A SYSTEM FOR THE GENERATION AND IDENTIFICATION OF MUTANT DNA-BINDING PROTEINS

By

Scott Webster

Department of Chemistry

University of Edinburgh

A thesis submitted for the degree of Doctor of Philosophy at the

University of Edinburgh

August 1995

DECLARATION

I, Scott Webster, hereby certify that this thesis has been composed by myself, that it is a record of my own work, and that it has not been accepted in partial or complete fulfilment of any other degree of professional qualification.

University of Edinburgh

August 1995

ACKNOWLEDGEMENTS

Abstract

GCN4 is a transcriptional activatory protein found in *S. cerevisiae* and is a member of the basic region leucine zipper (bZip) family of DNA-binding proteins. The full length protein is 281 residues in length and consists of an activatory region and a DNA-binding region. The DNA-binding region of the protein consists of two separate regions; a leucine zipper through which the protein dimerises, and a basic region which contacts target DNA. Various studies have shown that it is possible to mutate residues in the basic region and create proteins with changed specificity. The aim of this project was to use λZAPII technology to create a large library of mutants and investigate the effects of single and double mutants in the basic region and linker segment of GCN4.

The DNA-binding (bZip) region of GCN4 was extracted from a plasmid containing the entire GCN4 gene and cloned into the *lacZ* gene of the bacteriophage vector λZAPII. The recombinant vector was infected into a strain of *E. coli* which allowed the bZip gene to be expressed as a fusion protein with β-galactosidase. *In vitro* screening experiments were carried out and any 'positive' clones which bound to a DNA target identified. The DNA sequence of each positive clone was then determined. These experiments demonstrated that the λZAPII system was an excellent vehicle for the expression and identification of a DNA library and that identical conditions could be used to screen a large library of mutants.

A phagemid containing the bZip gene was excised from λZAPII *in vivo* and infected into *E. coli*. These cells were used to overexpress the β-galactosidase-bZip fusion protein. The fusion protein was purified and shown to bind with high affinity to a specific DNA target.

Single-stranded phagemid DNA was prepared *in vivo* and used as a template to attempt *in vitro* mutagenesis across a large portion of the GCN4 basic region via phosphorothioate based spiked mutagenesis.

ABBREVIATIONS

ATF	activating transcription factor
bZip	basic region leucine zipper
CD	circular dichroism
C/EBP	CAAT-box enhancer binding protein
CREB	cyclic-AMP response element binding site/protein
dNTP	deoxynucleotide triphosphate
GCN4	general control non-derepressable 4
HEPES	N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulphonic acid)
IPTG	isopropylthiogalactoside
LMP	low melting point
NMR	nuclear magnetic resonance
PCR	polymerase chain reaction
RF	replicative form
TAF-1	transcriptional activatory factor-1
TBP	transcription factor binding protein
TFII(A-H)	transcription factor II (A-H)
Tris	tris(hydroxymethyl)aminomethane
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

CONTENTS

CHAPTER ONE
Introduction

1.1 TRANSCRIPTIONAL REGULATION.

In eukaryotic organisms, initiation and regulation of transcription is dependant on the interaction of several proteins with the promoter elements of DNA. RNA polymerase II (pol II), which is sufficient to initiate transcription in prokaryotes, cannot initiate transcription alone and is reliant on a number of other proteins known as transcripton factors[1].

Genes to be regulated contain promoter sequences which are composed, in general, of TATA and initiator elements. These elements are necessary for the initiation and accuracy of transcription. The regulation of genes may also be affected by sequences known as enhancers, which can act at great distances from the promoter elements. In yeast, sequences known as upstream activator sequences (UAS), which have an analogous function, are present[2].

Initiation of transcription occurs when a complex of initiation factors is formed on the TATA box. Initially, TFIID is bound to the TATA element of the promoter. This complex, which also consists of TBP and the TBP-associated factors (TAF's), acts as a binding site for TFIIB. Pol II and TFIIF are then recruited into the complex. To complete formation of the initiation complex TFIIE and TFIIH become associated and transcription may occur[3,4] (Fig.1.1).

The rate of transcription may be regulated by additional transcription regulatory proteins which recognise specific DNA sequences within enhancer or UAS elements. These proteins may then participate in highly specific interactions with the various components of the initiation complex[5,6].

2
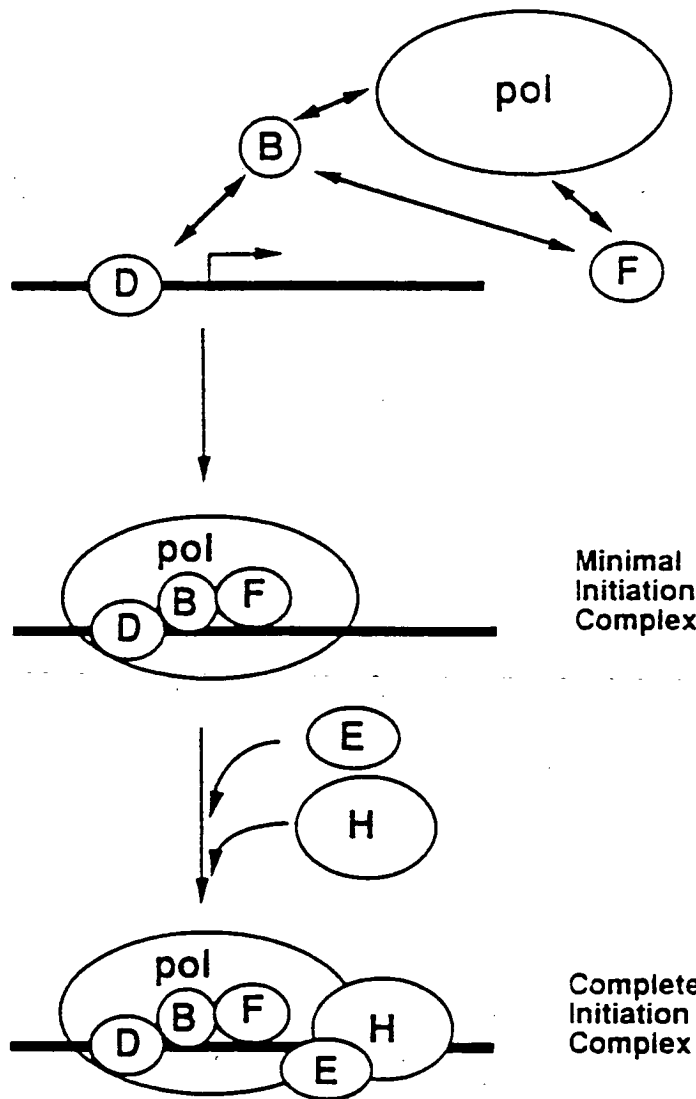
Transcription factors are represented by their letter designations. The bold line represents the promoter DNA while the bent arrow indicates the initiation site. Double-headed arrows indicate protein-protein interactions.

Fig.1.1. Transcription Initiation Complex Assembly.

These transcriptional regulatory proteins invariably consist of two distinct regions; an activation domain and a DNA-binding domain. Several types of activation domains have been identified and classified on the basis of their amino acid composition as acidic, glutamine rich and proline rich[7,8]. These domains may be relatively short and are often unstructured. A structural change my occur, however, upon association with one or more of the components of the initiation complex.

Structural and sequence comparisons of transcriptional regulatory proteins have revealed that the DNA-binding domains of many of these proteins form similar three dimensional structures. Studies have shown that it is possible to classify certain proteins according to their shape. As a result several structural families have been described whereby each member forms the distinct shape or motif of that family. Each of these motifs involves one or more units of secondary structure (usually an α-helix) that are complementary to the structure of B-DNA.

1.2 DNA BINDING MOTIFS.

The structure of numerous transcriptional regulatory proteins have been determined. In this introduction, I have attempted to give an overveiw of the best characterised DNA-binding motifs.
Several excellent reviews have been published which give an insight into the various structures adopted by these proteins[9,10,11].

1.3 THE HELIX-TURN-HELIX MOTIF.

The helix-turn-helix (HTH) motif was the first DNA-recognition motif discovered[12,13]. Most of the prokaryotic regulatory proteins belong to this group. Many protein structures containing the HTH motif have been solved by X-ray crystallography and NMR, including those of λ Cro protein, λ repressor and the trp repressor[14,15,16,17,18].
The HTH is generally defined as a 20-residue segment, with two α-helices that cross at an angle of about 120°. Additional helical residues may also be present at the N-terminus of the first helix or the C-terminus of the second. The HTH (unlike many other motifs) is not a separate, stable domain. It cannot fold or function by itself, but always occurs as part of a larger DNA-binding domain.
Comparisons of HTH protein-DNA complexes have shown that a number of common features exist. Repressor proteins bind to DNA as dimers with each monomer recognising a half-site. The first helix of the HTH unit does not contact the DNA, but the N-terminus of this helix contacts the DNA backbone. The second helix, however, fits into the major groove with the N-terminal portion being closest to the edges of the base pairs. Site-specific contacts with bases in the major groove are made by amino acid

5

side-chains. Recognition is also acheived via a network of hydrogen bonds between the protein and the DNA backbone (Fig.1.2).
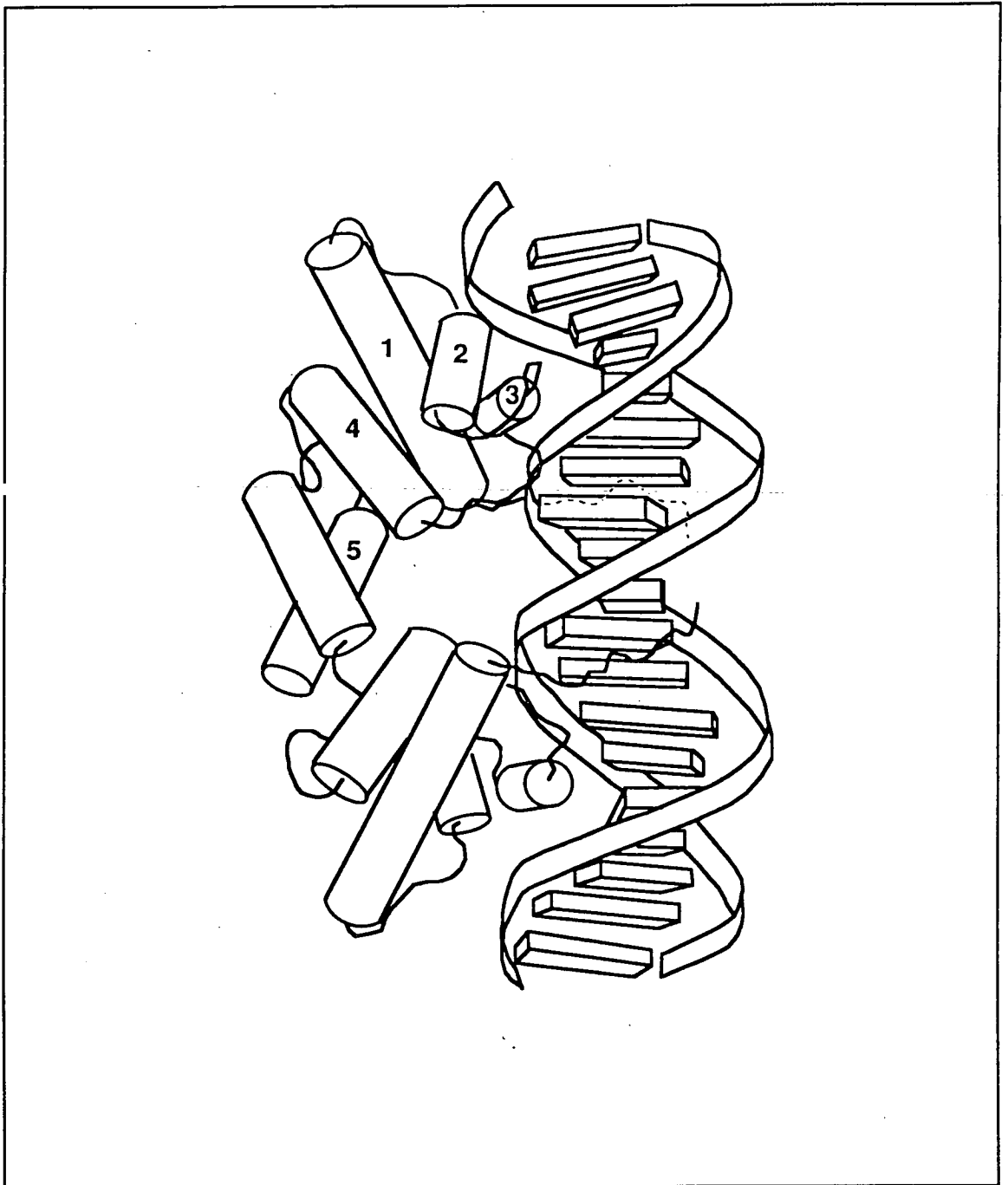
Fig.1.2. Schematic representation of a HTH-DNA complex.

1.4 HOMEODOMAINS.

Homeodomains are 60 amino acid DNA-binding domains which consist of three α-helices and an amino-terminal arm[19]. Structural studies of various homeodomains have shown that they contain a HTH motif. However, unlike the HTH, the 60-residue homeodomain forms a stable, folded structure and can bind to DNA as a monomer by itself.

The structures of the various homeodomains are very similar. Helix 1 and helix 2 pack against each other in an antiparallel arrangement, with helix 3 roughly perpendicular to them. The hydrophobic face of helix 3 packs against helices 1 and 2 to form the interior of the domain (Fig.1.3). Contacts with DNA are made by residues of helix 3, which fits into the major groove, and by residues in the extended N-terminal arm, which fits into the minor groove. Comparison of two homeodomains, MATα2 and the engrailed homeodomain, shows that the contacts made between the protein and the DNA are very similar[20,21].
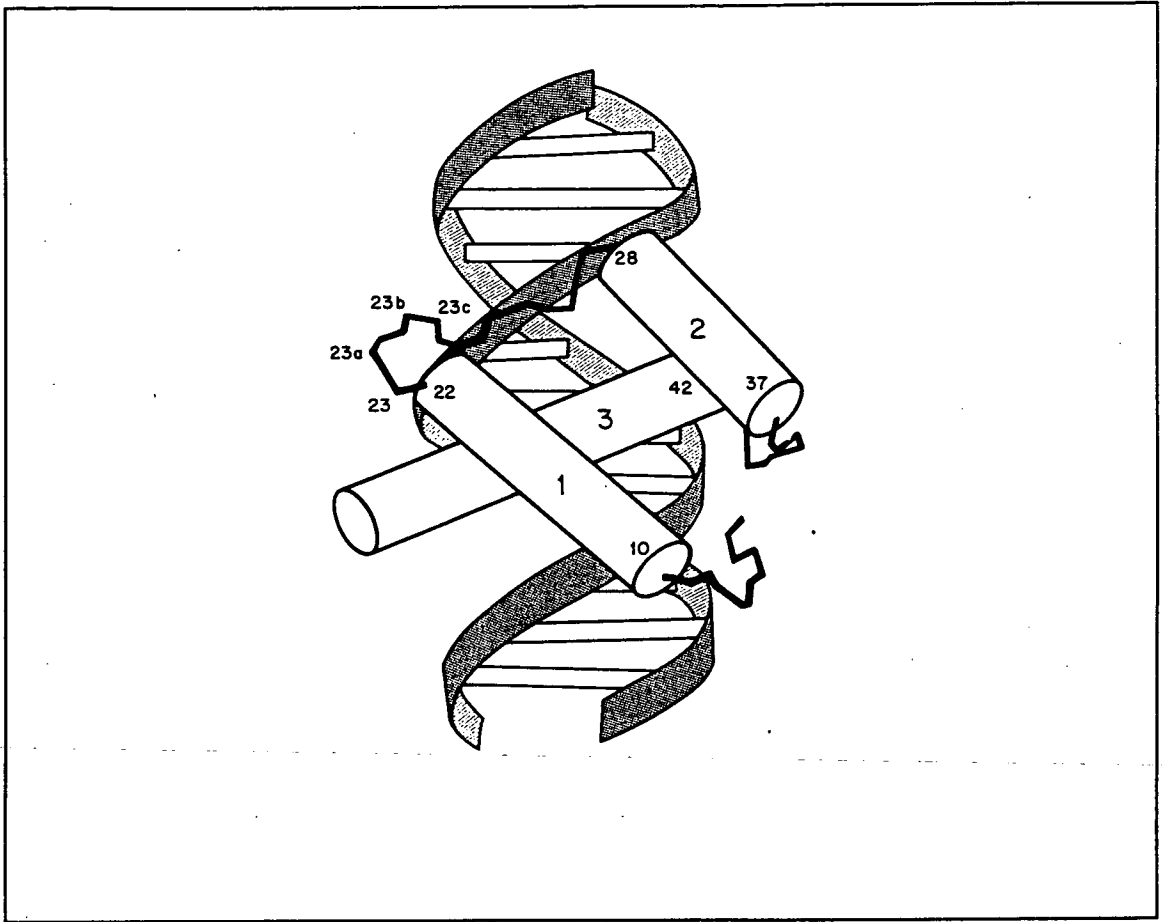
Fig.1.3. Schematic representation of a homeodomain-DNA complex.

1.5 ZINC BINDING DOMAINS.

Three classes of zinc binding domains have been identified, each with a different structure[22]. The first (Type I) is the original "Zn finger" which is a 30-residue peptide with two cysteines and two histidines coordinated to one Zn^{2+} ion[23]. These proteins usually contain the sequence pattern

$$Cys-X_{2or4}-Cys-X_{12}-His-X_{3-5}-His$$
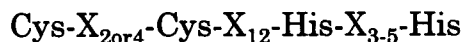
The majority of proteins containing these units have three or more fingers in direct succession. The basic structure consists of a 12-residue α-helix packed against a β-hairpin. Both the cysteines flank the turn in the hairpin, and the two Zn-binding histidines are on the inward facing side of the helix. This structure serves to hold the β sheet and the α-helix together to form a compact globular domain (Fig.1.4).

The second type of Zn-binding domain (Type II) contains two loop-helix elements, each with a Zn^{2+} ion liganded by two cysteines at the beginning of the loop and two others at the N-terminus of the α-helix[24]. The helices are packed against each other at approximately 90°. DNA-binding specificity is determined by residues near the N-terminus which lie in the major groove. Unlike the classical Zn finger, these proteins bind to DNA as dimers with residues at the beginning of the second loop responsible for dimerisation (Fig.1.4).

The final class of Zn-binding domain (Type III) has only been found in yeast. The most studied of these proteins is GAL4[25]. The DNA-binding domain of GAL4 is comprised of the 65 amino-terminal residues. This region of the protein retains its ability to dimerise and bind to DNA upon isolation from the intact protein. The core of the DNA-binding domain is folded around two Zn^{2+} ions which are tetrahedrally coordinated by six cysteine residues, forming what is known as a 'binuclear cluster'. The

metal binding domain mediates sequence-specific contacts by insertion of each monomer into the major groove on opposite faces of the DNA. The carboxy-terminal residues of each monomer form a two-stranded coiled-coil via association of each amphipathic α-helix (Fig.1.4).
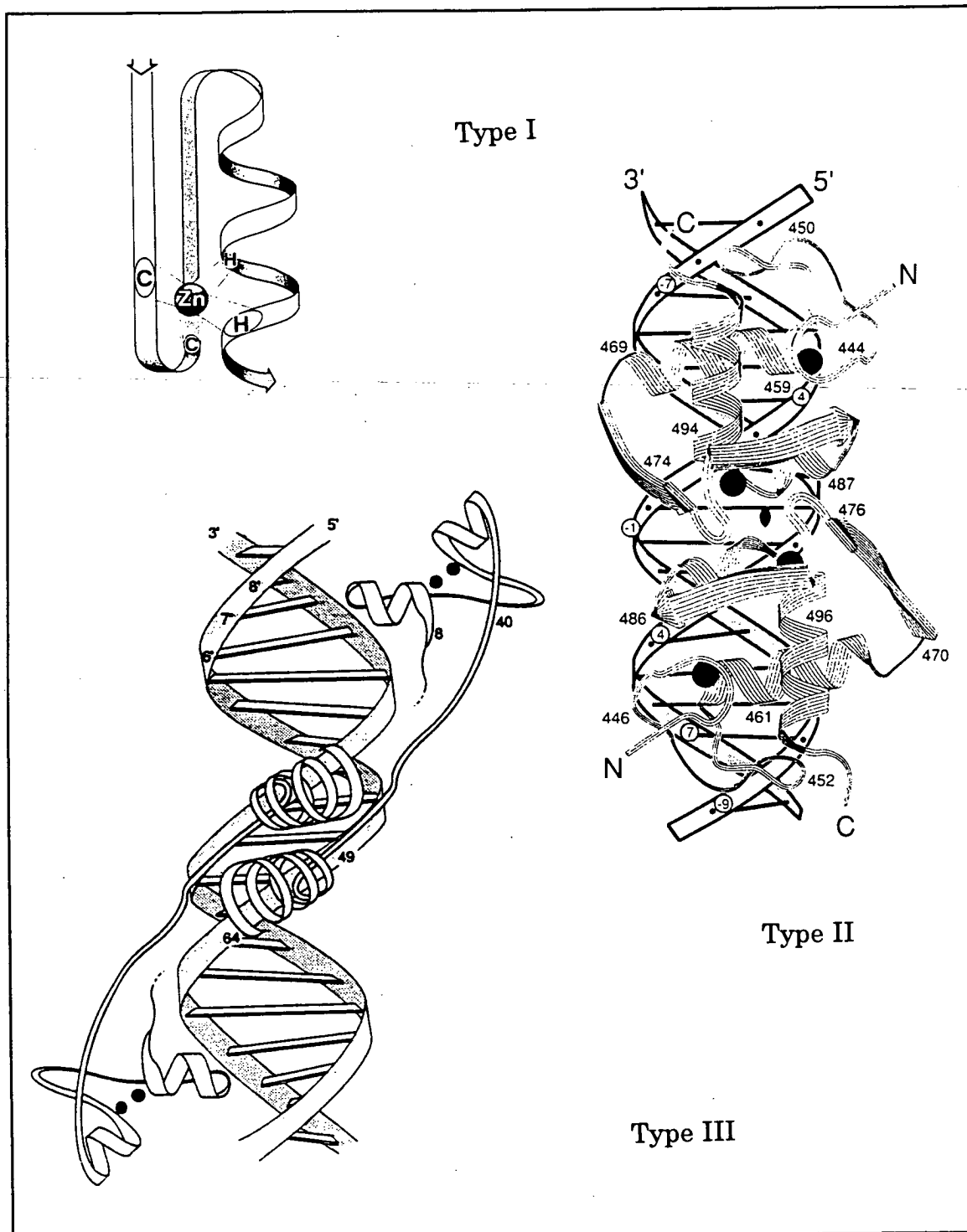
Fig.1.4. Zinc binding domains.

1.6 BASIC REGION LEUCINE ZIPPER PROTEINS.

One of the most recently discovered, and intensely studied, DNA-binding motifs is the basic region leucine zipper (bZip) motif[26]. Several excellent reviews describing its dimerisation and DNA-binding properties have been published[27,28,29,30].

1.6.1 Background.

Comparison of the amino acid sequences of a variety of bZip proteins shows that the DNA-binding domain of these proteins consists of two regions, the leucine zipper and basic region, which have distinct characteristics (Fig.1.5).

```
                 BASIC REGION                    LEUCINE ZIPPER

GCN4    DPAALKRARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVGER
cJun    LKREIRLMKNREAARECRRKKKEYVKCLEEKVKTLKAQNSELASTANHLREQVAQL
cFos    EKRRIRRERNKMAAAKCRNRRRELTDTLQAETDQLEDEKSALQTEIANLLKEKEKL
CREB    RKREVRLMKNREAARECRRKKKEYVKCLENRVAVLENQNKTLIEELKALKDLYCHK
C/EBP   NEYRVRRERNNIAVRKSRDKAKQRNVETQQKVLELTSDNDRLRKRVEQLSRELDTL
ATF-1   LKREIRLMKNREAARECRRKKKEYVKCLENRVAVLENQNKTLIEELKTLKDLYSNK
```

Fig.1.5. bZip amino acid sequences.

The leucine zipper region is characterised by the presence of a heptad repeat of leucine residues which are strongly conserved throughout the bZip family. Amino acids which face the dimerisation interface are more strongly conserved between closely related family members than amino acids that face away from the interface. The basic region is characterised by the presence of an abundance of basic amino acids several of which are

highly conserved within the bZip family of proteins. For example, an asparagine, two consecutive alanine residues and a cysteine or serine are conserved in the centre of the basic region. An invariant arginine and several conserved basic residues flank these neutral amino acids. The spacing between these amino acids is identical among virtually all proteins of the bZip family.

The role of each of the regions has been investigated in a number of studies. The leucine zipper model, described below, suggested that the leucine zipper was responsible for dimerisation of the bZip protein which allowed the dimeric protein to interact with its target DNA in some way[31]. Various studies involving molecular biological techniques have addressed the function of each of these regions. Experiments carried out by Kouzarides *et al* investigating the Fos-Jun heterodimer suggested that these regions were functionally very different[32]. Mutagenesis of the leucine residues of both Fos and Jun eliminated dimerisation indicating that the residues in this region were necessary for dimerisation. Mutations in the basic region, meanwhile, were shown to abolish DNA-binding of the heterodimer suggesting that the basic region interacted with DNA in some way. Further experiments involving mutagenesis, this time of the Fos leucine zipper and basic region, indicated the importance of the leucine zipper for dimer formation[33]. In this case it was demonstrated that mutants in the basic region did not disrupt the dimerisation of the proteins, but did affect DNA-binding indicating that the presence of the leucine zipper was soley for the purpose of forming dimers. A model was proposed whereby Fos and Jun complexed through the leucine zipper with each monomer interacting with one half-site of the palindromic DNA.

Domain swap experiments also hinted at the nature of the leucine zipper and basic region. One of the first studies was carried out by Sellers *et al* who generated chimeras between GCN4 and Fos[34]. The ability of the Fos-Jun complex and the GCN4 homodimer to bind to the same DNA-binding site was exploited to show that a chimeric protein consisting of the leucine zipper of GCN4 fused to the Fos basic region was able to dimerise not only with itself but also GCN4. Both these complexes were capable of binding to the same DNA target. Similar experiments involved the generation of chimeras between GCN4 and C/EBP[35]. The results of these experiments showed that the chimeras were able to homodimerise and bind to the target DNA specified by the basic region of the chimera. Both of these studies demonstrated the ability of the leucine zipper to confer dimerisation autonomously upon bZip proteins allowing the basic region to interact with DNA and determine the binding specificity.

1.6.2 The leucine zipper.

The structure of bZip proteins was first addressed by Landschultz and co-workers who termed the motif the leucine zipper[31]. By projecting the amino acid sequence of the rat liver protein C/EBP onto an idealised α-helix they were able to show that a 28 amino acid segment of the DNA-binding domain exhibited considerable amphipathy. This model showed that one side of the α-helix was composed predominantly of hydrophobic residues with leucine residues appearing at every seventh position. The opposite face of the helix was shown to consist mainly of amino acids with charged and polar side chains. Analysis of other proteins with significant sequence similarity such as GCN4, Fos, and Jun showed that segments within the DNA-binding domains of these proteins also formed

13

amphipathic α-helices when arranged on a helical wheel. Closer analysis of the model showed that the polar face of the amphipathic helix displayed charged residues configured in a manner which would allow salt bridge formation. This and the fact that the α-helices displayed a heptad repeat of leucine residues led the authors to postulate that these proteins formed a coiled-coil structure which is commonly found in fibrous proteins[36]. Using computer modelling the authors argued that the leucine residues of one protein monomer would be able to interdigitate with those of another in a parallel or antiparallel manner, although an antiparallel conformation was the favoured possibility.

```
            d e f g a b c d e f g a b c d e f g a b c d e f g a b c d

GCN4        L E D K V E E L L S K N Y H L E N E V A R L K K L V G E R
Jun         L E E K V K T L K A Q N S E L A S T A N H L R E Q V A Q L
JunB        L E D K V K T L K A E N A G L S S A A G L L R E Q V A Q L
Fos         L Q A E T D Q L E D E K S A L Q T E I A N L L K E K E K L
FRA1        L Q A E T D K L E D E K S G L Q R E I E E L Q K Q K E R L
C/EBP       T Q Q K V L E L T S D N D R L R K R V E Q L S R E L D T L
CREB        L E N R V A V L E N Q N K T L I E E L K A L K D L Y C H K
ATF-1       L E N R V A V L E N Q N K T L I E E L K T L K D L Y S N K
```

Fig.1.6. Leucine zipper amino acid sequences.

The structure of the leucine zipper region was refined in a number of later studies. Experiments by Gentz *et al*, who employed mutagenesis to study both the leucine zipper and the basic region of the oncoproteins Fos and Jun, indicated that the orientation of the α-helices in the Fos-Jun heterodimer was in fact parallel[37]. Study of the leucine zipper of GCN4 using CD also suggested that the leucine zipper dimer formed a coiled-coil of parallel α-helices[38]. Both studies presented evidence which suggested

that it was unlikely that the leucine zipper was formed via interdigitation of leucine residues (Fig.1.7).
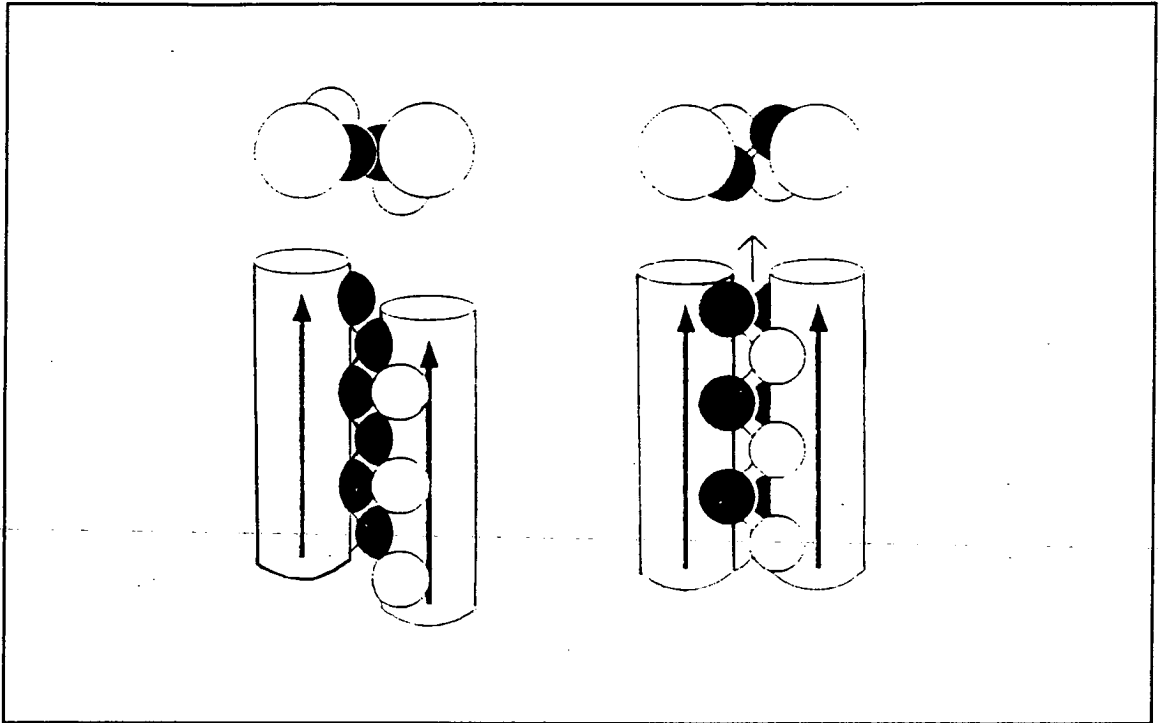
Fig.1.7. Interdigitation versus knobs into holes association.

It was noted in both studies that the association of α-helices in a typical coiled-coil was not the result of interdigitation, but via a "knobs into holes" association created by hydrophobic interactions occurring at the 4, 3 positions of each helix. This model was first proposed by Crick who realised that the structure of a coiled-coil could be stabilised by the packing of the hydrophobic side-chains of one helix into the spaces created between the side-chains of the neighbouring helix[36]. The pattern of residues in a coiled-coil may be described as $(\mathbf{abcdefg})_n$ whereby residues at positions **a** and **d** fall on the same face of an α-helix. Oppositely charged residues often occur at positions **e** and **g** of adjacent heptads allowing the formation of interhelical ion pairs. Comparison of the sequences of GCN4, Jun and C/EBP showed that the leucine zipper

region contained hydrophobic residues spaced every four and then three residues apart which was consistent with this model (Fig.1.6).
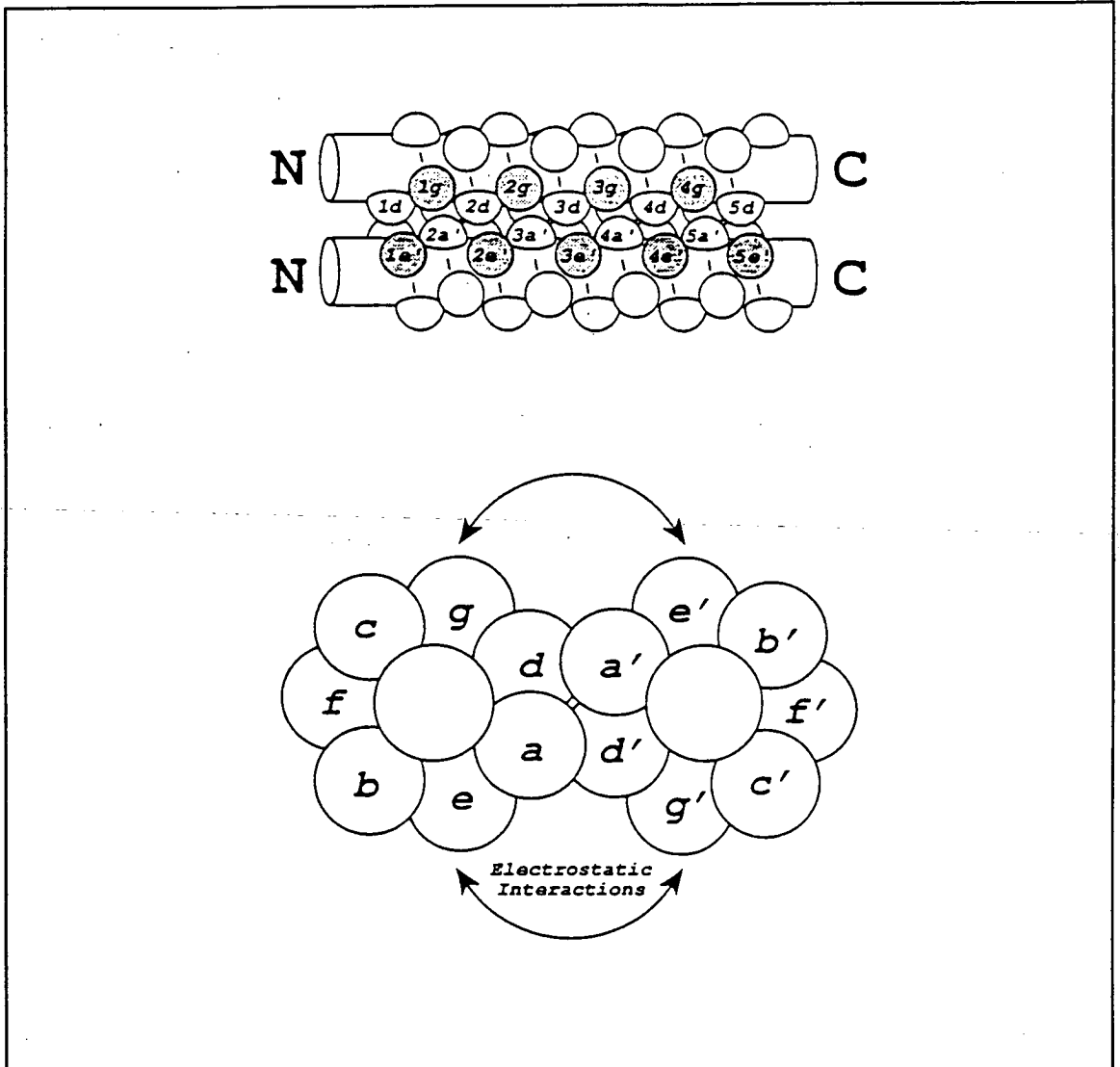
Fig.1.8. Coiled-coil packing interactions.

Physical experiments involving the use of NMR indicated that the leucine zipper region of GCN4 was α-helical[39]. The symmetrical nature of the spectra obtained suggested that the structure of the leucine zipper was that of a parallel coiled-coil. Exchange data suggested that certain residues such as val-257, leu-260, leu-261, ala-272, leu-274, lys-275, lys-

276, leu-277, val-278, and glu-279 were shielded from solvent indicating that these residues formed the hydrophobic core of the coiled-coil.

The structure of the leucine zipper region was finally elucidated with the publication of the X-ray structure of the GCN4 leucine zipper[40] (Fig.1.8). The study, whose results had been suggested by previous experiments, shows that the leucine zipper of GCN4 does in fact form a parallel coiled-coil[41]. The structure shows that the conserved leucines do not interdigitate as proposed in the original leucine zipper model. Instead it can be seen that the leucines and the amino acids in position **a** are surrounded by four residues from the neighbouring helix. Each leucine at position **d** packs against both the symmetry-related leucine (**d'**) and the side chain of the following residue (**g'**). These two types of layer alternate through the structure to form the hydrophobic interface between each α-helix.

The presence of interhelical ion pairs is suggested by distance calculations which indicate that salt bridge formation can occur between residues at position **g** of one heptad and position **e'** of the following heptad in the neighbouring helix.
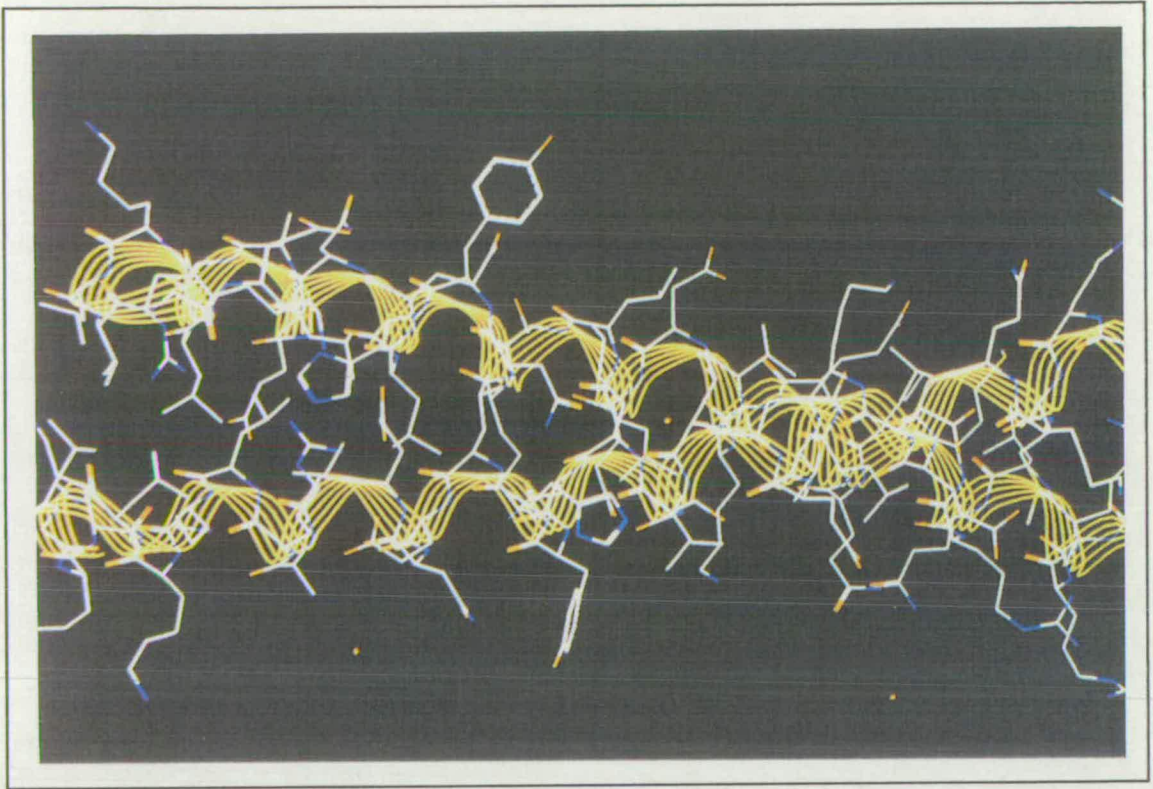
Fig.1.9. The GCN4 leucine zipper viewed along the coiled-coil axis.
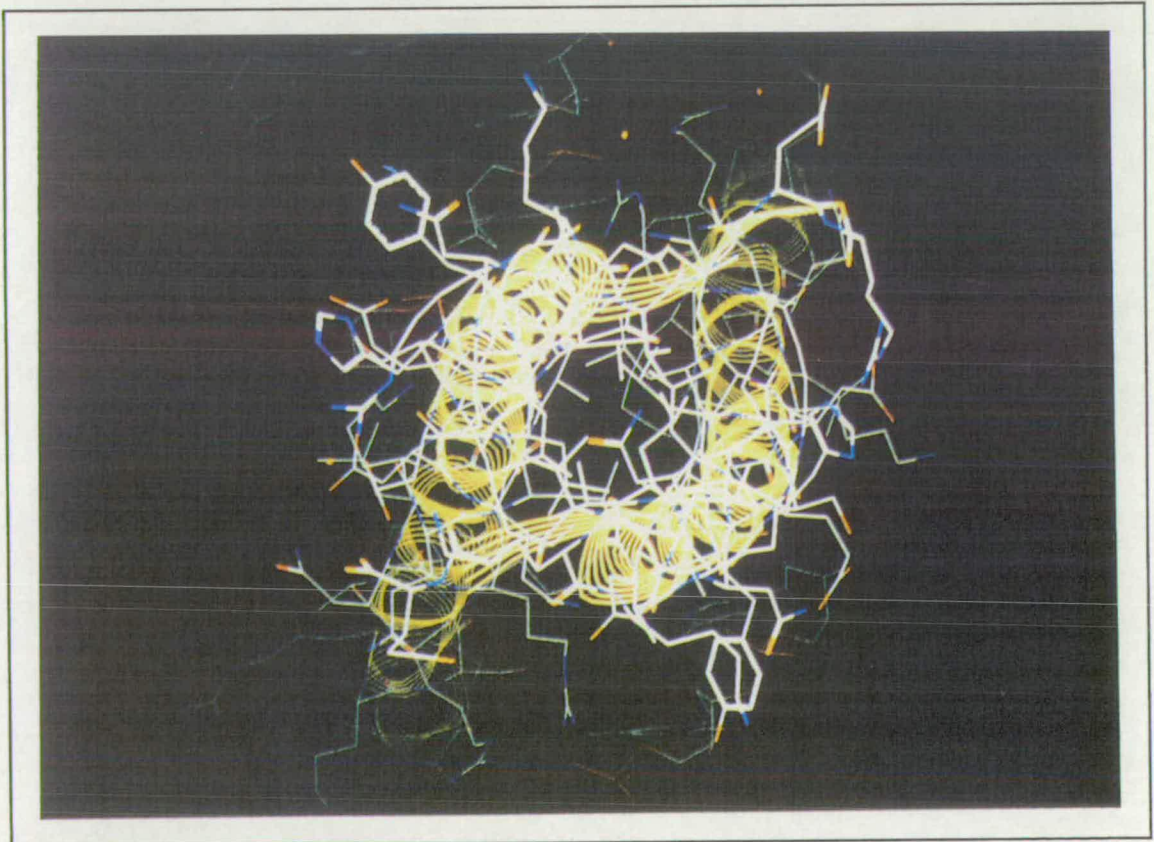
Fig.1.10. The GCN4 leucine zipper viewed down the coiled-coil axis.

1.6.3 Dimerisation specificity.

Comparison of the leucine zipper sequences of various bZip proteins shows that there are number of common features (Fig.1.6). A variety of hydrophobic residues are found at positions **a** and **d**. Closer inspection reveals that the most favoured residue at position **d** is leucine, while β-branched amino acids often occur at the **a** position. Polar residues are usually found at the **b**, **c**, **e**, **f**, and **g** positions. Asparagine in the central heptad is common, while residues at positions **e** and **g** are often oppositely charged. This configuration of amino acids gives each heptad a characterstic $(H.P.P.H.P.P.P)_n$ pattern, where H represents a hydrophobic residue and P a polar residue. The helices supercoil in a manner which places the hydrophobic residues on one side of the α-helix.
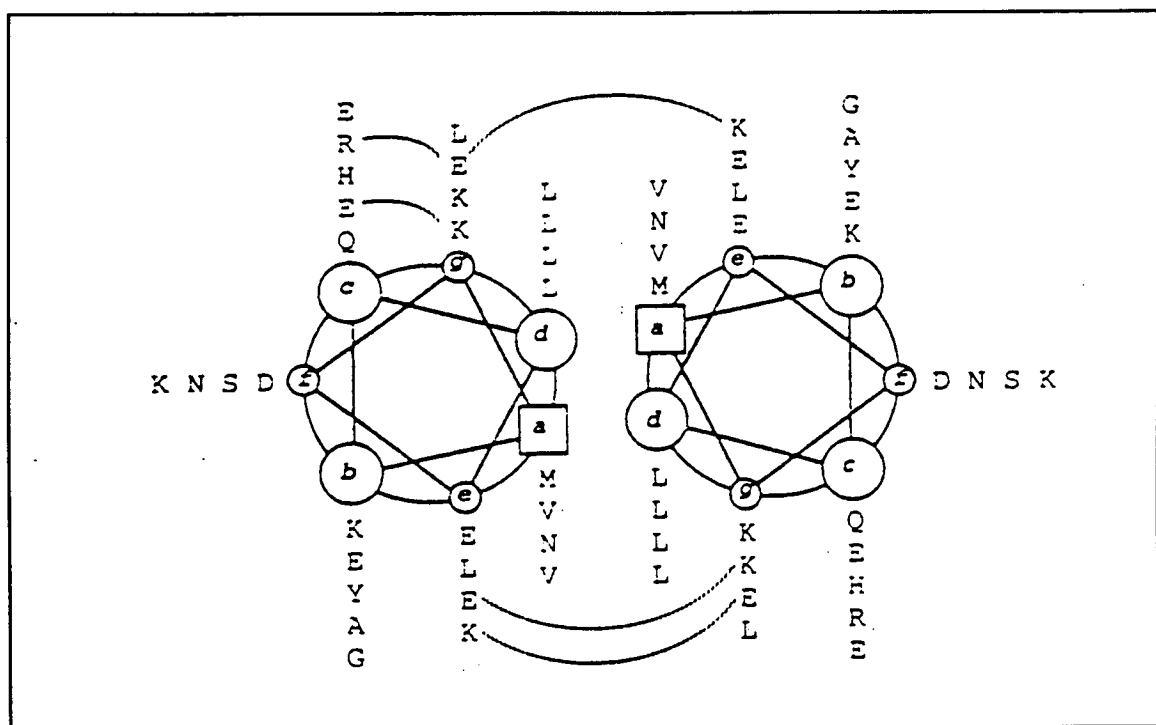
Fig.1.11. The GCN4 dimer helical wheel.

Single substitutions of residues at positions **a** and **d** can be made which do not alter the dimerisation specificity of the GCN4 leucine zipper[42,43]. It is possible to incorporate a wide variety of amino acids, including basic and acidic residues, at the **d** positions. However, substitution of two or more leucines at the **d** positions results in reduction or elimination of dimerisation[42,43]. The presence of leucine at postion **d** appears to maximise the stability of the bZip dimer in some way.

Although there is a marked conservation between the leucine zipper sequences within the bZip family, these proteins may associate as homodimers or heterodimers. Evidence suggests that residues **e** and **g** of the heptad repeats are of particular importance in influencing the dimerisation properties of certain bZip coiled-coils[44,45].

Perhaps the most well characterised of the heterodimeric bZip proteins is the Fos-Jun heterodimer (Fig.1.12). The Fos leucine zipper is very acidic, with a preponderance of acidic residues at positions **e**, **g**, and **b**. The leucine zipper of Jun, on the other hand, has a net positive charge due to the presence of basic amino acids at positions **e** and **g**. Studies have shown that the amino acids at these positions appear to influence whether each of these proteins will homodimerise or heterodimerise[46].

Experiments involving peptides have demonstrated that hybrid peptides containing the residues of Fos at the **e** and **g** positions are unable to associate, while the corresponding peptide containing Jun residues at these positions homodimerises and also heterodimerises with the Fos hybrids[44]. These results suggest that electrostatic attraction and repulsion between residues in these postions is a major determinant of the dimerisation state of Fos and Jun. Subsequent studies have demonstrated that a hybrid protein containing residues g_1 and e_2 of Fos, and a hybrid protein containing residues g_1 and e_2 of Jun is able to

heterodimerise, indicating that the electrostatic attraction between residues **g** and **e** is important in conferring heterodimerisation to these proteins[45]. Further experiments suggested that heterodimer formation was predominantly driven by destabilisation of Fos homodimers as a result of unfavourable electrostatic interactions at the **e** and **g** positions. The presence of these interhelical electrostatic interactions has been demonstrated in the X-ray structure of the Fos-Jun-DNA complex[48].
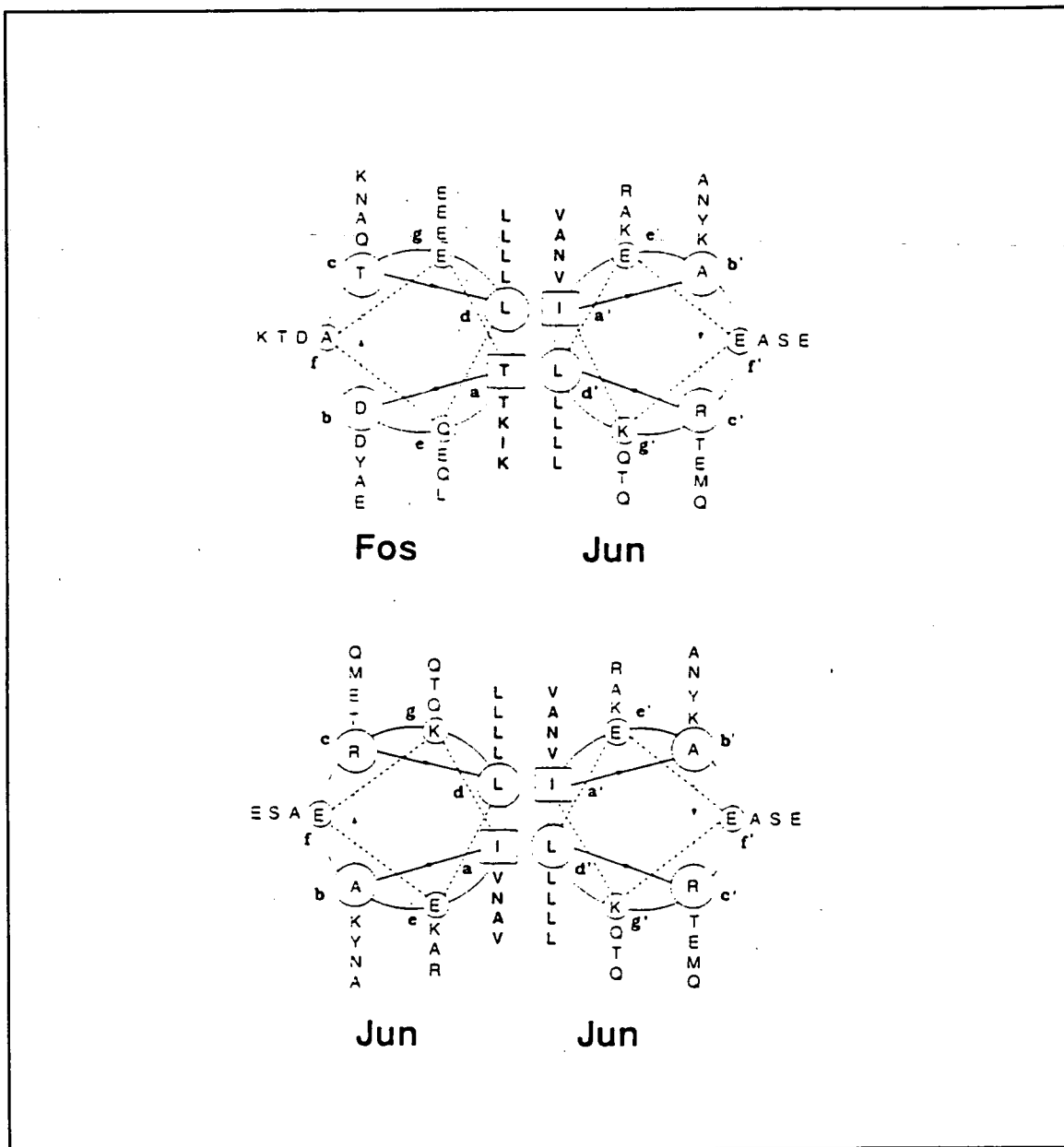
Fig.1.12. Fos-Jun and Jun-Jun helical wheels.

The tendency of salt-bridges to be formed between the **e** and **g** residues in adjacent heptad repeats was also noted by Vinson and co-workers[49]. They noted that two interhelical salt-bridges between residues in positions **e** and **g** were possible; one between the **g** of one helix and the **e'** of the other, the second between the **e** position of one helix and the **g'** of the opposing helix. These interactions were termed i+5 and i+2 salt bridges respectively (Fig.1.13). This 'salt-bridge rule' was used to design novel peptides, based on C/EBP, which would preferentially homo- or heterodimerise. Results showed that it was possible to predict how the leucine zipper regions of bZip proteins would associate simply by analysis of residues at positions **e** and **g**. It was proposed that the precise placement of these charged amino acids was of critical importance in determining whether a bZip leucine zipper would homo- or heterodimerise, or both.
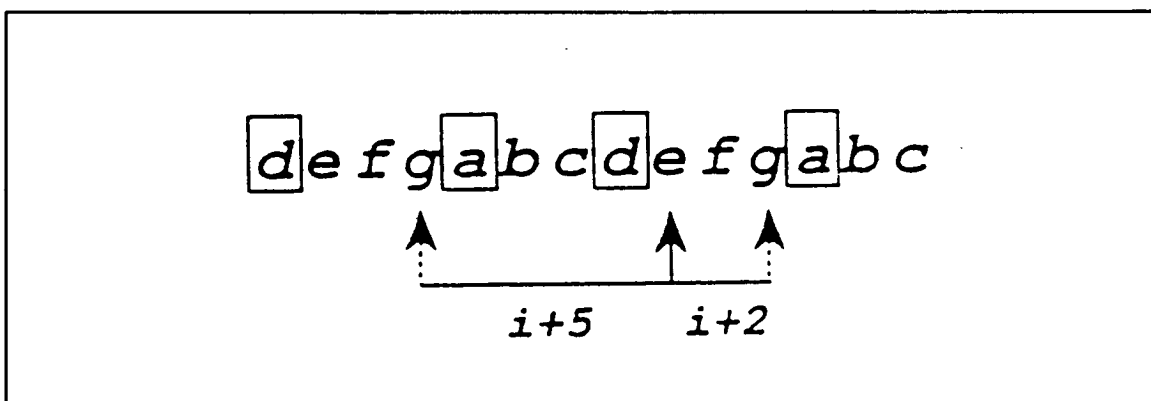
Fig.1.13. Salt bridges.

A thermodynamic scale determining the stability of the commonly occuring **e** to **g** interhelical ion-pairs was subsequently proposed[50]. The four most commonly occuring residues at the **e** and **g** positions of bZip proteins are lysine, arginine, glutamine, and glutamic acid. Thermal melting experiments suggested that the most stable helical pair occurred

between glutamic acid in the **e** position and arginine in the following **g'** position.

Complementary evidence has also been presented which suggests that the dimerisation properties of a given coiled-coil cannot be predicted solely on the basis of the interactions between **e** and **g** residues. A study employing combinatorial mutagenesis of residues at these positions in the GCN4 leucine zipper suggested that favourable charge-charge interactions are not required to form stable leucine zippers[47].

In an exhaustive study employing random selection, Pu *et al* generated a library of leucine zippers which were capable of forming homo- and heterodimers[52]. It was found that a diverse set of zippers were capable of forming homo- or heterodimers, and that charged residues were not essential for dimer stability. These results suggested that the factors governing the formation of homo- and heterodimers are complex.

The residues on the inside positions of bZip coiled-coils also affect the orientation of the α-helices and the oligomerisation state of a coiled-coil. A study investigating the orientation of the α-helices in two-stranded coiled-coils focussed on the function of residues at the **e** and **g** positions[53]. It was shown that attractive and repulsive interactions were of particular importance in determining helical orientation. As in the Fos-Jun heterodimer, attractive interactions between residues **e** and **g** seem to be of particular importance in parallel coiled-coil formation. If the attractive interactions occur between positions **g** and **g'** or **e** and **e'** then an antiparallel coiled-coil is the favoured product. Ion-pairing between **e** and **g** residues of a GCN4-lac repressor hybrid has also been shown to be a major determinant in the formation of parallel two-stranded coiled-coils[54].

Changing the **a** and **d** positions of the GCN4 leucine zipper can alter the oligomerisation state remarkably. Harbury *et al* have recently shown that replacing these residues with combinations of leucine, valine, or isoleucine results in dimeric, trimeric or tetrameric proteins[55]. A peptide containing isoleucine in the **a** positions exhibited a parallel two-stranded conformation, while a peptide containing isoleucine at all **a** and **d** positions formed parallel trimers. The X-ray structure of the isoleucine trimer showed that the three α-helical peptide monomers wrapped in a gradual left-handed superhelix with the isoleucine residues pointing into the core of the trimer[56]. The individual α-helices appeared to be virtually identical to the GCN4 dimer on a local scale.

A peptide containing leucine at all **a** positions and isoleucine at all **d** positions oligomerised as a parallel four-helical bundle. The X-ray structure showed that the four α-helices wrapped in a left-handed superhelix. As in the dimeric and trimeric structures the leucine and isoleucine side-chains at the **a** and **d** positions point into the centre of the structure.

The X-ray structures of each of the oligomers show that interhelical ion-pairs exist between the **e** and **g** residues of each helix. However, other salt-bridges are present in the tetramer between positions **g** to **b** and **c** to **e**. Each of the X-ray structures also shows that different 'knobs-into-holes' packing interactions occur. This pattern is reflected in the amino acid preference of each structure. In dimers, the **a** positions are populated with β-branched residues, whereas β-branched residues are disfavoured at the **d** positions. The presence of β-branched residues at both **a** and **d** positions favours trimer formation, while β-branched residues at the **a** positions disfavours tetramers.

1.6.4 The DNA-binding region.

In the absence of an X-ray or detailed NMR structure the mode of binding of bZip proteins to DNA was predicted using computer modelling techniques. The first model to propose a structure for the bZip-DNA complex was described by Vinson and co-workers, who termed their model the "scissors-grip" model[57]. The primary feature of this model, which was based on the protein C/EBP, was the presence of a coiled-coil in the leucine zipper region which allowed the dimer to align itself at the centre of the dyad axis of a palindromic DNA target. This left the basic region of each monomer free to track along the major groove of the B-form DNA generating base-specific contacts.

Sequence information was used to predict that the basic region formed an α-helix. However, it was suggested that the basic region would only become α-helical upon association with specific DNA.

The key feature of this model was the proposal that the basic region did not form a continuous α-helix when bound to DNA. Instead, it was suggested that a kink was formed at the invariant asparagine, via a structure known as an N-cap, which bent the α-helices towards the DNA allowing additional contacts[58]. Specific amino acid side-chain contacts with DNA were not predicted. However, the idea was suggested that the positively charged side-chains of the conserved basic residues could be arranged in a manner which would allow complementary interaction with the negatively charged phosphodiester backbone.

A second, more detailed, model which used footprinting and CD experiments as well as previous experimental evidence was proposed by O'Neil *et al*[26]. The model, known as the "induced-fork", was based on GCN4 and was in general agreement with the structure of the complex

proposed by Vinson *et al.* The model proposed that the leucine zipper of GCN4 also formed a coiled-coil structure which positioned the protein on the centre of the DNA recognition site allowing the arms of the basic region to make sequence specific contacts with the DNA bases in the major groove (Fig.1.16).

CD experiments were performed which suggested that the helicity of the basic region increased in the presence of specific DNA. The results of these experiments were used to predict that, unlike in the scissors-grip model, the basic region of the protein was entirely α-helical with no kinks. This view directly conflicted with the structure predicted previously and suggested that the region of the basic region involved in DNA contact was considerably smaller than in the scissors-grip model.

```
GCN4    D P A A L K R A R N T E A A R R S R A R K L Q R M K Q L
ATF-1   L K R E I R L M K N R E A A R E C R R K K K E Y V K C L
ATF-2   E K R R K V L E R N R A A A S R C R Q K R K V W V Q S L
C/EBP   N E Y R V R R E R N N I A V R K S R D K A K Q R N V E T
CREB    R K R E V R L M K N R E A A R E C R R K K K E Y V K C L
CREM    R K R E L R L M K N R E A A R E C R R K K K E Y V K C L
cFos    E K R R I R R E R N K M A A A K C R N R R R E L T D T L
cJun    I K A E R K R M R N R I A A S K C R K R K L E R I A R L
JunB    I K V E R K R L R N R L A A T K C R K R K L E R I A R L
```

Fig.1.15. Basic region amino acid sequences.

Perhaps the most striking feature of the induced-fork model was the prediction of the amino acids involved in DNA contact. Comparison of the amino acid sequences of a variety of bZip proteins showed that four residues at the centre of the basic region were conserved throughout the bZip family. The authors postulated that the short side-chains of asn-18, ala-15, ala-14 and cys-11 could directly interact with the DNA bases of

the recognition site. The positively charged residues flanking this quartet were proposed to add additional stability to the complex via interactions with the phosphodiester backbone. These predictions were later substantiated with the publication of the X-ray structures of the GCN4-DNA complexes and the Fos-Jun-AP-1 complex[48,59,60].
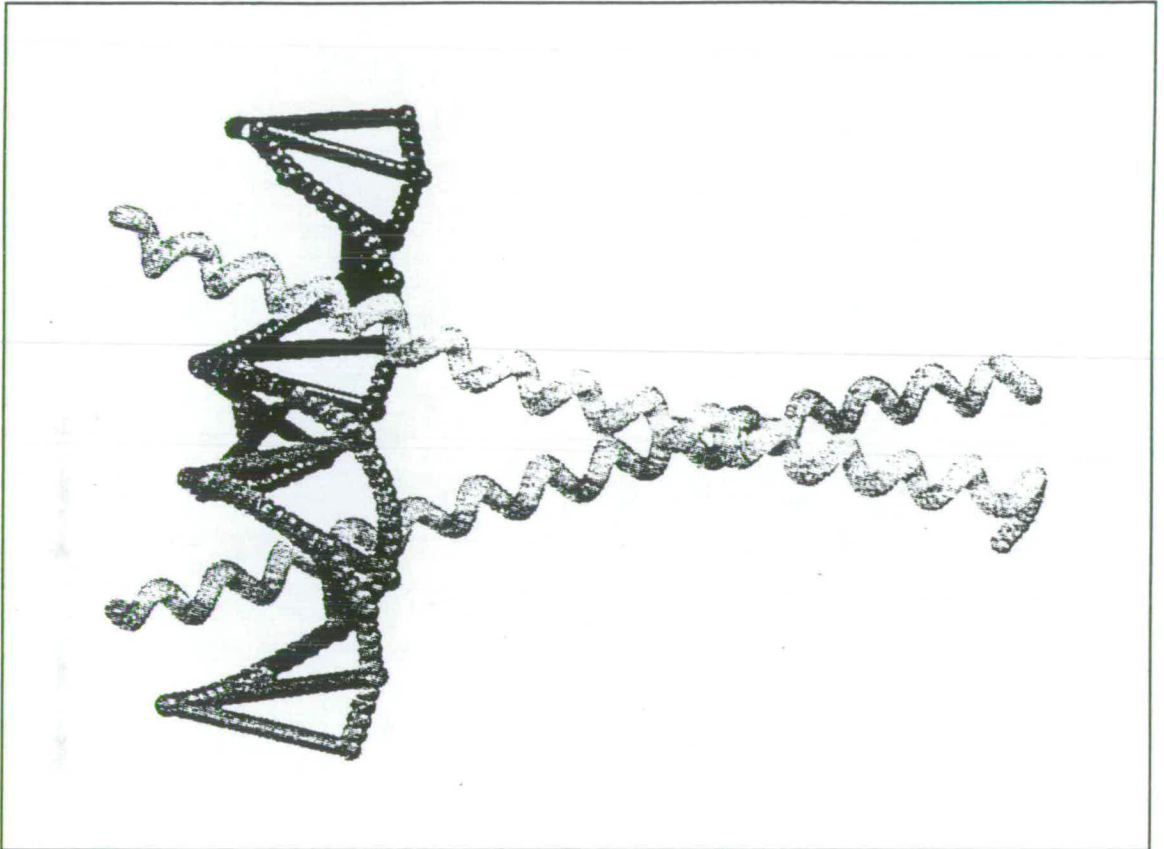
Fig.1.16. The induced fork model.

Subsequent studies involving biological and physical techniques have added credence to both of the above models. Experiments performed by Oakley and co-workers confirmed that the basic region α-helices of peptides based on GCN4 were located in the major groove of DNA[61]. This result, which was acheived using the technique of affinity cleavage, agreed with the evidence presented in both of the models. Similar results obtained using alkylation interference also suggested that the arms of the

basic region were located in the major groove of DNA[62]. The pattern of interference demonstrated that the dimeric protein contacted the DNA target with two-fold symmetry.

The helicity of the basic region was investigated in a number of studies. An ingenious study exploiting the preference of trypsin for unstructured protein showed that digestion of the C/EBP-DNA complex resulted in two fragments[63]. It was postulated that the basic region of C/EBP bound to the DNA in a manner that prevented proteolysis. This result was consistent with α-helical content suggesting that the basic region was entirely α-helical with no kinks.

Physical evidence complementing these results was presented by Saudek *et al* who used NMR and CD to demonstrate that the basic region of GCN4 was likely to be entirely α-helical[64].

Further experiments involving NMR and CD suggested that a considerable conformational change occurred when GCN4 bound to DNA[65]. In the absence of DNA it was shown that the helical content of the peptide is approximately 70% which corresponds to 40-43 residues. NMR evidence suggested that most of this could be attributed to the leucine zipper region of the peptide. Upon addition of target DNA the helicity of the peptide was shown to be 95-100% indicating that the basic region of the peptide undergoes a considerable conformational change. A study of the DNA-binding activity of Fos and Jun showed than these proteins also undergo a conformational change upon addition of target DNA indicating that the flexibility of the basic region is probably universal throughout the bZip family[66]. Similar results were presented by O'Neil *et al* who postulated a series of conformational changes which accompanied DNA binding[67]. In the absence of target DNA it was proposed that the basic regions of bZip proteins could adopt flexible non-

helical conformations that minimise any repulsive intramolecular electrostatic interactions between their positively charged side-chains. Addition of DNA could result in an electrostatic attraction between the protein and the DNA allowing a non-specific interaction between the basic region and the phosphodiester backbone. The strongly negative electrostatic potential created by the presence of the DNA could then induce α-helical conformation in the basic region which could be stabilised further by interactions with bases and the phosphodiester backbone.

1.6.5 The DNA.

bZip proteins bind to a wide variety of DNA sequences. However, despite this diversity, a number of common features exist between each DNA target. The dimeric nature of bZip proteins imposes a number of constraints upon the DNA recognition site. Each basic region tracks along the major groove in opposite directions, therefore binding sequences tend to be palindromic. These dyad-symmetric binding sites may be directly abutted, as in the CREB/ATF site, or overlapping, as in the AP-1 site[68,69]. The area of each basic region contacting the DNA is relatively small, therefore the target sequence is typically in the region of 8 to 10 base-pairs. DNA sequences outside this target region may also play a small part in the binding of bZip proteins, however, mutation or deletion of base-pairs within this region can affect binding markedly[70,71,72].

The local conformation of unbound DNA is sequence dependent and is determined by the preferred stacking interactions between adjacent base pairs[73]. Although the predominant forms of DNA are A- and B-form, DNA can adopt conformations which are intermediate between these. These

intermediate forms may have an enlarged major groove which is more accessible to the recognition helices of DNA-binding proteins[74]. The binding of proteins in the major groove may also be enhanced by bending of the DNA which may incline the base pairs towards amino acid side-chains involved in making sequence specific contacts[73].

The AP1-site.

```
+4 +3 +2 +1  0 -1 -2 -3 -4
A  T  G  A  C  T  C  A  T
T  A  C  T  G  A  G  T  A
-4 -3 -2 -1  0 +1 +2 +3 +4
```

The CREB/ATF site.

```
+4 +3 +2 +1  0  0 -1 -2 -3 -4
A  T  G  A  C  G  T  C  A  T
T  A  C  T  G  C  A  G  T  A
-4 -3 -2 -1  0  0 +1 +2 +3 +4
```

Fig.1.17. The AP-1 and CREB/ATF DNA sequences.

The AP-1 site is a target sequence for a number of bZip proteins[69]. In the unbound state this sequence adopts a straight B-form conformation[75]. In the bound state, however, the AP-1 site exhibits a variety of conformations which are stabilised by the bound protein[76]. The binding of GCN4, for example, does not alter the conformation whereas the binding of the Jun homodimer and the Fos-Jun heterodimer bends the DNA significantly[59,77]. Phasing analysis has shown that the Jun homodimer bends the DNA towards the major groove, while the Fos-Jun heterodimer bends the DNA towards the minor groove, away from the leucine

zipper[78]. An exhaustive study of DNA-bending by Fos and Jun related proteins demonstrated that the AP-1 site can adopt a variety of conformations upon association with a given dimer[76]. They demonstrated that the degree and direction of the protein-induced DNA-bending was dependent upon the basic region interacting with each DNA half-site.

Protein-induced bending by bZip proteins has also been shown to occur at the CREB/ATF site[79]. In the absence of protein, the CREB/ATF site is bent at angle of 11° towards the major groove[75]. However, upon association with the bZip protein CRE-BP1 the DNA displays no curvature indicating that the DNA is bent towards the minor groove. Binding of GCN4 also appears to induce a very slight curvature of this sequence, this time towards the major groove[75].

The sequence of the basic region appears to be important in influencing the degree of DNA-bending by bZip proteins. Comparison of bZip proteins reveals that there is a correlation between basic region sequence and the direction of protein-induced bending[75]. Residues at the very N-terminus of the basic region appear to be critical in this respect. Proteins inducing a minor groove bend contain three basic residues in this area of the basic region, whereas proteins that induce a major groove bend contain a pattern of charged and hydrophobic residues in these positions.

Sequence specific binding of bZip proteins to their DNA targets is therefore dependent upon a number of factors. DNA bending influences the binding specificity and it may be stabilised or induced upon association with a protein. This conformational change creates the correct spatial relationship for interaction between amino acid side-chains and the DNA bases and phosphodiester backbone. This association is dependent upon both the flexibility of the DNA and the protein, as well as the sequence of the basic region.

1.7 GCN4.

1.7.1 Background.

One of the most exhaustively studied members of the bZip family is the yeast transcriptional activator GCN4. GCN4 is a positive regulator of 30-40 unlinked genes encoding enzymes of amino acid biosynthetic pathways in *S. cerevisiae*[80]. GCN4 protein is synthesised only during conditions of amino acid deprivation via a novel translational control mechanism. The mechanism is complex and is thought to be mediated by 4 short open reading frames in the leader of GCN4 mRNA[81]. Amino acid biosynthetic enzymes are therefore only induced where cells are starved for amino acids. Studies have shown that GCN4 stimulates transcription by binding to short nucleotide sequences present upstream from each gene under its control.

The sequence of the GCN4 gene was published independently by two groups in 1984[82,83]. The amino acid sequence derived from this shows that the entire protein is 281 residues in length and that the carboxy-terminal end of the protein, which is essential for regulatory function, is rich in basic residues.

Analysis of truncated GCN4 proteins defined the regions of GCN4 responsible for DNA-binding and transcriptional activation[84] (Fig.1.18). A series of N-terminally deleted proteins were generated and assayed for their ability to bind to the *his3* DNA sequence[84]. Results demonstrated that the DNA-binding domain of GCN4 was contained entirely within the 60 carboxy-terminal amino acids. Further experiments with C-terminally deleted Lex-GCN4 fusion proteins showed that a fully functional GCN4 transcription region was located within the 125 N-terminal amino acids

of the native protein[84]. A 19 amino acid segment (107 to 125) within this acidic region was identified which was critical for transcriptional activation *in vivo*.
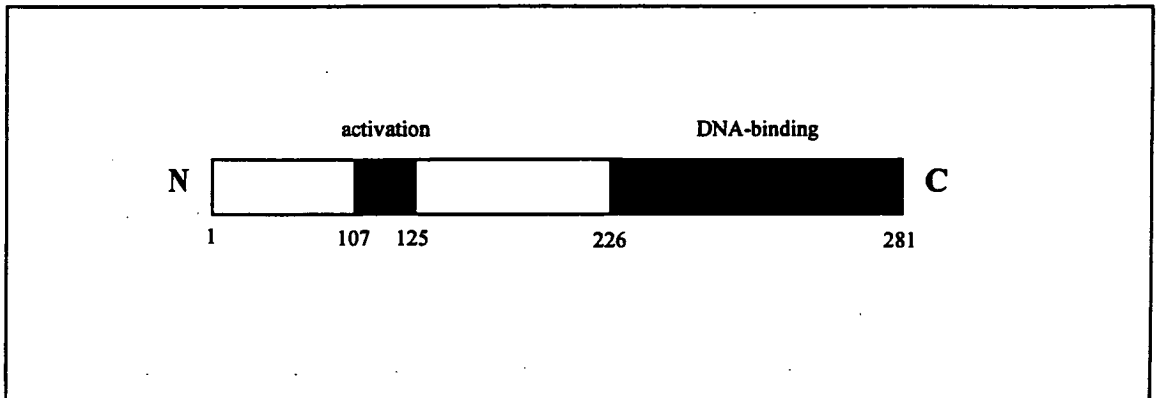
Fig.1.18. Schematic representation of the GCN4 protein.

The exact structure and function of the activation region is not known. Deletion studies have suggested that this region probably has no defined structure[85]. It is thought that the function of the activation region is to interact with one or more of the components of the preinitiation complex bound to the TATA box. Indeed, evidence has been presented which indicates that GCN4 is able to bind to Pol II *in vitro*[86]. However, these results indicated that the DNA-binding region interacts independently with Pol II suggesting that the activation region is unlikely to be involved in this interaction. The activation region may, however, contact another component of the transcriptional machinery.

The DNA-binding region of GCN4 has been studied intensively and it is the most well characterised of all the bZip proteins. The bZip region consists of residues 226 to 281. This region is able to fold independently from the rest of the protein and bind DNA. Like all bZip proteins this region consists of two distinct segments, a leucine zipper and a basic region, which perform separate functions[31,34,87]. The leucine zipper

region, which consists of four heptad repeats, forms a homodimeric coiled-coil and positions the bZip region correctly to bind DNA[31] (Fig.1.19). The region between the leucine zipper and the basic region is of critical importance for DNA binding. Deletion or addition of amino acids which distort the helical register abolish DNA binding[88]. Mutations in this region also alter DNA recognition markedly[89]. The basic region of GCN4 is involved in DNA recognition and numerous residues directly contact the DNA bases and phosphate backbone[59,60,90]. Mutation of amino acids in this region can result in proteins with altered specificity.

```
        a  b  c  d  e  f  g

          226-    D  P  A

        A  L  K  R  A  R  N    BASIC

        T  E  A  A  R  R  S    REGION

        R  A  R  K  L  Q  R


        M  K  Q  L  E  D  K

        V  E  E  L  L  S  K    LEUCINE

        N  Y  H  L  E  N  E    ZIPPER

        V  A  R  L  K  K  L

        V  G  E  R  -281
```

Fig.1.19. The GCN4 bZip region amino acid sequence.

Various studies have defined the sequences of the DNA targets bound by the bZip region of GCN4. Initial studies, carried out *in vitro*, identified that many yeast biosynthetic genes containing the core sequence TGACTC were bound by GCN4[91]. This study reinforced previous evidence

which had shown that GCN4 bound with high affinity to the *his3* site containing the palindromic sequence TGACTCA[92]. The optimal sequence for the binding of GCN4 was later defined as ATGACTCAT[93] (Fig.1.17). This palindromic sequence is known as the AP-1 site as it is also recognised by proteins of the AP-1 family of transcription factors[69]. Later studies revealed that GCN4 has almost equal affinity for the fully symmetric CREB/ATF site, which contains an additional G.C base pair at the centre[62] (Fig.1.17). GCN4 is unable, however, to bind to DNA targets containing a deletion at the central C.G base pair[92].

1.7.2 The GCN4 bZip region-DNA complex.

The crystal structures of the GCN4 bZip region bound to both the AP-1 site and the ATF/CREB site have recently been published[59,60]. The structures of both complexes are very similar and confirm much of the experimental evidence mentioned previously.

Each structure shows that the dimeric protein is entirely α-helical and that the leucine zipper is oriented perpendicular to the axis of the DNA (Fig.1.20). Comparison of each leucine zipper coiled-coil reveals that they are essentially superimposable and virtually identical to the X-ray structure of the isolated GCN4 leucine zipper[40]. Charged residues appearing at the e and g postions form only a small number of salt-bridges, indicating that the stability of the dimer results mainly from interactions between residues at the core of the coiled-coil.
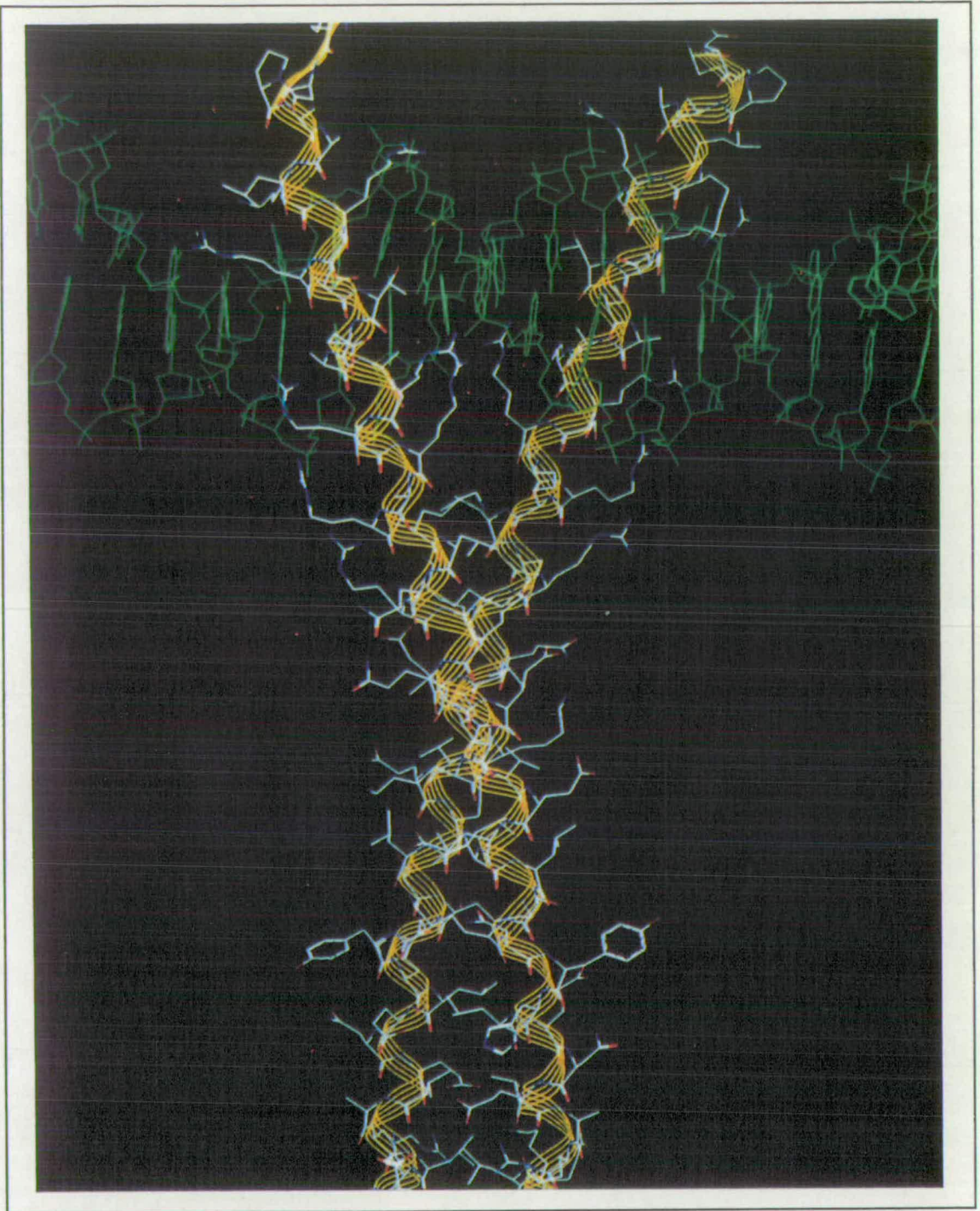
Fig.1.20. The GCN4-AP-1 DNA complex.

Note: The X-ray structure coordinates upon which the above and following pictures are based were kindly provided by T. Ellenberger.
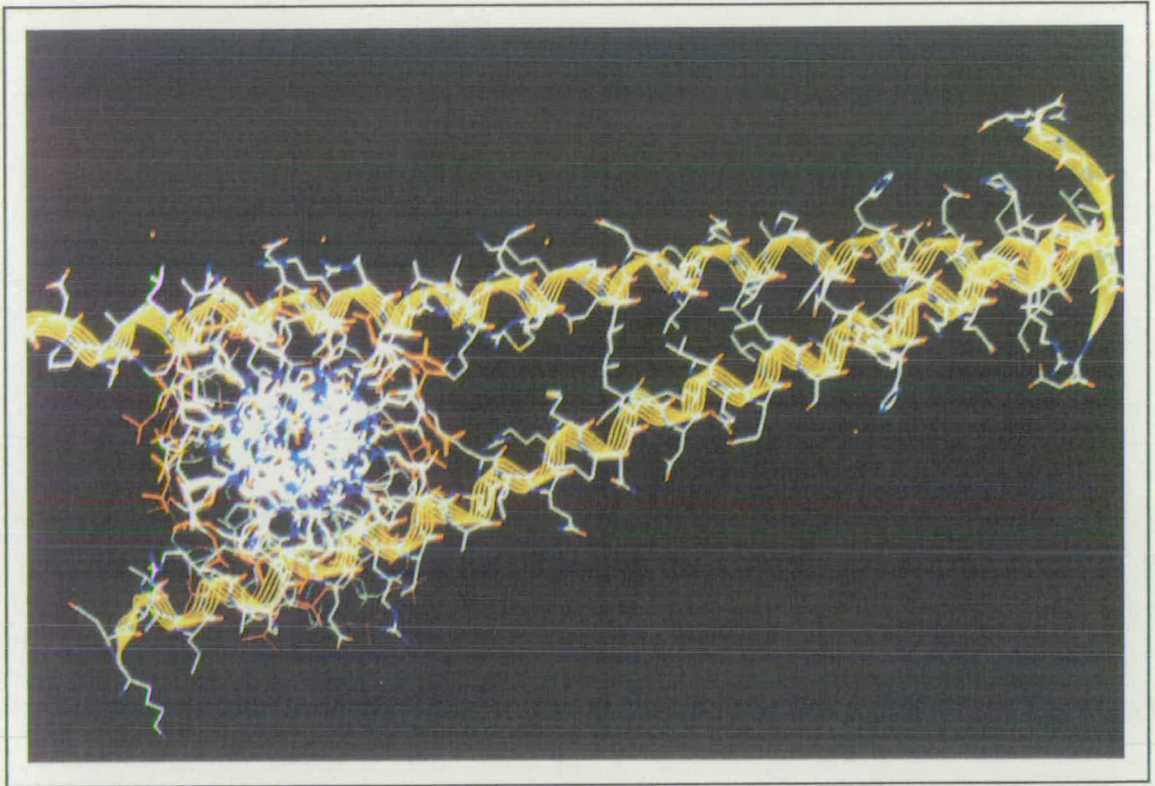
Fig.1.21. The GCN4-AP-1 DNA complex viewed down the DNA.

Fig.1.22. Amino acid-base contacts.
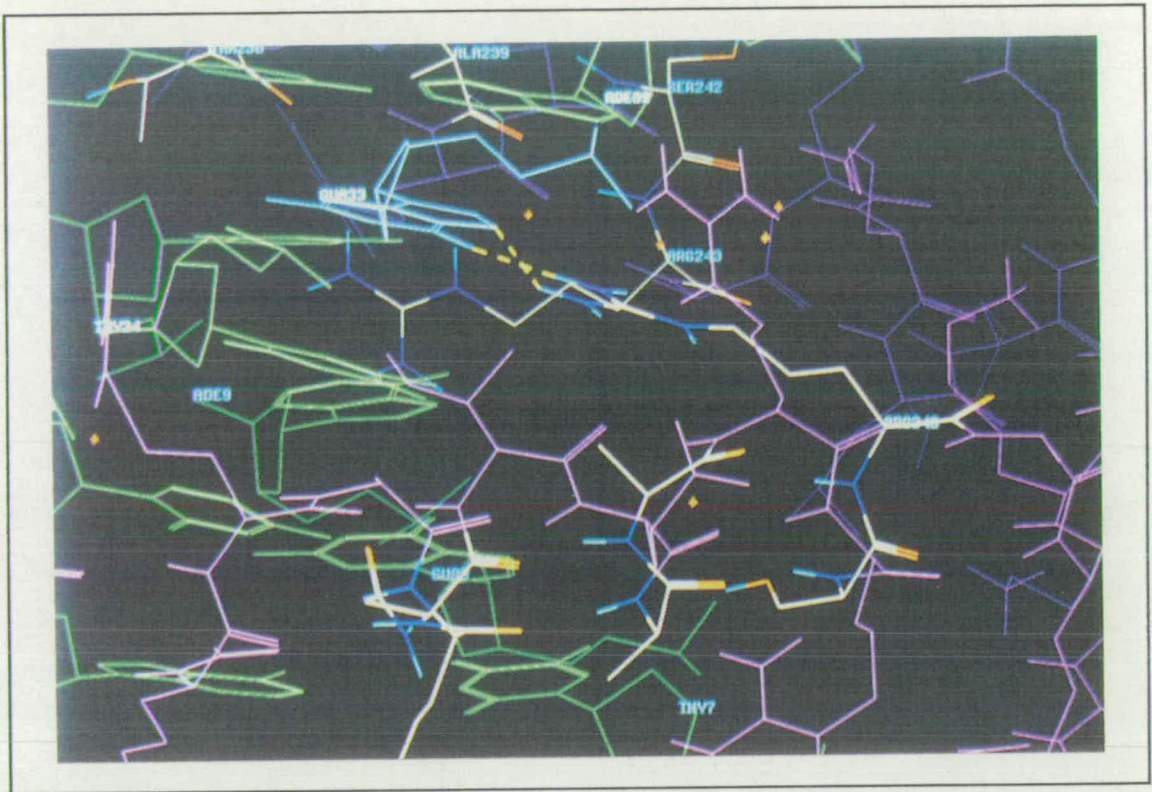
Fig.1.23. The arg-243-DNA interaction.
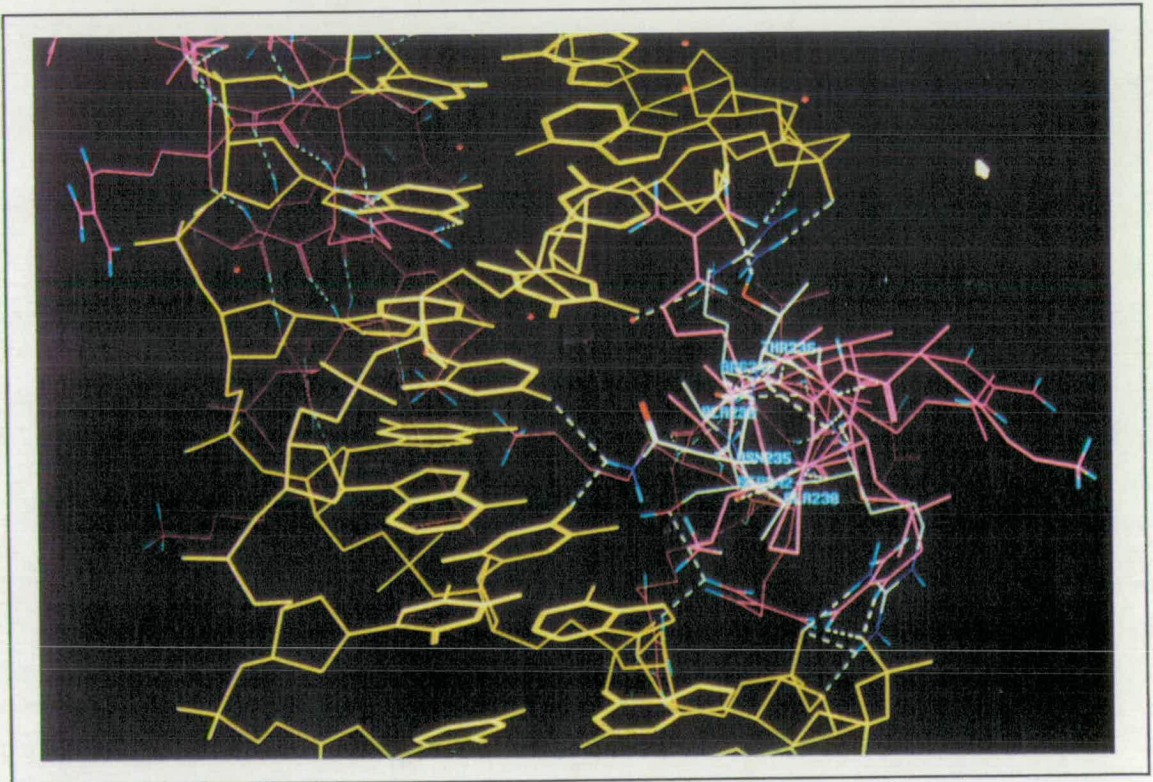
Fig.1.24. Amino acid-base interactions viewed down the basic region.

The main differences between each complex concern the basic region and the DNA. In both complexes the basic region of each bZip monomer contacts a half-site of the DNA target sequence. The majority of contacts made between amino acid side-chains and DNA are conserved between each structure, however, a number of differences are apparent.

In the GCN4-AP-1 complex the basic region contacts the DNA in the segment comprising residues 232 to 249 (Fig.1.22). Contacts to DNA bases are made by a number of residues. Asn-235, which is conserved throughout the bZip family, accepts a hydrogen bond from the N4 of C(-2) while donating a hydrogen bond to O4 of T(+3). Additional contact appears to be mediated through a water molecule which allows interaction with the N6 and N7 of A(+4). DNA base contacts are also made by ala-238 and ala-239 which are in van der Waals contact with the 5-methyl groups of T(+3) and T(-1) respectively. Ser-242, which contacts the phosphate backbone, is also in hydrophobic contact with the 5-methyl group of T(+3). The asymmetry of the AP-1 site causes GCN4 to contact the central base pair in an asymmetric manner. Arg-243 of the left half-site monomer donates hydrogen bonds to O6 and N7 of G(0) (Fig.1.23). However, arg-243 of the right half-site monomer is unable to interact with G(0) and donates a hydrogen bond to the unesterified phosphate oxygens of C(0) and A(+1) instead. All other contacts positioning the basic region on the AP-1 DNA are made by basic and polar residues which donate hydrogen bonds to the phosphate oxygens of the DNA backbone.

The interactions between the basic region of GCN4 and the ATF/CREB DNA are remarkably similar to those observed between GCN4 and the AP-1 site. Contacts between protein and DNA are observed in the region between lys-231 and lys-246. Hydrophobic interactions equivalent to those observed in the AP-1 complex are made between ala-238, ala-239

and ser-242 and the ATF/CREB DNA bases. An additional hydrophobic contact, not observed in the AP-1 complex, is also made between the Cγ atom of thr-236 and the C2 and C3 atoms of the G(0) deoxyribose moeity. The contact between asn-235 and T(+3) and C(-2) is equivalent to that seen in the AP-1 site. However, no water mediated contact is observed in the ATF/CREB complex. The interaction between arg-243 and the ATF/CREB DNA is markedly different to that observed in the AP-1 complex. The symmetrical nature of the ATF/CREB site allows arg-243 of each monomer to hydrogen bond with the N7 atom of G(0) and to a backbone phosphate. The remaining contacts between the basic region and DNA are mediated via interaction between amino acid side-chains and phosphate oxygens of the DNA backbone. The pattern of interactions is essentially similar to those observed in the GCN4-AP-1 complex.

The DNA conformations are different in each complex. The conformation of the DNA in the GCN4-AP-1 complex is straight, B-form DNA, and it is unaffected by the binding of the protein. This fact is reflected in the DNA-binding constant, which is higher than that for binding to the ATF/CREB site[72].

The conformation of ATF/CREB DNA is different to that of AP-1 DNA. In the GCN4-ATF/CREB complex the DNA is bent symmetrically by 20° towards the leucine zipper axis. In the unbound state ATF/CREB DNA is bent by approximately 11°, indicating that GCN4 induces a slight curvature in the DNA[75]. This distortion shifts the phosphate groups between A(+6) and G(+2) towards the leucine zipper axis placing them at essentially the same distance as those of the AP-1 site. Base-specific interactions are further enhanced by a slight curvature in the basic region between ala-233 and ala-239. This allows the base-specific interactions of asn-235 and ala-239 to be maintained.

The recognition of a specific DNA target by a protein clearly requires that both the protein and the DNA exhibit a degree of flexibility. In this case, GCN4 appears to exert its influence in two ways; stabilising the straight conformation of the AP-1 site and inducing a bend in the ATF/CREB DNA.

1.7.3 DNA-binding specificity.

The region of GCN4 N-terminal to the leucine zipper displays enormous flexibility. This flexibility is reflected in the diversity of sequences recognised by the protein. GCN4 displays maximum binding affinity for the AP-1 site, however, the fully symmetric ATF/CREB site is also recognised, albeit with slightly lower affinity[72]. This lower binding constant is probably a result of the distortion of the basic region and DNA required for sequence specific binding. Many other symmetrical sequences, which differ only slightly from those above are also recognised, but with much lower affinity[71,92,94].

```
        230       235       240       245       250       255
        AALKRARNTEAARRSRARKLQRMKQLEDKVEE
```
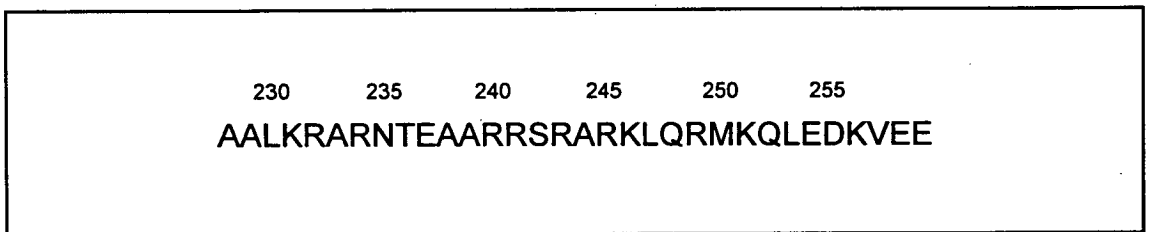
Fig.1.25. The GCN4 basic region and linker segment.

The binding specificity of the GCN4 bZip region has been explored in a number of ways including the generation of synthetic peptides, mutants and chimeric proteins.

1.7.4 Synthetic peptides.

In the wild-type protein the leucine zipper positions the basic region for sequence specific binding. However, peptides have been synthesised which possess only two basic regions joined via a disulphide link[95]. These proteins carry enough information to specifically bind to the AP-1 and ATF/CREB sites, although with a much reduced binding constant. The basic region of GCN4 can be further reduced without altering the sequence specific recognition properties of the basic region markedly[61,96]. Dipeptides consisting of only 20 residues of the basic region have been shown to bind both the ATF/CREB and AP-1 sites. Deletion experiments have suggested that the region between 231 and 245 is critical for DNA recognition[96]. Indeed the X-ray structures of the GCN4-DNA complexes show that the majority of contacts between amino acid side-chains and DNA bases occur in this region[59,60].

The region between residues 245 and 255, in which the basic region diverges from the leucine zipper, is particularly important in determining the half-site preferences of GCN4[59]. Peptides in which this region has been replaced by a bis(terpyridyl) iron II complex have been shown to bind preferentially to different DNA sites[97,98]. Complexes containing different linker molecules were shown to orientate the basic region in ways which altered its binding preferences. The results suggested that the amino acid composition of the linker segment was of critical importance for half-site recognition. This fact was reinforced by the observation that the linker domains are not conserved across bZip families, but conserved within them.

1.7.5 Single mutants.

Mutations of residues in the linker segment also alter the half-site preferences of GCN4. AP-1 proteins have an uncharged residue at position 247, whereas proteins of the ATF/CREB family have a charged residue at this position. Mutation of leu-237 of GCN4 to lys-247 or arg-247 results in proteins with comparable binding affinity to GCN4, whereas, proteins in which residues 247, 249 and 250 are replaced by those of CREB show a reduction in binding affinity for the AP-1 site such that the ATF/CREB site is preferred by a factor of two[89]. These results indicate that the composition of the linker region, which affects the angle of bifurcation of the basic region, is of importance in the selection of binding site and that the half-site preference of a protein can be changed by the mutation of residues within this region.

The basic region contacts DNA in the region spanning residues 231 and 249. Base-specific contacts are generated by asn-235, thr-236, ala-238, ala-239, ser-242 and arg-243. All other contacts occur via interactions between amino acid side-chains and the phosphodiester backbone of the DNA[59,60].

The majority of mutational studies of the GCN4 basic region have examined the effects of mutating the residues contacting the DNA bases. These studies have involved single and double point mutations, the generation of chimeric proteins and substitution of multiple residues of the GCN4 basic region with those of other bZip proteins.

Asn-235 is conserved throughout the entire bZip family. Mutation of this residue in GCN4 shows that very few substitutions are tolerated[99]. Experiments investigating the binding of mutants to the AP-1 site have shown that only one mutant, trp-235, binds with high affinity. Screening

with symmetrically mutated AP-1 sites revealed that the trp-235 mutant bound specifically to the symmetrically mutated TTGACTCAA sequence, which is recognised only weakly by GCN4[100].

Further experiments investigating the role of asn-235 revealed that a slightly larger number of mutants at this position could be generated[101]. However, the majority of these bound only weakly to the ATF/CREB site and symmetrically mutated sequences. A mutant protein containing tryptophan at position 235 was shown to bind with altered specificity to the sequence AGGACGTCCT. In contrast to the trp-235 mutant described above, which binds to a site changed at the 4 postion, this mutant recognised a binding site which was altered at the 3 position. The reasons for this apparent discrepancy are not clear. Identical contacts are made between asn-235 in the wild-type protein and both the AP-1 and ATF/CREB DNA sequences, therefore it would be reasonable to assume that similar contacts would be made between the mutant protein and its favoured DNA target[59,60]. The different sequence preferences of the trp-235 mutant may reflect changes in the half-site spacing of the ATF/CREB site and any DNA bending induced by the trp-235 mutant. These results demonstrate that the rules governing the binding specificities of mutant proteins are complex and cannot be easily predicted.

Ala-238 and ala-239 are in hydrophobic contact with T(+3) and T(-1) respectively. Identical contacts between these residues and the DNA bases are observed in both the AP-1 and ATF/CREB protein-DNA complexes[59,60]. Mutation of these residues has been carried out and the binding of the resulting mutants to both the AP-1 site and the ATF/CREB site investigated.

Studies show that relatively few mutations are tolerated at the 238 position. Substitution of serine or cysteine results in a protein which is

able to bind with high affinity to the AP-1 site[99]. However, identical mutant proteins have been shown to behave differently when incubated with DNA containing the ATF/CREB site[102]. The ser-238 protein binds with high affinity whereas the cys-238 protein binds only weakly. Mutants with extended aliphatic side-chains, such as leucine and valine, bind only weakly to both sites, indicating that mutants at this position interact with DNA bases hydrophobically and that the binding is constrained by the size of the binding pocket.

Mutation of ala-239 shows that a wide variety of mutants, which bind to the AP-1 and ATF/CREB sites, are tolerated[103]. These substitutions include basic and neutral amino acids such as glycine and valine. Studies have shown that insertion of serine at this position results in a protein which displays broadened specificity for ATF/CREB related sequences. Single substitutions also result in mutant proteins of changed specificity. An alanine to cysteine conversion results in a protein which binds to ATF/CREB sites changed at the 1 position, whereas substitution with aspargine favours a site altered at the 2 position.

A large number of mutations can be tolerated at the 242 position[102]. Mutant proteins containing cysteine, phenylalanine, histidine, leucine, methionine, glutamine, threonine, valine and tryptophan all bind to the ATF/CREB site with high affinity. A number of these mutants also display broadened specificity for symmetrically mutated ATF/CREB sites. Modelling studies suggest that hydrophobic interactions play an important part in the binding of these mutants to their targets.

Arg-243 is conserved throughout the bZip family. This fact is reflected in the small number of mutations allowed at this position. Only two weakly binding mutants, lys-243 and tyr-243, have been identified[101].

1.7.6 Double mutants.

Relatively few proteins containing double mutations in the basic region of GCN4 have been generated. The majority of studies have focussed on the amino acids involved in contacting the DNA bases.

Mutation of ala-238 and ser-242 results in a library of mutants in which very few actually bind to DNA[102]. Those that are able to bind show broadened specificity for a number of ATF/CREB related sequences. Similarly, mutation of ala-239 and ser-242 results in a large number of mutants which display broadened specificity. For example, a cys-239/trp-242 mutant, binds to four symmetrically mutated sites, whereas a ser-239/trp-242 mutant binds to nine[102].

Mutation of the conserved residues, asn-235 and arg-243, results in only a small number of weakly binding mutants[101]. However, a his-235/trp-243 mutant does show a degree of selectivity for an ATF/CREB site altered at the 2 position.

In general, the binding of double mutants follows the pattern demonstrated by the singly mutated proteins. A wide variety of sequences are recognised, with mutant proteins displaying broadened and altered specificities. As with singly mutated proteins the pattern of interactions between protein and DNA may differ significantly from that observed in the wild-type complexes. These factors make it extremely difficult to predict the binding specificity of a given protein, making the basic region of GCN4 an attractive target for random mutagenesis.

1.7.7 Chimeric proteins.

A different method of creating 'new' proteins involves the practise of generating chimeric proteins via recombinant DNA techniques. The area of the basic region responsible for specific recognition of a DNA target can often be identified in this way. Many researchers have employed this method of generating hybrid proteins to identify these recognition regions and create proteins which have changed specificities.

Suckow *et al* showed that replacement of residues 228 to 241 in the basic region of GCN4 with those of C/EBP resulted in a protein which bound to the C/EBP binding site[104]. Further analysis, via stepwise replacement of residues, revealed that the region between thr-236 and ala-239 contained enough information to distinguish between the ATF/CREB and C/EBP targets. Substitution with the corresponding residues of TAF-1 revealed that the specificity of the hybrid could be switched to that of TAF-1.
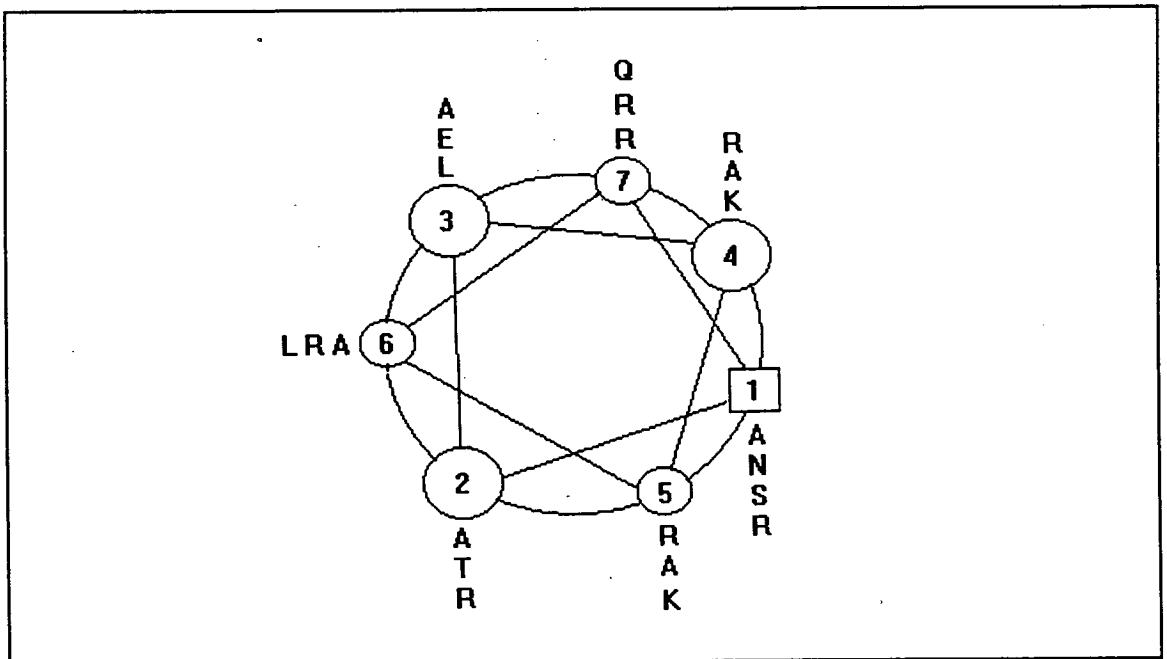
Fig.1.26. The GCN4 basic region helical wheel.

Similar studies investigating the DNA-binding specificity of GCN4-C/EBP chimeras were carried out by Johnson[105]. The experiments showed that as the fusion point of the chimeras was moved through the basic region, a transition between C/EBP and GCN4 binding specificity could be observed. Further experiments suggested that C/EBP and GCN4 use equivalent segments of the basic region to contact DNA. Substitution of thr-236 and ala-239 of GCN4 with the corresponding residues of C/EBP resulted in a protein with broadened specificity for the C/EBP site. In combination with ala-227, these mutations were shown to confer greater C/EBP-like character to the GCN4 hybrid, indicating that residues on the same face of the basic region α-helices are important in DNA recognition (Fig.1.26).

1.7.8 Future work.

As described above, various studies have addressed the effects of mutating residues in the GCN4 basic region. The majority of these studies have employed site-directed mutagenesis to randomly mutate specific residues and have assessed the binding characteristics of the resulting mutants by gel retardation assay.

As yet, a study employing random mutagenesis across a large area of the basic region has not been attempted. Previous evidence has indicated that the rules governing the binding of a bZip protein to DNA are complex and that residues outwith the conserved region may play an important part in determining binding specificity. Randomly mutating a large section of the basic region would therefore address the role of these residues and go some way towards simplifying the rules for the specific binding of a bZip protein to DNA.

CHAPTER TWO
Materials and Methods

2.1 Materials.

Sigma ammonium persulphate, ampicillin, DTT, EDTA, HEPES, IPTG, maltose, SDS, tetracycline, Tris, and most other chemicals.

Fluka acrylamide, bisacrylamide, guanidinium hydrochloride, sodium perchlorate, TEMED.

Difco bacto-agar, bacto-tryptone, yeast extract.

Gibco BRL restriction enzymes, agarose, LMP agarose, exonuclease III, ATP, caesium chloride.

Pharmacia Sephadex G-50, T4 DNA ligase, T4 polynucleotide kinase, Klenow fragment, DNA polymerase I, DNase I, dNTP's, dCTPαS, T7 sequencing kit.

Amersham [^{35}S]methionine, [α-^{35}S]dATP, [α-^{32}P]dCTP, [γ-^{32}P]ATP.

Stratagene λZAPII kit, Taq DNA polymerase.

USB T7 exonuclease.

2.2 Bacterial strains and plasmids.

The following strains of *E. coli* were used:

SOLR: e14⁻(*mcrA*), Δ(*mcr*CB-*hsd*SMR-*mrr*)171, *sbc*C, *rec*B, *rec*J, *umu*C::Tn5(kan^r), *uvr*C, *lac*, *gyr*A96, *rel*A1, *thi*-1, *end*A1, λ^R, [F', *pro*AB, *lac*I^{q}ZΔM15]. Su⁻ (non-suppressing)

XL1-Blue: *end*A1, *hsd*R17, *sup*E44, *thi*-1, *rec*A1, *gyr*A96, *rel*A1, *lac*, [F', *pro*AB, *lac*I^{q}ZΔM15, Tn10(tet^R)]

JM109:   recA1, endA1, gyrA96, thi, hsdR17, supE44, relA1, λ-,
           Δ(lac-proAB), [F', traD36, proAB, lacIqZΔM15]

TG1:   supE, hsdΔ5, thi, D(lac-proAB),
           [F', traD36, proAB+, lacIq, lacZΔM15]

pBluescript SK- was derived from λZAPII which was used to clone the
bZip and GCN4 genes[106,107].

pTZ19R is derived from pUC19 and was used to clone the OP1 DNA
fragment[108, 109, 110].

YCp88-GCN4 was obtained from Prof. K. Struhl and used to isolate the
GCN4 and bZip genes[84].

2.3 Maintenance of bacterial strains.

All strains were stored at -70ºC in 15% (v/v) glycerol. A single colony was
grown overnight at 37ºC in 3ml LB supplemented with the appropriate
antibiotics. To 850µl of this culture was added 150µl glycerol and the
culture stored at -70ºC. Strains which were used regularly were also
maintained on agar plates.

2.4 Gowth media and supplements.

LB (Luria Bertani) medium.

LB medium contained the following:

            10g   bacto-tryptone

10g	NaCl
5g	yeast extract

Water was added to a volume of 1 litre and the pH adjusted to 7.5 before autoclaving.

Bacto-agar (15g/l) was added for plates.

For the growth of XL1-Blue and SOLR cells, the medium was supplemented with magnesium sulphate to a final concentration of 10mM and maltose to a final concentration of 0.2%. The appropriate antibiotic was also added.

2xYT broth.

2xYT broth was used for the growth of XL1-Blue cells and contained the following:

16g	bacto-tryptone
10g	NaCl
10g	yeast extract

Water was added to a volume of 1 litre before autoclaving.

Top agar.

Top agar was used for the growth of bacterial plaques containing λZAPII. The media contained the following:

1g	bacto-tryptone

0.75g	bacto-agar
0.5g	NaCl

Water was added to a volume of 100ml before autoclaving.

M9 Minimal media.

Minimal media was used for the growth of TG1 cells. It contained the following:

50ml	10xM9 salts
1ml	$MgSO_4$ (1M)
5ml	20% glucose
50µl	$CaCl_2$ (1M)
200µl	thiamine (2mg/ml)

Water to a volume of 500ml was added before autoclaving. Bacto-agar (7.5g) was added for plates.

10xM9 salts contained the following:

30g	Na_2HPO_4
15g	KH_2PO_4
5g	NH_4Cl
2.5g	NaCl

Water to a volume of 500ml was added and the pH adjusted to 7.4 before autoclaving.

ampicillin	50mg/ml, filter sterilised, stored at 4°C for 14 days.
tetracycline	12.5mg/ml in ethanol, filter sterilised, stored at -20°C.
X-gal	250mg/ml in DMF, stored at -20°C.

2.5 Preparation of $CaCl_2$ competent cells.

From an overnight culture, 500µl of cells were suspended in 20ml of medium supplemented with the appropriate antibiotic. The cells were incubated at 37°C with shaking until the OD_{600} was between 0.37 and 0.5. The suspension was transferred to a 50ml Falcon tube and centrifuged at 4°C for 10 minutes at 12000g. The supernatant was removed and the cells resuspended in 10ml cold 100mM $CaCl_2$. After repeated inversion of the tube, the cells were stored on ice for 20 minutes. The cells were centrifuged as before and the supernatant removed. The cells were resuspended in 1ml cold 100mM $CaCl_2$. These $CaCl_2$ competent cells could be stored for up to 5 days at 4°C.

2.6 Transformation of $CaCl_2$ competent cells with plasmid DNA.

50µl of competent cells were placed in an Eppendorf tube. 1µl of plasmid DNA was added, before vortexing gently to mix the contents. The tube was placed in a water bath at 37°C for 30 seconds, before immediately transferring the tube to ice for 1 hour. The tube was then placed in a water bath at 42°C for 2 minutes. 800µl of medium supplemented with the appropriate antibiotic was then added and the tube incubated at 37°C for 50 minutes. The cells were centrifuged for 10 minutes at 12000g. The supernatant was removed and the cells resuspended in 100µl of medium. The tube was vortexed briefly and 50µl of the cells plated on agar plates

supplemented with the relevant antibiotic. The plates were incubated overnight at 37°C.

2.7 Small scale preparation of plasmid DNA (mini-prep).

A single colony from a plate containing the appropriate cells was suspended in 3ml of medium supplemented with the appropriate antibiotic. The bacteria were grown overnight at 37°C with shaking. 1ml aliquots of the culture were placed in Eppendorf tubes and centrifuged at 12000g for 5 minutes. The supernatant was removed and each pellet resuspended in 350μl of STET (100mM NaCl; 10mM Tris.HCl, pH 8.0; 1mM EDTA; 5% v/v Triton X-100). 25μl of freshly prepared lysozyme (10mg/ml in 10mM Tris.HCl, pH 8.0) was added and the mixture briefly vortexed. The tube was placed in a heating block at 100°C for exactly 40 seconds, before centrifuging at 12000g for 10 minutes. The supernatant was removed and 40μl of sodium acetate (3M, pH 5.5) added to the supernatant along with 420μl isopropanol. The solution was mixed by vortexing and stored at room temperature for 20 minutes. The pellet of nucleic acid was recovered by centrifugation at 12000g for 30 minutes at 4°C.

2.8 Purification of DNA using Prep-A-Gene matrix[112].

The pellet of DNA was resuspended in TE (10mM Tris.HCl, pH 8.0; 1mM EDTA) and centrifuged to remove any undissolved clumps of protein. Prep-A-Gene matrix was added to the solution. (Note: 1μl of matrix was added for every 0.2μg of DNA). Binding buffer (50mM Tris.HCl, pH 7.5; 1mM EDTA; 6M $NaClO_4$) equal to 3 times the combined volume of matrix

and DNA solution was added. The mixture was incubated at room temperature for 10 minutes with gentle mixing. The solution was centrifuged for 30 seconds before discarding the supernatant. The pellet was resuspended in 30 pellet volumes of binding buffer and vortexed briefly. The tube was centrifuged as before and the rinse with binding buffer repeated twice more. The pellet was then washed in 30 pellet volumes of wash buffer (40mM Tris.HCl, pH 7.4; 4mM EDTA; 800mM NaCl; 50% v/v ethanol) before centrifuging for 30 seconds and removing the supernatant. This step was repeated twice more, taking care to remove the last traces of wash buffer after the final wash. The bound DNA was eluted from the pellet in 2 pellet volumes of elution buffer (10mM Tris.HCl, pH 8.0; 1mM EDTA) by incubating for 5 minutes at 45°C. The tube was spun for 1 minute and the supernatant containing the purified DNA removed. A second extraction with 1 pellet volume of elution buffer was then carried out to recover the remaining DNA.

2.9 Phenol extraction of nucleic acid solutions.

The crude pellet of nucleic acids was resuspended in TE (10mM Tris.HCl, pH 8.0; 1mM EDTA). An equal volume of phenol/chloroform/isoamylalcohol (25:24:1) saturated in Tris.HCl (10mM, pH 8.0) was added and the mixture vortexed for approximately 1 minute. The phases were separated by centrifugation at 12000g for 2 minutes. The upper aqueous layer was collected, taking care not to disturb the white precipitate at the phase interface, and the process repeated until no white precipitate appeared. An equal volume of chloroform/isoamylalcohol (24:1) was added and the solutions mixed by

vortexing as above. The phases were separated by centrifugation as above and the upper aqueous layer containing the nucleic acids collected.

2.10 Ethanol precipitation of nucleic acids.

Nucleic acids were precipitated from aqueous solution by addition of sterile sodium acetate to a final concentration of 0.3M and 2 volumes of ethanol. After incubation on dry ice for 1 hour, the nucleic acid was pelleted by centrifugation for 30 minutes at 12000g and 4°C. The supernatant was removed and the pellet washed with 70% ethanol. The supernatant was removed and the pellet dried under vacuum. The pellet was then resuspended in the appropriate volume of sterile water and stored at -20°C.

2.11 Preparation of plasmid DNA from a large volume culture.

Late in the evening 2 x 500ml cultures of bacteria containing the relevant plasmid were started in media supplemented with the appropriate antibiotic. The cells were grown overnight at 37°C with shaking. Early in the morning the cells were harvested by centrifugation at 6000rpm for 10 minutes at 4°C in a GSA rotor. The supernatant was removed and the pellet resuspended in 100ml cold TE (10mM Tris.HCl, pH 7.5; 1mM EDTA). The cells were centrifuged again at 6000rpm for 10 minutes at 4°C in a GSA rotor. The pellet was retained and resuspended in 3ml TS solution (30mM Tris.HCl, pH 8.0; 25% sucrose). 1ml of lysozyme (10mg/ml) was added and the suspension incubated for 5 minutes at 4°C. 1ml of EDTA solution (500mM EDTA, pH 8.5) and 400µl of DNase free RNase (10mg/ml) were added and the solution incubated at 4°C for a

further 5 minutes. 5ml of Triton mix (1ml 10% Triton X-100; 12.5ml 500mM EDTA, pH 8.5; 5ml 1M Tris.HCl, pH 8.0) was added and the tube inverted several times before incubating at 4°C for 10 minutes until lysis occurred. The solution was then centrifuged at 15000rpm for 30 minutes at 4°C in a SS34 rotor. The supernatant was transferred to fresh tube and made up to 10ml with lysis mix (15ml TS solution; 5ml 500mM EDTA, pH 8.5; 25ml Triton mix; 7ml water). 9g of Caesium chloride was then added and the tube shaken gently to dissolve the Caesium chloride. The tube was stored on ice and 900μl of ethidium bromide (10mg/ml) added. The solution was mixed gently and the solution transferred to a heat seal tube using a syringe and needle. Checking that the tubes were balanced, each solution was centrifuged at 38000rpm for 60 hours at 18°C in a Ti50 rotor.

The DNA bands were visualised under UV light and the lower of the two bands containing the plasmid DNA removed using a syringe and needle.

2.12 Agarose gel electrophoresis of DNA.

DNA samples were routinely analysed by electrophoresis in 1% agarose gels. Gels were prepared by boiling 0.5g of agarose in 50ml of 1xTBE buffer. The solution was allowed to cool to approximately 45°C before adding 1μl of ethidium bromide (10mg/ml). The gel was poured into the casting tray. The comb was added and the gel allowed to set at room temperature. The comb was removed and the gel placed in the electrophoresis apparatus. 1xTBE was added so as to just cover the surface of the gel. The DNA solutions were carefully pipetted into each well and electrophoresed at a constant voltage of 60V for 2-3 hours. The samples were visualised under UV light.

Samples to be loaded in gels were mixed with the appropriate volume of 6x loading buffer II (0.25% bromophenol blue; 0.25% xylene cyanol; 15% Ficoll (Type 400) in water).

2.13 Isolation of DNA from LMP agarose gels.

Low melting point (LMP) agarose gels of the appropriate concentration were prepared as above using 1xTAE as a buffer.

After electrophoresis at 60V the DNA was visualised under UV. The band of interest was excised from the gel using a scalpel and placed in an Eppendorf tube.

The DNA was purified from the gel slice in one of three ways, depending on the size of the DNA fragment.

(i) Prep-A-Gene purification (for DNA fragments above 200bp)[112].

The volume of the gel slice was estimated and 3 volumes of Prep-A-Gene binding buffer added (section 2.8). The sample was heated at 50°C until all the agarose had dissolved. The appropriate volume of Prep-A-Gene matrix was then added and the solution vortexed briefly before incubating at room temperature for 10 minutes. The Prep-A-Gene DNA purification procedure was then followed as described previously.

(ii) Purification by Phenol extraction.

The gel slice containing the DNA was heated to 70°C for 5 minutes to melt the agarose before adjusting the volume to 500µl with TE (10mM Tris.HCl, pH 8.0; 1mM EDTA). An equal volume of phenol buffered in

Tris.HCl (10mM, pH 8.0) was added and the sample vortexed vigorously. The phases were separated by centrifugation at 12000g for 1 minute. The upper phase was removed and transferred to a clean tube. The DNA was purified further as described previously (section 2.9).

(iii) Purification via elution[113].

The gel slice was placed in an Eppendorf tube and incubated in 500μl water for 15 minutes. The gel was transferred to a 500μl Eppendorf tube containing a siliconised glass wool plug and frozen at -70ºC. The tube was punctured at the bottom and transferred to a 1.6ml Eppendorf tube. The tubes were centrifuged at 12000g for 10 minutes and the eluant containing the DNA recovered. The DNA was then precipitated in the usual way (section 2.10).

2.14 Native PAGE of nucleic acids.

For a 5% native gel the following were combined:

47.5ml	distilled water
15ml	5xTBE
7.5ml	50% acrylamide:bisacrylamide
400μl	10% ammonium persulphate

The solution was mixed and sonicated for 15 minutes. 30μl TEMED was added and the gel poured between the siliconised glass plates within vertical electrophoresis apparatus. A comb was added and the gel allowed

to set for 1-2 hours. The DNA samples were loaded and electrophoresed at 20-30mA for the appropriate amount of time.

2.15 Denaturing PAGE of nucleic acids.

Denaturing PAGE was carried out as above (section 2.14) except urea was added to a final concentration of 40% (w/v) and the solution made up to the appropriate volume with distilled water.

2.16 Dideoxy sequencing of DNA.

Sequencing of double and single-stranded DNA was carried out according to the method of Sanger *et al* using the T7 Sequencing Kit (Pharmacia)[114].

Before carrying out the sequencing reactions a 12% denaturing polyacrylamide gel was poured, allowed to set and pre-electrophoresed for 1 hour at 60W (section 2.15).

2.16.1 Annealing of primer to single-stranded template.

The sequencing primer was annealed to each DNA sample as follows in an Eppendorf tube:

10µl	template DNA (1-2µg)
2µl	primer DNA (4mg/ml)
2µl	annealing buffer (350mM NaCl; 100mM MgCl$_2$; 280mM Tris.HCl, pH 7.5)

Each tube was vortexed briefly and centrifuged before incubating at 60°C for 10 minutes. Each reaction was then incubated for a further 10 minutes at room temperature.

2.16.2 Annealing of primer to double-stranded template.

The concentration of the double-stranded template DNA was adjusted such that 8µl contained 1.5-2µg.

To denature the double-stranded template, the following were added to an Eppendorf tube:

8µl   template (1-2µg)

2µl   2M NaOH

The tube was vortexed briefly and the contents incubated at 37°C for 30 minutes.

3µl of sodium acetate (3M) was added and 7µl of distilled water. 60µl of ethanol was added and the mixture placed on dry ice for 15 minutes. The DNA was pelleted by centrifuging at 12000g for 30 minutes at 4°C. The supernatant was removed and the pellet washed in cold 70% ethanol before drying under vacuum. The dried pellet was resuspended in 10µl of distilled water.

To 10µl of the denatured pellet was added 2µl of annealing buffer (350mM NaCl; 100mM MgCl$_2$; 280mM Tris.HCl, pH 7.5) and 2µl of primer (10µg). The reaction was incubated at 37°C for 20 minutes. The tube was then placed at room temperature for 10 minutes before proceeding with the sequencing reactions (see below).

2.16.3 Sequencing reactions.

The sequencing reactions were carried out as follows:

Four wells of an Elisa plate were labelled 'A', 'C', 'G' and 'T' respectively. 2.5µl of 'A' mix-short, 'C' mix-short, 'G' mix-short and 'T' mix-short were pipetted into the corresponding wells.

(Note: each sequencing mix contained 15µM of the relevant ddNTP in addition to 150µM of each dNTP, 10mM $MgCl_2$, 50mM NaCl and 40mM Tris.HCl, pH 7.5)

'Enzyme Premix' was then prepared as follows in an Eppendorf tube,

water	n µl
Labelling mix-dATP	3n µl
T7 DNA Polymerase (1.5U/µl)	2n µl
[α-^{35}S]dATP (10µCi/µl)	n µl

where n = number of templates.

(Note: labelling mix-dATP contained 2µM dCTP, 2µM dGTP and 2µM dTTP)

6µl of 'Enzyme Premix' was then added to each of the annealed templates and primers. The solutions were mixed gently and incubated at room temperature for 5 minutes. Meanwhile, the four sequencing mixes were warmed at 37ºC.

After incubation was complete, 4.5µl of each of the labelling reactions was transferred into each of the four pre-warmed sequencing mixes and incubated at 37°C for 5 minutes.

5µl of Stop Solution (20mM EDTA, pH 7.5; 0.05% (w/v) xylene cyanol FF; 0.05% (w/v) bromophenol blue) was added to each well and the reactions heated at 80°C for 2 minutes.

1.5µl of each reaction was loaded onto the pre-electrophoresed gel and electrophoresed at 60W for approximately 3 hours.

The gel was washed in 10% methanol/10% acetic acid for 30 minutes and the gel dried for 3 hours at 80°C. The dry gel was autoradiographed for 36 hours at room temperature before developing the film and determining the DNA sequence.

2.17 Construction of pOP1.

pTZ19R was supplied on LB/ampicillin(50µg/ml) plates in TG1 cells. A bacterial colony was grown 5ml LB supplemented with ampicillin (50µg/ml) overnight at 37°C with shaking. The plasmid DNA was isolated as described previously (sections 2.7, 2.8) and prepared for cloning of the OP1 DNA fragment as follows:

2.17.1 Digestion with SmaI.

The following were placed in an Eppendorf tube:

10µl	pTZ19R DNA (10µg)
6µl	sterile water
2µl	10xReact4 buffer

2μl   SmaI (10U/μl)

The contents were mixed briefly and incubated at 37°C for 3 hours. The reaction was stopped by heating at 70°C for 10 minutes. The DNA was purified with Prep-A-Gene matrix (section 2.8) and precipitated (section 2.10). The pellet was resuspended in 25μl sterile water.

2.17.2 Digestion with EcoRI.

The following were combined in an Eppendorf tube:

20μl   pTZ19R/SmaI DNA (10μg)

2.5μl   10xReact 3 buffer

2μl   EcoRI (10U/μl)

0.5μl   sterile water

The contents were vortexed briefly and incubated at 37°C for 3 hours. The reaction was stopped by incubating at 70°C for 10 minutes. The DNA was purified using Prep-A-Gene (section 2.8) and precipited (section 2.10). The pellet was resuspended in 20μl sterile water and a 5μl sample removed for gel analysis.

2.17.3 Ligation of OP1 DNA into pTZ19R.

The following were combined in an Eppendorf tube:

2μl   pTZ19R/SmaI/EcoRI (50ng)

0.5μl   OP1 DNA (10ng)

0.5μl	T4 DNA ligase (1U/μl)
0.5μl	10mM ATP
0.5μl	10x ligation buffer
1μl	sterile water

The contents were mixed briefly and incubated at 4°C overnight. The recombinant plasmid DNA was transformed into fresh competent TG1 cells (sections 2.5, 2.6) and grown on LB/ampicillin(50μg/ml) plates. The clones were identified by dideoxy sequencing (section 2.16) and stored as a glycerol stock (section 2.3).

2.18 Labelling of OP1 DNA with [α-^{32}P]dCTP.

2.18.1 Preparation of OP1 DNA fragment.

A large scale culture of pOP1 DNA was prepared (section 2.11) and used to prepare the OP1 fragment.

2.18.2 Digestion with EcoRI.

The following were combined in an Eppendorf tube:

25μl	pTZ19R DNA (10μg)
3μl	10xReact3 buffer
2μl	EcoRI (10U/ml)

The contents were mixed briefly and incubated at 37°C for 3 hours. The reaction was stopped at 70°C for 10 minutes. The DNA was precipitated (section 2.10) and resuspended in 25μl of sterile water.

2.18.3 Digestion of XbaI.

The following were combined in an Eppendorf tube:

25μl	pTZ19R/EcoRI DNA (10μg)
3μl	10xReact2 buffer
2μl	XbaI (10U/μl)

The contents were mixed briefly and incubated at 37°C for 3 hours. The reaction was stopped by heating at 70°C for 10 minutes.

The DNA was purified via LMP agarose gel electrophoresis (section 2.13).

2.18.4 Radioactive labelling of OP1 DNA with [α-³²P]dCTP.

The following were combined in an Eppendorf tube:

20μl	OP1 DNA (1μg)
5μl	sterile water
3μl	10xReact3 buffer
1μl	[α-³²P]dCTP (10mCi/ml)
1μl	Klenow fragment (10U/μl)

The contents were mixed briefly and incubated at 37°C for 30 minutes. The reaction was stopped by heating at 70°C for 10 minutes.

The DNA was purified via two rounds of ethanol precipitation (section 2.10) and resuspended in 10μl sterile water and stored at -20°C in a perspex container.

(Note: All manipulations involving ^{32}P were carried out behind a perspex shield.)

2.19 PCR of the GCN4 bZip region from YCp88-GCN4.

Before carrying out the PCR, the plasmid DNA was linearised via digestion with BamHI. The restriction digest was analysed by agarose gel electrophoresis and the linearised DNA purified using Prep-A-Gene matrix (setion 2.8).

The following were then placed in a 500μl Eppendorf tube:

0.5μl	primer 575P (0.07μM)
0.5μl	primer 576P (0.07μM)
0.5μl	Taq Polymerase (5U/μl)
2μl	dNTP's (5mM)
5μl	10xPCR buffer
10μl	YCp88-GCN4 DNA (1μg)
31.5μl	sterile distilled water

The solutions were overlayed with 40μl mineral oil and centrifuged briefly. The tube was placed in a Techne PHC-3 Thermal Cycler. The following cycle was then implemented,

90ºC	2 mins	
90ºC	1 min	
54ºC	1 min	30 cycles
72ºC	2 mins	
72ºC	5 mins	

The product of the PCR was checked by electrophoresis at 60V on a 2% agarose gel.

5μl portions of the crude PCR products were digested with 10U EcoRI in React3 buffer. The reactions were incubated for 2 hours at 37ºC. The enzyme was deactivated by heating at 70ºC for 10 minutes.

The bZip DNA was separated from the other fragments via LMP agarose gel electrophoresis (section 2.13). The DNA was purified using Prep-A-Gene matrix (section 2.8).

2.20 PCR of the GCN4 gene from YCp88-GCN4.

The PCR of the GCN4 gene was carried out in an identical manner to that above, except primer 575P was replaced by primer 120X at an identical concentration.

2.21 Construction of λZAPII library.

The ligation and packaging reactions were carried out using a kit purchased from Stratagene.

2.21.1 Ligation of bZip gene into λZAPII arms.

The following were placed in an Eppendorf tube:

2μl	bZip DNA (5ng)
1μl	λZAPII/EcoRI arms (1μg)
0.5μl	10x ligation buffer
0.5μl	10mM ATP
0.5μl	T4 DNA ligase (4U/μl)
0.5μl	sterile water

The reaction was incubated overnight at 4°C.

2.21.2 Ligation of GCN4 gene into λZAPII arms.

The ligation of the GCN4 gene into λZAPII was carried out in an identical manner to that for the bZip region except the bZip DNA was replaced by 2μl GCN4/EcoRI DNA (20ng).

2.21.3 Packaging of λZAPII DNA.

This was carried out as follows using Gigapack II Plus packaging extract purchased from Stratagene:

Two extracts (1 yellow, 1 red) were removed from a -70°C freezer and placed on dry ice. At the same time Sonic extract (yellow tube) was thawed. The Freeze/Thaw extract (red tube) was quickly thawed between fingers until it just began to thaw. 3μl of the ligated λZAPII DNA was

added immediately to the Freeze/Thaw extract before placing on ice. 15μl of Sonic extract was added before stirring carefully to avoid air bubbles. The tube was centrifuged briefly. The mixture was then incubated at room temperature for 2 hours. 500μl of SM was added, before adding 20μl of chloroform. The tube was mixed gently before centrifuging briefly to sediment the debris. The supernatant was removed and stored at 4°C.

2.21.4 Preparation of host bacteria.

20ml of LB supplemented with maltose (0.2% w/v) and MgSO$_4$ (10mM) was inoculated with a single colony of XL1-Blue bacteria. The broth was grown for approximately 8 hours at 37°C with shaking until the OD$_{600}$ reached 1.0. The bacteria were then pelleted at 12000g for 10 minutes before resuspending the cells in 10ml of MgSO$_4$ (10mM). These cells were stored at 4°C for up to 2 days.

2.21.5 Infection of packaged λZAPII phage into *E.coli*.

The packaged λZAPII phage were diluted by a factor of 10^2 and 10^4 in SM (100mM NaCl; 10mM MgSO$_4$; 50mM Tris.HCl, pH 7.5; 0.01% w/v gelatin). 10μl of each sample was added to 200μl XL1-Blue cells before incubating at 37°C for 20 minutes. 3ml of top agar at a temperature less than 50°C was added to each and the mixture quickly poured onto LB/tetracycline (12.5μg/ml) plates. The plates were incubated overnight at 37°C and the phage titre recorded.

2.21.6 Amplification of λZAPII phage.

XL1-Blue cells were grown to an OD_{600}=0.5 and 100μl mixed with λZAPII phage (1×10^6 pfu). The suspension was incubated for 15 minutes at 37ºC and added to 4ml LB supplemented with $MgSO_4$ (10mM) and maltose (0.2% w/v). The cells were grown overnight at 37ºC with shaking. Chloroform (2 drops) was added and the incubation continued for a further 15 minutes at 37ºC. The cells were pelleted at 12000g for 20 minutes and the supernatant retained. Chloroform (1 drop) was added and the suspension stored at 4ºC.

2.21.7 Blue/white colour test for recombinant phage.

10μl of λZAPII phage were added to 200μl XL1-Blue cells (OD_{600}=0.5 in 10mM $MgSO_4$) and incubated for 20 minutes at 37ºC. 3ml of top agar, 15μl of IPTG (500mM) and 50μl of X-gal (250mg/ml) were added and plated on LB/tetracycline (12.5μg/ml) plates. The plates were incubated overnight at 37ºC.

2.22 *In vitro* screening of λZAPII library.

XL1-Blue host bacteria were prepared as previuosly described. 5μl batches of a 1×10^{-2} dilution of λZAPII-bZip phage in SM (100mM NaCl; 10mM $MgSO_4$; 50mM Tris.HCl, pH 7.5; 0.01% w/v gelatin) were added to 200μl of XL1-Blue cells and the broth incubated at 37ºC for 20 minutes. 3ml of top agar was added and the mixture poured onto LB/tetracycline (12.5μg/ml) plates. The plates were incubated at 37ºC for 4 hours until the plaques were barely visible. A nitrocellulose filter was soaked in

IPTG (10mM) and dried at room temperature before laying it over a plate. The plates were incubated for a further 4 hours at 37°C then chilled at 4°C for 30 minutes. The filter was marked by pushing a needle filled with ink through both filter and plate. The filter was peeled off and allowed to dry before being subjected to the following cycle of washing, denaturing, renaturing and probing:

2 x 5 mins	6M Gu.HCl in 1x buffer
2 x 5 mins	3M Gu.HCl in 1x buffer
2 x 5 mins	1.5M Gu.HCl in 1x buffer
2 x 5 mins	0.75M Gu.HCl in 1x buffer
5 mins	1x buffer
5 mins	1x buffer + 5% NFDM
5 mins	1x buffer + 0.25% NFDM
overnight	1x buffer, 0.25 NFDM, sonicated salmon sperm DNA(10mg/ml), 3µl of ^{32}P-labelled probe DNA.
3 x 5 mins	1x buffer + 0.25% NFDM

The nitrocellulose filters were allowed to dry at room temperature before laying on filter paper and establishing an autoradiograph.

Positive plaques were isolated by coring with a sterile pipette end. Each plaque was placed in an Eppendorf tube and dissolved in 500µl SM (100mM NaCl; 10mM $MgSO_4$; 50mM Tris.HCl, pH 7.5; 0.01% w/v gelatin) and 20µl chloroform. Each tube was vortexed to release the phage particles into the SM (100mM NaCl; 10mM $MgSO_4$; 50mM Tris.HCl, pH 7.5; 0.01% w/v gelatin). The samples were incubated at room temperature for 2 hours. Serial dilutions were made and stored at 4°C.

(Note: 1x buffer contained the following 25mM HEPES, pH 7.5; 10mM DTT; 4mM KCl; 3mM $MgCl_2$)

2.23 *In vivo* excision of pBluescript SK- phagemids.

The following were combined in an Eppendorf tube:

200μl	XL1-Blue cells (OD_{600}=0.5 in 10mM $MgSO_4$)
200μl	λZAPII phage stock (1.5×10^4 pfu/ml)
1μl	Exassist helper phage ($>1 \times 10^{10}$ pfu/ml)

3ml of 2xYT medium was added and each sample incubated at 37°C for 3 hours with shaking. The cells were then heated at 70°C for 20 minutes before centrifuging for 20 minutes at 12000g. The supernatant containing the pBluescript SK- phagemid was removed and stored at 4°C.
10μl of phage stock was added to 200μl SOLR cells (OD_{600}=0.5 in 10mM $MgSO_4$) in a sterile Eppendorf tube and incubated at 37°C for 15 minutes. 100μl of the suspension was plated on LB/ampicillin (50μg/ml)/kanamycin (50μg/ml) plates and incubated overnight at 37°C.

2.24 Preparation of pBluescript SK- single-stranded DNA.

Early in the morning, cultures of SOLR bacteria, containing the relevant pBluescript SK- phagemid, were started from single colonies taken from LB/ampicillin (50μg/ml) plates. The cultures were grown in LB medium supplemented with maltose (0.2% w/v), $MgSO_4$ (10mM), ampicillin (50μg/ml) and kanamycin (50μg/ml). Late in the afternoon 3ml batches of LB, supplemented as above, were inoculated with enough bacteria to

74

obtain an $OD_{600}=0.05$. The cultures were grown to an $OD_{600}=0.1$ before adding R408 helper phage ($>7.5 \times 10^{10}$ pfu/ml) to each culture at a multiplicity of infection (moi) of 1:1. The cells were grown overnight at $37^{\circ}C$ with shaking.

In the morning the cells were heated for 15 minutes at $65^{\circ}C$ before centrifuging at 12000g for 2 minutes to precipitate the cell debris. 1.2ml aliquots of the supernatants were then transferred to fresh tubes and the single stranded phage precipitated by addition of 300μl ammonium acetate (3.5M, pH 7.5); 20% PEG solution. The tubes were inverted to mix before incubating at room temperature for 30 minutes. The phage were pelleted by centrifugation at 12000g for 20 minutes. The supernatants were removed and the tubes spun for a further minute and any remaining supernatant removed. The pellets were resuspended in 200μl TE (10mM Tris.HCL, pH 8.0; 1mM EDTA) and stored at $-20^{\circ}C$.

The single-stranded DNA was purified either via extraction with phenol (section 2.9) or adsorption onto Prep-A-Gene matrix (section 2.8).

2.25 Helper phage amplification.

10ml XL1-Blue cells were grown in LB supplemented with maltose (0.2% w/v) and $MgSO_4$ (10mM) to an $OD_{600}=0.3$ at $37^{\circ}C$ with shaking. Helper phage at a multiplicity of infection of 20:1 was added and the cells grown for a further 8 hours at $37^{\circ}C$ with shaking. The cells were heated at $65^{\circ}C$ for 15 minutes and centrifuged at 12000g for 5 minutes. The supernatant was retained and stored at $4^{\circ}C$.

2.26 [^{35}S]Methionine/IPTG induction of fusion protein.

A bacterial colony of SOLR cells containing the relevant pBluescript SK-phagemid was grown in 10ml LB supplemented with maltose (0.2% w/v), $MgSO_4$ (10mM) and ampicillin (50μg/ml) at 37°C with shaking to an OD_{600}=0.4. IPTG to a final concentration of 10mM was added, along with 5μl of [^{35}S]Methionine (10mCi/ml). The culture was then incubated at 37°C with shaking. 1ml aliquots were removed at ½, 1, 2 and 3 hours. Each aliquot was centrifuged for 2 minutes at 12000g and the pellet resuspended in SDS loading buffer.

A 10-20% gradient SDS-polyacrylamide gel was poured and overlayed with a 3% stacking gel (section 2.30). The gel was allowed to set before loading each of the protein samples. The gel was electrophoresed at 30mA for 7 hours. The gel was stained in Coomassie Blue R250 then destained for 24 hours in 40% methanol/10% acetic acid/50% water at room temperature. The gel was dried at 80°C for 3 hours and overlayed with XOMAT film. The radioactive gel was autoradiographed for 36 hours, before developing the film to observe the image.

2.27 Large scale preparation of fusion protein.

A 20ml culture of SOLR cells containing the relevant pBluescript SK-phagemid was grown in LB supplemented with maltose (0.2% w/v), magnesium sulphate (10mM) and ampicillin (50μg/ml) overnight at 37°C with shaking. The culture was transferred to 500ml LB, supplemented as above, and grown for 2 hours at 37°C with shaking, to an approximate OD_{600} of 1.0. 50ml batches of this culture were transferred to flasks containing 500ml of LB, supplemented as above, and grown for

approximately 2 hours to an OD_{600} of 0.4. IPTG to a final concentration of 1mM was added to each flask and the cultures grown for a further 3 hours at 37°C with shaking until the OD_{600} was approximately 1.0.

The cells were then centrifuged at 5000rpm for 10 minutes at 4°C in a GSA rotor. The pellets were recovered and resuspended in 10ml HEPES (50mM, pH 7.5). Each sample was placed on ice and sonicated 15 times for 30 seconds at 30 second intervals. The suspension was then centrifuged at 14000rpm for 15 minutes at 4°C in a SS34 rotor. The supernatant was removed and 200μl DNase I (1mg/ml) added to each. Each sample was incubated at room temperature for 30 minutes. The samples were centrifuged at 14000rpm for 15 minutes at 4°C in a SS34 rotor and the supernatant transferred to a fresh tube.

2.28 Ammonium sulphate fractionation of cell free extract.

The cell free extract was fractionated by addition of ammonium sulphate to final concentration of 40% at 4°C with constant stirring. The sample was centrifuged at 15000rpm for 10 minutes at 4°C in a SS34 rotor and the supernatant retained. Ammonium sulphate was added to final concentration of 60% at 4°C with stirring. The sample was then centrifuged at 15000rpm for 10 minutes at 4°C as before and the pellet of protein resuspended in the minimum volume of HEPES (50mM, pH 7.5) and dialysed for 1-2 days at 4°C. Samples were stored as a 15% glycerol stock at -20°C.

2.29 Gel filtration chromatography.

A slurry of Sephadex G-50 resin in HEPES (50mM, pH 7.5) was prepared and poured into a glass column The resin was allowed to settle and then washed with 500ml buffer at 1ml/min. The column was loaded with 500µl cell free extract and eluted with HEPES (50mM, pH 7.5) at 1ml/min. The eluant was monitored at 280nm and the fractions collected at 15 minute intervals until no further proteins eluted from the gel. The protein fractions were identified and concentrated by ultrafiltration through a YM-3 membrane at 4°C. The concentrated samples were identified via SDS-PAGE and the concentration determined.

2.30 SDS-Polyacrylamide gradient electrophoresis.

Two different percentage (10% and 20%) polyacrylamide solutions were made as follows:

10%

10ml	water
8.3ml	30% acrylamide:bisacrylamide
6.3ml	Tris.HCl (1.5M, pH 8.8)
250µl	10% SDS
100µl	10% ammonium persulphate

20%

1.75ml	water
16.6ml	30%acrylamide:bisacrylamide
6.3ml	Tris.HCl(1.5M, pH 8.8)
250µl	10% SDS

100μl   10% ammonium persulphate

Each solution was sonicated for 15 mitutes before adding 7.5ml TEMED to each. The solutions were immediately poured into separate chambers of a gradient former which was connected to a peristaltic pump. The polyacrylamide solution was pumped between two glass plates and overlayed by water.

A 3% stacking gel was made as follows and poured onto the gradient gel after it had set:

<div align="center">

3%

</div>

6.1ml	water
2.5ml	Tris.HCl (0.5M, pH 6.8)
1.3ml	30% acrylamide:bisacrylamide
100μl	10% SDS
50μl	10% ammonium persulphate
10μl	TEMED

A comb was inserted into the top of the gel, between the glass plates, which was then left to polymerise.

The gradient gel was attatched to a vertical gel electrophoresis apparatus (Biorad Protean II Slab Gel) and running buffer (180mM Glycine; 25mM Tris; 0.1%(w/v) SDS) added.

The protein samples were suspended in an equal volume of sample buffer (100mM Tris.HCl, pH 6.8; 20% (v/v) glycerol; 15% β-mercaproethanol; 3% (w/v) SDS; 0.25% (w/v) bromophenol blue) and boiled for 5 minutes. The samples along with protein standards were loaded onto the gel and electrophoresed at 20mA for 4 hours.

The gel was removed from the apparatus and stained with Coomassie Blue R250 in 40% methanol/10% acetic acid for 30 minutes. The gel was destained in 40% methanol/10% acetic acid and the sample visualised. The gel was then dried under vacuum at 80°C for 1 hour.

2.31 Determination of protein concentration.

The procedure was carried out according to that previously described by Bradford[115].

50mg of Coomassie Blue G-250 was dissolved in 25ml of 95% ethanol. This was added to 48.5ml of 88% phosphoric acid and the solution made up to 500ml with distilled water.

BSA to a concentration of 1mg/ml was suspended in HEPES (50mM, pH 7.5) and a series of dilutions made. 5ml of the reagent solution was added to 100μl of each protein solution and the absorbance at 595nm measured.

A graph of protein concentration versus absorbance was plotted and used to determine the amount of protein from an unknown 100μl sample.

2.32 Gel retardation assay.

A series of dilutions of the purified protein were made and stored at 4°C. OP1 DNA was labelled with [α-32P]dCTP to a high specific activity (section 2.18) and a 5% native polyacrylamide gel poured (section 2.14). The following were then combined in an Eppendorf tube:

1μl   32P-labelled OP1 DNA

10µl	2x retard buffer (50mM HEPES, pH 7.5;
	80mM NaCl; 10mM MgCl$_2$; 0.2mM EDTA;
	1mM DTT; 8% Ficoll)
1µl	sonicated salmon sperm DNA (1mg/ml)
1-8µl	protein in HEPES (50mM, pH 7.5)

Each sample was made up to 20µl with sterile water, mixed and incubated on ice for 1 hour.

Each sample was loaded onto a pre-electrophoresed gel at 4°C and electrophoresed at 30mA until the samples had entered the gel, then at 20mA for 2 hours.

The gel was placed in a cassette and overlayed with XOMAT film for 2-3 hours at -70°C. The film was developed and the retarded bands identified.

(Note: all manipulations involving [32]-P were carried out behind a perspex shield.)

2.33 *In vitro* mutagenesis.

2.33.1 Phosphorylation of spiked oligonucleotide.

The following were combined in an Eppendorf tube:

14µl	sterile water
6µl	500mM Tris.HCl, 1mM EDTA (pH 8.0)
2µl	spiked primer (5 OD$_{260}$/ml)
3µl	100mM ATP
3µl	100mM MgCl$_2$

2µl	100mM DTT
1µl	T4 polynucleotide kinase (10U/µl)

The contents of the tube were vortexed briefly before incubating at 37°C for 30 minutes. The reaction was stopped by incubating at 70°C for 10 mimutes.

2.33.2 Labelling of spiked primer with [γ-^{32}P]ATP.

The spiked oligonucleotide was labelled under identical conditions to those used above (section 2.33.1) except the ATP was replaced by 1µl [γ-^{32}P]ATP (10mCi/ml) and 2µl of sterile water.

The DNA was loaded onto a denaturing polyacrylamide gel and electrophoresed (section 2.15). The gel was placed in a cassette and exposed to XOMAT film for 2-3 hours at -70°C. The film was developed and the bands visualised.

2.33.3 Preparation of RF-IV DNA.

Annealing of spiked primer to template DNA.

The following mixture was combined in a sterile Eppendorf tube:

10µl	500mM Tris.HCl, 1mM EDTA (pH 8.0)
10µl	500mM NaCl
10µl	single-stranded pbZip DNA (1µg/µl)
6µl	phosphorylated spiked primer

The mixture was incubated at 70ºC for 5 minutes, 55ºC for 20 minutes and 37ºC for 30 minutes in a thermal cycler. The solution was then placed on ice.

Polymerisation.

The following were added to the template/primer mixture after the annealing mixture had been cooled on ice:

2.5µl	10mM dATP
2.5µl	10mM dGTP
2.5µl	10mM dTTP
2.5µl	10mM dCTPαS
10µl	10mM ATP
10µl	100mM MgCl$_2$
3µl	500mM Tris.HCl, 1mM EDTA (pH 8.0)
1µl	Klenow fragment (10U/µl)
2µl	T4 DNA ligase (7U/µl)
8µl	sterile water

The contents were mixed and spun briefly before incubating for 40 hours at 16ºC. The reaction was stopped by heating at 70ºC for 10 minutes. A 10µl sample was removed for gel analysis.

2.33.4 Strand specific nicking of RF-IV DNA.

The polymerised RF-IV DNA was precipitated by addition of sodium acetate to a final concentration of 0.3M and 2 volumes of ethanol for 1

hour on dry ice. The DNA was pelleted at 12000g for 30 minutes at 4°C. The pellet was then washed in 70% ethanol and dried under vacuum. The following were added to the pellet:

38µl	sterile water
5µl	100mM DTT
3ml	100mM MgCl$_2$
3ml	500mM NaCl
1.2ml	500mM Tris.HCl, 1mM EDTA (pH 8.0)
3µl	NciI (7.5U/µl)

The reaction was incubated at 37°C for 3 hours. The enzyme was inactivated by heating at 70°C for 10 minutes. 5µl was removed for gel analysis.

2.33.5 Exonuclease III gapping reaction.

0.8µl NaCl (5M) was added to the nicked DNA solution (section 2.33.4) and the solution cooled on ice. 3µl exonuclease III (65U/µl) was added and the reaction vortexed briefly. The reaction was incubated at 37°C for 30 minutes before heating at 70°C for 10 minutes.
A 10µl sample was removed for agarose gel electrophoresis.

2.33.6 T7 exonuclease gapping reaction.

The nicked DNA (section 2.33.4) was precipitated (section 2.10) and resuspended in the following:

24μl	sterile water
3μl	10xReact8 buffer
3μl	T7 exonuclease (50U/μl)

The reactants were mixed briefly and incubated at 37°C for 50 minutes. The contents were heated at 70°C for 10 minutes and 10μl removed for gel analysis.

CHAPTER THREE

The λZAPII system

3.1 Overview.

The central aim of this project was to clone the bZip region of GCN4 into a vector which allowed expression of the protein, identification of clones and mutagenesis without the need for subcloning.

A vector system which was ideal for the experiments described herein was the λZAPII system[106,116]. This vector allows the cloned gene to be screened with radioactive DNA probes *in vitro* via induction of the cloned gene with IPTG and subsequent transfer of the fusion protein to nitrocellulose. Identification of DNA-binding proteins is facilitated by excising the pBluescript SK- phagemid, contained within the λZAPII genome, and sequencing the phagemid-derived single-stranded DNA produced.

Excision of the pBluescript SK- phagemid allows overexpression of the bZip fusion protein with IPTG. The binding characteristics of the fusion protein can then be investigated further via gel retardation assay.

Single-stranded phagemid DNA can then generated and used to carry out *in vitro* mutagenesis of the bZip gene. The mutant library can be recloned into λZAPII and any mutants which bind to target DNA characterised following the procedures described above.

A summary diagram detailing each step in the process is shown overleaf (fig. 3.1).
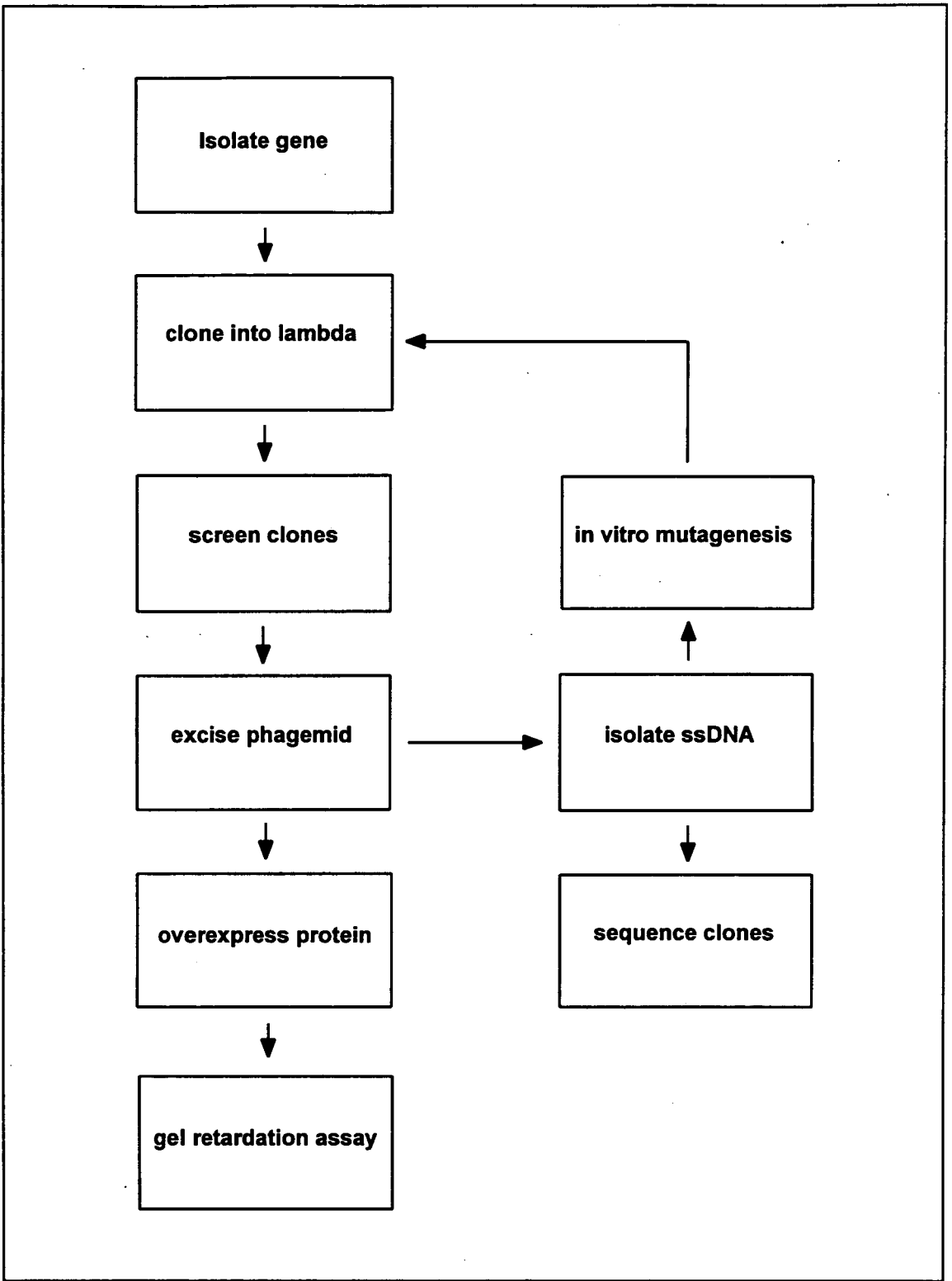
Fig.3.1. The λZAPII system.

3.2 Preparation of the GCN4 and bZip genes via PCR.

3.2.1 Introduction.

The first stage of the project was to isolate the genes of interest and manipulate them such that they could be cloned into a suitable vector, which in this case was λZAPII[106,116]. A convenient way of doing this was to amplify the appropriate DNA sequences via polymerase chain reaction (PCR) from the plasmid YCp88-GCN4 which contains the entire GCN4 gene[84,117].

PCR is a simple process which involves the use of synthetic oligonucleotides and a DNA polymerase enzyme to carry out the semi-synthetic amplification of the target DNA sequence. Briefly, short synthetic oligonucleotide primers designed to bind different strands of a larger, double-stranded target sequence are synthesised. The target DNA is mixed with a large excess of the primers, and the temperature raised to thermally denature the double-stranded target. The temperature is lowered, and the primers anneal to their respective sites on the target. A thermostable DNA polymerase enzyme and deoxynucleotide triphosphates (dNTP's) are present in the reaction mixture and DNA synthesis then proceeds from the 3'-ends of the primers, copying the target strand to which the primer is bound, until the end of the target is reached. This constitutes one cycle. The entire process can be repeated a number of times thereby exponentially amplifying the DNA target sequence.

PCR has a number of advantages over other methods of isolating DNA sequences from plasmids.

(i) Small quantities of DNA can be amplified easily.

(ii) The DNA can be manipulated such that restriction sites and other sequences not present in the wild-type sequence can be engineered into the final product.

(iii) The DNA sequences can be deleted to a suitable size simply by altering the primer sequence.

(iv) The amplified DNA can be purified easily.

3.2.2 The GCN4 gene.

The GCN4 gene is 281 residues in length and the region of interest, the bZip region, is situated at the amino-terminus between residues 226 and 281. This region is able to bind DNA independently from the rest of the protein[84] (Fig.3.2).
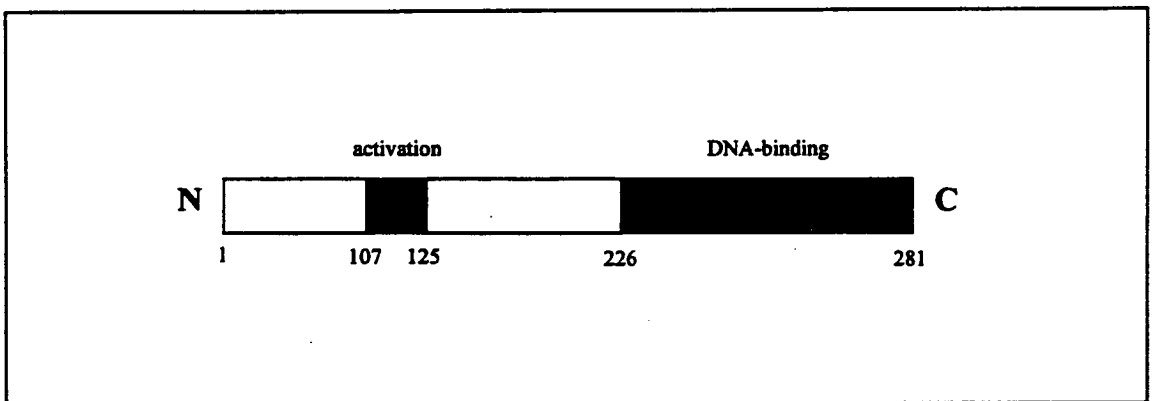
Fig.3.2. Schematic representation of the GCN4 protein.

It was decided to isolate and clone both the full length GCN4 gene and the GCN4 bZip region. Cloning the full length GCN4 gene into λZAPII would ensure that the β-galactosidase sequences would be separated from the DNA-binding region and would therefore be unlikely to interfere with the binding of the fusion protein to DNA.

3.2.3 Design of PCR primers.

A number of different criteria were considered when designing the oligonucleotide primers for each DNA sequence.

The bZip gene.

The bZip region is situated between residues 226 and 281 of the GCN4 gene. The PCR process requires two primers; one which anneals to the start of the gene (the start primer) and one which anneals to the end of the gene (the stop primer). Because the two DNA strands of the plasmid are denatured during the PCR process, a primer which is complementary to the start of the non-coding strand and a primer which is complementary to the end of the coding strand are required. Therefore, sequences which were complementary to these segments of the bZip region were included in each of the primers.

The sequence of each primer was also constrained by the nature of the cloning vector. Since the amplified bZip DNA was eventually cloned into an EcoRI site within the multiple cloning site of λZAPII, it was necessary to engineer an EcoRI sequence in each primer. It was also essential to ensure that the bZip DNA could be inserted into the *lacZ* gene of λZAPII within the correct reading frame for gene expression. This was achieved by analysing the sequences of the *lacZ* gene of λZAPII and those of the plasmid derived sequence and tailoring the sequence of each of the primers to allow in frame insertion of the bZip gene.

A further requirement was the insertion of a stop codon at the end of the bZip gene. This was necessary to ensure that only the bZip sequences would be expressed as a fusion protein with β-galactosidase upon

infection into *E. coli*. The stop codon sequence used was TAG. This was inserted between the bZip sequence and the EcoRI site of the stop primer. The final feature of each primer was the inclusion of a 'GC clamp' which comprised of GC repeats. This feature added stability to the PCR products and also provided an increased region of sequence for the binding of the EcoRI restriction enzyme. The GC clamp was placed at the 5'-end of each primer and removed prior to ligation into λZAPII (Fig.3.3).

The GCN4 gene.

The same criteria used for the design of the primers for amplification of the bZip gene were applied when determining the sequence of the primers for PCR of the GCN4 gene.

The stop primer was identical to that used for the bZip gene while the start primer simply included a sequence which was complementary to the start of the GCN4 gene. The sequences 5' to this region were identical to those used for the bZip region (Fig.3.3).

bZip gene start primer.

5'- CGCGC GAATTC GGTGCCCGAATCC -3'

| GC | EcoRI | bZip |
| clamp | site | sequence |
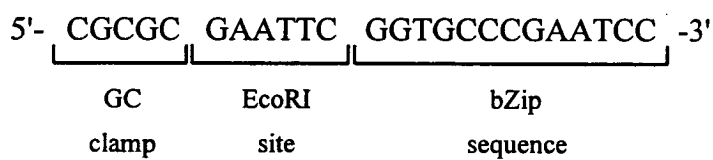
bZip/GCN4 gene stop primer.

5'- GCGCGC GAATTC TAG CGTTCGCCAAC -3'

| GC | EcoRI | stop | bZip |
| clamp | site | codon | sequence |

GCN4 gene start primer.

5'- GCGCGC GAATTC TAGCGTTCGCCAAC -3'

| GC | EcoRI | GCN4 |
| clamp | site | sequence |

Fig.3.3. PCR primers.

BamHI/NdeI

HindIII/ClaI

HindIII/EcoRI cen3

Asp718/HindIII ars1

ori

YCp88-GCN4

amp

ura3

EcoRI

HindIII gcn4
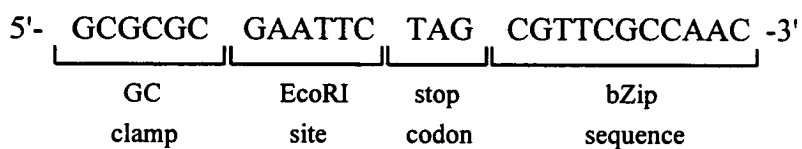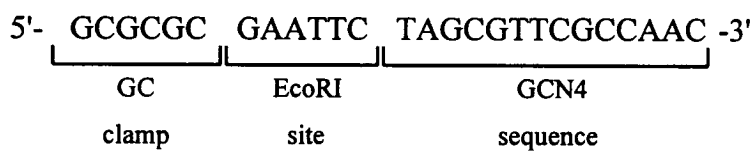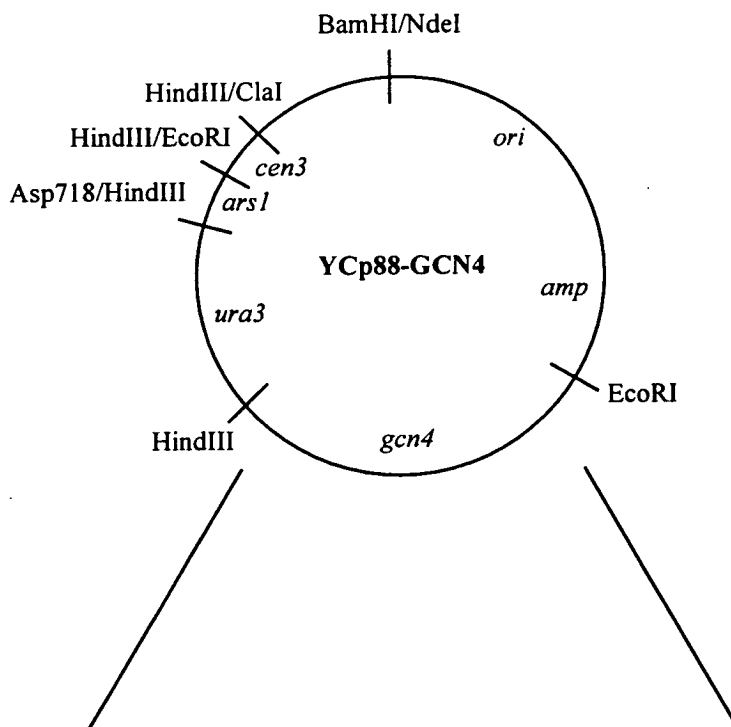
```
ATGTCCGAATATCAGCCAAGTTTATTTGCTTTAAATCCAATG
MetSerGluTyrGlnProSerLeuPheAlaLeuAsnProMet

GGTTTCTCACCATTGGATGGTTCTAAATCAACCAACGAAAATGTATCTCGCTTCCACTTCTACT
GlyPheSerProLeuAspGlySerLysSerThrAsnGluAsnValSerAlaSerThrSerThr

GCCAAACCAATGGTTGGCCAATTGATTTTTGATAAATTCATCAAGACTGAAGAGGATCCAATT
AlaLysProMetValGlyGlnLeuIlePheAspLysPheIleLysThrGluGluAspProIle

ATCAAACAGGATACCCCTTCGAACCTTGATTTTGATTTTGCTCTTCCACAAACGGCAACTGCA
IleLysGlnAspThrProSerAsnLeuAspPheAspPheAlaLeuProGlnThrAlaThrAla

CCTGATGCCAAGACCGTTTTGCCAATTCCGGAGCTAGATGACGCTGTAGTGGAATCTTTCTTT
ProAspAlaLysThrValLeuProIleProGluLeuAspAspAlaValValGluSerPhePhe

TCGTCAAGCACTGATTCAACTCCAATGTTTGAGTATGAAAACCTAGAAGACAACTCTAAAGAA
SerSerSerThrAspSerThrProMetPheGluTyrGluAsnLeuGluAspAsnSerLysGlu

TGGACATCCTTGTTTGACAATGACATTCCAGTTACCACTGACGATGTTTCATTGGCTGATAAG
TrpThrSerLeuPheAspAsnAspIleProValThrThrAspAspValSerLeuAlaAspLys

GCAATTGAATCCACTGAAGAAGTTTCTCTGGTACCATCCAATCTGGAAGTCTCGACAACTTCA
AlaIleGluSerThrGluGluValSerLeuValProSerAsnLeuGluValSerThrThrSer

TTCTTACCCACTCCTGTTCTAGAAGATGCTAAACTGACTCAAACAAGAAAGGTTAAGAAACCA
PheLeuProThrProValLeuGluAspAlaLysLeuThrGlnThrArgLysValLysLysPro

AATTCAGTCGTTAAGAAGTCACATCATGTTGGAAAGGATGACGAATCGAGACTGGATCATCTA
AsnSerValValLysLysSerHisHisValGlyLysAspAspGluSerArgLeuAspHisLeu

GGTGTTGTTGCTTACAACCGCAAACAGCGTTCGATTCCACTTTCTCCAATTGTGCCCGAATCC
GlyValValAlaTyrAsnArgLysGlnArgSerIleProLeuSerProIleValProGluSer

AGTGATCCTGCTGCTCTAAAACGTGCTAGAAACACTGAAGCCGCCAGGCGTTCTCGTGCGAGA
SerAspProAlaAlaLeuLysArgAlaArgAsnThrGluAlaAlaArgArgSerArgAlaArg

AAGTTGCAAAGAATGAAACAACTTGAAGACAAGGTTGAAGAATTGCTTTCGAAAAATTATCAC
LysLeuGlnArgMetLysGlnLeuGluAspLysValGluGluLeuLeuSerLysAsnTyrHis

TTGGAAAATGAGGTTGCCAGATTAAAGAAATTAGTTGGCGAACGCTGA
LeuGluAsnGluValAlaArgLeuLysLysLeuValGlyGluArg***
```

Fig.3.4. The YCp88-GCN4 plasmid.

3.2.4 Preparation of plasmid DNA.

Before attempting amplification of the GCN4 and bZip target sequences, the plasmid containing the GCN4 gene, YCp88-GCN4, was transformed into the *E. coli* strain JM109 via standard methods[111,118] (Fig.3.4). The transformed cells were maintained on ampicillin plates and used to amplify the YCp88-GCN4 plasmid DNA. The plasmid DNA was linearised prior to each PCR by treating it with the restriction enzyme BamHI. Linearising the plasmid DNA in this way minimised the possibility of any tertiary structure which could impede the annealing of the PCR primers to the DNA.

3.2.5 The heating cycle.

The ideal conditions for each of the PCR's were determined by experimentation. The reactants were subjected to an initial heating step at 90°C to ensure complete denaturation of each DNA strand. This step was followed by a cycle of denaturation, annealing and extension. A final annealing step was also included to stabilise the double-stranded product.

3.2.6 Results and discussion.

Agarose gel electrophoresis of the PCR products indicated that each of the reactions had been successful. The amplified bZip DNA could be observed as a single band at approximately 200 base pairs, and the amplified GCN4 gene as a band at 870 base pairs. These sizes compared favourably to the calculated size of each fragment. Faint bands at low molecular weight corresponding to the PCR primers were also observed.

The end-clamps at either end of the amplified genes were removed via digestion with the restriction endonuclease EcoRI. The 'sticky-ended' EcoRI fragments were separated from the end-clamps on LMP agarose gels and each DNA purified from the gel slices.

The removal of the end-clamps was difficult to monitor as the digested and undigested DNA molecules demonstrated extremely similar mobilities, even on high percentage agarose gels. A more reliable method to generate the sticky-ended molecules would have involved the creation of concatemers. This could have been achieved by addition of T4 polynucleotide kinase to the crude PCR products, resulting in phosphorylated molecules. These molecules could then have been ligated together with T4 DNA ligase resulting in long concatenated molecules of high molecular weight. Subsequent treatment with EcoRI would have resulted in sticky-ended molecules all of the same length, but with much smaller molecular weight, making the process easier to monitor via electrophoresis.

3.3 Ligation of the GCN4 and bZip genes into λZAPII.

3.3.1 Introduction.

The goals of this project required that the GCN4 and bZip genes be inserted into a high efficiency vector such that the genes could be selectively expressed and screened for their binding to DNA probes. A vector which ideally suited the purpose of the experiments described here was λZAPII[106,116]. The λZAPII vector contains a unique EcoRI site contained within the *lac*Z gene into which each gene can be cloned. If the gene is cloned within the correct reading frame of the *lac*Z sequence then a fusion protein with β-galactosidase may be expressed (Fig.3.5).

T3   MCS   T7

lacZ

| A ———— J | amp
COLE1 | c1857 |

Terminator   Initiator

1.Ligate bZip gene

2.Screen with DNA probe

T3   bZip gene   T7

lacZ

| A ———— J | amp
COLE1 | c1857 |

Terminator   Initiator

1.Coinfect E.coli with lambda

   and helper phage

T7   EcoRI

f1

bZip

pBluescript SK-

EcoRI

T3
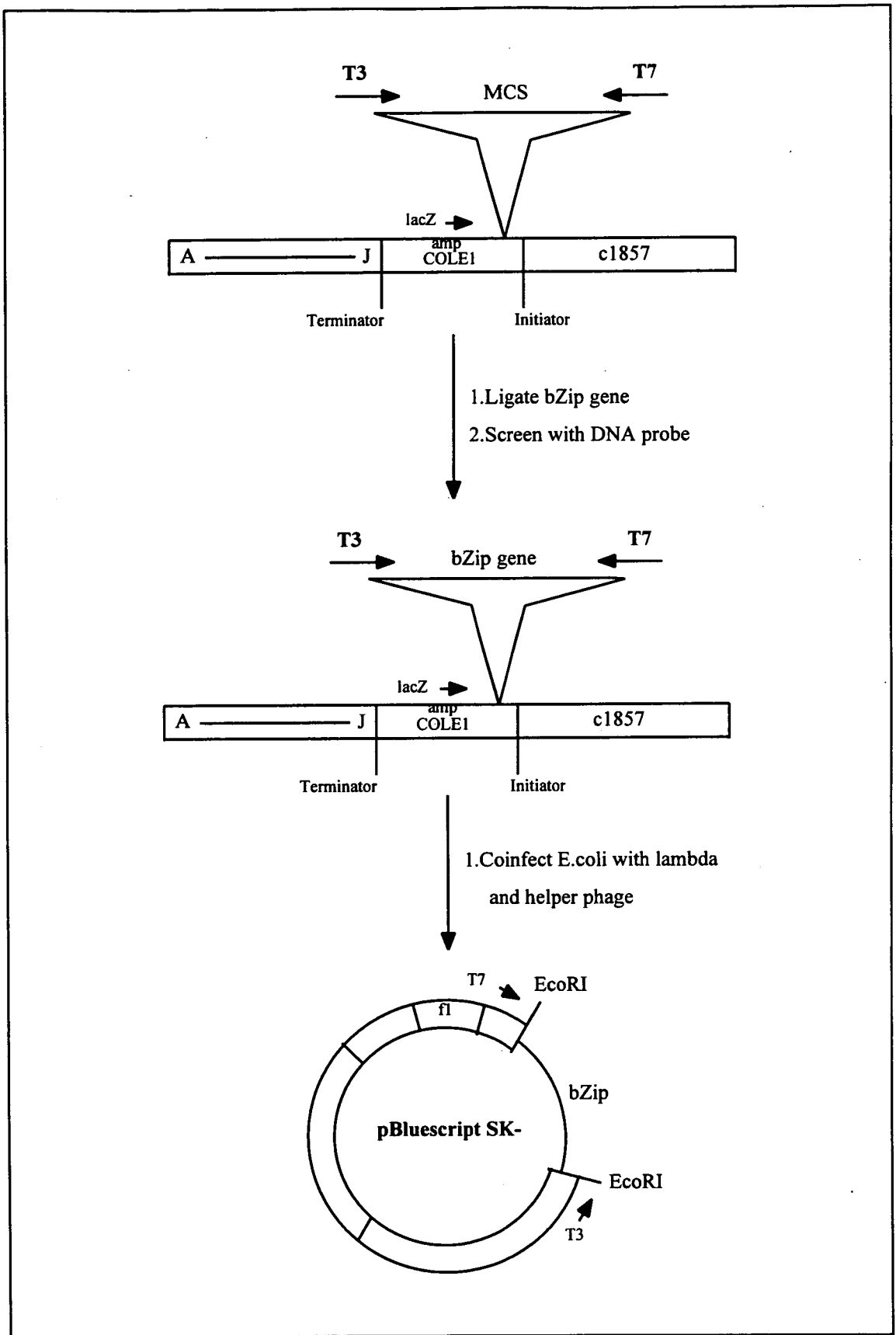
Fig.3.5. The λZAPII vector system.

3.3.2 Packaging of lambda.

Naked bacteriophage λ DNA cannot be transformed into cells directly and must be packaged with the relevant phage proteins prior to transfection[119]. *In vitro* packaging systems have been developed which allow efficient packaging of the λ DNA into phage particles[120,121,122]. These *in vitro* packaging systems usually involve the complementation of the products of two λ lysogens, each carrying a single amber mutation in a distinctly different gene. Neither of the lysogens is capable of packaging phage DNA alone, however, when the extracts are mixed, mature phage particles are produced by complementation. The phage particles are then able to infect *E. coli*.

3.3.3 Transfection of lambda into *E. coli*.

The genetics of the λZAPII bacteriophage require it to be infected into cells which allow correct expression of its genetic elements. The strain of *E. coli* used for initial transfection was XL1-Blue[123]. This strain has a number of features which make it a suitable host for λZAPII. The F' episome present in this strain is of particular importance. It allows blue/white colour selection for non-recombinant and recombinant phage and also *in vivo* excision of the pBluescript SK- phagemid contained within the λZAPII genome[106].

When XL1-Blue cells containing λZAPII clones are grown in the presence of the chromogenic substrate X-gal and the inducer IPTG, the full length β-galactosidase gene gives rise to blue colonies. However, if the *lacZ* gene is disrupted by insertion of foreign DNA, growth in the presence of IPTG and X-gal results in a white colour due to complementation between the

amino-terminal portion of the *lac*Z gene and the *lac*ZΔM15 gene which is located on the F' episome of XL1-Blue. Therefore, if cells containing the λZAPII clones are grown in the presence of IPTG and X-gal, expression of the *lac*Z gene can be monitored by the colour of the bacterial colony. Recombinant phage, which have a disrupted *lac*Z gene, can therefore be isolated as white colonies and any blue colonies containing non-recombinant phage rejected.

3.3.4 Results and discussion.

Prior to ligation, the concentration of each of the DNA sequences amplified by PCR was determined spectrometrically and adjusted such that there was an equimolar ratio of λZAPII arms to insert. An equimolar ratio makes it less likely that concatenated species will form during ligation.

Self ligation was avoided by cloning each of the inserts into λZAPII arms which had been dephosphorylated. T4 DNA ligase will only catalyse the formation of a phosphodiester bond between adjacent nucleotides if one nucleotide contains a 5'-phosphate group and the other a 3'-hydroxyl group, therefore the recombinant λ molecules were nicked. The nicks were repaired by *E. coli* enzymes after transfection.

Each of the ligation reactions was carried out at 4°C to avoid disruption of the base pairs formed at each EcoRI site. λZAPII clones containing the bZip genes were named λbZip, while clones containing the GCN4 gene were named λGCN4.

3.3.5 Transfection.

Prior to transfection, each of the λZAPII clones was packaged into infectious phage particles using the Gigapack II packaging extract (Stratagene).

After transfection into the *E. coli* strain XL1-Blue, the efficiency of each ligation reaction was checked via blue/white colour selection as described above. The cells were grown overnight in the presence of IPTG and X-gal. It was necessary to add a high concentration of X-gal because of the large polylinker that interrupts the *lacZ* gene and reduces the complementing activity of the *lacZ* gene with the *lacZ*ΔM15 gene. Observation of each plate revealed that the efficiency of each of the ligation reactions was significantly different. No blue plaques were observed on the agar plate containing λbZip phage, indicating that the bZip gene had been cloned into the λZAPII arms with a high degree of efficiency. However, only 75% of the plaques grown on the plate containing λGCN4 phage were white. This result indicated that the GCN4 DNA had not ligated into the lambda arms efficiently.

The packaging efficiency was monitored via measurement of the plaque titre of each sample. The results demonstrated that the λbZip DNA had packaged efficiently (2×10^5 pfu/ml). The GCN4 titre (1.5×10^4 pfu/ml), however, was substantially lower. This result reflected the fact that the GCN4 insert had not been cloned efficiently as recombinant molecules are packaged much more efficiently than non-recombinant molecules.

It was decided on the basis of these results that only the λbZip clones should be screened for *in vitro* binding activity of the β-galactosidase-bZip fusion protein.

The purity of the λGCN4 library was investigated by sequencing pBluescript DNA derived from white plaques.

3.3.6 The λGCN4 library.

As described above, initial results indicated that the ligation of the GCN4 gene into λZAPII had not been completely successful.

A number of white plaques were therefore cored from an agar plate and the pBluescript phagemid sequence containing the insert excised from the λ phage using Exassist helper phage (see appendix). After transfection into SOLR cells, single-stranded phagemid DNA was prepared via infection with R408 helper phage. This single-stranded DNA was purified and sequenced.

The results of the dideoxy sequencing showed that none of the samples cored from the plate contained λZAPII with a GCN4 gene in the correct orientation for gene expression. 80% of the phage were shown to contain a GCN4 gene which had been cloned upside down, in the wrong orientation for gene expression, while the remainder contained no DNA insert at all.

The reasons for the unidirectional cloning of the GCN4 gene are not clear, however, there are a number of possible explanations. It is possible that the region of the DNA at the beginning of the GCN4 gene may have been folded in such a manner which disfavoured ligation. It is also possible that the foreign DNA sequences of the GCN4 gene were simply not tolerated by the lacZ sequences within the multiple cloning site. Indeed, this has been shown to occur when cloning foreign DNA into filamentous bacteriophage vectors[111]. The unidirectional cloning may also have been due to incomplete removal of the GC end-clamp prior to ligation, although this is unlikely.

The presence of non-recombinants in the phagemid samples derived from the white coloured core samples may be attributed to host bacteria which have lost the F' episome through mutation. Subsequent pBluescript phagemids derived from λZAPII were therefore infected into SOLR cells and grown on ampicillin/kanamycin plates. This combination of antibiotics allowed for selection of both the phagemid and the F' episome.

3.4 *In vitro* detection of bZip clones.

3.4.1 Introduction.

Previous studies have demonstrated that it is possible to identify DNA-binding proteins *in vitro* by screening a library of lambda clones[124,125]. The method involves infecting a suitable strain of *E. coli* with a lambda library. The cells are then grown on an agar plate until the bacterial plaques are just visible before overlaying the surface of the agar with a nitrocellulose filter impregnated with IPTG. The overexpressed fusion proteins and cell debris stick to the nitrocellulose. These are then screened for their binding characteristics with a radioactively labelled DNA probe. Bacterial plaques expressing proteins which bind to the DNA probe can subsequently be identified via autoradiography.

In vitro methods for identifying DNA-binding proteins have a number of advantages over *in vivo* methods and other *in vitro* techniques:

(i) The use of radioactively labelled concatenated probes allows weakly binding proteins to be detected more readily.

(ii) Screening of mutants is straight forward.

(iii) No subcloning is required, therefore, identification of clones is swift.

(iv) No protein purification is required.

(v) A large library of mutants can be screened at once.

3.4.2 The λbZip library.

The bZip gene was cloned within the *lacZ* gene of the bacteriophage lambda vector λZAPII as described in section 3.3. This allowed expression

of the protein as a fusion with β-galactosidase upon induction with IPTG. λZAPII contains the phagemid sequence pBluscript SK- which can be excised *in vivo* (see appendix). This can be converted to single-stranded DNA which can be used for sequencing and *in vitro* mutagenesis. Any 'positive' clones can therefore be identified easily and rapidly.

3.4.3 The DNA probe.

The process of *in vitro* screening requires a radioactively labelled DNA probe. In this case, two complementary synthetic oligonucleotides containing concatenated AP-1 sites were synthesised. The oligonucleotides were annealed together and ligated into the phagemid vector pTZ19R using standard methods[108,111] (Fig.3.6). The recombinant plasmid DNA, pOP1, was transformed into TG1 cells and sequenced to ensure correct ligation had occurred. These cells were used for large scale preparations of plasmid DNA. The OP1 sequence containing the AP-1 binding sites was excised via treatment with the relevant restriction enzymes. The OP1 DNA was then purified and labelled with the appropriate $[\alpha\text{-}^{32}P]$dNTP.

Using DNA probes with concatenated binding sites has been shown to increase the signal produced upon screening *in vitro*, therefore, two AP-1 binding sites were included within the 70bp OP1 probe[124]. An SfiI site was included between each of the AP-1 sites (Fig.3.7). This opened up the possibility of inserting DNA sequences between the binding sites should the protein be unable to bind due to the close proximity of the AP-1 binding sites. Cloning the OP1 DNA into pTZ19R allowed the OP1 fragment to be isolated as a fragment bearing different restriction sites at

either end. This allowed the fragment to be labelled preferentially at either end of the molecule with various [α-^{32}P]dNTP's.
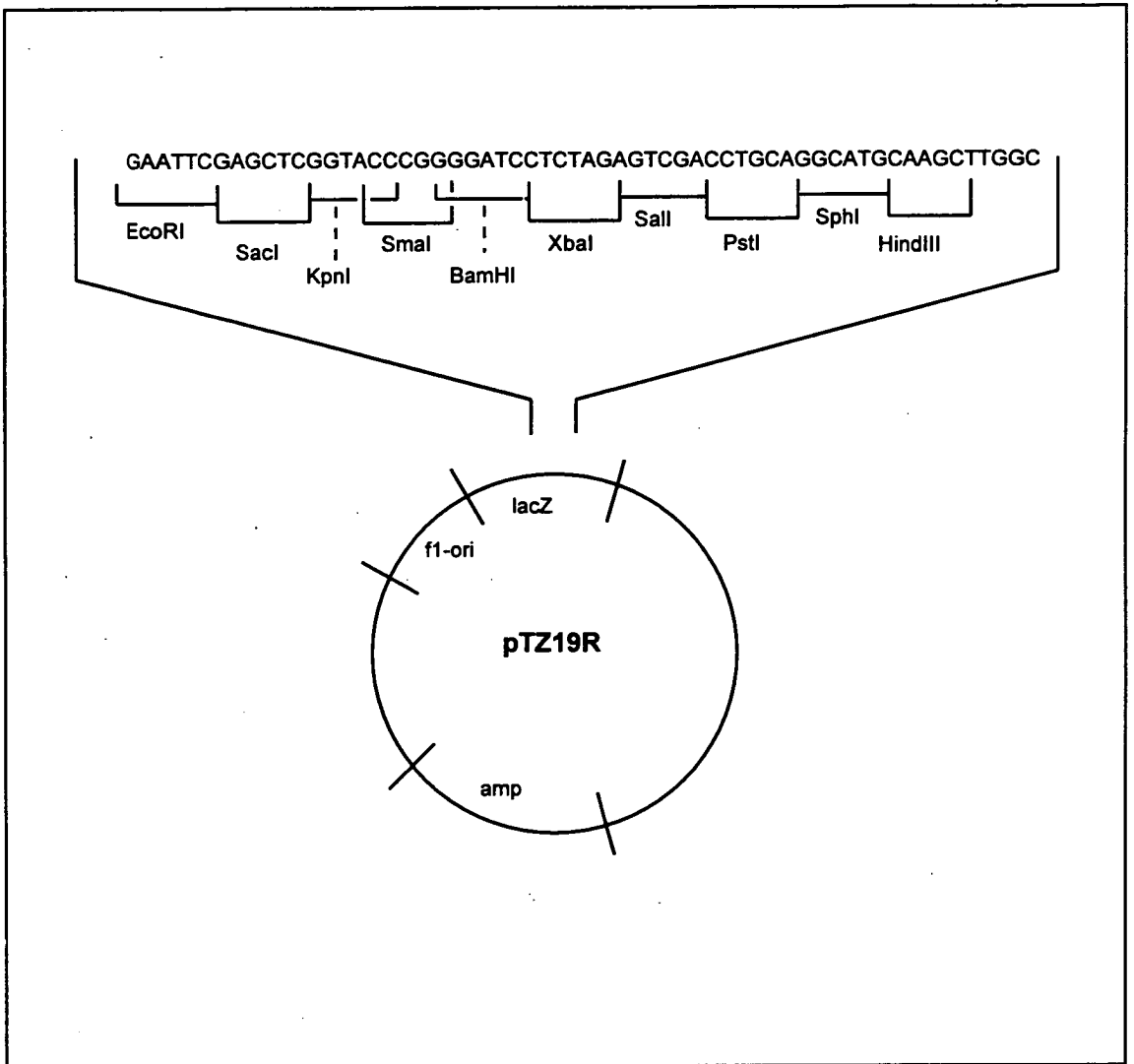
Fig.3.6. The pTZ19R plasmid.

```
                         AP-1
5'- ATTCCACAGTGCCT  ┌ATGACTCAT┐CCAGTGGCCT
3'-     GGTGCTACGGC │TACTGAGTA│GGTGACCGGA
                    └─────────┘
                         AP-1
   TTTTGGCCCCT  ┌ATGACTCAT┐CCAGTGGGGATCC  -3'
   AAAACCGGGGA  │TACTGAGTA│GGTCACCCCTAGG  -5'
               └─────────┘
```
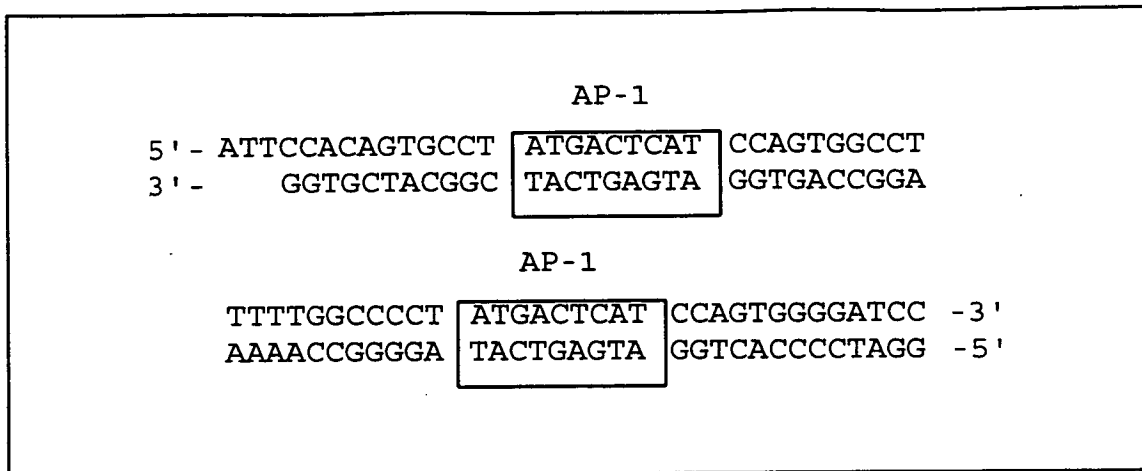
Fig.3.7. The OP1 probe.

3.4.4 Screening.

Before screening with radioactively labelled DNA the plaque titre was adjusted such that each of the bacterial plaques was well separated from its neighbours on an agar plate. Each of the plates was then incubated in the absence of IPTG until the plaques were just visible. The plates were removed and nitrocellulose filters impregnated with IPTG laid over the developing plaques. The plates were incubated for a further 3-4 hours before removing the filters.

The ideal conditions for the *in vitro* screening assay were determined by experimenting with denaturation and renaturation of the fusion protein bound to the nitrocellulose filters. Primary screening was carried out with and without denaturation/renaturation with guanidinium hydrochloride. The results indicated that the protein was bound to the nitrocellulose in such a way which hampered binding to DNA. Subsequent experiments were therefore carried out with denaturation and stepwise renaturation of the bound protein.

The λbZip library was plated in XL1-Blue cells as described above. After denaturation/renaturation with guanidinium hydrochloride the filters

carrying the fusion proteins were incubated with ^{32}P labelled OP1 DNA. Salmon sperm DNA was also included to act as a control for non-specific binding.

3.4.5 Identification of positive clones.

After autoradiography approximately 50% of the plaques were visible as black dots on a piece of film (Fig.3.8). Each of the positive plaques was cored from the agar plate and the phage particles lysed from the agar. The pBluescript SK- sequence containing the bZip gene was excised from each phage sample using the Exassist helper phage (see appendix). The phagemids were then transformed into the *E. coli* strain SOLR and maintained on ampicillin/kanamycin plates. Single-stranded DNA was then prepared for sequencing via infection with R408 helper phage. Single-stranded rescue of the pBluescript SK- phagemid produces the non-coding (-) strand packaged with its coat proteins.

Before sequencing each of the single stranded DNA samples, a short oligonucleotide primer was synthesised. The sequence of the primer was chosen such that it annealed to a sequence 102 bases upstream of the cloned gene.

Each of the samples was sequenced using the dideoxy method developed by Sanger *et al*[114].

The results of the sequencing experiments showed that all of the single-stranded DNA samples contained a bZip gene in the correct orientation for expression of a β-galactosidase-bZip protein. These results indicated that the screening method used to identify DNA-binding proteins was reliable and could be used in future experiments to identify novel DNA-binding proteins.
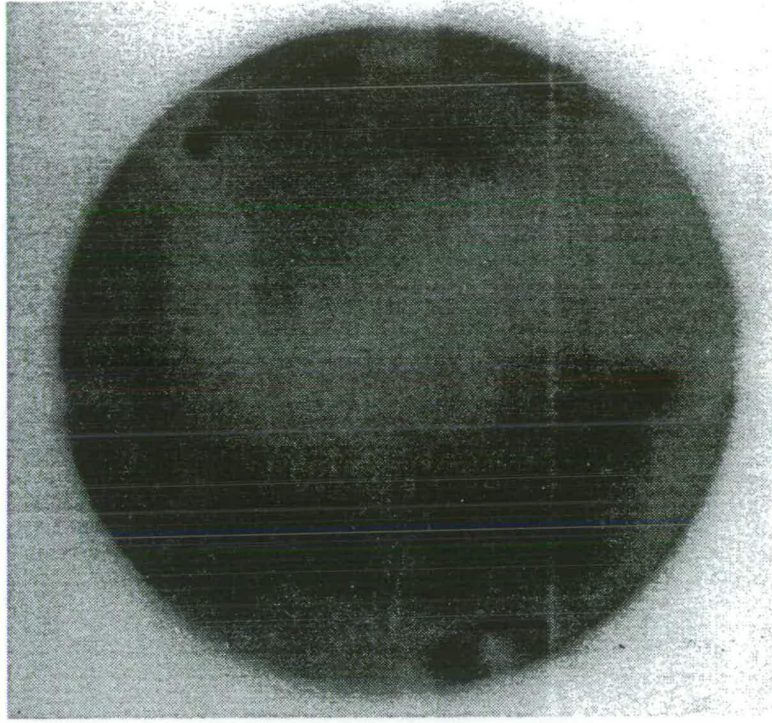
Fig.3.8. Autoradiograph of *in vitro* screen.

3.5 Overexpression of β-galactosidase-bZip fusion protein.

3.5.1 Introduction.

Since the bZip DNA had been cloned into a unique EcoRI site within the *lac*Z gene of pBluescript SK-, it was possible to overexpress the bZip protein as a fusion with β-galactosidase[106]. The *lac*Z gene, which codes for β-galactosidase, is transcribed via induction from the *lac* promoter. The *lac* promoter is regulated by a helix-turn-helix protein called the *lac* repressor. To abolish the repression of the *lac* promoter by the repressor protein, IPTG, which is an inducer of β-galactosidase, may be added to the cell extract. IPTG forms a complex with the *lac* repressor protein thus interrupting binding of the protein to the promoter DNA and inducing transcription of the *lac*Z gene. The β-galactosidase-bZip fusion protein can therefore be produced faster than other cellular proteins simply by the addition of IPTG of a suitable concentration (Fig.3.9).

```
ATG    ACC    ATG    ATT    ACG    CCA    AGC
met    thr    met    ile    thr    pro    ser

TCG    AAA    TTA    ACC    CTC    ACT    AAA
ser    lys    leu    thr    leu    thr    lys

GGG    AAC    AAA    AGC    TGG    AGC    TCC
gly    asn    lys    ser    trp    ser    ser

ACC    GCG    GTG    GCG    GCC    GCT    CTA
thr    ala    val    ala    ala    ala    leu

GAA    CTA    GTG    GAT    CCC    CCG    GGC
glu    leu    val    asp    pro    pro    gly

TGC    AGG    AAT    TCG
cys    arg    asn    ser

GTG    CCC    GAA    TCC    AGT    GAT    CCT
val    pro    glu    ser    ser    asp    pro

GCT    GCT    CTA    AAA    CGT    GCT    AGA
ala    ala    leu    lys    arg    ala    arg

AAC    ACT    GAA    GCC    GCC    AGG    CGT
asn    thr    glu    ala    ala    arg    arg

TCT    CGT    GCG    AGA    AAG    TTG    CAA
ser    arg    ala    arg    lys    leu    gln

AGA    ATG    AAA    CAA    CTT    GAA    GAC
arg    met    lys    gln    leu    glu    asp

AAG    GTT    GAA    GAA    TTG    CTT    TCG
lys    val    glu    glu    leu    leu    ser

AAA    AAT    TAT    CAC    TTG    GAA    AAT
lys    asn    tyr    his    leu    glu    asn

GAG    GTT    GCC    AGA    TTA    AAG    AAA
glu    val    ala    arg    leu    lys    lys

TTA    GTT    GGC    GAA    CGC
leu    val    gly    glu    arg
```

Fig.3.9. β-galactosidase-bZip fusion protein.

111

3.5.2 Protein overexpression.

A λZAPII clone containing the bZIP gene in the correct orientation for protein expression was identified as described previously and the pBluescript SK- phagemid, pbZip, containing the gene excised *in vivo*. After infecting into SOLR cells glycerol stocks of the culture were prepared and maintained at -70°C.

The overexpression of the fusion protein was investigated initially by growing the cells containing the phagemid in the presence of ^{35}S-labelled methionine and IPTG. After induction with IPTG aliquots were removed from the broth at prescribed time intervals and analysed via SDS-PAGE. Autoradiography of the radioactive gel revealed that the correct protein was being overexpressed relative to the other cellular proteins and that the mass of the protein was approximately 11kD. This result compared favourably with the calculated mass of 11155D.

Further experiments indentified the ideal conditions for the growth of the cells and induction with IPTG. These conditions were used to prepare large amounts of cell free extract containing the overexpressed fusion protein.

3.5.3 Protein purification.

Partial purification of the bZip fusion was acheived by first lysing the cells via sonication and isolating the cell free extract via centrifugation. DNA was removed via treatment with DNase I. The protein sample was then fractionated with ammonium sulphate. Experiments revealed that fractionation of the cell free extract with ammonium sulphate to an initial concentration of 40% followed by a 60% step resulted in

substantial removal of high molecular weight cellular proteins from the extract.

The relatively small size and poor overexpression of the bZip fusion protein excluded various methods which could be used for purification from the cell free extract. The initial purification step was therefore carried out via size exclusion chromatography using a Sephadex G-50 column. This allowed separation of the high molecular weight proteins from the lower molecular weight proteins (Fig.3.10). However, the major drawbacks of this technique were the low resolution, low loading capacity of the column and the subsequent dilution of the protein sample. The latter problem was overcome by concentrating the filtrate by ultrafitration through a 3000 molecular weight cut-off membrane, although this method was time consuming due to the low molecular weights involved.

The protein extract was not purified further. However, further purification could be achieved by ion-exchange chromatography. The β-galactosidase-bZip fusion protein is highly basic therefore purification by cation-exchange chromatography should be possible. Purification could also be achieved via affinity chromatography. However, this method would require the construction of an affinity matrix containing DNA covalently linked to a suitable support.
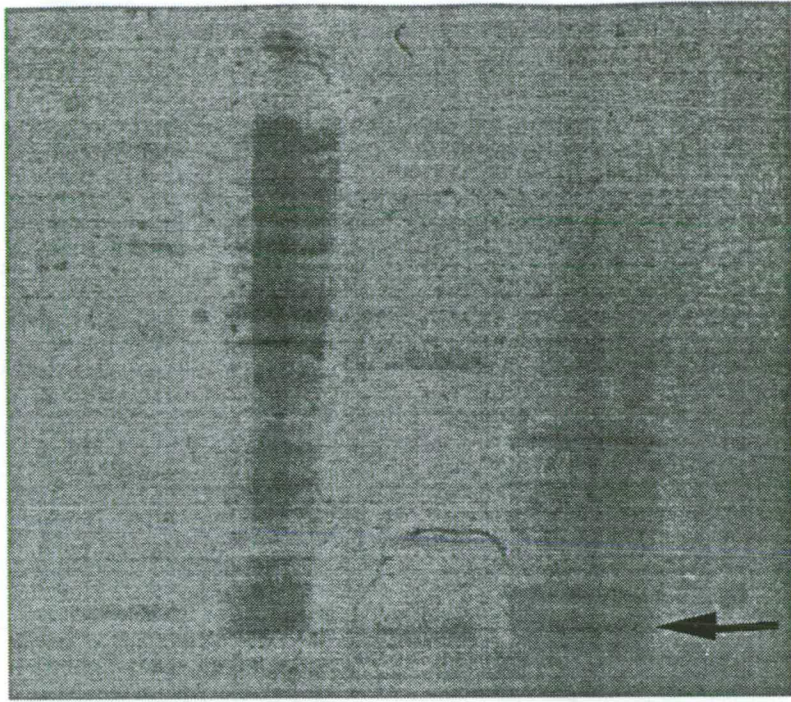
Fig.3.10. SDS-PAGE gel of protein purification.

3.6 Gel retardation assay.

3.6.1 Introduction.

A standard method for investigating the binding characteristics of DNA binding proteins is the gel retardation assay[126,127]. The technique involves incubating a protein sample with a radioactively labelled DNA probe containing a binding site and separating the bound complexes from the unbound DNA via electrophoresis. DNA molecules with protein bound are retarded relative to free DNA molecules due to the increased charge and molecular weight of the protein-DNA complex.

3.6.2 The DNA probe.

Before carrying out the assay, pOP1 DNA was digested with EcoRI and XbaI restriction endonucleases. The OP1 fragment was purified and labelled preferentially at one end with $[\alpha\text{-}^{32}P]CTP$ by the action of the Klenow fragment of DNA polymerase I.

3.6.3 Characterisation of protein-DNA complexes.

Partially purified extracts containing the bZip fusion protein were incubated with the ^{32}P-labelled DNA probe. Dissociation of protein-DNA complexes was minimised by incubating at $4^{o}C$ and separating the protein-DNA complexes from the free DNA at high voltage via electrophoresis on a native polyacrylamide.

Initial experiments revealed that the pH of the gel and running buffer should be adjusted to pH 7.5 to avoid aggregation of the protein in the

wells of the gel. It is likely that the aggregation of the protein in the wells was caused by precipitation of cellular proteins in the crude sample. The isoelectric point (pI) for the β-galactosidase-bZip fusion protein was calculated as 10.9 and it is therefore likely that the precipitation of the fusion protein may have been caused by aggregated cellular proteins at pH 8.3, the original pH of the electrophoresis buffer. Experiments also showed that the mobility of the fusion protein was sufficiently different to that of the probe to retard the DNA. If the mobility of the protein was too near to that of the free DNA then no retardation would have been observed.

The results of initial gel retardation experiments indicated the bZip fusion bound to the OP1 probe with high affinity. It could be observed that the occupation of both the AP-1 sites increased as the protein concentration was increased (Fig.3.11). However, at high protein concentration the band corresponding to the protein-DNA complex was promoted relative to that observed at lower protein concentration. The reason for this anomalous behaviour is not clear. The promoted band is likely to correspond to OP1 bound by two protein dimers. The binding of two protein dimers to the DNA probe would result in increased positive charge and size which, under the conditions of the experiment, would be expected to lead to increased retardation of the DNA probe. The increased mobility of the protein-DNA band may therefore reflect bending of the DNA which may have occurred as a result of the binding of two fusion protein dimers. Each of the retarded bands could be distinguished by carrying out DNaseI footprinting experiments across a range of protein concentrations.

protein–
DNA complex

unbound DNA

increasing protein concentration
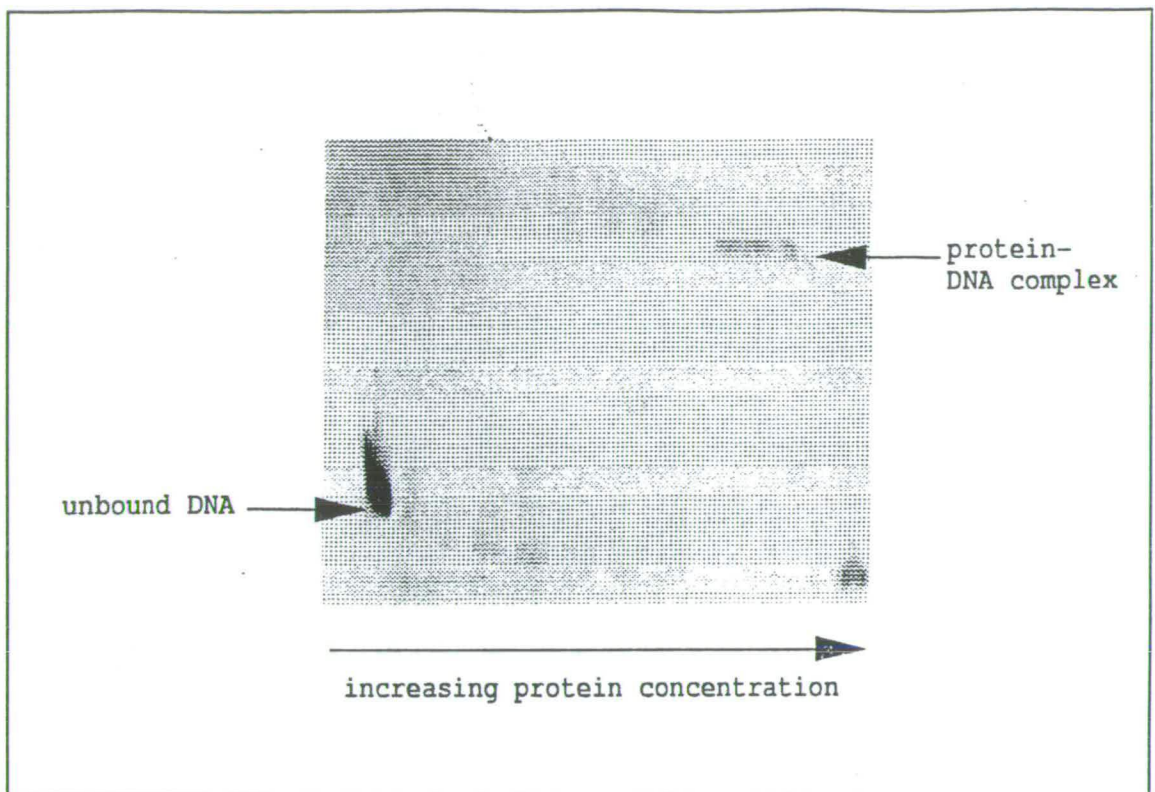
Fig.3.11. Gel retardation assay.

3.6.4 The DNA-binding affinity.

Gel retardation assays can be used to obtain an estimate of the DNA-binding constant of a protein. This is achieved by estimating the concentration of protein at which half the DNA is bound.. The estimation is only valid if the DNA concentration is at least 10 to 100-fold lower than that of the protein[127].

The concentration of the fusion protein extract was adjusted such that a wide range of concentrations were obtained. The protein samples were incubated with OP1 probe DNA and the complexes separated from unbound DNA via electrophoresis.

The binding constant, Kd, was estimated as ca. 10^{-10}M from an auroradiogram which showed that the concentration of protein required to bind half the DNA was approximately 2.12×10^{-10}M. This calculation

was based on the assumption that both AP-1 binding sites were independent of each other and that there was no cooperativity between proteins bound to each site. The estimated value for the DNA-binding constant of the fusion protein compared favourably with that calculated previously for the binding of GCN4 to the *his3* site $(10^{-10}M)$[84].

It must be stressed that the calculated value for Kd was only an estimate and its accuracy was dependent upon the measurement of protein concentration. The crude protein concentration was ascertained by Bradford's assay and the concentration of β-galactosidase-bZip fusion protein in the crude mixture estimated from SDS-PAGE gels. A more accurate determination of binding constant would require purification of the fusion protein to homogeneity. Accurate quantitation of the amount of radioactivity in each of the gel bands would also be required. This could be achieved by film densitometry or scintillation counting of excised DNA bands.

3.7 Conclusions.

It is clear from the experimental work described throughout this chapter that λZAPII is an excellent vector for the cloning, identification and expression of the cloned bZip gene.

The results of the *in vitro* screening process demonstrated that it is possible to screen a large library of recombinant phage and identify tightly binding proteins. These results indicated that the λZAPII system could be used to express a large library of mutants which could be screened in an identical manner to that employed for the λbZip library. The results also showed that expression of the bZip protein with β-galactosidase did not disrupt specific binding of the protein to target DNA.

Cloning of the bZip gene into λZAPII also allowed effective excision of the pBluescript SK- phagemid which was used to overexpress the β-galactosidase-bZip fusion protein. The overexpression, however, was poor but may have been improved by addition of a higher concentration of IPTG.

The gel retardation experiments demonstrated that the fusion protein was able to bind to DNA containing tandem AP-1 sites with high affinity. The mobility of the doubly bound probe, however, was anomalous.

CHAPTER 4

Mutagenesis

4.1 Mutagenesis of GCN4 basic region.

4.1.1 Background.

As outlined in Chapter 1, previous studies investigating the binding characteristics of mutants of the GCN4 basic region and linker segment have been carried out[89,99,100,101]. The majority of these studies have utilised site-directed mutagenesis to mutate specific residues and have focussed mainly on the residues involved in contact with the DNA bases. Numerous single and double mutants, which have changed specificity have been generated, although the effects of randomly mutating a section of the basic region have only been partially investigated. Kim *et al* used a degenerate oligonucleotide with an average of two base-pair substitutions to randomise a region between thr-236 and lys-246. Several single and double mutants which were able to bind wild-type and symmetrically mutated AP-1 sites were isolated[89]. The results indicated that residues not involved in DNA-base contact could influence specificity in combination with other base-contacting residues.

As yet, a thorough study investigating the effects of mutating a large section of the basic region has not been completed. Previous evidence has indicated that mutating residues in combination can influence the DNA-binding specificity of a mutant protein[101,102,103]. For example, mutating residues on the same face of the basic region helix can alter specificity, while other combinations of mutants may also influence specificity[105]. Although mutations in the linker segment have been shown to affect binding specificity, mutation of basic region residues in combination with mutations in the linker segment have not been investigated and these may also alter the binding specificity significantly. Mutations in either

segment may also distort the basic region α-helix such that the pattern of DNA contacts may be changed. The effects of disparate combinations of mutations across the basic region and linker segment may also be extremely significant in changing the specificity of a mutant protein. It would therefore be valuable to randomly mutate a large section of the GCN4 basic region.

Numerous techniques for the generation of random mutations in DNA have been devised. These include *in vivo* methods such as the use of mutator strains, *in vitro* enzymatic approaches, and procedures that use chemicals to damage the DNA[128,129,130]. However, it is known that mutagenesis by these procedures may result in an uneven distribution of mutations. A method which ensures complete randomness in the region of interest is the spiked oligonucleotide mutagenesis approach[131]. This procedure makes use of the phosphorothioate method for mutagenesis[132].

4.1.2 The phosphorothioate method for mutagenesis.

The phosphorothioate method for oligonucleotide directed mutagenesis of DNA overcomes the problems associated with techniques which require the transfection of heteroduplex species and which ultimately lead to low mutational efficiency[132]. In this procedure, a mismatched oligonucleotide primer is synthesised and annealed to a single-stranded DNA template containing the gene of interest. The primer DNA is extended by a polymerisation reaction in which one of the natural deoxynucleotide triphosphates is replaced by the corresponding deoxynucleotide 5'-O-(1-thiophosphate). Phosphorothioate groups are incorporated into one strand of the newly synthesised RF-IV DNA, resulting in a strand asymmetry which may be exploited.

Certain restriction enzymes cannot hydrolyse a phosphorothioate linkage, therefore, reaction of such DNA with a suitable restriction enzyme produces a nick in the strand[133,134]. The nick in the strand is extended into the gap by reaction with a suitable exonuclease which digests away the wild-type sequence opposite the mismatch introduced by the primer. On repolymerisation, the gapped DNA is repaired using the mutant-carrying strand as the template. The mutant sequence is now present in both strands of the DNA as a fully complementary homoduplex species. This DNA can be transfected with high mutational frequencies[132].

4.1.3 Spiked mutagenesis.

The spiked oligonucleotide mutagenesis procedure can be used to introduce random mutations across a large region of a gene (Fig.4.1)[131,135]. In this method, long oligonucleotide primers which span the region of interest are synthesised by a protocol that gives a defined probability of incorporating one of the three 'wrong' nucleotide bases at each position along the oligonucleotide. The spiked primer is then phosphorylated and annealed to a single-stranded DNA template. The phosphorothioate procedure for *in vitro* mutagenesis is then followed and a collection of mutants obtained that contain changes spanning the whole length of the primer. The mutants can then be transfected into cells and the transformants screened for their ability to bind DNA probes.
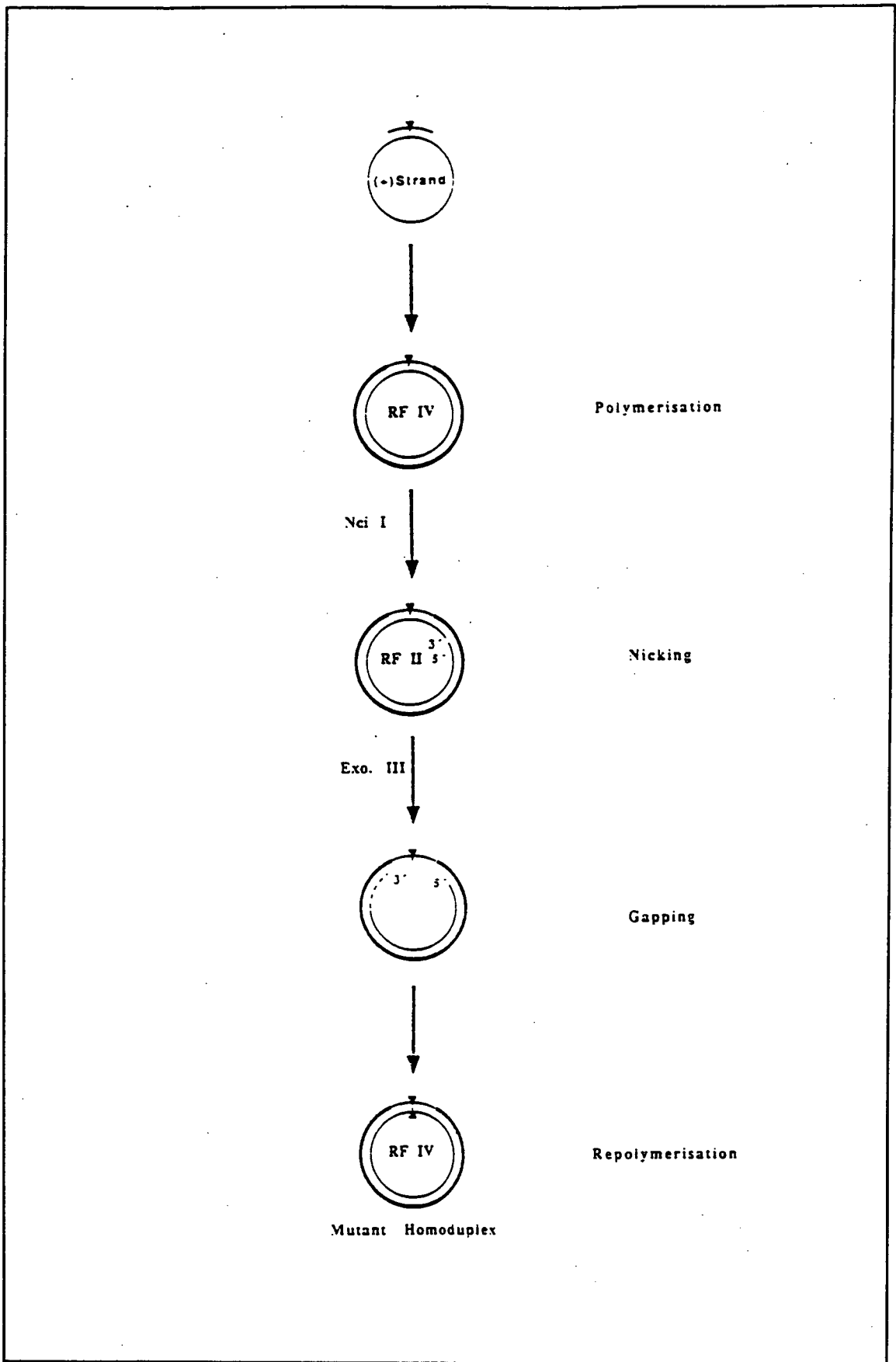
Fig.4.1. Phosphorothioate based spiked mutagenesis.

4.1.4 Mutagenesis of the GCN4 basic region.

As mentioned above, no studies have been carried out which have subjected a large section of the GCN4 basic region to random mutagenesis. The object of this project was to randomly mutate a large area spanning the basic region and linker segment via spiked mutagenesis and detect any single, double or triple mutants which bound to DNA containing an AP-1 site. Screening of the mutant library in an identical manner to that described in Chapter 3 should detect any novel mutants, as well as those previously characterised by other procedures.

It was decided to randomly mutate a segment of the GCN4 bZip gene equivalent to the region between arg-234 and asp-255 (Fig.4.2). This area covers all residues involved in contacting the DNA bases, as well as the majority of those making contacts to the phosphate backbone[59]. Residues in the linker segment, which comprises the residues between arg-245 and asp-255, would also be mutated. These residues have been shown to be particularly important in determining half-site preferences[89].

```
    230     235     240     245     250     255
    AALKRARNTEAARRSRARKLQRMKQLEDKVEE
```
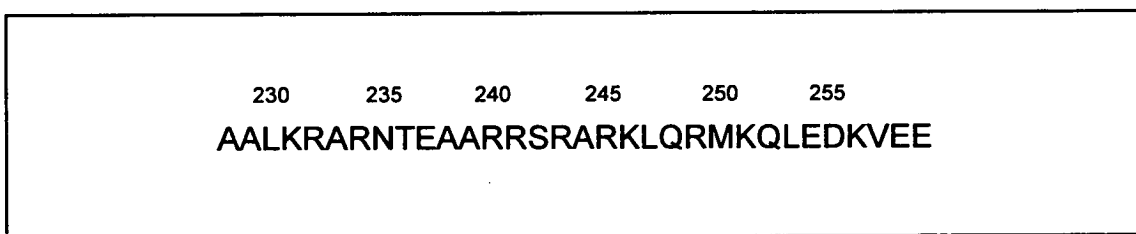
Fig.4.2. The GCN4 basic region and linker segment.

4.1.5 Design of the mutagenic primer.

A 76-base spiked oligonucleotide primer corresponding to the region between arg-234 and asp-255 was synthesised. The oligonucleotide, which

was phosphorylated chemically at the 5'-end, contained two small 5-base unspiked regions at the 5'- and 3'-ends to ensure efficient annealing to the single stranded template DNA. The 66-base spiked region was synthesised such that each primer had a high probability of containing a single or double mutation. This was acheived by spiking each phosphoramidite used for the synthesis of the oligonucleotide with a calculated percentage of the three 'wrong' phosphoramidites. Since a binomial distribution is observed for the number of mutations per oligonucleotide it was calculated that each phosphoramidite should be spiked with 0.83% of each of the wrong phosphoramidites to maximise the number of single and double mutations in each primer molecule.

The probability P of finding n errors in an m-long oligonucleotide that is synthesised with a fraction α of the three wrong nucleotides at each position can be calculated using the following formula[131]:

$$P(n,m,\alpha) = [m!/(m-n)!n!][\alpha]^n[1-\alpha]^{m-n}$$

Therefore, the probability of a spiked primer containing one mutation at any position is approximately 31%, while the probability of a primer containing two mutations is approximately 9%.

If the population of mutated DNA molecules are transformed efficiently then 130000 transformants will cover >99% of the possible one-base changes and approximately 30% of the possible two-base changes.

4.1.6 Mutagenesis.

The phagemid, pbZip, which contains the bZip region of GCN4 was converted to a single-stranded form via infection with a helper phage (see

appendix). The purified (-) strand DNA was then used as a template for the spiked mutagenesis procedure.

The phosphorothioate mutagenesis procedure is a multistep process. The first stage of the procedure involves annealing of a mutagenic oligonucleotide primer to a single-stranded DNA template. The mutagenic strand is then polymerised by the action of Klenow fragment and the advancing 3'-end ligated to the 5'-end by the action of T4 DNA ligase in the presence of ATP. The template strand is then nicked by a suitable restriction enzyme and the template strand digested by the action of exonuclease III. The final step of the process involves repolymerisation of the template strand with DNA polymerase I. The mutated gene can then be excised and cloned into λZAPII, allowing screening of mutant proteins after transfection into *E. coli*.

4.1.7 Preparation of RF-IV DNA.

Before carrying out the polymerisation step, the concentration of both the template DNA and the primer DNA were determined spectrometrically and adjusted such that the template:primer ratio was approximately 1:2. This ratio was necessary to avoid non-specific annealing which can impede polymerisation. The annealing step was then carried out by incubating the template/primer mixture successively at 70°C, 55°C and 37°C to ensure efficient annealing. The ratio of Klenow fragment to T4 DNA ligase was adjusted to avoid strand displacement synthesis which can occur if a large excess of Klenow fragment is present in the reaction mixture. This was minimised further by carrying out the reaction at 16°C.

Initial attempts to follow the phosphorothioate method failed at the polymerisation stage. Agarose gel electrophoresis of the products revealed that the mutagenic primer annealed to the pbZip template efficiently and that polymerisation had taken place. However, no ligation occurred as indicated by the presence of a strong band corresponding to nicked RF-II DNA. Subsequent experiments revealed that the absence of ligation was not due to inferior T4 DNA ligase or unsuitable buffer conditions, but was likely to be due to the oligonucleotide primer. Since the primer DNA had been phosphorylated synthetically, it was thought that the absence of ligation may have been due to inefficient phosphorylation. However, attempts to rephosphorylate the primer DNA with T4 polynucleotide kinase proved unsuccessful.

Attempts to purify the oligonucleotide by denaturing polyacrylamide gel electrophoresis revealed that the DNA was highly impure and consisted of numerous smaller fragments. The presence of these fragments may have been due to truncated failure sequences or as a consequence of DNA degradation which may have been caused by the presence of harmful chemical impurities. The failure sequences present in the DNA mixture may have been removed by HPLC if the oligonucleotide had not been phosphorylated, however, phosphorylated oligonucleotides cannot be purified in this manner. It was decided, therefore, to discard the oligonucleotide primer and synthesise a fresh, highly purified one.

The new primer was synthesised with an identical sequence to that used previously. In this case the phosphorylated DNA was purified via capillary electrophoresis. The DNA was stored in aliquots at -20°C and all manipulations carried out under sterile conditions to avoid any biological contamination.

The annealing and polymerisation reactions were carried out in an identical manner to those above. Agarose electrophoresis of the products revealed that the polymerisation reaction had proceeded with high efficiency and that approximately 80% of the DNA was present as closed circular RF-IV DNA.

4.1.8 Strand specific nicking of RF-IV DNA.

The second stage of the phosphorothioate procedure for mutagenesis involves the nicking of the template DNA strand. The pbZip phagemid sequence contains six NciI restriction sites at positions outwith that covered by the mutagenic primer. It was therefore decided to nick the (-) strand with the NciI enzyme[136]. After incubation in the presence of the enzyme, agarose gel electrophoresis revealed that a single band corresponding to nicked RF-II DNA was present and that the reaction had gone to completion.

4.1.9 The gapping reaction.

The penultimate stage of the mutagenesis procedure involves digestion of the nicked (-) strand by a suitable exonuclease[137]. Exonuclease III is an enzyme which digests a free 3'-terminus in the 3' to 5' direction and was ideal for the purpose of these experiments. Initial attempts at the gapping reaction indicated that the exonuclease III was unable to digest the nicked DNA strand. Exonuclease III is known to be very sensitive to buffer conditions, therefore, the reaction was attempted in a number of different buffers. However, despite varying the buffer conditions no gapping of the (-) strand could be observed.

Test experiments in a variety of buffers revealed that linear plasmid derived DNA could be digested completely over a period of 3 hours, indicating that the enzyme used was active and that the inactivity of the enzyme in the mutagenesis procedure must have been due to an inhibitor present in the nicked DNA mixture. Purification of the nicked DNA by a variety of methods failed to improve the efficiency of the reaction, therefore, T7 exonuclease, which is more tolerant to variations in buffer concentration was used.

Initial results of reaction with T7 exonuclease indicated that the nicked DNA had been digested relatively efficiently. However, subsequent attempts to generate RF-IV DNA via polymerisation of the mutagenic primer proved impossible, despite identical reaction conditions to those used previously.

The purity of the primer DNA was therefore checked by labelling the 5'-end with [γ-^{32}P]ATP via incubation with T4 polynucleotide kinase. Denaturing polyacrylamide electrophoresis revealed that the primer DNA had undergone decomposition into several smaller fragments. The reason for this is unclear as the DNA was stored at -20°C in small aliquots. It is likely that the degradation of the DNA was a result of destructive impurities present in the storage buffer. These impurities may also have been responsible for inhibiting the action of exonuclease III in the gapping reaction.

It was decided to attempt the mutagenesis procedure with a third mutagenic primer which was not chemically phosphorylated. In this case, the primer sequence was not identical to that of the previous two. Financial constraints meant that a shorter completely random primer, spanning the region between ala-233 and met-250, be synthesised. Although this primer was shorter than the previous two, the area of the

basic region mediating DNA contact as well as part of the linker segment were covered by the primer.

Before carrying out the polymerisation an aliquot of the oligonucleotide primer was labelled at the 5'-end with [γ-^{32}P]ATP as above. Denaturing polyacrylamide electrophoresis revealed that the DNA was pure and could be observed as a single band after autoradiography.

Prior to attempting the polymerisation stage, the spiked primer was phosphorylated by the action of T4 polynucleotide kinase in the presence of ATP. The polymerisation reaction was then carried out in an identical manner to that used previously. However, agarose gel electrophoresis revealed that no RF-IV DNA had been produced. A band at high molecular weight could be observed indicating that strand displacement synthesis may have occurred (Fig.4.3). It is unlikely that this was caused by an excess of primer DNA as the concentration of both primer and template were adjusted to avoid this.
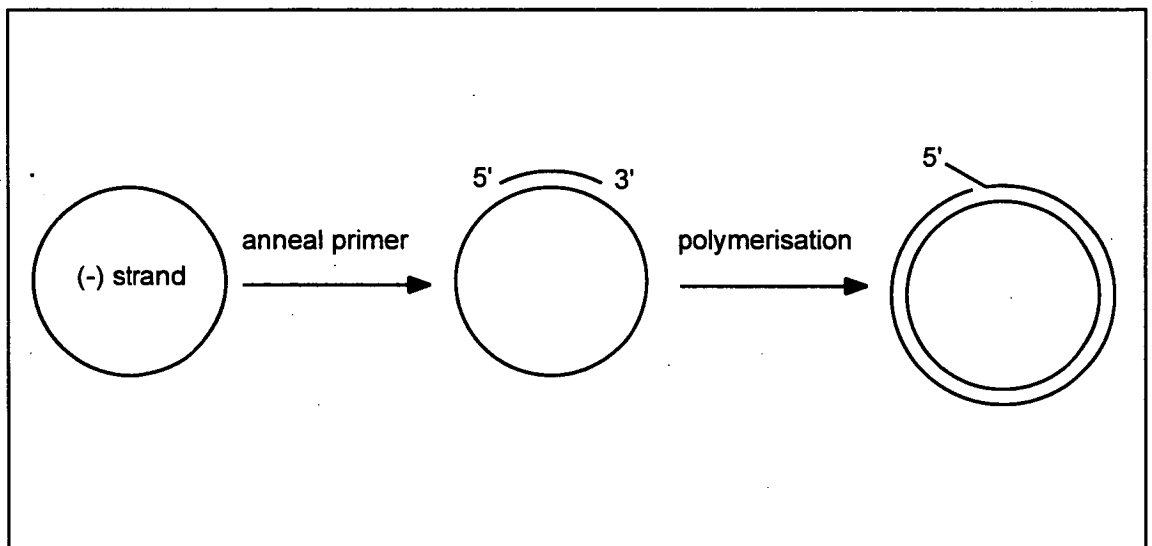
Fig.4.3. Strand displacement synthesis.

Further experiments investigating the reason for the lack of polymerisation suggested that the template DNA was responsible. The

single-stranded template DNA was therefore sequenced to check for the presence of mutations which may have led to incorrect annealing. The dideoxy sequencing experiment revealed that a deletion of two guanines in the centre of the region covered by the spiked primer had occurred (see appendix). The presence of this deletion in the single-stranded template would disrupt the annealing of the primer such that an overhang would be created. Subsequent polymerisation of the primer partially annealed to the mutated template DNA would, therefore, result in strand diplacement synthesis making ligation of 3'- and 5'-ends impossible.

Although sequencing of the single-stranded pbZip template revealed that a deletion in the bZip gene had occurred, the question remained as to whether this had resulted before or after infection of the cells containing the phagemid with helper phage. The cells from a glycerol stock were therefore plated and used to prepare a large scale culture of pbZip double-stranded DNA. The double-stranded DNA was sequenced using the dideoxy method which revealed that exactly the same deletion of two guanines in the centre of the basic region had occurred.

The reliability of the phosphorothioate procedure as a technique for *in vitro* mutagenesis was evaluated by carrying out test reactions using a control template and primer from an *in vitro* mutagenesis kit. Each step of the procedure was monitored by agarose electrophoresis which demonstrated that the procedure was reliable and efficient.
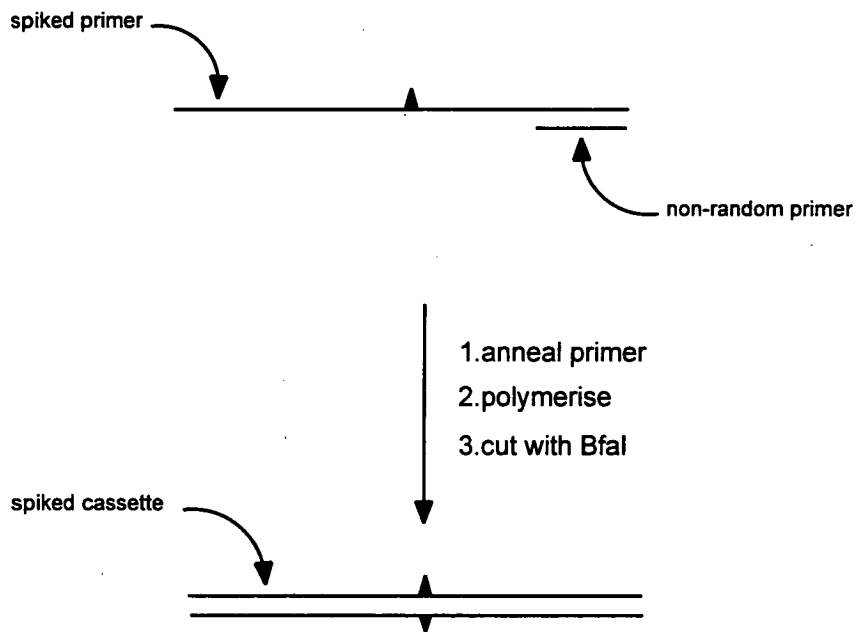
Other methods for generating the mutant library were considered, and these are discussed below. However, financial and time constraints precluded the implementation of these procedures.

4.1.10 Alternative procedures for mutagenesis.

Two viable alternative methods of generating the random mutants were considered. The first of these involved the recloning of the entire GCN4 bZip gene into pBluescript SK-. This could be acheived by first treating the pbZip phagemid DNA with EcoRI restriction enzyme which would excise the pbZip gene. The phagemid DNA could then be separated from the bZip DNA on a LMP agarose gel and the pBluescript phagemid DNA purified. The bZip gene, isolated previously by PCR, could then be ligated into the phagemid and transformed into SOLR cells. After sequencing to isolate a correct clone the spiked mutagenesis procedure could be followed as described above.

A second possibility of generating the random mutations involves cloning of the spiked primer DNA as a cassette into the pbZip phagemid (Fig.4.4). Computer analysis of the bZip region DNA sequence revealed that two unique restriction sites, BbsI and BfaI, not found in pBluescript SK- were present. Closer analysis revealed that the BfaI site lay within the region covered by the spiked primer, while the BbsI site was 3' to the spiked region.
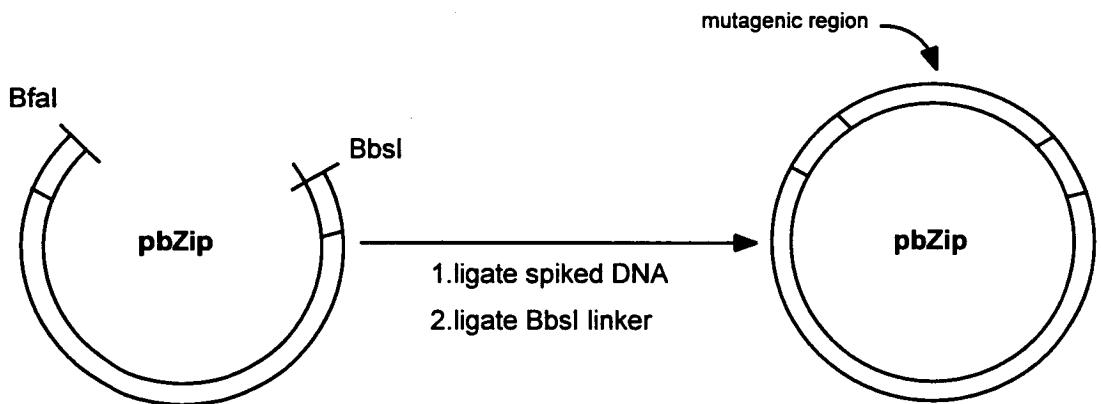
Fig.4.4. Cassette mutagenesis of pbZip.

This method for creating mutants relies on the spiked primer being converted to a double-stranded form before ligation. This could be achieved via mutually primed synthesis using a DNA primer which anneals to approximately 15 bases at the 3'-end of the primer. The BfaI site could be created by digestion with BfaI restriction enzyme and the double-stranded spiked fragment ligated into pbZip which had been digested with BfaI and BbsI. A double stranded linker segment would also have to be included in the ligation reaction to fill in the gap produced. The recombinant plasmid containing the mutated region could then be treated with EcoRI and XhoI to excise the mutated bZip gene. After purification this could be cloned into λZAPII and the mutants screened for their ability to bind DNA probes *in vitro*.

In addition to the time involved, this procedure has a number of disadvantages associated with it. The random nature of a 15-base segment corresponding to residues 245 to 250 will be lost. Also, cloning such small DNA fragments may be difficult and inefficient leading to a reduction in the mutational frequency. Finally, the BfaI site in the spiked primer may be difficult to cut due to the small amount of sequence at this end of the primer.

Numerous other procedures to mutate DNA have been developed. However, relatively few of these techniques allow for random mutation across a large region of DNA. Two viable alternatives to the phosphorothioate method do exist.

The first of these techniques makes use of PCR and the concept of overlap extension for introducing mutations into a region of a gene (Fig.4.5)[138,139]. The procedure requires the synthesis of four oligonucleotide primers; two mutagenic primers which cover the region of interest, and two non-random primers which complement either end of

the gene. Three separate PCR steps are then carried out with the final reaction producing the mutated gene. The gene is excised with the relevant restriction enzymes and cloned into the appropriate vector.

This method would be ideal for the mutagenesis of the GCN4 basic region. After carrying out the PCR steps the mutated gene could be excised via treatment with EcoRI and XhoI. The DNA fragment could then be directionally cloned into λZAPII and the mutants screened.

A second method for the generation of random mutations involves the use of enzymes[140]. The procedure requires transformation of the phagemid containing the gene of interest into a *dut⁻ ung⁻* strain of *E. coli* which allows production of uracil-containing single-stranded DNA. A DNA primer is then annealed upstream of the region to be mutated and four separate polymerisation reactions carried out, each with a limiting concentration of one of the four natural dNTP's. This step generates truncated fragments and is followed by a misincorporation step in which the three wrong bases are incorporated via polymerisation with reverse transcriptase. The mismatched plasmid library is then transformed into a *dut⁺ ung⁺* strain of *E. coli.*

This method has numerous disadvantages associated with it which makes it unsuitable for the mutagenesis of the GCN4 basic region. Firstly, only single mutations can be generated. Secondly, the length of the mutagenic region cannot be controlled. Finally, the procedure requires re-transformation and a subsequent cloning step which would reduce the number of mutants significantly.
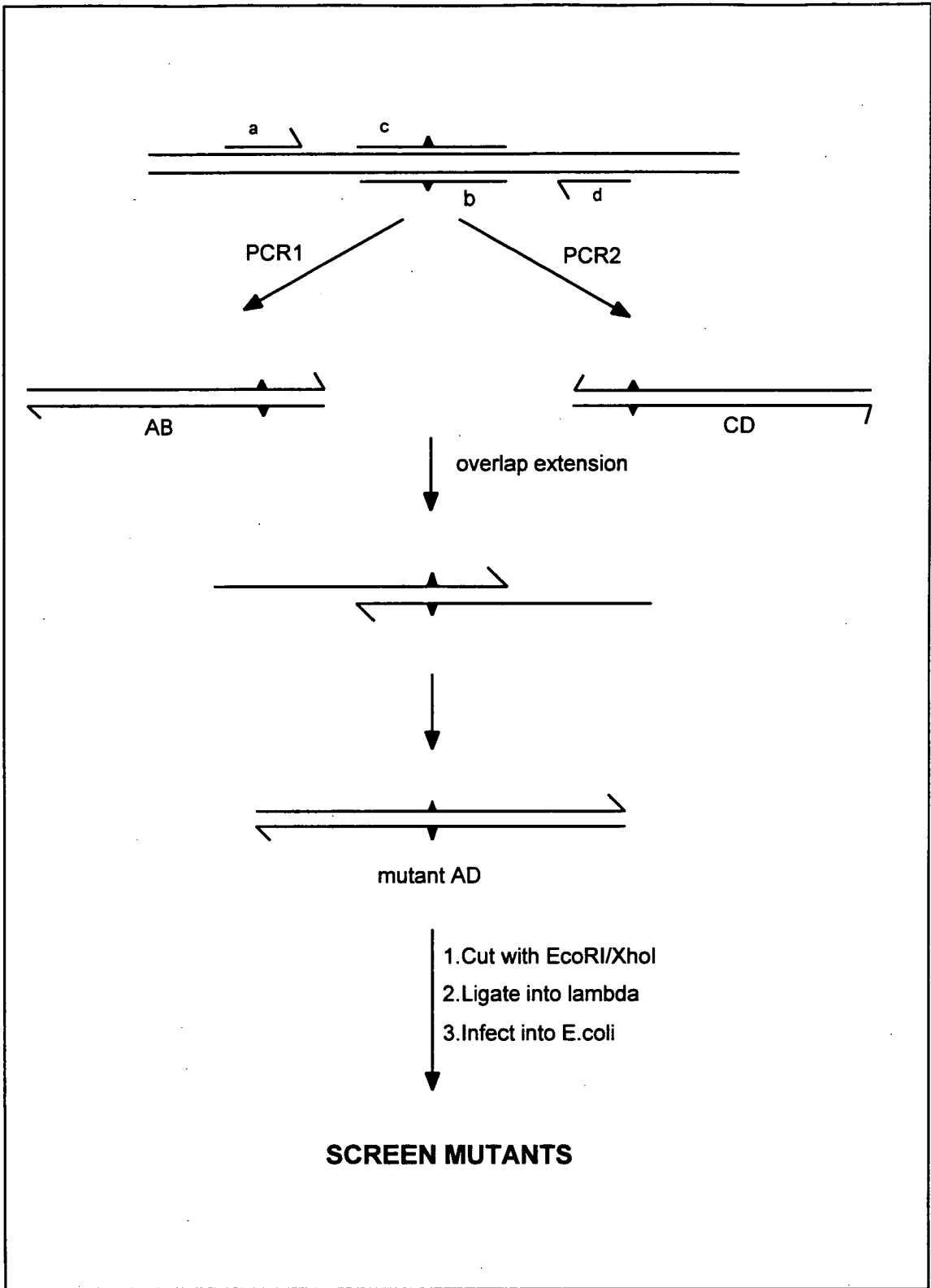
Fig.4.5. The overlap extension method for mutagenesis.

4.1.11 Future work.

Assuming the random mutations in the GCN4 basic region have been created the next stage of the project would involve cloning of the mutant DNA library into λZAPII. This could be acheived by directionally cloning into λZAPII arms which had been previously digested with EcoRI and XhoI. However, it may be necessary to 'rescue' the DNA prior to cloning by transforming the pbZIP mutant library into SOLR cells. This is dependent upon the ability of EcoRI and XhoI to hydrolyse phosphorothioate linkages. The mutant (+) strand contains only one phosphorothioate base, namely dCTPαS. Since EcoRI hydrolyses the linkage between guanine and thymine, the plasmid will be linearised. However, XhoI hydrolyses the linkage between cytosine and thymine. It is therefore possible that only the (-) strand will be nicked making direct cloning imposssible and rescue necessary. If this occurs then the mutational efficiency may be reduced significantly. The mutational efficiency of the procedure can be monitored readliy by sequencing 10-20 phagemids derived from the λZAPII library.

Once the mutant DNA library has been cloned into λZAPII screening with radioactive DNA probes containing the AP-1 site can proceed. The ideal conditions for these experiments have been determined previously. Care must be taken in adjusting the plaque titre such that the maximum number of single and double mutants can be screened. If we assume a high mutational efficiency then screening 130000 plaques should cover 99% of all single mutations and 30% of all double mutations (see section 4.1.5).

Identification and characterisation of mutants should be acheived initially by screening with DNA containing with the wild-type AP-1 site.

However, subsequent screening experiments should involve incubation with symmetrically mutated AP-1 probes as well as a probe containing the ATF/CREB site to detect any mutants with altered half-site specificity.

Prior to further characterisation of mutant proteins they should be overexpressed in *E. coli* as described previously. However, previous work has suggested that efficient overexpression may be difficult to acheive making purification difficult. An alternative to overexpression would involve production of the mutant proteins via *in vitro* translation. pBluescript SK- contains T7 and T3 polymerase binding sites which flank the cloned gene. These can be used to produce RNA transcripts from which protein can be translated *in vitro*. This method would therefore produce pure protein which could also be labelled *in vitro* using ^{35}S-methionine.

The binding of mutant proteins should be investigated further via gel retardation assay. These should be carried out with a range of DNA probes and the binding constant calculated.

Characterisation of protein-DNA interactions could be acheived by modelling studies. The X-ray structure of the GCN4-AP-1 complex has been solved and this could be modified accordingly to model interactions between DNA bases and amino acid side-chains. Modelling studies may provide valuable information which could simplify the rules for DNA-binding of a bZip protein.

Finally, a succesful conclusion to this work could lead to the investigation of other DNA-binding proteins in a similar manner.

REFERENCES

1. Sawadago, M., Senetac, A. (1990) *Annu. Rev. Biochem.* **59**, 711-754.

2. Struhl, K. (1987) *Cell* **49**, 295-297.

3. Buratowski, S. (1994) *Cell* **77**, 1-3.

4. Roedar, R. G. (1991) *TIBS* **16**, 402-408.

5. Tijian, R., Maniatis, T. (1994) *Cell* **77**, 5-8.

6. Burley, S. K. (1994) *Curr. Opin. Struct. Biol.* **4**, 3-11.

7. Hope, I. A., Mahadevan, S., Struhl, K. (1988) *Nature* **333**, 635-640.

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

9. Johnson. P. F., McKnight, S. L. (1989) *Annu. Rev. Biochem.* **58**, 799-839.

10. Pabo, C. O., Sauer, R. T. (1984) *Annu. Rev. Biochem.* **53**, 293-321.

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

12. Brennan, R. G. (1991) *Curr. Opin. Struct. Biol.* **1**, 80-88.

13. Brennan, R. G. (1992) *Curr. Opin. Struct. Biol.* **2**, 100-108.

14. Anderson, W. F., Ohlendorf, D. H., Takeda, Y., Matthews, B. W. (1981) *Nature* **290**, 744-749.

15. Brennan, R. G., Roderick, S. L., Takeda, Y., Matthews, B. W. (1990) *Proc. Natl. Acad. Sci.* **87**, 8165-8169.

16. Jordan, S. R., Pabo, C. O. (1988) *Science* **242**, 893-899.

17. Pabo, C. O., Lewis, M. L. (1982) *Nature* **298**, 443-447.

18. Otwinowski, Z., Schevitz, R. W., Zhang, R., Lawson, C. L., Joachimak, A., Marmostein, R. Q., Luisi, B. F., Sigler, P. B. (1988) *Nature* **335**, 321-329.

19. Laughon, A. (1991) *Biochemistry* **30**, 11357-11367.

20. Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B., Pabo, C. O. (1990) *Cell* **63**, 579-590.

21. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., Pabo, C. O. (1991) *Cell* **67**, 517-528.

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

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

24. Hard, T., Kellenbach, E., Boelens, R., Maler, B. A., Dahlman, K., Freedman, L. P., Carlsted-Duke, J., Yamamoto, K. R., Gustafsson, J., Kaptein, R. (1990) *Science* **249**, 157-160.

25. Marmorstein, R., Carey, M., Patshne, M., Harrison, S. C. (1992) *Nature* **356**, 408-414.

26. O' Neil, K. T., Hoess, R. H., Degrado, W. F. (1990) *Science* **249**, 774-778.

27. Ellenberger, T. (1994) *Curr. Opin. Struct. Biol.* **4**, 12-21.

28. Pathak, D., Sigler, P. B. (1992) *Curr. Opin. Struct. Biol.* **2**, 116-123.

29. Kerppola, T. K., Curran, T. (1991) *Curr. Opin. Struct. Biol.* **1**, 71-79.

30. Alber, T. (1992) *Curr. Opin. Struct. Biol.* **2**, 205-210.

31. Landschultz, W. H., Johnson, P. F., McKnight, S. L. (1988) *Science* **240**, 1759-1764.

32. Kouzarides, T. Ziff, E. (1988) *Nature* **336**, 646-651.

33. Neuberg, M., Schuermann, M., Hunter, J. B., Muller, R. (1989) *Nature* **338**, 589-590.

34. Sellers, J. W., Struhl, K. (1989) *Nature* **341**, 74-76.

35. Agre, P., Johnson, P. F., McKnight, S. L. (1989) *Science* **246**, 922-925.

36. Crick, F. H. C. (1953) *Acta Crystallogr.* **6**, 689-697.

37. Gentz, R., Rauscher , F. J., Abate, C., Curran, T. (1989) *Science* **243**, 1695-1699.

38. O' Shea, E. K., Rutkowski, R., Kim, P. S. (1989) *Science* **243**, 1695-1699.

39. Oas, T. G., McIntosh, L. P., O' Shea, E. K., Dahlquist, F. W. (1990) *Biochemistry* **29**, 2891-2894.

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

41. Rasmussen, R., Benvegnu, D., O' Shea, E. K., Kim, P. S., Alber, T. (1991) *Proc. Natl. Acad. Sci.* **88**, 561-564.

42. van Heeckeren, W. J., Sellers, J. W., Struhl. K. (1992) *Nucleic Acids Res.* **20**, 3721-3724.

43. Hu, J. C., O' Shea, E. K., Kim, P. S., Sauer, R. T. (1990) *Science* **250**, 1400-1403.

44. O' Shea, E. K., Rutkowski, R., Kinm, P. S. (1992) *Cell* **68**, 699-708.

45. John, M., Briand, J. P., Granger-Schnarr, M., Schnarr, M. (1994) *J. Biol. Chem.* **269**, 16247-16253.

46. Schuermann, M., Hunter, J. B., Hennig, J. B., Hennig, G., Müller, R. (1991) *Nucleic Acids Res* **19**, 739-746.

47. Hu, J. C., Newell, N. E., Tidor, B., Sauer, R. T. (1993) *Protein Science* **2**, 1072-1084.

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

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

50. Krylov, D., Mikhailenko, I., Vinson, C. R. (1994) *EMBO J.* **13**, 2849-2861.

51. O' Shea, E. K., Lumb, K. J., Kim, P. S. (1993) *Curr. Biol.* **3**, 658-667.

52. Pu, W. T., Struhl, K. (1993) *Nucleic Acids Res.* **21**, 4348-4355.

53. Monera, O. D., Kay, C. M., Hodges, R. S. (1994) *Biochemistry* **33**, 3862-3871.

54. Alberti, S., Oehler, S., von Wilcken-Bergmann, B., Muller-Hill, B. (1993) *EMBO, J.* **12**, 3227-3236.

55. Harbury, P. B., Zhang, T., Kim, P. S., Alber, T. (1993) *Science* **262**, 1401-1407.

56. Harbury, P. B., Kim, P. S., Alber, T. (1994) *Nature* **371**, 80-83.

57. Vinson, C. R., Sigler, P. B., McKnight, S. L. (1989) *Science* **246**, 911-916.

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

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

60. Konig, P. Richmond, T. J. (1993) *J. Mol. Biol.* **233**, 139-154.

61. Oakley, M. G., Dervan, P. B. (1990) *Science* **248**, 847-850.

62. Nye, J. A., Graves, B. J. (1990) *Proc. Natl. Acad. Sci.* **87**, 3992-3996.

63. Shuman, J. D., Vinson, C. R., McKnight, S. L. (1990) *Science* **249**, 771-774.

64. Saudeck, V., Pasley, H. S., Gibson, T., Gausepohl, H., Frank, R., Pastore, A. (1991) *Biochemistry* **30**, 1310-1317.

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

66. Patel, L., Abate, C., Curran, T. (1990) *Nature* **347**, 572-575.

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

68. Hai, T., Lui, F., Allegretto, E. A., Karin, M., Green, M. R. (1988) *Genes Dev.* **2**, 1216-1226.

69. Angel, P., Allegretto, E. A., Okino, S. T., Hattori, K., Boyle, W. J., Hunter, T., Karin, M. (1988) *Nature* **332**, 166-171.

70. Gartenberg, M. R., Ampe, C., Steitz, T. A., Crothers, D. M. (1990) *Proc. Natl. Acad. Sci.* **87**, 6034-6038.

71. Oliphant, A., Brandle, C. J., Struhl, K. (1989) *Mol. Cell. Biol.* **9**, 2944-2949.

72. Sellers, J. W., Vincent, A. C., Struhl, K. (1990) *Mol. Cell. Biol.* **10**, 5077-5086.

73. Travers, A. A. (1992) *Curr. Opin. Struct. Biol.* **2**, 71-77.

74. Nekludova, L., Pabo, C. O. (1994) *Proc. Natl. Acad. Sci.* 91, 6948-6952.

75. Paolella, D. N., Palmer, R. C., Schepartz, A. (1994) *Science* 264, 1130-1133.

76. Kerppola, T. K., Curran, T. (1993) *Mol. Cell. Biol.* 13, 5479-5489.

77. Kerppola, T. K., Curran, T. (1991) *Science* 254, 1210-1214.

78. Kerppola, T. K., Curran, T. (1991) *Cell* 66, 317-326.

79. de Groot, P. P., Delwas, V., Sassone-Corsi, P. (1994) *Oncogene* 9, 463-468.

80. Hinnebusch, A. G. (1988) *Microbiol. Rev.* 52, 248-273.

81. Hinnebusch, A. G. (1990) *TIBS* 15, 148-152.

82. Hinnebusch, A. G. (1984) *Proc. Natl. Acad. Sci.* 81, 6442-6446.

83. Thireos, G., Driscoll Penn, M., Greer, H. (1984) *Proc. Natl. Acad. Sci.* 81, 5096-5100.

84. Hope, I. A., Struhl., K. (1986) *Cell* 46, 885-894.

85. Hope, I. A., Mahadevan, S., Struhl, K. (1989) *Nature* 333, 635-640.

86. Brandl, C. J., Struhl, K. (1989) *Proc. Natl. Acad. Sci.* 86, 2652-2656.

87. Kouzarides, T., Ziff, E. (1989) *Nature* 340, 568-571.

88. Pu, W. T., Struhl, K. (1991) *Proc. Natl. Acad. Sci.* 88, 6901-6905.

89. Kim, J., Tzamerias, D., Ellenberger, T., Harrison, S. C., Struhl, K. (1993) *Proc. Natl. Acad. Sci.* 90, 4513-4517.

90. Blatter, E. E., Ebright, Y. W., Ebright, R. H. (1992) *Nature* 359, 650-652.

91. Arndt, K., Fink, G. R. (1986) *Proc. Natl. Acad. Sci.* 83, 8516-8520.

92. Hill, D.E., Hope, I. A., Macke, J. P., Struhl, K. (1986) *Science* 234, 451-457.

93. Hope, I. A., Struhl, K. (1985) *Cell* 43, 177-188.

94. Mavrothalissitis, G., Beal, G., Papas, T. S. (1990) *DNA Cell Biol.* **9**, 783-788.

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

96. Talanian, R. V., McKnight, C. J., Rutkowski, R., Kim, P. S. (1992) *Biochemistry* **31**, 6871-6875.

97. Cuenod, B., Schepartz, A. (1993) *Science* **259**, 510-513.

98. Cuenod, B., Schepartz, A. (1993) *Proc. Natl. Acad. Sci.* **90**, 1154-1159.

99. Pu, W. T., Struhl, K. (1991) *Mol. Cell. Biol.* **11**, 4918-4926.

100. Pu, W. T., Struhl, K. (1992) *Proc. Natl. Acad. Sci.* **89**, 2007-2011.

101. Suckow, M., Schwamborn, K., Kisters-Woike, B., von Wilcken-Bergmann, B., Müller-Hill, B. (1994) *Nucleic Acids Res.* **22**, 2198-2208.

102. Suckow, M., Madan, A., Kisters-Woike, B., von Wilcken-Bergmann, B., Müller-Hill, B. (1994) *Nucleic Acids Res.* **22**, 2198-2208.

103. Suckow, M., von Wilcken-Bergmann, B., Müller-Hill, B. (1993) *Nucleic Acids Res.* **21**, 2081-2086.

104. Suckow, M., von Wilcken-Bergmann, B., Müller-Hill, B. (1993) *EMBO J.* **12**, 1193-1200.

105. Johnson, P. F. (1993) *Mol. Cell. Biol.* **13**, 6919-6930.

106. Short, J. M., Sorge, J. A. (1992) *Methods Enz.* **216**, 495-508.

107. Alting-Mees, M. A., Sorge, J. A., Short, J. M. (1992) *Methods Enz.* **216**, 483-495.

108. Mead, D. A. (1986) *Protein Engineering* **1**, 67.

109. Vieira, J., Messing, J. (1982) *Gene* **19**, 259-268.

110. Yanisch-Perron, C., Vieira, J., Messing, J. (1985) *Gene* **33**, 103-119.

111. Maniatis, T., Fritsch, E. F., Sambrook, J. (1989) *Molecular cloning: A Laboratory Mmanual.* Cold Spring Harbour Laboratory, Cold Spring Harbour, New York.

112. Willis, E. H., Mardis, E. R., Jones, W. L., Little, M. C. (1990) *Biotechniques* **9**, 92-99.

113. Tautz, D., Renz, M. (1983) *Anal. Biochem.* **132**, 14.

114. Sanger, F., Nicklen, S., Coulson, A. R. (1975) *Proc. Natl. Acad. Sci.* **74**, 5463-5467.

115. Bradford, M. M., (1976) *Anal. Biochem.* **72**, 248-254.

116. Short, J. M., Fernandez, J. M., Sorge, J. A., Huse, W. D. (1988) *Nucleic Acids Res.* **16**, 7583-7600.

117. Mullis, K., Faloona, F. (1987) *Methods Enz.* **155**, 335-350.

118. Reha-Krantz, L. J., (1985) *Gene* **38**, 275.

119. Chautiwale, V. M., Therwath, A., Deshpande, V.V. (1992) *Microbiol. Rev.* **56**, 577-591.

120. Becker, A., Murialdo, H., Gold, M. (1977) *Virology* **78**, 277-290.

121. Becker, A., Marko, M., Gold, M. (1977) *Virology* **78**, 291-305.

122. Hohn, B., Murray, K. (1977) *Proc. Natl. Acad. Sci.* **74**, 3259-3263.

123. Bullock, W. O., Fernandez, J. M., Short, J. M. (1987) *Biotechniques* **5**, 376-379.

124. Vinson, C. R., LaMarco, K. L., Landschultz, W. H., McKnight, S. L. (1988) *Genes. Dev.* **2**, 801-806.

125. Singh, H., LeBowitz, J. H., Baldwin, A. S., Sharp, P. A. (1988) *Cell* **52**, 415-423.

126. Lane, D., Prentki, P., Chandler, M. (1992) *Microbiol. Rev.* **56**, 509-528.

127. Carey, J. (1991) *Methods Enz.* **208**, 103-117.

128. Botstein, D., Shortle, D. (1985) *Science* **229**, 1193-1201.

129. Kadonga, J. T., Knowles, J. R. (1983) *Nucleic Acids Res.* **13**, 1733-1745.

130. Brown, D. M., MacNaught, A. D., Schell, P. D. (1966) Biochemical *Biophysics Res. Comm.* **24**, 967.

131. Hermes, J. D., Parekh, S. M., Blacklow, S. C., Koster, H., Knowles, J. R. (1989) *Gene* **84**, 143-151.

132. Taylor, J. W., Ott, J., Eckstein, F. (1983) *Nucleic Acids Res.* **13**, 8765-8785.

133. Potter, B. V. L., Eckstein, F. (1984) *J. Biol. Chem.* **259**, 14243-14248.

134. Taylor, J. W., Schmidt, W., Cosstick, R., Ozruszek, A., Eckstein, F. (1985) *Nucleic Acids Res.* **13**, 8749-8763.

135. Hermes, J. D., Blacklow, S. C., Knowles, J. R. (1990) *Proc. Natl. Acad. Sci.* **87**, 696-700.

136. Nakamaye, K. L., Eckstein, F. (1986) *Nucleic Acids Res.* **14**, 9679-9697.

137. Sayers, J. R., Schmidt, W., Eckstein, F. (1988) *Nucleic Acids Res.* **16**, 791-802.

138. Higuchi, R., Krummel, B., Saiki, R. (1988) *Nucleic Acids Res.* **16**, 7351-7367.

139. Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K., Pease, L. R. (1989) *Gene* **77**, 61-68.

140. Holm, L., Koivula, A. K., Lehtovaara, P. M., Hemminiki, A., Knowles, J. K. C. (1990) *Protein Engineering* **3**, 181.

APPENDIX

pBluescript SK- phagemids.

The λZAPII vector which was used to express oligonucleotide libraries, contains the phagemid sequence pBluescript SK-. This sequence is placed between the initiation and termination signals of the f1 origin of replication and can be excised *in vivo*. The pBluescript SK- phagemid containing the gene of interest can then be used to overexpress protein and also to generate single-stranded DNA. Single-stranded phagemid DNA was generated continuously during the course of this project for use in dideoxy sequencing and *in vitro* mutagenesis experiments.

In vivo excision of pBluescript SK- phagemids.

When XL1-Blue cells are infected with λZAPII, the pBluescript SK- sequence may be isolated as a phagemid. Since λZAPII contains an f1 origin of replication derived from a filamentous phage, infection with such a phage results in excision of the pBluescript phagemid from the λ vector. When cells containing λZAPII are infected by the filamentous phage, the phage is immediately converted to a double-stranded molecule that expresses phage proteins. The filamentous phage gene II protein recognises the f1 initiator signal and nicks the DNA strand. This nick is used as an initiation site for DNA synthesis and replication proceeds from this origin, displacing the (+) strand which is cleaved at the termination signal. Following nicking by the gene II protein at the termination site, the two ends of the single-stranded DNA are ligated to form a circular stranded of the (+) strand genome. Because both the filamentous helper phage and pBluescript contain the f1 phage packaging sequences, both molecules can be packaged into filamentous phage particles.

The recovery of the pBluescript phagemid is performed by heat inactivating the *E. coli* and temperature sensitive λ phage paticles, infecting a fresh culture of *E. coli* cells at low filamentous phage-to-cell ration and plating on ampicillin plates to select for the ampicillin resistant colonies containing the pBluescript phagemid.

The subsequent transformation of the excised phagemid is performed using *Su*⁻, λ resistant *E. coli* cells (SOLR) which are plated on ampicillin resistant plates. In this strain neither the λ phage nor the infecting helper phage is active because the cell is resistant to λ phage infection and the amber mutations for filamentous phage infection are not suppressed. The pBluescript phagemid survives in this strain due to the presence of the ColE1 origin and ampicillin resistance.

Single-stranded rescue of pBluescript phagemids.

pBluescript phagemid vectors do not enter the single-stranded state unless filamentous helper phage proteins are present. The presence of the f1 origin within phagemid vectors permits production of single-stranded DNA and phagemid particles when required.

The f1 origin in pBluescript contains all the cis-acting sequences required for packaging and replication of filamentous phage. All protein factors required in trans are supplied by helper phage infection of *E. coli* harbouring the pBluescript clone. On infection with helper phage, gene II is expressed by the helper phage and this promotes single-stranded replication of the pBluescript clone at the f1 origin. This single-stranded DNA is packaged by structural proteins also coded for by the filamentous helper phage and extruded, as are the helper phage, through the bacterial membrane into the growth medium. The *lacZ* coding strand is packaged

by the fl(+) origin and the non-coding strand is packaged from vectors containing the fl(-) origin.

To obtain single-stranded DNA it is essential to grow the clone in a strain of *E. coli* containing the F' episome coding for pili, because the filamentous helper phage require pili for infection.

DNA Primers.

Sequencing primer

    5'-AACAGCTATGACCATG-3'

PCR primer 575P

    5'-CGCGCGAATTCGGTGCCCGAATCC-3'

PCR primer 576P

    5'-GCGGCGAATTCTAGCGTTCGTTCGCCAAC-3'

PCR primer 120X

    5'-CGCGCGAATTCGATGTCCGAATATCAG-3'

Spiked primer 1

    5'-GTGCTAGAAACACTGAAGCCGCCAGGCGTTCTCGTGCGAG
    AAAGTTGCAAAGAATGAAACAACTTGAAGACAAGGT-3'

Note: underlined sequences are unspiked.

Spiked primer 2

    5'-GTGCTAGAAACACTGAAGCCGCCAGGCGTTCTCGTGCGAG
    AAAGTTGCAAAGAATG-3'

Mutation of pbZip.

The autoradiograph depicted below shows the location of the deletion within the basic region of the bZip gene.

Postgraduate courses attended

Organic Research Seminars, Department of Chemistry, University of Edinburgh (3 years attendance).

Basic Course in Radiation Protection in Laboratory Work in Science and Medicine, Dr. J. D. Simpson, Radiation Protection Service, University of Edinburgh (1991, 8 lectures).

Medicinal Chemistry, Prof. R. Baker and colleagues, Merck Sharp & Dohme (1992, 1993, 1994).

ICI Pharmaceuticals Presentation on Zoladex, Dr. B. J. A. Furr and colleagues, ICI (1992, 6 lectures).

Recent Advances in the Synthesis and Activity of Agrochemicals, Dr. I. Boddy and Dr. P. J. Dudfield, Schering Agrochemicals (1992, 6 lectures).

The Molecular Genetics of Bacteriophage λ, Prof. N. Murray, Department of Life Sciences, University of Edinburgh (1992, 6 lectures).

NMR and its Applications to Molecules of Biological Importance, Dr. I. H. Sadler and Dr. J. Parkinson, Department of Chemistry, University of Edinburgh (1993, 6 lectures).

Chemical Development in the Pharmaceutical Industry, Dr. J. P. Clayton and colleagues, Smithkline Beecham Pharmaceuticals (1993, 5 lectures).

Industrial Biocatalysis, Dr. J. T. Sime and colleagues, Smithkline Beecham Pharmaceuticals (1994, 4 lectures).

Genetic Control of Development: The Homeobox Story, Dr. W. Gehring, University of Basel (1994, 5 lectures).