Protein-protein interactions mediated by Cys2His2 zinc-fingers

DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.) DER NATURWISSENSCHAFTLICHEN FAKULTÄT III – BIOLOGIE UND VORKLINISCHE MEDIZIN DER UNIVERSITÄT REGENSBURG

vorgelegt von
Astrid Giesecke
aus Pfaffenhofen
Mai 2006

Promotionsgesuch eingereicht am: 24.05.06

Die Arbeit wurde angeleitet von Prof. Dr. Keith Joung

Prüfungsausschuss:

Vorsitzender: Prof. Dr. Stephan Schneuwly

Erster Gutachter: Prof. Dr. Reinhard Sterner

Zweiter Gutachter: Prof. Dr. Keith Joung

Dritter Prüfer: Prof. Dr. Herbert Tschochner

Die Arbeit wurde betreut von

Prof. Dr. Keith Joung

Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of Tables

Chapter 1

Chapter 3

Chapter 5

Chapter 6

Chapter 7

Chapter 1. Introduction.

1.1 Protein-protein interaction domains

Virtually all biological processes depend on specific interactions between different molecules such as DNA, RNA, proteins and lipids. Within a cell, specific binding is mediated by molecular recognition between various biomolecules. Proteins are the organizers of almost all cellular processes and have various functions both in the nucleus and the cytoplasm ranging from transcriptional regulation, DNA replication, splicing, to protein localization and trafficking, metabolism, and protein degradation. Furthermore, protein interactions are involved in formation of complex structures such as the cytoskeleton (e.g. Lodish *et al.*, 1996). These functions generally involve interactions with different macromolecules since proteins rarely work in isolation and all proteins in a given cell are believed to be connected through an extensive protein-protein interaction network (e.g. Giot *et al.,* 2003; Rual *et al.,* 2005). For example, DNA-binding transcription factors regulate transcription by either activating or repressing defined target genes. However, regulation of gene expression is very complex and protein-protein interactions play significant roles in this process by for example recruiting co-repressors or co-activators to the promoter of the target gene (for comprehensive reviews see Ptashne and Gann, 1997; Lemon and Tjian, 2000; Brivanlou and Darnell, 2002; Pawson *et al.*, 2002; Pawson and Nash, 2003). Many classes of DNA binding proteins are involved in contacting other proteins and there are several examples of transcription factors that have to dimerize in order to bind to the DNA. (e.g. Leucin zipper families of transcription factors including Fos/Jun [Kouzarides and Ziff, 1988; Kouzarides and Ziff, 1989] activating transcription factor (ATF)/cAMP responsive element (CRE) binding proteins (CREB) [Hai and Curran, 1991]).

Molecular recognition by different proteins is mediated by interacting motifs which consist of a small number of residues that fold into an autonomous domain. Many proteins contain more than one of such interaction domain (Reichmann *et al.*, 2005). This cassette-like modular behavior of these domains suggests an ideal mechanism to build complexes consisting of many different components. Although nature provides a highly complex and functionally diverse repertoire of protein interactions, the number of domain folds seems to be relative small. This suggests that many proteins must fold into similar structures and the large number

of specific interactions develop from displaying a diverse set of defined sequences upon these stable folds (reviewed in Koonin *et al.,* 2002). To date, structural information has been obtained for several of these protein-binding modules revealing how diverse these domain folds are. For some of the structural folds generalizations regarding the target recognition have been made. An example is the Src-homology-2 (SH2) domain which recognizes and binds short tyrosine-phosphorylated sequences, and thereby has a key role in tyrosine kinase signaling cascades (Schlessinger and Lemmon, 2003; reviewed in Pawson, 2004). Other protein-binding folds such as the ankyrin repeat (reviewed in Sedgwick and Smerdon, 1999; Mosavi *et al.*, 2004) or the leucine-rich repeat (reviewed in Buchanan and Gay, 1996 and in Kobe and Kajava, 2001) interact with diverse protein partners that do not exhibit common features and a universal recognition code has not been proposed yet for these interaction modules. Thus, analyzing protein-protein interaction domains to identify general binding and recognition modes is still a challenging task in modern biology.

1.2 Cys2His2 zinc finger proteins

1.2.1 General

The classical Cys2His2 zinc finger domain (C2H2 ZF) is the most prevalent protein motif in human cells (Venter *et al.*, 2001; Lander *et al.,* 2001) and represents the most common DNA binding domain found in eukaryotic transcription factors (Pellegrino and Berg, 1991; Jacobs, 1992; Venter *et al.,* 2001; Lander *et al.,* 2001). It is estimated that there are ~900 C2H2 ZFs encoded in the human genome constituting 2-3% of all genes (Tupler *et al.*, 2001; Venter *et al.*, 2001; Lander *et al.,* 2001; Huntley *et al.*, 2006). C2H2 ZFs also represent the largest transcription factor class in *Drosophila melanogaster*, *Caenorhabditis elegans* and *Saccharomyces cerevisiae* (Tupler *et al.*, 2001) but seem to be absent in prokaryotes (reviewed in Wolfe *et al.,* 2000). The abundance of these domains reflects their versatility evidenced by their abilities to recognize a wide variety of specific DNA sequences. Although originally discovered as DNA binding domains, C2H2 ZF have also been implicated in protein-contacts with various protein classes. This may provide an explaination for why some proteins have C2H2 ZFs that are evidently not involved in DNA-binding (reviewed in

Mackay and Crossley, 1998). Furthermore, the C2H2 ZF motif has also been shown to mediate protein-RNA intercations (reviewed in Hall, 2005; reviewed in Brown, 2005). Most C2H2 ZF proteins contain more then one C2H2 ZF frequently arranged in tandem arrays with some proteins having over 30 C2H2 ZFs (reviewed in Iuchi, 2001).

1.2.2 Discovery of the C2H2 ZF motif

The term C2H2 ZF was initially used to define a 30 residue repeated sequence motif that was discovered more then 20 years ago in the *Xenopus laevis* transcription factor TFIIIA (Miller *et al.*, 1985). During efforts to isolate and purify 7S particles of *Xenopus laevis* oocytes containing 5S RNA and TFIIIA, Miller and co-workers observed significant concentrations of zinc present in these particles. Furthermore, proteolytic digest of the complex of 5S RNA and TFIIIA consistently yielded several persistent small peptide fragments of size ~3 K Daltons. This suggested that TFIIIA contained several small stable protein domains. Subsequently, amino acid sequence analysis demonstrated that TFIIIA harbored nine sequence repeats of 30 residues containing two invariant cysteine and two invariant histidine residues, which were known to be the most common zinc ligands. These findings suggested that TFIIIA was composed of a linear arrangement of small conserved structural motifs that can bind zinc via their cysteines and histidines residues and this characteristic motif was termed as the C2H2 ZF motif (Miller *et al.*, 1985).

1.2.3 Structural properties

The C2H2 ZF motif is defined by a conserved amino acid sequence of the form (F/Y)-X-C-X_{2-5}-C-X_3-(F/Y)-X_5-Ψ-X_2-H-X_{3-5}-H, where X represents any amino acid residue and Ψ encodes a hydrophobic residue (note that the spacing between the two cysteines and two histidines is flexible). Structural studies using nuclear magnetic resonance (NMR) defined the C2H2 ZF motif as a β-hairpin followed by an α-helix (ββα-fold) which folds into an autonomous protein domain (Lee *et al.,* 1989). Each finger binds a single zinc ion by coordinating it tetrahedrally between the two cysteines at one end of the β-sheet termed as the "zinc knuckle" (turn containing the two cysteines; Grishin, 2001; Krishna *et al.*, 2003) and the two histidines at the carboxy-terminus of the α-helix (Figure 1.1). The overall structure of

the protein is stabilized by zinc binding as well as by the small well conserved hydrophobic core adjacent to the zinc binding side (Miller *et al.*, 1985; Lee *et al.,* 1989; reviewed in Wolfe *et al.*, 2000; Pabo *et al.*, 2001; Iuchi, 2001). C2H2 ZFs are completely unfolded in the absence of zinc but can also fold in the presence of cobalt *in vitro* (Frankel *et al.,* 1987). Cobalt is frequently substituted for zinc to study metalloproteins since its optical absorption spectrum is extremely sensitive to the coordination state, thus allowing one to determine the stoichiometry of a complex (Frankel *et al.,* 1987). Additional studies have shown that the zinc-coordinating C2H2 motif can be partially substituted by other zinc-ligating motifs (such as the Cys4 motif) (Krizek *et al.,* 1993; Green and Sarkar, 1998). Furthermore, Michael and co-worker (1992) have synthesized a "minimalist" C2H2 zinc finger peptide containing an alanine substitution at every residue position except at positions defining the C2H2 ZF motif (i.e. all Xs in the amino acid sequence shown above are replaced by an alanine). This peptide can still form complexes with cobalt *in vitro* (see above), indicating that the defined conserved residues of the C2H2 ZF motif are necessary but also sufficient for proper folding (Michael *et al.,* 1992).

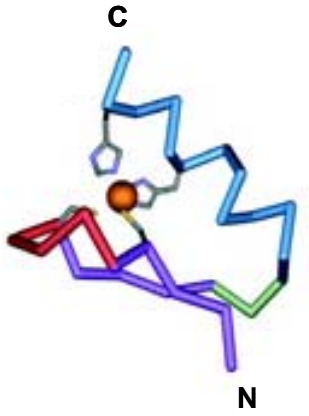

Figure 1.1 The C2H2 ZF motif. C2H2 ZFs are characterized by a β-hairpin that contains the two zinc-ligating cysteines followed by an α-helix that provides the histidine ligand pair. The α-helix is colored in blue and the β-hairpin is colored in purple. The "zinc knuckle" connecting the two β-sheets is shown in red. Green indicates a loop. Zinc is represented as a orange sphere and residues that bind zinc are shown as ball-and-stick. This Figure was reproduced from Krishna *et al.* (2003).

1.2.4 Interaction of C2H2 ZFs with DNA

C2H2 ZFs are well known for their ability to bind DNA and the molecular details of this interaction have been studied extensively. In general, three or more tandemly arranged C2H2 ZFs are required to recognize DNA motifs that are present in the promoter region of a wide variety of defined target genes. An exception is the GAGA factor from *Drosophila melanogaster*, which contains a single C2H2 ZF. GAGA can bind to its specific binding site

via its single C2H2 ZF and an amino-terminal extension containing a stretch of basic amino acids, but the C2H2 ZF alone is not sufficient for DNA-binding (Pedone *et al.,* 1996; Omichinski *et al.,* 1997). The interaction between C2H2 ZFs and DNA is usually very specific and occurs with high affinity. The crystal structure of Zif268 bound to DNA was solved 15 years ago and provided the first detailed insights into C2H2 ZF-DNA recognition (Pavletich and Pabo, 1991). The Zif268 DNA-binding domain (DBD) contains three C2H2 ZFs and binds to a 10 base pair (bp) target sequence. Each individual finger interacts with a sub-site of 3-4 bp by using the amino-terminal end of its α-helix (or recognition helix) to bind the major groove of the DNA (Figure 1.2). The orientation of the protein causes the amino-terminal finger to contact the 3' end while the carboxy-terminal finger binds to the 5' end of one strand of the DNA (termed the primary strand). Interestingly, each finger in the Zif268 structure binds to its sub-site in a uniform pattern of residue-base interactions: each finger utilizes amino acids at four key positions (-1, 2, 3 and 6) on the surface of the recognition helix to contact bases in the DNA (Figure 1.2).

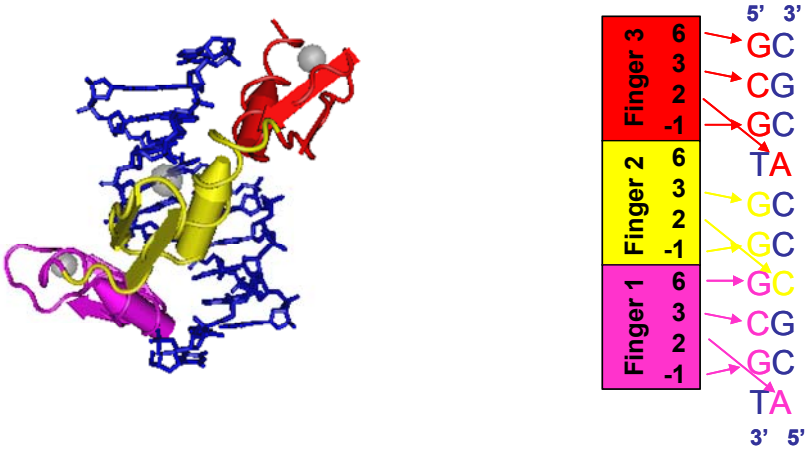

Figure 1.2 DNA-binding by C2H2 ZFs. Structure of Zif268 bound to its specific DNA binding site. Fingers 1, 2 and 3 of Zif268 are purple, yellow and red, respectively. DNA is colored in blue and zinc ions are represented as grey spheres. ZFs insert into the major groove by making specific base contacts with the DNA. These contacts are made from residue positions -1, 2, 3 and 6 of each α-helix as indicated in the scheme right of the structure. This Figure was adapted from Wolfe *et al.* (2000) and Pabo *et al.* (2001).

Additional binding affinity is provided by interactions with the phosphate backbone of the DNA as well as between neighboring C2H2 ZFs (Pavletich and Pabo, 1991; Elrod-Erickson *et al.*, 1996; reviewed in Wolfe *et al.*, 2000 and in Pabo *et al.*, 2001). The linker region that connects adjacent fingers plays a role in spacing and orientation of the individual fingers on

the DNA. Although flexible in the absence of DNA, it becomes well ordered upon DNA-binding (Foster *et al.,* 1997; Wuttke *et al.,* 1997). The most common linker sequence is TGEKP (Pellegrino and Berg, 1991) and mutational analysis has demonstrated that the linker is important for high affinity DNA binding (Choo and Klug, 1993). Although the structure of the Zif268-DNA complex still serves as the prototype for understanding DNA recognition, structural studies of other naturally occurring C2H2 ZFs bound to their DNA subsite have revealed important variations on this common pattern, implying that C2H2 ZF mediated DNA recognition is more complex than initially thought (reviewed in Klug and Schwabe, 1995; Wolfe *et al.*, 2000 and in Pabo *et al.*, 2001).

1.2.5 C2H2 ZF engineering

The simple modular behavior of certain C2H2 ZFs together with their versatility in recognizing a large number of specific DNA sites was soon availed for protein engineering purposes. Many studies from different labs have demonstrated that DNA binding specificity can be altered by simply changing key residues in the recognition helix of fingers from Zif268 and SpI. Combinations of rational design and selection methods have been successfully applied to create individual C2H2 ZF with novel DNA specificities (Desjarlais and Berg, 1992; Desjarlais and Berg, 1993; Choo and Klug, 1994; Jamieson *et al.*, 1994; Jamieson *et al.*, 1996; Wu *et al.*, 1995). To further create proteins able to bind to DNA with high specificity, these individual C2H2 ZFs peptides can be assembled into a protein that recognizes an entirely novel DNA site consisting of individual subsites. In an initial study, Choo and co-workers (1994) used three individually re-designed C2H2 ZFs and linked them together to create a three-finger DNA-binding protein that binds to a novel 9-bp site. This protein was able to specifically regulate expression of a reporter gene although the affinity for DNA-binding was low (Choo *et al.,* 1994). To account for context-dependent interactions between neighboring C2H2 ZFs and subsites that are evidently important for DNA-binding (Desjarlais and Berg, 1993), Greisman and Pabo (1997) developed a selection strategy where the desired three-finger protein is gradually assembled. In this approach, two wild-type C2H2 ZFs represent an “anchor” while the randomized third C2H2 ZF is used to perform selections for a target subsite. In the next steps the pool of selected reengineered C2H2 ZFs obtained in the first step is retained, which allows the selection of a new second (and subsequently third) C2H2 ZF that binds DNA in the context of a protein containing the first (and subsequently

second) selected finger. This strategy yield optimized C2H2 ZF proteins that display high levels of specificity and affinity (Greisman and Pabo 1997). This approach together with other optimized selection methods has subsequently been utilized many times to create several three-finger proteins with novel DNA-binding specificities (Isalan and Choo, 2000; Hurt *et al.*, 2003).

However, to increase the specificity and affinity of engineered C2H2 ZFs for their target site, one has to extend the interaction surface. This is especially desirable considering the fact that a DNA sequences has to be at least 16 bp long in order to represent a unique binding site in the context of the whole genome. (Calculations expect that there should be ~ 10000 identical 9 base pair sites present in the human genome while a sequence of 16-18 base pairs should occur only once [reviewed in Wolfe *et al.*, 2000 and in Choo and Isalan, 2000]). Thus, arrays of tandemly repeated C2H2 ZFs have also been constructed by simply linking three-finger domains together into six-finger or nine-finger proteins using linkers of the sequence TGEKP (Liu *et al.,* 1997; Kamiuchi *et al.,* 1998). Surprisingly, although these proteins are capable of recognizing their extended target site, the affinity enhancements for DNA-binding were only modest. In contrast, when a longer linker was used (consisting of eight residues) to connect these 3-finger domains the relative affinity could be dramatically increased (Kim and Pabo, 1998). Moore and co-workers (2000) used a different approach by linking three two-finger units together into a six-finger protein which displayed increased levels of specificity and affinity. In a parallel study, they also designed optimized “structural” linkers capable of connecting two three-finger peptides into a six-finger protein with significant affinity and specificity enhancements (Moore *et al.,* 2001).

It is noteworthy that in addition to these reengineered DNA-binding ZFs, naturally occurring single C2H2 ZFs can also be used to create DBDs with novel specificities. For example, Bae and co-workers (2003) screened C2H2 ZFs encoded in the human genome for their ability to bind to diverse DNA sites. These C2H2 ZFs were then used as modular “building blocks” to construct novel specific DBDs (Bae *et al.*, 2003). In summary, re-engineered single fingers together with naturally occurring fingers can generally be utilized to create multi-finger proteins with novel DNA-binding specificities by linking them together. Several of these polyfinger proteins have proven to bind to their targeted sequences with high affinity. Connecting these reengineered multi finger proteins to functional domains resulted in the design of a wide variety of synthetic transcription factors capable of regulating specific endogenous genes both in tissue culture and in whole organisms (Klug, 1999; Blancafort *et*

al., 2004; Blancafort *et al.*, 2005; for comprehensive reviewes see: Falke and Juliano, 2003; Jamieson *et al.*, 2003; Lee *et al.*, 2003; Jantz *et al.*, 2004).

1.2.6 Interaction of C2H2 ZFs with RNA

Although best known for their ability to bind to DNA, C2H2 ZFs were originally identified in TFIIIA, a protein that associates with 5S rRNA within the 7S particle in *Xenopus* oocytes (see section 1.2.2). TFIIIA binds specifically to the internal control region (ICR) of the 5S RNA gene, but can also directly interact with 5S RNA (Pelham and Brown, 1980; see also reviews by Shastry, 1996; Brown, 2005; and Hall, 2005). TFIIIA contains nine C2H2 ZFs and it has been demonstrated, that DNA binding is mainly accomplished by C2H2 ZF 1-3, while C2H2 ZF 4-6 are essential for RNA binding although these fingers can also bind to DNA (Hansen *et al.,* 1993; Nolte *et al.,* 1998; Neely *et al.,* 1999; Searles *et al.,* 2000; Lu *et al.,* 2003). The crystal structure of C2H2 ZFs 4-6 of TFIIIA bound to a fragment of the 5S rRNA was only recently solved and provided striking insights into how C2H2 ZFs mediate interactions with RNA (Lu *et al.,* 2003). The structural arrangement of RNA is generally more complex than that of DNA and includes the formations of internal loops and helices. The secondary structure in 5S RNA that is contacted by TFIIIA consists of loop E (bound by ZF4), helix V (bound by ZF5) and loop A (bound by ZF6) (Lu *et al.,* 2003). As expected, only a few contacts to nucleotides in the RNA are made (as in the case of DNA-binding C2H2 ZFs) and α-helices in the C2H2 ZFs are responsible for these contacts. Binding of ZF4 involves residue positions -2, -1, +1 and +2 of the α-helix, while ZF6 contacts RNA via residues at positions -1, +1 and +2 of the α-helix (Lu *et al.,* 2003; compare to Figure 1.2). Thus, both C2H2 ZFs require residues at position -1 and +2 which have also been shown to be important for DNA-binding (see section 1.2.4). Surprisingly, ZF5 does not directly contact nucleotides in the RNA but binds to its phosphate backbone via multiple contacts made by basic amino acid residues (Lu *et al.,* 2003). Thus, these studies revealed both similarities as well as differences in RNA-recognition mediated by C2H2 ZFs compared to DNA-binding.
Other examples of C2H2 ZF proteins capable of binding to both DNA and RNA have also been described. One example is the Wilms tumor 1 (WT1) protein, which contains four C2H2 ZFs and binds to DNA as well as to specific single-stranded RNAs using distinct C2H2 ZFs for different RNAs (Caricasole *et al.,* 1996; Bardeesy and Pelletier, 1998). Interestingly, alternative splicing of exon 9 inserts (or removes) three amino acids, lysine, threonine and

serine (commonly referred to KTS), between the third and the fourth C2H2 ZF. Isoforms containing this insertion are impaired in their ability to bind to DNA but can still interact with RNA (reviewed in Lee and Habor, 2001). In addition, C2H2 ZFs that bind primarily to RNA have also been found. Examples are the Double-stranded RNA-binding proteins zinc finger a (dsRBP-Zfa) which contains seven C2H2 ZFs and the Just another zinc finger (JAZ) protein which is composed of four C2H2 ZFs (Finerty and Bass, 1997; Yang *et al.,* 1999).
In summary, several C2H2 ZFs capable of binding to RNA have been described. The ability of C2H2 ZFs to bind to RNA provides another example of their versatility, especially considering the fact that RNAs provide a wide spectrum of secondary and tertiary structures including double-stranded RNA, single-stranded RNA and DNA-RNA duplexes (reviewed in Iuchi, 2001).

1.2.7 Interaction of C2H2 ZFs with other proteins

1.2.7.1 Examples

C2H2 ZF can also mediate protein-protein interactions with a wide variety of protein classes (reviewed in Mackay and Crossley, 1998). Interestingly, unlike DNA-binding C2H2 ZFs that recognize the defined structural motif of the double-stranded DNA, protein binding C2H2 ZFs can bind to a wide range of different structures including other C2H2 ZFs, other classes of ZFs as well as completely different protein motifs (reviewed in Matthews and Sunde, 2002). Increasing numbers of reports describe protein-protein interactions mediated by C2H2 zinc fingers and some of these are shown in Table 1.1. For example, the Repressor element-1 (RE-1) silencing transcription factor/neuronal restricted silencing factor (REST/NRSF, hereafter REST) can specifically bind to the co-repressor protein CoREST by utilizing one C2H2 finger (Andres *et al.*, 1999; see also chapter 7). Additional studies describe the importance of the C2H2 zinc finger domain for various interactions although it is not clear if the C2H2 ZFs always directly participate in the interaction. Examples are the WT1 protein that interacts with different classes of proteins including p53, p73, p63, CREB binding protein (CBP)/p300, and the Sex-determining region of the Y chromosome protein (SRY) (reviewed in Lee and Haber, 2001; reviewed in Scharnhorst *et al.*, 2001; Wang *et al.*, 2001; Matsuzawa-Watanabe *et al.*, 2003) as well as the Ying Yang 1 protein (YY1) which binds to several cellular factors including TATA binding protein (TBP), CBP/p300, TFIIB, E1A, c-

Myc, SpI and ATF/CREB (reviewed in Thomas and Seto, 1999). It has been demonstrated for both WT1 and YY1 proteins that the zinc finger domain is at least necessary for the respective interactions and in some cases a physical interaction directly involving the ZFs has been shown (e.g. Lee *et al.*, 1993; Zhou *et al.*, 1995; Matsuzawa-Watanabe *et al.*, 2003).

C2H2 ZF containing protein	Interaction partner	Reference
Ikaros	Ikaros	Sun *et al.*, 1996
REST	Co-REST	Andres *et al.*, 1999
YY1	SpI, ATF a2	Lee *et al.*, 1993; Zhou *et al.*, 1995
Roaz	Roaz	Tsai and Reed, 1998
Sryδ	sryδ	Payre *et al.*, 1997
BMZF2	WT1	Lee *et al.*, 2002
WT1	SRY	Matsuzawa-Watanabe *et al.*, 2003
ZNF74	RNA Polymerase II	Grondin *et al.*, 1997

Table 1.1 C2H2 ZFs involved in protein-protein interactions. Protein pairs are shown in the first two columns. For theses pairs a direct involvement of the ZF domain in the interaction has been demonstrated. REST, Repressor element-1 silencing transcription factor; YY1, Ying Yang 1; Roaz, Rat olfactory 1 / early B-cell factor –associated zinc finger protein; Sryδ, Serendipity δ; BMZF2, Bone marrow zinc finger 2; WT1, Wilms tumor 1; ZNF74, Zinc finger 74; Co-REST, Co-repressor of REST; ATF, Activating transcription factor; SRY, Sex-determining region of the Y chromosome.

1.2.7.2 Structures of protein-binding C2H2 ZFs

Unfortunately, only a handful of structures of C2H2 ZFs involved in protein binding have been described and none of these structures shows an actual C2H2 ZF mediated protein-protein interaction. Examples include the transactivation domain of ATF-2 containing a single C2H2 ZF (Nagadoi *et al.* 1999), the substrate-binding domain of Seven in absentia homolog 1a (Siah1a) containing one C2H2 ZF (Polekhina *et al.*, 2002) and a single C2H2 finger of the dimerization domain from the transcription factor Eos (Westman *et al.*, 2004). These structures have provided some insights about protein-contacting C2H2 ZFs including similarities as well as remarkable differences to DNA-binding C2H2 ZFs.

ATF-2:

Using NMR spectroscopy Nagadoi and co-workers (1999) have solved the solution structure of the transactivation domain of ATF-2 which contains a domain termed N-subdomain that shows high sequence homology to the C2H2 ZF motif. In fact, the structure of this domain is highly similar to the typical ββα motif of a C2H2 ZF. A comparison of the N-subdomain with two DNA-binding C2H2 ZFs, ZF1 of Zif268 (Elrond-Erickson *et al.,* 1996) and ZF3 of the

human glioblastoma protein (GLI) (Pavletich and Pabo, 1993) indicates that the backbone structure of the N-subdomain matches well with the two DNA-binding C2H2 ZFs. In particular, the arrangement of the hydrophobic core is almost identical to the DNA-binding C2H2 ZFs (Nagadoi *et al.* 1999). Interestingly, sequence comparison between the N-subdomain and various DNA-binding C2H2 ZFs demonstrated that residue positions known to bind to the phosphate backbone of DNA are only conserved in DNA-binding C2H2 ZFs while residue positions that are important for maintaining the typical structure of the C2H2 ZF are conserved in both the DNA-binding C2H2 ZFs and the N-subdomain (Nagadoi *et al.* 1999). Another notable feature that Nagadoi and co-workers (1999) described is the difference in surface charge distributions between protein-contacting C2H2 ZFs and DNA-binding C2H2 ZFs. While the surface of DNA-binding C2H2 ZFs (from GLI and Zif268) is highly positive, protein-contacting C2H2 ZFs are either neutral (in the case of N-subdomain) or negative (in the case of the protein-interacting ZF1 of GLI). This suggests that the charges present on the surface of a C2H2 ZF protein can be a determinant of whether the protein binds to DNA or other proteins (Nagadoi *et al.* 1999).

Eos:

The C2H2 ZFs of the transcription factor Eos that mediate dimerization have been investigated using a combination of different techniques including circular dichroism (CD), UV-Vis spectrophotometry and NMR spectroscopy (Westman *et al.,* 2004). Eos contains an amino-terminal (N-terminal) domain consisting of four C2H2 ZFs and a carboxy-terminal (C-terminal) domain which is composed of two C2H2 ZFs. While the N-terminal domain binds to DNA, the C-terminal domain is implicated in mediating protein contacts (Perdomo *et al.*, 2000; Westman *et al.,* 2003; Westman *et al.,* 2004; see also section 1.3.2). Initial UV-Vis experiments demonstrated that both C-terminal C2H2 ZFs from Eos can fold in the presence of zinc and are likely to take on the typical ββα structure. The solution structure of the second C-terminal C2H2 ZF was subsequently solved by NMR spectroscopy (Westman *et al.,* 2004). Surprisingly, two distinct sets of conformers (termed EosC2' and EosC2") were obtained that differ in the arrangement of the polypeptide backbone at the C-terminus (Figure 1.3). Both conformers consist of a loose β-hairpin-like fold that positions the two cysteines followed by a short (4-7 residues, depending on the conformer) but well-ordered α-helix, which contains the two histidines involved in zinc-binding. However, the positions of the two histidine side chains are reversed when comparing the two conformers (Figure 1.3). This suggested that the

structure of this C2H2 ZF displays some conformational flexibility (Westman *et al.,* 2004) which is in contrast to the well accepted view that C2H2 ZFs structures are highly ordered. The structure of the two conformers was further compared to the second C2H2 ZF of the DNA-binding protein MBP-1 (Omichinski *et al.,* 1992), indicating that the overall structural arrangement is similar although the positions of the two zinc-ligating histidines in EosC2" is swapped in comparison to the corresponding position in MBP-1 (Figure 1.3, Westman *et al.,* 2004). It is noteworthy, that the overall arrangement of the hydrophobic core in both conformers is conserved, which is somewhat surprising given that the generally invariant phenylalanine that is usually present after the second zinc-ligating cysteine in substituted by a serine (Westman *et al.,* 2004). The surface charge distributions of both conformers were further analyzed, indicating that no patches of positive charge are present which is consistent with the involvement of this C2H2 ZF in mediating protein-contacts rather the DNA-binding (Westman *et al.,* 2004; Nagadoi *et al.* 1999).

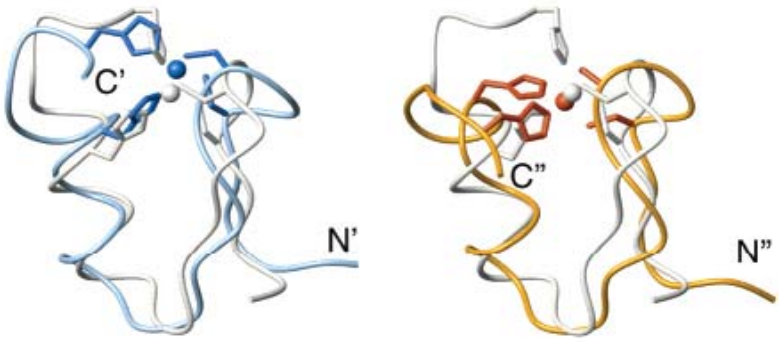

Figure 1.3 Structural comparison of the ZF motif in Eos with the ZF motif of a DNA binding ZF. The structures of the two conformers EosC2' (blue) and EosC2" (yellow) of the second C-terminal C2H2 ZF are shown overlaid with ZF2 of MBP-1 (white). The side chains of the zinc-ligating residues are shown as ball-and-stick in dark blue (for EosC2'), red (for EosC2") and white (for MBP-1). The corresponding zinc ion is shown as a colored sphere. N- and C-termini are also indicated. This Figure was taken from Westman *et al.* (2004).

In summary, these structures have provided initial insights into characteristic features of a protein-binding C2H2 ZF. In particular, they demonstrated that the fold of these C2H2 ZFs largely conforms to the typical ββα structure of a DNA-binding C2H2 ZFs although differences definitely exist. On the other hand, inspection of the charged residue distribution on the surface of these C2H2 ZFs clearly indicate that protein-binding C2H2 ZFs lack the characteristic regions of positive charges present on the surface of DNA-binding C2H2 ZFs. Furthermore, the charge distribution seems to be more flexible (both negative and neutral

charges are present) which could point towards a high level of complexity required for protein-interactions as opposed to binding to the regular structure of the negatively charged DNA.

1.2.7.3 Structures of protein-protein interactions mediated by C2H2 ZFs

Only limited structural information of protein-protein interactions mediated by C2H2 ZFs has been obtained. For example, the crystal structure of a complex containing the five C2H2 ZFs from GLI bound to their DNA site has revealed that ZF1 of GLI is not involved in contacting DNA (Pavletich and Pabo, 1993). Instead, ZF1 is packed against ZF2 of GLI and makes extensive protein-protein interactions with this finger, although the biological relevance of this intra-molecular interaction is unknown. This interaction involves hydrophobic contacts made by several residues at the interfaces of both fingers. Furthermore, the linker between ZF1 and ZF2 in GLI is two residues longer then the typical five residue linker and seems to provide more flexibility (Pavletich and Pabo, 1993). It is noteworthy that these studies are consistent with the findings of Nagadoi and co-workers (1999), which describe the surface of GLI ZF1 as highly negatively charged (see section 1.2.7.2) and therefore presumably involved in protein-protein interactions.

Structural information has also been obtained for an artificial peptide extension (consisting of 15 residues) that was linked to the amino-terminus of ZF1 and ZF2 of Zif268 (Wang *et al.,* 2001). This peptide mediates dimerization of the two C2H2 ZFs, thus permitting them to bind to a palindromic DNA-site. The crystal structure of this complex revealed that the peptide reaches across the DNA and extensively contacts a hydrophobic patch of residues present on the surface of the other C2H2 ZF (i.e. peptide extension of one monomer contacts the ZFs in the other monomer and vice versa). Closer inspection of these hydrophobic residue positions found in Zif268 ZF1 demonstrate that they match perfectly well to the patch of residues in GLI ZF1 and ZF2 identified by Pavletich and Pabo (1993) (Wang *et al.,* 2001). Thus, this exposed hydrophobic surface may be generally important in C2H2 ZFs for contacting other proteins (Wang *et al.,* 2001).

In summary, theses structures have provided some insights in the mechanism of protein binding mediated by C2H2 zinc fingers, although the molecular details of such interactions which are well characterized for DNA recognition by C2H2 ZFs (see section 1.2.4) are still missing. In addition, these studies demonstrate that C2H2 ZFs can interact with DNA and

proteins simultaneously (Wang *et al.,* 2001), which provides further indication that distinct protein surfaces are used for these different kinds of interactions.

1.2.8 Other ZF motifs

After the discovery and characterization of the classical C2H2 ZF motif, several other classes of zinc-ligating domains have been described and the term ZF is now most commonly defined as a protein motif that folds independently around one or more zinc cations. Different classes of ZF proteins vary in the nature of their zinc-binding residues, but they all have in common that they bind zinc ions in a purely structural manner rather then using it for catalytic processes (reviewed in Matthews and Sunde, 2002). Although originally classified according to the identity and geometry of the zinc binding ligands (reviewed in Mackay and Crossley, 1998), the growing number of structural reports on these proteins suggested that classification should be based on structural properties. Using this method, classes of ZF proteins are assorted into eight different fold groups (Krishna *et al.*, 2003). However, the majority of ZFs belong to two protein folds: the classical C2H2 like finger and the treble clef finger (Grishin, 2001; Krishna *et al.*, 2003). Although ZFs are involved in various different functions within the cell, they mediate contacts with other molecules such as DNA, RNA, proteins, and lipids. Most ZF proteins contain more than one ZF which are frequently arranged in tandem arrays suggesting that ZF proteins serve as a platform for assembling a range of different biomolecules (reviewed by Klug and Schwabe, 1995; and by Matthews and Sunde, 2002).

1.3 The Ikaros transcription factor family

1.3.1 The Ikaros protein

1.3.1.1 Discovery

Hematopoiesis is the process of producing a functional distinct set of cells that comprise the mature blood. These cells arise from pluripotent hematopoietic stem cells that successively become more specified by regulated division and differentiation steps. This process is tightly controlled and selective changes in gene expression patterns have shown to be important for

the transition of hematopoietic precursors to a highly differentiated cell (reviewed in Orkin, 1995; and in Cantor and Orkin, 2001). The regulation of this coordinated program of gene activation and silencing is mediated by transcription factors and the search for such factors started more then 15 years ago. In an attempt to isolate regulatory proteins involved in the control of differentiation of the hematopoietic T cell lineage, Georgopoulos and co-workers (1992) identified the Ikaros protein. The T cell lineage is characterized by the presence of the CD3-T cell receptor complex which is encoded by the CD3δ gene (Furley *et al.*, 1986; Haynes *et al.*, 1989). Ikaros was isolated as a factor that specifically binds to a G-rich sequence present in the regulatory elements of CD3δ (Georgopoulos *et al.*, 1992). This factor was also shown to encode the Lymphoid transcription factor 1 (LyF-1) protein, which binds to functionally important regulatory sites within the lymphocyte specific terminal deoxynucleotidyltransferase (TdT) promoter (Lo *et al.,* 1991; Ernst *et al.,* 1993; Hahm *et al.,* 1994). To further establish its function as a factor involved in the development of the hematopoietic system, the expression pattern of the Ikaros protein was analyzed. Ikaros is first detected in hematopoietic precursor populations and is later mainly present in mature T cells, B-cells and natural killer cells. In contrast, it is downregulated in most differentiated erythroid and myeloid lineages including mature monocytes, macrophages and erythrocytes (Georgopoulos *et al.*, 1992; Klug *et al.*, 1998). Thus, the temporal expression pattern of Ikaros in hematopoietic cell lines is consistent with the idea that it plays a role in lymphoid cell development Georgopoulos *et al.*, 1994; Klug *et al.*, 1998; reviewed by Georgopoulos *et al.*, 1997; Westman *et al.*, 2002; and by Cobb and Smale, 2005).

1.3.1.2 Ikaros isoforms

Ikaros is composed of seven exons from which at least eight isoforms (Ik-1 to Ik-8, Figure 1.4) can be generated by alternative mRNA splicing events (Hahm *et al.,* 1994; Molnar and Georgopoulos, 1994; Molnar *et al.*, 1996). As shown in Figure 1.4, each isoform encodes a distinct C2H2 ZF protein and most of these isoforms consist of two defined C2H2 ZF domains. While the C-terminal C2H2 ZF domain is present in all isoforms, the N-terminal C2H2 ZF domain contains different combinations of one to four C2H2 ZFs and two isoforms (Ik-6 and Ik-8) have no N-terminal C2H2 ZFs at all. It has been demonstrated that the N-terminal C2H2 ZFs are required for DNA binding. In addition, DNA specificity of the different isoforms was analyzed extensively using gel-shift assays and PCR site selections (Molnar and Georgopoulos, 1994) which showed that three N-terminal C2H2 ZFs are

necessary to bind to a single binding site of the conserved core motif GGGAA. Thus, only three of the Ikaros proteins (Ik-1, Ik-2 and Ik-3) which contain three or four N-terminal C2H2 ZFs, bind to this sequence. Ik-4 is composed of two N-terminal C2H2 ZFs and binds to a tandem recognition site containing an inverted repeat of the consensus motif. On the other hand, Ikaros isoforms with only one or no N-terminal C2H2 ZFs (Ik-5, Ik-6, Ik-7 and Ik-8) are not able to mediate interactions with DNA (Molnar and Georgopoulos, 1994).

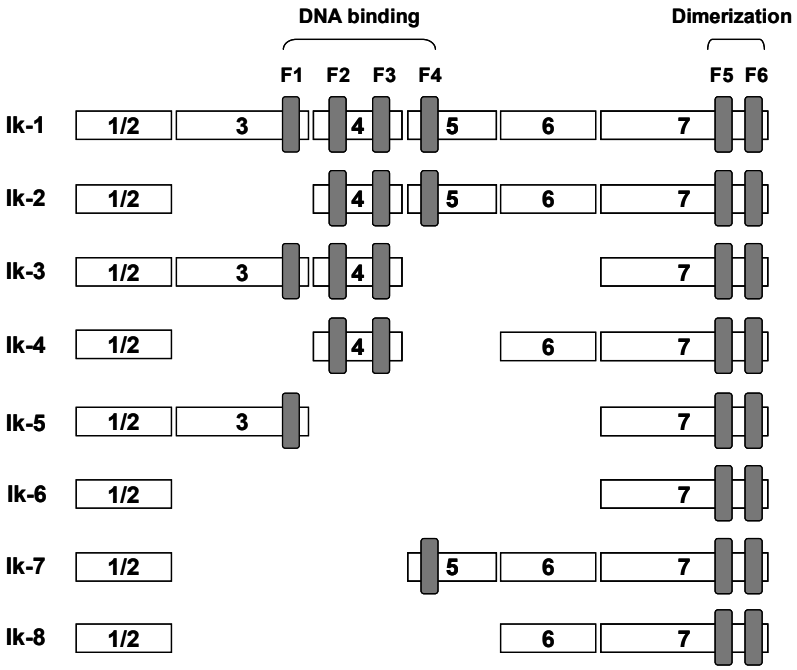

Figure 1.4 Schematic diagram of the Ikaros isoforms (Ik-1 - Ik-8). Exons are presented as white numbered boxes and individual ZFs are depicted as grey rounded rectangles. ZF domains involved in DNA-binding and dimerization are also indicated.

The C-terminal C2H2 ZF domain present in all Ikaros isoforms has been shown to mediate dimerization (Sun *et al.*, 1996; McCarty *et al.*, 2003) and is utilized to engage these various Ikaros proteins in homo- and heterodimeric complexes. Oligomerization between isoforms composed of N-terminal C2H2 ZFs capable of DNA-binding dramatically increases their DNA affinity. However, heterodimers consisting of Ikaros proteins with and without a functional DNA-binding domain can not bind to DNA. Furthermore, the formation of such complexes interferes with the transcriptional activity of DNA-binding isoforms which could provide a mechanism to control activity of these isoforms (Sun *et al.,* 1996).

1.3.1.3 Knock out studies of Ikaros

To evaluate the role of Ikaros in lymphoid cell development, knock out studies in mice were performed. Mice homozygous for a deletion of the N-terminal DNA-binding domain completely lack lymphoid progenitors as well as mature B and T lymphocytes and natural killer cells (Georgopoulos *et al.*, 1994). On the other hand, mice that are heterozygous for this mutation generate abnormal T cells and develop T-cell leukemias and lymphomas (Winandy *et al.*, 1995). This suggests that Ikaros is necessary for early development of lymphoid progenitors but also plays a role in later T cell maturation. Interestingly, mice homozygous for a deletion of the C-terminal C2H2 ZF domain in Ikaros display a phenotype less severe than the phenotype caused by the missing N-terminal domain (Wang *et al.,* 1996). This deletion is considered to be a null mutation resulting in a complete loss of Ikaros activity while the deletion of the N-terminal DNA binding domain is believed to fulfill a dominant-negative function. Thus, the severe phenotype observed in mice homozygous for the N-terminal C2H2 ZF domain can be explained by a mechanism in which this mutated protein dominantly effects and disrupts other proteins that could otherwise partially compensate for a loss of function of the Ikaros protein (see next section).

1.3.2 Ikaros related proteins

1.3.2.1 Discovery

Shortly after the discovery of Ikaros, several other proteins were identified that have functional properties similar to Ikaros suggesting that these proteins compose a family of related transcription factors. These proteins were isolated using different approaches that were initially aimed at identifying potential interacting targets of Ikaros (see section 1.3.1.3): Aiolos and Helios were amplified from a mouse cDNA library using degenerate oligonucleotides directed against the N- or C-terminal domain of Ikaros (Morgan *et al.*, 1997; Kelley *et al.*, 1998). In addition, a separate independent study identified Helios as a component of an Ikaros containing complex isolated by gel filtration chromatography (Hahm *et al.*, 1998). Mouse Eos was originally discovered in a screen for upregulated genes in long term cultured astrocytes (Honma *et al.*, 1999). At the same time, human Eos and Pegasus were identified in a Yeast two-hybrid screen using the C-terminal domain of Aiolos (Perdomo *et al.*, 2000). All of these proteins harbor four N-terminal C2H2 ZFs and two C-terminal C2H2 ZFs and are homologous to each other. The highest level of homology occurs

in the C-terminal C2H2 ZF domain (Figure 1.5) and it has been shown that this domain can mediate homo- and hetero-typic interactions among these various transcription factors as suggested by the way Eos and Pegasus were isolated.

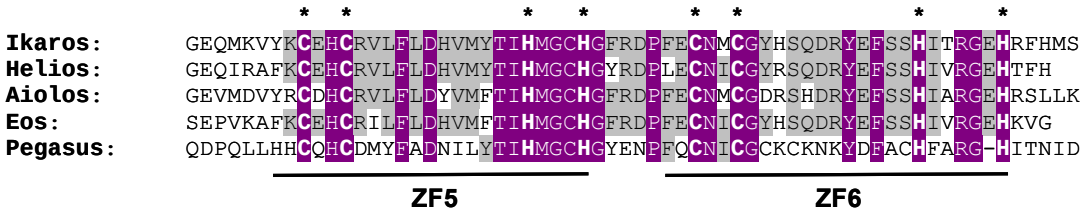

Figure 1.5 Amino acid alignment of the C-terminal ZF domain of Ikaros-family proteins. The C-terminal domain contains two ZFs (ZF5 and ZF6). Residue positions that are identical in all proteins are in purple while residue positions that a identical in three or more family members are shown in grey. Asterisks indicate putative zinc-ligating cysteine and histidine.

1.3.2.2 Similarities and differences among Ikaros family members

As in the case of Ikaros, alternative splice variants have been described for some of these family members. At least seven additional isoforms have been identified for Helios and five for Aiolos whereas no isoforms appear to exist for Eos and Pegasus (Kelley *et al.*, 1998; Perdomo *et al.*, 2000; Liippo *et al.,* 2001; Nakase *et al.*, 2002). All isoforms contain the C-terminal C2H2 ZFs while the arrangement of C2H2 ZFs at the N-terminus is variable. Like Ikaros, Aiolos and Helios are present in hematopoietic cell lines where they display a distinct and well defined expression profile. For example, Aiolos is detected at low levels in hematopoietic progenitor populations and is upregulated at intermediate and late stages of T and B cell development (Morgan *et al.,* 1997). On the other hand, Helios is primarily found in T cells (Hahm *et al.,* 1998; Kelley *et al.,* 1998). Eos and Pegasus are more broadly expressed and their expression is not restricted to the hematopoietic system although they also have overlapping expression patterns with Ikaros and Helios (Honma *et al.,* 1999; Perdomo *et al.,* 2000).

The DNA binding specificities of the Ikaros family members have been further analyzed. While Aiolos and Eos can bind to the conserved Ikaros binding site (GGGAA) with levels of specificity and affinity comparable to Ikaros, Helios and Pegasus appear to interact with distinct DNA sites. Binding sites for Helios are characterized by the core sequence of GGGA or GGAAAA (Hahm *et al.,* 1998), while Pegasus can bind to sites defined by a loose consensus of GNNNGNNG (Perdomo *et al.*, 2000).

Thus, certain members of the Ikaros family display some similar features consistent with the idea that they play a role in regulating lymphoid cell development. However, individual proteins also have unique characteristics, suggesting that each Ikaros-family protein is involved in distinct functions during hematopoiesis. Furthermore, since these proteins apparently possess partially similar DNA binding sites, they could, in principle, compete for target sites in the regulatory region of certain genes (Morgan *et al.,* 1997). Although such a mechanism has not been demonstrated yet, it may provide an additional strategy of regulating gene expression in the developing lymphoid system.

1.3.3 Mechanism of Ikaros action

The function of Ikaros-family members as master regulators of lymphoid cell development was proposed briefly after their discovery (Georgopoulos *et al.*, 1992; Hahm *et al.,* 1994; Molnar and Georgopoulos, 1994). These proteins are composed of C2H2 ZF domains which are most commonly found in DNA-binding transcription factors and putative target site have been identified in regulatory elements of several lymphoid-specific genes (Molnar and Georgopoulos, 1994; Christopherson *et al.*, 2001; Lopez *et al.*, 2002). The mechanism by which Ikaros-family proteins act on gene expression is likely to be very complex and both activation as well as repression has been observed for these proteins. For example, Ikaros and Aiolos can both activate and repress gene expression when recruited to DNA through a heterologous DNA-binding domain (Sun *et al.*, 1996; Morgan *et al.,* 1997; Brown *et al.*, 1997; Koipally *et al.*, 1999; Koipally and Georgopoulos, 2000; Koipally and Georgopoulos, 2002). Helios appears to solely act as an activator (Kelley *et al.,* 1998) while Eos and Pegasus have been shown to function as repressors (Perdomo *et al.*, 2000).
Although the details of transcriptional regulations mediated by the Ikaros-family is not well characterized, different lines of evidence have shown that localization of the transcription factors as well as protein-protein interactions with themselves and other proteins influence their regulatory function (see below).

1.3.3.1 Interactions with other proteins

To mediate their function in the hematopoietic system, members of the Ikaros-family interact with a wide variety of different proteins ranging from co-activators and co-repressors to

chromatin remodeling factors as well as histone modifying enzymes. By doing so, Ikaros proteins participate in different regulatory complexes that either active or repress certain genes. For example, Ikaros and Aiolos are present in the nucleosome remodeling and deacetylase (NuRD) complex that also contains the ATPase Mi-2 (Kim *et al.,* 1999). This complex is involved in transcriptional repression due to its association with histone deacetylases (HDACs). On the other hand, Ikaros and Aiolos have also been described as components of the SWI/SNF chromatin-remodeling complex which can mediate gene activation (Kim *et al.,* 1999). Finally, Ikaros can repress transcription through its interaction with the C-terminal binding protein (CtBP) and this mechanism of repression is independent of histone deacetylase activity (Koipally and Georgopoulos, 2000). Furthermore, Eos has also been found associated with CtBP (Perdomo and Crossley, 2002) whereas Aiolos and Helios do not seem to bind to this protein (Koipally and Georgopoulos, 2000).

1.3.3.2 Targeting of Ikaros to pericentromeric heterochromatin

The potential of the Ikaros proteins to mediate interaction with various proteins may provide a molecular mechanism for their ability to function as both activators and repressors. However, an additional notable feature of the Ikaros protein is the observation that it can be found in regions of condensed chromatin near the centromers termed the pericentromeric heterochromatin (PC-HC) which are often associated with inactive genes. In fact, immunofluorescence in situ hybridization (immuno-FISH) analysis has demonstrated that Ikaros colocalizes with centromeric chromatin and various transcriptionally silenced genes, suggesting that Ikaros is responsible for their relocalization in the nucleus (Brown *et al.,* 1997; Klug *et al.,* 1998). Additional studies have shown that targeting of Ikaros to PC-HC requires both its DNA-binding domain and its C-terminal dimerization domain (Cobb *et al.,* 2000). It is noteworthy that Koipally and co-workers (2002) have found a surprising correlation between the localization of Ikaros to PC-HC and its ability to activate transcription. These findings led to the proposal of a model where Ikaros can function as a "potentiator" of gene expression by for example squelching repressor complexes (such as the NuRD complex) away from defined target genes into regions of PC-HC (Koipally *et al.,* 2002).

1.3.4 The Dimerization Zinc Finger (DZF) domain

1.3.4.1 Importance of the C-terminal domain for the function of Ikaros family proteins

As described above, the main function of the C-terminal C2H2 ZF domain appears to be to support homo- and heterodimerization among the various encoded isoforms as well as among different Ikaros family members (Sun *et al.,* 1996;; Morgan *et al.* 1997; Kelley *et al.* 1998; Koipally *et al.* 1999; Perdomo *et al.,* 2000). There is some evidence that C-terminal C2H2 ZF domain can weakly bind to DNA but a consensus binding site has not been identified yet (Georgopoulos *et al.*, 1992; Molnar and Georgopoulos, 1994). Furthermore, this domain lacks the typical TGEKP linker that characterizes most of the C2H2 ZFs and which has been shown to be important for DNA binding. This suggested that the C-terminal C2H2 ZF domain of the Ikaros family members might be exclusively involved in mediating protein contacts. It has been described that the homo-typic oligomerization mediated by the various C-terminal domains is necessary for high affinity DNA binding and subsequent transcriptional regulation suggesting that transcription factors of the Ikaros family might oligomerize on DNA (Molnar and Georgopoulos, 1994; Cobb *et al.*, 2000; Trinh *et al.*, 2001). In addition, these homotypic interactions are essential for localizing Ikaros to PC-HC which results in either silencing or activation of Ikaros target genes (Brown *et al.*, 1997; Cobb *et al.*, 2000; Koipally *et al.,* 2002). It is noteworthy that the C-terminal domain is not directly involved in targeting to the PC-HC but appears to be only required for dimerization since targeting was also achieved when this domain was replaced by the leucine zipper dimerization domain (Cobb *et al.,* 2000).

1.3.4.2 Dimerization ZF domains are also found in other transcription factors

Interestingly, the Trichorhinophalangeal syndrome 1 (TRPS1) protein also contains a homologous dimerization domain at its C-terminus. TRPS1 is a transcription factor that is associated with a dominantly inherited human disease characterized by skeletal abnormalities (Momeni *et al.*, 2000) and is otherwise not believed to belong to the Ikaros family of transcription factors (Momeni *et al.*, 2000). Yeast two-hybrid and co-immunoprecipitation assays suggested that TRPS1 interacts with itself and with Eos via its C-terminal zinc fingers (McCarty *et al.*, 2003; Westman *et al.*, 2003; Westman *et al.*, 2004). Recently, a dimerization domain from the Ikaros homologue Hunchback, a segmentation gap gene involved in embryonic pattern formation in *Drosophila melanogaster*, has been described which can

mediate homodimerization but can not heterodimerize with Ikaros *in vitro* (Tautz *et al.*, 1987; McCarty *et al.*, 2003). Further *in vitro* studies on different dimerization domains have shown that both C-terminal C2H2 ZFs are required for efficient dimerization (McCarty *et al.*, 2003; Westman *et al.*, 2004). Thus, the two C-terminal zinc fingers found in different transcription factors from the Ikaros family, TRPS1 and Hunchback have been shown to be essential as well as sufficient for dimerization. These domains are therefore referred to as dimerization zinc finger (DZF) domain (McCarty *et al.*, 2003).

1.3.4.3 Specificity determinants of the DZF domain

Using directed mutagenesis together with chemical crosslinking and co-immunoprecipitations, McCarty and co-workers (2003) performed initial studies to define the molecular determinants of DZF domain mediated interactions. These studies demonstrated that interactions between the DZF domains of Ikaros, Hunchback and TRPS1 are highly specific since each of these DZF domains can interact with themselves but do not heterodimerize with one another. To identify specificity determinants for the Ikaros and Hunchback DZF domains, McCarty and co-workers (2003) constructed a series of Ikaros-Hunchback DZF chimeras by introducing increasing amounts of one protein in the other protein. These chimeras were then tested for their ability to interact with either the wild-type Ikaros or the wild-type Hunchback DZF domain. As shown in Figure 1.6, minimal regions required for selective dimerization for both Ikaros and Hunchback were defined in this way. These regions overlap but differ in length (Figure 1.6, McCarty *et al.,* 2003).

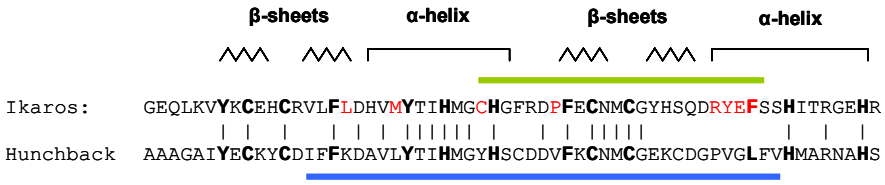

Figure 1.6 Minimal regions required for selective dimerization of Ikaros and Hunchback. Selectivity regions defined by analysis of chimeric proteins are indicated by a green line (for Ikaros) and a blue line (for Hunchback). Residue positions important for Ikaros dimerization are shown in red. Conserved residues are in bold. Residues that are identical in both proteins are connected by a line. Predicted secondary structures are indicated above the amino acid sequences. This figure was adapted from McCarty *et al.* (2003).

Interestingly, one of their chimeras was able to homo-dimerize but could not interact with the two wild-type DZF domains (from Ikaros and Hunchback) indicating that this chimeric DZF domain possesses a novel interaction specificity. Furthermore, the preferential homodimerization of this protein suggests that it interacts using a parallel interaction mode (i.e. the N-terminal C2H2 ZFs in each monomer interact with each other and C-terminal C2H2 ZFs in each monomer interact with each other). This conclusion was based on the reasonable assumption that the individual C2H2 ZFs in this hybrid protein harbor the same interaction specificities then the parental DZFs from which they were constructed (McCarty *et al.*, 2003).

To precisely define residues that are important for dimerization, McCarty and co-workers (2003) further introduced single or double mutations into the Ikaros DZF domain by replacing every amino-acid in Ikaros that differs between Ikaros and Hunchback with the corresponding Hunchback amino-acid at the same position. These "swap" mutants were then tested for their abilities to homodimerize as well as heterodimerize with the wild-type Ikaros DZF domain. As shown in Figure 1.6, eight residues were identified that are likely to be important for the Ikaros DZF domain interaction. Interestingly, many of these residues are present in the predicted α-helices, suggesting that this structure, like in the case of DNA-binding, is important for protein-recognition (McCarty *et al.,* 2003).

1.4 Goals of this thesis

Over the last decade, much effort has focused on studying DNA recognition by C2H2 ZF proteins while their role as protein-binding modules was revealed more recently (reviewed in Mackay and Crossley, 1998). The abundance and versatility of ZFs suggests that they comprise a favorable motif in nature for evolving a large number of functional properties in the cell that are not just limited to DNA recognition. The fold of ZFs itself is relatively simple suggesting that the variability within non-conserved regions of this domain must mediate binding specialization. Our understanding of how affinity and specificity is provided by such a stable framework is still limited. Such knowledge would provide useful information that will eventually allow us to predict protein interactions based on sequence information, and engineer synthetic interaction domains that can be used for applications in gene therapy and biotechnology.

The Ikaros family of transcription factor represents an example of C2H2 ZF proteins that can mediate interactions with DNA as well as with other proteins. While DNA-binding mediated by the N-terminal domain is likely to be similar to other characterized DNA-binding ZFs, relatively little is know about the details of the protein interactions mediated by the C-terminal domain (DZF domain). Therefore, we sought to address this lack of understanding by examining the DZF domains found in the Ikaros and Hunchback transcription factor family using a combination of genetic, biochemical and functional assays. This should add to the existing knowledge of both the ability of C2H2 ZFs to mediate protein-protein interactions in general as well as the biological importance of the DZF domain.

Chapter 2. Materials and Methods.

2.1 General techniques

Standard buffers and solutions were prepared as described in Sambrook and Russell (2001). Use of commercially available solutions and buffers is mentioned at the relevant steps. Unless otherwise stated Milli-Q water was used to prepare buffers and HCl or NaOH was used to adjust the pH of all buffers. Solutions were filter-sterilized through a 0.2 µm filter and media was autoclaved if not mentioned otherwise. Note that all reporter strains and constructs including the methods used to generate them are listed in the appendix.

2.1.1 Materials

2.1.1.1 Buffers and solutions

- 10 x Annealing buffer: 400 mM Tris (pH 8.0), 200 mM $MgCl_2$, 500 mM NaCl
- 2 x Quick ligation buffer: 132 mM Tris (pH 7.6), 20 mM $MgCl_2$, 2 mM ATP, 15 % Polyethylene glycol. Fresh DTT was added to a final concentration of 2 mM.
- DNA elution buffer: 0.5 M NH_4OAc, 10 mM MgOAc, 1 mM EDTA (pH 8.0), 0.1 % SDS
- Competent cell solution A: 10 mM $MnCl_2$, 50 mM $CaCl_2$, 10 mM MES (pH 6.3, pH was adjusted with KOH)
- Transfer buffer: 100 ml of 10 x Tris-Glycin (Biorad), 200 ml of Methanol. Distilled water was added to 1 liter.
- TBST: 20 ml of 1 M Tris-HCl (pH 7.5), 8 g of NaCl, 1 ml of Tween 20. Distilled water was added to 1 liter.

2.1.1.2 Antibiotics

Antibiotic	Stock solution	Final concentration in plates	Final concentration in liquid medium
Carbenicillin (Carb)	50 mg/ml in H_2O	100 µg/ml	50 µg/ml
Chloramphenicol (Cam)	30 mg/ml in ethanol	30 µg/ml	30 µg/ml
Kanamycin (Kan)	30 mg/ml in H_2O	30 µg/ml	30 µg/ml
Tetracycline (Tet)	12.5 mg/ml in 80 % EtOH	12.5 µg/ml	12.5 µg/ml

2.1.1.3 Media

LB agar plates and Liquid LB medium were prepared as described in Sambrook and Russell (2001).

Plates:

LB/C plates	LB plates with Carb
LB/CA plates	LB plates with Cam
LB/K plates	LB plates with Kan
LB/T plates	LB plates with Tet

2.1.1.4 PAGE Gels

Reagents for preparation of analytical and denaturing polyacrylamid gel electrophoresis (PAGE) gels were obtained from National Diagnostics. Note that a ratio of 29:1 Acrylamide/Bisacrylamide was applied for native gels while a ratio of 19:1 Acrylamide/Bisacrylamide was used for denaturing gels containing urea.

Analytical Gels:

Reagent	**Amount for 5 % gel**	**Amount for 10 % gel**
H_2O	40.5 ml	34.25 ml
10 x TBE buffer	2.5 ml	2.5 ml
10 % APS	700 µl	700 µl
TEMED	50 µl	50 µl
40 % Bisacrylamide Stock	6.25 ml	12.5 ml

Denaturing polyacrylamide urea gel:

Oligo length (in nucleotides)	**% Acrylamide**	**Reagent**				
		Diluent	**Buffer**	**APS**	**TEMED**	**25 % Concentrate**
≤ 25	19	7 ml	5 ml	400 µl	20 µl	38 ml
26-40	15	15 ml	5 ml	400 µl	20 µl	30 ml
41-100	12	21 ml	5 ml	400 µl	20 µl	24 ml

2.1.1.5 Bacterial strain

E. coli XL-1 Blue: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′ *proAB lacI*qZΔ*M15* Tn*10* (Tet^r)] (Stratagene).

2.1.2 Methods

2.1.2.1 PAGE purification of oligonucleotides

PAGE was used to purify oligonucleotides (oligos, primers) >40 bp using denaturing polyacrylamide urea gels at various concentrations (depending on the length of the oligo, see section 2.1.1.4) as described in Sambrook and Russell (2001) with a few differences. Eluted oligonucleotides were equally mixed with deionized formamide, incubated at 95°C for 2 min and stored at 4˚C. The samples along with a sample containing formamide with bromophenyl blue were loaded onto a denaturing polyacrylamide urea gel which had been pre-run for 30 min at 23 watts. Following loading of the sample, the gel was run at 23 watts until the bromophenyl blue sample was about 3/4 of the way to the bottom. The oligonucleotides were visualized using UV shadowing, excised and eluted by incubating them in 8 ml TE buffer at 37˚C for 4 hr or overnight (ON). To further purify the extracted oligonucleotides, a C18 Sep-Pak column (Waters) was prepared by attaching it to the barrel of a 10 ml polypropylene syringe and pushing 5 ml 70 % acetonitrile and subsequently 5 ml TE buffer through it. The oligonucleotides samples were applied to the column, washed with 2 ml TE and eluted in 3 ml 70 % acetonitrile. The samples were dried down using a SpeedVac, resuspended in TE buffer, EtOH precipitated and pellets were resuspended in TE buffer. The concentration of the oligonucleotide was determined by measuring the absorbance at 260 nm using a Beckman Coulter DU 640 Spectrophotometer.

2.1.2.2 Polymerase chain reaction (PCR)

Various PCR strategies were used to synthesize and clone the different target genes (Figure 2.1). The Expand High Fidelity PCR System (Roche) was used to perform PCR reactions. This system contains the thermostable *Taq* DNA Polymerase and the thermostable *Tgo* DNA polymerase, which provides proofreading activity (Expand High fidelity enzyme mix). The corresponding buffer system (10 x Expand buffer) contains $MgCl_2$ at a final concentration of 15 mM. Resulting final PCR products were purified using gel electrophoresis (see section 2.1.2.10).

Strategy A:

DNA fragments encoding the various target genes were amplified from either plasmid DNA or human cDNA library using PCR. Specific restriction sites used for further cloning procedures were incorporated in both PCR oligonucleotides at the 5' ends (Figure 2.1A). 5-10 cycles with a partial annealing temperature (T_pa) were performed to allow initial binding of the primers to the template. A subsequent step of 15-20 cycles using the full annealing temperature (T_fa) was carried out for full binding of the primers to the template. The following conditions were used for a 50 µl reaction:

Reaction:

Reagent	**Final concentration**	**Volume**
Template DNA	0.1–200 ng	Variable
10 x Expand buffer with 15 mM $MgCl_2$	1 x	5 µl
10 mM dNTP mix	200 µM of each dNTP	4 µl
Top strand primer (10 pmol/µl)	200 nM	1 µl
Bottom strand primer (10 pmol/µl)	200 nM	1 µl
Expand High fidelity enzyme mix	1.3 U/reaction	0.375 µl
Sterile dH_2O		Add up to 50 µl

Settings:

	Temperature	**Time**	**Cycles**
Initial Denaturation	95˚C	5 min	1 x
Denaturation Partial annealing Elongation	95˚C 45-60˚C 72˚C	30 s 30 s 1 min	5–10 x
Denaturation Full annealing Elongation	95˚C 55-70˚C 72˚C	30 s 30 s 1 min	15–25 x
Final Elongation	72˚C	5 min	1 x
Cooling	4˚C	Unlimited time	1 x

Strategy B:

For *de novo* synthesis of the template, various overlapping oligonucleotides (P1, P2, P3, P4, P5 and P6) were annealed using the following conditions for a 50 µl reaction (Figure 2.1B):

Reaction:

Reagent	Final concentration	Volume
10 x Expand buffer with 15 mM $MgCl_2$	1 x	5 µl
10 mM dNTP mix	200 µM of each dNTP	4 µl
Primer P1–P6 (10 pmol/µl)	200 nM of each primer	1 µl of each primer
Expand High fidelity enzyme mix	1.3 U/reaction	0.375 µl
Sterile dH_2O		Add up to 50 µl

Settings:

	Temperature	Time	Cycles
Initial Denaturation	95˚C	5 min	1 x
Denaturation Primer annealing Elongation	95˚C 45-55˚C 72˚C	30 s 30 s 1 min	10 x
Final Elongation	72˚C	5 min	1 x
Cooling	4˚C	Unlimited time	1 x

The annealed oligonucleotides were purified using a standard QIAgen PCR purification kit and eluted with 50 µl of 0.1 x elution buffer (EB buffer). The resulting product was then amplified with a pair of external oligonucleotides (P1 and P6) using the following conditions (Figure 2.1B):

Reaction:

Reagent	Final concentration	Volume
Template DNA	variable	5 µl
10 x Expand buffer with 15 mM $MgCl_2$	1 x	5 µl
10 mM dNTP mix	200 µM of each dNTP	4 µl
Top strand primer (10 pmol/µl)	200 nM	1 µl
Bottom strand primer (10 pmol/µl)	200 nM	1 µl
Expand High fidelity enzyme mix	1.3 U/reaction	0.375 µl
Sterile dH_2O		Add up to 50 µl

Settings:

	Temperature	Time	Cycles
Initial Denaturation	95˚C	5 min	1 x
Denaturation Full annealing Elongation	95˚C 55-70˚C 72˚C	30 s 30 s 1 min	20 x
Final Elongation	72˚C	5 min	1 x
Cooling	4˚C	Unlimited time	1 x

Strategy C:

A three step PCR approach was applied to create target genes consisting of two different fragments that are fused together. In the first step (step 1), the 2 fragments of interest were individually amplified using standard PCR conditions as described below. Primer P2 and P3 carried a common region at the 5' end (red box in Figure 2.1C) which was used for the subsequent fusion PCR step (Step 2). In a final step (Step 3), the resulting PCR fusion product was amplified with a pair of external oligonucleotides (P1 and P4).

Step 1 was performed as described in Strategy A. Both amplified fragments were then purified using a standard QIAgen PCR clean up kit and eluted with 50 µl of 0.1 x EB buffer. To anneal the resulting products together the following conditions for a 50 µl reaction were used (Step 2):

Reaction:

Reagent	Final concentration	Volume
PCR Fragment 1	Variable	5 µl
PCR Fragment 2	Variable	5 µl
10 x Expand buffer with 15 mM $MgCl_2$	1 x	5 µl
10 mM dNTP mix	200 µM of each dNTP	4 µl
Expand High fidelity enzyme mix	1.3 U/reaction	0.375 µl
Sterile dH_2O		Add up to 50 µl

Settings:

	Temperature	Time	Cycles
Initial Denaturation	95˚C	5 min	1 x
Denaturation Full annealing Elongation	95˚C 55-70˚C 72˚C	30 s 1 min 1 min	4 x
Final Elongation	72˚C	5 min	1 x
Cooling	4˚C	Unlimited time	1 x

The resulting PCR fusion product was purified using a standard QIAgen PCR purification kit, eluted with 50 µl of 0.1 x EB buffer and amplified with a pair of external oligonucleotides (P1 and P4) as described in Strategy B (Step 3).

Strategy D:

To introduce site directed mutations into the target sequence, a three step PCR approach was applied similar as described in Strategy C (Figure 2.1D). In the first step, the left and right cassette of the target gene was individually amplified using standard PCR conditions.

Overlapping sequence regions containing the desired point mutation were used as internal primers (P2 and P3) for the subsequent fusion PCR step. In a final step, the resulting PCR fusion product was amplified with a pair of external oligonucleotides (P1 and P4). All 3 steps were performed as described in Strategy C. Note that all constructs harboring single point mutations were generated using this strategy if not otherwise stated.

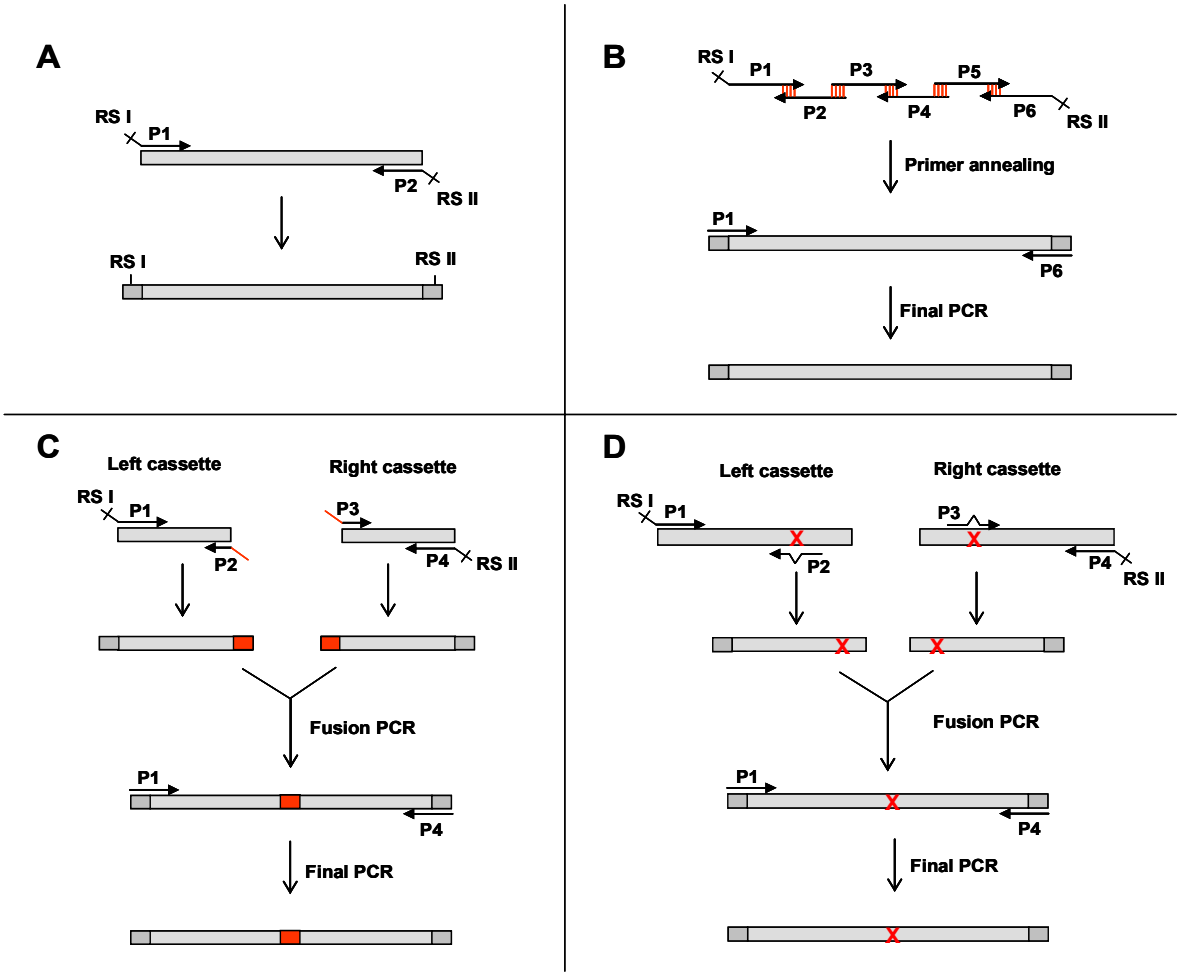

Figure 2.1 Scheme of PCR strategies used to synthesize and amplify target genes. (A) DNA fragments encoding the target gene were amplified from available template DNA. Specific restriction sites (RS I and RS II) used for further cloning procedures were incorporated in both PCR oligonucleotides (P1 and P2) at their 5' ends. **(B)** For *de novo* synthesis of the target gene, various overlapping oligonucleotides (P1, P2, P3, P4, P5 and P6, overlapping region depicted as red lines) were annealed together and amplified with a pair of external oligonucleotides (P1 and P6). **(C)** A target gene composed of two different fragments was generated in 3 steps. The individual fragments (left and right cassette) were amplified using standard PCR (as described in **A**). Primer P2 and P3 carried a common region at the 5' end (red box) which was used to fuse the two fragments together. The resulting fusion product was amplified with a pair of external oligonucleotides (P1 and P4). **(D)** To introduce site directed point mutations into the target sequence, a left and a right cassette of the target gene were individually amplified. Overlapping sequence regions containing the desired mutation (indicated by red X) were incorporated in the internal primers (P2 and P3) to subsequently fuse the two cassettes together. In a final step, the resulting PCR fusion product was amplified with a pair of external oligonucleotides (P1 and P4).

2.1.2.3 Primer annealing

To clone short DNA fragments consisting of only 10–30 base pairs, a primer annealing approach was used. Two complementary oligonucleotides were annealed together to create a double stranded DNA fragment bearing specific overhangs that were used for further cloning procedures. Annealing was performed using the following conditions for a 200 µl reaction:

Reagent	Final concentration	Volume
Oligonucleotide 1 (10 pmol/µl)	50 nM	1 µl
Oligonucleotide 1 (10 pmol/µl)	50 nM	1 µl
10 x Annealing buffer	1 x	20 µl
Sterile dH_2O		178 µl

Reaction was incubated at 95°C for 2 min on a heat block. The heat block was shut off allowing the reaction to slowly cool down to ~ 35°C. The samples were stored at 4˚C after the reaction.

2.1.2.4 Restriction endonuclease digestion of DNA

Digest reactions were performed using conditions recommended by the manufacturer for the specific enzymes. Enzymes and buffer systems were obtained from New England Biolab (NEB).

2.1.2.5 Making chemical competent cells

XL-1 Blue chemical competent cells were prepared as described in the following. 1 l of LB medium containing Tet and 15 mM $MgCl_2$ was inoculated with 1 ml of an ON culture of XL-1 Blue and the culture was incubated at 37˚C with shaking (250 rpm) until cell density reached OD_{600} = 0.3-0.6. Cells were centrifuged at 4000 rpm at 4˚C for 30 min and the cell pellet was resuspended in 300 ml of Solution A. The cell suspension was incubated on ice for 20 min, pelleted by centrifugation and resuspended in 60 ml of Solution A containing 15 % glycerol. The resulting chemical competent cells were aliquoted, frozen in a dry ice/ethanol bath and stored at -80°C.

2.1.2.6 Ligation and Transformation

Ligation reactions were performed in 20 µl using the following conditions:

Reagent	Final concentration	Volume
Purified Bkb DNA	Variable	1 µl
Purified Fragment DNA	variable	8 µl
2 x Quick ligation buffer	1 x	10 µl
T4 DNA Ligase (NEB)	400/reaction	1 µl

The ligation reaction was incubated at room temperature (RT) for 5 min and stored at 4˚C after the reaction for further transformation procedure. Ligations were transformed into chemically competent XL-1 Blue *E. coli* cells essentially as described (Sambrook and Russell, 2001).

2.1.2.7 Plasmid purifications

Plasmid DNA from ON cultures grown from single colonies was purified using the Qiagen Miniprep kit according to the manufacturer's protocol using 50 µl of 0.1 x EB buffer for elution. The DNA was verified by restriction analysis and sequencing (by sending them to the Massachusetts General Hospital (MGH) sequencing core facility). To obtain larger amounts of plasmid DNA, sequence verified plasmids were re-transformed to perform Midi- or Maxi-preps using the QIAgen Midi or Maxi kit, respectively.

2.1.2.8 DNA precipitation

Ethanol (EtOH) precipitation was used to concentrate DNA as described in Sambrook and Russell (2001).

2.1.2.9 Purification of DNA fragments and digests

To clean up PCR intermediate products and restriction digests of PCR fragments, a standard QIAgen PCR purification kit was used according to the manufacturer's protocol using 20-50 µl of 0.1 x EB buffer for elution.

2.1.2.10 Gelisolation of Plasmid/Fragment DNA and PCR Fragments

Plasmid digests and final PCR reactions were separated using Gel electrophoresis. We generally used agarose gel electrophoresis for fragments from > 1kb and polyacrylamide gel electrophoresis (PAGE) for smaller fragments (20 bp to 800 bp). Agarose gels were run in 1 x TAE buffer (Sambrook and Russell, 2001) at 100 V and separated fragments were extracted

and purified using the QIAgen Gel extraction kit according to the manufacturer's protocol. PAGE was performed in 0.5 x TBE buffer (Sambrook and Russell, 2001) at 200 V and the purification of extracted fragments from a PAGE gel was essentially done as detailed in Sambrook and Russell (2001). Briefly, DNA fragments were excised, crushed and extracted by eluting them in 700 µl of Elution buffer at 37˚C for 4 hr or ON. The samples were centrifuged twice and the supernatants were transferred to a new tube. An EtOH precipitation was performed and the resulting DNA pellets were resuspended in dH_2O.

2.1.2.11 SDS-Polyacrylamide Gel electrophoresis of proteins

Analytical electrophoresis of proteins was carried out using 10–20% precast gels (Biorad) which were run in 1 x TGS (Biorad) buffer at 100 V. Staining of SDS-polyacrylamide gels was performed using either Coomassie staining or silver staining (for more sensitive detection) following standard protocols (Sambrook and Russell, 2001).

2.1.2.12 Western blot analysis

To transfer proteins from SDS-polyacrylamide gels to a membrane, a Trans-Blot SD semi-dry transfer cell (Biorad) was used according to the manufacturer's instructions. Proteins were transferred at a constant voltage of 25 V for 1 hr in transfer buffer and the membrane was subsequently rinsed in TBST buffer. To block nonspecific binding sites, the membrane was submerged in 5 % milk (dissolved in TBST) and incubated at 4˚C ON. Following the blocking, the membrane was rinsed 3 x in TBST buffer and subsequently incubated with the primary antibody diluted in TBST/5 % milk for 1-2 hr at RT with slow shaking. Upon completion of the incubation, the membrane was rinsed 3 x with TBST and washed 3 x for 10 min at RT with shaking. Secondary antibodies were linked to the enzyme Horseradish Peroxidase (HRP) for subsequent detection. This antibody was added in TBST/5 % and the membrane was incubated for 1-2 hr at RT with slow shaking. Washing was performed as before and ECL Plus kit (Amersham) was used for visualization according to the manufacturer's protocol.

2.2 Genetic techniques

2.2.1 Materials

2.2.1.1 Buffers and solutions

- Z buffer: 16.1 g of $Na_2HPO_4 \cdot 7H_2O$, 5.5 g of $NaH_2PO_4 \cdot H_2O$, 0.75 g of KCl; 0.246 g of $MgSO_4 \cdot 7H_2O$. Distilled water was added to 1 liter. 2.7 µl β-mercaptoethanol was added per 1 ml prior to use.
- 5 x PEG/NaCl solution: 17.5 % PEG 8000, 12.5 % NaCl
- Amino acid mixture: Amino acid mixture containing 17 different amino acids (Phe 0.99 %, Lys 1.1 %, Arg 2.5 %, Gly 0.2 %, Val 0.7 %, Ala 0.84 %, Trp 0.41 %, Thr 0.71 %, Ser 8.4 %, Pro 4.6 %, Asn 0.96 %, Asp 1.04 %, Gln 14.6 %, Glu 18.7 %, Tyr 0.36 %, Ile 0.79 % and Leu 0.79 %) was prepared as described in Giesecke and Joung (2005).

2.2.1.2 Media

LB agar plates and liquid LB medium, liquid 2 x YT medium and SOC medium were prepared as described in Sambrook and Russell (2001).

Liquid NM medium: for 500 ml:

dH_2O	418 ml
10x M9 salts (Miller recipe)	50 ml
20% glucose	10 ml
20 mM adenine HCl	5 ml
Amino acid mixture (see above)	15 ml
1 M $MgSO_4$	0.5 ml
10 mg/ml thiamine	0.5 ml
10 mM $ZnSO_4$	0.5 ml
100 mM $CaCl_2$ (was added last)	0.5 ml

Medium was filter-sterilized through a 0.2-µm filter.

Plates:

LB/T/C plates	LB plates with Tet and Carb
LB/C/K plates	LB plates with Carb and Kan
LB/C/K^{70} plates	LB plates with Carb and Kan (70 µg/ml)
LB/T/K plates	LB plates with Tet and Kan

LB/T/K/S plates	LB plates with Tet, Kan and 5% sucrose
LB/CA/K plates	LB plates with Cam and Kan
LB/C/CA/K plates	LB plates with Carb, Cam and Kan

NM plates:

For 500 ml of plates, 418 ml of dH_2O with 7.5 g of Bacto-Agar was autoclaved in a flask containing a stir bar. Agar was cooled to 65–70^oC and the basic NM components (as for the liquid NM medium) were premixed and added to the agar.

NM/C/CA/K plates	NM plates with Carb, Cam and Kan
Selection plates	NM/C/C/K plates with isopropyl-beta-D-thiogalactopyranoside (IPTG, 50 μM), 3-Aminotriazole (3-AT, 25 mM) and streptomycin (40 μg/ml).
2 x YT/T plates	2 x YT plates with Tet

2.2.1.3 Bacterial strains

- CSH100: [F'*lacproA +, B+ (lacI^q lacPL8) / araD (gpt-lac)5*].
- *E. coli* KJ1C: ΔhisB463, Δ(gpt-proAB-arg-lac)XIII, zaj::Tn10 [F' Tet^r] (Joung *et al.*, 2000).

2.2.2 Methods

2.2.2.1 Construction of bacterial two-hybrid reporter strains

Overview:

To construct reporter strains, the promoter region containing the DNA binding site (DBS) for a DNA binding protein of interest is introduced onto a plasmid termed pSB: For constructing *lacZ* reporter strains, the DBS was assembled on the pSB-*lacZ* plasmid which harbors the *lacZ* gene to assess reporter activity by performing β-galactosidase assays (see section 2.2.2.2). For constructing *his3*/*aadA* selection strains, the DBS was introduced into the pSB-*his3*/*aadA* plasmid which also harbors the bacterial *aadA* gene and the *his3* gene (Joung *et al.,* 2000; Hurt *et al.*, 2003). While the *aadA* gene confers resistance to streptomycin, the *his3* gene permits histidine biosynthesis. Co-cistronically expression of these genes allows *E. coli* cells to grow on selective medium containing streptomycin but lacking histidine (see section 2.2.2.16).

After introducing the DBS onto the respective pSB plasmid, it is subsequently moved onto a F' episome using homologous recombination. In a final step, the F' episome is transferred into a "clean" background strain that represents the desired reporter strain (Whipple, 1998).
The pSB plasmid is a derivate of the pFW11 plasmid which contains the 3' end of the lacI^q gene and the 5' end of the *lacZ* gene (Whipple, 1998). The region between the *lacI*^{q} and *lacZ* fragments harbors a Kan resistance gene and an EcoRI-SalI segment. This segment also bears the -35 and -10 hexamers of the promoter that controls *lacZ* and a pair of SapI restriction sites. In addition, plasmid pSB carries the *sacB* gene outside the *lacI*^{q}*–lacZ* region which leads to sucrose sensitivity upon expression in *E. coli*. When pSB is transformed into CSH100 cells which contain an F' episome bearing the complete lac operon (including the *lacI*^{q} and *lacZ* genes), the region between the *lacI*^{q} and *lacZ* fragments can recombine onto the F' episome via double homologous recombination using the *lacI*^{q} and *lacZ* homology regions. Because single recombination events will also occur on the episome, one has to select for the double recombination events. Doubly F' recombinants can be distinguished from singly and non-recombinants because they will lose the *sacB* marker gene upon completion of the double recombination (in contrast to singly recombinants) but will obtain Kan resistance (in contrast to non-recombinants). Thus, in a final step the F' episome is moved into the Tet resistant recipient strain KJ1C (Joung *et al.*, 2000) via conjugation and the cells are plated on medium that contains Tet, Kan and sucrose. Only cells that have received the F' episome containing the double recombination product will be able to form colonies on these plates (Whipple, 1998; Joung *et al.*, 2000).

Crosses:
Crosses for reporter strain construction were essentially done as described in Whipple (1998) with a few differences. The pSB stuffer plasmids were transformed into CSH100 strains and plated on LB-agar plates containing Kan as the selective drug. The plates were incubated at 37°C for 12–18 hr. For each reporter strain, starter cultures of the plasmid-containing CSH100 strain were inoculated. To do this, the bacteria lawn was scraped off the plates with a sterile wooden stick, transferred to 10 ml LB medium and resuspended by vortexing. 0.2 ml of the resulting cultures was then used to inoculate tubes containing 10 ml LB_0 medium. An additional tube with 10 ml LB_0 medium was inoculated with 0.2 ml of a KJ1C ON culture. The CSH100 cultures, the KJ1C culture and a tube with only 10 ml LB_0 medium as a control were incubated at 37°C for 2 hr without agitation.

After the 2 hr incubation, crosses and cross controls were set up as followed:

- reporter containing CSH100 + KJ1C (actual crossing)
- reporter containing CSH100 + LB_0 medium (negative control)
- KJ1C + LB_0 medium (negative control)
- LB_0 medium + LB_0 medium (negative control)

To do this, 1 ml of each component was gently mixed together and incubated at 37°C for 1 hr without agitation. The mixtures were then placed on the wheel for 1.5 hr at 37°C. Upon completion of the 1.5 hr incubation, 0.3 ml of each actual cross was plated on LB/T/K plates as well as on LB/T/K/S plates. In addition, 20 µl of the various crossing controls were spotted on both type of plates. The plates were incubated at 37°C for 12–18 hr. After the ON incubation, there should be a >10-fold reduction in the number of colonies on the LB/T/K/S plates compared to the LB/T/K plates and the controls should be clean. If these two conditions were met, 2 candidates (A and B) for each reporter were picked from the LB/T/K/S plates and re-streaked twice on LB/T/K/S plates. Two individual tests were performed to check if the complete F' episome of strain CSH100 was transferred to KJ1C: (1) The original KJ1C strain harbors a deletion of genes involved in proline and arginine biosynthesis. Upon complete conjugation, KJ1C strains should obtain intact copies of these genes from the episome of the CSH100 strain and therefore be able to grow on plates without proline and arginine. Thus, colonies from the second re-streak were patched on plates lacking these amino acids to test whether they were able to grow on these plates. (2) ON cultures were inoculated with the same colonies and were then used as templates to perform PCR with specific primers that bind to the *lacZ* gene to test, if the complete *lacZ* gene was moved to the KJ1C strain. At the same time, glycerol stocks were made from the remaining ON cultures. If both test turned out positive, the reporter strains were made chemically competent.

2.2.2.2 β-galactosidase reporter assays

Double transformations into bacterial two-hybrid reporter strains and β-galactosidase assays were performed as previously described (Thibodeau *et al.*, 2004; Giesecke and Joung, 2005). Briefly, LB supplemented with Carb, Cam, Kan, 10 µM $ZnSO_4$, and 50 µM IPTG was inoculated with ON cultures grown from single colonies that harbored two plasmids. Logarithmic phase bacterial cultures were lysed using a commercially available lysis reagent (BugBuster, Novagen). The resulting extracts were assayed for β-galactosidase activity by

assessing the ability of this enzyme to cleave Ortho-Nitrophenyl-βD-galactopyranoside (ONPG) which results in a detectable color change. To do this, kinetic assays were performed in a 680 Microplate reader (Biorad) by measuring the absorbance at 415 nm (relative to a blank) for ~ 30 min.

β-galactosidase reporter assays for the Bacterial one-hybrid system were essentially performed as described above with a few differences. Plasmids encoding the various peptides were introduced by single transformation into reporter strain FW123. 20 µl of cell lysates from logarithmic phase cultures grown in LB containing Kan, Cam and IPTG at various concentrations were mixed with 1 x chlorophenolred-ß-D-galactopyranoside (CPRG, Stratagene) and kinetic assays were performed in a 680 Microplate reader by measuring the absorbance at 570 nm (relative to a blank) for ~ 30 min at 37˚C.

2.2.2.3 Western Blot analysis of β-galactosidase cultures

For Western blot analysis of the expression level of peptides expressed to perform β-galactosidase assays, whole cell lysates from 0.5 ml cultures used for β-galactosidase assays were obtained by pelleting the cells and resuspending the pellets in 60 µl of 2 x SDS sample buffer (Biorad). The amount of total protein loaded was normalized by the OD_{595} and were resolved on a SDS-polyacrylamide gel. Western blot analysis was performed using the anti-FLAG M2 monoclonal antibody (Sigma) as described in section 2.1.2.12.

2.2.2.4 Making electroporation competent cells

To prepare electroporation competent cells, a XL-1 Blue ON culture grown from a fresh single colony was used to inoculate 6 l of 2 x YT medium supplemented with Tet. The cultures were incubated at 18˚C with shaking (250 rpm) for ~ 2 days until cell density reached OD_{600} = 0.3-0.6. The cultures were centrifuged at 4000 rpm at 4˚C for 30 min using six pre-chilled centrifuge bottles. Each cell pellet was resuspended in 225 ml of cold miliQ water and washed twice using the same conditions. After the second wash step each cell pellet was resuspended in 225 ml of 10 % pre-chilled glycerol. The cells were centrifuged at 6000 rpm at 4˚C for 20 min, resuspended in a total volume of 50 ml of 10 % pre-chilled glycerol, centrifuged again and resuspended in 10 % glycerol using a final volume of 6 ml. The resulting electroporation competent cells were aliquoted, frozen in a dry ice/ethanol bath and stored at -80°C.

2.2.2.5 Growing M13K07 helper Phage

2 µl of M13K07 (NEB) helper phage was streaked out on 2 x YT/Tet plates and covered with ~ 3.5 ml of Top agar containing ~ 0.2 ml of a XL-1 Blue ON culture. Following an ON incubation at 37°C, a piece of agar containing dozens of phage plaques was scooped of the plate and used to inoculate 1 l of 2 x YT medium supplemented with Tet. After 2 hr incubation at 37˚C with shaking, Kan was added to the cells to a final concentration of 70 µg/ml and the culture was grown at 37°C for 24 hr. To harvest the phage, cells were pelleted by centrifugation and the supernatant containing the phage was filtered through a 0.2-µm PES filter.

To further concentrate the helper phage, PEG precipitation was used. 4 volumes of phage were added to 1 volume of 5 x PEG/NaCl solution and the samples were incubated on ice for ~ 16 hr. The phage was centrifuged at 10000 rpm at 4˚C for 45 min and the phage pellet was resuspended in a small volume of 2 x YT containing 15 % glycerol. The titer of the phage was determined as described (Giesecke and Joung, 2005).

2.2.2.6 Overview library construction

To construct libraries of shuffled dimerization zinc finger (ZF) domains (see section 3.3), subdomains encoding individual N-terminal and C-terminal ZF fragments were initially generated using standard PCR where the internal primers were phosphorylated at their 5' (Figure 2.2). These fragments were then reassembled by performing directional blunt end ligation via the phosphorylated internal primers. PCR was used to amplify products of the blunt end ligations. Shuffled DZFs were then ligated into the pACYC-α and pBR-UV5-Zif268 plasmids to express them as fusions to the *E. coli* RNAP α-subunit and the Zif268 DNA-binding domain (DBD), respectively (Figure 2.2). Ligations into the pACYC-α plasmids were performed after digesting both plasmid and the shuffled fragments with BamHI and XhoI. To clone the fragments as fusions to the Zif268 DBD, a Zif268 Bbs I "stuffer" plasmid was constructed and treated with the *Pfu* enzyme in order to create a cloning site. Shuffled fragments were processed in a similar manner to generate a complementary overhang for the plasmid cloning site. These fragments were then ligated into the Zif268 Bbs I "stuffer" plasmid.

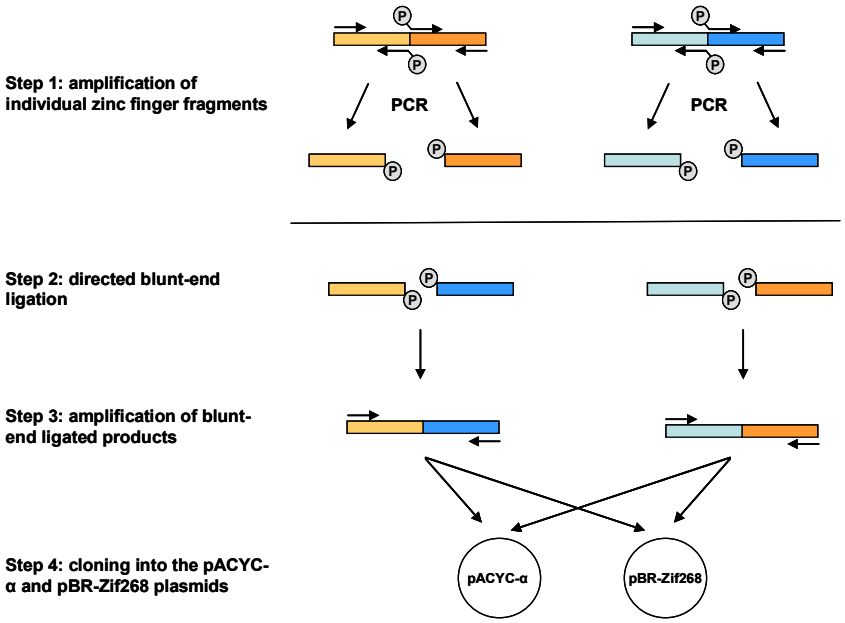

Figure 2.2 Schematic overview of library construction. Light and dark colored boxes represent N-terminal and C-terminal zinc finger, respectively. Different colors (orange and blue) depict domains derived from different proteins. Grey circles labeled with "P" represent phosphate groups at the 5' end of the oligonucleotides and PCR fragments. White circles depict plasmids. See text for details.

2.2.2.7 Primer phosphorylation and PCR

To perform blunt end ligations, the internal primer used to amplify individual ZFs from various DZF domains were phosphorylated at their 5' end using the following conditions for a 50 µl reaction:

Reagent	Final concentration	Volume
Primer (50 pmol/µl)	6 µM	6 µl
10 x T4 Polynucleotide Kinase Reaction buffer (NEB)	1 x	5 µl
10 mM dATP	1mM	5 µl
T4 Polynucleotide Kinase (PNK) (NEB)	10 Units/reaction	1 µl
Sterile dH_2O		33 µl

The reaction was incubated at 37°C for 30 min. The PNK was inactivated by heating the reaction at 65°C for 20 min. The various phosphorylated primers were then used to amplify the individual N-terminal and the C-terminal ZF fragments using standard PCR. The resulting products were purified using a standard QIAgen PCR purification kit as described in section 2.1.2.9 by eluting the DNA with 50 µl of 0.1 x EB buffer.

2.2.2.8 Blunt end ligation and end PCR

Equal volumes of N-terminal and C-terminal encoding PCR fragments were blunt end ligated using the following reaction conditions for a 30 µl reaction:

Reagent	Final concentration	Volume
Various PCR products	Variable	2 µl of each PCR
10 x T4 DNA Ligase reaction buffer with 10 mM dATP (NEB)	1 x	3 µl
T4 DNA Ligase Kinase (NEB)	4000 Units/reaction	2 µl
Sterile dH_2O		Add up to 30 µl

Reaction were incubated at RT (RT) for 2 hr and the resulting product were purified as described in section 2.1.2.9 by eluting the DNA with 50 µl of 0.1 x EB buffer. To amplify all products of the blunt end ligations, PCR was performed using specific external oligonucleotides. Aliquots of the resulting PCR products were resolved on a 5 % SDS PAGE gel and the corresponding bands were excised as a pool and gel-purified as described in section 2.1.2.10. The amount of gel-purified PCR products was quantified using a Beckman Coulter DU 640 Spectrophotometer by measuring absorbance at 260 nm.

2.2.2.9 Preparing plasmids and fragments for the Zif268 library

To clone each shuffled pool of ZFs into the pBR-UV5-Zif268 phagemid, a BbsI "stuffer" plasmid (see A1) containing two BbsI restriction sites was digested with Bbs I and purified as described in section 2.1.2.9. To clone the shuffled pool of ZFs into the Bbs I "stuffer" plasmid, we took advantage of the 3'- to 5' exonuclease activity of the *Pfu* polymerase which can be used to remove nucleotides from the 3' end of the DNA. In the presence of nucleotides, the polymerase activity will counter the exonuclease activity. However, if only one of the four nucleotides is present, the enzyme will continue degrading the DNA until it comes to a position where it can fill in the provided nucleotide, thereby producing a specific overhang. Thus, the exonuclease activity of the *Pfu* polymerase was used to create an overhang in the digested stuffer and a complementary second overhang in the fragment pool.
The following reaction conditions were applied to *Pfu* treated plasmids and fragments in a volume of 80 µl.

Plasmid DNA:

Reagent	Final amount / concentration	Volume
Plasmid "stuffer" DNA	800 fmol (2 µg)	variable
10 x *Pfu* buffer (Stratagene)	1 x	8 µl
10 mM dCTP	1 mM	8 µl
Pfu DNA polymerase (Stratagene)	8 Units/reaction	3.2 µl
Sterile dH_2O		Add up to 80 µl

Fragment DNA:

Reagent	Final amount / concentration	Volume
Pool of DNA fragments	800 fmol (0.1 µg)	variable
10 x *Pfu* buffer	1 x	8 µl
10 mM dGTP	1 mM	8 µl
Pfu DNA polymerase	8 Units/reaction	3.2 µl
Sterile dH_2O		Add up to 80 µl

The reactions were incubated at 72°C for 15 min and then at 4°C for 2 min. The resulting products were purified using a standard QIAgen PCR purification kit as described in section 2.1.2.9 and the DNA was eluted in 20 µl of sterile dH_2O

2.2.2.10 Ligations

Ligations for all libraries (pACYC-α and pBR-UV5-Zif268 libraries) were performed using the following reaction conditions for a 20 µl reaction:

Reagent	Final amount / concentration	Volume
Plasmid DNA	0.2 µg	variable
DNA fragment pool	0.05 µg	variable
10 x T4 DNA ligase buffer	1 x	2 µl
T4 DNA Ligase	4000 Units/reaction	2 µl
Sterile dH_2O		Add up to 20 µl

The Ligation reactions were incubated at 16°C ON. Following the ON incubation, the ligations were ethanol precipitated as described in section 2.1.2.8 and resuspended in 7 µl of sterile dH_2O.

2.2.2.11 Electroporation

Library ligations were used to electroporate various bacterial strains. The pBR-Zif268 ligations were transformed into the XL-1 Blue strain which possess an F factor bearing a Tet

resistance gene. pACYC-alpha ligations were introduced into the RP45 reporter strain harboring the Zif268 binding site. To do this, 70 µl of cells were added to the ligations and pulsed at 1.75 kV for a few milliseconds using a Biorad Gene pulser II. The pulsed cells were immediately transferred to 1 ml of SOC and incubated at 37°C for 1 hr on a roller wheel.

2.2.2.12 Amplification of the pBR-Zif268 library

The amplification of the pBR-Zif268 library was performed as previously described with a few minor differences (Giesecke and Joung, 2005). Following the 1 hr recover step, a small aliquot of cells were serially diluted 10^{-1} to 10^{-6} in 2 x YT. 5 µl of each dilution was spotted three times on LB/T/C (to assess number of transformants) and LB/C/K (to assess helper phage contamination). The plates were incubated at 37°C for 16-18 hr and the resulting titer representing the pre-amplification library size was assessed on the next day.

The reminder of the transformed cells was used to inoculate 9 ml 2 x YT supplemented with Carb (selects for presence of the library phagemids) and Tet (selects for the presence of the F factor). Following 2 hr growth at 37°C with shaking, a small aliquot of the cells was serially diluted, spotted and incubated as described above. The resulting titer of these plates representing the post-amplified library size was assessed on the next day. The remaining amplified cells were pelleted by centrifugation at 4°C, 4000 rpm for 30 min. The cell pellet was resuspended in 2 ml of 2 x YT containing 15 % glycerol, divided into 4 aliquots, frozen in an dry ice/ethanol bath and stored at -80°C.

2.2.2.13 Conversion of the pBR-Zif268 library into infectious phage particles

The conversion of the pBR-Zif268 library into phage was performed as previously described with minor differences (Giesecke and Joung, 2005). Briefly, 2 aliquots of the amplified cells were thawed and used to inoculate 10 ml 2 x YT supplemented with Carb and Tet. Following 1.5 hr incubation at 37°C with slow (125 rpm) shaking, the cells were infected with M13KØ7 helper phage at a multiplicity of infection (moi) of > 100:1 for 30 min at RT. The infected cells were then added to 90 ml 2 x YT containing Carb and Tet and incubated at 37°C with shaking. After 2 hr a small aliquot of the cells were serially diluted as described above and spotted on LB/T/C and LB/C/K^{70} plates. The plates were incubated at 37°C for 16-18 hr and the resulting titer which represented the infection efficiency was assessed on the next day.

Kan was added to the remaining culture to a final concentration of 70 µg/ml and the culture was grown at 37°C for 18 hr.

On the next day, the cells were filtered through a 0.2-µm PES filter in order to harvest the phage library. The titer of the phage library was determined by calculating the Carb-transduction units (CTUs). To do this, the phage was serially diluted 10^{-1} to 10^{-6} in 2 x YT. 10 µl of each dilution was used to infect 50 µl of an ON culture of RP45 that harbored a plasmid conferring Cam resistance. After 25 min incubation time at RT, 190 µl of 2 x YT was added to the cells and the cultures were left for 2 hr at 37°C without shaking. 5 µl of each infection was spotted three times on LB/CA/K (to assess number of cells) and LB/C/C/K (to assess CTUs) and the plates were incubated at 37°C for 16-18 hr.

2.2.2.14 Preparation of selection strain expressing the pACYC-alpha library

As described in section 2.2.2.11 pACYC-alpha ligations were introduced into the RP45 reporter strain using electroporation. Following the 1 hr recover step after Electroporation, a small aliquot of the cells was diluted by 1:10 in series to 10^{-6} in 2 x YT. 5 µl of each dilution was spotted three times on LB/CA/K (to assess number of transformants). The plates were incubated at 37°C for 16-18 hr and the resulting titer representing the library size was assessed on the next day. The remainder of the transformed cells (~ 800 µl) was aliquoted and stored as a frozen glycerol stock at -80°C.

After determining the library size, ~ 10^3 cells were plated on LB/CA/K and the plates were incubated at 37°C for 16-18 hr. On the next day, the cells were harvested by adding 3 ml prewarmed NM medium and a dozen sterile glass beads to each plate. The plates were gently agitated and the resuspended cells were transferred into a sterile glass tube. The resulting selection strain (containing the pACYC-alpha library) was used to inoculate 20 ml of NM medium containing Cam, Kan and IPTG[50] by diluting them by 1:1000–1:10000. The cells were incubated at 37°C with gently shaking (125 rpm) for 16-18 hr. The remaining cells were aliquoted and stored as a frozen glycerol stock at -80°C.

2.2.2.15 Introduction of the pBR-Zif268 library into the selection strain

1 ml of saturated ON cultures (consisting of ~ 1.4 x 10^9 cells; OD_{600} > 2.0; number of bacteria were calculated assuming that an OD_{600} value of 1.0 is equivalent to 7 x 10^8 cells) which harbored the pACYC-alpha-library were infected with ~1.25 x 10^8 CTU of phage

containing the phagemid pBR-Zif268 library. Following a 30 min incubation time at RT, the infected cells were added to 4 ml of prewarmed NM medium with Cam, Kan and IPTG^{50} and the cultures were incubated for 2 hr at 37°C with slow shaking (110 rpm).

2.2.2.16 Performing the Selection

Upon completion of the incubation, a small aliquot of the infected cells were serially diluted by 1:10 in series to 10^{-5} in prewarmed NM medium with Cam, Kan and IPTG^{50}. 5 µl of each dilution was spotted three times on LB/CA/K (to assess total number of cells), LB/CA/C/K (to assess number of transformed cells) and NM/CA/C/K/I^{50} (to assess number of transformed cells capable of growing on His-deficient medium). The plates were incubated at 37°C for 16-18 hr and the resulting titer was assessed on the next day. The remainder of the infected cells was plated on NM selection plates (see section 2.2.1.2) which were incubated at 37°C for 24-48 hr.

2.2.2.17 Plasmid linkage

4–12 colonies from the selection plates were used to inoculate 4 ml LB supplemented with Cam, Carb and Kan and the cultures were grown ON at 37°C with shaking. Plasmid DNA was isolated using the QIAGEN miniprep kit by employing a vacuum manifold essentially as described. After applying the vacuum, the columns were washed 3 x with PB buffer and 3 x with PE buffer. To elute the DNA from the column, 60 µl of prewarmed (60°C) 0.1 x EB buffer was used. To separate the two plasmids, the isolated DNA was diluted by 1:10 and transformed into XL-1 Blue cells. Cells were plated on LB/T/C plates and on LB/CA plates which selects only for the plasmid harboring the respective antibiotic resistance gene. 5 µl of single colony ON cultures were spotted on LB/T/C plates and on LB/CA plates (to assure presence of only one plasmid type). Plasmid DNA was isolated from cultures as described in section 2.1.2.7 and the same pairs of plasmid originally isolated from the selection strain were re-introduced into the bacterial two-hybrid reporter strain to perform β-galactosidase assays as described in section 2.2.2.2.

2.3 Techniques used for protein analysis in mammalian cells

2.3.1 Materials

2.3.1.1 Buffers and solutions

- 2 x HBS: 50 mM HEPES, 280 mM NaCl, 1.5 mM $Na_2HPO_4 \cdot 7H_2O$. Distilled water was added to 500 ml and pH of the solution was adjusted to 7.0.
- Buffer A: 10 mM HEPES (pH 7.8) 1.5 mM $MgCl_2$, 10 mM KCl, 20 uM $ZnCl_2$, 1 mM DTT, 1 mM Benzamidine, 1 tablet of protease inhibitor (Roche) per 5 ml
- Buffer B: 0.3 M HEPES (pH 7.9), 1.4 M KCl, 30 mM $MgCl_2$
- RIPA Buffer (McCarty *et al.*, 2003): 50 mM Tris (pH 7.5), 150 mM NaCl, 1 % NP40, 0.1 % SDS, 0.025 % Sodium Deoxycholate, 1 mM DTT

2.3.1.2 Cells

Flp-In TRex human embryonic kidney (HEK) 293 cells (Invitrogen).

2.3.2 Methods

2.3.2.1 Plating, transient transfection and induction of HEK 293 cells

HEK 293 cells expressing the Tet repressor were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10 % fetal bovine serum (FBS) in a 5 % CO_2 incubator at 37°C. To perform transient transfections, cells were plated in 24-well plates at a density of 150,000 cells per well 24 hours before transfection. Plasmids encoding fusion peptides (see section 3.5) were co-transfected into cells using Lipofectamine 2000 reagent (Invitrogen) according to the manufacturer's protocol using 1 µl of Lipofectamine and 0.5 µg total plasmid DNA per well. The medium was removed and replaced with fresh medium containing the Tet derivate doxycycline (1 µg/µl) ~ 16 h after transfection to induce the expression of the fusion proteins. 24 h after induction the culture medium was harvested to quantify secreted VEGF-A levels.

2.3.2.2 VEGF-A assay

The secreted VEGF-A protein levels in the culture medium were quantified using the human VEGF-A ELISA kit (R&D Systems). After harvesting the culture medium, the samples were centrifuged at 0.3 rpm for 10 min and the supernatant was used directly to perform VEGF-A ELISA assays according to the manufacturer's protocol. A plate washer was used to carry out the washing steps and the absorbance was measured at 450 nm with a correction at 570 nm (relative to a blank) using a microplate reader (Bio-Rad Model 680).

2.3.2.3 WST-1 proliferation assay

VEGF-A values of the individual samples were normalized to the number of viable cells in the culture which were determined using the cell proliferation reagent WST-1 (Roche). After harvesting the culture medium, fresh medium containing ~ 5 % WST-1 reagent was directly added to the cells and the cultures were incubated in a 5 % CO_2 incubator at 37°C for 45–60 min. The medium was harvest and the absorbance at 450 nm with a correction at 655 nm (relative to a blank) was measured using a microplate reader (Bio-Rad Model 680). OD values were scaled by normalizing the OD for each individual sample to the average OD value calculated from all samples. These scaled OD values were used to normalize the VEGF-A values.

2.3.2.4 Western blot analysis

Plasmids encoding the SpI-DZF and p65-DZF fusion proteins (each harboring a N-terminal FLAG tag, see A1 and section 3.5) were co-transfected into HEK293 cells using Lipofectamine 2000, 0.25 μg of the SpI-DZF plasmid and either 0.25 μg or 0.06 μg of the p65-DZF fusion plasmid (0.19 μg of pcDNA5 plasmid DNA was used to keep the total DNA amount of DNA added constant at 0.5 μg). Inductions were performed as described in section 2.3.2.1. 24 h after induction, the culture medium and the cells were harvested. Secreted VEGF-A levels in the culture medium were quantified using ELISA (R&D Systems).

For Western blot analysis of the hybrid protein expression levels, whole cell lysates were obtained as follows. Phosphate buffered saline (PBS) buffer was used to wash the cells twice and detach them from the plates. The samples were centrifuged at 13000 rpm at 4˚C for 2 min, the supernatant was removed and the cells were lysed using 2 x SDS sample buffer. The amount of peptide was normalized by the number of viable cells (determined using WST-1 reagent as described above) and equal amounts of lysate were resolved on a SDS-

polyacrylamide gel. Western blot analysis was performed using the anti-FLAG M2 monoclonal antibody (Sigma). Anti-mouse IgG (Amersham Biosciences) served as a secondary antibody and ECL (Amersham Biosciences) was used for visualization. Band intensities were quantified using a Biorad Fluor-S MultiImager and Quantity One software.

2.3.2.5 Co-immunoprecipitation assay

Coimmunoprecipitation assays were performed essentially as previously described (Cupit *et al.*, 2003; McCarty *et al.*, 2003). Briefly, confluent HEK 293 cells were transfected with pairs of plasmid DNAs (see A1 and section 3.5) encoding the DZF domain either as a fusion to the Ikaros isoform I (Ik I) or to a FLAG epitope tag (10 μg total) using the calcium phosphate method (Ausubel, 1995). Cells were harvested ~ 40 hours after transfection by washing them off the plates with PBS and spinning them at 1000 rpm at 4˚C for 5 min. The cell pellets were resuspended in cold Buffer A and were homogenized using a Dounce tissue homogenizer (Fisher). The nuclei were pelleted by centrifuging the samples at 1000 rpm at 4˚C for 5 min. To clarify the cytoplasmic extract, cold Buffer B was added to the supernatant and the resulting samples were centrifuged at 13000 rpm at 4˚C for 10 min. To perform co-immunoprecipitations, 10-30 μl of clarified cytoplasmic extracts were diluted in 200 μl of cold RIPA buffer, added to 15 μl of RIPA buffer-equilibrated Anti-FLAG M2 Affinity Gel (Sigma) and the resulting samples were incubated with over-head rotation for 1.5 h at 4°C. The beads were collected by centrifugation, washed three times with 1 ml of cold RIPA buffer, mixed with 30 μl 2 x SDS sample buffer, boiled, and centrifuged.

The resulting supernatants (output) and 15 μl of the clarified cytoplasmic extract mixed with 30 μl 2 x SDS sample buffer (input) were resolved by SDS-polyacrylamide gel electrophoresis. Western blot analysis was performed as described in section 2.1.2.12 using the Ikaros M-20 antibody (Santa Cruz Biotechnology) directed against the amino-terminal 53 amino acids of Ikaros. Monoclonal anti-goat IgG (Sigma) was used as secondary antibody.

2.4 Protein overexpression and purification

2.4.1 Materials

2.4.1.1 Buffers and solutions

- Lysis Buffer: 25 mM HEPES (pH 7.9), 100 mM NaCl, 1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride (PMSF), 10 mM dithiothreitol (DTT), 1 mM Benzamidine
- Wash buffer 1: 50 mM Tris (pH 8.0), 100 mM NaCl, 1 mM EDTA, 0.5 % Triton X-100, 1 mM Benzamidine, 20 mM DTT
- Wash buffer 2: 0.5 M Urea (freshly de-ionized with amberlite mixed bed resin), 50 mM HEPES or MES (pH X, X indicates that various pH were applied depending on the peptide, see Chapter 5, Table 5.1), 1 mM Benzamidine, 20 mM DTT
- Solubilization buffer 1: 9 M Urea (freshly de-ionized), 50 mM HEPES or MES (pH X), 150 mM DTT
- Solubilization buffer 2: 4 M Urea (freshly de-ionized), 50 mM HEPES or MES (pH X), 150 mM DTT
- Buffer A for FPLC: 8 M Urea (freshly de-ionized), 50 mM HEPES or MES (pH X), 10 mM DTT
- Buffer B for FPLC: 8 M Urea (freshly de-ionized), 50 mM HEPES or MES (pH X), 10 mM DTT, 1 M NaCl
- Buffer A for HPLC: 99.9 % MilliQ H_2O, 0.1 % TFA. Buffer was filtered through a 0.2 μm Nylon filter prior to use.
- Buffer B for HPLC: 90% acetonitrile (HPLC grade), 9.9% MilliQ H_2O, 0.1% TFA. Buffer was filtered through a 0.2 μm Nylon filter prior to use.

2.4.1.2 Plates

LB/C/CA plates	LB plates with Carb and Cam

2.4.1.3 Bacterial strain

BL21(DE3)pLysS.: *E. coli* B F^- *dcm ompT hsdS*(rB^- mB^-) *gal* λ(DE3) [pLysS Cam^R] (Stratagene).

2.4.2 Methods

2.4.2.1 Protein induction and expression (described for 1 liter bacterial cell culture)

Peptides to be purified were cloned into the pET3a expression plasmid (Novagen, see A1 and Chapter 5) and the resulting plasmids were transformed into the *E. coli* strain BL21 (DE3) pLysS. 50 ml LB supplemented with Carb was inoculated with a 2 ml BL21 starter culture grown from a single colony containing the pET3a plasmid and incubated with shaking at 250 rpm and 37°C until an OD_{600} of ~ 0.5 was reached. The 50 ml culture was transferred to a flask containing 950 ml of LB/Carb and incubated with shaking at 250 rpm and 37°C until OD_{600} reached 0.6. Expression was induced by adding IPTG to a final concentration of 0.4 mM and incubation was continued under the same conditions for 3 hr. The induced cells were harvested by centrifugation at 4000 rpm, 4°C for 30 minutes. The supernatant was decanted and the *E. coli* pellet was stored at -80°C. 200 µl samples of uninduced and induced cultures were collected, harvested by centrifugation at 8000 rpm for 1 min and the cell pellet was resuspended in a buffer containing cold 10 mM Tris (pH 8.0) and 2 x SDS sample buffer. The samples were boiled for 3-5 minutes and the resulting whole cell lysates were normalized for loading based on the OD_{600} at harvest and resolved by SDS-polyacrylamide gel electrophoresis. Coomassie Blue staining was used to visualize the over expressed peptides.

2.4.2.2 Inclusion body isolation

The cells were lysed using a freeze/thaw strategy as follows: The *E. coli* cell pellet consisting of induced cells was resuspended in Lysis buffer. To complete the lysis 0.3 % NP-40 and 0.5 mg/ml of lysozyme were added and the resulting mixture was incubated on ice with occasional swirling for 30 minutes. 6 mM $MgCl_2$, 1 mM $CaCl_2$ and deoxyribonuclease I (DNase I, Worthington) at 2 U/ml were added and the solution was stirred at 4°C for 1 hour to digest the DNA. Inclusion bodies containing the overexpressed peptides were harvested by centrifugation at 8000 rpm in a Sorvall GSA rotor for 15 minutes at 4°C.

To remove contaminants absorbed onto the hydrophobic inclusion bodies, the pellet was washed twice using buffers that contained detergents and chaotropic agents. For pellet wash 1, the inclusion body pellet was resuspended in Wash buffer 1. A DNase I digest was performed in 10 mM $MgCl_2$, 1 mM $CaCl_2$ and DNase I at 2 U/ml for 30 minutes at RT. The inclusion bodies were harvested by centrifugation at 11000 rpm in a GSA rotor for 15 minutes at 4°C. Following centrifugation, the inclusion body pellet was resuspended in Wash buffer 2. To harvest the inclusion bodies, the solution was centrifuged at 11.500 rpm in a SA600 rotor for 15 minutes at 4°C.

To solubilize the peptides, the inclusion body pellet was resuspended in Solubilization buffer 1. Solubilization buffer 2 was added and the resulting solution was incubated at 75°C for 20 min. The solubilized inclusion bodies were harvested by centrifugation at 16000 rpm in a GSA rotor for 15 minutes at 4°C and the supernatant containing the peptides was frozen in liquid nitrogen and stored at -80°C. 20 µl samples were collected at each step and added to 2 x SDS sample buffer for SDS-polyacrylamide gel electrophoresis.

2.4.2.3 Cation exchange chromatography

Cation exchange column:

Cation exchange chromatography was performed using a hand-packed SOURCE 15S column (Pharmacia). The SOURCE matrix is based on rigid polystyrene/divinyl benzene beads and the functional charged group is $–CH_2SO_3^-$ (methyl sulfonate). The maximum pressure for this column is 3.0 MPa and the flow rate limit is 6.0 ml/min. The Column Volume (CV) is 10.053 ml and the typical loading range is 10–25 mg of peptide.

Calculation of the pI value:

To calculate the isoelectric point (pI) value of the peptides used for cation exchange chromatography, the http://workbench.sdsc.edu/ webpage was used. Here, the net charge of the peptide is determined according to the amino acids in the sequence.

Instrument:

The instrument used to perform the cation exchange chromatography was an AKTA fast protein liquid chromatography (FPLC) and the software for analyzing the run was Unicorn version 3.2 (Amersham Pharmacia Biotech).

Run profile:

Before loading the sample, the column was equilibrated with 5 CV Buffer A, 10 CV Buffer B and 5 CV Buffer A at a flow rate of 6 ml/min. The solubilized peptide samples were thawed and loaded in Buffer A using a superloop at a flow rate of 6 ml/min. Usually, 10 ml for trial runs and 40 ml for preparative runs were loaded. The flow rate of the run and the pressure limit was set to 5.0 ml/min and 3.0 MPa, respectively. To differentially elute the bound peptides from the column, the ionic strength of the Buffer was changed by introducing an increasing gradient of Buffer B to the column. To do this, Buffer A and Buffer B were mixed together while changing the ratio linearly. 0 % Buffer B was used as starting conditions for 3 CV to wash out unbound sample. When doing a trial run, a gradient of 0 % to 40 % Buffer B over a volume of 16 CV was used for the first elution segment. The target concentration of the second segment was 100 % Buffer B over a volume of 4 CV. When doing a preparative run, the target concentration of the first segment was 8 % Buffer B over a volume of 2 CV and a gradient of 8 to 20 % Buffer B over a volume of 8 CV was applied for the second elution segment. A third elution segment was used that aimed for 100 % B over a volume of 6 CV. The run was monitored at a wavelength of 280 nm. Protein peaks were collected and immediately frozen on liquid nitrogen. Following the run, remaining bound substances were removed by washing the column with 100 % Buffer B for 1CV.

2.4.2.4 Reverse Phase chromatography

RP column:

Reverse phase chromatography was performed on a Vydac C4 reverse-phased preparative column (Grace Vydac). The hydrophobic (reversed) phase is attached to the silica consisting of butyl aliphatic groups by polyfunctional chlorobutylsilanes which results in a cross-linking or polymerization of the hydrophobic phase. The typical flow rate for this column is 10–30 ml/min and the maximum pressure is 1000 psi. The CV is 95 ml and the loading range is 5–200 mg of peptide.

Instrument:

Reverse phase chromatography was performed on a Beckman High Performance Liquid Chromatography (HPLC) and the software for analyzing the run was Karat32, 7.0.

Sample preparation:
Captured peptides from the FPLC run were thawed and acidified with a 10 % TFA solution to pH 2-3. 0-6 range pH paper strips were used to determine the pH of the peptide sample. Before applying it to the column, the sample was filtered through a 0.2 μm Nylon filter.

Run profile:
Prior to the run, the column was equilibrated at a flow rate of 5 ml/min with 50 ml of 100 % Buffer B and 100 ml of 100 % Buffer A until the monitored baseline was stable. The sample was applied manually to the column. For trial runs, 0 % Buffer B was run through the column for 10 min to wash out unbound sample. Conditions were then altered so that the bound peptides were eluted differentially using a Buffer B gradient. To do this, a 2 %/min gradient from 0 %-90 % Buffer B was chosen for the first elution segment. The target concentration of the second elution segment was 100 % B which was reached using a 1 %/min gradient. For a preparative run, the starting conditions were set on 0 % Buffer B for 15 min. A 2.8 %/min gradient from 0 %-50 % Buffer B was then applied, followed by a 1.6 %/min gradient from 50 %-90 % B and a 5 %/min gradient from 90 %-100 % Buffer B. The run was monitored at a wavelength of 210 nm and eluting peaks were collected and immediately frozen on liquid nitrogen. After the run, the column was washed with at least 100 ml Buffer B.

2.4.2.5 Storage of purified samples
Captured peptide from the HPLC run was lyophilized for 2–3 days and the final dried product was stored in an anaerobic chamber (Coy Laboratory Products) where the oxygen content was kept below 1 part per million.

2.4.2.6 Refolding
To refold the purified peptide (in an anaerobic chamber), the dried sample was carefully dissolved in water. Trial run samples were dissolved in about 50–100 μl H_2O and preparative run samples were dissolved in about 500 μl H_2O. The concentration of the peptide sample was calculated using Beer's law ($A = \varepsilon\ c\ l$) where ε corresponds to the extinction coefficient, c represents the concentration of the sample, l is the path length of the cuvette and A corresponds to the absorbance of the sample. To determine the absorbance of the sample, the OD_{210} and OD_{280} were assessed. The extinction factor for OD_{280} was calculated using the

http://workbench.sdsc.edu/ webpage. The extinction factor for OD_{210} is 2.2 x 10^{-2} ml x μl^{-1}. The path length for the cuvette was 1 cm. To refold the peptide, 1.5 molar equivalents of $CoCl_2$ or $ZnCl_2$ were added to the purified peptides and the pH was adjusted by slowly adding refolding buffer in two steps. In general the final volume of a folding reaction was 12.5 µl.

2.5 Techniques used for protein analysis in *Drosophila melanogaster*

2.5.1 Materials

2.5.1.1 Buffers and solutions

- Extraction buffer: 20 mM HEPES (pH 7.5), 100 mM KCl, 5 % glycerol, 0.05 % NP40, 1 mM DTT, 1 x complete protease inhibitor (Roche).
- 10 x PBS: 2.83 g of $NaH_2PO_4 \cdot H_2O$, 13.74 g of $Na_2HPO_4 \cdot 2H_2O$ and 90.0 g of NaCl. Distilled water was added to 1 liter.
- 0.2 % PBT: 100 ml 10 x PBS, 2 ml Triton X-100. Distilled water was added to 1 liter.

2.5.1.2 8 % Tris-glycine SDS-Polyacrylamide Gel

Reagent	Amount for 40 ml
H_2O	18.5 ml
30 % Acrylamide mix	10.7 ml
1.5 M Tris (pH 8.8)	10.0 ml
10 % SDS	0.4 ml
10% APS	0.4 ml
TEMED	0.024 ml

2.5.1.3 Fly stocks

Stock	Description	Reference
y w	Harbors mutation in white and yellow gene.	Lindsley and Zimm, 1992
Ki, Δ2-3	Stock harboring dominant markers and transposase gene.	Lindsley and Zimm, 1992
y w; Bl/CyO; +/+	Stock harboring dominant markers and balancer chromosome.	Lindsley and Zimm, 1992
y w; +/+; H/TM3	Stock harboring dominant markers and balancer chromosome.	Lindsley and Zimm, 1992
gmr-gal4	Eye-specific driver line that drives expression in the eye in all cells behind the morphogenetic furrow.	Freeman, 1996
engrailed-gal4	Neuroblast specific driver line.	Tabata *et al.*, 1995

2.5.2 Methods

2.5.2.1 Maintaining the flies

Flies were raised on a sucrose/cornmeal/yeast medium supplemented with the mold inhibitor Tegosept. Cultures were maintained in 12 hr:12 hr light-dark cycles at 25°C and 70 % relative humidity.

2.5.2.2 Generation of germline transformants

Stable transgenic flies were essentially generated as previously described using microinjections (Rubin and Spradling, 1982). 0.5-1.0 μg of midi plasmid DNA was used to inject dechorionized embryos collected from a cross between y w and Δ2-3 flies. The locations of the P-element insertions were determined genetically by performing standard mapping crosses of transformed flies to strains containing dominantly marked second and third chromosomal balancer chromosomes (Greenspan, 1997).

2.5.2.3 Preparation of fly head extracts and Western blot analysis

For each crossing, 40 heads isolated from frozen flies were placed in Eppendorf tubes and homogenized in 1 volume of Extraction buffer using an electronic pestle. The homogenates were centrifuged at full speed for 10 min and the clarified supernatant was removed to a new tube. Head extracts were mixed with 2 x SDS loading buffer, boiled and resolved by electrophoresis on a 8 % polyacrylamid SDS gel. Following electrophoresis, gels were

blotted onto a nitrocellulose membrane for 1 hr at 0.5 A using a semi dry blotting apparatus (Enprotech). Blocking, antibody incubation and washing were done as described in section 2.1.2.12 using either anti-c-myc (Roche) or anti-HA (Covance) as primary antibodies and anti-mouse IgG (Amersham Biosciences) as the secondary antibody.

2.5.2.4 Immunohistochemistry

ON egg collections at 29˚C were dechorionized and stained essentially as previously described (Nose *et al.*, 1992; reviewed in Patel, 1994). Briefly, embryos were dechorionized using 7.5 % bleach and fixed in fixative for ~ 20 min. To remove the vitelline membran the embryos were vortexed and washed several times in 100 % methanol. Methanol was subsequently removed by rinsing the embryos 5 times and washing them 3 times for 15 min in 1 x PBT.

The following primary antibodies were used to stain the embryos: guinea pig anti-Hunchback (1:1000, J. Urban), rat anti-Zfh-2 (1:200, J. Urban) and rabbit anti-Eve (1:5000, J. Urban). Primary antibody incubation was performed in 1 x PBT containing 0.03 % NaAzide ON at 4˚C. After rinsing and washing the embryos several times in 1 x PBT, the secondary antibodies, anti-guinea pig Cy5, anti-rat Cy3 and anti-rabbit-FITC (Dianova) were applied at a 1:250 dilution in 1 x PBT for 1,5 hr. The embryos were rinsed and washed again several times in 1 x PBT. Flat preparations of embryos were mounted in Vectashield mounting medium (Vector Laboratories) and analyzed with a confocal laser scanning microscope (Leica TCS SPII). Scanning images were processed with Adobe Photoshop and show projections of multiple focal planes.

Chapter 3. Synthetic protein-protein interaction domains created by shuffling C2H2 ZFs.

3.1 Introduction

The ability to construct complex synthetic cellular networks provides a strategy for biological engineering. Such artificial systems have applications in basic science for understanding natural phenomena as well as in biotechnology or medicine. However, in order to built these synthetic systems one requires a large set of macromolecular components including DNA-binding proteins and protein interaction domains. These "parts" have to be modular and must act independently without disturbing other cellular functions (reviewed in Sprinzak and Elowitz, 2005; Endy, 2005). Natural regulatory networks in living cells extensively use C2H2 ZFs as such "parts" and the C2H2 ZF motif has emerged as the most abundant molecular recognition domain in the human genome (Pellegrino and Berg, 1991; Venter *et al.*, 2001; Lander *et al.,* 2001). While C2H2 ZFs are best known for their ability to bind to DNA, they can also interact with RNA and other proteins, demonstrating their versatility. Because of its prevalence and functional diversity this motif was hypothesized to be a well suited scaffold for constructing artificial DNA-binding domains (DBD) with desired specificities. The simple modular structure of DNA-binding C2H2 ZFs, which use individual fingers to bind to individual "subsites" in the DNA sequence, has proven to be advantageous and prompted researchers to design single C2H2 ZFs capable of binding to desired subsites. Synthetic C2H2 ZFs and/or naturally occurring C2H2 ZFs can then be "mixed and matched" to create DBDs that possess novel binding specificities. These artificial DBDs can be fused to regulatory or effector domains for a wide variety of applications in biological research, systems biology, synthetic biology, and gene therapy (Klug, 1999; reviewed in Falke and Juliano, 2003; reviewed in Jamieson *et al.*, 2003, reviewed in Lee *et al.*, 2003; Blancaford *et al.*, 2004, reviewed in Jantz *et al.*, 2004). Expanding the current strategies for designing desired interaction specificities using these domains is still the focus of on-going research.

Although C2H2 ZFs can also interact with a variety of different protein partners, little is understood about the versatility of the C2H2 zinc finger domain for mediating protein-protein interactions and general strategies for designing novel interaction specificities have not been applied to construct artificial protein-protein interactions. To address these deficits, we

sought to investigate whether individual protein-interacting C2H2 ZFs units can be used to construct finger "arrays" with novel protein binding specificities by shuffling them in a manner similar to the DNA-binding C2H2 ZFs. For these studies, we chose to use C2H2 ZFs from the dimerisation zinc finger (DZF) domains of various proteins including members of the mammalian Ikaros family of transcription factors, the *Drosophila melanogaster* Hunchback protein, and the human TRPS1 protein. DZF domains are defined as two C2H2 ZFs connected by a short linker (Figure 3.1) that are sufficient to mediate homo- and hetero-typic interactions (see section 1.3.4 for details). Previously, a functional chimeric protein containing portions of the human Ikaros and *Drosophila* Hunchback DZFs has been described. This synthetic hybrid DZF can mediate homo-dimerization but does not interact with either the wild-type Ikaros or wild-type Hunchback DZF domain (McCarty *et al.*, 2003; see section 1.3.4.3). The ability of this hybrid protein to homodimerize together with the observation that Ikaros and Hunchback do not interact with each other suggested a "parallel" interaction mode with the N-terminal finger of one DZF contacting the N-terminal finger of the other DZF. In addition, since this protein can not interact with the wild-type DZFs, this result suggested that shuffling of DZF-derived C2H2 ZF domains can yield synthetic DZFs exhibiting novel protein-binding specificities.

This chapter describes the construction of synthetic DZF domains with novel protein-protein interaction specificities that were used to design artificial transcriptional activators which are functional in human cells. By linking these domains together to create synthetic arrays consisting of four fingers, we were also able to design extended protein-interaction surfaces. Analysis of these two and four finger arrays led to the proposal of an alternative anti-parallel interaction mode for DZF domains.

3.2 Analysis of DZF domains using the bacterial two-hybrid system

3.2.1 Sequence comparison of different DZF domains

Amino acid alignments of the human Ikaros family members Ikaros, Helios, Aiolos, Eos and Pegasus (Figure 3.1) demonstrate a high degree of sequence identity in the DZF domain. Ikaros, Eos, Helios and Aiolos are almost identical throughout the DZF domain whereas

Pegasus exhibits regions with a few distinct amino acids. Including the DZF domain of human TRPS1 in the alignment shows that it exhibits a considerable sequence homology with the Ikaros family members. Interestingly, the C-terminal zinc finger organization found in the various Ikaros like proteins also exists within the Hunchback protein of *Drosophila melanogaster*. This suggests that Ikaros evolved from a more ancient family of transcription factors and that additional Hunchback related proteins may exist in other insects. To identify novel DZF domains we searched the genome database for potential DZF domains in proteins from different organism using either the Ikaros or Hunchback DZF domain as a template. We found several Ikaros-like DZF domains from various vertebrate species which have already been shown to be highly conserved (Cupit *et al.*, 2003; Haire *et al.*, 2000). In addition, we were able to pull out several novel potential DZF domains present in various Hunchback proteins of different invertebrate species. Figure 3.1 shows some of the amino acid sequences we found and demonstrates sequence homology with the Hunchback DZF domain from *Drosophila*.

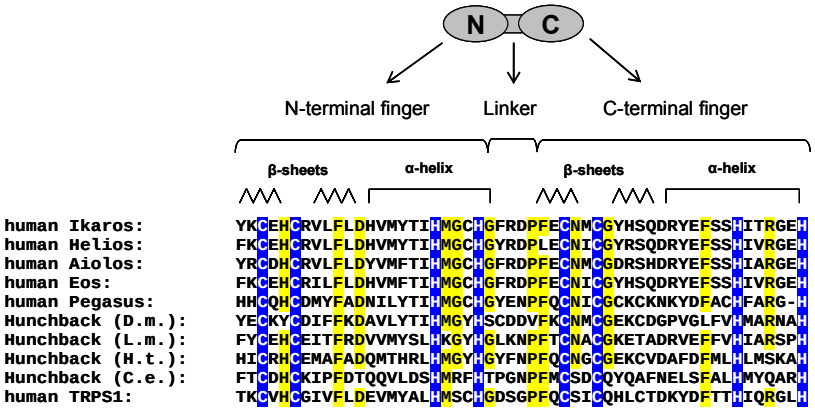

Figure 3.1 Alignment of DZF domains from various transcription factors. The top of the Figure shows a schematic of a DZF domain represented as double ovals that are connected by a bar. The bottom of the Figure presents an amino acid sequence alignment of DZFs from proteins indicated to the left. Organisms from which these proteins derive are also indicated. *D.m. = Drosophila melanogaster, L.m. = Locusta migratoria, H.t. = Helobdella triserialis* and *C.e. = Caenorhabditis elegans*. Secondary structures as defined for DNA binding C2H2 ZFs are shown on top of the amino acid sequence. Conserved cysteines and histidines are highlighted in blue. Positions showing 80% or greater conservation among the ten DZF domains are highlighted in yellow. This Figure was taken from Giesecke *et al.,* (2006).

Thus, we were interested in testing if these novel domains also support dimerization. The DZF domains we chose for this analysis are human Ikaros, Eos, Helios, Aiolos, Pegasus, TRPS1 and Hunchback from *Drosophila melanogaster* (*D.m.*), *Locusta migratoria* (*L.m.*) (Patel *et al.*, 2001), *Helobdella triserialis* (*H.t.*) (Savage and Shankland, 1996) and

Caenorhabditis elegans (*C.e.*) (Fay *et al.*, 1999). Each domain possesses different amino acid sequences and we were therefore interested in determining which of these domains can mediate homo- and/or heterotypic interactions.

3.2.2 The bacterial two-hybrid system

In analogy to the Yeast two-hybrid (Y2H) system, a Bacterial two-hybrid (B2H) system was recently developed as a convenient method to identify and analyze protein-protein interactions (Dove *et al.*, 1997; for comprehensive reviews see: Hu, *et al.*, 2000; Ladant and Karimova, 2000; Joung, 2001; Hu, 2001; Dove and Hochschild, 2004). This system is based on transcriptional activation of a reporter gene mediated by a protein-protein (or protein-DNA interaction).

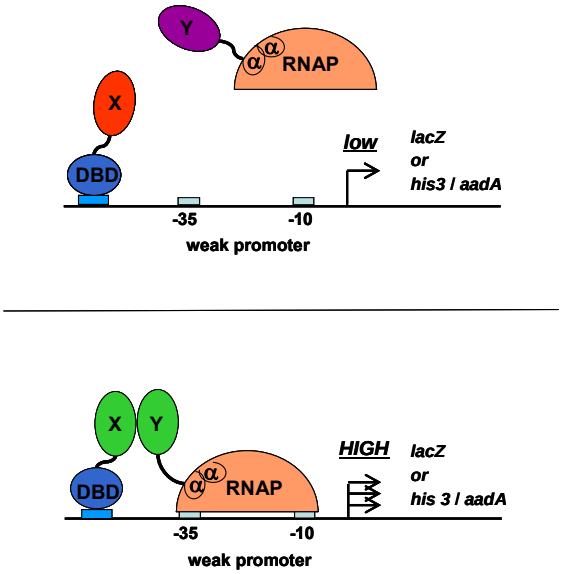

Figure 3.2 Schematic diagram of the B2H system. Proteins to be analyzed are represented by X and Y. Protein X is fused to a DNA binding domain (DBD) that binds to its respective DNA binding site upstream of the test promoter. Protein Y is fused to the alpha-subunit of the RNA polymerase (RNAP). In the absence of an interaction between protein X and Y, the expression of the reporter genes (*lacZ* or *his3/aadA*) is low (upper panel). However, if Protein X and Y do interact, RNAP is recruited to the weak promoter which results in the activation of the reporter gene (lower panel). This Figure was taken and adapted from Giesecke and Joung, (2005).

As shown in Figure 3.2, transcriptional activation can be achieved by recruiting the RNA polymerase (RNAP) to the promoter of a reporter gene(s) via the interaction between two proteins of interest. To accomplish this, two hybrid proteins are co-expressed in an *E. coli* cell: Protein X is linked to a DNA-binding domain (DBD), which binds to an engineered DNA-binding site (DBS) in the promoter region of the reporter gene and protein Y is connected to a subunit (e.g. the alpha subunit) of the RNAP. If protein X and Y interact, RNAP is recruited to the promoter, resulting in activation of the reporter gene(s).

The B2H system can be used as a reporter system for analyzing potential or known interactions at a molecular level by utilizing the *lacZ* gene (encoding β-galactosidase) as a reporter which can be easily quantified by performing β-galactosidase assays. Furthermore, the B2H system can also be used as a selection system for identifying interacting proteins from recombinant DNA libraries. Here, the reporter genes used are the yeast *his3* and the bacterial *aadA* genes. By using *E. coli* cells that are auxotrophic for histidine one can select for expression of HIS3 by growing these cells in the absence of this amino acid. The *aadA* gene confers resistance to streptomycin and is expressed co-cistronically with *his3*. Thus, by growing the *E. coli* cells in the presence of streptomycin the stringency of the selection can be increased (Joung *et al.,* 2000; reviewed in Joung, 2001; Hurt *et al.*, 2003; Giesecke and Joung, 2005). The crucial advantage of this bacterial based system compared with the yeast two hybrid method is the capability to screen very large libraries (>10^9) of interaction candidates (Joung *et al.*, 2000). Additional advantages address its usability due to the fast growth and high transformation efficiency obtained in bacteria and the lack of requirement for nuclear localization of the hybrid proteins involved (for reviews see e.g. Hu, *et al.*, 2000; Ladant and Karimova, 2000; Joung, 2001; Hu, 2001).

3.2.3 DZF interactions can be detected using the B2H system

To initially test whether the B2H system can be applied for studying DZF domain interactions, plasmids encoding fusion proteins for the B2H system setup were constructed. To do this, the Ikaros DZF domain was fused to the Zif268 DNA binding domain and to the RNAP α-subunit were constructed (Figure 3.3A, experiments performed by R. Fang). To test for interaction, plasmids encoding these hybrid proteins were transformed into the B2H reporter strain and β-galactosidase assays were performed (as described in section 2.2.2.2). (Note that this reporter strain bears the Zif 268 binding site at position -65 relative to the promoter of the reporter gene. The transcription startpoint of the reporter gene is defined as position +1 and the DBS is therefore placed 65 nucleotides upstream of the startpoint, see also Appendix A3). It was found that transcription of the *lacZ* reporter gene is elevated in cells expressing both the IkDZF-Zif268 and the RNAP α-IkDZF hybrid proteins suggesting that the Ikaros DZF domain mediates interactions between these proteins (Figure 3.3B, R. Fang). In contrast, control experiments expressing either the Zif268-hybrid or the α-hybrid protein alone did not show increased activation of *lacZ* indicating that both hybrid DZF

proteins must be present to activate the expression of the *lacZ* reporter gene (data not shown). Two mutations, D18Q and R47P (note that mutations are abbreviated following the pattern: wild-type amino acid, residue position, mutant amino acid), have been shown to disrupt homodimerization of the Ikaros DZF domain (McCarty *et al.*, 2003) and were subsequently introduced into the two fusion proteins. Both mutations decreased *lacZ* expression indicating that interaction between the Ikaros DZFs mediates activation of *lacZ* (Figure 3.3B, Rui Fang). These preliminary data demonstrated that DZF domain interactions can be studied in the B2H reporter system.

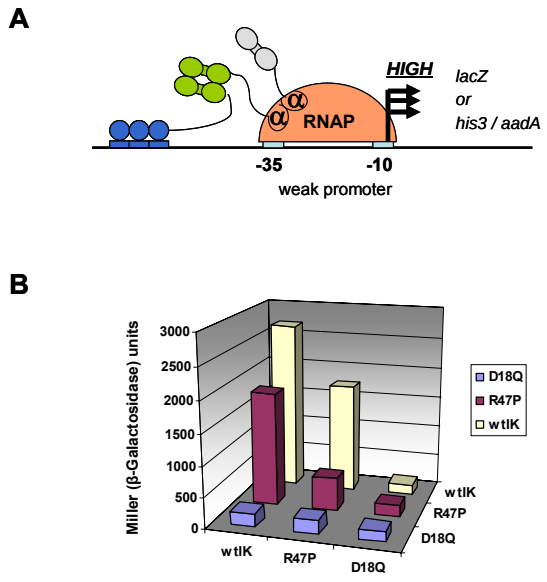

Figure 3.3 Analysis of dimerization mediated by the Ikaros DZF domain in the B2H system. (A) Adaptation of the B2H system to study DZF domain mediated interactions. The DBD of Zif268 (three blue circles) was applied to tether DZF1 to the weak *lacZ* promoter containing a Zif268 DBS (three blue boxes). DZF2 is fused to the RNAP α-subunit and interaction of the two DZF domains (indicated by green double ovals) will recruit the RNAP to the promoter and activate transcription of the reporter gene. Note that although two α-DZF fusions are present which could theoretically homo-dimerize, the DZF1-DZF2 interaction is dominant and can therefore be detected. This Figure was kindly provided by K. Joung **(B)** Pairwise combinations of plasmids encoding the DZF domain from wild-type (wtIk) and mutant (D18Q, R47P) Ikaros fused to both the Zif268 and the RNAP α-subunit were transformed into the B2H reporter strain and *lacZ* expression was measured by performing β-galactosidase assays (work performed by R. Fang).

3.2.4 Interaction specificity profiles of wild-type DZFs determined using the B2H system

The interaction specificities of DZF domains were examined by testing all pairwise interactions of ten different wild-type DZF domains using the B2H system to determine which of these domains can mediate homo- and/or heterotypic interactions. We constructed plasmids encoding fusions of each of these ten DZFs to the Zif268 DBD and to the RNAP α subunit and transformed pairwise plasmid combinations into a B2H reporter strain. β-galactosidase assays were performed to assess the potential interactions of all ten DZFs. The results of this assay confirmed all previously described DZF interactions identified by

biochemical or genetic methods except the homo-dimerization mediated by Aiolos (Figure 3.4A).

A

Zif268 hybrids

RNAP α hybrids	Ik	He	Ai	Eo	Pe	Tr	Hd	Hl	Hh	Hc
Ik	**12.5 (+/- 0.8)**	**7.2 (+/- 0.6)**	**4.8 (+/- 0.4)**	**15.8 (+/- 1.8)**	**6.6 (+/- 0.6)**	0.8 (+/- 0.0)	0.8 (+/- 0.1)	0.9 (+/- 0.1)	0.5 (+/- 0.1)	0.5 (+/- 0.0)
He	**8.3 (+/- 0.4)**	**5.9 (+/- 0.3)**	**3.5 (+/- 0.1)**	**12.9 (+/- 0.7)**	**7.7 (+/- 0.7)**	0.9 (+/- 0.1)	0.7 (+/- 0.2)	0.9 (+/- 0.1)	0.5 (+/- 0.0)	0.5 (+/- 0.0)
Ai	**12.0 (+/- 0.2)**	**5.8 (+/- 0.6)**	1.8 (+/- 0.2)	**14.8 (+/- 0.9)**	**9.0 (+/- 0.7)**	0.8 (+/- 0.1)	1.6 (+/- 0.2)	1.1 (+/- 0.1)	0.7 (+/- 0.1)	0.8 (+/- 0.1)
Eo	**10.7 (+/- 0.6)**	**7.1 (+/- 0.5)**	**4.8 (+/- 0.5)**	**12.7 (+/- 0.8)**	**11.0 (+/- 0.5)**	1.8 (+/- 0.1)	0.9 (+/- 0.0)	0.9 (+/- 0.0)	0.4 (+/- 0.1)	0.5 (+/- 0.1)
Pe	**5.3 (+/- 0.3)**	**4.3 (+/- 0.4)**	1.9 (+/- 0.2)	**7.0 (+/- 0.6)**	**8.2 (+/- 0.5)**	1.3 (+/- 0.1)	**5.3 (+/- 0.5)**	2.5 (+/- 0.1)	0.4 (+/- 0.1)	0.4 (+/- 0.1)
Tr	2.5 (+/- 0.1)	1.8 (+/- 0.1)	1.4 (+/- 0.1)	**3.5 (+/- 0.1)**	2.4 (+/- 0.2)	**3.3 (+/- 0.3)**	1.4 (+/- 0.1)	1.4 (+/- 0.1)	0.9 (+/- 0.1)	1.1 (+/- 0.2)
Hd	0.6 (+/- 0.1)	0.7 (+/- 0.1)	0.8 (+/- 0.0)	0.9 (+/- 0.1)	**13.2 (+/- 0.8)**	0.6 (+/- 0.1)	**11.9 (+/- 1.1)**	**10.6 (+/- 0.6)**	0.6 (+/- 0.0)	0.8 (+/- 0.1)
Hl	0.8 (+/- 0.1)	0.6 (+/- 0.0)	0.9 (+/- 0.0)	0.6 (+/- 0.1)	2.1 (+/- 0.2)	0.6 (+/- 0.0)	**8.8 (+/- 0.4)**	**10.4 (+/- 0.7)**	0.6 (+/- 0.0)	0.6 (+/- 0.1)
Hh	1.6 (+/- 0.1)	1.3 (+/- 0.1)	1.3 (+/- 0.1)	1.6 (+/- 0.1)	0.9 (+/- 0.1)	0.8 (+/- 0.1)	1.3 (+/- 0.1)	1.7 (+/- 0.2)	**5.9 (+/- 0.4)**	1.1 (+/- 0.1)
Hc	0.9 (+/- 0.0)	0.8 (+/- 0.1)	1.1 (+/- 0.1)	0.8 (+/- 0.0)	0.8 (+/- 0.1)	0.6 (+/- 0.1)	1.1 (+/- 0.1)	1.0 (+/- 0.0)	0.8 (+/- 0.0)	1.1 (+/- 0.2)

B

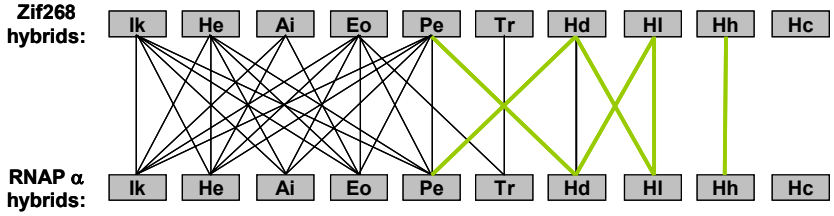

Figure 3.4 Dimerization specificities of wild-type DZFs determined using the B2H system. (A) Pairwise combinations of plasmids encoding ten wild-type DZFs were transformed into the B2H reporter strain and β-galactosidase activity was assessed. Values represent mean-fold activation of three individual experiments. Standard errors of the means are also shown. Interactions resulting in a >2.5-fold activation of *lacZ* expression were defined as positive and are highlighted in bold red text. This cutoff was chosen because the highest fold-activation obtained for a known non-interacting DZF pair (Ikaros and TRPS1; McCarty *et al.*, 2003) was 2.5. Ik = human Ikaros, He = human Helios, Ai = human Aiolos, Eo = human Eos, Pe = human Pegasus, Tr = human TRPS1, Hd = *Drosophila melanogaster* Hunchback, Hl = *Locusta migratoria* Hunchback, Hh = *Helobdella triserialis* Hunchback and Hc = *Caenorhabditis elegans* Hunchback. **(B)** A summary of interaction specificity profiles for the ten DZF domains is shown. Lines indicate significant interactions between two domains as determined in A. Black lines indicate interactions between DZF domains that have also been observed with other methods and green lines indicate novel interactions that have not been previously described. This Figure was taken and adapted from Giesecke *et al.*, (2006).

The reason for this discrepancy is unknown but we suggest that it may be due to a stability problem of the Zif268-Aiolos hybrid protein since this protein generally displays lower activity when tested in combination with other α-hybrid proteins (e.g. Ikaros). However, this further validates the use of the B2H system as a rapid method to assay DZF domain interactions. In addition, this experiment discovered novel homo- and heterotypic interactions

among these DZF domains. Examples for new homotypic interactions are Hunchback *L.m.* and Hunchback *H.t.* which are both able to mediate homodimerization. Interestingly, this result shows that Pegasus can interact with both Ikaros and Hunchback *D.m.* while these two do not interact with each other (Figure 3.4, McCarty *et al.*, 2003). Hunchback *D.m.* on the other hand can mediate interactions with both Pegasus and Hunchback *L.m.* which also do not interact with each other (Figure 3.4). Although these interactions mediated by domains from different species do not occur in nature, they constitute further evidence for the complex interaction specificities of DZF domain mediated dimerization. In addition, these preliminary results suggest that new interaction specificities might be identified using the B2H system.

3.3 Identification of interacting synthetic DZFs using B2H selections

Since the result of our interaction profile experiment clearly demonstrates that DZF domains are very diverse in their interaction specificities, we were interested in exploring whether shuffling C2H2 ZFs from different DZFs might generate synthetic DZFs with novel interaction specificities. This would suggest that protein-interacting ZFs, like their DNA-binding counterparts, can function in a modular fashion. To do this, libraries consisting of combinations of shuffled C2H2 ZFs derived from various wild-type DZFs were constructed and interacting pairs of synthetic DZFs from these libraries were identified using a B2H selection system.

3.3.1 Overview of selections

3.3.1.1 Construction of "shuffled" C2H2 ZF libraries

To create libraries of synthetic DZF domains we shuffled C2H2 ZFs derived from eight wild-type DZFs characterized in the B2H system (note that DZF domains from Aiolos and Helios were not included in this experiment since they are almost 100 % homologous to the Ikaros DZF). Two different shuffling approaches were used to create two pools of variant DZF domains: in one approach the inter-finger linker remained associated with the N-terminal finger whereas in the other approach the linker remained associated with the C-terminal finger. To create these pools of variant shuffled DZF domains, DNA fragments encoding

both the N-terminal C2H2 ZF and the C-terminal C2H2 ZF of defined DZF domains were amplified by PCR. All subdomains were reassembled using blunt end ligation in which the reassembly of non chimeric subdomains (which would result in wild-type proteins) was precluded. Each of the two reassembled pools represents one library and was then separately cloned into two different expression vectors that express the DZFs as fusions to the Zif268 DBD or the RNAP α-subunit. This strategy results in a "set" of four different plasmid libraries that can be subsequently applied to perform selections in the B2H system (Figure 3.5).

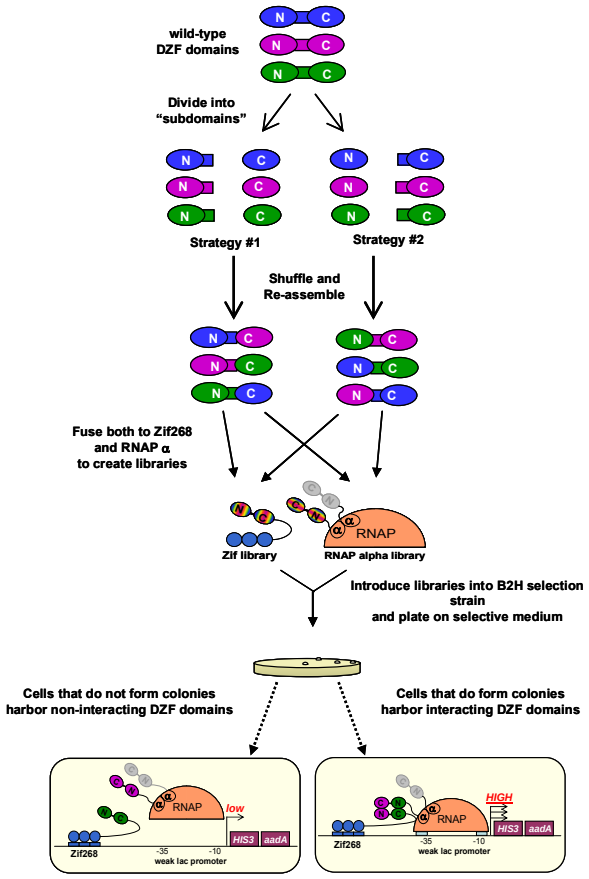

Figure 3.5 Schematic overview of shuffled DZF library construction and B2H selections to identify interacting synthetic DZFs. DZFs are represented as in Figure 3.1. Note that this approach results in a total set of four different plasmid libraries by combining two strategies for shuffling and reassembling the subdomains with the construction of two libraries (one Zif268 and one RNAP α-subunit library). See text for additional details. This Figure was taken from Giesecke *et al.*, (2006).

3.3.1.2 Bacterial two-hybrid selections

Four separate selection experiments were performed in which all possible pairwise combinations of the four libraries were screened for potential interactions. To perform selections, libraries pairs (i.e. one Zif268 library and one RNAP library) were introduced into B2H selection strains harboring the selectable, co-cistronic *his3* and *aadA* genes as reporters (Joung *et al.*, 2000; Hurt *et al.*, 2003). Cells expressing DZF-Zif268 and α-DZF hybrid proteins that interact with each other should activate transcription of the *His3* and *aadA*

genes, thereby permitting them to grow on selective media (Figure 3.5). Cells that were able to form colonies were picked from each selection plate and plasmids encoding the pairs of shuffled DZF domains were isolated. To verify that a survival phenotype is genuinely linked to the plasmid encoding the pair of selected DZF variants, the purified plasmids were transformed into a B2H reporter strain and β-galactosidase assays were performed (phenotype-, plasmid-linkage). Plasmid pairs that resulted in elevated level of β-galactosidase expression were then sequenced to determine the identities of N-terminal ZF, linker and C-terminal ZF present in the interacting pairs. Additional fragment linkage analyses were performed to confirm that activation in the B2H system was linked to specific chimeric DZF fragments isolated from the library. To do this, fragments encoding the different DZF variants were re-cloned into the B2H plasmids and were re-tested for their ability to activate reporter gene expression by performing β-galactosidase assays.

3.3.2 Individual selection experiments

Three different "sets" of libraries (A, B, and C) were constructed by shuffling combinations of C2H2 ZFs derived from various subsets of different wild-type DZFs (Table 3.1). Three independent selections were then performed, each using one of the three library "sets" (A, B, and C). For each of these three selections, all possible pairwise combinations of the four libraries in a "set" were analyzed for potential interactions.

3.3.2.1 Libraries

For constructing library set A we decided to use six subdomains derived from the human Ikaros, human Eos, human Pegasus, Hunchback *D.m.*, Hunchback *L.m.*, and Hunchback *H.t.* DZFs. Various subdomains from DZFs which are all know to interact with each other (e.g. Ikaros and Eos) were used as internal positive controls to test if it is generally possible to create chimeric DZFs that can mediate interaction. The expectation was to re-isolate several shuffled combinations of these domains.

Since the initial characterization of the ten DZF domains showed interesting specificity patterns for Pegasus, Hunchback *D.m.* and Hunchback *L.m.* (Hunchback *D.m.* can interact with both Pegasus and Hunchback *L.m.* while these two do not interact with each other) another library set B was constructed consisting of shuffled combinations of these three

DZFs. We were interested in testing if we could enrich for chimeric proteins containing these subdomains.

To create chimeric DZF domains with additional novel interaction specificities, a third library that only contained DZF domains that do not interact with each other was designed. We reasoned that chimeric proteins consisting of these subdomains would be more likely to possess novel specificities. Therefore DZF domains from human Eos, human TRPS1, Hunchback *D.m.*, Hunchback *H.t.* and Hunchback *C.e.* were applied to create shuffled library set C.

3.3.2.2 Selections

For library set A each pool of DNA fragments encoded 30 different shuffled DZF domains since reassembly of wild-type DZFs was precluded (6 x 6 = 36 - 6). Note that for each library, the number of transformants (10^3-10^4) exceeded the theoretical number of shuffled DZFs by ~100-fold. The potential number of possible pairwise combinations of shuffled DZFs for each selection experiment is 30 x 30 = 900. ~10^6 transformants were plated which outnumbered the theoretical number of potential combinations by at least 1000-fold (Table 3.1). Interestingly, a large number (~1000) of surviving colonies were obtained for each selection experiment.

	Theoretical # of potential DZFs in each library	**Actual # of transformants for each library**	**Theoretical # of combinations for each selection (library x library)**	**Actual # of combinations tested for each selection (library x library)**
Set A: Ik, Eo, Pe, Hd, Hl, Hh	30	10^3-10^4	900	~ 10^6
Set B: Pe, Hd, Hl	6	10^4-10^5	36	~ 5 x 10^6
Set C: Eo, Tr, Hl, Hh, Hc	20	10^3-10^4	400	~ 10^6

Table 3.1 Library sizes and B2H selection statistics. Data in columns 1 and 2 describe the theoretical and actual sizes of each library in "Sets" A, B, and C. Data in columns 3 and 4 describe the theoretical and actual number (#) of combinations tested in each selection performed for "sets" A, B, and C. Each "set" consists of four libraries (two Zif268 fusion libraries and two RNAP alpha-subunit fusion libraries, depicted in Figure 3.5) which are used to perform a corresponding "set" of four selections that tested all pairwise combinations of the four libraries. Abbreviations are as in Figure 3.4. See text for details.

In library set B each pool of DNA fragments encoded six shuffled DZF domains precluding the reformation of the wild-type DZFs (3 x 3 = 9 – 3). The number of transformants for each library was 10^4-10^5 which exceeded the theoretical number by ~1000-fold. We plated ~10^6 transformants, outnumbering the potential number of combinations (36) by at least 10000-fold (Table 3.1). A large number of colonies (~1000) survived the selections.

For library set C each fragment pool contained 20 (5 x 5 = 25 – 5) shuffled DZFs, again precluding the reassembly of wild-type domains. The number of transformants (10^3-10^4) for these libraries outnumbered the theoretical number of shuffled DZFs by about 100-fold. ~10^6 transformants were plated which exceeded the theoretical number of potential combinations (400) by at least 1000-fold (Table 3.1). Interestingly, less surviving colonies (~100) compared to the other selection sets were obtained suggesting that this selection yields only a couple of interacting DZF domains.

3.3.2.3 Plasmid linkage

For all three independent selection sets (A, B and C) 2 to 12 colonies were picked from each of the four selection plates. Plasmids encoding the interacting pairs of DZF variants were then isolated from these colonies.

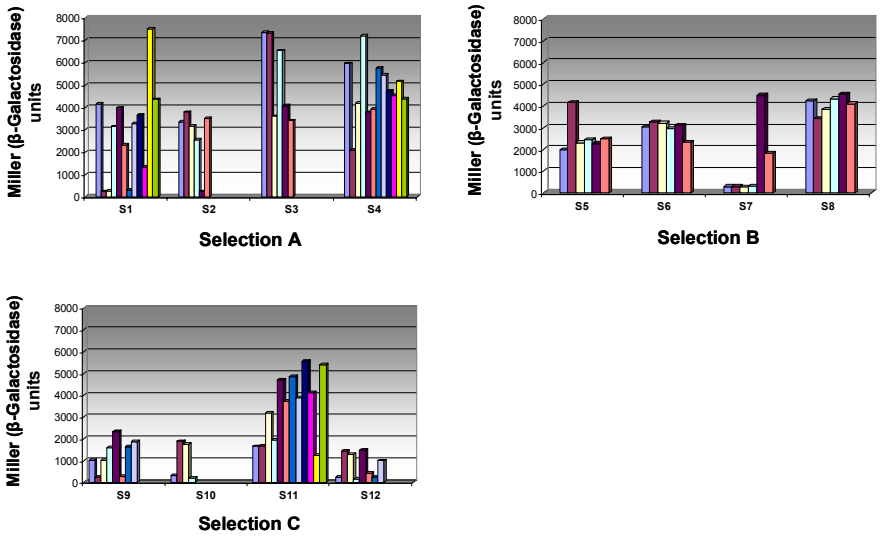

Figure 3.6 Plasmid linkage analysis for selected interacting pairs. Plasmids isolated from colonies that survived the selection were transformed into the B2H reporter strain and β-galactosidase assays were performed. Each graph corresponds to one selection "set" (selection A, B and C) which consists of 4 individual selection experiments (S1-S4, S5-S8, S9-S12). Note that each bar shown for an individual selection in a particular "set" corresponds to one colony picked from the respective selection plate.

For plasmid linkage analysis, each of these pairs of plasmids was transformed into the B2H reporter strain and β-galactosidase assays were performed. The result (Figure 3.6) of these assays demonstrated that for each selection experiment most (average of ~ 85 %) of the selected pairs were able to activate transcription in the B2H system. This confirms that the survival phenotype of the cells on the selection plates is linked to their ability to activate expression of the selection gene. The activation obtained for most of the selected pairs of DZF domains is similar to the activation for the wild-type DZFs. Interestingly, some of the novel DZF variants displayed higher β-galactosidase units than the wild-type proteins, implying that these novel pairs might interact more strongly than wild-type DZF domains (Figure 3.6 and data not shown).

3.3.2.4 Sequencing

Plasmid pairs that resulted in elevated level of β-galactosidase expression were sequenced to determine the identities of the individual C2H2 ZFs and linkers encoded. Sequencing revealed that the selection identified several novel heterodimeric pairs of interacting DZF domains which represent combinations of C2H2 ZFs from the parental DZFs used to construct the libraries (Table 3.2). Many of the interacting pairs contain C2H2 ZFs derived from wild-type DZFs which are known to interact with each other. For example, the majority of selected DZFs from library A are various shuffled combinations of the subdomains from Ikaros, Eos and Pegasus. It seemed likely that these shuffled DZFs would all interact with the parental DZF domains from which they are derived and would therefore not exhibit novel interaction specificities. Similarly, library B only harbored shuffled ZF combinations consisting of subdomains from either Pegasus and Hunchback *D.m.* or from Hunchback *D.m.* and Hunchback *L.m.*. We were not able to pull out chimeric proteins containing ZFs from Pegasus and Hunchback *L.m.* together in a single DZF domain. However, three of the identified pairs from Selection A consisted of DZF domains composed of subdomains from parental DZFs that do not interact with each other. Interestingly, selection C harbored only one pair of interacting DZF variants also composed of C2H2 ZF subdomains from wild-type DZFs which do not interact with each other (Table 3.2). These four pairs are likely to represent DZF domains with potentially novel specificities and were therefore chosen for additional analysis (see below).

	Zif268 hybrids	RNAP α hybrids	Frequency
Library Set A	Ik-Eo-Eo	Eo-Ik-Ik	2
	Ik-Eo-Eo	Eo-Eo-Ik	1
	Ik-Eo-Eo	Ik-Eo-Eo	4
	Ik-Eo-Eo	Ik-Ik-Eo	1
	Ik-Ik-Eo	Eo-Ik-Ik	2
	Ik-Ik-Eo	Eo-Eo-Ik	3
	Ik-Ik-Eo	Ik-Ik-Eo	1
	Ik-Ik-Eo	Ik-Eo-Eo	1
	Ik-Ik-Eo	Pe-Eo-Eo	2
	Ik-Ik-Eo	Pe-Ik-Ik	1
	Ik-Eo-Eo	Pe-Eo-Eo	1
	Ik-Eo-Eo	Eo-Pe-Pe	1
	Pe-Eo-Eo	Ik-Ik-Eo	3
	Pe-Pe-Hd	Hd-Hd-Pe	4
	Pe-Hd-Hd	**Hl-Eo-Eo**	**1**
	Ik-Ik-Hd	**Pe-Pe-Eo**	**1**
	Pe-Pe-Eo	**Eo-Eo-Hd**	**2**
	Ik-Ik-Hd	Pe-Pe-Eo	1
Library Set B	Hl-Hd-Hd	Hl-Hd-Hd	2
	Hd-Hl-Hl	Hd-Hl-Hl	1
	Pe-Hd-Hd	Hl-Hd-Hd	2
	Pe-Pe-Hd	Hl-Hl-Hd	2
	Pe-Pe-Hd	Hd-Pe-Pe	4
	Pe-Hd-Hd	Hd-Hd-Hl	1
	Pe-Pe-Hd	Hd-Hd-Pe	6
Library Set C	**Tr-Eo-Eo**	**Eo-Eo-Hd**	**3**

Table 3.2 Interacting pairs of synthetic shuffled DZF domains identified by genetic selection. Identities of the subdomains present in the synthetic DZF domains are shown. Subdomains are defined as N-terminal ZF, linker, and the C-terminal ZF (shown from left to right). The Selection "set" (A, B and C) from which these pairs were identified as well as the number of times each pair was identified are also indicated. Abbreviations for the subdomains are as defined in the legend to Figure 3.4. Location of the synthetic DZF on the Zif268 or RNAP a-subunit is shown on top of the table. Interacting pairs chosen for further characterization are marked in bold red text.

3.3.2.5 Linkage analysis

To further confirm that activation of reporter gene expression in the B2H system is linked to the presence of various shuffled DZF fragments that were isolated from the different selections, fragment analysis was performed. Thus, fragments encoding selected DZF variants were re-cloned into fresh B2H plasmids and were re-tested for their ability to activate reporter gene expression by performing β-galactosidase assays. It was found that all of the shuffled DZFs tested mediated interaction with their selected partner indicating that the activation phenotype in the B2H system is linked to the specific DZF domain (data not shown).

3.4 Analysis of synthetic DZF domains using the B2H system

3.4.1 Interaction specificities of selected synthetic DZFs

To further analyze the specificity of the synthetic DZF variants, we decided to focus on four interacting pairs we identified in which at least one of the DZF variants contains subdomains derived from wild-type DZFs that do not interact with each other (Table 3.2, highlighted in red; note that these four pairs comprise a total of 6 different DZF domains). The specificity profiles of the DZF domains in these four pairs were examined using the B2H system as a reporter.

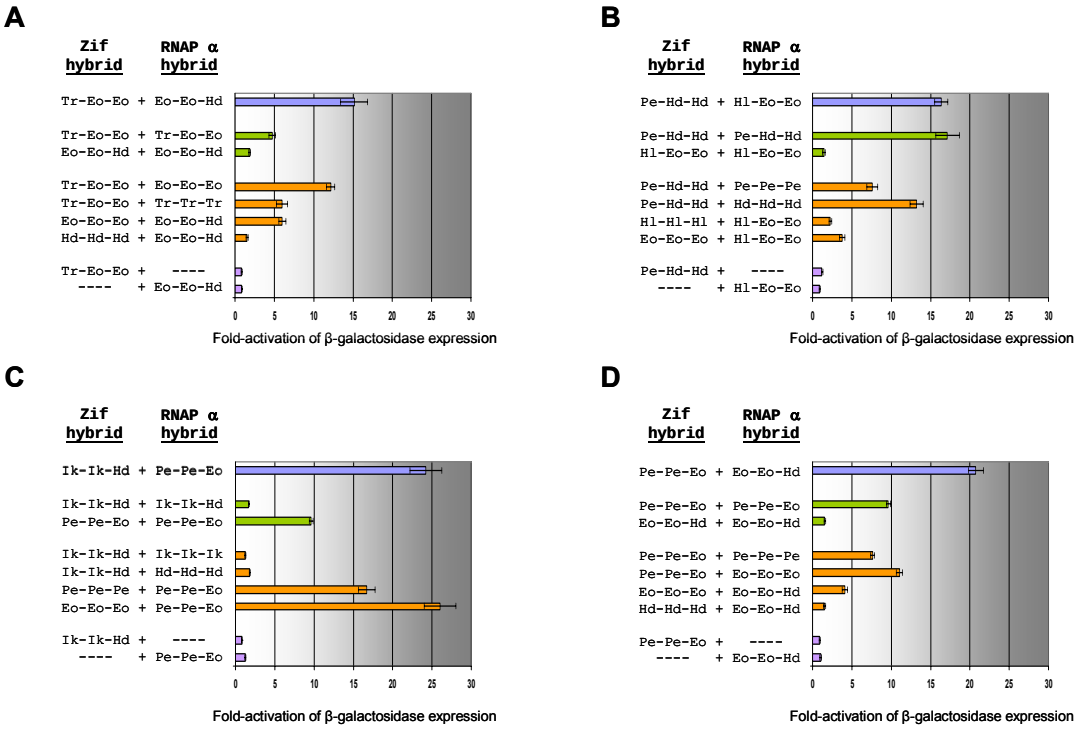

Figure 3.7 Interaction specificities of synthetic DZFs analyzed in the B2H system. DZFs from four pairs (**A, B, C** and **D**, respectively) were tested for hetero- (blue bars) and homo-typic (green bars) interaction as well as their ability to interact with the parental wild-type DZFs from which they were derived (orange bars). Assays were performed using the B2H system. Control experiments expressing combinations of the synthetic DZFs with either the Zif268 or the α-subunit protein with no DZF fused (indicated as '----') are also shown (purple bars). Bars shown represent mean fold-activations of transcription in the B2H system calculated from three independent experiments. Standard error of the mean is indicated by the error bars. Abbreviations for identities of the synthetic DZFs are defined in Figure 3.4.

We assayed the ability of each of the six DZF domains in the various pairs to mediate hetero- and homo-typic interactions and also assessed their interactions with the wild-type DZFs from which they derived. It was found that for all four heterodimeric pairs, one DZF domain in the pair is not able to interact with itself (homotypic interaction) nor does it interact with its parental wild-type DZFs, while the second domain does both interact with itself and with its parental DZFs (Figure 3.7). The latter DZF domain is the one consisting of subdomains from wild-type DZFs which are known to interact with each other. Closer inspection of the activity profile suggest that three of the four DZF pairs tested prefer to hetero-dimerize, although one DZF domain in each pair can still mediate homodimerization. The fourth pair contained one DZF domain (Pe-Hd-Hd) that mediates both homo- and hetero-typic interactions equally well while the other pair (Hl-Eo-Eo) mediated only heterotypic interaction.

Zif268 hybrids

RNAP α hybrids	Tr-Eo-Eo	Eo-Eo-Hd	Pe-Hd-Hd	Hl-Eo-Eo	Pe-Pe-Eo	Ik-Ik-Hd
Tr-Eo-Eo	4.8 (+/- 0.4)					
Eo-Eo-Hd	**15.1 (+/- 1.7)**	1.8 (+/- 0.1)				
Pe-Hd-Hd	1.5 (+/- 0.2)	1.4 (+/- 0.2)	17.1 (+/- 1.5)			
Hl-Eo-Eo	0.9 (+/- 0.1)	1.1 (+/- 0.2)	**16.3 (+/- 0.8)**	1.4 (+/- 0.2)		
Pe-Pe-Eo	**4.6 (+/- 1.2)**	**19.2 (+/- 1.6)**	2.5 (+/- 0.5)	2.8 (+/- 0.4)	9.6 (+/- 0.3)	
Ik-Ik-Hd	**9.5 (+/- 1.4)**	1.6 (+/- 0.2)	2.4 (+/- 0.4)	1.3 (+/- 0.3)	**20.7 (+/- 0.9)**	1.7 (+/- 0.0)

Table 3.3 Analysis of cross-interactions among the six synthetic DZFs in the B2H system. Pairs of plasmids encoding the six synthetic DZFs fused to either the Zif268 DBD or the RNAP α-subunit were co-transformed into the B2H reporter strain harboring the *lacZ* gene and reporter gene expression was assessed by performing β-galactosidase assays. Each combination was tested in triplicates with averages and standard errors of the mean shown. Positive interactions (as defined in Figure 3.4) are marked in colored text. Bold red text indicates hetero-typic interactions between DZFs that were selected with each other and red text depicts homo-typic interactions between these DZFs. Blue bold text indicates cross-reactivity between DZFs that were identified in different selection sets. Identities of subdomains in the synthetic DZFs are abbreviated as in Figure 3.4.

To further analyze the specificity of these six synthetic DZF domains, we tested for potential cross-interactions among them using the B2H system. The result of this experiment is summarized in Table 3.3 and indicates that all six DZF domains interact most strongly with the partners they were selected with, although some DZF domains also exhibit some cross-

reactivity and bind to at least one other synthetic DZF. Note that these DZFs were obtained from different selection sets and could not have been selected in these combinations.
Control experiments expressing either the Zif-hybrid or the α-hybrid protein alone did not show increased β-galactosidase Units (Figure 3.7). Furthermore, the previously described D18Q mutation was introduced into both the synthetic and the wild-type DZFs. This mutation has been shown to disrupt homodimerization of the Ikaros and Hunchback DZF domain (McCarty *et al.*, 2003). All DZF domains with the exception of wild-type Pegasus were affected by this mutation in their ability to mediate homo- and heterotypic interaction (data not shown). This indicates that both interacting DZFs have to be intact and present to activate the expression of the *lacZ* reporter gene.

3.4.2 Anti-parallel interaction mode for synthetic DZFs

Analysis of the identities and homo- and heterotypic interaction specificities of the synthetic DZFs suggests that these domains interact in an anti-parallel fashion (i.e. the N-terminal finger of one monomer interacts with the C-terminal finger in the other monomer and vice versa). As mentioned in Chapter 1 (section 1.3.4.3), previous studies suggested that DZFs dimerize in a parallel mode (i.e. the N-terminal C2H2 ZFs in each monomer interact with each other and C-terminal C2H2 ZFs in each monomer interact with each other) (McCarty *et al.*, 2003). This conclusion was based on the observation that a synthetic Ikaros-Hunchback hybrid DZF could efficiently homo-dimerize and on the assumption that the individual C2H2 ZFs in this synthetic hybrid display the same interaction specificities then the parental DZFs from which they were constructed (McCarty *et al.*, 2003). This analysis was applied to the various DZFs identified in the selections. Six pairs of synthetic DZF domains are composed of subdomains that indicate a specific orientation: five pairs suggest an anti-parallel orientation while one pair is more consistent with a parallel interaction mode (Figure 3.8 and Table 3.2). The remaining pairs of chimeric DZFs contain subdomains that do not provide information in terms of an interaction mode since they can interact using either a parallel or an anti-parallel mode (Table 3.2).

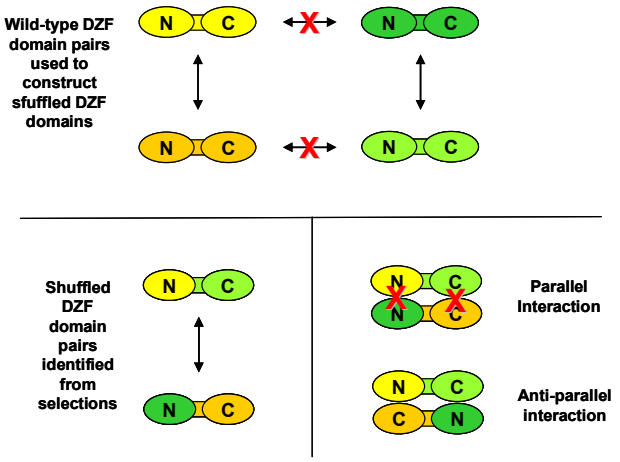

Figure 3.8 Schematic indicating how interaction specificities of synthetic DZF domains suggest an anti-parallel interaction mode. Wild-type DZF domains used to construct the synthetic DZFs and their interaction specificity are shown on top of the Figure. Interaction specificity of a pair of synthetic DZFs is shown on the bottom (left side). Occurring interaction suggests an antiparallel interaction mode for the interaction mediated by the two synthetic DZFs. Interactions are indicated by an arrow while red X's indicate that no interaction was obtained. Note that in this presentation the linker region was assigned to a certain finger but can also occur in the opposite constellation. See text for details.

However, additional confirmation of an anti-parallel interaction mode is provided by the specificity profiles of the four synthetic pairs. As mentioned above three of the four pairs prefer to interact in a heterotypic fashion and one DZF domain in each pair is not able to homodimerize at all which also rules out a strictly parallel interaction. For example, Ik-Ik-Hd and Pe-Pe-Eo can mediate robust heterodimerization but each individual DZF in this pair interacts less (Pe-Pe-Eo) or not at all (Ik-Ik-Hd) with itself.

3.4.3 Prediction and design of interactions between DZF domains

Since preliminary results demonstrated that shuffling of DZF-derived C2H2 ZFs can yield synthetic DZFs with new specificities we were also interested in creating new DZF domains by mixing certain ZFs of interest together. We decided to test whether the anti-parallel interaction mode could be used as a general rule for engineering DZF-mediated interactions. To do this, all combinations of C2H2 ZFs and linkers from the Ikaros and *Drosophila* Hunchback DZFs were systematically synthesized applying the same approach used to construct our libraries. Connecting these Ikaros-Hunchback chimeras to both the Zif268 as well as the α-subunit of the RNAP resulted in eight different fusion proteins that were tested for their ability to interact with each other using the B2H system as a reporter. It was found that none of them showed a significant interaction (data not shown). On the contrary, a functional chimeric protein containing portions of the human Ikaros and *Drosophila* Hunchback DZFs has been described recently (McCarty *et al.*, 2003; see also section 1.3.4.3).

The reason for this discrepancy remains unknown but in the approach of McCarty and coworkers (2003) the breakpoint for separating the protein into two subdomains was defined differently. Thus, residues from the respective other subdomain were included into their chimera which were not included in our chimeric peptides. It is also important to note, that we tested this chimeric protein in the B2H system and found that it interacted only weakly as judged by our assay.

One possible explanation for these findings is that DZF-derived C2H2 ZFs can not be considered as completely modular since not all combinations of fingers can produce functional interaction domains (see also section 3.8.1 below). This rules out that interactions between DZF domains can be consistently predicted and therefore deliberately designed and further emphasizes that methods such as selections currently provide the only way to obtain such synthetic interacting DZF pairs.

3.5 Analysis of DZF domains in mammalian cells

3.5.1 Synthetic DZF domains are functional in the nucleus of mammalian cells

3.5.1.1 Assay for DZF interactions

To further test whether interacting DZFs can mediate protein-protein interactions in the nucleus of a mammalian cell we developed an "activator reconstitution" assay using a previously described artificial transcriptional activator of the endogenous human VEGF-A gene. This activator protein consists of a synthetic DNA-binding domain that binds within a region of accessible chromatin in the human VEGF-A promoter (originally termed "VZ+434b", hereafter termed DBD) (Liu *et al.*, 2001) and a NF-κB p65 transcriptional activation domain (termed "p65"). We divided this synthetic activator in its two domains and fused individual DZFs to each of the two fragments. Dimerization mediated by the DZF domains should then reconstitute this bi-partite transcriptional activator and therefore lead to an activation of the endogenous VEGF-A gene (Figure 3.9A, Pollock *et al.*, 2002).

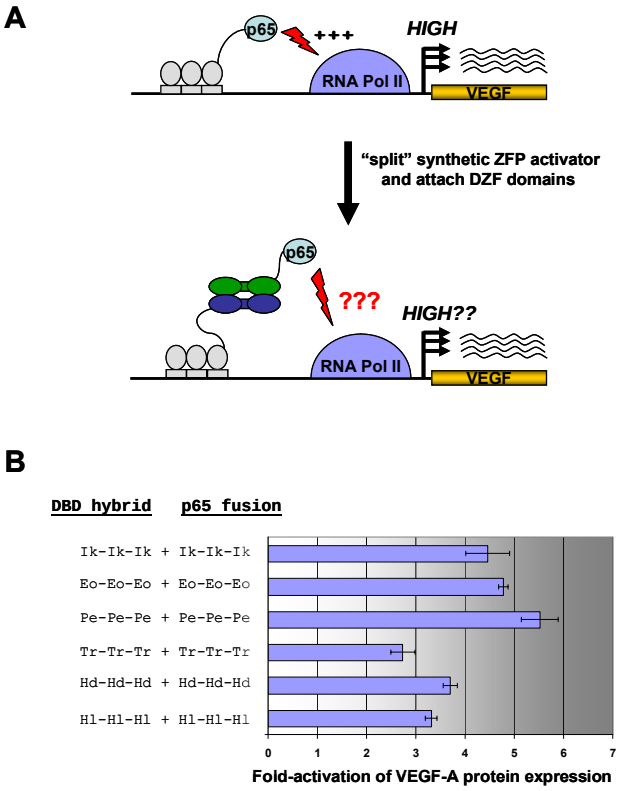

Figure 3.9 DZFs can be used to assemble an artificial bi-partite transcriptional activator in human cells. (A) A Schematic overview of the mammalian cell-based "activator reconstitution" assay for testing DZF interactions is presented. The originally described artificial transcriptional activator of the endogenous human VEGF-A gene is shown on top (Liu *et al.*, 2001). By splitting this zinc finger protein (ZFP) into its DBD and its transcriptional activation domain (p65) and attaching DZF domains to these fragments, this setup can be used to test if two DZF domains can mediate interaction. Interactions of DZFs (green and dark blue ovals) should mediate reconstitution of this synthetic activator protein which should result in an activation of VEGF-A expression (bottom panel). **(B)** Interactions mediated by wild-type DZFs were assessed using the activator reconstitution assay. Fold-activation of VEGF-A protein expression was calculated by measuring the VEGF-A contents in the culture medium using ELISA. Bars shown represent mean fold-activations of VEGF-A expression determined from three independent experiments and error bars indicate standard errors of the mean.

To test whether expression of various DBD-DZF and p65-DZF fusion proteins could promote a dimerization dependent activation of VEGF-A, plasmids encoding these hybrid proteins were introduced into Flp-In TRex HEK293 cells using transient transfection. Transfected cells were induced after 24h by adding IPTG (see section 2.3.2.1), and the level of VEGF-A expression was measured by ELISA (Liu *et al.*, 2001).

3.5.1.2 Synthetic DZFs can mediate assembly of heterologous protein domains in mammalian cells

Using the activator reconstitution assay, we initially assessed whether various wild-type DZF pairs could mediate reconstitution of this bi-partite transcriptional activator in the nucleus of a human cell. As shown in Figure 3.9B, all tested wild-type DZFs were able to mediate efficient activation of VEGF-A expression in this assay. In contrast, transfection of either hybrid protein alone did not activate VEGF-A expression indicating that activation depends on the presence of an interacting DZF domain on both hybrid proteins (data not shown). This suggests that the artificial bi-partite activator was reconstituted at the VEGF-A locus by the DZF domain interaction and could therefore mediate activation of the VEGF-A gene.

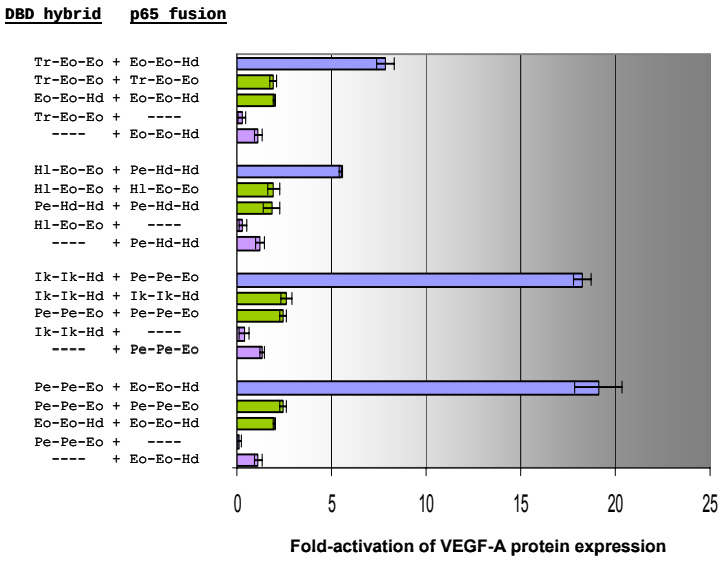

Figure 3.10 Analysis of homo- and hetero-typic interaction of synthetic DZFs using the activator reconstitution assay. DZFs derived from four pairs were tested for their ability to mediate hetero- (blue bars) and homo-typic (green bars) interactions. Control experiments expressing combinations of the synthetic DZFs with either the DBD or p65 with no DZF fused (indicated as '----') are also shown (purple bars). Each experiment was performed at least three times. Bars shown represent mean fold-activations of VEGF-A expression determined from three independent assays. Standard error of the mean is represented as error bars. Identities of subdomains present in the synthetic DZFs are abbreviated as in Figure 3.4.

We further tested whether the four chimeric DZF pairs we identified and characterized in the B2H system would also be able to mediate activation of VEGF-A. As shown in Figure 3.10, all four synthetic pairs could mediate activation of VEGF-A expression as measured by ELISA indicating that they were able to interact with each other. Control experiments expressing either the DBD-hybrid or the p65-hybrid protein alone did not show increased activation of VEGF-A expression demonstrating that both interacting DZFs have to be present to reconstitute the synthetic activator (Figure 3.10). In addition, we confirmed that the D18Q mutation which disrupted DZF mediated dimerization in the B2H system also affected dimerization of the various DZFs in the mammalian "reconstitution assay" (data not shown).

Next, the ability of the six individual DZFs to mediate homodimerization was examined and these results show that all DZF domains failed to interact with themselves and were only able to mediate heterotypic interactions (Figure 3.10). This result is different from previous experiments performed using the B2H reporter system where at least one of the DZFs in each pair could still mediate self interaction. It is unclear why the mammalian "activator reconstitution" assay does not detect self-interaction of the DZFs but one explanation could be that homodimeric interactions (as in the case for wild-type DZFs, Figure 3.9B) generally

do not mediate VEGF-A activation as sufficiently as heterodimeric interactions (e.g. as mediated by the synthetic DZFs, see Figure 3.10, see also section 4.8.3 below).
In summary, these results demonstrate that the synthetic DZF pairs are able to mediate heterotypic dimerization in this assay and therefore function in the context of a mammalian cell. In addition, these data show that DZF domains can be used to assemble an artificial bipartite activator at the endogenous human VEGF-A gene locus resulting in activation of VEGF-A.

3.5.1.3 Interactions are specific and do not depend on overexpression

To further test the applicability of these synthetic domains as protein-interaction modules, the expression levels were analyzed. These domains might also interact with certain endogenous proteins (e.g., other DZF-containing proteins) and the observed activity and specificity might depend on overexpressing them. Thus, experiments were performed to rule out the possibility that the interactions of synthetic DZFs are dependent on overexpression. FLAG epitope-tagged versions of DZF-DBD and DZF-p65 fusion proteins were constructed, whose expression could be quantified by Western blotting using an anti-FLAG monoclonal antibody. To test, whether the interaction of the synthetic DZFs depends on their over-expression, the expression level of the DZF-p65 fusion proteins in VEGF-A activator reconstitution experiments was lowered by decreasing the amount of DZF-p65-encoding plasmid used to transfect the cells. Western blot analysis demonstrated that for three of the four synthetic DZF pairs, the level of p65-DZF was decreased by 7-fold or more without affecting DZF-DBD expression levels (Figure 3.11A). Although the amount of p65 was significantly reduced, these proteins were still able to mediate robust activation of VEGF-A (Figure 3.11B). In the case of the Ik-Ik-Hd and Pe-Pe-Eo DZFs, the level of Pe-Pe-Eo DZF-p65 fusion protein could not be decreased even when the amount of plasmid encoding this protein was reduced by 10-fold. Further reduction of the plasmid amount resulted in an impaired VEGF-A activation (data not shown). Thus, three of our four synthetic DZF pairs were able to mediate interaction even when the expression level of one member of the pair was reduced by >7-fold suggesting that overexpression of our synthetic DZF domains is not required to achieve interaction.

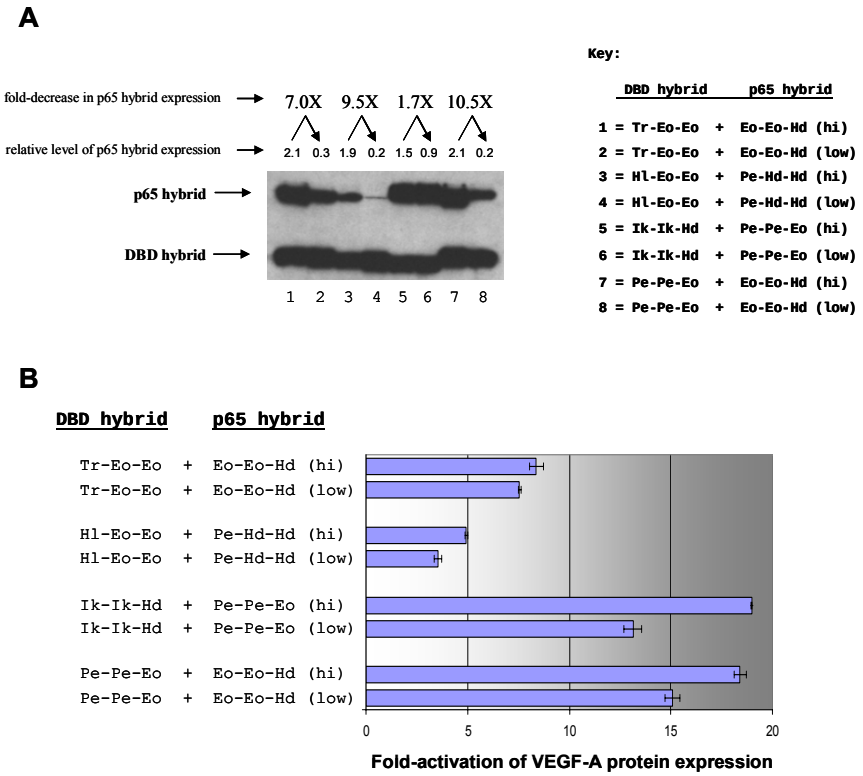

Figure 3.11 Synthetic DZF-DZF interactions do not critically depend upon protein over-expression. **(A)** Western blot analysis of human cells transfected with various amounts of DZF-encoding plasmids. Western blots were performed on 293 cells transfected with combinations of FLAG-tagged DBD-DZF and p65-DZF hybrid proteins using an anti-FLAG monoclonal antibody. For each pair of DZF hybrid, two different amounts of DZF-p65-encoding plasmid were used ("hi" = 0.25 ug DNA, "low" = 0.06 ug DNA). The amount of cell lysate loaded in each lane was normalized to the number of viable cells determined by WST1 assay (see section 2.3.2.3). Band intensities were quantified using BioRad Quantity One software and a BioRad Fluor-S MultiImager instrument. The relative intensities of each DZF-p65 band are shown above each lane and represent the average of band intensities from two assays. Relative values were calculated by normalizing the intensity of each DZF-p65 band to the associated DZF-DBD band. **(B)** VEGF-A assays were performed simultaneously on culture supernatants of cells used for the Western blot analysis. Fold-stimulation of VEGF-A protein expression was calculated by measuring the VEGF-A contents in the culture medium using ELISA. Bars shown represent mean fold-activations of VEGF-A expression determined from two independent experiments and error bars indicate standard errors of the mean. Identities of subdomains in the synthetic DZFs are abbreviated as in Figure 3.4. This Figure was taken from Giesecke *et al.* (2006).

An additional set of experiments was performed to test, if the specificities of DZFs can be maintained even when they and their potential competitive interaction partners are both overexpressed. It was found that each of the four synthetic DBD-DZF hybrids interacts specifically with a DZF-p65 fusion harboring its synthetic interacting partner DZF but fails to interact with DZF-p65 fusions harboring either a synthetic non-interacting DZF or the wild-type TRPS1 DZF (Figure 3.12A). Importantly, Western blot analysis of cells expressing these proteins revealed that within each set of experiments each DZF-DBD fusion protein and its

three corresponding p65-DZF fusions (harboring an interacting synthetic DZF, a non-interacting synthetic DZF and the DZF from wild-type TRPS1) were expressed at similar levels (Figure 3.12B). This suggested that the specificities of DZF interactions are maintained even when the DZF hybrid proteins are overexpressed. Furthermore, since TRPS1 is widely expressed in many different human cell types (Momeni *et al.*, 2000) these results also suggest that most of the synthetic DZFs do not effectively interact with potential endogenous competitors, even when these are very abundant.

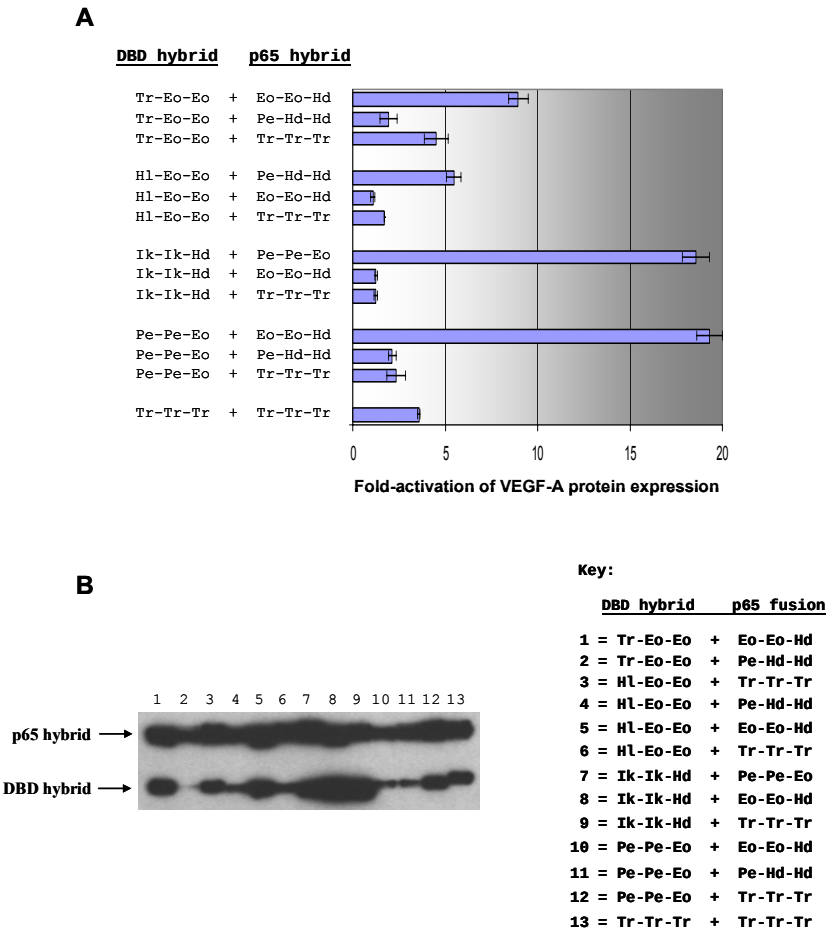

Figure 3.12 Overexpression of non-interacting synthetic DZFs does not result in an activation of VEGF-A. (A) Expression of VEGF-A from human 293 cells transfected with plasmids expressing various combinations of interacting and non-interacting DBD-DZF and p65-DZF hybrid proteins. Fold-stimulation of VEGF-A protein expression was calculated as described in Figure 3.11. Identities of subdomains present in the synthetic DZFs are abbreviated as in Figure 3.4. (**B)** Western blots of whole cell lysates prepared from the transfected cells used for ELISA assays in (A) are shown. Blots were performed as described in Figure 3.11. This Figure was taken from Giesecke *et al.*, (2006).

3.5.2 Synthetic DZF domains are functional in the cytoplasm of mammalian cells

To determine whether our synthetic DZF pairs can also mediate protein-protein interactions in the cytoplasm of a human cell, co-immunoprecipitation assays were performed. For these experiments a previously developed co-immunoprecipitation method to assess DZF interactions was used. In this assay, the different synthetic DZFs were co-expressed in HEK 293 cells as two fusion proteins of different molecular weights (Figure 3.13).

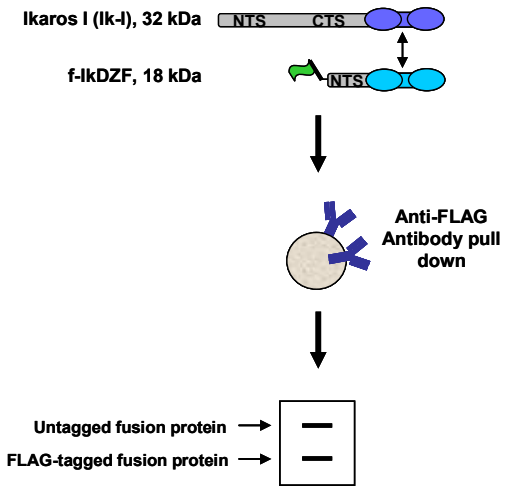

Figure 3.13 Scheme of the co-immunoprecipitation assay for analyzing DZF domain mediated interactions. FLAG tagged DZF proteins consisting of the Ikaros N-terminal sequence (NTS), the DZF domain (light blue ovals) and a N-terminal FLAG tag were co-expressed with untagged full-length Ikaros isoform I (Ik-I) consisting of NTS, C-terminal sequence (CTS) and the DZF domain (dark blue ovals) in HEK 293 cells. Cell lysates containing both FLAG-tagged and untagged DZFs were incubated with anti-FLAG-agarose beads. Proteins bound to these beads were visualized by Western Blot using antibodies that recognize the NTS. Detection of two bands indicates the interaction between the two DZFs since untagged fusion proteins (upper band) can only be pulled down by binding to FLAG-tagged proteins (lower band) (McCarty *et al.*, 2003).

The larger size protein represents the Ikaros isoform I (Ik-1, see section 1.3.1.2) containing a replacement of the C-terminal fingers with various DZFs. The smaller size protein consists of only 79 amino acids from the N-terminus of Ik-I fused directly to the various synthetic DZFs and harbors a FLAG epitope tag (McCarty *et al.*, 2003). None of the DZF fusion proteins used in this assay harbors a nuclear localization signal. To test for interaction, cytoplasmic cell lysates were analyzed by immunoprecipitation using an anti-FLAG antibody followed by extensive washing. The precipitated DZF fusion proteins were then visualized by Western blot analysis using Ikaros antibodies that recognize the common N-terminal domain present in both fusions.

As shown in Figure 3.14, it was found that the various tagged proteins efficiently co-immunoprecipitated the corresponding untagged fusion proteins suggesting that the synthetic DZF pairs can mediate interaction in this assay. Control experiments expressing only the untagged DZF fusion peptides demonstrated that these peptides did not bind to the anti-FLAG antibody in a nonspecific manner, indicating that the precipitation is due to the interaction between the different DZF domains (data not shown).

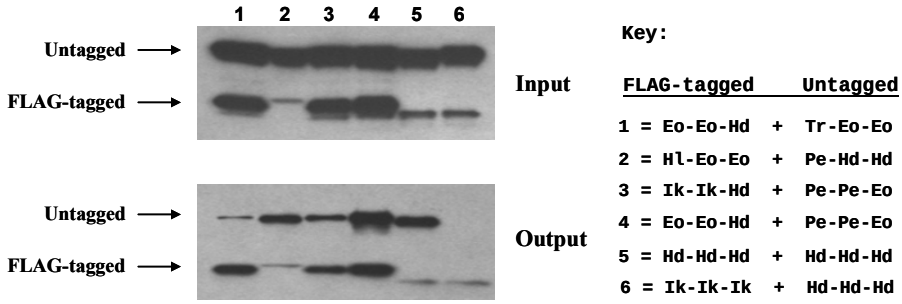

Figure 3.14 Synthetic DZF interactions analyzed using the coimmunoprecipitation assay. Mammalian cell extracts were co-transfected with plasmids encoding combinations of tagged and untagged DZFs (defined in the Key) and immunoprecipitated using an anti-FLAG antibody. The "input" represents fusion protein samples before incubating them with anti-FLAG antibody and the "output" shows proteins after the immunoprecipitation step. Western blots were performed with an antibody that recognizes the Ikaros NTS sequence. Controls testing the homotypic interaction of wild-type Hunchback (positive control) and the interaction between Ikaros DZF and Hunchback DZF domains (negative control) are also shown. Identities of subdomains in the synthetic DZFs are abbreviated as in Figure 3.4.

Thus, the results of this assay together with the results of the "activator reconstitution" experiments show that the synthetic DZF pairs are functional in the complex protein environment of a human cell and can interact in both the nucleus and the cytoplasm.

3.6 Engineering of more extended C2H2 ZF-mediated protein-protein interfaces

3.6.1 Design of extended interaction surfaces

In previous studies various groups have attempted to create high-affinity binding peptides by multimerizing two or three C2H2 DNA-binding ZFs (see section 1.2.5). Different strategies were used to create these longer arrays of ZFs capable of recognizing extended DNA sequences. For example, four- and six-finger peptides have been created by joining together units consisting of two ZFs with different types of linker. These tandem arrays of ZFs were capable of binding 12 and 18 base pair DNA sequences with high affinities and specificities (Moore *et al.*, 2001; Tan *et al.*, 2003; Urnov *et al.*, 2005). We were therefore interested in testing whether it would also be possible to design extended protein-protein interfaces by linking together DZF domains. The goal was to connect two synthetic DZFs to create double-

DZFs (dDZFs) consisting of four ZFs which are expected to mediate protein-protein interactions with higher affinity than the single DZF domains. Figure 3.15B shows the two pairs of synthetic DZFs that were chosen for designing the double-DZFs. The individual DZFs in these pairs prefer to heterodimerize with their respective partner and the interaction is very specific such that individual DZFs in each pair do not bind to DZFs present in the other pair (Figure 3.15A). Choosing these DZFs has the advantage that interactions occurring between the individual DZFs in these double-DZFs can be controlled and unwanted inter- and intra-molecular interactions can therefore be avoided.

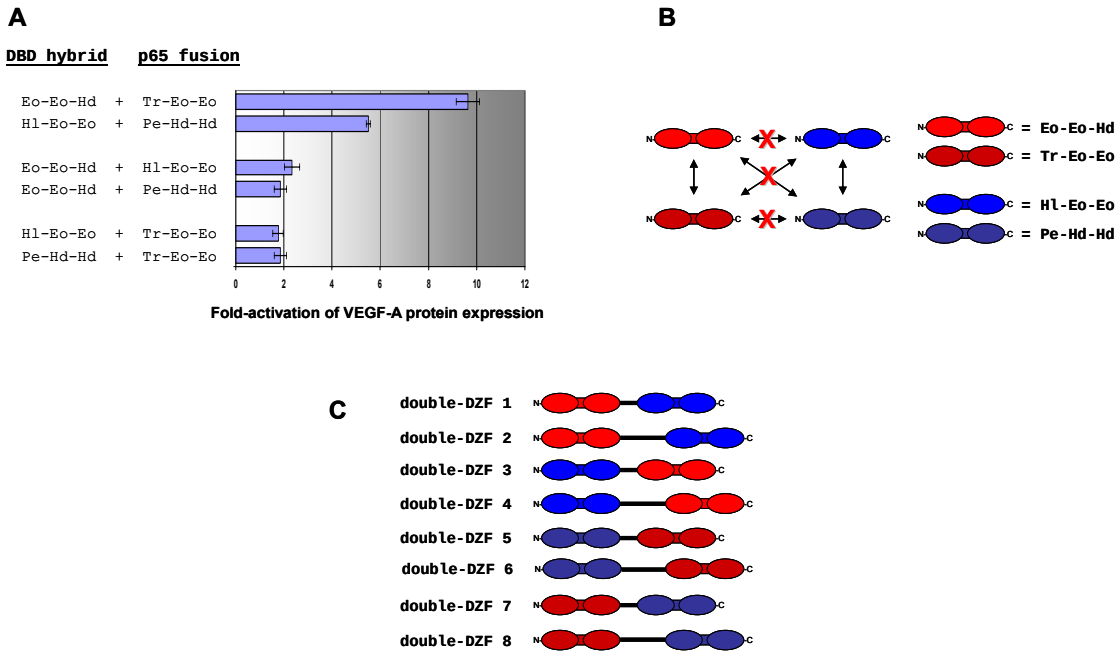

Figure 3.15 Construction of synthetic double-DZFs. (A) Pairwise interactions between four synthetic DZFs were tested using the mammalian cell-based activator reconstitution assay. Fold-stimulation of VEGF-A protein expression was calculated using ELISA measurements of secreted VEGF-A that were performed at least three times. Bars shown represent mean fold-activations of VEGF-A expression and error bars indicate standard errors of the mean. **(B)** Summary of interaction specificities of synthetic DZFs used to construct double-DZFs. Interaction specificities shown (arrows) were determined using the B2H system (Table 3.3) and the mammalian cell-based "activator reconstitution" assay (**A**). Red X's indicate that no interaction was detected. **(C)** Schematic of synthetic double-DZFs. Inter-DZF linkers are shown as black lines. Short lines represent the GEKP linker while long lines indicated the hinge linker (see main text for details). Double-DZFs 1-4 were connected to the DBD and double-DZFs 5-8 were fused to p65. This Figure was taken and adapted from Giesecke *et al.*, (2006).

Eight double-DZFs were designed by varying the linear order of the synthetic DZFs and using two kinds of linkers between them (Figure 3.15C). For one set of constructs the synthetic DZFs were connected with a linker consisting of the amino acids GEKP after the residue following the last histidine of their natural linker sequence. In a second set of

constructs a flexible hinge peptide (EFPKPSTPPGSSGGAP) was used as the linker region, which is derived from the murine IgG3 hinge region (Pluckthun and Pack, 1997; Deyev *et al.* 2003).

3.6.2 Characterization of various double-DZFs

To initially test whether double-DZFs could mediate specific interactions, the mammalian cell-based activator reconstitution assay was used. As shown in Figure 3.16, only double-DZF1/double-DZF5 and double-DZF4/double-DZF7 pairs were able to mediate robust activation of VEGF-A expression.

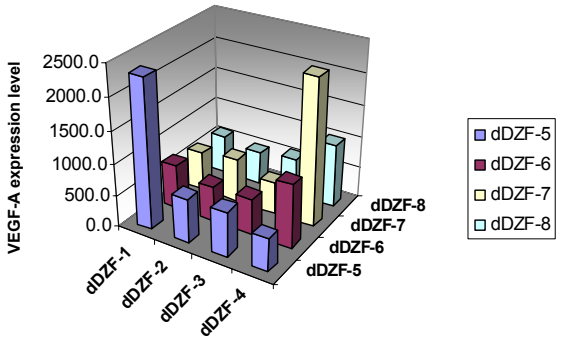

Figure 3.16 Synthetic double-DZFs interact in mammalian cells. Pairwise combinations of plasmids encoding double-DZFs (dDZF) were co-transformed into HEK293 cells and interactions of these pairs were assessed using the mammalian cell-based activator reconstitution assay. VEGF-A protein expression levels were calculated as absolute values using ELISA measurements of secreted VEGF-A that was performed two times. Identities of dDZF domains as indicated in Figure 4.15C.

By contrast, testing different combinations of the same double-DZFs (double-DZF1/double-DZF7 and double-DZF4/double-DZF5) resulted in a lower VEGF-A expression level similar to the one obtained with single DZFs. Interestingly, these pairs of double-DZFs only differed in the linear order of their constituent synthetic DZFs. The remaining double-DZFs exhibited a lower level of VEGF-A activation. Surprisingly, the only difference between some of these double-DZFs is the linker region between the single DZFs pairs. For example double-DZF1 and double-DZF5 mediated efficient activation of VEGF-A expression. The single DZFs in these two double-DZFs are linked by the amino acids GEKP. In double-DZF2 and double-DZF6 the same single DZFs were fused in the same linear order using the hinge linker and the activation of VEGF-A mediated by these two double-DZFs was less efficient. A simple explanation of these results is that the linker may play a role in orientating the two single DZFs in the double-DZFs and is therefore important for the ability of the double-DZFs to interact with another double-DZF. It can certainly be ruled out that the linker is part of the

interacting surface because both double-DZF1/double-DZF7 and double-DZF4/double-DZF5 pairs were not able to mediate a robust activation of VEGF-A, although they harbor the same linker region as double-DZF1/double-DZF5 and double-DZF4/double-DZF7 pairs which were able to active VEGF-A very efficiently.

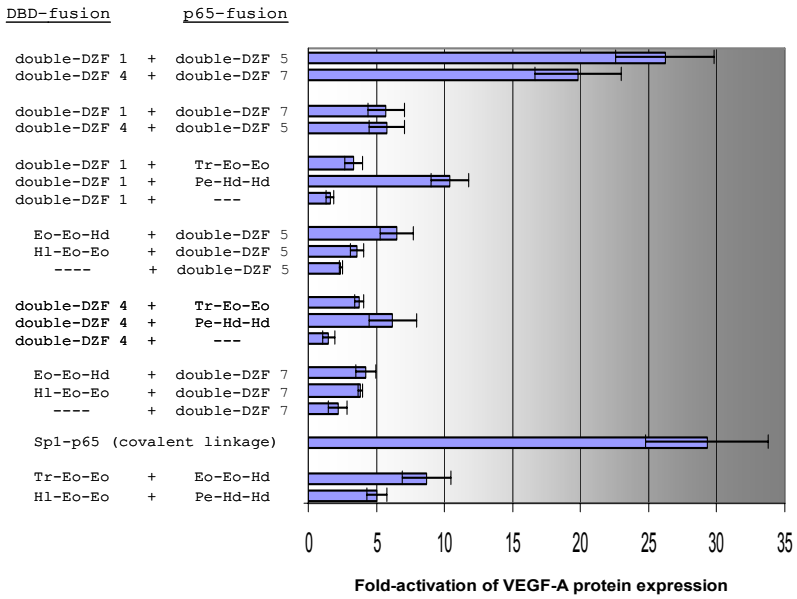

Figure 3.17 Analysis of the interactions mediated by double-DZFs in mammalian cells. The strength and mode of interactions mediated by the double-DZF1/double-DZF5 and double-DZF4/double-DZF7 pairs was assessed using the mammalian cell-based activator reconstitution assay. dDZFs were tested for their interaction with each other and with the single DZFs from which they were derived. Controls testing the interaction of single DZF pairs and of the intact synthetic VEGF-A activator SpI-p65 (consisting of the DBD that binds the VEGF-A promoter covalently fused to the p65 activation domain) are also shown. Fold-stimulation of VEGF-A protein expression was calculated by measuring the VEGF-A contents in the culture medium using ELISA. Bars shown represent mean fold-activations of VEGF-A expression calculated from three independent assays and error bars indicate standard errors of the mean. Identities of the synthetic DZFs are abbreviated as in Figure 3.4, and identities of dDZF domains are as indicated in Figure 3.15C.

We were interested in comparing the interaction strength of the double-DZFs with the level of interaction observed with the single DZF pairs that we used to construct these double-DZFs. As shown in Figure 3.17, both double-DZF pairs (double-DZF1/double-DZF5 and double-DZF4/double-DZF7) mediated very robust activation of VEGF-A expression, greater than the activation observed with the single DZF pairs from which they were constructed. The level of activation achieved by the double-DZFs is similar to that obtained with the originally described artificial VEGF-A activator (SpI-p65 in Figure 3.17; Liu *et al.*, 2001). Control experiments demonstrated that the expression of neither the DBD-double-DZF protein nor the p65- double-DZF protein alone activated VEGF-A expression. Interactions of

the double-DZFs with the individual single DZFs which they consist of caused a reduced activation of VEGF-A compared to the activation obtained with the double-DZFs. In addition, when the linear order of the synthetic DZFs was reversed, both double-DZF pairs (double-DZF1/double-DZF7 and double-DZF4/double-DZF5) stimulated VEGF-A expression less efficiently. This suggests that double-DZF-1/double-DZF-5 and double-DZF-4/double-DZF-7 interact stronger because they are able to utilize both single DZFs whereas double-DZF-1/double-DZF-7 and double-DZF-4/double-DZF-5 interact less strongly because they use only one of the two DZFs for mediating the interaction.

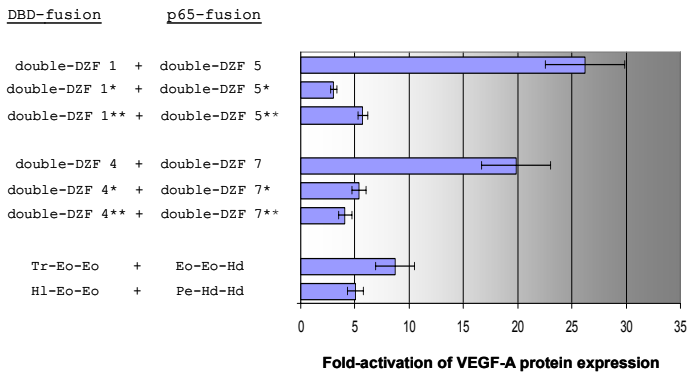

Figure 3.18 Analysis of the effect of the D18Q mutation introduced into the single DZFs present in the double-DZFs. The D18Q mutation was introduced into one of the single DZFs present in the DBD-dDZF hybrid and into the corresponding single DZF present in the p65-dDZF hybrid. One star (*) indicates that the mutation was introduced into the single DZFs Hl-Eo-Eo and Pe-Hd-Hd present in each of the dDZF domains, while two stars (**) indicate that the mutation was introduced into the Tr-Eo-Eo / Eo-Eo-Hd DZF domains. Note that all dDZFs consist of combinations and different arrangements of these single DZFs (see Figure 3.15C). Controls testing the interaction of the intact dDZFs and the single DZF pairs are also shown. Fold-stimulation of VEGF-A protein expression was calculated from three independent assays and error bars indicate standard errors of the mean.

To further confirm this, another set of double-DZFs was designed where each of the single DZFs in every double-DZF construct was individually mutated by introducing the D18Q mutation. A double-DZF construct that harbors such a mutation in one of its DZFs was then tested for its ability to interact with another double-DZF construct, which harbors a mutation in the corresponding interacting DZF domain (i.e., two single DZF domains that usually interact with each other). Both of the double-DZF pairs were impaired in their ability to activated VEGF-A expression when they harbored mutations in interacting single DZFs. The activation level was similar to that observed with the single DZF pairs suggesting that only the wild-type DZFs in the double-DZF constructs mediated the interaction (Figure 3.18). This confirms that the two wild-type double-DZF pairs are able to mediate a very robust activation

of VEGF-A by using both single DZFs for the interaction, thus displaying an extended interaction surface.

3.6.3 Double-DZF interaction confirms anti-parallel interaction mode for DZF domains

Analysis of the affinities of the various double-DZFs and the linear order of the single synthetic DZFs in these double-DZFs confirms an anti-parallel interaction mode for DZF domains (Figure 3.19). Assuming that the DZFs within the double-DZF interact in an anti-parallel fashion, double-DZF1/double-DZF5 and double-DZF4/double-DZF7 pairs would be expected to interact more strongly because they would be able to utilize both DZFs at the same time. On the other hand, the double-DZF1/double-DZF7 and double-DZF4/double-DZF5 pairs would not be able to use both DZF domains and are therefore expected to interact more weakly (Figure 3.19). Thus, the relative affinities of the different double-DZF pairs provided additional evidence for an anti-parallel interaction mode of the DZF domains.

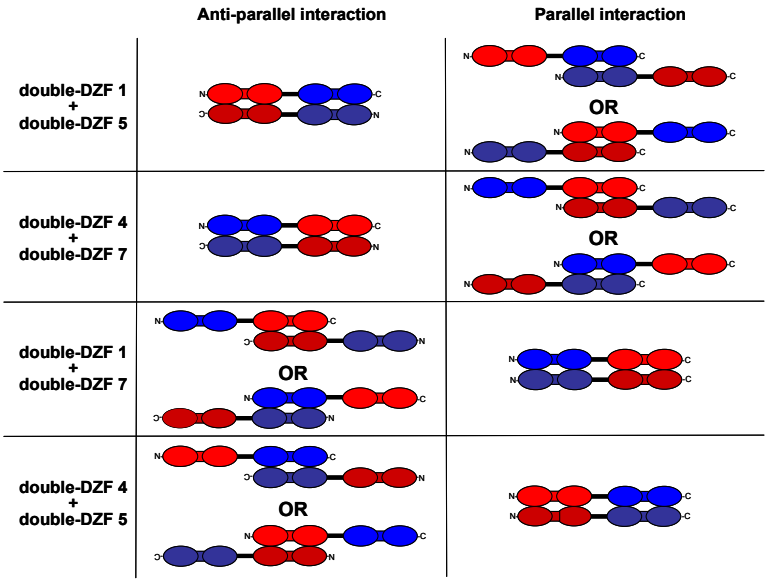

Figure 3.19 Potential interaction modes for synthetic double-DZFs. Possible anti-parallel and parallel interation modes applied to the individual DZFs within the double-DZFs result in various models for the interaction of the double-DZFs. Identities of double-DZFs and the single DZFs from which they were constructed are color-coded as in Figure 3.15B and C. This Figure was taken from Giesecke *et al.,* (2006).

3.6.4 Scaffold

We were interested in exploring whether it is possible to engineer synthetic "scaffolds" or "adaptors" (Ferrell, 2000; reviewed in Pawson and Scott, 1997) which can be used to tether various proteins that are linked to DZF domains. Initially we planed to design a transcriptional scaffold in which two transcriptional activators (input) are integrated into one output represented by activation of the VEGF-A expression. This should result in a synergetic transcriptional activation which is described as the process where several activators work together and support each other in activating the respective target gene. Synergetic activation generally results in a substantial increase of transcriptional activation. This level of activation is higher compared to a situation where the effect of two independent activators is purely additive (reviewed in Ptashne and Gann, 1997; Jackson *et al.*, 1998; Zhu *et al.*, 2000; Cassel *et al.*, 2002; reviewed in Remenyi *et al.*, 2004). Thus, the idea was to design constructs where two activation domains (p65 and VP16) are each fused to different synthetic DZF domains which both can interact with one of the single synthetic DZFs present in a double-DZF domain. Applying double-DZF domains fused to SpI should permit an assembling of two activation domains at the promoter which in turn should lead to a synergetic activation of VEGF-A (Figure 3.20).

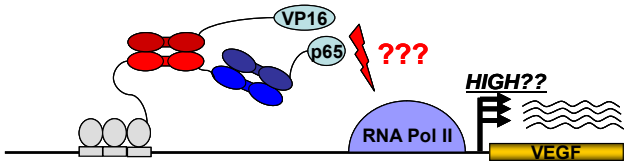

Figure 3.20 Schematic overview of a transcriptional scaffold. Interactions between the dDZF (light red and blue double ovals connected by a line) fused to a synthetic DBD that binds to a DBS in the VEGF-A promoter and two single DZFs (dark red and blue double ovals) fused to the p65 and the VP16 transcriptional activation domains should stimulate VEGF-A expression stronger than interactions with either p65 or VP16 DZF fusions alone.

For an initial set of constructs double-DZF1 and double-DZF4 were used since both worked well for designing an extended interaction surface. The two corresponding single partner DZFs were separately fused to the p65 and the VP16 activation domains. These constructs were introduced into our mammalian cell-based assay and tested for their ability to mediate activation of VEGF-A. As shown in Figure 3.21, the interaction between the double-DZF domains and both single synthetic DZFs fused to the activation domains was low and the activation of VEGF-A expression was not very efficient.

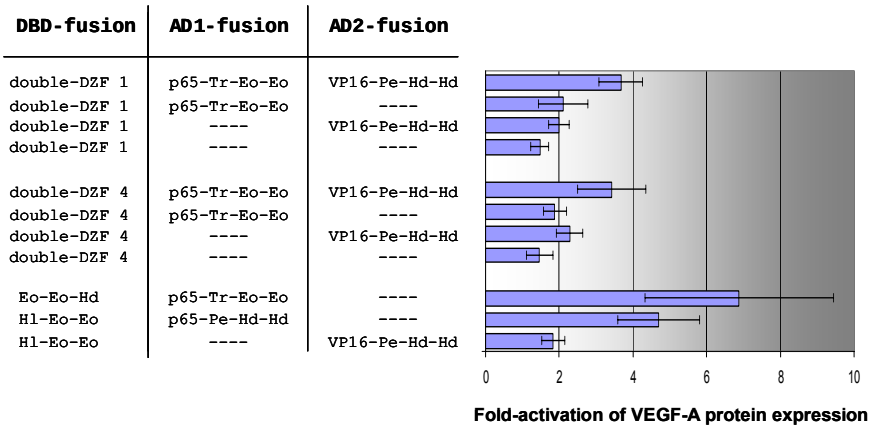

Figure 3.21 Interactions of double-DZFs with single DZFs fused to different activation domains (AD). dDZFs 1 and 4 are connected to the DBD (left column) while single DZFs are either fused to p65 or VP16 (AD1 or AD2, middle and right column). Various combinations of plasmids encoding these fusions (see columns) were transfected into HEK293 cells and interactions were assessed using the mammalian cell based activator reconstitution assay. Fold-stimulation of VEGF-A protein expression was calculated using ELISA measurements of secreted VEGF-A that were performed three times. Bars shown represent mean fold-activations of VEGF-A expression and error bars indicate standard errors of the mean. Identities of the synthetic DZFs are abbreviated as in Figure 3.4, identities of dDZF domains are as indicated in Figure 3.15C.

Control experiments showed that neither interactions of the double-DZFs with the individual single DZFs nor the double-DZF protein alone activated VEGF-A expression. In all cases the activation is clearly lower than that obtained with the single DZF pairs from which they were constructed. Interestingly, a significant difference in the level of VEGF-A expression was detected for the single DZF pairs depending on which activation domain was used for the activation. Fusing one of the DZFs to VP16 did not yield the same amount of VEGF-A activation as fusing the same DZF to p65 (data not shown). This suggested that the activation mediated by VP16 is generally lower then that mediated by p65 which could be one reason why the scaffold was not able to mediate robust activation of VEGF-A. In addition, we suspected that the double-DZF construct was not accessible and did not provide the right geometry for the two DZFs fused to the activation domains to achieve an optimal binding.

Therefore we decided to optimize the double-DZF construct by introducing different linkers in between the two synthetic DZFs. A very long and flexible linker consisting of the amino acids FHMSGGGGSGGGGS and a shorter FHMS linker were designed. Unfortunately, these new constructs did not improve the level of VEGF-A activation suggesting that the linker is either not important for the scaffold setup or does still not provide the right geometry. Interestingly, adding another p65 activation domain instead of using VP16 and therefore

recruiting two p65 peptides to the promoter caused an increased but still weak activation of VEGF-A expression (data not shown). Thus, we further tried to optimize the geometry of the scaffold using two p65 activation domains. We tested another scaffold setup where one of the p65 domains is fused to the N-terminus while the other p65 domain is fused to the C-terminus of the respective synthetic DZF domain. For this experiment double-DZF1 and double-DZF4 were used to recruit the two DZF-linked p65s. Unfortunately, this setup still did not lead to an increased level of VEGF-A activation (data not shown).

3.7 Using DZF domains to dimerize DNA-binding zinc-fingers

3.7.1 Overview

Previous studies have shown that DNA-binding C2H2 ZFs can be linked together into tandem arrays capable of binding to extended DNA sequences with increased affinity and specificity. However, only limited numbers of fingers can be linked together in order to create a functional protein. Several synthetic peptides consisting of more then three ZFs have been characterized but these peptides displayed only little enhancements in their DNA-binding affinity (reviewed in Wolfe *et al.*, 2000; reviewed in Pabo *et al.*, 2001; see also section 1.2.5). These limitations are likely to be due to structural or energetical problems that emerge when more then three fingers are present. To avoid this problem, alternative approaches were taken to design DNA-binding proteins with higher and more specific DNA binding affinity. One successful strategy involved the attachment of a dimerization domain to DNA-binding ZFs which enables these ZF to bind DNA in a cooperative manner. Different dimerization systems have been described which assemble DNA-binding ZFs on the DNA and enhance their specificities and affinities (Pomerantz *et al.*, 1998; Wang and Pabo, 1999; Wolfe *et al.*, 2000; Wolfe *et al.*, 2003). Thus, we were interested in testing if DZF domains can also be used as modules that could be applied to dimerize DNA binding C2H2 ZFs.

3.7.2 Setup

As an initial test of whether DZF domains can be used as dimerization elements we applied these domains to cooperatively assemble DNA-binding ZFs on the DNA. For some of the DZF domains (e.g. - Ikaros family of transcription factors) it has been shown that they mediate dimerization of their N-terminal DNA-binding zinc finger domain (Sun *et al.*, 1996; Morgan *et al.*, 1997; Kelley *et al.*, 1998). Thus, we hypothesized that these domains should be able to mediate dimerization of heterologous DNA-binding C2H2 ZFs as well. To begin exploring the possibility of dimerizing DNA-binding ZFs using DZF domains, the B2H system was adapted so that it could be used as a reporter system for detecting protein-DNA interactions mediated by dimerization.

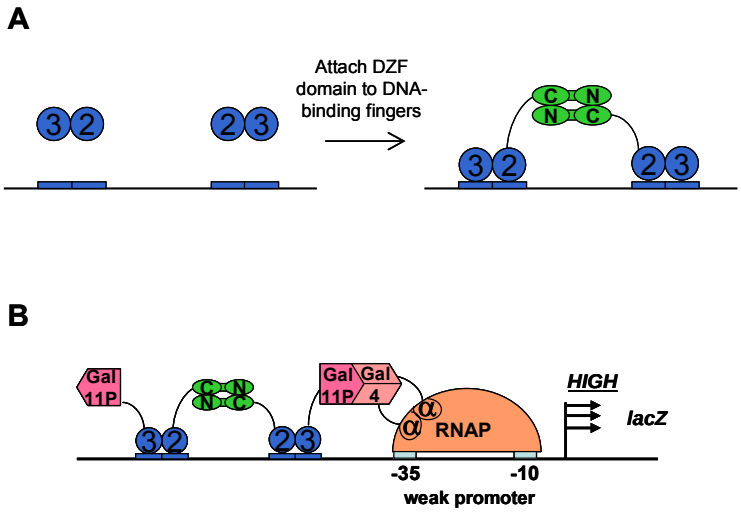

Figure 3.22 Schematic overview of the B2H system for testing if DZF domains can dimerize DNA-binding ZFs. (A) DNA-binding domains consisting of F2 and F3 of Zif268 but lacking F1 (numbered blue circles) fail to bind to their specific DBS (left side of Figure). Attachment of a dimerization element represented by the DZF domain (green double ovals) should mediate dimerization of the two DNA-binding ZFs, thus permitting them to bind to their composite DBS (right side of Figure). **(B)** The B2H system was used to detect cooperative binding of DZF-ZF23 fusion proteins to a composite DBS. In this setup Gal11P (dark pink) was fused to the C-terminus of F3 from Zif268 and its interacting partner Gal4 (light pink) was fused to the RNAP α-subunit. Assembling of the DNA-binding ZFs (numbered blue circles) on the composite DBS placed upstream of the *lacZ* reporter gene by DZF mediated dimerization should recruit the RNAP to the promoter via the interaction between Gal11P and Gal4, which should result in an activation of *lacZ* transcription. This Figure was kindly provided by K. Joung.

The idea was to construct a fusion of finger 2 and 3 from Zif268 (ZF23) to the DZF domain of Ikaros. Finger 2 and 3 from Zif268 were chosen because previous studies have shown that these two fingers are not able to bind tightly to their specific DNA site due to low binding affinities. However, attaching a dimerization domain can mediate the cooperative binding of

these two fingers to a composite DNA side consisting of two specific "half-sites" (Figure 3.22A, Wolfe *et al.*, 2000). To detect the protein-DNA interactions, a B2H reporter strain must be used which bears a composite DNA-binding side consisting of two "half-sites" known to be bound by ZF23 in the promoter region of the *lacZ* reporter gene. Furthermore, a fragment of the yeast Gal11P protein (GP) was connected to the C-terminus of the DZF-ZF23 fusion protein which will allow detection of DNA binding through its interaction with the yeast protein Gal4 (Joung *et al.*, 2000; Hurt *et al.*, 2003). To initiate activation of the *lacZ* reporter gene, the B2H reporter strain has to co-express Gal4 fused to the RNAP α-subunit with the DZF-Zif23-GP fusion protein. Cooperative binding of two DZF-Z23-GP fusion proteins to their composite DNA-binding site should then enable GP to recruit the RNAP to the promoter via its interaction with Gal4, which should in turn activate the expression of *lacZ* (Figure 3.22B).

3.7.3 Characterization of Ik-Zif268-Gal11P and Ik-Z23-Gal11P in the B2H system

Initial experiments to determine whether the Ik-Z23-GP fusion protein can be studied using this B2H setup were performed. First, we tested if the GP fragment is functional in the context of our Ikaros-Zif268 hybrid protein. Thus, the Ik-Zif268 fragment (containing all three DNA-binding ZFs) was fused to a fragment of the yeast GP protein using a flexible linker of the sequence GGGGS to create the Ik-Zif268-GP fusion. A C-terminal FLAG tag epitope was also added for Western blot analysis. Plasmids encoding the Ik-Zif268-GP fusion and the α-Gal4 hybrid protein were then introduced in the B2H strain harboring one Zif268 binding side and β-galactosidase assays were performed to assess the binding of this fusion protein to its DNA side. As shown in Figure 3.23, the Ik-Zif268-GP fusion protein can bind to its specific DNA site and efficiently activates transcription of the *lacZ* reporter gene compared with the Ik-Zif268 only control. This suggests that the GP protein is functional in the context of our fusion protein and can generally be used to detect DNA-binding through its interaction with the yeast protein Gal4.

F1 of the Zif268 protein was then deleted by PCR to create the Ik-Z23-GP hybrid protein. As expected, in contrast to the Ik-Zif268-GP, Ik-Z23-GP was not able to mediate activation of *lacZ* in the B2H system (Figure 3.23) suggesting that Finger 2 and 3 of Zif268 were not able to bind to the promoter consisting of one Zif268 DBS.

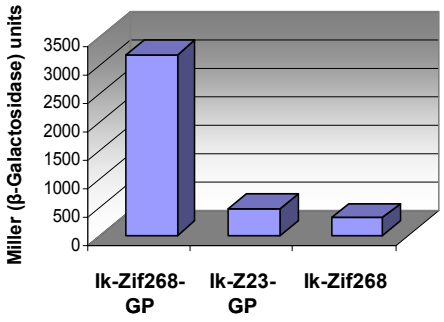

Figure 3.23 The B2H system can be used to study DZF mediated dimerization of DNA-binding ZFs. Plasmids encoding fusions between the Ik DZF domain, Gal11P (GP) and either Zif268 (Ik-Zif268-GP) or the Zif268 protein with a deleted F1 (Ik-Z23-GP) were co-transformed with the RNAP α-Gal4 plasmid into the reporter strain harboring only one Zif268 DBS and β-galactosidase assays were performed to assess the activity of these constructs. A control expressing the Ik DZF domain fused to Zif268 lacking the GP protein (Ik-Zif268) is also shown.

To further test whether the Ik-Z23-GP protein is stably expressed, Western blot analysis was performed. This should rule out that the abolished activation of *lacZ* is due to a lack of protein. Thus, whole cell lysates from the cultures used for the β-galactosidase assays were resolved by SDS-polyacrylamide gel electrophoresis. Western blot analysis was performed to visualize the hybrid proteins using an anti-FLAG antibody that recognizes the C-terminal FLAG tag epitope. We found that the Ik-Z23-GP protein is stably expressed (data not shown) and concluded that Ik-Z23-GP could not bind with high affinity to the Zif268 binding side because it contained only two DNA-binding ZFs.

3.7.4 Dimerization of Ik-Z23-GP using the DZF domain

We next wanted to determine whether the attached DZF domains can be used to mediate cooperative binding of two 2-Finger Zif268 domains to a composite DNA side composed of two half-sites. To assess cooperative DNA-binding by the Ik-Z23-GP protein, appropriate B2H reporter strains were constructed that harbor composite DNA-binding sites in the promoter region of the *lacZ* reporter gene. In these strains the six-base-pair binding sites ("half-site") for each Ik-Z23-GP monomer are arranged in an inverted orientation. A series of B2H reporters were generated containing various spacings of the half-sites (Table 3.4) to find an arrangement optimal for DNA-binding (termed as 0, 1, 3 ... sites). To test whether Ik-Z23-GP can dimerize and subsequently bind to the inverted half-sites, plasmids expressing Ik-Z23-GP together with a plasmid encoding the α-GAL4 were introduced into each of the B2H reporter strains to assess β-galactosidase activities. It was found that none of the reporters exhibited high activation of the *lacZ* gene indicating that Ik-Z23-GP was not able to bind cooperatively as a dimer. Increasing the concentration of IPTG to produce more protein did

not improve the ability of Ik-Z23-GP to activate *lacZ* (data not shown).

Spacing (bp)	Binding site
0	5′ – GCGTGGCCACGC – 3′
1	5′ – GCGTGGCCCACGC – 3′
3	5′ – GCGTGGGTCCCACGC – 3′
5	5′ – GCGTGGGTGGCCCACGC – 3′
7	5′ – GCGTGGGTGTAGCCCACGC – 3′
9	5′ – GCGTGGGTGGGTAGCCCACGC – 3′
11	5′ – GCGTGGGTGGGCTTAGCCCACGC – 3′
13	5′ – GCGTGGGTGGGCGATTAGCCCACGC – 3′
16	5′ – GCGTGGGTGGGCGACTATTAGCCCACGC – 3′
22	5′ – GCGTGGGTGGGCGACGCACAGTATTAGCCCACGC – 3′

Table 3.4 Composite DBSs for cooperative binding of F2 and F3 of Z23. The sequences of the different DBSs used to construct B2H reporter strains are given in the right column. The six base-pair binding sites (half-sites, 5' – GCGTGG – 3') recognized by the ZFs in each Ik-Z23-GP monomer are highlighted in red and green where the red triplet is recognized by F2 and the green triplet is recognized by F3 of Z23. The spacing between two half-sites is shown in the left column (indicated as numbers of base-pairs present in between the two half-sites). Note that only one DNA strand is shown and the second half-site is actually on the other strand (indicated by a reverse complement display of the actual half-site).

We suspected that our setup configuration might not provide the right geometry for the Ik-Z23 to bind cooperatively. Thus, we decided to adjust the configuration of the Ik-Z23 protein by optimizing the linker region between the Ikaros DZF domain and the Z23 DNA-binding domain. Different linkers of various length and flexibility (Table 3.5) were introduced since the geometry of our dimeric DNA-binding complex was unknown.

Linker name	Linker sequence
L1	RFHMSGGSRVRTGSKTPPHERP
L2	AAA
L3	GGGGS
L4	RFHMS
L5	RFHMSGGGGS
L6	TGEKP

Table 3.5 Linkers used to connect the Ikaros DZF domain to the Z23. The right column shows the amino-acid sequences of the various linkers. Residues that represent the authentic C-terminal end of the Ikaros protein are highlighted in red and were only present in L1, L4 and L5. L2, L3 and L6 were directly fused to the final zinc-binding histidine of the DZF domain.

Plasmids encoding the different linker constructs were introduced into each of the B2H reporter strains and β-galactosidase activities were assessed. None of the various proteins in

combination with the different binding sites we tested was able to activate *lacZ* (data not shown). However, linker L6 (canonical TGEKP-type linker) improved the ability of Ik-Z23-GP to bind to the 0 site which stimulated transcription of *lacZ* gene nearly two-fold (Figure 3.24). Again, this effect could not be improved by adding more IPTG (thereby increasing the amount of protein, data not shown).

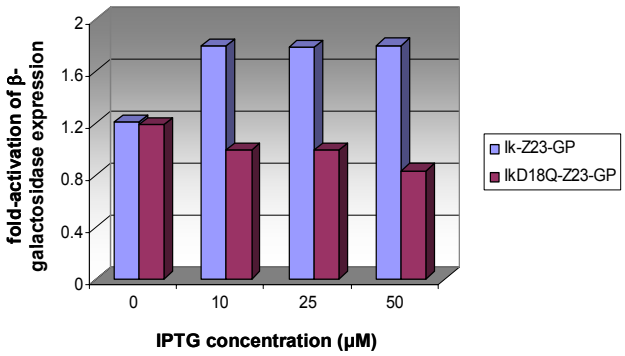

Figure 3.24 Analysis of DZF mediated dimerization of DNA-binding ZFs. Plasmids encoding Z23-GP fused to either the wild-type Ik DZF domain (blue bars) or the Ik DZF domain harboring the D18Q mutation (purple bars) together with the RNAP α-Gal4 expression plasmid were introduced into the 0-site B2H reporter strain harboring a composite DBS (see Table 3.4) and activity of these constructs was assessed over a range of IPTG concentrations.

To further verify that this increased activation of *lacZ* depends on a cooperative binding of Ik-Z23-GP mediated by dimerization, the D18Q mutation was introduced into the attached DZF domain. The effect of this mutation was assessed in the 0 site reporter using various concentrations of IPTG. As shown in Figure 3.24, D18Q completely abolished the ability of Ik-Z23-GP to activate *lacZ*. This suggested that the DZF domain mediated a cooperative binding of Ik-Z23 to the two half-sites by dimerizing two Ik-Z23 molecules. We concluded that it was in principle possible to use the DZF domains as a dimerization element to assemble DNA-binding ZFs on specific DNA-sites although the cooperative binding of Ik-Z23 was not very strong. This again demonstrated the versatility of the C2H2 ZF motif.

3.8 Discussion

3.8.1 DZF-derived C2H2 ZFs can be "mixed and matched"

The principal goal of DNA-binding ZF design was to engineer ZFs that can bind any DNA site of interest. To achieve this, different groups took advantage of the simple modular structure of DNA-binding ZFs and applied various methods including modeling, sequence

comparison and selections to alter the specificity of individual fingers within a multifinger protein. These fingers together with naturally occurring ZFs can then be combined by "mixing and matching" to create designer ZFs capable of recognizing novel DNA sequences. In addition, various ZFs can be linked together into tandem arrays that are capable of recognizing extended DNA sequences. Before this study, it was not known whether protein-binding ZFs could be mixed and matched as well in order to create synthetic proteins with novel interaction specificities. It has previously been shown that the two ZFs in the Pegasus and Eos DZF domain are separable and capable of folding independently of each other (Westman *et al.*, 2004). These data suggested that ZFs in the DZF domain may behave in a modular manner. In addition, a functional hybrid DZF domain consisting of two individual ZFs derived from different wild-type proteins has been described (McCarty *et al.*, 2003). Thus, we decided to systematically investigate whether C2H2 ZFs and DZFs could be "mixed and matched" to create domains with novel interaction specificities. Our results showed that shuffling of DZF-derived C2H2 ZFs can yield synthetic DZFs with new specificities that can then be linked together into more extended interaction interfaces. Interestingly, only a small number (less then 0.5%, 26 selected DZF pairs/5344 total potential pairs) of the potential combinations of DZFs were identified as positive for interactions. Since the theoretical number of combinations in all libraries was completely over sampled, and because our sequencing results revealed that we identified many of the interacting pairs multiple times, we are confident that we identified nearly all interacting DZFs. In addition, attempts to create synthetic DZFs without using the B2H selection strategy by simply constructing all possible combinations of C2H2 ZFs and linkers from the Ikaros and *Drosophila* Hunchback DZFs were unsuccessful and none of these synthetic domains could mediate interaction with each other. Thus, we conclude that individual fingers in the DZF domain do not always operate as completely modular units and, as is the case for DNA-binding C2H2 ZFs (Isalan *et al.*, 1997; Isalan *et al.*, 1998; Wolfe *et al.*, 1999; Hurt *et al.*, 2003; see also section 1.2.5), context-dependent interactions are important for the binding affinity. This also emphasizes the importance and necessity of using selection methods to identify synthetic DZF pairs that are fully optimized for protein-binding.

3.8.2 Anti-parallel interaction mode

Analysis of the interaction specificities of the synthetic DZFs revealed a novel antiparallel interaction geometry for DZF domains. Previous studies suggested that DZF domains interact in a parallel fashion based on the interaction specificity of an engineered chimeric Ikaros/Hunchback DZF (McCarty *et al.*, 2003). However, the identities of our selected synthetic DZFs together with their homo- and heterotypic interaction specificities indicated that these domains interact in an anti-parallel manner. Furthermore, the relative interaction affinities of the double-DZFs are most consistent with a model in which the component synthetic DZFs interact in an anti-parallel fashion. It is noteworthy that our selections also identified a synthetic DZF pair that suggests a parallel interaction mode (Pe-Hd-Hd interacting with Hd-Hd-Hl). Thus, both parallel and anti-parallel interaction geometries seem to exist for synthetic DZFs which further highlight the functional versatility of C2H2 ZFs. It is worth noting that these interaction modes were defined for synthetic isolated DZFs and may not reflect the geometry of naturally occurring DZFs in the context of a full-length protein.

3.8.3 Applications of synthetic DZFs

Our synthetic DZF domains have proven to be efficient in mediating assembly of a bi-partite transcriptional activator capable of stimulating expression of the endogenous VEGF-A gene in human cells. This represents an example of synthetic control as we managed to by-pass the normal regulatory signals such as hypoxia to activate VEGF-A. Furthermore, because this activation depends upon the presence of two DZF-linked proteins, it provides a mechanism for making VEGF-A expression dependent on two inputs (as in an 'AND' gate circuit, Kramer *et al.*, 2004). These synthetic DZF domains can also be used to activate expression of a specific gene in a bacterial cell suggesting that they may also have implications for constructing synthetic circuits in bacteria.

The fact that our synthetic DZF domains mediate preferentially heterotypic interactions provides important advantages compared to naturally occurring wild-type DZFs. Heterotypic interactions may be particularly useful for applications requiring asymmetric complex assembly. In fact, these synthetic DZFs are more efficient in mediating assembly of a bi-partite transcription factor in our "activator reconstitution" assay than for example naturally

occurring homotypic DZFs. On explanation for this observation could be that the formation of unwanted homodimers of DNA-binding domain fusions (or activation domain fusions) is less likely to occur since our domains prefer heterotypic interactions. These undesired homodimers could compete with the formation of the wanted DNA-binding domain/activation domain heterodimer and would thereby impair the ability of the DZF domains to mediate activation of VEGF-A.

Furthermore, our synthetic DZF domains demonstrated a high level of specificity which may have helped avoiding any unwanted cross-interactions with endogenous competitors. In fact, our synthetic DZFs were functional in three cellular compartments (i.e. the cytoplasm of bacterial and mammalian cells and the nucleus of mammalian cells) indicating that none of the potential competitors present in these cellular contexts was able to interfere with the interactions. Moreover, the specificity did not depend on over-expression of the synthetic DZF domains since both lowering the expression levels of these DZFs and over-expression of potential interfering proteins did not impair the ability to interact with defined partners in human cells.

Another notable feature of our synthetic DZF domains is that they can be linked together into more extended arrays. This suggests that DZF domains can be used as modules to create synthetic multi-DZF 'scaffold' or 'adaptors' upon which various DZF-linked proteins might be assembled. Our initial attempts to design a transcriptional scaffold consisting of two transcriptional activators has proven to be challenging and the obtained activation of the VEGF-A expression was very weak. It will be interesting to explore whether it is possible to engineer a scaffold in the cytoplasm of a human cell by applying our synthetic DZF domains. Such a scaffold could be used to create novel synthetic signaling pathways by simple tethering defined signaling molecules of interest (Park *et al.*, 2003; Harris *et al.*, 2001).

3.8.4 Future directions

Finally, these findings have important implications for future prospects to design C2H2 ZFs capable of interacting with any target protein of interest. The versatility and modularity of the DZF domain suggest potential strategies for re-engineering zinc finger protein-protein interfaces. Thus, like for their DNA-binding counterparts, it may be possible to randomize residues within the DZF domain that are known to be important for mediating protein-protein interactions (see Chapter 4) to create C2H2 ZFs with completely novel specificities. Such an

approach depends on a precise structural and biochemical information about the interaction surface. Mutagenesis as well as structural analysis of DZF domains is necessary and will help narrow down the choice of residues feasible for randomization (see Chapters 4 and 5). Resulting re-engineered ZFs in combination with naturally occurring ZFs could then be shuffled and linked together to create finger arrays with novel interaction specificities.

In the long-term, the capability to engineer synthetic C2H2 ZFs with desired protein-protein interaction specificities, together with existing strategies for engineering designer C2H2 ZF DNA-binding proteins, should yield a powerful toolbox useful to construct artificial cellular networks. Hence, the development of these technologies should have significant applications in synthetic biology as well as biomedical research.

Chapter 4. Genetic analysis of various DZF domains using mutagenesis.

4.1 Introduction

Despite the functional importance of the DZF as a novel protein recognition domain, relatively little is currently known about how this domain mediates protein-protein interactions. DZFs exist in several proteins which all display substantial sequence homologies at the amino acid level. This suggests that the DZF domain exhibits a uniform fold on which the dimerization presumably is based. Despite the high degree of homology, studies have shown that the interaction mediated by distinct DZF domains have differences in specificities. Recent studies by McCarty and coworkers (2003) have provided initial molecular determinants of the DZF domain interactions. The results of these studies demonstrated that dimerization mediated by the DZF domain of Ikaros and Hunchback is highly specific since both domains showed homodimerization but were not able to heterodimerize with each other. Furthermore, by analyzing a series of Ikaros-Hunchback DZF chimeras for their abilities to interact with either wild-type Hunchback or wild-type Ikaros, McCarty and co-workers (2003) defined the boundaries of specificity determining residues (see section 1.3.4.3 and Figure 1.6 for details). To localize "specificity determinants" on the interaction surface of the Ikaros DZF domain, McCarty and coworkers constructed several double or single "swap residue" mutants based on the amino-acid sequence of the Hunchback DZF domain. Residues that are different in Hunchback were introduced into the Ikaros DZF at the corresponding position. The ability of these "swap" mutants to interact with the wild-type Ikaros DZF were tested using co-immunoprecipitations. Mutants identified in this initial screen were then assessed for homo-dimerization using chemical crosslinking assays. This experiment identified eight residues that seemed to be important for Ikaros mediated dimerization (McCarty *et al.*, 2003; see also section 1.3.4.3 and Figure 1.6). Although this study provided some insights into the molecular mechanism through which DZF domains mediate specific protein-interactions, additional studies of the protein interface are required to define and distinguish residues responsible for both specificity and affinity on the uniform defined fold likely to be present in all DZF domains.

To begin to address these points, a detailed mutational analysis of the Ikaros, Pegasus and Hunchback *Drosophila* DZF domain interfaces was performed. We chose these three DZFs because a comparison of the amino acid sequences of these domains reveals both similarities and differences (Figure 4.1). In addition, they display very selective interaction profiles: Although all three DZF domains can mediate homodimerization, Hunchback and Ikaros fail to heterodimerize with each other. Pegasus on the other hand can interact with both Ikaros and Hunchback (see section 3.2.4 for details). We reasoned that detailed analysis of these three domains should help define the nature of specific homo- and hetero-typic interactions at the interface.

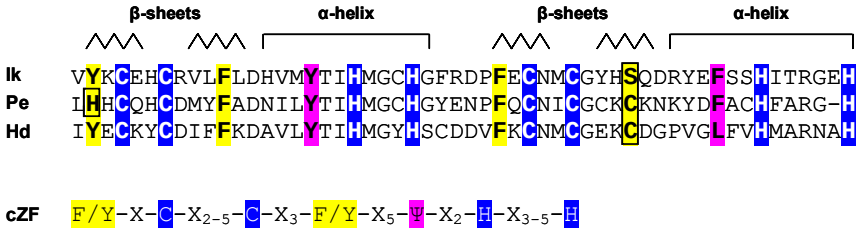

Figure 4.1 Comparison of the amino acid sequences of the DZF domain from Ikaros (Ik), Pegasus (Pe) and Hunchback *Drosophila* (Hd). Conserved cysteines and histidines involved in zinc binding are highlighted in blue. Residues that are in positions that are normally highly conserved in classical C2H2 ZFs are highlighted in yellow (Phe/Tyr) and pink (Ψ, hydrophobic residues). The consensus sequence motif of a classical DNA binding C2H2 ZF (cZF) is presented below the DZF sequences for comparison. Deviations from the consensus sequences are indicated by boxes. Secondary structures defining the DNA binding C2H2 ZFs are indicated on top of the amino acid sequences.

To investigate systematically which amino acids are critical for dimerization, the DZF domains of Ikaros, Hunchback and Pegasus were studied using alanine scanning mutagenesis. In general, this approach is used to identify functionally important residues of a protein by sequentially substituting each residue with an alanine and assessing the effect of this replacement. Substitution of the DZF domain residues by alanine may identify contact sites for homodimerization because protein-protein interaction surfaces are known to be accomplished by side-chain interactions between certain amino-acids (Ashkenazi *et al.*, 1990). Since alanine only has a CH_3 group instead of a longer side chain it does not alter the main-chain conformation. Moreover, alanine is the most prevalent amino acid found at positions buried inside the protein as well as at exposed positions of the protein structure. Thus, alanine is not expected to destroy the overall structural features of the protein, and

therefore provides an ideal replacement residue (Cunningham and Wells, 1989; Thukral *et al.*, 1991).

In another set of experiments, a similar strategy as previously described was applied by constructing several "swap" mutants for the Ikaros and Hunchback DZF domains (McCarty *et al.*, 2003). Since Ikaros and Hunchback DZFs do not heterodimerize with each other, we reasoned that swapping a residue important for making contact in one DZF with the corresponding residue in the other DZF domain is likely to disrupt this specific contact and thereby the ability to dimerize. Our hope was that this approach would help identifying specific residues that are important for the dimerization of e.g. the Ikaros DZF but not for Hunchback and vise versa. Thus, this approach in combination with the alanine scanning mutagenesis should help determining residue positions that are necessary for specific DZF mediated interactions.

4.2 DZF domain mediated homodimerization can be studied in the Bacterial one-hybrid system

4.2.1 The Bacterial one-hybrid (B1H) system

In addition to the B2H system a B1H system has been described which can be used to study and identify dimeric proteins (Hu *et al.*, 1990). This system is based on the observation that the N-terminal DNA binding domain (NTD) of the bacteriophage lambda repressor (λcI) requires a dimerization domain in order to bind efficiently to its specific DNA site, the λ operator. Thus, fusing a domain of interest (X) to the λcI NTD will permit the λcI-X fusion protein to bind to the λ operator as long as X can mediate dimerization. The dimerization can be detected by using an appropriate *E. coli* reporter strain. As shown in Figure 4.2, in this strain, the λ operator is placed in between the -10 and -35 region of a strong promoter and binding of the λcI-X fusion protein to its site will repress the promoter by blocking binding of the RNAP. By using *lacZ* as the reporter gene, repression of the promoter can be easily assessed performing quantitative β-galactosidase assays.

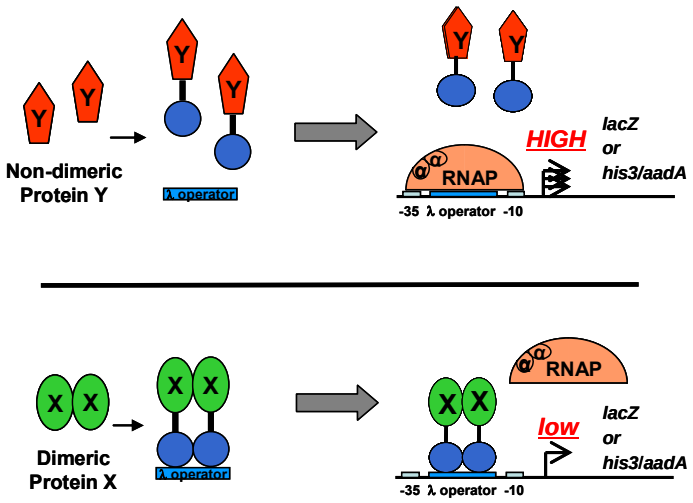

Figure 4.2 Schematic of the Bacterial one-hybrid (B1H) system. A dimeric protein X can mediate dimerization of the N-terminal DBD of λcI (λcI-NTD, blue circles) resulting in the binding of the λcI-X fusion protein to the λ operator site (lower panel, left side). In an appropriate B1H reporter strain the λ operator is placed between the -10 and -35 sequences. Thus, binding of the dimeric λcI-X fusion protein competes with binding of the RNAP which results in a repression of the reporter gene expression (lower panel, right side). In contrast, non-dimeric proteins fused to λcI-NTD will abolish the ability of λcI to bind to the λ operator site, thus resulting in an activation of reporter gene expression (upper panel). This figure was kindly provided by K. Joung.

4.2.2 Validation of the B1H system for studying homodimeric DZF domain interactions

To initially test whether homodimeric interactions of the DZF domain can be detected in the B1H system, we generated plasmids encoding the DZF domains from Ikaros (work performed by R. Fang), Pegasus and *Drosophila* Hunchback fused to the C-terminus of the λcI DBD (consisting of the NTD and the linker that connects the NTD with the naturally C-terminal domain of λcI). Each of these plasmids (encoding the λcI-DZF fusion proteins) was transformed into a B1H reporter strain containing a λcI operator positioned between the -35 and -10 hexamers of the lac UV5 promoter. To test for homodimerization and DNA-binding mediated by these different DZF domains, β-galactosidase assays were performed. As expected, it was found that cells which expressed the λcI NTD+linker alone (lacking the C-terminal λ dimerization domain) showed high levels of transcription of the *lacZ* reporter gene. In contrast, transcription of *lacZ* was repressed in cells expressing the λcI-DZF fusion proteins suggesting that all three DZF domains mediate efficient homodimerization of the chimeric protein (Figure 4.3A).

To further test, if the repression of *lacZ* is dependent on the expression of the different λcI-DZF fusion proteins, an IPTG titration experiment was performed where the expression of the proteins was induced using various amounts of IPTG.

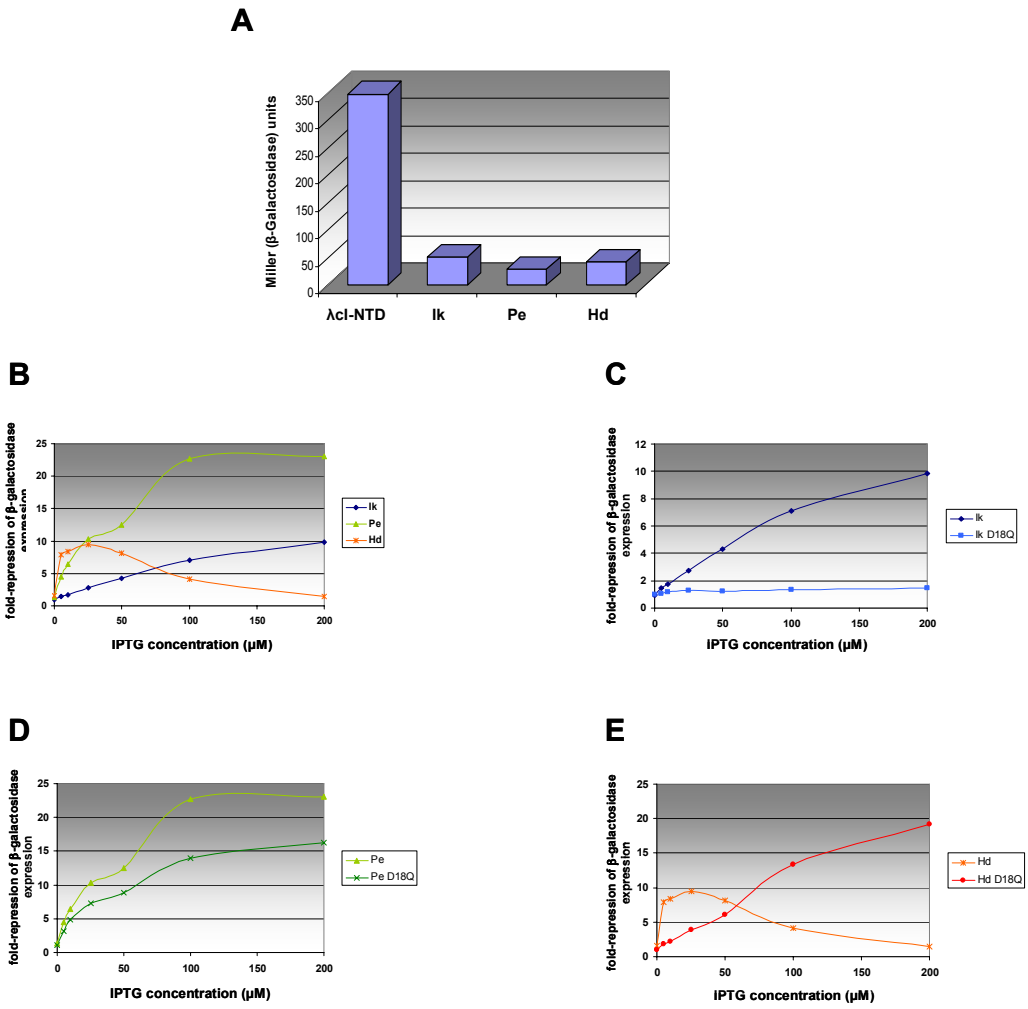

Figure 4.3 Analysis of the Ikaros, Pegasus and Hunchback *Drosophila* DZF domains in the B1H system. (A) B1H reporter strains expressing the three DZF domains fused to λcI were assayed for β-galactosidase activity. A control expressing the λcI-NTD protein is also shown. In this assay an IPTG concentration of 50 μM was used for inducing the expression of the fusion proteins. **(B, C, D** and **E)** The activity of these wild-type constructs together with constructs harboring the D18Q mutation in the DZF domains were assessed using increasing concentrations of IPTG. The three wild-type proteins **(B)**, wild-type and D18Q Ikaros **(C)**, wild-type and D18Q Pegasus **(D)** as well as wild-type and D18Q Hunchback **(E)** were group wise compared to each other.

As shown in Figure 4.3B, repression of *lacZ* was directly related to the protein expression level as long as the IPTG concentration stayed below a certain threshold. Interestingly, the effect of increased IPTG concentration was different for the individual fusion proteins. The Pegasus DZF domain reached a maximum level of repression at an IPTG concentration of

100 μM, whereas for the Ikaros DZF domain the repression level increased linearly with the IPTG concentration even up to high levels of IPTG (Figure 4.3B and data not shown, R. Fang). Overexpression of the Hunchback DZF domain was toxic to the bacterial strains and the fusion protein could not mediate repression when the IPTG concentration went above 25 μM. The toxicity may have forced the bacterial cells to select for a Hunchback mutant in order to survive and grow. This would explain why these cells did not repress *lacZ* when high levels of proteins were present. On the other hand, these hybrid proteins might have dimerized detached from the DNA and were therefore not able to mediate transcriptional repression at the promoter.

To further validate the B1H system as a genetic method to study DZF mediated homodimerization an additional series of constructs was designed where the previously described D18Q mutation (McCarty *et al.*, 2003) was introduced into the DZF domains. Plasmids encoding this mutation were transformed into the B1H reporter strain and β-galactosidase assays were performed using cultures grown in various concentrations of IPTG. It was found that the three proteins showed a different response to the D18Q mutation at various concentrations of IPTG. For the Ikaros DZF domain, the ability of the protein to repress *lacZ* was completely abolished by the D18Q mutation even at very high levels of IPTG (Figure 4.3C, R. Fang). In contrast, overexpression of the mutated Hunchback DZF domain abolished repression only at IPTG concentrations below 10 μM, indicating that the effect of the mutation is very sensitive to the amount of expressed protein (Figure 4.3E). In the case of the Pegasus DZF domain, the D18Q mutation resulted in a small reduction of repression (Figure 4.3D) which is consistent with the data obtained in the B2H system (see section 3.4.1 and data not shown).

In summary, the results of these assays demonstrated that the repression observed in the B1H assay is due to expression of the different domains since the level of repression depends on the amount of protein expressed in the cell. Furthermore, the D18Q mutation affected dimerization of Ikaros and Hunchback but not of Pegasus which was also found when this mutation was tested in the B2H system (see section 3.4.1). These results confirm the validity of using the B1H system to analyze dimerization mediated by the DZF domain.

4.3 Analysis of the Ikaros DZF domain

4.3.1 Alanine scanning mutagenesis

To investigate systematically which amino acids of Ikaros are critical for dimerization, the Ikaros DZF domain was studied using alanine-scanning mutagenesis (work performed by R. Fang). This approach should reveal residue positions that contribute to dimerization.

To perform an alanine-scan of the Ikaros DZF domain a set of plasmids encoding mutant proteins was generated and assayed using the B1H system. PCR-mediated mutagenesis was used to introduce individual alanine substitutions at 47 different residue positions within the Ikaros DZF domain except at conserved positions that mediate zinc binding (cysteines and histidines). To analyze the effects of these mutations on homodimerization, plasmids were constructed encoding each of these 47 different mutants fused to λcI and introduced into the reporter strain to measure their β-galactosidase activity (R. Fang).

Figure 4.4 shows the results of the alanine-scan. The cutoff for defining a dimer-defective mutant was set to >1.5 fold the β-galactosidase activity obtained with wild-type Ikaros DZF domain. This cutoff was chosen to include the standard error of these experiments which is typically about 10%. Thus, mutants displaying β-galactosidase units 1.5-fold higher than wild-type were considered to possess a significant dimerization defect. Mutations in 21 of the 47 positions tested displayed β-galactosidase units comparable to that of the wild type (<1.5 fold activation), suggesting that those residues are not directly involved in dimerization. 10 of these alanine substitution mutants exhibited slightly higher dimerization activity. Alanine substitutions in the remaining 26 positions tested significantly affected dimerization. The nature and importance of the various residue positions are described in more detail below.

Predicted conserved hydrophobic residues:

Alanine substituted mutants Y07A, F16A, Y22A, F35A and F50A affected residues which are likely to form the hydrophobic core of the C2H2 ZF since these residue positions are defined as conserved in the typical C2H2 ZF motif (Figure 4.1, see also section 1.2.3). These mutants were all severely defective in dimerization. It is noteworthy that residue position 44 is also defined as highly conserved in the classical DNA binding C2H2 ZF and is typically a

phenylalanine. Interestingly, this residue is substituted by a serine in the Ikaros DZF domain and was not affected by an alanine mutation (see also section 4.7.3 below).

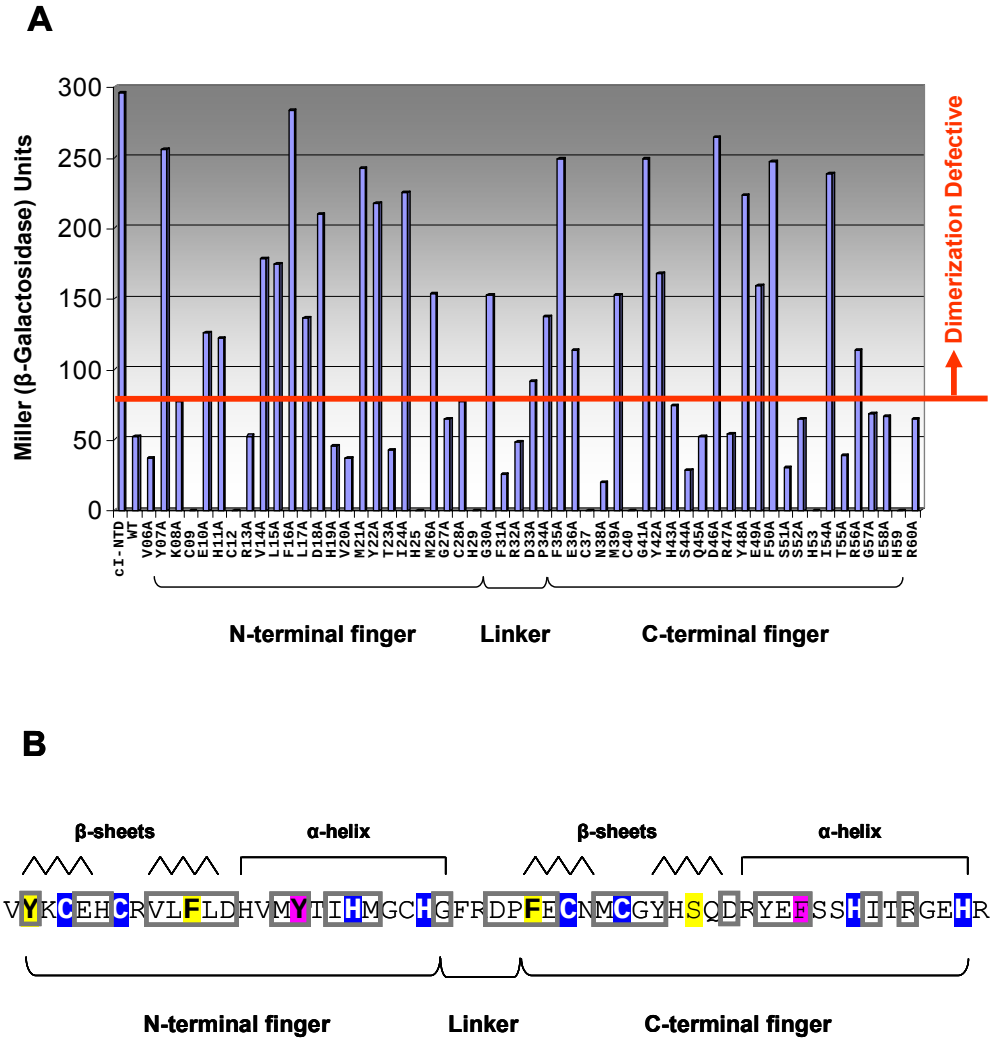

Figure 4.4 Alanine scan analysis of the Ikaros DZF domain in the B1H system. (A) Ikaros DZF mutants were tested for their ability to mediate homodimerization in the B1H system by performing β-galactosidase assays. The cutoff for defining dimer-defective mutants was set to >1.5-fold the β-galactosidase activity of wild-type Ikaros and is indicated by the red line. This cutoff was chosen to take account of the standard error of these experiments which is typically about 10%. Residues replaced by an alanine are shown below the graph and each substitution corresponds to the respective bar. Note that conserved cysteines and histidines are also shown but were not mutated. Residues were numbered according to McCarty *et al.*, 2003 where the conserved tyrosine 2 residues before the first Zinc-ligating cysteine is defined as residue position 7. Controls expressing λcI and wild-type Ikaros are also shown. Regions defining the N-terminal ZF, linker and C-terminal ZF are indicated on the bottom of the residue sequence. Values are presented as the mean calculated from three independent experiments. **(B)** Summary of residues that were affected by an Alanine-substitution. Zinc ligating cysteines and histidines are highlighted in blue and conserved residues are highlighted in yellow (Phe/Tyr) and pink (hydrophobic residues). Residues that abolished dimerization when mutated are boxed. Typical secondary structures are indicated above the amino-acid sequence. Figure 4.4A was kindly provided by K. Joung.

Predicted β-sheets:

Mutants E10A, H11A, V14A, L15A, L17A and D18A lie in the predicted ββ-sheet region of the first C2H2 ZF motif and were all dimerization defective, whereas mutants K08A and R13A did not affect dimerization. The ββ-sheet region in the second ZF contained five more residue positions that were affected by an alanine substitution (E36A, M39A, G41A, Y42A and D46A). Although we can not rule out that G41 participates in the interaction, we suspect that because of its features this residue is likely to play a structural role by for example providing flexibility in the β-sheet region of the second ZF. Mutations in the other residues in this region (N38A, H43A, S44A and Q45A) did not affect dimerization.

Predicted α-helices:

Residues 19-29 and 47-59 in fingers 1 and 2, respectively, are predicted to form the α-helix. Mutants M21A, I24A, M26A, I54A and R56A in these regions significantly disrupt the dimerization while mutations in the other residues of the α-helices (H19A, V20A, T23A, G27A, C28A, R47A, S51A, S52A, T55A, G57A and E58A) did not affect dimerization.

Inter-finger linker:

The region between the two individual C2H2 zinc fingers is termed the linker region and contains a highly diverse sequence throughout the different DZF domains. Mutations in three of these residues G30A, D33A and P34A, disrupted dimerization whereas mutations F31A and R32A had no effect.

Overall, this preliminary study of the Ikaros DZF domain identified several residue positions that may contribute to dimerization. Theses residues are spread out throughout the whole DZF domain and could be directly involved in the interaction but could also affect the affinity indirectly, by for example changing the confirmation or stability of Ikaros. In fact, Western Blot analysis of the expression level of these mutated peptides in *E. coli* indicated that many of the mutants that are severely impaired in their ability to dimerize, are expressed at lower levels compared to the wild-type Ikaros DZF domain (R. Fang, personal communication).

4.3.2 Residue "swap" scanning mutagenesis

To further narrow down specific residue positions in Ikaros important for dimerization, a similar strategy already described by McCarty and co-workers (2003) was taken. In this study, proteins harboring either single or double mutation were tested for their ability to heterodimerize with wild-type Ikaros using co-immunoprecipitation assays. In contrast, we decided to systematically analyze DZFs harboring single "swap" mutations for their ability to homodimerize, which is expected to be more stringent then examining heterodimerization with the wild-type protein (note that McCarty and co-workers further analyzed their identified mutants for homo-dimerization using chemical crosslinking assays which is a less stringent than e.g. co-immunoprecipitation). To do this, we constructed a large series of 32 Ikaros "swap" mutants in which single residues in the Ikaros DZF domain were replaced with residues that correspond to the same position in the Hunchback DZF domain (residue "swap" mutants, Figure 4.5A). All mutants were expressed as a fusion to λcI in the appropriate reporter strain and assessed for their capability to homodimerize by performing β-galactosidase assays (Figure 4.5B, work performed by R. Fang).

The results are summarized in Figure 4.5C and agreed well with the previously described result. Mutations in 18 of the 32 positions tested displayed β-galactosidase units comparable to that of the wild type (<1.5 fold activation), suggesting that these positions are not important for dimerization. Swap mutations in the other 13 positions tested abolished dimerization. Interestingly, five of these mutants (Q45D, D46G, R47P, Y48V and E49G) are adjacent to one another and seem to form a patch in the second ZF. Q45D lies at the end of the predicted ββ-sheet, while the remaining residues (D46G, R47P, Y48V and E49G) are part of the predicted α-helix (with residue D46G representing position -1 of the α-helix). Note that position F50 is also affected by the swap mutation but is defined as conserved in the typical C2H2 ZF motif and is therefore likely to play a structural role in the C2H2 ZF fold. The remaining eight defective mutations are distributed throughout the DZF domain with mutations E10K, L17K and Y42E affecting residues lying in the predicted ββ-sheets, while mutations M21L, C28Y, G30S and G57N destroy residues that are part of the predicted α-helices. Only one residue position in the linker region was found to be impaired by a mutation (P34V). In summary, this study of the Ikaros DZF domain exhibits additional insights into residues important for dimerization.

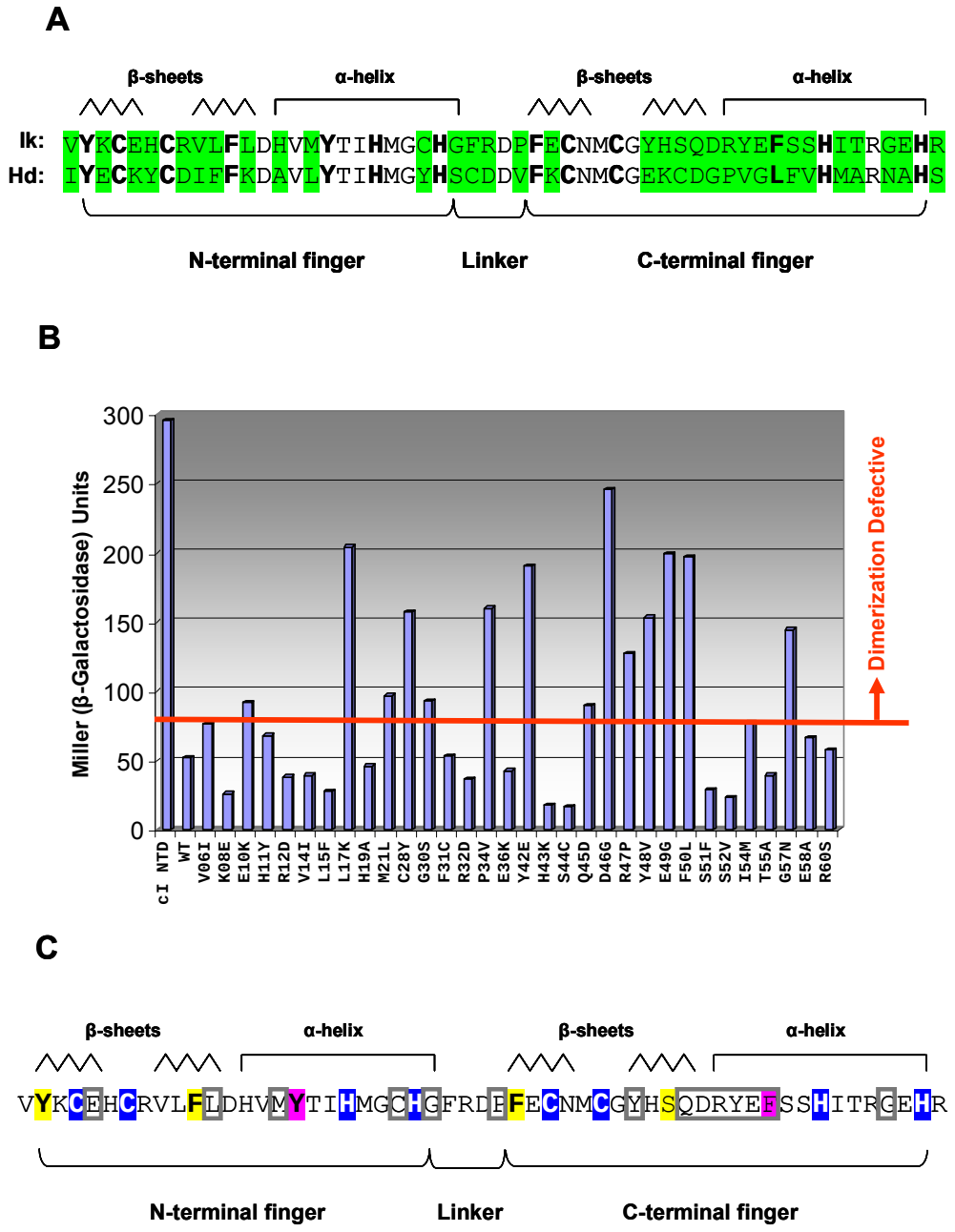

Figure 4.5 "Swap" scan analysis of the Ikaros DZF domain in the B1H system. (A) Sequence alignment of the Ikaros and Hunchback DZF domain. Residues that are different in both DZFs are highlighted in green and were used to construct "swap" mutants. See text for details. **(B)** "Residue swap" mutants were tested for their ability to abolish dimerization by introducing them into the B1H reporter strain and assessing their β-galactosidase activities. The cutoff for defining dimer-defective mutants is as defined in Figure 4.4 indicated by a red line. "Swap" mutations are indicated below the graph. Values represent the mean of three assays. **(C)** Summary of residues that were affected by a "swap" mutation. Zinc ligating cysteines and histidines together with conserved residues are highlighted in blue, yellow, and pink. Residue positions that were affected by a "swap" mutation are boxed. Predicted secondary structures are indicated above the amino-acid sequence and individual ZF are shown below. Figure 4.5B was kindly provided by K. Joung.

4.3.3 Comparison of mutants identified by alanine scan and "swap" scan analysis

32 of the 47 alanine scan mutants bore substitutions at residue positions also tested in the swap mutation experiment described above. Comparison of the ability to dimerize for these 32 pairs of mutants reveals that 23 pairs show the same dimerization phenotype (Figure 4.6). Substitutions in 13 of these residue positions consistently showed no effect on dimerization and are certainly not part of the interaction surface whereas mutations in the remaining ten positions significantly affected the ability to dimerize. However, nine residue positions yielded different results depending on which amino acids were used for substitution. For example, five positions were only impaired by an alanine replacement whereas for the other four positions only the "swap" mutation had an impact on dimerization. This suggested that in some cases the nature of the substituted side chain has to be considered since it may have an effect on the confirmation or the chemical properties of the whole protein.

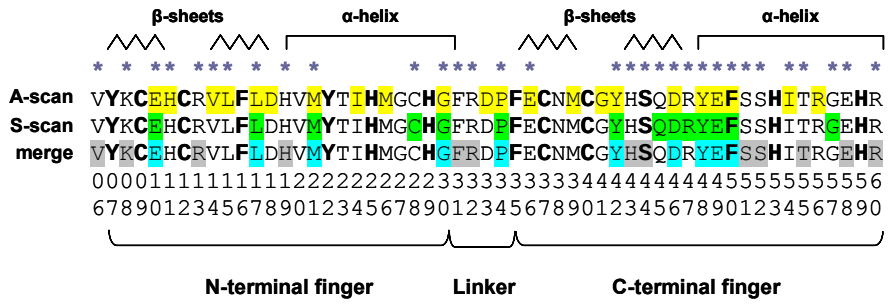

Figure 4.6 Comparison of the residues identified by alanine scan and "swap" scan analysis as important for dimerization of the Ikaros DZF domain. The results are projected on the amino acid sequence of the Ikaros DZF domain which is shown three times. Residues identified by alanine scan (A-scan) analysis are highlighted in yellow and residues identified by "swap"-scan (S-scan) analysis are highlighted in green. A merge of both results is shown below, where residues identified as important in both scans are highlighted in turquoise while residues that consistently showed no effect are highlighted in grey. Blue asterisks mark residue positions that were analyzed by both scans. Bold letters indicated conserved residues. Numbering of residues is shown as defined in Figure 4.4. Individual fingers and secondary structures are also shown.

The remaining 15 of the alanine-scan mutants introduced substitutions at residue positions that are the same in both the Ikaros and Hunchback DZF domains and were therefore not tested in the residue "swap" experiments. 11 of these residue positions (Y7, F16, D18, Y22, I24, M26, D33, F35, M39, G41, and R56) were strongly affected by a mutation. These residues may participate in contacts at the dimer interface but could also destabilize the protein and therefore its ability to dimerize.

4.4 Analysis of the Hunchback DZF domain

4.4.1 Alanine scanning mutagenesis

An alanine scan was also performed with the DZF domain of *Drosophila* Hunchback PCR-mediated mutagenesis was applied to introduce individual alanine substitutions at 49 different residue positions within the DZF domain except at positions occupied by the conserved cysteines and histidines. This set of mutant proteins was then assayed using the B1H system.
As demonstrated in Figure 4.7, mutations in 18 of the 49 positions tested displayed β-galactosidase units comparable to that of wild-type, suggesting that those residues are not directly involved in dimerization. Three of these alanine substitution mutants exhibited slightly higher dimerization activity. Alanine substitutions in the remaining 31 positions tested impaired dimerization.

Predicted conserved hydrophobic residues:
Mutations in the conserved hydrophobic residues (Y07A, F16A, Y22A, F35A and L50A) all severely affected the dimerization as was the case for Ikaros. Interestingly, residue C44 was also impaired by an alanine. As mentioned above, this residue position is normally conserved in the C2H2 ZF motif (and is typically a phenylalanine) but no effect was detected for the Ikaros DZF domain when this residue was mutated.

Predicted β-sheets:
Seven mutations (K10A, Y11A, C12A, I14A, F16A, K17A and D18A) in the predicted ββ-sheet of the N-terminal finger and 5 mutations (K36A, C40A, G41A, D45A and G46A) in the ββ-sheet of the C-terminal finger disrupted dimerization whereas the remaining residue positions (E08, D13, N38, E42 and K43) in this region in both C2H2 ZFs seemed not to be impaired by alanine substitutions.

Predicted α-helices:
Affected residue positions in the predicted α-helices are L21, Y22, I24, M26, G27, Y28, V48, V52, M54, A55, R56 and A58 while the remaining positions (A19. V20, T23, S30, P47, G49, F51 and N57) in this secondary structure were not affected by alanine substitutions.

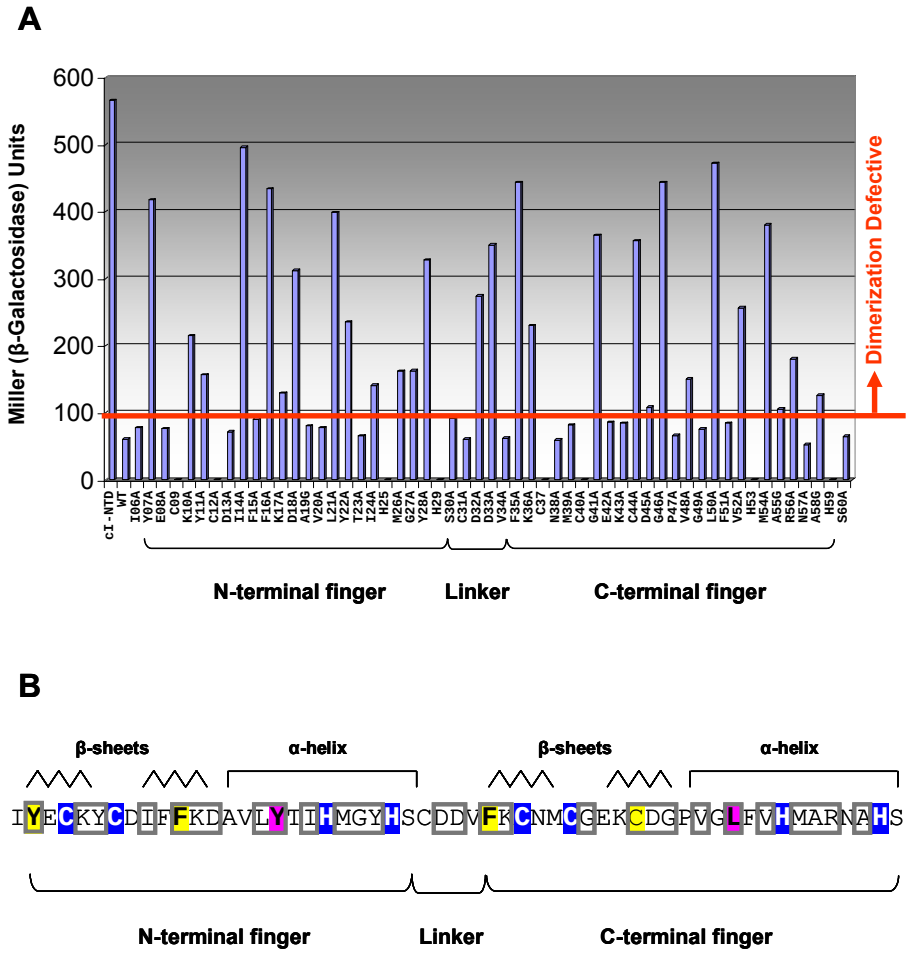

Figure 4.7 Alanine scan analysis of the Hunchback DZF domain in the B1H system. (A) Hunchback DZF mutations were tested for their ability to abolish homodimerization in the B1H system by performing β-galactosidase assays. The cutoff is indicated by the red line and defined as in Figure 4.4. Residues replaced by an alanine are shown below the graph. Note that alanine residues in the original DZF domain were always replaced by a glycine. Zinc binding cysteines and histidines are also shown but were not mutated. Residues are numbered as in Figure 4.4. Controls expressing λcI and wild-type Hunchback are also shown. Values represent the mean calculated from three independent assays. **(B)** Summary of residues that were affected by an alanine-substitution. Conserved residues are highlighted in blue, yellow and pink as described in Figure 4.1. Residues that effected dimerization when mutated are boxed. Predicted secondary structures and individual ZF with linker region are also indicated.

Inter-finger linker:

The linker region between the N-terminal and the C-terminal finger contained 2 additional mutations D32A and D33A that displayed severe dimerization defects.

Thus, this study identified several residue positions that may contribute to dimerization mediated by the Hunchback DZF domain. These residues could be directly involved in making contacts but could also be important for stability of the protein. Although some of

these residues are adjacent to one another, overall they are spread out throughout the whole domain.

4.4.2 Residue "swap" scanning mutagenesis

To further investigate dimerization mediated by the DZF domain "swap" mutation analysis was also applied to the *Drosophila* Hunchback DZF.

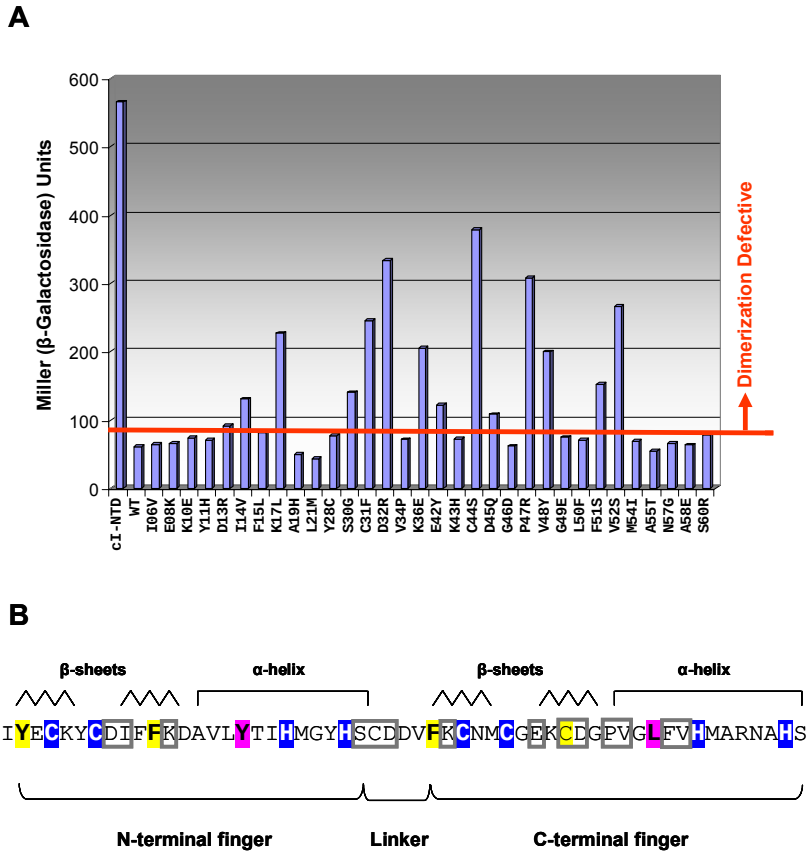

Figure 4.8 "Swap" scan analysis of the Hunchback DZF domain in the B1H system. (A) "Residue swap" mutants are defined as in Figure 4.5A and were tested for their ability to abolish dimerization by performing β-galactosidase assays. The cutoff for "dimerization defective" is indicated by a red line and defined in Figure 4.4. "Swap" mutations are indicated below the X-axis. Values represent the mean of three experiments. **(B)** Summary of residues that were affected by a "swap" mutation depicted on the amino acid sequence of the Hunchback DZF domain. Conserved residues are marked in blue, yellow and purple and residue positions that were impaired by a mutation are indicated by boxes. Predicted secondary structures and individual ZFs are also shown.

To do this, 32 Hunchback "swap" mutants in which single residues were replaced with the corresponding residues from the Ikaros DZF domain were constructed. As described for the Ikaros DZF domain, we used PCR mutagenesis to individually introduce these mutations

within the Hunchback DZF domain and subsequently assessed the ability of each of these swap mutants to homodimerize using the B1H system. It was found that 18 mutants yielded β-galactosidase activity <1.5 fold of the wild-type activity, suggesting that they did not affect dimerization of Hunchback (Figure 4.8). Mutations in the remaining 14 residue positions weakened or abolished the interaction (Figure 4.8). Some of the disruptive "swap" substitutions occurred within the inter linker region (S30G, C31F and D32R) and at the predicted transition between the second β-sheet and the α-helix of the C-terminal finger (C44S, D45Q, P47R and V48Y). The remaining seven defective mutations (D13R, I14V, K17L, K36E, E42Y, F51S and V52S) are distributed throughout the DZF domain. Overall, these 14 residues may participate in contacts at the dimerization interface that are responsible for specificity of the Hunchback DZF domain.

4.4.3 Comparison of mutants identified by alanine scan and "swap" scan analysis

32 of the 49 residue positions differ in both Ikaros and Hunchback and were tested in both the alanine scan as well as in the residue swap experiments. Comparing the dimerization phenotype for these 32 pairs demonstrated that 18 pairs behaved the same in their ability to dimerize. Mutations in eight of the positions caused dimerization defects whereas mutations in the remaining ten positions did not affect dimerization. However, substitutions in 14 residue positions displayed different effects when replaced with alanine or swapped with the corresponding Ikaros amino acid. Eight of these residue positions were affected by an alanine but not by a swap substitution, whereas six positions were only impaired by the "swap" substitution.

The remaining 17 alanine mutations are at residue positions that are the same in both the Hunchback and Ikaros DZF domains and were therefore not tested in the residue swap experiments. 11 of these mutations (Y7A, F16A, D18A, L21A, I24A, M26A, G27A, D33A, F35A, G41A and R56A) possessed significant dimerization defects and may provide additional information about residues involved in contacts at the dimer interface.

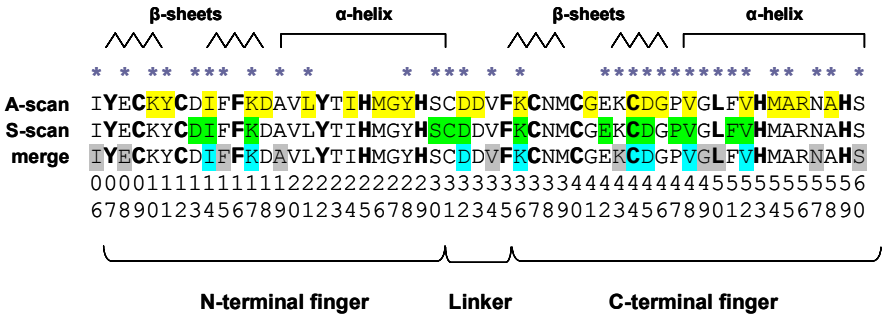

Figure 4.9 Comparison of the residues identified by alanine scan and "swap" scan analysis as important for dimerization of the Hunchback DZF domain. The results are summarized on the amino acid sequence of the Hunchback DZF domain. Residues identified by A-scan analysis are highlighted in yellow in the first sequence and residues identified by S-scan analysis are highlighted in green in the second sequence. A merge of both results is shown below, where residues identified as important in both scans are highlighted in turquoise while residues that consistently showed no effect are highlighted in grey. Blue asterisks mark residue positions that were analyzed in both scans. Conserved residues are marked in bold letters. Numbering of residues is shown as defined in Figure 4.4. Individual fingers and typical secondary structures are also shown.

4.5 Analysis of the Pegasus DZF domain by alanine scanning mutagenesis

To perform an alanine scan of the Pegasus DZF domain, plasmids encoding 47 mutant proteins were generated as described for Ikaros and Hunchback. As before, this set of mutants was subsequently assayed using the B1H system.

As shown in Figure 4.10, three types of mutants emerged from the alanine-scan. Mutations in 22 of the 47 positions tested were silent having no effect on dimerization. Ten of the alanine substitution mutants exhibited higher dimerization activity and seemed to stabilize the protein or the interaction. Alanine substitutions in the other 15 positions tested weakened or abolished dimerization. The most disruptive alanine substitutions affected the conserved hydrophobic residues (H07A, F16A, Y22A, F35A, C44A and F50A). Since these residue positions define the predicted hydrophobic core it is likely that mutations in those amino acids destabilize the protein and therefore indirectly affect dimerization. Another severe mutation affected position N46, the last residue in the predicted second β-sheet of the C-terminal ZF. The remaining 9 alanine substitutions only weakened the dimerization. Mutants H08A, H11A, D13A, M14A, Y15A and D18A are present in the predicted ββ-sheet region of the first zinc finger motif and mutant K43A lies in the ββ-sheet of the second zinc finger. An additional mutation at position L6 which is the first amino acid of the DZF domain abolished

dimerization. Neither any residue positions in the α-helices nor in the inter-finger linker were affected by an alanine substitution.

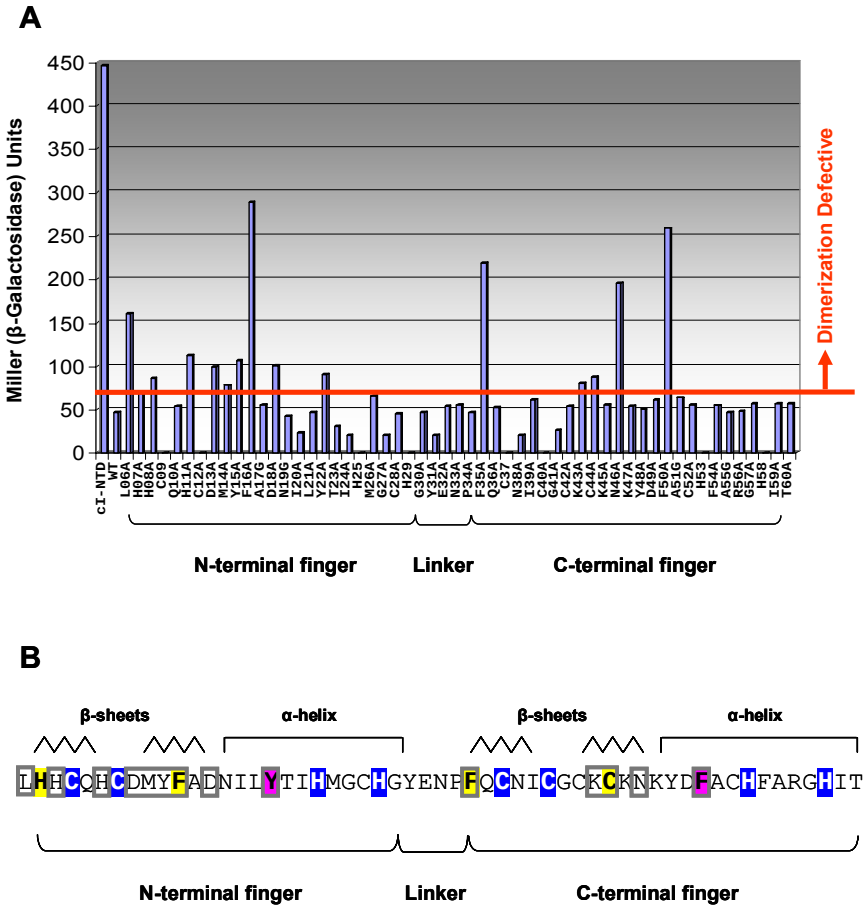

Figure 4.10 Alanine scan analysis of the Pegasus DZF domain in the B1H system. (A) Pegasus DZF mutations indicated below the graph were tested for their ability to abolish homodimerization in the B1H system by assessing their β-galactosidase activities. The cutoff is indicated by the red line and defined as in Figure 4.4. Note that alanine residues in the original DZF domain were replaced by a glycine. Zinc binding cysteines and histidines were not mutated. Residues were numbered as in Figure 4.4. Controls expressing λcI and wild-type Pegasus are also shown. Values represent the mean determined from three independent assays. **(B)** Summary of residues that were affected by an alanine-substitution depicted on the amino acid sequence of the Pegasus DZF domain. Conserved residues are highlighted in blue, yellow and pink as described in Figure 4.1. Residues that abolished dimerization when mutated are boxed. Secondary structures and individual ZFs with linker region are also indicated.

Overall, the alanine scan mutagenesis for the Pegasus DZF domain revealed only a few important amino acids which most likely are structural residues. Individual point mutations seemed not to be effective enough to disrupt dimerization of the Pegasus DZF domain. Note that the Pegasus DZF domain was not further analyzed by "swap" scan analysis. Since it appears that the interaction mediated by Pegasus is not selectively specific towards

Hunchback and Ikaros (Pegasus can mediate interaction with both Hunchback and Ikaros), swapping in residues from these two proteins will not offer new information about the specificity of this DZF domain-mediated interaction.

4.6 Comparison of results obtained for different DZF domains

4.6.1 Alanine scan mutagenesis for Ikaros, Hunchback and Pegasus

Comparison of the alanine scan results for the Ikaros and Hunchback DZF domains reveals partially overlapping but also distinct sets of residues important for homodimerization (Figure 4.11A). Besides the predicted conserved structural residue positions (Y7, F16, Y22, F35 and F/L50, bold in Figure 4.11A) there are several residue positions (10, 11, 14, 17, 18, 21, 24, 26, 33, 36, 41, 46, 48, 55 and 56) in both DZF domains that are affected by an alanine substitution. These residues are spread out across the DZF domain and are found in the predicted β-sheets, α-helices and the inter-finger linker region. Four mutations at residues 10, 11, 17 and 18 are adjacent to one another and lie on the β-sheets of the C-terminal zinc finger. However, 14 alanine-scan mutants yielded different results for the Ikaros and Hunchback DZF domain. Of the 14 positions, six (15, 30, 34, 39, 42 and 49) were only affected in the Ikaros DZF domain whereas eight positions (27, 28, 32, 44, 45, 52, 55 and 58) had the reverse phenotype and were only impaired in the Hunchback DZF domain. These residues may define positions that are specifically important for the respective DZF.

Including Pegasus in this comparison adds some supplementary information for understanding dimerization. In general the Pegasus DZF domain was only weakly affected by the alanine substitutions. Besides mutations in the conserved hydrophobic residue positions there were some additional mutations (L6, H8, D13 and K43) localized in the predicted ββ-sheet region of the first and second zinc finger motif that disrupted dimerization. Interestingly, these positions were affected in neither the Ikaros nor the Hunchback DZF domain, suggesting these residues are specifically important for dimerization mediated by the Pegasus DZF domain. Otherwise, the result obtained with the alanine scan mutant at position 14 was consistent with the result obtained in Ikaros while the effect of alanine at conserved position 44 was consistent with the effect this mutation caused in Hunchback. Substitutions in

the remaining four residue positions (11, 14, 18 and 46) disrupted dimerization which was also the case for Ikaros and Hunchback.

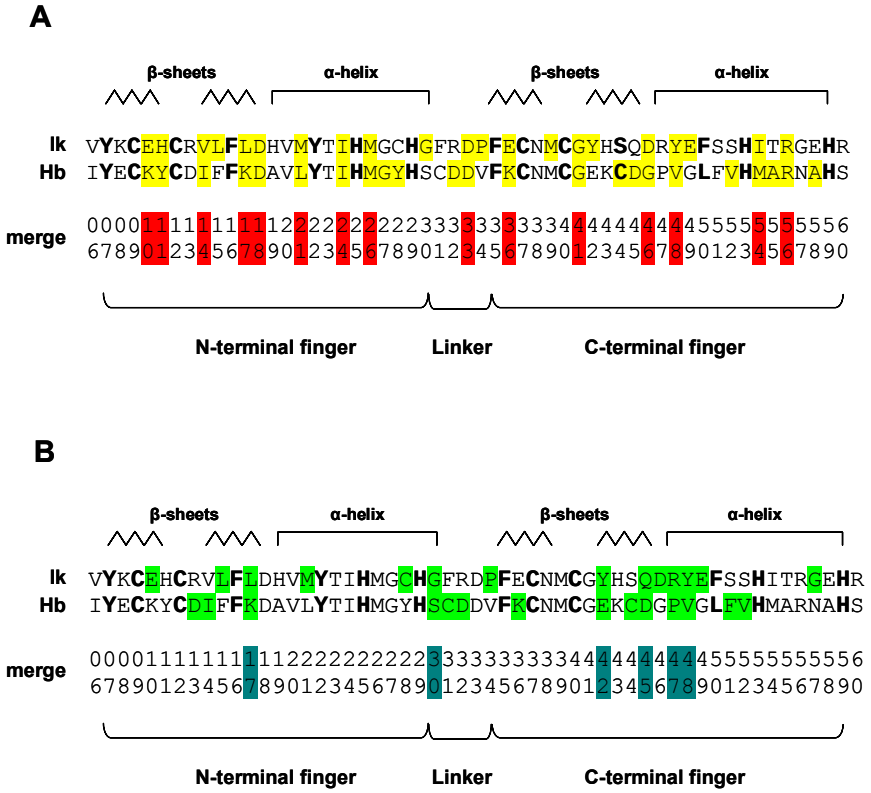

Figure 4.11 Residues identified as important for dimerization for the Ikaros DZF domain compared to residues identified for the Hunchback DZF domain. The results are summarized and projected on the amino acid sequence of the Ikaros and Hunchback DZF domain. Conserved residues are marked in bold letters. Numbering of residues is shown as defined in Figure 4.4. Individual fingers and typical secondary structures are also shown. **(A)** Residues identified by A-scan analysis are highlighted in yellow. Identified residue positions found in both DZF domains are highlighted in red in the numerical identifier below. **(B)** Residues identified by "swap" scan analysis are highlighted in green and residue positions identified in both DZF domains are summarized in blue below.

4.6.2 Residue "swap" scan mutagenesis for Ikaros and Hunchback

Comparison of the swap mutant results for the Ikaros and Hunchback DZF domains reveals some insights in how these DZF domains mediate dimerization (Figure 4.11B). Mutations in six of these residue positions (17, 30, 42, 45, 47 and 48) disrupted dimerization of both the Ikaros and the Hunchback DZF domain. Positions 10, 15, 21, 28, 34, 46, 49 and 57 were affected by a mutation only in the Ikaros DZF domain while positions 13, 14, 31, 32, 36, 44, 51 and 52 showed the reverse result and were only affected by mutations in the Hunchback

DZF domain. Four Ikaros residues (Q45, D46, R47, Y48 and E49) which seemed to be very important for selectively were adjacent to one another and formed a patch on the predicted α-helix of the C-terminal finger. A similar patch of residues (C44 D45 P47 and V48) was found in Hunchback. Here, additional residues (S30, C31 and D32) were present at the inter linker region which were not present in the Ikaros DZF domain. The rest of the affected residues were spread out for both DZF domains. Overall, this result confirmed the result of the alanine-scan and emphasizes that these two DZF domains share partially overlapping regions but use different specificity determinants.

4.6.3 Comparison with result previously obtained for different DZF domains

Other groups have identified residues in the DZF domains from Ikaros and Eos that are important for dimerization and the results of these studies compared to our findings will be discussed below (Sun *et al.*, 1996; McCarty *et al.*, 2003; Westman *et al.*, 2004). As mentioned above, McCarty and co-workers performed a "swap" mutation analysis for the Ikaros DZF domain using mammalian-cell based chemical crosslinking or co-immunoprecipitation assays to analyze the effect of these mutations. Sun and co-workers replaced 12 single residue positions in the Ikaros DZF domain with glycine (C9, C12, D18, H25, H29, C37, C40, D46 and E49), proline (V20 and M21) and histidine (G57) and used Y2H assays for analyzing these mutants. The last group (Westman *et al.*, 2004) introduced pairwise alanine substitutions into all residues in the Eos DZF domain with the exception of defined structural amino acids and used Y2H assays to test for the abrogation of homo-dimerization.

Generally, our results for Ikaros matched well with those from McCarty and co-workers, although we found one mutation, D46Q that significantly abolished homodimerization in our system that was not detected in the previous study using co-immunoprecipitation. This divergent result may be due to the different methods used for analyzing the mutants. For example, the B2H system may generally be more sensitive for detecting dimerization defects. On the other hand, this mutant may have been expressed at lower levels in *E. coli* which resulted in an abrogation of homo-dimerization. Discrepancies are definitely present in the study of Sun and co-workers since none of their mutations (except in the structural residues) seemed to have an effect on dimerization. We obtained a similar result for the Pegasus DZF domain where, besides the structural amino acids, only few other residues were identified as

important. Consistent with Ikaros, important residues identified for Eos concern mainly the α-helix in the first ZF, although additional residues were found that were not important for Ikaros dimerization (Westman *et al.*, 2004). On the other hand, almost none of the residues in the second ZF were identified as important for Eos which is clearly different from the result obtained with Hunchback and Ikaros. Thus, differences appear for the various DZF domains and discrepancies occur when different methods are used (see also section 4.7.6).

4.7 Discussion

4.7.1 Overall fold of the DZF domain is expected to be similar to that of the classical C2H2 ZFs

The work in this chapter was aimed at identifying residues important for affinity and specificity of DZF domain mediated homodimerization. The DZF domains from Ikaros, Pegasus and Hunchback were chosen because they interact in very specific and distinctive patterns despite a high degree of homology in their amino acid sequences. Despite their functions as mediators of protein-protein interactions (instead of DNA binding) both fingers in all three DZF domains match the conserved motif of classical C2H2 zinc fingers. Exceptions are the missing Phe/Tyr residue in the second finger (four positions after the second C, Figure 4.1) which is part of the predicted hydrophobic core and the linker between the two fingers, which differs from the highly conserved linker that usually connects DNA-binding ZFs. Although these variations exist, studies of the individual C-terminal fingers from Eos and Pegasus using Circular Dichroism (CD), UV-Vis and NMR have shown that both domains individually fold in a manner similar to that of the classical C2H2 ZFs (Westman *et al.*, 2004). In addition, the structure of the second finger in Eos was determined by NMR spectroscopy and revealed that the overall fold is similar to the typical ββα structure of other C2H2 ZFs although some significant differences exist (Westman *et al.*, 2004; see also section 1.2.7.2). For example, the structure displays some conformational flexibility which could result in a rearrangement of the backbone and thereby of the zinc ligating residues. However, this finger contains a well organized hydrophobic core which is believed to stabilize the overall ββα-structure of classical C2H2 ZFs. This is somehow surprising given the fact that the invariant phenylalanine (four positions after the second C) of the

classical C2H2 ZF motif is missing (Westman *et al.*, 2004). Closer examination of the hydrophobic core suggested that the loss of this bulky hydrophobic side-chain is partly compensated by the side-chains of a tyrosine (two positions after the second zinc ligating cysteine) and the invariant phenylalanine (2 positions before the first zinc ligating cysteines) as well as by the presence of a phenylalanine three residue positions before the first zinc ligating histidine (note that this residue position is defined as an invariant hydrophobic position in the classical C2H2 ZF motif). These three residues are positioned closer to the center of the hydrophobic core and may help maintaining the overall structural integrity of this C2H2 ZF (Westman *et al.*, 2004). Because of the high sequence homology between the various DZF domains (see section 3.2.1) it may be reasonable to assume that the overall fold of the DZF domain is constant for the different proteins and represents a scaffold upon which the residues important for contacting other proteins in a specific manner are displayed.

4.7.2 Mutational analysis narrowed down residue positions that might be important for dimerization

To identify residues that are important for dimerization affinity, a mutational analysis of the DZF domains from Ikaros, Pegasus and Hunchback was performed using alanine and "swap" mutation scans. Note that the Pegasus DZF domain was only analyzed by alanine scan mutagenesis and only a few residue positions were found to be affected by such a mutation. Individual point mutations seemed not to be effective enough to disrupt dimerization. A possible reason for this could be that the binding energy of this interaction is generally higher compared to Ikaros and Hunchback. However, since we can not rule out that the introduced mutations affected stability of the protein, this may just indicate that Pegasus is more stable and the structure of the peptide can be maintained despite the presence of mutations.

For Ikaros and Hunchback, several residues could be identified that were affected in their ability to mediate homodimerization when replaced by a different amino acid. Mutations at these positions may disrupt specific contacts at the dimerization surface, indicating that these positions are essential for dimerization, but we can not rule out that these positions were simply important for stability of the structure or rather solubility of the protein. In addition, some mutated proteins might have been toxic to the cells and were not expressed at high levels which would also result in a diminished ability of the domain to dimerize. In any case, several residues were not defective in their ability to homodimerize when replaced by a

mutant and are therefore unlikely to contribute substantially to the interaction surface (see Figures 4.6 and 4.9).

4.7.3 Results of alanine scan and "swap" scan analysis are generally consistent

The DZF domains of Ikaros and Hunchback were analyzed by both alanine scans and "swap" scans. Although the results of these two scans were generally very consistent for each DZF domain, there are some residue positions where only one of the two introduced mutations showed an effect on dimerization. For example, in both DZF domains several positions were only affected by an alanine and not by the corresponding swap residue. Introducing "swap" residues that are not directly involved in mediating contacts may be less severe since these residues supposedly help maintaining the structural integrity of another (either Ikaros or Hunchback) DZF.

At other positions the opposite phenotype was observed and only the swap mutation had an effect on dimerization. For example, in the Ikaros DZF domain positions 45 and 47 are both affected by the respective swap mutation, suggesting that they are important for the interaction. In addition, both positions are adjacent to other residues identified as important for dimerization. However, introducing an alanine at these positions did not cause a dimerization defect. An explanation for this phenomenon could be that in these cases the swap residue may have had an effect on neighboring amino acid residues by, for example, changing the arrangement of these residues, which in turn caused changes in the orientation of the interaction surface. A similar argument can be made for residues 47 and 51 in the Hunchback DZF domain, which are only affected by the corresponding swap mutation. Both positions are close to other residues defined as important and mutations at these positions may act on nearby essential residue positions.

Thus, it is noteworthy that in some cases the nature of the substituted side chain has to be considered since it may have an effect on the confirmation or the chemical properties of the whole protein. Some of the side chains in the "swap" mutations may have been inappropriate for testing the potential function and importance of the original side chain present in the wild-type protein. Thus, results from both scans have to be considered and compared before one can make conclusions about potential important residue positions. Although it is difficult to define residue positions that make contacts at the interaction interface, those amino acids

which were not defective in dimerization when replaced by an alanine or "swap" residue are definitely not important for dimerization.

4.7.4 Several structural and hydrophobic residue positions were affected by a mutation

Residues identified by alanine scan and "swap" analysis were generally distributed throughout the whole domain except for Pegasus where most of the residues were found in the predicted β-sheets of the N-terminal ZF. Many of them are defined as likely structural amino acids important for maintaining the ββα structure (e.g. positions 07, 16, 22, 35 and 50). Position 44 is a phenylalanine in the classical C2H2 ZF motif and is believed to be involved in maintaining the hydrophobic core of the ββα fold. Interestingly, as described for Eos, this residue is substituted by a serine in the Ikaros DZF domain and was not affected by a mutation. On the other hand, residue positions 35, 42 and 48 were impaired by a mutation and these residues are believed to be involved in maintaining the hydrophobic core in the Eos C-terminal C2H2 ZF (see section 4.7.1). This suggests that the hydrophobic core and the overall structure of the C-terminal C2H2 ZF of Ikaros are similar to the corresponding Eos ZF. On the other hand, both Hunchback and Pegasus harbor a cysteine at position 44 and introducing a mutation at this position affected the ability of these proteins to dimerize. Thus, different residue positions may be involved in maintaining the hydrophobic core of these two proteins.

Besides these defined hydrophobic and aromatic residue positions a few other hydrophobic amino acids were identified as important in the DZF domains, which are mainly present in the N-terminal C2H2 ZF. Hydrophobic residues have been found to be abundant in various protein-protein interaction surfaces suggesting that they are sufficient to stabilize protein complexes (Lo Conte *et al.*, 1999; reviewed in Jones and Thornton, 1996). In fact, analysis of residue pairing preferences at protein interaction interfaces has shown that the most prevalent pairing involves amino acid interactions between hydrophobic residues (Glaser *et al.*, 2001). Thus, hydrophobic residues identified as important for the DZF domain are likely to be involved in the interaction surface, possibly by providing the necessary binding energy for the DZF interaction in general. They could be participating in making initial contact with other DZF domains. Using van der Waals interactions they may stabilize this primary dimerization complex upon which the specific tight interactions are eventually established.

On the other hand, they may just contribute indirectly to the dimerization by both orientating and stabilizing the protein backbone in a way that the interaction surface is exposed.

4.7.5 Residue positions important for specific dimerization are mainly located in the predicted α-helices of the DZF domains

Importantly, the result of the "swap" scan analysis demonstrated that some of the identified residues in the Ikaros DZF domain are adjacent to one another and have been suggested to form clusters on the surface of the protein using homology modeling (McCarty *et al.*, 2003). Most striking is the patch in the C-terminal finger formed by five adjacent residue positions (Q45, D46, R47, Y48 and E49) that were all affected by a mutation and are very likely to cluster on the surface of the domain. Within the N-terminal finger, residue positions 17, 21 and 28 were expected to cluster as well and these three positions were all important for dimerization as judged by our mutational analysis. Interestingly, both regions correspond to the α-helix which is known to contact specific DNA sites in DNA binding ZFs (McCarty *et al.*, 2003; reviewed in Wolfe *et al.*, 2000; see also section 1.2.4). These base contacts are mainly made by amino acids at positions -1, 2, 3 and 6 of the α-helix, although variations of this pattern have been described as well. Examples are ZFs 4 and 5 from GLI that use α-helix residues at positions 1 and 5, respectively to contact specific DNA-sites (Pavletich and Pabo, 1993). In addition, the residue at position 10 has been described to form important base contacts for ZF3 of TFIIIA (Wuttke *et al.*, 1997). Within the C-terminal DZF finger of Ikaros, the cluster of residues that is suggested to contribute to selectivity lies between positions -2 and +4, although position 4 is occupied by a hydrophobic core residue (F50) and is therefore not directly involved in making protein contacts (McCarty *et al.*, 2003). The candidate residues found in the predicted α-helix of the N-terminal ZF in the Ikaros DZF domain correspond to positions -2, +3 and +10 of the recognition helix (McCarty *et al.*, 2003). It is noteworthy that besides these two clusters in the predicted α-helices several other positions were affected which are distributed throughout the whole domain including the β-sheets and the linker region.

The hunchback DZF domain displays a similar but distinct group of adjacent residues (C44, D45, G46, P47 and V48) in the second ZF located in the predicted C-terminal end of the β-sheet and in the predicted N-terminal α-helix. These residues correspond to positions -3 to +3 of the recognition helix and may also form an exposed cluster on the surface of the protein

assuming that the overall ββα fold of this domain is similar to the one modeled for Ikaros (McCarty *et al.*, 2003). There was no obvious cluster found in the α-helix of the N-terminal ZF which is in contrast to the result obtained for Ikaros. Rather, additional residues (S30, C31 and D32) are present within the inter linker region and these positions were not identified as being important for dimerization in the Ikaros DZF domain. The remaining affected residues were mainly found in the predicted β-sheet region of the domain.

4.7.6 A potential role for the linker in mediating specific dimerization

Another notable feature of these results is the potential role of the linker region for dimerization specificity. In general, the linker connecting the individual fingers in the DZF domain motif is different from the conserved TGEKP linker that normally links DNA-binding zinc fingers. Furthermore, three residues inclined to be important for dimerization are found as a cluster in the linker region of the Hunchback DZF domain although it is not clear if these residues are directly involved in making specific contacts with other DZF domains. Ikaros on the other hand contains a rigid proline residue at position 34 which appears to be necessary for dimerization. This residue may play a role in decreasing general flexibility of the protein (note that Hunchback has a valine residue at this position). Thus, it is possible that the linker is essential to maintain a defined orientation of the two ZFs in the DZF domain which in turn is important for selectivity of the interaction.

4.7.7 Different DZF domains are likely to use different residue positions for mediating specific dimerization

Comparing the results obtained for the Ikaros and Hunchback DZF domain highlights both similarities and differences. For instance, the result for both DZF domains suggests that the predicted N-terminus of the α-helix in the second C2H2 ZF is important for dimerization specificity (McCarty *et al.*, 2003). On the other hand, an additional important cluster of residues is found in the predicted α-helix of the first ZF in the Ikaros DZF domain and this cluster is missing in the Hunchback protein. Thus, this cluster may be important for dimerization of Ikaros, but not for Hunchback. Other groups have identified residues in the DZF domains from Ikaros and Eos that are important for dimerization (McCarty *et al.*, 2003; Westman *et al.*, 2004). In general, our results matched well with those obtained in these

studies, indicating, that the predicted N-terminal end of the α-helix is generally important for dimerization. While Hunchback dimerization relies mainly on the predicted α-helix in the second finger, Eos seems to require the α-helix of the first ZF. Ikaros on the other hand seems to use the α-helices of both fingers. Thus, it appears that differences exist for the various DZF domains. Although it is difficult to interpret these variations, it may indicate that different molecular mechanisms are used for mediating dimerization which in turn contributes to the specificity of the DZF domain and might explain the various interaction specificities between them (see Chapter 3). Hence, DZF domains may generally share overlapping regions for mediating dimerization but use different specificity determinants for their interactions. It is important to note that recent studies by Westman and co-workers (2003) suggested that the Eos DZF domain mediates the formation of multimeric complexes consisting of as many as ten molecules. In addition, Ikaros has been reported to form a multimer as well (Trinh *et al.*, 2001; McCarty *et al.*, 2003). Thus, the different interactions surfaces described for the various DZF domains may account for different multimerization states of the respective DZFs.

Chapter 5. Steps towards determining the structure of the DZF domain.

5.1 Introduction

Although numerous examples of protein-protein interactions mediated by C2H2 ZFs have been described, there is still relatively little structural information available about how C2H2 ZFs mediate protein-protein interactions. Unfortunately, only a few structures of C2H2 ZFs involved in protein binding have been described (see also section 1.2.7.2). For example, the structure of the C-terminal C2H2 ZF of Eos which is part of the DZF domain has been solved using NMR spectroscopy. This structure revealed both similarities and differences to the typical ββα fold of the C2H2 ZFs (Westman *et al.*, 2004; see section 1.2.7.2) but does not provide information about C2H2 ZF mediated dimerization. Other initial attempts to obtain biochemical information about C2H2 ZF mediated protein-protein interactions have been performed, but no structure of a C2H2 ZF interacting with another protein has been described to date. In contrast, much progress has been made in understanding DNA recognition mediated by C2H2 ZFs and structures of various DNA-binding C2H2 ZFs bound to their DNA sites have been described in detail (Pavletich and Pabo, 1991; Pavletich and Pabo, 1993; Elrod-Erickson *et al.*, 1996; Houbaviy *et al.*, 1996; Kim and Berg, 1996; Nolte *et al.*, 1998; reviewed in Wolfe *et al.*, 2001). We therefore attempted to use X-ray crystallography to obtain detailed structural information of a dimeric or multimeric DZF complex. This structure together with mutational information (see Chapter 4) should identify specific surfaces and residues within the C2H2 ZFs that are involved in this protein-protein interaction.

To determine high-resolution crystallographic structures of DZF domains, we sought to purify milligrams of highly pure peptides from different DZF domains which could then be used for X-ray crystallography (in collaboration with Dr. Robert Grant of the X-Ray Crystallography Core Facility in the Department of Biology at the Massachusetts Institute of Technology [MIT]). We decided to follow a previously described strategy which was successfully used to purify various C2H2 DNA-binding ZF domains containing 2 to 5 fingers for crystallographic studies (Figure 5.1, Pavletich and Pabo, 1991; Pavletich and Pabo, 1993; Elrod-Erickson *et al.*, 1996; Elrod-Erickson *et al.*, 1998).

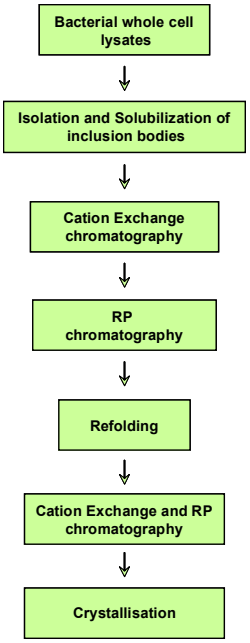

Figure 5.1 Purification strategy for C2H2 ZFs. This strategy was successfully applied to purify various DNA-binding ZFs for setting up crystallization trials. RP, reverse phase. See text for details.

In this approach, ZF domains are over-expressed in *E. coli* and isolated from inclusion bodies. The presence of peptides in inclusion bodies can generally be beneficial since the over-expressed protein often dominates in inclusion bodies taking up > 50% of the total cellular protein content. In addition, they provide protection for the peptide from proteolytic degradation as well as for the cell against the potential toxicity of the peptide (reviewed in Lilie *et al.*, 1998; Clark, 1998). ZF peptides are subsequently solubilized under denaturing and reducing conditions. To further purify solubilized C2H2 DNA-binding ZFs, successive rounds of cation exchange (using Fast Protein Liquid Chromatography [FPLC]) and reverse-phase (using High Performance Liquid Chromatography [HPLC]) chromatography were used (Elrod-Erickson *et al.*, 1998). Cation exchange chromatography was carried out under denaturing conditions (using a Source 15S column) and the denaturant was subsequently removed by reverse-phase (RP) chromatography (using a C4 column). After these various successive purification steps, the biological activity of the peptide has to be recovered by refolding it back to its native form. Since ZF proteins are prone to oxidation, care must be taken to refold them under anaerobic conditions. So far, various C2H2 ZFs were successfully re-folded by applying different buffer-conditions (Pavletich and Pabo, 1991; Pavletich and Pabo, 1993; Elrod-Erickson *et al.*, 1996; Elrod-Erickson *et al.*, 1998). Folding reactions can be performed in the presence of cobalt instead of zinc. Coordinated binding of the Co^{2+} ion by the cysteines and histidines leads to the reduction of cobalt which results in a color change

of the sample from clear to blue. Thus, the color change can be used as a preliminary indicator of successful folding events (Pavletich and Pabo, 1991). Mobility shift experiments were performed to show, that the refolded DNA-binding C2H2 ZF was active and able to bind to its specific DNA binding site. To separate active peptides from inactive forms, two additional purification steps were performed. The folded C2H2 ZF was first purified on a MonoS cation exchange column with a NaCl gradient and was subsequently loaded onto a C4 RP HPLC column to perform another round of RP chromatography. Resulting pure and active peptides were then used for setting up crystals (Pavletich and Pabo, 1991).
This chapter describes the over-expression and partial purification of DZF domains from several Ikaros family members. It further delineates various attempts to refold these DZFs into stable active peptides.

5.2 Overexpression and purification of the Pegasus DZF domain

5.2.1 Overexpression of the Pegasus DZF domain

Overexpression of peptides was performed in the pET *E. coli* expression system under the control of an IPTG-inducible promoter. We constructed a pET3a derived expression plasmid which expressed a segment encoding the DZF domain of human Pegasus under control of a strong T7 phage promoter. This plasmid was introduced into the *E. coli* strain BL21(DE3)pLysS, in which the T7 RNA polymerase expression is controlled by the Lac repressor. To test, whether this peptide can be overexpressed, IPTG was added to the *E. coli* cells to induce expression. Analysis of whole cell lysates from uninduced and IPTG induced cells using SDS PAGE revealed that a peptide of the predicted molecular weight of 7.14 kD (Table 5.1) is expressed at high levels (Figure 5.2A, lane 2). Control experiments demonstrated that the expression of this peptide is dependent upon induction by IPTG since uninduced cells did not express the peptide (Figure 5.2A, lane 1). In addition, SDS-PAGE analysis of soluble fractions and inclusion bodies from lysates of the induced cultures indicate that the Pegasus DZF peptide is primarily localized within the inclusion body fraction (Figure 5.2B, lane 2).

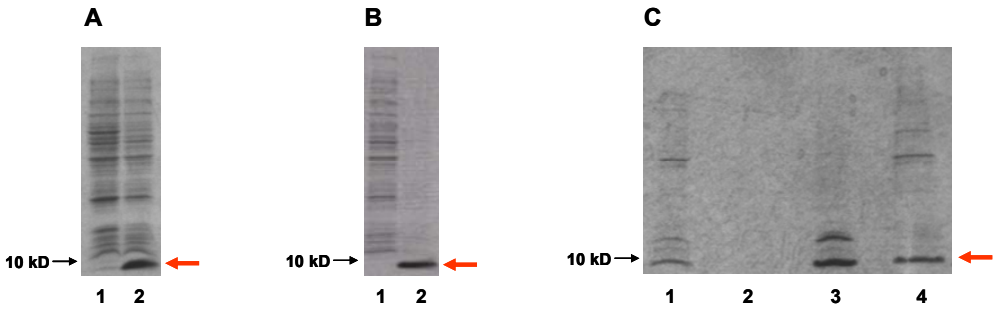

Figure 5.2 SDS-PAGE analysis of the over-expression and purification of the Pegasus DZF. (A) Analysis of whole cell lysates from uninduced (lane 1) and induced (lane 2) bacterial cell cultures transformed with a plasmid encoding the Pegasus DZF domain. Induction was performed at 37˚C for three hours by adding 0.4 mM IPTG. **(B)** Analysis of soluble fractions (lane 1) and inclusion bodies (lane 2) from whole cell lysates obtained in A. (**C**) Analysis of samples taken during the inclusion body purification and solubilization. Lane 1 = wash step 1, lane 2 = wash step 2. lane 3 = solubilized peptide, lane 4 = inclusion bodies. Red arrow indicates band which corresponds to the Pegasus DZF peptide. Black arrow indicates 10 kD band of Standard (Precision Plus protein standard, Biorad) which was run simultaneously with the samples (data not shown). See text for details.

5.2.2 Purification of inclusion bodies and solubilization

To further investigate whether the Pegasus DZF peptide can be captured by isolating inclusion bodies, we harvested 2 liter cultures of *E. coli* cells expressing the Pegasus DZF domain peptide. The cells were lysed using a freeze/thaw protocol and by adding detergents (see section 2.4.2.2). After centrifugation, the inclusion body pellet was washed with a buffer containing chaotropic agents and detergents. To solubilize the peptides, inclusion bodies were resuspended in a buffer containing a strong denaturant (urea) and a reducing agent (DTT). SDS-PAGE analysis of the purified and solubilized inclusion bodies demonstrated that most of the Pegasus peptide was soluble (Figure 5.2C lane 3) under denaturing conditions although some peptide was still trapped in the insoluble pellet (Figure 5.2C lane 4). Samples of the supernatant were taken after every wash step (pellet wash 1 and 2, see section 2.3.2.2) and also analyzed by SDS-PAGE (Figure 5.2 lane 1 and 2). While the supernatant of pellet wash 1 contained some contaminants that probably absorbed onto the hydrophobic inclusion bodies, the supernatant of pellet wash 2 was clean. Thus, isolation of inclusion bodies can be used to capture the Pegasus DZF peptides from cell lysates.

Peptide	Molecular weight (kD)	pI value	Composition of buffer A for loading onto column	Conc of buffer B required for eluting peptide of column (source 15S)	Conc of buffer B required for eluting peptide of column (C4)	Extinction coefficient at 280 nm (units of M^{-1} cm^{-1})
Pe	7135.5	6.7	MES 6.0	12-13.5 %	68-72 %	5120
Tr	7056.4	6.0	MES 5.0	10-12 %	74-79 %	2560
Ik	7700.2	6.7	MES 5.0	Not recorded	Not recorded	5120
Hd	6587.0	N.d.	N.d.	N.d.	N.d.	N.d.
Hc	6898.1	6.6	MES 5.5	10-12 %	68-72 %	3840
Tr-Eo-Eo	7066.4	5.8	MES 5.0	8-10 %	72-76 %	3840
Zif268-Pe	18626.3	8.5	HEPES 7.5	29-31 %	65 %	6400
Pe-Zif268	19449.2	N.d.	N.d.	N.d.	N.d.	N.d.
Ik N-finger	4156.2	N.d.	N.d.	N.d.	N.d.	N.d.
Ik C-finger	4397.1	N.d.	N.d.	N.d.	N.d.	N.d.
Hd N-finger	3752.5	N.d.	N.d.	N.d.	N.d.	N.d.
Hd C-finger	3503.2	N.d.	N.d.	N.d.	N.d.	N.d.

Table 5.1 Biochemical properties of peptides and applied buffer conditions. This table summarizes information about the various peptides (shown in first column) required to perform the different purification steps and folding reactions. Cells displaying N.d. (not determined) belong to peptides that were not further purified. Abbreviations are as defined in Figure 3.4 (Chapter 3). N-finger, N-terminal ZF; C-finger, C-terminal ZF; Conc, concentration. Source 15 S and C4 indicate the respective column (see section 2.4.2.3 and 2.4.2.4). See text for details.

5.2.3 Ion exchange chromatography

To purify the solubilized Pegasus DZF peptide, cation exchange chromatography was used under denaturing conditions. Separation in ion exchange chromatography is based on the reversible binding of a charged peptide to an oppositely charged medium. Cation exchangers are negatively charged and have positively charged counter-ions (cations). Thus, peptides to be bound to the exchanger have to carry a "net" positive charge. Since most proteins are both positively and negatively charged the "net" surface charge of a protein is usually influenced by the composition of the surrounding medium. Thus, at a certain pH value termed as the isoelectric point (pI) a protein will have zero "net" charge. Below its pI a protein has a "net" positive charge and can bind to cation exchangers. To assure that Pegasus binds reliable to the negatively charged exchangers, we decided to initially choose a buffer which was at least 1 pH unit below the pI of Pegasus. The estimated pI value for the Pegasus DZF peptide was 6.7 (Table 5.1). Thus, the peptide was loaded on a Source 15S cation exchange column in a buffer of pH 5.0 that also provided denaturing and reducing conditions (Buffer A). Conditions were then altered so that the bound molecules were eluted differentially using a

NaCl gradient (Buffer B) where increasing salt concentration gradually displaced the peptide on the column.

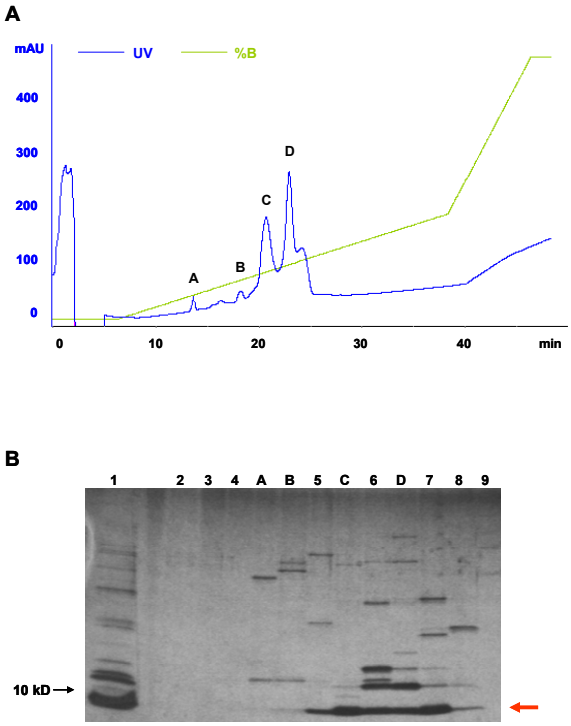

Figure 5.3 Analysis of the initial ion exchange chromatography run applied to purify the Pegasus DZF domain. (A) Analytical FPLC trace for the DZF domain from Pegasus. Peptides were loaded onto the column in buffer A (pH 5.0). Bound peptides were eluted over ~40 min using a Buffer B gradient (shown in green). UV trace measuring the Absorbance units (mAU) is shown in blue. Peaks obtained are numbered A, B, C and D. **(B)** SDS-PAGE analysis of the FPLC run from (**A**). Peak fractions A-D together with several fractions (2-9) taken at different time points were analyzed. Lane 1 represents a sample of the solubilized peptide before it was loaded onto the column. Red arrow indicates Pegasus DZF domain band and black arrow indicates 10 kD band deduced from Standard. See text for details.

A trial chromatography run showed 4 peaks (2 small and 2 major peaks), which eluted at different concentrations of buffer B (Figure 5.3A). To find out, which peak contains the Pegasus peptide, the peak fractions and several fractions adjacent to the peaks were collected and analyzed by SDS-PAGE and subsequent silver staining. As shown in Figure 5.3B, the Pegasus DZF peptide eluted in a very broad range between 12-25% of buffer B, although the majority was located in peak C. This suggested that the peptide bound too tightly to the negatively charged exchangers and could therefore not be eluted efficiently in a single fraction. In addition, the Pegasus peptide eluted close to another protein which decreased the purity of the preparation.

Therefore, we decided to increase the pH of Buffer A to 6.0 with the hope that the Pegasus DZF peptide would bind less tightly to the column. As shown in Figure 5.4A, under these conditions four well-defined peaks eluted at different concentrations of buffer B. Analysis of the chromatography run by SDS-PAGE and silver staining demonstrated that the Pegasus peptide eluted as a single peak between 12-13.5% of buffer B, although there is still some peptide eluting even at lower concentrations of buffer B. The captured sample contained mainly the Pegasus peptide as judged by silver staining (Figure 5.4B). Increasing the pH of Buffer A even further did not improve the elution profile and the majority of the peptide did not bind to the column at all (data not shown).

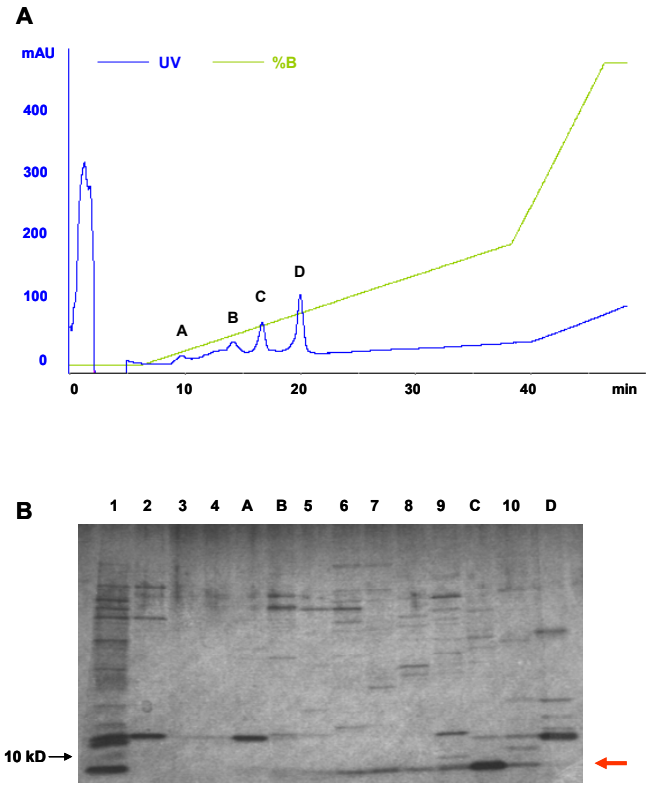

Figure 5.4 Analysis of an ion exchange chromatography run applied to purify the Pegasus DZF domain. (A) Analytical FPLC trace for the DZF domain from Pegasus. Peptides were loaded onto the column in buffer A (pH 6.0). Bound peptides were eluted using a Buffer B gradient (shown in green). UV trace is shown in blue. Peaks are numbered A, B, C and D. **(B)** SDS-PAGE analysis of the FPLC run from (**A**). Peak fractions A-D together with several sample fractions (2-10) were analyzed. Lane 1 represents the solubilized peptide. Red arrow indicates Pegasus DZF domain band (contained in sample C) and black arrow indicates 10 kD band of Standard.

We tested if larger amounts of the Pegasus peptide could be purified when the protein purification was scaled up. For this experiment, pH 6.0 for Buffer A was used and a shallower gradient of buffer B was used for eluting the protein. Analysis of the

chromatography run demonstrated that the Pegasus peptide eluted as a single peak between 11-12% of buffer B (data not shown). The amount of protein was increased compared to the small scale run as judged by the absorbance units and the captured volume.

5.2.4 Reverse phase chromatography

To further purify the Pegasus DZF peptide, RP chromatography (using HPLC) was applied. Using this approach, peptides are separated based on their "hydrophobic character". Peptides adsorb to the hydrophobic nonpolar surface of the column after applying them in a polar mobile phase by interaction of the nonpolar components of the proteins with the hydrophobic nonpolar stationary phase. Compounds are eluted from RP HPLC columns by decreasing the polarity of the mobile phase using organic solvents which results in desorption of the protein. To assure that separation is solely based on hydrophobicity, the effective charges of the peptide have to be reduced before loading it onto the column. This can be accomplished by reducing the pH (typically to 2) and at the same time providing an ion-pairing reagent (e.g. TFA) that "hides" the resulting positive charges by forming "ion pairs" with them.
Thus, the sample was acidified with 10 % TFA to pH 2-3 to provide binding to the hydrophobic stationary phase of the C4 reverse phase column that is solely based on the hydrophobic character of the peptide. The protein was then loaded onto the column using nonpolar buffer conditions (Buffer A, consisting of water with 0.1 % TFA). Molecules were eluted differentially with an organic mobile phase (Buffer B, containing acetonitrile with 0.1 % TFA) that was introduced gradually. As shown in Figure 5.5A, three major peaks eluted at different concentrations of buffer B. Several fractions together with the peaks were captured and analyzed by SDS-PAGE and subsequent silver staining demonstrating that the Pegasus DZF domain eluted as a single peak between 68-71 % of buffer B (Figure 5.5B). In addition, the captured sample was highly pure and contained only the Pegasus DZF domain. (Note that the other peaks did not contain any proteins and are likely to represent buffer and DTT eluting from the column). The same buffer conditions were then used to scale up the sample amount using a shallower gradient for eluting the peptide. As expected, this chromatography run yielded more protein as judged by the absorbance units and the captured volume (data not shown). The HPLC peak fraction corresponding to the Pegasus DZF peptide was subsequently lyophilized and the dried product was moved to an anaerobic chamber.

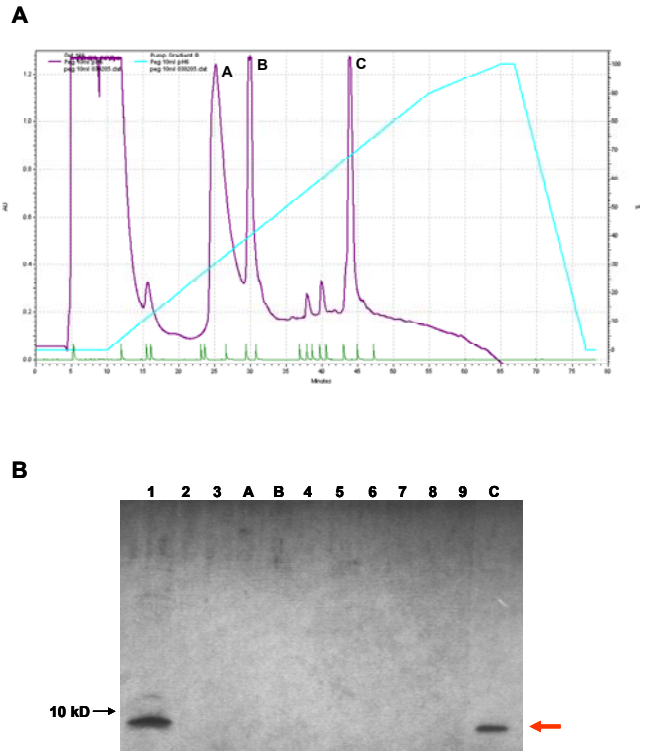

Figure 5.5 Analysis of the Reverse Phase (RP) chromatography run applied to further purify the Pegasus DZF. (A) Analytical HPLC trace for the DZF domain from Pegasus. Bound peptides were eluted over ~ 60 min using a Buffer B gradient (shown in turquoise). UV trace is shown in purple. Collected peaks are numbered A, B, C. **(B)** SDS-PAGE analysis of the HPLC run from (**A**). Peak fractions A-C together with several fractions (2-9) taken at different time points were collected and analyzed. Lane 1 represents a sample of the acidified peptide before it was loaded onto the column. The red arrow indicates the Pegasus DZF domain band and the black arrow indicates where the 10 kD band of the Standard ran.

5.2.5 Refolding

After these successful purification steps, we attempted to refold the Pegasus peptide with either cobalt or zinc. To do this, the dried peptide was resuspended in water (in an anaerobic chamber) and the protein concentration was determined (~ 25 mg/ml in this initial experiment). To attempt refolding the peptide in its active form, 1.5 molar equivalents of cobalt were added and the pH was adjusted by introducing different buffers. The solution of dissolved Pegasus peptide immediately turned blue after adding the buffer, suggesting that the peptide was coordinating the cobalt ion (data not shown). However, insoluble aggregates accumulated immediately following appearance of the blue color. Spinning down the sample demonstrated that the aggregates represented the refolded peptide since the precipitate was blue and the supernatant did not harbor any detectable protein as judged by UV absorbance

measurement. Using zinc instead of cobalt or performing the folding reaction on ice also resulted in a precipitation of protein following adjustment of the pH.

The efficiency of refolding depends in general on the competition between correct folding events and aggregation (reviewed in Lilie *et al.*, 1998). Thus, it is very important to slow down the aggregation process in order to obtain correct folded and biological active protein. There a various biochemical variables that influence the formation of properly folded protein which were gradually evaluated as described below.

5.2.5.1 Protein concentration

A very direct way of minimizing aggregation is by reducing the concentration of the protein, since aggregation usually happens at high concentrations of proteins (reviewed in Clark, 1998). Thus, the protein concentration of the Pegasus DZF peptide was decreased stepwise by serially diluting purified unfolded peptide with water. The following final concentrations of protein were tested in this experiment: 10, 0.1, 0.05, 0.001 and 0.0001 mg/ml. We attempted to re-fold these diluted peptides using the method described above and the color change was monitored. The result of this experiment showed that reduction of the protein concentration still led to formation of precipitates. Going below a certain protein concentration (< 0.001 mg/ml) did not permit detectable levels of blue color anymore (as determined by eye).

5.2.5.2 Folding buffer composition (pH, ionic strength)

Other variables that influence the stability of the proper folded state are the pH and the ionic strength of the folding buffer (reviewed in Clark, 1998; reviewed in Lilie *et al.*, 1998). Thus, different buffer conditions (pH 5.0 to pH 8.0) were used and salt was added at various concentrations to refold the protein. As shown in Table 5.2, the peptide folded but precipitated at pH 7.0 - 8.0 but then was not able to fold at pH 5.0 or pH 6.0 as judged by the color change or lack thereof. Adding salt at different concentrations did not prevent precipitation once the protein was folded (Table 5.3). Introducing salt to the folding reaction before adding cobalt resulted in an immediate precipitation and the color did not change, even after introducing cobalt.

Buffer, pH	Folding phenotype
0.5 M MES, pH 5.0	No change in color. No precipitation
0.5 M MES, pH 6.0	No change in color. No precipitation
1 M HEPES, pH 7.0	Blue precipitate
1 M BTP, pH 7.0	Blue precipitate
1 M HEPES, pH 7.5	Blue precipitate
1 M BTP, pH 7.5	Blue precipitate
1 M HEPES, pH 8.0	Blue precipitate
1 M BTP, pH 8.0	Blue precipitate

Table 5.2 Evaluation of different folding buffers with different pH values used to perform the refolding reaction.

5.2.5.3 Urea

Urea has been proven to inhibit aggregation by increasing the solubility of unfolded proteins and decreasing non-specific hydrophobic interactions which results in a general increase of correctly folded protein (Orsini and Goldberg, 1978). Thus, urea was added at various concentrations (0-2 M) to refold the protein either in combination with salt or without adding salt. As shown in Table 5.3, even after introduction of urea we still observed blue precipitates.

Buffer	Salt	Urea	Folding phenotype
1 M HEPES, pH 7.5	0 mM NaCl, KCl or NH_4OAc	0 M	Blue precipitate
		1 M	
		2 M	
	10 mM NaCl, KCl or NH_4OAc	0 M	Blue precipitate
		1 M	
		2 M	
	200 mM NaCl, KCl or NH_4OAc	0 M	Precipitation before adding cobalt, no color change after adding cobalt
		1 M	
		2 M	

Table 5.3 Evaluation of different folding buffer compositions (pH, ionic strength and addition of urea) used to perform the refolding reaction. 1 M HEPES, pH 7.5 was used for all reactions. Three different salts at three different concentrations (column 2) were tested either in combination with urea (used in two concentrations) or without adding urea.

5.2.5.4 Additives

It has been shown that the use of refolding additives can prevent aggregation by interfering with intermolecular hydrophobic interactions. A variety of additives have been described that prevent aggregation by stabilizing the proper folded state, by destabilizing incorrect folded

peptides, and by enhancing the solubility of either folding intermediates or unfolded peptides. Examples for such additives are detergents, surfactants and sugars which have proven to minimize aggregation and increase the yield of properly folded protein (Maeda *et al.*, 1996; reviewed in Clark, 1998). Thus, we tested a large series of different additives which were introduced to the folding reaction in order to prevent aggregation. To do this, a commercially available detergent screen kit was used which provided 72 unique detergents (Hampton Research, detergent screen 1, 2 and 3). None of the provided detergents was able to prevent aggregation during the refolding step (data not shown).

Since we were not able to find proper folding conditions for the Pegasus DZF peptide, we decided to purify the DZF domain from human Ikaros (work performed by R. Fang), human TRPS-1, Hunchback *D.m.* and Hunchback *C.e.,* as well as the synthetic DZF domain Tr-Eo-Eo applying the same purification strategy used for the Pegasus DZF domain. The hope was that using different proteins might help to solve the aggregation problems since the aggregation could be due to various surface-exposed hydrophobic amino acids in the Pegasus peptide. Another reason for the aggregation could be the formation of higher order oligomers as described for Eos (Westman *et al.*, 2003) and Ikaros (McCarty *et al.*, 2003) which in turn could result in a precipitation of the proteins. Thus, the DZF domain from Hunchback *C.e.* and the synthetic Tr-Eo-Eo domain were chosen because they do not mediate homodimerization at all or only weakly, respectively, as judged by the B2H (see Chapter 3, sections 3.2.4 and 3.4.1) and B1H system (data not shown).

5.3 Overexpression and purification of various DZF domains

5.3.1 Purification of the Ikaros, TRPS-1, Hunchback *C.e.* and Tr-Eo-Eo DZF domains

Since purification of Pegasus resulted in a very clean and stable product, we applied the same purification approach to purify the DZF domains from Ikaros (work performed by R. Fang), TRPS-1, Hunchback *D.m.*, Hunchback *C.e.* and the synthetic DZF Tr-Eo-Eo. pET3a derived expression plasmids were constructed which expressed peptides encoding these DZFs and were introduced into the *E. coli* strain BL21(DE3)pLysS. Analysis of whole cell lysates from

uninduced and IPTG induced cells showed that nearly all DZF domains (except Hunchback *D.m.*, see below) were expressed at high levels and that expression depended upon induction with IPTG (data not shown). The calculated molecular weights of these peptides (Table 5.1) corresponded well with the relative sizes of the peptides as judged by their electrophoretic mobility. However, it was found that the Hunchback DZF domain from *D.m.* could not be induced at all and efforts to overcome this problem using various IPTG concentrations and different induction times as well as temperatures were unsuccessful (data not shown).

As in the case of the Pegasus DZF, the remaining DZF peptides were localized in inclusion bodies which were isolated to capture the desired peptides. To do this, 400 ml of induced starter culture were used and the peptides were solubilized in 10 ml denaturing buffer (Table 5.1). Analysis of the solubilized inclusion bodies demonstrated that for all four peptides the majority of the sample was soluble (data not shown).

To further purify the peptides, cation exchange chromatography was applied using conditions for a trial run. The estimated pI value for the four peptides and the individual buffer conditions used for loading the peptides onto the Source 15S cation exchange column are shown in Table 5.1. The peptide samples were subsequently eluted with a NaCl gradient (Buffer B). Analysis of the chromatography run by SDS-PAGE and silver staining demonstrated that all four peptides were able to bind to the column and eluted as single peaks (data not shown). The captured samples were very clean consisting mainly of the desired peptides as judged by silver staining (data not shown). Table 5.1 summarizes the concentrations of Buffer B required to elute the individual peptides from the column.

A subsequent reverse phase chromatography step was used to further purify the protein samples. The peptides were acidified with 10 % TFA to pH 2-3, loaded onto the column in a nonpolar buffer (Buffer A) and eluted differentially with a organic mobile phase (Buffer B). SDS-PAGE analysis of collected samples demonstrated that the peptides eluted as single peaks and the resulting samples were highly pure as judged by silver staining (data not shown). Elution conditions for the individual samples are shown in Table 5.1. HPLC peak fractions representing the individual peptides were dried by lyophilization and stored in an anaerobic chamber.

5.3.2 Refolding of the Ikaros, TRPS-1, Hunchback *C.e.* and Tr-Eo-Eo DZF domains

We attempted to re-fold the various purified DZF peptides by adding 1.5 molar equivalents of cobalt and adjusting the pH by introducing different buffers. All peptides turned blue after refolding but insoluble aggregates accumulated in the resulting samples as was observed with the Pegasus DZF peptide. Several of the folding conditions described above (e.g. pH, urea, protein concentration and selected detergents) were evaluated but none of these prevented aggregation during the refolding step.

Thus, the aggregation seemed to be a general issue of the DZF domain. Although we also analyzed DZF domains that are not able to mediate dimerization as judged by B2H and B1H assays, we can not rule out that the aggregation is caused by the process of dimerization itself. Folding reactions are generally performed at high protein concentrations and the DZFs may have dimerized at these high concentrations.

5.4 Overexpression and purification of the Pegasus DZF domain linked to Zif268

5.4.1 Inclusion body isolation and solubilization of the Zif268-Pegasus fusion peptide

After these various unsuccessful attempts to avoid aggregation, we decided to link the DZF domain to a soluble protein, hoping that this additional protein might help prevent the DZF domain from aggregation. We chose to link the DZF domain of Pegasus to Zif268 which has been shown to be soluble after refolding and was used for X-ray crystallography before (Pavletich and Pabo, 1991). In addition, we know that Pegasus is active and can mediate dimerization when fused to Zif268 since it was used in this configuration in the B2H system (see Chapter 3, section 3.2.4).

To determine whether the Pegasus DZF peptide could be purified as a fusion to the Zif268 DNA-binding domain using the purification strategy described above, we initially constructed two pET3a derived expression plasmids which expressed the DZF domain from Pegasus as an N- or C-terminal fusion to Zif268 (Table 5.1).

These plasmids were introduced into the *E. coli* strain BL21(DE3)pLysS and IPTG was added to the cells to induce the expression. Analysis of whole cell lysates from uninduced and IPTG induced cells using SDS PAGE reveals that peptides of the expected molecular weights (18.6 kD for the C-terminal fusion and 19.4 kD for the N-terminal fusion, Table 5.1) are expressed at high levels (Figure 5.6A and data not shown).

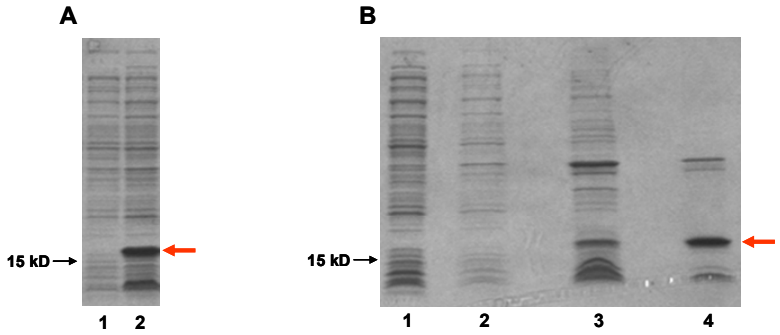

Figure 5.6 SDS-PAGE analysis of the overexpression and purification of the Zif268-Pegasus DZF fusion protein. (A) Analysis of whole cell lysates from uninduced (lane 1) and induced (lane 2) cell cultures transformed with a plasmid encoding the Pegasus DZF domain fused to the C-terminus of Zif268. **(B)** Analysis of samples taken during the inclusion body purification and solubilization. Lane 1 = wash step 1, lane 2 = wash step 2. lane 3 = inclusion bodies, lane 4 = solubilized peptide. Red arrow indicates band which corresponds to the Zif268-Peg fusion protein. Black arrow indicates 15 kD band of the Standard which was run simultaneously with the samples (data not shown).

To further test whether the fusion peptides can be captured by preparing inclusion bodies, 400 ml of induced starter culture were used to isolate inclusion bodies and the peptides were subsequently solubilized in 10 ml of denaturing buffer (Table 5.1). SDS-PAGE analysis of the solubilized inclusion bodies demonstrated that the majority of the Zif268-Pegasus fusion protein was soluble (Figure 5.6B). In contrast, all of the Pegasus-Zif268 fusion protein (data not shown) was trapped in the inclusion bodies and was therefore considered to be insoluble.

5.4.2 Ion exchange chromatography

We decided to continue purifying the Zif268-Pegasus peptide by cation exchange chromatography. The estimated pI value for the Zif268-Pegasus peptide is 8.5 and the peptide was loaded on the column in a buffer of pH 7.5 (Buffer A) followed by differential elution with a gradient of Buffer B (Table 5.1). Surprisingly, the chromatography run bore only 1 well defined peak (Figure 5.7A).

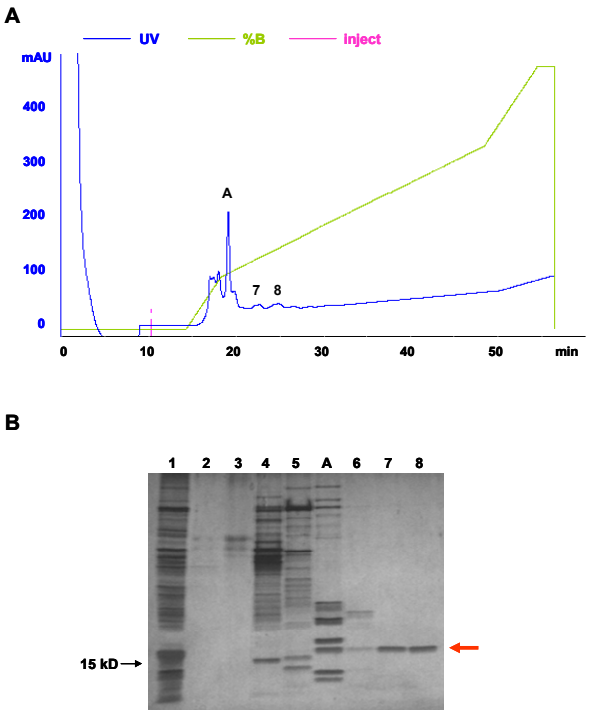

Figure 5.7 Analysis of the ion exchange chromatography run to purify the Zif268-Pegasus fusion protein. (A) Analytical FPLC trace for the Zif268- Pegasus fusion protein. Buffer B gradient applied to elute bound peptides is shown in green. UV trace is shown in blue. Collected peak and two of the collected fractions are indicated (A, 7, and 8). **(B)** SDS-PAGE analysis of the FPLC run from (**A**). Peak fraction A together with several fractions (2-8) taken at different time points were collected and analyzed. Lane 1 represents the solubilized peptide before loading it onto the column. Red arrow indicates Pegasus DZF domain band and black arrow shows where the 15 kD band of the Standard ran. See text for detail.

An aliquot of this peak and several 20 µl fractions at various steps of the run were collected and analyzed by SDS-PAGE and subsequent silver staining. As shown in Figure 5.7B, the Zif268-Pegasus peptide was only partly present in the collected peak fraction but eluted from the column in a very broad range between 29-31 % of buffer B without giving any FPLC signal. Using different loading buffers to improve the binding/elution conditions could not improve the elution profile and the peptide still did not elute in a defined peak. Furthermore, increasing the amount of loaded sample did not result in the occurrence of a clear peak suggesting, that the Zif268-Pegasus peptide can not be detected using a wavelength of 280 nm.

5.4.3 Reverse phase chromatography

Since we were able to capture some of the eluting Zif268-Pegasus peptide, we decided to apply it to a C4 reverse phase column in order to further purify it using RP chromatography. The peptide was acidified to pH 2-3, loaded onto the column in a nonpolar buffer and eluted differentially with a organic mobile phase. The chromatography run bore 4 very small peaks, which eluted at different concentrations of buffer B (Figure 5.8). Peak fractions together with several fractions adjacent to the peaks were captured and analyzed by SDS-PAGE and subsequent silver staining demonstrating that the Zif268-Pegasus peptide eluted without giving a clear signal. Peak C contained some of the sample but did not give a strong signal (data not shown). Thus, as before, the sample could not be detected although a wavelength of 210 nm was used for monitoring the run. The HPLC fraction corresponding to the Zif268-Pegasus peptide was subsequently lyophilized but the dried product did not contain any protein.

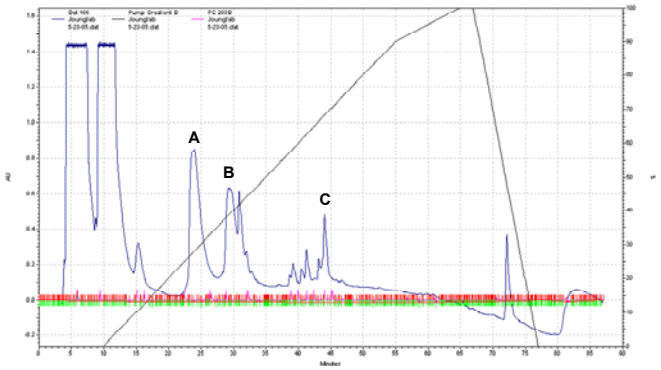

Figure 5.8 Analytical RP-HPLC trace for the Zif268-Pegasus fusion protein. Bound peptides were eluted over a ~60 min period using a Buffer B gradient (shown in black). UV trace is shown in blue. Collected peaks are numbered A, B, C but did not contain the expected peptide.

5.5 Attempted purification of single finger domains from DZFs

Since the purification of Zif268-DZF fusion proteins proved to be very challenging we took a different approach and decided to attempt to purify single finger domains derived from various DZFs for use in X-ray crystallography studies. We constructed four pET3a derived expression plasmids which expressed single N- or C-terminal ZFs of the Ikaros and the

Hunchback *Drosophila* DZF domains (Table 5.1). These plasmids were introduced into the *E. coli* strain BL21(DE3)pLysS. To test, whether any of these peptides could be overexpressed we examined whole cell lysates from uninduced and IPTG induced cells using SDS PAGE and found that the peptides could not be appreciably detected and are therefore not expressed at high levels (data not shown). Efforts to overcome this problem by using different concentration of IPTG, increasing and decreasing induction time or adding glucose were not successful and the expression level could not be improved (note that plasmids containing toxic target genes can be destabilized in stationary phase cultures by expression of the T7 RNA polymerase. This expression is initiated by cAMP mediated derepression of the lacUV5 promoter and can therefore be avoided by adding glucose to the media which inhibits cAMP production).

5.6 Discussion

In order to obtain structural information about a C2H2 ZF mediated protein-protein interaction, we attempted to purify the DZF domains from several Ikaros/Hunchback transcription factors for use in crystallographic studies. When overexpressed, these DZFs accumulated in inclusion bodies and could be solubilized using denaturants (urea) and reducing agents (DTT). Further purification was successfully achieved by ion-exchange and RP-HPLC chromatography and resulted in a very clean sample containing only the respective DZF domain as judged by silver staining. Simultaneously, the denaturant was removed during the HPLC run to permit subsequent refolding of the solubilized proteins. To perform the folding reaction, the peptides were rapidly diluted into folding buffer containing either cobalt or zinc. However, although highly purified peptides were successfully obtained, various attempts to refold these peptides into active domains resulted in the formation of precipitates containing the various DZFs. We can not rule out that aggregation of these DZF peptides is due to an unknown misfolding event but believe that these peptides may just fold fine (as indicated by the blue color). The correctly folded DZFs may aggregate due to the dimerization surface, and DZFs that are usually not able to mediate dimerization may actually dimerize at these high protein concentrations (that are present in the folding reactions). Although we do not know the precise multimerization state of the various DZF domains,

aggregation may just be a consequence of the formation of higher order oligomers as previously described for Eos (Westman *et al.*, 2003) and Ikaros (McCarty *et al.*, 2003).

In general, misfolding as well as aggregation competes with the correct folding event resulting in a dilution of the amount of active peptides. For example, aggregation processes can be caused by nonspecific, hydrophobic interactions of mainly unfolded polypeptide chains. Moreover, correct folding of peptides also depends on correct regeneration of covalent disulfide bonds and it is known that the presence of free thiols can cause complications during the re-folding process due to oxidation problems (reviewed in Rudolph and Lilie, 1996). Thus, various properties of the peptide to be folded can influence the folding process and have to be evaluated. In fact, analysis of the amino acid sequences of the various DZFs that were unsuccessfully refolded indicates that they all contain a highly hydrophobic N-terminal finger consisting of numerous aromatic and aliphatic residues. These residues may have decreased the solubility of the peptides during refolding which in turn resulted in aggregation of folding intermediates. Introducing silent mutations into these hydrophobic residues may help preventing aggregation. In addition, at least one cysteine residue was present in these DZFs predicted not to be involved in zinc ion co-ordination. These cysteines provide free thiol groups that may have formed incorrect disulfide bonds resulting in misfolded peptide aggregates (reviewed in Rudolph and Lilie, 1996). Thus, it may be reasonable to silently mutate all cysteine residues that are not involved in zinc binding.

Several attempts to address the problem of aggregation were undertaken that were all aimed to directly influence aggregation during the folding event. For example, because aggregation usually appears at high protein concentrations (reviewed in Rudolph and Lilie, 1996; reviewed in Rudolph *et al.*, 1998; reviewed in Clark, 1998), renaturation was performed at high dilutions of the protein. Furthermore, various refolding conditions were tested including variables such as buffer composition (pH and ionic strength of the folding buffer) and temperature. Several additives know to enhance the folding process were added to the folding reaction as well. Examples are urea (Orsini and Goldberg, 1978; Maeda *et al.*, 1996), ionic and non-ionic detergents (Tandon and Horowitz, 1987) and sugars (Maeda *et al.*, 1996; Ahn *et al.*, 1997). Although the precise mechanism of action for these additives is not known they are believed to prevent aggregation by either destabilizing incorrect folded intermediates or by stabilizing the correct folded product (reviewed in Clark, 1998). However, none of the tested additives in combination with several buffer conditions was able to decrease the amount of aggregated DZF domains. Other additives such as L-Arginine/HCl (Buchner and

Rudolph, 1991; Brinkmann *et al.*, 1992) or the use of chaperones (Thomas *et al.*, 1997; Altamirano *et al.*, 1997) have also been described as enhancers of the folding reactions but were not tested in this study.

However, since none of the attempts to directly influence the success of the folding reaction worked, it is very likely that this problem can only be solved by using a different folding method. Various methods for refolding of proteins have been described and may be helpful for future attempts to renature the DZF domain: Dialysis is probably the most common method to remove denaturing and reducing agents and therefore allowing the peptide to renature. In doing so, the concentration of the solubilizing agent decreases slowly which allows the protein to refold properly (reviewed in Clark, 1998; reviewed in Rudolph and Lilie, 1996). Another method to remove the denaturant is pulse renaturation. Here, aliquots of denatured protein are added to the renaturation buffer at defined time points, so that the concentration of unfolded protein is kept low. This strategy is based on the observation that during refolding only the concentration of unfolded and not that of correct folded protein is critical for aggregation. The process is stopped when the concentration of denaturant reagent introduced into the renaturation buffer reaches a critical level at which even native peptides tend to aggregate (reviewed in Lilie *et al.*, 1998).

Finally, one can try to address the solubilization problem right at the beginning by avoiding a situation where the DZF peptides aggregate in inclusion bodies in the first place. ZF proteins are generally insoluble but can be linked to a soluble peptide that forces the fusion peptide to stay in solution. In fact, previous studies described various DZF domains fused to the maltose-binding protein (MBP) which were soluble and could be used for biochemical analysis (Westman *et al.*, 2003). In addition, subsequent studies done by this group using single C2H2 ZFs of Pegasus and Eos fused to glutathione S-transferase (GST) resulted in the NMR structure of the C-terminal ZF of the DZF domain from Eos (Westman *et al.*, 2004). However, despite extensive efforts this group did not succeed in crystallizing the complete Eos DZF domain or oligomers generated by this domain. In addition, other groups reported similar technical difficulties including the insolubility of DZFs (Westman *et al.*, 2004, 2003; McCarty *et al.*, 2003). Thus, the crystallization of the DZF domain has proven to be very challenging and will remain the goal of future studies.

Chapter 6. Functional analysis of the Hunchback DZF domain in *Drosophila melanogaster*.

6.1 Introduction

The DZF domain of Ikaros family members has been shown to mediate homo- and hetero-oligomerization and this process is believed to be important to support high affinity DNA binding. This domain is highly conserved among these family members but has also been found in the Hunchback transcription factors from various species. In order to understand the biological role of the Hunchback DZF domain in greater detail, we sought to perform a functional analysis of this domain. We decided to focus on the Hunchback transcription factor from *Drosophila melanogaster*, an organism which has proven to be an ideal model system for studying many biological processes.

In addition to its important function in formation of the anterior-posterior axis in the fly embryo (Lehmann and Nuesslein-Volhard, 1987; Bender *et al.*, 1988; Tautz, 1988; Wimmer *et al.*, 2000; Schroeder *et al.*, 2004), Hunchback has also been shown to be involved in regulating cell fates in the developing central nervous system (CNS) of the *Drosophila* embryo (Brody and Odenwald, 2000; Isshiki *et al.*, 2001; Novotny *et al.*, 2002; Pearson and Doe, 2003; reviewed in Brody and Odenwald, 2002). Development of the CNS requires the generation of a highly complex network of cells with divergent and well defined functions. The question of how one proliferating cell changes over time to generate a sequential ordered series of specified cell types is still one of the most challenging and interesting questions in developmental biology. The stem cell-like precursors of the CNS, termed neuroblasts (NBs), give rise to characteristic invariant cell lineages that produce a diverse population of neurons and glia cells (Bossing *et al.*, 1996; Schmidt *et al.*, 1997). NBs undergo multiple rounds of asymmetric cell divisions to generate a renewed NB and one ganglion mother cell (GMC) at each division, which subsequently produces two neurons and/or glia cells. It has been demonstrated, that the identity of each GMC is determined by its “birth” order within the NB lineage (Doe and Goodman, 1985) suggesting that a NB must constantly change its properties to generate different cells at each division. Recently, four transcription factors have been identified that are expressed in several NB lineages and the spatial and temporal expression pattern of these genes functions as a determinant for the temporal identity of GMC (Isshiki *et*

al., 2001; Kambadur *et al.*, 1998). NBs sequentially express Hunchback (Hd) → Krüppel (Kr) → POU domain protein (Pdm) → Castor (Cas), with GMC and their progenies maintaining the transcription factor profile present at their birth (Figure 6.1). The precise timing of this sequential expression in the NB is essential for proper CNS development (Isshiki *et al.*, 2001).

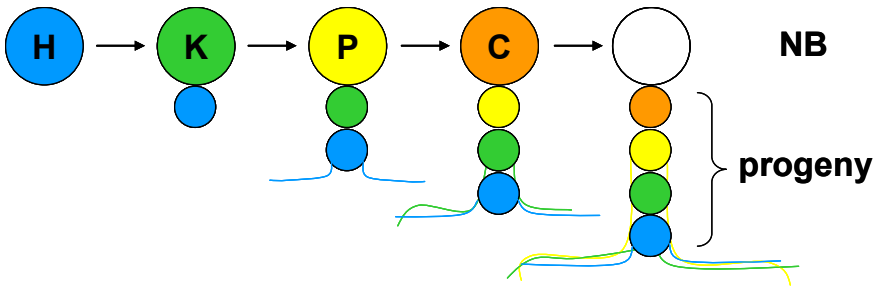

Figure 6.1 Sequential transition in NB gene expression. NBs transiently express Hd → Kr → Pdm → Cas, but their progeny maintain expressing the same transcription factor present during their birth. Cells that express Hunchback (dark blue) are pushed into deeper layers of the developing CNS upon birth of subsequent progeny (Isshiki *et al.*, 2001). The expression pattern of these genes is believed to act as a competence factor, determining the ability of the NB to generate progeny with distinct functions (e.g. neurite projection pattern). Hunchback (H) expressing cells are shown in dark blue, Krueppel (K) expressing cells are shown in green, POU domain protein (P) expressing cells are yellow and Castor (C) expressing cells are orange. White cell indicates that the expressed transcription factor is unknown. This figure was adapted from Brody and Odenwald (2002).

A very well characterized NB lineage is the early forming NB 7-1 (Figure 6.2A). NB 7-1 gives rise to more then 20 GMCs, but only the first five express the nuclear marker Even-skipped (Eve) (Bossing *et al.*, 1996). Each of these five Eve^+ GMCs gives rise to one Eve^+ motoneuron (U1, U2, U3, U4 and U5; note that each GMC also generates a second Eve^- neuron which can not be tracked because no molecular marker is present). It has been demonstrated that the first two GMCs and their two Eve^+ motoneuron progeny (U1 and U2) express Hunchback while later progeny do not show Hunchback expression. Hunchback mutants lack these two first-born Eve^+ GMCs which results in a loss of Eve^+ U1 and U2 motoneurons. In contrast, over-expression of Hunchback in NB 7-1 results in an excess of early born GMCs at the expense of later-born GMCs and all differentiate as early-born U1/U2 motoneurons (Figure 6.2B; Isshiki *et al.*, 2001). Interestingly, ectopic expression of Hunchback containing a point mutation in one of the zinc ligating cysteines of the DZF domain results in a phenotype similar to the wild-type phenotype with motoneurons U1-U5 being generated again (personal communication J. Urban). Thus, Hunchback is necessary for

specifying early-born cell fates and the C-terminal DZF domain in Hunchback seems to be important for maintaining this function. However, despite the importance of the DZF domain, remarkably little is know about the mechanism involved.

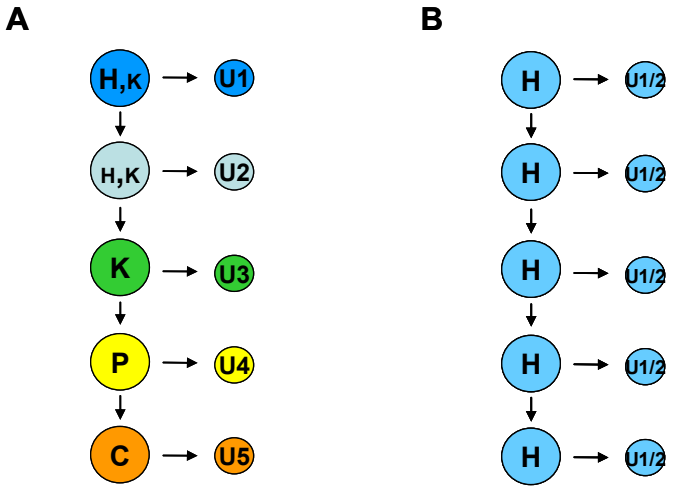

Figure 6.2 Characteristic transcription factor expression pattern of the early NB 7-1 lineage. (A) In wild-type embryos, the NB 7-1 initially performs five cell divisions resulting in five GMCs and each of these GMCs give rise to one motoneuron (U1-U5). The expression pattern present in the NB is retained in the motoneuron (as indicated by the colors) and determines the identity of the respective motoneuron. Note that in order to simplify matters, the GMCs are not shown. Each dividing GMC produces two siblings but only one can be identified and is therefore shown (U1-U5, respectively). Big letters indicate high expression level of the respective transcription factor and small letters indicate low levels of expression. Abbreviations are as defined in Figure 6.1. **(B)** Ectopic expression of Hunchback throughout the NB 7-1 results in an over-production of early born motoneurons (U1 and U2, indicated in blue) at the cost of later born neurons.

The experiments of this chapter are aimed at exploring if Hunchback uses its DZF to dimerize in *Drosophila melanogaster* and if this dimerization is essential for the function of the protein. Thus, constructs encoding Hunchback with various natural and modified DZF domains were generated. Both, DZF domains that mediate homodimerization as well as DZFs that can only heterodimerize were used and transgenic flies containing these constructs were established. Over-expression of these constructs in NB 7-1 was then used as an assay to analyze whether swapping the natural Hunckback C-terminus with DZFs that either can or can not dimerize will restore the function of the wild-type protein. Furthermore, Hunchback variants that prefer to heterodimerize are expected to cause a mutant phenotype when expressed individually but should be able to rescue the wild-type function of Hunchback upon co-expressing them simultaneously. With this approach we sought to answer the question if DZF mediated homodimerization of the Hunchback protein is important for its

biological function. In addition, transgenic flies expressing tagged versions of the various constructs were generated to eventually investigate whether Hunchback homodimers could be identified in the fly.

6.2 Generating transgenic flies

6.2.1 Overview: The GAL4-UAS system

The GAL4-UAS system was used to express Hunchback in a cell- and tissue specific pattern. The principle of this system is demonstrated in Figure 6.3 and described below.

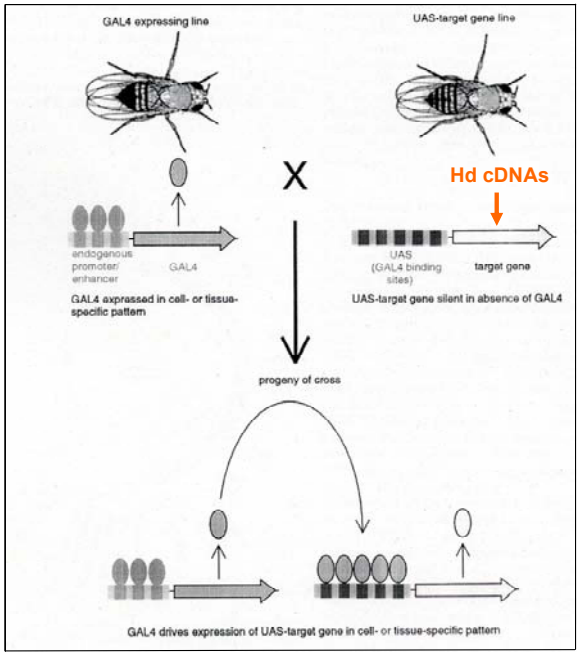

Figure 6.3 The GAL4/UAS system. The GAL4-line expresses the GAL4 activator under the control of a cell- or tissue specific promoter (left) while the UAS-line (right) harbors the target gene (e.g. Hunchback) downstream of the upstream activation sequence (UAS). Crossing these two transgenic lines together generates progeny which carry both elements (GAL4 and the UAS controlled target gene). In these flies, the cell-specific promoter drives expression of GAL4 which in turn activates the transcription of the target gene (e.g. Hunchback) in the same set of cells. This Figure was taken and adapted from Phelps and Brand (1998).

This bipartite ectopic expression system utilizes the yeast transcription factor GAL4 which binds to its target DNA sequence, the upstream activation sequence (UAS), thereby activating gene transcription. GAL4 can be expressed in a cell specific pattern by placing it under the control of various *Drosophila melanogaster* promoter (or enhancer) sequences. Subsequently, GAL4 activates transcription of the GAL4-responsive (UAS) target gene in an identical set of cells. The GAL4 gene and the UAS-target gene are maintained in two distinct transgenic

lines. Thus, the GAL4 activator protein is inactive in one line (GAL4-line) while the target-gene is not expressed in the other line (UAS-line) because GAL4 is missing. After crossing the GAL4 line to the UAS-line the target gene is turned on in the resulting progeny (Brand and Perrimon, 1993; Phelps and Brand, 1998).

6.2.2 Description of pUAST-Hunchback constructs

In order to characterize the function of the DZF domain in *Drosophila melanogaster*, constructs encoding full-length Hunchback proteins with either natural or synthetic DZF domains at the C-terminus were constructed. The DZF domain from wild-type Ikaros and two synthetic DZFs (Tr-Eo-Eo and Eo-Hd-Hd) were used to replace the Hunchback DZF domain in the wild-type full-length protein. While the Ikaros DZF domain can mediate homo-dimerization, Tr-Eo-Eo and Eo-Hd-Hd prefer to hetero-dimerize and interact either only weak or not at all with themselves (see chapter 3, section 3.4.1). Care was taken to keep the introduced DZF domains as small as possible so that they all started with the conserved Tyr/Phe two residues N-terminal of the first zinc-ligating Cys (see Figure 3.1, chapter 3). All fragments continued to the authentic C-terminus of the respective DZFs (as defined in Chapter 3) terminating 1-5 residues after the final zinc-binding histidine. For each fragment, additional versions containing either a myc- or a HA-tag were designed by fusing these tags to the C-terminus of the DZF domains using a GEKP linker.

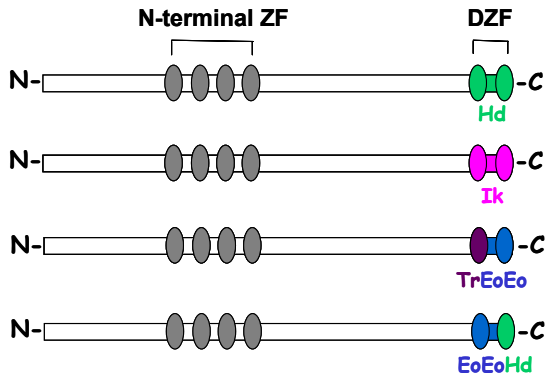

Figure 6.4 List of full-length Hunchback germ-line transformation constructs containing swapped DZF domains. The DZF domain from Hunchback (green ovals) was replaced with the DZF domain from wild-type Ikaros (pink ovals) as well as with the synthetic DZFs Tr-Eo-Eo (purple-blue ovals) and Eo-Eo-Hd (blue-green ovals). Note that tagged versions of wild-type Hunchback (myc and HA tag), Hunchback Ik (myc and HA tag), Hunchback TrEoEo (myc tag) and Hunchback EoEoHd (HA tag) were also generated.

For expression in the GAL4-UAS system, the resulting 10 fragments (Figure 6.4 and not shown) were fused to the N-terminal domain of wild-type Hunchback and cloned into the

pUAST plasmid placing them under the control of the UAS promoter (Brand and Perrimon, 1993; see section 6.2.1).

6.2.3 *P*-element transformation

P-element transformation was performed to introduce the various Hunchback constructs into flies. To do this, the pUAST plasmids containing *P*-elements up- and downstream of the UAS-*hunchback* DNA were injected into the germline of *white*$^{-}$ fly embryos that express the transposase gene (Greenspan, 1997). These *P*-elements can then "randomly" insert themselves stably into genomic DNA. Since pUAST also expresses the *white*$^{+}$ gene in-frame with the various *hunchback* fragments, transformants were selected using eye color as a marker for successful insertion of the *P*-element (expression of the *white*$^{+}$ gene results in flies with red eyes). To produce stable stocks, flies containing the *P*-elements were further crossed to strains containing dominantly marked balancer chromosomes and the location of the insertions was thereby mapped to a particular chromosome (Table 6.1; Greenspan, 1997).

cDNA	Number of transgenic lines			
	X	II	III	total
Hd	1	3	8	12
Hd-myc	2	2	1	5
Hd-HA	0	4	6	10
Hd Ik	0	1	1	2
Hd Ik-myc	1	1	2	4
Hd Ik-HA	1	3	4	8
Hd TrEoEo	1	8	6	15
Hd TrEoEo-myc	0	9	6	15
Hd EoEoHd	0	3	5	8
Hd EoHdHd-HA	1	6	5	12

Table 6.1 Summary of transgenic flies obtained for each Hunchback construct. Locations of inserted constructs (left column) were mapped to a particular chromosome as indicated in column 2, 3 and 4. X indicates the X-chromosome, II indicates the second chromosome and III indicates the third chromosome. Note that flies carrying the constructs both homozygous and heterozygous were obtained, although this is not further specified here.

6.3 Expression analysis of the constructs

6.3.1 Rough eye phenotype

The location of the inserted *P*-element can greatly affect the expression level of the transformed gene and hence any resulting phenotype can vary even when a binary system like the GAL4/UAS system is used. Consequently, the levels of UAS-cDNA expression vary among lines harboring the same construct in the presence of equal amount of GAL4 activator. Thus, the expression level of various independent *P*-element lines was evaluated. Initially, we focused on the constructs containing either a myc- or a HA-tagged DZF, which could be used for Western-blot analysis using anti-myc or anti-HA antibodies, respectively. We decided to over-express these constructs in the *Drosophila* eye, which provides an excellent tissue to produce a lot of protein, even if this protein would otherwise be harmful for the fly. Although several genes have been described which affect eye development upon ectopic expression (Hay *et al.*, 1997; Rørth *et al.*, 1998), we were not expecting to see any side effects from over-expressing our proteins, since Hunchback has not been reported to be involved in eye development.

Figure 6.5 Overexpression of Hunchback in the eye causes a rough eye phenotype. Ectopic expression of Hunchback using *gmr-gal4* as a driver results in a rough eye phenotype with various levels of severity from mildly rough, almost normal sized eyes to very rough or glassy eyes ("++++" right column). The middle column shows a phenotype with medium ("++") severity present in most of the flies after overexpression the various Hunchback constructs. A control harboring only the *gmr-gal4* driver is shown in the left column. F-2 / F-3 indicate the respective construct which was overexpressed in the shown flies. These pictures were taken from female flies with a Canon EOS 30D camera using an MP65 object lens.

The *gmr-gal4* line is a commonly used eye-specific driver line that contains five copies of the Glass response element from the Rhodopsin 1 gene and drives expression in the eye in all cells behind the morphogenetic furrow (Freeman, 1996). It has been shown that *gmr-gal4* homozygotes have a visible rough eye phenotype (characterized by a smaller size of the eye) whereas heterozygotes appear normal (Freeman, 1996). Transgenic flies containing the tagged UAS-constructs were crossed to the *gmr-gal4* driver to initiate expression of these constructs in the *Drosophila* eye. Interestingly, for some transgenic lines progenies of these crossings had a rough eye phenotype that was visible under the dissecting microscope (Figure 6.5). This phenotype was only seen in flies containing both the *gmr-gal4* and the UAS-Hunchback constructs, indicating that the rough eye is due to GAL4 initiated expression of Hunchback in the eye. The visible phenotypes varied in severity from mildly rough, almost normal sized eyes to very rough or glassy eyes, which were reduced to almost 1/2 of the normal size (Figure 6.5, Table 6.2 and Table 6.3). Progenies with very severe eye phenotypes were usually obtained at lower frequency suggesting that the crossing was semi-lethal. Thus, regardless of the mechanism by which Hunchback disrupts eye development, it is apparent that it has an effect when ectopically expressed.

cDNA	Intensity of phenotype				
	-	+	++	+++	++++
Hd	0	2	3	1	1
Hd-myc	0	0	2	0	0
Hd-HA	0	0	1	2	0
Hd Ik	0	0	2	0	0
Hd Ik-myc	0	0	0	2	0
Hd Ik-HA	2	1	3	0	0
Hd TrEoEo	1	4	5	1	0
Hd TrEoEo-myc	1	1	7	1	0
Hd EoEoHd	0	3	3	0	1
Hd EoHdHd-HA	0	2	4	2	0

Table 6.2 Analysis of eye phenotype severity obtained for the different Hunchback constructs. Expression level of constructs was estimated by analyzing the resulting eye phenotype after overexpressing the constructs in the eye using the *gmr-gal4* driver. Digits indicate numbers of transgenic lines obtained with the particular phenotype. Eye phenotype was gradually and subjectively judged where "–" indicates normal wild-type eyes and "++++" indicates a severe phenotype with eyes almost 1/2 of the size of wild-type eyes. Note that not all transgenic flies obtained for a certain cDNA construct were tested in this assay.

In general, there was no correlation between severity of the phenotype and a particular construct. In fact, different insertions lines of a particular construct seemed to cause different

levels of severity, although the same Hunchback protein is over-expressed in these lines (Table 6.2). This suggests that the severity of the phenotype depended on the expression level of the Hunchback protein, which in turn was depending on the location of the P-element insertion.

6.3.2 Western blot analysis

To further verify this hypothesis, Western blot analysis was performed to analyze the expression level of the tested transgenic lines which carried a tag epitope. Lysates from fly head extracts were resolved by SDS-polyacrylamide gel electrophoresis and the proteins were visualized using anti-myc or anti HA antibodies that recognize the respective C-terminal tag epitope. It was found that most of the peptides were stably expressed (Figure 6.6 and data not shown). Judged by their electrophoretic mobility, the relative weight of these peptides is ~ 105 kD which is higher then the calculated molecular weight of Hunchback (~85 kD). This could be due to modifications of the protein in the fly. Control experiments using constructs without tags did not result in any bands demonstrating that the antibody stained specifically the respective tag (data not shown).

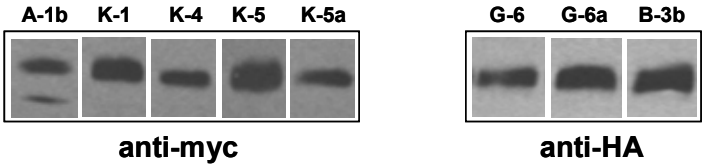

Figure 6.6 Western blot analysis determining the expression levels of different Hunchback constructs. Protein-extracts obtained from 40 fly heads for each transgenic type were resolved by SDS-polyacrylamide gel electrophoresis and the proteins were visualized using anti-myc or anti-HA antibodies. Construct numbers correspond to a particular Hunchback cDNA as indicated in Table 6.3. Note that the bands shown (either stained with anti-myc or anti-HA) were obtained from one blot and can therefore be compared to each other.

As summarized in Table 6.3, for most of the peptides the expression level of the various constructs is consistent with the severity of the rough eye phenotype. Thus, the higher the amount of expressed peptide was, the more severe was the phenotype this peptide had caused (Table 6.3). An Exception was protein UAS-Hd Ik-myc, which could not be detected by Western blot although this peptide caused a severe rough eye phenotype suggesting that it was expressed at a high level. Furthermore, Western blot analysis of wild-type Hunchback

tagged with myc exhibited a second band not present in the other samples (Table 6.3 and Figure 6.6). Thus, over-expressing Hunchback variants in the eye and subsequent examination of the eye phenotype can generally be used as a method to estimate the expression level of the respective constructs. Accordingly, transgenic lines with similar expression levels of the respective constructs as judged by the eye phenotype were selected for further analysis.

cDNA	Transgenic line	Intensity of rough eye phenotype	Band intensity of Western blot
Hd--myc	A-1	++	+
	A-1b	++	++
Hd--HA	B-2a	+++	+++
	B-2b	++	++
	B-3b	+++	+++
Hd--Ik--myc	E-1	+++	-
	E-1a	+++	-
Hd--Ik--HA	H-1c	+	+
	H-1d	+	+
	H-2	-	-
Hd--TrEoEo--myc	K-1	+++	+++
	K-4	++	++
	K-5	+++	+++
	K-5a	++	++
Hd--EoHdHd--HA	G-6	++	++
	G-6a	+++	+++

Table 6.3 Analysis of the expression level for the different Hunchback constructs. Comparison of the resulting eye phenotype after overexpressing the respective construct with the band intensity after Western blot analysis. Eye phenotype was judged as described in Table 6.2 and Figure 6.5. The band intensity obtained by Western blot analysis was judged in a similar manner with "–" indicating that no band was detected while "+++" indicates that band intensity was strong. Note that only constructs tested for both, the eye phenotype and Western blot band intensity are shown.

6.4 Overexpression in Neuroblast

6.4.1 Triple staining of a wild-type embryo

As described above, over-expression of Hunchback and its variants can be used as a "functional" assay for analyzing the importance of DZF mediated dimerization. For these experiments, the well characterized NB lineage 7-1 was chosen as a model system which will be introduced in the following section. 7-1 neurons can be easily identified in stage 14

embryos since they express a combination of specific molecular markers as well as by their characteristic location in the embryo (Isshiki *et al.*, 2001). A typical staining of a wild-type embryo is shown in Figure 6.7.

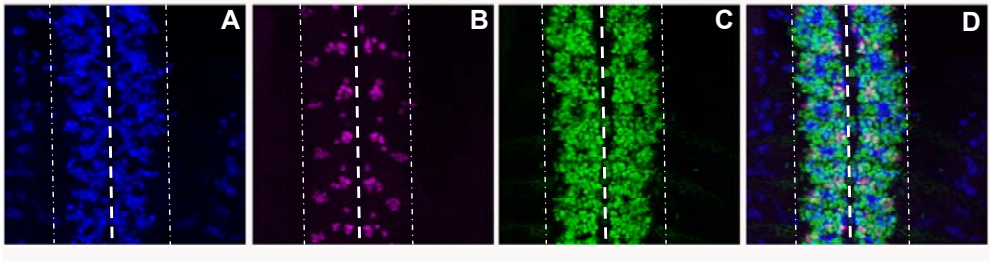

Figure 6.7 Triple staining of a wild-type embryo at stage 14. Dorsal view of a wild-type embryo stained for Hunchback (**A**, in blue), Eve (**B**, in magenta) and Zfh-2 protein (**C**, in green). (**D**) shows a merge of the three stainings. Sections of the confocal images were combined representing various layers from dorsal to ventral, anterior is up. Thick dashed lines indicate the ventral midline and thin dashed lines define lateral borders of the ventral nerve cord. Note that the total width of the ventral nerve code (which corresponds to the region within the lateral borders) is approximately 60 µm. See text and Chapter 2, section 2.5.2.4 for details.

Panel A represents a staining using an antibody against Hunchback. This staining generally served as a control for the expression level and location of the various Hunchback variants (note that in this wild-type embryo Hunchback was not over-expressed, thus, reflecting the background of Hunchback staining). As described in the introduction, NB 7-1 generates a set of defined cells that express the marker eve. Thus, by using an antibody against eve, NB 7-1 and its progenies could be easily identified (Figure 6.7, panel B, stained in pink). Another molecular marker that labels late sub-lineage neurons is the transcription factor Zinc finger homeodomain 2 (Zfh-2) which is stained in panel C (Figure 6.7, in green) using an anti-Zfh-2 antibody. Panel D in Figure 6.7 shows a merge of these three antibody stainings.

6.4.2 Hunchback represses expression of Zfh-2

To begin investigating whether dimerization of Hunchback is necessary for its function in NB development, the Hunchback variants were over-expressed in NB 7-1. *Engrailed-gal4* (*en-gal4*, Tabata *et al.*, 1995) was used as a driver since it drives expression in defined NBs. Thus, *en-gal4* was crossed to the different UAS-Hunchback variants to initiate expression of these constructs and resulting embryos together with wild-type embryos were collected after an overnight incubation to allow development to the desired stage. These collections were

then stained, dissected and analyzed as detailed in Chapter 2, section 2.5.2.4. Staining against Zfh-2 in stage 14 embryos demonstrated that ectopic wild-type Hunchback expression resulted in an absence of Zfh-2 expression in Hunchback expressing neurons (Figure 6.8B and J. Urban, personal communication). While in wild-type embryos Zfh-2 is equally expressed in all late sub-lineage neurons, Hunchback clearly seemed to inhibit this expression (Figure 6.8, compare A and B). Regions where Zfh-2 expression is inhibited corresponded well to rows where Hunchback was over-expressed as labeling for this protein confirmed (data not shown). When NB 7-1 is forced to continuously express Hunchback containing either the Ikaros or the synthetic Eo-Eo-Hd DZF domain, Zfh-2 expression was also absent in these neurons (Figure 6.8C, E). In contrast, embryos over-expressing Hunchback with the synthetic Tr-Eo-Eo DZF domain displayed a different phenotype with Zfh-2 equally expressed as it was the case for wild-type embryos (Figure 6.8D). This suggested that the C-terminal domain in Hunchback can be replaced by another dimerization domain although it remains unknown if this domain has to be functional. Both the Eo-Eo-Hd and the Tr-Eo-Eo DZF domain have been shown not to mediate homo-dimerization and were therefore expected to display a similar mutant phenotype when introduced into the Hunchback protein. However, only Tr-Eo-Eo affected the Hunchback function of inhibiting the Zfh-2 expression.

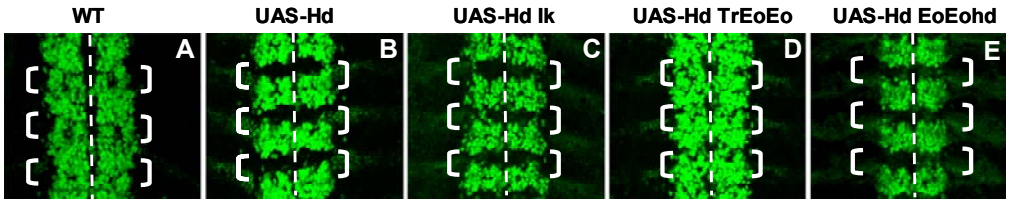

Figure 6.8 Ectopic expression of Hunchback in the NB 7-1 inhibits Zfh-2 expression. Dorsal view of embryos at stage 14 stained for the Zfh-2 protein (green). Anterior is up, dashed lines indicate the ventral midline. Square brackets indicate the *engrailed* expression domain where Hunchback expression was driven. **(A)** In wild-type embryos Zfh-2 is equally expressed in all segments. **(B)** Ectopic expression of Hunchback using *engrailed* (*en*) *gal4* as a driver diminishes expression of Zfh-2 in these segments. **(C)** Over-expression of Hunchback harboring the DZF domain of Ikaros also inhibits Zfh-2 expression in the *en-gal4* expression domain. **(D)** Forcing the NB to express Hunchback containing the TrEoEo DZF domain results in a wild-type like phenotype and Zfh-2 is equally expressed in all segments. **(E)** Over-expressing Hunchback harboring the EoEoHd DZF domain abolishes Zfh-2 expression in the ectopic expression segments. Figure 6.8A was kindly provided by Ulricke Mettler.

6.4.3 Hunchback misexpression changes identity of Motoneurons

To further investigate whether the C-terminal DZF domain in Hunchback has to be functional for dimerization, double stainings with antibodies against Zfh-2 and eve were analyzed. The molecular marker Eve is only present in five motor neurons which are derived from NB 7-1.

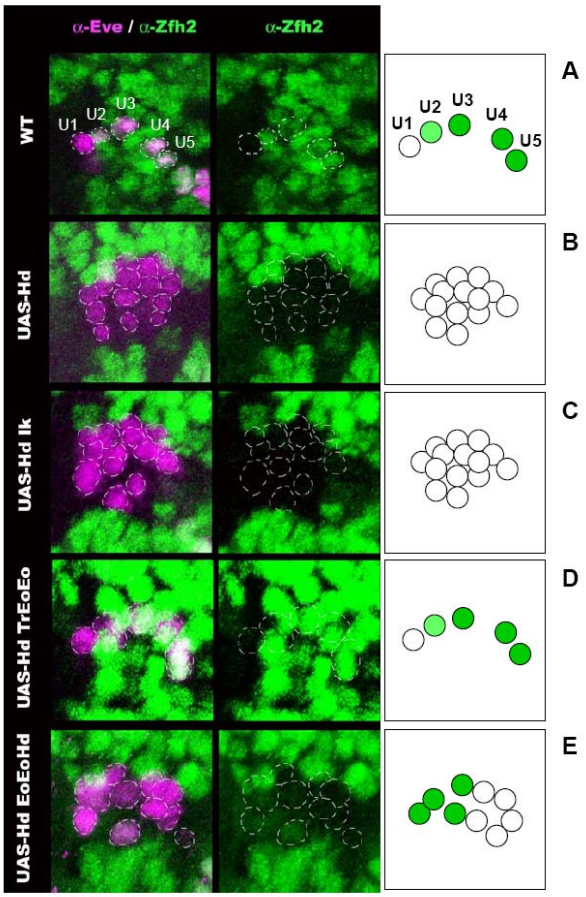

Figure 6.9 Ectopic expression of Hunchback in the NB 7-1 results in the over-production of early born motoneurons. Embryos at stage 14 either double labeled for the 7-1 progeny marker Eve (pink) and Zfh-2 (green) are shown in the left panel. The same images are shown again as single staining labeled for Zfh-2 in the middle panel where cells expressing eve are indicated by white dashed circles (Note that these cells are not visible in these images because the staining was performed only against Zfh-2 and the two proteins are not co-expressed in the same cells). A summary of the phenotypes is indicated in the right column. White circles represent cells which do not express Zfh-2, while green circles indicate cells that express either weak (light green) or normal amounts (dark green) of Zfh-2. All panels represent combined confocal images consisting of multiple focal planes. Anterior is up, midline is to the left. Note that the width of the region comprehending the group of five U-motoneurons is approximately 15 µm. **(A)** In wild-type embryos, motoneuron U1-U5 can be identified by Eve expression (magenta) and their typical position (see also top right panel). U1 and U2 motoneurons are known to express Hunchback (Isshiki *et al.*, 2001) which inhibits the expression of Zfh-2 completely (U1) or partly (U2). **(B, C)** Ectopic expression of wild-type Hunchback **(B)** or Hunchback harboring the Ikaros DZF domain **(C)** results in the production of extra Eve^{+} cells and the expression of Zfh-2 in these cells is inhibited. **(D)** Ectopic expression of Hunchback containing the TrEoEo DZF domain produces five motoneurons and 1 or 2 of these cells are $Zfh\text{-}2^{-}$. **(E)** Over-expressing of Hunchback containing the EoEoHd DZF domain reduces the amount of Eve^{+} cells compared to ectopic expression of wild-type Hunchback, although there are still excessive motoneurons present. These cells partly express Zfh-2. Figure 6.9A was kindly provided by Ulricke Mettler.

Thus, in wild-type embryos these five neurons (U1, U2, U3, U4 and U5) can be uniquely identified by Eve expression as well as their stereotyped cell position (Figure 6.9A, Isshiki *et al.*, 2001). The first two motor neurons U1 and U2 are Hunchback^{+} while later progeny (U3, U4 and U5) do not express Hunchback. Closer examination of these two neurons in wild-type embryos indicates that they did not (U1) or only mildly (U2) express Zfh-2 whereas the remaining motor neurons were Zfh-2^{+} (Figure 6.9A). This further confirmed that Hunchback represses the expression of Zfh-2. When NB 7-1 was forced to continuously express wild-type Hunchback, additional Eve^{+} cells were found (average of 14 cells, Table 6.4), which were all Zfh-2^{-} (Figure 6.9B). Thus, ectopic expression of Hunchback resulted in an over-production of early-born neurons at the cost of later-born cells. Over-expression of Hunchback with the swapped Ikaros DZF had the same phenotype producing extra Eve^{+} cells that did not express Zfh-2 (Figure 6.9C, Table 6.4). In contrast, when Hunchback containing the synthetic Tr-Eo-Eo DZF domain was misexpressed a phenotype similar to wild-type embryos was obtained (Figure 6.9D, Table 6.4). Again, five Eve^{+} motor neurons were detected with the first two not expressing Zfh-2. Forcing the NB to over-express Hunchback with the synthetic Eo-Eo-Hd DZF domain resulted in extra Eve^{+} cells although the number was reduced compared to overexpression of wild-type Hunchback (average of 9 cells for Eo-Eo-Hd compared to 14 for wild-type Hunchback, Table 6.4). Furthermore, several of these cells expressed Zfh-2 which was not the case after ectopic expression of wild-type Hunchback (Figure 6.9E). Thus, closer examination of the overexpression responses for the various Hunchback variants indicates that the DZF domain had to be functional to replace the wild-type Hunchback DZF. While the Ikaros DZF could fully replace the Hunchback DZF domain, the two synthetic DZFs could not, suggesting that DZF mediated dimerization of Hunchback may be important for regulation of NB competence.

cDNA	Number of U-neurons	Zfh-2^{-}
--	5	1.5
Hd	13,9	13.9
Hd Ik	12.9	12.6
Hd TrEoEo	5.5	1.8
Hd EoEoHd	8.9	4.4

Table 6.4 Average number of U-neurons in NB 7-1. Number of motoneurons (U-neurons) present in NB 7-1 were counted after ectopic expression of the different Hunchback constructs (left column). Zfh-2^{-} neurons in these sets of motoneurons are also indicated (right column). These numbers represent the mean of 18 scored hemisegments. Note that wild-type flies are presented in the first row indicated by a "--" since no cDNA was overexpressed in these flies.

6.5 Discussion

6.5.1 The DZF domain is important for regulating NB competence in *Drosophila*

This chapter aimed to understand the importance of DZF mediated dimerization for Hunchback in terms of its function as a regulator of CNS development. Hunchback is involved in specifying early-born temporal identity in *Drosophila* neural stem cell lineages and the C-terminal domain in Hunchback seems to be important for maintaining this function (Isshiki *et al.*, 2001; Novotny *et al.*, 2002; Pearson and Doe, 2003; J. Urban, personal communication). This is in contrast to its function in early embryonic development where the C-terminal C2H2 ZFs are not required (J. Urban, personal communication). To distinguish whether these C-terminal C2H2 ZFs must mediate dimerization for biological function or are merely required for the stability of Hunchback, we constructed plasmids expressing Hunchback in which the C-terminal DZF domain was replaced with other DZFs that either do or do not support dimerization. The expectation was that these domains would fold in a manner similar to the Hunchback DZF and therefore should not disturb the overall folding and integrity of the Hunchback protein. Thus, importance of dimerization itself could be analyzed by testing the ability of these DZF variants to fulfill the biological function of wild-type Hunchback. Ectopic expression of wild-type Hunchback in NBs can transform all GMCs towards a first-born fate resulting in an overproduction of neurons with early fates at the expense of later born neurons. Interestingly, this phenotype was also reported when the NB was forced to express Hunchback containing the homo-dimer forming Ikaros DZF domain. However, misexpression of a Hunchback construct in which the C-terminal domain was replaced by a DZF domain that did not support homodimerization (Tr-Eo-Eo) resulted in generation of later-born neurons, representing a phenotype similar to the one seen in wild-type embryos. In summary, these results lead to two major conclusions: First, the DZF domain in Hunchback participates and is necessary for maintaining the competence of NBs to generate early born progeny during CNS development. Introduction of “mutant” DZFs by replacing the wild-type DZF domain with a nonfunctional DZF abolished this function. Second, the importance of the DZF domain is an outcome of its ability to mediate dimerization since it can be swapped by another functional dimerization domain. Thus, the C-terminal C2H2 ZFs in the Hunchback protein appear to be required for generating first-born cell fates only because dimerization is essential for this process.

6.5.2 Subjects for future studies to confirm these findings

It is important to note that when the NB was forced to express Hunchback containing the synthetic Eo-Eo-Hd DZF, there were still some extra early-fate cells although this domain should have displayed a similar phenotype to cells misexpressing Hunchback Tr-Eo-Eo. We do not have an explanation for this, but speculate that the expression level of our constructs may have influenced the outcome of the experiment. Although care was taken to choose transgenic lines displaying a similar expression level, we can not rule out that some of these constructs produced more protein then others due to position effects (see Tables 6.2 and 6.3). Another concern was that the amount of protein expressed in the NBs was generally too high to be sensitive enough to detect "mutant" phenotypes. In fact, a similar problem was obtained in our original attempts to perform an alanine-scan of the Hunchback DZF domain where the expression level was too high to detect any mutations in the B1H assay (see chapter 4 and data not shown). Thus, additional experiments will be required to reliably uncover a mutant phenotype. More transgenic lines have to be tested and it might be reasonable to decrease the temperature for inducing the expression because it is known that higher temperatures increase the expression level in the GAL4/UAS system (Brand and Perrimon, 1993; Phelps and Brand, 1998).

It is noteworthy, that initial experiments co-expressing the two heterodimeric Hunchback variants containing the synthetic DZFs (Eo-Eo-Hd and Tr-Eo-Eo) in NBs were performed. The result of these experiments suggested that the two proteins may mediate heterodimerization since co-expression resulted in a phenotype similar to that obtained when over-expressing wild-type Hunchback (data not shown). On the other hand, this phenotype might have been caused by the high amount of protein present (because two proteins were simultaneously expressed) that might have resulted in the same phenotype compared to the one obtained after overexpressing the wild-type Hunchback protein. Thus, controls expressing equally high amounts of the same Hunchback protein (either Hunchback Eo-Eo-Hd or Hunchback Tr-Eo-Eo) have to be tested to rule out the possibility of such a dosis effect.

Furthermore, it will also be important to determine if a Hunchback dimer exists in *Drosophila* since this would further confirm the importance of DZF mediated dimerization. Transgenic flies expressing tagged versions of the various Hunchback constructs were generated in this study and have been shown to be stably expressed. Thus, these lines will be

subject to future co-immunoprecipitation experiments aimed to reveal if Hunchback homo-dimers exist *in vivo*.

6.5.3 Biological role of dimerization at a mechanistic level

It is interesting to speculate on the biological role of Hunchback dimerization. The competence of NB 7-1 to respond to Hunchback has been further characterized (Pearson and Doe, 2003; Grosskortenhaus *et al.,* 2005; Cleary and Doe, 2006). NB 7-1 shows progressive limitations in its ability to produce early fates if only pulses of Hunchback are provided at different time points of the lineage (Pearson and Doe, 2003). However, if Hunchback levels are kept constantly high throughout the lineage, the NB will extend its competence and will produce early fates even after several cell divisions. After subsequent down-regulation of Hunchback the NB takes on the normal fate again by producing later born type neurons. Thus, the NB responds to Hunchback in a well defined manner by extending its ability to react to Hunchback over-expression but also by maintaining full competence to generate later-born neurons (Grosskortenhaus *et al.,* 2005). The ability of Hunchback to keep the NB in a "young" state despite many cell divisions raises the question of how Hunchback acts at a mechanistic level. In the case of Ikaros, the mammalian homologue of *Drosophila* Hunchback, several models have been proposed of how transcriptional regulation occurs. It has been shown that Ikaros associates with chromatin and remodeling proteins and is essential for silencing gene expression in mature B cells during hematopoiesis (Kim *et al.*, 1999; Sabbattini *et al.*, 2001; see also section 1.3.3.1). Furthermore, Ikaros seems to co-localize to pericentromeric heterochromatin (PC-HC) together with transcriptionally silenced genes suggesting that it recruits these genes to heterochromatin regions through an unknown mechanism (Brown *et al.*, 1997; Klug *et al.*, 1998; Cobb *et al.*, 2000; Koipally *et al.*, 2002). This process is dependent on both DNA binding and dimerization of Ikaros (Cobb *et al.*, 2000; Koipally *et al.*, 2002; see also section 1.3.3.2). Hunchback may act in a similar matter to prevent expression of later temporal identity genes by repositioning these genes to transcriptional inaccessible regions within the nucleus and thereby silencing them. Thus, dimerization might permit the protein to interact simultaneously with the target genes and regions of pericentromeric heterochromatin (Figure 6.10A).

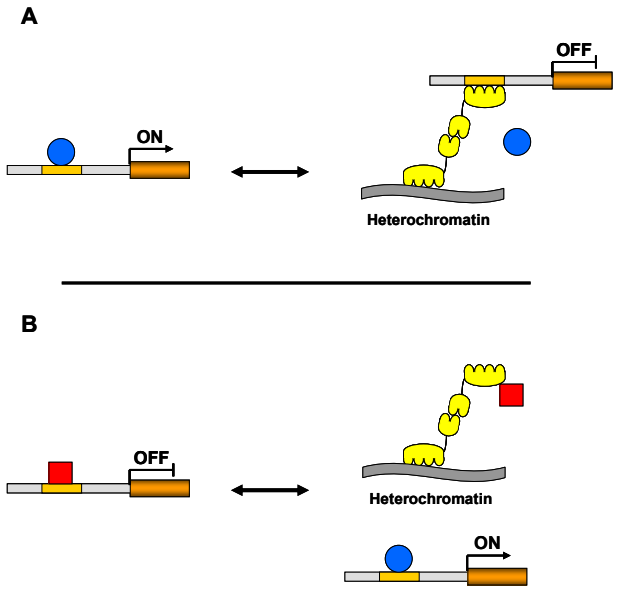

Figure 6.10 Models for the biological role of a potential Hunchback homodimer. (A) Hunchback (yellow) may target a specific gene (orange) to transcriptionally inactive pericentromeric heterochromatin (PC-HC, dark grey wave) through dimerization, whereby one Hunchback monomer binds centromeric sequences while the other monomer interacts with sequences in a specific target gene. This would relocate this gene to regions that are inaccessible to transcriptional regulators including activators (blue circle). (B) Alternatively, the Hunchback dimer may stimulate gene activity by recruiting repressor complexes (red rectangle) to PC-HC. Activators would then be able to bind to specific genes and drive transcription. This figure was adapted from Cobb *et al.*, (2000) and Koipally *et al.*, (2002).

On the other hand, it has also been suggested that Ikaros enables gene expression by trapping repressor complexes at promoters of target genes and relocating them to PC-HC, thereby allowing potential activators to bind to these target sites again (Koipally *et al.*, 2002). Thus, to further maintain plasticity of gene expression which keeps the NB in a multipotent state Hunchback may also utilize such a mechanism. In doing so, Hunchback may act as a dimer, with one subunit bound to centromeric regions and the other available for interactions with potential repressor complexes (Figure 6.10B). A combination of these mechanisms --partly accomplished by the DZF domain-- would allow Hunchback both to prevent expression of later temporal identity genes as well as maintain the multipotent state of the neuroblast.

6.5.4 Using the eye phenotype to identify components of Hunchback regulated processes

Finally, it is noteworthy that ectopic expression of Hunchback in the *Drosophila* eye caused a rough eye phenotype. Although the mechanism and biological significance of this result is not clear, it may point towards a role of Hunchback in eye development. In the past decade the *Drosophila* eye has proven to be a powerful model system to study the function of any gene of interest. While *in vivo* disruption of target genes has proven to be difficult, misexpression analysis in the *Drosophila* eye seems to be a convenient strategy aimed to understand the function of any gene of interest (reviewed in Thomas and Wassarman, 1999).

Although eye development is usually severely influenced by misexpression of a gene of interest this has no consequences for the viability of the fly. Thus, lessons learnt from studying genes in the eye have provided additional understanding of their function in other tissues as well. In general, the goal of such studies is to produce an eye phenotype that can then be used to screen for modifiers of this phenotype which are likely to be involved in the same biological process (Kurada and White, 1998; Carrera *et al.*, 1998). Thus, by performing screens for genes that dominantly modify the rough eye phenotype caused by misexpressing Hunchback, new components involved in the biological pathways mediated by Hunchback may be identified.

Chapter 7. Analyzing protein-protein interactions mediated by different ZF motifs using the B2H system.

7.1 Introduction

Biological processes ranging from gene transcription and protein synthesis to cellular signaling and protein degradation are regulated by protein-protein interactions. Identification and characterization of these various protein interactions constitutes one of the most important goals of modern biology. Both biochemical and genetic methods have been described to study protein-protein interactions. Using biochemical assays (e.g. chromatography, co-immunoprecipitations) interacting proteins can be identified in crude extracts by their ability to co-purify, while in genetic assays the formation of a protein contact results in a detectable alteration in a phenotype of a cell (reviewed in Hu *et al.*, 2000). The best characterized genetic assay is the Yeast two-hybrid (Y2H) system which was developed more then 15 years ago as a convenient method to identify and analyze protein-protein interactions (Fields and Song, 1989; reviewed in Serebriiskii *et al.*, 2001; reviewed in Fields, 2005). This system was recently translated into *E. coli* cells resulting in the development of the Bacterial two-hybrid (B2H) system (e.g. Hu, *et al.*, 2000; Joung, 2001; see also section 3.2.2). The feasibility of the B2H system for analyzing protein-protein interactions has been demonstrated with many protein pairs suggesting that, in theory, any protein pair of interest can be analyzed in this system (Dove *et al.*, 1997; Dove and Hochschild, 1998; Blum *et al.*, 2000; Joung *et al.*, 2000; Shaywitz *et al.*, 2000; Dove and Hochschild, 2001; Saito *et al.*, 2004). Furthermore, this system can detect relatively weak interactions possessing equilibrium dissociation constants in the micromolar range and the magnitude of reporter gene activity increases accordingly with the strength of a particular interaction (Dove *et al.*, 1997).

Although various interacting protein-protein pairs have been demonstrated to work in the B2H system prior to the work described in this thesis, protein-protein interactions mediated by ZF domains have not been investigated in this system. As described in Chapter 1 (section 1.2.8), in addition to the classical C2H2 ZF motif, several other classes of zinc-ligating domains have been identified (Matthews and Sunde, 2002). Some of these ZFs are well known for theit ability to mediate protein-protein interactions although relatively little is

known about the molecular details of these interaction interfaces. Since we successfully validated the B2H system both as a reporter as well as a selection system for analyzing DZF mediated dimerization (Chapter 3), we were also interested in testing if the B2H system could provide a means to characterize protein-protein interactions mediated by a variety of different other ZF proteins (Figure 7.1).

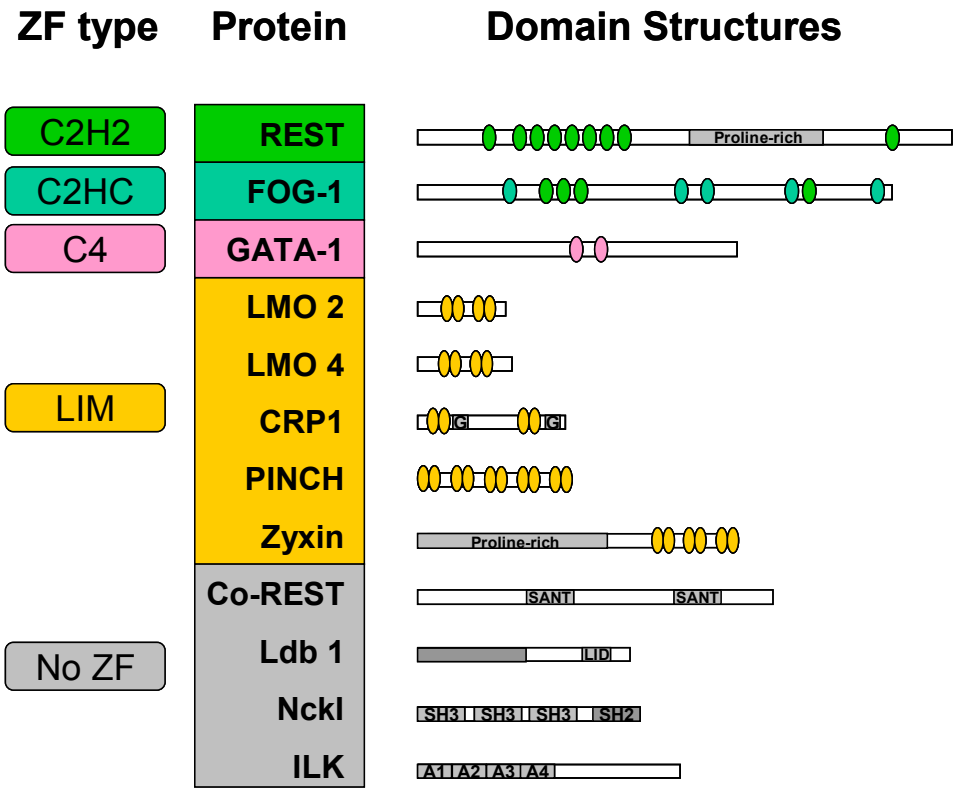

Figure 7.1 Proteins analyzed in the B2H system. The domain structures of the various proteins tested in the B2H system are shown. The colored boxes indicate type of the respective ZF motif (Note that proteins without ZFs are not further classified and are summarized as "No ZF" group). Individual ZFs are shown as colored ovals or as double ovals for the LIM domains. Additional domains or typical motifs are shown as grey boxes. REST, RE-1 silencing transcription factor; FOG, Friend of GATA; LMO, LIM only; CRP, cysteine-rich protein; PINCH, particulary interesting new cysteine and histidine-rich protein; Co-REST, Co-repressor of REST; Ldb, LIM domain binding protein, ILK, Integrin-linked kinase; G, Glycin-rich region; SANT, SW13/ADA2/NCoR/TFIIIB domain; LID, LIM interaction domain; SH2/3, Src-homology 2/3; A, Ankyrin repeat.

7.2 Interaction between the ZF proteins REST and CoREST

7.2.1 Background

The transcriptional repressor REST is a biologically important example of a C2H2 protein that mediates protein-protein interaction. REST represses expression of neuronal genes in non-neuronal cells by binding to a conserved DNA sequence (RE-1, repressor element 1). In addition to its DNA binding domain which consists of a cluster of 8 ZFs, REST harbors two additional C2H2 domains at its N- and C-termini (Figure 7.1). While the N-terminal domain associates with the mSin3/histon deacetylase 1, 2 (HDAC1, 2) complex to mediate repression, the C-terminal domain mediates gene silencing by interacting with the co-repressor CoREST (reviewed in Lunyak *et al.*, 2004; reviewed in Lunyak and Rosenfeld, 2005). CoREST is characterized by two SANT (SW13/ADA2/NCoR/TFIIIB) domains (Figure 7.1, Aasland *et al.*, 1996, Andres *et al.*, 1999) that may be involved in stabilizing the binding of REST to its DNA site. The REST interaction domain in CoREST was mapped by Y2H analysis and revealed that two domains (aa 102-195 and 145-225, including SANT motif 1) were sufficient for the interaction with REST (Ballas *et al.*, 2001). Interestingly, the interaction domain of REST requires only the single C-terminal C2H2 zinc finger and point mutations in this motif disrupted the interaction between REST and CoREST and abolished REST repressor activity (Andres *et al.*, 1999). However, little is known about the molecular details of how the C2H2 ZF of REST interacts with CoREST. We were therefore interested in testing if the B2H system could provide a means to further characterize this interaction.

7.2.2 Validation of the B2H system for studying REST/CoREST interactions

To test whether the REST/CoREST interaction might be detected in the B2H system we constructed several different plasmids encoding fusion proteins required for testing in the B2H system setup. In an initial experiment the C-terminal C2H2 zinc finger (F9, residues 1008-1097) of REST (previously shown to interact with CoREST) was fused to the DNA-binding domain of the dimeric bacteriophage lambda cI (λcI) protein. In addition, two fragments of the CoREST N-terminus (CoREST A, residues 102-195 and CoREST B, residues 130-180) were fused to the α-subunit of the *E. coli* RNA polymerase. To test the

interaction of the hybrid proteins a reporter strain harboring the binding site for the λcI protein upstream of the weak *lacZ* promoter was constructed. Combinations of plasmids expressing the different hybrid proteins were transformed into this reporter and β-galactosidase activity was assessed. Neither CoREST A nor CoREST B was able to mediate transcriptional activation of *lacZ* when co-expressed with the REST F9 fusion protein (data not shown). Using different IPTG concentrations to increase the expression level of the fusion proteins was not successful (data not shown).

Although the λcI protein has previously been used successfully to make DBD hybrid proteins (Dove *et al.*, 1997) we were concerned that it might not bind properly to its binding side or did not provide the right conformation for REST to be able to interact with CoREST. Therefore, it might have been unable to recruit the *E. coli* RNA polymerase to the promoter. Thus, we decided to construct additional plasmids encoding various DBD-fusion proteins. One construct expressed REST F9 as a direct fusion to the REST F3-8 DBD which incorporated the F9 into the context of its specific natural DBD. Two additional constructs consisted of two segments of REST F9 (F9 long, residues 1008-1097 and F9 short, residues 1049-1097) fused to the DNA binding domain of the Zif268 protein (see Chapter 3, section 3.2). Combinations of plasmids expressing these hybrid proteins together with the α-CoREST fusion proteins were introduced into appropriate reporter strains bearing either the REST F3-8 or the Zif268 binding site and β-galactosidase activity was assessed. Unfortunately none of these combinations were able to activate expression of *lacZ* (data not shown). Because efficient recruitment of the RNAP by the α-subunit depends on the right distance and orientation of the DNA binding site (DBS) relative to the transcription start point of the *lacZ* promoter a range of different reporter strains bearing the Zif268 binding site positioned at various distances and orientations of the promoter were further tested. None of these reporters demonstrated interactions of REST/CoREST as evidenced by the lack of activation of *lacZ* (data not shown).

Since the interaction between REST and CoREST has been previously demonstrated, this suggests that these proteins might not be expressed sufficiently in bacteria which would explain why their interaction could not be detected in the B2H system.

7.3 Interaction between the ZF proteins GATA-1 and FOG-1

7.3.1 Background

Interaction of the mammalian FOG-1 protein with the GATA-1 protein is a biologically important example of a protein-protein interaction mediated by zinc fingers. GATA-1 belongs to a small family of transcription factors that were originally found in erythroid cell lines (Evans and Felsenfeld 1989; Tsai et al., 1989; Trainor *et al.*, 1990; Pevny et al., 1991). FOG-1 is co-expressed with GATA-1 during embryonic and hematopoietic development and functions as an *in vivo* cofactor for GATA-1. It has been shown that the GATA-1/FOG-1 interaction is necessary for proper hematopoiesis and thrombopoiesis (reviewed in Cantor *et al.*, 2002). The GATA-1/FOG-1 interaction is mediated by two different zinc finger motifs: GATA-1 contains two zinc fingers of the Cys4 type (where zinc ligands are represented by four cysteines) that fold into a structure termed the treble clef motif which is composed of two irregular β-hairpins and an α-helix (e.g. Kowalski *et al.*, 1999; Grishin, 2001; Krishna *et al.*, 2002; Figure 7.1). While the C-terminal finger (GATA-1 CF) binds to DNA, the N-terminal finger (GATA-1 NF) mediates interaction with FOG-1 (Kowalski *et al.*, 2002).
FOG-1 contains nine putative ZFs (F1-F9) that are of two different types: four belong to the classical C2H2 finger-type (F2, F3, F4 and F8) whereas the other five are of the C2HC type (F1, F5, F6, F7 and F9) (Figure 7.1). The sequence of the C2HC ZFs conforms well to the C2H2 ZF consensus with the exception of a cysteine replacing the final histidine (Tsang *et al.*, 1997). Structural studies of C2HC fingers indicate that their overall fold is similar to the ββα fold of C2H2 fingers but with a more extended structure at the C-terminal end (Liew *et al.*, 2000). Different reports suggest that C2HC fingers do not bind to DNA but instead mediate interactions with other proteins (Matthews *et al.*, 2000).
It has previously been shown that the GATA-1 NF can interact with the C2HC fingers F1, F6, or F9 of FOG-1(Fox *et al.* 1999). By contrast, GATA-1 CF fails to bind to the FOG-1 zinc fingers. Interestingly, FOG-1 fingers that are involved in protein contact are exclusively of the C2HC type, whereas the four C2H2 FOG-1 fingers are unable to mediate protein-protein interactions with GATA-1. Using the Y2H system and alanine scanning mutagenesis, several residues important for this interaction were defined for both GATA-1 and FOG-1 (Fox *et al.*, 1999; Fox *et al.*, 1998). Although these amino acid residues were identified, additional study

and definition of the protein interface is still required to define the specific amino acid-amino acid interactions that occur at the interaction interface. Thus, we wished to determine whether this interaction can also be detected in the B2H system to further analyze and characterize the interaction surface.

7.3.2 Validation of the B2H system for studying GATA/FOG interactions

7.3.2.1 GATA-1 and FOG-1 interactions can be detected in the B2H system

To test whether GATA/FOG interactions can be detected in the B2H system a plasmid encoding a GATA-1/DBD hybrid protein consisting of the N-terminal zinc finger of GATA-1 (residues 200-254) was constructed. The DBD protein in this hybrid protein is the λcI protein. Additionally, three plasmids encoding hybrid proteins consisting of either FOG-1 zinc finger 1 (residues 241-295), 6 (residues 677-760), or 9 (residues 945-995) fused to the α-subunit of the *E. coli* RNA-polymerase were created. All of these zinc fingers have been shown to be involved in protein-protein contact between GATA-1 and FOG-1 in yeast (see section 7.3.1). Combinations of plasmids expressing these hybrid proteins were then introduced into a reporter strain bearing the λcI binding site positioned upstream of the *lacZ* gene and β-galactosidase activity was measured. None of the FOG-1 fingers was able to mediate transcriptional activation of *lacZ* when co-expressed with the GATA-1 fusion protein (data not shown). To vary the expression level of the fusion proteins, different concentrations of IPTG were used to induce the expression of the constructs in the reporter cells. It was found that IPTG concentrations above 30 μM inhibited the growth of the bacteria and the optical densities of these cells were extremely low even after 6 hr of growth. This suggested that high levels of the fusion proteins might be toxic to the bacterial strains. Experience with the B2H system has demonstrated that certain λcI-hybrid proteins can be toxic to bacterial strains (K. Joung, personal communication).

Therefore we decided to construct an additional plasmid encoding GATA-1 fused to the DNA binding domain of the Zif268 protein (see Chapter 3, section 3.2). Combinations of the α-fusion proteins together with this new plasmid were then introduced into a reporter strain containing the binding site for the Zif268 protein and β-galactosidase activity was assessed. As shown in Figure 7.2A two of the FOG-1 fingers (F1 and F6) mediated transcriptional activation of *lacZ* when co-expressed with the GATA-1 NF fusion protein. The strongest

activation (4-fold) was observed with GATA-1 NF and FOG-1 F1 (Figure 7.2A). Control experiments expressing either the Zif-hybrid or the α-hybrid protein alone did not show increased β-galactosidase expression (data not shown).

To further test if activation of *lacZ* is depended on the expression of the two fusion proteins, we performed an IPTG titration experiment where the expression of the fusion proteins was induced using increasing concentrations of IPTG. Increased *lacZ* expression only occurred when the amount of IPTG used for induction reached a certain level (>25 μM) indicating that the activation depended on expression of both fusion proteins (Figure 7.2B).

Since activation of *lacZ* depends on the distance and orientation of the DBD-binding site relative to the promoter, the interaction between GATA-1 NF and FOG-1 F1 was further analyzed using various Zif268 binding site reporter strains. As shown in Figure 7.2C, the strongest activation was observed with reporter strains bearing the Zif268 binding site at position -61 and -65 relative to the promoter of the reporter gene. Further decreasing or increasing of the distance relative to the promoter did abolish the ability of GATA-1 and FOG-1 to activate *lacZ* (Figure 7.2C and data not shown). Interestingly, the binding site at -61 was positioned such that the C-terminal part of the Zif268 fusion protein is most proximal to the promoter while the orientation of the -65 binding site brought the N-terminal part of the Zif268 fusion protein close to the promoter. Although different from each other, both orientations seemed to provide the essential geometry for the GATA-1-Zif268 fusion protein to mediate interaction with FOG-1 (Figure 7.2C and data not shown).

7.3.2.2 Mutations in GATA-1 and FOG-1 disrupted the interactions

To obtain additional evidence that the transcriptional activation detected in the B2H system reflects the physiological GATA-1/FOG-1 interaction, various GATA-1 NF and FOG-1 F1 mutations known to disrupt their interaction (Fox *et al.*, 1998; 1999) were introduced into our hybrid proteins using site-directed PCR mutagenesis (as described in section 2.1.2.2). We tested whether these mutations would affect the ability of GATA-1 NF and FOG-1 F1 to mediate transcriptional activation of the *lacZ* reporter gene. Indeed, substitution of FOG residue I262 with alanine resulted in a strong inhibition of the interaction with wild type GATA-1 whereas replacement of R265A had a weaker impact on binding (Figure 7.2D). The same differential effects were found when these mutants were tested in the Y2H system (Fox *et al.*, 1999). Similarly, substitution of the GATA-1 residues E203 with valine and V205 with threonine also disrupted the interaction just as they did in the yeast-based system, with the

mutation at residue position E203 displaying a stronger impact on interaction then the mutation at residue position 205 (Figure 7.2D, Fox *et al.,* 1998).

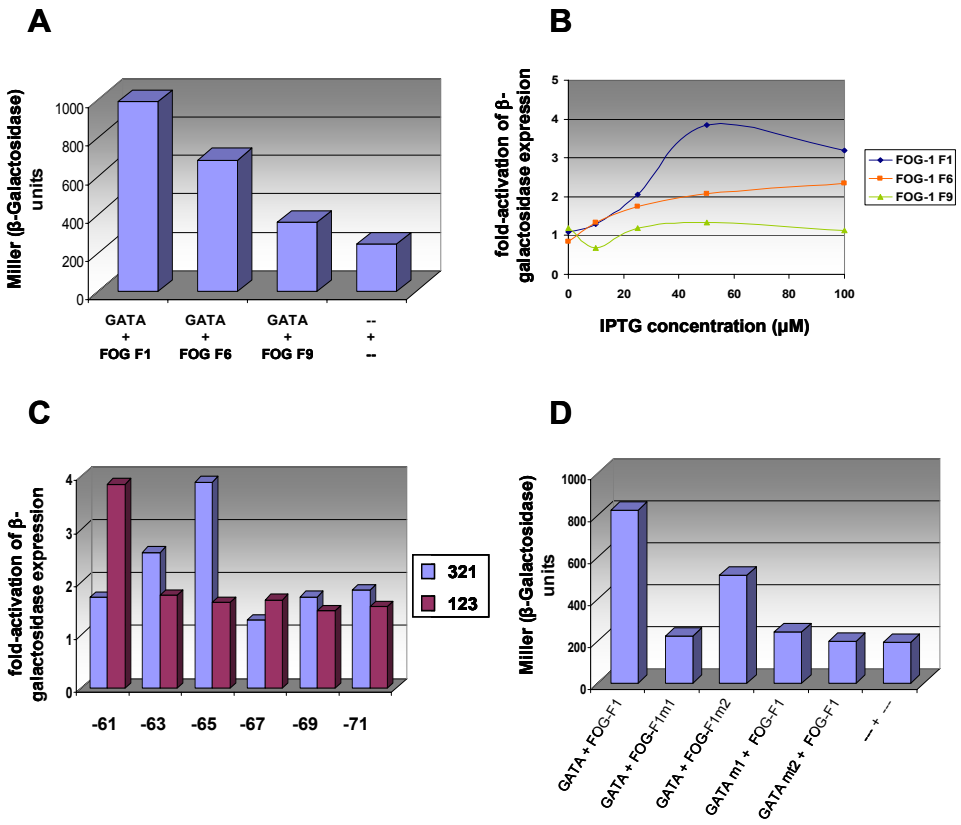

Figure 7.2 Analysis of the interaction between GATA-1 and FOG-1 in the B2H system. B2H reporter strains expressing pairwise combinations of plasmids encoding domains from GATA-1 and FOG-1 were assayed for β-galactosidase activity. In these assays GATA-1 NF was fused to the DBD of Zif268 while ZFs 1, 6 and 9 from FOG-1 were connected to the α-subunit of the RNAP. If not mentioned otherwise an IPTG concentration of 50 μM was used. **(A)** The reporter strain harboring the Zif268 DBS at position -61 in orientation 123 relative to the promoter was transformed with combinations of plasmids and β-galactosidase assays were performed. A control expressing Zif268 and the α-subunit alone is shown to the right. **(B)** β-galactosidase activities of -61 reporter cultures grown at various IPTG concentrations were measured to analyze the effect of different hybrid-protein expression levels. Fold activation was normalized to cells expressing only Zif268 and the α-subunit. **(C)** The interaction between GATA-1 NF and FOG-1 F1 was further analyzed using reporter strains that differ in their location and orientation of the Zif268 DBS relative to the promoter. Position of the binding site is indicated on the X-axis. Purple bars represent reporter strains with the Zif268 binding site in orientation 123 and blue bars indicate orientation 321. **(D)** Specific mutations in FOG-F1 (m1=I262A, m2=R265A) and in GATA-1 (m1= E203V, m2= V205T) were analyzed for their ability to abolish the interaction between GATA-1 NF and FOG-1 F1. In this assay, the reporter strain harboring the Zif268 DBS at position -65 in orientation 321 relative to the promoter was used.

Taken together, these results suggest that interactions between GATA-1 NF and fingers from FOG-1 interaction can mediate transcriptional activation in the B2H system. Importantly, these experiments further establish the feasibility of using the B2H system for studying protein-protein interactions mediated by different types of ZF proteins.

7.4 LIM domain mediated protein-protein interactions

7.4.1 Background

The LIM domain (termed after the initials of the founding members of this class, Lin-11, Isl1 and MEC-3) is defined as a cysteine-histidine rich domain consisting of two tandemly repeated zinc-coordinating fingers (for comprehensives reviews see: Dawid *et al.*, 1998; Jurata and Gill, 1998; Bach, 2000; Kadrmas and Beckerle, 2004). Because of their abundance and diversity, LIM domain proteins are further classified by sequence similarity and on the basis of the presence and nature of associated domains (Dawid *et al.*, 1998; Bach, 2000; Kadrmas and Beckerle, 2004). For example, LIM homeodomain (LHX), LIM only (LMO) and LIM kinase (LIMK) proteins all contain paired LIM domains at the N-terminus and form Group 1. Group 2 proteins are mainly composed of the LIM domains. LIM proteins of Group 3 are more heterogeneous in sequence than Group 1 and 2 and the LIM domains are localized at the C-terminus. Like GATA-1, the LIM domain ZF motif belongs to the structural treble clef motif consisting of two treble-clef fingers that are separated by two amino acids (Kadrmas and Beckerle, 2004). Although LIM proteins share several structural similarities with GATA ZFs, they do not seem to bind DNA. Rather, LIM domains have been implicated to mediate protein-protein interactions and are found in proteins with a wide variety of functions (Bach, 2000; Kadrmas and Beckerle, 2004). Many different types of proteins are described that can interact with LIM domain proteins including kinases, transcription factors and other LIM domains. In addition, a few LIM proteins have been reported to mediate homodimerization. While some LIM proteins seem to be exclusively involved in protein contacts, others have additional separate functional domains. Examples for such functional domains are the homeodomain in LHXs or the kinase domain in LIMK proteins.

Since the LIM domain proteins provide an example of ZF proteins that are thought to exclusively mediate protein-protein interactions we were interested in analyzing the LIM domain using the B2H system. We chose several well characterized LIM domains and their partners that all mediate interaction through LIM domains in at least one of the partners. These proteins partly belong to different classes of LIM domain proteins providing a representative selection of LIM domain proteins (Figure 7.1, summarized in Table 7.1).

LIM protein	Location	Interacting partner	Class / Description	Interaction	References
CRP1 (2)	Cytoplasmatic	CRP1	LIM domain	CRP1 proteins have been reported to form homodimers using either one or both LIM domains.	Feuerstein *et al.*, 1994
Zyxin (3)	Cytoplasmatic/ nuclear	CRP1	LIM domain	Interaction of Zyxin with CRP1 is mediated by LIM domain 1. Neither LIM domain of CRP1 alone is sufficient for Zyxin binding but the region in between the two domains can be deleted without abolishing its capability to bind to Zyxin.	Schmeichel and Beckerle, 1994 Schmeichel and Beckerle, 1998
LMO2/4 (1)	Nuclear	Ldb1	Cofactor for nuclear LIM proteins	Ldb1 interacts with LMO LIM proteins using its C-terminal LIM interaction domain (LID) which consists of 39 residues. Solution structures have shown that LID binds to the N-terminal LIM domain of LMO2 and LMO4 and identified important residues at the interaction surface.	Agulnick *et al.*, 1996 Jurata *et al.*, 1996 Deane *et al.*, 2003 Jurata and Gill, 1997 Breen *et al.*, 1998
PINCH (3)	Cytoplasmatic/ nuclear	ILK	Integrin-linked kinase	The PINCH-ILK interaction is important for proper localization of ILK to focal adhesions and its function in integrin signaling. PINCH-ILK interaction involves the LIM1 domain of PINCH and the ankyrin repeat present in ILK. The NMR structure of PINCH LIM1 has been solved and interaction between this domain and the ankyrin repeat of ILK has been characterized showing that it is mainly ZF2 of LIM1 which mediates this interaction.	Li *et al.*, 1999 Tu *et al.*, 1998 Tu *et al.*, 1999 Velyvis *et al.*, 2001
PINCH (3)	Cytoplasmatic/ nuclear	NckI	Adaptor protein	PINCH interacts with the third SH3 domain of Nck2 using its LIM4 domain. Nck2 is involved in growth factor signaling cascades. The NMR structure of LIM4 domain was determined and important residues that mediate the interaction with SH3 of Nck2 were mapped.	Velyvis *et al.*, 2003

Table 7.1 Protein-protein interactions mediated by LIM domains. Protein pairs are shown in column 1 and 3. The number in parentheses in the first column indicates the group to which the respective LIM protein belongs. Column 4 classifies the interacting partner. Information about the importance and details of interactions are provided in column 5. See Figure 7.1 for abbreviations.

7.4.2 Validation of the B2H system for studying LIM domain mediated interactions

To test whether LIM domain mediated interactions can be detected in the B2H system we constructed several different plasmids encoding various fusion proteins. Combinations of plasmids expressing these hybrid proteins were then introduced into the different reporter strains bearing the Zif268 binding site at various positions and orientations relative to the test promoter, and β-galactosidase activity was measured. Although these proteins have been previously described and characterized as interacting domains, most of the tested combinations were not able to mediate transcriptional activation of the *lacZ* reporter (Table 7.2 and data not shown). However, co-expression of LMO2 and LMO4 with the LID fusion protein was able to activate *lacZ* transcription. The strongest activation was observed for LMO4 LIM1 and LID with the reporter strain that bears the Zif268 binding site at position -65 (orientation 321). In this initial experiment an IPTG concentration of 50 μM was used which inhibited the growth of the bacteria suggesting that the produced level of fusion proteins was toxic to the bacterial strains.

DNA-binding fusion protein	RNAP fusion protein	Activation of reporter
CRP1	CRP1	-
CRP1	CRP1 LIM2	-
CRP1 LIM2	CRP1	-
CRP1 LIM2	CRP1 LIM2	-
CRP1	h-zyxin	-
CRP1	c-zyxin	-
h-zyxin	CRP1	-
c-zyxin	CRP1	-
PINCH LIM1	ILK	-
PINCH LIM4	Nck2	-
LMO2	LID	+
LMO2 LIM1	LID	+
LMO4	LID	+
LMO4 LIM1	LID	+

Table 7.2 Analysis of interactions mediated by LIM domains in the B2H system. Tested protein pairs are shown in the first two columns. Column 3 summarizes the result of interactions assessed in duplicates using various conditions. + indicates increased reporter gene activity while – indicates that no reporter gene activity was detected. An IPTG concentration of 50 μM was used for all assays and different reporter strains were analyzed.

Therefore, an IPTG titration experiment was performed using various amounts of IPTG below 50 μM. As shown in Figure 7.3, the highest *lacZ* expression for all four tested

interactions occurred with 10 µM IPTG and higher concentrations inhibited growth. Interestingly, reporter activation was already observed even without induction of expression suggesting that basal expression levels of the hybrid proteins are sufficient to activate *lacZ* expression.

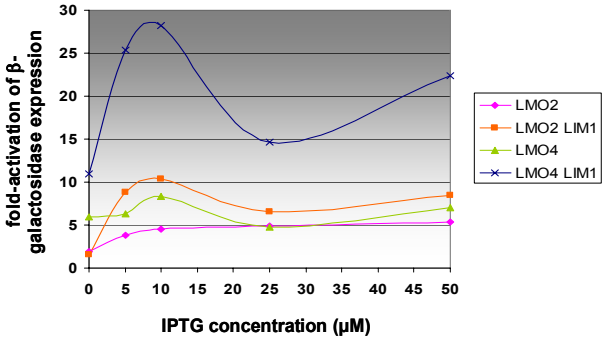

Figure 7.3 Analysis of the interaction between LMO and LID in the B2H system. The -65 (orientation 321) reporter was transformed with plasmids encoding the appropriate fragments and β-galactosidase activity was assessed over increasing IPTG concentrations. Fold activation was normalized to cells expressing Zif268 together with the α-subunit.

Control experiments expressing either the Zif-hybrid or the α-hybrid proteins alone did not show increased β-galactosidase Units (data not shown). Thus, the B2H can generally be used to study certain LIM domains although most of the tested pairs were not able to mediate transcriptional activation.

7.5 Discussion

7.5.1 Protein-protein interactions mediated by ZF proteins can be studied in the B2H system

The work in this chapter aimed to further validate the B2H system for analyzing protein-protein interactions mediated by ZF proteins. Different ZF-protein pairs know to interact with each other were introduced into the B2H system and their abilities to activate a reporter gene were analyzed. These pairs consisted of different classes of ZF proteins with distinct structural and functional features. To test for transcriptional activation of the reporter gene, one of the potentially interacting peptides was fused to the α-subunit of the RNAP while the other peptide was connected to various DBDs. Although only some of the previously described interactions could be detected in the B2H system, the results of this chapter (together with Chapter 3) provide further evidence that it is in principle possible to study

protein-protein interactions mediated by ZF proteins using the B2H system. Results obtained for particular pairs that did work in this system were generally very reliable and consistent with previously described observations obtained using different methods including the Y2H system. For example, mutations characterized in the Y2H system capable of abolishing a defined interaction displayed a similar effect when analyzed in the B2H (e.g. mutations that destroyed the interaction between GATA-1 and FOG-1). Furthermore, negative controls expressing either the DBD-hybrid or the α-hybrid protein alone did not show increased activation of *lacZ* suggesting that the activation depends on the presence of the two interacting proteins. Hence, the activation of the reporter gene is due to an interaction with similar characteristics compared to those described in previous studies. This work (together with Chapter 3) provided examples of protein-protein interactions mediated by ZF proteins analyzed in the B2H system. In addition, it confirmed that particular ZF proteins can be involved in contacting other proteins.

7.5.2 B2H versus Y2H

Unfortunately, not all of the tested pairs could activate the reporter gene, although these pairs were previously tested for their ability to interact with each other using different methods. Several reasons might explain this lack of detection in the B2H system. For example, some peptides may have been unable to fold properly or may have been insoluble or toxic to bacterial cells (as opposed to yeast cells) and were therefore not expressed at high levels. Some peptides may have required posttranslational modifications which the bacteria cells were not able to provide. Other reasons could be that they did not provide the right geometry for the existing setup which could be due to certain properties of the peptides, including size or folding conformation. On the other hand, these pairs may just be unable to interact at all in bacteria or the interactions are too weak to be detected in the B2H system.

Assuming that it is not possible to test every single interacting pair in the B2H system, why is it still worth trying it? It definitely provides some advantages compared to the well described Y2H system, mainly the fast growth and high transformation efficiency obtained in bacteria. Furthermore, some proteins can not be analyzed in yeast because they may require nuclear localization or activate transcription of the reporter in an unspecific manner (reviewed in Hu *et al.*, 2000; reviewed in Hu, 2001). In addition, certain cellular processes can not be studied in yeast because of the capability of endogenous yeast proteins to influence these events. For

example, it has proven difficult to study the effect of phosphorylation on protein-protein interactions in yeast due to present endogenous kinases. Bacteria on the other hand provide an isolated system where these effects could easily be studied (Shaywitz *et al.*, 2002; Shaywitz *et al.*, 2000). Thus, each system has advantages and disadvantages and can not be considered as better or worse but rather will provide complementary information (reviewed in Hu, 2001; Serebriiskii *et al.*, 2005).

Summary

The C2H2 ZF motif is a compact ~ 30 amino acid molecular recognition domain that comprises a β-hairpin followed by an α-helix (ββα fold). In proteins, these domains are typically found as tandem arrays that mediate specific interactions with various macromolecules including DNA, RNA and other proteins. Although very well characterized as a DNA-binding domain, relatively little is currently understood about the molecular details of protein-protein interactions mediated by C2H2 ZFs. The Ikaros and Hunchback transcription factor family provides an ideal model system for studying ZF mediated protein-protein interactions. Ikaros, the founding member of this family is defined as a classical C2H2 ZF protein composed of a cluster of four C2H2 ZFs at the N-terminus and two additional C2H2 ZFs at the C-terminus. While the N-terminal ZFs are involved in specific DNA recognition, the C-terminal domain (termed as Dimerization Zinc Finger or DZF domain) has been shown to mediate the homo- and hetero-typic interactions.

In this thesis, the DZF domains found in the Ikaros and Hunchback transcription factor family have been examined using a combination of genetic, biochemical and functional assays. We first established a bacterial-based genetic system for studying C2H2 ZF mediated protein-protein interactions. This system is more rapid then previously described methods and allows the performance of complex genetic selections.

To test, if protein-interacting C2H2 ZFs, in analogy to DNA-binding ZFs, can be used to create novel protein-protein interaction specificities, we constructed libraries of synthetic DZFs by shuffling individual C2H2 ZFs from DZF domains found in the human Ikaros and other related transcription factors. Using a bacterial-based selection system, we identified synthetic heterodimeric DZFs that can mediate activation of a single copy reporter gene in bacterial cells. These synthetic protein-protein interaction domains can also be used to reconstitute a synthetic bi-partite activator in the nucleus of a human cell which results in a transcriptional activation of the endogenous VEGF-A gene. In addition, these synthetic two-finger domains can be linked together to create more extended protein-protein interaction interfaces. These results demonstrate that certain protein-interacting C2H2 ZFs (like their DNA binding counterparts) can function in a modular fashion. Furthermore, analysis of the interaction specificities of these synthetic domains led to the discovery of a novel anti-parallel interaction mode for the DZF domain.

The homo-typic interaction mediated by different DZF domains was examined in greater detail using mutational analysis. These studies narrowed down residues that are likely to be important for dimerization mediated by the Hunchback DZF domain. Comparing these amino acids to residue positions previously identified as important for dimerization of Ikaros and Eos highlights both similarities and differences. To obtain further information about the physical and chemical interaction surface we attempted to purify active peptides consisting of different DZF domains for X-ray crystallography. Although highly purified DZF peptides were successfully obtained, various attempts to refold these peptides into active domains resulted in the formation of aggregates consisting of the various DZFs.

Based on findings in the bacterial and cell culture systems, we started exploring if Hunchback dimerizes in *Drosophila melanogaster* using its DZF domain and if dimerization is essential for the function of the protein. Therefore, constructs encoding the full-length Hunchback protein harboring various natural and modified DZF domains were generated and expressed in transgenic flies. These transgenics were used to perform functional *in vivo* studies of the Hunchback DZF domain in Neuroblast specification during *Drosophila melanogaster* development. We confirmed previous studies that the C-terminal domain in Hunchback is important for maintaining the function of Hunchback in specifying early-born temporal identity in *Drosophila* neural stem cell lineages. Importantly, our results indicate that this domain can be functionally replaced with a heterologous (i.e.: non fly) DZF domain, suggesting that the importance of the DZF domain is due to its ability to mediate dimerization.

References

Agulnick A.D., Taira M., Breen J.J., Tanaka T., Dawid I.B. and Westphal H. (1996). Interactions of the LIM-domain-binding factor Ldb1 with LIM homeodomain proteins. *Nature.* **384**: 270-272.

Ahn J.H., Lee Y.P. and Rhee J.S. (1997). Investigation of refolding condition for Pseudomonas fluorescens lipase by response surface methodology. *J Biotechnol.* **54**: 151-160.

Altamirano M.M., Golbik R., Zahn R., Buckle A.M. and Fersht A.R. (1997). Refolding chromatography with immobilized mini-chaperones. *Proc Natl Acad Sci U S A.* **94**: 3576-3578.

Andres M.E., Burger C., Peral-Rubio M.J., Battaglioli E., Anderson M.E., Grimes J., Dallman J., Ballas N. and Mandel G. (1999). CoREST: a functional corepressor required for regulation of neural-specific gene expression. *Proc Natl Acad Sci U S A.* **96**: 9873-9878.

Ashkenazi A., Presta L.G., Marsters S.A., Camerato T.R., Rosenthal K.A., Fendly B.M. and Capon D.J. (1990). *Proc Natl Acad Sci U S A.* **87**: 7150-7154.

Ausubel F.M., Brent R., Kingston R.E., Moore D.D., Seidman J.G., Smith J.A., Struhl K. (1996). Current Protocols in Mol. Biol. *New York: John Wiley & Sons*.

Bach I. (2000). The LIM domain: regulation by association. *Mech Dev.* **91**: 5-17.

Bae K.H., Kwon Y.D., Shin H.C., Hwang M.S., Ryu E.H., Park K.S., Yang H.Y., Lee D.K., Lee Y., Park J., Kwon H.S., Kim H.W., Yeh B.I., Lee H.W., Sohn S.H., Yoon J., Seol W. and Kim J.S. (2003). Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat Biotechnol*. **21**: 275-280.

Ballas N., Battaglioli E., Atouf F., Andres M.E., Chenoweth J., Anderson M.E., Burger C., Moniwa M., Davie J.R., Bowers W.J., Federoff H.J., Rose D.W., Rosenfeld M.G., Brehm P. and Mandel G. (2001). Regulation of neuronal traits by a novel transcriptional complex. *Neuron.* **31**: 353-365.

Bardeesy N. and Pelletier J. (1998). Overlapping RNA and DNA binding domains of the wt1 tumor suppressor gene product. *Nucleic Acids Res.* **26**: 1784-1792.

Bender M., Horikami S., Cribbs D. and Kaufman T.C. (1988). Identification and expression of the gap segmentation gene hunchback in *Drosophila melanogaster*. *Dev Genet.* **9**: 715-732.

Blancafort P., Segal D.J. and Barbas C.F. 3rd. (2004). Designing transcription factor architectures for drug discovery. *Mol Pharmacol.* **66**: 1361-1371.

Blancafort P., Chen E.I., Gonzalez B., Bergquist S., Zijlstra A., Guthy D., Brachat A., Brakenhoff R.H., Quigley J.P., Erdmann D. and Barbas C.F. 3rd. (2005). Genetic reprogramming of tumor cells by zinc finger transcription factors. *Proc Natl Acad Sci U S A.* **102**: 11716-11721.

Blum J.H., Dove S.L., Hochschild A. and Mekalanos J.J. (2000). Isolation of peptide aptamers that inhibit intracellular processes. *Proc Natl Acad Sci U S A.* **97**: 2241-2246.

Bossing T., Udolph G., Doe C.Q. and Technau G.M. (1996). The embryonic central nervous system lineages of *Drosophila melanogaster*. I. Neuroblast lineages derived from the ventral half of the neuroectoderm. *Dev Biol.* **179**: 41-64.

Brand A.H. and Perrimon N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development.* **118**: 401-415.

Breen J.J., Agulnick A.D., Westphal H. and Dawid I.B. (1998). Interactions between LIM domains and the LIM domain-binding protein Ldb1. *J Biol Chem.* **273**: 4712-4717.

Brinkmann U., Buchner J. and Pastan I. (1992). Independent domain folding of Pseudomonas exotoxin and single-chain immunotoxins: influence of interdomain connections. *Proc Natl Acad Sci U S A.* **89**: 3075-3079.

Brivanlou A.H. and Darnell J.E. Jr. (2002). Signal transduction and the control of gene expression. *Science.* **295**: 813-818.

Brody T. and Odenwald W.F. (2000). Programmed transformations in neuroblast gene expression during *Drosophila* CNS lineage development. *Dev Biol.* **226**: 34-44.

Brody T. and Odenwald W.F. (2002). Cellular diversity in the developing nervous system: a temporal view from *Drosophila*. *Development.* **129**: 3763-3770.

Brown K.E., Guest S.S., Smale S.T., Hahm K., Merkenschlager M. and Fisher A.G. (1997). Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell.* **91**: 845-854.

Brown R.S. (2005). Zinc finger proteins: getting a grip on RNA. (2005). *Curr Opin Struct Biol.* **15**: 94-98.

Buchanan S.G. and Gay N.J. (1996). Structural and functional diversity in the leucine-rich repeat family of proteins. *Prog. Biophys. Mol. Biol.* **65**: 1-44.

Buchner J. and Rudolph R. (1991). Renaturation, purification and characterization of recombinant Fab-fragments produced in *Escherichia coli*. *Biotechnology (N Y).* **9**: 157-162.

Campos Ortega J.A. and Hartenstein V. (1985). The embryonic development of *Drosophila melanogaster*. *Berlin: Springer Verlag*

Cantor A.B. and Orkin S.H. (2001). Hematopoietic development: a balancing act. *Curr Opin Genet Dev*. **11**: 513-519.

Cantor A.B. and Orkin S.H. (2002). Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene* **21**, 3368-3376.

Caricasole A., Duarte A., Larsson S.H., Hastie N.D., Little M., Holmes G., Todorov I. and Ward A. (1996). RNA binding by the Wilms tumor suppressor zinc finger proteins. *Proc Natl Acad Sci U S A*. **93**: 7562-7566.

Carrera P., Abrell S., Kerber B., Walldorf U., Preiss A., Hoch M. and Jackle H. (1998). A modifier screen in the eye reveals control genes for Kruppel activity in the *Drosophila* embryo. *Proc Natl Acad Sci U S A*. **95**: 10779-10784.

Cassel T.N., Berg T., Suske G. and Nord M. (2002). Synergistic transactivation of the differentiation-dependent lung gene Clara cell secretory protein (secretoglobin 1a1) by the basic region leucine zipper factor CCAAT/enhancer-binding protein alpha and the homeodomain factor Nkx2.1/thyroid transcription factor-1. *J Biol Chem.* **277**: 36970-36977.

Choo Y., Sanchez-Garcia I. and Klug A. (1994). In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature.* **372**: 642-645.

Choo Y. and Klug A. (1994). Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proc Natl Acad Sci U S A.* **91**: 11168-11172.

Christopherson I., Piechoki M., Liu G., Ratner S. and Galy A. (2001). Regulation of L-selectin expression by a dominant negative Ikaros protein. *J Leukoc Biol*. **69**: 675-683.

Clark E.D.B. (1998). Refolding of recombinant proteins. *Curr Opin Biotechnol.* **9**: 157-163.

Cleary M.D. and Doe C.Q. (2006). Regulation of neuroblast competence: multiple temporal identity factors specify distinct neuronal fates within a single early competence window. *Genes Dev.* **20**: 429-434.

Cobb B.S., Morales-Alcelay S., Kleiger G., Brown K.E., Fisher A.G., Smale S.T. (2000). Targeting of Ikaros to pericentromeric heterochromatin by direct DNA binding. *Genes Dev.* **14**: 2146-2160.

Cobb B.S. and Smale S.T. (2005). Ikaros-family proteins: in search of molecular functions during lymphocyte development. *Curr Top Microbiol Immunol.* **290**: 29-47.

Cunningham B.C. and Wells J.A. (1998). High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science.* **244**: 1081-1085.

Cupit P.M., Hansen J.D., McCarty A.S., White G., Chioda M., Spada F., Smale S.T. and Cunningham C. (2003). Ikaros family members from the agnathan Myxine glutinosa and the urochordate Oikopleura dioica: emergence of an essential transcription factor for adaptive immunity. *J Immunol.* **171**: 6006-6013.

Dawid I.B., Breen J.J. and Toyama R. (1998). LIM domains: multiple roles as adapters and functional modifiers in protein interactions. *Trends Genet.* **14**: 156-162.

Deane J.E., Mackay J.P., Kwan A.H., Sum E.Y., Visvader J.E. and Matthews J.M. (2003). Structural basis for the recognition of ldb1 by the N-terminal LIM domains of LMO2 and LMO4. *EMBO J*. **22**: 2224-2233.

Desjarlais J.R. and Berg J.M. (1992). Toward rules relating zinc finger protein sequences and DNA binding site preferences. *Proc Natl Acad Sci U S A*. **89**: 7345-7349.

Desjarlais J.R. and Berg J.M. (1993). Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. *Proc Natl Acad Sci U S A*. **90**: 2256-2260.

Deyev S.M., Waibel R., Lebedenko E.N., Schubiger A.P. and Pluckthun A. (2003). Design of multivalent complexes using the barnase*barstar module. *Nat Biotechnol.* **21**: 1486-1492.

Doe C.Q. and Goodman CS. (1985). Early events in insect neurogenesis. I. Development and segmental differences in the pattern of neuronal precursor cells. *Dev Biol.* **111**: 193-205.

Doe C.Q. and Goodman C.S. (1985). Early events in insect neurogenesis. II. The role of cell interactions and cell lineage in the determination of neuronal precursor cells. *Dev Biol.* **111**: 206-219.

Dove S.L. and Hochschild A. (1998). Conversion of the omega subunit of *Escherichia coli* RNA polymerase into a transcriptional activator or an activation target. *Genes Dev.* **12**: 745-754.

Dove S.L. and Hochschild A. (2001). Bacterial two-hybrid analysis of interactions between region 4 of the sigma(70) subunit of RNA polymerase and the transcriptional regulators Rsd from *Escherichia coli* and AlgQ from Pseudomonas aeruginosa. *J Bacteriol.* **183**: 6413-6421.

Dove S.L. and Hochschild A. (2004). A bacterial two-hybrid system based on transcription activation. *Methods Mol Biol.* **261**: 231-246.

Dove S.L., Joung J.K. and Hochschild A. (1997). Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature.* **386**: 627-630.

Elrod-Erickson M., Benson T.E. and Pabo C.O. (1998). High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure.* **6**: 451-464.

Elrod-Erickson M., Rould M.A., Nekludova L. and Pabo C.O. (1996). Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions. *Structure.* **4**: 1171-1180.

Endy D. (2005). Foundations for engineering biology. *Nature.* **438:** 449-453.

Ernst P., Hahm K. and Smale S.T. (1993). Both LyF-1 and an Ets protein interact with a critical promoter element in the murine terminal transferase gene. *Mol Cell Biol*. **13**: 2982-2992.

Evans T. and Felsenfeld G. (1989). The erythroid-specific transcription factor Eryf1: a new finger protein. *Cell*. **58**: 877-885.

Falke D. and Juliano R.L. (2003). Selective gene regulation with designed transcription factors: implications for therapy. *Curr Opin Mol Ther*. **5**: 161-166.

Fay D.S., Stanley H.M., Han M. and Wood W.B. (1999). A Caenorhabditis elegans homologue of hunchback is required for late stages of development but not early embryonic patterning. *Dev Biol.* **205**: 240-253.

Ferrell J.E. Jr. (2000). What do scaffold proteins really do? *Sci STKE.* **52**: PE1.

Feuerstein R., Wang X., Song D., Cooke N.E. and Liebhaber S.A. (1994). The LIM/double zinc-finger motif functions as a protein dimerization domain. *Proc Natl Acad Sci U S A*. **91**: 10655-10659.

Fields S. and Song O. (1989). A novel genetic system to detect protein-protein interactions. *Nature.* **340**: 245-346.

Fields S. (2005). High-throughput two-hybrid analysis. The promise and the peril. *FEBS J.* **272**: 5391-5399.

Finerty P.J. Jr. and Bass B.L. (1999). Subsets of the zinc finger motifs in dsRBP-ZFa can bind double-stranded RNA. *Biochemistry*. **38**: 4001-4007.

Foster M.P., Wuttke D.S., Radhakrishnan I., Case D.A., Gottesfeld J.M. and Wright P.E. (1997). Domain packing and dynamics in the DNA complex of the N-terminal zinc fingers of TFIIIA. *Nat Struct Biol*. **4**: 605-608.

Fox A.H., Kowalski K., King G.F., Mackay J.P. and Crossley M. (1998). Key residues characteristic of GATA N-fingers are recognized by FOG. *J Biol Chem.* **273**: 33595-33603.

Fox A.H., Liew C., Holmes M., Kowalski K., Mackay J. and Crossley M. (1999). Transcriptional cofactors of the FOG family interact with GATA proteins by means of multiple zinc fingers. *EMBO J*. **18**: 2812-2822.

Frankel A.D., Berg J.M. and Pabo C.O. (1987). Metal-dependent folding of a single zinc finger from transcription factor IIIA. *Proc Natl Acad Sci U S A*. **84**: 4841-4845.

Freeman M. (1996). Reiterative use of the EGF receptor triggers differentiation of all cell types in the *Drosophila* eye. *Cell.* **87**: 651-660.

Furley A.J., Mizutani S., Weilbaecher K., Dhaliwal H.S., Ford A.M., Chan L.C., Molgaard H.V., Toyonaga B., Mak T., van den Elsen P., *et al.* (1986). Developmentally regulated rearrangement and expression of genes encoding the T cell receptor-T3 complex. *Cell*. **46**: 75-87.

Georgopoulos K., Bigby M., Wang J.H., Molnar A., Wu P., Winandy S. and Sharpe A. (1994). The Ikaros gene is required for the development of all lymphoid lineages. *Cell.* **79**: 143-156.

Georgopoulos K. (2002). Haematopoietic cell-fate decisions, chromatin regulation and ikaros. *Nat Rev Immunol.* **2**: 162-174.

Georgopoulos K., Moore D.D. and Derfler B. (1992). Ikaros, an early lymphoid-specific transcription factor and a putative mediator for T cell commitment. *Science.* **258**: 808-812.

Georgopoulos K., Winandy S. and Avitahl N. (1997). The role of the Ikaros gene in lymphocyte development and homeostasis. *Annu Rev Immunol.* **15**: 155-176.

Giesecke A.V., Fang R. and Joung J.K. (2006). Synthetic protein–protein interaction domains created by shuffling Cys_2His_2 zinc-fingers. *Molecular Systems Biology* 2 doi:10.1038/msb4100053

Giesecke A.V. and Joung J.K. (2005) A bacterial two-hybrid system for studying and Modifying protein-protein interactions. In: Protein-Protein Interactions: A Molecular Cloning Manual, 2nd ed., E.A. Golemis, editor. *Cold Spring Harbor Laboratory Press*: 195-216.

Giot L., Bader J.S., Brouwer C., Chaudhuri A., Kuang B., Li Y., Hao Y.L., Ooi C.E., Godwin B., Vitols E., Vijayadamodar G., Pochart P., Machineni H., Welsh M., Kong Y., Zerhusen B., Malcolm R., Varrone Z., Collis A., Minto M., Burgess S., McDaniel L., Stimpson E., Spriggs F., Williams J., Neurath K., Ioime N., Agee M., Voss E., Furtak K., Renzulli R., Aanensen N., Carrolla S., Bickelhaupt E., Lazovatsky Y., DaSilva A., Zhong J., Stanyon C.A., Finley R.L. Jr., White K.P., Braverman M., Jarvie T., Gold S., Leach M., Knight J., Shimkets R.A., McKenna M.P., Chant J. and Rothberg J.M. (2003). A protein interaction map of *Drosophila melanogaster*. *Science.* **302**: 1727-1736.

Glaser F., Steinberg D.M., Vakser I.A. and Ben-Tal N. (2001). Residue frequencies and pairing preferences at protein-protein interfaces. *Proteins.* **43**: 89-102.

Greenspan R.J. (1997). Fly pushing: The theory and practice of *Drosophila* genetics. Cold Spring Harbor Laboratory Press. New York.

Greisman H.A. and Pabo C.O. (1997). A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science.* **275**: 657-661.

Grishin N.V. (2001). Treble clef finger--a functionally diverse zinc-binding structural motif. *Nucleic Acids Res.* **29**: 1703-1714.

Grondin B., Cote F., Bazinet M., Vincent M. and Aubry M. (1997). Direct interaction of the KRAB/Cys2-His2 zinc finger protein ZNF74 with a hyperphosphorylated form of the RNA polymerase II largest subunit. *J Biol Chem.* **272**: 27877-27885.

Grosskortenhaus R., Pearson B.J., Marusich A. and Doe C.Q. (2005). Regulation of temporal identity transitions in *Drosophila* neuroblasts. *Dev Cell.* **8**: 193-202.

Green A. and Sarkar B. (1998). Alteration of zif268 zinc-finger motifs gives rise to non-native zinc-co-ordination sites but preserves wild-type DNA recognition. *Biochem J*. **333**: 85-90.

Hahm K., Ernst P., Lo K., Kim G.S., Turck C. and Smale S.T. (1994). The lymphoid transcription factor LyF-1 is encoded by specific, alternatively spliced mRNAs derived from the Ikaros gene. *Mol Cell Biol.* **14**: 7111-7123.

Hahm K., Cobb B.S., McCarty A.S., Brown K.E., Klug C.A., Lee R., Akashi K., Weissman I.L., Fisher A.G. and Smale S.T. (1998). Helios, a T cell-restricted Ikaros family member that quantitatively associates with Ikaros at centromeric heterochromatin. *Genes Dev.* **12**: 782-796.

Hai T. and Curran T. (1991). Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity. *Proc Natl Acad Sci U S A.* **88**: 3720-3724.

Haire R.N., Miracle A.L., Rast J.P. and Litman G.W. (2000). Members of the Ikaros gene family are present in early representative vertebrates. *J Immunol.* **165**: 306-312.

Hall T.M. (2005). Multiple modes of RNA recognition by zinc finger proteins. (2005). *Curr Opin Struct Biol.* **15**: 367-373.

Hansen P.K., Christensen J.H., Nyborg J., Lillelund O. and Thogersen H.C. (1993). Dissection of the DNA-binding domain of Xenopus laevis TFIIIA. Quantitative DNase I footprinting analysis of specific complexes between a 5 S RNA gene fragment and N-terminal fragments of TFIIIA containing three, four or five zinc-finger domains. *J Mol Biol.* **233**: 191-202.

Harris K., Lamson R.E., Nelson B., Hughes T.R., Marton M.J., Roberts C.J., Boone C. and Pryciak P.M. (2001). Role of scaffolds in MAP kinase pathway specificity revealed by custom design of pathway-dedicated signaling proteins. *Curr Biol.* **11**: 1815-1824.

Hay B.A., Maile R. and Rubin G.M. (1997). P element insertion-dependent gene activation in the *Drosophila* eye. *Proc Natl Acad Sci U S A.* **94**: 5195-5200.

Haynes B.F., Denning S.M., Singer K.H. and Kurtzberg J. (1989). Ontogeny of T-cell precursors: a model for the initial stages of human T-cell development. *Immunol Today.* **10**: 87-91.

Heldin C.H., Miyazono K. and ten Dijke P. TGF-beta signalling from cell membrane to nucleus through SMAD proteins. *Nature.* **390**: 465-471.

Honma Y., Kiyosawa H., Mori T., Oguri A., Nikaido T., Kanazawa K., Tojo M., Takeda J., Tanno Y., Yokoya S., Kawabata I., Ikeda H. and Wanaka A. (1999). Eos: a novel member of the Ikaros gene family expressed predominantly in the developing nervous system. *FEBS Lett.* **447**: 76-80.

Houbaviy H.B., Usheva A., Shenk T. and Burley S.K. (1996). Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proc Natl Acad Sci U S A.* **93**: 13577-13582.

Hu J.C., O'Shea E.K., Kim P.S. and Sauer R.T. (1990). Sequence requirements for coiled-coils: analysis with lambda repressor-GCN4 leucine zipper fusions. *Science.* **250**: 1400-1303.

Hu J.C., Kornacker M.G. and Hochschild A. (2000). *Escherichia coli* one- and two-hybrid systems for the analysis and identification of protein-protein interactions. *Methods.* **20**: 80-94.

Huntley S., Baggott D.M., Hamilton A.T., Tran-Gyamfi M., Yang S., Kim J., Gordon L., Branscomb E. and Stubbs L. (2006). A comprehensive catalog of human KRAB-associated zinc finger genes: Insights into the evolutionary history of a large family of transcriptional repressors. *Genome Res*. **16**: 669-677.

Hurt J.A., Thibodeau S.A., Hirsh A.S., Pabo C.O. and Joung J.K. (2003). Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc Natl Acad Sci U S A.* **100**: 12271-12276.

Isalan M., Choo Y. and Klug A. (1997). Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc Natl Acad Sci U S A.* **94**: 5617-5621.

Isalan M., Klug A. and Choo Y. (1998). Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. *Biochemistry.* **37**: 12026-12033.

Isshiki T., Pearson B., Holbrook S. and Doe C.Q. (2001). *Drosophila* neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny. *Cell.* **106**: 511-521.

Iuchi S. (2001). Three classes of C2H2 zinc finger proteins. *Cell Mol Life Sci.* **58**: 625-635.

Jackson P., Mastrangelo I., Reed M., Tegtmeyer P., Yardley G. and Barrett J. (1998). Synergistic transcriptional activation of the MCK promoter by p53: tetramers link separated DNA response elements by DNA looping. *Oncogene.* **16**: 283-292.

Jacobs G.H. (1992). Determination of the base recognition positions of zinc fingers from sequence analysis. *EMBO J.* **11**: 4507-4517.

Jamieson A.C., Kim S.H. and Wells J.A. (1994). In vitro selection of zinc fingers with altered DNA-binding specificity. *Biochemistry.* **33**: 5689-5695.

Jamieson A.C., Miller J.C. and Pabo C.O. (2003). Drug discovery with engineered zinc-finger proteins. *Nat Rev Drug Discov.* **2**: 361-368.

Jamieson A.C., Wang H. and Kim S.H. (1996). A zinc finger directory for high-affinity DNA recognition. *Proc Natl Acad Sci U S A.* **93**: 12834-12839.

Jantz D., Amann B.T., Gatto G.J. Jr. and Berg J.M. (2004). The design of functional DNA-binding proteins based on zinc finger domains. *Chem Rev.* **104**: 789-799.

Jones S. and Thornton J.M. (1996). Principles of protein-protein interactions. *Proc Natl Acad Sci U S A.* **93**: 13-20.

Joung J.K., Ramm E.I. and Pabo C.O. (2000). A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc Natl Acad Sci U S A.* **97**: 7382-7387.

Joung J.K. (2001). Identifying and modifying protein-DNA and protein-protein interactions using a bacterial two-hybrid selection system. *J Cell Biochem Suppl.* **37**: 53-57.

Jurata L.W. and Gill G.N. (1997). Functional analysis of the nuclear LIM domain interactor NLI. *Mol Cell Biol.* **17**: 5688-5698.

Jurata L.W. and Gill G.N. (1998). Structure and function of LIM domains. *Curr Top Microbiol Immunol.* **228**: 75-113.

Jurata L.W., Kenny D.A. and Gill G.N. (1996). Nuclear LIM interactor, a rhombotin and LIM homeodomain interacting protein, is expressed early in neuronal development. *Proc Natl Acad Sci U S A.* **93**: 11693-11698.

Kadrmas J.L. and Beckerle M.C. (2004). The LIM domain: from the cytoskeleton to the nucleus. *Nat Rev Mol Cell Biol.* **5**: 920-931.

Kambadur R., Koizumi K., Stivers C., Nagle J., Poole S.J. and Odenwald W.F. (1998). Regulation of POU genes by castor and hunchback establishes layered compartments in the *Drosophila* CNS. *Genes Dev.* **12**: 246-260.

Kamiuchi T., Abe E., Imanishi M., Kaji T., Nagaoka M. and Sugiura Y. (1998). Artificial nine zinc-finger peptide with 30 base pair binding sites. *Biochemistry*. **37**: 13827-13834.

Kelley C.M., Ikeda T., Koipally J., Avitahl N., Wu L., Georgopoulos K., Morgan B.A. (1998). Helios, a novel dimerization partner of Ikaros expressed in the earliest hematopoietic progenitors. *Curr Biol.* **8**: 508-515.

Kim C.A. and Berg J.M. (1996). A 2.2 A resolution crystal structure of a designed zinc finger protein bound to DNA. *Nat Struct Biol.* **3**: 940-945.

Kim J., Sif S., Jones B., Jackson A., Koipally J., Heller E., Winandy S., Viel A., Sawyer A., Ikeda T., Kingston R. and Georgopoulos K. (1999). Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. *Immunity.* **10**: 345-355.

Kim J.S. and Pabo C.O. (1998). Getting a handhold on DNA: design of poly-zinc finger proteins with femtomolar dissociation constants. *Proc Natl Acad Sci U S A*. **95**: 2812-2817.

Klug C.A., Morrison S.J., Masek M., Hahm K., Smale S.T. and Weissman I.L. (1998). Hematopoietic stem cells and lymphoid progenitors express different Ikaros isoforms, and Ikaros is localized to heterochromatin in immature lymphocytes. *Proc Natl Acad Sci U S A.* **95**: 657-662.

Klug A. and Schwabe J.W. (1995). Protein motifs 5. Zinc fingers. *FASEB J.* **9**: 597-604.

Klug A. (1999). Zinc finger peptides for the regulation of gene expression. *J Mol Biol.* **293**: 215-218.

Kobe B. and Kajava A.V. (2001). The leucine-rich repeat as a protein recognition motif. *Curr. Opin. Struct. Biol.* **11**: 725-732.

Koipally J. and Georgopoulos K. (2000). Ikaros interactions with CtBP reveal a repression mechanism that is independent of histone deacetylase activity. *J Biol Chem*. **275**: 19594-19602.

Koipally J. and Georgopoulos K. (2002). Ikaros-CtIP interactions do not require C-terminal binding protein and participate in a deacetylase-independent mode of repression. *J Biol Chem*. **277**: 23143-23149.

Koipally J., Heller E.J., Seavitt J.R. and Georgopoulos K. (2002). Unconventional potentiation of gene expression by Ikaros. *J Biol Chem.* **277**: 13007-13015.

Koipally J., Renold A., Kim J. and Georgopoulos K. (1999). Repression by Ikaros and Aiolos is mediated through histone deacetylase complexes. *EMBO J*. **18**: 3090-3100.

Koonin E.V., Wolf Y.I. and Karev G.P. (2002). The structure of the protein universe and genome evolution. *Nature.* **420**: 218-223.

Kouzarides T. and Ziff E. (1988). The role of the leucine zipper in the fos-jun interaction. *Nature.* **336**: 646-651.

Kouzarides T. and Ziff E. (1989). Leucine zippers of fos, jun and GCN4 dictate dimerization specificity and thereby control DNA binding. *Nature.* **340**: 568-571.

Kowalski K., Czolij R., King G.F., Crossley M. and Mackay J.P. (1999). The solution structure of the N-terminal zinc finger of GATA-1 reveals a specific binding face for the transcriptional co-factor FOG. *J Biomol NMR.* **13**: 249-262.

Kowalski K., Liew C.K., Matthews J.M., Gell D.A., Crossley M. and Mackay J.P. (2002). Characterization of the conserved interaction between GATA and FOG family proteins. *J Biol Chem.* **277**: 35720-35729.

Kramer B.P., Fischer C. and Fussenegger M. (2004). BioLogic gates enable logical transcription control in mammalian cells. *Biotechnol Bioeng.* **87**: 478-484.

Krishna S.S., Majumdar I. and Grishin N.V. (2003). Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res.* **31**: 532-550.

Krizek B.A., Zawadzke L.E. and Berg J.M. (1993). Independence of metal binding between tandem Cys2His2 zinc finger domains. *Protein Sci.* **2**: 1313-1319.

Kurada P. and White K. (1998). Ras promotes cell survival in *Drosophila* by downregulating hid expression. *Cell.* **95**: 319-329.

Ladant D. and Karimova G. (2000). Genetic systems for analyzing protein-protein interactions in bacteria. *Res Microbiol.* **151**: 711-720.

Lee J.S., Galvin K.M. and Shi Y. (1993). Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. *Proc Natl Acad Sci U S A.* **90**: 6145-6149.

Lee M.S., Gippert G.P., Soman K.V., Case D.A. and Wright P.E. (1989). Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science.* **245**: 635-637.

Lee S.B. and Haber D.A. (2001). Wilms tumor and the WT1 gene. *Exp Cell Res.* **264**: 74-99.

Lee D.K., Seol W. and Kim J.S. (2003). Custom DNA-binding proteins and artificial transcription factors. *Curr Top Med Chem.* **3**: 645-657.

Lehmann R. and Nusslein-Volhard C. (1997). hunchback, a gene required for segmentation of an anterior and posterior region of the *Drosophila* embryo. *Dev Biol.* **119**: 402-417.

Lemon B. and Tjian R. (2000). Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev.* **14**: 2551-2569.

Liew C.K., Kowalski K., Fox A.H., Newton A., Sharpe B.K., Crossley M. and Mackay J.P. (2000). Solution structures of two CCHC zinc fingers from the FOG family protein U-shaped that mediate protein-protein interactions. *Structure.* **8**: 1157-1166.

Li F., Zhang Y. and Wu C. (1999). Integrin-linked kinase is localized to cell-matrix focal adhesions but not cell-cell adhesion sites and the focal adhesion localization of integrin-linked kinase is regulated by the PINCH-binding ANK repeats. *J Cell Sci.* **112**: 4589-4599.

Liippo J., Nera K.P., Veistinen E., Lahdesmaki A., Postila V., Kimby E., Riikonen P., Hammarstrom L., Pelkonen J. and Lassila O. (2001). Both normal and leukemic B lymphocytes express multiple isoforms of the human Aiolos gene. *Eur J Immunol*. **31**: 3469-3474.

Lilie H., Schwarz E. and Rudolph R. (1998). Advances in refolding of proteins produced in *E. coli*. *Curr Opin Biotechnol.* **9**: 497-501.

Lindsley D.L. and Zimm G.G. (1992). The genome of *Drosophila melanogaster*. *Academic Press, San Diego*

Liu P.Q., Rebar E.J., Zhang L., Liu Q., Jamieson A.C., Liang Y., Qi H., Li P.X., Chen B., Mendel M.C., Zhong X., Lee Y.L., Eisenberg S.P., Spratt S.K., Case C.C. and Wolffe A.P. (2001). Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. *J Biol Chem.* **276**: 11323-11334.

Liu Q., Segal D.J., Ghiara J.B. and Barbas C.F. 3rd. (1997). Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc Natl Acad Sci U S A*. **94**: 5525-5530.

Lo Conte L., Chothia C. and Janin J. (1999). The atomic structure of protein-protein recognition sites. *J Mol Biol.* **285**: 2177-2198.

Lodish H., Baltimore D., Berk A., Zipursky L.S., Matsudaira P., Darnell J. (1996). Molekulare Zellbiologie. *Walter de Gruyter. Berlin, New York.*

Lo K., Landau N.R. and Smale S.T. (1991). LyF-1, a transcriptional regulator that interacts with a novel class of promoters for lymphocyte-specific genes. *Mol Cell Biol*. **11**: 5229-5243.

Lopez R.A., Schoetz S., DeAngelis K., O'Neill D. and Bank A. (2002). Multiple hematopoietic defects and delayed globin switching in Ikaros null mice. *Proc Natl Acad Sci U S A*. **99**: 602-607.

Lu D., Searles M.A. and Klug A. (2003). Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. *Nature.* **426**: 96-100.

Lunyak V.V. and Rosenfeld M.G. (2005). No rest for REST: REST/NRSF regulation of neurogenesis. *Cell.* **121**: 499-501.

Lunyak V.V., Prefontaine G.G. and Rosenfeld M.G. (2004). REST and peace for the neuronal-specific transcriptional program. *Ann N Y Acad Sci.* **1014**: 110-120.

Mackay J.P. and Crossley M. (1998). Zinc fingers are sticking together. *Trends Biochem Sci.* **23**: 1-4.

Maeda Y., Yamada H., Ueda T. and Imoto T. (1996). Effect of additives on the renaturation of reduced lysozyme in the presence of 4 M urea. *Protein Eng.* **9**: 461-465.

Matsuzawa-Watanabe Y., Inoue J. and Semba K. (2003). Transcriptional activity of testis-determining factor SRY is modulated by the Wilms' tumor 1 gene product, WT1. *Oncogene.* **22**: 7900-7904.

Matthews J.M., Kowalski K., Liew C.K., Sharpe B.K., Fox A.H., Crossley M. and MacKay J.P. (2000). A class of zinc fingers involved in protein-protein interactions biophysical characterization of CCHC fingers from fog and U-shaped. *Eur J Biochem.* **267**: 1030-1038.

Matthews J.M. and Sunde M. (2002). Zinc fingers--folds for many occasions. *IUBMB Life*. **54**: 351-355.

McCarty A.S., Kleiger G., Eisenberg D. and Smale S.T. (2003). Selective dimerization of a C2H2 zinc finger subfamily. *Mol Cell.* **11**: 459-470.

Michael S.F., Kilfoil V.J., Schmidt M.H., Amann B.T. and Berg J.M. (1992). Metal binding and folding properties of a minimalist Cys2His2 zinc finger peptide. *Proc Natl Acad Sci U S A*. **89**: 4796-4800.

Miller J., McLachlan A.D. and Klug A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J.* **4**:1609-1614.

Molnar A. and Georgopoulos K. (1994). The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins. *Mol Cell Biol.* **14**: 8292-8303.

Molnar A., Wu P., Largespada D.A., Vortkamp A., Scherer S., Copeland N.G., Jenkins N.A., Bruns G. and Georgopoulos K. (1996). The Ikaros gene encodes a family of lymphocyte-restricted zinc finger DNA binding proteins, highly conserved in human and mouse. *J Immunol.* **156**: 585-592.

Momeni P., Glockner G., Schmidt O., von Holtum D., Albrecht B., Gillessen-Kaesbach G., Hennekam R., Meinecke P., Zabel B., Rosenthal A., Horsthemke B. and Ludecke H.J. Mutations in a new gene, encoding a zinc-finger protein, cause tricho-rhino-phalangeal syndrome type I. *Nat Genet.* **24**: 71-74.

Moore M., Klug A. and Choo Y. (2001). Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. *Proc Natl Acad Sci U S A.* **98**: 1437-1441.

Morgan B., Sun L., Avitahl N., Andrikopoulos K., Ikeda T., Gonzales E., Wu P., Neben S. and Georgopoulos K. (1997). Aiolos, a lymphoid restricted transcription factor that interacts with Ikaros to regulate lymphocyte differentiation. *EMBO J.* **16**: 2004-2013.

Mosavi L.K., Minor D.L. Jr. and Peng Z.Y. (2002). Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl. Acad. Sci. U S A.* **99**: 16029-16034.

Moore M., Choo Y. and Klug A. (2001). Design of polyzinc finger peptides with structured linkers. *Proc Natl Acad Sci U S A.* **98**: 1432-1436.

Nagadoi A., Nakazawa K., Uda H., Okuno K., Maekawa T., Ishii S. and Nishimura Y. (1999). Solution structure of the transactivation domain of ATF-2 comprising a zinc finger-like subdomain and a flexible subdomain. *J Mol Biol.* **287**: 593-607.

Nakase K., Ishimaru F., Fujii K., Tabayashi T., Kozuka T., Sezaki N., Matsuo Y., Harada M. (2002). Overexpression of novel short isoforms of Helios in a patient with T-cell acute lymphoblastic leukemia. *Exp Hematol.* **30**: 313-317.

Neely L.S., Lee B.M., Xu J., Wright P.E. and Gottesfeld J.M. (1999). Identification of a minimal domain of 5 S ribosomal RNA sufficient for high affinity interactions with the RNA-specific zinc fingers of transcription factor IIIA. *J Mol Biol.* **291**: 549-560.

Nolte R.T., Conlin R.M., Harrison S.C. and Brown R.S. (1998). Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex. *Proc Natl Acad Sci U S A.* **95**: 2938-2943.

Nose A., Mahajan V.B. and Goodman C.S. (1992). Connectin: a homophilic cell adhesion molecule expressed on a subset of muscles and the motoneurons that innervate them in *Drosophila*. *Cell.* **70**: 553-567.

Novotny T., Eiselt R. and Urban J. (2002). Hunchback is required for the specification of the early sublineage of neuroblast 7-3 in the *Drosophila* central nervous system. *Development.* **129**: 1027-1036.

Omichinski J.G., Clore G.M., Robien M., Sakaguchi K., Appella E. and Gronenborn A.M. (1992). High-resolution solution structure of the double Cys2His2 zinc finger from the human enhancer binding protein MBP-1. *Biochemistry.* **31**: 3907-3917.

Omichinski J.G., Pedone P.V., Felsenfeld G., Gronenborn A.M. and Clore G.M. (1997). The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode. *Nat Struct Biol.* **4**: 122-132.

Orkin S.H. (1995). Hematopoiesis: how does it happen? *Curr Opin Cell Biol*. **7**: 870-877.

Orsini G. and Goldberg M.E. (1978). The renaturation of reduced chymotrypsinogen A in guanidine HCl. Refolding versus aggregation. *J Biol Chem.* **253**: 3453-3458.

Pabo C.O., Peisach E. and Grant R.A. (2001). Design and selection of novel Cys2His2 zinc finger proteins. *Annu Rev Biochem.* **70**: 313-340.

Park S.H., Zarrinpar A. and Lim W.A. (2003). Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms. *Science.* **299**: 1061-1064.

Patel N.H. (1994). Imaging neuronal subsets and other cell types in whole-mount *Drosophila* embryos and larvae using antibody probes. *Methods Cell Biol.* **44**: 445-487.

Patel N.H., Hayward D.C., Lall S., Pirkl N.R., DiPietro D. and Ball E.E. (2001). Grasshopper hunchback expression reveals conserved and novel aspects of axis formation and segmentation. *Development.* **128**: 3459-3472.

Pavletich N.P. and Pabo C.O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science.* **252**: 809-817.

Pavletich N.P. and Pabo C.O. (1993). Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science.* **261**: 1701-1707.

Pawson T. and Nash P. (2003). Assembly of cell regulatory systems through protein interaction domains. *Science.* **513**: 445-452.

Pawson T., Raina M. and Nash P. (2002). Interaction domains: from simple binding events to complex cellular behavior. *FEBS Lett.* **300**: 2-10.

Pawson T. and Scott J.D. (1997). Signaling through scaffold, anchoring, and adaptor proteins. *Science.* **278**: 2075-2080.

Pawson T. (2004). Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell.* **116**: 191-203.

Payre F., Buono P., Vanzo N. and Vincent A. (1997). Two types of zinc fingers are required for dimerization of the serendipity delta transcriptional activator. *Mol Cell Biol.* **17**: 3137-3145.

Pearson B.J. and Doe C.Q. (2003). Regulation of neuroblast competence in *Drosophila*. *Nature.* **425**: 624-628.

Pedone P.V., Ghirlando R., Clore G.M., Gronenborn A.M., Felsenfeld G. and Omichinski J.G. (1996). The single Cys2-His2 zinc finger domain of the GAGA protein flanked by basic residues is sufficient for high-affinity specific DNA binding. *Proc Natl Acad Sci U S A*. **93**: 2822-2826.

Pelham H.R. and Brown D.D. (1980). A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. *Proc Natl Acad Sci U S A*. **77**: 4170-4174.

Pellegrino G.R. and Berg J.M. (1991). Identification and characterization of "zinc-finger" domains by the polymerase chain reaction. *Proc Natl Acad Sci U S A.* **88**: 671-675.

Perdomo J. and Crossley M. (2002). The Ikaros family protein Eos associates with C-terminal-binding protein corepressors. *Eur J Biochem*. **269**: 5885-5892.

Perdomo J., Holmes M., Chong B. and Crossley M. (2000). Eos and pegasus, two members of the Ikaros family of proteins with distinct DNA binding activities. *J Biol Chem.* **275**: 38347-38354.

Pevny L., Simon M.C., Robertson E., Klein W.H., Tsai S.F., D'Agati V., Orkin S.H. and Costantini F. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature*. **349**: 257-260.

Phelps C.B. and Brand A.H. (1998). Ectopic gene expression in *Drosophila* using GAL4 system. *Methods.* **14**: 367-379.

Pluckthun A. and Pack P. (1997). New protein engineering approaches to multivalent and bispecific antibody fragments. *Immunotechnology.* **3**: 83-105.

Polekhina G., House C.M., Traficante N., Mackay J.P., Relaix F., Sassoon D.A., Parker M.W. and Bowtell D.D. (2002). Siah ubiquitin ligase is structurally related to TRAF and modulates TNF-alpha signaling. *Nat Struct Biol.* **9**: 68-75.

Pollock R., Giel M., Linher K. and Clackson T. (2002). Regulation of endogenous gene expression with a small-molecule dimerizer. *Nat Biotechnol.* **20**: 729-733.

Pomerantz J.L., Wolfe S.A. and Pabo C.O. (1998). Structure-based design of a dimeric zinc finger protein. *Biochemistry.* **37**: 965-970.

Ptashne M. and Gann A. (1997). Transcriptional activation by recruitment. *Nature.* **386**: 569-577.

Reichmann D., Rahat O., Albeck S., Meged R., Dym O. and Schreiber G. (2005). The modular architecture of protein-protein binding interfaces. *Proc Natl Acad Sci U S A.* **102**: 57-62.

Remenyi A., Scholer H.R. and Wilmanns M. (2004). Combinatorial control of gene expression. *Nat Struct Mol Biol.* **11**: 812-815.

Rorth P., Szabo K., Bailey A., Laverty T., Rehm J., Rubin G.M., Weigmann K., Milan M., Benes V., Ansorge W. and Cohen S.M. (1998). Systematic gain-of-function genetics in *Drosophila*. *Development.* **125**: 1049-1057.

Rual J.F., Venkatesan K., Hao T., Hirozane-Kishikawa T., Dricot A., Li N., Berriz G.F., Gibbons F.D., Dreze M., Ayivi-Guedehoussou N., Klitgord N., Simon C., Boxem M., Milstein S., Rosenberg J., Goldberg D.S., Zhang L.V., Wong S.L., Franklin G., Li S., Albala J.S., Lim J., Fraughton C., Llamosas E., Cevik S., Bex C., Lamesch P., Sikorski R.S., Vandenhaute J., Zoghbi H.Y., Smolyar A., Bosak S., Sequerra R., Doucette-Stamm L.,

Cusick M.E., Hill D.E., Roth F.P. and Vidal M. (2005). Towards a proteome-scale map of the human protein-protein interaction network. *Nature.* **437**: 1173-1178.

Rubin G.M. and Spradling A.C. (1982). Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218**: 348-353.

Sabbattini P., Lundgren M., Georgiou A., Chow C., Warnes G. and Dillon N. (2001). Binding of Ikaros to the lambda5 promoter silences transcription through a mechanism that does not require heterochromatin formation. *EMBO J.* **20**: 2812-2822.

Saito Y., Doi K., Yamagishi N., Ishihara K. and Hatayama T. (2004). Screening of Hsp105alpha-binding proteins using yeast and bacterial two-hybrid systems. *Biochem Biophys Res Commun.* **314**: 396-402.

Sambrook J. and Russel D.W. (2001). Molecular cloning, Third edition. *Cold Spring Harbor Laboratory Press, Cold Spring Harbor*, *New York*.

Savage R.M. and Shankland M. (1996). Identification and characterization of a hunchback orthologue, Lzf2, and its expression during leech embryogenesis. *Dev Biol.* **175**: 205-217.

Scharnhorst V., van der Eb A.J. and Jochemsen A.G. (2001). WT1 proteins: functions in growth and differentiation. *Gene.* **273**: 141-161.

Schlessinger J. and Lemmon M.A. (2003). SH2 and PTB domains in tyrosine kinase signaling. *Sci. STKE.* **191**: RE12.

Schmeichel K.L. and Beckerle M.C. (1994). The LIM domain is a modular protein-binding interface. *Cell.* **79**: 211-219.

Schmeichel K.L. and Beckerle M.C. (1998). LIM domains of cysteine-rich protein 1 (CRP1) are essential for its zyxin-binding function. *Biochem J.* **331**: 885-892.

Schmidt H., Rickert C., Bossing T., Vef O., Urban J. and Technau G.M. (1997). The embryonic central nervous system lineages of *Drosophila melanogaster*. II. Neuroblast lineages derived from the dorsal part of the neuroectoderm. *Dev Biol.* **189**:186-204.

Schroeder M.D., Pearce M., Fak J., Fan H., Unnerstall U., Emberly E., Rajewsky N., Siggia E.D. and Gaul U. (2004). Transcriptional control in the segmentation gene network of *Drosophila*. *PLoS Biol.* **2**: E271.

Searles M.A., Lu D. and Klug A. (2000). The role of the central zinc fingers of transcription factor IIIA in binding to 5 S RNA. *J Mol Biol*. **301**: 47-60.

Sedgwick S.G. and Smerdon S.J. (1999). The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem. Sci.* **24**: 311-316.

Serebriiskii I.G., Fang R., Latypova E., Hopkins R., Vinson C., Joung J.K., and Golemis E.A. (2005). A combined yeast/bacteria two-hybrid system: development and evaluation. *Mol Cell Proteomics*. **4**: 819-826.

Serebriiskii I.G., Toby G.G., Finley R.L. Jr. and Golemis E.A. (2001). Genomic analysis utilizing the yeast two-hybrid system. *Methods Mol Biol.* **175**: 415-454.

Shastry B.S. (1996). Transcription factor IIIA (TFIIIA) in the second decade. *J Cell Sci*. **109**: 535-539.

Shaywitz A.J., Dove S.L., Greenberg M.E. and Hochschild A. (2002). Analysis of phosphorylation-dependent protein-protein interactions using a bacterial two-hybrid system. *Sci STKE.* **142**: PL11.

Shaywitz A.J., Dove S.L., Kornhauser J.M., Hochschild A. and Greenberg M.E. (2000). Magnitude of the CREB-dependent transcriptional response is determined by the strength of the interaction between the kinase-inducible domain of CREB and the KIX domain of CREB-binding protein. *Mol Cell Biol.* **20**: 9409-9422.

Sprinzak D. and Elowitz M.B. (2005). Reconstruction of genetic circuits. *Nature*. **438:** 443-448.

Sun L., Liu A. and Georgopoulos K. (1996). Zinc finger-mediated protein interactions modulate Ikaros activity, a molecular control of lymphocyte development. *EMBO J.* **15**: 5358-5369.

Tabata T., Schwartz C., Gustavson E., Ali Z. and Kornberg T.B. (1995). Creating a *Drosophila* wing de novo, the role of engrailed, and the compartment border hypothesis. *Development.* **121**: 3359-3369.

Tandon S. and Horowitz P.M. (1987). Detergent-assisted refolding of guanidinium chloride-denatured rhodanese. The effects of the concentration and type of detergent. *J Biol Chem.* **262**: 4486-4491.

Tan S., Guschin D., Davalos A., Lee Y.L., Snowden A.W., Jouvenot Y., Zhang H.S., Howes K., McNamara A.R., Lai A., Ullman C., Reynolds L., Moore M., Isalan M., Berg L.P., Campos B., Qi H., Spratt S.K., Case C.C., Pabo C.O., Campisi J. and Gregory P.D. (2003). Zinc-finger protein-targeted gene regulation: genomewide single-gene specificity. *Proc Natl Acad Sci U S A.* **100**: 11997-12002.

Tautz D., Lehmann R., Schnürch H., Schuh R., Seifert E., Kienlin A., Jones K., Jäckle H. (1987). Finger protein of novel structure encoded by hunchback, a second member of the gap class of *Drosophila* segmentation genes. *Nature.* **327**: 383-389.

Tautz D. (1988). Regulation of the *Drosophila* segmentation gene hunchback by two maternal morphogenetic centres. *Nature.* **332**: 281-284.

Thibodeau S.A., Fang R. and Joung J.K. (2004). High-throughput beta-galactosidase assay for bacterial cell-based reporter systems. *Biotechniques*. **36**: 410-415.

Thomas B.J. and Wassarman D.A. (1999). A fly's eye view of biology. *Trends Genet.* **15**: 184-190.

Thomas J.G., Ayling A. and Baneyx F. (1997). Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from *E. coli*. To fold or to refold. *Appl Biochem Biotechnol.* **66**: 197-238.

Thomas M.J. and Seto E. (1999). Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key? *Gene.* **236**: 197-208.

Thukral S.K., Morrison M.L. and Young E.T. (1991). Alanine scanning site-directed mutagenesis of the zinc fingers of transcription factor ADR1: residues that contact DNA and that transactivate. *Proc Natl Acad Sci U S A.* **88**: 9188-9192.

Trainor C.D., Evans T., Felsenfeld G. and Boguski M.S. (1990). Structure and evolution of a human erythroid transcription factor. *Nature.* **343**: 92-96.

Trinh L.A., Ferrini R., Cobb B.S., Weinmann A.S., Hahm K., Ernst P., Garraway I.P., Merkenschlager M. and Smale S.T. (2001). Down-regulation of TDT transcription in CD4(+)CD8(+) thymocytes by Ikaros proteins in direct competition with an Ets activator. *Genes Dev.* **15**: 1817-1832.

Tsai R.Y. and Reed R.R. (1998). Identification of DNA recognition sequences and protein interaction domains of the multiple-Zn-finger protein Roaz. *Mol Cell Biol.* **18**: 6447-6456.

Tsai S.F., Martin D.I., Zon L.I., D'Andrea A.D., Wong G.G. and Orkin S.H. (1989). Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature.* **339**: 446-451.

Tsang A.P., Visvader J.E., Turner C.A., Fujiwara Y., Yu C., Weiss M.J., Crossley M., Orkin S.H. (1997). FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell.* **90**: 109-119.

Tupler R., Perini G. and Green M.R. (2001). Expressing the human genome. *Nature.* **409**: 832-833.

Tu Y., Li F., Goicoechea S. and Wu C. (1999). The LIM-only protein PINCH directly interacts with integrin-linked kinase and is recruited to integrin-rich sites in spreading cells. Mol Cell Biol. 19: 2425-2434.

Tu Y., Li F. and Wu C. (1998). Nck-2, a novel Src homology2/3-containing adaptor protein that interacts with the LIM-only protein PINCH and components of growth factor receptor kinase-signaling pathways. *Mol Biol Cell.* **9**: 3367-3382.

Urnov F.D., Miller J.C., Lee Y.L., Beausejour C.M., Rock J.M., Augustus S., Jamieson A.C., Porteus M.H., Gregory P.D. and Holmes M.C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature.* **435**: 646-651.

Velyvis A., Vaynberg J., Yang Y., Vinogradova O., Zhang Y., Wu C. and Qin J. (2003). Structural and functional insights into PINCH LIM4 domain-mediated integrin signaling. *Nat Struct Biol.* **10**: 558-564.

Velyvis A., Yang Y., Wu C. and Qin J. (2001). Solution structure of the focal adhesion adaptor PINCH LIM1 domain and characterization of its interaction with the integrin-linked kinase ankyrin repeat domain. *J Biol Chem.* **276**: 4932-4939.

Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G., Smith H.O., Yandell M., Evans C.A., Holt R.A., et al. (2001). The sequence of the human genome. *Science.* **291**: 1304-1351.

Wang B.S., Grant R.A. and Pabo C.O. (2001). Selected peptide extension contacts hydrophobic patch on neighboring zinc finger and mediates dimerization on DNA. *Nat Struct Biol*. **8**: 589-593.

Wang B.S. and Pabo C.O. (1999). Dimerization of zinc fingers mediated by peptides evolved in vitro from random sequences. *Proc Natl Acad Sci U S A*. **96**: 9568-9573.

Wang J.H., Nichogiannopoulou A., Wu L., Sun L., Sharpe A.H., Bigby M. and Georgopoulos K. (1996). Selective defects in the development of the fetal and adult lymphoid system in mice with an Ikaros null mutation. *Immunity*. **5**: 537-549.

Wang W., Lee S.B., Palmer R., Ellisen L.W. and Haber D.A. (2001). A functional interaction with CBP contributes to transcriptional activation by the Wilms tumor suppressor WT1. *J Biol Chem.* **276**: 16810-16816.

Westman B.J., Mackay J.P. and Gell D. (2002). Ikaros: a key regulator of haematopoiesis. *Int J Biochem Cell Biol.* **34**: 1304-1307.

Westman B.J., Perdomo J., Matthews J.M., Crossley M. and Mackay J.P. (2004). Structural studies on a protein-binding zinc-finger domain of Eos reveal both similarities and differences to classical zinc fingers. *Biochemistry.* **43**: 13318-13327.

Westman B.J., Perdomo J., Sunde M., Crossley M. and Mackay J.P. (2003). The C-terminal domain of Eos forms a high order complex in solution. *J Biol Chem.* **278**: 42419-42426.

Whipple F.W. (1998). Genetic analysis of prokaryotic and eukaryotic DNA-binding proteins in *Escherichia coli*. *Nucleic Acids Res.* **26**: 3700-3706.

Wimmer E.A., Carleton A., Harjes P., Turner T. and Desplan C. (2000). Bicoid-independent formation of thoracic segments in *Drosophila*. *Science.* **287**: 2476-2479.

Winandy S., Wu P. and Georgopoulos K. (1995). A dominant mutation in the Ikaros gene leads to rapid development of leukemia and lymphoma. *Cell.* **83**: 289-299.

Wolfe S.A., Grant R.A., Elrod-Erickson M. and Pabo C.O. (2001). Beyond the "recognition code": structures of two Cys2His2 zinc finger/TATA box complexes. *Structure.* **9**: 717-723.

Wolfe S.A., Grant R.A. and Pabo C.O. (2003). Structure of a designed dimeric zinc finger protein bound to DNA. *Biochemistry.* **42**: 13401-13409.

Wolfe S.A., Greisman H.A., Ramm E.I. and Pabo C.O. (1999). Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code. *J Mol Biol.* **285**: 1917-1934.

Wolfe S.A., Nekludova L. and Pabo C.O. (2000). DNA recognition by Cys2His2 zinc finger proteins. *Annu Rev Biophys Biomol Struct.* **29**: 183-212.

Wolfe S.A., Ramm E.I. and Pabo C.O. (2000). Combining structure-based design with phage display to create new Cys(2)His(2) zinc finger dimers. *Structure.* **8**: 739-750.

Wu H., Yang W.P. and Barbas C.F. 3rd. (1995). Building zinc fingers by selection: toward a therapeutic application. *Proc Natl Acad Sci U S A.* **92**: 344-348.

Wuttke D.S., Foster M.P., Case D.A., Gottesfeld J.M. and Wright P.E. (1997). Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: determinants of affinity and sequence specificity. *J Mol Biol.* **273**: 183-206.

Wuttke D.S., Foster M.P., Case D.A., Gottesfeld J.M. and Wright P.E. (1997). Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: determinants of affinity and sequence specificity. *J Mol Biol.* **273**: 183-206.

Yang M., May W.S. and Ito T. (1999). JAZ requires the double-stranded RNA-binding zinc finger motifs for nuclear localization. *J Biol Chem*. **274**: 27399-27406.

Zhou Q., Gedrich R.W. and Engel D.A. (1995). Transcriptional repression of the c-fos gene by YY1 is mediated by a direct interaction with ATF/CREB. *J Virol.* **69**: 4323-4330.

Zhu X.S., Linhoff M.W., Li G., Chin K.C., Maity S.N. and Ting J.P. (2000). Transcriptional scaffold: CIITA interacts with NF-Y, RFX, and CREB to cause stereospecific regulation of the class II major histocompatibility complex promoter. *Mol Cell Biol.* **20**: 6051-6061.

Appendix.

Plasmid name	Description / purpose	Source / Method of constructing	Cloning sites used for further cloning procedures
pBR-UV5-λcI	Phagemid used in the B2H system to clone protein X as a C-terminal fusion to the DBD of the λcI protein. Expression is driven by the IPTG-inducible *lac*UV5 promoter; plasmid has f1 ori of replication.	Kindly provided by K. Joung. See also Giesecke and Joung (2005).	BglII-XbaI
pBR-UV-REST F3-8	Phagemid used in the B2H system to clone protein X as a C-terminal fusion to the DBD of REST represented by ZF3-8. Expression is driven by the IPTG-inducible *lac*UV5 promoter; plasmid has f1 ori of replication.	Plasmid was constructed by replacing the SalI-BamHI fragment of the pBR-GP-Z12BbsI phagemid (Joung *et al.*, 2000) with a PCR amplified fragment consisting of the UV5 promoter region (starting at the SalI site) of the pBR plasmid and the F3-8 fragment. An additional NotI cloning site was included at the C-terminus (N-terminal to the BamHI site).	NotI-BamHI
pBR-UV5-Zif268	Phagemid used in the B2H system to clone protein X as an N-terminal fusion to the DBD of the murine Zif268 (residues 327-421). Expression is driven by the IPTG-inducible *lac*UV5 promoter; plasmid has f1 ori of replication.	Plasmid was generated by replacing the SalI-NotI fragment of the pBR-GP-Z123 (Joung *et al.*, 2000) with a short DNA fragment consisting of two annealed oligonucleotides that incorporated a NdeI site.	NdeI-XhoI. Note that there are two more NdeI sites.
pACYC-α	Plasmid used in the B2H system to clone protein Y as a C-terminal fusion to the α-subunit (residues 1-248) of the *E. coli* RNAP. Expression is driven by the IPTG-inducible tandem *lpp/lac*UV5 promoter.	Original plasmid pACYC-α (Joung *et al.*, 2000) was modified by introducing a short DNA fragment consisting of two annealed oligonucleotides that incorporated a BamHI, KpnI and XhoI site into the NotI-BglII site of the original plasmid (R. Fang).	BamHI-XhoI

pBR- UV5-Zif268 BbsI stuffer	Phagemid used to create libraries of shuffled DZF domains for B2H selections. Shuffled fragments were cloned into the BbsI site, thereby creating N-terminal fusions to the DBD of the murine Zif268 (residues 327-421). Expression is driven by the IPTG-inducible *lac*UV5 promoter; plasmid has f1 ori of replication.	A BbsI "stuffer" fragment containing two BbsI restriction sites was designed by annealing two complementary oligonucleotides together. In addition, a 5' Sal I and a 3' BspEI overhang was created which was used to clone the annealed product into the Sal I and BspEI sites of the pBR-UV5-Zif268 plasmid.	BbsI
pSB stuffer	Plasmid used to assemble DBS to construct reporter strains for the B2H system.	Plasmid is a derivate of pFW11 (Whipple, 1998) and was kindly provided by K. Joung.	EcoRI-SalI or SapI
pcDNA5-p65	Plasmid used in the mammalian cell-based activator reconstitution assay to clone protein X as a C-terminal fusion to the human NF-κB p65 subunit (residues 283-551). Plasmid harbors an N-terminal nuclear localization signal (NLS) from SV40 large T antigen and expression is driven by a modified cytomegalovirus (CMV) promoter that can be repressed by the tetracycline repressor.	pcDNA5 was obtained from Invitrogen. Plasmid pcDNA5-p65 was made by amplifying the p65 fragment using PCR with two specific external primers. The Top strand PCR primer introduced a nuclear localization signal from the simian virus 40 (SV40) large T antigen at the N-terminus of p65 and incorporated a HindIII site. A BamHI and XhoI site for further cloning procedures was introduced at the C-terminus of p65 by the bottom strand primer. The amplified DNA was cloned into the HindIII and XhoI sites of the pcDNA5 expression vector.	BamHI-XhoI
pcDNA5-SpI	Plasmid used in the mammalian cell-based activator reconstitution assay to clone protein Y as a C-terminal fusion to the synthetic VZ+434b DBD, (Liu *et al.*, 2001) that binds to the human VEGF-A gene. Plasmid harbors an N-terminal NLS from SV40 large T antigen and expression is driven by a modified CMV.	Plasmid pcDNA5-SpI was made by replacing the p65 fragment of the pcDNA5-NA-VZ+434b plasmid (K. Joung) with a BamHI-XhoI cloning site. Complementary oligonucleotides were annealed together to generate a DNA fragment that bear BamHI and XhoI sticky ends which was then cloned into the BamHI and XhoI sites of the pcDNA5-NA-VZ+434b plasmid vector.	BamHI-XhoI
pcDNA5-FLAG-p65	Plasmid is the same as pcDNA5-p65 but harbors a N-terminal FLAG-tag to	The FLAG-tag encoding sequence was incorporated in one of the external primers used to amplify a left cassette fragment.	BamHI-XhoI

	perform quantitative Western blot analysis.	This fragment was then fused to the right cassette fragment encoding the p65 protein. The resulting PCR fragments were cloned into the HindIII and BamHI sites of the pcDNA5-p65 plasmid.	
pcDNA5-FLAG-SpI	Plasmid is the same as pcDNA5-SpI but harbors a N-terminal FLAG-tag to perform quantitative Western blots.	Fragment encoding FLAG-SpI was made as described for FLAG-p65 (see above) and cloned into the HindIII and BamHI sites of the pcDNA5-SpI plasmid.	BamHI-XhoI
pcDNA3.1-IkI	Plasmid used in mammalian cell based co-immunoprecipitation assay to express protein X as a fusion to the C-terminus of the Ikaros isoform I (Ik-I, residues 1-222). Expression of the fusion protein is driven by a modified CMV promoter and no NLS is present.	pcDNA3.1 was obtained from Invitrogen. The Ik I was assembled from overlapping oligonucleotides using PCR and cloned into the EcoRI-BamHI site of pcDNA3.1 (R. Fang).	KpnI-HindIII
pcDNA3.1-FLAG-NTS	Plasmid used in mammalian cell based co-immunoprecipitation assay to express protein Y as a fusion to the N-terminal fragment of Ik-I (NTS) Plasmid also contained a N-terminal FLAG-tag. Expression is driven by a modified CMV promoter and the NLS is missing.	NTS was assembled from oligonucleotides and cloned into the BamHI-KpnI site of pcDNA3.1. An assembled FLAG tag was included at the N-terminus of the fusion protein by cloning it into the EcoRI-BamHI (R. Fang).	KpnI-HindIII
pACYC-λcI	Plasmid used in the B1H system to express protein X as a C-terminal fusions to the N-terminal DNA binding domain of the lambda cI repressor (λcI NTD). Expression is driven by the IPTG-inducible *lac*UV5 promoter.	Plasmid was originally made in Ann Hochschilds lab and was further modified by introducing a KpnI-HindIII cloning site incorporated in overlapping oligonucleotides which was cloned into the NotI-BamHI site of pACYC-λcI (Rui Fang).	KpnI-HindIII
pET3a	Plasmid used to express peptides under the control of an IPTG-inducible promoter.	pET3a was obtained from Novagene	NdeI-BamHI
pUAST	Plasmid used to express cDNAs in the GAL4-UAS system under control of the	pUAST containing the wild-type Hunchback cDNA (including extra non translated C-terminal sequence) was	XhoI-KpnI

	UAS promoter (Brand and Perrimon, 1993). Plasmid also contained P-elements, the Hsp70 promoter, a Poly-adenylation site from SV40 and the white^{+} gene (in-frame with the cDNA of interest).	kindly provided by J. Urban.	

A1. Description of primary plasmids. The plasmid name is indicated in the first column. A brief description of the plasmid is shown in the second column. The third column describes the source of the plasmid or techniques employed to construct the plasmid. Note that PCR and primer annealing strategies are described in Chapter 2, sections 2.1.2.2 and 2.1.2.3. Restriction sites used for further cloning procedures (see A2) are indicated in the last column.

Fragment name	Plasmid	Amino acids	Method of gene synthesis and cloning
REST F9	pBR-UV5-λcI pBR-UV-REST F3-8 pBR-UV5-Zif268	Human REST: 1008-1097	Amplified from IMAGE clone (ATCC).
REST F9 long	pBR-UV5-Zif268	Human REST: 1008-1097	*De novo* synthesis using PCR strategy B. Codons optimized for *E. coli.*
REST F9 short	pBR-UV5-Zif268	Human REST: 1049-1097	*De novo* synthesis using PCR strategy A. Codons optimized for *E. coli.*
CoREST A	pACYC-α	Human CoREST: 102-195	Amplified from human cDNA (Panomics).
CoREST B	pACYC-α	Human CoREST: 130-180	Amplified from human cDNA (Panomics).
GATA-Nf	pBR-UV5-λcI pBR-UV5-Zif268	Mouse GATA-1: 200-254	*De novo* synthesis using PCR strategy B.
FOG-1 F1	pACYC-α	Mouse FOG-1: 241-295	Amplified from IMAGE clone (ATCC).

FOG-1 F6	pACYC-α	Mouse FOG-1: 677-760	*De novo* synthesis using PCR strategy B.
FOG-1 F9	pACYC-α	Mouse FOG-1: 945-995	Amplified from IMAGE clone (ATCC).
CRP1	pBR-UV5-Zif268 pACYC-α	Human CRP1: 6-175	Amplified from human cDNA (Panomics).
CRP1 LIM2	pBR-UV5-Zif268 pACYC-α	Human CRP1: 115-175	Amplified from human cDNA (Panomics).
Zyxin LIM1	pBR-UV5-Zif268 pACYC-α	Human Zyxin: 351-409	Amplified from human cDNA (Panomics).
Chicken Zyxin LIM1	pBR-UV5-Zif268 pACYC-α	Chicken Zyxin: 351-409	*De novo* synthesis using PCR strategy B.
PINCH LIM1	pBR-UV5-Zif268	Human PINCH: 6-68	Amplified from human cDNA (Panomics).
PINCH LIM4	pBR-UV5-Zif268	Human PINCH: 189-248	Amplified from human cDNA (Panomics).
LMO2	pBR-UV5-Zif268	Human LMO2: 22-150	Amplified from human cDNA (Panomics).
LMO2 LIM1	pBR-UV5-Zif268	Human LMO2: 22-88	Amplified from plasmid encoding LMO2.
LMO4	pBR-UV5-Zif268	Human LMO4: 15-145	*De novo* synthesis using PCR strategy B. Codons optimized for *E. coli.*
LMO4 LIM1	pBR-UV5-Zif268	Human LMO4: 15-81	Amplified from plasmid encoding LMO4.
ILK	pACYC-α	Human ILK: 1-162	Amplified from human cDNA (Panomics).
Nck2	pACYC-α	Human Nck2: 161-273	Amplified from human cDNA (Panomics).
LID	pACYC-α	Human ldb1: 299-238	*De novo* synthesis using PCR strategy B.
Ik	pBR-UV5-Zif268 pACYC-α	Human Ikaros: 456-519	Amplified from human cDNA (Panomics) (R. Fang).

Ai	pBR-UV5-Zif268 pACYC-α	Human Aiolos: 446-509	*De novo* synthesis using PCR strategy B. Codons optimized for *E. coli.*
He	pBR-UV5-Zif268 pACYC-α	Human Helios: 465-526	Amplified from human cDNA (Panomics).
Eo	pBR-UV5-Zif268 pACYC-α	Human Eos: 471-532	Amplified from human cDNA (Panomics).
Pe	pBR-UV5-Zif268 pACYC-α	Human Pegasus: 358-420	Amplified from human cDNA (Panomics), used Strategy D to silence internal NdeI site.
Tr	pBR-UV5-Zif268 pACYC-α	Human TRPS-1: 1209-1272	*De novo* synthesis using PCR strategy B. Codons optimized for *E. coli.*
Hd	pBR-UV5-Zif268 pACYC-α	Hunchback from *D.m.*: 699-758	*De novo* synthesis using PCR strategy B.
Hl	pBR-UV5-Zif268 pACYC-α	Hunchback from *L.m.*: 453-512	*De novo* synthesis using PCR strategy B.
Hh	pBR-UV5-Zif268 pACYC-α	Hunchback from *H.t.*: 409-468	*De novo* synthesis using PCR strategy B.
Hc	pBR-UV5-Zif268 pACYC-α	Hunchback from *C.e.*: 923-982	*De novo* synthesis using PCR strategy B. Codons optimized for *E. coli.*
Tr-Eo-Eo	pBR-UV5-Zif268 pACYC-α	TRPS-1: 1209-1238 Eos: 504-532	Amplified from respective plasmid which was isolated from B2H selection strain.
Eo-Eo-Hd	pBR-UV5-Zif268 pACYC-α	Eos: 471-504 Hunchback *D.m.*: 733-758	“
Pe-Hd-Hd	pBR-UV5-Zif268 pACYC-α	Pegasus: 358-386 Hunchback *D.m.*: 728-758	“
Hl-Eo-Eo	pBR-UV5-Zif268 pACYC-α	Hunchback *L.m.*: 453-481 Eos: 504-532	“
Ik-Hd-Hd	pBR-UV5-Zif268 pACYC-α	Ikaros: 456-484 Hunchback *D.m.*: 728-758	“
Pe-Pe-Eo	pBR-UV5-Zif268 pACYC-α	Pegasus: 358-391 Eos: 509-532	“

Ik	pcDNA5-SpI pcDNA-p65	Human Ikaros: 456-519	Subcloned from pACYC construct containing the respective DZF domain using BamHI and XhoI.
Eo	pcDNA5-SpI pcDNA-p65	Human Eos: 471-532	“
Pe	pcDNA5-SpI pcDNA-p65	Human Pegasus 358-420	“
Tr	pcDNA5-SpI pcDNA-p65 pcDNA-FLAG-SpI pcDNA-FLAG-p65	Human TRPS-1: 1209-1272	“
Hd	pcDNA5-SpI pcDNA-p65	Hunchback from *D.m.*: 699-758	“
Hl	pcDNA5-SpI pcDNA-p65	Hunchback from *L.m.*: 453-512	“
Tr-Eo-Eo	pcDNA5-SpI pcDNA-p65 pcDNA-FLAG-SpI pcDNA-FLAG-p65	TRPS-1: 1209-1238 Eos: 504-532	“
Eo-Eo-Hd	pcDNA5-SpI pcDNA-p65 pcDNA-FLAG-SpI pcDNA-FLAG-p65	Eos: 471-504 Hunchback *D.m.*: 733-758	“
Pe-Hd-Hd	pcDNA5-SpI pcDNA-p65 pcDNA-FLAG-SpI pcDNA-FLAG-p65	Pegasus: 358-386 Hunchback *D.m.*: 728-758	“
Hl-Eo-Eo	pcDNA5-SpI pcDNA-p65 pcDNA-FLAG-SpI pcDNA-FLAG-p65	Hunchback *L.m.*: 453-481 Eos: 504-532	“
Ik-Hd-Hd	pcDNA5-SpI pcDNA-p65	Ikaros: 456-484 Hunchback *D.m.*: 728-758	“

	pcDNA-FLAG-SpI pcDNA-FLAG-p65		
Pe-Pe-Eo	pcDNA5-SpI pcDNA-p65 pcDNA-FLAG-SpI pcDNA-FLAG-p65	Pegasus: 358-391 Eos: 509-532	"
dDZF-1	pcDNA5-SpI	Residues as described for the individual synthetic DZFs.	Construct was made using PCR Strategy C. Individual DZF fragments were first amplified by PCR and identical regions in the internal oligonucleotides were used for a subsequent fusion PCR. BamHI and XhoI sites were incorporated in the external primers which were applied to clone the various PCR fragments as fusions to SpI.
dDZF-2	pcDNA5-SpI	"	"
dDZF-3	pcDNA5-SpI	"	"
dDZF-4	pcDNA5-SpI	"	"
dDZF-5	pcDNA-p65	"	Construct was made as described for dDZF-1 and cloned as a fusion to p65.
dDZF-6	pcDNA-p65	"	"
dDZF-7	pcDNA-p65	"	"
dDZF-8	pcDNA-p65	"	"
VP16- Pe-Hd-Hd	pcDNA-p65	Herpes simplex virus VP16 activation domain: 414-491 Pegasus: 358-386 Hunchback *D.m.*: 728-758	Fragment was made using PCR Strategy C. Individual fragments encoding the VP16 and the Pe-Hd-Hd peptides were amplified first. An identical linker region encoding a BamHI site was used as internal primers for the subsequent fusion PCR step. An N-terminal HindIII site followed by the NLS from the simian virus 40 (SV40) large T antigen and a C-terminal XhoI site were introduced by the external primers. The resulting PCR fragment was used to replace the HindIII - XhoI fragment (including p65) of pcDNA5-p65.
Tr-Eo-Eo	pcDNA3.1-IkI	TRPS-1: 1209-1238 Eos: 504-532	Subcloned from pACYC construct containing the respective DZF domain using KpnI and HindIII.

Eo-Eo-Hd	pcDNA3.1-FLAG-NTS	Eos: 471-504 Hunchback *D.m.*: 733-758	“
Pe-Hd-Hd	pcDNA3.1-IkI	Pegasus: 358-386 Hunchback *D.m.*: 728-758	“
Hl-Eo-Eo	pcDNA3.1-FLAG-NTS	Hunchback *L.m.*: 453-481 Eos: 504-532	“
Ik-Hd-Hd	pcDNA3.1-FLAG-NTS	Ikaros: 456-484 Hunchback *D.m.*: 728-758	“
Pe-Pe-Eo	pcDNA3.1-IkI	Pegasus: 358-391 Eos: 509-532	“
Ik-Zif268-GP-FLAG	pBR-UV5-Zif268	Human Ikaros: 456-519 Murine Zif268: 334-421 Yeast Gal11P: 263-352	Construct was made using PCR Strategy C. Individual fragments expressing the Zif268 and the GP domain were amplified first. The identical linker region encoding the GGGGS linker including a BamHI site was used as internal primers for the fusion PCR step. An N-terminal XhoI site and a C-terminal FLAG tag followed by a XbaI site were introduced by the external primers. The resulting PCR product was used to replace the XhoI-XbaI fragment of the pBR-Ik-Zif268 plasmid in order to create the pBR-Ik-Zif268-GP-FLAG plasmid.
Ik-Z23-GP-FLAG	pBR-UV5-Zif268	Human Ikaros: 456-519 Murine Zif268: 367-421 Yeast Gal11P: 263-352	Plasmid pBR-Ik-Z23-GP carried a deletion of F1 of Zif268 and was made by replacing the XhoI-BamHI segment containing the Zif268 domain with a PCR amplified fragment consisting of F23 of Zif268.
Ik-(L1-L6)-Z23-GP-FLAG	pBR-UV5-Zif268	Human Ikaros: 456-519 Murine Zif268: 367-421 Yeast Gal11P: 263-352	Plasmids expressing various linker in between the Ik-DZF domain and F23 of Zif268 were constructed using PCR Strategy C. Left cassette fragments expressing the Ik-DZF domain including region up to the SalI site and right cassette fragments encoding the Z23 fragment were fused together using the various linker as identical regions. The resulting PCR fragments were then cloned into the SalI-BamHI site of the pBR-Ik-Z23-GP plasmid.
Ik	pACYC-λcI	Human Ikaros: 456-519	Subcloned from pACYC construct containing the respective DZF domain using KpnI and HindIII.
Pe	pACYC	Human Pegasus: 358-420	“

Hd	pACYC	Hunchback from *D.m.*: 699-758	"
Ik	pET3a	Human Ikaros: 457-519	Amplified from plasmid encoding the respective DZF domain (R. Fang).
Pe	pET3a	Human Pegasus: 360-420	"
Tr	pET3a	Human TRPS-1: 1211-1272	"
Hd	pET3a	Hunchback from *D.m.*: 702-758	"
Hc	pET3a	Hunchback from *C.e.*: 925-982	"
Tr-Eo-Eo	pET3a	TRPS-1: 1211-1238 Eos: 504-532	"
Zif268-Pe	pET3a	murine Zif268: 333-421 Pegasus: 360-420	Constructed using PCR Strategy C. Left cassette fragments encoding the Zif268 domain and right cassette fragments encoding the Pegasus DZF domain were fused together via an identical region in the internal primer. The resulting PCR fragment was cloned into the pET3a expression plasmid.
Ik Nf	pET3a	Human Ikaros: 457-489	Amplified from plasmid DNA.
Ik Cf	pET3a	Human Ikaros: 485-519	"
Hd Nf	pET3a	Hunchback from *D.m.*: 702-732	"
Hd Cf	pET3a	Hunchback from *D.m.*: 728-758	"
Hd	Original pUAST-Hd cDNA	Hunchback *D.m.*: 1-758	Constructed using PCR Strategy C. Left cassette fragments encoding the Hunchback upstream of and including an internal XhoI site and right cassette fragments encoding the respective DZF domain were fused together via an identical region in the internal primer. The fused PCR fragment was then used to replace the XhoI-KpnI fragment in the original

			pUAST-Hunchback cDNA plasmid.
Hd-myc	pUAST	Hunchback *D.m.*: 1-758	Constructed as described above (for Hd) whereas the external C-terminal primer introduced an additional myc-tag.
Hd-HA	pUAST	Hunchback *D.m.*: 1-758	Constructed as described above (for Hd-myc) whereas the external C-terminal primer incorporated a HA-tag encoding sequence.
Hd Ik	pUAST	Hunchback *D.m.*: 1-704 Ikaros: 462-519	Constructed as described for Hd using the Ikaros DZF domain as a template for generating the right cassette.
Hd Ik-myc	pUAST	Hunchback *D.m.*: 1-704 Ikaros: 462-519	Constructed as described for Ik whereas the external C-terminal primer introduced an additional myc-tag.
Hd Ik-HA	pUAST	Hunchback *D.m.*: 1-704 Ikaros: 462-519	Constructed as described for Ik whereas the external C-terminal primer introduced an additional HA-tag.
Hd TrEoEo	pUAST	Hunchback *D.m.*: 1-704 TRPS-1: 1215-1238 Eos: 504-532	Constructed as described for Hd using the Tr-Eo-Eo DZF domain as a template for generating the right cassette.
Hd TrEoEo-myc	pUAST	Hunchback *D.m.* 1-704 TRPS-1: 1215-1238 Eos 504-532	Constructed as described for Tr-Eo-Eo whereas the external C-terminal primer incorporated a myc-tag encoding sequence.
Hd EoHdHd	pUAST	Hunchback *D.m.*: 1-704 Eos: 477-504 Hunchback *D.m.*: 733-758	Constructed as described for Hd using the Eo-Eo-Hd DZF domain as a template for generating the right cassette.
Hd EoHdHd-HA	pUAST	Hunchback *D.m.*: 1-704 Eos: 477-504 Hunchback *D.m.*: 733-758	Constructed as described for Eo-Eo-Hd whereas the external C-terminal primer introduced an additional HA-tag.

A2. Description of constructs. The fragment name and plasmid it was cloned into are indicated in the first two columns. The third column shows the residue numbers the fragment corresponds to in the full-length protein. The organism the fragment is derived from is also indicated. A description if the cloning techniques used to generate the construct is shown in the last column. PCR strategies are indicated and described in section 2.1.2.2. *D.m.* = *Drosophila melanogaster, L.m.* = *Locusta migratoria, H.t.* = *Helobdella triserialis, C.e.* = *Caenorhabditis elegans*, Ik = human Ikaros, Eo = human Eos, Pe = human Pegasus, Tr = human TRPS1, Hd = *D.m.* Hunchback, Hl = *L.m.* Hunchback, Hc = *C.e.* Hunchback.

Reporter name	Description / of reporter strain	Source / Method of constructing the respective pSB stuffer plasmid
RP28	Expression of the *lacZ* reporter gene is controlled by a promoter harboring the λcI DBS at position -62 relative to this promoter.	The pSB plasmid containing the λcI binding site at position -62 was made by subcloning the λcI binding site (derived from plasmid KJ1567 provided by K. Joung) into the EcoRI-SalI site of pSB. The corresponding reporter strain was constructed by R. Fang.
KJST178	Expression of the *lacZ* reporter gene is controlled by a promoter harboring the REST F3-8 DBS in orientation 345678 relative to the promoter.	Reporter strain was kindly provided by S. Thibodeau and K. Joung.
KJ1697, KJ1701	Expression of the *lacZ* reporter gene is controlled by a promoter harboring the REST F3-8 DBS at two positions in orientation 876543 relative to this promoter.	Reporter strains were kindly provided by K. Joung.
KJAG275A	Expression of the *lacZ* reporter gene is controlled by a promoter harboring the Zif268 DBS at position -46 in orientation 321 relative to this promoter.	SapI "stuffer" plasmids containing the SapI restriction sides at various positions relative to the promoter were designed. To do this, 4 overlapping primers (P1, P2, P3 and P4) were annealed and amplified with a pair of external primers (P1 and P4) using PCR. While P1 and P2 were constant primers that encoded the right cassette of the construct, P3 and P4 represented unique primers that created various left cassettes of the constructs by introducing ascending numbers of spacer nucleotides in between the SapI restriction side and the promoter. An EcoRI and a SalI restriction site were incorporated in the external primers. Each resulting stuffer cassette was cloned into the EcoRI and SalI sites of the pSB stuffer plasmid. To create the Zif 268 binding side, two complementary primers were annealed together and cloned into the Sap I sites of the pSB stuffer plasmids constructed as described above.
KJAG293A	Reporter strain harboring the Zif268 DBS at position -48 in orientation 321 relative to the *lacZ* promoter.	"
KJAG324A	Reporter strain harboring the Zif268 DBS at position -50 in orientation 321 relative to the *lacZ* promoter.	"
CC2-A	Reporter strain harboring the Zif268 DBS at position -61 in orientation 321 relative to the *lacZ* promoter.	Reporter strain was kindly provided by K. Joung.

KJAG232A	Reporter strain harboring the Zif268 DBS at position -63 in orientation 321 relative to the *lacZ* promoter.	“
KJAG242A	Reporter strain harboring the Zif268 DBS at position -65 in orientation 321 relative to the *lacZ* promoter.	“
KJAG338B	Reporter strain harboring the Zif268 DBS at position -67 in orientation 321 relative to the *lacZ* promoter.	“
KJAG311A	Reporter strain harboring the Zif268 DBS at position -69 in orientation 321 relative to the *lacZ* promoter.	“
KJAG319B	Reporter strain harboring the Zif268 DBS at position -71 in orientation 321 relative to the *lacZ* promoter.	“
KJAG290A	Reporter strain harboring the Zif268 DBS at position -42 in orientation 123 relative to the *lacZ* promoter.	“
KJAG267A	Reporter strain harboring the Zif268 DBS at position -44 in orientation 123 relative to the *lacZ* promoter.	“
KJAG274A	Reporter strain harboring the Zif268 DBS at position -46 in orientation 123 relative to the *lacZ* promoter.	“
KJAG294A	Reporter strain harboring the Zif268 DBS at position -48 in orientation 123 relative to the *lacZ* promoter.	“
KJAG326A	Reporter strain harboring the Zif268 DBS at position -50 in orientation 123 relative to the *lacZ* promoter.	“
KJAG386A	Reporter strain harboring the Zif268 DBS at position -55 in orientation 123 relative to the *lacZ* promoter.	“
KJAG383A	Reporter strain harboring the Zif268 DBS at position -57 in orientation 123 relative to the *lacZ* promoter.	“
KJAG378A	Reporter strain harboring the Zif268 DBS at position -59 in orientation 123 relative to the *lacZ* promoter.	“
CC1-A	Reporter strain harboring the Zif268 DBS at position -61 in orientation 123 relative to the *lacZ* promoter.	Reporter strain was kindly provided by K. Joung.
KJAG237A	Reporter strain harboring the Zif268 DBS at position -63 in orientation 123 relative to the *lacZ* promoter.	“
KJAG243A	Reporter strain harboring the Zif268 DBS at position -65 in orientation 123 relative to the *lacZ* promoter.	“

KJAG336A	Reporter strain harboring the Zif268 DBS at position -67 in orientation 123 relative to the *lacZ* promoter.	"
KJAG300B	Reporter strain harboring the Zif268 DBS at position -69 in orientation 123 relative to the *lacZ* promoter.	"
KJAG302A	Reporter strain harboring the Zif268 DBS at position -71 in orientation 123 relative to the *lacZ* promoter.	"
KJ0A	Expression of the *lacZ* reporter gene is controlled by a promoter harboring a composite DBS for binding of Finger 2 and 3 of Zif268 with 0 bp present in between the two half-sites..	Complementary primers encoding the half-sites with various spacings were annealed together to make double stranded DNA fragments bearing SapI sticky ends. The annealed oligonucleotides were cloned into the Sap I sites of the pSB stuffer plasmid.
KJ1A	Reporter strain harboring a composite DBS with 1 bp in between the two half-sites.	"
KJ3A	Reporter strain harboring a composite DBS with 3 bp in between the two half-sites.	"
KJ5A	Reporter strain harboring a composite DBS with 5 bp in between the two half-sites.	"
KJ7A	Reporter strain harboring a composite DBS with 7 bp in between the two half-sites.	"
KJ9A	Reporter strain harboring a composite DBS with 9 bp in between the two half-sites.	"
KJ11B	Reporter strain harboring a composite DBS with 11 bp in between the two half-sites.	"
KJ13B	Reporter strain harboring a composite DBS with 13 bp in between the two half-sites.	"
KJ16A	Reporter strain harboring a composite DBS with 16 bp in between the two half-sites.	"
KJ22A	Reporter strain harboring a composite DBS with 22 bp in between the two half-sites.	"
FW123	Expression of the *lacZ* reporter gene is controlled by a promoter harboring the λcI DBS positioned between the -35 and -10 promoter region.	Reporter strain was kindly provided by F. Whipple.
RP45	Selection strain in which the expression of the *his3* and	Reporter strain was kindly provided by R. Fang.

	the *aadA* reporter genes is controlled by a weak promoter bearing an upstream Zif268 binding side at position -65 in orientation 321.	

A3. Description of reporter strains. The reporter name is indicated in the first column. A brief description of the reporter is shown in the second column. Note that the DNA binding sites (DBSs) for the various DNA-binding domains are placed at various distances and orientations relative to the promoter that controls *lacZ* expression. The transcription startpoint of the reporter gene is defined as position +1 and the DBS sits therefore variable numbers of nucleotides upstream of the startpoint Individual fingers in the DBD bind to corresponding subsite marked by a number. For example, finger 1 of Zif268 binds to subsite 1. Thus, in orientation 321, the binding site is positioned in a way that the N-terminal Zif268 finger (finger 1) points toward the promoter while orientation 123 positions the C-terminal finger (finger 3) close to the promoter. The third column describes the source of the plasmid or techniques employed to construct the plasmid used to make the respective reporter strain (see section 2.2.2.1).

Acknowledgements

I am greatly thankful for the supervision I have received from Keith Joung during the course of the last three years. He always supported me with guidance, advice and enthusiasm even when projects did not work out immediately. I also admire his ability to keep track of every project that goes on in the lab and for finding time to answer whatever question I had no matter how busy his schedule was.

I thank Magdalena, Shondra, Edouard, Tao and especially Rui for being great lab-mates and the laughs we had especially during cookie-hours. I also thank Stacy for teaching me different techniques as well as keeping the lab running and Andy for helpful discussions during lab meetings. I am grateful for the help from Robert Grant and Christy Taylor with problems regarding protein-purification and for providing the facilities to perform these experiments in the first place. I have also learnt a lot from Joachim Urban, Ulricke Mettler and Ralf Stanewsky, and I am very grateful for their advice and help with everything related to flies. I thank Jeffrey Hall and Joachim Urban for the opportunity to conduct my fly project in their labs at Brandeis University and the University of Mainz. Finally, I also thank Gisela Szabo and Erhard Strohm for help with last minute experiments.

Personal, I am thankful for all the support I have received from my family and friends over the last three years. Especially, I really appreciate the encouragement and interest my parents show in whatever I do and I bet that nobody has practiced how to pronounce “Drosophila melanogaster” as often as my mom has. Finally, I thank Ralf for all his patience and for being there for me during the course of the last three years.

Declaration

I hereby declare that to the best of my knowledge all the experiments described in this Dissertation were performed by me unless stated otherwise. Any known concepts or literature from others used in this work are acknowledged.

This Dissertation or any part of it has not been submitted for the purpose of obtaining any other degree.

Regensburg,
May, 2006

Astrid Giesecke