PHOTOCONTROL OF

ARTIFICIAL

TRANSCRIPTION FACTORS

by

Seni Chanapai

DECLARATION

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ………………………………   Date …………………………

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed ……………………………….   Date …………………………

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.

Other sources are acknowledged by explicit references.

Signed ……………………………….   Date …………………………

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ……………………………….   Date …………………………

Acknowledgements

I would like to thank my supervisor, Prof. Rudolf Allemann for giving me the opportunity to join his research group and work on this project. I wish to express my gratitude to Dr Robert Mart for his help, patience and the proof reading of this thesis. Thanks to Dr Thomas Fricke, Dr Amang Li, Dr Alicja Antonczak, Dr Ania Piasecka, Andrew Wood, Alina Sikora, Stella Matthews, Dilruba Meah and especially Rima Trofimovaite for their friendship and support. I would also like to thank Dr Mahmoud Akhtar, Prof. Gerald Richter, Dr James Redman and Dr Simon Pope for their guidance and helpful discussions.

Finally I would like to thank my parents for their encouragement throughout these years and for believing in me.

Abstract

The design of a photoswitchable homeodomain artificial transcription factor (PATF), modelled on an engrailed homeodomain, for the purpose of controlling DNA binding affinity and controlling the transcription process in cells using light has been investigated. This study was conducted using a 3,3'-bis(sulfo)-4,4'bis(chloroacetamino)azobenzene crosslinker, alkylated between two cysteine residues with different spacings (*i*, *i*+4, *i*, *i*+7 and *i*, *i*+11) and either a rigid or flexible linker domain. In previous studies, basic leucine zipper transcription activators have been photocontrolled in living cells by incorporating a photoswitchable azobenzene crosslinker.

Circular dichroism spectroscopy showed the conformation of crosslinked PATF (XL-PATF) peptides (*i*, *i*+11 spacing) containing rigid and flexible linkers could be controlled reversibly by light. Fluorescence anisotropy experiments using labelled DNA confirmed the *in vitro* DNA binding affinity of PATF was considerably higher with the crosslinker in the *trans* (ground state) configuration than in the *cis* (photoexcited state) configuration. Further studies of peptides with *i*, *i*+4 and *i*, *i*+7 spacings with a semi-rigid and rigid linker domains showed increased binding affinity with the crosslinker in the *cis* configuration.

Initiation of transcription was investigated by an *in vitro* transcription assay to measure the ability of PATF molecules to moderate the production of RNA by irradiation with UV light. PATF molecules with *i*, *i*+11 spacings showed increased transcriptional activation with the crosslinker in the ground state configuration and *i*, *i*+4 and *i*, *i*+7 spacings resulted in increased transcription activation with the crosslinker in the excited state conformation. Control of 50% of transcriptional activity was achieved for *i, i*+11

spacings, and PATFs with a rigid linker domain were more effective switches than those with flexible linkers. Using i, $i+4$ and i, $i+7$ spacings in PATFs resulted in a lower degree of control but, as anticipated, transcriptional activation was increased after irradiation.

Table of contents

List of figures

List of tables

List of abbreviations

A   adenine

AD   activation domain

AdML   adenovirus major late promoter

AH   amphipathic α-helix

Ala (A)   alanine

AP2   activating protein 2

aPP   avian pancreatic polypeptide

Arg (R)   arginine

Asn (N)   asparagine

Asp (D)   aspartic acid

ATP   adenosine triphosphate

BSBCA   3,3'-bis-(sulfo)-4,4'-bis(chloroacetoamino)azobenzene

bHLH   basic helix-loop-helix

bHLH/z   basic helix-loop-helix zipper

BME   β-mercaptoethanol

BSA   bovine serum albumin

bZIP   basic leucine zipper

C   cytosine

cAMP   cyclic adenosine monophosphate

CAP   catabolite activator protein

CD   circular dichroism

CHAPS	3-[(3-cholamidopropyl)-dimethylammonio]propanesulfonate
CMV	cytomegalovirus immediate-early promoter
CTF/NF1	CAAT box-binding transcription factor/nuclear factor-1
Cys (C)	cysteine
dad-HDH-3	dark-adapted HDH-3
dad-XL-PATF	dark-adapted XL-PATF
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
dGTP	deoxyguanosine triphosphate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DBD	DNA binding domain
dNTP	deoxynucleoside triphosphate
dsDNA	double-stranded DNA
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
ESI	electrospray ionisation
FAM	fluoresceinamide
FynSH3	Fyn-tyrosine kinase
G	guanine
GCN4	general control of nitrogen and purine metabolism
Gdn-Cl	guanidinium chloride

Gln (Q)   glutamine

Glu (E)   glutamic acid

Gly (G)   glycine

HEPES   4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

His (H)   histidine

HLH   helix-loop-helix

HPLC   high performance liquid chromatography

HSE   heat shock element

HSF   heat shock factor

Hsp   heat shock protein

HTH   helix-turn-helix

Ile (I)   isoleucine

IPTG   isopropyl-β-D-thiogalactopyranoside

irrad-HDH-3   irradiated HDH-3

KSI   ketosteroid isomerase

K_D apparent dissociation constant

LB   lysogeny broth

Leu (L)   leucine

LBD   ligand binding domain

Luc   luciferase reporter

Lys (K)   lysine

MALDI-TOF   matrix assisted laser desorption ionisation-time of flight

MCK   muscle creatine enhancer

MEF2	myocyte enhancer factor 2
Met (M)	methionine
MOPS	3-(N-morpholino) propanesulfonic acid
mRNA	messenger RNA
MTS	methyl alkanethiosulfonate
MWCO	molecular weight cut off
MyoD	myogenic determination factor D
NMR	nuclear magnetic resonance
nNOS	neuronal nitric oxide synthase
NTP	nucleoside triphosphate
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PATF	photocontrollable homeodomain artificial transcription factor
PCR	polymerase chain reaction
Phe (F)	phenylalanine
PMSF	phenylmethanesulfonyl fluoride
PP	oligo-prolines
PPI	polyproline type I
PPII	polyproline type II
Pro (P)	proline
RHR	Rel homology region
RNA	ribonucleic acid
rP172	recombinant porcine amelogenin

rpm	revolutions per minute
rRNA	ribosomal RNA
RRE	Rev response element
RT	room temperature
SDS	sodium dodecyl sulfate
Ser (S)	serine
snRNA	small nuclear RNA
SRF	serum response factor
sp1	specificity Protein 1
ssDNA	single strand DNA
T	thymine
TAE	tris-acetate/EDTA
TB	terrific broth
TBE	tris-borate/EDTA
TBP	TATA-box binding protein
TCEP	tris-(carboxyethyl)phosphine
TE	tris-EDTA
TEMED	N, N, N', N'- tetramethylethylenediamine
TFA	trifluoroacetic acid
TFE	2, 2, 2-trifluoroethanol
TFII	transcription factor of RNA polymerase II
Thr (T)	threonine
Tris	tris(hydroxymethyl)-aminomethane

tRNA   transfer RNA

Trp (W)   tryptophan

Tyr (Y)   tyrosine

U   uracil

UV   ultraviolet

UV-Vis   ultraviolet-visible

Val (V)   valine

XL-PATF   crosslinked-PATF

Chapter 1
Introduction

Preface

Transcription and translation are the essential processes required for synthesis of specific proteins from DNA. As such they are fundamental to every living organism and impact on all biological processes, including growth, development and the ability of an organism to respond to environmental changes.

Transcription is the process, during which a particular segment of DNA is copied to create a messenger RNA (mRNA), and is used as a key point for regulation of the expression of genetic information. Proteins known as transcription factors, which bind to DNA and initiate the transcription process, regulate transcription. The mRNA molecules produced are then used as templates for protein synthesis during translation and the protein products derived from them can alter behaviour in particular cell types or in response to a particular stimulus.

The proliferation of light-controllable photonic compounds that can be incorporated into biomolecules provides an alternative strategy to rapidly and reversibly control transcription and subsequent protein signaling pathways. In conjunction with macromolecular function in biological systems, this technology may in the future be used to develop new artificial photoreceptors with medical and biosensor applications.

1.1 Transcription of Genes

Gene transcription is a process in which mRNA is synthesised from a specific DNA template by enzymes known as RNA polymerases. The process begins with RNA polymerases binding to a target DNA sequence known as a promoter close to the beginning of the gene and is controlled by proteins known as transcription factors, which bind to specific DNA sequences located at the promoter sites of the target genes. Motor proteins termed helicases (belonging to the translocase family of proteins) move directionally along nucleic acid stands unwinding the DNA, allowing the RNA polymerases to move along the gene copying the antisense (3' end) of the DNA resulting in the production of mRNA composed of complementary bases to those in the template DNA strand (**Figure 1.1**).
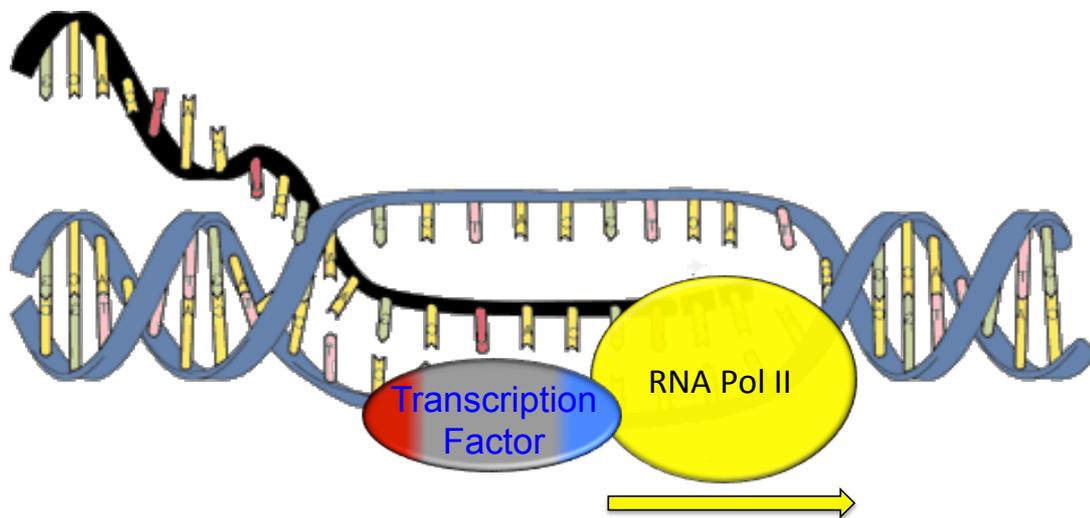
Figure 1. 1. *Basic transcription scheme with binding of transcription factor binding to target DNA and recruiting RNA polymerase enzyme.*

Transcription factors contain structural elements known as DNA-binding domains (DBD), which recognise double or single stranded DNA in either a DNA sequence-

specific or non-specific manner.[1] Transcription factors also contain a separate activation domain (AD). The AD dictates whether to activate or repress transcription by recruiting components associated with the transcription machinery or the repression machinery respectively. Regulation of gene expression is an important biological process because it allows cells to produce proteins appropriate for a given situation. Organisms as diverse as bacteria, animals, and plants respond to elevated temperatures by a rapid and transient increase in the synthesis of conserved polypeptides collectively referred to as heat shock proteins (Hsp). [2] When eukaryotic cells are exposed to elevated temperatures, heat shock factor (HSF) is activated and binds to heat shock element (HSE) promoters to up-regulate production of Hsp70 to help fold newly synthesised proteins and stabilise existing ones to overcome heat stress. [1] This is essential for eukaryotes, prokaryotes and viruses as the ability to control transcription increases the adaptability of an organism.

1.1.1 Prokaryotic Transcription

The prokaryotic genome is considerably smaller than that of eukaryotes. In prokaryotes, virtually all RNA is synthesised by a single DNA-dependent RNA polymerase. The only exception is the synthesis of short RNA primers complementary to ssDNA template strands by an enzyme known as primase. In prokaryotes there is no nucleus to separate the processes of transcription and translation, so they happen simultaneously in the cellular cytoplasm; translation of an mRNA strand often occurs even while downstream mRNA is still being synthesised. In bacteria, genes encoding enzymes of a particular metabolic pathway are usually grouped adjacent to one another in a cluster on

the chromosome, allowing all the genes in the group to be expressed in a coordinated fashion through transcription into a single polycistronic mRNA (a single RNA transcript that encodes more than one polypeptide). [3] This allows a single control point for an entire metabolic pathway.

In *E. coli*, the RNA polymerase holoenzyme is a complex multimeric protein (~450 kDa). It is made up of 4 main subunits, two α subunits, β, β' and interchangeable σ subunits ($α_2ββ'σ$). The largest subunit, β' (~155 kDa), is used for DNA binding while β (151 kDa) binds nucleoside triphosphate substrates and interacts with the σ subunits. A number of related proteins, known as σ-factors, can serve as σ-subunits. The different σ-factors allow the RNA polymerase holoenzyme to recognise different DNA promoter sequences that identify transcription start sites. [3] Both β and β' contribute to the formation of the catalytic site for RNA synthesis, while the two α subunits are essential for assembly of the enzyme and activation by regulatory proteins.

1.1.1.1 Transcription initiation

Transcription process is made up of four stages: binding of RNA polymerase at promoter site, initiation of polymerisation, chain elongation and chain termination. The process begins with the σ-subunit of RNA polymerase recognising and binding to the promoter sequence. The RNA polymerase and the promoter form a closed promoter complex, where the dsDNA has not yet been unwound so that the RNA polymerase can read the base sequence of the DNA template strand and transcribe it into a complementary RNA sequence. RNA polymerase binds an 'envelope', a nucleotide sequence spanning the region from -70 to +20, where the +1 position is defined as the transcription start site (the base in DNA that specifies the first base in the RNA transcript).

Prokaryotic promoters vary in sizes from 20 to 200 base pairs, but typically consist of a 40 base pair region located to the 5'-side of the transcription start site. The promoter has two consensus sequence elements. The Pribnow box containing a conserved hexameric sequence (5'-TATAAT-3') is usually located 10 nucleotides upstream from the transcription start site, and another hexameric sequence (5'-TTGACA-3') located 35 nucleotides upstream from the transcription start site. [4, 5] The σ-subunits of RNA polymerase bind to these consensus elements as well as other sequences that provide specificity.

Next, the DNA duplex must be 'opened' so that RNA polymerase can gain access to the single stranded template in order for transcription to begin. The efficiency of transcription initiation is inversely proportional to the melting temperature (T_m) in the Pribnow box, which suggests that the A:T-rich nature of this region is suited for easy unwinding of the DNA duplex and creation of the open promoter complex. [3] The σ-subunit of RNA polymerase is directly involved in unwinding the dsDNA, and its interaction with the non-template strand holds the complex formed between RNA polymerase and promoter DNA open, with the σ-subunit acting as a sequence-specific single-stranded DNA-binding protein. The RNA polymerase has two binding sites for nucleoside triphosphates (NTPs) during initiation step; the initiation site and the elongation site. The first nucleotide binds at the initiation site, base-pairing with the exposed base at the +1 position within the opened promoter complex. The second incoming nucleotide binds at the elongation site, base-paring with the +2 base. The polymerase then catalyses nucleophilic attack by the 3'-OH of the growing RNA chain on the α phosphorus atom of a complementary ribonucleoside triphosphate.

1.1.1.2 Chain Elongation and Termination

The σ-subunit dissociates once a short oligonucleotide chain has been synthesised and subsequent elongation of the RNA transcript is governed solely by the core polymerase ($\alpha_2\beta\beta'$). As the RNA transcript gets longer, tangling of the RNA transcript to the non-template DNA strand can occur and topoisomerases are employed to unknot tangled DNA segments during the transcription process. [3] Once the desired mRNA strand has been completely transcribed, prokaryotic cells use two methods to terminate elongation; intrinsic termination governed by the presence of a specific sequence that terminates transcription [2] or a protein called ρ termination factor. ρ Termination factor binds to mRNA as a homohexamer and posess RNA-dependent ATPase and helicase activities. [2] It actively terminates transcription, using the energy from ATP hydrolysis to mediate the dissociation of the nascent transcript. [6] In *E. coli,* ρ termination factors are used at the end of genes and operons and at regulatory sites (attenuators) preceding genes. [6] Intrinsic termination, (also known as ρ-independent termination) on the other hand can terminate transcription by the mRNA forming hairpin structures. A protein bound to RNA polymerase known as *nus*A then binds to the hairpin structure tightly enough to halt the polymerase, ultimately terminating transcription. [7, 8]

1.1.1.3 Regulation of Prokaryotic Transcription

Although transcription in prokaryotes is highly efficient, it needs to be regulated to enable adaptation to environment changes and stress. In bacteria, genes encoding enzymes are often grouped adjacent to one another resulting in a polycistronic mRNA encoding many of the enzymes of the metabolic pathway. The rate of transcription is only limited by the accessibility of the promoter sequence to RNA polymerase,

therefore regulation of transcription is carried out by transcription factors acting as repressors (or negative regulators), which bind tightly to specific sites located next to the promoter (operator sites). Interaction of a regulatory protein with the operator controls transcription of the gene cluster by controlling access of RNA polymerase to the promoter. These combinations of operator and promoter sequences are known as operons. [3] An example of such a repressor is bacteriophage λ repressor. Bacteriophage λ is a virus that infects *E. coli* and has two modes of infection; the *lytic pathway* which destroys its host, or the *lysogenic pathway* where it becomes part of its host. In the lytic pathway, viral functions are fully expressed; viral DNA and proteins are rapidly produced and packaged together, leading to eventual lysis of the host cell and the release of approximately 100 progeny virions. In the lysogenic pathway, the phage DNA becomes integrated into the host-cell genome and can be replicated together with host-cell DNA for many generations (**Figure 1.1.1.3.1**). [1] Environmental changes can trigger the expression of this dormant viral DNA, which leads to the formation of progeny virus and lysis of the host. For example, studies have shown that when lysogenic *E. coli* are irradiated with UV light almost the entire population will undergo lysis. [5]
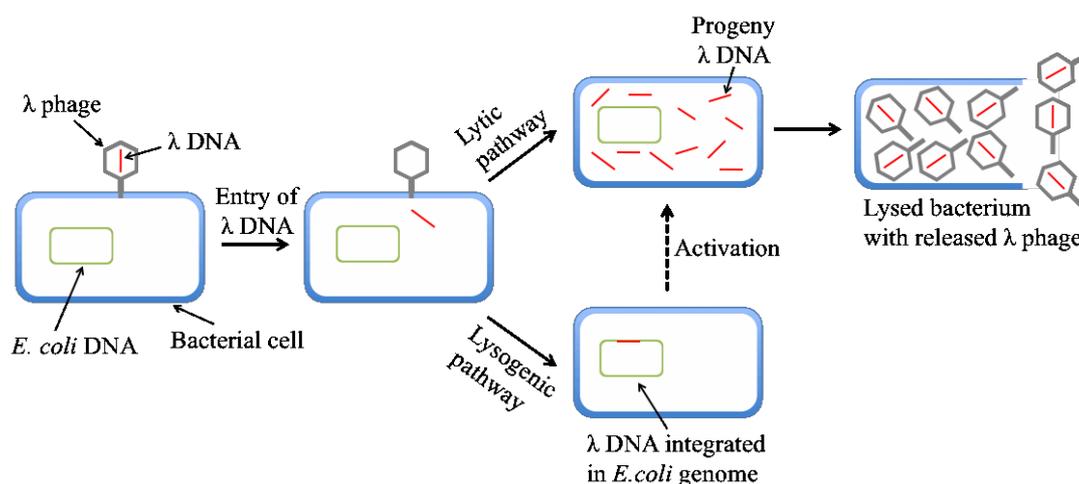
Figure 1.1.1.3. 1. *Diagram of the lytic and lysogenic pathways of bacterial λ phage.* [1]

In a lysogenic bateria, a single λ phage gene is actively transcribed and translated to a single protein – the λ repressor. This repressor works as both a positive and negative regulator. It binds to operators (O_R1, O_R2 and O_R3) on the host-integrated λ DNA, turning its own gene 'on' and repressing all other phage genes. UV light inactivates the λ repressor, resulting in the synthesis of a second phage regulatory protein known as Cro. [5] Cro, a protein essential to lytic growth, binds to the same sequence as the λ repressor, but has the opposite effect. The promoter sequences for repressor and Cro are adjacent to each other and overlap the operator sites (**Figure 1.1.1.3.2**).
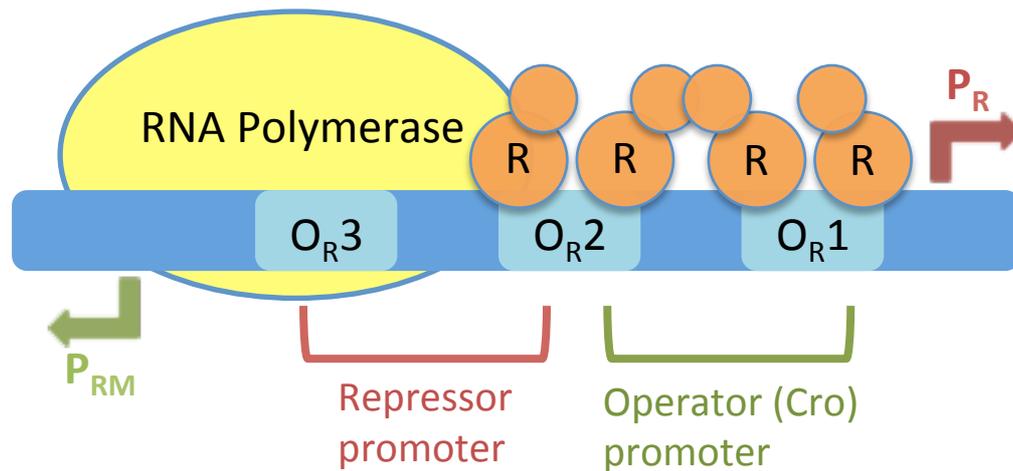
Figure 1.1.1.3. 2 . *Transcription of λ repressor. O_R1, O_R2 and O_R3 are the binding sites for λ repressor (R) and Cro. Cro, a repressor, works in opposition to the phage's repressor to control the genetic switch that determines whether a lytic or lysogenic cycle will follow infection. Cro and repressor compete for control of an operator region containing three operators (O_R1, O_R2 and O_R3) that determines the state of the lytic/lysogenic genetic switch. If this competition is won by the repressor, R, transcription of the Cro gene is blocked and repressor synthesis is maintained, resulting in lysogeny. If the competition is won by Cro, the late gene of λ phage will be expressed, resulting in lysis. In this case, Cro blocks transcription that occurs from P_{RM}, the promoter that is responsible for the maintenance of repressor transcription.* [5]

These two regulatory proteins in combination with RNA polymerase and their promoter and operator sites on the DNA constitute a regulatory 'switch' and exemplify negative transcription control in prokaryotes.

1.1.2 Eukaryotic Transcription

Transcription in eukaryotes is a much more complex process than in prokaryotes. In prokaryotes transcription and translation are closely coupled, in eukaryotes transcription occurs inside the nucleus and translation happens in the rough endoplasmic reticulum. This spatial and temporal separation of transcription and translation enables eukaryotes to regulate gene expression in much more intricate ways, contributing to the richness of eukaryotes. [1]

In bacteria, polypeptide chains are encoded by triplet codons in DNA, which is also the case for higher organism such as eukaryotes. It was discovered that newly synthesised RNA chains isolated from the nuclei of eukaryotic cells are much larger than the mRNA molecules that result from them. This is due to the presence of sequences known as intervening sequences, most commonly known as *introns*, that are removed from the primary transcript by a process known as RNA splicing, leaving only sequences that are found in mature RNA known as *exons* (for expressed region, **Figure 1.1.2.1**). The splicing of introns is carried out by a complex of small nuclear RNA (snRNA) and protein subunits collectively known as the spliceosome, which recognises trigger sequences in nascent RNA that specify splice sites. [1] Introns nearly always begin with GU and end with an AG that is proceeded by a pyrimidine-rich tract (**Figure 1.1.2.2**). Individual exons encode discrete structural and functional units of proteins, for example the central exon of myoglobin and haemoglobin genes encodes a heme-binding region that can reversibly bind O_2. [1] Other exons specify α-helical segments that anchor proteins in cell membrane. An entire domain of protein may be encoded by a single exon.
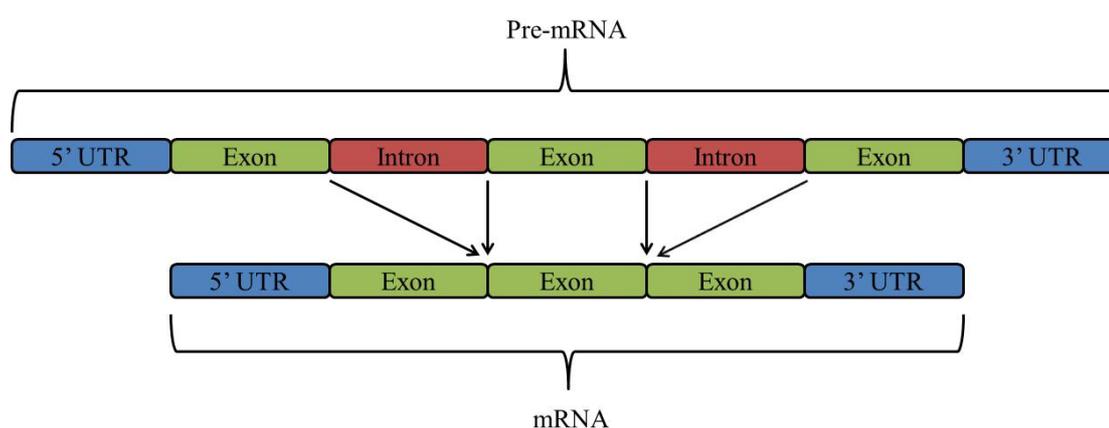
Figure 1.1.2. 1. *Diagram of exons and introns in pre-mRNA and mRNA after splicing by spliceosomes. The UTRs are non-coding parts of exons at the end of the mRNA.*
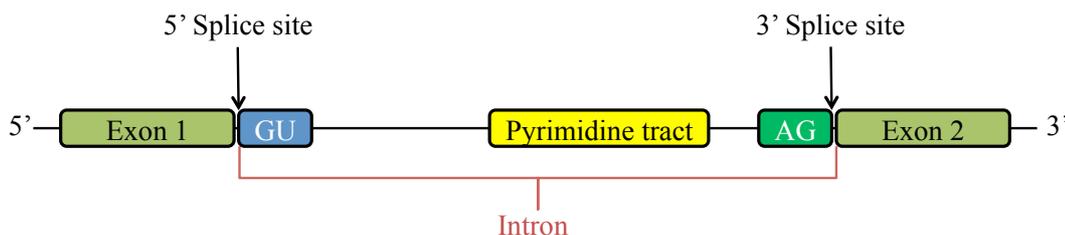
Figure 1.1.2. 2. *Consensus sequence for the splicing of mRNA precursors.*

While prokaryotes use a single type of polymerase for the synthesis of almost all RNA, the nucleus of eukaryotes contains three types of RNA polymerases differing in template specificity, localisation and susceptibility to inhibitors. The three types of RNA polymerases are: RNA polymerase I is found in the nucleoli and specifically transcribes the genes for 18S, 5.8S and 28S ribosomal RNA. RNA polymerase II, located in the nucleoplasm, is used to synthesise messenger RNA (mRNA); and RNA polymerase III, also located in the nucleoplasm, is used to synthesise all transfer RNA (tRNA) and other ribosomal RNA molecules. [1] Similarly to prokaryotes, eukaryotic RNA polymerases do not require a double stranded primer and synthesis proceeds in the 5' -> 3' direction according to instructions given by an antiparallel DNA template strand (the coding strand).

Type II eukaryotic RNA polymerases contain between 8 and 12 subunits, just over double that of their prokaryotic counterpart. Two large subunits known as RPB1 (~220 kDa) and RPB2 (~220 kDa) correspond to the catalytic β and β' subunits in the core of the prokaryotic enzyme. To initiate transcription RNA polymerase II promoter sequences are required. These are located about 25 bases upstream of the transcription start site and are known as the TATA box. The TATA box is present is nearly all eukaryotic organisms that give rise to mRNA. Studies on the TATA box sequences have demonstrated that mutation of a single base pair strongly impairs promoter

activity, highlighting the importance of these RNA polymerase II specific recognition sequences. [9] Most TATA boxes are flanked by GC-rich sequences.

Many promoters contain a CAAT box, and some contain a GC box and sequences containing a TATA box and the loci of these motifs, where present, in the upstream region collectively define the promoter. The presence of a CAAT box, usually located around -80 relative to the transcription start site, signifies a strong promoter. [3] One or more copies of the sequence GGGCGG or its complement (also known as the GC box) have been found upstream from the transcription start sites of 'housekeeping genes'. Housekeeping genes encode proteins commonly present in all cells and essential to normal function. [3] Constitutively expressed genes, which are continuously expressed rather than regulated, tend to include GC boxes located around the 35 bases upstream of the transcription start site.

Transcription Factor II (TFII) complexes guide RNA polymerase II to the transcription start sites. Transcription is initiated by the TFIID complex binding to the TATA box, where TFIIA is recruited, followed by TFIIB (**Figure 1.1.2.3**). RNA polymerase II and then TFIIE join the other factors to form a complex called the basal transcription apparatus. [1] This complex can transcribe DNA, albeit slowly, but additional transcription factors are required for selectivity and to achieve higher levels of mRNA synthesis. The key initial event is the recognition of the TATA box by the TATA box binding protein (TBP, section **1.2.7.1**, **Figure 1.2.7.1**), an approximately 30 kDa component of the 700 kDa TFIID. TBP is known to bind 10^5 fold more tightly to the TATA box than to non-cognate sequences and the dissociation constant of the specific complex is approximately 1 nM. TBP consists of two similar domains each containing two α helices and a five-stranded antiparallel β sheet. The bases of the TATA box bind to the concave surface of TBP and the saddle-shaped protein induces large

conformational changes in the DNA strand. A sharp kink is produced at each end of the seven base pair TATA box, causing a partial unwinding of the double helix and a widening of its minor groove, enabling extensive contacts with the anti-parallel β-sheets on the concave side of TBP. The resulting complex features many hydrophobic interactions at this interface, with four phenylalanines intercalated between base pairs of the TATA box. The flexibility of AT-rich sequences is exploited to bend the DNA locally whilst the B-DNA structure is conserved outside the TATA box. This complex is asymmetric, creating a unique start site and assuring that transcription proceeds unidirectionally.

The complex machinery described above is only sufficient for basal transcription; many more transcription factors are involved in the initiation and regulation of the specific additional genes required in each particular cell type and/or in response to other signals. To understand these greater levels of complexity it is important to examine the methods used by the proteins involved to recognise and bind to specific DNA sequences, as well as how these proteins interact with one another, and how such interactions influence which genes are transcribed and the level of transcription.
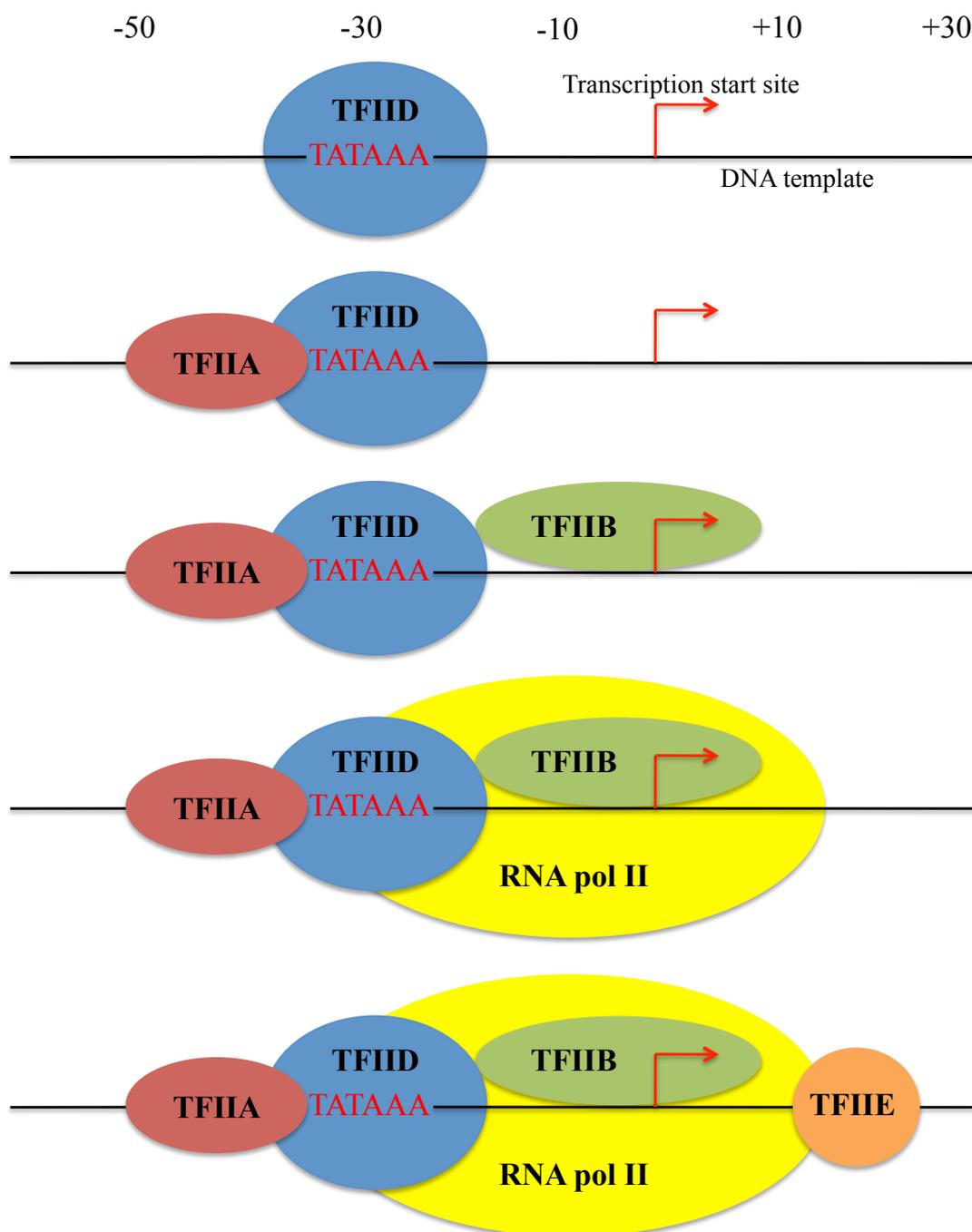
Figure 1.1.2.3. *Diagrams of the pathways in the assembly of the basal transcriptional complex. Transcription factors TFIIA, B, D, and E are essential in initiating transcription by RNA polymerase II. The stepwise assembly of these general transcription factors begins with the binding of TFIID to the TATA box. This is in turn recruits TFIIA, B and D, followed by RNA polymerase II.*

1.2 Eukaryotic Transcription Factors

The structures of many transcription factors and their complexes with DNA have been obtained using X-ray crystallography and NMR spectroscopy and have revealed conserved structural elements that make direct contact with DNA (DNA-binding domains [DBD]). [10] Transcription factors can be classified into families based on the overall structure of their DBDs which include: homeodomains, zinc finger proteins, steroid receptors, MADS-box proteins, basic zippers (bZIP), basic helix-loop-helix (bHLH) proteins and β-sheet motifs.

1.2.1 Homeodomains

Many eukaryotic transcription factors make use of the highly conserved homeodomain structure to recognise specific DNA sequences. Homeodomains are approximately 60 amino acid residues long and are encoded by a 180-bp DNA sequence known as homeobox, or homeotic genes. Homeoboxes were first identified in many of the genes that help control embryogenesis in the fruit fly *Drosophila melanogaster*. [11] Mutation of these genes can result in altered distribution of *Drosophila* body parts; for example, by introducing mutations to *Antennapedia* homeotic gene, it was possible to grow an extra leg in the region which contains the antenna in a normal fly. [12, 13] Other genes such as *Ultrabithorax* are responsible for repression of the formation of wings in the third thoracic segment of the fly. [14, 15] The engrailed homeodomain is responsible for segment polarity in *Drosophila* during embryogenesis, where it defines the anterior and posterior compartments both in segmentation and in limb specification. Engrailed is known to

bind preferentially to DNA sequences in the promoter of the *engrailed* gene for positive feedback. [16] Correct function of these genes are key to development in *Drosophila*.

Homeodomains are characterised by a helix-turn-helix (HTH) motif that makes contact with the major and minor grooves of its DNA target recognition site. [17] X-ray crystallography shows homeodomains to be globular and contain three α-helices, with helix I and II almost exactly antiparallel to each other, whereas helix III crosses over the remaining two helices at an angle of approximately 120 ^{o}C (**Figure 1.2.1.1**). [16]
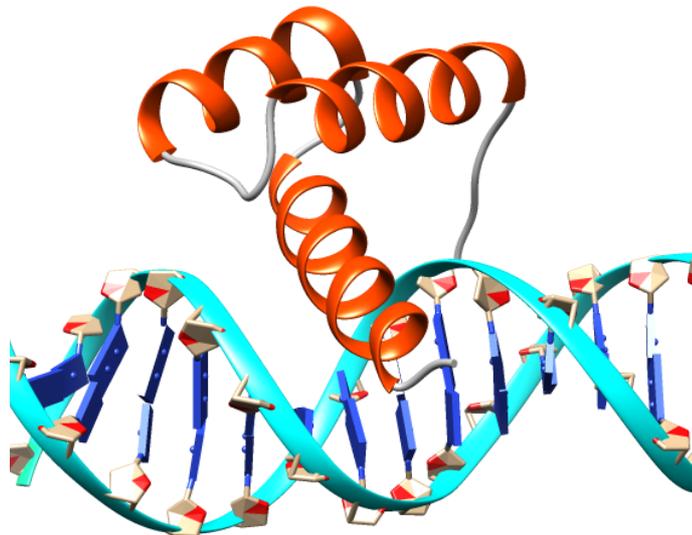
Figure 1.2.1.1. *Crystal structure of engrailed homeodomain-DNA complex (3HDD.pdb). [18] Helix-III can be clearly seen lying inside the major groove of the DNA. In turn this helix-III is stabilised by perpendicular helices I and II, which are anti parallel to each other forming a hydrophobic core.*

One face of helix-III is hydrophobic and packs against helices I and II to form the hydrophobic core of the protein domain. The opposite face side of helix-III is hydrophilic, with multiple positive charges; it fits into the major groove of DNA making contacts with the bases and sugar backbone phosphates and forms the DNA binding domain of homeodomain. [16] Seven amino acids from helix-III make contacts

with the phosphate groups in the DNA backbone and align the recognition helix in the major groove so that base-specific contacts can be made by other residues. For example, hydrogen bonds are made with DNA bases by Asn51 and Gln50, whilst Ile47 makes hydrophobic contacts with a thymine (**Chapter 3**). [16] The N-terminal flexible arm makes additional specific contacts to bases in the minor groove, allowing residues 3, 5 and 7 to make base-specific contacts; whilst residue 6 interacts with the sugar-phosphate backbone. The loop between helices I and II interacts with the DNA backbone on the other side of the major groove. [19] Attempts to reduce the size of the wild type engrained homeodomain by removing parts of the N-terminal resulted in a 10-fold reduction of binding affinity to the target DNA, emphasising the importance of the N-terminal region to the stability of the protein. [20]

Crystal structures of λ Cro protein and the DNA binding domain of λ repressor revealed a conserved recognition motif consisting of an α-helix, a turn, and a second α-helix. [21] However, these prokaryotic DBDs tend to be small parts of much larger proteins and the HTH cannot fold or function on their own. In constrast, homeodomains can fold properly and bind to target DNA even in the absence of the remainder of the protein. [21] Homeodomains are known to interact with other proteins for gene regulation. For example, the MAT α2 homeodomain protein found in yeast interacts with two different DNA-binding proteins MCM1 (a protein from the MADS-box family of transcription factors) and MAT α1 (another homeodomain protein) in order to successfully bind to operator sequences and regulate the transcription of cell type-specific genes. [19, 22]

1.2.2 Basic Zippers

The basic-region-leucine-zipper (bZip) domain is one of the simplest protein structures used for DNA recognition and binding. They consist of a region rich in amino acid residues with basic sidechains that bind to DNA and a leucine-rich region known as leucine zipper, that causes dimerization of bZip proteins.[17] This type of domain is found in many eukaryotic transcription factors including the yeast transcription activator GCN4 [23] and the transforming oncoproteins Fos and Jun that together make up the AP-1 family of factors.[24]

A typical leucine zipper region is approximately 30-40 residues long with a leucine at every third and seventh residue, forming a heptad repeat. Studies have demonstrated that these leucines drive dimerization by forming a parallel coiled-coil structure.[25] **Figure 1.2.2.1** illustrates the end view of a leucine zipper coiled-coil and the seven unique amino acid postions (a, b, c, d, e, f and g) found in each heptad.[26, 27] Residues 'a' and 'd' are typically hydrophobic along the dimerization interface to create a hydrophobic core that contributes most of the energy needed to stabilise the zipper.[27, 28] Leucine is the favoured hydrophobic residue for the 'd' position and isoleucine is the favoured residue for the 'a' position.[29, 30] The 'g' and 'e' position of the leucine zipper frequently contain charged amino acids.[31] Studies on the atomic structure of leucine zipper dimers have shown that oppositely charged amino acids in the 'g' positions and the following 'E' position lie across the hydrophobic interface and interact inter-helically, leading to the stabilisation of the structure this helping to regulate the specificity of bZip protein dimerisation.[26, 32-35, 36]
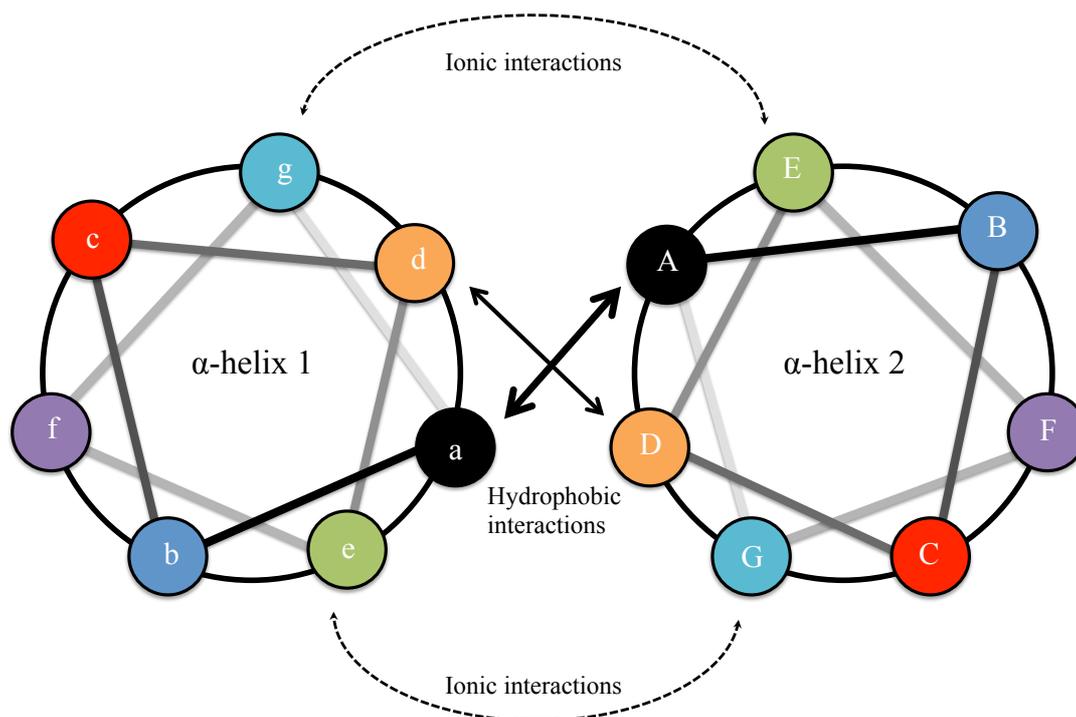
Figure 1.2.2.1. *End-view diagram of coiled-coil structure. Looking down the helix axis from N-terminus to C-terminus, with the amino acids in the seven unique positions (a, b, c, d, e, f and g) of the heptad presented as circles, oriented as the amino acid side chains would exit the α-helix. Amino acids on the second helix of the dimer are denoted A, B, C, D, E, F and G. The potential electrostatic interactions between the g of one helix and the following E postion of the opposite position (g <-> E) are indicated by the black dashed-arrow. The opposite side of a leucine zipper coiled-coil is identical in a homodimer but different in a heterodimer. [27]*

NMR and X-ray crystallographic studies on Fos-Jun heterodimer and GCN4 homodimer complexes demonstrate how a pair of zippers wrap around each other resulting in a parallel, two-stranded α-helical coiled-coil with a continuous hydrophobic interface. [16, 25 36] Dimerization of leucine zippers is essential for DNA binding with mutations that prevent dimerization also preventing DNA binding. [37] Dimerization facilitates DNA binding by organising two DBD α-helical structures to bind to a palindromic DNA recognition site; clamping around the major groove of the DNA like a pair of scissors (**Figure 1.2.2.1**). [5, 36]
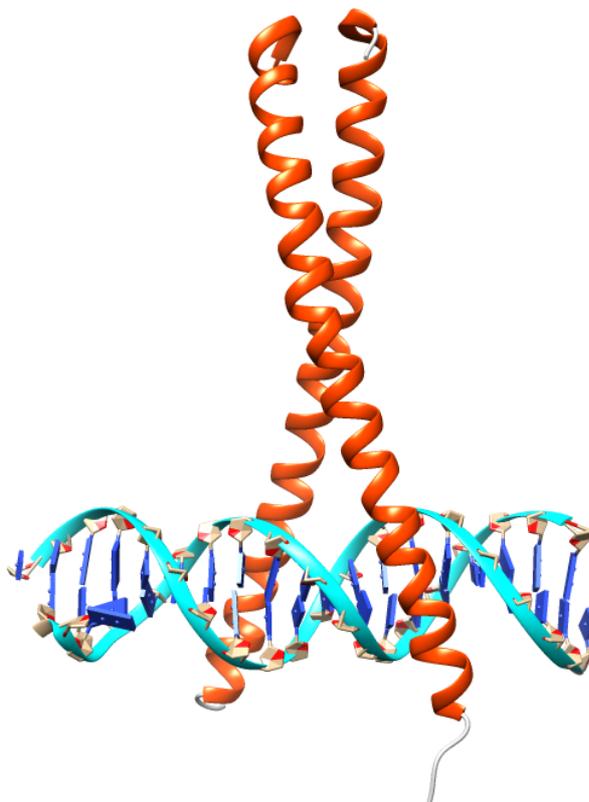
Figure 1.2.2.2. *Crystal structure of GCN4 (orange) leucine zipper (upper part) bound to DNA (1ysa.pdb).* [26]

1.2.3 Zinc Fingers

Zinc finger proteins were first described in transcription factor TFIIIA of *Xenopus*, a protein which plays a critical role in regulating the transcription of the 5S ribosomal RNA genes by RNA polymerase III. [4] Upon purifying this transcription factor a repeated structure was found, with between seven and eleven atoms of zinc associated per molecule of purified protein. [38] Cloning the gene encoding TFIIIA revealed that this repeated structure contained two invariant cysteine and two invariant histidine residues, which were each thought to bind to a single zinc atom. [39, 40] NMR and X-ray studies showed that the two cysteine, two histidine finger is a compact globular domain made up of a 12-residue α-helix packed against an irregular β-sheet and the zinc sandwiched

between them. [40] The histidines are located on the inside face of the helix, whereas the

cysteines flank a hairpin turn (**Figure 1.2.3.1**).
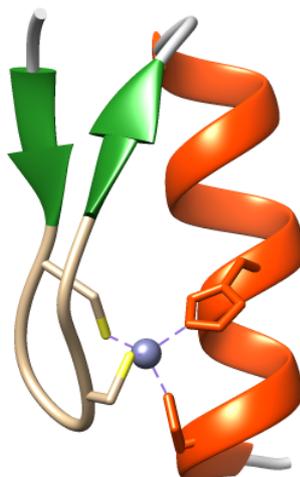
Figure 1.2.3.1. *Diagram of a zinc finger motif. The zinc ion is shown in purple, the α-helix is shown in orange, the β-strands are in green and the turns are in cream. The zinc atom can be seen coordinating with two histidines and two cysteines (1aay.pdb).* [41]

Zinc co-ordination *via* cysteine and histidine residues serves as a scaffold for the helical

region, which comes into direct contact with the DNA, [5] two anti-parallel β-sheets also

pack against the α-helix to help stabilise it. [42] When bound to DNA, the α-helix lies in

the major groove of the DNA and makes sequence specific interactions with the bases

of DNA whilst the β-sheets lie further away from the helical axis of the DNA and

contact the phosphate backbone (**Figure 1.2.3.2**). [40, 43]
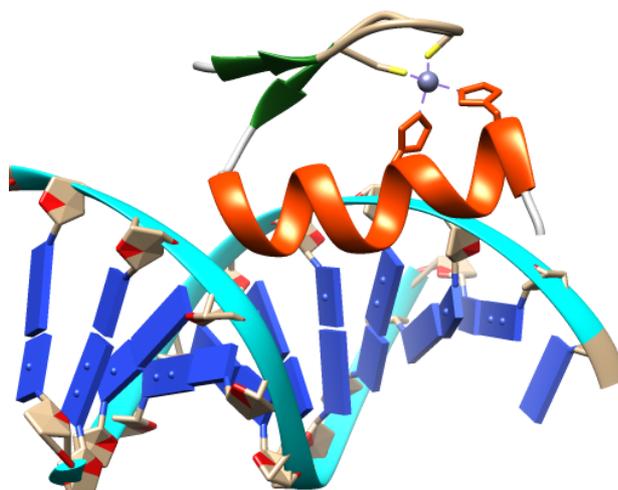
Figure 1.2.3.2. *Crystal structure of zinc finger protein Zif268 bound to the target DNA (1aay.pdb).* [41]

1.2.4 Basic Helix-Loop-Helix Proteins

The Basic Helix-Loop-Helix domain (bHLH) is a structural motif common to a family of transcription factors. These include the E12 and E47 human transcription factors, which play a key role in the development of the nervous system, [44] MyoD which activates genes specific to skeletal muscle, [5] and MASH-1 which is a central factor for the regulation of the differentiation of committed neuronal precursor cells of the peripheral nervous system. [45] The motif consists of two amphipathic helices, containing all the charged amino acids on one side of the helix, separated by a non-helical loop of variable length [44] with one of the helices containing a basic region responsible for binding to a DNA target known as E-box. [46] The bHLH motif works in a similar fashion to the leucine zipper, causing dimerization of the transcription factor molecule and facilitating DNA binding by basic sidechains (**Figure 1.2.4.1**). [47]
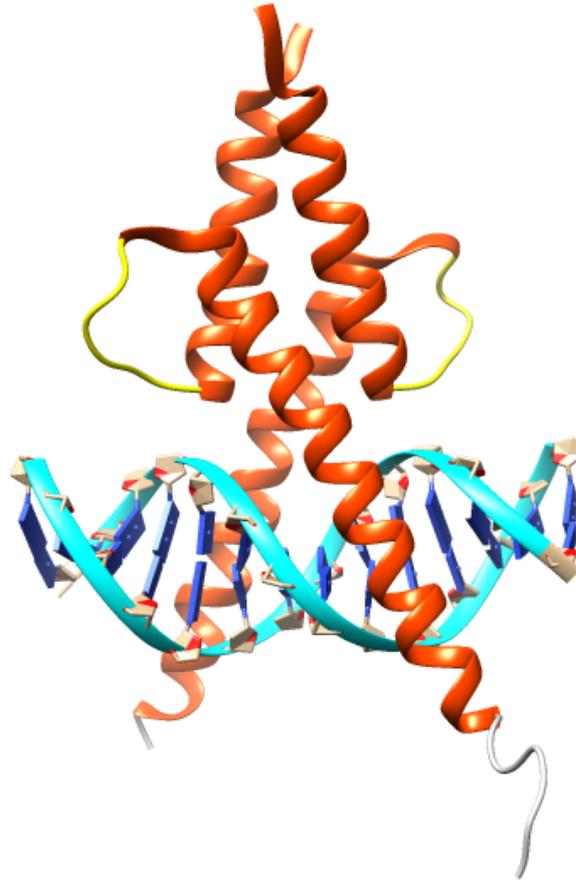
Figure 1.2.4.1. *Crystal structure of bHLH protein MyoD bound to its target DNA (1mdy.pdb).* [48] *The α-helices are shown in orange and the loops are in yellow.*

Accordingly, deletion or mutation of the MyoD basic domain protein does not prevent dimerization but does prevent binding of the protein to target DNA, [49] whereas mutations or deletions in the helix-loop-helix region, prevent both dimerization and DNA recognition. [5] When the DNA binding domain of MyoD was deleted, the resulting protein lost specific DNA binding ability. Selectivity was restored by substituting the basic domain of MyoD with the DBD domain of E12 protein. [50] However, the hybrid protein does not activate muscle-specific gene expression, suggesting that in addition to mediating DNA binding, the basic region of MyoD also contains additional elements involved in the activation of muscle-specific genes. [5, 50] Replacing three amino acids within the E12 basic region to their MyoD equivalents allowed the E12 basic region to

activate muscle-specific gene expression after DNA binding by the protein. [51] The crystal structure of MyoD bound to DNA (**Figure 1.2.4.1**) [48] suggests that these amino acids play a role in allowing the MyoD basic region to assume a particular conformation in which it interacts with other transcription factors. [5] Kanshal *et al.*, have demonstrated that the substitution of these same amino acids in E12 for their MyoD equivalents allows the mutant E12 protein to bind to another muscle-specific transcription factor MEF2A, an interaction normally a unique property of MyoD, [52] suggesting that the basic domain functions both as a DNA binding domain and as a site for protein-protein interactions critical for transcriptional activation.

1.2.5 Nuclear Receptors

Nuclear receptors are the largest family of eukaryotic transcription factors and are responsible for sensing steroidal and non-steroidal hormones. The binding of these hormones to their cognate receptors activates the receptor and allows them to bind to a number of specific sites on DNA. [16] Related receptors include retinoic-acid receptors and a legion of orphan receptors (receptors that have not yet been paired with physiological ligands, such as the nerve growth factor-induced receptor NGFI), which play major roles in intracellular signaling and carcinogenesis. [53] The activated forms of these receptors bind to target DNA and activate transcription of genes. [5]

In common with other most transcription factors, the nuclear receptors have structural features in common comprising a central, sequence specific, DNA binding domain and a ligand binding domain (LBD) contained in the C-terminal half of the receptor and recognises specific hormonal and non-hormonal ligands. [54] These receptors contain variable N-terminal and C-terminal domains, as well as a variable length hinge region

between the DBD and LBD. Nuclear receptors can exist as homo- or heterodimers with each partner binding to specific regulatory element sequences that exist as half-sites separated by variable length nucleotide spacers between direct or inverted half-site repeats. [54] These nuclear receptors can also be classified into two different groups: Class 1 receptors (steroid hormone receptors), function as homodimers and bind to half-site regulatory element inverted repeats, and Class 2 receptors which exist as heterodimers with Retinoid X Receptor partners and function in a ligand dependent manner (**Figure 1.2.5.1**) (**Figure 1.2.5.2**). [55]
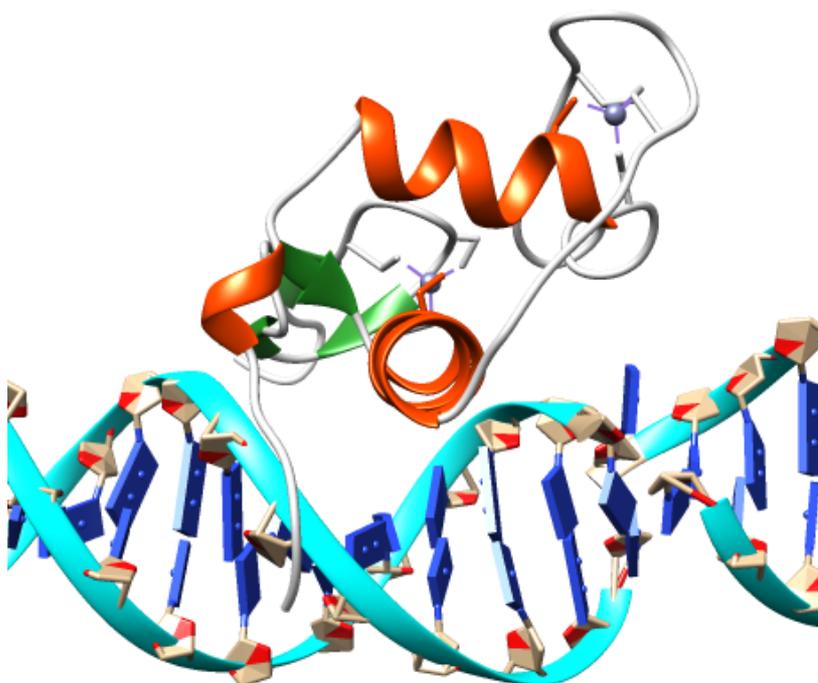
Figure 1.2.5.1. *Structure of Retinoid X Receptor bound to DNA (1by4). The α-helices are shown in orange, the turns are in grey, the β-strands are in green and the zinc atoms are in purple.*

The DBDs of nuclear receptors have structures reminiscent of zinc finger proteins where each DBD contains two tetrahedrally coordinated zinc ions, but with four cysteine residues rather than the two cysteine and two histidine residues acting as the

ligands; an arrangement seen in some sub-families of zinc finger proteins. [39] NMR and X-ray crystallography shows these domains contain two loop-helix motifs that are folded together to form a single structural unit, where each motif has a zinc ion liganded by two cysteines from the start of the loop and two cysteines from the N-terminus of the α-helix (**Figure 1.2.5.2**). [16, 56] The α-helices in the domain are highly amphipathic and the hydrophobic surfaces in each pack together to produce a stable and extensive hydrophobic core that stabilises the structure of the domain.
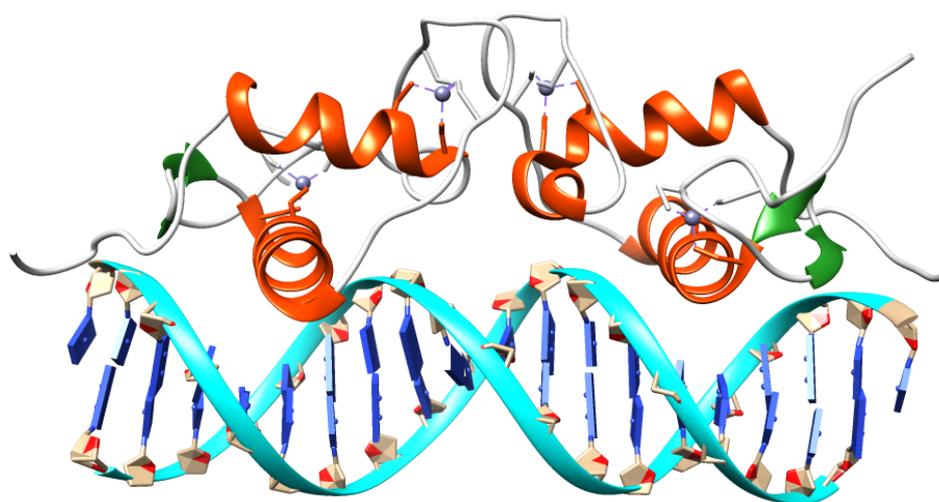
Figure 1.2.5.2. *Structure of the glucocorticoid receptor bound to target DNA (1r40.pdb). It binds to DNA as a dimer, the first helix of each subunit makes contacts with the bases on the DNA, while the second helix makes backbone contacts and acts a dimerisation interface. The α-helices are shown in orange, the turns are in grey, the β-strands are in green and the zinc atoms are in purple.* [56]

1.2.6 MADS-box Proteins

MADS-box proteins are DNA binding domains named after the first four transcription factors in which they were discovered: Mcm1 (minichromosome maintenance 1) in yeast, [57] AG (*Agamous*) and DEFA (*Antirrhinium majus*) in plants, [58] and SRF (Human serum response factor) in vertebrates. [59] These domains bind to DNA as a dimer and are organised in three layers (**Figure 1.2.6.1**). [59] The N-terminal region of the MADS-box is

predominantly hydrophilic with a high proportion of basic residues, whereas the C-terminal region is predominantly hydrophobic. [60] The N-terminal region binds to DNA through an anti-parallel coiled-coil of two amphipathic α-helices, one from each monomer of the dimer. The coiled-coil is oriented parallel to the minor groove and the DNA is bent around the protein with each α-helix binding in adjacent major grooves. The next layer of the domain is a four-stranded antiparallel β-sheet, packed against the coiled-coil, followed by an irregular coiled region and a short α-helix. [17] Myocyte Enhancer Factor 2 (MEF2) is a transcription factor belonging to this family of MADS-box proteins involved in the regulation of a number of muscle-specific genes and is postulated to potentiate the transcriptional activity of myogenic bHLH proteins such as MyoD, E12 and E47 proteins by recognising the basic region of these factors. [61, 62, 63, 64]
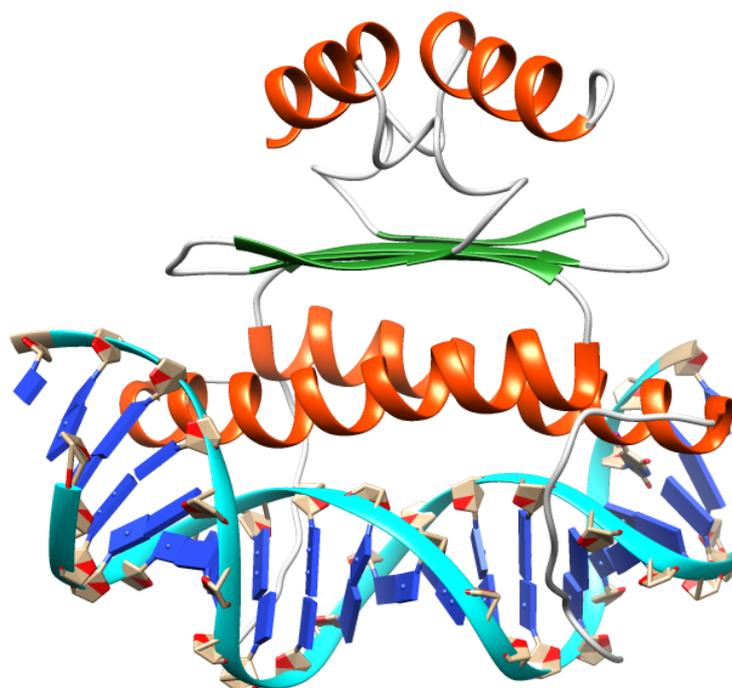
Figure 1.2.6.1. *Structure of SRF-DNA complex (1srs.pdb). [59] α-helices are in orange and β-sheets in green. The binding of the protein causes bending of the DNA by 72 °.*

1.2.7 β-Sheet Motifs

1.2.7.1 TATA Binding Proteins

TATA-box binding protein (TBP) is required by all three eukaryotic RNA polymerases for correct initiation of transcription of ribosomal, messenger, small nuclear and transfer RNAs. [65] The TBP binds DNA as a monomer consisting of a curved antiparallel β-sheet buttressed by helices. The concave surface of the β-sheet binds to the minor groove of the TATA box, and the DNA is kinked in two places resulting in DNA bending by approximately 100 ^{o} (**Figure 1.2.7.1**). [17, 65] Two pairs of phenylalanine residues make contact with the DNA, inserted between adjacent base-pairs resulting in the kinks in the DNA structure; other, polar side-chains make contacts with the base pairs in the minor groove and there are further hydrophobic interactions between TBP and the minor groove of the DNA. [66, 67]
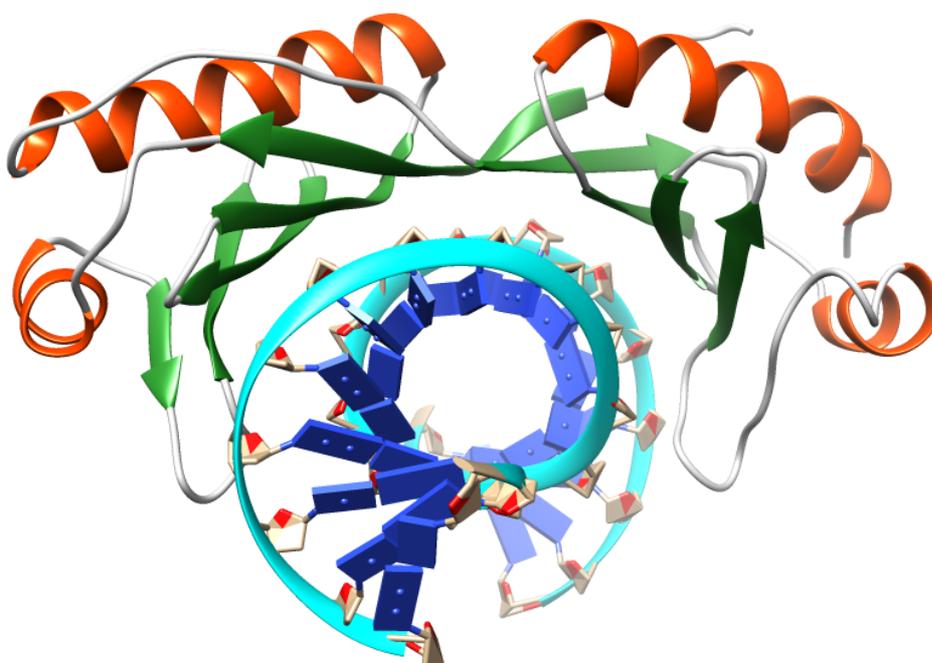
Figure 1.2.7.1. *Structure of TBP bound to target DNA (1cdw.pdb).* [65] *The α-helices are shown in orange, the β-sheets in green and coils in grey.*

1.2.7.2 Rel Family

The Rel family plays a major role in the immune response in higher organisms. The members of this family all contain the Rel homology region (RHR) that is responsible for protein dimerisation and DNA recognition. The structure consists of two β-sandwich domains of which the N-terminal domain binds to the DNA bases using well-defined loops between the β-strands (**Figure 1.2.7.2**). The C-terminal is the dimerization domain and is involved in binding to the phosphate backbone. [68, 69]
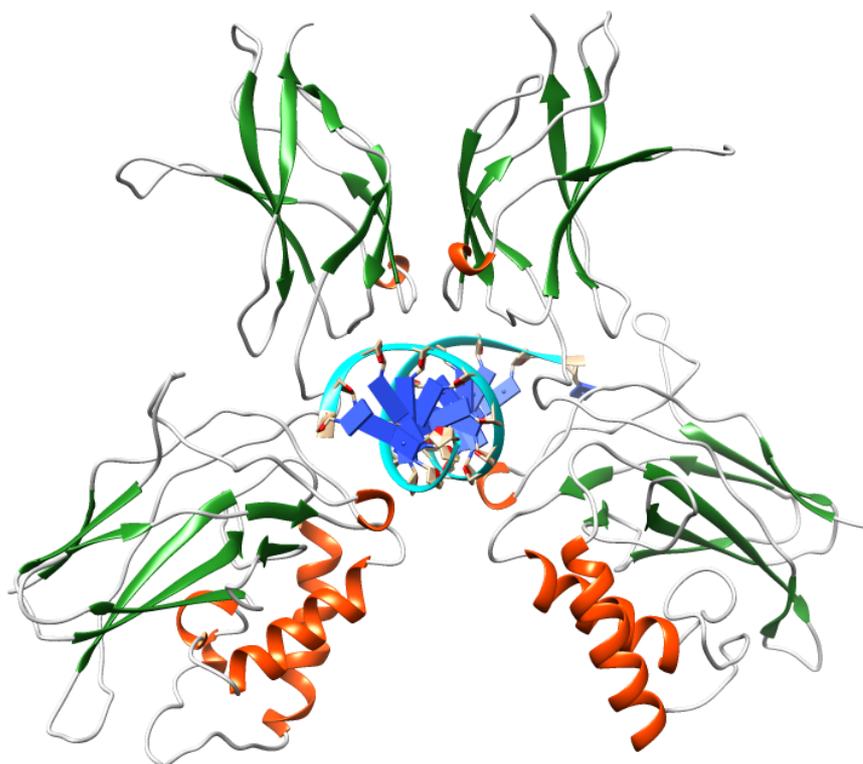
Figure 1.2.7.2. *Structure of the homodimeric NF-κB p50 transcription factor in complex with a κB site (1nfk.pdb).* [68] *The α-helices are in orange, the β-sheets are in green and coils are in grey.*

1.3 DNA Binding Specificity of Proteins

Many DNA-binding proteins make use of α-helices to make contact with major grooves of DNA. [65, 70] However, an isolated α-helix from these known motifs cannot bind DNA in a site-specific fashion or perform regulatory functions without the surrounding regions of the protein. [21] The orientation of α-helices to the major groove (or minor groove) also plays an important part on how these proteins bind onto their target DNA and when the particular pattern of amino acids on the surface of the protein matches the pattern of groups on the surface of the DNA double helix, binding takes place through the formation of hydrogen bonds and van der Waals contacts. [71] High specificity requires recognition of a number of bases in a DNA, although DNA binding domains of proteins often only require three to four bp of nucleotide sequence for recognition. Such short sequences occur too frequently to be unique and uniquely recognisable. In order to achieve higher specificity, several classes of DNA-binding proteins, such as bZIP proteins require, as previously discussed, dimerization of their monomer, in order to increase the total number of nucleotides recognised as well as the possibility of a variation in spacing between the two half-sites. Both of these measures increase the rarity of occurrence of the particular sequence and spacing of bases, increasing specificity. Protein-protein interactions are therefore vital for specific DNA binding and activation of gene transcription.

Direct contact with DNA bases is vital for site-specific binding but the methods used to contact bases differ greatly between DNA-binding protein families. Structural studies have shown direct hydrogen bonds between the protein side chains and the bases, occasional hydrogen bonds between the polypeptide backbone and the bases, hydrogen bonds mediated by water molecules and hydrophobic contacts are all used to achieve

specificity. [21] An example of water-mediated specific contact can seen in the bHLH domain of MyoD, where a nitrogen atom from Arg 111 side chain makes contact with oxygen of a guanine in the binding site through a water molecule (**Figure 1.3.1**). [72]
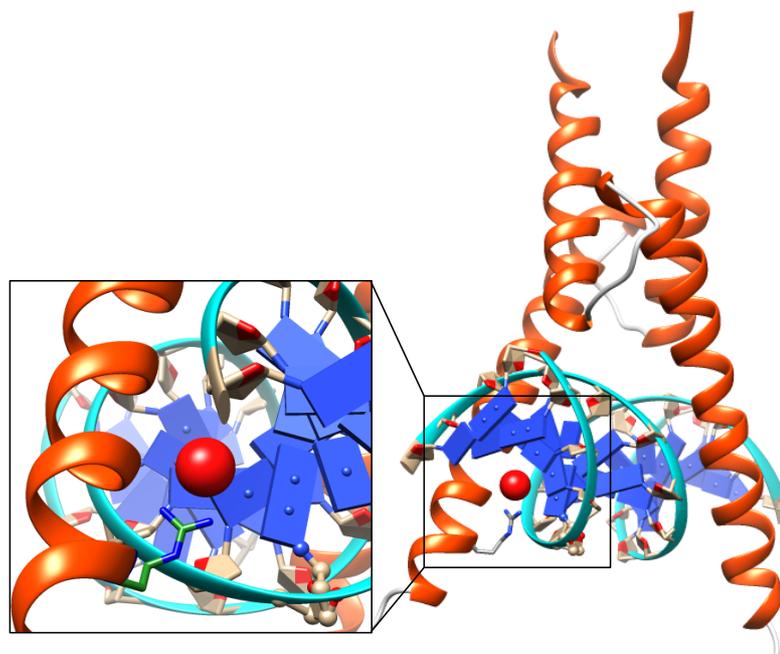
Figure 1.3.1. *Diagram of MyoD-DNA complex with the water molecule (red sphere) mediating the contact between the protein and DNA via Arginine 111 (green and blue) (1mdy.pdb).* [72]

Contacts made by proteins with the DNA backbone can also play an integral role in site-specific recognition. In the most well characterised complex, such as helix-turn-helix motif and zinc finger proteins, approximately half of all the hydrogen bonds are involved in making contacts with the DNA backbone; almost all of these contacts involve the phosphodiester oxygens. [21] There is no obvious pattern to which residues make contact with the backbone, with many different side chains' functionalities, including the –NH of the polypeptide backbone, known to hydrogen bond to the phosphodiester oxygens.

In addition, the DNA structure can also influence specificity. There are many examples where the binding of protein causes bending of DNA, these include TATA binding

proteins where the concave surface of the β-sheet causes bending of the DNA of approximately 100 ° upon binding to the minor groove, and the MADS-box serum response factor (SRF) protein that also causes DNA to bend by 72 °. [65, 59]

The double-stranded DNA structure can be described as a polymer with uniform structure, containing negatively charged sugar phosphate backbone and a core of stacked base pairs whose edges are exposed in the major and minor grooves. Each base pair has a set of characteristic functional groups, by which each DNA sequence can be described as having a chemical 'signature' characterized by the pattern of these groups exposed in the DNA grooves. [73] DNA binding proteins, such as transcription factors, recognise this chemical surface combined with its sequence-dependent variations in structure and flexibility by having a surface that is chemically complementary to that of the DNA, forming a series of favourable electro static and van der Waals interactions between the protein and the base pairs.

The vast majority of proteins recognise functional groups in the major groove of the DNA, as it is here that each base pair can be uniquely distinguished. The pattern of hydrogen bond donors and acceptors is less varied in the minor groove, with A-T similar to T-A and G-C similar to C-G. [73] It has also been suggested that sequence specificity may arise from hydrogen bonding interactions between the protein and DNA, due to the requirement for near collinear apposition of donor and acceptor groups. [73] Bidentate interactions, by which a single side chain forms two or more hydrogen bonds with the DNA allow an even greater degree of specificity than a single hydrogen bonds to a side chain. In order to achieve this interaction, amino acids must contain more than one hydrogen-bonding atom. [74] Arginine for example can recognise a guanine base through bidentate interactions. There are two possible arginine-guanine interaction conformations, the end-on approach interacts with one or both of the distal

nitrogen atoms (**Figure 1.3.2**.**A-C**) and the side-on approach uses the N^{ε} and $N^{\eta 1}$ atoms (**Figure 1.3.2**.**D**). [73, 74] It has been suggested that the first interaction conformation is preferred over the second due the easier access of arginine side chain as probes compared to extending along the groove floor. [74] Van der Waals interactions also play an important part in specificity, where they impose steric constrains on the types of bases that can be accommodated at particular positions, therefore also playing a role in sequence selectivity. [74]
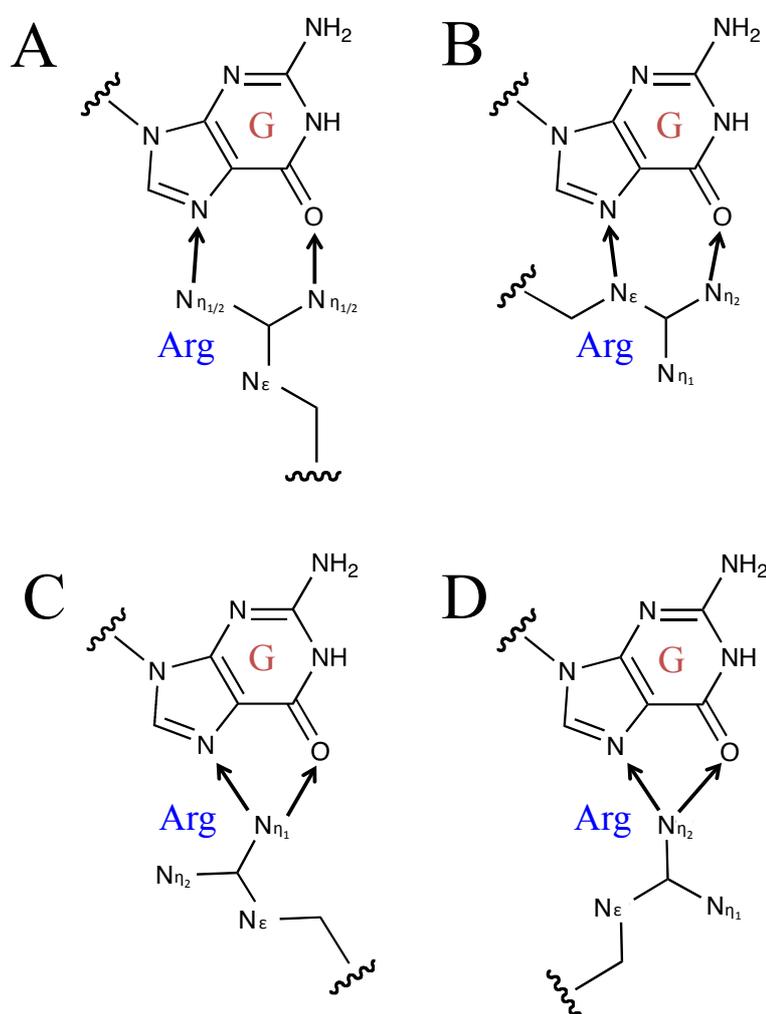
Figure 1.3.2. *Diagrams of bidentate interactions. The arrows are drawn between interacting atoms, pointing from the donor to the acceptor.* [74]

The specificity of a DNA binding protein to its cognate site can be determined by the difference in ΔG obtained for binding to a specific (S) and a non-specific (NS) site (**Equation 1.3.1**), or the ratio of the apparent dissociation constants (K_D) for those sites (**Equation 1.3.2**).

$$\Delta\Delta G = \Delta G^{NS} - \Delta G^{S} \qquad\qquad\qquad Equation\ 1.3.1$$

$$Specificity = K_D^{NS} / K_D^{S} \qquad\qquad\qquad Equation\ 1.3.2$$

Specificity of several prokaryotic and eukaryotic transcription factors has been determined using the above methodology (**Table 1.3**).

Protein	Specificity
λ Cro (O_R1)	1.8×10^5
lac repressor	2.9×10^7
MASH-1	3.2
GCN4	31
engrailed homeodomain	26

Table 1.3. *Specificity values of λ Cro,* [75] *lac repressor,* [76] *MASH-1,* [77] *GCN4* [78] *and engrailed homeodomain* [79] *transcription factors.*

1.4 Stabilisation of DNA-binding proteins and α-helices in vitro

Small peptides corresponding to the isolated helical or other motifs are usually disordered in solution, since the folding process is thermodynamically disfavoured. Naturally occurring proteins overcome this barrier using complex folds to display a small number of functionally important amino acids. This pre-organisation of the

conformation of small numbers of functional amino acids into the recognition motif is important for potent recognition of the binding domain of the protein.

The α-helix is stabilised by hydrogen bonding between the amide NH and the carboxylic (C=O) groups of the peptide backbone. The C=O group of one amino acid (*i*) makes a H-bond to the NH group situated four residues ahead in the sequence (*i+4*) (**Figure 1.4.1.A**).
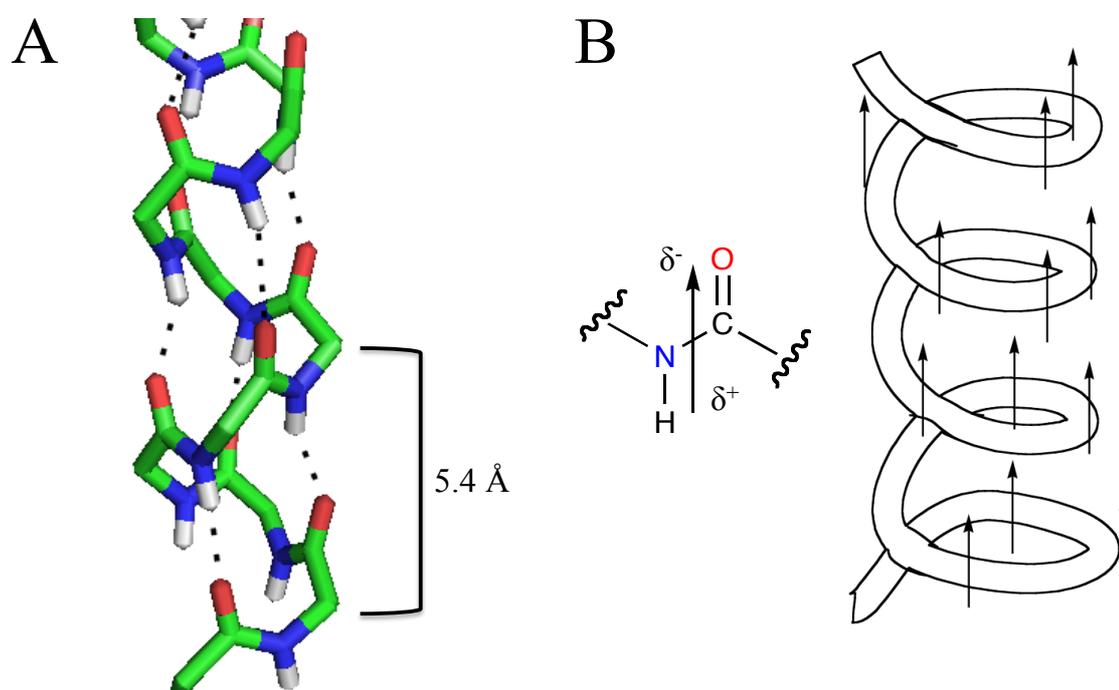
Figure 1.4.1. *Diagrams of an α-helix. A) Cartoon representation of a helix back bone to the helix axis. Hydrogen bonds are represented in black dotted lines. Other atoms are in standard colours. B) Diagram of dipoles on the helix.* [80]

The C=O group is almost parallel to the axis of the helix pointing towards the NH group and hydrogen bonded to it, while the side chains point away from the axis. One turn of the helix incorporates 3.6 amino acid residues, involving 13 atoms from the O to the H of the H bond. Each amino acid residue extends the helical axis by 1.5 Å. With 3.6 residues per turn, this is equal to 5.4 Å of the travel along the helix axis per turn which is also referred to as the translation distance or the 'pitch' of the α-helix. Since the first

and last four residues differ from the interior ones as they cannot make the intrahelical hydrogen bonds between the backbone C=O groups of one turn and the NH group of residues in the next, they therefore require hydrogen bonds to be made with either solvent or other residues in the protein. The first and last residues of a helix are frequently residues whose sidechains can provide such hydrogen bonds. [80] Helices are often amphipathic, where one side has hydrophobic residues that bind to hydrophobic areas in the protein, and the other has hydrophilic residues that interact with water. The alignment of dipoles of the amide bond backbone nearly parallel to the axis of an α-helix causes a net dipole moment with its positive pole at the N-terminus and negative pole at the C-terminus (**Figure 1.4.1.B**). [80]

In order for an α-helix to form from a disordered coil, four amino acids must adopt an α-helical conformation, making the $i + 4$ hydrogen bond. This nucleates the formation of further hydrogen bonds, propagating the α-helix at a rate and extent dependent on the intrinsic helical propensity of each amino acid residue. [81] However, disordered polypeptide chains in solution have considerable conformational freedom and high entropy. If side chains contain hydrogen-bonding groups, then these groups make bonds with the polypeptide backbone, if favourable, or with solvent molecules. Formation of a helix results in substantial loss in conformational freedom of the peptide due to the rigidity enforced by intra-helical hydrogen bonding, but this is compensated for by the release of water previously hydrogen bonded to the amides. [80]

The helical propensities of amino acids were initially investigated by examining the frequency that each amino acid occurs in α-helical, β-sheet, and random coil conformations. Latterly, the helical propensity was refined by studying the helical content of designed peptides comprised mainly of alanine. Alanine has the smallest side

chain of the L-amino acids and does not make electrostatic interactions with solvent molecules and therefore serves as a good model for calculating relative energies. [80] Using alanine-based peptides, Charkrabartty *et al.* measured the α-helical propensity of all amino acids, and developed a theory to describe the α-helix-random coil transition of polypeptides (**Table 1.4.1**). [82, 83] It was concluded from the values obtained that most amino acids oppose folding, where arginine and leucine residues are helix-indifferent and only alanine strongly favours helix formation. [83]

Residue	Helix Propensity	Residue	Helix Propensity
Ala	1.54	Cys	0.33
Arg^{+*}	1.04	Asn	0.29
Leu	0.92	Asp	0.29
Lys$^+$	0.78	Asp$^-$	0.29
Glu	0.63	Trp	0.29
Met	0.60	Phe	0.28
Gln	0.53	Val	0.22
Glu$^-$	0.43	Thr	0.13
Ile	0.42	His$^+$	0.06
Tyr	0.37	Gly	0.05
His	0.36	Pro	~ 0.001
Ser	0.36		

Table 1.4.1. *Helical propensity values of amino acids measured at 273 K; a higher value is indicative of the likelihood of this amino acid accruing in an α-helix. [83] (*) arginine is protonated and positively charged.*

The large helical propensity of alanine compared to glycine (which lies on the other end of the spectrum) is due to several reasons. Firstly, glycine is inherently more helix destabilising than alanine because glycine has greater conformational freedom in the unfolded state; assuming a helical fold is therefore more entropically unfavourable. Secondly, on folding from an extended conformation alanine buries more solvent-accessible hydrophobic surface area than glycine. Thirdly, the side chain of alanine can hinder the solvation of NH and C=O groups in the unfolded state, reducing the number of hydrogen bonds made with solvent, which is important for the *C*- and *N*-termini of helices. [80] Therefore, alanine generally imparts more stability in the middle positions of an α-helix because it buries more solvent-accessible area, while glycine is preferred at the termini as it allows less hindered solvation of the amine and carboxylate groups. This effect is most noticeable at the *N*-terminus because the side chain points backwards along the helix. [80] Amino acids such as glutamic acid, aspartic acid, threonine and serine are also preferred at the *N*-terminus (*N*-cap) because they act as hydrogen bond acceptors for a free NH_2 group, stabilising the helix. Similarly, residues that can act as hydrogen bond donors, such as arginine or lysine, are improve helix stability when placed at the *C*-terminus. [80, 84, 85]

Synthetic strategies aimed towards stabilising the helical structure include acylation or succinylation of the *N*-terminus, [84, 86] lactam bridging, [87] disulfide formation, [88, 89] multiple salt bridges, [90] hydrazone linkages and use of unnatural amino acids (**Figure 1.4.3**). [91, 92, 93]
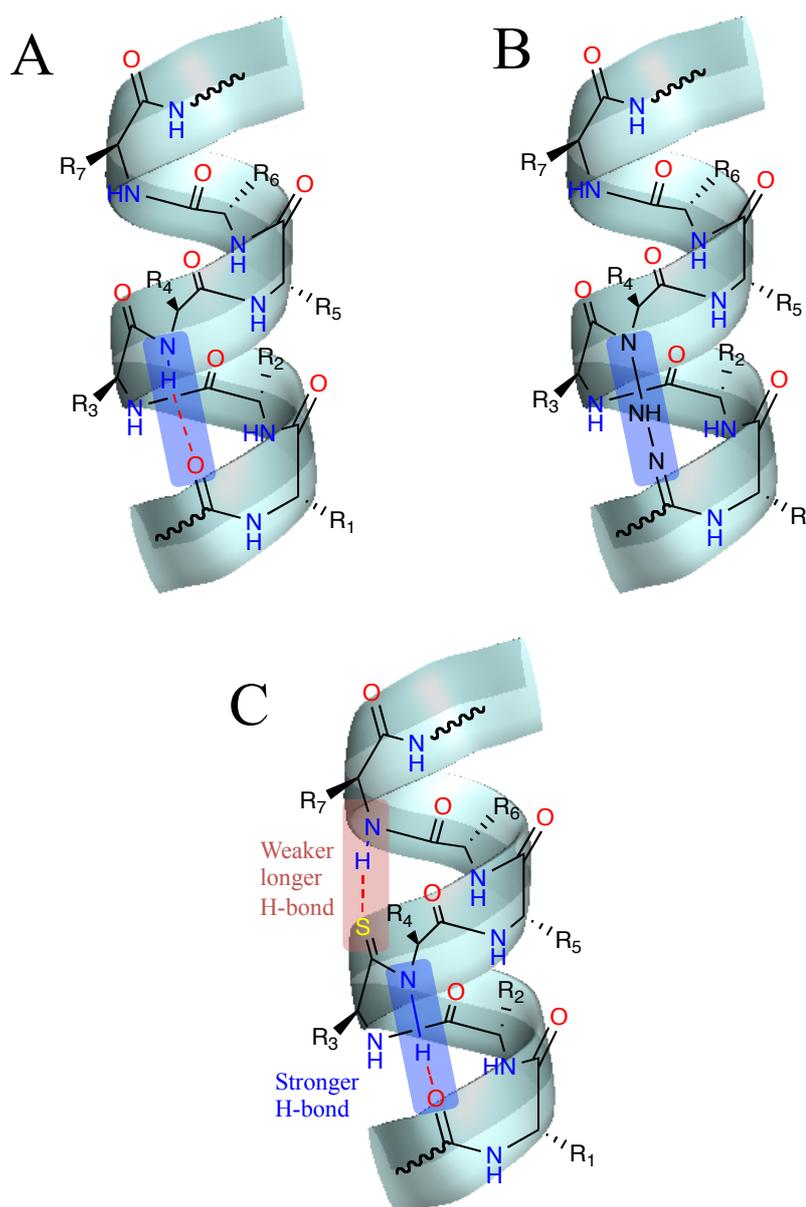
Figure 1.4.3. *Diagrams of the α-helical hydrogen bonding network and two different strategies of stabilising this structure. A) unmodified α-helix (H-bond interaction highlighted in blue), B) hydrazone bond (blue) and C) α-helix with a thioxo amino acid. Note that the H-bond formed by the thiol group (red) is weaker and longer, and the amine H-bond (blue) is stronger than the H-bonds usually formed in helices.*

Miwa *et al.* demonstrated that a thioxo amino acid can be used to stabilise the yeast transcription factor GCN4.[93] The thioxo amino acid was incorporated into a helix where the sulfur formed a longer and weaker hydrogen bond with the amine of the $i + 4$ residue while the amine of the thioxo amino acid formed a stronger hydrogen bond than

the natural occurring hydrogen bonds in α-helices (**Figure 1.4.3**). However, it was demonstrated that these linkages do not otherwise affect the formation of the helix, or the oligomerisation of the protein, but did endow peptides with higher thermal stability. [93]

Another method of using unnatural amino acid to stabilise the helical structure was demonstrated by Schafmeister *et al.*[94] A pair of α-methyl amino acids bearing olefinic side chains of varying length amino acids with *R* or *S* stereochemistry (**Figure 1.4.4**) were incorporated into peptides based on the C-peptide, the 13 *N*-terminal residues of RNAse A, at the *i* and either *i* + 4 or *i* + 7 positions and cyclised *via* olefin metathesis to cross-link one or two turns, respectively, of the helix. [94, 95, 96] Hydrocarbon stapling increased the helicity and resistance to proteases of the synthetic peptides.

A

B

n = 1, 2, 3, 4, 6   n = 1, 3, 4
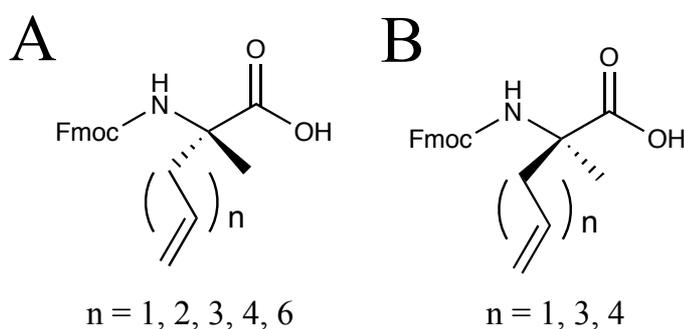
Figure 1.4.4. *Chemical structures of modified A) S- and B) R-amino acids with olefinic side chains of variable length.* [94]

Many investigations have sought to stabilise small helical peptides by pre-organising the helix to reduce the entropic cost of folding. [81, 97, 98, 99] One strategy to achieve helical stability in smaller peptides is to introduce the key functional residues (for example for DNA recognition and binding) into natural scaffolds with a well-defined and stable helical structure. Zondlo *et al.* presented an example of this use of natural scaffolds to create miniature proteins with novel activities using avian pancreatic polypeptide (aPP).

A small, well-folded protein consisting of a single α-helix stabilised by hydrophobic interaction with a type II polyproline helix, aPP was modified to contain residues used by the yeast transcription factor GCN4 to recognise specific DNA target site by the process known as protein grafting to contruct a 42 amino acid peptide that exhibited high DNA affinity and specificity at 4 °C (**Figure 1.4.5**). [100, 101] However, the activity was lost at room temperature, observed *via* electrophoretic mobility shift assay. [100] In order to overcome this temperature dependency, a phage library containing peptides with random mutations on the poly-proline helix was employed to select peptides active at room temperature. [101] The resulting p007 peptide demonstrated high affinity for the GCN4 cognate sequence at 25 °C. Further investigation of the 29 amino acids of p007 to the DNA affinity, specificity and secondary structure by alanine-scanning mutagenesis, it was concluded that all residues on the α-helix and most of the residues on the poly-proline helix were important for DNA-affinity and specificity. [102] The same system was also employed to synthesize a miniature homeodomain with 100 times increased affinity for its cognate DNA relative to the orginal recognition helix alone. [103]

Weston *et al.* designed an 18-residue miniature decarboxylate enzyme, Apoxaldie-1, based on a neurotoxic peptide apamin found in bee venom, in which a helical region is stabilised by two disulfide linkages. The helix displays three lysine residues on the solvent exposed side of an α-helix that are catalytically active (**Figure 1.4.6**). [104] Apoxaldie-1 demonstrated high stability to thermal denaturation and to high concentrations of the denaturant guanidinium chloride due to the presence of two disulfide linkages found in apamin. [105] The decarboxylation rate of this miniature enzyme was found to be approximately four orders of magnitude faster than *n*-butylamine. [104]
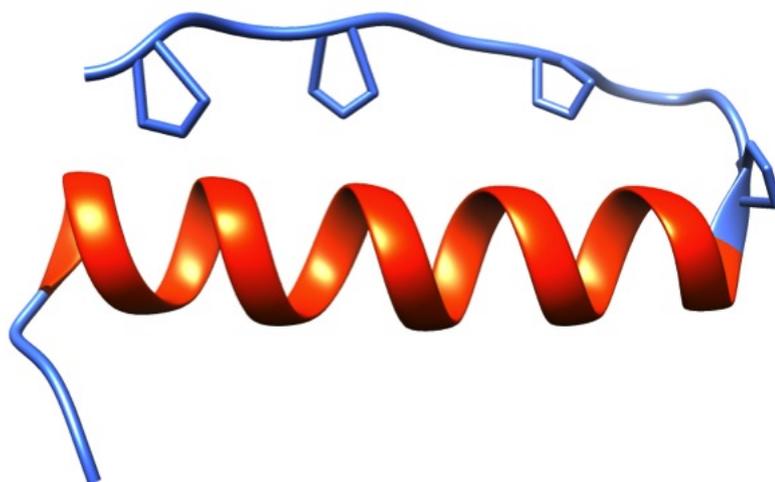
Figure 1.4.5. *X-ray diffraction structure of aPP (2bf9). The α-helix is shown in orange and the proline residues in combination with other residues adopt a proline-type II structure that is shown in blue (proline residues are shown as sticks).*
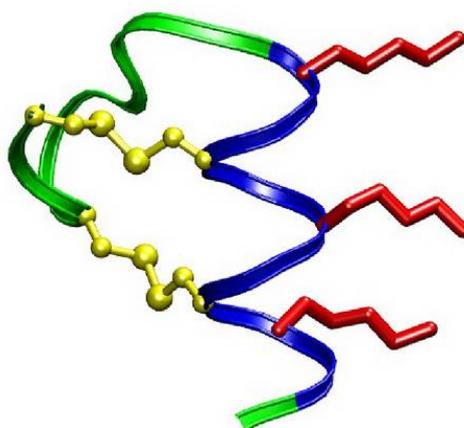
Figure 1.4.6. *Diagram of Apoxaldie-1.* [104] *The disulfide bonds are represented in yellow, the catalytic lysine side chains are in red and the α-helix is shown in blue.*

Talanian *et al.* have demonstrated that removal of the leucine zipper region from the bZIP protein GCN4 could be compensated for by using a covalent disulfide bond to effect dimerisation of the remaining basic domain. The resulting peptides proved capable of binding to DNA, with a dissociation constant of approximately 10 nM (*via* EMSA), which is of similar affinity and specificity to those of parent protein (**Figure 1.4.7**). [106]
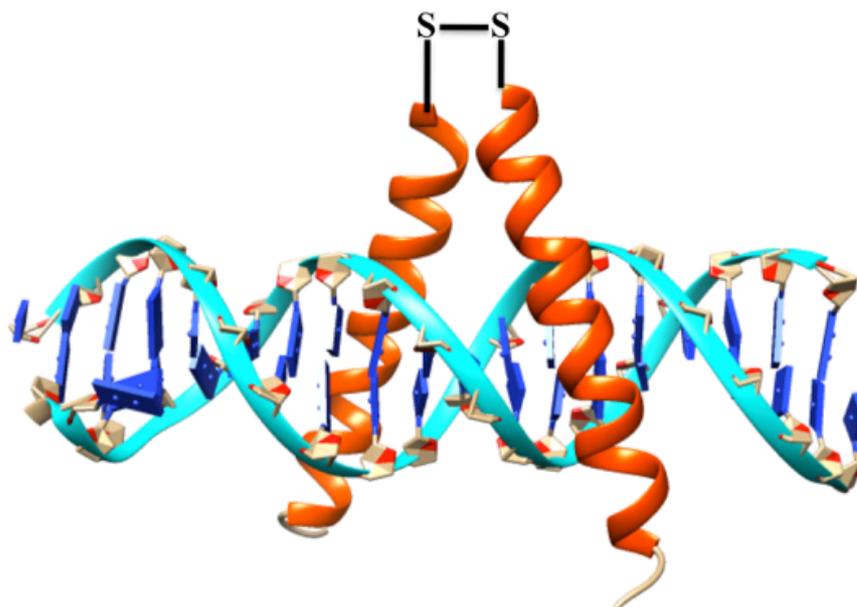
Figure 1.4.7. *Diagram of a substitution of the leucine zipper dimerisation motif by a disulfide bridge.* [106]

Helices may also be stabilised by inducing conformational contraints through the use of cross-linkers. For example, *para*-disubstituted benzene rings were used to link the side chains of residues in a peptide with an *i*, *i*+7 relationship. [107] Acetylenic crosslinkers of different lengths with varying degrees of flexibility (**Figure 1.4.8**) were used to modify lysine residues in either *i*, *i*+4 or *i*, *i*+7 spacings in order to stabilise α-helical structures in short peptides. Improved stabilization was confirmed by comparison to the same peptide cross-linked only with alkyl or oxyethylene skeletons. [108]
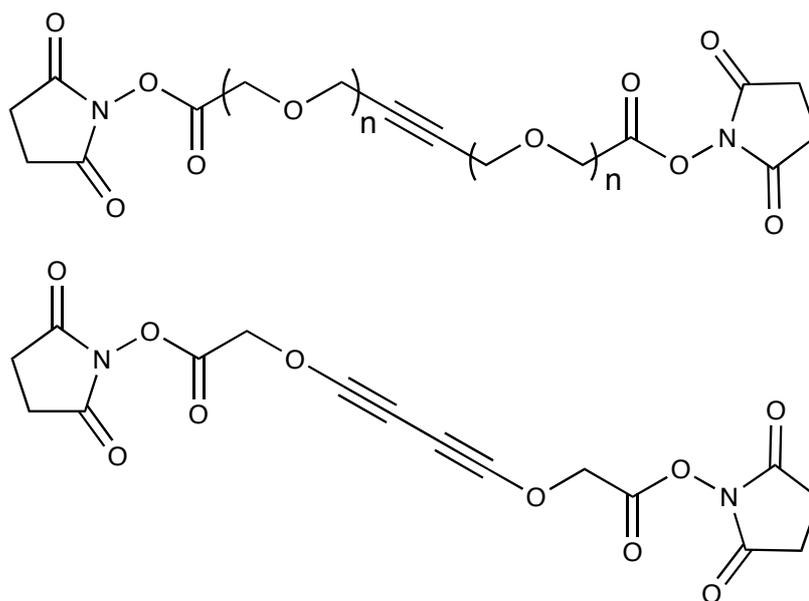
Figure 1.4.8. *Chemical structures of acetylenic crosslinking agents used to stabilise α-helices, where n corresponds to 1 or 2.* [108]

1.5 Photocontrol of Peptide Conformation

Rather than simply enforce secondary structures by the use of a covalently bound crosslinker, photocontrollable crosslinkers can facilitate reversible, light-dependent regulation of peptide or protein conformation. One such photocontrollable crosslinker is a spiropyran compound, which was introduced onto cysteine residues mutated into the sequence of the MscL channel protein found in *E. coli*. [109] Irradiation of the crosslinker with UV light (366 nm) led to photochemical ring opening process, with the formation of a charged zwitterionic merocyanine state (**Figure 1.5.1**).

This isomerisation occurring inside the narrow protein pore causes hydration of the dye and weakening of the hydrophobic interactions used to maintain the closed channel conformation, so that the pore opens. When the crosslinker was exposed to visible light, the compound returned to its closed ring, uncharged state, allowing reversible control of the channel conductivity upon an optical signal. However, the amount of merocyanine

formed on irradiation consistently dropped after the first open-close illumination cycle and the isomerisation process was slow. [109]

Spiropyran state   **Merocyanine state**
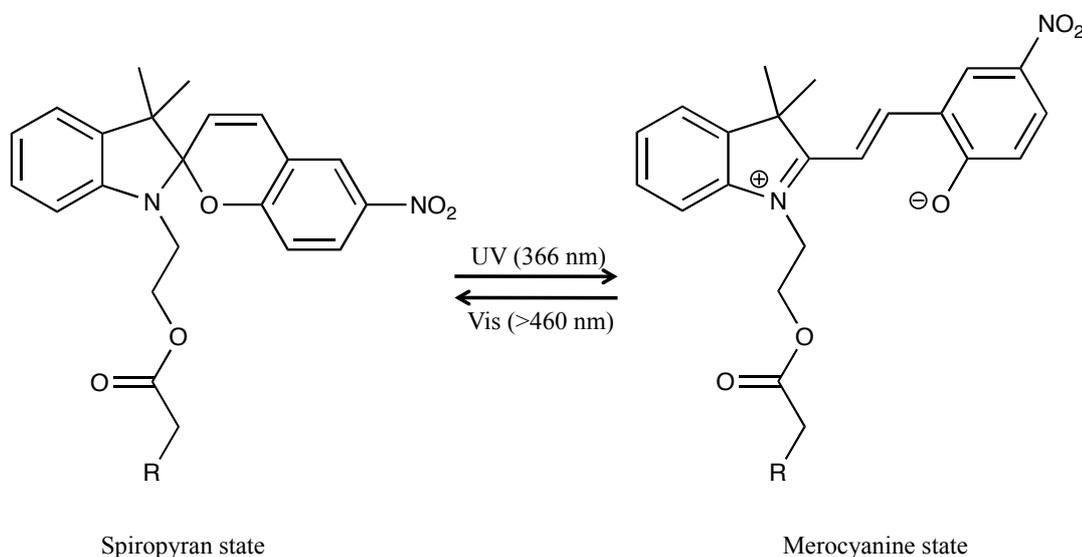
Figure 1.5.1. *Structures of the spiropyran and merocyanine states of the 3',3'-dimethyl-1'-(2-iodoacetyloxyethyl)-6-nitrospiro[2H-1-benzopyran-2,2'-indoline] photoswitch (R = iodine or MscL protein).* [109]

Azobenzene is the most widely used photoswitchable crosslinker, it can adopt two conformations, *cis* (a non-planar conformation with C_2 symmetry) and *trans* (a planar structure with C_{2h} symmetry). [110, 111] The *trans* conformation of azobenzene is 10-12 kcal mol^{-1} more stable than the *cis* isomer so that, at equilibrium in the dark, *trans* is the dominant isomer. [112, 113] The dark equilibrium state of azobenzene is >99% *trans* isomer (at 298K), whereas the amount of *cis* isomer that can be achieved upon irradiation is typically 70%-90% depending on the system. [114]

The azobenzene crosslinker undergoes a reversible photoisomerisation between both states, with *trans* to *cis* isomerisation occuring upon irradiation with UV light at 365 nm and a *cis* to *trans* isomerisation takes place at 420 nm or thermally (**Figure 1.5.2**). [110]
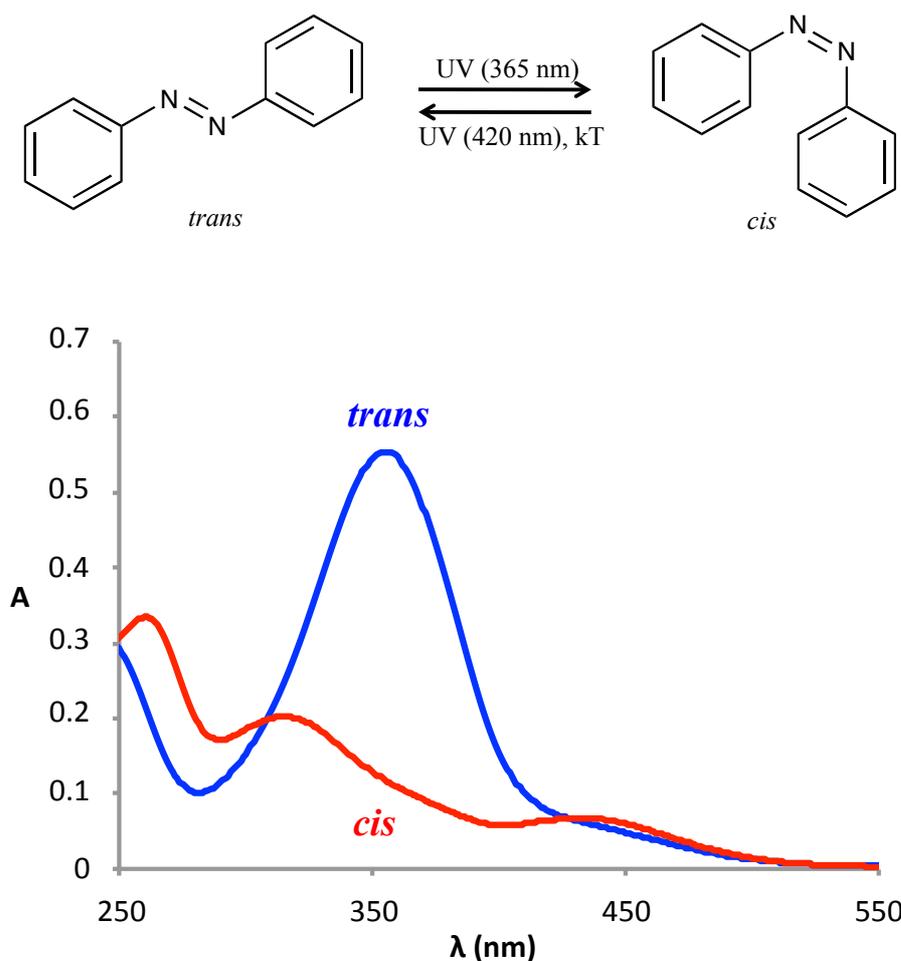
Figure 1.5.2. *(top) Isomerisation of the azobenzene and (bottom) electronic absorption spectra of the trans (blue) and cis (red) isomers of azobenzene.*

Four different pathways have been proposed as possible pathways for the photoisomerisation of the azobenzene crosslinker: rotation, inversion, concerted inversion, and inversion-assisted rotation (**Figure 1.5.3**). [111, 115, 110, 116] The rotational pathway involves loss of the N=N π-bond to allow free rotation about the N-N bond. Rotation changes the C-N-N-C dihedral angle while the N-N-C angle remains fixed at approximately 120°. [116] In the inversion mechanism, one N=N-C angle increases to 180° while the C-N=N-C dihedral angle remains fixed at 0°, resulting in a transition state with one sp hybridised azo-nitrogen atom. [117] For isomerisation to occur by concerted inversion, both N=N-C bond angles need to increase to 180° in order to generate a

linear transition state. In inversion-assisted rotation, large changes in the C-N=N-C dihedral angle and smaller but significant changes in the N=N-C angles occur simultaneously. [111] The transition state formed in concerted inversion has no net dipole moment, whereas the other three pathways possess polar transition states. Relaxation from all four transition states can afford either the *cis* or *trans* isomer; therefore, all four mechanisms predict photostationary states consisting of both isomers. [111]
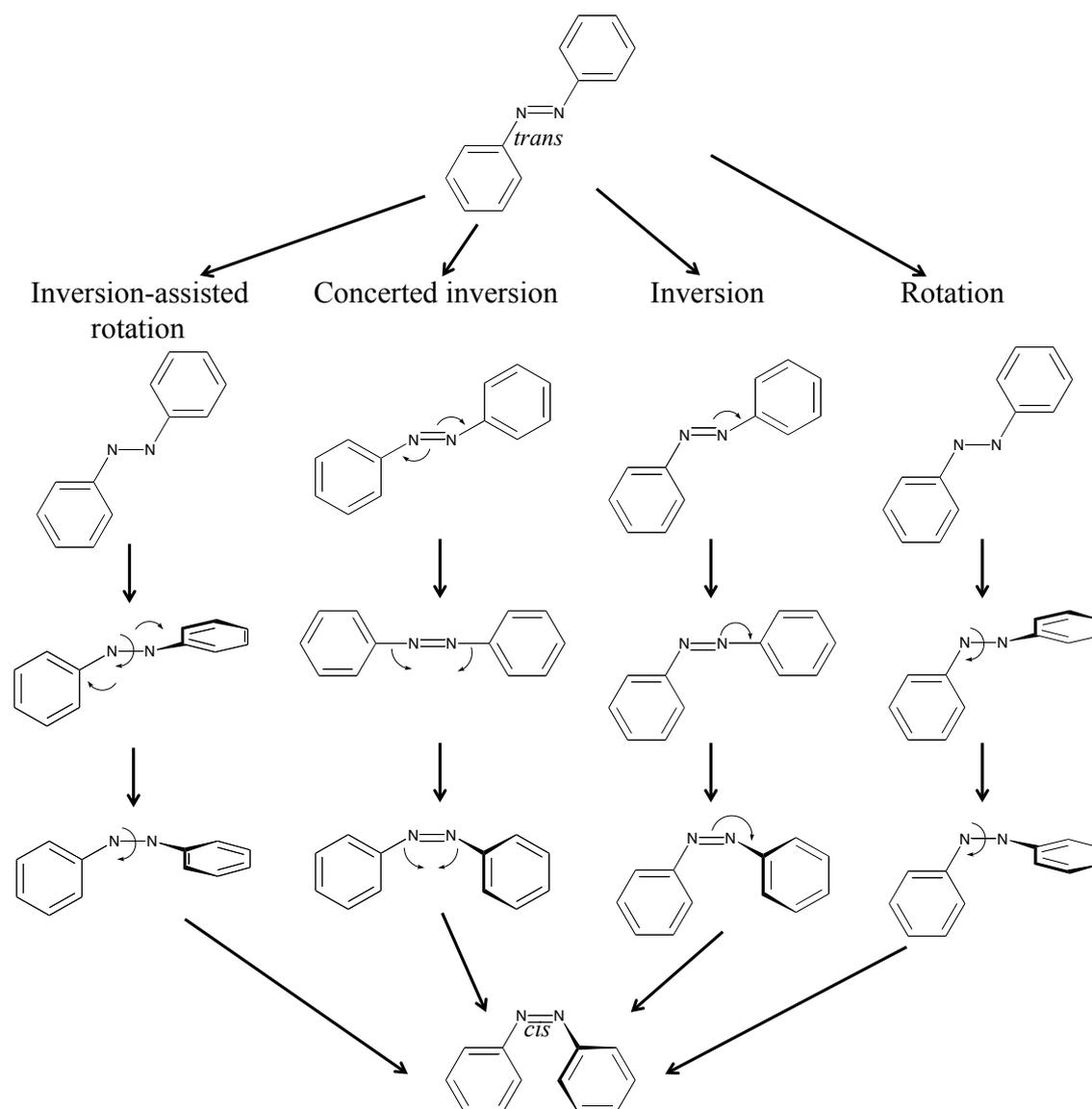
Figure 1.5.3. *Schematic diagram to show the suggestive pathways for trans-cis isomerisation of the azobenzene: Rotation, inversion, concerted inversion and inversion-assisted rotation.* [111]

Azobenzene-based photoswitches require UV light for photoisomerization, which can limit their application in biological systems, where UV light can trigger unwanted responses such as apoptosis. [118, 119, 120, 121] However, the presence of substituents on the phenyl rings of the azobenzene crosslinker may strongly influence the positions and shapes of azobenzene absorption bands. Different amino groups at the *ortho* or *para* positions can drastically red-shift the spectrum. [112] The incorporation of a donating group at one *para* position and an acceptor group at the other *para* position ('push-pull' azobenzenes) can lead to further red-shifts. [112, 122] This can be explained by seeing the excited state as having more dipolar character than the ground state; [123] substituents that can donate electron density inductively or *via* resonance will stabilise the excited state more than the ground state, resulting in red-shifts. [124] Beharry *et al.* have developed an azobenzene crosslinker that is photoswitched by green light (530-560 nm) by substituting all four *ortho* positions of a 4,4'-diamidoazobenzene core (**Figure 1.5.4. A)**) with methoxy groups (**Figure 1.5.4. B)**). [125] However, this modification increased sensitivity towards reduction under biomimetic conditions with a half-life of only 1.5 hours in a 1 mM reduced glutathione solution, limiting the potential for biological applications. [126] Substituents on the phenyl ring can also influence the thermal relaxation rate from *cis*-to-*trans*, which can be understood in terms of how the groups affect the degree of single bond character in the azo moiety. [127] Low pH can also increase the rate of thermal reversion to the *trans* isomer due to protonation at the azo moiety reducing the barrier to rotation. [128] The relaxation rates are also solvent dependent, with many substituents showing an increase in reversion rate with increasing polarity. [129, 130, 131, 132]
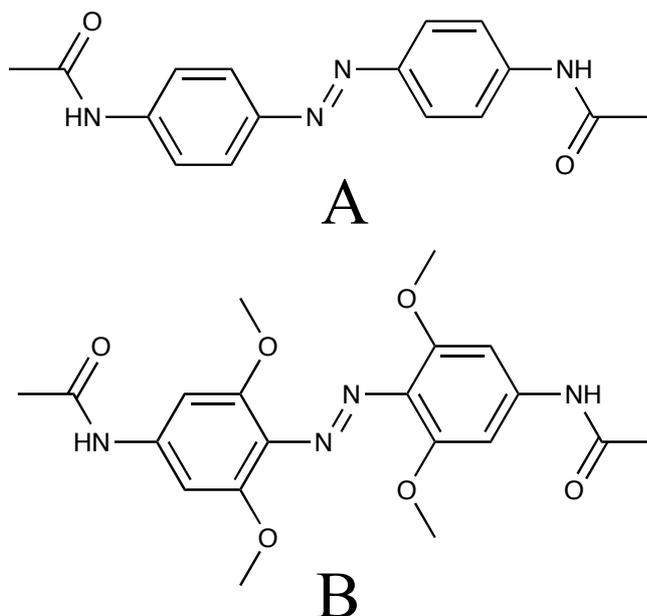
Figure 1.5.4. *A) Structure of parent 4,4'-diamidoazobenzene, B) structure of trans-2,2',6,6'-tetramethoxy-4,4'-diacetamidoazobenzene..The absorption of these compounds in the trans state are very similar to those of the parent azobenzene A. Irradiation of trans B with green light (530-560 nm) produced a mixture with approximately 80% fraction of the cis isomer, which is comparable to the fraction produced with the parent 4,4'-diamidoazobenzene.* [125]

Azobenzenes are useful photoswitches because their isomerisation is rapid (nanoseconds) and can result in a strong change in geometry upon switching. The end-to-end distance of each isomer is substantically different; the distance between the carbons at the *para* positions of the rings changes by approximately 3.5 Å. [133] Azobenzene possesses high photostability and repeated photoisomerisation does not degrade or destroy the compound. [134, 135] Whilst the structural changes induced by photoswitching depend on the magnitude of the end-to-end distance change of the azo compound, simply increasing the end-to-end distance of azobenzene may not necessarily increase the change upon isomerisation if the extension is too flexible. [127] For example, azobenzenes *para*-substituted with urea or carbamate groups may have longer end-to-end distances compared to 3,3'-bis-(sulfo)-4,4'-bis(chloroacetoamino)azobenzene crosslinker (discussed later), but their flexibilities

reduce the calculated net change in distance when alkylated onto peptide upon switching. [136] Isomers with overlapping end-to-end distance ranges will also reduce the degree of functional change observed, for example Beharry *et al.* have shown that minimal changes in helical content of a targeted peptide were observed for a dialkylamino-azobenzene unless the attachment points on the peptide were carefully optimised (**Figure 1.5.5**). [137]
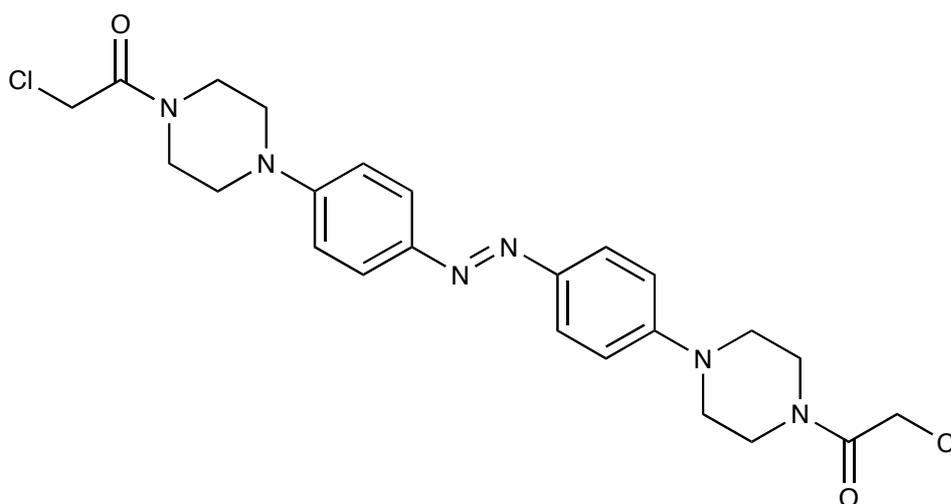
Figure 1.5.5. *Structure of the dialkylamino azobenzene crosslinker.* [137]

Azobenzene photoswitches can regulate the conformations of many different structural motifs. In proteins, β-hairpins are often identified as binding epitopes involved in protein-protein and protein-DNA interactions. However, β-hairpins are relatively unstable and difficult to characterise. Dong *et al.* incorporated an azobenzene crosslinker into a β-hairpin motif and were able to characterise the motif by NMR spectroscopy, demonstrating the successful folding of β-structures upon switching the azobenzene *trans* to *cis*. *Trans*-azo isomer exhibited a mostly disordered structure and *cis*-azo isomer led to the formation of the β-hairpin structure (**Figure 1.5.6**). [138]
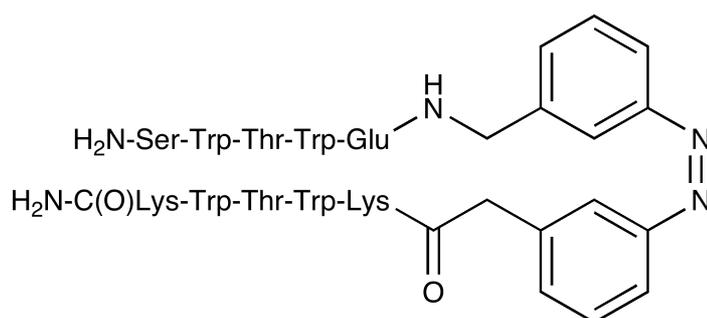
Figure 1.5.6. *Structure of the β-hairpin peptide with the azobenzene crosslinker.* [138]
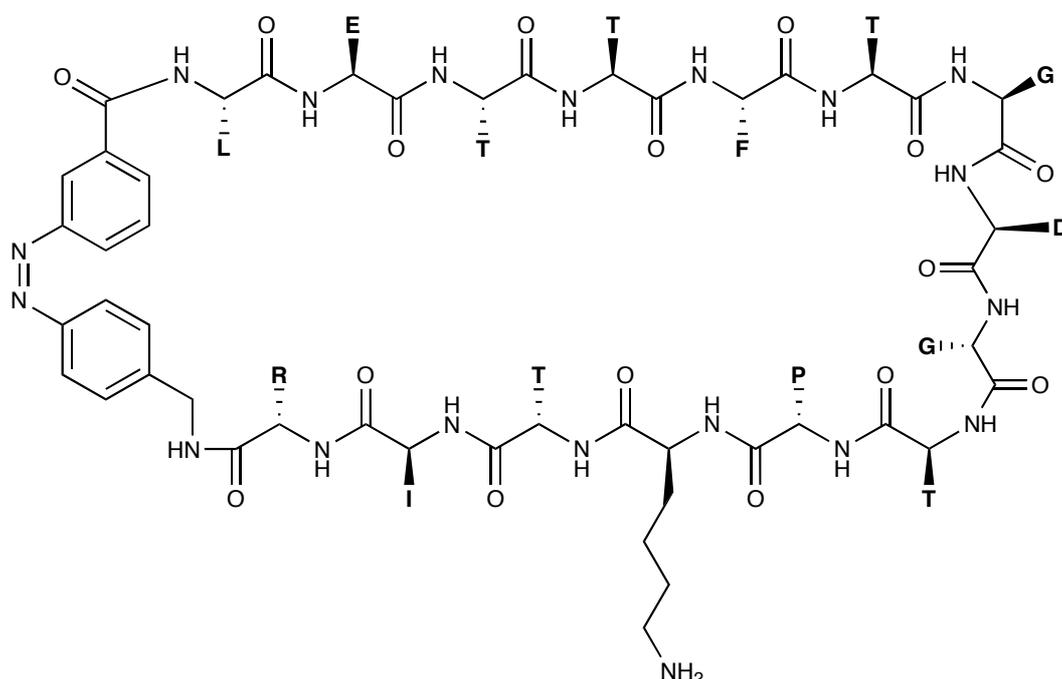
Figure 1.5.7. *Structure of synthetic cyclic peptide based on a β-finger peptide of neuronal nitric oxide synthase (nNOS) with amino acids glycine (G), leucine (L), Glutamic acid (E), threonine (T), phenyl alanine (F), aspartic acid (D), proline (P), isoleucine (I) and arginine (R) modified with an azobenzene crosslinker. The lysine side chain is for immobilisation on the sensor chip for binding studies using SPR techniques.* [139]

Photocontrol of a β-hairpin peptide was also demonstrated by Hoppmann *et al.*, who incorporated an azobenzene photoswitch into a synthetic cyclic peptide related to a β-finger peptide of neuronal nitric oxide synthase (nNOS) which mediates the formation of a variety of multi protein complexes in the cell (**Figure 1.5.7**). [139, 140] nNOS interacts with syntrophin PDZ domain to mediate the production of nitric oxide to coordinate

muscle contraction.[141] The binding of this crosslinked nNOS to target syntrophin PDZ was observed to change upon photoisomerisation. [139]

Woolley *et al*. have developed an extensive family of azobenzene crosslinkers, which can be attached *via* peptide side-chains rather than the backbone and which feature other substituents to modulate optical or physical properties. [142, 143, 144, 145] They all contain α-haloacetamido groups, allowing selective reaction with cysteine residue side chains in the presence of other amino acids *via* an S_N2 reaction. The switches are symmetrical so a single species is formed upon crosslinking. The most commonly used member of the family (**Figure 1.5.8**) features a minimum number of single bonds between the azo moiety and the peptide backbone to enhance coupling of the change in end-to-end distance of the switch to the peptide backbone and sulfonate groups to increase water solubility. [142, 146]
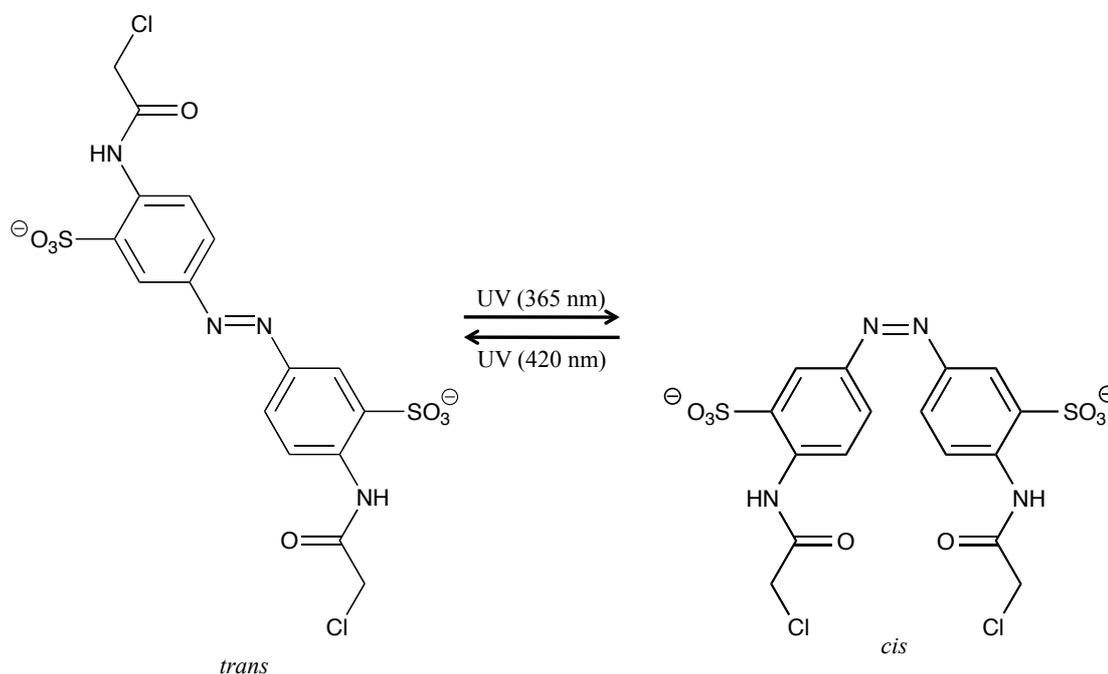
Figure 1.5.8. *The photoisomerisation of Woolley et al. designed 3,3'-bis-(sulfo)-4,4'-bis(chloroacetoamino)azobenzene (BSBCA) crosslinker.*

The ability to react with cysteine side chains allows this family of crosslinkers to be used to control α-helicity. [12416] Molecular modeling suggested that in order to stabilise the helical motif at *trans* conformation, an *i, i* +11 spacing of cysteine residues inserted into a sequence is the most appropriate to match the *trans* configuration crosslinker length of approximately 16.8 Å. To match the *cis* conformation length of approximately 11.3 Å to a helix the cysteines need to be in the *i, i* +4 or *i, i* +7 spacings (**Figure 1.5.9**).[114, 147]

Time-resolved optical rotatory dispersion experiments on a crosslinked peptide demonstrated that the complex undergoes subnanosecond photoisomerisation to the *cis* form and re-isomerises on time scale of minutes. [148] Time-resolved IR spectroscopy experiments were used to examine the folding and unfolding of a photoswitchable 16-residue alanine-based α-helical peptide and refined the isomerisation timescale to picoseconds with folding and unfolding occurring on a 100 ns to μs timescale. [149, 150, 151] This is interpreted as the crosslinker tilting the energy landscape without altering the intrinsic conformational dynamics of the peptide, [127] an approach generally applicable to a variety of peptides and proteins. [152] Zhang *et al.* introduced an azobenzene crosslinker to the SH3 domain from Fyn-tyrosine kinase (FynSH3) and demonstrated that the energy difference between the *trans* and *cis* isomers of the crosslinker (approximately 10-12 kcal mol^{-1}) was greater than the folding free energy of uncrosslinked FynSH3, suggesting that photoisomerisation should be sufficient to drive the folding/unfolding of other globular proteins. [153]
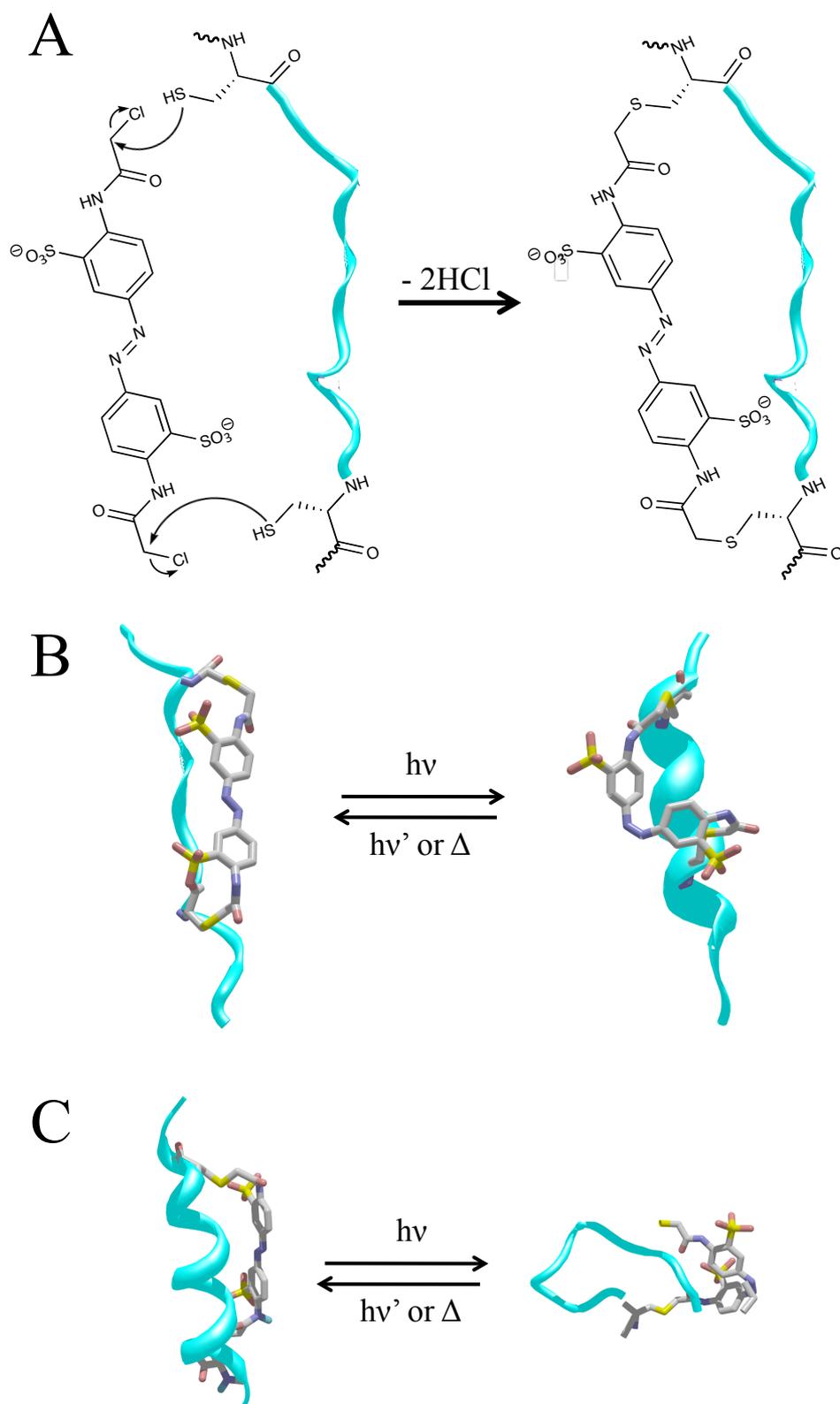
Figure 1.5.9. *A) crosslinking reaction of azobenzene and peptide, B) isomerisation scheme of i, i+7 spacings, C) isomerisation scheme of i, i+11 spacings.*

By photocontrolling helix formation in small peptides, it is possible to control the conformation and binding of helical elements of transcription factors' DBDs. Guerrero *et al*. reported that introducing an azobenzene crosslinker into an 18-residue polypeptide based on the recognition helix of the Q50K engrailed homeodomain in an *i*, *i* +11 spacing caused the peptide to be more helical and have greater DNA binding affinity when the crosslinker is in the *trans* configuration. [154] Upon irradiating the crosslinked HDH-3 peptide, the *cis* conformation disfavoured the helical motif, reducing binding of the peptide to DNA. Guerrero *et al*. also demonstrated the ability to regulate the binding specificity of the muscle-specific transcription factor MyoD. [155] As previously discussed, MyoD belongs to a family of transcription factors that rely on a basic helix-loop-helix domain for DNA binding. By substituting Met 116 and Ser 123 residues located on the water-exposed face of the recognition helix of MoyD with cysteines in an *i*, *i* +7 spacing, it was possible to introduce a BSBCA crosslinker, thus enabling the MyoD DNA recognition domain to be photoswitchable. [155] Ultimately, the binding control of this photoswitchable MyoD to target DNA was achieved upon isomerization. Woolley *et al*. demonstrated reversible control of a modified GCN4-bZIP protein using this crosslinker binding to its target DNA. [156] Alkylating the azobenzene on the cysteines spaced *i*, *i*+7 residues apart and away from the dimer interface resulted in the *cis* isomer enhancing the degree of helicity of the coiled-coil region. [156] This work was extended to manipulate the binding affinity of the AP-1 transcription factor Fos *in vivo*. [157] The crosslinker was again attached to *i*, *i*+7 spaced cysteine residues introduced on the face of the helix away from the heterodimerisation interface. Upon photoactivation, the helicity of this crosslinked AFosW peptide is enhanced, leading to the formation of coiled-coil Fos-Jun heterodimers. This in turn reduces the population of Jun-Jun homodimers and reduces gene expression. [157] Crosslinked XAFosW reduced binding of

AP-1 to DNA up to 10-fold after UV irradiation *in vitro*. [157] The *in vivo* activity of the photoswitchable dominant negative XAFosW to AP-1 was carried out in 293T cells. Luciferase reporter (AP-1-Luc) under the control of an AP-1 promoter together with a β-galactosidase reporter under the control of a constitutive rous sarcoma virus (RSV) promoter (pRSV-Gal) was used in order to permit quantitative analysis. [157] Ratios of luciferase activity to β-galactosidase activity were used as an indication of specific inhibition of AP-1 activity. In order to minimise possible photodamage to cells, the isomerisation of *trans* XAFosW to *cis* XAFosW was done by irradiating the cells with a 365 nm light-emitting diode in short bursts so that at the end of the incubation period, at least 50% of the linker moieties were expected to be present in the *cis* form in the irradiated cells. [157] When non-crosslinked AFosW was used, the activity of AP-1 was inhibited in a light-independent manner. However, when XAFosW was used, it was observed that the activity was light-dependent, which resulted in a decrease in AP-1 activity upon photoisomerisation by as much as approximately 40%. [157] The effectiveness of XAFosW in the photocontrol of AP-1 activity in living cells was also confirmed by the use of AP-1 GFP reporter in combination with flow cytometry measurements confirming that *cis* XAFosW is a more potent AP-1 inhibitor than its *trans* isomer. [157] *Mart et al.* developed photoswitchable RNA binding peptides based on Rev-RRE complex, a regulator of virion protein (Rev), which interacts with the stem-loop IIB of the Rev response element (RRE) of HIV type 1. [158] Rev binds to RRE containing mRNA, removing it from the nucleus, thus allowing expression of unspliced or singly spliced mRNA into different protein products. This is an important interaction as Rev-RRE interaction controls the infection cycle of HIV type 1, as the resulting products are needed for virus particle assembly. [158] Two Rev-based peptides (residues 33-50 of Rev) $Rev^{i,i+7}$ and $Rev^{i,\,i+11}$ were designed by replacing two appropriately spaced

amino acids (*i, i*+7 and *i, i*+11 spacings) located on the solvent-exposed face of Rev α-helix and opposite the residues involved with RNA bindings, with cysteine residues in order to accommodate the BSBCA photoswitchable crosslinker (Rev$^{i,i+7}$-XL-SO$_3$H and Rev$^{i, i+11}$-XL-SO$_3$H). [158] Upon photoisomerisation, slower relaxation rate of Rev$^{i,i+7}$-XL-SO$_3$H from *cis* to *trans* (UV-visible spectroscopy, 299 min) compared to Rev$^{i, i+11}$-XL-SO$_3$H (135 min) was observed. CD spectroscopy of the peptides at dark-adapted state (dad) also demonstrated that Rev$^{i, i+11}$-XL-SO$_3$H adopted an α-helical conformation, whilst Rev$^{i,i+7}$-XL-SO$_3$H adopted a less α-helical structure in dark-adapted state, but was mostly α-helical after photoisomerisation with 363 nm light. In order to perform binding studies, the peptides were labelled with fluoresceinamide (FAM) on their N-termini, allowing FRET measurements of RNA binding affinities from the peptides to a RNA molecule labelled with cyanine dye CY3 acceptor fluorophore. [158] However binding studies of Rev$^{i, i+11}$-XL-SO$_3$H and Rev$^{i,i+7}$-XL-SO$_3$H could not be performed due to the low solubility of the fluorescently labelled and alkylated peptides. Hence unlabelled Rev$^{i, i+11}$-XL-SO$_3$H and Rev$^{i,i+7}$-XL-SO$_3$H were instead used to displace FAM Rev$_{wt}$ from a preformed complex with 5'-Cy3-labelled RRE. *trans* Rev$^{i, i+11}$-XL-SO$_3$H demonstrated lower binding affinity to RRE (247 nM) than the Rev$_{wt}$-RRE complex (5.8 nM), whereas no binding as observed for Rev$^{i,i+7}$-XL-SO$_3$H in either *trans* or *cis* conformations. Molecular dynamics simulations revealed that the sulfonate groups disturbed the α-helical structure of Rev$^{i, i+11}$-XL-SO$_3$H for all conformers of the crosslinker in a way that resulted in the bending of the peptide backbone. The presence of sulfonate groups on the crosslinker seems to cause steric hindrance with the phosphate backbone at the rim of the binding pocket. Therefore a non-sulfonated BSBCA was used, resulting in the production of Rev$^{i,i+7}$-XL-H and Rev$^{i, i+11}$-XL-H peptides. [158] *trans* Rev$^{i, i+11}$-XL-H demonstrated improved affinity for RRE (24 nM)

whereas *cis* Rev$^{i,\ i+11}$-XL-H demonstrated a decrease in affinity (66 nM). The dissociation constant (K$_D$) *trans* Rev$^{i,\ i+7}$-XL-H-RRE complex was obtained (1 μM), whereas *cis* Rev$^{i,\ i+7}$-XL-H-RRE complex demonstrated higher binding affinity (607 nM).

Wysoczanski *et al*. reported a NMR structure of a photoswitchable pro-apoptotic Bak-based peptide in complex with pro-surviving protein Bcl$_{XL}$. [159] The photoswitchable peptide was based on the sequence of a 16 residues Bak wild type (Bak$_{wt}$) with the I81F substitution known to increase the affinity for Bcl$_{XL}$. [160] The BSBCA photoswitchable crosslinker was introduced to two cysteine residues in an *i*, *i*+11 spacing, on the face of the helix predicted to orient away from the binding region, generating Bak$_{I181F}$$^{i,\ i+11}$-XL. [159] NMR studies have demonstrated that upon binding to the surface groove of Bcl$_{XL}$, Bak$_{I181F}$$^{i,\ i+11}$-XL is coiled into a slightly distorted helix and the binding appeared to be driven almost exclusively by hydrophobic interactions. Despite the fact that the presence of the crosslinker on Bak$_{I181F}$$^{i,\ i+11}$-XL had caused distortion on the complex, forcing the hydrophobic side chains out of the position adaopted by Bak$_{wt}$, the binding site of Bcl$_{XL}$ demonstrated a large degree of flexibility by remodelling itself to make favourable interactions with the crosslinker peptide.

To date, most applications of azobenzene photoswitches to controlling biomolecular functions have been carried *in vitro*. The ultimate aim would be taking this a step further in achieving photocontrol of biological function *in vivo*. There are two possible ways of achieving this: firstly by introducing a biomolecule modified with a photoswitch to the living system by for example, microinjection or using cell-penetrating peptide sequences, [157, 161] secondly where the target biomolecule is selectively modified *in vivo*.

In order to function *in vivo*, the azobenzene photoswitch must be chemically stable in reducing intracellular environments. [122] The intracellular redox potentials is maintained by the tripeptide glutathione present in its reduced form at approximately 1-10 mM concentration. [162] There have been indications that other azobenzene switches will be stable to intracellular reducing conditions. [119, 143] Sadovski *et al*. and Beharry *et al*. have demonstrated that diaminoazobenzene and BSBCA do not seem to reduce when incubated overnight with high concentrations of glutathione under *in vitro* physiological conditions. [119, 137, 163] Zhang *et al*. also reported a modified photoswitchable versions of cyclosporine A (an immunosuppressive drug) containing diamidoazobenzene-based switches that demonstrated resistance to glutathione reduction, which could also photoisomerise in human whole blood lysates. [164]

Bose *et al.* have also demonstrated that it was possible to incorporate a photoswitchable amino acid into proteins in *E. coli*. [165] Nonsense codon (amber TAG codon) suppression was used to introduce phenylalanine-4'-azobenzene (AzoPhe) into the *E. coli* catabolite activator protein (CAP) which appeared stable *in vivo* during the expression period (**Figure 1.5.10**). [165] UV-visible spectrum of the mutant CAP protein showed a distinct absorbance peak at 334 nm, corresponding to *trans* isomer of the AzoPhe. Irradiation of the mutant CAP with 334 nm light led to a reduction in the 334 nm peak and an increased absorbance at 420 nm. [165] The mutant CAP was finally irradiated with 420 nm, resulting in complete conversion back to the 334 nm band, suggesting the successful incorporation of AzoPhe into the CAP protein. Since the binding of cyclic adenosine monophosphate (cAMP) to CAP results in conformational changes in the protein that increase its binding affinity to its promoter, resulting in enhanced transcription from CAP-dependent promoters, binding studies of mutant CAP were performed. It was postulated that *trans* and *cis* isomers of AzoPhe would differentially

affect the binding affinity of cAMP to CAP and, as a consequence, the affinity of CAP for its lac promoter. In the *trans* configuration, the mutant CAP demonstrated lower binding affinity to its promoter by approximately 4 fold compared to wild type CAP in the presence of cAMP (EMSA). Upon photoswitching, the *cis* mutant CAP demonstrated even lower binding affinity by approximately 4 fold compared to the *trans* mutant CAP, suggesting that upon isomerisation, mutant CAP/cAMP interaction is reduced, resulting in lower binding affinity to target promoter. [165]
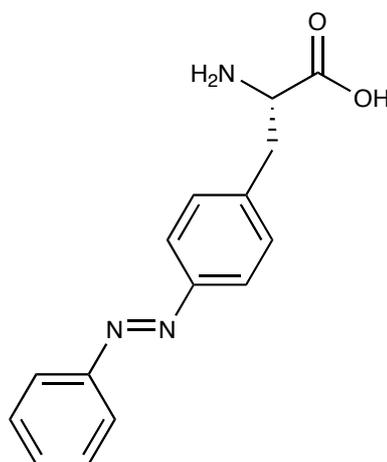
Figure 1.5.10. *Structure of phenylalanine-4'-azobenzene.*

Beharry *et al.* created a fluorescent peptide capable of reporting the state of an attached azobenzene crosslinker though fluoresence quenching (**Figure 1.5.11**). [163] The peptide was synthesised with all D-amino acids to make it resistant to proteolytic degradation and was microinjected into zebra fish embryos. Photoswitching could be detected throughout the organism for at least 2 days. It was found that the switching rates and quantum yields were similar to those found *in vitro*, although the magnitude of the change in fluorescence was diminished (from 40% to 20%) and the relaxation rate to the *trans* form was accelerated. Nevertheless, these findings offer evidence to suggest that

appropriately chosen azobenzene derivatives will be practical tools for spatio-temporally probing processes *in vivo*. [127]
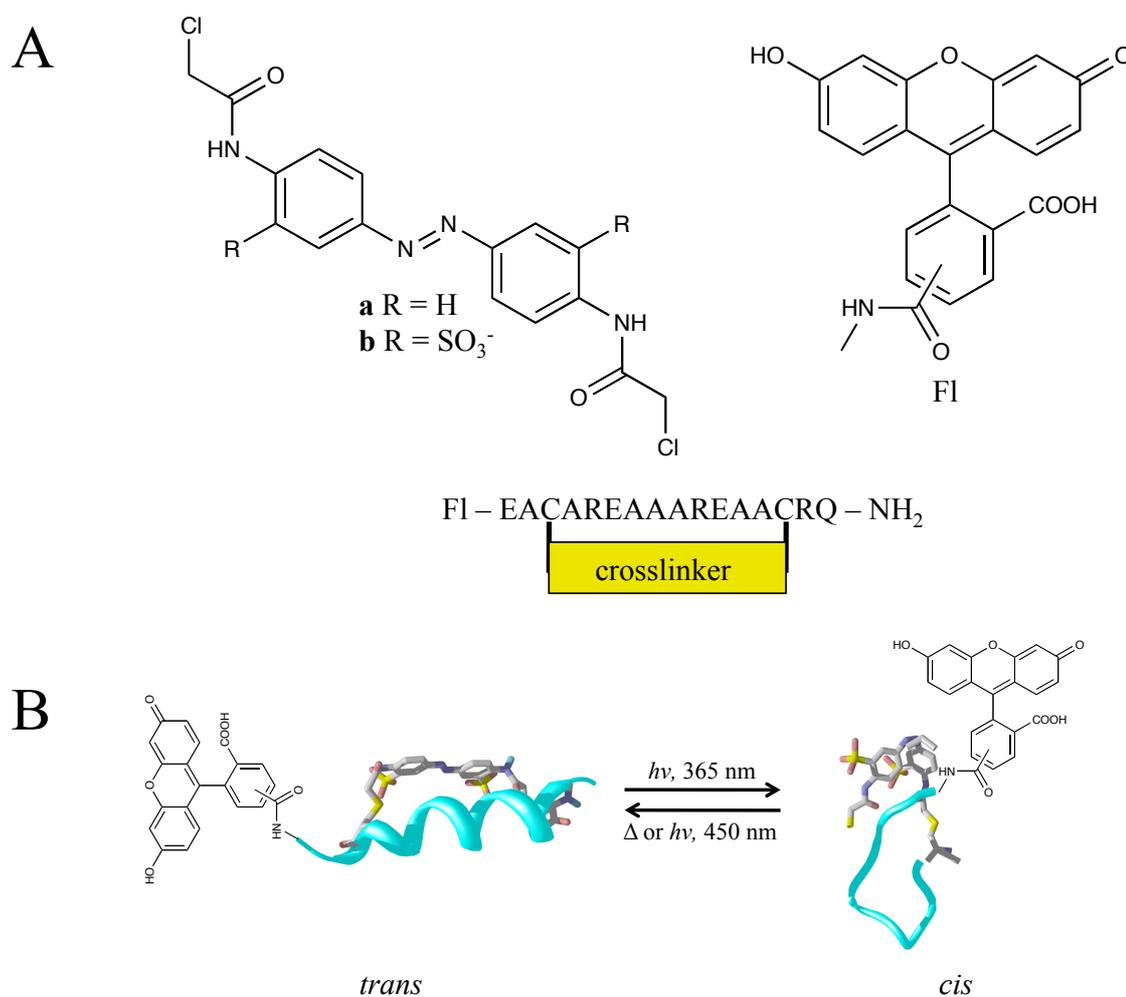
Figure 1.5.11. *A) Structure of azobenzene photoswitch and the fluorescent peptide reporter. B) trans-to-cis isomerisation of the fluorescent peptide reporter. [163] Irradiation of the crosslinked peptide at 365 nm causes trans-to-cis isomerisation of the photoswitch as well as fluorescein excitation. Irradiation with blue light (440-490 nm) causes cis-to-trans isomerisation as well as fluorescein excitation. Since the cis isomer of the crosslinked peptide has a lower quantum yield for fluorescence than the trans isomer, irradiation of a dark-adapted solution of reporter peptide with UV light produces a time-dependent decrease in fluorescence. Irradiation of cis peptide with blue light causes a time-dependent increase in fluorescence. The rates at which the switching processes is dependent on the intensity and the wavelength of irradiation. [163]*

1.6 Project Aims

The aim of the project was to develop novel miniature photoswitchable artificial transcription factors (PATFs) for the modulation of protein-DNA interactions, and to demonstrate the ability to control the transcription process using these PATFs in an *in vitro* environment. The DNA recognition helix of a homeodomain transcription factor and an amphipathic α-helix activating domain (AH) were selected as a model system due to the high level of photocontrol in previous studies and the potency at initiating the transcription process of the activating domain. [151, 166]

The DNA recognition helix was modified at different residue spacings with azobenzene cross-linkers and linkers regions of different lengths and flexibilities were use to connect the DNA recognition helix and the activating domains modules. The effects of these changes on the resulting variants structure and function were investigated.

Chapter 2

Materials and Methods
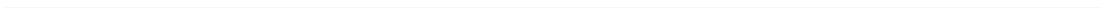
2.1 Materials

All chemicals and oligonucleotides were purchased from Fisher, Sigma-Aldrich, New England Biolabs, Eurofins, Promega or Perkin Elmer.

2.1.1 Culture Media

2.1.1.1 Luria Bertani (LB) medium

Bacto tryptone (10 g l^{-1}), yeast extract (5 g l^{-1}) and NaCl (10 g l^{-1}) were dissolved in deionised water (1 l) and the medium was autoclaved for 1 hour and 30 minutes at 15 lb/in^2.

2.1.1.2 Terrific Broth (TB) medium

Bacto typtone (12 g), yeast extract (24 g) and glycerol (4 ml) were dissolved in deionised water (1 l) and the medium was autoclaved for 1 hour and 30 minutes at 15 lb/in^2 on a liquid cycle. Prior to use, 100 ml of a sterile filtered solution of 0.17 M KH_2PO_4 and 0.72 M K_2HPO_4 were added.

2.1.2 Agar Plates

Agar plates were prepared with autoclaved LB media containing agar (15 g l^{-1}) and the appropriate concentration of antibiotic (eg. ampicillin 100 mg l^{-1}) when required. The solution was poured into 90 mm Petri dishes using aseptic techniques. Once set, the agar plates were stored inverted at 4 °C. When needed, the solution of bacterial culture was spread over the plate with a sterile glass rod or loop and incubated at 37 °C overnight.

2.1.3 *E. coli* Strains

The *E. coli* expression strains used in this work were: BL21(DE3), BL21(DE3)RIL, BL21(DE3)RP, BL21(DE3)* and XL1-Blue.

Strain	Genotype
BL21(DE3)	F⁻ *omp*T *hsd*S$_B$ (r$_B$⁻ m$_B$⁻) *gal dcm* (DE3)
BL21-CodonPlus-(DE3)RIL	F⁻ *omp*T *hsd* S$_B$ (r$_B$⁻ m$_B$⁻) *dcm*+ Tet^r *E.coli* gal λ (DE3) *end*A Hte [*argU ileY leuW* Cam^r]
BL21-CodonPlus-(DE3)RP	F⁻ *omp*T *hsd* S$_B$ (r$_B$⁻ m$_B$⁻) *dcm*+ Tet^r gal λ (DE3) *end*A Hte [*argU proL* Cam^r]
BL21(DE3)*	F⁻ *omp*T *hsd*S$_B$ (r$_B$⁻ m$_B$⁻) *gal dcm rne*131 (DE3)
XL1-Blue	*end*A1 *rec*A1 *gyr*A96 *thi*-1 *hsd*R17 *sup*E44 *rel*A1 *lac* [F' *pro*AB *lacI^q*ZΔM15 Tn10 (Tet^r)]

Table 2.1.3. *Genotype of the E.coli strains used in this work.*

The BL21 strains are designed for protein expression. The DE3 designation denotes that the strain contains the λDE3 lysogen carrying the gene for T7 RNA polymerase under control of the *lac*UV5 promoter allowing induction of the expression of recombinant proteins with isopropyl-1-thio-β-D-galactopyranoside (IPTG). The BL21-CodonPlus-RIL strains contain extra copies of the *argU, ileY,* and *leuW* tRNA genes to allow higher expression level of heterologous proteins with high amount of AGA, AGG, AUA and CUA codons in the gene sequence. The BL21-CodonPlus-RP strains contain extra copies of the *argU* and *proL* tRNA genes to allow higher expression level of heterologous proteins with high amount of AGA, AGG and CCC codons in the gene sequence. The BL21* strains contain a mutation in the gene encoding RNase E, an

enzyme that degrades mRNA, so a higher level of mRNA is available for protein translation. The XL1-Blue strain is a host for optimal propagation of plasmids. This strain is used for cloning where high transformation efficiencies are required.

2.1.4 Oligonucleotide sequences for PATF peptides

Oligonucleotide sequences used in the cloning of PATF peptides are as follow:

PATF-1 *for1*: 5'-
GAAGCGCAGTGTAAAATCGCGGCTAAGAACGCACGTGCGAAATGCAAGAAAGCTGGATCCG
GTGGCACTGGTGGCGGTTCTCCGGAATTCCCGGGTAT-3'

PATF-1 *for2*: 5'-TGAGCTGCAAGAACTGCAAGAGCTGCAGGCCCTCCTGCAGCAAATG-3'

PATF-1 *rev1*: 5'-
TTGCTGCAGGAGGGCCTGCAGCTCTTGCAGTTCTTGCAGCTCAATACCCGGGAATTCCGGAG
AACCGCCACCAGTGCCACCGGATCCAGCTTTCT-3'

PATF-1 *rev2*: 5'-TGCATTTCGCACGTGCGTTCTTAGCCGCGATTTTACACTGCGCTTCCAT-3'

PATF-2 *for1*: 5'-
GAAGCGCAGTGTAAAATCGCGGCTAAGAACGCACGTGCGAAATGCAAGAAAGCTCCGCCGC
CACCGCCTCCGCCACCGCCTCCGCCACCGCCGGAATTCCCGGGTAT-3'

PATF-2 *for2*: 5'-TGAGCTGCAAGAACTGCAAGAGCTGCAGGCCCTCCTGCAGCAAATG-3'

PATF-2 *rev1*: 5'-
TTGCTGCAGGAGGGCCTGCAGCTCTTGCAGTTCTTGCAGCTCAATACCCGGGAATTCCGGCG
GTGGCGGAGGCGGTGGCGGAGGCGGTGGCGGCGGAGCTTTCT-3'

PATF-2 *rev2*: 5'-TGCATTTCGCACGTGCGTTCTTAGCCGCGATTTTACACTGCGCTTCCAT-3'

PATF-3 *for1*: 5'-
GAAGCGCAGTGTAAAATCGCGGCTAAGAACTGCCGTGCGAAAATTAAGAAAGCTCCGCCGC
CACCGCCTCCGCCACCGCCTCCGCCACCGCCGGAATTCCCGGGTAT-3'

PATF-3 *for2*: 5'-TGAGCTGCAAGAACTGCAAGAGCTGCAGGCCCTCCTGCAGCAAATG-3'

PATF-3 *rev1*: 5'-
TTGCTGCAGGAGGGCCTGCAGCTCTTGCAGTTCTTGCAGCTCAATACCCGGGAATTCCGGCG
GTGGCGGAGGCGGTGGCGGAGGCGGTGGCGGCGGAGCTTTCT-3'

PATF-3 *rev2*: 5'-TAATTTTCGCACGGCAGTTCTTAGCCGCGATTTTACACTGCGCTTCCAT-3'

PATF-4 *for1*: 5'-
GAAGCGCAGATCAAAATCTGTGCTAAGAACTGCCGTGCGAAAATTAAGAAAGCTCCGCCGC
CACCGCCTCCGCCACCGCCTCCGCCACCGCCGGAATTCCCGGGTAT-3'

PATF-4 *for2*: 5'- TGAGCTGCAAGAACTGCAAGAGCTGCAGGCCCTCCTGCAGCAAATG-3'

PATF-4 *rev1*: 5'-
TTGCTGCAGGAGGGCCTGCAGCTCTTGCAGTTCTTGCAGCTCAATACCCGGGAATTCCGGCG
GTGGCGGAGGCGGTGGCGGAGGCGGTGGCGGCGGAGCTTTCT-3'

PATF-4 *rev2*: 5'-TAATTTTCGCACGGCAGTTCTTAGCACAGATTTTGATCTGCGCTTCCAT-3'

2.1.5 Preparation of antibiotic solutions

Stock solutions were prepared by dissolving each antibiotic in deionised water and filter sterilising using a 0.2 μm syringe filter before aliquoting for storage at -20 °C. Ampicillin stocks were prepared to a final concentration of 50 mg ml^{-1} in deionised water and used at a concentration of 100 μg ml^{-1}.

2.1.6 50× TAE (Tris-acetate/EDTA) electrophoresis buffer

Per litre:

242 g Tris base (pH 8.0)

27.1 ml glacial acetic acid

100 ml 0.5 M ethylene diamine tetraacetic acid (EDTA)

2.1.7 10× SDS running buffer

Per litre:

30.3 g Tris base (pH 8.3)

144 g glycine

10 g sodium dodecyl sulfate (SDS)

2.1.8 SDS gel-loading buffer

Per 10 ml:

1.25 ml 0.5 M Tris-HCl (pH 6.8)

2 ml 10% SDS

0.2 ml bromophenol blue

2.5 ml glycerol

0.5 ml β-mercaptoethanol

2.1.9 Staining buffer for SDS-PAGE

0.25 % (w/v) Coomassie brilliant blue

45% (v/v) methanol

10% (v/v) glacial acetic acid

All above in deionised water.

2.1.10 Destaining buffer for SDS-PAGE

10% (v/v) glacial acetic acid

12% (v/v) isopropanol

All above in deionised water.

2.1.11 Gel-loading buffer for agarose gel electrophoresis

0.25% (w/v) bromophenol blue

15% glycerol

All above in deionised water.

2.1.12 IPTG

100 mg/litre IPTG was dissolved in deionised water and sterilised by filtration through a 0.2 μm syringe filter.

2.1.13 Lysis buffer

50 mM potassium phosphate (pH 7)

100 mM sodium chloride

0.5 % Tween 20

All above dissolved in deionised water.

2.1.14 Guanidinium hydrochloride solution (Buffer A)

6 M Guanidinium hydrochloride dissolved in deionised water.

2.1.15 Ni^{2+}-NTA binding buffer (Buffer B)

50 mM Tris-HCl (pH 8.3)

300 mM sodium chloride

All above dissolved in 6 M guanidinium hydrochloride solution.

2.1.16 Ni NTA eluting buffer (Buffer C)

50 mM Tris-HCl (pH 8.3)

500 mM sodium chloride

500 mM imidazole

All above dissolved in 6 M guanidinium hydrochloride solution.

2.1.17 Peptide re-suspension solution prior to HPLC

40 % acetonitrile

0.1 % trifluoroacetic acid (TFA)

All above dissolved in deionised water.

2.1.18 HPLC buffers

Buffer A: HPLC grade acetonitrile

      0.1 % TFA

Buffer B: Degassed deionised water

      0.1 % TFA

2.1.19 Azobenzene alkylation buffer [154]

50 mM Tris base (pH 8.3)

200 mM triscarboxyethylphosphine (TCEP)

All above dissolved in deionised water.

2.1.20 Fluorescence Anisotropy assay buffer [154]

5 mM Tris-HCl (pH 7.9)

150 mM sodium chloride

6 mM magnesium chloride hexahydrate

15 % glycerol

All above dissolved in deionised water.

2.1.21 Circular Dichroism and UV spectroscopy buffer

5 mM potassium phosphate, pH 7.5.

2.1.22 PATF in vitro binding assay [167]

50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.0

100 mM potassium acetate

5 mM calcium chloride

6.5 % glycerol

1 mM dithiothreitol (DTT)

All above dissolved in deionised water.

2.1.23 *Hela in vitro* transcription buffers and reagents

The Promega HeLaScribe was used to carry out nuclear extract in vitro transcription. The procedure was followed according to the manufacturer's instructions.

2.1.23.1 rNTP stocks

25 × rNTP stock:

10 mM rATP

10 mM rCTP

10 mM rGTP

0.4 mM rUTP

Final 1 × rNTP stock:

0.4 mM rATP

0.4 mM rCTP

0.4 mM rGTP

16 μM rUTP

2.1.23.2 HeLa nuclear extract 1× transcription buffer

20 mM HEPES (pH 7.9 at 25 °C)

100 mM potassium chloride

0.2 mM ethylenediaminetetraacetic acid

0.5 mM dithiothreitol

20 % glycerol

All above dissolved in deionised water.

2.1.23.3 *Hela* extract stop solution

0.3 M Tris-HCl (pH 7.4 at 25 °C)

0.3 M sodium acetate

0.5 % sodium dodecyl sulfate

2 mM ethylenediaminetetraacetic acid

3 µg/ml tRNA

All above dissolved in deionised water.

2.1.23.4 Gel loading dye

98 % formamide

10 mM ethylenediaminetetraacetic acid

0.1 % xylene cyanol

0.2 % bromophenol blue

All above dissolved in deionised water.

2.1.23.5 TBE 10× buffer

0.9 M Tris-base

0.9 M boric acid

0.5 M ethylenediaminetetraacetic acid

All above dissolved in deionised water.

2.1.23.6 TE buffer

10 mM Tris-HCl (pH 8.0 at 25 °C)

1 mM ethylenediaminetetraacetic acid

All above dissolved in deionised water.

2.1.23.7 TE saturated phenol

Equal amounts of TE buffer and phenol were mixed and the phases were allowed to separate. This was repeated until the aqueous phase reached a pH of ~8.0.

2.1.23.8 TE saturated phenol:chloroform:isoamyl alcohol (25:24:1)

Equal amounts of TE buffer and phenol were mixed and the phases were allowed to separate. The lower part phenol phase (1 part) was mixed with (1 part) chloroform:isoamyl alcohol (24:1).

2.1.23.9 Transcription assay mixture

$(11 - x)$ μl *Hela* nuclear extract 1 × transcription buffer, where x is the volume of nuclear extract used:

1.5 μl (50 mM) magnesium chloride

1 μl 25 × rNTP mix

4 μl DNA template + PATF peptide mix

1 μl [α-^{32}P]rUTP (3000 Ci/mmol, 10 mCi/ml)

Nuclease-free water was added to a give a final volume after addition of the nuclear extract of 25 μl.

2.2 Molecular biology methods

2.2.1 Preparation of super competent cells

The desired *E. coli* XL1-blue strain was inoculated overnight in 10 ml LB media without antibiotics at 37 °C overnight. A volume (1 ml) of the overnight culture was grown in 100 ml LB medium (in a 250 ml conical flask) at 37 °C with shaking at 180 rpm, until an OD_{600} ~0.5 was reached (approximately 3 hours). The culture was then transferred into sterile 50 ml conical centrifuge tubes and was spun for 10 minutes at 2000 *g* at room temperature. The supernatant was discarded. The pellet was resuspended in 20 ml of cold buffer A (30 mM potassium acetate, 100 mM rubidium chloride, 10 mM calcium chloride, 50 mM manganese chloride, glycerol 15%, pH 5.8) and left on ice for 15 minutes. The solution was centrifuged as before and the pellet resuspended again, this time in 5-10 ml of cold buffer B (10 mM 3-(N-morpholino)propanesulfonic acid (MOPS), 75 mM calcium chloride, 10 mM rubidium chloride and 15% glycerol, pH 6.5), and left on ice for 1 hour. Aliquots (300 µl) were taken and stored at -80 °C. Buffers A and B were prepared in advance and sterilised by passing through a 0.22 µm filter.

2.2.2 Transformation

A sample of plasmid DNA (1 µl- 3 µl) and 100 µl of competent cells were mixed in an Eppendorf tube and incubated on ice for 30 minutes. The cells were subjected to a heat shock (40 °C) for 35-45 seconds and then placed again on ice for another 2 minutes. LB medium (1 ml) was added to the cells and the suspension was incubated at 37 °C for 1 hour with shaking at 180 rpm. The solution was then centrifuged for 1 minute (13,000 g) in a bench top microcentrifuge and the supernatant was discarded. The cells were re-

suspended in fresh LB medium (~200 µl) and the solution was plated on an agar plate containing the appropriate antibiotic. The plate was incubated at 37 °C overnight.

2.2.3 DNA isolation and purification

2.2.3.1 QIAprep spin miniprep kit

The QIAprep Spin Miniprep Kit was used for purification of 300 µg of high-copy plasmid DNA from 15 ml overnight cultures of *E. coli* in LB medium. The procedure was carried out according to the manufacturer's instructions.

2.2.3.2 Agarose gel electrophoresis

Agarose gels were used for analysis and isolation of DNA fragments. 1% Agarose gels were made up from 1× TAE buffer using a microwave oven to dissolve the agarose and the mixture was poured into the minigel kit and left to set. Samples were mixed with the gel-loading buffer and loaded into the gel. Gels were run in 1× TAE buffer at 100 V for 60 minutes after which the DNA was stained with ethidium bromide and visualised with a UV lamp at 254 nm. The DNA fragment of interest was extracted from an agarose gel with a clean, sharp scalpel. The QIAquick Gel Extraction Kit was used according to the manufacturer's instructions (QIAGEN).

2.2.4 Quantification of DNA and oligonucleotides in solution

The concentration of nucleic acids was determined spectrophotometrically using a Thermo Scientific NanoDrop 1000 fluorospectrometer, by measuring the optical density (OD) of a solution of DNA or oligonucleotide at 260 nm.

2.2.5 Storage of plasmids

2.2.5.1 Ethanol precipitation of Nucleic Acids

To precipitate the purified DNA, 50 µl ammonium acetate (7.5 M) and 2.5 volumes of absolute ethanol were mixed with the DNA solution and the mixture was left for 10 minutes at RT. The solution was then centrifuged for 15 minutes at maximum speed (~16,000 g) in a bench top microfuge. The supernatant was discarded and the pellet was washed with 250 µl of 80% ethanol. The solution was centrifuged for 10 minutes as before and the supernatant discarded. The pellet was dried at 37 °C in an oven and stored at -20 °C.

2.2.5.2 Glycerol stock

To prepare glycerol stocks a sample (0.8 ml) of cells from an overnight culture was aseptically mixed with 0.2 ml of sterile glycerol. Glycerol stocks were stored at -80 °C.

2.2.6 Digestion with restriction enzymes

Digestion reactions were performed with the desired volume of DNA (usually 10 µl) and the restriction endonucleases in their recommended buffers (New England Biolabs). The total volume for the reactions was 12 µl. Reactions were incubated for 2 hours at 37 °C. For double digestions the buffer used was the recommended one that gave the highest digestion efficiency for both enzymes. The DNA product for the digestion was analysed by agarose gel electrophoresis.

2.2.7 Dephosphorylation of DNA fragments

Following digestion with a single restriction enzyme, the vector should be dephosphorylated at the 5'-end with antarctic phosphatase to avoid self-ligation. DNA (0.5 µg 10 µl^{-1}) was suspended in 1× NEBuffer (by New England Biolabs) and 5 units of antarctic phosphatase were added to the mixture. The solution was incubated for 15 minutes at 37 °C. The enzyme was heat inactivated (at 65 °C) for 5 minutes, followed by purification of the DNA using QIAquick PCR purification protocol.

2.2.8 Phosphorylation of DNA fragments (Kinase reaction)

Forward oligonucleotides stock solutions (*for1* and *for2*, 100 pmol) were diluted to 50 pmol with deionized water. Equal volume quantities of 50 pmol *for1* and *for2* were then combined prior to phosphorylation. The same procedures were also applied to reverse oligonucleotides. Combined oligonucleotides were phosphorylated at the 5'-end with 0.5 µl T4 polynucleotide kinase (20,000 U ml^{-1}, New England Biolabs) in T4 ligase buffer (50 mM Tris-Cl (pH 7.5 at 25 °C), 10 mM magnesium chloride, 10 mM dithiothreitol, 1 mM ATP, 25 µg ml^{-1} BSA) and 2 mM magnesium chloride. The solutions were incubated at 37 °C for 2 hours, and subsequently purified using the QIAquick PCR purification protocol and stored at -20 °C for future use. The phosphorylation of *for1:for2* and *rev1:rev2* mixtures were done in separate tubes.

2.2.9 Annealing Oligonucleotides

Equimolar quantities of oligonucleotides were mixed in an Eppendorf tube that was placed in a thermal cycler at 95 °C. Gradually the temperature was reduced until the final temperature was at 15 °C over a period of 15 hours. The annealed oligonucleotides were stored at -20 °C.

2.2.10 Ligation Reaction

Ligation was carried out to insert the desired gene into an expression vector. The digested plasmid and the annealed oligonucleotides were mixed together with 1 μl T4 DNA ligase (10 U μl^{-1}) and 1 μl 10× T4 DNA ligase buffer (400 mM Tris-Cl (pH 7.8 at 25 °C), 100 mM magnesium chloride, 100 mM dithiothreitol, 5 mM ATP). The solution was then incubated at RT overnight or at 16 °C for 4 hours. Products were stored at -20 °C.

2.2.11 DNA sequencing

All sequencing reactions were carried out by Cardiff University sequencing services using T7 promoter and T7 terminator primers.

2.3 General methods for protein preparation and purification

The expression of peptides was done using Novagen pET31b expression system. The procedure was conducted according to the manufacturer's instructions.

2.3.1 Growth of bacterial cultures

Overnight cultures were grown in sterile conditions by inoculating a single colony from an agar plate into 5 ml LB medium containing the appropriate antibiotic. The cultures were incubated overnight at 37 °C with constant shaking at 180 rpm.

2.3.2 Protein expression using the T7 system

The desired competent cells were transformed with plasmid DNA. For large scale expression, the colonies were grown overnight in 10 ml LB medium containing the appropriate antibiotic. The entire contents of each of the overnight cultures were transferred to separate 550 ml of fresh LB medium containing antibiotic in 2 litre conical flasks, and incubated until they reached an $OD_{600} \sim 0.6\text{-}0.8$. The cells were then induced with IPTG to a final concentration of 0.8 mM, and left to grow for 4 to 6 hours. Samples (1 mL) were taken from each flask (including a sample prior to induction) and a SDS-PAGE was run to analyse the extent of the expression of the target protein. The cells were then centrifuged at 6,000 g for 30 minutes at 4 °C and pellets were stored at -20 °C.

2.3.3 Purification of PATF proteins

2.3.3.1 Lysis of cells

Frozen pellets from a large-scale expression were thawed on ice and resuspended by vortexing in the appropriate lysis buffer. The suspension was sonicated for 10 minutes (5 s on, 5 s off) on ice using a Sonicator W-37 (Heat Systems Ultrasonics Inc.). The resulting lysate was centrifuged (40,000 g, 30 minutes) and the supernatant removed and kept for reference. A further 30 ml of lysis buffer was added to resuspend the pellet and the cells were then centrifuged at 17,500 g for 20 minutes. This step was repeated (obtaining the supernatant 2 and 3 from each step that were kept for reference). To the pellet, 20 ml Buffer A was added and resuspended for 60 minutes at 4 °C. The suspension was then centrifuged at 25,000 g for 30 minutes. The final pellet was also kept for reference, and solubilised protein was kept for further purification.

2.3.3.2 Purification of peptides by Ni NTA

Solubilised protein was diluted with Ni NTA binding buffer B to 1× concentration. The mixture was loaded onto Ni-NTA agarose (QIAGEN) packed column pre-washed in Ni-NTA binding buffer B. After passing the solubilised protein through the column, this was washed several times with buffer B to wash any unbound proteins. After washing several times the bound protein was then eluted from the column with eluting buffer C.

2.3.3.3 Dialysis of the fusion protein

In order to remove salt and urea, the protein was dialysed using Medicell International Ltd. dialysis membrane (12,000-14,000 MWCO) in deionised H_2O overnight. Dialysis was carried out at 4 °C until the fusion protein precipitated out of the solution. The precipitated fusion protein was centrifuged at 5,000 g, the supernatant was kept for reference and the pelleted fusion protein was stored at -20 °C.

2.3.3.4 Cyanogen bromide cleavage of fusion protein

The pellet was dissolved in 80% formic acid (10 ml) and cyanogen bromide (0.3 g) was added to the mixture. This cleaved the desired peptide from the His-tag and KSI protein by reacting with methionine sidechains to truncate the polyamide chain with concomitant conversion of methionine into homoserine (**Chapter 3**). The flask containing this mixture was wrapped in aluminium foil and stirred under nitrogen for 18-20 hours. The solvent was evaporated and the product resuspended in acetonitrile/water (1:1).

The mixture was stirred for 1 hour and centrifuged (20 min, 4 °C, 25,000 g). The supernatant solution was purified by semi-preparative reverse phase HPLC (C_{18}

column) eluting with a 10-100% gradient of aqueous acetonitrile containing 0.1% TFA over 1 hr. The sample was freeze-dried and analysed by MALDI-TOF.

2.3.4 Mass spectrometry

Matrix Assisted Laser Desorption Ionisation-Time of Flight (MALDI-TOF) mass spectrometry were performed to identify the purified proteins. α-cyano-4-hydroxycinnamic acid (**Figure 2.3.4**) containing 1:1 MeCN/H$_2$O was the matrix used for MALDI-TOF experiments.
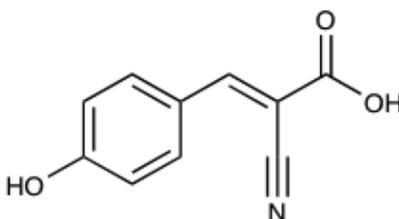
Figure 2.3.4. *α-Cyano-4-hydroxycinnamic acid.*

2.3.5 Measurement and calculation of protein concentration
2.3.5.1 Determination of the concentration of unalkylated PATF

The concentrations of PATFs were determined spectrophotometrically. Absorbance measurements were taken at 210 nm, 215 nm and 220 nm using a Jasco V-650 UV-Vis spectrophotometer. A clean quartz cuvette was used to hold the samples. The measurements were repeated three times, taking the mean value to calculate the concentration, and were carried out in the following order:

1 ml deionised water + 5 µl reference buffer (5 mM potassium phosphate)

Cuvette was carefully washed with deionised water

1 ml deionised water + 5 µl peptide solution

Peptide concentration was calculated using the equation:

$$C \ (\text{mg ml}^{-1}) = [(\text{Peptide}–H_2O) – (\text{Buffer}–H_2O) \times 201] / X_\lambda \qquad \textit{Equation 1}$$

Where $X_{210} = 20$, $X_{215} = 15$ and $X_{220} = 11$.

2.3.5.2 Determination of the concentration of Sulfo-azobenzene alkylated peptides

The extinction coefficient of the sulfonate-azobenzene cross-linker is 24,000 M^{-1} cm^{-1} at 363 nm [1] and was used to determine the concentration of cross-linked PATF peptides using Thermo Scientific NanoDrop 1000 spectrophotometer.

2.3.5.3 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

A volume (10 ml) of 15 % resolving gel (2.4 ml deionised water, 5 ml 30 % degassed acrylamide/bisacrylamide, 2.5 ml 1.5 M Tris-Cl (pH 8.8), 0.1 ml 10% w/v sodium dodecylsulfate) was mixed with polymerising agents ammonium persulfate (100 μl, 10 % w/v in water) and TEMED (10 μl). The solution was immediately poured between the assembled glass plates and left to polymerise at RT. An aliquot (3 ml) of 5 % stacking gel (5.7 ml deionised water, 1.7 ml 30 % degassed acrylamide/Bis acrylamide, 2.5 ml 0.5 M Tris-Cl (pH 6.8), 0.1 ml 10 % w/v SDS) containing was mixed with 10 % ammonium persulfate (100 μl) and TEMED (10 μl) and pipetted on top of the resolving gel. A comb was immediately inserted between the plates and the stacking gel was left to polymerise at RT. The comb was then removed and the wells rinsed with 1× SDS running buffer. Samples, mixed with 1× SDS gel-loading buffer, were loaded into the gel (15 μl in the 10-well gels and 8 μl in the 15-well gels) and subjected to 180 V potential for 60 minutes. After electrophoresis, the gel was stained and destained using the corresponding buffers, followed by visualisation of the protein bands on a light box equipped with a UV lamp.

2.3.6 Crosslinking the PATF peptides with 3,3'-bis(sulfo)-4,4'-bis(chloroacetamido) azobenzene crosslinker

The appropriate PATF (1 mg) was dissolved in Tris-HCl buffer (3 ml, 50 mM, pH 8.3) and *tris*(2-carboxyethyl)phosphine (TCEP) stock solution (30 μl, 200 mM) was added. Crosslinker (4 eq, approx 1.5 mg) was dissolved in Tris-HCl buffer (1.2 ml, 50 mM, pH 8.3), and was added in 3 aliquots to the protein solution at 20 minutes interval and stirred overnight at 5 °C. The cross-linked protein was purified by semi-preparative reverse phase HPLC (C$_{18}$ column) eluting with 10-100% aqueous acetonitrile containing 0.1% TFA over 1 hour. The mass of the crosslinked PATF was confirmed by MALDI-TOF mas spectrometry.

2.4 Photoisomerisation

Photoisomerisation of dark-adapted (dad) XL-PATFs was achieved by irradiating a solution of the peptide with a 250 W metal halide UV Light Point Source (UV-P 280) coupled to a 360 nm band pass filter (10 nm bandwidth). Photoisomerisation was complete in less than 6 minutes as judged by the absence of further changes in the UV/Vis spectra. The percentage of isomerisation was calculated by using the respective extinction coefficients for the pure *cis* from (ε_{363} = 1100 M^{-1} cm^{-1}) [2] and the *trans* form to determine the ratio from the obtained absorbance values before and after irradiation (A_{dad} and A_{lid}).

$A_{lid} = A_{cis} + A_{trans}$

$A_{lid} = x\, \varepsilon_{cis}\, d\, [BPN] + (1-x)\, \varepsilon_{trans}\, d\, [BPN]$ (x = fraction of *cis* isomer)

$$x = \frac{A_{lid} - \varepsilon_{trans} d[BPN]}{(\varepsilon_{cis} - \varepsilon_{trans})d[BPN]}$$

$$x = \frac{A_{lid} - A_{dad}}{(\varepsilon_{cis} - \varepsilon_{trans})d[BPN]} \qquad\qquad \textit{Equation 2}$$

2.5 UV/Visible absorption experiments

UV-Visible absorption experiments were carried out using a JASCO V-660 UV/Vis Spectrophotometer with a 1 cm path length cuvette. Samples were dissolved in 5 mM potassium phosphate (pH 8.0) to a final concentration of 2 µM and experiments were run at 25 °C. Assuming first order kinetics for the thermal *cis*-to-*trans* relaxation of the crosslinker, the half-life ($t_{1/2}$) of the *cis*-isomer could be calculated with the equation:

$$t_{1/2} = \ln 2 \, / \, k$$

Where k is the 1st order rate constant for the thermal relaxation process, which was calculated by plotting the natural logarithm of the percentage of *cis* isomer (calculated from the absorbance at 363 nm) versus time:

$$\ln (\%A^{cis}_{363}) = -k \, t + c$$

The activation energy for the isomerisation reaction was determined using the Arrhenius equation:

$$k = Ae^{\frac{-Ea}{RT}}$$

Where Ea is the activation energy and A is the pre-exponential factor.

2.6 Circular dichroism (CD) spectroscopy

CD spectra were recorded on a ChiraScan spectrometer. The temperature-dependent scans were performed at 25 °C using a 1 mm path length cuvette. The mean residue ellipticities $[\Theta]_r$ of different states of PATFs were calculated according to the equation:

$$[\Theta]_r = \Theta / (10 \, n \, c \, l)$$

Where n is the number of backbone amide bonds (46 in this case), c is the concentration and l is the pathlength (0.1 cm). The $[\Theta]_r$ at 222 nm of dark adapted and irradiated XL-PATF were found to be -12872 and -11735 deg cm^2 dmol^{-1} respectively.

The α-helical content of both dark adapted and irradiated XL-PATF were calculated by the following equation: [168]

$$\% \text{ helicity} = (-100 \times n \times [\Theta]_{r, \, 222}) / [40000 \, (n-4)]$$

2.7 DNA Binding specificity measured by Fluorescence Anisotropy

Fluorescence anisotropy measurements were performed at 15 °C on a Perkin Elmer Luminescence Spectrometer LS55 arranged in L format (494 nm excitation; 525 nm emission) using a 1 ml fluorescence quartz cuvette with samples dissolved in 5 mM Tris-HCl (pH 7.9), 150 mM sodium chloride, 6 mM magnesium chloride, 15% glycerol, containing 2.5 mM TCEP for measurements carried out with the unalkylated PATFs. Titrations of PATFs and XL-PATFs at concentrations 0-5,000 nM were successively added to 0.1-5 nM of fluorescein (5-FAM) labelled DNA in a total volume of 1 ml. The total volume varied less than 6% for each titration therefore the concentration of oligonucleotide can be assumed to have remained constant. For each anisotropy value ten measurements were taken using an integration time of 5 s.

The G factor (ratio of sensitivities of the monochromator for horizontally and vertically polarised light) can be calculated using the equation: [169]

$$G = I_{\perp} / I_{\parallel}$$

where $I_{\parallel}$ and $I_{\perp}$ are the intensities of the fluorescent emissions in parallel and perpendicular planes, respectively, to the excitation plane. The G factor value was always close to 1.14 for both PATFs and XL-PATFs. Values for fluorescence anisotropy (*A*) were then determined from the equation: [170]

$$A = (I_{\parallel} - G\, I_{\perp}) / (I_{\parallel} + 2GI_{\perp})$$

Fluorescence anisotropy data were expressed as:

$$F = (A - A_D) / (A_{PD} - A_D)$$

where *A* denotes the fluorescence anisotropy in the presence of the indicated concentration of protein, A_D denotes the fluorescence anisotropy in the absence of protein, and A_{PD} denotes the fluorescence anisotropy at saturation. [170]

The data were fit to the Langmuir isotherm:

$$F_{Fit} = 1 / (1 + K_D^{\,n} / [P]^{n})$$

where K_D is the apparent dissociation constant, [P] is the concentration of protein and *n* is the number of binding sites.

2.8 Hela in vitro transcription assay

The Promega HelaScribe kit was used to carry out nuclear extract in vitro transcription. The procedure was carried out according to the manufacturer's instructions.

2.8.1 Incubation of PATFs with pUCML53 (and doubly digested pUCML53) plasmid template

XL-PATF peptides (5 µL, 500 nM) were incubated with pUCML53 plasmid template (5 µL 100 ng) at room temperature for 75 minutes to a total volume of 10 µL. For photo-switched (irrad) XL-PATFs, the peptides were irradiated with UV source for 6 minutes prior to addition to the plasmid template. Upon addition of the irrad XL-PATFs, the mixture was irradiated with UV light for the remaining 75 minutes at room temperature.

2.8.2 In vitro transcription reaction

Incubated XL PATF and pUCML53 mixture was added to the transcription assay solution. 8 µl nuclear extract was then added to the mixture in order to start transcription to a total volume of 25 µl. The reaction mixture was incubated for 60 minutes at 30 °C. For irrad XL-PATFs, the reaction mixture was irradiated for 3 minutes every 15 minutes in the course of 1 hour reaction. After 1 hour, RNase T1 (10 units, Thermo Scientific) was added to the reaction mixture and incubated for further 15 minutes at 30 °C in order to cleave RNA transcripts other than the G-less cassette. [203] The reaction was terminated by the addition of the stop solution (175 µl). In order to improve reproducibility, the Hela extract stop solution was warmed at 25 °C for one hour and vortexed well to mix prior to addition.

2.8.3 Isolation of RNA transcript

TE-saturated phenol (200 µl) was added to the reaction, and the resulting mixture was then vortexed for 60 seconds and centrifuged at 14,000 g for 5 minutes. The upper aqueous phase (150 µl) was transferred to a clean tube. The phenol phase was extracted again by addition of the stop solution (200 µl) and the mixture was vortexed as

described above. The aqueous layers were combined and extracted with chloroform:isoamyl alcohol (24:1) (300 μl), vortexed and spun at 14,000 g for 5 minutes. The aqueous layer was then transferred to a clean Eppendorf tube. Alternatively, the RNA was mixed with phenol:chloroform:isoamyl alcohol (25:24:1) (200 μl), vortexed for 60 seconds and centrifuged at 14,000 g for 5 minutes. The upper, aqueous phase was then transferred to a clean Eppendorf tube. In order to precipitate RNA, 100 % ethanol (700 μl) was added to the aqueous layer, the resulting solution was mixed by vortexing and placed in dry ice bucket for 15 minutes before putting the sample in -80 °C freezer overnight. After overnight incubation, the tube was spun at 14,000 g for 10 minutes at 4 °C by positioning the tube in the microcentrifuge so that the pellet will form in known orientation. The supernatant was carefully removed by pipetting, making sure not to disturb the pellet. Sometimes the pellet was not clearly visible. The pellet was then dried in a vacuum desiccator containing silica gel (260 mm diameter).

2.8.4 Urea gel preparation

For a single 7.5 % urea gel, a solution composed of 9.38 ml 40% acrylamide/bisacrylamide (ratio 23:1) solution, 2.5 ml 10 × TBE buffer, 7 M urea (21.02 g), and deionised H_2O (38.12 ml) was made and carefully mixed so as not to introduce air bubbles. Polymerising agents, TEMED (300 μl) and ammonium persulfate (500 μl) were added. The mixture was immediately poured between the assembled glass plates (front plate: 16 × 19.7 cm; back plate: 19 × 19.7 cm; spacers: 1 mm) and fixed together using bulldog clips. A comb (20 teeth) was inserted and the plates were left to polymerise horizontally at room temperature, then stored at 4 °C. Gels were freshly prepared on the day of each experiment. Prior to use, the gels were pre-run in a vertical

gel electrophoresis apparatus (V1517, BRL Life technologies, Inc.) by subjecting them to a 250 V potential for 30 minutes at 4 °C eluting with 0.5 × TAE running buffer.

2.8.5 Gel analysis of RNA transcripts and visualization

After the RNA sample was dried of ethanol, the pellet carefully resuspended in 10 μl of nuclease-free water. An equal volume of loading dye was added to the resuspended pellet and the sample was heated at 90 °C for 10 minutes prior to loading onto a gel. The gel well was flushed with 0.5 × TBE buffer prior to loading of the samples. The gel was then run at a potential of 250 volts for 3 hours or until the bromophenol blue dye had dye just run off the gel. The gel was carefully removed from the plate and soaked in deionised water for at least 20 minutes in order to remove urea present in the gel which adversely affects the resolution of the subsequent imaging. The gel was then exposed to phosphorimaging screen for at least 3 hours before visualising on phosphorimaging scanner (Typhoon 9400, GE Healthcare Life Sciences). Quantification of the RNA transcript was done by densitometry using ImageJ software. Each transcription experiment had a positive control (CMV promega or pMLΔ53) and the quantification of each band was compared to the positive control.

2.9 Synthesis of 3,3'-bis(sulfo)-4,4'-bis(chloroacetamido)azobenzene crosslinker [146]

2.9.1 Synthesis of 2-acetylamino-5-aminobenzenesulfonic acid

2,5-Diaminobenzenesulfonic acid (15 g, 79.7 mmol, 1 equiv.) was added to glacial acetic acid (150 ml), the mixture was heated to 94 °C with stirring. Acetic anhydride (9.24 ml, 98 mmol 1.2 equiv.) was added dropwise to the reaction mixture over a period of 5 minutes. The mixture was left to stir at 94 °C for 2 hours. After 2 hours, the reaction flask was left to cool to room temperature, the precipitate filtered under

vacuum and rinsed with acetic anhydride (3 × 10 ml). A purple solid was obtained (4.91g, 26%). **¹H NMR** (400 MHz, DMSO) δ 10.36 (s, 1H, NH), 8.34 (d, J = 8.7 Hz, 1H), 7.70 (d, J = 2.7 Hz, 1H), 7.35 (dd, J = 2.7 Hz, 8.8 Hz, 1H), 2.08 (s, 3H, CH$_3$). **¹³C NMR** (100 MHz, DMSO) δ 26.1, 121.8, 122.9, 125.4, 126.5, 136.0, 137.2, 169.0, **HRMS** EI+ve calculated 230.0361 for C$_8$H$_{10}$N$_2$O$_4$S, observed 230.0365.

2.9.2 Synthesis of sodium 3,3'-bis(sulfonato)-4,4'bis(acetamido)azobenzene

2-Acetylamino-5-aminobenzenesulfonic acid (4.91g, 21.31 mmol, 1 equiv.) was dissolved in H$_2$O (51 ml), the pH of the solution was adjusted to pH 8.5 *via* addition of sodium carbonate (0.1 M) dropwise. This turned the solution from light pink to brown. The mixture was then cooled in an ice bath and sodium hypochlorite (available chlorine ≥ 4%, 54 ml, 0.65 mol) was added dropwise over 5 minutes. The reaction flask was left to stir at -5 °C in an ice-cooled H$_2$O/NaCl bath for 2 hours. Finally the mixture was wrapped in aluminium foil and left to incubate at 4 °C for 3 days. The precipitate was filtered and washed with hot ethanol to yield 3,3'-bis(sulfonato)-4,4'-bis(amino)azobenzene (1.26g, 12%). **¹H NMR** (400 MHz, D$_2$O) δ 8.18 (d, J = 2.3 Hz, 1H), 7.98 (d, J = 8.7 Hz, 1H), 7.85 (dd, J = 2.4 Hz, 8.7 Hz, 1H), 2.12 (s, 3H, CH$_3$). **¹³C NMR** (100 Mhz, DMSO) δ 2.2, 121.9, 124.2, 125.9, 133.8, 136.3, 147.8, 172.4, **HRMS** calculated 455.0336 for C$_{16}$H$_{15}$N$_4$O$_8$S$_2^-$, observed 455.0318.

2.9.3 Synthesis of sodium 3,3'-bis(sulfonato)-4,4'-bis(amino)azobenzene

Sodium 3,3'-bis(sulfonato)-4,4'-bis(acetamido)azobenzene (0.050 g, 0.11 mmol, 1 equiv.) was mixed with H$_2$O (2 ml). Concentrated hydrochloric acid (0.56 ml) was added to the mixture and the reaction flask was heated with stirring at 110 °C for 2.5 hours. The mixture was then left to cool to room temperature, and the solvent rotary

evaporated. The solids were re-dissolved in H$_2$O (8 ml) and pH adjusted *via* addition of NaOH (0.1 M). The mixture was then freeze dried overnight resulting in solid sodium 3,3'-bis(sulfonato)-4,4'-bis(amino)azobenzene (0.028 g, 69%). **^{1}H NMR** (400 MHz, DMSO) δ 7.90 (d, *J* = 2.3 Hz, 1H), 7.50 (dd, *J* = 2.3 Hz, 8.7 Hz, 1H), 6.71 (d, *J* = 8.7 Hz, 1H) , 6.20 (s, 2H, NH$_2$). **^{13}C NMR** (100 MHz, DMSO) δ 115.1, 121.4, 124.6, 129.6, 141.6, 147.0, **HRMS** ESI –ve calculated 371.0125 for C$_{12}$H$_{11}$N$_4$O$_6$S$_2^-$ observed 371.0129.

2.9.4 Synthesis of 3,3'-bis(sulfo)-4,4'-bis(chloroacetamido)azobenzene crosslinker

Sodium 3,3'-bis(sulfonato)-4,4'-bis(amino)azobenzene (0.028 g, 0.075 mmol, 1 equiv.) was mixed with chloroacetic acid (0.191 g, 2.02 mmol, 25 equiv.) and chloroacetic anhydride (0.346 g, 2.03 mmol 25 equiv.). The mixture was heated at 87 °C with stirring overnight, left to cool to room temperature and continue to cool on ice for another 30 minutes to precipitate 3,3'-bis(sulfo)-4,4'-bis(chloroacetamido)azobenzene crosslinker. The precipitate was washed with excess dichloromethane in order to remove any residual chloroacetic acid and chloroacetic anhydride. The crosslinker was finally filtered under vacuum and the precipitate collected (0.023 g, 59%). **^{1}H NMR** (400 MHz, DMSO) δ 11.20 (s, 1H, NH), 8.52 (d, *J* = 8.8 Hz, 1H), 8.22 (d, *J* = 2.4 Hz, 1H), 7.98 (dd, *J* = 2.4 Hz, 6.4 Hz, 1H), 4.23 (s, 2H, CH$_2$). **^{13}C NMR** (100 MHz, DMSO) δ 44.2, 120.2, 121.1, 126.8, 137.4, 137.6, 147.7, 165.7, **HRMS** ESI -ve calculated 522.9557 for C$_{16}$H$_{13}$N$_4$O$_8$S$_2$Cl$_2^-$, observed 522.9578.

Chapter 3

Photocontrol of Artificial Miniature Engrailed Homeodomain Transcriptional Activators

3.1 Introduction

This chapter describes the extension of the use of an azobenzene crosslinker to stabilise the recognition helix of a homeodomain transcription factor by adding a linker and activating domain to investigate whether the resulting molecule can act as a photoswitchable artificial transcription factor (PATF). The basis for the DBD element of these molecules will be a homeodomain. [171] DNA binding by full-length homeodomains relies on interactions between helix-3 of the helix-turn-helix motif and the major groove of DNA. Helix-3 in turn is stabilised by two anti-parallel helices, helix-1 and helix-2, which in turn are perpendicular to helix-3, forming a hydrophobic core between the helices (**Figure 3.1.1**).
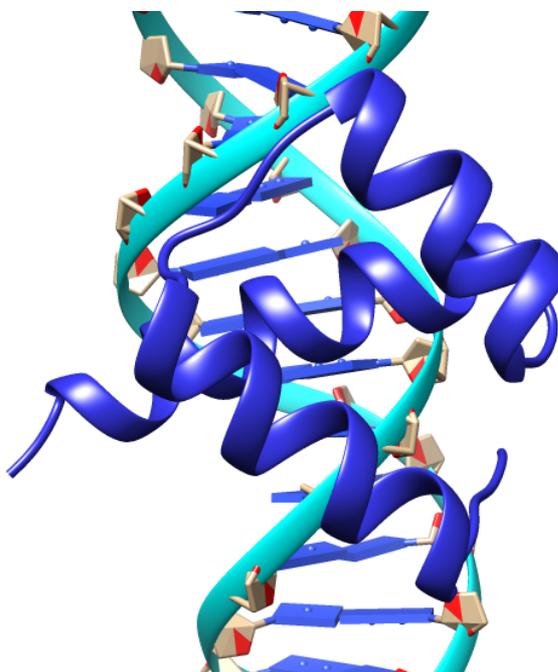
Figure 3.1.1. *Model of engrailed homeodomain with the helices-1 and -2 anti-parallel to each other and perpendicular to helix-3 binding to the major groove of DNA (3HDD.pdb).*

This packing stabilises the recognition helix. The N-terminal arm of the complex (**Figure 3.1.2**), which is unstructured in the absence of DNA, wraps over the phosphate backbone to make DNA contacts in the minor groove. [172] Initial attempts to reduce the wild-type homeodomain to just the helix-turn-helix motif by removing the N-terminal arm resulted in a reduction in binding affinity to specific target DNA by up to 10 fold, most likely due to the loss of the N-terminal contacts with minor-groove edges of the first two bases of the core sequence TAAT that are common to the DNA binding sites of homeodomain proteins (**Figure 3.1.2**). [20]
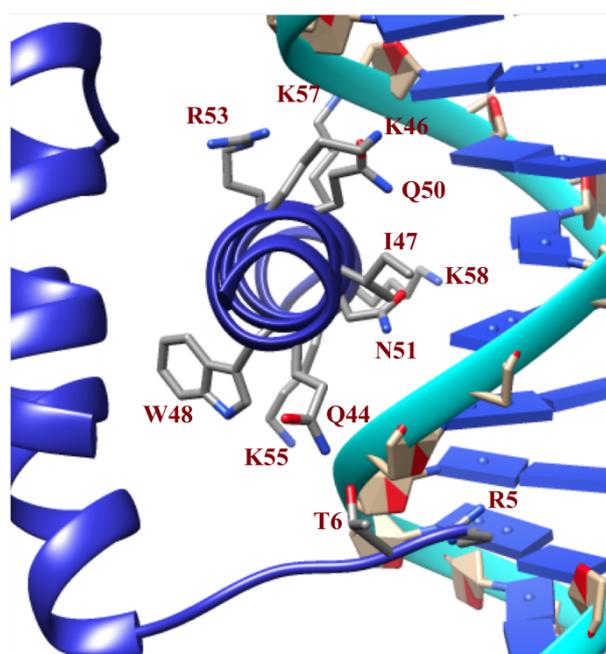
Figure 3.1.2. *Binding of residues from the engrailed homeodomain that contact the DNA (1HDD).* [173]

Residue 50 of homeodomains appears to be a key determinant of DNA-binding specificity, helping to distinguish between binding sites of the form TAAT<u>NN</u>. [174] Glutamine (Q) is the most commonly occurring residue at position 50, but other amino acids are found in various subfamilies. It was found that the tightest and most specific

binding to its cognate DNA sequence occurs when lysine (K) is present at position 50. Engrailed Q50K variants bind more tightly to TAAT<u>CC</u> (QRE sequence) than wild-type engrailed binds to TAAT<u>TA</u> (ERE sequence). [174] The Lys50 side chain projects directly into the major groove of the DNA making a series of hydrogen bonds with the guanine at the base pairs 5 and 6 of the optimal TAAT<u>CC</u> binding site (**Figure 3.1.3**). [174, 175]
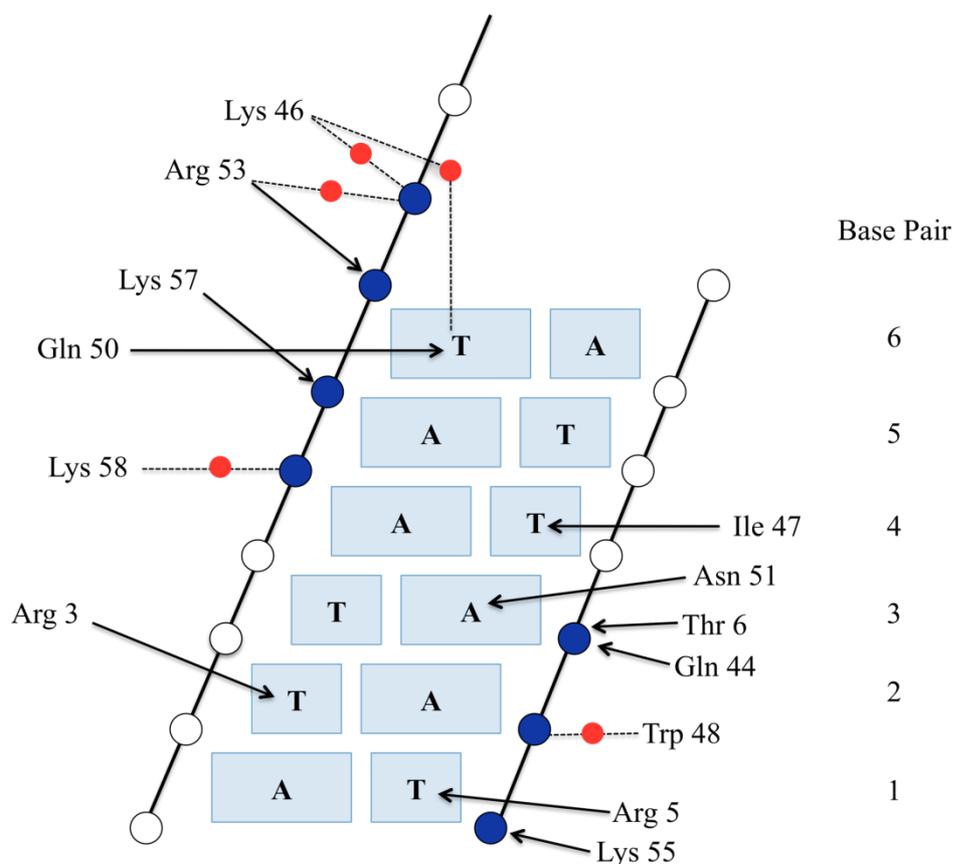
Figure 3.1.3. *Schematic showing contacts between engrailed homeodomain and its cognate DNA. Phosphates that are contacted by the homeodomain are shown in blue and hydrogen-bond intermediate water molecules are shown in red.* [18]

Montclare *et al*. studied the isolated recognition helix (helix-3) of Q50K engrailed homeodomain and found that, without stabilisation, its binding to target DNA sequence TAATCC was reduced by approximately 800 fold (1.7 µM) compared to the wild type protein (2.1 nM). [103, 79] Subsequent work detailed how a miniature engrailed

homeodomain peptide can be stabilised by alkylating thiols of appropriately spaced cysteine residues within the section of a DNA binding protein (**Figure 3.1.4**) with the azobenzene-derived photoisomerisable cross-linker 3, 3'-bis(sulfo)-4,4'-bis(chloroacetamido)azobenzene (BSBCA). [154, 176]
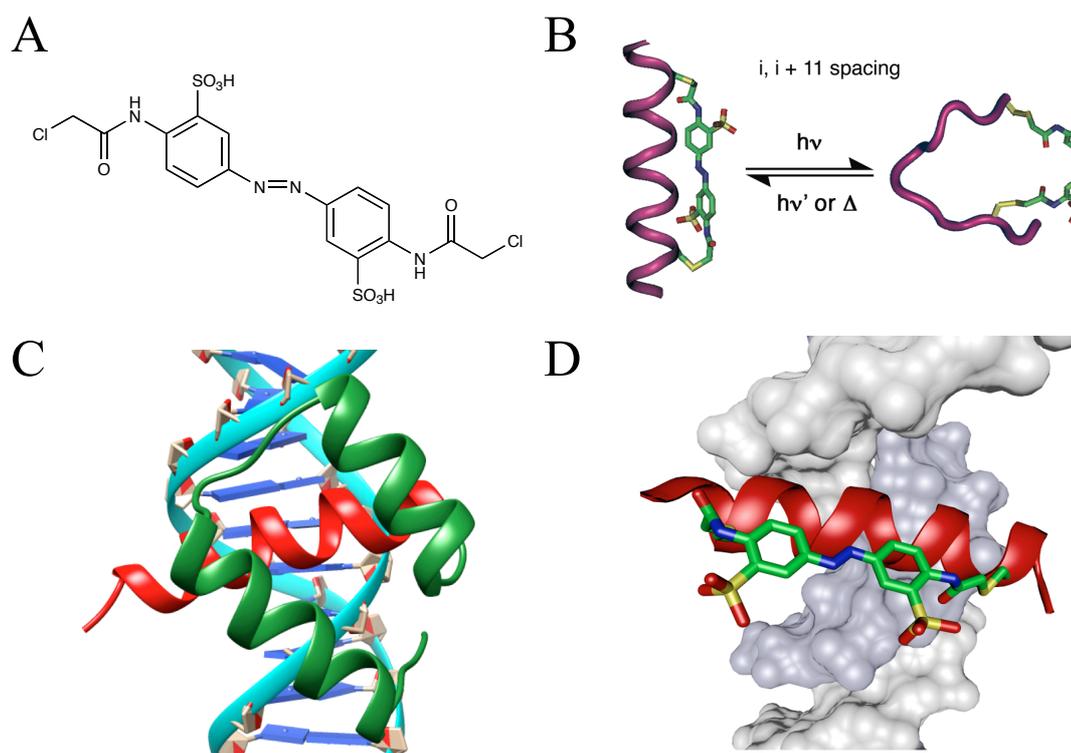
Figure 3.1.4. *A) Structure of crosslinker 3,3'-bis(sulfo)-4,4'-bis(chloroacetamido)azobenzene (BSBCA). B) Photoisomerisation of BSBCA linked to a protein structure with* via *the sidechains of two cysteine residues in an i, i+11 spacing. [154] C) Model of engrailed homeodomain with the helices-1 and -2 antiparallel (green) to each other and perpendicular to helix-3 (red) binding to the major groove of DNA (3HDD.pdb). D) Cartoon of modified helix-3 with an azobenzene crosslinker (right).*

In order to accommodate the azobenzene cross-linker the miniature engrailed homeodomain (HDH-3) was reduced to an 18-residue peptide, Ile45 and Ile56 were substituted with cysteines residues and the bulky Trp48, Phe49 and Lys52 of Q50K engrailed were replaced with helix-stabilising alanine residues in order to avoid steric conflict with the cross-linker. The C-terminal Ser59 was also substituted with the helix-stabilising Ala (**Figure 3.1.5.**). [154,177]

A

		45	50	55

Helix-3 Q50K eng   EAQ IKIWFKNKRAK IKKS
HDH-3   Ac - EAQCKIAAKNARAKCKKA - NH₂
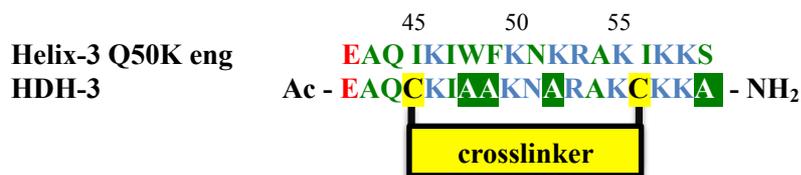
B
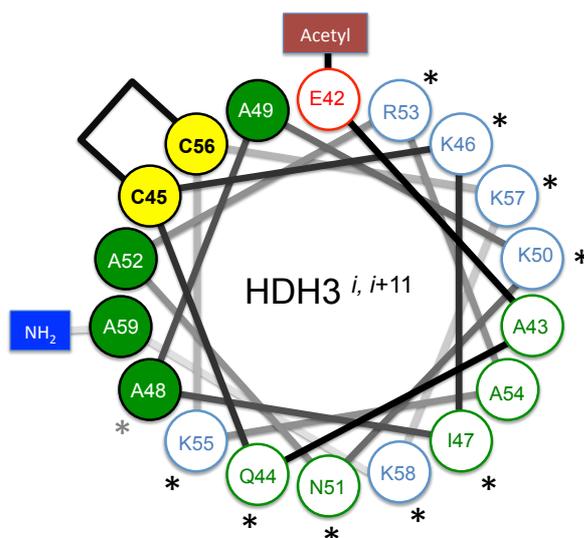
Figure 3.1.5. *Schematic of HDH-3. A) Sequences of wild type engrailed homeodomain compared to HDH-3 peptide with the modified amino acids. B) Diagram of the helical wheel of cross-linked HDH-3. Residues that are in contact with DNA are marked with asterisk, blue and red circles represent charged residues. Residues that have been substituted in order to avoid steric clash with the cross-linker are marked in filled green circles.* [154, 175]

An *i, i*+11 spacing of cysteine residues on helix-3 was used, corresponding to three turns of an α-helix and matching the *trans* configuration the azobenzene with length of approximately 16.8 Å between the *para*-acetoamido groups (**Figure 3.1.4**). This arrangement stabilised the helical structure of helix-3 in the dark state. Upon irradiation, the *cis*-conformation of the crosslinker is too short (approximately 11.3 Å) to readily accommodate an α-helical conformation, [142, 146] resulting in distortion of the helix and a reduction in DNA binding affinity. Dark adapted (dad) HDH-3 bound to its natural DNA target sequence QRE with a dissociation constant (K_D) of 7.5 +/- 1.3 nM, making

the complex 200 times more stable than that formed between the isolated helix-3 of engrailed and QRE DNA, [154, 103] although it still binds approximately 1000 fold less tightly that the full-length protein.[15] Upon irradiation, the dissociation constant increased by nearly 20 fold to 140 +/- 25 nM. [154]

3.2 Results and Discussion

3.2.1. Design of i, i+ 11 Homeodomain Artificial Transcription Factors (PATFs)

Giniger *et al.* designed a 15-residue sequence containing a core repeat of glutamic acid, glutamine and leucine, which was designed to form a four-turn, negatively-charged, amphipathic α-helix named AH (**Table 3.2.1.1**). This AH and a control sequence that should not form an amphipathic α-helix were attached to the yeast regulatory protein GAL4. [166] As predicted, AH functioned as an activating domain in a yeast extract cellular system and was as effective an activator as the natural yeast activator Gal4, whereas the variant with the non-amphipathic control sequence did not initiate the transcription process, suggesting that the amphipathic nature of activators is necessary for transcription activity. It is not known exactly how amphipathic α-helices function as transcriptional activating regions in eukaryotes, but they are believed to interact with RNA polymerase in a similar fashion to the negatively charged amphipathic α-helices found in transcription activators in bacteriophage λ and its close relative bacteriophage 434. [178] Other activating domains use similar residues; for example, the *Sp1* (Specificity Protein 1) transcription factor belongs to a family of glutamine-rich activating domains. The activating domain of *Sp1* contains very few negatively charged residues but approximately 25% glutamine residues that are essential for function, deletion abolishes the ability of SP1 to initiate transcription. [179] Indeed, Gerber *et al.* demonstrated that a continuous run of glutamine residues with no other amino acids fused to the DNA

binding domain of GAL4 factor can activate transcription. [180] Inspired by CTF/NF1 factor (or the CAAT box-binding transcription factor/nuclear factor-1) [181] which contain proline-rich (25%) activating domains, [182] Gerber *et al.* also demonstrated that a continuous run of proline residues fused to the DNA binding domain of GAL4 was also able to initiate transcriptional activation. [180] Other proline-rich transcription factors include the oncogene product Jun, AP2 (Activating Protein 2), and the C-terminal activation domain of the Oct2 homeodomain. [183]

Activation Domain	Sequence
AH	ELQELQELQALLQQQ

Table 3.2.1.1. *Sequence of peptide designed to form an amphipathic helix (AH) and function as an artificial activation domain. In most E. coli-encoded activating peptides the hydrophobic face of the amphipathic helix is extended one helical turn beyond the final acidic residue of the negatively-charged face, and bears two hydrophobic residues in the final amphipathic turn. The ALL residues were added to mimic this and the final three Q residues were added to stabilise the carboxy-terminal end of the helix.* [166]

Previous work in this field has demonstrated that artificial transcriptional activators can perform as well as their natural counterparts as long as key DBD and AD components are present. Dervan et al. synthesised a synthetic hairpin polyamide composed of *N*-methylpyrrole and *N*-methylimidazole amino acids known to bind to the minor groove of DNA and fused it to the amphipathic helix (AH) activating domain described above. This molecule was able to stimulate promoter specific transcription by this miniature artificial transcription activator *in vitro* to 40 fold above the basal level. [167] The DBD and AH domains were initially linked by a dimerisation domain from yeast GCN4 protein (residues 251-281) known to form coiled-coil structure, but the distance between the DBD and AH was also varied by the use of a flexible polyether linker. The length of the linker domains were discovered to affect the degree of activation; longer linkers resulted in low levels of transcription in a yeast nuclear extract assay, likely

because of the flexibility of the linker moiety projecting the AH too far away from DNA. Removing the linker domain entirely gave 25% of the activity of the DBD/GCN4 dimerisation domain/AH construct, suggesting that either the dimerisation domain was important to the efficiency of activation or spatial separation of the activation module from DNA also plays a role in the efficiency of activation. [167] Dervan *et al*. later replaced the flexible linker with rigid oligo-prolines of varying incremental lengths (18-45 Å) [184] and found the optimal length of the linker domain lay between 36-45 Å (12-15 proline residues), where the linker length of 12 proline residues (36 Å) generated the largest degree of activation.

Photocontrollable Homeodomain Artificial Transcription Factors (PATFs) were designed to consist of helix-3 of engrailed homeodomain with cysteine residues in order to attach an azo-benzene crosslinker at *i*, *i* + 11 spacing linked to an AH helix sequence by two different types of linker, a flexible MM8 linker or a rigid oligo-proline linker (PP) (**Table 3.2.1.2**) to assess the effect of the linker on the degree of activation (**Figure 3.2.1.2 and 3.2.1.3**). [185, 45]

Name	Sequences
PATF-1 *i, i*+11	EAQCKIAAKNARAKCKKA-GSGGTGGGS-PEFPGIELQELQELQALLQQ
PATF-2 *i, i*+11	EAQCKIAAKNARAKCKKA-PPPPPPPPPPPP-PEFPGIELQELQELQALLQQ

Table 3.2.1.2. *Amino acid sequences of PATF peptides with flexible linker (PATF-1) and rigid linker (PATF-2).*
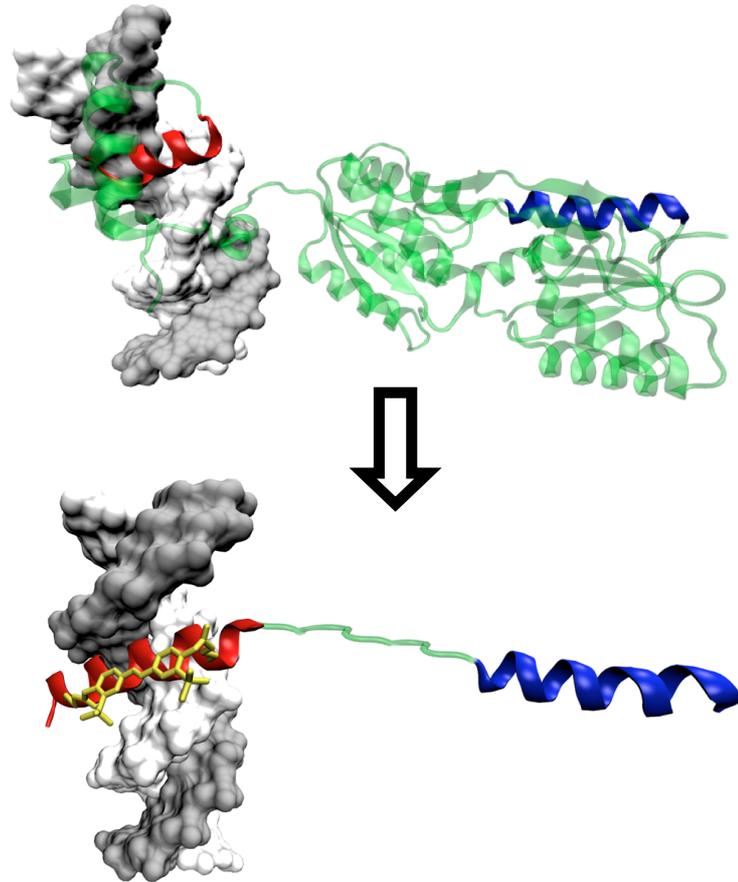
Figure 3.2.1.2. *Cartoon representation of PATF (top) with all the wild type components engrailed homeodomain HDH3 (red) as DNA binding domain, an activating domain AH (blue), and a linker connecting the two domains together (green).*
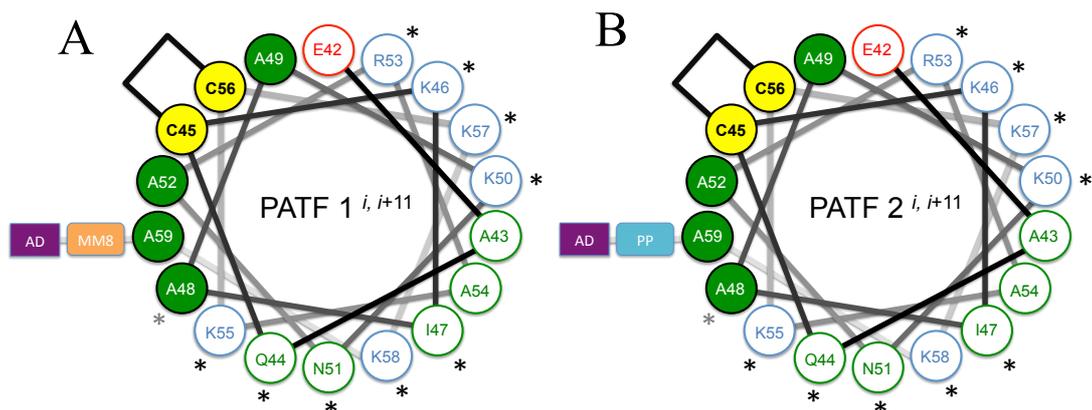
Figure 3.2.1.3. *Schematic of PATFs. A) Diagram of the helical wheel of cross-linked PATF-1 with i, i+11 spacing and MM8 flexible linker, B) Diagram of the helical wheel of cross-linked PATF-2 with i, i+11 spacing and oligo-proline linker (PP). Residues that are in contact with DNA are marked with asterisk, blue and red circles represent charged residues based on parent HDH-3 DNA binding domain. Residues that have been substituted in order to avoid steric clash with the cross-liner are marked in filled green circles.* [154, 175]

3.2.2 Construction of of PATF-1 and PATF-2

The oligonucleotides encoding PATF were purchased from Eurofins (pre-phosphorylation of oligonucleotide procedures are discussed in Chapter 2, Section 2.1.4 and Section 2.2.8). The oligonucleotides (2 µl, 25 pmol, *for1:for2* & *rev1:rev2* in separate tubes) were phosphorylated at the 5' end using T4 polynucleotide kinase (2.5 µl, 10,000 U ml^{-1}) in TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 25 °C). The solutions were incubated at 37 °C for 60 min. The oligos were then annealed by mixing equimolar quantities of forward and reverse oligos in a heat block at 95 °C. After 5 min at 95 °C, the heat block was turned off and solution was left to cool to room temperature.

The phosphorylated oligonucleotides (6 µl, 0.078 pmol) were ligated into a pET31b(+) vector (1:3 vector to insert ratio) *via* ligation of the desired gene with *Alw*NI digested dephosphorylated pET31b(+) (2 µl, 0.026 pmol) by mixing them together with T4 DNA ligase (0.5 µl) and 1 × ligase buffer (200 mM Tris-HCl, pH 7.6, 50 mM MgCl$_2$, 100 mM DTT, 10 mM ATP). The mixture was then incubated at 16 °C overnight and the resulting construct was used to transform *E. coli* XL1 Blue cells. The clones were checked by either re-digestion of plasmids with *Alw*NI to detect the removal of *Alw*NI restriction site upon successful insertion, or by PCR analysis (using T7 *prom* and *term* as primers) and then by DNA sequencing (**Figure 3.2.2**).
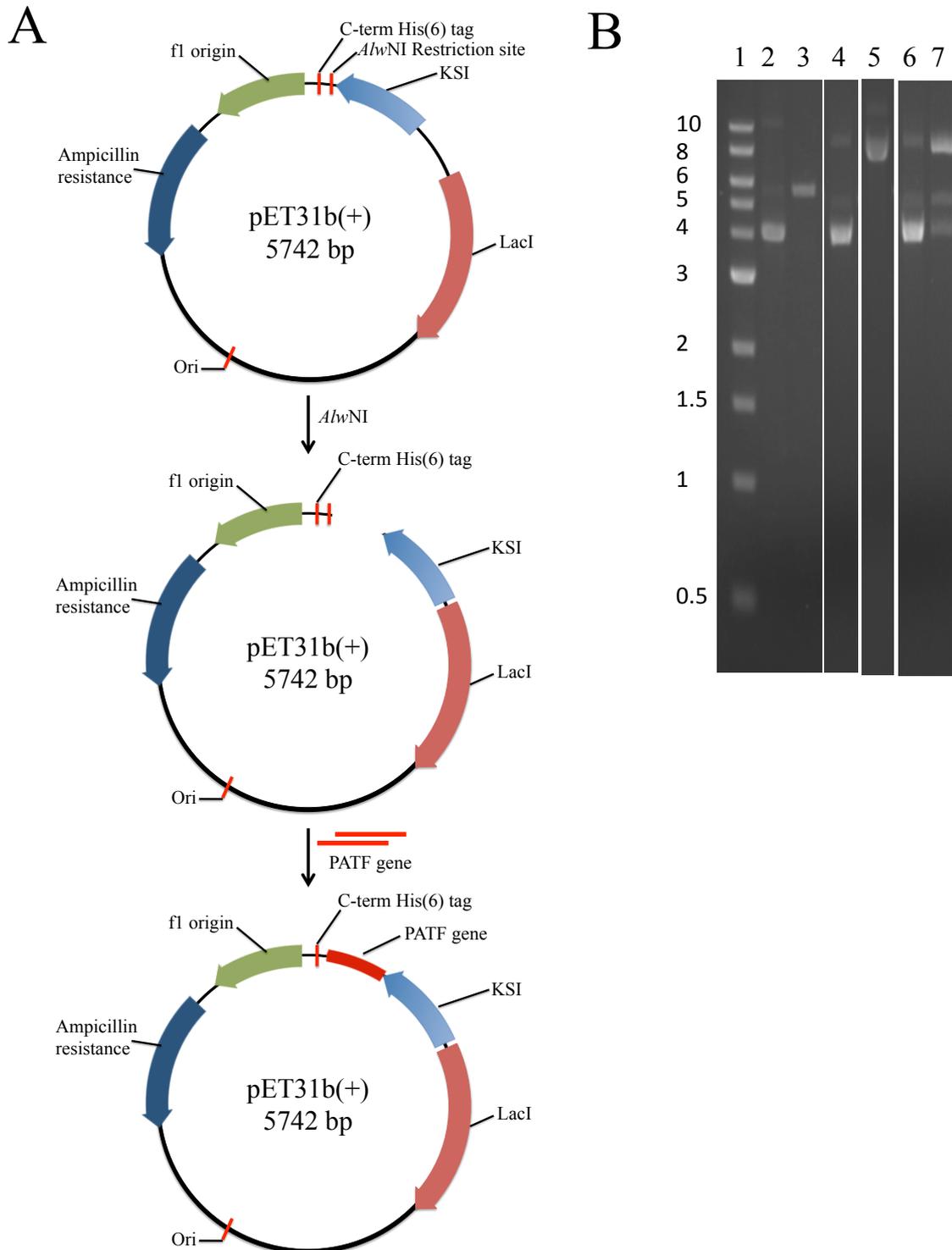
Figure 3.2.2. *A) Map of the pET31b(+) expression vector and cloning charts of PATF-1 and PATF-2 peptides. B) Agarose gels stained with ethidium bromide. Lane 1: 1 kb DNA ladder (fragment size is indicated in kbp), lane 2: circular pET31b(+), lane 3: empty AlwNI digested pET31b(+), lane 4: circular plasmid containing PATF-1 gene, lane 5: plasmid containing PATF-1 gene after AlwNI treatment (undigested), lane 6: circular plasmid containing PATF-2 gene, lane 7: plasmid containing PATF-2 gene after AlwNI treatment (undigested).*

Sequencing results of chosen clones have confirmed the desired sequence for PATF-1 (**Table 3.2.2**), however this was not the case for PATF-2 (**Table 3.2.2**) where one proline residue within the rigid oligo-proline linker was missing, making this an eleven-proline linker rather than the initial twelve. This type of error can arise when there are high numbers of the same amino acids in a row. This could be due to the high amount of codon usage for a particular amino acid, where by having large quantity of residue of the same sequence, this can result in errors or some sequences not being translated. [186] The other alternative explanation could be due to the repetitive nature of the proline sequences used in combination with polymerase slippage leading to some proline residues not being translated. Despite this error, the rigid linker of PATF-2 is still within the optimal length for the high degree of transcription activation as described previously. [184] The PATF-2 plasmid containing the missing proline residue was therefore used for protein expression (**Table 3.2.2**).

Name	Sequences
PATF-1 i, i+11	gaagcgcagtgtaaaatcgcggctaagaacgcacgtgcgaaatgcaagaaagctggatccggtggcactggt
	E A Q C K I A A K N A R A K C K K A G S G G T G
	ggcggttctccggaattcccgggtattgagctgcaagaactgcaagagctgcaggccctcctgcagcaa
	G G S P E F P G I E L Q E L Q E L Q A L L Q Q
PATF-2 i, i+11	gaagcgcagtgtaaaatcgcggctaagaacgcacgtgcgaaatgcaagaaagctccgccgccaccgcctccg
	E A Q C K I A A K N A R A K C K K A P P P P P P
	ccaccgccgccaccgccggaattcccgggtattgagctgcaagaactgcaagagctgcaggccctcctgcag
	P P P P P P E F P G I E L Q E L Q E L Q A L L Q
	caa
	Q

Table 3.2.2. *Sequencing results of PATF-1 and PATF-2 clones. The obtained sequences are in minuscule letters (top) and the translated sequences are in capital letters (bottom).*

3.2.3 Expression and purification of PATFs

To produce proteins, *E. coli* BLR(DE3) cells were transformed with a pET31b(+) plasmid harbouring a gene encoding the desired PATF product. The PATF gene is under the control of a promoter only recognised by T7 polymerase, which is itself under the control of a modified *lac* promoter (*lac*UV5). Isopropyl-β-D-1-thiogalactopyranoside (IPTG) was added to growing cells in LB media, to allow expression of the T7 RNA polymerase gene, and, in turn, PATF. [1, 187] IPTG addition once cultures reached an OD_{600} of 0.6 rapidly induced synthesis of a fusion protein between bacterial ketosteroid isomerase (KSI) and PATF with a C-terminal His-tag. (**Figure 3.2.2.1**).
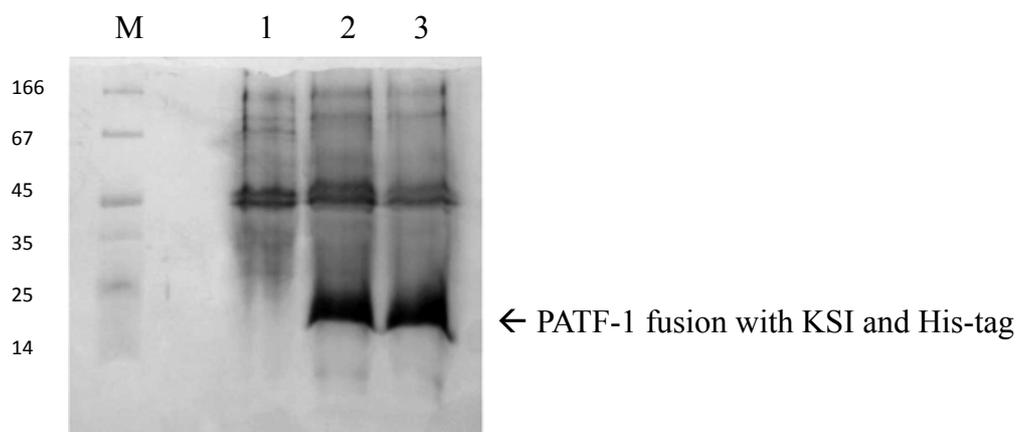
Figure 3.2.3.1. *SDS-PAGE analysis of the initial stages of PATF-1 purification. M: Protein marker (Fermentas), lane 1: Total cellular protein with no induction (negative control), lane 2: Total cellular protein 2 hour after induction, lane 3: Total cellular protein 5 hours after induction.*

Following 5 hours of expression, the cells were harvested by centrifugation, resuspended in lysis buffer and lysed by ultrasonication for 20 minutes (pulse on for 5 seconds, pulse off for 5 seconds) at 40% intensity. The suspension was clarified by centrifugation (20 mins, 4 °C, 9,000 × g), then the supernatant was removed and the pellet resuspended in further lysis buffer with added sodium chloride (to 1 M). After

stirring for 2 hours the solution was centrifuged (20 min, 4 °C, 12,000 × g), and the supernatant discarded. The pellet was dissolved at room temperature in guanidine hydrochloride (6M), stirred overnight and centrifuged (30 min, 4 °C, 25,000 × g) to remove membrane debris. The supernatant was dialysed against H_2O (3 L) using a 12-14 kDa cut-off dialysis membrane at 4 °C overnight and centrifuged (20 min, 4 °C, 12,000 × g). The pellet was then redissolved in 80% formic acid (10 ml) and cyanogen bromide (0.3 g) was added to cleave the His-tag and KSI protein by converting the methionine residues to homoserine with concurrent breakage of the following amide bond (**Figure 3.2.3.2**). The solution was stirred in a foil-wrapped flask under nitrogen for 18-20 hours.
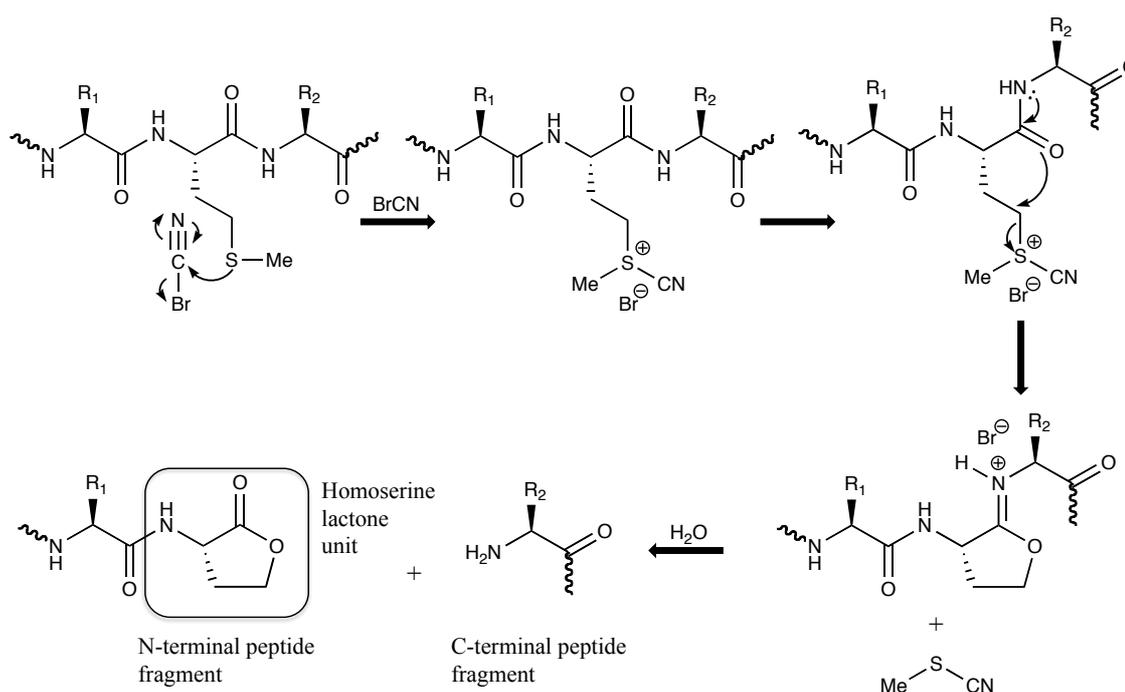
Figure 3.2.3.2. *The mechanism of peptide cleavage by cyanogen bromide.* [188]

The solvent was evaporated under reduced pressure and the product proteinaceous film was resuspended in acetonitrile/water (1:1). The mixture was stirred for 1 hour and centrifuged (20 min, 4 °C, 25,000 × g). The supernatant solution was purified by semi-preparative reverse phase HPLC (Phenomenex Gemini C_{18} column) eluting with 10-

100% acetonitrile/0.1% TFA, H_2O/0.1% TFA gradient. The desired product peak was identified by MALDI-TOF, then corresponding fractions were pooled and freeze-dried. Reinjection of pure PATF peptides revealed that PATF-1 eluted at approximately 19 minutes, whereas PATF-2 eluted at approximately 21 minutes using a 10-100% acetonitrile:water (containing 0.1 % TFA) linear gradient over the course of 60 minutes at a flow rate of 5 ml min^{-1} (**Figure 3.2.3.3**). PATF-1 differs in retention time compared to PATF-2 due to its more hydrophilic linker domain resulting in faster elution rate. The masses of PATF-1 and PATF-2 were measured by MALDI-TOF MS (**Table 3.2.3.1**).
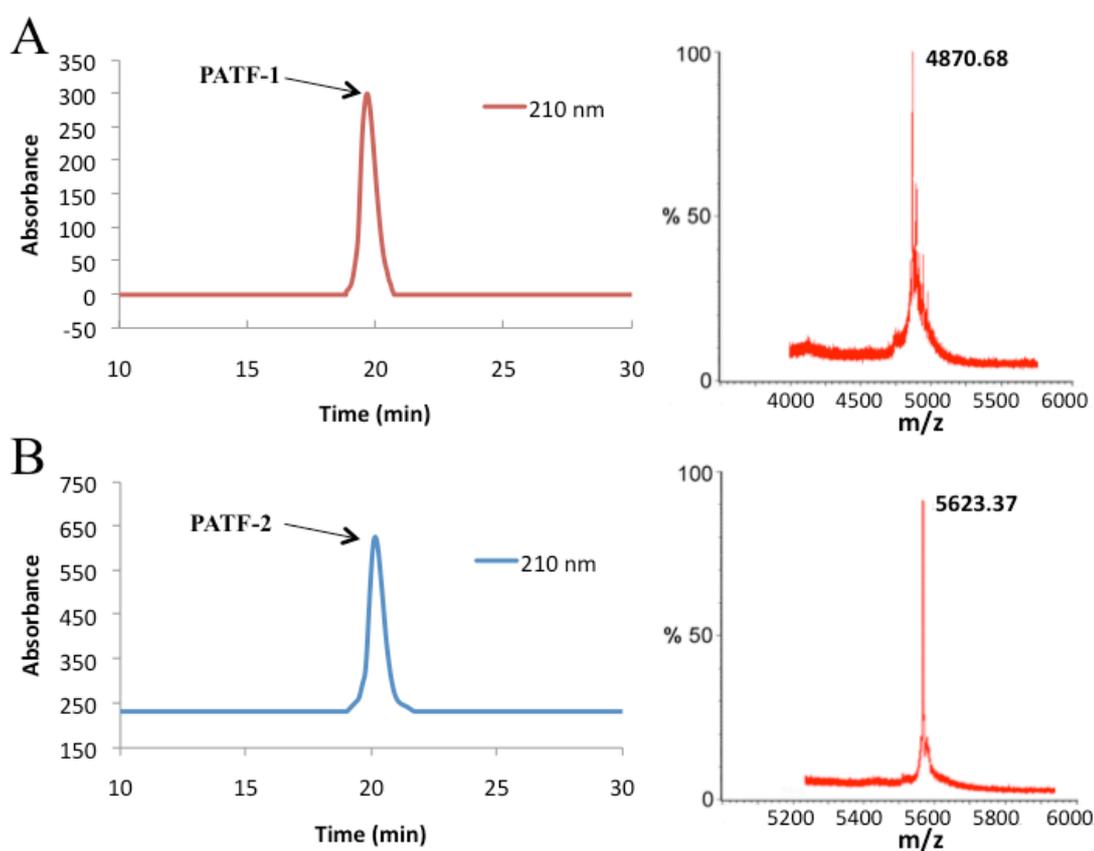
Figure 3.2.3.3. *Chromatogram from the analytical HPLC analysis and MALDI-TOF mass spectrum of A) PATF-1 and B) PATF-2.*

PATF	Monoisotopic calculated mass	Observed m/z
PATF-1	4936.54	4870.68
PATF-2	5386.88	5623.37

Table 3.2.3.1. *Monoisotopic calculated masses and observed masses of PATF-1 and PATF-2 peptides.*

The masses observed for PATF-1 and PATF-2 were deviated from the monoisotopic calculated masses by approximately 65.86 and 236.49 respectively. The reasons for this deviation stems from the fact that the MALDI-TOF mass spectrometer may have not been not calibrated properly before each measurement. It has also been suggested that the room temperature where the mass spectrometer is located can greatly affect the measurement. Fluctuation in the observed mass could be seen as the temperature of the room increases. Despite this deviation, we can assume that the correct peptides were obtained due to the confirmation from the sequencing results.

After CNBr cleavage and purification *via* HPLC initial protein expression yielded only 2 mg of PATF-1from 6 × 550 ml of media after 5 hours of expression at 37 $^{\circ}$C. Three different DE3 cell lines were used to express the protein to attempt to improve the yield. PATF-1 proteins were expressed in BL21 (DE3) CodonPlus RIL, CodonPlus RP and Star lines (**Figure 3.2.3.4**). Expression of PATF-1 in RIL and RP from 6 × 550 ml culture resulted in 1-2 mg of PATF-1. Expression of PATF-1 in BL21 Star resulted in approximately 4 mg of PATF-1 from 3.3 litres of culture medium, after HPLC purification. Comparing Star cells to RIL and RP by SDS gel electrophoresis (**Figure 3.2.3.4**), it seemed that more than 4 mg of proteins would have been present and that perhaps the problem was not from the level of expression but the method that was used to extract the protein.

After CNBr cleavage and solvent evaporation, the gelatinous substance was initially dissolved in acetonitrile/water (1:1). However, the peptide was poorly soluble in this mixture, acidified (0.1 % TFA) acetonitrile/water (2:3) was used, according to manufacturer's recommendation. This, however, did very little to improve the final yield of the protein. Another possibility was that the insoluble gel of cleaved KSI protein physically entrapped some of the PATF-1 proteins, rendering it insoluble in acetonitrile/water solvent. It was decided to dissolve the gelatinous substance in 8 M urea and dialyze the substance in a 500 Da cut off membrane leaving, in theory, the PATF-1 in the membrane. Unfortunately, the HPLC trace (not shown) revealed no trace of PATF-1 from this method. It was therefore decided to continue expressing PATF-1 using BL21(Star) cells. PATF-2 peptide was also expressed in BL21(Star) cells (**Figure 3.2.3.5**).
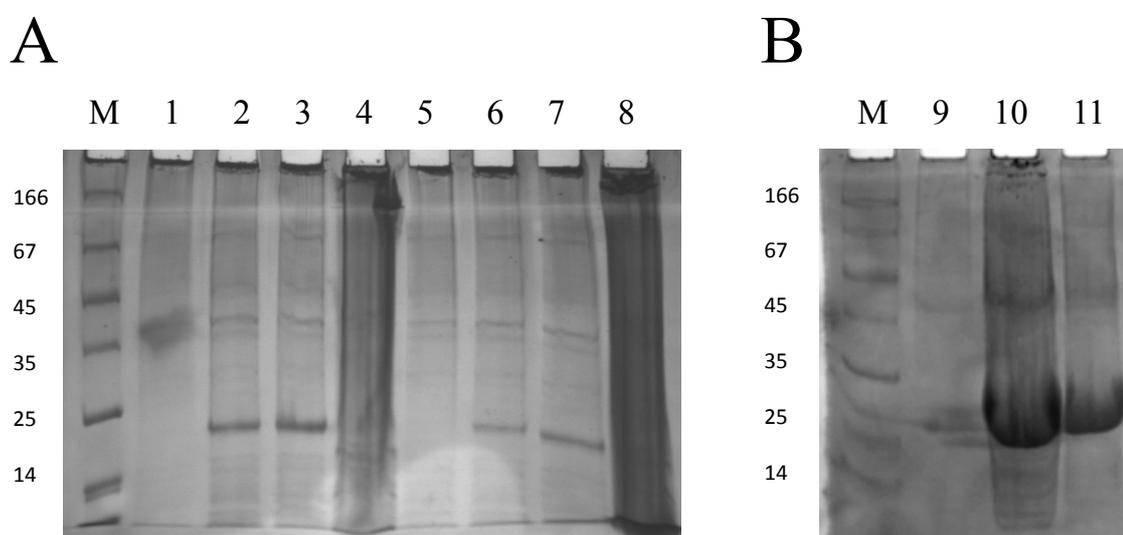
Figure 3.2.3.4. *SDS-PAGE analysis of PATF-1 expression in A) RIL and RP cells and B) Star cells. M: Protein marker (fragment size is indicated in kDa, Fermentas), lane 1: No induction RIL, lane 2: 2 hours induction RIL, lane 3: 5 hours induction RIL, lane 4: Overnight induction RIL, lane 6: 2 hours induction RP, lane 7: 5 hours induction RP, lane 8: Overnight induction RP, lane 9: No induction Star, lane 10: 2 hours induction Star, lane 11: 5 hours induction Star.*
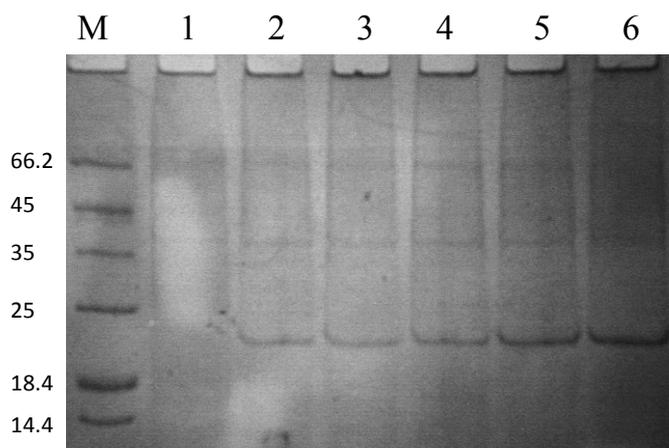
Figure 3.2.3.5. *SDS-PAGE analysis of PATF-2 expression in Star cells. M: Protein marker (fragment size is indicated in kDa, Fermentas), lane 1: No induction PATF-2, lane 2: 1 hours induction of PATF-2, lane 3: 2 hours induction of PATF-2, lane 4: 3 hours induction of PATF-2, lane 5: 4 hours induction of PATF-2, lane 6: 5 hours induction of PATF-2.*

3.2.4 Synthesis of the 3,3'–bis-(sulfo)-4,4'-bis (chloroacetamino)azobenzene (BSBCA) crosslinker

The 4 step synthesis of the 3,3'–bis-(sulfo)-4,4'-bis (chloroacetamino)azobenzene (BSBCA) crosslinker was carried out according to literature procedures. [146] 2,5-diaminobenzene sulfonic acid was acetylated with acetic anhydride in hot (94 °C) acetic acid. The monoacetylated product was then oxidised to an azobenzene using sodium hypochlorite solution (>4% chlorine availability). The de-acetylation of the azobenzene compound was performed using concentrated hydrochloric acid and water, where the mixture was heated under reflux for 2.5 hours, reduced to dryness followed by treatment with dilute sodium hydroxide. Finally, the azobenezene was chloroacetylated by chloroacetic acid and chloroacetic anhydride to produce the BSBCA crosslinker.
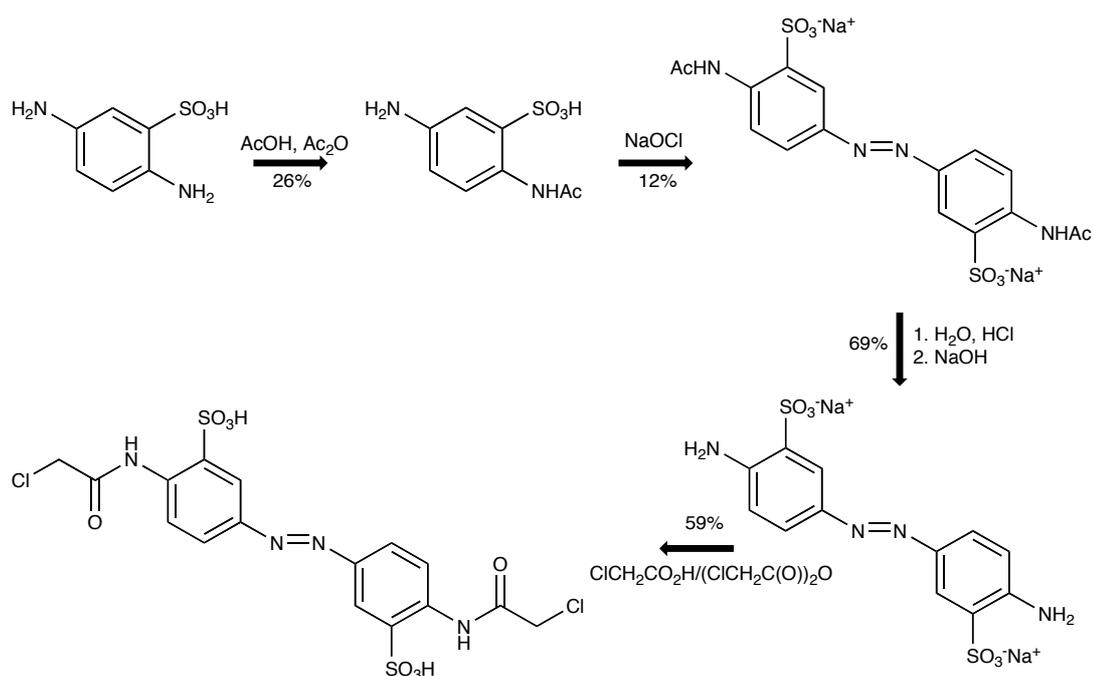
Figure 3.2.4. *3,3'-bis(sulfo)-4,4'-bis(chloroacetamido)azobenzene synthesis.* [146]

Products were characterised by ^{1}H and ^{13}C NMR spectroscopy and mass spectrometry and the data obtained were concordant with literature values.

3.2.5 Cross-linking of PATF-1 and PATF-2 with BSBCA crosslinker

PATF cysteine sidechains were alkylated with BSBCA by a nucleophilic substitution reaction. At pH 8.3 the sulfhydryl groups are mostly deprotonated and the resulting thiolate groups act as nucleophiles towards the crosslinker methylene groups (**Figure 3.2.5.1**).

The HPLC method used for PATFs was reapplied for the crosslinked (XL-PATFs) peptides. The HPLC chromatographs show 360 nm absorption from the *trans* form of BSBCA coinciding with peptide peaks at 18 minutes for PATF-1 and around 19 minutes for PATF-2 (**Figure 3.2.5.2**). The identity of the species were established by MALDI-TOF MS (**Table 3.2.5.1**). Similar deviation in masses by approximately 65.64

and 236.48 from the calculated monoisotopic masses of XL-PATF-1 and XL-PATF-2 respectively were also observed. It was also assumed that this deviation stems from the fact that the MALDI-TOF mass spectrometer may have not been not calibrated properly before each measurement and the room temperature fluctuation may have also contributed to this deviation. Despite this deviation, we can also assume that the correct peptides were obtained due to the confirmation from the sequencing results.
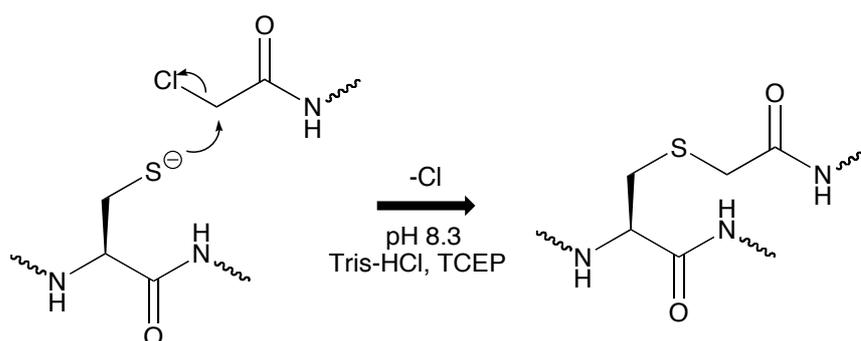
Figure 3.2.5.1. *Alkylation mechanism.*

XL-PATF	Calculated monoisotopic mass	Observed mass
XL-PATF-1	5388.55	5322.91
XL-PATF-2	5838.89	6075.37

Table 3.2.5.1. *Monoisotopic calculated masses and observed masses of XL PATF-1 and XL PATF-2 peptides.*
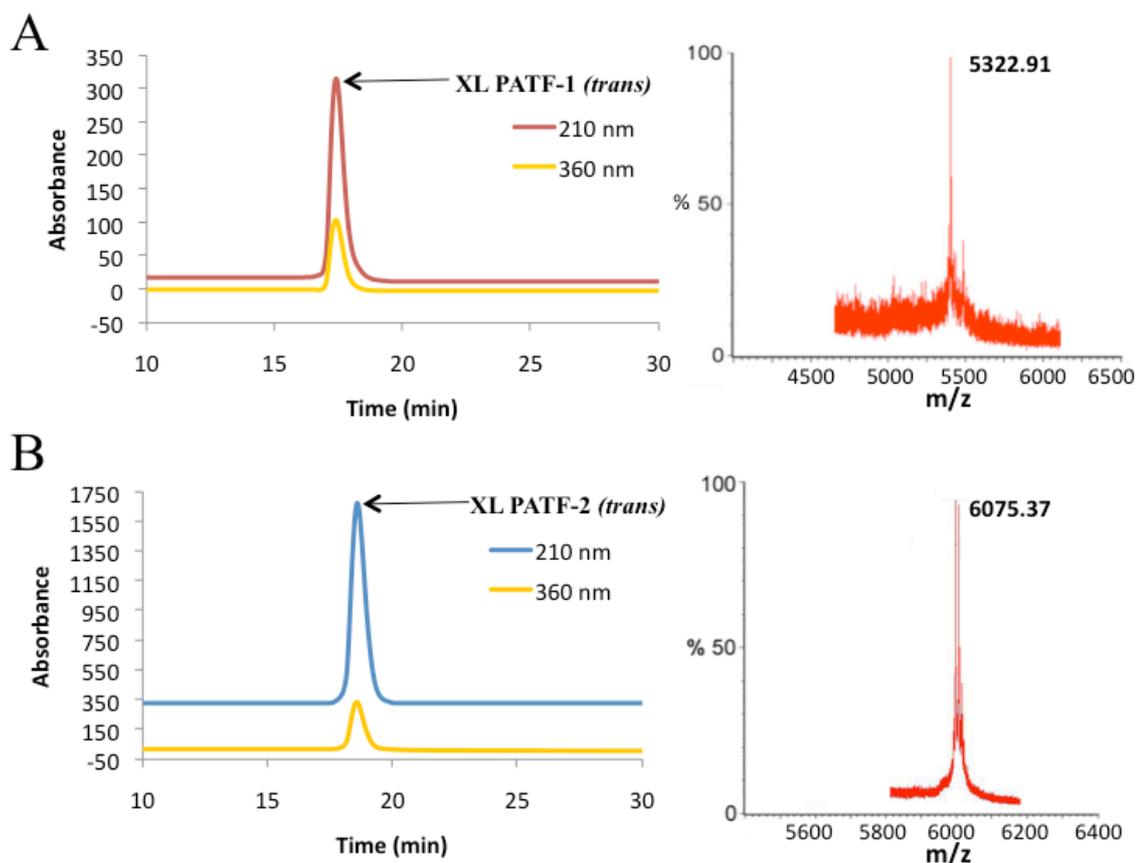
Figure 3.2.5.2. *Chromatogram from the analytical HPLC analysis and MALDI-TOF mass spectrum of A) XL PATF-1 and B) XL PATF-2*.

3.2.6 Spectroscopic characterisation of XL-PATF-1 and XL-PATF-2

3.2.6.1 UV-Visible absorption

The change in conformation of the crosslinker is readily observed by UV spectroscopy. Experiments at different temperatures were used to characterise the thermal isomerisation of irradiated BSBCA crosslinker, PATF-1 and PATF-2 from *cis* to *trans*. Both crosslinker and peptides were irradiated for 6 minutes under UV light (280 W lamp with 360 +/- 10 nm filter). The absorption spectrum of dark-adapted PATFs, where the azobenzene crosslinker is in the thermally stable *trans*-configuration were characterised by strong a maximum at 363 nm typical of the π- π* transition in amide-substituted *trans*-azobenzenes. Irradiation with 360 nm light led to a significant

reduction of this maximum. The difference in absorption between the *cis* and *trans* conformations can be clearly seen (**Figure 3.2.6.1.1**).

Irradiated crosslinker and XL PATF peptides revert to the dark-adapted state in a non-photochemical process characterised by isosbestic points at 250, 317 and 434 nm, suggesting the presence of only two species in equilibrium. Assuming first order kinetics for the thermal relaxation *cis*-to-*trans* of the peptides, the rate constant *k* for the thermal relaxation process was calculated by plotting the natural logarithm of the percentage of *cis*-isomer (calculated from the absorbance at 363 nm) versus time (**Figure 3.2.6.1.2**). This allowed subsequent calculation of the half-life (t½) of the *cis*-isomer (**Chapter 2, section 2.5**). The half-lives of XL PATF *cis*-to-*trans* thermal isomerisation were found to be 88 min and 55 min at 15 °C for XL PATF-1 and XL PATF-2 respectively (**Table 3.2.6.1.1**). The PATFs unexpectedly have significantly shorter half-lives than the HDH-3 DBD peptide upon which they are based (150 min at 15 °C). The presence of the linker and AD must cause an increase in the relaxation rate. In particular, significant changes in relaxation rates for XL PATF-1 compared to XL PATF-2 at different temperatures suggest the different linkers exert a strong influence.
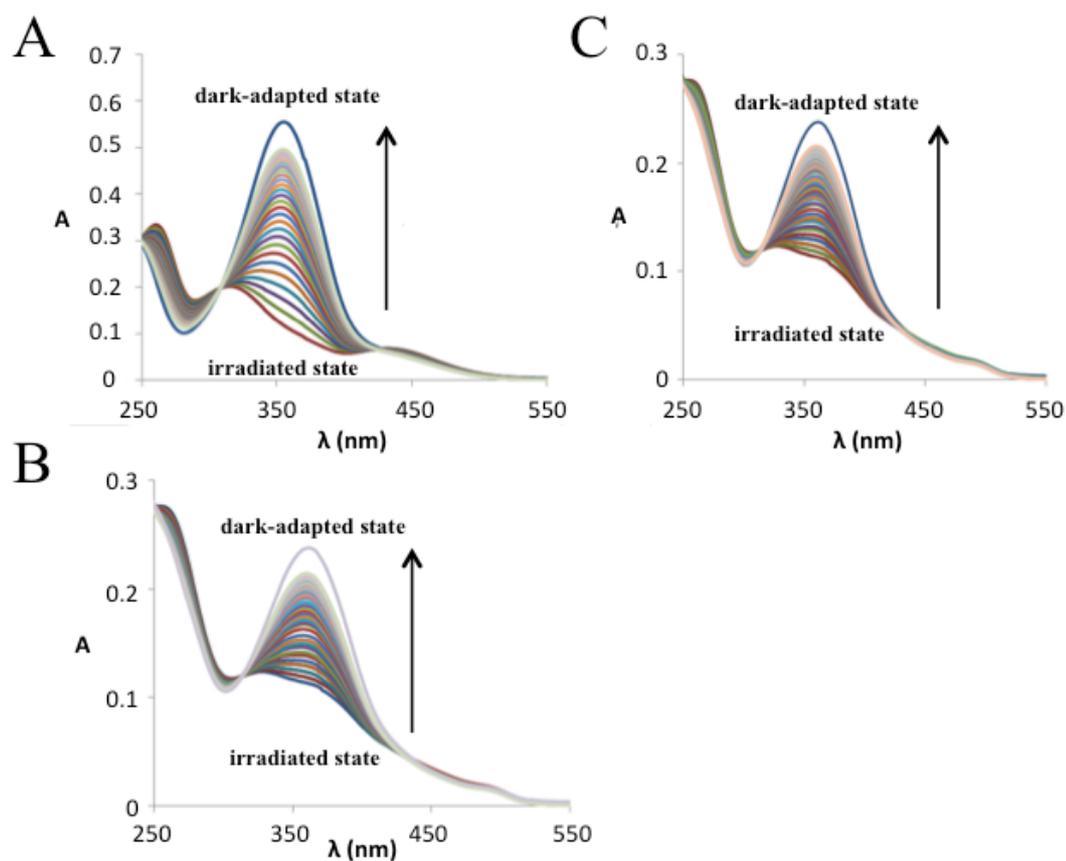
Figure 3.2.6.1.1. *UV/Vis spectra of A) crosslinker, B) XL PATF-1 and C) XL PATF-2 at 25 °C. Spectra were acquired on the dark-adapted state (trans), the irradiated state (cis) and 2 minute intervals after irradiation.*

Previous studies on a modified pro-apoptotic peptide $Bak_{72\text{-}87}{}^{i,i+11}$ have shown that more helical peptides have significantly shorter half-lives compared to less helical peptide HDH-3. [189] It was also suggested that high α-helical content of the stabilised peptide form presumably leads to very high relaxation rates in the case of *i, i+ 11* spacings and slower reversion rate in other spacings. [190]

A) $\ln (A-Ao) = -0.039\ T - 0.8055$

B)
$\ln (A-Ao) = -0.020\ T - 2.0414$
$\ln (A-Ao) = -0.038\ T - 0.8232$
$\ln (A-Ao) = -0.088\ T - 2.2145$

C)
$\ln (A-Ao) = -0.012\ T - 2.1553$
$\ln (A-Ao) = -0.026\ T - 2.1540$
$\ln (A-Ao) = -0.045\ T - 2.0827$
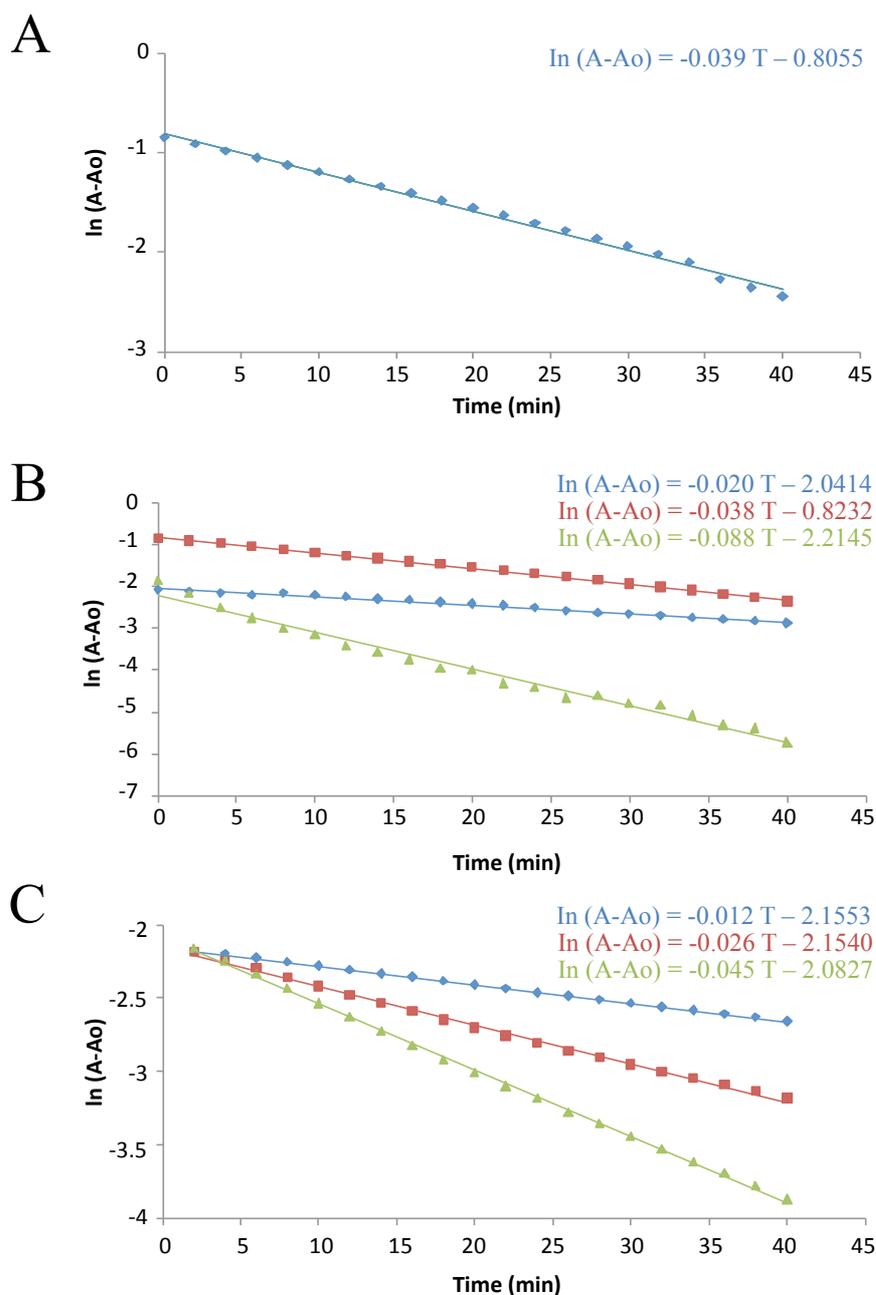
Figure 3.2.6.1.2. *Relaxation of different PATF peptides as a function of time in the dark as monitored by UV (363 nm). A) BSBCA crosslinker, B) XL PATF-1 and C) XL PATF-2. Blue: 15 ^oC, Red: 25 ^oC Green: 37 ^oC.*

Three different conditions were selected to determine the activation energy of reversion of XL PATF peptides, and the temperature dependence of the rates compared in an Arrhenius plot ($\ln k$ versus $1/T$) (**Chapter 2, section 2.5**). Isomerisation was followed by UV spectroscopy (363 nm) at 2 minute intervals for a period of 40 minutes and the

half-life ($t_½$) was calculated. The thermal relaxation experiments were performed in triplicate for each of the peptide. These data were used to construct Arrhenius plots of the XL PATF peptides which showed similar activation energies for both peptides (**Figure 3.2.6.1.3**) (**Table 3.2.6.1.2**). The isomerisation of the peptides proved to be fully reversible and temperature dependant as expected. The half-life of isolated crosslinker was 35 min at 15 °C (**Figure 3.2.6.1.1**) (**Figure 3.2.6.1.2**), a value concordant with previously reported values. [175]

Peptide	$t_½$ at 15 °C (min)	$t_½$ at 25 °C (min)	$t_½$ at 37 °C (min)
XL PATF-1	88 +/- 4.0	32 +/- 3.1	18 +/- 1.5
XL PATF-2	55 +/- 3.6	26 +/- 3.1	15 +/- 2.5

Table 3.2.6.1.1. *Half-life of the cis-isomer for PATF-1 and PATF-2 peptides.*
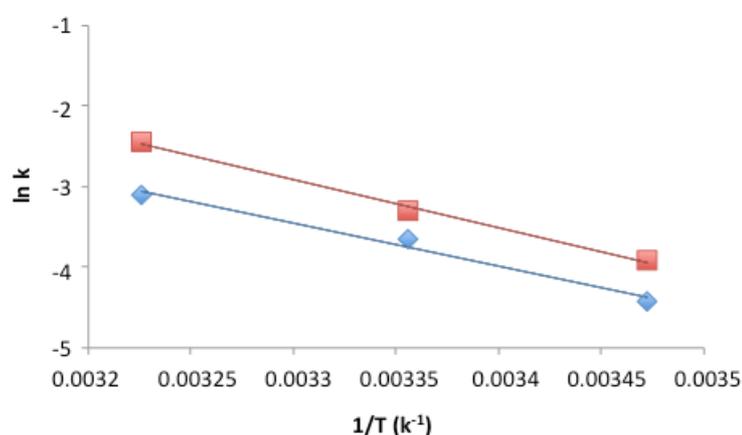
Figure 3.2.6.1.3. *Arrhenius plot of the UV relaxation rates of XL PATF-1 (blue) and XL PATF-2 (red).*

Peptide	E_a (kJ mol^{-1})	A (s^{-1})
XL PATF-1	49.7 +/- 0.5	1.99×10^7 +/- 1.19×10^4
XL PATF-2	44.4 +/- 1.1	1.43×10^6 +/- 4.13×10^4

Table 3.2.6.1.2. *Arrhenius parameters for the relaxation of different XL PATF peptides.*

3.2.6.2 Circular Dichroism (CD) spectroscopy of PATF peptides

CD spectra were measured to analyse the secondary structure of XL PATF peptides and observe changes induced by photoswitching. A typical CD spectrum of an α-helical protein possess a maximum at 190 nm with minima of approximately equal intensity at 208 and 222 nm. Dark-adapted XL PATF-1 gave a spectrum resembling that expected for a helical structure (**Figure 3.2.6.2.1**). [36] In contrast to the HDH-3 DBD peptide, the minimum at 208 nm is more intense due to a greater contribution from random coil conformations (**Figure 3.2.6.2.2**). Upon irradiation, the CD value at 190 nm clearly decreases whilst that at 222 nm increases and that at 208 remains roughly the same, suggesting that the α-helical contribution has diminished with more random coil present. The observed mean residue ellipticities at 222 nm of -50445 deg cm^2 $dmol^{-1}$ for dad- XL-PATF-1 and -40035 deg cm^2 $dmol^{-1}$ for irrad- XL-PATF-1 correspond to approximately 13.8% and 11.3% helicities respectively, which is a smaller percentage than found in previous work carried out on XL HDH-3 (40 % helicity for dad- and 18.6% helicity for irrad), [154] suggesting that the linker and AD attached to the DBD contribute to a decrease in overall helicity and reduced structural changes upon irradiation of XL-PATF-1 with 360 nm UV light. Interestingly, the observed mean residue ellipticities at 222 nm of -38658 deg cm^2 $dmol^{-1}$ and -38642 deg cm^2 $dmol^{-1}$ for uncrosslinked PATF-1 and PATF-2 correspond to approximately 9.9% and 9.8% helicities respectively are similar to their irradiated state, suggesting less helical conformations.
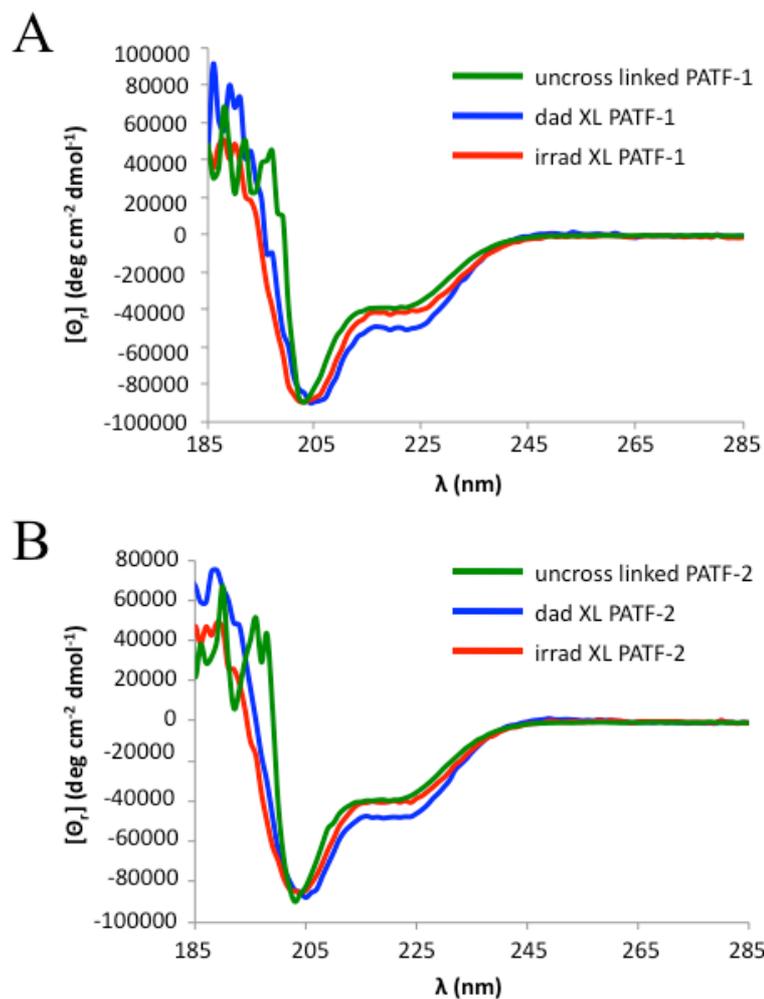
Figure 3.2.6.2.1. *CD spectra of dad-(blue), irrad-(red) and uncrosslinked-(green) A) PATF-1 and B) PATF-2 peptides carried out in 5 mM potassium phosphate (pH 8.0) at 25 ^oC.*

CD spectra of XL PATF-2 were similar to those of XL PATF-1, upon irradiation, the CD value at 190 nm decreases whilst that at 222 nm increases and that at 208 remains roughly unchanged (**Figure 3.2.6.2.1**), suggesting that the use of different linker has little or no effect on the overall structure of the peptide. It is also important to mention that polyproline can adopt two structured conformations; polyproline type I (PPI) and polyproline type II (PP2) helices. PPII helices form predominantly short stretches of regular structure characterised by the repetitive pairs of Ramachandran ϕ and ψ angles. The PPII helix is an extended (3.1 Å per residue compared to 1.5 Å in the α-helix) left-handed helix defined by the ϕ, ψ torsional angle cluster with the distribution maximum

at -75 ° and 145 °. [191] The right-handed polyproline I (PPI) on the other hand is less common since the prolines are in higher energy *cis* conformation than its *trans* PPII counterpart. PPI also has similar dihedral angles (-75 °, 160 °) to PPII. [192] Studies of polyproline helices by circular dichroism has shown that they result in similar spectra to those of unfolded proteins with minima at approximately 198 nm and maxima of 218 nm. [191] The observed mean residue ellipticities for XL PATF-2 at 222 nm of -47961 deg cm^2 $dmol^{-1}$ for dark adapted XL-PATF-2 and -40272.5 deg cm^2 $dmol^{-1}$ for irrad- state correspond to approximately 13.1% and 11% helicities respectively, extending the similarity between XL PATF-1 and XL PATF-2 and reinforcing that the cross-linker has a relatively subtle effect on the overall structure of both peptides upon irradiation.
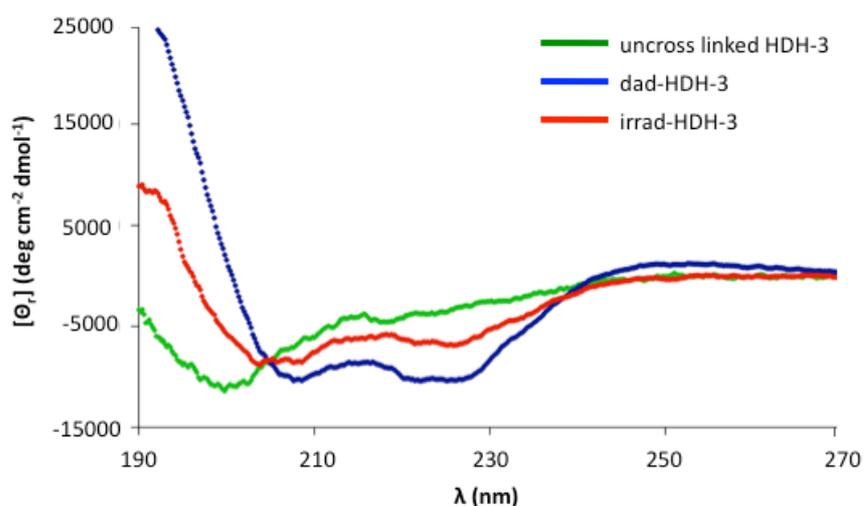
Figure 3.2.6.2.2. *CD spectra representation of uncross linked-(green), dad-(blue) and irrad-(red) crosslinked HDH-3 peptide.* [175]

Structural changes of XL PATF-1 can also be induced by addition of 2,2,2-trifluoroethanol (TFE), which is known to stabilise α-helical structures. The use of TFE can be considered as a way of weakening nonlocal interactions but favouring local interactions of protein structures. [193, 194] The change in structure of dad- XL PATF-1

with addition of 5-20% co-solvent TFE was investigated using CD spectroscopy (**Figure 3.2.6.2.3**).
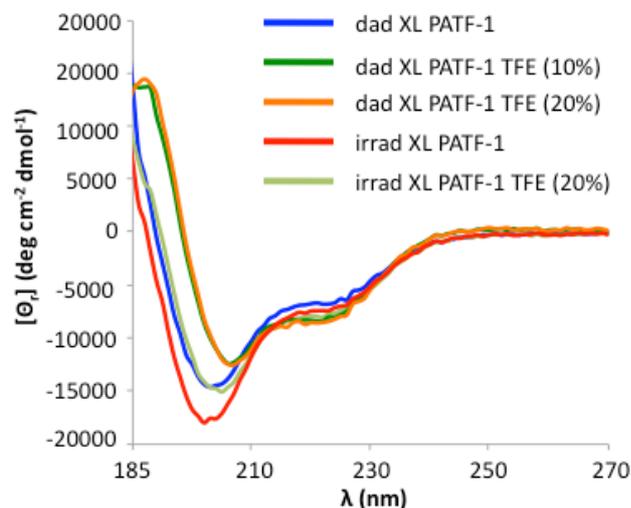
Figure 3.2.6.2.3. *CD spectra of dad- XL PATF-1 in the presence of TFE.*

Upon gradual addition of TFE the minima around 210 and 222 nm increased and the maximum at 190 nm increased, which is expected as random coil contribution to the CD spectrum is gradually reduced and helicity is increased. This gradual change was also observed in previous work on Bak and HDH-3 peptides. Even though a change in CD absorption is observed, the changes seem rather smaller than expected for a large change in protein conformation. This may suggest that crosslinked PATF-1 is close to achieving its maximal helicity in its dark state. Studies have demonstrated that TFE has little or no effect to the stabilisation of polyproline helices. Ndao *et al.* have demonstrated that by introducing TFE (10%-70%) to a recombinant porcine amelogenin (rP172) protein, which contains a polyproline type II (PPII) region, did not respond to TFE addition with conformational changes. [195] Furthermore, it is believed that the PPII residues in rP172 act as conformational stabilisers and offer resistance to TFE-induced folding.

3.2.7 DNA Binding Specificity measured by Fluorescence Anisotropy

Fluorescence anisotropy measurements were carried out to quantify the DNA binding affinities of the uncrosslinked PATF-1 and dark adapted and irradiated XL PATF peptides for DNA sequences containing QRE binding sites (**Table 3.2.7.1**). [169] QRE is known to contain the specific binding site for the Q50K engrailed homeodomain upon which the PATF peptides are based. [196, 197] Fluoresence anisotropy measures differences in the plane of excited and emitted light to determine differences in rates of rotation during the lifetime of excited dye molecules between free fluorescein-labelled oligonucleotides and oligonucleotide-peptide complexes. Pleasingly, in spite of the low molecular weight of the peptide compared to the dye-bearing oligonucleotide, which could make the assessment of change on anisotropy measurements difficult, the technique proved to be useful in allowing the determination of the dissociation constants (K_D) of the DNA complexes of dark-adapted and irradiated PATF-XL peptides.

Name	Sequences
QRE	5'-CGCAGTG<u>TAATCC</u>CCTCGAC-3'
	3'-GCGTCAC<u>ATTAGG</u>GGAGCTG-3'
ERE	5'-CGCAGTG<u>TAATTA</u>CCTCGAC-3'
	3'-GCGTCAC<u>ATTAAT</u>GGAGCTG-3'
MCK-S	5'-CAGGCAGCAGGTGTTGG-3'
	3'-GTCCGTCGTCCACAACC-3'

Table 3.2.7.1. *QRE, ERE and MCK-S DNA sequences. PATF peptides binding sequences are underlined.*

The data was fitted to the Langmuir isotherm equation (**Chapter 2, Section 2.7**) with no restriction on the number of binding sites present. The best fit of the experimental data for dark-adapted XL PATF-1 was obtained for a binding mode where one peptide is bound to one DNA molecule. Dark-adapted XL PATF-1 bound tightly to its QRE target site with an equilibrium dissociation constant of 43.2 nM at 15 °C (**Table 3.2.7.2**). Compared to the uncrosslinked PATF-1-QRE complex K_D of 702.2 nM at 15 °C (**Table 3.2.7.2**), the dark adapted XL PATF-1-QRE complex was nearly 20 times more stable. Unlike the XL PATF-1-QRE complex, PATF-1-QRE complex could not be properly fitted to the Langmuir isotherm equation assuming one binding site, the best fit to experimental data was obtained assuming that four PATF-1 peptides are simultaneously bound to QRE DNA; a low-affinity binding mode clearly distinct from the high-affinity binding mode observed for dad-XL PATF-1. Similar binding behaviour had previously been observed for the non-crosslinked HDH3-QRE complex.[154] The binding curve demonstrated a low binding degree up to approximately 400 nM followed by a relatively steep increase, suggestive of cooperative binding (**Figure 3.2.7.1**).
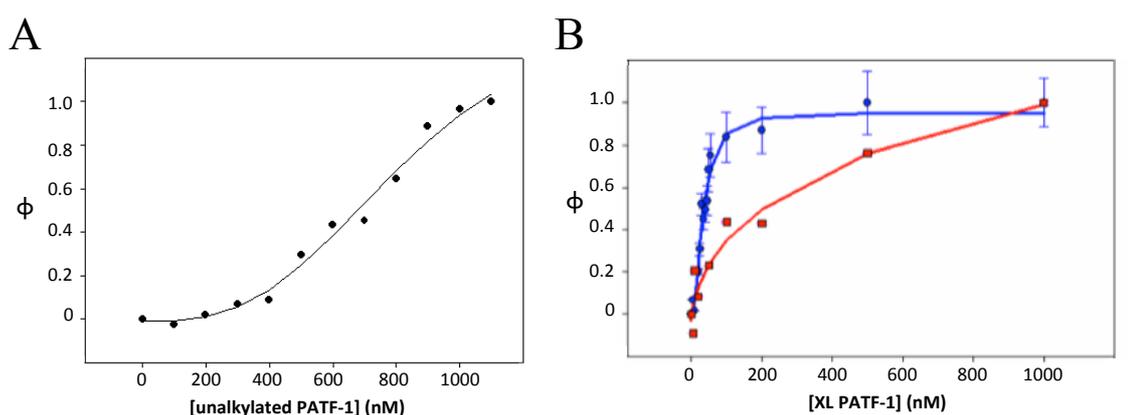
Figure 3.2.7.1. *Fractional occupancy curves obtained from fluorescence anisotropy measurements of the A) un-alkylated PATF-1 to QRE, B) dad- (blue) XL PATF-1 and irrad- (red) XL PATF-1 to QRE.*

XL PATF-1 does not bind as tightly as the reported affinity of the XL HDH-3 peptide-QRE complex (7.5 nM for dad-XL HDH-3), nor indeed as tightly as un-alkylated HDH-

3 (200 nM). The lower binding affinity of PATF-1 peptide to QRE DNA compared to XL HDH-3 must reflect the effect of the additional linker and AD on the overall structure of the DBD helix. Despite the lower binding affinity of XL PATF-1 when compared to XL HDH-3, XL PATF-1 forms a more stable complex with DNA than the isolated wild type helix-3 of Q50K engrailed-QRE complex (1.7 μM, approximately 30 times weaker, albeit this binding was measured *via* EMSA) [103] but binds less strongly than full-length Q50K engrailed homeodomain (approximately 170 times tighter than dad XL PATF-1, measurements done *via* EMSA). [196, 79]

The complex between irrad-XL PATF-1 and QRE was more than 90 times weaker than dad-XL PATF-1 (K_D = 3875 nM) (**Table 3.2.7.2** and **Figure 3.2.7.1**). Comparing the CD spectra of irrad- XL PATF-1 with dad- XL PATF-1, where only slight changes in helicity between *cis* and *trans* conformations are evident, the large changes in binding affinity suggest that only small structural changes of the α-helix are required in order to reduce the ability of a peptide to bind tightly to target DNA. The anisotropy data obtained from the binding of irrad- XL PATF-1 to QRE DNA appeared not to reach saturation even at 6000 nM irrad- XL PATF-1. Individual titrations gave widely ranging anisotropy values for free DNA and at saturation, resulting in each run being separately fitted. For each fit K_D, free DNA and saturation were varied, allowing each run to produce a separate value for K_D. The individual values of K_D were averaged to give the overall values for the dissociation constant. After these fits, data from separate experiments were combined by normalising the experimentally measured anisotropy to the fit parameters for graphical display.

The DNA binding specificity of the HDH3-derived DBD was investigated by measuring the K_D of XL-PATF 1 in complexes with DNA sequences containing the natural target wild-type engrailed (ERE) (**Table 3.2.7.2**), [79] and a completely unrelated

DNA sequence for which engrailed homeodomains have little affinity (MCK-S) (**Table 3.2.7.1**). [46] No binding to MCK-S could be detected for concentrations up to 6 µM for both dad- and irrad- XL PATF-1 (**Table 3.2.7.2**), suggesting that the peptide discriminated fully between the typical DNA targets of homeodomain proteins and heterologous sequences.
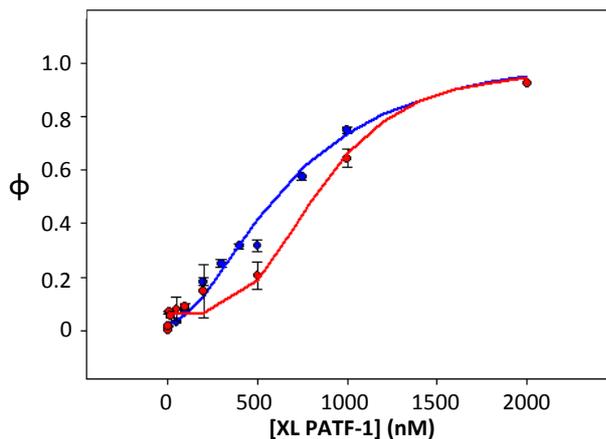
Figure 3.2.7.2. *DNA binding curves obtained from fluorescence anisotropy measurements of dad- (blue) XL PATF-1 and irrad- (red) XL PATF-1 to ERE.*

Peptide	K_D (nM)		
	QRE	ERE	MCK-S
dad-XL PATF-1	43.2 +/- 13.1	694.3 +/- 117.0	> 6000
irrad-XL PATF-1	3875.2 +/- 2871.7	779.9 +/- 256.1	> 6000
PATF-1	702.1 +/- 131.2	ND	ND
dad-XL HDH-3	7.5 +/- 1.3	140.0 +/- 11.0	> 6000
irrad-XL HDH-3	140.0 +/- 25.0	160.0 +/- 15.0	> 6000
HDH-3	200.0 +/- 11.0	175.0 +/- 16.0	ND

Table 3.2.7.2. *K_D values of dad-, irrad- and un-crosslinked PATF-1 and HDH-3 to DNA.* [175]

Binding of XL PATF-1 to ERE DNA showed no significant light-responsiveness, with dad- and irrad- dissociation constants of 694 nM and 779 nM repectively (**Table 3.2.7.2** and **Figure 3.2.7.2**). The binding of uncross-linked HDH-3 to ERE previously found to be similar to that with QRE, suggesting a non-specific binding mode for the unstructured peptide. [154] dad- XL PATF-1 displayed significant binding specificity and successfully discriminated between QRE and ERE in a similar fashion to XL HDH-3 (**Table 3.2.7.2**). [175]

Despite similar CD spectra to XL PATF-1, both dad- and irrad- XL PATF-2 were found not to bind as tightly to QRE DNA with K_D values of 765 nM and 16210 nM respectively. (**Table 3.2.7.3** and **Figure 3.2.7.3**).

	QRE K_D (nM)	
Peptide	dad	irrad
XL PATF-2	765.6 +/- 282.2	16210.4 +/- 326.0

Table 3.2.7.3. *K_D values of dad- and irrad- XL PATF-2 to QRE DNA. Data fitted with single binding site.*
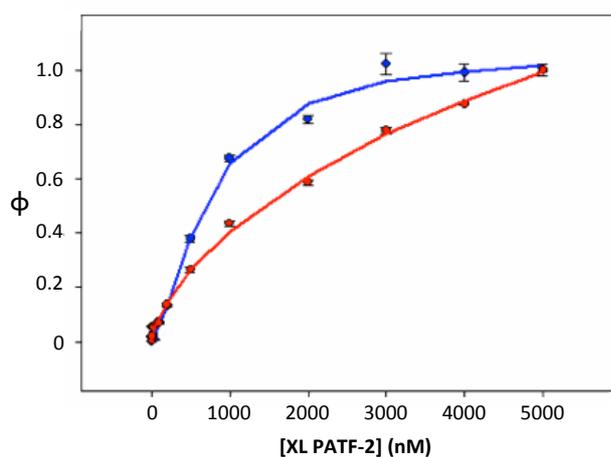
Figure 3.2.7.3. *DNA binding curves obtained from fluorescence anisotropy measurements of the XL PATF-2 to QRE DNA. dad-state is represented in blue and irrad-state in red.*

With identical DBD and AD regions, this must be due to the different linker region. Even though both linkers are approximately the same length, the flexible linker used in XL PATF-1 may allow the DBD to better fit the target DNA than the more rigid poly-proline linker. The large alterations in binding affinities compared to small changes in helicity seen in CD experiment reiterate how sensitive DNA binding is to small structural changes. In contrast, while previous studies carried out by Arora *et al*. and Mapp *et al.* using flexible polyethers, dimerisation domain from yeast GCN4 and rigid polyproline linkers on their transcription factor model demonstrated that the use of different linker affected their binding affinity to target DNA, [167, 184, 198] their transcription factor showed the tightest binding to target DNA with the rigid polyproline linker (7.7 nM). Transcription factors constructed using GCN4 linker and polyether linkers demonstrated slightly lower binding affinity to target DNA (11 nM and 32 nM respectively, measurements made by EMSA). Larger differences were noted in *in vitro* transcriptional activation, with 90% activation observed for the transcription factor with GCN4 linker, 70% for the rigid polyproline linker and 40% activation with flexible polyether linker. [167, 184, 198] Tighter binding to target DNA therefore does not necessarily result in better transcriptional activation, with activation favoured by domains that can cause dimerisation through coiled-coil structures.

3.2.8 *In Vitro* Transcription of PATF peptides

In vitro transcription assays with HeLa nuclear extracts were used to investigate the efficacy of PATF peptides as transcription factors. Nuclear extracts contain all the additional factors, including RNA polymerase II, needed to initiate the process of transcription and have been successfully used by Mapp *et al*. and Kwon *et al*. to

demonstrate *in vitro* transcription using a small molecule transcription factor.[167, 199] To create a transcriptional activation assay a technique described by Sawadogo *et al.* was adapted; the template plasmid pUCMLΔ53 was created based on the pMLΔ53 plasmid with a 75-bp × 5 repeat QRE binding sequences located 50-bp upstream of the TATA box (**Figure 3.2.8.1**).[200] An adenovirus major late promoter (AdML) sequence, an extensively studied promoter believed to be the strongest promoter available for most *in vitro* systems, lies 30-bp upstream of a 388-bp G-less cassette.[201] With this template, the specific transcripts initiated from the AdML promoter are templated upon a synthetic 388-bp DNA fragment that lacks cytidines, ensuring a large tract of the transcribed RNA strand will contain guanosine residues (**Figure 3.2.8.2**). *In vitro* transcription is then carried out in the presence of $[\alpha\text{-}^{32}\text{P}]$rUTP, a β-emitter with a half life of 14.3 days, and either in the absence of GTP or in the presence of RNase T1 and the chain terminator 3'-*O*-methyl-GTP. Under these conditions the only radioactively labelled RNA transcripts that can accumulate are the 388 nucleotide, RNase T1 resistant transcripts resulting from accurate initiation at the AdML promoter. These products can later be visualised *via* gel electrophoresis/autoradiography.[202] This assay system allows a rapid, direct and quantitative detection of promoter-dependent *in vitro* transcription by RNA polymerase II.

For each reaction, 100 ng of pUCMLΔ53 plasmid was pre-incubated with a dad XL-PATF (500 nM) peptide for 75 minutes at room temperature. irrad- XL PATFs were irradiated with 360 nm UV light for 6 minutes prior to addition to plasmid, which was then followed by continuous irradiation for the remaining 75 minutes. The plasmid-peptide mixture was then added to HeLa nuclear extract (8 µL, Promega) in a 25 µL reaction volume and processed as described in Chapter 2.[203] RNA transcripts were resolved on 7.5% polyacrylamide gels containing 7 M urea. Gels were soaked for at

least 20 minutes in deionised water in order to remove urea, then dried and exposed to photo-stimulatable phosphorimaging plates (Kodak). The storage plates were visualised by using Typhoon 9400 (GE Healthcare Life) followed by quantitation using ImageJ (**Figure 3.2.8.3**).
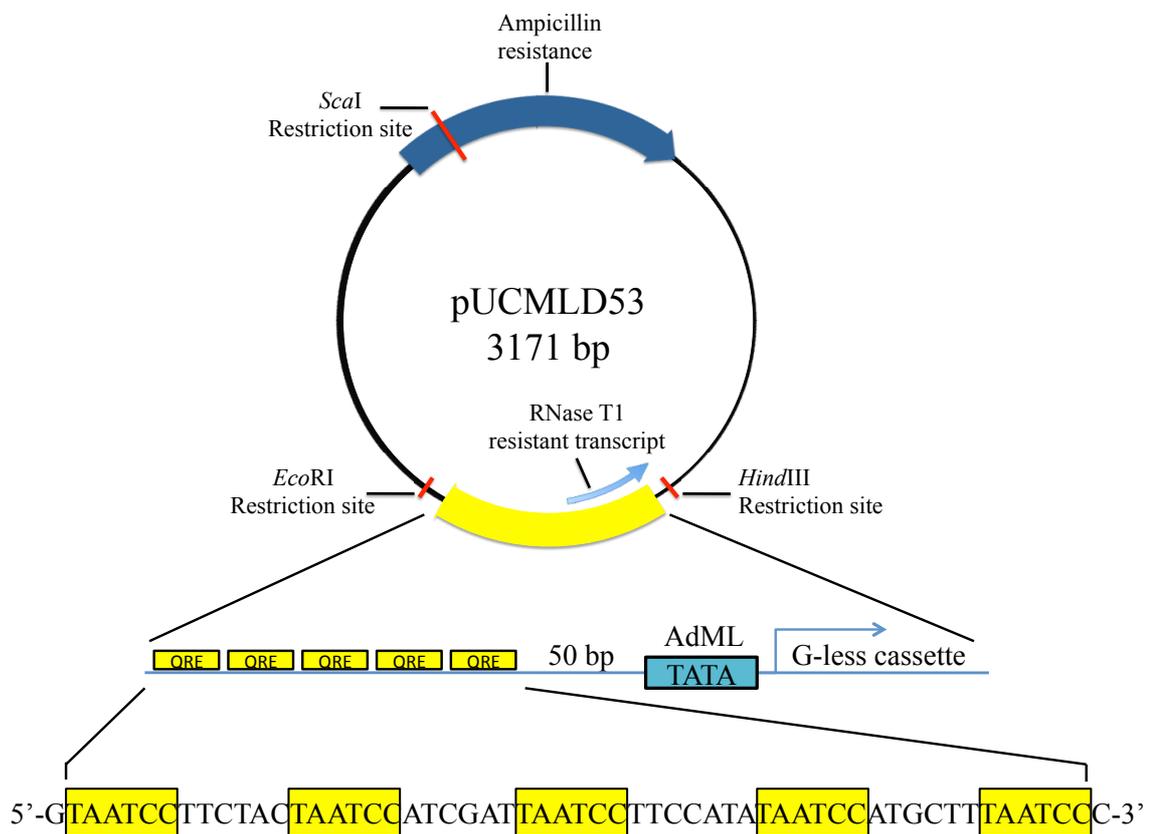
Figure 3.2.8.1. *Diagram of the pUCMLΔ53 plasmid DNA template used for in vitro transcription experiments. The promoter region is based on the Adenovirus major late promoter (AdML), and contains five cognate binding sites for PATF peptides upstream of G-less cassette reporter.*

Upstream

```
1      gaattcgcag tgTAATCCtt ctacTAATCC atcgatTAAT CCttccataT AATCCatgct

61     tTAATCCcct cgaccgcggt accctatccc gcgccccggg tgttcctgaa gggggcTAT

121    AAAAgggggt ggggcgcgt tcgtcctcAc tctcttccc
```

G-less cassette

```
1      ctccataccc ttcctccatc tataccaccc tactctcctt tcctcattat tcctcctatt

61     atcttctcct cttctctcct tcttctatat ttcccaaatc tatcatcatt cactctcatc

121    ccctcttcct tcactcccat tctattctac tcctttccct ttccatatcc cctccacccc

181    ccttcctccc ctctttcaat cttatcccca atcataaaat tatctcaatt atattctcct

241    tccataccccc ctatcatcct catccctatc accccctact cacccaatac tccctactca

301    tctcatatat ccttatcctc tcctcacctc tccctcctct atctcccccc ctcacactca

361    tttctcattc cactccc
```

Figure 3.2.8.2. *DNA sequences of cognate binding sites for PATF peptides, AdML promoter and G-less cassette. Upstream, sequences present upstream of the G-less cassette. PATF binding sequences are in capitals and highlighted in yellow, TATA box is in capital letters and highlighted in cyan, and the start site of transcription is highlighted in green. G-less cassette, nucleotide sequences of the long G-less cassette (388 bp).* [201]
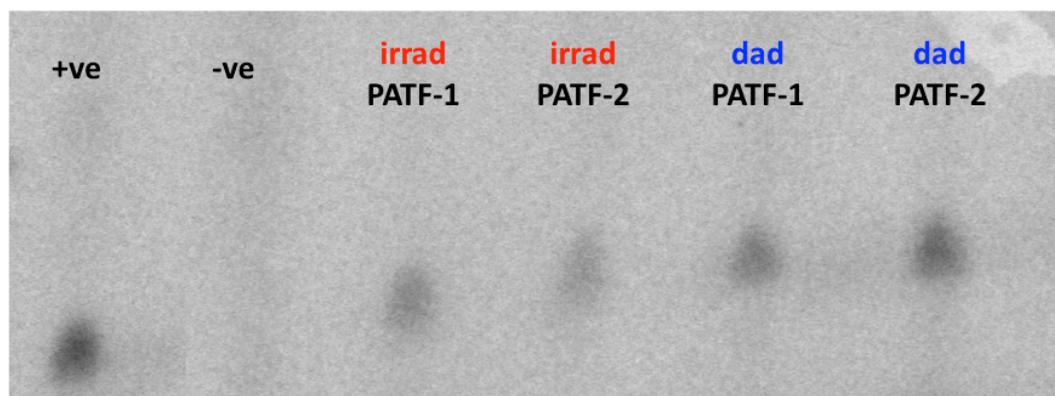
Figure 3.2.8.3. *Storage phosphor showing autoradiogram showing the results of in vitro transcription induced by XL PATF-1 in dark adapted and irradiated (dad and irrad) and XL PATF-2 (dad and irrad) at 500 nM concentration (+ve: positive control, -ve: negative control). Transcription reactions were performed for 60 minutes at 30 ^{o}C. The slant in the gel was due to uneven electric current running through the gel.*
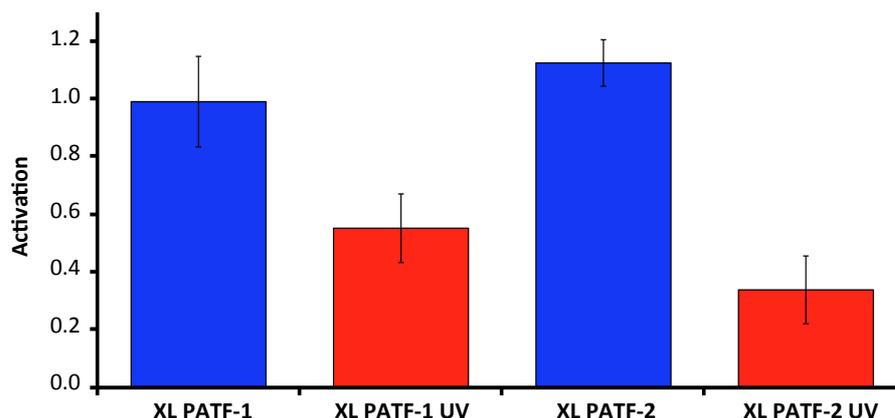
Figure 3.2.8.4. *Photoswitching of in vitro transcriptional activation by XL PATF-1 (dad and irrad) and XL PATF-2 (dad and irrad). The amount of transcript obtained from each reaction was compared with cytomegalovirus immediate-early promoter (CMV, Promega) as positive control to give a fractional activation value.*

Transcription reactions using XL PATF-1 and XL PATF-2 show that they produce similar levels of transcription in their dark adapted forms, however irradiation decreases transcription by XL PATF-2 by over three fold, while XL PATF-1 only shows a twofold reduction in activation. Work on a similar polyproline linker and activating domain artificial transcription activators carried out by Ansari *et al*. have previously demonstrated equivalent activation potency using this linker/activating domain design.[184]

However, a very dense band was noted at the top of the gel (**Figure 3.2.8.5** *A)*). This was initially believed to be due to the RNA transcript not being cleaved properly by T1 RNase treatment. After several attempts at optimising the reaction it was concluded that the T1 RNase was cleaving the RNA transcripts correctly, since the CMV positive control DNA creates two distinct bands (1.2 kb and 363 b). The prominent band in the positive control of 363 nucleotides, which is clearly visible on **Figure 3.2.8.5**, is similar in size as the G-less cassette nucleotides. One other possibility for the identity of the dark band derives from the circular nature of the pUCMLΔ53 plasmid. There could be many TATA binding sites scattered around the plasmid resulting in unspecified RNA

being transcribed. Furthermore, being a circular plasmid, transcription of these TATA sites might result in many very large RNA transcripts being produced. It was therefore decided to linearize the plasmid and monitor its effect on the appearance of the dense band. Three different designs of the linear pUCMLΔ53 were tested: pUCMLΔ53 doubly digested with *Sca*I and *Hin*dIII (approximately 1.5 kbp in size) and DNA singly digested with both *Hin*dIII and *Sca*I (**Figure 3.2.8.5 *B*)**). The plasmid was designed so that a *Hin*dIII restriction site is directly few base pairs downstream of the G-less cassette, so that transcription of the G-less cassette should stop immediately after the end of the cassette if the DNA is cleaved at this site. The other restriction site chosen was *Sca*I as it lies upstream of the AdML promoter in the ampicillin resistance gene.

Transcription tests were then performed using dad- XL PATF-1 on new DNA transcription template, and as expected the dark band observed with the circular plasmid was greatly reduced (**Figure 3.2.8.5 *B*)**). Singly digested plasmid with *Sca*I and *Hind*III show some reduction in the large dense band, but are unable to eliminate it. The doubly digested plasmid shows no dense band with only the G-less cassette transcript evident. Circular pUCMLΔ53 with no PATF-1 peptide present used as a negative control confirms that transcription of the G-less cassette relies on the PATF peptide in order to 'activate' transcription. In the absence of peptide transcription is minimal or even absent (* **Figure 3.2.8.5 *B*)**). To confirm that the remaining band represented the correct size for the G-less cassette, a denatured single strand 500 bp DNA ladder (**Figure 3.2.8.5 *C*)**) was run along the transcription experiment and stained in ethidium bromide for UV visualisation. It was hypothesised that denatured single strand DNA should have similar gel migrating properties as RNA and that a 500 bp DNA ladder denatured into single strands would best estimate the speed at which the RNA transcript migrates. Once gel electrophoresis was performed, the R_f values of each DNA sizes and G-less

transcript were calculated and a graph plotted using R_f values obtained from the DNA ladder to estimate the mass of the G-less cassette transcript. The R_f value of the G-less cassette was calculated to be 0.47, which corresponded to approximately 291 nucleotides on the DNA ladder graph. However, since SDS gels are known to have 10 % error, this would put our G-less cassette transcript closer to the 388 nucleotides initially predicted, suggesting that this band does indeed represent the RNA transcript.
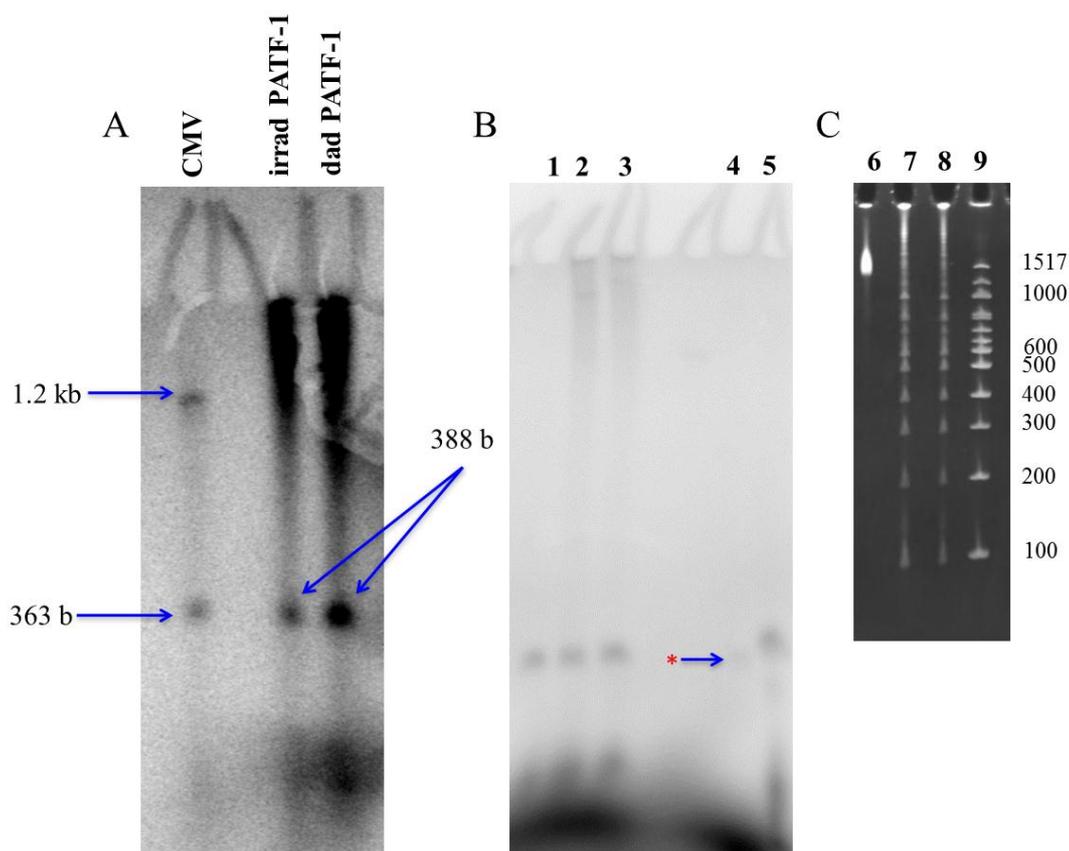
Figure 3.2.8.5. *A) Transcription of dad- and irrad- XL PATF-1 with CMV positive control. Two bands can be clearly seen for CMV representing 1.2 kb and 363 b nucleotide transcripts. Dark patches can also be seen for dad- and irrad- XL PATF-1 as well as the transcript from the G-less cassette (388 b), which runs at approximately the same speed as the CMV positive control. B) Transcription of dad-XL PATF-1 using 1) doubly ScaI and HindIII digested pUCMLΔ53, 2) singly Hind III and 3) singly ScaI digested pUCMLΔ53. 4) circular pUCMLΔ53 without PATF-1 peptide as negative control. Note: a small amount of background transcription (*) can still be seen despite the absence of PATF-1 peptide and 5) CMV as a positive control. C) Urea gel of 6) ScaI and HindIII digested pUCMLΔ53 plasmid, 7) & 8) denatured single strand 500 bp DNA ladder (NEB) and 9) 500 bp DNA ladder (NEB). Gel was stained in ethidium bromide after electrophoresis and visualised under UV light.*

3.2.9 Designing and cloning of *i, i + 7* and *i, i + 4* PATF peptides

Since the *i, i*+11 spacing has successfully controlled the conformation and the DNA binding properties of PATF-1 and PATF-2 as well as their *in vitro* transcription activities, *i, i*+7 and *i, i*+4 cysteine spacings were also investigated to offer the possibility of positive photoregulation. Previously, the azobenzene cross-linker in the *trans* conformation has shown to be too long to allow the *i, i*+7 residues and intervening residues to adopt an α-helix conformation, whilst in the *cis* conformation its length better matches the spacing of *i, i*+7 or *i, i*+4 cysteines spacing. [142, 114] Thus PATF-3 (*i, i*+7) and PATF-4 (*i, i*+4) are expected to increase transcription upon isomerisation with 363 nm UV light. The AH activation domain used on PATF-1 and PATF-2 was retained for PATF-3 and PATF-4 along with the better performing rigid polyproline linker of PATF-2 (**Figure 3.2.9.1**) (**Table 3.2.9.1**).
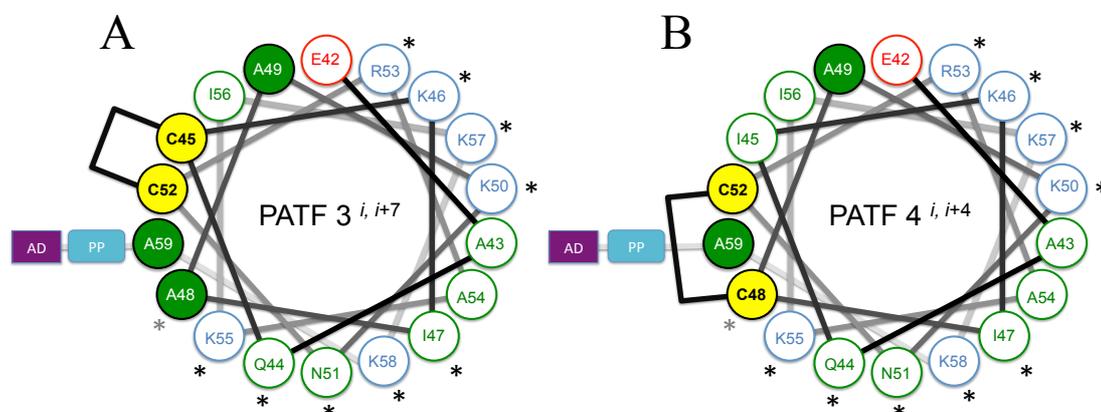
Figure 3.2.9.1. *Schematic of PATFs. A) Diagram of the helical wheel of cross-linked PATF-3 with i, i+7 spacing and PP linker. B) Diagram of the helical wheel of cross-linked PATF-4 with i, i+4 spacing and PP linker. Residues that are in contact with DNA in the parent homeodomain structure are marked with asterisk, blue and red circles represent charged residues based on parent HDH-3 DNA binding domain. Residues that have been substituted in order to avoid steric clash with the cross-liner are marked in filled green circles.* [154, 175]

Figure 3.2.10. *A) Map of the pET31b(+) expression vector and cloning charts of PATF-3 and PATF-4 peptides. B) Agarose gels stained with ethidium bromide. Lane 1: marker (fragment size is indicated in bp), lane 2: PATF-1 PCR product (+ve control), lane 3: empty pET31b(+) with no insert (-ve control), lane 4: PATF-3 PCR product, lane 5: PATF-4 PCR product.*

Sequencing results of chosen PATF-3 and PATF-4 clones demonstrated several deviations from the original sequences intended (**Table 3.2.10**). In this case PATF-3 has

four proline residues within the rigid oligo-proline linker missing, and PATF-4 has mutations within the rigid oligo-proline linker (**Table 3.2.10**) different from the original intended sequences (Section 3.2.9). As with the similar error observed for PATF-2 clone, it is believed that the error could arise when there are high numbers of the same amino acids in a row previously discussed in section 3.2.2. Despite this deviation, both PATF-3 and PATF-4 still have the correct DBD and AD sequences for DNA recognition and transcription activation. It was predicted that PATF-3, having shorter rigid oligo-proline linker than PATF-2, may not be as effective at promoting transcription as PATF-2 according to previous studies. [184] As for PATF-4, having four mutations within the rigid oligo-proline linker could affect the binding affinity to target DNA and promoting the transcription process differently to PATF-2 and PATF-3 with rigid linkers. Despite the mutations observed, clones of PATF-3 and PATF-4 obtained were used for protein expression.

Name	Sequences
PATF-3 *i, i*+7	gaagcgcagtgtaaaatcgcggctaagaactgccgtgcgaaaattaagaaagctccgccgccaccgcctccg E A Q C K I A A K N C R A K I K K A P P P P P ccaccgccggaattcccgggtattgagctgcaagaactgcaagagctgcaggccctcctgcagcaa P P P E F P G I E L Q E L Q E L Q A L L Q Q
PATF-4 *i, i*+4	gaagcgcagatcaaaatctgtgctaagaactgccgtgcgaaaattaagaaagctccgccgccacgccccgcc E A Q I K I C A K N C R A K I K K A P P P R P A accgctccgccaccgccggaattcccgggtattgagctgcaagaactgcaagagctgcaggccctcctgcag T A P P P P E F P G I E L Q E L Q E L Q A L L Q caa Q

Table 3.2.10. *Sequencing results of PATF-3 and PATF-3 clones. The obtained sequences are in minuscule letters (top) and the translated sequences are in capital letters (bottom). Residue mutations of PATF-4 are highlighted in blue.*

3.2.11 Expression and purification of PATF-3 and PATF-4

PATF-3 and PATF-4 peptides were expressed in BL21 Star cells using the same conditions as for PATF-1 and PATF-2. Expression of the peptides was assessed by SDS-PAGE (**Figure 3.2.11.1**).
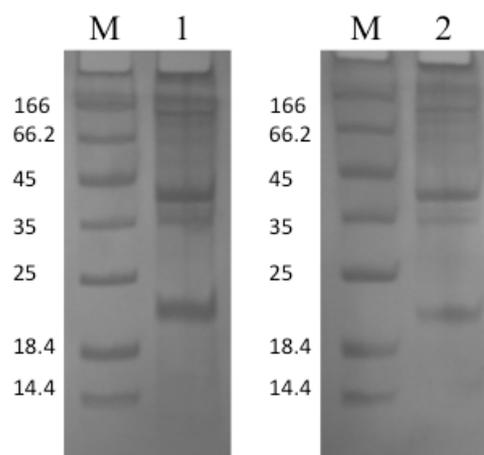
Figure 3.2.11.1. *SDS-PAGE analysis of PATF-3 and PATF-4 expression in Star cells. M: Protein marker (fragment size is indicated in kDa, Fermentas), lane 1: 3 hours induction of PATF-3, lane 2: 3 hours induction of PATF-4.*

KSI-PATF-3-His$_6$ and KSI-PATF-4-His$_6$ fusion protein inclusion bodies were solubilized in lysis buffer, re-precipitated and the KSI and His-tag were removed by cyanogen bromide cleavage. Purification was carried out by reverse phase HPLC (Phenomenex Gemini C$_{18}$ column). PATF-3 and 4 both eluted at approximately 21 minutes using a 10-100% acetonitrile:water (containing 0.1 % TFA) linear gradient over the course of 60 minutes at a flow rate of 5 ml min^{-1} (**Figure 3.2.11.2**). PATF-3 and PATF-4 have similar retention times to PATF-2 due to their shared linker. The masses of PATF-3 and PATF-4 were confirmed by MALDI-TOF MS (**Table 3.2.11.1**). Deviation in masses by approximately 281.72 and 47.96 from the calculated monoisotopic mass of PATF-3 and PATF-4 respectively were also observed. Similar causes for the deviation were previously discussed in Section 3.2.3. Despite this

deviation, we can also assume that the correct peptides were obtained due to the confirmation from the sequencing results.
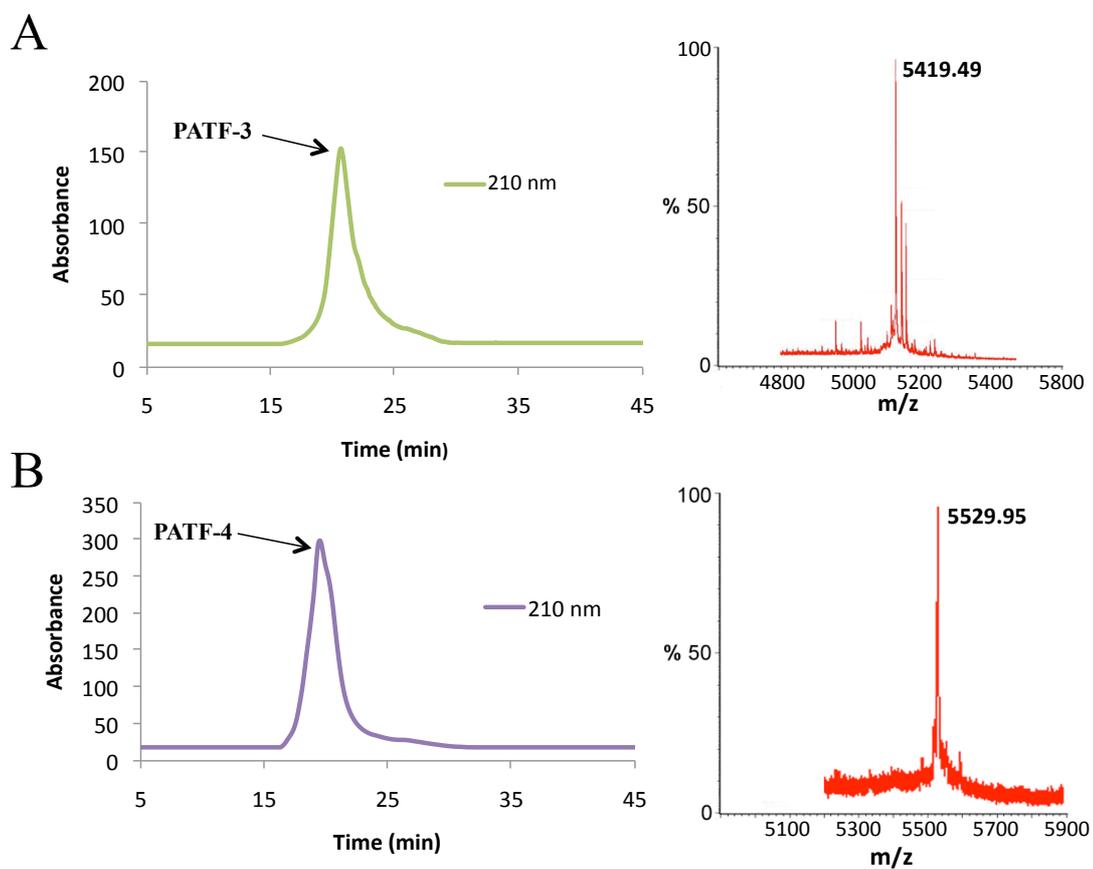
Figure 3.2.11.2. *Chromatogram from the analytical HPLC analysis and MALDI-TOF mass spectrum of A) PATF-3 and B) PATF-4.*

Peptide	Calculated monoisotopic mass	Observed mass
PATF-3	5137.77	5419.49
PATF-4	5481.99	5529.95

Table 3.2.11.1. *Calculated monoisotopic masses and observed masses of PATF-3 and PATF-4 peptides.*

3.2.12 Cross-linking of PATF-3 and PATF-4 with 3,3'–bis-(sulfo)-4,4'-bis (chloroacetamino)azobenzene

PATF-3 and PATF-4 were alkylated with **BSBCA** and purified as previously. The HPL chromatogram shows absorption at 360 nm at around 19 minutes for XL-PATF-3 XL PATF-4 (**Figure 3.2.12.1**). The masses were measured by MALDI-TOF MS (**Table 4.2.12.1**). Deviation in masses by approximately 281.71 and 47.96 from the calculated monoisotopic mass of XL-PATF-3 and XL-PATF-4 respectively were also observed. Similar causes for the deviation were previously discussed in Section 3.2.3. Despite this deviation, we can also assume that the correct peptides were obtained due to the confirmation from the sequencing results.
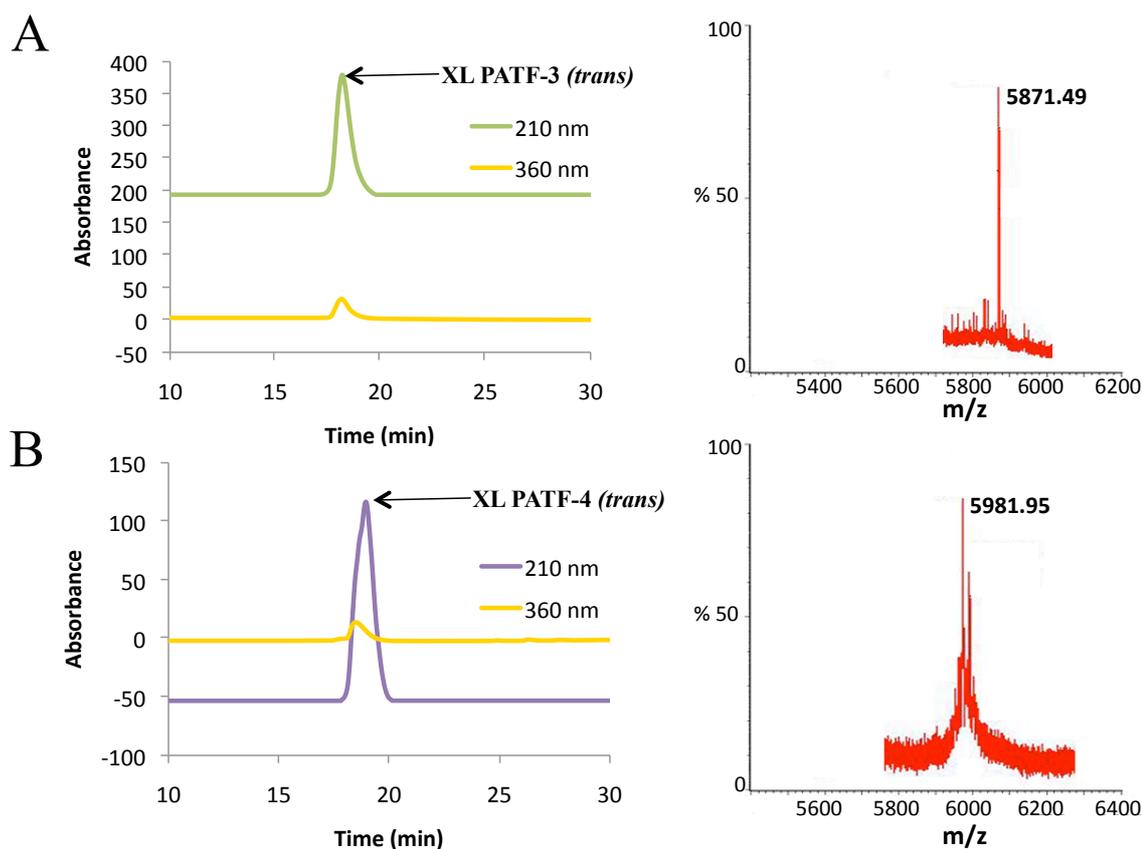
Figure 3.2.12.1. *Chromatogram from the HPLC analysis and MALDI-TOF mass spectrum of A) XL PATF-3 and B) XL PATF-4.*

Peptide	Calculated monoisotopic mass	Observed mass
XL PATF-3	5589.78	5871.49
XL PATF-4	5933.99	5981.95

Table 3.2.12.1. *Calculated monoisotopic masses and observed masses of XL PATF-3 and XL PATF-4 peptides.*

3.2.13 Spectroscopic characterisation of XL-PATF-3 and XL-PATF-4

3.2.13.1 UV-Visible absorption

The change in crosslinker conformation of XL-PATF-3 and XL-PATF-4 upon irradiation was observed by UV spectroscopy and at different temperatures as previously. Both peptides were irradiated for 6 minutes under UV light (363 nm) and the absorption spectra of dark-adapted XL PATF-3 and XL PATF-4 were recorded at different timepoints (**Figure 3.2.13.1.1**).
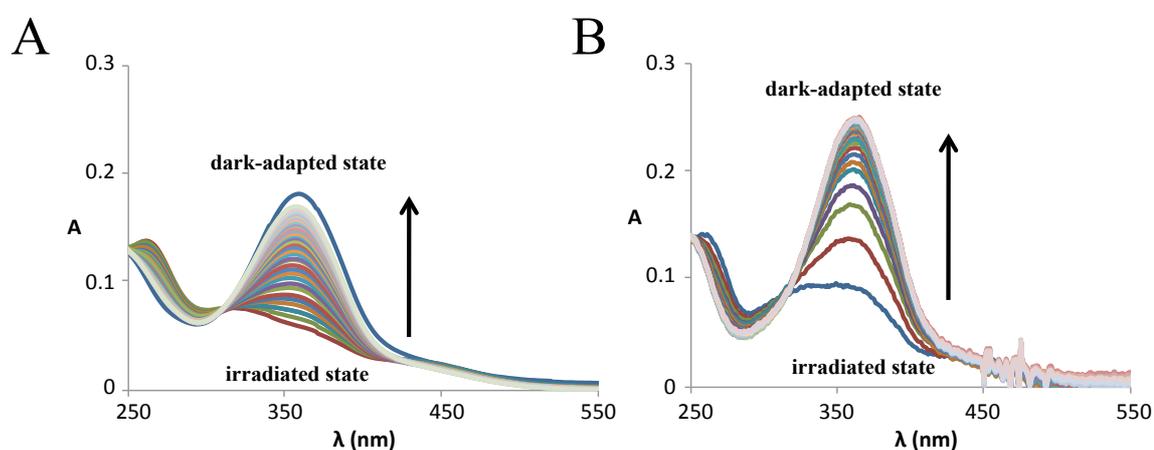
Figure 3.2.13.1.1. *UV/Vis spectra of A) XL PATF-3 and B) XL PATF-4 at 25 ^{o}C. Spectra were acquired on the dark-adapted state (trans), the irradiated state (cis) and 2 minute intervals after irradiation.*

Irradiated XL PATF-3 and XL PATF-4 peptides reverted to the dark-adapted state with similar half-lives of 45 and 40 mins respectively at 15 °C despite the fact that XL PATF-4 has mutations withing the polyproline linker (**Figure 3.2.13.1.2**). XL PATF-3

and XL PATF-4 also possess significantly shorter half-lives than the XL HDH-3 peptide, consistent with this reduction being a by-product of the attached linker and activating domain. Additional support for this hypothesis is lent by the similarity of the relaxation rates of XL PATF-3 to XL PATF-2, which shares similar rigid linkers. Interestingly, despite the mutations withing the rigid polyproline linker, XL PATF-4 demonstrated similar relaxation rates to XL PATF-3, suggesting that the position of the cysteine residues maybe have greater effect on the relaxation of the peptide than the linker used to link DBD and AD domains.
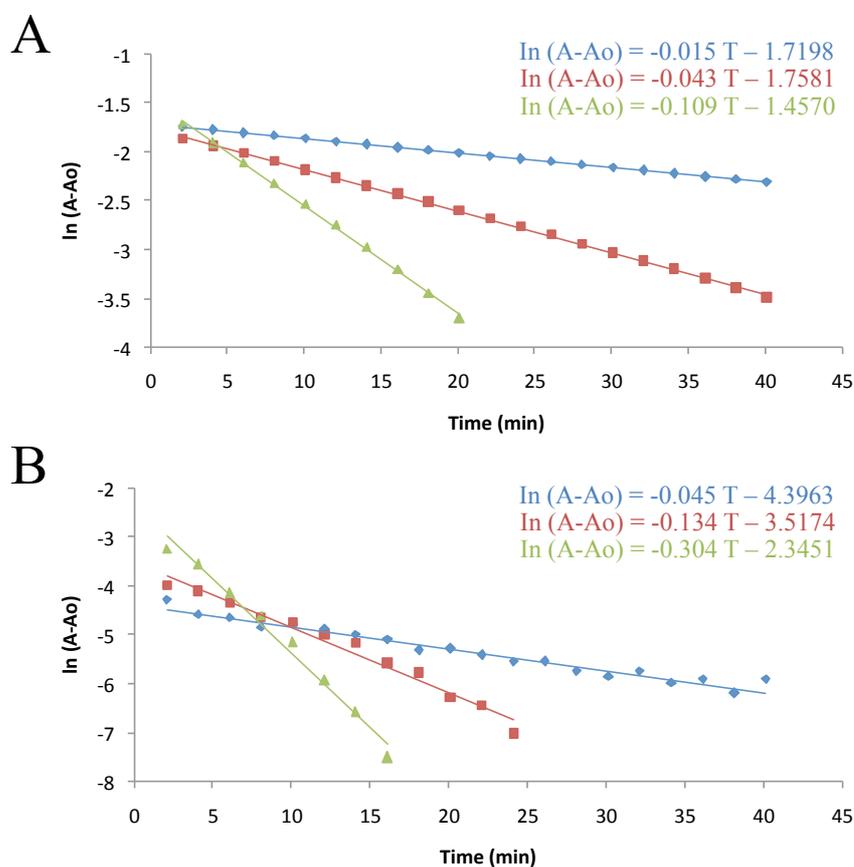
Figure 3.2.13.1.2. *Relaxation of different PATF peptides as a function of time in the dark as monitored by UV (363 nm). A) XL PATF-3 and B) XL PATF-4; Blue: 15 °C, red: 25 °C and green: 37 °C.*

Three different temperatures were selected and the temperature dependence of the rates compared in Arrhenius plots, which show XL PATF-3 and XL PATF-4 peptides to have similar E_a values, but both are higher than found for the *i*, *i* + 11 peptides (**Figure 3.2.13.1.3**) (**Table 3.2.13.1.1**).
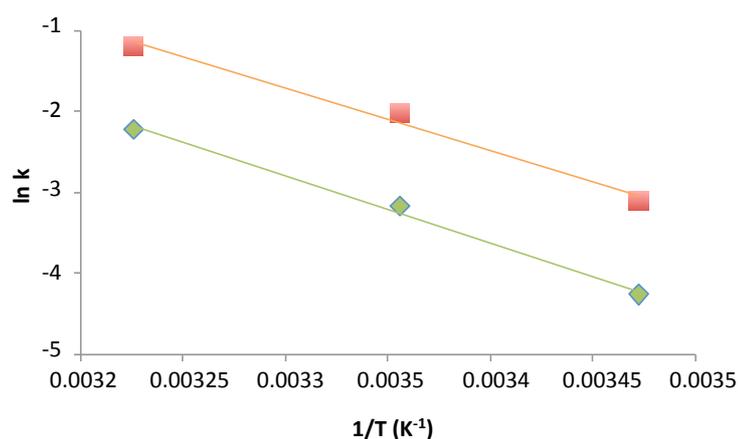
Figure 3.2.13.1.3. *Arrhenius plot of the UV relaxation rates of XL PATF-3 (green) and XL PATF-4 (orange).*

Peptide	E_a (kJ mol^{-1})	A (s^{-1})
XL PATF-3	69.8 +/- 1.8	4.97×10^{10} +/- 1.26×10^9
XL PATF-4	64.2 +/- 1.5	2.12×10^{10} +/- 1.07×10^9

Table 3.2.13.1.1. *Arrhenius parameters for the relaxation of XL PATF-3 and XL PATF-4 peptides.*

Similar observations have been made with crosslinked Bak peptides where Bak$_{72-87}^{i,i+7}$ has significantly higher activation energy (E_a) than Bak$_{72-87}^{i,i+11}$. [147] The rates of thermal relaxation of the irradiated XL PATF-3 and XL PATF-4 were calculated from triplicate UV spectroscopy experiments. Both peptides have similar and shorter relaxation rates compared to XL PATF-2 (**Table 3.2.6.1.3**).

Peptide	t½ at 15 °C (min)	t½ at 25 °C (min)	t½ at 37 °C (min)
XL PATF-3	45 +/- 2	16 +/- 0.2	5 +/- 0.8
XL PATF-4	40 +/- 4	16 +/- 2	2 +/- 1.5

Table 3.2.6.1.2. *Half-life of the cis-isomer for XL PATF-3 and XL PATF-4 peptides.*

3.2.13.2 Circular Dichroism (CD) spectroscopy of XL PATF-3 and XL PATF-4

CD spectra were measured to analyse the helicity of XL PATF-3 and XL PATF-4 peptides and investigating changes in structure induced by photoswitching. The CD spectra of XL PATF-3 and XL PATF-4 showed almost no difference between the dark adapted and irradiated states eventhough XL PATF-4 having mutations within the polyproline linker (**Figure 3.2.13.2.1**). Both XL PATF-3 and XL PATF-4 display high intensity minima peaks at approximately 205 nm, but only XL PATF-4 posessed a minima at 222 nm. The lack of clear minima at 208 and 222 nm (corresponding to peaks observed in α-helical structures), suggest that XL PATF-3 is mostly comprised of random coil structures. Circular Dichroism software (Dichroweb) was used to analyse the structure, which also confirms a predominantly random coil structure for XL PATF-3 (Fitted to SELCON3, model: Set 6 [optimised for 185-240 nm], 0% helix, 1.9% sheet, 21% coil) [204] The observed mean residue ellipticities of XL PATF-4 at 222 nm of -2160.3 deg cm^2 dmol^{-1} (dark adapted state) and -2570.6 deg cm^2 dmol^{-1} (irradiated state) which correspond to 5.9% and 7% helicities. As with XL PATF-3, CD spectra also suggest that the structure is mostly random coil. Circular Dichroism software suggested that XL PATF-4 the structure is also mostly random coil (Fitted to SELCON3, model: Set 6 [optimised for 185-240 nm], 0% helix, 1.9% sheet, 22% coil). Different cysteine spacings employed to disfavour a helical conformation in the *trans* azobenzene

conformation (*i, i+7* spacing on XL PATF-3 and *i, i+4* spacing on XL PATF-4) seem to have successfully distorted the structure, but have limited use in reinstating helicity in the peptides upon irradiating with UV light.
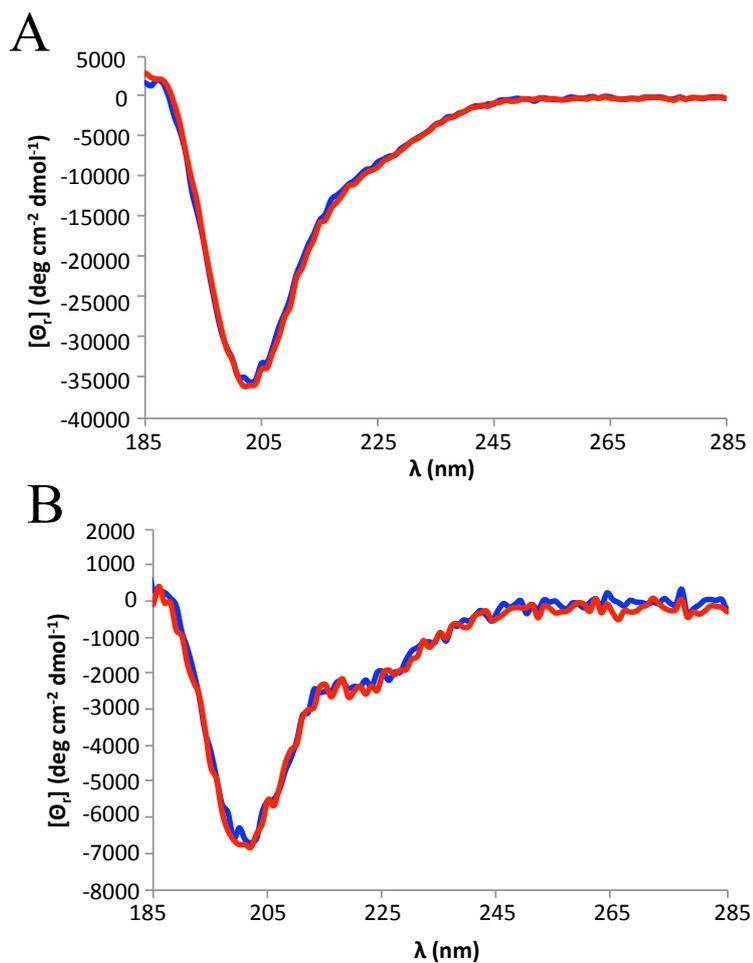
Figure 3.2.13.2.1. *CD spectra of dad-(blue) and irrad-(red) A) XL PATF-3 and B) XL PATF-4 peptides carried out in 5 mM potassium phosphate (pH 8.0) at 25 °C. The introduction of the BSBCA crosslinker seems to have little to no effect to the structure of both peptides upon isomerisation.*

3.2.14 DNA Binding Specificity measured by Fluorescence Anisotropy

Fluorescence anisotropy measurements were carried out to quantify the DNA binding affinities of dark adapted and irradiated XL PATF-3 and XL PATF-4 peptides for sequences containing QRE binding sites. XL PATF-3-QRE and XL PATF-4-QRE

complexes were both successfully fitted to the Langmuir isotherm equation assuming one binding site as previously done for dad- XL PATF-1-QRE and dad- XL PATF-2-QRE complexes. XL PATF-3 and XL PATF-4 demonstrated similar complex stabilities in their dad- and irrad-forms (**Figure 3.2.14.1**).
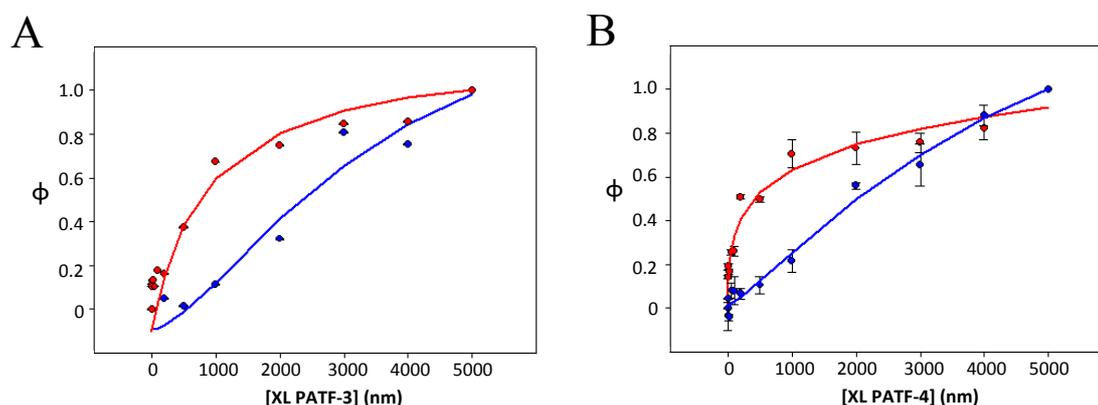
Figure 3.2.14.1. *DNA binding curves obtained from fluorescence anisotropy normalised to fit parameters measurements of the A) XL PATF-3, and B) XL PATF-4 to QRE DNA. dad-state is represented in blue and irrad-state in red.*

	QRE KD (µM)	
	dad	irrad
XL PATF-3	2.4 +/- 0.3	1.7 +/- 1.1
XL PATF-4	4.8 +/- 2.9	1.3 +/- 0.7

Table 3.2.14.1. *K_D values of dad- and irrad- XL PATF-3 and XL PATF-4 peptides to QRE DNA. Data fitted with single binding site model.*

The presence of a *trans* azobenzene cross-linker prevented XL PATF-3 (K_D = 2.4 µM) and XL PATF-4 (K_D = 4.8 µM) from forming high affinity complexes (**Table 3.2.14.1**). After irradiation, the stability of complexes formed between QRE DNA and the irrad-forms of XL PATF-3 and XL PATF-4 were increased by almost 2 fold (K_D = 1.7 µM) and 4 fold (K_D = 1.3 µM) respectively. This was anticipated since *i, i*+4 spacing on

PATF-4 is shorter than *i, i*+7 of PATF-3, allowing more effective distortion the DBD helix, as previously observed with alkylated BH3 peptides. [189] Slightly lower K_D value observed for XL PATF-4 could be due to the presence of mutations within the polyproline linker, possibly making the linker a little more flexible than the polyproline linker in XL PATF-3, thus allowing the DBD of XL PATF-4 to better fit the target DNA as previously observed with XL PATF-1. The relatively small K_D differences between *trans*- and *cis*- configurations were commensurate with the small changes in the CD spectra of the peptides that indicated few signs of structural changes upon irradiation and with the low degree of helicity. Previous work on crosslinked BH3 peptides with *i, i*+7 and *i, i*+4 spacings yielded approximately 20 fold changes between dad- and irrad- forms in *in vitro* binding studies [205] and despite only small changes being evident in CD spectra these peptides proved capable of inducing light-dependant biochemical responses in permeabilised cells. [206]

3.2.15 *In vitro* transcription of XL PATF-3 and XL PATF-4 peptides

In vitro transcription assays by HeLa nuclear extracts were used to investigate the efficiency of XL PATF-3 and XL PATF-4 peptides as transcription factors using doubly digested *Sca*I and *Hin*dIII pUCMLΔ53 plasmid as a template. It was hypothesised that *i, i*+7 and *i, i*+4 cysteine spacings would disfavour DBD binding to target DNA and would reduce the transcription activation of the peptide *in vitro* in the dark, whereas the irradiated state should show higher levels of transcription. Having mutations within the polyproline line linker could also affect the transcription activity of XL PATF-4 since the linker may not be fully rigid as with XL PATF-3. Experimental results showed dad- XL PATF-3 and dad- XL PATF-4 displayed similar transcription levels with small

increases in activation for both irrad- XL PATF-3 and irrad- XL PATF-4 peptides (**Figure 3.2.15.1**).

In this case the pMLΔ53 plasmid, generously donated by Prof. Robert Roeder, was used as the positive control instead of CMV. pMLΔ53 also contains a G-less cassette and can be transcribed by RNA polymerase II without the use of other artificial transcription factors. [200] The efficiency of pMLΔ53 is very similar to CMV (14557 +/- 1610 for CMV, 14239 +/-5479 for pMLΔ53, obtained *via* densitometry using ImageJ software) (**Figure 3.2.15.2**). Irradiated XL PATF-4 showed the higher activation level of the two peptides, possibly due to the linker used in XL PATF-4 being longer than the rigid linker in XL PATF-3 since it was suggested that longer linkers (12-15 residues) were more effective at promoting transcription. [184] Also not being fully rigid could also mean that this had allowed better fit of XL PATF-4's DBD to target DNA. However, neither seems to be able to match the dynamic range of XL PATF-1 and XL PATF-2, with changes in activity below 50% (**Figure 3.2.15.3** *A)* and *B))*.

The reduced transcriptional activity of dad- XL PATF-3 and dad- XL PATF-4 is concordant with their CD spectra and fluorescence anisotropy results. Despite little or no changes in structural conformation for both peptides observed by CD these techniques the K_D values the irrad- states of both peptides suggest a measure of optical control over the binding to target DNA. Furthermore, the CD spectrum of XL PATF-4 suggested more helical structure than XL PATF-3 (a more obvious minima at 222 nm, possibly due to the presence of alanine and arginine residues within the linker), corresponding to higher binding affinity to QRE in the irrad-state, might explain the larger increase in activation upon irradiation than XL PATF-3. This could also suggest that rigidity may not be the only factor contributing to increasing the transcription

activity but also the length of the linker. Since the half-life of XL PATF-3 and XL PATF-4 are relatively short, photoswitching of transcriptional activation could be augmented by using lower temperatures during *in vitro* transcription, although this would be offset by the decreased rate of RNA polymerase II reaction.
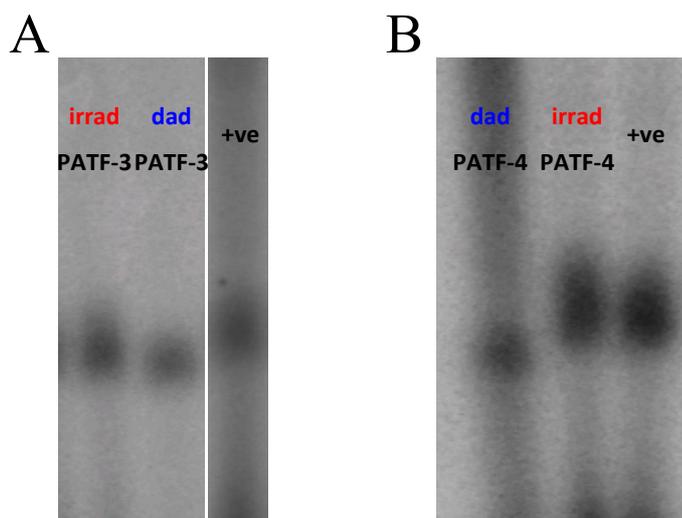
Figure 3.2.15.1. *Storage phosphor autoradiogram showing an in vitro transcription of A) XL PATF-3 (dad and irrad) and B) XL PATF-4 (dad and dad) at 500 nM concentration. Transcription reactions were performed for 60 minutes at 30 ^{o}C.*
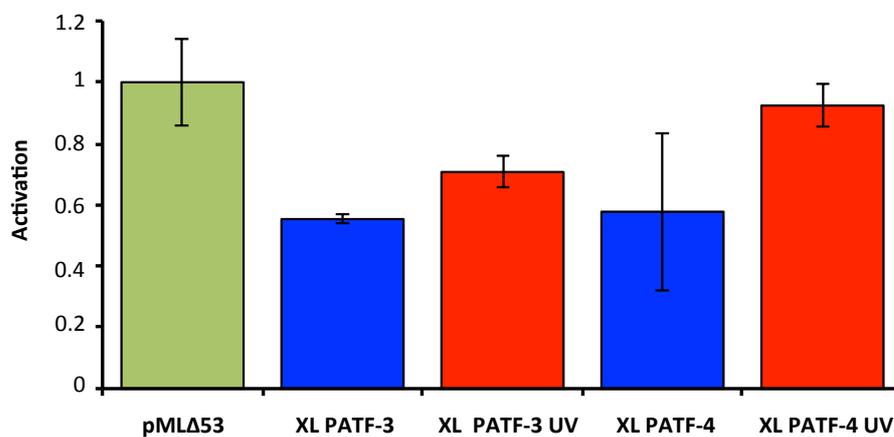
Figure 3.2.15.2. *In vitro transcriptional activation of XL PATF-3 (dad and irrad) and XL PATF-4 (dad and irrad) normalised to the transcription product of pMLΔ53.*
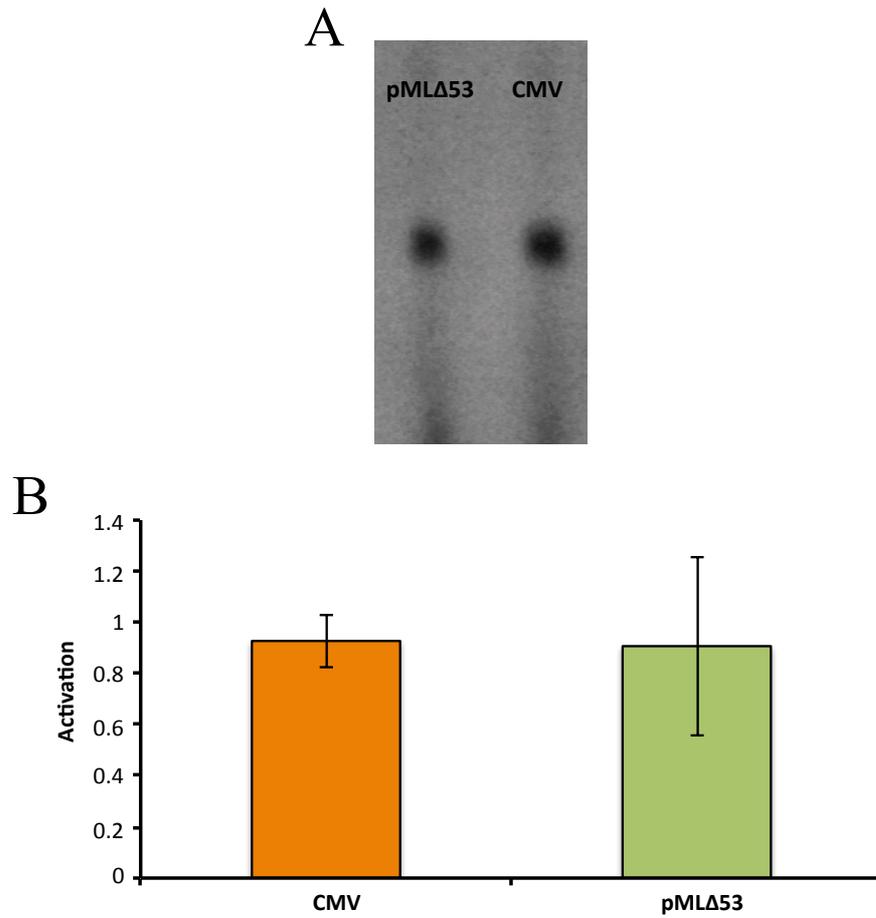
Figure 3.2.15.3. *A) Storage phosphor autoradiogram showing in vitro transcription pMLΔ53 and CMV. B) In vitro transcriptional activation of CMV and pMLΔ53.*

3.3 Summary

In summary, four light-responsive artificial transcriptional activators based on an 18 amino acid DNA binding domain of an engrailed homeodomain with azobenzene-derived photo-activatable cross-linkers were fused with an amphipathic activating domain (AH) connected by different linkers to regulate the conformation of peptides. The azobenzene proved able to not only control the binding specificity of these peptides to target DNA, but also demonstrated their potency at activating and deactivating the transcription process in an *in vitro* environment.

XL PATF-1 and XL PATF-2 were more potent transcription activators/deactivators when compared to XL PATF-3. XL PATF-4, with presence of mutations and therefore possibly affecting the rigidity of the linker, has demonstrated similar character to XL PATF-3 in terms of photoswitching, K_D and transcription activation. It would be interesting to investigate how full-length polyproline XL PATF-4 perform.

Despite smaller changes in K_D values between dad- and irrad states, XL PATF-2 seem to be the best design for a miniature artificial transcriptional factor due to its higher dynamic range in *in vitro* transcription assays. Previous work carried out by Woolley *et al*. on an artificial transcription activator has shown that the *in vivo* activity of their peptide was decreased by approximately 40% upon irradiation. [157] However, our XL PATF-2 has shown over 60% decrease in *in vitro* activation upon irradiation, which could be very promising for *in vivo* studies and a very good model for the design of future photo-switchable artificial activators.

Further *in vivo* studies of these peptides would enable us to investigate how potent these peptides are in live cell environment and potentially pave a way for the designs of more potent miniature artificial transcription activators.

Chapter 4
Conclusion

4. Conclusion

Incorporation of a photoswitchable azobenzene crosslinker to an PATF artificial transcription factor demonstrated that it is possible to reduce the size of the protein DBD motif by providing alternative means of stabilising the secondary structure. It also was possible to control the binding affinity of this miniature artificial transcription factor to its cognate DNA and to control the transcription process *in vitro*. Studies have also demonstrated that types of linkers connecting the DBD and AD domains can be crucial at promoting the transcription process and the binding affinities do not directly drive dynamic range in *in vitro* transcription assays.

However, there are also drawbacks to this system; UV light is required to isomerise the photoswitch, which could limit its use in *in vivo* environments as UV light can trigger unwanted responses such as apoptosis. One way around this would be the use of non-UV switchable azobenzene crosslinker developed by Beharry *et al*. However these non-UV switchable azobenzene crosslinkers have short half-life due to sensitivity to reducing conditions, which can be a drawback for *in vivo* environment. [125] More recent developments of a new family of azobenzene crosslinker could offer a way to effectively using these switches *in vivo*. For example, Bleger *et al.* have synthesised a new class of azobenzene with fluorine atoms *ortho* to the azo moiety, allowing the ability to reversibly isomerise the two isomers in the visible region of the spectrum. [207] Isomerisation from *trans*-to-*cis* of the tetrafluoro-substituted azobenzene was done by exposure to green light ($\lambda > 450$ nm), and the reverse from *cis*-to-*trans* was done by exposure to blue light (410 nm). The isomerisation with green light also produced 91% *cis* conformation, which is greater than previously observed for any azobenzene designs (70-80% *cis*). Moreover, this photoswitch has an exceptionally long half-life (700 days

at 25 ^{o}C) for an azobenzene derivative, which could make it an attractive system to use to photocontrol *in vivo*. Samanta *et al*. have also developed an azobenzene in which all four *ortho* positions to the azo group have been chloro-substituted. [208] This azobenzene moiety demonstrated effective *trans*-to-*cis* switching with red light (630-660 nm), a wavelength range that is orders of magnitude more penetrating through tissue than other parts of the visible spectrum as well as stability to reduction by glutathione under reduced condition. The azobenzene switch was also tested *in vivo* by microinjecting into zebrafish embryos, where it was possible to photoswitch the azobenzene with red-light without causing any morphological changes or apparent toxicity to the embryos. These results showcase the delivery of a specific biomolecule modified with a photoswitch to an organism using microinjection. Cell-penetrating peptide sequences could also be used to allow non-specific peptide delivery, followed by directed photoactivation. [157, 161]

The results obtained might be extended by additional studies exploring the use of more potent activating domains such as the natural viral protein VP16 or the yeast activator Gal4 to potentially achieve higher transcriptional activation. Since different linker domains seem to strongly affect the level of transcription activation, this might also be a fruitful route to improve these artificial transcription factors. Combinations of improvements in visible light photoswitching, activation activity and linker optimisation could ultimately lead to the creation of more potent and easily controllable artificial transcription factor of broad utility in live cells.

References

1) L. Stryer, *Biochemistry*, W. H. Freeman and Company, U.S.A., 4th edn., 1995.

2) M. S. Ciampi, *Microbiology*, 2006, **152**, 2515-2528.

3) C. Grisham, R. Garrett, *Biochemistry*, Thomson Brook/Cole, Canada, 5th edn., 2013.

4) D. Pribnow, *Proc. Natl. Acad. Sci. USA.*, 1975, **72**, 784-788.

5) M. Ptashne, *A genetic switch: Gene Control and Phage λ*, Cell Press & Blackwell Scientific Publications, 1986.

6) J. P. Richardson, *Biochim. Biophys. Acta*, 2002, **1577**, 251-260.

7) P. J. Farnham, T. Platt, *Nucleic Acids Res.*, 1981, **9**, 563-578.

8) K. S. Wilson, P. H. von Hippel, *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 8793-8797.

9) M. Sawadogo, R. G. Roeder, *Cell*, 1985, **43**, 165-175.

10) S. C. Harrison, *Nature*, 1991, **353**, 715-719.

11) W. McGinnis, M. S. Levine, E. Hafen, A. Kuroiwa, W. J. Gehring, *Nature*, 1984, **308**, 428-433.

12) S. Schneuwly, R. Klemenz, *Nature*, 1987, **325**, 816-818.

13) J. H. Postlethwait, H. A. Schneiderman, *Proc. Natl. Acad. Sci. USA*, 1969, **64**, 176-183.

14) P. A. Lawrence, G. Morata, *Cell*, 1994, **78**, 181-189.

15) W. McGinnis, R. Krumlauf, *Cell*, 1992, **68**, 283-302.

16) R. J. White, *Gene Transcription: Mechanism and Control*, Blackwell Science, UK, 2001.

17) J. Locker, *Transcription Factors*, BIOS Scientific Publishers Limited, UK, 2001.

18) E. Fraenkel, M. A. Rould, K. A. Chambers, C. O. Pabo, *J. Mol. Biol.*, 1998, **284**, 351-361.

19) W. J. Gehring, M. Affolter, T. Bürglin, *Annu. Rev. Biochem*. 1994, **63**, 487-526.

20) Y. Q. Qian, D. Resendez-Perez, W. J. Gehring, K. Wüthrich, *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 4091-4095.

21) C. Pabo, R. T. Sauer, *Annu. Rev. Biochem.*, 1992, **61**, 1053-1095.

22) C. Wolberger, *Annu. Rev. Biophys. Biomol. Struct.*, 1999, **28**, 29-56.

23) I. A. Hope, K. Struhl, *EMBO J.*, 1987, **6**, 2781-2784.

24) M. Karin, Z. G. Liu, E. Zandi, *Curr. Opin. Cell. Biol.*, 1997, **9**, 240-246.

25) M. A. Weiss, T. E. Ellenberger, C. R. Wobbe, J. P. Lee, S. C. Harrison, K. Struhl, *Nature*, 1990, **347**, 575-578.

26) T. E. Ellenberger, C. J. Brandl, K. Struhl, S. C. Harrison, *Cell*, 1992, **71**, 1223-1237.

27) C. Vinson, A. Acharya, E. J. Taparowsky, *Biochim. Biophys. Acta*, 2006, **1759**, 4-12.

28) K. S. Thompson, C. R. Vinson, E. Freire, *Biochemistry*, 1993, **32**, 5491-5496.

29) J. Moitra, L. Szilák, D. Krylov, C. R. Vinson, *Biochemstry*, 1997, **36**, 12567-12573.

30) A. Acharya, S. B. Ruvinov, J. Gal, J. R. Moll, C. R. Vinson, *Biochemistry*, 2002, **41**, 14122-14131.

31) C. R. Vinson, T. Hai, S. M. Boyd, *Genes Dev.*, 1993, **7**, 1047-1058.

32) E. K. O'Shea, J. D. Klemm, P. S. Kim, T. Alber, *Science*, 1991, **254**, 539-544.

33) Y. H. Chen, J. T. Yang, K. H. Chau, *Biochemistry*, **13**, 1974, 3350-3359.

34) P. Lavigne, M. P. Crump, S. M. Gagné, R. S. Hodges, C. M. Kay, B. D. Sykes, *J. Mol. Biol.*, 1998, **281**, 165-181.

35) T. Alber, *Curr. Opin. Genet. Dev.*, 1992, **2**, 205-210.

36) J. N. M. Glover, S. C. Harrison, *Nature*, 1995, **373**, 257-261.

37) W. H. Landschulz, P. F. Johnson, S. L McKnight, *Science*, 1989, **243**, 1681-1688.

38) J. Miller, A. D. McLachlan, A. Klug, *EMBO*, 1985, **4**, 1609-1614.

39) R. M. Evans, S. M. Hollenberg, *Cell*, 1988, **52**, 1-3.

40) A. Klug, J. W. R. Schwabe, *FASEB Journal*, 1995, **9**, 597-604.

41) M. Elrod-Erickson, M. A. Rould, L. Nekludova, C. O. Pabo, *Structure*, 1996, **4**, 1325-1336.

42) M. S. Lee, G. P. Gippert, K. V. Soman, D. A. Case, P. E. Wright, *Science*, 1989, **245**, 635-637.

43) A. Klug, *Proc. Japan Acad.*, **81**, 2005 Ser. B, 87-102.

44) C. Murre, P. Schonleber-McCaw, D. Baltimore, *Cell*, 1989, **56**, 777-783.

45) M. Sieber, R. K. Allemann, *Nucleic Acids Res.*, 2000, **28**, 2122-2127.

46) A. B. Lassar, J. N. Buskin, D. Lockshon, R. L. Davis, S. Apone, S. D. Hauschka, H. Weintraub, *Cell*, 1989, **58**, 823-831.

47) C. Murre, P. Schonleber-McCaw, H. Vaessin, M Caudy, L. Y. Jan, Y. N. Jan, C. V. Cabrera, J. N. Buskin, S. D. Hauschka, A. B. Lassar, H. Weintraub, D. Baltimore, *Cell*, 1989, **58**, 537-544.

48) P. C. M. Ma, M. A. Rould, H. Weintraub, C. O. Pabo, *Cell*, 1994, **77**, 451-459.

49) T. J. Brennan, T. Chakraborty, E. N. Olson, *Proc. Natl. Acad. Sci. USA*, 1991, **88**, 5675-5679.

50) R. L. Davis, P. F. Cheng, A. B. Lassar, H. Weintraub, *Cell*, 1990, **60**, 733-746.

51) R. L. Davis, H. Weintraub, *Science*, 1992, **256**, 1027-1030.

52) S. Kanshal, J. W. Schneider, B. Nudal-Ginard, V. Mahdavi, *Science*, 1994, **266**, 1236-1240.

53) R. V. Weatherman, R. J. Fletterick, T. S. Scanlan, *Annu. Rev. Biochem.*, 1999, **68**, 559-581.

54) J. M. Olefsky, *J. Biol. Chem.*, 2001, **276**, 36863-36864.

55) D. J. Mangelsdorf, C. Thummel, M. Beato, P. Herrlich, G. Schütz, K. Umesono, B. Blumberg, P. Kastner, M. Mark, P. Chambon, R. M. Evans, *Cell*, 1995, **83**, 835-839.

56) B. F. Luisi, W. X. Xu, Z. Otwinowski, L. P. Freedman, K. R. Yamamoto, P. B. Sigler, *Nature*, 1991, **352**, 497-505.

57) S. Passmore, R. Elble, B. K. Tye, *Genes Dev.*, 1989, **3**, 921-935.

58) Z. Schwarz-Sommer, P. Huijser, W. Nacken, H. Saedler, H. Sommer, *Science*, 1990, **250**, 931-936.

59) L. Pellegrini, S. Tan, T. J. Richmond, *Nature*, 1995, **376**, 490-498.

60) P. Shore, A. D. Sharrocks, *Eur. J. Biochem.*, 1995, **229**, 1-13.

61) Y. T. Yu, R. E. Breitbart, L. B. Smoot, V. Mahdavi, B. Nadal-Ginard, *Genes & Dev.*, 1992, **6**, 1783-1798.

62) J. D. Molkentin, B. L. Black, J. F. Martin, E. N. Olson, *Cell*, 1995, **83**, 1125-1136.

63) J. D. Molkentin, E. N. Olson, *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 9366-9373.

64) B. L. Black, J. D. Molkentin, E. N. Olson, *Mol. Cell. Biol.*, 1998, **18**, 69-77.

65) D. B. Nikolov, H. Chen, E. D. Halay, A. Hoffmann, R. G. Roeder, S. K. Burley, *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 4862-4867.

66) J. L. Kim, D. B. Nikolov, S. K. Burley, *Nature*, 1993, **365**, 520-527.

67) Y. Kim, J. H. Geiger, S. Hahn, P. B. Sigler, *Nature*, 1993, **365**, 512-520.

68) G. Ghosh, G. V. Duyne, S. Ghosh, P. B. Sigler, *Nature*, 1995, **373**, 303-310.

69) C. W. Müller, F. A. Rey, M. Sodeoka, G. L. Verdine, S. C. Harrison, *Nature*, 1995, **373**, 311-317.

70) C. O Pabo, L. Nekludova, *J. Mol. Biol.*, 2000, **301**, 597-624.

71) A. Klug, *Gene*, 1993, **135**, 83-92.

72) P. C. M. Ma, M. A. Rould, H. Weintraub, C. O. Pabo, *Cell*, 1994, **77**, 451-459.

73) C. W. Garvie, C. Wolberger, *Mol. Cell*, 2001, **8**, 937-946.

74) N. M. Luscombe, R. A. Laskowski, J. M. Thornton, *Nucleic Acids Res.*, 2001, **29 (13)**, 2860-2874.

75) Y. Takeda, A. Sarai, V. M. Rivera, *Proc. Natl. Acad. Sci. USA*, 1989, **86**, 439-443.

76) D. E. Frank, R. M. Saecker, J. P. Bond, M. W. Capp, O. V. Tsodikov, S. E. Melcher, M. M. Levandoski, M. T. Record jr., *J. Mol. Biol.*, 1997, **267**, 1186-1206.

77) D. Meierhans, C. el-Ariss, M. Neuenschwander, M. Sieber, J. F. Stackhouse, R. K. Allemann, *Biochemistry*, 1995, **34**, 11026-11036.

78) T. Lundback, T. Hard, *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 4754-4759.

79) S. E. Ades, R. T. Sauer, *Biochemistry*, 1994, **33**, 9187-9194.

80) A. Fersht, *Structure and mechanims in protein sceince: A guide to enzyme catalysis and protein folding*, W. H. Freeman and Company, USA, 1998.

81) J. M. Scholtz, R. L. Baldwin, *Annu. Rev. Biomol. Struct.*, 1992, **21**, 95-118.

82) A. J. Doig, A. Chakrabartty, T. M. Klingler, R. L. Baldwin, *Biochemistry*, 1994, **33**, 3396-3403.

83) A. Chakrabartty, T. Kortemme, R. L. Baldwin, *Protein Science*, 1994, **3**, 843-852.

84) L. G. Presta, G. D. Rose, *Science*, 1988, **240**, 1632-1641.

85) J. S. Richardson, D. C. Richardson, *Science*, 1988, **240**, 1648-1652.

86) A. J. Doig, R. L. Baldwin, *Protein Sci.*, 1995, **4**, 1325-1336.

87) G. Osapay, J. W. Taylor, *J. Am. Chem. Soc.*, 1992, **114**, 6966-6973.

88) D. Y. Jackson, D. S. King, J. Chmielewski, S. Singh, P. G. Schultz, *J. Am. Chem. Soc.*, 1991, **113**, 9391-9392.

89) J. H. B. Pease, R. W. Storrs, D. E. Wemmer, *Proc. Natl. Acad. Sci. USA*, 1990, **87**, 5643-5647.

90) L. Mayne, S. W. Englander, R. Qui, J. Yang, Y. Gong, E. J. Spek, N. R. Kallenbach, *J. Am. Chem. Soc.*, 1998, **120**, 10643-10645.

91) E. Cabezas, A. C. Satterthwait, *J. Am. Chem. Soc.*, 1999, **121**, 3862-3875.

92) K. R. Rajashankar, S. Ramakumar, R. M. Jain, V. S. Chauhan, *J. Am. Chem. Soc.*, 1995, **117**, 10129-10130.

93) J. H. Miwa, L. Pallivathucal, S. Gowda, K. E. Lee, *Org. Lett.*, 2002, **4**, 4655-4657.

94) C. E. Schafmeister, J. Po, G. L. Verdine, *J. Am. Chem. Soc.*, 2000, **122**, 5891-5892.

95) T. Alber, *Curr. Opin. Genet. Dev.*, 1992, **2**, 205-210.

96) J. E. Brown, W. A. Klee, *Biochemistry*, 1971, **10**, 470-476.

97) M. E. Vázquez, A. M. Caamaño, J. L. Mascareñas, *Chem. Soc. Rev.*, 2003, **32**, 338-349.

98) B. C. Cunningham, J. A. Wells, *Curr. Opin. Struct. Biol.*, 1997, **7**, 457-462.

99) P. A. Nygren, M. Uhlen, *Curr. Opin. Struct. Biol.*, 1997, **7**, 463-469.

100) N. J. Zondlo, A. Schepartz, *J. Am. Chem. Soc.*, 1999, **121**, 6938-6939.

101) J. W. Chin, A. Schepartz, *J. Am. Chem. Soc.*, 2001, **123**, 2929-2930.

102) L. Yang, A. Schepartz, *Biochemistry*, 2005, **44**, 7469-7478.

103) J. K. Montclare, A. Schepartz, *J. Am. Chem. Soc.*, 2003, **125**, 3416-3417.

104) C. J. Weston, C. H. Cureton, M. J. Calvert, O. S. Smart, R. K. Allemann, *ChemBioChem*, 2004, **5**, 1075-1080.

105) C. Granier, E. Pedroso-Muller, J. Van Rietschoten, *Eur. J. Biochem.*, 1978, **82**, 293-299.

106) R. V. Talanian, C. J. McKnight, P. S. Kim, *Science*, 1990, **249**, 769-771.

107) C. Yu, J. W. Taylor, *Bioorg. Med. Chem.*, 1999, **7**, 161-175.

108) K. Fujimoto, N. Oimoto, K. Katsuno, M. Inouye, *Chem. Comm.* 2004, **11**, 1280-1281.

109) A. Kocer, M. Walko, W. Meijberg, B. L. Feringa, *Science*, 2005, **309**, 755-758.

110) C. R. Crecca, A. E. Roitberg, *J. Phys. Chem. A*, 2006, **110**, 8188-8203.

111) H. M. D. Bandara, S. C. Burdette, *Chem. Soc. Rev.*, 2012, **41**, 1809-1825.

112) H. Rau, *Angew. Chem. Int. Ed.*, 1973, **12**, 224-235.

113) A. R. Dias, M. E. Minas da Piedade, J. A. Martinho Simões, J. A. Simoni, C. Teixeira, H. P. Diogo, Y. Meng-Yan, G. Pilcher, *J. Chem. Thermodyn.*, 1992, **24**, 439-447.

114) D. G. Flint, J. R. Kumita, O. S. Smart, G. A. Woolley, *Chemistry & Biology*, **9**, 391-397.

115) H. Rau, E. Lueddecke, *J. Am. Chem. Soc.*, 1982, **104**, 1616-1620.

116) J. L. Magee, W. Shand, Jr. H. Eyring, *J. Am. Chem. Soc.*, 1941, **63**, 677-688.

117) D. Y. Curtin, E. J. Grubbs, C. G. McCarty, *J. Am. Chem. Soc.*, 1966, **88**, 2775-2786.

118) G. Hallas, M. A. Jalil, *Dyes Pigm.*. 1996, **32**, 129-133.

119) O. Sadovski, A. A. Beharry, F. Zhang, G.A. Woolley, *Angew. Chem. Int. Ed.*, 2009, **48**, 1484-1486.

120) Q. Dong, K. Svoboda, T. R. Tiersch, W. T. Monroe, *J. Photochem. Photbiol., B*, 2007, **88**, 137-146.

121) J. Forman, M. Dietrich, W. T. Monroe, *Photochem. Photbiol. Sci.*, 2007, **6**, 649-658.

122) A. A. Beharry, G. A. Woolley, *Chem. Soc. Rev.*, 2011, **40**, 442-4437.

123) G. Hallas, R. Marsden, J. D. Hepworth, D. Mason, *J. Chem. Soc. Perkin Trans. II*, 1986, 123-126.

124) J. Dokić, M. Gothe, J. Wirth, M. V. Peters, J. Schwarz, S. Hecht, P. Saalfrank, *J. Phys. Chem. A*, 2009, **113**, 6763-6773.

125) A. A. Beharry, O. Sadovski, G. A. Woolley, *J. Am. Chem. Soc.*, 2011, **133**, 19684-19687.

126) H. A. Wegner, *Angew. Chem. Int. Ed.*, 2012, **51**, 4787-4788.

127) A. A. Beharry, G. A. Woolley, *Chem. Soc. Rev.*, 2011, **40**, 442-4437.

128) N. J. Dunn, W. Humphries, A. R. Offenbacher, T. L. King, J. A. Gray, *J. Phys. Chem. A*, 2009, **113**, 13144-13151.

129) R. J. W. Le Fèvre, J. Northcott, *J. Chem. Soc.*, 1953, 867-870.

130) W. R. Brode, J. H. Gould, G. M. Wyman, *J. Am. Chem. Soc.*, 1952, **74**, 4641-4646.

131) W. R. Brode, J. H. Gould, G. M. Wyman, *J. Am. Chem. Soc.*, 1953, **75**, 1856-1859.

132) K. S. Schanze, T. F. Mattox, D. G. Whitten, *J. Org. Chem.*, 1983, **48**, 2808-2813.

133) H. Fliegl, A. Köhn, C. Hättig, R. Ahlrichs, *J. Am. Chem. Soc.*, 2003, **125**, 9821-9827.

134) R. Behrendt, C. Renner, M. Schenk, F. Wang, J. Wachtveitl, D. Oesterhelt, L. Moroder, *Angew. Chem. Int. Ed.*, 1999, **38**, 2771-2774.

135) C. Renner, R. Behrendt, S. Sporlein, J. Wachtveitl, L. Moroder, *Biopolymers*, 2000, **54**, 489-500.

136) N. Pozhidaeva, M. E. Cormier, A. Chaudhari, G. A. Woolley, *Bioconj. Chem.*, 2004, **15**, 1297-1303.

137) A. A. Beharry, O. Sadovski, G. A. Woolley, *Org. Biomol. Chem.*, 2008, **6**, 4323-4332.

138) S. L. Dong, M. Loweneck, T. E. Schrader, W. J. Schreier, W. Zinth, L. Moroder, C. Renner, *Chem. Eur. J.*, 2006, **12**, 1114-1120.

139) C. Hoppmann, S. Seedorff, A. Richter, H. Fabian, P. Schmieder, K. Rück-Braun, M. Beyermann, *Angew. Chem. Int. Ed.*, 2009, **48**, 6636-6639.

140) B. Z. Harris, W. A. Lim, *J. Cell. Sci.*, 2001, **114**, 3219-3231.

141) J. E. Brenman, D. S. Chao, H. Xia, K. Aldape, D. S. Bredt, *Cell*, 1995, **82**, 743-752.

142) J. R. Kumita, O. S. Smart, G. A. Woolley, *Proc. Natl. Acad. Sci.*, 2000, **97**, 3803-3808.

143) G. A. Woolley, *Acc. Chem. Res.*, 2005, **38**, 486-493.

144) D. C. Burns, D. G. Flint, J. R. Kumita, H. J. Feldman, L. Serrano, Z. Zhang, O. S. Smart, G. A. Woolley, *Biochemistry*, 2004, **43**, 15329-15338.

145) V. Borisenko, G. A. Woolley, *J. Photochem. Photobiol., A*, 2005, **173**, 21-28.

146) Z. Zhang, D. C. Burns, J. R. Kumita, O. S. Smart, G. A. Woolley, *Bioconjugate Chem.*, 2003, **14**, 824-829.

147) S. Kneissl, *Photocontrol of Biomacromolecular Interactions*, 2009, Thesis (School of, Cardiff University).

148) E. Chen, J. R. Kumita, G. A. Woolley, D. S. Kliger, *J. Am. Chem. Soc.*, 2003, **125**, 12443-12449.

149) J. Bredenbeck, J. Helbing, J. R. Kumita, G. A. Woolley, P. Hamm, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 2379-2384.

150) J. A. Ihalainen, B. Paoli, S. Muff, E. H. Backus, J. Bredenbeck, G. A. Woolley, A. Caflisch, P. Hamm, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 9588-9593.

151) J. A. Ihalainen, J. Bredenbeck, R. Pfister, J. Helbing, L. Chi, I. H. van Stokkum, G. A. Woolley, P. Hamm, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 5383-5388.

152) J. R. Kumita, D. G. Flint, O. S. Smart, G. A. Woolley, *Protein Eng.*, 2002, **15**, 561-569.

153) F. Zhang, A. Zarrine-Afsar, M. S. Al-Abdul-Wahid, R. S. Prosser, A. R. Davidson, G. A. Woolley, *J. Am. Chem. Soc.*, 2009, **131**, 2283-2289.

154) L. Guerrero, O. S. Smart, G. A. Woolley, R. K. Allemann, *J. Am. Chem. Soc.*, 2005, **127**, 15624-15629.

155) L. Guerrero, O. S. Smart, C. J. Weston, D. C. Burns, G. A. Woolley, R. K. Allemann, *Angew. Chem. Int. Ed.*, 2005, **44**, 7778-7782.

156) G. A. Woolley, A. S. I. Jaikaran, M. Berezovski, J. P. Calarco, S. N. Krylov, O. S. Smart, J. R. Kumita, *Biochemstry*, 2006, **45**, 6075-6084.

157) F. Zhang, K. A. Timm, K. M. Arndt, G. A. Woolley, *Angew. Chem. Int. Ed.*, 2010, **49**, 3943-3946.

158) R. J. Mart, P. Wysoczanski, S. Kneissl, A. Ricci, A. Brancale, R. K. Allemann, *ChemBioChem.*, 2012, **13**, 515-519.

159) P. Wysoczanski, R. J. Mart, E. J. Loveridge, C. Williams, S. B. M. Whittaker, M. P. Crump, R. K. Allemann, *J. Am. Chem. Soc.*, 2012, **134**, 7644-7647.

160 J. W. Chin, A. Schepartz, *Angew. Chem. Int. Ed.*, 2001, **40**, 3806-3809.

161) T. Fricke, R. J. Mart, C. L Watkins, M. Wiltshire, R. J. Errington, P. J. Smith, A. T. Jones, R. K. Allemann, *Bioconjugate Chem.*, 2011, **22**, 1763-1767.

162) C. Boulegue, M. Loweneck, C. Renner, L. Moroder, *ChemBioChem*, 2007, **8**, 591-594.

163) A. A. Beharry, L. Wong, V. Tropepe, G. A. Woolley, *Angew. Chem. Int. Ed.*, 2011, **50**, 1325-1327.

164) Y. Zhang, F. Erdmann, G. Fischer, *Nat. Chem. Biol.*, 2009, **5**, 724-726.

165) M. Bose, D. Groff, J. Xie, E. Brustad, P. G. Schultz, *J. Am. Chem. Soc.*, 2006, **128**, 388-389.

166) E. Giniger, M. Ptashne, *Nature*, 1987, **330**, 670-672.

167) A. K. Mapp, A. Z. Ansari, M. Ptashne, P. B. Dervan, *Proc. Natl. Acad. Sci.*, 2000, **97**, 3930-3935.

168) D. Y. Jackson, D. S. King, J. Chmielewski, S. Singh, P. G. Schultz, *J. Am. Chem. Soc.*, 1991, **113**, 9391-9392.

169) J. R. Lakowicz, *Principles of fluorescence spectroscopy.*, 1999, NY, Plenum Press.

170) T. Heyduk, Y. Ma, H. Tang, R. H. Ebright, *Methods Enzymol.*, 1996, **274**, 492-503.

171) A. K. Mapp, A. Z. Ansari, *Curr. Opin. Chem. Biol.,* 2002, **6**, 765-772.

172) N. D. Clarke, C. R. Kissinger, J. Desjarlais, G. L. Gilliland, C. O. Pabo, *Protein Sci.*, 1994, **3**, 1779-1787.

173) K. T. O'Neil, J. D, Shuman, C. Ampe, W. F. Degrado, *Biochemistry*, 1991, **30**, 9030-9034.

174) L. Tucker-Kellogg, M. A. Rould, K. A. Chambers, S. E. Ades, R. T. Sauer, C. O. Pabo, *Structure*, 1997, **5**, 1047-1054.

175) L. Guerrero, *Photocontrol of the Specificity of DNA Binding Proteins*, 2005, Thesis (Chemistry department, University of Birmingham).

176) D. C. Burns, F. Zhang, G. A. Woolley, *Nat. Protoc.,* 2007, **2**, 251-258.

177) C. A. Rohl, W. Fiori, R. L. Baldwin, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 3682-3687.

178) F. D. Bushman, M. Ptashne, *Proc. Natl. Acad. Sci. USA*, 1986, **83**, 9353-9357.

179) A. J. Courey, R. Tjian, *Cell*, 1988, **55**, 887-898.

180) H. P. Gerber, K. Seipel, O. Georgiev, M. Höfferer, M. Hug, S. Rusconi, W. Schaffner, *Science*, 1994, **263**, 808-811.

181) A. Gaussin, U. Modlich, N. J. Niederländer, A Schambach, C. Duros, A. Artus, C. Baum, O. Cohen-Haguenauer, N. Mermod, *Gene Therapy*, 2012, **19**, 15-24.

182) N. Mermod, E. A. O'Neill, T. J. Kelly, R. Tjian, *Cell*, 1989, **58**, 741-753.

183) P. J. Mitchell, R. Tjian, *Science*, 1989, **245**, 371-378.

184) A. Z. Ansari, A. K. Mapp, D. H. Nguyen, P. B. Dervan, M. Ptashne, *Chem. Biol.*, 2001, **8**, 583-592.

185) M. Sieber, R. K. Allemann, *Nucleic Acids Res.*, 1998, **26** (**6**), 1408-1413.

186) M. Archetti, *J. Mol. Evol.*, 2004, **59**, 400-415.

187) B. D. Spangler, *Methods in Molecular Biology and Protein Chemistry*, Wiley, U.S.A., 2002.

188) R. Kaiser, L. Metzka, *Anal. Biochem.*, 1999, **266**, 1-8.

189) S. Kneissl, E. J. Loveridge, C. Williams, M. P. Crump, R. K. Allemann, *ChemBioChem*, 2008, **9**, 3046-3054.

190) G. A. Woolley, *Acc. Chem. Res.*, 2005, **38**, 486-493.

191) A. A. Adzhubei, M. J. E. Sternberg, A. A. Makarov, *J. Mol. Biol.*, 2013, **425**. 2100-2132.

192) A. L. Rucker, T. P. Creamer, *Protein Science*, 2002, **11**, 980-985.

193) A. Starzyk, W. Barber-Armstrong, M. Sridharan, S. M. Decatur, *Biochemistry*, 2005, **44**, 369-376.

194) K. Gast, D. Zirwer, M. Müller-Frohne, G. Damaschun, *Protein. Sci.*, 1999, **8**, 625-634.

195) M. Ndao, K. Dutta, K. M. Bromley, R. Lakshminarayanan, Z. Sun, G. Rewari, J. Moradian-Oldak, J. S. Evans, *Protein Sci.*, 2011, **20**, 724-734.

196) S. E. Ades, R. T. Sauer, *Biochemistry*, 1995, **34**, 14601-14608.

197) S. E. Harding, B. Z. Chowdhry, *Protein-ligand interactions: Structure and Spectroscopy*, 2001, Oxford University Press.

198) P. S. Arora, A. Z. Ansari, T. P. Best, M. Ptashne, P. B. Dervan, *J. Am. Chem. Soc.*, 2002, **124**, 13067-13071.

199) Y. Kwon, H. D Arndt, Q. Mao, Y. Choi, Y. Kawazoe, P. B. Dervan, M. Uesugi, *J. Am. Chem. Soc.*, 2004, **126**, 15940-15941.

200) M. Sawadogo, R. G. Roeder, *Cell*, 1985, **43**, 165-175.

201) J. C. Wang, M. Sawadogo, M. W. Dyke, *Biochim. Biophys. Acta*, 1998, **1397**, 141-145.

202) M. Sawadogo, R. G. Roeder, *Proc. Natl. Acad. Sci. USA*, 1985, **82**, 4394-4398.

203) N. F. Lue, P. M. Flanagan, P. M. III, Kelleher, A. M. Edwards, R. D. Kornberg, *Methods Enzymol.*, 1991, **194**, 545-551.

204) B. A Wallace, *Dichroweb*, 2013, dichroweb.cryst.bbk.ac.uk (Institute of Structural and Molecular Biology, Birbeck College, University of London).

205) R. J. Mart, P. Wysoczanski, S. Kneissl, A. Ricci, A. Brancale, R. K. Allemann, *ChemBioChem*, 2012, **13**, 515-519.

206) R. J. Mart, R. J. Errington, C. L Watkins, S. L. Chappell, M. Wiltshire, A. T. Jones, P. J. Smith, R. K. Allemann, *Mol.Biosyst.*, 2013, DOI: 10.1039/c3mb70246d.

207) D. Bléger, J. Schwarz, A. M. Brouwen, S. Hecht, *J. Am. Chem. Soc.*, 2012, **134**, 20597-20600.

208) S. Samanta, A. A. Beharry, O. Sadovski, T. M. McCormick, A. Babalhavaeji, V. Tropepe, G. A. Woolley, *J. Am. Chem. Soc.*, 2013, **135**, 9777-9784.