Using Homologous Recombination to Manipulate the Genome of Human Somatic Cells

MATTHEW PORTEUS

Departments of Pediatrics and Biochemistry, UT Southwestern Medical Center, Dallas, TX 75214

Introduction

Gene targeting by homologous recombination is a powerful tool to precisely manipulate the genome for experimental uses. It has been used extensively in yeast, murine embryonic stem cells, and the chicken DT40 cell line but has not been widely used in vertebrate somatic cells. The reason gene targeting has not been used in vertebrate somatic cells is that the relative rate of gene targeting (the ratio of homologous integrants to random integrants) is very low. In the last several years, recombinant AAV and DNA double-strand breaks have been developed as technologies to increase the relative rate of gene targeting in vertebrate somatic cells. These improvements have resulted in rates of gene targeting that are useful for experimental purposes and may even be useful for gene therapy.

The manipulation of the genome of cells has been a fundamental experimental technique in biomedical research. Furthermore, the manipulation of the human genome for therapeutic purposes is the basis for the field of gene therapy. There are a number of different ways that the genome can be altered for experimental purposes. These include the use of random mutagens, uncontrolled integration of transgenes, and homologous recombination. Homologous recombination represents the most precise way to manipulate the genome because defined changes, both large and small, can be

Abbreviations: rAAV, recombinant adeno-associated virus; DNA, deoxyribonucleic acid; DSBs, double-strand breaks; SDSA, synthesis dependent strand annealing; ES, embryonic stem; ESC, embryonic stem cell; HPRT, hypoxanthine phosphoribosyltransferase; HAT, hypoxanthine-aminopterin-thymidine; ITR, inverted terminal repeat; bp, basepair; ZFN, zinc finger nuclease; SCID, severe combined immunodeficiency.
introduced into specific genomic locations while leaving the rest of the genome unperturbed.

Fundamentally homologous recombination is a natural cellular process that is used for a variety of purposes across all life forms including phage, prokaryotic cells, and eukaryotic cells. The detailed mechanism of homologous recombination has been studied for decades and models for the mechanism of homologous recombination continue to be developed (these models and a more detailed discussion of the mechanism of homologous recombination have been reviewed elsewhere) (Szostak et al., 1983; West, 2003). It is likely that no single model explains all aspects of homologous recombination as homologous recombination serves different purposes under different circumstances. Conceptually, homologous recombination can be thought of as either an exchange of homologous DNA fragments (cross-over recombination) or the transfer of DNA information from one DNA fragment to another (gene conversion).

In mitotic eukaryotic cells, the primary use of homologous recombination is to repair DNA double-strand breaks (DSBs). The SDSA model of homologous recombination is the current favored model for repair of DSBs and is schematized in Figure 1. In meiotic eukaryotic cells, homologous recombination is the way cells exchange genetic material between homologous chromosomes to create genetic diversity and is the way that cells synapse homologous chromosomes to allow proper chromosomal segregation. In special circumstances, homologous recombination is used in mitotic cells to rearrange their own genomes. In yeast, for example, homologous recombination is used during mating-type conversion (Haber, 1998). In certain vertebrates (e.g., chickens and rabbits) homologous recombination is used during somatic hypermutation to create diversity in the immune system (McCormack et al., 1993). Finally, a class of fascinating genetic parasites, called homing endonucleases, use homologous recombination as way of expanding and maintaining themselves (Chevalier and Stoddard, 2001).

While homologous recombination is a natural way that nature rearranges genetic material, homologous recombination has become a powerful scientific tool. Figure 2 shows a general schematic, for how homologous recombination is used experimentally. In this chapter, I will briefly discuss the use of homologous recombination as a tool for genetic manipulation in yeast, murine embryonic stem cells, the chicken DT40 cell line, and then discuss the recent developments that have been made to use it as a way to broaden its use into cells that have been previously thought to be resistant to experimental manipulation by homologous recombination. Furthermore, if the rate of homologous recombination can be increased sufficiently, the precision of the process would potentially make it an ideal form of gene therapy. That is, one can easily envision using homologous recombination to correct disease causing mutations, such as point mutations that cause sickle cell diseases and cystic fibrosis, while leaving the rest of the genome undisturbed.

**Homologous recombination in yeast**

In 1978, Hinnen et al. (1978) applied molecular biological techniques to the newly developed technique of *S. cerevisiae* (yeast) transformation by plasmid DNA (Hinnen et al., 1978). They found that a large frequency the plasmid integration events were
Using homologous recombination to manipulate the genome . . . 197

Figure 1. Schematic Representation of the Synthesis Dependent Strand Annealing (SDSA) Model of Homologous Recombination. In this model, homologous recombination is initiated by a DNA double-strand break (DSB). The DSB is then processed to generate long 3’ single-stranded ends. These ends are used by the homologous recombination machinery for strand invasion into an undamaged, homologous double-stranded DNA molecule. The undamaged homologous DNA serves as the template for repair. DNA polymerase then uses the 3’ end of the invading strand as a primer and the undamaged DNA strand as a template for new DNA synthesis. After some point, synthesis is stopped, the strands of DNA are paired back with their original partner. DNA polymerase then fills in the remaining gaps in the damaged strand and DNA ligase seals the ends. In the normal repair of DSBs, the sister-chromatid is used as the repair template. In gene targeting, the cell inadvertently uses transfected fragment of DNA that contains sequence differences as the repair template. When the cell uses this alternative template, allele S can be changed into allele A. The sequence difference in allele A is depicted as a hatched region.

by homologous recombination. Depending on the structure of the transformed plasmid, the percentage of targeted integrations ranged from 30-85% (Orr-Weaver et al., 1981). That is, the relative rate of gene targeting (the ratio of homologous integrants to random integrants) in yeast is very high. This high relative rate of gene targeting has developed into a powerful and amazingly simple tool for yeast geneticists. There are multiple experimental ways of making use of this high relative rate of gene targeting, including the ability to make null alleles (“knockouts”) in genes in a simple one-step gene replacement procedure (Adams et al., 1997; Rothstein, 1983). The ease of doing gene targeting by homologous recombination in yeast, has resulted in a goldmine of thousands of strains being made with precisely defined changes, both subtle and complex, in the genome.

Limits to homologous recombination in mammalian somatic cells

If gene targeting by homologous recombination is such a valuable tool in yeast, why
Figure 2. Schematic Representation of Gene Targeting by Homologous Recombination. The left depicts how homologous recombination can be used to insert a selectable marker into a precise location. A targeting vector in which a selectable marker is inserted between “D” and “E” is introduced into cells. The genomic target is represented by the dotted box and the sequence schematized as “Y….L.” When homologous recombination occurs, depicted by the crossed lines, the selectable marker will be inserted precisely into the genomic locus between “D” and “E.” The right depicts how homologous recombination can be used to replace a segment of genomic DNA with a selectable marker. In this targeting vector, the selectable marker is flanked by sequence A-D on the left homology arm and sequence I-K on the right homology arm. These sequences are derived from the endogenous genomic target. The genomic target is represented by the dotted box and the sequence schematized as “Y….L.” When homologous recombination occurs, depicted by the crossed lines, the selectable marker will be inserted into the gene and the sequence E-H of the gene will be replaced by the selectable marker. This general format is how almost all gene targeting experiments are done. More sophisticated targeting experiments (such as has been done with zinc finger nucleases) involve merely changing single or a small number of nucleotides rather than inserting a relatively large selectable marker.

has it not been developed in mammalian somatic cells? The reason is that the relative rate of gene targeting in most mammalian cells is too low to be of experimental utility. In a gene targeting experiment, the targeting DNA is introduced into the cell. It can then integrate into the genome of the cell in one of two possible ways: randomly or homologously. In yeast, for reasons that are still not completely understood and described above, the DNA integrates almost exclusively in a homologous fashion; that is, it rarely undergoes random integration. In mammalian somatic cells, however, the converse is true; the target DNA primarily integrates almost exclusively in a non-homologous (or random) fashion. For every 1 gene targeting event, there will 10-20,000 random integrants in mammalian cells (Porteus and Baltimore, 2003; Sedivy and Sharp, 1989). Such a low ratio of targeted to non-targeted events means that for the most part it is experimentally impractical to use gene targeting in mammalian somatic cells.

Despite the low relative rate of gene targeting in mammalian cells, there have been several examples of using it. The first example of gene targeting using homologous
recombination was by Smithies and colleagues who showed that they could target the human β-globin gene in a human somatic cell line (Smithies et al., 1985). This seminal work established that gene targeting in homologous recombination was possible in mammalian somatic cells. Later, Sedivy and colleagues used a sophisticated combination of positive and negative selection and screening, to identify cell lines that had undergone gene targeting events (Seedivy and Sharp, 1989). Using this strategy they were able to knockout both alleles of the p21 gene in primary human diploid fibroblasts before replicative senescence occurred (Brown et al., 1997). In general, this technique has not been widely adopted, but has been used in several instances, notably by Vogelstein and colleagues, to make knockouts in HCT116 cells, a colon cancer cell line (Bunz et al., 1998).

**Homologous recombination in murine ES cells**

Gene targeting by homologous recombination was developed as a technique in mammalian cells in the 1980s by the groups of Mario Cappechi and Oliver Smithies (Capecchi, 1989; Koller and Smithies, 1992). Murine embryonic stem cells (mESCs) are cells that are derived from the inner cell mass of blastocysts and can be grown indefinitely in cell culture under defined conditions. In cell culture they can be induced to differentiate into specific cell types from all three major cell lineages (ectodermal, mesodermal, and endodermal) and when injected back into the inner cell mass can also give rise to cells in all three cell lineages and germ cells. If gene targeting using homologous recombination could be performed in mESCs then one would not only create immortal cells with defined genomic changes but also might be able to create mice with specific genetic mutations.

The first successful gene targeting experiments in mESCs was reported in the mid-1990s. In these experiments the HPRT gene was targeted. HPRT was an ideal target for these experiments because it is on the X chromosome and in male cells only one allele would have to be targeted. Furthermore, HPRT positive cells can be selected for by growing in HAT media while HPRT deficient cells can be selected for because they are resistant to treatment with 6-thioguanine (6-TG). In these first experiments, gene targeting was used in mESCs to either mutate the HPRT gene or correct a mutated HPRT gene (Doetschman et al., 1987; Doetschman et al., 1988; Thomas and Capecchi, 1987). These targeted mESCs were then used to create mice with a defined change in the HPRT locus. Since those experiments, the technology of gene targeting in mESCs has been widely developed using both positive and negative selectable markers and the number of mES cell and mouse lines created using this technology is in the thousands. One of the interesting findings from these studies is that the relative rate of gene targeting in mESCs seems to be 10-100 fold higher than that found in mammalian somatic cells. The mechanism for this higher relative rate of gene targeting remains unclear.

**Homologous recombination in chicken DT40 cells**

With the development of gene targeting by homologous recombination in mESCs, researchers could create cells, mice, and then from those mice, cell lines with defined
genetic mutations. Thus, it became possible, in a somewhat roundabout way, to create genetically modified mouse somatic cell lines, the first way to create vertebrate cells with specific genetic changes made by homologous recombination. The second way it became possible to make vertebrate cells by homologous recombination was described in 1991 by Buerstedde and Takeda (Buerstedde and Takeda, 1991). They were studying the chicken DT40 cell line. DT40 cells are chicken pre-B cells that have been immortalized by the avian leukosis virus. In normal B-cell development, chickens create immunoglobulin diversity by somatic hypermutation using homologous recombination. DT40 cells seem to have been immortalized at this stage in B-cell development (Buerstedde et al., 1990). Buerstedde and Takeda introduced a standard gene targeting vector that contained a selectable marker into DT40 cells, identified individual clones that integrated the selectable marker, and analyzed the resulting clones for the site of integration (Buerstedde and Takeda, 1991). In mammalian somatic cells, one would have expected, as described above, to screen ten-twenty thousand clones before finding a single clone that had undergone gene targeting. Remarkably, they found that 50-90% of the clones, depending on the gene targeted, had undergone gene targeting. This rate of targeting is approximately three-orders of magnitude greater than would have been expected. Further studies of DT40 cells have shown that this high rate of gene targeting is not the result of a high intrinsic rate of gene targeting, but instead is the result of a low intrinsic rate of random integration. In this respect, one can think of DT40 cells as a vertebrate form of yeast.

Since Buerstedde and Takeda’s discovery, hundreds of DT40 cell lines have been made by gene targeting to study a wide range of cellular processes including B-cell development and signaling, genomic stability, and DSB repair (Yamazoe et al., 2004). The limit of this system, of course, is that DT40 cells are a highly specialized cell type from a single species.

Development of gene targeting in human somatic cells

Even with the development of gene targeting in mESCs and DT40 cells, it still was not possible to use homologous recombination in human somatic cells. A backbone of biomedical research is that the genetic similarities between humans and “model” organisms is sufficient that understanding processes in these model systems help us understand normal human physiology, human diseases and lead to improved therapies. Nonetheless, it would be extremely useful to be able to do gene targeting by homologous recombination in human somatic cells. In the last few years, two different strategies have been developed to increase the rate of gene targeting in human somatic cells: the use of recombinant adeno-associated virus and the stimulation of gene targeting by DNA double-strand breaks.

Homologous recombination by rAAV

The first indication that gene targeting in mammalian somatic cells might be possible at frequencies that could be experimentally useful came from the work of David Russell and his colleagues at University of Washington using recombinant adeno-associated virus (rAAV) (Russell and Hirata, 1998). rAAV is a single-stranded DNA
Using homologous recombination to manipulate the genome... virus of the parovirus family that depends on cells being co-infected with either adenovirus or herpes virus to be replicated. The normal nucleic acid structure of AAV consists of a Rep gene and Cap gene flanked by inverted terminal repeats (ITR’s). In rAAV, the Rep and Cap genes are replaced by an expression cassette including a promoter and a gene of interest. The recombinant virus is then made in HEK-293 cells by providing the Rep and Cap protein products in trans. rAAV is capable of infecting a variety of cell types and has been studied as a possible vector for many gene therapy trials, including as a treatment for hemophilia B (Manno et al., 2006; Manno et al., 2003). Once rAAV infects cells it can be maintained episomally in non-dividing cells or can integrate randomly into the genome, just like other fragments of DNA can. Russell and his colleagues found, however, that rAAV integrates via homologous recombination at a much higher frequency than anticipated (Russell and Hirata, 1998). The rate of gene targeting by homologous recombination was directly related to the multiplicity of infection; that is, the more virus that infected the cell, the higher the rate of gene targeting. Under certain circumstances, rAAV can achieve gene targeting rates of 1% or greater. This rate of targeting could be achieved in primary human diploid cells, at a variety of genomic loci, and can create a variety of different genomic changes (including point mutations, small insertions or deletions, or small insertion or deletions) (Russell et al., 2002). The use of gene targeting by rAAV has even been used in a pre-clinical study of gene therapy for the autosomal dominant disorder osteogenesis imperfecta (Chamberlain et al., 2004). The mechanism by which rAAV causes a high rate of gene targeting remains a mystery. Nonetheless, gene targeting by rAAV continues to be developed as a technology and a protocol to streamline the production of rAAV virus for gene targeting has been established (Kohli et al., 2004). There are good reasons to expect that this technology could be broadly useful for mammalian geneticists but it remains to be seen if it becomes widely adopted.

Homologous recombination induced by DNA double-strand breaks

A second way to stimulate gene targeting in mammalian somatic cells is to create DNA double-strand breaks (DSBs) in the genomic target. As discussed above, homologous recombination is an evolutionarily conserved mechanism to repair DSBs. In the normal repair of DSBs by homologous recombination, the sister-chromatid, an exact duplicate of the damaged DNA, serves as the template for homologous recombination. In this way, homologous recombination is considered the most accurate form of DSB repair. If the DSB were repaired by homologous DNA that was not the sister-chromatid, such as an extra-chromosomal fragment of DNA, it might be possible to “trick” the cell into undergoing gene targeting. In the mid-1990’s several groups explored this possibility by using the I-SceI endonuclease. I-SceI endonuclease (“Sce”) is a yeast homing endonuclease (Brenneman et al., 1996; Jasin, 1996). Sce is different from standard restriction endonucleases (which have recognition sites of 4, 6, or 8 basepairs) by having a recognition site of 18 basepairs. This long recognition site is why homing endonucleases are sometimes referred to as “meganucleases.” When Sce is expressed in a vertebrate cell that has a Sce recognition site (“Sce site”) integrated into its genome, Sce will create a DSB at its recognition site (Rouet et al., 1994b). That is, Sce is active on genomic DNA with its accompanying chromatin modifications in vertebrate cells. If Sce could create a DSB in the genome, then
could that DSB serve as a stimulus for gene targeting by homologous recombination? To address this question, Sce sites were inserted into mutated reporter genes that were then integrated into the genome of different mammalian cells. A plasmid that expressed Sce and a plasmid that could serve as a repair template were then introduced into these cells. If homologous recombination occurred between the repair template and the mutated integrated reporter gene, then the reporter gene would become functional. The frequency of such spontaneous gene targeting is approximately one in a million; that is, for every million cells that the repair plasmid was introduced into, one cell would undergo gene targeting (Porteus and Baltimore, 2003). If, however, Sce was expressed simultaneously with the introduction of the repair plasmid, the rate of gene targeting could be increased by 1000-50,000-fold (up to 5% of transfected cells), depending on the conditions of the experiment (Choulika et al., 1995; Porteus and Baltimore, 2003; Rouet et al., 1994a; Sargent et al., 1997). Variations of this system, including different reporter genes, have been used to study a variety of cell types, including mESCs (Cohen-Tannoudji et al., 1998; Donoho et al., 1998). The range of different cell types suggest that the stimulation of gene targeting by a DSB is a universal characteristic and could be used in a general fashion. The limitation, of course, is that a Sce site has to be introduced into the desired target gene beforehand. Since no mammalian gene has an endogenous Sce site within it, this posed a potentially severe limitation to the application of DSB-induced gene targeting. To solve this problem, a method of creating DSBs at specific genomic sites needed to be developed. One possible way to create such DSBs is to re-design meganucleases, such as Sce, to recognize novel target sites (Porteus and Carroll, 2005). Progress is being made in this protein engineering problem (Ashworth et al., 2006). More progress, however, has been made using zinc finger nucleases to solve this problem and the development and uses of zinc finger nucleases will be discussed further below.

Using zinc finger nucleases to create site-specific DSBs

Type IIS restriction endonucleases are enzymes that bind a specific DNA recognition site but cleave DNA at a short distance from that site. FokI is an example of such a restriction enzyme: it binds to 5’-GGATG-3’ but creates a DSB nine bp away from the binding site. Proteolysis studies of FokI showed that its DNA binding domain and its DNA cutting (nuclease) domain were separable. Chandrasegaran and his colleagues hypothesized that if they fused a new DNA binding domain to the nuclease domain they could create a restriction enzyme with a novel DNA recognition site (reviewed in (Durai et al., 2005)). In a series of papers they fused the nuclease domain to first a homeobox DNA binding domain and then to a zinc finger DNA binding domain (Kim et al., 1996; Kim and Chandrasegaran, 1994; Kim et al., 1998). In both instances they created a protein that would cut DNA near the site bound by the DNA binding domain. These new proteins were initially called “chimeric nucleases.” Later those that were a fusion between a zinc finger DNA binding domain and the nuclease domain were called “zinc finger nucleases” or ZFNs.

The potential of ZFNs was in the nature of the zinc finger DNA binding domain. The zinc finger DNA binding domain was first identified by Klug and colleagues and the first crystal structure by Pabo and colleagues (Diakun et al., 1986; Pavletich and
Using homologous recombination to manipulate the genome . . .

The zinc finger DNA binding domain is a 30 amino acid domain that has a ββα structure in which two cysteines and two histidines chelate a single zinc ion, thereby stabilizing the entire domain. Each zinc finger recognizes a triplet of nucleotides. Stringing a series of individual zinc fingers together creates a protein that is capable of binding longer recognition sites. If, for example, a protein has three zinc fingers in a row, then it will bind a 9 bp site. There are over 700 proteins in the human genome that contain zinc finger domains although only a small fraction have had their cognate DNA binding site identified. The zinc finger crystal structure demonstrated that it binds DNA in a seemingly modular structure. Base contacts were made by individual amino acids and each finger seemed to bind its cognate three bp recognition site independently of its neighboring finger. By altering the individual amino acids within a finger and by combining different fingers with each other, one could create zinc finger proteins with new binding sites. It has been shown that zinc finger proteins can be designed to recognize a wide variety of sequences and suggested that they can be designed to recognize nearly every sequence (Pabo et al., 2001; Segal and Barbas, 2001). The ability to re-design zinc finger proteins to recognize new sites with relative ease is unique among the different classes of DNA binding domains. The combination that ZFNs can cut DNA and the ability to redesign zinc fingers to recognize new target sites suggested that one might be able to design ZFNs to create DSBs at a variety of different DNA sequences.

The initial in vitro studies of ZFNs suggested that a ZFN monomer could cut DNA but follow-up studies showed that efficient cleavage of DNA occurred when the nuclease domain dimerized (Smith et al., 2000). Thus to get efficient cutting of DNA, the ZFNs had to bind DNA in a particular orientation (see Figure 3). Two consequences of this required dimerization for cutting were: 1) that it increased specificity; and 2) it required that two ZFNs be made for each potential target site. The collaborative work of Carroll, Chandrasegaran and their colleagues showed that ZFNs could cut naked DNA in vitro and could cut DNA and stimulate a form of homologous recombination, called single-strand annealing, in Xenopus oocytes (Bibikova et al., 2001; Smith et al., 2000). The question was whether the ZFNs could create DSBs in mammalian genomic DNA and thereby stimulate gene targeting. To test this hypothesis, Porteus and Baltimore developed a green fluorescent protein (GFP) based gene targeting reporter system (Porteus and Baltimore, 2003). In this system, the GFP gene was mutated by the insertion of recognition sites for Sce and for a pair of ZFNs. This reporter was integrated as single copy in the genome of a human somatic cell line (HEK-293 cells). The reporter cell line was then transfected with a plasmid (repair template) that would correct the GFP mutation in the integrated reporter and either an expression plasmid for the Sce nuclease or expression plasmids for the ZFNs. They found that the ZFNs stimulated gene targeting as efficiently as Sce in this reporter system. In this work, Porteus and Baltimore used ZFNs for which the target binding sites were already known (Porteus and Baltimore, 2003).

The next step in the development of the ZFNs was to show that ZFNs could be prospectively designed to recognize novel target sites and stimulate gene targeting at those sites. Several groups have now shown that is possible and the most striking example was by Urnov and coworkers in 2005 (Alwin et al., 2005; Porteus, 2006; Urnov et al., 2005). In that work, ZFNs were designed to recognize a target site within exon 5 of the IL2RG gene. Mutations in the IL2RG are the most common...
cause of severe combined immunodeficiency (SCID) and exon 5 is a hotspot for disease causing mutations. They found in K562 cells, a human erythroleukemia cell line, that the IL2RG ZFNs could stimulate gene targeting at one allele of the IL2RG gene in 11% of cells and both alleles of the gene in 6% of the cells. Using these ZFNs they showed that they could first mutate both alleles of the gene in a single step and then correct both alleles in a second step. Finally, they showed that the ZFNs could stimulate gene targeting up to 5% in primary human T-cells. These results showed that DSB-mediated gene targeting by ZFNs could achieve rates that not only would be useful in an experimental setting but might also be useful as gene therapy for genetic diseases.

In addition to their development in mammalian cells, ZFNs are also being developed to stimulate gene targeting in a number of other systems that have also been refractory to experimental genomic manipulation by homologous recombination. Dana Carroll and his colleagues have continued to make important contributions in the development of ZFN technology and have used them to mediate gene targeting in D. melanogaster and to create DSB’s (the initiating event in gene targeting) in somatic cells of C.
Using homologous recombination to manipulate the genome...

elegans (Beumer et al., 2006; Bibikova et al., 2003; Morton et al., 2006). Other investigators have shown that ZFNs can be used to stimulate targeting in plants (Lloyd et al., 2005; Wright et al., 2005). The general nature of DSB stimulated gene targeting throughout the animal and plant kingdom suggest that as researchers make progress in applying ZFNs in one system, that those findings may readily translate to advances in other systems.

Combining DNA double-strand breaks and rAAV for gene targeting

Can one combine the use of rAAV and DSBs to stimulate gene targeting in mammalian somatic cells? In fact, combining the two technologies seems to be synergistic (Miller et al., 2003; Porteus et al., 2003). While the rate of gene targeting using rAAV is about 10-100-fold higher than using naked plasmid DNA, the rate of rAAV mediated gene targeting can be increased a further 100-fold by inducing a DSB in the target gene. These experiments were done using Sce as the nuclease to create the DSB but there is every reason to expect that if the DSB was created by ZFNs instead, a similar level of enhancement in gene targeting would result. This raises the possibility that the optimal way to stimulate gene targeting in human somatic cells is to infect cells with rAAV that provides the repair template and expresses the gene specific ZFNs. Hopefully in the next several years research supporting this technological paradigm will be published.

Future directions

The last several years have resulted in multiple publications that suggest that ZFN induced gene targeting has a promising feature. But the technology is still not widely used in mammalian cells because of several unresolved issues. Below I discuss three: 1) What is the best way to make the zinc finger portion of the ZFN?; 2) How to limit the cytotoxicity and off-target effects of ZFNs?; and 3) Expanding the use of ZFNs to different gene targets and in different cell types.

1. WHAT IS THE BEST WAY TO MAKE THE ZINC FINGER DNA BINDING DOMAIN?

The seemingly modular nature of zinc finger binding to DNA suggested that if one could make individual fingers that bound each of the 64 possible triplets that one could assemble a zinc finger protein that could recognize any target. Barbas and his colleagues have focused on this problem and have published sets of fingers for most CNN, ANN, and GNN triplets (Dreier et al., 2001; Dreier et al., 2005; Segal et al., 1999). In addition, Sangamo Biosciences has also published a set of fingers for the 16 GNN triplets (Liu et al., 2002). Several investigators have used these published data sets to assemble zinc finger proteins (hereafter, called a “modular-assembly” approach) and made new zinc finger proteins (either transcription factors or nucleases) that are active (Alwin et al., 2005; Beumer et al., 2006; Porteus, 2006; Segal et al., 2003). These successes have been limited to sequences that contain mostly or entirely GNN triplets as part of their target site. There may be a publication bias in these results, however, as unsuccessful attempts to make new zinc finger proteins would
not be published and an unbiased study of the success of making proteins using the modular-assembly approach has not been performed. Thus, it is unclear whether one should expect nearly every modular-assembly zinc finger protein to work or whether only a small fraction will work. While generating this data is not particularly glamorous, it is critical for the further advancement of the field. There is now a published protocol about how to use modular-assembly to make zinc finger proteins and zinc finger nucleases and two websites, referenced within that protocol, that also provides guidance if researchers are interested in pursuing this approach (Wright et al., In press).

There are several possible reasons that the modular-assembly approach might not work. The first is that the original collection fingers may not make fingers that make specific contacts with each of the target nucleotides. Segal and his colleagues recently published a crystal structure of a six-finger protein made by modular-assembly that binds its eighteen bp target sequence with sub-nanomolar affinity (Segal et al., 2006). When they examined the structure, however, they found that of the 18 possible DNA-amino acid contacts only half had the expected hydrogen bond interaction. Of the 9 contacts that had the expected hydrogen bond interactions that seemed to contribute strongly to binding strength, 5 of them were to guanines. Thus, it is possible that despite the extensive work that went into selecting fingers that bind to ANN and CNN triplets, that the current set of zinc finger modules still may not be optimal for target sites of the non-GNN variety. There are many natural zinc finger containing proteins, such as nuclear hormone receptors, that bind target sites rich in A:T basepairs and perhaps better modules can be developed using these natural zinc fingers as a scaffold to select for ones with new binding site targets.

The second reason that modular-assembly may not work as efficiently as desired is that the binding of zinc finger proteins to their target site may not be completely modular. That is, there may be context dependence of the binding if the zinc finger to its binding site. This context dependence can be at the level of individual contacts between nucleotides and amino acids (the effect of one amino acid-DNA contact within a finer on another amino acid-DNA contact) and at the level of different fingers binding (the effect of one zinc finger binding its cognate target on the ability of another finger within the protein to bind its target site) (Pabo et al., 2001; Wolfe et al., 2000a; Wolfe et al., 2000b). If context is critically important in zinc finger binding, then the modular-assembly approach may have a low efficacy in generating high affinity zinc finger proteins. Instead of modular-assembly, selection based approaches, in which entire zinc finger proteins are selected in the context of the neighboring fingers. There are several different selection approaches available, including phage display, bacterial one-hybrid, and bacterial two-hybrid techniques (Durai et al., 2006; Greisman and Pabo, 1997; Hurt et al., 2003). We are presently comparing zinc finger nucleases made by modular-assembly vs. selected by bacterial 2-hybrid to generate data on the utility of these different approaches.

Another question about zinc finger nucleases is the optimal configuration. In several papers, ZFNs with three zinc fingers were used but in the Urnov et al. (2005) paper, ZFNs with four fingers were used (Urnov et al., 2005). Moreover, in the Urnov paper, they used a 4 amino acid linker between the zinc finger and the nuclease to target a binding site in which there was a five-nucleotide spacer between sites. Other work has generally used amino acid linkers of 5 or longer to target binding sites with
a six-nucleotide spacer between ZFN binding sites. Determining the optimal configuration of amino acid linker and nucleotide spacer will be an important advance in the field.

2. HOW TO LIMIT THE TOXICITY OF ZINC FINGER NUCLEASES?

Most zinc finger nucleases have three-zinc fingers and thus are designed to recognize a nine bp target site. Such target sites would be expected to occur ~11,000 times in a 3 billion bp genome. Fortunately, efficient cutting of DNA only occurs when the nuclease domain is able to dimerize, requiring two ZFNs to bind in a precise orientation. This increases the binding site to 18 bp, a length that statistically would be unique in a 3-billion bp genome. Unfortunately, zinc finger proteins also bind to sequences that are not exact matches. The consequence of this imperfection is that ZFNs have cytotoxicity when expressed. This cytotoxicity has been demonstrated in both mammalian cells and in D. melanogaster and are at least in part the result of creating off-target DSBs. For experimental uses, such off-target DSBs can probably tolerated, but if ZFN mediated homologous recombination is to be used for gene therapy in humans, these off-target DSBs must be essentially eliminated. Two ways of decreasing the cytotoxicity of ZFNs are to: 1) Improve the specificity of the zinc finger domain; and 2) limit the duration of expression of the ZFN.

3. EXPANDING THE USE OF ZINC FINGER NUCLEASES

An important future direction of research in ZFN mediated gene targeting is to expand its use. So far, only a handful of genes, and some of these are reporter genes, have been targeted using the technology. It is likely that as attempts are made at targeting other genes new problems will arise. Some of these problems might be anticipated, such as the effect of chromatin structure on the ability of ZFNs to bind and cut DNA and the effect of chromatin structure and cellular context (such as cell cycle stage) on whether cells repair the DSB by homologous recombination using the repair plasmid as template or not. In addition, unanticipated problems may occur in the future. Finally, most studies of ZFNs have been done in tissue culture cell lines. An intriguing application of ZFN mediated gene targeting is to be able to do somatic cell genetics in primary human cells, including stem cells of different ilk. Thus, an important area of future research is to directly study how ZFNs work in cells of this type.

Summary

Homologous recombination is an ancient natural process that cells use to both repair DNA double-strand breaks and to shuffle DNA. Initially homologous recombination was a powerful tool for experimentalists studying phage, bacteria and yeast but slowly the application of homologous recombination as an experimental tool to vertebrate and mammalian cells developed. While still generally limited to murine embryonic stem cells and chicken DT40 cells, the development of rAAV and the use of ZFNs are making it increasingly likely that homologous recombination will eventually become a tool for scientists studying mammalian somatic cells. There is even the
possibility that gene targeting by homologous recombination may become efficient and safe enough to use in gene therapy. For many, the use of homologous recombination to treat human disease would represent the pinnacle built on a mountain of work in the fields of the homologous recombination, protein engineering, gene delivery, and stem cells. While the pinnacle is now in site, there remains substantial work to be done to reach its top. Let's hope we get there in a careful fashion as fast as possible.

Acknowledgements

The work in the Porteus laboratory is supported by grants from the NHLBI, March of Dimes, Burroughs-Wellcome Fund, State of Texas and UT Southwestern Medical Center.

References


CAPECCHI, M. (1989) Altering the genome by homologous recombination. Science 244,
Using homologous recombination to manipulate the genome . . . 209

1288-1292.


Nature Biotechnol, 23, 967-973.


