DEVELOPMENT AND CHARACTERISATION OF A ZINC FINGER NUCLEASE SPECIFIC FOR THE HUMAN β-GLOBIN GENE

A thesis submitted to Imperial College London

for the degree of Doctor of Philosophy

By

Tommaso Vannocci

Gene Targeting Group,
Department of Haematology, Faculty of Medicine,
Imperial College London,
Hammersmith Hospital
Du Cane Road, London, W12 0NN

Declaration of Originality

I, Tommaso Vannocci, declare that the work presented in this thesis is my own and that it has not been submitted elsewhere for any award. Where appropriate, I have referenced all sources and clearly indicated the work done by others.

Abstract

β-thalassemia and sickle cell disease, which are caused by mutations in the β-globin gene, are two of the most common single gene disorders worldwide and the only available cure is allogeneic bone marrow transplantation that is limited by donor availability. Gene therapy, by delivery of a β-globin expression vector into autologous haematopoietic stem cells, is a valuable alternative but the technique is affected by unpredictable protein expression levels and, more significantly, by random integration of the vector and the risk of insertional oncogenesis.

Gene correction by Homologous Recombination (HR) with a DNA repair template would avoid the above-mentioned issues, becoming ideal therapeutic approach. Although spontaneous HR events are very rare, Double Strand Breaks (DSBs) at the target locus can greatly stimulate them. Therefore, the development of Zinc Finger Nucleases (ZFNs), which are customised endonucleases capable of cleaving any desired DNA sequence, has created the opportunity to design molecular tools to treat many rare monogenic disorders.

The aim of this thesis has been to develop a ZFN-based gene targeting at the β-globin locus as a basis for β-thalassaemia gene-correction therapy. Initially, six ZF domains specific for the β-globin gene were designed, assembled and tested using publicly available reagents, but these were found to have a low binding efficiencies. Therefore, a commercial approach was used to obtain a functional ZFN (ZFN4) and this was shown to produce DSBs at the β-globin gene in ≈ 1% of the transfected human erythroleukemia cells without detectably cleaving the most closely related sequence in the human genome (the δ-globin gene). Using a reporter assay, ZFN4 was also shown to promote gene correction through intrachromosomal HR but was found to be 20 times less efficient than the homing endonuclease I-SceI.

Ultimately, ZFN4-stimulated targeted integration of a drug resistance marker at the endogenous β -globin gene locus was tested: 95% of the drug-resistant cells were targeted while the absolute frequency compared to the whole cell population resulted to be 0.1% The collected data show that ZFN4-

mediated gene targeting of the β-globin locus is possible but further studies are required in order address the discrepancy between cutting and targeting efficiencies and to increase the absolute frequency of gene targeted cells.

Acknowledgements

I would like to thank my co-supervisor Prof Irene Roberts for her helpfulness on the clinical aspect of the project and especially my supervisor Dr Andy Porter for his on-going support, for sharing his knowledge and experience, for his ability to inspire in me the joy and pleasure of doing research and for being always willing to discuss about science also outside the project.

My gratitude also goes to UK Thalassaemia Society, Leuka and CHAMPS for funding this project.

I would like to thank all the people of the Gene Targeting group for their friendship and help. A special thank to Hitoshi Kurata for the advices and support in the first stages of the project.

Special thanks to my family: to my parents and my brother for their unconditional trust in me and to my girlfriend Camilla without whom I would feel lost. Thanks also to my extended group of friends for their support and for being by safety valve.

Abbreviations:

AAV	Adeno-Associated Virus
Amp	Ampicillin
BLAST	Basic Local Alignment Search Tool
bp	Base pair
Cam	Chloroamphenicol
CoDA	Context Dependent Assembly
DSB	Double Strand Break
dsDNA	Double stranded DNA
ES cells	Embryonic Stem cells
FACS	Fluorescence activated cell sorting
GFP	Green Fluorescent Protein
GT	Gene Targeting
HbA	Haemoglobin A
HBB	Human Betaglobin
HPRT	Hypoxanthine phosphoribosyl transferase
HDR	Homology Directed Repair
HR	Homologous Recombination
HS	HyperSensitive
HSC	Haematopoietic Stem Cells
IDLV	Integrase- Defective Lentiviral Vector
iPS cells	Induced Pluripotent Stem cells
ITR	Inverted Terminal Repeat
LB	Luria Bertani
LCR	Locus Control Region
LTR	Long Terminal Repeat
Kan	Kanamycin
Neo	Neomycin
NHEJ	Non-Homologous End Joining
OD	Optical Density
OPEN	Oligomerized Pool ENgineering

ORF	Open Reading Frame
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
SDSA	Synthesis Dependent Strand Annealing
SNP	Single Nucleotide Polymorphism
SSA	Single Strand Annealing
ssDNA	Single stranded DNA
ssONs	Single Stranded Oligo Nucleotides
TALEN	Transcription Activator-Like Effector Nuclease
TE	Tris EDTA
TS	Target Sequence
Zeo	Zeocin
ZF	Zinc Finger
ZF-DBD	Zinc Finger DNA Binding Domain
ZFN	Zinc Finger Nuclease

Table of Contents

List of Figures

List of Tables

1 Chapter 1: Introduction

Zinc Finger Nucleases (ZFNs) are a class of engineered, gene-specific, DNA cleavage proteins that have been selected as the preferred molecular tool to produce targeted gene editing in a variety of animal or human cell types and animal models. ZFNs produce Double Strand Breaks (DSBs) at specific target sequences that are then repaired by one of the cellular DNA-repair pathways. Researchers can take advantage of these repair mechanisms to induce a range of different gene modifications for scientific or therapeutic purposes.

This thesis concerns the development of ZFNs for promoting gene correction of disease-causing mutations at the human β-globin gene with a particular focus on β-thalassaemia mutations.

As an introductory overview, the nature of the β-thalassaemia disease, of gene therapy in general and of gene correction in particular will be considered.

1.1 Thalassaemias

Thalassaemias are hereditary autosomal recessive anaemias characterised by aberrations in synthesis and structure of the two different globin chains that compose the quaternary structure of the haemoglobin protein. Normal adult haemoglobin (HbA, $\alpha_2\beta_2$) is a tetramer formed by two different sub-units: α-globin and β-globin. Between 5% and 7% of the world population is a carrier for a globin variant including 1.7% of individuals who are heterozygous for α- or β-thalassaemia traits (Rund and Rachmilewitz 2005; Muncie and Campbell 2009; Higgs, Engel et al. 2012). It is estimated that 60,000 individuals (or 4.4 individuals every 10,000 births) are born every year with severe homozygous or compound heterozygosity for α- or β-thalassaemia traits (Rund and Rachmilewitz 2005; Muncie and Campbell 2009; Higgs, Engel et al. 2012) making thalassaemias one of the most common genetic diseases in the world.

Most of the individuals who are born with one of these hereditary diseases are from the Mediterranean, Middle East, Central Asia, India and China where

thalassaemia traits reach frequencies of 5-30% in specific ethnic groups (Muncie and Campbell 2009; Higgs, Engel et al. 2012). The greatest burden of disease is found in low- and middle-income countries where diagnostic, therapeutic and management approaches are poor and the frequencies of thalassaemias are probably underestimated due to the high rate of infant mortality that affected most of these countries till recent times (Weatherall 2011; Higgs, Engel et al. 2012). The diseases are therefore expected to rise in frequency due both to falling childhood mortality and also to the increase in migration.

The reasons for the high frequencies of thalassaemias found in tropical or sub-tropical areas of the globe are to be found in the selective advantage that this type of inherited haemoglobin disorders confers to the carriers that are more protected against malaria, a disease that is, or was, endemic in these sorts of climate (Weatherall 2011; Higgs, Engel et al. 2012).

1.2 β-thalassaemia

β-thalassaemia is one of the most common genetic diseases and is caused by any of more than 200 point mutations or deletions in the β-globin gene. When the mutation completely removes any type of β-globin expression the disease is called β^0-thalassaemia while if the mutation reduces the production of the protein, the disease is know as β^+-thalassaemia. The clinical effects, in the most severe cases, are chronic haemolysis, ineffective erythropoiesis and transfusion dependent anaemia (Maniatis, Fritsch et al. 1980; Higgs, Engel et al. 2012).

Although the severity of the disease may vary greatly from patient to patient even in the event of a shared β-globin genotype, β-thalassaemia can be divided into three broad categories: asymptomatic β-thalassaemia minor (also known as β-thalassaemia trait), in which the subject carries only one mutated copy of the gene (β/β^T), an intermediate disorder with some need for blood transfusion known as thalassaemia intermedia, which is usually caused by mutations on both alleles but at least one of them is mild (β^t/β^T or β^+/β^0), and thalassaemia major (β^T/β^T or β^0/β^0) in which the subject is a homozygote

(same mutation on both chromosomes) or a compound heterozygote (two different types of severe mutations) and all patients are blood transfusion-dependent (Weatherall and Clegg 2001; Rund and Rachmilewitz 2005; Higgs, Engel et al. 2012).

Patients with β-thalassaemia major are asymptomatic for the first three-six months of life due to the persistence of foetal haemoglobin (HbF- $\alpha_2\gamma_2$) but become progressively transfusion-dependent after that. Patients usually present with pallor and growth retardation but in more severe cases, where the diagnosis is delayed, they may also have abdominal swelling, skeletal abnormalities due to increase bone marrow haematopoiesis and jaundice due to haemolytic anaemia (Rund and Rachmilewitz 2005). If the patient is not transfused, death usually occurs by the age of ten years and, instead, if blood transfusion is carried out regularly, iron overload progresses and patients develop endocrinopathies (such as hypogonadism or diabetes mellitus), splenomegaly (enlargement of the spleen) and, later in life, most of the deaths occur due to cardiac complication secondary to accumulation of iron in the heart (Rund and Rachmilewitz 2005; Muncie and Campbell 2009).

Two principal treatments are now available for β-thalassaemia: blood transfusion and bone marrow transplantation. Although blood transfusion increases the quality and expectancy of life of affected individuals, it is only a palliative treatment that has to be carried out for all the life of the patient. In addition to the risk of transfusion transmitted infections, the treatment is also time consuming, with prolonged sessions every month, and leads to iron overload (Sadelain, Boulad et al. 2007). Iron overload is characterised by the accumulation of iron inside the heart, liver and endocrine glands leading to the complications described above (Rund and Rachmilewitz 2005; Muncie and Campbell 2009). For this reason blood transfusion is always supported by iron chelation therapies that increase the life expectancy of the patients but may also be cumbersome and painful with serious side effects (Rund and Rachmilewitz 2005) .

Finally, although blood transfusion with chelation allows patients with β-thalassaemia major to reach adulthood, this therapy presents major drawbacks that reduce greatly the quality of life and are economically expensive (Rund and Rachmilewitz 2005; Sadelain, Boulad et al. 2007).

Economic adversities are especially relevant when it is considered that thalassaemias are characteristic diseases of African, Middle Eastern, South-East Asian and Mediterranean areas where the cost of diagnosis and therapies are far too burdensome for the population and children with thalassaemia are often untreated or under-treated (Weatherall and Clegg 2001; Muncie and Campbell 2009).

Bone marrow transplantation with related donor haematopoietic stem cells is the only reliable cure but it is greatly limited by donor availability (only 30% of the patients with β-thalassaemia major has a matched sibling donor) and by economical issues (Sadelain, Boulad et al. 2007). Although low-risk patients (that is, patients with low levels of complications derived by other conditions) showed great improvements after bone marrow transplantation, mortality caused by transplant-related complications in patients with severe complications ranges between 5% and 30% depending on the age of the patients and their health conditions, especially in the case of patients with iron overload and hepatitis where the risk of graft rejection is increased (Rund and Rachmilewitz 2005; Sadelain, Boulad et al. 2007).

The lack of suitable donors may be partially solved by the use of matched unrelated transplantation, by the use of related or unrelated umbilical-cord-blood transplantation and by the creation of bone marrow donor registries; however, the mortality and morbidity of these approaches greatly limits their use for the majority of patients (Rund and Rachmilewitz 2005; Sadelain, Boulad et al. 2007).

1.3 The β-globin locus

The β-globin gene is located in the β-globin locus (HBB locus), a region of roughly 65 kb on the short arm of chromosome 11 (11p15.5) that comprises 5 different genes and expresses 4 different globins in addition to the β-globin protein.

These genes are all under the control of the same Locus Control Region (LCR, composed by five distinct DNase I HyperSensitive regions HS1, 2, 3, 4 and 5) and are expressed at different times during embryonic and foetal

development (Levings and Bungert 2002). The spatial distribution of the β–like globin genes in the chromosome reflects the order of expression of the globins during development. Epsilon globin is the closest to the LCR sequence and is expressed in the very early stages of the embryo while β-globin is the last of the globin cluster and is expressed in post-natal life (Efstratiadis, Posakony et al. 1980; Lawn, Efstratiadis et al. 1980; Maniatis, Fritsch et al. 1980; Levings and Bungert 2002) (Fig. 1.1).

Studies conducted by measuring the sensitivity of both the entire locus, and the LCR region, to DNAse I treatment show that the locus is generally DNAse I-resistant in cells that do not express β-like globin proteins indicating that the chromatin presents a condensed DNA structure. In erythroid cells, however, the β-globin locus is more accessible to nuclease treatment showing a higher degree of accessibility of the chromatin to possible transcription factors (Bulger and Groudine 1999; Levings and Bungert 2002). While the LCR locus is DNase I-sensitive at all stages of erythroid development, the genes of the β-globin locus are sequentially nuclease-sensitive with the ε-globin been highly accessible in the embryonic stage and the β and δ-globin being nuclease-sensitive in erythroid cells of adults (Gribnau, Diderich et al. 2000).

Although the role of the LCR region in opening the chromatin structure of the β-globin-like genes is not completely understood, the formation of a holocomplex (an higher order structure made of different proteins such as GATA-1 and EKLF that recruits activators, chromatin remodelling and transcription factors) at the LCR locus is vital for tight regulation and enhancement of the expression of β-globin-like genes (Levings and Bungert 2002). The Chromosome Conformation Capture (or 3C) assay allowed Kim et al. to show in hybrid MEL (Murine Erythroleukemia) cells containing a complete human chromosome 11 how two regions of the LCR (more precisely HS2 and HS3) are spatially close to highly transcribed β- and δ-globin genes and recruit transcription factors that interact with the two globin genes (Kim, Kim et al. 2012). The role of the LCR in the regulation of the β-globin locus has fundamental consequences for the development of gene therapy for both

β-thalassaemia and sickle cell anaemia, which will be discussed in Section 1.6.

All five β–like globin genes share the same basic structure and gene organisation with 3 exons separated by 2 introns at almost identical positions. This seems to indicate a common origin for the β–like globin genes and evolutionary studies indicated that the divergence between β/δ and ε/γ globin genes started around 200 million years ago, after the divergence between birds and mammals, while divergence between β and δ occurred much later, around 40 million years ago (Efstratiadis, Posakony et al. 1980). The first intron, of roughly 120-150 bp, is located between codons 30 and 31 while the second one, of 800-900 bp, is located between codons 104 and 105 (Maniatis, Fritsch et al. 1980). These two introns are conserved in both α- and β- globin genes of humans, mice, rabbits and chicken probably indicating that these interruptions were characteristic of the original proto-globin gene that, around 500 millions years ago, differentiated into the α- and β-like globin genes (Efstratiadis, Posakony et al. 1980; Maniatis, Fritsch et al. 1980).

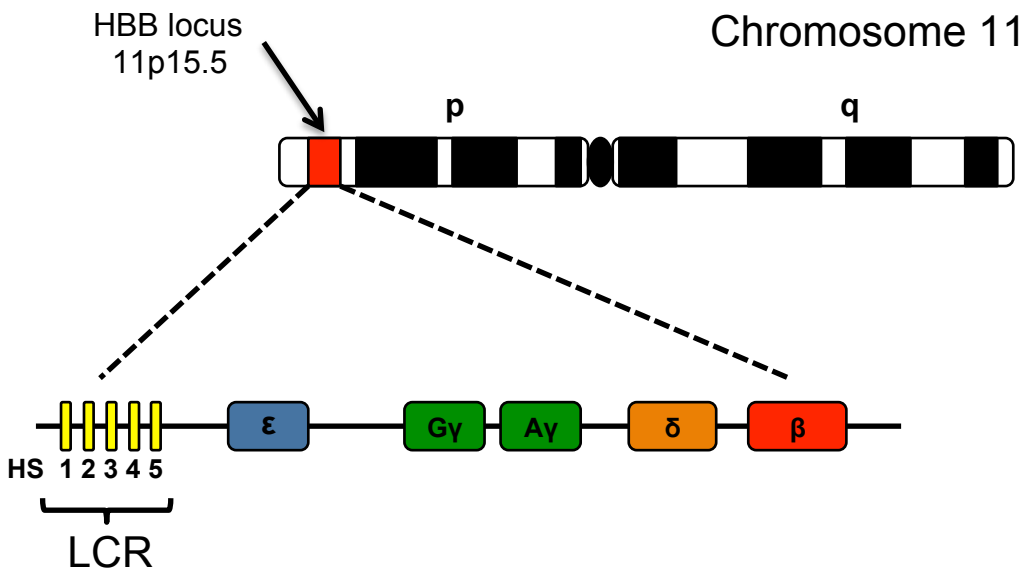


Figure 1.1 HBB locus.
Graphic representation of Chromosome 11 with highlighted HBB locus. The locus is composed by 5 different globin genes (ε, Gγ, Aγ, δ and β) that are sequentially expressed from the very beginning of embryo growth, through fetal development and, finally, to adult life. Expression of globin proteins is regulated by a single Locus Control Region (highlighted in yellow) characterised by 5 HS sites (HyperSensitive sites) all required for tight regulation and high expression of the different globins.

1.4 Gene Therapy

Traditional gene therapy involves the delivery of genetic materials in the cells or tissues of individuals with some kind of genetic disorder or acquired disease. The expression of the novel exogenous genetic material should compensate for the disease-causing mutation and therefore restore the normal activity of the target cells or tissue.

The initial hint of a possible gene therapy approach for the treatment of hereditary disease was envisioned in the late 1960s when studies on marked cell lines (Szybalska and Szybalski 1962) and on the mechanism of action of the SV40 polyomavirus (that stably integrates its own DNA in the genome of the target cell) (Sambrook, Westphal et al. 1968) indicated the possibility of introducing and expressing the desired genetic information into mammalian cells. Unfortunately, the techniques to handle and modify desired genes to be transfected in target cells by viral methods were still at the early stages (Friedmann 1992). However, the idea of gene therapy turned into a concrete possibility when molecular cloning techniques became the base for the creation of recombinant DNA that could be tailored to satisfy the needs of possible gene therapy approaches (Friedmann 1992).

Controversy about the feasibility of the approach and ethical debates about the possible use of gene therapy in human subjects divided the scientific community (Friedmann and Roblin 1972) but, thanks to successful preliminary results obtained by transfection of murine bone marrow cells by the non-viral calcium phosphate method (Cline, Stang et al. 1980; Mercola, Stang et al. 1980), Cline et al. decided to treat two β-thalassaemia patients' bone marrow cells with a vector carrying the β-globin gene and then to auto-transplant them back into the patients (Friedmann 1992). The decision to carry on with the trial was not backed by Cline's university (UCLA) and the resultant outrage of both the general public and the scientific community ruined Cline's reputation and greatly impaired the acceptability of gene therapy in the eyes of the world.

Arguably, the first successful gene therapy trial was carried out much later (2000) by Cavazzana-Calvi et al. who successfully treated two children (aged

11 and 8 months) affected by Severe Combined Immunodeficiency-XI (SCID-XI), an X-linked disease resulting from mutation of the gene encoding the interleukin-2 receptor gamma chain (IL2R-γc). Thus a retroviral vector capable of expressing the ILRR-γc gene was transduced ex-vivo into the patients' bone marrow cells that were then reinfused into the patients (Cavazzana-Calvo, Hacein-Bey et al. 2000). Ten months after treatment the two patients showed signs of clinical improvement and the levels of T, B and NK cells detected were found to be comparable to those of control subjects. Following this success, a further twenty children suffering from SCID-X1 were treated with the same gene therapy approach. Although seventeen out of twenty were successfully treated, showing signs of recovery similar to the ones found in the previous two patients, 31-68 months later five of the twenty developed T cell leukaemia due to the activation of a proto-oncogene (LMO2) by the enhancer located in the therapeutic vector which had integrated adjacent to the LMO2 gene (Hacein-Bey-Abina, Garrigue et al. 2008; Fischer, Hacein-Bey-Abina et al. 2010). The five patients were treated with chemotherapy and four of them recovered without losing the beneficial effect of the gene therapy but, unfortunately, the fourth child died. Traditional gene therapy through virus-mediated gene insertion had previously been linked to the risk of insertional oncogenesis. The random integration of the vector in the host genome can lead to misregulation of critical cellular pathways with later development of cancer (Sadelain, Boulad et al. 2007; Persons 2009).

Although the leukaemias associated with the SCID-XI trial increased concerns about the safety of gene therapy and indicated the need for a better understanding of the mechanisms involved in the integration of therapeutic vectors in the human genome, the SCID-X1 gene therapy programme had a much lower mortality (5%) than conventional bone marrow transplantation (25% mortality), the only other cure for this disease (Sheridan 2011).

A promising alternative to traditional gene therapy is *in utero* gene therapy. Foetal gene therapy has been considered a controversial approach with serious ethical and clinical issues such as the risk of germ-line modifications, risks to foetal and post-natal development and risks to the mother; usually the alternatives of embryo-selection and abortion are considered more favourably

due to their perceived reduced risk and also for ethical reasons (Coutelle and Rodeck 2002; Coutelle 2008). However, foetal gene therapy may offer practical advantages including an increased ability to target inaccessible tissues, the need for relative low doses of vector, permanent gene expression in stem cells and an increased immunological tolerance of the subject to vector antigens and therapeutic gene product (Coutelle and Rodeck 2002). At the moment, extensive studies have been carried out only on small and large animal models but promising results have been obtained for example on sheep's foetuses (a suitable model for human prenatal gene therapy) using ultrasound-guided transthoracic injection to deliver an adenoviral vector encoding a β-galactosidase (used as a marker for a proof of concept assessment) (Peebles, Gregory et al. 2004).

1.5 Viral vectors and non-viral delivery methods

Viral vectors have been, since the early stages of gene therapy, the preferred method (due to their naturally developed system for gene delivery) for the delivery of therapeutic genes into target cells and many different types of viral strains have been adapted for efficient transduction. The first step in the design of novel viral vectors has been the need to limit the ability of the virus to propagate or to be pathogenic. The initial cycle of all viruses can be divided into two different phases: an infection phase where the virus "connects" with the host cell and introduces its own genomic DNA (or RNA in the case of retroviruses) and a replication phase where the genetic information of the virus is replicated due to both regulatory viral genes and to the cell machinery itself. At a later stage, viral structural genes induce the cell in producing viral particles for the encapsulation of novel viral genetic information and then a final lytic stage allows the discharge of novel viral particle and the infection propagates.

The use of viral vectors for gene therapy focuses on the first infection phase used to deliver the desired DNA sequence into the target cell and, at the same time, aims to avoid uncontrolled replication and propagation of the viral vector. In order to do so, the standard approach has been to remove the

pathogenic genes and to separate viral coding genes from the cis-acting sequences that flank the viral DNA. The coding genes, which are essential for the replication and packaging of viral DNA, are expressed by plasmids or are directly integrated in the genome of producer cells (or packaging cells) while the therapeutic gene is cloned in the middle of the cis-acting sequences (Kay, Glorioso et al. 2001). In this way the therapeutic gene can be safely packaged into replication-deficient viral vectors that preserve only the infection properties of the original virus, do not uncontrollably replicate in the target cells of the patient and do not induce long term/chronic immunological response triggered by the delivery of those viral genes that encode for antigens.

Two main types of viral vectors exist: integrating and non-integrating vectors. Both of them present advantages and disadvantages. Integrating viral vectors are necessary for ensuring continued gene expression in actively replicating target cells; however, uncontrolled integration into the host cell genome is associated with a risk of insertional oncogenesis. By definition, non-integrating viral vectors avoid the dangers of random integration but can only be transduced effectively into quiescent cells because in replicating cells the therapeutic genetic material would be lost (or "diluted") during subsequent cell generations (Verma and Weitzman 2005; Sheridan 2011). Three of the main viral vectors used for gene therapy are described below.

Retroviral vectors: Retroviral vectors contain a single-stranded RNA molecule of roughly 11 kb that is retro-transcribed in a double stranded DNA molecule and than integrated randomly into the genome (Kay, Glorioso et al. 2001). In a normal retrovirus the genetic information is composed of three viral genes *gag* (encoding for capsid proteins), *pol* (encoding for the protease, reverse transcriptase and integrase) and *env* (encoding for the envelope glycoproteins) and these are flanked by two Long Terminal Repeats (LTRs) (Kay, Glorioso et al. 2001; Verma and Weitzman 2005). In order to generate a retroviral vector, the three *gag*, *pol* and *env* genes are moved from the viral genome to the genome of a packaging cell line while the therapeutic genetic material is cloned between the two flanking LTRs sequences (see Fig. 1.2). Thus, when transfected with the latter DNA construct, the packaging cells can

produce viral particles carrying an RNA genome encoding the therapeutic protein along with protein precursors of the protease, reverse transcriptase and integrase essential for the efficient integration of the vector into the host cell genome (Verma and Weitzman 2005). Retroviral vectors such as the Moloney retrovirus have been already used in clinical trials (Cavazzana-Calvo, Hacein-Bey et al. 2000) and lentiviral vectors (retroviral vectors based on the HIV-type viruses) are extensively developed due to the apparent reduced risk of insertional oncogenesis (Montini, Cesana et al. 2006). Lentiviral vectors have been recently used in a first series of clinical trials for AIDS gene therapy (Manilla, Rebello et al. 2005), for X-linked adrenoleukodystrophy, for SCID-XI (Cavazzana-Calvo, Hacein-Bey et al. 2000; Cartier, Hacein-Bey-Abina et al. 2009) and for β-thalassaemia (Bank, Dorazio et al. 2005).

Adenovirus vectors: Human adenoviruses are a broad family of fifty or more serotypes capable of targeting a variety of tissues, such as the eye, respiratory tract, urinary bladder and liver (Verma and Weitzman 2005). The viral genome is a double stranded DNA sequence of ≈ 36 kb encoding more than fifty different proteins. In adenovirus-based vectors, the bulk of the "early genes" E1, E2 and E4 (responsible for viral genome replication) are removed to make space to the therapeutic gene and also to avoid viral replication in the host cell. The recombinant therapeutic gene is flanked by Inverted Terminal Repeats (ITR) and a packaging recognition signal that are required for efficient formation of viral particles (Kay, Glorioso et al. 2001). Unfortunately, this type of vector has been associated with the first person to die in a gene therapy trial. The patient was suffering from a nitrogen metabolism disorder (ornithine transcarbamylase -OT- disorder) and was treated with a gene encoding OT via transduction with adenovirus and died of a severe inflammatory response to adenovirus which caused progressive organ failure (Couzin and Kaiser 2005). More recently, adenoviruses have been used in clinical trials for the treatment of cancers where immunogenicity is less of a concern (Ganly, Kirn et al. 2000; Heise and Kirn 2000).

Adeno-associated viruses: Adeno-associated viruses (AAVs) are a non-pathogenic family of human parvoviruses that need a helper virus like adenovirus to successfully replicate. Many different serotypes are known with the AAV2 serotype being the one that has been mostly used for the creation of recombinant viral vectors. The genetic information is carried as a linear single stranded DNA sequence of ≈ 5 kb and, in a AAV viral vector, only the two ITR sequences are left while the original viral genes (*rec* and *cap*) are replaced by the therapeutic cassette. Transduction with AAV vectors usually gives rise to an episomal therapeutic gene but can also randomly integrate in the genome and this integration has been associated with a slow but steady increase in the expression of the therapeutic gene that reaches a plateau after several weeks (Kay, Glorioso et al. 2001; Verma and Weitzman 2005). Initial clinical trials based on AAV2 vectors were carried out on cystic fibrosis patients by intranasal administration of a viral vector carrying the cystic fibrosis trans-membrane conductance regulator (CFTR) gene. Transduction efficiency and expression of the therapeutic gene were found to be satisfactory but high levels of gene expression lasted only for 60 days and due to antibody responses to AAV2 capsids developed by the patients, the treatment could not be repeated (Mueller and Flotte 2008; Conese, Ascenzioni et al. 2011). A similar transient expression was noticed in a clinical trial for haemophilia B where patients were transduced with an AAV2-FIX (factor IX) vector (Mueller and Flotte 2008). More encouraging results were obtained in patients suffering from Leber Congenital Amaurosis (LCA), an inherited retinopathy, where an AAV2 carrying the RPE65 gene were transduced by sub-retinal injection. These results showed no evidence that AAV2 gene therapy was associated with serious complications in humans and also showed that vision in the patients was improved although not with the efficacy found in dog models (Colella and Auricchio 2012) (see Figure 1.2).

Viruses are highly developed, naturally occurring systems for the accurate delivery of exogenous genetic material in host cells therefore have been extensively researched to be used in possible therapeutic applications but many aspects of this delivery method have to be improved. The obstacles that should be overcome are the possible harmful effects of an immune response to viral capsid proteins, to transduced cells that the organism does not

recognise anymore as part of the original tissue or the immune response to the product of the therapeutic gene. Additionally, there is still the risk of insertional oncogenesis that could be overcome by the development of vectors capable of integrating at safe harbours in the genome of the target cell or by the use of episomes that do not integrate but can still replicate with the host cell. The small proportion of host cells that are transduced in the target tissue, the possible low levels or unregulated expression of the therapeutic gene and the risk of targeting aspecifically other tissues (including the germ lines) are reasons for concerns as well (Kay, Glorioso et al. 2001). Tissue-specific viral vectors can be engineered by modification of the proteins that recognise specific cell receptors or, alternatively, by using tissue-specific promoters upstream of the therapeutic gene.

Non-viral delivery methods: although viral vectors are the preferred method for gene delivery, alternatives have been sought in order to overcome the risks and downfalls related to the use of viruses. One approach used novel S/MAR (scaffold/matrix attachment region) minicircle vectors with reduced CpG bacterial sequences (to avoid the immune response triggered by the Toll-like receptors) which were successfully delivered into the liver of a mouse model by hydrodynamic injection in the tail vein. This led to higher expression of a luciferase marker gene as well as more prolonged expression compared to other similar vectors (Argyros, Wong et al. 2011). Alternatively, cationic lipids, such as GL67, have been used for the delivery of a CFTR therapeutic gene in cystic fibrosis patients (Alton, Stern et al. 1999) or “lipidoids” developed by Anderson et al. have been used for the delivery of RNA interference molecules (Akinc, Zumbuehl et al. 2008).

Alternatives also exist to traditional DNA therapeutic genes like the use of non-viral delivered mRNA molecules that would be transiently expressed directly in the cytosol without having to penetrate the nuclear membrane (Yamamoto, Kormann et al. 2009) or exon-skipping techniques based on the delivery of antisense oligonucleotides. This last approach can be successfully used, for example, in the case of muscular dystrophy where mutations that disrupt the reading frame of the dystrophin gene can be “avoided” by alternative splicing of the mRNA molecule mediated by antisense

oligonucleotides (Foster, Popplewell et al. 2012). Exon skipping for Duchenne muscular dystrophy has been successfully tested in animal models (Malerba, Sharp et al. 2011; Yokota, Hoffman et al. 2011) and now clinical trials on human patients have been started (Cirak, Arechavala-Gomeza et al. 2011).

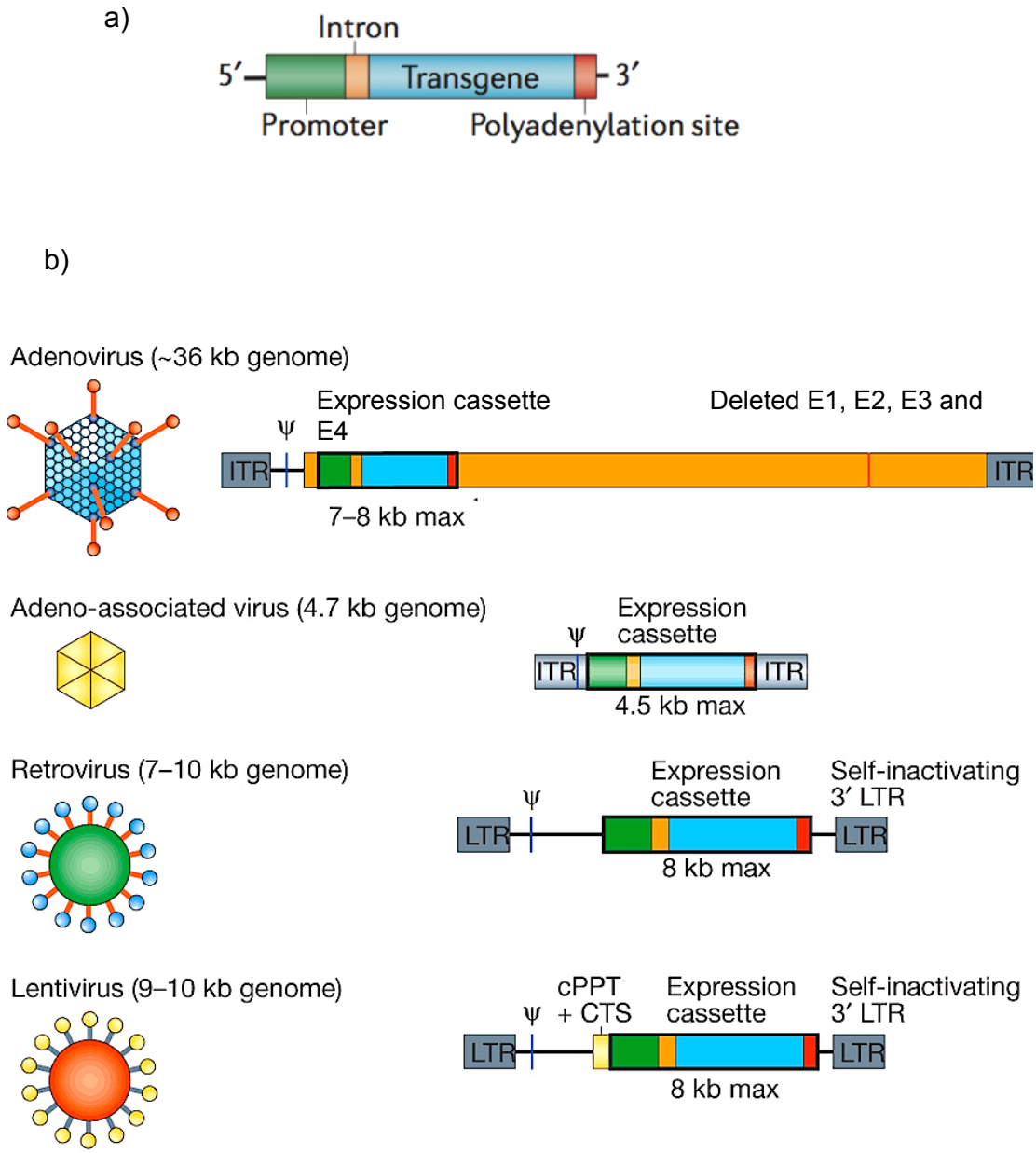


Figure 1.2 Viral vectors commonly used in gene therapy.
a) Standard expression cassette. From 5' end to 3' end: promoter in green, spacer sequence (intron) in orange and polyadenilation site for the termination of transcription in red. **b)** Types of viral vectors. The adenovirus has the dsDNA flanked on both sides by the ITR sequences. The packaging signal (Ψ) is immediately upstream of the E1, E2, E3 and E4 early genes that have been deleted to accommodate the expression cassette. The adeno-associated virus has the genomic ssDNA substituted by the expression cassette that is flanked by the two ITR sequences and anticipated by the packaging signal (Ψ). The retrovirus and lentivirus have the expression cassette in place of the RNA genome. Both viruses are made replication defective thanks to the removal in the 3' LTR of the promoter and enhancer regions. The figure is a modified picture from O'Connor et al. Nature Rev. 2006.

1.6 Gene therapy for β-thalassaemia

Therapies for β-thalassaemic patients are limited to blood transfusion, which is only palliative and has to be administered for the lifetime of the patient or to bone marrow transplantation, which is mainly limited to the 25-30% of patients with HLA-matched donors as well as a 5-10% risk of fatal complications (Sadelain, Boulad et al. 2007). Gene therapy, therefore, is a promising alternative cure for β-thalassaemia. Due to the basic principle of the technique that involves the introduction of a functional β-globin gene into haematopoietic stem cells of the patient and their auto-transplantation, gene therapy avoids the major drawbacks of both blood transfusion and allogeneic transplantation. In theory, if the therapeutic gene is correctly delivered and stably expressed by the target cells, the treatment should be necessary only once in the life of the patient without the need for subsequent administrations (unlike blood transfusion) and such autologous transplantation should completely avoid the risk of graft rejection or the need for immune suppression treatments that are characteristic of allogeneic bone marrow transplantation.

The first study showing an efficient gene therapy approach for β-thalassaemia was carried out on transgenic mice in 2002 by May et al. (May, Rivella et al. 2002). Previous experiments had shown that transduction with a viral vector containing the β-globin gene with its own proximal promoter and enhancers sequences was not sufficient to induce expression of the gene at therapeutic levels (Bender, Gelinas et al. 1989; May, Rivella et al. 2002). On the other hand, the incorporation of distal regulatory sequences, normally located 25 kb upstream of the endogenous β-globin gene, was found to be essential for the high-level and long-term expression of the β-globin gene at therapeutic levels (May, Rivella et al. 2002). Thus portions of the Locus Control Region (LCR, see Section 1.2), were introduced into a viral vector along with the β-globin gene to induce long-term gene therapy in mice (May, Rivella et al. 2002).

The recombinant lentivirus TNS9 was used to transduce and efficiently integrate in β-thalassaemic heterozygous mice ($Hbb^{th3/+}$) a construct made of the human β-globin gene, specific enhancers and promoter and the H2, H3 and H4 sequences of the LCR region which are essential for prolonged and

stable expression of the β-globin protein (May, Rivella et al. 2000) (see Fig. 1.3 a). The $Hbb^{th3/+}$ mice are the most accurate model of the disease and display a β-thalassaemia intermedia trait with chronic anaemia and abnormal red cell size and shape. Fifteen weeks after being transplantated with TNS9-transfected bone marrow cells, however, these mice showed an improvement in the anaemia and aberrant red cell morphology. The expression of TNS9 human β-globin in thalassaemic mice was of therapeutic relevance with 19% to 22% of total haemoglobin containing human β-globin 40 weeks after transplantation and stable levels of the β-globin protein detectable for up to one year (May, Rivella et al. 2000; May, Rivella et al. 2002).

The success of these preliminary studies in mice led in 2005 to the development of a phase I-II clinical trial of β-globin gene therapy for β-thalassaemia in humans (Bank, Dorazio et al. 2005). The study was started in Paris and it is still underway; it uses lentiviral vectors in order to deliver a β-globin gene, regulated by the previously described LCR region, in harvested CD34+ hematopoietic stem/progenitor cells from patients with β-thalassaemia or sickle cell disease (Bank, Dorazio et al. 2005). The so-called LentiGlobin vector (a self-inactivating lentivirus, see Fig. 1.3 b) was chosen due to its superior qualities compared to previously tested oncoretroviral vectors. The lentivirus has several advantages. These include an increased viral titre, the ability to target non-dividing cells and the capacity to carry larger recombinant constructs (an important factor in gene therapies for globin pathologies).

LentiGlobin drives the production of a modified human β-globin chain that displays a single point mutation in codon 87 ($\beta^{A\text{-}T87Q}$). The mutation, which is characteristic of the wild type form of the foetal γ-globin gene, was introduced for two main reasons: first of all, it has been shown to have a reduced interaction with sickle-cell haemoglobin (making it an anti-sickling protein) and therefore is useful in the treatment of sickle cell patients (Pawliuk, Westerman et al. 2001). Secondly, this modified protein has the same oxygen transport activity and haemoglobin assembly efficiency as the normal β-globin but after transplant it can be easily discriminated from the wild type β-globin by simple electrophoresis analysis. This last aspect has been fundamental for the assessment of the efficacy of the treatment (Bank, Dorazio et al. 2005). In order to reduce the risk of insertional oncogenesis (Hacein-Bey-Abina,

Garrigue et al. 2008), the LTRs of the LentiGlobin vector have been inactivated by mutation and the only promoter and enhancers are those in the recombinant β-globin gene. In addition, the chicken cHS4 chromatin insulator () (Evans-Galea, Wielgosz et al. 2007) has been introduced at the end of the 3' LTR sequence and, after reverse transcription, should flank both extremities of the vector further increasing the safety of the construct (Bank, Dorazio et al. 2005).

To date 2 patients have been treated with the LentiGlobin treatment. The first patient suffered severe complications and had to be transplanted with back-up autologous CD34+ cells for haematopoietic rescue. The second patient, an adult suffering from a severe form of β-thalassaemia (β^E/β^0 trait) and dependent on blood transfusion since childhood, received the treatment in June 2007 and in mid-2008 became transfusion-independent. The measurement of haemoglobin levels in peripheral blood was stable and quantitative analysis of β-globin chains showed that one third was made of the globin variant produced by the vector (Persons 2009; Cavazzana-Calvo, Payen et al. 2010).

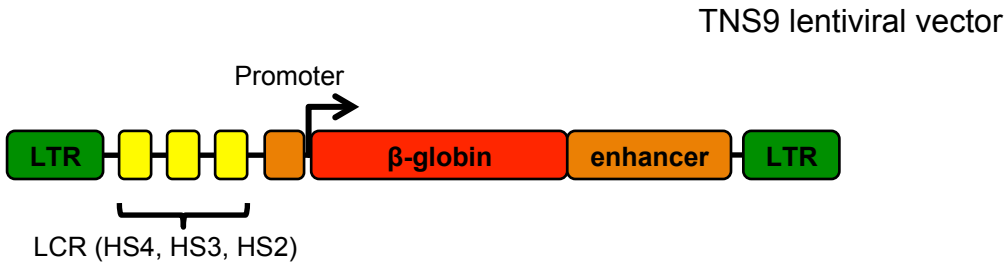
Although the results obtained in this trial were promising, in May 2009 the investigators reported to the NIH Office of Biological Activities (OBA) the presence of an excess in the transplanted cell population, from patient 2, of a dominant clone in which the vector had integrated into the HMGA2 gene (OBA ; Kaiser 2009; Persons 2009; Cavazzana-Calvo, Payen et al. 2010).

The HMGA genes (High Mobility Group, HMGA1 and HMGA2) are responsible for the expression of the so called AT-hooks proteins, a group of architectural transcription factors. The HMGA proteins regulate in both positive and negative manners the expression of many genes involved in cell growth, proliferation and differentiation by altering the chromatin conformation. This family of proteins is usually expressed during the development of the embryo while suppressed in adult life, however, HMGA2 misregulation in adults has been associated with a number of different tumours like breast cancer, lung cancer and sarcomas while rearrangements of the gene have been linked to the onset of benign tumours.

The over-expression of the gene in the Lentiglobin patient seems to be due to the loss, on one side of the LentiGlobin vector that has integrated close

to the HMGA2 gene, of one of the cHS4 insulators. However, because of the tissue-specific LCR regulator, the HMGA2 gene is only over-expressed in erythroid cells and the clone does not yet seem to have acquired a malignant status (Cavazzana-Calvo, Payen et al. 2010). The high number of mRNA copies in erythroblasts can also be ascribed to the lack of the 3' distal portion of the mRNA, which is known to contain binding sites for microRNAs that drive its degradation. The truncation of the mRNA seems to be due to the alternative splicing of the immature messenger mediated by the presence of the Lentiglobin (Cavazzana-Calvo, Payen et al. 2010). The reason for the dominant presence of this specific clone has not been fully explained but it is probably due to a stochastic effect due to the low number of transduced bone marrow cells or due to a proliferative advantage conferred by the HMGA2 gene (Cavazzana-Calvo, Payen et al. 2010).

a)



b)

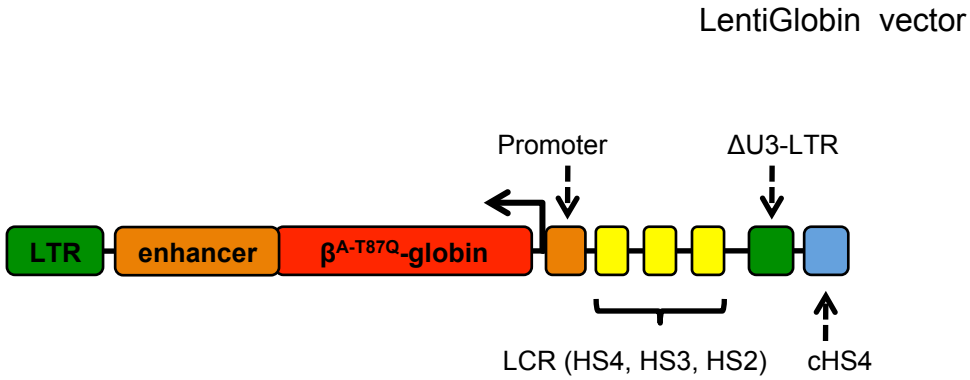


Figure 1.3 Lentiviral vectors for the gene therapy of β-thalassaemia.

a) TNS9 lentiviral vector used in a mouse model of β-thalassaemia. The vector carries a β-globin expression cassette flanked by specific promoter and enhancer. Part of the LCR region (HS2, 3 and 4 in yellow) is located upstream of the expression cassette. **b)** LentiGlobin vector used in the first clinical trial for the gene therapy of β-thalassaemia. The vector presents similarities to the TNS9 vector with the β-globin expression cassette flanked by specific promoter and enhancer and preceded by part of the LCR region (HS2, 3 and 4 in yellow). The β-globin carries a mutation ($\beta^{A\text{-}T87Q}$) for easy detection in the patient blood samples. In addition, the 3'LTR has the promoter and enhancer deleted in order to make the vector replication defective and the cHS4 (chromatin insulator, in blue) sequence has been introduced downstream of the LTR as an insulator in order to increase the safety of the vector.

1.7 Gene therapy by gene targeting/correction

An alternative approach to traditional gene therapy is gene correction, a specific application of gene targeting. Gene targeting comprises a series of techniques that allow the specific “editing” of a desired gene in host cells or organisms. Gene correction is based on the modification of the mutant gene rather than the addition of a therapeutic gene. A DNA repair template, made of the desired mutation to be introduced flanked by two DNA sequences homologous to the target, is introduced into the cell to replace the homologous chromosomal segment through Homologous Recombination (HR, see Fig. 1.4). Gene targeting presents many theoretical advantages compared to traditional gene therapy: 1) its therapeutic action is not affected by dominant mutations; 2) it is not influenced by limited vector capacity because the technique does not require the delivery of a complete and sometimes very large functional gene; 3) the challenge of recreating appropriate physiological level/patterns of gene expression is eliminated because gene correction, by definition, restores the disease gene to its normal state and 4) gene targeting presents a reduced risk of insertional mutagenesis and oncogenesis that have been major issues at the centre of the debate about gene therapy clinical trials (Vega 1991; Yanez and Porter 1998; Lanzov 1999).

Gene targeting achieves gene correction through a form of HR, catalysed by the cellular DNA repair machinery, in which a DNA-fragment homologous to the target sequence is used as a repair template (Lanzov 1999) (Fig. 1.4).

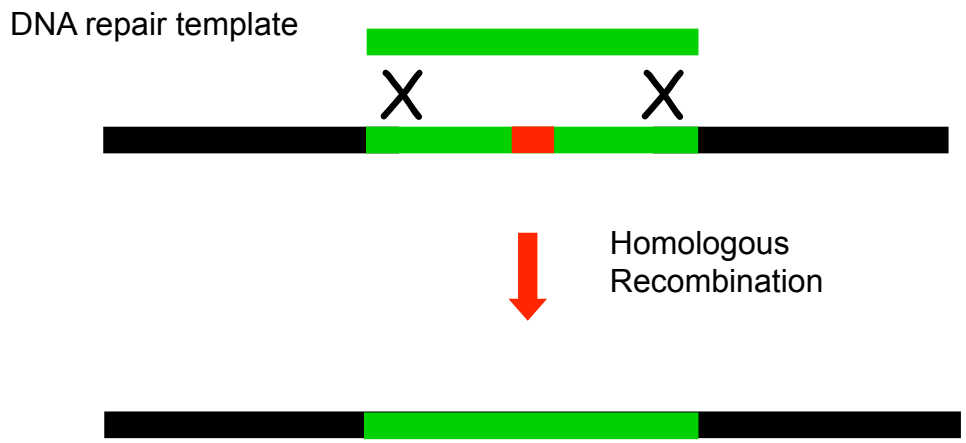


Figure 1.4 Gene correction by Homologous Recombination.
A DNA repair template (green fragment) homologous to the target sequence is used to correct possible mutation (in red) through HR (represented by the two crosses).

The basis for the development of gene targeting techniques have been initially laid in the late 1970s with *Saccharomyces cerevisiae* yeast models (Scherer and Davis 1979) where, compared to other eukaryotic models, gene targeting is easily obtained. Subsequently, targeted gene modifications in mammalian models have been achieved during the 1980s by Smithies et al. in cancer cell lines (Smithies, Gregg et al. 1985) and by the pioneering experiments performed by Capecchi et al. in murine cells (Thomas, Folger et al. 1986; Thomas and Capecchi 1987) leading to mouse gene knockout technology.

Initially, Smithies et al. were able to obtain gene targeting of the human β-globin gene in different cancer cell lines using a circular targeting construct based on a neomycin resistance cassette and a 4.6 kb β-globin homology region. The authors of the research found also that gene targeting frequencies were significantly increased if the homology region was cut by a restriction enzyme creating a linear targeting construct with a selectable marker flanked by two human β-globin homology arms (Smithies, Gregg et al. 1985).

Thanks to the use of positive selection induced by selectable markers, Capecchi et al. showed how a faulty neomycin resistance cassette randomly integrated in the genome of murine cells could be efficiently repaired by HR using as a repair template a second non-functional neoR cassette (carrying a different mutation) with frequencies of 1 cell in 10^3 acquiring neomycin resistance (Thomas, Folger et al. 1986). In a later stage the use of double selection based on the targeting of the HPRT gene (whose inactivation confers resistance to 6-thioguanine) by a functional neoR cassette (called gene addition, see Fig. 1.5) proved the feasibility of producing specific mutations at an endogenous locus in murine embryonic stem cells (Thomas and Capecchi 1987). These gene targeting experiments, coupled with the ability to produce embryonic stem cells from mouse blastocysts developed by Evans et al. (Evans and Kaufman 1981), paved the way to the creation of a robust and valuable method for the manipulation of the murine genome and the generation of “knockout” mice, one of the most useful models for the study of the functions of genes. Capecchi, Evans and Smithies were awarded the Nobel Prize for this work in 2007.

Although gene targeting has been widely used in mouse embryonic stem cells for study of specific cell functions (by selectively knocking out desired genes), the technique also had the potential to be used for the correction of many hereditary diseases in humans. This slow development is due to the inefficiency of gene targeting and progress is this area has depended on, and will continue to depend on, an understanding of the underlying mechanisms of HR.

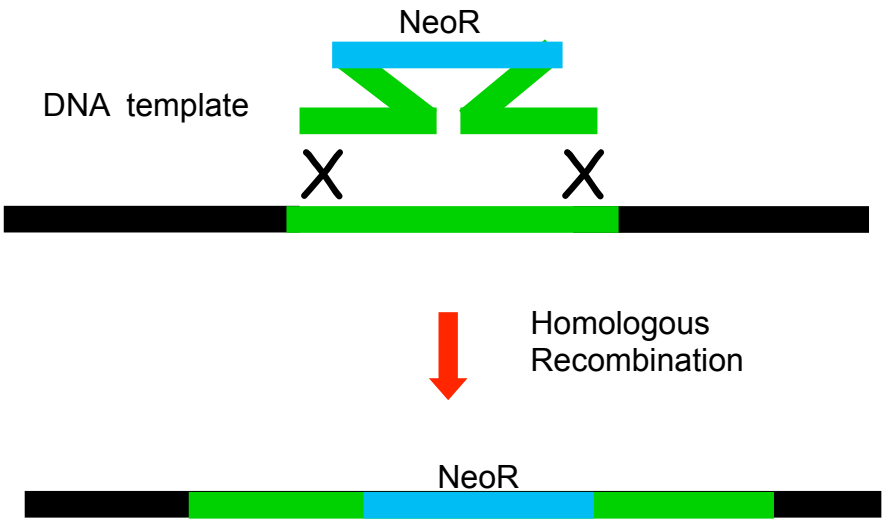


Figure 1.5 Gene addiction mediated by HR.
A further application of gene targeting is the introduction of an exogenous gene (in this case a Neomycin resistance cassette) in a specific target locus. A DNA template composed by the desired gene (in blue) flanked by two homologous regions (in green) is introduced in the cell. HR (crosses) occurring between the homologous regions allows the introduction of the construct in the desired locus.

1.8 Double Strand Breaks and Homologous Recombination

HR is an error-free mechanism for repairing chromosomal Double Strand Breaks (DSB) that usually occurs during or after DNA replication (i.e. during S phase and G2 phase of the cell cycle). HR is thus essential in order to maintain genomic integrity. In this process the broken chromosome is repaired by the cellular machinery using the "sister" chromatid as a template and the information stored is recovered without loss of functionality (Yanez and Porter 1998; Lanzov 1999). Gene targeting takes advantage of this HR mechanism; an exogenous DNA fragment, with the desired modification flanked by sequences homologous to the target locus, is delivered to cells where it is used as a template for HR mediated repair, instead of the sister chromatid, and so introduces the desired alteration into the target locus (Yanez and Porter 1998) (Fig. 1.4 and 1.5).

Such HR-based gene modification events, however, occur in human cells with a very low frequency (typically fewer than 1 cell in 10^5 transfected cells undergoes the desired genetic modification). These frequencies, though useful for research purposes (e.g. production of knockout and knockin transgenic mice using murine embryonic stem cells) are too low to be useful in the context of gene therapy (Porter 2005). Crucially, however, it has been shown that the introduction of a DSB within the target gene greatly enhances the frequency of gene targeting.

Due to the high risk of serious genome rearrangements, DSBs are probably the most dangerous type of DNA damage that can occur to a cell and, if not immediately repaired, they can easily trigger apoptosis: the last line of defence of an organism against oncogenesis derived by extensive genomic rearrangements (Kaina 2003). DSBs naturally occur by means of environmental factors, such as ionising irradiation, chemical agents and ultra violet light (UV) or by means of endogenous cellular products, such as reactive oxygen species. In addition, during DNA replication, any "nick" (single strand DNA damage) in the genome can be easily converted by the

replication fork in a DSB increasing the risk of genome instability (Kuzminov 2001; Helleday, Lo et al. 2007).

Two major repair pathways are used by cells to repair DSBs and maintain genome stability: Non Homologous End Joining (NHEJ) and Homologous Recombination (HR). NHEJ is the most direct way that the cell has to repair DSBs and is based, in its simplest form, on the direct ligation of the two severed DNA ends, if they are blunt or compatible ends. This simple process most of the time is associated with faithful repair of the DNA sequence without the introduction of mutations. However, most of the DSB events are also associated with damage at the severed DNA ends, such as aberrant phosphate groups, damaged backbone sugars and damaged DNA bases that make the two DNA ends incompatible which therefore have to be processed before DNA repair (Helleday, Lo et al. 2007). Repairs by NHEJ of these types of complex damage are quite often error-prone and can introduce mutations and deletions at the DSB site with the obvious risk of damaging important cell functions or inducing cancer. The NHEJ pathway involves many different proteins of which the most important are Ku70/80 that binds both DNA ends and recruits DNA-PK (DNA protein kinase) that interacts with Ku70/80 proteins and brings the two DNA ends close together and, finally, the ligase IV complex that binds the two DNA extremities and repairs the DSB (Helleday, Lo et al. 2007) (see Fig. 1.6).

Where possible, that is if a DNA template is available, DSBs will be accurately repaired by the error-free pathway HR. The preferred repair template used by this pathway is the sister chromatid and therefore this repair mechanism is limited to the S/G2 phases of the cell cycle where the repair template is readily available. The first step of the HR pathway is the resection in the 5’ to 3’ direction of the two DNA ends with formation of long 3’ stretches of single stranded DNA (that are stabilised and “protected” by the RPA protein). From this point, three main alternative modes are possible for the repairing of the DNA damage: the synthesis-dependent strand-annealing pathway (SDSA), the double-Holliday junction pathway and the single strand annealing pathway (SSA) (Helleday, Lo et al. 2007).

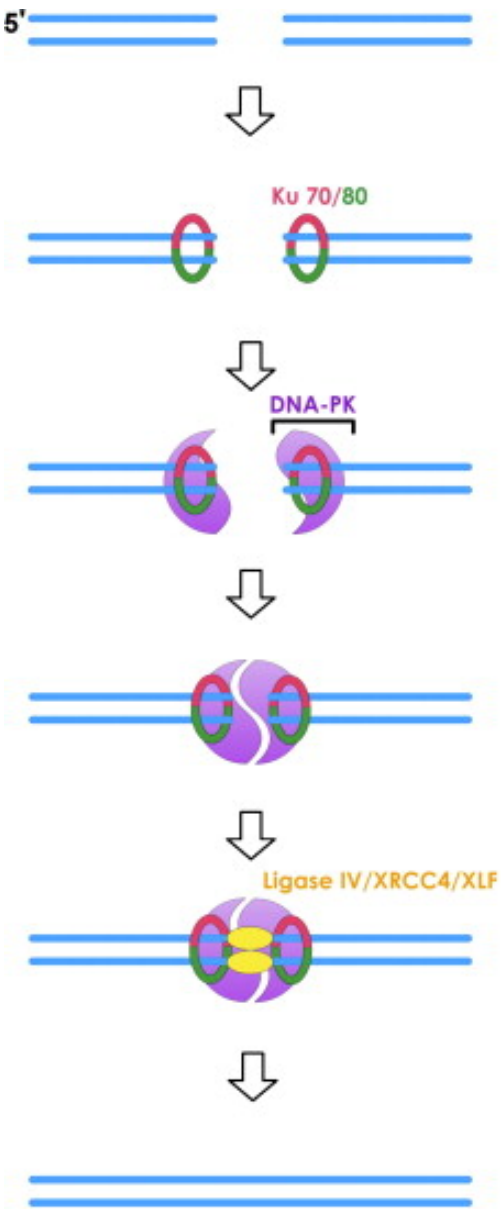


Figure 1.6 Model of the for Non-Homologous End Joining (NHEJ).
DNA ends created by a DSB are recognised and bound by the Ku70/80 proteins. Then the DNA-PKc proteins are recruited and interact with Ku70/80 in order to bring close together the two DNA ends. Finally, the ligase IV complex (comprised of ligase IV, XRCC4 and XLF) binds the two extremities and repair the DSB. Picture obtained from Helleday, Lo et al. 2007.

Synthesis-dependent strand-annealing pathway (Fig. 1.7): SDSA is probably the most common way the cell has to repair DSBs and involves the single strand DNA invading the sister chromatid (mediated by the RAD51 protein) in order to form heteroduplex DNA. After this initial step DNA synthesis elongates the single strand DNA (carried out by DNA polymerase η) on the other side of the original DSB using the sister chromatid as a template while the single Holliday junction that has been formed at the other side of the elongation point slides in the same direction of the replication (branch migration) in order to release the single strand DNA. The newly created single strand DNA at this point can anneal back to its original single strand DNA counterpart (probably facilitated by the RAD52 protein). Finally, the excess of DNA produced by the replication is removed by endonucleases while the missing “gap” left on the other single strand originated in the initial processing of the DSB is filled by DNA replication recreating the error-free double stranded DNA (see Fig. 1.7).

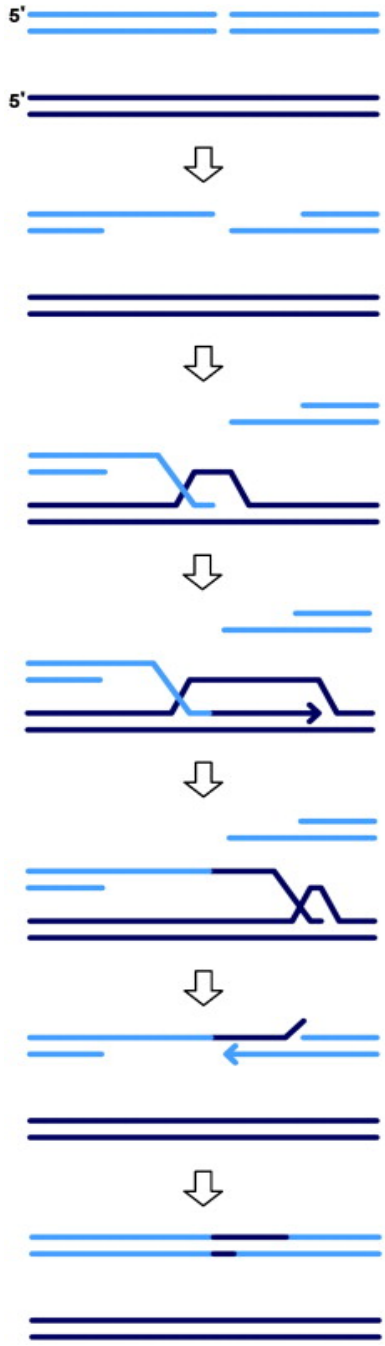


Figure 1.7 Model of the Synthesis-Dependent Strand-Annealing pathway (SDSA).

The pathway involves the ssDNA invading the sister chromatid in order to form an heteroduplex DNA. After this initial step DNA synthesis is carried out using as a primer the invading ssDNA. Branch migration slides the single Holliday junction in the same direction of the replication in order to release the single strand DNA. The newly created ssDNA anneals to its original ssDNA counterpart. Finally, the excess of DNA produced by the replication is removed by endonucleases while the missing "gap" left on the other single strand originated in the initial processing of the DSB is filled by DNA replication recreating the error-free dsDNA. Picture obtained from Helleday, Lo et al. 2007.

Double-Holliday junction pathway (Fig. 1.8*)*: The DSB may also be repaired by the “classic” double-Holliday junction pathway. This mechanism is not so commonly used to repair DNA damages but is probably more prevalent during meiosis. The pathway mainly follows the steps highlighted for the SDSA pathway but in this case both of the single DNA strands invade the template and DNA replication is carried out in both directions with the formation on other sides of two Holliday junctions. Depending on how the two Holliday junctions are resolved, this pathway can originate a simple gene conversion or a crossing-over event. Importantly, crossing-over events during DSBs repair are very rare and tend to occur more frequently during meiosis (Helleday, Lo et al. 2007) (see Fig. 1.8).

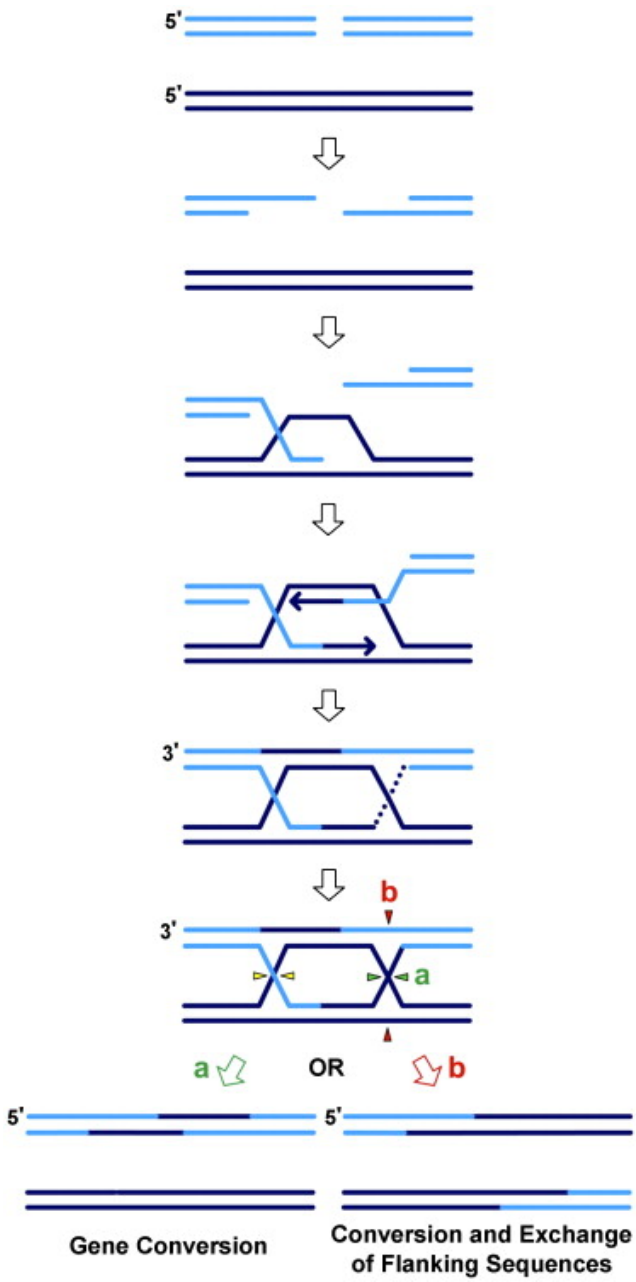


Figure 1.8 Model of the "classic" double-Holliday junction pathway.
The pathway involves the two single DNA strands invading the sister chromatid in order to form two heteroduplex DNA regions. DNA synthesis is carried out on both strands with the formation, on other sides, of two Holliday junctions. Depending on how the two Holliday junctions are resolved, this pathway can originate a simple gene conversion or a crossing-over event. Picture obtained from Helleday, Lo et al. 2007.

Single strand annealing pathway (Fig. 1.9): while the pathways described above are characterised by faithful repair of DNA damage without loss of genetic information (conservative HR). In contrast, the third pathway, SSA, is associated with the loss of part of the DNA sequence (non-conservative HR). SSA uses two homology regions found on both sides of the DSB to repair the DNA damage by a mechanism that involves RAD52 and RPA but not RAD51. The two DNA ends are resected beyond the two regions of homology and the two newly formed 3' overhangs are simply annealed thanks to the homologous sequences. At this point the excess of DNA is cleaved by endonucleases on both sides and gaps are filled repairing the damage. Unfortunately, with this pathway all the information stored in the sequence between the two homology regions is lost (see Fig. 1.9).

Extensive studies on the repair mechanisms of DSBs in mammalian cells have been carried out at the beginning of the 1990s by Maria Jasin's group (Memorial Sloan-Kettering Cancer Center). The group focused particularly on HR pathways and over the years has developed a set of repair assays for the easy detection of HR events. In one experiment the rare cutting endonuclease I-SceI was used in mammalian cells to stimulate HR by the means of a DSB (Jasin 1996). This endonuclease has a very long recognition sequence of 18 nucleotides that enables it to produce a DSB in specific loci without the risk of producing damage in the rest of the genome of the host. The stimulated gene correction event was used to restore the expression of a selectable marker gene (a neomycin cassette in the first experiments and subsequently a GFP expression plasmid) in order to identify possible gene correction events. Results showed how gene targeting can be stimulated 50 to 1000 fold by a DSB and how there is an inverse correlation between the frequency of gene correction and the distance between the cutting site and the sequence modification (Jasin 1996; Elliott, Richardson et al. 1998).

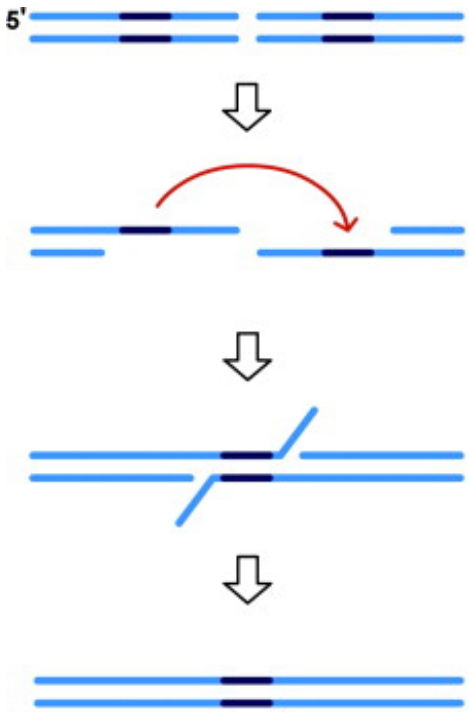


Figure 1.9 Model of the Single Strand Annealing (SSA) pathway.
The SSA pathway uses two homology regions (in dark color) found on both sides of the DSB to repair the DNA damage. The two DNA ends are resected beyond the two regions of homology and the two newly formed 3' overhangs are simply annealed thanks to the homologous sequences. At this point the excess of DNA is cleaved by endonucleases on both sides and gaps are filled repairing the damage. Picture obtained from Helleday, Lo et al. 2007.

The ability to produce DSBs to stimulate HR, coupled with the introduction in the target host of a targeting construct to be used as a repair template for the correction of the induced DNA damage, has defined the current concept of targeted gene alteration. Depending on the type of targeting construct used, gene targeting may be used to inactivate a specific gene for genomic studies (gene knockout), to produce gene correction of a disease-causing mutation or to introduce, at specific loci, functional exogenous genes (gene addition) to exert a novel function in the host cell (Urnov, Rebar et al. 2010).

Although different types of targeting constructs (or donor DNA) have been developed over the years, most studies have focused mainly on two types: double stranded DNA and single stranded DNA constructs. Single stranded oligonucletides (ssONs) of ≈ 100 bp complementary to one of the two strands of a possible target gene have been successfully used to produce targeted gene modifications. Although the mechanism by which the ssONs induce the gene editing is not clear, it seems that DNA replication is involved in the process (Jensen, Dalsgaard et al. 2011). McLachlan et al. showed in a recent study that the ssON-mediated repair of a faulty HPRT gene in mammalian cells was increased in frequency during the S phase of the cell cycle and that the frequency of gene editing was drastically reduced in cells deficient in HR but not in cells deficient in other DNA repair pathways (McLachlan, Fernandez et al. 2009).

The double stranded donor DNA, instead, usually carries the desired point mutation or gene to be introduced flanked by two extended “homology arms” (each one of them usually between 500 bp to 2 kb in length) that mediate the homology directed recombination between the targeting construct and the endogenous locus. If the construct contains only a point mutation between the homology arms, donor DNA can be used to create or correct mutations. On the other hand, the targeting construct could carry a complete functional gene for the introduction of a novel gene function or could carry a therapeutic, promotorless cDNA fragment that could be introduced close to the endogenous promoter of the gene to be corrected. This last type of donor DNA has the advantage of being able to correct all possible mutations of the target gene and at the same time to be regulated by the endogenous

promoter. A cDNA targeting construct has been used successfully in a recent study for the correction of human haemophilia in a mouse model. As a result of DSB-induced gene replacement, a disrupted human factor IX gene located in the genome of liver cells of mice was corrected, in vivo, by the introduction at the end of exon 1 of a splicer acceptor immediately followed by a partial factor IX cDNA. Levels of human factor IX were detected in the circulation after the treatment and remained constant for at least thirty weeks (Li, Haurigot et al. 2011).

Due to the low frequencies of gene targeting in the absence of an induced DSB at the target locus, most of the targeting constructs used for research purposes also carry between the homology arms a selectable marker allowing for the drug-mediated selection of cells that have incorporated the donor DNA (Yanez and Porter 1998). A more sophisticated approach makes use of positive-negative selection where in addition to the positive-selection marker a negative selection marker is located outside the homology arms. In this way the researcher is able to select not only the cells that incorporated the targeting construct, but also to select only the cells that introduced the donor DNA by means of HR (which do not incorporate the negative selectable marker) by excluding any possible random integration events (that do include the negative selectable marker). From a therapeutic point of view, due to the small number of cells that are usually targeted, the possibility of using a selection marker should be ideal for the enrichment of targeted stem cells for transplantation to correct genetic disorders although it is possible that the drug resistance gene might interfere with the normal homeostasis of the cell or could trigger gene silencing of the neighbouring genes. A possible alternative is based on the removal of the selection marker immediately after positive selection of targeted events. This approach has been used by Sebastiano et al. to correct the sickle-cell mutation in iPS (induced Pluripotent Stem) cells. The targeting construct used had a 1.6 kb β-globin sequence with a puromycin cassette flanked by two loxP sites located in the first intron. Using the Cre recombinase, Sebastiano et al. were able to remove the drug resistance cassette from corrected iPS cells leaving behind only a copy of the loxP site as a residual “scar” sequence (Sebastiano, Maeder et al. 2011).

For this observation to be useful for therapeutic gene targeting, however, a method had to be found for the development of customized endonucleases capable of targeting and cleaving any specific sequence in the genome in order to achieve, through stimulation via DSBs, acceptable frequencies of gene targeting. In order to be of any value these endonucleases have not only to be able to target a wide range of possible sequences but also need to recognise a specific target without producing DSB in other regions of the genome (Porter 2005). The engineering of customised endoncleases may address all of these requirements and establish a versatile and precise tool for future gene therapy.

Customised endonucleases: Since the initial studies based on the I-SceI endonuclease, researchers have developed several different types of engineered nucleases based on naturally occurring proteins. Initially, researchers focused on the well-characterised meganucleases (of which I-SceI is an example) by attempting to change their nucleotide binding specificity in order to increase the number of possible target sites. The process of engineering novel binding specificities for this class of enzymes has been particularly difficult for two main reasons. Firstly, the complexity of the tertiary structure of the proteins prevented the possibility of predicting the amino-acid sequence that would confer novel binding specificities. Secondly, the strong connection in meganucleases between the binding and cleavage domains meant that any modification of one of the two would greatly affect the efficiency of the other (Paques and Duchateau 2007). However, due to intensive structural characterisation of the meganuclease family, the modification of critical amino-acid residues and the development of highly efficient screening methods, the commercialisation of engineered homing endonuclease technology has now become feasible (Grizot, Epinat et al. 2010; Arnould, Delenda et al. 2011). Two other classes of tailored nucleases have proved to be more flexible for therapeutic gene targeting: Zinc Finger Nucleases (ZNFs; described in detail in the following sections) and Transcription Activator-Like Effector Nucleases (TALENs).

TALEN are the most recent class of efficient gene-editing molecules to have been developed and are very promising due to their highly modular

nature. The TAL effectors are DNA binding proteins found in *Xanthomonas* (a bacterial plant pathogen) and modulate gene expression in the host organism. These effectors are composed by modules of ≈ 33-35 conserved amino-acids and each module binds specifically to only one base pair due to only two variable residues (repeat variable di-residues RVD) in position 12 and 13 (Cermak, Doyle et al. 2011). In contrast to the difficulties encountered during the development of novel ZFNs and meganucleases, TAL effectors, due to their ability to bind single nucleotides, have proved to be easily assembled in customised DNA binding proteins capable of recognising long DNA sequences and their fusion to the *FokI* nuclease domain has created a new family of nucleases that have already been used to cleave DNA sites with high specificity both in vitro and in vivo (Li, Huang et al. 2011; Miller, Tan et al. 2011). The scientific community has taken great interest in this new technology and both commercial (Cellectis) and public (Cermak, Doyle et al. 2011) approaches have been developed for the creation of customised TALENs.

1.9 Zinc finger nucleases

The most versatile molecular tools developed to date for targeted editing of DNA sequences in different organisms are ZFNs, which are fusion proteins composed of two different subunits: a zinc finger binding domain and a non-specific nuclease *FokI* domain. The initial idea of using the nuclease domain for the production of targeted DSBs was conceived by Chandrasegaran's group in 1992 when they realized that the type II restriction enzyme *FokI* was composed by two distinct units: one capable of specifically binding the DNA sequence and one that was catalysing the cleavage (Li, Wu et al. 1992). By altering the DNA binding domain it was possible to produce DSBs in different sites and therefore researchers focused on the search for optimal binding domains. The most effective was found to be a zinc coordinated finger-like structure called zinc finger (Kim, Cha et al. 1996).

Zinc fingers (Cys_2His_2) are repeat units found in a wide range of naturally occurring proteins (mostly in the DNA-binding domains of transcription

factors) usually made of 30 amino-acids (arranged in an α-helix and two β-sheets configuration) that are able to bind a wide range of trinucleotides. Structural stability is provided by a zinc ion coordinated in a tetrahedral pattern by two cysteine residues on the β-sheets and by two histidine residues on the α-helix (Wolfe, Nekludova et al. 2000). The DNA-binding feature is due to the insertion of the α-helix in the major groove of the double helix. A combination of key amino-acids in positions -1, 2, 3 and 6 on the α-helix are responsible for the interaction (through hydrogen bonds or hydrophobic interactions) with three consecutive nucleotides of the DNA binding site (Choo and Klug 1997; Wolfe, Nekludova et al. 2000) (Fig. 1.10).

One of the first steps towards the use of engineered zinc finger proteins was obtained by Klug et al. in 1994. The researchers created a three-finger protein domain that specifically recognised a 9 bp sequence in the *bcr-abl* oncogenes and showed that the novel protein was able to efficiently bind its target sequence in vitro. Subsequently, mouse cells carrying a copy of the *bcr-abl* gene were used to prove the ability of the ZF protein to bind, in vivo, the target sequence and suppress gene expression of the *bcr-abl* gene (Choo, Sanchez-Garcia et al. 1994).

Natural zinc finger-based proteins are able to recognise different DNA sequences of different lengths through the combination of different zinc fingers in a tandem configuration (Wolfe, Nekludova et al. 2000). This “modular” configuration has inspired researchers to develop libraries of zinc fingers specific for all 64 possible trinucleotides and to use these to design new molecular tools capable of targeting any desired target site. The fusion of a specific binding domain with the non-specific cleavage domain of the endonuclease *FokI* gives rise to a brand new category of molecular tools able to produce DSBs in any locus of interest, thereby stimulating the HR pathway for possible gene correction at any desired disease-related gene (Kim, Cha et al. 1996; Porter 2005).

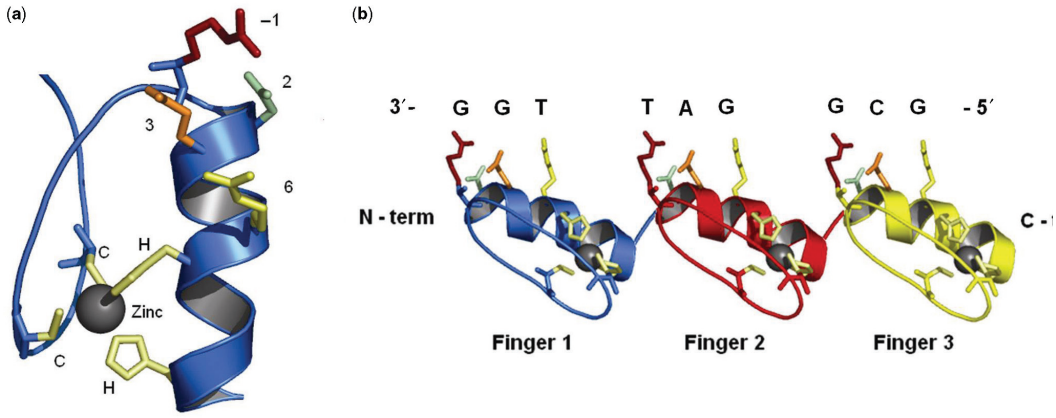


Figure 1.10 Zinc finger molecular structure.
a) Structure of the ZF molecule with two β-sheets and α-helix coordinated by a zinc ion and key residues (-1, 2, 3 and 6) for the binding of the trinucleotides sequence. **b)** ZF domain composed of three different ZFs and the 9 nucleotides target sequence. Picture from Sander et al. 2007.

Typically a ZFN is made up of three or four zinc fingers arranged in a row and then linked to the *FokI* domain. The DNA binding domain of the ZF guides the non-specific nuclease activity of *FokI* to a specific DNA target site. The *FokI* domain is active only in a dimer configuration, therefore two monomers use their ZF domains to recognize half-sites on either side of a “spacer” sequence (usually 6 nucleotides) where the dimerised *FokI* domains produce the DSB (Cathomen and Joung 2008) (see Fig. 1.11).

A first study involving the expression of ZFNs in eukaryotes was carried out in 2002. A vector carrying the gene for zinc finger nucleases driven by a heat-shock promoter (Hsp70) was introduced into the genome of *Drosophila Melanogaster* (Bibikova, Golic et al. 2002). The ZFNs were able to recognize a target sequence on the *y* gene of Drosophila and to produce a DSB. Genetic analysis showed the presence of somatic mutations (deletions or small insertions) at the level of the target gene. These mutations are the expected outcome for repaired DSBs through NHEJ the mechanism by which the DNA is repaired in absence of a homologous DNA repair template (Bibikova, Golic et al. 2002). In a subsequent study from the same group a complete gene targeting procedure was performed in Drosophila. In this case *y* and *ry* genes were targeted by ZFNs and the DSBs produced were used to stimulate HR events with a marked donor DNA fragment as repair template. Of the 25% mutant progeny, produced after the mating of transfected parents, most of the offspring presented the gene correction expected by HR (Beumer, Bhattacharyya et al. 2006).

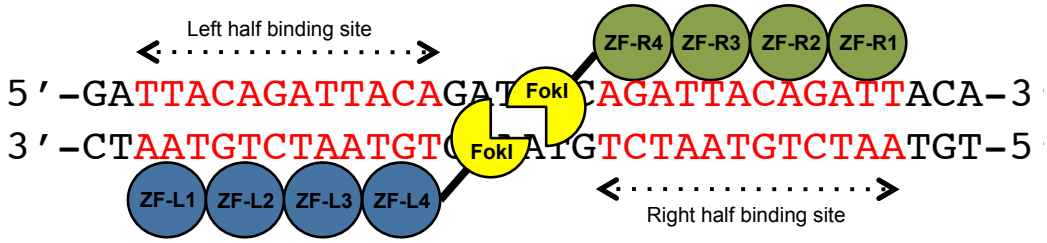


Figure 1.11 ZFN binding to target sequence and formation of FokI dimer.
ZFs are numbered from 1 to 4 starting from the N-terminus, the *FokI* monomer is linked to the C-terminus of the ZF domain that binds to the target sequence in a "3' to 5'" orientation.

Another remarkable experiment was carried out in human cells for the correction of a mutation on the IL2Rγ gene that causes SCID-XI. In the early stage of the study, K562 cells were transfected with expression vectors for the two ZFN sub-units and a DNA repair template homologous to IL2Rγ. The expression of ZFNs stimulated gene targeting events at the endogenous locus with 11% of the transfected cells carrying one targeted allele and 6% of them both targeted alleles without selection (Urnov, Miller et al. 2005). In order to test the efficacy of targeted gene modification in primary cells, the experiment was then carried out using $CD4^+$ T cells and 5% of these cells were successfully targeted. The final outcome showed how the permanent modification of the endogenous gene through gene targeting was effective in up to 20% of the chromatids and that 7% of the treated cells were homozygous for the exogenous sequence in absence of selection (Urnov, Miller et al. 2005).

The ZFNs technique was later tested in combination with an Integrase-Defective Lentiviral Vector (IDLV) in order to increase the efficacy of delivery into target cells. IDLV viral particles were loaded with two types of vector, one expressing ZFNs and the other containing the donor sequence and the efficiency of gene targeting was tested in different cell lines (Lombardo, Genovese et al. 2007). Correction of the IL2Rγ gene was achieved with an efficiency between 13% and 39%, depending on the type of cell line, while targeted integration of a GFP construct in the chemokine receptor 5 (CCR5) locus produced up to 50% of GFP positive cells in K562 and Jurkat cell lines (integration of the GFP gene in the target locus was confirmed by PCR analysis). These results led to similar experiments in human stem cells.

The same GFP construct specific for the CCR5 gene was delivered with ZFNs expression vectors into $CD34^+$ haematopoietic progenitor cells and in human Embryonic Stem cells (ES) HUES-3 and HUES-1 with 0.11% and 5.3% of them expressing GFP, respectively. Furthermore the induced differentiation of HUES cells into neural progenitor cells was not affected by the newly introduced construct and the expression of GFP was confirmed by confocal fluorescence microscopy (Lombardo, Genovese et al. 2007).

ZFN-mediated gene targeting was also used in a study where, for the first time, both human iPS and ES cells were targeted (Zou, Maeder et al. 2009). ZFN expression vectors and DNA repair templates were introduced into both cell types through nucleofection and ZFN-mediated HR was used to correct a disrupted GFP reporter gene that had been randomly integrated in the chromosomes of both iPS and ES cells. GFP expression was enhanced 2400-fold in ES cells transfected with both DNA repair template and ZFN expression vectors compared to cells transfected with only the DNA repair template. In iPS cells, GFP gene correction was increased 1400-fold when ZFNs expression vectors were co-transfected with the repair template (Zou, Maeder et al. 2009). In order to assess ZFN-stimulated gene correction at an endogenous locus, both ES and iPS cells were transfected with ZFNs specific for the endogenous *PIG-A* gene and a DNA repair template having a hygromycin cassette flanked by two *PIG-A* homologous regions. The use of ZFNs increased the frequency of HR-mediated gene correction 200-fold relative to cells transfected with the DNA repair template alone (Zou, Maeder et al. 2009).

A case study carried out using probably one of the most effective ZFNs so far described highlights the challenges of applying ZFN technology to different types of cells and organisms. Li et al. used a very efficient ZFN (F9-ZFN) specific for the human factor IX gene in K562 (erythroleukaemia), Hep3B (human hepatocyte) cells and a humanised haemophilia mouse model (Li, Haurigot et al. 2011). The cutting efficiency of the F9-ZFN was initially tested in K562 cells by measurement of the frequency of “indels” (small deletions or insertions that are the products of DSBs repaired by NHEJ) produced at the endogenous gene. The F9-ZFN was found to be highly effective in producing DSBs with 45% of the treated cells showing the presence of indels. Frequencies of ZFN-induced HR events at the target locus were measured by detection of a stably integrated novel NheI restriction site (carried by the NheI-targeting construct) and 18% of the alleles were found to be targeted in both K562 and Hep3B cells. For the in vivo study, Li et al. created a humanised mouse model carrying a defective factor IX gene. Initially, in order to test the F9-ZFN activity in mouse liver, the mice were injected through the tail vein with an AAV vector (AAV8) expressing, under a liver-specific promoter, the

F9-ZFN. Cutting frequencies, measured as the percentage of detected indels, were found to be as high as 47%, indicating that the combination of AAV8 vector and the liver specific promoter results in very efficient ZFN induced gene modifications. To test the efficiency of gene targeting, Li et al. designed an AAV8 vector carrying donor DNA based on a partial cDNA of the human factor IX (the targeting construct already referred to in Section 1.8) and the two AAV8 vectors (F9-ZFN and donor DNA) were introduced by intraperitoneal injection into neonatal mice. The different injection technique was used due to the higher survival rate in young mice (compared to tail vein injections) but was also shown to induce lower frequencies of F9-ZFN-induced indels (16% against the very high 47% previously detected). Gene targeting frequencies, measured by PCR assay, were found to be around 1-3% (Li, Haurigot et al. 2011). The results obtained in this study show the feasibility of producing gene correction directly in vivo without the need for the isolation and ex vivo treatment of cells and subsequent auto-transplantation in the host but they also indicate a wide variability in cutting and gene targeting frequencies depending on the targeted host (cell lines or mice), delivery methods (transfection of cells or AAV8 vector) and, more importantly, methods of administration of the AAV8 vectors in mice (tail injection or intraperitoneal injection). The great variability in cutting and gene targeting frequencies observed in this study highlights the need for robust and highly efficient ZFNs in order to obtain acceptable therapeutic levels of gene targeting. Similar considerations can be applied also to the results in cell lines used for research purposes (like K562 or HEK293 cells) compared to more complex primary cell types like, for example, stem cells or iPS cells.

iPS cells have been used in two recently published studies on gene targeting of the endogenous β-globin gene for the development of gene therapy for sickle-cell anaemia (Sebastiano, Maeder et al. 2011) and β-thalassaemia (Zou, Mali et al. 2011). The two studies showed effective gene targeting of the β-globin gene in iPS cells but at very low frequencies. A more detailed description of the results of these two papers is reported in Section 7.3.2). Factors that may affect the targeting efficiencies are the transcriptional status of the target locus, and therefore the accessibility of the gene to ZFNs; the levels of HR that depend on both the cell type and replication status of the

cells used; and the possible silencing effect that can suppress the expression of novel exogenous genes or corrected genes (Cathomen and Joung 2008; Hockemeyer, Soldner et al. 2009; Carroll 2011).

DNA hypermethylation-mediated silencing of targeted genes that had undergone HR events is probably one of the greatest obstacles that will have to be addressed in order to develop efficient gene targeting techniques for therapeutic applications. This phenomenon was examined by Cuozzo et al. in a 2007 study where they used a Homology Directed Repair (HDR) reporter assay developed by the Jasin group (the reporter assay is described at length in Chapter 6) in Hela cells (Cuozzo, Porcellini et al. 2007). The reporter assay is carried out by transfecting cells that harbour a GFP expression plasmid disrupted by the I-SceI target site with an I-SceI expression plasmid. Cells will only express functional GFP proteins if the I-SceI-cleaved GFP plasmid is repaired by HR using a repair template located downstream of the GFP gene. Cuozzo et al. showed how the GFP reporter plasmids subjected to I-SceI-induced HR were subsequently silenced by methylation and the levels of GFP expression tended to drop during the days after transfection. This effect was reversed if the cells carrying the reporter plasmid were treated with 5-aza-2-deoxycytidine for 48h after transfection with the I-SceI expression plasmid. The silencing effect was not observed in Hela cells transfected with a wild type GFP expression vector and therefore this effect could only be related to GFP genes that were targeted by HR events (Cuozzo, Porcellini et al. 2007).

1.10 Engineering ZFNs

Since the discovery of the potential of ZF proteins as components of molecular tools for gene correction, gene knockout and gene addition, different techniques/reagents have been developed in order facilitate the synthesis of functional ZFNs with high specificity for novel target sequences.

Sangamo BioSciences was one of the first to create a method to produce highly efficient ZF proteins capable of targeting endogenous genes in different types of cells (Urnov, Miller et al. 2005; Miller, Holmes et al. 2007; Moehle, Rock et al. 2007; Liu, Chan et al. 2009; Cavazzana-Calvo, Payen et al. 2010). Although the precise details of the technique that Sangamo uses to produce

effective ZFNs is not fully disclosed, due to intellectual property protection, it seems that a two-step process is involved. In the first step, four-finger domains are created by assembling a pair of two-finger units, chosen from a pre-characterised archive, and tested for their specificity to the target sequence. In the second step, the most efficient domains have their binding activity increased by optimisations made following the guidelines of an algorithm-based approach (Cathomen and Joung 2008). Sangamo's ZFN engineering platform is extensively described in Chapter 4 of this thesis. Sigma Aldrich, which has acquired the licence for Sangamo's proprietary ZFN technology, has made a variety of gene-specific ZFN by this method and they can be purchased "off the peg". Alternatively, customised ZFNs can be ordered from Sigma Aldritch (which is at the moment the only provider), but these are extremely expensive (£20,000) and for this reason not fully accessible to the scientific community.

On the other hand, a group of academic laboratories has founded the Zinc Finger Consortium in order to promote the development of ZF engineering techniques that are reliable, user-friendly and freely available to the academic community. The "Modular Assembly" approach was the first technique developed by the Consortium and the protocol was published in 2006 while all reagents, including an archive of more than 140 plasmids, each encoding a different ZF 'module' and designed for convenient ligation at the DNA level. These reagents are available from the non-profit plasmid repository Addgene (Wright, Thibodeau-Beganny et al. 2006). As the name suggests, the process involves the linkage of different ZF modules, each one of them specific for a trinucleotide, in order to create a "ZF array" (a multi-finger domain, usually made of three or four ZFs) with the desired binding affinity for the target sequence. A web-based software package helps the identification of possible target loci in a given sequence that are suitable for the targeting through arrays created from the available ZF-module archives. A Bacterial Two Hybrid assay (B2H) enables a rapid and efficient test for the binding activity of the produced ZF array. In the end, successful ZF arrays are cloned into a vector carrying the FokI domain that allows the expression of complete ZFNs in mammalian or plant cells (Wright, Thibodeau-Beganny et al. 2006). This approach was successfully used in a series of experiments that targeted

genes *in vitro* (Segal, Beerli et al. 2003), in *Drosophila melanogaster* (Beumer, Bhattacharyya et al. 2006) and in *Caenorhabditis elegans* (Morton, Davis et al. 2006). A more recent large-scale evaluation of the Modular Assembly technique, however, showed a higher failure rate than expected: a set of ZF arrays designed for 104 different target sites was tested and the overall efficiency, measured as the binding efficiency of the ZF arrays in the B2H assay, was calculated to be less than 25% (Ramirez, Foley et al. 2008). This low rate of success for the Modular Assembly approach is probably due to a "context dependent" effect: neighbouring ZF units are not always independent from one another in their binding activity because of an overlapping effect in which one zinc finger "invades" the neighbouring trinucleotide influencing the activity of the following ZF (Cathomen and Joung 2008; Ramirez, Foley et al. 2008). It seems that some ZFs have, in addition to the standard binding of the three-nucleotide, a propensity to make contact also with the base, on the parallel strand, complementary to the base immediately 5' of its trinucleotide binding site (Isalan, Choo et al. 1997). Depending on the adjacent base, this 'cross-strand' interaction can impair the affinity of the ZF for its trinucleotide binding site making trinucleotide recognition sequence context-dependent (Klug 2010).

In order to overcome these problems the Zinc Finger Consortium developed a new strategy in late 2008: The Oligomerized Pool ENgineering (OPEN) system (Maeder, Thibodeau-Beganny et al. 2008). This approach has the advantage of taking notice of the context dependent effect through a "context-sensitive selection strategy"; an archive of ZF pools, each one of them containing ZF variants specific for a trinucleotide, is used to assemble a small library of different ZF arrays all specific for the target sequence. A B2H selection system, in which the binding of ZFs activate a selectable marker, is then used to search the newly produced library for ZF arrays with increasing binding activity (Maeder, Thibodeau-Beganny et al. 2008). OPEN was used in a first study to produce 37 efficient ZFNs that were able to modify 11 different target sites in three endogenous human genes (VEGF-A, HoxB13, CFTR) and one endogenous tobacco plant gene (SuRA); up to 50% of the transfected cells presented the specific gene modification. When ZF arrays produced with the Modular Assembly approach and with the OPEN system were compared,

only 18% of ZF arrays produced with the first method were effective in human cells while the second method gave a 75% efficiency (Maeder, Thibodeau-Beganny et al. 2008).

The OPEN system protocol was published in late 2009 and it is supported by a web-based software that, as in the previous protocol, helps to identify possible target sequences. Although this protocol yields a higher percentage of functional ZFNs, it requires high levels of expertise with phage-display technology and with the screening of combinatorial libraries and therefore it is not easily employed by research groups with limited experience or equipment (Maeder, Thibodeau-Beganny et al. 2009).

In 2011 researchers from the Zinc Finger Consortium published a third protocol that should be easily applied by any laboratory but, at the same time, should also take in account possible context-dependent effects between neighbouring ZFs and therefore it should provide a more robust approach than the modular assembly protocol (Sander, Dahlborg et al. 2011). The Context Dependent Assembly (CoDA) protocol does not consider ZFs as single, independent modules but instead is based on ZF domains each one made of three different ZFs: ZF1-ZF2-ZF3 with ZF1 and ZF3 being at the N and C termini, respectively. The archive provided by the consortium contains 319 different ZF1 and 344 different ZF3 that work efficiently when coupled on both sides of 18 fixed ZF2 units. New domains with binding affinity for different DNA sequences can be created by linking together different ZF1 and ZF3 units interspaced by a common ZF2 unit (see Fig. 1.12) (Sander, Dahlborg et al. 2011). Although this protocol has still to be fully assessed and it is limited to the creation of functional ZFNs based on subunits made only of 3 ZFs each, it is a simple alternative to more complex (OPEN) or expensive (Sangamo-Sigma) approaches and at the same time promises to be more effective than the original modular assembly protocol.

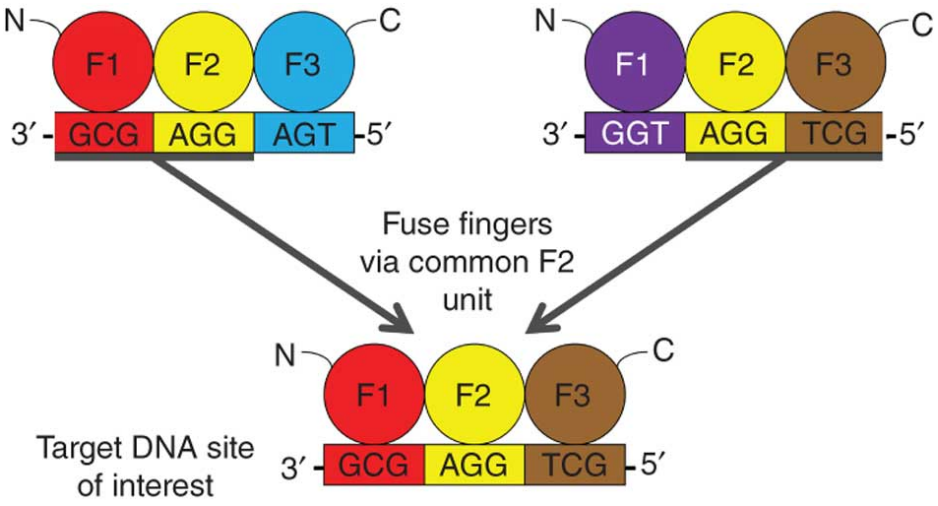


Figure 1.12 CoDA approach for the creation of novel 3 fingers ZF domains.
Two different 3 fingers domains are shown on top. Each one binds a specific 9-bp target site but both of them have in common a middle F2 unit. The two domains can be rearranged in order to create a novel ZF domain with new specificity where F1 and F3 are derived from the two different domains but can be joined together thanks to the common middle F2 unit. Picture from Sander et al. 2011.

1.11 Aims

The aim of this project was to develop ZFNs capable of promoting correction of specific mutations in the β-globin gene in order to develop a reliable gene-correction therapy for patients with β-thalassaemia or sickle cell anaemia.

The first step was to identify suitable target sequences in the β-globin gene to be targeted by customized ZFNs and to design and synthesise ZF domains specific for the selected target sites through a modular assembly approach. The ZF arrays would then be assessed for sequence-specific binding activity in a bacterial two hybrid assays (B2H) in order to identify the most promising ZFN to be used in subsequent experiments.

The next step was to take possible ZF arrays that showed a low binding efficiency and improve them through one of the two directed evolution systems developed throughout the project. These systems would be based on the creation of libraries of mutated ZF arrays with different binding activities; members of these libraries would then be selected for efficient binding activities through newly designed bacterial selection methods based on a zeocin selection cassette or on a double selectable system to take advantage of a streptomycin resistance cassette and a histidine pathway gene.

Alternatively, a ZFN specific for the β-globin gene would be acquired from Sigma Aldrich in order to increase the chances of obtaining efficient ZFNs to be used in the following gene targeting experiments and possibly to be used in future therapeutic applications.

All ZFNs created during the development of this project would be tested in three different assays:

1. ZFN cutting efficiencies would be assessed through the Cel-I assay in two different cell lines (K562 and HT1080) to detect “indels”, small mutations or deletions derived from the correction of possible DSBs through the non-conservative NHEJ repair pathway.
2. The ability of the ZFNs to promote HR would be tested using a Homology Directed Repair (HDR) assay based on a GFP reporter plasmid to compare the HR levels induced by ZFNs with the “gold standard” I-SceI endonuclease.

3. Finally, successful ZFNs would be used to produce gene targeting at the β-globin endogenous locus in K562 cells using as a repair template a puromycin cassette flanked by two homology arms for easy selection of successful targeted events.

2 Chapter 2: Materials and Methods

2.1 Preparation of nucleic acids

2.1.1 Bacterial plasmid isolation

Plasmid DNA extraction and purification from liquid cultures was carried out through alkaline lysis, high-salt precipitation and DNA isolation using Zyppy Plasmid Miniprep Kit (ZYMO RESEARCH CORP.) for up to 10-15 μg of plasmid DNA and E.Z.N.A. Plasmid Maxi Kit (OMEGA bio-tek) for up to 100 μg of plasmid DNA. Isolated DNA was re-suspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0).

2.1.2 Ethanol precipitation and phenol extraction of DNA samples

Ethanol precipitation and phenol/chloroform extraction techniques were used for the purification of DNA samples.

Ethanol precipitation: 1/10 (vol/vol) of 3 M sodium acetate and 3/1 (vol/vol) of cold 100% ethanol (Fisher Scientific) were added to the DNA solution. The sample was centrifuged at 15000 g for 30 min at 4^0C, the supernatant was removed and the pellet was washed with 70% ethanol, centrifuged at 15000 g then dried and re-suspended in suitable amounts of TE buffer.

Phenol extraction: 1/1 (vol/vol) of phenol solution (25:24:1 phenol:chloroform:iso-amyl alcohol, by volume, Sigma) was added to the DNA solution, vortexed for 1min and then centrifuged at 13000 g for 5 min in order to separate the inorganic/organic phases. Inorganic phase containing the DNA was then collected and ethanol precipitated.

2.1.3 Mammalian genomic DNA isolation

"*Simplified mammalian DNA isolation procedure*": DNA from adherent cells HT1080 was isolated following the Liard et al. protocol (Laird, Zijderveld et al. 1991). In brief: a confluent 10 cm diameter petri dish (~1-2x10^6 cells) was washed with PBS and 10 ml of lysis buffer (100 mM Tris.HCl pH 8.5, 5 mM EDTA, 0.2% SDS, 200 mM NaCl, 100 µg Proteinase K/ml) was added to the dish and incubated at 37°C on a rocking table for 5 hours. 10 ml of isopropanol (Fisher Scientific) were then added to the lysate and the plate was agitated until white fibres of genomic DNA became visible. The DNA fibres were then recovered with a disposable pipette tip and the excess of fluid was removed from the fibres by dabbing them on the inside of the petri dish lid. Genomic DNA was then dissolved in a tube containing 100 µl TE buffer.

"*Commercial genomic DNA purification kit*": Genomic DNA from ~1x10^6 K562 cells was isolated using Wizard Genomic DNA Purification Kit (Promega) following the manufacturers instructions. The kit is based on a four-step protocol: lysis of cells and nuclei, RNase treatment, salt-precipitation of proteins and isopropanol precipitation of genomic DNA. Genomic DNA was then resuspended in 100 µl of TE buffer. Assuming ~100% recovery of genomic DNA, 1 million cells should generate ~ 6 µg of genomic DNA.

2.1.4 DNA purification from agarose gel or from solution

DNA bands resolved on an 1 or 2% agarose gel were isolated using a sharp blade and the gel fragment containing the DNA band was melted in 600 µl of ADB buffer (ZYMO RESEARCH CORP.) at 37°C. The DNA sample was isolated using Zyppy Plasmid Gel-extraction Kit (ZYMO RESEARCH CORP.). In alternative to ethanol precipitation, in order to exchange solution or to concentrate DNA samples were purified adding 2 volumes of DNA Binding Solution (ZYMO RESEARCH CORP.) to the samples and then using the Zyppy Plasmid Gel-extraction Kit

2.2 Handling of bacteria (*Escherichia coli*)

2.2.1 Liquid and solid media

Standard Luria Bertani (LB) liquid medium or LB agar plates (9 cm petri dishes, Sterilin) were used for the growth of bacterial strains throughout the project unless differently specified (J. Sambrook 1982). All liquid cultures were grown with aeration at 37°C in a shaker-incubator and bacterial growth was monitored by measuring the Optical Density (OD) at 600 nm using a spectrophotometer (BioPhotometer, Eppendorf). LB agar plates or liquid LB were used in conjunction with different antibiotics for the selection of bacterial transformants (see Section 2.2.4). Depending on the transformed plasmid, the following antibiotics and relative concentrations were used:

- 100 μg/ml ampicillin (Sigma)
- 12.5 μg/ml chloramphenicol (Sigma)
- 30 μg/ml kanamycin (Sigma)
- 12.5 μg/ml tetracycline (Sigma)

Low salt LB liquid media with pH corrected to 7.5 with 1 M NaOH and low salt LB agar plates were used in conjunction with 25-50 μg/ml Zeocin (Invitrogen) during the creation and testing of the Zeocin selectable system.

M9 minimal medium agar plates were used for the selection of CSH100 bacterial strain. A 500 ml solution was prepared as follows: 439 ml of H_2O with 7.5 g of Bacto-Agar (Becton-Dickinson) were autoclaved; when the solution was cooled at 65^0C, 50 ml of 10x M9 salts (1l of solution contain: 70 g of Na_2HPO_4-$7H_2O$ (BDH), 30 g of KH_2PO_4 (BDH), 5 g of NaCl (BDH) and 10 g of NH_4Cl (BDH)), 1 ml of 1 M $MgSO_4$ (BDH), 10 ml 20% (wt/vol) glucose (Sigma) and 0.5 ml of 100 mM CaCl (BDH) were added. The solution was then poured into plates.

Once set, all agar plates were dried by removing their lids in a laminar flow hood for 30 min, and then stored at 4°C.

LB agar plates with 12.5 μg/ml tetracycline, 30 μg/ml kanamycin and 5% (wt/vol) sucrose were used for the selection of conjugation products between KJ1C and CSH100 bacterial strains during the OPEN selectable system.

2.2.2 Bacterial strains

The following *Escherichia coli* (*E.coli*) strains were used throughout the development of the project:

- KJBAC1 (*F- lacI^q ΔhisB463 Δ(gpt-proAB-arg-lac)XIII zaj::Tn10*) provided by Addgene was used in the B2H assay and in the Zeocin selectable system.
- CSH100 (*F' lacproA+proB+(lacIq lacPL8)/ara ∆(gpt-lac)5*) provided by Addgene was used in the OPEN selectable system.
- KJ1C (*F- ∆hisB463 ∆(gpt-proAB-arg-lac)XIII zaj::Tn10*) provided by Addgene was used in the OPEN selectable system.
- Transformax EPI300 (*F$^-$ mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ$^-$ rpsL (Str^R) nupG trfA dhfr*) provided by Epicentre Biotechnologies was used for the amplification of the low copy plasmid pBAC-LacZ.
- NEB 10-β (*araD139 ∆(ara,leu)7697 fhuA lacX74 galK16 galE15 mcrA f80d(lacZ∆M15)recA1 relA1 endA1 nupG rpsL rph spoT1∆(mrr-hsdRMS-mcrBC))* provided by *NEB was used in all standard cloning procedures.*
- *JM109 (endA1, recA1, gyrA96,thi, hsdR17(r_k^-, m_k^+), relA1, supE44, Δ(lac-proAB), (F' traD36, proAB, laqIqZΔM15)) provided by NEB*.
- *NEB 5-α (F' I^q (F´ proA$^+$B$^+$ lacI^q ∆(lacZ)M15 zzf::Tn10 (Tet^R) / fhuA2∆(argF-lacZ)U169 phoA glnV44 Φ80Δ(lacZ)M15 gyrA96 recA1 endA1 thi-1 hsdR17))* provided by *NEB was used for the amplification of pGP-FF-ZF plasmids.*

2.2.3 Preparation of competent cells for chemical transformation

The protocol used for creating bacterial competent cells was derived from Wright et al. (Wright, Thibodeau-Beganny et al. 2006). 1ml of saturated bacterial inoculum from an overnight 4 ml culture was added to 50 ml of LB media containing 15 mM $MgCl_2$ (BDH) and incubated with agitation for 1.5 hours at 37˚C. The cells were then pelleted by centrifuging at 5000 g for 20 min at 4˚C. The pellet was re-suspended in 3 ml of ice-cold, filter sterile Solution A (10 mM $MnCl_2$ (BDH), 50 mM $CaCl_2$ (BDH), 10 mM MES (2-(N-morpholino)ethanesulfonic acid) pH 6.3 (Sigma)) containing 15% glycerol and it was left in ice for 20 min. The solution was then aliquoted into sterile tubes, frozen in dry ice and then stored at -80˚C.

2.2.4 Chemical transformation

The protocol used for the chemical transformation of bacteria cells was obtained from Sambrook et al. (J. Sambrook 1982). 50-100 ng of DNA plasmids were added to 50 µl ice-cold competent cells and left on ice for 30 min. Samples were heat-shocked for 30-45 sec in water-bath at 42˚C and then returned to ice for 2 min. The cells were then recovered with 500 µl LB at 37˚C for 1.5 hours and then plated on LB plates containing a suitable antibiotic. Single colonies were collected after overnight incubation at 37°C and were expanded overnight in 4 ml of liquid LB with a suitable antibiotic at 37°C with agitation. Around 3 ml of culture were used for the isolation of plasmids (see Section 2.1.1) while 0.5-1 ml of the same culture were used for long-term storage. Bacterial transformants were stored in 20% glycerol (v/v) (Sigma) at -80°C.

2.3 Gel electrophoresis

2.3.1 Agarose gel electrophoresis of DNA (DNA-AGE)

Casting of 1% and 2% agarose (w/vol) gel was carried out using Agarose Molecular Grade (Bioline) or agarose "Iberose" (WebScientific) melted in 1x TAE with addiction of 1x Syber Safe DNA stain (Invitrogen). DNA samples were prepared by mixing the DNA solutions with 6x loading buffer (10 ml of a 6x stock solution: 3 ml glycerol, 25 mg bromophenol blue, 7 ml H_2O) and then loading them on agarose gel immersed in 1x TAE running buffer in a horizontal electrophoresis unit Horizon 58 (Life Technologies) with an applied voltage of 50-100 Volts. Samples dimensions were calculated using as reference 1 kb DNA ladder (New England Biolabs or Invitrogen) or 100 bp DNA ladder (NEB). DNA bands were visualized by exposing agarose gels with UV light on a transilluminator (BioDoc-It imaging system, UVP).

2.3.2 Polyacrylamide electrophoresis of DNA (DNA-PAGE)

Polyacrylamide gel electrophoresis was used for higher resolution of small DNA fragments. 12 ml of 8% polyacrylamide gel solution was obtained by mixing 3.2 ml of 29:1 acrylamide:bis-acrylamide solution (Sigma), 8.6 ml of H_2O, 1.2 ml of 10x TBE, 200 µl of 10% (wt/vol) ammonium persulfate (Sigma) and 10 µl of TEMED (Sigma). The gel was cast on a vertical electrophoresis unit (Biorad) and then immersed in 1x TBE running buffer. DNA samples with 6x loading buffer were loaded along 100 bp DNA ladder and a voltage of 50 Volts was applied for 3 hours. The obtained gel was then stained in a solution of 1xTBE plus 1x ethydium bromide (Sigma) and DNA bands were visualized by exposing gels with UV light on a transilluminator (BioDoc-It imaging system, UVP).

2.3.3 ^{32}P-labelling of 1kb DNA ladder

1µg of 1 kb ladder (NEB) was mixed with dNTPs (dATP, dGTP and dTTP 50 µM final concentration (NEB)), 0.5 µl α-^{32}P-dCTP (10 mCi/ml Perkin Elmer), buffer 2 (NEB) and 5 U Klenow fragment (3'-5' exo$^-$) in a 20 µl reaction volume and incubated at 37°C for 30 min. Labelled ladder was separated from the reaction mix using a IllustraTM ProbeQuantTM G50 micro columns (GE Healthcare) and resuspended in 1x loading buffer.

2.3.4 Probe creation and labelling

The probe to be used during the Southern blot analysis (see Chapter 6) was created by PCR amplification using Phusion Hot Start II high-fidelity DNA polymerase from Finnzymes (Thermo Fisher Scientific) and K562 genomic DNA as a template.

The primers used were:

Primer name	Sequence (5'-3')
HBB-P-F	GATTTGAAACTGAGGCTCTGACC
HBB-P-R	CAAGACCCTGTTTCACATCCCTG

And the PCR conditions were:

Initial denaturation	98˚C	5 min
Denaturation	98˚C	1 min
Annealing	57˚C	30 sec
Extension	72˚C	1min
Final extension	72˚C	10 min

Number of cycles: 30

100 ng of purified probe (see Section 2.1.4) and 250 ng of random hexamer primers (Promega) were boiled in 20 µl of TE for 2 min and then quickly placed in ice. Afterwards, dNTPs (dATP, dGTP, dTTP 50 µM final concentration (NEB)), 3 µl of α-^{32}P-dCTP (10 mCi/ml), 2 µl buffer 2 (NEB) and

5 U Klenow fragment (3'-5' exo$^-$) were added in a final volume of 20 µl solution. The reaction mix was incubated for 1h at 37°C.

The probe was separated from the reaction mix using the previously mentioned mini column and the percentage of radiolabelling was estimated measuring counts before and after purification. Around 50% of the α-^{32}P-dCTP was incorporated in the probe.

2.4 Molecular cloning and manipulation of DNA samples

2.4.1 Plasmid digestion with restriction enzymes

All restriction enzymes used in this project were obtained from New England Biolabs. Digestions were performed following protocols provided by the supplier. Unless differently stated, reactions were carried out in a 20 µl volume containing the appropriate 1x buffer, ≈ 1 µg of DNA sample to be digested and 2-3 units of restriction enzyme. Digestions were carried out usually for 1-2 hours.

2.4.2 Dephosphorylation of DNA 5'-ends

Removal of phosphate groups from the 5' ends of DNA samples was carried out using Antarctic Phosphatase (New England Biolabs). DNA plasmids were incubated at 37˚C with 5 U of Antarctic Phosphatase for 15 min and then the reaction was heat inactivated at 65˚C for 5 min.

2.4.3 Ligation

Standard ligations were performed with T4 DNA Ligase (NEB) following manufacturer's protocol. Usually, 10-20 ng of digested vector and an excess of DNA insert were ligated at room temperature for 30 min in a volume of 10 µl. Ligations of blunt products were carried out as follow: 10-20 ng of linear

vector was ligated with an excess of DNA insert in a reaction volume of 10 µl containing 1 µl (3 U) of T4 DNA ligase for 16 hours at 15˚C.

2.4.4 End-filling of 5' ends

End-filling of 5' overhangs was carried out using Klenow enzyme (NEB) and following manufacturer's instruction. Briefly, 20 µl reactions were carried out at room temperature for 15 min. Each reaction was supplied with 0.5 µg of DNA sample supplemented with 33 µM dNTPs and 1 unit of Klenow enzyme.

2.4.5 Sequencing

300-600ng of DNA samples containing 3.2 pmoles of primers were suspended in 10 µl of H_2O and sent to the Sequencing Service, MRC CSC Genomics Laboratory (Hammersmith Hospital, London). Sequencing is performed by an automated fluorescent DNA sequencing instrument ABI3730XL. Results were received as .seq files and analyzed with SerialCloner 1.3 software or Ape Plasmid Editor 2.0.37.

2.4.6 Polymerase Chain Reaction (PCR)

All PCR reactions were carried out using either a Multi Gene thermo-cycler TC9600-G (Labnet Internetional Inc.) or a PTC-200 DNA Engine thermo-cycler (MJ Research). Unless otherwise stated, standard PCR reactions were carried out using Taq DNA Polymerase with Standard buffer (NEB) or DreamTaq DNA Polymerase (Fermentas) and the following conditions:

- 5 U of Taq DNA Polymerase, 5 µl 10x buffer, 200 µM (final concentration) dNTPs, 0.4 µM (final concentration) primers, 39 µl H_2O, ≈10 ng of DNA sample.

Initial denaturation	95˚C	1/5 min
Denaturation	95˚C	30 sec
Annealing	50-60˚C*	30 sec
Extension	68˚C	1 min/kb
Final extension	68˚C	5 min

Number of cycles: 30

* Annealing temperatures varied depending on primers

- 2.5 U DreamTaq polymerase, 5 µl 10x DreamTaq buffer, 200 µM (final concentration) dNTPs, 0.4 µM (final concentration) primers, 39.5 µl H_2O, ≈10 ng of DNA sample.

Initial denaturation	95˚C	1/5 min
Denaturation	95˚C	30 sec
Annealing	50-60˚C*	30 sec
Extension	72˚C	1min/kb
Final extension	72˚C	5 min

Number of cycles: 30

* Annealing temperatures varied depending on primers

High-fidelity PCR was carried out using the Phusion Hot Start II high-fidelity DNA polymerase from Finnzymes (Thermo Fisher Scientific) using the following condition: 1 U Phusion polymerase, 10 µl 5x PCR buffer, 200 µM (final concentration) dNTPs, 0.4 µM (final concentration) primers, 32.5 µl of H_2O, ≈10 ng of DNA sample.

Initial denaturation	98˚C	1/5 min
Denaturation	98˚C	1 min
Annealing	50-60˚C*	30 sec
Extension	72˚C	1min/kb
Final extension	72˚C	5 min

Number of cycles: 30-40

* Annealing temperatures varied depending on primers

2.4.7 Colony PCR

$1x10^3$ cells from HT1080 or K562 cells were collected and mixed with 25 µl lysis solution (1.125 µl of 10% NP40, 1.125 µl of 10% Tween, 0.8 µl Proteinase K (Invitrogen), 2.5 µl 10x Dream Taq PCR buffer (Fermentas) and 19.45 µl of H_2O). The samples were then incubated at 50˚C for 1 hour and then at 94˚C for 10 min.

In a second step 25 µl of PCR solution (5 U of Dream Taq (Fermentas), 2.5 µl PCR buffer (Fermentas), 200 µM (final concentration) dNTPs, 0.4 µM (final concentration) primers, 16.5 µl of H_2O) was added to the samples. PCR reactions were carried out as follow:

Initial denaturation	95˚C	5 min
Denaturation	95˚C	1 min
Annealing	57˚C	30 sec
Extension	72˚C	1min
Final extension	72˚C	10 min

Number of cycles: 30

The primers used were the same described in Section 2.7.4.

2.5 Construction of novel plasmids

2.5.1 pBAC-ZFTS-LacZ reporter plasmids

Oligonucleotides used as target sites (TSs) in the B2H assay (see Section 2.7.1) were as follows:

- TS-1(L) target sequence: forward oligonucleotide 5'-tgtggaagatcttcgacagccctgggcaggtt-3' and reverse oligonucleotide 5'-taatgtaataacctgcccagggctgtcga-3'.
- TS-1(R) target sequence: forward oligonucleotide 5'-tgtggaagatcttcgacagcaccaacttcatt-3' and reverse oligonucleotide 5'-taatgtaataatgaagttggtgctgtcga-3'.
- TS-2(L) target sequence: forward oligonucleotide 5'-

tgtggaagatcttcgacatccaagctaggccc-3' and reverse oligonucleotide 5'-taatgtaatgggcctagcttggatgtcga-3'.

- TS-2(R) target sequence: forward oligonucleotide 5'-tgtggaagatcttcgacagaataatccagcct-3' and reverse oligonucleotide 5'-taatgtaataggctggattattctgtcga-3'.
- TS-3(L) target sequence: forward oligonucleotide 5'-tgtggaagatcttcgacacaatagaaactggg-3' and reverse oligonucleotide 5'-taatgtaatcccagtttctattgtgtcga-3'.
- TS-3(R) target sequence: forward oligonucleotide 5'-tgtggaagatcttcgacagaataatccagcct-3' and reverse oligonucleotide 5'-taatgtaatagacaggtttaagctgtcga-3'.

For each TS, equal amounts (50 fmol) of forward and reverse oligonucleotide were annealed in 100 µl of annealing buffer (40 mM Tris pH 8, 20 mM $MgCl_2$, 50 mM NaCl) by heating the solution at 95°C for 2 minutes and then slowly (around 1 hour) cooling it to 25°C. Each pBAC-ZFTS-LacZ plasmid was made by linearization of the vector pBAC-LacZ (1µg) with 0.5 µl of BsaI (10 U/µl) at 50°C for 1 hour. Annealed oligonucleotides were ligated to the digested pBAC-LacZ in order to form the pBAC-ZFTS-LacZ reporter plasmids (ligations were carried out following the protocol of Section 2.4.3).

2.5.2 pDR-GFP-ZFN4 reporter plasmid

The GFP reporter plasmid for the testing of ZFN4-induced HR was created by modification of pDR-GFP-HR plasmid previously developed by Maria Jasin's group (Pierce, Johnson et al. 1999). In order to introduce the specific target sequence for ZFN4, two oligonucleotides were designed:

Oligos name	Sequence (5'-3')
pDR-TS-4-Fwd	CAGGGTAATATAGGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAATAA
pDR-TS-4-Rev	TCATCCACGTTCACCTTGCCCCACAGGGCAGTAACGGCAGACCTATATTACCCTGTTAT

The two oligos were annealed in presence of 1x Annealing buffer (40 mM Tris pH 8, 20 mM $MgCl_2$, 50 mM NaCl) by incubation in a thermocycler:

Denaturation 95˚C	2 min
95˚C ramping to 85˚C	-2˚C/sec
85˚C ramping to 25˚C	-0.1˚C/sec
4˚C	hold

1 µg of pDR-GFP-HR was linearised by enzymatic digestion with 5 units of I-SceI (NEB) in presence of 1x I-SceI buffer. The reaction was carried out at 37˚C for 1h and then the enzyme was inactivated at 65˚C for 20 min. ≈ 0.2 µg of the linearised plasmid were subsequently ligated with an excess of annealed oligos following the procedure described in Section 2.4.3. The oligos were designed in a way that, when ligated in the pDR-GFP-HR plasmid, would reconstitute the I-SceI site on the left of TS-4. Successful introduction of TS-4 was confirmed by sequencing with the primer:

Primer name	Sequence (5'-3')
TS-4-pc	CTCTGCTAACCATGTTCATGC

2.5.3 pBL-TV-TC, pTV-TC4 targeting constructs and pBL-TV-pc positive control

The left and right homology arms were created by PCR amplification using a high fidelity DNA polymerase with K562 genomic DNA as a template (see Section 2.4.6).

Primer name	Sequence (5'-3')
HAL-F	CATTGTGCGGCCGCATATCAGGGATGTGAAACAGG
HAL-R	CATTGTGAATTCTCCTCAGGAGTCAGATGCAC
HAR-NF	TCCTGAGGAGAAGTCTGCCGTTAC
HAR-R	CATTGTGGTACCTGACTGGGAGAGAGGACAAGGAC
HAR-F	CATTGTATCGATCCTGGGCAGGTTGGTATCAAG
PC-F	CATTGTGAGCTCGGTGACAATTTCTGCCAATCAGGAC TAGTATATCAGGGATGTGAAACAGGGTC
PC-R	CATTGTTCTAGACTGCATTCTAGTTGTGGTTTGTCC

HAL-F and HAL-R were used to amplify the 2 kb left homology arm and at the same time to introduce a NotI and EcoRI restriction sites (underlined in the above table) upstream and downstream of the sequence, respectively.

The right homology arm was created through a nest PCR approach. Initially a 2489 bp fragment was amplified using the primers HAR-NF and HAR-R. This PCR product was then used as a template to amplify the final 2420 bp right homology arm using the primers HAR-F and HAR-R. The primers also introduced the two restriction sites ClaI and KpnI upstream and downstream of the PCR product (restriction sites are underlined in the above table).

Cloning of the left and right homology arms in the pBL-Puro plasmid using the aforementioned restriction sites created the initial targeting construct pBL-TV-TC where the Puromycin resistance cassette is flanked by the two homology arms.

The second and final targeting construct pTV-TC4 (a shortened version of pBL-TV-TC) was created by restriction digestion with ApaI and KpnI of pBL-TV-TC plasmid, isolation of the fragment carrying the $Puro^R$ cassette flanked by the two homology arms and cloning it in the pBSKS+ plasmid. pTV-TC4 has a left homology arm of 1.3 kb and a right homology arm of 2.5 kb.

The PC-F and PC-R primers were used during the creation of the positive control pBL-TV-pc plasmid to amplify the left homology arm of pBL-TV-TC and at the same time to introduce a 23 bp sequence to the end of the left

hand homology arm homologous to the PT1-F primer and including a SpeI site immediately upstream of the left homology arm. The two primers also introduced the unique SacI and XbaI restriction sites (underlined in the above table). Enzymatic digestion of pBL-TV-TC with SacI and XbaI was used to remove the left homology arm and to replace it with the newly amplified "extended" left homology arm creating the pBL-TV-pc positive control plasmid.

In order to create cell populations to be used as a control during the PCR screening of puromycin resistant cells, pBL-TV-pc was linearised with PvuI and then stably transfected in both HT1080 and K562 cells (see Section 2.6.6).

2.5.4 Zinc Finger DNA Binding Domains (ZF-DBDs) created by Modular assembly approach or commissioned to Mr Gene Company

Four of the ZF-DBDs (ZF1-R, ZF1-L, ZF2-R and ZF3-R) described in Chapter 3 were created following the protocol described by Wright et al. (Wright, Thibodeau-Beganny et al. 2006) while the remaining two (ZF2-L and ZF3-L) were synthesised by Mr Gene Company (now a subsidiary of GeneArt, http://order.geneart.com/).

The ZF plasmids (pc3XB) archive, containing 141 different zinc finger modules, used for the cloning of ZF1-R, ZF1-L, ZF2-R and ZF3-R was obtained from Addgene (http://www.addgene.org/). The plasmid used were:

- ZF1-R: pc3XB60, pc3XB71, pc3XB106 and pc3XB94.
- ZF1-L: pc3XB92, pc3XB103, pc3XB89 and pc3XB90.
- ZF2-R: pc3XB73, pc3XB104, pc3XB83 and pc3XB93.
- ZF3-R: pc3XB106, pc3XB75, pc3XB82 and pc3XB77.

The sequences of the ZF2-L and ZF3-L were assembled *in silico* using data from the web archive provided by Addgene and were sent to Mr Gene. The company synthesised the desired sequences, linking synthetic oligonucleotides and amplifying them by PCR, and the final product, flanked

by two restriction sites compatible with the modular assembly protocol, was then cloned into a standard plasmid and shipped.

2.6 Mammalian cell culture

2.6.1 Culturing

All mammalian cells were cultured at 37˚C and 5% CO_2 in a humidified incubator NAPCO 5410 (Precision Scientific). Handling was carried out in a laminar flow hood (Faster BHA36, Jencons-PLS) using sterile techniques. Three different human cell lines were used throughout the development of this project: HT1080 adherent cells (fibrosarcoma cell line), K562 suspension cells (erythroleukemia cell line) and DG75 suspension cells (Burkitt lymphoma cell line).

2.6.2 Adherent cells

HT1080 adherent cells were cultured in 10 cm plates (Corning) with DMEM medium from Gibco supplemented with 10% fetal calf serum FCS (heat inactivated, Gibco), 10 mM sodium pyruvate (Gibco), 20 mM L-glutamine, 400 I U/ml Penicillin-Streptomycin (Invitrogen) and 10 ml of a 100x MEM-NEAA solution (Gibco). Cells were passaged at 80% confluency washing first with PBSA (137 mM NaCl, 10 mM $Na_2HPO_4 \cdot 2\ H_2O$, 2.7 mM KCl, and a pH of 7.4) and then using 1 ml of 0.5% 10x trypsine-EDTA (Gibco) in order to detach the cells. Afterwards, cells were diluted 1:8 in fresh DMEM medium with supplements and plated.

2.6.3 Suspension cells

K562 and DG75 suspension cells were cultured in 75 cm^2 or 25 cm^2 flasks (Corning) using RPMI 1640 medium (Gibco) supplemented with 10% fetal calf serum FCS (heat inactivated, Gibco), 10 mM sodium pyruvate (Gibco), 20 mM

L-glutamine, 400 I U/ml Penicillin-Streptomycin (Invitrogen) and 10 ml of a 100x MEM-NEAA solution (Gibco). Cells were subcultured by removing part of the culture and adding fresh RPMI medium with supplements.

2.6.4 Long term storage of mammalian cells

Long-term storage of ≈10^6 cells was carried out resuspending cells in 0.5 ml of Freezing media (40% [v/v] standard DMEM medium with supplements, 50% [v/v] FBS and 10% [v/v] DMSO) and transferring them to cryotubes (Cryo Tube, Nunc). Vials were slowly frozen at -80°C by means of a polystyrene holder and after 24-48h were transferred to a liquid nitrogen container for indefinite storage.

2.6.5 Cell counting and vital staining assay

Counting of HT1080, K562 and DG75 cells was performed by adding ~ 10 µl of medium containing cells to a haemocytometer (Superior, Marienfeld) and the number of cells was visually established under an inverted phase contrast microscope Olympus CK2. K562 and DG75 cell viability was assessed through vital staining adding 50% (v/v) of Trypan Blue (Sigma) and then visualising the sample with the haemocytometer under the inverted microscope. Death of cells was measured by counting the number of cells that incorporated the staining.

2.6.6 Transfection by electroporation

Stable transfections of HT1080, DG75 and K562 cells were carried out through electroporation of ≈ 5x10^6 cells. Cells were centrifuged at 400 g for 5 min in order to remove the media and then washed once in PBS, centrifuged again and resuspended in 0.8 ml of cold PBS. 10 µg of vector to be transfected was linearised using a suitable restriction site (PvuI unless differently specified) and then was ethanol precipitated and resuspended in ≈

10 µl of sterile TE buffer. Chilled cells (in 0.8 ml PBS) were then mixed with the linearised DNA and immediately transferred to an electroporation cuvette (0.8 ml with 4 mm gap BTX Harvard Apparatus). Electroporation was achieved with a Gene-Pulser (BioRad): HT1080 cells were electroporated at 400 V/250 µF while K562 and DG75 cells at 240 V/960 µF. Immediately after the electroporation, cells were resuspended in warm media and left to recover at 37°C and 5% CO_2 for 48h before starting antibiotic selection.

Transient transfections in DG75 and K562 cells were carried out through electroporation of 2, 4, 6 or $8x10^6$ cells. Cells were centrifuged at 400 g for 5 min in order to remove the media and then washed once in PBS, spun again and resuspended in 0.4 ml of cold PBS. Chilled cells (in 0.4 ml PBS) were then mixed with 6 µg of DNA and immediately transferred to a 0.4 cm cuvette (BTX Harvard Apparatus). Electroporation was obtained with a Gene-Pulser (BioRad) at 240 V/960 µF. Cells were resuspended in warm media and left to recover at 37°C and 5% CO_2 for 48h before flow cytometry analysis.

2.6.7 Transfection by lipofection

HT1080 and DG75 cells were transfected using Lipofectamine2000 (Invitrogen) according to manufacturers instructions. The day before transfection, cells were plated on 24 wells plate at a density of 0.5, 1, 1.5 or $2x10^5$ cells per well using DMEM (HT1080) or RPMI (DG75) medium without antibiotics (penicillin or streptomycin). Unless differently specified, 0.8 µg of DNA (for each well to be transfected) were mixed in 50 µl of opti-MEM reduced-serum medium (Gibco). At the same time 2 µl Lipofectamine2000 were added to 50 µl opti-MEM. The two solutions were incubated for 5 min, mixed together and then left to incubate for another 20 min. The 100 µl solution was then added to each well of the 24 wells plate and cells were left in incubation at 37°C and 5% CO_2 for 5 hours before replacing the medium with standard DMEM or RPMI medium. Alternatively, lipofection of HT1080 cells was scaled-up plating $4x10^5$ cells in a 6-wells plate and transfection was carried out the day after using 4 µg of DNA. Transfection reactions were performed following the above protocol but, in order to take into account the

increased number of cells and µg of vector, the amount of lipofection reagents was modified following the manufacturer's guidelines.

2.6.8 Transfection by nucleofection

Transfections of K562 suspension cells were carried out using two different sets of reagents. The Ingenio Electroporation Kit (Mirus) was used in combination with the Nucleofector II apparatus (Lonza-Amaxa) set using the T-016 program (suggested by Lonza for this specific cell line). Briefly, $1x10^6$ (per reaction) of actively dividing cells were spun at 1500 rpm for 5 min, were resuspended in PBS and spun again to remove traces of medium. Cells were then resuspended in 100 µl Ingenio Solution, mixed with 2 µg of DNA (unless differently specified) and transferred in a 0.2 cm cuvette (Eppendorf). Cells were electroporated with the Nucleofector II machine and immediately transferred to 5 ml pre-warmed RPMI medium. Transfection efficiencies, measured by nucleofecting 2 µg of pmaxGFP (Lonza-Amaxa), resulted lower than expected (≈40% transfection efficiency) and therefore the Mirus reagent was exchanged for the Amaxa Nucleofector Kit V (including also Amaxa cuvettes). The protocol used was the same as the one used for the Mirus reagent. Transfection efficiencies increased with respect to previous experiments, reaching at least 70-80% GFP expression.

2.6.9 Transfection efficiency assay and flow cytometry

Transfection efficiencies were measured by transfecting cells with pmaxGFP plasmid (Lonza-Amaxa), assessing GFP expression after 48h with flow-cytometry. 0.8 µg or 2 µg pmaxGFP were used for lipofection or nucleofection, respectively. Efficiencies were measured for each transfection experiment and were carried out, on the same day, using the same cell population.

Flow cytometry was used to measure the percentage of GFP positive cells in sample populations during the Homology-directed Repair Assay and to generally assess transfection efficiencies during nucleofection, lipofection and

electroporation experiments. Each measurement was performed on 20-50x10^3 cells per sample (depending on cell viability). HT1080 cells were trypsinised and resuspended in DMEM medium before performing flow cytometry while K562 and DG75 cells were counted (with haemocytometer) and directly used for the assay. All measurements were carried out using a FACSCalibur machine (BD). Data were acquired with the CellQuest Pro software (BD) and analysed with the Flowjo software package (Treestar).

2.6.10 Selection of clones

HT1080 clones were obtained by plating HT1080 cells on a 10 cm plate immediately after stable transfection (see Section 2.6.6) and by leaving them to recover for 48h. After this period of time, 0.4 µg/ml puromycin selection was carried out for 15 days (DMEM medium with antibiotic was exchanged every 3 days) and then single colony forming units were selected and expanded in new 10 cm plates with fresh DMEM medium supplemented with puromycin.
Potential K562 clones from experiments 1,2 and 3 (Chapter 6) were obtained by limiting dilution: transfected cells (see Section 2.6.8) were left to recover for 48h after transfection and then plated on a 96 wells plate at a density of 0.5-1x10^3 cells/well (depending on the amount of transfected construct) in presence of 0.4 µg/ml puromycin. Selection was carried out for 15 days and then Puro resistant wells were expanded in fresh RPMI medium with puromycin for further studies.

2.6.11 Limiting dilution and Poisson distribution

Frequencies of gene targeting events in K562 cells during experiments 1, 2 and 3 (Chapter 7) were measured by limiting dilution analysis. 48h after transfection 10^3 cells/well were plated on 96 wells plates in presence of 0.4 µg/ml puromycin. Selection was carried out for 15 days and then Puro resistant wells were counted. In order to estimate the average of Puro resistant cells per well the formula $n = -\ln F_0$ was used (Poisson distribution).

F_0 is the number of wells not containing resistant cells while n is the average of Puro resistant cells per well.

2.7 Assays for targeted DNA binding or mutagenesis

2.7.1 Bacterial Two Hybrid (B2H) assay

Binding affinity of the produced ZF arrays were tested by the use of the B2H assay based on the β-galactosidase assay. The experiments were carried out following the protocol described by Wright et al. (Wright, Thibodeau-Beganny et al. 2006). For each KJBAC1 bacterial strain harbouring pGP-FF-ZF, pAC-KAN-alphaGal4 and pBAC-ZFTS-LacZ 4 ml of liquid mini-cultures containing 100 µg/ml ampicillin, 12.5 µg/ml cloramphenicol, 30 µg/ml kanamycin, 500 µM IPTG and 10 µM $ZnSO_4$ were genertaed by inoculation with a single colony and overnight incubation. Control cultures contained pGP-FF (without any ZF domain), pAC-KAN-alphaGal4 and pBAC-ZFTS-LacZ and were grown in liquid mini-cultures containing 100 µg/ml ampicillin, 12.5 µg/ml cloramphenicol, 30 µg/ml kanamycin, 500 µM IPTG and 10 µM $ZnSO_4$. The minicultures were sub-cultured by dilution (1:40) in 4 ml containing 100 µg/ml ampicillin, 12.5 µg/ml chloramphenicol, 30 µg/ml kanamycin, 500 µM IPTG and 10 µM $ZnSO_4$, the growth was monitored by measuring the OD at 600 nm and the cells were harvested at log phase (OD=0.3-0.8) and immediately lysed (see following description). The OD_{600} at the time of harvesting was recorded for each sample and the cultures were lysed in a 96-well plate by adding 100 µl of culture to 11 µl of Lysis Master Mix (10:1 [v:v} mixture of Popculture reagent to 400 U R-lysozyme, both reagents are from Novagen). Lysis was carried out for 30 min at room temperature. 15 µl of each lysate was added to 135 µl of Z-buffer (1 l of Z-buffer was prepared with 16.1 g of Na_2HPO_4-$7H_20$, 5.5 g of NaH_2PO_4-H_20, 0.75 g of KCl, 0.246 g of $MgSO_4$-$7H_20$ and 2.7 ml of β-mercaptoethanol dissolved in ddH_20 and filter sterilized) and 30 µl of 4 mg/ml ONPG (ortho-Nitrophenyl-β-galacotiside). The 96-wells plate containing the samples was then placed in a plate reader at room temperature and serial measurements of absorbance at 420 nm were taken every minute

during a period of 30 min.

Velocity of cleavage of ONPG molecules was calculated plotting the values of the absorbance at 420 nm against time and then calculating the slope of the line. Binding efficiencies of the ZF arrays were expressed as the ratio of velocities from the experimental and control cultures, after normalising each to account for any differences in cell density. Thus binding was measured using the formula $[V/OD_{600}]/[V^{NC}/OD^{NC}_{600}]$ where V and V^{NC} are the rates of increase in OD_{420} obtained from the experimental and control cultures, respectively, while OD_{600} and OD^{NC}_{600} are the OD_{600} measurements at the time of harvesting the experimental and control cultures, respectively.

2.7.2 Surveyor Assay

Surveyor mutation detection kit was provided by Transgenomic and is used to detect Single Nucleotide polymorphisms (SNPs), point mutations, insertions and deletions in a specific DNA sequence. The assay uses the Cel-I endonuclease from celery that is able to recognise small mismatch regions in a heteroduplex DNA fragment. The protocol is based on 3 main steps: PCR amplification of the desired DNA sequence, creation of mismatch regions and Cel-I digestion.

PCR reactions were performed using the Phusion Hot Start II high-fidelity DNA polymerase from Finnzymes (Thermo Fisher Scientific). Reaction reagents were: 0.5 µl Phusion polymerase, 10 µl 5x PCR buffer, 200 µM (final concentration) dNTPs, 0.4 µM (final concentration) primers, 32.5 µl of H_2O, 4 µl genomic DNA originated from cells transfected with ZFN4 expression plasmids and mRNA (Chapter 4).

The primers used to amplify the region of interest in the β-globin gene were:

Primer name	Sequence (5'-3')
HBB-F	AGGGTTGGCCAATCTACTCC
HBB-R	CAAAGAACCTCTGGGTCCAA

The primers used for the δ-globin region were instead:

Primer name	Sequence (5'-3')
HBD-F	AACTGCTGAAAGAGATGCGGTGG
HBD-R	AATGTGGGAGAAGAGCAGGTAGG

Cycles conditions:

Initial denaturation	98˚C	5 min
Denaturation	98˚C	1 min
Annealing	57(β)-58(δ) ˚C	30 sec
Extension	72˚C	1min
Final extension	72˚C	10 min

Number of cycles: 30

PCR products were purified using the commercial DNA recovery kit from NEB (previously described, Section 2.1.4) and resuspended in 20 µl of 1x Annealing buffer (40 mM Tris pH 8, 20 mM $MgCl_2$, 50 mM NaCl). The amplified DNA fragments were hybridised in order to form heteroduplex using a thermocycler:

Denaturation 95˚C	2 min
95˚C ramping to 85˚C	-2˚C/sec
85˚C ramping to 25˚C	-0.1˚C/sec
4˚C	hold

Newly formed heteroduplexes were subject to digestion with Cel-I following the manufacturer's instructions. Possible DNA fragments derived by digestion of mismatched regions were then visualised on an 8% polyacrylamide gel (previously described, Section 2.3.2).

Estimation of the frequencies of heteroduplexes formation was used as an indicator of ZFN4 activity at the target locus. To evaluate the frequencies, a known amount of Cel-I digested product (≈ 1 µg) was loaded on a 8% polyacrylamide gel in conjunction with a known amount of 100 bp ladder (0.5 µg, NEB). The estimation was carried out by comparing the relative intensities of the two digested products (219 bp and 175 bp) and the ones of the 100,

200 and 300 bp bands of the ladder. By manufacturer's specification, the µg of the 100, 200 and 300 bp bands in 0.5 µg of ladder are 48 ng, 25 ng and 29 ng respectively.

2.7.3 MwoI restriction digestion

In order to confirm, by alternative means, the presence of ZFN4 induced gene modifications at the β-globin locus, PCR reactions were carried out as described in Section 2.7.2 using the same primers and genomic DNA.

PCR products were purified using the commercial DNA recovery kit (previously described, Section 2.1.4) and resuspended in 20 µl H_2O. Restriction digestions with MwoI were performed following the manufacturer's protocol (NEB) within the space of 2 hours. Restriction digestion products were then visualised on an 8% polyacrylamide gel (described in Section 2.3.2) and ZFN4 cleavage efficiencies were estimate as in Section 2.7.2.

2.7.4 Homology Direct Repair (HDR) assays

Plating densities for HT-DR-ZFN4, DG-DR-ZFN4 and HT-DR-GFP cells in lipofection experiments were chosen based on the ones suggested by the manufacture's protocol (Lipofectamine 2000, Invitrogen). Experiments performed on DG-DR-ZFN4 and K-DR-ZFN4 by use of electroporation followed a similar timeline but the protocol was modified to take in account the differences in transfection technique (see below for further description of the used protocol).

Briefly, on day 1 HT-DR-ZFN4, HT-DR-GFP or DG-DR-ZFN4 cells were plated on 24-well plates at different densities: $1x10^5$ and $1.5x10^5$ cells per well for HT-DR-ZFN4 and HT-DR-GFP and $4x10^5$ and $6x10^5$ cells per well for DG-DR-ZFN4. On day 2 cells were transiently transfected by use of lipofection (see Section 2.6.7) with 0.4 µg of ISceI expression plasmid and 0.4 µg of a control plasmid (pGP-FF) or with 0.4 µg of both ZFN4 expression plasmids. Negative controls were performed by transfection of $1x10^5$ and $1.5x10^5$ HT-DR-ZFN4 and HT-DR-GFP cells or $4x10^5$ and $6x10^5$ DG-DR-ZFN4 cells with

0.8 µg of pGP-FF control plasmid. Transfection efficiencies for each cell line were assessed by transfection of 0.8 µg of pmaxGFP plasmid (Lonza).

Using an alternative approach, $2x10^6$ DG-DR-ZFN4 and K-DR-ZFN4 cells were electroporated on day 2 with 3 µg of ISceI expression plasmid and 3 µg of a control plasmid (pGP-FF) or with 3 µg of each ZFN4 expression plasmid. Negative controls were performed transfecting the same amount of of DG-DR-ZFN4 and K-DR-ZFN4 cells with 6 µg of pGP-FF control plasmid and transfection efficiencies were assessed by transfection of 6 µg of pmaxGFP plasmid (for the electroporation protocol see Section 2.6.6). All samples were left to recover for 48h at 37°C in presence of 5% CO_2 and then, on day 4, GFP expression levels were measured by flow cytometry (see Section 2.6.9 for the flow cytometry protocol).

2.7.5 PCR screening of genomic DNA from puromycin resistant cells

Screening of puromycin resistant cells for the detection of possible targeting events was carried out using DreamTaq DNA Polymerase (Fermentas) with the following reaction conditions:

0.5 µl DreamTaq polymerase, 5 µl 10x DreamTaq buffer, 200 µM (final concentration) dNTPs, 0.4 µM (final concentration) primers, 37.5 µl H_2O and 4 µl of genomic DNA.

Initial denaturation	95˚C	1 min
Denaturation	95˚C	30 sec
Annealing	57˚C	30 sec
Extension	72˚C	1min
Final extension	72˚C	5 min

Number of cycles: 30

The primers used in the PCR reactions were:

Primer name	Sequence (5'-3')
PT1-F	GGTGACAATTTCTGCCAATCAGG
PT2-F	TAGGCTCCAGATAGCCATAGAAG
PT3-F	AGAGCTGAAAGGAAGAAGTAGGAG
PT1-R	CCCGCAAGCCCGGTGCCTGA
PT2-R	GCATTCTAGTTGTGGTTTGTCC

PT1-F, PT2-F and PT1-R were initially used in the preliminary PCR tests using pBL-TV-pc positive control plasmid as a template (see Section 2.5.3). PT3-F and PT2-R were selected to be used during PCR screening of puromycin resistant cells that were transfected with pTV-TC4 (see Sections 2.5.3).

2.7.6 Southern blot analysis

6 µg of genomic DNA to be analyzed were digested overnight with 100 U of HindIII in presence of 5 mM Spermidine (Sigma) in 100 µl volume. 1 µg of radiolabelled 1kb ladder and the digested DNA were then separated on a 0.8% agarose gel at 15 V overnight. The gel was then immersed in denaturation solution (1.5 M NaCl, 0.5 M NaOH) for 45 min, twice in neutralization solution (1 M Tris-HCl, 1.5 M NaCl, pH 7.4) for 20 min and then briefly rinsed in deionized water. The denaturated DNA in the gel was then blotted onto a positively charged nylon membrane (HybondTM-N+, Amersham) by capillary transfer. The transfer was carried out by following the protocol described by Sambrook et al. (J. Sambrook 1982). The DNA sample was left to transfer overnight. The day after the nylon membrane was washed with 2x SSC, baked for 1h at 80°C and then the transferred DNA was cross-linked to the membrane using a Stratalinker UV Crosslinker 1800 (Stratagene) for 2 min.

The membrane was then pre-hybridized with 50 ml of Church mix (0,25 M Sodium Phosphate pH 7.2, 1 mM EDTA, 1% BSA, 7% SDS) in a roller bottle at 68°C for 1h. The ^{32}P-labelled probe was boiled for 5 min and then

added to the Church mix in the roller bottle. The solution was left to incubate over night at 68°C. The following day the excess of probe was washed away at 68°C with first 2x SSC, two times with 50 ml of 1x SSC and in the end with 50 ml of 0.5x SSC. A Hyperfilm MP (GE Healthcare) was set with the membrane between two layers of intensifying screens, exposed for 4 days at -80°C and then developed.

3 Chapter 3: Creation of zinc finger binding domains by modular assembly and development of a "directed evolution approach"

3.1 Overview

At the beginning of this project, ZFNs specific for the β-globin sequence were designed and developed using a modular assembly approach protocol that was published in 2006 by a group participating in the Zinc Finger Consortium (Wright, Thibodeau-Beganny et al. 2006). This protocol, in theory, allows users to generate ZFNs specific for any desired sequence through a simple succession of standard molecular cloning steps in which single ZF modules are linked together to form a DNA Binding Domain (DBD). The ZF-based DBDs are tested for their ability to bind the desired target sequences through a Bacterial 2 Hybrid assay. Pairs of DBDs that bind their target sequences well are then fused to the FokI nuclease domain and tested for their ability to cleave the target locus in human cell lines.

The first step of the protocol is to identify potential target sequences in a desired region of the genome that can be targeted by ZFNs produced using this specific approach. This process is based on the use of the ZiFiT (Zinc Finger Targeter, http://zifit.partners.org/ZiFiT/) web-based software that was designed initially to be used in combination with the modular assembly protocol but that has been updated in order to support recent and more efficient approaches like the OPEN or CoDA protocols (Sander, Zaback et al. 2007; Sander, Maeder et al. 2010). The program was developed as a web tool capable of analyzing a specific DNA sequence and, scanning through a depository of ZF modules, to identify possible combinations of ZF modules specific for small sequences of 6, 9 or 12 nucleoitdes in the desired genomic region.

In a second step, DNA encoding the ZF-DBDs is assembled using standard cloning techniques. Plasmids, based on the vector pc3XB (Wright,

Thibodeau-Beganny et al. 2006) and encoding each single ZF module (capable of recognising a specific trinucleotide), are provided by the Zinc Finger Consortium through Addgene. Each ZF module sequence is flanked by two restriction enzyme sites (XmaI, BamHI) that are used to isolate and then sub-clone the ZF module DNA. In this way, DNA encoding the different modules is assembled, by DNA ligation, in order to form a sequence encoding the desired DBD composed of 3 or 4 ZF modules (Fig. 3.1 a).

The newly created ZF-DBDs are then tested for their ability to bind their specific target sequence by use of a Bacterial 2 Hybrid assay (B2H assay) that was developed to easily test protein-protein interactions or DNA-protein interactions (Joung, Ramm et al. 2000; Wright, Thibodeau-Beganny et al. 2006). The assay is based on the use of the KJBAC1 bacteria strain transformed with three different plasmids (Fig. 3.1 b). DNA for each assembled ZF-DBD is cloned into an expression plasmid (pGP-FF). The resulting plasmid (pGP-FF-ZF) confers resistance to ampicillin and encodes a fusion protein with the Gal11 subunit (ZF-Gal11) which is expressed in an IPTG-inducible manner. A second plasmid (pAC-KAN-αGal4) confers resistance to kanamycin and expresses the fusion protein αGal4-RNA polymerase, also in response to IPTG. Finally, a third plasmid (pBAC-lacZ) confers resistance to chloramphenicol and is used to clone the target sequence for the ZF-DBD to be tested close to a lacZ cassette (Fig. 3.1 b). The assay tests the ability of ZF-Gal11 to bind to its target sequence in pBAC-lacZ and so to recruit the αGal4-RNA polymerase that is able to promote the expression of the lacZ cassette (Fig. 3.1 c).

In this way, the binding of the ZF-DBD to its target sequence is measured as the expression of the lacZ cassette. For this a standard β-galactosidase assay is used in which the velocity of cleavage of ONPG (a colorimetric artificial substrate of the β-galactosidase) is measured spectrophotometrically as increasing absorbance at 420nm with time.

a)

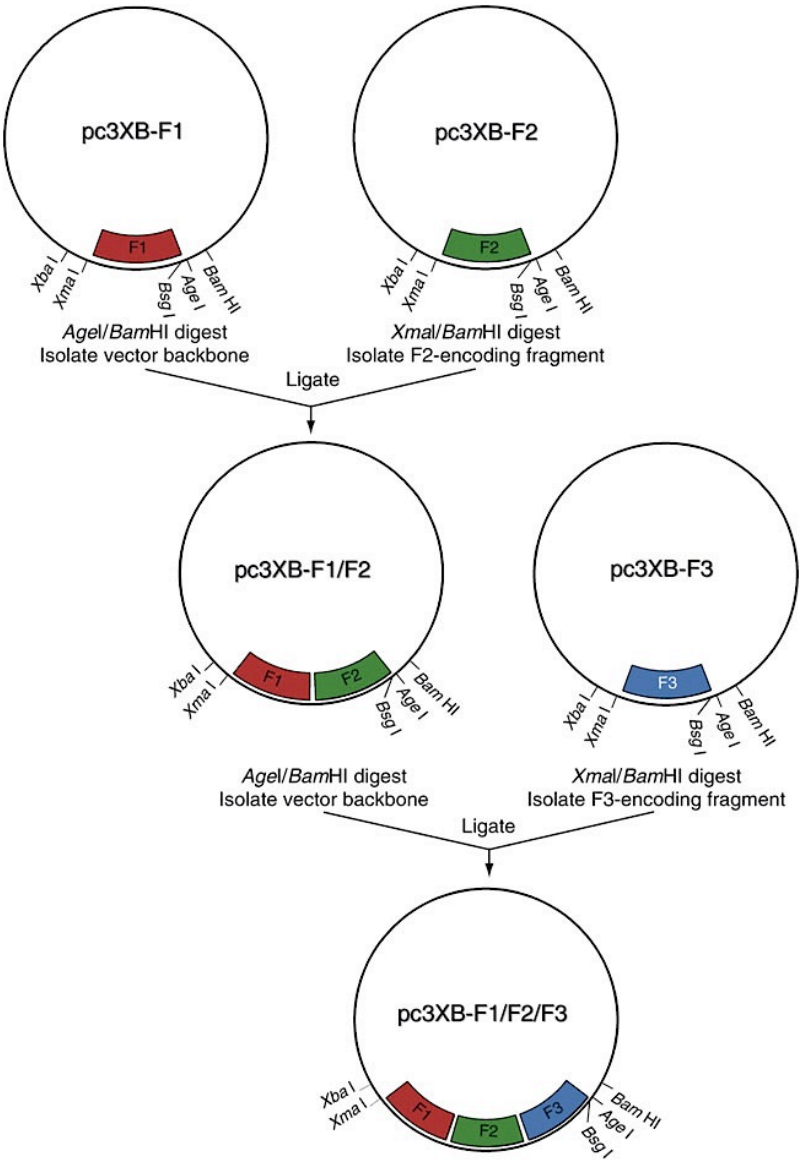


Figure 3.1 (a) Overview of the modular assembly approach.
Restriction digestion and ligation of ZF modules (F1-F2-F3). The example shows the successive cloning steps for the assembly of DNA encoding ZF-DBD. The pc3XB-F1 plasmid is digested with AgeI and BamHI to create the recipient vector while pc3XB-F2 is digested with XmaI and BamHI to isolate the F2 fragment. The vector and the ZF2 fragment are then ligated to create the pc3XB-F1-F2. Similar steps can be repeated in order to add one ZF at a time and so to build up ZF-DBDs made of 2, 3 or 4 ZF modules (picture obtained from Wright et al. 2006).

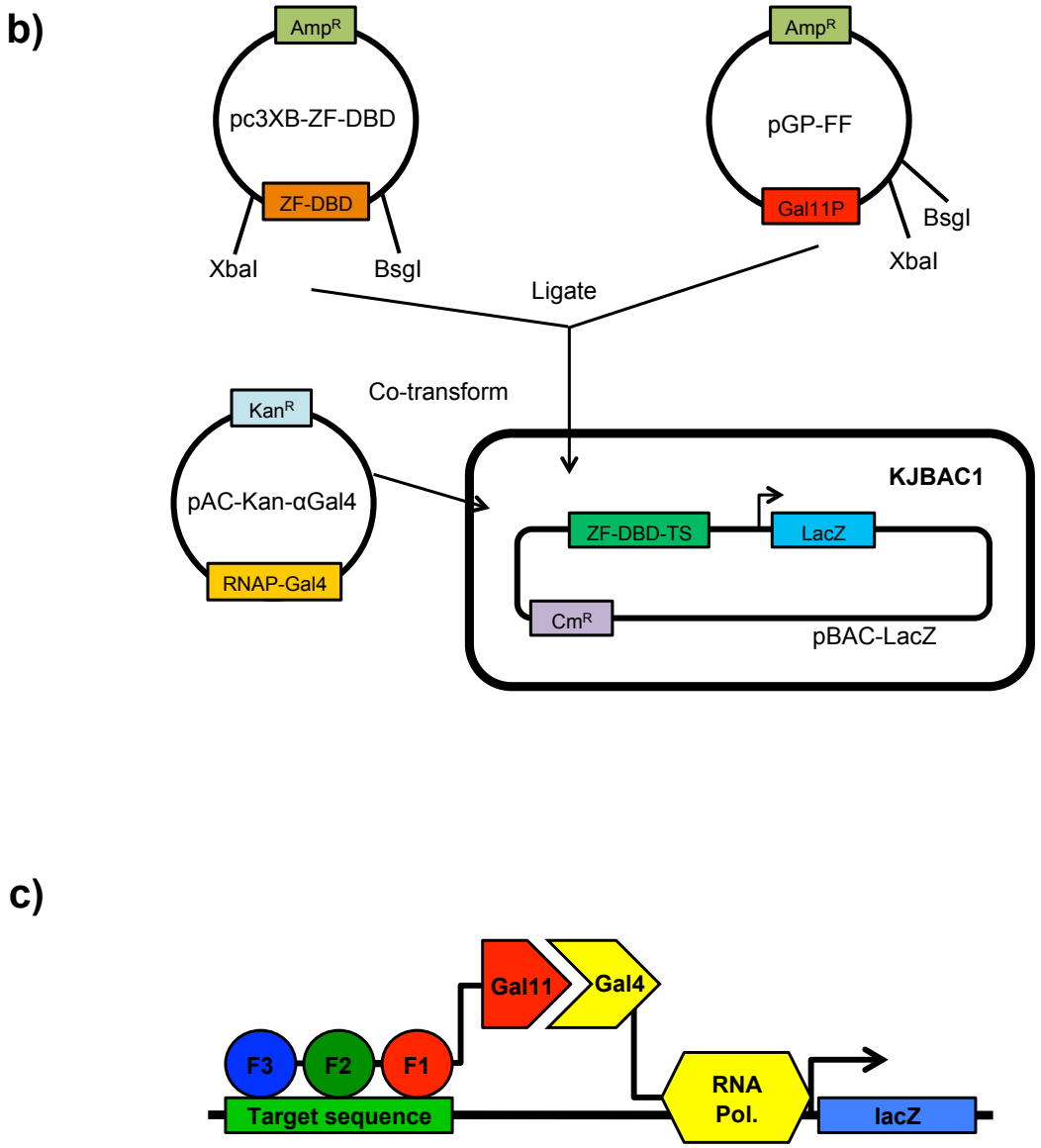


Figure 3.1 (b and c) Overview of the modular assembly approach.
b) Creation of the B2H reporter system. The XbaI-BsgI digested pGP-FF plasmid is ligated to the XbaI-BsgI fragment ZF-DBD (encoding for F1-F2-F3) creating the IPTG-inducible expression vector pGP-FF-F1-F2-F3. The expression vector is then co-transformed with pAC-KAN-alphaGal4 in the KJBAC1 bacteria strain already harbouring the reporter plasmid pBAC-ZFTS-LacZ. **c)** Baterial two hybrid (B2H) assay. Co-expression of ZF-F1-F2-F3-Gal11 and Gal4-RNA-polymerase is induced by IPTG. Successful binding of the ZF domain to the target sequence recruits the fusion protein Gal4-RNA-polymerase that activates the expression of the LacZ reporter gene. A standard colorimetric assay is then used to quantify the binding efficiencies of any produced ZF domain.

3.2 Results

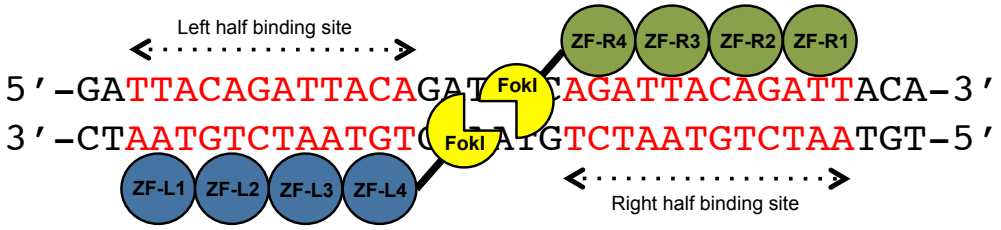
3.2.1 Identification of target sequences in the human β-globin gene

The first step of the modular assembly approach was to identify possible target sequences in the human β-globin gene that could be targeted by ZFNs produced following this protocol. It was decided to look for 12-nucleotide sequences for each half site (4 ZF units) in order to increase the specificity of the binding and to avoid possible off-target cleavage. ZFNs produced in this way would target a sequence of 24 nucleotides: left and right ZF domains would bind to 12 nucleotides on each sides of a 6-nucleotide spacer sequence where the FokI domain would produce a DSB (Fig. 3.2 a).

The ZiFiT software identified 50 possible targets; from this group, three (TS-1, TS-2 and TS-3) were selected on the basis of their positions relative to known β-thalassemia mutations (TS-1 being close also to the sickle cell mutation) and for their reduced risk of possible off-target cleavage events (Fig. 3.2 b). The risk of off-target cleavage is a particularly limiting factor in the development of ZFNs specific for the β-globin gene because it has high homology with the other globin genes, especially with the δ-globin gene. A search for possible homologies was carried out by use of the BLAST algorithm (Basic Local Alignment Search Tool) (BLAST). The BLAST query was performed by search of the “human genomic database” for regions of partial homology with the human β-globin sequence (obtained from the NCBI Nucleotide database as a FASTA format sequence) with the settings optimised for “highly similar sequences” (megablast). As expected, the result of the query indicated the δ-globin gene as the one with the highest level of homology with the β-globin gene. This partial homology was especially significant in the regions of the three exons (for a more detailed analysis see Appendix 1). TS-2 and TS-3 were considered sufficiently different from their corresponding δ-globin analogues and therefore a reduced risk of off-target cleavage was expected. On the other hand, TS-1 differed from its own δ-

globin analogue by only one bp indicating the risk of potentially unwanted gene modifications that could affect the overall safety and efficacy of a possible therapeutic application for a ZFN specific for this target sequence (the alignment of the three target sequences to their δ-globin analogues is shown in Appendix 1). However, the synthesis of the ZF-DBDs for TS-1 was still carried forward because of the opportunity of studying off-target cleavage events in the highly homologous globin family and also for the possibility, in the future, to improve the specificity and safety of ZFNs.

a)



b)

cacctgtggagccacaccctagggttggccaatctactcccaggagcagggagggcaggagccaggg
ctgggcataaaagtcagggcagagccatctattgcttacatttgcttctgacacaactgtgttcacta
gcaacctcaaacagacACCATTTGCTTCTGACACAACTGTGTTCACTATAGCAACCTCAAACAGACAC
CATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATG
AAGTTGGTGGTGAGGCCCTGGGCAGgtaggtatcaaggttacaagacaggtttaaggagaccaataga
aactggcatgtggagacagagaagactcttgggtttctgataggcactgactctctctgcctattggg
tctattttcccacccttagGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGG
GATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCTCGG
TGCCTTTAGTGATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGC
ACTGTGACAAGCTGCACGTGGATCCTGAGAACTTCAGGgtgagtctatgggacgcttgatgttttctt
tccccttcttttctatggttaagttcatgtcataggaaggggataagtaacagggtacagtttagaat
gggaaacagacgaatgattgcatcagtgtggaagtctcaggatcgttttagtttcttttatttgctgt
tcataacaattgttttcttttgtttaattcttgctttctttttttttcttctccgcaatttttactat
tatacttaatgccttaacattgtgtataacaaaaggaaatatctctgagatacattaagtaacttaaaa
aaaaaactttacacagtctgcctagtacattactatttggaatatatgtgtgcttatttgcatattca
taatctccctactttatttcttttatttttaattgatacataatcattatacatatttatgggttaaa
agtgtaatgttttaatatgtgtacacatattgaccaaatcagggtaattttgcatttgtaatttttaaa
aaatgctttcttcttttaatatactttttgtttatcttatttctaatactttccctaatctctttct
ttcagggcaataatgatacaatgtatcatgcctctttgcaccattctaaagaataacagtgataattt
ctgggttaaggcaatagcaatatctctgcatataaatatttctgcatataaattgtaactgatgtaag
aggtttcatattgctaatagcagctacaatccagctaccattctgcttttatttttatggttgggataa
ggctggattattctgagtccaaagctaggcccttttgctaatcatgttcatacctcttatcttcctccc
acagCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAGAATTCACCCCACCAG
TGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCCACAAGTATCACTAAGCT
CGCTTTCTTGCTGTCCAATTTCTATTAAAGGTTCCTTTGTTCCCTAAGTCCAACTACTAAACTGGGGGG
ATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAATAAAAAACATTTATTTTCATTGCAAtgatg
tatttaaattatttctgaatattttactaaaaagggac

Figure 3.2 ZFN binding and cleavage example and β-globin gene sequence.

a) ZFN binding and cleavage overview. The picture represents a generic ZFN: half binding sites are coloured in red and indicated by dashed lines. ZF left and right domains are coloured in blue and green, respectively. The FokI domain, in its dimeric form is represented in yellow. b) β-globin gene and ZFNs binding sites. Exons are in black capitalized characters while introns are in blue lower case characters. Residues highlighted in orange are known sites for mutations-causing β-thalassaemia and the sickle-cell mutation is highlighted in purple. Target sequences TS-1, TS-2, TS-3 are highlighted in red, green and yellow, respectively. Left and right half-sites to be bound by the proposed ZF domains are underlined.

3.2.2 Modular assembly of DNA encoding ZF-DBDs for TS-1, TS-2 and TS-3

Six different ZF-DBDs (2 for each target sequence) were assembled through a process of restriction enzyme digestions and ligations where each single ZF module was isolated and then linked to one another in order to obtain functional recognition domains (Fig 3.1 a). In order to reduce the time spent in the creation of ZFNs, it was decided to commission the synthesis of two ZF domains (ZF2-L and ZF3-L) to a third party: the Mr Gene company (see Section 2.5.4). Although the use of a third party increased the cost of production of the ZF-DBDs, this approach greatly reduced the time spent in the assembly process.

Plasmids encoding the four remaining ZF-DBDs were created following the standard modular assembly protocol. Each ZF module was provided, in a standard pc3XB plasmid, by the Zinc Finger Consortium through Addgene (see Section 2.5.3). The plasmids encoding ZF1-R and ZF1-L DBDs, specific for the TS-1 target sequence, were assembled and tested for their binding activity by Hitoshi Kurata, a post-doctoral researcher in the Porter lab who was involved in the first stages of the project.

A summary of the six ZF-DBDs is reported in Table 3.1 while the complete sequence for each one of the ZF domains is shown in Appendix 1.

Table 3.1 Nomencaltue and composition of ZF-DBDs and their target sequences in the β-globin gene

[a] ZF -DBD	[b] ZF modules	[c] Target sequences name	[d] DNA sequence
ZF1-R	ZF60-71-106-94	TS-1 (R)	5′-TGGACGGGTCCC-3′
ZF1-L	ZF92-103-89-90	TS-1 (L)	5′-TACTTCAACCAC-3′
ZF2-R	ZF73-104-83-93	TS-2 (R)	5′-CCGGATCGAACC-3′
ZF2-L	ZF73-93-77-77	TS-2 (L)	5′-CCGACCTAATAA-3′
ZF3-R	ZF106-75-82-77	TS-3 (R)	5′-GGTCAAAGATAA-3′
ZF3-L	ZF69-96-74-103	TS-3 (L)	5′-CTGTCCAAATTC-3′

[a] Right and left ZF DBDs are designated by R and L, respectively.
[b] ZF modules (arranged from N terminus to C terminus) used during the assembly of each domain. Labelling follows the Addgene catalogue's nomenclature (see Appendix 1 for the complete sequence of the ZF-DBDs).
[c] Half target sites corresponding to each ZF domain. The highlights are color-coded as in Fig. 3.2
[d] Actual DNA sequences of each target site.

3.2.3 Assessing ZF-DBD binding activity in the B2H assay

Binding efficiencies of ZF-DBDs created with this type of approach can vary greatly due to unwanted interactions between one module and the opposite strand of the following binding site. For this reason, each domain's binding activity must be individually tested by use of the bacterial two-hybrid assay (Joung, Ramm et al. 2000; Wright, Thibodeau-Beganny et al. 2006).

The six half-site target sequences (Table 3.1), in the form of short oligonucleotides, were cloned in the pBAC-lacZ (carrying a chloramphenicol resistance cassette) reporter plasmid immediately upstream of the lacZ cassette (pBAC-TS-1(R)-lacZ, pBAC-TS-1(L)-lacZ, pBAC-TS-2(R)-lacZ, pBAC-TS-2(L)-lacZ, pBAC-TS-3(R)-lacZ and pBAC-TS-3(L)-lacZ). Each pBAC-TS-lacZ was then individually transformed in the bacterial reporter strain KJBAC1 (Fig. 3.1 b, see Section 2.5.1 for details).

Following the modular assembly protocol (Wright, Thibodeau-Beganny et al. 2006), the six ZF-DBD's coding sequences were isolated from the pc3XB

plasmid by digestion with restriction enzymes and then cloned into the expression plasmid pGP-FF (carrying an ampicillin resistance cassette), which expresses each domain as a fusion protein with a Gal11 subunit, to generate pGP-FF-ZF1-R, pGP-FF-ZF1-L, pGP-FF-ZF2-R, pGP-FF-ZF2-L, pGP-FF-ZF3-R and pGP-FF-ZF3-L (Fig. 3.1 b).

Each newly created pGP-FF-ZF domain was then individually co-transformed with the pAC-KAN-αGal4 (carrying an kanamycin resistance cassette) plasmid into the KJBAC1 reporter strain, already harbouring the pBAC-lacZ plasmid (e.g. pBAC-TS1-lacZ) with the appropriate target sequence, selecting for bacteria that were resistant to all three antibiotics: ampicillin, kanamycin and chloramphenicol (Fig. 3.1 b).

The transformed KJBAC1 cells were cultured overnight in the presence of IPTG and antibiotics in order to express the two fusion proteins (ZF-Gal11P and the αGal4-RNA polymerase) and the cultures were tested for expression of β-galactosidase (Section 2.7.1). In order to quantify the binding efficiency, each experiment was carried out in parallel with a negative and a positive control and was repeated multiple times (Fig. 3.3). The negative control was created simply by co-transforming the KJBAC1 strain (carrying the relative pBAC-ZFTS-LacZ reporter plasmid) with the empty pGP-FF vector and pAC-KAN-αGal4. Expression of these two plasmids enables the interaction between the Gal11P subunit and αGal4-RNA polymerase, but in the absence of any ZF-DBD these direct minimal expression of the lacZ cassette. As a positive control, a pGP-FB-BA plasmid (Addgene) was used in place of the experimental pGP-FF-ZF plasmids. The positive control is provided by Addgene and it is based on a previously tested ZF domain that has been shown to be highly efficient (Wright, Thibodeau-Beganny et al. 2006). The positive control was created by transforming KJBAC cells with the relative plasmids (pGP-FB-ZF, pAC-KAN-αGal4 and pBAC-BA-LacZ) as previously described. The pGP-FB-ZF plasmid expresses a highly efficient ZF domain linked to the standard Gal11 subunit which specifically recognise its own target sequence carried by pBAC-BA-LacZ (Fig. 3.3).

The optical density at 600nm of each culture was determined, as a measure of cell density, as well as the β-galatosidase actvity, as measured by the velocity of ONPG cleavage (methods). Data collected were analyzed

using the formula $[V/OD_{600}]/[V^{NC}/OD^{NC}_{600}]$ where V is the slope of the line (velocity) obtained from kinetic data of an experimental culture (such as those in Fig 3.3 a), V^{NC} is the equivalent velocity derived from the negative control, and OD_{600} and OD^{NC}_{600} are the optical density measurements at 600nm of the two cultures (sample and negative control, respectively) at the time of the experiment. This ratio thus measures fold-increase, relative to the negative control, in the β-galactosidase activity for each test culture, normalising for any differences in cell density, and the ratio can be taken as a measure of the ability of the ZF-DBD in that culture to bind its target sequence. (Fig. 3.3, see also Section 2.7.1).

It is suggested in the modular assembly protocol that, in order to be effective in a mammalian model, the binding efficiency of each ZF domain, as measured by this ratio, should have a value of at least 3 (Wright, Thibodeau-Beganny et al. 2006). Unfortunately, of all six ZF-DBDs tested, only one (ZF1-R) showed a highly satisfactory ratio of 7 (compared with a ratio of 5 for the positive control). Of the other five ZF-DBDs, ZF3-R showed a ratio of 2.7 while the remaining four had a binding efficiency no higher than 1, i.e. no better than the negative control (data are summarized in Fig. 3.3).

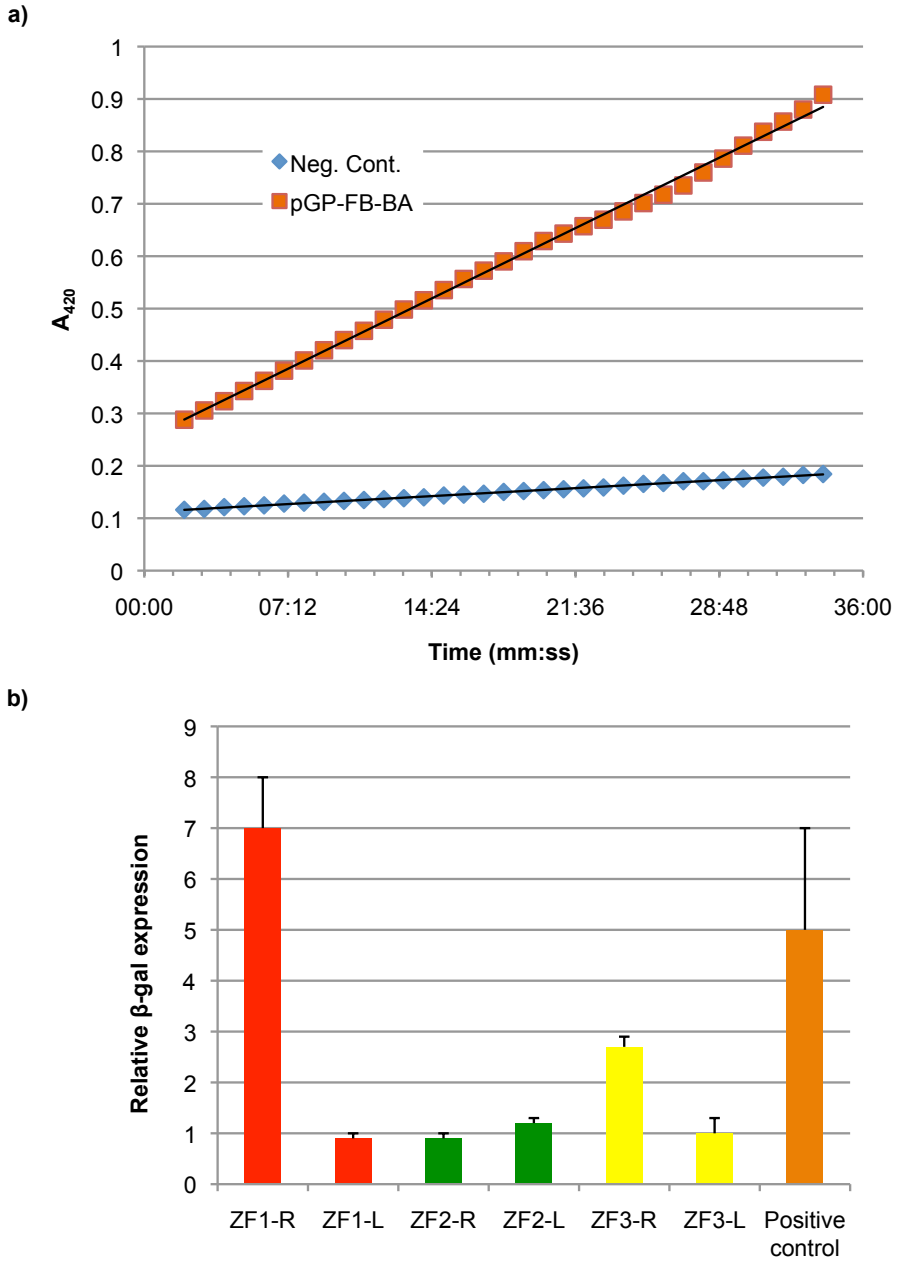


Figure 3.3 β-galactosidase assay and ZF domains' binding efficiencies.

a) β-galactosidase assay example. The velocity of ONPG cleavage is measured as the variation of absorbance at 420nm during time. The positive control pGP-FB-BA and the relative negative control are reported in orange and blue, respectively. **b)** ZF domains' binding efficiencies. This graph represents the binding efficiency of all ZF domains (colour coded as in table 3.1) plus the positive control expressed as the ratio $[V/OD_{600}]/[V^{NC}/OD^{NC}_{600}]$ with standard deviation. Each experiment was repeated 4 times except for ZF1-R that was repeated 5 times and for the positive control that was repeated 6 times.

3.2.4 Development of a directed evolution system

Although two early large-scale surveys of the modular assembly approach identified a success rate in producing effective ZF domains of 100% and 60% (Bae, Kwon et al. 2003; Mandell and Barbas 2006), a more recent and wider analysis highlighted how these figures were actually completely inaccurate (Ramirez, Foley et al. 2008). The study, which tested a greater number of ZF-DBDs and a wider spectrum of target sequences, showed a failure rate that was calculated to be around 76% with the obvious outcome of a high risk of producing ineffective ZF domains.

Unfortunately this high failure rate matched perfectly with the data obtained assessing the newly created ZF domains specific for the β-globin gene: only 1 domain out of 6 showing satisfactory binding, indicating a failure rate of 83%. In order to obtain more efficient domains and to take advantage of the ones already produced, it was decided to modify the newly created ZF domains with the aim of enhancing their binding efficiencies.

A two-step protocol was designed where, in the first step, a library of mutated ZF domains, based on the ones already produced, would be created through random mutagenesis. Thus, error-prone PCR would introduce random mutations in the coding sequence generating a number of similar ZF domains with a wide range of binding efficiencies that could then, in a the second step, be screened to select the most effective ones. This second step would be carried out through a selection system based on the B2H assay where a higher binding efficiency of the ZF protein would confer a survival advantage to the cells expressing it (Fig. 3.4).

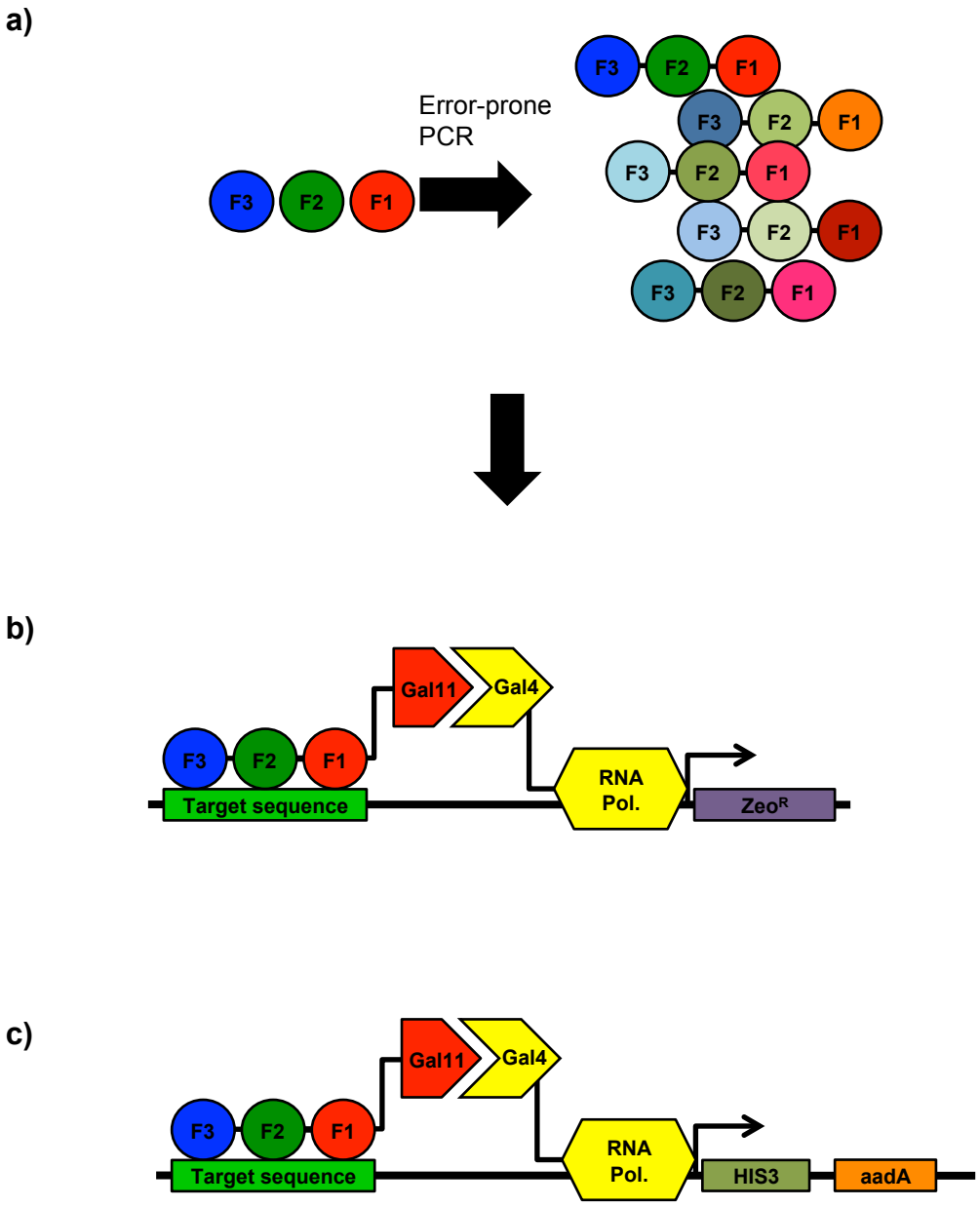


Figure 3.4 Directed evolution system.
a) ZF domain variants library. The first step of the directed evolution approach involves the creation of a wide library of ZF domain mutants through error prone PCR. **b)** Zeocin selectable system. In the second step, the newly created library is screened for ZF domains with improved binding efficiencies. A modified B2H system carrying a Zeo^R selectable marker in place of the original lacZ cassette is used to rapidly select between different ZF domains. Efficient binding of a possible domain leads to the expression of the Zeo^R cassette providing a survival advantage to the cell expressing it. **c)** OPEN selectable system. This alternative version of the selectable system was based on a double selection approach where two different markers (the recovery of the histidine pathway by HIS3 and streptomycin resistance by aadA) should increase the efficiency of the screening process.

Zeocin selection approach: The first selection system to be tested was based on the activation, through the binding of ZF domains, of a zeocin resistance cassette (Fig. 3.4 b). The lacZ cassette from pBAC-TS-1(R)-lacZ and pBAC-TS-1(L)-lacZ reporter plasmids was removed by restriction enzyme digestion and substituted with a zeocin resistance cassette creating the pBAC-TS-1(R)-ZeoR and pBAC-TS-1(L)-ZeoR plasmids that were then transformed into the standard KJBAC1 strain (this cloning step was performed by Hitoshi Kurata, a post doc working on the initial phase of this project).

In order to confirm the ability of the system to effectively discriminate between different ZF binding activities, the pGP-FF-ZF1-R and pGP-FF-ZF1-L were selected as positive and negative controls, respectively, based on their behaviour in the B2H assay (fig 3.3 b). Each plasmid was co-transformed with pAC-KAN-αGal4 into KJBAC1 cells harbouring the appropriate pBAC-ZeoR selection plasmid. pGP-FF-ZF1-R was paired with the pBAC-ZeoR containing the TS-1(R) target sequence (pBAC-TS-1R-ZeoR) while pGP-FF-ZF1-L was paired with the one containing the TS-1(L) target sequence (pBAC-TS-1L-ZeoR). Transformants (from now on called KJ-Zeo-pos and KJ-Zeo-neg for positive and negative controls, respectively) were selected overnight on agar plates with all three antibiotics (ampicillin, chloramphenicol and kanamycin). To test for ZF-DBD dependent zeocin-resistance, the day after, individual colonies were picked and grown overnight in 4 ml low-salt LB (high ionic strength inhibits the antibiotic activity) containing ampicillin, chloramphenicol kanamycin and IPTG (from now on shorten to amp-kan-cam-I) but not zeocin. The day after, 10 µl of the inoculum were collected and used to start a fresh culture in 4 ml low-salt LB (amp-kan-cam-I) and 25 µg/ml zeocin (Section 2.2.1).

Bacteria growth was monitored by periodically measuring (by means of a spectrophotomiter) the OD of the culture at 600 nm until OD reached values of 0.8. Unfortunately both controls grew in presence of zeocin without any significant distinctions between positive and negative. For this reason, different concentrations of zeocin (between 25 and 50 µg/ml, as suggested by the manufacturer) were used in order to increase the stringency of the

selection but the negative control was still growing in presence of the antibiotic. The genotype of KJBAC1 (obtained from Addgene and reported in Section 2.2.2) seems to indicate a lack of inactivation of the recA1 pointing out the risk of internal recombination between the three different plasmids (pGP-FF-ZF1, pAC-KAN-αGal4 and pBAC-TS-1-ZeoR). In order to rule out any possible acquired constitutive zeocin resistance derived from a recombination event, a different cell line (JM109, a known recA1 strain) was tested. JM109 cells were transformed with pBAC-ZeoR harbouring the two different target sequences (creating novel positive and negative control) and the test was repeated following the previously described approach.

Unfortunately, also in this case, both positive and negative controls grew in presence of zeocin reaching OD_{600} 0.8 at the same time. The results at this point apparently indicated that the problem was related not to the host strain or to the antibiotic concentration, therefore it was supposed that the ineffective selection was due to some characteristic of the medium itself. The manufacturer (Invitrogen) protocol for the zeocin states that the antibiotic is stable and effective in a very narrow pH window (around pH 7.5). Therefore it was decided to test the pH of the medium by measuring it using a pH meter (Hanna Instruments, HI 931410). The pH of the low-salt LB medium, before any bacteria growth was started, was found to be around pH 6.3 and thus identified as a possible cause of the antibiotic inefficiency. The pH of the low-salt LB was therefore corrected to the desired value of pH 7.5 with NaOH and the medium was then used in the next series of experiments using fresh KJ-Zeo-pos and KJ-Zeo-neg transformants. Using this medium it became possible to discriminate between positive and negative controls but this test was carried out, due to the small volume of the culture (4 ml), only during a period of 4-5 hours therefore a prolonged test was carried out overnight in order to assess the stability of the zeocin selection during a prolonged period of time.

The novel test was carried out for both positive and negative controls by picking single colonies from fresh KJ-Zeo-pos and KJ-Zeo-neg transformants and by directly inoculating 4 ml low-salt LB (amp-kan-cam-I) containing 25 μg/ml zeocin and incubating overnight. Unfortunately, the pH was shown (by use of pH paper indicators, Whatman narrow range indicator) to change

during the night with an acidification of the medium. This pH variation probably reduced the effectiveness of zeocin and, although the negative control didn’t grow during the first 16-18h (the overnight incubation), presence of bacteria growth was identified immediately after (20 to 24 h) by direct observation and confirmed by OD600 measurement. This result was considered unacceptable for a strong and reliable selection system and for this reason it was decided to move the selection step in solid medium.

Low-salt LB agar plates were shown, by use of pH paper indicators, to have a pH of ≈ 7 and to maintain it constant during a period of 48 h in presence of both the three antibiotics (ampicillin, kanamycin and chloramphenicol) and in presence of KJ-Zeo-pos and KJ-Zeo-neg.

A new set of experiments was carried out by picking single colonies from KJ-Zeo-pos and KJ-Zeo-neg transformants and starting overnight inoculi (in 4 ml low salt LB amp-kan-cam-I). The day after, 10^5 cells from these overnight cultures were plated on low-salt LB agar plates (amp-kan-cam) containing 25 µg/ml zeocin with or without IPTG (to test if the system was still under the control of IPTG).

Plates were incubated at 37°C for 16-18h and then the number of colonies was counted. This experiment was repeated 4 times in order to confirm reproducibility. The number of colonies for each experiment is shown in Table 3.2.

Table 3.2 Zeocin selection test on low-salt LB agar plates

	Sample	+IPTG	-IPTG	[d] Age of transformants
[a] Experiment 1	[b] Positive control	143	25	1 day
	[c] Negative control	0	0	
[a] Experiment 2	[b] Positive control	100	20	1 day
	[c] Negative control	0	0	
[a] Experiment 3	[b] Positive control	>100	>100	6 days
	[c] Negative control	>100	>100	
[a] Experiment 4	[b] Positive control	30	11	2 days
	[c] Negative control	0	1	

[a] Positive and negative controls were grown at 37°C for 16-18h on low-salt LB agar (amp-kan-cam), pH 7 in presence 25 μg/ml zeocin with or without IPTG.
b KJ-Zeo-pos.
c KJ-Zeo-neg.
d This column reports the age of the KJ-Zeo-pos and KJ-Zeo-neg transformants at the time when the four different selection experiments were carried out.

The data obtained from experiment 1, 2 and 4 showed how the selectable system on a solid-state medium was able to discriminate between positive and negative control and between presence and absence of IPTG, however, the number of colonies, for the positive control, was extremely low compared to the number expected from a plating of 10^5 cells probably suggesting low levels of activation of the zeocin resistance cassette by the ZF1-R domain.

Furthermore, unclear results were found in experiment 3 where six days old KJ-Zeo-pos and KJ-Zeo-neg transformants were used (measured as the time elapsed between transformation of KJBAC1 cells with the three pGP-FF-ZF1, pAC-KAN-αGal4 and pBAC-TS-1-ZeoR plasmids and the carrying out of the zeocin selection experiment). Colonies were obtained on all plates without any distinction between positive or negative control and presence/absence of IPTG and the number of colonies was higher than in the other experiments.

These results probably suggest that zeocin resistant cells developed with time in a manner that does not require IPTG indicating that, perhaps, DNA

rearrangements of the zeocin resistance cassette were occurring conferring a constitutive resistance to the antibiotic.

Although the reasons are not clearly explained, the modular assembly protocol states that KJBAC1 cells transformed with the three plasmid pBAC-lacZ, pGP-FF and pAC-KAN-αGal4 are stable only for a week, after which, the cells grow poorly or not at all (Wright, Thibodeau-Beganny et al. 2006). Although experiment 3 showed the opposite outcome, it was decided to check further the importance of time elapsed between the transformation and the starting of the zeocin selection of each single experiment (Table 3.2).

In order to confirm a possible connection between the age of the cultures and the acquired zeocin resistance, one-week-old KJ-Zeo-pos and KJ-Zeo-neg transformants, created by transforming KJBAC1 cells with the three pGP-FF-ZF1, pAC-KAN-αGal4 and pBAC-TS-1-ZeoR plasmids and then storing the agar plates containing the transformants at 4°C for 7 days, were used to repeat the experiment previously described in Table 3.2. The results obtained from this novel experiments mirrored the one found for experiment 3 (Table 3.2) and therefore confirming the initial observation that old transformants were all resistant to zeocin showing a possible constitutively expressed resistance cassette.

Although the selectable system proved to be more effective in a solid medium support than in a liquid one, in addition to the very low number of colonies detected on the positive controls for experiments 1, 2 and 4 (≈ 100 colonies compared to the 10^5 cells plated), a new reason for doubting the reliability of the whole process was identified by examining the shape and size of colonies obtained during zeocin selection. Colonies growing on the positive controls from the previous 4 experiments did not present homogeneous sizes, on the contrary, 3 different types of colonies were identified: small, medium and big. This could have reflected different expression of zeocin resistance gene in different clones, but when different colonies were picked, cultured overnight in 4 ml low-salt LB (amp-kan-cam-l) and then plated again on low-salt LB agar plates (amp-kan-cam-l) containing 25 μg/ml zeocin the same mixed population of small, medium and big colonies was detected on all plates without any distinction. It therefore seemed that the ability of clones to

grow in zeocin was variable during time and that single cells in the same “clone population” acquire different sensitivity to the antibiotic.

The testing of the zeocin selectable model identified a series of flaws that seriously reduced the efficiency of the selection system. The stability of the antibiotic plus the different sizes of the colonies and the actual inefficiency of the selection (with a very small number of zeocin resistant colonies for 10^5 plated cells in the positive control) pointed towards a possible genomic instability of the KJBAC1 strain, to plasmid incompatibility problems or to some inefficiency of the zeocin resistance cassette and were considered enough to decide to discard the zeocin selection in favour of a more reliable selectable marker.

OPEN selection approach: At the same time that the zeocin selectable system was developed, an alternative B2H selection system was designed and published by the same group that created the modular assembly protocol (Maeder, Thibodeau-Beganny et al. 2009). Although it was designed to screen large libraries of ZF domains produced through a phage expression technology, the system employed an attractive selection strategy, where the resistance to streptomycin and the rescue of the histidine pathway confer a powerful double selection, that was considered as a potential improvement on the zeocin selection used in the previous section.

This new selectable system is based on a modified B2H system were the pBAC-lacZ plasmid is substituted by an F’ episome containing the possible ZF target site close to a HIS3 and aadA (histidine pathway gene and streptomycin resistance, respectively) cassette (Fig. 3.4). As in the standard B2H assay, pGP-FF and pAC-αGal4 (an analogue of pAC-KAN-αGal4 that differs from it for having a chloramphenicol resistance instead of kanamycin resistance) plasmids are co-transformed in cells containing the F’ selection episome: co-expression of the pGP-FF and pAC-αGal4 plasmids, followed by the effective binding of the ZF domain and recruitment of the αGal4-RNA polymerase fusion protein, leads to the transcription of the selectable markers conferring resistance to streptomycin and, at the same time, rescuing the histidine pathway.

In order to assess the efficiency of this model, the same pGP-FF-ZF1-R and pGP-FF-ZF1-L were used as positive and negative controls, respectively. The first step involved the creation of the two F’ episomes carrying the target sequences for both positive and negative controls. TS-1(R) and TS-1(L) sequences were initially cloned in the pKJ1712 plasmid close to the HIS3 and aadA cassette creating pKJ1712-TS-1(R) and pKJ1712-TS-1(L). This region of the plasmid is flanked by a lacIq and lacZ genes that are homologues to the lacIq and LacZ sequences found on the F’ episome harboured by the CSH100 bacteria strain (Fig. 3.5 a).

Transformation of the CSH100 strain with the pKJ1712 plasmid allows double crossover events to occur between the homology regions of both pKJ1712 and the F’ episome and, therefore, the transfer of the target sequence HIS3 and aadA cassette (Fig. 3.5 a). As a final step, the mating of the newly created CSH100 cells with the KJ1C (F-) bacterial strain, through conjugation, permitted the transfer of the episome and the creation of the final KJ1C selectable strain. Presence of F’ episomes that underwent effective double-crossover events was ensured by double selection with kanamycin and sucrose. The presence of the KanR cassette would confer resistance only to cells harbouring the newly created HIS3-aadA cassette while negative selection through the sacB gene would exclude any F’ episome that underwent only one crossover event with the pKJ1712 plasmid (Fig. 3.5 b).

The positive control was created co-transforming pGP-FF-ZF1-R (ampR) and pAC-αGal4 (CamR) in the KJ1C strain harbouring the TS-1(R) F’ episome (KanR) and the negative control was created co-transforming pGP-FF-ZF1-L (ampR) and pAC-αGal4 (CamR) in the KJ1C strain harboring the TS-1(L) F’ episome (KanR).

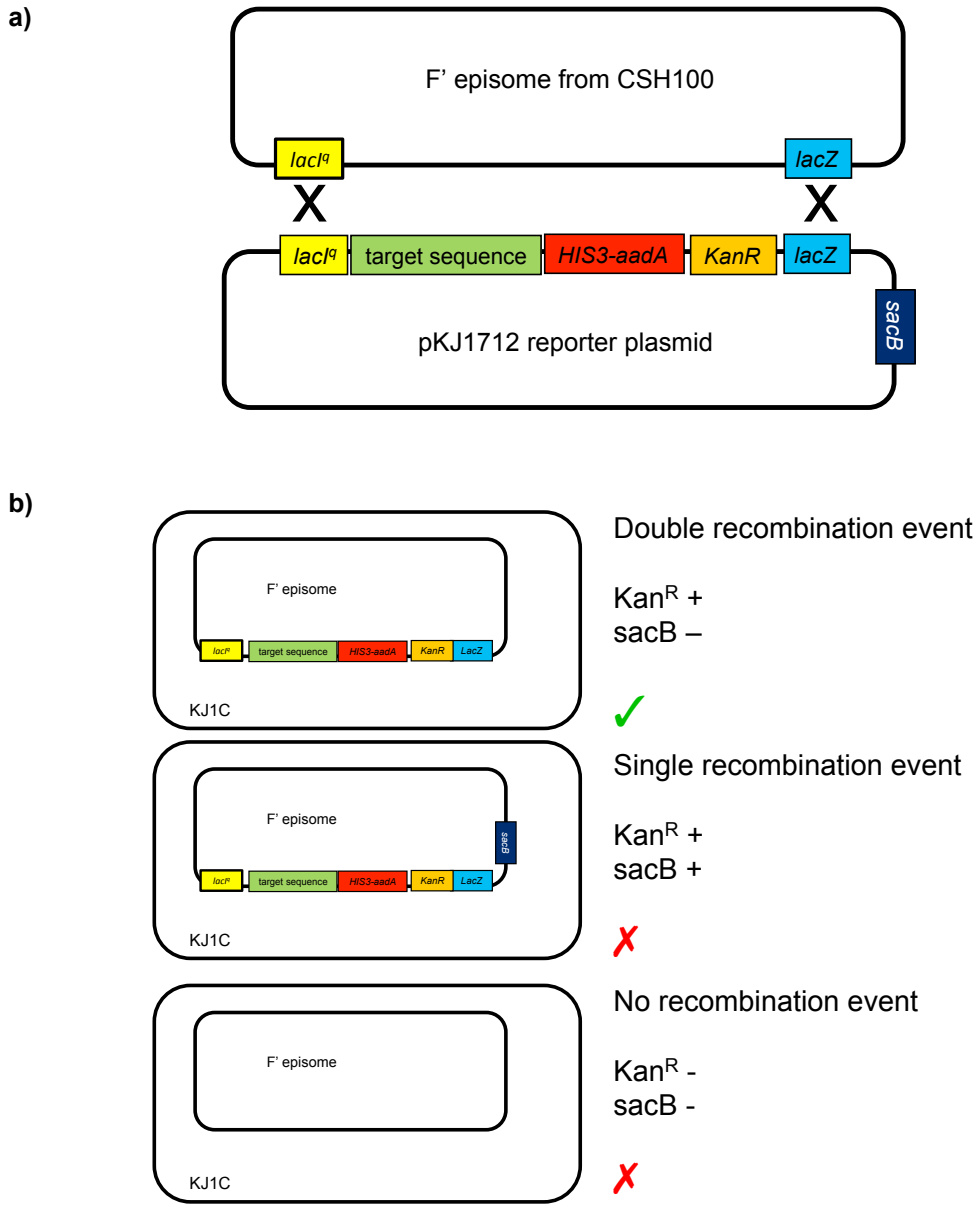


Figure 3.5 OPEN selection system.
a) Creation of the F’ selectable episome. Double cross-over event between the identical $lacI^q$ and LacZ present on both the pKJ1712 plasmid and F’ episome creates the final selectable episome carrying the target sequence and the two selectable markers HIS3-aadA. Kan^R selectable marker is used to positively select for recombinant in KJ1C bacterial strain after conjugation, sacB instead is used as negative selection for single recombinant. **b)** KJ1C bacterial strain after conjugation with CHS100. The mating between the two bacterial strains permits the transfer of the F’ selectable episome to the KJ1C cell and creation of the final selectable bacterial strain. Conjugated KJ1C cells are then selected in kanamycin and sucrose for the acquisition of a double recombinant F’ episome. Only KJ1C carrying a double recombinant episome will be able to grow thanks to the acquired Kan^R and the absence of the counter-selectable marker sacB that kills cells in presence of sucrose.

The reliability of the system was assessed through a series of preliminary tests in order to identify possible problems like the one encountered during the zeocin approach. Positive and negative controls were picked up as single colonies from fresh transformants plates and were grown overnight in a 4 ml liquid culture in presence of ampicillin, kanamycin, chloramphenicol and IPTG in order to start to express the ZF domains. The day after ≈ 10^5 cells were plated on minimal medium agar plates in presence of 20 μg/ml streptomycin and incubated at 37°C (minimal medium was used in order to enforce the HIS3 selection step). The OPEN protocol suggests carrying out selection on minimal plates for 48h. After 18h it was not possible to identify any kind of growth on the positive and negative controls but after 48h, instead, ≈ 200 colonies of variable sizes started to form on the positive control. Unfortunately, it was immediately evident that also this new approach suffered of selection problems: some kind of bacterial growth was detected also on the negative control in the form of a very thin layer of bacteria. This thin layer was not growing enough to form colonies but was still discernible from samples plated with ≈ 10^5 cells of untransformed KJ1C where no kind of growth was detected.

In order to assess the contribution of the levels of antibiotic during the selection, standard LB agar plates with different concentrations of streptomycin (between 5 μg/ml and 40 μg/ml) were used. Positive and negative controls were picked up as single colonies from fresh transformants plates and were grown overnight in a 4 ml liquid culture in presence of ampicillin, kanamycin, chloramphenicol and of IPTG. The day after ≈ 10^5 cells were plated and then incubated for 24h. Also in this case the results obtained were unclear: a small number of colonies (≈ 100) were detected on both negative and positive controls for antibiotic concentrations below 15 μg/ml and no colonies at all were detected for concentrations above 20 μg/ml. The OPEN protocol suggests to proceed with two steps of selection: the first with the expression of the ZF-DBD plus the αGal4-RNA polymerase driven by IPTG and the second, more stringent, without IPTG induction. For this reason, the selection test described above was repeated using a streptomycin concentration of 20 μg/ml in absence of IPTG induction. In this case a small

number of small colonies (≈ 20) were detected after 48h on both positive and negative controls without distinction.

At this point it was decided to rule out possible effects due to the presence of IPTG. Cells from both fresh positive and negative controls transformants were grown overnight in a 4 ml liquid culture in presence of ampicillin, kanamycin, chloramphenicol but in absence of IPTG. The day after ≈ 10^5 were plated on standard LB agar plates (with ampicillin, kanamycin, chloramphenicol) in presence or absence of IPTG and without streptomycin. While cells from negative control were growing on both ± IPTG plates (more than 500 colonies for each plate), cells from the positive control were growing in absence of IPTG (more than 500 colonies) but, in presence of IPTG, showed a highly reduced growth rate (≈ 100 colonies). This last experiment highlighted a possible toxic effect probably due to the expression and binding of the ZF domain to its target sequence.

Due to the ongoing problems that occurred during the creation of the two selectable systems, it was decided to seek an alternative approach to obtain functional ZFNs specific for the β-globin gene. After raising enough funds, it was possible to take advantage of a commercially available platform offered by Sigma-Aldrich that, during that period of time, had acquired the patent from Sangamo Bioscience to use their technologies to commercially produce ZFNs under the name of CompoZr. Although the service provided by Sigma is highly expensive (£20,000 at the time of the purchase), the company supplies, in a short amount of time, functional ZFNs designed for any desired gene and that have been already validated in a mammalian model. For this reason and also for the increasing amount of time required to optimize the selectable systems, it was decided to indefinitely postpone the development of the directed evolution approach and concentrate the effort towards the characterization of the newly acquired ZFN.

3.3 Discussion

3.3.1 Identification of target sequences in the human β-globin gene

A traditional induced gene correction is achieved thanks to homologous recombination between the desired targeted gene and a DNA donor template that carries the desired mutation (or correction) to be introduced, flanked by two regions homologues to the target sequence. Typically, only a small tract of the gene is corrected and, if the genetic disease to be cured can be caused by any one of possible mutations localized in the gene (as is the case for β-thalassaemia), multiple ZFNs and donor templates have to be created (one for each mutation or cluster of mutations). An alternative to this approach is to introduce, through targeted gene insertion into a suitable location, an expression cassette for the therapeutic gene. This approach may present some problems in the regulation of protein levels and also because the defective endogenous protein is still expressed. The fine regulation of expression of β-globin protein and the balance with the α-globin are of paramount importance for the effective development of functional hemoglobin molecules. For this reason it was decided to use the first type of approach and to introduce corrections in the endogenous gene in order to restore the natural levels of β-globin protein.

The choice of the nuclease (ZFN) target sequences is a critical step because the efficiency of correcting a specific mutation by HR decreases with increasing distance between the DSB produced by the ZFN and the location of the mutation. β-thalassemia is caused by more than 200 point mutations and deletions that are located along the β-globin gene not evenly but clustered mostly in the region of the 3 exons. Of the 50 possible target sequences, identified using the ZiFiT software, three were chosen for their specific position in the β-globin gene. It has been shown that, although the frequency of gene correction decreases with an increase of the distance

between mutation and DSB, gene correction of mutations as far as 400 bp upstream or downstream of the DSB produced by a ZFN can be corrected (Porteus 2006). The three target sequences chosen are actually no further away than 200 bp from all known β-thalassemia mutations and therefore are potentially useful for a possible therapeutic approach.

TS-1 and TS-2 are located in the first exon and first intron respectively close to a know cluster of β-thalassemia mutations and, in the case of TS-1, also close to the sickle cell mutation (allowing the possible correction of two different diseases). TS-3, instead, is found at the end of the second intron, close to a second group of mutations located on the third exon.

A further consideration that affected the selection of the target sequences was the risk of possible off-target cleavages. The β-globin gene shares a strong homology to the other family of the human globin genes and for this reason there is a high risk that poorly designed ZFNs could produce DSBs also in other region of the chromosome, inducing a higher rate of cell death after transfection and, more importantly, the risk of chromosome rearrangements and oncogenesis following DNA damage. In order to assess this problem, the chosen target sequences were analyzed, using the BLAST algorithm, for possible homologies or partial homologies with other regions of the human genome. Although a BLAST query is able to detect regions of partial homology, the algorithm was not specifically developed to identify possible ZFNs off-target sites and therefore could miss potential critical sites. An alternative to BLAST queries was developed in recent years (2011) in the form of the web-based software ZFN-site (Cradick, Ambrosini et al. 2011). This program has an improved algorithm that helps in the search for ZFN cleavage sites and at the same time identifies a wider range of possible off-target cleavage using “degenerate specificities” that take in account ambiguities in the binding activity of each ZFN subunit.

3.3.2 Modular assembly of DNA encoding ZF-DBDs for TS-1, TS-2 and TS-3.

The process of assembling ZF domains, although being based on simple steps involving standard cloning techniques, is time consuming and, due to the many cloning steps involved, can be affected by errors especially during the enzymatic digestion and ligation steps. For this reason each cloning step had to be checked for possible inaccuracy by sequencing and, when detected, errors had to be corrected by stepping back to the previous cloning step increasing the amount of time spent in the creation of the ZF-DBDs. If it is considered that, in order to employ the ZFNs in a mammalian model, the ZF domains have to be produced by time consuming cloning techniques and have to be tested in a bacterial two hybrid assay (that has itself to be created) the time spent in the process is substantial, especially considering that 6 different ZF domains (2 for each target sequence) had to be created. For this reason it was decided that, although more expensive, two of the ZF domains (ZF2-L and ZF3-L) would be produced by a company effectively reducing the time employed in creating all ZF domains and reducing also the risk of introducing errors in the final ZF domain expression plasmids. On the other hand, the kit provided by Addgene has a complete library of ZF modules that can be used to create multiple binding domains specific for almost any possible target sequence and therefore, after an initial period of optimisation of the protocol, a research group could be able, in theory, to create a wide range of ZFNs without the need of external support or new materials, reducing greatly the costs.

3.3.3 Assessing ZF-DBD binding activity in the B2H assay

The modular assembly protocol was chosen for this project mainly for its affordability and (perceived) simplicity in creating effective ZF domains. The choice was supported even more by the, supposedly, high efficiency in creating effective ZFNs highlighted in two different papers (Bae, Kwon et al. 2003; Mandell and Barbas 2006). Unfortunately, both assessments largely overestimated the real efficiency of the approach due to limited sampling of all possible ZF modules. Although these two studies examined the binding activity of a wide number of ZF domains made of 3 modules (specific for a 9 bp target sequence), most of the domains were composed by at least 2 modules that would recognize a GNN (where NN are any bases) sub-site. A later paper, which was published concurrently with the development of this project, assessed the discrepancy between the expected efficiency and the actual higher failure rate (Ramirez, Foley et al. 2008). Ramirez et al examined 168 different ZF domains capable of binding also target sites containing 1 or no GNN sub-sites (that are the majority of all possible target sequences made of 9 bp). The study highlighted how ZF domains, created through the modular assembly approach and that would bind to target sequences containing no GNN sub-sites, would be highly ineffective reducing the overall efficiency of the approach to a mere 25%.

These results were unintentionally confirmed by the assessment of the newly created ZF domains specific for the β-globin gene, where only one showed a satisfactory binding efficiency. Considering that a ZFN in order to produce a DSB has to be composed of two different ZF domains, the overall efficiency in producing effective ZFNs will be reduced to 6%. Even these low estimate, which is based on the B2H assay, may be misleading because in mammalian cells it could be even lower due to difference of binding activity of ZF domains in bacterial and mammalian cells (Cathomen and Joung 2008).

A second factor that almost certainly contributes to poor binding activity of ZF domains, may be the inability of the modular assembly approach to take account of the so called “context dependent effect” (explained in more details

in Section 1.10) that involves the interaction of neighbouring ZF modules with the final outcome of a general reduction of the binding efficiency of the domain (Cathomen and Joung 2008).

3.3.4 Development of a directed evolution system

In order to assess the problems identified in the modular assembly protocol, members of the Zinc Finger Consortium developed in 2009 a completely different approach for the creation of new ZFNs: the OPEN system (Maeder, Thibodeau-Beganny et al. 2009). This protocol still has the advantage of being completely open-source and affordable but at the same time has a higher success rate due to a totally different approach in selecting effective ZF domains. While the modular assembly protocol considered each single ZF module more or less as an independent unit when binding to the target sequence, the OPEN approach takes in account the possible interactions between neighbouring modules and the effect on their binding efficiency.

Although this newly developed approach promised to be a reliable alternative to the modular assembly protocol, when the ZiFiT software (that has been adapted to be used in conjunction with this new protocol) was used to identify possible target sequences in the β-globin gene, it was not able to identify any. The reason for this was that the ZF libraries to be used were still under development and the choice of possible ZF modules was still very limited. Furthermore, the techniques involved in the OPEN protocol are time consuming and labour-intensive requiring the creation of phage-based combinatorial ZF libraries that would take expertise and time to set up.

Instead, the approach that was taken, which was conceived before the OPEN protocol was published, did not involve small ZF libraries but took advantage of the already created ZF domains that would not be discarded but instead would be improved through a directed evolution system. The idea was based on the combination of two techniques: the creation of pools of randomly mutated ZF-DBDs, based on the ones that were already produced by the modular assembly approach, and a selectable system where effective binding would confer a survival advantage to the cell expressing it. The latter was a

simple extension of the B2H system replacing the lacZ cassette with a gene conferring an antibiotic resistance. This would have the advantage of reducing the time spent in generating new ZF domains and could overcome the problematic of context dependent effects. In the case of this project this approach was potentially attractive because one of the six ZF domains already showed a very high binding efficiency (ZF1-R) and the possibility of improving the corresponding domain (ZF1-L) would have lead to the quick and effective creation of a functional ZFN.

The creation of mutated ZF libraries is a method that has already been used with success in order to produce the wide range of ZF modules capable of targeting all possible 64 trinucleotides. Binding specificity of ZF proteins is determined by a very small number of amino acid residues (see Section 1.9) located in the N-terminus of the ZF α-helix and the modification of these key residues has been shown to greatly affect binding activities (Rebar and Pabo 1994; Segal, Dreier et al. 1999).

The selectable system developed was based on zeocin selection and, although it was able to discriminate between positive and negative controls, it presented a series of problems that cast doubts on the usefulness of the system. Although low-salt LB, optimized pH and solid media were used in order to increase the effectiveness of the zeocin selection, the results obtained during the preliminary tests where unsatisfactory due to the lack of stability of the antibiotic at different pH values, to the variable response of the bacteria cells to the antibiotic (identified by different sizes of colonies) and due to a low growth also from the ZF1-R-DBD positive control (that, by contrast, responded incredibly well in the standard B2H assay).

A deeper analysis of the problems would be required in order to understand clearly all the factors involved in the poor efficiency of the selectable system but a possible explanation could be the presence, in the bacterial strain, of multiple plasmids carrying different antibiotic resistances (to ampicillin, kanamycin, chloramphenicol and zeocin) and a possible genetic re-arrangement between this multiple DNA molecules. A possible sign of genetic instability is the presence on plates of colonies of different sizes and their ability to form also from one single clone. A possible internal re-arrangement could be also explained through the mechanism of action of zeocin antibiotic

that is a copper-chelated glycopeptide that, after activation inside the cell through the removal of the copper ion, binds and cleaves the DNA producing extended DNA damage. If the antibiotic, for any reason, is not able to kill immediately the cells, it could instead promote a possible genetic re-arrangement between the different plasmids present in the cell.

For these reasons the zeocin approach was discarded in favour of a more complex, and possibly more robust, selection system based on the one already developed by the Zinc Finger Consortium group for the OPEN protocol (Maeder, Thibodeau-Beganny et al. 2009). This approach also resembles the original B2H system but it relies on the action of two selectable markers combined together.

Although this double selection has been shown to be effective for the OPEN protocol, it was clear that, in the context of the directed evolution approach, it presented similar difficulties to the zeocin system. The problems involved in the double selection, the presence of a partial bacterial growth on the negative control and the unclear effect of IPTG on the positive control are factors that have to be addressed in order to create a robust selection method. From the data collected, it seems that expression of the ZF domain and its binding to the target sequence were connected to a decrease in bacterial growth, probably due to toxic effects related to the expression of the selectable markers or expression of the two fusion proteins expressed by pGP-FF and pAC-αGal4.

The difficulties experienced during the development of the two directed evolution approaches and the increasing amount of time spent were considered excessive and a different system was sought in order to obtain functional ZFNs. The arrival of new fundings and, at the same time, the creation of a commercially available platform by Sigma for the creation and testing of novel ZFNs were two decisive factors in the decision to set aside the directed evolution approach and concentrate the effort not in the actual creation of the ZFNs proteins but in the development and characterization of a feasible technique to introduce gene corrections at the level of the β-globin gene.

4 Chapter 4: Assessing the ability of commercially sourced zinc finger nuclease to cleave the human β-globin gene

4.1 Overview

The process of producing and testing ZF-DBDs using the modular assembly approach for the creation of novel ZFNs specific for the β-globin gene has been time consuming and, more importantly, ineffective. Although the Zinc Finger Consortium groups have worked towards the creation of an improved protocol (Maeder, Thibodeau-Beganny et al. 2009) that would be more effective in the creation of functional ZFNs, the archive of novel ZF modules was still under development when this project was in need of an alternative approach for the creation of β-globin-specific ZFNs and therefore the acquisition of a commercially produced ZFN was chosen as an alternative. Commercial ZFNs are now produced and sold by Sigma-Aldrich using technologies and resources developed by the American biotech company Sangamo Bioscience.

Although during the 1990s various institutions and private companies have steadily filed patents based on ZF proteins technologies, Sangamo Bioscience has proved to be the more competitive and effective company in the field of ZF proteins and, since then, has acquired most of the licences like the one owned by Chandrasegaran for the fusion of DNA binding proteins to the cleavage domain of the *FokI* endonuclease (Scott 2005) or directly buying entire companies like Gendaq Ltd, a British biotech company founded by Aaron Klug and that had patents on methods for the creation of ZF binding proteins based on libraries of “two-zinc-finger“ domains (Chandrasekharan, Kumar et al. 2009).

Sangamo Bioscience is a biotech company founded in 1995 with the aim of developing and commercialising DNA binding proteins for regulation and modification of plant and animal genomes in the research and therapeutic

fields. Their strategy was based on the creation of a proprietary technology platform based on natural occurring ZF proteins: DNA binding proteins that are present, as transcriptional factors, in many different higher organisms. In the following years Sangamo tested its technology platform creating ZF proteins specific for novel DNA target sequences with the aim of creating two main molecular tools: ZF protein activators/repressors to be used as therapeutic gene regulators and ZFN to be used as gene-editing tools (Sangamo-Bioscience).

The Sangamo approach has proved to be highly successful and the developed technology has led to creation of a therapeutic program in advanced stage (a phase 2 trial for the treatment of HIV/AIDS), a therapeutic program for haemophilia (in a preclinical stage) and research programs in the treatment of cancer, Parkinson's disease, monogenic disease and neuropathic pain (Sangamo-Bioscience 2012).

Sangamo has established commercial partnership with different companies including, in 2007, a deal with Sigma-Aldrich Corporation for research purposes in order to develop Zinc-Finger based laboratory reagents (Sangamo-Bioscience 2007).

In 2008, Sigma-Aldrich, under the name CompoZr, started to commercialize the production of customised ZFNs to be used for production of gene knock-outs, gene corrections or gene additions (see Section 1.10). The custom design technology provided by the company allows the creation of ZFNs for specific target sequences following a 4 stages process. In a first step, the Sangamo proprietary algorithm (Super Finder Algorithm) is used to design a group of possible ZFNs specific for the desired genomic region. A selection of target candidates is then sent back to the custumer for approval. ZFN candidates are then assembled using ZF modules derived from Sangamo's archive and, in a final step, the candidates are tested and validated in a relevant cell line using a mismatch assay (Cel-I assay). The process typically takes 10-15 weeks, after which the best performing ZFN pair is sent to the costumer to be used for research purposes (Sigma-Aldrich).

Ordering functional ZFNs directly from a company such as Sigma is financially demanding but has a series of advantages: while the "DIY" approaches, such as the Modular Assembly or the OPEN systems, are

cheaper and can be used in theory to produce ZFNs specific for a great variety of possible target sequences, the techniques involved are time consuming and the process requires the creation of multiple ZFNs for the same gene due to the risk of low cutting efficiencies. These limiting steps are particularly onerous for small research groups with limited experience and a small workforce. In contrast, Sigma has a dedicated and experienced team that, in theory, in a short period of time is able to design, test and deliver functional ZFNs reducing the workload for the researcher. Two additional advantages make Sigma's approach a more valuable option: first of all Sigma uses for the creation of novel ZFNs the Sangamo's proprietary archive of "two-zinc-finger" domains that seems to work better than assembling single module, probably because its use minimises the context dependent effect (described in Section 1.10 and Chapter 3). A second important advantage is the type of engineered obligate heterodimer *FokI* domain that the company use and that was not included in the original Modular Assembly approach.

Creating DSBs at specific loci has opened the way to potentially engineer the genome of many different organisms but, unfortunately, expression of ZFNs can lead to high levels of cytotoxicity that are probably due to off-target cleavage events (Porteus and Baltimore 2003; Alwin, Gere et al. 2005; Beumer, Bhattacharyya et al. 2006). Off-target cleavage is a major concern in the development of a successful and safe gene-editing technique for two main reasons. First, increased cytotoxicity in treated cells reduces the number of viable cells and therefore the overall efficiency of gene editing. Second, unwanted ZFN-induced modifications at off-target loci increase the risk of mutagenesis and, in a clinical setting, oncogenesis (Cathomen and Joung 2008). The risk of off-target cleavage is greatly increased by the formation of ZFN homodimers (i.e. left-left or right-right ZFNs) that depend on the interaction between wild-type *FokI* subunits (Fig. 4.1 a). With the aim of reducing the risk of homodimer interactions, different groups have developed a variety of approaches to modify the protein structure of the *FokI* domain in order to create obligate heterodimer-ZFNs with reduced cytotoxicity (Miller, Holmes et al. 2007; Szczepek, Brondani et al. 2007; Doyon, Vo et al. 2011).

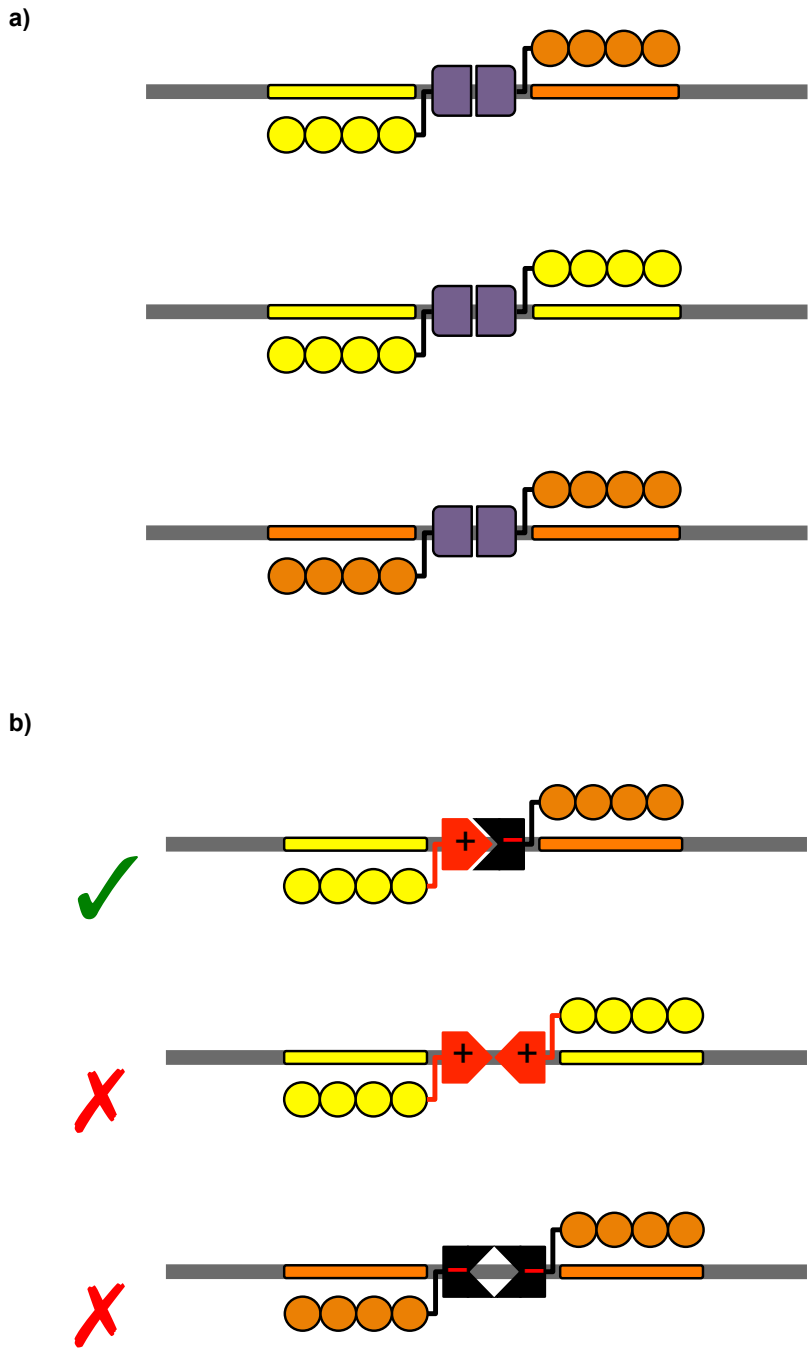


Figure 4.1 Homo- and heterodimerisation of FokI domains in ZFNs.
a) Possible dimers involving the normal of FokI domain: heterodimers (top) or homodimers (middle, bottom). Left and right ZFN domains (circles) and their target sequences (colored lines) are shown in yellow and orange, respectively. The homodimer FokI domain is colored in purple. **b)** Use of modified FokI domains limits heterodimer formation. Smbols are as in a) except the normal FokI domain are replaced by two modified domains one(ref) linked to ZFN-L the other (black) to ZFN-R. The + and – symbols indicate amino acid residues with opposite charges. With these modified FokI domains heterodimers (top)are stable but not homdimers (middle, bottom) are not.

Although the approaches were different, these three groups identified a set of key amino acids, located at the dimerisation interface, that when modified greatly affected the interaction of the two *FokI* subunits. These residues were changed in order to introduce charge-charge interactions where amino acids positively and negatively charged would only allow obligate heterodimer configurations eventually reducing the risk of off-target cleavage (Fig. 4.1 b).

During the development of ZFNs one of the first steps immediately after the synthesis of the novel nuclease is the assessment of its ability to produce DSBs at the desired target locus. Cutting efficiency of novel ZFNs is assessed using the Surveyor Mutation Detection kit (Transgenomic).

This technique is based on the use of the Cel-I nuclease, a specific mismatch endonuclease that is able to recognise and cleave mismatches in heteroduplexes containing single nucleotide polymorphisms (SNPs) or small deletions/insertions in one strand (Oleykowski, Bronson Mullins et al. 1998). In the absence of a DNA repair template, ZFN-induced DSBs are preferentially repaired by NHEJ but this pathway, due to its non-conservative nature, also tends to introduce deletions or insertions (indels) at the cutting site. These mutations can be used in the Surveyor assay to detect, and possibly quantify, the cutting activity of novel ZFNs. After initial expression of ZFNs in a cell population, genomic DNA is extracted and the region of interest (the ZFN target and surrounding sequences) is amplified by PCR reaction (Fig. 4.2 a and b). At this point the sample will contain a heterologous population of PCR products: some will be amplified from unmodified target loci, others from modified target loci with ZFN-induced indels. PCR products are next subjected to a melting and re-annealing step in order to promote the formation of heteroduplex DNA between unmodified and modified PCR products (Fig. 4.2 c). Finally, the samples are subjected to enzymatic Cel-I digestion and the digested products are then visualised by DNA PAGE electrophoresis (Fig. 4.2 d and e).

The following chapter describes the β-globin specifc ZFN that was commissioned from Sigma-Aldrich and the use of the Cel-I assay to assess its efficiency and specificity.

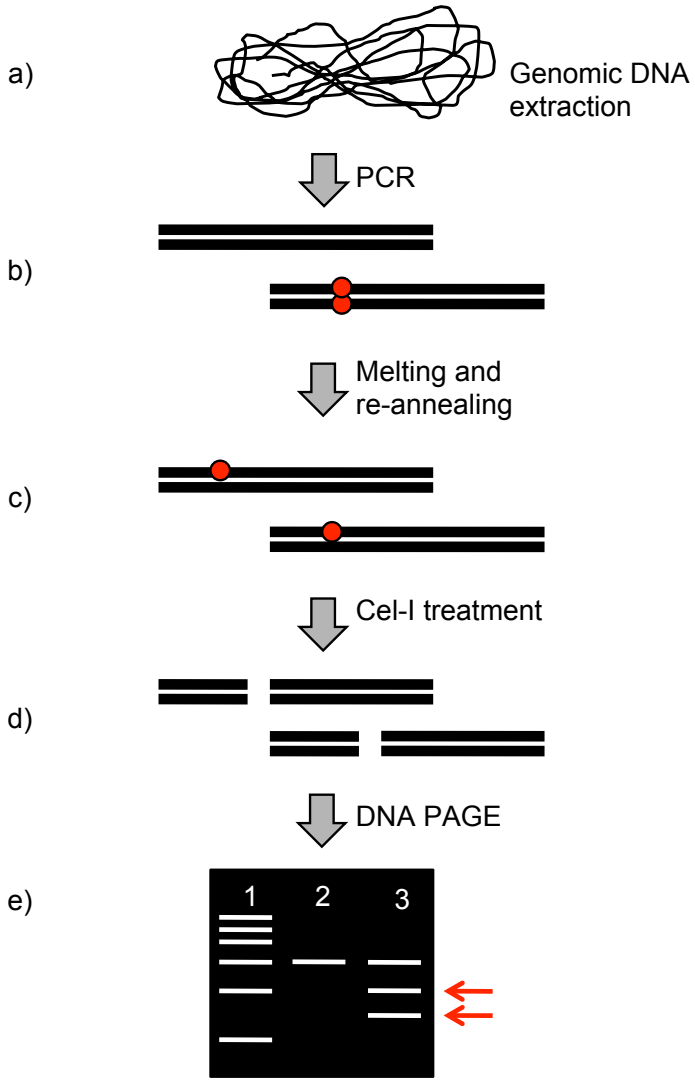


Figure 4.2 Cel-I assay schematics.
a) Genomic DNA extraction from ZFNs-expressing cells. **b)** PCR amplification of the region encompassing the ZFN target site. The PCR sample is composed by a mixed population of wild-type and mutated sequences (derived from un-cleaved and cleaved genes, respectively). A mutation generated by NHEJ-mediated repair of a ZFN-induced DSB is depicted as a red circle. **c)** Melting and re-annealing of the PCR products allows the formation of heteroduplexes from wild-type and mutated strands. **d)** The Cel-I nuclease recognises and cleaves these mismatch substrates but does not cut reformed homoduplexes. **e)** Digested samples are resolved by DNA PAGE as represented in this example. Lane 1) DNA ladder. Lane 2) Negative control derived from cells not treated with ZFNs: the only band is the full length homoduplex PCR product. Lane 3) Products derived from ZFNs-treated cells: a mix of uncut homoduplex (top band) and Cel-I-digested heteroduplex (arrowed bands) is seen, the latter indicating ZFN-induced cleavage at the desired locus.

4.2 Results

4.2.1 Choice of the target sequence, ZFN development and delivery

Initially, a specific region located in the first exon of the β-globin gene was chosen for analysis by Sigma's bio-informatics team in search of possible sequences that could be targeted by ZFNs designed with Sangamo's proprietary platform (Fig. 4.3 a and b). The region was selected for its proximity to both the sickle cell mutation and a cluster of known β-thalassemia mutations, and also for the presence of heterologies with the δ-globin gene (Fig. 4.3 b). This last aspect was of paramount importance to reduce to a minimum the risk of off-target cleavages that could occur at the δ-globin locus that has a high degree of homology with the β-globin gene. When the sequence was analysed with the SuperFinder algorithm, 26 possible ZF-DBDs were identified that could be combined in order to create 16 different ZFNs (Fig. 4.3 c). Sigma assembled expression plasmids for the 26 the candidate ZFN monomers using Sangamo's proprietary ZF-archive, and then transfected them in the 16 combinations in a suitable human cell line (K562). Using the mismatch-sensitive Cel-I assay, ZFN activity was estimated as the ability to produce DSBs and promote NHEJ at the target locus. The best ZFN (ZFN4), with an efficiency in promoting NHEJ estimated by Sigma at 10%, was supplied to the Gene Targeting group in the form of both expression plasmids and mRNA molecules (for the map and sequence of the ZFN4 expression plasmids see Appendix 2).

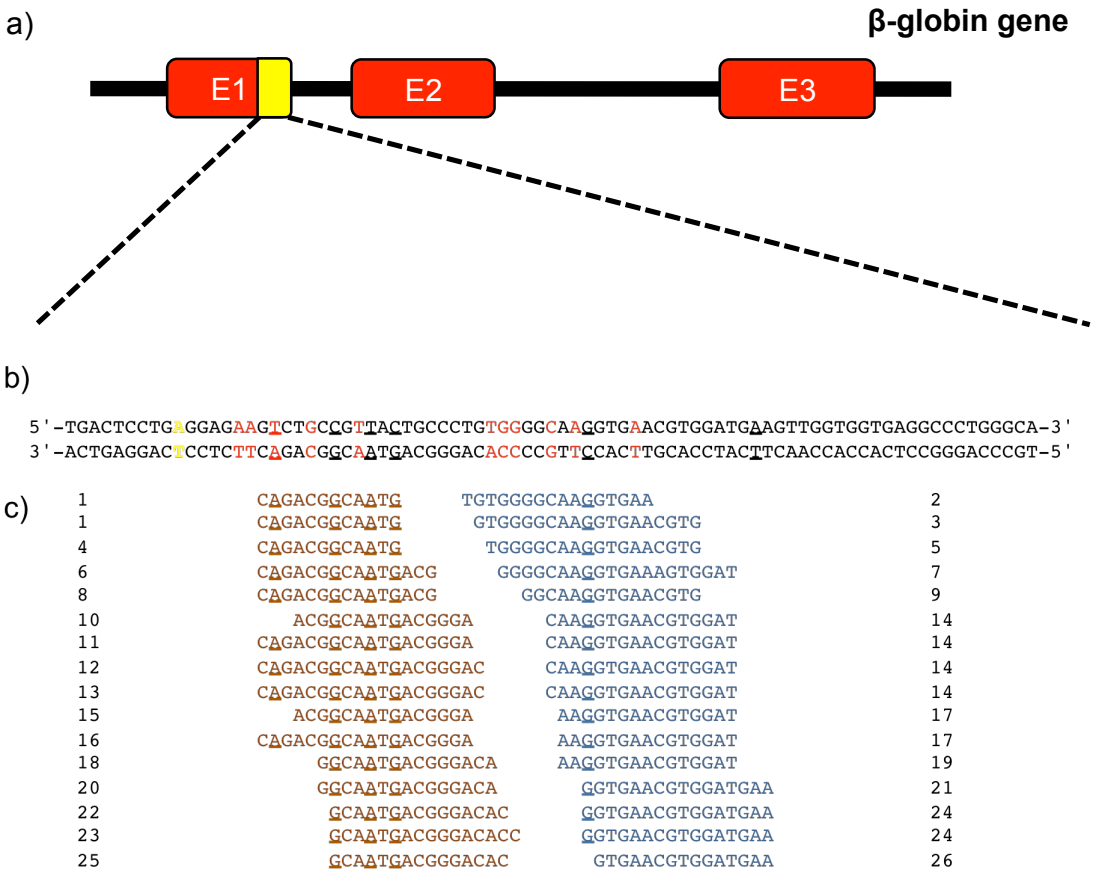
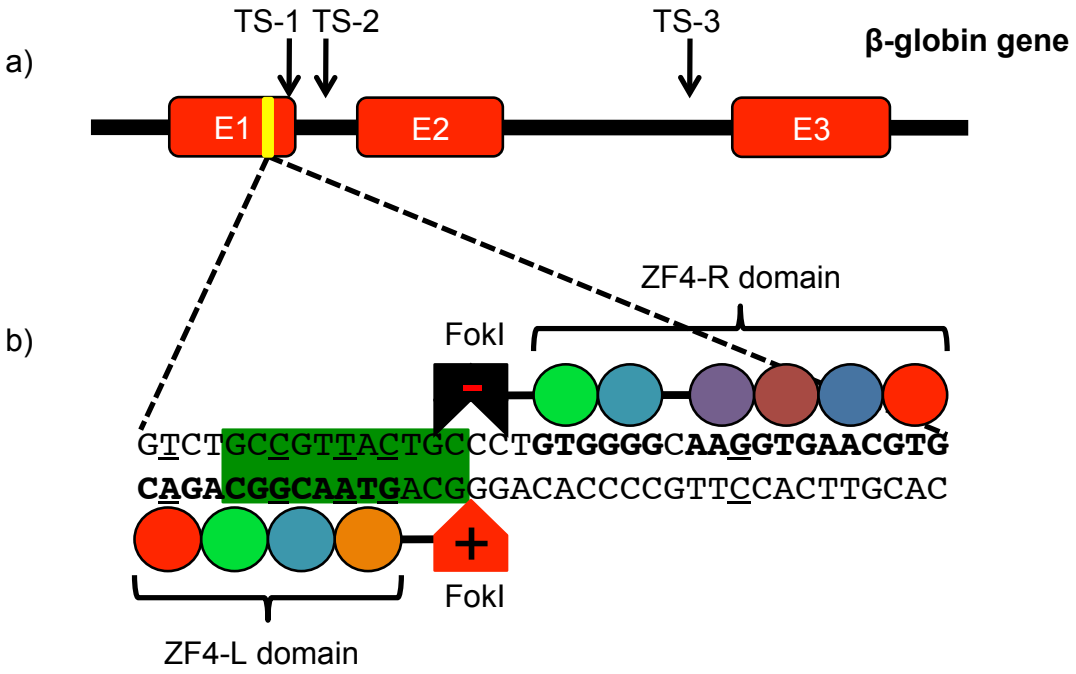


Figure 4.3 Potential β-globin ZFN target sequences identified by the SuperFinder algorithm.
a) Graphic representation of the β-globin gene with exons coloured in red. The region (81 bp) chosen for analysis by the SuperFinder algorithm is highlighted in yellow. **b)** The sequence examined by the Sigma bio-informatics team includes nucleotides mutated in sickle-cell disease (yellow) and β-thalassaemia (red). Nucleotides that differ from the equivalent region of the δ-globin sequence are underlined. **c)** The SuperFinder algorithm identified 26 different ZF-DBDs that combined together could give 16 different ZFNs. Left and right target sequences for each ZFN are displayed in brown and blue, respectively. Numbers to left and right of the sequences represent the possible different ZF-DBDs that could be created for that specific target sequences.

The newly created ZFN4 is composed by two different subunits: ZFN4-L made of 4 ZF modules and ZFN4-R made of 6 ZF modules. The engineered nuclease targets a 37 nucleotides sequence (from now on called TS-4) located 22 bp downstream of the sickle cell mutation. ZFN4-L binds to a 12 bp recognition sequence while ZFN4-R binds 18 bp. More specifically, the right binding site is made by a 12 bp sequence, that is recognised by the first 4 ZF modules of ZFN4-R, followed by a 1 bp spacer and then other 6 bp that are bound by the last 2 ZF modules (Fig. 4.4 b).

4.2.2 Transient expression of ZFN4 in human cells

In order to confirm the ability of ZFN4 to specifically edit the desired target sequence in the human β-globin locus, adherent HT1080 fibrosarcoma cells and suspension K562 myeloid leukemia cells were transiently transfected with expression plasmids or in vitro synthesised mRNA encoding for ZFN4-L and ZFN4-R. Cells were transfected with 2μg of each ZFN4 expression plasmid or 2μg of each mRNA. Lipofection and nucleofection were used to transfect $4x10^5$ HT1080 cells and $1x10^6$ K562 cells, respectively (see Section 2.6.7 and 2.6.8 for transfection protocols). After transfection, cells were incubated for 48h at 37°C in presence of 5% CO2 and afterwards genomic DNA was extracted with a commercial genomic isolation kit (Section 2.1.3). To detect a possible increase in ZFN4-activity during time, an additional experiment was carried out transfecting HT1080 cells with ZFN4-encoding plasmids and extracting genomic DNA after 72h instead of the customary 48h. Transfection efficiencies were assessed by transfecting HT1080 and K562 cells with 2 μg of pmaxGFP expression plasmid (Lonza). GFP expression was measured on FACS machine at 48h after transfection and the efficiency was expressed as the percentage of GFP positive cells among the total population (Section 2.6.9). Transfection efficiencies in both HT1080 and K562 experiments were estimated to be 80% or more (data not shown).



c)

ZF-DBD	Binding site
ZF4-L	3′-CAGACGGCAATG-5′
ZF4-R	3′-GTGCAAGTGGAA-GGGGTG-5′

Figure 4.4 ZFN4 overview.
a) β-globin gene representation with exons in red and the ZFN4 target site (TS-4) in yellow. The target sequences TS-1, TS-2 and TS-3 for the ZF domains generated with the modular assembly approach are also indicated (arrows). **b)** TS-4 sequence and binding of ZFN4. The ZFN4-L DBD is composed of four fingers and binds a 12 bp half site (in bold) while the ZFN4-R DBD has six fingers and binds a 18 bp half site (in bold). Heterology with the δ-globin gene is indicated by underlined nucleotides. The MwoI restriction site (GCNNNNNNNGC) used to confirm ZFN4's cleavage activity is highlighted in green (see Section 4.1.3 for the assay). The heterodimerising FokI domains (red/- and black/+) are represented as in Fig. 4.1. **c)** Summary of binding sites for ZFN4-DBDs. Sequences are shown on the 3'-5' orientation as this corresponds to the N-to-C terminal orientation of the ZFN amino-acid sequence.

4.2.3 Detection of ZFN4-induced gene modifications at the β-globin endogenous locus by the use of Surveyor assay

Due to ZFN4's activity and the following NHEJ events used to repair possible DSBs, the genome of each transfected cells should contain modified (mono- or bi-allelic modifications) or unmodified β-globin loci. In order to detect possible ZFN4-induced indels, a 394 bp region located in the β-globin locus and encompassing TS-4 was PCR-amplified using as a template the genomic DNA from transfected HT1080 and K562 cells (see Section 2.7.2). For each experiment, negative controls were carried out using as a PCR template the genomic DNA from HT1080 or K562 cells transfected with a control plasmid (pGP-FF). The primers used were provided by Sigma along with the ZFN4 expression vectors (for the complete sequence of the two primers see Section 2.7.2). The PCR products were subjected to a cycle of melting and reannealing in order to promote the formation of heteroduplexes between unmodified and modified PCR products. Afterwards, the mixture of homo and heteroduplexes was treated with the Cel-I nuclease and the digested products were visualised on a 8% polyacrylamide gel (DNA-PAGE) (see Section 2.3.2 for the DNA-PAGE protocol and Section 2.7.2 and Fig. 4.2 for the Cel-I assay protocol and schematic).

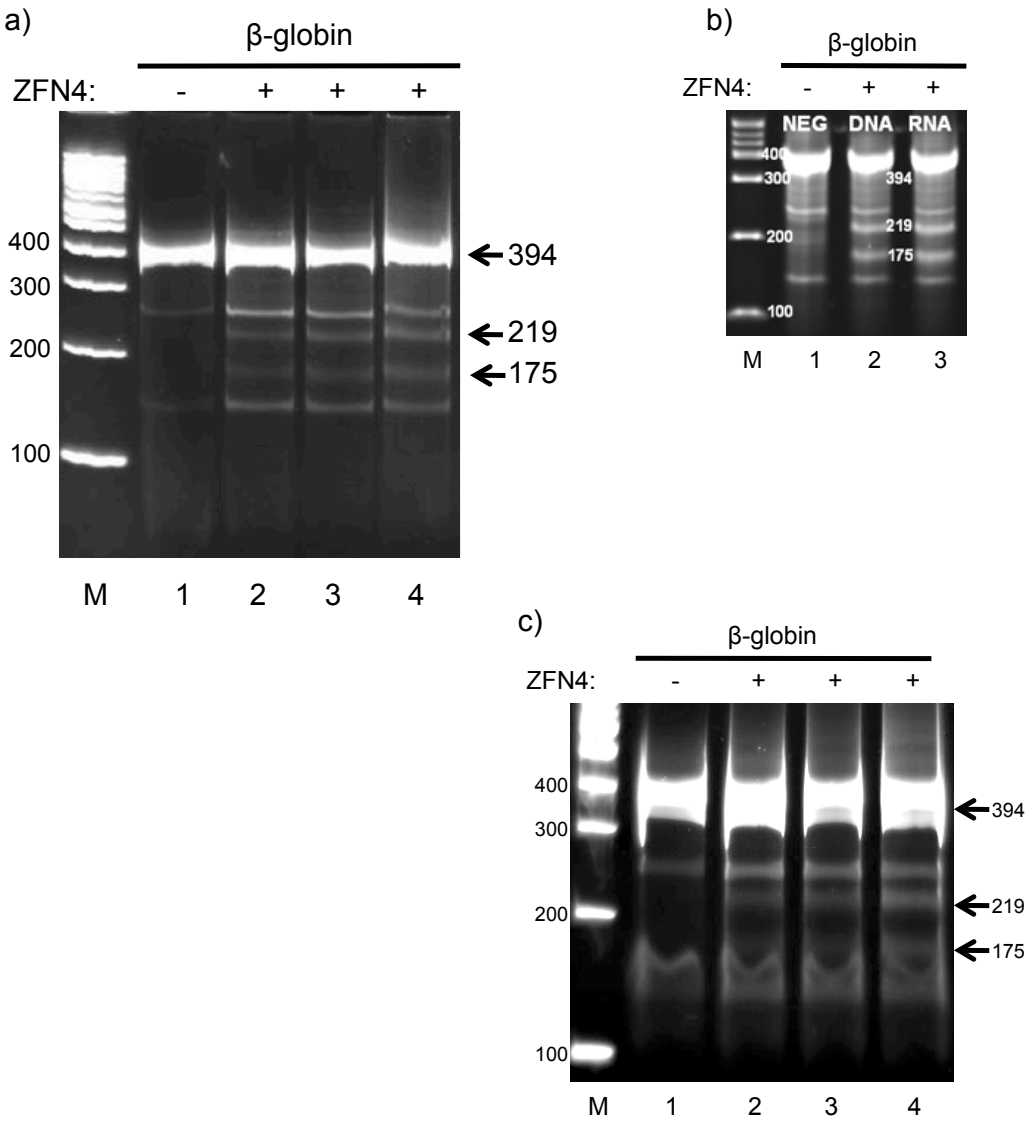


Figure 4.5 Cel-assays detect ZFN4-induced NHEJ events at the β-globin locus in human K562 and HT1080 cells.
a) Analysis of ZFN4 activity at the β-globin locus of K562 cells. Lane M is a 100 bp ladder (NEB). Cells were transfected with control DNA (lane 1), ZFN4 expression plasmids (lane 2) or ZFN4-encoding mRNA (lane 3 and 4). Genomic DNA for lanes 3 and 4 was obtained from two separated transfection. Genomic DNA isolated from cells 48h after transfection was used to amplify a 394 bp β-globin gene fragment carrying TS-4. All PCR products were digested with the Cel-I endonuclease before visualisation by DNA-PAGE (8% polyacrylamide gel). **b)** Analysis by the Sigma team (DNA-PAGE carried out on a 10% polyacrylamide gel). Lane 1: un-transfected K562 cells, K562 cells transfected with ZFN4 expression plasmids (lane 2) or with ZFN4-encoding mRNA (lane 3). **c)** Detection of ZFN4 activity at the β-globin locus in HT1080 cells. Cells were transfected with control DNA (lane 1), ZFN4 expression plasmids (lane 2 and 3) or ZFN4-encoding mRNA (lane 4). Genomic DNA was isolated 48h (lanes 1,2 and 4) or 72 h (lane 3) after transfection.

After the Cel-I treatment, the PCR sample should contain a mixed population made of uncut 394 bp fragments derived from homoduplexes and two distinct fragments of 219 and 175 bp derived from Cel-I-digested heteroduplexes. However, when the digested products were visualised on polyacrylamide gels, two additional bands of ≈ 150 and ≈ 240 bp were identified (Fig. 4.5 a and c). These bands were found also in negative control samples indicating the possible presence of a Single Nucleotide Polymorphism (SNP) carried by one of the two alleles in both K562 and HT1080 cells. A similar bands pattern was also identified by the independent testing conducted by Sigma's team (Fig. 4.5 b). The possible presence of SNPs in the exon 1 was confirmed by examining the map of the human β-globin gene found in the web archive of the NCBI website (NCBI Gene ID: 3043 (NCBI)). In addition to a graphic representation of the exons, to the complete sequence of the gene and to a list of clinically relevant mutations, the map also lists all the known SNPs with their relative position in the gene. The NCBI archive indicated the presence of a wide family of SNPs located in the region of the β-globin gene's exon 1. Although the presence of too many SNPs in the genomic region that has been PCR-amplified for the Cel-I assay prevented the possibility of establishing the correct SNP responsible for the digestion pattern noticed in the assay, the possible link between the additional bands in the assay and the presence of a SNP was partially confirmed.

In addition to these constitutive ZFN4-independent products, ZFN4-dependent Cel-I products of the expected sizes (219 bp and 175 bp) were detected. No substantial difference in the amount of these products was noticed between cell lines (HT1080 and K562) or between modes of ZFN4 expression (plasmid and mRNA forms) (Fig. 4.5 a and c). Moreover, no increase in ZFN4 activity (i.e. intensity of Cel-I products) was noticed in genomic DNA obtained from transfected HT1080 cells that were incubated for 72h before genomic DNA isolation (Fig. 4.5 c, lane 3).

Frequencies of heteroduplex formation (used as an evaluation of ZFN4 activity) were estimated by comparing the relative intensity of the digested products with known amounts of 100 bp ladder (NEB) (details in Section 2.7.2). The frequency was estimated to be ≈ 1-2% in both HT1080 and K562

cells, much lower than the estimate of 10% made by the Sigma's team. The assays described in this thesis were carried out employing the same experimental conditions as the ones used in the Sigma assay and transfection efficiencies were consistently high during all experiments, therefore the reason for this discrepancy is unclear. It could be ascribed to possible differences in the two batches of K562 cells used or to a suboptimal performance of the Cel-I assay in detecting and cleaving heteroduplexes. It is noticeable that, in the analysis by Sigma (Fig 4.5 b, lanes 2 and 3), the intensity of the ZFN4-dependent Cel-I products is clearly greater than that of the constitutive products, whereas this is not the case in the analysis carried out in this project (Fig 4.5 a, lanes 2-4). This observation argues against variations between the two analyses in the efficiency of Cel-I digestion, because reduced Cel-I activity should reduce the intensities of all Cel-I so that the ratio of intensities between constitutive and ZFN4-induced products should remain constant. Differences between K562 cells used in the two different analyses therefore seems the more likely explanation for the different estimates of ZFN4-mediated cleavage. ZFNs cutting efficiencies may vary greatly from one cell line to another due to differences in the chromatin structure and accessibility of the target sequence (Urnov, Rebar et al. 2010; Carroll 2011). Although the cell line used in both experiments was nominally the same, K562 is known to be particularly prone to genomic rearrangements that could create a great level of heterology between different laboratory stocks of K562 (Dimery, Ross et al. 1983; Wu, Voelkerding et al. 1995). Sigma's team, when consulted about the discrepancy between the two Cel-I assay results, also suggested a reduced ZFN4 activity due to possible differences in the two K562 cell lines. To rule out possible contamination of the K562 cell population with other cell lines or possible substantial genomic rearrangements that could have affected the frequency of ZFN4-induced DSBs, 107 K562 cells were handed over to the Leukaemia Cytogenetics Unit at Hammersmith Hospital (London) for Fluorescence In Situ Hybridisation (FISH) by use of a BRC/ABL dual-colour probe (Kreatech Diagnostic) and for karyotyping analysis (G-banded chromosome analysis). The two analyses excluded a possible contamination of the K562 cell population by other cell types and the few chromosomal abnormalities identified were subjected to

cell-to-cell variation and were largely consistent with the published karyotype of K562 cells (Naumann, Reutzel et al. 2001).

4.2.4 Alternative detection of gene modification by MwoI restriction digestion

Thanks to an MwoI restriction site located within TS-4 (Fig. 4.4 b), it was possible to independently confirm ZFN4's cleavage activity by an alternative approach to the Cel-I assay. The 394 bp TS-4 region of the β-globin gene was PCR-amplified using the same primers used in the Cel-I assays and, as a template, the same genomic DNA from K562 cells transfected with ZFN4 expression plasmids or mRNA used in the Cel-I assay previously described. Theoretically, all of the 394 bp PCR product from untransfected cells should be digested by MwoI to form two fragments of 164 bp and 230 bp. By contrast a proportion of the 394 bp PCR products derived from ZFN4-treated cells should be resistant to MwoI digestion due to the loss of the MwoI site resulting from NHEJ-mediated repair of ZFN4-induced DSBs.

MwoI restriction digestion was carried out on PCR products amplified from the same genomic DNA used for the Cel-I assay (see Fig. 4.5 a). Afterwards, digestion products were visualised by DNA-PAGE on an 8% polyacrylamide gel (for the full description of the MwoI experiment see Section 2.7.3). MwoI restriction digestion patterns (Fig. 4.6 a) were compared with the results obtained in the Cel-I assay and the relative intensity of the uncut 394 bp product was estimated by the same approach used in Section 4.1.2. The ZFN4 cutting efficiencies estimated in the MwoI experiment were found to be comparable to the ones obtain by the Cel-I assay confirming the initial assessment of a 1-2% ZFN4 cutting efficiency.

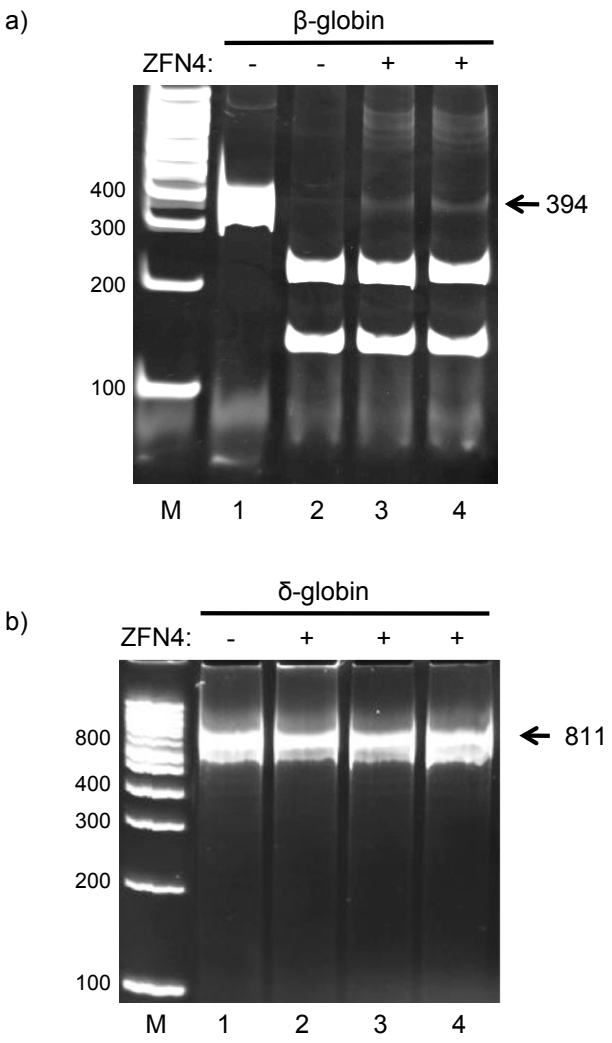


Figure 4.6 MwoI assay detects ZFN4-induced NHEJ events at the β-globin locus while CeI-I assay does not at the δ-globin locus in human K562 cells.

Lanes M of both pictures are a 100 bp ladder (NEB) a) Detection by MwoI assay of ZFN4 activity at the β-globin locus. K562 cells were transfected with control DNA (lane 1 and 2), ZFN4 expression plasmids (lane 3) or ZFN4-encoding mRNA (lane 4). Lane 1: MwoI-undigested PCR product; Lanes 2-4: PCR products were digested with MwoI before DNA-PAGE. b) Analysis of ZFN4 activity at the δ-globin locus of K562 cells. Cells were transfected with control DNA (lane 1), ZFN4 expression plasmids (lane 2) or ZFN4-encoding mRNA (lane 3 and 4). Lanes 3-4 were from two separated transfections. Genomic DNA isolated from cells 48h after transfection was used to amplify a 811 bp fragment carrying the homologous region of the δ-globin gene. All PCR products were digested with the CeI-I endonuclease before visualisation by DNA-PAGE (8% polyacrylamide gel).

4.2.5 Testing for ZFN4-induced off-target cleavage events at the δ-globin locus

As mentioned in Section 4.1.1, the risk of ZFN-induced off-target cleavage is one of the major problems related to the use of ZFNs in a possible therapeutic approach.

Although the use of *FokI* domains with an obligate heterodimer conformation (see Section 4.1.1) should greatly reduce it, the risk of off-target cleavage events could still be substantial if the chosen target sequence shares some degree of homology with other regions of the genome. This risk is particularly high for ZFNs designed to target the β-globin locus because the gene is part of a family of globin genes that share a high degree of homology (see Section 1.3).

In order to identify possible sequences partially homologous to TS-4, a BLAST query of the human genome was performed following the same criteria used in Section 3.2.1. As expected, the query identified as the most probable off-target site a 40 bp region of the δ-globin gene (GACTGCTGTCAATGCCCTGTGGGGCAAAGTGAACGTGGAT). The sequence shares a high degree of homology with the β-globin target site with only 5 bases (underlined in the above sequence) out of 40 that are different from TS-4 (for a complete comparison of the β- and δ- globin sequences see Appendix 1).

To identify off-target cleavage events at the δ-globin locus, genomic DNA samples that showed ZFN4 activity at the β-globin locus in the Cel-I assay (Section 4.1.2 Fig. 4.5 a lanes 1-4) were used as templates for the PCR amplification of an 811 bp region of the δ-globin gene encompassing the possible off-target site (for the primers used see Section 2.6.1). In the case of undesired ZFN4-induced cleavage at this site, the Cel-I assay would generate, on the polyacrylamide gel, two bands of ≈ 322 and 489 bp.

The Cel-I assay was performed on the amplified δ-globin PCR products and the DNA-PAGE of the treated samples (Fig. 4.6 a, lanes 1-4) showed the characteristic 811 bp band of undigested product but failed to show any sign

of digestion products. ZFN4-mediated cleavage at δ-globin gene was thus undetectable, and certainly less than at the β-globin locus.

4.3 Discussion

4.3.1 Detection of ZFN4-induced gene modifications at the β-globin endogenous locus by the use of Surveyor assay and MwoI restriction digestion

Although Sigma's CompoZr Custom ZFN Service provided a "ZFN-candidate validation" step where ZFN4's cleavage efficiency was tested in a suitable cell line (in this case K562 cells) using the mismatch detection assay (Cel-I assay), the decision to independently assess ZFN4's cleavage activity at the target locus was taken in order to identify possible discrepancies between the two different K562 populations (the one used by Sigma and the one used in this project) and also to test ZFN4 in a different cell line (HT1080). In addition, Sigma's CompoZr Custom ZFN Service was a new and expensive technology/service and was therefore important to confirm that ZFN4 worked as claimed.

As expected, cleavage at the target locus was identified, by Cel-I assay, only in cells transfected with ZFN4-encoding plasmids or mRNA. The manufacturer's technical manual, provided along with the ZFN4 kit, suggests the testing of both ZFN expression forms because of possible variable levels of DNA/mRNA uptake in different cell types and also because of the different modalities of protein expression. While the expression plasmids are supposed to generate ZFNs at a constant level during the period of 48-72h, the mRNA should produce a short burst of ZFNs reducing the exposure of the cells to the proteins. Moreover, the lack of promoters in the mRNA form is an additional advantage when testing ZFNs in different cell lines due to the reduced variability in levels of protein expression, variability that is usually induced by compatibility issues between promoters and cell types.

The experiments presented in Fig. 4.5 show that intensities of heteroduplex-digested bands were approximately equivalent (at least for

genomic DNA extracted after 48h) for both types of ZFN4-encoding nucleic acid (plasmids or mRNA) and between different cell lines (K562 and HT1080). These observations confirmed the initial validation step carried out by Sigma and also proved the ability of ZFN4 to promote gene modifications with a similar efficiency in a completely different cell line.

However, when the ZFN4 cutting efficiencies were estimated by comparing the intensities of cleaved and uncleaved PCR products, they were found to be up to one order of magnitude lower than the efficiency estimated by the Sigma team (1-2% cutting efficiency compared to Sigma's 10%). The difference in efficiencies was substantial therefore an in depth analysis of the factors that could affect the assay was carried out.

Transfection efficiencies were monitored during all the experiments and were always found to be above 80% therefore the reduced ZFN4 activity could not be ascribed to ineffective delivery of expression vectors or mRNA. Another possible reason for the discrepancy could have been the necessary modifications made to the Cel-I assay protocol. In theory, the Cel-I manual provided by Transgenomics recommends to perform the Cel-I digestion directly in the PCR buffer, immediately after the melting/reannealing step. However, as stated in the manufacturer's protocol, the Cel-I endonuclease efficiently cleaves heteroduplexes only in a very narrow pH window and can be also affected by additives present in the PCR buffer. The PCR buffer used in this project was not compatible with the reaction conditions required by the Cel-I assay (and described in the manufacturer's protocol) therefore the PCR products had to be purified and resuspended in a suitable buffer prior to Cel-I digestion (for the complete protocol see Section 2.7.2). This additional step and possibly a reduced rate of heteroduplex formation could have affected the assay in a significant way. The MwoI approach (MwoI restriction digestion, Section 4.1.3) does not require the formation of PCR heteroduplexes therefore it was a very useful method to confirm the initial Cel-I assessment. However, when the presence of ZFN4-induced gene modifications at the target locus were investigated with this totally different approach the intensities of undigested bands, characteristic of modified β-globin loci, were approximately equivalent to the one found in the Cel-I assay (Fig. 4.6 b).

When consulted about the possible reasons for the perceived discrepancies between the two Cel-I assays, Sigma's ZFN Technical Liaison suggested a possible difference, due to genomic instabilities, in the two K562 populations used. Nevertheless, FISH and karyotyping analyses conducted on the cell population by an independent group (see Section 4.1.2) failed to identify substantial differences between the K562 cell line used in this project and the published karyotype of K562 cells (Naumann, Reutzel et al. 2001).

Sigma does not specify which method is used to quantify the cutting efficiencies of novel ZFNs therefore one possible explanation to the lower level of ZFN4 activity found in this project could be derived from differences in the NHEJ quantification methods used by the two groups. However in an informal conversation with members of the Sigma team during a recent ZFN conference (EBML, Heidelberg, Dec 2011) Dr Porter (personal communication) was told that further Cel-I assays of ZFN4 at Sigma gave cleavage estimates of approximately 2%. Thus, although this statement has not been formally reported, it seems likely that the estimate of 1-2% cleavage identified in this project may be more reliable than the original Sigma estimate of 10%.

4.3.2 Detection of possible ZFN4-induced off-target cleavage events at the δ-globin locus

The risk of off-target cleavage has become an increasing concern for the development of a safe and reliable therapeutic approach based on ZFN-induced gene correction but still a definitive method capable of identifying all possible off-target events occurring *in vivo* in the entire genome has to be created. The current approaches are mainly based on an "in silico" method based on the identification of possible targets by the use of search algorithms like BLAST and on "in vitro" selection methods like SELEX (see Section 6.3.3) (Perez, Wang et al. 2008; Hockemeyer, Soldner et al. 2009; Li, Haurigot et al. 2011; Sebastiano, Maeder et al. 2011).

In the case of the β-globin gene, the risk of undesired gene modifications at additional loci is elevated due to the wide similarities shared by all the

globin genes (β,ε,γ and δ-globin). However, this high degree of homology was also useful to test the specificity of ZFN4. By the use of "in silico" analysis (see Section 4.1.4) it was clear that a 40 bp sequence of the δ-globin locus was the most similar site with a homology rate of ≈ 87% (only 5 bp out 40 are different). To assess ZFN4's specificity, genomic DNA samples that showed evidence of ZFN4-induced indels were subjected to an additional Cel-I screening but this time focused on the δ-globin gene. The experiment failed to detect cleavage events at this site or at least not at the same frequencies as the ones found for the β-globin gene. Unfortunately, the possibility of off-target cleavage at this locus cannot be completely excluded due to the nature of the assay used. As stated in the manufacture's manual, the Cel-I assay is effective in detecting gene modification at a specific locus only if they occur with a frequency of at least 1% (percentage of mutated genes respect to the whole gene population). In the case of ZFN4, cutting efficiencies at the target site were already at the limit of detection therefore very low frequencies of off-target cleavage would pass unnoticed. However, this partial assessment of the risk of off-target cleavages can be supported by the data found by Zou et al. (as fully described in Section 6.3.3). This group of researchers used a GFP correction assay (as the one described in Chapter 5) to investigate possible ZFN4-induced gene modifications at highly homologous sites of the ε,γ and δ-globin genes but they could not detect any (Zou, Mali et al. 2011). More in-depth analysis (possible approaches are examined in Section 6.3.3) is required because in a possible therapeutic approach also very low levels of off-target cleavage could compromise the efficacy of the technique. If it is considered that, in a hypothetical gene therapy approach for β-thalassaemia or sickle cell anaemia, bone marrow cells obtained by the patient would be genetically corrected and then transplanted back into the patient a possible ZFN-induced mutation at the level of an oncogene of a single cell could confer, back in to the patient, a proliferative advantage that could promote the onset of leukaemia.

In a study carried out by Pattanayak et al. in 2011, a novel approach was used to extensively test the specificities of active ZFNs (Pattanayak, Ramirez et al. 2011). Traditional methods (Segal, Dreier et al. 1999; Bulyk, Huang et al. 2001; Meng, Thibodeau-Beganny et al. 2007) have been used to

characterise the binding specificity of monomeric ZF-DBDs in search for possible off-target half sites but these approaches are based on the assumption that binding of one ZF-DBD monomer would not affect the binding of the other ZF-DBD monomer. The new approach developed by Pattanayak et al., instead, tried to assess cleavage specificities of full ZFNs heterodimers by the use of an *in vitro* selection method that is able to interrogate a vast library of partially randomised ZFNs target sites (these libraries should contain all possible DNA sequences that differ from the wild type target site by 7 or fewer point mutations). The experiment was performed using two well-characterised ZFNs: the CCR5-224-ZFN (part of an HIV therapeutic approach that is currently at the stage of clinical trials) and the VF2468-ZFN (part of an anti-cancer therapeutic approach). In the case of the CCR5-224-ZFN, the study identified, in vitro, 37 sites of the human genome that could be cleaved by the ZFN. Of these 37 sites 9 were actively cleaved in vivo in K562 cells and of these 9 sites only one, the CCR2 gene, was previously identified by traditional methods as a possible off-target site. The study interestingly indicated that off-target cleavage is directly proportional to the ZFN concentration, that ZFNs can cleave, at significant frequencies, any sequence that differs from the intended target sequence by three or less nucleotides and, more importantly, that there is a compensation effect between half sites of a target sequence where perfect binding of a ZF-DBD can promote tolerance in the binding affinity of the other ZF-DBD reducing, therefore, the overall specificity of the ZFN (Pattanayak, Ramirez et al. 2011).

Off-target cleavage is probably the main drawback for the creation of a reliable therapeutic approach based on ZFNs and therefore the development of safer molecular tools is of the highest importance. One promising approach is based on the implementation of a nickase to stimulate gene targeting. Nickases are a family of enzymes capable of producing Single Strand Breaks (SSB or “nicks”) at specific loci. Although not efficiently as DSBs, nicks have been shown to stimulate HR and could be used as an alternative to traditional ZFNs and TALENs technologies based on the use of *FokI* domains (Lee, Neiditch et al. 2004; van Nierop, de Vries et al. 2009). A recent study from Ramirez et al. indicated that the inactivation of the catalytic residues in one of the *FokI* monomer of standard ZFNs could generate efficient nickases

capable of targeting any desired locus in the genome. A striking feature of this novel molecular tool is the reduced rate of mutations induced by NHEJ at both on and off target sites (Ramirez, Certo et al. 2012). The newly created nickase used in the study was able to promote HR in a GFP reporter assay (as the one used in Chapter 5) even though at a lower frequency than its parental ZFN (0.14% and 0.29%, respectively) but when it was tested in the Traffic Light Reporter assay (TLR), a fluorescent repair assay capable of measuring HR and NHEJ frequencies at the same time (Certo, Ryu et al. 2011), the results showed that, while ZFN-induced DSBs are preferentially repaired by NHEJ than by HR, nicks at the target locus are instead preferentially repaired by HR. Indirectly, this study suggests also a possible reduction in the frequencies of mutations at off-target sites where nickase-induced SSBs would be repaired by HR using the sister chromatid as a template.

5 Chapter 5: Development and use of an assay for ZFN4-induced homology-directed repair

5.1 Overview

After the initial assessment of ZFN4-induced NHEJ events by the use of the Cel-I assay, a Homology Directed Repair (HDR) assay was designed in order to measure ZFN4's ability to promote HR in human cell lines and, at the same time, to compare its efficiency to I-SceI endonuclease.

I-SceI is a well-characterized rare homing endonuclease that recognizes a long DNA sequence of 18 bp (TAGGGATAACAGGGTAAT). This long target sequence has the advantage of occurring only once in every $7x10^{10}$ bp (of random sequence), a frequency so low that the sequence is expected to occur only once every 20 mammalian-sized genomes conferring I-SceI a high degree of specificity (Jasin 1996). Therefore, thanks to its unique specificity, this homing endonuclease has been extensively used to study the gene repair pathways by the introduction of its target sequence in mammalian cell models and it has been shown to greatly promote HR by the production of DSBs (Jasin 1996; Pierce, Johnson et al. 1999; Pierce, Hu et al. 2001).

The HDR assay developed in this project was based on the existing reporter system designed and published by Maria Jasin's group (Pierce, Johnson et al. 1999). In these studies, a modified GFP expression plasmid (pDR-GFP-HR) was designed to carry two disrupted copies of the eGFP gene separated by a puromycin resistance cassette (pgk-puro): one copy, called SceGFP and under the control of a CMV (CytoMegaloVirus) promoter, is a complete eGFP gene that has been disrupted by an I-SceI target site and two in-frame termination codons, while the second copy, iGFP, is missing the promoter region at the 5' end and is also truncated at the 3' end (see Fig. 5.1 a). The reporter plasmid pDR-GFP-HR is able to express fully functional GFP only if a gene conversion event occurs between the SceGFP gene and the iGFP where the iGFP sequence is used as an intra-chromosomal repair

template (Fig. 5.1 b). False positives derived by the competing NHEJ and Single Strand Annealing (SSA) pathways, which could re-arrange the two GFP genes in a functional one, are avoided thanks to the double truncation at the 5' and 3' ends of the iGFP repair template (Fig. 5.1 c) (Pierce, Johnson et al. 1999).

The reporter plasmid pDR-GFP-HR is stably transfected in a suitable mammalian cell line and the assay is carried out transiently transfecting the newly created cell line with an I-SceI expression plasmid. Levels of HR events induced by DSBs produced by I-SceI can be easily quantified through Flow Cytometry, typically 48h after transfection. The percentage of GFP positive cells among the whole cell population provides a measure of the ability of I-SceI to cleave its chromosomal target and promote HDR of the resulting DSB.

In 1999 Jasin's group used the pDR-GFP-HR system in AA8 hamster cells and demonstrated that expression of I-SceI was able to promote HR at the target locus with a frequency of 5%. By comparison, un-transfected cells showed less than 0.01% of GFP positive cells (Pierce, Johnson et al. 1999).

To determine whether ZFN4 could similarly stimulate HDR, and compare its efficiency to I-SceI, it was decided to modify pDR-GFP-HR so that it included the target site for ZFN4.

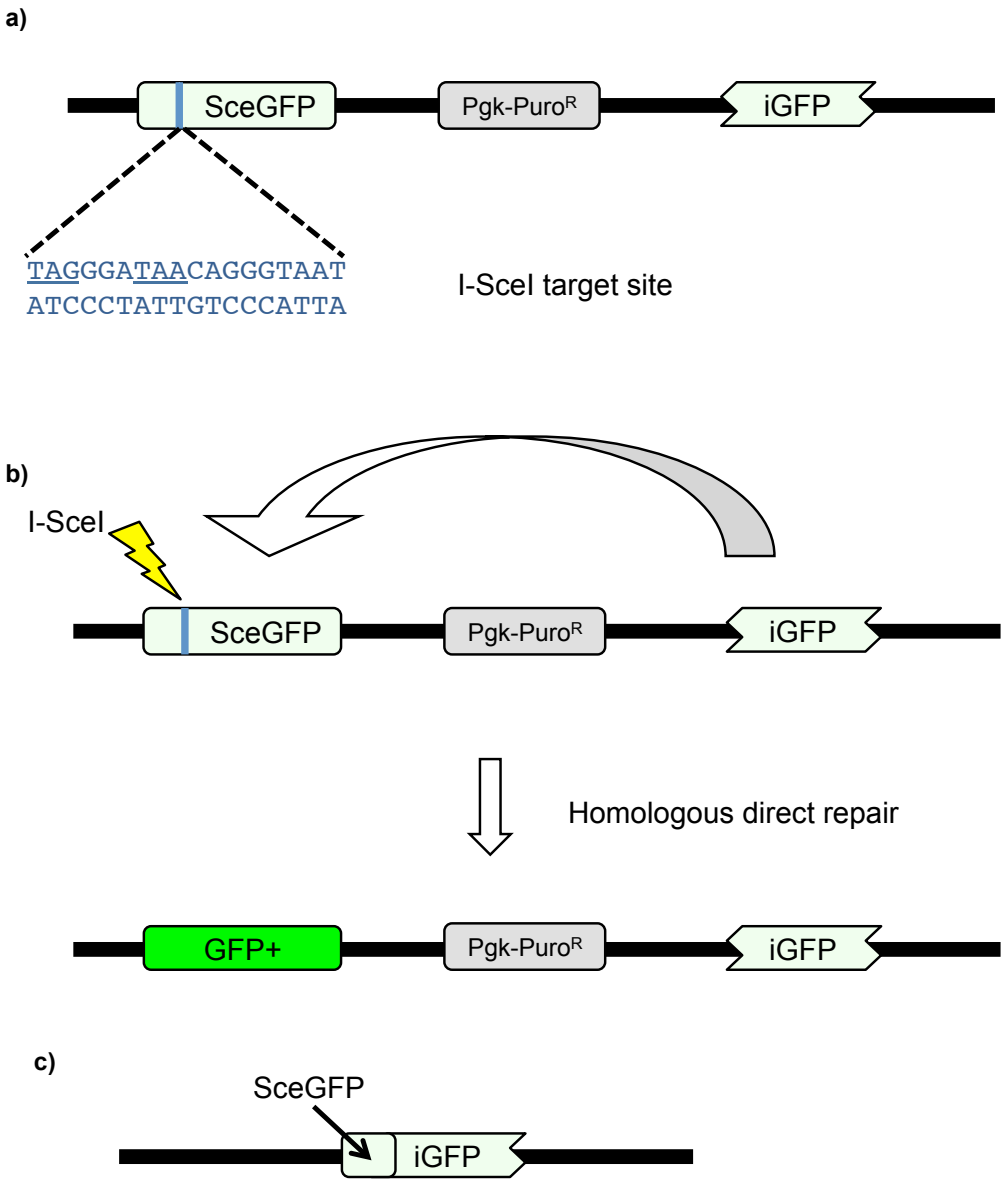


Figure 5.1 Principle of the DR-GFP reporter plasmid.
a) Graphic representation of pDR-GFP-HR. The plasmid carries two copies of the eGFP gene separated by a Pgk-Puro resistance cassette. The Sce-GFP is interrupted by the insertion of the I-SceI target sequence (shown in the blow up) which includes two in-frame termination codons (underlined). The iGFP gene instead is an eGFP truncated at the 5' and 3' ends and functions as a intra-chromosomal repair template during HR events. **b)** I-SceI-induced homology directed repair. An I-SceI induced DSB at the target locus is repaired by intra-chromosomal HR using the iGFP gene as a repair template. Expression of functional GFP is then easily monitored by flow cytometry. **c)** Non-conservative repair pathways. Fusion of the two GFP genes derived by NHEJ or SSA or non-conservative HR rearrangements does not produce a functional GFP gene due to the truncated 3' end of the iGFP sequence.

5.2 Results

5.2.1 Creation of the reporter plasmid pDR-GFP-ZFN4

To assess the ability of ZFN4 to promote HR a modified pDR-GFP-HR reporter plasmid was designed to be used in HT1080, K562 and DG75 cells. The advantage of using the well characterised pDR-GFP-HR was due to the possibility of flanking the already existing I-SceI target sequence with TS-4 (ZFN4's target sequence) and, therefore, to be able to assess ZFN4-induced HR and at the same time to directly compare the frequency of events with the one induced by I-SceI. The first step involved the introduction of the 40 bp target sequence for the ZFN4 (TS-4) into the pDR-GFP-HR reporter plasmid. The pDR-GFP-HR plasmid, (kindly donated by Maria Jasin) was linearised by digestion with the I-SceI endonuclease at the level of the SceGFP gene. A short stretch of double stranded DNA (i.e. annealed synthetic oligonucleotides) was designed to introduce the new TS-4 target sequence and, at the same time, to regenerate the I-SceI site (for the detailed procedure see Section 2.5.2). The DNA sequence carrying TS-4 was cloned in the linearised SceGFP locus of the pDR-GFP-HR reporter plasmid immediately downstream of the reconstituted I-SceI target site generating the pDR-GFP-ZFN4 reporter plasmid (Fig. 5.2 for a graphic representation of the construct and Appendix 3 for a partial sequence).

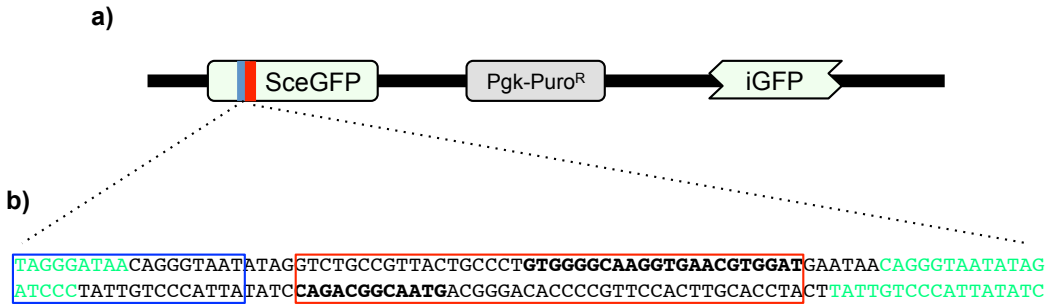


Figure 5.2 Structure of the pDR-GFP-ZFN4 reporter plasmid.
a) The arrangement of cassettes in pDR-GFP-ZFN4 is shown and is similar to that in pDR-GFP-HR (see Fig 5.1), but includes a ZFN4 target site (TS-4; red bar) downstream of an I-SceI target site (blue bar). **b)** The nucleotide sequence of the region including I-SceI and ZFN4 recognition sequences. The ZFN4 target sequence (TS-4; red box) was cloned into the cleaved I-SceI site of pDR-GFP-HR (green nucleotides) on a pair of annealed oligonucleotides (black nucleotides). This regenerated the original I-SceI recognition site (blue box). The binding site for ZFN4-L and ZFN4-R are in bold.

5.2.2 Stable transfection of adherent and suspension cells and selection of suitable clones

The reporter plasmid pDR-GFP-ZFN4, linearised at a PvuI site, was stably transfected into HT1080 cells using electroporation (for the transfection protocol see Section 2.6.6). 48h after transfection puromycin selection was started and, thanks to the pgk-puro cassette located between the SceGFP and the iGFP sites, 12 colonies were isolated after 13 days and were singularly expanded. The remaining colonies left on plate (≈ 100 colonies) were trypsinised and plated in fresh media in presence of puromycin in order to create a pool of cells carrying the pDR-GFP-ZFN4 plasmid. The clones and pool of cells were renamed HT-DR-ZFN4 (see Table 5.1) and liquid nitrogen stocks were created for long-term storage to be used in the experiments described in the following chapter (for the creation of liquid nitrogen stocks see Section 2.6.4).

To assess the ability of ZFN4 to promote HR in blood-derived cell lines, K562 (myeloid leukaemia) and DG75 (Burkitt lymphoma) cells were also stably transfected with pDR-GFP-ZFN4. Stable transfections were performed as for HT1080 cells (see Section 2.6.6) and puromycin selection was started after 48h but was carried out only on pools of K562 and DG75 cells without selection of single clones. These pools of stably transfected K562 and DG75 were dubbed K-DR-ZFN4 and DG-DR-ZFN4, respectively (Table 5.1).

As a form of control, HT1080 cells were also stably transfected with the original pDR-GFP-HR plasmid, were selected in puromycin and 10 clones plus a pool of cells (obtained by collecting the remaining ≈ 100 colonies) were isolated after 15 days. This set of stable transfectants was named HT-DR-GFP, with a numerical suffix to indicate the clone number or a -P suffix to indicate a pool of clones (Table 5.1). HT-DR-GFP cells were used to compare the data published by Pierce et al (Pierce, Johnson et al. 1999) on AA8 hamster cells with the ones obtained in this project for the HT1080 cell line and at the same time to identify possible discrepancies in HR frequencies

between pDR-GFP-HR and pDR-GFP-ZFN4 due to the presence of 40 additional bp (TS-4) in the SceGFP gene of pDR-GFP-ZFN4.

Table 5.1 Cell lines generated for the GFP reporter assay

Parental cell line	Transfected plasmid	Cells selected in puromycin	Name of novel cell lines
HT1080	pDR-GFP-HR	10 clones	HT-DR-GFP-1, -2, -3 etc.
	pDR-GFP-HR	Pooled clones	HT-DR-GFP-P
	pDR-GFP-ZFN4	12 clones	HT-DR-ZFN4-1, -2 ,-3 etc.
	pDR-GFP-ZFN4	Pooled clones	HT-DR-ZFN4-P
K562	pDR-GFP-ZFN4	Pooled clones	K-DR-ZFN4-P
DG75	pDR-GFP-ZFN4	Pooled clones	DG-DR-ZFN4-P

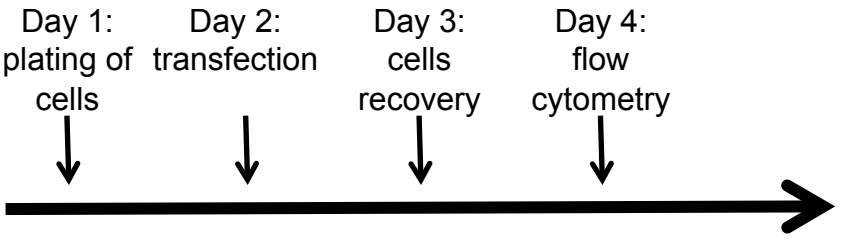
5.2.3 Homology-directed repair assays in pooled clones

Initially, preliminary tests were carried out using the newly generated pools HT-DR-ZFN4-P, HT-DR-GFP-P, K-DR-ZFN4-P and DG-DR-ZFN4-P. I-SceI or ZFN4 expression plasmids were delivered to the HT1080 derivatives (HT-DR-ZFN4-P and HT-DR-GFP-P) by lipofection, to K562 derivatives (K-DR-ZFN4-P) by electroporation and to DG75 derivatives (DG-DR-ZFN4-P) by either lipofection or electroporation (Table 5.2)

A schematic of the HDR experiments time line is shown in Fig. 5.3 and full description of the protocol is reported in Section 2.7.4. Unless differently specified, this protocol was used to perform all the following HDR experiments.

The results of the initial set of experiments carried out on the pooled clones are shown in Table 5.2.

a)



b)

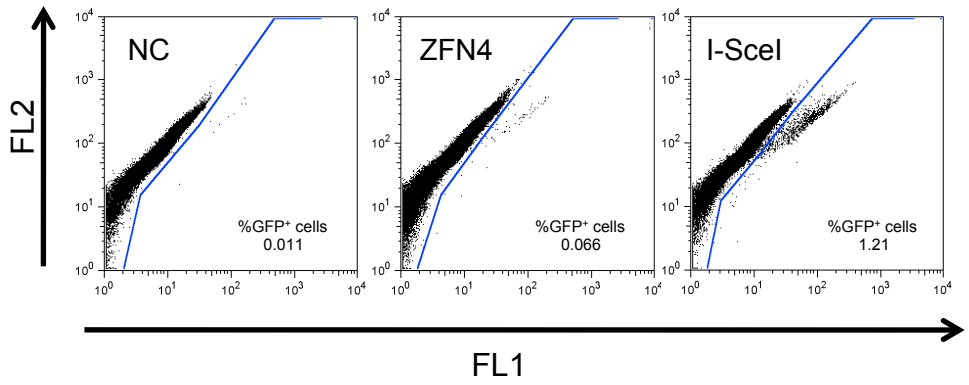


Figure 5.3 HDR assay schematic.
Each HDR experiment described in this chapter was performed, unless differently specified, following the steps described in this figure. **a)** Day 1: HT-DR-GFP, HT-DR-ZFN4, K-DR-ZFN4 and DG-DR-ZFN4 cells (derived from pools of cells or from single clones) were plated (HT-DR-GFP and HT-DR-ZFN4) or freshly sub-cultured (K-DR-ZFN4 and DG-DR-ZFN4) the day before transfection. Day 2: cells were transfected (lipofection or electroporation as specified in Section 2.7.4) with four different sets of plasmids as described in Section 2.7.4. Day 3: cells were left to recover at 37°C in presence of 5% CO_2. Day 4: (48h after transfection) GFP expression for each sample was assessed by flow cytometry of $5x10^4$ cells per sample. **b)** Flow cytometry example. Data were collected by two-colour fluorescent analysis (FL1 green fluorescence and FL2 orange fluorescence) and induced GFP expression was measured as the percentage of events falling below the diagonal (represented in the graph as a blue line).

Table 5.2 HDR assay results for pooled clones

Host cells	[a] Transf technique	[b] no. of cells x10^5	% GFP$^+$ cells 48h post-transf [c] NC	[d] I-SceI	[e] ZFN4	[f] TE
HT-DR-ZFN4-P	LF	1	0.02	0.15	0.01	84
HT-DR-ZFN4-P	LF	1.5	0.01	0.05	0.01	62
HT-DR-GFP-P	LF	1	0.31	0.39	-	84
HT-DR-GFP-P	LF	1.5	0.19	0.25	-	63
DG-DR-ZFN4-P	LF	4	0.04	0.03	0.03	0.5
DG-DR-ZFN4-P	LF	60	0.03	0.04	0.03	0.5
DG-DR-ZFN4-P	EP	20	0.01	0.01	0.01	12
K-DR-ZFN4-P	EP	20	0.03	0.05	0.04	40

[a] LF, lipofection; EP, electroporation. See section 2.6.6 and 2.6.7 for detailed protocols.
[b] Number of cells electroporated or plated the day before lipofection.
[c] Negative controls. As a control, cells were transfected with the bacterial pGP-FF plasmid (see Section 3.1) that is not expressed in mammalian cells .
[d] Cells transfected with I-SceI expression plasmid (pCβASce)
[e] Cells transfected with ZFN4 expression plasmids.
[f] Transfection efficiency: cells were transfected with a GFP expression plasmid (pmaxGFP; Lonza).

Unfortunately, two main problems were noticed. First, although transfection efficiencies for HT-DR-GFP and HT-DR-ZFN4 were satisfactory (62-84%), transfection efficiencies following lipofection of DG-DR-ZFN4 (0.5 %) and electroporation of DG-DR-ZFN4 and K-DR-ZFN4 (12-40%) were relatively low (Table 5.2). The protocol used for the electroporation of DG-DR-ZFN4 and K-DR-ZFN4 was based on experiments of a former PhD student in our group (Jennifer Zobel; personal communication) who achieved transfection efficiencies of 70% or more in both cell lines. During the electroporation of DG-DR-ZFN4, Trypan blue staining (Section 2.6.5) indicated a high level of cell death 48h after transfection. Initially, such cell death was considered responsible for the low transfection efficiency.

To reduce cell death and increase transfection efficiencies, a series of tests were carried out on DG-DR-ZFN4, increasing of the amount of cells used in the electroporation. Thus cell density was increased from 2x10^6 till 4, 6 or 8x10^6 and cells were electroporated with pmaxGFP plasmid following the electroporation protocol described in Section 2.6.6. Although cell death was remarkably reduced, the transfection efficiencies were usually even lower

than the initial electroporation experiments: the highest efficiency obtained (for $4x10^6$ DG-DR-ZFN4 cells) was still only 0.5%.

An alternative electroporation protocol was also tested on K-DR-ZFN4. Briefly, 3 $x10^6$ cells (instead of 6 $x10^6$) were electroporated with 6 µg of pmaxGFP plasmid following the protocol described in Section 2.6.6. The resulting transfection efficiency (43%) was an improvement. Nevertheless, given the apparently low level of ZFN4 activity suggested by the Cel-I assay (see Section 4.1.2), it was decided to pursue only methods that gave the highest transfection efficiencies (> 60%).

A second problem was that, although transfection efficiencies for HT-DR-GFP-P (and HT-DR-ZFN4-P) were satisfactory (see Table 5.2) the levels of GFP expression induced by I-SceI (0.05%- 0.15%) (or by ZFN4) were lower than those (~ 6%) obtained by Pierce et al. (Pierce, Johnson et al. 1999).

Moreover, a relative high background of basal GFP expression was detected in most of the negative controls (Table 5.2). The high GFP background was close to the levels of GFP expression induced by I-SceI and more especially to the ones induced by ZFN4. The GFP basal expression was also detected by Pierce et al. in their 1999 paper (Pierce, Johnson et al. 1999) and could be explained by possible spontaneous HR events occurred in some of the cells that, subsequently, started to constitutively express functional GFP proteins.

Due to the low levels of induced GFP expression and the relative high background detected in the negative controls, the use of pools of cells was discarded in favour of the use of single HT-DR-ZFN4 clones.

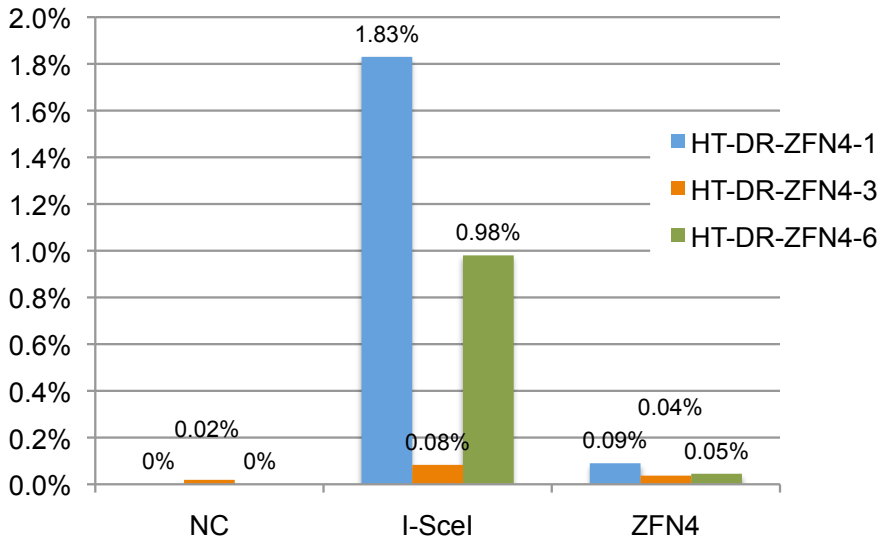
5.2.4 Screening of individual clones in the HDR assays

Ten out of the twelve HT-DR-ZFN4 clones that were previously selected (Section 5.3.2) were used in the HDR assay in order to identify the most suitable clone for more detailed analyses. The tests were performed by lipofection as before (section 2.7.4) at two different cell plating densities ($1.0x10^5$ and $1.5x10^5$) on day 1 and the experiments were repeated twice for each clone at each cell density.

The data collected from the screening of HT-DR-ZFN4 clones highlighted key variables affecting the HDR assay. First, although they all carried the same reporter HT-DR-ZFN4, the different clones displayed different I-SceI- and ZFN4-induced GFP expressions. Most of them (HT-DR-ZFN4-2, -4, -5, -7, -8, -9 and -10) showed very low levels of I-SceI- and ZFN4-induced GFP expression (on average ≈ 0.04% and ≈ 0.02% for I-SceI and ZFN4, respectively). These GFP levels were very close to un-induced GFP levels (negative controls: average ≈ 0.02%) and therefore these clones were discarded in favour of the remaining three (HT-DR-ZFN4-1, -3 and -6) whose I-SceI-induced GFP expression was more clearly in excess of the negative controls (Figure 5.4). This clonal variation in nuclease-inducible GFP expression probably reflects "chromosomal position effects" due to the random integration of the pDR-GFP-ZFN4 reporter plasmid in the genome indicating that different regions of the chromatin are more or less accessible to the nucleases used in these experiments or to HR following cleavage.

A second important factor that greatly affected the assay was the density of cells plated at day 1. It is well know that cell density affects the efficiency of lipofection. However, the effect of cell density on nuclease-induced GFP expression appeared to be distinct from its effect on transfection efficiency (as confirmed by the experiments summarised in Table 5.4). During the preliminary tests, in order to identify the optimal cell density for the HDR assay, on day 1 two different cell densities were used: $1x10^5$ and $1.5x10^5$ per well. Due to the poor levels of GFP expression, an additional density of $4x10^5$ cells per well, was also tested for three of the clones that previously performed poorly in the assay (HT-DR-ZFN4-3, -4 and -5). Transfection efficiencies were consistently high (usually above 90%) but I-SceI- and ZFN4-induced GFP expression was notably better at the lower cell densities. The best results were obtained with a density of $1x10^5$ cells per well (see Figure 5.4) while HT-DR-ZFN4-3, -4 and -5 clones at a density of $4x10^5$ cells per well showed values of induced GFP expression similar the ones obtained for the relative negative controls (on average ≈ 0.02%).

a)



b)

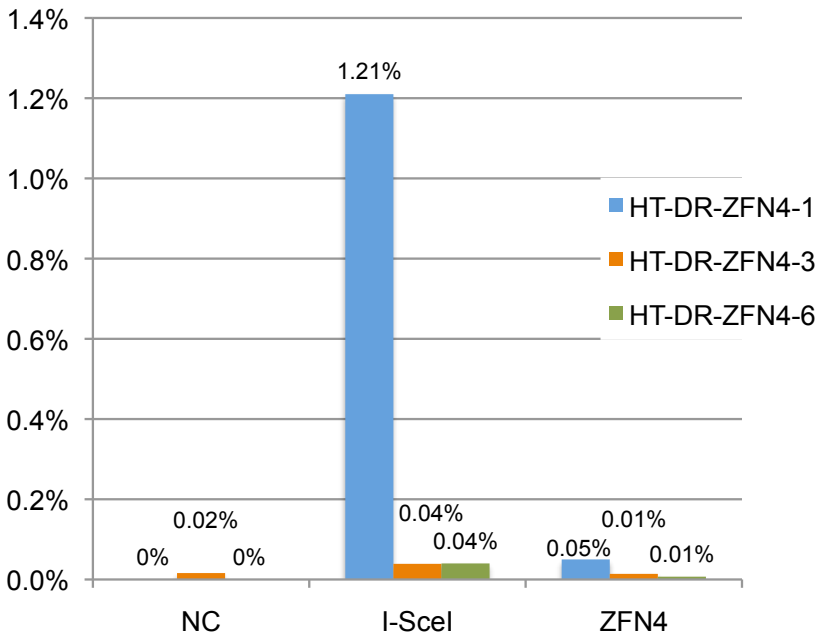


Figure 5.4 I-SceI and ZFN4 induced GFP expression levels in three HT-DR-ZFN4 clones.
Examples of GFP expression in the negative control or induced by I-SceI or ZFN4 are reported in the two diagrams as the percentage of GFP positive cells on the whole sample population ($5x10^4$ cells per sample). The HDR assay was performed for each clone at two different cell densities: **a)** $1.0x10^5$ cells per well and **b)** $1.5x10^5$ cells per well. (N=1)

Although all the experiments were carried out using cells originated from new liquid nitrogen stocks in order to reduce to a minimum the number of cell passages (and therefore avoiding possible silencing of the HDR reporter gene), of the three clones that responded satisfactorily to the I-SceI treatment in the preliminary screening (HT-DR-ZFN4-1, -3 and -6) the HT-DR-ZFN4-6 clone showed, in subsequent experiments, a substantial decrease in I-SceI and ZFN4 induced GFP expression. The initial test carried out at a cell density of $1.0x10^5$ cells per well showed an I-SceI-induced GFP expression of 0.98% while two two following experiments, carried out at the same cell density, highlighted an I-SceI induced GFP expression of 0.32% and 0.05%, respectively. A similar reduction was noticed also for HT-DR-ZFN4-6 cells transfected with the ZFN4 expression plasmids. Due to the reduction in induced GFP expression and the difficulties in addressing the origins of the problem, the HT-DR-ZFN4-6 was not used in the following experiments. Some possible explanation for this decrease is examined in Section 5.4.2.

The levels of I-SceI- or ZFN4-induced GFP expression during the experiments carried out in this project were consistently lower than the 5% GFP expression reported by Pierce et al. (Pierce, Johnson et al. 1999). In order to identify possible differences due to the presence of the additional 40 bp ZFN4's target sequence (TS-4), two clones (HT-DR-GFP-7 and -8) carrying the original pDR-GFP-HR reporter plasmid, which harbours only the I-SceI target sequence in the Sce-GFP gene, were used in a control test. The experiments were carried out following the same protocol previously described (see Figure 5.3) with cell densities per well plated at day 1 of $1x10^5$ and $1.5x10^5$ for the HT-DR-GFP-8 and $1.5x10^5$ for HT-DR-GFP-7 (see Table 5.3). The results showed levels of I-SceI-induced GFP expression similar to the ones found for the HT-DR-ZFN4-1 clone arguing against a possible negative effect due to the presence of the additional ZFN4's target sequence.

Table 5.3 HDR assay of HT-DR-GFP clones

Cell line	[a] Transf. technique	[b] no. of cells x10^5	% GFP^+ cells 48h post-transf		
			[c] NC	[d] I-SceI	[e] TE
HT-DR-GFP-8	LF	1	0	1.18	92
HT-DR-GFP-8	LF	1.5	0	0.5	73
HT-DR-GFP-7	LF	1.5	0	0.02	63

[a] LF, lipofection.
[b] Number of cells per well plated the day before transfection.
[c] Negative controls. Cells transfected with pGP-FF.
[d] Cells trasnsfected with I-SceI expression plasmid (pCβASce)
[e] Transfection efficiency: cells were transfected with a GFP expression plasmid (pmaxGFP; Lonza).

5.2.5 Detailed analysis of the HT-DR-ZFN4-1 clone

To increase confidence in the collected data, and to fully analyse the effect of the cell density on the levels of induced GFP expression, the HT-DR-ZFN4-1 clone was selected to be used in following experiments due to its relative high response to I-SceI and ZFN4 treatment (see Fig. 5.5).

In order to assess the possible effect of the cell density on the frequency of induced HR, four different cell densities were used to perform the HDR reporter assay. The day before each transfection (day 1), HT-DR-ZFN4-1 clone was plated in a 24-well plate at increasing densities of 0.5x10^5, 1x10^5, 1.5x10^5 and 2x10^5 cells per well (see Table 5.4). All the experiments were conducted on three different days indicated in Table 5.4 as Exp. 1, 2 and 3.

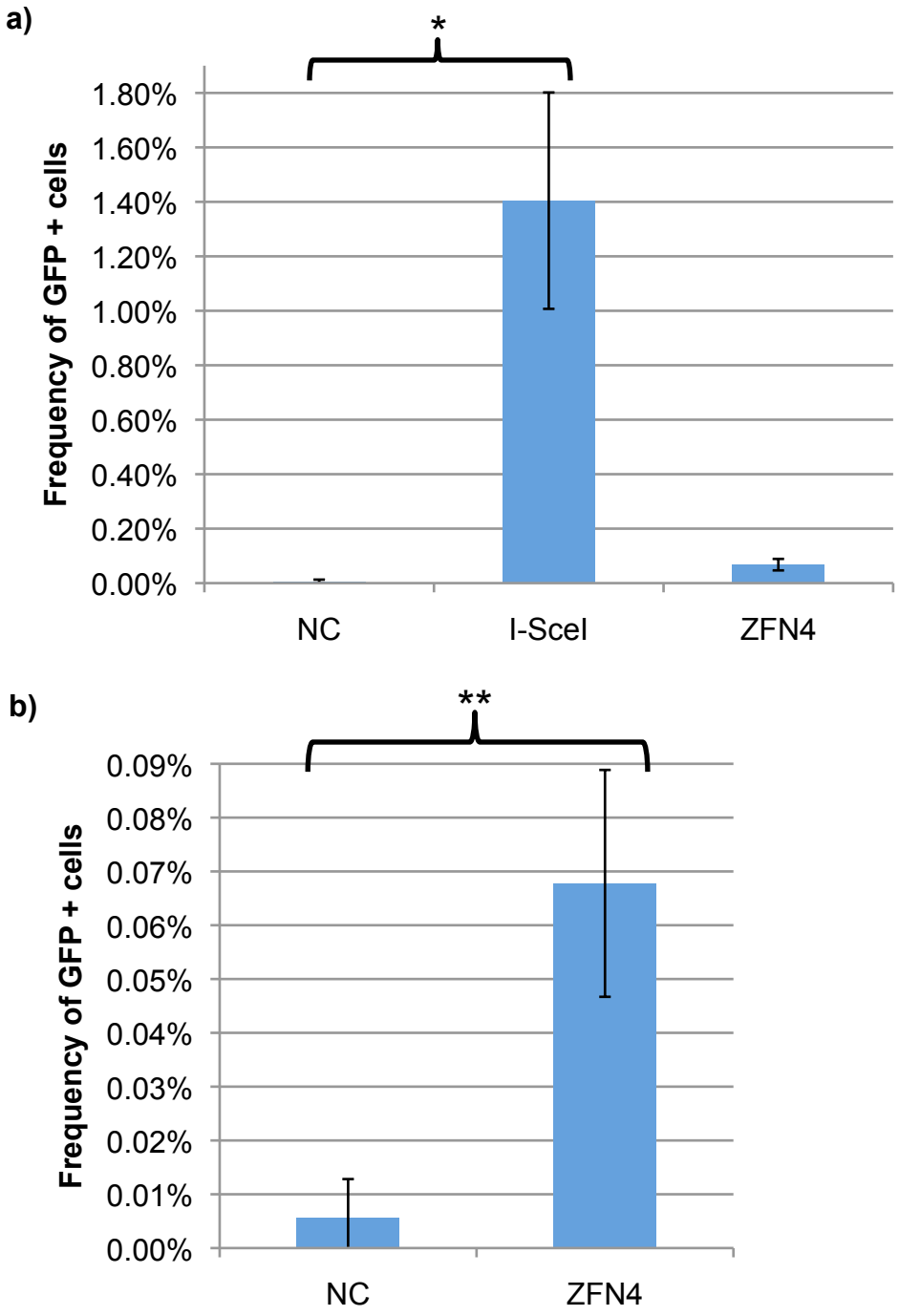


Figure 5.5 HDR assays of clone HT-DR-ZFN4-1.
a) and **b)** The percentages of GFP positive cells for the negative control (NC), I-SceI and ZFN4 are reported in this diagrams as the mean obtained by the analysis of data collected from experiments carried out at cell densities of 0.5, 1.0 and 1.5x10^5 cells per well. The mean and relative error bars shown in the two diagrams were obtained by pooling data obtained during the course of nine different experiments carried out at the above mentioned cell densities (statistical relevance of the assay was confirmed by the measuring of the p values * p= 3.8x10^{-7} and ** p=2.1x10^{-7}).

Table 5.4 HDR assays of clone HT-DR-ZFN4-1

[a] no. of cells x10^5	[b] Expt	% GFP^+ cells 48h post-transf			
		[c] NC	[d] I-SceI	[e] ZFN4	[f] TE
0.5	1	0.01	1.77	0.09	97
0.5	1	0	1.92	0.1	nd
0.5	2	0	1.13	0.05	95
0.5	2	0.01	1.35	0.04	nd
1	3	0.02	1.56	0.07	99
1	3	0	1.83	0.09	nd
1	1	0	0.71	0.06	95
1.5	3	0.01	1.16	0.06	98
1.5	3	0	1.21	0.05	nd
2	3	0.01	0.46	0.04	98
2	3	0.01	0.52	0.04	nd

[a] Number of cells per well plated the day before lipofection.
[b] The three experiments were carried out on three different days.
[c] Negative controls. cells transfected with pGP-FF.
[d] Cells trasnsfected with I-SceI expression plasmid (pCβASce)
[e] Cells transfected with ZFN4 expression plasmids.
[f] Transfection efficiency: cells were transfected with a GFP expression plasmid (pmaxGFP; Lonza, see Section 2.6.9). nd: not determined.

Nuclease-induced GFP expression levels were found to be inversely correlated to the cell densities (Table 5.4 and Fig. 5.6), with the lowest GFP expression found at a density of 2x10^5 cells (with an average of 0.5% GFP expression for I-SceI and 0.04% for ZFN4). The effect of the cell density on the HDR assay is shown in Figure 5.6. This dependency of the levels of HR on the cell density can be explained by the difference in the two main DNA repair pathways (HR and NHEJ) used by the cell to correct possible DSBs where HR is preferentially used by cells to repair damage during DNA replication (more precisely in the S and G2 phases of the cell cycle) and therefore during active cell growth while the NHEJ pathway is mostly used by the cell during the G1/G0 phase of the cell cycle (Branzei and Foiani 2008). A possible hypothesis that could explain the dependence of HR levels to the cell density is that cells plated at low densities (e.g. 0.5x10^5 cells per well) on day 1 were in an active replication state at the time of transfection (day 2) and were still growing when I-SceI or ZFN4 were expressed (day 2 and 3). On the other hand, cells plated at 2.0x10^5 cells per well probably would have reached

confluency before I-SceI or ZFN4 were expressed and would have already been in the G0 phase of the cell cycle.

Despite the fact that GFP expression greatly varied with the changing of cell density, the ratio between levels of I-SceI and ZFN4 induced HR for each experiment was constant suggesting that the ability of ZFN4 to produce DSBs at the target locus and to promote HR is on average 20-fold lower than the one displayed by I-SceI (Figure 5.5). However, it is not possible to rule out that better expression of I-SceI (perhaps because it requires delivery of only a single expression plasmid) may have contributed to the higher proportion of GFP positive cells and expression levels of both I-SceI and ZFN4 should be assessed for example by Western blot.

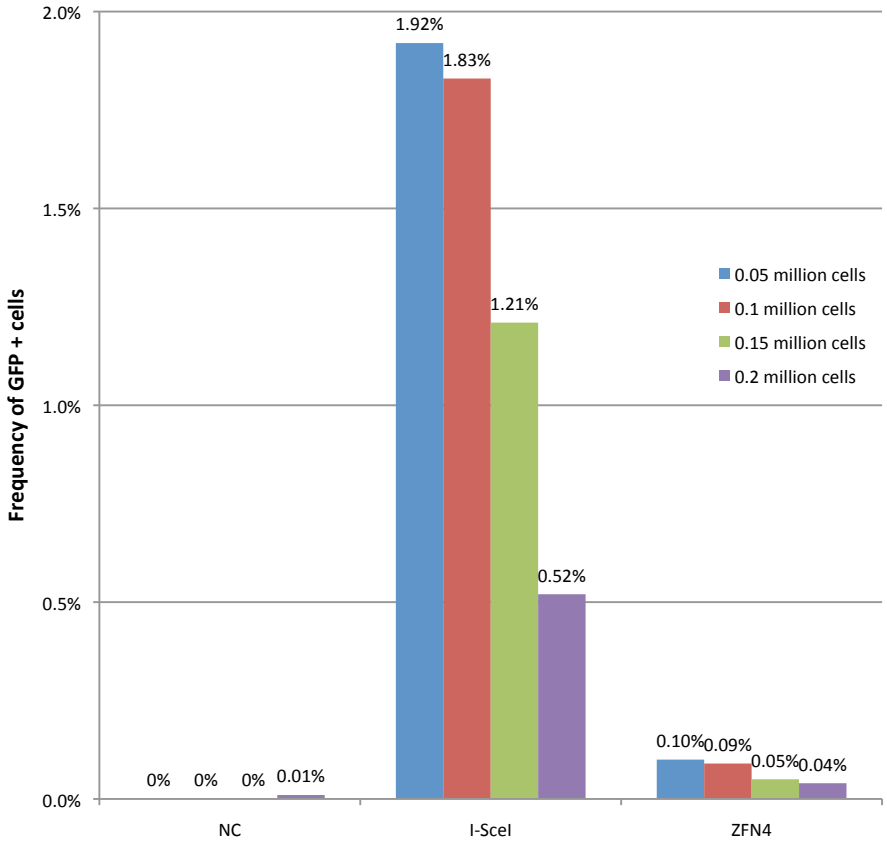


Figure 5.6 Effect of the cell density on the levels of induced GFP expression for the HT-DR-ZFN4-1 clone.
The diagram shows the difference in GFP expression induced in the negative control or by I-SceI and ZFN4 at cell densities of 0.5, 1.0, 1.5 and $2x10^5$ cells per well. Each value is reported above the relative column as the percentage of GFP + cells on the whole cell population ($5x10^4$ cells per sample) measured by flow cytometry.

5.3 Discussion

5.3.1 Creation of the reporter plasmid pDR-GFP-ZFN4

The use in this project of the HDR assay based on the reporter plasmid pDR-GFP-HR was thought to be a valuable approach to assess the frequencies of HR induced by ZFN4 and at the same time to compare them with the well characterised activity of the I-SceI endonuclease. Comparing the activity of different ZFNs or comparing different customised endonucleases like ZFNs and TALEN can be useful to assess the state of novel gene modification techniques and, more importantly, to identify the complex factors that can affect the frequencies of gene targeting in different cell lines. Unfortunately, many different factors affect the efficiency of engineered nucleases and therefore it is relatively difficult to compare different gene targeting techniques. The factors affecting frequencies of gene targeting can be broadly collected in three categories: the type of nuclease used, the type of delivery method and the type of cell line. Since the discovery that DSBs produced in the genome can greatly stimulate the frequencies of HR (Rouet, Smih et al. 1994; Choulika, Perrin et al. 1995), different engineered nucleases like ZFNs (Wright, Thibodeau-Beganny et al. 2006; Maeder, Thibodeau-Beganny et al. 2009; Sander, Dahlborg et al. 2011), TALENs (Cermak, Doyle et al. 2011; Li, Huang et al. 2011) and meganucleases (Grizot, Smith et al. 2009) have been developed as possible tools for gene targeting. Each of these nucleases has widely different binding and cutting efficiencies that affect their specificity and ability to promote HR and a wide range of binding efficiencies are also found within the same category of nucleases due to their modular nature (ZFNs and TALEN). Different delivery methods have been shown to affect the frequencies of gene targeting (Yanez and Porter 1998; Yanez and Porter 1999; Carroll 2011) and different cell lines or organisms, where the gene modification has to be produced, require different types of delivery methods

for both the nucleases expression vectors and DNA targeting construct. Finally, the same nuclease can induce widely different levels of gene targeting when applied in different cell lines (Li, Haurigot et al. 2011; Zou, Mali et al. 2011).

The use of an HDR assay based on the pDR-GFP-HR reporter plasmid could be useful to compare the activity of different nucleases by expression of their efficiencies as the ratio between the percentage of GFP positive cells induced by the nuclease and the one induced by the "golden standard" I-SceI.

5.3.2 Homology-directed repair assay

Preliminary results carried out on pools of HT-DR-ZFN4 and HT-DR-GFP cells showed that the use of pools of cells was not ideal for assessing the efficiency of ZFN4 in inducing GFP expression mainly for two reasons. The results obtained from HT-DR-GFP pool showed relative high levels of GFP expression also in the negative controls making it difficult to measure any I-SceI-induced GFP positive cells. Although not as high as in HT-DR-GFP cells, the GFP values obtained from the negative controls of HT-DR-ZFN4 were found to be close to those induced by ZFN4. This basal GFP expression can be easily explained by a possible spontaneous recombination, in some of the cells, between the two GFP cassettes that would at that point start to constitutively express the fluorescent protein. Due to the low levels of ZFN4 cutting efficiency measured by the Cel-I assay (Section 5), the ZFN4 induced GFP expression was expected to be low and therefore GFP levels in the negative controls of the HDR assay had to be minimal in order to be able to easily discriminate between spontaneous HR events from the ones induced by either I-SceI or ZFN4.

The second reason why the pools of HT-DR-ZFN4 cells were discarded in favour of individual clones was the relative low GFP expression induced by both I-SceI and ZFN4. The discrepancies found between the data collected in this project and the ones from Pierce et al. could be ascribed to differences in the different cell lines used (HT-1080 cells used in this project and AA8 hamster cells used by Pierce et al.) or to differences in the transfection

methods (lipofection and electroporation, respectively). Although the differences in methodology and cell lines used in the two studies could explain them, the low levels of induced GFP expression made difficult to assess the activity of both I-SceI and ZFN4. By the following screening of ten individual HT-DR-ZFN4 clones it was clear that the low induced GFP expression measured in the pool was due to the fact that the majority of HT-DR-ZFN4 cells had a very low response to treatment with I-SceI and ZFN4 with only three clones out of ten showing an acceptable level of response to the nucleases treatment. Although disappointing, these data could indicate a dependence of the nucleases ability to cut their specific target sequence to the chromatin structure where the pDR-GFP-ZFN4 reporter plasmid has randomly integrated. DNAse I treatment of genomic DNA has been extensively used to identify regions that are actively transcribed (euchromatin) from regions of the genome that are inactive (heterochromatin) (Kornberg and Lorch 1999). The difference in DNAse I accessibility to euchromatin and heterochromatin regions of the genome could be also extended to other types of nucleases and therefore, depending on the position where the pDR-GFP-ZFN4 reporter plasmid randomly integrated, some HT-GFP-ZFN4 clones were more responsive to I-SceI and ZFN4 treatment than others.

An additional factor that could have contributed to the low levels of induced GFP expression is the hypermethylation of pDR-GFP-ZFN4 reporter plasmids that underwent HR events. This phenomenon was examined by Cuozzo et al. in a 2007 study where they used the same HDR reporter assay developed by the Jasin group in Hela cells (Cuozzo, Porcellini et al. 2007). In their study, Cuozzo et al. showed how pDR-GFP-HR reporter plasmids that were subjected to I-SceI-induced HR were subsequently silenced by methylation and the levels of GFP expression tended to drop during the days after transfection. This effect was reversed if the cells carrying the reporter plasmid were treated with 5-aza-2-deoxycytidine for 48h after transfection with the I-SceI expression plasmid. The silencing effect was not observed in Hela cells transfected with a wild type GFP expression vector and therefore this effect could only be related to GFP genes that were targeted by HR events (Cuozzo, Porcellini et al. 2007).

A reduction in GFP expression levels was noticed in this project for the HT-DR-ZFN4-6 clone. The initial, relatively high, GFP expression induced by I-SceI or ZFN4 was followed, in subsequent experiments, by a drastic reduction in the percentage of GFP positive cells. This reduction can not be explained by a silencing of the pDR-GFP-ZFN4 reporter plasmid because of the continuous selection pressure imposed by puromycin that should have prevented the silencing of the reporter plasmid and also because each experiment carried out on HT-DR-ZFN4 cells was performed using fresh cell samples (and therefore subjected to low numbers of cell passages) originated from liquid nitrogen stocks. However, it is possible that the HT-DR-ZFN4-6 clone was not, in reality, a population of cells originated from a single colony but a cell population originated from multiple clones. If these multiple clones had the pDR-GFP-ZFN4 reporter plasmid randomly integrated in different regions of the chromatin, the relatively high GFP expression detected in the first experiments would have been produced by clones with different levels of response to the I-SceI and ZFN4 treatment. It is possible that in subsequent experiments one of the clones (one with low levels of I-SceI- and ZFN4-induced GFP expression) could have acquired a growth advantage reducing the overall response of HT-DR-ZFN4-6 to I-SceI and ZFN4 treatment.

The data collected on experiments carried out on the HT-DR-ZFN4-1 clone indicated that I-SceI (≈ 1.4%) was 20 times more efficient than ZFN4 (≈ 0.07%) in promoting HR at the target locus. The 20-fold difference between I-SceI and ZFN4 induced GFP positive cells was constant throughout all the experiments and was not affected by the different cell densities that were tested. By comparison, Zou et al., who used ZFN4 in a study published in 2011 (Zou, Mali et al. 2011), carried out a similar GFP correction assay in 293T cells and detected a percentage of ZFN4-induced GFP positive cells of 0.16%. The data collected in this project and the one found by Zou et al. are difficult to compare due to the different cell lines used (HT1080 and 293T cells, respectively) and due to the different types of GFP reporter plasmids used (the one used by Zou et al. has a different promoter upstream of the GFP gene and the repair template was not included in the same reporter plasmid but was transiently transfected along the ZFN4 expression plasmids). However, the differences in ZFN4's efficiencies reported in the two studies

show how the activity of the engineered nucleases varies greatly depending on both the cell line and the type of assay used to test them and highlight the need for a more standardised approach to evaluate different nucleases activities. This last consideration supports the idea of using I-SceI as a "gold standard" that would allow expressing the efficiencies of possible customised nucleases not as a simple percentage of induced GFP positive cells, which varies greatly depending on the cell type, on the transfection technique and the reporter assay used, but as a ratio between the activity of I-SceI and the nuclease to be tested. All the factors that affect the frequencies of ZFN induced HR events would also affect the ones induced by I-SceI therefore the ratio between the two values should be more or less constant also if the experimental conditions varies.

Different cell densities greatly affected the overall frequency of HR events induced by both I-SceI and ZFN4. Cells at lower densities were more responsive to treatment with the nucleases indicating that cells in a state of active replication are more prone to correct possible DSBs by the HR pathway. The dependence of the ZFNs' efficiencies on the cell cycle, on the accessibility of the DNA target locus (chromatin structure) and the possible silencing effect observed on genes that underwent HR (Cuozzo, Porcellini et al. 2007) are all limiting factors that have to be addressed in order to develop an effective gene targeting approach for the treatment of any possible genetic disease. These factors are even more relevant for the development of a ZFN-based gene therapy for β-thalassemia or sickle cell anaemia where the gene correction of disease-causing mutation has to be made in undifferentiated HSCs or iPS cells. Two papers (Sebastiano, Maeder et al. 2011; Zou, Mali et al. 2011) already showed that ZFN-induced gene targeting of the β-globin locus in iPS cells occurs with very low frequencies even when using drug resistance on the targeting construct to induce positive selection. Although the development of novel high-affinity ZFNs or the use of alternative nucleases like TALEN will improve the efficiency of gene targeting in HSCs and iPS cells, in order to create an effective and, more importantly, safe therapeutic approach there will still be the need to better understand all the complex factors that could affect HR and therefore the gene correction of possible genetic diseases in relevant cell types.

6 Chapter 6: Targeted integration at the endogenous locus

6.1 Overview

After assessing ZFN4's cutting efficiency and its ability to promote HR in a model GFP-based assay (Chapter 5), the next step was to demonstrate the efficacy of ZFN4 to promote gene targeting at the endogenous HBB locus. This requires cells to be transiently transfected with ZFN4 expression plasmids in conjunction with a DNA repair template homologous to the targeted region. Although the expression of ZFNs should greatly increase the frequency of HR between the repair template and the target locus, the low cutting efficiency estimated for ZFN4 suggested the need for a reliable and quick method to select and analyze putative targeting events. For this reason it was decided to develop a β-globin gene targeting construct in which a selectable marker gene ($Puro^R$, conferring resistance to puromycin) is flanked by two extended regions of homology with the β-globin gene. The construct was designed so that the puromycin resistance cassette is positioned within the ZFN4 target sequence: in this way, possible targeted clones would no longer be susceptible to cleavage by ZFN4 excluding the risk of damaging the selectable marker. Alternative targeting construct designs were possible but were rejected because they were considered inefficient to use. The $Puro^R$ cassette, for example, could be located in other positions of the construct leaving the ZFN4 target site intact but this could lead to cleavage by ZFN4 not only of the targeting construct before HR occurs (which may reduce targeting efficiency) but also of targeted loci following HR with the risk of damaging, through NHEJ, the $Puro^R$ cassette if close to the target site. Although this problem could be avoided by introducing silent mutations in the ZFN4 target sequence of the targeting construct, this would not avoid other potential problems such as the chance of crossover events between the $Puro^R$ cassette and the mutated ZFN4 site so that the resistance cassette does not integrate (Fig. 6.1). Other factors that could reduce the frequency of gene targeting are the amount of heterology between the target gene and the targeting construct

and the distance between the ZFN target site and the site where the heterology region (the resistance cassette) should be introduced (Elliott, Richardson et al. 1998; Moehle, Rock et al. 2007). Therefore, placing the resistance cassette (that has a size of 1.4 kb) precisely at the centre of the target site flanked by extended regions of homology should partially compensate for this problem (more details in discussion, Section 6.3.1).

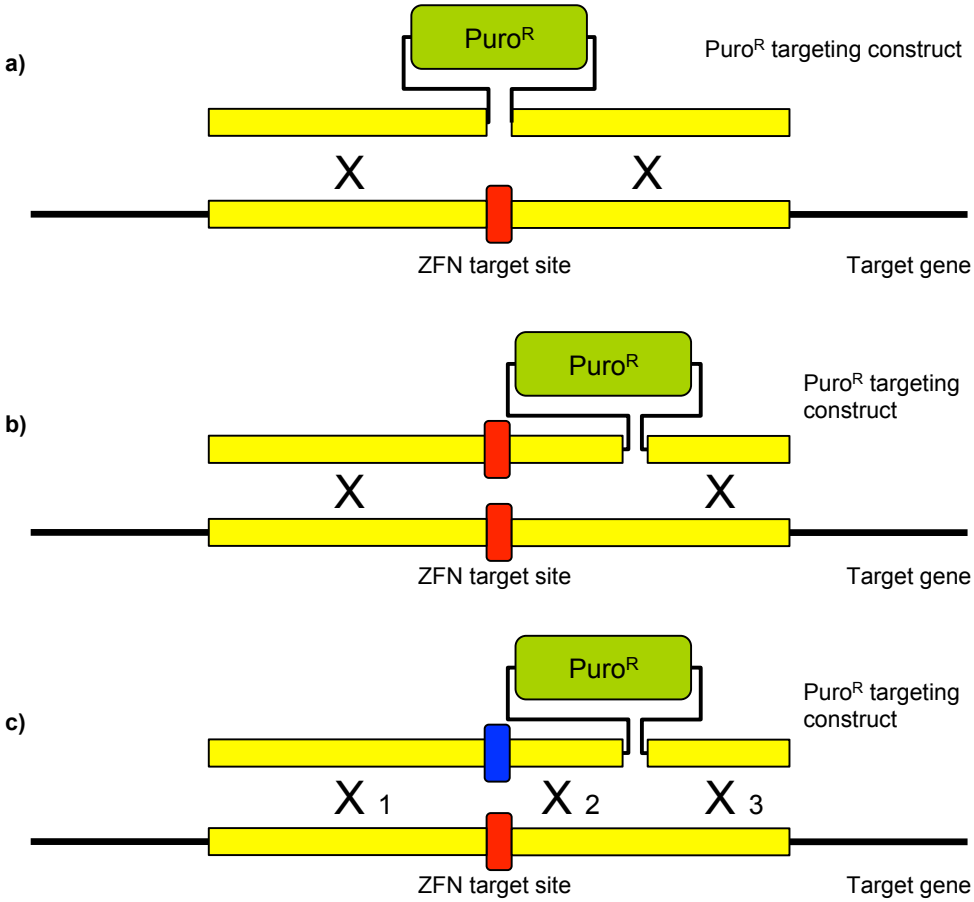


Figure 6.1 Alternative targeting construct designs.
a) The two homology regions of the targeting construct and of the target gene are in yellow. The ZFN target site is in red while the puromycin cassette is in green. A ZFN-induced DSB and successive HR between the homology regions will introduce the Puro^R cassette at the target locus and at the same time will disrupt the ZFN target site preventing further ZFN cleavage of the successfully targeted gene. **b)** In this construct-design the position of the Puro^R cassette in the targeting construct is shifted leaving the ZFN target site intact. In this type of design ZFNs could cleave the targeting construct before HR occurs and could also cleave possible targeted genes with the risk of disrupting the Puro^R cassette through NHEJ. **C)** This alternative design introduces in the construct a modified target site (in blue) that cannot be cleaved by ZFNs. This would circumvent the problematic examined in b) but the targeting construct would still be inefficient: the PuroR cassette would be integrated in the gene only if a double-crossover event would occur between points 1 and 3 (indicated in figure) but not in the case of a double crossover between points 1 and 2.

6.2 Results

6.2.1 Creation of a targeting construct specific for the β-globin gene

Following the above considerations, an initial targeting construct (pBL-TV-TC, Fig. 6.2 c) was designed as a puromycin resistance cassette (driven by the SV40 promoter) flanked by a left homology arm of 2 kb and a right homology arm of 2.5 kb. Two sets of different primers (see Section 2.5.3) were designed in order to PCR amplify the two arms of homology and to introduce at the same time suitable restriction sites for easy cloning. Genomic DNA from K562 cells was used as a template for the PCR reactions while a high-fidelity Taq polymerase (Phusion from Thermo Scientific) was used in order to reduce the risk of introducing unwanted mutations that could reduce the rate of homology between construct and target site. The first pair of primers (HAL-F and HAL-R) amplified a 2 kb region 3 bp upstream of the TS-4 while the second pair of primers (HAR-F and HAR-R) amplified a 2.5 kb region 15 bp downstream of the target site (Figure 6.2 b). The newly created homology arms were then digested with the specific restriction enzymes (described in Section 2.5.3) and cloned in the pBL-Puro plasmid in two sites flanking the resistance cassette. This first targeting construct was called pBL-TV-TC and had a size of 8.8 kb (Fig. 6.2 c).

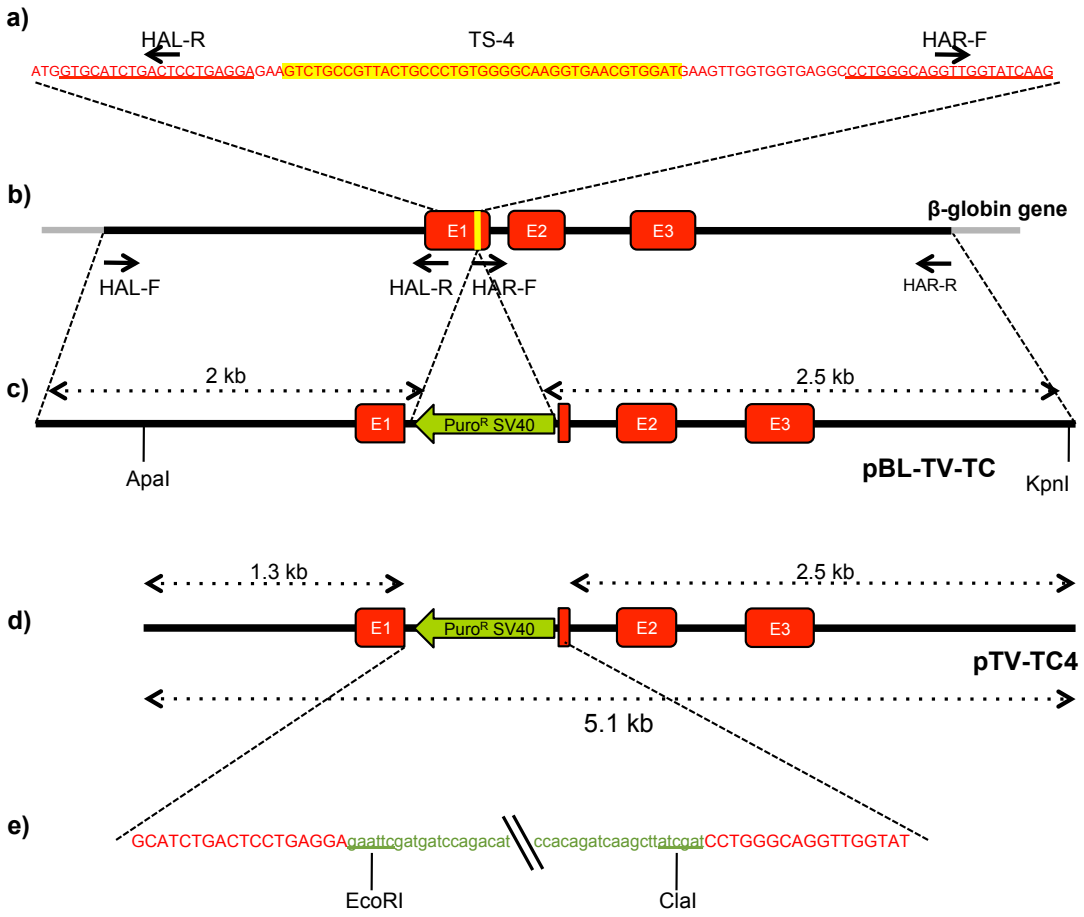


Figure 6.2 Design of targeting constructs.
a) Part of HBB exon1 sequence (red) showing the TS-4 target site (highlighted yellow) and the HAL-R and HAR-F primers (underlined) used to amplify the homology arms. **b)** Map (roughly to scale) of the β-globin gene region showing exons 1, 2 and 3 (red boxes), TS-4 (yellow segment) and the locations and orientations of PCR primers (arrows) used to amplify the homology arms. Chromosomal regions flanking the left and right regions of homology are depicted in grey. **c)** Map of the initial targeting construct (pBL-TV-TC; vector DNA omitted; for a full map see Appendix 1) with DNA depicted as in b) showing the puromycin resistance cassette (green arrow) located between the two homology arms. Regions of homology between the pBL-TV-TC and the target locus are indicated by dashed lines between b) and c). The ApaI and KpnI sites, used to generate the final targeting construct (pTV-TC4) are indicated. **d)** Map of the final targeting construct pTV-TC4 (vector DNA omitted; for a full map see Appendix 3). Due to difficulties originated during the detection by PCR screening (see Section 6.2.1) of the positive control pBL-TV-pc construct when integrated in the genomic DNA of transfected cells, the original pBL-TV-TC targeting construct had to be modified by the shortening of the left homology arm by taking advantage of the two ApaI and KpnI sites indicated in c). **e)** DNA sequence of junctions between the Puro^R cassette (green) and homology arms (red) in both pTV-TC4 and pBL-TV-TC. Sites for EcoRI and ClaI used when cloning the Puromycin cassette are underlined (additional information in Materials and Methods; Section 2.5.3).

To detect and confirm targeting events at the β-globin locus, a specific PCR assay was developed based on primers PT1-F and PT1-R. The target-specific primer (PT1-F) binds to the globin locus immediately upstream of the left homology arm of pBL-TV-TC, while the construct-specific primer (PT1-R) binds within the Puro^R cassette (Fig 6.3 a). In this way only authentic targeting events should generate a characteristic 2.5 kb PCR product (the PCR assay is described further in Fig. 6.3). In order to optimise and test the efficiency of the assay, a positive control plasmid (pBL-TV-pc; Fig 6.3 b) was generated from pBL-TV-pc by adding a 23 bp sequence to the end of the left hand homology arm (see Section 2.5.3). This 23 bp sequence was homologous to the PT1-F primer but included a novel SpeI site immediately upstream of the left homology arm. While the pBL-TV-pc amplification products is almost identical to that expected from a targeted clone, it differs in that it is susceptible to cleavage by SpeI. Thus any false-positive PCR assays resulting from contamination of test assays with pBL-TV-pc can be eliminated by showing that the PCR is sensitive to digestion with SpeI.

Initially, amplification (PCR condition reported in Section 2.5.3) from pBL-TV-pc alone was used to confirm the correct size (2.5 kb) of the PCR product (Fig 7.3 d, lane 3). The PCR reaction was then repeated (same PCR condition of Section 2.5.8) mixing 7 pg pBL-TV-pc with 4 µl (≈ 0.3 µg) of genomic DNA isolated from untransfected K562, in order to asses the ability of the assay to detect small copy numbers of target in the presence of genomic DNA. 7 pg of pBL-TV-pc was calculated to be around $7x10^5$ copies of the plasmid. In comparison, 4 µl of genomic DNA were considered to have $4x10^4$ copies of each β-globin allele (Genomic DNA was extracted from 10^6 cells and resuspended in 100 µl therefore 4 µl would contain $4x10^4$ copies of each allele assuming 100% recovery). The expected 2.5 kb PCR product was detected on agarose gel in both reactions (plasmid alone or plasmid plus genomic DNA) Fig. 6.3.

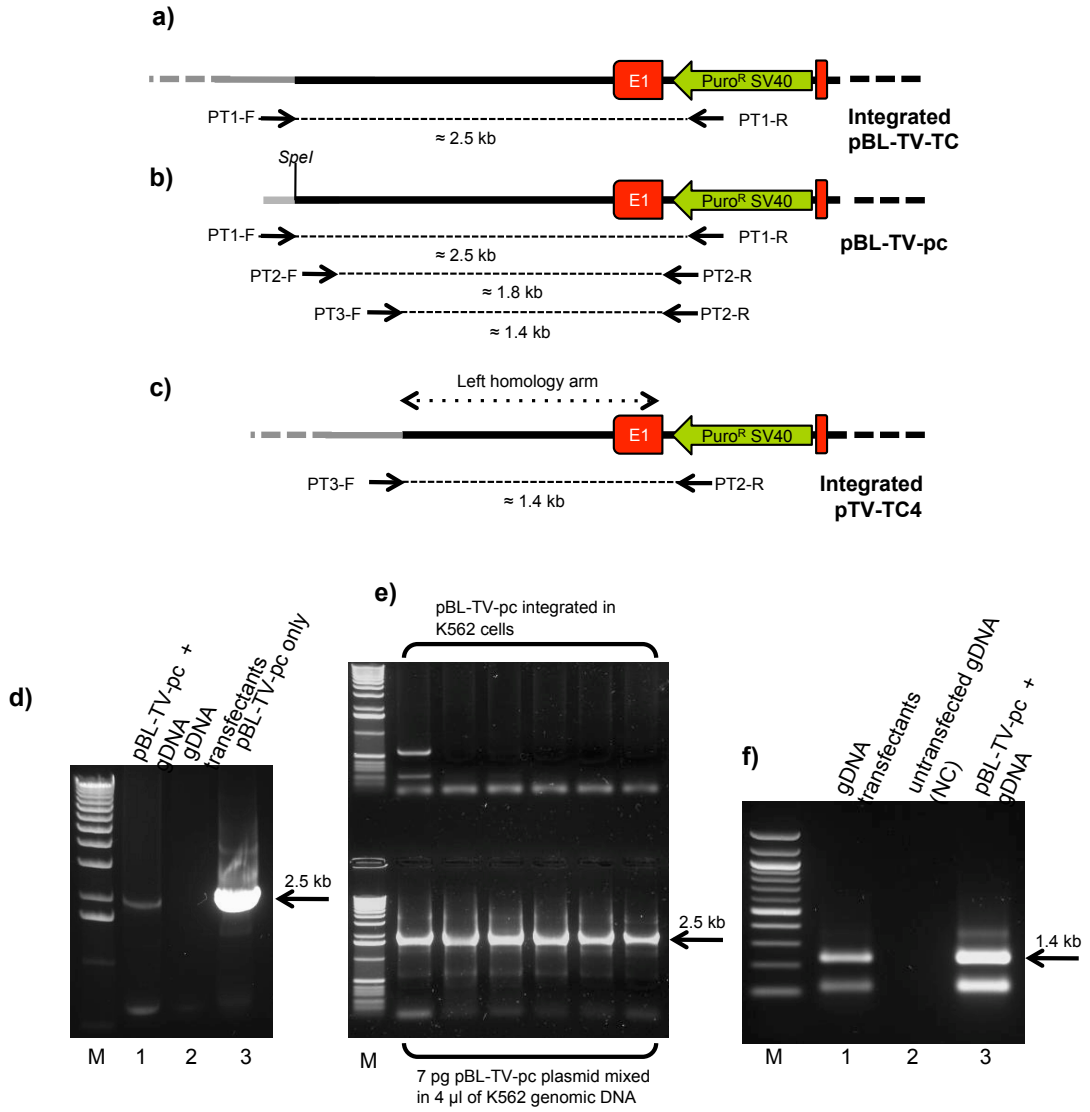


Figure 6.3 PCR assays for targeted integration.

a) Expected structure of pBL-TV-TC (left hand region only) after targeted integration. DNA is represented as in Fig 6.2**.** A 2.5 kb PCR product diagnostic for targeted integration is expected for primers PT1-F/PT1-R, as indicated. **b)** Structure of pBL-TV-pc (left hand region only). pBL-TV-pc has the same structure as pBL-TV-TC except for an additional 29 nucleotides (grey), cloned immediately upstream of the left homology arm, that binds to PT1-F and includes an SpeI site, as indicated. Expected PCR products using the alternative set of primers are shown below pBL-TV-pc map. **c)** Expected structure of the left hand portion of pBL-TV-TC after targeted integration. The PCR product from primers PT3-F and PT2-R diagnostic for targeted integration of pTV-TC4 is shown below the map. **d)** PCR reactions using PT1-F and PT1-R primers. Lane 1: 7 pg of pBL-TV-pc plasmid mixed in 4 µl (0.3 µg) of K562 genomic DNA. Lane 2: pBL-TV-pc integrated in K562 genomic DNA (0.3 µg). Lane 3: pBL-TV-pc plasmid alone (7 pg). **e)** Temperature gradient (50°C to 60°C) PCR using PT1-F and PT1-R primers. The PCR reaction resulted positive only for samples containing 7 pg pBL-TV-pc plasmid

mixed in 4 µl of K562 genomic DNA (bottom part of the picture). When pBL-TV-pc was integrated in K562 genomic DNA the PCR resulted negative (top part of the picture) except for a band in the first line (temp. 50°C) due to possible aspecific binding of the primers. **f)** Successful PCR reaction using PT3-F and PT2-R primers gave the characteristic 1.4 kb PCR product. Lane 1: pBL-TV-pc integrated in K562 genomic DNA (0.3 µg). Lane 2: genomic DNA from untransfected K562 cells (0.3 µg). Line 3: positive control, 7 pg pBL-TV-pc plasmid mixed in 4 µl of K562 genomic DNA.

In a third stage, pBL-TV-pc was stably integrated in the genome of both K562 and HT1080 cells (see Section 2.6.6). Cells were selected in puromycin for 15 days and then genomic DNA was extracted. 0.3 ug of each of these two genomic DNA samples were then used as a template in order to test the ability of the PCR assay to detect the presence of the construct when integrated in the genome. Unfortunately, no product was detected either for the K562 cells (Fig. 6.3 d, lane 2) or the HT1080 cells (not shown). To address the problem, different PCR conditions were explored: in a first attempt, a gradient of annealing temperature (between 50°C and 60°C) was used to optimise the reaction but the PCR still resulted negative (Fig. 6.3 e). Next, in the attempt to resolve possible secondary structure that could interfere with the reaction, 1.5 µl of Dimethyl Sulfoxide (DMSO) were added to the 50 µl PCR reaction. Also in this case the PCR reaction resulted negative (data no shown). As a last attempt, the concentration of $MgCl_2$ was increase from 0.5 mM to a maximum of 1.5 mM but, still, the PCR did not produce any product (data not shown). During the creation of the targeting construct pBL-TV-TC, the PCR reaction used to amplify the left homology arm resulted to be highly efficient therefore it was assumed that problems during PCR amplification using PT1-F and PT1-R were due to poor designing of the primers, to differences between polymerases (high-fidelity Phusion used in the creation of pBL-TV-TC vs Dream Taq used in the PCR screening) or to some secondary structure (e.g. hairpin-like structures) of the DNA region immediately upstream of the left homology arm (Nelms and Labosky 2011). Secondary structure are usually associated to high GC content therefore the sequence to be amplified by PT1-F and PT1-R was inspected for a possible high GC-content that could have affected the PCR reaction but the overall percentage of GC was found to be ≈ 35% (determined by the Ape plasmid editor software). The sequence was also searched for smaller specific GC-rich regions but none were identified (sequences of pBL-TV-TC and of the β-globin gene in Appendix 3). Therefore, in order to overcome these difficulties, it was decided to design a new set of primers and consequently to modify the original pBL-TV-TC targeting construct.

A new reverse primer (PT2-R) that would bind immediately upstream of the Puro^R cassette's ORF and two other forward primers (PT2-F and PT3-F) that would bind downstream of the first PT1-F (Fig 6.3 b) were created. The PCR reaction was repeated on the previously extracted genomic DNA (containing the integrated pBL-TV-pc) using primer pairs PT2- F/PT2-R or PT3-F/PT-2R. Under the PCR conditions chosen (see Section 2.5.3) the 1.8 kb PCR product from primers PT2-F/PT2-R was not detected (data not shown) but the expected 1.4 kb PCR product from primers PT3-F/PT2-R was readily detected (Fig. 6.3 f). Based on this result, a new targeting construct, pTV-TC4 (Fig 6.3 c), was generated from pBL-TV-TC. Apart from a shortened left homology arm of 1.4 kb that no longer extends far enough upstream to include the PT3-F sequence, pTV-TC4 is identical to pBL-TV-TC.

As described in detail in the following sections, gene targeting experiments were carried out co-transfecting cells with ZFN4-encoding sequences, in the form of mRNA or expression plasmids, and with variable amount of pTV-TC4 targeting construct. Cells were then selected in puromycin for 15 days and possible targeting events were detected through two different methods: PCR screening with primers PT3-F/PT2-R and Southern blot.

6.2.2 Transfection of K562 cells with pTV-TC4 and selection of Puro-resistant clones

K562 cells were transfected, by nucleofection (Section 2.6.8), with ZFN4 in the form of expression plasmids or mRNA, and with increasing amounts of pTV-TC4 targeting construct, in order to identify a possible correlation between the frequency of targeting events and the concentration of pTV-TC4. Three experiments were carried out with three different amounts of pTV-TC4 (2,4 and 8 µg in experiments 1, 2 and 3, respectively, Table 6.1). Each experiment involved two nucleofections, one in which nucleic acid encoding both ZFN4 subunits was used, the other, a negative control, in which 4 µg of an unrelated plasmid (pGP-FF) (experiments 1 and 2) or only 4 µg of ZFN4L (experiment 3) was used. When ZFN4(L) and ZFN4(R) were co-transfected, the nucleic acid was mRNA (experiments 1 and 2) or DNA (experiment 3).

The mRNA was provided by Sigma in the form of pre-mixed aliquotes containing 2 µg of each expression construct (ZFN4(L) and ZFN4(R)). In each nucleofection, one million K562 cells were nucleofected. Transfected cells were left to recover for 48h in an incubator at 37 °C in presence of 5% CO_2 and then puromycin selection was started. Transfected cells were divided into two batches: bulk selection on the first batch of around half a million cells was used in order to create a pool of puromycin resistant clones. The resulting population was split using some for storage in liquid nitrogen and some to prepare genomic DNA. The second batch of cells, instead, was plated in 96-wells plates at limiting dilution in presence of puromycin in order to obtain wells of cells derived from individual (or a very few) $Puro^R$ cells. Cells transfected with 2 and 4 µg of pTV-TC4 (experiments 1 and 2) were plated at density of 10^3 cells per well while cells transfected with 8 µg of pTV-TC4 (experiment 3), which were expected to have a higher percentage of targeting events, were plated at a density of $0.5x10^3$ cells per well.

Experiments 1 and 2 were carried out in parallel using cells from a single culture and nucleofection reagents from Lonza (Section 2.6.8). Expt 3 was carried out with a distinct culture, at a later date, and using nucleofection reagents from Mirus (2.6.8). To determine the transfection efficiency, 1 million cells from each of these two cultures were nucleofected with a GFP expression plasmid (pMaxGFP; Lonza) in parallel with the experimental nucleofections and under conditions that were otherwise identical. The percentage of GFP positive cells was determined by flow cytometically 48h after nucleofection (Section 2.6.8). Thus transient transfection efficiencies of 72% (expts 1 and 2) and 47% (expt 3) were determined.

Puromycin selection of the three targeting experiments was carried out for 15 days. Afterwards, the number of wells containing $Puro^R$ cells was counted for each transfection (Table 6.1, row 9). Using Poisson statistics these numbers were converted into the frequency of $Puro^R$ clones generated (Table 6.1, row 10 and 11). In experiments 1 and 2, expression of ZFN4 stimulated the frequency of $Puro^R$ cells 11.8- and 35.3-fold, respectively. This indicated that ZFN4 promotes the integration of pTV-TC4 into the genome (presumably at the β-globin locus, although this was yet to be confirmed) and suggested a possible correlation between the amount of targeting construct and the

number of ZFN4-dependent integration events. In experiment 3, however, ZFN4 expression resulted in only a 1.4-fold increase in the frequency of $Puro^R$ cells, largely as a result of the relatively high frequency in the absence of ZFN4 expression (14 to 22-fold higher than in experiments 1 and 2 [using row 11 data]). Thus, compared to experiments 1 and 2, the relatively high concentration of pTV-TC4 used in experiment 3 (or possibly the use of modified nucleofection conditions) appears to have promoted ZFN4-independent integration of pTV-TC4 much more than the integration in ZFN4-expressing cells which was almost unchanged compared to experiment 2. At this point it was not known whether $Puro^R$ cells generated when ZFN4 was expressed represented random or targeted integrations. To address this question, and to determine whether ZFN4-induced increases in $Puro^R$ cells represented increases in targeted integrations, it was necessary to test $Puro^R$ cells for targeted integration at the β-globin locus.

Table 6.1 Effect of ZFN4 on HBB gene targeting in K562 cells

		Expt 1		Expt 2		Expt 3	
1	ZFN4L mRNA	-	2	-	2	-	-
2	ZFN4R mRNA	-	2	-	2	-	-
3	ZFN4L DNA	-	-	-	-	4	2
4	ZFN4R DNA	-	-	-	-	-	2
5	pTV-TC4	2	2	4	4	8	8
6	Unrelated plasmid (pGP-FF)	4	-	4	-	-	-
7	Transfection efficiency (%)	72	72	72	72	47	47
8	Number of plated cells/well	$1x10^3$	$1x10^3$	$1x10^3$	$1x10^3$	$0.5x10^3$	$0.5x10^3$
9	$Puro^R$ wells/96-well plate	5/96	43/96	3/96	63/96	33/96	47/96
10	Average $Puro^R$ cells/well	0.053	0.59	0.032	1.06	0.34	0.49
11	Frequency of $Puro^R$ cells (%)	0.005	0.059	0.003	0.106	0.068	0.098
12	PCR+ clones/$Puro^R$ clones	0/5	19/20	1/3	19/20	0/18	16/20
13	Relative targeting frequency (%)	0	95	33	95	0	80
14	Absolute targeting frequency (%)	0	0.077	0.001	0.139	0	0.167

[1-6] 1 million cells were transfected with the indicated nucleic acids (ug), distributed in 96-well plates and selected in puromycin

[7] Percentage of GFP+ cells at 48h in a parallel nucleofection (see text).

[10] $-\ln F_0$ where F_0= fraction of wells with no $Puro^R$ cells (Poisson distribution).

[11] (row 10/row 8) x 100

[13] Row 12 expressed as percentage

[14] Product of rows 11 and 13 divided by row 7

6.2.3 PCR screening of puromycin resistant cells

The PCR assay based on primers PT3-F and PT2-R (section 6.2, Fig. 6.3 c) was then used to screen Puro^R cells and discriminate genuine targeting events from Puro^R cells generated by the random integration of pTV-TC4.

In order to reduce both time and costs, it was decided to screen for possible targeting events using a colony PCR approach where lysates obtained from a small number of puromycin resistant cells would be used as a direct template for the PCR reaction (Section 2.4.7). Increasing numbers (10^3, 10^4 and 10^5) of K562 cells stably transfected with pBL-TV-pc positive control (described in Chapter 6.2) were used to identify the minimum amount of cells to be used in the assay. PCR reactions using lysates from 10^3 cells gave a very faint signal while lysates from 10^5 cells originated extensive smears when the PCR products were run on agarose gel (data not shown). Of this three, the most effective number of cells for the detection of targeting events was found to be 10^4 cells and therefore 10^4 of Puro^R cells from wells of the 96-well plates in experiments 1 and 2 (Table 6.1) were analysed by PCR.

In practice, however, it was found that the diagnostic 1.4 kb PCR product, when detected by colony PCR, was often very faint (Fig 6.4 a), in contrast to the signal obtained in the same assay using positive control cells (K562 with randomly integrated pBL-TV-pc) The colony PCR assay was thus close to its limit of sensitivity in detecting targeted cells and it was necessary to carry out the colony PCR assay multiple times in order to distinguish those cells that were consistently unable to generate the 1.4 kb product (PCR-negative) from those that could, albeit inconsistently, generate the 1.4 kb product (PCR-positive). Although the PCR-positive cells were almost all from cells that had been transfected with ZFN4-encoding mRNA, the colony PCR assay was considered too time consuming and it was decided instead to expand cells until there were enough (1 million cells) to prepare genomic DNA and then perform a standard PCR reaction.

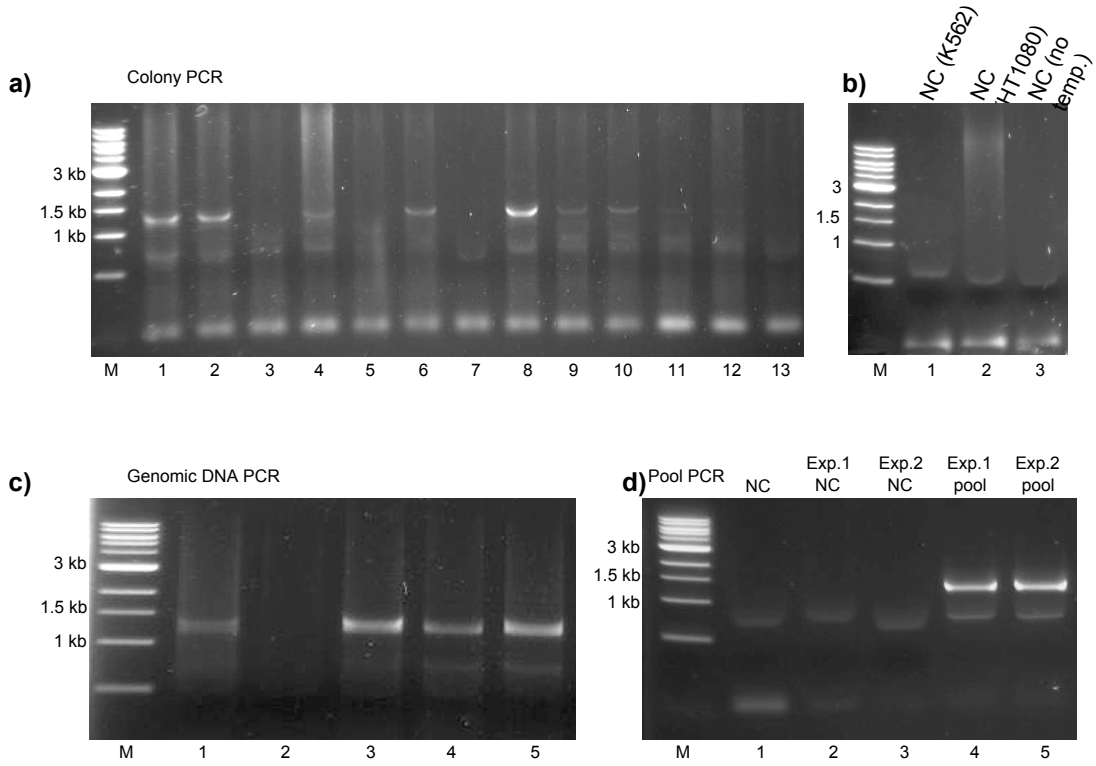


Figure 6.4 PCR assay for the detection of targeting events.

a) Example of screening for targeting events using colony PCR assays. An agarose gel electrophoretic analysis is shown of 13 assays, each on 10^4 cells from a single well of a 96-well plate in experiment 2 (Table 7.1, + ZFN4 mRNA). Although the diagnostic 1.4 kb PCR product was not detected in samples 3,5,7 and 13, when the analysis was repeated a second time the samples resulted positive (not shown). **b)** Example of negative controls during PCR screening. Each PCR reaction was paired with negative controls in order to rule out possible false positives due to contamination. Line 1: genomic DNA from untransfected K562. Line 2: genomic DNA from untransfected HT1080. Line 3: PCR reaction without any template. **c)** Example of screening for targeting events using PCR assays on genomic DNA isolated from cultures originating from single wells from experiment 3 (Table 7.1; + ZFN4 plasmids). Of the 5 cultures analysed only one (lane 2) was negative and when the analysis was repeated (not shown) this sample was again negative, confirming it to be a non-targeted clone. **d)** PCR assay carried out on genomic DNA originated from pools of cells. Lane 1: negative control, untransfected K562 cells. Lane 2: K562 transfected with only 2 µg of pTV-TC4 (second negative control). Lane 3: K562 transfected with only 4 µg of pTV-TC4 (third negative control). Lane 4 and 5: K562 transfected with ZFN4(L) +ZFN4(R) and 2 or 4 µg pTV-TC4, respectively.

When the PCR assay was repeated on genomic DNA the 1.4 kb bands present on the agarose gel were clearer and the intensity was more constant between samples (Fig. 6.4 c). To address the possibility that cultures testing positive reflected contamination (either with a plasmids such as pBL-TV-pc, or genomic DNA from a targeted culture) negative control assays were carried out in genomic DNA from untransfected K562 and HT1080 cells and these always failed to generate the 1.4kb PCR product. This excludes contamination of PCR reagents but cannot exclude the possibility that some genomic DNA preps from non-targeted cultures were contaminated. Although the original targeting construct (pBL-TV-TC) was designed so that contamination with pBL-TV-pc could be indicated by sensitivity to SpeI-digestion, this was not possible with the final targeting construct, (pTV-TC4). Subsequent analyses by Southern blots (see Section 6.2.4) however suggest that contamination was not a problem.

Initial PCR screening of genomic DNA was carried out on the pools of Puro^R cells from experiments 1 and 2 (section 6.2.2). The 1.4 kb band, characteristic of possible targeting events, was found only in samples coming from cells transfected with both ZFN(L) and ZFN(R) and pTV-TC4 (Fig. 6.4 d)., This result clearly suggests that the increased frequencies of Puro^R cells seen when cultures were transfected with ZFN4 mRNA were indicative of increased frequencies of targeted integration.

To estimate the proportion of Puro^R cells that had undergone gene targeting as a result of ZFN4 expression the same PCR assay was used to screen genomic DNA obtained from the cells in a selection of puromycin resistant wells from each of the three experiments (Table 6.1, row 12). Based on the proportion of PCR-positive wells detected in this way, the relative targeting efficiency (% of Puro^R cells that were targeted) and absolute targeting efficiency (% of transfected cells that were targeted) were estimated for each nucleofection (Table 6.1, rows 13 and 14, respectively). In the negative controls (no ZFN4 expression or expression of only ZFN4(L)) the relative targeting efficiency was generally low: 0/5 (0%), 1/3 (33%), and 0/18 (0%) in experiments 1, 2 and 3, respectively. Given the small sample size, particularly in experiment 2, these values are necessarily rough estimates.

Nevertheless, the value for experiment 3 (0/18) provides strong evidence that the relatively high frequency of Puro^R cells obtained in the absence of ZFN4 (0.068%) reflects random integration of pTV-TC4. When ZFN4 was expressed, however, the relative targeting frequency was high (80-95%), and the absolute targeting efficiencies were hugely stimulated, in all three experiments. These data therefore clearly suggest a strong correlation between the expression of ZFN4 and the frequency of gene targeting of the β-globin gene.

6.2.4 Confirmation of targeting events by Southern Blot analysis

Although they are widely used to detect targeting events, PCR assays can give false positives, for instance by contamination. Even if contamination is ruled out, it has been shown that under some conditions a clone with a randomly integrated targeting construct can generate the PCR product indicative of a targeting event (see Section 6.3.2) (Frohman MA and Martin GR 1990). It was therefore important to confirm by another method that putative targeted clones identified by PCR were genuinely targeted. For this reason, targeting of the β–globin gene was confirmed using Southern Blot analysis. A specific radioactive probe of ≈ 600 bp was designed to bind immediately upstream of the region of homology between the β-globin gene and the left arm of the pTV-TC4 targeting construct (Section 2.3.4). Digestion of genomic DNA with HindIII restriction enzyme would generate, in the untargeted β-globin gene, a fragment of 7.8 kb while digestion of successful targeted loci would yield a fragment of 3.7 kb due to the introduction of a novel HindIII sites located immediately upstream of the puromycin resistant cassette (Figure 6.5 a). Both of these fragments should bind the probe.

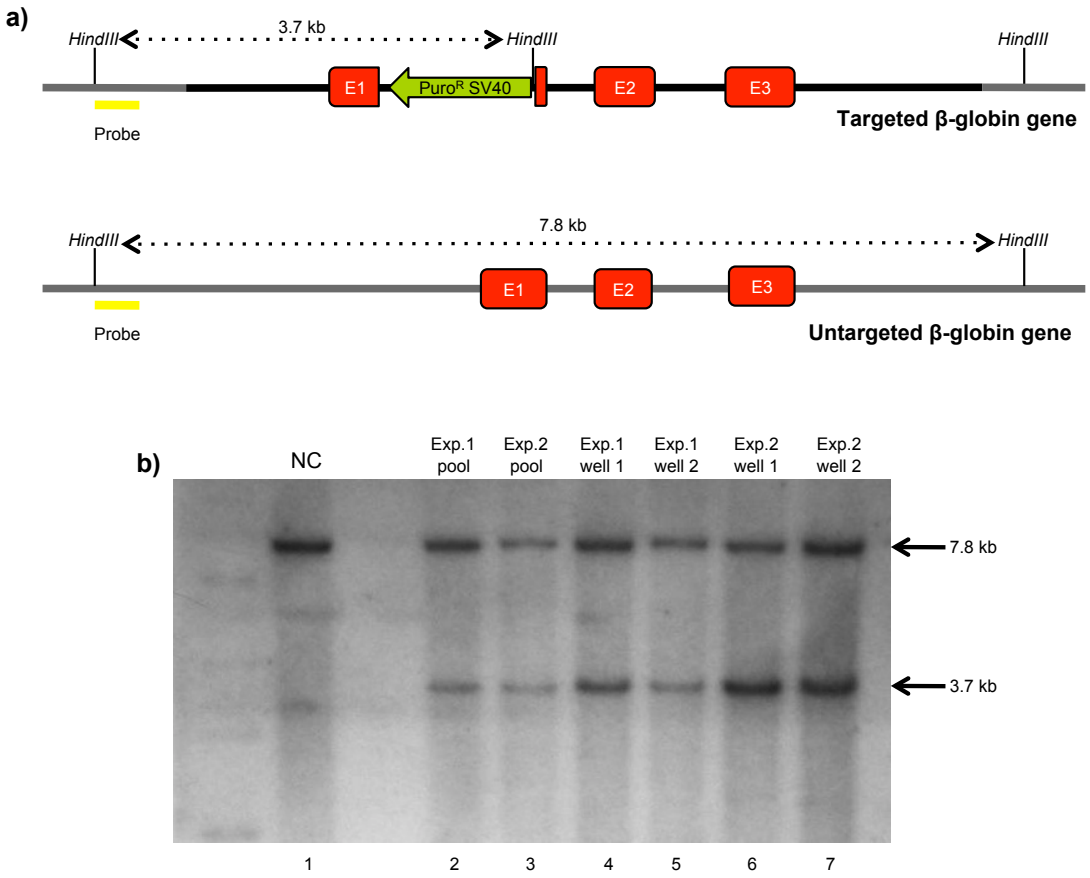


Figure 6.5 Southern blot analysis of K562 and PCR-positive Puro^R derivatives.

a) Map of β-globin gene that has (top) or has not (bottom) undergone gene targeting with pTV-TC4. DNA is depicted as in Fig 6.2 and the position of the probe used for Southern blotting (yellow) and diagnostic HindIII fragments are indicated. **b)** Autoradiograph of an agarose gel that was blotted and probed with the ^{32}P-labelled probe shown in a). Samples loaded and electrophoresed on the gel were HindIII-digested genomic DNA from untransfected K562 cells (lane 1, NC) or from Puro^R cells obtained after transfections of K562 (see Table 6.2) with ZFN4 expressing mRNA and either 2 µg (Experiment 1: lanes 2,4,5) or 4 ug (Experiment 2: lanes 3,6,7) of targeting construct. The genomic DNA was isolated from pools of Puro^R cells (lanes 2,3) or from individual wells obtained after limiting dilution (lanes 4-7).

Southern blot analysis was carried out on 6 PCR-positive samples obtained following the expression of ZFN4 in experiments 1 and 2 (Table 6.1). For each experiment genomic DNA was isolated from a pool of $Puro^R$ cells and from two of the wells of PCR positive $Puro^R$ cells. The DNA was digested overnight with HindIII and then hybridized with the ^{32}P-labelled probe (see Section 2.4.5). The Southern blot confirmed targeting for all samples tested, showing the expected bands for wild type (7.8kb) and targeted β-globin (3.7 kb) alleles while the negative control (genomic DNA originated from a pool of cells from experiment 1 that were transfected with the negative control plasmid pGP-FF and pTV-TC4) yielded only the wild-type band at 7.8 kb (Figure 6.5). The relative intensity of the two bands was generally similar suggesting that, as expected for clones that had undergone gene targeting at one of their two β-globin alleles, or for populations consisting mostly of such clones. Not so for lane 1 where the 7.8 kb band resulted to be more intense that the 3.7 kb band. This is due probably to the fact that the cell population of this specific pool of cells was not originated by only targeting events but also by $Puro^R$ cells where pTV-TC4 was randomly integrated in the genome. But this does not agree with Table 6.2 that shows relative targeting frequencies were 95% in both expt 1 and expt 2.

6.2.5 Measure of cell viability after transfection with ZFN4

From the experiments carried out during the GFP correction assay (Chapter 5) an increased tendency towards cell death was noticed for samples expressing ZFN4 relative to negative controls samples that were not expressing any nucleases. These observations (based on the appearance of cells under phase contrast microscopy 48 h after transfection) were highly qualitative. Nevertheless, the GFP assay was repeated many times and the cell-death phenomenon was consistently observed in each sample expressing ZFN4, but not in the negative controls. Cyto-toxicity induced by ZFNs has been previously reported in several studies and has been measured using a range of different assays including cell viability measurement, quantification of

apoptosis or through the use of specific antibodies for phosphorylated histone H2AX or p53 tumour suppressor-binding protein 1 (Bibikova, Golic et al. 2002; Alwin, Gere et al. 2005; Beumer, Bhattacharyya et al. 2006; Miller, Holmes et al. 2007; Szczepek, Brondani et al. 2007; Cathomen and Joung 2008). In order to quantify a possible negative effect of ZFN4 on cell viability, HT1080 cells were transfected (using nucleofection) either with pZFN4(L) and pZFN4(R) (2 μg each) or with 4 μg of pZFN4L alone. Immediately after transfection the cells were counted with a haemocytometer (Section 2.6.5) and then were plated with serial dilutions onto 10 cm plates. Following 10 days culture without selection, numbers of colonies were counted and used as an indication of cell viability/colony forming units (Table 6.2). The collected data suggested that co-expression of the two ZFN4 sub-units could have a negative effect on cell survival: plates containing cells that were expressing only one of the two sub-units had almost double the amount of colonies. These figures suggest a possible apoptotic effect due to DSBs produced by ZFNs but, with this single experiment, it was not possible to properly assess and quantify it. The experiment should be repeated several times in order to exclude possible experimental errors due to differences in transfection conditions or errors during the cell counting.

Table 6.2 Cell viability assay

[a] number of cells plated	[b] ZFN4(L)	[c] ZFN4(L) + ZFN4(R)
200	29	15
500	58	27

[a] 1 million HT1080 cells were transfected (nucleofection) with only one (ZFN4(L)) or with both two sub-units (ZFN4(L) + ZFN4(R)), were counted immedialty after using an haemocytometer and were plated at two different densities (200 cells or 500 cells per plate). Colony forming units were counted after 10 days.
[b] Samples were transfected with 4 μg of ZFN4(L)
[c] Samples were transfected with 2 μg of ZFN4(L) and 2 μg of ZFN4(R)

6.3 Discussion

6.3.1 Creation of a targeting construct specific for the β-globin gene

Detection of HR events, and therefore detection of gene targeting, in mammalian cells can be quite difficult due to the very low frequencies usually obtained. Although the use of ZFNs should increase the frequency of gene targeting of around 1000-fold, the low cutting efficiency of ZFN4 noticed during the Cel-I assay (Chapter 4) suggested that a correspondingly low frequency of targeting events might be obtained. For this reason a specific targeting construct carrying a selectable marker flanked by two extended regions of β–globin homology was developed to allow rapid and easy screening of possible targeting events. The targeting construct was designed taking in account two different aspects that could have affected the frequency of gene targeting: the amount of heterology between the target gene and the targeting construct and the distance between the ZFN cleavage site and the position of the resistance cassette to be introduced. The efficiency of gene addition at a specific locus is in inverse proportion to the length of foreign DNA that has to be introduced: a study from 2007 showed that targeted integration efficiencies at the same locus were affected by the length of the heterology in the constructs: the introduction of a simple 12 bp tag sequence would occur with a 15% frequency but the introduction of a 1.5 kb gene would only occur with a 5% frequency (Moehle, Rock et al. 2007). An extended stretch of uninterrupted homology flanking the heterology region should improve the frequency of HR (Waldman and Liskay 1988) and therefore it was decided that the $Puro^R$ cassette would be flanked by two long left and right homology arms. The relationship between the distance of the DSB in the target gene and the heterology region to be introduced has been analysed in a paper from 1998 (Elliott, Richardson et al. 1998). In the study, a targeting construct was used to correct a defective Neo^R cassette integrated in mouse embryonic stem cells (ES cells) and at the same time to introduce silent mutations (unique restriction sites) at different distances from the DSB. Elliot

et al. showed how, already at a distance of 100-200 bp from the DSB, mutations carried by the targeting construct are introduced with very low frequencies (only 16-17% of the Neo^R colonies incorporated the mutation) compared to mutations only 8 bp away (83% incorporated the desired mutation) (Elliott, Richardson et al. 1998). For this reason, the two homology arms were design to introduce, during HR, the $Puro^R$ cassette directly at the locus of the DSB (Fig. 6.1 a and 6.2).

Although this type of design should maximise the frequency of gene conversion at the target locus, a lasting factor could negatively affect the efficiency of HR. When the DSB is introduced by ZFN4, the two ends at the cleavage site will have in total a 60 bp region of heterology with the targeting construct (see Fig. 6.2 a). During the HR repair pathway, the two regions of dsDNA at the DSB will be subject to 5' to 3' end resection in order to produce long stretches of 3'-end ssDNA that will then invade the DNA repair template during the D-loop formation and be used as a priming agent during DNA synthesis (more details about DNA repair mechanism in Section 1.8). However, in order to be effective as a primer, the 3'-end should be completely homologous to the repair template (Paques and Haber 1999). In the case of the β-globin targeting construct the presence of this 60 bp heterology could in theory reduce greatly the frequencies of gene targeting. Although the cellular mechanism is not completely clear, studies in two model organisms (*S. Cerevisiae* and *S. Pombe*) have identified two sets of proteins involved in the processing of heterologous 3'-end ssDNA: Rad1p-Rad10p and Rad16p-Swi10p in *S. Cerevisiae* and *S. Pombe*, respectively. After invasion of the D-loop by the 3'-end ssDNA, these two sets of proteins (and probably their human equivalent XPF-ERCC1) have the function of recognising and remove the regions of mismatch (Paques and Haber 1999; Farah, Cromie et al. 2009). Although their effect on HR frequencies is not yet clear, the presence of heterology regions at the ZFNs' cleavage sites is going to be a constant feature of most of the gene targeting constructs. This is due to the fact that, after gene targeting has occurred, additional ZFN cutting has to be avoided in order to reduce the risk of damaging the newly corrected gene. This is usually accomplished by introducing silent mutations at the ZFN binding site that will

prevent the ZFN to cut both the targeting construct and the corrected endogenous gene.

In order to discriminate between genuine targeting events and random integration, a specific PCR reaction was developed. A forward primer (PT1-F) binds outside the left region of homology and a reverse primer (PT1-R) binds immediately upstream of the selectable marker (puromycin). Unfortunately, when the PCR reaction was tested using the positive control pBL-TV-pc, it was immediately clear that the PCR reaction was able to detect the positive control in its plasmid form but not when integrated in the genome of transfected cells. To circumvent this problem different PCR conditions and reagents were tested (section 6.2.1) but still the PCR reaction resulted to be unreliable. GC-rich sequences tend to reduce the efficiency of the PCR amplification due to the formation of secondary structures but when the sequence to be amplified was checked for possible GC-rich regions none were found. The use of the high fidelity Phusion polymerase was considered but the possible beneficial advantage in using this more efficient reagent was counterbalanced by the relative high costs of the product. This factor was considered especially relevant due to the amount of puromycin resistant cells to be screened. Therefore, in order to overcome the difficulties in the PCR amplification, a new set of primers (PT2-F, PT3-F and PT2-R) was designed (for more details see Section 6.2.1 and Fig. 6.3). The newly created PT2-F and PT2-R were tested but failed in producing any PCR product. On the other hand, when PT3-F and PT2-R primers were used, the PCR reaction became more efficient and reproducible on samples containing pBL-TV-pc integrated in the genome. Although this last PCR reaction was finally effective in producing the desired product (1.4 kb), it was still not clear which factors were affecting the two previous sets of primers (PT1-F, PT1-R and PT2-F, PT2-R). Following these results the original targeting construct (pBL-TV-TC) was modified (to make pTV-TC4) reducing the length of the left homology arm from 2 kb to 1.3 kb in order to take in account the modified PCR assay.

6.3.2 PCR screening, Southern blot and gene targeting frequencies

Initially, the decision of performing the PCR screening directly on cell-lysates was made with the aim of reducing the time spent in expanding the cells (10^4 cells for the colony PCR compared to ≈ 10^6 cells for genomic extraction) and the costs for the genomic extraction. Unfortunately, although the preliminary tests carried out on K562 cells stably transfected with pBL-TV-pc resulted consistent in producing the characteristic 1.4 kb band, when the colony PCR was performed using cells from puromycin resistant wells the signal greatly varied between different samples (Fig. 6.4 a) and in many cases was found to be at the limit of detection or completely absent. Two possible factors could be responsible for the discrepancy: the preliminary tests and the successive positive controls were performed using a pool of K562 cells with the pBL-TV-pc plasmid randomly integrated in the genome. It could be possible that some of the K562 cells had multiple copies of the pBL-TV-pc plasmid integrated in the genome or that some of the pBL-TV-pc molecules integrated in regions of the genomic DNA that were more “accessible” to the PCR screening compared to the β-globin locus. A second factor affecting the PCR screening could be related to human errors occurred during the preparation of the samples: the PCR screening was performed on a substantial number of samples and errors in the counting of the cells or during the isolation of the pellet could have affected the performance of the assay. The first step of the colony PCR protocol (Section 2.4.7) involved the spinning of the sample to remove the media, washing with PBS and again spinning and removal of the buffer. Due to the small size of the cell pellet, it is possible that part of the cell population was lost during these initial steps effectively reducing the number of genomic template.

Gene targeting frequencies were assessed by transfecting K562 cells with ZFN4 and increasing amounts of targeting construct pTV-TC4 (section 6.3). The relative frequencies of targeting events for these experiments were used as a way to assess the risk of random integration during a possible

therapeutic approach. Data collected during limiting dilution in multi-wells plates of cells transfected with 2 or 4 μg of targeting construct (experiments 1 and 2) showed that the expression of ZFN4 was responsible for the increase in number of $Puro^R$ colonies. Out of a total of 192 wells plated at limiting dilution (Tabe 6.2 experimets 1 and 2) only 8 became $Puro^R$ after transfection with only the targeting construct, compared with 106 wells when transfected with the targeting construct and ZFN4 mRNA. Moreover, the PCR assay showed that a high percentage (≈ 95%, Table 6.1, row 13) of $Puro^R$ wells derived by cells transfected with both ZFN4 and targeting construct were due to genuine targeting events. The data show that, at these concentrations of pTV-TC4 (2 and 4 μg), gene targeting of the β-globin gene would not be greatly affected by the risk of oncogenesis induced by random integration. Random integration of the targeting construct could occur at a sensitive genomic site leading to the activation of a possible proto-oncogene. This occurrence has been reported in a famous case study where children with severe combined immunodeficiency (SCID) were treated with a retrovirus-mediated gene therapy approach (for more details Section 1.4). Although the technique was effective in curing their disease, 4 patients developed leukemia from insertional mutagenesis (Hacein-Bey-Abina, Garrigue et al. 2008). Despite the fact that gene targeting does not rely on the random integration of a therapeutic vector, the risk of insertional mutagenesis is still a factor to take in account due to the large number of DNA template present in the cell during gene correction. This negative aspect could be possibly seen in experiment 3: the difference between cells expressing or not expressing ZFN4 in presence of 8 μg of pTV-TC4 (Table 6.1 expt 3) was much less pronounced than the previous experiments. In a therapeutic context, this increase in random integration would mean a higher risk of insertional oncogenesis. The use of high concentrations of targeting construct to promote higher frequencies of gene targeting must therefore be limited to ensure that such risks are avoided.

Although the PCR screening assay was based on specific primers that should only promote amplification when HR events occurred, in reality a number of artefacts could produce false positives affecting the reliability of the assay. Contamination of the samples to be screened with the pBL-TV-pc positive control or with genomic DNA from other targeted cells is the main risk

but also possible false positive due to Polymerase Halt-Mediated Linkage of Primers (PHLOP) should be considered. During the initial phases of the PCR reaction, if the Taq polymerase is prematurely interrupted during synthesis, it is possible that the two truncated DNA strands (one for each primer) could share regions of partial homology. The annealing of the two truncated strands in the following cycles could be used as a template creating false positives (Martin 1990). To rule out possible PCR artefacts due to contamination of the samples with the positive control or due to PHLOP artefact, it was decided to corroborate the PCR assay with Southern Blot analysis and a selection of PCR positive samples were selected for further investigation (Chapter 6.2.5). The use of Southern blot analysis is widely employed in conjunction to PCR assays and DNA sequencing to confirm targeting events and has been used also by Zou et al. and Sebastiano et al. in their β-globin targeting experiments (Sebastiano, Maeder et al. 2011; Zou, Mali et al. 2011). The selection of samples that resulted positive in the PCR assay were found also positive in the Southern blot assay while the negative control showed only the 7.8 kb band characteristic of the wild type β-globin gene. The relative similar intensity of the bands for the 7.8 kb and the 3.7 kb bands indicated that, as expected, the targeting of the β-globin locus in this cells was limited to only one of the two alleles. This was not true for the pool of cells derived by experiment 1 (sample in lane 2, Fig. 6.5) where the 7.8 kb band (wild type β-globin gene) resulted more intense than the 3.7 kb band. This fact may suggest a higher presence of $Puro^R$ cells due to random integration of the targeting construct but, at the same time, would contradict the data obtained with the PCR assay (Section 6.2.3, Table 6.1 row 13). In order to clarify the reason of this discrepancy the Southern blot assay should be repeated and, to increase the strength of the statistical analysis, the PCR assay should be extended at a wider population of $Puro^R$ wells.

Despite the fact that gene targeting levels are affected by the type of cell line to be used, due to both varying levels of HR and ZFN accessibility to the target site, it is useful to compare the data obtained in this research and the one published by Zou et al. in 2012 (Zou, Mali et al. 2011). Zou et al. used the same ZFN pair (ZFN4) to produce gene targeting in iPS cells derived from bone marrow stromal cells (MSCs) from a patient with sickle cell disease

(SCD). Despite the use of a targeting construct containing a negative selection for the counter-selection of randomly integrated vectors, only 4 out of 300 selected clones (a relative targeting frequency ≈1%) resulted positive in the PCR screening for targeted integration at the β-globin locus (Zou, Mali et al. 2011) and only 1 of these resulted positive in the Southern blot analysis (the other 3 clones stopped proliferating or did not show up in the Southern blot analysis). In a similar study, Sebastiano et al. targeted the β-globin locus in iPS cells using a different set of 3 different ZFNs (ZFNA, B and C, see Table 6.3) (Sebastiano, Maeder et al. 2011). Although the relative frequency of targeting events was 10-fold higher than the frequency measured by Zou et al, of all selected clones, on average, only 13% resulted positive in the PCR screening and the frequency was significantly less than the figure of 95% obtained for this project in K562 cells. These striking differences in targeting frequencies can not be explained only by differences in transcriptional status of the targeted gene because the β-globin locus is silent in both K562 (Rutherford, Clegg et al. 1981) and iPS cells. Zou et al. suggest, as an explanation, a possible silencing effect of the selectable markers at the target locus. Possible epigenetic modification could target preferentially unfavourable loci (in this case the β-globin gene) and this could explain the high number of resistant clones due to randomly integrated vectors compared to the number of targeted ones. These low levels of relative frequency could be explained by a generally low level of HR in iPS cells or a relative low frequency of HR events at loci that are not actively expressed (Hockemeyer, Soldner et al. 2009).

Table 6.3 Frequencies of ZFN-induced gene modification at the β-globin locus

ZFN	[a] Cell type	Cel-I assay	[b] Selection	[c] Gene targeting		Ref.
				Rel.freq.	Abs.freq.	
[d] ZFN4	K562	1%	Puro	95%	0.08%	1
[e] ZFN4	K562	1%	Puro	95%	0.14%	1
[f] ZFN4	K562	1%	Puro	80%	0.17%	1
ZFN4	iPSC	-	Hygro-GCV	0.3%	-	2
ZFNA	293T	13%	-	-	-	3
ZFNB	PriFib	1%	-	-	-	3
ZFNC	PriFib	2%	-	-	-	3
ZFNA	iPSC	-	Neo	8%	-	3
ZFNB	iPSC	-	Neo	21%	-	3
ZFNC	iPSC	-	Neo	10%	-	3

[a] K562 (erythroleukeia), 293T (embryonic kidney), iPSC (induced pluripotent stem cells from SCD patients), PriFib (primary fibroblasts)
[b] Puro (puromycin), Hygro (hygromycin) GCV (gancyclivir), Neo (neomycin)
[c] Gene targeting frequencies: Rel.freq. (relative frequency), Abs.freq. (absolute frequency)
[d] ZFN4 co-transfected with 2μg targeting construct (pTV-TC4)
[e] ZFN4 co-transfected with 4μg targeting construct (pTV-TC4)
[f] ZFN4 co-transfected with 8μg targeting construct (pTV-TC4)
1 Data reported in this project
2 (Zou, Mali et al. 2011)
3 (Sebastiano, Maeder et al. 2011)

Absolute frequencies of targeting events at the β-globin gene were highly stimulated by expression of ZFN4 (from 0.001% in absence of ZFN4 to values of 0.16% in cells expressing ZFN4). These frequencies are ≈10 times lower than the frequency of cleavage of the β-globin gene as estimated by Cel-I assay (Chapter 4). These values were lower than expected: targeting of a different locus (F9 gene) in K562 cells by Li et al. were only 3 times lower than the frequency of cleavage (Li, Haurigot et al. 2011). Although frequencies increased with the increase in amount of targeting construct (from 0.07% to 0.16%), also the number of randomly integrated events tended to increase therefore ZFNs with increased cutting efficiencies will be necessary to improve absolute frequencies.

Although the data collected in this project, and those described by Zou et al. and Sebastiano et al., are encouraging for the development of a therapeutic gene targeting approach, the low absolute frequencies obtained will require the use of targeting constructs with selectable markers that can be used to easily isolate possible targeted clones. The absolute frequencies for the Zou and Sebastiano papers were not reported but it could be assumed that they were lower than the one identified in this project due to the difficulties encountered when working with iPS cells.

As a way to compare the aim and results of this thesis with the actual status of ZFNs induced gene-targeting techniques, it could be useful to briefly examine the results obtained by Li et al. The study was based on the delivery of ZFN and targeting construct through an AVV8 delivery method on a mice model of haemophilia. Therefore the conditions were quite different from the ones examined in this project (Li, Haurigot et al. 2011). Nevertheless, the differences between cleavage and targeting efficiency and the differences of targeting efficiencies between different cell lines can be compared to the data obtained in this project and in Zou et al. paper. The ZFN used by Li et al. had a cleavage efficiency, measured by Surveyor assay, of 45% (one of the highest reported till now). When ZFN-induced gene targeting was measured in K562, the absolute frequency was found to be 18% but when the targeting efficiency was measured in the mice model it dropped to ≈2%. These figures show the actual difficulties encountered when moving from a model study to a more complex system or when moving from one cell type to another and suggest the need of a better understanding of the processes and conditions that affect HR in different cell lines. Increasing the efficiencies of novel ZFNs is only one of many aspects to be addressed: improved targeting construct design, improved delivery methods and a deeper understanding of HR mechanics are all factors to be considered in order to improve the safety and efficiency of future therapeutic approaches.

6.3.3 Measure of cell viability after transfection with ZFN4

Various studies have reported a tendency of cells expressing functional ZFNs to undergo apoptosis and different approaches have been used to investigate and measure the induced cell death (Bibikova, Golic et al. 2002; Alwin, Gere et al. 2005; Beumer, Bhattacharyya et al. 2006; Miller, Holmes et al. 2007; Szczepek, Brondani et al. 2007; Cathomen and Joung 2008). In the case of this project, possible ZFN4-induced cell death was noticed during by phase-microscopic examination of cells expressing ZFN4, both during the GFP correction assay (described in Chapter 5) and also during the targeting experiments described in this chapter. The first observations had only a qualitative aspect and until then it was not possible to confirm or quantify the possible negative effect on cell survival. In order to confirm the occurrence of ZFN4-induced cell death, and obtain an estimate of its extent, a simple cell-viability assay based on serial dilutions was used. The results (Table 6.2) suggested a ~50% reduction in plating efficiency as a result of ZFN4 expression. This assay is by no means a definitive approach to investigate ZFN-induced apoptosis, and further and more precise analysis would be required. The experiments should be repeated in order to acquire enough data to produce a statistical significance and a second approach, like the employment of apoptotic markers, should be used to confirm the initial observation. Although important for a possible transition to a therapeutic approach, the in-depth assessment of a possible negative effect of ZFNs on cell survival was considered less relevant for the experiments detailed in this chapter and therefore was limited to this brief test. The peak in ZFNs expression, and therefore the possible apoptotic effect is thought to occur during the first 48h immediately after transfection and should be taken in consideration when analysing data from experiments with adherent cells. This was not considered relevant for the experiments that were carried out on K562 suspension cells. As fully explained in Section 6.2.3, 1 million K562 cells were transfected (using nucleofection) with or without both pZFN4(L) and pZFN4(R) and different amounts of targeting construct, they were left to

recover for 48h and then, before been subject to limiting dilution and puromycin selection, they were counted with a haemocytometer in combination with a cell-viability dye (Section 2.6.5). This approach has been used to take into account the different levels of cell survival and to plate, after limiting dilution, the same number of living K562 cells before starting selection. Nevertheless the 2-fold loss of viabilty is of interest. Usually ZFN-induced cell death has been ascribed to excessive cleavage at off-target sites in the genome (Cathomen and Joung 2008) but the use of forced heterodimers in the FokI domain (as was the case in ZFN4) should reduce the frequency of off-target cleavage. Low frequencies of off-target cleavage by ZFN4 were suggested, by different means, both in this project (in the case of δ-globin; Section 4.2.5, Fig. 4.6 b) and by Zou et al. (in ε,γ and δ-globin gene). The Surveyor assay used in this project (Chapter 4) is by no means a definitive way to assess off-target cleavage events at the δ-globin locus due to its low accuracy (the assay was already barely able to detect on-target cleavages) but the approach taken by Zou et al. supports the idea of low levels of ZFN4-induced off-target cleavages. In their paper, a GFP correction assay, similar to the one described in Section 5 of this thesis, was used to investigate cleavage events at the β-globin gene and at highly homologous sites of the β-locus genes (ε,γ and δ-globin genes). Significant levels of ZFN4-induced cleavage were detected only for the β-globin site and not for the ε,γ and δ-globin sites (Zou, Mali et al. 2011). It is therefore difficult to explain the two-fold increase in cell death in terms of off-target cleavages and more in depth analysis would be required. Until now, the search of possible off-target sites has been implemented through queries on BLAST (Basic Local Alignment Search Tool) algorithm but this tool is not adequate for the task and may miss possible off-target sites. A recent paper (2011) introduced a new web-based software (ZFN-Site) that has been specifically developed for this type of task (Cradick, Ambrosini et al. 2011). The software is able to carry out two different type of searches: one “basic search” where the intended target site is used as a base for the query and one “relaxed specificity search” where information’s about each ZFN subunits’ specificity (left and right) is taken into account. This second function allows for the search of possible off-target sites implementing the query with information from SELEX studies (Systematic

Evolution of Ligands by Exponential Enrichment) or similar approaches where single ZF proteins specificity has been examined (Cradick, Ambrosini et al. 2011).

7 Chapter 7: Final discussion

7.1 Overview

The development of a potential gene targeting therapy for the cure of β-thalassaemia and sickle cell anaemia would be a tremendous benefit for the world population and, more importantly, for those countries that are, at the same time, more affected by these type of diseases and do not have enough financial resources to provide adequate healthcare to all afflicted individuals.

Due to the strict relationship with *Plasmodium falciparum* (the cause of malaria), these diseases are prevalent in sub-tropical or tropical areas where economic struggle has hampered efforts to address the needs of globinopathies patients. Furthermore, only two therapeutic solutions are available for β-thalassaemia and sickle cell anaemia: one, blood transfusion, is only a palliative approach that is affected by serious contraindications while the other, bone marrow transplantation is limited by donor availability (see Section 1.1).

This project has tried to address all these problems by developing the basis for a potential therapeutic approach for monogenic β-globin diseases based on gene correction of the β-globin locus mediated by ZFNs.

A gene-targeting based therapy would be more affordable than life-long blood transfusion/iron chelation therapies because it would require only a single treatment; at the same time, it would solve the lack of bone marrow donors because the hypothetical patient would be the source of his/her own donor cells (after the duly correcting the disease-causing mutation).

7.2 Open-source or commercial approach?

In the course of this project, both open-source and commercial methods were used to obtain a functional ZFN specific for the β-globin gene. This double strategy has created the opportunity to compare the two widely different approaches from both a practical and ethical point of view.

The development by the Zinc Finger Consortium and other groups of protocols and reagents for the design and production of ZFNs has created the opportunity for research groups to generate ZFNs at an affordable price, while also contributing to the improvement of the technologies involved. An open-source approach has the advantage of promoting a collective effort where shared ZF archives and knowledge are vital for the fast and efficient development of novel ZFN platforms. This process has also the advantage of reducing the production costs allowing a greater number of researchers to use this technology. Cost-related issues are of fundamental importance to patients that could be able to benefit from a ZFN-based therapy. A wide range of genetic diseases (known as rare diseases) occurs with a very low frequency affecting a small percentage of the world population. These low incidences have the effect of limiting the amount of funding devoted to the research into potential therapies (due to the low impact of the disease on the world population) and also reduce the interest of the private sector due to the restricted size of the market and the expected low profit from the sale of rare-disease drugs. In contrast, non-commercial, low-cost ZFN technologies are more likely to, be implemented by rare-disease research groups with limited funding and the cost-benefit would be also transferred to the patient where the costs of any resulting ZFN-therapy would not be increased by the price of royalties to be paid to the owners of the underlying intellectual properties.

Unfortunately, the first open-source approaches, like the one used in this project, are difficult and time-consuming to implement and the low rate of effective ZFNs produced with these methods has been a limiting step to the creation of reliable therapeutic procedures. The interest of the scientific community, however, has brought increasing numbers of researchers into the field and new, robust and easier protocols like the OPEN or CoDA systems have been developed (Maeder, Thibodeau-Beganny et al. 2009; Sander, Dahlborg et al. 2011).

On the other hand, the monopoly on the ZFN market imposed by Sangamo Bioscience (that has the majority of patents on zinc finger technologies) has, in the opinion of some researchers, promoted the progress of the field and reduced the costs of acquiring licences from different parties (Chandrasekharan, Kumar et al. 2009). The centralization resulting from

Sangamo's monopoly and the associated extensive financial effort has brought a large number of scientists to work within the same company, promoting close collaboration and accelerating the development of new technologies. Chandrasekharan argues also that a monopoly of the ZFNs market would not affect public researchers due to the fact that "academic scientists routinely ignore patents and private-sector patentees correspondingly refrain from enforcing their patents" (Chandrasekharan, Kumar et al. 2009). For the sake of transparency, it should be noted that Chandrasekharan was one of the first researcher to sell a ZFN-license to Sangamo and sat on the company's advisory board for several years (Scott 2005).

Undoubtedly, purchasing ZFNs from Sigma (that has acquired the license from Sangamo) has the advantage of greatly reducing the effort and the time spent in creating and testing ZFNs. The company designs, validates and provides the final product in less than four months allowing the researcher to dedicate more time to the actual application of the nuclease and the development of a hypothetical therapy. Unfortunately the costs of purchasing a ZFN is still prohibitive (£20,000 for a custom nuclease) and, unlike the open-source approaches that allows the creation of multiple ZFNs specific for any desired gene with only one kit, Sigma's prices are related to individual ZFN requiring a new purchase for every new gene or DNA sequence to be targeted.

In the case of this project, additional doubts were cast on the transparency and reliability of the company. Sigma offers ZFNs under two different prices: "off the shelf" ZFNs that have been already created are sold for £9,000 while custom nucleases, due to the costs of development, are sold at full price. In case of two different research groups simultaneously requesting a ZFN for the same target sequence, Sigma demands the full price from both groups reasoning that, in order to preserve the privacy of all involved parties, they cannot disclose purchase requests made by other costumers. This was probably the case for this project were the custom ZFN4 was also purchased by another lab around the same time and without the knowledge of the two research groups involved in the acquisition (Zou, Mali et al. 2011). Although the privacy and interest of the costumer has to be preserved at all costs, the

secrecy adopted by the company allowed them to double the profit at the expense of publicly funded groups.

A second issue is related to the validation step performed by Sigma: while the company, at the time of the purchase, reported a 10% cutting efficiency of ZFN4 measured by Cel-I assay, a later conversation with an employee of the company revealed a totally different cutting efficiency with a reduction of 5 fold on the previous assessment (see Section 4.2.1). In the case of a big company like Sigma, only a rough estimation of ZFNs' efficiencies can reasonably be expected due to the large number of molecules that has to be processed everyday, but such a 5 fold overestimate can greatly affect the work of the research group that ordered the nuclease, especially if the financial effort required to acquired a tailored ZFN is considered.

An additional point can be made about overestimation in cutting efficiencies: although the development of ZFN-based therapeutic treatments is still at its infancy and substantial work has yet to be done by researchers, it has been increasingly clear that highly efficient ZFNs will be required in the future in order to develop a reliable approach for the cure of rare monogenic diseases. While ZFNs with low cutting efficiencies can still be used for research purposes to obtain valuable results, thanks to the use of efficient transfection and selection methods in well characterized research models (e.g. cell lines like K562, HEK293 cells or model organisms like mice), such low-efficiencies are likely to be compounded during clinical applications by numerous factors such as sub-optimal transfection efficiencies, limited accessibility of target stem cells (e.g. HSCs or iPS cells) and low frequencies of HR in such cells. Some of the factors that could affect future therapeutic applications will be addressed in the following sections.

In view of these considerations, a low cutting efficiency of 10% (the figure suggested by the Sigma team during their initial assessment) for ZFN4 would probably not be of therapeutic value but could at least have been a solid base for subsequent developments. Instead, the 2% cutting efficiency identified here for ZFN4 was already at the limit of detection in the Cel-I assay and was similarly limiting in both the HDR reporter assay and in the targeting of the endogenous gene in K562 cells. Sigma-Aldrich states in its commercial advert for custom ZFNs that thanks to their nucleases "single or biallelic edits occur

in 1-20% of clone population" (Sigma-Aldrich). Although the Sigma service is directed towards research applications, ZFNs need to move, in the near future, to clinical settings where frequencies of 1-20% will not be enough. In my opinion, in addition to improved delivery methods and a better understanding of the factors that determine gene targeting efficiencies in different tissues and cell types, it will be necessary to design multiple ZFNs for each possible monogenic disease gene in order to find the most effective one. If such a set of ZFNs had to be ordered from Sigma at the price of £20,000 for each nuclease, the relative costs of the research process would be too high for the small charities and funding bodies that finance the research of rare monogenic diseases. The high costs of the final ZFN-based gene therapy would be also transfered to the single patient or the National Health System reducing, in the end, the number of treated patients.

7.3 Challenges for gene targeting in HSCs and iPS cells

Haematopoietic Stem Cells (HSCs) would be the preferred target cells in a possible ex-vivo therapeutic approach of β-thalassaemia based on ZFNs. The safe and specific gene editing of HSCs could in theory not only cure thalassaemias but could also be used to treat a wide range of genetic or acquired diseases including SCID, sickle cell anaemia or AIDS. Unfortunately, both traditional gene therapy and gene targeting in HSCs have encountered many obstacles that have delayed the potential therapeutic benefits of these two, otherwise promising, approaches. In addition to the now well known risks of inducing leukaemias by the integration of traditional viral vectors close to proto-oncogenes (as the SCID clinical trial showed (Hacein-Bey-Abina, Garrigue et al. 2008)), the clinical application of HSC gene therapy/gene targeting is affected by a series of technical difficulties derived by the intrinsic complexity of working with HSCs. One main obstacle is the limited number of HSCs that can be obtained from a patient: the CD34 marker is usually used to select and "enrich" stem cell populations but it seems that only a fraction of CD34+ cells are genuine HSCs. Therefore there is a continuing need to develop more effective methods to expand, ex-vivo, these limiting number of

HSCs (Kiem, Jerome et al. 2012). The inherent difficulties of expanding stem cells while, at the same time, maintaining their self-renewal ability and engraftment capacity are also factors that limit the efficacy of many other possible clinical applications and therefore, in order to obtain long-term therapeutic advantages, the auto-transplant of HSCs that do not have a competitive advantage requires preventive and possibly toxic myeloablative treatments like chemotherapy or irradiation (Riviere, Dunbar et al. 2012).

The limited number of HSCs that can be used also requires a high level of transfection efficiency (obtained by viral vector delivery or other means like nucleofection) in order to maximise the chances of reaching sufficient levels of ZFN-induced gene correction with effective therapeutic advantages and transfection issues are also increased by the need of co-transfecting along the ZFN expression vectors the DNA repair template (Urnov, Rebar et al. 2010). The second limiting step in obtaining sufficient levels of gene correction is inherently linked to the ZFN's cutting efficiency and ability to promote HR events in HSCs.

The first effective gene targeting mediated by ZFNs of an endogenous gene in HSCs has been carried out in 2007 by Lombardo et al. This group, using an integration-defective lentiviral vector as delivery method, managed to obtain targeted gene addition by HR of a GFP construct into the CCR5 locus with frequencies up to 0.11% compared to frequencies of 0.005% in the absence of the ZFN (Lombardo, Genovese et al. 2007). Since then, researchers have put a lot of effort towards the development of improved gene targeting approaches for HSCs, but frequencies are still relatively low for actual therapeutic applications. The only ZFN-based approach in HSCs that has reached a clinical trial stage is a ZFN-gene therapy for AIDS but this approach involves the knocking-out (by NHEJ-mediated DSB repair) of the CCR5 gene that is relatively easy compared to HR-mediated gene correction (Kiem, Jerome et al. 2012). The CCR5 gene expresses a non-functional receptor protein that is used by the HIV virus to enter CD4+ T-cells. The lack of the receptor has no known negative effects and, moreover, individuals that are homozygous for a CCR5 mutation seems to be HIV resistant. The connection between CCR5 mutations and HIV resistance was confirmed even further by the so called "Berlin patient" case where an HIV-positive patient

was transplanted with HSCs from an individual that was homozygous for a CCR5 mutation. Three and an half years later the “Berlin patient” was completely free of any traceable HIV infection even without antiretroviral therapy. This case opened also the way to possible HIV treatments based on the transplant of CCR5-mutated HSCs. In order to obviate to the very low number of CCR5- individuals that could be used as HSCs donors, Sangamo developed an alternative approach based on a ZFN specific for the CCR5 gene (Maier, Brennan et al. 2013). The technique, which is now in a Phase2 clinical trial, involves the delivery of a CCR5-specific ZFN in CD4+ T-cells from HIV-positive patients by transduction with an Ad5/F35 adenoviral vector. Expression of the ZFN will produce DSBs at the CCR5 locus that, in absence of repair template, would be repaired by the error-prone NHEJ pathway and subsequent disruption of the CCR5 gene (Maier, Brennan et al. 2013). Transduced CD4+ cells will be then auto-transplanted back into the patients. Initial reports on the first six patients to be treated with the CCR5-specific ZFN showed promising results. After being infused with 10 billion transduced cells, the patients had the antiretroviral treatments suspended for twelve weeks and the levels of CD4+ cells were monitored. The number of CD4+ T-cells increased during time and the transduced cells were found to be persistent in the circulation also after one year (AIDS patient care report, 2011, “HIV/AIDS and STD updates”).

A possible alternative to HSCs is the use of iPS cells derived from, for example, skin cells obtained from the patient to be treated. Gene targeting of iPS cells has the advantage of having a potentially unlimited amount of treatable cells. Although frequencies resulted to be low, successful ZFN-induced gene targeting of the human β-globin gene in iPS cells has been recently obtained by two different groups (Sebastiano, Maeder et al. 2011; Zou, Mali et al. 2011). Unfortunately, the use of iPS cells in possible clinical applications is quite controversial: additional studies have to be carried out in order to address the risk related to the “reprogramming” techniques used to create iPS cells and this type of cell has still to be fully characterised in order to identify possible genotoxic effects. Additionally, safe methods to induce HSC differentiation into erythroid lineages have yet to be found and the ZFN-treated cells that are successfully transplanted back in to the patient have to

express durable and acceptable levels of the therapeutic or corrected gene without suffering gene silencing (Sebastiano, Maeder et al. 2011; Zou, Mali et al. 2011; Riviere, Dunbar et al. 2012).

7.4 Closing comments

The data reported in this study and the ones published by Sebastiano et al. and Zou et al. (Sebastiano, Maeder et al. 2011; Zou, Mali et al. 2011) show that the successful ZFN-induced gene targeting of the human β-globin gene is feasible and the results are encouraging for the development of a possible therapeutic treatment for both β-thalassaemia and sickle cell anaemia.

Unfortunately the absolute gene targeting frequency of 0.1% obtained with ZFN4 and reported in this thesis is far too low to be of any use without a selectable marker. One possibility, explored successfully by Li et al. in their ZFN-induced gene targeting of a haemophilia mouse model (Li, Haurigot et al. 2011), is the use of a targeting vector that carries a selectable marker flanked by two loxP sites. The drug selection can be used to "enrich" the cell population of successfully targeted cells and then can be easily removed by Cre-induced recombination. This approach, though, would be effective only on iPS cells due to the need for prolonged expansion in order to obtain acceptable amounts of corrected cells to be of any therapeutic value.

Ultimately, in order to create a successful therapeutic approach for β-globin disorders based on any nucleases technology, a series of technical advances have to be acquired.

First, new and more efficient engineered nucleases have to be developed based on traditional ZFNs or on novel systems like TALENs. Such nucleases should be not only more effective in producing DSBs and inducing HR events but should also be safer with reduced risks of off-target cleavages. In order to make them safer, improved validation systems have to be developed based on both bioinformatic approaches for the prediction of probable off-target sites and also on effective assays for the detection of unwanted DSBs events (Urnov, Rebar et al. 2010).

Second, a better understanding of the cellular mechanisms that affect the frequencies of HR and NHEJ have to be acquired in order to improve the low frequencies of gene targeting events noticed in HSCs and iPS cells. Third, in the case of HSCs, additional methods for the identification and isolation of stem cells should be developed in order to increase the number of possible target cells for the therapy (Kiem, Jerome et al. 2012).

Finally, long term expression of therapeutic or corrected genes should be obtained and maintained constantly in order to produce effective therapeutic approaches that do not require successive and repeated treatments. Gene silencing induced by epigenetic modifications has been shown to affect exogenous/ectopic genes but, more recently, also to affect endogenous genes that underwent gene correction by HR (Cuozzo, Porcellini et al. 2007). Therefore additional research has to be carried out in order to understand the link between epigenetic modifications and silencing of both exogenous DNA sequences used for therapeutic gene augmentation and endogenous genes that have undergone HDR.

References

(2013). "HIV/AIDS and STD Updates." AIDS Patient Care STDS **27**(3): 201-203.

Akinc, A., A. Zumbuehl, et al. (2008). "A combinatorial library of lipid-like materials for delivery of RNAi therapeutics." Nat Biotechnol **26**(5): 561-569.

Alton, E. W., M. Stern, et al. (1999). "Cationic lipid-mediated CFTR gene transfer to the lungs and nose of patients with cystic fibrosis: a double-blind placebo-controlled trial." Lancet **353**(9157): 947-954.

Alwin, S., M. B. Gere, et al. (2005). "Custom zinc-finger nucleases for use in human cells." Mol Ther **12**(4): 610-617.

Argyros, O., S. P. Wong, et al. (2011). "Development of S/MAR minicircles for enhanced and persistent transgene expression in the mouse liver." J Mol Med (Berl) **89**(5): 515-529.

Arnould, S., C. Delenda, et al. (2011). "The I-CreI meganuclease and its engineered derivatives: applications from cell modification to gene therapy." Protein Eng Des Sel **24**(1-2): 27-31.

Bae, K. H., Y. D. Kwon, et al. (2003). "Human zinc fingers as building blocks in the construction of artificial transcription factors." Nat Biotechnol **21**(3): 275-280.

Bank, A., R. Dorazio, et al. (2005). "A phase I/II clinical trial of beta-globin gene therapy for beta-thalassemia." Ann N Y Acad Sci **1054**: 308-316.

Bender, M. A., R. E. Gelinas, et al. (1989). "A majority of mice show long-term expression of a human beta-globin gene after retrovirus transfer into hematopoietic stem cells." Mol Cell Biol **9**(4): 1426-1434.

Beumer, K., G. Bhattacharyya, et al. (2006). "Efficient gene targeting in Drosophila with zinc-finger nucleases." Genetics **172**(4): 2391-2403.

Bibikova, M., M. Golic, et al. (2002). "Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases." Genetics **161**(3): 1169-1175.

BLAST. from http://blast.ncbi.nlm.nih.gov/Blast.cgi.

Branzei, D. and M. Foiani (2008). "Regulation of DNA repair throughout the cell cycle." Nat Rev Mol Cell Biol **9**(4): 297-308.

Bulger, M. and M. Groudine (1999). "Looping versus linking: toward a model for long-distance gene activation." Genes Dev **13**(19): 2465-2477.

Bulyk, M. L., X. Huang, et al. (2001). "Exploring the DNA-binding specificities of zinc fingers with DNA microarrays." Proc Natl Acad Sci U S A **98**(13): 7158-7163.

Carroll, D. (2011). "Genome engineering with zinc-finger nucleases." Genetics **188**(4): 773-782.

Cartier, N., S. Hacein-Bey-Abina, et al. (2009). "Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy." Science **326**(5954): 818-823.

Cathomen, T. and J. K. Joung (2008). "Zinc-finger nucleases: the next generation emerges." Mol Ther **16**(7): 1200-1207.

Cavazzana-Calvo, M., S. Hacein-Bey, et al. (2000). "Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease." Science **288**(5466): 669-672.

Cavazzana-Calvo, M., E. Payen, et al. (2010). "Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia." Nature **467**(7313): 318-322.

Cermak, T., E. L. Doyle, et al. (2011). "Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting." Nucleic Acids Res **39**(12): e82.

Certo, M. T., B. Y. Ryu, et al. (2011). "Tracking genome engineering outcome at individual DNA breakpoints." Nat Methods **8**(8): 671-676.

Chandrasekharan, S., S. Kumar, et al. (2009). "Proprietary science, open science and the role of patent disclosure: the case of zinc-finger proteins." Nat Biotechnol **27**(2): 140-144.

Choo, Y. and A. Klug (1997). "Physical basis of a protein-DNA recognition code." Curr Opin Struct Biol **7**(1): 117-125.

Choo, Y., I. Sanchez-Garcia, et al. (1994). "In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence." Nature **372**(6507): 642-645.

Choulika, A., A. Perrin, et al. (1995). "Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae." Mol Cell Biol **15**(4): 1968-1973.

Cirak, S., V. Arechavala-Gomeza, et al. (2011). "Exon skipping and dystrophin restoration in patients with Duchenne muscular dystrophy after systemic phosphorodiamidate morpholino oligomer treatment: an open-label, phase 2, dose-escalation study." Lancet **378**(9791): 595-605.

Cline, M. J., H. Stang, et al. (1980). "Gene transfer in intact animals." Nature **284**(5755): 422-425.

Colella, P. and A. Auricchio (2012). "Gene therapy of inherited retinopathies: a long and successful road from viral vectors to patients." Hum Gene Ther **23**(8): 796-807.

Conese, M., F. Ascenzioni, et al. (2011). "Gene and cell therapy for cystic fibrosis: from bench to bedside." J Cyst Fibros **10 Suppl 2**: S114-128.

Coutelle, C. (2008). "Why bother? Is in utero gene therapy worth the effort?" Mol Ther **16**(2): 219-220.

Coutelle, C. and C. Rodeck (2002). "On the scientific and ethical issues of fetal somatic gene therapy." Gene Ther **9**(11): 670-673.

Couzin, J. and J. Kaiser (2005). "Gene therapy. As Gelsinger case ends, gene therapy suffers another blow." Science **307**(5712): 1028.

Cradick, T. J., G. Ambrosini, et al. (2011). "ZFN-site searches genomes for zinc finger nuclease target sites and off-target sites." BMC Bioinformatics **12**: 152.

Cuozzo, C., A. Porcellini, et al. (2007). "DNA damage, homology-directed repair, and DNA methylation." PLoS Genet **3**(7): e110.

Dimery, I. W., D. D. Ross, et al. (1983). "Variation amongst K562 cell cultures." Exp Hematol **11**(7): 601-610.

Doyon, Y., T. D. Vo, et al. (2011). "Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures." Nat Methods **8**(1): 74-79.

Efstratiadis, A., J. W. Posakony, et al. (1980). "The structure and evolution of the human beta-globin gene family." Cell **21**(3): 653-668.

Elliott, B., C. Richardson, et al. (1998). "Gene conversion tracts from double-strand break repair in mammalian cells." Mol Cell Biol **18**(1): 93-101.

Evans, M. J. and M. H. Kaufman (1981). "Establishment in culture of pluripotential cells from mouse embryos." Nature **292**(5819): 154-156.

Evans-Galea, M. V., M. M. Wielgosz, et al. (2007). "Suppression of clonal dominance in cultured human lymphoid cells by addition of the cHS4 insulator to a lentiviral vector." Mol Ther **15**(4): 801-809.

Farah, J. A., G. A. Cromie, et al. (2009). "Ctp1 and Exonuclease 1, alternative nucleases regulated by the MRN complex, are required for efficient meiotic recombination." Proc Natl Acad Sci U S A **106**(23): 9356-9361.

Fischer, A., S. Hacein-Bey-Abina, et al. (2010). "20 years of gene therapy for SCID." Nat Immunol **11**(6): 457-460.

Foster, H., L. Popplewell, et al. (2012). "Genetic therapeutic approaches for duchenne muscular dystrophy." Hum Gene Ther **23**(7): 676-687.

Friedmann, T. (1992). "A brief history of gene therapy." Nat Genet **2**(2): 93-98.

Friedmann, T. and R. Roblin (1972). "Gene therapy for human genetic disease?" Science **175**(4025): 949-955.

Ganly, I., D. Kirn, et al. (2000). "A phase I study of Onyx-015, an E1B attenuated adenovirus, administered intratumorally to patients with recurrent head and neck cancer." Clin Cancer Res **6**(3): 798-806.

Gribnau, J., K. Diderich, et al. (2000). "Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus." Mol Cell **5**(2): 377-386.

Grizot, S., J. C. Epinat, et al. (2010). "Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds." Nucleic Acids Res **38**(6): 2006-2018.

Grizot, S., J. Smith, et al. (2009). "Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease." Nucleic Acids Res **37**(16): 5405-5419.

Hacein-Bey-Abina, S., A. Garrigue, et al. (2008). "Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1." J Clin Invest **118**(9): 3132-3142.

Heise, C. and D. H. Kirn (2000). "Replication-selective adenoviruses as oncolytic agents." J Clin Invest **105**(7): 847-851.

Helleday, T., J. Lo, et al. (2007). "DNA double-strand break repair: from mechanistic understanding to cancer treatment." DNA Repair (Amst) **6**(7): 923-935.

Higgs, D. R., J. D. Engel, et al. (2012). "Thalassaemia." Lancet **379**(9813): 373-383.

Hockemeyer, D., F. Soldner, et al. (2009). "Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases." Nat Biotechnol **27**(9): 851-857.

Isalan, M., Y. Choo, et al. (1997). "Synergy between adjacent zinc fingers in sequence-specific DNA recognition." Proc Natl Acad Sci U S A **94**(11): 5617-5621.

J. Sambrook, E. F. F., Tom Maniatis (1982). Molecular Cloning: a Laboratory Manual, Cold Spring Harbor

Jasin, M. (1996). "Genetic manipulation of genomes with rare-cutting endonucleases." Trends Genet **12**(6): 224-228.

Jensen, N. M., T. Dalsgaard, et al. (2011). "An update on targeted gene repair in mammalian cells: methods and mechanisms." J Biomed Sci **18**: 10.

Joung, J. K., E. I. Ramm, et al. (2000). "A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions." Proc Natl Acad Sci U S A **97**(13): 7382-7387.

Kaina, B. (2003). "DNA damage-triggered apoptosis: critical role of DNA repair, double-strand breaks, cell proliferation and signaling." Biochem Pharmacol **66**(8): 1547-1554.

Kaiser, J. (2009). "Gene therapy. Beta-thalassemia treatment succeeds, with a caveat." Science **326**(5959): 1468-1469.

Kay, M. A., J. C. Glorioso, et al. (2001). "Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics." Nat Med **7**(1): 33-40.

Kiem, H. P., K. R. Jerome, et al. (2012). "Hematopoietic-stem-cell-based gene therapy for HIV disease." Cell Stem Cell **10**(2): 137-147.

Kim, S., Y. W. Kim, et al. (2012). "Chromatin structure of the LCR in the human beta-globin locus transcribing the adult delta- and beta-globin genes." Int J Biochem Cell Biol **44**(3): 505-513.

Kim, Y. G., J. Cha, et al. (1996). "Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain." Proc Natl Acad Sci U S A **93**(3): 1156-1160.

Klug, A. (2010). "The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation." Q Rev Biophys **43**(1): 1-21.

Kornberg, R. D. and Y. Lorch (1999). "Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome." Cell **98**(3): 285-294.

Kuzminov, A. (2001). "Single-strand interruptions in replicating chromosomes cause double-strand breaks." Proc Natl Acad Sci U S A **98**(15): 8241-8246.

Laird, P. W., A. Zijderveld, et al. (1991). "Simplified mammalian DNA isolation procedure." Nucleic Acids Res **19**(15): 4293.

Lanzov, V. A. (1999). "Gene targeting for gene therapy: prospects." Mol Genet Metab **68**(2): 276-282.

Lawn, R. M., A. Efstratiadis, et al. (1980). "The nucleotide sequence of the human beta-globin gene." Cell **21**(3): 647-651.

Lee, G. S., M. B. Neiditch, et al. (2004). "RAG proteins shepherd double-strand breaks to a specific pathway, suppressing error-prone repair, but RAG nicking initiates homologous recombination." Cell **117**(2): 171-184.

Levings, P. P. and J. Bungert (2002). "The human beta-globin locus control region." Eur J Biochem **269**(6): 1589-1599.

Li, H., V. Haurigot, et al. (2011). "In vivo genome editing restores haemostasis in a mouse model of haemophilia." Nature **475**(7355): 217-221.

Li, L., L. P. Wu, et al. (1992). "Functional domains in Fok I restriction endonuclease." Proc Natl Acad Sci U S A **89**(10): 4275-4279.

Li, T., S. Huang, et al. (2011). "Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes." Nucleic Acids Res **39**(14): 6315-6325.

Liu, P. Q., E. M. Chan, et al. (2009). "Generation of a triple-gene knockout mammalian cell line using engineered zinc-finger nucleases." Biotechnol Bioeng.

Lombardo, A., P. Genovese, et al. (2007). "Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery." Nat Biotechnol **25**(11): 1298-1306.

Maeder, M. L., S. Thibodeau-Beganny, et al. (2008). "Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification." Mol Cell **31**(2): 294-301.

Maeder, M. L., S. Thibodeau-Beganny, et al. (2009). "Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays." Nat Protoc **4**(10): 1471-1501.

Maier, D. A., A. L. Brennan, et al. (2013). "Efficient Clinical Scale Gene Modification via Zinc Finger Nuclease-Targeted Disruption of the HIV Co-receptor CCR5." Hum Gene Ther **24**(3): 245-258.

Malerba, A., P. S. Sharp, et al. (2011). "Chronic systemic therapy with low-dose morpholino oligomers ameliorates the pathology and normalizes locomotor behavior in mdx mice." Mol Ther **19**(2): 345-354.

Mandell, J. G. and C. F. Barbas, 3rd (2006). "Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases." Nucleic Acids Res **34**(Web Server issue): W516-523.

Maniatis, T., E. F. Fritsch, et al. (1980). "The molecular genetics of human hemoglobins." Annu Rev Genet **14**: 145-178.

Manilla, P., T. Rebello, et al. (2005). "Regulatory considerations for novel gene therapy products: a review of the process leading to the first clinical lentiviral vector." Hum Gene Ther **16**(1): 17-25.

Martin, M. F. a. G. (1990). PCR Protocols. New York, Academic Press.

May, C., S. Rivella, et al. (2000). "Therapeutic haemoglobin synthesis in beta-thalassaemic mice expressing lentivirus-encoded human beta-globin." Nature **406**(6791): 82-86.

May, C., S. Rivella, et al. (2002). "Successful treatment of murine beta-thalassemia intermedia by transfer of the human beta-globin gene." Blood **99**(6): 1902-1908.

McLachlan, J., S. Fernandez, et al. (2009). "Specific targeted gene repair using single-stranded DNA oligonucleotides at an endogenous locus in mammalian cells uses homologous recombination." DNA Repair (Amst) **8**(12): 1424-1433.

Meng, X., S. Thibodeau-Beganny, et al. (2007). "Profiling the DNA-binding specificities of engineered Cys2His2 zinc finger domains using a rapid cell-based method." Nucleic Acids Res **35**(11): e81.

Mercola, K. E., H. D. Stang, et al. (1980). "Insertion of a new gene of viral origin into bone marrow cells of mice." Science **208**(4447): 1033-1035.

Miller, J. C., M. C. Holmes, et al. (2007). "An improved zinc-finger nuclease architecture for highly specific genome editing." Nat Biotechnol **25**(7): 778-785.

Miller, J. C., S. Tan, et al. (2011). "A TALE nuclease architecture for efficient genome editing." Nat Biotechnol **29**(2): 143-148.

Moehle, E. A., J. M. Rock, et al. (2007). "Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases." Proc Natl Acad Sci U S A **104**(9): 3055-3060.

Montini, E., D. Cesana, et al. (2006). "Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration." Nat Biotechnol **24**(6): 687-696.

Morton, J., M. W. Davis, et al. (2006). "Induction and repair of zinc-finger nuclease-targeted double-strand breaks in Caenorhabditis elegans somatic cells." Proc Natl Acad Sci U S A **103**(44): 16370-16375.

Mueller, C. and T. R. Flotte (2008). "Clinical gene therapy using recombinant adeno-associated virus vectors." Gene Ther **15**(11): 858-863.

Muncie, H. L., Jr. and J. Campbell (2009). "Alpha and beta thalassemia." Am Fam Physician **80**(4): 339-344.

Naumann, S., D. Reutzel, et al. (2001). "Complete karyotype characterization of the K562 cell line by combined application of G-banding, multiplex-fluorescence in situ hybridization, fluorescence in situ hybridization, and comparative genomic hybridization." Leuk Res **25**(4): 313-322.

NCBI. "Beta-globin gene." from www.ncbi.nlm.nih.gov/gene/3043.

Nelms, B. L. and P. A. Labosky (2011). "A predicted hairpin cluster correlates with barriers to PCR, sequencing and possibly BAC recombineering." Sci Rep **1**: 106.

OBA. from http://oba.od.nih.gov/oba/RAC/meetings/Dec2009/Leboulch_RACDec09.pdf.

Oleykowski, C. A., C. R. Bronson Mullins, et al. (1998). "Mutation detection using a novel plant endonuclease." Nucleic Acids Res **26**(20): 4597-4602.

Paques, F. and P. Duchateau (2007). "Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy." Curr Gene Ther **7**(1): 49-66.

Paques, F. and J. E. Haber (1999). "Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae." Microbiol Mol Biol Rev **63**(2): 349-404.

Pattanayak, V., C. L. Ramirez, et al. (2011). "Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection." Nat Methods **8**(9): 765-770.

Pawliuk, R., K. A. Westerman, et al. (2001). "Correction of sickle cell disease in transgenic mouse models by gene therapy." Science **294**(5550): 2368-2371.

Peebles, D., L. G. Gregory, et al. (2004). "Widespread and efficient marker gene expression in the airway epithelia of fetal sheep after minimally invasive tracheal application of recombinant adenovirus in utero." Gene Ther **11**(1): 70-78.

Perez, E. E., J. Wang, et al. (2008). "Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases." Nat Biotechnol **26**(7): 808-816.

Persons, D. A. (2009). "Hematopoietic stem cell gene transfer for the treatment of hemoglobin disorders." Hematology Am Soc Hematol Educ Program: 690-697.

Pierce, A. J., P. Hu, et al. (2001). "Ku DNA end-binding protein modulates homologous repair of double-strand breaks in mammalian cells." Genes Dev **15**(24): 3237-3242.

Pierce, A. J., R. D. Johnson, et al. (1999). "XRCC3 promotes homology-directed repair of DNA damage in mammalian cells." Genes Dev **13**(20): 2633-2638.

Porter, A. C. (2005). "Two hands make light work of gene modification." Rejuvenation Res **8**(4): 211-215.

Porteus, M. H. (2006). "Mammalian gene targeting with designed zinc finger nucleases." Mol Ther **13**(2): 438-446.

Porteus, M. H. and D. Baltimore (2003). "Chimeric nucleases stimulate gene targeting in human cells." Science **300**(5620): 763.

Ramirez, C. L., M. T. Certo, et al. (2012). "Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects." Nucleic Acids Res **40**(12): 5560-5568.

Ramirez, C. L., J. E. Foley, et al. (2008). "Unexpected failure rates for modular assembly of engineered zinc fingers." Nat Methods **5**(5): 374-375.

Rebar, E. J. and C. O. Pabo (1994). "Zinc finger phage: affinity selection of fingers with new DNA-binding specificities." Science **263**(5147): 671-673.

Riviere, I., C. E. Dunbar, et al. (2012). "Hematopoietic stem cell engineering at a crossroads." Blood **119**(5): 1107-1116.

Rouet, P., F. Smih, et al. (1994). "Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease." Mol Cell Biol **14**(12): 8096-8106.

Rund, D. and E. Rachmilewitz (2005). "Beta-thalassemia." N Engl J Med **353**(11): 1135-1146.

Rutherford, T., J. B. Clegg, et al. (1981). "Embryonic erythroid differentiation in the human leukemic cell line K562." Proc Natl Acad Sci U S A **78**(1): 348-352.

Sadelain, M., F. Boulad, et al. (2007). "Therapeutic options for patients with severe beta-thalassemia: the need for globin gene therapy." Hum Gene Ther **18**(1): 1-9.

Sambrook, J., H. Westphal, et al. (1968). "The integrated state of viral DNA in SV40-transformed cells." Proc Natl Acad Sci U S A **60**(4): 1288-1295.

Sander, J. D., E. J. Dahlborg, et al. (2011). "Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA)." Nat Methods **8**(1): 67-69.

Sander, J. D., M. L. Maeder, et al. (2010). "ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool." Nucleic Acids Res **38**(Web Server issue): W462-468.

Sander, J. D., P. Zaback, et al. (2007). "Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool." Nucleic Acids Res **35**(Web Server issue): W599-605.

Sangamo-Bioscience. from http://www.sangamo.com/index.html.

Sangamo-Bioscience. (2007). "Sigma-Aldrich and Sangamo BioSciences Announce Alliance to Develop Zinc Finger-Based Laboratory Research Reagents." from http://investor.sangamo.com/releasedetail.cfm?ReleaseID=253118.

Sangamo-Bioscience. (2012). "Product Pipeline." from http://www.sangamo.com/pipeline/index.html.
Scherer, S. and R. W. Davis (1979). "Replacement of chromosome segments with altered DNA sequences constructed in vitro." Proc Natl Acad Sci U S A **76**(10): 4951-4955.
Scott, C. T. (2005). "The zinc finger nuclease monopoly." Nat Biotechnol **23**(8): 915-918.
Sebastiano, V., M. L. Maeder, et al. (2011). "In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases." Stem Cells **29**(11): 1717-1726.
Segal, D. J., R. R. Beerli, et al. (2003). "Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins." Biochemistry **42**(7): 2137-2148.
Segal, D. J., B. Dreier, et al. (1999). "Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences." Proc Natl Acad Sci U S A **96**(6): 2758-2763.
Sheridan, C. (2011). "Gene therapy finds its niche." Nat Biotechnol **29**(2): 121-128.
Sigma-Aldrich. "CompoZr® Custom ZFN Service." from http://www.sigmaaldrich.com/life-science/zinc-finger-nuclease-technology/custom-zfn.html.
Smithies, O., R. G. Gregg, et al. (1985). "Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination." Nature **317**(6034): 230-234.
Szczepek, M., V. Brondani, et al. (2007). "Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases." Nat Biotechnol **25**(7): 786-793.
Szybalska, E. H. and W. Szybalski (1962). "Genetics of human cess line. IV. DNA-mediated heritable transformation of a biochemical trait." Proc Natl Acad Sci U S A **48**: 2026-2034.
Thomas, K. R. and M. R. Capecchi (1987). "Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells." Cell **51**(3): 503-512.
Thomas, K. R., K. R. Folger, et al. (1986). "High frequency targeting of genes to specific sites in the mammalian genome." Cell **44**(3): 419-428.
Transgenomic. from www.transgenomic.com.
Urnov, F. D., J. C. Miller, et al. (2005). "Highly efficient endogenous human gene correction using designed zinc-finger nucleases." Nature **435**(7042): 646-651.
Urnov, F. D., E. J. Rebar, et al. (2010). "Genome editing with engineered zinc finger nucleases." Nat Rev Genet **11**(9): 636-646.
van Nierop, G. P., A. A. de Vries, et al. (2009). "Stimulation of homology-directed gene targeting at an endogenous human locus by a nicking endonuclease." Nucleic Acids Res **37**(17): 5725-5736.
Vega, M. A. (1991). "Prospects for homologous recombination in human gene therapy." Hum Genet **87**(3): 245-253.
Verma, I. M. and M. D. Weitzman (2005). "Gene therapy: twenty-first century medicine." Annu Rev Biochem **74**: 711-738.

Weatherall, D. (2011). "The inherited disorders of haemoglobin: an increasingly neglected global health burden." Indian J Med Res **134**: 493-497.

Weatherall, D. J. and J. B. Clegg (2001). "Inherited haemoglobin disorders: an increasing global health problem." Bull World Health Organ **79**(8): 704-712.

Wolfe, S. A., L. Nekludova, et al. (2000). "DNA recognition by Cys2His2 zinc finger proteins." Annu Rev Biophys Biomol Struct **29**: 183-212.

Wright, D. A., S. Thibodeau-Beganny, et al. (2006). "Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly." Nat Protoc **1**(3): 1637-1652.

Wu, S. Q., K. V. Voelkerding, et al. (1995). "Extensive amplification of bcr/abl fusion genes clustered on three marker chromosomes in human leukemic cell line K-562." Leukemia **9**(5): 858-862.

Yamamoto, A., M. Kormann, et al. (2009). "Current prospects for mRNA gene delivery." Eur J Pharm Biopharm **71**(3): 484-489.

Yanez, R. J. and A. C. Porter (1998). "Therapeutic gene targeting." Gene Ther **5**(2): 149-159.

Yanez, R. J. and A. C. Porter (1999). "Influence of DNA delivery method on gene targeting frequencies in human cells." Somat Cell Mol Genet **25**(1): 27-31.

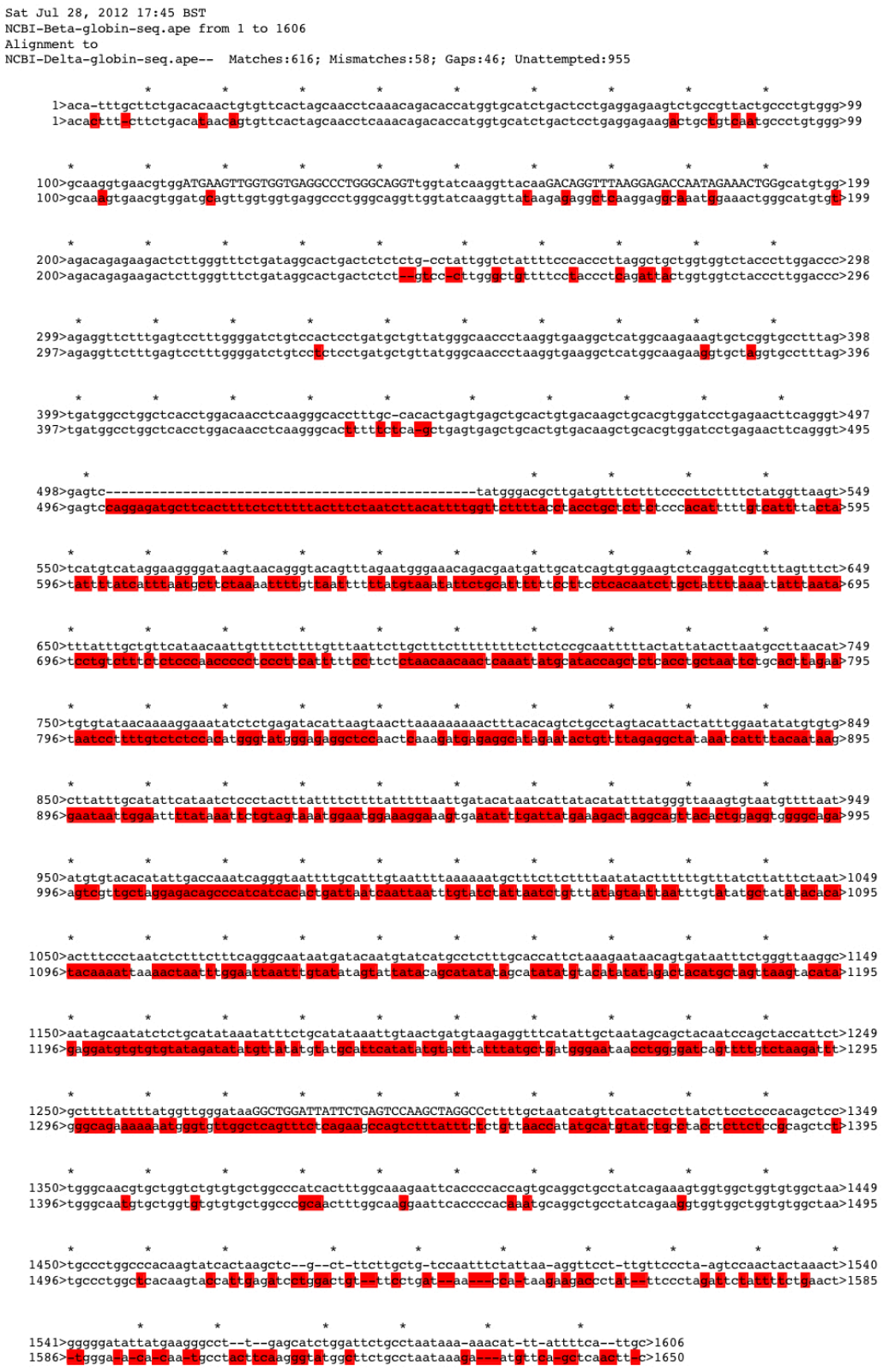
Yokota, T., E. Hoffman, et al. (2011). "Antisense oligo-mediated multiple exon skipping in a dog model of duchenne muscular dystrophy." Methods Mol Biol **709**: 299-312.

Zou, J., M. L. Maeder, et al. (2009). "Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells." Cell Stem Cell **5**(1): 97-110.

Zou, J., P. Mali, et al. (2011). "Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease." Blood **118**(17): 4599-4608.

Appendix 1

a. β- and δ-globin sequences alignment



The two β and δ DNA sequences reported in this appendix were obtained from the NCBI web archive and were aligned (only for graphic representation) using the Ape-A plasmid editor software. The actual search for possible homologies was carried out by use of the BLAST algorithm from NCBI as described in Section 3.2.1. The three target sequences (from top to bottom) TS-1, TS-2 and TS3 are in capitol letters.

b. ZF-DBDs coding sequences

In the following section the coding sequences for the six ZF-DBDs are reported. Each ZF-DBD is made of four different ZF modules and each ZF module is composed by a constant sequence (small letters) and a variable sequence (capital letters in bold) that codes for the amino acids responsible for the binding specificity.

ZF1-R:

ZF-60-71-106-94, sequence:

5′-tctagacccggggagaagccttacaaatgcccagaatgtggaaagagttttagc
ACTTCTGGACATCTTGTGAGAcaccagagaacacataccggggagaagccttacaaa
tgcccagaatgtggaaagagttttagc**CAGTCTGGAGACCTTAGAAGA**caccagaga
acacataccggggagaagccttacaaatgcccagaatgtggaaagagttttagc**AGG
TCTGACCATCTCACCACA**caccagagaacacataccggggagaagccttacaaatgc
ccagaatgtggaaagagttttagc**CTAAGAAACATCTCGCAGAA**caccagagaacac
ataccggtggatcctctgcacagt-3′

ZF1-L:

ZF-92-103-89-90, sequence:

5′-tctagacccggggagaagccttacaaatgcccagaatgtggaaagagttttagc
ACTTCTGGAAATCTCACAGAAcaccagagaacacataccggggagaagccttacaaa
tgcccagaatgtggaaagagttttagc**ACTACCGGAGCACTCACAGAA**caccagaga
acacataccggggagaagccttacaaatgcccagaatgtggaaagagttttagc**GAA
AGGTCTCATCTCAGAGAG**caccagagaacacataccggggagaagccttacaaatgc
ccagaatgtggaaagagttttagc**TCTAAGAAAGCACTCACAGAA**caccagagaaca
cataccggtggatcctctgcacagt-3′

ZF2-R:

ZF-73-104-83-93, sequence:

5′–tctagacccggggagaagccttacaaatgcccagaatgtggaaagagttttagc
GATTGTAGGGATCTCGCTAGAcaccagagaacacataccggggagaagccttacaaa
tgcccagaatgtggaaagagttttagc**AGGGAAGATAATCTCCATACA**caccagaga
acacataccggggagaagccttacaaatgcccagaatgtggaaagagttttagc**GAA**
AGGTCTCATCTCAGAGAGcaccagagaacacataccggggagaagccttacaaatgc
ccagaatgtggaaagagttttagc**ACTTCTCATTCCCTCACAGAA**caccagagaaca
cataccggtggatcctctgcacagt–3′

ZF2-L:

ZF-73-93-77-77, sequence:

5′–tctagacccggggagaagccttacaaatgcccagaatgtggaaagagttttagc
GATTGTAGGGATCTCGCTAGAcaccagagaacacataccggggagaagccttacaaa
tgcccagaatgtggaaagagttttagc**ACTTCTCATTCCCTCACAGAA**caccagaga
acacataccggggagaagccttacaaatgcccagaatgtggaaagagttttagc**ACT**
ACCGGAAATCTCACAGTGcaccagagaacacataccggggagaagccttacaaatgc
ccagaatgtggaaagagttttagc**ACTACCGGAAATCTCACAGTG**caccagagaaca
cataccggtggatcctctgcacagt–3′

ZF3-R:

ZF-106-75-82-77, sequence:

5′–tctagacccggggagaagccttacaaatgcccagaatgtggaaagagttttagc
AGGTCTGACCATCTCACCACAcaccagagaacacataccggggagaagccttacaaa
tgcccagaatgtggaaagagttttagc**GATTCTGGAAATCTCAGAGTG**caccagaga
acacataccggggagaagccttacaaatgcccagaatgtggaaagagttttagc**CAA**
CTTGCACATCTCAGAGCTcaccagagaacacataccggggagaagccttacaaatgc
ccagaatgtggaaagagttttagc**ACTACCGGAAATCTCACAGTG**caccagagaaca
cataccggtggatcctctgcacagt–3′

ZF3-L:

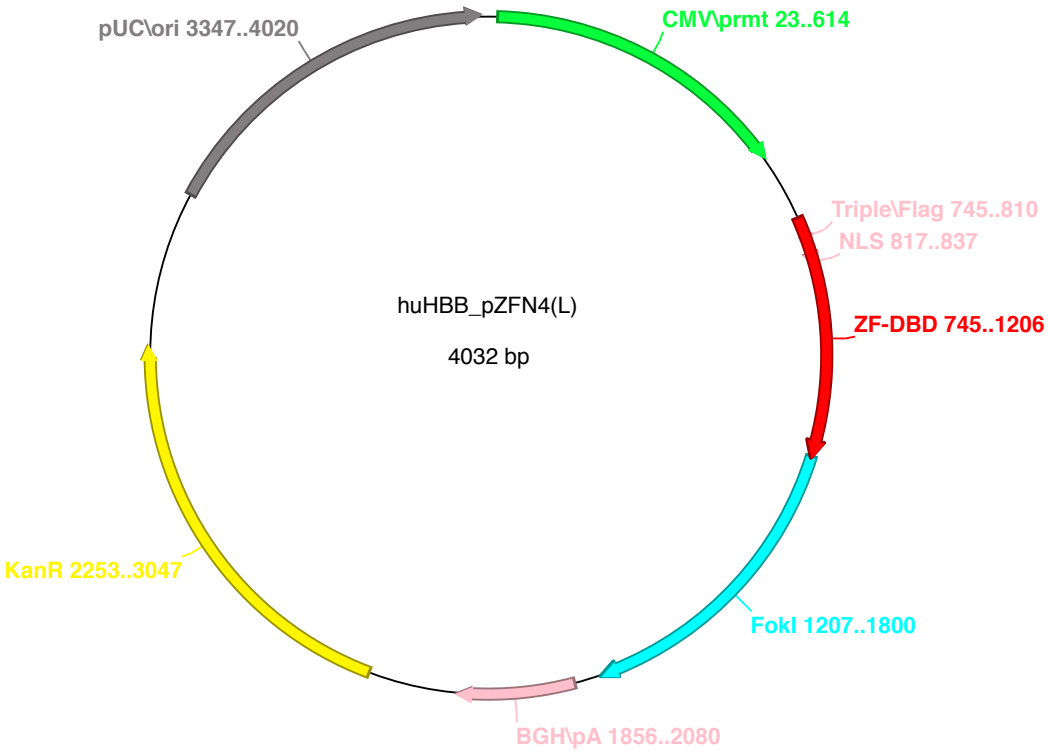
ZF-69-96-74-103, sequence:

5′–tctagacccggggagaagccttacaaatgcccagaatgtggaaagagttttagc
GATCCTGGAGCACTCGTGAGAcaccagagaacacataccggggagaagccttacaaa
tgcccagaatgtggaaagagttttagc**ACTAAGAACTCTCTCACAGAA**caccagaga
acacataccggggagaagccttacaaatgcccagaatgtggaaagagttttagc**CAA**
AGGGCAAATCTCAGAGCTcaccagagaacacataccggggagaagccttacaaatgc
ccagaatgtggaaagagttttagc**ACTACCGGAGCACTCACAGAA**caccagagaaca
cataccggtggatcctctgcacagt–3′

Appendix 2

Human HBB ZFN4-(L)

Map:



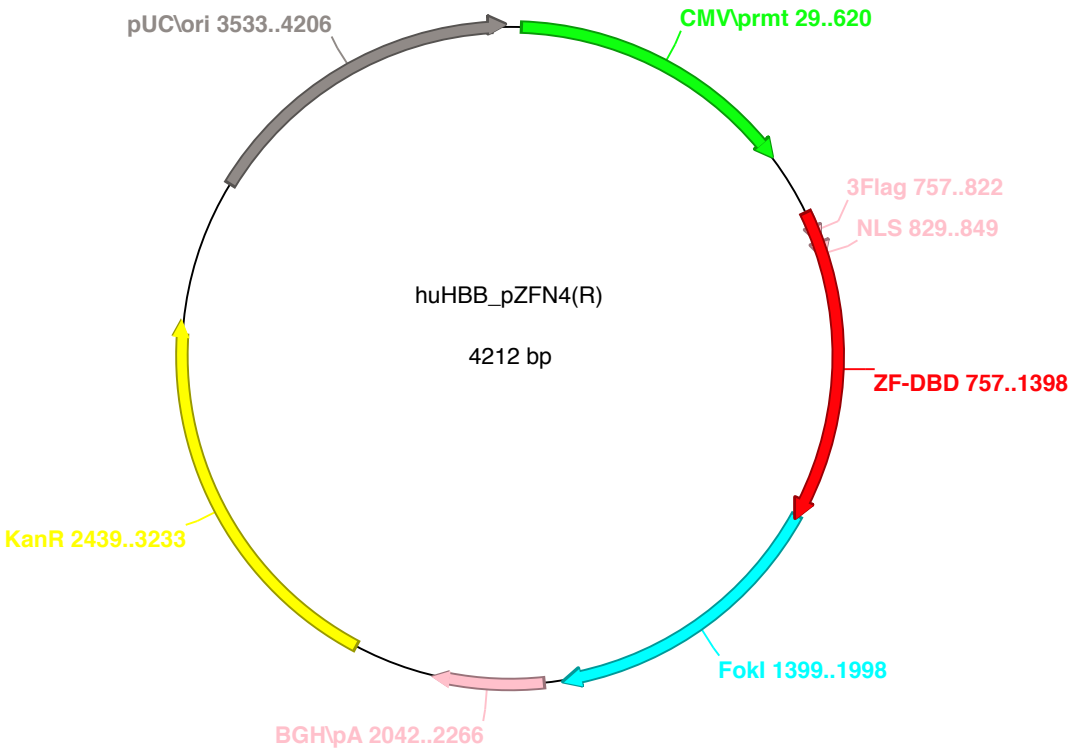
Sequence (colour coding follows the one used in the map):

tcgcgatgtacgggccagatatacgcgttgacattgattattgactagttattaata
gtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacata
acttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtc
aataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatg
ggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgcc
aagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgccca
gtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgc
tattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttga
ctcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggca
ccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaat
gggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactag
agaacccactgcttactggcttatcgaaattaatacgactcactatagggagaccca
agctggctagcgtttaaacttaagctgatccactagtccagtgtggtggaattcgcc
atggactacaaagaccatgacggtgattataaagatcatgacatcgattacaaggat
gacgatgacaagatggcccccaagaagaagaggaaggtgggcatccacggggtaccc
gccgctatggctgagaggcccttccagtgtcgaatctgcatgcgtaacttcagtgac
cgctccaacctgtcccgccacatccgcacccacaccggcgagaagccttttgcctgt
gacatttgtgggaggaaatttgcccagtccggcgacctgacccgccataccaagata
cacacgggatctcagaagcccttccagtgtcgaatctgcatgcgtaacttcagtcgc

tccgacaccctgtcccagcacatccgcacccacaccggcgagaagccttttgcctgt
gacatttgtgggaggaaatttgcccagtccggctccctgacccgccataccaagata
cacctgcggggatcccagctggtgaagagcgagctggaggagaagaagtccgagctg
cggcacaagctgaagtacgtgccccacgagtacatcgagctgatcgagatcgccagg
aacagcacccaggaccgcatcctggagatgaaggtgatggagttcttcatgaaggtg
tacggctacaggggaaagcacctgggcggaagcagaaagcctgacggcgccatctat
acagtgggcagccccatcgattacggcgtgatcgtggacacaaaggcctacagcggc
ggctacaatctgcctatcggccaggccgacgagatggagagatacgtggaggagaac
cagacccggaataagcacctcaaccccaacgagtggtggaaggtgtaccctagcagc
gtgaccgagttcaagttcctgttcgtgagcggccacttcaagggcaactacaaggcc
cagctgaccaggctgaaccacatcaccaactgcaatggcgccgtgctgagcgtggag
gagctgctgatcggcggcgagatgatcaaagccggcaccctgacactggaggaggtg
cggcgcaagttcaacaacggcgagatcaacttcagatcttgataactcgagtctaga
gggcccgtttaaacccgctgatcagcctcgactgtgccttctagttgccagccatct
gttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtc
ctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctatt
ctgggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcagg
catgctggggatgcggtgggctctatggcttctactgggcggttttatggacagcaa
gcgaaccggaattgccagctggggcgccctctggtaaggttgggaagccctgcaaag
taaactggatggctttctcgccgccaaggatctgatggcgcaggggatcaagctctg
atcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcag
gttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaa
tcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttt
ttgtcaagaccgacctgtccggtgccctgaatgaactgcaagacgaggcagcgcggc
tatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactg
aagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcat
ctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgc
atacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagc
gagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagc
atcaggggctcgcgccagccgaactgttcgccaggctcaaggcgagcatgcccgacg
gcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaa
atggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatc
aggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctg
accgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttct
atcgccttcttgacgagttcttctgaattattaacgcttacaatttcctgatgcggt
attttctccttacgcatctgtgcggtatttcacaccgcatacaggtggcacttttcg
gggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgta
tccgctcatgagacaataaccctgataaatgcttcaataatagcacgtgctaaaact
tcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaa
aatcccttaacgtgagttttcgttccactgagcgtcagacccgtagaaaagatcaa
aggatcttcttgagatccttttttttctgcgcgtaatctgctgcttgcaaacaaaaaa
accaccgctaccagcggtggtttgtttgccggatcaagagctaccaactcttttcc
gaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagcc
gtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgct
aatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttgga
ctcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtg
cacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtga
gctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaag
cggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggta
tctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatg
ctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggtt
cctgggctttgctggccttttgctcacatgttcttgactct

Human HBB ZFN4-(R)

Map:



Sequence (colour coding follows the one used in the map):

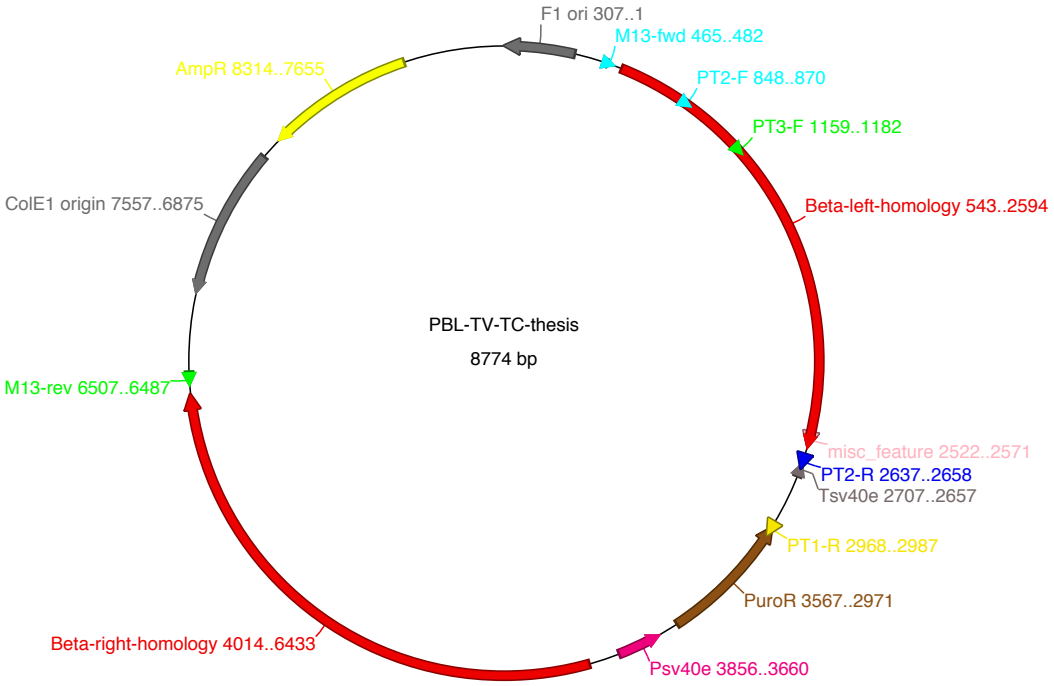
gactcttcgcgatgtacgggccagatatacgcgttgacattgattattgactagtta
ttaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgt
tacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccatt
gacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacg
tcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatca
tatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcatta
tgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagt
catcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcg
gtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgtt
ttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgac
gcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggct
aactagagaacccactgcttactggcttatcgaaattaatacgactcactataggga
gacccaagctggctagcgtttaaacttaagctgatccactagtccagtgtggtggaa
ttcgccatgagatctgactacaaagaccatgacggtgattataaagatcatgacatc
gattacaaggatgacgatgacaagatggcccccaagaagaagaggaaggtgggcatt
catggggtacccgccgctatggctgagaggcccttccagtgtcgaatctgcatgcgt
aacttcagtcgtagtgacagcctgagccggcacatccgcacccacacaggcgagaag
ccttttgcctgtgacatttgtgggaggaaatttgccgacagcagcaaccgcaaaacg
cataccaagatacacacgggcggacaacggcccttccagtgtcgaatctgcatgcgt
aacttcagtcgctccgccgccctgtcccgccacatccgcacccacaccggcgagaag

ccttttgcctgtgacatttgtgggaggaaatttgcccgcctggacaaccgcaccgcc
cataccaagatacacacgggcggaggcggatctcagaagcccttccagtgtcgaatc
tgcatgcgtaacttcagtcgctccgcccacctgtcccgccacatccgcacccacacc
ggcgagaagccttttgcctgtgacatttgtgggaggaaatttgcccgctccgacgcc
ctggcccgccataccaagatacacctgcggggatcccagctggtgaagagcgagctg
gaggagaagaagtccgagctgcggcacaagctgaagtacgtgccccacgagtacatc
gagctgatcgagatcgccaggaacagcacccaggaccgcatcctggagatgaaggtg
atggagttcttcatgaaggtgtacggctacaggggaaagcacctgggcggaagcaga
aagcctgacggcgccatctatacagtgggcagccccatcgattacggcgtgatcgtg
gacacaaaggcctacagcggcggctacaatctgcctatcggccaggccgacgagatg
cagagatacgtgaaggagaaccagacccggaataagcacatcaaccccaacgagtgg
tggaaggtgtaccctagcagcgtgaccgagttcaagttcctgttcgtgagcggccac
ttcaagggcaactacaaggcccagctgaccaggctgaaccacaaaaccaactgcaat
ggcgccgtgctgagcgtggaggagctgctgatcggcggcgagatgatcaaagccggc
accctgacactggaggaggtgcggcgcaagttcaacaacggcgagatcaacttctga
taactcgagtctagagggcccgtttaaacccgctgatcagcctcgactgtgccttct
agttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggt
gccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagt
aggtgtcattctattctgggggggtggggtggggcaggacagcaagggggaggattgg
gaagacaatagcaggcatgctggggatgcggtgggctctatggcttctactgggcgg
ttttatggacagcaagcgaaccggaattgccagctggggcgccctctggtaaggttg
ggaagccctgcaaagtaaactggatggctttctcgccgccaaggatctgatggcgca
ggggatcaagctctgatcaagagacaggatgaggatcgtttcgcatgattgaacaag
atggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgact
gggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcagg
ggcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaactgcaag
acgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgc
tcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggc
aggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatg
caatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcga
aacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatg
atctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaagg
cgagcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccga
atatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtg
tggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttg
gcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgc
agcgcatcgccttctatcgccttcttgacgagttcttctgaattattaacgcttaca
atttcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatac
aggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaat
acattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataata
gcacgtgctaaaacttcatttttaatttaaaaggatctaggtgaagatccttttttga
taatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccc
cgtagaaaagatcaaaggatcttcttgagatccttttttttctgcgcgtaatctgctg
cttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagct
accaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgt
ccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctac
atacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtg
tcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctg
aacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgag
atacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcgga
caggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagg
gggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcg

tcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgc
ggccttttacggttcctgggcttttgctggccttttgctcacatgttctt

Appendix 3

pBL-TV-TC map:



pBL-TV-TC partial sequence (regions of interest):

cacgacgttgtaaaacgacggccagtgagcgcgcgtaatacgactcactatagggcg
aattggagctccaccgcggtggcggccgcATATCAGGGATGTGAAACAGGGTCTTGA
AAACTGTCTAAATCTAAAACAATGCTAATGCAGGTTTAAATTTAATAAAATAAAATC
CAAAATCTAACAGCCAAGTCAAATCTGCATGTTTTAACATTTAAAATATTTTAAAGA
CGTCTTTTCCCAGGATTCAACATGTGAAATCTTTTCTCAGGGATACACGTGTGCCTA
GATCCTCATTGCTTTAGTTTTTTACAGAGGAATGAATATAAAAAGAAAATACTTAAA
TTTTATCCCTCTTACCTCTATAATCATACATAGGCATAATTTTTTAACCTAGGCTCC
AGATAGCCATAGAAGAACCAAACACTTTCTGCGTGTGTGAGAATAATCAGAGTGAGA
TTTTTTCACAAGTACCTGATGAGGGTTGAGACAGGTAGAAAAAGTGAGAGATCTCTA
TTTATTTAGCAATAATAGAGAAAGCATTTAAGAGAATAAAGCAATGGAAATAAGAAA
TTTGTAAATTTCCTTCTGATAACTAGAAATAGAGGATCCAGTTTCTTTTGGTTAACC
TAAATTTTATTTCATTTTATTGTTTTATTTTATTTTATTTTATTTTATTTTGTGTAA
TCGTAGTTTCAGAGTGTTAGAGCTGAAAGGAAGAAGTAGGAGAAACATGCAAAGTAA
AAGTATAACACTTTCCTTACTAAACCGACATGGGTTTCCAGGTAGGGGCAGGATTCA
GGATGACTGACAGGGCCCTTAGGGAACACTGAGACCCTACGCTGACCTCATAAATGC
TTGCTACCTTTGCTGTTTTAATTACATCTTTTAATAGCAGGAAGCAGAACTCTGCAC
TTCAAAAGTTTTTCCTCACCTGAGGAGTTAATTTAGTACAAGGGGAAAAAGTACAGG
GGGATGGGAGAAAGGCGATCACGTTGGGAAGCTATAGAGAAAGAAGAGTAAATTTTA
GTAAAGGAGGTTTAAACAAACAAAATATAAAGAGAAATAGGAACTTGAATCAAGGAA
ATGATTTTAAAACGCAGTATTCTTAGTGGACTAGAGGAAAAAAATAATCTGAGCCAA
GTAGAAGACCTTTTCCCCTCCTACCCCTACTTTCTAAGTCACAGAGGCTTTTTGTTC
CCCCAGACACTCTTGCAGATTAGTCCAGGCAGAAACAGTTAGATGTCCCCAGTTAAC
CTCCTATTTGACACCACTGATTACCCCATTGATAGTCACACTTTGGGTTGTAAGTGA

CTTTTTATTTATTTGTATTTTTGACTGCATTAAGAGGTCTCTAGTTTTTTATCTCTT
GTTTCCCAAAACCTAATAAGTAACTAATGCACAGAGCACATTGATTTGTATTTATTC
TATTTTTAGACATAATTTATTAGCATGCATGAGCAAATTAAGAAAAACAACAACAAA
TGAATGCATATATATGTATATGTATGTGTGTATATATACACACATATATATATATAT
TTTTTCTTTTCTTACCAGAAGGTTTTAATCCAAATAAGGAGAAGATATGCTTAGAAC
CGAGGTAGAGTTTTCATCCATTCTGTCCTGTAAGTATTTTGCATATTCTGGAGACGC
AGGAAGAGATCCATCTACATATCCCAAAGCTGAATTATGGTAGACAAAACTCTTCCA
CTTTTAGTGCATCAACTTCTTATTTGTGTAATAAGAAAATTGGGAAAACGATCTTCA
ATATGCTTACCAAGCTGTGATTCCAAATATTACGTAAATACACTTGCAAAGGAGGAT
GTTTTTAGTAGCAATTTGTACTGATGGTATGGGGCCAAGAGATATATCTTAGAGGGA
GGGCTGAGGGTTTGAAGTCCAACTCCTAAGCCAGTGCCAGAAGAGCCAAGGACAGGT
ACGGCTGTCATCACTTAGACCTCACCCTGTGGAGCCACACCCTAGGGTTGGCCAATC
TACTCCCAGGAGCAGGGAGGGCAGGAGCCAGGGCTGGGCATAAAAGTCAGGGCAGAG
CCATCTATTGCTTACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACA
GACACCATGGTGCATCTGACTCCTGAGGAgaattcgatGATCCAGACATGATAAGAT
ACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTT
GTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAG
TTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGG
TTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGCTGATTATGATCCTCTAG
AGTCGGTGGGCCTCGGGGGCGGGTGCGGGGTCGGCGGGGCCGCCCGGGTCGGCTTCG
GTCGGAGCCATGGGGTCGTGCGCTCCTTTCGGTCGGGCGCTGCGGGTCGTGGGGCGG
GCGTCAGGCACCGGGCTTGCGGGTCATGCACCAGGTGCGCGGTCCTTCGGGCACCTC
GACGTCGGCGGTGACGGTGAAGCCGAGCCGCTCGTAGAAGGGGAGGTTGCGGGGCGC
GGAGGTCTCCAGGAAGGCGGGCACCCCGGCGCGCTCGGCCGCCTCCACTCCGGGGAG
CACGACGGCGCTGCCCAGACCCTTGCCCTGGTGGTCGGGCGAGACGCCGACGGTGGC
CAGGAACCACGCGGGCTCCTTGGGCCGGTGCGGCGCCAGGAGGCCTTCCATCTGTTG
CTGCGCGGCCAGCCGGGAACCGCTCAACTCGGCCATGCGCGGGCCGATCTCGGCGAA
CACCGCCCCCGCTTCGACGCTCTCCGGCGTGGTCCAGACCGCCACCGCGGCGCCGTC
GTCCGCGACCCACACCTTGCCGATGTCGAGCCCGACGCGCGTGAGGAAGAGTTCTTG
CAGCTCGGTGACCCGCTCGATGTGGCGGTCCGGGTCGACGGTGTGGCGCGTGGCGGG
GTAGTCGGCGAACGCGGCGGCGAGGGTGCGTACGGCCCGGGGGACGTCGTCGCGGGT
GGCGAGGCGCACCGTGGGCTTGTACTCGGTCATGGAAGGTCGTCTCCTTGTGAGGGG
TCAGGGGCGTGGGTCAGGGGATGGTGGCGGCACCGGTCGTGGCGGCCGACCTGCAGG
CATGCAAGCTTTTTGCAAAAGCCTAGGCCTCCAAAAAAGCCTCCTCACTACTTCTGG
AATAGCTCAGAGGCCGAGGCGGCCTCGGCCTCTGCATAAATAAAAAAAATTAGTCAG
CCATGGGGCGGAGAATGGGCGGAACTGGGCGGAGTTAGGGGCGGGATGGGCGGAGTT
AGGGGCGGGACTATGGTTGCTGACTAATTGAGATGCATGCTTTGCATACTTCTGCCT
GCTGGGGAGCCTGGGGACTTTCCACACCTGGTTGCTGACTAATTGAGATGCATGCTT
TGCATACTTCTGCCTGCTGGGGAGCCTGGGGACTTTCCACACCCTAACTGACACACA
TTCCACAGatcaagcttatcgatCCTGGGCAGGTTGGTATCAAGGTTACAAGACAGG
TTTAAGGAGACCAATAGAAACTGGGCATGTGGAGACAGAGAAGACTCTTGGGTTTCT
GATAGGCACTGACTCTCTCTGCCTATTGGTCTATTTTCCCACCCTTAGGCTGCTGGT
GGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGA
TGCTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCTCGGTGCCTT
TAGTGATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAGTGA
GCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAACTTCAGGGTGAGTCTATGGGA
CGCTTGATGTTTTCTTTCCCCTTCTTTTCTATGGTTAAGTTCATGTCATAGGAAGGG
GATAAGTAACAGGGTACAGTTTAGAATGGGAAACAGACGAATGATTGCATCAGTGTG
GAAGTCTCAGGATCGTTTTAGTTTCTTTTATTTGCTGTTCATAACAATTGTTTTCTT
TTGTTTAATTCTTGCTTTCTTTTTTTTTCTTCTCCGCAATTTTTACTATTATACTTA
ATGCCTTAACATTGTGTATAACAAAAGGAAATATCTCTGAGATACATTAAGTAACTT
AAAAAAAAACTTTACACAGTCTGCCTAGTACATTACTATTTGGAATATATGTGTGCT

TATTTGCATATTCATAATCTCCCTACTTTATTTTCTTTTATTTTTAATTGATACATA
ATCATTATACATATTTATGGGTTAAAGTGTAATGTTTTAATATGTGTACACATATTG
ACCAAATCAGGGTAATTTTGCATTTGTAATTTTAAAAAATGCTTTCTTCTTTTAATA
TACTTTTTTGTTTATCTTATTTCTAATACTTTCCCTAATCTCTTTCTTTCAGGGCAA
TAATGATACAATGTATCATGCCTCTTTGCACCATTCTAAAGAATAACAGTGATAATT
TCTGGGTTAAGGCAATAGCAATATCTCTGCATATAAATATTTCTGCATATAAATTGT
AACTGATGTAAGAGGTTTCATATTGCTAATAGCAGCTACAATCCAGCTACCATTCTG
CTTTTATTTTATGGTTGGGATAAGGCTGGATTATTCTGAGTCCAAGCTAGGCCCTTT
TGCTAATCATGTTCATACCTCTTATCTTCCTCCCACAGCTCCTGGGCAACGTGCTGG
TCTGTGTGCTGGCCCATCACTTTGGCAAAGAATTCACCCCACCAGTGCAGGCTGCCT
ATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCCACAAGTATCACTAAGCTC
GCTTTCTTGCTGTCCAATTTCTATTAAAGGTTCCTTTGTTCCCTAAGTCCAACTACT
AAACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAATAAAAAACA
TTTATTTTCATTGCAATGATGTATTTAAATTATTTCTGAATATTTTACTAAAAAGGG
AATGTGGGAGGTCAGTGCATTTAAAACATAAAGAAATGAAGAGCTAGTTCAAACCTT
GGGAAAATACACTATATCTTAAACTCCATGAAAGAAGGTGAGGCTGCAAACAGCTAA
TGCACATTGGCAACAGCCCCTGATGCATATGCCTTATTCATCCCTCAGAAAAGGATT
CAAGTAGAGGCTTGATTTGGAGGTTAAAGTTTTGCTATGCTGTATTTTACATTACTT
ATTGTTTTAGCTGTCCTCATGAATGTCTTTTCACTACCCATTTGCTTATCCTGCATC
TCTCAGCCTTGACTCCACTCAGTTCTCTTGCTTAGAGATACCACCTTTCCCCTGAAG
TGTTCCTTCCATGTTTTACGGCGAGATGGTTTCTCCTCGCCTGGCCACTCAGCCTTA
GTTGTCTCTGTTGTCTTATAGAGGTCTACTTGAAGAAGGAAAAACAGGGGTCATGGT
TTGACTGTCCTGTGAGCCCTTCTTCCCTGCCTCCCCCACTCACAGTGACCCGGAATC
TGCAGTGCTAGTCTCCCGGAACTATCACTCTTTCACAGTCTGCTTTGGAAGGACTGG
GCTTAGTATGAAAAGTTAGGACTGAGAAGAATTTGAAAGGCGGCTTTTTGTAGCTTG
ATATTCACTACTGTCTTATTACCCTGTCATAGGCCCACCCCAAATGGAAGTCCCATT
CTTCCTCAGGATGTTTAAGATTAGCATTCAGGAAGAGATCAGAGGTCTGCTGGCTCC
CTTATCATGTCCCTTATGGTGCTTCTGGCTCTGCAGTTATTAGCATAGTGTTACCAT
CAACCACCTTAACTTCATTTTTCTTATTCAATACCTAGGTAGGTAGATGCTAGATTC
TGGAAATAAAATATGAGTCTCAAGTGGTCCTTGTCCTCTCTCCCAGTCAggtaccca
gcttttgttccctttagtgagggttaattgcgcgcttggcgtaatcatggtcatagc
tgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaa

Red: Homology arms, left and right

Light blue: PT2-F

Green: PT3-F

Yellow: PT1-R

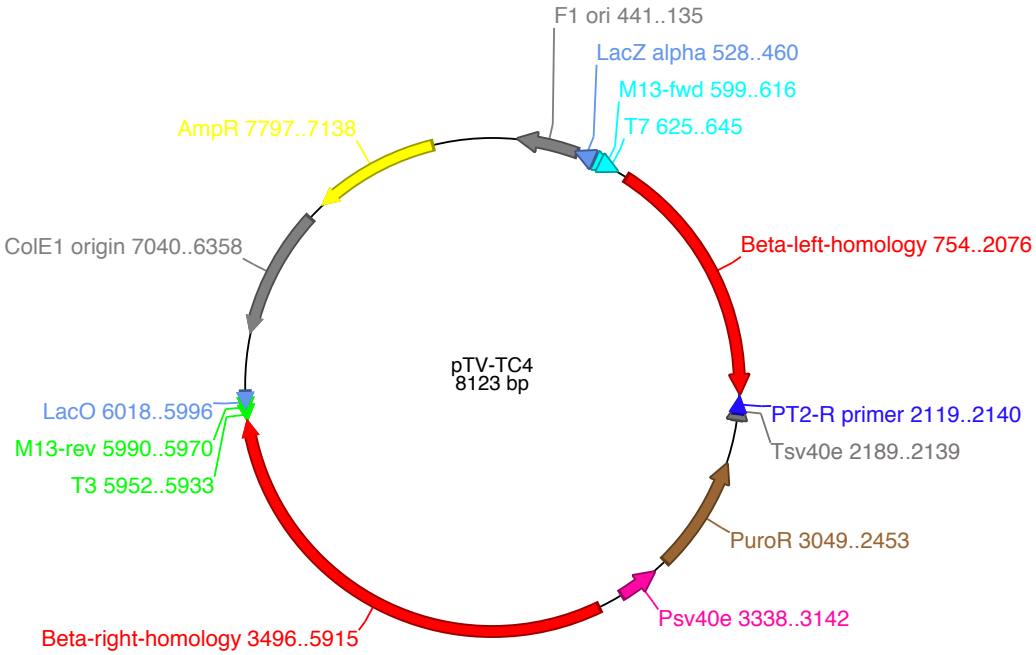
Dark blue: PT2-R

Light brown: Puro^{R} cassette

Dark grey: Tsv40e terminator

Pink: Psv40e

pTV-TC4 map:



pTV-TC4 partial sequence (regions of interest):

ttggagctccaccgcggtggcggccgctctagaactagtggatccccgggctgcag
gaattcgatatcaagcttatcgataccgtcgacctcgaggggggggccCTTAGGGAAC
ACTGAGACCCTACGCTGACCTCATAAATGCTTGCTACCTTTGCTGTTTTAATTACAT
CTTTTAATAGCAGGAAGCAGAACTCTGCACTTCAAAAGTTTTTCCTCACCTGAGGAG
TTAATTTAGTACAAGGGGAAAAAGTACAGGGGGATGGGAGAAAGGCGATCACGTTGG
GAAGCTATAGAGAAAGAAGAGTAAATTTTAGTAAAGGAGGTTTAAACAAACAAAATA
TAAAGAGAAATAGGAACTTGAATCAAGGAAATGATTTTAAAACGCAGTATTCTTAGT
GGACTAGAGGAAAAAAATAATCTGAGCCAAGTAGAAGACCTTTTCCCCTCCTACCCC
TACTTTCTAAGTCACAGAGGCTTTTTGTTCCCCCAGACACTCTTGCAGATTAGTCCA
GGCAGAAACAGTTAGATGTCCCCAGTTAACCTCCTATTTGACACCACTGATTACCCC
ATTGATAGTCACACTTTGGGTTGTAAGTGACTTTTTATTTATTTGTATTTTTGACTG
CATTAAGAGGTCTCTAGTTTTTTATCTCTTGTTTCCCAAAACCTAATAAGTAACTAA
TGCACAGAGCACATTGATTTGTATTTATTCTATTTTTAGACATAATTTATTAGCATG
CATGAGCAAATTAAGAAAAACAACAACAAATGAATGCATATATATGTATATGTATGT
GTGTATATATACACACATATATATATATATTTTTTCTTTTCTTACCAGAAGGTTTTA
ATCCAAATAAGGAGAAGATATGCTTAGAACCGAGGTAGAGTTTTCATCCATTCTGTC
CTGTAAGTATTTTGCATATTCTGGAGACGCAGGAAGAGATCCATCTACATATCCCAA
AGCTGAATTATGGTAGACAAAACTCTTCCACTTTTAGTGCATCAACTTCTTATTTGT
GTAATAAGAAAATTGGGAAAACGATCTTCAATATGCTTACCAAGCTGTGATTCCAAA
TATTACGTAAATACACTTGCAAAGGAGGATGTTTTTAGTAGCAATTTGTACTGATGG
TATGGGGCCAAGAGATATATCTTAGAGGGAGGGCTGAGGGTTTGAAGTCCAACTCCT
AAGCCAGTGCCAGAAGAGCCAAGGACAGGTACGGCTGTCATCACTTAGACCTCACCC

TGTGGAGCCACACCCTAGGGTTGGCCAATCTACTCCCAGGAGCAGGGAGGGCAGGAG
CCAGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGA
CACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTGAG
GAgaattcgatGATCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAA
CTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTAT
TTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTA
TGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACA
AATGTGGTATGGCTGATTATGATCCTCTAGAGTCGGTGGGCCTCGGGGGCGGGTGCG
GGGTCGGCGGGGCCGCCCGGGTCGGCTTCGGTCGGAGCCATGGGGTCGTGCGCTCCT
TTCGGTCGGGCGCTGCGGGTCGTGGGGCGGGCGTCAGGCACCGGGCTTGCGGGTCAT
GCACCAGGTGCGCGGTCCTTCGGGCACCTCGACGTCGGCGGTGACGGTGAAGCCGAG
CCGCTCGTAGAAGGGGAGGTTGCGGGGCGCGGAGGTCTCCAGGAAGGCGGGCACCCC
GGCGCGCTCGGCCGCCTCCACTCCGGGGAGCACGACGGCGCTGCCCAGACCCTTGCC
CTGGTGGTCGGGCGAGACGCCGACGGTGGCCAGGAACCACGCGGGCTCCTTGGGCCG
GTGCGGCGCCAGGAGGCCTTCCATCTGTTGCTGCGCGGCCAGCCGGGAACCGCTCAA
CTCGGCCATGCGCGGGCCGATCTCGGCGAACACCGCCCCCGCTTCGACGCTCTCCGG
CGTGGTCCAGACCGCCACCGCGGCGCCGTCGTCCGCGACCCACACCTTGCCGATGTC
GAGCCCGACGCGCGTGAGGAAGAGTTCTTGCAGCTCGGTGACCCGCTCGATGTGGCG
GTCCGGGTCGACGGTGTGGCGCGTGGCGGGGTAGTCGGCGAACGCGGCGGCGAGGGT
GCGTACGGCCCGGGGGACGTCGTCGCGGGTGGCGAGGCGCACCGTGGGCTTGTACTC
GGTCATGGAAGGTCGTCTCCTTGTGAGGGGTCAGGGGCGTGGGTCAGGGGATGGTGG
CGGCACCGGTCGTGGCGGCCGACCTGCAGGCATGCAAGCTTTTTGCAAAAGCCTAGG
CCTCCAAAAAAGCCTCCTCACTACTTCTGGAATAGCTCAGAGGCCGAGGCGGCCTCG
GCCTCTGCATAAATAAAAAAAATTAGTCAGCCATGGGGCGGAGAATGGGCGGAACTG
GGCGGAGTTAGGGGCGGGATGGGCGGAGTTAGGGGCGGGACTATGGTTGCTGACTAA
TTGAGATGCATGCTTTGCATACTTCTGCCTGCTGGGGAGCCTGGGGACTTTCCACAC
CTGGTTGCTGACTAATTGAGATGCATGCTTTGCATACTTCTGCCTGCTGGGGAGCCT
GGGGACTTTCCACACCCTAACTGACACACATTCCACAGatcaagcttatcgatCCTG
GGCAGGTTGGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAAACTGGGCA
TGTGGAGACAGAGAAGACTCTTGGGTTTCTGATAGGCACTGACTCTCTCTGCCTATT
GGTCTATTTTCCCACCCTTAGGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTT
TGAGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAGGTGAA
GGCTCATGGCAAGAAAGTGCTCGGTGCCTTTAGTGATGGCCTGGCTCACCTGGACAA
CCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTGTGACAAGCTGCACGTGGA
TCCTGAGAACTTCAGGGTGAGTCTATGGGACGCTTGATGTTTTCTTTCCCCTTCTTT
TCTATGGTTAAGTTCATGTCATAGGAAGGGGATAAGTAACAGGGTACAGTTTAGAAT
GGGAAACAGACGAATGATTGCATCAGTGTGGAAGTCTCAGGATCGTTTTAGTTTCTT
TTATTTGCTGTTCATAACAATTGTTTTCTTTTGTTTAATTCTTGCTTTCTTTTTTTT
TCTTCTCCGCAATTTTTACTATTATACTTAATGCCTTAACATTGTGTATAACAAAAG
GAAATATCTCTGAGATACATTAAGTAACTTAAAAAAAAACTTTACACAGTCTGCCTA
GTACATTACTATTTGGAATATATGTGTGCTTATTTGCATATTCATAATCTCCCTACT
TTATTTTCTTTTATTTTTAATTGATACATAATCATTATACATATTTATGGGTTAAAG
TGTAATGTTTTAATATGTGTACACATATTGACCAAATCAGGGTAATTTTGCATTTGT
AATTTTAAAAAATGCTTTCTTCTTTTAATATACTTTTTTGTTTATCTTATTTCTAAT
ACTTTCCCTAATCTCTTTCTTTCAGGGCAATAATGATACAATGTATCATGCCTCTTT
GCACCATTCTAAAGAATAACAGTGATAATTTCTGGGTTAAGGCAATAGCAATATCTC
TGCATATAAATATTTCTGCATATAAATTGTAACTGATGTAAGAGGTTTCATATTGCT
AATAGCAGCTACAATCCAGCTACCATTCTGCTTTTATTTTATGGTTGGGATAAGGCT
GGATTATTCTGAGTCCAAGCTAGGCCCTTTTGCTAATCATGTTCATACCTCTTATCT
TCCTCCCACAGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCA
AAGAATTCACCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTA
ATGCCCTGGCCCACAAGTATCACTAAGCTCGCTTTCTTGCTGTCCAATTTCTATTAA

AGGTTCCTTTGTTCCCTAAGTCCAACTACTAAACTGGGGGATATTATGAAGGGCCTT
GAGCATCTGGATTCTGCCTAATAAAAAACATTTATTTTCATTGCAATGATGTATTTA
AATTATTTCTGAATATTTTACTAAAAAGGGAATGTGGGAGGTCAGTGCATTTAAAAC
ATAAAGAAATGAAGAGCTAGTTCAAACCTTGGGAAAATACACTATATCTTAAACTCC
ATGAAAGAAGGTGAGGCTGCAAACAGCTAATGCACATTGGCAACAGCCCCTGATGCA
TATGCCTTATTCATCCCTCAGAAAAGGATTCAAGTAGAGGCTTGATTTGGAGGTTAA
AGTTTTGCTATGCTGTATTTTACATTACTTATTGTTTTAGCTGTCCTCATGAATGTC
TTTTCACTACCCATTTGCTTATCCTGCATCTCTCAGCCTTGACTCCACTCAGTTCTC
TTGCTTAGAGATACCACCTTTCCCCTGAAGTGTTCCTTCCATGTTTTACGGCGAGAT
GGTTTCTCCTCGCCTGGCCACTCAGCCTTAGTTGTCTCTGTTGTCTTATAGAGGTCT
ACTTGAAGAAGGAAAAACAGGGGTCATGGTTTGACTGTCCTGTGAGCCCTTCTTCCC
TGCCTCCCCCACTCACAGTGACCCGGAATCTGCAGTGCTAGTCTCCCGGAACTATCA
CTCTTTCACAGTCTGCTTTGGAAGGACTGGGCTTAGTATGAAAAGTTAGGACTGAGA
AGAATTTGAAAGGCGGCTTTTTGTAGCTTGATATTCACTACTGTCTTATTACCCTGT
CATAGGCCCACCCCAAATGGAAGTCCCATTCTTCCTCAGGATGTTTAAGATTAGCAT
TCAGGAAGAGATCAGAGGTCTGCTGGCTCCCTTATCATGTCCCTTATGGTGCTTCTG
GCTCTGCAGTTATTAGCATAGTGTTACCATCAACCACCTTAACTTCATTTTTCTTAT
TCAATACCTAGGTAGGTAGATGCTAGATTCTGGAAATAAAATATGAGTCTCAAGTGG
TCCTTGTCCTCTCTCCCAGTCAggtaccccagcttttgttcccttta gtgagggtta
attgcgcgcttggcgt

Red: Homology arms, left and right

Dark blue: PT2-R

Light brown: Puro^{R} cassette

Dark grey: Tsv40e terminator

Pink: Psv40e

β-globin sequence plus surrounding:

aagcttgaatagttttcaggaactttgaatgctgatttagatttgaaactgaggct
ctgaccataaccaaatttgcactatttattgcttcttgaaacttatttgcctggtat
gcctgggcttttgatggtcttagtatagcttgcagccttgtccctgcagggtattat
gggtaatagaaagaaaagtctgcgttacactctagtcacactaagtaactaccattg
gaaaagcaacccctgccttgaagccaggatgatggtatctgcagcagttgccaacac
aagagaaggatccatagttcatcatttaaaaaagaaaacaaaatagaaaaaggaaaa
ctatttctgagcataagaagttgtagggtaagtctttaagaaggtgacaatttctgc
caatcaggatttcaaagctcttgctttgacaattttggtctttcagaatactataaa
tataacctatattataatttcataaagtctgtgcatttctttgacccaggatattt
gcaaaagacatattcaaacttccgcagaacactttatttcacatatacatgcctctt
atatcagggatgtgaaacagggtcttgaaaactgtctaaatctaaaacaatgctaat
gcaggtttaaatttaataaaataaaatccaaaatctaacagccaagtcaaatctgca
tgttttaacatttaaaatattttaaagacgtcttttcccaggattcaacatgtgaaa
tcttttctcagggatacacgtgtgcctagatcctcattgctttagtttttacagag
gaatgaatataaaaagaaaatacttaaattttatccctcttacctctataatcatac

ataggcataatttttaacctaggctccagatagccatagaagaaccaaacactttc
tgcgtgtgtgagaataatcagagtgagatttttcacaagtacctgatgagggttga
gacaggtagaaaaagtgagagatctctatttatttagcaataatagagaaagcattt
aagagaataaagcaatggaaataagaaatttgtaaatttccttctgataactagaaa
tagaggatccagtttcttttggttaacctaaatttttatttcattttattgttttatt
ttatttttatttttatttttatttgtgtaatcgtagtttcagagtgttagaagctgaaag
gaagaagtaggagaaacatgcaaagtaaaagtataacactttccttactaaaccgac
atgggtttccaggtaggggcaggattcaggatgactgacagggccttagggaacac
tgagaccctacgctgacctcataaatgcttgctacctttgctgttttaattacatct
tttaatagcaggaagcagaactctgcacttcaaaagtttttcctcacctgaggagtt
aatttagtacaaggggaaaaagtacaggggatgggagaaaggcgatcacgttggga
agctatagagaaagaagagtaaatttagtaaaggaggtttaaacaaacaaaatata
aagagaaataggaacttgaatcaaggaaatgattttaaaacgcagtattcttagtgg
actagaggaaaaaaataatctgagccaagtagaagacctttcccctcctacccta
ctttctaagtcacagaggcttttgttcccccagacactcttgcagattagtccagg
cagaaacagttagatgtccccagttaacctcctatttgacaccactgattaccccat
tgatagtcacactttgggttgtaagtgacttttattttatttgtatttttgactgca
ttaagaggtctctagtttttatctcttgtttcccaaaacctaataagtaactaatg
cacagagcacattgatttgtatttattctatttttagacataatttattagcatgca
tgagcaaattaagaaaaacaacaacaaatgaatgcatatatatgtatatgtatgtgt
gtatatatacacacatatatatatatttttttcttttcttaccagaaggttttaat
ccaaataaggagaagatatgcttagaaccgaggtagagttttcatccattctgtcct
gtaagtattttgcatattctggagacgcaggaagagatccatctacatatcccaaag
ctgaattatggtagacaaaactcttccacttttagtgcatcaacttcttatttgtgt
aataagaaaattgggaaaacgatcttcaatatgcttaccaagctgtgattccaaata
ttacgtaaatacacttgcaaaggaggatgtttttagtagcaatttgtactgatggta
tggggccaagagatatatcttagagggagggctgagggtttgaagtccaactcctaa
gccagtgccagaagagccaaggacaggtacggctgtcatcacttagacctcaccctg
tggagccacaccctagggttggccaatctactcccaggagcagggagggcaggagcc
agggctgggcataaaagtcagggcagagccatctattgcttACATTTGCTTCTGACA
CAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTGAGGA
GAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGA
GGCCCTGGGCAGgttggtatcaaggttacaagacaggtttaaggagaccaatagaaa
ctgggcatgtggagacagagaagactcttgggtttctgataggcactgactctctct
gcctattggtctattttcccacccttagGCTGCTGGTGGTCTACCCTTGGACCCAGA
GGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTA
AGGTGAAGGCTCATGGCAAGAAAGTGCTCGGTGCCTTTAGTGATGGCCTGGCTCACC
TGGACAACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTGTGACAAGCTGC
ACGTGGATCCTGAGAACTTCAGGgtgagtctatgggacgcttgatgttttctttccc
cttcttttctatggttaagttcatgtcataggaaggggataagtaacagggtacagt
ttagaatgggaaacagacgaatgattgcatcagtgtggaagtctcaggatcgtttta
gtttcttttatttgctgttcataacaattgttttcttttgtttaattcttgctttct
ttttttttcttctccgcaatttttactattatacttaatgccttaacattgtgtata
acaaaaggaaatatctctgagatacattaagtaacttaaaaaaaaactttacacagt
ctgcctagtacattactatttggaatatatgtgtgcttatttgcatattcataatct
ccctactttatttcttttatttttaattgatacataatcattatacatatttatgg
gttaaagtgtaatgttttaatatgtgtacacatattgaccaaatcagggtaattttg
catttgtaatttaaaaaatgctttcttcttttaatatacttttttgtttatcttat
ttctaatactttccctaatctctttctttcagggcaataatgatacaatgtatcatg
cctctttgcaccattctaaagaataacagtgataatttctgggttaaggcaatagca
atatctctgcatataaatatttctgcatataaattgtaactgatgtaagaggtttca
tattgctaatagcagctacaatccagctaccattctgcttttatttatggttggga

taaggctggattattctgagtccaagctaggccctttgctaatcatgttcatacct
cttatcttcctcccacag**CTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCAC**
TTTGGCAAAGAATTCACCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGT
GTGGCTAATGCCCTGGCCCACAAGTATCACTAAGCTCGCTTTCTTGCTGTCCAATTT
CTATTAAAGGTTCCTTTGTTCCCTAAGTCCAACTACTAAACTGGGGGATATTATGAA
GGGCCTTGAGCATCTGGATTCTGCCTAATAAAAAACATTTATTTTCATTGCaatgat
gtatttaaattatttctgaatattttactaaaaagggaatgtgggaggtcagtgcat
ttaaaacataaagaaatgaagagctagttcaaaccttgggaaaatacactatatctt
aaactccatgaaagaaggtgaggctgcaaacagctaatgcacattggcaacagcccc
tgatgcatatgccttattcatccctcagaaaaggattcaagtagaggcttgatttgg
aggttaaagttttgctatgctgtattttacattacttattgttttagctgtcctcat
gaatgtcttttcactacccatttgcttatcctgcatctctcagccttgactccactc
agttctcttgcttagagataccacctttcccctgaagtgttccttccatgttttacg
gcgagatggtttctcctcgcctggccactcagccttagttgtctctgttgtcttata
gaggtctacttgaagaaggaaaaacaggggtcatggtttgactgtcctgtgagccct
tcttccctgcctcccccactcacagtgacccggaatctgcagtgctagtctcccgga
actatcactctttcacagtctgctttggaaggactgggcttagtatgaaaagttagg
actgagaagaatttgaaaggcggctttttgtagcttgatattcactactgtcttatt
accctgtcataggcccaccccaaatggaagtcccattcttcctcaggatgtttaaga
ttagcattcaggaagagatcagaggtctgctggctcccttatcatgtcccttatggt
gcttctggctctgcagttattagcatagtgttaccatcaaccaccttaacttcattt
ttcttattcaatacctaggtaggtagatgctagattctggaaataaaatatgagtct
caagtggtccttgtcctctctcccagtcaaattctgaatctagttggcaagattctg
aaatcaaggcatataatcagtaataagtgatgatagaagggtatatagaagaatttt
attatatgagagggtgaaaccctcaaaatgaaatgaaatcagacccttgtcttacac
cataaacaaaaataaatttgaatgggttaaagaattaaactaagacctaaaaccata
aaaatttttaaagaaatcaaaagaagaaaattctaatattcacgttgcagccgtttt
ttgaatttgatatgagaagcaaaggcaacaaaaggaaaaataaagaagtgaggctac
atcaaactaaaaaatttccacacaaaaaacaaaacaatgaacaaatgaaaggtga

Bold uppercase: exons

Red: regions homologous with the left and right homology arms of pTV-TC4

Yellow: ZFN4 target sequence

Blue: PT3-F primer (PCR assay)

Green: Southern-blot probe's binding region