ZINC FINGER NUCLEASE-INDUCED DOUBLE-STRAND BREAKS MEDIATE TARGETED MUTAGENESIS IN *CAENORHABDITIS ELEGANS*

by

John Jason Morton

A dissertation submitted to the faculty of The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry

The University of Utah

May 2007
SUPervisory Committee APPROval

of a dissertation submitted by

John Jason Morton

This dissertation has been read by each member of the following supervisory committee and by majority vote has been found to be satisfactory.

Chair: Dana Carroll
Tim Formosa
Erik Jorgensen
Michael S. Kay

1/19/07
To the Graduate Council of the University of Utah:

I have read the dissertation of John Jason Morton in its final form and have found that (1) its format, citations, and bibliographic style are consistent and acceptable; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the supervisory committee and is ready for submission to The Graduate School.

Dana Carroll
Chair: Supervisory Committee

Approved for the Major Department

Dana Carroll
Chair/Dean

Approved for the Graduate Council

David S. Chapman
Dean of The Graduate School
Zinc finger nucleases (ZFNs) are chimeric proteins composed of a DNA-binding domain comprised of several tandem Cys$_2$His$_2$ zinc fingers and a nonspecific endonuclease domain from the Type II s restriction enzyme FokI. When expressed in the nematode *Caenorhabditis elegans*, these molecules bind its DNA at a sequence specified by the zinc fingers and produce double-strand breaks (DSBs) at this target site in somatic tissues. These breaks activate the nematode DNA repair mechanisms, which attempt to repair the lesions either by homologous recombination (HR), using an unbroken homologous DNA template, or by nonhomologous end joining (NHEJ), by which the broken ends are simply rejoined without regard for sequence conservation. The latter process will, at some frequency, produce targeted mutations at the break site.

The work presented below characterizes ZFN-mediated mutagenesis in the somatic tissue of the nematode. Expression of sets of ZFNs, designed to specifically cleave either a synthetic site on an extrachromosomal array or a genomic target, resulted in a spectrum of NHEJ-mediated target site mutations at a frequency of about 20% at each locus. Elimination of the canonical NHEJ pathway in the nematode significantly altered the frequency and types of mutations observed, indicating that this process is fundamental in repairing ZFN-induced DSBs. It is hoped that the information gained from these studies will serve as a guide in the use of ZFNs for germline mutagenesis and gene targeting in the nematode. Use of this technology
together with an introduced altered DNA template to direct HR-mediated DSB repair
will herald the arrival of a long-sought method of efficient gene targeting for C.
elegans.
To my wife,

Angie Morton
# TABLE OF CONTENTS

ABSTRACT ............................................................................................................... iv

LIST OF FIGURES .................................................................................................... ix

LIST OF TABLES ...................................................................................................... xi

ACKNOWLEDGMENTS .......................................................................................... xii

INTRODUCTION ........................................................................................................ 1

References ................................................................................................................ 7

Chapter

1. ESSENTIALS OF ZFN-MEDIATED GENE TARGETING ................................. 10

   Part I: Gene targeting strategies ................................................................. 10
   - Gene targeting in other model organisms ................................................ 10
   - Other gene targeting strategies ............................................................... 19
   - Prior targeting attempts in *C. elegans* .................................................... 20

   Part II: Zinc finger nucleases ................................................................. 23
   - Zinc fingers .............................................................................................. 23
   - Zinc fingers in artificial transcription factors ........................................... 37
   - Polydactyl zinc finger proteins ................................................................. 40
   - *FokI* restriction enzyme ........................................................................ 42
   - Early ZFN studies .................................................................................... 45
   - ZFN-mediated gene targeting in other organisms .................................... 47

   Part III: Adapting ZFNs for gene targeting in nematodes ....................... 56
   - Transgenic DNA in nematodes ............................................................... 56
   - Mechanisms of DSB repair ..................................................................... 63

   References ...................................................................................................... 70

2. ZFN-INDUCED SOMATIC CLEAVAGE AND REPAIR ................................. 90

   Introduction .................................................................................................... 90
   Materials and methods .............................................................................. 91
   Results ......................................................................................................... 93
### 3. THE ROLE OF THE NHEJ MACHINERY IN ZFN-INDUCED DSB REPAIR... 118

- **Introduction** ................................................................. 118
- **Materials and methods** ................................................... 119
- **Results** ........................................................................... 120
- **Discussion** ..................................................................... 127
- **Current model of DSB repair** .......................................... 136
- **Appendix** ........................................................................ 139
- **References** ..................................................................... 144

### 4. TOWARD GERMLINE ZFN EXPRESSION ................................. 146

- **Introduction** ..................................................................... 146
- **Materials and methods** ................................................... 147
- **Results and discussion** .................................................... 151
- **References** ..................................................................... 165

**CONCLUDING REMARKS** ...................................................... 168
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>Gene targeting in yeast</td>
<td>12</td>
</tr>
<tr>
<td>1-2</td>
<td>Gene targeting in mouse ES cells</td>
<td>14</td>
</tr>
<tr>
<td>1-3</td>
<td>Gene targeting in flies</td>
<td>17</td>
</tr>
<tr>
<td>1-4</td>
<td>Structure of the zinc finger motif</td>
<td>25</td>
</tr>
<tr>
<td>1-5</td>
<td>Crystal structure of the Zif268 zinc fingers, bound to their DNA target</td>
<td>27</td>
</tr>
<tr>
<td>1-6</td>
<td>Zinc finger DNA-binding orientation</td>
<td>28</td>
</tr>
<tr>
<td>1-7</td>
<td>Structure of FokI bound to its DNA target</td>
<td>43</td>
</tr>
<tr>
<td>1-8</td>
<td>The ZFN QQR bound to its DNA target</td>
<td>48</td>
</tr>
<tr>
<td>1-9</td>
<td>Scheme of the yellow donor experiment</td>
<td>51</td>
</tr>
<tr>
<td>1-10</td>
<td>RNAi pathway in the nematode</td>
<td>60</td>
</tr>
<tr>
<td>1-11</td>
<td>Depiction of the NHEJ pathway</td>
<td>64</td>
</tr>
<tr>
<td>1-12</td>
<td>Scheme of the HR pathway</td>
<td>66</td>
</tr>
<tr>
<td>2-1</td>
<td>Synthetic QQR target</td>
<td>95</td>
</tr>
<tr>
<td>2-2</td>
<td>Representation of the nematode heatshock construct</td>
<td>96</td>
</tr>
<tr>
<td>2-3</td>
<td>Gel of MluI digestions of the QQR target</td>
<td>99</td>
</tr>
<tr>
<td>2-4</td>
<td>Nowhere target sequence</td>
<td>102</td>
</tr>
<tr>
<td>2-5</td>
<td>Nw ZFNs and their backbone</td>
<td>104</td>
</tr>
<tr>
<td>2-6</td>
<td>Gel of HindIII digestions of the Nowhere targets</td>
<td>108</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3-1</td>
<td>Gel of <em>MluI</em> digestions of the short QQR target from wild-type nematodes</td>
<td>121</td>
</tr>
<tr>
<td>3-2</td>
<td>Gel of <em>MluI</em> digestions of the short QQR target from <em>lig-4</em> nematodes</td>
<td>122</td>
</tr>
<tr>
<td>3-3</td>
<td>Gel of <em>HindIII</em> digestions of the Nowhere target in <em>lig-4</em> nematodes</td>
<td>126</td>
</tr>
<tr>
<td>3-4</td>
<td>Model of the ZFN-induced DSB repair pathway in <em>C. elegans</em></td>
<td>137</td>
</tr>
<tr>
<td>3-5</td>
<td>Cross scheme to create <em>cku-80</em> QQR nematodes</td>
<td>140</td>
</tr>
<tr>
<td>3-6</td>
<td>Gel of <em>MluI</em> digestions of the short QQR target in <em>cku80</em> nematodes</td>
<td>141</td>
</tr>
<tr>
<td>4-1</td>
<td>The <em>ben-1</em> target sequence</td>
<td>152</td>
</tr>
<tr>
<td>4-2</td>
<td>In vitro assays of Ben1A and B ZFN target activity</td>
<td>154</td>
</tr>
<tr>
<td>4-3</td>
<td>In vitro assays of Ben1A and B ZFN target specificity</td>
<td>157</td>
</tr>
<tr>
<td>4-4</td>
<td>Flp recombination products</td>
<td>161</td>
</tr>
<tr>
<td>4-5</td>
<td>The RFP-GFP Flp reporter</td>
<td>162</td>
</tr>
<tr>
<td>4-6</td>
<td>Flp reporter activity in nematodes</td>
<td>163</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>Proposed degenerate recognition code for zinc finger-DNA interactions</td>
</tr>
<tr>
<td>1-2</td>
<td>Zinc finger-DNA specificities</td>
</tr>
<tr>
<td>1-3</td>
<td>Targeting frequencies in Drosophila</td>
</tr>
<tr>
<td>2-1</td>
<td>QQR target mutations</td>
</tr>
<tr>
<td>2-2</td>
<td>Nowhere target homologs</td>
</tr>
<tr>
<td>2-3</td>
<td>Nowhere target mutations</td>
</tr>
<tr>
<td>3-1</td>
<td>Short QQR target mutations in wild-type nematodes</td>
</tr>
<tr>
<td>3-2</td>
<td>Short QQR target mutations in <em>lig-4</em> nematodes</td>
</tr>
<tr>
<td>3-3</td>
<td>Nowhere target mutations in <em>lig-4</em> nematodes</td>
</tr>
<tr>
<td>3-4</td>
<td>Microhomologies flanking deletions in wild-type and <em>lig-4</em> nematodes</td>
</tr>
<tr>
<td>3-5</td>
<td>Short QQR target mutations in <em>cku-80</em> nematodes</td>
</tr>
<tr>
<td>4-1</td>
<td>Alignment of the <em>ben-1, tbb-1, tbb-2,</em> and <em>mec-7</em> target sites</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

I would like to thank the members of my lab for all of their assistance during my graduate career: Kelly Beumer, Jon Trautman, and Ana Bozas provided valuable help and insight on an almost daily basis. Josh Checketts and Erin Fuchs also helped perform many of the experiments described in this dissertation. Finally, I would like to thank my advisor, Dana Carroll, for all of this support and encouragement.

Several figures in this dissertation have been reproduced, with permission, from previously published works. I would like to thank Annual Reviews, Elsevier Ltd, Nature Publishing Group, and The National Academy of Sciences of the United States of America for their kind cooperation.

Finally much of this work was supported by the National Institutes of Health Genetics Training Grant T32 GM07464, for which I am most grateful. I would also like to thank David Stillman, administrator of the grant, for his insight, suggestions, and encouragement.
INTRODUCTION

Since the first sequences of genes were determined, a central question among geneticists has been how an intentional change in a gene will affect its function. This is, perhaps, one of the most fundamental questions in all genetics, which seeks to discover what genes do by manipulating them. The enormity of such a task can be best exemplified by examining the genome sizes of different organisms. Among popular model organisms, the *Saccharomyces cerevisiae* yeast genome contains 6000 genes in 12 million base pairs (bp) (Goffeau et al., 1996), the *Caenorhabditis elegans* nematode genome has more than 19,000 genes in 100 million bp (1998; Hillier et al., 2005), the *Drosophila melanogaster* fruitfly genome has 14,000 genes in 180 million bp (Adams et al., 2000), the mouse has approximately 20,000 genes in 2.5 billion bp (Waterston et al., 2002), and the human genome contains 20,000 – 25,000 genes within its 3 billion bp (2004; Venter et al., 2001). Within these vast genomes, the exact sequence of the particular gene to be manipulated must be located and specifically modified without damaging other regions of the genome, which could obscure any phenotypic changes arising from alterations to the target DNA sequence.

One of the first real advances in the ability to manipulate DNA came with the discovery of restriction enzymes, which could recognize and cleave specific DNA sequences (Smith and Wilcox, 1970). However, the size of the genome and the inflexibility of the sequences recognized by the enzymes still left the manipulation of
genes in almost all organisms out of reach. Even when genes were removed from the genome and placed in plasmids, engineering specific changes in their sequence was laborious (Hershfield et al., 1974). Furthermore, when genes were successfully modified, it was still necessary to return them to their genomic locus. This became a great challenge.

In mammalian cells, when a plasmid containing a modified gene is introduced into a cell, it will either remain unincorporated and be quickly lost, or it can be integrated randomly into the genome (Folger et al., 1982; Gordon and Ruddle, 1981; Robins et al., 1981; Tikchonenko et al., 1981). When integration does occur, the resulting improper genomic context of the modified gene will almost always result in misregulated expression, since it will now be driven by foreign regulatory elements or silenced by surrounding chromatin. Additionally, unless the genomic copy of this gene has been deleted, expression of the inserted gene may be entirely masked by wild-type gene expression. To accurately examine the consequences of altering a gene, it is necessary to either alter it in vivo or find a method of introducing an altered copy of the gene specifically to its genomic locus.

A process by which an altered gene on a plasmid could be inserted into the correct genomic locus was first discovered in yeast (Hinnen et al., 1978). Although the mechanism by which this event occurred was poorly understood, it was assumed that the gene replacement was the result of DNA recombination between the two copies of the gene. Later yeast studies showed that simply flanking an altered gene with sequences corresponding to those surrounding it at its genomic locus was sufficient to prompt its homologous integration into the genome at a relatively high frequency.
Using this technique, any gene in the yeast genome could now, in principle, be knocked out by simple deletion, or altered copies of a gene could be inserted into the genome at the correct locus (Huh et al., 2003; Winzeler et al., 1999). Thus was born the era of gene targeting.

Early yeast targeting studies demonstrated that linearized plasmids recombined with their target at a much higher frequency than circular molecules did (Orr-Weaver et al., 1981). An important clue to the mechanism by which these altered genes were inserted into the genome arose during studies of meiotic recombination, when it was noted that DNA double-strand breaks (DSBs) initiated crossovers between homologous chromosomes (Kolodkin et al., 1986). When it became clear that these DSBs stimulated a process known as homologous recombination (HR), by which information from one molecule of DNA could be copied to a homologous site on a separate piece of DNA, techniques of manipulating genes in other organisms were given additional direction.

After the first successful demonstration of gene targeting in yeast, work intensified in finding an analogous technique in other organisms. However, in all other organisms examined, simple flanking homology proved insufficient for homologous integration of the altered genes. Even when flanked by extensive homologies, most introduced genes were still either lost or integrated at random points in the genome (Folger et al., 1982). Such integration was not homology-dependent but, instead, seemed driven by an alternative and competing repair process known as nonhomologous end joining (NHEJ), in which broken DNA ends are simply rejoined without regard to homology (Jackson, 2002). Even if a gene was occasionally
integrated at its correct genomic locus, the difficulty of distinguishing these events from
the accompanying random insertions precluded the possibility of identifying these rare
events. Clearly, additional tools were needed for this technique to be successful in
higher organisms.

Acting on the assumption that introduced genes were, at some frequency, integrated by HR even in complex genomes, Mario Capecchi developed powerful selective processes to identify the rare integrants in mouse embryonic stem cells (Mansour et al., 1988; Thomas and Capecchi, 1987). The triumph of this selection process was manifest by the first successful gene targeting experiments in mice (Thompson et al., 1989). Several years later, Kent Golic was also able to introduce altered genes into the genome of Drosophila by HR, using a linearized template sequence that had been previously placed in the genome on a transposon (Rong and Golic, 2000).

Employing the procedures developed for mice and flies, gene targeting by HR is an extremely rare event, and powerful selection and/or rigorous screening are needed to identify true targeting events. The utility of gene targeting makes further development of the technology highly desirable, so that additional targeted alterations in these organisms can be more easily obtained. Furthermore, these successes may indicate that the introduction of altered genes into the genomes of many organisms may occur by HR at a very low frequency. However, for many popular experimental organisms, including the zebrafish Danio rerio and the nematode Caenorhabditis elegans, no good technique for replacing genes at their genomic loci exists. New strategies must be
developed to increase the efficiency of gene targeting, so that an effective technique can be found to target genes in these organisms.

Recently, a novel molecular tool has been developed which can aid in the development of efficient gene targeting strategies. Zinc finger nucleases (ZFNs) were first described a decade ago (Kim et al., 1996). These engineered proteins, composed of modifiable zinc finger DNA-binding domains coupled to a nuclease domain, have been shown capable of cleaving specific and arbitrary genomic targets in vivo (Bibikova et al., 2003; Bibikova et al., 2002; Urnov et al., 2005). The resulting chromosomal DSBs are then repaired by the cell, either by NHEJ to produce random mutations in the targeted gene or by HR using an altered gene on a template molecule.

This technique has several advantages over traditional targeting strategies: First, since cells can be easily transformed to express ZFNs, many gene targeting experiments can be performed entirely in vivo. Second, since the ZFN-mediated DSBs stimulate the cellular repair machinery and convene it at a specific site, ZFNs can produce targeted events at much higher frequencies in flies and in mammalian cells than traditional targeting techniques. Third, ZFNs appear to be active in all organisms in which they have been tested. These characteristics indicate that ZFNs may herald the advent of a new technology, adept at specifically and efficiently altering the sequence of arbitrary genes in many different organisms.

This dissertation describes the application of ZFN technology to the nematode Caenorhabditis elegans. The first chapter introduces the challenges that must be overcome to employ ZFNs to target genes in the nematode. It discusses previous attempts at targeting genes in various model organisms, including the nematode. It
summarizes the discovery of the zinc finger motif, the subsequent development of zinc finger technology, and the invention and successful use of the ZFN in gene targeting experiments. Finally, it reviews what is known about DSB repair in the nematode and discusses how the nematode cellular machinery can be used to express ZFNs and repair the resultant DSBs to generate mutations.

The subsequent chapters describe the application of ZFN technology to generate targeted mutations in the nematode. Chapter 2 details how the ZFNs have been introduced in the nematode and activated to stimulate cleavage of their intended target site DNA in somatic tissues. It also describes the mutations resulting from NHEJ-mediated repair of these DSBs. In Chapter 3, the repair products of targeted DSBs induced in nematodes lacking various NHEJ components are examined, and the apparent function of these components in DSB repair is addressed. Finally, Chapter 4 of this dissertation discusses the first steps taken in activating ZFNs in the germline to induce mutations in an arbitrary gene among the progeny of a targeted nematode. When coupled with repair by an altered template molecule, ZFN-induced DSBs in the germline should result in a viable gene targeting technique in the nematode *C. elegans*. 
References


CHAPTER 1

ESSENTIALS OF ZFN-MEDIATED GENE TARGETING

Part I: Gene targeting strategies

Gene targeting in other model organisms

Gene targeting may occur by roughly the same mechanism, albeit at differing frequencies, in the three model organisms in which it has been demonstrated. In yeast, introduced genes are incorporated into the homologous locus at a relatively high frequency, so that HR integration predominates over NHEJ-based integration. The ease with which HR-based integration occurs may be a consequence of the less functional NHEJ pathway in yeast (Daley et al., 2005b).

Early studies showed that yeast transformed by nonreplicative plasmids either lost the plasmids or integrated them into the genome, often at a chromosomal locus corresponding to yeast genes contained on the plasmid (Hinnen et al., 1978). Subsequent experiments showed that these plasmid sequences were incorporated into the yeast genome at their target by HR (Scherer and Davis, 1979). Furthermore, the frequency of this targeted integration could be increased by linearizing the plasmid prior to transformation (Orr-Weaver et al., 1981). If the gene was linked to a selectable marker, simple selection of transgenic organisms was then sufficient to identify those
carrying the altered gene at the proper genomic locus (Rothstein, 1991; Rothstein, 1983).

Accordingly, before being introduced into yeast, the gene being studied would be ligated into a plasmid, adjacent to a selectable marker and flanked by short homologous sequences which guided its integration (Baudin et al., 1993). Later, it was found that altered genes could simply be amplified by PCR, using primers containing ~50 bp of homology to the intended genomic target (Baudin et al., 1993; Gray and Honigberg, 2001; Lorenz et al., 1995; Manivasakam et al., 1995). Eventually, PCR-based modules were developed which allowed any gene for which primers could be synthesized to be deleted, tagged, or overexpressed in vivo (Longtine et al., 1998; Wach et al., 1994) (Fig. 1-1). This process has become so routine that almost every gene in the yeast genome has now been knocked out, and most genes have also been tagged for expression studies (Huh et al., 2003; Winzeler et al., 1999).

These advances exemplify the complete success of gene targeting in yeast. Unfortunately, application of this process in higher organisms has been more difficult to achieve. When altered genes are introduced into mammalian cells, they are almost always integrated into the chromosomes via NHEJ, even when flanked with extensive sequences homologous to those surrounding the targeted gene (Folger et al., 1982). In the rare event that an altered gene is recombined into the proper locus, the difficulty in screening for this occurrence from the hundreds or thousands of accompanying NHEJ-based insertions makes its identification almost impossible.
Any yeast gene can be easily targeted and tagged using a PCR product in which the terminal sequences (blue) contain ~50 bp of homology with the genomic locus. These sequences can be synthesized as the 5' portion of primers designed to specifically amplify the plasmid-based marker and altered gene to be inserted into the yeast genome. Yeast can then be transformed by the linear PCR product, which will be incorporated into the genome at the proper locus by the yeast recombination machinery (Huh et al., 2003).

Figure 1-1: Gene targeting in yeast.
To overcome this obstacle, during early targeting experiments in mice the altered gene was not only inserted with a selectable marker, such as neomycin resistance (neo\(^R\)), and flanked by homologous DNA, but a counterselectable marker, such as herpes simplex virus thymidine kinase (HSV-tk), was placed outside this region of homology (Capecchi, 1989; Capecchi, 2005; Mansour et al., 1988) (Fig. 1-2). In this case, when the altered gene is inserted into the genome via NHEJ, the HSV-tk is integrated with the construct. Selection with neomycin and an antiviral nucleoside analogue such as gancyclovir or, in more recent experiments, fialuridine (FIAU) will kill the transgenic cell.

When the altered gene is incorporated via HR, HSV-tk will not be integrated into the genome, since it is outside the region copied by the recombination machinery. Now the cell is both neomycin-resistant and FIAU-resistant and will survive to be identified as a targeted cell. At this point, the neomycin resistance gene, which may alter expression of the gene, can be removed if it is flanked by \(\text{loxP}\) sequences. Expression of the bacteriophage Cre recombinase will catalyze circularization of all DNA between these \(\text{loxP}\) repeats and subsequent loss of this molecule (Abuin and Bradley, 1996; Araki et al., 1995; Sauer and Henderson, 1989). The resulting altered gene then carries only one residual recombined \(\text{loxP}\) site, which is unlikely to dramatically affect its expression (Abuin and Bradley, 1996).

When these manipulations are performed on isolated embryonic stem (ES) cells, they can then be incorporated into a developing mouse embryo, which, after implantation into a surrogate mother, will produce a transgenic mosaic mouse (Capecchi, 1989; Thompson et al., 1989). If all goes well and some proportion of this
The targeting construct contains a neomycin resistance (Neo\(^R\)) gene, flanked by several kilobases of sequence homologous to the locus to be targeted. A herpes simplex virus thymidine kinase (HSV-tk) gene is attached to one terminus of this cassette. When introduced into the cell, this construct can replace the targeted gene by homologous recombination. In this case, the terminal HSV-tk gene is not introduced into the targeted locus and is lost. Such cells will become resistant to both neomycin and FIAU. Alternatively, targeting constructs are frequently randomly integrated into cell lines by NHEJ. In this case, the terminal HSV-tk gene is also incorporated into the chromosome, making cells neomycin-resistant, but sensitive to FIAU (Capecchi, 2005).
mouse's germline is also transgenic, some F1 mice will be heterozygous for this altered
gene at the correct locus. Inbreeding of these mice will produce homozygous F2
offspring containing the modified target gene.

Even though homologous integration represents only a small fraction of
integrations into the genome, the efficiency of this targeting strategy is dependent on
the amount of homology between the target site and the sequences flanking the gene to
be inserted (Deng and Capecchi, 1992; Deng et al., 1993; Thomas et al., 1992). The
greater the extent of homology between these two sites, the more efficiently the target
can be replaced by the modified gene. This contrasts sharply with observations in
yeast, which indicate that extended homology is not a significant factor in HR-mediated
transgene integration (Lorenz et al., 1995; Manivasakam et al., 1995).

A similar scheme was utilized in flies to target genes. In this organism, gene
targeting experiments had been initially hampered by the inability of introduced foreign
DNA to integrate into the chromosome by any mechanism (Fox et al., 1970; Fox and
Yoon, 1966). Luckily, studies of transposons, known as P-elements, gave researchers
an alternative means of randomly integrating foreign DNA into the fly genome
(Spradling and Rubin, 1982). In the presence of transposase, the P-elements can hop in
and out of the genome at random, moving via a cut-and-paste mechanism (Beall and
Rio, 1997; Kaufman and Rio, 1992). The P-element normally encodes its own
transposase, so its expression catalyzes its eventual excision and insertion into a new
genomic location (Kaufman et al., 1989; Kaufman and Rio, 1992; Rio et al., 1986).
However, it was discovered that, when an alternate source of transposase was supplied,
transposons in which the transposase has been replaced with an exogenous sequence could still be prompted to hop (Robertson et al., 1988).

Additionally, when a P-element hops out of the genome, the DSB it creates is usually repaired using the homolog (Engels et al., 1990; Gloor et al., 1991) but can occasionally be repaired from a homologous sequence present on a P-element at some other location (Engels et al., 1994; Johnson-Schlitz and Engels, 1993). Furthermore, later studies determined that DSBs induced by transposon excision could be repaired by homologous sequences from injected plasmids (Keeler et al., 1996). This body of knowledge eventually led to the development of a viable gene targeting strategy.

A P-element was created which contained an altered gene and its flanking DNA along with a visible marker (Rong and Golic, 2000). This entire sequence was surrounded by two FRT sites, which can be recombined by yeast recombinase Flp in a reaction analogous to that of the loxP-Cre recombination in mice (Broach et al., 1982). This construct was introduced into the fly genome and subsequently activated in transgenic animals by Flp expression from a different immobile P-element. Recombination of the FRT sites removed all of the altered gene and its flanking DNA from the P-element and circularized it within the nucleus.

The resulting circular molecule acts much like a plasmid and should be able to recombine into a homologous chromosomal target when activated for recombination by a DSB (Keeler et al., 1996). This circular template was stimulated for recombination with its genomic target by cleavage at an engineered I-SceI site within the modified gene (Fig. 1-3) or (as shown in subsequent experiments) near the remaining FRT site opposite this gene (Gong and Golic, 2003). I-SceI is a homing endonuclease, also
Targeting construct

\[ \text{P-element insertion site} \quad \text{Donor gene (5' and 3' frag.)} \quad \text{Marker} \quad \text{P-element insertion site} \]

**Flp and I-SceI expression**

**Target locus**

**Targeted gene**

Target gene at its chromosomal locus

Homologous recombination

Target gene, interrupted by the integrated targeting construct.

Figure 1-3: Gene targeting in flies.

A P-element containing an altered gene, an internal I-SceI site, and an identifiable marker, all of which are flanked by FRT sites, is hopped into a fly chromosome. Subsequent expression of Flp and I-SceI removes this targeting cassette from its P-element insertion, then linearizes it to activate it for recombination with a target gene. It can then be incorporated into the genome by homologous recombination (Rong et al., 2002).
commonly called a meganuclease, since it recognizes an 18-bp sequence instead of the 5-6 bp sequences typically recognized by most restriction enzymes (Plessis et al., 1992). The resulting DSB linearized the circular DNA and allowed the now linear molecule to replace its genomic target by HR at a low frequency (Rong and Golic, 2000; Rong and Golic, 2001; Rong et al., 2002). When this happened in the germ tissues, some flies in the subsequent generation will be heterozygous for the altered gene. Identification and crossing of these flies produced offspring homozygous for the targeted gene.

Although the specifics vary, the general principle behind the successful targeting techniques developed in mice and flies is identical: Long stretches of homologous sequence between template and target will, at some frequency, direct the HR mechanisms of the cell to incorporate the altered gene into the correct genomic locus. The frequencies at which genes are targeted vary between these two methods. In mice, it is estimated that homologous integration occurs less than once per $1 \times 10^6$ ES cells (Capecchi, 1989; Mansour et al., 1988). In flies, more than 20,000 offspring may need to be screened to identify a mutant (Rong and Golic, 2001), although targeting events can also occur much more frequently, depending on the locus (Rong et al., 2002). In any case, however, it is clear that targeting frequencies in these organisms are often very low and that a more efficient targeting technique would be quite valuable. Since these gene targeting strategies rely on HR to guide the proper integration of a modified sequence into the genome, any process which increases HR in a cell may also increase the efficiency of these targeting strategies.
Other gene targeting strategies

Studies have indicated that HR is a primary cellular mechanism used to repair different DNA lesions, the most serious of which is the DSB (Pastink et al., 2001). It follows, then, that induction of DSBs will stimulate HR in cells (Choulika et al., 1995; Taghian and Nickoloff, 1997). Several cellular processes naturally produce DSBs, such as VDJ rearrangement (Landree et al., 1999) and crossing over during meiosis (Bassing and Alt, 2004). Additionally, any event that damages DNA, such as some types of chemical modification or ionizing radiation, can create DSBs at high frequency (Bradley and Taylor, 1981; Dugle et al., 1976). In some cases, reagents have been characterized which can even induce DSBs at a specific genomic site.

Transposons have been shown repeatedly in several organisms to create single genomic DSBs upon excision (Beall and Rio, 1997; Ding et al., 2005; Emmons et al., 1986; Engels et al., 1990; Plasterk, 1991). Homing endonucleases, such as I-SceI and the yeast mating factor HO, also produce DSBs at their recognition sites (Plessis et al., 1992; Rudin and Haber, 1988). In constructs engineered to contain transposons or nuclease recognition sites, DNA repair by HR is highly stimulated after induction of the DSB (Gloor et al., 1991; Rouet et al., 1994a; Rouet et al., 1994b). The disadvantage of using these tools together with a gene targeting strategy is that the gene to be targeted must either naturally contain such a site or must be manipulated previously to acquire a site. Such prior manipulation generally defeats the purpose of gene targeting in the first place. At present, gene targeting remains largely inefficient due to this obstacle.

Several different synthetic systems for inducing targeted DSBs have been developed over the past decade, with some success (Vasquez et al., 2001).
Oligonucleotides can be synthesized that will recognize long pyrimidine tracts in DNA and bind to these sites, forming DNA triplexes. When these oligonucleotides are covalently linked to psoralen and exposed to UV light, they form adducts, which act as mutagenic hotspots since they stimulate the DNA repair machinery (Faruqi et al., 2000; Faruqi et al., 1996; Segal et al., 1997; Vasquez et al., 2000; Wang et al., 1995). Although these oligonucleotides have been shown to induce specific mutations at their targets at a low frequency, this system has the major disadvantage of only being able to target pyrimidine tracts in a genome, so it has not been extensively used. Alternative systems employ pyrole-imidazole polyamines and modified peptide amino acids in much the same scheme (Demidov et al., 1993; Kaihatsu et al., 2004; Wurtz and Dervan, 2000). Due to the difficulty of synthesizing these molecules and employing them to target arbitrary chromosomal sequences, these systems have also not been widely employed. Finally, there has been some recent effort in modifying the DNA recognition residues in some of the homing endonucleases, like I-SceI, to change the DNA sequence to which it will bind and cleave (Arnould et al., 2006; Chames et al., 2005; Epinat et al., 2003). These studies are in their infancy but may expand the use of these enzymes in gene targeting, although they will probably never be able to target arbitrary sequences.

Prior targeting attempts in C. elegans

Although these studies have increased the activity of the HR machinery at engineered sites in model organisms, the efficiency of true gene targeting in mice and flies still remains very low. However, at least some method of producing targeted
mutations exists for these organisms. In contrast, no reliable system of gene targeting exists for the nematode *Caenorhabditis elegans*. Early nematode experiments, based on the successes enjoyed in flies and mice, attempted to target genes based on their homology. In nematodes, however, introduced genes will not integrate into the chromosomes, but instead form long concatamers, which then replicate as semistable extrachromosomal arrays (Stinchcomb et al., 1985). Since these arrays form primarily by homologous recombination between injected DNA molecules, altered genes included in them are typically not integrated into the corresponding chromosomal locus by HR (Mello et al., 1991). Subsequent gene targeting experiments sought to overcome this challenge.

Initially, studies of the nematode Tc1 transposon indicated that after a transposon-induced DSB at a genomic locus, altered genes from an array could, at extremely low frequency, undergo HR with their genomic homolog to produce the desired changes in the nematode genome (Plasterk and Groenen, 1992). In an attempt to generalize this mechanism of gene targeting, these transposons were jumped randomly throughout the nematode genome in the attempt to characterize transposon insertions in as many genes as possible. It was planned that, in future studies on any of these genes, the transposon could be hopped from this locus to create specific DSBs, which could then be repaired from a modified template either on an array (Plasterk, 1992) or on another homologous transposon (Fischer et al., 2003). Work on the creation of this and similar types of transposon insertion libraries is ongoing (Barrett et al., 2004; Granger et al., 2004; Zwaal et al., 1993).
Alternately, transgenes on an array could be induced to recombine with a homologous genomic target by exposing the transgenic nematodes to high doses (6000 rad) of gamma-radiation (Stuart K. Kim, personal communication). Although the frequency of homologous integration of a single copy of the transgene was quite low ($5 \times 10^{-5}$), this technique represented another possible strategy of gene targeting.

Other research focused on techniques whereby the DNA destined for incorporation in an extrachromosomal array could be directed to a chromosomal homologous target. In one early study, a tRNA sequence encoding an amber mutation, lethal in high concentrations, was included during the injection of transgenic DNA to discourage the formation of arrays, in the hopes that introduced DNA would, instead, integrate into the chromosomes – possibly even at the correct locus (Fire, 1986). This technique eventually produced two nematode lines in which an altered transgene had replaced its genomic copy by homologous recombination (Broverman et al., 1993). Although this success was greeted with a great deal of enthusiasm, it could not be repeated consistently. Other studies demonstrated that when random single-stranded oligonucleotides were included in standard nematode injections, in addition to the extrachromosomal arrays normally seen, a low frequency of randomly integrated transgenes was repeatedly observed, although no instances of homologous integration using this process have been documented (Mello et al., 1991).

Although these initial experiments have produced some encouraging results, gene targeting in nematodes remained notoriously inefficient and somewhat unpredictable. After the discovery of RNA interference (RNAi), it seemed most reasonable and simple to transiently knock out or knock down expression of a gene of
interest (Agrawal et al., 2003; Fire et al., 1998; Grishok, 2005; Kamath et al., 2001; Montgomery et al., 1998). For this reason, research into a technique to specifically and permanently target genes in nematodes has slowed in recent years.

Recently, however, a new procedure, known as biolistic transformation, has been used to integrate transgenes into the genome and, in some cases, produce targeted nematode mutants (Berezikov et al., 2004; Jantsch et al., 2004; Praitis et al., 2001; Wilm et al., 1999). Modified plasmid DNA is linearized and mixed with tiny gold beads, which are then fired by pressurized helium into a plate containing thousands of nematodes. Some beads penetrate the nematodes, possibly producing genomic DSBs, and release their DNA. Since it is present in a very low concentration, the DNA can sometimes be integrated as a low-copy insertion in a chromosome. Although integration occurs infrequently, large numbers of nematodes can be bombarded, and a marker can be included to aid in the identification of transformants (Wilm et al., 1999). Usually, the foreign DNA integrates randomly, but there have been reports of integration by HR (Jantsch et al., 2004). Currently, its main use lies in simply integrating a transgene in low copy so it will not be silenced in the germline, as it would in an array (Dernburg et al., 2000). Because of this limited success, a new and more efficient targeting technique is eagerly anticipated among nematode researchers.

Part II: Zinc finger nucleases

Zinc fingers

The ZFN is a chimeric protein composed of two domains: a Cys$_2$His$_2$ zinc finger DNA-binding domain and a nonspecific endonuclease domain. The Cys$_2$His$_2$ zinc
finger is a naturally occurring protein motif and is common throughout the genome of most eukaryotes (Rubin et al., 2000). The first identified Cys\textsubscript{2}His\textsubscript{2} zinc finger was characterized from a transcription factor in *Xenopus laevis* called TFIIIA (Miller et al., 1985). This transcription factor contains nine zinc fingers, which act cooperatively to bind both DNA and RNA. Subsequently, it was discovered that this protein motif was the most common DNA-binding motif in all metazoa and, in fact, is the most abundant DNA-binding motif in the human genome (Tupler et al., 2001).

A Cys\textsubscript{2}His\textsubscript{2} zinc finger is composed of approximately 30 amino acids, running in two antiparallel β sheets followed by an extended α helix, and has the consensus amino acid sequence X\textsubscript{2}–C–X\textsubscript{2,4}–C–X\textsubscript{12}–H–X\textsubscript{3,5}–H (Pabo et al., 2001). The structure of the motif is stabilized by hydrophobic interactions and a coordinated zinc ion, which interacts with the two cysteine residues near the turn in the β sheet and the two histidine residues found at the C-terminus of the α helix (Berg, 1988; Lee et al., 1989; Pavletich and Pabo, 1991) (Fig. 1-4.) These contacts position the amino acids in the N-terminal portion of the α helix in such a manner that they can slide into the major groove of a DNA molecule, allowing each finger to make contact primarily with three consecutive bases of DNA (Nardelli et al., 1991). In this orientation, the amino acids at critical points within the α helix are able to confer specificity for a particular DNA triplet.

Through this interaction, zinc fingers are able to bind DNA in a sequence-specific manner. Typically, a single zinc finger cannot bind to DNA with sufficient affinity to be biologically useful, so most naturally occurring proteins with this motif contain at least three zinc fingers acting in tandem to bind sequential DNA triplets. Hundreds of zinc fingers, often in transcription factors, have been identified.
Figure 1-4: Structure of the zinc finger motif.

The specificity determining residues (-1, 2, 3, 6, numbered with respect to the start of the α-helix) are depicted extending from the protein backbone in blue. The two histidine (red) and two cysteine (yellow) residues are shown coordinating the zinc (grey) ion. Figure modified from (Pabo et al., 2001), with permission from Annual Reviews.
Although the first zinc fingers were identified in *Xenopus laevis*, the best characterized come from the mouse. Screens designed to search mouse cells for transcriptional regulators identified a cDNA clone, named Zif268, which contained three tandem zinc finger repeats, indicative of a transcription factor (Christy et al., 1988). The crystal structure of the zinc fingers from Zif268, bound to their target DNA sequence, was subsequently solved, and the modular nature of zinc finger-DNA interaction became apparent (Elrod-Erickson et al., 1996; Pavletich and Pabo, 1991) (Fig. 1-5).

Each of the zinc fingers interacts primarily with a DNA triplet on the same strand of the cocrystallized DNA target sequence, and the zinc finger residues that play the greatest role in DNA binding were identified. These residues are found exclusively along the N-terminal region of the α helix of the zinc finger, in positions -1, 3, and 6 (numbered with respect to the beginning of the α helix) (Pavletich and Pabo, 1991). It is important to note that zinc fingers bind DNA in a 3'-5' direction. For a given finger, the residue at position -1 binds the last base in a DNA triplet, read in the standard 5'-3' direction, while the residue at position 6 binds the first base in the triplet (Fig. 1-6).

There is also evidence that the amino acid at position 2 can interact with both the amino acid at residue -1 and with the base immediately downstream and on the complementary DNA strand from the base bound by residue -1. This is a nonmodular interaction, since the complement of this base must interact with residue 6 of a previous zinc finger when several zinc fingers are linked in tandem (see Fig. 1-6).

This interaction is not universal, being largely restricted to situations in which the amino acid residue 2 is an aspartate, and the amino acid at residue -1 is arginine.
Figure 1-5: Crystal structure of the Zif268 zinc fingers, bound to their DNA target.

Finger 1 is shown in red, finger 2 in yellow, and finger 3 in purple. The coordinating zinc ions are shown as silver spheres. The α-helix of each finger can dock in the major groove of the DNA (blue), creating modular sequence-specific contacts between the molecules. Figure from (Elrod-Erickson et al., 1996), with permission from Elsevier Ltd.
The primary strand of the Zif268 DNA target extends 5'-3'. For all of the fingers, residue -1 of the α-helix binds to the 3' nucleotide of the corresponding DNA triplet. Residue 2 binds to the residue immediately upstream from this triplet and on the secondary strand. Residue 3 binds to the central nucleotide of the triplet. Residue 6 binds to the 5' nucleotide of the triplet. (The crystal structure of the Zif268-DNA complex indicates that residue 6 of finger 2 does not contact the DNA. This non-interaction is not typical of all zinc fingers.) Nucleotides that interact directly with Zif268 are in black, while others are in grey.
(Elrod-Erickson et al., 1996). However, it does affect the potential modularity of zinc finger binding in two ways: First, the aspartate at residue 2 strengthens the hydrogen bond between arginine at residue -1 and the DNA base (typically a guanine, as will be discussed later) with which it interacts. Second, studies on different zinc fingers have indicated that the aspartate on residue 2 can bind preferentially with either a cytosine or an adenine in the complementary DNA strand. In this case, zinc fingers with an arginine at residue -1 and an aspartate at residue 2 may have an exceptionally strong affinity for DNA targets of the form 5' NNGG 3' or 5' NNGT 3'(Pabo et al., 2001; Segal et al., 1999). This phenomenon, known as target site overlap, has shaped the strategies used for the selection of new zinc finger sequences with altered DNA-binding specificities. However, this nonmodular interaction appears to be limited to the specific residues described above. For this reason, as well as the fact that this interaction has only modest effects on the affinity of the triplet to the zinc finger, its importance has been minimized in many applications of the zinc fingers (Segal et al., 2003).

After these crystal studies indicated that each zinc finger binds a specific DNA triplet in a largely modular fashion, it became clear that, if modular zinc fingers could be designed to bind specific triplets, these fingers could then be linked together to create novel DNA-binding proteins that would recognize arbitrary DNA sequences. Many subsequent experiments were conducted to identify the amino acid-nucleotide combinations which lead to efficient DNA binding (Desjarlais and Berg, 1992; Desjarlais and Berg, 1993). Using consensus information from Zif268, the recently characterized zinc finger-containing human transcription factor Sp1 (Kadonaga et al., 1987), and a database of additional zinc finger sequences (Berg, 1990; Jacobs, 1992),
early studies attempted to alter the DNA affinity of newly constructed zinc fingers by systematically changing amino acids at the key -1, 3, and 6 positions in the α helices. In many instances, however, it also proved necessary to change the identity of neighboring residues to improve the specificity of a zinc finger created after mutation at a key residue. Even after these changes, many of the resulting zinc finger proteins did not bind to their proposed target sequences with high specificity or great affinity.

Due to these types of unpredictable context effects, no clear-cut code has emerged from these studies, although there is some evidence that a redundant code exists from which general binding trends can be drawn. For example, an arginine at residue 6 in the α helix will often bind most strongly to a guanine in the 5’ position of the target site. When residue 3 in the zinc finger α helix is histidine, it will often bind guanine in the central position of the target site. If it is asparagine it seems to preferentially bind adenine at this position, and if it is aspartate, it may bind cytosine. Likewise, at residue -1 on the α helix, arginine seems, once again, to favor guanine binding (especially when aspartate is located at position 2) and aspartate often binds to cytosine. Also, in this context, glutamine often binds with adenine. A complete description of the proposed recognition code is presented in Table 1-1 (Choo and Klug, 1994a; Desjarlais and Berg, 1992; Pabo et al., 2001; Wolfe et al., 2000)

As it became increasingly clear that there was no code to predict which zinc finger sequences could be created to specifically and tightly bind to different DNA triplets, new strategies were devised to characterize novel zinc fingers of desired specificity and high affinity. First among these was phage display (Scott and Smith, 1990; Smith, 1985). The DNA sequences encoding the specificity-determining α-
Table 1-1: Proposed degenerate recognition code for zinc finger-DNA interactions.

<table>
<thead>
<tr>
<th>Nucleotide at target site</th>
<th>6 (5' nt)</th>
<th>3 (middle nt)</th>
<th>2 (3' nt*)</th>
<th>-1 (3' nt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Q</td>
<td>N/H</td>
<td>D/N</td>
<td>Q</td>
</tr>
<tr>
<td>C</td>
<td>E</td>
<td>D/T/E/S</td>
<td>D/E</td>
<td>D/H/E</td>
</tr>
<tr>
<td>G</td>
<td>R</td>
<td>H/K</td>
<td>S</td>
<td>R</td>
</tr>
<tr>
<td>T</td>
<td>K/T</td>
<td>S/A</td>
<td>T</td>
<td>T/L/H</td>
</tr>
</tbody>
</table>

The α-helical residues -1, 2, 3, and 6 have been shown to be most critical in determining the specificity of the DNA target site. For a given nucleotide, residues at these positions are likely to have the indicated identities. The most commonly observed amino acid at a given position is in bold.

The residue at position 2 also interacts with the nucleotide on the opposite DNA strand located just upstream from the 3' target nucleotide. In this context, the residue at position 2 can also influence the specificity of an N-terminal zinc finger. Table information compiled from (Sera and Uranga, 2002; Wolfe et al., 2000).
helical residues -1, 2, 3, and 6 in the first finger of the well-characterized zinc finger protein Zif268 were randomized, and the modified Zif268 constructs were cloned into the pIII gene of an M13 filamentous phage vector (Jamieson et al., 1994; Rebar and Pabo, 1994). Since the protein product of the pIII gene is expressed on the surface of the phage, these phage produced several copies of the modified Zif268 protein, containing two wild-type fingers and a randomly altered finger.

These phage were then mixed with oligonucleotides containing the wild-type binding sites for the second and third Zif268 fingers and several altered sequences at the first finger position. Only phage in which all three fingers were capable of binding were retained when the oligonucleotide targets were washed to remove nonspecific protein-DNA interactions. These phage were then eluted from their targets, amplified, and subjected to additional rounds of selection. This process eventually produced Zif268 variants capable of specifically binding to the sequences 5’-GCGTGGGAC-3’ and 5’-GCGTGGGCA-3’, among others, with affinities comparable to that of Zif268 (Choo and Klug, 1994b; Jamieson et al., 1994; Rebar and Pabo, 1994).

The success of these studies spawned a wave of research designed to identify zinc finger sequences, either in a Zif268 or an Sp1 backbone, which would bind to all possible DNA triplets. To accomplish this, it was necessary to make the assumption that when a zinc finger sequence was selected as having high affinity for a DNA triplet as a first (or, in many later studies, a second) finger, it would retain this affinity when placed in a different context. Although some of the results of the early studies designed to identify a zinc finger binding code indicated that different zinc fingers may alter the specificity of neighboring domains (Desjarlais and Berg, 1992; Desjarlais and Berg,
1993), the phage display experiments have shown that zinc fingers are largely modular and, aside from concerns arising when an arginine and an aspartate occupy α helical positions -1 and 2, respectively, can be moved to any position in a zinc finger protein at will (Elrod-Erickson et al., 1996; Segal et al., 2003).

Since these early phage display studies were most successful in generating zinc fingers that bound to guanine-rich sequences, later studies increased the zinc finger binding variability by randomizing the zinc finger sequences encoding α-helical residues -1, 1, 2, 3, 4, 5, and 6. The studies have led to the identification of zinc fingers that bind with high affinity and good specificity to all GNN, most ANN and CNN, and a few TNN DNA triplets (Dreier et al., 2001; Dreier et al., 2005; Dreier et al., 2000; Liu et al., 2002; Segal et al., 1999; Wu et al., 1995). Thus, a library of zinc fingers, which can be assembled arbitrarily to create zinc finger proteins capable of specifically binding to almost any DNA sequence, has been created. The specificity-determining residues of the zinc fingers in this library are presented in Table 1-2.

As the rules governing the creation and subsequent use of zinc fingers were being developed, several groups also began trying to construct zinc finger proteins to bind specific biologically relevant DNA targets, including sequences in the TATA box, a p53 binding site, and a nuclear receptor element (Greisman and Pabo, 1997; Wolfe et al., 1999). Since a library of characterized zinc fingers was still in its infancy, the zinc fingers used these studies were identified by a technique known as sequential selection.

When using this strategy to construct zinc fingers for the TATA target, this site was fused to the Zif268 target site in such a manner that the second and third fingers of
Table 1-2: Zinc finger-DNA specificities.

<table>
<thead>
<tr>
<th>DNA Target</th>
<th>ZF Sequence (Barbas group)</th>
<th>Rating</th>
<th>Alternate Sequence (Sangamo)</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAA</td>
<td>QSSNLVR</td>
<td>++</td>
<td>QSGNLAR</td>
<td>++</td>
</tr>
<tr>
<td>GAC</td>
<td>DPGNLVR</td>
<td>++</td>
<td>DRSNLTR</td>
<td>++</td>
</tr>
<tr>
<td>GAG</td>
<td>RSDNLVR</td>
<td>+++</td>
<td>RSDNLAR</td>
<td>+++</td>
</tr>
<tr>
<td>GAT</td>
<td>TSGNLVR</td>
<td>+++</td>
<td>TSANLSR</td>
<td>+++</td>
</tr>
<tr>
<td>GCA</td>
<td>QSGDLRR</td>
<td>++</td>
<td>QSGDLTR</td>
<td>++</td>
</tr>
<tr>
<td>GCC</td>
<td>DCRDLAR</td>
<td>+++</td>
<td>DRSDLTR</td>
<td>++</td>
</tr>
<tr>
<td>GCG</td>
<td>RSDDLVR</td>
<td>+</td>
<td>RSDDLQQR</td>
<td>++</td>
</tr>
<tr>
<td>GCT</td>
<td>TSGELVR</td>
<td>++</td>
<td>QSSDLTR</td>
<td>+++</td>
</tr>
<tr>
<td>GGA</td>
<td>QRAHLER</td>
<td>+++</td>
<td>QSGHLQQR</td>
<td>++</td>
</tr>
<tr>
<td>GGC</td>
<td>DPGHLVR</td>
<td>++</td>
<td>DRSHLAR</td>
<td>+</td>
</tr>
<tr>
<td>GGG</td>
<td>RSDKLVR</td>
<td>+++</td>
<td>RSDHLSR</td>
<td>+++</td>
</tr>
<tr>
<td>GGT</td>
<td>TSGHLVR</td>
<td>++</td>
<td>TSGHLVR</td>
<td>++</td>
</tr>
<tr>
<td>GTA</td>
<td>QSSSLVR</td>
<td>+++</td>
<td>QSGALAR</td>
<td>+</td>
</tr>
<tr>
<td>GTC</td>
<td>DPGALVR</td>
<td>+++</td>
<td>DRSALAR</td>
<td>+</td>
</tr>
<tr>
<td>GTG</td>
<td>RSDELVR</td>
<td>++</td>
<td>RSDALTR</td>
<td>++</td>
</tr>
<tr>
<td>GTT</td>
<td>TSGSLVR</td>
<td>+</td>
<td>TSGALTR</td>
<td>+</td>
</tr>
<tr>
<td>DNA Target</td>
<td>ZF Sequence</td>
<td>Rating</td>
<td>DNA Target</td>
<td>ZF Sequence</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>--------</td>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td>AAA</td>
<td>QRANLRA</td>
<td>++</td>
<td>CAA</td>
<td>QSGNLTE</td>
</tr>
<tr>
<td>AAC</td>
<td>DSGNLRV</td>
<td>+++</td>
<td>CAC</td>
<td>SKKALTE</td>
</tr>
<tr>
<td>AAG</td>
<td>RKDNLKN</td>
<td>++</td>
<td>CAG</td>
<td>RADNLTE</td>
</tr>
<tr>
<td>AAT</td>
<td>TTGNLTV</td>
<td>+</td>
<td>CAT</td>
<td>TSGNLTE</td>
</tr>
<tr>
<td>ACA</td>
<td>SPADLTR</td>
<td>+++</td>
<td>CCA</td>
<td>TSHSLTE</td>
</tr>
<tr>
<td>ACC</td>
<td>DKKDLTR</td>
<td>++</td>
<td>CCC</td>
<td>SKKHLAE</td>
</tr>
<tr>
<td>ACG</td>
<td>RTDTLRD</td>
<td>+</td>
<td>CCG</td>
<td>RNDTLTE</td>
</tr>
<tr>
<td>ACT</td>
<td>THLDLIR</td>
<td>+++</td>
<td>CCT</td>
<td>TKNSLTE</td>
</tr>
<tr>
<td>AGA</td>
<td>QLAHLRA</td>
<td>++</td>
<td>CGA</td>
<td>QSGHLTE</td>
</tr>
<tr>
<td>AGC</td>
<td>-</td>
<td>-</td>
<td>CGC</td>
<td>HTGHLLE</td>
</tr>
<tr>
<td>AGG</td>
<td>RSDHLTNT</td>
<td>++</td>
<td>CGG</td>
<td>RSDKLTE</td>
</tr>
<tr>
<td>AGT</td>
<td>HRTTTLTN</td>
<td>++</td>
<td>CGT</td>
<td>SRRTCRA</td>
</tr>
<tr>
<td>ATA</td>
<td>QKSSLIA</td>
<td>++</td>
<td>CTA</td>
<td>QNSTLTE</td>
</tr>
<tr>
<td>ATC</td>
<td>-</td>
<td>-</td>
<td>CTC</td>
<td>-</td>
</tr>
<tr>
<td>ATG</td>
<td>RRDELNV</td>
<td>+</td>
<td>CTG</td>
<td>RNDALTE</td>
</tr>
<tr>
<td>ATT</td>
<td>HKNALQN</td>
<td>++</td>
<td>CTT</td>
<td>TGGALTE</td>
</tr>
</tbody>
</table>

The specificity-determining residues (-1, 1, 2, 3, 4, 5, 6 of the α helix) of the zinc finger found to bind most specifically to each of the DNA target triplets. Two different studies were conducted to determine which finger sequences bind best to the 5'GNN3' triplets. Since the attributes of the fingers selected in these studies cannot be easily compared, both are reported. Additionally, an arbitrary rating (+, ++, ++++) of the apparent combined affinity and specificity (and, thus, usefulness) of the fingers is included. The TNN section of the table has been omitted, since no comprehensive studies to identify zinc fingers which bind to these sequences have been published. Table information compiled from (Carroll et al., 2006; Dreier et al., 2001; Dreier et al., 2005; Dreier et al., 2000; Liu et al., 2002; Segal et al., 1999).
the nascent protein could bind to their natural Zif268 triplets while the randomized residues on the first finger would be selected for their ability to bind to the initial triplet of the TATA target. After this chimeric protein was identified, the third finger was removed so the new third (formerly second) finger could bind to the Zif268 target and the second (formerly first) finger could bind to its triplet on the TATA sequence while a new first finger was selected which bound to the next portion of the TATA box target. In the third round, the final Zif268 finger was removed, and selection for a new first finger guided by the previously selected TATA fingers. This method, although arduous, ensured that each finger was selected in the context of its neighbors so that the final protein would have maximum affinity to the desired target.

An alternate approach was used in identifying zinc fingers which bound to the HIV-1 promoter (Isalan et al., 2001). In this study, the sequential selection strategy was modified by dividing the middle finger of Zif268 in half and performing only two rounds of sequential selection. This bipartite selection reduced the work required to identify a zinc finger protein of novel specificity while still accounting for interfinger context effects.

Additionally, many novel zinc finger target sequences have been identified through a strategy known as “GeneGrip,” which examines the target affinity of the natural zinc finger proteins found throughout the human genome (Bae and Kim, 2006; Bae et al., 2003). DNA encoding previously identified human zinc finger motifs was cloned into the third finger of Zif268, and this hybrid protein was placed upstream of a Gal4 transcriptional activation domain. Reporter plasmids were then constructed by placing Zif268 targets, in which the initial triplet had been randomized, immediately...
upstream of either a *his3* or *lacZ* gene. Yeast matings identified which zinc finger proteins bound tightly to the hybrid targets by reporter activation. Since all of these zinc finger motifs occur naturally in the human genome, it was hoped that they would bind specific DNA triplets without many of the context effects seen in libraries constructed by phage display. Although the majority of the zinc fingers identified bound most strongly to GNN targets, many novel zinc fingers have been characterized by this technique.

Work on this and other zinc finger identification strategies is ongoing, and much progress has been made in the last several years. Presently, by drawing upon the pooled data from all of the selection studies described, it is possible to design zinc fingers with good affinity and high specificity to almost any DNA triplet.

**Zinc fingers in artificial transcription factors**

Many of the studies designed to characterize zinc fingers were conducted in an effort to harness one of the most powerful potential uses of zinc finger proteins: the ability to act as specific artificial transcription factors. This application is not surprising, since most naturally occurring zinc fingers are found in transcription factors, binding to a specific DNA sequence to recruit and activate RNA polymerases to nearby genes (Urnov, 2002). The modular DNA-binding nature of zinc fingers makes them especially suited to this role, since they can recognize specific asymmetric DNA sequences so that either a linked or separate modulator protein can act on the DNA in a sequence-specific manner. It is for this purpose that the most artificial zinc fingers proteins have been constructed. The ability to specifically activate or repress a gene
has made possible a new form of gene targeting which manipulates the expression of a specific gene by altering its promoter activity without altering its sequence.

Consequently, as the rules regulating zinc finger binding were discovered, many such studies were performed over the course of only a few years. Initial studies usually measured the ability of zinc finger proteins bound at a promoter to modify the expression of a downstream gene in a transfected reporter construct. For example, in one series of experiments it was shown that Zif268, along with novel zinc fingers selected by phage display for desired target sequences, could bind downstream from the TATA-binding protein to efficiently repress luciferase expression (Kim et al., 1997a; Kim and Pabo, 1997; Pomerantz et al., 1995). Additional studies advanced this concept by demonstrating that novel zinc fingers could also be fused to activation domains, such as VP16 (Sadowski et al., 1988), to activate expression of transiently transfected CAT reporter genes (Choo et al., 1997).

The next great advance in zinc finger studies came when it was shown that artificially constructed zinc fingers were capable of altering gene expression on a chromosomal target (Choo et al., 1994). In this case, a completely novel nine-base target, 5' - GCAGAAGCC - 3', was identified in an oncogene on a mouse chromosome, and zinc fingers were selected by phage display to identify candidates that would specifically bind this sequence. Mouse cells were transfected with these fingers, and their effectiveness at inhibiting transcription from this gene was monitored by northern blot. Reduced mRNA levels from transfected cells indicated that, when zinc fingers were bound, they actually repressed transcription in vivo.
The success of these initial studies encouraged many additional experiments. Most of these subsequent experiments employed chimeric proteins, consisting of a zinc finger DNA-binding domain and an effector domain to specifically alter activity of the targeted gene. Initially, transcription activation domains, such as the VP16 activation domain from the herpes simplex virus, were used to show that zinc fingers could initiate transcription of otherwise quiescent genes downstream from their target site. Likewise, repressor domains, such as the KRAB repressor domain (Margolin et al., 1994), can be used to effectively shut down transcription at a zinc finger-targeted locus. Studies targeting the proto-oncogenes erbB-2 and erbB-3 with chimeric zinc finger activators and repressors strongly indicated the utility of specifically modifying the activity of a gene at a specific locus in this manner (Beerli et al., 2000; Beerli et al., 1998). Furthermore, since the zinc finger binding sites selected at these two loci were quite similar, this study also demonstrated that the zinc finger constructs were able to recognize their targets with remarkable specificity, indicating that this type of gene manipulation could be done at an arbitrary and specific locus in a large genome. The success of these studies prompted additional demonstrations of specific in vivo modulation of genes by zinc fingers attached to effector domains (Kang and Kim, 2000; Papworth et al., 2003; Ren et al., 2002).

As these studies conclusively demonstrated that zinc finger constructs were capable of initiating transcription on unpackaged DNA, additional studies sought to increase the utility of the zinc finger system. Experiments in which zinc finger modulators were designed to activate the erythropoietin (epo) and vascular endothelial growth factor A (vegfa) genes indicated that, although successful in activating their
target genes, the zinc finger proteins were less effective modulators when the DNA was condensed in chromatin (Liu et al., 2001; Urnov, 2002; Urnov et al., 2002; Zhang et al., 2000). A study using zinc fingers attached to a Koxl repression domain to specifically reduce the expression of the *chk2* gene employed a microarray to demonstrate that this repression was specific to the target gene (Tan et al., 2003). Additionally, zinc fingers have been attached to histone and chromatin-modifying domains to modify gene expression at targeted loci (Corbi et al., 2004; Isalan and Choo, 2000; Jamieson et al., 2003; Snowden et al., 2002), and zinc fingers have been attached to the Oct-1 homeodomain to create a target-specific drug delivery system (Pomerantz et al., 1995). Finally, studies in mice have shown that zinc fingers are able to specifically activate the *vegfa* gene, resulting in its normal expression in vivo, demonstrating that the goal of gene therapy using zinc finger promoters is within reach (Rebar et al., 2002).

**Polydactyl zinc finger proteins**

During the course of these studies, it was found that, in many cases, a three-fingered protein was sufficient to achieve specific binding to the target sequence (Choo et al., 1994; Liu et al., 2001; Zhang et al., 2000). However, in a large genome, the 9-bp DNA target sequence identified by a three-fingered protein is certainly not unique to the specified genomic locus (Liu et al., 1997). To resolve this dilemma, different strategies to add additional fingers to the protein were undertaken (Segal, 2002). The additional specificity conferred by a six-fingered, or polydactyl, protein would, theoretically, also increase the affinity of the protein. However, this did not always prove true. Although modeling studies based on the crystal structure of Zif268 bound
to its DNA target indicated that six zinc fingers could be canonically linked together to bind an extended target, it was found these consecutive zinc finger motifs did not always increase affinity significantly (Beerli et al., 1998; Liu et al., 1997; Segal et al., 2003). It was thought that this effect might be due to the steric restrictions imposed on the DNA by the rigid structure of an extended zinc finger protein, which would counteract the greater affinity of the combined fingers.

Studies examining the zinc finger-DNA complex indicated that binding by zinc fingers contorted the DNA molecule, resulting in a complex in which the major groove of the DNA was enlarged (Nekludova and Pabo, 1994) and the DNA partially unwound (Shi and Berg, 1996). To resolve this problem, artificial linkers were constructed to increase the flexibility of the protein, resulting in polydactyl finger binding proteins with better affinity (Beerli et al., 1998; Kim and Pabo, 1998; Moore et al., 2001a; Moore et al., 2001b). Using these constructs, two or more sets of two or three zinc fingers can be linked to target a specific binding site, although, until a complete library of specific zinc fingers is available, the lower frequency of these longer target sites will reduce the utility of these polydactyl zinc fingers. In spite of these obstacles, this field of research has the potential of developing effective strategies for controlling gene expression, and these successes are almost certainly only a small example of the potential of this system. Work in this area remains very active, and new strategies for in vivo alteration of gene expression are undoubtedly forthcoming. Furthermore, in addition to their role as gene modulators, zinc fingers have become a central component in a new strategy aimed at not merely modifying gene expression but at actually targeting specific genes in vivo by inducing specific DSBs at their target sites.
**FokI restriction enzyme**

In 1996, Chandrasegaran et al. attached the nuclease domain of the type II restriction enzyme *FokI* to a set of three zinc fingers of known affinity (Kim et al., 1996). The *FokI* restriction enzyme had been first described in 1981 (Sugisaki and Kanazawa, 1981) and was subsequently purified and characterized (Kaczorowski et al., 1989; Kita et al., 1989; Looney et al., 1989) (Fig 1-7).

Like other type II restriction enzymes, *FokI* contains a DNA recognition domain and a nuclease domain and forms a homodimer to cleave its target DNA. However, unlike the many closely related enzymes such as *EcoRI* and *BamHI*, in which the two domains are closely linked, these domains form separate entities in *FokI* (Li et al., 1992; Li et al., 1993; Waugh and Sauer, 1993). The DNA-binding domain recognizes the nonpalindromic sequence 5' - GGATG - 3'. Once bound, the nuclease domain can dimerize with that of another molecule to cleave the DNA 9 and 13 bases away from this target site. This unusual mode of DNA cleavage led to subsequent structural studies to better characterize the activity of this enzyme (Wah et al., 1997).

Whereas most type II enzymes cleave within a palindromic DNA sequence by forming tight homodimers that act in a concerted fashion to cleave their target, *FokI* does not usually dimerize in solution. It remains a monomer even at micromolar concentrations and can, apparently, bind its target sequence as a monomer. In this conformation, the nuclease domain is somehow sequestered by the recognition domain, so that the enzyme is inactive in solution (Wah et al., 1998). Once bound to its target site however, the nuclease domain is exposed, even though the *FokI*-DNA complex is catalytically inactive.
The recognition domain of a single molecule of FokI is composed of three subdomains (D1-D3), in magenta, green, and white. Domains D1 and D2 recognize the 5' - GGATG - 3' sequence to which the enzyme binds, while D3 interacts with the cleavage domain, in blue, sequestering it when the enzyme is not bound to DNA. When FokI binds to DNA, the cleavage domain is released and can swing into the major groove to cleave the DNA at a point 13-bp downstream from the recognition site. A second FokI molecule must then be recruited to complete the DSB (Wah et al., 1998; Wah et al., 1997), with permission from Nature Publishing Group.
Studies employing FokI mutants lacking the DNA-binding domain have shown that, in order to cleave DNA, this complex must dimerize with another molecule of FokI, which has either bound to a nearby recognition site or which is still in solution, although the latter reaction occurs quite slowly (Bitinaite et al., 1998). Alternately, two distant recognition sites can coordinate enzyme activity by looping out the intervening DNA sequences to bring the two FokI monomers into close proximity. Finally, two FokI monomers bound to their targets can associate in trans. (Catto et al., 2006; Vanamee et al., 2001). In this case, when both monomers are bound to DNA, dimerization is much more favorable. Once dimerization occurs, target site cleavage can commence. Thus, through one of these mechanisms, FokI dimers can cleave DNA containing only a single recognition site, even though the monomers will not form stable dimers in solution.

Prior to his work with zinc fingers, Chandrasegaran had sought different ways of targeting cleavage of this nuclease domain to novel sites. In earlier experiments he had increased the distance between the binding and cleavage domains of FokI to alter the number of nucleotides between its recognition site and the cleavage site on target DNA. In subsequent studies, he had also linked the FokI endonuclease domain with the Drosophila helix-turn-helix homeobox domain to alter its specificity (Kim and Chandrasegaran, 1994; Li and Chandrasegaran, 1993). Other researchers also explored similar strategies and would also eventually link the Zα DNA-binding domain from human double-stranded RNA adenosine deaminase to the FokI nuclease domain to produce a Z-DNA-specific endonuclease (Kim et al., 1997b). Zinc fingers seemed much more flexible than any of these tools, so they became the subject of continuing
research (Kim et al., 1996; Kim et al., 1997c). When they were attached to the endonuclease domain of FokI, it was found that the resulting chimeric proteins contained a DNA recognition domain that could direct cleavage by the nuclease domain to a nearby DNA sequence. The utility of such a sequence-specific endonuclease was realized and the requirements and capabilities such an enzyme were characterized (Durai et al., 2005; Mani et al., 2005a; Mani et al., 2005b; Smith et al., 1999).

**Early ZFN studies**

In order to examine the potential usefulness of this new system, preliminary studies were performed to characterize the DNA-binding and cleavage requirements of the ZFNs. Initial experiments, using a ZFN named QQR, indicated that, as in the natural FokI enzyme, two nuclease domains had to dimerize in order to create a DSB containing a 4 bp 5' staggered overhang in their substrate DNA (Smith et al., 2000). This could be easily accomplished by creating two inverted binding sequences on a synthetic target DNA where two QQR molecules could bind and bring their nuclease domains into close proximity. In concurrent experiments, QQR and a related ZFN with an altered DNA recognition domain, known as QNK, were also shown to bind and cleave synthetic targets containing inverted binding sites for the heterodimer.

It was discovered that, as with the natural FokI enzyme, the dimer interface between the nuclease domains was so weak that individual ZFNs would not dimerize in solution. They first had to be brought into close proximity at their cleavage site by the zinc fingers. Furthermore, the replacement of the FokI DNA-binding domain with the zinc finger domain appeared to largely eliminate the ability of the nuclease to cleave
the target when only one molecule was bound. Thus, DNA target specificity of the zinc fingers equated to cleavage specificity of the ZFN pair. This characteristic actually improves that specificity of the ZFN system: Each ZFN binds to a specific 9-bp sequence of DNA. Along a random strand of DNA, a specific 9-bp sequence should occur once every 262,000 bp. However, as a dimer, the ZFNs bind to a specific 18-bp sequence, which should occur once every $6.8 \times 10^{10}$ bp. Thus, a sequence selected for ZFN-mediated targeting should be unique in even a large genome. Nonspecific cleavage of the genomic DNA should be minimized since the zinc fingers used generally display good specificity for their target and homodimer targets should occur as infrequently as the target site. Furthermore, using ZFNs containing only three zinc fingers eliminates the steric restriction imposed by the imperfect register between the zinc fingers and DNA major groove in many of the studies in which additional zinc fingers were linked (Liu et al., 1997). This combination of characteristics means that, when active, ZFNs should be optimal targeting tools.

As these initial experiments were being completed, other studies examining the ability of the ZFN-mediated DNA cleavage to stimulate homologous recombination were also underway (Bibikova et al., 2001). Target plasmids containing a 1.25-kb direct repeat, separated by different ZFN targets, were injected into *Xenopus laevis* oocyte nuclei, where they were assembled into chromatin. After 3-4 hours, QQR and/or QNK were injected, and after several additional hours the target plasmids were recovered and assayed for DSB-induced recombination between repeats. Several related ZFN constructs, varying only in the number of amino acids linking the zinc fingers to the nuclease domain, were tested. Eventually it was found that the ZFNs
could most efficiently induce recombination between repeats when this variable linker was arbitrarily fixed at zero and their target sites were placed in an inverted orientation and separated by 6 bp (Fig. 1-8). Modeling of the ZFNs bound to DNA indicated that this orientation would maximally align the nuclease domains, while minimizing steric interactions between protein and DNA, and produce most effective cleavage.

These experiments defined the parameters under which ZFNs should be most capable of locating a specific chromosomal target and inducing a specific DSB at the site at high frequency. Subsequent activation of the cellular repair mechanisms should produce targeted mutations at that site. The next step was to attempt to recapitulate these results in vivo using novel ZFNs employing zinc fingers from the existing library assembled specifically to bind to target sequences chosen in a gene of interest.

**ZFN-mediated gene targeting in other organisms**

The first ZFN in vivo targeting experiment was attempted in the fruit fly *Drosophila melanogaster* (Bibikova et al., 2002). The *yellow* (y) gene was selected since it is on the X chromosome and, in males, its mutation would be easily observable. Its sequence was examined to locate DNA sequences for which ZFNs could be constructed. Since the GNN fingers had been most thoroughly characterized, target site searches were restricted to the degenerate sequence 5' NNC NNC NNC NNNNNN GNN GNN GNN 3.' Such a site would enable two ZFNs to be constructed that bound to GNN triplets on apposing sequences and were separated by a 6 bp DNA spacer – optimal conditions based on the in vitro experiments completed previously.
The zinc fingers of two molecules of QQR are in blue and are aligned with their target DNA triplet. The FokI nuclease domains, in purple, straddle the DNA between these binding sites and can dimerize at this location to stimulate the observed DSBs and subsequent 4-bp 5' overhangs at the target (Bibikova et al., 2001; Smith et al., 2000).

Figure 1-8: The ZFN QQR bound to its DNA target.
In the end, the sequence 5' GCC TAC CGC ATTAAA GTG GAT GAG 3' was chosen, and novel ZFNs (yA and yB), designed to bind to each of the 9-bp binding sites, were constructed and placed in the Zif268 backbone. Once constructed, the activity of these ZFNs was verified in an in vitro assay using the purified yA and yB proteins and a plasmid containing the yellow gene (Carroll et al., 2006). Flies were then transformed by ZFN-containing P-element vectors, which were expressed via heatshock in larvae. Adult transgenic flies were scored for somatic mutations, and several male flies could be seen bearing somatic yellow patches, indicating that yA and yB had indeed cleaved their target in yellow. Furthermore, when these flies were crossed with a stock designed to reveal mutations, 5.7% of them produced yellow sons (0.44% of all progeny.)

Interestingly enough, this high level of mutation was observed even though expression of one of the ZFNs resulted in some lethality among transgenic flies. This lethality has been shown to be the result of off-target cleavage by the ZFN, even though no canonical target sites can be identified in the fly genome in an orientation that would predict cleavage (Beumer et al., 2006). This indicates the following: (1) even though the zinc fingers have been shown to be fairly specific, complete specificity is not guaranteed, and (2) in spite of some nonspecific ZFN cleavage, moderation of induction of the ZFNs can attenuate this lethality so it is not an insurmountable problem and mutant progeny can still be obtained.

The yellow mutations induced by these ZFNs were products of inaccurate DNA repair in the fly cells. The DSBs stimulated by the ZFNs must be repaired for the cell to remain active. There are two main strategies for DNA repair: NHEJ and HR. Since
the NHEJ components simply jam the broken DNA ends together without regard for sequence conservation, missense and nonsense mutations resulting from substitutions, deletions, and insertions frequently occur. These types of mutations led to the yellow phenotype seen in the mutants. The alternative DNA repair pathway, HR, uses the homologous chromosome or sister chromatid as a template to repair the DSB and results in perfect reconstruction of the sequence. When this type of repair occurs, no mutations will result.

The next series of experiments in flies was designed to capture at least a subset of HR-mediated repair events by employing an exogenous marked DNA sequence that could be used as a template for directed DNA repair via HR (Bibikova et al., 2003) (Fig. 1-9.) In this case, the exogenous sequence – or donor – was also introduced on a P element, then either allowed to act as a template from its genomic locus or liberated using the Flp/FRT system and I-SceI linearization.

In this series of experiments, somatic mutations were not examined, but germline mutations arose in 20% of heatshocked males and 14% of heatshocked females. When the donor was linearized, it was most effective as a template for repair of the DSB. In this case, more than 2% of all sons of males were mutant, and 63% of these mutations were the products of HR. From heatshocked females, 73% of mutants arose from HR using the marked donor. Thus, through these series of experiments, a specific way to induce mutations in an arbitrary gene was developed and a new gene targeting strategy was born in Drosophila. Furthermore, this targeting strategy was about two orders of magnitude more efficient than the targeting strategy pioneered by Kent Golic. Its success is due to the ability of the ZFN to induce a specific DSB at a predetermined
The *yellow* target, bound and cleaved by the *yA* and *yB* ZFNs, is shown centrally. Depicted above, the *yellow* donor - excised from its P-element and linearized by I{-SceI} - can act as a template to repair the *yellow* gene. Repair via HR using this donor molecule will result in mutation of the gene, since the donor has been engineered to contain two in-frame stop codons at the target site. Additionally, *yellow* mutant flies resulting from donor repair can be distinguished from mutant flies resulting from NHEJ processes by diagnostic digestion since a recognition site for the restriction enzyme *XhoI* has been inserted immediately downstream of the stop codons (Beumer et al., 2006).
target, which then activates the cellular DNA repair mechanisms and stimulates incorporation of a marked donor.

Since this initial success, additional genes have been targeted, including *brown* (*bw*) and *rosy* (*ry*) (Beumer et al., 2006) (Table 1-3). At the *rosy* locus, results have been especially encouraging, since no ZFN-related lethality was observed, allowing ZFNs to be induced at a high level. In this case, even when no donor is present, ~25% of heatshocked female flies produce mutant offspring. It seems as if ZFNs have the ability to induce mutations and target arbitrary genes at a very high efficiency. The question now becomes: can ZFNs be used to target genes in other organisms also?

It has been recently shown that ZFNs are functional in human cell lines. In an initial study, QQR, driven by a CMV promoter, was cotransfected with a repair template into HEK cells containing an integrated inactive GFP construct in which two inverted QQR binding sites had been inserted (Porteus and Baltimore, 2003). The repair template contained sequences that, if incorporated into the inactive GFP by HR, would activate it. It was discovered that 0.15% of the transfected cells became positive for GFP, indicating that the DSB induced by QQR at the target site in the inactive GFP had been repaired by HR, using the homology supplied by the repair template. Furthermore, in a second experiment, using a target cleavable by a QQR-Zif268FN heterodimer, even higher levels of GFP rescue were obtained. Over the course of several days, it was observed that 70% of the rescued cells died, indicating that overexpression and off-target cleavage of the ZFNs were issues in human cells also.

In a separate experiment, ZFNs were designed to target the human IL2Rγ gene, in which a specific mutation results in X-linked SCID (Urnov et al., 2005). In this case,
Table 1-3: Targeting frequencies in Drosophila

<table>
<thead>
<tr>
<th>Target gene</th>
<th>Target sequence (5' → 3')</th>
<th>Mutant progeny (%)</th>
<th>Total Mutant Progeny (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>yellow</td>
<td>GCCTACCCGC</td>
<td>attaaa</td>
<td>7.0</td>
</tr>
<tr>
<td></td>
<td>GTGGATGAG</td>
<td></td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8.9</td>
</tr>
<tr>
<td>brown</td>
<td>CCCATCATC</td>
<td>aggcgg</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td>GAGGTTGGGC</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>rosy</td>
<td>AGCTACTAC</td>
<td>acgaat</td>
<td>15.0</td>
</tr>
<tr>
<td></td>
<td>GGCCTGGGA</td>
<td></td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>25.0</td>
</tr>
</tbody>
</table>

The yellow, brown, and rosy target sequences in Drosophila, along with the percentage of offspring with targeted mutations arising from the female germline after the mothers were heatshocked at 35°C (for brown) or 37°C (for yellow and rosy). Information from (Beumer et al., 2006; Bibikova et al., 2002), as well as from Kelly Beumer and Dana Carroll, unpublished.

At the brown locus, expression of one of the nucleases, bwB, was lethal at high levels in the flies. In subsequent experiments, in which bwB lethality was tempered in a new ZFN construct, mutant progeny were recovered at levels comparable to those seen at the yellow and rosy loci (Dana Carroll and Kelly Beumer, personal communication).
two polydactyl ZFNs, each consisting of two sets of two fingers connected by extended linkers (Moore et al., 2001b), were used to stimulate DSBs at their target in K562 cells. Interestingly, this target site consisted of finger binding sites separated by only 5 bp, indicating that dimerization requirements may not be as strict as observed in Xenopus oocyte nuclei. A modified, but unprotected, template plasmid was introduced with the ZFNs to act as a template for HR. These constructs rescued 12.2% of the cells at a single locus and 2.4% of the cells at both endogenous loci, as assayed by restriction digest. Furthermore, in experiments performed in CD4⁺ T cells, HR-mediated gene correction was still observed at a frequency of 5.3%. The design changes made to the zinc fingers seem to have eliminated the apparent toxicity of their expression (possibly by increasing their specificity and affinity to their target by the addition of the fourth finger on each set), although some level of off-target activity may still occur. This study is especially significant since it indicates that ZFNs may be useful tools in gene therapy treatments in humans, where hematopoietic stem cells removed from a patient can be modified in vitro at a high frequency and returned to the body.

ZFNs have also been employed successfully both in the experimentally popular Arabidopsis thaliana and in tobacco plants. In Arabidopsis, QQR was placed behind a heatshock promoter in a vector which also contained a QQR target and introduced into young plants by Agrobacterium (Lloyd et al., 2005). After induction of QQR, the QQR target was amplified from transgenic plants, and it was discovered that 7.9% of all target sites had been cleaved and repaired inaccurately. Furthermore, when the next generation of transgenic plants was examined, 10% of these plants contained target site mutations acquired through germline transmission of the mutated targets. This
experiment demonstrated that ZFNs can induce DSBs in Arabidopsis, and the NHEJ-mediated DNA repair can induce mutations at a high frequency in the organism.

In tobacco, not only was it was shown that ZFNs could induce DSBs, but that these DSBs could also be repaired by HR using a exogenous template - an indication that ZFNs could be used in true gene targeting experiments in plants (Wright et al., 2005). This is significant because, as in higher metazoans, most exogenous DNA introduced into plants is integrated by NHEJ. There are typically $10^3$ - $10^7$ illegitimate recombination events for every HR-mediated event. An artificial target site for Zif268FN fingers was placed within a defective GUS:NPTII reporter (which is a gene fusion that, when active, acts analogously to the LacZ reporter in bacteria when incubated with 5-bromo-4-chloro-3-indolyl glucuronide and also provides kanamycin resistance to the cell). This construct was introduced into tobacco chromosomes via Agrobacteria. Transgenic tobacco protoplasts were selected, into which the ZFNs were introduced along with a repair template (as in the HEK GFP experiments), and rescue of the reporter gene was monitored. In this case, repair by HR occurred in about 10% of the transformed tobacco plants protoplasts. Among 20% of these protoplasts, DSB-induced HR repaired the reporter gene without accompanying NHEJ-mediated events. Thus, it can be seen that ZFN technology can be used in a wide variety of organisms, since DSB repair proceeds by a widely conserved mechanism.

These experiments indicate that the specific DSBs induced by ZFNs are capable of increasing the frequency of gene targeting by orders of magnitude over traditional targeting techniques. Using this technique, it should be possible to target almost any gene in Drosophila, Arabidopsis, and even in human cells with relative ease.
Furthermore, there is no reason why other organisms, in which no gene targeting strategy exists, cannot profit from this strategy. In addition to the work in nematodes described in this dissertation, efforts are underway to show that ZFNs can induce DSBs in zebrafish, and proposals have been issued to examine their effectiveness in planaria.

Part III: Adapting ZFNs for gene targeting in nematodes

In order to examine the effectiveness of ZFNs in inducing targeted mutations in *C. elegans*, an efficient method for delivering these tools to the nematode genome and expressing them in the germline must be established. Although ZFNs have been shown to induce DSBs in several different organisms, there is no universal mode of delivery for these enzymes. Even though the fundamental DNA repair processes seem to be conserved, each organism will process incoming transgenic DNA (containing the ZFNs and modified repair template) distinctly. This, in turn, will dictate how this transgenic DNA can be activated to achieve targeted cleavage and repair. In Drosophila, P-elements can be introduced and information on them can be expressed in the germline (Gloor et al., 1991). In human cells, including harvested white blood cells that can be returned to their hosts, transgenes can be introduced by transfection (Urnov et al., 2005). In plants, introduction by means of an agrobacterial vector or via protoplast transformation seems to be sufficient (Lloyd et al., 2005; Wright et al., 2005).

Transgenic DNA in nematodes

Many different experiments have examined the fate of foreign DNA in the nematode germline. It is widely known that foreign DNA can be introduced stably into
nematodes simply by injecting it into their gonads, where it will form extrachromosomal arrays in some proportion of their offspring (Fire, 1986; Stinchcomb et al., 1985). A typical hermaphrodite nematode has two large gonads extending both anterior and posterior from a central vulva, from which self-fertilized eggs emerge (Brenner, 1974). In each arm of the gonad, oocyte development begins distally from a small group of stem cells that continually divide to produce diploid nuclei in a common syncytium (Riddle, 1997; Wood, 1988). As these nuclei are pushed toward the vulva they cellularize and undergo meiosis until the mature oocytes are fertilized in the proximal gonad just prior to being laid. When foreign DNA is injected into the common syncytium in the distal gonad, it is packaged into long concatemers, consisting of randomly joined pieces of DNA (Mello et al., 1991). These concatemers associate with the nuclei, where they can then form stable extrachromosomal arrays that can be transmitted as the nuclei divide during meiosis and subsequent cell divisions.

This phenomenon appears to be unique to nematodes, possibly because the chromosomes in this organism are largely holocentric (Herman et al., 1982). In any case, if large enough, an extrachromosomal array acts like an artificially constructed chromosome and becomes a semistable feature in nematodes that acquire it. If it contains no selectable markers, it can be lost at some frequency without detrimental effect to the nematode, but can also be passed from generation to generation indefinitely. Furthermore, if this extrachromosomal array contains functional genes, they can be expressed in what appears, in many cases, to be a roughly wild-type pattern. This feature has allowed nematode geneticists to knock out genes through
random mutagenesis or RNAi and create artificial constructs to examine gene function and rescue (Clark and Baillie, 1992; Kelly et al., 1997; McDowall and Rose, 1997).

However, there appears to be a very strict limitation to this phenomenon: the expression of any genes present in the extrachromosomal array is very efficiently silenced in the germline (Dernburg et al., 2000; Kelly et al., 1997; Pirrotta, 2002). The DNA sequence is not modified, and, once nematodes hatch, gene expression from the array appears normal, but while in the germline, no expression of genes on an array occurs in most cases. This characteristic presents a formidable challenge, since the introduction and expression of ZFNs in the nematode germline are absolutely necessary to generate targeted offspring. Although the reasons for this strict silencing remained unknown for many years, recent studies seem to shed some light on this event.

Experiments during which injected exogenous RNA was introduced into nematodes resulted in inadvertent inactivation of the genes being studied (Fire et al., 1998; Montgomery et al., 1998; Timmons and Fire, 1998). After further examination of the process, it was eventually determined that this phenomenon was related to the recently characterized post-transcriptional gene-silencing events seen in plants, wherein high copy gene inserts resulted in inactivation of the gene (Matzke and Matzke, 1998). The RNA injected into the nematodes was contaminated with small amounts of antisense RNA that could form double-stranded (ds) RNA in the nematodes. This dsRNA would, in turn, initiate some uncharacterized translational silencing mechanism, resulting in inactivation of the gene by the widespread destruction of the all homologous RNA molecules (Sharp, 1999). Eventually, this phenomenon, called RNA interference (RNAi), was attributed to a sort of viral immune response by the
nematodes (Plasterk, 2002). Invading viruses create dsRNA during the reverse transcription process necessary to insert their genome into the nematode chromosome. The nematode, in turn, can respond by silencing all RNA with the same sequence to evade infection by the virus.

Much of the mechanism responsible for RNA interference has since been characterized (Agrawal et al., 2003; Grishok, 2005; Mello and Conte, 2004; Sijen et al., 2001) (Fig. 1-10). Any dsRNA molecules present are bound by an enzyme called Dicer, which, as implied by its name, chews them up into 21-bp segments known as short interfering (si) RNAs (Bernstein et al., 2001). These siRNA molecules are next unwound by the RISC proteins, and the resulting single-stranded (ss) RNA molecules both direct cleavage of mRNA to which they are complementary and become primers for an RNA-dependent RNA polymerase, which proceeds to synthesize additional copies of the dsRNA target from any homologous mRNA present, thus amplifying the signal generated by the invading dsRNA in a catalytic manner (Sijen et al., 2001; Smardon et al., 2000). Through this mechanism, all RNA, exogenous or endogenous, which has sequence similarity to the detected dsRNA can be effectively eliminated. This mechanism is especially potent in the nematode germline.

It seems as though injected DNA in an extrachromosomal array in nematodes is subject to extensive RNAi in the germline – a process known as cosuppression (Dernburg et al., 2000; Ketting and Plasterk, 2000; Robert et al., 2005). This makes sense when it is remembered that an array is composed of multiple copies of a transgene joined together in random orientation and subject to little regulation of expression. Transcriptional read-through of a transgene may lead to inadvertent
When a cell is exposed to dsRNA, Dicer, together with several cofactors, cut the RNA into 21-bp fragments, known as siRNAs. The siRNA molecules are unwound by the Mut-7 protein, and the single-stranded molecules direct RISC-mediated cleavage of complementary mRNA and act as primers to subsequently amplify homologous mRNA by an RNA-dependent RNA polymerase, which creates additional dsRNA molecules. This process is active in extrachromosomal arrays from which dsRNA is transcribed (Grishok, 2005).
transcription of the complementary sequence of another copy of the gene in reverse orientation. The mRNA produced would then be likely to form an extended hairpin, creating dsRNA and initiating the RNAi mechanism in the nematode. Such a model is supported by experiments done on the nematode transposon Tc1, which is somatically active, but is silenced in the germline (Emmons et al., 1986; Sijen and Plasterk, 2003).

This transposon is present in about 20 copies, randomly distributed throughout the nematode genome (Emmons et al., 1983). Like many other known transposons, Tc1 operates by a cut-and-paste mechanism. The transposon encodes only one gene, flanked by terminal inverted repeats (Plasterk, 1991; Vos and Plasterk, 1994). Expression of the gene results in a protein that will recognize these terminal sequences, clip out the intervening DNA, and paste it into a new locus (Vos et al., 1996). While frequent hopping of the transposons has been observed in nematode somatic tissue, activity in the germline is generally not observed (Emmons and Yesner, 1984).

In an effort to understand this phenomenon, researchers searched nematodes for dsRNA corresponding to Tc1 sequences (Sijen and Plasterk, 2003). It was hypothesized that transposase transcription occasionally proceeds beyond the end of the gene, so that the resultant mRNA contains both recognition sequences. These can then form stem-loop structures containing a short region of dsRNA, which activates the RNAi pathway and silences transposition in the germline. Since the somatic tissues are not as sensitive to RNAi, the small amount of dsRNA is, apparently, tolerated in these tissues and transposition can occur. In additional experiments performed on nematodes to examine this phenomenon, animals were screened for mutants in which germline transposition of Tc1 was no longer silenced. Many, but not all, of the mutants
characterized by these experiments also exhibited other RNAi defects (Ketting et al., 1999; Tabara et al., 1999; Vastenhouw et al., 2003).

The RNAi mechanism has become a useful tool for nematode geneticists, since it can be employed to knock out or knock down the expression of uncharacterized genes so their function can be studied. Initially, dsRNA containing the sequence of the target gene was injected into nematode gonads, and the resulting RNAi progeny were studied (Fire et al., 1998). Eventually, it was determined that a sufficient RNAi effect could be obtained simply by feeding nematodes bacteria expressing dsRNA from a transformed plasmid (Kamath et al., 2001; Timmons and Fire, 1998). Today, plasmid RNAi libraries exist so that almost any nematode gene, known or unknown, can be studied individually, or genomic trends can be examined to find multiple genes that contribute to a process (Kamath et al., 2003; Rual et al., 2004).

Since the discovery of RNAi, many experiments have been conducted to identify all of the components of this pathway. Ironically, many of these experiments are conducted by RNAi, examining genes for which introduced dsRNA will inhibit subsequent RNAi against a reporter gene (Kim et al., 2005). These studies have identified many of the genes involved in RNAi, and, interestingly enough, in nematodes mutant for many of these genes, germline silencing is also reduced or eliminated. In worm mutants for such genes as mes-4, mut-6, mut-7, mut-16, and many others, transposons become active, and reporter genes on arrays are expressed in the germline (Garvin et al., 1998; Kelly and Fire, 1998; Ketting et al., 1999; Tabara et al., 1999).

Armed with this knowledge, there appear to be several potential solutions to the obstacle of germline expression of ZFNs in the nematode. Initially, ZFN expression
can be induced solely in somatic tissues to ensure that these proteins are capable of stimulating DSBs and producing mutations in nematodes. Additionally, genes for ZFNs can be mixed with genomic DNA and injected into nematode gonads. In this context, the extrachromosomal arrays formed will be less subject to silencing since they will contain less repetitive DNA. In conjunction with this, several nematode gene promoters have been shown to be active in the nematode germline (Bessereau et al., 2001; Strome et al., 2001). Driving the ZFNs with one of these promoters may stimulate their expression. Finally, injecting the ZFNs into nematodes mutant for one of the RNAi components (either genetically or by RNAi against one of the genes) may produce offspring in which ZFNs are expressed in the germline. These alternatives will be discussed more explicitly in Chapter 4.

**Mechanisms of DSB repair**

As stated, initial ZFN experiments can be performed in somatic tissue to verify that ZFNs are indeed capable of stimulating DSBs in nematodes. Although nothing about nematode biology indicates that DSB repair occurs differently in these organisms, little information about this process has been published. The expressed ZFNs are able to generate mutations via DSB repair through either of two potentially mutagenic pathways: NHEJ or HR.

Mutations which arise through NHEJ repair result from the process by which nematodes rejoin broken DNA without regard to the original sequence (Jackson, 2002). This process has been well characterized in other organisms, and there is evidence that it proceeds in a similar fashion in nematodes (Clejan et al., 2006) (Fig. 1-11). A DSB
Figure 1-11: Depiction of the NHEJ pathway.

After a DNA molecule suffers a DSB, the ku70 and ku80 heterodimers localize to its broken ends, then recruit the catalytic subunit of the DNA-protein kinase (DNA-PKcs). This complex undergoes autophosphorylation, which recruits the XRCC4 and ligase IV proteins. This large complex is then capable of joining the broken ends of the DNA together, repairing the break (Sekiguchi and Ferguson, 2006).
results in the recruitment of the ku70/80 heterodimer – loop-like proteins which thread onto each end of the break to orient the ends, then associate with one another to bridge the broken DNA (Cary et al., 1997; Walker et al., 2001; Wang et al., 1998; Yoo et al., 1999). These proteins, in turn, activate the catalytic subunit of the DNA Protein Kinase (DNA-PKcs), which when bound to them, forms the complete DNA-PK (Chan et al., 1999). This complex recruits the XRCC4 protein, which, in conjunction with ligase IV (Koch et al., 2004; Modesti et al., 2003), seals the broken ends back together (Calsou et al., 2003; Ramsden and Gellert, 1998). Recently, additional proteins, named Artemis and Cernunnos, have also been implicated in the NHEJ repair pathway, although their roles seem to be of lesser importance (Ahnesorg et al., 2006; Ma et al., 2005).

When DNA is repaired through the NHEJ pathway, it is often subject to exonuclease activity, which can remove damaged bases and result in frequent deletions. Additionally, error-prone polymerases are often used to resynthesize excised bases and can often result in sequence alterations at the break site (Daley et al., 2005a; Ma et al., 2004). If this repair occurs in a gene, these deletions and substitutions can easily alter the reading frame of the gene, introduce missense or nonsense mutations, and produce functional gene knock-outs. This is a frequent consequence of ZFN activity seen in other organisms, including plants and flies (Bibikova et al., 2002; Lloyd et al., 2005). Although not true gene targeting, it is useful in knocking out a gene, a process also useful in studying its function.

When ZFNs generate mutations through the HR repair pathway, the repair process is more complex (Valerie and Povirk, 2003; van den Bosch et al., 2002) (Fig 1-12). Limited research indicates that, again, most of the components of this pathway are
After the DNA suffers a DSB, the Mre11-Rad50-Nbs1 complex initiates the HR mechanism. Exonucleases chew back the broken ends, and Rad52 assists Rad51 in binding the ssDNA and directing it to a homologous sequence. Once a heteroduplex forms, Rad52 and Rad54 direct elongation of the single-stranded DNA past the break point, after which it can withdraw, reanneal with its complement, and be ligated. This process can result in gene conversion (as depicted) (Valerie and Povirk, 2003).
basically conserved in nematodes (Alpi et al., 2003; Boulton et al., 2002; Boulton et al., 2004; Rinaldo et al., 2002). After a DSB is formed, typically the 5' strand at each end is resected to produce single-stranded 3' tails, which are quickly coated with Replication Protein A (RPA) (Hellday, 2003). The tails then interact with the HR protein Rad52, which signals Rad51 and helps it to displace RPA, bind the single-stranded DNA ends, and guide their invasion into an unbroken homologous sequence, concurrently unwound by Rad54 (Martin et al., 2005; New et al., 1998). Other proteins, such as ATM, ATR, Blm, Brca1, and Brca2, have also been implicated in this pathway (Valerie and Povirk, 2003). Once in this duplex, DNA polymerase will extend the single-stranded DNA past the break point, after which the ends can withdraw from the duplex, anneal together, and be ligated into a complete chromosome (Hirano and Sugimoto, 2006; Kawamoto et al., 2005; McIlwraith et al., 2005).

It is by homologous recombination that genuine gene targeting can be achieved. In this case, not only must the ZFNs induce a targeted DSB, but an ectopic template sequence, which has been previously modified to contain the changes desired at the targeted locus, must also be present and available to participate in the repair process. Typically, the template for repair during HR is the sister chromatid, so that the broken chromosome will be repaired with the exact sequence it contained (Kadyk and Hartwell, 1992). When no sister chromatid is available, the homologous chromosome can be used, which can result in some sequence alteration if this alternate template contains a different allele than that present on the broken chromosome. Such repair is known as gene conversion (Elliott et al., 1998). This type of gene conversion also
occurs when information from a modified ectopic template is used to repair the broken chromosome, resulting in the type of gene targeting possible with ZFNs.

In order to target genes in this manner, some knowledge of mechanism by which the cell dictates the mode of DNA repair is necessary. When a DSB is induced, one of the first proteins to localize at the break is the MRN (Rad50-Mre11-Nbs1) complex, which may direct subsequent events and dictate which repair process is to be used (van den Bosch et al., 2003) (see Fig. 1-12). Alternatively, recent studies indicate that repair may result from simple competition between Rad52 and Ku70/80 binding to broken DNA ends. In this case, whichever protein localized first would dictate the subsequent mechanism of repair (Di Virgilio and Gautier, 2005; Lee and Paull, 2004). Finally, the stage of the cell cycle may affect which type of DNA repair predominates: cells in G1 may preferentially repair DSBs by NHEJ since there is no sister chromatid to act as a template, while cells in G2 may typically repair by HR since a template is readily available (Takata et al., 1998).

To generate simple deletion mutants by ZFN-induced DSBs in nematodes, ZFNs can simply be targeted to a chromosomal locus. Even though many of these DSBs may be repaired by HR, these events will simply restore the target site, which can be subsequently recleaved. Any DSBs that are repaired by NHEJ are subject to mutation. When generating targeted mutations via HR in nematodes, a modified plasmid can be included on an array to act as a template for repair, since previous studies have shown that sequences on extrachromosomal arrays are capable of performing this task as some frequency (Plasterk and Groenen, 1992). Since this plasmid will compete with sister chromatids and homologous chromosomes as the repair template, it may be
advantageous to increase its availability by subsequent manipulation, as has been done in flies. If all of these factors can be manipulated, genuine gene targeting in *C. elegans* should be possible.
References


homologous recombination between injected plasmid DNA molecules. Mol Cell Biol 2, 1372-1387.


panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. J Biol Chem 276, 11323-11334.


Segal, D. J. (2002). The use of zinc finger peptides to study the role of specific factor binding sites in the chromatin environment. Methods 26, 76-83.


CHAPTER 2

ZFN-INDUCED SOMATIC CLEAVAGE AND REPAIR

Introduction

Although ZFNs have been extensively tested in many different organisms, there is no guarantee that they will effectively bind, cleave, and produce mutations at their targets in nematodes. Furthermore, the nematode transgene silencing machinery may make it difficult to express the ZFNs in the nematode germline to produce mutant offspring. In this case, ZFN activity would be difficult to verify, since a lack of mutant progeny may indicate that the ZFNs are not functional, but also may simply mean that their expression is silenced in the germ tissues. To circumvent this problem, initial studies with ZFNs in the nematode can be conducted in somatic tissue, so ZFN expression and activity can be monitored without regard to germline silencing. Such a study would necessitate that evidence of ZFN-induced DSBs and inaccurate repair be identified independently at each target, since somatic mutations will rarely have an observable phenotype. To this end, a series of experiments was designed to measure ZFN activity through cleavage and inaccurate repair at synthetic and genomic targets containing diagnostic restrictions sites that are easily observable after PCR amplification from the nematode.
Materials and methods

The ZFNs employed in these studies were expressed from the plasmid pJM1 (Morton et al., 2006). This plasmid is a derivative of pJL44.2, in which the 16-48 heatshock promoter drives expression of MosI transposase (Bessereau et al., 2001). The NdeI and SpeI sites in pJL44.2 were eliminated by PCR-directed mutagenesis. The transposase-encoding sequence was then excised by MluI + Nhel digestion and replaced with a cassette containing a four-adenosine translation initiation consensus sequence, an ATG codon, a nuclear localization signal encoding the amino acids DPKKKRKY, (Adam et al., 1989; Kalderon et al., 1984), and the coding sequence for the Ben1A ZFN (see Chapter 4). In this construct, the DNA sequence encoding the zinc fingers can be excised by NdeI + SpeI digestion, and alternate zinc fingers can be inserted.

To create pJM1-QQR, the sequence encoding the zinc fingers was excised from the Zif-QQR-FN (L0) plasmid (Bibikova et al., 2001) by NdeI + SpeI digestion and placed in the pJM1 backbone. To create pJM1-NwA and pJM1-NwB, the sequences for complete sets of zinc fingers were synthesized by PCR from four overlapping oligonucleotides (Segal, 2002). The PCR products were then digested with NdeI and SpeI and placed in the pJM1 backbone.

The pJM2-QQRt plasmid was made by annealing the complementary oligonucleotides (5'-CTAGCTTCTTCCCCACGCGTGGGGAAGAA-3' and 5'-AGCTTTTCTTCCCCACGCGTGGGGAAGAAG-3') to create inverted QQR binding sites separated by an MluI site and flanked by sequences compatible to HindIII and Nhel cut sites. The annealed fragment was ligated into HindIII + Nhel-digested pLA440, originally constructed by Andrew Fire.
The studies described in this chapter were performed using the wild-type N2 nematode strain. These nematodes were transformed by gonad injection, as described (Mello et al., 1991). Plasmid DNA was linearized with Scal prior to injection. A DNA mixture of pJM1-QQR (5 ng/µl), pJM2-QQRt (15 ng/µl), pPD118.33 (Pmyo-2::GFP; 1 ng/µl), and 1-kb ladder (80 ng/µl; Invitrogen, Carlsbad, CA) was used to produce the strain EG3526 oxEx654 that carries both the QQR expression construct and the QQR target. A DNA mixture of pJM1-NwA (5 ng/µl), pJM1-NwB (5 ng/µl), pPD118.33 (1 ng/µl), and 1-kb ladder (100 ng/µl) was used to create the strain EG3839 oxEx706 that carries both the NwA and NwB ZFN expression constructs.

To induce ZFN expression, L2-L3 transgenic nematodes were heat shocked in a 35°C water bath for 1 hour on two consecutive days, then allowed to recover at room temperature for several hours before further analysis. Each nematode was then frozen at -80°C in lysis buffer [50 mM KCl, 10 mM Tris•HCl (pH 8.3), 2.5 mM MgCl₂, 0.45% Nonidet P-40, 0.45% Tween-20, 0.01% gelatin, 200 µg/ml Proteinase K], then lysed at 65°C for 1 hour and 95°C for 15 minutes. The appropriate ZFN target was then amplified by PCR using Taq polymerase (New England Biolabs, Beverly, MA). To recover the 1006 bp QQR targets, the primers 5’-AACCGGGCGGTGGTACCTACAGCGTGAGC-3’ and 5’-TTAACCAGGTGCTGAGTGCCTGAAACGCAC-3’ were used. To recover the 752 bp Nowhere target, the primers 5’-AACCGGGAATTCCCAATCTATTTTTCTGTGTAACGTG-3’ and 5’-AATTCCCGGATCCACAAAATTTGGCTTTTCTTGTAACC-3’ were used.

The amplified target DNA was purified using MinElute columns (Qiagen, Valencia, CA), digested with MluI (QQR target) or HindIII (Nowhere target), and
analyzed by gel electrophoresis in 1.5 or 2% agarose gels. Target DNA resistant to restriction digest was excised from the gel, purified on a MinElute column, and reamplified by PCR. It was then purified on a MinElute column and cloned for analysis of individual target mutations. When the mutation frequency at each of the targets was examined, digestion and subsequent reamplification of the mutant DNA sequences were not performed. Amplified target DNA was purified and cloned into the appropriate vector without prior manipulation.

The QQR target DNA was ligated into the pGEM-T vector backbone (Promega, Madison, WI). The Nowhere target DNA was digested with EcoRI and BamHI, sites for which had been introduced by the primers used for amplification (see above, in bold), and ligated into a pBluescript backbone previously digested with the same enzymes. The ligation mixtures were transformed into E. coli SS320 electrocompotent cells, which were incubated overnight at 37°C on plates containing 100 μg/ml ampicillin. Transformant colonies were screened by colony PCR (Sambrook, 2001) and restriction digest. Those identified as containing mutant target sequences were cultured, and their plasmid DNA was purified with Qiagen columns and sequenced at the University of Utah Core Sequencing Center.

Results

As a preliminary examination of the efficacy of the ZFN system in nematodes, a somatic assay was designed to evaluate the ability of a previously characterized ZFN, QQR, to stimulate a DSB at a synthetic target. Since QQR has been repeatedly shown to bind and cleave the DNA sequence 5’-GGGGAAGAA-3’ (Bibikova et al., 2001;
Kim et al., 1997; Lloyd et al., 2005; Porteus and Baltimore, 2003; Shi and Berg, 1995; Smith et al., 2000), an artificial target was constructed, containing two inverted repeats of the QQR binding site, separated by the 6-bp MluI restriction site (Fig. 2-1). In this arrangement, cleavage by QQR and subsequent NHEJ-mediated repair of the target can introduce mutations at the break point. These mutations can be identified by loss of the diagnostic MluI restriction site. The QQR nuclease was placed in a vector designed to regulate its inducible expression in nematodes (Bessereau et al., 2001). In this context, ZFN expression was driven by the C. elegans 16-48 heatshock promoter (Candido et al., 1989), followed by a four-adenosine consensus translation initiator sequence (Riddle, 1997). An SV40 large T-antigen nuclear localization sequence was also added to the N-terminus of this construct (Adam et al., 1989; Kalderon et al., 1984) (Fig. 2-2).

Once constructed, the plasmids containing the inducible QQR nuclease, its target, and a GFP marker were linearized and injected into the distal gonads of young adult nematodes, using standard techniques (Mello et al., 1991; Stinchcomb et al., 1985). Assembly of these components in the nematode gonad produces an extrachromosomal array containing multiple copies of both the QQR-encoding sequences and its target (see Fig. 2-1). Transgenic offspring were isolated and separately propagated to identify stable transformant lines. The presence of both QQR and its target was then verified by PCR (data not shown.) Several L2-L3 larvae from a stable transgenic line were heatshocked at 35° C for one hour on two consecutive days to induce expression of QQR. After the nematodes had recovered for several hours from their final heatshock, a 1-kb region of the QQR target DNA was amplified from single nematodes using PCR (Morton et al., 2006).
A construct containing two inverted QQR binding sites (in red), separated by the 6-bp MluI diagnostic restriction site (in blue), was introduced together with the QQR ZFN in an extrachromosomal array to assay for QQR activity in nematode somatic tissue. After heatshock, QQR is expressed and can cleave its target. Inaccurate repair of the target will destroy the restriction site, rendering target mutations observable by PCR followed by restriction digest.
Figure 2-2: Representation of the nematode heatshock construct.

The DNA and protein sequences of the QQR nuclease in the nematode expression vector are shown. The 16-48 heatshock promoter is in lowercase. The four-adenosine consensus sequence is in lowercase gray. The nuclear localization sequence is underlined. The specificity-determining residues of the three zinc fingers of QQR in its modified Sp1 consensus backbone (black) are bolded. The FokI nuclease domain is underlined with dashes.
DNA Sequence:
catgattgtagttgagtttgaagatttcacaattagagtgaatggtcattcctttag
aacattcagtcctcctttgcaaaagggggeactcacaattcagaaatattgttttggttactgaagaacccaga
tactttttcacttctccttttgaactfaggtttgtatttgaatgtcattcagacccctttagaacattcacaacgtgc
gagatgcggtctataacatgtctttgcacctatggggtgtattttgaaatgaatgcatctaggacccctttagaacattcacaacgtgc
gagatgcggtctataacatgtctttgcacctatggggtgtattttgaaatgaatgcatctaggacccctttagaacattcacaacgtgc

Protein sequence:
MGSDPKKRRKVHMEKLNRSGGDIPKDKQHACPECAGKSSFSQSSLNKHQQRSTGKPYKCPECGKSFSSQSLQKH
QRTHTGEKPYKCEPCGKSFSSRDHLSRHRQTHQNKKQL
F*
Amplified target DNA from several nematodes was subsequently digested with $MluI$ and separated by gel electrophoresis to look for the presence of an $MluI$-resistant population of target molecules. A picture of the gel is shown in Fig. 2-3. The DNA amplified from an unheatshocked transgenic nematode (noHS) and the QQR target plasmid (+) cut basically to completion, while the QQR target DNA amplified from each of the heatshocked nematodes (1-5) contained a large proportion of $MluI$-resistant molecules. Thus, it appears that QQR is active in nematode somatic tissue.

In order to examine the $MluI$-resistant targets more closely, QQR target DNA amplified from additional heatshocked nematodes was ligated into pGEM-T© plasmid backbone and cloned into bacteria so that individual mutations could be inspected. The QQR target DNA from randomly selected transformed bacterial colonies was analyzed by colony PCR (Sambrook, 2001) and restriction digest to determine which of the QQR targets had been mutated. Bacteria containing an $MluI$-resistant target were cultured to isolate the plasmid DNA, which was sequenced to determine the specific type of mutation it carried. By this process, it was determined that 26% of the QQR targets harbored some mutation. A complete list of the mutations identified is reported in Table 2-1. A wide range of insertions, deletions, substitutions, and complex mutations, containing a combination of features observed in other types of mutation, were catalogued.

After determining that ZFNs could stimulate DSBs at a synthetic locus in the somatic tissues of nematodes, we next searched the nematode genome for a promising chromosomal locus to target with ZFNs. Since the modular nature of the GNN-binding
The QQR target is 1006 bp long and contains a diagnostic *MluI* restriction site at position 548. In both an unheatshocked nematode and a plasmid control, digestion with *MluI* cleaves the QQR target basically to completion, creating fragments of 546 and 460 bp. In the heatshocked nematodes, a significant proportion of the QQR target DNA has been rendered resistant to *MluI* digestion, indicating that mutation has eliminated the restriction site from the target: L = DNA ladder, noHS = unheatshocked QQR nematode, + = QQR target plasmid (WT), 1-5 = heatshocked QQR nematodes. Figure adapted from (Morton et al., 2006). Copyright 2006. The National Academy of Sciences of the United States of America.
These mutations were sequenced from the nematode QQR target DNA cloned into bacteria. The majority of the characterized mutations consist of 4-bp insertions, indicating that blunt-ended DNA is preferentially joined by NHEJ in the extrachromosomal arrays. The remaining mutations are mostly deletions, which may or may not be the products of blunt-ended ligation.

Table 2-1: QQR target mutations

<table>
<thead>
<tr>
<th>QQR Target</th>
<th>TCCTCCCCC ACCGC CGT GGGGAAGAA</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insertions</td>
<td>ACGCCGCGTG</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>ACGCGCGTG</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ACGCCGCGTG</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ACGCGCGTG</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ACGCGCGTG</td>
<td>1</td>
</tr>
<tr>
<td>Deletions</td>
<td>A-GCGT</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>374bp &lt;-- ------ --&gt; 239bp</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>256bp &lt;-- ------ --&gt; 18bp</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>243bp &lt;-- ------ --&gt; 223bp</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>62bp &lt;-- ------ --&gt; 403bp</td>
<td>1</td>
</tr>
<tr>
<td>Substitutions</td>
<td>TCGCGT</td>
<td>1</td>
</tr>
<tr>
<td>Complex</td>
<td>490bp &lt;-- ------ --&gt; A 109bp</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>T-C CCC ACGCG- --&gt; 126bp</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>TCC ACGCGCGGA</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>CC- ACGCGCGTG</td>
<td>1</td>
</tr>
</tbody>
</table>
zinc fingers was well characterized (Dreier et al., 2000; Liu et al., 2002; Segal et al., 1999) and prior success in Drosophila had been achieved with GNN-binding fingers (Beumer et al., 2006; Bibikova et al., 2003; Bibikova et al., 2002), we restricted our search to the sequences: 5'- NNC NNC NNC NNNNNN GNN GNN GNN - 3' in which the central 6-N repeat consisted of a well-characterized restriction site. These target characteristics would enable us to create two new ZFNs which bind to three consecutive GNN triplets and whose nuclease domains could dimerize and cleave within the 6-bp restriction site separating the ZFN binding sites. Such an orientation produces maximal ZFN activity and, when the induced DSB is repaired by NHEJ, any resultant mutations in the repair process would destroy the restriction site and allow the mutation to be easily identified, as in the QQR assay previously performed.

After a computer-assisted search of the nematode genome, many potential targets on each of the six chromosomes were identified. Eventually, a target, beginning at nucleotide 3008453 of the X chromosome, was chosen www.wormbase.org (Fig. 2-4). This target consists of two zinc finger binding sites, each composed entirely of consecutive GNN triplets, which are separated by a HindIII restriction site. Although an additional HindIII site resides several hundred nucleotides from the target, it is far enough away that its presence should not interfere with an analysis of ZFN-induced mutations at this target. Additionally, this target site is located over 1 kb away from the nearest putative transcribed region of the X chromosome. Since this assay does not rely on any nematode phenotype, the location of the target is not important. It must, however, be accessible to ZFN binding and cleavage to be of value.
Figure 2-4: Nowhere target sequence

This DNA sequence is located between nucleotides 3008453 – 3008476 of the X chromosome. The sequences to which NwA was constructed to bind are in purple (top strand). The sequences to which NwB was constructed to bind are in pink (bottom strand). The naturally occurring HindIII restriction site is in blue. Cleavage and inaccurate repair of the target will destroy the HindIII site, allowing mutant targets to be identified by PCR and restriction digest.
In several previous studies, when zinc finger transcription factors were employed to modulate expression of a nearby gene, they were discovered to be less effective in regions of compact chromatin (Liu et al., 2001; Urnov, 2002; Urnov et al., 2002; Zhang et al., 2000). However, other experiments using zinc finger transcription factors successfully modulated target expression without regard to local chromatin structure (Beerli et al., 2000; Dreier et al., 2001). Additionally, ZFNs in both *Xenopus* oocytes and *Drosophila* have not indicated that the chromatin state of the target prohibits ZFN activity (Bibikova et al., 2001; Bibikova et al., 2002). On this basis, we decided to construct novel ZFNs capable of stimulating DSBs at this target, which we named “Nowhere” (Nw), due to its apparent isolation.

Two new ZFNs were constructed, each containing the correct finger-determining sequences in a Zif268 backbone (Segal et al., 2003). Although this backbone is different from the modified Sp1 backbone of QQR, prior successes using ZFNs containing this backbone prompted its inclusion in these ZFNs (Beumer et al., 2006; Bibikova et al., 2003; Bibikova et al., 2002). Each construct was subsequently amplified by PCR using primers corresponding to the ends of its sequence and containing restriction sites that would allow the fingers to replace the QQR fingers in the nematode heatshock vector. (See Fig. 2-5 for the zinc finger coding regions of the two ZFNs, as well as the DNA and protein sequences of the Zif268 backbone, and an alignment of the QQR and NwA zinc finger domains.) The complete zinc finger coding sequence for each ZFN was ligated into the previously created nematode expression vector (see Fig. 2-2), which was then cloned and sequenced to verify its identity. The new ZFNs were named “Nowhere A” (NwA) and “Nowhere B” (NwB).
The amino acids from the specificity-determining regions (α-helix positions -1, 1, 2, 3, 4, 5, 6) for the zinc fingers of NwA and NwB are shown in the box at the top. Since the zinc fingers bind anti-parallel to the DNA, finger 1 of each protein binds to the 3' triplet of its target. These fingers are in a Zif268 backbone, different in sequence to that used in QQR, shown in Fig. 2-2. The DNA and amino acid sequences of this backbone, as well as an alignment of the QQR and NwA zinc finger regions, are also shown.
DNA sequence of the Zif 268 backbone:

ATGGAGCCGTATGCTTGCCCTGTCGAGTCCTGCGATCGACGATTTTCT
CATATCCGAATCCACACAGGAGCTTAACATCAGCTTTTC
CACATGCCAATGCTCAACTCCACACAGGAGCTTAACATCAGCTTTTC
AAGAAAGTTTTCC

Protein sequence of the Zif268 backbone:

MEPYACPVESCDRRFS

Alignment of the QQR and NwA zinc finger domains:

QQR: MEKLRNGSGDPGKKQHKACPECCKSF
NwA: MEPYACPVESCDRRFS

<table>
<thead>
<tr>
<th>ZFN</th>
<th>Finger 1</th>
<th>Finger 2</th>
<th>Finger 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nowhere A</td>
<td>QSSDLTR</td>
<td>RSDALTR</td>
<td>RSDHLSR</td>
</tr>
<tr>
<td>Nowhere B</td>
<td>TSGHLVR</td>
<td>RSDALTR</td>
<td>QSGHHLQR</td>
</tr>
</tbody>
</table>
These new ZFN-expression vectors and a GFP marker were linearized and injected into the nematode gonad to form an extrachromosomal array. Transformant nematodes were selected and stable lines isolated. When the presence of both the NwA and NwB constructs had been verified by PCR (data not shown), several L2-L3 nematodes were heatshocked at 35°C for one hour on two consecutive days to express the ZFNs. After the final heatshock, the nematodes were then allowed to recover at room temperature for several hours, and the Nowhere target DNA was amplified by PCR (Morton et al., 2006).

Initially, the activity of the ZFNs was verified by digesting the Nowhere target with HindIII and separating it via gel electrophoresis to detect a HindIII-resistant DNA population. After several preliminary attempts at this failed, it was discovered that there were several Nowhere target homologs scattered throughout the genome www.wormbase.org (Table 2-2). Although none of these homologs contains both the NwA and NwB binding sites, or the HindIII site, they are similar enough to prevent specific amplification of the genuine Nowhere target using the initial PCR primers chosen. After alignment of these homologs, a second set of PCR primers was chosen that was more specific. Gels of Nowhere target amplified from heatshocked nematodes with these primers and then cut by HindIII, clearly show that heatshocked nematodes contain a greater proportion of HindIII-resistant target DNA than do transgenic nematodes which were not heatshocked to activate NwA and NwB expression (Fig. 2-6.) Thus, it appears as though ZFNs are also active at a genomic target.

In order to examine the Nowhere target mutations more closely, targets from additional heatshocked nematodes were amplified, ligated into a pBluescript backbone,
Table 2-2: Nowhere target homologs

<table>
<thead>
<tr>
<th>Target</th>
<th>Location (Chromosome: Nucleotide)</th>
<th>Target sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nowhere</td>
<td>X: 3,008,453</td>
<td>ACC CAC TCC AAGCTT GGG GTG GCT</td>
</tr>
<tr>
<td>Homolog 1</td>
<td>I: 11,508,156</td>
<td>ATC CTT TCC AAGTCTT GGG GTG GGG</td>
</tr>
<tr>
<td>Homolog 2</td>
<td>III: 3,192,626</td>
<td>ATC CTT TCC AAGTCTT GGG GTG GCT</td>
</tr>
<tr>
<td>Homolog 3</td>
<td>IV: 15,186,903</td>
<td>ATC CTT TCC AAGTCTT GGG GTG GCT</td>
</tr>
<tr>
<td>Homolog 4</td>
<td>V: 5,334,292</td>
<td>ATC CTT TCC AAGTCTT GAG GTG GCT</td>
</tr>
<tr>
<td>Homolog 5</td>
<td>X: 8,842,611</td>
<td>ATC CTT TCC AAGTCTT GAG GTG GCT</td>
</tr>
</tbody>
</table>

The genomic location and target site sequence of the Nowhere target and its five homologs. Sequence differences at the zinc finger binding sites and the diagnostic HindIII restriction site are highlighted in red. [www.wormbase.org](http://www.wormbase.org), release WS162.
The target is 741 bp long, and the restriction site is located at base 288. Cleavage by HindIII produced bands of 288 and 453 bp. The faint HindIII-resistant bands on the unheatshocked nematodes are presumably the product of amplification of the Nowhere target homologs. The proportion of HindIII-resistant DNA is noticeably greater in the heatshocked nematodes, indicating that the Nowhere ZFNs induced mutations at the target that eliminated the diagnostic restriction site: L = DNA ladder, noHS = unheatshocked nematodes, 1-2 heatshocked NwAB nematodes. Figure adapted from (Morton et al., 2006). Copyright 2006. The National Academy of Sciences of the United States of America.
and cloned. Nowhere targets containing *HindIII* mutations were identified by PCR (Sambrook, 2001). By this process, it was determined that 18% of the Nowhere targets harbored some mutation. The bacteria containing these targets were cultured, and their plasmid DNA isolated and sequenced to determine the precise nature of the mutation at the Nowhere target.

To supplement the mutations observed in this experiment, a separate analysis, in which Nowhere target DNA was enriched for mutations by *HindIII* digestion prior to cloning, was also performed. A complete list of all the mutations is given in Table 2-3. Although some of the 4-bp insertions seen among the QQR mutants were observed, the majority of the mutations categorized consisted of either small or large deletions. These mutations also indicate that the NwA and NwB ZFNs also appear to be quite active in the somatic tissues – even at a chromosomal locus.

**Discussion**

The QQR analysis described above provides two types of information about ZFN activity in the somatic tissues of the nematode. First, the percentage of the target sequences containing a mutation is indicative of the relative activity of QQR in the nematode. Of the 167 targets analyzed from four worms, 43 contained some mutation, implying that 26% of the target sites in the nematode were cleaved, repaired by NHEJ, and, in the process, underwent some type of mutation. By comparison, of the 154 targets examined among bacteria containing unenriched Nowhere targets, 27 contained a mutation either at or spanning the *HindIII* restriction site. This implies that 18% of the Nowhere targets are cleaved by the NwA, NwB ZFN heterodimer and inaccurately
These mutations were sequenced from the nematode Nowhere target cloned into bacteria. Although several of the characterized mutations are of 4-bp insertions, other types of mutations predominate, indicating that fill-in of the staggered ends and subsequent blunt-ended ligation is not the preferential repair pathway at a chromosomal locus. Many of the mutations are deletions, implying that extensive chew-back of the DNA often occurs before NHEJ repairs the broken ends.

<table>
<thead>
<tr>
<th>Nowhere target</th>
<th>ACCCACTCC A^TAGCTT GGGGTGGCT</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Insertions</strong></td>
<td>AAGC TAGCTT</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>AAGC TGCTT</td>
<td>1</td>
</tr>
<tr>
<td><strong>Deletions</strong></td>
<td>-AGCTT</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>AAG--T</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>AAGC T</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>A---TT</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>-AGCTT</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>AAG--</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>- ---</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>AAGC T --&gt; 53bp</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>-- --- ---</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>--&gt; 44bp</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>56bp &lt;= ---CTT</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>66bp &lt;= -AGCTT</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>11bp &lt;= ------ --&gt; 80bp</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>40bp &lt;= ------ --&gt; 83bp</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>211bp &lt;= ------TT</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>A----- --&gt; 238bp</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>AAGC T --&gt; 248bp</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>AAGC --&gt; 253bp</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>276bp &lt;= -AGCTT</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>297bp &lt;= -AGCTT</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>326bp &lt;= ------TT</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>390bp &lt;= ------CTT</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>424bp &lt;= ------CTT</td>
<td>1</td>
</tr>
<tr>
<td><strong>Substitutions</strong></td>
<td>AAGCCT</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>AAGCTC</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>AGGCTT</td>
<td>1</td>
</tr>
<tr>
<td><strong>Complex</strong></td>
<td>AAG--T GGA</td>
<td>1</td>
</tr>
</tbody>
</table>
repaired via NHEJ. A statistical analysis of the apparent activity of the ZFNs at this locus and at QQR locus indicates that there is no significant difference between their ability to induce mutations at these targets (p = 0.10).

These frequencies, although impressive, actually underestimate the true activity of ZFNs in nematode somatic tissues. When a target is cleaved by ZFNs, in addition to being mutated via NHEJ repair, it can be accurately repaired by either ligation or HR. In this latter case, the target could then be potentially recleaved by the ZFNs, indicating that the high frequency of mutation observed could be the product of repeated cleavage-repair cycles, which would be interrupted only by the creation of a mutation at the break site that prevented subsequent cleavage. In other cases, accurate repair of a cleaved target may never result in mutation, so the activity of the ZFNs are not accurately represented in the mutation frequency of the target site.

The second type of information produced by this experiment regards the types of mutations induced by the ZFNs at their target sites. Of the 43 mutations characterized during analysis of the QQR target, 26 (or 60%) consisted of a 4-bp insertion at the break site. This can most easily be explained as arising from a fill-in of the 4-bp staggered single-strand 5’ overhangs produced by QQR-induced cleavage of each strand (see the ZFN section of Chapter 1 and (Kim et al., 1996; Smith et al., 2000)), followed by ligation of the now blunt-ended DNA molecules. This method of repair seems to be overwhelmingly favored at this synthetic target. This may be because most other repair events are processed through the HR pathway. It has been shown that extrachromosomal arrays arise primarily by homologous recombination and that DSBs induced in the arrays will further stimulate HR between identical sequences (Mello et
al., 1991). If a ZFN-induced DSB is chewed back from the break site, the ends can search for homologous DNA sequences to act as a template for HR-mediated repair. Since the extrachromosomal array in which the QQR target is located contains many additional copies of the target, templates for repair are common and many breaks are probably repaired perfectly by this process.

Alternatively, the broken ends can be directly ligated by the NHEJ components. Since several deletions were also recovered, traditional NHEJ-mediated repair must also play a role in repairing the break. There is no way of knowing what form the ends of the DNA molecules take, but an overall absence of microhomologies flanking these deletions may be an indication that these mutations also arose from blunt-end ligation.

Of interest, and possible relevance, is the fact that, when DNA is injected into nematode gonads, it is often linearized with a blunt-cutting restriction enzyme, such as EcoRV, PvuII, or Scal, so that plasmids will be evenly distributed throughout the array. If circular plasmids are injected, they are concatenated into arrays through repeated cycles of HR (Kelly et al., 1997). Since HR so efficiently incorporates DNA into arrays, it is possible that DSBs with staggered ends, which may often result in extended single-stranded overhangs, are preferentially repaired by this pathway. For this reason, many of the mutations observed at the QQR target site may be a consequence of the ligation of blunt-ended molecules.

An examination of the types of mutations characterized at the Nowhere target indicates that, even though the QQR and NwA and NwB ZFNs appear to induce mutations at about the same frequency, the types of mutations to which they give rise are markedly different. At the QQR target, 60% of mutations observed were 4-bp
insertions. At the Nowhere target 4-bp insertions accounted for only 6 of 54, or 11%, of the mutations characterized. This type of mutation, although seen, does not predominate at the Nowhere target. Instead, large deletions, spanning all or some of the restriction site, account for the majority of the observed mutations.

These differences may highlight the different chromosomal context of each target and the relative availability of the components of the two major DNA repair pathways at these locations. Whereas the abundance of homologous templates allowed HR to play a major role on an extrachromosomal array, the availability of only a single additional copy of the Nowhere target on the homologous X chromosome limits the frequency at which HR can be employed to repair a break at the Nowhere locus. Instead, after a break is induced, the ends are chewed back as they attempt to locate a homologous template. If the copy is found, HR can be used to repair the broken DNA. However, if no intact template can be located, as may occur if both copies of the Nowhere target are broken, the broken ends of the DNA are simply forced back together by the NHEJ machinery, creating the frequently seen extended deletions. As observed at the QQR target, the 4-bp insertions at the Nowhere target arise from fill-in of the single-strand overhangs generated by the ZFNs and ligation of the blunted ends by the NHEJ machinery. It seems reasonable to assume that this type of mutation should arise with equal frequency at both loci examined. In this case, they are simply not observed as frequently at the Nowhere target because other NHEJ-mediated events (which cause mutations) are more common, whereas at the QQR target, HR-mediated events (which do not cause mutations) predominate.
This leads to a final question about the frequencies of mutations observed at the two loci. If nonmutagenic HR predominates at the QQR target and mutagenic NHEJ is more active at the Nowhere target, why are the targeting frequencies at the two targets so similar? One possibility is that the homodimer QQR is more active than the heterodimer NwA and NwB ZFNs. In this case, even a lower percentage of NHEJ-induced mutations at the synthetic target may yield a higher apparent activity for the ZFN. Alternatively, it is possible that some of the observed 4-bp insertions at the QQR target are actually the result of HR-mediated repair. Since the ZFNs at the target site are active over an undefined period, DSBs are almost certainly produced and repaired at different targets at different times. If an initial DSB is blunted and ligated to produce a 4-bp insertion at the target, it is possible that this mutated target site could be used as the template for a subsequent HR-mediated repair event. Such a mechanism would increase the apparent frequency of the NHEJ-mediated insertion, even though HR would remain the main repair pathway at the target. Thus, it is possible that the two different sets of ZFNs do have similar activity in nematode somatic tissue, although they generate different types of mutations through diverse mechanisms.
References


Segal, D. J. (2002). The use of zinc finger peptides to study the role of specific factor binding sites in the chromatin environment. Methods 26, 76-83.


CHAPTER 3

THE ROLE OF THE NHEJ MACHINERY
IN ZFN-INDUCED DSB REPAIR

Introduction

The different types of mutations characterized in the initial experiments examining ZFN activity in nematode somatic tissues indicated that there is a competition between DSB repair pathways at both the synthetic and genomic targets. To examine more fully the nature of this interplay, we designed a series of experiments to closely examine the role of the NHEJ machinery in repairing ZFN-induced DSBs at these loci. For these experiments, we chose to examine the effects of expressing the ZFNs in nematodes lacking the NHEJ component DNA ligase IV. The absence of this enzyme should completely inactivate the classical NHEJ pathway in the nematodes (Clejan et al., 2006). (See Fig. 1-11.) The fact that the ligase IV-deficient (lig-4) nematodes do not have an obvious phenotype implies that there may be some redundant mechanism that can repair spontaneously occurring DSBs in the organisms. In this case, these experiments may also help elucidate this alternate NHEJ pathway.
Materials and methods

In addition to the N2 nematodes described previously, experiments were performed in the DNA ligase IV mutant strain RB873 *lig-4(ok716)* III and in the *ku80* mutant strain RB964 *cku-80(ok861)* III. A DNA mixture of pJM1-QQR (5 ng/µl), pJM2-QQRt (15 ng/µl), pPD118.33 (*Pmyo-2::GFP*; 1 ng/µl), and 1-kb ladder (80 ng/µl; Invitrogen, Carlsbad, CA) was injected into RB873 to produce the strain EG3521 *lig-4(ok716)* III; *oxEx653* that carries both the QQR expression construct and the QQR target. A DNA mixture of pJM1-NwA (5 ng/µl), pJM1-NwB (5 ng/µl), pPD118.33 (1 ng/µl), and 1-kb ladder (100 ng/µl) was used to create the strain EG4315 *lig-4(ok716)* III; *oxEx827* that carries the NwA and NwB ZFN expression constructs.

Additionally, N2 males were crossed with hermaphrodites from the EG3526 *oxEx654* strain to produce male offspring containing the QQR expression construct and its target. Transgenic male offspring were crossed to RB964 hermaphrodites and transgenic progeny were transferred to new plates and allowed to reproduce. Single transgenic F2 offspring were then separated and allowed to self-propagate. A transgenic line, homozygous for the *cku-80(ok861)* allele was identified by PCR, using the primers 5’-CTCGAGTTGCGAGAATTGGCAG-3’ and 5’-CCACCTTCGACCAGTGAGAGAGG-3’, which amplify a 517-bp product from the *cku-80(ok861)* allele and an 1848-bp product from the N2 allele.

Analysis of the transgenic nematodes was conducted as previously described, except that the primers 5’-CCTGGCTTATCGAAATTAATACGA-3’ and 5’-CTATAGGGCGAATTGGGTACC-3’ were used to amplify a 207-bp fragment from the QQR target. When individual target mutations were examined at this short target, it
was digested with NotI + XhoI, which produces an 88-bp fragment in the unmodified target, and ligated into a similarly digested pBluescript backbone.

**Results**

An experiment examining QQR activity at the QQR target (similar to the previous QQR experiment performed) was conducted in both wild-type and lig-4

nematodes. In this case, a shorter section of the target (207 bp) was examined. When this short QQR target was amplified from several heatshocked wild-type nematodes, digested with MluI, and visualized by gel electrophoresis, a proportion of MluI-resistant DNA consistent with that seen at the full-length QQR target was observed (Fig. 3-1; compare with Fig. 2-3.) Several L2-L3 lig-4

nematodes from a stable transgenic line containing QQR and its target were also heatshocked. After identical processing, the QQR targets from these lig-4

nematodes yielded only a very small proportion of MluI-resistant DNA (Fig. 3-2.) This indicates that, in the absence of ligase IV, very little of the target is processed by a mutagenic repair mechanism.

In order to identify individual QQR target mutations, targets from additional heatshocked wild-type and lig-4

nematodes were amplified, and the 88-bp section spanning the ZFN binding sites was ligated into a pBluescript backbone and cloned. From each group, targets containing MluI mutations were identified by colony PCR (Sambrook, 2001) and restriction digest. Among targets from the wild-type nematodes, 11% contained a mutation, whereas in the targets amplified from lig-4

nematodes, only 0.4% harbored a mutation. These bacteria were cultured, and their plasmid DNA isolated and sequenced to determine the precise nature of the mutations. Since a very
The target is 230 bp long, and the *MluI* restriction site is located at base 145. Cleavage at this site results in 145 and 85 bp DNA fragments. A plasmid containing the QQR target (+) has been cut to completion, while a substantial proportion of the QQR target amplified heatshocked nematodes has become *MluI*-resistant, indicating that inaccurate repair of QQR-induced cleavage has induced mutations at its target site: L = DNA ladder, + = QQR target plasmid (WT), - = N2 nematode, included to validate the specificity of QQR target amplification, 1-5 = heatshocked QQR nematodes. Figure adapted from (Morton et al., 2006) Copyright 2006. The National Academy of Sciences of the United States of America.
The target is 230 bp long, and the MluI restriction site is located at base 145. Cleavage at this site results in 145 and 85 bp DNA fragments. A plasmid containing the QQR target has been cut to completion, and only a very small proportion of the targets from the lig-4 nematodes have become MluI-resistant, indicating that, in the absence of ligase IV, little mutagenic repair of the target occurs: L = DNA ladder, + = QQR target plasmid (WT), 1-9 = heatshocked lig-4 QQR nematodes. Figure adapted from (Morton et al., 2006) Copyright 2006. The National Academy of Sciences of the United States of America.
small proportion of the QQR targets from the lig-4 nematodes appeared to be MluI-resistant, a separate analysis, in which target DNA was enriched for mutations by MluI digestion prior to cloning, was also performed in both wild-type and lig4 nematodes.

A complete list of the mutations identified in targets from wild-type nematodes is reported in Table 3-1. Much as seen at the longer QQR target, a majority of the mutations were 4-bp insertions, although a variety of substitutions, deletions, and complex mutations were also observed. A list of the mutations characterized from targets in lig-4 nematodes is reported in Table 3-2. In this case, no insertions were observed, and most mutations were substitutions, deletions, and complex mutations.

From these results, it is clear that lig-4 nematodes process the ZFN-induced DSBs on an extrachromosomal array in a very different fashion than wild-type nematodes. Since previous work had indicated that DSBs are processed differently on an array and at a genomic locus, we conducted a third set of experiments to examine the types of mutations observed at a genomic locus in lig-4 nematodes.

The NwA and NwB plasmids were injected into lig-4 nematodes, and larvae from a stable transgenic line were heatshocked to activate ZFN expression. The Nowhere target DNA was subsequently amplified from these nematodes, digested with HindIII, and examined by gel electrophoresis to estimate the proportion of Nowhere targets in which a mutation had destroyed this restriction site (Fig. 3-3). (Compare this with Fig. 2-7, which shows the identical experiment in wild-type nematodes, where it was experimentally determined that 18% of the Nowhere targets had been mutated. In this case, it appears as though only 5-10% of the Nowhere target is uncut by HindIII.) Much as was witnessed at the QQR target, a smaller proportion of HindIII-resistant
Table 3-1: Short QQR target mutations in wild-type nematodes

<table>
<thead>
<tr>
<th>QQR target</th>
<th>TTCTTCCCC A^CGCGT GGGGAAAGAA</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Insertions</strong></td>
<td>ACGCCGCGGT</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>ACGCGCGGT</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>ACGCGCGGT</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>ACGCGCGGT</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>ACGCAGCGGT</td>
<td>1</td>
</tr>
<tr>
<td><strong>Deletions</strong></td>
<td>-CGCGT</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>ACGC-T</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>ACGG-T</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>--GCGT</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>-----T</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>---- -----T</td>
<td>1</td>
</tr>
<tr>
<td><strong>Substitutions</strong></td>
<td>ACGCGC</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>GCCTCGG</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>ATGCCT</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ACACGT</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ACGCAT</td>
<td>1</td>
</tr>
<tr>
<td><strong>Complex</strong></td>
<td>---- CT</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>CCT ACGCCGCGGT</td>
<td>1</td>
</tr>
</tbody>
</table>

These mutations were sequenced from the short QQR target DNA cloned into bacteria. As previously observed at the full-length version of this target, a majority of the characterized mutations consist of 4-bp insertions, indicating that blunt-ended DNA is preferentially joined by NHEJ in the extrachromosomal arrays. The remaining mutations include mainly deletions and substitutions, which may arise by blunt-ended ligation or an alternative repair pathway.
These mutations were sequenced from the short QQR target DNA amplified from *lig-4* nematodes and cloned into bacteria. There are no insertions, implying that ligase IV may be strictly necessary for the ligation of blunt-ended DNA. The mutations consist entirely of small deletions and substitutions, as well as several complex mutations. These mutations may arise by an alternate end-joining repair mechanism or may be the result of homologous recombination utilizing an error-prone polymerase.
Full-length target amplified from nematodes is 741 bp, and the HindIII restriction site is located at base 288. Digestion products are 288 and 453 bp. A plasmid containing the QQR target indicates that HindIII digestion has not cut its target site to completion. However, there is still a very small proportion of HindIII-resistant DNA among the Nowhere targets amplified from the lig-4 nematodes: L = DNA ladder, - = blank, included to validate the specificity of Nowhere target amplification, + = Nowhere target plasmid (WT), 1-6 = heatshocked lig-4 NwAB nematodes.
DNA observed in the *lig-4* nematodes indicates that DSB repair events are also significantly altered at a genomic locus.

Once again, to examine the Nowhere target mutations from the *lig-4* nematodes more closely, targets from additional heatshocked nematodes were amplified, ligated into a pBluescript backbone, and cloned. Targets containing *HindIII* mutations were identified by colony PCR (Sambrook, 2001) and restriction digest. The bacteria containing these targets were cultured, and their plasmid DNA isolated and sequenced to determine the precise nature of the mutations. A complete list of all the mutations observed in the *lig-4* nematodes is given in Table 3-3. Overall, the types of mutations seen are similar to those predicted in *lig-4* nematodes, based on observations at the QQR target. Most of these mutations are either substitutions or deletions, although there is a single 4-bp insertion. There are also several complex mutations, containing satellite substitutions.

**Discussion**

Among the 239 sequences examined from bacteria containing unenriched QQR targets amplified from wild-type nematodes, 25 mutations were observed, indicating that 11% of the targets were cleaved by QQR and repaired inaccurately. Although this frequency is much lower than that observed for the full-length target, differences in apparent enzyme activity can be reconciled by realizing that, in the short target PCR, the large deletions commonly seen at the full-length target have been excluded. Furthermore, a majority of mutations characterized from all wild-type nematodes were 4-bp insertions. In total, 89 of the 127 (or 70%) of the mutations resulted from fill-in of
Table 3-3: Nowhere target mutations in lig-4' nematodes

These mutations were sequenced from lig-4' nematode Nowhere target DNA cloned into bacteria. A single 4-bp insertion is present, although a lack of any other insertions implies that blunt-end joining cannot readily occur in the absence of ligase IV. The remaining mutations consist of small deletions and substitutions, along with three complex mutations, which could have arisen by an alternate end-joining process or by homologous recombination via an error-prone polymerase.
the staggered DSB and blunt-end ligation (compared to 60% at the longer QQR target). This increased frequency of 4-bp insertions can also be attributed to the exclusion of large deletions by the primers used to amplify this shorter target.

When 232 sequences were screened from the bacteria containing unenriched QQR targets amplified from lig-4 nematodes, only one mutation was found. This implies that only 0.4% of the targets cleaved by QQR were repaired via a mutagenic pathway. Consistent with the types of mutations characterized from the lig-4 nematodes, this mutation was a substitution of one base in the MluI site. Statistical analysis comparing the frequency of mutations in the two strains indicates that there is a significant difference in the ZFN-induced mutation rate (p = 5 x 10⁻⁶).

Further, among mutations characterized at the QQR target amplified from lig-4 nematodes, no 4-bp insertions were observed, indicating that ligase IV may be absolutely necessary for blunt-end joining, as has been seen in other systems (van Heemst et al., 2004). Most of the mutations consisted of simple substitutions or small deletions at the restriction site. Interestingly, several mutations contained secondary or satellite mutations. In this case, the defining mutation at the MluI site was accompanied by an additional, apparently unrelated, mutation at a point close by. Although such satellite mutations were occasionally seen among sequences from wild-type nematodes at both the full-length and short QQR targets (Tables 2-1 and 3-1), their frequency jumped several-fold in the lig4 nematodes (Table 3-2).

Finally, although an extensive analysis of Nowhere target mutations was not conducted in lig-4 nematodes, it appears as if the types of mutations observed correlate well with those seen in lig-4 nematodes at the QQR target - with one exception.
Among Nowhere target mutations, there was a single 4-bp insertion, indicative of repair of blunt-ended DNA. Although the exact mechanism by which this mutation arose is unclear, the fact that it was only observed once indicates that this repair product is not typical of those produced in the absence of DNA ligase IV.

There are three major differences between the frequencies and types of mutations generated in wild-type and lig-4 nematodes, which may help clarify the role of NHEJ in ZFN-induced DSB repair. First, when examined at the synthetic target, the frequency of mutations recovered from lig-4 nematodes is much lower than seen in wild-type nematodes. This indicates that, in the absence of a functional NHEJ repair pathway, the majority of the DSBs induced by ZFNs must either be repaired by an alternate nonmutagenic pathway or left unrepaired. Second, at the synthetic target, the 4-bp insertions characteristic of blunt-ended DNA ligation are one of the most common mutations observed in wild-type nematodes, while in the lig-4 nematodes, such insertions are completely absent. Thus, it appears that the ligase IV protein is required for the repair of blunt-ended DSBs, consistent with the observations in mammalian cells (van Heemst et al., 2004). Finally, even the types of mutations characterized in both strains of nematodes have different distinguishing features. In the lig-4 nematodes, the deletions and substitutions that constitute the majority of all observed mutations are frequently accompanied by satellite mutations away from the diagnostic restriction site. The deletions are also more frequently bounded by small stretches of microhomology in the target sequence, although additional deletions would need to be recovered to verify this statistically (Table 3-4.) Taken together, however, these
Table 3-4: Microhomologies flanking deletions in wild-type and \textit{lig-4} nematodes

<table>
<thead>
<tr>
<th>Target</th>
<th>Nematode strain</th>
<th>Deletions (&gt; 1bp)</th>
<th>Microhomologies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WT</td>
<td>CCCCACGCGT</td>
<td>1 bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCCCACGCGCTGG</td>
<td>1 bp*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCCCACGCGTG</td>
<td>1 bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CTTCCCACGCGTG</td>
<td>1 bp</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>\textit{lig4}^-</td>
<td>CCCCACGCGT</td>
<td>1 bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TCCCACGCGT</td>
<td>1 bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CGCGTCGGGGAAAG</td>
<td>1 bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCCCACGCGTG</td>
<td>1 bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCCCACGCCGGGGG</td>
<td>1 bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCCCACGCCGTGGG</td>
<td>1 bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTCTCCACGCGTG</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Short QQR</td>
<td>ATAAGCTTTGTGGGGAA</td>
<td>1 bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTTCTTTCAAAGAGCT</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTTCCTCCAAAGTACC</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTCCCTCACCTAGCCAT</td>
<td>3 bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AAGCTTTCTAAGCTAGCC</td>
<td>5 bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCCCACGCGTG</td>
<td>2 bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TAAGCTTTGAAGTACC</td>
<td>2 bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GATAAGCTGCTAGCCA</td>
<td>2 bp</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>15</td>
<td>10</td>
</tr>
</tbody>
</table>

* microhomology based on substitution
Microhomologies flanking deletions catalogued in wild-type and *lig-4* nematodes at the QQR and Nowhere targets. The four nucleotides flanking the deletion are in black. The four terminal nucleotides of the deletion are in gray. Microhomologies are determined by aligning flanking sequences with the terminal deleted sequences, where single-stranded DNA on the opposite strand may exist. Typically, these microhomologies extend from 1-4 nucleotides. Presumably, they form the basis for interactions between single-stranded DNA ends so that ligation or primed synthesis of the broken strands can occur. The interacting DNA ends cannot be blunt for this type of interaction to occur.
features imply that mutations are being created by a different mechanism than that observed in the wild-type nematodes.

There are several possible related mechanisms by which these observed differences may arise in the lig-4 nematodes: First, it is possible that ligase IV is only active in rejoining blunted DNA ends and that other types of joints are sealed by different ligases. When ligase IV is absent, the role of these alternative ligases becomes more apparent. Such a mechanism would readily explain the dramatically reduced mutation frequency in lig-4 nematodes at the QQR target in the array, where much of the repair of the staggered DNA ends may proceed by homologous recombination. In this case, it might also be predicted that the mutation frequency at the Nowhere target in lig-4 nematodes will be higher, since NHEJ-related processes should play a bigger role in repairing staggered DSBs at this locus.

The satellite mutations frequently observed in lig-4 nematodes also support this mechanism. These mutations indicate that repair of the DSB in many of the lig-4 nematodes is somehow associated with error-prone DNA synthesis. This type of synthesis-dependent end joining has been specifically linked to DSB repair in Xenopus nuclear extracts (Lehman et al., 1994). Reports of error-prone polymerases associated with different ligases, such as DNA ligase I, lend additional evidence for such a mechanism (Showalter et al., 2006; Tomkinson et al., 2006). Finally, it has been shown that, when gaps exist at potential joints between two DNA sequences, error-prone polymerases in the PolX family bridge the gaps before ligation occurs (Daley et al., 2005). It is possible that a similar pathway may also be responsible for producing the substitutions seen at the MluI and HindIII sites in the lig-4 nematodes.
If the mutations seen in the lig-4 nematodes result from a secondary, synthesis-dependent DNA repair pathway, this pathway must also be active in wild-type nematodes. Here, however, satellite mutations are rarely observed, indicating that either the polymerase which acts in conjunction with ligase IV-mediated repair faithfully incorporates the correct bases in the DNA strands it synthesizes or that little polymerase activity is necessary when the DNA is repaired via a ligase IV-dependent pathway. However, if all of the insertions thought to result from ligase-IV-dependent repair are discounted, so that only substitutions or deletions of the type seen in lig-4 nematodes are considered, wild-type nematodes yield only six mutations in the 239 sequences examined. This frequency (2.6%) is not statistically different from the frequency (0.4%) seen in lig-4 nematodes (p = 0.14), implying that the same alternative repair pathway may, in fact, be active in the two strains.

The microhomologies often observed in the mutations from lig-4 nematodes (Table 3-4) are also consistent with this alternate repair mechanism. The DNA ends produced by a DSB are often subjected to degradation by exonucleases. During repair by the canonical NHEJ pathway, the DNA ends can simply be blunted and joined by ligase IV together to produce a deletion. In the lig4 nematodes, however, when DNA ligase IV is not available to rejoin the blunt ends, these ends must be repaired by an alternate mechanism. In this case, resection of the 5’strand of DNA continues until a small stretch of microhomology between the overhanging 3’ ends allows them to anneal and transiently stabilize the junction. The small duplex formed by the overhanging strands primes subsequent DNA synthesis to restore the resected DNA,
after which it is repaired by an alternate ligase. This process results in a deletion that is flanked by microhomologies in the molecule.

Similar results have been reported in studies examining DSB repair in the germline of wild-type and *lig-4* nematodes (Morton et al., 2006; Robert and Bessereau, 2006). When the repair products of linear plasmids injected into the syncytial gonad were recovered by PCR and analyzed, it was found that, in wild-type nematodes, the plasmid ends were typically blunted or partially blunted before being joined. In *lig-4* nematodes, however, many of the ends had been resected. These recovered deletions were also frequently flanked by microhomologies. Similar results were also seen in footprints generated by *MosI* excision from the germline, indicating the that same repair processes are active in both somatic and germ tissues.

Although the alternate DNA repair mechanisms proposed above are similar to the canonical NHEJ pathway, it is also possible that homologous recombination plays a role in generating the observed mutations in *lig-4* nematodes. Several error prone polymerases have been identified and specifically associated with the HR repair process (Kawamoto et al., 2005; McIlwraith et al., 2005). When DSBs are created, the broken ends of the DNA are often subjected to exonuclease activity, which can degrade the 5’ strands at broken ends, freeing the 3’ DNA ends to search for a homologous repair template. Once found, they will invade this template to create a heteroduplex, and a polymerase will extend the sequence past the break, after which the DNA strand can withdraw and re-anneal with its complement. If, during extension, an error-prone polymerase creates a mutation, the diagnostic restriction site can be destroyed. The same process can also create the observed satellite mutations. If this process does play
a major role in target repair, it may explain the higher frequency of satellite mutations observed at the QQR target, where HR may play a greater role in DNA repair.

Although the experiments performed have provided tantalizing clues as to the repair process that may be active in the absence of ligase IV, there are many unanswered questions about the mechanism that is actually employed. Of particular interest is the possible role of HR in this process. To examine the role of HR in repairing the ZFN-induced DSBs at the targets, a marked template, homologous to the ZFN target, could be created. When introduced together with the ZFNs and a synthetic target (for QQR), this target would then be available for repair, providing a mechanism to track which types of mutations are actually the products of HR-mediated repair. This experiment would answer many of the remaining questions about how the elevated 4-bp insertion frequency arises in an array and, possibly, in what context the observed satellite mutations in the lig-4 nematodes arise.

**Current model of DSB repair**

The experiments and results described in this and the previous chapter have all provided information about DSB repair in *C. elegans* that can be harmonized in the following model (Fig. 3-4). When a set of ZFNs induces a DSB at its target site in nematode somatic tissue, the lesion can either be repaired by HR, if a homologous template exists, or by NHEJ. It appears as though HR is the preferred repair pathway, since many of the mutations recovered at the synthetic target on an extrachromosomal array, which contains many homologous templates, seemed to contain features indicative of HR-mediated repair. Such a preference has also been previously reported.
Figure 3-4: Model of the ZFN-induced DSB repair pathway in *C. elegans*.

This model describes the three possible fates of ZFN-induced DSBs in nematodes. The DNA ends can be blunted, then joined by DNA ligase IV. The broken ends can be degraded by the repair machinery, then aligned via microhomologies to initiate synthesis of the sequences surrounding the break. Repair by an alternate DNA ligase creates deletions at the DSB site. Finally, the broken ends can be degraded and repaired by homologous recombination, possibly using an error-prone polymerase, resulting in the substitutions seen at the DSB site.
(E. Jorgensen and J. Bessereau, personal communication). If an error-prone polymerase is associated with HR, many of the substitutions and satellite mutations seen at the ZFN targets could arise from this repair pathway.

The role of the NHEJ pathway in DSB repair can be best intuited by examining the changes in the types of mutations observed when this pathway no longer functions, as in the *lig-4* nematodes. In this case, the most common target site mutation seen in the wild-type nematodes, the 4-bp insertion, completely disappears, as do other insertions. Instead, deletions and substitutions prevail. Furthermore, the deletions are more often associated with flanking microhomologies.

These characteristics all indicate that the classical DNA ligase IV-dependent NHEJ pathway is largely responsible for repairing blunted ends. When this pathway is disabled, blunt ends either remain unrepaird (not a great option for the cell) or are chewed back to either (1) create staggered ends which can be ligated by an alternate ligase, possibly associated with an error-prone polymerase, or (2) expose sequences for which a homologous template exists, so that either accurate or error-prone HR can proceed.

This model seems to reconcile the observed data quite well but remains quite vague about mutations that may or may not result from homologous recombination. Clearly, more experiments must be conducted to determine how the substitutions and satellites so frequently seen really arise in the mutated targets. As previously mentioned, experiments of this nature, employing a marked homologous template, are in preparation, and their completion should provide welcome information to help complete this model.
Appendix

Similar experiments have also been initiated in a *cku-80* nematode strain using QQR and the shorter QQR PCR product. Like ligase IV, *ku80* is a protein involved in the classical NHEJ DNA repair pathway (see Fig. 1-11) (Jones et al., 2001; Ramsden and Gellert, 1998; Wang et al., 1998). After DNA has been subjected to a DSB, the free ends can either invade a homologous sequence to initiate homologous recombination (van den Bosch et al., 2002), or they can be brought into close proximity for NHEJ repair via ligase IV (Wilson et al., 1997). The *ku80* protein, acting as a heterodimer with the closely related *ku70* protein, binds the free ends of broken DNA in a sequence-independent manner (Cary et al., 1997). The bound heterodimers then interact, forcing the DNA ends close together, into a conformation in which they can be joined by ligase IV (Doherty and Jackson, 2001; Walker et al., 2001).

Since its activities are so closely coordinated with those of ligase IV, we hypothesized that, in its absence, mutations similar to those found in *lig-4* nematodes should arise. At this time, only very preliminary results have been obtained. This experiment became problematic because the injected ZFN plasmids could not be made to form a stable array. After several attempts, the extrachromosomal array from the wild-type QQR nematode strain was crossed into the *cku-80* strain. The required crosses are outlined in Fig. 3-5.

A homozygous *cku-80* QQR line was obtained and verified by PCR (data not shown). Larvae from this line were then heatshocked as previously described, and truncated QQR targets were amplified from individual nematodes by PCR. This DNA was digested with *MluI*, then visualized by gel electrophoresis (Fig. 3-6).
Figure 3-5: Cross scheme to create *cku-80* QQR nematodes.

\[ \text{N2} \bigcirc \times \text{oxEx654} \bigoplus \]
\[ \downarrow \]
\[ \text{oxEx654} \bigcirc \bigcap \times \text{cku80} \bigoplus \]
\[ \downarrow \]
\[ \text{cku80/+, oxEx654} \bigoplus \]
\[ \downarrow \]
\[ \text{cku80, oxEx654} \bigoplus \]

*oxEx654* = the extrachromosomal array containing QQR, its target, and the GFP marker. Wild-type (N2) males were crossed with the wild-type nematode strain carrying the transgenic array. Transgenic male offsprings were then crossed to *cku-80* nematodes. Transgenic hermaphrodites from this cross were *cku-80* heterozygotes. These animals were allowed to self-fertilize to produce *cku-80* homozygotes, which were identified by PCR.
The target is 230 bp long, and the *MluI* restriction site is located at base 145. Cleavage at this site results in 145 and 85 bp DNA fragments. A plasmid containing the QQR target (+) has been cut to completion, while a very small proportion of the QQR target amplified from the heatshocked nematodes has become *MluI* resistant, indicating that inaccurate repair of QQR-induced cleavage has induced mutations at its target site. 

\[\text{L = DNA ladder, } + = \text{QQR target plasmid (WT), } - = \text{N2 nematode, included to validate the specificity of QQR target amplification. (A small amount of DNA from the plasmid control has spilled into this well, resulting in faint *MluI*-sensitive bands.) } 1 = \text{heatshocked } cku-80', \text{QQR nematode.}\]
Much as was observed in the *lig-4* nematodes, very little *MluI*-resistant DNA is present after heat shock, indicating that the disrupted mutagenic NHEJ DNA repair pathway is incapable of readily repairing the ZFN-induced breaks. Additional nematodes were heatshocked, and their truncated QQR targets were subsequently amplified by PCR, digested with *MluI* to enrich for mutations at the target site, and cloned into bacteria as previously described. Colony PCR of the bacterial colonies identified several mutations (Sambrook, 2001), which were sequenced and are shown in Table 3-5. All of the mutations identified are substitutions, similar to the majority of the mutations seen among the *lig-4* nematodes. This is not surprising, since the ligase IV and *ku80* proteins are both critical components of the same repair pathway. The similarity in the types of mutations observed after their deletion indicates that the repair pathway becomes nonfunctional when these proteins are absent. Work is ongoing to characterize additional mutations from the *ku-80* nematodes and determine the frequency at which mutations arise in this strain.
Table 3-5: Short QQR target mutations in *cku-80* nematodes.

<table>
<thead>
<tr>
<th>QQR target</th>
<th>TTC</th>
<th>TTC</th>
<th>CCC</th>
<th>ATCGCGT</th>
<th>GGG</th>
<th>GAA</th>
<th>GAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insertions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deletions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substitutions</td>
<td></td>
<td></td>
<td></td>
<td>GCGCGT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ATCGCGT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ACACGT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ACGTGT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ACGCGC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ACGCGG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complex</td>
<td></td>
<td></td>
<td></td>
<td>GCGCGT</td>
<td>GGG</td>
<td>GGA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ATGCGT</td>
<td>GGA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These mutations were sequenced from *cku-80* nematode short QQR target DNA cloned into bacteria. There are no insertions, implying that the ku80 protein, together with ligase IV may be strictly necessary for the ligation of blunt-ended DNA. The mutations consist entirely of small substitutions, including several complex mutations. These mutations may arise by an alternate end-joining repair mechanism or may be the result of homologous recombination utilizing an error-prone polymerase.
References


CHAPTER 4

TOWARD GERMLINE ZFN EXPRESSION

Introduction

The studies described in this dissertation show that ZFNs are capable of inducing DSBs at their target sites in the somatic tissues of *C. elegans*. These results are significant in two ways: First, they show that ZFNs are active and effective in yet another organism, lending evidence to the hypothesis that these tools may become the basis for a universal gene targeting system. They also show that ZFNs can efficiently locate and cleave specific targets, both on an extrachromosomal array and in the genome, with high efficiency in nematodes. This process has also proven quite useful when examining the DSB repair machinery in *C. elegans*, about which relatively little is known. Future studies, using a marked template to examine homologous repair, will provide additional insight into these processes.

The ability to produce a targeted somatic DSB is a critical first step in the creation of an efficient gene targeting strategy in nematodes but is, in itself, of limited usefulness. To become an effective tool in nematode genetics, ZFN activity must be induced in the germ tissues of the nematode to generate targeted mutant offspring. Working toward this goal, several preliminary experiments have been performed to examine the possible strategies of germline ZFN activation.
Materials and methods

The zinc finger coding sequences for both Ben1A and Ben1B were created via PCR mutagenesis of existing zinc fingers. The specificity-determining regions of the Ben1A fingers (QSSSLVR RSDELVR DKKDLTR), which recognize the DNA sequence 5'-GTA GTG ACC-3', were modified from the yB zinc fingers (RSDERKR QSSSLVR PDGHLVR) in the Zif268 backbone, which recognize the DNA sequence 5'-GCG GTA GGC-3' (Bibikova et al., 2002). The specificity-determining regions of the Ben1B fingers (QSSHLVR DCRDLAR QSSNLVR), which recognize the DNA sequence 5'-GGA GCC GAA-3', were created from the pMal-c2 E3-HS1 zinc fingers (QSSHLVR DCRDLAR QSSHLVR) in an Sp1C backbone, which recognize the DNA sequence 5'-GGA GCC GGA-3', provided by David Segal at the Scripps Research Institute in La Jolla, CA (Dreier et al., 2001; Dreier et al., 2000; Liu et al., 2002).

The Ben1A and Ben1B finger-encoding sequences were digested with Ndel + SpeI and ligated into a similarly digested pET15bZifA backbone, which contains IPTG-inducible promoter and an N-terminal histidine tag, and had been previously constructed to express ZFNs in E. coli and purify them with a His-bind column and a heparin-Sepharose column (Smith et al., 2000). The purified Ben1A and Ben1B ZFNs were used in an in vitro assay to examine cleavage of the ben-1 target, as well as the similar tbb-1, tbb-2, and mec-7 targets (Bibikova et al., 2001; Carroll et al., 2006).

Plasmids containing the targets for the in vitro assay were constructed from the pL4440 plasmid, provided by Susan Mango. The primers 5'-ATAAAGATGCGG-CCGCATCTAAAGATAAAGCACGTCTTATTATT3' and 5'-GCCGCTCGAGTTACAGCAATCGATGAACTTAC-3' were used to amplify the target region of tbb-1 gene.
from the N2 nematode genome. The primers 5'-ATAAGAATGCggccgccgcccGGCGC-GAGCTGCTTCCGTCTCCT-3' and 5' -GCCGCTCGAGCTTGAAAGAAATTAATTTTAC-3' were used to amplify the target region of the tbb-2 gene from the N2 nematode genome. The primers 5'-ATAAGAATGCggccgccgcccGGCGC-GAGCTGCTTCCGTCTCCT-3' and 5' -GCCGCTCGAGCTTGAAAGAAATTAATTTTAC-3' were used to amplify the target region of the mec-7 gene from the N2 nematode genome. All PCR reactions were performed using proofreading Pfu polymerase (Promega, Madison, WI). Since these primers contain XhoI and NotI restriction sites (in bold, above), the PCR products generated were digested with these enzymes, then ligated into a pL4440 backbone which had been similarly digested. The pL4440ben-1 plasmid, which contains the cDNA sequence of the ben-l gene in the pL4440 backbone, was a gift from Susan Mango. The plasmids were linearized with Scal before addition to the in vitro reaction mix.

The pJM1 vector was designed to express the ZFNs in nematodes, using the 16-48 heatshock promoter. The Ben1A ZFN-encoding sequence was placed in this vector in two steps. Initially, the N-terminal histidine tag in the pET15bBen1A vector was excised by digestion with Ncol + Not1, and an oligonucleotide encoding a nuclear localization signal from the SV40 large T-antigen, constructed by annealing the primers 5'-CATGGGCAGCGATCCAAAAAAGAAGAGAAAGGTACA-3' and 5'-TGGGGTACCTTTCTCTTCTTTTTGGATCGCTGCC-3', was inserted in its place. Next, the sequence encoding the NLS and Ben1AFN was amplified from the plasmid with the proofreading Pfx polymerase (Invitrogen Carlsbad, CA), using the primers 5'-GCCGACGCGTTAAAAATGGGCAGCGATCCAAAAAAG-3' and 5'-GCGCGC-
GCTAGCTTAAAGTTTATCTGCCTATTATT-3', which add the upstream four-
adenosine repeat and the MluI and NheI restriction sites necessary to place this
fragment into the digested pJM1 backbone. Alternate zinc finger-encoding sequences
can be substituted into this plasmid by NdeI and SpeI digestion (as done for QQR and
NwA and NwB in Chapter 2.) The Ben1B finger sequences were digested with Ndel +
SpeI and ligated into the similarly digested pJM1 backbone, to create the plasmid
pJM1-Ben1B.

The studies in this chapter were performed using the wild-type N2 nematode
strain. These nematodes in which ben-1 targeting was examined were transformed by
gonad injection, as previously described (Mello et al., 1991). Plasmid DNA was
linearized with ScaI prior to injection. A DNA mixture of pJM1-Ben1A (5 ng/µl),
pJM1-Ben1B (5 ng/µl), pPD118.33 (Pmyo-2::GFP; 1 ng/µl), and 1-kb ladder (80
ng/µl; Invitrogen, Carlsbad, CA) was used to produce the strain EG4336 ox837 that
carries the Ben1AFN and Ben1BFN constructs on an extrachromosomal array.

To induce ZFN expression, L2-L3 transgenic nematodes were heat shocked in a
35°C water bath for one hour on two consecutive days, then allowed to recover at
room temperature for several hours before further analysis. Each nematode was then
frozen at -80°C in lysis buffer [50 mM KCl, 10 mM Tris•HCl (pH 8.3), 2.5 mM
MgCl₂, 0.45% Nonidet P-40, 0.45% Tween-20, 0.01% gelatin, 200 µg/ml Proteinase
K], then lysed at 65°C for one hour and 95°C for 15 minutes. The ben-1 target was
then amplified by PCR using Taq polymerase (New England Biolabs, Beverly, MA)
and the primers 5’-TGTTGTCTCCTTGCGCATCG-3’ and 5’-AATATTCATT-
AAATTCAATTAATCC-3’.
To create the constructs used in the Flp reporter studies described, the Flp gene and an upstream four-adenosine translation initiation consensus sequence were excised from the plasmid pWD79hsp2uflpRV, provided by Wayne Davis, via *Mlu*1 + *Nhe*1 digestion. This fragment was then ligated into a similarly digested pJL44.2 backbone to create the pJM3Flp plasmid, in which Flp expression is driven by the 16-48 heatshock promoter.

The pJM4RFP-GFP Flp reporter was built from the pPD118.33 plasmid. Two FRT sequences were inserted into the MCS found between the end of *myo-2* promoter and the beginning of the GFP gene. The first was produced by annealing the oligonucleotides 5’-TCGACGAAGTTCCTATTCTCTAGAAAGTATAGGAACCTTCGGATCCGC-3’ and 5’-GGCCGCGGATCCGAAGTTCCTATACTTTCTAGAGATAGGAACCTTCG-3’ and ligating the product into a *Sal*1 + *Not*1 digested backbone. The second was produced by annealing the oligonucleotides 5’-CGCGTGAAAGTTCCTATTCTCTAGAAAGTATAGGAACCTTCG-3’ and 5’-CCGGGGAAGTTCCTATACTTTCTAGAGATAGGAACCTTCG-3’ and ligating the product into the *Mlu*1 + *Xma*1 digested backbone. The RFP sequence and its accompanying 3’UTR were amplified from the plasmid prSETBmRFP1, provided by Wayne Davis, using the primers 5’-CGCGCGGGATCCAAAAATGGCCTCCTCCGAGGAC-3’ and 5’-CGCGCGACGCAGTCGTGGACTCCAACGTCAAAC-3’, which add *Bam*HI and *Mlu*1 restriction sites (in bold) necessary to insert the fragment between the FRT sites.

The nematodes in which Flp activity at the Flp reporter was examined were transformed by gonad injection, as previously described (Mello et al., 1991). Plasmid DNA, except that of pRF4 (Mello et al., 1991), was linearized with *Sal*1 prior to
injection. A DNA mixture of pJM3Flp (5 ng/μl) pJM4GFPRFP (1 ng/μl), pRF4 (rol-6; 15 ng/μl) and 1-kb ladder (80 ng/μl; Invitrogen, Carlsbad, CA) was used to produce the strain EG4526 oxEx966 carrying Flp and the Flp reporter on an extrachromosomal array. Transgenic nematodes from stable lines were heatshocked for one hour at 35°C to express Flp, allowed to recover, then examined for RFP and GFP expression.

Results and discussion

Ideally, the gene to be targeted in the nematode germline should produce an easily observable phenotype among progeny in which it has been mutated. To this end, several different genes were initially considered for targeting, including unc-22, unc-29, unc-38, and ben-1. The sequence of each of these genes was searched for possible ZFN targets, consisting of two inverted three-finger binding sites composed of either GNN or ANN triplets, separated by a 6-bp spacer. Potential targets with these characteristics could be identified by a computer-assisted search for the degenerate sequence 5’-NNY NNY NNY NNNNNN RNN RNN RNN -3’. The candidate targets were then manually examined to identify those composed entirely of triplets for which fingers existed that had been previously characterized as being highly specific and having good affinity to their targets, based on information from the phage display experiments previously published (Dreier et al., 2001; Dreier et al., 2000; Segal et al., 1999).

Ultimately, a target was chosen in the second exon of the ben-1 gene (Fig. 4-1). This gene, which encodes a β-tubulin, is closely related to the other β-tubulin genes
Figure 4-1: The *ben-1* target sequence.

This sequence is located at nucleotide 1325 in exon 2 of the *ben-1* gene on chromosome III. The binding site of Ben1B is in pink. The binding site for Ben1A is in lavender. The 6-bp spacer, in which ZFN-induced cleavage should occur, is in blue.
*tbb-1, tbb-2,* and *mec-7*. These proteins are major components of microtubules and appear to play a partially redundant role (*Ellis* et al., 2004; *Wright* and *Hunter*, 2003). As such, *ben-1* can be mutated without adverse effect, but its absence makes the nematodes resistant to benzimidazole antimicrobial agents, since it is the only member of the β-tubulin family affected by these drugs (*Driscoll* et al., 1989; *Kwa* et al., 1995). A single functional copy of the gene, however, is not sufficient to paralyze nematodes in the presence of benzimidazoles. Thus, a mutation at *ben-1* gives dominant resistance to benzimidazoles, whereas all nematodes with two wild-type copies of the gene are effectively paralyzed and capable of little coordinated movement.

After selection of a target, sequences encoding two ZFNs, named Ben1A and Ben1B, were created via PCR mutagenesis from existing zinc fingers. This was done by consulting a database of readily available zinc finger proteins, then modifying the sequences of DNA-binding residues as needed to alter their affinity. The Ben1A fingers were kept in the Zif268 backbone in which their parent fingers were located, while the Ben1B fingers were maintained in an Sp1C backbone.

Since it was unclear how well these ZFNs would bind with their target sites, they were first expressed in *E. coli*, purified by column chromatography, and assayed in vitro using a linearized plasmid containing their target sequence (*Carroll* et al., 2006). At high concentrations, the ZFNs cleaved DNA in a nonspecific manner, but, as their concentration was reduced, specific cleavage at the target site was observed (Fig. 4-2). As an additional test of the specificity of Ben1AFN and Ben1BFN, target plasmids containing the highly similar targets found in the related β-tubulins *tbb-1, tbb-2,* and *mec-7* were constructed (Table 4-1).
Figure 4-2: In vitro assays of Ben1A and B ZFN target activity.

This assay examines *ben-1* target cleavage by purified Ben1A and B ZFNs. Each ZFN was serially diluted, then mixed with a linear plasmid target to determine the concentration at which the enzyme does not cleave DNA indiscriminately. Once correct ZFN concentrations had been identified (lanes 4 and 8), both enzymes were mixed together with the target (lanes 9-12) to assay for specific target site cleavage (lane 12.) The linearized plasmid containing the *ben-1* target has a length of 4154 bp. Cleavage at the target site results in bands of lengths 3073 and 1081 bp.
Table 4-1: Alignment of the *ben-1*, *tbb-1*, *tbb-2*, and *mec-7* target sites.

<table>
<thead>
<tr>
<th>Target</th>
<th>Location (Chromosome: Nucleotide)</th>
<th>Target sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>ben-1</em></td>
<td>III: 3,540,229</td>
<td>GGT CAC TAC ACCGAA GGA GCC GAA</td>
</tr>
<tr>
<td><em>tbb-1</em></td>
<td>III: 10,741,929</td>
<td>GGA CAT TAC ACCGAG GGA GCA GAA</td>
</tr>
<tr>
<td><em>tbb-2</em></td>
<td>III: 4,017,193</td>
<td>GGT CAC TAC ACAGAA GGA GCT GAG</td>
</tr>
<tr>
<td><em>mec-7</em></td>
<td>X: 7,775,998</td>
<td>GGA CAC TAC ACCGAG GGA GCA GAA</td>
</tr>
</tbody>
</table>

The 6-bp spacer between finger binding sites is in the center of the sequence. Differences in the binding sequences in the target homologs are in bold.
Additional in vitro cleavage assays indicated that, at concentrations shown to specifically cleave the *ben-1* target, these similar targets were not cut (Fig. 4-3). This seemed to indicate that Ben1AFN and Ben1BFN would perform well in nematodes, so their ZFN coding regions were transferred to the nematode heatshock vector (see Fig. 2-2) and injected into the distal gonad of young adult nematodes.

When injected DNA forms an extrachromosomal array in nematodes, individual molecules are ligated together in random orientation. Thus, transcription will often proceed through a coding region and into a section of DNA in the opposite orientation. The resulting RNA molecule can form a hairpin of dsRNA which will, in turn, activate the RNA interference pathway in the nematodes and induce silencing of the transgenic DNA in the nematode germline (Dernburg et al., 2000). This phenomenon is especially common when the DNA in the extrachromosomal array is quite repetitive.

The situation is actually aggravated during most injections since the foreign DNA of interest is usually mixed with highly repetitive DNA sequences, often the DNA ladder used as a standard in agarose gels. This carrier DNA increases the ease of array formation and seems to aid in their stability. It is commonly used because many studies employing extrachromosomal arrays examine the expression of transgenes only in somatic tissues, where silencing is not as severe.

It was feared that this phenomenon would be a barrier to germline expression of the Ben1A and Ben1B nucleases, and, indeed, screening of several thousand offspring of heatshocked transgenic nematodes produced no mutants. While work toward germline expression of the Ben1A and Ben1B ZFNs continued, initial efforts to identify targeted mutation at the *ben-1* target in somatic tissues were undertaken.
In vitro assays of ZFN-induced cleavage at the *ben-1* vs. the *tbb-1, tbb-2,* and *mec-7* targets by purified Ben1A and B ZFNs to examine nuclease specificity. The ZFNs were initially assayed separately at several different concentrations to ensure that, at the concentrations tested, no nonspecific DNA cleavage occurred (boxed lanes: *ben-1; A only,* *ben-1; B only*). At these concentrations (indicated by the additional boxed lanes), the ZFNs were shown to be able to induce specific cleavage at the *ben-1* target (resulting in the faint 3073- and 1081-bp bands), but not at any of the homolog targets. (The contrast has been enhanced to maximize target visibility.)
These were hampered by an inability to effectively screen mutant sequences, since no unique restriction site exists at the target. Eventually examination of this locus was dropped in favor of the more easily assayable Nowhere target.

Several possible strategies can be used to decrease the germline silencing from an extrachromosomal array. First, a less repetitive carrier DNA, such as predigested nematode genomic DNA, can be used in the injections. In this case, arrays are much less subject to germline silencing but are less commonly formed in the nematode (Kelly et al., 1997). Initial attempts to place the Ben1A and Ben1B encoding sequences into this context were unsuccessful, and additional attempts are underway.

There are several mutant strains of nematode in which germline silencing is inactive. Among the genes inactivated in these various strains are mut-7, mes-4, mut-16, rde-1 (Garvin et al., 1998; Ketting et al., 1999; Tabara et al., 1999). Injections into these strains may produce transgenic offspring, although outcrossing may be necessary since the strains tend to accumulate mutations via transposon hopping. Alternately, RNAi could be used to transiently eliminate expression of one of these genes in a strain carrying the ZFN plasmids on an array (Timmons and Fire, 1998). Heatshock to activate the ZFNs while the transgenic nematodes are undergoing RNAi may also produce targeted offspring.

Although the nematode 16-48 heatshock promoter has been shown to be active in the germline (Bessereau et al., 2001), alternate promoters may be more effective in driving germline expression of the ZFNs. Principal among these is the pie-1 promoter. The Pie-1 protein is expressed in the earliest stages of the developing egg and regulates the activity of many of the genes involved in early development (Mello et al.,
Studies focusing on transgene expression in the germline have shown that a *pie-1* promoter can remain active in the germline over several generations, especially when the nematodes are incubated at a slightly increased temperature (Strome et al., 2001). Although ZFN expression driven by a *pie-1* promoter is not inducible, even unregulated germline expression from an array is desirable since the mutant progeny created could be identified regardless of continuous ZFN expression. Finally, within the last several years, gene gun technology has been adapted so that foreign DNA can be introduced into nematodes via biolistic transformation (Berezikov et al., 2004; Jantsch et al., 2004; Praitis et al., 2001; Wilm et al., 1999). Gold beads coated with DNA are shot into nematodes to produce transgenic animals in which this DNA is usually integrated into a chromosome at low copy number. When expressed from a chromosomal locus, transgenes are less subject to silencing. In this context, the ZFNs, driven by either a heatshock or *pie-1* promoter, may be active in the germline and should produce targeted mutations.

These experiments are designed to increase expression of the ZFN transgenes, but their efficacy remains questionable. To determine the optimal conditions for transgene expression, an assay for germline expression of a transgene is needed. To this end, we have designed a Flp reporter system, which uses an RFP-GFP reporter plasmid to detect germline transgene expression. Flp is a yeast recombinase, shown to be active in several systems (Broach et al., 1982; Cox, 1983; Rong and Golic, 2000). This enzyme recognizes a 34-bp sequence, known as the Flp recombinase target (FRT) (Senecoff et al., 1985). When two such sequences are bound in *cis*, the intervening
DNA is either inverted, if the FRTs are in opposite orientation, or excised and circularized, if the FRTs are in the same orientation (Fig. 4-4).

The RFP-GFP reporter is a plasmid in which the pharynx-specific myo-2 promoter drives expression of an RFP gene, flanked by repeated FRT sequences. A GFP gene lies downstream of the final FRT sequence (Fig. 4-5). In the absence of Flp, the myo-2 promoter drives RFP expression, while its 3' UTR effectively eliminates any GFP expression. When Flp is introduced, it excises the RFP and 3' UTR sequences, so that the myo-2 promoter is now able to efficiently drive GFP expression.

To validate the Flp reporter, Flp was first cloned into the nematode heatshock vector (Fig. 2-2), so it would be expressed in the same manner as the ZFNs. It was then injected with the reporter to form an extrachromosomal array. Transgenic animals could be identified by their pharyngeal RFP fluorescence (Fig. 4-6a). These animals were subsequently heatshocked to induce Flp expression, which should allow the myo-2 promoter to express GFP. As predicted, when these nematodes were examined, they all clearly expressed GFP in the pharynx (Fig. 4-6b). Some residual RFP expression remained, and it is unclear if this was due to the half-life of the protein or to incomplete excision of the RFP sequence from all cells expressing Flp in the pharynx. In any case, this assay provides proof of the activity of Flp in the nematode.

The more valuable use for this assay, however, is as a means to detect the conditions necessary for transgene activity in the germline. If Flp is active in the germline, the offspring of the GFP-expressing transgenic nematodes described above should express pharyngeal GFP. If it is not, their pharynxes will be marked with RFP.
Schematic of Flp recombination products resulting from possible FRT orientations. When two FRT sites are oriented in a direct repeat, Flp catalyzes circularization of the intervening sequence, removing it, along with one FRT site, from the molecule. When two FRT sites are in inverted repeat, Flp catalyzes the inversion of the intervening sequence but retains it, along with both FRT sites, in the molecule.
This plasmid was constructed from a *C. elegans* myo2::GFP marker plasmid (pPD118.33.) The *myo-2* promoter drives expression of the GFP protein in the pharynx of the nematode, in essence causing its head to glow. In the multicloning site (MCS) separating the promoter from the GFP, a construct consisting of two repeated FRT sequences flanking an RFP reporter and 3' UTR, was inserted. When present, the *myo-2* promoter now drives RFP expression, while the 3' UTR prevents the expression of GFP. When Flp is active at the reporter, the RFP and its accompanying 3' UTR are excised from the construct, and GFP is expressed from the promoter.
Figure 4-6: Flp reporter activity in nematodes.

a. Picture of a nematode expressing RFP from the Flp reporter. The myo-2 promoter drives expression of the RFP gene in the nematode pharynx, while the downstream 3'UTR inhibits GFP expression. 

b. Picture of the same nematode expressing GFP from the reporter after Flp expression. The presence of Flp has excised the RFP gene and the downstream UTR, allowing the myo-2 promoter to now drive expression of GFP. There is some residual expression of RFP (shown in (a)), indicating that either RFP has an extended half-life or that Flp expression incompletely activates all of the reporter plasmids.
In this case, when Flp was expressed from a heatshock promoter in a repetitive array, all transgenic nematode offspring showed no indication of Flp activity. However, studies are proceeding in which Flp has been placed either in a nonrepetitive array or behind a pie-1 promoter. Additional experiments in which these constructs are to be integrated into the genome are also progressing. Together, these studies should address and resolve the issues involving transgene expression in the germline. Once the optimal conditions for expression have been determined, Flp expression can be replaced by ZFN expression, and targeted mutations in nematode offspring should result.
References


recognizing each of the 5'-GNN-3' DNA target sequences. Proc Natl Acad Sci U S A 96, 2758-2763.


CONCLUDING REMARKS

Since the first genes were targeted in yeast almost 30 years ago, technology has evolved so that it is now possible to specifically target genes in several different organisms. Early targeting techniques in yeast were improved by the use of PCR to rapidly create multiple targeting constructs. Fly targeting became even more useful with the establishment of an ends-out donor system. Targeting in mice has seen improvements in the capabilities of the drugs used to select cells in which homologous integration of the modified gene has occurred and in the employment of the Cre-\textit{loxP} system to excise unwanted DNA sequences from an integrated vector or create tissue-specific gene knock-outs.

The discovery of the zinc finger motif, the characterization of its DNA-binding specificity, and its subsequent use in the zinc finger nuclease represents the next steps in the ongoing evolution of gene targeting. In flies, plants, and human cell lines, the specific target-site DSBs created by ZFNs have been shown to dramatically increase the frequency with which genes can be knocked out. Furthermore, in all of these organisms, modified homologous templates introduced with ZFNs can be used to direct repair of these DSBs, resulting in genuine gene targeting at a high frequency. Future experiments will undoubtedly extend the utility of this system and increase the frequency at which targeting can be achieved.
This dissertation has shown that ZFN technology can also be adapted to induce targeted mutagenesis in the somatic tissue of the nematode, *Caenorhabditis elegans.* The induction of ZFNs constructed to bind and cleave a unique target site on either an extrachromosomal array or at a genomic locus results in mutations at about 20% of either of these targets. The mechanism whereby many of these mutations arise has also been studied. The cell often repairs the DSBs created by the ZFNs using the canonical NHEJ machinery. Elimination of one of the major components of this pathway, DNA ligase IV, significantly alters the frequency and types of mutations generated by the ZFNs at both targets, indicating that other, presently uncharacterized, processes also play a role in ZFN-induced DSB repair. Future studies can be designed to reveal more about the nature of these mechanisms.

Additionally, the experiments described in this dissertation provide a solid foundation from which future studies designed to activate ZFNs in the nematode germline can be conducted. Once the obstacle of germline transgene cosuppression has been overcome by one of the strategies described above, ZFNs can be expressed in the germline to produce targeted knockouts among a nematode’s progeny. If the somatic activity of the ZFNs is any indication of their germline activity, targeted mutant nematodes may also arise at a high frequency. Once achieved, germline mutation can be supplemented by the use of an array-based modified template to generate the long-awaited procedure for genuine gene targeting in the nematode.