ZINC-FINGER NUCLEASES: THE PAST AND FUTURE OF *IN VIVO* GENOME EDITING

Kaitlin VanderWeide

Advised by Dr. Jean Cardinale

Department of Biology, Alfred University College of Liberal Arts and Sciences

Contents

Figures

Abstract

The race to discover the solution to genetic disease has been a long one; only recently has it come to look like it is a race that may actually be won. Zinc-Finger Nucleases (ZFNs) are synthetic enzymes that can be designed to generate targeted double-stranded breaks in DNA, giving the scientific community a way to effectively stimulate genomic changes. With the exquisite control granted by the use of custom ZFN technology, the door has been opened to the age of *in vivo* genome editing. Through examination of the development of this synthetic nuclease technology, as well as its past and present uses, conjecture can be made regarding future use. This study uses Cystic Fibrosis as a model to which ZFNs may be applied, examining the possibility of using this technology as a treatment for this disease in the future.

Introduction

In this age of advanced medical practices, where vaccines and antibiotics have made it possible to treat most infectious disease, the eyes of the medical scientists of the world have turned to genetic disease as their next target. 11.1% of pediatrics and 12% of adult hospital admissions are due to genetic disorders [1]. Many of these people can now receive treatment to alleviate symptoms, but most genetic diseases are chronic in nature with acute recurring episodes, and sometimes are fatal. Devastating effects are seen in infant deaths: 3-5% of all births have a congenital malformation, 20-30% of all infant deaths are due to such disorders, and 30-50% of post-neonatal deaths are caused by genetic defects[1]. The numbers show that genetic disease has significant impact on society. There are, as with any kind of disease, two treatment options: treat the symptoms, or treat the disease by seeking a cure. We can treat many symptoms of genetic disease, but only recently have efforts to actually reverse it by changing a patient's genome been pursued.

Early attempts at gene therapy involved episomal (outside the chromosomes) application via various vectors to cells *ex vivo* and *in vitro*. In these studies, genes or gene fragments were forced into cell nuclei. The hope was that the DNA fragments would somehow be incorporated into the cells' genomes by molecular machinery. The results often showed that the fragments would be expressed very briefly, and then degraded over time. In other cases, the fragments were

incorporated, but there was no way to control where it was inserted, how much was inserted, or really, anything about the incorporation at all. This was extremely dangerous because random gene insertion can lead to cancers in whole organisms and cytotoxicity in many cells[2]. What scientists needed was a method to make targeted additions and deletions. A new possibility emerged with the growing understanding of DNA repair mechanisms.

Two extremely important mechanisms of DNA repair are the Non-Homologous End Joining (NHEJ) and the Homologous Recombination (HR) pathways. These mechanisms specifically repair double-stranded breaks in DNA. In eukaryotes, particularly in vertebrate species, NHEJ occurs 1000 times more often than HR [3]. NHEJ is an imprecise method of repair because it reconnects the two strands of DNA independent of the terminal sequences [3]; in other words, the repair of the break is random and error prone. If this should happen in a functional section of the genome, such as a gene or promoter region, there is a chance normal function could be disrupted by mutations induced by error prone NHEJ. HR, on the other hand, is an extremely precise repair mechanism that relies on the presence of a homologous sequence. The appropriate sequence for the locus can then be copied from the homologous sequence accurately and without inducing mutations as NHEJ does[3]. The HR pathway is also significant in that it can use a donor sequence (a sequence provided by an injected DNA fragment) to fill in the break if there are similar sequences at the ends of the donor sequence and on either side of the break[2]. In this way, a new DNA fragment could be effectively, permanently, and accurately integrated into a cell's genome.

Altering the genome by either disrupting a gene or by adding a DNA fragment of some sort now appeared feasible. All that was needed was a way to induce a targeted double-stranded break, some enzyme that could be customized to cut at extremely specific loci on demand. In theory, the shorter the recognition sequence is for an enzyme, the more likely it is for there to be multiple sites at which it could cut. There are multitudes of naturally found enzymes that can induce a double-stranded break in DNA, but these have set specificities and are generally not easy to alter. The odds of finding a natural enzyme that cuts in the precise desired location for gene therapies is fairly unlikely. A super enzyme with particular requirements would be needed for this. The criterion for such a protein would be first, that it be easily to manipulate in the lab, and second, be able to recognize a rather long, specific DNA sequence. The length of the

recognition site is important because, the longer the sequence of DNA base pairs is, the less likely it is for there to be an undesired off-target alternate binding site somewhere else in the genome.

What if such a super enzyme existed? With the advent of Zinc-Finger Nucleases (ZFNs) to the world of genetic research, the theoretical ideal became a reality. ZFNs consist of two domains: a non-specific nuclease (cutting) domain and a highly specific, DNA-binding Zinc-Finger protein domain. The benefit of building a nuclease using Zinc-fingers is that this particular protein is made of individual finger motifs that are easy to manipulate in a lab setting and can be linked together to have very long binding specificities. When this precision is combined with the versatility of the FokI nuclease domain that is not specific in its own right, but can make a double-stranded break in DNA, targeting sites in the genome at which NHEJ or HR could occur becomes a reality. The possibilities this new tool bring to light could, and have, revolutionized the genetic research field. To understand all the implications of, and uses for, ZFNs, one must start almost twenty years ago and examine the story of how this amazing new technology came to be.

Zinc-Finger Proteins

The journey to *in vivo* editing began in 1985 with the discovery of Zinc-Finger Proteins. While studying oocytes from *Xenopus laevis*, Miller, McLachlan and Klug discovered a repeating motif consisting of a zinc centered domain containing repeating cystidine and histidine residues[4]. It wasn't until sometime after this study that these proteins began to be referred to as Zinc-Fingers.

Zinc-Finger proteins (ZFPs) are the most common DNA binding motif in eukaryotic organisms, and are primarily found in transcription factors[5]. The fact that ZFPs are commonly found in eukaryotic transcription factors makes it likely that they would be an efficient tool to use for *in vivo* gene modification in many organisms of interest; if they already exist in whole living organisms, reintroducing them should not cause as many issues as a foreign or uncommon

substance could. A single ZFP consists of a zinc ion tetrahedrally coordinated between a two stranded β-sheet and a short α-helix (a ββα structure)[5]. Specific contacts are usually made with DNA bases at amino acid positions -1, 2, 3, and 6 in the α-helix in such a way that the protein finger motif essentially recognizes and binds a DNA triplet[6]; Adjacent ZFPs recognize adjacent DNA triplets (see Figure 1). Often, in engineered ZFP arrays, three fingers are linked together; thus the typical recognition site for a constructed ZFP is 9 base pairs long.

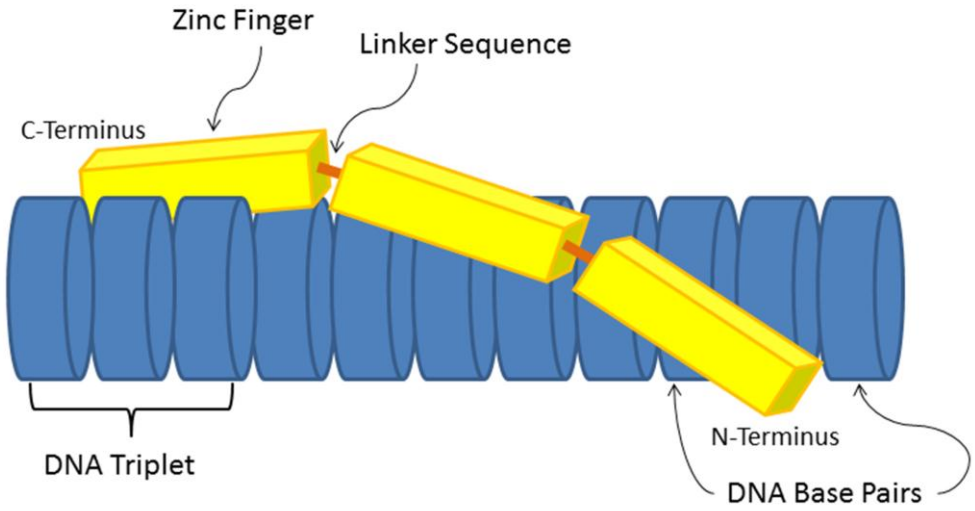

Figure 1: Zinc-Finger DNA Binding. Each finger motif contacts three DNA base pairs and is joined to other fingers in the array by a linker sequence. Each yellow box represents a single, complete finger motif within the zinc-finger protein. In other words, one yellow box represents two β-sheets and an α-helix.

Part of the energetics of ZFP binding depends on contact with phosphates in the DNA backbone. Eighty percent of the time, this contact is made by the position 7 histidine residue in the α-helix to the phosphate just to the 5' side of the binding site[7]. When following the reading conventions of 5'→3' in DNA and N→C terminus in proteins, ZFPs can be said to bind DNA in an antiparallel manner: the C-terminus of the protein points in the 5' direction and the N-terminus of the protein is oriented in the 3' direction [6]. Adjacent ZFPs are connected and held together by a conserved linker region[7]. It is as yet inconclusive if there is special significance to

this linker region with regards to DNA binding capabilities, but the region tends to be highly conserved[7]; sequences that are common- or "conserved"- across species tend to signal that they have a particular function. The high conservation of linker sequences in wild-type ZFPs indicates there might be some significance to this short polypeptide holding the ZFPs together.

As interest in ZFPs increased, it became apparent that the next step had to be forming an understanding of how these proteins bind specifically to DNA triplets. What amino acids in what locations in the helix bind to which DNA bases in the target site? In 1992, Desjarlais and Berg began using mutagenesis studies (See appendix A) to try to understand which amino acids in the key binding positions would recognize which DNA bases[8]. The advent of these studies necessitated a consistent and reliable system by which the binding of the new Zinc-Fingers could be analyzed and new mutations selected and purified. The most common system still used is the Phage Display system developed by Rebor and Pabo in 1994[9]. Progress continued relatively quickly in the generation of new ZFPs; the goal was to develop a novel Zinc-Finger motif to recognize each of the 64 possible DNA triplets, one for each possible combination of the four nucleotides (A, G, C, and T) and sets of three. By 1999, designer ZFP domains had been generated for each of the 16 GNN triplets[10], and by 2001, the ANN family had also been characterized[11] (The "N" standing for any nucleotide- A, G, C, or T). All that remained now was to put them to use.

The most significant usage of ZFPs in recent years has been as the DNA binding domain in Zinc-Finger Nucleases. Recalling the criterion previously laid out for the ideal nuclease for use in genome editing, ZFPs provide the manipulatable, high specificity required. The second key element, a non-specific DNA cutting domain, was discovered almost 5 years after ZFPs.

FokI Nuclease

The enzyme FokI was isolated from the bacteria *Flavobacterium okeanokoites* in 1989 [12]. It is classified as a Class IIS restriction enzyme [12]. These enzymes cut at a specified distance from their recognition site; that is, the actual cut in the DNA happens a certain number of nucleotides away from the site where the enzyme has bound to the DNA. The implication of such a mechanism is that the protein contains two separate domains: a recognition domain that

binds to the DNA strand and a cleavage domain where the actual cut of the phosphate backbone of the DNA occurs [13]. For FokI, the cleavage site on the DNA sequence was first thought to be nine to thirteen nucleotides from its recognition site [13]; however, like many early observations about the enzyme, this would prove to be not entirely true. Class IIS enzymes are also known to be modular in nature; they consist of distinct c-terminal cleavage and n-terminal recognition domains that can be manipulated separately [13]. This means that FokI's non-specific cleavage domain can be separated from its wild type DNA recognition domain and a new one with *any theoretical specificity* can be connected to the FokI cleavage domain to build a new enzyme (See Figure 2).

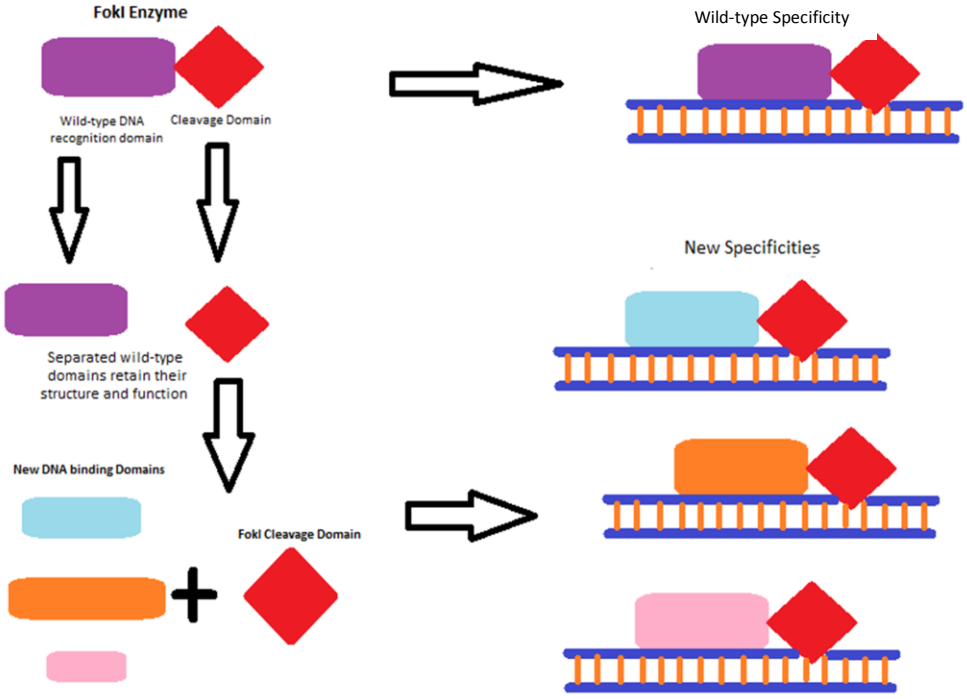

Figure 2: FokI Domain Recombination. Based on early studies, the cleavage and DNA recognition domains of wild-type FokI are separated and new DNA binding domains can be attached to the FokI cleavage domain to generate new enzymes with new specificities.

The first attempt to fuse the C-terminal FokI cleavage domain to a new N-terminal DNA recognition domain was in 1993, when researchers replaced FokI's wild-type N-terminal with an ultrabithorax homeodomain from *Drosophila* [13], generating a chimeric nuclease that had the recognition domain specific to *Drosophila* and an otherwise non-specific cleavage domain from FokI; Instead of cutting at a locus in *Flavobacterium*, the nuclease now would cut a locus in the

Drosophila genome. This was primarily a theoretical study with very little in the way of practical application. There was little significance in the choice of new domain except to prove the recombination could be accomplished. In order to generate a truly useful customized nuclease, the DNA recognition domain would need to be composed of a readily available molecule that could be easily manipulated experimentally to build highly specific recognition motifs, and, of course, said molecule must be compatible with the FokI cleavage domain. Zinc-finger proteins seemed to fit the bill nicely. The first time Zinc-Finger domains were used with FokI's cleavage domain to build a chimeric nuclease was in 1995 [14]. They have continued to be the popular partner to FokI in Zinc-Finger Nucleases ever since.

It is not only FokI's modular structure that makes it an excellent choice for creating novel nucleases: this enzyme has another advantage- two FokI molecules must dimerize at the cleavage site in order for cleavage to occur [15] (Figure 3).

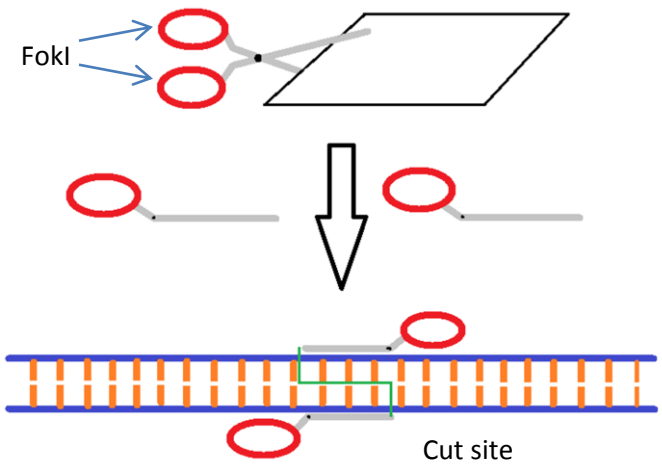

Figure 3: FokI Nuclease Activity. FokI operates under a similar principle to scissors: two halves must come together to cut both sides of a target. Two FokI molecules, unlike scissors, must orient anti-parallel to each other across the DNA strand and their nicks in the phosphate backbone of the DNA result in overhangs.

When FokI was originally isolated, it was assumed to cleave as a monomer because it exists in solution as a monomer [16]. It wasn't until 1998 that a ground breaking study proved definitively that two FokI domains must dimerize in order for DNA cleavage to occur [15]. This effectively doubles the number of nucleotides that are specifically recognized by the nuclease as a whole unit, increasing the ability to make an incredibly specific cut. Recall that a single Zinc-Finger motif binds a DNA triplet. In ZFNs, three to four fingers are linked together, and then attached to

FokI. This means the recognition site for one monomer ranges from 9 to 12 DNA bases, and a dimer's recognition site ranges from 18 to 24 bases plus a small gap of 4-6 bases (discussed later). Because of the gap between recognition and cleavage sites previously mentioned, there is a gap of several bases between the recognition sequences of the two monomers. This is called the spacer sequence [15]. The two cleavage domains overlap in such a way that four nucleotide 5' overhangs (also known as "sticky ends") are generated when the double stranded break is made[17].

This dimerizing ability, while greatly enhancing FokI's potential as a customizable nuclease domain, also has an added complication: Wild type FokI can act as either a homodimer or heterodimer. This means that for every intentionally generated Left/Right pair, there also is a chance of two other dimers forming and thus binding and cleaving at and undesirable off target site [18]. This creates a serious issue in designs meant for *in vivo* editing: any possibility of off target cleavage could result in cytotoxicity that could hurt the organism (Figure 4).

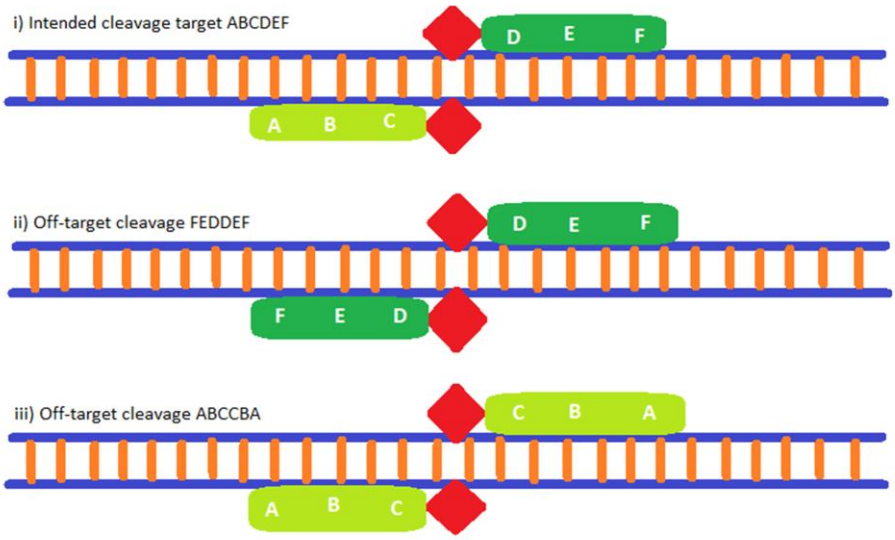

Figure 4: Heterodimers and Homodimers. For every cleavage site correctly recognized by a heterodimer, there are two possible off-target sites that can be recognized by the two possible homodimer pairings.

This became a problem of interest in several labs. In 2007, Miller et. al. generated +/- variants on wild type FokI. In these variants, amino acids at the positions in the FokI cleavage domain where dimerization occurs were substituted for positively and negatively charged amino

acids resulting in obligate heterodimer formation through charged interactions [18]; this operates on the same principle as magnets- positive repels positive and negative repels negative so there is only one way they will line up. This work was continued by Ramalingham et. al. in 2010, further working to refine the cleavage domain of FokI to force heterodimerization and thus greater binding specificity for the nuclease overall [19]. In 2011, Doyon et. al. introduced a variant on FokI with orthogonal (right angled) domains in which the shape of the enzyme- not just charged interactions- forced obligate heterodimer formation [20]. This study boasted that their new approach allowed for such exquisite specificity that researchers could be confident of targeting two sites within the genome at once [20], a truly ambitious statement. This would greatly widen the scope of ZFN based genome cleavage, opening the possibility of excising large chunks of DNA at a time and replacing entire genes.

Now that ZFPs and FokI had met, it was time to explore the possibilities of their pairing. ZFNs grew rapidly in repute and versatility over the years following their first appearance.

Zinc-Finger Nucleases

To orient to ZFNs as a whole unit, terminology is key. There are four key terms that are used to distinguish the pieces of the ZFN molecule and their placement on DNA for cutting. First, the nucleotide sequence of the DNA that the Zinc-Finger assembly binds to is called the recognition sequence. Due to the dimerization of the nuclease to generate its genomic cut, there is also a Left Module and Right Module of the nuclease. Between the binding Zinc-Fingers of these two modules is a gap of about six base pairs termed the spacer sequence; it is within this gap that the break is generated (See Figure 5).

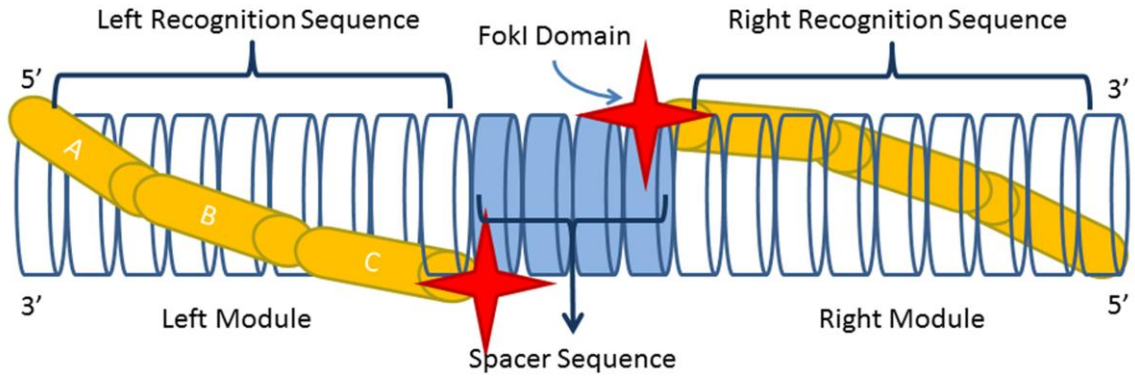

Figure 5: ZFN Terminology. The transparent cylinders represent double-stranded DNA labeled for 5′ and 3′ ends. The yellow cylinders are Zinc-Finger Proteins, and the Red star represents the FokI nuclease domain.

The process of designing and assembling ZFNs takes almost a month in the lab, assuming, that is, that everything goes smoothly the first time. The process involves six main steps [17]. First, a specific genetic target must be selected and searched for possible Zinc-Finger binding sites: both the Left and Right Modules must be accounted for and an appropriately sized spacer sequence must be present, as previously explained (Figure 5). Second, Zinc-Finger arrays must be selected for the chosen site, and their own coding sequences determined. These first two steps were done by hand at the time of the first publication of universal protocols, with researchers having to painstakingly search a target sequence using nothing but their own eyes and knowledge of existing Zinc-Finger binding specificities. Then, they had to use mutagenesis studies (Appendix A) to generate Zinc-Fingers to match their target. More recently, the Zinc Finger Consortium has brought together several online databases to make this process much easier [21]. Fu et. al. described a database in 2009 called ZiFDB (Zinc Finger Data Base) in which information about known Zinc-Finger specificity could be stored [22]. In 2010, an article appeared by Sander et. al. describing the latest advancement: a database known as ZiFiT which combined ZiFDB with NCBI's BLAST search tool [24]. This allows a target sequence to be input to the site and searched for possible sites that could be recognized by existing Zinc-Fingers combined into ZFNs, including the appropriate spacer sequence [23]. This greatly shortens the process of designing custom ZFNs, greatly increasing the efficiency with which they may be implemented.

The third step in the construction of ZFNs is to obtain coding sequences for the oligonucleotides that would encode the actual individual Zinc-Fingers. This is one of the longest and trickiest steps in the process of assembling custom ZFNs. First, remember that there is more to a Zinc-Finger than just the piece that binds the DNA; there is also the conserved framework consisting of the β sheets that are a large part of the finger's structure. Currently, two possible frameworks may be chosen from: the first comes from the mouse transcription factor Zif268 [17]. This framework has already been widely used and is extremely well characterized and therefore easy to work with. The other option for the framework is human Sp1C, based on a consensus sequence from a human transcription factor [17]. Sp1C-based Zinc-Fingers have been shown to have a slightly higher binding affinity in humans than the mouse framework. Once the framework is selected, the specific sequence devised from the Zinc-Finger identified in step 2 must be integrated into the framework. This results in the completion of the sequence of a single finger.

The fourth step in ZFN production is to clone the Zinc-Finger sequences in order in frame with the sequence for the FokI nuclease domain. This involves a PCR technique involving many overlapping primers (up to 7), and results in the generation of an expression plasmid that will encode the entire nuclease [17]. The fifth step follows logically: the ZFNs must be expressed and purified to test the viability of the protein. The sixth and final step of ZFN construction is *in vitro* testing of cleavage activity. This step requires the selection of delivery method as well, dependent upon what the use of the ZFNs will be. The very simplest method to deliver the expression plasmid is by direct injection into the nucleus of a cell [25]. This allows the most precise delivery possible, but only works well on small scale where one would work with a single cell at a time, making it really only a viable choice for some *in vitro* techniques. The second method of delivery is electroporation, where an electric pulse is used to force the expression plasmids into cells in solution [26]. This is slightly better on larger scale experiments, but is not viable for anything but *in vitro* or, in some select cases, *ex vivo* studies. It also bears a risk of cytotoxicity, and, by nature of how the plasmids enter the cell, risks damage to the cells being manipulated. For larger scale and more specific, targetable delivery, viral vectors may be used. The most commonly used right now are the Adenovirus and Integrase-Defective Lentivirus (IDLV) [27, 28]. The Adenovirus vector is highly customizable and is by far the most popular method of ZFN

delivery *in vitro* and *in vivo*. There are many databases online offering design and construction options [29].

The best way to understand these protocols is to actually perform them, but this is clearly impractical without access for a sophisticated lab that specializes in molecular manipulation. What can be done is closely review current studies and perform a theoretical experiment that would only exist on paper. The remainder of this review will attempt to accomplish both of these tasks.

Current Research

After years of research building up, it wasn't until the mid to late 2000s that the use of ZFNs became an accepted procedure. Zinc-Finger Nucleases are currently used in both research and clinical applications, a few examples of which are reviewed below:

Research Application: Generating Knock-out Rats (Geurts et. al. 2009)

The species *Rattus norvegicus*- better known simply as the Rat- has been a useful tool in disease research since the nineteenth century. In past genetic research, however, rats have had a severe limitation: there has not been a way to introduce site-specific, heritable mutations into the rat genome to study their effects [25]. This is much simpler in bacteria, plants, and amphibians, but they are not as good a model for human diseases and dysfunctions. Until now, alteration of the rat genome, much like that of the human genome, has been impractical due its large size and the lack of precision in genomic targeting tools.

With the advent of customizable ZFNs, such a targeting tool was now attainable. Geurts et. al. set out to prove that this was indeed a viable strategy. The goal of the study was to use the ZFNs to generate double-stranded breaks at targeted loci. Non-Homologous End Joining would then, being the more prevalent repair mechanism in vertebrates, incorrectly repair the break, thereby essentially "knocking-out" the targeted gene and making it non-functional. Being able to

do this allows the effects of loss of the gene to be studied, as well as providing a model on which different therapies may be tried.

The scientists who performed this study designed ZFNs to target the Immunoglobulin M (IgM) and Rab38 genes [25]. The team chose to directly inject DNA and mRNA encoding designed 5- and 6-finger ZFNs into the cytoplasm of single-celled rat embryos. Other studies had demonstrated that such intraplasmic injection allows the donor DNA to be expressed temporarily within the cells [25]. In this case, the ZFNs expressed succeeded in targeting the desired genes as shown using PCR techniques and DNA sequencing. As the rat embryos grew and divided, any genomic change were copied to the next line of daughter cells; thus mutations were distributed to all cells within the developing organism. The rats that developed from the modified embryos were labeled "founder animals."

It was found that 35 of 295 founder animals (12%) carried mutations in the targeted genes [25]. The IgM and Rab38 mutants produced had 25-100% disruption of gene expression, having from 3-187 base pairs deleted [25]. Variation in the size of the deleted section of DNA resulted from the imprecise nature of repair by NHEJ. Interestingly, no ZFN disruption was detected at any of the predicted off-target sites elsewhere in the genome (again analyzed using PCR and DNA sequencing techniques) [25]. From these initial knock-out rats, true breeding homozygous lines were raised.

This study was a great success and clearly opened the door for many new developments in the genetic research field. Knock-out rats allow us to explore the effects of gene loss and are useful model organisms for human genetic disease that allow scientists to discover and explore new treatments. This study then opens the door to a new age of genetic studies.

Clinical Application: ZFNs and HIV Treatment (Perez et. al. 2008, and Holt et. al. 2010)

Since its outbreak in the 1980s, Human Immune-deficiency Virus (HIV) has become an issue of global importance. Due to the high mutation rate of the virus, it has been nearly an impossible feat to develop an effective vaccine. However, there is an existing human mutation that grants virtual HIV immunity to homozygotes. A 32 base pair deletion in the CCR5 gene that is naturally found in many people of Western European descent seems to prevent the virus from entering T-cells without altering host immune function [30]. In theory then, if a mutation similar to

the delta 32 deletion could be achieved on demand in HIV patients, the virus could not reproduce and would eventually die out in the host, thereby eradicating the infection. To date, two approaches have been tried in association with Sangamo labs, CA.

In 2008, Perez et. al. set out to design ZFNs that would target the CCR5 gene in human T-cells at a site just upstream from the naturally occurring delta 32 deletion locus [31]. They generated four-finger nucleases specific to their chosen site with a goal of stimulating the NHEJ pathway to disrupt the gene [31]. They used an adenovirus vector to transduce a line of cells that was designed to report HIV infection [31]. The expression of the ZFNs within successfully transduced cells resulted in permanent disruption of the CCR5 gene as predicted. Before challenging the cells with HIV, CCR5 surface expression was analyzed and found to be reduced ten-fold in the ZFN treated cells [31]; this ten-fold reduction, though it may not sound like much, is still a measurable loss of CCR5 and thus should be a measurable loss of the ability of HIV to infect transduced cells. Correspondingly, the reduction in infection frequency among these cells compared to the controls was statistically significant (reported p-value <0.001) [31], suggesting there would also be clinical significance. Genetic modification at the target site was confirmed by PCR amplification and DNA sequencing.

Now that the scientists' theory had been proven with the easily quantifiable reporter cell line, it was time to test the ZFNs against CCR5 in a model line of CD4+ T-cells *in vitro*. This time, the researchers chose to transduce these cells via electroporation with small amounts of ZFN expression plasmid rather than the adenoviral vector used previously [31]. Post-transduction analysis of the cells revealed ~2.4% allele disruption [31]. Interestingly, this level persisted at just 2.3% in control cells not challenged by HIV, but in the culture in which the viral challenge exerted a selective pressure, a 30-fold enrichment in cells bearing CCR5 disruption was observed by day 52 of culture [31]; in other words, the cells with CCR5 disruption were selected for over the controls when pressured by HIV infection. Neither group exhibited any adverse consequences of transduction- cell viability remained high.

Yet another version of the experiment was repeated using primary CD4+ T-cells harvested from a variety of donors (rather than an identical cell line to give a more varied sample), again using an adenoviral vector (Ad5/35) to transduce the cells with the ZFN expression plasmid [31]. The results were rather spectacular: 40-60% of CCR5 alleles were

disrupted after ZFN expression and selective pressure was applied by introducing HIV [31]. It is not only this high rate of disruption that is exciting, however, but also the fact that the results were consistent across the board with many different donors.

This final treatment method has recently reached phase 2 clinical trials under supervision of Sangamo Labs [32]. T-cells are harvested from HIV-positive patients, and are then transduced *ex vivo* to disrupt the CCR5 gene using the method described above. The T-cells are then re-infused to circulation in the patient. Specific statistical results have not been released, but presumably, with the study recently advancing to phase 2 clinical trials, significant results are occurring, otherwise the funding would be cut. With the limited life span of T-cells in circulation, it is also presumable that this is not an absolute "cure" for the participants in the study; eventually, the conferred resistance should dissipate.

Unlike the Perez lab, Holt and associates chose to work with human hematopoietic stem/progenitor cells (HSPCs) [26]: the kind of stem cell found in red marrow from which blood cells (such a T-cells) are produced. The theory behind this approach was that, if the original stem cells could be modified to have a disrupted CCR5 gene, all cells arising from these modified HSPCs would theoretically inherit the same disruption and therefore be resistant to HIV infection. This was demonstrated in an HIV-infected patient who received a marrow transplant from an individual homozygous for the delta 32 mutation during treatment for leukemia. It was found that the patient receiving the transplant regained and maintained normal CD4+ T-cell levels without continuing on an antiretroviral drug regimen, indicating what amounted to a cure for his disease [26]. However, since homozygous delta-32 donors are in rather short supply, this is not a fix for all HIV patients. This is why the Holt study is significant: using their techniques, scientists can create "donor" cells from a patient's own HSPCs.

After testing several methods for delivery of ZFN plasmids to the HSPCs, the researchers settled on nucleofection via electroporation as the best route of transduction despite some loss in cell viability after treatment. This method achieved high levels of CCR5 disruption (~17%, with ~30-40% of those disruptions were bi-allelic) and thus seemed the most efficient choice [26]. A mouse model was used to test the system. Modified HSPCs were engrafted into 1-day old mice that had been treated with radiation to kill existing stem cells in their bone marrow. Eight weeks after transplantation, 40% of cells observed in circulation were human, indicating a

successful engraftment of the treated cells [26]. There were no apparent ill health effects seen in the mice. Eighteen weeks after transplantation, marrow harvested showed 11% disruption of the CCR5 gene [26]. It was time to challenge the mice with the virus. A group of mice with non-ZFN modified HSPCs was used to determine significance of ZFN disruption by acting as a control with which to compare the treated group. It was found that, following HIV challenge, mice with ZFN modified HSPCs had much higher levels of circulating and tissue-associated T-cells than the control mice [26]. Normal levels of T-cells in the treated mice remained constant long after infection while the control group experienced the sharp drop in T-cells normally seen in HIV infection.

With apparent success being observed in relation to both techniques, reports about clinical trials should reach public ears sometime in the near future. Due to the research of these labs, the future for people suffering with HIV infections looks much brighter.

Clinical Application: *In Vivo* ZFN Treatment for a Mouse Model of Haemophilia (High et. al. 2011)

In affiliation with the Division of Hematology of the Children's Hospital of Philadelphia, PA, High et al. set out to see if ZFN editing could be used to mitigate, if not correct, the blood disease Haemophilia B [27]. This so called "bleeding disease" is characterized by a greatly increased clotting time, and is the most common of the haemophilias.

A defective F9 gene that codes for Clotting Factor IX is the genetic cause of Haemophilia B. Ninety-five percent of the mutations in the F9 gene are located in exons 2-8 [27]. Due to the broad stretch of DNA that can contain the mutations, it was not possible to target a single section of the gene: all effected exons needed to be replaced. In order to manage this, the team designed and produced four-finger ZFNs that would cut within the first intron of the F9 gene [27]. The hope was that the Homologous Recombination pathway would allow DNA to be inserted in front of the effected exons by being read during transcription before the mutated section of the gene, and result in production of functional factor IX. Factor IX is primarily produced in the liver; thus it would be most efficient for this study to focus its editing efforts on liver tissue rather than every cell in the body. To target the liver specifically, a hepatotrophic (liver-seeking) adenovirus (AAV8-ZFN) was generated to deliver ZFNs to that organ and transduce only liver cells [27]. This

virus was injected simultaneously with a second engineered hepatotrophic virus that delivered donor DNA containing the functional wild-type F9 exons 2-8 (AAV8-donor) [27].

The mice used in this experiment had to have a human F9 (hF9) mini gene inserted into the murine gene coding for factor IX at the ROSA26 locus [27]. This is because mice lack the target site in the first intron of hF9 that the ZFNs were designed to recognize. This strain of mice was renamed hF9mut/HB, and would serve as a "humanized" model for research purposes. On day two of life, the hF9mut/HB mice were injected peritoneally with the combined AAV8-ZFN/AAV8-donor viruses [27]. The co-delivery of both ZFNs and donor DNA by the combined viral vectors was meant to stimulate genome repair in liver cells via the Homologous Recombination (HR) mechanism: after the attack of the ZFNs created a DOUBLE-STRANDED BREAK in the genome, the donor DNA would be incorporated into the genome. It was found at week 10 of life in the hF9mut/HB mice that there were circulating Factor IX levels as high as 2-3% that of normal in mice [27], bearing in mind that the model mice originally had no detectable circulating Factor IX, effectively converting a severe haemophilia to a milder condition. Repeated treatments could, perhaps, raise levels even more, but at this point, that is pure speculation as this is currently the only published study of its kind.

Cystic Fibrosis: A Theoretical Approach

Cystic Fibrosis (CF) is a life-threatening genetic disease that affects nearly 30,000 Americans [33]. In the 1950s, the life expectancy for someone diagnosed with CF was only about 10 years old. Currently, with many up and coming therapies available, life expectancy has increased to 30 years old and beyond [33]. The disease is caused by mutations in the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) gene which encodes a chloride ion channel expressed in the epithelial cells of the airway and digestive tract [34].

CFTR is a membrane spanning chloride ion (Cl^-) transport channel protein (See Figure 6)[34]. It is part of the family known as ATP-binding Cassettes, proteins that bind Adenosine Tri-Phosphate (ATP) to induce a conformational change that allows specific solutes to pass through the protein. In the case of CFTR, the conformational change triggered by ATP binding opens the channel of the protein to allow chloride ions to diffuse along the concentration gradient (from

areas of high concentration to areas of low concentration) [34]. In epithelial cells, CFTR keeps chloride concentration at a stable level in mucus secreted by these tissues. This, in turn, helps regulate mucus amount and viscosity. In individuals with CF, the loss of chloride regulation results in the production of large amounts of thick, sticky mucus being produced [33, 34]. In the digestive tract, this thick mucus occludes the secretion of necessary digestive enzymes in the intestines, as well as clogging the duct delivering that delivers enzymes produced by the liver and pancreas. The mucus especially interferes with the delivery of pancreatic enzymes, resulting in a condition known as pancreatic insufficiency [34]. It can also cause intestinal blockages on occasion and sterility in men due to blockage of sperm from passing from the testicles to the penis for ejaculation [34]. In the airways, the mucus can interfere with gas exchange in the tissues of the lungs; however, this is not what makes this disease often fatal. The mucus in the airways is an ideal breeding ground for many dangerous bacteria. It is these opportunistic infections that usually kill CF patients [33, 34].

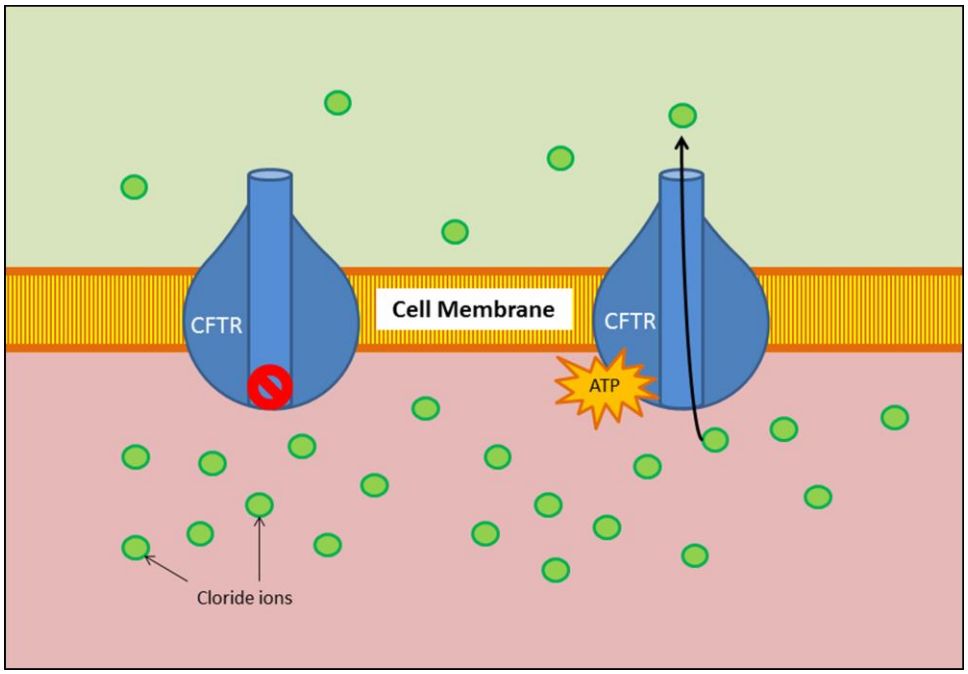

Figure 6: Normal CFTR Function. CFTR does not allow Cl- to cross the cell membrane until ATP binds. The Cl- ions then may diffuse along the concentration gradient through the protein channel.

CF is inherited as a recessive trait, meaning that a person must possess two copies (alleles) of the defective CFTR gene to have the disease [34]. Even one wild-type CFTR allele will produce enough functional protein channels to avoid the disease. That being said, roughly 1 in every 31 Americans carries one defective copy of the gene [33]. This puts the odds of two carriers meeting and producing offspring at about 1 in 961. The odds of two carriers having a child with CF are 1 in 4 due to the recessive inheritance pattern. Therefore, the odds in America of offspring have CF is 1 in 3844 based on numbers alone. In reality roughly 1000 people are newly diagnosed with CF every year [33].

Part of the challenge facing gene therapies for this disease is that there are 1906 different disease causing mutations known in the CFTR gene. These mutations can pertain to either protein functionality at the membrane or trafficking during processing in the endoplasmic reticulum (See Figure 7) [35]; functional mutations affect the ability of the channel to perform its function while trafficking mutations effect the protein during processing and packaging in the endoplasmic reticulum, causing the molecule to be degraded before it can be transported to the cell membrane [34].

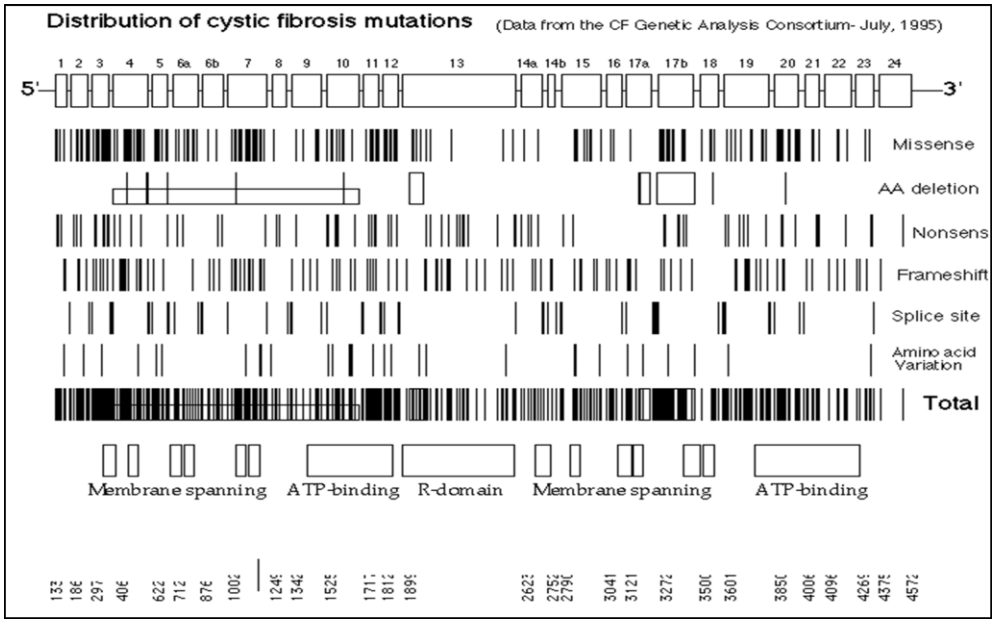

Figure 7: Distribution of Mutations on the CFTR Gene. This illustration is reproduced from the Cystic Fibrosis Mutation Database [35]. Each line depicts a known mutation within the CFTR gene corresponding to the type listed to the right hand side of the image.

Within each patient, one or many of these mutations may be present in the CFTR gene. Adding complication, the two CF alleles inherited may not contain the same set of mutations, so a patient may carry two different sets of mutations. Any potential gene therapy must therefore be tailored specifically to each and every unique patient, to target their own specific set of CFTR mutations, or else it must be tailored to replace the entire defective gene. The largest successful integrations using ZFNs have been less than one kilo-base pair (1000 nucleotides) [2]; the CFTR gene is 6129 base pairs long [36]. The length of the sequence to be inserted and the idea of completely removing the defective gene leaves some question as to whether complete gene replacement is feasible. The second difficulty facing the design of genetic therapies is the high turnover rate of epithelial cells as they are shed and replaced frequently in the body [37]; thus, any genetic changes made to an epithelial tissue would be transient by nature unless it could be somehow delivered to the germ layer. Since this is not, at this time, feasible, gene therapy must be targeted to the epithelial layer itself.

An attempt was made in 1999 by Harvey et. al. to deliver wild-type CFTR cDNA directly to airway epithelium where they hoped it would be expressed [37]. They used an adenoviral vector to deliver the normal copies of the gene directly into the airway epithelium, spraying recombinant virus containing the gene into the lower airway [37]. They found that there were no apparent ill-effects from the treatment, a very important factor when dealing with DNA. However, only about 5% expression of CFTR mRNA was observed in cells, and this expression only lasted about 30 days [37]. Another issue seen was that repeating the dosage of the vector eventually no longer produced detectable expression. The mixed results of this study are encouraging and disappointing; encouraging in that there was a short period with some success, disappointing in that the treatment was ultimately a failure. The CFTR DNA needs to be integrated into the epithelial genome for a more permanent expression.

The availability of wild-type CFTR DNA, and the ability of the gene to be delivered to the lung endothelium by an adenoviral vector (as demonstrated in the Harvey study described above) makes the application of ZFNs a logical next step; adenoviruses have already been shown to be effective vectors for ZFN application to tissues, and the wild-type CFTR DNA can act as a donor gene to be incorporated by HR. In theory then, ZFNs can be used to stimulate complete replacement of the CFTR gene. The experiment would follow a similar format to the High et. al.

2011 study of haemophilia [27]. First, a target must be established; in this case, due to the size of the proposed insertion, I propose searching for a cut site on either side of the gene, in both the 5' and 3' untranslated regions (UTR). The search would be done using the ZiFiT database available to the public online [38]. The results from this search would provide the design for the Zinc-Finger array from which the actual DNA coding sequence for the ZFNs could be established for reproduction. This leaves actual application which would involve many *in vitro* and *ex vivo* tests before being attempted *in vivo*. Then, depending on the results of these tests, clinical trials could potentially follow. For the sake of this paper, however, only the design part of this experiment will be carried out (See figure 8).

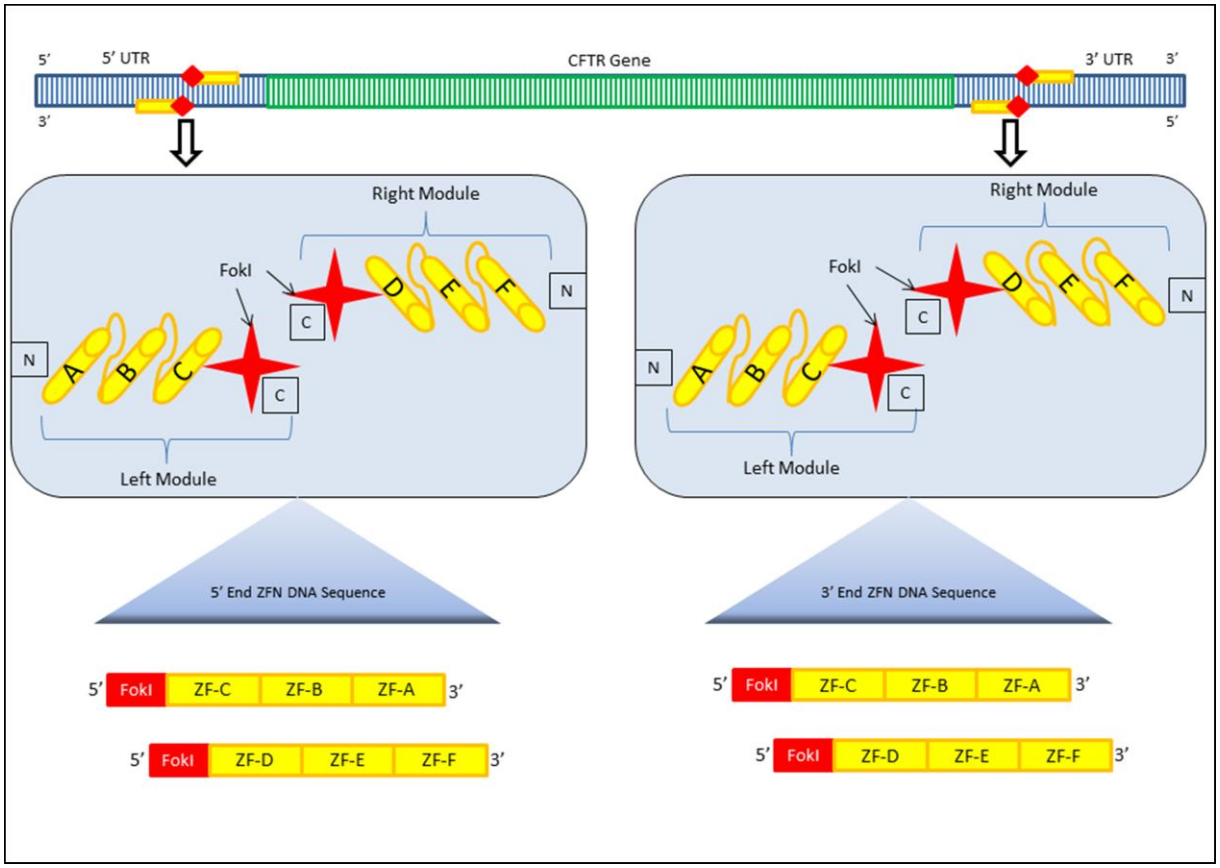

Figure 8: Experimental Design for CFTR replacement Therapy. After locating possible ZFN cut sites in the 5' and 3' UTR, the ZFP sequence can be determined. From the ZFP sequence, the DNA sequence for the entire ZFN can be derived.

The CFTR gene sequence was collected from the Cystic Fibrosis Mutation Database website [35]. The sequences for the 5' and 3' UTRs were collected. Using ZiFiT's online software, the sequences were searched for possible ZFN cutting sites. The search parameters were set for 3-finger nucleases with a 5, 6, or 7 base pair spacer sequence. Results were sorted by "Target Prediction," meaning they were sorted by the certainty of having the nuclease be active with confidence rated on a scale of 1-9 (9 being the best) for each module [39]. Graphic results from ZiFiT are shown in Figure 9 for the best match for each end. ZiFiT allows a user to use NCBI's BLAST tool to search for possible off-target cleavage site for any given set of ZFNs. The top two hits for the 5' UTR were examined using BLAST [24]. The first possible site was found to have one possible off-target cleavage site on Chromosome 19 near the ZFP 507 gene on the 3' side (Data not shown). The second possible ZFN arrangement was found to have no off-target cleavage sites detected by BLAST. This, then, was the set of proteins selected for the 5' UTR in the design. The 3' UTR proved trickier. The top four possible target sites were examined. The first was located in exon 27 of the CFTR gene, and was therefore not feasible. The second had two possible off-target sites: one on Chromosome 1 near the angiopoietin-related protein 1 precursor gene on its 5' side, and one on Chromosome X near NTF2-related export protein 2 isoform 3 on the 5' side. The third and fourth had more than 5 possible off-target sites each found by BLAST; this was deemed an unacceptable risk, and so they were discounted. Therefore, though 2 off-target sites was less than ideal, the second set of ZFNs was chosen for use in the 3' UTR.

ZiFiT is linked to ZiFDB in such a way that one should be able to click on the colored letters representing the Zinc-Finger arrays generated in ZiFiT and be taken to the DNA sequence that would encode the array. However, due to unknown technical difficulties, the ZiFDB has been down for the latter half of this study, so the rest of the proposed procedure cannot be continued at this time.

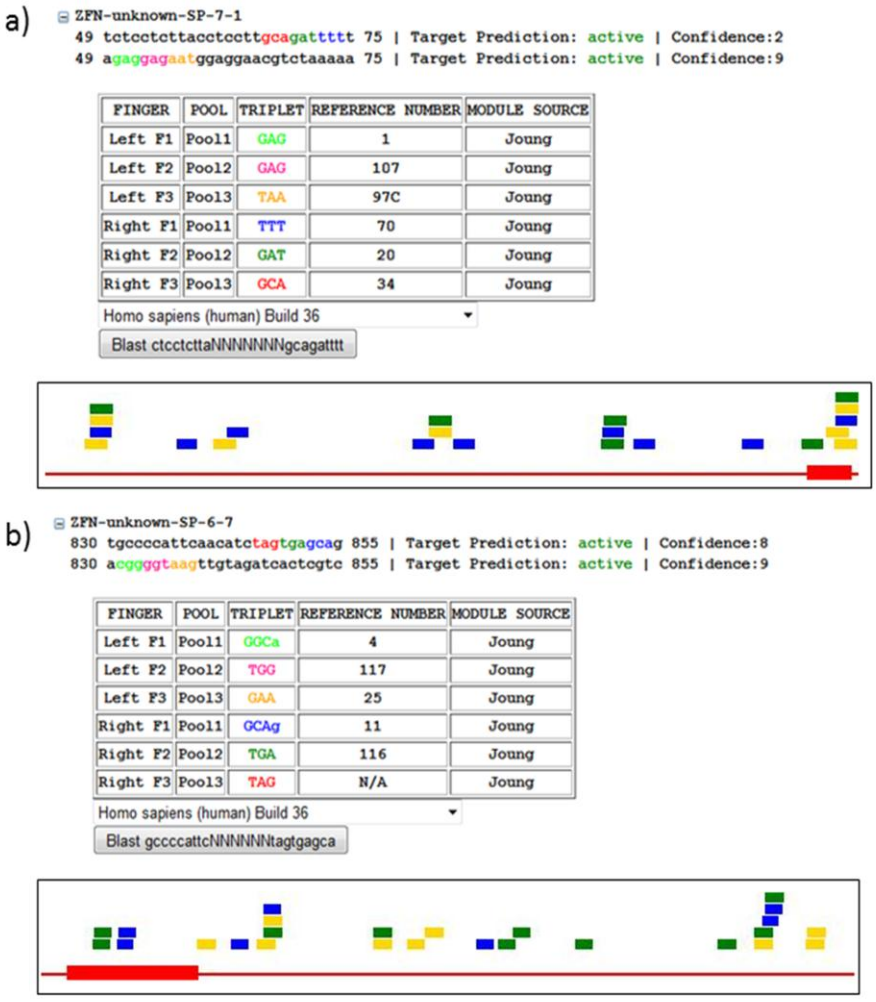

Figure 9: ZiFiT Results. The results given by the search of the 5' and 3' UTR of the CFTR include a representation of the potential sequence with the triplets to be recognized by known ZFPs each represented by a different color text. Results were sorted by Target Prediction, with the"active" denomination being preferable to "inactive." Confidence is rated 1-9 with 9 being the best chance of accurate binding. The included tables list the components of the generated arrays in the form of reference numbers and source. The BLAST link to check for off target sites is found beneath. The second piece of data is the graphic analysis of the query sequence with the thin red line representing the 5' and 3' UTR, the thick line representing exons. Blue bars represent a 5 base pair spacer sequence, Green a 6 base pair spacer, and Gold a 7 base pair spacer. A) Results for the 5' UTR. B) Results for the 3' UTR.

Conclusion

The 20 year journey from the discovery of ZFPs and FokI to the implementation of ZFNs staggers the imagination with its possibilities. As demonstrated in this study, many people have put countless hours into advancing this technology with these possibilities in mind. ZFNs open many doors in research and clinical applications for the study and treatment of both genetic and infectious disease. It is possible within this lifetime that diseases like HIV that have plagued the human race may become a thing of the past. Applications in cancer research and treatment can also be imagined with little difficulty.

It is possible that ZFNs compose one of the most significant current research fields existing. This being said, there is a surprisingly small amount of information about the technology available to the public. It is my belief that this lack of transparency will make the reception of ZFN treatment options hesitant at best and hostile at worst. The public opinion is already wary of genetic modification due to mass media science fiction representations. The popular game BioShock for example depicts a dystopian world where genetic experiments have gone horribly wrong, leaving most of the occupants as mindless zombies. As the protagonist, players can inject themselves with "plasmids," a technology that modifies the user's hand into weapons like a flame thrower [40]. Another, less far-fetched, example is the book *Beggars in Spain*, by Nancy Kress. This book opens with a married couple in their doctor's office designing their child. They pick everything about her appearance, and then add a new modification that will make the child never need to sleep [41]. The book explores the social implications of living as a known genetically modified human being, and puts a truly negative spin on it. With images like these pervading our culture, the biggest obstacle for ZFN stimulated genetic modification will be public opinion.

However, if this can be overcome, the possibilities are endless. If the genetic mechanism of a disease is known, a cure can be proposed. With all the questions out there to answer, one thing is quite clear: this is a very exciting time to be in the biological sciences field.

References

[1] Prevalence of Genetic Conditions/Birth Defects [Internet]. [Updated 2012]. Kansas University Medical Center; [Cited 17 May 2012]. Available from http://www.kumc.edu/gec/prof/prevalnc.html.

[2] Wu, Kandavelou, and Chandrasegaran. Custom-designed zinc finger nucleases: what is next?. Cellular and Molecular Life Sciences. 2007; 64(22): 2933-44.

[3] Takata et. al. Homologous recombination and non-homologous end-joining pathways of DNA double-stranded break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. The EMBO Journal. 1998; 17(18): 5497-5508.

[4] Miller, McLachlan and Klug. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. The EMBO Journal. 1985; 4(6): 1609-14.

[5] Klug and Schwabe. Zinc Fingers. FASEB Journal. 1995; 9 (8): 597-604.

[6] Pavletich and Pabo. Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Angstroms. Science. 1991; 252(5007): 809-817.

[7] Wolfe, Nekludova, and Pabo. DNA Recognition by Cys_2His_2 Zinc Finger Proteins. Annual Review of Biophysics and Biomolecular Structure. 1999; 3: 183-212.

[8] Dejarlais and Berg. Toward rules relating zinc finger protein sequences and DNA binding site preferences. Proceedings of the National Academy of Science of the USA. 1992; 89: 7345-49.

[9] Rebar and Pabo. Zinc Finger Phage: Affinity Selection of Fingers with New DNA-Binding Specificities. Science. 1994; 263(5147): 671-673.

[10] Segal et. al. Toward controlling gene expression at will: Selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. Proceedings of the National Academy of Science of the USA. 1999; 96: 2758-63.

[11] Segal and Barbas. Custom DNA-binding proteins came of age: polydactyl zinc-finger proteins. Current Opinion in Biotechnology. 2001; 12: 632-37.

[12] Kaczorowski, Skowron, and Podhajska. Purification and Characterization of the *FokI* restriction endonuclease. Gene. 1989; 80(2): 209-216.

[13] Kim and Chandrasegaran. Chimeric restriction endonuclease. Proceedings of the National Academy of Science of the USA. 1994; 91: 883-887.

[14] Kim, Cha, and Chandrasegaran. Hybrid restriction enzymes: Zinc finger fusions to *FokI* cleavage domain. Proceedings of the National Academy of Science of the USA. 1996; 93: 1156-60.

[15] Bitinaite et. al. *FokI* dimerization is required for DNA cleavage. Proceedings of the National Academy of Science of the USA. 1998; 95: 10570-75.

[16] Kaczorowski, Skowron, and Podhajska. Purification and Characterization of the *FokI* restriction endonuclease. Gene. 1989; 80(2): 209-216.

[17] Carroll et. al. Design, construction and *in vitro* testing of zinc finger nucleases. Nature Protocols. 2006; 1(3): 1329-41.

[18] Miller et. al. An improved zinc-finger nuclease architecture for highly specific genome editing. Nature Biotechnology. 2007; 25(7): 778-785.

[19] Ramalingam et. al. Creating Designed Zinc-Finger Nucleases with Minimal Cytotoxicity. Journal of Molecular Biology. 2011; 405: 630-641.

[20] Doyon et. al. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nature. 2011; 8(1): 74-79.

[21] Zinc Finger Consortium Software Tools [Internet]. [Updated 2012]. The Zinc Finger Consortium; [Cited 17 April 2012]. Available from www.zincfingers.org/software-tools.htm.

[22] Fu et. al. Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays. Nucleic Acids Research. 2009; 37: 279-283.

[23] Sander et. al. ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. Nucleic Acids Research. 2010; 38: 462-468.

[24] Basic Local Alignment Search Tool (BLAST) [Internet]. [Updated March 2012]. NCBI; [cited 2012 April 18]. Available from: http://blast.ncbi.nlm.nih.gov/.

[25] Guerts et. al. Knockout Rats Produced Using Designed Zinc Finger Nucleases. Science. 2009; 325(5939): 433-445.

[26] Holt et. al. Zinc finger nuclease-mediated CCR5 knockout hematopoietic stem cell transplantation controls HIV in vivo. Nature Biotechnology. 2010; 28(8): 839-847.

[27] High et. al. *In vivo* genome editing restores haemostasis in a mouse model of haemophilia. Nature. 2011; 475: 217-223.

[28] Lombardo et. al. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nature Biotechnology. 2007; 25: 1298-1306.

[29] ViraQuest [Internet]. [Updated 2011]. ViraQuest, Inc; [cited 2012 April 2012]. Available from http://viraquest.com/?referrer=adwords.

[30] O'Brien et. al. HIV-1 infection in a man homozygous for CCR5 delta 32. The Lancet. 1997; 349: 1219.

[31] Perez et. al. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nature Biotechnology. 2008; 26(7): 808-816.

[32] Sangamo BioSciences Announces Initiation of Two New Phase 2 HIV Clinical Trials [Internet]. [Updated 2012 Jan 9]. Sangamo BioSciences, Inc; [cited 2012 April 18]. Available from http://investor.sangamo.com/releasedetail.cfm?ReleaseID=637760.

[33] About Cystic Fibrosis [Internet]. [Updated 2012]. Cystic Fibrosis Foundation; [cited 2012 April 18]. Available from http://www.cff.org/AboutCF/.

[34] Chiaw, Eckford, and Bear. Insights into the mechanisms underlying CFTR channel activity, the molecular basis for cystic fibrosis and strategies for therapy. Essays in Biochemistry. 2011; 50: 233-248.

[35] Cystic Fibrosis Mutation Database [Internet]. [Updated 2011 April 25]. Cystic Fibrosis Consortium; [cited 2012 April 18]. Available from http://www.genet.sickkids.on.ca/Home.html.

[36] Cystic fibrosis transmembrane conductance regulator [Homo sapiens] [Internet]. [Updated 1991]. NCBI; [cited 2012 April 18]. Available from http://www.ncbi.nlm.nih.gov/protein/AAB20219.2.

[37] Harvey et. al. Airway epithelial CFTR mRNA expression in cystic fibrosis patients after repetitive administration of a recombinant adenovirus. Journal of Clinical Investigations. 1999; 104: 1245-55.

[38] ZiFiT [Internet]. [Updated April 2012]. Zinc Finger Consortium; [cited 2012 April 18]. Available from http://zifit.partners.org/ZiFiT/Introduction.aspx.

[39] Sander et. al. Predicting success of oligomerized pool engineering (OPEN) for zinc finger target site Sequences. BMC Bioinformatics. 2010; 11(543): 1-11.

[40] Bioshock [Internet]. [Updated 2012 April 9]. Gamespot; [cited 2012 April 18]. Available from http://www.gamespot.com/bioshock/.

[41] Beggars in Spain [Internet]. [Updated 2012]. Good Reads; [cited 2012 April 18]. Available from http://www.goodreads.com/book/show/68333.Beggars_in_Spain.

Appendix A: Mutagenesis Studies and Novel Zinc-Finger Specificities

Zinc-finger proteins (ZFPs) consist of two β-sheets and an α-helix (See **Zinc-Finger Proteins** section). The α-helix is where the specific protein-DNA interactions occur. The amino acids that construct the α-helix determine what unique DNA triplet the Zinc-Finger will bind to [6]. It has been the quest of scientists since the discovery of ZFPs to decode the relationship between the amino acid sequence of the α-helix and the triplet of DNA bases to which it will bind.

Mutagenesis studies have been the primary method of generating new ZFPs to study the relationship between α-helix composition and DNA triplet recognition [8]. In these studies, Polymerase Chain Reaction (PCR) techniques are used to generate point mutations in the form of substituting a single nucleotide in the α-helix region of the ZFP DNA sequence (See figure A.1). Overlapping primers are generated which contain the targeted change so that, as the PCR reaction progresses, the change will be integrated into the products. This method allows controlled, targeted mutations to be made to the gene one nucleotide at a time. Each new ZFP generated by this method can then be tested for functionality and specificity of binding.

Each novel ZFP provides more data regarding the code that links amino acid composition in the α-helix to the DNA base that it can bind. Every step towards understanding that concept leads scientists closer to having full control over their prospective tools.

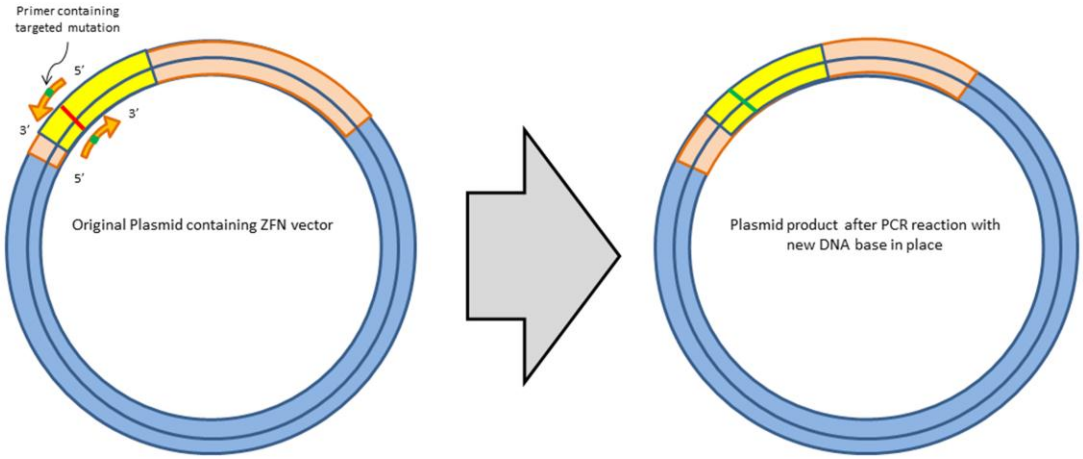

Figure A.1: PCR Mutagenesis studies. These studies are often performed with circular plasmids. On the plasmids drawn here, the pink arc represents the structural components of the ZFP. The yellow arc represents the α-helix encoding region. The red stripe is the nucleotide to be replaced. On each orange primer arrow, a green stripe represents the new nucleotide being added. The product plasmid shows a green stripe substituted for the original red to show nucleotide substitution.

Appendix B: Glossary

Allele*:* A variation on a particular heritable trait found on homologous chromosomes. (For example, there is an allele for color-blindness and an allele for normal color vision that can exist on the X chromosome. The inheritance of the alleles determines the fate of color vision in the individual.)

cDNA: Also called "copy DNA," it is a DNA copy of a gene that is reverse-transcribed from mRNA so that it contains only exons, no introns.

Conformational change: The conformation of a protein refers to its shape. A conformational change is a change in protein shape, often caused by the binding of another molecule to the protein. In the case of ZFNs, the acts of binding to the DNA recognition sequence and the dimerization of two FokI domains cause a conformational change in FokI that actually generates the nick in the DNA backbone.

Cytotoxicity: Literally "cell-toxic," a condition that kills a cell in some way. ZFNs can be cytotoxic if they cleave off target in the genome and disrupt a critical gene.

DNA Structure: DNA (Deoxyribose Nucleic Acid) is composed of subunits called nucleotides. A single nucleotide consists of a Deoxyribose sugar molecule attached to a phosphate group and a nitrogenous base. The base can be one of four possible structures represented by the letters A, G, C, or T. Bonds form between the sugar of one nucleotide and the phosphate of another to form the sugar-phosphate backbone of the DNA molecule. Two strands of DNA align antiparallel to one another and are held together by the pairing of nitrogenous bases: A to T, and G to C. Two nucleotides joined by a hydrogen bond between their nitrogenous bases are referred to as a base pair.

Domain: A unit of a protein that maintains a specific function. Example: Nucleases have a DNA binding domain and a DNA cleavage domain.

Double-Stranded Break: A break in a strand of DNA that cuts both strands.

Episome: A piece of DNA in the nucleus of a cell that is not part of a chromosome.

Epithelium: Tissue lining organs that are connected to the external environment (i.e.; the airways, the digestive tract, and the skin).

Exons: The parts of a gene that code for the protein product.

***Ex vivo*:** Referring to an experiment in which a tissue or cell sample is taken from an organism, then worked with outside the body.

Heterozygous: Referring to the possession of two or more different alleles for the same gene.

Homologous Chromosomes (also "homology"): During fertilization, an individual receives one of each chromosome from each parent. The same chromosomes received from each parent are referred to as "homologous chromosomes." Homology refers to the same gene sequence in different locations (i.e.; the same sequence on the other chromosome in a homologous pair).

Homozygous: Referring to the possession of two identical alleles for the same gene.

Introns: The parts of a gene that do not code for the protein product (note: also known as "junk DNA"). Generally contains the genetic instructions for splicing together exons.

***In vitro*:** Term for an experiment that takes place entirely outside the body (also referred to as a "test tube experiment").

***In vivo*:** Term for an experiment that takes place entirely within a body, usually studying the activity of one or more tissues under differing experimental conditions.

Locus: The position of something within a DNA sequence, gene, or chromosome.

Motif: The smallest unit of a protein that still retains function (in ZFPs, this would be a single finger).

Nuclease: An enzyme that cuts DNA.

Plasmid: A circular piece of DNA.

Polymerase Chain Reaction (PCR): A technique for amplifying a particular section of DNA or for assembling a longer DNA sequence from several shorter ones. This technique is used to

assemble the coding sequence for ZFNs by joining the coding sequences for the finger motifs in order with the coding sequence for FokI cleavage domain.

Residue: Term referring to a subunit of a chemical structure.

Transcription: The process in protein assembly where the DNA coding is copied into an RNA messenger.

Transduction: The act of transforming a cell with outside DNA, often with a plasmid.

Translation: The step in protein production when amino acids are assembled in order according to the sequence of the RNA messenger.

Vector: Something that can transfer DNA into cells (i.e.; a virus).