CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Cystic fibrosis gene repair: correction of ΔF508 using ZFN and CRISPR/Cas9 guide RNA gene editing tools
Authors	Hollywood, Jennifer A.
Publication date	2013
Original Citation	Hollywood, J. 2013. Cystic fibrosis gene repair: correction of ΔF508 using ZFN and CRISPR/Cas9 guide RNA gene editing tools. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2013, Jennifer Hollywood - http://creativecommons.org/licenses/by-nc-nd/3.0/
Download date	2026-10-13 07:28:22
Item downloaded from	https://hdl.handle.net/10468/1407

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Cystic fibrosis gene repair: Correction of ΔF508 using ZFN and CRISPR/Cas9 guide RNA gene editing tools

Jennifer Hollywood B.Sc.

Thesis submitted for the degree of Ph.D. to the National University of Ireland, Cork

Department of Physiology & Department of Microbiology

Heads of Department: Prof. Ken O' Halloran, Prof. Ger Fitzgerald

Research Supervisors: Dr. Patrick Harrison, Dr. Martina Scallan

Submitted October 2013

Table of Contents

Abstract ..... **x**

Abbreviations ..... **xi**

1 Introduction **1**

1.1 Cystic Fibrosis 1

1.1.1 A brief history 1

1.1.2 Cloning of the gene 2

1.1.3 Cystic Fibrosis Transmembrane Regulator 3

1.1.4 Airway Surface Liquid (ASL) regulation 5

1.1.5 CF today 6

1.1.6 Diagnosis 7

1.1.7 Bacterial Infections and Inflammation 7

1.2 CF Mutations 10

1.2.1 Mutations and classification 10

1.2.2 Mutations and disease severity 12

1.2.3 ΔF508 14

1.3 Mutation Specific Therapies 14

1.3.1 Aminoglycosides for Premature Stop Codons 14

1.3.2 Correctors 15

1.3.3 Potentiators 16

1.3.4 Lung tissue engineering 17

1.4 Gene Therapy for CF 18

1.4.1 CF: A good and bad candidate for gene therapy 18

1.4.2 Which cells/ part of the lung to target? 20

1.4.3 Efficacy is difficult to detect 21
1.4.4 Animal models to study the disease 21
1.4.5 CF gene therapy to date 22
1.5 Gene addition vs Gene repair 27
1.6 Genome editing 29
1.6.1 DNA repair mechanisms 29
1.6.2 Homologous Recombination 31
1.6.3 Single Strand Annealing 32
1.6.4 Non-Homologous End Joining 32
1.6.5 Regulation of repair pathway choice 33
1.6.6 Homology-directed repair (HDR) 34
1.7 ZFN, TALEN and CRISPR/Cas-based tools for genome editing 37
1.7.1 Zinc Finger Proteins (ZFPs) 37
1.7.2 Zinc finger nucleases: designer nucleases for HDR 40
1.7.3 Zinc Finger Protein Design 42
1.7.4 Modular Design 43
1.7.5 Rational Design 44
1.7.6 ZFP Oligomerised Pool Engineering “OPEN” Selection 45
1.7.7 Context-dependent assembly (CoDA) 46
1.7.8 Transcription Activator-like effectors (TALE) 46
1.7.9 CRISPR/Cas system 48
1.8 Gene modifications using programmable and RNA-guided nucleases 51
1.8.1 Gene Disruption 52
1.8.2 Gene Correction 54
1.8.3 Gene Correction and iPs cells 55
1.9 Off-Target activity 59
1.9.1 FokI heterodimerisation 59

1.9.2 Nickases 61
1.10 Aims: 62
1.11 Objectives: 62
2 Materials and Methods 63
2.1 Cell Lines 63
2.2 Cell Culture Reagents 63
2.2.1 Cell Culture Medium 63
2.2.2 Other reagents 64
2.3 Bacterial Strains 64
2.4 Bacterial Growth Media and Antibiotics 65
2.4.1 Growth Media 65
2.4.2 Antibiotics 65
2.5 Buffers, solutions and drugs 66
2.6 Molecular biology reagents 67
2.7 Chemical Reagents 68
2.8 Other Reagents 68
2.9 Plasmids 69
2.9.1 Plasmids provided 69
2.9.2 Plasmids constructed 69
2.10 Kits 70
2.11 Oligonucleotides 71
2.12 Equipment 74
2.13 Software 74
2.14 Cell culture protocols 75
2.14.1 Subculturing cells 75
2.14.2 Cell quantification and seeding density 75
2.15 Bacterial protocols 76

2.15.1 Inoculation of LB broth 76
2.15.2 Bacterial transformation 76
2.16 DNA and RNA isolation 76
2.16.1 Plasmid extraction (Qiagen Plasmid Plus Purification Kit) 76
2.16.2 Plasmid extraction (Qiagen Miniprep Kit) 77
2.16.3 Plasmid extraction (Qiagen Maxiprep Kit) 77
2.16.4 Measuring DNA concentration 78
2.16.5 RNA extraction 78
2.16.6 Measuring RNA concentration 79
2.17 DNA and RNA manipulation 79
2.17.1 Restriction digests 79
2.17.2 Preparing DNA agarose gels 79
2.17.3 Loading and running agarose gels 80
2.17.4 RNA gel 80
2.17.5 Gel purification 80
2.17.6 Annealing oligonucleotides 81
2.17.7 Ligation 81
2.17.8 De-phosphorylation 81
2.17.9 Golden gate cloning 81
2.17.10 Site-directed mutagenesis 81
2.17.11 Reverse Transcription 82
2.17.12 Polymerase chain reaction 83
2.17.13 High Fidelity Polymerase Chain Reaction 83
2.17.14 PCR purification 83
2.17.15 Cel-1 (Surveyor) Nuclease Assay 84
2.17.16 T7E1 assay 84
2.17.17 In vitro transcription translation reaction 84

2.18 Transfection 85
2.18.1 Transfection of cell lines 85
2.19 Fluorescence Microscopy 85
2.20 Flow Cytometry 85
2.21 Genotyping 85
2.22 Immuno-staining 85
2.23 Enrichment of GFP positive cells 86
2.24 Next Generation Sequencing 87
2.25 Miscellaneous methods 88
2.25.1 Dual-luciferase reporter assay 88
3 Design and construction of pITR-donor and pITR-donor-XC donor repair plasmids, and correction of ΔF508 mutation in the CFTE cell line 89
3.1 Aims: 89
3.2 Objectives: 89
3.3 Design and construction of pITR-donor and pITR-donor-XC 89
3.3.1 Cloning of CFTR into AAV backbone 89
3.3.2 Site-Directed Mutagenesis on pITR-donor to yield pITR-donorXC . 92
3.4 Development of a wild type specific RT-PCR assay to detect homology directed repair with pITR-donor 94
3.5 CFi9-L^{EL}/CFi9-R^{KK} ZFNs mediate HDR in the presence of pITR-donor or pITR-donor-XC 94
3.6 Quantification of the level of repair 97
3.6.1 XhoI/ClaI digests: 97
3.6.2 Genotyping 98
3.6.3 Next generation sequencing (NGS) data 99
3.7 Conclusions 100

4 Mini-gene Repair Strategy using ZFNs 103

4.1 Aims: 103

4.2 Objectives: 103

4.3 Design and construction of three *Nhe*I donor plasmids with homology arms of 4.3, 1.5 and 1 kb. 104

4.3.1 pITR-CFTE and pITR-CFTE-NheI 105

4.3.2 Construction of pUC-CFTE-NheI-1 kb and pUC-CFTE-NheI-1.5 kb donors 107

4.4 Development of detection assays to determine if *Nhe*I is introduced into the genome following ZFN mediated HDR with *Nhe*I donor plasmids 109

4.4.1 NheI-specific primer reaction 110

4.4.2 NheI quantification assay 111

4.5 Successful incorporation of *Nhe*I tag into CFTR using CFi9-L^{EL}/CFi9-R^{KK} ZFNs and pUC-CFTE-*Nhe*I-1.5 kb donor 112

4.6 Design and construction of mini-gene construct to correct >80% of CF-causing mutations by ZFN mediated HDR 115

4.6.1 Design of mini-gene donor 115

4.6.2 Construction of mini-gene donor 116

4.7 Detection of HDR with CFi9-L^{EL}/CFi9-R^{KK} ZFNs and pUC-mini-gene donor 121

4.7.1 Transfection with ZFNs and mini-gene donor produce GFP positive cells 121

4.7.2 Enrichment of GFP positive cells 123

4.7.3 Fluorescent Activated Cell Sorting (FACs) analysis 124

4.7.4 Does addition of SAHA increase HDR as determined by FACs analysis? 125

4.7.5 PCR-based assays to determine if HDR occurred at the correct location 126

4.7.6 Next Generation Sequencing sheds light on mini-gene HDR..... 127

4.8 Conclusions 127

5 CRISPR/Cas9/gRNA system as an alternative to ZFNs to mediate HDR in the presence of donor plasmids 132

5.1 Aims: 132

5.2 Objectives: 132

5.3 Design of two CRISPR gRNAs to target and cleave CFTR 133

5.3.1 One-step cloning of CRISPR IN9 and CRISPR EX10 134

5.4 Determination of cleavage efficiency of CRISPR IN9 and CRISPR EX10..... 135

5.4.1 NGS data for CRISPR cleavage..... 137

5.5 CRISPR/Cas9 system can mediate HDR in the presence of appropriate donor plasmids. 138

5.5.1 CRISPR EX10 can mediate HDR with pITR-donor-XC..... 139

5.5.2 NGS data for CRISPR ex10 and pITR-donor-XC 141

5.5.3 CRISPR IN9 gRNA HDR with the pUC-mini-gene donor 142

5.6 Conclusions: 143

6 Discussion: 145

7 Bibliography 152

8 Appendix: 171

8.1 Functional Assay: Immunofluorescence for detection of CFTR protein trafficking 171

8.2 Transfection of CFTE cells with pCEP-CFTR and pEGFP 172

Declaration:

This thesis is the candidate's own work and has not been submitted for another degree, either at University College Cork or elsewhere.

Signed: _____

Peer Reviewed Publications generated from this work:

- C.M. Lee, R. Flynn, **J.A. Hollywood,** M.F. Scallan, P.T. Harrison (2012). Correction of the ΔF508 mutation in the cystic fibrosis transmembrane conductance regulator gene by zinc-finger nuclease homology-directed repair. *BioResearch Open Access* **1(3)**: 99-108.

Manuscript in preparation:

- **J.A. Hollywood,** CM. Lee, MF. Scallan, PT. Harrison. Genome editing of cystic fibrosis cells using gRNA-Cas9 system to correct the ΔF508 mutation

Abstracts and Poster Presentations:

- **J.A. Hollywood**, K. Kaschig, C.M. Lee, M.F. Scallan, P.T. Harrison (September 2010). *ctns* gene repair using AAV mutagenesis. 6th International Cystinosis Conference, Lignano, ITALY. Poster presentation.

- **J.A. Hollywood**, C.M. Lee, K. Kaschig, M.F. Scallan, P.T. Harrison (October 2011). Correction of the DF508 mutation in the CFTR gene in Epithelial Cells using Zinc Finger Nuclease Homology-Directed Repair. European Society of Gene and Cell Therapy conference in Brighton Poster presentation.

- **J.A. Hollywood**, C.M. Lee, K. Kaschig, M.F. Scallan, P.T. Harrison (July 2011). Correction of the DF508 mutation in the CFTR gene in Epithelial Cells using Zinc Finger Nuclease Homology-Directed Repair. Physiological Society Annual Meeting, Oxford 201.1 Poster presentation. (Shortlisted for poster prize)

- **J.A. Hollywood**, C.M. Lee, K. Kaschig, M.F. Scallan, P.T. Harrison (April 2012). Correction of the DF508 mutation in the CFTR gene in Epithelial Cells using Zinc Finger Nuclease Homology-Directed Repair. 9th European CF Basic Science meeting, Paris 2012. Poster presentation and invited 2 min talk.

- **J.A. Hollywood**, C.M. Lee, K. Kaschig, M.F. Scallan, P.T. Harrison (May 2012). Correction of the DF508 mutation in the CFTR gene in Epithelial Cells using Zinc Finger Nuclease Homology-Directed Repair. 6th European CF Young Investigator Meeting, Paris 2012. Oral presentation.

- **J.A. Hollywood**, C.M. Lee, K. Kaschig, M.F. Scallan, P.T. Harrison (June 2013). Correction of >80% CF-causing mutations by ZFN Homology-Directed Repair. 37th European Cystic Fibrosis Society Meeting, Lisbon, 2013.

Abstract

Cystic Fibrosis (CF) is an autosomal recessive monogenic disorder caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene with the ΔF508 mutation accounting for approximately 70% of all CF cases worldwide. This thesis investigates whether existing zinc finger nucleases designed in this lab and CRISPR/gRNAs designed in this thesis can mediate efficient homology-directed repair (HDR) with appropriate donor repair plasmids to correct CF-causing mutations in a CF cell line.

Firstly, the most common mutation, ΔF508, was corrected using a pair of existing ZFNs, which cleave in intron 9, and the donor repair plasmid pITR-donor-XC, which contains the correct CTT sequence and two unique restriction sites. HDR was initially determined to be <1% but further analysis by next generation sequencing (NGS) revealed HDR occurred at a level of 2%.

This relatively low level of repair was determined to be a consequence of distance from the cut site to the mutation and so rather than designing a new pair of ZFNs, the position of the existing intron 9 ZFNs was exploited and attempts made to correct >80% of CF-causing mutations. The ZFN cut site was used as the site for HDR of a mini-gene construct comprising exons 10-24 from CFTR cDNA (with appropriate splice acceptor and poly A sites) to allow production of full length corrected CFTR mRNA.

Finally, the ability to cleave closer to the mutation and mediate repair of CFTR using the latest gene editing tool CRISPR/Cas9 was explored. Two CRISPR gRNAs were tested; CRISPR ex10, which targets within exon 10, 85 bp upstream of the ΔF508, was shown to cleave at an efficiency of 15% and CRISPR in9 targets intron 9 and cleaved at 3%. Both CRISPR gRNAs mediated HDR with appropriate donor plasmids at a rate of ~1% as determined by NGS. This is the first evidence of CRISPR induced HDR in CF cell lines.

Abbreviations

AAV- Adeno-associated virus

ABC – ATP binding cassette

ADP- Adenosine diphosphate

AdV – Adenovirus

ASL – Airway surface liquid

ATP – Adenosine triphosphate

B2H – Bacterial-2-hybrid

BRCA1 – Breast cancer 1, early onset

BRCA2 – Breast cancer 2, early onset

CaCC – Calcium-activated chloride channels

cAMP – cyclic adenosine monophosphate

CAT – Chloramphenicol acetyltransferase

CAR – Coxsackie B and Adenovirus type 2 and 5 receptor

CCR5 – Chemokine co-receptor 5

CCR2 – Chemokine co-receptor 2

cDNA- complimentary DNA

CDK – Cyclin-dependent protein kinases

CF – Cystic fibrosis

CFBE – Cystic fibrosis bronchial epithelium

CFTE – Cystic fibrosis tracheal epithelium

CFTR – Cystic fibrosis transmembrane conductance regulator

CMV - Cytomegalovirus

CoDA – Context-dependent assembly

CRISPR – Clustered regularly interspaced short palindromic repeats

crRNA – CRISPR RNA

DNA-PKCS – DNA protein kinase catalytic subunit

DSB – Double-stranded break

EGFP – Enhanced green fluorescent protein

ENaC – Epithelial sodium channel

ER – Endoplasmic reticulum

ES cell – Embryonic stem cell

FEV1 – Forced expired volume in one second

GFP – Green fluorescent protein

gRNA – guide RNA

H2AX – H2A histone family member X

HAE – Human airway epithelium

HBE – Human bronchial epithelium

HDR – Homology directed repair

HEK293 – Human embryonic kidney 293

HIV – Human immunodeficiency virus

HTS- High throughput screening

HR – Homologous recombination

HSC – Hematopoietic stem cell

HSPC – Haematopoietic stem and progenitor cell

IL - Interleukin

IL2RG – Interleukin-2 receptor gamma

iPS – induced pluripotent stem

ITR – Inverted terminal repeat

KO – Knock out

Ku70 – ATP-dependent DNA helicase 2 subunit 1

Ku80 – ATP-dependent DNA helicase 2 subunit 2

LTR – Long terminal repeat

MHC- Major histocompatibility complex

MRN – Complex containing MRE11, Rad50, and NBS1

MSD – membrane spanning domain

NADPH – Nicotinamide adenine dinucleotide phosphate

NBD – Nucleotide binding domain

NHEJ – Non-homologous end joining

NLS – Nuclear localisation sequence

NPD – Nasal potential difference

P. aeruginosa – Pseudomonas aeruginosa

PAM – Protospacer adjacent motif

P2Y2-R – Purinoceptor 2

PCL – Periciliary layer

PD- Potential difference

PCR – Polymerase chain reaction

PFU – Plaque forming units

PIV3 – Parainfluenza virus 3

PKA – Protein kinase A

PKC – Protein kinase C

PolX – DNA polymerase X

R – Regulatory

rAAV – Recombinant Adeno-associated virus

Rad1 – Radiosensitive mutant 1

Rad10 – Radiosensitive mutant 10

RAD50 – Rad50 homolog

RAD51 – Rad51 homolog

RAD54 – Rad54 homolog

RFLP - restriction fragment length polymorphism

RPA – Replication protein A

RT – Reverse transcriptase

RSV – Respiratory syncytial virus

RVD – repeat variable domain

Sa –Staphylococcus aureus

S. cerevisiae – Saccharomyces cerevisiae

SCID – Severe combined immune deficiency

SDSA – Synthesis-dependent strand annealing

SNP – Single nucleotide polymorphism

SSA – Single strand annealing

ssDNA- single stranded DNA

SV40 – Simian virus 40

TALEN – Transcription activator-like effectors nuclease

TFIIIA – Transcription factor IIIA

TMD – Transmembrane domain

trancrRNA - transactivating crRNA

Ubx - Ultrabithorax

WT – Wild type

X. laevis – *Xenopus laevis*

XRCC4 – X-ray repair cross-complementing 4

ZF – Zinc finger

ZFP – Zinc finger protein

ZFN – Zinc Finger Nuclease

1 Introduction

1.1 Cystic Fibrosis

1.1.1 *A brief history*

"Woe to the child which when kissed on the forehead tastes salty. He is bewitched and soon must die". This proverb from early English folklore is the first reference to the disease known today as Cystic Fibrosis (CF). The disease was first recognised in modern times by Dr. Dorothy H. Anderson in 1939, who from studying the autopsies and case histories of children, described the symptoms and organ changes that occurred (Anderson, 1938). She noted that most destruction occurred in the pancreas and so called it "cystic fibrosis of the pancreas". During the 1940s physicians were becoming more aware of the clogging of ductal systems, such as the pancreas, with thick secretions which would impair the body's ability to break down food and absorb nutrients. They also became more aware of the impaired breathing of CF patients due to these mucus secretions and inflammation caused by bacterial infections living in the lungs. By 1946 researchers had deduced from studying the pattern of disease inheritance in families that CF was a recessive disease whereby two copies of the mutant gene were needed to cause the disease. Then during a heat wave in New York in 1949, hospitals saw a huge increase in the treatment of children with CF for dehydration compared to other children. Paul di Saint'Agnese and colleagues (including Dr. Dorothy Anderson) noticed that children with CF lost an excessive amount of salt in sweat. Even though the reason for this increase in saltiness would not be discovered for many years, the observation was of great clinical benefit and resulted in the development of the first method of CF diagnosis: measurement of Cl^- in sweat which remains today the cornerstone of diagnosis (see section 1.1.6). During the 1970s scientists began to realise that a malfunction in the epithelial cells was at the heart of every tissue in CF patients. In 1973 Paul M. Quinton found that the CF epithelia lining the sweat glands failed to take up chloride ions efficiently which led to the salty skin of patients. This was the first indication that it was a Cl^- flux problem; however, the underlying cause was still not known (Quinton and Philpott, 1973).

1.1.2 *Cloning of the gene*

The 1980's was a time when researchers were trying to map the genetic location of the CF-causing gene. A number of novel molecular techniques were used such as restriction fragment length polymorphism (RFLP) analysis. This tool had been used to map the β-globin gene and led to predictions of variants which caused sickle cell anaemia (Kan & Dozy, 1978). In 1985, the CF locus was mapped using this technique to human chromosome 7q31 (White et al., 1985). This chromosomal mapping provided a guide for locating the gene and the methods of chromosomal walking and jumping, and complementary DNA (cDNA) hybridisation were used to isolate DNA sequences from the CF region to create a physical map of the gene, a sequence of 250,000 base pairs (bp) (Rommens et al., 1989). cDNA clones from the sequence were used to probe a cDNA library created from colon carcinoma cell lines, CF sweat gland cells and adult lung tissue leading to the isolation of 18 further cDNA clones and prediction of a 1480 amino acid protein (Riordan et al., 1989). This group went on to name the gene the cystic fibrosis transmembrane conductance regulator (CFTR). Finally the cause of CF was known to be linked to mutations in this protein. The first mutation to be identified was the ΔF508 class II mutation (Kerem et al, 1989), (Riordan et al., 1989), (Rommens et al., 1989). In 1990, Drumm et al. used retroviruses to transduce a functional CFTR cDNA into a CF cell line. Agents that increase intracellular cAMP stimulated ^{125}I efflux in clones exposed to CFTR expressing virus but not in those transduced with control. Patch-clamping analysis of whole cell clones showed that the anion efflux responses were due to cAMP stimulation of Cl^- conductance (Drumm et al., 1990). That same year Rich et al. showed that delivery of wild-type (WT) CFTR cDNA but not mutant corrected the channel defect, demonstrating a causal relationship between mutations in the CFTR gene and defective Cl^- transport was the cause of the disease (Rich et al., 1990). With the developments in gene discovery and the indication that cDNA addition could resolve the ion flux defect in CF cell lines, people were excited that possible gene-based treatments were not far away. However, as discussed below, the path to CF-gene based treatments has proved difficult to navigate and has still not succeeded.

1.1.3 *Cystic Fibrosis Transmembrane Regulator*

The CFTR gene comprises 180,000 base pairs, is located on the long arm of chromosome 7 and encodes a 1,480 amino acid protein. Wild type CFTR glycoprotein is trafficked to the apical membrane of airway epithelial cells to allow the transport of chloride out of the cell. It is a member of the ATP-binding cassette (ABC) transporter family and consists of two homologous cytoplasmic nucleotide-binding domains (NBD) linked to two membrane spanning domains (MSD) and separated by a regulatory (R) domain. CFTR is unique within the ABC family in that it functions as an ion channel and has a distinctive R domain and it is the phosphorylation of this domain by cAMP-dependent PKA which is a prerequisite for channel gating by ATP, see figure 1.1. Phosphorylation of the R domain alters its conformation and exposes an ATP hydrolysis site on the first NBD (NBD1). Hydrolysis of ATP at NBD1 opens the Cl^- pore and the NBD2 catalytic site, allowing dissociation of ADP from NBD2 and permitting binding of ATP. This binding closes NBD1 catalytic site along with its hydrolysis product ADP, maintaining the open channel state. Hydrolysis of ATP at NBD2 opens NBD1 allowing release of ADP from NBD1, which then closes the channel to complete the cycle. Cl^- transport is essential in maintaining ion and fluid homeostasis in airway epithelial cells, and absence of CFTR at the apical membrane leads to dehydration of the airway surface liquid and accumulation of macromolecular secretions.

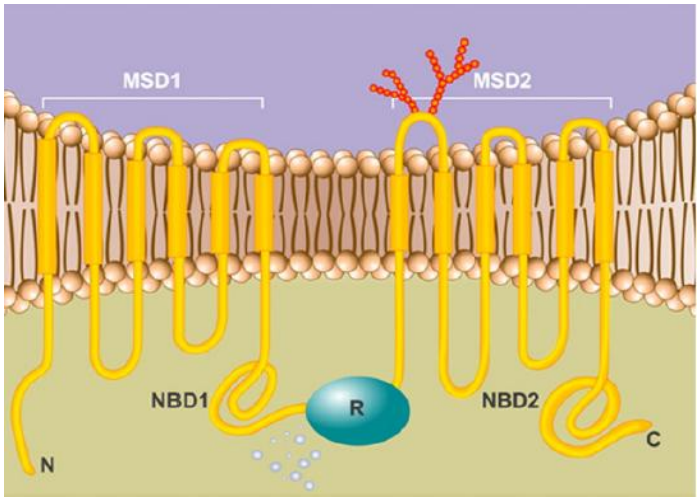


Figure 1.1: Predicted topology and protein structure of CFTR, (Lubamba et al., 2012).

Before the CFTR protein can reach the apical membrane a number of cellular processes must occur. Firstly the DNA is transcribed and spliced in the nucleus to

mRNA. This mRNA then moves from the nucleus to the ribosomes and then to the endoplasmic reticulum (ER) where the nascent protein is translated. Further protein maturation occurs in the ER lipid bilayer leading to complex folding of the CFTR protein (see lane 1, figure 1.4). Final CFTR processing occurs in the Golgi with a glycosylation step and involves conversion of mannose-enriched side chain to a mature complex oligosaccharide attached at the asparagine residues in the fourth extracellular loop within MSD2, (see figure 1.1). Western blot analysis is a common technique used to determine the maturation of CFTR using specific antibodies and by measuring the distance of migration of specific bands one can deduce the amount of CFTR that achieves this final step. Band C at 180-kDa represents a fully mature CFTR protein while the lower band B is representative of the less complex core-glycosylated form that has not reached the golgi. Figure 1.2 is an example of a western blot showing both mature and immature versions of the protein. This process of protein folding has been found to be highly inefficient with mis-folding occurring in more than one-half of wild-type CFTR, (Ward & Kopito, 1994). Despite the low level of abundance and low expression levels in wild-type CFTR expression systems, the protein is fully active functionally and maintains the desired state of ion and fluid homeostasis. These early observations raised the possibility that even correction of a small number of cells could lead to wild-type-like lung function.

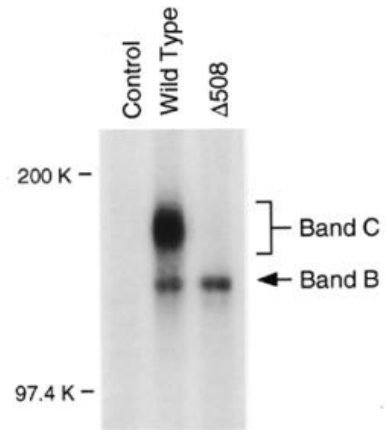


Figure 1.2: Western blot showing the mature band C for fully glycosylated CFTR and immature band B for the immature CFTR (O'Riordan et al., 2000).

1.1.4 *Airway Surface Liquid (ASL) regulation*

The apical surface of airway epithelium is lined with a thin fluid coating called the airway surface liquid (ASL). This is a two layer system that interfaces with the beating cilia on the epithelial surface. The upper mucus layer is composed of the secreted high molecular weight mucins, MUC5AC and MUC5B and functions primarily to trap inhaled particles. Typically this layer is composed of 1% mucins, 1% salt and 98% water, and varies from 7-70 μm in height (Sims & Horne, 2013), with 7 μm being the height of the cilia (Harvey et al., 2011). The lower phase, or periciliary layer (PCL) is a polyionic water layer that permits interactions between the hydration status of the two layers (Tarran et al., 2001) and keeps the mucus layer at a sufficient distance from the underlying airway epithelium to optimise mucus clearance. Normal airway epithelia employ the epithelial sodium channel (ENaC), CFTR and calcium-activated chloride channels (CaCC) for the maintenance of PCL volume. The hydration/volume of the normal airway surface is maintained in the highly water permeable airway epithelia by active transport processes that control the mass of salt (NaCl) on airway surfaces, with water following by osmosis (Matsui et al., 2000). Regulation of the balance between absorption and secretion determines the net transport of ions across the epithelium and hence the mass of salt on the epithelial surface. CFTR has been shown to have inhibitory effects on the ENaC channel (Rubenstein et al., 2011) and the lack of CFTR in CF patients is thought to cause ENaC hyperactivity leading to increased Na^+ absorption followed by increased water absorption out of the ASL and into the cell by osmosis which causes dehydration of the ASL (figure 1.3). There is some data demonstrating that re-hydration of ASL might be achieved by either inhibiting ENaC (Althaus, 2013) or by up-regulating calcium-activated chloride channels, however there has been no clinical benefit to date and so much attention is still directed towards repairing the CFTR defect.

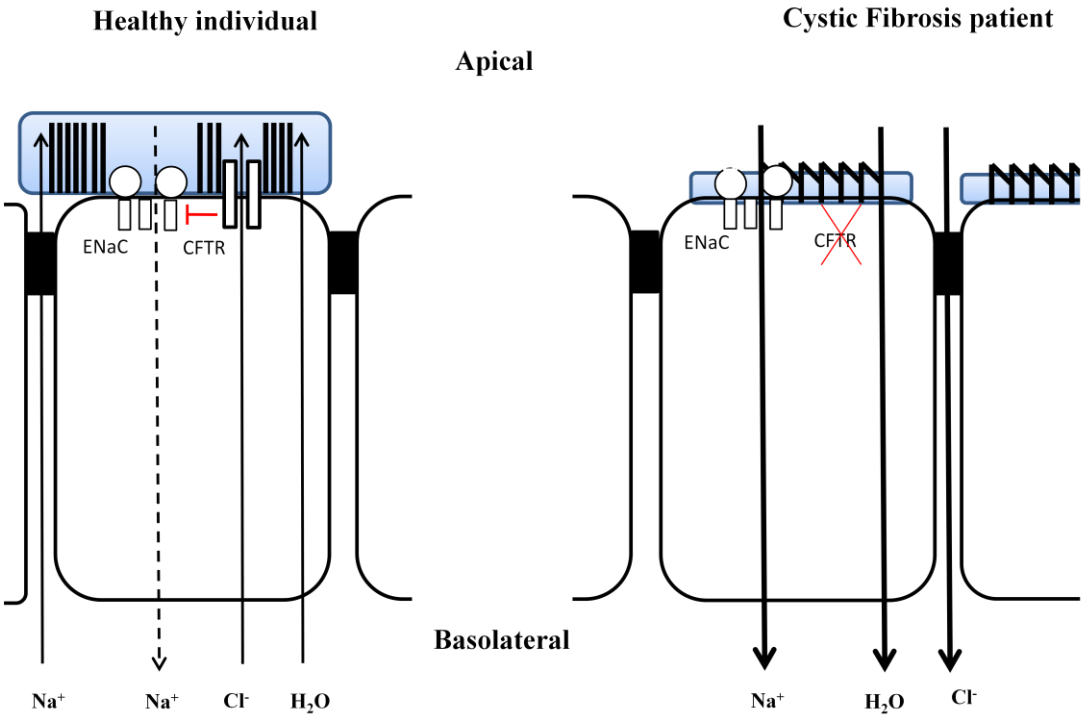


Figure 1.3: Representation of normal and CF airway surface epithelia. In normal cells, Cl^- efflux to the ASL regulates the total Na^+ influx resulting in water transport and mucus clearance. However, in CF epithelia which don't contain a functional CFTR channel, Na^+ hyperabsorption leads to abnormal water transport, resulting in dehydrated ASL and abolition of mucus transport. Thickness of lines indicates the amount of ions/molecules transported across the membrane.

1.1.5 *CF today*

CFTR is expressed in all secretary epithelia. Thus, in CF patients where both copies of the gene are defective, multiple organs including the respiratory, gut, pancreas and reproductive systems are affected. As the old folklore saying suggests, children affected with the disease would die in infancy but today advances in therapy over the past few decades enable CF patients to survive into their late twenties and beyond. Today pancreatic failure is rarely life-threatening due to replacement of digestive enzymes with capsules (Heubi, 2007) however, disease of the lung is still the main cause of fatality where defective CFTR leads to dehydration of the airway surface liquid, leading to lung disease characterised by mucus plugging, infections and sustained inflammation. Treatments for the lung include; daily sessions of physiotherapy to allow postural drainage, DNase inhalation treatment to break up mucus (Shak et al., 1990) and antibiotics to treat the repeated bacterial infections. It is these repeat cycles of infection and inflammation which cause irreparable damage to the lung tissue which eventually leads to patient death (see section 1.1.7) unless a

lung transplant (see section 1.3.4) is possible. But none of these treatments address the genetic cause of the disease.

1.1.6 *Diagnosis*

Measurement of sweat electrolyte levels is the optimal diagnostic test for CF. Sodium and chloride concentrations are raised in patients with the disease, >60 mmol/L is diagnostic, 40-60 mmol/L intermediate and <40 mml/l is normal. Another test measures the nasal potential difference (NPD) to assess altered salt transport. As of the 1st of July 2011, newborn screening for CF is offered in Ireland using the Guthrie blood spot test, www.hse.ie/go/newbornscreening. It screens initially for raised concentrations of immunoreactive trypsinogen and positive results are tested for CFTR mutations. Progress in the understanding of the disease and better management of the condition means the predicted median for survival for babies born with CF in the 21st century is now more than 50 years (Dodge et al., 2007).

1.1.7 *Bacterial Infections and Inflammation*

As a consequence of changes in the ion and fluid homeostasis in the lung of CF patients, changes in airway microbiology also occur with age. In healthy respiratory systems, the upper respiratory tract is colonised by a wide variety of microorganisms comprising the normal flora while the lower tract is kept sterile by various innate host defence mechanisms. These include physical barriers and enodocytic/phagocytic mechanisms.

In CF patients these defences fail and result in an increased susceptibility to infection from a very early age. Initially, *Staphylococcus aureus* (Sa) in the first 6 months, *Haemophilus influenza* peaks in prevalence at 7-18 months then *Pseudomonas aeruginosa* (Pa) is usually detected and is chronic for the life-time of the patient. Over 30% of patients at 3 years of age and 80% of young CF adults are chronically infected with Pa. The long term immune responses to these chronic infections progressively destroy the lung and ultimately lead to respiratory failure. The chronic inflammation is thought to be a consequence of the mutations in CFTR. A number of theories have been put forward as to why the lungs of CF patients become so overwhelmed with infection and inflammation.

- "High Salt" hypothesis

One hypothesis is that the abnormal salt composition of the ASL is a host factor predisposing CF patients to colonisation with Pa. In 1996, Smith et al studied the effects of salt concentration in the ASL on bacterial killing and found that the higher salt concentration in CF compared to normal reduced the bactericidal activity of antibacterial peptides in ASL and thus allowed colonisation of Pa (Smith et al., 1996).

A recent study on CFTR knock-out pigs found no changes in electrolyte composition of the ASL between KO and WT pigs; however, they did detect altered pH of the ASL suggesting a role of CFTR in HCO_3^- secretion. They hypothesise that changes in pH composition of the ASL impair bacterial killing in the CF lung (Pezzulo et al., 2012). Inhaled bicarbonate clinical trial for CF patients has been completed but no results were posted (ClinicalTrials.gov: NCT00177645).

-"Low volume" hypothesis

The high salt theory was disputed earlier on by Matsui et al., (1998) who found no difference in ASL salt concentration between CF and normal airways. They suggested that it was the volume of the ASL that mattered and the low ASL in CF patients caused by the enhanced airway liquid absorption leads to impaired mucociliary clearance and increased bacterial growth.

The mucociliary action is the most prominent host innate defence mechanism of the airway epithelium (Clarke & Pavia, 1980). Clearance relies on two features; 1) the ciliated apical surface and 2) a mucus layer that lines the airway lumen. The cilia beat synchronously, creating a steady current that continually moves the mucus layer upward toward the nasopharynx. The mucus layer is biphasic, consisting of an upper, viscous layer that can trap particles, pathogens and debris and a lower, more fluid layer in which the cilia beat. Normally, this clearance system traps microorganisms in the mucus and carries them to the nasopharynx where they are exacerbated and swallowed. In CF, defective CFTR decreases the height of the ASL in this airway lumen, meaning cilia are unable to beat as normal and the upper mucus layer becomes thick and stagnant, creating an attractive environment for pathogens.

These theories from the 1990s offered an explanation as to why the lungs are susceptible to infection but were not able to explain why the predominant pathogen detected in CF lungs is Pa.

-Hypoxic mucus and Pa response

A more recent theory is able to link mucociliary clearance, hypoxic mucus and the establishment of Pa in the CF lung. Worlitzsch et al. (2002) found that a steep oxygen gradient is present in the mucus lining the airways of CF patients caused by impaired mucociliary clearance leading to stagnant mucus and increased oxygen consumption by airway epithelia resulting in ENaC over activity. They found that the opportunistic Pa could penetrate this thick mucus layer and survive in the hypoxic environment. They also found that Pa form a biofilm under these conditions, leading to speculation that this might be a bacterial stress response. The extracellular polysaccharide matrix of which the biofilms are encased in protect bacteria from antibiotics which cannot penetrate this layer (Mah & O' Toole, 2001).

-Inflammatory response

It has been demonstrated that the inflammatory response is exaggerated in CF patients and that this inflammation starts early in infancy and can even be detected without any sign of infection (Balough et al., 1995). CF pulmonary inflammation is characterised by increased neutrophils, elevated interleukin (Il) -8 concentration and neutrophil elastase (McElvaney et al., 1992). Neutrophils provide the first line of defence by killing and digesting phagocytosed bacteria, however, in CF, neutrophils fail to eradicate infections in the lung. Up until recently it was neutrophil dysfunction in CF airways which was linked to neutrophil necrosis and release of proteolytic enzymes which had a role in destroying the tissue of the lung by overwhelming the antiprotease defences in the lung (McElvaney et al., 1991). However, now there is a shift towards emphasis on defective immune cell function and in particular, dysregulated neutrophil activity. Research is accumulating indicating functional and signalling changes in CF neutrophils that could impact upon clinical prognosis and lung disease severity. Micro-array analysis demonstrated that CF neutrophils display a distinct gene expression profile when compared to healthy controls, with up-regulation of genes encoding for chemokines

CCL17 and CCL18, interleukin receptors IL-3,-8,-10 and -12 and colony stimulating factors (Adib-Conquy et al., 2008). In this study, a comparison was made between blood and airway neutrophils with limited difference in gene expression observed and results suggest CF neutrophils have a perturbed inflammatory profile.

Measurement of IL-6, IL-8 and IL-10 levels in normal and CF isolated HBE cells by Bonfield et al., (1999), revealed that normal individuals secreted IL-10 but no IL-8/6 whereas cells from CF patients had no detectable levels of IL-10 but did produce IL8/6. As IL-10 inhibits the production of anti-inflammatory cytokines, this non-existent to low level resulted in an imbalance in inflammatory and pro-inflammatory cytokines and led to excessive and persistent inflammation in the CF airways. DiMango et al. (1995) had shown previously that gene products of Pa stimulated respiratory epithelial cells to secrete IL-8, and that CF cell lines produced four times as much compared to control cell lines. If CF patients are predisposed to increased inflammatory responses from birth, before infection then neonatal screening is of upmost importance, as the earlier the detection the earlier intervention can begin.

1.2 CF Mutations

1.2.1 *Mutations and classification*

More than 1,900 different mutations in CFTR have been identified and can be divided into six main classes based on their effects on CFTR production, trafficking, stability and the amount of residual CFTR function (see table 1.1 and figure 1.4). This classification of different genotypes has been shown to be linked to disease severity (Hubert et al., 1996). The most common mutation, ΔF508, is the deletion of phenylalanine at codon 508 which accounts for ~70% of all CF alleles. This belongs to the class II mutation group whereby the CFTR protein is transcribed but is unstable and is degraded before it reaches the apical membrane. There are four specific mutations besides ΔF508 that reach a frequency of 1% to 3%: G551D (Glycine->Aspartic), W1282X (TGG->TGA), G542X (GGA->TGA), and N1303K (Asparagine->Lysine). Only another 20 specific mutations reach a threshold

frequency >0.1% (Cystic Fibrosis Genetic Analysis Consortium database, www.genet.sickkids.on.ca/cftr). The prominence of CF in Europe (carrier frequency of ~2%) is thought to be a result of the tuberculosis (TB) pandemic at the beginning of the 17th century as CF patients and carriers have some resistance to TB (Poolman & Galvani, 2007). This pandemic allowed CF mutations to establish themselves and to reach their current levels in Europe.

Mutation Class	Description	Disease Severity
Class I	No protein is produced. The most common types of mutations in the group are nonsense mutations and frameshift mutations which result in premature stop codons, e.g. G542X	Severe
Class II	Defective protein processing. Full length CFTR is produced but cannot exit the endoplasmic reticulum (ER). The most common mutation ΔF508 falls in this group.	Severe
Class III	Block in CFTR regulation. The CFTR protein is produced and delivered to the apical membrane, However, mutations such as G55ID prevent ATP binding and channel gating. The corrector VX-770 is used to treat this mutation.	Severe
Class IV	Altered conductance. The CFTR is correctly inserted into the apical membrane but mutations prevent the channel from fully opening or from closing, e.g. R117H	Mild
Class V	Reduced synthesis. Mutations in the promoter region or ribosomal binding site result in a reduction of protein synthesis. However, any CFTR protein produced is fully functional, e.g. 2789 + 5G->A	Mild
Class VI	Functional but unstable with decreased half life at the cell surface, e.g. Gln1412X	Severe

Table 1.1: List of CFTR mutation classes.

1.2.2 *Mutations and disease severity*

As outlined in table 1.1, mutations in the gene encoding the CFTR protein lead to disruption of the cellular processes mentioned in section 1.1.3 which are necessary for production of functional CFTR. Depending on what stage of production the protein reaches determines the severity of the disease. For example in class I no transcript is made due to nonsense mutations meaning no protein at all is produced, in class II the protein is incorrectly folded leading to degradation before it reaches the membrane and in class III the protein is made and correctly trafficked to the apical membrane but is unable to open to allow ion efflux making these three classes the most severe. The least severe classes are IV and V, in which the channel is functional but at a lower efficiency due to either altered conductance or reduced synthesis. Finally class VI mutations are the rarest and result in unstable proteins and are considered severe. Treatments that can restore function of the protein are needed and much work has been done for classes III, IV and V whereby the protein is at the surface but not functional. But for the more severe class, whereby no protein at all is made, a gene based treatment such as cDNA addition or gene repair would appear the only option.

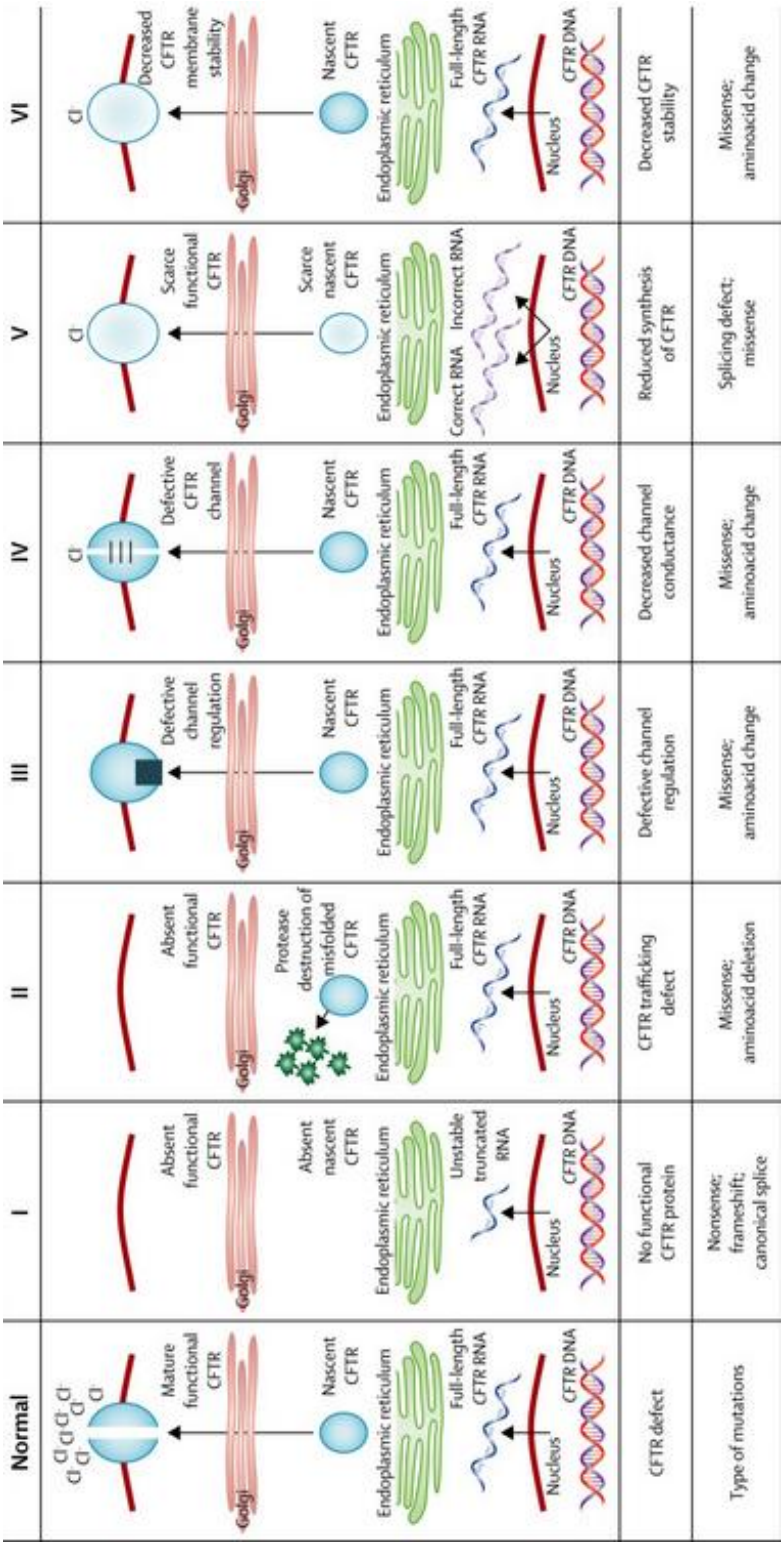


Figure 1.4: Schematic representation of the six different CFTR mutation classes, taken from (Boyle & De Boeck, 2013). Normal: CFTR reached the apical membrane and allows Cl^- flux. Class I: A full length transcript is disrupted by premature stop codons, no protein is made. Class II: Protein is transcribed but is mis-folded and becomes trapped in the ER leading to degradation, meaning no protein reaches apical surface. Class III: Protein reaches the apical membrane but the channel is defective, allowing no Cl^- flux. Class IV: Channel is present at the apical membrane but has defective conductance allowing partial Cl^- movement. Class V: Channel is at the apical membrane but at low abundance leading to reduced total Cl^- flux. Class VI: Channel at apical membrane but is unstable leading to little Cl^- movement.

1.2.3 *ΔF508*

The missense class II mutation ΔF508 is by far the most common CF-causing mutation, affecting 70% of all CF patients. It is caused by the deletion of a phenylalanine at position 508 of the CFTR protein which leads to misfolding and degradation before it can reach the membrane (Kerem et al., 1989). F508 resides in the first nucleotide binding domain (NBD) of CFTR near a predicted interface with the fourth intercellular loop-ICL4. Identification and assessment of positions statistically coupled to F508 using evolutionary sequence analysis revealed that both NBD1 folding and interaction with ICL4 are altered by the mutation (Mendoza et al., 2012). It was determined that correction of both defects is needed to restore CFTR function. Emerging therapies which can re-route the misfolded protein to the plasma membrane are of great interest as it has been shown that once the ΔF508 CFTR protein is at the surface it exhibits suboptimal gating and activation relative to wild-type (Denning et al., 1992). These types of therapies are called 'correctors'. Once at the membrane, 'potentiators' increase opening probability of the protein and increase function, making them of use to mutations like G55ID, a class III mutation.

1.3 Mutation Specific Therapies

1.3.1 *Aminoglycosides for Premature Stop Codons*

Nonsense mutations occur when a single base change produces an in-frame stop codon producing a highly unstable transcript and no protein synthesis. For the ~10% of patients with the nonsense mutations of class I, a possible treatment would be to promote read-through of the premature stop codons to enable translation of the entire sequence and synthesis of full length CFTR protein. It was shown that aminoglycoside antibiotics can interact with the small ribosomal RNA subunit-18S and alter normal ribosomal proofreading during protein translation in eukaryotic cells, (Purohit & Stern, 1994). Sometimes this interaction inserts an incorrect amino acid at the premature stop codon and allows production of a full length protein, albeit with a single incorrect amino acid, in other words it can suppress the premature stop codons. Two of the first aminoglycosides to be tested to read through such a mutation were G-418 and Gentamicin, a clinically approved antimicrobial. They exhibited suppressive activity in CF cell models with stop

mutations, and restored expression of functional CFTR, (Howard et al., 1996). Although a 14-day trial involving nine CF patients in 2000 provided promising results indicating that gentamicin applied directly to the nasal mucosa of CF patients with class I mutation could improve CFTR Cl^- transport, (Wilschanski et al. 2000) a trial in 2007 on a larger number of patients was unable to demonstrate significant alteration in nasal PD following 28 days of nasally administered gentamicin, (Clancy et al., 2007).

Using high-throughput screening (HTS) technology a novel compound PTC124 (Ataluren) has been identified, a non-aminoglycoside, which is able to read through premature stop codons but not normal stop codons and has no evidence of renal toxicity. PTC124 acts by inactivating the nonsense-mediated mRNA decay pathway which usually degrades truncated proteins and so by disrupting this pathway, nonsense containing transcripts are stabilised (Welch et al., 2007). It has been shown to be superior to gentamicin and well tolerated in mouse studies with another premature stop codon disease, Duchenne muscular dystrophy, where it induced production of full length proteins, (Welch et al., 2007). In a CF mouse model, expressing the human CFTR-G542X transgene, oral administration of PTC124 suppressed the G542X nonsense mutation and restored a significant amount of CFTR protein at the apical surface as seen by immunofluorescence staining and also CFTR function saw a restoration of 24-29 % of the average cAMP-stimulated Cl^- currents observed in WT following treatment, (Du et al., 2008). Phase II trials with PTC124 determined the drug to be well tolerated with mild side effects. However, the 2012 phase III trial of ataluren in 238 patients with cystic fibrosis with stop mutations did not reach its primary endpoint of improvement in forced expiratory volume in 1 s (FEV1) at 48 weeks, except in a subset of individuals who were not simultaneously taking nebulised aminoglycoside antibiotic (ECFS meeting, 2013). Total absence of CFTR protein in class I mutations make it the most challenging defect to address and a gene therapy approach is most likely to benefit this class the most.

1.3.2 *Correctors*

The class II mutation ΔF508 can be rescued to the cell surface in cell cultures by reducing temperature to 27 °C, (Denning et al., 1992). However, this technique is

not practical for patients. Correctors are compounds that can enhance the productive trafficking of these mutated CFTR proteins. Using the HTS technique compounds were identified which acted as potent correctors. One such compound was VX-809. In pre-clinical models of CF, western blot analysis of HBE cell lysates following treatment showed an increase in the amount of band C correlating to an increase in CFTR maturation to the apical membrane. Cl^- secretions were also increased to ~15% of that seen in WT HBE cells (Van Goor et al., 2011). The compound proceeded to phase II clinical trials in adults with CF homozygous for ΔF508 and was well tolerated when taken once over 28 days, with adverse effects similar to placebo. It reduced sweat chloride levels in a dose-dependent manner but there was no significant improvement in CFTR function in the nasal epithelium as measured by PD changes nor were there significant changes in lung function or patient report outcomes. No maturation of immature ΔF508 was detected in rectal biopsies, (Clancy et al., 2007). Despite these disappointing phase II results, VX-809 is entering into phase III clinical trials to be used as a combination therapy with the potentiator VX-770 (see 1.3.3) as it was seen in ΔF508 HBE cells, the level of VX-809 corrected ΔF508 could be doubled by addition of VX-770, (Van Goor et al., 2011). Another corrector entering phase III combination therapy trials with VX-770 is the compound VX-661. Both trials are still recruiting volunteers and so results are pending.

1.3.3 ***Potentiators***

In January 2012, a huge advance was made for people suffering from the class III gating mutation, G551D, which affects 4-5% of people with CF. The potentiator, VX-770 was approved by the FDA and was available for treatment. This small molecule acts by opening the CFTR channel that is already present at the apical membrane. It was identified following HTS of 280,000 chemically diverse drug-like and lead-like compounds using a cell based fluorescence membrane potential assay designed specifically to identify CFTR potentiators, (Van Goor et al., 2009).

They found that VX-770 increased CFTR open probability in recombinant cell lines with either ΔF508-CFTR (after biosynthetic rescue at 27ºC) or G551D-CFTR mutations. All experimental cell lines/cultures were pre-treated with forskolin to activate cAMP prior to VX-770 treatment as it is not an activator. Assessment of Cl^-

secretion in primary HBE cultures with one ΔF508 allele and one G511D allele saw an increase to ~50% of that observed in HBE non-CF cultures. Loss of CFTR function causes an increase in ENaC channel activity leading to an increase in Na^+ absorption and by osmotic gradient an increase in water absorption resulting in a dehydrated airway surface liquid. When they assessed the Na^+ absorption in VX-770 treated G551D/ΔF508 HBE and WT HBE cultures by measuring potential difference (PD), they saw a decrease in the PD compared to baseline. Likewise, measurement of fluid levels as determined by airway surface liquid volume revealed an increase to about half of that observed in WT-HBE. An increase in cilia beating was also observed. The effects on ΔF508-CFTR cells were only seen following rescue of the channel indicating that in order for the potentiator to work the channel needs to be at the cell surface. Also the channel needs to be activated by endogenous cAMP/PKA signalling pathways as seen when no positive CFTR function affects were seen when cells weren't first stimulated with forskolin.

1.3.4 *Lung tissue engineering*

Lung transplantation is the only definitive treatment for end-stage CF lung disease, but as limited supply of donor lungs and survival statistics for lung transplants are lower than other organ transplants, this makes it a last resort (Nichols et al., 2012). However, if successful it can be highly effective and the new lungs will be mutant free meaning normal ion and water transport will occur in the lungs. Other tissues will still be affected by CF but these effects can be well managed.

Groups have begun to use tissue-engineering strategies to develop autologous bio-artificial lungs that could potentially overcome the need for donors. Peterson et al., (2010) first generated a whole-lung scaffold by perfusion and de-cellularisation of adult rat lung. They then reseeded the endothelial and epithelial surfaces of the scaffold with new cells. Following implantation of the new lung into the rat for short time intervals, gas exchange occurred in the engineered lungs (Petersen et al., 2010). By correcting CF patients own cells by gene repair *ex vivo* and then reseeding these cells onto the scaffold, a CF-free lung could be made available that could be transplanted back into the patient and would not be rejected. A similar strategy has been used previously to treat patients with severe combined immuno-deficiency (SCID) disease. Bone marrow cells were corrected *ex vivo* and then re-

introduced into the patient leading to full correction of the disease phenotype (Cavazzana-Calvo, et al., 2000). There is plenty of work that needs to be done before tissue engineering can be used in a clinical setting but if combined with gene repair therapies the need for donor lungs for CF patients could be made redundant.

1.4 Gene Therapy for CF

The recent drug discoveries mentioned before show great promise for CF patients who suffer from the relevant mutations, however, there are still 5-10% who suffer from mutations where no protein at all is produced, who benefit from no available treatment apart from traditional methods. For these patients other forms of treatment are needed. Gene therapy offers an alternative to these drug based methods. Furthermore, if one could repair the genetic mutation that gives rise to the disease, for example correcting the three base pair deletion in ΔF508, then the corrected DNA would give rise to functionally active protein for the life-time of the cell.

1.4.1 *CF: A good and bad candidate for gene therapy*

A number of factors make CF an attractive candidate for gene therapy. It is a monogenic disease and so only requires correction of one gene. In addition, the levels of full-length CFTR mRNA vary considerably in phenotypically normal individuals suggesting that a high level of transgene expression is not required to restore normal phenotype. There is some variance in studies reporting the amount of transcript needed to restore function. Chu et al., (1991) found when analysing the CFTR mRNA transcripts of freshly isolated HBEs from 12 normal adults that all had some CFTR mRNA transcripts with exon 9 completely deleted. They concluded that individuals can have up to 66% of bronchial CFTR mRNA transcripts that lack exon 9, a region representing 21% of the sequence coding for the NBD1 region of the protein, and show no symptoms of CF disease (Chu et al., 1991). In a further study, they deduced that only between 8-16% bronchial epithelial cells need to express normal exon 9 CFTR mRNA to maintain normal phenotype as deduced from the observation that an individual with 92% of transcripts lacking exon 9 showed no symptoms of CF (Chu et al., 1992). Another group developed a PCR based method to quantify CFTR transcripts and applied this

to the analysis of nasal epithelium RNA of five patients with CF and the 3272-26A>G/ΔF508 genotype. The 3272-26A>G mutation creates an alternative splice acceptor site in intron 17 that competes with the normal one. The use of this alternative splice site causes a reduction in normal transcript levels and as a consequence normal protein (Beck et al., 1999). They calculated that 8.2 ± 0.1% of the total CFTR RNA was normal full-length CFTR in these five patients. When the reduced number of mutant CFTR transcripts that result from the ΔF508 allele were also taken into account, it was calculated that these patients had only 4.7 ± 0.5% normal CFTR mRNA relative to normal individuals. All five patients had mild CF compared to ΔF508 homozygotes therefore they concluded that 5% of normal CFTR mRNA is enough to ameliorate the severity of the disease (Ramalho et al., 2002). A more recent study looked at the functional effects of correcting CFTR on the levels of ASL. This *in vitro* study involved delivering CFTR cDNA in a parainfluenza virus to a model of the human CF ciliated surface epithelium and measuring the effects on mucus transport. They found that 25% of cells needed to be corrected to restore mucus transport rates to levels comparable to non-CF airways (Zhang et al., 2009).

A lot of studies look at the Cl^- defect but what about the Na^+ abnormalities seen in CF? It is unclear whether both Na^+ and Cl^- abnormalities need to be corrected for clinical benefit. A study addressed this by introducing varying numbers of non-CF cells with CF cells and assessed the resultant degree of correction of both ion abnormalities. It showed that a mixture of ~ 5% of non-CF cells corrected the Cl^- defect completely but nearly 100% of cells needed to be non-CF in order to restore Na^+ transport (Johnson et al., 1995).

The natural defence barriers presented by the lung to foreign particles makes delivery of gene therapy tools difficult. Firstly the mucociliary mechanism mentioned earlier can inhibit gene transfer as CF patients produce more mucus and increase the viscosity and penetrability of this layer. To counter viscous CF sputum, agents such as rhDnase and gelsolin have been used effectively to reduce viscosity and increase transduction efficiency (Stern et al., 1998). The host's innate immune system creates another barrier making repeat dosing with some viral vectors impossible, as recognition of viral coat proteins results in production of neutralising

antibodies (Zabner et al., 1996). Cao et al., demonstrated that transient immunosuppression could significantly enhance the efficiency of transgene expression and facilitated repeat dosing of helper-dependant (HD) adenovirus vectors to mouse lungs (Cao et al., 2011).

1.4.2 *Which cells/ part of the lung to target?*

The optimal cell type to target is unclear and remains a topic of debate. Some think the cells with maximal CFTR expression are in the submucosal glands in bronchial tissues notably in the serous component of the secretary tubules (Engelhardt et al., 1992). This group used *in situ* hybridisation and immunochemistry to characterise the distribution of CFTR in cells. Another group have reported WT CFTR protein was localized to the apical membrane of ciliated cells within the superficial epithelium and gland ducts (Kreda et al., 2005). These conflicting reports make it difficult to ascertain which cells to target for gene therapy. Delivery via nebulisation would hit the surface epithelium but would not reach the submucosal glands. Also treating the surface epithelia, which are terminally differentiated, would mean gene expression would be lost with the death of the cell. The life-span of human epithelia is still unknown, however studies in mouse models have shown that ciliated airway cells have an average half-life of six months in the trachea and 17 months in the lung (Rawlins & Hogan, 2008). Targeting stem cells of the respiratory tract such as basal cells, would be much more difficult to transduce but if successful gene transfer would be passed on to progeny cells and lead to long-term CFTR expression. There is evidence that a number of epithelial cell types in the lung are slow-cycling/progenitor cells that have the capacity to proliferate *in vivo* following lung injury (Liu et al. 2009). Studies of gut epithelia which have a high turnover rate identified the crypt base columnar cell as the intestinal stem cell. As the stem cells proliferate, some become displaced from the niche and commit to differentiation leaving the overall stem cell number approximately conserved. The clone size undergoes a random sequence of expansions and contractions around the crypt base. Sooner or later the entire crypt base stem cell population derives from the last remaining clone. Targeting the stem cells of CF patients with integrating vectors, like AAV, could lead to long term correction.

1.4.3 *Efficacy is difficult to detect*

Measurements of transgene mRNA and protein levels will determine if gene transfer is successful. As CFTR is normally expressed at a low level in healthy lungs, the level of transgene mRNA and protein will be similarly low, and assays used to detect these levels may not be sensitive enough. Also, detection of protein may not always correlate to improved function, as has been seen in some previous clinical trials. Transmembrane potential difference (PD) is used as an end point in nasal trials of CF gene transfer and novel pharmacological agents (Wilschanski et al., 2000). More recently, tracheal and bronchial PD measurement techniques have been developed and used in clinical trials of liposome-mediated CFTR gene transfer, (Alton et al., 1999). Function can also be deduced from more downstream measures, like inflammation. For example, IL-6 /8 and neutrophils are increased in CF lungs, if there was a decrease in these levels following gene transfer one could infer that function has been restored. Forced expiratory volume (FEV_1) is another measure of efficacy outcome. A number of measurements would need to be taken into account in order to achieve a clear picture of whether a gene therapy trial has clinical benefits.

1.4.4 *Animal models to study the disease*

The development of novel CF therapies has been hindered by the lack of animal models that recapitulate the key features of lung and other organ disease pathogenesis. Transgenic mice with CFTR null alleles and a variety of other mutations have been available since the early 1990s but as these do not spontaneously develop a CF lung phenotype their use has been limited. In 2008, two groups used AAV vectors, followed by nuclear transfer and cloning, to develop CF pigs (Rogers et al., 2008) and CF ferrets (Sun et al., 2010). Pigs share many anatomical and physiological features with humans. Newborn pigs lacking CFTR exhibited defective Cl^- transport and developed meconium ileus, exocrine and pancreatic destruction all of which are seen in human newborns with CF (Rogers et al., 2008). Like humans, null pigs' lungs were free of inflammation at birth but manifest a bacterial host defence defect without prior infection. Null ferrets develop multi-organ system disease and neonatal animals manifest a pulmonary host disease defect associated with colonisation of bacteria (Sun et al., 2010). Both of these new

models of CF disease are very attractive for gene therapy studies, however, cost and handling procedures mean it will be several years before these models are available to the wider research community.

1.4.5 *CF gene therapy to date*

-Viral vectors

Since cloning of the CF gene in 1989, twenty-five Phase I/II clinical trials have been carried out using a variety of viral and non-viral gene transfer agents, (Griesenbach & Alton, 2009). In this section some of the findings of these trials are discussed.

The first trial in 1993 used E1-deficient adenovirus (AdV) vectors to deliver CFTR cDNA to the nose of three CF patients. The outcome of treatment was assessed by measuring PD in the nasal epithelium in response to low Cl^- solution. An increase in PD was observed in all three patients suggesting the CF Cl^- defect had been partially corrected. No detection of increased levels of CFTR mRNA or protein was reported. Patients suffered nasal congestion and mild rhinorrhoea but symptoms subsided within 24 hrs of administration of AdV-CFTR and were attributed to the method of delivery (Zabner at al., 1993).

In 1994, the first clinical trial to deliver CFTR to human lung was carried out using AdV (Crystal et al. 1994). In this study four patients, all homozygous for ΔF508, were administered with a replication deficient AdV containing the 4.5 kb of CFTR cDNA. Approximately 5-14% of bronchial cells tested were shown to express AdV derived CFTR after treatment but no increase in lung function (FEV_1) was detected. The level of CFTR positive cells was quite high and may have been a result of bronchial brushings being taken from the site of AdV vector delivery. This study demonstrated that it was safe to deliver AdV expressing CFTR to the lung over a 1,000-fold range from 2×10^6 to 2×10^9 plaque forming units (PFU) with no sign of virus shredding or replication competent virus. However, one patient did suffer adverse effects having received a dose of 2×10^9 which lasted 14 days post administration. The problem was attributed to vector-induced inflammation. No replication-competent or deficient virus was observed in nasal, pharyngeal, rectal,

blood or urine samples. There was a significant increase in IL-6 in one patient, and lower levels of IL-6 detected in the other three suggesting IL-6 levels were dose-dependent and that there was a threshold for symptomatic responses to IL-6.

The following year a group conducted a trial on 12 CF patients, again using AdV, to deliver CFTR cDNA. The nasal epithelium was the site of delivery, with six patients receiving AdV-CFTR and six receiving control. Reverse-transcriptase (RT) PCR detected molecular evidence of gene transfer in 5/6 of the patients treated with AdV-CFTR, however, the level of epithelial cells transduced was very low (<1%) with no restoration of Cl^- conductance or normalisation of Na^+ transport as determined by nasal PD. It is thought the absence of AdV tropism for surface columnar cells which normally express CFTR (a factor that should have been considered before clinical trials were started) may explain this low level. No toxic effect was seen at the lower doses, however, like before there was some mucosal inflammation in 2/3 patients at the highest dose of 2×10^{10} PFU meaning the problem of low gene transfer cannot be overcome by increasing the dose, Although there were no phenotypic changes detected, the safety profile of AdV as a vector was demonstrated (Knowles et al., 1995).

Much attention has focused on the problem of repeat administration of AdV; as AdV expression is transient it would be necessary to repeatedly deliver CFTR cDNA for the lifetime of the patient. Zabner et al. (1996) showed that repeat administration was safe in a group of six patients who received increasing doses of up to 10^{10} infectious units with no detection of adverse effects; however, they did develop additional humoral immune responses. Partial correction of the Cl^- defect was observed in some subjects, but this was decreased in subsequent administrations thought to be because the immune response limited gene transfer.

In a lower airway study (Harvey et al., 1999), Ad-CFTR was delivered at doses ranging from 3×10^6 to 2×10^9 PFU via an endobronchial spray in three doses over a nine month period to 14 CF patients. Bronchoscopic assessment was performed 3 and 30 days post-administration. The data from this study demonstrates that; (i) the strategy appears to be safe; (ii) 3 days after the first administration, vector-derived CFTR cDNA expression was detected in the CF airway epithelium in a dose-related

pattern, with greater than 5% endogenous CFTR mRNA levels at the higher doses; (iii) expression is transient; second administration resulted in some expression that was not dose-related and the third administration produced no expression in any sample; (iv) the lack of expression with repeat dosing does not correlate with induction of systemic anti-Ad neutralising antibodies. Although repeat administration is safe, efficacy is severely compromised. As gene expression was not seen in patients with high pre-existing levels of antibody it is thought that the response may be dictated by these pre-existing titers and modified by route of administration but is not dose dependant (Harvey et al. 1999).

The poor uptake of AdV vectors by apical membrane of airway epithelia may be explained by the lack of AdV receptors on the cell surface. The AdV fibre protein receptors, the Coxsackie B and Adenovirus type 2 and 5 receptor (CAR) and major histocompatibility complex (MHC) class 1 α-2 domain are all located on the basolateral surface of epithelial cells (Pickles et al., 1998). This is reinforced by the studies of Cohen et al., (2001), who observed that when AdV is applied to the basolateral surface of cells or when gap junctions are disrupted allowing access to the basolateral surface, the level of AdV transduction is increased. The inability of rAdV to replicate means it cannot infect the lung as effectively as WT AdV, which is commonly associated with lung infections as only a small number of WT virus particles are needed to initiate infection.

Another virus vector regularly used in CF clinical trials is recombinant Adeno-Associated virus (rAAV). Twenty-five CF patients were treated with AAV-serotype- 2 via intranasal and endobronchial administration in a phase I clinical trial (Flotte et al., 2003). Doses of rAAV-CFTR vector ranged from 3×10^1 to 1×10^9 replication units (RU). One serious adverse event was recorded and judged to be vector related. Vector shredding was minimal and serum-neutralising antibodies were detected after vector delivery of the highest doses. PCR was used to measure gene transfer and was not observed in nasal and bronchial epithelia until a dose of 1×10^7 was administered. No change in nasal PD was reported. In another study by Moss et al., (2007), repeat administration of AAV-CFTR was investigated. In this study 102 CF patients received two doses AAV-CFTR or matching placebo administered 30 days apart via a nebuliser. In contrast to a smaller phase II trial

carried out previously, Moss et al., (2004) reported that there was no difference observed in FEV_1 or IL-8 levels between groups. It did demonstrate the safety of AAV vector re-administration but also suggests the need to improve gene transfer technology to increase clinical efficacy outcome.

These studies have shown;

a) Both AdV and AAV viral vectors are safe delivery vehicles

b) Repeat dosing is necessary as expression is transient

c) Although some functions are improved following first dose, no benefit has been seen as subsequent dosing induces an immune response limiting gene transfer

There have been no new studies using virus vectors for CF patients reported by clinicaltrials.gov since 2005.

-Non-viral vectors

An alternative to using virus vectors for gene transfer is to use non-viral or synthetic vectors. As mentioned above, the repeat administration of viral vectors seems to be limited by the production of antibodies in response to vector delivery or the existence of pre-existing neutralising antibodies. Non-viral vectors may be able to overcome these limitations.

The majority of trials to date have used lipid-based gene transfer agents. The first non-viral clinical trial assessed the safety and efficacy of a single nasal dose delivery of plasmid DNA containing CFTR cDNA under the control of SV40 promoter, complexed with (3β[N-(N',N'-dimethylaminoethane)-carbomoyl] cholesterol:dioleoylphosphatidylethanolamine (DC-Chol/DOPE) liposomes. Fifteen CF patients were tested, with 9 receiving the CFTR cDNA and 6 receiving just liposome. No adverse defects were detected and there was a 20% partial correction of Cl^- defect observed in nasal PD. This was maximal at day three and had reverted to pre-treatment levels by day seven, (Caplen et al., 1995). Alternative formulations were investigated in which plasmid DNA containing the CMV promoter complexed with dioleoyl trimethyl ammonium propane (DOTAP) liposomes and were shown

to have similar effects. In one study in which eight CF patients were treated with pCMV-CFTR-DOTAP, transgene DNA was detected in 7/8 patients up to 7 days following treatment and for up to 28 days in two of these seven patients. Vector derived CFTR mRNA was observed in 2/7 patients at days 3 and 7 post-treatment (Porteous et al., 1997). Although delivery of these complexes was well tolerated any electrophysiological defect correction was small and variable. By studying the gene transfer ability of different commercially available cationic lipids to the lung of mice *in vivo* Lee et al. (1996) found that a novel lipid GL-67 in combination with the neutral co-lipid DOPE, could enhance gene transfer by 1,000 fold as compared with DNA administered alone. Co-formulation of GL67 (GL67A) with the neutral lipid DOPE, (which is thought to facilitate pDNA endosomal escape), along with small amounts of a polyethylene glycol-containing lipid, DMPE-PEG5000, was found to stabilise formulations at concentrations sufficient for aerosol delivery to the lung (Eastman et al., 1997). Delivery was safe and a single nasal dose of GL67A/pDNA produced vector derived mRNA expression and an overall correction of ~20% of the Cl^- defect in the nose (Zabner et al. 1997). A double-blind placebo-controlled trial was performed by Alton et al., (1999) on eight CF patients to assess the safety and efficacy of GL67-CFTR (Alton et al. 1999). The CFTR cDNA was under the control of the CMV promoter and the DNA-lipid was delivered by nebulisation to the lung. Seven of the eight patients receiving the active complex experienced flu-like symptoms that were resolved within 36 hrs. Pulmonary administration resulted in a significant degree of correction of Cl^- abnormality that equated to 25% restitution towards normal values as determined by *in vivo* PD. However, no effect was seen on the increased Na^+ absorption experienced by CF patients. Encouragingly bacterial adherence was decreased following treatment in five out of six patients. No vector specific CFTR mRNA was detected in any patient sample, however vector-derived DNA was detected at levels of 10^6 copies or more per sample in all eight patients at day two eluding to the possibility that RNA integrity was compromised or there was inadequate sample.

To overcome the initial flu-like symptoms experienced in this trial a number of modifications were made. Firstly, the pro-inflammatory CpG motifs were removed, and secondly the viral CMV promoter was exchanged for a humanised one (hybrid

elongation factor-1a) capable in preclinical trials of sustaining prolonged gene expression (Hyde et al., 2008). They showed that even a single CpG motif in the plasmid DNA was sufficient to elicit an inflammatory response in a mouse model. When this newly improved plasmid containing CFTR cDNA (pGM169) was complexed to GL67A, it led to > 4 week expression in murine lung model upon single dosing.

This study has lead to the formation of The UK Cystic Fibrosis Gene Therapy Consortium and major funding to support current trials. One such trial is currently assessing GL67A/pGM169 in an ongoing multi-dose trial at the Royal Brompton Hospital in London and at the Western General Hospital and Royal Hospital for Sick Children in Edinburgh (www.cfgenetherapy.org.uk). An update of the trial was presented at the 37th ECFS meeting held in Lisbon by Dr. Jane Davis. It was reported that the delivery by aerosol of the GL67A/pGM169 complex under the control of the humanised promoter showed enhanced duration when compared to the CMV driven complex. Reduced inflammation was observed due to removal of the CpGs and there was an improvement in Lung Clearance Index (LCI).

As well as DNA-lipid complexes, some groups are using nanoparticles to deliver CFTR cDNA. Compacted DNA nanoparticles were delivered to the nasal epithelium in twelve CF patients (Konstan et al., 2004). Some Cl^- transport was observed in nasal PD and at day 14 vector PCR analysis showed a mean value in patients receiving the highest dose of 8.0 mg or 0.58 copies per cell. Corrections persisted for up to six days following gene transfer.

1.5 Gene addition vs Gene repair

Although millions of dollars have been spent on these studies there are still little or no reports of success to date in these clinical trials. As they are focusing on delivering CFTR cDNA in a transient manner, the limitations of repeat dosing are going to have to be overcome in order for patients to benefit from treatment. The developments of lipid-complexes which do not produce an immune response are promising. Viral vectors which target specific regions and cell types are being studied and may be of clinical benefit. Another obstacle with cDNA addition is that once the corrected cell dies so does the CFTR. If one could permanently correct the

DNA mutation in a stem/progenitor cell, rather than adding cDNA, then the correct DNA sequence would be passed on to daughter cells. Another advantage of such a gene repair strategy would be that the corrected gene would be under the control of the endogenous promoter and so spatiotemporal expression would be maintained unlike cDNA addition where the gene is under the control of an exogenous promoter and so expression levels are typically higher than normal.

The field of genome editing is described in the following section.

1.6 Genome editing

DNA damage is constantly occurring in cells due to both environmental and endogenous insults at a rate of 10^3 - 10^6 lesions per cell per day (Lodish et al., 2004). A double-stranded break (DSB) is the most cytotoxic form of DNA damage. It can occur during the normal course of DNA replication, through exposure to by-products of cellular metabolism like oxygen free radicals (Valko, et al., 2006) and ionising radiation (Jackson & Bartek, 2009). To combat threats posed by DNA damage, cells have evolved with mechanisms to detect DNA lesions, signal their presence and promote their repair (Rouse & Jackson, 2002).The ability of a cell to repair these breaks is critical for the preservation and faithful propagation of genetic information. Deficiencies in DSB repair can lead to mutations and chromosomal rearrangements which may ultimately result in genomic instability and tumourigenesis. *Xeroderma pigmentosum* is an autosomal recessive disease in which the ability of the DNA repair system to repair damage caused by exposure to UV light is deficient and leads to skin cancers.

1.6.1 *DNA repair mechanisms*

There are three major pathways used by cells to repair DSBs; the error-prone non-homologous end joining (NHEJ) pathway (Lieber, 2008), the faithful single strand annealing (SSA) pathway, and homologous recombination (HR) pathway, which can also be referred to as homology directed repair (HDR) (San Filippo, et al., 2008). In eukaryotes, homologous recombination is restricted to the S and G_2 phase of the cell cycle when the sister chromatid is in close proximity (Sonoda et al., 2006) whereas NHEJ can occur throughout the cell cycle because it does not require a homologous chromosome. The three pathways are described in detail below and the regulation of choice is discussed in the next section.

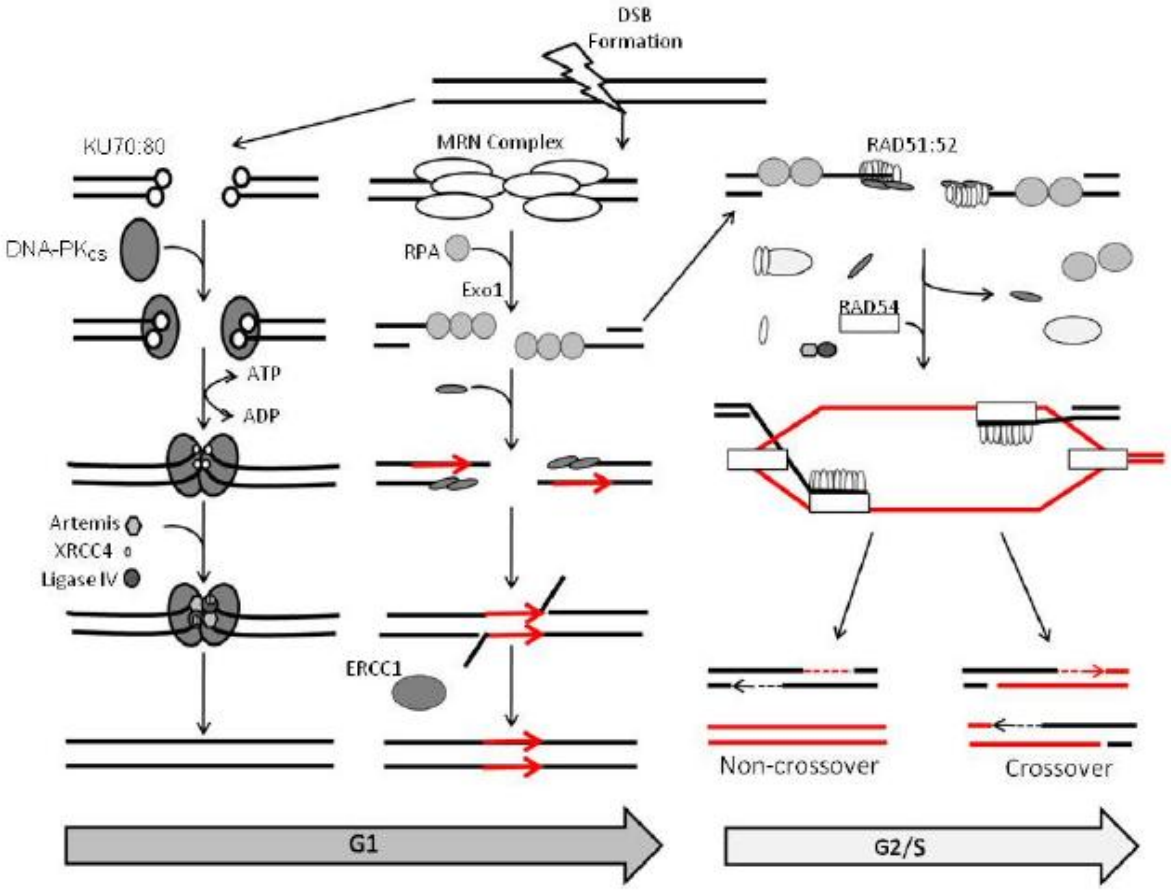


Figure 1.5: DSB repair pathways. The major pathway during the G1 growth phase is NHEJ shown in the left column. Following DSB formation the KU70:80 dimer is recruited to the free DNA ends and recruits DNA-PKcs. DNA-PKcs recruits the nuclease Artemis, XRCC4 and DNA ligase IV. XRCC4 acts as a scaffold to enhance Artemis activity which results in small deletions prior to both ends being joined by DNA ligase IV. The middle column demonstrates the single strand annealing pathway. This pathway is used to repair DSBs that occur near repeat sequences. It is initiated by the binding of the MRN complex to the free DNA ends. The EXO1 nuclease resects back the both sides of the break to create 3' ssDNA overhangs. RPA binds to the ssDNA. Single stranded repeat sequences either side of the DSB (shown on red) can bind via complementary base pairing creating 3' heterologous tails which are removed by ERCC1. The final major pathway which cells employ to repair DSBs is the homologous recombination pathway shown in the far right column. End resection occurs in a similar manner to that in single strand annealing. Following recruitment of RPA to 3' ssDNA overhangs RAD51/52 is recruited to the overhangs. Interaction between RAD51/52 and RAD54 facilitated strand invasion of a homologous DNA duplex. DNA synthesis then occurs creating a D-loop which can be resolved in different ways resulting in a non-crossover or crossover event. (Adapted from Vasileva & Jessberger, 2005)

1.6.2 *Homologous Recombination*

HR is an important repair mechanism in mitotic cells to correct DSBs and is essential during meiosis to promote pairing and segregation of chromosome homologs at the first meiotic division. It relies on the presence of a homologous template from which to repair from and this is generally the sister chromatid in mitotic cells or the chromosome homolog in meiosis (Mimitou & Symington, 2009).

HR is initiated by resection of the DNA ends at the DSB site to yield 3' single-stranded DNA (ssDNA) overhangs which are capable of invading duplex DNA containing a homologous sequence (Mimitou & Symington, 2009). The MRN complex (MREII, RAD50 and NBS1) together with CtIP (Sartori et al., 2007) is required for the initial end processing step of HR in human cells. The 5'-3-exonuclease EXO1 (Mimitou & Symington, 2008) or the combined helicase/nuclease activities of SGS1/DNA2 (Zhu et al., 2008) are involved in more extensive processing. The ssDNA is bound initially by a replication protein A (RPA), which removes any secondary DNA structures. RPA is displaced by BRCA2 which recruits the filament forming protein RAD51 via its BRC repeats (Yuan et al., 1999). RAD51 is a DNA-dependent ATPase that forms nucleoprotein filaments with DNA. Once recruited to the DSB, RAD51 catalyses strand exchange during which ssDNA invades homologous duplex DNA forming a displacement loop (D-loop). The mechanism by which strand invasion occurs is not fully understood, however, the DNA strand exchange activity of RAD51 is strongly stimulated by the helicase RAD54 which displays dsDNA-dependent ATPase (Petukhova et al., 1998), DNA translocase (Heyer et al., 2006) and DNA supercoiling and chromatin remodelling activities (Tan et al., 2003). Once formed, the D-loop has multiple fates (San Filippo et al., 2008). In the primary pathway in mitotic cells, termed synthesis-dependent strand annealing (SDSA), the 3' D-loop is extended by repair synthesis, and then the newly synthesised DNA strand dissociates to anneal to the other 3' end to complete the reaction. If the second 3' overhang is "captured" by the D-loop, a double Holliday junction forms that can potentially be resolved by several different proteins, GEN1 (Mimitou & Symington, 2009) and SLX1/SLX4 (Klein & Symington, 2009). As double Holliday junction

resolution can occur in different ways, crossover and non-crossover products are possible. Crossover products play an important role in facilitating chromosome segregation during meiotic recombination (Cole et al., 2010). Crossovers occurring during mitotic recombination may have deleterious effects, including loss of heterozygosity (Moynahan & Jasin, 2011). However, proteins such as BLM suppress mitotic crossovers by disrupting Holliday junctions, thereby decreasing the risk of genomic instability (Chu & Hickson, 2009).

1.6.3 *Single Strand Annealing*

Another repair pathway that makes use of sequence homology, but is distinct to HR, is single strand annealing (SSA) (Lin et al., 1984). SSA can occur after end resection if sequence-repeats exist on both ends of the DSB. The complementary single strands formed at the repeats then re-anneal and flaps formed from the annealing reaction are trimmed off, resulting in loss of sequence between repeats. Compared to HR, SSA is far more mutagenic because it involves loss of genetic information. Proteins identified to promote SSA in mammalian cells include RAD52 (annealing), ERCC1 and RAD1/RAD10 (flap endonuclease) (Stark et al., 2004).

1.6.4 *Non-Homologous End Joining*

NHEJ is initiated by DNA end-binding proteins Ku70 and Ku80 heterodimer, which rapidly associate with exposed DNA breaks. NHEJ can join DNA ends together with a number of different structures and as a result makes use of a number of processing steps that may include cleavage and gap filling prior to ligation. Artemis and DNA-PKcs exist as a complex within cells and binds to the Ku-DNA end complexes. Upon complex formation, DNA-PKcs phosphorylate Artemis, and Artemis acquires endonucleolytic activity on the 5’ and 3’ overhangs, as well as hairpins (Ma et al., 2002). This cleavage may result in gaps in the DNA that need to be filled in with polymerases. This gap filling reaction is most likely carried out by members of the PolX family which includes polymerases μ and λ, which interact with the Ku-DNA complex via BRCT domains (Ma et al., 2004). The modification of DNA ends prior to end joining by these processing steps can lead to deletions and insertions accounting for the more error-prone nature of NHEJ compared to HR. Ultimately, Ku/DNA-PKcs complex recruits DNA ligase IV/XRCC4 complex

which completes repair of the break. The XRCC4 acts as a scaffold that forms interactions with both Ku and DNA and both stabilises and stimulates ligase activity of DNA ligase IV (Grawunder et al., 1997).

Although NHEJ is imprecise at the local sequence level, it is efficient in restoring chromosomal integrity that would otherwise result in the loss of hundreds of genes on entire chromosomal arms or segments. Vertebrates have taken advantage of the error prone nature of NHEJ in the generation of antigen receptors for the adaptive portion of the immune system. In V(D)J recombination, the imprecision at the V-to-D and D-to-J joining sites markedly increases the amount of potential diversity that would otherwise be limited simply to the various combinations of V, D and J segments. The major disadvantage of NHEJ is the accumulation of randomly located mutations over time in the genome.

1.6.5 *Regulation of repair pathway choice*

In 1998, it was shown that the stage of the cell cycle is a decisive factor in the control of DSB repair (Takata et al., 1998). In this genetic study it was found that when the chicken cell line DT40 lacked Ku70, NHEJ was defective whereas when the cell line lacked Rad54, HR was moderately impaired. Both ΔKu70 and ΔRad54 cell lines were more sensitive to killing by ionising radiation relative to the WT DT40 cell line, suggesting both pathways contribute to repair of DSBs. A phase-specific sensitivity profile to ionising radiation clearly showed that HR and NHEJ are differentially employed during the cell cycle, with ΔRad54 cell lines showing ionising sensitivity during the S to G_2 phase and the ΔKu70 showing sensitivity at the G_1 phase. The restriction of HR to the S/G2 phases of the cell cycle makes sense in that the primary repair template in mammalian cells is the sister chromatid, which is not present in G1 cells.

The choice between repair pathways is also governed by cyclin-dependent protein kinases (CDKs), with DSB resection being a major site of control, an event necessary for HR but not for NHEJ. In yeast, CDK activity is required for efficient end resection of DSBs, and hence, HR, during S/G2 (Ira et al., 2004). They reported that DNA damage checkpoint activation by a DSB requires CDK1 in budding yeast. CDK1 was also found to be required for DSB-induced HR as inhibition of CDK1

resulted in a compensatory increase in NHEJ. CDK1 is required for efficient 5' to 3' resection of DSB ends and for the recruitment of RPA and Rad51 proteins.

More recent data have identified the role of a DNA damage response protein 53BP1. In response to DSBs, 53BP1 binds to chromatin at damaged sites and blocks end resection, thereby favouring repair by NHEJ (Bothmer et al., 2010). Pro-NHEJ functions require the direct physical association of 53BP1 with DNA ends and the DSB-induced phosphorylation of its N-terminal ATM/ATR kinase sites (Bothmer et al., 2011). The DNA damage response protein RIF1 has recently been identified as an essential factor recruited by phosphorylated 53BP1 to promote NHEJ and block HR (Chapman et al., 2012).

Many groups have attempted to manipulate the cell cycle and/or the concentrations of relevant proteins in order to discriminate between one pathway and another. A study involving the introduction of a point mutation into the αB-crystallin gene in one-celled mouse embryos found that by co-injection of ssDNA along with antibodies against Ku70/80, or by supplementing the system with hRad51/54, resulted in an increase of the rate of HR to enable targeted mutagenesis (Morozov & Wawrousek, 2008). A few studies have reported that a transient cell cycle arrest in the G2 phase using microtubule inhibitors vinblastine and nocodazole can increase targeted HR up to seven-fold (Urnov et al., 2005; Maeder et al., 2008). However, these compounds are highly toxic to cells and therefore would not be of use in a therapeutic setting. Recently, the cyclin-dependent kinase inhibitor indirubin-3'-monoxime has been shown to increase transduction of AAV and nuclease-mediated gene targeting (Rahman et al., 2013). The transient cell cycle arrest induced by indirubin increased AAV-mediated homology directed repair up to six-fold in human cell lines and ten-fold in umbilical cord-derived mesenchymal stromal cells. Minimal cytotoxicity was observed and so indirubin may prove useful for future clinical applications.

1.6.6 *Homology-directed repair (HDR)*

DSBs are repaired efficiently as they arise naturally during the life cycle of a cell. The repair process of HR using the sister chromatid as a repair template gave way

to the hypothesis that if artificial DNA repair templates were introduced, could they be used to repair mutations in the genome. That is, if you provide a donor plasmid containing the correct sequence of DNA and homology arms, would recombination be significantly increased with the generation of DSB close to the mutation locus?

The first methods for homologous genome editing were described in yeast whereby a stable leu-2^- yeast strain was transformed to leu-2^+ by using a chimeric ColE1 plasmid carrying the yeast leu-2 gene (Hinnen et al., 1978). This simple method of genome modification demonstrated that plasmid DNA and yeast sequences could integrate into the yeast chromosome, not only at the intended leu-2 region, implying homology plays a role in the insertion event but also in several other chromosomal locations albeit to a lesser extent than the intended location.

The first evidence of extrachromosomal homologous recombination in mammalian somatic cells was provided by Kucherlapati and colleagues (Kucherlapati et al., 1984), thus showing that an exogenous donor sequence can be used for HR repair rather than relying on the sister chromatid. They demonstrated this principle using two plasmids, each containing a different mutation in a neomycin resistance gene such that either plasmid alone does not confer resistance to this drug on cells (human bladder carcinoma cell line), but the presence of both results in HR between the plasmids to create a functional neomycin resistant cDNA sequence and thereby confers neomycin resistance to the cells. This application of HR for the repair of mutations has been named homology-directed repair to distinguish it from endogenous HR processes.

-Length of homology arms and sequence divergence

Multiple groups have explored the optimal length of homology required for efficient HR. Thomas and Capecchi first described an exponential relationship between length of homology and targeting frequency in the HPRT locus (Thomas and Capecchi, 1987). A later study comparing the frequency of gene targeting with varying lengths of homology revealed that the minimum length required for HR to occur was 1.7 kb (Hasty et al., 1991). They found that an additional 225 bp increased the frequency by at least five-fold. An increase in homology from 1.9 to 4.2 kb and from 4.2 to 6.0 kb resulted in a further 16-fold and 3-fold increase

respectively, with a total difference of 240-fold efficiency between 1.7 and 6 kb. Of note in this study, all vectors used to introduce the target site were linearised prior to electroporation into mouse embryonic stem (ES) cells. They determined that the 1.3 kb vector contained only 132 bp of homology on the 5' end, which may be too little homology for efficient HR to occur. Homology arms of equal length on either side of the target sequence may be more efficient. A comparison of vectors with total arm length of 6-8 kb to vectors with arms longer than 18.4 kb revealed that increasing the length higher than 8 kb did not increase the efficiency of HR. Again all vectors were linearised prior to transfection (Lu et al., 2003). As well as homology arm length the degree of homology between the template sequence and the sequence to be repaired is also important. A study by Elliott et al., (1998) revealed that 1.2% sequence divergence resulted in a six-fold decrease in recombination (Elliott et al., 1998). This study also gave some insight into the effect of distance from the DSB/cut site on repair, with the majority of events (~80%) occurring at 58 bp or less.

-Increasing the rate of HDR by inducing a DSB

The rate of HR in mammalian cells typically falls in the range of 10^{-5} to 10^{-6} events per cell per generation, however, a DSB can stimulate HR by >1000 fold (Sun et al., 2012). It had been shown that linear vectors could increase the level of HDR in cells, leading to the hypothesis that DSBs in the DNA are highly recombinogenic and that broken ends of DNA molecules can interact directly with homologous chromosomal regions (Orr-Weaver et al., 1981). However, to harness the recombinogenic potential of a DSB it is necessary to create a unique DSB at the genomic site of interest. Early studies used the homing endonuclease I-SceI from *S. cerevisiae*, which recognises a unique 18 bp nonpalindromic sequence (Colleaux et al., 1988) creating a DSB on average once every 7×10^{10} bp (4^{18}). Rouet et al., expressed a modified I-SceI endonuclease in COS1 mammalian cells, resulting in cleavage and enhanced extrachromosomal recombination (Rouet et al., 1994). Recombination was measured by chloramphenicol acetyltransferase (CAT) activity which can only be expressed following successful HR. CAT activity was very low when the recombination substrates were transfected by themselves. However a ten-fold increase was detected following transfection along with I-SceI. As there are

only a handful of homing nucleases like I-SceI and a limited number of recognition sequences, the chances of cleaving in a gene are virtually zero. Also the limitation of having to pre-insert the I-SceI site into the genome makes it unsuitable for more advanced gene modification experiments. Another study in the same year demonstrated that the non-specific nuclease domain *Flavobacterium okeanokoites* endonuclease (*Fok*I) could be removed and retargeted to bind a different DNA sequence (Kim & Chandrasegaran, 1994). *Fok*I is a type II restriction endonuclease that recognises a nonpalindromic sequence in duplex DNA and cleaves nine and 13 bp away from the recognition site. They reported the presence of two distinct domains within this enzyme; a sequence-specific domain and a non-specific nuclease domain (cleavage domain). By linking the *Drosophila* Ultrabithorax (Ubx) homeodomain to the cleavage domain of *Fok*I they were able to redirect *Fok*I cleavage to the Ubx recognition site indicating that the non-specific domain could be used in conjuction with DNA binding proteins to target any sequence in the genome. As random introduction of genetic material in higher organisms can lead to unintended effects due to insertional mutagenesis, targeted genome modification to a specific location is desired.

1.7 ZFN, TALEN and CRISPR/Cas-based tools for genome editing

The previous section showed the importance of a DSB to induce HDR. The way forward is to use 'programmable enzymes' to create specific DSBs at any desired location. This section describes the advances in the last 8-10 years in this area.

1.7.1 *Zinc Finger Proteins (ZFPs)*

The Cys_2-His_2 (C_2H_2) zinc-finger (ZF) domain is a type of DNA-binding motif found in eukaryotes and is the second most frequently encoded protein motif in the human genome accounting for ~2% of all genes (Beerli & Barbas, 2002). The C_2H_2 class of zinc fingers were first identified in *Xenopus laevis* transcription factor IIIA (TFIIIA) (Miller et al., 1985). TFIIIA was shown to contain nine tandem C_2H_2 zinc fingers each consisting of 30 amino acids, containing two invariant pairs of cysteines and histidines, the most common Zn^{2+} ligands. X-ray crystallography

confirmed that a ZF was composed of two beta strands and an alpha helix (Lee et al., 1989). The two cysteines are near a turn in the anti-parallel beta sheet, and the two histidine residues are in the C-terminal portion of the alpha helix. By studying the X-ray crystal structure of DNA complex with Zif268 (a mouse immediate early protein) at 2.1Å resolution, Pavletich et al., were able to show that ZFs bind to the major groove in a simple tandem arrangement with a semi-conserved pattern of interactions in which three/four amino acids from the alpha helix contact three/four adjacent bases in DNA (Pavletich & Pabo, 1991). The most specific amino-DNA interactions occur at positions -1, +3 and +6 from the alpha helix and a fourth interaction between the amino acid at position +2 and a base on the opposite DNA strand (Figure 1.6a) (Jacobs, 1992). The overall binding arrangement has the first (N-terminal) finger near the 3' end of the recognition site and the last (C-terminal) finger near the 5' end of the recognition site. Typically the amino acid residue at -1 bind the 3rd base, the +3 binds the middle base and the +6 residue binds the 5' base with the +2 residue binding the 4th base on the opposite strand.

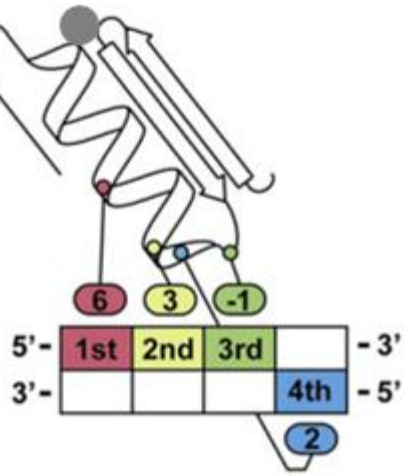


Figure 1.6a: Interaction of a zinc-finger domain with 4-bp DNA target observed in the complex. In the complex, one-to-one interactions between all recognition amino acids and DNA bases at specific positions were observed for the first time. (Adapted from Sera, 2009).

As it appeared that each ZF works as an individual module, it was proposed that fingers with different triplet specificities could be combined to create a zinc finger protein (ZFP) that could recognise longer DNA sequences (Miller et al., 1985). They were shown to be usually linked by the short sequence TGEKP (Pellegrino & Berg, 1991). Studies have indicated that this linker sequence is not just a passive structural element but can actually stabilise the alpha helix assisting in binding the protein to DNA (Laity et al., 2000). The first example of ZFPs with altered binding specificities was described in 1992. In this study they mutagenised the amino acid at position three of the 2nd finger of Sp1 based on a database of zinc finger sequences available at the time, and presented the first partial recognition code of GNG and GNT triplets (Desjarlais & Berg, 1992). Since these early manipulations a number of methods are now available to design specific ZFPs to target any sequence in the genome (figure 1.6b) see section 1.7.3.

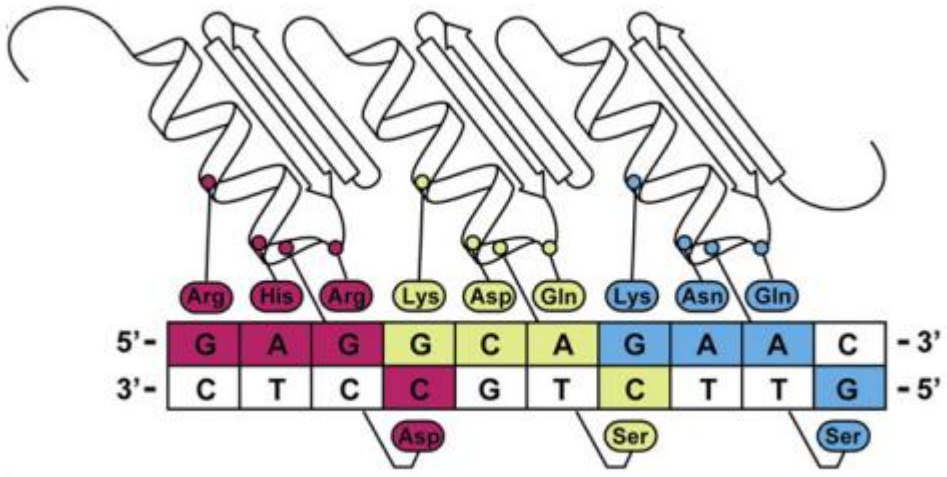


Figure 1.6b: Interaction of 3-finger ZFP with double stranded DNA. The 4 bp DNA targets are overlapped. (Sera, 2009)

1.7.2 *Zinc finger nucleases: designer nucleases for HDR*

The first group to link the cleavage domain of the *Fok*I endonuclease to two different ZFPs and create a zinc finger nuclease (ZFN) fusion protein showed that cleavage was indeed directed towards the recognition site of the ZFPs (Kim, et al., 1996). In this study they designed and tested two ZFNs *in vitro* with the λ phage genome (~48 kb). The targeted ZFN site was successfully cleaved in a sequence-specific manner with a cleavage efficiency of >95%. In 1998, a study determined that WT *Fok*I acts as a dimer. Introduction of mutations into the nuclease domain interface of *Fok*I, making it defective for dimerisation, revealed greatly impaired DNA cleavage (Bitinaite, et al., 1998). It was also demonstrated in this study that free nuclease may interact with bound nuclease to mediate DNA cleavage when the nuclease is in excess. In order to cleave at a desired location two 9 bp DNA sequences in inverted orientation and separated by 6-35 bp should be selected, as reported by Smith et al., (2000). They determined that when ZFNs bind in a tail to tail fashion (see figure 1.7) the rate of DNA cleavage occurs almost to completion with cleavage observed at a 1:1 ratio of ZFN to DNA. No cleavage was observed when only one ZFN was used, however at a ratio of 5:1 some cleavage occurred *in vitro*.

A)

B)

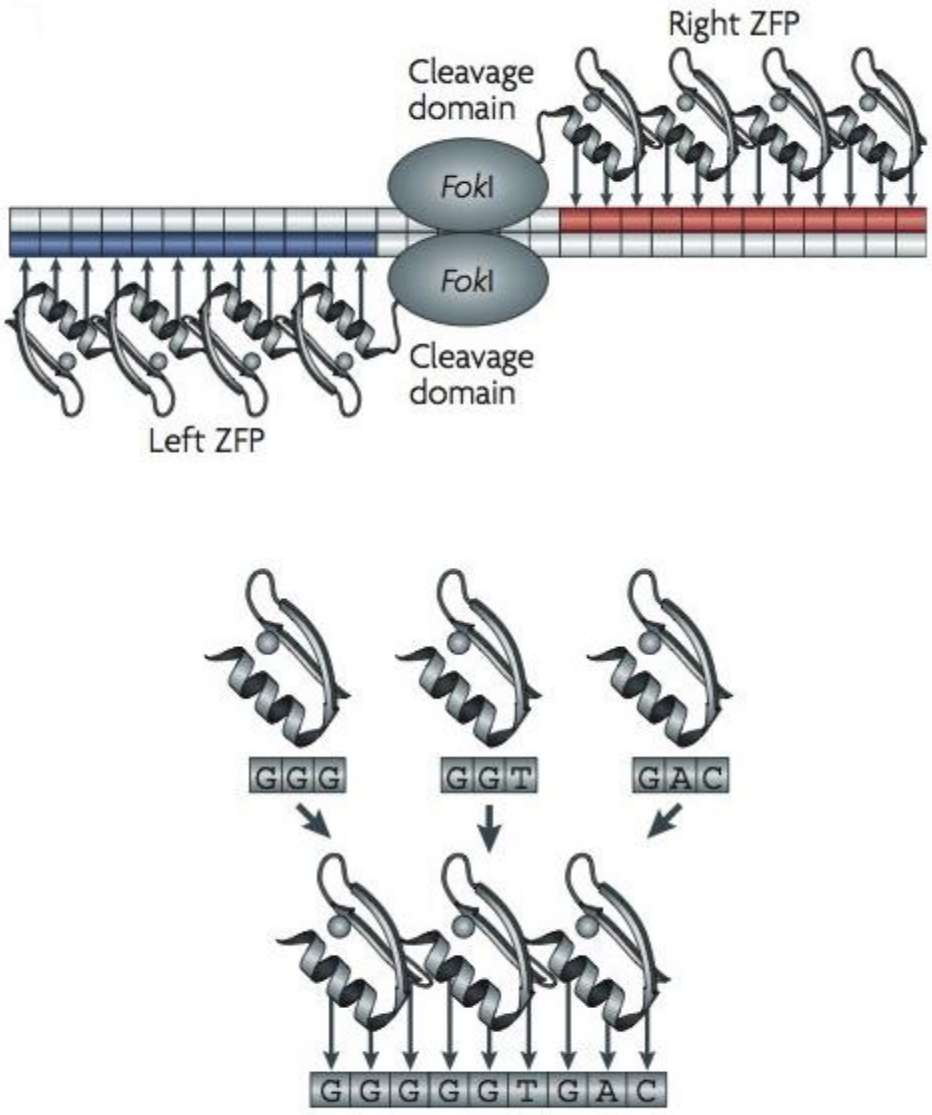


Figure 1.7: A) A pair of ZFNs bound to their respective sites in a tail-to-tail fashion and on opposite strands. This conformation enables efficient *Fok*I nuclease domain dimerisation resulting in DSB formation. **B)** The modular assembly method uses phage display and gel mobility assays to isolate ZF motifs for specific target triplets. These motifs are then linked to create ZFPs with longer DNA recognition sequences. This method is limited to designing 3-finger ZFPs. (Adapted from Rebar & Miller, 2004 and Miller et al., 2007).

The first report of the ability of ZFNs to create a DSB and induce homology-directed repair was demonstrated in *Xenopus laevis* (Bibikova et al., 2001). The group microinjected plasmids containing two 1.25 kb repeats separated by ZFN binding sites into *X. laevis* oocyte nuclei along with the ZFNs. Repair of DSB formation by ZFNs in these cells would proceed by the single-strand annealing mechanism used to repair repeat sequences (see section 1.6.3). Homology-directed

repair was only observed when two ZFN binding sites were orientated in a tail to tail fashion with no repair detected with a single copy of the ZFN binding site indicating that conditions for cleavage in oocyte are stricter than *in vitro.* This study also looked at the effect of spacer length between the two ZFN binding sites as well as the length of the linker between the ZFP and *Fok*I nuclease domain on DNA cleavage. Unlike the *in vitro* data reported previously by Smith et al., (2000) whereby all spacers from 6 to 35 bp were effectively cleaved when ZFN protein was present in modest excess (Smith et al., 2000); the effective spacer in oocytes was 6 to 18 bp with a 6 bp spacer being the most efficiently cleaved. The most effective linker between a ZFP and *Fok*I nuclease domain was found to be between 0 and 6 amino acids.

A study in 2009 tested several linker sequences of different lengths with different spacer regions (Shimizu, et al., 2009). They used a plasmid-based single strand annealing (SSA) reporter assay in HEK 293 cells. Briefly, a luciferase gene was disrupted by a stop codon and a ZFN target site. A ZFN-induced DSB would allow SSA HR with an 870 bp region of homology in direct repeat orientation resulting in active luciferase activity. They showed that a 6 aa linker TGAAAR enhanced activity of ZFNs separated by a 6 bp spacer sequence and was most specific. An earlier study had reported highly-efficient HR in human cells using a ZFN with 4 aa linker on a target with a 5 bp spacer (Urnov et al., 2005). The exact spectrum of activities is most likely to depend on a variety of factors including cell type, assay used, composition and concentrations of the ZFNs and target sites. Of note, ZFs are not restricted to just three fingers, up to six fingers can be linked together to target a total of 36 bp.

1.7.3 *Zinc Finger Protein Design*

This section will look at a number of methods that have been developed to enable scientists to design and create their own ZFNs in the lab. Today ZFNs are readily available commercially from Sangamo Biosciences in partner with Sigma-Aldrich however pricing makes these ZFNs unattainable for some research groups who opt to design their own.

1.7.4 *Modular Design*

There are 64 possible DNA triplets which can be recognised by ZFs. One design strategy was to identify highly specific fingers for all 64 triplets in order to establish a library for rapid construction of ZFPs (Pavletich & Pabo, (1991), Liu et al., (2002), Dreier et al., (2001, 2005)). Then via modular assembly ZFs for specific triplets can be linked together to recognise longer DNA sequences.

From phage display selection and studying the Zif268 crystal structure several proposed codes emerged for the binding of ZFPs to DNA targets (Choo & Klug, 1994). Screening methods based on gel mobility shift assays were used to isolate ZF motifs that not only bind all sixteen 5'-GNN-3' triplets but were optimised to bind all sixteen triplets at the first, second and third position of a ZFP (Liu et al., 2002). Gel mobility shift assays were used to isolate ZFPs with high affinity for GNN triplets. Selected ZFPs were then re-tested up to five times via gel mobility shift assays with a degenerate oligo pool to select for ZFPs that showed both high specificity and high affinity. Most ZFP sequences identified matched those obtained by phage display in a previous study (Segal et al., 1999). Other phage display studies have isolated ZF motifs that bind twelve of all possible sixteen 5'-ANN-3' (Dreier, et al., 2001) and fifteen 5'-CNN-3' triplets (Dreier et al., 2005). Only a handful of ZF motifs that bind 5'-TNN-3' triplets have been published, the best characterised being TGG which is the second finger of Zif268 (Pavletich & Pabo, 1991). It is important to note that apart from the Liu et al. 2002 study in which they looked at the positional effect of a ZF motif on amino acid sequence, all other phage display methods selected ZFs for triplets at the second position. The availability of ZF libraries allows assembly of ZFPs that can recognise a diverse range of genomic targets.

In spite of this large pool of resources, it has been reported that such designed ZFPs have an unexpectedly high failure rate (Ramirez et al., 2008). In this study 168 ZFPs designed via modular assembly were tested against 104 DNA target sites in a bacterial 2 hybrid (B2H) assay. This system can accurately identify arrays that lack ZFN activity in human cells as determined by the expression of β-galactosidase. Of the 104 target sites tested, they did not obtain a single three finger array that scored positive in the B2H assay for 79 sites, (76% failure rate). The preference of ZFs for

guanine rich sequences was noted early (Pavletich & Pabo, 1991) and interestingly, when the authors grouped target sites based on number of GNN triplets present, they found that the number of GNN triplets had a remarkable effect on the success of ZFP binding efficiency. Only 41% of target sites containing three GNN triplets failed to bind a ZFP. This failure rate rose to 71%, 88% and 100% with 2 GNN, 1 GNN and 0 GNN respectively. Therefore, even with three GNN triplets in the target site, the chance of designing an efficient pair of ZFPs is only 60%.

1.7.5 *Rational Design*

In 2002, Sera and Uranga published a non-degenerate recognition code table to allow rapid creation of numerous ZFPs with satisfactory binding properties. The table was based on a designed ZFP with three ZFs that recognise overlapping 4 bp sites in a 10 bp target sequence. Known and potential DNA base-amino acid interactions derived from X-ray crystal structures allowed identification of an amino acid for each position of the α-helical region (-1, 2, 3, and 6) of the ZF domain which contact DNA bases at specific positions in a regular fashion implicating four amino acids in DNA recognition (Sera & Uranga, 2002).

	1st Base	2nd Base	3rd Base	4th Base
G	Arg	His	Arg	Ser
A	Gln	Asn	Gln	Asn
T	Thr	Ser	Thr	Thr
C	Glu	Asp	Glu	Asp
	Position 6	Position 3	Position -1	Position 2

Figure 1.8: Non degenerate recognition table design (Sera & Uranga, 2002). List of amino acid residues which make contact with the desired DNA bases at each position of a ZF recognition site.

For the most part, amino acids with shorter and smaller side chains were used for recognition of the second and fourth DNA bases and amino acids with larger and longer side chains were used for recognition of the first and third bases. Threonine was used for thymine recognition at positions -1, 2 and 6 as the methyl group of threonine interacts with the methyl group of thymine via hydrophobic interactions (Aggarwal et al., 1988) and this interaction could be strengthened by the potential hydrogen-bonding of the hydroxyl group of threonine and the O_4 of thymine. Serine was used at position 3. In this study, they used the non-degenerate recognition table to design ten 3-finger ZFPs, five bound to their targets with high affinity and specificity while the other five did not. It was reported that the five successful ZFPs had at least three guanines at various positions in the first nine bps of DNA, whereas the unsuccessful ZFPs had less than two guanines in the first 9 bases reiterating the finding by Pavletich & Pabo, (1991) that the number of guanines in the DNA target has a remarkable effect on ZF binding.

1.7.6 *ZFP Oligomerised Pool Engineering "OPEN" Selection*

A potentially more effective way to generate multi-finger arrays is the OPEN method, however, this method is much more labour intensive compared to modular assembly as it requires construction and interrogation of large randomised libraries (typically $>10^8$ constructs). Screening methods have focused on screening Zif268 variants in which ZF1, ZF2 and ZF3 was randomised at the alpha helix positions -1, 1, 2, 3, 5, and 6 (Maeder et al., 2008). Zif268 variants were tested against all sixteen GNN triplets' at all three positions and between 5 to 7 TNN triplets at each position in a B2H assay using Zif268 target sites with two Zif268 recognition triplets and the target triplet. Ninety-five ZFs for each triplet at each position were then pooled to create ZF pools. These pools were recombined to create a ZFP library consisting of $95^3 = 8.6 \times 10^5$ ZFPs which were interrogated using a B2H selection system against a 9 bp target site. Successful ZFPs were determined by the expression of the histidine synthesising gene *his*3B which allowed growth of bacteria on histidine deficient media. Selection of highly active ZFPs was achieved with a gradient of 3-aminotriazole (3-AT), a competitive inhibitor of *his*3B. In order for OPEN selection to target any genomic sequence, 192 finger pools need to be established. However once established, this method yields highly active and specific ZFNs.

1.7.7 *Context-dependent assembly (CoDA)*

CoDA is a publically available platform for engineering ZFPs using standard cloning techniques. Three-finger arrays are assembled using one of the 319 N-terminal and one of the 344 C- terminal fingers that have been previously identified in other arrays engineered to function when positioned adjacent to one of the fixed 18 middle fingers. In contrast to OPEN and modular assembly methods, CoDA does not treat the fingers as independent modules but instead accounts for context-dependent effects between adjacent fingers, thereby, increasing the probability that a ZFN will function efficiently (Sander et al., 2010). Target sites for ZFPs designed using the CoDA method are reported to be found in diverse sequences on an average of every 500 bp.

1.7.8 *Transcription Activator-like effectors (TALE)*

TALE proteins are naturally occurring proteins from the plant pathogenic bacteria genus *Xanthomonas* that act as transcriptional activators in the plant cell nucleus. Binding to DNA is through a central domain of tandem repeats. Boch et al. (2009) were able to decipher the DNA recognition code that allows this binding and found that the DNA-binding domains consisted of a series of 33-35 amino acid repeat domains that recognises a single base pair of DNA (Boch et al., 2009). The specificity is determined by two hypervariable amino acids (aa 12 and 13), known as repeat-variable di-residues (RVDs). The four most common RVDs are shown in table 1.2 and figure 1.9a with their most frequently associated nucleotide.

Repeat Variable Domain	**Associated nucleotide**
HD	C
NI	A
NG	T
NN	G

Table 1.2: The four most common RVDs with their associated nucleotides

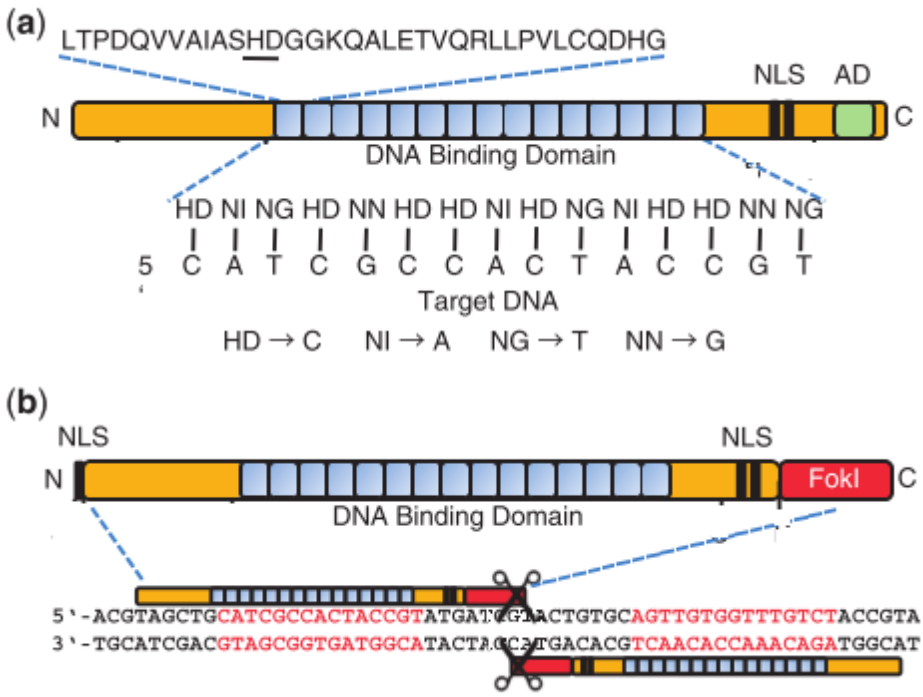


Figure 1.9: (a) Structure of the naturally occurring TAL effector. A consensus repeat sequence is shown with the sequence determining variable-repeat di-residues underlined. The four most common RVDs are shown with their most frequently associated nucleotide. **(b)** Structure of a TALEN. Two monomeric TALENs bind DNA to allow *Fok*I to dimerise and cleave the target site. NLS; nuclear localisation signal, AD; transcriptional activation domain. (Adapted from Cermak et al., 2011).

Like ZFs, TALE proteins can be linked together to recognise DNA sequences and then fused to a cleavage domain like *Fok*I to make a TALE nuclease (TALEN) (Christian et al., 2010). Each target half-site is typically constructed to bind 14-20 bp with a spacer of 12-19 bp between binding sites. However, in contrast to ZFPs there is no reengineering necessary of the linkage between repeats to enable construction of long arrays of TALEs with the ability of targeting unique sites in a genome. The only targeting limitation for TALE arrays is that TALE binding sites should start with a T base (Boch et al., 2009). The single base recognition of TALE-DNA binding repeats allows greater flexibility than triplet-confined ZFPs. However the cloning of repeat TALE arrays presents a technical challenge due to extensive repeat sequences. Numerous strategies have been developed to overcome this

challenge such as ‘Golden Gate’ molecular cloning which uses a type IIS restriction endonuclease to cleave outside the recognition sequence to create 4 bp overhangs. Cloning is then expedited by digesting and ligating in the same reaction mixture as correct cloning eliminates the enzyme recognition site (Cermak et al., 2011). Using this cloning system, Kim et al. (2013) constructed a library of TALENs spanning the human genome by assembling TALEN plasmids for 18,740 protein-encoding genes (Kim et al., 2013). Several protocols are available describing rapid assembly of custom TALENs using publically available reagents from Addgene (Cermak et al., 2011). Reyon et al. (2012) recently presented the largest collection of TALENs to date using the fast ligation-based automatable solid-phase high-throughput (FLASH) system (Reyon et al., 2012), whereby DNA fragments are assembled on solid-phase magnetic beads rather than in solution. They assembled a total of 144 TALEN pairs; 48 pairs were tested in a human cell-based EGFP reporter system and were all found to process efficient gene-modification activities, 96 pairs were targeted against endogenous human genes implicated in cancer regulation and 88% successfully cleaved their target site.

1.7.9 *CRISPR/Cas system*

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR associated protein Cas9 is the latest gene editing tool. It is a prokaryotic RNA-programmable nuclease system that provides natural immunity in bacteria against invading viruses and plasmids (Wiedenheft et al., 2012). This defence system relies on small RNAs for sequence-specific detection and silencing of foreign nucleic acids. An *in vitro* study of the *S. pyogenes* type II CRISPR system demonstrated that CRISPR RNA (crRNA) fused to normally trans-encoded trans-activating crRNA (tracrRNA) is sufficient to direct Cas9 protein to sequence-specifically cleave DNA sequences matching the crRNA (Jinek et al., 2012). This raised the possibility that the approach could be used for targeted gene editing *in vivo.* The target sequence is specified by a 20 bp CRISPR RNA and the only constraint associated with this sequence is the requirement for the bases NGG, known as the protospacer adjacent motif (PAM), immediately following the 20 bp target sequence. In January 2013, it was reported that the type II CRISPR system could be used to reengineer the genome in human cells. A human codon optimised version of

Cas9 was cloned into a mammalian expression system and the human U6 polymerase III promoter was used to express crRNA-tracrRNA fusion transcripts, called guide RNAs (gRNAs). The U6 promoter requires a 'G' in order to initiate transcription and along with the requirement for the PAM sequence, any genomic site of the form $GN_{20}GG$ can potentially be targeted (Mali et al., 2013). A major advantage of this CRISPR system over ZFNs/TALENs is that in order to target a specific sequence a simple 20 bp oligo is designed according to $GN_{20}GG$ and cloned into the U6 expression plasmid followed by co-transfection along with the Cas9 protein compared to the labour intensive design and assembly methods described above. As depicted in figure 1.10 RNA-guided gene targeting in human cells involves co-expression of the Cas9 protein bearing a SV40 nuclear localisation signal at the C-terminus with a guide RNA expressed from the human U6 promoter. Cas9 unwinds the DNA duplex and cleaves both strands upon recognition of the target sequence by the gRNA. Another advantage of the CRISPR system is that the architecture of natural CRISPR loci with multiple tandem guide templates makes it possible to target many targets in a single experiment as was seen in a study by Wang et al. 2013. In this study five genes were simultaneously disrupted in embryonic mouse cells using the CRISPR/Cas system (Wang et al., 2013). A follow on study from this reported the one-step generation of mice carrying reporter constructs in three different genes (Yang et al., 2013). They also produced conditional mice by microinjecting two Mecp2-specific gRNAs, Cas9 mRNA and two oligos encoding loxP sites into fertilised eggs. Usually the development of a new mouse model for a disease involves careful breeding of multiple generations and can take at least a year, with CRISPR technology; a new mouse model can be ready for testing in a matter of weeks.

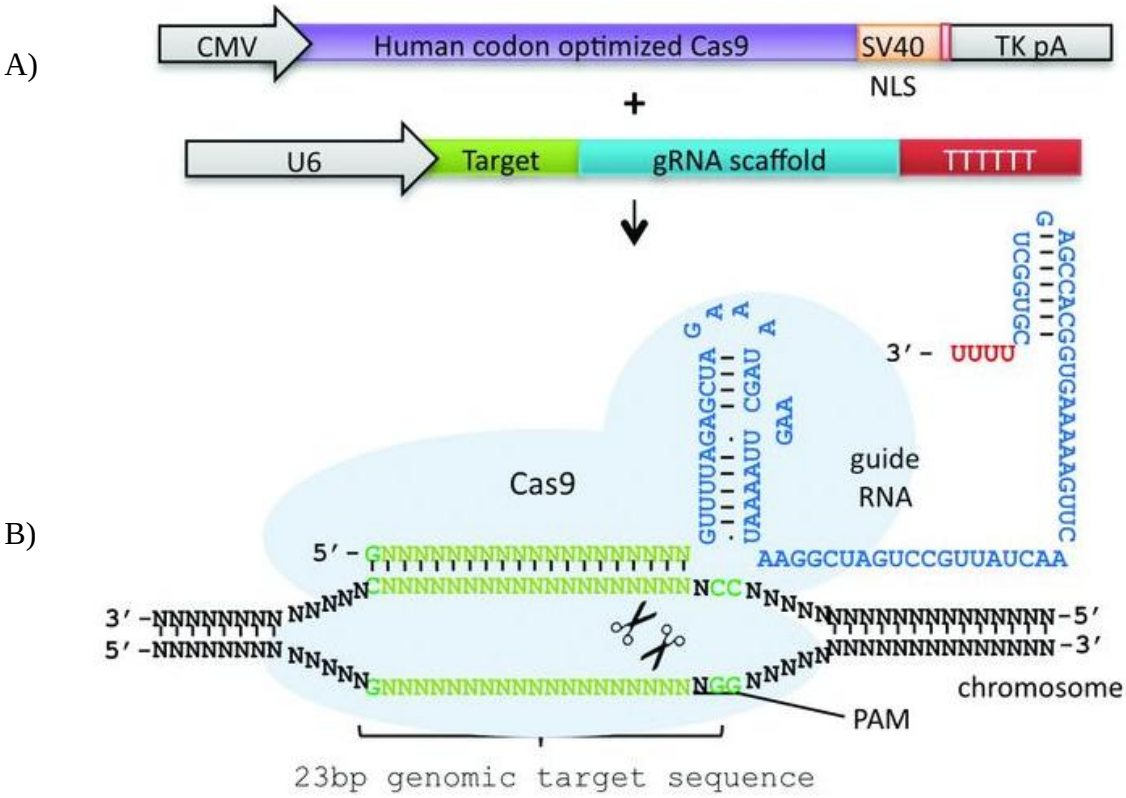


Figure 1.10: Schematic representation of gene editing in human cells using type II CRISPR system. A) Human codon optimised Cas9 under the control of the CMV promoter and NLS at the C terminal is co-transfected into cells with gRNA expressed from U6 promoter. B) Inside the cell, Cas9 unwinds the DNA and gRNA binds and cleaves DNA to create a DSB. (Mali et al., 2013).

In terms of flexibility in binding DNA sequences ZFNs still suffer from context effects on DNA recognition and a limited range of targetable sequences, due to the lack of availability of fingers for all possible DNA triplets. TALE nucleases (TALENS) appear to be less subject to these constraints due to greater simplicity of their DNA-binding domains, with one well defined module for each of the four bases. However, some degree of ambiguity emerged recently when it was shown that the module used to recognise G also binds A. To increase specificity of these G recognising modules it was shown that a module with a RVD NK sacrifices affinity to gain specificity, while modules with an NH RVD show improved biological activity and specificity (Cong et al., 2013). The new CRISPR/Cas system with its efficiency and ease of use looks set to be the most flexible out of all the gene editing tools in target recognition, however, the incidence of off-target events have still not been fully explored and the outcome of these studies will ultimately decide the future of this system (Pennisi, 2013).

1.8 Gene modifications using programmable and RNA-guided nucleases

The advances made in efficient gene targeting have enabled investigators to manipulate many genes in a diverse range of cells and organisms. Gene disruption involves the use of ZFN/TALEN/CRISPR (nucleases) to target and abolish the function of an endogenous locus, creating a gene knockout. Nucleases delivered to a cell, in the absence of a donor, result in a DSB which will be repaired by NHEJ leading to insertions/deletions (indels) of nucleotides. If this cleavage occurs in a coding region of a gene, these indels will usually result in loss of gene function. Gene correction involves delivery of nucleases along with donor DNA which result in precise modification of a targeted locus. Nucleases are designed to create a DSB near the site of the gene mutation, and the donor DNA is used as a template to correct this DSB by HDR (figure 1.11). In this section I will discuss a number of advances made in the last decade using nuclease-induced gene modifications.

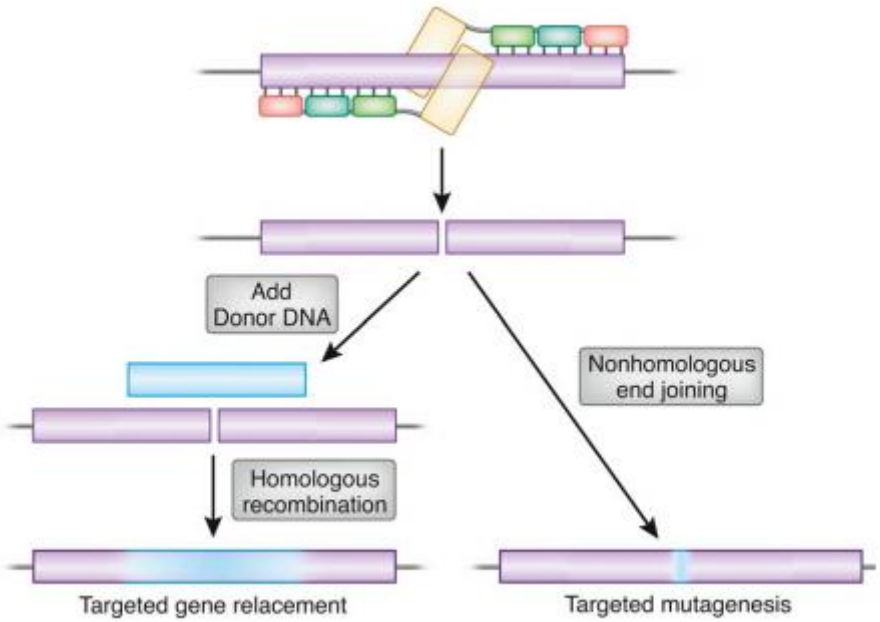


Figure 1.11: Repair outcomes following a DSB induced by nucleases. The nuclease binds the target DNA and creates a DSB. If a homologous donor DNA is present repair can proceed through the HR pathway using the donor DNA as a template. Alternatively the break can be repaired by NHEJ resulting in mutations at the cleavage site (Carroll, 2011).

1.8.1 *Gene Disruption*

This technique has been applied to the CCR5 locus of human CD4+ T cells in an effort to make them resistant to HIV infection. CCR5 is the major co-receptor for HIV-1 entry (Deng et al., 1996). It was discovered that cells with a 32 bp deletion in both CCR5 alleles confers resistance to HIV-1 (Samson et al., 1996) and so CCR5 became a target for HIV therapy. ZFNs targeting the CCR5 gene resulted in non-functional gene products that would fail to be expressed at the cell surface and so prevent HIV entry. These ZFNs were delivered *ex vivo* using Ad5/35 vector to primary CD4+ T cells and resulted in >50% gene disruption of transduced cells (Perez et al., 2008). CCR5 was also successfully disrupted in human CD34+ haematopoietic stem/progenitor cells at a rate of 17% of the total alleles in the population (Holt et al., 2010). These modified cells were then engrafted into mouse HIV models and when challenged with CCR5-tropic HIV-1 showed significantly lower (10-fold decrease) HIV-1 levels compared to controls at 6-weeks post-infection. Clinical trials are underway using this approach with the goal to improve HIV positive patients' immune system by allowing their CD4+ T cells to survive longer (ClinicalTrials.gov. Trial numbers: NCT00842634, NCT01044654, and NCT01252641). Data from the first phase I clinical trial demonstrated improvement in several clinical parameters while being well tolerated (Perez-Pinera et al., 2013).

TALENs have also been designed to target the CCR5 gene. A study in 2011 compared the efficiency and specificity of both ZFNs and TALENs targeting this gene. Specificity is an important parameter of designer nucleases and will be discussed in more detail in section 1.9. Off-target events are very undesirable as they can cause mutagenesis which can result in cancer. A major off-target locus of the CCR5 binding nucleases is the gene CCR2 which shares a high degree of sequence identity with CCR5 (figure 1.12).

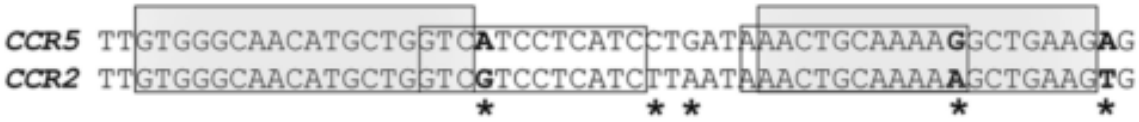


Figure 1.12: Sequences of CCR5 and CCR2 genes. Asterisks indicate the differences between the two target sequences. Letters in bold indicate the nucleotides affecting binding by the designer nucleases used in the study. Target sites are highlighted by grey shaded (TALEN) or black outlined boxes (ZFN). (Mussolino et al., 2011).

In this study, the ZFN target site in CCR5 differed from the CCR2 site by two nucleotides, one in each half site. In contrast, the entire 19 bp target sequence of the left TALEN subunit was found in CCR2 while the right target half site varied at one position (Mussolino et al., 2011). Measurement of indels in the CCR2 gene following transfection with either ZFNs or TALENS specific to the CCR5 locus revealed mutations at 11% and 1% respectively. Even though gene disruption at the intended CCR5 target was similar with either ZFNs or TALENs the off target events at CCR2 were much lower with TALENs. This study was pivotal in demonstrating that the TALEN platform enables the design of nucleases with single nucleotide specificity.

Generation of knock-out (KO) animals has been made easier and more efficient due to the advancements in gene editing tools. ZFNs targeting the interleukin 2 receptor gamma (IL2RG) were used to successfully create KO rats which were used as models to study X-linked severe combined immunodeficiency (X-SCID) (Mashimo et al., 2010). ZFNs were injected as mRNA into the pronucleus of fertilised oocytes and yielded genetically modified offspring at rates greater than 20% which processed a number of indel mutations thus disrupting the gene. These genetic changes were faithfully transmitted to the next generation along with the SCID phenotype. TALENs have been used to disrupt a number of genes in a wide variety of model organisms including the zebrafish. Cade et al. (2012) constructed TALENs to target ten different endogenous genes of the zebrafish and found that 7/10 pairs of TALENS induce indels at the correct location with efficiencies ranging from 2-76% (Cade et al., 2012).

The latest gene editing technology CRISPR/Cas system has proven to be very efficient at generating mutations in multiple genes. A recent report demonstrated that a single CRISPR array encoding two loci separated by 119 bp could be cleaved simultaneously in cultured human cells at a relatively low efficiency of 1.6% (Cong et al., 2013). Another group tested the potential to disrupt five genes simultaneously in mouse ES cells. Guide RNAs targeting five genes, Tet1, Tet2, Tet3, Sry and Uty were designed, pooled and co-transfected with a Cas9 plasmid into ES cell. Of 96 clones screened, 10% carried mutations in both alleles of the five genes demonstrating the capacity of CRISPR/Cas system as a powerful tool to efficiently target multiple genes in a single experiment (Wang et al., 2013).

1.8.2 *Gene Correction*

Gene correction is the precise modification of a targeted locus using custom nucleases and donor DNA. Following the creation of a DSB in the target locus close to the mutation by a nuclease, homologous recombination with the Donor DNA (which contains homology arms complementary to the target sequence flanking the mutated gene and a central region containing the correct sequence) can result in gene correction.

The potential of ZFNs to create DSBs and increase the efficiency of HDR in human disease was highlighted in an experiment that corrected a mutation in the IL2RG gene in a human cell line (Urnov et al., 2005). As already stated mutations in the IL2RG gene give rise to X-SCID (see 1.8.1). First, to rapidly assess the potential for ZFN-driven editing of the genome, ZFNs were initially tested in a novel *ex vivo* EGFP repair assay. A mutated EGFP construct containing the ZFN target site was stably integrated into the genome of HEK293 cells. ZFNs were introduced into cells along with a donor plasmid containing the wild type EGFP sequence flanked by 1.5 kb total homology arms. The level of ZFN induced HDR was assessed by flow cytometry. The optimised ZFNs were shown to be five times more potent at EGFP correction than donor only controls. To determine if ZFNs can evoke a comparable increase in HR frequency at the IL2RG locus, a pair of four finger ZFs were designed to recognise a total of 24 bp of the IL2RG sequence which were optimised *in vitro* by guided single amino-acid substitutions to yield more efficient ZFNs. The IL2RG donor sequence contained a point mutation which introduced a novel *Bsr*BI

site into a 1.5 kb sequence homologous to wild type IL2RG to allow easy detection of HDR. 10% of cells underwent HDR and this was increased to 18% when cells were arrested in the G2/M phase for 30 hrs by addition of vinblastine, (a cell cycle arresting drug). These HDR events were stable for at least one month and corrected cells showed no signs of reduced fitness.

A follow up study expanded on this by using the IL2RG gene as a potential 'safe harbour' to insert several expression cassettes (Moehle et al., 2007). Conventional gene addition is widely used in gene therapy and is usually achieved via random integration of the transgene (Section 1.4). This presents safety concerns due to the possibility of insertional mutagenesis and may result in poor levels of expression due to chromatin-based effects on transgene expression (Kwaks & Otte, 2006). Moehle and colleagues achieved ZFN-mediated gene addition at a specific locus using a donor construct containing 1.5 kb EGFP expression cassette flanked by 750 bp homology arms and ZFNs targeting the IL2RG gene. The EGFP cassette was successfully added into the IL2RG locus at a frequency of 6%. They then constructed a donor with 750 bp homology arms to the IL2RG gene interrupted with a 7,762 bp insert carrying three separate promoter-transcription units, two which encoded the heavy and light chains of the human IgG antibody molecule and the third ecoding a marker. A qualitative PCR-based assay confirmed ZFN and donor treated cells acquired transgene specific sequences in the X chromosome and measurable levels of human IgG was detected in the media harbouring cells exposed to ZFNs and donor construct (0.131 ng/ml) compared to donor only treated cells (0.012 ng/ml).

1.8.3 *Gene Correction and iPs cells*

By correcting the patients' own cells *ex vivo* and then re-introducing them back into the body overcomes the problems of immunological incompatibility between patient and donor cells. The first demonstration of a significant clinical benefit from conventional gene therapy was reported by Cavazzana-Calvo et al., (2000) to treat X-SCID in two infant patients. Bone marrow from the patients was exposed to a retroviral vector carrying the cDNA of the human γc gene and then returned to the patients. After a 10-month follow up period both patients developed normal numbers of T lymphocytes and natural killer cells, which showed evidence of

immunologic function (Cavazzana-Calvo et al., 2000). However, some problems were encountered in later trials. In France, an X-SCID gene therapy trial was carried out on nine patients. Eight of the nine established a working immune system following treatment. However, four developed leukaemia due to random integration of the retrovirus vector and one patient subsequently died. The other three patients survived following chemotherapy treatment and maintained a healthy immune system. Following the death of a patient all X-SCID trials were stopped (Hacein-Bey-Abina et al., 2011). Another disease which has been successfully treated using conventional gene therapy is β-thalassaemia. Sufferers of β-thalassaemia are transfusion-dependant for life as the disease results in a profound decrease in β-globin synthesis. An adult patient with severe β^{E}/β^{O}-thalassaemia received a lentiviral β-globin gene transfer and was transfusion independent at a follow up 21 months post-treatment (Cavazzana-Calvo et al., 2010).

Precise editing of the genome coupled with the availability of stem cells offers a whole new therapeutic opportunity. The development of induced pluripotent stem (iPS) cells has raised prospects for patient-specific therapies (Yamanaka, 2007). Human iPS cells can differentiate into all somatic tissues, can grow to unlimited numbers and can be derived from a variety of somatic cells such as fibroblasts from skin biopsies (Lowry et al., 2008). This strategy was used to correct a mouse model of human sickle cell anaemia with iPS cells derived from autologous skin. Firstly fibroblasts from adult mice were reprogrammed into iPS cells by transducing the cells with virus vectors expressing four genes Oct4, Sox2, Klf4 and c-Myc. The β-globin gene was added by electroporation of cells with a target construct containing the human β-globin gene. Modified iPS cells were allowed to differentiate *in vitro* before being transplanted into affected donor mice after irradiation (Hanna et al., 2007). Treated mice demonstrated stable engraftment for at least 12 weeks post-transplantation. An indicator of sickle cell disease activity and severity is the elevated level of reticulocytes in peripheral blood, which are immature red blood cells (RBCs), reflecting increased production of RBCs to overcome their chronic loss. Mice that had received corrected iPS cells showed a dramatic reduction in reticulocyte count indicating rescue from the disease. In 2011, this technique was transferred to human iPS cells containing the mutation for sickle cell anaemia.

Sebsatiano et al. generated iPS cell lines from fibroblasts of sickle cell anaemia patients and subsequently corrected the genetic disease-causing mutation using ZFN HDR (Sebastiano et al., 2011). They used the OPEN method to design two pairs of ZFNs, one pair 24 bp upstream from the sickle cell mutation and one directly at the mutation site. The donor repair plasmid contained 1.6 kb of the β-globin gene sequence centred approximately on the sickle cell causing mutation E6V. A puromycin resistant gene 82 bp was added downstream of the E6V mutation to allow drug selection of corrected clones. Two silent mutations were also introduced into the donor construct to prevent cleavage by the ZFNs in an attempt to increase repair efficiency by preventing cleavage of the repaired sequence. PCR and Southern blot analysis revealed successful correction of the β-globin gene with targeting efficiencies as high as 37.9% with no detection of off-target events, even in highly related DNA sequences of the δ- and γ- genes. Furthermore, the puromycin gene was introduced 82 bp away from one ZFN site, indicating that ZFNs can mediate cassette insertion even when it is nearly 100 bp away from the DSB.

By correcting a patient's own cells *ex vivo* means there is no requirement for immunosuppressive drugs to prevent rejection and risks associated with off target binding events can be screened for and controlled prior to transplantation. However, some genetic diseases affect organ systems in which *ex vivo* manipulation of target cells is not feasible such as the liver or the lungs. Kathy High's group reported an exciting strategy to overcome this challenge (Li et al., 2011) whereby they restored haemostasis in a mouse model of haemophilia using ZFNs and a donor vector carrying part of wild-type F9 (figure 1.13). Haemophilia B is caused by deficiency of blood coagulation factor IX, encoded by the F9 gene. Most disease causing mutations in humans are found in exons 2-8, therefore, to allow complete coverage of a wide spectrum of mutations they constructed a promoterless donor plasmid containing a partial cDNA of wild type exons 2-8 flanked by a splice acceptor and poly-adenylation sites and homology arms. They designed a ZFN pair to target intron 1 of F9 so that following successful cleavage and HR with the donor, splicing of wild type exon 1 with the introduced exon 2-8 would lead to expression of functionally active factor IX and rescue of the defect caused by most mutations.

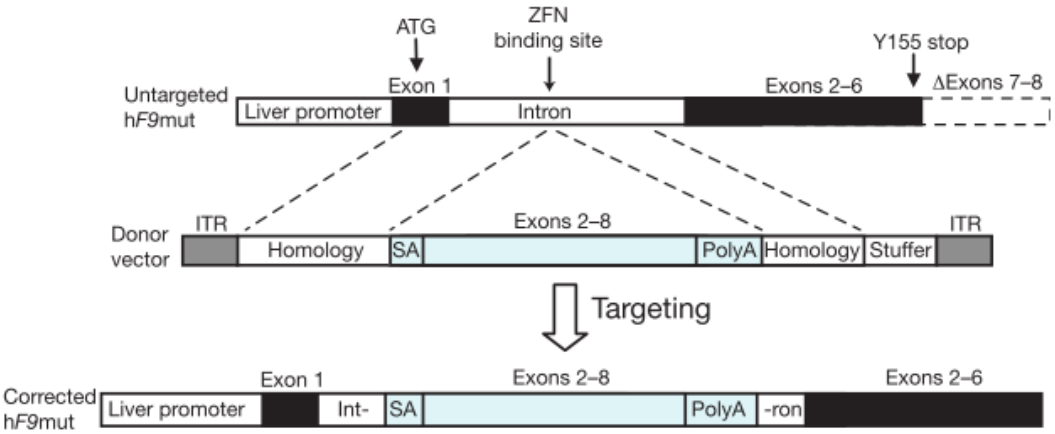


Figure 1.13: Schematic representation of the gene repair strategy using a mini-gene donor to correct F9 gene. ZFN binding site in intron 1 creates a DSB. The donor vector contains partial F9 cDNA (exons 2-8), splice acceptor and polyA sites and homology arms to intron 1. HDR with the donor vector results in correct splicing from endogenous exon 1 into introduced exons 2-8 and production of F9 under the control of the endogenous liver promoter. (Li et al., 2011).

AAV8 expressing ZFNs were injected into the tail vein of mice expressing the human F9 mutation. Cleavage was determined to range from 34 – 47%. Mice were injected intraperitonealy at day 2 of life with AAV8-ZFN and AAV8-donor. DNA from the liver was extracted at week 10 and HDR was detected at a targeting efficiency range of 1-3%. Human factor IX levels in the plasma of mice treated with ZFN and donor averaged 166-354 ng/ml which corresponds to 3-7% of normal levels which was sufficient to correct the haemophilia B phenotype. This strategy overcomes the possibility of over-expression of a protein when introduced as a full length cDNA with its own promoter as is the case with most cDNA addition trials. Furthermore, as it is targeted to a specific location in the genome, the risks of oncogene activation are also reduced as was seen following viral vector delivery of IL2RG to bone marrow cells of X-SCID patients (Hacein-Bey-Abina et al., 2008).

As well as ZFNs, TALENs have also been used to edit the genomes of organisms including the zebrafish *in vivo*. As well as using donor DNA plasmids as repair templates, recent work has shown that single-stranded DNA (ssDNA) can be an effective donor for HDR based genome editing at a ZFN–induced DSB (Chen et al., 2011). Bedell et al. (2012) used TALENs and ssDNA to precisely modify sequences in the zebrafish genome through HDR (Bedell et al., 2012). They introduced a custom-designed *Eco*RV site and a modified loxP sequence into somatic tissue *in*

vivo and went on to show successful germline transmission of both of these markers. Human genes have also been successfully targeted using TALENS. Sun et al. (2012) optimised a pair of TALENs to target the β globin gene in human cells. By varying the N- and C-terminal extensions flanking the central repeat domain and the spacer length between each effector binding element of the DNA substrates they were able to increase the cleavage efficiency of the TALENs. They also discovered that some TALEN architects are capable of binding sequences not preceded by a 5'-T (Sun et al., 2012).

CRISPR/Cas system is still in the early stages of experimentation but some groups have already reported efficient genome editing in zebrafish *in vivo* (Hwang et al., 2013). In this study, ten gRNAs were constructed to target ten endogenous zebrafish genes. These gRNAs were co-injected along with the Cas9 plasmid into cell-one stage zebrafish embryos and the frequency of altered alleles was measured using the T7 endonuclease assay. High frequencies of indel formation were detected at eight of the ten sites in individual embryos tested with a mean of 24.1-59.4%. Two genes which they were unable to alter using TALENs were successfully targeted using their designed gRNAs.

1.9 Off-Target activity

The possibilities of genome editing are endless and the tools to carry out these changes are constantly being developed. In order for customisable nucleases to carry relevance for genetic analysis and clinical application, they must demonstrate strict specificity toward their intended DNA targets. However, complex genomes like ours often contain multiple copies of sequences that are identical or highly homologous to the intended target, leading to off-target activity and cellular toxicity. These off-target events can lead to mutagenesis and cancers if oncogenic genes are switched on. A number of adjustments have been applied to these nucleases in order to prevent off target cleavage and improve specificity.

1.9.1 *FokI heterodimerisation*

ZFN and TALEN cleavage requires homodimerisation of the *Fok*I endonuclease which is inactive as a monomer. An advantage of this is that cleavage requires simultaneous binding of both ZFNs/TALENs to their respective half sites.

However, it also introduces a problem in that protein-protein interactions mediated by the WT *Fok*I cleavage domain are not themselves selective for heterodimer species. Therefore, the expression of any ZFN/TALEN heterodimer (L/R) also yields two side product homodimers (L/L and R/R) which are of no use for gene modification but may limit the safety and effectiveness by cleaving off-target sequences. This phenomenon was demonstrated in a 2006 study whereby two out of three designed ZFNs exhibited toxicity due to homodimer formation of a single ZFN (Beumer et al., 2006). To overcome this problem, both Szczepek et al. (2007) and Miller et al. (2007) modified the dimer interface of *Fok*I by substituting one amino acid for another to yield a progressive improvement in the specificity of heterodimer formation. Using *in silico* approaches, Szczepek et al., (2007) generated compatible *Fok*I cleavage domain variants (RR:DD) by remodelling the dimerization interface of the native enzyme (Szczepek et al., 2007). In Miller's study, the first developmental cycle identified a variant with a glutamic acid to lysine mutation at position 490 (E490K) that exhibited modest preference for heterodimerisation with the WT *Fok*I domain. Subsequent cycles yielded variant cleavage domains with the double mutation E490K:I538K and Q486E:I499L (+ and -) which exhibited strong preference for heterodimerisation in a GFP reporter assay (figure 1.14). When tested for frequency of gene correction R+/L- stimulated gene correction at a rate of 1.98%. Reversal of the ZFN cleavage domain pairings to R-/L+ yielded a 1.71% correction rate. Both modified *Fok*I combinations were more efficient than WT *Fok*I which was 1.27% (Miller et al., 2007). Another group sought to enhance ZFNs by integrating the *Sharkey* framework via directed evolution. *Sharkey* consists of two amino acid mutations, S418P and K441E, which were shown to increase cleavage both *in vitro* and *in vivo* (Guo et al., 2010) and in some cases >15 fold increase in activity compared to traditional ZFNs was observed.

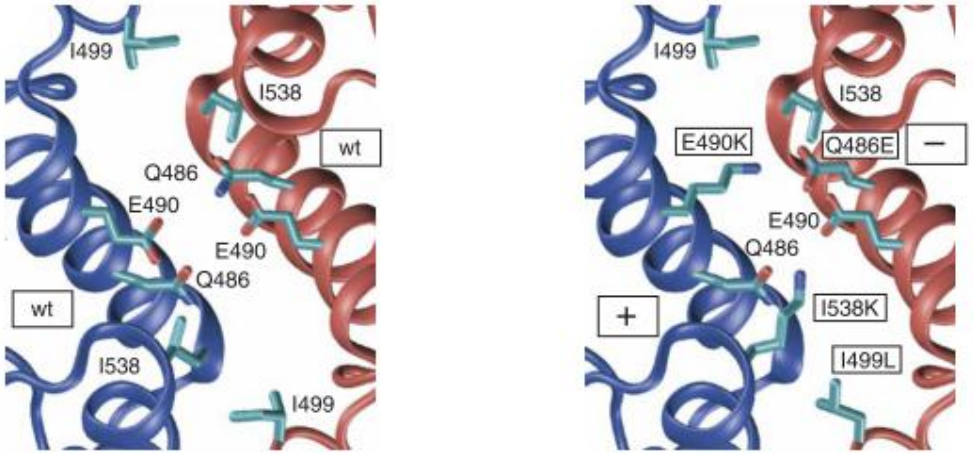


Figure 1.14: *Fok*I active domain. Left shows WT structure. Right shows introduced mutations KK and EL (Miller et al., 2007).

1.9.2 ***Nickases***

The discovery that nicked DNA could stimulate HR without activating the NHEJ repair pathway (McConnell Smith et al., 2009) led to the development of ZFNickases (Kim et al., 2012). Rather than producing a DSB, nickases cut the DNA on one strand only which means the chromosome is still intact so NHEJ is not activated. In this study, the group introduced a mutation (Asp450 to Ala) into the active site of the *Fok*I domain of one ZFN so that it could no longer cleave. Subsequently, upon dimerisation at the intended DNA sequence only one ZFN would cleave to produce a single stranded break. Consequently this approach leads to fewer off-target events than conventional DSB induction, however, the frequency of HDR with nickases remains lower than those achieved with traditional ZFNs.

In the CRISPR/Cas9 system, the Cas9 protein has been modified so that it only cleaves one strand of DNA and so acts like a nickase (Jinek et al., 2012). Point mutations were introduced into the codons of the catalytic residues of the HNH and RuvC-like domains of Cas9 and the plasmids were purified. Incubation of variant Cas9 with native plasmid DNA showed that RNA guided mutant Cas9 proteins yielded nicked circular plasmids whereas WT Cas9 yielded linear DNA products.

1.10 Aims:

The aim of this body of work was to develop an efficient process to correct CF-causing mutations using nuclease-directed HDR. We first set out to correct the most common mutation, ΔF508, using existing ZFNs, which cleave in intron 9, and donor plasmids made in this thesis. The rate of HDR was successful, but initially determined to be of relatively low efficiency, so two new approaches for correction were explored. First, to take advantage of the position of our ZFNs, we investigated the HDR of a mini-gene construct into the ZFN cut site. As a proof of principle we introduced a seven base pair tag into this site. Following successful incorporation of the tag, we sought to repair almost 80% of mutations using a mini-gene construct containing WT exons 10-24, splice acceptor and polyA sites and a GFP tag. Secondly, the ability to cleave closer to the ΔF508 mutation was determined using the latest gene editing tool, the CRISPR/Cas9 system. Two sites were targeted with CRISPR guide RNAs and used with existing donor plasmids to induce HDR.

1.11 Objectives:

Chapter 3:

Repair the ΔF508 mutation in CFTE cells using existing ZFNs targeting intron 9 and donor repair plasmids pITR-donor and pITR-donor-XC

Chapter 4:

Introduce unique seven base pair tag into the ZFN cut site to prove the principle of mini-gene recombination

Design and construct a mini-gene donor containing WT exons 10-24 to correct >80% of CF-causing mutations

Chapter 5:

Design and synthesise CRISPR gRNAs to target CFTR and measure cleavage efficiency

Induce HDR using CRISPR gRNAs and appropriate donor plasmids

2 Materials and Methods

2.1 Cell Lines

16HBE14o-: Human bronchial epithelial cells, endogenous wildtype CFTR (Cozens et al., 1994)

CFTE29o-: Cystic fibrosis tracheal epithelial cells, homozygous for ΔF508 (Chastre et al., 1991)

Cell lines were generous gifts from Prof. D. Gruenert (California Pacific Medical Centre Research Institute, San Francisco, CA).

2.2 Cell Culture Reagents

The following reagents were used in growth and maintenance of cell lines:

2.2.1 *Cell Culture Medium*

Growth medium (Epithelia)

500 ml Eagle's minimum essential medium (Earle's salts, 25mM HEPES and sodium bicarbonate, without L-glutamine; Sigma-Aldrich)
50mM Foetal calf serum (FCA; PAA Laboratories)
5 ml Penicillin-Streptomycin (100 units/ml penicillin and 100 µg/ml Streptomycin final concentrations; Invitrogen)
5 ml L-glutamine (2mM final concentration; Invitrogen)

Seeding medium (Epithelia)

500 ml Eagle's minimum essential medium (Earle's salts, 25mM HEPES and sodium bicarbonate, without L-glutamine; Sigma-Aldrich)
50 ml Foetal calf serum (FCS; PAA Laboratories)
5 ml L-glutamine (2mM final concentration; Invitrogen)

Transfection medium 1

500 ml Eagle's minimum essential medium (Earle's salts, 25mM HEPES and sodium bicarbonate, without L-glutamine; Sigma-Aldrich)

5 ml L- glutamine (2mM final concentration; Invitrogen)

50 ml Foetal calf serum (FCA; PAA Laboratories)

Transfection medium 2

500 ml Eagle's minimum essential medium (Earle's salts, 25mM HEPES and sodium bicarbonate, without L-glutamine; Sigma-Aldrich)

5 ml L- glutamine (2mM final concentration; Invitrogen)

2.2.2 ***Other reagents***

Trypsin-EDTA solution (2.5 g/L Trypsin and 0.38 g/L EDTA; Invitrogen)

1X PBS (137mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4,

1.47 mM KH_2PO_4, pH of 7.4)

2.3 Bacterial Strains

All bacterial strains used in this work are listed in Table 2.1.

Strain	Genotype and Source
DH5α	F^- φ80*lacZ*ΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *nd*A1 *hsd*R17(r_k^-, m_k^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1λ^- (Invitrogen)
XL10	Tet^rΔ (mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 *end*A1 supE44 thi-1 *rec*A1 *gyr*A1 *lac* The [F' proAB lac^qZΔM15 Tn10 (Tet^r) Amp Cam^r] (Stratagene)

Table 2.1. Bacteria, genotypes and sources

2.4 Bacterial Growth Media and Antibiotics

2.4.1 *Growth Media*

LB Agar (Lennox L agar)	LB Broth (Lennox L Broth)
9.14 g/L enzymatic digest of casein	10 g/L enzymatic digest of casein
4.57 g/L yeast extract	5 g/L yeast extract
4.57 g/L sodium chloride	5 g/L sodium chloride
13.72 g/L bacteriological agar	

Both media were prepared from tablets (Sigma-Aldrich). Each tablet was dissolved in 50 ml of deionised water and autoclaved (15 minutes, 121 °C, 15 psi).

2.4.2 *Antibiotics*

Ampicillin

A 100 g/ml stock solution of ampicillin (Sigma-Aldrich) was prepared by dissolving 1 g of powder in 10 ml of deionised water and filter sterilising through a 0.2 µm pore. Ampicillin was used at a final concentration of 100 µg/ml.

Kanamycin

Kanamycin (10 mg/ml; Sigma-Aldrich) was used at a final concentration of 30 µg/ml. Prepared as per ampicillin.

Tetracycline

Tetracycline (5 mg/ml; Sigma-Aldrich) was used at a final concentration of 20 µg/ml. Prepared as per ampicillin.

2.5 Buffers, solutions and drugs

Phosphate buffered saline (PBS)

10X PBS (138 mM NaCl, 2.7 mM KCl; Sigma-Aldrich) was diluted 1:10 with autoclaved deionised water under a laminar flowhood to make 1X PBS.

Tris-acetate-EDTA (TAE)

50X TAE comprised 242 g of Trizma base, 18.6 g EDTA and 57.1 ml glacial acetic acid made up to 1 L with deionised water and adjusted to pH 7.8. 1X TAE was prepared by diluting the 50X solution with deionised water.

X-gal

A 100 mg/ml stock solution of X-gal (5-bromo-4chloro-3-indolyl-beta-D-galactopyranoside; Sigma-Aldrich) was prepared by dissolving 1 g of X-gal in dimethylformamide and making up to a volume of 10 ml. X-gal was used at a final concentration of 80 µg/ml.

IPTG

A 1 M stock solution of IPTG (Isopropyl β-D-1-thiogalactopyranoside; Sigma-Aldrich) was prepared by dissolving 2.38 g of IPTG in autoclaved deionised water and making up to a volume of 10 ml. IPTG was used at a final concentration of 20 mM.

Calcium Chloride ($CaCl_2$)

A 1 M stock solution of $CaCl_2$ (Calcium Chloride-Sigma-Aldrich) was prepared by dissolving 147 g of $CaCl_2$ in autoclaved deionised water and making up to a volume of 500 ml. $CaCl_2$ was used at a final concentration of 0.3mM.

2 X HBS

2 X HBS (280mM NaCl, 1.5mM Na_2HPO_4, 50mM HEPES; Sigma Aldrich) was made up to a volume of 500 ml with autoclaved deionised water and pH was adjusted to 7.1.

Benzonase buffer

Bezonase buffer (50mM Tris HCl pH 8, 1mM $MgCl_2$; Sigma-Aldirch, 0.1 mg/ ml BSA; New England Biolabs was made up to a volume of 1 ml with autoclaved deionised water.

Benzonase Nuclease

Benzonase nuclease (250 units/µl; Sigma-Aldrich) was used at 50U/ml.

2.6 Molecular biology reagents

All reagents used for molecular biology are listed in table 2.2.

Reagents	Source
10X buffer 1/2/3/4 for restriction digests	New England Biolabs
10X ligation buffer	New England Biolabs
Lambda DNA/EcoRI+HindIII marker	Fermentas
MassRuler Express Reverse DNA ladder	Fermentas
MassRuler Express LR Forward DNA ladder	Fermentas
6X blue loading dye	Fermentas
Tritrack	Fermentas
Lipofectamine 2000	Invitrogen
Platinum Taq Polymerase	Invitrogen
Taq Polymerase	Promega
Platinum Pfx Polymerase	Invitrogen
Restriction enzymes	New England Biolabs
RNA Loading buffer	Sigma
T4 DNA Ligase	New England Biolabs

Table 2.2. Molecular biology reagents and sources

2.7 Chemical Reagents

Chemical reagents not listed in Table 2.3. were supplied by Sigma-Aldrich

Chemical	Sources
Glacial acetic acid	BDH
Trizol reagent	Molecular Research Centre
Ethidium Bromide	Sigma- Aldrich
Boric Acid	Sigma-Aldrich
Saponin	Sigma-Aldrich
Formalin solution	Sigma-Aldrich
Goat serum donor herd	Sigma-Aldrich

Table 2.3. Chemical reagents and sources

2.8 Other Reagents

Other reagents used in this work are listed in Table 2.4. All plasticware was supplied by Starstedt unless otherwise stated.

Consumable	Source
20 ml syringes	BD
Disposable culture tube 12 x 75 mm	Fischer Scientific
Uvette	Eppendorf
Glass coverslips	Corning
Microscope slides	Corning
Parafilm	American Can Co.
Scalpels	Swann-Morton
Syringe filters (pore size 0.2 µm)	Whatman

Table 2.4. Consumables and sources

2.9 Plasmids

2.9.1 *Plasmids provided*

Commercially obtained plasmids and constructs that were not produced within this work are listed in Table 2.5.

Plasmids	Sources
pGL3-Control	Promega
phRL-CMV	Promega
pcDNA3.1(+)	Invitrogen
pUC19	Invitrogen
pWhitescript	Stratagene
pEGFP-N1	Clontech
hCas9	Addgene
hCas9 D10A	Addgene

Table 2.5. Plasmids and respective sources

2.9.2 *Plasmids constructed*

Constructs created over the course of this project are listed in Table 2.6.

Plasmid	Details
pITR-donor	pAAV containing *cftr* DNA from HBE cells
pITR-donor-XC	pAAV containing *cftr* DNA with restriction enzyme sites added
pUC-donor	pUC19 containing 4.3 kb segment of *cftr*
pUC-donor-XC	pUC19 containing 4.3 kb segment of *cftr* with added restriction sites
pITR-CFTE	pAAV containing *cftr* DNA from CFTE cells
pITR-CFTE-*Nhe*I	pITR-CFTE with added *Nhe*I site
pUC-CFTE-NheI-1kb	pUC19 containing 1kb fragment of pITR-CFTE-*Nhe*I
pUC-CFTE-*Nhe*I-1.5 kb	pUC19 containing 1.5 kb fragment of pITR-CFTE-*Nhe*I

pUC-mini gene	pUC19 containing appropriate splice acceptor, exons 10-24 CFTR cDNA WT, T2A linker and GFP with Poly A tail.
pEX-A (pGuide)	pEX-A plasmid containing the U6 promoter and *BseR*I sites for cloning of gRNAs
pCEP-CFTR	CFTR cDNA in a pCEP backbone under the control of a CMV promoter
CRISPR ex10	gRNA targeting exon10 of CFTR in pGUIDE backbone
CRISPR in9	gRNA targeting intron 9 of CFTR in pGUIDE backbone

Table 2.6. Plasmid constructs and respective details

2.10 Kits

Kits used in this work and sources are listed below in Table 2.7.

Kits	Sources
QIAprep Miniprep Kit	Qiagen
Qiagen Plasmid Maxi Kit	Qiagen
Qiagen Plasmid Plus Purification Kit	Qiagen
StartaPrep DNA Gel Extraction Kit	Stratagene
QuikChange XL Site-Directed Mutagenesis Kit	Stratagene
Dual-Luciferase Reporter Assay System	Promega
GoTaq PCR Core System	Promega
Improm-II Reverse Transcription System	Promega
Virakit	Virapur
Topo TA Cloning	Invitrogen
QIAquick PCR purification Kit	Qiagen
Platinum Taq DNA Polymerase	Invitrogen
PfuTurbo polymerase	Invitrogen

Table 2.7. Kits and Sources

2.11 Oligonucleotides

Name and Purpose c=cloning, s=sequencing, sdm=site-directed mutagenesis p=primer Underlined=restriction site	(5'->3') Sequence
CFTR- Fwd- *Bst*EII (c)	TGACTAGGTAACCGAGCCTTTTCTTTTTCATTG
CFTR-Rev-*Mlu*I (c)	CAACGCGTAGGCCCTAACTTTCATTCTTG
Fwd SD *Xho*I (sdm)	GATTATGGGAGAACTCGAGCCTTCAGAGGG
Rev SD *Xho*I (sdm)	CCCTCTGAAGGCTCGAGTTCTCCCATAATC
Fwd SD *Cla*I (sdm)	GTGTTTCCTATGATGAATATCGATACAGAAGCGTCATCAAAG
SD *Cla*I Rev (sdm)	CTTTGATGACGCTTCTGTATCGATATTCATCATAGGAAACAC
Fwd SDM 1 MUT (sdm)	CTTGACCAGCGTGACTGCAAGCTTGATGTAGGAGAAG
Rev SDM 1 MUT (sdm)	CTTGACCAGCGTGACTGCAAGCTTGATGTAGGAGAAG
Fwd SDM 4 MUT (sdm)	CTCTTCTGCTTCTCCTATAATAAGCTTGCAGTCACGCTG
Rev SDM 4 MUT (sdm)	CAGCGTGACTGCAAGCTTATTATAGGAGAAGCAGAAGAG
MluI/BstEII linker f (c)	CTAGAGTCAGCACGCGTGCTGTACGATGGTAACCCGGACCGTCACCGA
MluI/BstEII linker r (c)	AGCTTCGGTGACGGTCCGGGTTACCATCGTACAGCACGCGTGCTGACT
LITR CFTR (s)	TGGAGCGAACGACCTACACC
RITR CFTR (s)	TGTACTAGGAGCTGGAGATG
RITR F1 CFTR (s)	CACTAAATCGGAACCCTAAA
RITR CF CFTR (s)	TTAAGAGGAAAAGACTTCAG
Fwd-CFTR-NHEJ (p)	ATCATGTGCCCCTTCTCTGT
Rev CFTR-NHEJ (p)	TGCTTTGATGACGCTTCTGTAT
Fwd GAPDH (p)	CGCGGGGCTCTCCAGAACATCAT
Rev GAPDH (p)	GCCAGCCCCAGCGTCAAAGGTG

Fwd RT-PCR (p)	AGAACTGAAACTGACTCGGAAGG
Rev RT-PCR (p)	CTATATTCATCATAGGAAACACCAAAG
FP-i9.1	ACTTGCTTTGCCATTAACAGA
RP i10.1	GTCGTGTTCTAGGTATATGACC
Donor HDR (s)	CTAATGGTGATTATGGGAGAAC
Fwd gDNA (p)	ACTTGCTTTGCCATTAACAGA
Rev gDNA (p)	GCAGATCAATGCTCATTCCA
FAM-Fpe10.2 (p)	FAM-CAGTTTTCCTGGATTATGCCTGGCA
RP i10.1 (p)	AGTTGGCATGCTTTGATGACGCT
Fwd NHEJ Probe (p)	FAM-TGCATAGCAGAGTACCTGAAACA
Rev NHEJ (p)	CAAATCACGCTCAGGATTCA
Fwd_*Eco*RI (p)	GTCAGGACATATAACATATTTAAAC
Rev_*Nhe*I (p)	ATGGCAGCTAGCAAGCAGAAGTGTATATTC
Fwd_*Nhe*I (p)	CTAGTCGCTAGCAAGGATGATAATTGGGGGCAAGTG
Rev_*Btg*I (p)	TATCTAATCCACGGTTTGC
Fwd_*Kas*I (p)	CTAGTCGGCGCCGTCAGTAGTATTATCTTT
Rev_*Xba*I (p)	ATGGCATCTAGATCACAATTTTACCCCTCTAA
Fwd_*Nhe*I specific primer (p)	ACACTTCTGCTTGCTAGC
Rev_*Nhe*I-specific primer (p)	CAATTATCATCCTTGCTAGC
CMW_sequence primer (s)	GTACGGGCCAGATATAC
*Fok*I_sequence primer (s)	GGCAAACAACAGATGGCTGG
Fwd_NheI assay primer (p)	CTAGCCTGGGTGACAGAGTG
Fwd_eGFP seq primer (s)	CACTCTCGGCATGGACGAG

Rev_eGFP (s)	CTGAAGCACTGCACGCCG
Fwd_eGFP/pA	ATCAGATCTCGTGAGCAAGGGCGAGGAGC
Rev_eGFP/pA	AGCTAGCCCACAACTAGAATGCAGTG
Fwd_cDNA primer (c)	GTTTTCCTGGATTATGCCTGGC
Rev_cDNA primer (c)	GTCACTCGAGAAGCCTTGTATCTTGCACCT
Fwd_SA primer (c)	GTCAGCTAGCCTAATAATGATGGGTTTTATTTCCAG
Rev_SA primer (c)	CTCTTCTAGTTGGCATGCTTTG
T2A(+) (c)	TCGAGGGCAGAGGAAGTCTTCTAACATGCGGTGACG TGGAGGAGAATCCCGGGCCTA
T2A(-) (c)	GATCTAGGCCCGGGATTCTCCTCCACGTCACCGCAT GTTAGAAGACTTCCTCTGCCC
Fwd_recombination primer (p)	GTAGGCTATATGTATCTGTGAG
Rev_ex11 primer (p)	CGAGCAACTGGAGGTGAGTCAC
Fwd_GFP_HDR_(1) primer (p)	CACTGCATTCTAGTTGTGGTTTG
Rev_in10 primer (p)	GTCGTGTTCTAGGTATATGACC
Fwd_GFP_HDR_(2) primer (p)	CACATGGTCCTGCTGGAGTT
CRISPR IN9 (+) (c)	ATAATGACCTAATAATGATGT
CRISPR IN9 (-) (c)	ATCATTATTAGGTCATTATCG
CRISPR EX10 (+) (c)	GAGAACTGGAGCCTTCAGAGT
CRISPR EX 10 (-) (c)	TCTGAAGGCTCCAGTTCTCCG

Table 2.8. Names and sequences of oligonucleotides used in this project

2.12 Equipment

All equipment used in this work is listed in Table 2.9.

Equipment	Sources
Biophotometer	Eppendorf
Bright-Line haemocytometer	Sigma-Aldrich
EasyCast gel tray (Owl)	World Precision Instruments
IX70 inverted microscope	Olympus
Luminometer	Berthold Technologies
Mastercycler Gradient	Eppendorf
Orbital shaker	Stuart Scientific
pH meter	Hanna
Lightcycler	Roche
Fluorescent Microscope	Martina Scallan's Lab
Immuno-fluorescent Microscope	Martina Scallan's Lab
Genetic Analyser 3130 (Applied Biosystems).	Collette Hand Lab, CUH
Zeiss Laser Scanning Microscope 510	Martina Scallan's Lab

Table 2.9. Equipment and sources

2.13 Software

The following software was used over the course of this project:

Olympus analysis software

UVIPhotoMW software

DNASTAR Lasergene sequence analysis software

Primer X (www.bioinformatics.org/primerx/)

Ensembl genome browser (www.ensembl.org)

Pubmed (http://www.ncbi.nlm.nih.gov/pubmed)

ClustalW (http://www.genome.jp/tools/clustalw/)

2.14 Cell culture protocols

2.14.1 *Subculturing cells*

Cells were passaged when approximately 80% confluent. Cells were washed with 1X PBS and incubated with 2 ml of 0.25 % trypsin-EDTA. Detached cells were re-suspended in 8 ml of growth medium and transferred to a 15 ml tube and centrifuged at 1000 rpm for 5 minutes at room temperature. The cell pellet was re-suspended in 6 ml of growth medium and 1 ml used to seed a new flask. Growth medium was added to make up a final volume of 20 ml (75 cm^2 flask) or 25 ml (175 cm^2 flask).

2.14.2 *Cell quantification and seeding density*

10 µl of re-suspended cells was transferred to each chamber of a Bright- Line haemocytometer. The number of cells within each of the 25 1 mm squares in both chambers (50 in total) was counted under the microscope. If the number of cells in the 50 squares was greater than 500, the sample was diluted and recounted. The number of cells per millilitre of suspension was determined according to the formula:

$$[\text{Cells/ml} = \text{average count per square} \times \text{dilution factor} \times 10^4]$$

For 6-well plates, 300,000 cells were seeded and the volume made up to 2 ml with growth medium. For 100 cm dishes, 4×10^6 cells were seeded and the volume made up to 10 ml with growth medium.

2.15 Bacterial protocols

2.15.1 *Inoculation of LB broth*

Individual colonies were transferred from agar plates to 5 ml (for Miniprep) or 50 ml (for Midiprep) or 150 ml (for Maxiprep) of LB broth containing the relevant antibiotic using a disposable inoculation loop.

2.15.2 *Bacterial transformation*

50 µl aliquots of DH5α competent cells were thawed on ice. 1 to 10 ng (maximum 5 µl) of DNA was added to a 50 µl aliquot of cells and incubated on ice for 30 minutes. The cells were incubated at 42 °C for 2 minutes, then 2 minutes on ice. 100 µl of LB broth was added to the cells and transferred to an orbital shaker at 200 rpm and 37 °C for 1 hour. 100 µl of the cell suspension was plated on a LB agar plate containing the appropriate selective antibiotic. Plates were incubated at 37 °C for 16-20 hours.

2.16 DNA and RNA isolation

2.16.1 *Plasmid extraction (Qiagen Plasmid Plus Purification Kit)*

Plasmid DNA for transfection and ligations was extracted from *E. coli* DH5α cells using a Plasmid Plus Midi Kit as per the manufacturer's instructions. 50 ml of LB broth containing the appropriate antibiotic was inoculated with the transformed bacteria and incubated at 37 °C on an orbital shaker at 200 rpm for 16 hours. The culture was centrifuged at 6,000 x rpm at 4°C for 15 minutes. The pellet was re-suspended in 4 ml of RNase Buffer P1. Four millilitres of lysis Buffer P2 was added, mixed by inversion and incubated at room temperature for 5 minutes. Four millilitres of Buffer S3 was added, mixed by inversion, and filtered through a QIAfilter cartridge to remove proteins and genomic DNA. The filtrate was transferred to a fresh 50 ml tube and 2 ml of Buffer BB was added and mixed by inversion. Lysate was transferred to a Midi spin column with a tube extender attached and solution was drawn through the column using a vacuum. DNA was

washed with 700 µl of Buffer ETR, followed by further washing with 700 µl of Buffer PE and centrifuged at 13,000 x rpm for 1 minute. DNA was eluted from the column by adding 200 µl of Buffer EB, incubating at room temperature for 5 minutes and centrifuged as before. DNA obtained was stored at 4 °C.

2.16.2 *Plasmid extraction (Qiagen Miniprep Kit)*

Plasmid DNA for restriction analysis was obtained from individual bacterial colonies using a QIAprep Miniprep Kit according to the manufacturer's guidelines. Individual colonies were grown overnight in 5 ml of LB broth containing a suitable plasmid selective antibiotic at 37 °C on an orbital shaker at 200 rpm. 1 ml of bacterial suspension was centrifuged at 6,000 x rpm for 5 minutes at room temperature. The pellet was re-suspended in 250 µl of Buffer P1 containing RNase and cells lysed with 250 µl of Buffer P2. Lysis was terminated after 5 minutes with 350 µl of Buffer N3 and mixture centrifuged at 13,000 x rpm for 10 minutes at room temperature. The supernatant was applied to a QIAprep spin column and centrifuged at 13,000 x rpm for 1 minute. 750 µl of Buffer PE was added to the column which was centrifuged as before. DNA was eluted from the column by adding 50 µl of deionised water, incubating at room temperature for 5 minutes and centrifuging at 13,000 x rpm for 1 minute. DNA obtained was stored at 4 °C.

2.16.3 *Plasmid extraction (Qiagen Maxiprep Kit)*

Plasmid DNA for virus production was obtained from individual colonies grown for 16 hrs in 150 ml LB broth containing the appropriate antibiotic. Cells were centrifuged at 6000 x rpm for 15 minutes at room temperature. The supernatant was discarded and the cell pellet re-suspended in 10 ml of RNase Buffer P1. Cells were lysed with 10 ml Buffer P2. 10 ml of Buffer P3 was added after 4 minutes incubation at room temperature. Lysate was poured into a QIAfilter Cartridge and incubated at room temperature for 10 minutes. A HiSpeed Maxi Tip was equilibrated by adding 10 ml QBT Buffer and allowing column to empty by gravity. Lysate was filtered through the pre-equilibrated HiSpeed Maxi Tip. QIAGEN-Tip was washed with 60 ml Buffer QC and DNA eluted with 15 ml Buffer QF. DNA was precipitated by adding 10.5 ml room-temperature isopropanol, mixed and

incubated for 5 minutes at room temperature. A QIAprecipitator Maxi Module was attached to the outlet nozzle of a 30 ml syringe and the elute/isopropanol mixture added and filtered through under constant pressure. 2 ml of 70% ethanol was added to the syringe and filtered through to wash the DNA. The QIAprecipitator was attached to a 5 ml syringe and 1 ml of Buffer TE was added to the barrel. The DNA was eluted into a 1.5 ml collection tube and stored at 4 °C.

2.16.4 *Measuring DNA concentration*

The Biophotometer spectrophotometer was calibrated with a 1:10 dilution of Buffer TE before measuring the sample. DNA was diluted 1:10 (10 µl + 90 µl deionised water) and transferred to a cuvette. The dilution values were entered into the spectrophotometer and absorbance of the sample at 260 nm and 280 nm measured. An absorbance value of 1.0 at a wavelength of 260 nm indicates a DNA concentration of 50 µg/ml. The spectrophotometer automatically calculates concentration and adjusts for dilution. DNA purity was assessed by dividing absorbance at 260 nm by absorbance at 280 nm. A value between 1.8 and 2 indicated an acceptable level of purity.

2.16.5 *RNA extraction*

Cells were washed with PBS then re-suspended/lysed with 1 ml Trizol reagent. The lysate was transferred to a 1.5 ml tube and incubated for 5 minutes at room temperature. 200 µl of chloroform was added and the tube shaken vigorously by hand for 15 seconds, incubated for 2.5 minutes at room temperature, then centrifuged at 12,000 x g for 15 minutes. The clear upper phase was removed to a fresh 1.5 ml tube, and the RNA precipitated with the addition of 750 µl of isopropanol. After 10 minutes incubation, the tube was centrifuged at 12,000 x g for 10 minutes and the supernatant discarded. 1 ml of 75 % ethanol was added and the tube centrifuged at 7,600 x g for 5 minutes. The supernatant was discarded and the RNA pellet allowed to air- dry for 5 minutes. RNase free water was added and the RNA re-suspended by heating at 55 °C for 5 minutes. RNA was stored at -80 °C.

2.16.6 *Measuring RNA concentration*

The Biophotometer spectrophotometer was calibrated with 100 µl of deionised water before measuring the sample. RNA was diluted 1:10 (10 µl RNA + 90 µl deionised water) and transferred to a cuvette. The dilution values were entered into the spectrophotometer and absorbance of the sample at 260 nm and 280 nm measured. An absorbance value of 1.0 at a wavelength of 260 nm indicates an RNA concentration of 40 ng/ml. The spectrophotometer automatically calculates concentration and adjusts it for dilution. RNA purity was assessed by dividing absorbance at 260 nm by absorbance at 280 nm. A value between 1.8 and 2 indicated an acceptable level of purity.

2.17 DNA and RNA manipulation

2.17.1 *Restriction digests*

In a typical digest, 1 µg of plasmid DNA was cleaved by combining it with 5-10 units of each required enzyme, 2 µl of the appropriate 10X buffer and deionised water to a final volume of 20 µl in a 0.5 ml tube. The digest was incubated at 37 °C for 1-2 hours.

2.17.2 *Preparing DNA agarose gels*

To prepare 1 % agarose gels, 1 g of agarose was mixed with 100 ml of 1X TAE buffer in a conical flask and heated in a microwave until the agarose had completely dissolved. When the flask was cool to touch, 8 µl of 10 mg/ml ethidium bromide was added and the agarose was poured into an EasyCast gel tray containing a comb to form wells and allowed to set for 20 minutes. The comb was removed, the tray placed in the electrophoresis chamber and 1 X TAE added to cover the gel.

2.17.3 *Loading and running agarose gels*

Up to 20 µl of DNA was mixed with 6X Tritrack loading dye (20 µl DNA + 4 µl dye). The electrophoresis chamber was attached to a power supply which was set at 120 Volts. When the yellow band reached the end of the gel, the tray was removed from the electrophoresis chamber and the bands viewed using a gel documentation system. Images were acquired using UVIPhotoMW software.

2.17.4 *RNA gel*

1 % TAE gels were prepared as previously described in section 2.17.2, 15 µl of RNA was mixed with 5 µl of RNA loading buffer and heated at 65 °C for 15 minutes. The sample was then cooled for 10 minutes on ice and loaded onto the gel. The gel was run and photographed using a gel documentation system as described in section 2.17.3.

2.17.5 *Gel purification*

The agarose gel was viewed on top of an ultraviolet transilluminator and the band of interest excised with a scalpel blade and transferred to a 1.5 ml tube. A StrataPrep DNA Gel Extraction Kit was used to isolate the DNA according to the manufacturer's guidelines. 300 µl of DNA extraction buffer was added for every 100 µg of gel slice and the sample heated at 50 °C for 10 minutes. The melted gel was transferred to a microspin cup and centrifuged at 13,000 x rpm for 1 minute at room temperature. DNA remained bound to the fibre matrix within the cup while the waste solution passed through the matrix to a receptacle tube and was discarded. The microspin cup was washed with 750 µl of wash buffer and centrifuged as before. The flow-through was discarded, the microspin cup centrifuged again to remove any trace of ethanol and transferred to a fresh receptacle tube. 50 µl of deionised water was added to the fibre matrix and incubated for 5 minutes. The DNA solution was collected by centrifugation at 13,000 x rpm for 1 minute at room temperature.

2.17.6 *Annealing oligonucleotides*

5 µl each of sense and anti-sense single stranded oligonucloetides (1 µg/µl) were added to 85 µl of DEPC and 5 µl of ligase buffer, heated at 95 °C for 5 minutes and allowed to cool to room temperature.

2.17.7 *Ligation*

Digested plasmid and insert DNA were mixed at different ratios to a total volume of 8 µl (1:3, 1:1, 3:1). 2 µl of 10 X ligation buffer and 1 µl (400 units) of T4 DNA ligase were added to each and the reaction incubated at room temperature for 15 minutes.

2.17.8 *De-phosphorylation*

Following digestion of plasmid, it was sometimes necessary to de-phosphorylate the backbone prior to ligation. 16 µl of plasmid backbone was added to 2 µl Antarctic phosphatase buffer and 2 µl Antarctic phosphatase and incubated at 37 °C for 30 mins followed by 20 mins at 65 °C to inactivate the enzyme.

2.17.9 *Golden gate cloning*

Plasmid pEX-A and CRISPR gRNA oligos were digested and ligated simultaneously in one reaction. 150 ng of plasmid was mixed with 15/150 ng CRISPR gRNA oligo, 1 µl *Bse*RI, 1 µl T4 ligase, 2 µl ligase buffer and dH_2O to a total volume of 20 µl. The reaction was incubated at 37 °C for 5 min, then 16 °C for 10 min for 10 cycles, followed by 30 mins at 37 °C and finally 20 mins at 80 °C to heat inactivate the enzymes. XL-Blue bugs were transformed with 2 µl of the reaction.

2.17.10 *Site-directed mutagenesis*

Mutations were introduced into plasmid DNA using the QuikChange XL Site-Directed Mutagenesis Kit. All primers were designed using PrimerX and ordered online from MWG Biotech (www.mwg-biotech.com) The reaction components were mixed together in thin-walled PCR tubes in the following order: 5 µl 10X reaction buffer, 1 µl template DNA, 1.25 µl primer 1, 1.25 µl primer 2, 1 µl dNTP mix and autoclaved deionised water to a final volume of 49 µl. 1 µl of *PfuTurbo*

polymerase (2.5 units) was added, and the tubes transferred to a Mastercycler Gradient thermalcycler programmed as shown in table 2.10.

Cycles	**Temperature**	**Time**
1	95 °C	1 minute
12	95 °C 55 °C	30 seconds 1 min/kb plasmid length
1	68 °C	5 minutes

Table 2.10. Thermalcycler guidelines for site-directed mutagenesis

DNA was digested by adding 1 µl of *Dpn* I and incubating at 37 °C for 1 hour. Forty-five micro-litres of XL-10 ultracompetent *E.coli* cells were thawed on ice and treated with 2 µl of β-mercaptoethanol. After adding 1 µl of DNA, the cells were incubated on ice for 30 minutes. The cells were heat-shocked in a 42 °C heat block for 2 minutes and then incubated on ice for 2 minutes. 450 µl of pre-heated LB broth was added and the tubes places on an orbital shaker in an incubator at 200 rpm and 37 °C for 1 hour. The transformed bacteria were spread on LB agar plates containing the appropriate selective antibiotic.

A control mutagenesis supplied with the kit allowed for the determination of reaction efficiency. Re-activation of the β-galactosidase gene in pWhitescript through site-directed mutagenesis using the supplied primers, could be screened for by growing transformed cells on LB containing X-gal and IPTG. Colonies containing the corrected β-galactosidase gene appear blue as β-galactosidase cleaves X-gal resulting in an insoluble blue product. Colonies containing the original pWhitescript will appear white. Comparing numbers of blue and white colonies gives an indicator of the efficiency of the mutagenesis.

2.17.11 ***Reverse Transcription***

Reverse transcription of mRNA was performed using the ImProm-II Reverse Transcription System. 1 µg of RNA extract and 1 µl of Oligo$(dT)_{15}$ (0.5 µg) primer were mixed and heated at 70 °C for 5 minutes and incubated on ice for 5 minutes. The remaining components (4 µl 5X reaction buffer, 3.2 µl $MgCl_2$ (25mM stock), 1

µl dNTPs (40 mM stock), 1 µl reverse transcriptase) were added to a final volume of 20 µl. The reaction was incubated at 42 °C for 1 hour. The reverse transcriptase was inactivated by heating at 70 °C for 15 minutes. This product then served as template for PCR.

2.17.12 *Polymerase chain reaction*

PCR was carried out using the GoTaq PCR Core System. Template DNA (5 µl in the case of RT product) was mixed with reaction components (10 µl 5X Colourless GoTaq Flexi buffer, 4 µl $MgCl_2$ (25 mM stock), 1 µl dNTPs (40 mM stock), 1 µl forward primer (50 mM stock), 1 µl reverse primer (50 mM stock), 0.25 µl (5 units) GoTaq DNA polymerase, autoclaved deionised water to a final volume of 50 µl) in a thin-walled PCR tube. Samples were loaded into Mastercycler Gradient thermacycler.

2.17.13 *High Fidelity Polymerase Chain Reaction*

High fidelity PCR was carried out using the Platinum pfx polymerase. Template DNA (100 ng for genomic DNA, and 10 ng for plasmid DNA) was mixed with reaction components (10 µl 10x PCR Buffer, 1.5 µl of 10 mM dNTP mix, 1 µl of 50mM $MgSO_4$, 1.5 µl of each primer, 0.5 µl of Platinum pfx polymerase, and deionised water up to 50 µl) in a thin walled PCR tube.

2.17.14 *PCR purification*

PCR products were purified as per manufacturer's instructions. Five volumes of Buffer PB was added to 1 volume of PCR reaction and mixed. If the colour of the reaction was orange or violet, then 10 µl of 3 M sodium acetate, pH 5.0 was added and mixed and colour of mixture turns yellow. A QIAquick column was placed in a 2 ml collection tube. Sample was applied to QIAquick column and centrifuged for 1 min at 13,000 rpm to bind DNA. Flow-through was discarded. To wash DNA, 0.75 ml of Buffer PE was added to column and centrifuged as before. DNA was eluted from the column by adding 50 µl deionised water or EB buffer, incubating at room

temperature for 5 min and centrifuging as before. DNA was stored at 4 °C until needed.

2.17.15 *Cel-1 (Surveyor) Nuclease Assay*

30 µl of PCR product was incubated in a heat block at 95 °C for 5 minutes. The heat block was removed and left to cool to room temperature allowing heteroduplex formation. 16 µl of the heteroduplex solution was incubated with 2 µl of surveyor nuclease and 2 µl of enhancer solution at 42 °C for 20 min. The surveyor nuclease was inactivated by adding 2 µl stop solution.

2.17.16 *T7E1 assay*

Two hundred nanograms of purified PCR product was added to 2 µl NEB buffer 2 and deionised water to a total of 19 µl volume. Solution was incubated in a heat block at 95°C for 5 min. Heat block was removed from and left to cool to room temperature facilitating heteroduplex formation. 1 µl (10U) T7E1 was added to heteroduplex solution and incubated at 37°C for 15 min. Two microlitres of 0.5 M EDTA was added following incubation.

2.17.17 *In vitro transcription translation reaction*

In vitro transcription translation (IVTT) reactions were carried out according to manufacturer's specifications. 2 µg of T7 expression plasmid was incubated with 1 mM methione and 1X IVTT reaction mix in a 50 µl reaction. All reactions were incubated at 30°C for 2 hours. A Renilla luciferase expression plasmid (pHRL) was used as a positive control. Renilla protein concentrations were determined by a luciferase assay with 100,000 relative light units (RLU) corresponding to 1 ng of Renilla protein. ZFN protein concentrations were approximated to the Renilla values after accounting for differences in peptide length.

2.18 Transfection

2.18.1 *Transfection of cell lines*

Cells were transfected with 4 µg of DNA and 10 µl of Lipofectamine 2000. Cells were plated in transfection media 1, 24 hours in advance of transfection as described in section 2.14. Plasmid DNA and Lipofectamine 2000 were each mixed separately with transfection medium 2, combined, and incubated at room temperature for 20 minutes. Mixture was transferred to cells and incubated at 37 °C and 5 % CO_2.

2.19 Fluorescence Microscopy

Cells were transfected with pEGFP and incubated for 24 hours to allow for enhanced green fluorescent protein (EGFP) expression. Cells were washed with 1 x PBS. Cells were viewed on an Olympus 1x70 inverted microscope equipped with a green fluorescent filter set (500±10 nm excitation and 535±15 nm emission). Images were acquired using Olympus analysis software.

2.20 Flow Cytometry

Cells were harvested with 500 µl of trypsin. 500 µl of media was added and the cells suspension was centrifuged at 1200 rpm for 5 mins. The cell pellet was washed with 1 ml of cold 1XPBS and centrifuged again at 1200 rpm. The cell pellet was re-suspended in 1 ml PBS plus FCS. Cells were sorted and GFP positive cells were detected using a green fluorescent protein filter set (500±10 nm excitation and 535±15 nm emission).

2.21 Genotyping

Twelve microlitres of formalin and 0.5 µl of a known marker were added to 1 µl PCR sample and heated to 95 °C for 5 mins and cooled for 10 mins. The sample was transferred to a 96-well plate and the proportion of PCR product at 101 and 103 bp was analysed using a Genetic Analyser 3130 (Applied Biosystems).

2.22 Immuno-staining

Cells were grown on 13 mm coverslips for 72 hrs at a density of 2 x 10^5 cells/well in a 12-well plate. Following washing with 1 X PBS, cells were fixed with 300 µl of

4% paraformaldehyde for 10 mins. Cells were washed thoroughly and permeabilised with 300 µl of 0.1% Saponin for 30 mins. Cells were washed and blocked with 300 µl of 1% bovine serum albumin and 5% goat serum for 1 hr. Cells were washed. Primary CFTR 596 antibody was added at 1:1000 and incubated for 1 hr at room temperature. Cells were washed. A secondary antibody, goat anti-mouse IgG Cy3 (abcam), was added at 1:5000 as well as the nuclear stain DAPI (1mg/ml) at 1:500 in a dark room and incubated for 1 hr. A single drop of mounting media was added to glass slides and sterilised forceps were used to position the coverslip upside down onto the mounting media making sure to avoid air bubbles. Slides were incubated in the dark at room temperature until ready to view.

2.23 Enrichment of GFP positive cells

Cells were initially transfected in a 6-well plate. In wells containing GFP positive cells, one-hundred micro-litres of the cell suspension was re-seeded 7-days post-transfection into a new 6-well plate. At day 12, the complete cell population of GFP positive cell containing wells was transferred evenly onto a 96-well plate in an attempt to decrease the surface growth area and increase GFP positive cell growth. Again the complete cell population of any GFP containing wells in the 96-well plate was transferred into its own 96-well plate at day 14.

To increase growth, the concentration of foetal calf serum (FCS) was increased by 5% and cells were seeded into a new 96-well plate at day 21. Cells were transferred to a single well of a 24-well plate. The cell population from this single well was then transferred to three new wells of a 24-well plate. Serum starvation was applied to cells to induce a stress response which would lead to increased GFP positive cell survival. The concentration of FCS was altered with cells receiving 0, 5 or 10% total FCS for 24 hours. At day 22 the FCS was increased to 4% in well with 0% FCS. At day 23, each of the three wells was re-seeded into three new wells increasing the number of wells to nine, with each set of three supplemented with varying concentrations of FCS, either 1, 2, 5 or 10%. Cells kept in 1% FCS showed the highest increase in GFP from 4 cells at day 26 to 10 on day 27. At day 29 no further increase was detected and a yeast infection determined the end of the enrichment.

2.24 Next Generation Sequencing

Next generation sequencing (NGS) was performed by Eurofins MWG. PCR samples were amplified using labelled For-CFTR-NHEJ and Rev-CFTR-NHEJ primers shown in table 2.11. PCR products of 452 bp were pooled together into a single eppendorf at a final concentration of 100 ng per sample.

Sample	Labelled Primer
Untransfected A	CGTAGACTAGATCATGTGCCCCTTCTCTGT
Untransfected B	TACGAGTATGTGCTTTGATGACGCTTCTGTAT
CFi9-L^{EL}/CFi9-R^{KK} A	TACTCTCGTGATCATGTGCCCCTTCTCTGT
CFi9-L^{EL}/CFi9-R^{KK} B	TAGAGACGAGTGCTTTGATGACGCTTCTGTAT
CFi9-L^{EL}/CFi9-R^{KK} Donor-XC A	TCGTCGCTCGATCATGTGCCCCTTCTCTGT
CFi9-L^{EL}/CFi9-R^{KK} Donor-XC B	ACATACGCGTTGCTTTGATGACGCTTCTGTAT
CFi9-L^{EL}/CFi9-R^{KK} Mini-gene A	ACGCGAGTATATCATGTGCCCCTTCTCTGT
CFi9-L^{EL}/CFi9-R^{KK} Mini-gene B	ACTACTATGTTGCTTTGATGACGCTTCTGTAT
CFi9-L^{EL}/CFi9-R^{KK} Donor-*Nhe*I1.5 kbA	ACTGTACAGTATCATGTGCCCCTTCTCTGT
CFi9-L^{EL}/CFi9-R^{KK} Donor-*Nhe*I1.5 kb B	AGACTATACTTGCTTTGATGACGCTTCTGTAT
CRISPR In9 A	AGCGTCGTCTATCATGTGCCCCTTCTCTGT
CRISPR In9 B	AGTACGCTATTGCTTTGATGACGCTTCTGTAT
CRISPR In9 Mini-gene A	ATAGAGTACTATCATGTGCCCCTTCTCTGT
CRISPR In9 Mini-gene B	CGCTACGTTGCTTTGATGACGCTTCTGTAT
CRISPR Ex10 A	CAGTAGACGTATCATGTGCCCCTTCTCTGT
CRISPR Ex10 B	CGACGTGACTTGCTTTGATGACGCTTCTGTAT
CRISPR Ex10 Donor- XC A	TACACACACTATCATGTGCCCCTTCTCTGT
CRISPR Ex10 Donor- XC B	TACACGTGATTGCTTTGATGACGCTTCTGTAT

Table 2.11: List of primers used for NGS. Underlined sequence indicates sequencing label used to identify samples. A: forward sequence, B: reverse sequence.

2.25 Miscellaneous methods

2.25.1 *Dual-luciferase reporter assay*

Luciferase assays were performed using the Dual-Luciferase Reporter Assay System according to manufacturer's guidelines. Cells were transfected 24 hours prior to assaying with separate plasmids containing genes for firefly and *Renilla* luciferase. Cells were washed with 1X PBS, then 500 µl of passive lysis buffer was added and the dishes placed on an orbital shaker for 15 minutes at 100 rpm. 20 µl of lysate was transferred to a borosilicate glass tube and 100 µl of Luciferase Assay Reagent was added to this and mixed by pipetting. The tube was placed in a single-sample luminometer and the firefly luminescence measured for 10 seconds. 100 µl of Stop & Glo Reagent was added and the tube briefly vortexed. *Renilla* luminescence was measured for 10 seconds. Luminescence ratio was calculated according to:

[Luminescence ratio = Firefly luminescence/*Renilla* luminescence]

3 Design and construction of pITR-donor and pITR-donor-XC donor repair plasmids, and correction of ΔF508 mutation in the CFTE cell line

3.1 Aims:

The aim of this chapter was to design donor repair plasmids for use with exisitng CFi9-L^{EL}/CFi9-R^{KK} ZFNs to correct the ΔF508 mutation in the CFTR gene. These ZFNs cleave the CFTR gene to create a double stranded break (DSB) 203 bp upstream of the CTT deletion (Lee et al., 2012). To test if AAV-ITR sequences affect HDR and to facilitate downstream use of AAV as delivery vectors, donor sequences were flanked by AAV-ITRs.

3.2 Objectives:

- Amplify a 4.3 kb region of the CFTR gene, and clone it into a plasmid flanked by inverted terminal repeats (ITRs) from Adeno-Associated virus (AAV) to make pITR-donor
- Introduce two unique restriction sites into pITR-donor, to help detect HDR events, using site directed mutagenesis to make pITR-donor-XC
- Transfect CFTE cells with donor plasmids and the CFi9-L^{EL}/CFi9-R^{KK} ZFNs to effect HDR
- Detect repair using an RT-PCR based assay
- Measure the level of repair using restriction digests and genotype analysis

3.3 Design and construction of pITR-donor and pITR-donor-XC

3.3.1 *Cloning of CFTR into AAV backbone*

To create the pITR-donor donor plasmid, a 4,343 bp PCR product was generated from genomic DNA from HBE cells using high fidelity PCR and the primers CFTR-Fwd-*Bst*EII and CFTR-Rev-*Mlu*I, which introduce a unique *BstE*II site at the 5'end and a *Mlu*I site at the 3' end. This PCR product contains wild type exon 10 and 4.3 kb of homology to the flanking intronic regions.

Optimal temperature conditions for the PCR were identified as seen in figure 3.1. The 4,343 bp band was extracted from the gel and purified for cloning. This WT CFTR sequence from the HBE cell line differs from the mutant CFTE cell line sequence by a well characterised single nucleotide polymorphism (SNP) A/G rs213950 in exon 10 as well as the CTT.

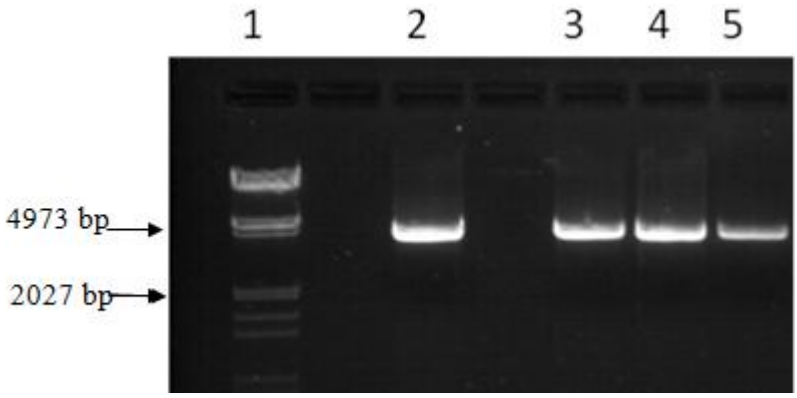


Fig 3.1: Temperature optimisation of CFTR-Fwd-*Bst*EII and CFTR-Rev-*Mlu*I primers. Lane 1: *Eco*RI/*Hind*III lambda ladder, band sizes 2027 and 4973 bp indicated by arrows. Lanes 2-5: 20µl PCR product of genomic HBE DNA at 51.6, 56.8, 54, 60 °C.

The PCR product in lane 2 was gel extracted, purified then cleaved with *BstE*II and *Mlu*I and ligated into the *BstE*II/*Mlu*I backbone of the pAAV-IRES-GFP vector as depicted in figure 3.2.

Following transformation, two potential clones were analysed by restriction analysis using *EcoR*I, a successful clone would yield a single band of 7,251 bp. As seen in figure 3.3 both clones were correct and subsequent sequencing of ITRs and exon 10, confirmed successful cloning.

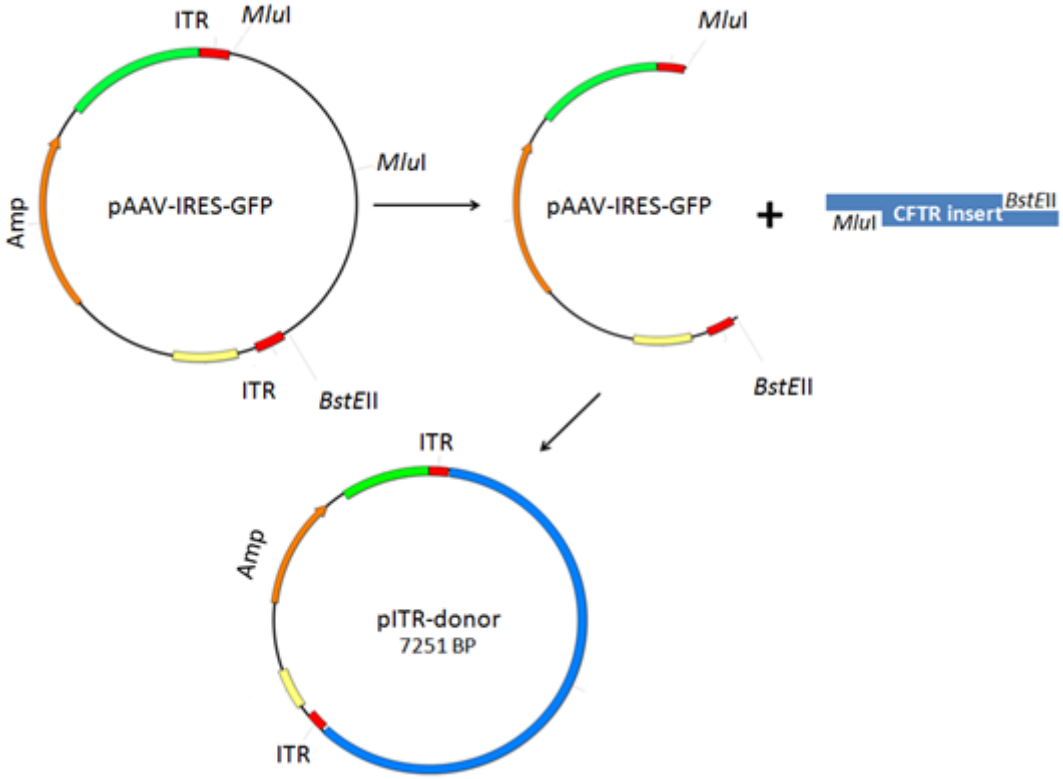


Figure 3.2 Schematic representation of pAAV-IRES-GFP and pITR-donor constructs. Location of *BstE*II and *Mlu*I restriction sites used to clone CFTR insert are indicated.

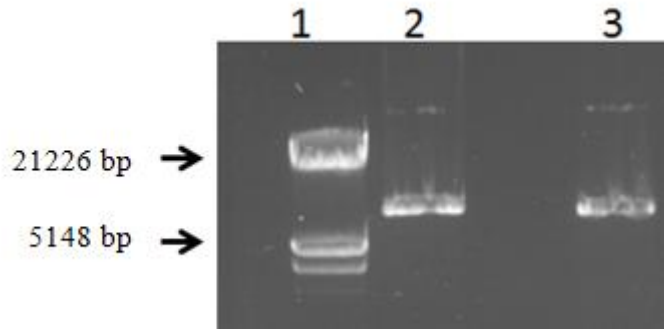


Figure 3.3 Analysis of two potential pITR-donor clones. Lane 1 contains lambda DNA *Eco*RI/*Hind*III marker, band sizes 5148 and 21,226 bp indicated by arrows. Lane 2 and 3 contain potential pITR-donor digested with *EcoR*I.

Use of the CFi9-L^{EL}/CFi9-R^{KK} ZFNs and this donor plasmid containing exon 10 derived from HBE cells should be able to correct the CTT deletion, and would also introduce the G variant of the well characterised SNP, both these changes occur within exon 10 so would be incorporated into the *CFTR* mRNA, see section 3.5.

3.3.2 *Site-Directed Mutagenesis on pITR-donor to yield pITR-donorXC*

To make detection of HDR easier, a donor containing *Xho*I and *Cla*I restriction sites was constructed (pITR-donor-XC) by site-directed mutagenesis. The *Xho*I site was 92 bp upstream of the CTT and the *Cla*I site was 21 bp downstream of the CTT sequence (see figure 3.5). To verify successful introduction of both sites, 5 clones were digested with *Xho*I and *Cla*I independently as seen in figure 3.4. Clone 3 was positive for both restriction sites as seen in lanes 7 and 8 of figure 3.4 and these were further confirmed by sequencing.

Following successful repair with this donor plasmid, these unique restriction sites will be incorporated into the genome of CFTE cells and cleavage of PCR products using these enzymes will allow a quantification of repair.

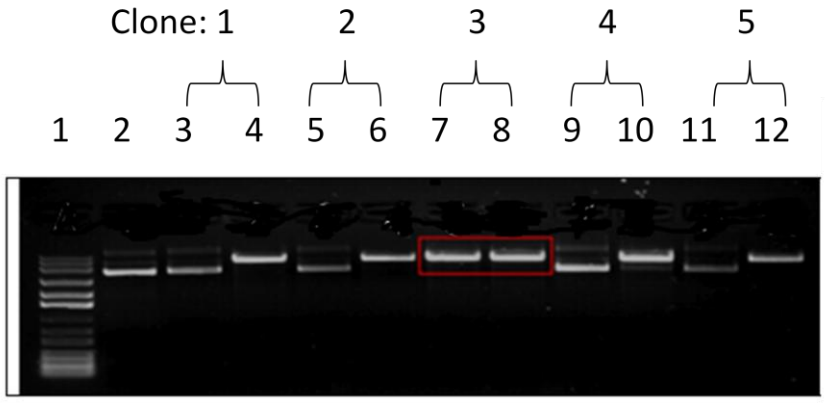


Figure 3.4: Clones digested with enzymes *Xho*I and *Cla*I. Lane 1: Mass Ruler Express Reverse DNA Ladder. Lane 2: Uncut clone; negative control. Lanes 3,5,7,9 and 11 were digested with *Xho*I, lanes 4, 6,8,10 and 12 with *Cla*I. Clone 3 was positive for both restriction sites.

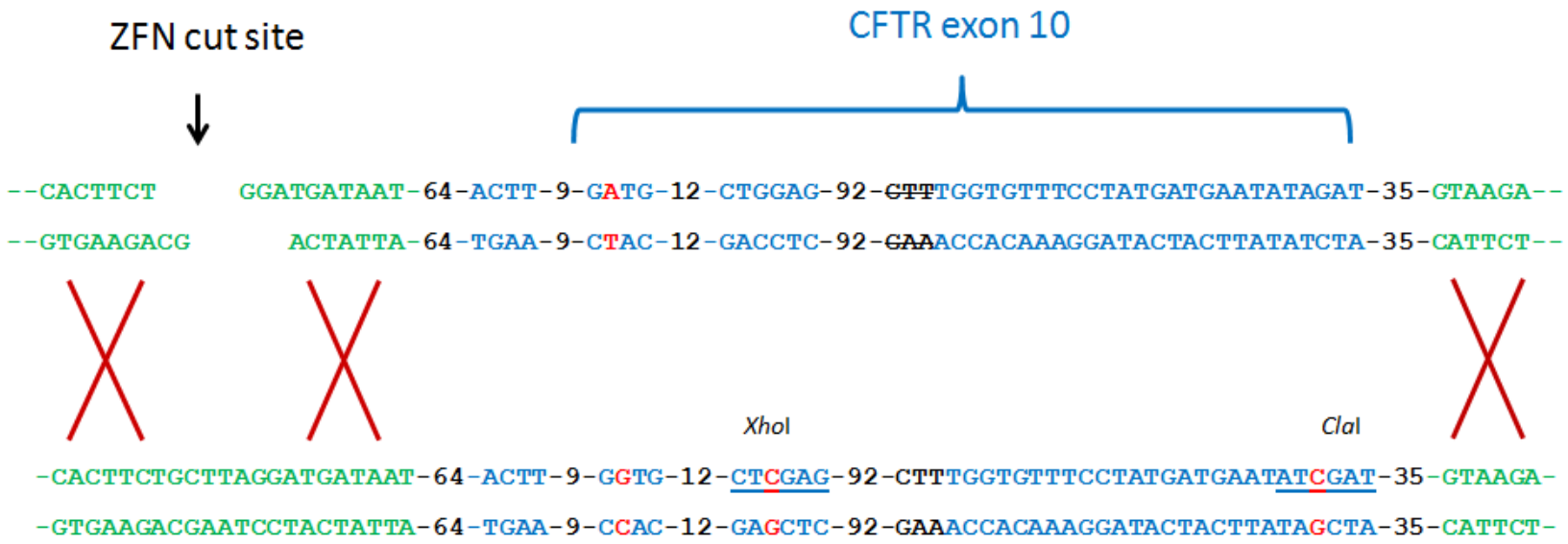


Figure 3.5: Schematic representation of CFTR sequence in CFTE cells (top) and donor plasmids (bottom).Base changes for introduced restriction sites and A/G SNP shown in red. Numbers represent intervening base pairs. Large red X's represent homologous recombination.

3.4 Development of a wild type specific RT-PCR assay to detect homology directed repair with pITR-donor

If HDR is successful it should correct the CTT in exon 10 and add CUU in the mRNA. To detect repair, an RT-PCR assay previously established in the lab to generate a product using a primer pair that can only bind the corrected mRNA was used (Lee et al., 2011). Successful amplification via WT specific RT-PCR should yield a 675 bp band.

3.5 CFi9-L^{EL}/CFi9-R^{KK} ZFNs mediate HDR in the presence of pITR-donor or pITR-donor-XC

To test the hypothesis that HDR using our donor plasmids and ZFNs can correct the genetic mutation in CF cells, CFTE cells were transfected with CFi9-L^{EL} and CFi9-R^{KK} and pITR-donor or pITR-donor-XC as outlined in table 3.1.

Assay / **Plasmid (µg)**	1	2	3	4	5	6
CFi9-L^{EL}	0	1	0	0	1	1
CFi9-R^{KK}	0	1	0	0	1	1
pITR-donor	0	0	2	0	2	0
pITR-donor-XC	0	0	0	2	0	2
pcDNA3.1	4	2	2	2	0	0

Table 3.1: Amount of plasmids transfected for CFi9-L^{EL}/CFi9-R^{KK} HDR assays with pITR-donor or pITR-donor-XC

RT-PCR was performed on total RNA extracted 72 hrs post-transfection using forward and reverse RT-PCR primers, GAPDH primers were used as a positive control.

.

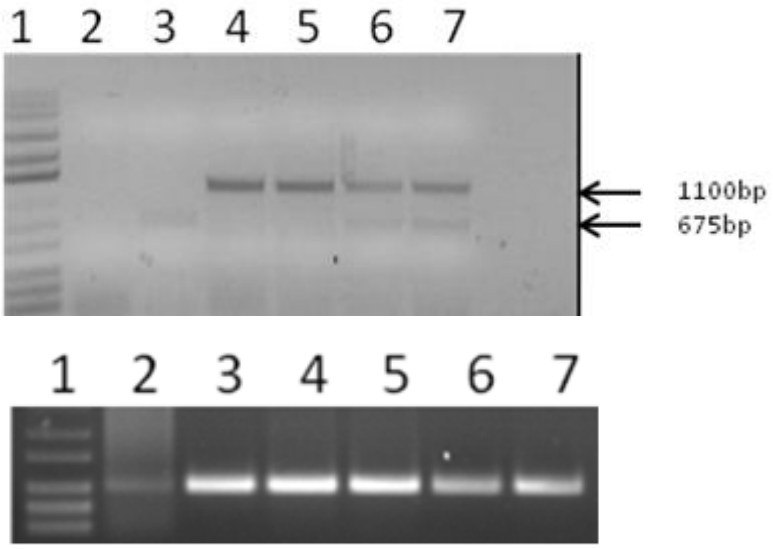


Figure 3.6: Upper panel: Results from WT specific RT-PCR assay. Lane 1: Mass Ruler Express Reverse DNA Ladder. Lanes 2-7 contain samples 1-6 from table 3.1. Lower panel: GAPDH RT-PCR reactions of same samples.

The presence of a 675 bp band in lanes 6 and 7 in figure 3.6 indicates repair when either pITR-donor or pITR-donor-XC donor plasmids are used with the ZFNs. (The band at 1100bp seen in lanes 4-7 transfected with donor plasmids is an artefact that was also observed in PCR optimisation). Cells transfected with the donor should contain the CTT and G SNP. Cells transfected with pITR-donor-XC should also contain the introduced *Xho*I and *Cla*I sites. The 675 bp band in lane 7 from the gel in figure 3.6 was extracted and cloned into a TOPO vector and sequenced. As the RT-PCR was performed using a reverse primer which binds the CTT at the 3'end, one would expect to see this sequence in all products, therefore the G SNP and restriction enzyme sites were used as indicators of homologous recombination. Figure 3.7 shows the presence of the *Xho*I site and G SNP which are 92bp and 103 bp upstream of the CTT deletion respectively in one of the sequenced clones. As this sequence can only be derived from an mRNA molecule that has been synthesised from ZFN-HDR corrected genomic DNA, this data confirms that CFTR gene repair can occur.

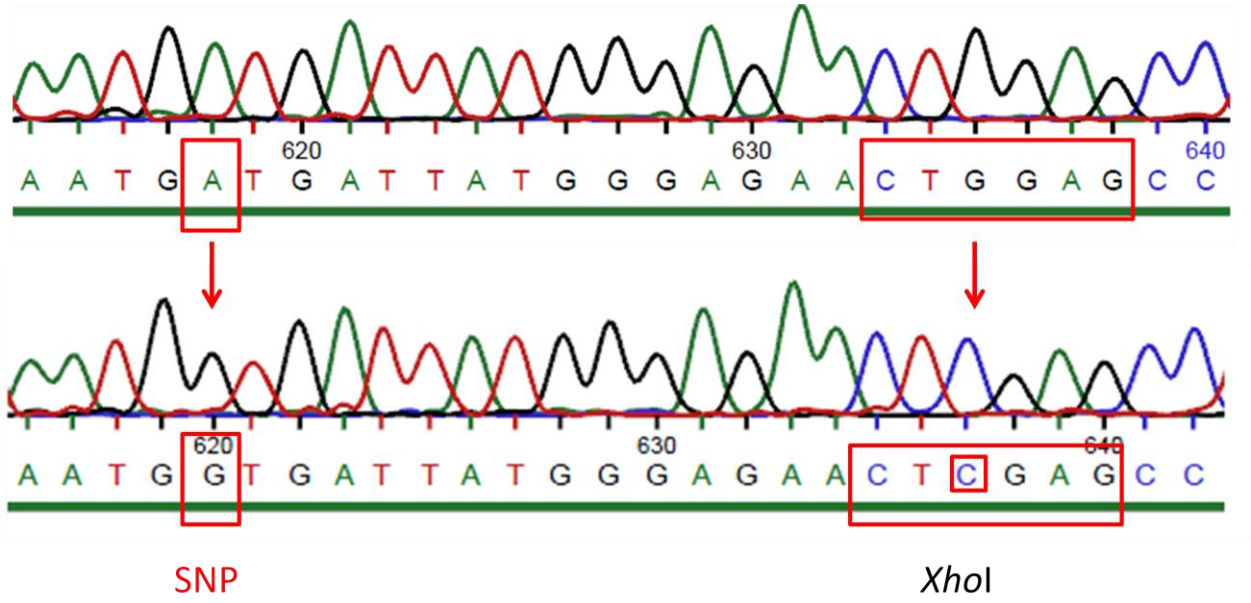


Figure 3.7: Alignment of part of exon 10 of *CFTR* cDNA from CFTE cells transfected with pCFi9-L^{EL}/pCFi9-R^{KK} and pITR-donorXC (bottom) versus mock transfected CFTE cells (top). The corrected sequence contains both the G SNP and the complete *Xho*I site, whereas the wild-type has the A SNP and an incomplete *Xho*I site.

However, of 16 clones sequenced, the remainder showed no incorporation of the SNP or *Xho*I but the CTT was present, suggesting the possibility of mis-priming. There is a faint band of 675 bp in lane 2 of figure 3.6 containing sample transfected with CFi9-L^{EL}/CFi9-R^{KK} ZFN only. This may be due to mis-priming of the Rev-RT-PCR primer with the mutant ΔF508 sequence, as the sequence to which the primer binds is almost an exact match regardless of the absence of the CTT triplet in the wild type sequence as seen in figure 3.8. The second base A at the 3' end is the only mismatch in the mutant ΔF508 sequence when compared to the wild type sequence.

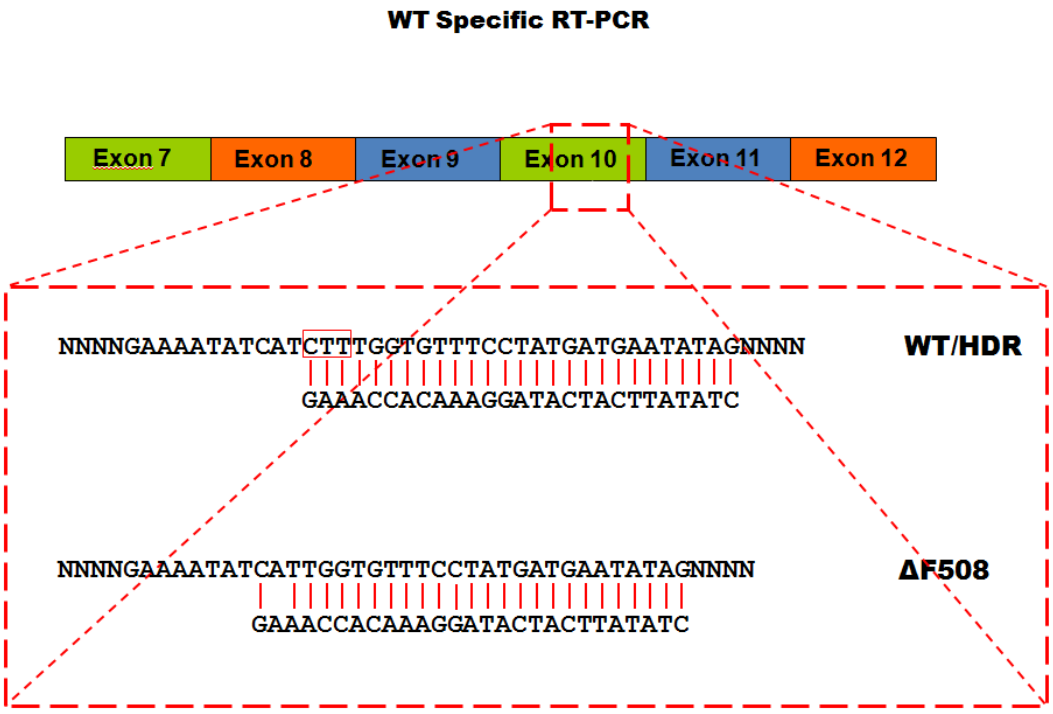


Figure 3.8: Diagram demonstrating the possibility of mis-priming occurring between the CTT specific reverse primer and the mutant sequence.

3.6 Quantification of the level of repair

3.6.1 *XhoI/ClaI digests:*

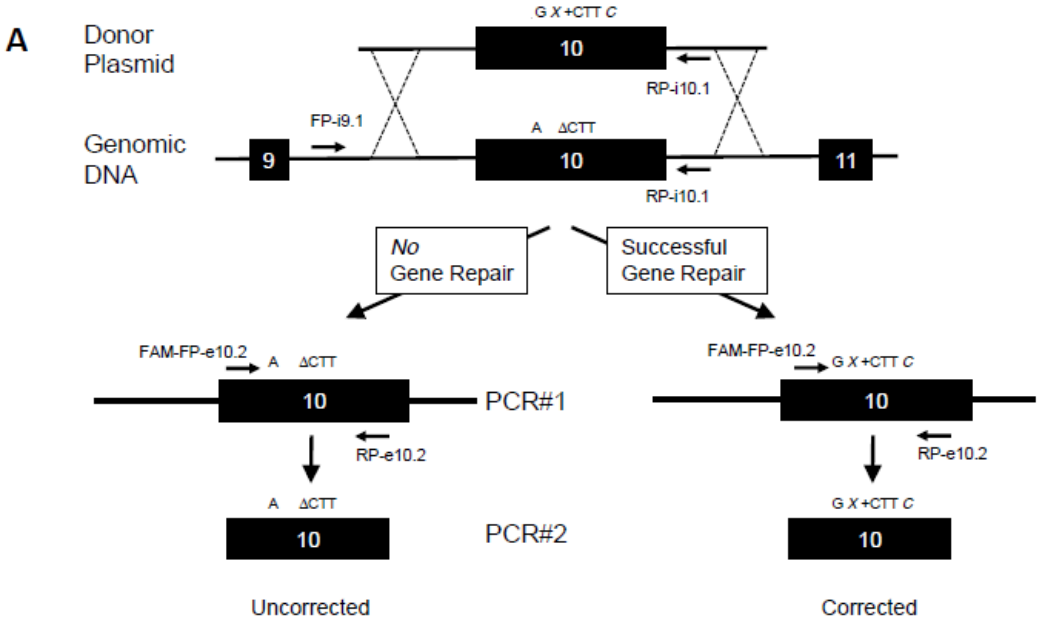
Having shown that CFTR gene correction was possible it was necessary to determine the level of repair. The pITR-donor-XC donor provides a potentially easy method for quantification of repair as detection of either restriction site by digest analysis on the DNA from corrected cells indicates that HDR has occurred. Previous studies using this approach could detect as low as 2.6% (Urnov et al., 2005). Seventy-two hours post-transfection of CFTE cells with ZFNs and pITR-donor-XC donor, genomic DNA was extracted, and a 2.6 kb PCR product was generated using a forward primer in intron 9 FP-i9.1 (outside donor plasmid region to eliminate plasmid carry-over) and reverse primer, RP i10.1, in intron 10 located approximately 490 bp downstream of the CTT site. The 2.6 kb product was gel extracted and purified for cloning into the TOPO TA vector system. Following transformation 96 clones were cultured and digested with either *Xho*I or *Cla*I. No clones were successfully digested, suggesting that the level of repair was below 1%.

3.6.2 *Genotyping*

A potentially more sensitive assay was developed involving a two-step fluorescence PCR-based assay to quantify the percentage of cells in which this ZFN-mediated HDR had occurred. Using this method a 101 bp product represents uncorrected DNA whereas a 104 bp product indicates gene correction. The percentage of products that are 104 bp will indicate the percentage of cells in which at least one ΔF508 CFTR allele has been corrected. To determine the level of detection of the assay, different amounts of PCR amplicons derived from HBE and CFTE cells, corresponding to corrected and uncorrected sequences respectively, were mixed at a number of different ratios and screened using a Genetic Analyser 3130 (Applied Biosystems). As seen in figure 3.9B, at 50% HBE to 50% CFTE, corrected: uncorrected, the system was able to detect the corrected PCR product. At 1% HBE to 99% CFTE, a corrected product is still detectable. However, at 0.5% there was no detection, indicating the limit of detection using this method was approximately 1%.

Having verified the assay to detect correction of genomic DNA, we used it to analyse DNA treated with ZFNs and/or the donor, extracted 72 hrs post-transfection. A 2.6 kb PCR product (PCR#1) was generated using forward primer FP-i9.1 and reverse primer RP i10.1, see figure 3.9A and then a second PCR (PCR#2) amplified using primers, FAM-FPe10.2 (FAM labelled) and RP e10.2, both within exon 10, was performed on these 2.6 kb products.

No PCR#2-WT product was detected in genotyping analysis of mock transfected CFTE cells, CFTE cells transfected with ZFNs alone, or pITR-donor/pITR-donor-XC only. No PCR#2-WT product was detected with ZFNs and pITR-donor-XC (figure 3.9C) suggesting that the level of repair is below 1%. We know that repair has occurred as seen by PCR and sequencing data but the level is below the detection limit of the Genetic Analyser 3130. We can conclude from these results that repair of the CTT deletion in CFTE cell line is possible using ZFN and donor plasmids; however, the efficiency of repair is low.



B

CFTE:HBE ratio	PCR #2 Uncorrected	PCR #2 Corrected
100:1	147,407	0
99:1	44,416	147
50:50	87,238	30,535

C

CFTEs Transfection	PCR #2 Uncorrected	PCR #2 Corrected
pITR-donorXC only	42,977	0
pITR-donorXC + ZFNs	10,137	0

Figure 3.9 Two-step fluorescence-based assay to detect HDR. **(A)** PCR#1 should amplify a 2.6-kb region of genomic DNA without creating a product from pITR-donor-XC. Using the 2.6-kb product as a template, the second PCR (PCR#2) should generate a 101-bp uncorrected amplicon (A: ΔCTT) in the absence of ZFN-HDR, or a 104-bp amplicon if the ΔCTT has been repaired; this corrected amplicon should also contain *Xho*I and *Cla*I restriction sites (G X+ CTT C). **(B)** Relative levels of uncorrected and corrected PCR#2 amplicons generated from genomic DNA from HBE and CFTE cells mixed at percentages shown. **(C)** Relative levels of uncorrected and corrected PCR#2 amplicons generated from genomic DNA from CFTE cells transfected as shown. HDR, homology directed repair.

3.6.3 *Next generation sequencing (NGS) data*

Shortly before submission of this thesis, the opportunity to perform a small-scale deep sequencing analysis arose. NGS was performed on a single sample obtained from cells transfected with ZFNs and pITR-donor-XC. Data was obtained from 629 sequences, of these, 12 had the correct CTT, G->C (*Xho*I) and A->C (*Cla*I). Figure 3.10 below displays some of the sequence data. The arrows indicate the base pair changes that occur following HDR with pITR-donor-XC. As already seen in section 3.5 our ZFNs can mediate HDR with our donor plasmids but the level of repair as determined by digest and genotyping analysis was <1%. However, in this one

sample the level of repair was 2% suggesting the sensitivity of our detection assays is such that they cannot measure the level of repair. More samples will need to be analysed to obtain a more accurate value but the data clearly shows that HDR is occurring in these cells.

Figure 3.10: Sample sequences of 629 PCR products from CFTE cells transfected with ZFNs and pITR-donor-XC confirming incorporation of *Xho*I, CTT and *Cla*I sites from donor into genome. The arrows indicate the base pair changes following HDR with pITR-donor-XC. G->C for *Xho*I site, CTT addition and A->C for *Cla*I site.

3.7 Conclusions

The aim of this chapter was to correct the ΔF508 mutation present in the CFTE cell line using a pair of ZFNs made previously in the host lab and donor plasmids described in this chapter containing the correct sequence.

Evidence of precise repair in cells was initially detected using an RT-PCR based assay three days post-transfection with ZFNs and pITR-donor/pITR-donor-XC, (figure 3.6, lanes 6 and 7). Subsequent sequencing data of clones derived from the DNA in lane 7 confirmed successful repair by homologous recombination with the pITR-donor-XC plasmid by detection of the G SNP and the *Xho*I restriction site (figure 3.7).

Further study to quantitate repair efficiency indicated it was below 1%. This low level of repair, despite cleavage efficiency by ZFNs at 7%, may be due to the distance of the CTT deletion from the ZFN target cut site which is 203bp downstream. To date, efficient ZFN-HDR has only been reported at or within a few base pairs of the ZFN-induced DSB (Urnov et al., 2005) (Maeder et al., 2008). It has been shown previously (Elliott et al., 1998) that the distance between the repair site and cut site is proportional to the level of repair. It was shown that 80% of

gene-targeting events extended no further than 45 bp from the DSB, although gene repair tracts did extend as far as 228 bp in 8% of corrected clones. In the same study, even a low level of heterology (1.2%) between donor and target was shown to decrease the level of repair by approximately six fold. Therefore, we concluded that the 203bp might be too far for efficient correction at the CTT target site and the various differences between pITR-donor-XC sequence and the target region (i.e. a SNP, two base pair changes to create *Xho*I and *Cla*I sites, and the 3 bp CTT insertion) may also contribute to the low level of HDR.

Nonetheless, this initial observation showed for the first time correction of the endogenous CFTR gene in the CFTE cell line following transfection with our ZFNs targeting intron 9 and donor plasmids. Correction of *CFTR* using ZFN-HDR has not previously been reported, although a pair of ZFNs which can cleave human *CFTR* in 1.2 % of cells has been described using the OPEN method of ZFN construction (Maeder et al., 2008), and a pair of ZFNs which can cleave mouse *cftr in vitro* only have also been reported (Mani et al., 2005).

As shown in section 3.6.3, the efficiency is potentially higher than originally envisaged at 2% by NGS. Still, the overall efficiency of repair was found to be low and needs to be improved if this technology is ever to be of potential clinical use. An obvious next step would be to design a new pair of ZFNs closer to the CTT deletion as distance was determined to be a limiting factor. However, this would be time consuming and ZFN design has been shown to have a low success rate.

In this thesis, two different options were subsequently explored:

1) to take advantage of the position of the ZFNs in intron 9 and use this site for insertion of a mini-gene construct comprising exons 10-24 from CFTR cDNA (with appropriate splice acceptor and poly A sites) to allow production of full length corrected CFTR mRNA. This approach has been used successfully in cells to correct the IL2RG locus (Lombardo et al., 2007) and *in vivo* to correct human F9 in a murine model of haemophilia that subsequently restored haemostasis (Li et al., 2011). This strategy would enable the correction of a much wider range of CF causing mutations and will be described in chapter 4.

2) Use other gene-editing tools, more amenable to changing location, such as the CRISPR/Cas9 system, as described in chapter 5.

4 Mini-gene Repair Strategy using ZFNs

4.1 Aims:

The data from chapter three showed that our ZFNs effectively cleave intron 9 of the CFTR gene in 7% of cells. Genotyping data initially suggested CFTR gene repair by HDR occurred, but in less than 1% of treated cells, though a follow-up deep sequencing study showed nearly 2% could be obtained. Some groups have reported HDR levels of up to 20%, but only when the cut site and HDR site coincide; thus, it seemed plausible that this low level could be a result of the distance from our ZFN cut site to the CTT site of repair (~203 bp) which is perhaps too far for really efficient HDR to occur. Rather than design a new pair of ZFNs closer to CTT it was decided to take advantage of the efficient cleavage activity of our ZFNs and use them to explore an alternative strategy whereby they would be used to recombine a partial cDNA cassette (exons 10-24) with appropriate splice acceptor and poly-A sites into the ZFN cut site (intron 9) which should result in a full length WT CFTR transcript following correct splicing. Firstly, to test whether repair is possible at the cleavage site, a new donor plasmid was designed which targets a seven base pair tag containing a unique *Nhe*I restriction site to intron 9 where the ZFNs cleave. Successful recombination with this donor would indicate that mini-gene recombination would be feasible at this location. The donor would then be modified to incorporate the partial cDNA, splice acceptor, poly-A site and a GFP tag (for identification of corrected cells). If this could be recombined in at this location, this would correct all cystic fibrosis-causing mutations occurring in exons 10-24 which accounts for ~80% of all known mutations (Cystic Fibrosis Mutation Database).

4.2 Objectives:

- Design and construct four new donor plasmids;

  - pITR-CFTE, (4.3 kb region of CFTR gene amplified from CFTE gDNA and cloned into pITR-IRES-GFP)

  - pITR-CFTE-*Nhe*I, (two step PCR to introduce unique *Nhe*I site into pITR-CFTE)

- pUC-CFTE-*Nhe*I-1 kb, (1 kb of homology to CFTR gene centered on *Nhe*I in a pUC vector)

- pUC-CFTE-*Nhe*I-1.5 kb, (1.5 kb of homology to CFTR gene centered on *Nhe*I in a pUC vector)

- Detect recombination of *Nhe*I into ZFN cut site by PCR and digest analysis and identify the most suitable donor for mini-gene repair
- Construct mini-gene donor plasmid
- Detect recombination of mini-gene by PCR
- Assess GFP expression from mini-gene

4.3 Design and construction of three *Nhe*I donor plasmids with homology arms of 4.3, 1.5 and 1 kb.

Three *Nhe*I plasmids were constructed to determine which length of homology would be optimal for the mini-gene donor. As sequence divergence between donor DNA and genomic DNA has been shown to lower the efficiency of repair (Elliott et al., 1998) all new plasmids were derived from CFTE genomic DNA to avoid any sequence discrepancies such as SNPs which exist between the CFTR sequences from HBE and CFTE cell lines. A schematic outlining the cloning plan is shown in figure 4.1.

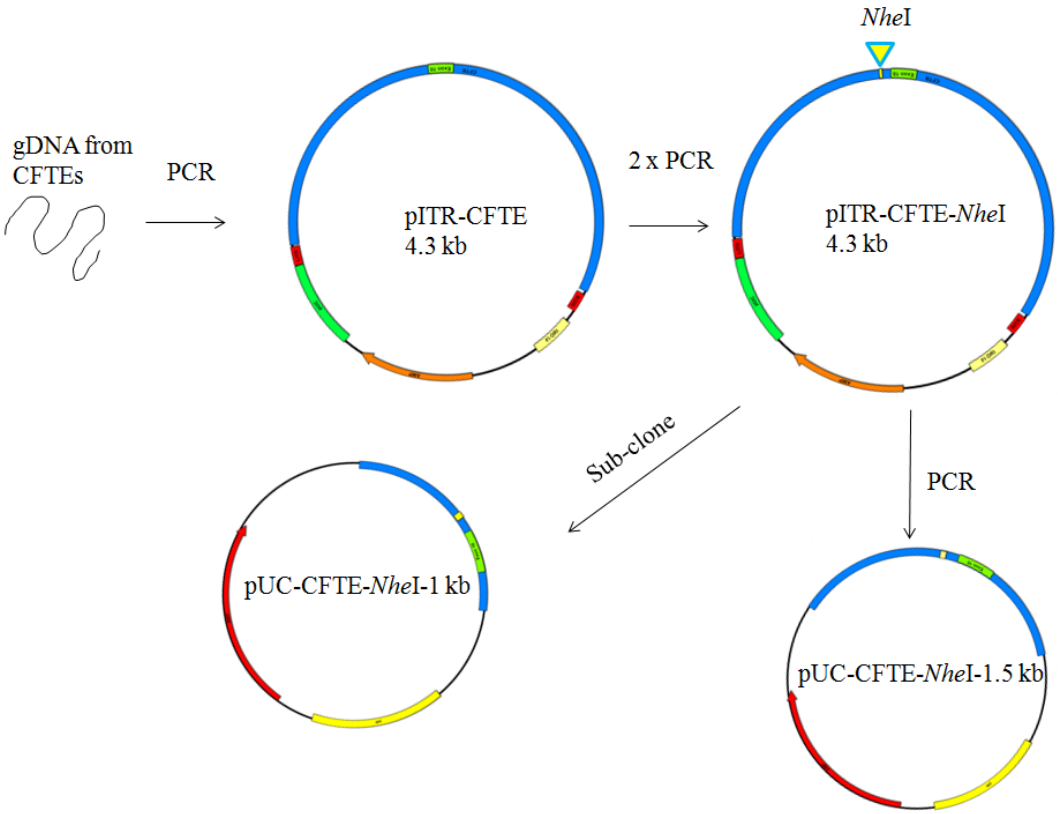


Figure 4.1: Schematic representation of the construction of the new *Nhe*I plasmids. Backbone and length of homology is shown. PCR on CFTE gDNA to create pITR-CFTE. Introduce *Nhe*I site by two-step PCR to create pITR-CFTE-*Nhe*I. Subclone part of pITR-CFTE-*Nhe*I into pUC backbone to create pUC-CFTE-*Nhe*I-1kb. PCR a region of pITR-CFTE-*Nhe*I to create pUC-CFTE-*Nhe*I 1.5kb.

4.3.1 *pITR-CFTE and pITR-CFTE-NheI*

The plasmid pITR-CFTE was created in the same manner as pITR-donor in section 3.3, except the 4.3 kb region of genomic DNA was amplified from CFTE genomic DNA rather than from HBE DNA. This 4.3 kb fragment was then cloned into the pITR-IRES-GFP plasmid backbone using restriction sites *Mlu*I and *Bst*EII.

This plasmid was then modified by introducing a unique tag sequence, 5'-GCTAGCA-3', containing a restriction site for easy detection, into the spacer region of the ZFN target site in pITR-CFTE. Inserting the sequence here will disrupt the ZFN target site in the plasmid and thereby prevent cleavage of repaired alleles. The restriction enzyme *Nhe*I (underlined) was chosen as it is unique to pITR-CFTE. To introduce this tag sequence at the ZFN cut site in the pITR-CFTE plasmid site-directed mutagenesis was initially performed; however after several unsuccessful attempts a PCR strategy was used instead. Four primers were designed

which amplify short regions of DNA and incorporate the *Nhe*I unique site and also restriction sites *Eco*RI and *Btg*I needed for cloning into pITR-CFTE plasmid. Figure 4.2A shows temperature optimisation of both PCR products using the pITR-CFTE plasmid as a template (see figure 4.1). PCR 1 at 544 bp in lane 2 was produced using the primers Fwd_*Eco*RI and Rev_*Nhe*I and PCR 2 at 1178 bp in lanes 5-7 were amplified using Fwd_*Nhe*I and Rev_*Btg*I. The bands of correct size were gel extracted and purified.

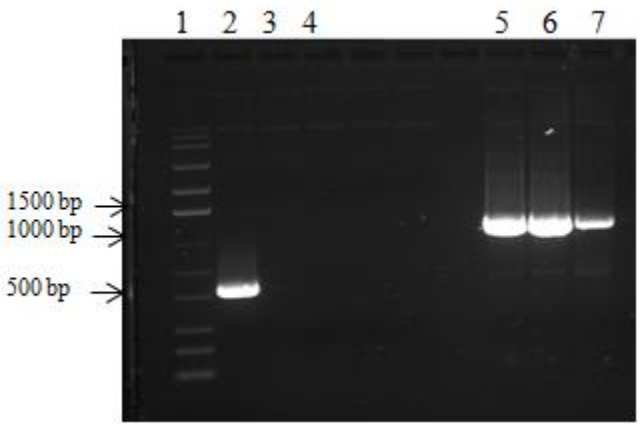


Figure 4.2A: Temperature optimisation of PCR 1 with primer Fwd_ *Eco*RI and Rev_ *Nhe*I. and PCR 2 with primers Fwd_*Nhe*I and Rev_*Btg*I. Lane 1: Mass ruler express reverse DNA ladder. Lanes 2-4: 20 µl PCR product 1 at 55, 60.6, 65.9 °C. Lanes 5-7: 20 µl PCR product 2 at 55, 60.6, 65.9 °C

To create the ITR backbone, pITR-CFTE was digested with *Eco*RI and *Btg*I, (see Figure 4.2B), and the band at 5581 bp was extracted and purified.

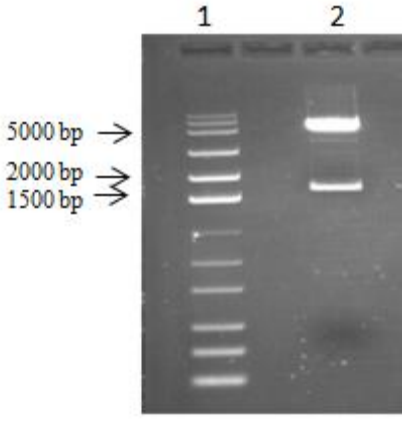


Figure 4.2B: pITR-CFTE digested with *Eco*RI and *Btg*I. Lane 1: Mass ruler express reverse DNA ladder. Lane 2: pITR-CFTE digested producing two bands at 5581 bp and 1567 bp.

A three-way ligation was performed with the PCR products 1 and 2 from Figure 4.2A and the ITR backbone from Figure 4.2B to create the pITR-CFTE-*Nhe*I plasmid. Four potential clones were tested for the introduced *Nhe*I site by restriction digest with *Hind*III and *Nhe*I, figure 4.3. Positive clones should result in 3 bands at 5049, 1766 and 441 bp. Digest analysis revealed three of the four clones contained the *Nhe*I site.

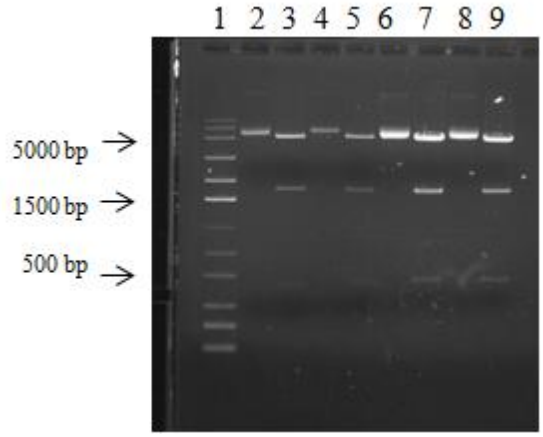


Figure 4.3: Restriction enzyme analysis to identify clones containing the *Nhe*I site. Lane1: Mass ruler express reverse ladder. Lanes 2, 4, 6 and 8: Uncut samples. Lanes 3,5,7 and 9: samples digested with *Hind*III and *Nhe*I. Bands at 5049, 1766 and 441 are positive for the *Nhe*I site. Lanes 3, 7 and 9 display the correct banding pattern.

DNA sequencing was used to confirm the introduction of the *Nhe*I site.

4.3.2 *Construction of pUC-CFTE-NheI-1 kb and pUC-CFTE-NheI-1.5 kb donors*

To test whether the length of homology arms has an effect on the rate of HDR, two smaller donor plasmids containing the *Nhe*I site were constructed using pUC19 as a backbone, one with roughly 1 kb homology and the other with 1.5 kb homology to the target region of CFTR.

Construction of pUC-CFTE-*Nhe*I-1 kb donor plasmid was straightforward by subcloning the 1 kb *Eco*RI/*Hind*III insert of pITR-CFTE-NheI into pUC19 (see figure 4.1). Cloning using these endogenous enzyme sites resulted in a right homology arm of 435 bp and a left homology arm of 544 bp. Following sequential

digest (figure 4.4) the band of correct size in lane 5 was extracted and purified and cloned into the pUC plasmid backbone from lane 3.

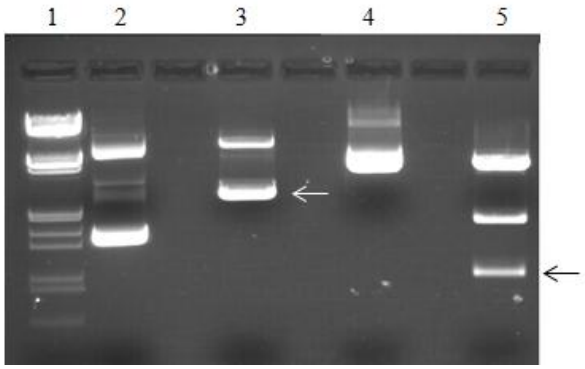


Figure 4.4: Lane 1: *Eco*RI/*Hind*III lambda ladder. Lanes 2 and 4 contain uncut pUC 19 and pITR-CFTE-*Nhe*I plasmids respectively. Sequential digestion with *Eco*RI and *Hind*III of pUC19 and pITR-CFTE-NheI results in the banding pattern shown in lanes 3 and 5. As there are two *Hind*III sites in pITR-CFTE-NheI, three bands are expected at 4536, 1761 and 979 bp. The 979 bp band has the desired sequence and *Nhe*I site. The arrows indicate which bands were extracted for cloning.

Construction of pUC-CFTE-*Nhe*I-1.5 kb donor required amplification of a 1.5 kb fragment of pITR-CFTE-*Nhe*I with primers Fwd_*Kas*I and Rev_*Xba*I (which introduce *KasI* and *Xba*I restriction sites) cloned into pUC19 vector. This cloning strategy produces homology arms of exactly 750 bp either side of the *Nhe*I site.

Digest analysis of 8 clones with *Hind*III and *Nhe*I, which should give a banding pattern of 3216, 435 and 334 bp, revealed 8/8 clones contained the correct insert, figure 4.5. Sequencing data confirmed the correct location of *Nhe*I.

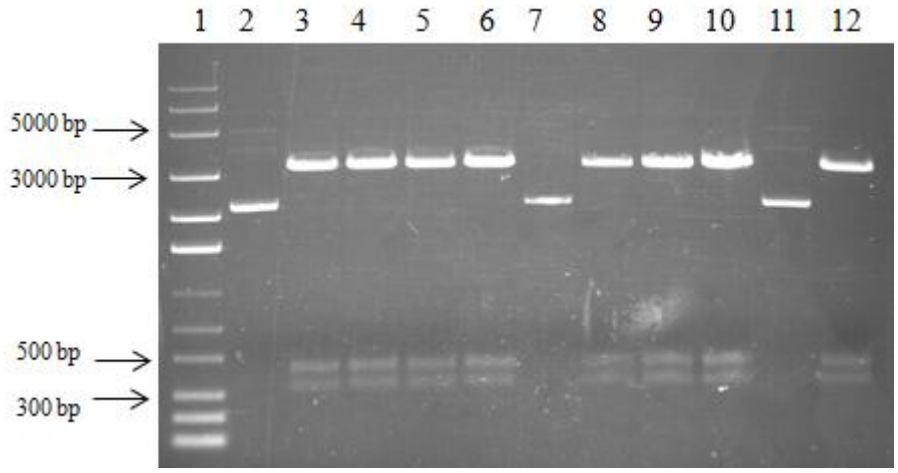


Figure 4.5: Digest analysis of 8 potential clones for *Nhe*I insert. Lane 1: Mass ruler express reverse ladder, Lanes 2, 7 and 11 are uncut plasmid controls. Lanes 3, 4, 5, 6, 8, 9, 10 and 12 were digested with *Nhe*I and *Hind*III. All show the expected banding pattern of 3216, 435 and 334 bp.

4.4 Development of detection assays to determine if *Nhe*I is introduced into the genome following ZFN mediated HDR with *Nhe*I donor plasmids

If HDR with any of the *Nhe*I donor plasmids occurs, the unique seven base pair sequence should be introduced into the genomic DNA within the spacer region of the ZFN cut site. A problem experienced in chapter 3 when detecting the CTT sequence in corrected cells was the occurrence of mis-priming which was explained in section 3.5. The new donor plasmids constructed in this chapter have an advantage in that the sequence to be detected is seven base pairs making the chance of mis-priming very unlikely. Also the fact that a unique restriction site was incorporated into the seven base pairs should enable quantification of repair through digest analysis, as was attempted with the *Xho*I and *Cla*I sites in section 3.6.1. A further advantage with these new donors is that both a positive (pITR-CFTE-*Nhe*I) and negative control (pITR-CFTE) were available to optimise the detection assays.

Two detection assays were developed; firstly, a PCR assay with a *Nhe*I-specific reverse primer and secondly to quantify HDR events a *Nhe*I digest assay.

4.4.1 *NheI-specific primer reaction*

A *Nhe*I-specific PCR using genomic DNA as template should produce a band of 564 bp if successful recombination has occurred with the *Nhe*I donor plasmids as shown in figure 4.6.

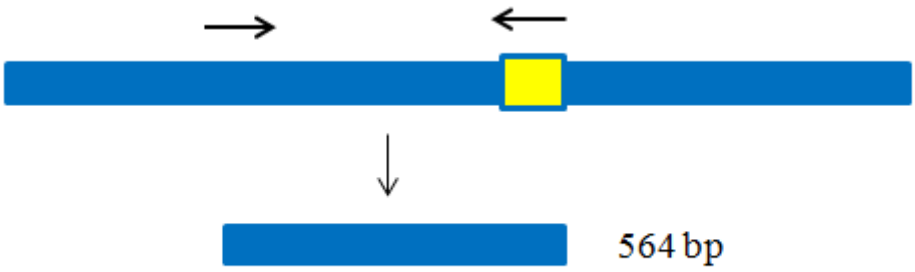


Figure 4.6: Schematic representation of PCR strategy used to detect introduction of *Nhe*I site (yellow box) with *Nhe*I specific reverse primer to give a band of 564 bp.

The Fwd_*Eco*RI primer and Rev_*Nhe*I-specific primer were temperature optimised using proof-reading polymerase Pfx and the plasmids pITR-CFTE and pITR-CFTE-*Nhe*I as templates. A band of 564 bp indicates that *Nhe*I is present in the DNA sample. As seen in Figure 4.7, at 59.5 °C there was a band in both the negative and positive controls. At a higher temperature of 62 °C there was only a band in the positive control therefore this was the temperature used for all *Nhe*I specific PCRs.

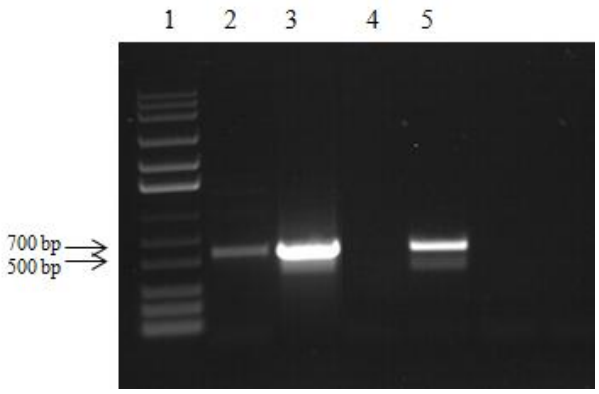


Figure 4.7: Temperature optimisation of NheI specific PCR. Lane 1: Mass ruler express reverse ladder. Lane 2 and 4: Negative control plasmid pITR-CFTE used as a template at 59.5 and 62 °C. Lanes 3 and 5: Positive control plasmid pITR-*Nhe*I used as a template at 59.5 and 62 °C.

4.4.2 *NheI quantification assay*

To quantify the level of recombination with the *Nhe*I donor plasmids a digest detection assay was used. A 2.6 kb PCR was generated from gDNA and subsequently digested with *Nhe*I. Comparing the intensity of the two bands produced against the remaining undigested 2.6 kb band would give an indication of the level of recombination.

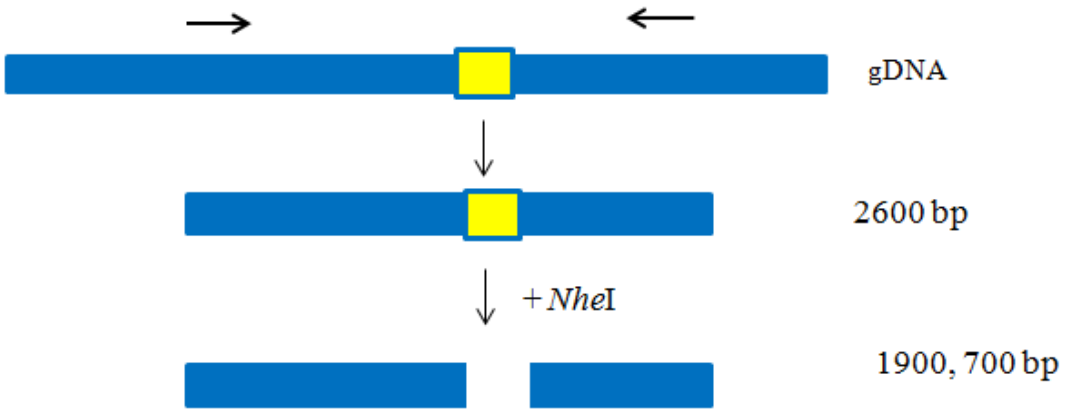


Figure 4.8: Schematic representation of *Nhe*I digest quantification assay. PCR on gDNA produces a band of 2.6 kb. Successful NheI digest produces two bands of 1900 and 700 bp.

As the *Nhe*I product above does not occur in nature, the plasmids pITR-CFTE and pITR-CFTE-*Nhe*I were used in order to determine the level of detection by *Nhe*I digest. Different amounts of pITR-CFTE and pITR-CFTE-*Nhe*I were mixed together and made up to a total of 10 ng and analysed. PCR was performed using the primers Fwd_*Eco*RI and Rev_*Btg*I and go Taq polymerase. Following PCR, products were purified and digested with *Nhe*I and ran on a 5% polyacrylamide gel. This PCR should produce a product of 1729 bp and following successful digest with *Nhe*I should produce two bands of 1180 and 549 bp. In figure 4.9, pITR-CFTE-*Nhe*I was mixed at 50, 20, 10, 5, 2, and 1% of pITR-CFTE. The limit of detection was taken to be 1% as the bands in lane 7 are very weak meaning at ratios below 1% it would be difficult to ascertain whether digest was successful.

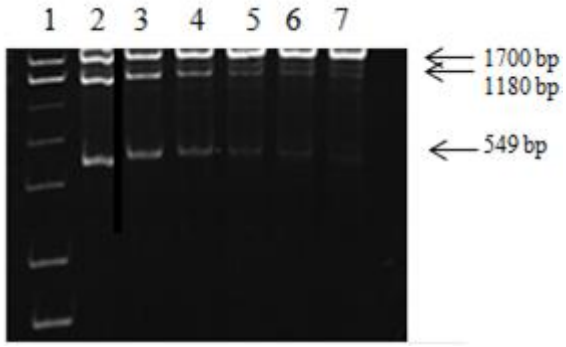


Figure 4.9: *Nhe*I digestion assay to determine the level of detection. Lane 1: Mass ruler express reverse ladder. Lanes 2-7: pITR-CFTE-*Nhe*I at 50, 20, 10, 5, 2 and 1% of pITR-CFTE. Arrows indicate expected bands following successful digestion.

When using this quantification assay to analyse DNA from cells transfected with the *Nhe*I donor plasmids it was necessary to use different primers Fwd_gDNA and Rev_gDNA to avoid false positive results from donor plasmid carry-over in DNA extracts. This PCR product of 2.6 kb upon *Nhe*I digestion will produce two bands of 1900 and 700 bp as shown in figure 4.8.

4.5 Successful incorporation of *Nhe*I tag into CFTR using CFi9-L^{EL}/CFi9-R^{KK} ZFNs and pUC-CFTE-*Nhe*I-1.5 kb donor

To test the hypothesis that HDR can occur at the ZFN cut site using our *Nhe*I donor plasmids and ZFNs, CFTE cells were transfected with ZFNs CFi9-L^{EL} and CFi9-R^{KK} and the various donors as shown in table 4.1.

Assay / Plasmid (µg)	1	2	3	4	5	6
CFi9-L^{EL}	0	1	0	0	1	1
CFi9-R^{KK}	0	1	0	0	1	1
pITR-CFTE-*Nhe*I (4.3 kb)	0	0	0	2	0	2
pUC-CFTE-*Nhe*I-1.5 kb	0	0	2	0	2	0
pcDNA3.1	4	2	2	2	0	0

Table 4.1: Amount of plasmids transfected for CFi9-L^{EL}/CFi9-R^{KK} HDR assays with pITR-CFTE-*Nhe*I or pUC-CFTE-*Nhe*I-1.5 kb.

Seventy-two hours post-transfection, a two-step PCR was performed on total DNA to eliminate false positives from donor plasmid carry over. In step one, high-fidelity PCR with primers Fwd_gDNA (binds outside plasmid) and Rev_gDNA produced a 2.6 kb product (figure 4.10a). This product was purified and used as template for step two.

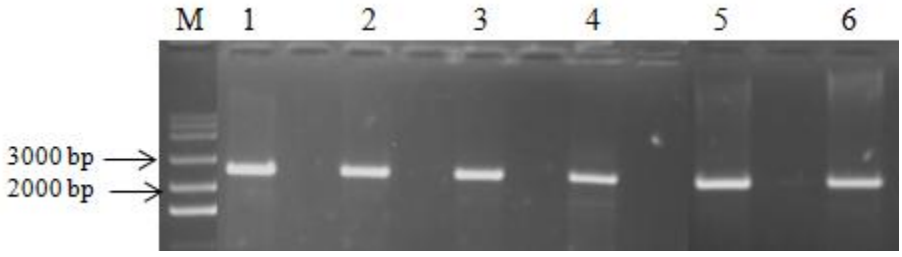


Figure 4.10a: Step one: 2.6 kb PCR product from samples described in table 4.1. M = Mass ruler express reverse ladder. Lanes 1-6 contain samples 1-6 from table 4.1.

In step two, PCR was performed using primers Fwd_*Eco*RI primer and Rev_*Nhe*I-specific primer. A band of 564 bp should only be produced if successful recombination has occurred. Plasmids pITR-CFTE-*Nhe*I and pITR-CFTE were used as positive and negative controls respectively for the *Nhe*I specific PCR, (figure 4.10b).

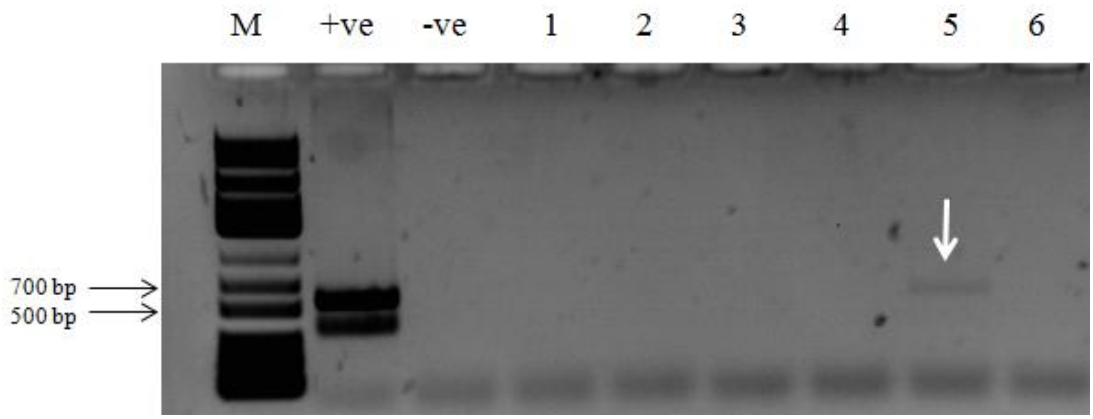


Figure 4.10b: Step two**:** *Nhe*I specific PCR. M = Mass ruler express reverse ladder. +ve = positive control, -ve = negative control, lanes 1-6 contains samples 1-6 from table 4.1. White arrow in sample 5 indicates correct band of 564 bp.

The absence of a 564 bp band in sample 6 indicates that the combination of ZFNs and pITR-CFTE-*Nhe*I donor did not induce a detectable level of HDR. No band was present in samples treated with ZFNs and pUC-CFTE-*Nhe*I 1 kb donor either (data not shown).

However, the presence of a 564 bp band in sample 5 indicates that HDR has occurred in cells transfected with ZFNs and pUC-CFTE-*Nhe*I-1.5 kb. This shows that it is possible to insert a unique seven base tag into the ZFN site in exon 10 of CFTR genomic DNA.

To determine the level of recombination a quantification assay was performed as described earlier in section 4.4.2. Several attempts were made to quantify recombination but no detectable cleavage was observed (data not shown). Based on the data from the sensitivity assay (figure 4.9), it was initially deduced that the level of HDR at the ZFN cut site was below 1%.

However, NGS performed on a single sample obtained from cells transfected with ZFNs and pUC-*Nhe*I-1.5 kb revealed the level of HDR to be ~2.7% implying that

the efficiency of repair at the ZFN cut site is higher than previously thought. This indicates that the detection limit of the sensitivity assay when using genomic DNA is >3%. Figure 4.11 shows an alignment of two DNA sequences obtained by deep sequencing.

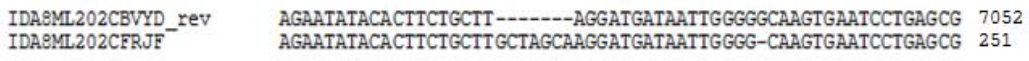


Figure 4.11: Sample sequences of 7,303 PCR products from CFTE cells transfected with ZFNs and pUC-CFTE-*Nhe*I 1.5 kb confirming incorporation of *Nhe*I site from donor into genome.

4.6 Design and construction of mini-gene construct to correct >80% of CF-causing mutations by ZFN mediated HDR

Having proved the principle that it is possible to add a unique seven base pair sequence into the ZFN cut site, the possibility of recombining in a mini-gene donor was explored. It was decided to use the pUC-CFTE-NheI-1.5 kb plasmid as the backbone for the construct as this length of homology gave the only positive result.

4.6.1 *Design of mini-gene donor*

The aim was to construct a donor repair plasmid containing partial CFTR WT cDNA exons (10-24) with appropriate splice acceptor and poly-A sites, and also a GFP tag to allow visualisation of corrected cells. A T2A linker was used to link the CFTR cDNA and GFP sequences. The diagram in figure 4.12 depicts the hypothesis under test. Following HDR at the ZFN site in intron 9 with this mini-gene donor, endogenous exon 9 should be spliced together with the introduced WT exons10-24. If this occurs as planned, then splicing will result in a full length CFTR transcript that is under the control of the endogenous promoter thereby maintaining normal tissue-specific expression levels. The transcript should then be translated into a fully functional CFTR protein which can be trafficked to the apical membrane and allow correct Cl^- flux.

The T2A linker is positioned such that production of GFP should only occur if CFTR mRNA is produced first. This GFP protein should be present in all corrected cells and so should make detection of corrected cells easier. This experiment was based on the design used in Li et al. (2011) where they used a similar cDNA construct to target WT exons 2-8 to intron 1 of the F.IX gene (Li et al., 2011).

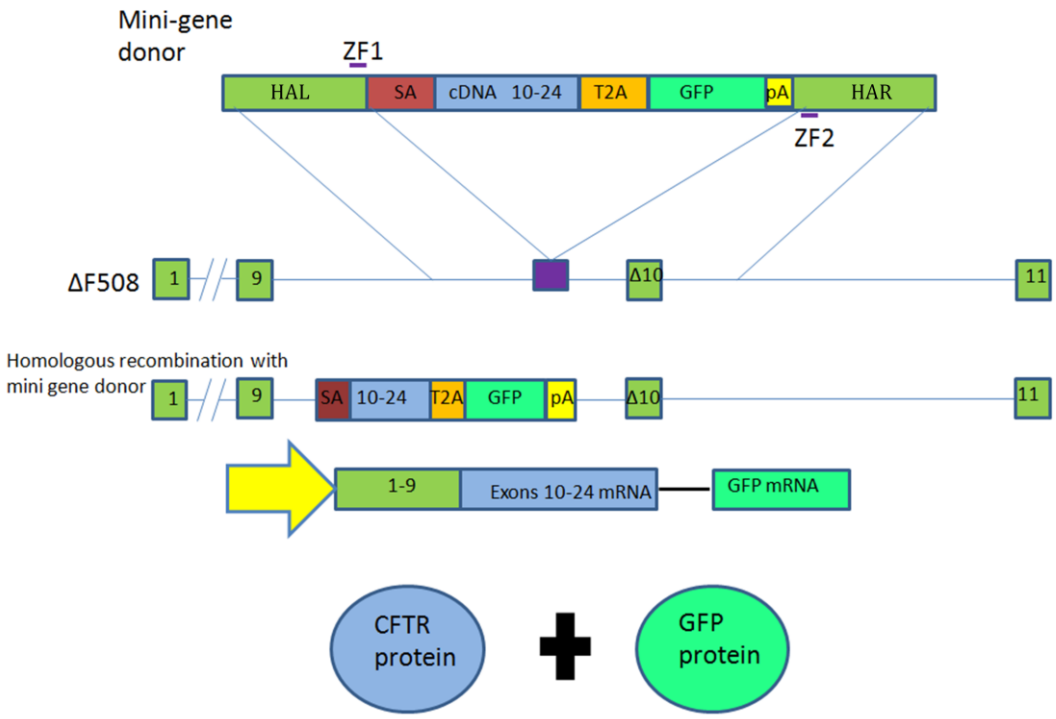


Figure 4.12: Schematic representation of what is expected to occur in the cell following HDR with mini-gene donor mediated by ZFN cleavage. Mini-gene donor is shown with cDNA exons 10-24 in blue, splice acceptor (SA) in red, T2A linker in orange, GFP in fluorescent green and 750 bp homology arms left and right (HAL, HAR) in green. Upon ZFN cleavage in intron 9 (purple box), homologous recombination should occur with the mini-gene donor to repair the DSB. This HR will result in recombination of the entire mini-gene construct into intron 9 of the CFTR gene. During mRNA production exons will be spliced together, exon 9 will splice into the WT cDNA of the mini-gene due to the presence of the correct splice acceptor to produce a full length and functional CFTR protein. The T2A linker allows production of GFP mRNA only following production of CFTR mRNA. If no CFTR is made then no GFP will be made. The final products are two separate proteins, CFTR and GFP.

4.6.2 *Construction of mini-gene donor*

Construction of the mini-gene donor required a number of cloning steps. Firstly, a synthetic polylinker was designed containing all necessary restriction enzymes for mini-gene assembly and cloned into the MCS of pUC19 to form pUC-linker. Then

each of the four parts of the mini-gene was amplified from the relevant DNA/plasmid using appropriate primers with added restriction sites. Cloning was carried out in "reverse" order with the GFP-pA fragment cloned first followed by T2A, partial cDNA and finally the SA fragment into the linker region of pUC. When all fragments were correctly assembled the entire mini-gene construct was excised from the pUC-linker plasmid and cloned into the *Nhe*I site of pUC-CFTE-*Nhe*I-1.5 kb which provides the 1.5 kb homology arms. This section will briefly describe each cloning step performed.

-Step 1: Cloning of pUC-Linker

Two 76 bp oligos, containing the nine restriction sites necessary for cloning and the extra bases needed for efficient binding of enzymes, were designed and synthesised. The oligos were annealed and cloned into sites *Xba*I and *Hind*III of pUC19. Figure 4.13 shows the sequence of one strand of the linker oligo with restriction sites underlined and relevant fragments to be cloned in bold above the sequence.

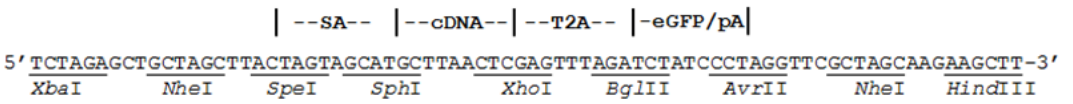


Figure 4.13: One strand of the linker oligo designed to clone the four mini-gene donor fragments. Restriction sites are underlined and mini-gene donor fragments are above the sequence in bold. SA: Splice Acceptor; cDNA: exons 10-24 of CFTR; T2A: linker; EGFP/pA: GFP sequence and polyA site.

-Step 2: Cloning of GFP and poly-A site into pUC-linker to create pUC-GFP/pA

As the *Bgl*II restriction site necessary for cloning the eGFP/pA fragment was present in the cDNA sequence it was necessary to clone the GFP and poly-A site into pUC-linker first so as to avoid later disruption of the cDNA fragment. The plasmid eGFP-N1 was used as a template to create the eGFP/pA fragment and was amplified by primers Fwd_eGFP/pA and Rev_eGFP/pA which introduced *Bgl*II and *Nhe*I restriction sites to the 5' and 3' ends of the 955 bp fragment respectively. The eGFP/pA fragment was then cloned into the *Avr*II and *Bgl*II sites of pUC-linker. *Avr*II and *Nhe*I produce the same overhangs, which allow ligation to be performed without disrupting downstream cloning steps which will need to use the *Nhe*I site.

-Step 3: Cloning of T2A linker into pUC-GFP/pA to create pUC-T2A-GFP/pA

The 57 bp T2A linker was designed and assembled from two single stranded oligos with restriction site compatible overhangs, *Xho*I at the 5' end and *Bgl*II at the 3' end, as shown in figure 4.14.

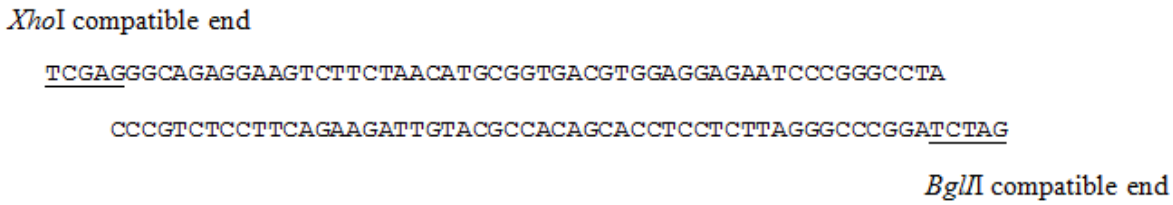


Figure 4.14: T2A linker sequence. Restriction site compatible ends are underlined.

This linker should allow production of the GFP mRNA following transcription of CFTR mRNA. The plasmid pUC-GFP/pA was digested with the same enzymes to produce the backbone to be used for ligation with the T2A insert.

-Step 4: Cloning of WT cDNA exons 10-24 and Splice Acceptor (SA) into pUC-T2A-GFP/pA to create pUC-SA-cDNA-T2A-GFP/pA

The strategy to assemble a splice acceptor (SA) sequence with exons 10-24 cDNA sequences is shown in figure 4.15a. It was decided to use the endogenous intron SA sequence. A ~ 200 bp sequence containing the SA and first ~175 base pairs of exon 10, including a *Sph*I site, was amplified from genomic HBE DNA to produce SA/ex10. Wild-type exons 10-24 were amplified as a ~4136 bp fragment from the pCEP-CFTR plasmid as shown. Relevant restriction sites were added to the fragments during PCR and used to clone the cDNA exons 10-24 and SA/ex10 fragments into pUC-T2A-GFP/pA. Both fragments are shown in figure 4.15b.

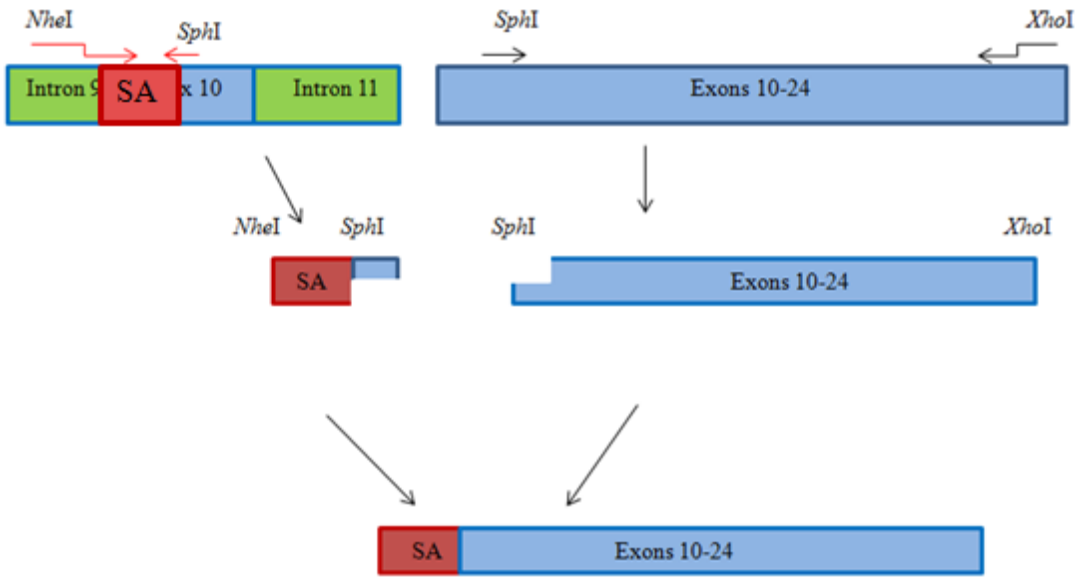


Figure 4.15a: Schematic representation of cloning strategy to clone SA/ex10 in-frame with exons 10-24. SA in red is amplified from HBE DNA with primers that add a *Nhe*I restriction site to the 5' end and include the endogenous *Sph*I site in exon 10. Exons 10-24 are amplified from the pCEP-CFTR plasmid starting within exon 10 close to the *Sph*I site. Following digest with indicated enzymes, the overhangs produced allow ligation of the SA and cDNA fragments so that they are in-frame and will allow correct splicing of CFTR exons.

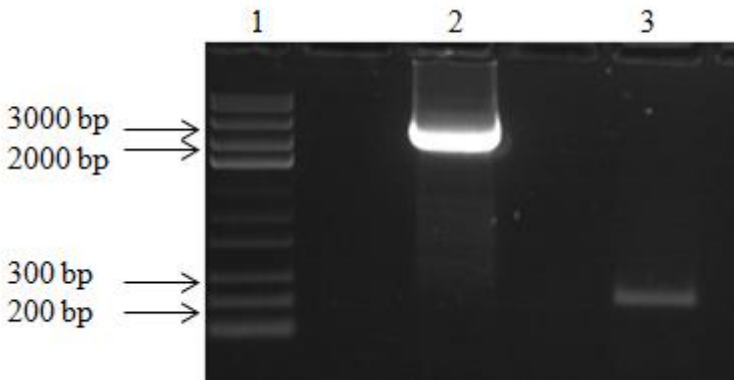


Figure 4.15b: cDNA and SA/ex10 fragments for mini-gene cassette. Lane 1: Mass ruler express reverse ladder with relevant bands indicated with black arrows. Lane 2: cDNA fragment amplified from the pCEP-CFTR plasmid. Lane 3: SA/ex10 fragment amplified from HBE genomic DNA. Both bands were gel extracted and purified.

The cDNA fragment, (figure 4.15b, lane 2) following purification and digestion, was ligated into the *Sph*I and *Xho*I sites of the linker region of pUC-T2A-GFP/pA to create pUC-cDNA-T2A-GFP/pA.

pUC-cDNA-T2A-GFP/pA was digested with *Spe*I and *Sph*I to create a backbone into which the SA/ex10 fragment was ligated using the *Nhe*I (same overhang as *Spe*I) and *Sph*I sites. Sequencing data and digest analysis confirmed insertion of the cDNA and SA fragments thus completing the mini-gene cassette and yielding pUC-SA-cDNA-T2A-GFP/pA.

-Step 5: Cloning of SA-cDNA-T2A-GFP/pA mini-gene cassette into pUC-CFTE-*Nhe*I-1.5 kb

The entire mini-gene cassette was excised from of the pUC backbone as a 4091 bp *Nhe*I fragment. The pUC-CFTE-NheI-1.5kb plasmid was also digested with *Nhe*I to yield a backbone of 4001 bp and de-phosphorylated to avoid self-ligation before insertion of the mini-gene cassette. *Hind*III was used to analyse clones derived from the reaction. Bands at 5048, 1898, 829 and 334 bp indicate that the mini-gene cassette has been successfully cloned into the pUC-CFTE-*Nhe*I-1.5kb plasmid to yield the complete mini-gene donor. Figure 4.16 shows the correct banding pattern in lane 3.

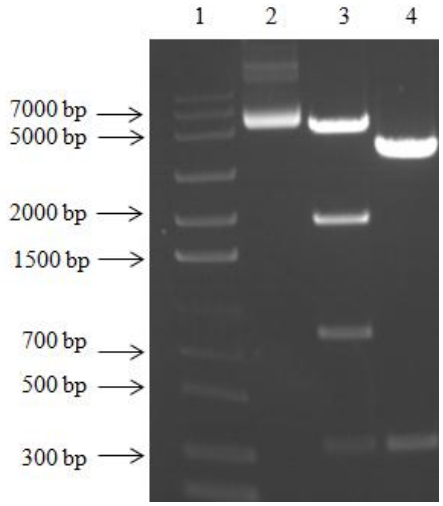


Figure.4.16: Digest analysis of potential clones derived from ligation of mini-gene cassette and pUC-CFTE-*Nhe*I-1.5 kb. Lane 1: Mass ruler express reverse ladder with relevant bands indicated with black arrows. Lane 2: Uncut control of sample in lane 3. Lane 3: Sample derived from ligation digested with *Hind*III which yields the expected banding pattern of 5048, 1898, 829 and 334 bp. Lane 4: pUC-CFTE-*Nhe*I-1.5 kb digested with *Hind*III to show banding pattern if mini-gene cassette has not been inserted correctly. There are two *Hind*III sites present in this plasmid yielding two bands of 3667 and 334 bp.

The final plasmid pUC-SA-cDNA-T2A-GFP/pA (pUC-mini-gene) contains the entire mini-gene cassette described in section 4.6 flanked by 750 bp homology arms in a pUC backbone. This plasmid will be used to test the hypothesis that our ZFNs can induce HDR of exons 10-24 of the CFTR gene in the CFTE cell line.

4.7 Detection of HDR with CFi9-L^{EL}/CFi9-R^{KK} ZFNs and pUC-mini-gene donor.

As discussed in previous chapters PCR based screening methods had been successfully used to determine if repair had occurred, but had not been successful in accurately quantifying the level of repair. The design of the mini-gene donor construct means that GFP should only be produced if CFTR is corrected by HDR. If a cell is corrected it should produce both CFTR and GFP, allowing an indirect measurement of HDR. Thus, measuring the % of GFP positive cells would give an estimate of the level of repair. Furthermore, this marker allows parameters of transfection such as the ratio of ZFN:donor plasmid or use of histone deacetylase inhibitor drugs like SAHA, to be varied and any subsequent changes in GFP easily monitored. In addition, cells could be enriched by FACs to perform functional assays to determine if CFTR in corrected cells was trafficked to the apical membrane rather than trapped in the ER.

In this section I will discuss attempts to detect HDR of the mini-gene.

4.7.1 *Transfection with ZFNs and mini-gene donor produce GFP positive cells.*

The hypothesis is based on the design of the mini-gene donor and is that GFP positive cells indicate that HDR has occurred. Firstly, following transfection, cells were analysed by fluorescent microscopy. However, it is important to rule out false positive results, not just with ZFN only or donor only treated samples but for ZFN and donor treated samples also. Ultimately, HDR will need to be confirmed by PCR and sequencing results.

CFTE cells were typically transfected with CFi9-L^{EL} and CFi9-R^{KK} and mini-gene donor as outlined in table 4.2. The ratio of ZFN:donor was varied to test which ratio gives rise to GFP positive cells.

Assay / **Plasmid** (µg)	**1**	**2**	**3**	**4**	**5**	**6**
Ratio ZFN:Donor	N/A	N/A	N/A	1:1	1:3	3:1
CFi9-L^{EL}	0	1.5	0	1	0.5	1.5
CFi9-R^{KK}	0	1.5	0	1	0.5	1.5
pUC-mini-gene donor	0	0	3	2	3	1
eGFP	3	0	0	0	0	0
pcDNA 3.1	1	1	1	0	0	0

Table 4.2: Amount of plasmids transfected for CFi9-L^{EL}/CFi9-R^{KK} HDR assays with mini-gene donor.

Fluorescent microscopy was used to visualise transfected cells 3, 6, 14, 21 and 33 days post-transfection. GFP positive cells were detected in all wells transfected with ZFNs and mini-gene donor, with the highest amount seen at a ratio of 1:3 ZFN:donor. This ratio was used for the remainder of experiments, see figure 4.17. GFP was only detected in cells transfected with both ZFNs and donor, no GFP was observed in ZFN only or donor only transfected cells. The visualisation of GFP implied that correction has occurred and the CFTR protein was also being produced. However, as this observation could also be explained if the mini-gene had integrated in a random manner close to a promoter, it was necessary to confirm the recombination events by PCR. Prior to using PCR-based assays to verify, it was decided to try and enrich the GFP positive population.

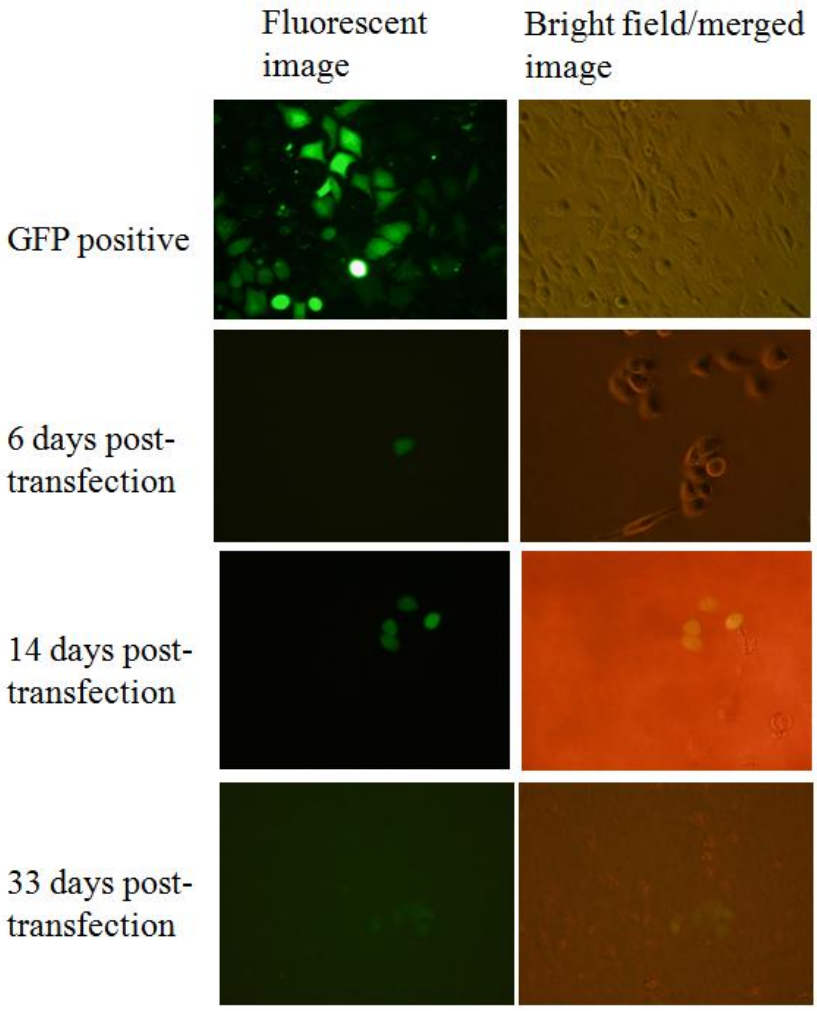


Figure 4.17: Fluorescent images of treated cells as outlined in table 4.2. GFP positive control: cells transfected with eGFP plasmid only, (table 4.2, assay 1). Images of cells taken at various time points, transfected with ZFN and mini-gene donor at a ratio of 1:3.

4.7.2 *Enrichment of GFP positive cells*

Several attempts were made to enrich GFP positive cells with the view to increase the potential readout for PCR and FACs analysis. In order to enrich corrected cells, the overall density of cells was decreased by transferring treated cells into culture plates with a smaller surface area/well to promote growth. Cells were subjected to serum starvation in order to induce a stress response to promote corrected cell growth over uncorrected cells (section 2.23).

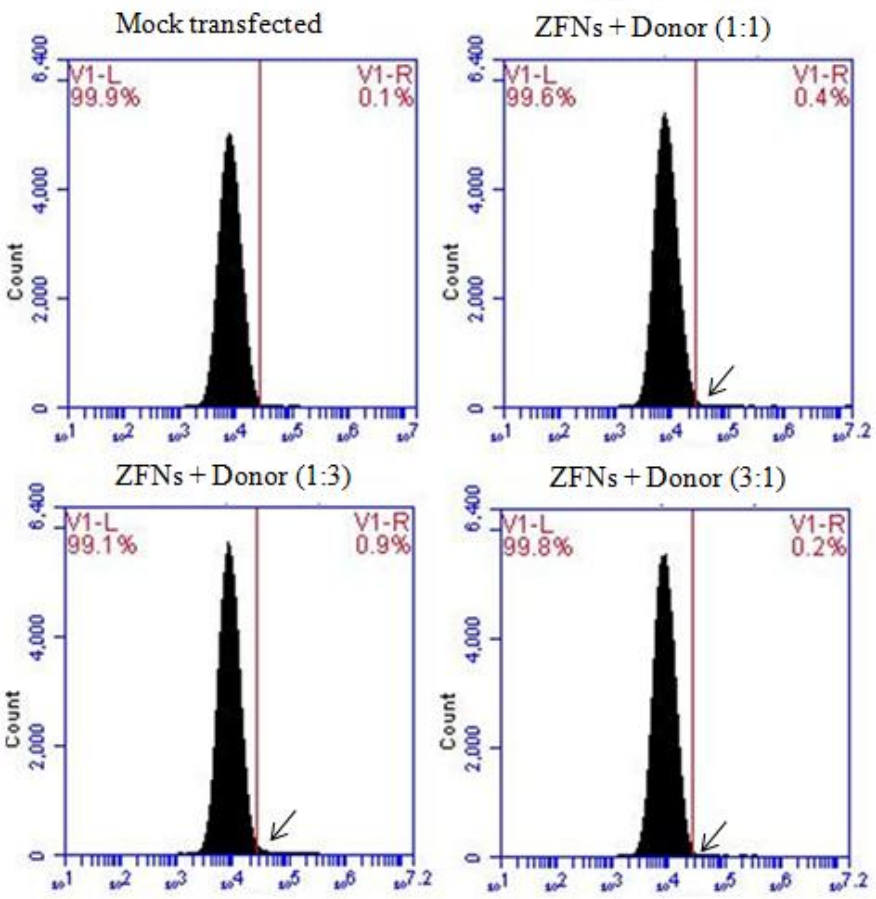
Although there was a slight increase in GFP positive cell number following enrichment, it could not be further increased or maintained. Furthermore,

throughout the enrichment process the brightness of GFP was notably fainter than pEGFP control transfected cells. We hypothesised that following correct HDR the GFP protein will be under the control of the naturally weak endogenous CFTR promoter and this may be the reason for the weak GFP signal which may be difficult to detect using a conventional fluorescent microscope. The observation that GFP is detected up to 29 days post-transfection is consistent with the long-term correction known to occur following mini-gene HDR (Li et al., 2011).

4.7.3 *Fluorescent Activated Cell Sorting (FACs) analysis*

FACs was used to detect an increase in GFP expression in a treated population of cells. A shift in GFP above that seen in mock transfected cells would give an indication of the number of cells that had been corrected by HDR. Typically 500 µl of total cells were trypsinised and DNA extracted three days post-transfection and used for FACs analysis as described in section 2.20. Data obtained is provided table 4.3B below.

A)



B)

Sample	% GFP positive cells
Mock transfected	0.3 ± 0.2
ZFNs and donor (1:1)	0.4 ± 0.3
ZFNs and donor (1:3)	0.8 ± 0.6
ZFNs and donor (3:1)	0.1 ± 0.1

Table 4.3: A) Representative sample FACs traces. **B)** Average percentage of GFP positive cells ± SEM obtained from FACs analysis performed on 70,000 cells. N=3

As stated above, if the mini-gene had correctly recombined, GFP would be expected to be expressed at low levels. Thus, the fluorescence was not expected to be much above background. As shown in table 4.3, cells transfected with ZFNs and donor at a ratio of 1:3 did contain a greater number of GFP positive cells than mock transfected population. Indeed, a simple analysis of the data in table 4.3B suggests 0.5% of cells express GFP which is roughly the level of HDR that might be expected. However, given the low level of green fluorescence detected, it was decided to try and increase the level of GFP expression to see if it was possible to improve the signal to noise ratio (see below).

4.7.4 *Does addition of SAHA increase HDR as determined by FACs analysis?*

The US FDA approved Histone Deacetylase (HDAC) inhibitor, SAHA, has been shown to increase stabilisation, trafficking and activity of ΔF508 CFTR by decreasing the expression of HDAC7. As a consequence SAHA was shown to increase the yield of CFTR trafficking as well as increase CFTR mRNA (Hutt et al., 2010). It was hypothesised that addition of SAHA would increase the expression levels of CFTR mRNA and in turn GFP mRNA, thus increasing signal to noise ratio, which should improve the accuracy of the FACs analysis. SAHA was added 24 hrs post-transfection at a concentration of 5 μM and removed either 24 or 72 hrs later. Cells incubated with SAHA for longer than 72 hrs all underwent apoptosis

(data not shown). Cells incubated for the shorter time of 24 hrs survived and GFP expression was measured by FACs. A shift in GFP was observed in cells treated with SAHA compared to untreated cells, however, there was no difference in GFP levels in SAHA treated-mock transfected cells compared to SAHA treated-ZFN and mini-gene donor transfected cells (data not shown).

These observations, along with problems experienced in enrichment suggested the possibility that the GFP signal may not be a consequence of mini-gene HDR as proposed in figure 4.12, rather it may be a consequence of low level random integration.

4.7.5 *PCR-based assays to determine if HDR occurred at the correct location*

A convincing readout from GFP experiments described above was not obtained, so it was decided to perform PCR to detect the presence of the mini-gene construct at the ZFN cut site in the CFTR gene. If HDR had occurred as predicted, then WT exons 10-24 would have been inserted into the ZFN cut site in intron 9. Figure 4.18 shows where the four sets of primers used to try and detect this bind.

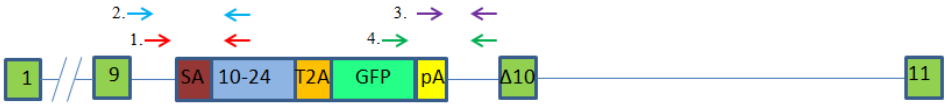


Figure 4.18: Four sets of primers were used to produce a PCR product that would confirm whether HDR with the MG donor had occurred at the correct location. Firstly detection of recombination was investigated at the 5' end of the construct using primer sets 1 (Fwd_gDNA and Rev_ex11) and 2 (Fwd_recombination and Rev_ex11). When these proved unsuccessful, the 3' end was investigated using primer sets 3 (Fwd_GFP_HR_1 and Rev_ex10) and 4 (Fwd_GFP_HR_2 and Rev_ex10).

Repeated attempts with the four primer sets failed to reveal evidence of HDR. The inability to obtain a PCR product to show correct recombination of the mini-gene donor into the ZFN cut site, strongly suggested that mini-gene HDR had not occurred as predicted and that any GFP expression was most likely a consequence of low level random integration.

4.7.6 *Next Generation Sequencing sheds light on mini-gene HDR*

One round of deep-sequencing was carried out on a DNA sample extracted from cells transfected with ZFNs and mini-gene donor (1:3). If recombination has been successful at the ZFN cut site the sequence should now contain a 7 bp tag, GCTAGCT (underlined T is from the linker described in fig. 4.13), followed by a Δ32 bp due to the way the mini-gene donor construct was cloned into the *Nhe*I site of pUC-CFTE-*Nhe*I-1.5 kb. This would be followed by corrected CTT, mini-gene exons 10-24 and GFP. Deep-sequencing data obtained from a single sample is shown in figure 4.20.

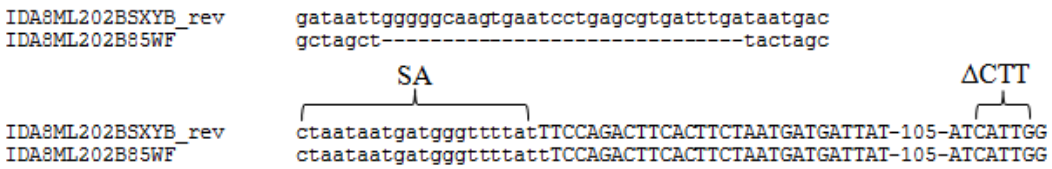


Figure 4.20: Sample sequences of PCR products from CFTE cells transfected with ZFNs and pUC-mini-gene confirming incorporation of 7 bp tag and 32 bp deletion from donor into genome. Approximately 1% of sequences had this sequence. Lower case indicates intronic sequence and higher case indicates exon 10. SA sequence and position of CTT addition is indicated.

The sequencing data clearly shows the expected 7 bp tag and 32 bp deletion following HDR with the mini-gene donor. However, from this data it can be seen that recombination only occurred at this site as no CTT was detected in these altered sequences. If the DSB had been repaired fully with the mini-gene donor the CTT sequence should be in this data but it is not present.

It was originally hypothesised that recombination would occur as depicted in figure 4.12. However evidence to support this has not been obtained in this body of work. However, as discussed below, this has allowed the design of an alternative strategy.

4.8 Conclusions

The aim of this chapter was to explore one of the two alternative strategies generated from results obtained in chapter three. Thus, rather than design a new pair of ZFNs closer to CTT it was decided in this chapter to take advantage of the

efficient cleavage activity of our ZFNs and use them to explore an alternative strategy using a CFTR cDNA mini-gene.

Firstly, to test that HDR could occur at the cut site and to determine what length of homology arm was optimal, three new *Nhe*I donor plasmids were constructed containing different lengths of homology. HDR was detected when cells were transfected with ZFNs and pUC-CFTE-*Nhe*I-1.5 kb as shown by the 564 bp *Nhe*I-specific PCR product (section 4.5). Several attempts were made to quantify the level of HDR by in-house detection assays with no success. It was thought that either the detection system was not sensitive enough, as was seen previously in chapter three, or else the level of HDR was below the limit of detection estimated for these methods.

When samples were subject to NGS analysis, it was confirmed that the 7 bp tag had been added and occurred at a level of ~2.7%. This value is slightly higher than the 2% HDR value for CTT correction using the same ZFNs. However, as it was only possible to perform both NGS experiments on DNA from a single transfection, it was not possible to determine the significance of this difference. Likewise, whilst tempting to speculate the increased value of HDR in this experiment is because the *Nhe*I tag is recombined at the ZFN cut site, compared with the CTT recombination some 203 bp downstream of the cut site, more experiments would be needed to confirm this. Also, a formal comparison of homology arm length and the role of AAV-ITRs would require additional experiments to be performed. Regardless, the *Nhe*I specific PCR successfully detected *Nhe*I in genomic DNA from cells transfected with ZFNs and pUC-CFTE-*Nhe*I-1.5 kb which proved the principal that a unique sequence can be introduced into the ZFN cut site in our cell line.

Having shown that HDR occurs at the cut site, the second aim of this chapter was to correct CF-causing mutations using our ZFNs and mini-gene construct. To date, only two groups have successfully reported mini-gene recombination (Lombardo et al., (2007) and Li et al., (2011)). The strategy described herein required a more complex donor than these previously described and involved a number of intricate steps. Following transfection with ZFNs and pUC-mini-gene, GFP was detected using fluorescent microscopy up to 33 days post-transfection. The hypothesis based

on our mini-gene design is that GFP production would indicate that HDR had occurred, and that WT CFTR was being produced. However, the difficulties experienced when attempts were made to enrich the GFP positive cells, pre-treat with SAHA to increase expression or test for HDR by FACs, and failure of PCR to generate a product providing evidence of HDR as predicted indicated a problem in the design strategy and suggested that the GFP visualised may be an artefact. The lack of HDR as fully predicted was confirmed by data obtained from deep sequencing of a sample treated with ZFNs and pUC-mini-gene.

However, this NGS data generated very valuable insight regarding what was happening following DNA cleavage by the ZFNs and also helped explain why we never detected a PCR product from our GFP positive samples. It also confirmed that the GFP detected in FACs and fluorescent microscopy was an artefact. If HDR with the mini-gene occurs there will be: 1) a 7 bp tag followed by a deletion of ~32 bp due to the mini-gene disrupting the sequence as it is cloned into the *Nhe*I site of pUC-CFTE-*Nhe*I-1.5 kb and 2) CTT will be present in exon 10. The NGS data did confirm that HDR had occurred between our donor plasmid and CFTR sequence, as seen from the 7 bp tag and Δ32 bp; however, homologous recombination did not incorporate the entire donor construct as deduced from the absence of CTT in exon 10.

The design of our construct may be the limiting factor here as a number of features of the construct may actually be preventing the HDR process as required. Firstly, the SA fragment contains the exact same sequence as the endogenous SA in CFTE cells. Secondly, the left homology arm contains mutant exon 10 which together with the SA fragment may provide enough homology to repair the break without using the right homology arm. Figure 4.21 depicts what is thought to be happening with our mini-gene construct. The dashed red lines show the possible homology arms being used to repair the ZFN induced DSB. The overlap of sequence in the construct with sequence in the right homology arm could be the reason why HDR does not extend to include the entire mini-gene. However, HDR did occur as shown in our sequencing data by the introduction of the 7 bp tag and 32 bp deletion, section 4.7.6.

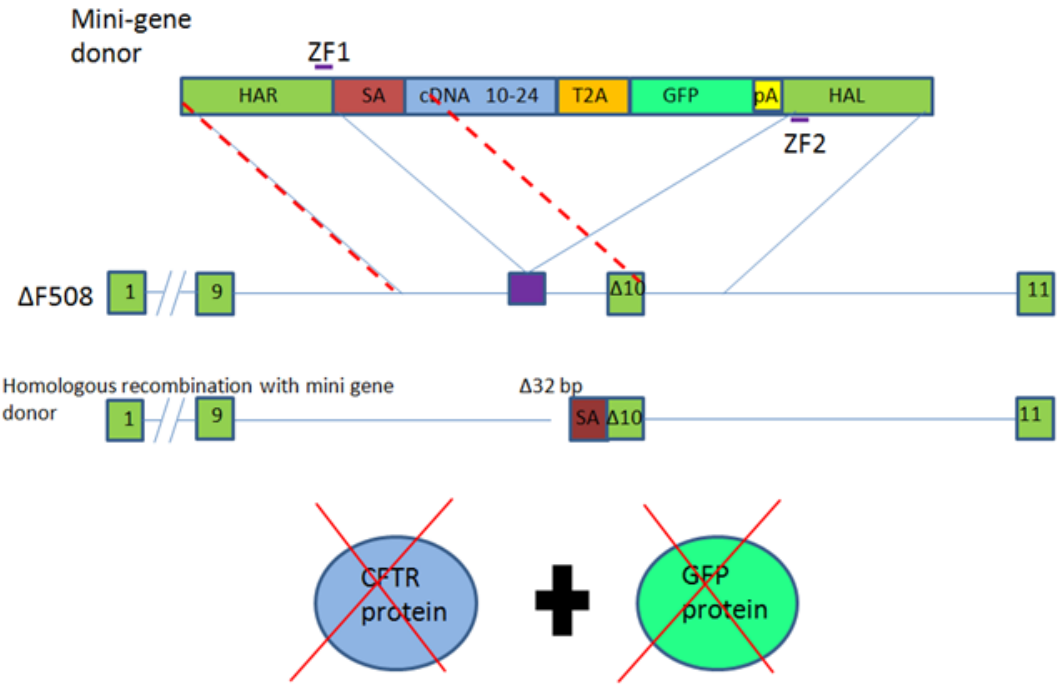


Figure 4.21: Schematic representation of HDR theory following ZFN mediated HDR with pUC-mini-gene. Red dashed lines indicate what homology is being used to repair ZFN induced DSB. Following this method of HDR, there will be a 32 bp deletion before the SA sequence.

To overcome this problem in the future, a SA fragment will be designed to contain the essential splice machinery sequence but different intervening sequences so as to avoid homology overlap. Also the cut site will be moved further upstream in intron 9 to avoid any overlapping homology arms, i.e. exon 10 coding sequence. NGS will be used to confirm correct recombination with the whole mini-gene. This new cut site will be CRISPR based or possibly double nickase based (see final discussion).

Another issue that will have to be examined is the extent of donor sequence incorporated at the target. This depends on a number of factors; 1) the extent of synthesis after strand invasion, 2) whether the invading 3' end had been resected and 3) the direction of mismatch repair in the heteroduplex formed by annealing (Carroll, 2011). These factors still remain largely unknown however the lengths of conversion tracts have been measured and were found to be quite long in Drosophila where several kilobases of donor was incorporated (Nassif et al.,1994), whereas in mammalian cells only short tracts of between 100 and 200 bp from the

break were incorporated after repair (Elliott et al., 1998). To test this, new donor plasmids could be constructed containing multiple restriction site markers at various distances from the cut site, much like the pITR-donor-XC plasmid. Co-transfection of these plasmids with ZFNs and analysis by NGS would then give an idea of the effects of distance on HDR efficiency. To avoid too much sequence divergence, multiple plasmids with one restriction site marker in each could be used. If only 100-200 bp is the limit for repair it implies our mini-gene may not be fully incorporated as it is over 200 bp. However, the study by Li et al., (2011) on which we based our design, showed successful incorporation of their mini-gene donor and production of the desired protein, factor IX (Li et al., 2011) which suggests much longer tracts can be incorporated when using mini-genes. Indeed, up to 8 kb can be incorporated (Moehle et al., 2007). Of note, the homology arms to incorporate F.IX targeted an intronic region. This further suggests that we need to target further upstream in intron 9 to exclude exon 10.

To summarise, successful HDR was achieved using the *Nhe*I donor, proving the principle that recombination can occur at the ZFN cut site. Data from the mini-gene experiments indicate that an intron target will be more likely to succeed and this will require CRISPR gRNAs, which will be described in chapter five.

5 CRISPR/Cas9/gRNA system as an alternative to ZFNs to mediate HDR in the presence of donor plasmids

5.1 Aims:

From the conclusions drawn in chapter three it was hypothesised that distance from the cut site to the mutation may be a limiting factor determining the level of HDR. Therefore, the second option to be explored was to use other gene-editing tools, more amenable to changing location, such as the CRISPR/Cas9/gRNA system. This system is the latest tool for genome editing and as described in section 1.7.9, there are very simple guidelines to follow in order to create a CRISPR gRNA to target a specific sequence. Two new sequences were chosen to be cleaved and were of the form $GN_{20}GG$, one in exon 10, 85 bp upstream of the CTT deletion and one in intron 9 close to the original ZFN target site (see figure 5.1). CRISPR ex10 will be used in conjunction with the original ΔF508 donors, pITR-donor and pITR-donor-XC, while CRISPR in9 will be used with pUC-mini-gene donor to mediate HDR.

5.2 Objectives:

- Design and construct two CRISPR guide RNAs to target CFTR
- Determine the level of cleavage/NHEJ repair at the target sites
- Detect HDR events using CRISPR/Cas9/gRNA system and donor plasmids

5.3 Design of two CRISPR gRNAs to target and cleave CFTR

In order to target a specific sequence a simple 20 bp oligo is designed according to $GN_{20}GG$ and cloned into the U6 expression plasmid followed by co-transfection along with the Cas9 protein. Two sites were chosen within the CFTR gene (see figure 5.1). The G's necessary for gRNA binding are provided by the sequence being targeted and so are not included in the 20 bp oligo.

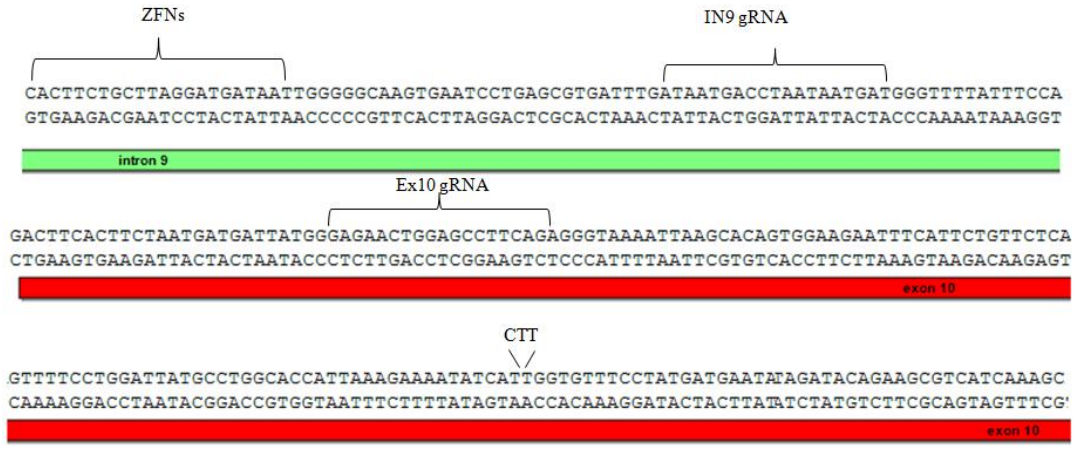


Figure 5.1: Schematic representation of a section of the CFTR sequence displaying the position of the target sites for the ZFNs, intron 9 gRNA and exon 10 gRNA and the location of the CTT deletion

For cloning into the U6-expressing plasmid, *Bse*RI overhangs were added to the 20 bp oligos (see figure 5.2).

CRISPR In9 Target oligos

ATAATGACCTAATAATGATGT

GCTATTACTGGATTATTACTA

CRISPR Ex10 Target oligos

GAGAACTGGAGCCTTCAGAGT

GCCTCTTGACCTCGGAAGTCT

Figure 5.2: Two 20 bp sequences for the two CRISPR gRNAs are shown with the added bases necessary for *Bse*RI cloning underlined.

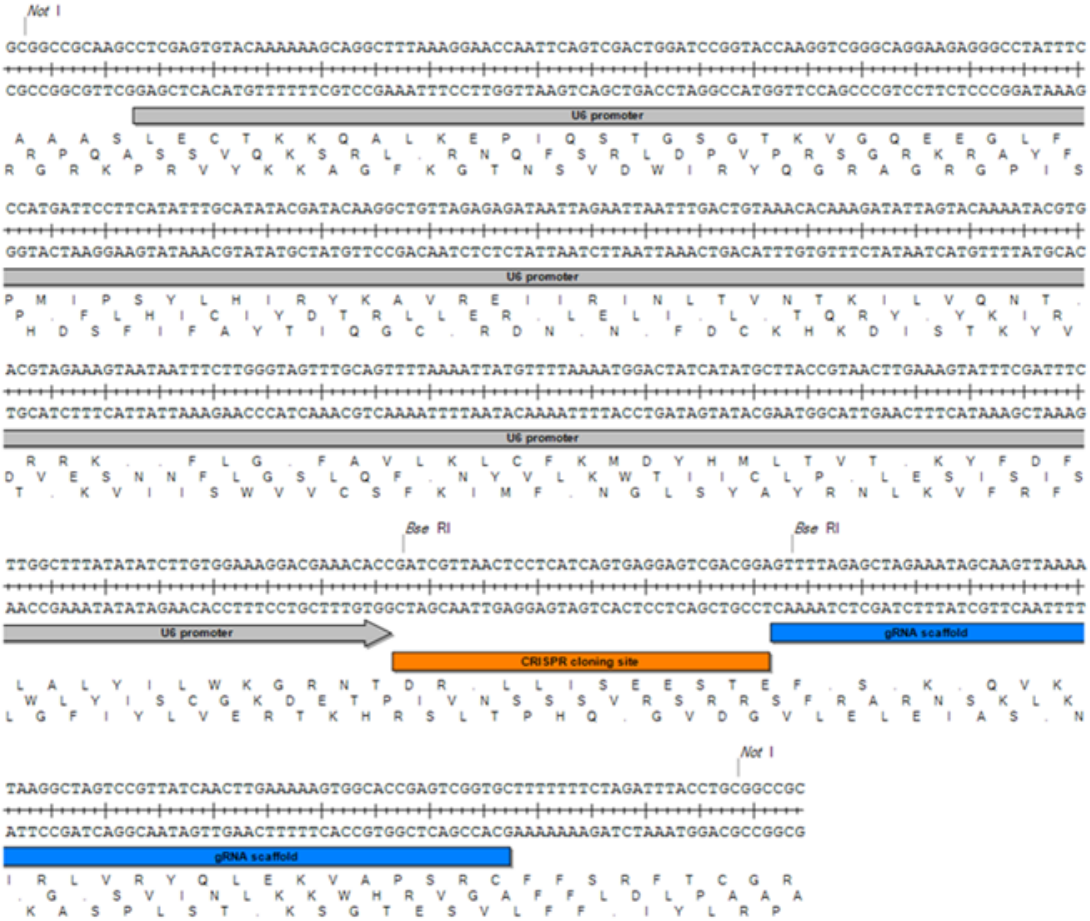


Figure 5.3: pEX-A plasmid. U6 promoter is in grey, CRISPR cloning site in orange and gRNA scaffold is in blue.

5.3.1 *One-step cloning of CRISPR IN9 and CRISPR EX10*

A plasmid was designed to provide the gRNA scaffold and promoter necessary for CRISPR gRNA expression. A 474 bp sequence was designed, synthesised and cloned into the *Not*I sites of the pEX-A plasmid. This sequence consisted of the U6 promoter, gRNA scaffold and *BseR*I restriction sites necessary for CRISPR gRNA cloning (see figure 5.3).

The two short single stranded oligos (figure 5.2) were annealed and cloned into the *BseR*I sites of the pEX-A plasmid by the golden-gate method. Potential clones were digested with *Xho*I and *BseR*I. Successful ligation results in the *BseR*I sites being destroyed and so only *Xho*I will cleave to produce a single band. Figure 5.4 shows the results of digestion.

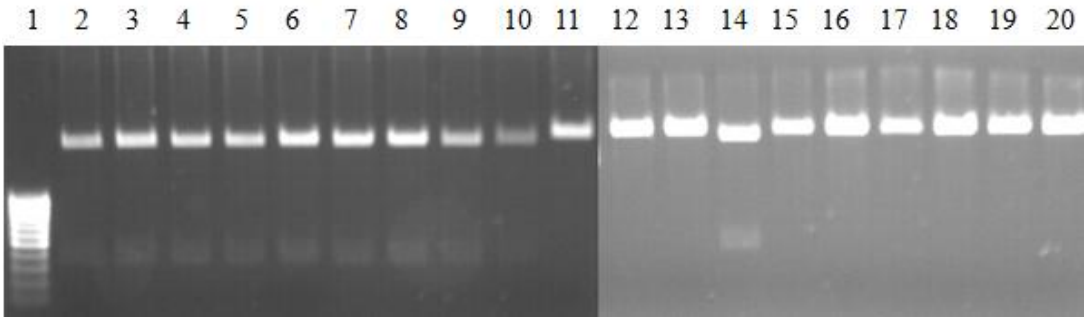


Figure 5.4: Digest analysis of clones derived from golden gate cloning of CRISPR gRNAs into pEXA. Lane 1: Low range Ladder. Lanes 2-11: Potential clones derived from CRISPR in9 ligation. Lanes 12-20: Potential clones derived from CRISPR ex10 ligation. Successful cloning produces a single band as seen in lane 11 and in all lanes except 14 in lanes 12-20.

One out of ten clones were successful for CRISPR in9, while eight out of nine was successful for CRISPR ex10.

5.4 Determination of cleavage efficiency of CRISPR IN9 and CRISPR EX10

The CRISPR/Cas system can be implemented in mammalian cells by co-expressing the bacterial Cas9 nuclease along with the guide RNA. Cells were transfected with ZFNs, CRISPR in9/hCas9, CRISPR ex10/hCas9 or CRISPR ex10/hCas9D10A (see table 5.1). The mutant Cas9 cleaves only one strand of DNA and so acts as a nickase which have been shown to increase the rate of HR by decreasing NHEJ.

Assay / **Plasmid (µg)**	1	2	3	4
CRISPR in9	0.0	0.0	0.5	0.0
CRISPR ex10	0.5	0.0	0.0	0.5
hCas9	0.5	0.0	0.5	0.0
hCas9 D10A	0.0	0.0	0.0	0.5
CFi9-L^{EL}	0.0	0.5	0.0	0.0
CFi9-R^{KK}	0.0	0.5	0.0	0.0
pcDNA3.1	3.0	3.0	3.0	3.0

Table 5.1: Amounts of plasmids used in transfection of CFTE cells

Proof-reading PCR using Pfx and CFTR-NHEJ primers was performed on total DNA extracted 72 hrs post-transfection and NHEJ events were detected using the T7E1 assay. Samples were run on a 5% polyacrylamide gel. Results obtained are shown in figure 5.5. Successful cleavage of DNA with CRISPR ex10 should produce bands at ~450, 290 and 140 bp. CRISPR in9 should produce bands of ~430, 230 and 200 bp while ZFNs produce bands of ~430, 250 and 180 bp.

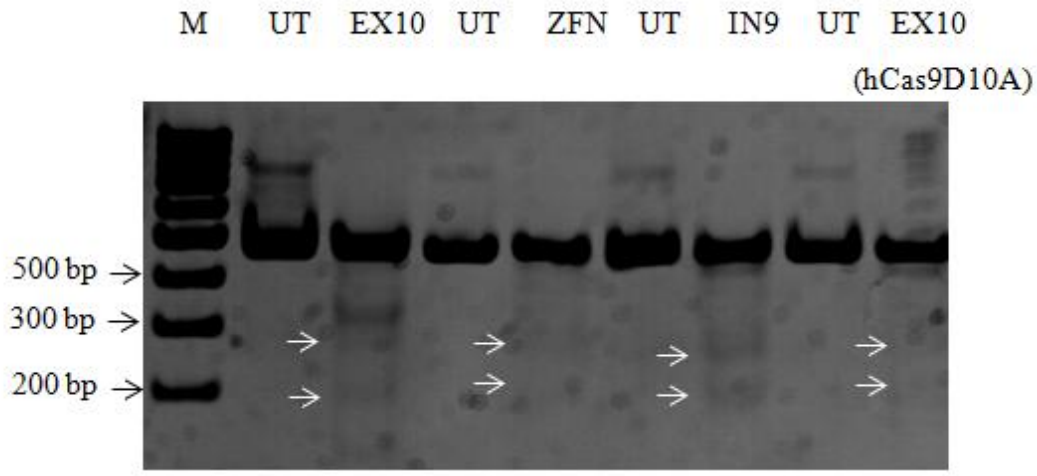


Figure 5.5: T7E1 assay results on a 5% polyacrylamide gel. M: Mass ruler express reverse ladder. UT: samples untreated with T7E1. EX10: CRISPR gRNA ex10/hCas9 treated with T7E1. IN9: CRISPR gRNA in9/ hCas9 treated with T7E1. EX10/hCasD10A: CRISPR gRNA ex10/ hCasD10A treated with T7EI

Results from the T7E1 assay in figure 5.5 show that both CRISPR gRNAs are able to recognise the CFTR gene and cleave DNA to produce a DSB as indicated by the desired banding pattern (shown by white arrows) when co-transfected with hCas9/hCas9D10A plasmid. Densitometry analysis of the relevant bands was used to determine the level of cleavage of the two gRNAs. CRISPR ex10 with hCas9 cleaved at 16%, ZFNs at 3%, CRISPR in9 at 5% and CRISPR ex10 with hCas9 D10A at 3% which is lower than CRISPR ex10 and hCas9 as expected.

5.4.1 *NGS data for CRISPR cleavage*

NGS data on DNA treated with CRISPR ex10/Cas9 and CRISPR in9/Cas9 provided further evidence that cleavage had occurred. The data shows that following cleavage, a mix of deletions and insertions are produced by NHEJ. Below in figure 5.6 is some of the deep sequencing data obtained from DNA treated with CRISPR ex10/hCas9. The level of NHEJ for this particular sample was 15.5%. This value correlates closely to the value obtained using the T7E1 assay. To determine if this correlation is reproducible more experiments will need to be performed.

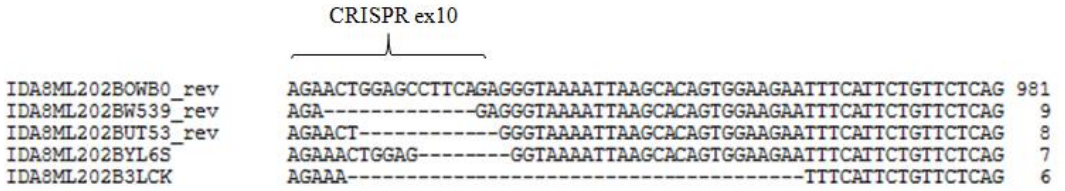


Figure 5.6: Sample sequences of 1,346 PCR products from CFTE cells transfected with CRISPR ex10 and Cas9 confirming cleavage of the genome. Deletions and insertions indicate successful cleavage of CRISPR and Cas9 plasmids at the desired location. Sequence for CRISPR ex10 target site indicated.

The CRISPR in9 gRNA also showed evidence of gene targeting as seen in figure 5.7. Large deletions and some small insertions are evident in the deep sequencing data obtained from a single sample. The level of cleavage in this sample was 2.3%.

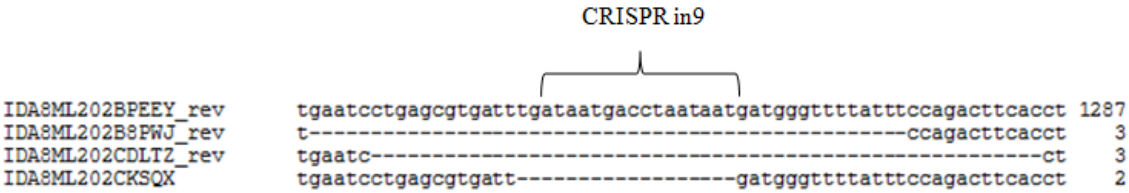


Figure 5.7: Sample sequences of 1,609 PCR products from CFTE cells transfected with CRISPR in9 and Cas9 confirming cleavage of the genome. Deletions and insertions indicate successful cleavage of CRISPR and Cas9 plasmids at the desired location. Sequence for CRISPR in9 target site indicated.

Of note both CRISPR create indels approximately 3-4 bases upstream of the PAM sequence as reported by Mali et al. (2013). This is the first evidence of CFTR gene targeting with CRISPR gRNAs and the levels of NHEJ detected are comparable to levels reported by other groups targeting other genes.

5.5 CRISPR/Cas9 system can mediate HDR in the presence of appropriate donor plasmids.

Having shown that both CRISPR gRNAs were able to efficiently cleave DNA, the ability of CRISPR gRNA-induced DSBs to mediate HDR with donor plasmids was investigated.

5.5.1 *CRISPR EX10 can mediate HDR with pITR-donor-XC*

CRISPR ex10 cleaves roughly 85 bp upstream of the CTT deletion (see figure 5.1) and so was co-transfected with pITR-donor-XC. A typical transfection of CFTE cells is outlined in table 5.2.

Assay **Plasmid (µg)**	1	2	3	4
CRISPR ex10	0.5	0.0	0.5	0.5
hCas9	0.0	0.0	0.5	0.0
hCas9D10A	0.5	0.0	0.0	0.5
pITR-donor-XC	0.0	3.0	3.0	3.0
pcDNA3.1	3.0	1.0	0.0	0.0

Table 5.2: Amount of plasmids transfected for CRISPR EX10 HDR assays with pITR-donor-XC

Total DNA was extracted and analysed for HDR using essentially the same PCR-based method described in section 3.6.1. Briefly, a 2.6 kb PCR was generated to exclude plasmid carry over using Fwd and Rev_gDNA primers. PCR products are shown in figure 5.8.

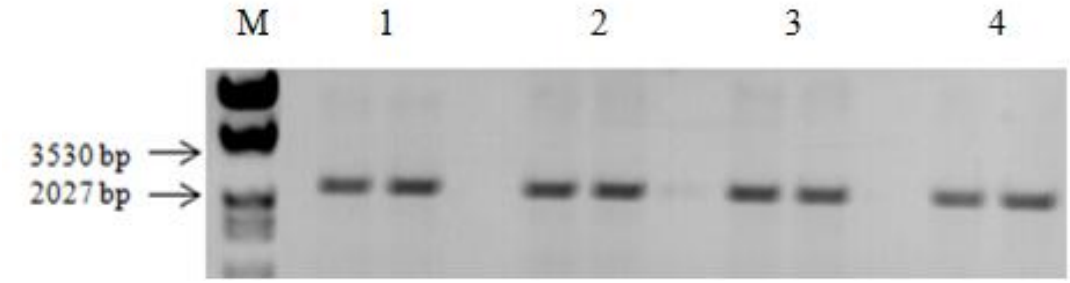


Figure 5.8: 2.6 kb PCR products generated from DNA samples outlined in table 5.2. Lane 1: *Eco*RI/*Hind*III lambda ladder. Lanes 2-5: 40 µl of PCR product generated from samples 1-4 in table 5.3 respectively. The entire PCR product was extracted to obtain a high yield for digest analysis.

The PCR products in figure 5.8 were extracted and purified, then digested with *Xho*I and *Cla*I. Following digest, as well as the original 2.6 kb band, some PCR products should be cleaved to produce bands; 2040 and 606 bp for *Xho*I digests or 2157 and 489 bp for *Cla*I digests if HDR has been successful. As a positive control to show that PCR products can be successfully digested following purification, a 432 bp fragment was amplified using pITR-donor-XC as a template across the region containing the two restriction sites. The expected two bands of 290 and 140 bp were produced following *Xho*I digest and bands of 411 and 23 bp following *Cla*I digest.

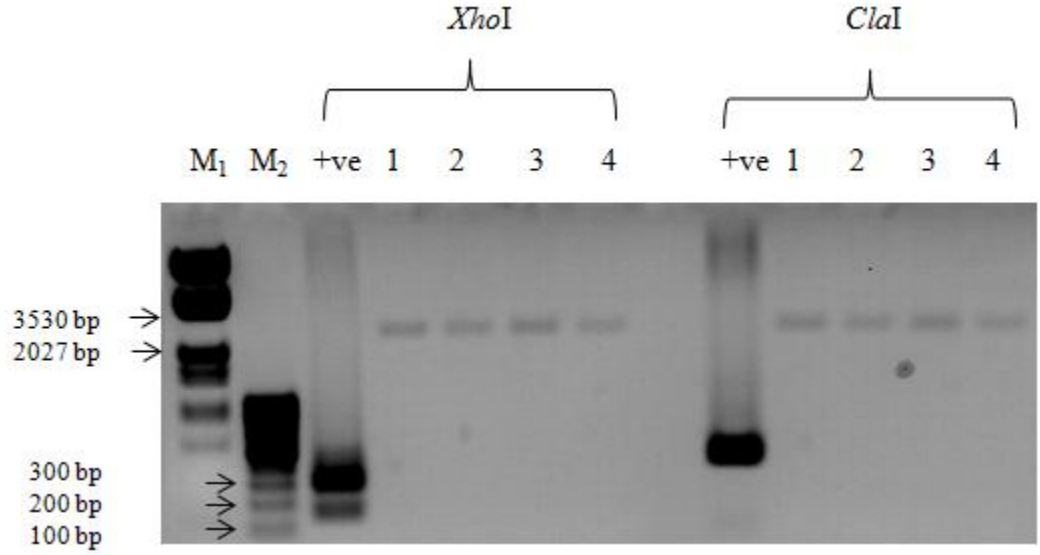


Figure 5.9: Digest analysis of 2.6 kb PCR products in figure 5.8. Lane M1: *Eco*RI/*Hind*III lambda ladder. M2: Low range ladder. +ve: Positive control digested with *Xho*I. 1-4: Samples 1-4 as in table 5.2 digested with *Xho*I. +ve: Positive control digested with *Cla*I. Lanes 1-4: Samples 1-4 as in table 5.2 digested with *Cla*I.

Figure 5.9 shows the results from these digests. All 2.6 kb PCR products remained undigested. Based on data from chapter 3, the inability to detect digestion suggested HDR, if occurring was <3%.

5.5.2 *NGS data for CRISPR ex10 and pITR-donor-XC*

Although HDR could not be detected by the PCR-based method above, it was decided to analyse a sample of CRISPR ex10/hCas9/pITR-donor-XC treated cells to see if HDR could be detected by NGS. A 432 bp PCR product was generated and analysed. A sample of the data obtained is shown in figure 5.10.

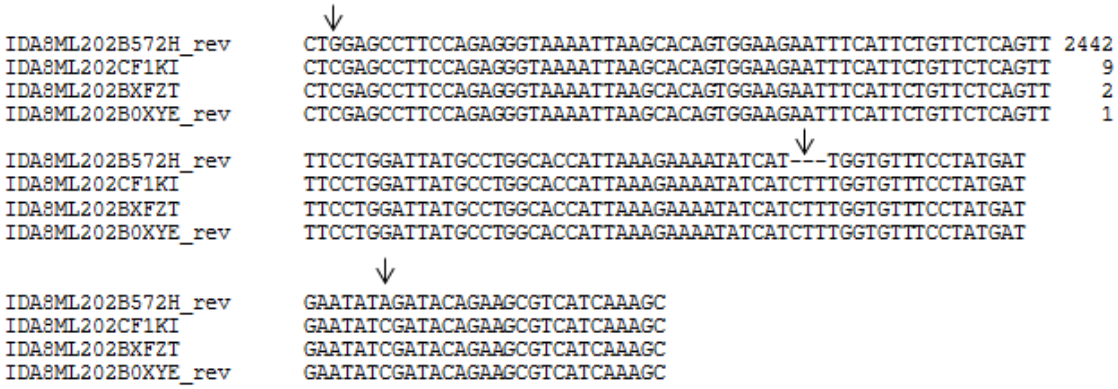


Figure 5.10: Sample sequences of 3,244 PCR products from CFTE cells transfected with CRISPR ex10/hCas9 and pITR-donor-XC confirming incorporation of *Xho*I, CTT and *Cla*I sites from donor into genome. The arrows indicate the base pair changes following HDR with pITR-donor-XC. G->C for *Xho*I site, CTT addition and A->C for *Cla*I site.

The data in figure 5.10 shows that CRISPR ex10 did mediate HDR with pITR-donor-XC. The black arrows indicate the base changes, G->C for *Xho*I, addition of CTT and A->C for *Cla*I. The overall level of HDR was determined to be 1% in this sample. Interestingly, it was found that some sequences contained the CTT addition and the *Cla*I site but no *Xho*I. This is odd as the *Xho*I site is located at the CRISPR ex10 cut site. Perhaps strand invasion during repair excluded the *Xho*I site in the donor. The sequence data for this particular sample is evidence that HDR has definitely occurred and shows that the CRISPR/Cas9/gRNA system can mediate repair. This is the first evidence that CRISPR gRNAs can mediate HDR in a CF cell line.

5.5.3 *CRISPR IN9 gRNA HDR with the pUC-mini-gene donor*

CRISPR in9 cleaves 40 bp downstream of the ZFN cut site (see figure 5.1) and so was co-transfected with the mini-gene donor plasmid. A typical transfection of CFTE cells with CRISPR in9 gRNA, hCas9 (WT/nickase) and pUC-mini-gene is outlined in table 5.3. ZFNs were used as controls.

Assay / **Plasmid (µg)**	1	2	3	4	5	6
CRISPR in9	0.0	0.5	0.0	0.0	0.5	0.5
hCas9	0.0	0.5	0.0	0.0	0.5	0.0
hCasD10A	0.0	0.0	0.0	0.0	0.0	0.5
CFi9-L^{EL}	0.0	0.0	0.0	0.5	0.0	0.0
CFi9-R^{KK}	0.0	0.0	0.0	0.5	0.0	0.0
pUC-mini-gene	0.0	0.0	3.0	3.0	3.0	3.0
eGFP N1	3.0	0.0	0.0	0.0	0.0	0.0
pcDNA 3.1	1.0	3.0	1.0	0.0	0.0	0.0

Table 5.3: Amounts of plasmids used in transfection of CFTE cells

These experiments were performed before full analysis of the ZFN and mini-gene experiment described in chapter 4 was completed. As before, a low level of GFP could be observed in transfections with CRISPR gRNAs, Cas9 and donor but not with donor only (data not shown). However, as shown in chapter 4, repeated attempts to generate a PCR product to confirm correct recombination could not be obtained; leading to the conclusion that HDR had not occurred completely as originally planned. However, when the DNA from cells treated with CRISPR IN9 gRNA, hCas9D10A and mini-gene donor was used for deep sequence analysis, both the GCTAGCT tag and expected deletion of ~32 bp were detected, but no HDR occurred far enough to incorporate the WT cDNA.

The data gathered demonstrates the ability of both our CRISPR/Cas9 gRNAs to mediate HDR, and confirms independently our conclusion that HDR with this existing mini-gene will need to be re-designed. However, the fact that two CRISPR gRNAs were designed so easily and shown to work efficiently to mediate both NHEJ and HDR (in the presence of a donor plasmid), even without full optimisation, is very encouraging for this study.

5.6 Conclusions:

The aim of chapter five was to establish the latest gene editing tool, CRISPR/Cas9 as a cheap and efficient alternative to our ZFNs to enable relocation of the cut site as distance was determined to be a limiting factor for HDR in chapter three. CRISPR gRNAs are easily designed and can be constructed within one week, compared to the costly design methods involved in making ZFNs.

Here, we designed two gRNAs targeting intron 9 and exon 10 of CFTR. Both CRISPR gRNAs were able to efficiently cleave the desired DNA sequence as determined by T7E1 data (section 5.4) and deep sequencing results (section 5.4.1). The overall cleavage efficiency of CRISPR in9 was found to be ~2-3% while CRISPR ex10 cleaved at a rate of ~15-16%.

When these CRISPR gRNAs were co-transfected with the necessary hCas9 plasmid and the relevant donor, HDR events were observed. CRISPR ex10 used in conjunction with pITR-donor-XC resulted in HDR at a rate of 1% as determined by deep sequencing of a single sample. Again no HDR was detected when using our PCR and digest analysis assays showing that these assays are not sensitive enough to measure HDR and so new assays will need to be developed in the future.

The CRISPR in9 gRNA performed precise HDR by incorporating a 7 bp tag and 32 bp deletion at an efficiency of ~1%.

This data shows that HDR does occur in CFTE cells when transfected with our CRISPR/Cas gRNAs and donor plasmids.

For future work the obvious next step is to re-design the repair strategy so as to avoid any exon 10 in the homology arms of the mini-gene and to target an intronic region further upstream of the original ZFN cut site which will be made much easier using CRISPR/Cas technology. As discussed in chapter four, we initially created the mini-gene construct with the homology arms described because our ZFNs cleave close to exon 10 and so to include enough homology it was necessary to include this sequence. It is now possible to target a much wider range of sequences of the form $GN_{20}GG$ using CRISPR gRNAs and so we will target one of the 161 sites of this form in the intron 9 sequence further upstream of exon 10 so as to avoid any crossover of homology with construct sequence. Then following cleavage, HDR should occur with the mini-gene cassette which will be incorporated into the intronic region to produce full length CFTR protein following correct splicing.

The CRISPR/Cas system can also be used for the repair strategy using obligate ligation-gated recombination discussed in detail in the final discussion. This strategy involves adding identical CRISPR gRNA sites to the 5' and 3' end of the mini-gene cassette without homology arms. Then following CRISPR gRNA cleavage of the genome (same sequence as mini-gene gRNAs) and the mini-gene plasmid, the resulting overhangs should be compatible and so the mini-gene cassette should be inserted into intron 9 of CFTR by ligation and rely on the NHEJ repair pathway which is six times more efficient than HR. This strategy was used successfully to insert a 15 kb gene cassette into human cell lines earlier this year (Maresca et al., 2013).

The data presented in this chapter clearly demonstrates the ability of our CRISPR/Cas gRNAs to mediate HDR in a CF cell line to correct CF-causing mutations.

6 Discussion:

Cystic Fibrosis (CF) is caused by mutations in the CFTR gene which leads to defective chloride and bicarbonate transport and dysregulation of sodium transport via loss of epithelium sodium channel (ENaC) inhibition by CFTR protein. The most serious manifestation of CF is reduced mucociliary clearance in the lung. This results in the development of viscous hypoxic mucus which promotes recurrent *P. aeruginosa* infections and concomitant loss of lung function. Damage to the lung architecture occurs from both *P. aeruginosa* virulence factors and increased neutrophil mediated necrosis resulting in inflammation.

CF presents as a target for novel therapeutics due to reports that low level correction of CFTR transcript levels (5-10%) (Ramalho et al., 2002) or of defective cells (25%) (Zhang et al., 2009) may be sufficient to restore a normal phenotype. It is one of the most common monogenic diseases and since the CFTR gene was cloned in 1989 many research efforts have been made worldwide to understand and correct this disease. All CFTR mutants produce mRNA transcript and aside from class I mutations, all other CFTR mutants produce CFTR protein. Both the transcript and the protein have been targeted by novel drugs such as PTC124, VX-770 and VX-809.

VX-770 offers help to a small subset of CF patients (4-5%) that have the G551D class III mutation and proved the principles that direct targeting of CF at its root cause and restoring CFTR protein function can be clinically beneficial (Van Goor et al., 2009). Interestingly, there are reports of increased efficacy of this drug in homozygous patients versus heterozygotes for G551D (Harrison et al., 2013). However, the drug is of no benefit to the majority of CF patients who have mutations from the other classes. Other novel compounds targeted to a subset of class I mutants (PTC124) and class II mutants (VX-809) have shown promise in small scale studies but results from large scale studies were disappointing and did not meet target endpoints. A combination therapy with VX-770 and VX-809 is currently entering phase III clinical trials. Another compound VX-661 is also under trial in combination with VX-770 and results released in April, 2013 from a phase II trial involving 128 ΔF508 homozygote CF patients showed significant improvements in lung function in those who received the highest dose of treatment

(cff.org). Development of CF drugs is a main interest of major pharmaceutical companies at present and given the proof-of-principle from VX-770, other drug companies such as Pfizer, Genzyme and Novartis are all striving to find new drugs to treat the disease. This interest offers new hope for CF patients.

In contrast to development of small molecules, gene therapy trials have attempted to complement the defective endogenous CFTR gene with a WT cDNA copy. This method, if successful, would benefit all CF patients since it would restore functional CFTR protein to all treated cells. It would however, require repeat dosing for the lifetime of the CF patient. Despite numerous clinical trials over the last 20 years in CF patients no clinical breakthrough has been observed. The early trials did demonstrate that viral delivery of transgenes to patients could be achieved. However, the AAV and AdV vector used did not effectively deliver the CFTR transgene to a significant number of cells, nor was the response sustainable with repeat dosing. The low level of transduction of these viral vectors was attributed to the lack of specific AdV/AAV receptors on the apical surface of the cell and the onset of an immune response to these virus vectors made repeat dosing difficult. To avoid repeat dosing an option would be to target progenitor cells since the corrected cells at the lung surface will be lost quickly, this would allow long term correction. Recently, a study has shown that both rAAV1 and rAAV5 are capable of preferentially transducing conducting airway epithelial progenitor cells that had the ability to clonally expand *in vivo* following lung injury in a mouse model (Liu et al., 2009).

Non-viral vectors are being explored as an alternative delivery method to viral vectors. The UK CF gene therapy consortium is continuously conducting CF gene therapy clinical trials and has developed a CFTR-expressing plasmid complexed with a cationic lipid. GL67A was the most effective agent, and the CFTR expression vector has been modified to induce prolonged expression (up to at least 2 months) and rendered CpG free to avoid an immune response. A current phase IIB clinical trial is assessing the clinical benefit of repeat dosing of nebulised CFTR gene therapy over a 12 month period (Alton et al., 2013).

An alternative approach to cDNA addition is to try and correct the endogenous defective gene in the affected tissue using nuclease-mediated homology directed repair (HDR). Designer nucleases, such as zinc finger nucleases (ZFNs) and the RNA-guided CRISPR/Cas9 system, can be used to create site-specific double-stranded breaks (DSB). This DSB can then be repaired using the cells own repair machinery, (homologous recombination) and a donor DNA sequence to result in the incorporation of the correct sequence into the genome. This method will still require efficient delivery of the nucleases and donor DNA sequence to progenitor cells to provide long term correction of the lung. However, once the endogenous gene of a progenitor cell is corrected it is permanent for the lifetime of that cell, meaning there is no requirement for repeat administration to that particular cell making it an attractive alternative to cDNA addition. The repair of an endogenous locus also conserves tissue specific and temporal control of expression as well as expression of natural splice variants.

This body of work provides the first clear evidence that ZFNs and CRISPR gRNAs can successfully induce HDR between a donor repair plasmid and genomic DNA in a CF cell line. In chapter three, correction of the most common CF causing mutation, ΔF508, was achieved using our ZFNs and pITR-donors at a level of ~2% of transfected cells. This level of HDR was relatively successful for a first attempt and an obvious next step was to design a new pair of ZFNs closer to the CTT deletion to improve on this level as distance was determined to be a limiting factor. However, as this would be time consuming and ZFN design has a low success rate two different options were subsequently explored in this thesis:

1) the position of the ZFNs in intron 9 was exploited and used as a site for recombination of a mini-gene construct comprising exons 10-24 from CFTR cDNA (with appropriate splice acceptor and poly A sites) to allow production of full length corrected CFTR mRNA and enable the correction of a much wider range of CF-causing mutations as described in chapter 4.

2) Another gene-editing tool, CRISPR/Cas9 RNA-guided system, which is more amenable to changing location, was explored as described in chapter 5.

The principle of recombination at the ZFN cut site was established by successfully introducing a 7 bp tag into this site in chapter four. NGS data from subsequent transfections with the ZFNs and mini-gene donor showed that some HDR had occurred with the mini-gene donor as determined from introduction of the 7 bp tag and 32 bp deletion as predicted but CTT remained uncorrected suggesting that HR had not extended far enough to incorporate the entire construct. We know now from this data that in order for our mini-gene strategy to be successful, the cut site needs to be moved further into intron 9 see figure 6.1.

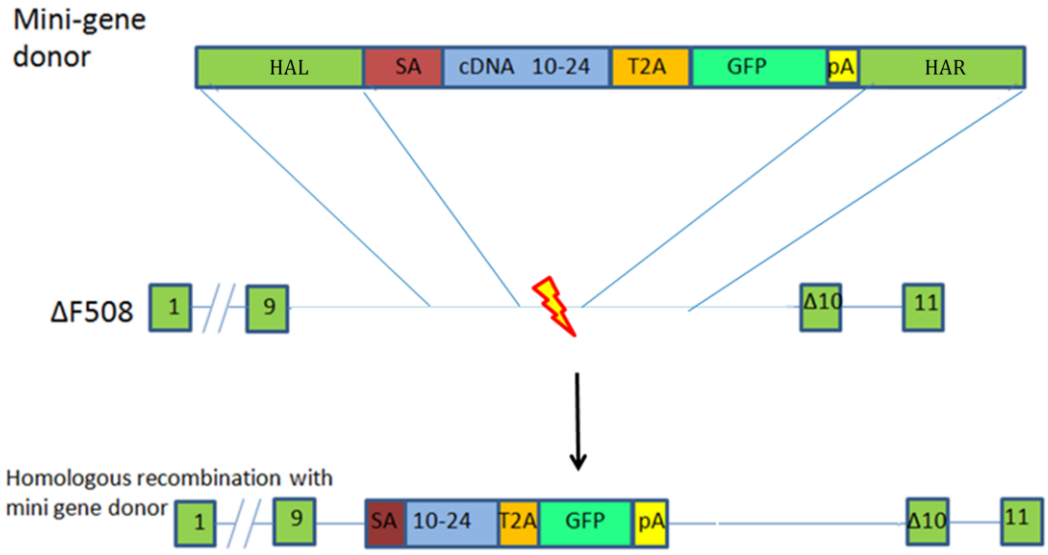


Figure 6.1: Schematic representation of new mini-gene repair strategy.

This new mini-gene construct will be designed to contain homology arms to intron 9 only and a new cut site will be introduced using CRISPR gRNAs. The ease of design and use of CRISPR/Cas was displayed in chapter five whereby two new target sites were successfully cleaved and used with existing donor repair plasmids to induce HDR to repair the ΔF508 mutation in the CF cell line. This is the very first evidence of successful use of the CRISPR/Cas9 system to correct the ΔF508 mutation in a CF cell line.

As an alternative to the strategy above, in future studies the use of the so-called ObLiGaRe strategy to integrate the mini-gene into the target site by NHEJ will be explored. This method takes advantage of the overhangs produced following cleavage. If each end of the mini-gene contains the same CRISPR gRNA target

sequence to that in the genome then upon successful cleavage of all three sites by CRISPR/Cas9, complementary overhangs will be produced allowing direct ligation of the mini-gene into the cut site. Using this strategy eliminates the need for homology arms and relies on the NHEJ pathway which occurs six times more frequently than the HR pathway.

To determine whether distance from cut site is a limiting factor in HDR, new donors could be constructed containing multiple restriction site markers at various distances from the cut site, much like the pITR-donor-XC plasmid. Co-transfection of these plasmids with ZFNs and analysis by NGS would then give an idea of the effects of distance on HDR efficiency. To avoid too much sequence divergence, multiple plasmids with one restriction site marker in each could be used.

To be of future clinical benefit, the possibility of off-target events will need to be assessed. As discussed in section 1.9, off-target cleavage can lead to mutagenesis and cancers. Similar sequences, differing from one or more base pairs, to the one being targeted are potential off-target sites, therefore a number of modifications have been developed to try and prevent the *Fok*I domain of ZFNs and TALENs from off-target cutting. The *Fok*I domain of the ZFNs used in this project were modified by a former colleague in such a way so that they can only cleave as heterodimers and a subsequent toxicity assay showed them to be non-toxic in cells (Lee et al., 2012). It will be necessary in the future to look at the incidence of off-target sites induced by our ZFNs. A number of methods have been developed to do this. An *in vitro* cleavage site selection assay was developed by Pattanayak et al., (2011) to identify sequences from a large partially degenerate library, based on intended target DNA sites that can be cleaved by ZFNs targeting the CCR5 gene. They identified 36 potential off-target sites and analysis of these sites in the human genome revealed that nine off-target sites were cleaved (Pattanayak et al., 2011). Another method exploits the incorporation of integrase-deficient lentivirus DNAs into nuclease-induced DSBs to map ZFN cleavage sites in human cells. Gabriel et al., (2011) applied this approach to the same CCR5-targeted ZFNs characterised in the Pattanayak study and identified four off-target genomic sites (Gabriel et al., 2011). Only one site was common in both of these studies suggesting that neither method can identify all possible off-target sites. A recent study has shown the *in*

silico method to provide an improved approach to screen the genome for potential off-target sites. Looking at the same ZFNs as the previous two groups, they identified the previously described sites as well as dozens of additional sites (Sander et al., 2013). Therefore, it may take a combination of methods in order to identify all possible sites.

As for the new CRISPR/Cas9 system, there have been varying reports regarding specificity. A recent study suggests that CRISPR gRNAs can be highly active even at imperfectly matched DNA interfaces (Fu et al., 2013). If these nucleases are to be of clinical benefit in the future then the specificity will need to be improved. An obvious approach would be to increase the length of the target site; however, studies have shown that increasing the guide sequence mediates similar levels of modification and does not improve targeting specificity (Ran et al., 2013). The same group went on to explore the potential of Cas9 nickases to increase specificity. They reasoned that two Cas9-nicking enzymes directed by a pair of gRNAs targeting the opposite strands of a target locus could mediate DSBs and reduce off-target activity. The double nickase technique was able to generate similar levels of on-target modifications as produced with the WT Cas9. However, when they looked at the same sites which had been cleaved off-target by WT Cas9 they found that double nickase did not generate detectable cleavage, suggesting that this technique can potentially reduce off-target events. Our future studies on CF will explore the use of the double nickase technique.

In our approach to correct CF causing mutations, the use of positive selection was avoided so that the donor plasmids would only contain the correct sequence of the genomic DNA being corrected. This makes enrichment and screening of cells that have been corrected difficult, hence one reason why the mini-gene was tagged with a GFP sequence. If patients' own cells were to be treated *ex vivo* then a selection cassette could be useful in order to target and enrich corrected cells. Conventional selection cassettes such as antibiotic resistance genes are flanked by loxP or Flp/FRT sites so that upon correction these cassettes can be excised from the genome. This presents its own problems in that small ectopic sequences that remain can interfere with the transcriptional regulatory elements of surrounding genes (Meier et al., 2010). A recent study combined ZFNs and *piggyBac* technology to

correct a point mutation in human iPS cells. *PiggyBac* works by converting selection cassettes into transposons and so enables the removal of cassettes flanked by *PiggyBac* inverted repeats without leaving any residual sequences (Yusa et al., 2011). One could speculate that the same could be done using CRISPR gRNAs.

Having shown that ZFN and CRISPR-mediated HDR repair of ΔF508 is possible, there are a few experiments which appear to be the next obvious step. First of all, treated cells should be probed to see if the corrected CFTR is expressed at the apical cell surface. Antibody staining followed by Z-stacking multiple microscopic images can determine whether CFTR has translocated to the cell surface, something that ΔF508 CFTR cannot do. Some preliminary studies were performed using CFTR expressing plasmids and GFP to ascertain whether our CFTR antibodies work and whether it can be detected at the apical membrane, (see appendix). The next step would be to validate the activity of corrected CFTR using functional assays to measure: 1) Cl^- flux by patch clamp analysis to directly measure channel activity in individual cells and Ussing chambers to measure changes in short circuit current on a population of corrected cells and 2) ASL height using confocal microscopy. Both the Ussing chamber and ASL assays could be used to determine the minimum number of cells which need to be corrected to reach normal values.

Having shown herein that our nuclease-mediated HDR method can correct CFTR at the genomic level, the data gathered from the assays above will validate that this correction can restore normal Cl- flux to CF cells. The ability to restore normal Cl^- flux is crucial for establishing normal mucociliary clearance in the CF lung. This will ultimately decide whether this approach can be of clinical benefit. Correction of the underlying genetic defect in CFTR via nuclease-mediated HDR presents a significant first step towards achieving a realistic alternative to gene complementation.

7 Bibliography

Adib-Conquy, M., Pedron, T., Petit-Bertron, A.-F., Tabary, O., Corvol, H., Jacquot, J., Clément, A., et al. (2008). Neutrophils in cystic fibrosis display a distinct gene expression pattern. *Molecular medicine (Cambridge, Mass.), 14*(1-2), 36–44.

Aggarwal AK, Rodgers DW, Drottar M, Ptashne M, Harrison SC. (1988). Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science*. 11;242(4880):899-907.

Althaus, M. (2013). ENaC inhibitors and airway re-hydration in cystic fibrosis: state of the art. *Current molecular pharmacology*, *6*(1), 3–12.

Alton, E W F W, Stern, M., Farley, R., Jaffe, A., Chadwick, S. L., Phillips, J., Davies, J., 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. *The Lancet*, *353*.

Alton, Eric W F W, Boyd, a C., Cheng, S. H., Cunningham, S., Davies, J. C., Gill, D. R., Griesenbach, U., et al. (2013). A randomised, double-blind, placebo-controlled phase IIB clinical trial of repeated application of gene therapy in patients with cystic fibrosis. *Thorax*.

Andersen DH. (1939) Cystic Fibrosis of the pancreas and its relation to celiac disease: A clinical and pathologic study. *Am J Dis Child*. 1938;56(2):344–399.

Balough K, McCubbin M, Weinberger M, Smits W, Ahrens R, Fick R.(1995). The relationship between infection and inflammation in the early stages of lung disease from cystic fibrosis. *Pediatr Pulmonol.*20(2):63-70

Beck, S., Penque, D., Garcia, S., Gomes, a, Farinha, C., Mata, L., Gulbenkian, S., et al. (1999). Cystic fibrosis patients with the 3272-26A-->G mutation have mild disease, leaky alternative mRNA splicing, and CFTR protein at the cell membrane. *Human mutation*, *14*(2), 133–44.

Bedell, V. M., Wang, Y., Campbell, J. M., Poshusta, T. L., Starker, C. G., Krug, R. G., Tan, W., et al. (2012). In vivo genome editing using a high-efficiency TALEN system. *Nature*, *491*(7422), 114–8.

Beerli, R. R., & Barbas, C. F. (2002). Engineering polydactyl zinc-finger transcription factors. *Nature biotechnology*, *20*(2), 135–41.

Beumer, K., Bhattacharyya, G., Bibikova, M., Trautman, J. K., & Carroll, D. (2006). Efficient gene targeting in Drosophila with zinc-finger nucleases. *Genetics*, *172*(4), 2391–403.

Bibikova, M., Carroll, D., Segal, D. J., Jonathan, K., Smith, J., Kim, Y., Trautman, J. K., et al. (2001). Stimulation of Homologous Recombination through Targeted Cleavage by Chimeric Nucleases Stimulation of Homologous Recombination through Targeted Cleavage by Chimeric Nucleases. *Mol. Cell. Biol.* 21**:** 289–297.

Bitinaite, J., Wah, D. a, Aggarwal, a K., & Schildkraut, I. (1998). FokI dimerization is required for DNA cleavage. *PNAS 95*(18), 10570–5.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., et al. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science (New York, N.Y.)*, *326*(5959), 1509–12.

Bonfield,, T.L et al., 1999. Altered respiratory epithelial cell cytokine production in cystic fibrosis. *J Allergy Clin Immunol.* 104(1):72-8.

Bothmer, A., Robbiani, D. F., Di Virgilio, M., Bunting, S. F., Klein, I. a, Feldhahn, N., Barlow, J., et al. (2011). Regulation of DNA end joining, resection, and immunoglobulin class switch recombination by 53BP1. *Molecular cell*, *42*(3), 319–29.

Bothmer, A., Robbiani, D. F., Feldhahn, N., Gazumyan, A., Nussenzweig, A., & Nussenzweig, M. C. (2010). 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. *The Journal of experimental medicine*, *207*(4), 855–65.

Boyle, M. P., & De Boeck, K. (2013). A new era in the treatment of cystic fibrosis: correction of the underlying CFTR defect. *The Lancet Respiratory Medicine*, *1*(2), 158–163.

Cade, L., Reyon, D., Hwang, W. Y., Tsai, S. Q., Patel, S., Khayter, C., Joung, J. K., et al. (2012). Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. *Nucleic acids research*, *40*(16), 8001–10.

Cao, H., Yang, T., Li, X.-F., Wu, J., Duan, C., Coates, a L., & Hu, J. (2011). Readministration of helper-dependent adenoviral vectors to mouse airway mediated via transient immunosuppression. *Gene therapy*, *18*(2), 173–81.

Caplen NJ, Alton EW, Middleton PG, Dorin JR, Stevenson BJ, Gao X, Durham SR, Jeffery PK, Hodson ME, Coutelle C, et al. (1995). Liposome-mediated CFTR gene transfer to the nasal epithelium of patients with cystic fibrosis. *Nat Med.*1(1):39-46.

Carroll, D. (2011). Genome engineering with zinc-finger nucleases. *Genetics*, *188*(4), 773–82.

Cavazzana-Calvo, M., Lagresle, C., Hacein-Bey-Abina, S., & Fischer, and A. (2000). Gene Therapy of Human Severe Combined Immunodeficiency (SCID)-X1 Disease. *Science*, *288*(5466), 669–672.

Cavazzana-Calvo, Marina, Payen, E., Negre, O., Wang, G., Hehir, K., Fusil, F., Down, J., et al. (2010). Transfusion independence and HMGA2 activation after gene therapy of human β-thalassaemia. *Nature*, *467*(7313), 318–22.

Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. a, et al. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic acids research*, *39*(12), 82.

Chapman, J. R., Taylor, M. R. G., & Boulton, S. J. (2012). Playing the end game: DNA double-strand break repair pathway choice. *Molecular cell*, *47*(4), 497–510.

Chastre, E., Gioia, Y. Di, Barbry, P., Mornet, E., Fanen, P., Champigny, G., Emami, S., et al. (1991). Functional Insertion of the SV40 Large T Oncogene in Cystic Fibrosis Intestinal Epithelium. *J. Biol. Chem.* 266: 21239

Chen, F., Pruett-Miller, S. M., Huang, Y., Gjoka, M., Duda, K., Taunton, J., Collingwood, T. N., et al. (2011). High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nature methods*, *8*(9), 753–5.

Choo, Y. E. N., & Klug, A. (1994). Toward a code for the interactions of zinc fingers with DNA : Selection of randomized fingers displayed on phage. *PNAS 91*, 11163–11167.

Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A. J., et al. (2010). TAL Effector Nucleases Create Targeted DNA Double-strand Breaks. *Genetics*.

Chu, C. S., Trapnell, B. C., Murtagh, J. J., Moss, J., Dalemans, W., Jallat, S., Mercenier, a, et al. (1991). Variable deletion of exon 9 coding sequences in cystic fibrosis transmembrane conductance regulator gene mRNA transcripts in normal bronchial epithelium. *The EMBO journal*, *10*(6), 1355–63.

Chu, C., Trapnell, B. C., Curristin, S. M., Cutting, G. R., Crystal, R. G., Dalemans, W., Jallat, S., et al. (1992). Extensive posttranscriptional deletion of the coding sequences for part of nucleotide-binding fold 1 in respiratory epithelial mRNA transcripts of the cystic fibrosis transmembrane conductance regulator gene is not associated with the clinical manifestations, *J Clin Invest* 90(3): 785–790

Chu, W. K., & Hickson, I. D. (2009). RecQ helicases: multifunctional genome caretakers. *Nature reviews. Cancer*, *9*(9), 644–54.

Clancy, J. P., Rowe, S. M., Bebok, Z., Aitken, M. L., Gibson, R., Zeitlin, P., Berclaz, P., et al. (2007). No Detectable Improvements in Cystic Fibrosis Transmembrane Conductance Regulator by Nasal Aminoglycosides in Patients with Cystic Fibrosis with Stop Mutations. *American journal of respiratory cell and molecular biology*, *37*, 19–22.

Clarke, S. W., & Pavia, D. (1980). Lung mucus production and mucociliary clearance: methods of assessment. *British journal of clinical pharmacology*, *9*(6), 537–46

Cohen, C. J., Gaetz, J., Ohman, T., & Bergelson, J. M. (2001). Multiple regions within the coxsackievirus and adenovirus receptor cytoplasmic domain are required for basolateral sorting. *The Journal of biological chemistry*, *276*(27), 25392–8.

Cole, F., Keeney, S., & Jasin, M. (2010). Comprehensive, fine-scale dissection of homologous recombination outcomes at a hot spot in mouse meiosis. *Molecular Cell*, *39*(5), 700–710.

Colleaux, L., D'Auriol, L., Galibert, F., & Dujon, B. (1988). Recognition and cleavage site of the intron-encoded omega transposase. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(16), 6022–6.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, N.Y.)*, *339*(6121), 819–23.

Cozens, A.L., Yezzi, M J, Kunzelmann, K, Ohrui, T, Chin, L, Eng, K, Finkbeiner, W E, Widdicombe, J H and Gruenert DC. (1994). CFTR expression and chloride secretion in polarized immortal human bronchial epithelial cells. *American Journal of Respiratory Cell and Molecular Biology*, Vol. 10, No. 1 (1994), pp. 38-47.

Crystal, R. G., McElvaney, N. G., Rosenfeld1, M. A., Chu1, C.-S., Mastrangeli, A., Hay, J. G., Brody, S. L. H., et al. (1994). Administration of an adenovirus containing the human CFTR cDNA to the respiratory tract of individuals with cystic fibrosis. *Nature genetics*. **8**, 42 - 51

Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, Di Marzio P, Marmon S, Sutton RE, Hill CM, Davis CB, Peiper SC, Schall TJ, Littman DR, Landau NR. (1996). Identification of a major co-receptor for primary isolates of HIV-1. *Nature*.381(6584):661-6.

Denning, G.M. et al., 1992 Processing of mutant cystic fibrosis transmembrane conductance regulator is temperature-sensitive. *Nature* 358, 761-764

Desjarlais, J., & Berg, J. M. (1992). Toward rules relating zinc finger protein sequences and DNA. *Proc Natl Acad Sci USA*, *89*, 7345–7349.

DiMango, E., Zar, H. J., Bryan, R., & Prince, A. (1995). Diverse Pseudomonas aeruginosa gene products stimulate respiratory epithelial cells to produce interleukin-8. *Journal of Clinical Investigation*, *96*(5), 2204–2210.

Dodge, J. A., Lewis, P. A., Stanton, M., & Wilsher, J. (2007). Cystic fibrosis mortality and survival in the UK: 1947–2003, 522–526.

Dreier, B, Beerli, R. R., Segal, D. J., Flippin, J. D., & Barbas, C. F. (2001). Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *The Journal of biological chemistry*, *276*(31), 29466–78.

Dreier, Birgit, Fuller, R. P., Segal, D. J., Lund, C. V, Blancafort, P., Huber, A., Koksch, B., et al. (2005). Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *The Journal of biological chemistry*, *280*(42), 35588–97.

Drumm, M.L. et al., 1990. Correction of the cystic fibrosis defect in vitro by retrovirus-mediated gene transfer. *Cell*. 1990 Sep 21;62(6):1227-33.

Du, M., Liu, X., Welch, E. M., Hirawat, S., Peltz, S. W., & Bedwell, D. M. (2008). PTC124 is an orally bioavailable compound that promotes suppression of the human CFTR-G542X nonsense allele in a CF mouse model, *PNAS* 12;105(6):2064-9

Eastman SJ, Lukason MJ, Tousignant JD, Murray H, Lane MD, St George JA, Akita GY, Cherry M, Cheng SH, Scheule RK. (1997). A concentrated and stable aerosol formulation of cationic lipid:DNA complexes giving high-level gene expression in mouse lung. *Hum Gene Ther*. 1997 Apr 10;8(6):765-73.

Elliott, B., Richardson, C., Winderbaum, J., Jac, A., Jasin, M., & Nickoloff, J. A. C. A. (1998). Gene Conversion Tracts from Double-Strand Break Repair in Mammalian Cells Gene Conversion Tracts from Double-Strand Break Repair in Mammalian Cells. *Mol Cell Biol*, *18* 93-101.

Engelhardt JF, Yankaskas JR, Ernst SA, Yang Y, Marino CR, Boucher RC, Cohn JA, Wilson JM. (1992). Submucosal glands are the predominant site of CFTR expression in the human bronchus. *Nature* ;2(3):240-8

Flotte, T. R., Zeitlin, P. L., Reynolds, T. C., Heald, A. E., Pedersen, P., & Beck, S. (2003). Phase I trial of intranasal and endobronchial administration of a recombinant adeno-associated virus serotype 2 (rAAV2)-CFTR vector in adult

cystic fibrosis patients: a two-part clinical study. *Hum Gene Ther*, *14*(11), 1079−1088.

Fu, Y., Foden, J. a, Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology*, *31*(9), 822–6.

Gabriel, R., Lombardo, A., Arens, A., Miller, J. C., Genovese, P., Kaeppel, C., Nowrouzi, A., et al. (2011). An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nature biotechnology*, *29*(9), 816–23.

Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T. E., Mann, M., & Lieber, M. R. (1997). Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature 388*, 3–6.

Griesenbach, U., & Alton, E. W. (2009). Gene transfer to the lung: lessons learned from more than 2 decades of CF gene therapy. *Adv Drug Deliv Rev*, *61*(2), 128−139.

Guo, J., Gaj, T., & Barbas, C. F. (2010). Directed Evolution of an Enhanced and Highly Efficient FokI Cleavage Domain for Zinc Finger Nucleases. *Journal of molecular biology*. 400(1):96-107.

Hacein-Bey-Abina, S., Garrigue, A., Wang, G. P., Soulier, J., Lim, A., Morillon, E., Clappier, E., et al. (2008). Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1, *J. Clin. Invest*.118, 3132-3142.

Hanna, J., Wernig, M., Markoulaki, S., Sun, C.-W., Meissner, A., Cassady, J. P., Beard, C., et al. (2007). Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science*, *318*(5858), 1920–3.

Harrison, M. J., Murphy, D. M., & Plant, B. J. (2013). Ivacaftor in a G551D Homozygote with Cystic Fibrosis. *The New England journal of medicine*, *369*(13), 1280.

Harvey, B., Hackett, N. R., El-sawy, T., Rosengart, T. K., Hirschowitz, E. A., Lieberman, M. D., Lesser, M. L., et al. (1999). Variability of Human Systemic Humoral Immune Responses to Adenovirus Gene Transfer Vectors Administered to Different Organs, *J. Virol 73*(8), 6729–6742.

Harvey, B., Leopold, P. L., Hackett, N. R., Grasso, T. M., Williams, P. M., Tucker, A. L., Kaner, R. J., et al. (1999). Airway epithelial CFTR mRNA expression in cystic fibrosis patients after repetitive administration of a recombinant adenovirus, *J. Clin. Invest. 104*(9), 1245–1255.

Harvey, P. R., Tarran, R., Garoff, S., & Myerburg, M. M. (2011). Measurement of the airway surface liquid volume with simple light refraction microscopy. *American journal of respiratory cell and molecular biology*, *45*(3), 592–9.

Hasty, P., Rivera-Pérez, J., & Bradley, a. (1991). The length of homology required for gene targeting in embryonic stem cells. *Molecular and cellular biology*, *11*(11), 5586–91.

Heubi, J. E. (2007). Cystic Fibrosis Treatment of Exocrine Pancreatic Insufficiency in Cystic Fibrosis Patients. *Cystic Fibrosis*, 48–50.

Heyer, W.-D., Li, X., Rolfsmeier, M., & Zhang, X.-P. (2006). Rad54: the Swiss Army knife of homologous recombination? *Nucleic acids research*, *34*(15), 4115–25.

Hinnen, a, Hicks, J. B., & Fink, G. R. (1978). Transformation of yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *75*(4), 1929–33.

Holt, N., Wang, J., Kim, K., Friedman, G., Wang, X., Taupin, V., Crooks, G. M., et al. (2010). Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nature biotechnology*, *28*(8), 839–47.

Hubert, D., Bienvenu, T., Desmazes-Dufeu, N., Fajac, I., Lacronique, J., Matran, R., Kaplan, J. C., et al. (1996). Genotype-phenotype relationships in a cohort. *European Respiratory Journal*, *9*(11), 2207–2214.

Hutt, D. M., Herman, D., Rodrigues, A. P. C., Noel, S., Joseph, M., Matteson, J., Hoch, B., et al. (2010). Reduced histone deacetylase 7 activity restores function to misfolded CFTR in cystic fibrosis. *Nature Chem Biol.*, *6*(1), 25–33.

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., Peterson, R. T., et al. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology*, 31, 227-229

Hyde, S. C., Pringle, I. A., Abdullah, S., Lawton, A. E., Davies, L. A., Varathalingam, A., Nunez-alonso, G., et al. (2008). CpG- free plasmids confer reduced inflammation and sustained pulmonary gene expression. *Nature biotechnology*, *26*(5)

Ira, G., Pellicioli, A., Balijja, A., Wang, X., Fiorani, S., Carotenuto, W., Liberi, G., et al. (2004). DNA end resection , homologous recombination and DNA damage checkpoint activation require CDK1. *Nature*, 431, 1011–1017.

Jackson, S. P., & Bartek, J. (2009). The DNA-damage response in human biology and disease. *Nature*, *461*(7267), 1071–8.

Jacobs, G. H. (1992). Determination of the base recognition positions of zinc fingers from sequence analysis. *The EMBO journal*, *11*(12), 4507–17.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. a, & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816–21.

Johnson, L. G., Boyles, S. E., Wilson, J., & Boucher, R. C. (1995). Normalization of Raised Sodium Absorption and Raised Calcium-mediated Chloride Secretion by Adenovirus-mediated Expression of Cystic Fibrosis Transmembrane Conductance Regulator in Primary Human Cystic Fibrosis Airway Epithelial Cells. *J. Clin. Invest*, *95*, 1377–1382.

Kan, Y. W., & Dozy, a M. (1978). Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. *Proceedings of the National Academy of Sciences of the United States of America*, *75*(11), 5631–5.

Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC (1989). Identification of the cystic fibrosis gene: genetic analysis. *Science.* 1989 Sep 8;245(4922):1073-80.

Kim, E., Kim, S., Kim, D. H., Choi, B.-S., Choi, I.-Y., & Kim, J.-S. (2012). Precision genome engineering with programmable DNA-nicking enzymes. *Genome research*, *22*(7), 1327–33.

Kim, Y. G., & Chandrasegaran, S. (1994). Chimeric restriction endonuclease. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(3), 883–7.

Kim, Y. G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to FokI cleavage domain. *Proc. Natl Acad. Sci. USA*, *93*, 1156–1160.

Kim, Y., Kweon, J., Kim, A., Chon, J. K., Yoo, J. Y., Kim, H. J., Kim, S., et al. (2013). A library of TAL effector nucleases spanning the human genome. *Nature biotechnology*, *31*(3), 251–8.

Klein, H. L., & Symington, L. S. (2009, July 10). Breaking up just got easier to do. *Cell*. doi:10.1016/j.cell.2009.06.039

Knowles, M. R., Hohneker, K. W., Zhou, Z., Olsen, J. C., Noah, T. L., Hu, P. C., Leigh, M. W., et al. (1995). A controlled study of adenoviral-vector-mediated gene transfer in the nasal epithelium of patients with cystic fibrosis. *The New England journal of medicine*, *333*(13), 823–31.

Konstan, M. W., Davis, P. B., Wagener, J. S., Hilliard, K. A., Stern, R. C., Milgram, L. J. H., Kowalczyk, T. H., et al. (2004). Compacted DNA Nanoparticles Administered to the Nasal Mucosa of Cystic Fibrosis Subjects Are Safe and Demonstrate Partial to Complete Cystic Fibrosis Transmembrane Regulator Reconstitution, *1269*(December), 1255–1269.

Kreda, S. M., Mall, M., Mengos, A., Rochelle, L., Yankaskas, J., Riordan, J. R., & Boucher, R. C. (2005). Characterization of Wild-Type and ⌬ F508 Cystic Fibrosis Transmembrane Regulator in Human Respiratory Epithelia, *Mol. Biol. Cell 16*, 2154–2167.

Kucherlapati, R. S., Eves, E. M., Song, K. Y., Morse, B. S., & Smithies, O. (1984). Homologous recombination between plasmids in mammalian cells can be enhanced by treatment of input DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *81*(10), 3153–7.

Kwaks, T. H. J., & Otte, A. P. (2006). Employing epigenetics to augment the expression of therapeutic proteins in mammalian cells. *Trends in biotechnology*, *24*(3), 137–42.

Laity, J. H., Dyson, H. J., & Wright, P. E. (2000). DNA-induced alpha-helix capping in conserved linker sequences is a determinant of binding affinity in Cys(2)-His(2) zinc fingers. *Journal of molecular biology*, *295*(4), 719–27.

Lee, C. M., Flynn, R., Hollywood, J. A., Scallan, M. F., & Harrison, P. T. (2012). Correction of the ΔF508 Mutation in the Cystic Fibrosis Transmembrane Conductance Regulator Gene by Zinc-Finger Nuclease Homology-Directed Repair. *BioResearch open access*, *1*(3), 99–108.

Lee ER, Marshall J, Siegel CS, Jiang C, Yew NS, Nichols MR, Nietupski JB, Ziegler RJ, Lane MB, Wang KX, Wan NC, Scheule RK, Harris DJ, Smith AE, Cheng SH (1996). Detailed analysis of structures and formulations of cationic lipids for efficient gene transfer to the lung. *Hum Gene Ther.* 1996 Sep 10;7(14):1701-17.

Li, H., Haurigot, V., Doyon, Y., Li, T., Wong, S. Y., Bhagwat, A. S., Malani, N., et al. (2011). In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature*, *475*(7355), 217–21.

Lieber, M. R. (2008). The mechanism of human nonhomologous DNA end joining. *The Journal of biological chemistry*, *283*(1), 1–5.

Lin, F. L., Sperle, K., & Sternberg, N. (1984). Model for homologous recombination during transfer of DNA into mouse L cells: role for DNA ends in the recombination process. *Molecular and cellular biology*, *4*(6), 1020–34.

Liu, Q., Xia, Z., Zhong, X., & Case, C. C. (2002). Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *The Journal of biological chemistry*, *277*(6), 3850–6.

Liu, X., Luo, M., Guo, C., Yan, Z., Wang, Y., Lei-Butters, D. C. M., & Engelhardt, J. F. (2009). Analysis of adeno-associated virus progenitor cell transduction in mouse lung. *Molecular therapy : the journal of the American Society of Gene Therapy*, *17*(2), 285–93.

Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky L, Darnell J. (2004). Molecular Cell Biology. 5th Edition. W.H. Freeman.

Lombardo, A., Genovese, P., Beausejour, C. M., Colleoni, S., Lee, Y. L., & Kim, K. A. (2007). Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol*, *25*, 1298–1306.

Lowry, W. E., Richter, L., Yachechko, R., Pyle, a D., Tchieu, J., Sridharan, R., Clark, a T., et al. (2008). Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(8), 2883–8.

Lu, Z. H., Books, J. T., Kaufman, R. M., & Ley, T. J. (2003). Long targeting arms do not increase the efficiency of homologous recombination in the beta-globin locus of murine embryonic stem cells. *Blood*, *102*(4), 1531–3.

Lubamba, B., Dhooghe, B., Noel, S., & Leal, T. (2012). Cystic fibrosis: Insight into CFTR pathophysiology and pharmacotherapy. *Clinical Biochemistry*.

Ma, Y., Lu, H., Tippin, B., Goodman, M. F., Shimazaki, N., Koiwai, O., Hsieh, C.-L., et al. (2004). A biochemically defined system for mammalian nonhomologous DNA end joining. *Molecular cell*, *16*(5), 701–13.

Ma, Y., Pannicke, U., Schwarz, K., & Lieber, M. R. (2002). Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell*, *108*(6), 781–94.

Maeder, M. L., Thibodeau-Beganny, S., Osiak, A., Wright, D. a, Anthony, R. M., Eichtinger, M., Jiang, T., et al. (2008). Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Molecular cell*, *31*(2), 294–301.

Mah, T. C., & O' Toole, G. A. (2001). Mechanisms of biofilm resistance to antimicrobial agents, *9*(1), 34–39.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., et al. (2013). RNA-guided human genome engineering via Cas9. *Science (New York, N.Y.)*, *339*(6121), 823–6.

Mani, M., Kandavelou, K., Fei, J. D., Durai, S., & Chandrasegaran, S. (2005). Design, engineering and characterization of zinc finger nucleases. *Biochem. Biophys. Res. Commun*, *335*, 447–457.

Maresca, M., Lin, V. G., Guo, N., & Yang, Y. (2013). Obligate Ligation-Gated Recombination (ObLiGaRe): Custom-designed nuclease-mediated targeted integration through nonhomologous end joining, *Genome Res*. 23, 539–546.

Mashimo, T., Takizawa, A., Voigt, B., Yoshimi, K., Hiai, H., Kuramoto, T., & Serikawa, T. (2010). Generation of knockout rats with X-linked severe

combined immunodeficiency (X-SCID) using zinc-finger nucleases. *PloS one*, *5*(1), e8870.

Matsui, H, Davis, C. W., Tarran, R., & Boucher, R. C. (2000). Osmotic water permeabilities of cultured, well-differentiated normal and cystic fibrosis airway epithelia. *The Journal of clinical investigation*, *105*(10), 1419–27.

Matsui, H, Grubb, B. R., Tarran, R., Randell, S. H., Gatzy, J. T., Davis, C. W., Boucher, R. C., et al. (1998). Evidence for Periciliary Liquid Layer Depletion , Not Abnormal Ion Composition , in the Pathogenesis of Cystic Fibrosis Airways Disease. *Cell*, *95*, 1005–1015.

McConnell Smith, A., Takeuchi, R., Pellenz, S., Davis, L., Maizels, N., Monnat, R. J., & Stoddard, B. L. (2009). Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(13), 5099–104.

McElvaney NG, Hubbard RC, Birrer P, Chernick MS, Caplan DB, Frank MM, Crystal RG.(1991). Aerosol alpha 1-antitrypsin treatment for cystic fibrosis. *Lancet*. 1991 16;337(8738):392-4.

McElvaney, N. G., Nakamura, H, Birrer, P., Hébert, CA, WL, W., M, A., JB, B., et al. (1992). Modulation of Airway Inflammation in Cystic Fibrosis v. *The Journal of Clinical Investigation,*, *90*(October), 1296–1301.

Meier, I. D., Bernreuther, C., Tilling, T., Neidhardt, J., Wong, Y. W., Schulze, C., Streichert, T., et al. (2010). Short DNA sequences inserted for gene targeting can accidentally interfere with off-target gene expression. *FASEB journal*, *24*(6), 1714–24.

Mendoza, J. L., Schmidt, A., Li, Q., Nuvaga, E., Barrett, T., Bridges, R. J., Feranchak, A. P., et al. (2012). Requirements for efficient correction of ΔF508 CFTR revealed by analyses of evolved sequences. *Cell*, *148*(1-2), 164–74.

Miller, J, McLachlan, a D., & Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *The EMBO journal*, *4*(6), 1609–14.

Miller, JC, Holmes, M., & Wang, J. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. *Nature Biotechnology*, *25*, 778–785.

Mimitou, E. P., & Symington, L. S. (2008). Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature*, *455*(7214), 770–4.

Mimitou, E. P., & Symington, L. S. (2009). Nucleases and helicases take center stage in homologous recombination. *Trends in biochemical sciences*, *34*(5), 264–72.

Moehle, E. a, Moehle, E. a, Rock, J. M., Rock, J. M., Lee, Y.-L., Lee, Y. L., Jouvenot, Y., et al. (2007). Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(9), 3055–60.

Morozov, V., & Wawrousek, E. F. (2008). Single-strand DNA-mediated targeted mutagenesis of genomic DNA in early mouse embryos is stimulated by Rad51/54 and by Ku70/86 inhibition. *Gene therapy*, *15*(6), 468–72

Moss, R. B. (2004). Repeated Adeno-Associated Virus Serotype 2 Aerosol-Mediated Cystic Fibrosis Transmembrane Regulator Gene Transfer to the Lungs of Patients With Cystic Fibrosis: A Multicenter, Double-Blind, Placebo-Controlled Trial. *Chest*, *125*(2), 509–521.

Moss, Richard B, Milla, C., Colombo, J., Accurso, F., Zeitlin, P. L., Clancy, J. P., Spencer, L. T., et al. (2007). Repeated Aerosolized AAV-CFTR for Treatment of Cystic Fibrosis : A Randomized Placebo-Controlled Phase 2B Trial, *Human Gene Therapy* 18 726–732.

Moynahan, M. E., & Jasin, M. (2011). Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat Rev Mol Cell Biol*, *11*(3), 196–207.

Mussolino, C., Morbitzer, R., Lütge, F., Dannemann, N., Lahaye, T., & Cathomen, T. (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic acids research*, *39*(21), 9283–93.

Nassif, N., Penney, J., Pal, S., Engels, W. R., & Gloor, G. B. (1994). Efficient Copying of Nonhomologous Sequences from Ectopic Sites via P-Element-Induced Gap Repairt, *Mol. Cell Biol. 14*(3), 1613–1625.

Nichols, J. E., Niles, J. a, & Cortiella, J. (2012). Production and utilization of acellular lung scaffolds in tissue engineering. *Journal of cellular biochemistry*, *113*(7), 2185–92.

O'Riordan, C. R., Lachapelle, a L., Marshall, J., Higgins, E. a, & Cheng, S. H. (2000). Characterization of the oligosaccharide structures associated with the cystic fibrosis transmembrane conductance regulator. *Glycobiology*, *10*(11), 1225–33.

Orr-Weaver, T. L., Szostak, J. W., & Rothstein, R. J. (1981). Yeast transformation: a model system for the study of recombination. *Proceedings of the National Academy of Sciences of the United States of America*, *78*(10), 6354–8.

Pattanayak, V., Ramirez, C. L., Joung, J. K., & Liu, D. R. (2011). Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nature methods*, *8*(9), 765–70.

Pavletich NP, Pabo CO. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science*. 10;252(5007):809-17

Pellegrino, G. R., & Berg, J. M. (1991). Identification and characterization of "zinc-finger" domains by the polymerase chain reaction. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(2), 671–5.

Pennisi, E. (2013). The CRISPR craze. *Science (New York, N.Y.)*, *341*(6148), 833–6.

Perez, E. E., Wang, J., Miller, J. C., Jouvenot, Y., Kim, K. a, Liu, O., Wang, N., et al. (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nature biotechnology*, *26*(7), 808–16.

Petersen, T. H., Calle, E. a, Zhao, L., Lee, E. J., Gui, L., Raredon, M. B., Gavrilov, K., et al. (2010). Tissue-engineered lungs for in vivo implantation. *Science (New York, N.Y.)*, *329*(5991), 538–41.

Petukhova, G, Stratton & Sung, P. (1998). Catalysis of homologous DNA pairing by yeast Rad51 and Rad54 proteins.*Nature*, *393*, 91–94.

Pezzulo, A. a, Tang, X. X., Hoegger, M. J., Alaiwa, M. H. A., Ramachandran, S., Moninger, T. O., Karp, P. H., et al. (2012). Reduced airway surface pH impairs bacterial killing in the porcine cystic fibrosis lung. *Nature*, *487*(7405), 109–13.

Pickles, R. J., Carty, D. M. C., Matsui, H., Hart, P. J., Randell, S. H., & Boucher, R. C. (1998). Limited Entry of Adenovirus Vectors into Well-Differentiated Airway Epithelium Is Responsible for Inefficient Gene Transfer. *Journal of virology*, *72*(7), 6014–6023.

Poolman, E. M., & Galvani, A. P. (2007). Evaluating candidate agents of selective pressure for cystic fibrosis. *Journal of the Royal Society, Interface / the Royal Society*, *4*(12), 91–8.

Porteous, D. J., Dorin, J. R., Mclachlan, G., Davidson, H., Stevenson, B. J., Naujoks, K., Ho, L., et al. (1997). Evidence for safety and efficacy of DOTAP cationic liposome mediated CFTR gene transfer to the nasal epithelium of patients with cystic fibrosis, *Gene Ther*. 4, 210–218.

Purohit P & Stern S. (1994). Interactions of a small RNA with antibiotic and RNA ligands of the 30S subunit. *Nature.* 25;370(6491):659-62.

Quinton, P.M. & Philpott C.W. 1973 A role for anionic sites in epithelial architecture. Effects of cationic polymers on cell membrane structure. *J Cell Biol.* 1973 Mar;56(3):787-96.

Rahman, S. H., Bobis-Wozowicz, S., Chatterjee, D., Gellhaus, K., Pars, K., Heilbronn, R., Jacobs, R., et al. (2013). The nontoxic cell cycle modulator

indirubin augments transduction of adeno-associated viral vectors and zinc-finger nuclease-mediated gene targeting. *Human gene therapy*, *24*(1), 67–77.

Ramalho, A. S., Beck, S., Meyer, M., Penque, D., Cutting, G. R., & Amaral, M. D. (2002). Five percent of normal cystic fibrosis transmembrane conductance regulator mRNA ameliorates the severity of pulmonary disease in cystic fibrosis. *American journal of respiratory cell and molecular biology*, *27*(5), 619–27.

Ramirez, C., Foley, J., Wright, D., Müller-Lerch, F., SH, R., Cornu, T., Winfrey, R., et al. (2008). Unexpected failure rates for modular assembly of engineered zinc fingers, *Nat. Methods 5*(5), 374–375.

Ran, F. A., Hsu, P. D., Lin, C.-Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. a, et al. (2013). Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell*, *154*(6), 1380–9.

Rawlins, E. L., & Hogan, B. L. M. (2008). Ciliated epithelial cell lifespan in the mouse trachea and lung. *American journal of physiology. Lung cellular and molecular physiology*, *295*(1), L231–4.

Rebar, E. J., & Miller, J. C. (2004). Design and applications of engineered zinc finger proteins, *Gene.*17;366(1):27-38

Reyon, D., Tsai, S. Q., Khayter, C., Foden, J. a, Sander, J. D., & Joung, J. K. (2012). FLASH assembly of TALENs for high-throughput genome editing. *Nature biotechnology*, *30*(5), 460–5. doi:10.1038/nbt.2170

Rich DP, Anderson MP, Gregory RJ, Cheng SH, Paul S, Jefferson DM, McCann JD, Klinger KW, Smith AE, Welsh MJ. (1990). Expression of cystic fibrosis transmembrane conductance regulator corrects defective chloride channel regulation in cystic fibrosis airway epithelial cells. *Nature.* 1990 27;347(6291):358-63.

Riordan, J. R., Rommens, J. M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., et al. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science (New York, N.Y.)*, *245*(4922), 1066–73.

Rogers, C. S., Hao, Y., Rokhlina, T., Samuel, M., Stoltz, D. A., Li, Y., Petroff, E., et al. (2008). Production of CFTR-null and CFTR-ΔF508 heterozygous pigs by adeno-associated virus – mediated gene targeting and somatic cell nuclear transfer, *J. Clin. Invest. 118*(4) 1571-1577

Rommens, J. M., Iannuzzi, M. C., Kerem, B., Drumm, M. L., Melmer, G., Dean, M., Rozmahel, R., et al. (1989). Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science*, *245*(4922), 1059–1065.

Rouet, P., Smih, F., & Jasin, M. (1994). Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(13), 6064–8.

Rouse, J., & Jackson, S. P. (2002). Interfaces between the detection, signaling, and repair of DNA damage. *Science (New York, N.Y.)*, *297*(5581), 547–51.

Rubenstein, R. C., Lockwood, S. R., Lide, E., Bauer, R., Suaud, L., & Grumbach, Y. (2011). Regulation of endogenous ENaC functional expression by CFTR and ΔF508-CFTR in airway epithelial cells. *American journal of physiology. Lung cellular and molecular physiology*, *300*(1), L88–L101.

Salima Hacein-Bey-Abina, Pharm.D., Ph.D., Julia Hauer, M.D., Annick Lim, M.Sci., C., Picard, M.D., Ph.D. et al., (2011). Efficacy of Gene Therapy for X-Linked Severe Combined Immunodeficiency, *N. Eng. J Med 363*(4), 355–364.

San Filippo, J., Sung, P., & Klein, H. (2008). Mechanism of eukaryotic homologous recombination. *Annual review of biochemistry*, *77*, 229–57.

Sander, J. D., Dahlborg, E. J., Goodwin, M. J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S. J., et al. (2010). Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nature Methods*,. 8, 67-69

Sander, J. D., Ramirez, C. L., Linder, S. J., Pattanayak, V., Shoresh, N., Ku, M., Foden, J. a, et al. (2013). In silico abstraction of zinc finger nuclease cleavage profiles reveals an expanded landscape of off-target sites. *Nucleic acids research*, (9), 1–7.

Sartori, A. A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Baer, R., Lukas, J., et al. (2007). Human CtIP promotes DNA end resection. *Nature*, *450*(7169), 509–514.

Sebastiano, V., Maeder, M. L., Angstman, J. F., Haddad, B., Khayter, C., Yeo, D. T., Goodwin, M. J., 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 (Dayton, Ohio)*, *29*(11), 1717–26.

Segal, D. J., Dreier, B., Beerli, R. R., & AND Barbas II, C. F. (1999). Toward controlling gene expression at will: Selection and design of zinc finger domains recognizing each of the 5-GNN-3 DNA target sequences. *Biochemistry*, *96*(March), 2758–2763.

Sera, T, & Uranga, C. (2002). Rational design of artificial zinc-finger proteins using a nondegenerate recognition code table. *Biochemistry*, *41*, 7074–7081.

Sera, Takashi. (2009). Zinc-finger-based artificial transcription factors and their applications. *Advanced drug delivery reviews*, *61*(7-8), 513–26.

Shak, S., Capon, D. J., Hellmiss, R., Marsters, S. a, & Baker, C. L. (1990). Recombinant human DNase I reduces the viscosity of cystic fibrosis sputum. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(23), 9188–92.

Shimizu, Y., Bhakta, M. S., & Segal, D. J. (2009). Restricted spacer tolerance of a zinc finger nuclease with a six amino acid linker. *Bioorganic & medicinal chemistry letters*, *19*(14), 3970–2.

Sims, D. E., & Horne, M. M. (1997). Heterogeneity of the composition and thickness of tracheal mucus in rats Heterogeneity of the composition and thickness of tracheal mucus in rats. *Am. J. Physiol*. 273, 1036-1041

Smith, J., Bibikova, M., Whitby, F. G., Reddy, A. R., Chandrasegaran, S., & Carroll, D. (2000). Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res*, *28*, 3361–3369.

Smith, J. J., Travis, S. M., Greenberg, E. P., & Welsh, M. J. (1996). Cystic fibrosis airway epithelia fail to kill bacteria because of abnormal airway surface fluid, *Cell 85*, 229–236.

Sonoda, E., Hochegger, H., Saberi, A., Taniguchi, Y., & Takeda, S. (2006). Differential usage of non-homologous end-joining and homologous recombination in double strand break repair. *DNA repair*, *5*(9-10), 1021–9.

Stark, J. M., Pierce, A. J., Oh, J., Pastink, A., & Jasin, M. (2004). Genetic Steps of Mammalian Homologous Repair with Distinct Mutagenic Consequences, Mol. Cell Biol. *24*(21), 9305–9316.

Stern, M., Caplen, N. J., Browning, J. E., Griesenbach, U., Sorgi, F., Huang, L., Gruenert, D. C., et al. (1998). The effect of mucolytic agents on gene transfer across a CF sputum barrier in vitro. *Gene therapy*, *5*(1), 91–8.

Sun, N., Liang, J., Abil, Z., & Zhao, H. (2012). Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Molecular bioSystems*, *8*(4), 1255–63.

Sun, X., Sui, H., Fisher, J. T., Yan, Z., Liu, X., Cho, H., Joo, N. S., et al. (2010). Technical advance Disease phenotype of a ferret CFTR-knockout model of cystic fibrosis, *J. Clin. Invest. 120*(9), 3149–3160.

Szczepek, M., Brondani, V., Buchel, J., Serrano, L., Segal, D. J., & Cathomen, T. (2007). Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases ¨. *Nature Biotechnology*, *25*(7), 786–793.

Takata, M., Sasaki, M. S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., et al. (1998). Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have

overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *The EMBO journal*, *17*(18), 5497–508.

Tan, T. L., Kanaar, R., & Wyman, C. (2003). Rad54, a Jack of all trades in homologous recombination. *DNA Repair*, *2*(7), 787–794.

Tarran, R., Grubb, B. R., Gatzy, J. T., Davis, C. W., & Boucher, R. C. (2001). The relative roles of passive surface forces and active ion transport in the modulation of airway surface liquid volume and composition. *The Journal of general physiology*, *118*(2), 223–36.

Thomas KR, Capecchi MR. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell.* 1987 Nov 6;51(3):503-12.

Urnov, F. D., Miller, J. C., Lee, Y. L., Beausejour, C. M., Rock, J. M., & Augustus, S. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, *435*, 646–651.

Valko, M., Rhodes, C. J., Moncol, J., Izakovic, M., & Mazur, M. (2006). Free radicals, metals and antioxidants in oxidative stress-induced cancer. *Chemico-biological interactions*, *160*(1), 1–40.

Van Goor, F., Hadida, S., Grootenhuis, P. D. J., Burton, B., Cao, D., Neuberger, T., Turnbull, A., et al. (2009). Rescue of CF airway epithelial cell function in vitro by a CFTR potentiator, VX-770. *PNAS 106*(44), 18825–30.

Van Goor, F., Hadida, S., Grootenhuis, P. D. J., Burton, B., Stack, J. H., Straley, K. S., Decker, C. J., et al. (2011). Correction of the F508del-CFTR protein processing defect in vitro by the investigational drug VX-809, PNAS, 108, 18843-18848.

Vasileva, A., & Jessberger, R. (2005). Precise hit: adeno-associated virus in gene targeting. *Nature Microbiology*, *3*(11), 837–47.

Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., & Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, *153*(4), 910–8.

Ward, C. L., & Kopito, R. R. (1994). Intracellular Turnover of Cystic Fibrosis Transmembrane Conductance Regulator. Inefficient processing and rapid degradation of wild-type and mutant proteins. J. Biol. Chem. *269*(41), 25710–25718.

Welch, E. M., Barton, E. R., Zhuo, J., Tomizawa, Y., Friesen, W. J., Trifillis, P., Paushkin, S., et al. (2007). PTC124 targets genetic disorders caused by nonsense mutations, *Nature 447*, 87-91

Wiedenheft, B., Sternberg, S. H., & Doudna, J. a. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature*, *482*(7385), 331–8.

Wilschanski, M., Famini, C., Blau, H., Rivlin, J., & Augarten, A. (2000). A Pilot Study of the Effect of Gentamicin on Nasal Potential Difference Measurements in Cystic Fibrosis. *Am J respiratory Crit Care Med*, (10), 860–865.

White R, Woodward S, Leppert M, O'Connell P, Hoff M, Herbst J, Lalouel JM, Dean M, Vande Woude G. (1985). A closely linked genetic marker for cystic fibrosis. *Nature.* 1985 Nov 28-Dec 4;318(6044):382-4.

Worlitzsch, D., Tarran, R., Ulrich, M., Schwab, U., Cekici, A., Meyer, K. C., Birrer, P., et al. (2002). Effects of reduced mucus oxygen concentration in airway Pseudomonas infections of cystic fibrosis patients. *J. Clin. Invest*, *109*(3), 317–325.

Yamanaka, S. (2007). Strategies and new developments in the generation of patient-specific pluripotent stem cells. *Cell stem cell*. 1, 39-49

Yang, H., Wang, H., Shivalila, C. S., Cheng, A. W., Shi, L., & Jaenisch, R. (2013). One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell*, *154*(6), 1370–1379.

Yuan, S. F., Lee, S., Chen, G. Song, M., Tomlinson, GE. & Lee, EY. (1999). BRCA2 is required for ionizing radiation-induced assembly of Rad51 complex in vivo *Cancer Res.* 3547–3551.

Yusa, K., Rashid, S. T., Strick-Marchand, H., Varela, I., Liu, P.-Q., Paschon, D. E., Miranda, E., et al. (2011). Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. *Nature*, *478*(7369), 391–4.

Zabner J, Couture LA, Gregory RJ, Graham SM, Smith AE, Welsh MJ. (1993) Adenovirus-mediated gene transfer transiently corrects the chloride transport defect in nasal epithelia of patients with cystic fibrosis. *Cell.* 22;75(2):207-16.

Zabner, J, Ramsey, B. W., Meeker, D. P., Aitken, M. L., Balfour, R. P., Gibson, R. L., Launspach, J., et al. (1996). Repeat administration of an adenovirus vector encoding cystic fibrosis transmembrane conductance regulator to the nasal epithelium of patients with cystic fibrosis. *The Journal of clinical investigation*, *97*(6), 1504–11.

Zabner, Joseph, Cheng, S. H., Meeker, D., Launspach, J., Balfour, R., Perricone, M. A., Morris, J. E., et al. (1997). Comparison of DNA – Lipid Complexes and DNA Alone for Gene Transfer to Cystic Fibrosis Airway Epithelia In Vivo. *J. Clin. Invest*, (16), 1529–1537.

Zhang, L., Button, B., Gabriel, S. E., Burkett, S., Yan, Y., Skiadopoulos, M. H., Dang, Y. L., et al. (2009). CFTR delivery to 25% of surface epithelial cells

restores normal rates of mucus transport to human cystic fibrosis airway epithelium. *PLoS biology*, *7*(7),

Zhu, Z., Chung, W.-H., Shim, E. Y., Lee, S. E., & Ira, G. (2008). Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell*, *134*(6), 981–94.

8 Appendix:

8.1 Functional Assay: Immunofluorescence for detection of CFTR protein trafficking

The CFTE cell line is homozygous for the ΔF508 mutation which causes CFTR to become trapped in the ER and degraded before reaching the apical membrane. We know from chapters 3-5 that our ZFNs and CRISPR gRNAs can mediate HDR with the appropriate donor plasmids but it remains to be formally established if this genetic repair leads to functional CFTR protein trafficking from the ER to the apical membrane that will allow normal ion flux and restore ASL.

To monitor CFTR trafficking we sought to stain the protein with appropriate antibodies and visualise its expression in cells using confocal microscopy. As a large population of cells would need to be screened with only a few being corrected given that the HDR rate is at 1-2%, it would be quite difficult and labour intensive to identify corrected cells, which is in part why we developed the mini-gene donor. Originally it was thought that using our mini-gene donor which contains the GFP tag would allow easier detection of CFTR corrected cells as GFP will be made following CFTR production and so a green cell should contain corrected CFTR at the apical membrane. However, the deep sequencing results revealed that the GFP signal detected in this instance was an artefact, and so could not be used.

Nonetheless, attempts were made to establish the assay using the pCEP-CFTR plasmid, (which persists in cells and contains the CFTR cDNA under the control of a CMV promoter), transfected in CFTE cells homozygous for the ΔF508 mutation. pCEP-CFTR was co-transfected with pEGFP, the idea being that if a cell took up the EGFP plasmid then it most likely took up the pCEP-CFTR plasmid also and so green cells should show some CFTR at the apical membrane.

CFTR antibodies have a reputation for being difficult to use and distinguishing real versus background signal can be problematic. The naturally low abundance of WT CFTR can also make it difficult to detect. This section describes our attempts to establish a CFTR trafficking assay using the CFTR 596 antibody, obtained from

Jack Riordan's lab in North Carolina, to be used for future functional studies and some of the problems encountered.

8.2 Transfection of CFTE cells with pCEP-CFTR and pEGFP

CFTE cells were seeded on 13mm coverslips and transfected with 2 μg of pCEP-CFTR and 2 μg of pEGFP. Following 72 hrs incubation, cells were stained for CFTR using primary CFTR 596 antibody and the goat anti-mouse IgG Cy3 secondary antibody.

CFTE cells which take up pCEP-CFTR plasmid should now express full length CFTR which should be trafficked correctly to the apical membrane of the cell. Cells transfected with both pCEP-CFTR and pEGFP should contain some cells which express both plasmids and so a green cell will should also contain CFTR at the apical membrane. Cells transfected with both pEGFP and pCEP-CFTR were used to represent what is expected to be observed if the mini-gene donor was correctly recombined into the genome to produce functional CFTR and also GFP due to the T2A linker. It is highly likely that the GFP positive cells also contain pCEP-CFTR.

In order to visually ascertain which cells contain WT CFTR, the distribution of the protein was used as a marker. The idea was that mutant CFTR will be trapped in the ER and so should be located close to the nucleus whereas corrected/pCEP-CFTR derived CFTR should be more spread out in the cell and mainly located at the apical membrane.

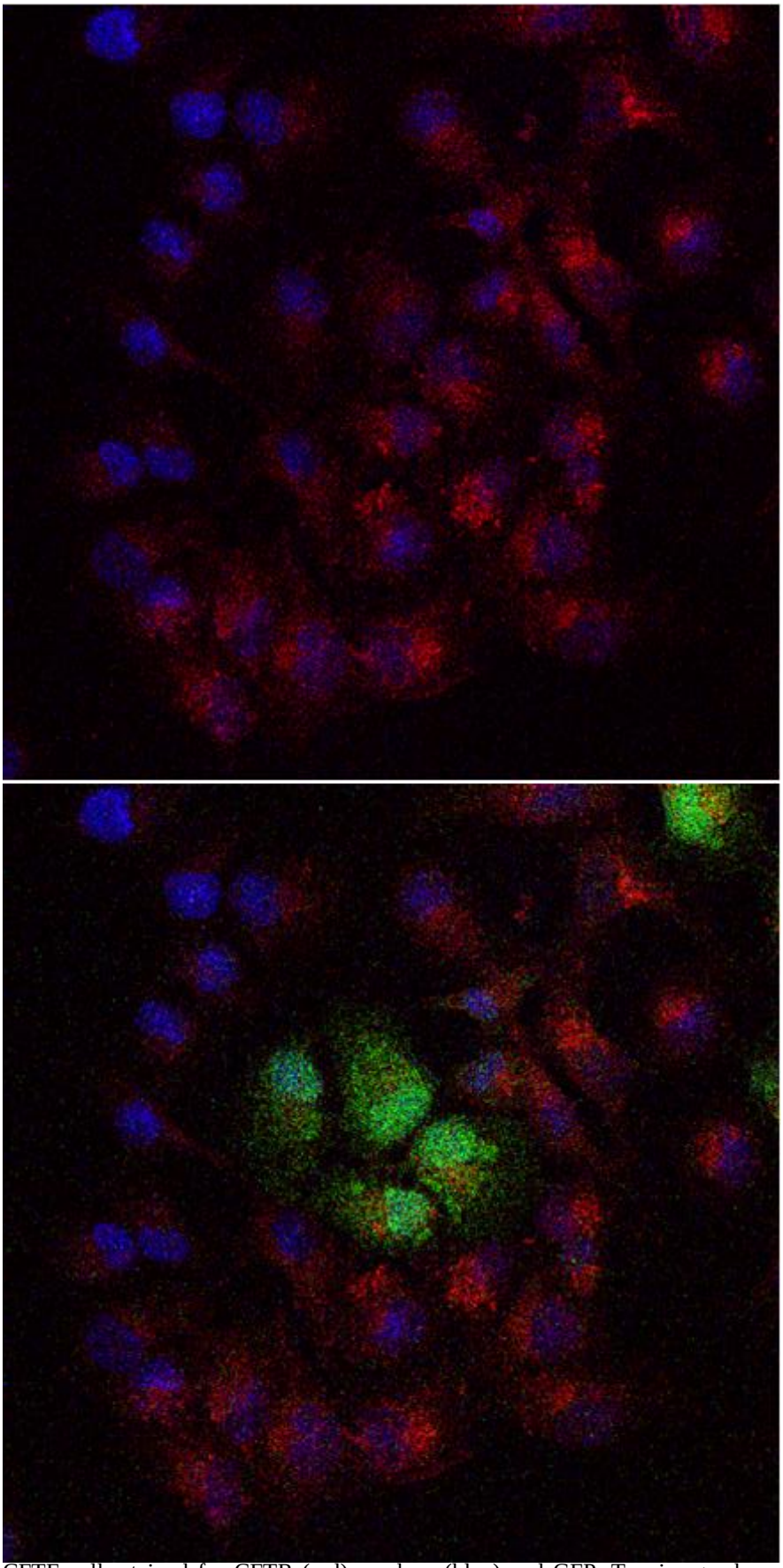


Figure 8.1: CFTE cells stained for CFTR (red), nucleus (blue) and GFP. Top image shows cells stained for CFTR and DAPI only. Bottom image shows same cells with GFP fluorescence added.

The images in figure 8.1 represent CFTE cells transfected as outlined above. The secondary Cy3 stains CFTR red and can be seen in all transfected cells as expected. DAPI stains the nucleus blue, while GFP fluoresces under the correct UV light. In the top image, where only CFTR and nuclei are stained, there is evidence of CFTR trapping in some cells whereby there is an abundance of CFTR stained in close proximity to the nucleus. In the bottom image, four GFP positive cells can be seen in the centre of the image. When these four cells are compared to the same four cells in the top image there seems to be a different distribution of CFTR. Rather than being packed tightly close to the nucleus, it now appears to be more spread out throughout the cell. This implies that these cells contain both pEGFP and pCEP-CFTR and so WT CFTR is being trafficked to the apical membrane. However, there are some cells where the signal is ambiguous with a similar appearance to the GFP positive cells but contain no GFP.

It was hoped that the visual difference in distribution of CFTR would be used as a marker when searching for corrected cells in a population and the GFP tag in the newly designed mini-gene construct will allow rapid and accurate identification of repair. This work is not conclusive and further analysis needs to be done before it can be used as a functional assay, however, it does confirm that the CFTR antibody works and can be used to stain for CFTR in our CFTE cell line.

To improve the assay further, polarised cells could be used and a tight junction staining for ZO-1 applied in order to locate the apical membrane and determine how much CFTR is present. Z-stacking of multiple microscopic images would give a 3D view of the cell and allow better visualisation of the CFTR at the membrane. As the number of corrected cells will be low, the GFP tag in the mini-gene donor will greatly aid detecting and monitoring the trafficking of CFTR.

If time allowed I would like to explore an alternative assay for CFTR functional studies, that of measuring ASL (see section 1.1.4). By measuring the height of ASL, one could determine whether CFTR is functional at the apical membrane. WT CFTR allows correct movement of Cl^- ions and by osmosis water through the membrane both which determine the ASL height. Following transfection with our nucleases and donor plasmids the ASL could be monitored for height changes by

immune-staining. Alternatively, by adding a dye like phenol red into the apical compartment and monitoring changes in the O.D. reading of the fluid could give an indication of whether more fluid is in the ASL following transfection due to nuclease induced HDR.

Acknowledgments:

First and foremost, I would like to thank my two supervisors Dr. Patrick Harrison and Dr. Martina Scallan for all their help and expertise during this project. Both were instrumental in securing funding for this work and were always readily available to offer advice and support. A special thanks to Paddy, who introduced me to the world of gene editing during my undergrad and for giving me the opportunity to pursue a career in this field.

I would like to thank the respective physiology and microbiology heads of department, Prof. Ken O' Halloran and Prof. Gerald Fitzgerald, for financial support of the entire project.

Sincere thanks to the College of Medicine and Health, (in particular, Michael Berndt, Geraldine McCarthy and John Higgins), Physoc and ECFS for affording me the opportunity to travel and present my work at various conferences which have proved invaluable. Thanks also to Cystinosis Ireland for providing access to the GX lab chip and funding my first scientific meeting. The support staff in the physiology department has been exceptional and essential for the smooth running of the lab, therefore I would like to thank Miriam O' Sullivan, Stephen Dineen, Tony Mulvihill, Kieran McDonnell and Drs Greg Jasionek and Marcella Burke. I would also like to thank Jackie, Nora, Verna, Nicola and Maude. A special thanks to my post graduate committee, Dr. Ahmad Ahmeda, Prof. Ken O' Halloran and Dr. Gordon Reid for some great advice over the years. Thanks also to the Microbiology department, in particular Conal Martin, Maurice O' Donoghue and Paddy O' Reilly.

For their help in generating the genotyping data, I would like to extend my gratitude to Dr. Collette Hand and David Shilling at C.U.H. I would also like to thank Drs. Steven Land and Anil Mehta (University of Dundee) for giving me the opportunity to test our techniques on primary CF cells and agreeing to collaborate with our group in the future. Thanks to Christine O' Carroll for teaching me how to use the flow cytometer.

To my fellow post-grads, Ciaran, Katrin, Kate, Will, Limian, Ciara, Jenny, Robbie, Philip, Elaine, Andrew, Aidan, Maria, Chrissy and Dee thanks so much for making

my PhD experience enjoyable. A massive thanks to my past colleagues Drs. Ciaran Lee and Katrin Kaschig who were instrumental in driving this work forward and who made the lab a fun place to be.

Finally, I would like to thank my family and friends for their love and support, especially my parents for all they've done for me throughout the years, and to Pa for always being there with words of encouragement and an open ear.