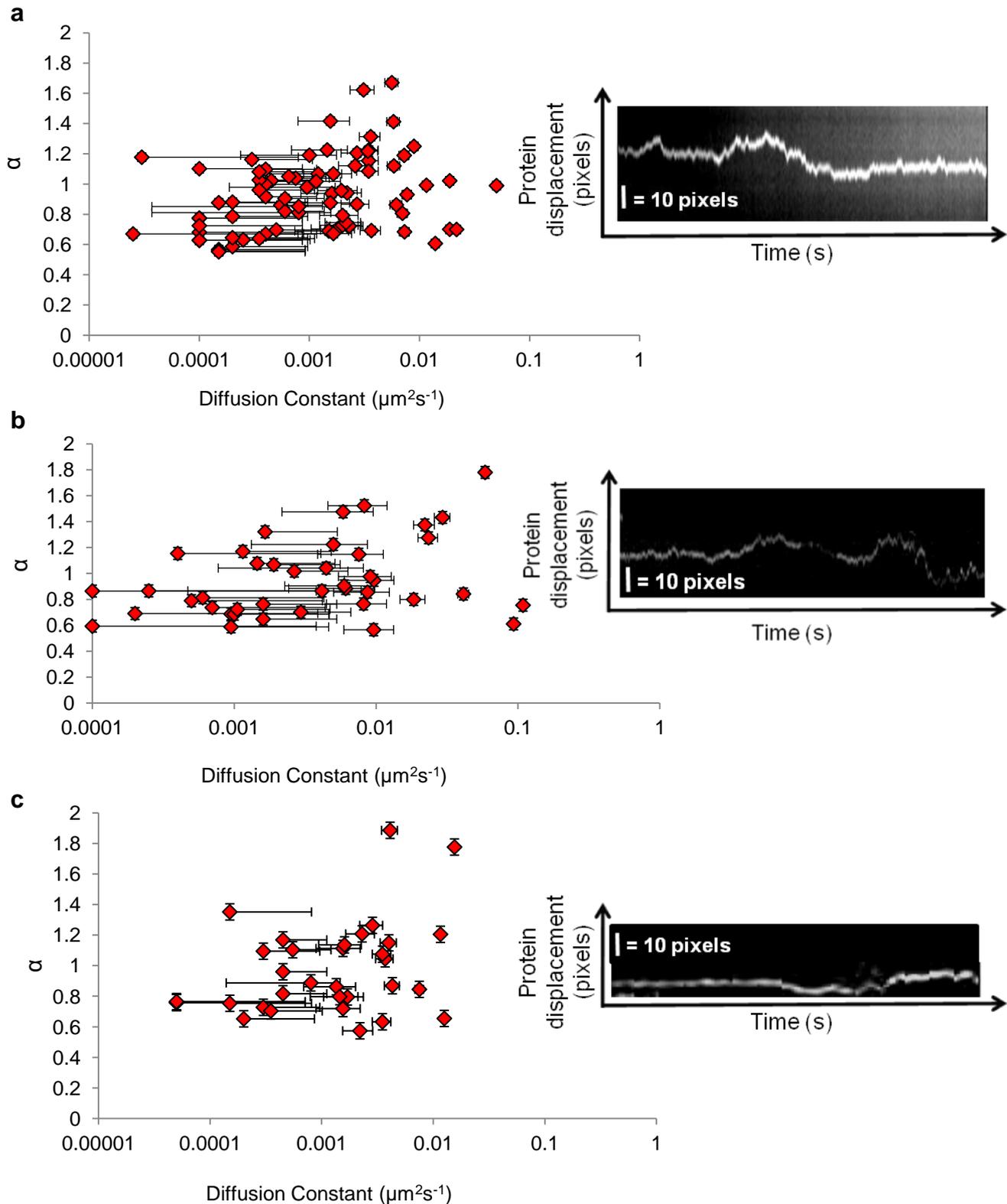
As  $\lambda$  DNA contains methylated bases, non-methylated  $\lambda$  DNA was also used to understand whether these  $\text{CH}_3$  groups hindered protein movement. cJun was loaded onto the DNA as this molecule was the most mobile protein. 244 cJun molecules were observed bound to the DNA. When cJun was loaded onto non-methylated  $\lambda$  DNA,  $60 \pm 3.0\%$  moved in comparison to cJun on methylated  $\lambda$  DNA (t-test p-value = 0.6).

The next part of this study was to identify the diffusion constants and  $\alpha$  values for cFos, cJun, and FosW that travelled along  $\lambda$  DNA while in a high salt buffer by using Equation 1 (as described in method 2.2.16). cFos produced diffusion constants ranging from  $1 \times 10^{-5}$  to  $1 \times 10^{-1} \mu\text{m}^2\text{s}^{-1}$ , cJun  $1 \times 10^{-4}$  to  $1 \times 10^{-1} \mu\text{m}^2\text{s}^{-1}$  and FosW  $1 \times 10^{-5}$  to  $1 \times 10^{-2} \mu\text{m}^2\text{s}^{-1}$ . This meant that the average diffusion constants for cFos, cJun and FosW were  $3.4 \times 10^{-3} \pm 8 \times 10^{-4} \mu\text{m}^2\text{s}^{-1}$ ,  $1.2 \times 10^{-2} \pm 4 \times 10^{-3} \mu\text{m}^2\text{s}^{-1}$  and  $2.8 \times 10^{-3} \pm 7 \times 10^{-4} \mu\text{m}^2\text{s}^{-1}$ , respectively. These findings have been shown in Figure 28.

To further understand how the proteins travelled along the DNA, their  $\alpha$  values were determined. This provided an insight into the mode of travel used when locating an AP1 or variant AP1 target site. cFos produced  $\alpha$  values ranging from 0.6 to 1.7, cJun 0.6 to 1.8 and FosW 0.6 to 1.9. Their average  $\alpha$  values were 0.9 (SEM = +/- 0.03), 1.0 (SEM = +/- 0.05) and 1.0 (SEM = +/- 0.05) respectively. Displaying an  $\alpha$  value of  $<1$  is indicative of the protein stringently searching the DNA for target sites using random diffusion. These can be seen in Figure 28.

	$\lambda$ DNA
<b>cFos</b>	<b>45 ± 3.7% (n = 518)</b>
<b>cJun</b>	<b>71 ± 3.1% (n = 266)</b>
<b>cFos:cJun Heterodimers</b>	<b>39 ± 6.1% (n = 326)</b>
<b>FosW</b>	<b>55 ± 11.0% (n = 206)</b>
<b>FosW + cJun</b>	<b>39 ± 11.0% (n = 112)</b>

**Table 3:** A table showing the percentage of cFos, cJun, cFos:cJun, FosW and FosW + cJun complexes that were observed diffusing along  $\lambda$  DNA. All cFos:cJun and FosW:cJun complexes that presented as dual colour were quantified as Heterodimeric. All complexes were observed while high salt buffer.



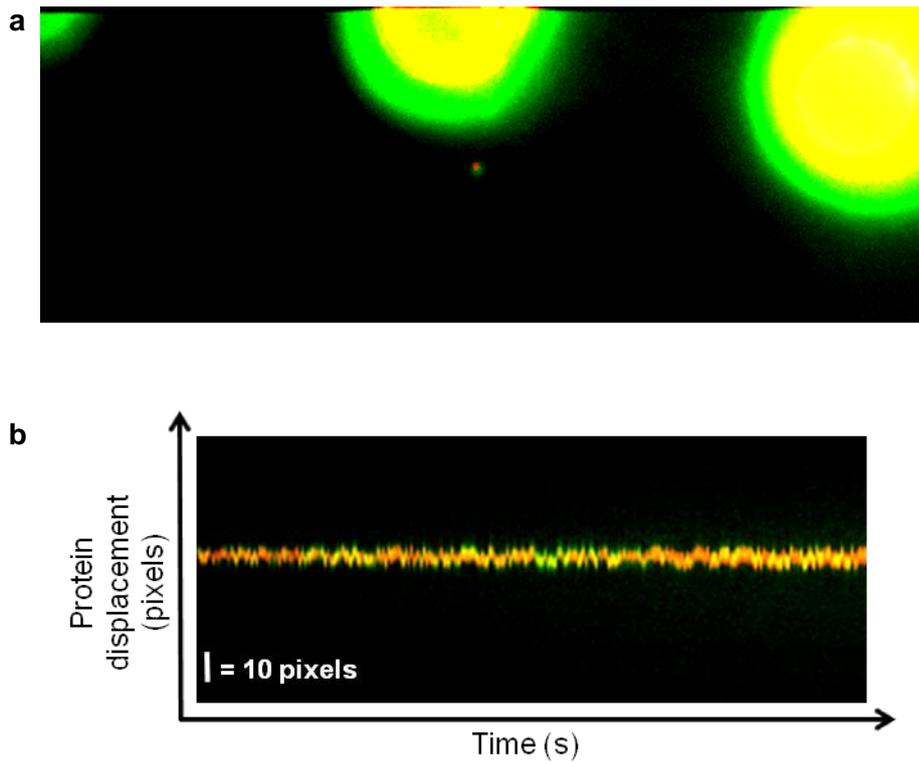
**Figure 28:** Graphs showing diffusion constants ( $\mu\text{m}^2\text{s}^{-1}$ ) against their corresponding  $\alpha$  values for cFos, cJun and FosW while in a high salt buffer. (a) cFos ( $n = 80$ ) displays a range of diffusion constants between  $1 \times 10^{-5}$  and  $1 \times 10^{-1} \mu\text{m}^2\text{s}^{-1}$  and  $\alpha$  values between 0.5 and 1.9. (b) cJun ( $n = 41$ ) shows a range of diffusion constants between  $1 \times 10^{-4}$  and  $1 \times 10^{-1} \mu\text{m}^2\text{s}^{-1}$  and  $\alpha$  values between 0.6 and 1.8. (c) FosW ( $n = 33$ ) shows a range of diffusion constants between  $1 \times 10^{-5}$  and  $1 \times 10^{-2} \mu\text{m}^2\text{s}^{-1}$  and  $\alpha$  values between 0.6 and 1.9. A kymograph representing each proteins movement is situated to the right of each graph. The scale bar on each kymograph represents 10 pixels which is  $0.819 \mu\text{m}$ , the equivalent of 2410 bp.

### 2.3.2: Protein colocalisation

To visualise heterodimeric and homodimeric complexes that had bound to  $\lambda$  DNA, a dual colour labelling technique was used. cFos and FosW were labelled with red Qdots and cJun was labelled with green Qdots. After the dual labelled protein mixture was loaded onto the DNA, all of the single and dual coloured complexes observed were quantified. An example kymograph showing a dual coloured cFos:cJun heterodimer can be seen in Figure 29.

One question that arose from the protein interaction data, as presented in section 2.3.1, was whether cFos remained monomeric or formed a homodimeric complex. cFos labelled with red and green Qdots were incubated together. 369 molecules were analysed; 66% ( $n = 242$ ) were red, 8% ( $n = 29$ ) were green and 27% ( $n = 98$ ) were dual coloured complexes. Visualising cFos as a dual coloured complex answers this question and clearly shows that cFos is capable of homodimerising.

The colocalisation of cJun was also visualised. 209 molecules were analysed; 60% ( $n = 125$ ) were coloured red, 13% ( $n = 28$ ) were coloured green and 27% ( $n = 56$ ) were dual coloured. To visualise heterodimeric complex formation, cFos and cJun were mixed together; 326 molecules were analysed; 43% ( $n = 139$ ) were red cFos, 25% ( $n = 82$ ) were green cJun and 32% ( $n = 105$ ) were dual coloured cFos:cJun heterodimers. The antagonistic peptide inhibitor FosW was also mixed with its target cJun protein and 112 molecules were analysed; 54% ( $n = 60$ ) were red FosW, 21% ( $n = 24$ ) were green cJun and 25% ( $n = 28$ ) were dual coloured FosW+cJun complexes. ~30% of all complexes were visualised as dual colour. These results can be seen in Table 4.



**Figure 29:** (a) An image showing a cFos:cJun heterodimer in dual colour. (b) An example kymograph showing the dual coloured cFos:cJun heterodimer in Figure 29a diffusing along  $\lambda$  DNA. The scale bar represents 10 pixels.

	Colocalised % on $\lambda$ DNA
<b>cFos</b>	<b>27 <math>\pm</math> 6.4% (<i>n</i> = 369)</b>
<b>cJun</b>	<b>27 <math>\pm</math> 10.4% (<i>n</i> = 209)</b>
<b>cFos:cJun heterodimers</b>	<b>32 <math>\pm</math> 4.3% (<i>n</i> = 326)</b>
<b>FosW + cJun</b>	<b>25 <math>\pm</math> 11.0% (<i>n</i> = 112)</b>

**Table 4:** A table showing the percentage of colocalised dual coloured cFos, cJun, cFos:cJun and FosW:cJun complexes (protein labelling has been described in method 2.2.9) that were bound to  $\lambda$  DNA. All complexes were observed while in high salt buffer.

### 2.3.3: Protein binding positions along $\lambda$ DNA

As previously mentioned, studies have shown that cJun and cFos:cJun heterodimers bind to their AP1 target site only (Mason et al. 2006). Many studies concluded this by performing an Electrophoretic Mobility Shift Assay (EMSA). These assays used oligonucleotides that contained a centrally positioned AP1 target site. This meant that the ratio of protein-DNA complexes and unbound DNA could only be quantified via a gel matrix (Orchard and May 1993).

Seldeen *et al* (2009) followed up these initial EMSA results by using isothermal titration calorimetry and structural modelling to study the possibility of cFos:cJun heterodimers binding to synthetic dsDNA oligonucleotides that contained single nucleotide variations of the AP1 target site sequence TGA**CT**CA. They found that the heterodimers bound to CGACTCA, TGA**CT**CG, TGTCTCA, TGACACA, TGCCTCA and TGACGCA variant target sites.

However, this thesis has demonstrated that cFos, cJun and cFos:cJun heterodimers can associate and travel along  $\lambda$  DNA while in a high salt buffer. In addition, pausing was observed as shown in Figure 12. The bottom protein shows clear pausing, whereas the upper protein trace shows an example of pausing that is less obvious and more difficult to quantify. Therefore, any cFos, cJun and cFos:cJun molecules that produced a diffusion constant ( $\mu\text{m}^2\text{s}^{-1}$ ) and  $\alpha$  value were further analysed to identify where these proteins may be pausing along  $\lambda$  DNA.

There are 103 AP1 and variant AP1 target sites in total that span the length of a single piece of  $\lambda$  DNA. These sites have also been listed in Table 2. The number of pauses were quantified using Method 2.2.19. The distance (bp) and configuration between more than 2 pause sites were then calculated and these values were input into the Binding Site Position Finder. The possibility of a protein visiting particular sites was recorded and the total percentage was calculated.

Table 6 shows the sites along  $\lambda$  DNA where cFos paused. 26 molecules were analysed. These provided 126 distances (bp) between pause locations that in turn produced 20,636 binding site possibilities. This data shows that the percentage of cFos binding into their AP1 target sites

varied from 0.4-1.5%. The remaining cFos molecules that bound to each variant AP1 target sites varied from 0.4-2.3%. The most commonly visited variant AP1 target site was TGCCTCA. Similarly, Table 7 shows where cJun paused as these molecules travelled along  $\lambda$  DNA. 17 molecules were analysed. These provided 126 distances (bp) between pause locations and the number of binding site possibilities produced was 20,378. The percentage of cJun molecules that bound to their AP1 target sites varied from 0.3-1.7%. Also, the percentage of cJun that bound to each variant AP1 target sites ranged from 0.2-2.3%. Like cFos, the most common pause site was the variant AP1 target site TGCCTCA.

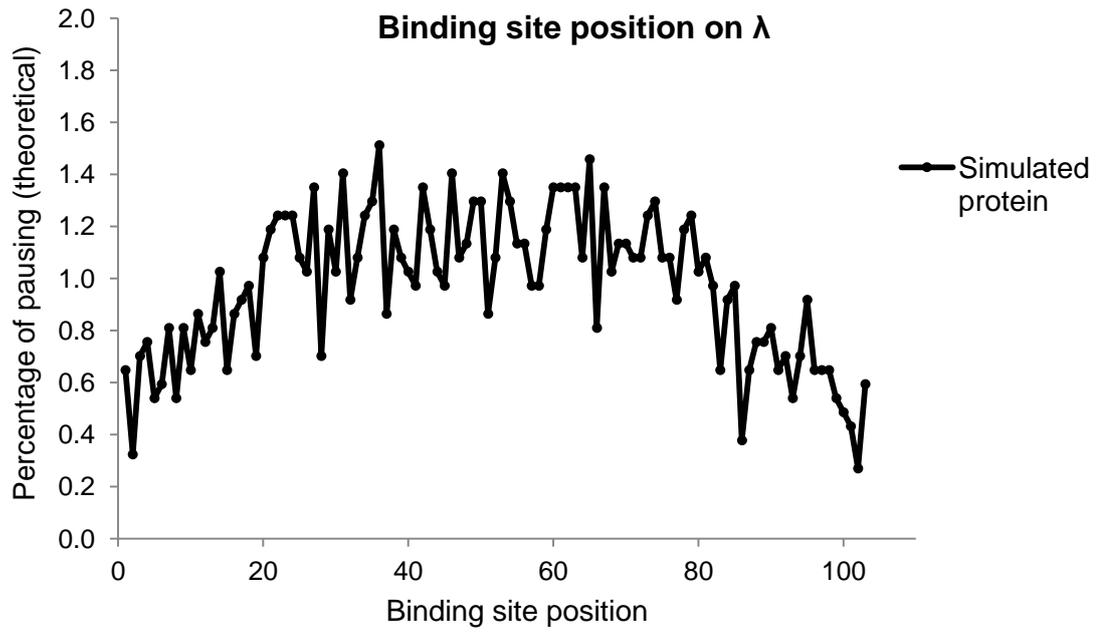
Table 8 however shows where cFos:cJun heterodimers paused as they travelled along  $\lambda$  DNA. 7 molecules were analysed and these provided 13 distances between pause locations. 1,940 binding site possibilities were produced. It was found that the percentage of cFos:cJun heterodimers that bound to their AP1 target sites varied between 0.1-1.3%. The remaining molecules that bound to the variant AP1 target sites ranged between 0.1-2.8%. Unlike cFos and cJun, the most popular binding site that cFos:cJun heterodimers paused in was TGACTTA.

To identify whether any of the AP1 and variant AP1 target sites were preferred by cFos, cJun and cFos:cJun heterodimers, a correlation graph (as shown in Figure 31) was plotted. This shows the percentage of pausing at each of the 103 target sites by all 3 proteins. cFos (blue) and cJun (pink) are visually similar. The cFos:cJun heterodimers (green) however show no correlation with cFos or cJun proteins.

To validate the Binding Site Position Finder, random numbers between 1000 and 40,000 ( $\lambda$  DNA is 48,502 bp long) were generated in Excel using a random numbers formula. These numbers mimicked the distances between pauses taken by the simulated protein as it travelled along  $\lambda$  DNA. When the simulated protein travelled along pUC19 DNA, 20 distances were entered into the Binding Site Position Finder and these generated 1852 binding site possibilities. When this simulation is compared with cFos, cJun and cFos:cJun heterodimers pausing within their AP1 and variant AP1 target sites, no correlation can be seen. This simulated data can be seen in Table 5 and Figures 30 and 32.

Simulated protein on $\lambda$								
Binding site	Start positions with error	%	Binding site	Start positions with error	%	Binding site	Start positions with error	%
TGACTCT	313	0.6	TGACGCA	20564	1.5	TAACTCA	29638	1.1
TCACTCA	708	0.3	TGACGCA	20962	0.9	TGACGCA	30321	1.1
TGACGCA	1545	0.7	TGACTAA	21030	1.2	TGAATCA	30593	1.2
TGGCTCA	1658	0.8	TGACTGA	21144	1.1	CGACTCA	30761	1.3
TGCCTCA	2465	0.5	TGACTCA	21153	1.0	TAACTCA	31170	1.1
TGACACA	2799	0.6	TGTCTCA	21420	1.0	TGACGCA	31344	1.1
TGACTGA	3092	0.8	TGACTCC	21657	1.3	TGACACA	32092	0.9
TAACTCA	3507	0.5	TCACTCA	21766	1.2	TGAATCA	32338	1.2
TGACTCA	3843	0.8	TCACTCA	21986	1.0	TGAATCA	32825	1.2
TGGCTCA	4995	0.6	TGACGCA	22056	1.0	TGAATCA	33563	1.0
AGACTCA	5352	0.9	TGCCTCA	22155	1.4	CGACTCA	33769	1.1
TGAATCA	7093	0.8	TGACTCA	22682	1.1	TGCCTCA	35793	1.0
TGACTGA	7097	0.8	AGACTCA	22764	1.1	TGACTCA	36560	0.6
TGTCTCA	8114	1.0	TGAATCA	22851	1.3	TGGCTCA	37410	0.9
TGACTTA	8919	0.6	TGGCTCA	22934	1.3	TGACTCC	38677	1.0
GGACTCA	9110	0.9	TGAATCA	23490	0.9	TGACTCA	41893	0.4
TGACCCA	9321	0.9	TGACTAA	23531	1.1	TGAATCA	43861	0.6
TGACTCT	9701	1.0	TGAATCA	23812	1.4	CGACTCA	43939	0.8
TGACTCC	9827	0.7	TGACTAA	23911	1.3	TGACTCT	43963	0.8
AGACTCA	10576	1.1	TACTCA	24693	1.1	TGACTCT	44467	0.8
TGACTCT	10614	1.2	TCACTCA	24705	1.1	TGACTTA	44553	0.6
TGACTCG	11096	1.2	TACTCA	25299	1.0	TGACTCA	44581	0.7
TCACTCA	11100	1.2	TGAATCA	25507	1.0	CGACTCA	44805	0.5
AGACTCA	11671	1.2	TGTCTCA	25603	1.2	TGACTCG	45437	0.7
TGACTCT	12058	1.1	GGACTCA	26157	1.3	TGACTCT	45972	0.9
TGACTGA	12289	1.0	TGACCCA	26191	1.3	TGACGCA	46132	0.6
TGGCTCA	12776	1.3	TGACTCG	26281	1.3	TGACTGA	46262	0.6
TGACACA	14134	0.7	TGACTGA	26744	1.3	TGGCTCA	47445	0.6
TCACTCA	14361	1.2	TGAATCA	27614	1.1	TAACTCA	47817	0.5
TGACCCA	16343	1.0	TGACTGA	27794	1.5	TGCCTCA	47884	0.5
TGGCTCA	18179	1.4	TGGCTCA	28285	0.8	TGACTGA	47938	0.4
TGACTCA	19119	0.9	TAACTCA	28718	1.3	TGACCCA	48180	0.3
TGGCTCA	19341	1.1	TGACTTA	28923	1.0	TGACGCA	48391	0.6
TGACTGA	20083	1.2	TGACTCG	29270	1.1			
TGACTCA	20096	1.3	TGACTGA	29450	1.1			

**Table 5:** A table showing the percentage of pausing within each AP1 and variant AP1 target site by a simulated protein as it travelled along the  $\lambda$  DNA. The 103 AP1 (highlighted yellow) and variant AP1 target sites have been listed in columns 1, 4 and 7 in the order in which they appear along the DNA. The starting position for each corresponding site is listed in columns 2, 5 and 8. The number of times a protein paused within each site has been shown as a percentage. These can be seen in columns 3, 6 and 9. Pausing within the AP1 and variant AP1 target sites varied from 0.6-1.3% and 0.3-1.5%, respectively.



**Figure 30:** A graph representing the percentage of pausing at each of the 103 target sites by a computer simulated protein as it travelled along  $\lambda$  DNA.

cFos binding positions								
Binding site	Start position on $\lambda$	%	Binding site	Start position on $\lambda$	%	Binding site	Start position on $\lambda$	%
TGACTCT	313	0.5	TGACGCA	20564	1.2	TAACTCA	29638	1.0
TAAGTCA	708	0.6	TGACGCA	20962	1.1	TGACGCA	30321	0.9
TGACGCA	1545	0.7	TGACTAA	21030	1.1	TGAATCA	30593	1.0
TGGCTCA	1658	1.1	TGACTGA	21144	1.0	CGACTCA	30761	1.4
TGCCTCA	2465	1.2	<b>TGACTCA</b>	21153	1.3	TAACTCA	31170	1.3
TGACACA	2799	0.6	TGTCTCA	21420	2.1	TGACGCA	31344	1.2
TGACTGA	3092	0.6	TGACTCC	21657	1.6	TGACACA	32092	0.8
TAAGTCA	3507	0.9	TAAGTCA	21766	1.3	TGAATCA	32338	0.4
<b>TGACTCA</b>	3843	0.5	TAAGTCA	21986	1.6	TGAATCA	32825	0.7
TGGCTCA	4995	0.6	TGACGCA	22056	1.3	TGAATCA	33563	0.6
AGACTCA	5352	0.7	<b>TGCCTCA</b>	22155	2.3	CGACTCA	33769	1.0
TGAATCA	7093	0.8	<b>TGACTCA</b>	22682	1.5	TGCCTCA	35793	0.6
TGACTGA	7097	0.5	AGACTCA	22764	0.8	<b>TGACTCA</b>	36560	0.5
TGTCTCA	8114	0.4	TGAATCA	22851	1.4	TGGCTCA	37410	0.6
TGACTTA	8919	0.9	TGGCTCA	22934	1.6	TGACTCC	38677	0.4
GGACTCA	9110	1.2	TGAATCA	23490	1.5	<b>TGACTCA</b>	41893	0.4
TGACCCA	9321	1.0	TGACTAA	23531	1.3	TGAATCA	43861	0.9
TGACTCT	9701	1.0	TGAATCA	23812	1.0	CGACTCA	43939	1.0
TGACTCC	9827	1.0	TGACTAA	23911	1.2	TGACTCT	43963	1.0
AGACTCA	10576	1.0	TTACTCA	24693	0.8	TGACTCT	44467	1.2
TGACTCT	10614	1.0	TAAGTCA	24705	0.7	TGACTTA	44553	1.7
TGACTCG	11096	0.9	TTACTCA	25299	1.0	<b>TGACTCA</b>	44581	0.6
TAAGTCA	11100	1.0	TGAATCA	25507	1.0	CGACTCA	44805	1.1
AGACTCA	11671	0.8	TGTCTCA	25603	1.2	TGACTCG	45437	1.1
TGACTCT	12058	0.9	GGACTCA	26157	0.7	TGACTCT	45972	1.0
TGACTGA	12289	0.5	TGACCCA	26191	0.8	TGACGCA	46132	0.8
TGGCTCA	12776	1.4	TGACTCG	26281	1.1	TGACTGA	46262	0.6
TGACACA	14134	1.0	TGACTGA	26744	0.5	TGGCTCA	47445	1.1
TAAGTCA	14361	0.6	TGAATCA	27614	0.7	TAAGTCA	47817	1.0
TGACCCA	16343	0.5	TGACTGA	27794	0.9	TGCCTCA	47884	0.9
TGGCTCA	18179	0.7	TGGCTCA	28285	0.7	TGACTGA	47938	0.9
<b>TGACTCA</b>	19119	0.7	TAAGTCA	28718	1.4	TGACCCA	48180	1.3
TGGCTCA	19341	1.3	TGACTTA	28923	1.4	TGACGCA	48391	1.0
TGACTGA	20083	0.5	TGACTCG	29270	1.6			
<b>TGACTCA</b>	20096	0.9	TGACTGA	29450	1.1			

**Table 6:** A table showing the percentage of moving cFos molecules ( $n = 26$ ) that bound to all AP1 (highlighted yellow) and variant AP1 target sites as they travelled along  $\lambda$  DNA. Columns 1, 4 and 7 show the 103 AP1 and variant AP1 target sites in the order in which they appear along a piece of  $\lambda$  DNA. Columns 2, 5 and 8 show the starting position (including the 80 bp error margin) of each target site and

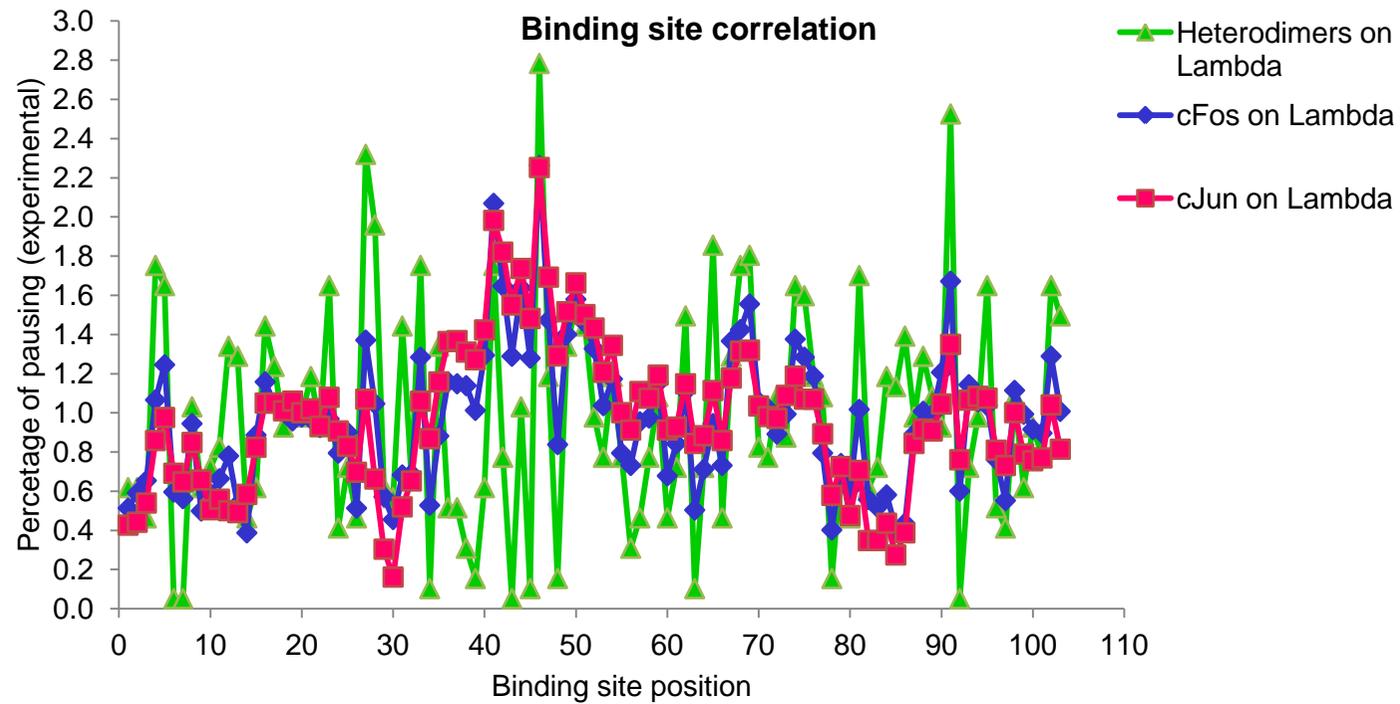
columns 3, 6 and 9 show the percentage of molecules that bound to each site. 2.3% of cFos visited the variant AP1 target site TGCCTCA (highlighted green), making this the most commonly visited site. 0.4-1.5% of the molecules visited their AP1 target sites.

cJun binding positions								
Binding site	Start position on $\lambda$	%	Binding site	Start position on $\lambda$	%	Binding site	Start position on $\lambda$	%
TGACTCT	313	0.4	TGACGCA	20564	1.4	TAACTCA	29638	1.0
TCACTCA	708	0.4	TGACGCA	20962	1.4	TGACGCA	30321	1.0
TGACGCA	1545	0.5	TGACTAA	21030	1.3	TGAATCA	30593	1.1
TGGCTCA	1658	0.9	TGACTGA	21144	1.3	CGACTCA	30761	1.2
TGCCTCA	2465	1.0	<b>TGACTCA</b>	21153	1.4	TAACTCA	31170	1.1
TGACACA	2799	0.7	TGTCTCA	21420	2.0	TGACGCA	31344	1.1
TGACTGA	3092	0.6	TGACTCC	21657	1.8	TGACACA	32092	0.9
TAACTCA	3507	0.8	TCACTCA	21766	1.6	TGAATCA	32338	0.6
<b>TGACTCA</b>	3843	0.7	TCACTCA	21986	1.7	TGAATCA	32825	0.7
TGGCTCA	4995	0.5	TGACGCA	22056	1.5	TGAATCA	33563	0.5
AGACTCA	5352	0.6	<b>TGCCTCA</b>	22155	2.3	CGACTCA	33769	0.7
TGAATCA	7093	0.5	<b>TGACTCA</b>	22682	1.7	TGCCTCA	35793	0.3
TGACTGA	7097	0.5	AGACTCA	22764	1.3	<b>TGACTCA</b>	36560	0.3
TGTCTCA	8114	0.6	TGAATCA	22851	1.5	TGGCTCA	37410	0.4
TGACTTA	8919	0.8	TGGCTCA	22934	1.7	TGACTCC	38677	0.3
GGACTCA	9110	1.1	TGAATCA	23490	1.5	<b>TGACTCA</b>	41893	0.4
TGACCCA	9321	1.1	TGACTAA	23531	1.4	TGAATCA	43861	0.8
TGACTCT	9701	1.0	TGAATCA	23812	1.2	CGACTCA	43939	0.9
TGACTCC	9827	1.1	TGACTAA	23911	1.3	TGACTCT	43963	0.9
AGACTCA	10576	1.0	TTACTCA	24693	1.0	TGACTCT	44467	1.0
TGACTCT	10614	1.0	TCACTCA	24705	0.9	TGACTTA	44553	1.3
TGACTCG	11096	0.9	TTACTCA	25299	1.1	<b>TGACTCA</b>	44581	0.8
TCACTCA	11100	1.1	TGAATCA	25507	1.1	CGACTCA	44805	1.1
AGACTCA	11671	0.9	TGTCTCA	25603	1.2	TGACTCG	45437	1.1
TGACTCT	12058	0.8	GGACTCA	26157	1.9	TGACTCT	45972	1.1
TGACTGA	12289	0.7	TGACCCA	26191	0.9	TGACGCA	46132	0.8
TGGCTCA	12776	1.1	TGACTCG	26281	1.1	TGACTGA	46262	0.7
TGACACA	14134	0.7	TGACTGA	26744	0.8	TGGCTCA	47445	1.0
TCACTCA	14361	0.3	TGAATCA	27614	0.9	TAACTCA	47817	0.8
TGACCCA	16343	0.2	TGACTGA	27794	1.1	TGCCTCA	47884	0.8
TGGCTCA	18179	0.5	TGGCTCA	28285	0.9	TGACTGA	47938	0.8
<b>TGACTCA</b>	19119	0.7	TAACTCA	28718	1.2	TGACCCA	48180	1.0
TGGCTCA	19341	1.1	TGACTTA	28923	1.3	TGACGCA	48391	0.8
TGACTGA	20083	0.9	TGACTCG	29270	1.3			
<b>TGACTCA</b>	20096	1.2	TGACTGA	29450	1.0			

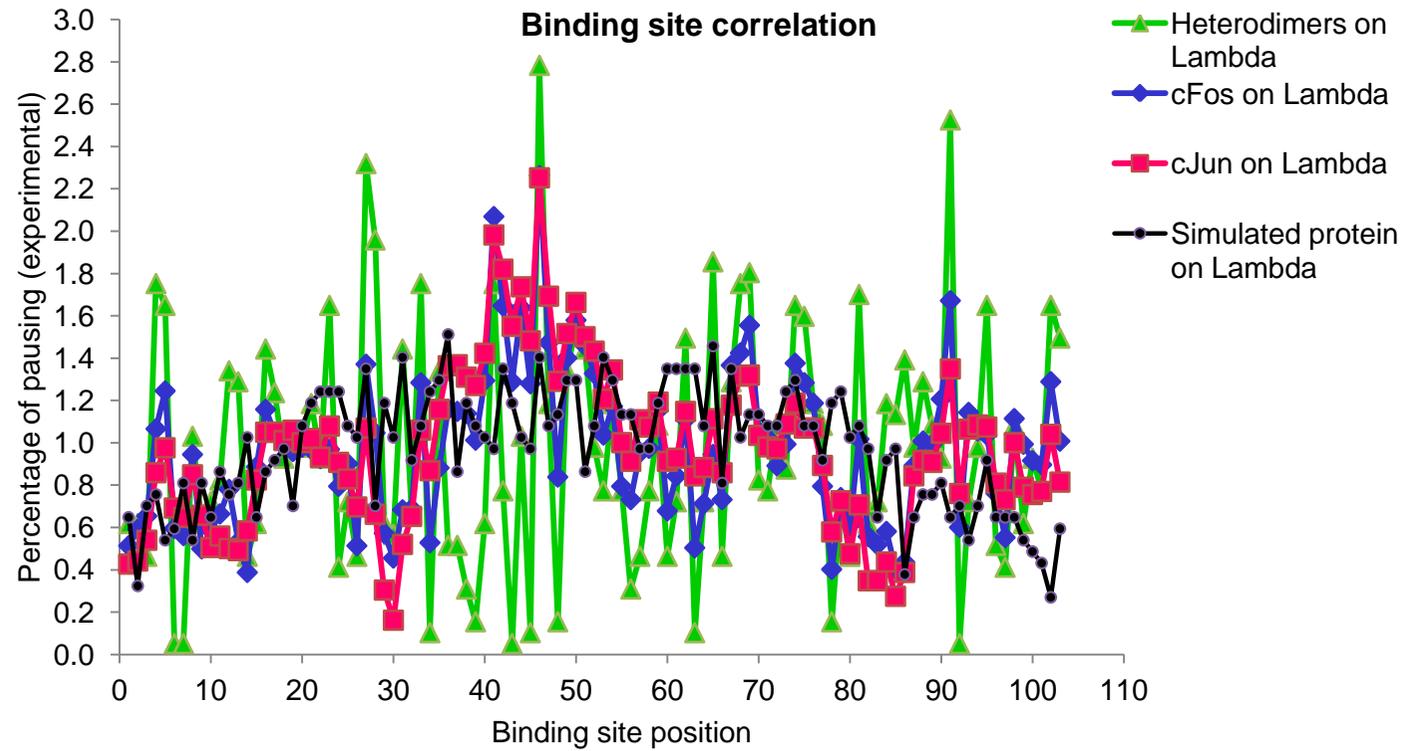
**Table 7:** A table showing the percentage of moving cJun (n = 17) molecules that bound to AP1 (highlighted yellow) and variant AP1 target sites as they travelled along  $\lambda$  DNA. Like cFos, 2.3% of all cJun molecules bound to the variant target site TGCCTCA (highlighted green), making this the most commonly visited site. 0.3-1.7% of all cJun molecules bound to their AP1 target sites.

cFos:cJun binding positions								
Binding site	Start position	%	Binding site	Start position	%	Binding site	Start position	%
TGACTCT	313	0.6	TGACGCA	20564	0.5	TAACTCA	29638	0.8
TCACTCA	708	0.6	TGACGCA	20962	0.5	TGACGCA	30321	1.1
TGACGCA	1545	0.5	TGACTAA	21030	0.3	TGAATCA	30593	0.9
TGGCTCA	1658	1.8	TGACTGA	21144	0.2	CGACTCA	30761	1.6
TGCCTCA	2465	1.6	<b>TGACTCA</b>	21153	0.6	TAACTCA	31170	1.6
TGACACA	2799	0.1	TGTCTCA	21420	1.8	TGACGCA	31344	1.2
TGACTGA	3092	0.1	TGACTCC	21657	0.8	TGACACA	32092	1.1
TAACTCA	3507	1.0	TCACTCA	21766	0.1	TGAATCA	32338	0.2
<b>TGACTCA</b>	3843	0.6	TCACTCA	21986	1.0	TGAATCA	32825	0.7
TGGCTCA	4995	0.7	TGACGCA	22056	0.1	TGAATCA	33563	0.5
AGACTCA	5352	0.8	<b>TGCCTCA</b>	22155	2.8	CGACTCA	33769	1.7
TGAATCA	7093	1.3	<b>TGACTCA</b>	22682	1.2	TGCCTCA	35793	0.6
TGACTGA	7097	1.3	AGACTCA	22764	0.2	<b>TGACTCA</b>	36560	0.7
TGTCTCA	8114	0.5	TGAATCA	22851	1.3	TGGCTCA	37410	1.2
TGACTTA	8919	0.6	TGGCTCA	22934	1.5	TGACTCC	38677	1.1
GGACTCA	9110	1.4	TGAATCA	23490	1.4	<b>TGACTCA</b>	41893	1.4
TGACCCA	9321	1.2	TGACTAA	23531	1.0	TGAATCA	43861	4.0
TGACTCT	9701	0.9	TGAATCA	23812	0.8	CGACTCA	43939	1.3
TGACTCC	9827	1.0	TGACTAA	23911	1.2	TGACTCT	43963	1.1
AGACTCA	10576	1.0	TTACTCA	24693	0.8	TGACTCT	44467	0.9
TGACTCT	10614	1.2	TCACTCA	24705	0.3	TGACTTA	44553	2.5
TGACTCG	11096	0.9	TTACTCA	25299	0.5	<b>TGACTCA</b>	44581	0.1
TCACTCA	11100	1.6	TGAATCA	25507	0.8	CGACTCA	44805	0.7
AGACTCA	11671	0.4	TGTCTCA	25603	1.1	TGACTCG	45437	1.0
TGACTCT	12058	0.7	GGACTCA	26157	0.5	TGACTCT	45972	1.6
TGACTGA	12289	0.5	TGACCCA	26191	0.7	TGACGCA	46132	0.5
TGGCTCA	12776	2.3	TGACTCG	26281	1.5	TGACTGA	46262	0.4
TGACACA	14134	2.0	TGACTGA	26744	0.1	TGGCTCA	47445	1.1
TCACTCA	14361	0.6	TGAATCA	27614	0.7	TAACTCA	47817	0.6
TGACCCA	16343	0.6	TGACTGA	27794	1.9	TGCCTCA	47884	0.9
TGGCTCA	18179	1.4	TGGCTCA	28285	0.5	TGACTGA	47938	0.8
<b>TGACTCA</b>	19119	0.7	TAACTCA	28718	1.3	TGACCCA	48180	1.6
TGGCTCA	19341	1.8	TGACTTA	28923	1.8	TGACGCA	48391	1.5
TGACTGA	20083	0.1	TGACTCG	29270	1.8			
<b>TGACTCA</b>	20096	1.3	TGACTGA	29450	0.8			

**Table 8:** A table showing the percentage of moving cFos:cJun (n = 7) molecules that bound to AP1 (highlighted yellow) and variant AP1 target sites as they travelled along  $\lambda$  DNA. 2.8% of the heterodimeric complexes commonly visited the variant AP1 target site TGCCTCA (highlighted green). 0.1-1.3% of the molecules visited their AP1 target sites.



**Figure 31:** This graph shows the correlation between the percentage of cFos ( $n = 26$ ), cJun ( $n = 17$ ) and cFos:cJun heterodimeric ( $n = 7$ ) moving molecules that paused within AP1 and variant AP1 target sites while travelling along  $\lambda$  DNA (as shown by Tables 6, 7 and 8). cFos and cJun molecules are visually similar. However cFos:cJun heterodimers do not follow the same correlation pattern.



**Figure 32:** A graph showing a comparison between cFos, cJun and cFos:cJun heterodimers that bound to each of the target sites on  $\lambda$  DNA with the simulated protein. No correlation can be seen when comparing the proteins with the simulated protein.

## **2.4: Discussion**

The way cJun and cFos:cJun bind to DNA and locate their AP1 target sites has not been studied before to our knowledge. Using a single molecule tightrope approach, this study provides a direct insight into how these transcriptional regulators search the DNA for their AP1 target sites.

### **2.4.1: cFos, cJun and cFos:cJun heterodimers interacting with $\lambda$ DNA**

It was important to understand how many proteins bound to DNA.  $\lambda$  DNA was used in this study as it is 48,502 bp long, was readily available and also contains 8 AP1 target sites. As expected, cJun and cFos:cJun bound to the DNA. To test the canonical view that cFos is unable to bind to DNA alone, cFos was loaded onto  $\lambda$ . Substantial binding was observed and cFos homodimers as dual labelled complexes were seen. This contradicted previous studies which showed that cFos cannot homodimerise but can only heterodimerise with cJun (Hai and Curran 1991, Schuermann et al. 1989, Woolfson and Alber 1995, Mason et al. 2006). The way cFos binds was the investigated.

This thesis provides direct visual evidence that cFos is capable of homodimerising on DNA, however the structural mechanism requires analysis of cFos DNA binding domain compared with cJun. In terms of arginine and lysine enrichment only, cFos is not too dissimilar to cJun. cFos however has 24 amino acid residues within its DNA binding domain (as shown in Figure 5). Three of which are arginine and seven lysine. cJun however has 22 amino acid residues: Four are arginine and seven are lysine. These positively charged amino acid residues enable electrostatic interactions with the negatively charged phosphate backbone of the DNA forming a protein-DNA complex. However, to fully understand how cFos interacts and binds to the DNA, x-ray crystallography studies should be performed to determine this protein-DNA interaction at a structural level (Dey et al. 2012, Rohs et al. 2010).

The percentage of cFos, cJun and cFos:cJun complexes that were observed travelling along  $\lambda$  DNA tightropes were quantified. cJun was the most mobile complex as 71% were observed travelling along the DNA. Although cFos was observed bound to DNA, 45% (t-test p-value = 0.03) of these proteins were also diffusing. Therefore, like cFos, performing x-ray

crystallography structural studies may give a greater perspective of not only how these proteins are able to bind to the DNA but also of how they move. From the colocalised heterodimeric complexes, 39% (t-test p-value = 0.06) were observed diffusing in comparison to cJun. Despite the limited cFos:cJun heterodimer sample size ( $n = 41$ ), a clear trend can be seen here although this result may not be statistically significant. This data implies that the cFos:cJun heterodimeric complex is so stable that it also becomes rigid and unable to bind and move along the DNA as efficiently as cJun. This finding suggests that residues inside of the leucine zipper region may be contributing towards how the protein is able to mobilise while bound to the DNA. Also, electrostatic interactions between arginine and lysine within the DNA binding domain with DNA are not completely responsible for protein movement.

FosW and FosW+cJun complexes that were observed travelling along the DNA tightropes were quantified. FosW, however has the same DNA binding domain as cJun that is enriched with arginine and lysine. It was therefore predicted to move in the same manner as cJun. However, only 55% of FosW were seen moving when compared with 71% mobile cJun. When FosW formed a heterodimeric complex with cJun, 39% (t-test p-value = 0.7) were observed moving and this is not statistically significant when compared with FosW alone. The leucine zipper region may be contributing towards the movement of the complex. This result does show that FosW is capable of locating and bind to its target cJun (Worrall and Mason 2011). Therefore, two things can be suggested. Firstly, as hypothesised, it was believed that FosW has a much stronger affinity for its target cJun, more so than cJun has for itself (Mason et al. 2006). Secondly, increasing the ratio of inhibitor to protein may enable 100% formation of FosW+cJun complexes.

#### **2.4.2: Protein colocalisation**

As described in section 1.3, it was previously believed that cJun is capable of homodimerising in addition to heterodimerising with cFos (Walters et al. 2014). This theory was confirmed as 27% cJun were observed as dual coloured complexes when studied alone.

As previously mentioned, cFos was believed to be incapable of homodimerising (Schuermann et al. 1989). This study clearly shows that 27% of cFos can homodimerise due to the observation of dual coloured complexes. This result therefore contradicts other studies which concluded that cFos is unable to homodimerise (Walters et al. 2014, Eferl and Wagner 2003). In addition, this result confirms the original proposition made by Landschultz *et al* in 1988 suggesting that cFos is capable of homodimerising.

When red labelled cFos and green labelled cJun were mixed together, all complexes observed as dual colour were quantified. This was important as 3 different coloured complexes were observed, that is, red labelled cFos, green labelled cJun and dual labelled cFos:cJun heterodimers by using the technique described in method 2.2.9. All proteins that showed colocalisation appeared dual coloured. Therefore, any observed dual coloured complexes were quantified as heterodimers. cFos:cJun heterodimers showed 32% colocalisation. This low colocalisation percentage (t-test p-value = 0.4) was unexpected as it was previously thought that cFos and cJun had the greatest affinity for each other out of all of the complexes formed (Mason et al. 2006, Nakabeppu et al. 1988).

#### **2.4.3: Interaction between FosW and cJun**

The antagonistic peptide inhibitor FosW was also loaded onto the DNA to observe binding events prior to quantifying how many complexes it formed with its target cJun. Visualising FosW is important due to its potential cancer therapeutic effects that could eventually lead to the development of a cancer fighting drug. Primarily, this peptide inhibitor was designed to target cJun, preventing it from homodimerising. FosW also has a more stable interaction with cJun than cJun has for itself (Mason et al. 2006).

Traditionally, this inhibitor does not contain a biotin tag or a DNA binding domain and differs structurally from wild-type cFos (Worrall and Mason 2011). However for the purpose of this study, a cJun DNA binding domain was attached and cloned into the sequence enabling any binding events to be visualised and quantified. A cJun DNA binding domain was chosen as cJun is the most mobile protein and during the development of this inhibitor, it was believed that

cFos could not bind to DNA or homodimerise. This study has shown that FosW is also capable of binding to the DNA, which was expected due to its cJun DNA binding domain.

The amount of colocalisation observed between FosW and its target cJun was 25%. As a primary result, this suggests that the ratio of FosW to cJun needs to be increased, especially when cJun is in excess during tumorigenesis. The mechanism used by the AP1 transcription factors and the biological peptide inhibitor FosW when swapping partners is currently unknown.

#### **2.4.4: Movement along DNA**

Upon binding, cFos, cJun, cFos:cJun and FosW were able to randomly diffuse along the DNA 1-Dimensionally as indicated by their  $\alpha$  values. This has also been simulated by a computerised protein taking a random walk as shown in section 2.2.16. cJun diffuses along the DNA more than cFos (t-test p-value = 0.02). FosW diffusion is similar to cFos (t-test p-value = 0.5) although there is no statistical significance. This again coincides with the possibility that core residues inside of the leucine zipper regions of cFos and FosW may determine how the protein is able to bind and move along DNA. However, when FosW was compared with cJun (t-test p-value = 0.01), as both have the same DNA binding domain, these are statistically significant. These findings suggest that cFos, cJun, cFos:cJun and FosW use 3-Dimensional diffusion to locate and bind to a non-specific part of the DNA sequence. Once bound, these proteins and the inhibitor then use 1-Dimensional diffusion to search the DNA for their target sites (Kampmann 2005).

As the proteins search the DNA, they may be using macroscopic association and dissociation events, such as jumping. This diffusion mechanism not only expands the proteins search radius over a greater distance, it also prevents the proteins from unnecessarily searching stretches of DNA that is target free. When a protein jumps, it uses 3-Dimensional diffusion to bind to a site further away on another part of the DNA (Kampmann 2004, Berg et al. 1981, Von Hippel 1994, Halford and Marko 2004). In contrast to macroscopic binding events such as jumping, the proteins may also be using a series of microscopic association and dissociation events such as hopping. The protein will either re-associate itself with the DNA at a site close by that it has a high affinity for, or it will dissociate and reassociate itself with the same area of binding

(Kampmann 2004). Jumping is one mechanism used by the proteins as they use 1-Dimensional diffusion to randomly search along the DNA backbone for the AP1 target sites.

#### **2.4.5: Pausing along the DNA**

cFos, cJun and cFos:cJun heterodimers were all observed to bind non-specifically along the DNA and pause. To investigate why this occurs, a Binding Site Position Finder was generated. This determined the possibility of where cFos, cJun and cFos:cJun heterodimers paused during their travel along the DNA, even though there was no defined sequence context to work on. Less than 2% of all proteins bound to each AP1 target site present along  $\lambda$  DNA. This was an unexpected result and contradicts studies that suggest cJun and cFos:cJun heterodimers are able to bind directly to AP1 only (Jochum et al. 2001, Leaner et al. 2003, Mason et al. 2006). However, cFos is capable of locating its AP1 target site and binding to it. Now the question remains as to where the proteins are pausing in addition to their AP1 target sites. Based on the Seldeen *et al* (2009) study, as described in section 2.1, cFos, cJun and cFos:cJun heterodimers paused within all of the variant AP1 target sites and this varied from 0.1-2.8%. The most commonvariant AP1 target site was TGCCTCA for all 3 proteins. It was previously reported that cFos:cJun heterodimers have a high affinity for these variant AP1 target sites (Seldeen et al. 2009). However this study also shows that cFos and cJun are capable of binding to this particular sequence. The relative binding of the proteins in two or more target sites provides a fingerprint for which target site sequences the proteins pause within.

This data generated from the Binding Site Position Finder is important as it suggests that the proteins may be exerting alternative biological responses within the variant AP1 target sites in addition to binding to their AP1 target sites. This data is also in keeping with the “waiting room model.” This can be defined as target areas of DNA that the proteins have a high affinity for in addition to AP1. The proteins may intermittently bind and pause within these target areas until an AP1 target site becomes available, resulting in the exertion of their biological function. This may be a way of controlling transcription.

#### **2.4.6: Conclusions**

Within this Chapter, it can be concluded that cFos is capable of homodimerising. This previously unseen observation was shown by the colocalisation study. Further structural studies will need to be performed in order to understand how, although this most likely occurs at the dimerisation interface between the two cFos proteins leucine zipper region. Similarly, the antagonistic peptide inhibitor FosW can also bind to its target protein cJun as expected. In the future, a study could be performed to understand what ratio of the inhibitor FosW is required to form complete complex formation with its target cJun when cJun is in excess using OAF microscopy.

All cFos, cJun, cFos:cJun, FosW and FosW+cJun complexes are capable of binding to DNA and undergoing 1-Dimensional sliding. FosW has the same diffusion constant as cFos, even though it contains a cJun DNA binding domain. This may be due to core residues inside of the leucine zipper region controlling DNA binding. Further studies will need to be performed to gain a greater understanding of how this occurs.

Finally, <2% of cFos, cJun and cFos:cJun heterodimers bind to their AP1 target sites and binding to each variant AP1 target site varies between 0.1-2.8%. Further studies could be performed in the future to understand why these proteins pause within these variant sites and whether this is a mechanism of controlling transcription, as described by the “waiting room model.”

### **Chapter 3: cFos, cJun and cFos:cJun interaction with different DNA substrates**

This chapter aims to study how cFos, cJun, cFos:cJun heterodimers and FosW +/- cJun interact with pUC19, pUCap1 and TF $\lambda$  DNA tightropes while being visualised using OAF microscopy. The proteins mode of diffusion, diffusion constant and  $\alpha$ -value will be determined using MSD analysis from data obtained from the proteins kymographs. Any colocalisation observed between AP1 family members will also be determined based on whether the proteins have formed homodimers or heterodimers once bound to the DNA.

An additional aim in this chapter is to determine where the proteins bind and pause along the DNA tightropes using Gaussian distribution plots and the Binding Site Position Finder. This approach will help to identify whether the proteins are just pausing within the AP1 target sites or whether they are also pausing within the variant AP1 target sites as well.

#### **3.1: Introduction**

Confirming previous reports, it was shown in the previous chapter that cJun can homodimerise or heterodimerise with cFos. cFos can also homodimerise which has previously not been seen. cFos, cJun and cFos:cJun heterodimers can locate and bind to areas of non-cognate DNA that they have a high affinity for using 3-Dimensional diffusion. Once bound, the proteins use random 1-Dimensional diffusion in search of their AP1 target site TGACTCA where they can exert their biological function that initiates cellular proliferation, differentiation and apoptosis (Jochum et al. 2001, Leaner et al. 2003, Campbell 2005), as described in Chapter 2. These cellular functions become initiated through the MAP kinase cellular signalling pathway (Shaulian and Karin 2001, Brown 2007). However, when the p53 gene becomes antagonised downstream, this causes the suppression of apoptosis and the survival of tumour cells due to the cell cycle being disrupted (Brown 2007, Shaulian and Karin 2001). As a result, cFos causes osteosarcomas (Wagner and Eferl 2005, Wang et al. 1995, Karin et al. 1997) and cJun initiates hepatocellular carcinoma and squamous cell carcinoma (Zhang et al. 2006), as described in Chapter 1. The mechanism used during the proliferation of tumorigenesis is currently unknown (Wagner 2001).

Chapter 2 has also demonstrated that while the proteins did search the DNA, less than 3% of cFos, cJun and cFos:cJun heterodimers were found pausing within their AP1 target sites. This result disputed bulk-phase experiments, such as EMSA studies, that can only quantify the ratio of protein-DNA binding events (Orchard and May 1993). Even though non-specific pausing was observed in Chapter 2, this was also observed within this chapter when the proteins interacted with pUC19 and pUCap1 Target-Free  $\lambda$  (TF $\lambda$ ) and DNA substrates. To understand how the proteins diffused and where they were pausing, the fore-mentioned DNA substrates were used. Based on the study by Seldeen *et al* (2009), all single nucleotide substitutions within the AP1 target sequence, known in this study as variant AP1 target sites, were identified (as described in Chapter 2 section 1.6). The locations of each of these variant AP1 target sites along the fore-mentioned DNA substrates were determined. As pUC19 only contains variant AP1 target sites (pUCap1 is composed of pUC19 with the addition of a single AP1 target site), any pausing may be in keeping with the “waiting room model” as described in Chapter 2 section 1.8. To determine where the proteins were pausing, an adapted binding site position finder was used (original model described in Chapter 2 section 2.218) to incorporate all variant AP1 target sites present on pUC19 and pUCap1 DNA substrates. However, TF $\lambda$  DNA is both variant and AP1 target site free. The rationale here is to determine how cJun will diffuse along the DNA in the absence of these target sites.

In this work, some of these variant AP1 target sites, were also integrated alongside the AP1 target sites. Any additional pausing was identified as being within these variant AP1 target sites. Therefore, pausing within these variant AP1 target sites may be a mechanism used to by the proteins to control transcription, described as the “waiting room model.” This is where the proteins may wait until an AP1 target site becomes available. Alternatively, cFos and cJun may be exerting their biological function within these variant AP1 sites. It has previously been suggested that the cFos:cJun heterodimers may be regulating specific genes within these variant AP1 target sites (Seldeen et al. 2009). This is currently unknown, however the research here could lead to the establishment of these sites as important in cellular pathways that control the transcription of other genes in addition to the biological functions already controlled by AP1.

The aim of this chapter is to perform some of the same experiments, as mentioned in Chapter 2, to determine whether partner swapping still occurs on different types of DNA and how cFos, cJun and cFos:cJun heterodimers interact with pUC19, pUCap1 and Target-Free  $\lambda$  (TF $\lambda$ ) DNA substrates. The plasmid pUC19 is 2686 bp long and does not contain any AP1 target sites. However it does contain 11 variant AP1 target sites. pUCap1 DNA is a variation of pUC19 whereby a single AP1 target site has been cloned at position 417 of the plasmid. These DNA substrates were used to understand whether the absence of AP1 target sites hindered or promoted protein movement. The final type of DNA used was TF $\lambda$ , a 962 bp section of  $\lambda$  DNA that is free from AP1 and variant AP1 target sites. It was therefore hypothesised that cJun could continuously move along the DNA without pausing in search of its target sites.

This chapter shows that cFos, cJun and cFos:cJun heterodimers interact with different DNA substrates. It was observed that when cFos, cJun and cFos:cJun heterodimers interacted with pUC19 and pUCap1 DNA, the proteins bound, moved and paused within the AP1 and variant AP1 target sites. An unexpected observation was cJun was mostly static when interacting with TF $\lambda$ .

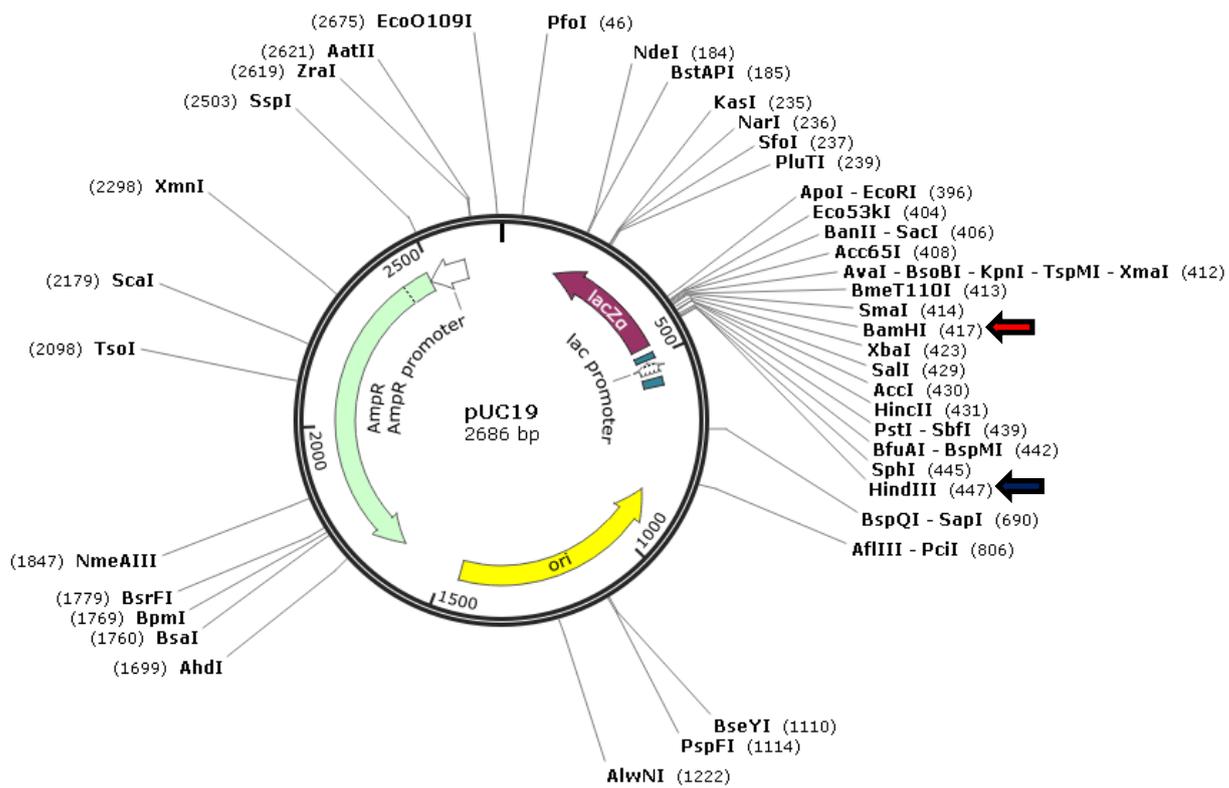
## 3.2: Materials and methods

### 3.2.1: Digestion of pUC19

pUC19 is a commercially available small, circular DNA plasmid that is 2,686 bp in length. This plasmid provides resistance to the antibiotic Ampicillin (Amp<sup>1</sup>). Within the plasmid there is a Multiple Cloning Site (MCS) that contains many restriction endonuclease cleavage sites. There are no WT AP1 target sites present in the pUC19 sequence. Therefore, this makes pUC19 a good negative control when comparing protein movement on DNA with and without the presence of AP1 target sites.

The restriction endonuclease BamHI (10 u/μL), was chosen to cleave pUC19 at its 5'- GGATCC -3' recognition sequence (indicated by the red arrow at position 417 bp in Figure 33), forming a 'sticky' 5'- cohesive overhang of 5'- GATC -3'. BamHI has no star activity meaning it will not cleave at sites other than its recognition sequence. The restriction digest mixture was left to incubate for 2-16 hours at 37 °C and heat inactivated at 80 °C for 20 minutes (ThermoFisherScientific 2015). The incubation time depended on the amount of DNA to be cleaved as 1 unit of restriction endonuclease can digest 1 μg of DNA in 1 hour (Fermentas).

The restriction endonucleases BamHI (10 u/μL) and HindIII (10 u/μL) cleaved DNA simultaneously (indicated by the blue arrow. Recognition sequence is 5'- AAGCTT -3' as shown in Figure 33) and their compatibility and efficiency with each other was ensured to be between 50-100% within the same digestion buffer. Like single cutters, dual cutting reactions were incubated for 2 - 16 hours depended on the efficiency of each cutter and the amount of DNA to be cleaved.



**Figure 33:** A circular plasmid map of Amp<sup>r</sup> resistant pUC19. Recombinant pUC19 is a high copy number cloning vector that is 2,686 bp in length. Most pUC19 restriction sites are located within the MCS region. The blue arrow shows HindIII restriction site at position 447 and the red arrow shows BamHI restriction site at position 417. pUC19 contains no AP1 target sites.

### 3.2.2: AP1 oligonucleotide design

In order to introduce a single WT AP1 target site with the recognition sequence 5'-TGACTCA-3' into pUC19 (that is WT target site free), an oligonucleotide was designed using an oligonucleotide properties calculator (Northwestern.University 2007). The oligonucleotide is 27 bp dsDNA designed to incorporate a WT AP1 target site in the middle (shown in red), a HindIII restriction site (shown in green) and a BamHI restriction site (shown in blue). Each restriction site contains a 4 base overhang. This enables ligation to occur with another complementary sequence due to being phosphorylated at its 5'- end. Shown in black is a sequence of non-cognate DNA between the AP1 target site and restriction sites. These additional bases increase the length of the oligonucleotide and its melting temperature. The oligonucleotides were designed so that they were completely complementary to each other and had a low probability of intramolecular hairpin formation.

As there are 2 different restriction sites at either end of the oligonucleotide, this results in a 50% chance of forming tandem repeats with neighbouring pUC19 DNA strands with complementary ends. Successful ligation also depends on the correct orientation of the oligonucleotide. By having different restriction sites, this will reduce the chance of the oligonucleotides forming concatemers by 50%.

In order to construct the oligonucleotide (shown in Figure 34), 20  $\mu$ L of 100  $\mu$ M sense and antisense oligonucleotides (each oligonucleotide was reconstituted in sterile H<sub>2</sub>O) were heated up to 95 °C for 5 minutes in a sterile microcentrifuge tube causing the strands to separate. These volumes were used as during heating some condensation can occur, reducing the final volume. The oligonucleotides were then slowly cooled to room temperature, enabling the oligonucleotides to self anneal (Sigma-Aldrich 2015). The newly annealed oligonucleotides were centrifuged for 5 seconds at 12,300 xg (xg calculated using the formula:  $G = 1.18 \times R \times \left(\frac{\text{RPM}}{1000}\right)^2$ . G is the G-force, R is the radius of rotation in mm and RPM is revolutions per minute) to ensure the solution was at the bottom of the microcentrifuge tube before storing at 4 °C. To avoid

repeated freeze-thaw cycles, 20  $\mu\text{L}$  of each oligonucleotide was removed from the original stocks and aliquoted into sterile microcentrifuge tubes. These were stored at  $-20\text{ }^{\circ}\text{C}$  until used.

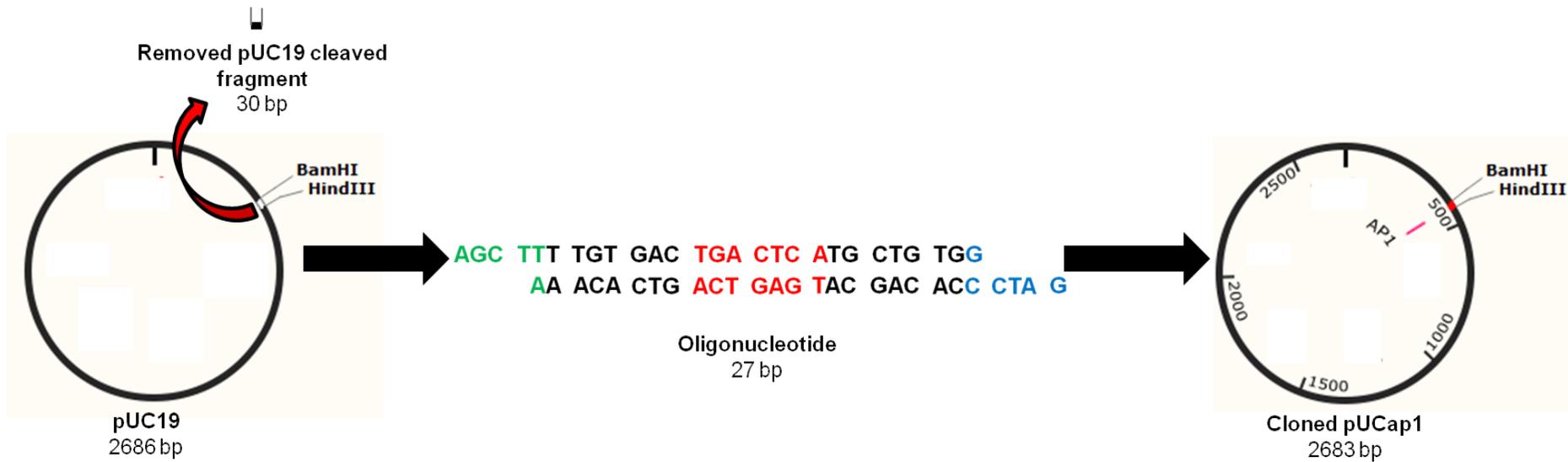
AGC    TTT    TGT    GAC    TGA    CTC    ATG    CTG    TGG  
         AA    ACA    CTG    ACT    GAG    TAC    GAC    ACC    CTA    G

**Figure 34:** A 27 bp self annealed oligonucleotide that contains a HindIII restriction site (green), a BamHI restriction site (blue) and a single AP1 target site (red). Each oligonucleotide contains no hairpins and cannot anneal to itself. The oligonucleotide is phosphorylated at its 5'- end.

### 3.2.3: Making pUCap1

pUCap1 was constructed (as shown in Figure 35) by first double-digesting pUC19 with BamHI (10 u/μL) and HindIII (10 u/μL) restriction endonucleases at 37 °C for 2-16 hours (ThermoFisherScientific 2015). To ensure that the double cleaved vector could not self anneal and ligate with itself during ligation with T4 DNA Ligase, phosphate groups from the 5'- end were removed using Fast Alkaline Phosphatase (FastAP. (ThermoFisherScientific 2015). This mixture was incubated for 15 minutes at 37 °C. FastAP was heat inactivated at 80 °C in accordance with the manufacturer guidelines. This temperature is higher than the 65 °C required to heat inactivate BamHI and HindIII. This meant that the restriction endonucleases were also inactivated.

In the heat inactivated digestion mixture, there are large 2656 bp long pUC19 fragments and small 30 bp fragments. It was imperative that these smaller fragments were removed using a PCR clean up kit (Qiagen), otherwise they would be competing against the larger pUC19 fragments for the donor AP1 oligonucleotide during the construction of pUCAP1 (Figure 34). After using the PCR clean up kit, this resulted in the 30 bp pUC19 fragment remaining only. To construct pUCap1, A 20 μL ligation mixture containing a 3:1 self annealed oligonucleotides to vector ratio respectively (based on their molarity ratio) was placed inside a sterile PCR tube and incubated in a PCR thermocycler (Techne TC-3000) at 16 °C for 15 minutes prior to adding T4 DNA ligase (Thermo Scientific). This ensured that the components within the ligation mixture were equilibrated to 16 °C. After the addition of T4 DNA ligase, the mixture was incubated at 16 °C overnight for 16 hours in a PCR thermocycler. To quench the reaction, a 10 μL stop buffer containing 2.4 M KCl was added to the reaction mixture. The KCl then diluted to 800 mM. This concentration can completely inhibit T4 DNA Ligase as shown in Figure 37, and quench the reaction. The newly created recombinant plasmid was transformed into Top10 Chemically Competent E. Coli (method 3.2.6).



**Figure 35:** A diagram showing the construction of pUCap1 (adapted from SnapGene), a circular DNA plasmid 2683 bp in length. Firstly, pUC19 was doubly cleaved with BamHI and HindIII restriction endonucleases followed by dephosphorylation with FastAP. This prevents plasmid religation and/or recircularisation upon ligation. The 30 bp cleaved fragment was removed using a PCR clean up kit. The self annealed donor 27 bp oligonucleotide (Figure 34) was inserted using a 3 oligos:1 vector ratio and ligated at 16 °C for 16 hours overnight using T4 DNA Ligase. The reaction was quenched with 800 mM KCl. The newly created recombinant plasmid was transformed into Top10 Chemically Competent *E. coli*.

### 3.2.4: Sequencing of pUCap1

To confirm that the correct pUCap1 construct had been made (as demonstrated in Figure 36), 100ng/ $\mu$ L was sent for sequencing (GATC) using one of their available primers, EGF2, a 17 base long primer with the sequence 5'-GGGGATGTGCTGCAAGG-3'. This primer (highlighted in green) was chosen as it annealed 82 bases away from the inserted AP1 oligonucleotide. This meant that the primer was far enough away from the insert, therefore avoiding nucleotide misinterpretation from the DNA polymerase and a high signal due to multiple primers binding to the same site (Abd-Elsalam 2003).

The sequence showed that a single AP1 oligonucleotide (highlighted in pink) was successfully inserted between BamHI and HindIII restriction sites.



**Figure 36:** Part of the sequence of pUCap1 showing the successful insertion of the AP1 oligonucleotide (shown by the pink arrow) between the HindIII and BamHI restriction sites. The EGF2 primer used during the sequencing reaction is shown in green. This primer annealed 82 bases away from the inserted oligonucleotide.

### 3.2.5: Making chemically competent cells

From an agar plate containing the antibiotic Amp<sup>1</sup>, a single colony that had grown for 16 hours at 37 °C was transferred to 100 mL of LB broth. The culture was incubated at 37 °C with shaking at 250 rpm until the cells reached an OD<sub>600</sub> of 0.4 - 0.6 (IKA KS i control shaking incubator). The cells were then transferred to sterile 30 mL centrifuge tubes and incubated on ice for 10 minutes. After incubation, the cells were pelleted by centrifugation at 4086 xg at 4 °C for 10 minutes. The supernatant was decanted and the cells were gently resuspended in 30 mL of ice-cold sterile MgCl<sub>2</sub> - CaCl<sub>2</sub> solution (80 mM MgCl<sub>2</sub> and 20 mM CaCl<sub>2</sub>). The magnesium and calcium together will cause *E. coli* to become competent, making it readily take up foreign DNA (Tsen et al. 2002). The cells were pelleted by centrifugation at 4086 xg at 4 °C for 10 minutes and the supernatant was decanted. The cell pellet was gently resuspended in 2 mL of ice-cold sterile 0.1 M CaCl<sub>2</sub>. The calcium present within the solution will counteract any electrostatic effects between the DNA and the cell membrane, increasing transformation efficiency (Suga and Hatakeyama 2003).

Prior to storage, 140 µl dimethylsulfoxide (DMSO) was added per 4 mL of cells and gently mixed. This is a cryopreservative agent and prevents water crystallising which could potentially damage and lyse the cells (Liseth et al. 2005). The cells were incubated on ice for 15 minutes before a further 140 µl DMSO was added and the cells were again incubated on ice for another 15 minutes. To prevent the cells from going through potentially damaging freeze-thaw cycles, 100 µl of suspension was aliquoted into sterile microcentrifuge tubes and immediately snap-frozen in liquid Nitrogen. All cell aliquots were stored at -80 °C until use.

### 3.2.6: Transformation of chemically competent *E. coli* cells

100 µL of chemically competent cells within a sterile microcentrifuge tube were thawed on ice for precisely 15 minutes. 10% (10 µL) of the cellular volume was made up of DNA (100 ng) and water. This was added to the cells and gently mixed, ensuring the cells were not damaged. This cell and DNA mixture was incubated on ice for 30 minutes and heat shocked at 42 °C using a heat block or water bath for precisely 30 seconds. This sudden increase in temperature allows pores within the *E. coli* plasma membrane to open, allowing the DNA to enter. Immediately

cooling the cells for 2 minutes after heat shocking closes these pores and the DNA cannot be released. Aseptic technique was used to transfer 900  $\mu\text{L}$  of sterile LB broth, a nutrient growth medium, to the cells prior to horizontal mixing using a shaking incubator at 37 °C for 75 minutes at 250 rpm. As these chemically competent cells are homemade, the incubation time was increased to compensate for a possible lower efficiency when compared with the commercially prepared cells. This incubation time enabled the cells to multiply. Agar plates containing 50  $\mu\text{g}/\text{mL}$  of Amp<sup>1</sup> were prepared aseptically. The plates were inverted and stored at room temperature until they were ready to be streaked.

Post incubation, a 'low concentration' plate was created. 75  $\mu\text{L}$  of cell suspension was pipetted onto a pre-made agar plate and evenly spread forming a monolayer. This was performed aseptically. For the 'high concentration' plate, the remaining suspension was centrifuged for 1 minute at 11,406 xg. The supernatant was decanted apart from 100  $\mu\text{L}$ . This remaining supernatant was used to resuspend the bacterial pellet. 75  $\mu\text{L}$  of cell suspension was pipetted and spread to create an even monolayer using the spreading and plate inverting techniques as mentioned above. The inverted plates were incubated at 37 °C overnight for 16 hours.

### **3.2.7: Preparing a bacterial culture**

After the chemically competent *E. coli* cells incubation at 37 °C had finished, a single isolated colony was chosen and collected using the end of a sterile pipette tip. The tip was placed into 10 mL of sterile L.B broth containing 50  $\mu\text{g}/\text{mL}$  Amp<sup>1</sup>. This culture was gently mixed overnight for 16 hours at 37 °C on a shaking incubator at 250 rpm.

### **3.2.8: Preparing a 1% (w/v) agarose gel**

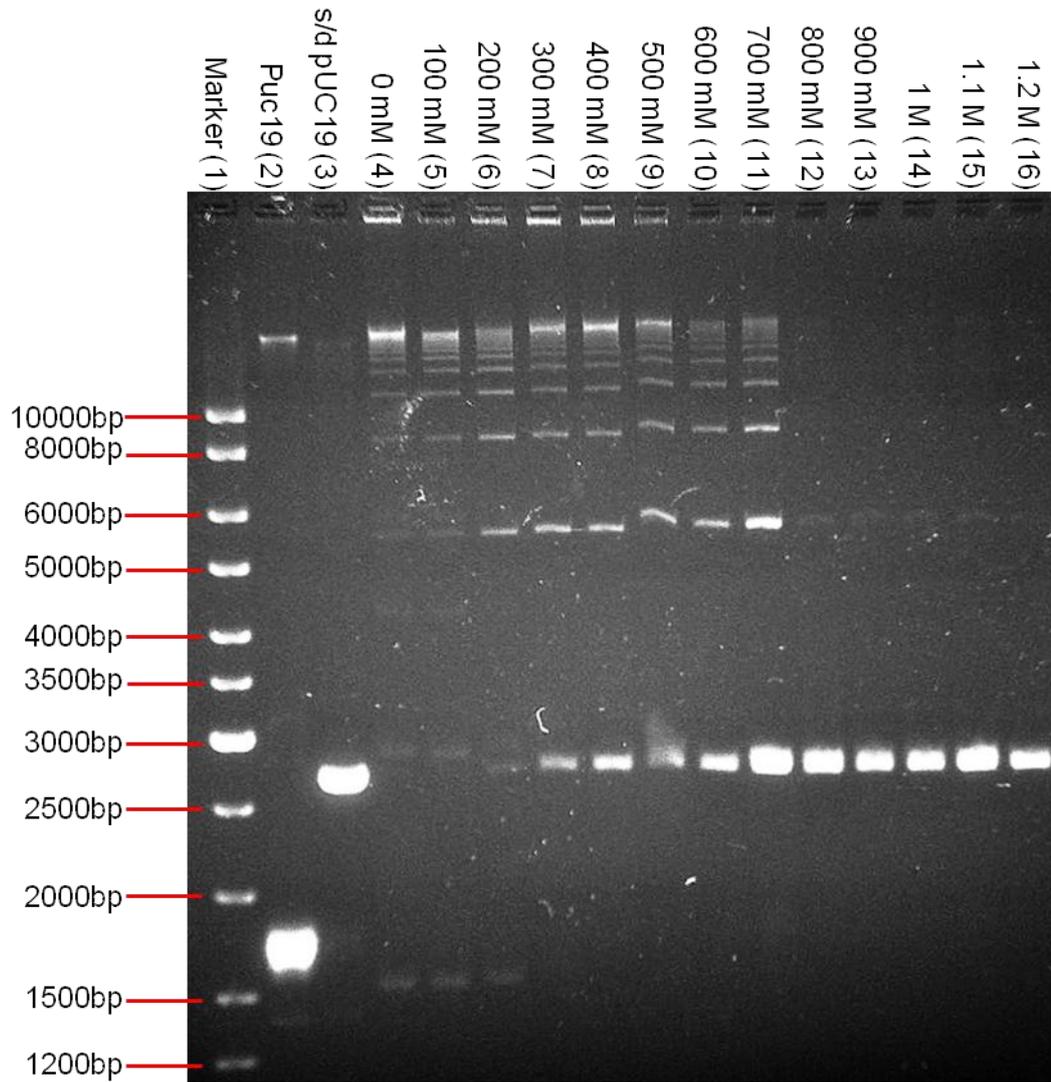
To resolve fragments between 500bp to 10kb, a 1% (w/v) standard agarose gel was used (Sambrook 2001). This gel percentage was used unless otherwise stated. Firstly, 1g of standard agarose powder was added to a conical flask containing 99 mL of 1x TAE buffer (40 mM Tris-acetate and 1 mM Ethylenediaminetetraacetic acid (EDTA)). The agarose mixture was gently heated in a microwave on a medium setting and swirled intermittently. This assisted the agarose powder to dissolve. Once the solution had cooled to 65 °C, Ethidium Bromide (EtBr) was added to give a final concentration of 0.5  $\mu\text{g}/\text{mL}$  and gently mixed before being poured into a pre-

taped gel tray containing a comb. EtBr was used as it intercalates in between the DNA base pairs and fluoresces under ultraviolet light (Reinhardt and Krugh 1978).

The agarose gel was allowed to completely set at room temperature before the tape was removed and the tray was submerged in 1x TAE buffer contained within a gel electrophoretic tank. Additional 1x TAE buffer was added to the tanks maximum fill lines. This ensured that the gel was completely covered with buffer. The gel comb was then carefully removed to ensure the wells were not distorted or damaged. Prior to loading DNA samples into each well, the DNA was gently mixed with homemade 1x bromophenol blue loading dye (1.66 mM Tris-HCl (pH 7.5), 0.03% (w/v) bromophenol blue, 10% (v/v) glycerol and 10 mM EDTA). This dye allowed the visual tracking of DNA migration during electrophoresis (Maniatis et al. 1975). The presence of glycerol enabled the DNA to sink to the bottom of the wells.

### **3.2.9: Inactivating T4 DNA Ligase using KCl**

Heat causes PEG<sub>6000</sub> to undergo conformational changes to its shape and transform from a free-flowing powder into a solid that has a wax-like consistency (Han et al. 1997). Therefore, an alternative method was found to inhibit T4 DNA Ligase using KCl. Various 20 µL ligation mixtures made within sterile PCR tubes were equilibrated to 16 °C for 15 minutes prior to the addition of T4 DNA Ligase. Each tube contained 350 ng of pUC19 cleaved with BamHI (10 u/µL), 15% (w/v) PEG<sub>6000</sub> (method 3.2.10 & Figure 38) and varying concentrations of KCl increasing from 0 M to 1.2 M in 100 mM increments (Figure 37). The aim was to find which KCl concentration quenched the reaction by completely inhibiting T4 DNA Ligase. Upon the addition of T4 DNA Ligase, each ligation mixture was incubated at 16 °C for exactly 30 minute. This ligation time was chosen based on the 1% (w/v) agarose gel results as shown in Figure 39. All samples were stored at -20 °C until use. The 1% (w/v) agarose gel shows that at a KCl concentration of 800 mM, ligation stopped. This means that 800 mM KCl is a sufficient concentration to completely inhibit T4 DNA Ligase.

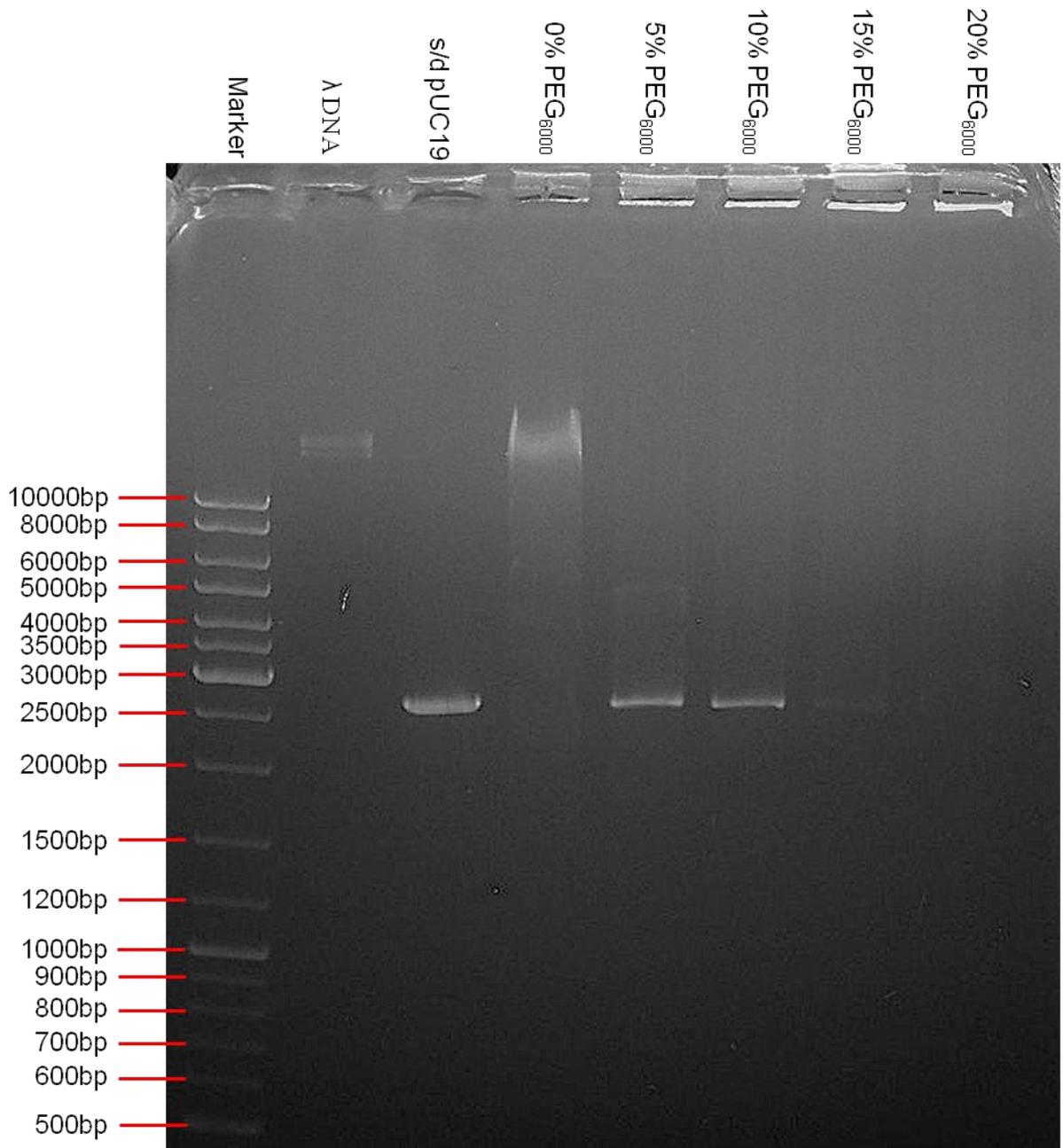


**Figure 37:** A standard 1% agarose gel image showing single cleaved pUC19 (with BamHI) that was ligated with T4 DNA ligase in the presence of 15% PEG<sub>6000</sub> (Figure 38) and varying concentrations of KCl. The same stock of single cleaved pUC19 was used. From this, 13 aliquots, each containing 350 ng of DNA were ligated in the conditions mentioned above for 30 minutes. This time was chosen due to the results shown from the concatemerisation of pUC19 at different time points gel (Figure 39). Lane 2 shows undigested pUC19 and lane 3 shows cleaved pUC19. Lanes 4-11 shows various species of concatemered pUC19 that formed in the presence of 0-700 mM KCl, some of which are comparable with  $\lambda$  DNA. Lanes 12-16 show that at 800 mM KCl, ligation stops. This means that 800 mM KCl is a sufficient concentration to completely inhibit T4 DNA Ligase.

### 3.2.10: PEG<sub>6000</sub> concentration for optimal DNA ligation

Polyethylene Glycol<sub>6000</sub> (PEG<sub>6000</sub>), a high molecular weight macromolecule that was used to promote the end-to-end ligation of single cleaved pUC19 by crowding around the DNA (Zimmerman and Pfeiffer 1983). PEG<sub>6000</sub> is not incorporated into the reaction but does act as background macromolecules. To identify which percentage of PEG<sub>6000</sub> was required to promote the concatemerisation of pUC19, a concentration gradient was performed.

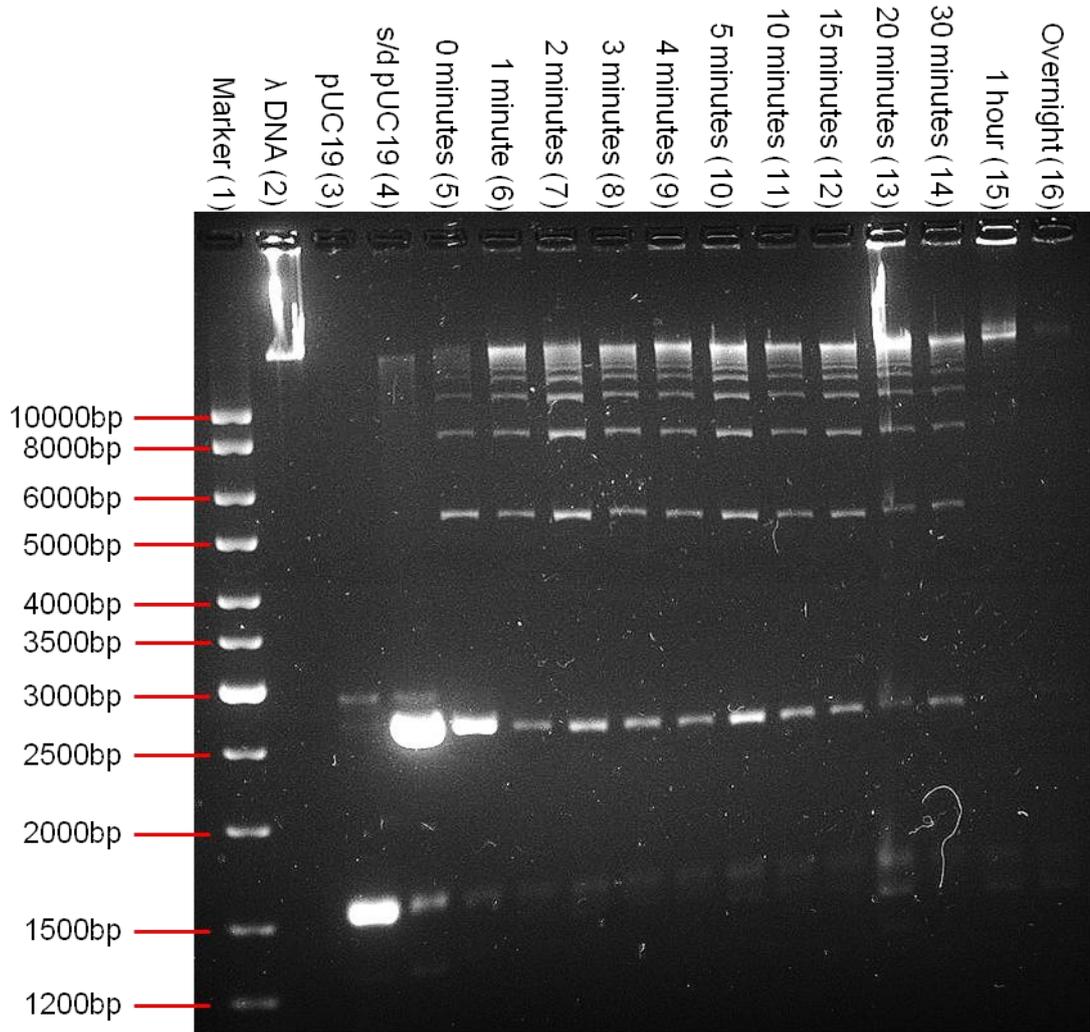
A 1 mL 50% (w/v) stock of PEG<sub>6000</sub> was made and stored at room temperature by adding PEG<sub>6000</sub> crystals to water until dissolved by vigorous vortexing. In a final volume of 20 µL each, several ligation mixtures were made containing 1.5 µg of single cleaved pUC19 and varying percentages of PEG<sub>6000</sub> within sterile PCR tubes. These percentages (w/v) were 0%, 5%, 10%, 15% and 20%. Prior to the addition of T4 DNA Ligase, the ligation mixtures were equilibrated to 16 °C for 15 minutes to ensure that each component was at this temperature. 5 Weiss U/ µL of T4 DNA Ligase was added and the ligation mixtures were incubated overnight for precisely 16 hours. After incubation, as PEG<sub>6000</sub> cannot be heated, ligation was terminated by adding a 10 µL of 3x KCl stop buffer (2.4 M KCl and water). 800 mM KCl (as shown in Figure 37) is a sufficient concentration to completely chemically inhibit T4 DNA Ligase. The samples were then immediately stored at -20 °C until use. The analytical gel in Figure 38 shows that 15% PEG<sub>6000</sub> is the optimum percentage to use as this generated the longest DNA concatemers.



**Figure 38:** A standard 1% (w/v) agarose gel image showing single cleaved pUC19 (with BamHI) that has been ligated for 16 hours at 16 °C with T4 DNA ligase in the presence of varying concentrations of PEG<sub>6000</sub>. Lane 2 shows λ DNA. This served as a size marker. Lanes 3-8 show pUC19 that was ligated together while in 0%, 5%, 10%, 15% and 20% (w/v) PEG<sub>6000</sub>. These PEG<sub>6000</sub> concentrations were used to identify which concentration promoted DNA concatemerisation that was comparable with the length of λ DNA or longer. 15% (w/v) PEG<sub>6000</sub> was chosen as the DNA remains in the well, meaning that it is longer than λ. This DNA substrate serves as a negative control as there are no WT AP1 target sites present.

### **3.2.11: A time course to create pUC19 tightropes that are comparable with the length of $\lambda$ DNA**

To identify the time duration needed to ligate pUC19 so that it could concatemerise to the same length as  $\lambda$  DNA which is 48,502 bp in length, a time course experiment was performed. 12 separate ligation mixtures containing 1.5  $\mu\text{g}$  of single cleaved pUC19 (with BamHI (10  $\text{u}/\mu\text{L}$ )) were firstly incubated in sterile PCR tubes at 16  $^{\circ}\text{C}$  in the presence of 15% (w/v) PEG<sub>6000</sub>. The mixtures were left to equilibrate for 15 minutes inside of the PCR thermocycler. Figure 39 shows the time course used to determine which time proved sufficient to form pUC19 concatemers that are comparable with the length of  $\lambda$  DNA. 5 Weiss U/  $\mu\text{L}$  T4 DNA Ligase was added to each ligation mixture. After each time duration had been reached, T4 DNA Ligase was chemically inactivated with 800 mM KCl (as shown by Figure 37). The zero time point was created by directly adding T4 DNA Ligase to the 1x KCl stop buffer which was then immediately added to the ligation mixture. The results of the level of ligation at each time point can be seen in Figure 39 and shows that a minimum of 30 minutes is sufficient time for pUC19 to ligate and form concatemers that are comparable with  $\lambda$  DNA.



**Figure 39:** A 1% (w/v) agarose gel showing the concatemerisation of 2686 bp long pUC19 at different time points using T4 DNA Ligase. Lane 2 shows  $\lambda$  DNA. This was used as a size marker. Lane 3 shows pUC19 plasmid. Lane 4 shows pUC19 that was cleaved with BamHI. Lanes 5-16 show the time course. At precisely each time point, T4 DNA Ligase was inhibited with 800 mM KCl (method 3.2.9). During the time course, different species of DNA can be seen, ranging from 5.5 kb (2mer) to 48 kb (18mer). This gel shows that a ligation time of 30 minutes is optimal (see lane 14).

### **3.2.12: Creating pUC19 concatemers.**

DNA Tightropes were created using pUC19 that was cleaved once with BamHI (10 u/μL) as described in method 3.2.1. This type of DNA was chosen as it does not contain any WT AP1 target sites. This results in pUC19 serving as a negative control when compared to λ DNA that contains 8 WT AP1 target sites (as shown in Chapter 2, Figure 11).

To create pUC19 DNA tightropes, firstly T4 DNA Ligase was used to ligate 1.5 μg of pUC19 for 30 minutes at 16 °C while in the presence of 15% (w/v) PEG<sub>6000</sub>. This was performed inside a PCR thermocycler. T4 DNA Ligase was added after the DNA mixture was equilibrated to 16 °C for 15 minutes, ensuring that each component was at this temperature. These optimal conditions were chosen based on the results shown in the agarose gels represented in Figures 36 and 37. At precisely 30 minutes, T4 DNA Ligase was inactivated using 800 mM KCl. This KCl concentration is needed to completely inhibit T4 DNA Ligase as shown by the agarose gel in Figure 37. Next, 500 ng of ligated pUC19 containing 800 mM KCl (total volume of 10 μL) was added to 1x ABC buffer as it contains 50 mM KCl (total volume of 90 μL), making a total volume of 100 μL. This buffer was chosen as the final KCl molarity reduced from 800 mM to 125 mM, less than the high salt buffer used for imaging that contains 150 mM KCl. The adapted DNA tightrope method, as described in method 3.2.13, was used to create pUC19 tightropes inside a flow cell chamber.

### **3.2.13: Creating pUC19 tightropes.**

This method is an adaptation of Chapter 2's method 2.2.7.

Firstly, beads were introduced into the flow cells chamber. The chamber was then equilibrated with 50 mM KCl by flushing through with 1x ABC buffer as described in method 2.2.5.1. Next, a 10 μL volume containing 500 ng of newly ligated pUC19 within 800 mM KCl (as described in method 3.2.9) was added to 90 μL of 1x ABC buffer that contained 50 mM KCl (as previously mentioned). This made a final KCl molarity of 125 mM. Once the DNA was introduced into the chamber, it became equilibrated with 125 mM KCl. This KCl concentration still remains less than the 150 mM KCl concentration used during imaging.

The method as described in 3.2.12 was followed to create pUC19 tightropes. However, the 1x ABC buffer previously used during bi-directional flow during the creation of tightropes was replaced with high salt buffer instead of low salt buffer. The high salt buffer equilibrated the chamber to 150 mM KCl. High salt buffer was also used during DNA labelling with YOYO-1 dye (as shown in method 3.2.14.2) and to flush the chamber prior to imaging.

### **3.2.14: Buffers**

#### **3.2.14.1: High salt buffer containing ATP**

T4 DNA Ligase (ThermoFisherScientific 2015), an Adenosine Triphosphate (ATP)-dependent enzyme, is able to form an enzyme-DNA complex upon binding to DNA (Wilkinson et al. 2001). In order for this complex to form, ATP is utilised by the enzyme. If the amount of ATP present is less than the amount required to complete its self-adenylation reaction, this can result in the reaction slowing down, eventually causing the enzyme to stall on the DNA (Rossi et al. 1997).

T4 DNA Ligase is used with a 1x reaction buffer containing 25 mM Tris-HCl (pH 7.6), 5 mM MgCl<sub>2</sub>, 0.5 mM ATP, 0.5 mM DTT and 2.5% (w/v) PEG<sub>8000</sub> (ThermoFisherScientific 2015). 1 mM ATP was added to the high salt buffer. By maintaining the ATP concentration at 1 mM, this will ensure that the enzymes adenylation reaction can complete. This means that the enzyme will be free from stalling on the DNA and the enzyme-DNA complex can readily dissociate.

#### **3.2.14.2: High salt buffer containing YOYO-1**

To observe the formation of TFA DNA (method 3.2.17) tightropes inside the flow cell chamber using OAF microscopy, high salt buffer (as described in method 2.2.5.1) was adapted by substituting 20 µL water with YOYO-1 dye to give a 1:50 YOYO-1 iodide dye to buffer ratio. This made a final YOYO-1 dye concentration of 20 nM. This buffer shall be referred to as high salt buffer + YOYO-1.

### **3.2.15: Phenol:Chloroform extraction and DNA precipitation**

During protein translocation along DNA, it appeared that adding ATP to the high salt buffer did not aid the complete dissociation of T4 DNA Ligase from the DNA as protein movement appeared inhibited (as shown in Table 9). Therefore, a Phenol:Chloroform extraction (PCE) and

precipitation of DNA (Palmiter 1974) method was performed as this would completely remove T4 DNA Ligase from the DNA, allowing proteins free movement again.

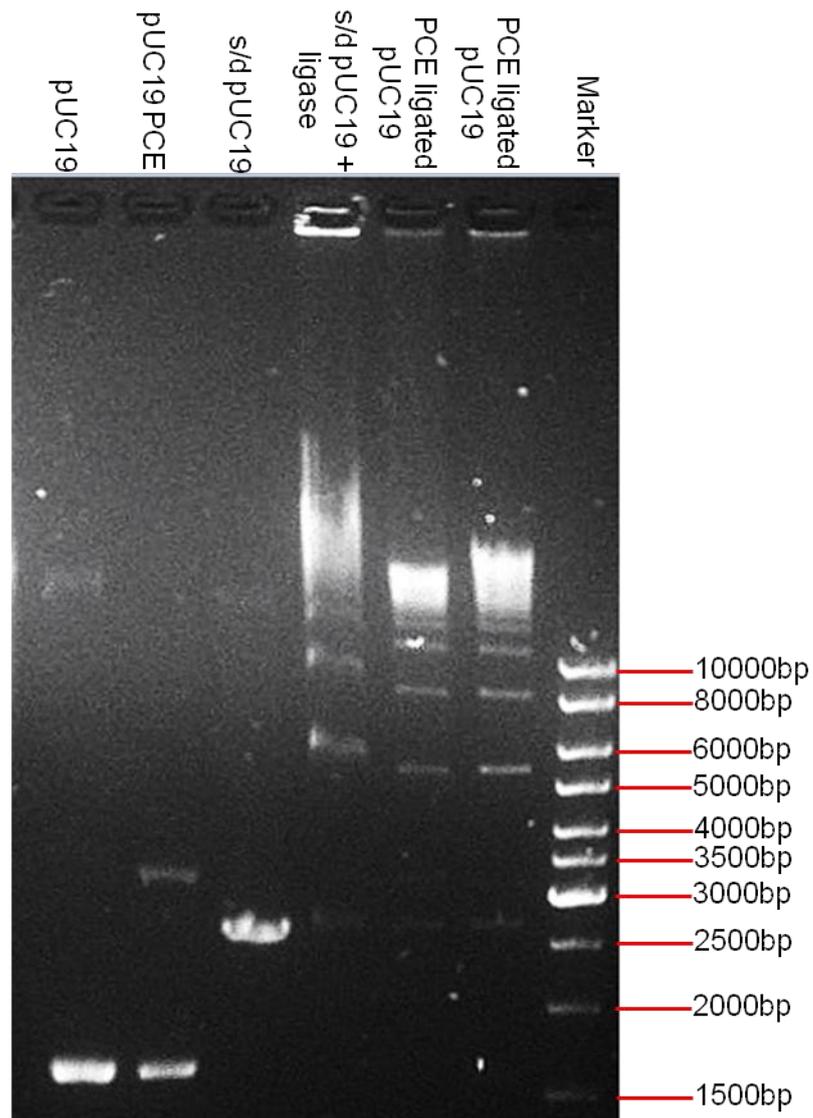
Firstly, a mixture containing equal volumes of Phenol and Chloroform were prepared in a microcentrifuge tube, making a final volume of 200  $\mu\text{L}$ . In another microcentrifuge tube, a 40  $\mu\text{L}$  volume containing 3  $\mu\text{g}$  of newly ligated DNA was added to 160  $\mu\text{L}$  of water. This made an equal volume to the PCE mixture. These volumes were chosen to ensure pipetting accuracy.

Next, the PCE mixture was vigorously vortexed for 15 seconds to ensure that the Phenol and Chloroform had thoroughly mixed. Immediately after vortexing, the DNA was added. The PCE + DNA mixture was then placed on a rotator for precisely 1 hour. The rotator is a custom made device whereby a ~3 inch diameter Perspex wheel was attached to a motor and clamped at a 39° angle using a clamp and stand. The wheel was set to slowly turn at 43 rpm and when combined with the tilt, this aided gentle sample mixing. This method for mixing was chosen as long DNA is prone to fragmenting and shearing, especially during vortexing. Gently inverting the sample in the afore-mentioned way prevented the DNA from becoming damaged. The mixture was separated into 3 layers: the aqueous layer (top), interphase (middle) and organic layer (bottom) by centrifugation at 11,406 xg for 1 minute. If the layers had not completely separated, the mixture was recentrifuged. The top aqueous layer that contained the DNA was aspirated and placed in a microcentrifuge tube and its volume measured. An equal volume of Chloroform was then added. The Chloroform and DNA mixture was then gently mixed on the rotator for 1 hour. After mixing, the Chloroform and DNA mixture was phase-separated into 3 layers by centrifugation at 11,406 xg for 1 minute. If the layers had not fully formed, this step was repeated. The top aqueous layer containing the DNA was aspirated, placed in a microcentrifuge tube and its volume measured. 0.111 volumes of 3 M sodium acetate at pH 5.2 and 2.5 volumes of ice-cold 100% (v/v) ethanol was added to the DNA. This mixture was incubated on ice for 30 minutes to precipitate the DNA.

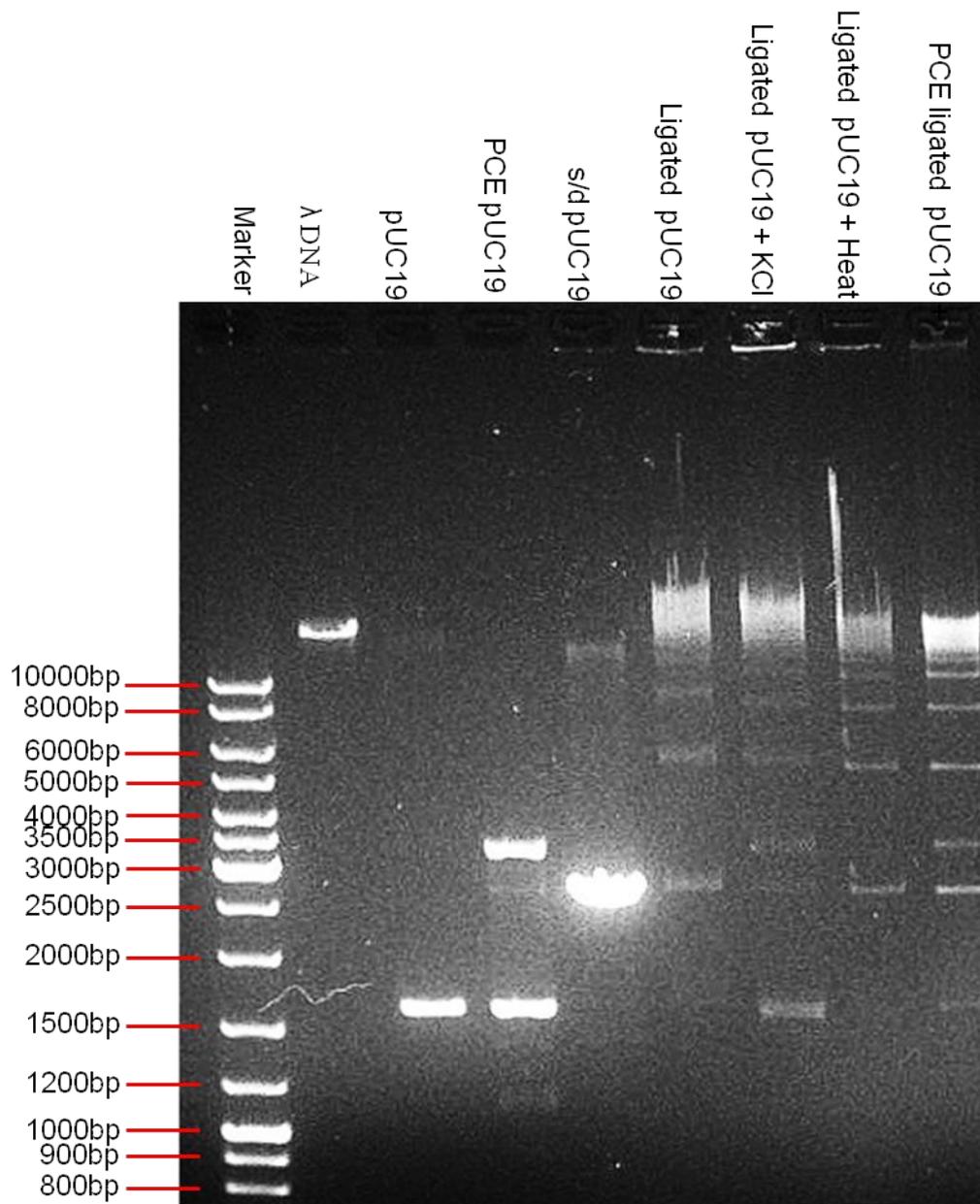
Post incubation, the DNA was pelleted by centrifugation at 11,406 xg for 20 minutes at 4 °C. The ethanol was aspirated and 500  $\mu\text{L}$  of 70% (v/v) ethanol was carefully added to the DNA

pellet. This was then centrifuged at 11,406 xg for 10 minutes at 4 °C. The ethanol was carefully aspirated and the DNA pellet was allowed to air dry at room temperature to remove any residual ethanol. The DNA pellet was dissolved in an appropriate volume of water when no ethanol droplets were observed within the microcentrifuge tube.

The gel images in Figures 40 and 41 both show the successful removal of T4 DNA Ligase from ligated pUC19 as a band shift can be seen after the DNA was extracted using Phenol:Chloroform. This occurs due to the T4 DNA Ligase denaturing and partitioning into the organic layer during the extraction (Palmiter 1974). The DNA is also free from shearing and severing. PCE and DNA precipitation were chosen instead of heat inactivation as PEG<sub>6000</sub> cannot be heated. This is present during the ligation of DNA (as described in method 3.2.10).



**Figure 40:** A 1% (w/v) standard agarose gel showing ligated pUC19 before and after PCE and DNA precipitation. Lane 1 shows uncut pUC19 plasmid. Similarly this is shown in lane 2, however the plasmid was extracted using Phenol:Chloroform and its DNA was precipitated. Lane 3 shows a linearised pUC19 plasmid that was cleaved once with the restriction enzyme BamHI. The next 3 lanes show pUC19 that has been ligated and tandemised over a 30 minute period. 4 species can be clearly identified across the lanes. These range from 2686 bp (single pUC19) to over 10,000 bp in length. The smears at the top are long concatemerised DNA that contains various different species and their lengths cannot be identified. Lane 4 shows ligated pUC19 that contains bound T4 DNA Ligase. A clear band shift can be seen when compared to the PCE DNA seen in lanes 5 and 6. This shows that the enzyme has been successfully removed. Lanes 5 and 6 also show that post PCE, the DNA has not been broken or sheared.



**Figure 41:** A 1% (w/v) standard agarose gel showing pUC19 that was ligated together at 16 °C for 30 minutes using T4 DNA Ligase. This gel determines which method was best at completely removing T4 DNA Ligase from the DNA, as described in 3.2.15, and without altering the integrity of the DNA. Lane 2 contains  $\lambda$  DNA. This was used as a size marker. Lane 3 contains pUC19 plasmid. This was also used as a size marker. Lane 4 contains pUC19 that was first treated with Phenol:Chloroform and then precipitated using method 3.2.23. This checked for any DNA shearing or damage. None can be seen here. Lane 5 shows pUC19 that was single cleaved with the restriction endonuclease BamHI (as described in method 3.2.1). Lane 6 shows ligated pUC19 that still contained active T4 DNA Ligase. Four clear bands can be seen. These show unligated pUC19 (2686 bp), open circle DNA, 3mers (8058 bp) and 10mers (26,860 bp). The smear shows pUC19 that has concatemerised to the length of  $\lambda$  DNA. Lane 7 shows ligated pUC19 containing 800 mM KCl. This salt concentration completely inhibits T4 DNA Ligase (as shown in Figure 37). The same DNA species can be seen as shown in Lane 6. Lane 8 shows ligated pUC19. T4 DNA Ligase was completely inhibited with 65 °C heat. The four DNA species have a lower molecular weight, suggesting that the T4 DNA Ligase has been removed from the DNA. Lane 9 shows Ligated pUC19 that was treated using PCE and the DNA was precipitated. The same DNA species can be seen. Like the DNA species shown in lane 8, a band shift can be seen. This again shows the complete removal of T4 DNA Ligase from the DNA.

### **3.2.16: Adaptation of Phenol:Chloroform extraction and DNA precipitation**

Initially, the PCE and DNA precipitation method as described in 3.2.15 was used to remove T4 DNA Ligase from TF $\lambda$  DNA. However, this method frequently produced yields as low as 10%. Due to the minimal amount of TF $\lambda$  obtained after gel extraction (as shown in Figures 42), this meant that the PCE and DNA precipitation method needed to be adapted. To lose such a large amount of ligated TF $\lambda$  during the removal of T4 DNA Ligase from the DNA not only made the generation of TF $\lambda$  a time consuming process but also effected experiments downstream due to limited DNA availability. Therefore, this method has been adapted and successfully increased the DNA yield to approximately 90% post PCE and DNA precipitation.

Firstly, equal volumes of Phenol and Chloroform were added to a microcentrifuge tube, making a total volume of 200  $\mu$ L. An 80  $\mu$ L volume that contained 2.8  $\mu$ g of newly ligated TF $\lambda$  (as described in method 3.2.15) was added to 120  $\mu$ L of water. The PCE mixture was vigorously vortexed for 15 seconds to ensure that the Phenol and Chloroform had thoroughly mixed. Immediately after vortexing, the DNA was added to the PCE mixture and gently inverted 15 times to mix. Next, the the PCE and DNA mixture was placed on the rotator for 15 minutes at 43 rpm. The mixture was then phase-separated into 3 layers using centrifugation at 11,406 xg for 1 minute. If the layers had not completely separated, the mixture was recentrifuged.

The top aqueous layer containing the DNA was aspirated and placed in a microcentrifuge tube and its volume measured. An equal volume of Chloroform was added. The Chloroform DNA mixture was gently inverted 30 times to mix prior to centrifugation. The DNA was then precipitated at -20  $^{\circ}$ C for 16 hours overnight instead of 30 minutes on ice.

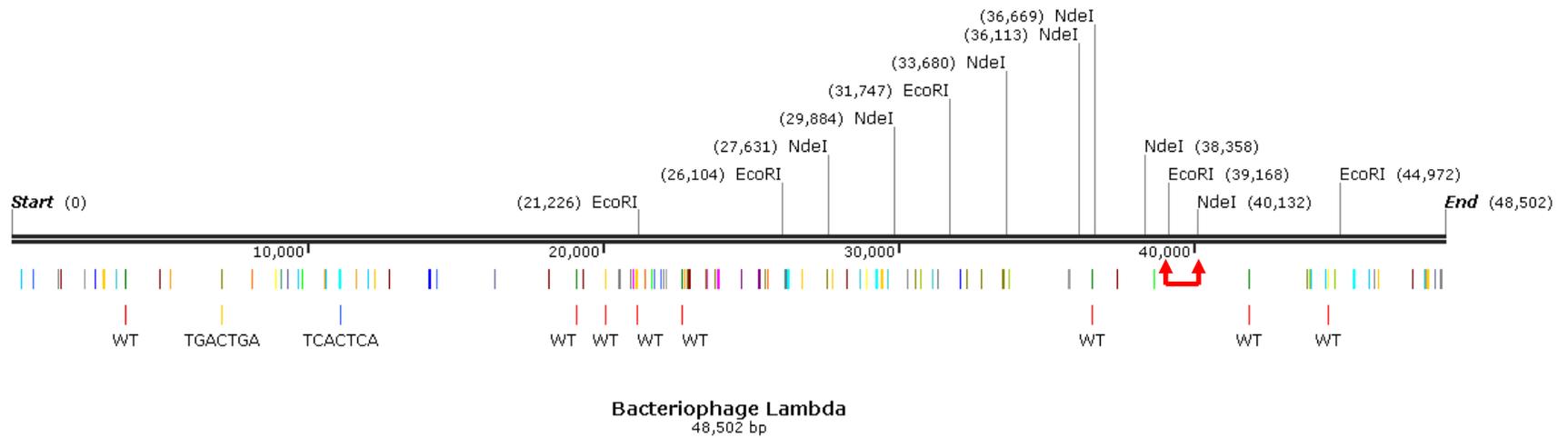
### **3.2.17: Identifying target free $\lambda$ (TF $\lambda$ ) DNA**

To help identify an area of  $\lambda$  DNA that did not contain any WT AP1 target sites or variant AP1 target sites, the program SnapGene (SnapGene 2013) was used. This target free  $\lambda$  DNA (TF $\lambda$ ) region would allow DNA bound proteins to be studied without the presence of their AP1 target and variant AP1 target sites, therefore serving as a true negative control.

The restriction endonucleases NdeI (recognition sequence 5'- CATATG -3'. Fermentas) and EcoRI (recognition sequence 5'- GAATTC -3'. Fermentas) were used to cleave  $\lambda$  DNA. By using

2 different restriction endonucleases, this increased the probability of locating a target free area along the DNA. NdeI and EcoRI could cleave the DNA simultaneously with 100% efficiency and with no star activity while in restriction buffer O (ThermoFisherScientific 2015). The restriction digest was performed according to the method supplied by the manufacturer.

Equal amounts (10u/  $\mu$ L) of NdeI and EcoRI were added to the digest mixture that contained 3  $\mu$ g of  $\lambda$  DNA. This amount of DNA was cleaved due to the number of DNA fragments that would be produced and the DNA was left incubating at 37 °C for 16 hours overnight to ensure complete DNA cleaving (as shown in Figure 43). Both enzymes were heat inactivated at 80 °C for 20 minutes. This 962 bp TFL region has been indicated by the red arrows in Figure 42.



**Figure 42:** A diagram representing a single piece of  $\lambda$  DNA (created using SnapGene). The location of all NdeI and EcoRI restriction sites can be seen in addition to the WT AP1 target sites (red) and variant AP1 target sites (multicolour). The red arrows highlight part of the  $\lambda$  DNA that is free from all of these target sites. This target free region also has a single NdeI and EcoRI restriction site at either end.

### **3.2.18: Double digestion of pUC19 for $\lambda$ pUC construction**

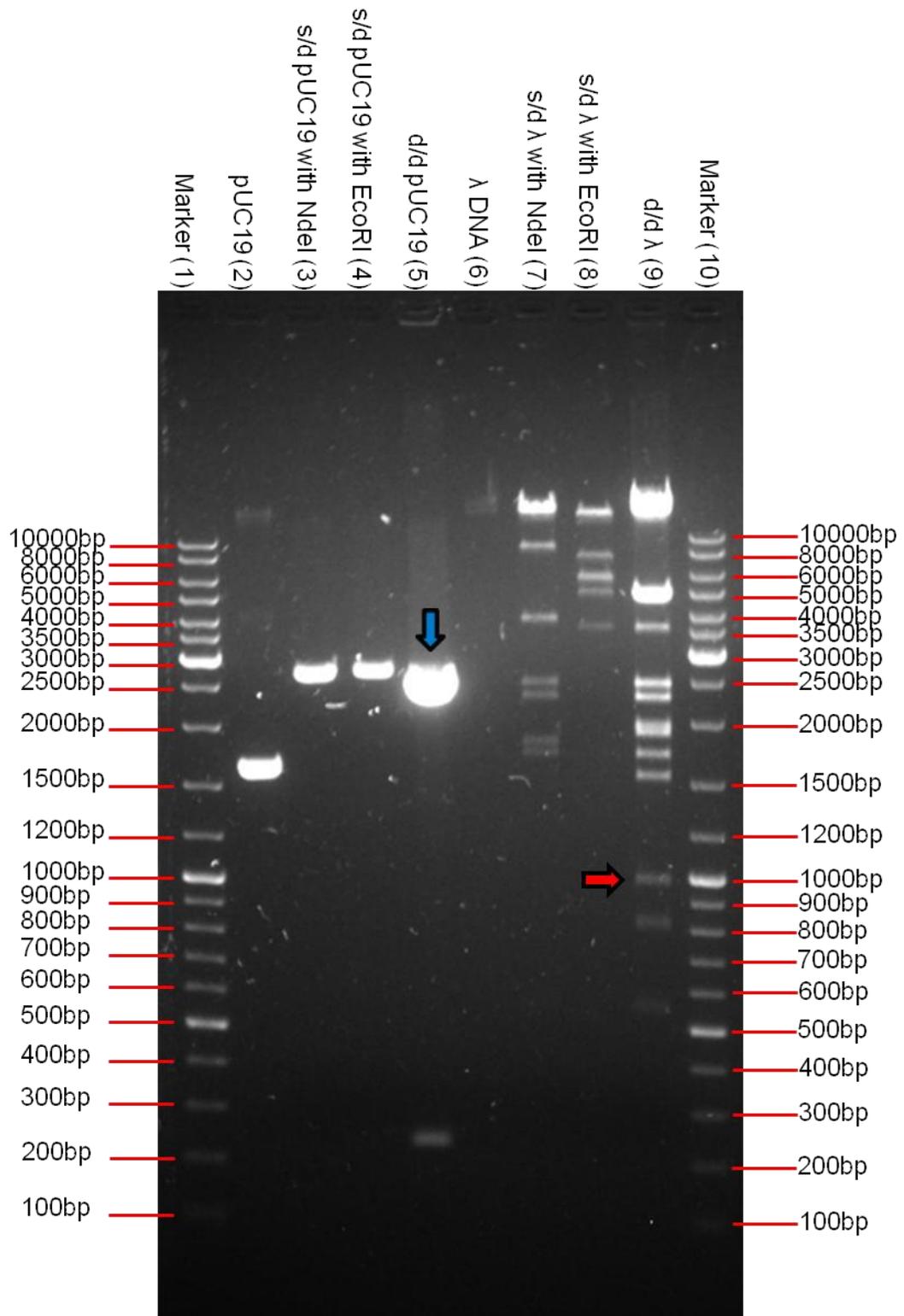
In preparation for the construction of  $\lambda$ pUC (as shown in Figure 43), 3  $\mu$ g of pUC19 was simultaneously cleaved with the restriction endonucleases NdeI (10  $\mu$ L) and EcoRI (10  $\mu$ L) (Fermentas) for 2-16 hours at 37 °C as described in method 3.2.24. To ensure that the double cleaved vector could not self-anneal during ligation, phosphate groups from the 5'- end were removed using Fast Alkaline Phosphatase (FastAP) (ThermoFisherScientific 2015). This mixture was incubated for 15 minutes at 37 °C. FastAP was heat inactivated at 80 °C in accordance with the manufacturer's guidelines. This temperature is higher than the 65 °C required to heat inactivate NdeI and EcoRI. This meant that both of the restriction endonucleases were also inactivated.

### **3.2.19: Constructing $\lambda$ pUC**

$\lambda$ pUC is a 3436 bp circular DNA plasmid. This can be seen in Figure 44. It was constructed using pUC19 and the  $\lambda$  DNA fragment TF $\lambda$  that were doubly cleaved with restriction endonucleases NdeI (10  $\mu$ L) and EcoRI (10  $\mu$ L) (as described in method 3.2.17).

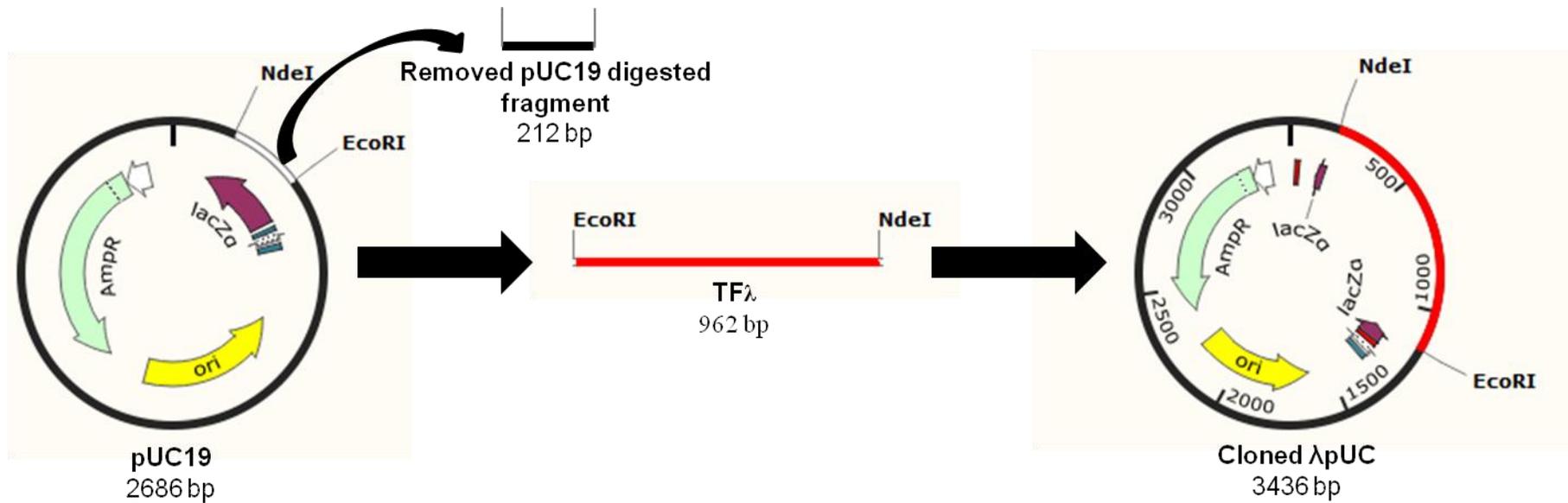
Prior to DNA construction, both doubly digested pUC19 and  $\lambda$  DNA were run on a 1% (w/v) agarose gel containing 1x TAE buffer and 0.5  $\mu$ g/mL of EtBr to separate out the fragments based on their molecular weight. This can be seen in Figure 43. 1.5  $\mu$ g of cleaved pUC19 and  $\lambda$  DNA was loaded onto the gel. This would increase the amount of DNA present in each band. The 2474 bp long pUC19 fragment and 962 bp long TF $\lambda$  DNA fragment were then excised and gel extracted using a gel extraction kit (Qiagen 2015).

Next,  $\lambda$ pUC was constructed by creating a 20  $\mu$ L ligation mixture containing a ratio of 3 TF $\lambda$  DNA fragments to 1 pUC19 vector (based on their molarity ratio). This ligation mixture was placed in a sterile PCR tube and incubated inside of the PCR machine at 16 °C for 15 minutes prior to adding T4 DNA Ligase. This ensured that the components within the ligation mixture were equilibrated to 16 °C. After the addition of T4 DNA Ligase, the mixture was incubated at 16 °C for 16 hours. This was then immediately stored at 4 °C. The newly created recombinant plasmid was transformed using Top10 Chemically Competent *E. coli* (method 3.2.6). This construction process has been represented by the diagram shown in Figure 44.



**Figure 43:** A 1% (w/v) agarose gel showing doubly cleaved pUC19 and λ DNA with the restriction endonucleases NdeI and EcoRI (Fermentas. Method 3.2.17). Lane 2 shows pUC19 plasmid. Lanes 3 and 4 show pUC19 that has been single cleaved with NdeI and EcoRI respectively. This shows that the restriction endonucleases are cleaving and no star activity has occurred. Lane 5 shows pUC19 that has

been doubly cleaved with both NdeI and EcoRI. Two bands can be seen. These are 2474 bp, indicated by the blue arrow, and 212 bp in length. Lane 6 shows  $\lambda$  DNA that has been cleaved with NdeI. Seven DNA species can be seen. These range from 1700 bp to 48,000 bp in length. Lane 7 shows  $\lambda$  DNA that has been cleaved with EcoRI. Five fragments were produced. These range from 4000 bp to 48,000 bp in length. Lane 8 shows  $\lambda$  DNA that was doubly cleaved with NdeI and EcoRI. This produced several fragments consisting of single and doubly cleaved ends. The target free DNA fragment of interest, as shown in Figure 42, is 962 bp in length. This has been indicated by the red arrow.



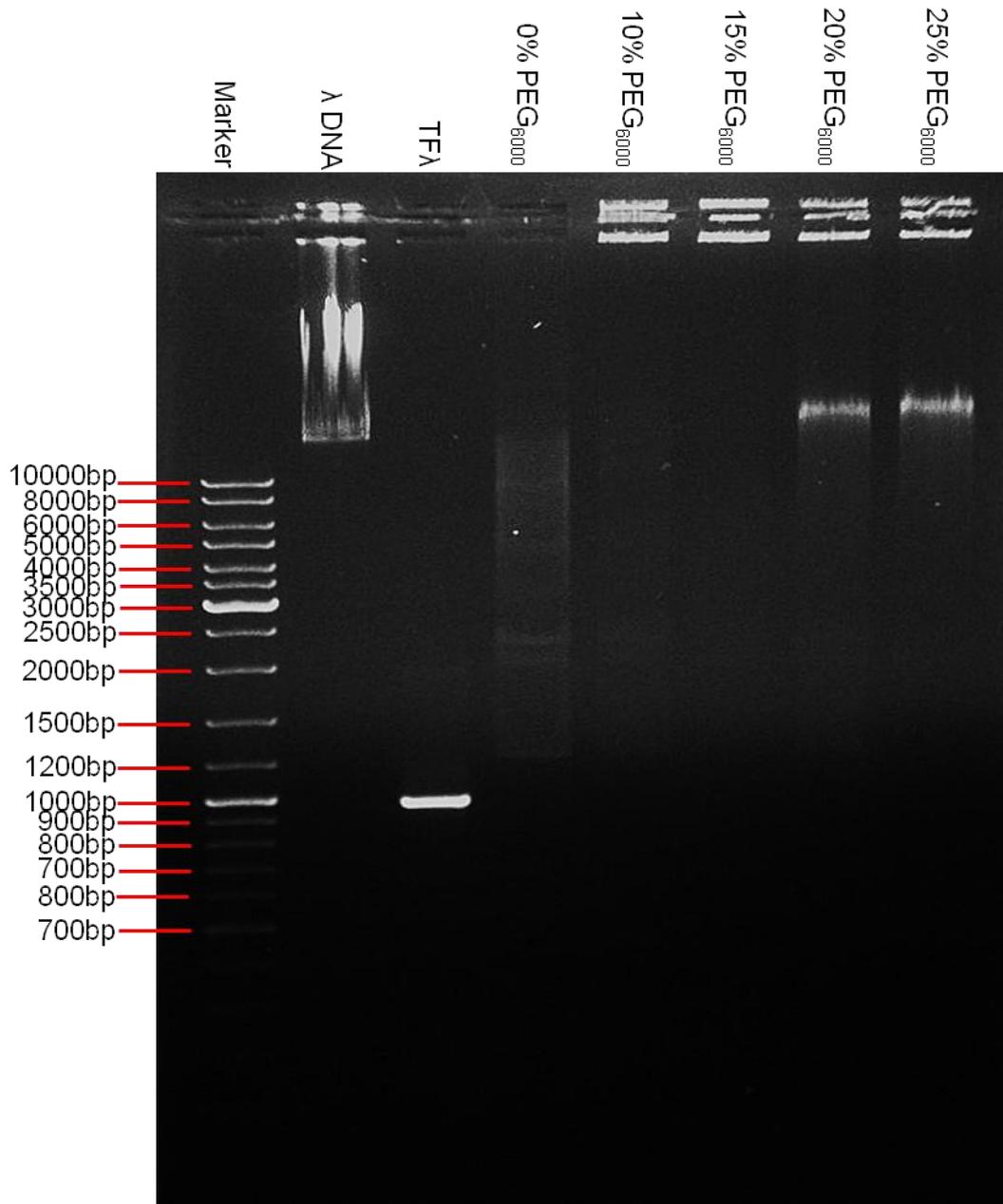
**Figure 44:** A diagram showing the construction of the 3436 bp long circular DNA plasmid  $\lambda$ pUC (adapted from SnapGene). pUC19 was first doubly cleaved with the restriction endonucleases NdeI and EcoRI and dephosphorylated with FastAP before being heat inactivated and gel extracted. The gel extracted donor 964 bp  $\lambda$  DNA fragment was inserted using a 3 fragments:1 vector ratio. These fragments were ligated together using T4 DNA Ligase at 16 °C for 16 hours. The newly created recombinant plasmid was transformed using Top10 Chemically Competent *E. coli*.

### **3.2.20: PEG<sub>6000</sub> concentration for optimal TFλ DNA ligation**

To identify which PEG<sub>6000</sub> concentration would promote end-on-end ligation of the 962 bp long TFλ fragment during ligation with T4 DNA Ligase, a concentration gradient was performed. This method is an adaptation of the method 3.2.19.

Firstly, a 1mL stock of 60% (w/v) PEG<sub>6000</sub> was made. The percentage of PEG<sub>6000</sub> was increased due to the minimal space available within the ligation mixture. This is due to TFλ having a low concentration, therefore requiring a larger proportion of the mixture. In a final volume of 20 μL each, several ligation mixtures were made containing 700 ng of TFλ and varying percentages of PEG<sub>6000</sub> within sterile PCR tubes. These percentages were 0%, 10%, 15%, 20% and 25% (w/v). Prior to the addition of T4 DNA Ligase, the ligation mixtures were equilibrated to 16 °C for 15 minutes to ensure that each component was at this temperature. 1 μL of T4 DNA Ligase was added and the ligation mixtures were incubated overnight for precisely 16 hours. Post incubation, a 10 μL 3x stop buffer (2.4 M KCl and water) was added. Used at 1x, 800 mM KCl inactivated T4 DNA Ligase. The samples were then immediately stored at -20 °C until use.

The analytical gel image represented in Figure 45 shows that 20% (w/v) PEG<sub>6000</sub> is the optimum percentage to use as this facilitated formation of the longest TFλ DNA concatemers.

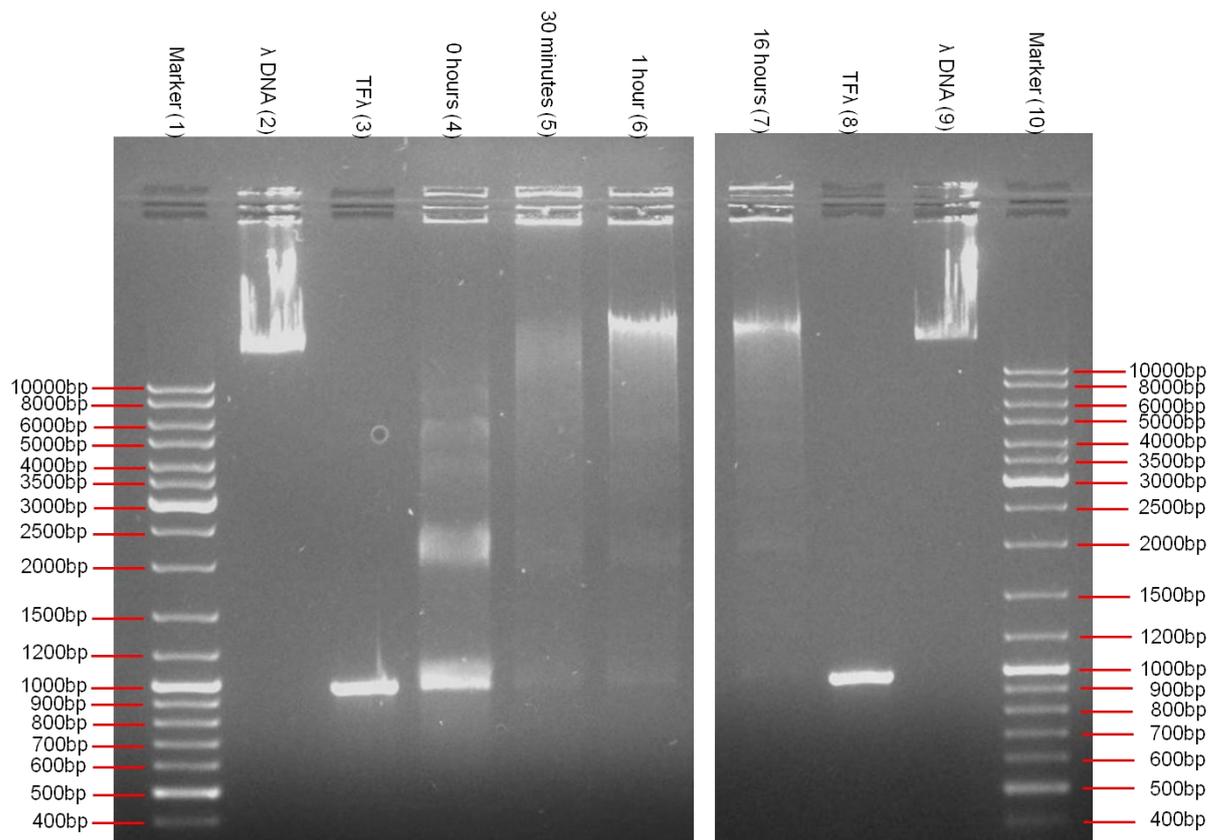


**Figure 45:** A 1% (w/v) agarose gel image showing TFL that had been ligated for 16 hours at 16 °C with T4 Ligase in the presence of varying concentrations of PEG<sub>6000</sub>. Lane 1 shows λ DNA. This was used as a size marker. Lane 3 shows TFL before ligation. Lanes 4-8 show ligated TFL in the presence of 0%, 10%, 15%, 20% and 25% (w/v) PEG<sub>6000</sub>. These PEG<sub>6000</sub> concentrations were used to identify which concentration promoted DNA concatemerisation. 20% (w/v) PEG<sub>6000</sub> was chosen as the DNA is shown to be comparable with λ and longer. Unlike pUC19 that contains variant AP1 target sites; this DNA construct serves as a true negative control as it is completely free from AP1 and variant AP1 target sites.

### **3.2.21: A time course to create TF $\lambda$ tightropes that are comparable with the length of $\lambda$ DNA**

A time course experiment was performed to identify which time duration was required to ligate and concatemerise TF $\lambda$  to the same length as  $\lambda$  DNA which is 48,502 bp in length. The method used here is an adaptation of method 3.2.11. Inside of the PCR thermocycler, 6 separate ligation mixtures, each containing 700 ng of TF $\lambda$  DNA and 20% (w/v) PEG<sub>6000</sub>, were incubated in sterile PCR tubes at 16 °C for 15 minutes to equilibrate. Upon the addition of 5 Weiss u/  $\mu$ L of T4 DNA Ligase to 2 ligation mixtures simultaneously, their ligation time commenced. TF $\lambda$  DNA was ligated for 30 minutes, 1 hour and 16 hours. Time point 0 hours was created by firstly adding T4 DNA Ligase to the 1x KCl stop buffer which was then immediately added to the ligation mixture. T4 DNA Ligase was inactivated 800 mM KCl after each time duration.

The gel shown in Figure 46 shows the level of ligation at each time point. A ligation time of 16 hours was chosen to concatemerise TF $\lambda$ . Firstly, 16 hours is a convenient time as the ligation could occur overnight and secondly, this provided sufficient time for TF $\lambda$  to concatemerise to a length that is comparable with  $\lambda$  DNA.



**Figure 46:** A 1% (w/v) agarose gel showing the concatemerisation of 92 bp long fragment TF $\lambda$  DNA at different time points using T4 DNA ligase. T4 DNA Ligase was inactivated with 800 mM KCl after each time duration. Lane 2 shows  $\lambda$  DNA. This was used as a size marker. Lane 3 shows unligated TF $\lambda$  DNA. Lane 4 shows ligation at time 0 minutes. DNA species of varying length can be seen. Lanes 5 also shows a vast range of DNA species. Most of the DNA has remained in the well, suggesting that it has formed extremely long concatemers. Lane 6 shows TF $\lambda$  DNA that has formed a variety of species, however a prominent band can be seen at 48,000 bp. This shows that the DNA has concatemerised to the length of  $\lambda$  DNA after 1 hour. Lane 7 shows similar results to lane 6, whereby a 16 hour time duration produces DNA species of varying length and a distinct tandem DNA species that is comparable in length to  $\lambda$  DNA.

### **3.2.22: Adapted TF $\lambda$ DNA ligation method**

$\lambda$ PUC was cleaved with the restriction endonucleases NdeI (10 u/ $\mu$ L) and EcoRI (10 u/ $\mu$ L). This yielded a small quantity of TF $\lambda$  DNA fragments that were 962 bp in length (as described in section 3.3.1). As TF $\lambda$  is much smaller when compared with other DNA substrates used in this study, extra care was taken to promote the fragments tandem end-on-end ligation, producing a DNA substrate that was similar in length to  $\lambda$  DNA. The following steps were taken to achieve this.

Firstly, a ligation mixture containing 20% (w/v) PEG<sub>6000</sub> (as described in method 3.2.20) was incubated in a PCR thermocycler for 15 minutes at 16 °C. This ensured that the mixture was equilibrated to this temperature. During equilibration, 700 ng of TF $\lambda$  DNA was placed in a sterile PCR tube. To self-anneal the fragments, the DNA was heated at 60 °C for precisely 5 minutes prior to cooling in an ice bath for 1 minute. The self-annealed TF $\lambda$  DNA fragments were added to the ligation mixture. T4 DNA Ligase (5 Weiss u/  $\mu$ L) was added to the mixture 10 minutes after the addition of the DNA fragments. This ensured that the mixture was equilibrated to 16 °C. Ligation occurred for 16 hours. This was immediately followed by PCE and DNA precipitation (as described in method 3.2.23).

### **3.2.23: The Binding site Finder for pUC19**

This Binding Site Position Finder, inclusive of the 80 bp error margin, functions in the same way as the Site Finder used to determine where cFos, cJun and cFos:cJun heterodimers were pausing as they diffused along  $\lambda$  DNA, both with and without variant AP1 target sites as described in Chapter 2, Figures 24 and 25.

pUC19 however is 2686 bp in length, 18 times smaller than  $\lambda$  DNA. There are no AP1 target sites present along pUC19, however, there are 11 variant AP1 target sites. Due to the size of pUC19, 3 pieces of linear DNA have been consecutively tandemised so that the DNA is longer in length. This ensures that any distances between pause locations that are greater than 2686 bp can be quantified.

All “Yes” answers that appeared within the first piece of pUC19 were quantified only. This is due to a repeating pattern of “Yes” being seen in strands 2 and 3. If, for example, a “Yes” response within strand 1 showed that the protein paused at binding sites where one site was on the first DNA strand and the second site was on the second strand, these were also quantified. A section of this pUC19 Binding Site Position Finder can be seen in Figure 47.

112      =IF(AND(\$D\$3<ABS(G13-E12),\$D\$3>ABS(E13-G12)),"Yes","No")

User input					
	Distance between 2 bound proteins:	355.0373227			
	AP1 length:	7			
	Error:	80			

Result					
Where have the 2 proteins bound?					
Binding site	Start position	Start position with error	End position	End position with error	
TGACACA	1	33	-47	39	119
TCACTCA	2	559	479	565	645
TAACTCA	3	568	488	574	654
TGACTCT	4	705	625	711	791
TCACTCA	5	750	670	756	836
GGACTCA	6	1176	1096	1182	1262
TTACTCA	7	1605	1525	1611	1691
TGCCTCA	8	1644	1564	1650	1730
TGACTCC	9	1693	1613	1699	1779
TCACTCA	10	2101	2021	2107	2187
TGTCTCA	11	2530	2450	2536	2616
TGACACA	12	33	-47	39	119
TCACTCA	13	559	479	565	645
TAACTCA	14	568	488	574	654
TGACTCT	15	705	625	711	791
TCACTCA	16	750	670	756	836
GGACTCA	17	1176	1096	1182	1262
TTACTCA	18	1605	1525	1611	1691
TGCCTCA	19	1644	1564	1650	1730

**Figure 47:** The adapted Binding Site Position Finder for pUC19 DNA. All 11 variant AP1 target sites that appear on a single piece of pUC19 DNA have been incorporated into this Site Finder. 3 pieces of pUC19 have been tandemised. Like the previous Binding Site Position Finders mentioned in Chapter 2, the user simply types the known distance in the red outlined box and presses enter.

### **3.2.24: The Binding site Finder for pUCap1**

The pUCap1 Binding Site Position Finder functions in the same way as the pUC19 Site Finder as described in method 3.2.23. All data obtained from this Site Finder is analysed in the same manner as well. Although pUCap1 is 3 bp shorter than pUC19, the fundamental difference is that pUCap1 contains a single AP1 target site. This meant that there were 12 target sites of interest. Figure 48 shows a section of this pUCap1 Binding Site Position Finder.

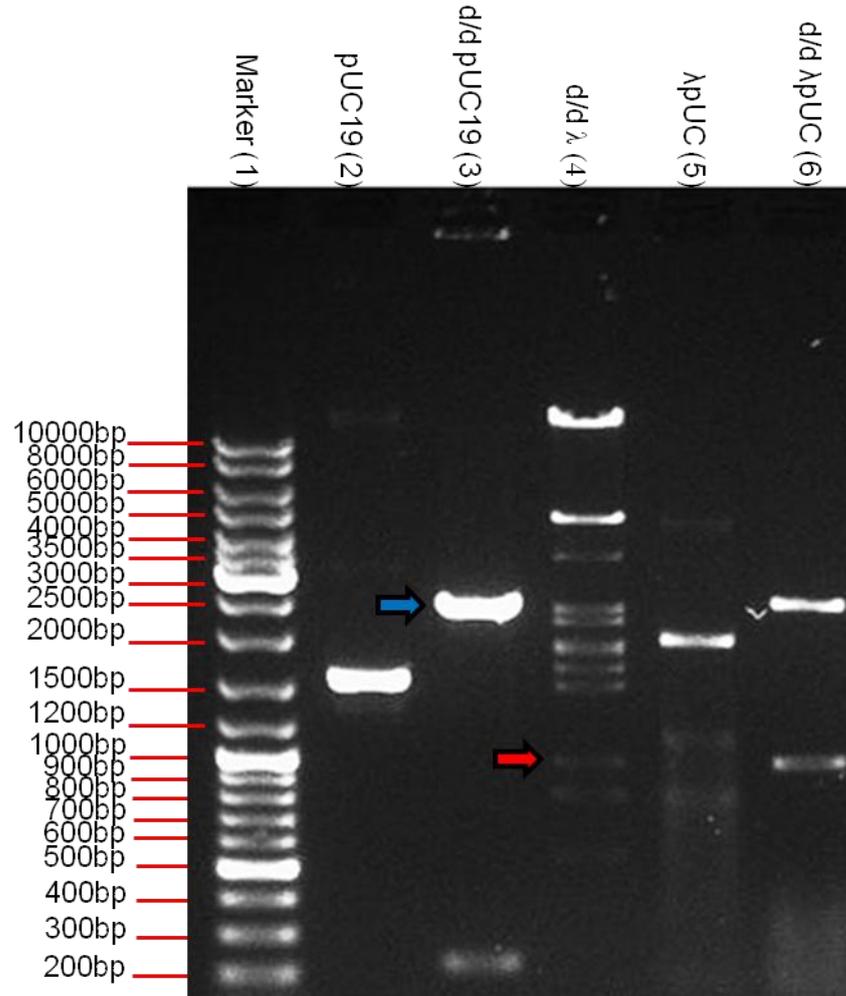


### **3.3: Results**

This chapter aimed to gain a greater understanding of how cFos, cJun and cFos:cJun heterodimers interact with different DNA substrates. The first DNA substrate used in this study was pUC19. This DNA is free from AP1 target sites; however there are 11 variant AP1 target sites present. The second type of DNA investigated was pUCap1. This DNA is composed of pUC19 that has a single AP1 target site cloned into it. Lastly, TF $\lambda$  is a DNA substrate that is completely free from AP1 and variant AP1 target sites. Using these different types of DNA will enable studies to be performed to understand whether the presence or absence of AP1 and variant AP1 target sites affects the proteins complex formation, diffusion characteristics (represented by their  $\alpha$  values), diffusion constants ( $\mu\text{m}^2\text{s}^{-1}$ ) and to quantify in which AP1 and variant AP1 target sites the proteins pause.

#### **3.3.1: Restriction mapping $\lambda$ pUC**

To confirm that the correct doubly digested pUC19 fragment and  $\lambda$  DNA fragment were ligated together forming  $\lambda$ pUC, a  $\lambda$ pUC plasmid was simultaneously cleaved with the restriction endonucleases NdeI (10 u/ $\mu$ L) and EcoRI (10 u/ $\mu$ L). This was performed as described in methods 3.2.17. The plasmid was incubated at 37 °C for 2 hours. Both of the restriction endonucleases were heat inactivated at 65 °C for 20 minutes. This was performed in accordance with Fermentas guidelines. Figure 49 shows  $\lambda$ pUC before its construction and Figure 44 shows the completed construct (schematic diagram). To further confirm the correct DNA components were used to construct  $\lambda$ pUC, the plasmid was sequenced.



**Figure 49:** A 1% (w/v) agarose gel showing the DNA components used to create  $\lambda$ pUC and its diagnostic cleavage with the restriction endonucleases NdeI and EcoRI. Lane 2 shows pUC19 plasmid. Lane 3 contains pUC19 that has been doubly cleaved with the restriction endonucleases NdeI and EcoRI. This produced a large fragment of 2474 bp, indicated by the blue arrow, and a small fragment of 212 bp. Lane 4 shows  $\lambda$  DNA that was cleaved with NdeI and EcoRI. This produced several fragments consisting of single and double cleaved ends. The 962 bp  $\lambda$  DNA fragment that contains no WT or variant AP1 target sites has been indicated by the red arrow. Lane 5 shows uncut  $\lambda$ pUC. To show that the correct fragments had been ligated together to form  $\lambda$ pUC, a  $\lambda$ pUC plasmid was doubly cleaved with the restriction endonucleases NdeI and EcoRI. The correct pUC19 fragment of 2474 bp and  $\lambda$  DNA fragment of 962 bp are shown. This can be seen in lane 6. Sequencing the  $\lambda$ pUC plasmid also confirmed the correct insertion of the donor  $\lambda$  DNA fragment TFL into pUC19.

### 3.3.2: Protein interaction with pUC19 DNA.

As described in Chapter 2 section 2.2.8, all of the proteins used within this study were biotinylated as handles for attaching Qdots. cFos was labelled with red 655 nm Qdots and cJun was labelled with green 525 nm Qdots using a 1:4 peptide to Qdot ratio. Dual colour experiments provided a host of red, green and dual coloured complexes. It is these dual coloured complexes that indicated heterodimeric complex formation. This colour labelling strategy was adopted throughout the entire duration of this study, ensuring that all proteins were fluorescently tagged prior to being loaded onto the DNA. All studies were performed in a high salt buffer. Proteins that were bound to DNA tightropes were analysed providing they were not close to any beads, the tightropes were not swaying or overcrowded with proteins as described in Chapter 2 section 3.1.1.

The first study performed aimed to identify what percentage of cJun moved while on ligated pUC19 tightropes. cJun was chosen for this study as it was the most mobile protein observed on  $\lambda$  DNA. 119 pUC19 tightropes were studied.  $54 \pm 3.7\%$  ( $n = 242$ . t-test p-value = 0.06) of cJun molecules were observed moving. This is 17% less than the percentage of cJun moving along  $\lambda$  DNA, as shown in Chapter 2 section 3.1.1.

This result raised the possibility that T4 DNA Ligase could be inhibiting protein movement. To determine this, 1  $\mu$ M ATP was added to the ligated DNA tightropes (107 tightropes were analysed) as T4 DNA Ligase utilises this during adenylation (Rossi et al. 1997). The aim here was to remove T4 DNA Ligase from the DNA and keep it in solution. Upon the addition of ATP,  $37 \pm 3.1\%$  ( $n = 229$ . t-test p-value = 0.9) of cJun molecules were now observed moving. This percentage shows a large reduction in movement, increasing the possibility that T4 DNA Ligase was in fact inhibiting protein movement. To solve this problem, 118 tightropes were studied where T4 DNA Ligase was completely removed from the DNA post ligation using PCE and DNA precipitation (as described in method 3.2.15). Afterward,  $58 \pm 0.4\%$  ( $n = 269$ . t-test p-value = 0.1) of cJun molecules were observed to diffuse freely along PCE pUC19 tightropes.

Next, cFos and cFos:cJun heterodimers were independently introduced to PCE pUC19 DNA tightropes and the number of moving complexes were quantified. 104 tightropes containing cFos were studied and  $49 \pm 1.2\%$  ( $n = 183$ ) of the molecules were observed to diffuse freely. This result can be compared with the 45% of cFos (t-test p-value = 0.35) that moved along  $\lambda$  DNA. 46 tightropes containing a mixture of red cFos ( $n = 23$ ), green cJun ( $n = 26$ ) and dual coloured cFos:cJun ( $n = 34$ ) complexes were also studied.  $47 \pm 2.2\%$  ( $n = 16$ , t-test p-value = 0.54) of cFos:cJun heterodimers were also observed moving along the DNA, 8% greater than cFos:cJun moving along  $\lambda$  DNA.

The introduction of a single AP1 target site into pUC19, termed pUCap1 (as described in method 3.2.3), enabled cFos, cJun and cFos:cJun to be studied when in the presence of 1 AP1 and 11 variant AP1 target sites. This was to see whether the presence of an AP1 target site would promote or hinder protein movement. This DNA substrate was studied alongside cJun on ligated pUC19.

Firstly, cJun was loaded onto ligated pUCap1 and the number of moving proteins was quantified. 109 tightropes were studied.  $46 \pm 4.5\%$  ( $n = 161$ ) cJun molecules were observed to diffuse freely. Next, 1  $\mu$ M ATP was added to the ligated DNA tightropes and 118 tightropes were analysed. Again, like with pUC19 ligated tightropes, this was to remove T4 DNA Ligase from the DNA.  $42 \pm 3.9\%$  ( $n = 169$ ) of cJun molecules were observed moving along pUCap1+ATP. To prevent protein inhibition while moving, T4 DNA Ligase was removed from the DNA using PCE and DNA precipitation. On this clean DNA substrate 62 tightropes were analysed and  $52 \pm 2.6\%$  ( $n = 195$ ) of cJun were observed to diffuse freely. This can be compared with cJun (t-test p-value = 0.78) moving along PCE pUC19 DNA.

cFos was the next protein to be loaded PCE pUCap1 DNA tightropes. 88 tightropes were studied and  $55 \pm 1.8\%$  ( $n = 145$ ) were observed moving. This observation can be compared with the percentage of cFos (t-test p-value = 0.20) moving on both  $\lambda$  (45%) and PCE pUC19. Finally, pre-mixed cFos and cJun were loaded onto the DNA enabling the number of moving cFos:cJun heterodimers to be quantified. 71 tightropes containing a mixture of red cFos ( $n = 53$ ), green cJun ( $n = 90$ ) and dual coloured cFos:cJun ( $n = 99$ ) complexes were also studied.  $32 \pm 1.6\%$  ( $n$

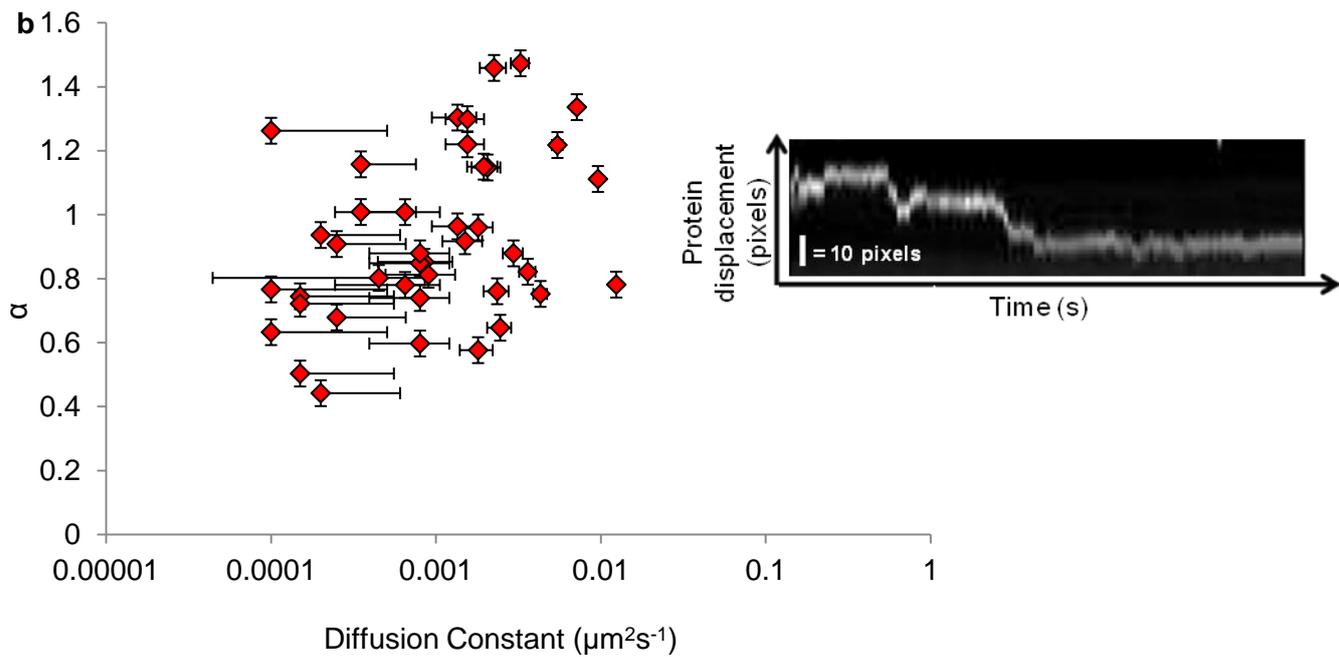
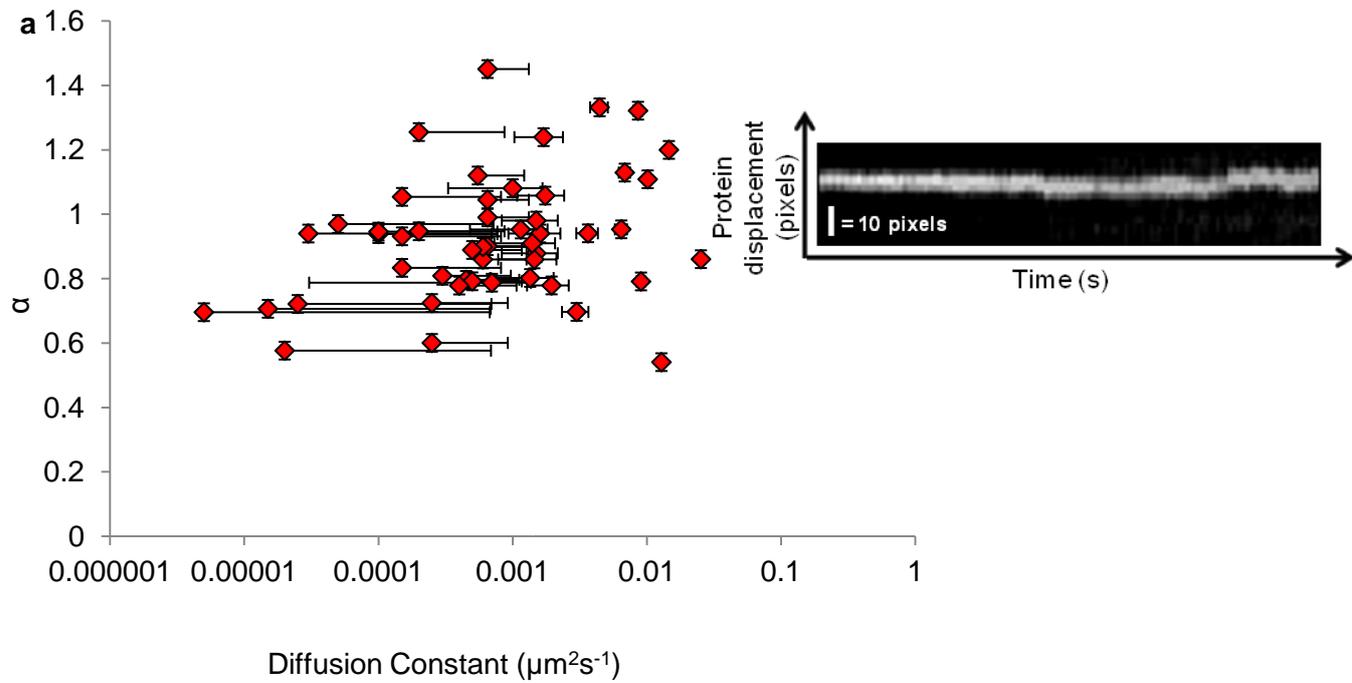
= 32) were moving dual coloured cFos:cJun heterodimeric complexes. This percentage is comparable with  $\lambda$  DNA although there was 15% less diffusion when compared with cFos:cJun moving on PCE pUC19 DNA. Lastly, cJun was loaded onto TFL DNA. Only 3% ( $n = 79$ , t-test p-value = 0.80) of cJun molecules were observed to diffuse when compared with cJun diffusing along  $\lambda$  DNA. This data can be seen in Table 9.

The next part of this study was to identify the diffusion constants and  $\alpha$  values (as described in Chapter 2) for cFos and cJun that travelled along PCE pUC19 and PCE pUCap1 DNA substrates while in a high salt buffer. The first DNA substrate to be analysed was PCE pUC19. On this DNA, cFos produced diffusion constants ranging from  $1 \times 10^{-5}$  to  $1 \times 10^{-2} \mu\text{m}^2\text{s}^{-1}$  and cJun  $1 \times 10^{-6}$  to  $1 \times 10^{-1} \mu\text{m}^2\text{s}^{-1}$ . This meant the average diffusion constants for cFos and cJun were  $1.5 \times 10^{-3} \pm 3 \times 10^{-4} \mu\text{m}^2\text{s}^{-1}$  and  $2.6 \times 10^{-3} \pm 6.7 \times 10^{-4} \mu\text{m}^2\text{s}^{-1}$ , respectively (as shown in Figure 50). Next, PCE pUCap1 DNA was analysed. cFos produced diffusion constants ranging from  $1 \times 10^{-4}$  and  $1 \times 10^{-2} \mu\text{m}^2\text{s}^{-1}$  and cJun from  $1 \times 10^{-6}$  and  $1 \times 10^{-2} \mu\text{m}^2\text{s}^{-1}$ . This resulted in the average diffusion constants for cFos and cJun being  $1.96 \times 10^{-3} \mu\text{m}^2\text{s}^{-1}$  and  $2.46 \times 10^{-3} \mu\text{m}^2\text{s}^{-1}$ , respectively (as shown in Figure 51).

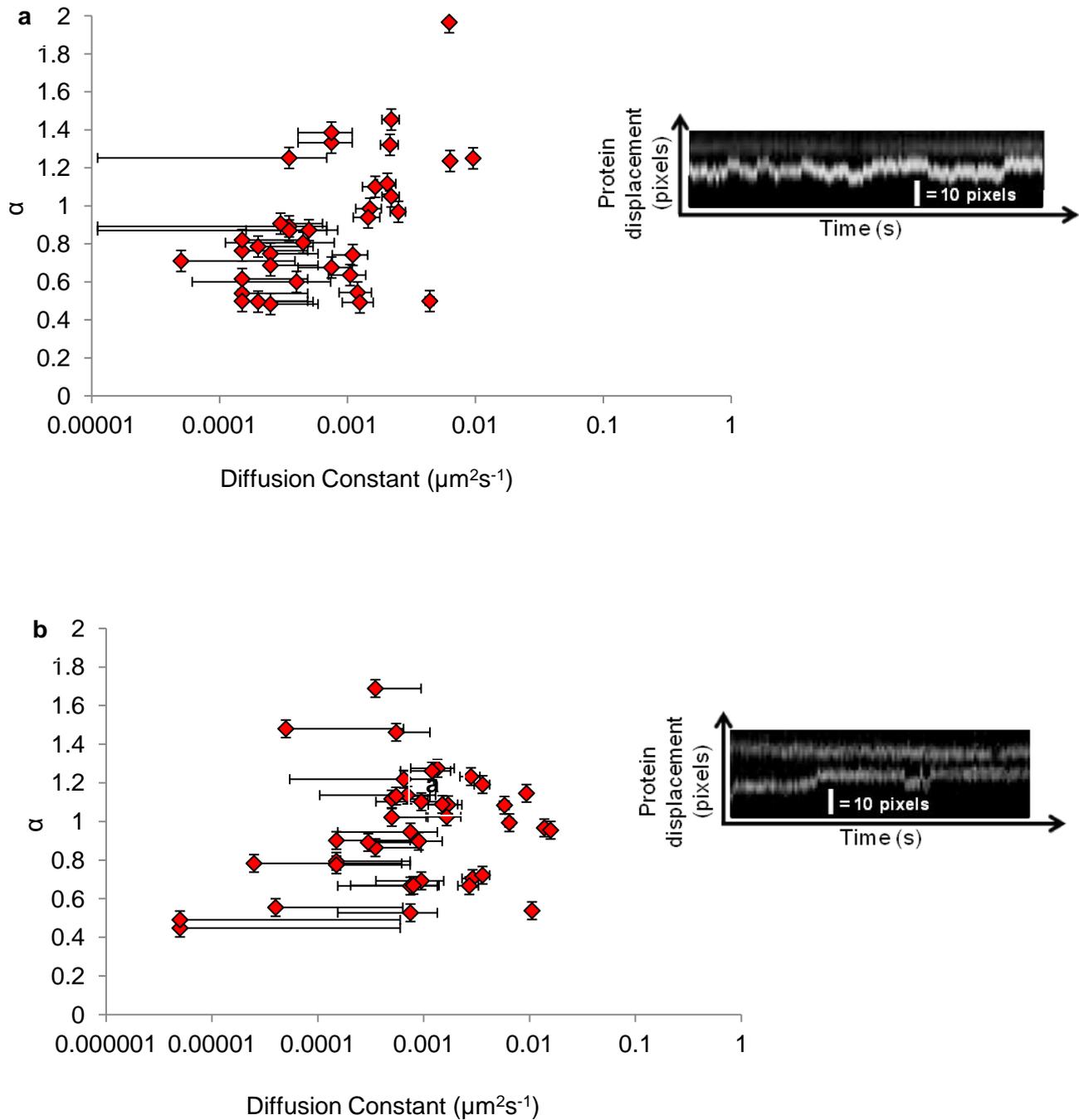
To understand the proteins mode of travel as they moved along PCE pUC19 and PCE pUCap1 DNA tightropes in search of their target sites, their  $\alpha$  values (as described in Chapter 2) were calculated. Firstly, cFos produced  $\alpha$  values ranging from 0.4 to 1.9 and cJun 0.5 to 1.4 with PCE pUC19. Their average  $\alpha$  values were  $0.9 \pm 0.05$  and  $0.9 \pm 0.02$  respectively. Lastly, cFos and cJun's  $\alpha$  values were determined as they travelled along PCE pUCap1 DNA tightropes. cFos produced  $\alpha$  values ranging from 0.4 to 1.4 and cJun 0.4 to 1.6. Their average  $\alpha$  values were  $0.9 \pm 0.04$  and  $1.0 \pm 0.04$ , respectively. The proteins displaying  $\alpha$  values of  $<1$  indicates that an anomalous sub-diffusion process may be occurring.

	pUC19	pUC19 + ATP	PCE pUC19	pUCap1	pUCap1 + ATP	PCE pUCap1	PCE TFλ
<b>cFos</b>	****	****	49 ± 1.2% ( <i>n</i> = 183)	****	****	55 ± 1.8% ( <i>n</i> = 145)	****
<b>cJun</b>	54 ± 3.7% ( <i>n</i> = 242)	37 ± 3.1% ( <i>n</i> = 229)	58 ± 0.4% ( <i>n</i> = 269)	46 ± 4.5% ( <i>n</i> = 161)	42 ± 3.9% ( <i>n</i> = 169)	52 ± 2.6% ( <i>n</i> = 195)	3 ± 1.5% ( <i>n</i> = 79)
<b>cFos:cJun Heterodimers</b>	****	****	47 ± 2.2% ( <i>n</i> = 34)	****	****	32 ± 1.6% ( <i>n</i> = 99)	****

**Table 9:** A table showing the percentage of cFos, cJun and cFos:cJun complexes that were observed moving along pUC19, pUCap1 and TFλ. 1 μM ATP was added to pUC19 and pUCap1 DNA substrates. This dissociated T4 Ligase from the DNA and kept it in solution, preventing it from inhibiting protein movement. To completely remove T4 DNA Ligase from the DNA, PCE and DNA precipitation was used.



**Figure 50:** Graphs showing diffusion constants ( $\mu\text{m}^2\text{s}^{-1}$ ) against their  $\alpha$  values for cFos and cJun as they travelled along PCE pUC19 DNA while in a high salt buffer. (a) cFos interacting with PCE pUC19 DNA ( $n = 37$ ) displays a range of diffusion constants between  $1 \times 10^{-5}$  and  $1 \times 10^{-2} \mu\text{m}^2\text{s}^{-1}$  and  $\alpha$  values between 0.4 and 1.9. (b) cJun interacting with PCE pUC19 DNA ( $n = 50$ ) shows a range of diffusion constants between  $1 \times 10^{-6}$  and  $1 \times 10^{-1} \mu\text{m}^2\text{s}^{-1}$  and  $\alpha$  values between 0.5 and 1.4. 10 pixels correspond to  $0.819 \mu\text{m}$ , the equivalent of 2410 bp (represented by the scale bar).



**Figure 51:** Graphs showing diffusion constants ( $\mu\text{m}^2\text{s}^{-1}$ ) against their  $\alpha$  values for cFos and cJun as they travelled along PCE pUCap1 DNA while in a high salt buffer. (a) cFos interacting with PCE pUCap1 DNA ( $n = 41$ ) displays a range of diffusion constants between  $1 \times 10^{-4}$  and  $1 \times 10^{-2}$   $\mu\text{m}^2\text{s}^{-1}$  and  $\alpha$  values between 0.4 and 1.5. (b) cJun interacting with PCE pUCap1 DNA ( $n = 40$ ) shows a range of diffusion constants between  $1 \times 10^{-6}$  and  $1 \times 10^{-2}$   $\mu\text{m}^2\text{s}^{-1}$  and  $\alpha$  values between 0.4 and 1.7. 10 pixels correspond to  $0.819 \mu\text{m}$ , the equivalent of 2410 bp (represented by the scale bar).

### **3.3.3: Protein colocalisation**

To determine whether DNA other than  $\lambda$  affected the formation of cFos:cJun heterodimeric complexes, PCE pUC19 and pUCap1 DNA substrates were used as these were completely free from T4 DNA Ligase. 83 proteins were observed bound to PCE pUC19. 28% were red cFos ( $n = 23$ ), 31% were green cJun ( $n = 26$ ) and 41% ( $n = 34$ ) were dual coloured cFos:cJun heterodimers. 242 proteins were also observed bound to PCE pUCap1. 23% ( $n = 53$ ) were red cFos, 37% ( $n = 90$ ) green cJun and 41% ( $n = 84$ ) were dual coloured cFos:cJun heterodimers. These results can be seen in Table 10.

	Colocalised % on PCE pUC19 DNA	Colocalised % PCE pUCap1 DNA
<b>cFos:cJun heterodimers</b>	<b>41 ± 1.8% (n = 83)</b>	<b>41 ± 1.8% (n = 242)</b>

**Table 10:** A table showing the percentage of colocalised, dual coloured cFos:cJun heterodimeric complexes (protein labelling described in chapter 2, method 3.1.2) that were bound to PCE pUC19 and PCE pUCap1 DNA substrates. All complexes were observed while in a high salt buffer.

### 3.3.4: Protein binding positions along Phenol:Chloroform extracted pUC19 and pUCap1 DNA

This part of the study follows on from Chapter 2 section 2.2.18.2 regarding where cFos, cJun and cFos:cJun heterodimers were pausing along  $\lambda$  DNA and to quantify the number of times the proteins paused within each AP1 and variant AP1 target sites. In this part of the study, the DNA substrates PCE pUC19 and PCE pUCap1 were used. All proteins were observed while in a high salt buffer.

As previously mentioned, originally it was thought that pUC19 was free from AP1 target sites and therefore target free. However the proteins were pausing. The question that arose was where? By revisiting the DNA sequence, and using the single nucleotide substitutions within the AP1 target sequence TGACTCA by Seldeen *et al* (2009), 11 variant sites were found. These are: TGACACA, TCACTCA, TAACTCA, TGACTCT, TCACTCA, GGACTIONCA, TTACTIONCA, TGCCTCA, TGACTCC, TCACTCA and TGTCTCA. These findings can be seen in Figure 55.

To find out whether cFos:cJun heterodimers bound to specific AP1 and variant AP1 target sites, the binding affinity ( $K_d$ ) for each site was determined using ITC by Seldeen *et al* (2009). In this study, the following sites with the highest  $K_d$  were located along  $\lambda$  DNA (DNA substrates pUC19 and pUCap1 could not be used due to a very small sample size) to determine which sites had the greatest occupancy by cFos:cJun heterodimers. There were 19 sites in total and these include; TGACTCA ( $K_d/\mu\text{M} = 0.21 \pm 0.01$ ), TGACGCA ( $K_d/\mu\text{M} = 0.64 \pm 0.01$ ) and TGACACA ( $K_d/\mu\text{M} = 0.92 \pm 0.03$ ). The sites that had the greatest cFos:cJun occupancy were TGACGCA and TGACTCA. These findings can be seen in Table 11 and Figure 52.

After the distances (bp) between more than 2 pause sites were identified and all configurations were calculated (as described in Chapter 2, method 2.2.19) to cover all possible distances, these distances were input into the Binding Site Position Finder (methods 3.2.23 and 3.2.24). The possibility of a protein visiting particular sites was recorded and the total percentage was calculated.

Table 14 shows cFos, cJun and cFos:cJun heterodimers that bound to variant AP1 target sites along PCE pUC19 DNA. cFos provided 218 distances (bp) between pause locations that in turn produced 5,614 binding site possibilities. This data shows that the percentages of cFos molecules bound to the variant AP1 target sites varied between 8.0-10.6%. The most highly occupied pause site was GGACTCA (highlighted green). cJun provided 686 distances (bp) between all pause locations and 18,314 binding site possibilities were produced. The percentage of cJun binding to variant AP1 target sites varied from 8.3-10.3%. Like cFos, the most highly occupied site was GGACTCA. cFos:cJun molecules provided 4 binding distances between pause locations and 114 binding site possibilities. Unlike cFos and cJun, cFos:cJun's frequently visited the site TGTCTCA (highlighted green).

Table 15 shows cFos, cJun and cFos:cJun heterodimers that bound to variant AP1 target sites along PCE pUCap1 DNA. cFos provided 405 distances (bp) between pause locations that in turn produced 12,718 binding site possibilities. The percentage of cFos molecules that bound to the AP1 target site was 9.1%. However, the percentage of cFos that bound to the variant AP1 target sites varied from 8.1-8.9%. Like cFos on PCE pUC19 DNA, the most highly occupied variant target site was GGACTCA (highlighted green). cJun molecules provided 393 distances between pause locations and therefore produced 12,356 binding site possibilities. 9.1% of cJun bound to the AP1 target site. The remaining cJun molecules that bound to the variant AP1 target sites varied from 7.5-9.1%. Like cFos, the most highly occupied site was GGACTCA. cFos:cJun heterodimers produced 159 distances that provided 4,800 binding site possibilities. 9.8% of cFos:cJun bound to its AP1 target site. The remaining molecules that bound to the variant AP1 target sites varied from 7.5-9.4%. Here, like cFos and cJun, cFos:cJun's most frequently visited site is GGACTCA. Overall, the most highly occupied site was AP1.

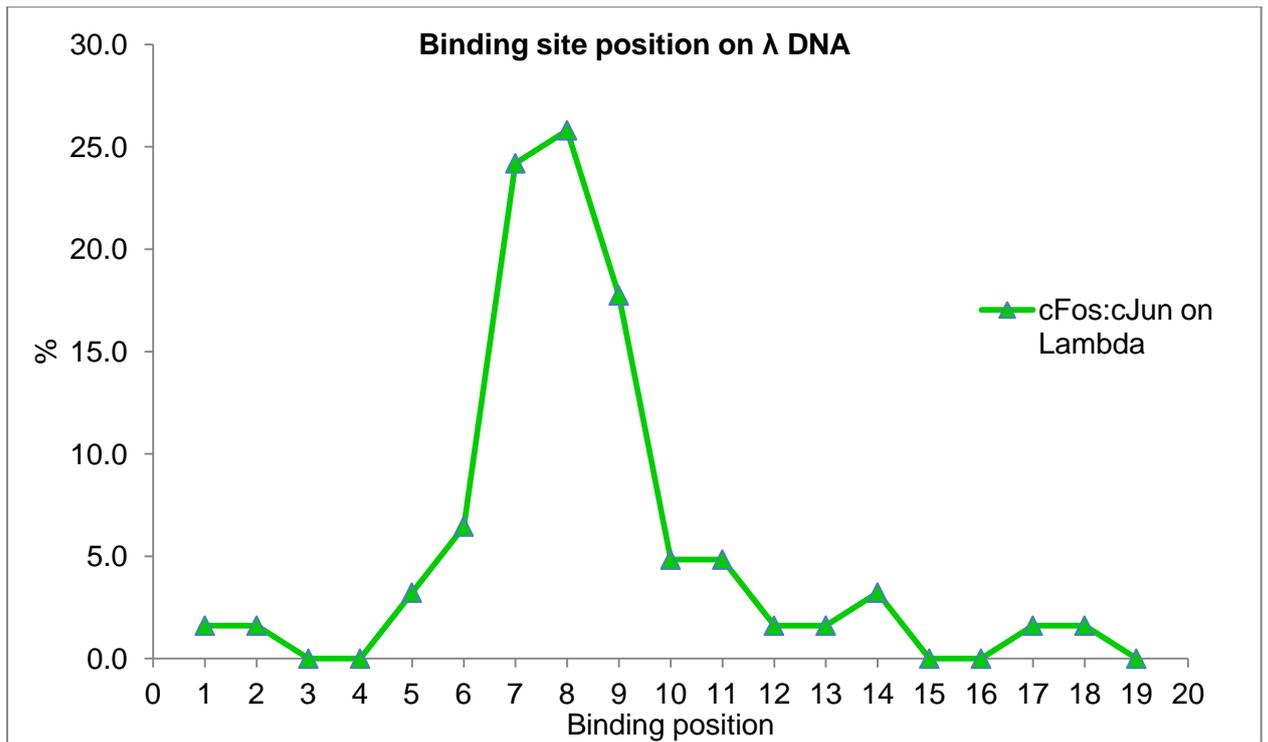
Correlation graphs was then plotted to identify which target sites were preferred by cFos, cJun and cFos:cJun heterodimers (as shown in Figure 54). These graphs show the percentage of pausing by all of the 3 proteins at each of the target sites. The first graph (a) shows cFos (blue), cJun (pink) and cFos:cJun (green). cFos and cJun appear to correlate visually, however the cFos:cJun heterodimers only show correlation at binding sites TCACTCA (site 5), TGCCTCA

(site 8) and TCACTCA (site 10). The second correlation graph (**b**) appears to show visual correlation between all 3 proteins.

As described in Chapter 2 section 3.1.3, the Binding Site Position Finder was validated using random numbers between 800 and 2600 (pUC19 and pUCap1 are 2686 bp and 2683 bp long, respectively) that were generated in Excel. These random numbers simulated the distances between pauses that a simulated protein took as it travelled along each of these DNA substrates. When the simulated protein travelled along pUC19 DNA, 36 distances were typed into the Binding Site Position Finder and these generated 1134 binding site possibilities. Similarly, when this protein travelled along pUCap1 DNA, 37 distances produced 1264 binding site possibilities. When this simulation is compared with cFos, cJun and cFos:cJun heterodimers pausing within their AP1 and variant AP1 target sites, visual correlation can be seen at positions 2, 3, 8 and 9 on pUC19 with cFos and cJun. Visual correlation can also be seen at positions 3 and 4 on pUCap1 with cFos, cJun and cFos:cJun heterodimers. This simulated data can be seen in Tables 12, 13 and Figures 53 and 55.

cFos:cJun heterodimers on $\lambda$ DNA			
	Start positions with error	n	%
TGACGCA	1465	1	1.6
TGACACA	2719	1	1.6
TGACTCA	3763	0	0.0
TGACACA	14054	0	0.0
TGACTCA	19039	2	3.2
TGACTCA	20016	4	6.5
TGACGCA	20484	15	24.2
TGACGCA	20882	16	25.8
TGACTCA	21073	11	17.7
TGACGCA	21976	3	4.8
TGACTCA	22602	3	4.8
TGACGCA	30241	1	1.6
TGACGCA	31264	1	1.6
TGACACA	32012	2	3.2
TGACTCA	36480	0	0.0
TGACTCA	41813	0	0.0
TGACTCA	44501	1	1.6
TGACGCA	46052	1	1.6
TGACGCA	48311	0	0.0

**Table 11:** Table showing the percentage of pausing within variant and AP1 target sites that cFos:cJun heterodimers ( $n = 62$ ) occupied with high binding affinity. Column 1 shows the 19 target sites in the order that they appear along  $\lambda$  DNA. All AP1 target sites have been highlighted yellow. Column 2 shows the start position of each site, including the error. Column 3 shows the number of times a cFos:cJun heterodimer occupied each site and column 4 shows each sites occupancy as a percentage. This varied between 1.6-25.8%.

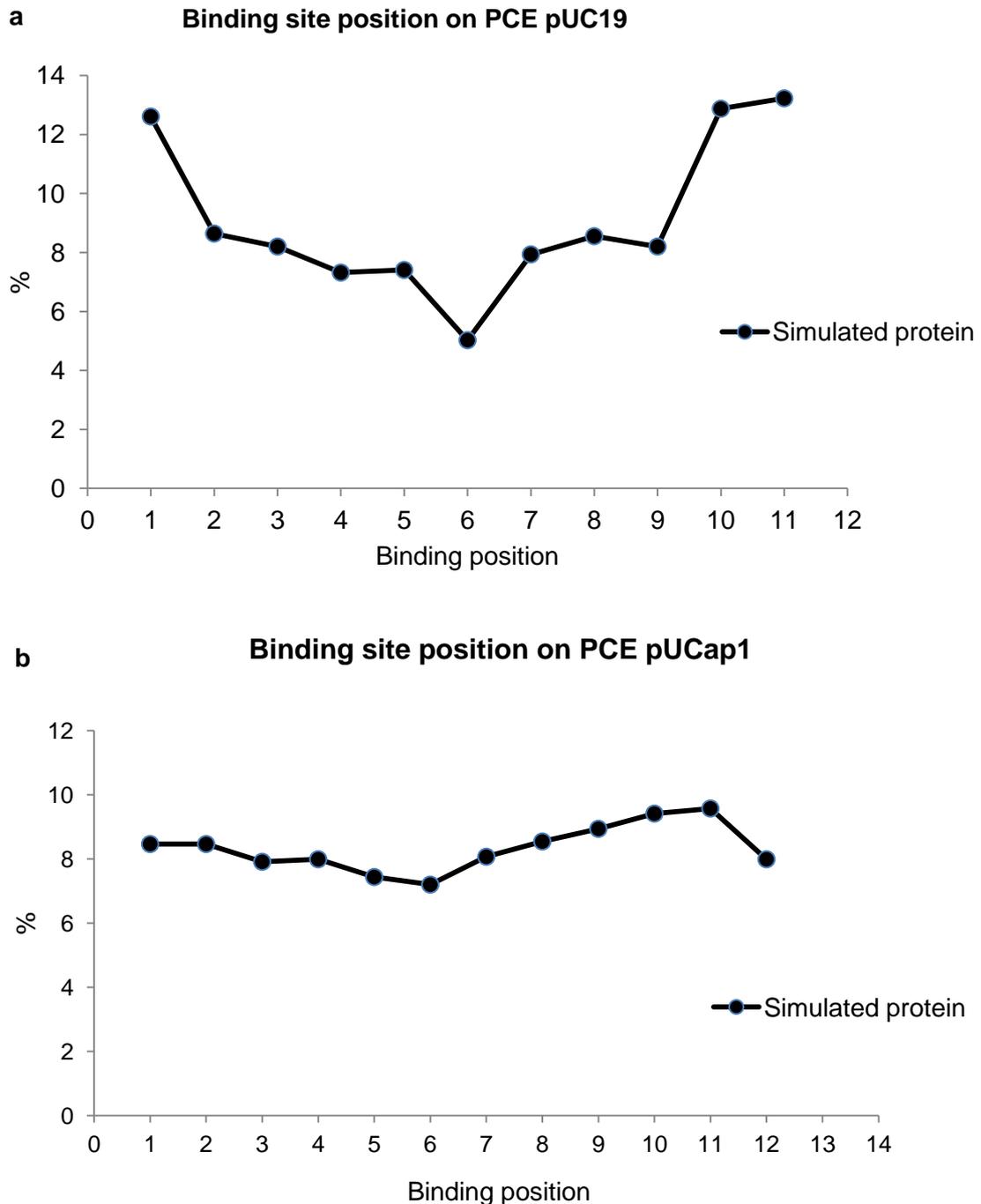


**Figure 52:** A graph representing the percentage of occupancy at each of the 19 high  $K_d$  variant and AP1 target sites (shown in Table 11) by cFos:cJun heterodimers ( $n = 62$ ).

Simulated protein on PCE pUC19			
	Start positions with error	n	%
TGACACA	-47	143	12.6
TCACTCA	479	98	8.6
TAACTCA	488	93	8.2
TGACTCT	625	83	7.3
TCACTCA	670	84	7.4
GGACTCA	1096	57	5.0
TTACTCA	1525	90	7.9
TGCCTCA	1564	97	8.6
TGACTCC	1613	93	8.2
TCACTCA	2021	146	12.9
TGTCTCA	2450	150	13.2

Simulated protein on PCE pUCap1			
	Start positions with error	n	%
TGACACA	-47	107	8.5
TGACTCA	165	107	8.5
TCACTCA	479	100	7.9
TAACTCA	488	101	8.0
TGACTCT	625	94	7.4
TCACTCA	670	91	7.2
GGACTCA	1096	102	8.1
TTACTCA	1525	108	8.5
TGCCTCA	1564	113	8.9
TGACTCC	1613	119	9.4
TCACTCA	2021	121	9.6
TGTCTCA	2450	101	8.0

**Tables 12 and 13:** Tables showing the percentage of pausing within each target site as a simulated protein travels along pUC19 (top table) and pUCap1 (bottom table) DNA substrates. The target sites have been listed in column 1 as they appear along the DNA and column 2 shows the corresponding start position, with error, for each site. Column 3 shows the number of times a protein visited each site and column 4 shows the number of visits per site as a percentage. Table 11 shows the percentage of pausing within each target site along pUC19. This varied between 5.0-13.2%. Table 12 shows the percentage of pausing within each target site along pUCap1. This varied from 7.2-9.6%.



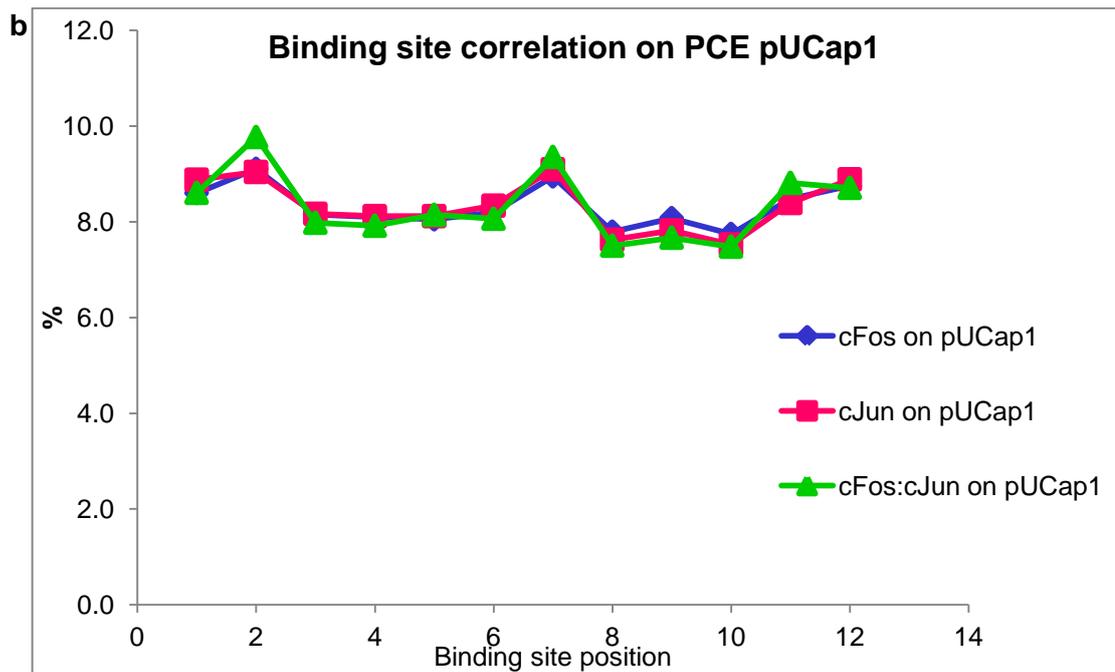
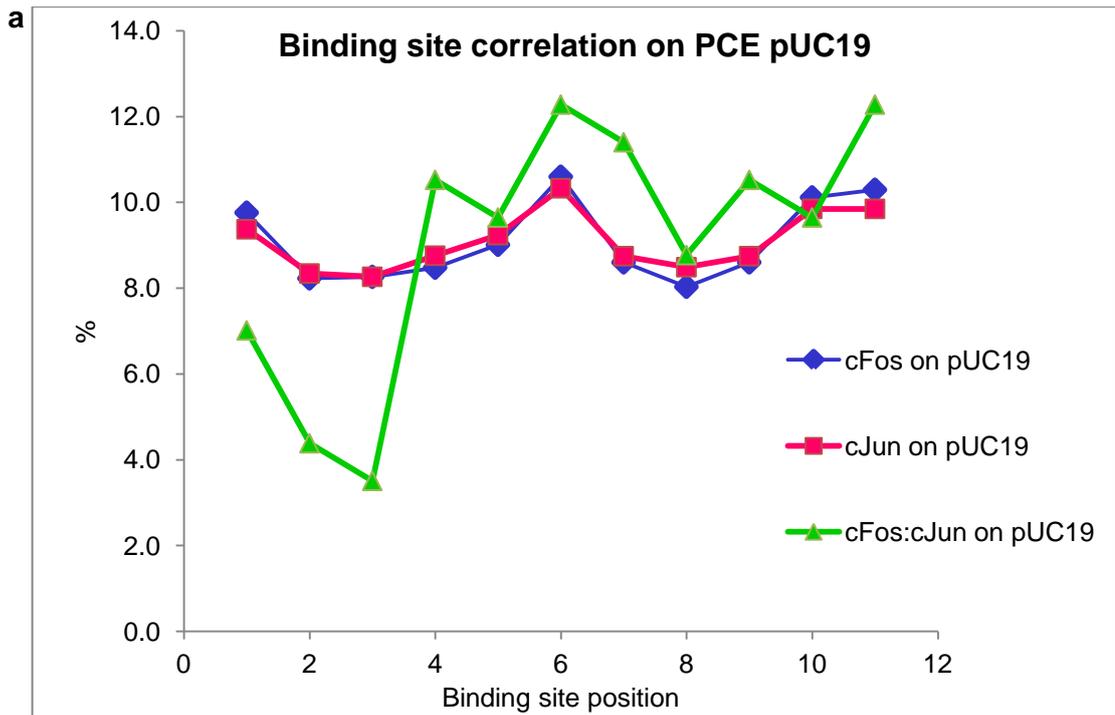
**Figure 53:** Graphs representing the percentage of pausing at each target site by a computer simulated protein as it travelled along DNA. (a) A computer simulated protein pausing within each of the 11 variant AP1 target sites as it travelled along pUC19 DNA. (b) A computer simulated protein pausing within the AP1 target site, located at binding position 2, and within the 11 variants AP1 target sites. The percentage of binding at each target site for pUC19 and pUCap1 DNA can be seen in Tables 12 and 13.

		A		B		C	
		cFos on PCE pUC19 P:C		cJun on PCE pUC19 P:C		cFos:cJun on PCE pUC19 P:C	
	Start positions with error	n	%	n	%	n	%
TGACACA	-47	548	9.8	1717	9.4	8	7.0
TCACTCA	479	462	8.2	1528	8.3	5	4.4
TAACTCA	488	464	8.3	1514	8.3	4	3.5
TGACTCT	625	476	8.5	1604	8.8	12	10.5
TCACTCA	670	506	9.0	1692	9.2	11	9.6
<b>GGACTCA</b>	1096	595	10.6	1892	10.3	14	12.3
TTACTCA	1525	483	8.6	1602	8.7	13	11.4
TGCCTCA	1564	451	8.0	1555	8.5	10	8.8
TGACTCC	1613	483	8.6	1602	8.7	12	10.5
TCACTCA	2021	568	10.1	1804	9.9	11	9.6
<b>TGTCTCA</b>	2450	578	10.3	1804	9.9	14	12.3

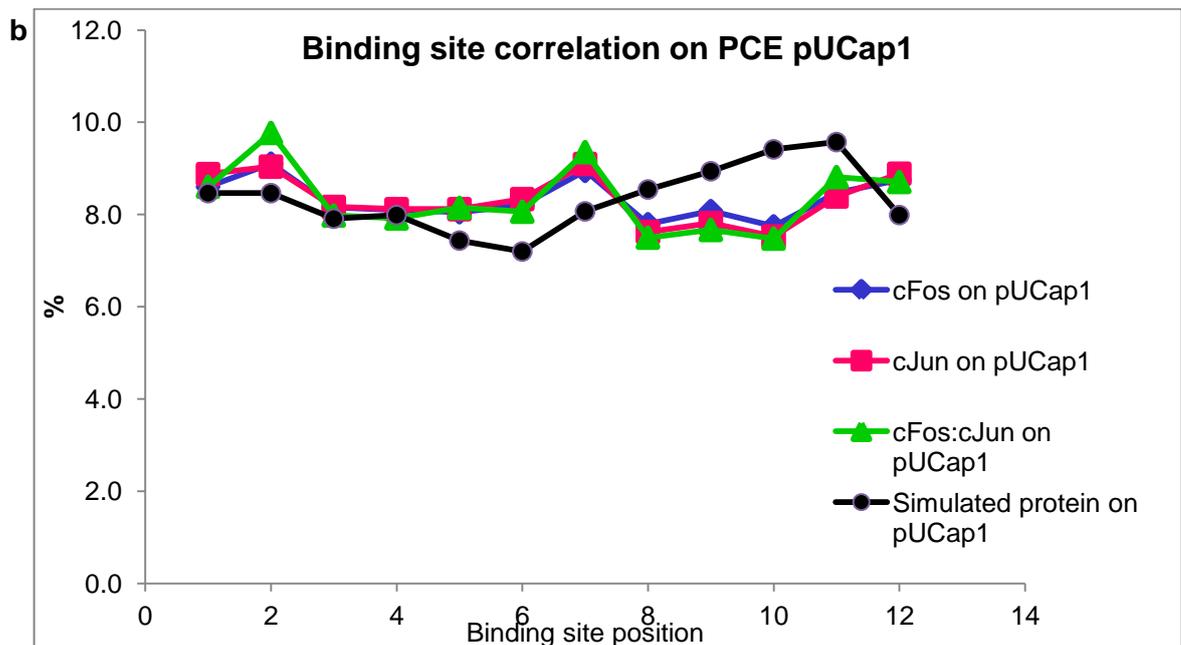
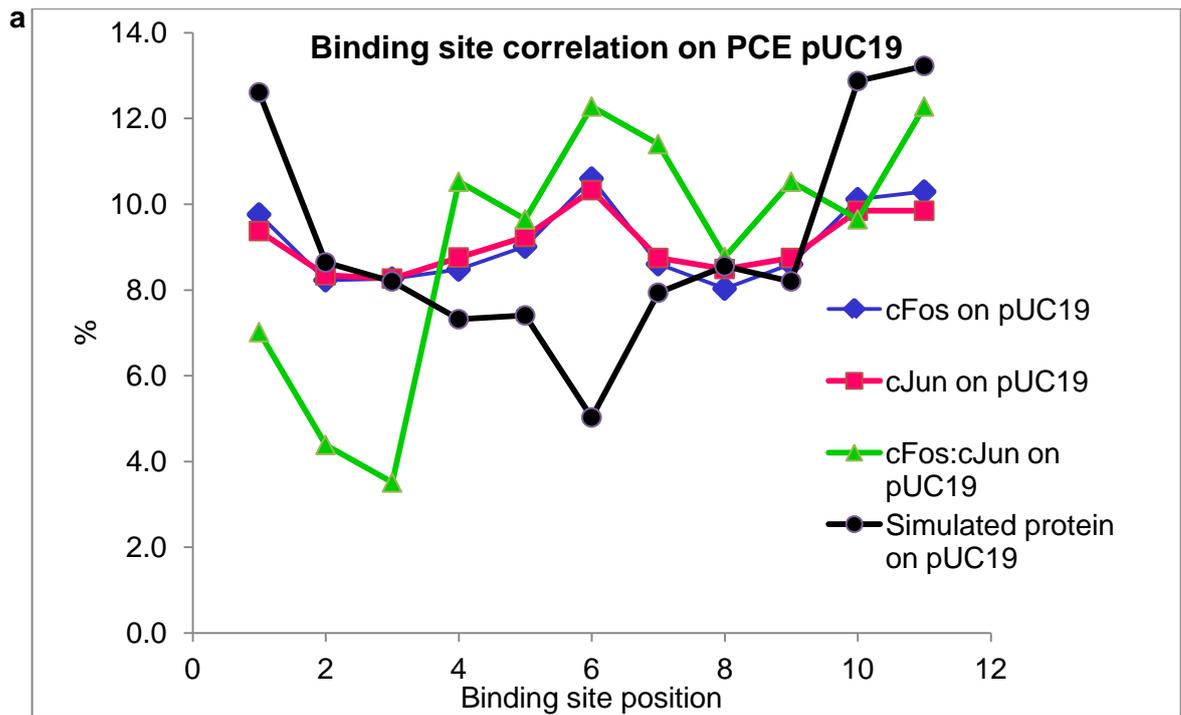
**Table 14:** A table showing the percentage of moving cFos, cJun and cFos:cJun molecules that bound to the variant AP1 target sites as they travelled along PCE pUC19 DNA. Column 1 shows the variant AP1 target sites in the order in which they appear along a piece of pUC19 DNA. Column 2 shows the starting position (including the 80 bp error margin) of each target site. Columns 3, 5 and 7 show the number of molecules that bound to each site and columns 4, 6 and 8 show this as a percentage. 10.6% of cFos ( $n = 52$ ) visited the variant AP1 target site GGACTCA (highlighted green), making this the most commonly visited site. 10.3% of cJun ( $n = 75$ ) also visited this variant AP1 target site GGACTCA (highlighted green). Like cFos, this was also cJun's most popular site. 12.3% of the cFos:cJun heterodimers ( $n = 2$ ) bound to both GGACTCA and TGTCTCA (highlighted green), therefore making these the most commonly visited sites.

		A		B		C	
		cFos on PCE pUCap1 P:C		cJun on PCE pUCap1 P:C		cFos:cJun on PCE pUCap1 P:C	
	Start positions with error	n	%	n	%	n	%
TGACACA	-47	1094	8.6	1097	8.9	413	8.6
<b>TGACTCA</b>	165	1158	9.1	1117	9.0	469	9.8
TCACTCA	479	1037	8.2	1009	8.2	383	8.0
TAACTCA	488	1030	8.1	1003	8.1	380	7.9
TGACTCT	625	1024	8.1	1003	8.1	391	8.1
TCACTCA	670	1043	8.2	1030	8.3	387	8.1
<b>GGACTCA</b>	1096	1138	8.9	1124	9.1	449	9.4
TTACTCA	1525	991	7.8	942	7.6	360	7.5
TGCCTCA	1564	1027	8.1	966	7.8	368	7.7
TGACTCC	1613	985	7.7	930	7.5	359	7.5
TCACTCA	2021	1078	8.5	1037	8.4	423	8.8
TGTCTCA	2450	1113	8.8	1098	8.9	418	8.7

**Table 15:** A table showing the percentage of moving cFos, cJun and cFos:cJun molecules that bound to AP1 (highlighted in yellow) and the variant AP1 target sites as they travelled along PCE pUCap1 DNA. Column 1 shows the AP1 and variant AP1 target sites in the order in which they appear along a piece of pUCap1 DNA. Column 2 shows the starting position (including the 80 bp error margin) of each target site. Columns 3, 5 and 7 show the number of molecules that bound to each site and columns 4, 6 and 8 show this as a percentage. Panel A) 9.1% of cFos ( $n = 52$ ) bound to their AP1 target site. 8.1-8.6% of cFos bound to the variant AP1 target sites. The most popular variant AP1 target site is GGACTCA (highlighted green). Panel B) 9.0% cJun ( $n = 46$ ) bound directly to their AP1 target site. 7.5-9.1% cJun also bound to variant AP1 target sites. Like cFos, cJun have the same popular AP1 variant target site GGACTCA. Panel C) 9.8% cFos:cJun heterodimers ( $n = 14$ ) bound to their AP1 target site. 7.5-9.4% bound to the variant AP1 target sites. CFos:cJun heterodimers also share the same popular variant AP1 target site GGACTCA with cFos and cJun.



**Figure 54:** These graphs show the correlation between the percentage of cFos, cJun and cFos:cJun heterodimeric moving molecules that paused within their target sites while travelling along PCE pUC19 and PCE pUCap1 DNA substrates. (a) cFos ( $n = 52$ ) and cJun ( $n = 75$ ) molecules appear to visually correlate. However cFos:cJun heterodimers ( $n = 2$ ) do not follow the same correlation pattern. (b) cFos ( $n = 52$ ), cJun ( $n = 46$ ) and cFos:cJun heterodimers ( $n = 14$ ) appear to visually correlate.



**Figure 55:** Graphs showing the percentage of pausing at each site by cFos, cJun, cFos:cJun heterodimers and the simulated protein on PCE pUC19 and pUCap1 DNA to observe any correlation. (a) Visually, correlation appears to occur at positions 2, 3, 8 and 9 between the simulated protein and cFos, cJun and cFos:cJun heterodimers. (b) Apparent visual correlation between the proteins and the simulated protein can be seen at positions 3 and 4.

### 3.4: Discussion

#### 3.4.1: Protein movement along pUC19, pUCap1 and TF $\lambda$

pUC19 was the first DNA substrate used to form tightropes as it did not contain any AP1 target sites and was therefore believed to be target free. As pUC19 is only 2686 bp long, the DNA was tandemised using T4 DNA Ligase, enabling tightropes to be created matching the length of  $\lambda$  DNA or longer. This DNA substrate was used to understand how proteins would interact with DNA when their AP1 target sites were not present.

The DNA substrate pUCap1 was also studied in parallel with pUC19 DNA. This is a variation of pUC19 whereby a single AP1 target site was introduced at position 417 within the plasmid pUC19. This DNA was constructed to understand whether a single AP1 target site located every 2683 bp (length of pUCap1) would affect protein interaction with the DNA and whether the proteins would pause at this specific site only.

The first protein used in this study was cJun as it was the most diffusive protein out of all of the complexes (as shown in Chapter 2, section 3.1.1). 242 cJun proteins were observed bound to pUC19 tightropes and 54% were seen moving. This percentage of mobility is 17% less (t-test p-value = 0.06) than when this protein travelled along  $\lambda$  DNA. Although this is not a statistically significant result, it shows a trend between the percentage of cJun diffusing along pUC19 and  $\lambda$  DNA. In a parallel study, 161 cJun proteins were also observed bound to pUCap1 DNA and 46% (t-test p-value = 0.72) were seen moving. Like cJun on pUC19, this was also an unexpected finding as the percentage of moving cJun along pUCap1 DNA was now similar to cFos (t-test p-value = 0.72) as it moved along  $\lambda$  DNA.

There are two main hypotheses that could explain pausing in the absence of specific sites. Firstly, it may be possible that T4 DNA Ligase was inhibiting protein movement by acting as an obstacle. Secondly, pUC19 and pUCap1 have regions of target free DNA that vary between 9 bp and 526 bp long. When compared with  $\lambda$  DNA where the longest non-specific region is 3216 bp, these regions are much smaller resulting in the proteins having less non-cognate DNA to search for their target sites and may pause within regions where they have high affinity.

To determine whether T4 DNA Ligase was inhibiting cJun from travelling along pUC19 and pUCap1 DNA, 1  $\mu$ M ATP was added to the high salt buffer to remove the Ligase from the DNA. 37% of cJun travelled along pUC19 DNA and 42% (t-test p-value = 0.82) travelled along pUCap1 DNA. Although these results are not statistically significant when compared, they do show that cJun is less diffusive. This suggests that T4 DNA Ligase is preventing the protein from moving. Adding ATP to the imaging buffer aimed to prevent the T4 DNA Ligase from binding to the DNA. However instead, ATP may have been evoking T4 DNA Ligase to deadenylate, causing it to form a stable enzyme-DNA complex and stall rather than remaining in solution (Rossi et al. 1997).

PCE and DNA precipitation was used to completely remove the enzyme from the DNA (Palmiter 1974). This meant that the DNA would be free from obstacles and cJun could move freely. 58% (t-test p-value = 0.55) of cJun now moved on pUC19 and 52% (t-test p-value = 0.54) moved on pUCap1 when compared with the 71% of cJun that travelled along  $\lambda$  DNA. These differences are not statistically significant. This indicates that pUC19 and pUCap1 DNA substrates may contain non-specific regions that cJun has a higher affinity for in addition to the known AP1 and variant AP1 target sites. This may account for why there is less movement when compared with cJun diffusing along  $\lambda$  DNA.

cFos was also loaded onto pUC19 and pUCap1 DNA that was free from T4 DNA Ligase. 49% and 55% of cFos (t-test p-value = 0.56) travelled along pUC19 and pUCap1 DNA, respectively. Despite the limited sample size, no significant difference could be seen when cFos moved along these DNA substrates. There is little variation between cFos moving on pUC19 (t-test p-value = 0.35) and pUCap1 (t-test p-value = 0.20) when compared with the percentage of movement along  $\lambda$  DNA. These comparisons show no statistical significance. To understand how cFos is capable of homodimerising, structural studies such as computational analysis will need to be performed. This involves studying the coiled coil structure, and alternative structures generated by algebraic parameters, to explore the significance of conformational specificity and any physiological changes upon binding (Kortemme and Baker 2004). Another structural study that could be performed involves using x-ray crystallography to understand how the protein binds to

and interacts with DNA (Rohs et al. 2010). 47% and 32% (t-test p-value = 0.13) of cFos:cJun heterodimers were observed moving along PCE pUC19 and pUCap1 DNA, respectively.

### **3.4.2: Protein colocalisation**

Colocalisation studies, were performed to quantify the amount of cFos:cJun heterodimeric complexes. As described in Chapter 2, red labelled cFos was mixed with green labelled cJun. This produced 3 differently coloured complexes: red cFos, green cJun and dual coloured cFos:cJun heterodimers. 41% colocalisation was observed between cFos with cJun on both pUC19 (t-test p-value = 0.54) and pUCap1 (t-test p-value = 0.18) DNA substrates. This is a 9% increase when compared to colocalised cFos:cJun heterodimers that were observed bound to  $\lambda$  DNA. This data indicates that partner swapping may be taking place within the AP1 family (Turner 1989). Prior to visualisation using OAF microscopy in this study, previous studies using EMSAs have shown that partner swapping is capable within the AP1 family due to the protein complexes that had bound to DNA (Hellman and Fried 2007), however, how this occurs is yet to be understood.

### **3.4.3: Movement along DNA**

To understand how cFos, cJun and cFos:cJun heterodimers move along both pUC19 and pUCap1 DNA substrates in search of their AP1 target sites, their diffusion constants and  $\alpha$  values were determined. While travelling along PCE pUC19 DNA, cFos and cJun had diffusion constants of  $1.5 \times 10^{-3} \mu\text{m}^2\text{s}^{-1}$  and  $2.6 \times 10^{-3} \mu\text{m}^2\text{s}^{-1}$ , meaning that both proteins could diffuse along these different DNA substrates. The corresponding  $\alpha$  values for cFos and cJun were both 0.9. These  $\alpha$  values indicate that cFos and cJun are both randomly diffusing along the DNA in search of their AP1 target sites while potentially using anomalous sub-diffusion since their values are  $<1$  (Saxton 1997). When cFos and cJun diffused along PCE pUCap1, their diffusion constants were  $1.9 \times 10^{-3} \mu\text{m}^2\text{s}^{-1}$  and  $2.4 \times 10^{-3} \mu\text{m}^2\text{s}^{-1}$ , respectively. Like pUC19 DNA, cJun (t-test p-value = 0.54) is more mobile on pUCap1 DNA in comparison to cFos. This may be due to the stable core Leucine residues inside of the cJun dimerisation region (Wagner 2001). The corresponding  $\alpha$  values for cFos and cJun were 0.9 and 1.0, respectively. These values are also  $<1$  indicating that when the proteins are on pUCap1, also randomly diffuse. Therefore, as

described in Chapter 2, the proteins may be using 3-Dimensional and 1-Dimensional searching mechanisms (Kampmann 2004, Gorman and Greene 2008) to locate their AP1 target sites along the DNA.

#### **3.4.4: Pausing along the DNA**

There are no AP1 target sites present along pUC19 DNA, however cJun were observed pausing. Based on the study by Seldeen *et al* (2009) who made single nucleotide substitutions throughout the AP1 sequence (as described in sections 2.1, 2.2.6.2 and 2.3.3), pUC19 and pUCap1 DNA sequences was checked for the presence of these variant AP1 target sites. Along both DNA substrates, there were 11 variant AP1 target sites.

As cFos, cJun and cFos:cJun heterodimers diffused along PCE pUC19 and pUCap1 DNA substrates, they were observed to pause. Similarly, as described in Chapter 2, the binding site position finder was used to determine whether the proteins were pausing non-specifically, within an AP1 target site (with the exception of pUC19 DNA) or within the 11 variant AP1 target sites.

The percentage of cFos and cJun that bound to the variant AP1 target sites located along PCE pUC19 DNA varied from 8.1-8.9% and 8.3-10.3% respectively. cFos:cJun heterodimers binding to the variant AP1 target sites varied from 7.5-9.4% favouring TGTCTCA. The percentage of cFos, cJun and cFos:cJun heterodimers that bound to their AP1 target sites located on PCE pUCap1 DNA were 9.1%, 9.1% and 9.8% respectively. The percentage of cFos, cJun and cFos:cJun heterodimers that bound to the variant AP1 target sites varied from 8.1-8.9%, 7.5-9.1% and 7.5-9.4%. These results show little variation in the percentage of pausing within their variant AP1 or AP1 target sites while diffusing along pUC19 and pUCap1 DNA substrates.

These results support the “waiting room model”, as described in section 1.8 and Figure 9, a mechanistic approach whereby the variant AP1 target sites may be used by the proteins to control transcription. How this occurs is currently unknown, however protein-DNA interaction studies such as EMSAs (Hellman and Fried 2007), Filter binding assays (Rio 2012), x-ray crystallography (Rohs *et al.* 2010) and computational analysis (Kortemme and Baker 2004) could be performed to understand how this occurs.

The final DNA substrate engineered was TF $\lambda$ . This was created to understand how cJun interacted with DNA that was completely AP1 and variant AP1 target site free. Only 2 cJun proteins were observed to diffuse out of the 79 that were observed bound to this DNA moved. This result was unexpected as it contradicts 1-Dimensional mechanistic theories suggesting that a protein will search the DNA for its intended target site (Wang et al. 2006, Li et al. 2009). To ensure that cJun had not degraded, the same batch of proteins was loaded onto  $\lambda$  DNA as previously 71% were moving (as described in Chapter 2). The protein was observed diffusing along  $\lambda$  DNA which showed that cJun had not degraded. This result therefore suggests that cJun does have high affinity binding sites on TF $\lambda$  as multiple bound proteins were observed.

As cJun was observed bound to the DNA, this suggests that the protein does have an affinity for the target free substrate. Based on studies regarding 3-Dimensional DNA searching methods (Gorski et al. 2006, Li et al. 2009, Mirny et al. 2009b), it may be possible that cJun used 3-Dimensional searching to locate high affinity areas along the DNA. Once bound, why cJun was predominately static currently remains unknown, especially as this protein was the most diffusive out of all the complexes used throughout this study. To my knowledge, there is no evidence in the literature about how AP1 proteins bind and diffuse along non-cognate areas of DNA.

Therefore, to understand whether a lack of AP1 and variant AP1 target sites could be contributing towards cJun becoming static when bound to non-cognate areas of DNA, in the future a single AP1 target site could be cloned into TF $\lambda$ . This future study may assist with determining whether this target site does promote protein movement along the DNA.

### **3.4.5: Conclusion**

The results in this chapter have shown that cFos, cJun and cFos:cJun heterodimers are able to bind onto different DNA substrates in addition to  $\lambda$  DNA (as mentioned in Chapter 2) and approximately half of the observed protein population moved. However, in contrast, cJun became predominately static when observed bound to TF $\lambda$ . This may be due to the DNA having a lack of AP1 and variant AP1 target sites. Why this occurs may be due to evolution, whereby

point mutations, deletions and substitutions within the DNA binding domains of these transcription factors results in a high affinity towards their AP1 target sites or any site that has a similar sequence. Although the leucine zipper remains conserved, the DNA binding domain is synonymous (Mechta-Grigoriou et al. 2001). This may occur through positive selection, improving the efficiency of the protein selecting specific binding sites (Pál et al. 2006). To understand why, future studies could be performed such as inserting an AP1 target site into TF $\lambda$  to determine whether more protein movement will occur. In addition, studies can be performed to understand which areas of non-specific DNA cFos, cJun and cFos:cJun heterodimers have a high affinity for and whether this also promotes protein movement.

#### **Chapter 4: Final conclusions**

Since the discovery of cancer, medical treatments have evolved and the first known treatment to arrest cellular growth was cytotoxic nitrogen mustard during World War II (Ewig and Kohn 1977). This became the first chemotherapeutic agent. Now, modern day medicine has evolved whereby chemotherapy and radiotherapy have been developed to try and preserve human life by attempting to eradicate cancer.

Since the discovery of the AP1 family of transcription factors (Bohmann et al. 1987, Vinson et al. 1993), it has been important to understand their role in tumorigenesis (Zhang et al. 2007, Latchman 1995). It has been well documented in cancer research that when in excess, cFos causes osteosarcomas and cJun causes hepatocellular carcinoma and squamous cell carcinoma (Eferl and Wagner 2003, Shaulian and Karin 2001, Wang et al. 1995, Zhang et al. 2006).

Understanding the mechanisms used by cFos and cJun as they interact with themselves and the DNA is vital for the development of site specific pharmaceutical agents. The antagonistic peptide inhibitor FosW was designed to interact with cJun (Mason et al. 2007a, Mason et al. 2006). In the future, FosW can potentially lead to a pharmaceutical agent that could reduce the excess quantities of cJun and hopefully free patients of hepatocellular carcinoma and squamous cell carcinoma.

In this thesis, Chapter 2 aimed to identify how cJun, cFos:cJun heterodimers and FosW+cJun interacted with  $\lambda$  DNA and each other in real time. In addition, whether cFos could form protein-DNA complexes and diffuse was also explored. Table 3 has shown the percentage of diffusion by each of the protein complexes as they search the DNA for target sites. cFos, cJun, cFos:cJun heterodimers, FosW and Fos+cJun diffused 45%, 71%, 39%, 55% and 39%, respectively. As the proteins searched the DNA, the diffusion constants for cFos, cJun and FosW were  $3.4 \times 10^{-3} \mu\text{m}^2\text{s}^{-1}$ ,  $8 \times 10^{-4} \mu\text{m}^2\text{s}^{-1}$  and  $1.2 \times 10^{-2} \mu\text{m}^2\text{s}^{-1}$ , respectively. These complexes also used anomalous sub-diffusion due to  $\alpha$  values being  $<1$ . This data can be seen in Figure 28. Upon loading the proteins onto the DNA tightropes, evidence of colocalisation and partner

swapping (shown in Table 4) was observed on the DNA. cFos, cJun, cFos:cJun heterodimers and FosW+cJun showed 27%, 27%, 32% and 25% colocalisation, respectively.

The final part of Chapter 2 looked at whether cFos, cJun and cFos:cJun heterodimers on paused within their AP1 target sites where they are known to exert their biological function. Tables 6, 7 and 8 show that the proteins not only paused within their AP1 target sites, they also paused within variant AP1 target sites as well. This data shows that each site had ~2% occupancy. All of the aims itemised in this chapter were achieved.

Chapter 3 expanded on the aims mentioned in Chapter 2 by observing how cFos, cJun and cFos:cJun heterodimers interacted with different DNA substrates. These included; AP1 target site free pUC19, pUCap1, a variation of pUC19 incorporating a single AP1 target site, and AP1 and variant AP1 target site free TF $\lambda$  DNA. Partner swapping was also determined on pUC19 and pUCap1 DNA. Firstly, Table 9 showed that 49% cFos, 58% cJun and 47% cFos:cJun heterodimers diffused along PCE pUC19. 55% cFos, 52% cJun and 32% cFos:cJun heterodimers also diffused along PCE pUCap1 DNA. Secondly, as cFos and cJun diffused along PCE pUC19 DNA in search of target sites, their diffusion constants were  $1.5 \times 10^{-3} \mu\text{m}^2\text{s}^{-1}$  and  $2.6 \times 10^{-3} \mu\text{m}^2\text{s}^{-1}$ . When cFos and cJun searched PCE pUCap1 DNA for target sites, their diffusion constants were  $1.96 \times 10^{-3} \mu\text{m}^2\text{s}^{-1}$  and  $2.46 \times 10^{-3} \mu\text{m}^2\text{s}^{-1}$ . When searching both pUC19 and pUCap1 DNA, cFos and cJun used anomalous sub-diffusion due to  $\alpha$  values being  $<1$  (shown in Figures 48 and 49).

To determine whether cFos and cJun were able to heterodimerise on pUC19 and pUCap1 DNA substrates, the proteins were loaded onto each DNA type. 41% colocalisation was observed on both PCE pUC19 and PCE pUCap1. This data is shown in Table 10.

The next part of this chapter was to observe how cJun interacted with TF $\lambda$ . Only 3% of molecules diffused which was unexpected since this DNA substrate is completely AP1 and variant AP1 target site free. Why this occurs is currently unknown and to my knowledge, there is no evidence in the literature suggesting why this may occur when the protein stall on non-cognate DNA. Therefore in the future, an AP1 target site could be cloned into TF $\lambda$  to determine whether AP1 and variant AP1 target sites do promote protein movement.

The final part of this chapter identified the percentage of cFos, cJun and cFos:cJun heterodimers that occupied each AP1 and variant AP1 target site located along pUC19 and pUCap1. This data showed that each pUC19 target site was ~12% occupied and ~10% of each site located along pUCap1 DNA also showed occupancy. This data is shown in Tables 14 and 15. All of the aims within this chapter have been achieved, however further work will need to be performed on cJun forming a protein-DNA complex with TF $\lambda$  as described above.

The data shown in Chapters 2 and 3 regarding the binding position of cFos, cJun and cFos:cJun heterodimers along each DNA substrate appears to support the “waiting room” model. However, future studies should be performed to support this model such as Filter Binding Assays. Instead of using an oligonucleotide containing a single AP1 target site, each short section of DNA should contain each variant AP1 target sites. cFos, cJun and cFos:cJun protein complexes should interact with each oligonucleotide variation during EMSA studies.

Therefore, within this final chapter, the major conclusions drawn from this study will be discussed and any work that could be performed in the future to enhance this study and its findings will also be discussed.

#### 4.1: Major conclusions from this thesis

This study has shown that cJun can bind to  $\lambda$  DNA and diffuse along it, homodimerise or heterodimerise with cFos, as suggested in the literature (Landschulz et al. 1988, Hai and Curran 1991, Ito et al. 2001). Dimerisation was observed by performing dual coloured colocalisation studies. Surprisingly, cFos is capable of binding to the DNA and diffusing along it, albeit more slowly than cJun, and homodimerising with itself. This was a previously unseen observation in earlier *in vitro* studies such as EMSAs as cFos was found to be less stable in comparison to cJun and cFos:cJun heterodimers while showing no homoassociation (Porte et al. 1995). This lack of association may have occurred due to the forces experienced by unstable cFos molecules as they travel through the gel matrix (Orchard and May 1993). However, a recent study by Szalóki *et al* (2015) used Förster resonance energy transfer (FRET) on live HeLa cells containing cFos and showed that cFos is capable of homodimerising (Szalóki et al. 2015). This recent evidence suggests that cFos is capable of binding to chromatin and may possibly act as a transcriptional regulator, especially when over-expressed during oncogenesis (Szalóki et al. 2015).

When the biological peptide inhibitor FosW, a modified version of cFos, was loaded onto the DNA, 25% colocalisation was observed between itself and its intended target cJun. This shows that FosW is able to heterodimerise with its intended target cJun for which it has a high affinity, more so than cJun has for itself, therefore confirming the original hypothesis. Upon binding, FosW renders cJun inactive as static complexes were seen. It was also found that FosW diffuses along the DNA in search of its intended target slower than cJun itself.

Traditionally FosW does not contain a DNA binding domain. However, for the purpose of this study a cJun DNA binding domain was cloned into FosW. This enabled any interaction with cJun to be visualised. Unlike wild-type cFos:cJun heterodimers that has a melting temperature ( $T_m$ ) 16 °C, FosW+cJun has a  $T_m$  63 °C, indicating that this complex has high stability, specificity and a low dissociation constant ( $K_d$ ) in comparison to wild-type cFos:cJun heterodimers; this has been previously determined by Circular Dichroism (Mason et al. 2006). This high stability and specificity that FosW has for cJun may be due to the optimised core packing,

electrostatic interactions and natural tendency to form an  $\alpha$ -helical structure due to residues within the dimerisation domains at positions (See Figure 3) being substituted with Leucine and Isoleucine (Mason et al. 2006) and may contribute towards slower 1-Dimensional diffusion.

While diffusing along the DNA in search of their AP1 target sites, cFos, cJun and cFos:cJun heterodimers were observed to pause within variant AP1 target sites in addition to AP1 binding sites. By incorporating a standard error margin of 80 bp either side of each AP1 and variant AP1 target site, the Binding Site Position Finder determined whether the distance between pause sites, as a protein diffused along the DNA, corresponded to binding into AP1 or variant AP1 target sites. There was little difference in the percentage of pausing within AP1 and the variant AP1 target sites. This suggests a “waiting room model” where the proteins may travel in and out of the variant AP1 target sites until an AP1 target site becomes free so that they can exert their biological function. This may be a way of controlling transcription.

To understand how the proteins interacted with different DNA substrates, pUC19, pUCap1 and TF $\lambda$  DNA templates were used. The DNA substrate pUC19 is canonical AP1 binding site free, however it does contain 11 variant AP1 target sites. The proteins were observed to pause on this template. A similar observation was seen when the proteins were loaded onto pUCap1. This data is in keeping with the “waiting room model” as the proteins were non-specifically pausing in the variant AP1 target sites regardless of whether an AP1 target site was present or not.

When cJun was loaded onto TF $\lambda$  DNA, the majority of the proteins were static despite the fact that on  $\lambda$  the same protein preparation on the same day was seen diffusing. Why this occurs is currently unknown and to my knowledge, there is no evidence in the literature suggesting why this may happen. Therefore, to determine whether the target sites may be contributing towards protein movement, a single AP1 target site could be cloned into TF $\lambda$  DNA.

## 4.2: Future work

This study has been able to answer some questions but at the same time has raised even more. Throughout the different sections of this thesis, these questions have been outlined and methods have been suggested to try to answer them. Therefore, this final section of the thesis aims to propose ideas that may be performed in the future to advance this study.

In Chapter 2 it was highlighted that cFos is capable of homodimerising and this was confirmed by dual labelling studies. These dual labelling studies also confirmed the formation of cJun homodimers and cFos:cJun heterodimers. Partner swapping within the AP1 family clearly occurs, although how each protein selects its intended partner is currently unknown. To gain an insight into how the proteins partner swap, the binding affinity of each protein could be determined using Fluorescence Correlation Spectroscopy (FCS). This technique uses a confocal microscope to image fluorescently labelled molecules by focusing the laser right down to the optical resolution limit. A solution volume well below micromolar range becomes confined (Bacia and Schwille 2007). Any fluorescence emitted from the molecules as they pass through this tiny space enables their physics parameters to be determined. These parameters include; molecule concentrations; mobility; rate constants; interactions and dynamics of the molecules. Any intensity in fluorescence is due to the molecules fluctuating as they diffuse through the detection volume. Using an extension of FCS such as dual FCS allows two corresponding binding partners to be cross-correlated, enabling the physics parameters between partners to be determined. The proteins binding affinity can be calculated based on the increased fluorescence (Ries and Schwille 2012).

The use of dual FCS in relation to partner swapping within the AP1 family will prove a useful imaging tool. This will give an insight into the rate of reaction and the dynamics behind how these proteins interchange partners.

Chapters 2 and 3 show that as cFos, cJun and cFos:cJun heterodimers searched the DNA, they paused within AP1 and variant AP1 target sites. As previously described, the “waiting room model” may provide one explanation as to why these proteins pause in these variant sites in

addition to AP1. This scenario suggests that the proteins may intermittently bind and pause at these variant sites. The proteins leave these variant sites when an AP1 target site becomes available so that they can exert their biological function. To understand the interaction between a protein with DNA, computational biology studies using bio-modelling could be performed. This approach could help to identify whether the dimers undergo any conformational changes upon binding to each of the target sites (Kortemme and Baker 2004). This may shed light on which sites are preferred by the proteins and may help support the “waiting room model.” X-ray crystallography could also be performed to identify the protein-DNA structure (Leonard et al. 1997) upon binding to each variant AP1 target site as single molecule studies are unable to do this.

It was hypothesised in Chapter 3 that cJun would continuously search non-cognate regions of DNA for AP1 target sites. To test this hypothesis, the DNA substrate TF $\lambda$  was created and used. cJun bound to the DNA, however minimal movement was observed. One study that could be performed is to clone a single AP1 target site into TF $\lambda$  to see whether cJun will use 1-Dimensional diffusion.

The antagonistic peptide inhibitor FosW still has scope for future experiments to be performed. In this study, 39% colocalisation between FosW and its target cJun was observed when bound to the DNA. In the future, the ratio of the inhibitor FosW to cJun would have to be optimised in relation the amount of over-expressed cJun and the progression of hepatocellular carcinoma and squamous cell carcinoma. In addition, this study only looked at FosW and cJun together on DNA tightropes. Whether FosW is capable of locating and binding to cJun when cFos is present could be studied using triple colour experiments, an enhanced version of the method used in this study to visualise colocalised complexes in dual colour. Imaging in triple colour would involve fluorescently labelling each protein and FosW with a different coloured Qdot (as described in method 2.2.9), loading the proteins and inhibitor onto the DNA tightropes and imaging using OAF (described in method 2.2.11). All coloured complexes could be identified and quantified. This method could determine whether FosW can disrupt cFos:cJun heterodimers as well as cJun homodimers.

## References

- Abate, C., Luk, D. and Curran, T. (1991) 'Transcriptional regulation by Fos and Jun in vitro: interaction among multiple activator and regulatory domains', *Molecular and cellular biology*, 11(7), 3624.
- Abd-Elsalam, K. A. (2003) 'Bioinformatic tools and guideline for PCR primer design', *african Journal of biotechnology*, 2(5), 91-95.
- Alivisatos, A. P. (1996) 'Perspectives on the physical chemistry of semiconductor nanocrystals', *The Journal of Physical Chemistry*, 100(31), 13226-13239.
- Angel, P., Allegretto, E. A., Okino, S. T., Hattori, K., Boyle, W. J., Hunter, T. and Karin, M. (1988) 'Oncogene jun encodes a sequence-specific trans-activator similar to AP-1'.
- Angel, P. and Karin, M. (1991) 'The role of Jun, Fos and the AP-1 complex in cell-proliferation and transformation', *Biochimica et biophysica acta*, 1072(2-3), 129.
- Antelman, J., Ebenstein, Y., Dertinger, T., Michalet, X. and Weiss, S. (2009) 'Suppression of Quantum Dot Blinking in DTT-Doped Polymer Films<sup>†</sup>', *The Journal of Physical Chemistry C*, 113(27), 11541-11545.
- Arndt, K. and Fink, G. R. (1986) 'GCN4 protein, a positive transcription factor in yeast, binds general control promoters at all 5'TGACTC 3'sequences', *Proceedings of the National Academy of Sciences*, 83(22), 8516-8520.
- Arndt, K. M., Pelletier, J. N., MuÈller, K. M., Alber, T., Michnick, S. W. and PluÈckthun, A. (2000) 'A heterodimeric coiled-coil peptide pair selected in vivo from a designed library-versus-library ensemble', *Journal of molecular biology*, 295(3), 627-639.
- Bacia, K. and Schwille, P. (2007) 'Fluorescence correlation spectroscopy', *Lipid rafts*, 73-84.
- Bénichou, O., Chevalier, C., Meyer, B. and Voituriez, R. (2011) 'Facilitated diffusion of proteins on chromatin', *Physical review letters*, 106(3), 038102.
- Berg, H. C. (1993) *Random Walks in Biology*, United States of America: Princeton University Press.
- Berg, O. G., Winter, R. B. and Von Hippel, P. H. (1981) 'Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory', *Biochemistry*, 20(24), 6929-6948.
- Berger, S. L. (2007) 'The complex language of chromatin regulation during transcription', *Nature*, 447(7143), 407-412.

- Bohmann, D., Bos, T. J., Admon, A., Nishimura, T., Vogt, P. K. and Tjian, R. (1987) 'Human proto-oncogene c-jun encodes a DNA binding protein with structural and functional properties of transcription factor AP-1', *Science*, 238(4832), 1386-1392.
- Bradham, C. and McClay, D. R. (2006) 'p38 MAPK in development and cancer', *Cell cycle*, 5(8), 824-828.
- Brown, T. A. (2007) *Genomes 3*, London: Garland Science.
- Bruchez, M., Moronne, M., Gin, P., Weiss, S. and Alivisatos, A. P. (1998) 'Semiconductor nanocrystals as fluorescent biological labels', *Science*, 281(5385), 2013.
- Burkhard, P., Meier, M. and Lustig, A. (2000) 'Design of a minimal protein oligomerization domain by a structural approach', *Protein Science*, 9(12), 2294-2301.
- Cairns, R. A., Harris, I. S. and Mak, T. W. (2011) 'Regulation of cancer cell metabolism', *Nature Reviews Cancer*, 11(2), 85-95.
- Campbell, N. A., Reece, J. E. (2005) *Biology, Biology*, London: Pearson.
- Carmeliet, P. and Jain, R. K. (2000) 'Angiogenesis in cancer and other diseases', *Nature*, 407(6801), 249-257.
- Champe, P. C., Harvey, R. A. and Ferrier, D. R. (2005) *Biochemistry*, 3rd ed., London: Lippincott Williams & Wilkins.
- Chan, W. C. and Nie, S. (1998) 'Quantum dot bioconjugates for ultrasensitive nonisotopic detection', *Science*, 281(5385), 2016-2018.
- Chen, L., Glover, J. N. M., Hogan, P. G., Rao, A. and Harrison, S. C. (1998) 'Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA', *Nature*, 392(6671), 42-48.
- Chen, L., Oakley, M. G., Glover, J. M., Jain, J., Dervan, P. B., Hogan, P. G., Rao, A. and Verdine, G. L. (1995) 'Only one of the two DNA-bound orientations of AP-1 found in solution cooperates with NFATp', *Current Biology*, 5(8), 882-889.
- Chial, H. (2008) 'Proto-oncogenes to oncogenes to cancer', *Nature Education*, 1(1), 33.
- Chinenov, Y. and Kerppola, T. K. (2001) 'Close encounters of many kinds: Fos-Jun interactions that mediate transcription regulatory specificity', *Oncogene*, 20(19), 2438-2452.

- Chiu, R., Boyle, W. J., Meek, J., Smeal, T., Hunter, T. and Karin, M. (1988) 'The c-Fos protein interacts with c-Jun/AP-1 to stimulate transcription of AP-1 responsive genes', *Cell*, 54(4), 541.
- Chow, A. (2010) 'Cell cycle control by oncogenes and tumor suppressors: driving the transformation of normal cells into cancerous cells', *Nature Education*, 3(7).
- Collado, M. and Serrano, M. (2010) 'Senescence in tumours: evidence from mice and humans', *Nature Reviews Cancer*, 10(1), 51-57.
- Crick, F. H. (1953) 'The packing of-helices: simple coiled-coils', *Acta crystallographica*, 6(8-9), 689-697.
- Das, M., Garlick, D. S., Greiner, D. L. and Davis, R. J. (2011) 'The role of JNK in the development of hepatocellular carcinoma', *Genes & development*, 25(6), 634-645.
- Dey, B., Thukral, S., Krishnan, S., Chakrobarty, M., Gupta, S., Manghani, C. and Rani, V. (2012) 'DNA–protein interactions: methods for detection and analysis', *Molecular and cellular biochemistry*, 365(1-2), 279-299.
- Downward, J. (2003) 'Targeting RAS signalling pathways in cancer therapy', *Nature Reviews Cancer*, 3(1), 11-22.
- Duffy, M., McGowan, P. and Gallagher, W. (2008) 'Cancer invasion and metastasis: changing views', *The Journal of pathology*, 214(3), 283-293.
- Eckhoff, K., Flurschütz, R., Trillsch, F., Mahner, S., Jänicke, F. and Milde-Langosch, K. (2013) 'The prognostic significance of Jun transcription factors in ovarian cancer', *Journal of cancer research and clinical oncology*, 139(10), 1673-1680.
- Eferl, R. and Wagner, E. F. (2003) 'AP-1: a double-edged sword in tumorigenesis', *Nature Reviews Cancer*, 3(11), 859-868.
- Elf, J., Li, G. W. and Xie, X. S. (2007) 'Probing transcription factor dynamics at the single-molecule level in a living cell', *Science*, 316(5828), 1191.
- Ellenberger, T. E., Brandl, C. J., Struhl, K. and Harrison, S. C. (1992) 'The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted [alpha] helices: crystal structure of the protein-DNA complex', *Cell*, 71(7), 1223-1237.
- Evan, G. I. and Vousden, K. H. (2001) 'Proliferation, cell cycle and apoptosis in cancer', *Nature*, 411(6835), 342-348.

- Ewig, R. A. and Kohn, K. W. (1977) 'DNA damage and repair in mouse leukemia L1210 cells treated with nitrogen mustard, 1, 3-bis (2-chloroethyl)-1-nitrosourea, and other nitrosoureas', *Cancer research*, 37(7 Part 1), 2114-2122.
- Fausto, N. (1998) *Mouse liver tumorigenesis: models, mechanisms, and relevance to human disease*, translated by 243-252.
- Fazio, T., Visnapuu, M. L., Wind, S. and Greene, E. C. (2008) 'DNA curtains and nanoscale curtain rods: high-throughput tools for single molecule imaging', *Langmuir*, 24(18), 10524-10531.
- Fish, K. N. (2009) 'Total internal reflection fluorescence (TIRF) microscopy', *Current protocols in cytometry*, 12.18. 1-12.18. 13.
- Frémin, C. and Meloche, S. (2010) 'From basic research to clinical development of MEK1/2 inhibitors for cancer therapy', *J Hematol Oncol*, 3(8), 1-11.
- Freyer, M. W. and Lewis, E. A. (2008) 'Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions', *Methods in cell biology*, 84, 79-113.
- Funatsu, T., Harada, Y., Tokunaga, M., Saito, K. and Yanagida, T. (1995) 'Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution', *Nature*, 374(6522), 555-559.
- Geertsema, H. J., Schulte, A. C., Spenkelink, L. M., McGrath, W. J., Morrone, S. R., Sohn, J., Mangel, W. F., Robinson, A. and van Oijen, A. M. (2015) 'Single-molecule imaging at high fluorophore concentrations by local activation of dye', *Biophysical journal*, 108(4), 949-956.
- Glover, J. N. M. and Harrison, S. C. (1995) 'Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA'.
- Gorman, J. and Greene, E. C. (2008) 'Visualizing one-dimensional diffusion of proteins along DNA', *Nature structural & molecular biology*, 15(8), 768-774.
- Gorman, J., Plys, A. J., Visnapuu, M. L., Alani, E. and Greene, E. C. (2010) 'Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice', *Nature structural & molecular biology*, 17(8), 932-938.
- Gorski, S. A., Dundr, M. and Misteli, T. (2006) 'The road much traveled: trafficking in the cell nucleus', *Current opinion in cell biology*, 18(3), 284-290.
- Granéli, A., Yeykal, C. C., Prasad, T. K. and Greene, E. C. (2006) 'Organized arrays of individual DNA molecules tethered to supported lipid bilayers', *Langmuir*, 22(1), 292-299.

- Haber, C. and Wirtz, D. (2000) 'Magnetic tweezers for DNA micromanipulation', *Review of Scientific Instruments*, 71(12), 4561-4570.
- Hagemann, U. B., Mason, J. M., Muller, K. M. and Arndt, K. M. (2008) 'Selectional and mutational scope of peptides sequestering the Jun-Fos coiled-coil domain', *Journal of molecular biology*, 381(1), 73-88.
- Hager, G. L., McNally, J. G. and Misteli, T. (2009) 'Transcription dynamics', *Molecular cell*, 35(6), 741-753.
- Hahn, S. (2004) 'Structure and mechanism of the RNA polymerase II transcription machinery', *Nature structural & molecular biology*, 11(5), 394.
- Hai, T. and Curran, T. (1991) 'Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity', *Proceedings of the National Academy of Sciences*, 88(9), 3720.
- Halazonetis, T. D., Georgopoulos, K., Greenberg, M. E. and Leder, P. (1988) 'c-Jun dimerizes with itself and with c-Fos, forming complexes of different DNA binding affinities', *Cell*, 55(5), 917-924.
- Halford, S. E. and Marko, J. F. (2004) 'How do site-specific DNA-binding proteins find their targets?', *Nucleic acids research*, 32(10), 3040-3052.
- Hamacher, R., Bakiri, L. and Wagner, E. (2013) 'THE FUNCTION OF Fos AND Fos~ Jun DIMERS IN LIVER CANCER DEVELOPMENT', *Journal of Hepatology*, 58, S25-S44.
- Han, S., Kim, C. and Kwon, D. (1997) 'Thermal/oxidative degradation and stabilization of polyethylene glycol', *Polymer*, 38(2), 317-323.
- Hanahan, D. and Weinberg, R. A. (2011) 'Hallmarks of cancer: the next generation', *Cell*, 144(5), 646-674.
- Håndstad, T., Rye, M. B., Drabløs, F. and Sætrum, P. (2011) 'A ChIP-Seq benchmark shows that sequence conservation mainly improves detection of strong transcription factor binding sites', *PLoS One*, 6(4), e18430.
- Healthwise (2008) 'Basal and squamous cell carcinoma', [online], available: <http://www.revolutionhealth.com/conditions/cancer/skin-cancer-nonmelanoma/diagnose/squamous-basal-carcinoma-images> [accessed 12/03/2012].
- Heffler, M. A. and Walters, R. D. (2012) 'Using electrophoretic mobility shift assays to measure equilibrium dissociation constants: GAL4-p53 binding DNA as a model system', *Biochemistry and Molecular Biology Education*, 40(6), 383-387.

- Hellman, L. M. and Fried, M. G. (2007) 'Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions', *Nature protocols*, 2(8), 1849-1861.
- Hilario, J. and Kowalczykowski, S. C. (2010) 'Visualizing protein-DNA interactions at the single-molecule level', *Current opinion in chemical biology*, 14(1), 15-22.
- Hinshaw, J. L., Laeseke, P. F. and Lee, F. T. (2012) 'The Role of Image-guided Tumor Ablation in the Management of Liver Cancer', [online], available: <http://www.cancernews.com/data/Article/504.asp> [accessed 12/03/2012].
- Hughes, C. D., Wang, H., Ghodke, H., Simons, M., Towheed, A., Peng, Y., Van Houten, B. and Kad, N. M. (2013) 'Real-time single-molecule imaging reveals a direct interaction between UvrC and UvrB on DNA tightropes', *Nucleic acids research*, gkt177.
- Ishijima, A. and Yanagida, T. (2001) 'Single molecule nanobioscience', *Trends in biochemical sciences*, 26(7), 438-444.
- Ito, T., Yamauchi, M., Nishina, M., Yamamichi, N., Mizutani, T., Ui, M., Murakami, M. and Iba, H. (2001) 'Identification of SWI· SNF complex subunit BAF60a as a determinant of the transactivation potential of Fos/Jun dimers', *Journal of Biological Chemistry*, 276(4), 2852.
- Jaiswal, J. K., Mattoussi, H., Mauro, J. M. and Simon, S. M. (2002) 'Long-term multiple color imaging of live cells using quantum dot bioconjugates', *Nature biotechnology*, 21(1), 47-51.
- Jochum, W., Passegué, E. and Wagner, E. F. (2001) 'AP-1 in mouse development and tumorigenesis', *Oncogene*, 20(19), 2401-2412.
- Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C. and Ha, T. (2008) 'Advances in single-molecule fluorescence methods for molecular biology', *Annu. Rev. Biochem.*, 77, 51-76.
- Juven-Gershon, T. and Kadonaga, J. T. (2010) 'Regulation of gene expression via the core promoter and the basal transcriptional machinery', *Developmental biology*, 339(2), 225-229.
- Kad, N. M., Wang, H., Kennedy, G. G., Warshaw, D. M. and Van Houten, B. (2010) 'Collaborative dynamic DNA scanning by nucleotide excision repair proteins investigated by single-molecule imaging of quantum-dot-labeled proteins', *Molecular cell*, 37(5), 702-713.
- Kalodimos, C. G., Biris, N., Bonvin, A. M., Levandoski, M. M., Guennuegues, M., Boelens, R. and Kaptein, R. (2004) 'Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes', *Science*, 305(5682), 386-389.
- Kampmann, M. (2004) 'Obstacle bypass in protein motion along DNA by two-dimensional rather than one-dimensional sliding', *Journal of Biological Chemistry*, 279(37), 38715.

- Kampmann, M. (2005) 'Facilitated diffusion in chromatin lattices: mechanistic diversity and regulatory potential', *Molecular microbiology*, 57(4), 889-899.
- Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J. F. and Wyczalkowski, M. A. (2013) 'Mutational landscape and significance across 12 major cancer types', *Nature*, 502(7471), 333-339.
- Karin, M., Liu, Z. and Zandi, E. (1997) 'AP-1 function and regulation', *Current opinion in cell biology*, 9(2), 240-246.
- Keller, S., Vargas, C., Zhao, H., Piszczek, G., Brautigam, C. A. and Schuck, P. (2012) 'High-precision isothermal titration calorimetry with automated peak-shape analysis', *Analytical Chemistry*, 84(11), 5066-5073.
- Kharchenko, P. V., Tolstorukov, M. Y. and Park, P. J. (2008) 'Design and analysis of ChIP-seq experiments for DNA-binding proteins', *Nat Biotechnol*, 26(12), 1351-1359.
- Khatter, H., Myasnikov, A. G., Mastio, L., Billas, I. M., Birck, C., Stella, S. and Klaholz, B. P. (2014) 'Purification, characterization and crystallization of the human 80S ribosome', *Nucleic acids research*, 42(6), e49-e49.
- Klausner, R. D. (2002) 'The fabric of cancer cell biology-Weaving together the strands', *Cancer Cell*, 1(1), 3.
- Kohn, W. D., Mant, C. T. and Hodges, R. S. (1997) ' $\alpha$ -Helical protein assembly motifs', *Journal of Biological Chemistry*, 272(5), 2583.
- Kortemme, T. and Baker, D. (2004) 'Computational design of protein-protein interactions', *Current opinion in chemical biology*, 8(1), 91-97.
- Kouzarides, T. (2007) 'Chromatin modifications and their function', *Cell*, 128(4), 693-705.
- Krylov, D., Mikhailenko, I. and Vinson, C. (1994) 'A thermodynamic scale for leucine zipper stability and dimerization specificity: e and g interhelical interactions', *The EMBO Journal*, 13(12), 2849.
- Landschulz, W. H., Johnson, P. F. and McKnight, S. L. (1988) 'The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins', *Science*, 240(4860), 1759.
- Latchman, D. (1995) *Gene Regulation 2nd edition*, London: Chapman.

- Lauer, S. A. and Nolan, J. P. (2002) 'Development and characterization of Ni-NTA-bearing microspheres', *Cytometry*, 48(3), 136-145.
- Leaner, V. D., Kinoshita, I. and Birrer, M. J. (2003) 'AP-1 complexes containing cJun and JunB cause cellular transformation of Rat1a fibroblasts and share transcriptional targets', *Oncogene*, 22(36), 5619-5629.
- Lei, G. and MacDonald, R. C. (2008) 'Effects on interactions of oppositely charged phospholipid vesicles of covalent attachment of polyethylene glycol oligomers to their surfaces: Adhesion, Hemifusion, Full Fusion and "Endocytosis"', *Journal of Membrane Biology*, 221(2), 97-106.
- Lemon, B. and Tjian, R. (2000) 'Orchestrated response: a symphony of transcription factors for gene control', *Genes & development*, 14(20), 2551.
- Leonard, D. A., Rajaram, N. and Kerppola, T. K. (1997) 'Structural basis of DNA bending and oriented heterodimer binding by the basic leucine zipper domains of Fos and Jun', *Proceedings of the National Academy of Sciences*, 94(10), 4913-4918.
- Leppä, S. and Bohmann, D. (1999) 'Diverse functions of JNK signaling and c-Jun in stress response and apoptosis', *Oncogene*, 18(45).
- Leutwyler, W. K., Bürgi, S. L. and Burgl, H. (1996) 'Semiconductor clusters, nanocrystals, and quantum dots', *Science*, 271(5251), 933-937.
- Li, G. W., Berg, O. G. and Elf, J. (2009) 'Effects of macromolecular crowding and DNA looping on gene regulation kinetics', *Nature Physics*, 5(4), 294-297.
- Liseth, K., Foss Abrahamsen, J., Bjørsvik, S., Grøtnebø, K. and Bruserud, Ø. (2005) 'The viability of cryopreserved PBPC depends on the DMSO concentration and the concentration of nucleated cells in the graft', *Cytotherapy*, 7(4), 328-333.
- Lu, C., Shen, Q., DuPré, E., Kim, H., Hilsenbeck, S. and Brown, P. H. (2005) 'cFos is critical for MCF-7 breast cancer cell growth', *Oncogene*, 24(43), 6516-6524.
- Lupas, A. N. and Gruber, M. (2005) 'The Structure of [alpha]-Helical Coiled Coils', *Advances in protein chemistry*, 70, 37-38.
- Maniatis, T., Jeffrey, A. and Van deSande, H. (1975) 'Chain length determination of small double- and single-stranded DNA molecules by polyacrylamide gel electrophoresis', *Biochemistry*, 14(17), 3787-3794.
- Mason, J. M. (2009) 'Electrostatic contacts in the activator protein-1 coiled coil enhance stability predominantly by decreasing the unfolding rate', *FEBS Journal*, 276(24), 7305-7318.

- Mason, J. M. and Arndt, K. M. (2004) 'Coiled coil domains: stability, specificity, and biological implications', *ChemBioChem*, 5(2), 170-176.
- Mason, J. M., Hagemann, U. B. and Arndt, K. M. (2007a) 'Improved Stability of the Jun-Fos Activator Protein-1 Coiled Coil Motif', *Journal of Biological Chemistry*, 282(32), 23015.
- Mason, J. M., Müller, K. M. and Arndt, K. M. (2007b) 'Positive aspects of negative design: simultaneous selection of specificity and interaction stability', *Biochemistry*, 46(16), 4804-4814.
- Mason, J. M., Schmitz, M. A., Müller, K. M. and Arndt, K. M. (2006) 'Semirational design of Jun-Fos coiled coils with increased affinity: Universal implications for leucine zipper prediction and design', *Proceedings of the National Academy of Sciences*, 103(24), 8989-8994.
- Mattheyses, A. L., Simon, S. M. and Rappoport, J. Z. (2010) 'Imaging with total internal reflection fluorescence microscopy for the cell biologist', *Journal of cell science*, 123(21), 3621-3628.
- McFadden, J. (2002) 'Osteosarcoma of the knee', [online], available: <http://www.mypacs.net/cases/OSTEOSARCOMA-OF-THE-KNEE-28947.html> [accessed 12/02/2012].
- Mechta-Grigoriou, F., Gerald, D. and Yaniv, M. (2001) 'The mammalian Jun proteins: redundancy and specificity', *Oncogene*, 20(19), 2378-2389.
- Medintz, I. L., Uyeda, H. T., Goldman, E. R. and Mattoussi, H. (2005) 'Quantum dot bioconjugates for imaging, labelling and sensing', *Nature materials*, 4(6), 435-446.
- Milev, S. (2013) 'Isothermal titration calorimetry: Principles and experimental design', *General Electric*, 9.
- Mirny, L., Slutsky, M., Wunderlich, Z., Tafvizi, A., Leith, J. and Kosmrlj, A. (2009a) 'How a protein searches for its site on DNA: the mechanism of facilitated diffusion', *Journal of Physics A: Mathematical and Theoretical*, 42(43), 434013.
- Mirny, L., Slutsky, M., Wunderlich, Z., Tafvizi, A., Leith, J. and Kosmrlj, A. (2009b) 'How a protein searches for its site on DNA: the mechanism of facilitated diffusion', *Journal of Physics A: Mathematical and Theoretical*, 42, 434013.
- Moras, D. and Gronemeyer, H. (1998) 'The nuclear receptor ligand-binding domain: structure and function', *Current opinion in cell biology*, 10(3), 384-391.

- Muller, P. A., Caswell, P. T., Doyle, B., Iwanicki, M. P., Tan, E. H., Karim, S., Lukashchuk, N., Gillespie, D. A., Ludwig, R. L. and Gosselin, P. (2009) 'Mutant p53 drives invasion by promoting integrin recycling', *Cell*, 139(7), 1327-1341.
- Muller, P. A., Vousden, K. H. and Norman, J. C. (2011) 'p53 and its mutants in tumor cell migration and invasion', *The Journal of cell biology*, 192(2), 209-218.
- Nakabeppu, Y. and Nathans, D. (1989) 'The basic region of Fos mediates specific DNA binding', *The EMBO Journal*, 8(12), 3833.
- Nakabeppu, Y., Ryder, K. and Nathans, D. (1988) 'DNA binding activities of three murine Jun proteins: stimulation by Fos', *Cell*, 55(5), 907-915.
- Narlikar, G. J., Fan, H.-Y. and Kingston, R. E. (2002) 'Cooperation between complexes that regulate chromatin structure and transcription', *Cell*, 108(4), 475-487.
- Neher, E. and Sakmann, B. (1976) 'Noise analysis of drug induced voltage clamp currents in denervated frog muscle fibres', *The Journal of physiology*, 258(3), 705.
- Newman, J. R. S. and Keating, A. E. (2003) 'Comprehensive identification of human bZIP interactions with coiled-coil arrays', *Science*, 300(5628), 2097.
- Northwestern.University (2007) 'Oligo Calc: Oligo Properties Calculator', [online], available: <http://www.basic.northwestern.edu/biotools/oligocalc.html> [accessed 2/2/2013].
- O'Neil, K. T., Hoess, R. H. and DeGrado, W. F. (1990) 'Design of DNA-binding peptides based on the leucine zipper motif', *Science*, 249(4970), 774-778.
- Olive, M., Krylov, D., Echlin, D. R., Gardner, K., Taparowsky, E. and Vinson, C. (1997) 'A dominant negative to activation protein-1 (AP1) that abolishes DNA binding and inhibits oncogenesis', *Journal of Biological Chemistry*, 272(30), 18586.
- Oliviero, S., Robinson, G., Struhl, K. and Spiegelman, B. (1992) 'Yeast GCN4 as a probe for oncogenesis by AP-1 transcription factors: transcriptional activation through AP-1 sites is not sufficient for cellular transformation', *Genes & development*, 6(9), 1799-1809.
- Orchard, K. and May, G. E. (1993) 'An EMSA-based method for determining the molecular weight of a protein--DNA complex', *Nucleic acids research*, 21(14), 3335.
- Pál, C., Papp, B. and Lercher, M. J. (2006) 'An integrated view of protein evolution', *Nature Reviews Genetics*, 7(5), 337-348.

- Palmer, B. R. and Marinus, M. G. (1994) 'The dam and dcm strains of *Escherichia coli*—a review', *Gene*, 143(1), 1-12.
- Palmiter, R. D. (1974) 'Magnesium precipitation of ribonucleoprotein complexes. Expedient techniques for the isolation of undegraded polysomes and messenger ribonucleic acid', *Biochemistry*, 13(17), 3606-3615.
- Pierce, D. W., Hom-Booher, N. and Vale, R. D. (1997) 'Imaging individual green fluorescent proteins', *Nature*, 388(6640), 338.
- PolySciencesLtd. (2011) 'Microspheres & Particles', [online], available: <http://www.polysciences.com/default/catalog-products/microspheres-particles> [accessed 11/10/2011].
- Porte, D., Oertel-Buchheit, P., Granger-Schnarr, M. and Schnarr, M. (1995) 'Fos leucine zipper variants with increased association capacity', *Journal of Biological Chemistry*, 270(39), 22721-22730.
- Powell, L. M., Dryden, D. T. and Murray, N. E. (1998) 'Sequence-specific DNA binding by EcoKI, a type IA DNA restriction enzyme', *Journal of molecular biology*, 283(5), 963-976.
- Qiagen (2015) 'QIAquick Gel Extraction Kit', [online], available: <https://www.qiagen.com/gb/shop/sample-technologies/dna-sample-technologies/dna-cleanup/qiaquick-gel-extraction-kit/> [accessed ]
- Raha, D., Wang, Z., Moqtaderi, Z., Wu, L., Zhong, G., Gerstein, M., Struhl, K. and Snyder, M. (2010) 'Close association of RNA polymerase II and many transcription factors with Pol III genes', *Proceedings of the National Academy of Sciences*, 107(8), 3639-3644.
- Rajaram, N. and Kerppola, T. K. (1997) 'DNA bending by Fos–Jun and the orientation of heterodimer binding depend on the sequence of the AP-1 site', *The EMBO Journal*, 16(10), 2917-2925.
- Reichert, W. and Truskey, G. (1990) 'Total internal reflection fluorescence (TIRF) microscopy. I. Modelling cell contact region fluorescence', *Journal of cell science*, 96(2), 219-230.
- Reinhardt, C. G. and Krugh, T. R. (1978) 'A comparative study of ethidium bromide complexes with dinucleotides and DNA: direct evidence for intercalation and nucleic acid sequence preferences', *Biochemistry*, 17(23), 4845-4854.
- Remy, I., Campbell-Valois, F. and Michnick, S. W. (2007) 'Detection of protein–protein interactions using a simple survival protein-fragment complementation assay based on the enzyme dihydrofolate reductase', *Nature protocols*, 2(9), 2120-2125.
- Ries, J. and Schwille, P. (2012) 'Fluorescence correlation spectroscopy', *Bioessays*, 34(5), 361-368.

- Riggs, A. D., Bourgeois, S. and Cohn, M. (1970) 'The lac repressor-operator interaction\* 1,\* 2:: III. Kinetic studies', *Journal of molecular biology*, 53(3), 401-417.
- Ringquist, S. and Smith, C. L. (1992) 'The Escherichia coli chromosome contains specific, unmethylated dam and dcm sites', *Proceedings of the National Academy of Sciences*, 89(10), 4539-4543.
- Rio, D. C. (2012) 'Filter-binding assay for analysis of RNA–protein interactions', *Cold Spring Harbor Protocols*, 2012(10), pdb. prot071449.
- Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B. and Mann, R. S. (2010) 'Origins of specificity in protein-DNA recognition', *Annual review of biochemistry*, 79, 233.
- Rossi, R., Montecucco, A., Ciarrocchi, G. and Biamonti, G. (1997) 'Functional characterization of the T4 DNA ligase: a new insight into the mechanism of action', *Nucleic acids research*, 25(11), 2106-2113.
- Sambrook, J., Russell, D. W. (2001) *Molecular Cloning. A laboratory manual*, 3rd ed., New York: Coldspring Harbour Laboratory.
- Santarpia, L., Lippman, S. M. and El-Naggar, A. K. (2012) 'Targeting the MAPK-RAS-RAF signaling pathway in cancer therapy', *Expert opinion on therapeutic targets*, 16(1), 103-119.
- Saxton, M. J. (1997) 'Single-particle tracking: the distribution of diffusion coefficients', *Biophysical journal*, 72(4), 1744.
- Schuermann, M., Neuberg, M., Hunter, J. B., Jenuwein, T., Ryseck, R. P., Bravo, R. and Müller, R. (1989) 'The leucine repeat motif in Fos protein mediates complex formation with Jun/AP-1 and is required for transformation', *Cell*, 56(3), 507.
- Seldeen, K. L., McDonald, C. B., Deegan, B. J. and Farooq, A. (2009) 'Single Nucleotide Variants of the TGACTCA Motif Modulate Energetics and Orientation of Binding of the Jun-Fos Heterodimeric Transcription Factor<sup>†</sup>', *Biochemistry*, 48(9), 1975-1983.
- Shaulian, E. and Karin, M. (2001) 'AP-1 in cell proliferation and survival', *Oncogene*, 20(19), 2390-2400.
- Shay, J. W. and Wright, W. E. (2010) 'Telomeres and telomerase in normal and cancer stem cells', *FEBS letters*, 584(17), 3819-3825.
- Shindo, H., Torigoe, H. and Sarai, A. (1993) 'Thermodynamic and kinetic studies of DNA triplex formation of an oligohomopyrimidine and a matched duplex by filter binding assay', *Biochemistry*, 32(34), 8963-8969.

- Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M. J., Davie, J. R. and Peterson, C. L. (2006) 'Histone H4-K16 acetylation controls chromatin structure and protein interactions', *Science*, 311(5762), 844-847.
- Shyy, J., Lin, M.-C., Han, J., Lu, Y., Petrine, M. and Chien, S. (1995) 'The cis-acting phorbol ester" 12-O-tetradecanoylphorbol 13-acetate"-responsive element is involved in shear stress-induced monocyte chemotactic protein 1 gene expression', *Proceedings of the National Academy of Sciences*, 92(17), 8069-8073.
- Sigma-Aldrich (2015) 'Custom DNA and RNA Oligos', [online], available: <http://www.sigmaaldrich.com/catalog/product/sigma/oligo?lang=en&region=GB> [accessed
- Simonson, M. S., Jones, J. and Dunn, M. (1992) 'Differential regulation of fos and jun gene expression and AP-1 cis-element activity by endothelin isopeptides. Possible implications for mitogenic signaling by endothelin', *Journal of Biological Chemistry*, 267(12), 8643.
- Sims, R. J., Nishioka, K. and Reinberg, D. (2003) 'Histone lysine methylation: a signature for chromatin function', *TRENDS in Genetics*, 19(11), 629-639.
- Smith, D. R., Pendry, J. B. and Wiltshire, M. C. (2004) 'Metamaterials and negative refractive index', *Science*, 305(5685), 788-792.
- SnapGene (2013) [online], available: <http://www.snapgene.com/> [accessed 3/1/2013].
- Song, D., Graham, T. G. and Loparo, J. J. (2016) 'A general approach to visualize protein binding and DNA conformation without protein labelling', *Nature Communications*, 7.
- Suga, M. and Hatakeyama, T. (2003) 'High-efficiency electroporation by freezing intact yeast cells with addition of calcium', *Current genetics*, 43(3), 206-211.
- Szalóki, N., Krieger, J. W., Komáromi, I., Tóth, K. and Vámosi, G. (2015) 'Evidence for Homodimerization of the c-Fos Transcription Factor in Live Cells Revealed by Fluorescence Microscopy and Computer Modeling', *Molecular and cellular biology*, 35(21), 3785-3798.
- Talanian, R. V., McKnight, C. J. and Kim, P. S. (1990) 'Sequence-specific DNA binding by a short peptide dimer', *Science*, 249(4970), 769-771.
- ThermoFisherScientific (2015) 'BamHI', [online], available: <https://www.thermofisher.com/order/catalog/product/ER0051?ICID=search-product> [accessed
- Thompson, R. E., Larson, D. R. and Webb, W. W. (2002) 'Precise nanometer localization analysis for individual fluorescent probes', *Biophysical journal*, 82(5), 2775-2783.

- Tonini, T., Rossi, F. and Claudio, P. P. (2003) 'Molecular basis of angiogenesis and cancer', *Oncogene*, 22(42), 6549-6556.
- Tsen, S.-D., Fang, S.-S., Chen, M.-J., Chien, J.-Y., Lee, C.-C. and Tsen, D. H.-L. (2002) 'Natural plasmid transformation in *Escherichia coli*', *Journal of biomedical science*, 9(3), 246-252.
- Turner, R. (1989) 'Leucine repeats and an adjacent DNA binding', *Science*, 2494701(1689), 243.
- van Dam, H. and Castellazzi, M. (2001) 'Distinct roles of Jun: Fos and Jun: ATF dimers in oncogenesis', *Oncogene*, 20(19), 2453-2464.
- van Oijen, A. M. (2011) 'Single-molecule approaches to characterizing kinetics of biomolecular interactions', *Current opinion in biotechnology*, 22(1), 75-80.
- Viadiu, H. and Aggarwal, A. K. (2000) 'Structure of BamHI Bound to Nonspecific DNA: A Model for DNA Sliding', *Molecular cell*, 5(5), 889-895.
- Vinson, C., Hai, T. and Boyd, S. (1993) 'Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design', *Genes & development*, 7(6), 1047-1058.
- Von Hippel, E. (1994) "'Sticky information" and the locus of problem solving: Implications for innovation', *Management science*, 429-439.
- Von Hippel, P. H. (2004) 'Completing the view of transcriptional regulation', *Science*, 305(5682), 350.
- von Hippel, P. H. and Berg, O. (1989) 'Facilitated target location in biological systems', *Journal of Biological Chemistry*, 264(2), 675.
- Wagner, E. (2002) 'Functions of AP1 (Fos/Jun) in bone development', *Annals of the rheumatic diseases*, 61(suppl 2), ii40-ii42.
- Wagner, E. F. (2001) 'AP-1-Introductory remarks', *Oncogene*, 20(19), 2334-2335.
- Wagner, E. F. and Eferl, R. (2005) 'Fos/AP-1 proteins in bone and the immune system', *Immunological reviews*, 208(1), 126-140.
- Walter, N. G., Huang, C. Y., Manzo, A. J. and Sobhy, M. A. (2008) 'Do-it-yourself guide: how to use the modern single-molecule toolkit', *Nature methods*, 5(6), 475-489.

- Walters, R. D., McSwiggen, D. T., Goodrich, J. A. and Kugel, J. F. (2014) 'Selection and Characterization of a DNA Aptamer That Can Discriminate between cJun/cJun and cJun/cFos'.
- Wang, H., Tessmer, I., Croteau, D. L., Erie, D. A. and Van Houten, B. (2008) 'Functional characterization and atomic force microscopy of a DNA repair protein conjugated to a quantum dot', *Nano letters*, 8(6), 1631-1637.
- Wang, Y., Austin, R. H. and Cox, E. C. (2006) 'Single molecule measurements of repressor protein 1D diffusion on DNA', *Physical review letters*, 97(4), 48302.
- Wang, Z. Q., Liang, J., Schellander, K., Wagner, E. F. and Grigoriadis, A. E. (1995) 'c-fos-induced osteosarcoma formation in transgenic mice: Cooperativity with c-jun and the role of endogenous c-fos', *Cancer research*, 55(24), 6244.
- Weinberg, R. A. (2007) *The Biology of Cancer*, USA: Garland Science.
- Whitmarsh, A. J. (2007) 'Regulation of gene transcription by mitogen-activated protein kinase signaling pathways', *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1773(8), 1285-1298.
- Wilkinson, A., Day, J. and Bowater, R. (2001) 'Bacterial DNA ligases', *Molecular microbiology*, 40(6), 1241-1248.
- Winter, R. B., Berg, O. G. and Von Hippel, P. H. (1981) 'Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor-operator interaction: kinetic measurements and conclusions', *Biochemistry*, 20(24), 6961-6977.
- Woolfson, D. N. and Alber, T. (1995) 'Predicting oligomerization states of coiled coils', *Protein Science*, 4(8), 1596-1607.
- Worrall, J. A. and Mason, J. M. (2011) 'Thermodynamic analysis of Jun–Fos coiled coil peptide antagonists', *FEBS Journal*, 278(4), 663-672.
- Zenz, R., Eferl, R., Scheinecker, C., Redlich, K., Smolen, J., Schonhaler, H. B., Kenner, L., Tschachler, E. and Wagner, E. F. (2008) 'Activator protein 1 (Fos/Jun) functions in inflammatory bone and skin disease', *Arthritis Research and Therapy*, 10(1), 10.
- Zhang, G., Luo, X., Sumithran, E., Pua, V., Barnetson, R. S. C., Halliday, G. and Khachigian, L. (2006) 'Squamous cell carcinoma growth in mice and in culture is regulated by c-Jun and its control of matrix metalloproteinase-2 and-9 expression', *Oncogene*, 25(55), 7260-7266.
- Zhang, Y., Pu, X., Shi, M., Chen, L., Song, Y., Qian, L., Yuan, G., Zhang, H., Yu, M. and Hu, M. (2007) 'Critical role of c-Jun overexpression in liver metastasis of human breast cancer xenograft model', *BMC cancer*, 7(1), 145.

Zhang, Y., Takahashi, S., Tasaka, A., Yoshima, T., Ochi, H. and Chayama, K. (2013) 'Involvement of microRNA-224 in cell proliferation, migration, invasion, and anti-apoptosis in hepatocellular carcinoma', *Journal of Gastroenterology and Hepatology*, 28(3), 565-575.

Zimmerman, S. B. and Pfeiffer, B. H. (1983) 'Macromolecular crowding allows blunt-end ligation by DNA ligases from rat liver or *Escherichia coli*', *Proceedings of the National Academy of Sciences*, 80(19), 5852-5856.

Zlatanova, J. and Leuba, S. H. (2003) 'Magnetic tweezers: a sensitive tool to study DNA and chromatin at the single-molecule level', *Biochemistry and cell biology*, 81(3), 151-159.