

DISS. ETH NO. 18246

INDUCIBLE GENE REPLACEMENT OF DNA POLYMERASE DELTA

A dissertation submitted to

ETH ZURICH

for the degree of

Doctor of Sciences

presented by

PATRICK ALBERT KEHL

Dipl. Natw. ETH

born 24.12.1975

citizen of Balgach (SG)

accepted on the recommendation of

Prof. Dr. Josef Jiricny

Prof. Dr. Fritz Thoma

Zurich, 2009

TO MY FAMILY

Summary

Deoxyribonucleic acid (DNA) polymerases ϵ and δ are responsible for leading and lagging strand synthesis during eukaryotic DNA replication. Because these enzymes have intrinsic error rates of 1.1×10^{-5} and 3.7×10^{-5} per base pair, error-free replication of genomes of these organisms can only be achieved with the help of auxiliary mechanisms that remove misincorporated nucleotides from newly-synthesized DNA prior to cell division. In all organisms studied to date, the replicative polymerases possess intrinsic proofreading exonucleases, which remove mispaired nucleotides from the 3' termini of newly-synthesized DNA and thus improve replication fidelity by ~ 2 orders of magnitude. Postreplicative mismatch repair (MMR) then mediates the removal of mispairs that escaped proofreading. This process improves the fidelity of DNA replication by up to 3 orders of magnitude, such that even genomes of higher eukaryotes can be replicated without errors. The consequences of missing MMR have been studied in great detail during the past decade, primarily because of a link to cancer. However, to date it is unclear whether malignancy arises as a result of the mutator phenotype on MMR-deficient cells, or whether it is linked to an as yet unknown function of the MMR proteins. In order to address this question, we set out to generate a cell system in which the mutator phenotype would not be induced by defective MMR. We wanted to influence the fidelity of DNA replication by introducing mutations into the polymerase and exonuclease domains of the replicating polymerases, but this approach is complicated by the fact that these enzymes are essential. The aim of my study was to develop a system that would overcome these difficulties and allow us to regulate replication fidelity by altering the base selectivity and/or the proofreading exonuclease activity of polymerase δ .

I focused on mutations that were already characterised in yeast and mice. However, in contrast to these studies, I wanted to devise an isogenic system, in which the variant enzymes would replace the endogenous activity by the combination of inducible expression and ribonucleic acid (RNA) interference mediated knock-down technologies. In this way, the wild type and mutator phenotypes could be studied in one and the same human cell line. Moreover, the flexibility of the inducible system should ensure that the exogenous variant is expressed in similar amounts to the endogenous wild type protein. I refer to the system as “gene replacement”.

Although, it does not involve manipulation of the endogenous gene in the genome, it replaces the genes encoded function with one encoded by a regulatable, stably-integrated expression vector.

Making use of the T-REx system, I was able to inducibly express the wild type p125, the large subunit of polymerase δ , and three variants: proofreading-deficient, error-prone, and both proofreading-deficient and error-prone. I confirmed that all four p125 variants had polymerase activity *in vitro* and that the mutant variants incorporated non-complementary nucleotides with higher efficiency. However, the cell lines expressing the variants in the presence of the wild type endogenous enzyme did not display reduced growth rates or morphological changes. This suggested that we have to eliminate the wild type enzyme, which I achieved with the help of inducible short hairpin RNA (shRNA) expression that targets exclusively the endogenous p125 messenger RNA. The current approach, which combines inducible expression of the variant with the concurrent expression of the shRNA, successfully replaces endogenous polymerase δ with a variant of choice. Cells in which the endogenous protein was replaced with the mutator variants displayed reduced viability. Whether the observed reduction in viability is indeed a consequence of the mutator phenotype requires further studies.

In order to simplify this strategy for the future, I created a construct, which allows the transfer of the above wild type or mutator variants of polymerase δ into any human cell line in a single step. Because this approach is applicable to any gene of interest, the work described in this thesis represents a proof of principle that opens a wide field of research possibilities and applications.

Zusammenfassung

Die Desoxyribonukleinsäure- (deoxyribonucleic acid, DNA) Polymerasen ϵ und δ sind verantwortlich für die Synthese des Leit- und Folgestranges während der DNA-Replikation in Eukaryoten. Da diese beiden Enzyme eine inhärente Fehlerrate von 1.1×10^{-5} beziehungsweise 3.7×10^{-5} pro Basenpaar haben, ist eine fehlerfreie Replikation von Genomen dieser Organismen nur möglich mit Hilfe von zusätzlichen Mechanismen, welche fehlerhaft eingebaute Nukleotide noch vor der Zellteilung aus der neu synthetisierten DNA entfernen. Die replikativen Polymerasen aller Organismen, welche bis anhin untersucht wurden, verfügen über eine Korrekturlesefunktion, welche fehlgepaarte Nukleotide vom 3' Ende der neu synthetisierten DNA entfernt und so die Genauigkeit der Replikation um ungefähr 2 Zehnerpotenzen erhöht. Die nachgeschaltete Fehlpaarungsreparatur (mismatch repair, MMR) entfernt Basenfehlpaarungen, welche der Korrekturlese entgangen sind. Dieser Prozess erhöht die Genauigkeit der DNA Replikation um bis zu 3 Zehnerpotenzen, so dass sogar Genome von höheren Eukaryoten ohne Fehler repliziert werden können. Eine Beeinträchtigung der MMR konnte mit Krebs in Verbindung gebracht werden, was dazu führte, dass die Konsequenzen einer fehlenden MMR in der letzten Dekade sehr ausführlich studiert wurden. Es ist trotz alledem immer noch unklar, ob bösartige Tumore das Resultat eines Mutatorphänotypen in MMR-defizienten Zellen sind, oder ob andere, bis anhin unbekannte Funktionen der MMR-Proteine dafür verantwortlich sind. Um diese Frage zu beantworten, haben wir uns entschlossen ein zelluläres System zu entwickeln, in welchem der Mutatorphänotyp nicht durch eine gestörte MMR induziert wird. Wir wollten die Genauigkeit der DNA-Replikation dadurch beeinflussen, dass wir Mutationen in die Polymerase- und Exonukleasedomäne der replikativen Polymerasen einfügen. Allerdings wurde dieser Ansatz dadurch erschwert, dass diese Enzyme lebensnotwendig sind. Das Ziel meiner Studie war es ein System zu entwickeln, das diese Schwierigkeiten überwindet und uns erlauben würde die Genauigkeit der Replikation durch modifizieren der Basenselektivität und/oder der Korrekturlesefunktion von Polymerase δ zu verändern.

Ich habe mich auf bekannte Mutationen aus Hefe und Maus konzentriert. Im Gegensatz zu jenen Studien, wollte ich ein isogenes System entwickeln, in welchem Enzymvarianten die endogene Aktivität durch eine Kombination von Technologien

zur induzierbaren Expression und Herabregulation durch Ribonukleinsäure- (ribonucleic acid, RNA) Interferenz ersetzen. Dadurch könnten Wildtyp- und Mutatorphänotyp in ein und derselben Zelllinie untersucht werden. Darüber hinaus sollte die Flexibilität des induzierbaren Systems sicherstellen, dass die exogene Variante in derselben Grössenordnung wie das endogene Wildtypprotein exprimiert wird. Ich verwende den Ausdruck Genaustausch für dieses System. Es findet zwar keine Manipulation des endogenen Gens im Genom statt, aber ein exogenes Gen in einem regulierbarem, stabil integriertem Expressionsvektor ersetzt dessen Funktion.

Durch Verwendung des T-REx Systems war ich in der Lage, den Wildtyp von p125, der grossen Untereinheit von Polymerase δ , und 3 Varianten mit Korrekturlesedefizienz, Fehleranfälligkeit oder der Kombination aus Korrekturlesedefizienz und Fehleranfälligkeit induzierbar zu exprimieren. Ich habe bestätigt, dass alle p125-Varianten Polymeraseaktivität *in vitro* aufweisen und dass die mutierten Varianten ein nicht-komplementäres Nukleotid mit grösserer Effizienz einbauen. Zelllinien, welche die verschiedenen Varianten zusammen mit dem endogenen Wildtypenzym exprimierten, zeigten dennoch keine reduzierten Wachstumsraten oder morphologische Veränderungen. Dies hat angedeutet, dass wir das Wildtypenzym entfernen müssen. Ich habe dies erreicht, indem ich kurze Haarnadelstruktur-RNA (short hairpin RNA, shRNA), welche ausschliesslich die endogene p125 Boten-RNA angreift, induzierbar exprimiert habe. Die entwickelte Methode kombiniert induzierbare Expression einer Variante mit gleichzeitiger Expression von shRNA und führt zu einem erfolgreichen Austausch von endogener Polymerase δ mit einer Variante der freien Wahl. Eine reduzierte Lebensfähigkeit wurde in Zellen beobachtet, in welchen das endogene Protein durch eine Mutatorvariante ersetzt wurde. Es wird weitere Studien benötigen um zu zeigen, ob die beobachtete reduzierte Lebensfähigkeit in der Tat eine Folge eines Mutatorphänotypen ist.

Um die Anwendung dieser Strategie für die Zukunft zu vereinfachen, habe ich einen Vektor konstruiert, welcher es erlaubt, den obigen Wildtyp oder Mutatorvarianten von Polymerase δ in einem einzigen Schritt in eine beliebige menschliche Zelllinie zu transferieren. Da diese Methode für jedes beliebige Gen anwendbar ist, stellt das Werk dieser These den Nachweis der grundsätzlichen Wirksamkeit, welcher ein weites Feld an Forschungsmöglichkeiten und Anwendungen eröffnet.

Table of contents

<i>Summary</i>	3
<i>Zusammenfassung</i>	5
1 Abbreviations	9
2 Introduction	11
2.1 DNA Replication	11
2.1.1 DNA structure forms the basis for replication	11
2.1.2 Prokaryotic DNA replication	12
2.1.3 Eukaryotic DNA replication	16
2.2 Mismatch Repair	23
2.2.1 Mismatch repair in E. coli	23
2.2.2 Mismatch repair in eukaryotes	24
2.3 Replication fidelity	27
2.3.1 Base selectivity and frameshift mutations	28
2.3.2 Proofreading	29
2.3.3 Mismatch repair	31
2.3.4 Mutator polymerases	34
2.4 Mutator phenotype and cancer	38
2.4.1 The chromosomal instability (CIN) mutator phenotype	38
2.4.2 Point mutation mutator phenotype.....	39
2.4.3 Microsatellite instability mutator phenotype	40
3 Aim of my studies	43
4 Results	44
4.1 Strategy rationale	44
4.2 Knock-down of p125 expression with small interfering RNA (siRNA)	46
4.3 Non-inducible p125 expression	48
4.4 Inducible POLD1 promoters	49
4.5 Inducible p125 expression	54
4.6 Polymerase activity	57
4.7 Small-hairpin RNA (shRNA) knock-down of p125 expression	62
4.8 Gene replacement	64

5	<i>Discussion</i>	69
5.1	Project overview	69
5.2	Future directions	72
5.2.1	U2OS T-REx clones	72
5.2.2	Mutation rates	73
5.2.3	Tumorigenesis	74
5.3	Expected mutator phenotypes	75
5.4	Possible applications	76
5.4.1	General gene replacement	76
5.4.2	<i>POLD1</i> gene replacement	77
6	<i>Conclusions</i>	79
7	<i>Materials and methods</i>	80
7.1	Vector construction	80
7.1.1	Inducible <i>POLD1</i> promoters	80
7.1.2	<i>POLD1</i> expression vectors	83
7.1.3	shRNA expression vectors	89
7.1.4	All-in-one <i>POLD1</i> gene replacement vectors	93
7.2	Cell culture	96
7.2.1	Cell lines	96
7.2.2	Vector transfection and isolation of stable clones	97
7.2.3	siRNA transfection	97
7.2.4	Dual luciferase promoter activity assays	97
7.3	Analytical procedures	98
7.3.1	Whole cell extracts	98
7.3.2	Western blot analysis	98
7.3.3	Gel sequencing	99
7.4	Polymerase δ analysis	99
7.4.1	Preparation of nuclear extracts from U2OS T-Rex cells	99
7.4.2	Immunoprecipitation of polymerase δ	100
7.4.3	Primer extension assay	100
8	<i>References</i>	102
9	<i>Acknowledgements</i>	119
	<i>Curriculum Vitae</i>	120

1 Abbreviations

A	deoxyadenylate
β-tub	β-tubulin
bp	base pair
C	deoxycytidylate
cDNA	complementary DNA
CIN	chromosomal instability
d.m.	double mutant
DNA	deoxyribonucleic acid
DMSO	dimethyl sulfoxide
dNTP	deoxyribonucleoside triphosphate
dox	doxycycline
dsDNA	double stranded DNA
e.p.	error prone
EXO1	exonuclease I
EV	empty vector
G	deoxyguanylate
GR	gene replacement
HMGB1	high-mobility group box 1
HNPCC	hereditary non-polyposis colon cancer
<i>hprt</i>	<i>hypoxanthine-guanine phosphoribosyltransferase</i>
indel	insertion and deletion
iRNA	initiator RNA
MCM	mini-chromosome maintenance
MEF	mouse embryonic fibroblast
<i>MLH</i>	<i>MutL</i> homolog
MMR	mismatch repair
mRNA	messenger RNA
<i>MSH</i>	<i>MutS</i> homolog
MSI	microsatellite instability
NMP	nucleoside monophosphates
NT	not transfected

ORF	open reading frame
p.def.	proofreading deficient
PCNA	proliferating cell nuclear antigen
PMS	postmeiotic segregation
PMSF	phenylmethylsulfonyl fluoride
pol	DNA polymerase
RFC	replication factor C
RFP	red fluorescent protein
RNA	ribonucleic acid
RNAi	RNA interference
RPA	replication protein A
shRNA	short hairpin RNA
siRNA	small interfering RNA
SSB	ssDNA binding protein
ssDNA	single stranded DNA
<i>SV40</i>	<i>Simian virus 40</i>
T	deoxythymidylate
TFIIH	transcription factor II H
tet	tetracycline
TETO	tet operator
TETO2	double tet operator
tetR	tet repressor
TGF	Transforming growth factor
UTR	untranslated region
wild type	wild type

2 Introduction

2.1 DNA Replication

2.1.1 DNA structure forms the basis for replication

Deoxyribonucleic acid (DNA) is a carrier of genetic information. Before a cell can divide, this information needs to be duplicated so that it can be passed on to the 2 daughter cells of the next generation. The structure of DNA, which was discovered over 5 decades ago by Watson and Crick [1], forms the basis for its replication. Pyrimidines and purines (Figure 1A) are the information carriers in DNA. The first nitrogen atom of a purine molecule and the 9th nitrogen of a pyrimidine bind to the first carbon atom of deoxyribose. Single-stranded DNA (ssDNA) is a linear polymer composed of monomers known as *nucleotides*. Adjacent nucleotides are linked by phosphodiester bonds (Figure 1B). The ssDNA polymer is characterized by *chemical directionality* represented by the presence of a hydroxyl group at the 3' end and a phosphate group at the 5' end. Two complementary strands of DNA are joined to form a double-helix. In this form, the single strands are arranged in an anti-parallel fashion, with the 5' end of one strand opposite the 3' end of its partner.

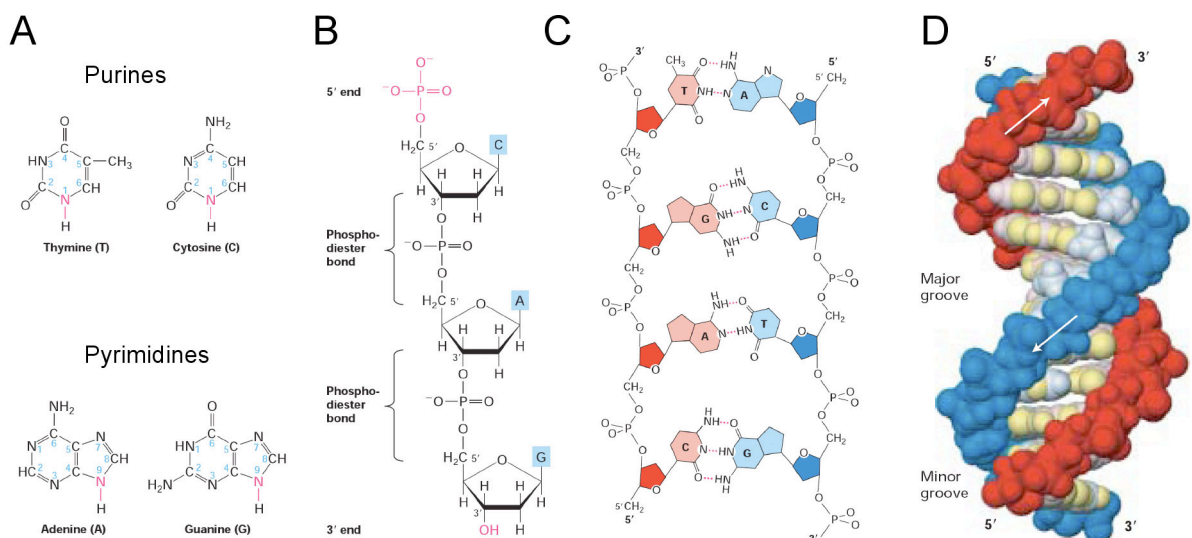


Figure 1: Structure of DNA

A: Pyrimidines and purines are the information carriers in DNA. B: Single-stranded (ss) DNA
C: Double-stranded (ds) DNA. D: The DNA double helix. (Adapted from [2])

The 2 strands of the double helix are held together by the formation of base pairs (bp). Deoxyadenylate (symbolized by the letter A) on one strand pairs with deoxythymidylate (T) on the opposite strand through the formation of 2 hydrogen bonds. In a similar manner, deoxyguanylate (G) pairs with deoxycytidylate (C) through the formation of 3 hydrogen bonds. (Figure 1C). The double helix formed by these entwined complementary strands is referred to as double stranded DNA (dsDNA). The 2 sugar-phosphate “backbones” are on the outer of the double helix with the bases projecting inward. The sides of the bases are accessible through gaps between the coils of the 2 backbones, which form the major and minor grooves on the surface of the molecule (Figure 1D).

DNA replication cannot begin until the 2 strands of the double helix have been separated, a process carried out by enzymes known as *helicases*. Once the hydrogen bond donor and acceptor groups on each base have been exposed, a base-pairing mechanism is used to “copy” the nucleotide sequence of each DNA strand. The resulting copy is actually a sequence that is complementary (rather than identical) to the original. The appropriate deoxyribonucleoside triphosphate (dNTP) is added to the new DNA chain by polymerization, a process catalyzed by enzymes known as DNA polymerases (pol). DNA polymerization always proceeds in a 5'-to-3' direction [2-4]. Consequently, the mechanisms underlying the synthesis of the 2 strands, which are anti-parallel, are characterized by substantial differences, which will be discussed below.

2.1.2 Prokaryotic DNA replication

DNA replication has been explored in great detail in the prokaryotic organism *Escherichia coli*. Its characteristics will be summarized in this section.

2.1.2.1 Replisome components

The multicomponent machinery responsible for DNA replication is collectively referred to as the *replisome*. Table 1 provides an overview of the replisome components that assemble at the replication fork, i.e., the point in the dsDNA at which the strands are separated to allow synthesis of a new complementary copy.

DNA polymerase III core

The catalytic core of *E. coli*'s replicative DNA polymerase III holoenzyme (pol III core) was first identified in a mutant *E. coli* strain that lacked pol I activity [5]. It is a heterotrimer composed of α , ϵ , and τ subunits [6]. The α subunit is responsible for the polymerase activity, and the ϵ subunit provides proofreading 3'-to-5' exonuclease activity. The τ subunit binds to and thereby activates the ϵ subunit [7].

The processivity of pol III is remarkably reduced in the absence of the ϵ subunit. This mechanism ensures that this proofreading subunit is present during DNA replication [8]. A correctly matched primer-template pair is a poor substrate for degradation by the exonuclease ϵ , but a good substrate for further polymerization by the α subunit. In contrast, a mismatched primer terminus is readily degraded by the ϵ exonuclease. Since it is also a poor substrate for polymerization by the α subunit, there is more time for its degradation by the exonuclease ϵ [9, 10]. The Pol III core is responsible for polymerizing both strands of the DNA.

Table 1: Components of the *E. coli* replisome^a

Replisome component [stoichiometry ^{b,d}]	Gene	Mol. wt. (kDa)	Function
Pol III holoenzyme ^c		791.5 ^c	Dimeric, ATP-dependent, processive polymerase/clamp loader ^c
Pol III star ^c		629.1 ^c	Dimeric polymerase/clamp loader ^c
Core ^c		166.0 ^c	Monomeric polymerase/exonuclease ^c
α [2]	<i>dnaE</i>	129.9	DNA polymerase
ϵ [2]	<i>dnaQ</i>	27.5	3'-5' Exonuclease
θ [2]	<i>holE</i>	8.6	Stimulates ϵ exonuclease
γ/τ complex ^c		297.1 ^c	ATP-dependent clamp loader ^c
γ/τ [1/2]	<i>dnaX</i>	47.5/ 71.1	ATPase, τ organizes Pol III star and binds DnaB
δ [1]	<i>holA</i>	38.7	Binds β clamp
δ' [1]	<i>holB</i>	36.9	Stator, stimulates γ ATPase in ATP site 1
χ [1]	<i>holC</i>	16.6	Binds SSB
ψ [1]	<i>holD</i>	15.2	Connects χ to clamp loader
β [2 dimers]	<i>dnaN</i>	40.6	Homodimeric processivity sliding clamp ^c
Primase [1]	<i>dnaG</i>	65.6	Generates RNA primers for Pol III holoenzyme
DnaB helicase [6]	<i>dnaB</i>	52.4	Unwinds duplex DNA 5'-3' ahead of the replication fork ^c
SSB [4]	<i>ssb</i>	18.8	Melts secondary structure in ssDNA, binds clamp loader through χ ^c

^aAbbreviations include SSB, single-stranded DNA-binding protein, and ssDNA, single-stranded DNA

^bRefers to the stoichiometry in the holoenzyme and replisome

^cRefers to a protein complex

^dColored boxes indicate a hierarchy of Pol III subassemblies that together comprise the holoenzyme

(from [11])

The sliding β clamp

The sliding β clamp is a homodimer of the beta clamp protein, which is encoded by the *dnaN* gene. It forms a ring with a large central channel, which encircles the DNA double helix [12]. The sliding β clamp confers processivity to pol III by binding it to the template DNA [13]. The closed ring of the clamp is very stable: its half life on the DNA helix is 72 min at 37° C [14]. The 2 β -protein monomers are bound to each other in a head-to-tail fashion, so the ring has 2 distinct faces. The C-terminal ends of the 2 monomers protrude on one side with the N-terminal ends on the other side [15]. The C-terminal side is involved in the β clamp protein's interaction with pol III and the clamp loader [16, 17].

The γ/τ clamp loader complex

The γ/τ clamp loader complex is responsible for loading the sliding β clamp onto the DNA. The clamp loader that is associated with DNA replication consists of 6 subunits: $\gamma_1\tau_2\delta_1\delta'_1\chi_1\psi_1$. The γ and τ subunits are encoded by the same gene, *dnaX*. The τ subunit is produced by translation of the full-length gene, whereas the γ subunit is a truncated version that is produced by a ribosomal frameshift [18, 19]. The C-terminal part of subunit τ contains 2 domains: one for binding the pol III core, the other for interaction with the replicative helicase, DnaB [20]. The 2 τ subunits of the clamp loader link the 2 pol III cores that simultaneously synthesize the 2 strands of the DNA. A clamp loader consisting only of $\gamma_3\delta_1\delta'_1$ subunits is still capable of clamp loading [21]. The 3 γ subunits bind and hydrolyze ATP. The δ subunit is capable of opening a β clamp on its own, and the δ' subunit is thought to act merely as a static body. The χ and ψ subunits are not involved in the clamp-loading mechanism [22]. The χ subunit mediates contact between the clamp loader and the ssDNA binding protein (SSB), a homotetramer present at the replication fork that serves to protect ssDNA and melt hairpins, and the ψ subunit connects the χ and γ subunits [23, 24].

Replicative helicase

The replicative helicase DnaB is assembled as a ring-shaped homohexamer, which encircles the lagging strand and unwinds the dsDNA [25]. The helicase also activates the ribonucleic acid (RNA) primase encoded by *DnaG* [26].

2.1.2.2 Replication model

The 2 DNA polymerases are connected by the clamp loader to the replicative helicase DnaB to form a complex known as the *replisome*. The replisome moves along the DNA molecule as it is being unwound by DnaB. Both strands are copied at the same time by 2 pol III core enzymes. These enzymes initiate DNA synthesis from short RNA primers, which are synthesized by a specialized RNA polymerase called *primase* [27]. Only a few priming events (sometimes only one) are required for synthesis of the so-called *leading strand*, which involves processive polymerization that occurs in the same direction as replication fork movement. In contrast, during synthesis of the *lagging strand*, the polymerase and the replication fork are moving in opposite directions. This strand must be synthesized as a series of separate sequences known as *Okazaki fragments*. Each fragment is initiated by synthesis of a short RNA primer. Interestingly, DnaG primase has to be recruited to the replication fork for each of the regular priming events that give rise to an Okazaki fragment. DnaG primase activity requires interaction with DnaB helicase [28]. The primase's interaction with the τ subunit of the clamp loader limits the primer size to 12 nucleotides [29]. DnaG binds to SSB on the adjacent ssDNA, attaching itself firmly to the primer until it is replaced with the β clamp by the clamp loader [30]. Once the β clamp has been loaded onto the primer, pol III core binds to the clamp and initiates DNA synthesis. Unbound β clamp has a higher affinity for the clamp loader, whereas DNA-bound β clamp has a higher affinity for the pol III core [17]. Synthesis continues until the pol III core reaches the downstream primer of the previous Okazaki fragment. Here, the polymerase dissociates at the nick [13], and the β clamp is left behind. After the polymerase dissociates from the lagging strand, it is available for binding to a β clamp at the beginning of another Okazaki fragment. If the polymerase is "stuck" and unable to reach the end of the last Okazaki fragment, another priming event can trigger the release of pol III core from the β clamp [31]. This mechanism ensures that the leading-strand and lagging-strand polymerases progress at the same speed (Figure 2). The initial RNA primer of each Okazaki fragment is later removed by the 5' to 3' exonuclease activity of pol I and replaced by simultaneous DNA synthesis mediated by the polymerase activity of pol I [32]. This process is known as *nick translation*. The nick is then sealed to ensure the integrity of the DNA. The enzyme responsible for this ligation is called *DNA ligase* [33].

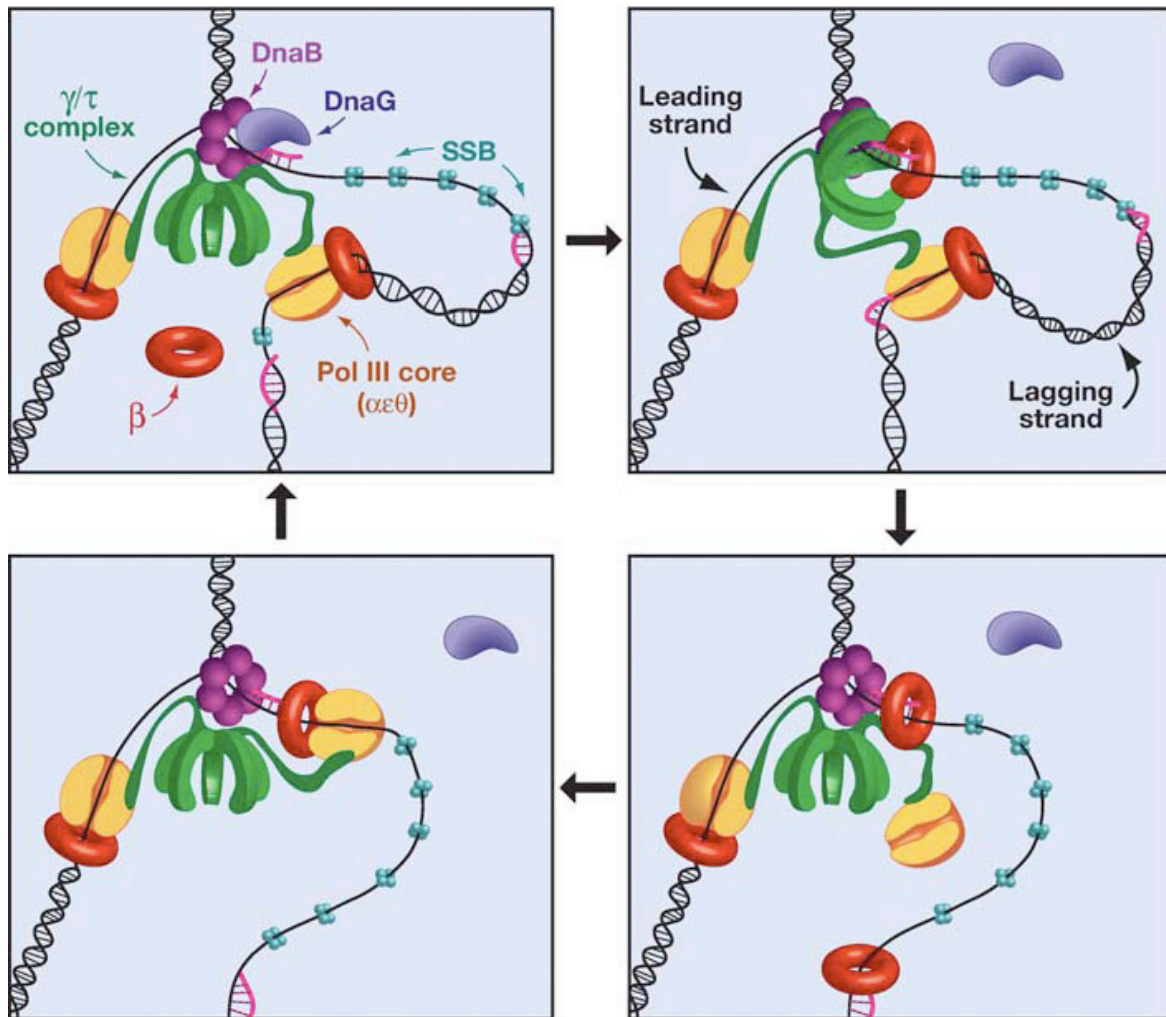


Figure 2: The *E. coli* replication fork

As the replisome advances, the clamp loader (γ/τ complex) loads a β clamp onto an RNA primer (pink) that has been synthesized by primase DnaG (upper right panel). When the lagging-strand polymerase reaches the end of ssDNA, it dissociates from the DNA and the β clamp (lower right panel) and cycles to newly loaded β clamp (lower left panel). (Adapted from [11])

2.1.3 Eukaryotic DNA replication

Eukaryotic DNA replication has been studied in greatest depth in yeast. The focus of this section will be the model system developed on the basis of these studies.

2.1.3.1 Replisome components

Replicative polymerases

The 3 polymerases needed for DNA replication (α , δ , and ϵ) are conserved in all eukaryotes. The replicative polymerases are enzymes composed of multiple

subunits (Table 2). Apart from nomenclature, the main difference between human polymerases and those of *Saccharomyces cerevisiae* is that the yeast polymerase δ has 3 rather than 4 subunits.

Table 2: Subunits of eukaryotic replicative DNA polymerases

Polymerase	Human			<i>S. cerevisiae</i>			Activity
	Gene	Protein	Mass ^a (kDa)	Gene	Protein	Mass ^a (kDa)	
Pol α - primase	<i>POLA</i>	p180	165.9	<i>pol1</i>	Pol1	166.8	DNA polymerase
	<i>POLA2</i>	p68	66.0	<i>pol12</i>	Pol12	78.8	
	<i>PRIM1</i>	p48	49.9	<i>pri1</i>	Pri1	47.7	RNA primase
	<i>PRIM2</i>	p55	58.8	<i>pri2</i>	Pri2	62.3	
Pol δ	<i>POLD1</i>	p125	123.6	<i>pol3</i>	Pol3	124.6	DNA polymerase, 3'-to-5' exonuclease
	<i>POLD2</i>	p50	51.3	<i>pol31</i>	Pol31	55.3	
	<i>POLD3</i>	p66	51.4	<i>pol32</i>	Pol32	40.3	
	<i>POLD4</i>	p12	12.4	No homolog present			
Pol ϵ	<i>POLE</i>	p261	261.5	<i>pol2</i>	Pol2	255.6	DNA polymerase, 3'-to-5' exonuclease
	<i>POLE2</i>	p59	59.5	<i>dpb2</i>	Dpb2	78.3	
	<i>POLE3</i>	p17	16.9	<i>dpb3</i>	Dpb3	22.7	
	<i>POLE4</i>	p12	12.2	<i>dbp4</i>	Dpb4	22.0	

^aThe subunit mass was deduced from the primary sequence.

For many years, the specific roles of polymerases δ and ϵ in leading- and lagging- strand synthesis remained unclear. A study conducted with proofreading-deficient mutants suggested that pol δ and ϵ proofread opposite strands in the replication fork [34]. There was also evidence that pol δ — but not pol ϵ — proofreads the errors made by pol α , which is naturally exonuclease-deficient [35] and known to be responsible for starting new Okazaki fragments on the lagging strand. The *pol1-L868M* mutation of pol α was associated with an increased mutation rate [36]. When this mutation was combined with mutations involving the proofreading exonuclease domains of polymerases δ and ϵ , only pol δ displayed synergistically increased mutation rates [35].

More light has recently been shed on the division of labor that occurs at the eukaryotic replication fork by 2 studies [37, 38] conducted with polymerases δ and ϵ

variants with mutations around the polymerase active sites. These alterations led to a slight increase in mutation rates with little or no effect on proofreading activity. Pursell *et al.* (2007) performed an *in vitro* fidelity analysis of pol ϵ harboring a methionine to glycine substitution at amino acid 644 (*pol2-M644G*) and found an error rate for T•dTTP that was at least 39 times higher than the reciprocal error rate for A•dATP [37]. They investigated the mutational signature of a haploid *pol2-M644G* mutant yeast strain using a *URA3* reporter gene, which was placed in the yeast genome in either a forward or reverse direction, close to the *ARS306* origin of replication. It had previously been shown [39] that the orientation of the replication fork with respect to this reporter can be clearly identified, because *ARS306* fires nearly every cell cycle. Pursell *et al.* analyzed mutational spectra at the *URA3* locus for AT to TA transversion mutations. These *in vivo* studies revealed transversions that were mainly the result of T•dTTP misincorporations, allowing the authors to attribute leading-strand synthesis to polymerase ϵ activity [37]. In a similar study with a mutant *Pol* δ allele (*pol3-L612M*), the likelihood of a T•dGMP mismatch was over 28 times higher than that of an A•dCMP error *in vitro*. *In vivo URA3* mutational analysis led to the conclusion that pol δ is responsible for lagging-strand replication [38].

DNA polymerase α -Primase

DNA polymerase α -primase is the only polymerase capable of initiating DNA replication in eukaryotic cells. It combines primase and DNA polymerase activities in one 4-subunit complex [40]. All 4 subunits are essential in *S. cerevisiae* [41]. Pol1 is capable of polymerase activity, but lacks exonuclease activity. The exonuclease domain is conserved, but catalytic residues in the exonuclease active site have been replaced by amino acids that cannot function in the proofreading reaction [42]. Pol12 binds tightly to Pol1 and stabilizes the catalytic subunit [43]. The small primase subunit Pri1 is known to be responsible for the synthesis of oligoribonucleotides. On its own, it is very unstable; it is stabilized by the second primase subunit Pri2, which also mediates contact between the primase complex and Pol1 [44]. Pri1 synthesizes RNA primers 9-10 nucleotides in length, which are extended by Pol1 [45]. Pol1 has been shown to interact with the catalytic subunit of pol δ , Pol3 [46].

DNA polymerase δ

As noted above, DNA polymerase δ (pol δ) is known to be responsible for lagging-strand replication [38]. The largest subunit of pol δ , Pol3, contains both the polymerase and 3'-to-5' exonuclease activities. Pol3 forms a stable complex with Pol31, and Pol32 is attached to this complex by its interaction with Pol31 [47]. The activity of Pol δ is stimulated by the clamp proliferating cell nuclear antigen (PCNA) [48]. DNA polymerase δ is a 3-subunit enzyme in *S. cerevisiae*, whereas the human homolog contains 4 subunits [49, 50]. The 4th subunit in the human enzyme is known to be involved in complex stabilization and the stimulation of polymerase activity [51, 52].

DNA polymerase ϵ

DNA polymerase ϵ , which is responsible for leading-strand replication [37], is a heterotetramer consisting of Pol2, Dpb2, Dpb3, and Dpb4 [53, 54]. PCNA stimulation produces only mild increases in pol ϵ activity (about 2-fold) since the enzyme is highly processive on its own [55].

Proliferating cell nuclear antigen

PCNA was first identified as nuclear antigen abundantly expressed in proliferating cells (as the name suggests) [56]). The PCNA clamp and β clamp are structurally similar, both consisting of a ring-shaped structure containing 6 domains. However, the β clamp is a homodimer consisting of a 3-domain subunit, while the PCNA clamp is a homotrimer consisting of a 2-domain subunit [57]. Compared with that of the β clamp, the closed ring of the PCNA clamp is a bit less stable on DNA, perhaps because its integrity depends on interaction with an additional subunit. The half life is 24 min at 37° C [14]. The 3 PCNA monomers bind to one another in a head-to-tail fashion, giving the ring 2 distinct faces. The C-terminal ends protrude on one side, and this face is involved in a variety of interactions whereby PCNA contributes not only to replication, but also to recombination, repair, and cell cycle control [58].

Clamp-loader complex

Replication factor C (RFC) is the eukaryotic clamp-loader complex. It was originally identified as a factor required for the *in vitro* replication of *Simian virus 40* (SV40) [59]. It consists of 5 subunits (Rfc1, Rfc2, Rfc3, Rfc4, and Rfc5), which are homologous to one other and to the γ/τ complex of *E. coli* [60]. RFC recognizes and loads PCNA onto the 3' termini of the template primer. DNA binding stimulates the ATPase activity of RFC, which is then released from the DNA [61]. Based on their analysis of the structure of an RFC:PCNA complex, Bowman *et al.* (2004) proposed the following model of the clamp-loading mechanism: Upon ATP-binding, RFC assumes a spiral conformation that allows it to attach to the end of the DNA double helix like a screw-cap. RFC binds to the C-terminal face of PCNA and loads it onto the dsDNA. After ATP hydrolysis, RFC is released, and PCNA is left with its C-terminal side facing the 3' terminus of the primer [62].

Replicative helicase

The mini-chromosome maintenance (MCM) 2-7 helicase is believed to function as the replicative helicase. MCM2-7 is a heterohexamer consisting of 6 essential subunits, which are numbered from 2 to 7 [63, 64]. The MCM2-7 complex of *Xenopus laevis* clearly displayed helicase activity when stimulated with another replication-fork protein, Cdc45 [65]. In *Drosophila melanogaster*, MCM2-7, together with PCNA, has been observed to follow BrdU-stained newly synthesized DNA *in vivo* replication [66]. A subset of the *S. cerevisiae* MCM2-7 complex consisting of subunits 4, 6, and 7 has displayed helicase activity *in vitro* characterized by 3'-to-5' polarity [67, 68]. The remaining MCM subunits (2, 3 and 5) are believed to have regulatory functions [69]. Recently, the MCM2-7 complex of *S. cerevisiae* was shown to have helicase activity *in vitro* [70]. The authors believe that a gap between Mcm2 and Mcm5 opens the toroid structure and inhibits helicase activity while allowing ssDNA binding. Closure of this gap and the initiation of helicase activity was found to be dependent on ATP and the anion content of the reaction buffer.

2.1.3.2 Replication model

Replication in eukaryotes appears to be more or less similar to that observed in prokaryotes. Not all functions have been proven, and additional proteins are believed to play roles [71], so the models discussed here are merely hypothetical.

Polymerases δ and ϵ were found to be responsible for lagging- and leading-strand synthesis, respectively [37, 38]. However, the catalytic domain harboring the polymerase activity of polymerase ϵ is not essential for cell growth and viability [72]. Therefore, some investigators have suggested that when polymerase ϵ is dysfunctional, an alternative replication fork can be assembled with polymerase δ on both the leading and the lagging strands (Figure 3; [73, 74]). It remains to be seen whether this alternate fork plays a role during normal DNA replication and which factors trigger its assembly. It could be a response to replicational stress or replication fork restart or depend on chromosomal context [73, 74].

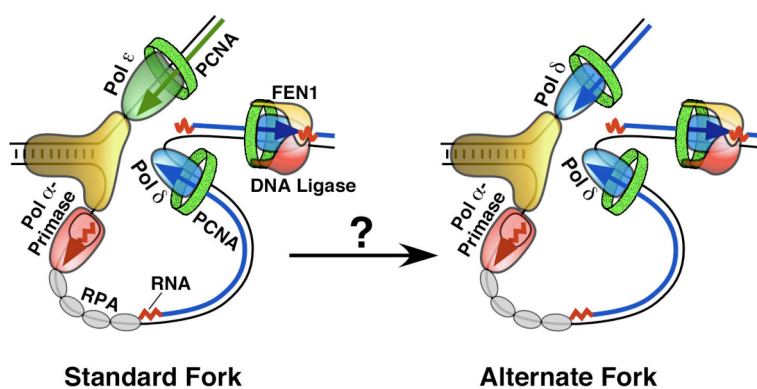


Figure 3: Eukaryotic DNA replication

Two proposed models of eukaryotic DNA replication. The conditions leading to transition from one model to another are currently unknown. (Adapted from [73])

The DNA needs to be unwound by the replicative helicase (in all probability MCM2-7). Single-stranded DNA is stabilized by the heterotrimeric ssDNA-binding protein known as replication protein A (RPA) [75]. The 2 strands are copied simultaneously by polymerase ϵ and δ , or by polymerase δ alone. DNA synthesis is initiated from short RNA/DNA hybrid primers, which are synthesized by polymerase α -primase. PCNA is loaded onto one of these primers by RFC and used as a clamp to tether polymerase δ and ϵ to the DNA template. DNA synthesis on the leading strand requires only a few priming events (sometimes only one) since polymerization is processive and occurs in the same direction as replication-fork movement. In contrast, the lagging strand is synthesized in a discontinuous manner because here, the polymerase and replication forks are moving in opposite directions. As a result, the lagging strand has to be synthesized as a series of discontinuous Okazaki

fragments. Each fragment is initiated by the synthesis of a short RNA/DNA hybrid primer. Once PCNA is loaded onto this primer, polymerase δ binds to the primer terminus and continues synthesis. When it encounters the RNA primer of the previous Okazaki fragment, it synthesizes 1 or 2 additional nucleotides, displacing a short flap of the primer (*strand displacement synthesis*). Polymerase δ is known to cooperate with the flap endonuclease FEN1 in degrading the initiator RNA (iRNA) during the maturation of Okazaki fragments [76-78]. FEN1 removes the small 5' flap and releases nucleoside monophosphates (NMP). Pol δ and FEN1 continue this process, which is known as *nick translation*, until all of the iRNA has been degraded. The resulting DNA-to-DNA nick is then sealed by DNA ligase I (Figure 4) [76]. Pol δ is able to offer FEN1 and ligase I different substrates: After limited strand displacement synthesis at the nick, the 3' terminus shifts from the polymerase active site to the exonuclease active site. The displacing strand is then degraded until the nick is reached again, and the 3' terminus shifts back to the polymerase active site. This iterative process is called *idling*. There are alternative pathways for Okazaki fragment maturation. *Rad27 Δ* yeast lacking FEN1 have been found to accumulate duplications up to 100 nt in length [79, 80]. These duplications seemed to arise from flaps up to 100 nt long. Flaps that are longer than 30 nt have been shown to bind RPA, which inhibits FEN1. However, nuclease Dna2 cleaves flaps with bound RPA, but leaves behind a 5' flap of 2-6 nt, which can be removed by pol δ idling [81, 82].

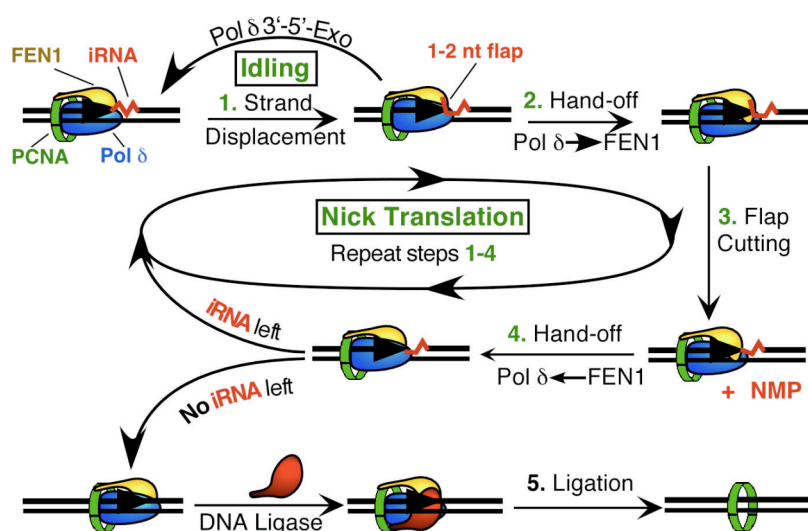


Figure 4: Okazaki fragment maturation

Details of the process are described in the text. (Adapted from [73])

2.2 Mismatch Repair

Postreplicative mismatch repair (MMR) is responsible for the correction of base-substitution errors, as well as insertions and deletions (indels), that arise during replication. MMR is highly conserved from bacteria to man. The chemical composition of a mismatch is indistinguishable from that of normal DNA. Effective MMR must include mechanisms for efficient identification of mismatches and for distinguishing between the parental and daughter strands. The former contains the correct genetic information and serves as the template during replication. The daughter strand contains the erroneous information [83]. In addition to mismatch errors, the MMR machinery also processes a number of other DNA lesions, such as UV photoproducts, DNA cross-links, and bases that have been alkylated, methylated, or oxidized [84-88].

2.2.1 Mismatch repair in *E. coli*

The MMR process has been characterized best in *E. coli*. The system has been reconstituted from defined components [89-91]. The MMR reaction can be divided into 3 steps: initiation, excision, and re-synthesis. It requires the activity of 11 protein complexes or proteins, 4 of which (MutS, MutL, MutH, and UvrD) are MMR-specific. The homodimeric MutS protein is responsible for mismatch or indel recognition during the initiation phase of MMR [92]. It binds to the replication error and recruits the homodimeric protein MutL in an ATP-dependent manner. The formation of this ternary MutS/MutL complex activates MutH endonuclease, which then incises the unmethylated strand of a hemimethylated GATC sequence. Only the newly synthesized strand is nicked, because methylation lags behind the replication fork [93, 94]. MMR has been shown to be bidirectional, so the incision can occur 3' or 5' to the mismatch [90, 91]. The MMR excision step begins with the unwinding of the dsDNA by DNA helicase II, starting from the nick. Interestingly, the unwinding process does not depend on MutL: it can be initiated by any nick within a distance of 1 kb from the mismatch [95]. As GATC methylation is observed only in Gram-negative bacteria, some investigators have suggested that the nicks between Okazaki fragments could serve as strand discrimination signals in Gram-positive bacteria and eukaryotes [96, 97]. The ssDNA containing the replication error is degraded by single-strand-specific exonucleases. When the nick is located 5' to the

mismatch, ExoVII and RecJ degrade the strand with 5'-to-3' polarity. When the nick is 3' to the mismatch, the strand is degraded by ExoI, ExoVII, and ExoX with 3'-to-5' polarity. (ExoVII, as we can see, contributes to both processes.) The degradation process is halted around 100 bp after the mismatch. [91, 98, 99]. The resulting ssDNA gap is stabilized by SSB until DNA polymerase III resynthesizes the DNA (Figure 5). Finally, the remaining nick is sealed by DNA ligase [89], and the MMR process is completed.

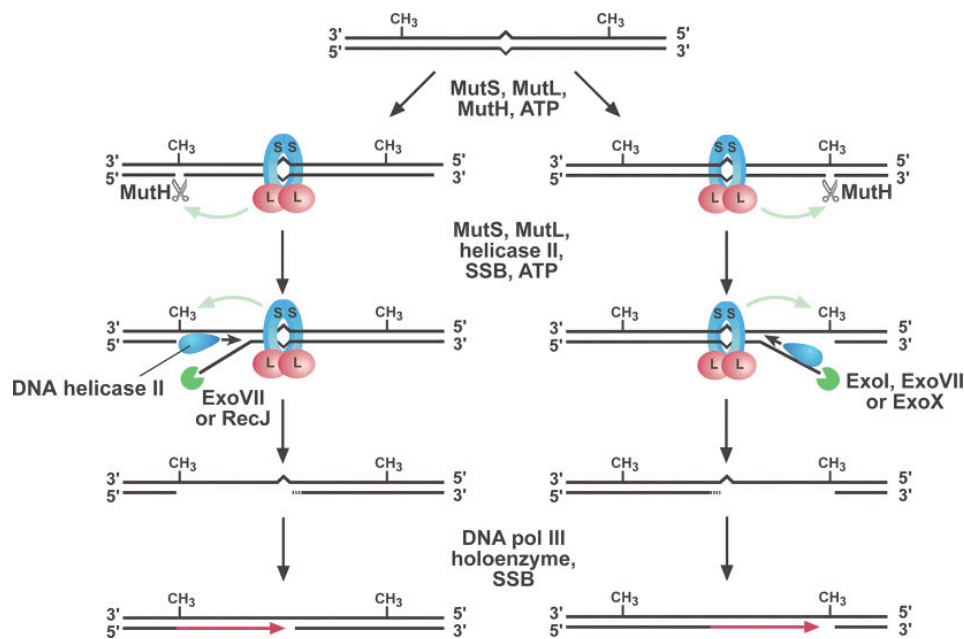


Figure 5: Mechanism of methyl-directed mismatch repair in *E. coli*

Green arrows indicate MutS- and MutL-dependent signaling between the mismatch and the nearest hemimethylated GATC site. Details of the reaction are described in the text. DNA ligase (not shown) restores the integrity of the repaired strand after DNA polymerase III holoenzyme fills in the gap. (Adapted from [100])

2.2.2 Mismatch repair in eukaryotes

Eukaryotes contain many *MutS* and *MutL* homologs, but no homologs of *MutH* have been identified thus far [101, 102]. Genetic studies in yeast have helped to elucidate the roles of eukaryotic MMR proteins. *Saccharomyces cerevisiae* contains 6 *MutS* homologs (*Msh1-6*) and 4 *MutL* homologs (*Mlh1-3* and *Pms1*). *Pms1* was initially identified in a yeast strain displaying increased postmeiotic segregation (hence its gene symbol: *Pms-1*), and this finding revealed the involvement of MMR in homologous recombination processes [103]. The Msh2 protein forms heterodimers

with Msh3 and Msh6, which are known, respectively, as MutS β and MutS α . MutS α recognizes base-substitution mismatches and single-nucleotide indels, whereas only the latter are recognized by MutS β [104-107]. The Msh1 protein is involved in the maintenance of mitochondrial DNA [108]. It is found exclusively in yeasts and is thought to be the founder of all eukaryotic MutS homologs [109]. The Msh4 and Msh5 proteins do not contribute to MMR, but they are involved in meiotic recombination [110, 111]. Mlh1 forms a complex with Pms1 and Mlh3 — MutL α — which is responsible for most MMR processing [112]. The MutL β complex formed by Mlh1 and Mlh3 is involved in certain MMR processes initiated by MutS β [113]. The role of the Mlh1/Mlh2 complex is less clear, but it seems to be involved in meiotic recombination [114].

Human cells express homologs of Msh2-6. The human MSH2/MSH6 complex, (hMutS α), for instance, can initiate the repair of eight different base-base mismatches and indels of up to 8 unpaired nucleotides. The MSH2/MSH3 complex referred to as hMutS β recognizes neither base-base mismatches nor single-nucleotide indels, but it is active on longer indels. Human MutS α and hMutS β are therefore redundant for the repair of indels of 2-8 nucleotides [115-117]. Interestingly, the overexpression of MSH3 results in a phenotype similar to that associated with MSH6 deficiency because it leads to the sequestration of MSH2 [118].

There are 4 human *MutL* homologs: *MLH1*, *MLH3*, *PMS1*, and *PMS2*. The *MLH1* protein is the key player since it forms complexes with each of the other 3 homologs. The *MLH1/PMS2* complex (known as hMutL α) supports repair activity initiated by hMutS α and hMutS β . hMutL α is responsible for most, if not all, MMR activity [119]. The other 2 complexes (hMutL γ formed by *MLH1* and *MLH3*, hMutL β formed by *MLH1* and *PMS1*) are present at low concentrations and have been shown to possess little or no repair activity *in vitro* [120, 121].

The current model of the human MMR system is based on *in vitro* reconstitutions [122, 123]. Mismatch repair is believed to be triggered by the binding of hMutS α to a mismatch or short indel or by hMutS β binding to longer indels. Human MutS α (or hMutS β) then recruits hMutL α . The ternary complex undergoes an ATP-driven conformational change, which allows it to move along the DNA strand like a sliding clamp until it encounters a strand break or nick. This strand break serves as a strand-discrimination signal. The essential set of proteins required for the excision

step depends on the location of the nick. If it is located 5' to the mismatch, the error can be excised by the concerted activities of hMutS α , RPA, and exonuclease I (EXO1), which is activated by the concomitant presence of hMutS α and a mismatch. Once the mismatch has been removed, RPA displaces EXO1, and further EXO1 activity is inhibited by the combined action of hMutS α and hMutL α [124, 125]. The efficiency with which EXO1 is loaded onto the DNA can be increased by the non-histone chromatin protein high-mobility group box 1 (HMGB1) [126].

The processing of a mismatch with a 3' nick requires a larger set of proteins: hMutS α , hMutL α , RPA, EXO1, PCNA, and RFC. RFC loads the PCNA clamp onto the DNA. The combined presence of the RFC and PCNA proteins, the hMutS α complex, and a mismatch stimulates the latent endonuclease activity of hMutL α , which then proceeds to incise the strand displaying discontinuity. This activity is made possible thanks to a single residue in PMS2, which is conserved not only in eukaryotic PMS2 homologs but also in archaeal and eubacterial mutL proteins [127]. The motif is absent, however, in MutL proteins from bacteria like *E. coli*, which rely on GATC methylation for strand discrimination, as discussed in the previous chapter.

Endonuclease activity introduces nicks 5' and 3' to the mismatch. This fact explains why the 5'-to-3' exonuclease EXO1 is able to degrade a substrate with an initial nick 3' to the mismatch. EXO1 does not possess 3'-to-5' exonuclease activity: as an entry site for degradation, it uses the nick introduced by PMS2 5' to the mismatch. The substrates with a 3' nick are converted into a substrates with nicks 3' and 5' to the mismatch (Figure 6). The topography of the replication fork suggests that the 3' terminus of the primer strand serves as the strand-discrimination signal on the leading strand, whereas on the lagging strand, this function is served by nicks between the Okazaki fragments.

EXO1 plays an essential role in the reconstituted MMR system *in vitro*, but its exonuclease function appears to be at least partially redundant *in vivo*, since MMR activity is largely preserved in Exo1^{-/-} mouse cells [128].

The final steps of the MMR process involve refilling of the single-strand gap by DNA polymerase δ and sealing of the nick by DNA ligase I.

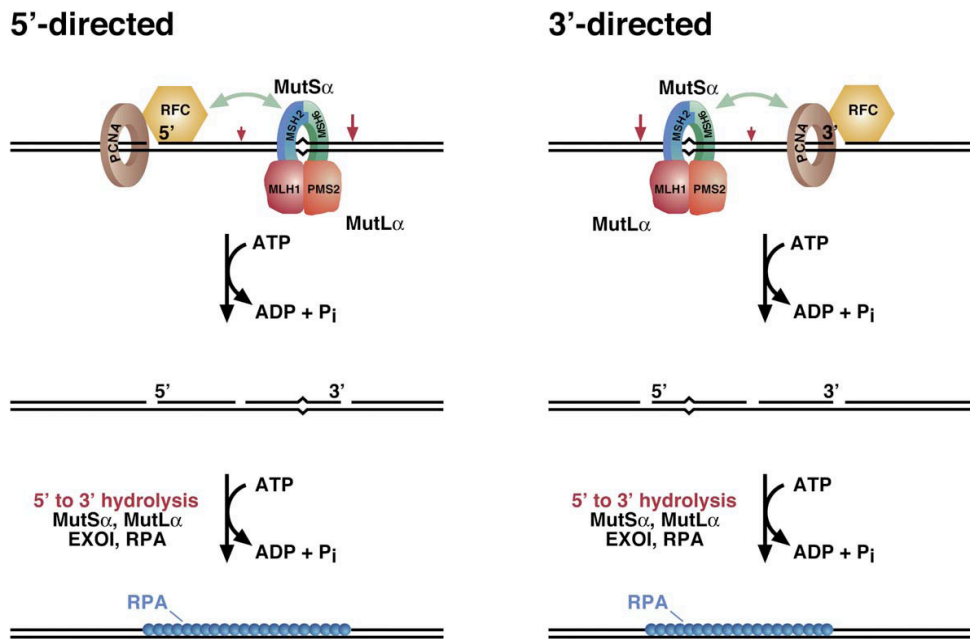


Figure 6: Incision of the discontinuous heteroduplex strand during MMR

MutS α , PCNA, and RFC activate latent MutL α endonuclease, which incises the discontinuous strand of 5' or 3' heteroduplex DNA in an ATP-dependent reaction. Incision displays a bias for the distal side of the mismatch (relative to the location of the original strand break) (large red arrows) but it can also occur proximal to the mispair (small red arrows). For a 3' heteroduplex, this process produces a new 5' terminus on the distal side of the mismatch. This serves as an entry site for MutS α -activated EXOI, which removes the mismatch in a 5'-to-3' hydrolytic reaction controlled by RPA. The strong bias for incision of the discontinuous strand implies signalling along the helix contour, which may involve ATP-promoted movement of MutS α or the MutS α •MutL α complex along the DNA double helix (not shown in the diagram). (Adapted from [129])

2.3 Replication fidelity

The DNA must be replicated exactly to avoid the introduction of unwanted changes into the genetic code. Such changes are referred to as mutations. DNA replication fidelity refers to the accuracy with which the genetic information is duplicated. It is a key determinant of genome stability [130]. Overall mutation rates in prokaryotes and eukaryotes have been estimated to be in the range 1 error per 10^9 - 10^{10} bp replicated [131, 132]. This high level of fidelity is achieved in several stages, as discussed below. The best characterized eukaryotic model system for the study of DNA replication fidelity is *S. cerevisiae*.

2.3.1 Base selectivity and frameshift mutations

The inherent nucleotide selectivity of DNA polymerases provides the greatest contribution to base-substitution fidelity. The fidelity of the 3 eukaryotic replicative polymerases (α , δ , and ϵ) found in eukaryotic cells has been assessed *in vitro* with proteins from *S. cerevisiae*. Polymerase α , which is naturally exonuclease-deficient, was compared with exonuclease-deficient mutant derivatives of polymerases δ and ϵ . Average error rates for single-base substitutions determined by a forward mutation assay were $1.1 \times 10^{-4} \text{ bp}^{-1}$ for pol α [133], $2.1 \times 10^{-4} \text{ bp}^{-1}$ for pol δ [134], and $1.1 \times 10^{-4} \text{ bp}^{-1}$ for pol ϵ [37]. These rates are all in the same general range, although pol δ emerges as a bit less accurate than the other 2 polymerases. Base substitutions account for most mutation events. Selectivity depends mainly on base-pair geometry. The A•T and G•C base pairs are similar in terms of their geometry, whereas the geometry of a mismatched base pair is distorted. The abnormal geometry is thought to cause steric clashes in and around the active pocket of the polymerase, which inhibit efficient catalysis of polymerization [135]. Watson and Crick hydrogen bonds seem to make a minor contribution to polymerase selectivity, since the free-energy differences between complementary and noncomplementary base pairs are relatively small in aqueous solution (0.2-0.4 kcal/mol). However, these differences are amplified when water is excluded from the active site.

Another common class of errors made by the replicative polymerases are indels (insertions and deletions). The consequences of indels in the open reading frame (ORF) of a gene are more severe than those of base-substitution mutations. This type of mutation can cause a shift in the reading frame, which leads to the production of a nonsense protein that eventually is terminated by a stop codon. The so-called *frameshift mutations* are caused by the presence of any indel whose length (in nucleotides) is not a multiple of 3. Single-nucleotide indels are most common. The primer and template strands dissociate and reanneal quite frequently. Single-nucleotide deletions are thought to be the result of *template slippage*, i.e., displacement of the template strand relative to the primer strand. Template slippage would be favored if a misinserted base in the primer strand were complementary to an adjacent base on the template strand [130]. Polymerase α , which is naturally exonuclease-deficient, has been compared with exonuclease-deficient mutant

derivatives of polymerase δ and ϵ . Average error rates for single-nucleotide deletions were $4.0 \times 10^{-5} \text{ bp}^{-1}$ for pol α [133], $4.0 \times 10^{-5} \text{ bp}^{-1}$ for pol δ [134], and $1.7 \times 10^{-5} \text{ bp}^{-1}$ for pol ϵ [37]. Pol ϵ is a bit more accurate than the other 2 polymerases. Primer slippage, i.e., displacement of the primer strand relative to the template strand, seems to be responsible for single-nucleotide insertions [130], which are less frequent than deletions. In the case of polymerase δ , no insertions were found with the assay used in this study. This does not necessarily mean that this polymerase does not make single-nucleotide insertions: its error rate may simply be below the detection limit of the assay. The average single-nucleotide insertion rates reported for the polymerases are $1.2 \times 10^{-6} \text{ bp}^{-1}$ for pol α [133], $\leq 7.9 \times 10^{-7} \text{ bp}^{-1}$ for pol δ [134], and $7.7 \times 10^{-6} \text{ bp}^{-1}$ for pol ϵ [37].

The likelihood of base-substitution and especially indel errors depends on the sequence context. When the primer and template strands dissociate temporarily, a repetitive sequence offers various possibilities for reannealing, and this is one of the reasons that the frequency of frameshift mutations is increased in this sequence context. Exonuclease-deficient *T7* polymerase was tested for single-nucleotide deletions in a homopolymeric run of deoxythymidylates in a reversion assay. Error rates increased with the length of the run, from $1.3 \times 10^{-4} \text{ bp}^{-1}$ for the 3 x T substrate to $1.0 \times 10^{-3} \text{ bp}^{-1}$ for the 7 x T substrate. This corresponded an 7.7-fold increase of the mutation rate in the same sequence context [136]. In the same reversion assay with the 7 x T substrate, error rates for exonuclease-deficient pols δ and ϵ ($4.0 \times 10^{-4} \text{ bp}^{-1}$ and $3.3 \times 10^{-4} \text{ bp}^{-1}$, respectively) were even higher than those for exonuclease-deficient *T7* pol [137].

2.3.2 Proofreading

Polymerases δ and ϵ both contain a functional 3'-to-5' exonuclease domain [42]. Compared with a matched primer, a mismatched primer in the polymerase active site is extended less efficiently. The 3' end of the primer becomes available for binding to the active pocket of the exonuclease. The DNA moves toward this pocket with a rotation in the double-helix axis. After the incorrect nucleotide has been removed, the primer end rotates back to the template strand and the polymerase active pocket [138].

The importance of proofreading becomes clear when one compares the fidelities of polymerases with and without exonuclease activity. For single-base substitutions, the average error rates determined in a forward mutation assay were $3.7 \times 10^{-5} \text{ bp}^{-1}$ for wild-type pol δ [134] and $1.1 \times 10^{-5} \text{ bp}^{-1}$ for pol ϵ [37]. According to these findings, pol ϵ is about 3 times more accurate than pol δ . When these rates are compared with those of proofreading-deficient polymerase mutants, it becomes clear that pol δ and pol ϵ proofread (and correct) 82% and 90% of all base-substitution errors, respectively. As for single-nucleotide deletions, average error rates of $3.8 \times 10^{-6} \text{ bp}^{-1}$ have been reported for pol δ [134] while those determined with the same approach for pol ϵ [37] fell below the detection limits of this assay ($\leq 6.0 \times 10^{-7} \text{ bp}^{-1}$). Therefore, polymerase δ edits 90% and pol ϵ at least 96% of all single-nucleotide deletions. For single-nucleotide insertions, both of these polymerases had error rates that were below detection limits: $\leq 4.7 \times 10^{-7} \text{ bp}^{-1}$ for pol δ [134] and $\leq 6.0 \times 10^{-7} \text{ bp}^{-1}$ for pol ϵ [37]. Based on these numbers, pol ϵ proofreading eliminates at least 92% of all single-nucleotide insertions. It is important to note that a proofreading-deficient pol δ variant has also been found to have a single-nucleotide insertion error rate below the detection limits [134], suggesting that proofreading may not be needed to avoid this type of error.

In a repetitive sequence context, single-nucleotide deletion error rates for pol δ are much higher than they are in an average sequence context (reported above). In a homopolymeric stretch of 7 deoxythymidylates, Fortune *et al.* (2005) found a single-nucleotide deletion error rate of $3.1 \times 10^{-4} \text{ bp}^{-1}$ for wild-type polymerase δ and $4.0 \times 10^{-4} \text{ bp}^{-1}$ for a proofreading-deficient variant of this polymerase [137]. Therefore, in this type of sequence context, pol δ proofreading repairs only 23% of single-nucleotide deletions.

When primer and template strands dissociate and re-anneal at the polymerase active site, there are numerous possibilities for generating a deletion loop. The presence of a mismatched primer terminus would slow polymerization and trigger excision of the loop by the exonuclease active site of the polymerase. In a repetitive stretch, however, a number of paired bases may lie between the mismatch and the primer terminus, and size of this number increases with the length of the repeat sequence [130]. This might explain why pol δ proofreading of frameshift mutations in repetitive sequences has displayed low efficiency.

The situation was different for pol ϵ : wild-type and proofreading-deficient variants of this enzyme were associated with error rates of $2.3 \times 10^{-5} \text{ bp}^{-1}$ and $3.3 \times 10^{-4} \text{ bp}^{-1}$, respectively [137]. If we analyze these numbers, we find that even in a repetitive sequence context, 93% of all single-nucleotide deletions were efficiently proofread by pol ϵ .

The different fidelities of pol ϵ and pol δ might have some important biological consequences. For example, given the latter polymerase's inferior performance in terms of single-nucleotide deletions within repeat sequences, it is not unreasonable to expect a much higher MMR workload on the lagging strand, which is synthesized by pol δ compared with the leading strand, which is the responsibility of pol ϵ .

2.3.3 Mismatch repair

Mismatch repair involves the removal of polymerization errors that escape the proofreading function of the replicative polymerases. Due to its complexity, the fidelity of MMR can be assessed only *in vivo*. Mutation rates depend in part on specific sequence contexts. Mutation rates for a mutational event at a given locus are calculated from the frequency of a selectable phenotype. However, at most of the loci commonly used for determination of mutation rates, frameshift mutations and more complex mutational events are detected more efficiently than base-pair substitutions. For example, it has recently been estimated that only 9% of base-pair substitutions at the *Can1* locus in yeast are detectable by selection for canavanine resistance [139]. Therefore, to estimate the contribution of MMR to replication fidelity, it is not enough to compare mutation rates based on phenotypic assays, which represent only the tip of the mutation iceberg. However, MMR might influence the ratio between base-substitution and frameshift mutations, since the latter mutations are mainly corrected through the MMR pathway, at least in a repetitive sequence context.

Using mutation rates reported in the literature, we estimated error rates for specific mutational events, such as single-base substitutions, single-nucleotide deletions, and single-nucleotide insertions (This was only possible when the reported rates were accompanied by details on the mutational spectrum.). Table 3 summarizes our calculations. The rates reported in the different studies displayed wide variation, and in some studies [80, 107], the number of mutant clones analyzed was quite low. Therefore, quantitative analysis of these data is unlikely to lead to

reliable conclusions. Qualitative analysis shows, however, that inactivation of MMR by *msh2* knockout increased indel error rates much more than base-substitution rates. Similarly, knockout of *mlh1* increased base-substitution rates more than 20-fold over those observed in wild-type yeast, but the increases it produced in single-nucleotide deletions and single-nucleotide insertions were much larger (160-fold and 224-fold, respectively). In short, our calculations indicate that indel repair by the MMR machinery is about 1 order of magnitude more efficient than the one of base-substitution errors. This difference may reflect stronger evolutionary pressure to repair frameshift mutations than those involving base-pair substitutions. In addition, generally speaking, single-nucleotide deletions were observed more frequently than single-nucleotide insertions, irrespective of the MMR status. This difference was already observed when error rates were calculated separately for polymerase selectivity and proofreading activity of DNA synthesis (see sections 2.3.1 and 2.3.2).

Within long homopolymer runs, frameshift mutations are increased dramatically *in vivo*. Frameshift reversion assays have been particularly useful in addressing this issue. Mutation rates have been shown to increase with the length of the run [140]. In *msh2*-deficient yeast, frameshift reversion rates in a homopolymeric run of deoxyadenylates was shown to increase from 4.6×10^{-8} for a 4 x A run to 1.6×10^{-3} for a 14 x A run. This 3.5-fold increase in run length was thus associated with a 35,000-fold increase in the mutation rate. In wild-type yeast, the same increase in run length also increased mutation rates but to a lesser extent (30-fold: from 5.4×10^{-9} for a 4 x A run to 1.6×10^{-7} for a 14 x A run). In this sequence context, MMR increases replication fidelity by 3 orders of magnitude. MMR is the only defense against replicative infidelity in homopolymeric runs, since proofreading is not effective in these sequences.

Table 3: Average error rates at the *Can1* locus in *S. cerevisiae*

Gene inactivated	Number analyzed	Can ^r mutation rate	Error rate [bp ⁻¹] (fold increase) for indicated mutations				Reported mutational spectra				Reference
			base-pair substitution ^a	single-nucleotide deletion ^b	single-nucleotide insertion ^c	complex event ^d	base-pair substitution	deletion -1	insertion +1	complex event	
wild type	227	1.52 x10 ⁻⁷	6.16 x10 ⁻¹⁰ (1.0)	2.08 x10 ⁻¹¹ (1.0)	3.02 x10 ⁻¹² (1.0)	5.29 x10 ⁻¹² (1.0)	150	55	8	14	[139]
<i>msh2</i>	20	4.0 x10 ⁻⁶	3.7 x10 ⁻⁹ (6.0) ^e	1.8 x10 ⁻⁹ (87) ^e	1.1 x10 ⁻¹⁰ (37) ^e	0	3	16	1	0	[107]
<i>msh6</i>	21	1.8 x10 ⁻⁶	9.5 x10 ⁻⁹ (15) ^e	1.5 x10 ⁻¹⁰ (7.0) ^e	0	0	18	3	0	0	
<i>msh3 msh6</i>	22	3.7 x10 ⁻⁶	7.1 x10 ⁻⁹ (12) ^e	1.3 x10 ⁻⁹ (63) ^e	9.4 x10 ⁻¹¹ (31) ^e	0	7	14	1	0	
wild type	20	2.8 x10 ⁻⁷	1.1 x10 ⁻⁹ (1.0)	2.4 x10 ⁻¹¹ (1.0)	1.6 x10 ⁻¹¹ (1.0)	1.6 x10 ⁻¹¹ (1.0)	13	3	2	2	[80]
<i>msh2</i>	20	2.9 x10 ⁻⁶	2.7 x10 ⁻⁹ (2.4)	1.2 x10 ⁻⁹ (52)	1.6 x10 ⁻¹⁰ (10)	0	3	15	2	0	
wild type	84	7.5 x10 ⁻⁸	3.7 x10 ⁻¹⁰ (1.0)	5.0 x10 ⁻¹² (1.0)	1.0 x10 ⁻¹² (1.0)	2.0 x10 ⁻¹² (1.0)	68	10	2	4	[141]
<i>msh3</i>	131	1.3 x10 ⁻⁷	3.7 x10 ⁻¹⁰ (1.0)	2.4 x10 ⁻¹¹ (4.7)	1.7 x10 ⁻¹² (1.7)	1.4 x10 ⁻¹¹ (6.9)	61	42	3	25	
<i>msh6</i>	101	6.7 x10 ⁻⁷	3.7 x10 ⁻⁹ (10)	1.5 x10 ⁻¹¹ (3.0)	2.2 x10 ⁻¹¹ (22)	3.7 x10 ⁻¹² (1.9)	90	4	6	1	
<i>mlh1</i>	56	3.2 x10 ⁻⁶	8.1 x10 ⁻⁹ (22)	8.1 x10 ⁻¹⁰ (160)	2.3 x10 ⁻¹⁰ (224)	3.2 x10 ⁻¹¹ (16)	23	25	7	1	
<i>mlh3</i>	72	1.1 x10 ⁻⁷	4.7 x10 ⁻¹⁰ (1.3)	1.5 x10 ⁻¹¹ (2.9)	0	4.3 x10 ⁻¹² (2.1)	50	17	0	5	

^a $\mu_{\text{bp|BPS}}^{\text{CAN1}} = (\mu_{\text{Can R}} * (\text{nr. BPS}/\text{total nr.})) / \tau_{\text{Can R|BPS}}^{\text{CAN1}}$; ^b $\mu_{\text{bp|del -1}}^{\text{CAN1}} = (\mu_{\text{Can R}} * (\text{nr. del -1}/\text{total nr.})) / \tau_{\text{Can R|indel}}^{\text{CAN1}}$; ^c $\mu_{\text{bp|ins +1}}^{\text{CAN1}} = (\mu_{\text{Can R}} * (\text{nr. ins +1}/\text{total nr.})) / \tau_{\text{Can R|indel}}^{\text{CAN1}}$; ^d $\mu_{\text{bp|CE}}^{\text{CAN1}} = (\mu_{\text{Can R}} * (\text{nr. CE}/\text{total nr.})) / \tau_{\text{Can R|indel}}^{\text{CAN1}}$; abbreviations: BPS, base-pair substitution; del -1, single-nucleotide deletion; ins +1, single-nucleotide insertion; CE, complex event; Nr., number of mutants; $\mu_{\text{Can R}}$ is the phenotypic mutation rate to canavanine resistance. $\tau_{\text{Can R|BPS}}^{\text{CAN1}}$ is the locus specific effective target size to canavanine resistance conditioned on a mutation at the CAN1 locus by a insertion, deletion or other DNA rearrangements - the target size is the whole gene (1773 bp). $\tau_{\text{Can R|indel}}^{\text{CAN1}}$ is the locus specific effective target size to canavanine resistance conditioned on a mutation at the CAN1 locus by a base-pair substitution - the value is 163 bp. $\mu_{\text{bp|BPS}}^{\text{CAN1}}$, $\mu_{\text{bp|del -1}}^{\text{CAN1}}$, $\mu_{\text{bp|ins +1}}^{\text{CAN1}}$ and $\mu_{\text{bp|CE}}^{\text{CAN1}}$ refer to the mutation rates per base pair for base substitutions, single-nucleotide deletions, single-nucleotide insertions, and any other more complex event, respectively. These mutation rates are referred to as *error rates* to distinguish them clearly from loss- or gain-of-function mutation rates. The number in parenthesis shows the fold-increase over the wild-type rate. ^eThe comparison refers to the wild-type data in lane 1, because no mutational spectra was provided for the wild-type control in the study. Calculations were done according to [139].

However, homopolymeric repeats are not the only sequences that are prone to primer/template slippage during DNA replication. Eukaryotic genomes contain large numbers of mono-, di-, and trinucleotide repeats that are collectively referred to as *microsatellites* [142, 143]. These sequence repeats were shown to be unstable in MMR-deficient yeast [144] and also in human MMR-deficient cancer cell lines [145, 146]. Microsatellite instability (MSI) is characterized by small deletions or expansions within short sequence repeats in tumor DNA, as compared with corresponding normal DNA [147]. In MMR-deficient cancer cell lines, spontaneous mutation rates at the hypoxanthine-guanine phosphoribosyltransferase (*hprt*) locus were increased 50- to 750-fold over those observed in MMR-proficient cancer cell lines [148, 149].

Cancer cells contain a huge variety of genetic alterations. More reliable comparisons are possible with genetically defined isogenic cell line pairs. Our laboratory has developed an isogenic system based on the human embryonic kidney cell line 293T in which *MLH1* expression can be regulated by exposure to doxycycline (dox). In the absence of *MLH1* expression, cells exhibited MSI and MMR-deficiency, whereas “reactivation” of this gene restored MMR proficiency and lead to the disappearance of MSI [150]. In another study, an *MLH1*-defective clone of the ovarian carcinoma cell line A2780 was used to generate an isogenic subclone that expressed *MLH1* [151]. A mutation rate of 1.8×10^{-6} at the *hprt* locus was reported for the MMR-deficient A2780MNU-Clone1. The MMR deficiency led to MSI at the *BAT26* locus (3.9×10^{-2}). *BAT26* consists of a homopolymeric run of 26 x A preceded by a shorter homopolymeric run of 5 x T. It is located in the fifth intron of the *MSH2* gene [152] and has been shown to be a very sensitive marker for MSI [147, 153].

MLH1 re-expression in the *MLH1*-1 subclone of A2780MNU-Clone1 reduced the *hprt* mutation rate 40-fold (to 4.5×10^{-8}), and the *BAT26* mutation rate dropped below the detection limit ($\leq 8 \times 10^{-5}$). These findings showed that MMR proficiency increased replication fidelity at the *BAT26* microsatellite at least 500 times [151].

2.3.4 Mutator polymerases

It is difficult to isolate the individual contributions of polymerase, proofreading, and MMR activities to replication fidelity *in vivo*. MMR can be assessed, as shown above, and proofreading-deficient mutants of the replicative polymerases are

available. However, assessing the contribution of proofreading to fidelity *in vivo* has proved to be difficult. A study in *E. coli* provides an excellent example of some of the difficulties involved in distinguishing the contribution of proofreading and MMR [154]. The pol III holoenzyme is the only processive DNA polymerase involved in DNA replication in *E. coli*, and loss of its proofreading activity caused by the introduction of 2 point mutations in the active site of the exonuclease subunit ϵ (*dnaQ926*) has been shown to be lethal. The lethality of the proofreading deficiency in *dnaQ926* mutant *E. coli* has been attributed to excessively high replication-error rates, a phenomenon known as *error catastrophe*. (Mathematically speaking, error catastrophe is unavoidable when the likelihood of inactivating an essential gene exceeds 50% per round of replication.) Surprisingly, transfection with plasmids containing the *E. coli mutL* gene restored viability in these cells. The error catastrophe thus appeared to result not only from the loss of proofreading activity, but also from the saturation of the MMR system caused by this loss. Prior to this study, a similar effect had been observed with the proofreading-impaired *E. coli* mutant *mutD5* [155]. This mutant was viable and showed a strong mutator phenotype. In addition to impaired proofreading activity, it also lacked MMR. The latter deficiency was attributed to saturation of the MMR system since normal MMR could be restored by transfection with plasmids encoding MMR proteins.

The situation is similar in eukaryotes, which have 2 processive replicative polymerases. Haploid yeast (*pol2-4 pol3-01*) with inactivating mutations in the proofreading exonuclease domains of polymerase δ (*pol3-01*) and ϵ (*pol2-4*) were found to be nonviable. However, the diploid double mutant (*pol2-4/pol2-4 pol3-01/pol3-01*) was viable [156]. In diploid cells, which contain 2 alleles per gene, genetic information is redundant, and higher mutation rates can thus be tolerated longer before an error catastrophe leads to cell death. Diploid yeast strains were compared in a frameshift reversion assay at the *his7-2* locus. Compared with wild-type yeast, mutation rates were increased 10-fold in *pol2-4/pol2-4* yeast, 490-fold in *pol3-01/pol3-01* yeast, and 1900-fold in *pol2-4/pol2-4 pol3-01/pol3-01* yeast [156]. In MMR-deficient *pms1/pms1* yeast, there was only a 250-fold increase. Diploid yeast that were heterozygous for the *pol3-01* allele had rates that were increased only 6-fold over those observed in the wild-type strain [157]. This low mutation rate contrasted sharply with that observed in diploid yeast that were homozygous for the *pol3-01* allele. The markedly elevated mutation rates found in these yeast seemed to

be caused by the exclusive presence of the proofreading deficient polymerase δ . Similarly, homozygosity – but not heterozygosity – for a mutation that inactivated the proofreading function of polymerase δ increased mutation rates in mouse cells compared to wild-type cells, as determined by resistance to ouabain [158].

We wondered if proofreading-deficient polymerase δ was being preferentially excluded from the replication fork or if MMR was simply able to correct the vast majority of mismatches generated by this polymerase. Data from yeast studies supported the latter possibility [157]. Diploid *pol3-01/+ pms1 Δ /pms1 Δ* yeast displayed a 2400-fold increase in the mutation rate at the *his7-2* locus. This showed that replication errors were indeed being generated, but they were efficiently repaired by MMR. In the same assay, the combination of polymerase δ proofreading deficiency and MMR deficiency in the diploid yeast strain *pol3-01/pol3-01 pms1 Δ /pms1 Δ* was associated with a 38,000-fold increase in the *his7-2* mutation rate [157]. The authors concluded that the MMR system was not being saturated, when this increase was compared with the 1900-fold increase observed with the *pol3-01/pol3-01* yeast and the 250-fold increase observed with *pms1 Δ /pms1 Δ* yeast, since the rate associated with the combined MMR and proofreading deficiency was roughly equal to the product of the rates associated with the individual deficiencies, as one would expect for 2 pathways that operate in a serial manner to correct an error.

Interestingly, the haploid *pol2-4 pms1 Δ* yeast was viable, although its growth was poor, whereas haploid *pol3-1 pms1 Δ* yeast was completely nonviable [157]. We can speculate that an alternative replication fork consisting of polymerase δ alone contributed to the viability of the haploid *pol2-4 pms1 Δ* strain. Haploid yeast strains were analyzed for frameshift reversion at the *his7-2* locus [157]. Compared with those observed in wild-type yeast, mutation rates were increased 12-fold in *pol2-4* yeast, 150-fold in *pms1 Δ* yeast, 240-fold in *pol3-01* yeast, and 1800-fold in *pol2-4 pms1 Δ* yeast. The multiplicative increase, observed in the double-mutant *pol2-4 pms1 Δ* strain, shows that proofreading and MMR act in serial order to avoid frameshifts during leading-strand replication.

Another class of mutator polymerases contains alterations in the polymerase active site. Li *et al.* (2005) introduced a mutation into the conserved active site of

polymerase δ , which rendered it sensitive to the antiviral drug, phosphonoacetic acid [159]. This *pol3-L612M* mutation was later used to demonstrate that pol δ is responsible for lagging-strand replication [38]. *In vitro* studies revealed 4-fold increase in the error rate for single-base substitutions ($1.4 \times 10^{-4} \text{ bp}^{-1}$) and 21-fold increase in the error rate for single-nucleotide deletions ($8.0 \times 10^{-5} \text{ bp}^{-1}$) compared with wild-type rates. Unlike wild-type pol δ , the mutant enzyme had a detectable error rate for single-nucleotide insertions: $1.1 \times 10^{-6} \text{ bp}^{-1}$ [134]. The *in vivo* mutation rate at the *Can1* locus was also increased 3.5-fold with respect to wild-type rates [159]. Interestingly, the mutation spectrum revealed no increase in the frequency of single-nucleotide deletions [160]. Inactivation of MMR by knockout of *msh2*, *mlh1*, or *pms1* resulted in 11-to-17-fold increases in mutation rates [159]. The combined presence of the *pol3-L612M* mutation with *msh2*, *mlh1*, or *pms1* knock-out produced 284- to 316-fold increases in mutation rates and reduced viability to 60%. The mutation rates in *msh6 Δ* cells and *msh6 Δ pol3-L612M* cells were increased, 6-fold and 134-fold, respectively. Viability was only mildly reduced (80%). There seemed to be a correlation between increases in the mutation rate and reductions in viability. The authors speculated that a portion of the cells died as a result of an error catastrophe. The triple mutant *msh3 Δ msh6 Δ pol3-L612M* cells displayed a 325-fold increase in the mutation rate, which was similar to that observed in the other MMR-deficient cells. The more limited increase in the mutation rate of the *msh6 Δ pol3-L612M* was therefore attributed to the residual MMR capacity of Msh3 [159]. The more than multiplicative increase in mutation rates observed with *pol3-L612M* and MMR deficiency indicated that mutations generated by Pol3-L612M were repaired efficiently by MMR and that the MMR system was not being saturated. This conserved residue in yeast pol δ was subjected to all 19 of the possible amino-acid substitutions [160]. Eight of the resulting mutants were viable as haploids, but only 3 had higher mutation rates at the *Can1* locus than *pol3-L612M*: *pol3-L612K*, *pol3-L612G*, and *pol3-L612N*, which displayed mutation rate increases of 13 fold, 17 fold, and 29 fold, respectively. These mutator strains were also associated with a higher number of aberrant cells during logarithmic growth. Pursell *et al.*, used a homologous mutation in the active site of polymerase ϵ to attribute leading-strand replication to pol ϵ [37]. *Pol2-M644G* mutant pol ϵ displayed a 3-fold increase in error rates for single-base substitutions ($3.4 \times 10^{-4} \text{ bp}^{-1}$), a 5-fold increase in error rates for single-

nucleotide insertions ($3.0 \times 10^{-6} \text{ bp}^{-1}$), and a detectable error rate for single-nucleotide deletions ($2.0 \times 10^{-6} \text{ bp}^{-1}$) *in vitro*. The mutation rate at the *Can1* locus was increased 3.9-fold for *pol2-M604G* yeast, 10-fold for *msh6Δ* yeast, and 610-fold for *msh6Δ pol2-M604G* over wild-type rates *in vivo*. Again, the more than multiplicative increase observed in the double-mutant *msh6Δ pol2-M604G* strain shows that proofreading and MMR act in serial order to avoid frameshifts during leading-strand replication.

2.4 Mutator phenotype and cancer

How is a mutator phenotype manifested in humans? It has long been speculated that a mutator phenotype is the basis of cancer development (or tumorigenesis) [161]. Cancer cells are derived from normal cells and are characterized by uncontrolled growth. Mathematical modeling of the generally increased cancer incidence in the elderly population has led some investigators to conclude that tumorigenesis is dependent on 4-7 rate-limiting steps [162]. Cells need to acquire alterations in at least 6 metabolic pathways in order to become pathogenic [163]. Most cancer cells have multiple alterations in their genomic sequence and continue to evolve towards a more malignant phenotype. Individual cells with acquired growth advantages expand into novel clonal populations, which are themselves subjected to environmental selection pressure within the tumor. Therefore, this evolution is based on the principles of clonal expansion and Darwinian selection [164]. It has been suggested that pre-cancerous cells already exhibit a mutator phenotype, which increases the probability that mutations will be generated in tumor-suppressor and tumor-promoter genes [165]. A mutator phenotype can be characterized by elevated chromosomal instability (CIN), point mutations, or MSI.

2.4.1 The chromosomal instability (CIN) mutator phenotype

This mutator phenotype leads to the accumulation of changes at the chromosomal level, including abnormal chromosome numbers and inversions, deletions, duplications, and translocations of chromosomal parts. These structural and numerical abnormalities are also known as *aneuploidy* [166]. Chromosomal instability (CIN) was assumed to occur in the majority of human cancers. CIN tumors were found to have a chromosomal gain and loss rate of 1×10^{-2} per chromosome and

round of replication, whereas the rate was not detectable for MSI tumors [167]. On the other hand, increased mutation rates at the *hprt* locus were observed in MSI-positive cancer cell lines but not in CIN cancer cell lines [168]. The molecular mechanisms underlying the CIN mutator phenotype were thought to be very heterogeneous, including genes involved in DNA damage checkpoints, chromosome metabolism, centrosome function, mitotic spindle checkpoint, and many more processes [169]. It is worth noting that an estimated 30% of all genes in the human genome encode proteins involved in the regulation of DNA fidelity [170].

2.4.2 Point mutation mutator phenotype

It has been shown that cancer cells contain a high number of point mutations involving tumor suppressor genes like p53 [171, 172]. Cancer cells also contain numerous random mutations. These mutations are not selected during tumor development since they do not confer any particular growth advantage, but they do lead to genomic differences between individual tumor cells. The frequency of point mutations in tumors was compared with that observed in normal tissues. Compared with the average mutation frequency of normal tissues ($\leq 1 \times 10^{-8} \text{ bp}^{-1}$), that of the tumor tissue samples was strikingly elevated ($2.1 \times 10^{-6} \text{ bp}^{-1}$ – an increase of at 210-fold) [173].

The mutation frequency differs from the error rate. The error rates that have been discussed thus far refer to rates per cell generation. The mutation frequency refers to the accumulated mutations present in a tissue *in vivo*, and it therefore reflects genomic changes during cancer development. The mutation frequency observed by Bielas *et al.* (2006) implied that each tumor cell contained more than 10^4 point mutations. Only a minority of these mutations were expected to be selected during clonal expansion of the tumor. Therefore, a huge reservoir of unnoticed point mutations was present in the tumors. It is possible that most of these have no relevance to tumorigenesis, but some of them might contribute to the development of resistance to drugs used to treat the cancer [174]. Closer monitoring of mutation rates is one way to assess the occurrence of random point mutations. There is a need for new cancer models that will allow us to elucidate the effects of these mutations on tumorigenesis. Ultimately, technology that would allow us to influence

mutation rates *in vivo* would contribute greatly to our understanding of malignant transformation.

Mutations in genes related to DNA repair can increase the point mutation rate [165]. Point mutations that alter DNA polymerase function are expected to contribute to increased mutation rates [175]. Indeed, an inactivating point mutation in the proofreading exonuclease domain of polymerase δ has been reported to cause increased mutation rates, tumorigenesis, and reduced survival in a homozygous murine model. Primary MEF derived from these mice underwent spontaneous immortalization when cultured. [158, 176]. A missense mutation in the palm domain of polymerase δ in a hepatoma cell line is known to be responsible for altered dNTP binding and lower fidelity in copying an oxidized variant of deoxyguanylate, O⁶-methyldeoxyguanylate [175]. Furthermore, mutations in polymerase δ were found in 5 of the 6 colon cancer cell lines studied by one group [177].

The fact that 3 of these 5 cell lines were MMR-deficient raises an interesting question: were the polymerase mutations the causes of this deficiency or consequences of an MMR mutator phenotype? Variants of polymerase β were reportedly expressed in 30% of the 149 tumors studied by one group [178]. Polymerase β is responsible for DNA resynthesis in the repair pathway that corrects damaged bases. A pol β variant found in a colon carcinoma [179] has been shown to induce a 2.5-fold increase in the mutation frequency when expressed in a mouse cell line [180]. Transient expression of this variant was also sufficient to cause immortalization of primary mouse cells [181]. These findings on the behavior of polymerase β strengthen the hypothesis that a mutation in a polymerase can lead to increased mutation rates and cellular transformation.

2.4.3 Microsatellite instability mutator phenotype

As noted above, MSI has been attributed to MMR deficiency. Defective MMR certainly does lead to increased instability within repeat sequences, but it increases the frequency of base substitutions and frameshift mutations in other sequence contexts as well. Monoallelic germ-line mutations involving an MMR gene have been associated with the cancer predisposition syndrome known as hereditary non-polyposis colon cancer (HNPCC). In carriers of such a mutation, inactivation of the intact allele of the mutated MMR gene results in an MSI mutator phenotype that

drives tumorigenesis [182-184]. The predisposition to cancer is inherited autosomally, and 80% of carriers develop cancer, which chiefly involves the colon, although cancers of the endometrium, stomach, ovaries, urinary tract and kidneys, biliary tract, pancreas, small intestine, brain, and skin are also found with increased incidence in HPNCC families [185, 186]. In a study conducted in the late 1990s, the majority of HNPCC mutations identified involved *MLH1* (50%), *MLH2* (40%), or *MSH6* (10%) [187]. Mutations in *MSH6* were associated with a higher median age at cancer diagnosis and slower cancer progression than mutations in *MLH1* or *MLH2* [188]. This was no surprise, since some of the functions of *MSH6* can be fulfilled by *MSH3*. HPNCC was shown to account for 5% to 8% of colorectal cancers [189]. On the whole, 15% of colorectal cancers reportedly display an MSI mutator phenotype [190]. In a study by Truninger *et al.*, loss of expression of *MSH2*, *MSH6*, *MLH1*, or *PMS2* was found in 1.4%, 0.5%, 9.8% and 1.5% of colorectal cancers, respectively. Interestingly, germ-line mutations of *PMS2* did not lead to autosomally inherited cancer. One explanation for this finding proposed by the authors was that pseudogenes of *PMS2* on the same chromosome would serve as a pool for recombination [191]. An MSI mutator phenotype is expected to be associated with increases in the number of mutations involving genes containing microsatellites. Indeed, a number of genes with microsatellites have displayed high-frequency mutation in tumors with MSI (Table 4, [169, 192]).

Table 4: Target genes for frameshift mutations in colon cancers with MSI

Gene	Microsatellite tract	Mutation frequency	Normal function
<i>TGFBR2</i>	A ₁₀	90%	Transforming growth factor (TGF) - β signaling
<i>ACVR2</i>	A ₉ and A ₈	86%	Activin signaling
<i>IGFIIR</i>	G ₈	10%	Insulin-like growth factor and TGF- β signaling
<i>BAX</i>	G ₈	50%	Apoptosis
<i>MSH3</i>	A ₈	50%	MMR
<i>MSH6</i>	C ₈	33%	MMR
<i>E2F-4</i>	(CAG) ₁₃	65%	Cell cycle control
<i>PTEN</i>	A ₆ and A ₆	19–34%	Growth regulation
<i>MBD4 (MED1)</i>	A ₁₀	40%	DNA repair and binding to methylated DNA
<i>TCF4</i>	A ₉	39%	Growth regulation
<i>CHK1</i>	A ₉	10%	G2 cell cycle checkpoint
<i>STK11</i>	C ₆	<2%	Signal transduction
<i>BLM</i>	A ₉	<18%	Chromosome stability, DNA repair, helicase
Caspase-5 (<i>ICÉrel-III</i>)	A ₁₀	62%	Apoptosis

<i>CDX2</i>	G6	<2%	Homeobox protein
<i>TBP</i>	(CAG) ₁₉ and (CAG) ₁₆	83%	TATA binding protein
<i>RIZ</i>	A ₈ and A ₉	26%	Interacts with retinoblastoma protein
<i>RAD50</i>	A ₉ and A ₈	31%	DNA repair
<i>SEC63</i>	A ₁₀	49%	Endoplasmic reticulum chaperone protein
<i>AIM2</i>	A ₁₀	48%	Interferon-inducible protein

(Adapted from [169])

It should be noted that heterozygous knockout of MMR genes in mice is not associated with tumor development. This observation was originally explained by the shorter life span and smaller size of the knockout animals, which made inactivation of the wildtype allele in somatic cells (*loss of heterozygosity*) and subsequent tumorigenesis unlikely. Homozygous knockout animals (*Msh2*^{-/-}, *Msh6*^{-/-}, and *Mlh1*^{-/-}) presented MSI and a high incidence of cancers, mostly lymphomas (cancers originating from blood cells) followed by carcinomas (cancers of epithelial origin) [193]. The order in which these cancers developed contrasted sharply with that observed in mice that expressed proofreading-deficient pol δ. These animals usually developed carcinomas followed by lymphomas [158]. What are the underlying mechanisms of this difference? Could it be linked to the different mutational spectra of point mutation and the MSI mutator phenotype in regard to the ratio between base-substitution and frameshift errors?

3 Aim of my studies

Mutations involving the MMR genes account for most but not all cases of HNPCC. We hypothesized that some of these cancers might be caused by alteration of "fidelity genes" involved in processes that occur during replication prior to MMR processing. Of particular interest are missense mutations involving the replicative polymerases: lesions that would not cause lethal impairment of the enzyme's catalytic activity or give rise to MSI. The main question we attempted to answer in this study was this: Can loss of proofreading, or loss of polymerase selectivity, or both lead to "saturation" of the MMR machinery, thereby generating a mutator phenotype capable of causing cancer? At the biotechnological level, we were also interested in developing a system in which a mutator phenotype could be reversibly induced to investigate the effects of hypermutation on cell-cycle progression and genomic stability. The major technical aim of this project was to devise a system in which an essential enzyme can be replaced with a mutant version of the same protein.

4 Results

4.1 Strategy rationale

To evaluate the hypothesis that loss of replication fidelity upstream of MMR could lead to MMR saturation, we created variants of human replicative DNA polymerase δ (a member of the B family of DNA polymerases) that altered the fidelity of DNA replication. Two different mutations were introduced into the human *POLD1* gene, which encodes p125, the large subunit of human DNA polymerase δ . The sites chosen for these mutations were located within the active polymerase (Region II) and active exonuclease (Exo II) sites of the polymerase, both of which are highly conserved among all B family polymerases [42]. (Figure 7A).

Figure 7B shows the 4 genetically engineered polymerases δ used in this study. The proofreading-deficient (p.def.) variant was created by insertion of the D402A mutation into the Exo II region. This alteration is homologous to the D400A mutation used by Goldsby *et al.* to inactivate the exonucleolytic proofreading function of murine polymerase δ [158]. Based on studies of its homolog in yeast [160], the second mutation, L606G, alters the polymerase's selectivity for incoming deoxyribonucleotide triphosphates (dNTPs). It was used to create the error-prone (e.p.) variant of polymerase δ . The third variant was a double-mutant (d.m.) polymerase containing both the D402A and L606G mutations, which was both error-prone and proofreading-deficient. The addition of a 3xFLAG tag increases the size of exogenous p125 proteins, allowing them to be distinguished from endogenous p125 by Western blot.

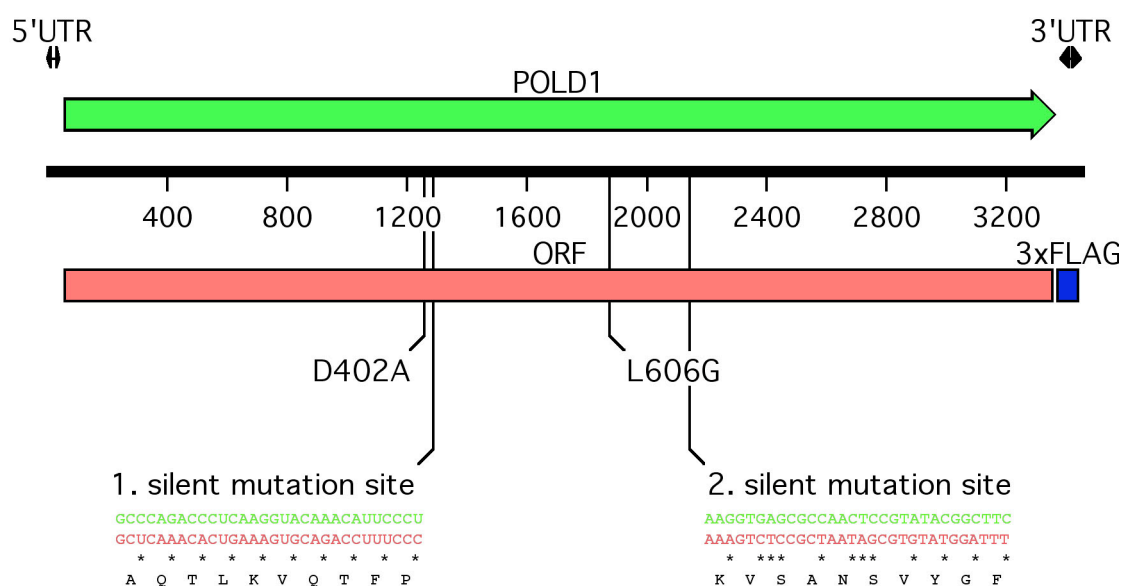


Figure 8: Schematic representation of the gene replacement procedure.

Knock-down of endogenous *POLD1* mRNA (depicted in green) is achieved by RNAi. The exogenous replacement *POLD1* mRNA (shown in red) includes D402A and/or L606G mutations, which alter the polymerase's fidelity; a 3xFLAG tag, which increases the size of the exogenous p125 protein so it can be distinguished from the endogenous form by Western blot; and silent mutations at 2 sites in the ORF (black stars), which prevent binding by small interfering RNA (siRNA) or short hairpin RNA (shRNA) targeting the endogenous mRNA.

Our objectives were: 1) to determine whether an essential mRNA can be substituted with a modified cDNA using a gene replacement approach; 2) to characterize the phenotype of cells stably expressing the engineered polymerase δ variants described above; 3) to determine whether the MMR system becomes saturated in the presence of a high mutation load; and 4) to find out if cells expressing our low-fidelity variant polymerases undergo malignant transformation.

4.2 Knock-down of p125 expression with small interfering RNA (siRNA)

Efficient knock-down of endogenous p125 expression was crucial for our gene replacement approach. We selected 5 target sites in the *POLD1* gene for RNAi-based knock-down: Two in the 5' UTR starting at positions 21 and 33 after the transcription start site (5'UTR-21 and 5'UTR-33) and 3 inside the *POLD1* ORF

starting at positions 1253, 1289, and 2149 after the transcription start site (ORF-1253, ORF-1289 and ORF-2149).

As shown in Figure 9, transfection of human HeLa cells with siRNAs targeting each of these sites markedly diminished the expression of p125.

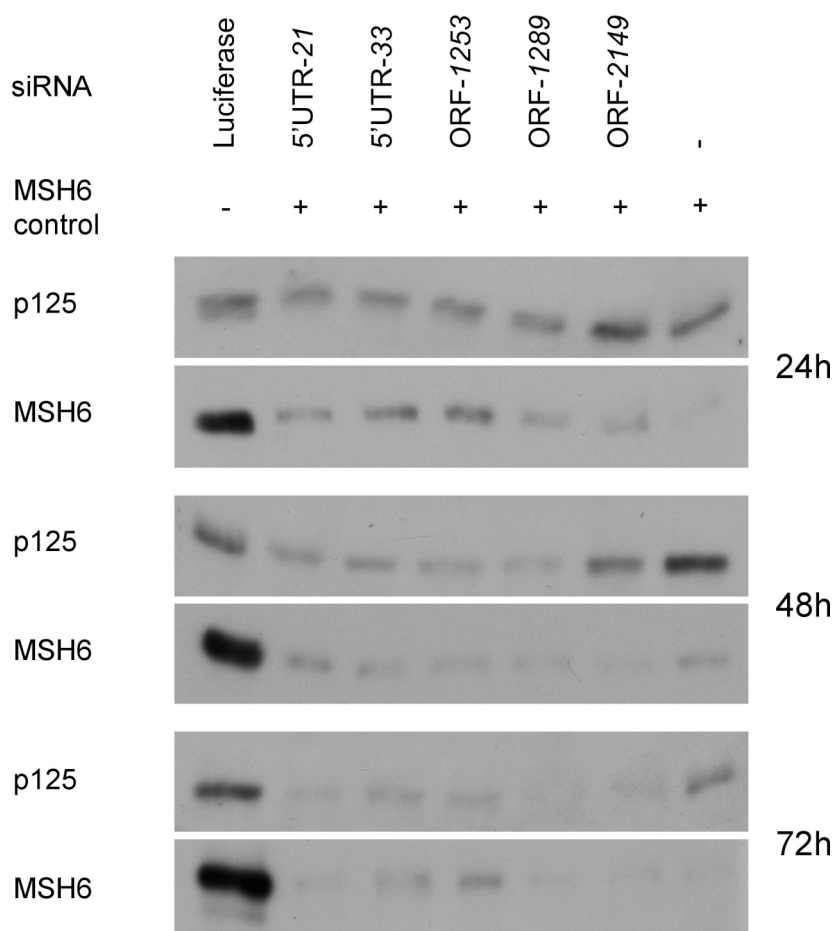


Figure 9: Western blots showing effective siRNA knock-down of p125 expression.

HeLa cells were transfected with siRNAs targeting different sites in the *POLD1* mRNA, luciferase (negative-transfection control), or *MSH6* (internal control for transfection efficiency).

It was important to identify several possible RNAi targets in the *POLD1* mRNA for 2 reasons. First, the results obtained with siRNA were no guarantee that all 5 of these sites would be effective targets for short hairpin RNA (shRNA)-mediated knock-down of endogenous p125 expression. Second, the possibility to reproduce knock-down via RNAi targeting a second or third sequence in the *POLD1* gene

provides a valuable means for confirming the specificity of the gene knock-down (i.e., excluding the possibility of off-target effects).

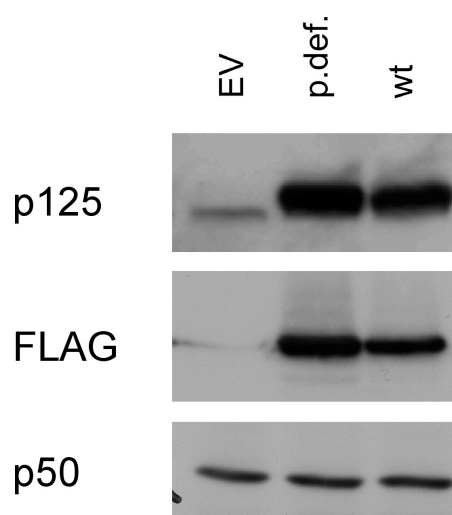
At this point, we had identified 5 sites in endogenous *POLD1* mRNA as potential RNAi targets. Since the 5' UTR is not included in our replacement *POLD1* mRNAs, they would be naturally immune to the effects of RNAi targeting either the 5' UTR21 or 5' UTR33 sequences. In contrast, RNAi targeting any of the 3 targets within the ORF might produce knock-down of both endogenous and exogenous p125 expression. To render the replacement mRNAs refractory to RNAi directed against the *ORF-1289* or *ORF-2149* sequences, we inserted silent mutations at both these sites. The D402A point mutation (Figure 8) in the p.def and d.m. *POLD1* variants was located at position 9 in the *ORF-1253* target. Schwarz *et al.* showed that a mutation at a corresponding position in mutant mRNA of human Cu, Zn superoxide dismutase was sufficient to prevent its reaction with siRNA targeting this site in the wild-type mRNA [194]. Therefore, replacement mRNAs for the p.def and d.m. *POLD1* variants (but not those for the wt or e.p. variants) should already be resistant to knock-down by RNAi targeting this site.

4.3 Non-inducible p125 expression

We initially tested 2 of the *POLD1* expression vectors created for use in our gene replacement experiments, one expressing wt p125 and the other the p.def. p125 variant. Human HEK293T cells were transiently transfected with each vector, and p125 expression was evaluated by Western blot. As shown in Figure 10, labeling with the anti-p125 antibody, which recognizes both the endogenous and exogenous forms of the protein, revealed marked increases (several-fold) in p125 levels in transfected cells, and these increases were largely due to expression of exogenous forms (p.def. or wt), as shown by the results of FLAG antibody labeling.

Figure 10: Transient expression of exogenous wild-type and proofreading-deficient p125.

HEK293T cells were transfected with p3xFLAG-CMV-14 (empty vector, EV); p3xFLAG-CMV-14-p125pdef (p.def.); and p3xFLAG-CMV-14-p125exo+ (wt). Cells were harvested 48h after transfection, and whole-cell extracts (50 μ g) were analyzed by Western blot. Blots were probed with anti-p125 antibody, which recognizes both the endogenous and exogenous forms of the protein; FLAG antibody, which recognizes only the exogenous forms of p125 (p.def. or wt); and with p50 antibody (loading control, which shows equal loading).



The overexpression of p125 observed in these experiments was unsuitable for our purposes: to ensure unbiased assembly of a 4-subunit polymerase δ , we needed to work with a level of exogenous p125 that resembles the endogenous level. To increase our control over the levels and also the timing of p125 expression, we set out to develop a system in which *POLD1* expression could be induced in a reversible and repeatable manner.

4.4 Inducible *POLD1* promoters

POLD1 expression is known to be upregulated during the S-phase of the cell cycle and following UV-induced DNA damage and downregulated after methyl methanesulfonate treatment [195-197]. In addition, binding sites for the transcription factors Sp1, Sp3, E2F, and p53 have been demonstrated in the *POLD1* promoter [197-199]. We constructed inducible *POLD1* promoters that were compatible with Invitrogen's T-REx system (Figure 11). This approach allowed us 1) to maintain control of *POLD1* expression throughout the cell cycle and preserve its response to DNA damage and 2) to culture cells with wild-type replication fidelity in the absence of induction.

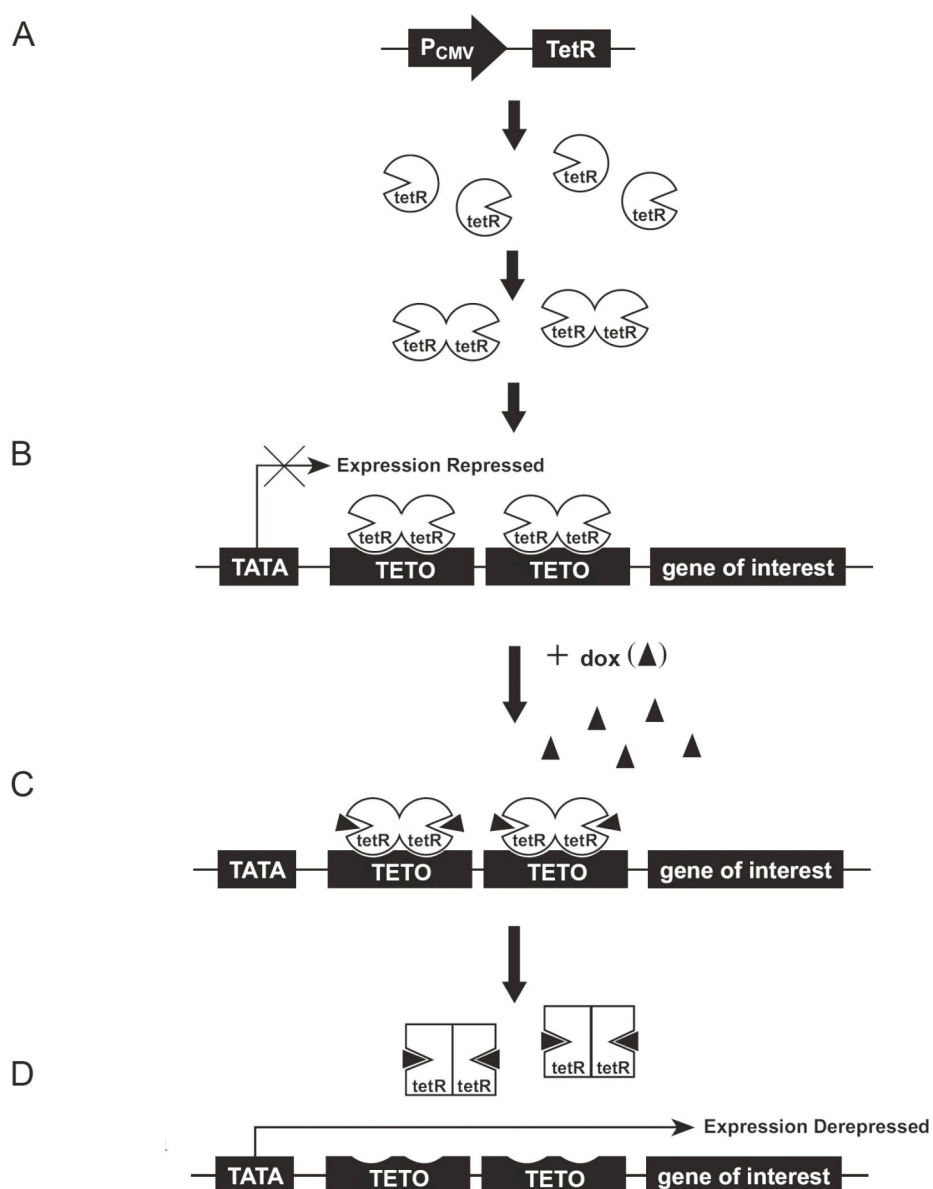


Figure 11: Components of the T-REx System for tetracycline-regulated mammalian gene expression.

In the original T-REx system, tetracycline (tet) serves as the inducing agent. In our experiments, tet was replaced with doxycycline (dox), which has a similar mechanism of action and a substantially longer half-life than tet (48 h vs. 24 h). (A) Tet repressor (tetR) protein expressed in cultured cells from a normal CMV promoter form homodimers, which (B) bind to tet operator sequences (TETO) in the inducible CMV promoter, repressing transcription of the gene of interest. C: When tetracycline — or in this case doxycycline (dox) — is added, it binds to the tetR homodimers, producing (D) a conformational change that causes them to be released from the TETO. As a result transcription of the gene of interest is derepressed. (Adapted from T-REx System user manual and [200])

We constructed 3 different *POLD1* promoters (Figure 12). The -275 promoter (Figure 12A) was cloned from the wild-type *POLD1* promoter starting 275 bp upstream from the major transcription start site. Important for the *POLD1* promoter activity are the SP1 binding site, the R1 and R2 11-bp direct repeats, as well as the E2F-like binding site adjacent to the major transcription-start site. The promoter is activated by transcription factors SP1 and SP3, which have been shown to bind to R1 and R2, and by E2F interaction with the E2F-like sequence [198]. The latter transcription factor is fully activated only during the transition from G1 to S phase. After DNA damage, interaction between p53 and the p53-binding sites on either side of R2 interferes with SP1 binding at R2, diminishing promoter activity [197, 199, 201].

Construction of the -275 *TETO-R1-TETO* promoter (Figure 12B) involved the introduction of a TETO sequence on each side of R1. The design allows transcriptional regulation analogous to that produced by p53 binding to sites next to R2. All mapped binding sites are left intact. A Kozak sequence was introduced at the end of the 5'UTR to ensure optimal translation. TetR homodimers bind to TETO and repress transcription. Addition of doxycycline (dox) leads to a conformational change of TetR homodimers and release from TETO.

A double tet operator (TETO2) sequence was introduced at the end of the -275 promoter to produce the -275 *TETO2* promoter (Figure 12C). This design leaves the original promoter sequence intact. TetR homodimers bind to TETO2 and repress transcription. TetR homodimers are released from TETO2 after addition of dox.

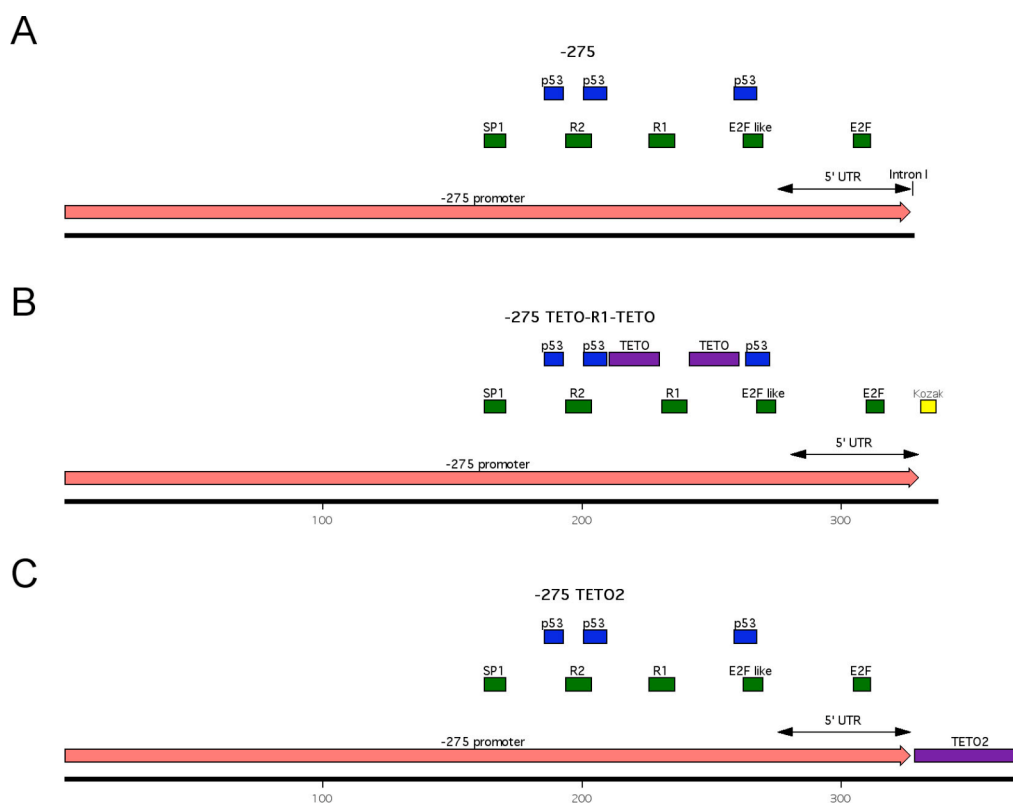


Figure 12: Inducible *POLD1* promoters

A: *The -275 promoter* – The *POLD1* promoter is shown starting 275 bp upstream from the major transcription start site. B: *The -275 TETO-R1-TETO promoter* – Insertion of a TETO sequence on each side of R1 C: *The -275 TETO2 promoter* – Introduction of a TETO2 sequence after the -275 promoter.

To evaluate their suitability for use in *POLD1* gene replacement experiments, we cloned the inducible promoters into pGL3 luciferase reporter vectors and tested their transcriptional activity in response to dox induction (Figure 13).

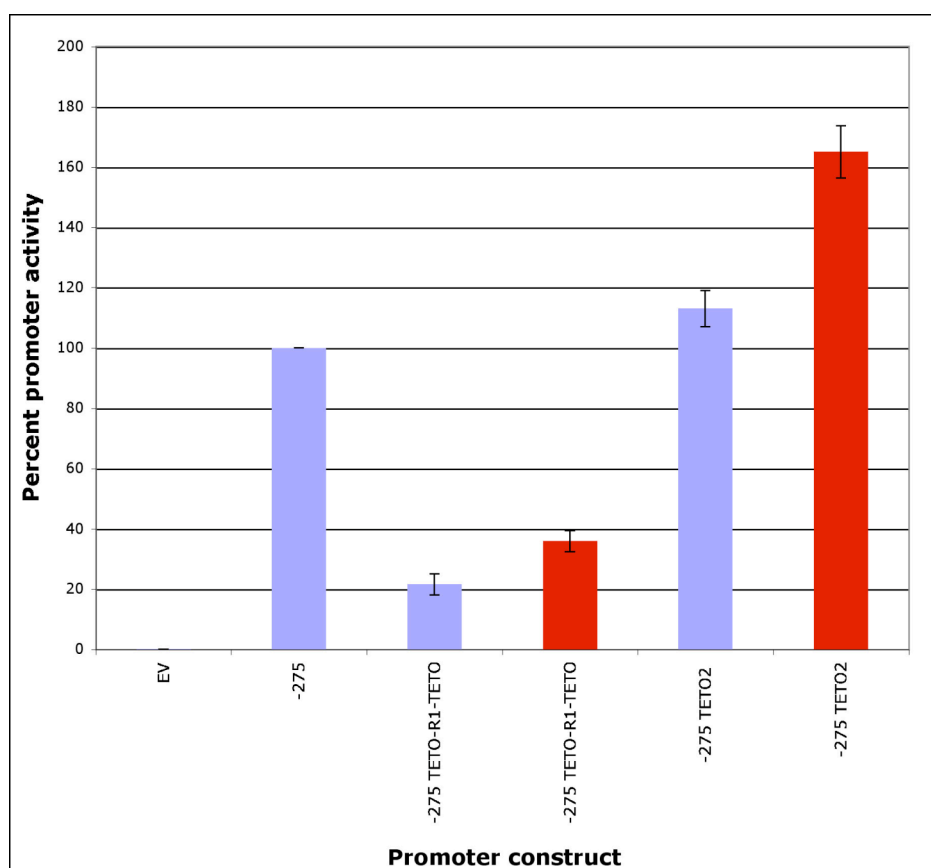


Figure 13: Activity of the inducible *POLD1* promoter constructs

The promoter activities of the -275, -275 TETO-R1-TETO and -275 TETO2 promoter in the pGL3 luciferase reporter vector were determined with the Dual-Luciferase Reporter Assay System in T-REx HeLa cells 30 h after transfection. Dox was added 6 hours after transfection when indicated. All samples were measured in triplicate. Activity is expressed as a percentage of unmodified -275 promoter activity (100%, second column from the left). Abbreviations: *EV*, empty pGL3-basic vector; -275, pGL3-delta(-275) vector; -275 *TETO-R1-TETO*, pGL3-delta(-275)TETO-R1-TETO vector; and -275 *TETO2*, pGL3-delta(-275)TETO2 vector. Red and blue columns show dox-induced and non-induced activity levels, respectively. Vertical bars represent standard deviations.

The inducible -275 TETO-R1-TETO and -275 TETO2 promoters were clearly more active after dox induction, but their activities were not completely repressed in the absence of dox. Therefore, these promoters were unsuitable for inducible *POLD1* gene replacement experiments. Since full repression of expression is indispensable for the development of timed mutational experiments, we decided to use the inducible CMV promoter of the T-REx system instead.

4.5 Inducible p125 expression

For these experiments, we chose human T-REx U2OS cells as the host cell line, because they are known to be well suited for microscopy and proficient in DNA damage signalling [202]. We constructed an inducible expression vector with a suitable selection marker for generation of stable T-REx U2OS clones and cloned our *POLD1* mutants into this vector. T-REx U2OS cells were transiently transfected with this construct, and p125 expression was assessed by Western blot (Figure 14).

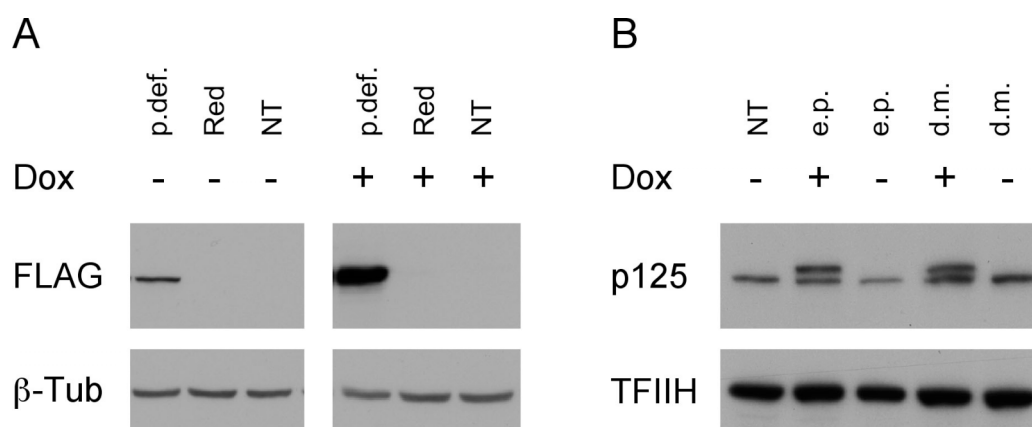


Figure 14: Inducible transient P125 expression

T-REx U2OS cells were transfected with pDsRed2CMVTET (Red), which expresses a red fluorescent protein (RFP) upon dox induction, or the same vector containing *POLD1* variants: pCMVTETp125pdefFLAG (p.def.), pCMVTETp125L606Gexo+FLAG (e.p.) and pCMVTETp125L606GpdefFLAG (d.m.). NT, not transfected. Cells were harvested 48h post-transfection. A: Exogenous p125 in whole-cell extracts (50 µg) was visualized by Western blot with the FLAG antibody. β-Tubulin (β-Tub) served as loading control. B: Exogenous and endogenous POLD1 were distinguished with the p125 antibody. The separation time was increased, and the protein load was decreased to 15 µg. Transcription factor II H (TFIIH) was used as loading control. The exogenous protein had lower mobility than its endogenous counterpart because the 3xFLAG tag increased its molecular weight.

Application of dox induced expression of the exogenous p125 variants at levels similar to those of the endogenous protein. The FLAG antibody displayed higher sensitivity than the p125 antibody in detecting the exogenous p125. FLAG-tagged p125 was detected even in the absence of induction, although at markedly reduced levels. The results of this transient expression experiment indicated that the inducible CMV promoter is well suited for p125 expression. Therefore, we proceeded

to generate stable T-REx U2OS clones expressing each of the 4 *POLD1* variants (referred to hereafter as the wt, p.def, e.p., and d.m. clones). The temporal characteristics and dose-dependence of dox-induced p125 expression were then assessed in each clone (Figure 15).

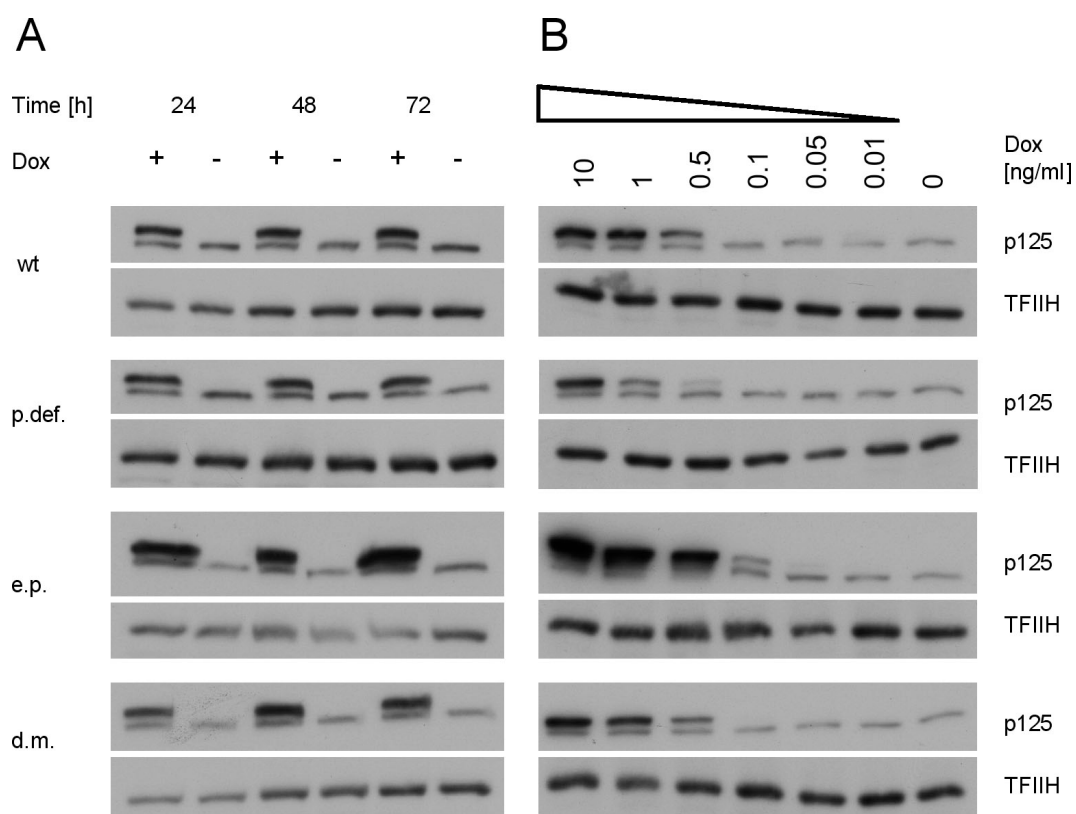


Figure 15: Effects of dox dose and duration of exposure on dox-inducible p125 expression from *POLD1* variants.

Western blots of protein extracts from stable clones of T-REx U2OS cells transfected with the plasmids pCMVTETp125exo+FLAG (wild-type [wt] clone), pCMVTETp125pdefFLAG (proofreading deficient [p.def] clone), pCMVTETp125L606Gexo+FLAG (error-prone [e.p.] clone), and pCMVTETp125L606GpdefFLAG (double-mutant [d.m.] clone). Cells were grown with or without dox at various concentrations and harvested at different times. A: Expression of p125 after 24, 48, and 72 h of dox induction at a concentration of 10 ng/ml. B: Expression of p125 after 48 h of dox induction at 6 different concentrations. TFIIH served as loading control.

The expression of exogenous p125 in the stable clones was tightly repressed in the absence of dox, and peak expression was observed even in cells exposed to only 24 h of dox induction (Figure 15A). Levels of p125 expression in the 4 cell lines differed, but identical levels of expression could be achieved by adjusting the dox

doses as follows: 0.5 ng/ml for the wt clone; 1.0 ng/ml for the p.def. clone; 0.1 ng/ml for the e.p. clone, and 0.5 ng/ml for the d.m cells (Figure 15B).

However, the half-life of dox is 48 hours. Consequently, changes in the drug concentration due to degradation were reflected in changing expression levels. To eliminate this problem, subsequent experiments were performed with a dox concentration of 10ng/ml for all cell lines. This experimental design was more robust. Each of the 4 clones was grown for 2 weeks in the presence or absence of dox. No differences in morphology or confluency were observed.

We expected to find the most severe mutator phenotype in the d.m. clone. To determine whether repair of the additional errors generated by the d.m. p125 might slow replication, creating a growth disadvantage, we compared the growth rates of the wt and d.m. clones (Figure 16).

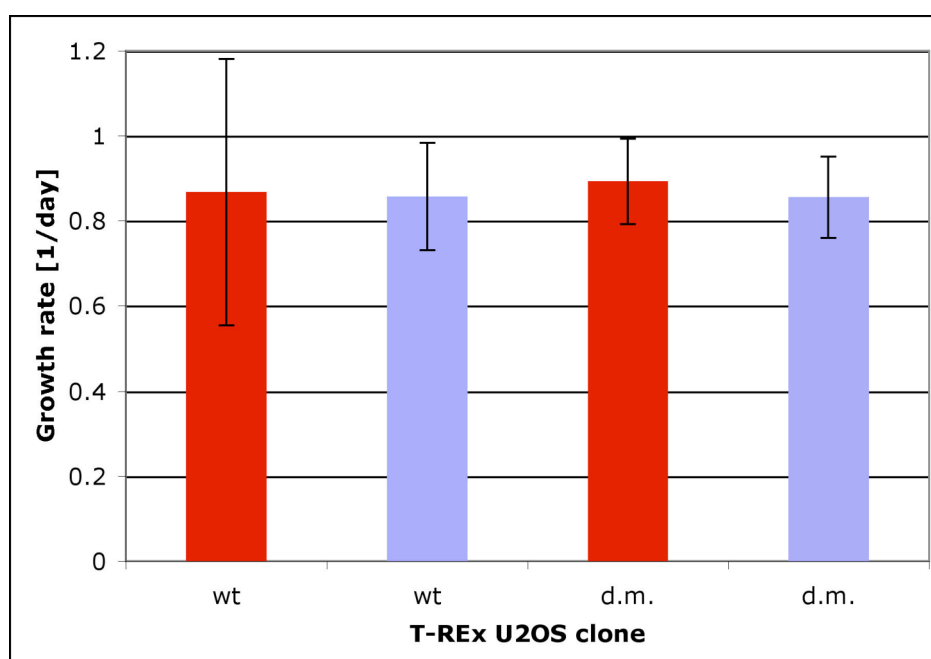


Figure 16: Growth rates of the wt and d.m. T-REx U2OS clones.

Cells of each clone (n=200,000) were seeded into a 10-cm dish and counted 4 days later. The growth rate was calculated according to the formula: $\text{growth rate} = \log_2(\text{counted cell number}/\text{initial cell number})/\text{days in culture}$. Values shown in the graph are the means of rates calculated in 4 different dishes for each condition. Error bars represent standard deviations. Red and blue columns represent growth rates before and after dox-induction.

As shown in Figure 16, however, the d.m. and wt cells displayed similar growth rates, before and after dox induction. We speculated that the absence of a growth

disadvantage in the d.m. clone might indicate that the activity of the d.m. p125 was being masked by that of the endogenous p125 and/or impaired in some way by its 3xFLAG tag.

4.6 Polymerase activity

To determine whether the exogenous p125 was interacting with the other subunits of polymerase δ , we immunoprecipitated the 3x-FLAG-tagged protein from nuclear extracts with the FLAG antibody (Figure 17).

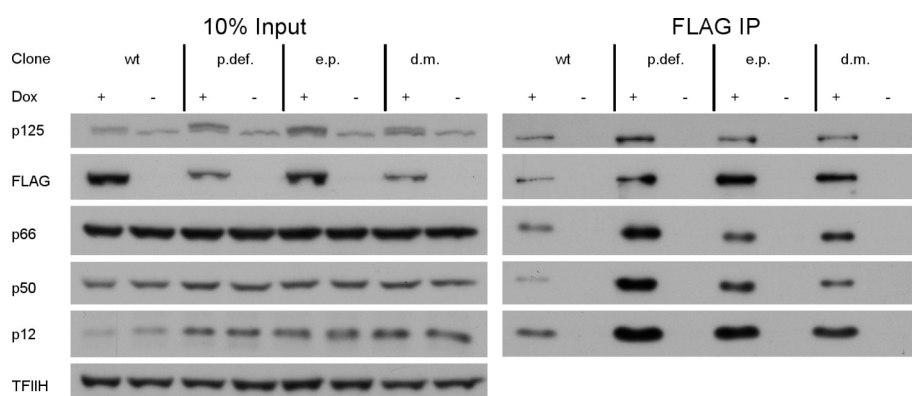


Figure 17: Immunoprecipitation of polymerase δ

Nuclear extracts were made from each of the 4 p125-expressing clones grown with and without dox. Polymerase δ was immunoprecipitated with FLAG antibody-coupled agarose beads. Polymerase δ was eluted from the beads under mild conditions by the addition of 3xFLAG peptide. Aliquots of this elution corresponding to 500 μ g input nuclear extract were subjected to Western blotting and compared with 50 μ g of nuclear extract. TFIIH was used as a loading control for nuclear extract.

As shown in Figure 17, all 4 subunits of polymerase δ were detected in the immunoprecipitate, confirming that the exogenous p125 was indeed interacting with the other subunits. Since the 4 variant polymerase expression clones expressed their respective p125 proteins at different levels (Figure 17, left panel), the levels of eluted p125 were also different (Figure 17, right panel). As for the other 3 polymerase δ subunits, p66, p50 and p12, which were expressed from wild-type genes, equal levels were observed in all 4 cell lines and regardless of dox induction (as expected). However, after immunoprecipitation and elution, the levels of the individual subunits showed equal ratios when the polymerases of the different cell lines were compared.

This observed ratio suggested that fully assembled polymerase δ was eluted in all cases.

Our next step was to develop an assay that would allow us to verify the activity of the eluted polymerase (Figure 18).

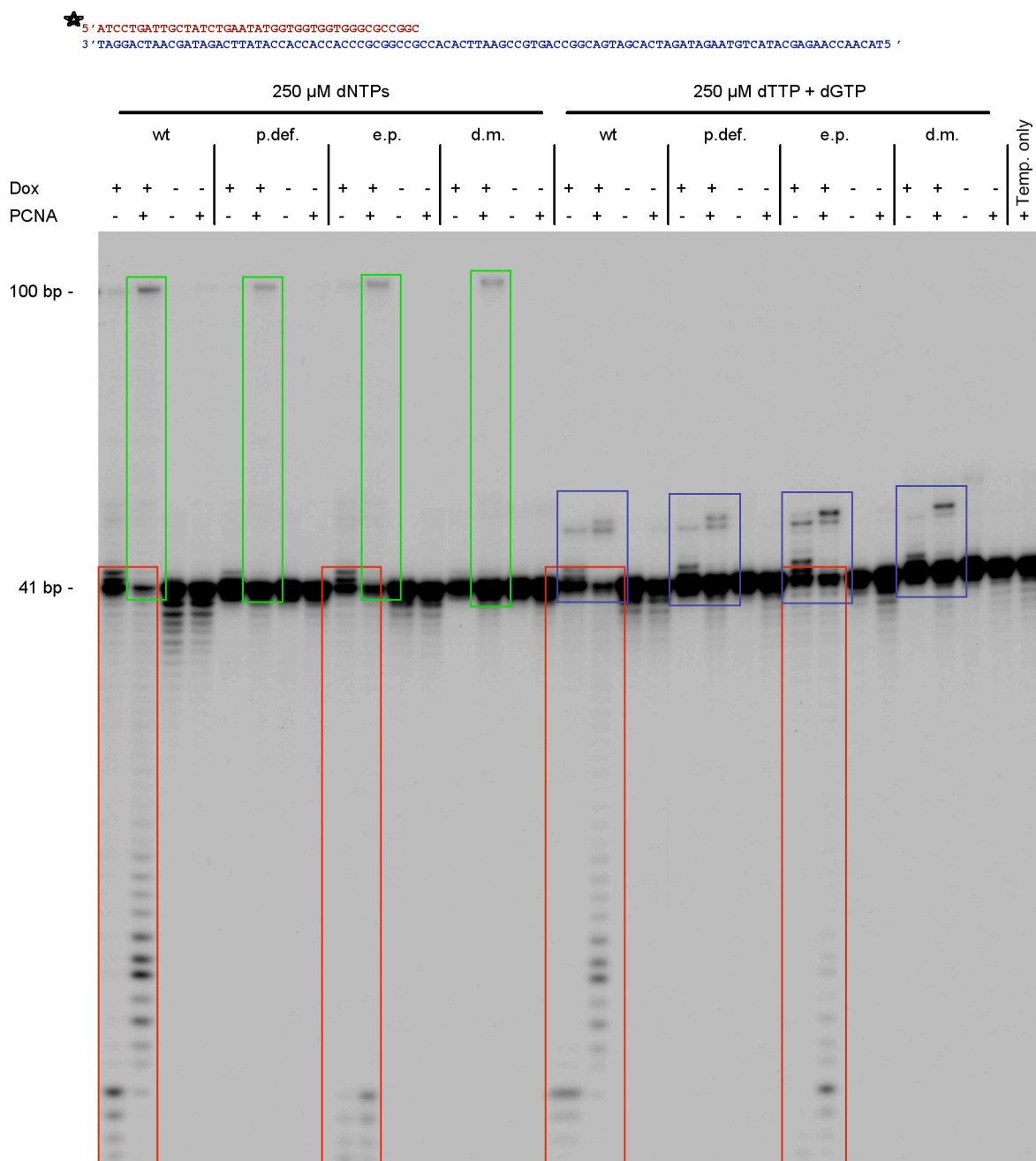


Figure 18: Results of primer extension assay I

Polymerase δ , eluted from 250 μ g of nuclear extract (Figure 17), was used in a primer extension assay with 25 fmol radioactively labeled template A. The reaction was then carried out for 10 min at 37° C at pH 7.0 in the presence of 10 mM MgCl₂ and 1 μ M PCNA, if indicated. B: The

products were precipitated and separated on a sequencing gel. The full 100-bp extension product was observed only when exogenous p125 expression was induced by dox and the reaction mixture contained PCNA (B, green rectangle). Degradation of the template took place only when the exonuclease domain was not mutated (B, red rectangle). The mutant polymerases incorporated an erroneous terminal dNTP with greater efficiency than wild type polymerase δ (blue rectangles and magnification of this area in Figure 19).

This assay demonstrated that all 4 of the polymerase δ variants had polymerase activity and confirmed that the p.def. and d.m. variants lack exonuclease activity. In the biased reaction, all nucleotides were present for the first 6 elongation steps, but the correct nucleotide for the 7th step, deoxyadenosine triphosphate, was absent. All of the polymerases δ variants incorporated an incorrect nucleotide at position 7, but the mutant variants did so with greater efficiency (Figure 19).

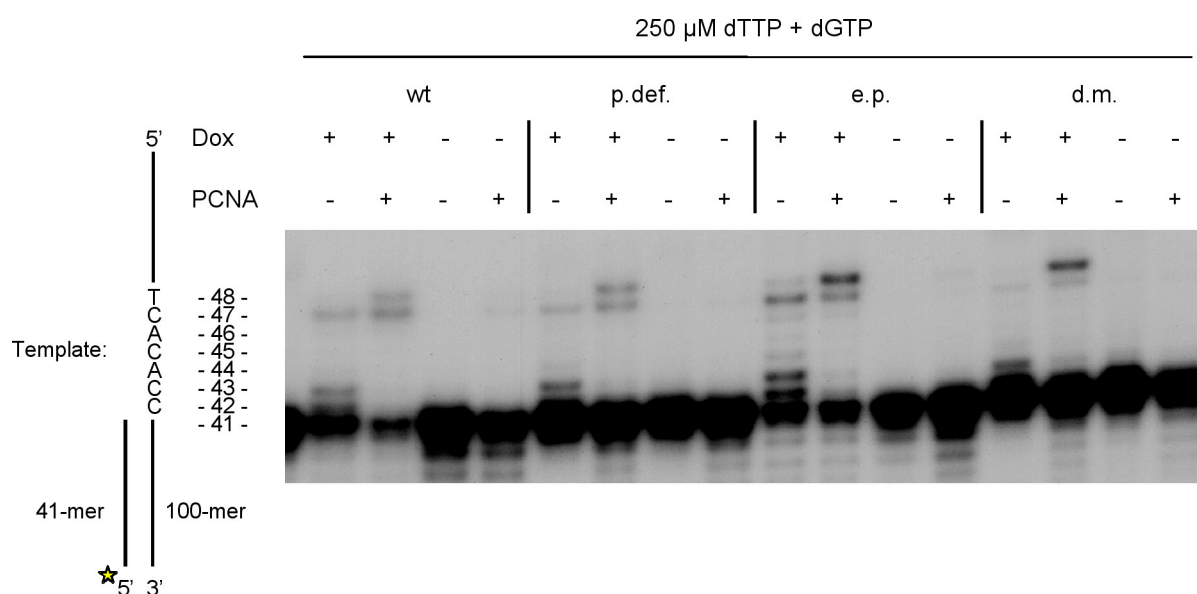


Figure 19: Detail of the results of primer extension assay I

Enlarged image of the sequencing gel shown in Figure 18. The corresponding area of the template was shown on the left side to identify the specific elongation product represented by each band. The correct dNTP for incorporation at position 7, dATP, was not present in the reaction. Compared with the wt polymerase, the mutant enzymes incorporated an incorrect nucleotide at this position more efficiently, as shown by the heavier bands opposite the template T in lanes p.def., e.p., and d.m. after dox induction.

As noted earlier, levels of eluted polymerase δ in these 4 reactions differed (Figure 17, right panel) because the variant polymerase expression clones expressed their

respective p125 proteins at different levels (Figure 17, left panel). We had developed a method for ensuring equimolar levels of polymerase expression (Figure 15B). However, the polymerase concentration differences were irrelevant for the purposes of the experiments shown here in Figure 18 and Figure 19 (to test polymerase activity and insertion of erroneous nucleotides). At this stage of assay development, we were more interested in producing a robust system. Therefore, induction was achieved with the maximal dox dose (10 ng/ml) in all clones. For these reasons, however, primer extension assay I provided qualitative rather than quantitative information on the primer extension properties of the polymerase δ variants.

The DNA precipitation step produced sharper bands in the sequencing gel, which facilitated differentiation of the individual extension products. However, it also eliminated low molecular weight radioactivity, thus preventing an accurate analysis of the degradation products of the reaction. Therefore, we repeated the polymerase activity assay without the precipitation step (Figure 20).

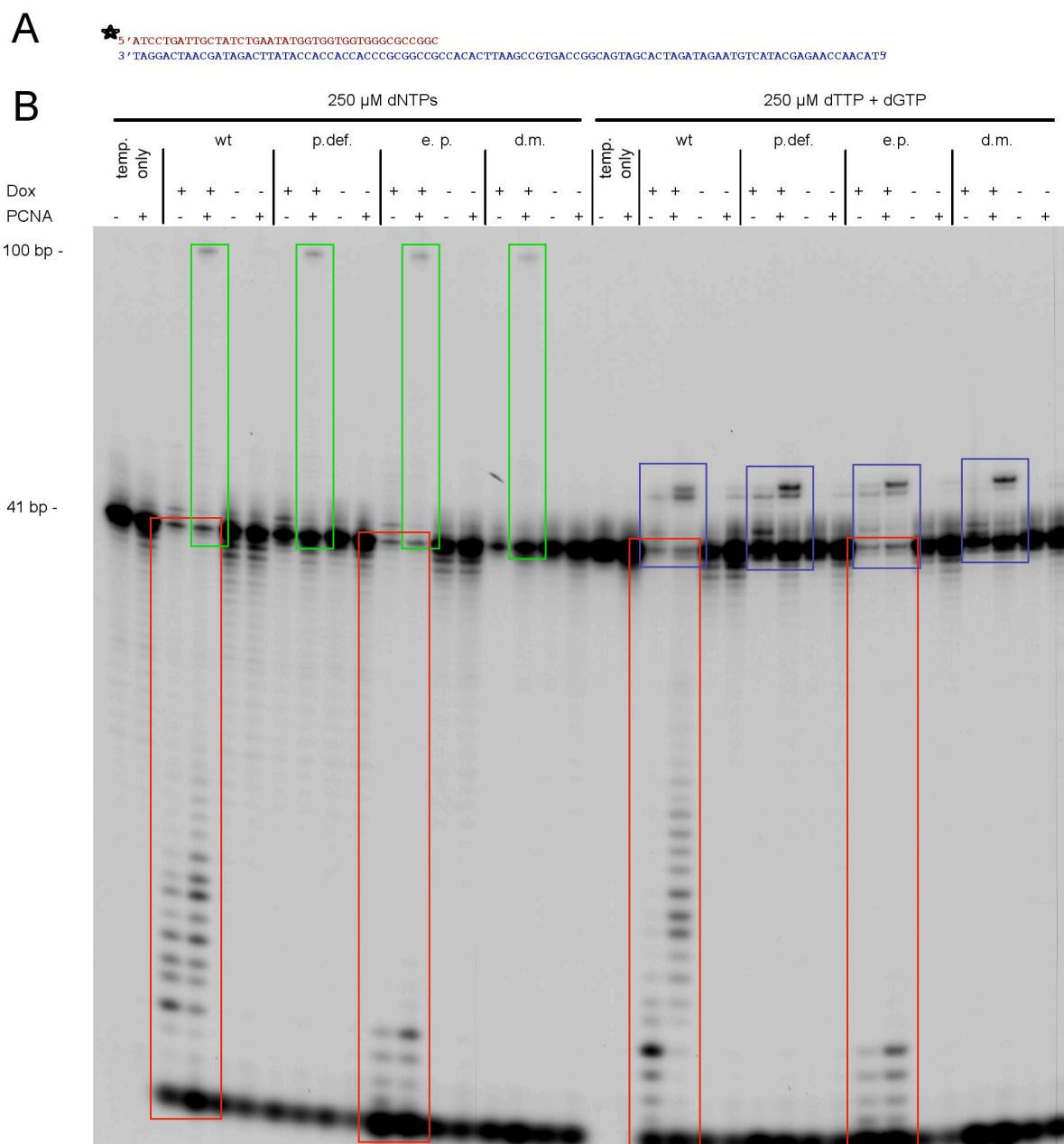


Figure 20: Results of primer extension assay II

Polymerase δ eluted from 250 μ g of nuclear extract (Figure 17) was used in a primer extension assay with 25 fmol radioactively labeled template A. The reaction was carried out for 10 min at 37° C at pH 7.0 in the presence of 10 mM $MgCl_2$ and, if indicated, 1 μ M PCNA. B: The products were separated on a sequencing gel without DNA precipitation. The full 100-bp extension product was observed only when exogenous p125 expression was induced by dox and the reaction mixture contained PCNA (B, green rectangle). Template degradation occurred only when the exonuclease domain was not mutated (B, red rectangle). The mutant polymerases incorporated an erroneous terminal dNTP with greater efficiency than wild-type polymerase δ (B, blue rectangles and magnification of this area in Figure 21).

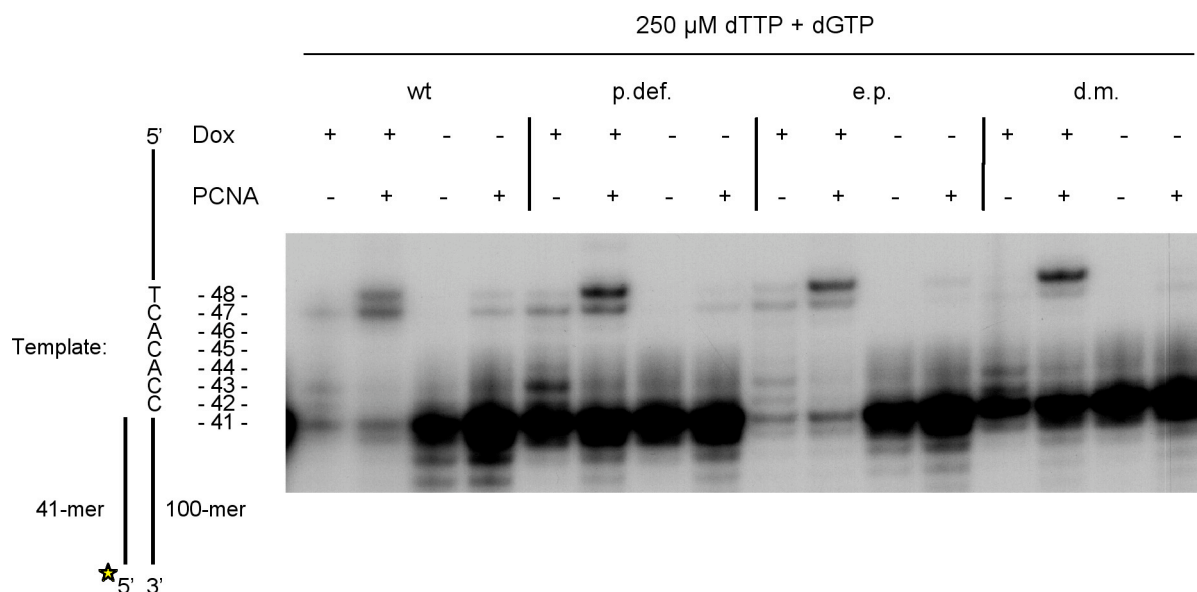


Figure 21: Detail of the results of primer extension assay II

Enlarged image of the sequencing gel shown in Figure 20. The corresponding area of the template was shown on the left side to identify the specific elongation product represented by each band. The correct dNTP for incorporation at position 7, dATP, was not present in the reaction. Compared with the wt polymerase, the mutant enzymes incorporated an incorrect nucleotide at this position more efficiently, as shown by the heavier bands opposite the template T in lanes p.def., e.p., and d.m. after dox induction.

The second primer extension assay confirmed the results of the first assay (Figure 20 and Figure 21). Since the precipitation step had been omitted, radioactive reaction products of all sizes were loaded. Compared with the wt polymerase δ , the e.p. variant was characterized by more advanced template degradation (Figure 20, red rectangles), which may have been related to the higher concentration of e.p. polymerase in the assay reaction, as discussed above (see also Figure 17, right panel). In contrast, no template degradation was seen with the p.def and d.m. variants, both of which harbor an inactivating mutation in the exonuclease domain.

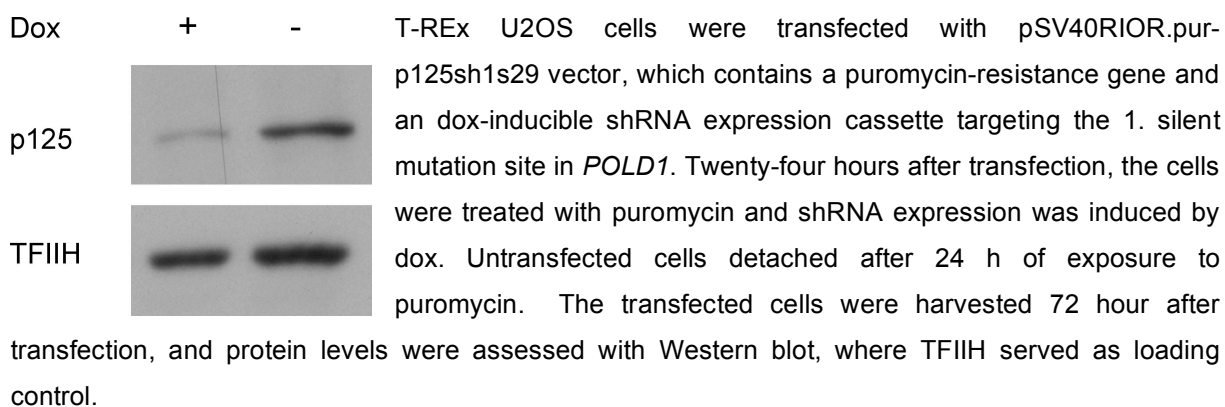
4.7 Small-hairpin RNA (shRNA) knock-down of p125 expression

In order to achieve a gene replacement in the p125 expression clones, we needed to construct an shRNA vector. Siolas *et al.* have shown that 29-mer shRNAs are more efficient than 19-mer shRNAs, and their microarray profiling studies indicated that 29-mer shRNA transfected cells are also more similar to the corresponding siRNA-transfected cells [203]. We therefore constructed a dox-inducible 29-mer shRNA

expression vector targeting the ORF-1289 site in the endogenous *POLD1* mRNA. As noted in section 4.2, siRNA targeting this sequence had been shown to produce effective knock-down of p125 expression, and a silent mutation had been introduced at this site in our replacement *POLD1* to render it refractory to knock-down. Consequently, this sequence is referred to hereafter as *1. silent mutation site*.

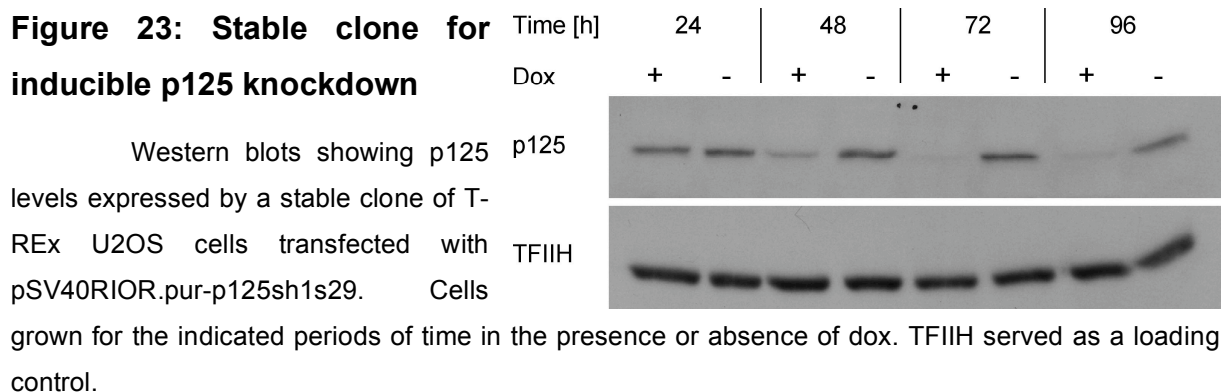
We tested the shRNA expression vector (pSV40RIOR.pur-p125sh1s29) in a transient transfection experiment to ensure that it produced efficient knock-down of p125 expression (Figure 22).

Figure 22: Transient knock-down of p125 expression with shRNA



The levels of p125 expressed by pSV40RIOR.pur-p125sh1s29-transfected cells were clearly reduced after Dox induction (Figure 22), confirming the functionality of the shRNA targeting the 1. silent mutation site of *POLD1*. We then generated a stable clone of these transfected cells that would allow us to analyze the behavior of the knock-down over time (Figure 23).

Figure 23: Stable clone for inducible p125 knockdown



Expression of p125 decreased progressively as shRNA induction times [i.e. duration of exposure to dox] increased. A reduction of p125 levels was associated with reduced viability manifested by the detachment of a high number of cells. This result was expected, because *POLD1* is an essential gene. The Western blot data indicated that we had achieved efficient inducible knock-down of p125.

4.8 Gene replacement

The stable wt, p.def., e.p., and d.m. clones express exogenous p125 upon dox induction, but endogenous p125 levels remained constant. To abolish endogenous p125 expression and achieve inducible gene replacement, we introduced the inducible shRNA expression vector (pSV40RIOR.pur-p125sh1s29) into these clones and selected one clone from each of the transfected lines. These gene-replacement (GR) clones were named according to the nature of the *POLD1* replacement gene they carried: wild-type gene replacement (wt GR), proofreading-deficient gene replacement (p.def. GR), and error-prone gene replacement (e.p. GR). Work is underway to generate a double-mutant gene replacement clone.

The generation of stable GR clones was very time-consuming, and additional studies might need to be performed in more than one cell type. While we were waiting for clones to grow, we began work on the development of novel, all-in-one gene-replacement vectors for each *POLD1* variant. These constructs, which combine the plasmid expressing shRNA against native *POLD1* with that expressing the engineered p125 variant (wt, p.def, e.p. or d.m.), have not been fully tested yet. In theory, however, the single-vector approach should considerably shorten the time needed to generate a stable *POLD1* gene replacement clone, and it might also facilitate the of *POLD1* gene replacement by transient transfection.

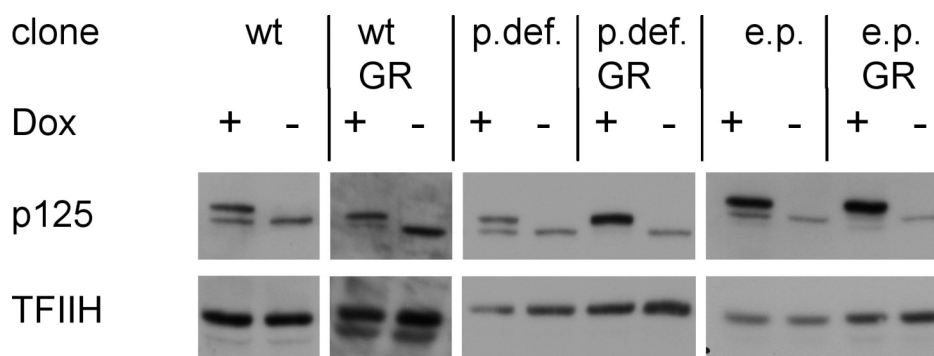


Figure 24: Inducible p125 expression in stable GR clones

Cells of each gene-replacement (GR) and parent-cell clone were seeded at equal cell densities and grown for 4 days in the presence and absence of dox. Protein levels were determined by Western blot and results for each GR clone were compared with those of the parent-cell line. TFIIH served as a loading control.

When our stable GR cell cultures had reached adequate size, we compared their expression of p125 with that of their parent-cell lines (Figure 24). In all 3 GR clones, exposure to dox resulted in simultaneous knock-down of endogenous p125 expression and induction of exogenous p125 expression. These results confirmed that we had achieved inducible gene replacement with wt, p.def. and e.p. *POLD1* variants. Interestingly, the expression levels of exogenous p125 in the p-def GR clone were increased over that observed in the parent-cell line (i.e., the p.def expression clone).

We monitored the expression levels of exogenous and endogenous p125 proteins in the 3 GR clones over the course of 8 days, during and after dox induction. (Figure 25).

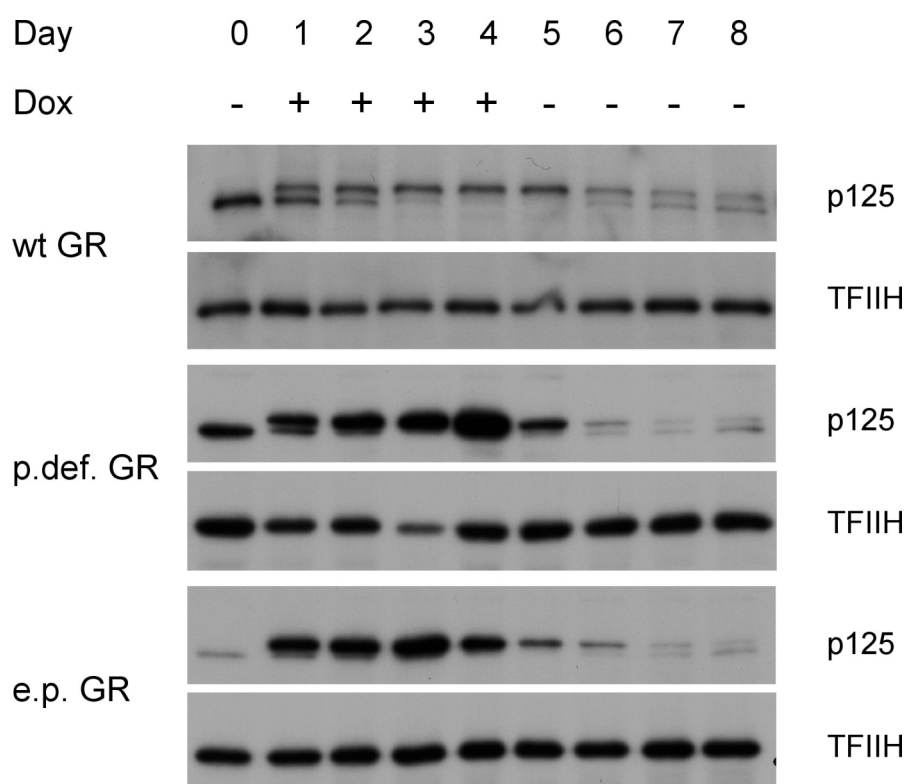


Figure 25: Time course of *POLD1* gene replacement experiments

Cells of each GR clone were grown in the presence (+) of dox (days 1-4). Dox was then removed, and growth was allowed to continue in the absence (-) of the induction agent (day 0 and days 5-8). Cells were harvested at the indicated time points. TFIIH served as loading control.

Generally speaking, in all 3 GR cell lines, endogenous p125 expression decreased progressively during days 1-3 and was no longer detectable on day 4. Expression of exogenous p125 peaked on day 2 and remained stable through day 4. The exception to this rule was the p.def GR clone, where there was a steady increase in exogenous protein expression during days 1-4. This finding was unexpected since it had not been observed in any of the parent-cell lines (Figure 15A) or in the wt GR or e.p. GR clone. The increasing expression of exogenous p125 in the p.def GR cells thus appears to be related specifically to the lack of proofreading exonuclease activity in this clone.

After removal of dox, exogenous p125 levels decreased in all 3 GR cell lines, but traces of the protein were still detectable on day 8. Dox removal also led to the reappearance of endogenous p125, but this response was not as immediate. Detectable levels were present by day 6, but expression was still far below pretreatment levels on day 8.

To determine how long it takes for these levels to normalize, we monitored the recovery process in one clone (p.def. GR) over a longer period of time (Figure 26). After approximately a week, the exogenous p125 had been completely degraded, and within another week, endogenous p125 expression reached levels similar to those found in the untreated control, thus confirming the reversibility of our gene replacement procedure. These cells were re-exposed to dox to confirm that the gene replacement was also repeatable. As shown in Figure 26, the effects observed were identical to those produced by the original induction of gene replacement in these cells.

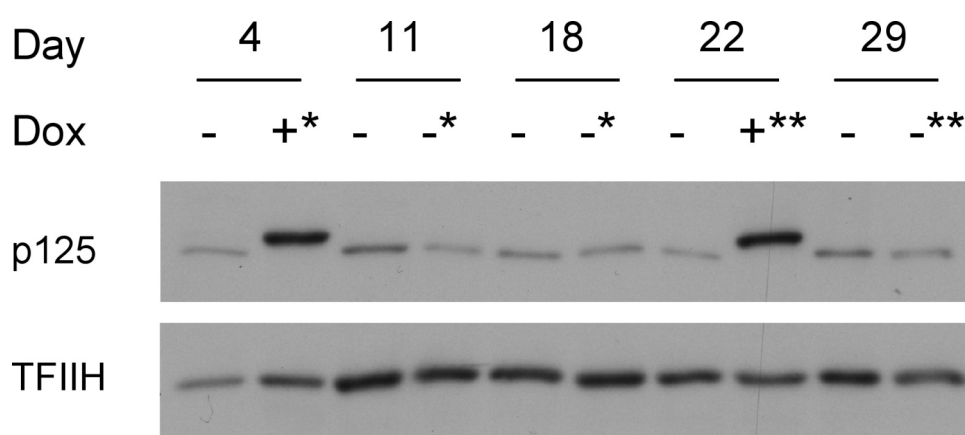


Figure 26: Reversibility and repeatability of the gene replacement procedure in the p.def GR clone

Cells of the p.def GR clone were grown in the presence of dox (days 1-4). Dox was then removed, and growth was allowed to continue in the absence of the induction agent (days 5-18). The same cells were then exposed to a second 4-day period of induction (days 19-22). Expression of p125 was assessed on the days indicated. Plus (+) and minus (-) signs indicate cells that were and were not being exposed to dox at the indicated time point. Asterisks indicate cells that had been exposed to dox induction on days 1-4 (*) and on days 19-22 (**). TFIIH served as loading control.

Microscopic examination of p.def GR cultures and e.p. GR cultures exposed to dox for 4 days revealed higher numbers of floating cells (compared with noninduced cells from the same clones). We seeded 1×10^6 cells of each clone into a 10-cm dish, let them grow with or without dox induction for 4 days, stained them with trypan blue, and counted the labeled cells with an automated cell counter (Countess from Invitrogen). The viability of wt GR cells was not reduced by dox induction. In contrast, the viability for induced cells from the p.def GR and e.p. GR clones were respectively

12% and 14% lower than those observed for their noninduced counterparts. We excluded the involvement of the selection agent puromycin in this phenomenon since the viability reductions were also observed in the absence of exposure to this agent. We concluded that the reduced viabilities of the p.def GR and e.p. GR clones were consequences of the defective proofreading and polymerase functions of polymerase δ in these cells. Further investigations will be needed to determine whether this diminished viability is a direct consequence of a mutator phenotype or related to other functions of polymerase δ .

5 Discussion

5.1 Project overview

The goal of my PhD project was to design a system for the induction of replication errors *in vivo* upstream of the MMR machinery. For this, we chose to focus on DNA polymerase δ , because this polymerase is responsible for the majority of cellular DNA replication. It was known that an inactivation of the proofreading activity of p125, the large subunit of DNA polymerase δ , conferred a defined mutator phenotype [158, 204]. Therefore, we have followed a stepwise approach in order to establish an inducible gene replacement of *POLD1*, which is the gene for p125. We cloned the wild type (wt) and proofreading-deficient (p.def.) *POLD1* together with a 3xFLAG at the C-terminus into a standard CMV promoter-driven expression vector. The tag was added in order to be able to distinguish the exogenous and endogenous proteins. In transient transfections, the protein was highly expressed and we therefore concluded that the tag interfered neither with its expression, nor with its stability. However, we wanted the expression of p125 to be as close as possible to the endogenous level, in order to ensure a proper assembly of the 4 subunits of polymerase δ . We also wanted to have an inducible expression system, in order to be able to fine-tune and time the expression. To this end, we modified the wild type *POLD1* promoter by the addition of tet operator sites in an attempt to combine wild type transcriptional control with a tight repression of promoter activity by tet repressor (tetR) binding in the absence of doxycycline (dox). Unfortunately, we obtained only promoter variants with a slightly attenuated expression in the uninduced state. As tight repression was a prerequisite for future mutagenesis studies, we decided to use a commercial promoter that was shown to be efficiently repressed in the absence of dox. No vector with a suitable selection marker was available at the time, so we modified our vector so as to make it compatible for use with the U2OS T-REx cell line, which already expresses tetR. Furthermore, we decided to incorporate the red fluorescent protein (RFP) into our new inducible expression vector, in order to be able to follow the induction of expression by fluorescence monitoring. Indeed, we found that RFP was expressed only in the presence of dox when the vector was transfected into U2OS-TREx cells. As the inducible system was functional, we transferred the wt and p.def. *POLD1* variants into the expression vector.

Recently, a conserved residue near the catalytic centre of the polymerase domain of polymerase δ was shown to decrease the replication fidelity of yeast [160]. Based on homology modelling with the known structure of the B-family polymerase RB69, the authors postulated that the yeast *pol3-L612G* mutation led to an enlargement of the active pocket of the polymerase, which diminished the discrimination against incorrect nucleotides. Although this mutation altered the fidelity of the polymerase, it may not have affected its proofreading ability. This finding offered us the possibility to further increase the mutator phenotype of the human polymerase by combining the 2 mutations in a single polypeptide. To this end, we incorporated the homologous mutation L606G into wt and p.def. *POLD1* in the inducible expression vector to generate the error-prone (e.p.) and double mutant (d.m.) *POLD1* variants. We first confirmed expression of the variants in transient transfections, and then went on to generate stable cell lines expressing each variant from the inducible *POLD1* expression plasmids. The resulting cell lines were named according to the nature of the expressed p125: wt, p.def., e.p., and d.m.. Since we were the first to express these polymerase δ variants in human cells, it was important to characterize the cell lines further.

The level of p125 expression was determined at different doses of dox and at different time points. In general, p125 was only expressed upon dox-induction and reached a stable maximal level already after 1 day. These experiments enabled us to induce equimolar levels of expression for the endogenous and exogenous p125 and thus avoid artefacts associated with overexpression of p125 expression. We then ensured by immunoprecipitation that the exogenous p125 could properly interact with the other 3 subunits of polymerase δ . In order to examine the polymerase activity of the polymerase variants, we developed a novel *in vitro* polymerase activity assay. Instead of resorting to DNA polymerases produced in heterologous systems such as in baculovirus-transfected insect cells, we immunoprecipitated and isolated the proteins bound to p125 *in vivo* in a human cell line. Our method has a number of advantages over other, commonly-used *in vitro* assays. The expression in a human cell line should ensure that all posttranslational modifications occur properly. Moreover, because we immunoprecipitated p125 with the anti-FLAG antibody, only the exogenous p125 was precipitated. Elution of the antibody-bound p125 was achieved under very mild conditions by competition with the 3xFLAG peptide. This mild elution procedure ensured that protein-protein interactions were maintained,

such that the fully-assembled polymerase δ was eluted in its natural form with known and possibly also unknown binding partners. Using this approach, we were able to confirm that all 4 polymerases δ variants were active. Furthermore, we confirmed that the wt and e.p., but not the p.def. and d.m polymerase δ variants retained their exonuclease activity, as expected.

We adapted the polymerase activity assay further to test the fidelity of the enzymes. By omitting 2 dNTPs in the primer extension assay, we could test the differences between the 4 polymerase variants in the incorporation of non-complementary nucleotides. We could show that p.def., e.p. and d.m. had an increased frequency of incorporation of erroneous terminal nucleotides as compared to wt polymerase δ . We therefore concluded that our mutant polymerases do indeed have altered polymerase fidelities *in vitro*. Somewhat unexpectedly, the morphology and growth rates of the cell lines were not changed by the expression of the exogenous enzymes. We therefore suspect that the wild type endogenous polymerase masks the mutator effect of the variant polymerases *in vivo*.

In order to overcome the above effect, it was necessary to eliminate the wt endogenous protein from the living cells. We anticipated this already at the onset of the study and designed the *POLD1* expression constructs to carry silent mutations that render them refractory to RNA interference (RNAi). Because we require to downregulate the expression of the endogenous polymerase over a long time period, we constructed a dox-inducible shRNA expression vector targeting the first silent mutation site of *POLD1* and confirmed endogenous p125 knockdown in both a transient transfection assay and in a stable clone, exclusively after dox induction. As we were able to combine the knockdown of the endogenous p125 with expression of the exogenous wt p125 variant, we used the same construct to generate also stable cell lines expressing p.def. and e.p.. These gene replacement cell lines were named wt GR, p.def. GR and e.p. GR. These cell lines represent a proof of principle for our gene replacement approach, namely that it is possible to replace an essential gene with an engineered variant through a combination of dox-inducible expression and knock-down.

In the course of these experiments, we made an interesting observation: in the p.def. GR cell line, the exogenous p125 protein was seen to accumulate after 4 days of dox induction. This effect was seen neither in the corresponding parental cell line

p.def., nor in wt GR and e.p. GR, or in their parental cell lines wt and e.p.. We postulated that accumulation of this p125 variant had to be linked to the action of the proofreading exonuclease activity of polymerase δ , because this was the only difference between the cell lines, however, this hypothesis needs to be verified experimentally in future studies.

In addition to the accumulation of p125, we observed a second phenomenon in the p.def. GR and e.p. GR cell lines. Namely, we noticed a reduced viability of these cells after 4 days of induction. We first considered that expression of the shRNA might have triggered an interferon response, but this possibility was ruled out, since the clone wt GR was expressing the same shRNA and did not show reduced viability. We therefore think that the effect could be linked to the mutator phenotype of the cells, but this also requires further investigation.

To obtain inducible gene replacement by 2 rounds of stable selection and clone generation was rather time consuming. In order to simplify this process, we constructed a novel all-in-one vector by combining the 2 plasmids expressing the p125 variant and the shRNA. This will facilitate the generation of future gene replacement clones significantly. The gene replacement will be inducible when the transfected cell line already expresses TetR, or when a vector encoding for TetR, like pcDNA6/TR, is co-transfected. It should be easy to adapt this vector into a general cloning vector for the gene replacement of any gene of interest.

5.2 Future directions

5.2.1 U2OS T-REx clones

Recently, heterozygous Pold1^{+/L604G} mice were generated [205]. Although they did not have a reduced life span or a significant increase in cancer incidence, MEFs of these animals displayed 5-fold higher mutation rates at the *hprt* locus and a 17-fold increase in the number of chromosome aberrations compared to wild type cells. This work was the first description of mammalian cells expressing a variant of a replicative polymerase. As described above, we have created a similar situation in a human cell line, in a clone that expresses a similar ratio of p125 containing the analogous mutation (L606G) and the wt enzyme (clone e.p.). As in the heterozygous Pold1^{+/L604G} mice, the cells appeared normal, at least as far as their growth characteristics and morphology are concerned. Whether they have a mutator

phenotype or display increased chromosomal instability (CIN) remains to be determined.

The relatively mild pathology of the heterozygous animals contrasted with the embryonal-lethal phenotype of homozygous *Pold1*^{L604G/L604G} mice [205]. A detailed study of the phenotypic consequences of the double knock-in in the murine system was therefore impossible. In our system, a cell line (e.p. GR) lacking endogenous polymerase δ could be generated and future studies should show whether its phenotype is restricted solely to replication-associated mutations, or whether the cells display also CIN as observed for heterozygous *Pold1*^{+L604G} mice. Should this be the case, we would like to elucidate the underlying mechanisms of this instability.

Another interesting question concerns the phenotype of mammalian cells expressing the doubly-mutant polymerase δ . To date, the only available data comes from yeast, where a haploid strain expressing polymerase δ carrying homologous mutations was not viable [160]. It would have been interesting to know whether a d.m. diploid yeast strain is viable, since diploid strains are able to tolerate higher mutation loads than haploids. However, there was no attempt to generate such a diploid yeast strain so far. We were able to express the d.m. polymerase δ *in vivo* and even characterize it *in vitro* for the first time. The d.m. human cell line displayed no obvious changes in morphology or growth rates, but it should be noted that these cells still expressed the wt endogenous protein. We are attempting to eliminate the wt polypeptide using our gene replacement approach, and work is underway to generate a stable clone, which is expected to have the strongest mammalian mutator phenotype known. It will be interesting to see whether such a high mutation load leads to an error catastrophe or whether it can be tolerated at least for several rounds of DNA replication.

5.2.2 Mutation rates

We would like to measure the *in vivo* mutation rates of our variants and compare them to those seen in yeast and mouse cells expressing polymerase δ variants carrying homologous mutations [158, 160, 204, 205]. One commonly-used approach to determine replication fidelity in mammalian cells is to measure mutation rates at the *hprt* locus [206]. Unfortunately, this assay depends on the presence in the cell of only a single copy of the *hprt* gene, which is located on the X-

chromosome. Because the U2OS cells used in our study are aneuploid, they are unsuitable for these experiments. We therefore plan to make use of a cell line that is more appropriate for the *hprt* mutation rate assay, such as A2780MNU-clone 1. This cell line offers the possibility of not only assessing the mutation rates, but also of determining MSI. More importantly, this cell line is MMR-deficient, but can be complemented with *MLH1* in order to become MMR-proficient [151]. Thus, It would allow us to investigate the contribution of MMR to the processing of the errors generated by polymerase δ variants. The dependence of mutation rates of polymerase δ variants on MMR has never been measured in mammalian cells. There could be several outcomes. In the absence of a functional MMR, the mutation load might be too high to allow cellular survival. The combination of p.def. polymerase δ with MMR deficiency led to an error catastrophe in haploid yeast, but not in a diploid strain [157]. We therefore expect that the p.def. *POLD1* gene replacement in an MMR-deficient background will be viable. It will be interesting to see whether the other 2 *POLD1* gene replacement mutants also give rise to an error catastrophe in a MMR-deficient background. Furthermore, a comparison of mutation rates in MMR-proficient and -deficient backgrounds will show whether a high mutation load can saturate MMR. Data from yeast suggest that MMR capacity is not saturated in a strain expressing p.def. polymerase δ [157], but this situation may be different in human cells. One way of determining MMR efficiency is by studying MSI. Cells deficient in MMR, or cells in which the capacity of this repair pathway is saturated, should display MSI, which could be measured for example at the *BAT26* locus.

5.2.3 Tumorigenesis

Tumorigenesis is still poorly understood and very far from preventable. The exact mechanisms and pathways by which a mutator phenotype is involved in cancer development still need further elucidation. Our gene replacement cell lines might offer the possibility to investigate tumorigenesis in a mouse xenograft model. U2OS cells were shown to be tumorigenic in nude mice [207], albeit with a latency period of almost 100 days. Because a mutator phenotype should accelerate tumour development, our U2OS clones expressing the mutator polymerase δ variants should give rise to tumors sooner than the parental cells. We plan to test this hypothesis by subcutaneously injecting the cell lines generated in this study into nude mice.

Rodrigues *et al.* showed that addition of dox to the drinking water of mice was sufficient to induce expression genes under dox control in tumour tissue in a xenograft model [208]. Our cell lines should therefore work as an isogenic system in xenograft tumorigenesis. Another possibility to study tumorigenesis would be to transfect human primary cells with our all-in-one *POLD1* gene replacement constructs. The use of primary cells might allow us to follow all the steps of malignant transformation.

5.3 Expected mutator phenotypes

The proofreading activities of the replicative polymerases δ and ϵ and MMR work together during DNA replication in order to remove base-substitution errors, or insertions and deletions (indels) generated during DNA polymerization. Indels are believed to arise through primer or template slippage during polymerization [209] (Figure 27). Proofreading activity of indels in microsatellites is poor, because indels situated more than 4 nucleotides from the primer terminus are not substrates for the exonuclease. As terminal misalignments can convert to internal ones in repetitive sequence contexts, and as the likelihood of these rearrangements increases exponentially with the length of the microsatellite, the contribution of proofreading to replication fidelity in long microsatellites is low [130]. For this reason, MMR is the only effective guardian against indels in repetitive sequence contexts. That is why MMR deficiency becomes readily apparent as MSI. However, the proofreading activities of polymerase δ and ϵ have a contribution towards indel fidelity outside of microsatellites [209].

Applied to our model system, expression of p.def. *POLD1* should lead to an increase of indels and base substitutions, while expression of the e.p. mutant should result in an increase in base substitutions, by analogy with the yeast enzyme.

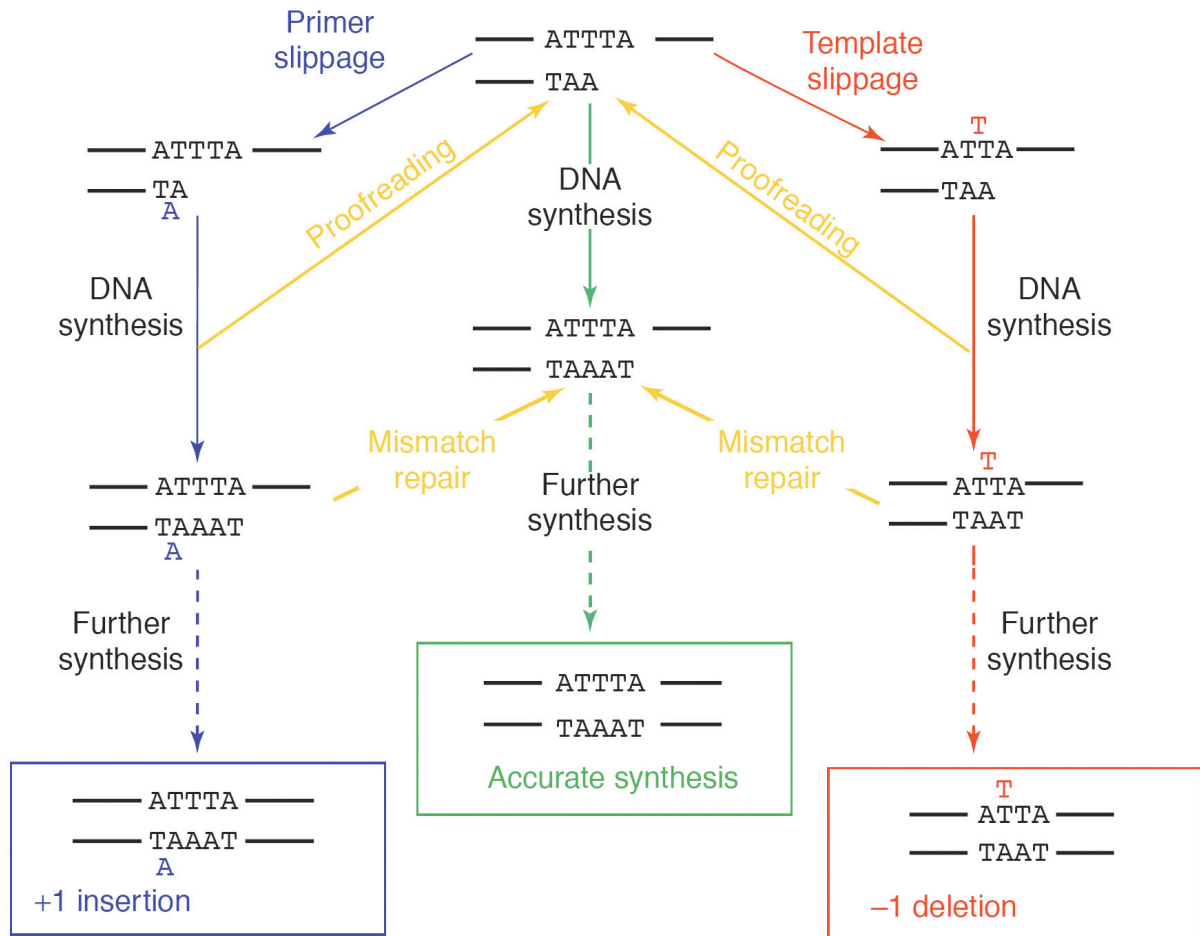


Figure 27: Determinants of indel fidelity

Indel mutations result from DNA strand slippage, which generates a misaligned intermediate with one (or more) unpaired nucleotides. These intermediates can spontaneously realign or can undergo proofreading. If further synthesis occurs on them and the unpaired nucleotide (or nucleotides) escapes postreplicative mismatch repair (yellow arrows), however, the result is an insertion or deletion depending on whether the unpaired nucleotide was located in the primer or template strand, respectively. (Adapted from [209])

Only if MMR were to become saturated, as observed in *E. coli* [155], should we observe MSI.

5.4 Possible applications

5.4.1 General gene replacement

Genetic manipulations in yeast or mice inactivate or replace the wild type gene in the genome. Although feasible, such methods are not efficient in human cell lines. However, RNAi technology allows the downregulation of any target gene and

standard expression technology allows the expression of mutant variants. We were able to show that it is possible to substitute an essential gene with a mutated variant in an inducible and reversible fashion. This approach should be applicable to any gene of interest. Long-term loss-of-function studies using RNAi are not possible for essential genes. That is why gene replacements with point mutants offer an elegant possibility to gain novel insights into the function of these key genes. Moreover, the gene replacement approach allows the “correction” of dominant-negative mutations. The general idea to achieve a gene replacement by combination of RNAi knockdown and expression of a RNAi-resistant replacement gene was not new when I started my PhD. Gene replacement was requested as the “ultimate” functional control for RNAi [210]. The idea was that as long as an observed phenotype of RNAi is specific for the target gene, the observed phenotype of RNAi should be reversed by using the same RNAi together with expression of the target gene in a form refractory to RNAi. The feasibility of the approach as a control for rescue of a knock-down phenotype was already confirmed the next year [211, 212]. Still nowadays, most RNAi experiments are carried out without this “ultimate” functional control for RNA specificity. One reason for this trend might be the lack of appropriate vector systems. A single lentiviral system that is suitable for inducible gene replacement was developed in the laboratory of Didier Trono [213]. However, not every laboratory is equipped with a biosafety level II facility, as required for the work with lentivirus. We think that our all-in-one gene replacement vector can easily be converted into a generally applicable cloning vector for gene replacement studies with mutant variants of any target gene and as a functional control of inducible shRNA expression by gene replacement of the wild type gene.

5.4.2 *POLD1* gene replacement

The specific *POLD1* gene replacement that we achieved in this work might have an industrial application. In bacteria, a similar approach has been used to improve strain properties by expression of the *MutD5* protein, the proofreading-deficient variant of the ϵ subunit of the replicative pol III holoenzyme [214]. Overexpression of this mutant was shown to confer a mutator phenotype even in the presence of the wild type protein. Following transfection with a plasmid for *MutD5* expression, bacterial strains could evolve under industrial production conditions. This

method for mutagenesis was superior to classical methods for strain improvement, which involve the use of UV light or chemical mutagens; such methods are usually discontinuous, inefficient and leading to substantial cellular damage. It is essential for industry that production strains are genetically stable. Therefore, only a temporary mutator phenotype is desirable. In the above example, loss of the expression vector restored genetic stability.

Our gene replacement constructs might be used in a similar fashion for the adaptation of mammalian cell lines to industrial applications. Transient transfection should induce a mutator phenotype that introduces novel traits to the cell line, while normal production conditions can be maintained. The plasmid should be lost after a while during replication and improved, as well as genetically-stable, clones might be selected. Recently, it was shown that proofreading-deficient polymerase δ is suitable for eukaryotic strain selection by evolving a proofreading-deficient yeast strain to higher temperature tolerance [215]. The availability of *POLD1* variants, which are expected to confer high mutation rates, together with the ability to selectively time, dose and reverse the gene replacement by adaptation of the corresponding dox treatment or modulation of the transfection procedure, should greatly contribute to fine tuning possibilities in an industrial selection process.

6 Conclusions

One goal of this thesis was to establish a novel approach to induce replication errors *in vivo*, upstream of the MMR machinery. We focused to study the effect of mutations in DNA polymerase δ , already known in yeast and mice to lead to reduced polymerase selectivity or loss of proofreading exonuclease function. Polymerase δ variants were expressed only upon dox induction and therefore a single human cell line could be used to compare the wild type state of the cell and a state characterized by a similar expression of the endogenous wild type and the exogenous variant. All our polymerase δ variants displayed polymerase activity *in vitro*, and the mutant variants incorporated non-complementary nucleotides with higher efficiency. However, the human cell lines expressing variants of p125, the large subunit of polymerase δ , in the presence of the endogenous enzyme did not display reduced growth rates or morphological changes. To gain further insights, we combined inducible shRNA expression, targeting exclusively the endogenous p125 mRNA, with inducible p125 expression. Cells in which the endogenous gene was replaced with the mutator variants displayed reduced viability, accompanied by accumulation of the exogenous proofreading-deficient but not the error-prone p125 variant. The observed reduction in viability might be a consequence of the mutator phenotype, but the accumulation of p125 seems to be a new phenomenon that merits further investigation. Finally, we have constructed a single-vector gene replacement construct, which should allow transferring the expected mutator phenotypes mediated by variants of polymerase δ into any human cell line. Because the gene replacement approach is generally applicable to any gene of interest, the presented work represents a proof of principle that opens a wide field of research possibilities and applications.

7 Materials and methods

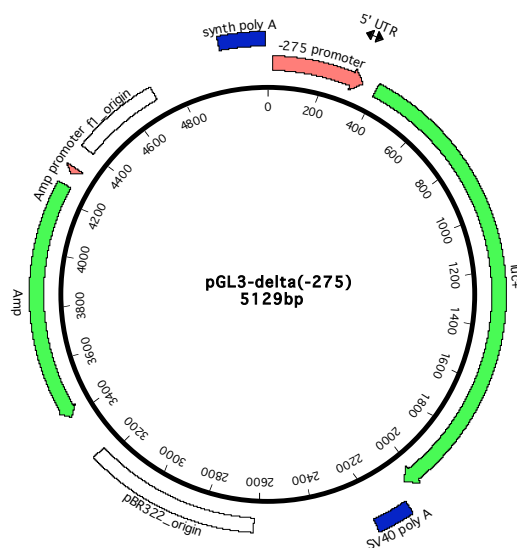
7.1 Vector construction

7.1.1 Inducible *POLD1* promoters

-275 promoter - The *POLD1* promoter sequence, beginning 275 bp upstream from the major transcription-start site and ending at the 3' end of the first exon, was PCR-amplified from 411 MI human genomic DNA (kindly provided by Dr. Giancarlo Marra). The reaction mixture contained 1x Thermopol buffer, 300 ng of template DNA, 1 μ M forward primer, 1 μ M reverse primer, 0.2 mM deoxynucleoside triphosphates (dNTPs), 8% dimethyl sulfoxide (DMSO), and *Taq* DNA polymerase (5 units/50 μ l) (New England Biolabs, Beverly, MA, USA). The primers (Microsynth, Balgach, Switzerland) were as follows: forward 275 Promoter: GGT GGT GAG CTC ATT AAT AGG GTG GGA GGA GAG AGA ACA GAA CCG CGG CGC; reverse 275 Promoter: GGT GGT AGA TCT CCC GCT TCA AAC AGC GTT TCC CGC CAC AGC CTA CG. The PCR product was purified by gel extraction, digested with *SacI* and *BglII* (All restriction enzymes were from New England Biolabs), and cloned into the corresponding restriction sites of a pGL3-Basic Luciferase Reporter vector (Promega, Madison, WI, USA). The result was the **pGL3-delta(-275)** vector shown in Figure 28.

Figure 28: The pGL3-delta(-275) vector

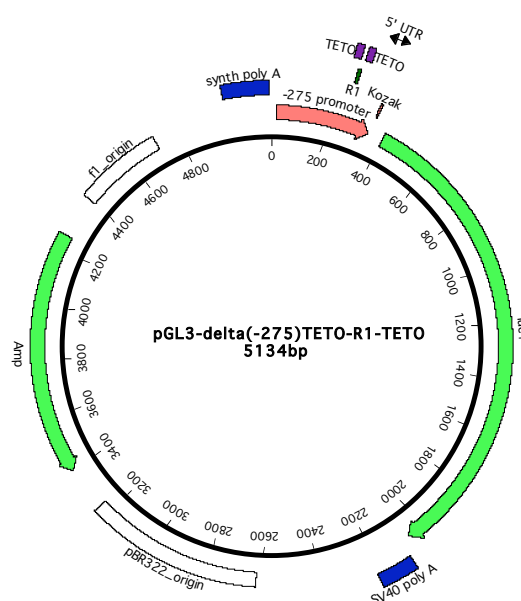
The *POLD1* promoter sequence starting 275 bp upstream from the major transcription-start site and ending at the 3' end of the first exon was cloned into the pGL3-Basic Luciferase Reporter Vector. The promoter (referred to hereafter as the -275 promoter) is located in front of a firefly luciferase (*luc+*) gene that has been optimized for monitoring transcriptional activity in transfected eukaryotic cells. The vector allows determination of promoter activity using the Dual Luciferase Reporter Assay System.



-275 TETO-R1-TETO promoter. A TETO sequence was inserted on either side of the R1 site in the -275 promoter: A megaprimer was generated by PCR with Supertet forward (CAA GCG GGG CGT GGC CTT GCC CTC CCT ATC AGT GAT AGA GAT GGG GCG TGG CCT CCC TAT CAG TGA TAG AGA TCT GGG CTT GCG CGC GCG GGA GTC); SuperProm reverse (CGG AAT GCC AAG CTT ACT TAG ATC GCA GCC ATG GTG CTT CAA ACA GCG TTT CCC GCC AC) (both from Metabion, Planegg, Germany); and pGL3-delta(-275). The reaction mixture contained 1x *Pfu* reaction buffer, 200 ng template DNA, 1 μ M forward primer, 1 μ M reverse primer, 0.25 mM dNTPs, 8% DMSO, 10% glycerol, and *Pfu* DNA polymerase (3 units/50 μ l) (Promega). The PCR program was as follows: 97°C for 2 min, (97°C for 1 min, 60°C for 1 min, 72°C for 4 min 30 s)₂₅, 72°C for 10 min. The megaprimer was gel-purified and used to mutate the pGL3-delta(-275) vector in an *in vitro* mutagenesis reaction mixture containing 1x *Pfu* reaction buffer, 50 ng template DNA, 220 ng megaprimer, 0.25 mM dNTPs, 8% DMSO, and *Pfu* DNA polymerase (3 units/50 μ l). The amplification protocol was as follows: 95°C for 2 min, (95°C for 30 s, 55°C for 1 min, 72°C for 10 min 30 s)₃₀, and 72°C for 20 min. Amplified DNA was selected by *DpnI* digest. The resulting vector, **pGL3-delta(-275)TETO-R1-TETO**, is shown in Figure 29.

Figure 29: The pGL3-delta(-275)TETO-R1-TETO vector

The -275 promoter was modified by the insertion of a TETO sequence on either side of the transcription factor binding site R1. The resulting promoter, -275 TETO-R1-TETO, with a Kozak sequence added, was cloned into the pGL3-Basic Luciferase Reporter Vector, in front of the luciferase (*luc+*) reporter gene. The vector allows assessment of promoter activity using the Dual Luciferase Reporter Assay System.

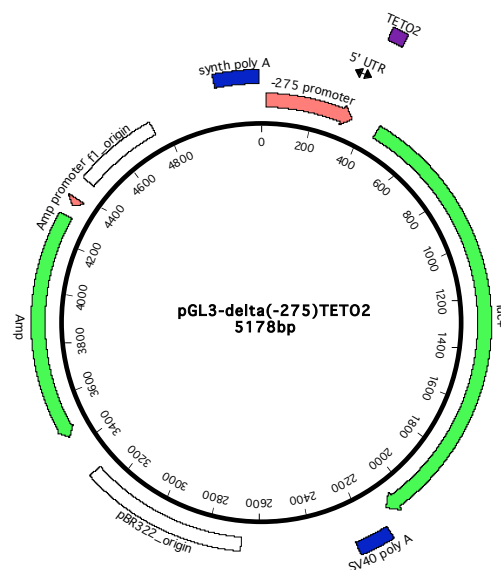


-275 TETO2 promoter - The TETO2 sequence was inserted after the -275 promoter, as follows: The -275 POLD1 promoter sequence described above was amplified from 411 MI human genomic DNA in a reaction mixture containing 1x Thermopol buffer, 150 ng template DNA, 1 μ M forward primer, 1 μ M reverse primer, 0.2 mM dNTPs, 9% DMSO, and *Taq* DNA polymerase (5 units/50 μ l). The primers (Operon, Cologne, Germany) were: 275 forward (AGG GTG GGA GGA GAG AGA AC) and 275 reverse (CCC GCT TCA AAC AGC GTT TCC). The PCR program was as follows: 95°C for 2 min, (95°C for 1 min, 61°C for 1 min, 72°C for 1 min)₃₀, 72°C for 10 min.

The PCR product was purified by gel extraction and used as a template in a second PCR reaction with 1x Thermopol buffer, 1 μ l template DNA, 1 μ M forward primer, 1 μ M reverse primer, 0.2 mM dNTPs, 8% DMSO, and *Taq* DNA polymerase (5 units/50 μ l). The primers were: forward 275 Promoter and reverse 275 tetO2 Promoter:GGT GGT AGA TCT GGA TCC GAT CTC TAT CAC TGA TAG GGA GAT CTC TAT CAC TGA TAG GGA GCC CGC TTC AAA CAG CGT TTC CCG CCA CAG CCT ACG (Microsynth). The PCR product was purified by gel extraction, digested with *Sac*I and *Bgl*II, and cloned into the corresponding restriction sites of a pGL3-Basic Luciferase Reporter Vector, generating the **pGL3-delta(-275)TETO2** vector (Figure 30).

Figure 30: The pGL3-delta(-275)TETO2 vector

TETO2 was inserted after the -275 promoter. This new promoter, -275 TETO2, was cloned into the pGL3-Basic Luciferase Reporter Vector, in front of the luciferase (*luc+*) reporter gene. The vector allows determination of promoter activity using the Dual Luciferase Reporter Assay System.



7.1.2 *POLD1* expression vectors

RNA interference (RNAi)-refractory *POLD1*-3xFLAG constructs - Silent mutations were introduced at two sites in *POLD1*, and a 3xFLAG tag was added to the C-terminus. The sequence-verified clone, IRAUp969B0431D6 (RZPD, Berlin, Germany), containing human *POLD1* cDNA in a pOTB7 vector (referred to hereafter as **pOTB7-POLD1**) was used as a template to generate a megaprimer in a PCR reaction. The reaction mixture contained 1x *Pfu* reaction buffer, 50 ng of template DNA, 1 μ M forward primer, 1 μ M reverse primer, 0.25 mM dNTPs, 8% DMSO, 10% glycerol, and *Pfu* DNA polymerase (3 units/50 μ l). The primers (Metabion, Planegg, Germany) were silent 2135 sense [GAC GGC AGC TGG CGC TGA AaG Tct cCG CtA Ata gCG TgT AtG GaT TtA CcG GCG CCC AGG TGG GCA AGT TGC CGT G] and silent 2655 antisense [GAC CAG CTG GGA GAT gTC aAT cCG aTT aCA gAG gAG aTC gGA aTT GAC GTC CTG TGC GTG AG] (The silent mutations are shown in lowercase letters.). The PCR protocol was as follows: 95°C for 2 min, (95°C for 40 s, 65°C for 1 min, 72°C for 2 min 30 s)₃₀, 72°C for 10 min. This megaprimer was gel-purified and used to mutate the pOTB7-POLD1 vector in an *in vitro* mutagenesis reaction containing 1x *Pfu* reaction buffer, 50 ng of template DNA, 75 ng megaprimer, 0.25 mM dNTPs, and *Pfu* DNA polymerase (3 units/50 μ l). The amplification protocol was as follows: 95°C for 2 min, (95°C for 1 min, 32°C for 2 min, 63°C for 25 min)₃₀, 68°C for 20 min. Amplified DNA was selected by *DpnI* digest. The resulting vector was named pOTB7-POLD1-53. The pOTB7-POLD1 plasmid was used as a template to generate a megaprimer in a PCR reaction containing 1x *Pfu* reaction buffer, 50 ng of template DNA, 1 μ M forward primer, 1 μ M reverse primer, 0.25 mM dNTPs, 8% DMSO and *Pfu* DNA polymerase (3 units/50 μ l). The primers (Metabion) were 1115 *NcoI* sense [CAG TCA CCC ACC GGA AGG GCC tTG GCA GCG CAT TGC GCC CTT G] and D402Asilent1282 antisense [CAC ACG GCC CAG GAA gGG aAA gGT cTG cAC tTT cAG tGT tTG AGC CCG AGA GAT GAG GTA CGG AAG GgC GAA GTT CTG]. The PCR protocol was as follows: 95°C for 2 min, (95°C for 1 min, 32°C for 1 min, 63°C for 25 min)₃₀, 68°C for 20 min. The 1115-1282 megaprimer was gel-purified and used with a modified *in vitro* mutagenesis method. A fraction of the colony used to prepare the pOTB7-POLD1-53 plasmid was used as a template in a megaprimer *in vitro* mutagenesis reaction containing 1x *Pfu* reaction buffer, 375 ng 1115-1282 megaprimer, 0.25 mM dNTPs, 8% DMSO, and *Pfu* DNA

polymerase (3 units/50 μ l). The amplification protocol was as follows: 95°C for 2 min, (95°C for 1 min, 32°C for 2 min, 63°C for 25 min)₃₀, 68°C for 20 min. Amplified DNA was selected by *DpnI* digest. The resulting vectors were named pOTB7-POLD1-534 and pOTB7-POLD1-532.

The pOTB7-POLD1-534 vector was then mutated with a variation of the megaprimer *in vitro* mutagenesis method. The reaction mixture contained 1x *Pfu* reaction buffer, 60 ng of template DNA, 1 μ M forward primer, 1 μ M reverse primer, 0.25 mM dNTPs, 8% DMSO, and *Pfu* DNA polymerase (3 units/50 μ l), and the primers were 1115 NcoI sense [CAG TCA CCC ACC GGA AGG GCC tTG GCA GCG CAT TGC GCC CTT G] (Metabion) 1413wt antisense [GAG AGA TGA GGT ACG GAA GGT CGA AGT TCT GGA TGT TGT AAC CG] (Microsynth). The amplification protocol was as follows: 95°C for 2 min, (95°C for 1 min, 32°C for 2 min, 63°C for 25 min)₃₀, 68°C for 20 min. Amplified DNA was selected by *DpnI* digest. The resulting vector was named pOTB7-POLD1-ND-43.

The pOTB7-POLD1-532 plasmid was used as a template to amplify the *POLD1* gene in a PCR reaction containing 1x *Pfu* reaction buffer, 50 ng of template DNA, 1 μ M forward primer, 1 μ M reverse primer, 0.25 mM dNTPs, 8% DMSO, and *Pfu* DNA polymerase (3 units/50 μ l reaction). The primers (Microsynth) were NCPoldstart forward [GGT GGA TCC GCG GCC GCT ACC ATG GAT GGC AAG CGG CGG CCA G] and CFLAGPold reverse [GGT CTC GAG TCT AGA CCA GGC CTC AGG TCC AGG GG]. The PCR protocol was as follows: 95°C for 2 min, (95°C for 1 min, 56°C for 1 min, 63°C for 20 min)₃₀, 63°C for 20 min. The PCR product was digested with *NofI* and *XbaI* and cloned into the corresponding restriction sites in a p3xFLAG-CMV-14 vector (Sigma-Aldrich, St. Louis, MO, USA) to generate the p3xFLAG-CMV-14-P534 vector.

The fragment of this vector between the *SacII* and *PmlI* sites was replaced with the corresponding fragment of pOTB7-POLD1 to produce the p3xFLAG-CMV-14-P5341 vector. The p3xFLAG-CMV-14-P5341 fragment between the *KpnI* and *SacII* sites was replaced with the corresponding fragment of pOTB7-POLD1-ND-43 or pOTB7-POLD1-534 to produce the **p3xFLAG-CMV-14-p125exo+** vector (Figure 31) and the **p3xFLAG-CMV-14-p125pdef** vector (Figure 32), respectively. These two vectors were verified by sequencing.

Figure 31: The p3xFLAG-CMV-14-p125exo+ vector

The CMV promoter allows mammalian expression of wild-type *POLD1* with silent mutations at two sites inside the ORF and a 3xFLAG tag added to the C-terminus. G418 can be used for selection of stable clones.

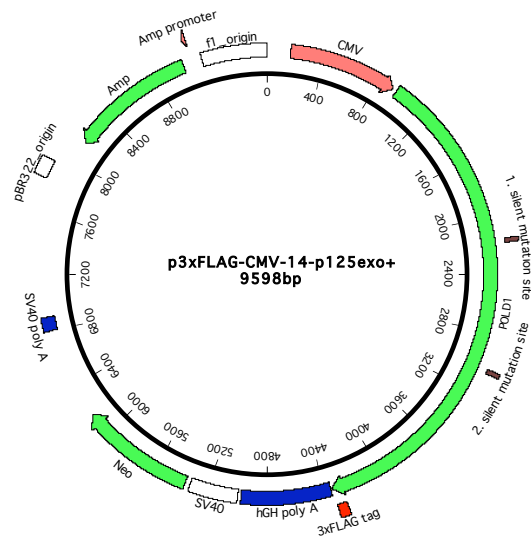
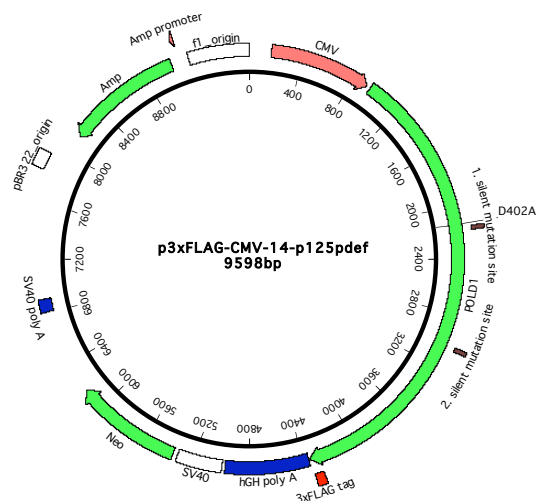


Figure 32: p3xFLAG-CMV-14-p125pdef

The CMV promoter allows mammalian expression of proofreading-deficient *POLD1* with silent mutations at two sites inside the ORF and a 3xFLAG tag added to the C-terminus. The D402A mutation in *POLD1* inactivates the exonuclease. G418 can be used for selection of stable clones.

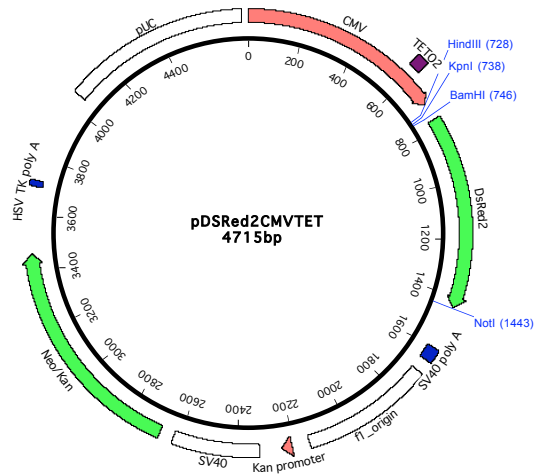


Inducible expression vector – To construct an inducible expression vector for use with T-REx U2OS cells, we replaced the fragment between the *Bam*HI and *Ase*I sites in a pDsRed2-Mito vector (Clonetech, Mountain View, CA, USA) with the corresponding fragment from a pcDNA5/TO vector (Invitrogen, Carlsbad, CA, USA). The result was the **pDsRed2CMVTET** vector shown in

Figure 33.

Figure 33: pDSRed2CMVTET

The inducible CMV promoter allows dox- or tet-induced mammalian expression of the *DsRed2* gene, which encodes a red fluorescent protein derived from *Discosoma sp.* in a cell line expressing tetR. *DsRed2* expression can be detected by fluorescence microscopy. Since repression in dox- or tet-absence is mediated by tetR, the vector allows to test for proper tetR expression. *DsRed2* can be replaced with any gene of interest or the gene of interest can be cloned to the N-terminus of *DsRed2*, forming a red fluorescent fusion protein.

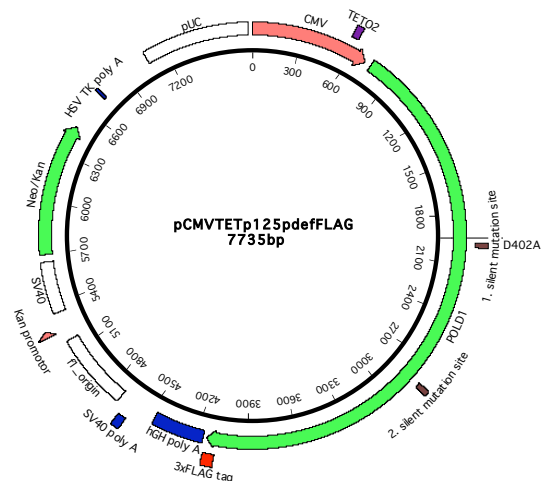


Inducible *POLD1* expression vectors - Each of the following 4 *POLD1* variants was cloned into the inducible expression vector pDSRed2CMVTET.

Proof-reading deficient POLD1. The p3xFLAG-CMV-14-p125pdef plasmid was digested with *HindIII* and *SphI*, and the *POLD1* ORF with a 3xFLAG tag was cloned into pDsRed2CMVTET. This vector was then subjected to *NotI* digestion, fill-in with *T4* DNA polymerase (New England Biolabs), and *HindIII* digestion. The resulting vector, **pCMVTETp125pdefFLAG**, no longer contained *DsRed2* (Figure 34).

Figure 34: pCMVTETp125pdefFLAG

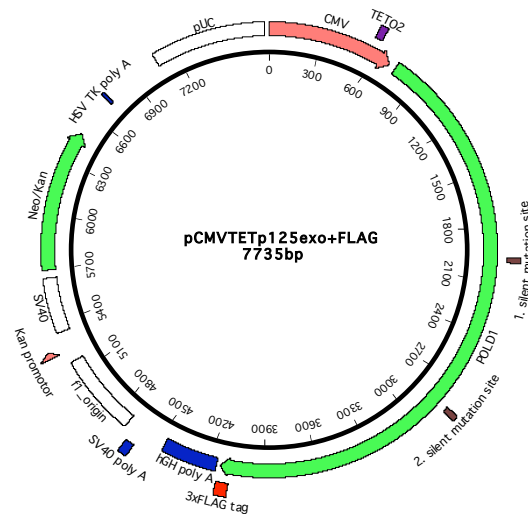
The inducible CMV promoter controls expression of proofreading-deficient *POLD1* with the D402A mutation, silent mutations at two sites, and a 3xFLAG tag. G418 can be used for selection of stable clones.



Wild-type POLD1. The fragment between the *KpnI* and *XbaI* sites in the pCMVTETp125pdefFLAG vector was then replaced with the corresponding fragment from the p3xFLAG-CMV-14-p125exo+ vector to generate the pCMVTETp125exo+FLAG vector (Figure 35).

Figure 35: The pCMVTETp125exo+FLAG vector

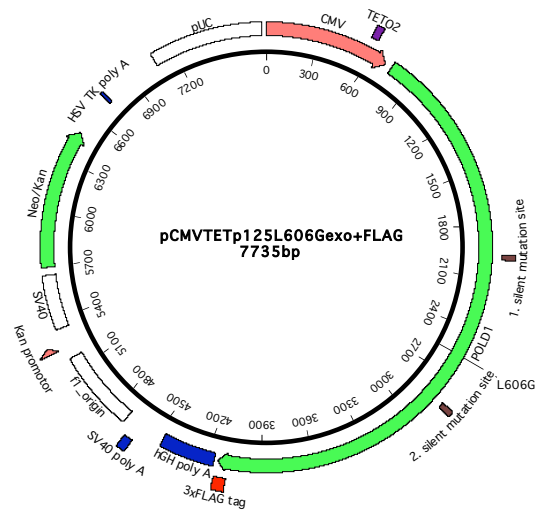
The vector contains wild-type *POLD1* with silent mutations at two sites and a 3xFLAG tag controlled by the inducible CMV promoter. G418 can be used for selection of stable clones.



Error-prone POLD1. The pCMVTETp125exo+FLAG vector was mutated in an *in vitro* mutagenesis reaction. The reaction mixture contained 1x *Pfu* reaction buffer, 35 ng of template DNA, 1 μ M sense primer, 1 μ M antisense primer, 0.25 mM dNTPs, 8% DMSO and *Pfu* DNA polymerase (3 units/50 μ l). The primers (from Microsynth) were p125 L606G sense [CTG GAC TTC TCC TCG gGC TAC CCG TCC ATC ATG ATG] and p125 L606G antisense [CAT CAT GAT GGA CGG GTA GCc CGA GGA GAA GTC CAG]. The L606G mutation is shown in lower-case letters. The amplification protocol was as follows: 95°C for 2 min, (95°C for 1 min, 55°C for 1 min, 63°C for 35 min)₃₀, 68°C for 20 min. Amplified DNA was selected by *DpnI* digest. The resulting vector, pCMVTETp125L606Gexo+FLAG, is shown in Figure 36.

Figure 36: pCMVTETp125L606Gexo+FLAG

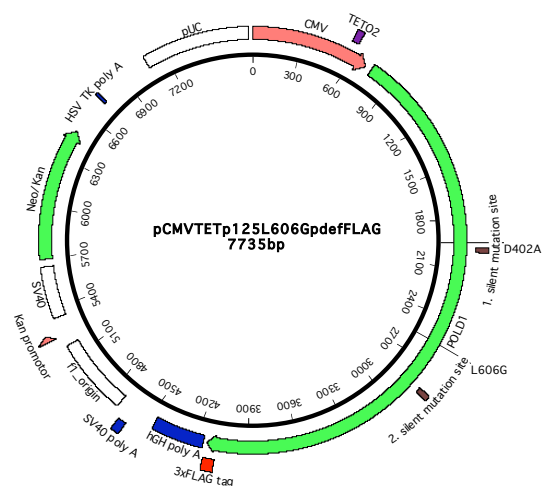
Error-prone *POLD1*, which includes the L606G mutation, silent mutations at two sites, and a 3xFLAG tag, is under control of the inducible CMV promoter. G418 can be used for selection of stable clones.



Double-mutant POLD1. The pCMVTETp125pdefFLAG vector was mutated in an *in vitro* mutagenesis reaction, as described in the previous paragraph, with p125 L606G sense and p125 L606G antisense primers. The resulting construct, **pCMVTETp125L606GpdefFLAG**, is shown in Figure 37.

Figure 37: pCMVTETp125L606GpdefFLAG

Double-mutant *POLD1*, which contains the D402A and L606G mutations, silent mutations at two sites, and a 3xFLAG tag, is under control of the inducible CMV promoter. G418 can be used for selection of stable clones.

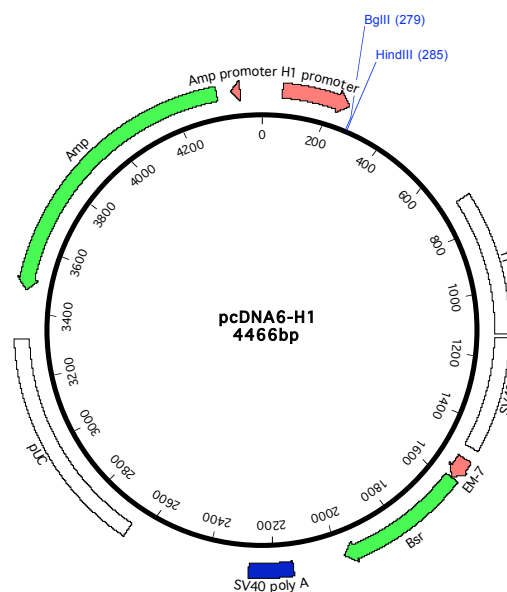


7.1.3 shRNA expression vectors

Inducible shRNA expression vector for selection with blasticidin – Vectors for the inducible expression of shRNA targeting *MSH2* and *POLD1* were constructed for use with T-REx U2OS cells and pDSRed2CMVTET-based expression vectors. The pcDNA6/TR vector (Invitrogen, Carlsbad, CA, USA) was digested with *Bgl*II, and blunt ends were generated by treatment with *T4* DNA polymerase. This linearized vector was gel-purified, digested with *Xho*I, and subjected to a second gel purification. The pSUPERIOR.neo+gfp vector (OligoEngine, Seattle, WA, USA) was digested with *Sph*I, and blunt ends were generated by treatment with *T4* DNA polymerase. This linearized vector was gel-purified and digested with *Xho*I. The fragment containing the H1 promoter was gel-purified and cloned into pcDNA6/TR to produce **pcDNA6-H1** (Figure 38).

Figure 38: The pcDNA6-H1 vector

ShRNA expression is under control of the inducible H1 promoter. Annealed DNA oligomers containing an shRNA sequence can be cloned into this vector with *Bgl*II and *Hind*III restriction sites. Blasticidin can be used for selection of stable clones.



The 1. site-19nt sense [GAT CCC gac cct caa ggt aca aac aTT CAA GAG Atg ttt gta cct tga ggg tcT TTT TA] and antisense [AGC TTA AAA Aga ccc tca agg tac aaa caT CTC TTG AAt gtt tgt acc ttg agg gtc GG] primers (Metabion) were annealed and cloned into pcDNA6-H1 with *Bgl*II and *Hind*III restriction sites. Lowercase letters denote the shRNA target sequence in the sense and antisense orientations. The

resulting vector, **pcDNA6-H1-p125sh1s19**, expresses a 19-mer shRNA that targets *POLD1* (1289 nucleotides after start of transcription).

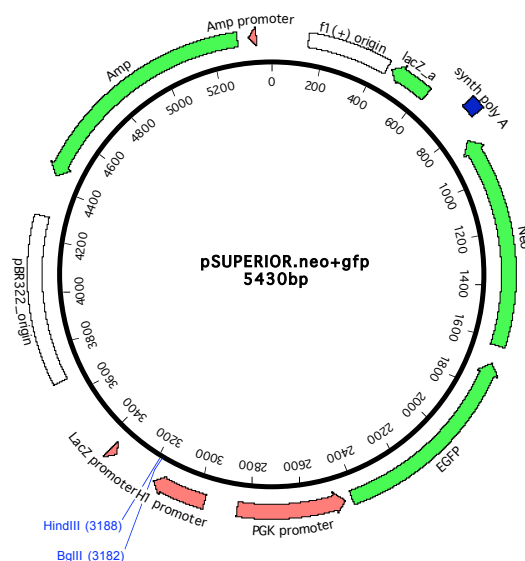
The 1. site-29nt sense [GAT CCC gac cct caa ggt aca aac att ccc ttt ccT CAA GAG gga aag gga atg ttt gta cct tga ggg tcT TTT TA] and antisense [AGC TTA AAA Aga ccc tca agg tac aaa cat tcc ctt tcc CTC TTG Agg aaa ggg aat gtt tgt acc ttg agg gtc GG] primers (Metabion) were annealed and cloned into pcDNA6-H1 with *Bgl*III and *Hind*III restriction sites, generating **pcDNA6-H1-p125sh1s29**. This vector expresses a 29-mer shRNA targeting *POLD1* (1289 nucleotides after start of transcription).

The shMSH2-2 21nt sense [GAT CCC aag ccc agg atg cca ttg tta TCA AGA Gta aca atg gca tcc tgg gct tTT TTT A] and antisense [AGC TTA AAA Aaa gcc cag gat gcc att gtt aCT CTT Gat aac aat ggc atc ctg ggc ttG G] primers (Metabion) were annealed and cloned into pcDNA6-H1 with *Bgl*III and *Hind*III restriction sites, generating **pcDNA6-H1-MSH2-2sh21**. This vector expresses a 21-mer shRNA that targets *MSH2* (1783 nucleotides after start of transcription). The target sequence had been used previously for siRNA-mediated *MSH2* knockdown [216].

Inducible shRNA expression vectors based on pSUPERIOR.neo+gfp –
We constructed inducible shRNA expression vectors targeting *MSH2*, *MSH6*, and *POLD1*. Figure 39 shows the basic properties of the empty vector pSUPERIOR.neo+gfp.

Figure 39: pSUPERIOR.neo+gfp

ShRNA expression is under control of the inducible H1 promoter. Annealed DNA oligomers containing an shRNA sequence can be cloned into pSUPERIOR.neo+gfp with *Bgl*III and *Hind*III restriction sites. G418 can be used for selection of stable clones. The concomitant expression of *EGFP* is a major advantage of the pSUPERIOR.neo+gfp for experiments involving transient shRNA expression, because it can be assessed with fluorescence microscopy to estimate transfection efficacy.



The 1. site-19nt sense and antisense primers were annealed and cloned into pSUPERIOR.neo+gfp with *Bgl*II and *Hind*III restriction sites. The **pSUPERIOR.neo+gfp-p125sh1s19** thus generated expresses a 19-mer shRNA that targets *POLD1* (1289 nucleotides after start of transcription).

The 1. site-29nt sense and antisense primers were annealed and cloned into pSUPERIOR.neo+gfp with *Bgl*II and *Hind*III restriction sites, generating **pSUPERIOR.neo+gfp-p125sh1s29**. This vector expresses a 29-mer shRNA that targets *POLD1* (1289 nucleotides after start of transcription).

The 5UTR-21-p125-19nt sense [GAT CCC gcg tag gct gtg gcg gga aTT CAA GAG Att ccc gcc aca gcc tac gcT TTT TA] and antisense [AGC TTA AAA Agc gta ggc tgt ggc ggg aaT CTC TTG AAt tcc cgc cac agc cta cgc GG] primers (Metabion) were annealed and cloned into pSUPERIOR.neo+gfp with *Bgl*II and *Hind*III restriction sites, generating **pSUPERIOR.neo+gfp-p125sh5UTR19**. This vector expresses a 19-mer shRNA that targets *POLD1* in the 5' UTR (21 nucleotides after start of transcription).

The shMSH2-2 21nt sense and antisense primers were annealed and cloned into pSUPERIOR.neo+gfp with *Bgl*II and *Hind*III restriction sites, generating **pSUPERIOR.neo+gfp-MSH2-2sh21**. This vector expresses a 21-mer shRNA that targets *MSH2* (1783 nucleotides after start of transcription).

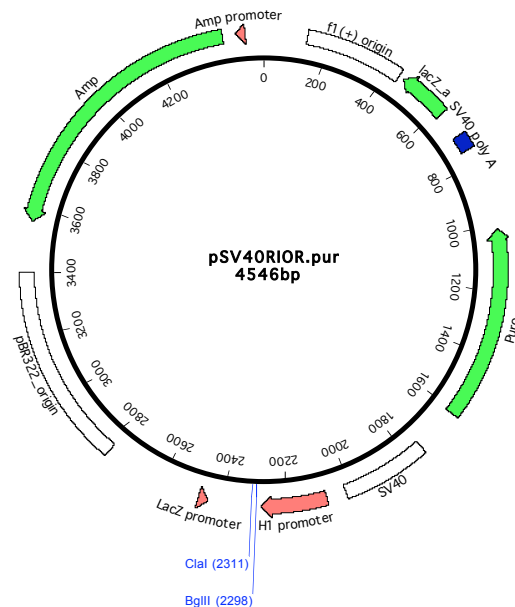
The MSH2-287 forward [GAT CCC Cga atc tgc aga gtg ttg tgT TCA AGA GAc aca aca ctc tgc aga ttc TTT TTG GAA A] and reverse [AGC TTT TCC AAA AAg aat ctg cag agt gtt gtg TCT CTT GAA cac aac act ctg cag att cGG G] primers (Microsynth) were annealed and cloned into pSUPERIOR.neo+gfp with *Bgl*II and *Hind*III restriction sites, generating **pSUPERIOR.neo+gfp-MSH2-287sh19**. This vector expresses a 19-mer shRNA, which targets *MSH2* (287 nucleotides after start of transcription).

The shMSH6 sense [GAT CCC cgc cat tgt tgc aga ttt aTT CAA GAG Ata aat ctc gaa caa tgg cgT TTT TA] and antisense [AGC TTA AAA Acg cca ttg ttc gag att taT CTC TTG AAt aaa tct cga aca atg gcg GG] primers (Metabion) were annealed and cloned into pSUPERIOR.neo+gfp with *Bgl*II and *Hind*III restriction sites, generating **pSUPERIOR.neo+gfp-MSH6sh19**. This vector expresses a 19-mer shRNA that targets *MSH6* (1881 nucleotides after start of transcription). The target sequence had been used previously for siRNA-mediated knockdown of MSH6 [217].

Inducible shRNA expression vector for selection with puromycin - Inducible shRNA expression vectors targeting *POLD1* were constructed for use with T-REx U2OS cells and pDSRed2CMVTET-based expression vectors. Figure 40 shows the basic properties of the empty vector pSV40RIOR.pur.

Figure 40: pSV40RIOR.pur

ShRNA expression is under control of the inducible H1 promoter. Annealed DNA oligomers containing an shRNA sequence can be cloned into pSV40RIOR.pur with *Bgl*II and *Cla*I restriction sites. Exposure of cells to puromycin for one day allows selection of a transfected cell population. Longer exposure can be used to generate stable clones of the transfected cells.



The pTREpur vector (Clontech, Palo Alto, CA, USA) was digested with *Xho*I, and blunt ends were generated by treatment with *T4* DNA polymerase. The puromycin-resistance gene was cloned into pSUPERIOR.neo+gfp-p125sh1s29 that was prepared by *Sph*I and *Bam*HI digest as well as *T4* DNA polymerase treatment. The resulting vector, **pSV40RIOR.pur-p125sh1s29**, expresses a 29-mer shRNA that targets *POLD1* (1289 nucleotides after start of transcription). The puromycin resistance gene of pTREpur was cloned into pSUPERIOR.neo+gfp-p125sh5UTR19 that was prepared by *Sph*I and *Bam*HI digest as well as *T4* DNA polymerase treatment. The resulting vector, **pSV40RIOR.pur-125sh5UTR19**, expresses a 19-mer shRNA, which targets *POLD1* in the 5' UTR (21 nucleotides after start of transcription). The pSUPERIOR.neo+gfp vector was digested with *Kpn*I and *Eco*RI, and the empty H1 promoter was cloned into the corresponding restriction sites of pSV40RIOR.pur-p125sh1s29 to produce **pSV40RIOR.pur**. pSV40RIOR.pur is the empty cloning vector of pSV40RIOR.pur-p125sh1s29 and pSV40RIOR.pur-125sh5UTR19. However since the *Hind*III restriction site is not unique in

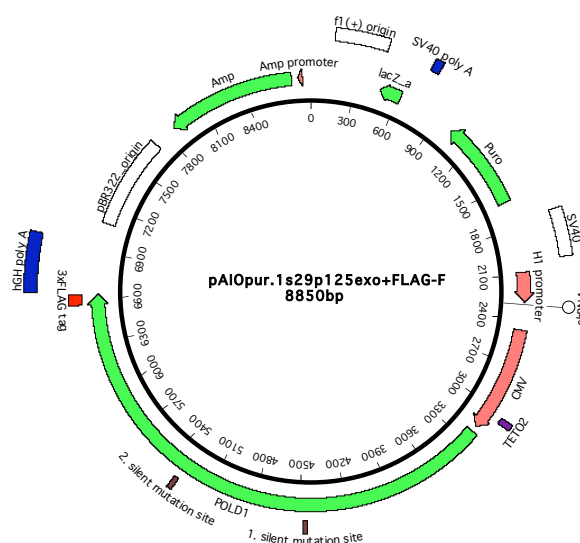
pSV40RIOR.pur, cloning from the on-hand annealed oligomers was not possible. In order to save time, the shRNA expression cassette was cloned from previously constructed shRNA expression vectors.

7.1.4 All-in-one *POLD1* gene replacement vectors

Inducible gene-replacement by a single vector (selection with puromycin) – In an attempt to facilitate gene replacement, all-in-one vectors were created, which combined inducible shRNA expression with inducible variant-gene expression. The pSV40RIOR.pur-p125sh1s29 vector was digested with *KpnI*, and blunt ends were generated by treatment with *T4* DNA polymerase. After heat inactivation of the *T4* polymerase, the plasmid was digested with *AflIII*. This linearized pSV40RIOR.pur-p125sh1s29 vector was used for the subsequent cloning. The vector pCMVTETp125exo+FLAG was digested with *MfeI*, and blunt ends were generated by treatment with *T4* DNA polymerase. After heat inactivation of The *T4* polymerase was heat-inactivated, and the plasmid was digested with *AflIII*. The fragment containing the inducible CMV promoter, the *POLD1* ORF, and the 3xFLAG tag was then cloned into the linearized pSV40RIOR.pur-p125sh1s29 vector to obtain the **pAIOpur.1s29p125exo+FLAG** construct (Figure 41).

Figure 41: The pAIOpur.1s29p125exo+FLAG vector

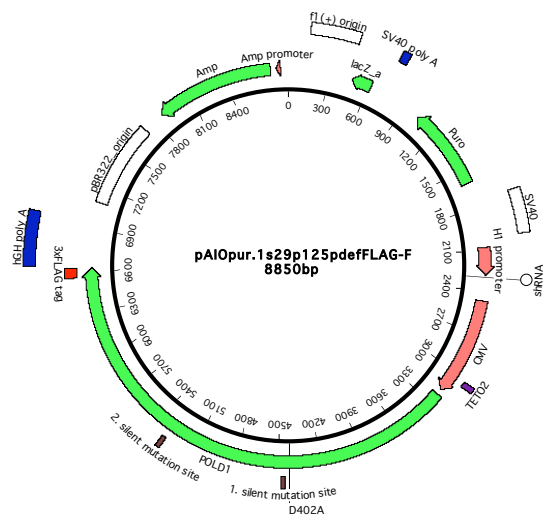
This all-in-one vector allows inducible one-step replacement of the endogenous *POLD1* gene with exogenous wild-type *POLD1*. Expression of the 29-mer shRNA targeting the first silent mutation site in *POLD1* is controlled by the inducible H1 promoter. Expression of wild-type *POLD1* with silent mutations at two sites and a 3xFLAG tag is under control of the inducible CMV promoter. Exposure of cells to puromycin for one day allows selection of a transfected cell population. Longer exposure can be used to generate stable clones of the transfected cells.



The same approach was used to clone the inducible CMV promoter with the *POLD1* ORF and the 3xFLAG tag from pCMVTETp125pdefFLAG into the linearized pSV40RIOR.pur-p125sh1s29 vector. The resulting vector **pAIOpur.1s29p125pdefFLAG** is shown in Figure 42.

Figure 42: pAIOpur.1s29p125pdefFLAG

This all-in-one vector allows inducible replacement of the endogenous *POLD1* gene with exogenous proofreading-deficient *POLD1*. Expression of the 29-mer shRNA targeting the first site of silent mutations in *POLD1* is under control of the inducible H1 promoter. Proofreading-deficient *POLD1* with the D402A mutation, silent mutations at two sites, and a 3xFLAG tag is controlled by the inducible CMV promoter. Exposure of cells to puromycin for one day allows selection of a transfected cell population. Longer exposure can be used to generate stable clones of the transfected cells.

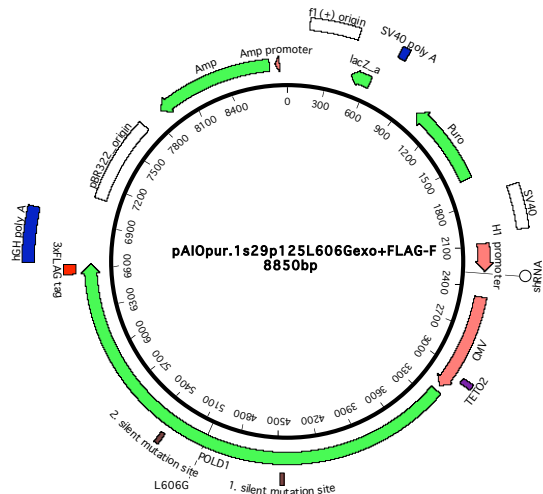


The inducible CMV promoter with the *POLD1* ORF and the 3xFLAG tag from pCMVTETp125L606Gexo+FLAG was cloned into the linearized pSV40RIOR.pur-p125sh1s29 with the same method. The resulting vector, **pAIOpur.1s29p125L606Gexo+FLAG**, is shown in

Figure 43.

Figure 43: pAIOpur.1s29p125L606Gexo+FLAG

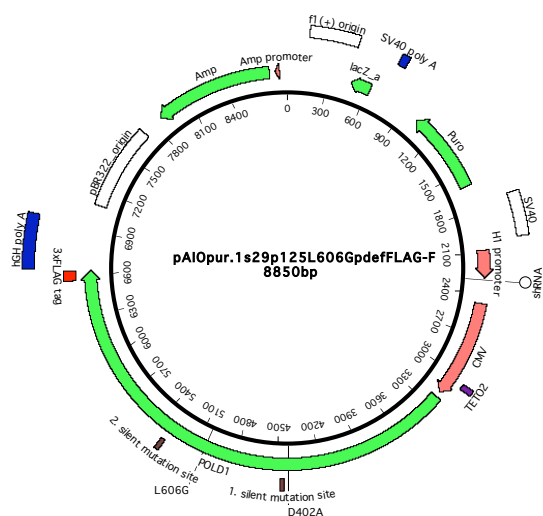
This all-in-one vector allows inducible replacement of the endogenous *POLD1* gene with exogenous error-prone *POLD1*. Expression of the 29-mer shRNA targeting the first site of silent mutations in *POLD1* is under control of the inducible H1 promoter. Error-prone *POLD1* with the L606G mutation, silent mutations at two sites, and a 3xFLAG tag is controlled by the inducible CMV promoter. Exposure of cells to puromycin for one day allows selection of a transfected cell population. Longer exposure can be used to generate stable clones of the transfected cells.



In the same manner, the inducible CMV promoter with the *POLD1* ORF and the 3xFLAG tag from pCMVTETp125L606GpdefFLAG was cloned into the linearized pSV40RIOR.pur-p125sh1s29 to produce the **pAIOpur.1s29p125L606GpdefFLAG** vector (Figure 44).

Figure 44: pAIOpur.1s29p125L606GpdefFLAG

This all-in-one vector allows inducible replacement of the endogenous *POLD1* gene with exogenous double-mutant *POLD1*. Expression of the 29-mer shRNA targeting the first site of silent mutations in *POLD1* is under control of the inducible H1 promoter. Double-mutant *POLD1*, which contains the D402A and L606G mutations, silent mutations at two sites, and a 3xFLAG tag, is under control of the inducible CMV promoter. Exposure of cells to puromycin for one day allows selection of a transfected cell population. Longer exposure can be used to generate stable clones of the transfected cells.



7.2 Cell culture

7.2.1 Cell lines

HEK293T cells [218] were cultivated in Dulbecco's modified Eagle's medium with Eagle salts supplemented with 5% FCS, 100 U/ml penicillin, and 100 µg/ml streptomycin (all from Gibco/Invitrogen, San Diego, CA, USA).

T-REx HeLa cells (Invitrogen) were grown in Minimum Essential Medium with Earle's Salts (Gibco/Invitrogen) supplemented with 10% Tet System-approved FCS (Clontech, Palo Alto, CA, USA), 100 U/ml penicillin (Gibco/Invitrogen), 100 µg/ml streptomycin (Gibco/Invitrogen), 2 mM L-glutamine, and 5 µg/ml blasticidin (Invitrogen).

T-REx U2OS cells (Invitrogen) were grown in Dulbecco's modified Eagle's medium with Eagle salts (Gibco/Invitrogen) supplemented with 10% Tet System-approved FCS (Clontech), 100 U/ml penicillin (Gibco/Invitrogen), 100 µg/ml streptomycin (Gibco/Invitrogen), and 50 µg/ml hygromycin B (Omnilab, Mettmenstetten, Switzerland). T-REx U2OS cells express the tetracycline repressor, tetR, encoded in the episomal plasmid pCEP4-tetR [200], which is stable under hygromycin B-selection.

Clones of T-REx U2OS were maintained in medium containing the selection agent, i.e., G418 (Invitrogen) at a concentration of 400µg/ml or puromycin (Invivogen, San Diego, CA, USA) at a concentration of 1µg/ml. In general, gene expression was induced with doxycycline (Sigma-Aldrich) at a concentration of 10ng/ml.

HeLa cells were grown in Dulbecco's modified Eagle's medium with Eagle salts supplemented with 10% Tet System-approved FCS, 100 U/ml penicillin, and 100 µg/ml streptomycin.

All cell lines were cultured at 37°C in a humidified atmosphere containing 5% CO₂ atmosphere.

7.2.2 Vector transfection and isolation of stable clones

For vector transfection, we used the FuGene 6 Transfection Agent (Roche, Basel, Switzerland) according to manufacturer's instructions. Stable clones were selected 2 days post-transfection with 400µg/ml G418 or 1µg/ml puromycin and isolated after 2-3 weeks.

7.2.3 siRNA transfection

Small interfering RNAs were purchased from Eurofins MWG Operon (Ebersberg, Germany), and all transfections were carried out with the Oligofectamine reagent (Invitrogen) in accordance with the manufacturer's instructions.

HeLa FMI cells were seeded into 10-cm dishes without antibiotics. The cells were transfected the following day with 350 pmol of siRNA directed against one of the five *POLD1* sequences listed in Table 5 and 350 pmol of siRNA directed against *MSH6*. Control cells were transfected with 700 pmol siRNA targeting either firefly luciferase or *MSH6* (CGC CAU UGU UCG AGA UUU A, [217]). After 24 hours, a portion of the cells was harvested; the remaining portion was split and harvested after 48 and 72 hours.

Table 5: siRNA targeting *POLD1*

Name	5'-to-3' sequence
5'UTR-21	GCG UAG GCU GUG GCG GGA A
5'UTR-33	GCG GGA AAC GCU GUU UGA A
ORF-1253	GAA CUU CGA CCU UCC GUA A
ORF-1289	GAC CCU CAA GGU ACA AAC A
ORF-2149	ACU CCG UAU ACG GCU UCA C

7.2.4 Dual luciferase promoter activity assays

T-REx HeLa cells were seeded without antibiotics into 24-well plates to reach confluency of about 30% the following day. The cells were then transfected (in triplicate) with the corresponding pGL3 luciferase reporter vector (10 ng) plus the internal control vector pRL-SV40 (0.5 ng). Six hours after transfection, the culture medium without antibiotics was replaced with complete medium and, when indicated, doxycycline was added at a concentration of 1µg/ml. Thirty hours after transfection,

the cells were washed once with PBS, incubated for 15 min at room temperature on a rocking platform with 50µl passive lysis buffer from the Dual Luciferase Reporter Assay System (Promega), and assayed for luciferase activity according to the manufacturer's instructions with a VERSAmax tuneable microplate reader.

7.3 Analytical procedures

7.3.1 Whole cell extracts

Cells were harvested and washed once with PBS (with centrifugation steps at 4° C and 240 x g for 5 min) before the pellet was dissolved in an appropriate volume of lysis buffer (50 mM Tris-HCl, pH 8.0; 125 mM NaCl; 1% NP40; 1 mM EDTA; 1 mM phenylmethylsulfonyl fluoride (PMSF); complete protease inhibitor [EDTA-free] (Roche, Basel, Switzerland) according to the manufacturer's instructions; 1 mM sodium orthovanadate; and 20 mM NaF). Lysis proceeded on ice for 60 min, and the extracts were then clarified by 15 min of centrifugation (20,000 x g) at 4°C. The protein-containing supernatant was snap-frozen and stored at -80°C.

7.3.2 Western blot analysis

Generally, 15 µg of whole-cell or nuclear extract was denatured, reduced, and subjected to 6.0%-12.5% SDS-PAGE. In preparation for western blot analysis, the proteins were transferred to Hybond-P PVDF membranes (Amersham Pharmacia Biotech, Little Chalfont, UK) with a semi-dry transfer protocol (0.8 mA/cm² for 90 min), and the membranes were blocked for 40 min with 5% non-fat dry milk / PBS-T (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 2 mM KH₂PO₄, 0.5 % Tween-20, pH 7.4). Primary antibodies in 2.5% non-fat dry milk / PBS-T were added, and the membranes were incubated for 1 h at room temperature or overnight at 4° C. The primary antibodies were polyclonal rabbit anti-POLD1 (p125) (a kind gift from Ulrich Hübscher, dilution 1:1000); monoclonal mouse anti-POLD1 (p125) (Abnova, Taipei, Taiwan; dilution 1:1000); polyclonal rat anti-POLD2 (p50) (a kind gift from Ulrich Hübscher, dilution 1:1000); monoclonal mouse anti-POLD3 (p68) (Abnova, dilution 1:1000); monoclonal mouse anti-POLD4 (p12) (Abnova, dilution 1:200); monoclonal mouse anti-FLAG (M2, Sigma-Aldrich; dilution 1:50 000); polyclonal rabbit anti-TFIIH (p89) (Santa Cruz Biotechnology Inc., Santa Cruz, CA; dilution 1:1000); monoclonal

mouse anti-MSH6 (BD Transduction Labs, Franklin Lakes, NJ, USA; dilution 1:2000); monoclonal mouse anti-MSH2 (Oncogene Research Products, San Diego, CA, USA; dilution 1:500); and monoclonal mouse anti- β -tubulin (Santa Cruz Biotechnology Inc., dilution 1:4000). The membranes were then washed three times in PBS-T and incubated with horseradish peroxidase-coupled secondary antibody in 2.5% non-fat dry milk / PBS-T for 1 h at room temperature or overnight at 4° C. After three other washes with PBS-T, immunoreactive proteins were detected with an enhanced chemiluminescence kit (ECL, Amersham Pharmacia Biotech, Little Chalfont, UK).

7.3.3 Gel sequencing

A 10% denaturing polyacrylamide gel was prepared as follows: Urea (42.04 g) was dissolved in 25 ml of 40% acrylamide (19:1 acrylamide/bis-acrylamide); 20 ml 5xTBE; and 23.75 ml H₂O. One milliliter of 10% ammonium persulfate and 100 μ l TEMED were added to start polymerization, and a 40-comb gel was poured. The gel was preheated for 30 min before samples were loaded. After 10 min of denaturing at 95°C, 2.5 μ l of each sample was loaded onto the gel. The reaction products were separated by electrophoresis. The gel was transferred to an X-ray cassette, and films were exposed for prolonged amounts of time at -80°C.

7.4 Polymerase δ analysis

7.4.1 Preparation of nuclear extracts from U2OS T-Rex cells

The selected clone was grown in complete medium, with or without dox (10 ng/ml), for at least 3 days, and log-phase cells (from 2 to 10 x 10⁸) were harvested from each 16 x 15 cm plate. The cells were centrifuged (500 x g at 4°C for 15 min). The pellet was resuspended in 24ml cold isotonic buffer (20 mM HEPES, pH 7.80; 5 mM KCl; 1.5 mM MgCl₂; 0.2 mM PMSF; 3 tablets/250 ml complete protease inhibitor [EDTA-free] (Roche); 0.5 μ g/ml leupeptin; 1 mM DTT; and 250 mM sucrose) and centrifuged at 500 x g at 4°C for 10 min. The cells were centrifuged and resuspended twice in 24 ml hypotonic buffer (isotonic buffer without sucrose) and centrifuged again at 500 x g at 4°C for 10 min. The pellet was resuspended in one pellet volume of hypotonic buffer, transferred to a tissue grinder, and lysed. The nuclei were spun down at 3000 x g for 7 min and resuspended in one pellet volume of extraction buffer (25 mM HEPES, pH 7.80; 10% sucrose; 1 mM PMSF; 0.5 mM DTT; and 1 μ g/ml

leupeptin). The pellet volume was measured, and 0.031 volumes of 5-M NaCl were added one drop at a time to avoid uneven increases in the salt concentration (final concentration: 0.155 M). The mixture was rotated for 1h at 4°C to allow the proteins to leave the nucleus, and the nuclear debris was then centrifuged at 14,500 x g at 4°C for 20 min. The supernatant was dialyzed twice (1 h each time) with 1 liter of dialysis buffer (25 mM HEPES, pH 7.80; 50 mM KCl; 0.1 mM EDTA; 10% sucrose; 1 mM PMSF; 2 mM DTT; and 0.1 µg/ml leupeptin). The extract was clarified by 15 min of centrifugation for 15 min at 20'000 x g at 4°C, and aliquots were snap-frozen in liquid nitrogen and stored at -80°C.

7.4.2 Immunoprecipitation of polymerase δ

ANTI-FLAG M2 agarose affinity gel beads (Sigma-Aldrich) were washed 3 times in dialysis buffer with centrifugation steps at room temperature and 30 x g for 1 min. The beads were resuspended in one pellet volume of dialysis buffer. Nuclear extracts were diluted in dialysis buffer, and 20 µl of a 50% M2 agarose affinity gel suspension was used per milligram of nuclear extract. After a 2-h incubation at 4°C on a tube rotator, the beads were washed 4 times with dialysis buffer. Centrifugation steps were carried out at 4°C and 30 x g for 1 min. The bound polymerase d was eluted from the beads by shaking them (300 rpm at 4°C for 45 min) with 20 µl elution buffer (300µg/ml 3xFLAG peptide in dialysis buffer) per milligram of the initial nuclear extract. The beads were removed by centrifugation at 110 x g at 4°C, and aliquots of the supernatant were snap-frozen in liquid nitrogen and stored at -80°C. Aliquots corresponding to 500µg of input of nuclear extract were analyzed by Western blot for the presence of each subunit of polymerase δ .

7.4.3 Primer extension assay

Labeling of the primer strand

Ten picomols of the DNA primer strand (41mer ATC CTG ATT GCT ATC TGA ATA TGG TGG TGG TGG GCG CCG GC from Microsynth) was incubated at 37° C with 30 µCi (1.11 MBq) of P³²- γ -ATP (Hartmann Analytic, Braunschweig, Germany) and 15 U *T4* polynucleotide kinase (New England Biolabs) in a reaction buffer containing 70 mM Tris-HCl (pH 7.6), 10 mM MgCl₂, and 10 mM DTT. After 45 min, another 10 U of *T4* polynucleotide kinase were added, and the labeling incubation

continued for 30 min. Water was added to the reaction mixture (20 μ l) to produce a final volume of 100 μ l. The labeled oligomer was purified by Sephadex G25 chromatography to eliminate excess P^{32} - γ -ATP. A similar labeling was done with the template strand (100mer TAC AAC CAA GAG CAT ACT GTA AGA TAG ATC ACG ATG ACG GCC AGT GCC GAA TTC ACA CCG CCG GCG CCC ACC ACC ACC ATA TTC AGA TAG CAA TCA GGA T, Microsynth) in order to obtain a 100 bp marker.

Annealing

Three picomols of the P^{32} - γ -ATP-labeled 41mer was annealed with 3.0 pmol of the 100mer in 1 x *T4* polynucleotide kinase reaction buffer heated to 95°C for 5 min and cooled overnight to room temperature.

Reaction

The elongation reaction mixture (volume, 10 μ l; pH 7.0) contained 50 mM Bis-Tris, 12.5 mM HEPES, 1.5 mM DTT, 0.25 mg/ml BSA, 150 μ g/ml 3xFLAG peptide, 0.05 mM EDTA, 0.05 μ g/ml leupeptin, 25 mM KCl, 0.5 mM PMSF, 5% sucrose, 10 mM $MgCl_2$ and, when indicated, 250 μ M of dTTP dATP, dCTP, dGTP (Sigma-Aldrich), 1 μ M PCNA along with 25 fmol template. The reaction was started by incubating 5 μ l of template solution (100mM Bis-Tris [pH 6.5], 2 mM DTT, 0.5 mg/ml BSA, 20 mM $MgCl_2$, 5 fmol/ μ l labeled template and, when indicated, 500 μ M of dTTP dATP, dCTP, dGTP and 1 μ M PCNA) with 5 μ l of eluate containing polymerase δ for 10 min at 37°C. The reaction was stopped by adding 10 μ l of gel loading buffer (95% formamide, 20mM EDTA, 0.02% bromophenol blue and 0.02% xylene cyanol) or 90 μ l precipitation buffer (200 μ g/ml tRNA, 300mM $NaCH_3COOH$ [pH 5.2], 35 mM EDTA). In the latter case, the samples were precipitated with ethanol, dried, checked with a Geiger counter, and dissolved in gel-loading buffer for an equal loading of radioactivity.

8 References

1. Watson, J.D. and F.H. Crick, *Genetical implications of the structure of deoxyribonucleic acid*. *Nature*, 1953. **171**(4361): p. 964-7.
2. Lodish, H., *et al.*, *Molecular Cell Biology*. W. H. Freeman and Company, 2004.
3. Kornberg, A. and T.A. Baker, *DNA Replication*. University Science Books. Vol. 2nd edition. 2005.
4. Alberts, B., *et al.*, *Molecular biology of the cell*. Garland Science, 2008. **5th Edition**.
5. Kornberg, T. and M.L. Gefter, *Purification and DNA synthesis in cell-free extracts: properties of DNA polymerase II*. *Proc Natl Acad Sci U S A*, 1971. **68**(4): p. 761-4.
6. McHenry, C.S. and W. Crow, *DNA polymerase III of Escherichia coli. Purification and identification of subunits*. *J Biol Chem*, 1979. **254**(5): p. 1748-53.
7. Taft-Benz, S.A. and R.M. Schaaper, *The theta subunit of Escherichia coli DNA polymerase III: a role in stabilizing the epsilon proofreading subunit*. *J Bacteriol*, 2004. **186**(9): p. 2774-80.
8. Studwell, P.S. and M. O'Donnell, *Processive replication is contingent on the exonuclease subunit of DNA polymerase III holoenzyme*. *J Biol Chem*, 1990. **265**(2): p. 1171-8.
9. Brenowitz, S., *et al.*, *Specificity and enzymatic mechanism of the editing exonuclease of Escherichia coli DNA polymerase III*. *J Biol Chem*, 1991. **266**(12): p. 7888-92.
10. Miller, H. and F.W. Perrino, *Kinetic mechanism of the 3'-->5' proofreading exonuclease of DNA polymerase III. Analysis by steady state and pre-steady state methods*. *Biochemistry*, 1996. **35**(39): p. 12919-25.
11. Johnson, A. and M. O'Donnell, *Cellular DNA replicases: components and dynamics at the replication fork*. *Annu Rev Biochem*, 2005. **74**: p. 283-315.
12. O'Donnell, M., *et al.*, *The sliding clamp of DNA polymerase III holoenzyme encircles DNA*. *Mol Biol Cell*, 1992. **3**(9): p. 953-7.
13. Stukenberg, P.T., J. Turner, and M. O'Donnell, *An explanation for lagging strand replication: polymerase hopping among DNA sliding clamps*. *Cell*, 1994. **78**(5): p. 877-87.

14. Yao, N., *et al.*, *Clamp loading, unloading and intrinsic stability of the PCNA, beta and gp45 sliding clamps of human, E. coli and T4 replicases*. *Genes Cells*, 1996. **1**(1): p. 101-13.
15. Kong, X.P., *et al.*, *Three-dimensional structure of the beta subunit of E. coli DNA polymerase III holoenzyme: a sliding DNA clamp*. *Cell*, 1992. **69**(3): p. 425-37.
16. Naktinis, V., *et al.*, *Assembly of a chromosomal replication machine: two DNA polymerases, a clamp loader, and sliding clamps in one holoenzyme particle. II. Intermediate complex between the clamp loader and its clamp*. *J Biol Chem*, 1995. **270**(22): p. 13358-65.
17. Naktinis, V., J. Turner, and M. O'Donnell, *A molecular switch in a replication machine defined by an internal competition for protein rings*. *Cell*, 1996. **84**(1): p. 137-45.
18. Flower, A.M. and C.S. McHenry, *The gamma subunit of DNA polymerase III holoenzyme of Escherichia coli is produced by ribosomal frameshifting*. *Proc Natl Acad Sci U S A*, 1990. **87**(10): p. 3713-7.
19. Tsuchihashi, Z. and A. Kornberg, *Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme*. *Proc Natl Acad Sci U S A*, 1990. **87**(7): p. 2516-20.
20. Gao, D. and C.S. McHenry, *tau binds and organizes Escherichia coli replication through distinct domains. Partial proteolysis of terminally tagged tau to determine candidate domains and to assign domain V as the alpha binding domain*. *J Biol Chem*, 2001. **276**(6): p. 4433-40.
21. Jeruzalmi, D., M. O'Donnell, and J. Kuriyan, *Crystal structure of the processivity clamp loader gamma (gamma) complex of E. coli DNA polymerase III*. *Cell*, 2001. **106**(4): p. 429-41.
22. Xiao, H., Z. Dong, and M. O'Donnell, *DNA polymerase III accessory proteins. IV. Characterization of chi and psi*. *J Biol Chem*, 1993. **268**(16): p. 11779-84.
23. Kelman, Z., *et al.*, *Devoted to the lagging strand-the subunit of DNA polymerase III holoenzyme contacts SSB to promote processive elongation and sliding clamp assembly*. *Embo J*, 1998. **17**(8): p. 2436-49.
24. Glover, B.P. and C.S. McHenry, *The chi psi subunits of DNA polymerase III holoenzyme bind to single-stranded DNA-binding protein (SSB) and facilitate*

- replication of an SSB-coated template*. J Biol Chem, 1998. **273**(36): p. 23476-84.
25. LeBowitz, J.H. and R. McMacken, *The Escherichia coli dnaB replication protein is a DNA helicase*. J Biol Chem, 1986. **261**(10): p. 4738-48.
26. Arai, N. and A. Kornberg, *Rep protein as a helicase in an active, isolatable replication fork of duplex phi X174 DNA*. J Biol Chem, 1981. **256**(10): p. 5294-8.
27. Masai, H. and K. Arai, *Leading strand synthesis of R1 plasmid replication in vitro is primed by primase alone at a specific site downstream of oriR*. J Biol Chem, 1989. **264**(14): p. 8082-90.
28. Tougu, K. and K.J. Marians, *The interaction between helicase and primase sets the replication fork clock*. J Biol Chem, 1996. **271**(35): p. 21398-405.
29. Zechner, E.L., C.A. Wu, and K.J. Marians, *Coordinated leading- and lagging-strand synthesis at the Escherichia coli DNA replication fork. III. A polymerase-primase interaction governs primer size*. J Biol Chem, 1992. **267**(6): p. 4054-63.
30. Yuzhakov, A., Z. Kelman, and M. O'Donnell, *Trading places on DNA--a three-point switch underlies primer handoff from primase to the replicative DNA polymerase*. Cell, 1999. **96**(1): p. 153-63.
31. Li, X. and K.J. Marians, *Two distinct triggers for cycling of the lagging strand polymerase at the replication fork*. J Biol Chem, 2000. **275**(44): p. 34757-65.
32. Okazaki, R., M. Arisawa, and A. Sugino, *Slow joining of newly replicated DNA chains in DNA polymerase I-deficient Escherichia coli mutants*. Proc Natl Acad Sci U S A, 1971. **68**(12): p. 2954-7.
33. Lehman, I.R., *DNA ligase: structure, mechanism, and function*. Science, 1974. **186**(4166): p. 790-7.
34. Shcherbakova, P.V. and Y.I. Pavlov, *3'-->5' exonucleases of DNA polymerases epsilon and delta correct base analog induced DNA replication errors on opposite DNA strands in Saccharomyces cerevisiae*. Genetics, 1996. **142**(3): p. 717-26.
35. Pavlov, Y.I., et al., *Evidence that errors made by DNA polymerase alpha are corrected by DNA polymerase delta*. Curr Biol, 2006. **16**(2): p. 202-7.

36. Niimi, A., *et al.*, *Palm mutants in DNA polymerases alpha and eta alter DNA replication fidelity and translesion activity*. Mol Cell Biol, 2004. **24**(7): p. 2734-46.
37. Pursell, Z.F., *et al.*, *Yeast DNA polymerase epsilon participates in leading-strand DNA replication*. Science, 2007. **317**(5834): p. 127-30.
38. Nick McElhinny, S.A., *et al.*, *Division of labor at the eukaryotic replication fork*. Mol Cell, 2008. **30**(2): p. 137-44.
39. Pavlov, Y.I., C.S. Newlon, and T.A. Kunkel, *Yeast origins establish a strand bias for replicational mutagenesis*. Mol Cell, 2002. **10**(1): p. 207-13.
40. Lehman, I.R. and L.S. Kaguni, *DNA polymerase alpha*. J Biol Chem, 1989. **264**(8): p. 4265-8.
41. Francesconi, S., *et al.*, *Mutations in conserved yeast DNA primase domains impair DNA replication in vivo*. Proc Natl Acad Sci U S A, 1991. **88**(9): p. 3877-81.
42. Pavlov, Y.I., P.V. Shcherbakova, and I.B. Rogozin, *Roles of DNA polymerases in replication, repair, and recombination in eukaryotes*. Int Rev Cytol, 2006. **255**: p. 41-132.
43. Brooke, R.G. and L.B. Dumas, *Reconstitution of the Saccharomyces cerevisiae DNA primase-DNA polymerase protein complex in vitro. The 86-kDa subunit facilitates but is not required for complex formation*. J Biol Chem, 1991. **266**(16): p. 10093-8.
44. Muzi-Falconi, M., *et al.*, *The DNA polymerase alpha-primase complex: multiple functions and interactions*. ScientificWorldJournal, 2003. **3**: p. 21-33.
45. Badaracco, G., *et al.*, *Initiation, elongation and pausing of in vitro DNA synthesis catalyzed by immunopurified yeast DNA primase: DNA polymerase complex*. Embo J, 1985. **4**(5): p. 1313-7.
46. Huang, M.E., *et al.*, *The Saccharomyces cerevisiae protein YJR043C (Pol32) interacts with the catalytic subunit of DNA polymerase alpha and is required for cell cycle progression in G2/M*. Mol Gen Genet, 1999. **260**(6): p. 541-50.
47. Gerik, K.J., *et al.*, *Characterization of the two small subunits of Saccharomyces cerevisiae DNA polymerase delta*. J Biol Chem, 1998. **273**(31): p. 19747-55.

48. Burgers, P.M. and K.J. Gerik, *Structure and processivity of two forms of Saccharomyces cerevisiae DNA polymerase delta*. J Biol Chem, 1998. **273**(31): p. 19756-62.
49. Zuo, S., et al., *Structure and activity associated with multiple forms of Schizosaccharomyces pombe DNA polymerase delta*. J Biol Chem, 2000. **275**(7): p. 5153-62.
50. Podust, V.N., et al., *Reconstitution of human DNA polymerase delta using recombinant baculoviruses: the p12 subunit potentiates DNA polymerizing activity of the four-subunit enzyme*. J Biol Chem, 2002. **277**(6): p. 3894-901.
51. Li, H., et al., *Functional roles of p12, the fourth subunit of human DNA polymerase delta*. J Biol Chem, 2006. **281**(21): p. 14748-55.
52. Liu, L., et al., *Identification of a fourth subunit of mammalian DNA polymerase delta*. J Biol Chem, 2000. **275**(25): p. 18739-44.
53. Dua, R., et al., *Subunit interactions within the Saccharomyces cerevisiae DNA polymerase epsilon (pol epsilon) complex. Demonstration of a dimeric pol epsilon*. J Biol Chem, 2000. **275**(37): p. 28816-25.
54. Chilkova, O., B.H. Jonsson, and E. Johansson, *The quaternary structure of DNA polymerase epsilon from Saccharomyces cerevisiae*. J Biol Chem, 2003. **278**(16): p. 14082-6.
55. Dua, R., et al., *In vivo reconstitution of Saccharomyces cerevisiae DNA polymerase epsilon in insect cells. Purification and characterization*. J Biol Chem, 2002. **277**(10): p. 7889-96.
56. Miyachi, K., M.J. Fritzler, and E.M. Tan, *Autoantibody to a nuclear antigen in proliferating cells*. J Immunol, 1978. **121**(6): p. 2228-34.
57. Krishna, T.S., et al., *Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA*. Cell, 1994. **79**(7): p. 1233-43.
58. Moldovan, G.L., B. Pfander, and S. Jentsch, *PCNA, the maestro of the replication fork*. Cell, 2007. **129**(4): p. 665-79.
59. Fairman, M., et al., *Identification of cellular components required for SV40 DNA replication in vitro*. Biochim Biophys Acta, 1988. **951**(2-3): p. 382-7.
60. Cullmann, G., et al., *Characterization of the five replication factor C genes of Saccharomyces cerevisiae*. Mol Cell Biol, 1995. **15**(9): p. 4661-71.

61. Gomes, X.V. and P.M. Burgers, *ATP utilization by yeast replication factor C. I. ATP-mediated interaction with DNA and with proliferating cell nuclear antigen.* J Biol Chem, 2001. **276**(37): p. 34768-75.
62. Bowman, G.D., M. O'Donnell, and J. Kuriyan, *Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex.* Nature, 2004. **429**(6993): p. 724-30.
63. Forsburg, S.L., *Eukaryotic MCM proteins: beyond replication initiation.* Microbiol Mol Biol Rev, 2004. **68**(1): p. 109-31.
64. Labib, K. and J.F. Diffley, *Is the MCM2-7 complex the eukaryotic DNA replication fork helicase?* Curr Opin Genet Dev, 2001. **11**(1): p. 64-70.
65. Masuda, T., S. Mimura, and H. Takisawa, *CDK- and Cdc45-dependent priming of the MCM complex on chromatin during S-phase in Xenopus egg extracts: possible activation of MCM helicase by association with Cdc45.* Genes Cells, 2003. **8**(2): p. 145-61.
66. Claycomb, J.M., *et al.*, *Visualization of replication initiation and elongation in Drosophila.* J Cell Biol, 2002. **159**(2): p. 225-36.
67. Ishimi, Y., *A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex.* J Biol Chem, 1997. **272**(39): p. 24508-13.
68. Lee, J.K. and J. Hurwitz, *Processive DNA helicase activity of the minichromosome maintenance proteins 4, 6, and 7 complex requires forked DNA structures.* Proc Natl Acad Sci U S A, 2001. **98**(1): p. 54-9.
69. Schwacha, A. and S.P. Bell, *Interactions between two catalytically distinct MCM subgroups are essential for coordinated ATP hydrolysis and DNA replication.* Mol Cell, 2001. **8**(5): p. 1093-104.
70. Bochman, M.L., S.P. Bell, and A. Schwacha, *Subunit organization of Mcm2-7 and the unequal role of active sites in ATP hydrolysis and viability.* Mol Cell Biol, 2008. **28**(19): p. 5865-73.
71. Bauerschmidt, C., *et al.*, *Interactions of human Cdc45 with the Mcm2-7 complex, the GINS complex, and DNA polymerases delta and epsilon during S phase.* Genes Cells, 2007. **12**(6): p. 745-58.
72. Kesti, T., *et al.*, *DNA polymerase epsilon catalytic domains are dispensable for DNA replication, DNA repair, and cell viability.* Mol Cell, 1999. **3**(5): p. 679-85.
73. Burgers, P.M., *Polymerase dynamics at the eukaryotic DNA replication fork.* J Biol Chem, 2008.

74. Kunkel, T.A. and P.M. Burgers, *Dividing the workload at a eukaryotic replication fork*. Trends Cell Biol, 2008. **18**(11): p. 521-7.
75. Iftode, C., Y. Daniely, and J.A. Borowiec, *Replication protein A (RPA): the eukaryotic SSB*. Crit Rev Biochem Mol Biol, 1999. **34**(3): p. 141-80.
76. Garg, P., et al., *Idling by DNA polymerase delta maintains a ligatable nick during lagging-strand DNA replication*. Genes Dev, 2004. **18**(22): p. 2764-73.
77. Jin, Y.H., et al., *The multiple biological roles of the 3'-->5' exonuclease of Saccharomyces cerevisiae DNA polymerase delta require switching between the polymerase and exonuclease domains*. Mol Cell Biol, 2005. **25**(1): p. 461-71.
78. Jin, Y.H., et al., *Okazaki fragment maturation in yeast. II. Cooperation between the polymerase and 3'-5'-exonuclease activities of Pol delta in the creation of a ligatable nick*. J Biol Chem, 2003. **278**(3): p. 1626-33.
79. Gordenin, D.A., T.A. Kunkel, and M.A. Resnick, *Repeat expansion--all in a flap?* Nat Genet, 1997. **16**(2): p. 116-8.
80. Tishkoff, D.X., et al., *A novel mutation avoidance mechanism dependent on S. cerevisiae RAD27 is distinct from DNA mismatch repair*. Cell, 1997. **88**(2): p. 253-63.
81. Kao, H.I., et al., *On the roles of Saccharomyces cerevisiae Dna2p and Flap endonuclease 1 in Okazaki fragment processing*. J Biol Chem, 2004. **279**(15): p. 15014-24.
82. Bae, S.H., et al., *RPA governs endonuclease switching during processing of Okazaki fragments in eukaryotes*. Nature, 2001. **412**(6845): p. 456-61.
83. Jiricny, J., *The multifaceted mismatch-repair system*. Nat Rev Mol Cell Biol, 2006. **7**(5): p. 335-46.
84. Duckett, D.R., et al., *Human MutSalpha recognizes damaged DNA base pairs containing O6-methylguanine, O4-methylthymine, or the cisplatin-d(GpG) adduct*. Proc Natl Acad Sci U S A, 1996. **93**(13): p. 6443-7.
85. Karran, P. and M.G. Marinus, *Mismatch correction at O6-methylguanine residues in E. coli DNA*. Nature, 1982. **296**(5860): p. 868-9.
86. Mazurek, A., M. Berardini, and R. Fishel, *Activation of human MutS homologs by 8-oxo-guanine DNA damage*. J Biol Chem, 2002. **277**(10): p. 8260-6.

87. Mello, J.A., *et al.*, *The mismatch-repair protein hMSH2 binds selectively to DNA adducts of the anticancer drug cisplatin*. *Chem Biol*, 1996. **3**(7): p. 579-89.
88. Mu, D., *et al.*, *Recognition and repair of compound DNA lesions (base damage and mismatch) by human mismatch repair and excision repair systems*. *Mol Cell Biol*, 1997. **17**(2): p. 760-9.
89. Lahue, R.S., K.G. Au, and P. Modrich, *DNA mismatch correction in a defined system*. *Science*, 1989. **245**(4914): p. 160-4.
90. Grilley, M., J. Griffith, and P. Modrich, *Bidirectional excision in methyl-directed mismatch repair*. *J Biol Chem*, 1993. **268**(16): p. 11830-7.
91. Cooper, D.L., R.S. Lahue, and P. Modrich, *Methyl-directed mismatch repair is bidirectional*. *J Biol Chem*, 1993. **268**(16): p. 11823-9.
92. Su, S.S. and P. Modrich, *Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs*. *Proc Natl Acad Sci U S A*, 1986. **83**(14): p. 5057-61.
93. Au, K.G., K. Welsh, and P. Modrich, *Initiation of methyl-directed mismatch repair*. *J Biol Chem*, 1992. **267**(17): p. 12142-8.
94. Welsh, K.M., *et al.*, *Isolation and characterization of the Escherichia coli mutH gene product*. *J Biol Chem*, 1987. **262**(32): p. 15624-9.
95. Dao, V. and P. Modrich, *Mismatch-, MutS-, MutL-, and helicase II-dependent unwinding from the single-strand break of an incised heteroduplex*. *J Biol Chem*, 1998. **273**(15): p. 9202-7.
96. Holmes, J., Jr., S. Clark, and P. Modrich, *Strand-specific mismatch correction in nuclear extracts of human and Drosophila melanogaster cell lines*. *Proc Natl Acad Sci U S A*, 1990. **87**(15): p. 5837-41.
97. Thomas, D.C., J.D. Roberts, and T.A. Kunkel, *Heteroduplex repair in extracts of human HeLa cells*. *J Biol Chem*, 1991. **266**(6): p. 3744-51.
98. Viswanathan, M., *et al.*, *Redundant exonuclease involvement in Escherichia coli methyl-directed mismatch repair*. *J Biol Chem*, 2001. **276**(33): p. 31053-8.
99. Burdett, V., *et al.*, *In vivo requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair*. *Proc Natl Acad Sci U S A*, 2001. **98**(12): p. 6765-70.
100. Iyer, R.R., *et al.*, *DNA mismatch repair: functions and mechanisms*. *Chem Rev*, 2006. **106**(2): p. 302-23.

101. Fishel, R. and R.D. Kolodner, *Identification of mismatch repair genes and their role in the development of cancer*. *Curr Opin Genet Dev*, 1995. **5**(3): p. 382-95.
102. Marti, T.M., C. Kunz, and O. Fleck, *DNA mismatch repair and mutation avoidance pathways*. *J Cell Physiol*, 2002. **191**(1): p. 28-41.
103. Kramer, W., et al., *Cloning and nucleotide sequence of DNA mismatch repair gene PMS1 from Saccharomyces cerevisiae: homology of PMS1 to procaryotic MutL and HexB*. *J Bacteriol*, 1989. **171**(10): p. 5339-46.
104. Alani, E., *The Saccharomyces cerevisiae Msh2 and Msh6 proteins form a complex that specifically binds to duplex oligonucleotides containing mismatched DNA base pairs*. *Mol Cell Biol*, 1996. **16**(10): p. 5604-15.
105. Habraken, Y., et al., *Binding of insertion/deletion DNA mismatches by the heterodimer of yeast mismatch repair proteins MSH2 and MSH3*. *Curr Biol*, 1996. **6**(9): p. 1185-7.
106. Iaccarino, I., et al., *MSH6, a Saccharomyces cerevisiae protein that binds to mismatches as a heterodimer with MSH2*. *Curr Biol*, 1996. **6**(4): p. 484-6.
107. Marsischky, G.T., et al., *Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair*. *Genes Dev*, 1996. **10**(4): p. 407-20.
108. Reenan, R.A. and R.D. Kolodner, *Characterization of insertion mutations in the Saccharomyces cerevisiae MSH1 and MSH2 genes: evidence for separate mitochondrial and nuclear functions*. *Genetics*, 1992. **132**(4): p. 975-85.
109. Culligan, K.M., et al., *Evolutionary origin, diversification and specialization of eukaryotic MutS homolog mismatch repair proteins*. *Nucleic Acids Res*, 2000. **28**(2): p. 463-71.
110. Hollingsworth, N.M., L. Ponte, and C. Halsey, *MSH5, a novel MutS homolog, facilitates meiotic reciprocal recombination between homologs in Saccharomyces cerevisiae but not mismatch repair*. *Genes Dev*, 1995. **9**(14): p. 1728-39.
111. Ross-Macdonald, P. and G.S. Roeder, *Mutation of a meiosis-specific MutS homolog decreases crossing over but not mismatch correction*. *Cell*, 1994. **79**(6): p. 1069-80.

112. Prolla, T.A., D.M. Christie, and R.M. Liskay, *Dual requirement in yeast DNA mismatch repair for MLH1 and PMS1, two homologs of the bacterial mutL gene*. Mol Cell Biol, 1994. **14**(1): p. 407-15.
113. Flores-Rozas, H. and R.D. Kolodner, *The Saccharomyces cerevisiae MLH3 gene functions in MSH3-dependent suppression of frameshift mutations*. Proc Natl Acad Sci U S A, 1998. **95**(21): p. 12404-9.
114. Wang, T.F., N. Kleckner, and N. Hunter, *Functional specificity of MutL homologs in yeast: evidence for three Mlh1-based heterocomplexes with distinct roles during meiosis in recombination and mismatch correction*. Proc Natl Acad Sci U S A, 1999. **96**(24): p. 13914-9.
115. Drummond, J.T., et al., *Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells*. Science, 1995. **268**(5219): p. 1909-12.
116. Genschel, J., et al., *Isolation of MutSbeta from human cells and comparison of the mismatch repair specificities of MutSbeta and MutSalpha*. J Biol Chem, 1998. **273**(31): p. 19895-901.
117. Palombo, F., et al., *hMutSbeta, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA*. Curr Biol, 1996. **6**(9): p. 1181-4.
118. Marra, G., et al., *Mismatch repair deficiency associated with overexpression of the MSH3 gene*. Proc Natl Acad Sci U S A, 1998. **95**(15): p. 8568-73.
119. Li, G.M. and P. Modrich, *Restoration of mismatch repair to nuclear extracts of H6 colorectal tumor cells by a heterodimer of human MutL homologs*. Proc Natl Acad Sci U S A, 1995. **92**(6): p. 1950-4.
120. Raschle, M., et al., *Identification of hMutLbeta, a heterodimer of hMLH1 and hPMS1*. J Biol Chem, 1999. **274**(45): p. 32368-75.
121. Cannavo, E., et al., *Expression of the MutL homologue hMLH3 in human cells and its role in DNA mismatch repair*. Cancer Res, 2005. **65**(23): p. 10759-66.
122. Constantin, N., et al., *Human mismatch repair: reconstitution of a nick-directed bidirectional reaction*. J Biol Chem, 2005. **280**(48): p. 39752-61.
123. Zhang, Y., et al., *Reconstitution of 5'-directed human mismatch repair in a purified system*. Cell, 2005. **122**(5): p. 693-705.
124. Genschel, J., L.R. Bazemore, and P. Modrich, *Human exonuclease I is required for 5' and 3' mismatch repair*. J Biol Chem, 2002. **277**(15): p. 13302-11.

125. Genschel, J. and P. Modrich, *Mechanism of 5'-directed excision in human mismatch repair*. Mol Cell, 2003. **12**(5): p. 1077-86.
126. Yuan, F., et al., *Evidence for involvement of HMGB1 protein in human DNA mismatch repair*. J Biol Chem, 2004. **279**(20): p. 20935-40.
127. Kadyrov, F.A., et al., *Endonucleolytic function of MutLalpha in human mismatch repair*. Cell, 2006. **126**(2): p. 297-308.
128. Wei, K., et al., *Inactivation of Exonuclease 1 in mice results in DNA mismatch repair defects, increased cancer susceptibility, and male and female sterility*. Genes Dev, 2003. **17**(5): p. 603-14.
129. Kadyrov, F.A., et al., *Saccharomyces cerevisiae MutLalpha is a mismatch repair endonuclease*. J Biol Chem, 2007. **282**(51): p. 37181-90.
130. Kunkel, T.A. and K. Bebenek, *DNA replication fidelity*. Annu Rev Biochem, 2000. **69**: p. 497-529.
131. Drake, J.W., *A constant rate of spontaneous mutation in DNA-based microbes*. Proc Natl Acad Sci U S A, 1991. **88**(16): p. 7160-4.
132. Drake, J.W., *The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes*. Ann N Y Acad Sci, 1999. **870**: p. 100-7.
133. Kunkel, T.A., et al., *Fidelity of DNA polymerase I and the DNA polymerase I-DNA primase complex from Saccharomyces cerevisiae*. Mol Cell Biol, 1989. **9**(10): p. 4447-58.
134. Nick McElhinny, S.A., et al., *Inefficient proofreading and biased error rates during inaccurate DNA synthesis by a mutant derivative of Saccharomyces cerevisiae DNA polymerase delta*. J Biol Chem, 2007. **282**(4): p. 2324-32.
135. Kunkel, T.A., *DNA replication fidelity*. J Biol Chem, 2004. **279**(17): p. 16895-8.
136. Kroutil, L.C., et al., *Exonucleolytic proofreading during replication of repetitive DNA*. Biochemistry, 1996. **35**(3): p. 1046-53.
137. Fortune, J.M., et al., *Saccharomyces cerevisiae DNA polymerase delta: high fidelity for base substitutions but lower fidelity for single- and multi-base deletions*. J Biol Chem, 2005. **280**(33): p. 29980-7.
138. Hubscher, U., G. Maga, and S. Spadari, *Eukaryotic DNA polymerases*. Annu Rev Biochem, 2002. **71**: p. 133-63.
139. Lang, G.I. and A.W. Murray, *Estimating the per-base-pair mutation rate in the yeast Saccharomyces cerevisiae*. Genetics, 2008. **178**(1): p. 67-82.

140. Tran, H.T., *et al.*, *Hypermutable of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants*. *Mol Cell Biol*, 1997. **17**(5): p. 2859-65.
141. Harrington, J.M. and R.D. Kolodner, *Saccharomyces cerevisiae Msh2-Msh3 acts in repair of base-base mispairs*. *Mol Cell Biol*, 2007. **27**(18): p. 6546-54.
142. Weber, J.L. and P.E. May, *Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction*. *Am J Hum Genet*, 1989. **44**(3): p. 388-96.
143. Hamada, H., M.G. Petrino, and T. Kakunaga, *A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes*. *Proc Natl Acad Sci U S A*, 1982. **79**(21): p. 6465-9.
144. Strand, M., *et al.*, *Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair*. *Nature*, 1993. **365**(6443): p. 274-6.
145. Parsons, R., *et al.*, *Hypermutable and mismatch repair deficiency in RER+ tumor cells*. *Cell*, 1993. **75**(6): p. 1227-36.
146. Umar, A., *et al.*, *Defective mismatch repair in extracts of colorectal and endometrial cancer cell lines exhibiting microsatellite instability*. *J Biol Chem*, 1994. **269**(20): p. 14367-70.
147. Dietmaier, W., *et al.*, *Diagnostic microsatellite instability: definition and correlation with mismatch repair protein expression*. *Cancer Res*, 1997. **57**(21): p. 4749-56.
148. Malkhosyan, S., *et al.*, *Differences in the spectrum of spontaneous mutations in the hprt gene between tumor cells of the microsatellite mutator phenotype*. *Mutat Res*, 1996. **316**(5-6): p. 249-59.
149. Glaab, W.E. and K.R. Tindall, *Mutation rate at the hprt locus in human cancer cell lines with specific mismatch repair-gene defects*. *Carcinogenesis*, 1997. **18**(1): p. 1-8.
150. Cejka, P., *et al.*, *Methylation-induced G(2)/M arrest requires a full complement of the mismatch repair protein hMLH1*. *Embo J*, 2003. **22**(9): p. 2245-54.
151. Blasi, M.F., *et al.*, *A human cell-based assay to evaluate the effects of alterations in the MLH1 mismatch repair gene*. *Cancer Res*, 2006. **66**(18): p. 9036-44.
152. Liu, B., *et al.*, *hMSH2 mutations in hereditary nonpolyposis colorectal cancer kindreds*. *Cancer Res*, 1994. **54**(17): p. 4590-4.

153. Hoang, J.M., *et al.*, *BAT-26, an indicator of the replication error phenotype in colorectal cancers and cell lines*. *Cancer Res*, 1997. **57**(2): p. 300-3.
154. Fijalkowska, I.J. and R.M. Schaaper, *Mutants in the Exo I motif of Escherichia coli dnaQ: defective proofreading and inviability due to error catastrophe*. *Proc Natl Acad Sci U S A*, 1996. **93**(7): p. 2856-61.
155. Schaaper, R.M. and M. Radman, *The extreme mutator effect of Escherichia coli mutD5 results from saturation of mismatch repair by excessive DNA replication errors*. *Embo J*, 1989. **8**(11): p. 3511-6.
156. Morrison, A. and A. Sugino, *The 3'-->5' exonucleases of both DNA polymerases delta and epsilon participate in correcting errors of DNA replication in Saccharomyces cerevisiae*. *Mol Gen Genet*, 1994. **242**(3): p. 289-96.
157. Morrison, A., *et al.*, *Pathway correcting DNA replication errors in Saccharomyces cerevisiae*. *Embo J*, 1993. **12**(4): p. 1467-73.
158. Goldsby, R.E., *et al.*, *High incidence of epithelial cancers in mice deficient for DNA polymerase delta proofreading*. *Proc Natl Acad Sci U S A*, 2002. **99**(24): p. 15560-5.
159. Li, L., *et al.*, *Sensitivity to phosphonoacetic acid: a new phenotype to probe DNA polymerase delta in Saccharomyces cerevisiae*. *Genetics*, 2005. **170**(2): p. 569-80.
160. Venkatesan, R.N., *et al.*, *Mutator phenotypes caused by substitution at a conserved motif A residue in eukaryotic DNA polymerase delta*. *J Biol Chem*, 2006. **281**(7): p. 4486-94.
161. Loeb, L.A., C.F. Springgate, and N. Battula, *Errors in DNA replication as a basis of malignant changes*. *Cancer Res*, 1974. **34**(9): p. 2311-21.
162. Renan, M.J., *How many mutations are required for tumorigenesis? Implications from human cancer data*. *Mol Carcinog*, 1993. **7**(3): p. 139-46.
163. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer*. *Cell*, 2000. **100**(1): p. 57-70.
164. Cahill, D.P., *et al.*, *Genetic instability and darwinian selection in tumours*. *Trends Cell Biol*, 1999. **9**(12): p. M57-60.
165. Venkatesan, R.N., J.H. Bielas, and L.A. Loeb, *Generation of mutator mutants during carcinogenesis*. *DNA Repair (Amst)*, 2006. **5**(3): p. 294-302.

166. Rajagopalan, H. and C. Lengauer, *Aneuploidy and cancer*. Nature, 2004. **432**(7015): p. 338-41.
167. Lengauer, C., K.W. Kinzler, and B. Vogelstein, *Genetic instability in colorectal cancers*. Nature, 1997. **386**(6625): p. 623-7.
168. Lengauer, C., K.W. Kinzler, and B. Vogelstein, *Genetic instabilities in human cancers*. Nature, 1998. **396**(6712): p. 643-9.
169. Grady, W.M. and J.M. Carethers, *Genomic and epigenetic instability in colorectal cancer pathogenesis*. Gastroenterology, 2008. **135**(4): p. 1079-99.
170. Herrmann, J.L., et al., *Implications of oncogenomics for cancer research and clinical oncology*. Cancer J, 2001. **7**(1): p. 40-51.
171. Levine, A.J., *p53, the cellular gatekeeper for growth and division*. Cell, 1997. **88**(3): p. 323-31.
172. Hussain, S.P., M.H. Hollstein, and C.C. Harris, *p53 tumor suppressor gene: at the crossroads of molecular carcinogenesis, molecular epidemiology, and human risk assessment*. Ann N Y Acad Sci, 2000. **919**: p. 79-85.
173. Bielas, J.H., et al., *Human cancers express a mutator phenotype*. Proc Natl Acad Sci U S A, 2006. **103**(48): p. 18238-42.
174. Loeb, L.A., J.H. Bielas, and R.A. Beckman, *Cancers exhibit a mutator phenotype: clinical implications*. Cancer Res, 2008. **68**(10): p. 3551-7; discussion 3557.
175. Popanda, O., et al., *A mutation detected in DNA polymerase delta cDNA from Novikoff hepatoma cells correlates with abnormal catalytic properties of the enzyme*. J Cancer Res Clin Oncol, 1999. **125**(11): p. 598-608.
176. Goldsby, R.E., et al., *Defective DNA polymerase-delta proofreading causes cancer susceptibility in mice*. Nat Med, 2001. **7**(6): p. 638-9.
177. Flohr, T., et al., *Detection of mutations in the DNA polymerase delta gene of human sporadic colorectal cancers and colon cancer cell lines*. Int J Cancer, 1999. **80**(6): p. 919-29.
178. Starcevic, D., S. Dalal, and J.B. Sweasy, *Is there a link between DNA polymerase beta and cancer?* Cell Cycle, 2004. **3**(8): p. 998-1001.
179. Wang, L., et al., *DNA polymerase beta mutations in human colorectal cancer*. Cancer Res, 1992. **52**(17): p. 4824-7.
180. Lang, T., et al., *A DNA polymerase beta mutant from colon cancer cells induces mutations*. Proc Natl Acad Sci U S A, 2004. **101**(16): p. 6074-9.

181. Sweasy, J.B., *et al.*, *Expression of DNA polymerase {beta} cancer-associated variants in mouse cells results in cellular transformation*. Proc Natl Acad Sci U S A, 2005. **102**(40): p. 14350-5.
182. Aaltonen, L.A., *et al.*, *Clues to the pathogenesis of familial colorectal cancer*. Science, 1993. **260**(5109): p. 812-6.
183. Ionov, Y., *et al.*, *Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis*. Nature, 1993. **363**(6429): p. 558-61.
184. Thibodeau, S.N., G. Bren, and D. Schaid, *Microsatellite instability in cancer of the proximal colon*. Science, 1993. **260**(5109): p. 816-9.
185. Lynch, H.T. and A. de la Chapelle, *Hereditary colorectal cancer*. N Engl J Med, 2003. **348**(10): p. 919-32.
186. Aarnio, M., *et al.*, *Cancer risk in mutation carriers of DNA-mismatch-repair genes*. Int J Cancer, 1999. **81**(2): p. 214-8.
187. Peltomaki, P. and H.F. Vasen, *Mutations predisposing to hereditary nonpolyposis colorectal cancer: database and results of a collaborative study. The International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer*. Gastroenterology, 1997. **113**(4): p. 1146-58.
188. Kolodner, R.D., *et al.*, *Germ-line msh6 mutations in colorectal cancer families*. Cancer Res, 1999. **59**(20): p. 5068-74.
189. Lynch, H.T. and A. de la Chapelle, *Genetic susceptibility to non-polyposis colorectal cancer*. J Med Genet, 1999. **36**(11): p. 801-18.
190. Boland, C.R., *et al.*, *A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer*. Cancer Res, 1998. **58**(22): p. 5248-57.
191. Truninger, K., *et al.*, *Immunohistochemical analysis reveals high frequency of PMS2 defects in colorectal cancer*. Gastroenterology, 2005. **128**(5): p. 1160-71.
192. Peltomaki, P., *Deficient DNA mismatch repair: a common etiologic factor for colon cancer*. Hum Mol Genet, 2001. **10**(7): p. 735-40.
193. Edelman, L. and W. Edelman, *Loss of DNA mismatch repair function and cancer predisposition in the mouse: animal models for human hereditary*

- nonpolyposis colorectal cancer*. Am J Med Genet C Semin Med Genet, 2004. **129C**(1): p. 91-9.
194. Schwarz, D.S., et al., *Designing siRNA that distinguish between genes that differ by a single nucleotide*. PLoS Genet, 2006. **2**(9): p. e140.
195. Zeng, X.R., et al., *DNA polymerase delta is involved in the cellular response to UV damage in human cells*. J Biol Chem, 1994. **269**(19): p. 13748-51.
196. Zeng, X.R., et al., *Regulation of human DNA polymerase delta during the cell cycle*. J Biol Chem, 1994. **269**(39): p. 24027-33.
197. Li, B. and M.Y. Lee, *Transcriptional regulation of the human DNA polymerase delta catalytic subunit gene POLD1 by p53 tumor suppressor and Sp1*. J Biol Chem, 2001. **276**(32): p. 29729-39.
198. Zhao, L. and L.S. Chang, *The human POLD1 gene. Identification of an upstream activator sequence, activation by Sp1 and Sp3, and cell cycle regulation*. J Biol Chem, 1997. **272**(8): p. 4869-82.
199. Chang, L.S., et al., *Structure of the gene for the catalytic subunit of human DNA polymerase delta (POLD1)*. Genomics, 1995. **28**(3): p. 411-9.
200. Yao, F., et al., *Tetracycline repressor, tetR, rather than the tetR-mammalian cell transcription factor fusion derivatives, regulates inducible gene expression in mammalian cells*. Hum Gene Ther, 1998. **9**(13): p. 1939-50.
201. Chung, D.W., et al., *Primary structure of the catalytic subunit of human DNA polymerase delta and chromosomal location of the gene*. Proc Natl Acad Sci U S A, 1991. **88**(24): p. 11197-201.
202. Su, T.T., *Cellular responses to DNA damage: one signal, multiple choices*. Annu Rev Genet, 2006. **40**: p. 187-208.
203. Siolas, D., et al., *Synthetic shRNAs as potent RNAi triggers*. Nat Biotechnol, 2005. **23**(2): p. 227-31.
204. Simon, M., L. Giot, and G. Faye, *The 3' to 5' exonuclease activity located in the DNA polymerase delta subunit of Saccharomyces cerevisiae is required for accurate replication*. Embo J, 1991. **10**(8): p. 2165-70.
205. Venkatesan, R.N., et al., *Mutation at the polymerase active site of mouse DNA polymerase delta increases genomic instability and accelerates tumorigenesis*. Mol Cell Biol, 2007. **27**(21): p. 7669-82.
206. Albertini, R.J., *HPRT mutations in humans: biomarkers for mechanistic studies*. Mutat Res, 2001. **489**(1): p. 1-16.

207. Manara, M.C., *et al.*, *Reversal of malignant phenotype in human osteosarcoma cells transduced with the alkaline phosphatase gene*. *Bone*, 2000. **26**(3): p. 215-20.
208. Rodrigues, M.A., *et al.*, *Nucleoplasmic calcium is required for cell proliferation*. *J Biol Chem*, 2007. **282**(23): p. 17061-8.
209. Garcia-Diaz, M. and T.A. Kunkel, *Mechanism of a genetic glissando: structural biology of indel mutations*. *Trends Biochem Sci*, 2006. **31**(4): p. 206-14.
210. *Whither RNAi?* *Nat Cell Biol*, 2003. **5**(6): p. 489-90.
211. Jiang, Y. and D.H. Price, *Rescue of the TTF2 knockdown phenotype with an siRNA-resistant replacement vector*. *Cell Cycle*, 2004. **3**(9): p. 1151-3.
212. Kojima, S., D. Vignjevic, and G.G. Borisy, *Improved silencing vector co-expressing GFP and small hairpin RNA*. *Biotechniques*, 2004. **36**(1): p. 74-9.
213. Szulc, J., *et al.*, *A versatile tool for conditional gene expression and knockdown*. *Nat Methods*, 2006. **3**(2): p. 109-16.
214. Selifonova, O., F. Valle, and V. Schellenberger, *Rapid evolution of novel traits in microorganisms*. *Appl Environ Microbiol*, 2001. **67**(8): p. 3645-9.
215. Shimoda, C., *et al.*, *Isolation of thermotolerant mutants by using proofreading-deficient DNA polymerase delta as an effective mutator in Saccharomyces cerevisiae*. *Genes Genet Syst*, 2006. **81**(6): p. 391-7.
216. Seifert, M., *et al.*, *The DNA-mismatch repair enzyme hMSH2 modulates UV-B-induced cell cycle arrest and apoptosis in melanoma cells*. *J Invest Dermatol*, 2008. **128**(1): p. 203-13.
217. Lin, Y., V. Dion, and J.H. Wilson, *Transcription promotes contraction of CAG repeat tracts in human cells*. *Nat Struct Mol Biol*, 2006. **13**(2): p. 179-80.
218. DuBridge, R.B., *et al.*, *Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system*. *Mol Cell Biol*, 1987. **7**(1): p. 379-87.

9 Acknowledgements

First of all, I would like to thank Prof. Dr. Josef Jiricny for giving me the opportunity to do my PhD in his laboratory, for all his support along the way, fruitful discussions and especially for encouraging me to design a project with a biotechnological focus.

His continuous efforts to provide the best possible environment for successful research indeed make the IMCR a place where everything is possible.

I also would like to thank current and past members of the laboratory and institute for all the support I received during my PhD and for the fantastic atmosphere that allowed me to grow in many ways.

I would like to thank especially the two technicians, Ms. Mariela Artola and Ms. Myriam Marti, who helped with some of the experiments, and Dr. Agnieszka Gembka for her introduction to primer extension assays.

And most importantly, I would like to thank my parents for their continuous support during all these years.

Curriculum Vitae

Patrick Kehl

Personal information

Name: **Patrick Albert Kehl**
Address: **Fasanenstr. 15**
CH - 9435 Heerbrugg, Switzerland
Telephone office: **+41 44 635 34 80**
Mobile phone: **+41 76 369 99 66**
Email: **kehl@imcr.uzh.ch**
Date of Birth: December 24th 1975
Marital Status: unmarried
Nationality: Swiss

Education

2004 – 2009 PhD thesis with Prof. Dr. Josef Jiricny at the Institute of Molecular Cancer Research, University of Zurich: „*Inducible gene replacement of DNA polymerase delta*“

May 2004 Graduation with *Eidg. Dipl. Natw. ETH*

February - August 2003 Diploma thesis at ETH: „*Coupling the biotransformation of toluene with the host metabolism*“

October – December 2002 Research stage at the Division of Chemical Engineering, University of Queensland, Brisbane, Australia: „*Metabolic engineering of hyaluronic acid production in Streptococcus zooepidemicus*“

1998 – 2004 Studies in biochemistry, genetics, biotechnology and bioinformatics at the Federal Institute of Technology Zurich (ETH), Switzerland

1983 – 1997 Primary school and grammar school in Widnau (Switzerland), secondary school and Matura in Heerbrugg (Switzerland)

Reference

Prof. Dr. Josef Jiricny
Professor and Chair
Institute of Molecular Cancer Research
University of Zurich
Winterthurerstr. 190
CH - 8057 Zürich, Switzerland
Telephone +41 44 635 34 50
Email jiricny@imcr.uzh.ch