Chapter 1. General introduction

1.1 Regulation of transcription in eukaryotes

The correct spatial and temporal development of a cell in an organism relies on the proper expression of the required genes. While each cell contains the same complement of genetic information, it is the subset of genes that is expressed that dictates the characteristics of a cell. The process of gene expression is therefore highly regulated and is governed through the combinatorial and synergistic effects of a conglomerate of regulatory proteins. When gene expression is uncontrolled, aberrant development can occur and this may be prevented in the future by understanding how gene expression is properly regulated.

While gene expression is regulated at each stage of the process, regulation at the first stage, transcription, is particularly important. The creation of an RNA message is firstly dependent on how accessible the gene is. Chromatin remodelling complexes containing acetylases (Howe et al., 1999), methylases (Chen et al., 1999), ubiquitinases (Pham and Sauer, 2000) and biotinidases (Stanley et al., 2001) play a role in modifying the structure and position of nucleosomes enabling the accessibility of the transcriptional start site (reviewed in Demeret et al., 2001). The next level of transcriptional regulation is imparted by the recruitment of RNA polymerase II to the start site, and this is mediated by regulatory proteins known as transcription factors.

1.1.1 Transcription factors in DNA-binding multi-protein complexes

RNA polymerase II requires the formation of a large protein complex referred to as the basal transcriptional apparatus to initiate transcription. This complex binds elements common to all gene promoters and gives rise to a basal level of transcription. A common core promoter element, the TATA box, is located ~30 base pairs upstream of the start site and is recognised by a component of the transcriptional apparatus, the TATA box-binding protein (TBP; Struhl, 1994).

Specificity of transcription required for the proper expression of the appropriate subset of genes is provided by gene-specific transcription factors. These recognise specific DNA sequences within gene promoters or enhancers and modify the
expression of these genes by recruiting the transcriptional apparatus and regulatory proteins such as chromatin remodeling complexes or other DNA-binding proteins (Figure 1.1).

**Figure 1.1. Schematic model of transcriptional activation.** Transcription is carried out by RNA polymerase II (red) with the aid of the basal transcriptional apparatus (blue). These proteins recognise DNA sequences at the transcriptional start site and the TATA box. Gene-specific transcription factors (yellow) that bind sequence-specific sites in the proximal promoter or enhancer regions further help to regulate the recruitment of the transcriptional apparatus.

Transcription factors are therefore often found in large DNA-binding multi-protein complexes, in which they interact with a number of various partners. For example, the GATA-1 protein is able to contact DNA at GATA and GATC sequences, as well as multiple protein partners such as FOG, LMO2 and CBP (reviewed in Cantor and Orkin, 2002). While some of these contacts may be simultaneous, others may occur in different spatial and temporal configurations. The manner by which transcription factors are able to contact multiple partners is by involving modular structures or domains. The following Sections will introduce some of the domains utilised by transcription factors to mediate protein-DNA and protein-protein interactions.

### 1.1.2 Domains involved in protein-DNA interactions

Most DNA-binding transcription factors recognise their DNA sites through small discrete domains that can function independently of the rest of the protein. Many structures of such domains bound to their cognate DNA sequences have been resolved, allowing an understanding of the molecular basis for these interactions. A few examples will be discussed here.
The homeodomain is a ~60-residue motif that was first recognised in proteins that regulate *Drosophila* development (reviewed in Kornberg, 1993), and has since been identified in many eukaryotic regulatory proteins. Homeodoms contain three α-helices, two of which lie almost perpendicularly to one another in a helix-turn-helix (HTH) motif that resembles those found in many prokaryotic proteins. While additional helices or other types of structure may be present, it is the second α-helix in the HTH motif that contacts the major groove of the DNA (Wintjens and Rooman, 1996; Figure 1.2A).

Basic leucine zipper (bZIP) domains contain heptad repeats with predominantly hydrophobic residues at the first and fourth positions. The domain dimerises (either with itself or with a partner protein) with the formation of a parallel α-helical coiled-coil; the formation of this structure positions the adjacent basic regions to make an effective interaction with the cognate palindromic DNA sequence (Ellenberger *et al.*, 1992; Figure 1.2B).

In the basic helix-loop-helix (bHLH) domain, the HLH region dimerises by forming a four-helix bundle. This dimerisation, in a similar way as the bZIP domain, is necessary for the interaction of the basic region with DNA (Ma *et al.*, 1994; Figure 1.2C).

The Ets domains bind to DNA containing a core GGA sequence (Mo *et al.*, 1998). The structure has an overall topology similar to that of a ‘winged helix-loop-helix’, and it is the third α-helix that contacts the DNA major groove (Figure 1.2D).

Zinc finger (ZnF) domains are known in many cases to interact with DNA, and these are discussed in more detail in Section 1.2.
Figure 1.2. Structures of DNA-binding domains bound to their cognate DNA. A. Homeodomain of vnd/NK-2 (PDB code 1NK3; Gruschus et al., 1997). B. Basic leucine zipper of GCN4 (PDB code 1YSA; Ellenberger et al., 1992). C. bHLH of MyoD (PDB code 1MDY; Ma et al., 1994). D. Ets domain of SAP-1 (PDB code 1BC7; Mo et al., 1998). The proteins are displayed in ribbon representations and DNA duplexes are in dark blue.

1.1.3 Mediators of protein-protein interactions

In addition to protein-DNA interactions, protein-protein interactions are also necessary for the formation of transcriptional regulatory complexes. In many cases, these interactions also appear to be mediated by discrete domains, although much less is known about the basis for these protein-protein interactions, in comparison to protein-DNA contacts. Some examples of what is known are presented below.

PDZ domains, named for the discovery of the domain in the PSD-95, DLG and ZO-1 proteins, are found in proteins involved in apoptosis, cytokinesis and signalling. The domain is able to recognise a short stretch of amino acids at the C-termini of target proteins. The PDZ domain contains two $\alpha$-helices and six $\beta$-strands arranged in a
β-barrel fold, and binds the protein ligand by a β-strand addition (Doyle et al., 1996; Figure 1.3A).

The ankyrin repeat is a 33-residue motif that folds into a helix-loop-helix structure. Several repeats pack side-by-side to form an extended curved structure with a groove running across the long concave surface (Krzywda et al., 2004). The concave surface with residues at the tips of the loops mediates the interactions with protein partners, as seen for example in the GABPβ-GABPα and IκBα-NFκB interactions (Batchelor et al., 1998, Huxford et al., 1998, Jacobs and Harrison, 1998; Figure 1.3B).

As noted above, basic leucine zipper (bZIP) domains also mediate protein-protein interactions facilitating the formation of dimeric α-helical coiled-coils. Homodimeric bZIP domains are found in transcription factors such as GCN4 (O'Shea et al., 1991), whereas protein pairs such as Fos/Jun (Glover and Harrison, 1995) and Myc/Max (Ferre-D'Amare et al., 1993) form heterodimeric coiled-coils.

Other domains, including homeodomains, Ets (Batchelor et al., 1998) and MADS box domains (Messenguy and Dubois, 2003), have been shown to be involved in mediating both protein and DNA interactions. Many zinc-binding domains, or zinc fingers (ZnFs) are also involved in mediating protein-protein interactions. These will be discussed in more detail in the following Section.

Figure 1.3. Domains that mediate protein-protein interactions. A. PDZ domain of PSD-95 complexed with a peptide derived from Cript (PDB code 1BE9; Doyle et al., 1996). B. Ankyrin repeats of GABPβ bound to the Ets domain of GABPα (PDB code 1AWC; Batchelor et al., 1998). The protein partners are displayed in magenta.
1.2 Zinc fingers (ZnFs)

In 1985, sequence analysis combined with proteolytic studies suggested the presence of modular domains in the DNA/RNA binding transcription factor TFIIIA from *Xenopus laevis* (Miller et al., 1985). Zinc ions were detected in the native protein, and the loss of structure with the addition of metal chelating agents implied a reliance on zinc for maintenance of structure. It was deduced that TFIIIA consists of nine tandem units, each ~30 residues in length and containing two invariant pairs of cysteine and histidine residues. The term ‘zinc finger’ was coined to describe the way adjacent domains of TFIIIA combine to ‘grip’ the DNA strands.

Nowadays, the definition of the term ‘zinc finger’ has become much broader, and it now defines a small, functional, independently folded domain that requires the coordination of one or more zinc ions to stabilise its structure. The zinc-ligating residues are usually cysteine and histidine residues, or occasionally aspartate or glutamate (Bach, 2000).

Zinc fingers are very common in eukaryotes. It has been estimated that ~3–4 % of human genes encode for ZnF-containing proteins (Lander et al., 2001). Further, this domain is among the most common structural motifs in the proteomes predicted from the genome sequences of *Saccharomyces cerevisiae*, *Drosophila melanogaster* and *Caenorhabditis elegans* (Rubin et al., 2000). Of those ZnF genes that have been characterised, most play roles in the regulation of gene expression.

1.2.1 The different classes of zinc fingers

ZnFs exhibit a variety of folds, and several studies have categorised them according to the identities and spacing of the zinc ligands (Harrison, 1991), zinc ligand geometries (Karlin and Zhu, 1997) and more recently, common structural features (Krishna et al., 2003). This Section describes some of the common classes of zinc fingers, and attempts to illustrate the diverse nature of these domains.

**Classical ZnFs**

Classical ZnFs are found in ~2 % of all human genes and are by far the most abundant DNA-binding domain found in human transcription factors (Tupler et al., 2001). This
domain often occurs in tandem repeats within a protein, and has the consensus sequence of \( \Phi-X-C-X_{2,4}-C-X_3-\Phi-X_5-\Phi-X_2-H-X_{2,5}-H/C \), where \( \Phi \) is hydrophobic and \( X \) is any residue.

Classical ZnFs are composed of a short \( \beta \)-hairpin followed by an \( \alpha \)-helix that forms a left-handed \( \beta\beta\alpha \)-unit (Figure 1.4A). The two zinc-ligating cysteines reside in the turn at the end of the \( \beta \)-hairpin while the histidines lie in the C-terminal end of the \( \alpha \)-helix. An extension of this simple \( \beta\beta\alpha \)-fold is utilised in baculovirus inhibitor of apoptosis repeat (BIR) domains found in many inhibitor of apoptosis (IAP) proteins (Figure 1.4B).

While the classical ZnF is well-recognised as a nucleic acid (DNA and RNA) binding domain (Section 1.2.2), an increasing number of examples has recently also established this domain as mediator of protein-protein interactions (Sun et al., 1996, Tsai and Reed, 1998, Lee et al., 1993, Merika and Orkin, 1995).

**GATA-type fingers**

These ZnFs were originally identified in the GATA superfamily of transcription factors. Four cysteine residues in a ‘treble clef’ fold (Krishna et al., 2003) ligate the central zinc ion, as part of the consensus \( C-X_{2,4}-C-X_{17-20}-C-X_2-C \). The treble clef fold consists of two irregular antiparallel \( \beta \)-sheets and an \( \alpha \)-helix; the \( \beta \)-sheet and \( \alpha \)-helix each contribute two ligands for zinc binding (Figure 1.5A). This fold is considered to be one of the major structural classes of ZnFs and is present in many different types of ZnF domains. GATA-type fingers are able to interact with DNA (Omichinski et al., 1993) and mediate interactions with other proteins (Fox et al., 1998).
Figure 1.5. Structures of treble clef zinc fingers. A. Single treble clef of the N-terminal GATA-type ZnF of GATA-1 (PDB code 1GNF; Kowalski et al., 1999). B. Tandem treble clefs of the C-terminal LIM domain from CRP (PDB code 1CTL; Perez-Alvarado et al., 1994). C. Overlapping treble clefs of the RING domain of equine herpes virus-1 (PDB code 1CHC; Barlow et al., 1994). D. ZnFs in the estrogen hormone receptor (PDB code 1HCP; Schwabe et al., 1990).

LIM domains

Named for the discovery of this domain in the Lin11, Isl-1 and Mec-3 homeodomain proteins (reviewed in Bach, 2000), the LIM domain has proved to be a conserved motif found in organisms ranging from ascidians to man. The LIM domain consists of two tandemly repeated zinc fingers with a consensus sequence of $\text{C}^\text{I}$-$\text{X}_2$-$\text{C}^\text{I}$-$\text{X}_{17-19}$-$\text{H}^\text{I}$-$\text{X}_2$-$\text{C}^\text{I}$-$\text{X}_2$-$\text{C}^\text{II}$-$\text{X}_2$-$\text{C}^\text{II}$-$\text{X}_{16-20}$-$\text{C}^\text{II}$-$\text{X}_2$-$\text{C}^\text{II}$, where I and II represents the zinc ion to which the ligand binds. This sequence folds into two tandem repeats of the treble clef fold (Figure 1.5B). Unlike GATA-type ZnFs however, LIM domains do not bind DNA; rather, they are thought to act as protein interaction motifs that act as bridges in large protein complexes.

RING fingers

The RING (really interesting new gene) finger is a cysteine-rich domain of 40–60 residues. It binds two zinc ions and is involved in protein-protein interactions. There are two different zinc-ligation topologies, $\text{C}_4\text{H}_3\text{C}_3$ and $\text{C}_3\text{H}_2\text{C}_3$, with the consensus sequence of $\text{C}^\text{I}$-$\text{X}_2$-$\text{C}^\text{I}$-$\text{X}_{9-39}$-$\text{C}^\text{II}$-$\text{X}_{1-3}$-$\text{H}^\text{II}$-$\text{X}_{1-3}$-$\text{C}^\text{I}$-$\text{X}_2$-$\text{C}^\text{II}$-$\text{X}_{4-48}$-$\text{C}^\text{II}$-$\text{X}_2$-$\text{C}^\text{II}$. An emerging role of RING-finger-containing proteins is in ubiquitination pathways.
where they are involved in the transfer of ubiquitin to a protein substrate, thereby targeting the substrate for destruction by the proteasome (Joazeiro and Weissman, 2000).

The interleaved nature of the ligation topology (where the first and third pair of ligands is bound to one zinc ion and the second and fourth pair bound to the other zinc; Figure 1.5C) is also a feature of the FYVE and PHD (plant homeodomain) zinc fingers. All these domains are considered to contain overlapping treble clef folds.

ZnFs in hormone steroid receptors

Receptors for steroids and related hormones-like molecules contain a ~80-residue zinc-binding domain with two zinc ions, each ligated by four cysteines (Freedman et al., 1988). Each zinc-binding motif is a typical treble clef where two zinc ligands reside in the α-helix and two more are in loop regions (Figure 1.5D). The α-helix in each zinc-binding motif are folded together such that they cross at right angles near their midpoints to form a single structural domain. This two-zinc domain then dimerises to bind palindromic DNA sequences (Lee et al., 1993).

1.2.2 Zinc finger-mediated functions

As already alluded to, different classes of zinc fingers are able to mediate many different types of interactions. These compact protein domains are able to function as sequence-specific DNA and RNA binding motifs, as well as mediate protein-protein and protein-lipid interactions.

Nucleic acid interactions

Many classes of ZnFs, including classical CCHH fingers, GATA-type fingers and the ZnFs in hormone receptors, recognise DNA.

Classical CCHH ZnFs, generally in tandem arrays of three or more, make specific contacts with DNA bases using key residues at the N-terminal end of their α-helix (Figure 1.6). These arrays are found in both activators and repressors of transcription, including the Kruppel-like factor family (Turner and Crossley, 1999), TFIIMA (Nolte et al., 1998) and the Wilms tumour protein 1 (Nakagama et al., 1995). Each ZnF
contacts three base pairs of DNA, and changes in the identities of residues that contact DNA alter the sequence specificity for binding. As a consequence, arrays of three classical ZnFs can be used to recognise an enormous variety of DNA sequences with high specificity and high affinity.

GATA-family transcription factors contain one or two GATA-type fingers that mediate the interaction between the protein and DNA elements containing a core GATA sequence (reviewed in Orkin, 1992). These domains also contact the major groove of DNA using their $\alpha$-helix, and make additional interactions with the minor groove via flanking basic regions (Omichinski et al., 1993, Starich et al., 1998).

A ZnF domain common to elongation factor TFIIS and several bacteriophage primases is able to recognise single-stranded DNA (Qian et al., 1993, Kusakabe et al., 1999). The ssDNA-binding function is vital to the role of the primases in unwinding double-stranded DNA and catalysing the template-directed synthesis of oligonucleotides on the ssDNA.

A number of classical ZnFs have also been shown to recognise both RNA and DNA-RNA hybrids (Shi and Berg, 1995). For example, TFIIIA fingers 4 to 7 and ZnFs in the p43 protein recognise stem loop structures of the *Xenopus laevis* oocyte 5S RNA (Theunissen et al., 1992, Zang and Romaniuk, 1995), and recently the structure was determined of the TFIIIA zinc fingers complexed to RNA (Lu et al., 2003).

**Figure 1.6. Classical ZnFs of Zif268 bound to its cognate DNA** (PDB code 1ZAA; Pavletich and Pabo, 1991). The ZnF domains wrap around the DNA major groove. The DNA is in *orange*, while the protein is displayed in ribbon representation. The zinc ligands are in *dark blue* and zinc ions in *magenta*. 

![Classical ZnFs of Zif268 bound to its cognate DNA](image)
The ZnF domains in retroviral nucleocapsid (NC) proteins from HIV and other related viruses are critical for viral replication and facilitate packaging of the viral genome by binding to specific sequences in the viral RNA (Figure 1.7).

**Protein interactions**

In addition to their DNA-binding abilities, both classical and GATA-type ZnFs also display the ability to bind proteins (reviewed in Mackay and Crossley, 1998). Classical fingers in the Ikaros proteins mediate both homo- and hetero-oligomerisation with other family members. These interactions modulate the DNA-binding and transcriptional activity of Ikaros-family proteins, which control gene expression in a range of different tissue types (Georgopoulos, 2002).

The GATA-type fingers in the erythroid transcription factor GATA-1 are the protein targets of many different partner proteins, including FOG, several of the Krüppel-like factor proteins, LMO2, PU.1 and CBP (reviewed in Cantor and Orkin, 2002). Interestingly, the myeloid-specific protein PU.1 appears to inhibit GATA-1 function by blocking GATA-1 DNA binding (Zhang *et al.*, 2000). Recently, the solution structure of the GATA-FOG complex was determined in our laboratory (A. Kwan, C. K. Liew and J. Mackay, unpublished results), revealing for the first time how classical and GATA-type zinc fingers can act as protein recognition motifs (Figure 1.8).
LIM proteins have been shown to mediate many different interactions, both in the nucleus and in the cytoplasm. For example, the LIM-only (LMO) family of proteins, members of which consist entirely of two LIM domains, are essential components of a number of transcriptional complexes. LMO2 forms a transcriptional activation complex with GATA-1, ldb1 and the bHLH proteins Tal1 and E2A (Wadman et al., 1997). The structural basis for the interaction of LIM domains in LMO2 and 4 with ldb1 was recently elucidated (Deane et al., 2003), revealing a novel β-zipper-type interaction in which a short segment of ldb1 adds an additional β-strand to the β-hairpins in the LIM domains (Figure 1.9).

Denatured proteins appear to be the binding target of the cysteine-rich ZnF domain of *E. coli* chaperone protein DnaJ. The domain binds two zinc ions to adopt a unique ZnF structure containing a V-shaped extended β-hairpin (Martinez-Yamout et al., 2000).
Lipid interactions

The ZnF domains in protein kinase C (PKCs) proteins mediate interactions with diacylglycerol and its structural analogs, phorbol esters (Zhang et al., 1995). Similarly, the FYVE zinc finger domain (named for Fab1p, YOTB, Vac1p and EEA1 proteins in which they were first discovered; Stenmark et al., 1996) targets proteins to specific membranes by the recognition of phosphatidylinositol-3-phosphate (Gaullier et al., 1998).

1.2.3 Protein engineering using zinc fingers

Consideration of the properties of classical ZnFs has led a number of groups to use these domains to create ‘designer’ DNA-binding proteins. Classical ZnFs have stable, well-defined structures and can tolerate a vast array of amino acid sequence changes, as evidenced by the large number of these domains that exist in nature. Furthermore, their modular DNA-binding mechanism allows a ‘mix and match’ approach to create ZnF arrays with arbitrary DNA-binding specificity.

In an early study, Choo and Klug created an M13 phage library coding for ZnFs with randomised DNA-binding residues. This method revealed DNA-binding preferences for particular ZnF sequences and led to the creation of a ZnF recognition “code” that correlated ZnF amino acid composition with a preferred target DNA sequence (Choo and Klug, 1994). While this code is not universal, it has been used successfully to design novel zinc fingers that are able to bind specifically to (G/A)NN triplet DNA sequences (where N is any of the four nucleotides; Segal et al., 1999, Dreier et al., 2001).

In order to increase specificity and affinity, several novel ZnF proteins have been designed to target long (>9 bp) DNA sequences. By joining domains together, up to nine ZnFs have been linked together using the traditional linker. While the nine-finger protein bound the expected 27-bp DNA site, the affinity of the interaction was similar to that displayed by the wild-type three-finger protein (Kamiuchi et al., 1998). However, by using longer linkers that would minimise the possibility of strain, a six-finger protein was found to bind a 18-bp DNA site with a $K_d$ of $2.1 \times 10^{-15}$ M$^{-1}$,
6000-fold stronger than that of the wild-type three-finger protein (Kim and Pabo, 1998).

Another method utilised to increase the length of the DNA target site is by attaching a dimerisation domain. In an alternative approach to tuning DNA-binding specificity, two ZnFs were linked to the GCN4 leucine zipper domain; the resultant dimeric protein bound a 10-bp DNA site with high specificity (Wolfe et al., 2003; Figure 1.10)

Figure 1.10. Structure of the chimeric Zif23-GCN4 protein. The designed protein contains two classical ZnFs of Zif268 fused to the leucine zipper of GCN4 (PDB code 1LLM; Wolfe et al., 2003).

The design of novel ZnF-based DNA-binding proteins has recently matured to the point where researchers are using these engineered proteins to regulate the expression of genes in model organisms to manipulate phenotype in a defined way. For example, an estrogen-based zinc finger system has been demonstrated in vivo with the regulation of an endostatin transgene in a mouse model (Xu et al., 2001). In this system, the administration of estrogen to the mice results in the dimerisation of the three-finger protein. The six-finger dimer then acts on a 18-bp DNA site to activate the expression of the endostatin transgene.

1.3 General aims of this study

Since the discovery of the first zinc finger domains in the early 1980’s, it has become clear that these domains are abundant in nature and highly versatile in their ability to mediate a myriad of interactions. This Thesis will focus on understanding the structure and function of a number of zinc fingers that are involved either in protein-protein interactions or both protein-protein and protein-DNA interactions.
Chapter 3 is centred on a GATA-type zinc finger domain, the N-terminal finger (or NF) of the transcription factor GATA-1. This protein domain is able to mediate interactions with both DNA and the coactivator FOG. These interactions are affected by mutations that have been identified in several families with inherited blood disorders. This study explores the effects of each mutation on the binding energetics of the interactions between the GATA-1 NF and its protein and DNA partners.

Chapter 4 investigates the molecular basis for the interaction between FOG (a protein partner of GATA-1) and TACC3, a protein implicated in chromosome dynamics. This recently discovered interaction is mediated by one of the classical zinc fingers in FOG and by a coiled-coil domain from TACC3. As noted above, little structural information is currently available on protein-protein interactions mediated by classical ZnFs, and the results presented here represent some of the first available data for such an interaction.

Chapter 5 explores the ligand requirements for the classical CCHH zinc finger fold. A number of ‘incomplete’ CCHX modules are found in sequence databases and, given that many zinc-containing enzymes function with only three protein ligands, this study investigates the ligand requirements for the formation of functional zinc-binding domains.

Together, these studies shed light on the structural and functional basis for a number of interactions mediated by zinc fingers.
Chapter 2. Materials and methods

All solutions were made up in Milli-Q® water (MQW).

2.1 Subcloning

Type II restriction endonucleases (BamHI and EcoRI; EC 3.1.21) and 100-bp DNA ladder were obtained from New England Biolabs (Beverly, MA). T4 DNA ligase (EC 6.5.1.1), Taq and Pfu turbo™ DNA polymerases (EC 2.7.7.7) were from Stratagene (La Jolla, CA). dNTPs were purchased from Boehringer Mannheim (Mannheim, Germany) and PCR primers were synthesised by Sigma Genosys (Castle Hill, NSW). Molecular biology grade agarose was obtained from Progen Industries Ltd (Darra, QLD) and Costar® Spin-X filter tubes were from Corning Incorporated (Corning, NY). Casein peptone pancreatic digest type M for bacterial culture, yeast extract and bacteriological agar were obtained from Amyl Media (Kings Langley, NSW). All other reagents were of analytical grade.

Plasmids containing DNA inserts corresponding to murine GATA-1 NF48 (200–248) and NF54 (200–254) in pGEX-2T vectors were gifts from Anthea Newton. She also kindly provided pGEX-2T vectors with inserts corresponding to V205M, G208S, R216Q, D218G and D218Y. DNA insert encoding murine FOG-1 finger 3 (328–360) was provided by David Bishop. Plasmids containing sequences encoding murine TACC3 residues 535–637, 561–637 and 591–637 (TACC103, TACC77 and TACC47, respectively) in pGAD10 vectors were gifts from Stella Lee. Robert Czolij provided pGEX-2T vectors containing BF1–3, BF1–2, BF1–3A and BF1–3(-HMLV) inserts.

While some plasmid gifts were used as provided, the majority was used as templates for the synthesis of required plasmids. The steps involved in the subcloning process are depicted in Figure 2.1 and Table 2.1 lists the plasmids that have been synthesised, together with the stage at which subcloning was started.
Figure 2.1. Steps involved in subcloning. Each step of the flowchart refers to the Section quoted in brackets. The letters (A to D) refers to the particular state of the plasmids: A. The required insert in the incorrect vector. B. Insert with some required sequences in a vector. C. The required DNA insert. D. The required DNA insert in the correct vector.
Table 2.1. Plasmids synthesised for protein overexpression purposes. *describes the step at which the synthesis of the plasmid began (please refer to Figure 2.1). °In-house plasmid database entry number.

<table>
<thead>
<tr>
<th>Insert</th>
<th>Vector</th>
<th>Starting stage*</th>
<th>Database #°</th>
</tr>
</thead>
<tbody>
<tr>
<td>NF48</td>
<td>pGEX-2T</td>
<td>D</td>
<td>649</td>
</tr>
<tr>
<td>NF54, V205M, G208S, R216Q, D218G, D218Y</td>
<td>pGEX-6P</td>
<td>A</td>
<td>76–80,650</td>
</tr>
<tr>
<td>FOG-F3</td>
<td>pGEX-2T</td>
<td>C</td>
<td>654</td>
</tr>
<tr>
<td>FOG-F3KRA, FOG-F3KRRRA</td>
<td>pGEX-2T</td>
<td>B</td>
<td>655, 656</td>
</tr>
<tr>
<td>TACC103, TACC77, TACC47</td>
<td>pGEX-6P</td>
<td>A</td>
<td>675–677</td>
</tr>
<tr>
<td>BF3</td>
<td>pGEX-2T</td>
<td>B</td>
<td>83</td>
</tr>
<tr>
<td>BF3A, D, E, N, Q, R</td>
<td>pGEX-2T</td>
<td>B</td>
<td>657–662</td>
</tr>
<tr>
<td>BF3(N)-H333A</td>
<td>pGEX-2T</td>
<td>B</td>
<td>673, 674</td>
</tr>
<tr>
<td>BF1–3 constructs</td>
<td>pGEX-2T</td>
<td>D</td>
<td>668–672</td>
</tr>
</tbody>
</table>

2.1.1 Site-directed mutagenesis using polymerase chain reaction (PCR)

PCR reactions were carried out in a 50-µl reaction volume containing 2.5 U of Pfu turbo™ DNA polymerase (Stratagene) and the supplied Pfu PCR buffer, dNTPs (each at 0.1 mM) and 10 pmol of each primer. The DNA template consisted of 1–2 µl of plasmid. A PCR Sprint thermal cycler (Hybaid, Middlesex, UK) was used for the 30-cycle reaction, in which each cycle consisted of one minute at 92 °C for denaturation, one minute at 55 °C for annealing and three minutes at 72 °C for extension. The reaction was stored at 4 °C upon completion.

The amplification of DNA sequences corresponding to FOG-F3KRA and FOG-F3KRRRA were performed using forward primers that contained the required mutations (‘F3triple-fwd’ for FOG-F3KRA and ‘F3quad-fwd’ for FOG-F3KRRRA; Table 2.2), reverse primer ‘A1137’ (from M. Crossley) and the plasmid template of DNA corresponding to FOG-F3 with the E354A mutation in pGAD10 vector (‘AS #6’; gift from S. Lee).

To amplify the wild-type and mutant BF3 genes, wild-type BF1–3 in pGEX-2T and ‘BKLFF3-2Tfwd’ was used as DNA template and forward primer, respectively. The reverse primer used in the amplification of the wild-type sequence was ‘BKLFF3-2Trev’, while those used for the mutants contained the codon corresponding to single amino acid substitution mutation. Reverse primers
‘3EcoBKLFF3Ala’, ‘Asp’, ‘Asn’, Glu’, ‘Gln’ and ‘Arg’ (#869 to #875 from M. Crossley) were used to synthesise the required BF3 mutant genes.

BF3-H333A and BF3N-H333A constructs were synthesised using an overlapping PCR strategy. For BF3-H333A, two PCR reactions were carried out concurrently to introduce the H333A mutation. Primers ‘BKLFF3-2Tfwd’ and ‘BF3A333-rev’ were used in one reaction, while the other utilised ‘BF3A333-fwd’ and ‘BKLFF3-2Trev’. The PCR products of the two reactions were pooled together for a third reaction, where primers ‘BKLFF3-2Tfwd’ and ‘BKLFF3-2Trev’ were used. A similar protocol was used for BF3N-H333A with one exception: primer ‘3EcoBKLFFAsn’ (#871 from M. Crossley) was used in place of ‘BKLFF3-2Trev’ in order to incorporate the H341N mutation. In all reactions, the pGEX-2T vector containing BF3 was used as the DNA template.

<table>
<thead>
<tr>
<th>Name</th>
<th>Purpose</th>
<th>Sequence</th>
<th>Oligo #</th>
</tr>
</thead>
<tbody>
<tr>
<td>F3triple-fwd</td>
<td>FOG-F3 KRA</td>
<td>CGGGATCCTCCGGTAAACG TCCGTTGTTTGCCGTATCT GCCTGTCGGCCTTC</td>
<td>736</td>
</tr>
<tr>
<td>F3quad-fwd</td>
<td>FOG-F3 KRA</td>
<td>CGGGATCCTCCGGTAAACG TCCGTTGTTTGCCGTATCTGC CGTTCGGCCTTCACCACCAAG</td>
<td>737</td>
</tr>
<tr>
<td>BKLFF3-2Tfwd</td>
<td>BF3 (wt + mut)</td>
<td>CGGGATCCATCAAAACCTTTCC AGTGTC</td>
<td>259</td>
</tr>
<tr>
<td>BKLFF3-2Trev</td>
<td>BF3 wt</td>
<td>GGAATTCATCATATTAGACTA GCATGTTGCCTTTT</td>
<td>260</td>
</tr>
<tr>
<td>BF3A333-fwd</td>
<td>BF3(N)-H333A</td>
<td>CGCTCTGACGCTCTTGCCCTA</td>
<td>565</td>
</tr>
<tr>
<td>BF3A333-rev</td>
<td>BF3(N)-H333A</td>
<td>TAGGGCAAGAGGCAGTCAGAGCG</td>
<td>566</td>
</tr>
<tr>
<td>GSTfwd</td>
<td>general</td>
<td>GGGCTGGCAAGACGCCGCTGTG</td>
<td>N/A</td>
</tr>
<tr>
<td>GSTrev</td>
<td>general</td>
<td>CCGGGAGCTGCAGTGTGTCAGAGG</td>
<td>N/A</td>
</tr>
</tbody>
</table>

2.1.2 Restriction enzyme digestion

Plasmids were digested with restriction enzymes BamHI and EcoRI (New England Biolabs, MA). The DNA (~5 µg, if concentration was known) was resuspended in a 30-µl reaction mixture containing 20 U of each restriction enzyme, the recommended buffer and 0.1 mg ml⁻¹ of bovine serum albumin. The mixture was then incubated at 37 °C for 1 h and a further 15 min at 85 °C to inactivate the enzymes.
2.1.3 **Purification of DNA by ethanol precipitation**

The DNA products were loaded onto a 2 % agarose gel and electrophoresed at 100 V for ~1 h. The agarose gel was prepared in TAE buffer (40 mM Tris, 1 mM EDTA, 0.1 % acetic acid) containing 2 µg ml⁻¹ ethidium bromide. The gel bands containing DNA were visualised by ultraviolet illumination, and fragments that corresponded to the correct size were excised and placed in a Spin-X® tube (Corning, NY). The gel extract was frozen and then centrifuged (10,000 rpm, 5 min, RT). The DNA in the filtrate was precipitated by adding 10 %(v/v) of 3 M sodium acetate and 2 volumes of ethanol, then placed on ice for 30 min. The DNA was pelleted by centrifugation (10,000 rpm, 10 min, 4 °C) and resuspended in sterile MQW.

2.1.4 **Ligation**

The *Bam*HI and *Eco*RI-digested fragment and vector were ligated using T4 DNA ligase (Strategene). The ligations were carried out with the gene inserts added in excess and allowed to incubate at RT for 12–24 h. A negative control, in which gene insert was not added, was carried out concurrently. The entire ligation mix was then used in the subsequent transformation.

2.1.5 **Transformation into *Escherichia coli* cells**

Luria-Bertoni (LB) medium (1 % casein peptone, 0.5 % yeast extract, 0.5 % sodium chloride) was autoclaved and stored at 4 °C. When required, ampicillin (50 µg ml⁻¹) was added. LB-agar plates were made with LB medium containing bacteriological agar (1.5 %). When the medium had cooled to ~40 °C after autoclaving, ampicillin (50 µg ml⁻¹) was added and the plates poured. These were stored at 4 °C until use.

Two bacterial strains of *Escherichia coli* were used in transformations. DH5α *(supE44, ΔlacU169 [φ80lac ZΔM15], hsdR17, recA1, endA1, gyrA96, thi-1, relA1)* cells were used in plasmid maintenance, whereas BL21 (DE3) (F* ompT [lon] hsdS*₂ (rB' mB')) cells were transformed when protein expression was intended.

A 100-µl aliquot of competent cells was allowed to thaw on ice and added to the plasmids or ligation reaction mix. The mixture was kept on ice for 10 min, then incubated at 37 °C for 5 min to facilitate uptake of the plasmids. LB medium (200 µl)
without ampicillin was added prior to a 90-min incubation at 37 °C. The mixture (50 µl of DH5α or 75 µl of BL21 cells) was streaked onto an LB-agar plate containing 50 µg ml⁻¹ ampicillin and then incubated overnight at 37 °C. Non-transformed cells, as well as those transformed with plasmids with no insert, were used as negative controls.

2.1.6 Colony screen

In order to identify cells that contain the required plasmids, individual colonies were placed in a PCR reaction mix and also applied onto a LB-agar plate in order to propagate the same colony. Primers ‘GSTfwd’ and ‘GSTrev’ and †Taq DNA polymerase were used to amplify the cloning sites of the pGEX vector (see Section 2.1.1). The PCR products were electrophoresed using a 2 % agarose gel as described in Section 2.1.3 and the presence of an insert was indicated by the size of the amplified fragment.

2.1.7 Extraction of plasmids from E. coli

The colony identified to harbor the insert-containing plasmid was used to inoculate LB medium (200 ml) with ampicillin (50 µg ml⁻¹) and the culture was incubated overnight at 37 °C with shaking. The plasmids were extracted using the Jetstar 2.0 midiprep kit (Genomed, Löhne, Germany), stored at –20 °C and used for transformation into BL21 (DE3) cells when protein production was required. Nucleotide sequencing was carried out by Supamac (Sydney University Prince Alfred Macromolecular Analysis Centre, Sydney University, Camperdown).

2.2 Protein overexpression

Isopropyl β-D-thiogalactopyranoside (IPTG) was purchased from Progen Industries Ltd (Darra, QLD) and ¹⁵N-ammonium chloride (>99 %) obtained from Spectra Stable Isotopes (Columbia, MD).

2.2.1 Overexpression using shaker flasks

A single E. coli BL21 (DE3) colony harboring insert-containing plasmids was used to inoculate LB medium (50 ml) with ampicillin (50 µg ml⁻¹), which was then grown
overnight at 37 °C with shaking. LB medium (1 L) with ampicillin (50 µg ml\(^{-1}\)) was then inoculated with the overnight culture to an \(A_{600\text{nm}}\) of 0.05 and then cultured at 37 °C with shaking.

For the expression of BKLF constructs, protein overexpression was induced with IPTG (0.4 mM) at 37 °C when the \(A_{600\text{nm}}\) reached ~0.6. For the other proteins (e.g., NF, FOG-1 and TACC proteins), when the \(A_{600\text{nm}}\) reached ~0.6, the temperature was lowered to 25 °C (Table 2.3). Protein overexpression was induced with IPTG (0.4 mM) at 25 °C. The cells were pelleted by centrifugation (6,000 rpm, 15 min, 4 °C) after 4 h when induced at 37 °C or after ~16 h when induced at 25 °C. The cell pellets were stored at –20 °C until required. One-ml samples were taken prior to induction and also before harvesting. These samples were pelleted by centrifugation (10,000 rpm, 5 min) and the cell pellets were subjected to glycine SDS-PAGE analysis to determine the outcome of the overexpression.

Shaker flasks were occasionally used to overexpress uniformly \(^{15}\)N-labelledd proteins and this was carried out in a manner described by Marley et al. (2001).

Table 2.3. Induction temperatures.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Induction temp (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NF48/NF54</td>
<td>25</td>
</tr>
<tr>
<td>FOG-F3</td>
<td>25</td>
</tr>
<tr>
<td>TACC proteins</td>
<td>25</td>
</tr>
<tr>
<td>BKLF proteins</td>
<td>37</td>
</tr>
</tbody>
</table>

### 2.2.2 Overexpression using a fermentor

The fermentor method was routinely used to produce uniformly \(^{15}\)N-labelled proteins according to a protocol adapted from Cai et al. (1998). This method was utilised to produce \(^{15}\)N-labelled proteins as it produces higher yields compared to that obtained with the use of shaker flasks. Media used are listed in Table 2.4. A 3-l sterile basic salt solution containing trace metal solution, ampicillin, thiamine, yeast extract and MgCl\(_2\) was prepared. One litre of this solution was inoculated with a 2-ml 12-h LB culture and grown overnight at 37 °C. The remaining 2 l of sterile solution was placed in the reaction vessel of a New Brunswick Scientific Bioflow III Fermentor. The program was set to monitor pH (maintained at 7), temperature (37 °C), dissolved O\(_2\)
Materials and methods

concentration and agitation rate (initially at 50 rpm and incremented by 2 rpm every 0.5 min if dissolved O₂ was less than 75 %). This solution was inoculated with the 1-l overnight culture and grown until a spike was observed in the dissolved O₂ level (indicating the depletion of NH₄Cl). At this point, ^1⁵N-NH₄Cl (0.75 g) was added and the cells were allowed to grow until another spike in the dissolved O₂ level was seen (indicating depletion of ^1⁵N-NH₄Cl). For proteins that required induction at 25 °C, the temperature was then lowered to 25 °C. When the culture reached the required temperature, protein overexpression was induced with IPTG and a further 3.75 g of ^1⁵N-NH₄Cl was added. The cells were then grown for 3 h (37 °C) or ~16 h (25 °C induction). The cells were harvested by centrifugation (6,000 rpm, 15 min, 4 °C) and stored at -20 °C until required.

Table 2.4. Recipe for 1-litre growth medium, used for overexpression in a fermentor (Cai et al., 1998).

<table>
<thead>
<tr>
<th>Solution Type</th>
<th>Concentration (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic salt solution (autoclaved)</td>
<td>13.0 g</td>
</tr>
<tr>
<td>KH₂PO₄</td>
<td></td>
</tr>
<tr>
<td>K₂HPO₄</td>
<td>10.0 g</td>
</tr>
<tr>
<td>Na₂HPO₄</td>
<td>9.0 g</td>
</tr>
<tr>
<td>K₂SO₄</td>
<td>2.4 g</td>
</tr>
<tr>
<td>NH₄Cl</td>
<td>1.0 g</td>
</tr>
<tr>
<td>Trace metal solution (sterile filtered)</td>
<td>0.60 g</td>
</tr>
<tr>
<td>FeCl₂.6H₂O</td>
<td></td>
</tr>
<tr>
<td>CaCl₂.2H₂O</td>
<td>0.60 g</td>
</tr>
<tr>
<td>MnCl₂.4H₂O</td>
<td>0.12 g</td>
</tr>
<tr>
<td>CoCl₂.6H₂O</td>
<td>0.08 g</td>
</tr>
<tr>
<td>ZnSO₄.7H₂O</td>
<td>0.07 g</td>
</tr>
<tr>
<td>CuCl₂.2H₂O</td>
<td>0.03 g</td>
</tr>
<tr>
<td>H₃BO₃</td>
<td>2 mg</td>
</tr>
<tr>
<td>Na₂MoO₄.2H₂O</td>
<td>0.025 g</td>
</tr>
<tr>
<td>EDTA</td>
<td>0.50 g</td>
</tr>
<tr>
<td>MgCl₂ (sterile filtered)</td>
<td></td>
</tr>
<tr>
<td>thiamine (5 mg ml⁻¹, sterile filtered)</td>
<td></td>
</tr>
<tr>
<td>ampicillin (100 mg ml⁻¹, sterile filtered)</td>
<td></td>
</tr>
<tr>
<td>D-glucose (12 g, sterile filtered)</td>
<td></td>
</tr>
<tr>
<td>10 % yeast extract (sterile filtered)</td>
<td></td>
</tr>
</tbody>
</table>
2.3 SDS-PAGE analysis

Two types of SDS-PAGE were used. Glycine gels (Sambrook et al., 1989) were used for resolving large proteins, where the minimum resolved size was ~10 kDa. Tricine gels (Schagger and von Jagow, 1987) were used for resolving small proteins (minimum size ~3 kDa). 1.5-mm gels were made unless stated otherwise. Mark12™ protein standards were supplied by Invitrogen Australia Pty Ltd (Mount Waverley, VIC) and 4–20 % pre-poured iGels were obtained from Gradipore (Frenchs Forest, NSW).

$5 \times$ SDS-PAGE loading buffer (0.3125 M Tris, 10 % SDS, 50 % glycerol, 25 % β-mercaptoethanol, 0.5 % bromophenol blue, pH 6.8) was added to all gel samples. The samples were then heated at 95 °C for 5 min prior to loading. SDS-PAGE were carried out using Hoefer Tall Mighty Small® or Small Mighty Small® apparatus (Amersham Biosciences, Castle Hill, NSW).

2.3.1 Glycine SDS-PAGE

Tris/glycine SDS-polyacrylamide gels comprised of a 4 % stacking gel (0.5 M Tris, pH 6.8) and a 15 % resolving gel (1.5 M Tris, pH 8.8; Sambrook et al., 1989). Gels were electrophoresed using running buffer (25 mM Tris, 250 mM glycine, 0.1 % SDS) at 40 mA until the dye front reached the bottom of the gel. Gels were then fixed in fixing solution (30 % methanol and 10 % acetic acid) for 20 min, stained in glycine stain (0.125 % Coomassie BBR, 40 % methanol, 7 % acetic acid) overnight and destained in fixing solution.

2.3.2 Tricine SDS-PAGE

Tris/tricine gels consisted of a 4 % stacking, 10 % spacer and 16.5 % resolving gel (pH 8.45; Schagger and von Jagow, 1987). Tricine gels were ran at 30 V for 1 h and then at 40 mA until the dye front reached the bottom of the gel. The cathode buffer consisted of 0.1 M Tris, 0.1 M tricine, 0.1 % SDS, pH 8.25, whereas the anode buffer was 0.2 M Tris, pH 8.9. Gels were fixed in a solution of 50 % methanol and 10 % acetic acid, stained overnight in 0.05 % Coomassie BBG with 10 % acetic acid and destained in 10 % acetic acid.
2.4 Protein purification

Acetonitrile (HPLC grade) was purchased from Allied Signal Inc. (Muskegon, MI). Glutathione-Sepharose® 4B beads and PreScission™ protease were obtained from Amersham Biosciences (Castle Hill, NSW). Reduced glutathione, trifluoroacetic acid (TFA), thrombin protease, lysozyme (from chicken egg white; EC 3.2.1.17) and phenylmethylsulfonylfluoride (PMSF) were supplied by Sigma Aldrich (St. Louis, MO). Tris(2-carboxyethyl)phosphine (TCEP) and dithiothreitol (DTT) were purchased from Progen Industries Ltd (Darra, QLD). DNase I (from bovine pancreas, Grade II) was obtained from Roche Diagnostics (Mannheim, Germany).

2.4.1 Glutathione (GSH) affinity chromatography

Cell pellets from protein overexpression were resuspended in lysis buffer (50–200 ml; 50 mM Tris, 50 mM NaCl, 1 % Triton-X, 1.4 mM PMSF, 0.1 µl ml⁻¹ β-mercaptoethanol, pH 8.0). Cell lysis was carried out with the addition of lysozyme and DNase I (100 µg and 10 µg per ml of lysate, respectively) and repeated rounds of freeze and thaw. The suspension was then centrifuged at 17,000 rpm, 4 °C for 15 min. The soluble fraction was immediately loaded onto pre-equilibrated glutathione-Sepharose® 4B beads (3–5 ml packed beads; Amersham Biosciences, Castle Hill, NSW).

The glutathione beads were initially rinsed with MQW (10 CV, column volume) and then pre-equilibrated with lysis buffer prior to sample loading. The soluble fraction was loaded onto the column at a rate of ~0.5 ml min⁻¹ and the collected flow-through was reloaded onto the column. Unbound molecules were removed with wash buffer (10 CV; 50 mM Tris, 100 mM NaCl, 10 % glycerol, 1.4 mM PMSF, 0.1 µl ml⁻¹ β-mercaptoethanol, pH 8.0), then with PBS (10 CV; 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 140 mM NaCl, 2.7 mM KCl, pH 8.0) and finally with either thrombin or PreScission buffer (10 CV; thrombin buffer was composed of 50 mM Tris, 150 mM NaCl, 2.5 mM CaCl₂, pH 8.0, whereas PreScission buffer was made up of 50 mM Tris, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.0).

Either thrombin (Sigma Aldrich) or PreScission™ protease (Amersham Biosciences) was used to release the protein from the GST moiety. The conditions of protease cleavages are set out in Table 2.5. When cleavage was carried out on beads, the
protease was added directly to the bead slurry. When in-solution cleavage was required, GST-fusion proteins were eluted from the beads with thrombin or PreScission buffer containing reduced glutathione (50–100 mM, pH ~8) and the protease was then added to the eluate.

Table 2.5. Conditions used in protease cleavage.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Protease</th>
<th>Amount of protease (U per l culture)</th>
<th>Cleavage conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>NF48</td>
<td>Thrombin</td>
<td>100</td>
<td>In solution, 4 °C, overnight</td>
</tr>
<tr>
<td>NF54 + mutants</td>
<td>PreScission™</td>
<td>80</td>
<td>On beads, 4 °C, overnight</td>
</tr>
<tr>
<td>FOG-F3 + mutants</td>
<td>Thrombin</td>
<td>50</td>
<td>In solution, 37 °C, 2 h</td>
</tr>
<tr>
<td>TACC3</td>
<td>PreScission™</td>
<td>80</td>
<td>On beads, 4 °C, overnight</td>
</tr>
<tr>
<td>BF3 + mutants</td>
<td>Thrombin</td>
<td>100</td>
<td>On beads, RT, overnight</td>
</tr>
</tbody>
</table>

The proteins cleaved on the beads were eluted with either thrombin or PreScission buffer in five 10-ml fractions, enriched with PMSF (1.4 mM) and β-mercaptoethanol (0.1 µl ml⁻¹) and stored at 4 °C. Proteins cleaved in solution were snap-frozen and stored at −20 °C until further purification. Samples (20 µl) of the bead slurry and flow-throughs were taken at each stage in order to monitor the purification by tricine SDS-PAGE (Section 2.3.2).

2.4.2 Reverse phase (rp) HPLC

All buffers and samples intended for HPLC use were filtered. Buffers were filtered using 0.45-µm membranes supplied by Supelco (Bellefonte, PA) and samples were filtered either with Millex® GP 0.22-µm filter units (for volumes > ~2 ml; Millipore, Billerica, MA) or Nanosep MF 0.45-µm filters (for volumes < 2 ml; Pall Corporation, Ann Arbor, MI).

Protein samples from GSH affinity chromatography were subjected to rpHPLC on either a Jupiter C18 (4.6 × 250 mm) or Vydac C18 (4 × 200 mm) column connected to a GBC (Dandenong, VIC) HPLC system. The injected fractions were eluted using a gradient of 5–95 % acetonitrile (0.1 % TFA) in 20 min with a flow rate of 1 ml min⁻¹. The eluate was monitored by measuring the absorbances at 215 and 280 nm, and the peaks were collected, lyophilised and stored at −20 °C.
2.4.3 Size exclusion chromatography of NF54

The fractions from GSH affinity chromatography were pooled and concentrated using a Microsep 1K concentrator (Pall Corporation). The pooled samples were loaded onto a Superdex™ 75 10/300 GL column (Amersham Biosciences) operating on a BioLogic (Bio-Rad Laboratories) HPLC system. The mobile phase (20 mM sodium acetate, 300 mM NaCl, 0.5 mM TCEP, 0.5 mM ZnSO₄, pH 6.1) was flowed at a rate of 1.5 ml min⁻¹. The eluate was monitored by absorbance at 280 nm and the peaks were collected and kept at 4 °C.

2.4.4 Size exclusion chromatography and multiangle light scattering of TACC77

Reverse phase HLPC-purified TACC77 was dissolved in 20 mM sodium acetate, 150 mM NaCl and 1 mM DTT. The pH was adjusted to ~5.0 using 0.1 M NaOH. The sample was loaded onto a Superose™ 12 HR 10/30 size exclusion column (Amersham Biosciences) operating on an AKTA (Amersham Pharmacia Biotech) HPLC system. The buffer (20 mM sodium acetate, 150 mM NaCl, 1 mM DTT, pH 5.0) was flowed at a rate of 0.5 ml min⁻¹. The size exclusion column was followed in-line by a mini-DAWN light scattering detector and an interferometric refractometer (Wyatt Technologies). Light scattering analysis was performed at a laser wavelength of 690 nm. The calibration of voltage and light-scattering intensity with toluene yielded a constant of $8.534 \times 10^{-6}$ for this study. Molecular weight calculations were performed with ASTRA software (Wyatt Technologies). The change in refractive index as a function of protein concentration is approximately constant for proteins and a value of 0.190 ml g⁻¹ was used.

2.4.5 Determination of protein concentration

The lyophilised proteins were dissolved in MWQ and the absorbance at the required wavelength was measured (Table 2.6). The concentrations were determined using their molar extinction coefficients (Table 2.6) according to the Beer-Lambert Law (Equation 2.1).

$$A = \varepsilon \cdot c \cdot l$$  \hspace{1cm}  (Equation 2.1)
where \( \varepsilon \) is the molar extinction coefficient (in \( \text{M}^{-1}\text{cm}^{-1} \)), \( c \) is the protein concentration (in \( \text{M} \)) and \( l \) is the pathlength (1 cm).

### Table 2.6. Molar extinction coefficients.

<table>
<thead>
<tr>
<th>Protein</th>
<th>( \lambda ) (nm)</th>
<th>( \varepsilon ) (M(^{-1})cm(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>NF48, NF54 (wild-type and mutants)</td>
<td>280</td>
<td>8,480</td>
</tr>
<tr>
<td>FOG-F3, FOG-F3(<em>{KRA}), FOG-F3(</em>{KRRA})</td>
<td>265</td>
<td>1,069</td>
</tr>
<tr>
<td>TACC103</td>
<td>280</td>
<td>4,470</td>
</tr>
<tr>
<td>TACC77</td>
<td>280</td>
<td>1,490</td>
</tr>
<tr>
<td>BF3 (wild-type and mutants)</td>
<td>230</td>
<td>19,500</td>
</tr>
<tr>
<td>GST-BF1-3</td>
<td>280</td>
<td>58,330</td>
</tr>
</tbody>
</table>

### 2.5 Large-scale DNA purification

This method was used to purify the DNA in NF:DNA interaction studies. The 16-mer oligonucleotides containing the core GATC sequence were purchased from Sigma Genosys in a 5 or 10 \( \mu \text{mol} \) scale with the trityl group attached. The sequence of the 5’ to 3’ strand is GCAACTGATCTGGACT.

The oligonucleotides were individually purified using reverse phase HPLC. 0.1 M triethylammonium acetate buffer (TEAAc) was prepared by diluting the required volume of triethylamine (HPLC grade, Sigma Aldrich) and slightly less than the required volume of acetic acid in water to the volume of 1 l. Acetic acid was added slowly to decrease the pH to 6.7. The crude trityl-on form of oligonucleotide was dissolved in TEAAc and injected onto a Jupiter C18 (4.6 \( \times \) 250 mm) column that had been equilibrated with 15 % acetonitrile in TEAAc. Initially, a 10-min period of isocratic elution was allowed for the release of free organics and detritylated products, then the trityl-on synthesis products were detritylated on column with a shallow 7-min 0–0.5 % trifluoroacetic acid gradient in MQW. The final products were eluted with a 20-min 0–50 % gradient of acetonitrile in TEAAc. The elution profile was monitored by following the absorbance at 215 and 280 nm and the major peak was collected and lyophilised.

The purified single-strand oligonucleotides were mixed together at approximate equimolar concentrations and annealed by heating to 95 °C for 5 min, followed by
Materials and methods

Slow cooling to room temperature. The annealed oligonucleotides were then subjected to size exclusion chromatography using either a Superdex™ 75 10/300 GL or Superose™ 12 HR 10/30 size exclusion column (Amersham Biosciences). The A_{280nm} was monitored and the major peak was collected. The fractions containing the duplex molecules were pooled together and concentrated using a Microsep 1K concentrator (Pall Corporation). The concentration of the single- and double-stranded molecules were determined spectrophotometrically using \( \varepsilon_{\text{ssDNA}} = 160,000 \text{ M}^{-1} \text{ cm}^{-1} \) and \( \varepsilon_{\text{dsDNA}} = 213,333 \text{ M}^{-1} \text{ cm}^{-1} \) at 260 nm (Equation 2.1).

2.6 Electrospray mass spectrometry

Data were recorded on a LCQ Electrospray Mass Spectrometer (Thermo Finnigan, CA). Protein samples were prepared by dissolving the rpHPLC-purified peptide in a 50 % acetonitrile and 1 % acetic acid solution, and directly infused at a flow rate of 5 \( \mu \text{l min}^{-1} \). The data, collected in positive ion mode, were analysed using Xcalibur® software (Thermo Finnigan, CA).

2.7 Circular dichroism (CD) spectropolarimetry

Ammonium \( d \)-camphor-10-sulfonate and 2,2,2-trifluoroethanol were obtained from Katayama Chemical (Hyogo, Japan) and Sigma Aldrich, respectively.

2.7.1 Using CD to determine secondary structure content

Far-ultraviolet CD spectra were collected on a Jasco J-720 spectropolarimeter equipped with a Neslab RTE-111 temperature controller. The instrument was calibrated with ammonium \( d \)-camphor-10-sulfonate. Using a 1-mm pathlength cell, spectra were recorded with a resolution of 0.5 nm and bandwidth of 1 nm over the wavelength range 184–260 nm. Final spectra represented the sum of three scans accumulated at a speed of 20 nm min\(^{-1}\) with a response time of 1 s. A baseline spectrum of buffer alone was recorded and subtracted from subsequent spectra.

The CD ellipticities are reported in the units of mean residue ellipticity (MRE), or \([\theta]\), in units of deg cm\(^{-2}\) dmol\(^{-1}\). These were calculated by the following equation:
\[ [\theta] = \frac{M_o \times \phi}{100 \times l' \times c} \]  
(Equation 2.2)

where \( M_o \) is the mean residue weight (usually 115 for proteins), \( \phi \) is the ellipticity (in deg), \( l' \) is the light path (in dm) and \( c \) is the protein concentration (in g ml\(^{-1}\)).

2.7.2 Using CD to determine the Zn\(^{2+}\)-binding affinities of BF3 and BF3N

For the Zn\(^{2+}\) titration experiments, aliquots of a ZnCl\(_2\) (13–20 mM; pH 5.5) solution were added to solutions of wild-type BF3 (13 µM) and mutant BF3N (20 µM), each containing 0.5 mM TCEP, pH 5.5. CD spectra were taken at each point in the titration, allowing 5 min for equilibration after each Zn\(^{2+}\) addition. Spectra were recorded over the wavelength range of 195–200 nm with a resolution of 1 nm and as the average of 50 scans.

2.8 Isothermal titration calorimetry (ITC)

2.8.1 Sample preparation

NF:DNA interactions

NF proteins were prepared by dissolving the lyophilised rhPLC-purified peptides in ~600 µl of the appropriate buffer. The pH was increased with the addition of 0.1 M NaOH to 6.5. Native NF54 samples were prepared by pooling the eluate from size exclusion chromatography and decreasing the volume using a Microsep 1K concentrator to ~600 µl. The DNA samples prepared using large-scale purification were concentrated to ~2 ml using Microsep 3K concentrators. Sample concentrations and buffer conditions of NF:DNA experiments are listed in Table 2.7.
Table 2.7. ITC conditions for titrations of wild-type and mutant NF into DNA. All experiments were performed in 10 mM sodium phosphate, 0.5–1 mM DTT and 0.1–0.3 mM ZnSO₄ (25 °C). The sequence of the top strand of DNA was 5’-GCAACTGATCTGGACT-3’.

<table>
<thead>
<tr>
<th>Protein</th>
<th>pH</th>
<th>[NaCl] (mM)</th>
<th>[NF] (µM)</th>
<th>[DNA] (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NF54</td>
<td>6.5</td>
<td>10</td>
<td>129</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>NF48</td>
<td>6.5</td>
<td>10</td>
<td>68</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>130</td>
<td>12</td>
</tr>
<tr>
<td>Native NF54</td>
<td>6.5</td>
<td>160</td>
<td>69</td>
<td>3</td>
</tr>
<tr>
<td>Native NF54</td>
<td>7.5</td>
<td>110</td>
<td>103</td>
<td>4</td>
</tr>
<tr>
<td>V205M</td>
<td>6.5</td>
<td>10</td>
<td>220</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>35</td>
<td>3</td>
</tr>
<tr>
<td>G208S</td>
<td>6.5</td>
<td>10</td>
<td>96</td>
<td>5</td>
</tr>
<tr>
<td>G208S</td>
<td>7.5</td>
<td>0</td>
<td>48</td>
<td>3</td>
</tr>
<tr>
<td>R216Q</td>
<td>6.5</td>
<td>10</td>
<td>47</td>
<td>3</td>
</tr>
<tr>
<td>R216Q</td>
<td>7.5</td>
<td>0</td>
<td>44</td>
<td>3</td>
</tr>
<tr>
<td>D218G</td>
<td>6.5</td>
<td>10</td>
<td>187</td>
<td>5</td>
</tr>
<tr>
<td>D218G</td>
<td>7.5</td>
<td>0</td>
<td>140</td>
<td>5</td>
</tr>
</tbody>
</table>

Ush-F1:NF interactions

The Ush-F1 was synthesised and purified as described in Liew et al. (2000). Ush-F1 was dissolved in ~600 µl and lyophilised NF proteins in ~2 ml of the appropriate buffer. The pH of the samples was adjusted to 5.0 using 0.1 M NaOH. The conditions of these ITC experiments are listed in Table 2.8.

Table 2.8. ITC conditions for titrations of Ush-F1 into wild-type and mutant NF. All experiments were carried out with 20 mM sodium acetate, 0.5 mM TCEP, 1.5 mM ZnSO₄, pH 5.0 and at 25 °C. The extinction coefficient for Ush-F1 is 2,740 M⁻¹ cm⁻¹ at 280 nm.

<table>
<thead>
<tr>
<th>Protein</th>
<th>[Ush-F1] (µM)</th>
<th>[NF] (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NF54</td>
<td>307</td>
<td>18</td>
</tr>
<tr>
<td>V205M</td>
<td>437</td>
<td>26</td>
</tr>
<tr>
<td>G208S</td>
<td>303</td>
<td>20</td>
</tr>
<tr>
<td>R216Q</td>
<td>669</td>
<td>39</td>
</tr>
<tr>
<td>D218G</td>
<td>669</td>
<td>50</td>
</tr>
</tbody>
</table>

FOG:TACC interaction

Lyophilised FOG-F3KRA and TACC77 were dissolved in 10 mM sodium acetate, pH 5.5 containing 1 mM TCEP and 2 mM ZnSO₄ (~500 µl and ~2 ml, respectively), and the sample pH was adjusted to 5.5 (using 0.1 M NaOH). In the two titrations that were
performed, concentrations of FOG-F3\textsubscript{KRA} and TACC77 were 320 \textmu M and 35 \textmu M, and 310 \textmu M and 35 \textmu M, respectively.

In all cases, the samples were extensively dialysed against the same reservoir of the appropriate buffer prior to the experiments using CelluSep H1 1K MWCO dialysis membrane (Membrane Filtration Products Inc, Seguin, Texas). The buffer was then kept and used in control experiments and sample dilution, when required. The concentrations of the samples were carefully determined spectrophotometrically (see Section 2.4.5).

### 2.8.2 ITC parameters

All experiments were performed at 25 °C on a VP-ITC microcalorimeter (MicroCal Inc., Northampton, MA). All samples were degassed by evacuation. Titrations consisted of 15–30 injections of 5–15 \textmu l with intervals of 250–300 s. The reference power was set at 10 \mu cal s\(^{-1}\) and the cell was stirred continuously at 310 rpm. The heat of dilution was measured in a separate experiment in which the high-concentration sample was injected into buffer. The data was analysed using the ORIGIN software supplied with the calorimeter (Microcal Software, Northampton, MA). The baseline was adjusted, when necessary, and the heat of dilution was subtracted from the sample titration. A non-linear least squares fit to a single binding site model utilising the Marquardt algorithm was used to obtain values for the binding constant (\(K_a\)), stoichiometry (\(n\)) and heat of binding (\(\Delta H\)).

### 2.9 Cross-linking studies

Lyophilised TACC77 and Hop (the control protein) were resuspended in 10 mM sodium acetate, 1 mM TCEP, 1 mM ZnSO\(_4\), and pH adjusted to ~5.5. The concentration of the samples were adjusted to ~120 \textmu M (Hop: \(\varepsilon_{280nm}=12,490\) M\(^{-1}\)cm\(^{-1}\)). Stock glutaraldehyde solution (1 %) was prepared by diluting 50 % (v/v) glutaraldehyde (Sigma Aldrich, St. Louis, MO) with buffer. Cross-linking reactions were carried out at room temperature and were started with the addition of 0.03 % glutaraldehyde. Samples were taken at varying time-points and SDS-PAGE loading buffer immediately added to stop the reaction. Equal volumes were loaded onto tricine SDS-PAGE (Section 2.3.2).
2.10 Sedimentation equilibrium

TACC103 samples were prepared in 10 mM sodium phosphate, 150 mM NaCl, 1 mM DTT, pH 7. The samples (9.3, 18.5 and 37.0 µM) were subjected to speeds of 16,000, 20,000 and 24,000 rpm at 4 °C. Absorbances were recorded at 230 and 360 nm. The partial specific volume determined from the amino acid sequence (Perkins, 1986) was 0.743 ml g⁻¹ and buffer density estimated using the SEDNTERP program (Hayes et al., 1995) was 1.006 g ml⁻¹.

BF3E (at concentrations of 8.6, 20.0 and 33.6 µM; all in 1 mM TCEP, 1 mM ZnSO₄, pH 5.6) was centrifuged against a matched buffer at 25 °C at 30,000 and 42,000 rpm. Absorbances were recorded at 230 and 360 nm. The partial specific volume of BF3E was 0.724 ml g⁻¹ and buffer density was 0.997 g ml⁻¹. The presence of a single Zn²⁺ ion was accounted for using the tabulated value for the density of Zn²⁺ (7.14 g ml⁻¹; Aylward and Findlay, 1994).

Sedimentation equilibrium experiments were performed using an Optima™ XL-A analytical ultracentrifuge (Beckman Instruments, Palo Alto, CA) equipped with an An-60ti rotor. Data were recorded in a double-sector cell as absorbance versus radius scans in 0.001-cm increments, and 10 scans were averaged for each dataset. Baseline correction was achieved by the subtraction of data collected at 360 nm. Scans were taken every 3 h and compared to ensure that equilibrium had been reached. Data analysis was performed using the NONLIN software (Johnson et al., 1981). The final parameters were determined by a non-linear least squares fit of the data to a single species model. Goodness-of-fit was determined by examination of the residuals derived from the fit.

2.11 GST-pulldowns

2.11.1 The FOG-F3 and TACC103 interaction

Wild-type FOG-F3 was overexpressed as a GST-fusion protein and immobilised onto glutathione-Sepharose® 4B beads. A negative control using GST alone was performed concurrently. The beads were then washed with wash buffer (50 mM Tris, 100 mM NaCl, 10 % glycerol, 1.4 mM PMSF, 0.1 µl ml⁻¹ β-mercaptoethanol, pH 8.0) and equilibrated with binding buffer (20 mM sodium acetate, 50 mM NaCl, 10 µM
ZnSO₄, 1 mM DTT, pH 5.0). Lyophilised TACC103 was resuspended in binding buffer and pH adjusted to 5.0. Equal amounts of TACC103 were added to beads containing either GST-FOG-F3 or GST. The binding reaction was carried out for ~3 h at 4 °C with shaking. Beads were gently pelleted by centrifugation (<1000 g, 30 s) and unbound proteins were then removed by repeated washing with binding buffer. Samples of the proteins prior to interaction, as well as the beads after washing, were taken for tricine SDS-PAGE analysis (Section 2.3.2).

2.11.2 The GATA-CF and BF3X interaction

Nonidet P-40 and Igepal CA-630 were purchased from Sigma Aldrich (St. Louis, MO).

Wild-type and mutant BF3 proteins (BF3A, D, E, N, Q and R) were overexpressed as GST-fusion proteins and immobilised onto glutathione-Sepharose® 4B beads. GST alone was also bound onto beads to serve as the negative control. The beads were then washed with wash buffer (50 mM Tris, 100 mM NaCl, 10 % glycerol, 1.4 mM PMSF, 0.1 µl ml⁻¹ β-mercaptoethanol, pH 8.0) and equilibrated with pulldown buffer (20 mM Tris, 150 mM NaCl, 0.5 % Nonidet P-40, 10 % glycerol, 10 µM ZnSO₄, 1 mM β-mercaptoethanol, 1 mM PMSF, pH 7.5). Immobilised protein levels were determined by analysing bead samples on glycine SDS-PAGE concurrently with BSA standards (1–10 µg).

Equal volumes of the MBP-GATA-1 CF sample were added to beads containing relatively equal amounts of GST-fusion BF3 proteins. The total volume of the binding reaction was increased 5-fold with the addition of pulldown buffer. The binding reaction was carried out for ~1 h at 4 °C with shaking. The beads were gently pelleted by centrifugation (<1000 g, 30 s) and washed extensively with pulldown buffer. For each reaction, the bead slurry was divided by half and all samples were incubated with 5× SDS-PAGE loading buffer. One sample set was analysed using glycine SDS-PAGE, while the other set was subjected to Western blot analysis.

For Western blot analysis, the samples were electrophoresed using 1-mm gels and the gels were not fixed or stained. The gels were transferred to BioTrace™ NT nitrocellulose transfer membranes (0.45 µm; Pall Corporation) using a Hoefer TE 22
Mighty Small Transphor system (Amersham Biosciences) at 50 mA in transfer buffer (25 mM Tris, 192 mM glycine, 20 % methanol). The following solutions were made in TBST buffer (50 mM Tris, 150 mM NaCl, 0.05 % Igepal CA-630, pH 7.4) and extensive washing with TBST buffer was carried out between steps. After the transfer, the membrane was blocked (3 % skim milk powder, 2 h, RT), exposed to anti-MBP antiserum (1:5000, 1 h, RT; New England Biolabs) and then anti-rabbit IgG-horseradish peroxidase conjugated antibody (1:7500, 1 h, RT; Amersham Biosciences). The resultant complex was visualised using the Renaissance® Western Blot Chemiluminescence Reagent Plus kit (NEN™ Life Science Products, Boston, MA) and images were developed on Kodak diagnostic X-omat™ film (Eastman Kodak Company, Rochester, NY).

2.12 NMR spectroscopy

2-2-Dimethylsilapentane-5-sulfonic acid (DSS) and D₂O (99.96 % isotopic purity) were obtained from Sigma Aldrich (St. Louis, MO). High-precision 528 PP 5-mm O.D. (outer diameter) NMR tubes were from Wilmad (Buena, NJ) and 5-mm O.D. susceptibility-matched microcells from Shigemi (Tokyo, Japan). All other reagents were of analytical grade. NMR experiments were performed on a Bruker AMX-600 (equipped with a 5-mm triple resonance probehead and three-axis pulsed field gradients). Sample temperature was regulated by a Bruker B-VT 1000 temperature control unit.

2.12.1 Spectral processing

NMR data were processed on Silicon Graphics O₂ workstations. Spectra were processed using XWINNMR (Bruker) and analysed using XEASY (Bartels et al., 1995) and SPARKY (Goddard and Kneller). The ¹H frequency scale was directly referenced to DSS at 0.00 ppm, while the ¹⁵N frequency scale was indirectly referenced to liquid ammonia using the ¹H frequency of the DSS resonance (Wishart et al., 1995). Spectral resolution in the directly detected dimension was enhanced by apodisation with a Lorentzian-Guassian window function (LB = 0.1, GB = –3), and shifted squared sine bell functions (60°) in the indirectly detected dimension. Digital resolution was enhanced by zero-filling once in each dimension and linear prediction
(in the case of $^{15}$N dimensions) before Fourier transformation. Polynomial baseline corrections were applied to the processed spectra.

2.12.2 Sample preparation

All NMR samples contained 5 % D$_2$O and 2–3 µM of DSS, and were stored at 4 °C until use.

Wild-type FOG-F3 and FOG-F3$^{KRA}$ used in 2D experiments

Unlabelled FOG-F3 (180 µM), unlabelled FOG-F3$^{KRA}$ (270 µM) and $^{15}$N-labelled FOG-F3$^{KRA}$ (245 µM) samples were prepared by resuspending the lyophilised proteins in 2 mM TCEP and 2 mM ZnSO$_4$. The pH of the samples was carefully increased to 5.0 (for wild-type FOG-F3) or 5.5 (for FOG-F3$^{KRA}$) using 0.1 M NaOH.

2.12.3 NMR experiments

Table 2.9 lists NMR experiments performed throughout this study. Spectral widths were 13 ppm for $^1$H and 35 ppm for $^{15}$N. Water suppression was achieved using pulsed-field gradients.

Table 2.9. NMR experiments. In all experiments, the $^1$H carrier frequency was set to the same frequency as the water peak and the relaxation delay was set to 1 s. These were performed on wild-type FOG-F3$^1$ and FOG-F3$^{KRA}$.$^2$

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Mixing time (ms)</th>
<th>Scans per $t_1$ increment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D</td>
<td>N/A</td>
<td>8–128</td>
</tr>
<tr>
<td>TOCSY (Bax and Davis, 1985)</td>
<td>70</td>
<td>700$^1$ or 512$^2$</td>
</tr>
<tr>
<td>DQF-COSY (Piantini et al., 1982)</td>
<td>N/A</td>
<td>512</td>
</tr>
<tr>
<td>NOESY (Kumar et al., 1980)</td>
<td>150$^1$ or 200$^2$</td>
<td>512</td>
</tr>
<tr>
<td>HSQC (Bax et al., 1990, Norwood et al., 1990)</td>
<td>N/A</td>
<td>128</td>
</tr>
<tr>
<td>3D HNHA (Vuister and Bax, 1993)$^2$</td>
<td>N/A</td>
<td>36</td>
</tr>
</tbody>
</table>

Titrations performed to study the NF:DNA interaction

$^{15}$N-NF48 (315 µM) was added to unlabelled double-stranded 16-bp DNA (210 µM; sequence of the top strand was 5'-GCAACTGATCTGGACT-3'). $^{15}$N-HSQC spectra were collected at protein:DNA ratios of 0, 0.25, 0.5, 0.75 and 1.0 (288 K). The
titration was performed in 10 mM sodium phosphate, 50 mM NaCl, 2.5 mM TCEP, 2.5 mM ZnSO₄, pH 5.5.

The titration of ¹⁵N-NF54 (180 µM) was carried out in a similar manner as described for the ¹⁵N-NF48 protein. ¹⁵N-HSQC spectra were recorded at protein:DNA ratios of 0, 0.5 and 1.0 (275 and 288 K). The buffer conditions were 10 mM sodium acetate, 10 mM NaCl, 0.5 mM ZnSO₄, 0.5 mM TCEP, pH 6.0. In both cases, titrations ended at stoichiometric ratio of 1:1 protein:DNA, as the presence of excess protein in the mixture resulted in precipitate formation.

**Titrations performed to study the FOG:TACC3 interaction**

In these experiments, the samples to be mixed together were dialysed against the same reservoir of buffer (10 mM sodium acetate, 2 mM TCEP, 1 mM ZnSO₄, pH 5.5) to ensure that any perturbations observed in spectra were not due to changing buffer conditions.

Unlabelled FOG-F₃KRA (185 µM) was added to ¹⁵N-labelled TACC77 (145 µM) with ¹⁵N-HSQC spectra collected at the FOG:TACC ratio of 0, 0.25, 0.5, 1.0 and 1.3 (298 K). The mixture was concentrated using a Microsep 1K concentrator (Pall Corporation) to increase the concentration of the sample. The titration was continued with the addition of more FOG-F₃KRA (200 µM) with spectra recorded at concentration ratios of 1.8, 2.6 and 6.3 (298 K). Spectra at 275 K were collected of the initial ¹⁵N-TACC77 sample and of the mix at the end of the titration.

Unlabelled TACC77 (330 µM) was added to ¹⁵N-FOG-F₃KRA (110 µM) with ¹⁵N-HSQC spectra recorded at 0, 0.5, 1.1, 1.6, 2.0, 2.5 and 3.0 molar equivalents of TACC77 to ¹⁵N-FOG-F₃KRA (298 K). This was carried out in 10 mM sodium acetate, 2 mM TCEP, 1 mM ZnSO₄, pH 5.5.

**2.12.4 Structure determination of FOG-F₃KRA**

**Resonance assignments**

The sequential resonance assignment method (Wüthrich, 1986) was used to obtain sequence-specific assignments of wild-type FOG-F3 and FOG-F₃KRA using DQF-COSY, TOCSY and NOESY experiments. The proton chemical shifts were
obtained by averaging the chemical shifts of all peaks with the same resonance assignment in the NOESY experiment.

**Interproton distance restraints**

Interproton distance restraints were derived from the volume of crosspeaks in the NOESY spectrum. Crosspeaks were integrated in XEASY and converted to upper distance limits using the CALIBA module of DYANA (Güntert et al., 1997). Calibration of peak integrals against distance was performed using standard DYANA parameters. The UNIX shell script ‘hbcorr’ (developed in-house) was used to add 0.5 Å to all distances involving Hβ atoms in order to account for spin diffusion.

**Dihedral angle restraints**

$^{3}J_{\text{NH}}\alpha$ coupling constants were measured by analysis of the HNHA spectrum (Vuister and Bax, 1993). Crosspeaks in this spectrum (arising from correlations between the HN and Hα protons of the residues) were assigned using the chemical shifts of HN and Hα protons identified during the sequential assignment process. Diagonal (HN) peaks and their corresponding crosspeaks (HN-Hα) were integrated in three dimensions using XEASY. The ratio of intensities is related to the coupling constant by the following equation (Vuister and Bax, 1993):

\[
\frac{I_{\text{cross}}}{I_{\text{diag}}} = -\tan^2\left(2\pi \cdot 3 J_{\text{NH}\alpha} \cdot \zeta\right)
\]  

(Equation 2.3)

where $I_{\text{cross}}$ and $I_{\text{diag}}$ are volumes of the crosspeak and diagonal peak, respectively, and $\zeta$ is a dephasing delay (13.05 ms). The calculated coupling constants were increased by 11% to account for the effects of cross-relaxation.

For residues whose $^{3}J_{\text{NH}\alpha}$ coupling constants could not be determined using the method described above, the INFIT algorithm (Szyperski et al., 1992) that is part of the XEASY package was used to estimate the coupling constant.

**Iterative assignment of NOESY spectra**

NOESY crosspeaks that were unambiguously assigned were used as input for initial structure calculations, which generated structures with a preliminary backbone fold. For the ambiguous NOEs, each possible assignment was checked against the
generated structures. Assignments consistent with the fold were used as input in subsequent structure calculations and restraints that did not give rise to structural violations were retained. The process of assigning ambiguous NOEs that remained after each round of structure calculations was repeated until as many ambiguous NOEs as possible had been assigned without generating significant structural violations.

**Initial structure calculations using DYANA**

All determined interproton distances and $^{3}J_{\text{H}_{\alpha}}$ coupling constants were used to generate a set of allowable $\phi$ dihedral angles in the GRIDSEARCH module of DYANA. The ANNEAL module was then used to generate 100–1000 structures from random starting conformations by simulated annealing using torsion angle dynamics. The 20–50 lowest energy structures were then examined in order to assign more NOEs. These structure calculations were carried out in the absence of zinc.

**Final structure calculations using ARIA 1.2**

Further structural refinement was performed in an automated manner using ARIA 1.2 (Nilges, 1995, 1997a) implemented in CNS 1.1 (Brünger et al., 1998). Zinc was incorporated into the calculations by introducing distance and angle restraints characteristic of a tetrahedral coordination sphere (Neuhaus et al., 1992). The zinc-ligating atoms were identified by examination of the initial structures calculated in the absence of zinc ($\text{S}^{\gamma}$ of Cys335, Cys338 and $\text{N}^{\varepsilon_{2}}$ of His351, 355). The bond lengths for $\text{S}^{\gamma}$-Zn(II) and $\text{N}^{\varepsilon_{2}}$-Zn(II) were fixed at 2.298 and 1.993 Å, respectively, with force constants of 250 kcal/mol$^{-1}$/Å$^{-2}$. Bond angles defining the Zn(II) coordination site were constrained to the following values: $\text{S}^{\gamma}$-Zn(II)-$\text{S}^{\gamma}$ = 112.0°, $\text{C}^{\beta}$-$\text{S}^{\gamma}$-Zn(II) = 108.0°, $\text{S}^{\gamma}$-Zn(II)-$\text{N}^{\varepsilon_{2}}$ = 111.0°, $\text{C}^{\varepsilon_{1}}$-$\text{N}^{\varepsilon_{2}}$-Zn(II) = 125.2°, $\text{C}^{\delta_{2}}$-$\text{N}^{\varepsilon_{2}}$-Zn(II) = 128.8° and $\text{N}^{\varepsilon_{2}}$-Zn(II)-$\text{N}^{\varepsilon_{2}}$ = 102.0°, each with a force constants of 50 kcal/mol$^{-1}$/deg$^{-2}$. The energies of these bond lengths and angles were modelled as harmonic quadratic functions.

In the first iteration, 50 structures were calculated using the manually assigned NOEs as soft restraints. The cut-off value of peak volume for ambiguous assignment was reduced from 1.01 for the first iteration to 0.80 in iteration 8 (Nilges et al., 1997b).
Calculations were performed in the simplified all-hydrogen PARALLHDG5.2 force field with non-bonded interactions modelled by PROLSQ force field (Linge and Nilges, 1999). Floating chirality assignment (Folmer et al., 1997) was used for all methylene and isopropyl groups, since no stereo-specific assignments were made. The final assignments made by ARIA 1.2 were checked and corrected manually where necessary. In the final set of calculation, the 100 lowest energy structures were refined in a 9-Å shell of water molecules (Jorgensen et al., 1983). The 20 conformers with the lowest value of $E_{\text{tot}}$ were visualised and analysed using the programs MOLMOL (Koradi et al., 1996) and PROCHECK (Laskowski et al., 1996). The family of 20 lowest energy structures has been deposited in the Protein Data Bank (PDB code 1SRK).

2.13 Atomic absorption spectrometry (AAS)

Lyophilised wild-type BF3 and mutant BF3D were dissolved in 10 mM sodium acetate, 1 mM ZnSO$_4$, 1 mM DTT and the pH of the samples was adjusted to 5.5. The samples were dialysed extensively against a zinc-free buffer (10 mM sodium acetate, 1 mM DTT, pH 5.5). The folded nature of both samples was confirmed by CD spectropolarimetry. The Zn$^{2+}$ contents of BF3 (26 µM) and BF3D (11 µM) were measured using a Varian SpectrAA 20 Plus atomic absorption spectrometer at 213.9 nm. Calibration standards were diluted manually from stock solution of ZnNO$_3$ Spectrosol® (AAS grade; BDH Chemicals).

2.14 Electrophoretic mobility shift assays (EMSAs)

DNA probes were end-labelled with $^{32}$P according to standard procedures (Sambrook et al., 1989) using T4 polynucleotide kinase (New England Biolabs, EC 2.7.1.78). To test for interaction with BKLF proteins, the sequence of the top strand of the double-stranded DNA probe was 5'-TAGAGCCACACCCTGGTAAG-3'.

In a normal EMSA, reactions were set up in a total volume of 30 µl, comprising of 0.1 pg of $^{32}$P-labelled DNA probe, ~100 ng of recombinant protein, 10 mM Hepes, pH 7.8, 50 mM KCl, 5 mM MgCl$_2$, 1 mM EDTA and 5 % glycerol. After incubation on ice for 10 min, the samples were loaded onto a 6 % native polyacrylamide gel made up in 0.5× TBE (45 mM Tris, 45 mM boric acid, 1 mM EDTA and 5 µM
ZnSO₄). The gel was subjected to electrophoresis at 250 V for 2 h at 4 °C. The gel was then transferred to Whatman 3MM paper and dried under vacuum on a gel drier for 30 min. Dried gels were scanned on a PhosphoImager™ screen (Molecular Dynamics, Sunnyvale, CA) and data analysed and quantified when necessary using ImageQuant (Molecular Dynamics).

In quantitative EMSAs, 67 pM ³²P-labelled DNA and 0–2000 nM wild-type BF3 or 0–4500 nM mutant BF3A protein were used. In addition to the buffer conditions described above, these reactions also included 1 mg ml⁻¹ dl-dC (Amersham Biosciences), 5 mg ml⁻¹ bovine serum albumin and 0.5 %(v/v) Nonidet P-40 (Sigma Aldrich).
Chapter 3. The role of the GATA-1 N-terminal zinc finger in transcriptional regulation

3.1 Introduction

3.1.1 The GATA protein superfamily

GATA proteins are a group of evolutionary conserved transcriptional regulators that contain either one or two zinc finger domains. These domains bind to (T/A)GATA(A/G) sequences in the regulatory regions of target genes. There are six vertebrate GATA factors. GATA-1, -2 and -3 are predominantly expressed in haematopoietic cells where they regulate gene expression in erythroid cells, megakaryocytes and T-lymphocytes (Orkin, 1992). GATA-4, -5 and -6 are detected in various mesoderm- and endoderm-derived tissues, such as the heart, liver, lung, gonads and gut (reviewed in Molkentin, 2000). In *Drosophila*, the GATA factors, Serpent and Pannier, are involved in the formation of the eye, heart, blood cells and in sensory bristle patterning (reviewed in Patient and McGhee, 2002).

The founding member of this family, GATA-1, is indispensable for the maturation of the erythroid and megakaryotic lineages (reviewed in Cantor and Orkin, 2002). Functionally important GATA DNA sequences have been identified in promoter or enhancer elements of virtually all erythroid- and megakaryocytic-expressed genes, and GATA-1 null mice show complete ablation of embryonic erythropoiesis due to arrested maturation and apoptosis of erythroid precursors (Fujiwara et al., 1996). Together, these data indicate a key role for GATA-1 in erythroid development.

3.1.2 The different roles of the zinc fingers in GATA-1

The two highly conserved GATA-type zinc fingers in GATA-1 are required for the recognition of both DNA and protein partners. The more C-terminal finger (C-finger or CF) recognises GATA motifs in DNA and can interact directly with PU.1 (Rekhtman et al., 1999), Krüppel-like factor proteins (Merika and Orkin, 1995) and CREB-binding protein (a histone acetylase known to associate with chromatin remodeling complexes; Blobel et al., 1998).
The role of the GATA-1 N-terminal zinc finger in transcriptional regulation

The N-terminal finger (N-finger or NF) contributes to the specificity and stability of interactions between GATA-1 and DNA containing tandem or overlapping GATA sites (Trainor et al., 1996, 2000), although it cannot bind GATA sites independently. Recently however, it was shown that N-finger can independently bind DNA sites containing GATC sequences (Newton et al., 2001). In addition to this DNA-binding ability, NF also mediates interactions with FOG (Tsang et al., 1997), Fli-1 (Eisbacher et al., 2003), c-Myb (Takahashi et al., 2000) and LMO2 (Osada et al., 1995).

3.1.3 FOG is a friend of GATA

FOG is co-expressed with GATA-1, -2 and -3 in developing haematopoietic cells (Tsang et al., 1997). Mice lacking FOG exhibit blocked erythropoiesis, similar to GATA-1-deficient mice (Tsang et al., 1998), as well as problems with megakaryocyte development. Direct evidence that FOG is an essential cofactor for GATA-1 in promoting erythroid development was provided by Crispino et al. (Crispino et al., 1999). Ablation of the GATA-FOG interaction through the introduction of mutations into GATA-1 abolished its ability to drive erythroid differentiation, and this activity was rescued with compensatory FOG mutations that reinstated the GATA-FOG interaction. Another vertebrate FOG protein, FOG-2, is coexpressed with GATA-4, -5 and -6, and its interaction with GATA-4 is necessary for proper heart morphogenesis (Lu et al., 1999, Svensson et al., 2000).

The Drosophila FOG protein Ushaped interacts with the GATA factors Pannier and Serpent to control the specification of heart and blood cells, respectively, in the fly (Haenlin et al., 1997, Fossett et al., 2001). A FOG homologue has also been identified in Xenopus, and its wide expression profile suggests that the two mammalian FOG proteins have evolved from a common ancestor that resembles the Drosophila and Xenopus FOG (Deconinck et al., 2000).

3.1.4 The GATA-FOG interaction

The N-finger of GATA proteins mediate the interaction between GATA-1 and FOG. The nature of this interaction appears to have been conserved during evolution, since GATA-1 NF has been shown to bind murine FOG and FOG-2 as well as Drosophila
Ushaped (Fox et al., 1999). The residues of GATA-1 NF that are important in its interaction with FOG have been identified through complementary mutagenesis and NMR experiments and they include V205, G208, A209, H222 and Y223 (Fox et al., 1998, Liew et al., 2000, Kowalski et al., 2002). It is notable that these residues are conserved among N-fingers and absent from C-fingers of GATA proteins, which explains why, despite ~50% identity between the two fingers, only N-finger is able to recognise FOG proteins.

FOG proteins contain multiple zinc finger domains. While all FOG fingers contain the conserved hydrophobic residues and ligand spacing characteristic of classical CCHH fingers, several of these domains contain a variant CCHC zinc ligation topology. Several ZnFs of FOG are able to bind GATA NF, and remarkably all GATA-interacting FOG fingers are of the CCHC type (Figure 3.1). The GATA-FOG interaction appears to be highly conserved, as mFOG and FOG-2 are able to bind NFs of mGATA-1, -2 and hGATA-3, and mGATA-1, -2, -3 and -4, respectively (Tsang et al., 1997, Tevosian et al., 1999).

![Figure 3.1. Zinc fingers of mFOG and Ushaped.](image)

The GATA-FOG binding interfaces have been delineated by mapping residues known to be essential for the interaction onto structures of GATA-1 NF and Ushaped finger 1 (Figure 3.2A). Because residues identified to be important are well-conserved and that fingers of related proteins show similar binding properties, these data strongly suggest that the identified binding interfaces represent the mode of interaction for all GATA-FOG protein complexes.
Figure 3.2. The GATA-FOG complex. A. The GATA-FOG binding interfaces. Residues necessary for the interaction are mapped onto the structures of mGATA-1 NF and Ushaped finger 1 using space-fill representation (PDB codes 1GNF and 1FV5; Fox et al., 1998, Fox et al., 1999, Liew et al., 2000, Kowalski et al., 1999). B. The solution structure of the GATA-FOG complex. GATA-1 NF (blue) is bound to Ushaped finger 1 (red; A. Kwan, C. K. Liew and J. Mackay, unpublished results). The proteins are displayed with their secondary structure elements, zinc ligands in blue and zinc ions in magenta.

Recently, the solution structure of the GATA-FOG complex was determined in our laboratory (A. Kwan, C. K. Liew and J. Mackay, unpublished results; Figure 3.2B). This structure not only supports earlier findings which delineated residues important for the GATA-FOG interaction, but also illustrates the relative orientation of the zinc fingers within the complex and reveals for the first time how classical and GATA-type zinc fingers can act as protein recognition motifs.

3.1.5 The GATA-DNA interaction

The surface used by N-fingers from GATA family proteins to contact DNA are likely to be the same as that used by the chicken GATA-1 C-finger (cCF; Omichinski et al., 1993), given the conservation of residues important for binding (Omichinski et al., 1993, Lowry and Atchley, 2000). The solution structure of the C-finger of chicken
GATA-1 bound to a 16-bp oligonucleotide revealed that both the α-helix and a loop region located between the two β-hairpins contact the DNA major groove with the C-terminal basic tail wrapping around into the minor groove (Omichinski et al., 1993; Figure 3.3). Overall, therefore, the surfaces used by the GATA-1 N-finger to contact both FOG and DNA are well understood.

3.1.6 Naturally occurring mutations of GATA-1 N-finger

The interactions mediated by the GATA-1 N-finger have become a topic of some interest with the recent association of several naturally occurring N-finger mutations with human familial blood disorders. The GATA-1 gene resides on the X chromosome, and males carrying defective GATA-1 alleles exhibit various haematological abnormalities, including anaemia, thrombocytopenia and β-thalassaemia (Table 3.1).

Table 3.1. Naturally occurring mutations of GATA-1 N-finger.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Phenotype</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>V205M</td>
<td>Severe thrombocytopenia (decrease in platelet levels) and dyserthropoietic anaemia (increased levels of immature erythroid precursors and mature erythrocytes were abnormal in size and shape)</td>
<td>Nichols et al., 2000</td>
</tr>
<tr>
<td>G208S</td>
<td>Thrombocytopenia and mild dyserthropoiesis</td>
<td>Mehaffey et al., 2001</td>
</tr>
<tr>
<td>R216Q</td>
<td>Thrombocytopenia and β-thalassaemia (decreased levels of β-globin)</td>
<td>Yu et al., 2002, Balduini et al., 2004</td>
</tr>
<tr>
<td>D218G</td>
<td>Thrombocytopenia and mild dyserthropoiesis</td>
<td>Freson et al., 2001</td>
</tr>
<tr>
<td>D218Y</td>
<td>Severe thrombocytopenia and marked anaemia</td>
<td>Freson et al., 2002</td>
</tr>
</tbody>
</table>
Biochemical binding studies show that the R216Q mutation affects the DNA-binding ability of GATA-1, while the remaining mutations appear to impair FOG binding (refer to references listed in Table 3.1). It is likely that an understanding of how the mutations differentially affect the interactions N-finger make with either DNA or FOG proteins will help explain the different phenotypes.

### 3.1.7 Specific aims of this study

The aim of this Chapter is to examine in detail the effects of the naturally occurring mutations on the binding properties of N-finger, focusing on interactions with DNA and a FOG-family CCHC ZnF, Ushaped finger 1 (Ush-F1). It is hoped that results obtained in this comparative study will improve the understanding of the molecular basis for the blood disorders.

### 3.2 Characterisation of wild-type N-finger

In order to study the interaction between N-finger and DNA, an N-finger construct that contains the C-terminal basic region was required. This was realised as GATA-1 NF that encompassed residues 200–243 did not exhibit DNA binding (M. Crossley, personal communication) and also, residues at the C-terminal region of chicken GATA-1 CF were required for DNA binding (Figure 3.4). Two protein constructs that differ in the length of their basic tails, NF48 (residues 200–248 of murine GATA-1) and NF54 (residues 200–254), were synthesised recombinantly.

**Figure 3.4. Sequence of murine GATA-1 NF.** Cysteine residues involved in zinc ligation are boxed in grey and natural mutations are shown by arrows. Residues implicated in mediating FOG binding are indicated by open circles, while those that align with the residues of cC-finger that contact DNA in the cCF:DNA complex (Omichinski et al., 1993) are indicated by closed circles. The positive residues at the C-terminal region are underlined.
3.2.1 Purification of NF48 and NF54

The proteins were expressed as GST-fusion proteins in either pGEX-2T (NF48) or pGEX-6P (NF54). High levels of overexpression were obtained for both proteins at 25 °C (Figure 3.5, lane 2 of both gels). After GSH-affinity chromatography, GST-NF48 was eluted from the beads and cleaved with thrombin in solution (on-column cleavage resulted in NF48 adhering to the Sepharose beads). GST-NF54 was successfully cleaved on-column with PreScission™ protease at 4 °C for ~48 h. Thrombin and PreScission™ proteases were able to cleave GST-NF48 and GST-NF54 specifically, releasing peptides corresponding to the full-length proteins. The protease treatments resulted in ~50 % cleavage for GST-NF48 using thrombin and ~95 % for GST-NF54 using PreScission™ (Figure 3.5, lane 8).

Both NF48 and NF54 were then purified to >95 % purity using reverse phase high performance liquid chromatography (rpHPLC). NF54 was also purified without denaturation by size exclusion chromatography. The molecular masses (M) of the purified proteins were confirmed by electrospray mass spectrometry (NF48: M_{obs.} = 5831.6 ± 0.7 Da, M_{theo.} = 5831.8 Da; NF54: M_{obs.} = 6754.8 ± 0.9 Da, M_{theo.} = 6753.9 Da).

---

**Figure 3.5.** Tricine SDS-PAGE analysis of glutathione-affinity chromatography of NF48 (left) and NF54 (right). Lane 1: MW standards, 2: total protein overexpressed, 3: insoluble fraction, 4: soluble fraction, loaded onto the glutathione beads, 5: sample flow through, 6: wash flow through, 7: beads prior to thrombin cleavage, 8: after cleavage and in the case of NF48, lanes 9–11: elution of free protein. Sizes of the standards, as well as the position of GST-NF fusion, GST and free NF proteins, are marked.
3.2.2 Circular dichroism (CD) analysis of NF48 and NF54

Far-UV CD spectropolarimetry is a form of optical spectrometry that is commonly used to measure the amount of secondary structure in proteins. The phenomenon of circular dichroism arises from differences in the way in which chromophores (such as the backbone amides in a protein) in a chiral environment absorb left- and right-circularly polarised light.

Idealised far-UV CD spectra of proteins containing only α-helices and only β-sheets, as well as that of a random polypeptide chain, are shown in Figure 3.6A. A CD spectrum may be deconvoluted into contributions from the various forms of secondary structure (Fasman, 1996). While a limitation of CD is the relatively low information content, which can affect the accuracy to which the amount of secondary structure is predicted, it is a fast and useful method to determine the approximate structure content of a protein, and requires only a small amount of material.

![Figure 3.6. CD spectra of NF48 and NF54. A. Idealised CD spectra of proteins containing none or only one type of secondary structure (adapted from Fasman, 1996). B. CD spectra of NF48 and NF54. Both protein samples contained 1 mM TCEP and 1 mM ZnSO\(_4\). The CD spectrum of the unfolded protein was obtained at pH ~3, while the pH was adjusted to ~5.5 for the spectra of the folded proteins.](image)

The lyophilised NF48 and NF54 proteins were resuspended in buffer containing reducing agent and Zn\(^{2+}\), and the pH of each sample was adjusted to ~5.5. These conditions were previously determined to be suitable for the folding of a smaller construct of mGATA-1 N-finger containing residues 200–243 (Kowalski et al., 1999). The CD spectra showed that the refolded NF54 and NF48 proteins both contained secondary structure compared to the apoprotein, as judged by the red-shift of the
minimum from ~198 to ~203 and ~205 nm, and the appearance of a shoulder at ~230 nm (Figure 3.6B). Further, the overall shapes of the CD spectra suggest that the two N-finger proteins contained both α-helical and β-sheet elements. The spectrum of NF48, which displayed a minimum at a longer wavelength (~205 nm), indicates the presence of a higher proportion of structure for this domain, most likely due to the longer unstructured tail region in NF54.

3.2.3 One-dimensional NMR spectra of NF48 and NF54

NMR spectroscopy relies on the absorption of radiofrequency radiation by the nuclei of elements such as 1H, 15N and 13C when samples containing these elements are placed in a magnetic field. The absorption frequency of a nucleus is highly sensitive to its physical and chemical environments, and a plot of signal intensity versus absorption frequency is termed a one-dimensional (1D) NMR spectrum.

1D 1H NMR spectra of refolded NF48 and NF54, as well as the NF54 that was purified by non-denaturing size exclusion chromatography, are shown in Figure 3.7. The signals are dispersed, suggesting that the proteins contained well-organised tertiary structure. Also, the linewidths of the signals are relatively sharp and indicate that the proteins were not undergoing aggregation and were probably monomeric in solution. The comparison also shows that the refolded proteins have taken up conformations that were very similar to that of the natively purified NF54, judged by the overall similarity of their spectra (Figure 3.7).

![Figure 3.7. 1D 1H NMR spectra of NF48 and NF54. The amide region of refolded NF48 and NF54 at pH 5.2 and 5.5, respectively, as well as that of the natively purified NF54 at pH 6.5 are displayed. The difference in sample pH may have contributed to minor differences in the spectra.](image-url)
3.3 The interaction of wild-type NF with DNA

Having determined that the recombinant NF domains appear to be folded in solution, we next assessed the ability of these proteins to bind DNA containing a core GATC sequence.

3.3.1 Isothermal titration calorimetry (ITC)

ITC is a powerful tool for direct thermodynamic measurements of binding reactions (Wiseman et al., 1989). A typical experiment is carried out by stepwise addition of one reactant (ligand) into the reaction cell containing the other reactant (macromolecule). The chemical reaction that ensues either releases or absorbs an amount of heat proportional to the fraction of bound ligand and the characteristic binding enthalpy. The ITC instrument measures the heat as the amount of power necessary to maintain constant temperature between the reaction and reference cells.

Figure 3.8 (upper panels) shows typical ITC profiles when NF proteins were injected into the reaction cell containing a 16-bp DNA duplex with a GATC motif. Injection pulses were initially negative, indicating an exothermic interaction. As the titration proceeded, the amount of free DNA in the reaction cell progressively decreased until complete saturation was reached. At this stage, subsequent injections produce similar sized pulses that correspond to dilution or mechanical effects. The heat evolved due to the NF:DNA interaction was obtained by integrating the peak area and subtracting away the heat of dilution. This was then plotted as a function of the molar ratio of NF to DNA duplex, as shown in the lower panels in Figure 3.8. The binding isotherms were fitted to a 1:1 binding model by using a non-linear least-squares method and the stoichiometry of binding, $n$, the binding constant, $K_a$, and the enthalpy change, $\Delta H$, were obtained.

Wild-type NF54 bound its cognate DNA sequence with an association constant of $(3.3 \pm 0.9) \times 10^7$ M$^{-1}$. This value was not significantly affected by the absence of five residues (including two basic residues) from the C-terminus, as the NF48 protein was able to bind DNA with a $K_a$ of $(2.0 \pm 10^6) \times 10^7$ M$^{-1}$. The interaction between N-finger and DNA was enthalpically driven as indicated by the negative $\Delta H$ values (Table 3.2).
The role of the GATA-1 N-terminal zinc finger in transcriptional regulation

Figure 3.8. ITC data showing titrations of A. NF54 and B. NF48 into DNA. Both experiments were carried out at pH 6.5, 10 mM NaCl and at 25 °C. Upper panels show the raw data with the integration baseline in red. Lower panels are data after peak integration and subtraction of the heat-of-dilution control, and the red lines represent the fit to a 1:1 binding model.

NF54 that was purified using non-denaturing methods was also tested for its DNA-binding properties. The native NF54 bound the DNA duplex with an association constant of $1.1 \times 10^6$ M$^{-1}$, 10-fold less than that of refolded NF. The pulses were negative and exothermic throughout the titration of native proteins, whereas titration of refolded proteins gave rise to exothermic pulses at the beginning that became positive and endothermic by the end. The endothermic pulses may represent a weak second binding reaction between DNA and refolded proteins, and may explain the 10-fold higher affinity displayed by the refolded proteins. It is notable that a primary strong and exothermic site and a secondary weak and endothermic site have been described for the trp repressor-operator and the GCN4-ATF/CREB site interactions (Ladbury et al., 1994, Berger et al., 1996).
3.3.2 N-finger binds DNA in pH- and salt-dependent manners

The DNA-binding activity of N-finger was dependent on both the ionic strength and pH. ITC experiments carried out at pH 6.5 with a 20 mM ionic strength indicated that NF bound DNA with a $K_a$ of $\sim 10^7$ M$^{-1}$ (NF48: $2.0 \times 10^7$ M$^{-1}$; NF54: $3.3 \times 10^7$ M$^{-1}$). When the ionic strength was increased to 80 mM, the binding was affected such that a well-defined binding isotherm was not observed, and DNA binding was abolished altogether at an ionic strength of 160 mM (at pH 6.5; Figure 3.9B). An increase in the ionic strength therefore reduced the affinity of the NF:DNA interaction substantially.

To determine the effect of pH on binding, an ITC experiment was performed at pH 7.5 and a 120 mM ionic strength. Under these conditions, native NF54 bound DNA with a $K_a$ of $1.9 \times 10^6$ M$^{-1}$ (Figure 3.9C), which is comparable to binding at pH 6.5 with 20 mM ionic concentration. This suggests that the DNA-binding affinity of N-finger increases with pH (from 6.5 to 7.5), and that this increase was able to compensate for the loss in affinity with increased ionic concentration. Note that pH 6.5 was used initially because the quality of NMR spectra of NF is substantially higher at the lower pH.

Overall, the affinity of GATA-1 NF for DNA containing a GATC motif under ‘physiological’ conditions is $10^6$–$10^7$ M$^{-1}$.
The role of the GATA-1 N-terminal zinc finger in transcriptional regulation

Figure 3.9. ITC data showing the dependence of DNA-binding on pH and ionic concentration. A. NF54 titrated into DNA at pH 6.5 with an ionic strength of 20 mM. B. NF48 titrated into DNA at pH 6.5 and 160 mM ionic concentration. C. NF54 titrated into DNA at pH 7.5 with an ionic strength of 120 mM. The raw (upper) and integrated (lower panels) data are displayed, with the fits to a 1:1 binding model also shown.

3.3.3 The NF:DNA complex was not amenable for NMR studies

Since the NF proteins were well-folded and monomeric in solution, we sought to study its interaction with DNA in more detail using NMR experiments. $^{15}$N-labelled NF48 and NF54 were produced in order to carry out chemical shift perturbation experiments (reviewed in Zuiderweg, 2002). A $^{15}$N-$^1$H heteronuclear single quantum correlation (or $^{15}$N-HSQC) spectrum is a two-dimensional map that correlates the chemical shifts of the amide protons with those of the attached amide nitrogens. Because the positions of signals in an NMR experiment are highly dependent on chemical environment, the addition of a binding partner will perturb the positions of signals corresponding to residues involved in binding. The titration can therefore provide a detailed picture of the binding interface, as long as significant structural rearrangements do not take place upon binding. $^{15}$N-HSQC spectra of $^{15}$N-NF48 and $^{15}$N-NF54 contain some well-separated signals (arising from structured parts of the proteins) with others clustered around frequencies observed for residues in random coil configurations (Figure 3.10, boxed regions). The spectra are therefore consistent with a folded ZnF domain flanked by flexible tail regions.
Because initial experiments revealed that the addition of DNA to protein resulted in precipitation, $^{15}$N-labelled NF proteins were added to the DNA in subsequent titrations. By adding no more than up to equimolar amounts of DNA, precipitation was minimised. A similar approach had been taken previously for the formation of the cCF:DNA complex (Omichinski et al., 1993).

The addition of DNA resulted in the disappearance of many peaks in the spectra of both $^{15}$N-NF48 and $^{15}$N-NF54 (Figure 3.10). In both cases, peaks that were clustered around the random coil frequencies remained, while many of the dispersed peaks broadened to such an extent that they could not be seen. These changes confirmed that the proteins interact with DNA, but precluded the identification of the residues involved in the binding surfaces. Attempts to improve the quality of the spectra by changing temperature (2–25 °C), pH (5.5–7.0) and salt concentration (0–150 mM) were unsuccessful, and time did not permit further attempts to optimise conditions for these experiments.
Figure 3.10. $^{15}$N-HSQC titrations of DNA into $^{15}$N-NF. A. $^{15}$N-HSQC spectrum of $^{15}$N-NF48 at pH 5.5 with 50 mM NaCl (315 µM, 15 °C). B. $^{15}$N-HSQC spectrum of $^{15}$N-NF54 at pH 6.0 with 10 mM NaCl (180 µM, 3 °C). Upper and lower panels show spectra of the proteins alone and with one molar equivalent of DNA duplex, respectively. The boxed regions contain signals at the random coil frequencies.
3.4 The structural and functional effects of naturally occurring mutations in GATA-1

To date, five mutations that reside in the N-finger domain of human GATA-1 have been identified. These mutations are V205M (Nichols et al., 2000), G208S (Mehaffey et al., 2001), R216Q (Yu et al., 2002, Balduini et al., 2004), D218G (Freson et al., 2001) and D218Y (Freson et al., 2002). In order to assess the effects of these mutations on the structure and function of GATA-1, proteins encompassing residues 200–254 of GATA-1 with single substitution mutations (V205M, G208S, R216Q, D218G, D218Y) were synthesised and purified in the same manner as wild-type NF54. The solution properties of these mutants were compared to that of the wild-type protein and four of the mutant proteins were assessed for their abilities to bind both DNA and a FOG-family finger, Ushaped finger 1 (Ush-F1).

The ZnF domain of Drosophila FOG (Ushaped) was used in these experiments because its solution properties were more favourable than those of ZnFs of murine FOG. Previous experiments in the laboratory showed that murine FOG fingers capable of interacting with GATA-1 (FOG-F1, -F6 and -F9) tended to homoaggregate in solution and exhibited poor solubility (Matthews et al., 2000). Drosophila Ushaped fingers however did not display these properties (Liew et al., 2000).

3.4.1 Mutant NF proteins were only partially structured

Far-UV CD spectra of the mutant proteins show that, while the mutants contained some secondary structure as compared to the unfolded protein, they were not as structured as the wild-type NF54. This was inferred from the observation that the wavelength of the minimum in each of the spectra of the mutants was between those of the folded and unfolded wild-type domain (Figure 3.11).
Next, the mutant NF proteins were assessed by 1D $^1$H NMR spectroscopy (Figure 3.12). Each showed some chemical shift dispersion in the downfield region of the spectrum (8.8–10.0 ppm), indicative of a degree of tertiary order, although, with the exception of R216Q mutant, the extent of this dispersion was less than for wild-type NF54. The presence of additional peaks around 10.2 ppm suggests that some regions of the mutant proteins existed in different conformational forms, similar to that reported for the wild-type protein (Kowalski et al., 1999).

**Figure 3.11. CD spectra of the mutant NF proteins.** The samples contained 1 mM TCEP and 1 mM ZnSO$_4$. The pH of the folded samples was ~5.5, while the unfolded sample had a pH of ~2.0.

**Figure 3.12. 1D $^1$H NMR spectra of mutant NF proteins.** The protein samples were 35–480 µM in concentration and contained 1 mM TCEP and 1 mM ZnSO$_4$, pH ~5.5. The spectra were all recorded at 15 °C.
Overall, the CD and NMR results are consistent with the mutant proteins being somewhat less structured than the wild-type NF54, with the R216Q and V205M proteins being the most and least structured, respectively. Having established the conformational preferences of these mutants, we next proceeded to assess their function.

3.4.2 The R216Q mutation impaired DNA binding

The ability of the partially structured mutant NF proteins to interact with DNA containing a GATC site was examined by ITC. These experiments were all carried out at 25 °C with NF proteins titrated into the reaction cell containing DNA.

The V205M protein bound DNA with a similar association constant to the wild-type NF54 ($K_a^{V205M:DNA} = 2.0 \times 10^7$ M$^{-1}$ and $K_a^{NF54:DNA} = 3.3 \times 10^7$ M$^{-1}$ at pH 6.5), indicating that DNA-binding ability was not compromised by the partially structured
nature of the mutant and that the sidechain substitution did not affect the ability of N-finger to bind DNA.

The titrations of G208S and D218G into DNA at pH 6.5 did not produce the characteristic sigmoidal binding isotherms, which indicate that the proteins did not interact with DNA. Since the wild-type NF displayed an increased DNA binding with the pH rise from 6.5 to 7.5, the G208S and D218G proteins were also tested for their ability to bind DNA at pH 7.5. At the more physiological pH of 7.5, the two proteins were able to bind DNA normally (Figure 3.13B and D). Curiously, the DNA-binding affinities of the two mutants were stronger than that of wild-type at pH 7.5 ($K_a^{G208S:DNA} = 9.1 \times 10^6$ M$^{-1}$, $K_a^{D218G:DNA} = 1.2 \times 10^8$ M$^{-1}$ versus $K_a^{NF54:DNA} = 1.9 \times 10^6$ M$^{-1}$). However, given that the buffer conditions of the NF mutants contained only 10 mM NaCl (while that of wild-type included 110 mM NaCl) and assuming that the DNA-binding abilities of the mutants were likely to be salt-dependent like the wild-type protein, then it is likely that the DNA-binding capabilities of G208S and D218G are comparable to that of the wild-type protein (at pH 7.5). The binding of these two NF mutants to DNA was accompanied by favourable net changes in both enthalpy and entropy (negative $\Delta H$ and positive $\Delta S$ terms).

In contrast to the other mutants, the R216Q protein was not able to bind DNA either at pH of 6.5 or 7.5 (Figure 3.13C). Therefore, while the R216Q protein was the most structured of the mutant proteins, the specific R216Q mutation clearly abolished the DNA binding ability of N-finger.

**Table 3.3. ITC statistics for mutant NF:DNA interactions.** All experiments were performed in 10 mM sodium phosphate, 0 or 10 mM NaCl, 0.5–1 mM DTT, 0.1–0.3 mM ZnSO$_4$ and at 25 °C. a average of two experiments.

<table>
<thead>
<tr>
<th>Protein</th>
<th>pH</th>
<th>n</th>
<th>$K_a$ (M$^{-1}$)</th>
<th>$\Delta H$ (kcal mol$^{-1}$)</th>
<th>$\Delta S$ (cal mol$^{-1}$ K$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>V205M</td>
<td>6.5</td>
<td>0.97 ± 0.03$^a$</td>
<td>(2.0 ± 1.6) × 10$^7$</td>
<td>-14.5 ± 5.5$^a$</td>
<td>-15.5</td>
</tr>
<tr>
<td>G208S</td>
<td>6.5</td>
<td>Binding isotherm not observed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G208S</td>
<td>7.5</td>
<td>1.46 ± 0.06</td>
<td>(9.1 ± 5.2) × 10$^6$</td>
<td>-4.2 ± 0.3</td>
<td>17.8</td>
</tr>
<tr>
<td>R216Q</td>
<td>6.5</td>
<td>Binding isotherm not observed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R216Q</td>
<td>7.5</td>
<td>Binding isotherm not observed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D218G</td>
<td>6.5</td>
<td>Binding isotherm not observed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D218G</td>
<td>7.5</td>
<td>0.53 ± 0.01</td>
<td>(1.2 ± 1.0) × 10$^8$</td>
<td>-9.8 ± 0.3</td>
<td>4.1</td>
</tr>
</tbody>
</table>
3.4.3 FOG-binding was affected by several of the mutations

The ability of the NF mutants to bind a FOG finger was assessed next. In these ITC experiments, Ush-F1 was titrated into the reaction cell containing each NF mutant. Note that these experiments were carried out at pH 5.0 because it has previously been established that the affinity for this interaction is not significantly pH dependent in the range 5.0–7.0 (Kowalski et al., 2002).

The titration of Ush-F1 into wild-type NF54 gave rise to a weak but reproducible binding isotherm (Figure 3.14A). The integrated data were fitted to a 1:1 binding model and the association constant was determined to be $4.5 \times 10^4$ M$^{-1}$ (25 °C). This value is slightly lower than that reported by Kowalski et al. (2.3 $\times 10^5$ M$^{-1}$, although the latter value was obtained at 5 °C; Kowalski et al., 2002).

Figure 3.14. ITC titrations of U-shaped finger 1 into A. Wild-type NF54. B. V205M. C. G208S. D. R216Q and E. D218G. All titrations were carried out at pH 5.0 and 25 °C. The raw and integrated data are displayed, as well as the fits to a 1:1 binding model.
The titration of Ush-F1 into V205M gave rise to pulses that remained similar in intensity (Figure 3.14B), which indicates that the proteins are not interacting. In the titration of Ush-F1 into G208S, although there appears to be a trend in the intensities of the integrated pulses (Figure 3.14C, lower panel), the data could not be fitted. These profiles implied that the two mutations, V205M and G208S, have essentially abolished the ability of N-finger to bind Ush-F1.

The titrations of Ush-F1 into R216Q and D218G, however, gave rise to clear binding isotherms (Figure 3.14D and E). The integrated data were fitted to a 1:1 binding model, giving association constants of $9.4 \times 10^4$ and $1.9 \times 10^4$ M$^{-1}$ for the Ush-F1:R216Q and Ush-F1:D218G interaction, respectively. These $K_a$ values, as well as the binding enthalpy and entropy, were all comparable to the interaction between Ush-F1 and wild-type NF54 (Table 3.4). Therefore, these two mutations have not compromised the ability of NF54 to bind a FOG finger.

**Table 3.4. ITC values of Ush-F1:NF interactions.** All ITC experiments were carried out with 20 mM sodium acetate, 1.5 mM ZnSO$_4$, 0.1 mM TCEP and at pH 5.0.

<table>
<thead>
<tr>
<th>Protein</th>
<th>$n$</th>
<th>$K_a$ (M$^{-1}$)</th>
<th>$\Delta H$ (kcal mol$^{-1}$)</th>
<th>$\Delta S$ (cal mol$^{-1}$ K$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NF54</td>
<td>0.55 ± 0.31</td>
<td>$(4.5 \pm 1.8) \times 10^4$</td>
<td>-8.0 ± 5.4</td>
<td>-5.7</td>
</tr>
<tr>
<td>V205M</td>
<td></td>
<td>Binding isotherm not observed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G208S</td>
<td></td>
<td>Binding isotherm not observed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R216Q</td>
<td>0.49 ± 0.01</td>
<td>$(9.4 \pm 0.3) \times 10^4$</td>
<td>-10.4 ± 0.2</td>
<td>-12.0</td>
</tr>
<tr>
<td>D218G</td>
<td>0.22 ± 0.10</td>
<td>$(1.9 \pm 0.3) \times 10^4$</td>
<td>-8.8 ± 4.3</td>
<td>-10.1</td>
</tr>
</tbody>
</table>
3.5 Discussion

3.5.1 The DNA-binding properties of N-finger

GATA-1 N-finger binds DNA containing a GATC sequence with an affinity of \( \sim 3 \times 10^7 \, \text{M}^{-1} \). This DNA-binding affinity is comparable to that of the GCN4 basic leucine zipper transcription factor binding its cognate DNA sites: the \( K_a \) for GCN4 binding the Ap-1 and CREB sites are \( 2.8 \times 10^7 \, \text{M}^{-1} \) and \( 3.7 \times 10^7 \, \text{M}^{-1} \), respectively (Berger et al., 1996). These association constants are weaker than those observed between arrays of classical CCHH zinc fingers and DNA (\( K_a \) for Zif268 and its cognate DNA is \( 1.5 \times 10^8 \, \text{M}^{-1} \); Yang et al., 1995), but are stronger than that reported for the association of the major ssDNA-binding protein of herpes simplex virus type I, ICP8 and its cognate ssDNA, \( 5.5 \times 10^5 \, \text{M}^{-1} \) (Gourves et al., 2000). The \( K_a \) of the NF:DNA interaction is \(~5\)-fold less than that of the interaction between chicken GATA-1 C-finger and DNA containing a GATA motif (\( 1.2 \times 10^8 \, \text{M}^{-1} \), Omichinski et al., 1993).

The data presented here show that the interaction of N-finger with DNA is dependent on both pH and ionic concentration. This is commonly observed among protein-DNA interactions (Lundback et al., 2000, Berger et al., 1996). In general, electrostatic interactions between the negatively charged DNA phosphate backbone and basic residues in a protein make substantial contributions to such interactions and binding affinities typically therefore decrease as the ionic concentration is increased. For example, in the solution structure of the complex between chicken C-finger and DNA (Omichinski et al., 1993), intermolecular contacts include three hydrogen bonds between residues conserved in murine NF (N186 and R176 of chicken CF) and the GAT sequence of the DNA, and these bonds may also be present in the NF:DNA complex.

The binding affinities of NF48 and NF54 were not substantially different. This is a somewhat surprising result, given that the minimal domain of CF required for DNA binding is longer than NF48 and that these additional residues in CF make specific contacts with DNA (Omichinski et al., 1993; Figure 3.15). This indicates that the most C-terminal residues of the NF54 do not contribute significantly to DNA binding,
unlike that of chicken CF, and may partially explain why NF binds DNA with a weaker association constant than CF.

![Figure 3.15. Minimal domains of mNF and cCF required for DNA-binding.](image)

The cysteine residues involved in zinc ligation are boxed in grey and the residues of cCF that contact DNA are marked with circles (Omichinski et al., 1993).

The full-length GATA-1 protein contains two zinc finger domains that contact DNA with differential specificity and affinity: NF binds GATC sites with a $\sim 10^7 \text{M}^{-1}$ affinity and CF at GATA sites with a $\sim 10^8 \text{M}^{-1}$ affinity. The behaviour of the fingers at double GAT sites is still unclear and may be dependent on the particular DNA sequence. At some bipartite and palindromic GATA sites, N-finger appears to increase the stability and specificity of the GATA-1:DNA complex (Martin and Orkin, 1990, Trainor et al., 1996). The full-length GATA-1 protein appears to bind DNA with double GATA sites with a $K_a$ of $\sim 10^9 \text{M}^{-1}$ (Omichinski et al., 1993), which suggests that the two ZnF domains do not necessarily cooperate in their interactions with DNA. Different DNA-binding conformations can potentially alter the extent and nature of GATA-1 interactions with other transcription factors. This may explain why the transcriptional activity of GATA proteins varies with the different promoter sequences.

### 3.5.2 Exchange processes affected NMR data

Attempts to acquire $^{15}$N-HSQC spectra of $^{15}$N-NF in the presence of DNA were unsuccessful. It is most likely that the loss of signal intensity observed arises from a chemical exchange process. Chemical exchange (Figure 3.16) refers to any process in which a nucleus exchanges between two or more environments, in this case, the free and bound forms. In the case of a bimolecular association, the off-rate for the complex is one of the determinants of the observed effects of the chemical exchange process. Generally, when the off-rate is fast compared to the chemical shift difference between the two forms (generally corresponding to lifetimes of as long as a few
microseconds), each nucleus gives rise to a single signal with a chemical shift that is the population-weighted average of the free and bound forms. This regime is often seen for very weak interactions, such as that between cytochrome c and adrenodoxin, with a $K_a$ of $\sim 4 \times 10^4$ M$^{-1}$ (Worrall et al., 2003). When the exchange process is much slower (seconds to hours), two sets of peaks are observed, one each for the free and bound forms (Figure 3.16). This was reported for example in the binding of fungal GATA AREA to DNA (Starich et al., 1998). In cases of intermediate chemical exchange (generally micro- to milliseconds), the frequencies of peaks can become poorly defined and extensive line broadening can occur. As observed in the titrations of NF into DNA in this study, the peaks became so broad that they essentially disappeared in the $^{15}$N-HSQC spectra.

Sometimes, by changing experimental parameters such as temperature, the exchange regime may be improved. Unfortunately, no temperatures between 275 and 298 K significantly improved the quality of the NF:DNA spectra. NMR experiments involving DNA constructs of different lengths and different surrounding sequences (around the core GATC motif) could conceivably affect the kinetics of the interaction with NF sufficiently to obtain sharper peaks. Alternatively, X-ray crystallography might prove to be a preferable option if this interaction were to be characterised in detail.

![Figure 3.16. Hypothetical NMR spectra showing the effects of chemical exchange.](image)

In the slow exchange regime, two peaks corresponding to the two distinct forms are observed, whereas one peak at the averaged frequency is observed under fast exchange. In the intermediate exchange regime, broadening of the signals is often observed (adapted from Roberts, 1993).
3.5.3 The effects of naturally occurring mutations on the DNA- and FOG-binding abilities of GATA-1 N-finger

Five naturally occurring single point substitution mutations residing in the NF region of GATA-1 have been identified. The structural stability of NF was somewhat affected by these mutations. CD and 1D $^1$H NMR data showed that the solution conformations of the mutants were less structured and organised than that of wild-type. Nonetheless, the partially structured nature of these mutant proteins did not hinder their ability to interact with either DNA or a FOG family protein.

The GATA-FOG interaction is a well-characterised system, in which molecular details are known concerning residues of NF that are important in its interaction with both DNA and FOG. Figure 3.17 (left) shows the FOG-binding residues delineated by mutagenesis and NMR experiments (Fox et al., 1998, Liew et al., 2000). The putative DNA-binding residues of NF are inferred by comparison to CF (Omichinski et al., 1993; Figure 3.17, right). Interestingly, the FOG- and DNA-binding surfaces lie on opposite sides of the NF molecule, suggesting that a ternary complex of GATA:FOG:DNA can be formed. Indeed, recent results from our laboratory have confirmed this proposal (F. Loughlin, unpublished data).

V205M was the first N-finger mutation identified (Nichols et al., 2000). In this study, it was found that the V205M mutation eliminated the ability of NF to interact with Ush-F1, a FOG finger, while having no effect on its DNA-binding ability. The ablation of FOG-binding ability correlates with the position of this residue in the center of the FOG-binding face of N-finger (Figure 3.17, red). Mutation of V205 into other residues such as threonine and glycine also abolishes FOG binding (Fox et al., 1998, Crispino et al., 1999). Patients carrying this mutation display severe thrombocytopenia, dyserythropoiesis and anaemia, highlighting the significance of the GATA-FOG interaction in proper differentiation of both red blood cells and platelets.

Similarly, experiments described in this Thesis found that the G208S mutant was unable to bind Ush-F1, but it retained the ability to interact with DNA at pH 7.5. Mehaffey et al. found that the FOG-binding ability of G208S depended on the FOG finger tested: the mutant was able to bind mFOG-F1 but not F9 (Mehaffey et al., 2001). Sequence comparison shows that Ush-F1 is more similar to FOG-F9 than to F1.
The role of the GATA-1 N-terminal zinc finger in transcriptional regulation (Figure 3.18), consistent with the results presented here. The G208 residue lies on one side of the putative FOG-binding face (Figure 3.17, black) and may be involved in only some of the interactions between NF and the different FOG fingers. In line with this possibility, the G208S mutation allows normal development of the erythroid lineage; however, the maturation of megakaryocytes has been disturbed with an accumulation of immature precursors. It is possible that this observation hints at specific roles for different FOG fingers in different contexts.

Figure 3.17. Naturally occurring mutations mapped onto the structure of NF. Structures on the left and right are interchangeable by a 180° rotation on the vertical axis. The FOG- and DNA-binding surfaces are coloured in orange and light purple, respectively (top). The residues are displayed as follow: V205 in red, G208 in black, R216 in brown and D218 in dark blue (middle panel). The D218 residue does not reside on either of the two faces. The lower panel shows the molecule in ribbon representation.
Figure 3.18. Sequence alignment of FOG-F1, F9 and Ush-F1. Residues identified by mutagenesis to be important for binding NF (Fox et al., 1999) are marked with open circles. Zinc ligands are boxed in grey. The NF protein containing the G208S mutation was able to bind FOG-F1, but not FOG-F9 or Ush-F1.

In contrast to the effects of V205M and G208S, the R216Q mutation eliminated the ability of NF to interact with DNA, while retaining the ability to bind Ush-F1. The putative DNA-binding face of N-finger determined by comparison to the known structure of the CF:DNA complex clearly encompasses R216, and a R→Q mutation would significantly affect the charge and size properties of the binding surface (Figure 3.17, brown). A close examination of the structure of the CF:DNA complex revealed that the corresponding arginine residue in CF forms an ion pair with the phosphate moiety of the guanine in the core GATA sequence (Omichinski et al., 1993). This highlights the significance of this residue in DNA binding and shows clearly why a mutation at this position may abolish the interaction of NF with DNA.

Interestingly, the D218G protein was able to interact with both DNA and Ush-F1. This result is supported by the position of this residue on the NF molecule: D218 does not lie on either the DNA- or FOG-binding surfaces (Figure 3.17, dark blue). Freson et al. reported that the D218G mutation affected the FOG-binding ability, with various FOG fingers affected differently (Freson et al., 2001). It remains unclear as to the molecular basis of the phenotypes observed with this particular mutation. Two situations are possible. First, the GATA-FOG interaction *is* affected by the D218G mutation and Ush-F1 used in this study does not sufficiently represent the FOG-finger that is involved. Second, GATA-FOG interaction *is not* affected and instead this mutation hinders the interaction of NF with another protein that also plays a role in the maturation of erythrocytes and megakaryocytes. This protein may be one of many that have already been identified as a protein partner of GATA-1 NF (e.g., Fli-1, c-Myb, LMO2), or it may be a novel GATA-interacting protein yet to be identified.

The most recently discovered mutation is D218Y (Freson et al., 2002). This particular mutation results in severe macrothrombocytopoenia and marked anaemia, which is a
stronger phenotype than the mutation of the same residue into glycine. Because this mutation was identified recently, only the solution properties of the D218Y protein were examined. Like the other mutant proteins, the D218Y protein was only partially structured. The partially structured nature of the other mutant proteins did not hinder their interactions with either DNA or Ush-F1, and therefore it is expected that this property of D218Y will not affect its binding ability. A recent study found that this mutation of N-finger affected the FOG-binding ability more significantly than the D218G mutation (Freson et al., 2002). While the D218 residue does not appear to reside on either the FOG- or DNA-binding surfaces of NF (Figure 3.17, dark blue), it is possible that the replacement of the aspartate sidechain with the bulkier aromatic sidechain of tyrosine may interfere with the FOG-binding surface. This position appears to accommodate some mutations as a D→V mutant is still able to bind FOG (Crispino et al., 1999). Together with the phenotypic differences observed in patients harbouring D218 mutations, it suggests that not only is the position of the mutated residue important, but at least in this case, the identity of the resultant mutation also plays a role.

In summary, biochemical and structural studies can allow us to understand some human disorders at the molecular level, and the work presented in this Chapter has shed some light on the molecular causes for several human blood disorders arising from point mutations in the N-terminal zinc finger of GATA-1. It is clear from this (and previous) work that the GATA-FOG interaction is intimately involved in the regulation of genes required for both erythropoiesis and megakaryopoiesis.

Many studies have revealed that GATA and FOG proteins make physical contacts with protein partners besides each other. However, most of these interactions are poorly characterised. In Chapter 4, a range of biochemical and biophysical methods are used to define the recently discovered interaction between FOG and a coiled-coil protein, TACC3.
Chapter 4. A novel zinc finger-mediated interaction between FOG and TACC3

4.1 Introduction

4.1.1 The GATA-independent roles of FOG

FOG-1 plays vital roles in the proper development of the erythroid and megakaryocytic lineages (Tsang et al., 1997). Mice lacking FOG displayed a marked blockage of erythropoiesis, reminiscent of GATA-1 deficiency. In contrast to mice lacking GATA-1, which showed a blocked development of their megakaryocytes in mid-maturation, FOG-deficient mice completely failed to develop megakaryocytes (Tsang et al., 1998, Fujiwara et al., 1996). This suggests that while FOG acts together with GATA-1 in erythroid cells, it appears to function during early stages of megakaryocytic development in a GATA-1-independent manner. The precise nature by which this occurs is unclear and it is envisaged that finding other proteins that interact with FOG will reveal some of its repertoire of functions.

4.1.2 Identification of TACC3 as a protein partner of FOG

There are multiple zinc finger domains in FOG-1. While several of these have been implicated in the interaction between FOG and GATA-1, the roles of others remain unknown (Figure 4.1). The presence of three classical ZnFs in a tandem array, fingers 2–4, is highly suggestive of a DNA-binding function. However, a number of studies have been unable to demonstrate unequivocally that this three-finger domain has the ability to recognise DNA (S. Orkin, unpublished results, M. Crossley, unpublished results, Kwan et al., 2003).

![Figure 4.1. Schematic of murine FOG-1.](image)

In an attempt to determine the function of fingers 2–4, a yeast two-hybrid screen of a murine erythroleukemia cell cDNA library (MEL) was carried out, using a bait encompassing fingers 1 to 4 (Natalie Bartle and Merlin Crossley, unpublished results).
Of six clones that emerged from this screen, three contained cDNA encoding murine GATA-1 (which can interact with FOG finger 1). The remaining three clones encoded for overlapping C-terminal regions of a protein known variously as TACC3, ERIC-1 and AINT (Genbank accession number AF247674). This experiment was also performed concurrently by a separate group and they too also identified TACC3 as a protein capable of interacting with FOG-1 (Garriga-Canut and Orkin, 2004).

### 4.1.3 TACC3 is a centrosomal protein implicated in the control of cell growth and differentiation

The third member of the transforming acidic coiled-coil containing protein family (TACC3) is a centrosomal or mitotic spindle-associated protein expressed in differentiating cell lines and tissues, such as the testis, spleen, thymus and peripheral blood leukocytes (McKeveney et al., 2001, Hao et al., 2002, Aitola et al., 2003). TACC3-deficient mice survive until mid-gestation with lethality involving several lineages of cells. Haematopoietic stem cells from these embryos were unable to be expanded, while still capable of terminal differentiation (Piekorz et al., 2002). Upregulation of the TACC3 gene has also been observed in various cancer cell lines (Still et al., 1999). These data, together with the observation that TACC3 is upregulated in erythroid progenitor cells, suggest that TACC3 plays a role during early stages of differentiation in haematopoiesis (McKeveney et al., 2001).

TACC1 overexpression was shown to transform primary mouse cells in culture, and all family members are defined by the presence of a putative C-terminal coiled-coil domain of ~200 residues (Figure 4.2). These characteristics give the proteins their name and support the suggestion that TACC proteins are involved in cancer (Still et al., 1999). TACC2 (or AZU-1) was recently identified as a potential tumour suppressor as restoration of protein levels reduced the malignant phenotypes of mammary epithelial cells (Chen et al., 2000).

![Figure 4.2. Schematic of TACC3.](image)

The ~200-residue putative coiled-coil domain at the C-terminus is shared by other members of the TACC family. This domain consists of two discrete putative coiled-coil regions predicted by Multicoil (Wolf et al., 1997).
4.1.4 Coiled-coil domains

Sequences at the C-terminus of TACC proteins contain the hallmark seven-residue (heptad) repeats found in coiled-coil domains. A coiled-coil is a bundle of as many as six α-helices wound into a superhelix. The positions in the repeat are traditionally labelled a to g, where positions a and d are usually filled by hydrophobic residues. These a and d residues align along the same side of the helix forming a hydrophobic stripe (Figure 4.3). Hydrophobic interactions between these stripes of neighbouring helices result in the oligomerisation of the domain. Further, because an α-helix has 3.6 residues per turn, the hydrophobic stripe drifts around the helix in a left-handed direction (Figure 4.3). As a consequence, coiled-coil helices supercoil, or wrap around each other, in order to maintain maximal contact between their hydrophobic stripes.

Two-stranded dimeric coiled-coils are found in a large number of proteins, including basic leucine zipper (bZIP) proteins that bind DNA to regulate many transcriptional processes (see Section 1.1.2; Figure 4.4A). Three-stranded trimeric coils have been identified in influenza haemagglutinin and envelope proteins of Moloney murine leukemia virus and HIV-1, and are thought to play a role in membrane fusion (Bullough et al., 1994, Fass et al., 1996, Chan et al., 1997; Figure 4.4B). Four-stranded coils are present in the Sendai virus phosphoprotein (Tarbouriech et al., 2000) and the synaptic fusion complex containing SNARE proteins (Sutton et al., 1998; Figure 4.4C).
Figure 4.4. Examples of coiled-coil domains. A. GCN4 dimer (PDB code 1YSA; Ellenberger et al., 1992). B. Influenza haemagglutinin trimer (PDB code 1HTM; Bullough et al., 1994). C. The tetrameric neuronal synaptic fusion complex (PDB code 1SFC; Sutton et al., 1998). Proteins are displayed in ribbon representations and DNA shown in dark blue.

4.1.5 The minimal regions required for the FOG:TACC3 interaction

The initial yeast two-hybrid screen utilised a region of FOG-1 containing fingers 1 to 4, and pulled out a C-terminal region (535–637) from TACC3 (N. Bartle and M. Crossley, unpublished results). As Figure 4.5 illustrates, deletion mutagenesis established the minimal region of FOG required to bind TACC3 as the third putative zinc finger (FOG-1 finger 3, or FOG-F3; N. Bartle, S. Lee and M. Crossley, unpublished results). FOG-F3 is a putative classical ZnF: its amino acid sequence aligns well with the consensus sequence of a classical ZnF (Figure 4.5).

Figure 4.5. The minimal region of FOG-1 required to bind TACC3. Different FOG constructs and their abilities to bind TACC3(349–637) in yeast two-hybrid assays are shown. The minimal region required for TACC-interaction encompasses finger 3, the sequence of which is illustrated together with the consensus sequence for a classical CCHH ZnF. C, H and Φ represent cysteine, histidine and hydrophobic residues.
Using the same strategy, the minimal domain of TACC3 required to bind FOG-F3 was determined to be a 47-residue region at the most C-terminus of the TACC3 protein (Figure 4.6; S. Lee and M. Crossley, unpublished results). This region lies within one of two putative coiled-coil domains predicted by MultiCoil (Wolf et al., 1997).

**Figure 4.6. The minimal FOG-binding region of TACC3.** The putative coiled-coil domains are shown in grey. Deletion constructs are listed with their FOG-binding abilities as determined by yeast two-hybrid assays (performed by S. Lee).

### 4.1.6 Specific aims of this study

The aims of the current study were to (i) characterise the solution properties of the individual domains required for the FOG:TACC3 interaction; and (ii) investigate the mechanisms by which the proteins interact with one another. This study will enhance our understanding not only of FOG biochemistry but also of the structural basis of protein-protein interactions mediated by zinc finger and coiled-coil domains.
4.2 Characterisation of FOG-1 finger 3

4.2.1 Wild-type FOG-F3 protein is folded and monomeric

DNA encoding FOG-F3 was subcloned into pGEX-2T, which enabled the overexpression of a GST-FOG-F3 fusion protein. This fusion protein was quite insoluble and precautions were taken in order to maximise the yield of soluble protein. These included performing the thrombin cleavage of the fusion protein in solution (rather than on beads) and adding dithiothreitol to the free FOG-F3 protein. FOG-F3 was then purified by reverse phase HPLC and its identity and purity was confirmed by electrospray mass spectrometry ($M_{\text{obs.}} = 3821.8 \pm 0.7$ Da, $M_{\text{theo.}} = 3822.4$ Da).

In order to determine whether the protein is a true zinc-binding domain, CD spectropolarimetry was used to monitor its secondary structure content in varying zinc concentrations and pH environments. As Figure 4.7 (left) shows, the CD spectrum of the wild-type FOG-F3 at low pH (excess $\text{Zn}^{2+}$) was indicative of a random polypeptide conformation. An increase of pH resulted in positive ellipticities at low wavelengths and a shift of the minimum from ~198 to ~207 nm suggesting that stable secondary structure was induced. Similarly, the protein was in an unfolded random coil configuration in the absence of $\text{Zn}^{2+}$, and stable secondary structure was obtained with the addition of $\text{Zn}^{2+}$ (Figure 4.7, right).

![Figure 4.7. pH- and Zn$^{2+}$-dependent folding of FOG-F3.](image)

**Figure 4.7. pH- and Zn$^{2+}$-dependent folding of FOG-F3.** The CD spectrum on the left shows FOG-F3 (9 $\mu$M) in the presence of 1 mM $\text{ZnSO}_4$ and 1 mM TCEP, at pH 2.9, 4.1 and 5.3. On the right are the spectra of a 16 $\mu$M sample at pH 5.5 with 1 mM TCEP and 0, 4 and 16 $\mu$M $\text{ZnSO}_4$. 
The folding process coupled to changes in zinc concentration and sample pH displayed by FOG-F3 is reminiscent of known zinc finger domains, such as classical fingers of TFIIIA and variant CCHC FOG fingers (Frankel et al., 1987, Matthews et al., 2000).

The 1D $^1$H NMR spectrum of wild-type FOG-F3 contained sharp and dispersed signals (Figure 4.8), which indicated that the protein was well-folded with a high degree of tertiary organisation. The linewidths of the signals further suggested that FOG-F3 was monomeric in solution.

![Figure 4.8. 1D $^1$H NMR spectrum of wild-type FOG-F3.](image)

The spectrum was recorded at 3 °C. The signal corresponding to water was removed for the purpose of clarity.

Two-dimensional COSY, TOCSY and NOESY NMR experiments were performed using wild-type FOG-F3 and sequence-specific assignments were made using standard procedures (Wüthrich, 1986). However, the low solubility of this protein (maximum concentration of ~100 µM at pH 5.0) made it very difficult to obtain good quality NOESY data and thereby determine the solution structure of the wild-type protein.

### 4.2.2 The design of FOG-F3 mutants with increased solubility

In order to study the FOG:TACC3 interaction using biophysical experiments, it was decided to design mutants of FOG-F3 with increased solubility. These mutants were required to retain (a) the ZnF fold and (b) the ability to interact with TACC3.

To improve solubility, two mutant proteins (FOG-F3$_{KRA}$ and FOG-F3$_{KRRA}$) were designed to exhibit higher isoelectric points (pI) than the wild-type domain. The removal of two acidic residues (E330 and E349) and introduction of either two or three basic arginine and lysine residues (Figure 4.9) shifted the pI from 8.0 to 10.3 and 10.7. These residues were chosen based on preliminary mutagenesis results
(performed by S. Lee) suggesting that they were not likely to be involved in the interaction.

**Figure 4.9. FOG-F3 mutants designed to increase solubility.** The amino acid sequences of wild-type, and triple and quadruple mutants are shown, with the substituted residues in grey. The numbers correlate to the sequence of mFOG-1. The first two residues (in *italics*) are remnants of the thrombin cleavage and are not part of the FOG-1 sequence.

The FOG mutants were overexpressed and purified in similar manners as the wild-type protein. Unlike the wild-type protein, the production of the mutants did not require precautions during purification to ensure that the proteins were soluble. As intended, solutions could be prepared of each purified domain at substantially higher concentrations than the wild-type domain: the maximum soluble concentration for FOG-F3_{KRA} was \(~700\ \mu\text{M}\).

The ability of each of the two mutants to fold was assessed using CD spectropolarimetry. Like wild-type FOG-F3, both triple and quadruple mutants gain substantial amounts of secondary structure with an increase in pH and in the presence of Zn^{2+} (Figure 4.10). This result indicates that the introduced mutations did not affect the structural integrity of the domain.

**Figure 4.10. Comparison of the CD spectra of wild-type FOG-F3 and the triple and quadruple mutants.** The samples contained 1 mM ZnSO_{4} and 1 mM TCEP, pH 5.1–5.4.

Next, 1D ^1\text{H} NMR spectra of the two solubility mutants were recorded. As Figure 4.11 shows, the spectra contained sharp and dispersed signals indicative of the
proteins being folded. Comparison of the spectra with that of the wild-type protein suggest that all three proteins assumed similar conformations in solution, as judged by similarities in the frequencies displayed by many of the highly shifted signals. Small differences observed are likely to be those corresponding to the mutated residues or those surrounding the substitution.

Figure 4.11. 1D $^1$H NMR spectra of wild-type FOG-F3, FOG-F3KRA and FOG-F3KRRA. The samples contained 1 mM ZnSO$_4$ and 1 mM TCEP, pH ~5.0. The spectra were recorded with 64 scans at 3 °C.

Having determined that both mutants were more soluble than wild-type and appeared to be zinc-binding domains, FOG-F3KRA, which carries one less mutation, was chosen for further experiments.

4.2.3 Wild-type and mutant FOG-F3 contain similar structural elements

Two-dimensional COSY, TOCSY and NOESY NMR experiments were carried out on FOG-F3KRA, and sequence specific assignments made (Wüthrich, 1986). The frequencies (chemical shifts) of nuclei are very sensitive to the physical environment of the nucleus concerned, and in some cases are additionally correlated to the type of secondary structure in which the nucleus is located. For example, the H$^\alpha$ chemical shift is shifted upfield when located within a helix, and downfield in a $\beta$-strand configuration. A comparison of H$^\alpha$ chemical shifts of wild-type and FOG-F3KRA (Figure 4.12) shows that the H$^\alpha$ shifts of the majority of residues in the two proteins were very similar. Further, the secondary structures predicted by the chemical shifts
illustrate that both proteins probably contain a short β-strand around residues 341–344 followed by a ~10-residue α-helix. Together, these data indicate that the wild-type and FOG-F3\textsubscript{KRA} proteins assumed similar tertiary conformations in solution.

Figure 4.12. H\textsuperscript{\alpha} chemical shift differences for wild-type FOG-F3 (black) and FOG-F3\textsubscript{KRA} (grey), compared to random coil shifts. The predicted secondary structures are displayed above, and amino acid sequences with residues that differ between wild-type and FOG-F3\textsubscript{KRA} marked with stars are shown below.

4.2.4 The solution structure of FOG-F3\textsubscript{KRA} shows that it is a true classical zinc finger

Having established that the triple mutant and wild-type proteins probably have comparable backbone conformations, the solution structure of FOG-F3\textsubscript{KRA} was determined. A total of 879 NOE constraints was extracted from a NOESY spectrum recorded at 278 K. Structure calculations were originally carried out using DYANA (Güntert \textit{et al.}, 1997), during which no zinc ion was included and the zinc coordination site was not restrained. On the basis of these structures, it was evident that thiol groups of C335 and C338 and N\textsuperscript{ε} atoms of H351 and H355 comprised the zinc coordination sphere. These structures were used as input for calculations using ARIA (Nilges, 1995). Experimental restraints included distance restraints determined from NOEs and φ angle restraints estimated from the HNHA experiment (Vuister and Bax, 1993) and the INFIT program (Szyperski \textit{et al.}, 1992). The 20 structures with lowest overall energies from a final refinement step carried out in water were used to represent the solution structure of FOG-F3\textsubscript{KRA} (Figure 4.13A). The structures display good covalent geometry and angles and good nonbonded contacts, as shown by the low value of the mean Lennard-Jones potential (Table 4.1). Analysis of the structures with PROCHECK-NMR (Laskowski \textit{et al.}, 1996) revealed that for well-ordered
A novel zinc finger-mediated interaction between FOG and TACC3 residues, 95.4% of backbone $\phi/\psi$ pairs lie within the most favoured or additionally allowed regions of the Ramachandran plot. The atomic coordinates for this family of conformers have been deposited with the Protein Data Bank (PDB accession code 1SRK).

Figure 4.13. Solution structure of FOG-F3 KRA. A. Ensemble of the best 20 structures overlaid over the backbone atoms of residues 333–355. The structured region is shown in orange, the zinc ion in magenta and zinc-chelating side chains in blue. B. Ribbon representation displaying elements of secondary structure as recognised in the program MOLMOL (Koradi et al., 1996). C. Overlay of FOG-F3 KRA and finger 3 of Sp1 (PDB code 1SP1; Narayan et al., 1997). FOG-F3 KRA(333–355) is displayed in orange, while Sp1 finger 3(3–25) is in green.

The overall topology of FOG-F3 KRA conforms to the expected fold of classical CCHH zinc finger domains: two short strands of antiparallel $\beta$-sheet linked by a rubredoxin-like $\beta$-turn (Dauter et al., 1996) and followed by an $\alpha$-helix (Figure 4.13B). The zinc-ligating cysteine residues are located in the $\beta$-turn, while the histidines reside at the C-terminus of the $\alpha$-helix. The short $\beta$-sheet encompasses residues 333–334 and 341–342, and a hydrogen bond is formed between the backbone amide proton of F342 and the carboxyl oxygen of F333. The $\beta$-turn linking these two strands has hydrogen bonds and dihedral angles consistent with the rubredoxin-like turn that is commonly found in zinc-binding domains. A type-VIII $\beta$-turn joins the $\beta$-sheet to the $\alpha$-helix (Hutchinson and Thornton, 1994). The $\alpha$-helix runs from residue 345 to 355, although both $d_{\alpha\beta}(i, i+3)$ and $d_{\alpha\beta}(i, i+4)$ NOEs are observed for residues 353–355, suggesting that the conformation of this region lies somewhere between an $\alpha$-helix and a $3_1\beta$ helix. This phenomenon was also reported for the last 3–4 residues
A novel zinc finger-mediated interaction between FOG and TACC3

in \( \alpha \)-helices of Sp1 fingers 2 and 3 (Narayan et al., 1997). This transition of an \( \alpha \)-helix to a 3\(_{10}\) helix is a general property shared by HX\(_3\)H subclass of classical ZnFs: structures with HX\(_{4-5}\)H spacing show no indication of a 3\(_{10}\) helix (Lee et al., 1989).

Overall, the structure of FOG-F3\(_{KRA}\) is very similar to other classical zinc fingers. It overlays with the third ZnF of Sp1 with a root mean square deviation of 1.2 Å over the ordered backbone atoms (Figure 4.13C; Narayan et al., 1997).

### Table 4.1. Structural statistics for the family of 20 FOG-F3\(_{KRA}\) structures.

<table>
<thead>
<tr>
<th>Experimental restraints</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total NOE restraints</td>
<td>879</td>
</tr>
<tr>
<td>Unambiguous NOE restraints</td>
<td>859</td>
</tr>
<tr>
<td>Ambiguous NOE restraints</td>
<td>20</td>
</tr>
<tr>
<td>Intraresidue distances ((i - j = 0))</td>
<td>368</td>
</tr>
<tr>
<td>Sequential distances (</td>
<td>i - j</td>
</tr>
<tr>
<td>Medium-range distances (</td>
<td>i - j</td>
</tr>
<tr>
<td>Long-range distances (</td>
<td>i - j</td>
</tr>
<tr>
<td>Dihedral (\phi) angles</td>
<td>31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PROCHECK statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Residues in most favoured regions</td>
<td>85.4 %</td>
</tr>
<tr>
<td>Residues in allowed regions</td>
<td>10.0 %</td>
</tr>
<tr>
<td>Residues in generously allowed regions</td>
<td>4.6 %</td>
</tr>
<tr>
<td>Residues in disallowed regions</td>
<td>0.0 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Atomic RMSD versus mean (Å)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Backbone atoms (333–355)</td>
<td>0.31 ± 0.08</td>
</tr>
<tr>
<td>All heavy atoms (333–355)</td>
<td>0.87 ± 0.19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean deviations from ideal covalent geometry</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond lengths (Å)</td>
<td>0.00558 ± 0.00037</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>0.717 ± 0.038</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean energies (kJ mol(^{-1}))</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(E_{\text{NOE}})</td>
<td>4.74 ± 0.81</td>
</tr>
<tr>
<td>(E_{\text{cdih}})</td>
<td>0.05 ± 0.10</td>
</tr>
<tr>
<td>(E_{\text{vdW}})</td>
<td>28.28 ± 0.92</td>
</tr>
<tr>
<td>(E_{\text{bonds}})</td>
<td>2.42 ± 0.20</td>
</tr>
<tr>
<td>(E_{\text{improper}})</td>
<td>2.46 ± 0.62</td>
</tr>
<tr>
<td>(E_{\text{angles}})</td>
<td>27.82 ± 1.00</td>
</tr>
<tr>
<td>(E_{\text{dihedral}})</td>
<td>72.33 ± 3.55</td>
</tr>
<tr>
<td>(E_{\text{total}})</td>
<td>143.02 ± 3.48</td>
</tr>
</tbody>
</table>
4.3 Characterisation of the TACC3 C-terminal coiled-coil

The region of murine TACC3 identified to be important for its interaction with FOG-F3 lies in one of two putative coiled-coil domains at the C-terminus of the protein. This Section describes biophysical experiments carried out to determine the solution behaviour of this domain of TACC3.

In coiled-coil domains, the identity of the residues at \( a \) and \( d \) positions significantly affects the packing geometry of the hydrophobic core. Heptad positions in adjacent helices align to form alternating \( a \) and \( d \) layers in the core of the supercoils and the geometries of these layers differ in two-, three- and four-stranded supercoils. It has been observed that when residues at position \( a \) are predominantly isoleucines, the helices pack together to form dimers, whereas tetramers are formed when isoleucines fill position \( d \) (Harbury et al., 1993). While isoleucines are not common in the sequence of the TACC3 domain, there are more located at position \( a \) than \( d \) (2 to 1; Figure 4.14). The lone position \( d \) isoleucine is located at the C-terminus of the domain and hence will probably not affect the packing geometry. The presence of two isoleucines at position \( a \) therefore hints towards the packing of TACC domain as a dimeric coiled-coil.

The TACC3 C-terminal putative coiled-coil region contains a heptad break consisting of a three-residue insertion at residues 573–575 (Figure 4.14). This type of break is referred to as a ‘stammer’ and is one of two types common in coiled-coils (Brown et al., 1996; the other being a four-residue insertion). These insertions lead to localised weakening of the coiled-coil interactions, which may result in either an over- or under-winding of the coiled helices and is accommodated by the domain (such as in the rat mannose-binding protein; Weis and Drickamer, 1994). Heptad breaks can also lead to unstructured regions, which accounts for the flexibility observed in long \( \alpha \)-fibrous molecules such as myosin (Offer, 1990) and the termination of coiled-coil regions (Carr and Kim, 1993).
A novel zinc finger-mediated interaction between FOG and TACC3

Figure 4.14. Heptad repeats of the C-terminal region of TACC3. The predicted coiled-coil domain starts at residue 562 and it encompasses a three-residue insertion (underlined). The residues in red and orange are situated at the a and d positions, respectively. TACC constructs used in the current study are TACC103 (535–637), TACC77 (561–637) and TACC47 (591–637).

Three different constructs of TACC3 were chosen for biophysical studies. The predicted coiled-coil domain stretches from residues 562 to the C-terminus (residue 637) and the longest construct, the 103-residue TACC103, encompasses this entire region and an extra 27 N-terminal residues (residues 535–637; Figure 4.14). TACC77 comprises just the predicted coiled-coil (residues 561–637), while TACC47 is the minimal region identified to be necessary for the interaction with FOG, the most C-terminal 47 residues of the protein (residues 591–637).

These three constructs were subcloned into protein expression vector pGEX-6P and overexpressed as GST-fusion proteins. The release of free proteins from the GST moieties were carried out with PreScission™ protease, and yielded large amounts of full-length TACC47, TACC77 and TACC103. Further purification using reverse phase HPLC was performed to obtain purified TACC proteins.

4.3.1 TACC3 C-terminal domains are highly α-helical in solution

The three TACC proteins showed typical α-helical CD spectra with double minima at 208 and 222 nm, a maximum at 190 nm and positive ellipticities below ~200 nm (Figure 4.15). The α-helical structure was robust, with CD spectra showing little change over the pH range of ~3 to ~9 and at temperatures from 8 to 25 °C (data not shown).
In order to assess whether or not these constructs form coiled-coils in solution, further CD data were collected. In a typical CD spectrum, the ellipticity at 222 nm, or $[\theta]_{222}$, is responsive to the overall $\alpha$-helical content, while $[\theta]_{208}$ is thought to be sensitive to whether the $\alpha$-helix is monomeric or involved in tertiary contacts (Cooper and Woody, 1990). Therefore the ratio of these ellipticities, $[\theta]_{208}/[\theta]_{222}$, is a useful criterion to evaluate the presence of coiled-coil structures. For a non-interacting $\alpha$-helix, the ratio has been shown to be ~1.2, while for helical coiled-coils, the ratio is generally 0.9–1.0 (Zhou et al., 1992).

In aqueous conditions, the ellipticities of TACC77 at 208 and 222 nm are $-24,300$ and $-24,400$ deg cm$^2$ dmol$^{-1}$, respectively, giving rise to a ratio of 1.00. The addition of 50 % TFE resulted in an increase in the amplitudes of both the molar ellipticities ($[\theta]_{208} = -36,400$ and $[\theta]_{222} = -31,600$ deg cm$^2$ dmol$^{-1}$) and the ellipticity ratio ($[\theta]_{208}/[\theta]_{222} = 1.15$; Figure 4.16A). TFE is known to disrupt the tertiary and quarternary structures stabilised by hydrophobic interactions and to induce ‘single-stranded’ $\alpha$-helices (Lau et al., 1984). The observed increase in the ellipticity ratio is therefore consistent with a loss of coiled-coil structure. In addition, the increase in amplitude at 222 nm (Figure 4.16A) indicated that the TACC77 protein was only partially $\alpha$-helical in aqueous condition. Estimations of $\alpha$-helical content using both CDstr (Johnson, 1999) and empirical methods (Chen et al., 1974) suggest that TACC77 is ~70 % helical in aqueous solution.
Changes in sample temperature between 7 and 25 °C appeared to have a small effect on the helical structure of TACC77 (Figure 4.16B). The ratio $\theta_{208}/\theta_{222}$ decreased marginally from 1.03 at 25 °C to 0.96 at 7 °C, suggesting an increased formation of coiled-coil structure with decreasing temperature. This trend was also observed in TACC47 (data not shown) and in the coiled-coil domain of human cartilage matrix protein (Beck et al., 1996).

Changes in sample concentration over the range 0.2 to 77 µM did not alter the coiled-coil content as measured by the $\theta_{208}/\theta_{222}$ ratio (data not shown), indicating either that these domains do not form multimeric coiled-coils or that the degree of self-association does not change over this concentration range.

Together, these CD studies suggest that the C-terminal domain of TACC3 is highly α-helical with some coiled-coil characteristics. It is possible that additional parts of the TACC3 sequence are required for full coiled-coil formation. Further, an interruption of the heptad repeat is present in the TACC77 protein (Figure 4.14) and this three-residue insertion may be partially responsible for a local unwinding of the supercoils (Brown et al., 1996).

These experiments also indicated that TACC47 was poorly soluble at pH > ~4.8. This observation is in line with the predicted isoelectric point of the construct (5.4), and subsequent experiments were therefore carried out with TACC77 and TACC103.
4.3.2 TACC77 exists in an equilibrium between multiple oligomeric states

In order to probe the oligomeric state of the TACC coiled-coil domain, chemical cross-linking studies were carried out on TACC77. In a cross-linking reaction, the chemical cross-linker glutaraldehyde reacts primarily with amine groups of lysine and N-terminal residues, forming bridging covalent bonds between reactive groups that are in close proximity. If the protein forms multimers, then these will be cross-linked and be detectable by SDS-PAGE analysis.

Figure 4.17 shows the gels obtained in a time-course experiment. It should be noted that monomeric TACC3 runs at an anomalously low apparent mass of ~7 kDa (lane 1; \( M_{\text{monomer}} = 9.3 \) kDa). The formation of higher order species is clearly observed, with an apparent dimer being the most prevalent at short reaction times. For comparison, a protein that is known to exist as a monomer in solution (Hop; Chen et al., 2002, Shin et al., 2002), was exposed to the same cross-linking conditions (Figure 4.17B); it is clear that TACC3 has a much higher propensity to self-associate.

Figure 4.17. SDS-PAGE analysis of cross-linked TACC77. A. Time-course experiment was carried out with 120 \( \mu \)M TACC77, 0.03 % glutaraldehyde in 10 mM sodium acetate, 1 mM TCEP, pH 5.5 at room temperature for 5, 10, 15, 20 and 30 minutes (lanes 1–5). Lane 6 is the MW standards. B. TACC77 and the control protein of ~8 kDa were cross-linked with 0.03 % glutaraldehyde at room temperature for 5 and 7 minutes (lanes 1 and 2).

While cross-linking is a useful method for rapidly assessing the propensity of a protein to self-associate, it is often difficult to interpret the results in detail. Thus, more accurate methods were employed to confirm the solution behaviour of the TACC3 domain.
4.3.3 TACC103 is dimeric in solution

In a sedimentation equilibrium experiment, an initially homogeneous solution is subjected to centrifugation such that the molecules begin to sediment towards the bottom of the cell. Once the sedimentation process is in equilibrium with the opposing process of diffusion, the net distribution of molecules becomes invariant with time, and an exponentially distributed concentration gradient is formed. The measurement of absorbance as a function of the position of the cell then provides information on the effective solution molecular weight of the molecules (Ralston, 1993). The distribution of a single, thermodynamically ideal species in a cell is described by equation 4.1.

\[ A_r = A_{r_0} \exp \left[ \frac{\omega^2}{2RT} M \left( 1 - \bar{\nu} \rho \right) \left( r^2 - r_{r_0}^2 \right) \right] + c \]  

(Equation 4.1)

where
- \( A_r \) - absorbance at radius \( r \)
- \( A_{r_0} \) - absorbance at a reference radius \( r_0 \)
- \( \omega \) - angular velocity (rad s\(^{-1}\))
- \( R \) - gas constant (8.314 J K\(^{-1}\)mol\(^{-1}\))
- \( T \) - temperature (K)
- \( M \) - molecular weight of the species (g mol\(^{-1}\))
- \( \bar{\nu} \) - partial specific volume of the species (ml g\(^{-1}\))
- \( \rho \) - density of solvent (g ml\(^{-1}\))
- \( c \) - baseline offset

Non-linear least-squares analysis is then used to fit the experimental data to this model, or if necessary to more complex models, and thereby obtain estimates for the solution molecular weight (Ralston, 1993). Datasets are normally recorded under a range of experimental conditions (such as varying rotor speeds and sample concentrations), in order to improve estimates for the fitted parameters. The goodness of a fit can be ascertained from the distribution of the differences between the experimental and calculated data (known as the residuals).

Sedimentation equilibrium experiments were used to determine the oligomeric state of TACC103. Samples of 9.3, 18.5 and 37.0 µM loading concentrations were subjected to three different rotor speeds (16,000, 20,000 and 24,000 rpm) to obtain concentration versus radial distance profiles (Figure 4.18). A global fit of the nine datasets to an ideal single species model using NONLIN (Johnson et al., 1981)
yielded a molecular weight of 24.5 ± 3.2 kDa and small random residuals (indicative of a good fit to the model). The experimental value is in good agreement with the theoretical mass of 24.7 kDa for dimeric TACC103.

Figure 4.18. Sedimentation equilibrium analysis of TACC103. The lower panel displays a plot of absorbance at 230 nm versus $r^2/2$ (cm$^2$), whereas the upper panel illustrates the residual deviations resulting from the fit of an ideal single species model to the data. Data were recorded on an 18.5 µM sample at 24,000 rpm and 4 °C.

Multiangle laser light scattering (MALLS) in conjunction with size exclusion chromatography (Superose™ 12 HR 10/30) was also used to assess the self-association state of TACC77 as it offers a second shape-independent approach for the determination of effective molecular mass (Wyatt, 1993).

Figure 4.19A shows the gel filtration elution profile together with in-line MALLS molecular mass determination. The profile shows that more concentrated samples of TACC77 gave rise to earlier eluting peaks from the column and also to higher MALLS-determined molecular masses. This suggests that the protein was forming higher order aggregates in a concentration-dependent manner. In each case, the elution of one peak (with a polydispersity value near unity) indicated rapid equilibration between the multiple oligomeric states (slower exchange between states would give rise to several peaks in the gel filtration profile). The determined masses (29–37 kDa; Figure 4.19B) suggest that TACC77 is exchanging between several oligomeric states ($M_{\text{monomer}} = 9.3$ kDa). Together with the cross-linking and sedimentation data, it can be concluded that that TACC77 exists primarily as a dimer at low concentrations, and undergoes exchange between dimer and higher order species at higher concentrations.
A novel zinc finger-mediated interaction between FOG and TACC3

Figure 4.19. Size exclusion chromatography and light scattering of TACC77. A. The lines represent elution profiles as detected by absorbance at 215 nm, while squares display the molecular masses at each point, calculated from MALLS data. The three different concentrations of TACC77, 124, 57 and 8 µg, are represented by red, black and green, respectively. The grey line marks the center of the peak corresponding to the most concentrated sample. B. MALLS-determined molecular masses for the different amounts of TACC77 detected by the in-line refractometer.

4.3.4 C-terminal TACC domains undergo exchange on the $^1$H chemical shift timescale

The 1D $^1$H NMR spectra of all three TACC proteins consisted of very broad signals (Figure 4.20), suggesting that all constructs were either undergoing chemical exchange on an intermediate ($\text{ms}–\mu\text{s}$) timescale (see Section 3.5.2) or forming large aggregates (or possibly both). This supports the previous observations that TACC77 exists in an equilibrium between multiple oligomeric states.
Figure 4.20. ID $^1$H NMR spectra of TACC3 domains. The samples were 160–330 µM in concentration, pH 5.0–5.6 and spectra were recorded at 25 °C. The intensity of the amide/aromatic regions (6–9.5 ppm) has been multiplied by 3 with respect to the downfield regions (–0.5–4.2 ppm) in order to clearly display the small signals.

Overall, the TACC3 C-terminal domain appears to be a highly α-helical region that most likely forms a dimeric coiled-coil. While the domain exists primarily as a dimer, it has tendencies to form higher order oligomers. This tendency was also manifested when the domain was overexpressed in HeLa cells; large highly ordered polymeric structures have been observed by electron microscopy (Gergely et al., 2000).

The solution properties of the individual components of the FOG:TACC3 interaction have now been elucidated. Finger 3 from FOG-1 is a true classical zinc finger, while the C-terminal TACC domain is most likely to be a dimeric coiled-coil domain. The next Section focuses on understanding the molecular basis underlying the interaction between FOG-1 and TACC3.
4.4 The FOG-TACC3 interaction

4.4.1 FOG-F3 bound the TACC domain in an *in vitro* GST-pulldown experiment

The first step in studying the interaction between FOG-F3 and the TACC domain (the C-terminal coiled-coil region) was to confirm and establish the interaction using purified components. This was carried out in an *in vitro* GST-pulldown experiment. The wild-type FOG-F3 was produced as GST-fusion protein, which enabled immobilisation onto glutathione-Sepharose® 4B beads. After incubation of the beads in a solution containing purified TACC103, the beads were washed extensively. As Figure 4.21 shows, GST-FOG-F3 was able to bind TACC103, while GST alone was not. This result confirmed a direct interaction between purified FOG and TACC3 proteins *in vitro*.

![Figure 4.21. GST-pulldown of GST-FOG-F3 and TACC103. SDS-PAGE analysis of GST and GST-FOG-F3 immobilised on beads in lanes 2 and 5, the input TACC103 sample (20 %) in lanes 3 and 6 and lanes 4 and 7 display the proteins left on the beads after extensive washing. Lane 1 is the MW standards and positions of GST, GST-FOG-F3 and TACC103 are labelled with arrows.](image)

4.4.2 The affinity of the FOG:TACC3 interaction

Isothermal titration calorimetry (ITC; Section 3.3.1) was used to determine the strength of the interaction. The titration of FOG-F3<sub>KRA</sub> into the sample cell containing TACC77 gave negative pulses (Figure 4.22A, *upper* panel), indicating an exothermic interaction. The intensity of these pulses decreased with FOG-F3<sub>KRA</sub> titration as binding sites on TACC77 became saturated. The integrated data or binding isotherm was a shallow exponential curve (Figure 4.22A, *lower* panels) suggestive of a weak interaction. The fit to the ‘one set of sites’ model was unable to converge due to the
shallowness of the binding isotherm, so in order to obtain an association constant, the stoichiometric ratio, \( n \), was constrained. Since the stoichiometry ratio of FOG-1 to TACC3 is not known, the \( n \) was constrained to different values in a stepwise fashion in order to determine the best fit for the data. Three different \( n \) values and their respective fits are shown in the lower panels of Figure 4.22A. It can be seen that the data clearly did not fit to a \( n = 2 \) constraint (grey line). While the data appeared to fit both \( n = 0.5 \) and \( n = 1 \) constraints well, the \( n = 0.5 \) constraint gave rise to a lower \( \chi^2 \) value and hence was a better fit to the experimental data. This stoichiometry ratio assumes one molecule of FOG-F3\( _{\text{KRA}} \) (the titrant) interacted with two molecules, or a dimer, of TACC77. Based on this fit, FOG-F3\( _{\text{KRA}} \) and TACC77 interacted with an association constant of \( 1.4 \times 10^4 \text{ M}^{-1} \) (Figure 4.22B). An affinity of this magnitude is consistent with experiments carried out using size exclusion chromatography, where coinjection of the two proteins resulted in the elution of the two individual components (data not shown).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( K_a(M^{-1}) )</th>
<th>( \Delta H(\text{kcal mol}^{-1}) )</th>
<th>( \Delta S(\text{cal mol}^{-1} \text{K}^{-1}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>( (1.4 \pm 0.7) \times 10^4 )</td>
<td>-5.9 ± 3.9</td>
<td>-0.9 ± 14.2</td>
</tr>
<tr>
<td>1.0</td>
<td>( (3.6 \pm 2.4) \times 10^4 )</td>
<td>-2.4 ± 1.7</td>
<td>-11.4 ± 7.2</td>
</tr>
</tbody>
</table>

Figure 4.22. ITC data for the titration of FOG-F3\( _{\text{KRA}} \) into TACC77. A. The upper panel shows the raw data, while lower panels display the integrated binding isotherm. The data were fitted to a single site model with the \( n \) value constrained to 0.5 (left, red line), 1 (right, red line) or 2 (right, grey line). B. The statistics of the fit with \( n = 0.5 \) and 1 constraints.
4.4.3 Analysis of the FOG:TACC3 interaction using $^{15}\text{N}$-HSQC NMR experiments

In order to determine the molecular details of the FOG-F3:TACC3 interaction, $^{15}\text{N}$-HSQC titrations were carried out. These were performed in two different 'orientations': unlabelled FOG-F3$_{KRA}$ was titrated into $^{15}\text{N}$-TACC77, and unlabelled TACC77 was added to $^{15}\text{N}$-FOG-F3$_{KRA}$.

Titration of unlabelled FOG-F3$_{KRA}$ to $^{15}\text{N}$-TACC77

$^{15}\text{N}$-HSQC experiments detect correlations between a proton and its attached nitrogen and therefore a $^{15}\text{N}$-HSQC spectrum of a $^{15}\text{N}$-labelled protein usually contains about as many signals as there are residues in the protein. The $^{15}\text{N}$-HSQC spectrum of $^{15}\text{N}$-TACC77 however contained very few signals (Figure 4.23, left). This stems from exchange processes occurring on an intermediate ($\mu$s–ms) timescale and is consistent with earlier findings that TACC77 undergoes exchange between multiple oligomeric states.

The addition of up to $\sim 6.3$ molar equivalents of unlabelled FOG-F3$_{KRA}$ resulted in the appearance of many signals in the $^{15}\text{N}$-HSQC spectrum (Figure 4.23, right). The appearance of signals suggests that the extent of the exchange process(es) was reduced (or perhaps that the kinetics of the process were altered), and is consistent with a situation in which the binding of FOG-F3 to TACC77 abrogates non-specific interactions between TACC77 dimers. However, even in the presence of a high concentration of FOG-F3$_{KRA}$, the reaction did not reach completion, and it remains unclear whether the spectrum of a fully bound complex would contain more signals. The clustered nature of reappeared signals supports the $\alpha$-helical character of the domain, as the regularity of $\alpha$-helical structure generally gives rise to limited chemical shift dispersion. Unfortunately, the quality of the spectrum of TACC77 in the presence of FOG was not adequate to undertake a more detailed structural analysis.
A novel zinc finger-mediated interaction between FOG and TACC3

Figure 4.23. $^{15}$N-HSQC spectra of $^{15}$N-TACC77 alone and in the presence of FOG-F3$^{\text{KRA}}$ (2 °C). On the left is the spectrum of $^{15}$N-TACC77 (145 μM), and the right spectrum was recorded at the end of the titration where ~6.3 molar equivalents of FOG-F3$^{\text{KRA}}$ were added. The samples contained 10 mM sodium acetate pH 5.5, with 1 mM TCEP and 1 mM ZnSO$_4$.

The addition of unlabelled TACC77 to $^{15}$N-FOG-F3$^{\text{KRA}}$

The $^{15}$N-HSQC spectrum of $^{15}$N-FOG-F3$^{\text{KRA}}$ (Figure 4.24A, black) yielded the expected number of signals and was assigned using a combination of the homonuclear chemical shift data (Section 4.2.3) and an HNHA spectrum. During the addition of unlabelled TACC77, a number of resonances disappeared, and several more reduced in intensity and/or underwent small shifts (Figure 4.24A, red). Thus, the kinetics of the interaction was such that the interaction was in the intermediate to fast exchange regime on the chemical shift timescale.

In a simple two-site exchange reaction, where the populations of each site are equal, the observed exchange regime is dependent on both the frequency difference between the two states and the exchange rate. The exchange rate at which the transition from slow to fast exchange occurs is related to the frequency difference as in Equation 4.2:

$$k_{ex} = \frac{\pi \Delta \nu}{\sqrt{2}}$$

(Equation 4.2)

where $k_{ex}$ is the rate of exchange and $\Delta \nu$ (=$|\delta_A-\delta_B|$) is the frequency difference between the two states. On the chemical shift timescale, slow exchange occurs when $k_{ex} << |\delta_A-\delta_B|$, intermediate exchange when $k_{ex} \approx |\delta_A-\delta_B|$ and fast exchange when $k_{ex} >> |\delta_A-\delta_B|$. This equation, while not accurate for a situation where the two sites are differentially populated, provides a useful approximation.
A novel zinc finger-mediated interaction between FOG and TACC3

Figure 4.24. 15N-HSQC spectra of 15N-FOG-F3KRA alone and in the presence of unlabelled TACC77. A. Spectra of 15N-FOG-F3KRA alone (black) and with 2.9 molar equivalents of TACC77 (red) are superimposed, and resonances are labelled with residue assignments. B. The boxed region in A displayed as a function of the concentration of added TACC77. Each spectrum has been shifted as indicated by the arrow in order to show the changes more clearly. The concentrations of TACC77 are 0, 82, 127 and 159 µM in the black, purple, red and orange spectra, respectively. The concentration of 15N-FOG-F3KRA was initially 110 µM. It can be seen that resonances from V354 and R350 disappear early in the titration.
In this titration experiment, the exchange process involves bound and unbound forms of $^{15}\text{N}$-FOG-F3$_{\text{KRA}}$. The residues that underwent line broadening are considered to be undergoing intermediate exchange, whereas resonances in fast exchange underwent smaller chemical shift changes during the transition from the free to the bound state.

The marked line broadening was specific to a few residues, and was distinct from smaller increases in linewidths observed for most resonances due to increase in correlation time on complex formation. In order to accurately define residues that underwent the most significant changes, changes in peak volume as a function of TACC77 concentration were plotted. Figure 4.25A shows the data for two resonances: T343 underwent dramatic line broadening, while S340 experienced only a moderate decrease in volume. An exponential function

$$\text{volume} = A \exp\left(-[\text{TACC77}] \times R \right)$$

(Equation 4.3)

where $A$ is a constant and $R$ is the rate of change, was used to fit the data for all resonances. As shown in Figure 4.25B, $R$ was largest for A341, T343, A346, R350, H351, V354 and D357. These residues are likely to be involved in the FOG:TACC3 contact surface.

![Figure 4.25. Quantification of peak volume changes in the $^{15}$N-FOG-F3$_{\text{KRA}}$:TACC77 titration. A. Changes in volume for T343 (filled) and S340 (open squares), fitted to a single exponential function. B. Rate of change (determined from the fitted function in A) of each residue of FOG-F3$_{\text{KRA}}$. The residues that underwent large changes in volume (above an arbitrary cutoff) are coloured grey. Residues at the termini, such as S328, G329, K330 and S360, and P332 were either absent from the spectra or increased slightly in peak volume, and therefore decay rates could not be calculated.](image-url)
The lifetime of the complex (\(\tau\)) may be estimated from chemical shift changes of resonances undergoing intermediate exchange. In the intermediate exchange regime, line broadening is most pronounced when the chemical shift difference between the two states approaches the lifetime of the complex. Hence the rate of exchange becomes limited by the off-rate, or dissociation rate constant (\(k_{\text{off}}\)), and therefore Equation 4.2 can be re-written as Equation 4.4.

\[
k_{\text{off}} = \frac{\pi \Delta \nu}{\sqrt{2}} \quad \text{where} \quad k_{\text{off}} = \frac{1}{\tau} \quad \text{(Equation 4.4)}
\]

For the FOG:TACC3 complex, a lower limit for the off-rate was estimated to be \(\sim 30 \text{ s}^{-1}\), corresponding to a lifetime of \(\sim 30 \text{ ms}\).

In order to obtain a measure of the association constant from the HSQC titration data, the chemical shift changes of several resonances that underwent fast exchange were examined. A compound chemical shift change was calculated according to the following equation:

\[
\Delta \delta_{\text{tot}} = \sqrt{\left(\Delta \delta_{HN} W_{HN}\right)^2 + \left(\Delta \delta_{N} W_{N}\right)^2} \quad \text{(Equation 4.5)}
\]

where \(\Delta \delta\) is the difference between the chemical shifts of nucleus \(i\) in the free and bound forms and \(W_i\) denotes a weighting factor (Ayed et al., 2001). The weighting factors were calculated from the ratio of the average variances of the amide proton and nitrogen chemical shifts observed for the 20 common amino acids in proteins deposited with the BioMagResBank (Seavey et al., 1991; http://www.bmrb.wisc.edu). The backbone amide \(^1\text{H}\) and \(^{15}\text{N}\) shifts dispersion are relatively independent of amino acid type and the average variances are \(\sigma_{HN}=0.66(\pm0.07)\) ppm and \(\sigma_N=4.30(\pm0.55)\) ppm, giving \(W_{HN}=1\) and \(W_N=0.1535\). The compound chemical shift changes were plotted against the concentration of added TACC77 (Figure 4.26) for several residues and a binding curve was fitted to the following model.

During the titration, FOG-F3KRA (\(F\)) and TACC77 (\(T\)) interact to form a complex (\(FT\)) with an association constant of \(K_4\).

\[
F + T \rightleftharpoons FT \quad \text{and therefore} \quad K_4 = \frac{[FT]}{[F][T]} \quad \text{(Equation 4.6)}
\]
Given that the total concentrations of F \((F_{\text{tot}})\) and T \((T_{\text{tot}})\) at every titration point are known, then \([FT]\) can be calculated:

\[ K_A = \frac{[FT]}{(F_{\text{tot}}-[FT])(T_{\text{tot}}-[FT])} \]  

(Equation 4.7)

\[ [FT] = \sqrt{\psi^2 - (4 \times F_{\text{tot}} \times T_{\text{tot}})} \]  

(Equation 4.8)

For resonances undergoing fast exchange, one signal \((\delta_{\text{obs}})\) is observed at the population-weighted averaged chemical shift of the nuclei in the free \((\delta_F)\) and bound \((\delta_{FT})\) states as in Equation 4.9.

\[ \delta_{\text{obs}} = \delta_F f_F + \delta_{FT} f_{FT} \]  

(Equation 4.9)

where \(f_F\) and \(f_{FT}\) are the fractions of free and bound FOG-F3KRA, respectively, \(f_F = [F]/F_{\text{tot}} = 1 - f_{FT}\) and \(f_{FT} = [FT]/F_{\text{tot}}\). Then, the observed chemical shift becomes dependent on the concentrations of the bound species, according to Equation 4.10.

\[ \delta_{\text{obs}} = \delta_F + (\delta_{FT} - \delta_F) \frac{[FT]}{F_{\text{tot}}} \]  

(Equation 4.10)

Thus, substitution of Equation 4.8 into Equation 4.10 yields a function that can be fitted to the chemical shift data, in which \(K_A\) and \(\delta_{FT} - \delta_F\) are fitted parameters and \(F_{\text{tot}}\) and \(T_{\text{tot}}\) are independent variables. Independent fitting of the chemical shift changes of T344, N347, C348, D357 and T358 gave an average \(K_A\) of \((1.2 \pm 1.1) \times 10^4\) and \((2.8 \pm 2.4) \times 10^4\) M\(^{-1}\) for FOG binding to monomeric and dimeric TACC77, respectively (Figure 4.26). Despite the fact that only a small part of the binding isotherm could be obtained, the calculated values compare well with those determined using isothermal titration calorimetry, \((3.6 \pm 2.4) \times 10^4\) and \((1.4 \pm 0.7) \times 10^4\) M\(^{-1}\), respectively.

**Figure 4.26.** Chemical shift changes of N347 and T358 as function of TACC77 concentration. The data were fitted to a binding model of FOG interacting with dimeric TACC77.
4.4.4 Residues necessary for binding were determined by site-directed mutagenesis

(Results described in this Section were obtained by Stella Lee in the laboratory of Merlin Crossley, but are described here for completeness.)

An alanine scanning strategy using yeast two-hybrid assays (Fields and Song, 1989, Chien et al., 1991) was carried out concurrently in order to identify the residues involved in the interaction and corroborate the $^{15}$N-HSQC titration data.

The screen was performed in both directions. First, TACC103 was tested for its ability to bind to a panel of FOG-F3 proteins containing single alanine mutations. Residues omitted in the mutational screen were those required for structural integrity of the zinc finger domain, such as the zinc-chelating residues and conserved hydrophobic residues. Second, wild-type FOG-F3 was screened against a panel of TACC47 mutants containing double alanine substitutions. Residues at heptad positions $a$ and $d$ were not mutated as they form interactions necessary for coiled-coil structure. The results are shown in Figure 4.27.

The mutation of several residues in either FOG-F3 or TACC3 were found to abrogate or severely weaken the formation of the FOG:TACC3 complex. For TACC3, the essential residues were located at the C-terminal region of the protein. In addition, substitutions of the most C-terminal residues severely compromised the interaction supporting earlier data that showed a truncated TACC domain without the C-terminal ten residues was unable to bind FOG.
Figure 4.27. Analysis of the FOG:TACC3 interaction by yeast two-hybrid assays. A. A panel of FOG-F3 mutants was screened for their ability to interact with wild-type TACC103. B. TACC47 mutants were tested for their binding ability for FOG-F3. Binding activities were judged by the relative growth of yeast after 10 days incubation at 29 °C. These experiments were carried out by S. Lee.

4.4.5 Mapping of the FOG and TACC3 binding surfaces

The residues implicated in the FOG:TACC3 interaction from the preceding studies are listed in Table 4.2.

Table 4.2. Residues implicated in the interaction between FOG-1 and TACC3.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Technique</th>
<th>Residues involved in interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOG-F3</td>
<td>NMR HSQC titration</td>
<td>A341, T343, A346, R350, H351, V354, D357</td>
</tr>
<tr>
<td></td>
<td>Yeast two-hybrid</td>
<td>T343, T344, N347, R350, V354</td>
</tr>
<tr>
<td>TACC47</td>
<td>Yeast two-hybrid</td>
<td>D621/E622, R625/I626, D628/D629</td>
</tr>
</tbody>
</table>
TACC-binding surface of FOG-F3

Figure 4.28 shows the solution structure of FOG-F3\textsubscript{KRA}, highlighting the residues implicated in its interaction with TACC3. The residues formed a contiguous surface along the length of the $\alpha$-helix, and the data from NMR (Figure 4.28A) and mutagenesis (Figure 4.28B) studies are consistent with one another, which serve to confirm the integrity of the results. The two techniques are clearly complementary. For example, while H351 was not tested in the mutagenesis screen (due to its requirement for ZnF structure), its involvement in TACC3 binding was determined in the NMR experiment. As Figure 4.28C illustrates, the three mutations in FOG-F3\textsubscript{KRA} that were introduced to improve protein solubility do not lie within the TACC3-binding surface.

![Figure 4.28. TACC3-binding surface of FOG-F3\textsubscript{KRA}. A. The residues delineated by $^{15}$N-HSQC titration are shown in orange. B. Residues implicated in yeast two-hybrid assays are in purple. C. FOG-F3\textsubscript{KRA} shown in the same orientation as in A and B. Residues in light blue are those mutated to produce the more soluble FOG-F3\textsubscript{KRA}. In all structures, zinc ions are shown in magenta and zinc ligands in dark blue.](image)

FOG-binding surface of TACC3

In order to illustrate the contact surface on TACC3, the residues implicated by mutagenesis were mapped onto a homology model of TACC77. This model was generated using the dimeric coiled-coil domain of cortexillin (PDB code 1D7M, Burkhard et al., 2000) as a template in the program SwissModel (Schwede et al., 2003). The template structure of cortexillin was chosen as it was sufficiently long to model the ~100 residues of the TACC3 domain. The modelled structure comprises...
residues 579–637 (I637 is the C-terminus of the protein) and does not include the three-residue insertion to avoid disturbing the coiled-coil structure. It should be noted that this insertion is well removed from the FOG-binding site. As shown in Figure 4.29, the results of the alanine scan implicate a 9–10 residue segment of the coiled-coil as the docking site for the $\alpha$-helix of FOG-F3. This surface is composed of $b, c, f$ and $g$ residues (of the coiled-coil heptad) and is negatively charged.

![Figure 4.29. FOG-binding surface of the TACC3 domain.](image)

The model is based on the template of cortexillin (PDB code 1D7M; Burkhard et al., 2000). Residues implicated for binding FOG-1 are mapped onto both $\alpha$-helices, hence two putative FOG-binding surfaces are displayed on the dimeric coiled-coil.

### 4.4.6 A model of the FOG:TACC3 complex

A recently described docking protocol HADDOCK (Domínguez et al., 2003) was used to generate a representation of the FOG:TACC complex. HADDOCK (the high ambiguity driven docking approach) utilises information concerning residues found to be important for binding in biochemical and biophysical interaction studies, such as NMR HSQC titrations and mutagenesis screens. This information is used as ambiguous interaction restraints (AIRs), which drive the docking of the two proteins in the program ARIA 1.2 (Nilges, 1995, 1997a). The structures of the free proteins (in this case, structure of FOG-F3$_{KRA}$ and the model TACC3 domain) and the residues demonstrated to be involved in the interaction (see Table 4.2) were used as input. AIRs were defined as ambiguous intermolecular distances with a maximum value of 3 Å between nominated residues of the two proteins. The structures were refined initially by rigid body energy minimisation, then with semi-rigid simulated annealing, during which amino acids at the interface (sidechain and backbone atoms) were allowed to move to in order to optimise the interface packing.
An overlay of the 10 lowest energy structures from this procedure showed that the
docking calculation converged well and reveals a representation of the structure of the
complex (Figure 4.30A). A total of \( \sim 1000 \text{ Å}^2 \) of surface area is buried in this model,
around half the average for protein complexes (Jones and Thornton, 1996). This is
consistent with the relatively weak association constant. A number of specific
electrostatic contacts are also observed along the length of the interface (Figure
4.30C). For example, the carbonyl oxygen of D357 in FOG-1 contacts the
guanidinium group of R625 in TACC3 and in a similar way, the guanidinium group
of R350 in FOG-1 interacts with the D629 sidechain carbonyl oxygens of TACC3.

Figure 4.30. Structural model of the FOG:TACC3 complex. A. The 10 lowest energy
conformers generated using HADDOCK. Only one half of the TACC3 dimer was used in the
calculations. B. Overall structure of the complex. The interfacial residues are displayed in
space-fill representation (yellow for FOG-F3 and green for TACC3). C. The FOG:TACC3
interface. The residues implicated in the interaction are labelled (red for FOG-F3 and blue for
TACC3).

The stoichiometry of the FOG:TACC3 complex remains unclear. While a complex of
one FOG protein bound to a dimeric TACC domain fitted the ITC data with the
lowest \( \chi^2 \) value, more direct data are needed before this conclusion could be
considered to be reliable. The docking model of the complex suggests that the dimeric
TACC domain will be able to accommodate the association of one or two FOG-F3
molecules.
4.5 Discussion

The FOG:TACC3 interaction is mediated by the third zinc finger of FOG-1. This domain is a true classical zinc finger and it interacts with TACC3 using its \(\alpha\)-helix. The region of TACC3 that contacts FOG is a short segment at the most C-terminus of the protein. This region is part of a coiled-coil domain that is primarily dimeric in solution. The interaction can be described in general terms as an interaction between \(\alpha\)-helices, those of a classical ZnF and a coiled-coil domain.

4.5.1 Classical zinc fingers as protein-binding domains

Classical CCHH zinc fingers are well-recognised DNA-binding domains (see Chapter 1). The \(\alpha\)-helices of these domains contact the major groove of DNA, and in particular residues at positions -1, 1, 2, 3 and 6 of the helix are important for making contacts with DNA (position 1 is the N-terminal residue of the helix).

While these domains are also known to act as protein recognition motifs in certain cases, the manner in which such interactions take place is not well understood. For example, some classical ZnFs of the Ikaros proteins are involved in the formation of oligomers with other family members to modulate the transcriptional activity of these proteins (Georgopoulos, 2002). Also, the classical ZnFs of the Krüppel-like factor proteins mediate protein-protein interactions with proteins such as GATA-1 (Turner and Crossley, 1999). While the molecular basis of these interactions is not known, the protein-protein interaction between a GATA ZnF and a variant CCHC FOG finger is now well understood (see Section 1.2.2 and Chapter 3). The results of this work therefore add to our understanding of protein-protein interactions by providing some of the first structural data available as to how a classical zinc finger contacts a protein partner.

FOG utilises its third classical ZnF to contact the TACC3 protein. Remarkably, residues that contact TACC3 lie in the same positions as those that are usually used in DNA recognition (Figure 4.31). It is notable that the surface used by FOG-F3 to contact TACC3 is more extensive than surface typically used in DNA binding. It extends further along the \(\alpha\)-helix, probably because the TACC3 coiled-coil is larger
than the major groove of double-stranded DNA, which does not permit contacts to be made by the entire α-helix of a classical ZnF.

Figure 4.31. Comparison of the binding surfaces of FOG-F3 and a DNA-binding classical ZnF. On the left, FOG-F3KRA is displayed with Cα atoms of TACC-binding residues shown as yellow spheres. The structure of the first ZnF of Zif268 bound to 4 bp of DNA (PDB code 1ZAA; Pavletich and Pabo, 1991) is shown on the right. Cα atoms of the residues that make the majority of the interactions with DNA are indicated with blue spheres. Only around half of the length of the α-helix is used in contacting the double-stranded DNA (shown in space-fill representation).

By analogy with the well-characterised DNA-binding properties of classical ZnFs, these results hint at the possibility of understanding the protein binding capacity of classical ZnFs in terms of understanding interactions between α-helices. However, it is likely that protein-protein interactions involving classical ZnFs will not be limited to interactions between α-helices, given the diverse and non-repetitive nature of protein structure.

The finding of a classical zinc finger that interacts with a protein partner underlines the significance of searching for potential protein binding activity as well as nucleic acid binding when studying novel classical ZnF proteins. This becomes increasingly important with the availability of large amounts of sequence data. In this context, it is noted that many classical ZnFs in protein sequences are not part of tandem arrays usually associated with DNA binding activity, and it seems likely that such domains may therefore act as protein recognition motifs. Such speculation then leads to the
question as to whether primordial classical ZnFs were binders of proteins or nucleic acids. The utilisation of only one ZnF domain for protein binding compared to an array required for DNA binding suggests that zinc finger progenitors may be protein binders.

### 4.5.2 Conservation of binding residues in other FOG and TACC proteins

**FOG-binding residues in TACC proteins**

Members of the TACC protein family are defined by homology in the ~200-residue coiled-coil domain at the C-termini. Sequences outside these regions are fairly divergent. The residues identified to be necessary for mTACC3 to bind mFOG-1 are located in the conserved coiled-coil region and these FOG-binding residues, like many residues in this region, are highly conserved throughout the different family members. All the residues are either fully conserved (Ile) or highly conserved (Asp→Glu, Arg→Lys), with the exception of one residue in *Drosophila* TACC (Asp→Ala; Figure 4.32). Therefore it is likely that FOG-1 is capable of interacting with other members of the TACC family.

![Sequence alignment of TACC domains](image)

**Figure 4.32. Sequence alignment of TACC domains.** Only the most C-terminal ~25 residues are shown. The FOG-binding residues are boxed in grey. Maskin is the *Xenopus* homologue of TACC3.

**TACC-binding residues in FOG proteins**

The FOG protein family includes mammalian FOG-1 and -2, *Xenopus* FOG and *Drosophila* Ushaped. As shown in Figure 4.33A, *Drosophila* Ushaped differs from the rest of the family in the relative positions of the classical and variant ZnFs. When the sequences of the classical zinc finger 3 of FOG-1, -2 and xFOG are compared, it is clear that finger 3 of murine and human FOG-1 and xFOG are highly conserved,
whereas the corresponding domain in FOG-2 appears to be completely divergent (Figure 4.33B). The sequence homology is consistent with the expression patterns of FOG proteins. In mammals, expression profiles of FOG-1 and -2 are different and non-overlapping (Tevosian et al., 1999, Lu et al., 1999, Svensson et al., 1999) suggesting that the two FOG proteins perform different functions. The expression profile of xFOG encompasses those of mammalian FOG-1 and –2, indicating that in Xenopus one FOG protein exists to carry out the functions of FOG (Deconinck et al., 2000). The conservation of the binding residues in murine, human and Xenopus FOG and TACC proteins suggests that a common FOG:TACC interaction may exist in frogs and mammals.

Figure 4.33. Sequence alignment of FOG proteins. A. The distribution of classical CCHH (white) and variant CCHC (grey) ZnFs differ in the various FOG proteins, with Drosophila Ushaped being the most different. B. Sequences of finger 3 of murine and human FOG-1 and -2, and Xenopus FOG are displayed, together with those of fingers 2 and 4 of murine FOG-1. The residues important for binding TACC3 are boxed in grey.

It is possible that other classical fingers may contact TACC3. Deletion mutagenesis of FOG-1 indicated that fingers 2 and 4 alone did not bind TACC3 (S. Lee and M. Crossley, unpublished results). However this does not eliminate the possibility of a cooperative binding event involving several FOG ZnFs and a larger portion of the TACC3 coiled-coil. None of the other classical FOG fingers appears to contain the required residues for TACC-binding. Some examples are shown in Figure 4.33B, where finger 4 of mFOG-1 is most conserved and contains three of eight TACC-binding residues. This search for TACC-binding residues in other ZnFs
however assumes that the fingers bind TACC3 in a similar manner to that elucidated for finger 3. In the situation where finger 3 together with adjacent ZnFs contact larger regions of the TACC3 coiled-coil (similar to the manner in which classical ZnFs in tandem arrays bind to consecutive basepairs of DNA), the other ZnFs are likely to utilise different residues to bind other segments of the TACC3 supercoils. Thus, the results presented here involve the minimal regions for interaction and it is possible that additional contacts made by other ZnFs may increase the binding association of the overall FOG:TACC3 interaction.

4.5.3 Weak interactions in biology

The affinity of the FOG:TACC3 interaction is relatively weak, but is comparable to that measured for the interaction between the variant CCHC FOG fingers and the N-terminal zinc finger of GATA-1 (see Section 3.4.3; Liew et al., 2000, Kowalski et al., 2002). The FOG:GATA interaction has been established as indispensable for normal erythropoiesis (Crispino et al., 1999). While it is possible that high local protein concentrations as a result of nuclear compartmentalisation, recruitment by other factors, stronger interactions in the context of full-length proteins or molecular crowding can increase the strength of these weak interactions substantially, there is no doubt that low affinity interactions can also be advantageous (Nooren and Thornton, 2003, Crowley and Ubbink, 2003). Weak associations potentially allow rapid exchange of binding partners and may be appropriate to effect rapid responses to transcriptional stimuli for the precise control of gene expression (Misteli, 2001).

4.5.4 Functional implications of a FOG:TACC3 complex

FOG proteins are essential protein coactivators of GATA factors in many biological processes (see Section 3.1.1). The finding that FOG-1 interacts with TACC3 is of particular interest as murine TACC3 was originally identified as erythropoietin-induced cDNA (ERIC) and is thus also implicated in the control of erythropoiesis (McKeveney et al., 2001). TACC3 knockout mice display defective haematopoietic stem cells (Piekorz et al., 2002) and forced expression of TACC3 in haematopoietic cell lines impaired terminal erythroid maturation (Garriga-Canut and Orkin, 2004). Together, these suggest a role for TACC3 in haematopoiesis.
To date, the biological function of the FOG:TACC3 interaction is unclear. Garriga-Canut and Orkin showed that TACC3 (a cytoplasmic protein) is able to retain FOG-1 in the cytoplasm and thus affect the amount of nuclear FOG-1 available to interact with GATA-1 (Garriga-Canut and Orkin, 2004). The observations that TACC3 is localised in the centrosome and is able to interact with structural proteins such as tubulin also raises the possibility that the FOG:TACC3 complex play a role in chromosomal movement. However other roles are possible. The *Xenopus* TACC protein Maskin appears to be involved in regulating translation of specific messenger RNAs in the developing frog oocyte (Stebbins-Boaz *et al.*, 1999). Murine TACC3 is also known as AINT (ARNT interacting protein) and is implicated in the regulation of transcription factor ARNT (aryl hydrocarbon nuclear translocator protein) (Sadek *et al.*, 2000). The coiled-coil domain of TACC3 also mediates its interaction with ARNT. Residue-specific information pertinent to this interaction has not been determined and therefore it remains unclear as to whether a ternary complex of FOG-1, TACC3 and ARNT is possible. ARNT is known to bind and translocate the AH receptor into the nucleus, where it binds xenobiotic response elements of target genes (Sadek *et al.*, 2000). In cells overexpressing TACC3, ARNT was no longer localised in the nucleus and instead was found in both the cytoplasmic and nuclear compartments (Sadek *et al.*, 2000). Significantly, the TACC:ARNT interaction indicates a direct role for TACC3 in the regulation of gene expression. Further experiments will be necessary to elucidate the multiple functions of TACC3, and in particular, how TACC3 cooperate with FOG to regulate gene expression in haematopoietic cells.

### 4.5.5 Summary

A novel interaction between FOG and TACC3 has been characterised at the molecular level. These studies have identified a novel mode of interaction for a classical zinc finger: the use of the α-helix to bind proteins. This extends the role of the α-helix as the contact site for DNA to include that of proteins as well.

An emerging property of the multi-zinc-finger protein FOG-1 is that various zinc fingers perform different roles; fingers 1, 5, 6 and 9 can bind GATA-1, while finger 3 can contact TACC3. It will be interesting to understand the roles of the remaining zinc
fingers of FOG and thereby build up a detailed picture of how these multi-zinc-finger proteins bring a number of functions together by the use of discrete domains.

It is notable that the only difference between the GATA-binding and TACC3-binding ZnFs of FOG lies in the identity of the last zinc-chelating residue. This observation raises several structural questions. How important is the identity of the last zinc ligand in classical ZnFs? Is the fold sufficiently robust to accommodate other residues beside cysteine and histidine? These issues are discussed in the next Chapter.
Chapter 5. Investigating the ability of CCHX sequences to form functional zinc-binding modules

5.1 Introduction

Classical zinc fingers are defined by their sequence similarity to the consensus of $\Phi$-X-$\Phi$-$X_{2,4}$-$C$-$X_{3}$-$\Phi$-$X_{5}$-$\Phi$-$X_{2}$-$H$-$X_{2,5}$-$H$, where $\Phi$ is a hydrophobic and X any residue. Thus, the zinc coordination sphere comprises two cysteines followed by two histidines (a CCHH topology) and the three hydrophobic residues, together with the cysteines, form the core of the domain. These domains have been characterised in detail over the last 20 years, and ~50 structures containing classical ZnF(s) have been determined. The recent realisation that the variant CCHC zinc fingers found in FOG family proteins share the same structure, despite utilising a cysteine as the fourth zinc ligand (Liew et al., 2000) raises the possibility that the identity of the last zinc ligand is a malleable criterion in the definition of a classical ZnF. Such a scenario has implications both for genome annotation (i.e., the identification of classical zinc fingers from sequence data) and for the design of mutagenesis experiments (i.e., whether or not the mutation of one of the zinc ligands likely to disrupt function).

5.1.1 BKLF finger 3: a model CCHH zinc finger

In order to address this issue, the transcriptional repressor protein basic Krüppel-like factor/Krüppel-like factor 3 (BKLF) (Crossley et al., 1996) was chosen as a model protein. BKLF contains a tandem array of three classical CCHH fingers at its C-terminus (Figure 5.1). These ZnFs are also known as Krüppel-like fingers as they contain significant homology to those found in the archetypal *Drosophila* regulatory protein Krüppel (Ruppert et al., 1988). BKLF utilises its three Krüppel-like fingers to bind with high affinity and specificity to DNA containing CACCC sequences (Crossley et al., 1996). In addition to their roles in binding DNA, the ZnFs also mediate the interaction between BKLF and the erythroid transcription factor GATA-1 (Turner, 2001). The biological roles of BKLF and other members of the Krüppel-like factor family are reviewed by Turner and Crossley (1999).
5.1.2 The aims of this study

The aim of the work described in this Chapter was to assess the degree of flexibility that exists in the identity of the final zinc-binding ligand. That is, to what extent can this final position be mutated without disrupting the structure and function of a classical zinc finger?

To address this, a panel of mutants containing single site mutations at the position of the fourth zinc ligand in BKLF finger 3 (BF3) was produced and these single finger CCHX mutants were tested to determine if they were able to bind zinc and form folded structures. In addition, binding experiments were performed to establish whether these CCHX modules were able to carry out functions of the wild-type protein, such as the ability to interact with DNA or to interact with a protein partner.

5.2 CCHX sequences in the database

A search of the Pfam database (http://www.sanger.ac.uk/Software/Pfam/; Bateman et al., 2002) using the search pattern Φ-X-C-X1–5-C-X3-Φ-X5-Φ-X2-H-X3–6-H/C identified 28,442 protein domains classified as CCHH zinc fingers (as of March 4, 2004). The number of classical ZnFs identified using this search pattern is currently increasing at a rapid rate, with ~8,000 new domains listed in the last 12 months (20,646 hits on March 10, 2003). This trend is in line with the growing amount of sequence information and confirms the status of zinc fingers as a very abundant protein domain. Within the 20,646 sequences identified in 2003, ~250 domains actually contain a cysteine in place of the final histidine residue. More interesting are the 82 sequences that lack either a cysteine or a histidine in this position (Figure 5.2).
Of these, 11 sequences contain a cysteine or histidine residue at nearby positions and it is conceivable that this residue acts as the fourth zinc ligand. The remaining 71 sequences do not exhibit a cysteine or histidine residue in an eight-residue range (positions 20–28 in Figure 5.2) and thus appear to be missing a zinc ligand at the fourth coordination site.

**Figure 5.2. Alignment of CCHX sequences.** Also included in the alignment are the consensus sequence of classical CCHH ZnFs and sequences of two classical ZnFs: BKLF finger 3 and FOG-1 finger 3. Positions of the conserved zinc-chelating residues are shaded in grey, with that of the fourth position boxed.

Given the dogma that four ligating amino acid sidechains are required for the formation of a classical zinc finger structure, the existence of these CCHX variants led us to ask whether such sequences might form functional zinc finger domains, or if instead they represent vestigial domains that can no longer function.
5.3 The zinc-binding properties of CCHX domains

5.3.1 Production of the mutant BF3X proteins

A panel of BKLF finger 3 (residues 316–344) mutants with single substitution mutation at the position of the fourth zinc-ligating residue (residue 341) was produced using standard overlap PCR mutagenesis. In total, six mutants were produced as GST fusion proteins (H341A/D/E/N/Q/R), and termed BF3X, where X represented the amino acid at the position of the fourth zinc ligand. Aspartate is known to act as a zinc ligand in some LIM domains (Bach, 2000). Glutamate, asparagine and glutamine were selected as residues with sidechains that could potentially act as ligands to Zn$^{2+}$, while alanine and arginine were chosen as negative controls.

![Figure 5.3. SDS-PAGE analysis showing the overexpression of GST-BF3X mutants. For each of the six mutants, cell contents prior to protein induction and at the end of overexpression (~3 h, 37 °C, 0.4 mM IPTG) are displayed in the left and right lanes, respectively.](image)

Each of the mutants was successfully overexpressed (Figure 5.3) and subjected to glutathione-affinity chromatography; thrombin was then used to release the BF3X domains. Treatment with thrombin, however, cleaved the BF3X proteins at several sites other than the engineered site, as judged by the elution of multiple protein fragments during subsequent reverse phase HPLC (Figure 5.4). This result suggested that single point mutations had increased the accessibility of the domain to the protease, compared to wild-type, which showed no such behaviour. Fortunately, useable amounts of intact BF3A, BF3D, BF3E and BF3N were isolated (Figure 5.4), although neither BF3Q nor BF3R could be isolated. The identities of the A, D, E and N mutants were confirmed with electrospray ionisation mass spectrometry.
5.3.2 Mutants form stable secondary structures in a Zn$^{2+}$-dependent manner

In order to ascertain whether the mutant domains were able to bind Zn$^{2+}$, far-UV CD spectra were recorded. The spectra of wild-type BF3 and each of the mutants at low pH (pH ~2) and in the presence of excess Zn$^{2+}$ were typical of unstructured polypeptides. Surprisingly, an increase of the pH to 5.5 resulted in noticeable increases in the secondary structure content of all four isolated mutants (Figure 5.5, coloured lines). This was manifested as a red shift of the minimum and, in the case of the wild-type protein, by the presence of positive ellipticity at low wavelengths. Spectra recorded at pH 5.5 in the absence of Zn$^{2+}$ were similar to the low pH spectra. The CD spectra indicated that all CCHX mutants, including the alanine mutant, were able to form stable secondary structure in the presence of Zn$^{2+}$ and at an appropriate pH.

Figure 5.4. Reverse phase HPLC chromatograms of BF3X mutants. All proteins were eluted with a 20-min linear gradient of 5→95 % acetonitrile starting at t = 6 min. Peaks marked with red circles were identified by mass spectrometry to correspond to the intact peptide.
Upon closer inspection of the BF3 sequence, a histidine residue representing a potential alternative zinc ligand was noticed at position 333 (Figure 5.6). Thus, two additional constructs were synthesised. The first corresponded to BF3 with a single H333A mutation (BF3-H333A), while the second additionally contained the H341N mutation at the position of the fourth zinc ligand (BF3N-H333A; Figure 5.6A). These two proteins were synthesised and purified as previously described for the BF3X mutants. When BF3-H333A and BF3N-H333A were placed in conditions containing excess Zn\(^{2+}\) and pH ~5, both proteins formed conformations with substantial secondary structure, as determined by CD spectropolarimetry (Figure 5.6B). The CD spectra contained minima at ~207 nm and positive ellipticity at low wavelengths, indicating that H333 is not compensating for the loss of H341 in the BF3X mutants.

![Figure 5.5. Far-UV CD spectra of wild-type and mutant BF3X proteins.](image)

Spectra recorded at pH 5.5 are shown as coloured lines, and the spectrum in black was collected at pH ~2 (25 °C).

![Figure 5.6. His333 is not involved in chelating Zn\(^{2+}\).](image)

A. Amino acid sequences of H333A mutants, with mutated residues shown in red. B. CD spectra of wild-type (red), H333A (green) and N-H333A (blue). Spectra were recorded at pH 5–6 (coloured spectra) and pH ~2 (black spectrum).
5.3.3 Mutant BF3N and wild-type proteins bound Zn\(^{2+}\) with similar affinities

Having determined that the CCHX sequences formed conformations with stable secondary structure in the presence of Zn\(^{2+}\), the Zn\(^{2+}\)-binding affinities of both wild-type and BF3N were determined in Zn\(^{2+}\) titrations monitored by CD. Zn\(^{2+}\) was titrated into solutions containing each domain, and changes in ellipticity at wavelengths between 195 and 200 nm were recorded (Figure 5.7). The change in ellipticity as a function of Zn\(^{2+}\) concentration was fitted using the model described below.

The addition of Zn\(^{2+}\) to the apoprotein resulted in formation of the holoprotein with a zinc-binding constant (\(K_{Zn}\)) defined according to Equation 5.1.

\[
P + Zn \rightleftharpoons P \cdot Zn \quad \text{where} \quad K_{Zn} = \frac{[P \cdot Zn]}{[P][Zn]} \quad \text{(Equation 5.1)}
\]

Equation 5.1 can be rewritten using total concentrations of protein (\(P_{tot}\)) and Zn\(^{2+}\) (\(Zn_{tot}\)) (Equation 5.2) and rearranged to form Equation 5.3 that describes the amount of holoprotein in the solution.

\[
K_{Zn} = \frac{[P \cdot Zn]}{(P_{tot} - [P \cdot Zn]) (Zn_{tot} - [P \cdot Zn])} \quad \text{(Equation 5.2)}
\]

\[
[P \cdot Zn] = \frac{\psi^2 - \sqrt{\psi^2 - (4 \times P_{tot} \times Zn_{tot})}}{2} \quad \text{where} \quad \psi = P_{tot} + Zn_{tot} + \frac{1}{K_{Zn}} \quad \text{(Equation 5.3)}
\]

The observed CD ellipticity (\(\theta\)) comprises population-weighted ellipticities of the apoprotein (\(\theta_p\)) and that of the holoprotein (\(\theta_{p \cdot Zn}\)), and is dependent on concentrations of the proteins as described below.

\[
\theta = \theta_p + \frac{[P \cdot Zn]}{P_{tot}} (\theta_{p \cdot Zn} - \theta_p) \quad \text{(Equation 5.4)}
\]

Equations 5.3. and 5.4 were fitted using non-linear least squares analysis. Independent fitting of the ellipticity changes at 195, 196, 197, 198, 199 and 200 nm revealed that wild-type and BF3N proteins bound Zn\(^{2+}\) with an affinity of \((4.6 \pm 1.4) \times 10^4 \text{ M}^{-1}\) and \((1.1 \pm 0.6) \times 10^5 \text{ M}^{-1}\), respectively. This indicated that the presence of a non-histidine residue at the position of the fourth zinc ligand did not compromise the zinc-binding affinity of the domain. Indeed, the affinity for the mutant appeared to be somewhat
higher; however, the errors associated with the titration are probably larger than estimated from the curve fit alone, and it is therefore reasonable to conclude that the zinc-binding affinity is essentially unchanged by the loss of H341.

5.3.4 The zinc:protein ratio of folded BF3D was 1:1

The stoichiometric ratios of protein to Zn\(^{2+}\) in wild-type and BF3D proteins were analysed using atomic absorption spectrometry (AAS). The proteins to be analysed were initially assessed by CD spectropolarimetry to ensure that they were folded in solution, and the protein solutions were then extensively dialysed against a buffer lacking Zn\(^{2+}\) to remove unbound Zn\(^{2+}\). The Zn\(^{2+}\) contents were then determined using AAS, which revealed that wild-type and BF3D proteins exhibited Zn\(^{2+}\) to protein ratios of 1.2:1 and 1.3:1, respectively. This suggests that both BF3 and BF3D ligate a single ion of Zn\(^{2+}\) per molecule.

5.3.5 CCHE is monomeric in solution

Having determined that the variant CCHX sequences were capable of binding Zn\(^{2+}\), one of the mutants was next subjected to sedimentation equilibrium experiments in order to determine its aggregation state in solution (see Section 4.3.3 for more information on this technique). The concentration versus radial distance profiles of mutant BF3E were obtained at two different rotor speeds, and non-linear least squares analysis showed that the data fitted very well to an ideal single species model with a molecular mass of 3,500 ± 200 Da (Figure 5.8). The value is in good agreement with the theoretical mass of 3,638 Da for BF3E containing one Zn\(^{2+}\) ion. This finding

![Figure 5.7. Zn\(^{2+}\) titrations of BF3 and BF3N. The affinities were determined from the observed changes in ellipticity with Zn\(^{2+}\) concentration (data at 199 nm are illustrated). The fitted values of $K_{Zn}$ were $(4.6 \pm 1.4) \times 10^4$ M\(^{-1}\) (wild-type, solid line) and $(1.1 \pm 0.6) \times 10^5$ M\(^{-1}\) (BF3N, dotted line).]
eliminated the possibility that one zinc ion may be bridging two protein molecules (using two ligands from each protein, for example).

![Figure 5.8. Sedimentation equilibrium data of BF3E.](image)

**Figure 5.8. Sedimentation equilibrium data of BF3E.** This data were recorded at 42,000 rpm at 25 °C. The *lower* panel displays a plot of absorbance at 230 nm versus $r^2/2$ (cm$^2$), whereas the *upper* panel shows the residual deviations resulting from the fit of an ideal single species model to the data.

### 5.3.6 Mutants appeared to be only partially folded

Next, 1D $^1$H NMR spectra of wild-type and BF3E were recorded in order to compare their structural properties. The spectrum of wild-type BF3 contained sharp and well-dispersed peaks (Figure 5.9A) indicative of a folded protein with a significant degree of tertiary structure. In contrast, the spectrum of BF3E contained broad signals (Figure 5.9B), although still exhibited a reasonable amount of chemical shift dispersion compared with the same protein at pH ~2 (Figure 5.9C). The observed broadness of the signals suggested that BF3E was undergoing exchange on the intermediate chemical shift timescale. Since BF3E is monomeric in solution, this exchange process may involve interconversion between a fully folded and a semi-folded or unfolded state. These data are consistent with the observation that CCHX domains were susceptible to multiple thrombin cleavages.
Figure 5.9. Comparison of 1D $^1$H NMR spectra of wild-type and BF3E. A. Spectrum of wild-type BF3 (100 μM) at pH 5.5. B. Spectrum of BF3E (85 μM) at pH 5.5 and C. Spectrum of BF3E (85 μM) at pH ~2. The spectra were recorded at 25 °C and only the amide/aromatic regions are shown for clarity.

Overall, mutant BKLF finger 3 proteins containing CCHX sequences were able to bind Zn$^{2+}$. At least one of these mutant proteins bound Zn$^{2+}$ with similar affinity to wild-type BF3, and all formed zinc-binding modules that contained substantial amounts of secondary structure. BF3E was monomeric in solution and contained one Zn$^{2+}$ ion per protein molecule. While the BF3X domains lacked the same degree of tertiary organisation as the wild-type protein, CCHX domains can indeed form discrete zinc-binding domains.

5.4 The protein- and DNA-binding properties of CCHX domains

The third classical ZnF of BKLF is an indispensable part of the BKLF protein. Two of the known functions of BF3 involve its interactions with the erythroid transcription factor GATA-1 (Crossley et al., 1996) and double-stranded DNA (Turner, 2001). The interaction BF3 mediates with DNA requires the presence of the two other Krüppel-like ZnFs in the protein: the three-finger array contacts DNA containing CACCC sequences (Crossley et al., 1996). In contrast, BKLF finger 3 alone is sufficient to mediate the interaction of BKLF with GATA-1 (Turner, 2001). The next issue to be addressed was therefore whether the semi-structured CCHX domains were able to interact with GATA-1 and/or DNA.
5.4.1 Can CCHX domains of BF3 bind GATA-1?

BKLF, like the homologous proteins Sp1 and EKLF, interacts with the GATA-1 protein through its three Krüppel-like ZnFs (Merika and Orkin, 1995, Gregory et al., 1996, Turner, 2001). Further, it is finger 3 of BKLF that has been shown to mediate the interaction of BKLF with the C-terminal zinc finger of GATA-1 (Turner, 2001, Cram, 2004). Therefore the CCHX:GATA interaction was probed in GST-pulldown experiments using GST-fusion BF3X proteins and MBP-fusion GATA-1 C-terminal zinc finger (C-finger or CF).

GST-BF3X proteins were initially bound to glutathione-Sepharose™ 4B beads, and a negative control was performed concurrently where GST alone was attached to the beads. The beads were then incubated with a solution containing MBP-GATA-CF and washed extensively to remove unbound proteins. As Figure 5.10 shows, wild-type and all GST-BF3X proteins were able to bind MBP-GATA-CF, whereas GST alone was not. The identities of the MBP-GATA-CF proteins were confirmed by using an antibody that detected MBP in a subsequent Western blot (Figure 5.10, lower panel). Overall, the GST-pulldown results indicated that, like the wild-type BF3 protein, the semi-structured CCHX domains were able to bind GATA-1.

Figure 5.10. GST-pulldown showing interactions between BF3X and GATA-1 C-finger. The GST-BF3X proteins were attached to glutathione beads, and used to pull out MBP-GATA-CF. The upper panel shows SDS-PAGE analysis (stained with Coomassie blue), while the lower panel shows a Western blot using anti-MBP antibodies.
5.4.2 Can CCHX domains bind DNA?

In order to assess the DNA-binding ability of CCHX sequences, constructs containing the three classical ZnFs of BKLF were synthesised. Similar to BF3 proteins, these BF1–3X proteins contained a single substitution mutation at the position of the fourth zinc ligand of finger 3 (Figure 5.11).

Figure 5.11. Amino acid sequence of BF1–3X. The position of the single point substitution in BF31–3X mutants is boxed in grey.

BF1–3X mutants were able to bind DNA

The BF1–3X mutants were tested in electrophoretic mobility shift assays (EMSAs) for their ability to bind the promoter region of the β-globin gene, which contains a typical CACCC box motif. As shown in Figure 5.12, all three fingers are required for binding, and, remarkably, all of the CCHX mutants retained near native DNA binding ability. GST alone and probe alone were also included as negative controls.

Figure 5.12. CCHX mutants bound DNA. The upper panel shows an EMSA testing the ability of GST-BF1–3X proteins to bind a CACCC box site. In lanes 1–7, minimal amounts of proteins were used to maximise the ability of the assay to detect differences in binding affinities. In lanes 8–10, DNA-binding activity of the two-finger construct BF1–2 was assessed, and more proteins were used to avoid overlooking any weak DNA-binding activity. The lower panel shows a SDS-PAGE that indicates that the proteins are essentially intact and present in similar concentrations (experiment was carried out by R. Czolij).
DNA-binding affinity was only moderately affected by H→A mutation

In order to quantify the effect of the point mutation on DNA binding, quantitative EMSAs were carried out using BF3 and the alanine mutant BF1–3A. To determine the DNA-binding affinity \( (K_a) \), a series of EMSA reactions containing a fixed concentration of DNA and variable protein concentrations were set up (Figure 5.13A and B). The fraction of DNA complexed to the BF1–3 proteins \( (f_{cx}) \) was calculated as follows.

\[
f_{cx} = \frac{I_{cx}}{I_{cx} + I_f}
\]

(Equation 5.5)

where \( I_{cx} \) and \( I_f \) are the intensities of bands corresponding to protein-complexed and free DNA, respectively (quantitated on the PhosphorImager™). In each EMSA reaction, the BF1–3 protein \( (P) \) interacts with DNA \( (D) \) to form a protein-DNA complex \( (PD) \), in a fashion analogous to that described in Equation 5.1 for the protein:Zn\(^{2+}\) interaction. The concentration of the protein-DNA complex \( ([PD]) \) becomes dependent on \( K_a \) and total concentrations of protein \( (P_{tot}) \) and DNA \( (D_{tot}) \) as described as follows.

\[
[P.D] = \frac{\psi - \sqrt{\psi^2 - (4 \times P_{tot} \times D_{tot})}}{2} \text{ where } \psi = P_{tot} + D_{tot} + \frac{1}{K_a}
\]

(Equation 5.6)

The DNA-binding affinity was obtained by fitting the experimentally-derived \( f_{cx} \) to Equations 5.5 and 5.6 using non-linear least squares methods (Figure 5.13C). This analysis showed that the wild-type protein binds the typical CACCC box site with a \( K_a \) of \( (2.5 \pm 0.4) \times 10^7 \text{ M}^{-1} \). Surprisingly, the H341A mutation had a relatively modest effect on DNA binding: the mutant BF1–3A bound DNA with a \( K_a \) of \( (2.16 \pm 0.22) \times 10^6 \text{ M}^{-1} \). Hence the removal of the fourth zinc ligand of the third and indispensable ZnF of BKLF only marginally decreased the affinity of the BKLF:DNA interaction.
A truncation mutant was able to bind DNA

Experiments were next carried out to determine whether a protein that was truncated at the C-terminus was able to bind DNA, in order to determine the extent to which DNA-binding relies on the presence of a residue at the position of the fourth zinc ligand. The truncation mutant, termed BF1–3(-HMLV), was truncated four residues from the C-terminus and hence is missing a residue at the fourth zinc coordination site of finger 3 (see Figure 5.11). Surprisingly, the truncation mutant was still able to interact with the DNA containing a CACCC box, albeit with a much weaker affinity (Figure 5.14). This data are consistent with the hypothesis that the final ZnF of BKLF does not require four zinc ligands to act as a functional DNA-binding domain.
Figure 5.14. DNA-binding ability of the BF1–3 truncation mutant. EMSAs testing the ability of three GST-BKLF constructs to bind CACCC sites are shown. From left to right, the constructs are wild-type BF1–3, BF1–3A and BF1–3(-HMLV). Protein concentrations used were 760, 500, 230 and 33 nM (lanes 1–4) for wild-type BF1–3 and BF1–3A, and higher concentrations of 8.2, 5.3, 2.5 and 0.4 µM (lanes 1–4) were used for BF1–3(-HMLV) to show the weaker binding.

5.5 Discussion

Zinc finger domains are known to form compact and stable structures in which the zinc ion is essential for structure formation. Functions performed by these domains, including the ability to bind nucleic acids, proteins and lipids, are thought to be dependent on the existence of stable structure, and all known zinc-binding domains dependent on zinc for structural integrity ligate the metal ion through four amino acid sidechains. The results presented here challenge these observations, showing that at least in the case of BKLF finger 3 only three amino acid sidechains were sufficient to form functional zinc-binding modules.

5.5.1 Zinc-binding properties of CCHX domains

A combination of CD, atomic absorption and 1H NMR data indicated that the CCHX mutants were capable of binding one molar equivalent of Zn$^{2+}$ to form substantial secondary structure. Sedimentation equilibrium experiments showed that the coordination sphere of each Zn$^{2+}$ ion only involved a single protein molecule. Also, the Zn$^{2+}$-binding affinity of the BF3 protein was not affected by the H341N mutation. Together, these data suggest that three protein ligands are sufficient (although not optimal) for Zn$^{2+}$-binding in BKLF finger 3.
This conclusion is supported by several previous findings. For example, a truncated CCHH ZnF that was missing its last zinc-ligating histidine was found to coordinate Co\(^{2+}\) in a tetrahedral manner and with an affinity comparable with that of the intact ZnF (Merkle et al., 1991). It was suggested that the fourth ligand of the zinc was likely to be a water molecule. In another study, it was observed that when one of three classical ZnFs of Zif268 was mutated to a CCHA configuration, the protein retained its ability to bind DNA (Green and Sarkar, 1998). Finally, Cook et al. (Cook et al., 1994) reported a mutant of the *Saccharomyces cerevisiae* transcriptional activator ADR1 in which the second histidine of the C-terminal CCHH finger (in a two-ZnF tandem array) was substituted to a tyrosine. This mutant retained the wild-type ability to activate transcription, and it was postulated that three protein ligands (and probably a water molecule) might be sufficient to bind Zn\(^{2+}\) and maintain the protein in its active form.

A novel function for tridentate zinc-coordinated ZnFs was recently reported (Nomura and Sugiura, 2004). In this study, a classical ZnF sequence was mutated to CCHX, XCHH and CXHH configurations. Similar to results presented here, the mutants retained the ability to bind one equivalent of Zn\(^{2+}\) and fold into compact structures. Interestingly, these proteins containing tetrahedral three-coordinate zinc sites were able to catalyse hydrolytic reactions, representing a novel designer function for zinc fingers and further supporting the results presented here that incomplete CCHX sequences could be functional.

### 5.5.2 Conformations of the CCHX domains

The CCHX mutants in BF3 did not form compact tertiary structures. This was evident from the CD and \(^1\)H NMR data, as well as the enhanced susceptibility of these mutants to proteolysis in *E. coli*. While the addition of Zn\(^{2+}\) resulted in the formation of conformations with substantial secondary structure, the CD spectra of the mutants remained different from that of the wild-type protein. The broad nature of the \(^1\)H NMR spectrum of BF3E indicated the existence of interconverting conformers in a chemical exchange process. Given that the zinc-binding affinity of the BF3N mutant was not significantly different from that of wild-type, it is likely that the chemical exchange arises from the loose packing of the CCHX mutants. This may involve the
molten globule state often discussed in the context of protein folding. This state is described as a compact, partially folded protein that contains some native secondary and tertiary structure but lacks extensive sidechain packing interactions of the native structure (Ptitsyn and Uversky, 1994). Interestingly, this partial formation of structure is sufficient to allow the recognition of both its protein and DNA partners, and it is possible that the CCHX domain forms additional regular structure concomitantly with binding.

5.5.3 DNA-binding properties of CCHX domains

The transcriptional repressor BKLF contains a tandem array of three classical CCHH ZnFs that bind to CACCC-containing DNA sequences. The binding affinity determined in this study ($3 \times 10^7 \text{ M}^{-1}$) is comparable with those reported for other known three-ZnF constructs binding to their cognate DNA sequence; for example, the three CCHH ZnFs of Sp1 bind DNA with a $K_a$ of $7 \times 10^7 \text{ M}^{-1}$ (Nagaoka et al., 2002). The presence of finger 3 of BKLF was necessary for the DNA-binding function of BKLF as fingers 1–2 alone could not bind DNA. However, when finger 3 was mutated to contain CCHX sequences, near native DNA-binding ability was observed. Indeed, the binding affinity was only reduced 10-fold when the fourth zinc ligand of finger 3 was mutated to alanine. Even more remarkable was the observation that a truncation protein lacking any residue at the fourth zinc-ligating position retained the ability to bind DNA. Together these results demonstrate that the DNA-binding contributions made by this ‘incomplete’ finger do not depend on the presence of four protein sidechains as zinc ligands.

The mechanisms underlying DNA-binding of classical ZnFs are known, due to the availability of a number of high-resolution structures of ZnF:DNA complexes (for a review, see Pabo et al., 2001). The $\alpha$-helix of each classical ZnF makes specific contacts with the major groove of DNA. Figure 5.15 shows the comparison of the DNA-binding residues on a classical ZnF of Sp1 bound to its cognate DNA with the corresponding residues on the solution structure of wild-type BF3 (Pavletich and Pabo, 1991, Simpson et al., 2003). It can be seen that the DNA-binding surface on Sp1 does not extend the length of the $\alpha$-helix and hence does not encompass the histidine residue at the fourth zinc coordination site. This comparison illustrates how
CCHX proteins and even a truncation mutant of BKLF missing several residues from the C-terminus could still participate in specific interactions with DNA.

Figure 5.15. Comparison of the putative DNA-binding residues in BF3 with those of Sp1. A. Solution structure of BF3 (Simpson et al., 2003). The histidine residue at the position of the fourth zinc coordination site is shown (sidechains in dark blue and Cα in red). Cα atoms of residues -1, 2, 3 and 6 of the α-helix corresponding to those implicated in the binding of classical ZnFs to DNA are shown as pink spheres. The three remaining zinc ligands are shown in light blue. B. The structure of the first ZnF of Zif268 bound to 4 bp of DNA (taken from the structure of a three-ZnF polypeptide bound to a 11-bp oligonucleotide; PDB code 1ZAA; Pavletich and Pabo, 1991). The C-terminal half of the α-helix is not used to contact the double-stranded DNA. DNA is shown in space-filling representation.

5.5.4 Implications for other studies

The identification of ZnFs from genomic sequence data relies in part on the presence of predictably spaced cysteine and histidine residues. While a number of variant CCHX zinc fingers were detected by the automated method used in this study, it is possible that others were overlooked, given the apparent plasticity of the requirements for a functional ZnF. Results presented here indicate that caution should be taken before presuming that CCHX zinc fingers found in putative proteins are non-functional or vestigial.

These data also cast doubt on the routine use of alanine substitutions to create null ZnFs in functional studies. Alanine-substituted ZnFs are generally regarded to be non-functional since structure and hence function of a ZnF is assumed to be dependent on the ligation of Zn$^{2+}$ by four ligands. This study indicates that, at least in
the context of some sequences, alanine substitution mutants may retain significant residual ZnF structure and activity.

The CCHH zinc finger fold is a common scaffold from which proteins with different DNA binding specificities have been generated. It is the simple structure and small number of residues required to structurally stabilise the domains that makes them particularly versatile and adaptable. The high number of CCHH zinc finger genes in eukaryotic genomes suggests that they may have evolved early in evolution, and an intermediate containing only three zinc ligands might well have played a role in their evolution. While CCHX sequences may represent remnants in the evolution of the modern day classical ZnFs, these domains where the presence of an amino acid other than histidine and cysteine as a zinc ligand (or a complete absence of a ligand) may conceivably constitute means by which ZnF proteins could be regulated. Whereas the CCHX ZnFs may still recognise their protein or DNA partner, their greater susceptibility to proteolytic degradation might reduce their cellular lifetime. This idea might be pertinent in the design of novel ZnFs as a method in which bioavailability of designed ZnFs might be controlled.
Chapter 6. General summary

6.1 Human mutations in the N-terminal ZnF domain of GATA-1

The N-finger of GATA-1 is an indispensable domain that mediates the interaction of the protein with DNA containing GATC motifs and with other proteins, including the coactivator protein FOG. The interaction of GATA-1 N-finger with DNA was characterised in this thesis (Chapter 3) and the binding affinity was determined to be $\sim 3 \times 10^7 \text{ M}^{-1}$. Further, the strength of this interaction, like many protein-DNA interactions, was dependent on ionic concentration and sample pH.

To date, five naturally occurring mutations that reside in the N-finger of GATA-1 have been identified in patients suffering from thrombocytopoenia and anaemia. When these single point substitution mutations were introduced into the N-finger domain, the mutants formed tertiary structures that were somewhat less structured than the wild-type. However, the partially structured nature of the mutants did not hinder their ability to interact with either DNA or a FOG-family protein.

It was found here that the R216Q mutation abolished the ability of N-finger to bind DNA, while FOG binding was not affected. On the other hand, the V205M and G208S mutations impaired FOG-binding and did not affect DNA binding. The D218G mutation however did not affect either DNA or FOG binding. This suggests that N-finger may associate with another protein that plays a role in the development of the blood cells, and that this interaction is compromised in the patients identified with the D218G mutation.

While this thesis examined GATA-1 N-finger mutations, other mutations of GATA-1 and other GATA family members have also recently been identified. Somatic mutations in the \textit{GATA-1} gene have been identified that lead to megakaryoblastic leukemia in sufferers of Downs syndrome (Wechsler \textit{et al.}, 2002). GATA-2 haploinsufficiency is associated with the hypoparathyroidism, deafness and renal dysplasia (HDR) syndrome (Van Esch \textit{et al.}, 2000) and GATA-4 hemizygosity has been observed in patients with congenital heart disease (Pehlivan \textit{et al.}, 1999).

Together, these show that mutations in GATA proteins lead to impaired biological processes. By understanding the molecular basis of these disorders, not only does that
allow us to begin to understand the normal roles that GATA proteins play in humans, but it may also aid in the development of treatments for these disorders.

### 6.2 A novel ZnF-mediated interaction between FOG and TACC3

While FOG was first identified as a protein partner of GATA-1, it has emerged that FOG also has some GATA-independent roles. One of these roles involves the centrosomal protein TACC3. While the biological role of the FOG:TACC3 interaction is as yet unclear, early signs suggest that the cytoplasmic TACC3 may be able to influence the cellular localisation of FOG, and therefore indirectly affect the ability of FOG to perform as a transcription factor in the nucleus (Garriga-Canut and Orkin, 2004).

The FOG and TACC3 interaction was characterised at a molecular level in this Thesis (Chapter 4). FOG utilises its third ZnF to interact with TACC3. The solution structure of this domain was determined using NMR spectroscopy: the domain is a true classical ZnF with a traditional left-handed \( \beta\beta\alpha \) fold. The region of TACC3 with which FOG finger 3 interacts appears to be a dimeric coiled-coil domain. A short 9–10 residue segment of the TACC3 coiled-coil contacts the \( \alpha \)-helix of the classical zinc finger 3 of FOG in the interaction. Thus, the FOG:TACC3 interaction appears to be mediated by \( \alpha \)-helices.

While the DNA-binding roles of classical ZnF are well-characterised (including the molecular details), its protein-binding capability has only recently been realised. Therefore, the FOG:TACC3 interaction is of particular interest as the data described here constitute some of the first available information as to the manner in which a classical ZnF contact a protein partner. Further, it is of note to find that the classical FOG finger 3 utilises its \( \alpha \)-helix to bind the TACC3 protein, as it is also the \( \alpha \)-helix of DNA-binding classical ZnFs that is utilised to interact with DNA.

### 6.3 ‘Incomplete’ ZnF sequences may be functional

The presence of variant CCHC classical ZnFs in FOG and the identification of several CCHX sequences in the protein database led to the question of whether functional
zinc-binding domains can be formed with sequences containing only three potential zinc ligands.

Data presented here (Chapter 5) show that indeed the CCHX ligation topologies introduced into the sequence of BKLF finger 3 were able to bind Zn$^{2+}$ and formed discrete zinc-binding domains. Further, these domains formed with only three protein ligands were able to interact with both DNA and protein partners.

These findings have several implications. Most notably, incomplete ZnF sequences should not be automatically assumed to be non-functional zinc-binding motifs, and their potential to form such domains needs to be tested. This therefore suggests that the number of zinc-binding domains in nature may be even higher than realised.

### 6.4 Zinc fingers: the versatile protein domain

What has emerged from the work presented in this Thesis is the versatility of zinc finger domains. The ZnF domain is one of the most common protein domains in eukaryotes, and it is able to mediate various types of interactions. While the ZnF fold is able to accommodate mutations without having its structural integrity affected (see Chapter 5), the mutations may hinder ZnF functions by affecting interactions that the domain mediates (as observed in the case of GATA-1 N-finger).

The ZnF fold has served as a template for the creation of many designer proteins. In particularly, the classical ZnF fold has been utilised to produce many novel DNA-binding proteins (see Section 1.2.3). It may soon be possible, taking into account the GATA-FOG interaction (see Chapter 3) and now the FOG-TACC3 interaction (Chapter 4), to design novel protein-protein interactions using ZnFs. This idea holds many possibilities as interactions between proteins, as well as those between proteins and DNA, play crucial roles in the development and maintenance of all organisms. The potential for zinc finger domains is only just beginning to be realised.
References


Cram, E. D. (2004). Structural studies of BKLF and characterisation of its interaction with GATA-1. (University of Sydney, Sydney)


Goddard, T. D. and Kneller, D. G. SPARKY 3. (University of California, San Francisco)


Turner, J. I. (2001). The role of the DNA-binding protein Basic Krüppel-like factor in transcription. (University of Sydney, Sydney)


erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J.* **16**, 3145-3157.


